# **UC San Diego**

# **UC San Diego Electronic Theses and Dissertations**

## **Title**

Structural and functional studies of the Drosophila melanogaster snRNA activating protein complex (DmSNAPc)

## **Permalink**

https://escholarship.org/uc/item/3w37c14s

### **Author**

Hung, Ko-Hsuan

## **Publication Date**

2011

Peer reviewed|Thesis/dissertation

# UNIVERSITY OF CALIFORNIA, SAN DIEGO SAN DIEGO STATE UNIVERSITY

# Structural and functional studies of the *Drosophila melanogaster* snRNA activating protein complex (DmSNAPc)

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Ko-Hsuan Hung

## Committee in Charge:

University of California, San Diego

Professor Katherine A. Jones Professor James T. Kadonaga

San Diego State University

Professor William E. Stumph, Chair Professor Sanford I. Bernstein Professor Robert W. Zeller

The Dissertation of Ko-Hsuan Hung is approved, and it is acceptable in quality at	nd
form for publication on microfilm and electronically:	
	_
	_
	_
	_
Chair	r

University of California, San Diego San Diego State University

2011

# **DEDICATION**

This dissertation is dedicated to my parents, my wife, and the rest of my family for their incredible support, continuous encouragement, and endless love.

# TABLE OF CONTENTS

Signature Page.	iii
Dedication Page	iv
Table of Contents	v
List of Figures.	V11
Acknowledgement	ix
Vita	Xi
Abstract	xiv
General Introduction	1
A: References	11
Chapter 1	
A map of <i>Drosophila melanogaster</i> small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunits assembly and DNA binding.	16
A: Abstract	
B: Introduction	
C: Material and Methods	
D: Results	
E: Discussion	
F: References	
Chapter 2	
Regulation of snRNA gene expression by the <i>Drosophila melanogaster</i> small nuclear RNA activating protein complex (DmSNAPc)	
A: Abstract	30
B: Introduction	30
C: Main body	31
D. Pafaranaa	12

# Chapter 3

	calization of residues in a novel DNA-binding domain of DmSNAP43 uired for DmSNAPc DNA-binding activity	46
	A: Abstract	47
	B: Introduction	48
	C: Material and Methods	52
	D: Results	54
	E: Discussion	62
	F: References.	63
Concluding	Remarks	
	A: References	72
Appendix		74
A:	Plasmid constructs for expressing various forms of the DmSNAP subunits in S2 cells	75
B:	List of stably transfected S2 cell lines	98
C:	High-salt FLAG purification of FLAG-tagged DmSNAPc from S2 cells	100
D:	Purification of His-tagged DmSNAPc from S2 cells by nickel-chelate chromatography	105
E:	Detailed protocol for electrophoretic mobility shift assay (EMSA)	115
F٠	Detailed protocol for chromatin immunoprecipitation assay (ChIP)	124

# LIST OF FIGURES

# General Introduction

•	Conserved structure of <i>Drosophila</i> snRNA gene promoters transcribed by Pol II and Pol III	5
	Comparison of <i>D. melanogaster</i> (Dm) and human (Hs) SNAP subunits	6
•	Working model for RNA polymerase specificity at snRNA promoters	9
Chapter 1		
_	Comparison of <i>D. melanogaster</i> (Dm) and human (Hs) SNAP subunits	21
	Domains of DmSNAP190 involved in assembly in vivo with DmSNAP50 and DmSNAP43	22
_	Domains of DmSNAP190 required for sequence-specific DNA binding by DmSNAPc	24
	Domains of DmSNAP50 involved in assembly in vivo with DmSNAP190 and DmSNAP43	24
	Domains of DmSNAP50 required for sequence-specific DNA binding by DmSNAPc	25
<u> </u>	Domains of DmSNAP43 involved in assembly in vivo with DmSNAP190 and DmSNAP50	26
	Domains of DmSNAP43 required for sequence-specific DNA binding by DmSNAPc	24
<u> </u>	Effect of DmSNAP mutations on the in vivo occupancy of a U1 snRNA gene promoter by DmSNAPc	25
<u> </u>	Summary of the effects of DmSNAP truncations on DmSNAPc assembly and DNA-binding	26
Chapter 2		
Figure 2.1: F	Promoter structure of snRNA genes	34
	Comparison of <i>D. melanogaster</i> (Dm), human (Hs), and <i>Trypanosoma brucei</i> (Tb) SNAP subunits	38

	Figure 2.3:	DmSNAPc subunit domains required for complex assembly and DNA binding	.40
	Figure 2.4:	Contact points between snRNA gene promoter sequences and DmSNAPc	34
	Figure 2.5:	Domains of DmSNAP50 and DmSNAP43 that cross-link to specific phosphate positions on U1 or U6 DNA probes	38
	Figure 2.6:	A working model illustrated how differential interactions of DmSNAPc with U1 or U6 PSEAs may lead to RNA polymerase specificity at <i>D. melanogaster</i> snRNA gene promoters	40
Chapte	er 3		
	Figure 3.1:	Identification of evolutionarily conserved amino acid residues within a DNA-binding domain of DmSNAP43	51
	Figure 3.2:	Mutations within two clusters of DmSNAP43 conserved residues have no affect on DmSNAP complex formation but result in reduced DmSNAPc DNA-binding activity	56
	Figure 3.3:	Alanine scanning mutagenesis within the DNA-binding domain of DmSNAP43 between residues 193 and 272	59
	Figure 3.4:	ChIP from stably-transfected S2 cell lines that over-express FLAG-tagged WT or mutant DmSNAP43 constructs that contain alanine substitutions	61

#### **ACKNOWLEDGEMENTS**

I would like to thank my thesis advisor Dr. Stumph for his guidance, encouragement, and continuous support during my stay in his lab. Without his mentorship I would not be able to develop into the scientist who I am today, and I will always be grateful for that. I would also like to thank Dr. Bernstein, Dr. Zeller, Dr. Kadonaga and Dr. Jones for serving on my dissertation committee and giving me invaluable suggestions on my projects to make this thesis possible.

I am grateful to all of the past and current members of the Stumph lab for their help and encouragement that facilitated the accomplishment of my projects and made this experience so much fun. I thank Mitchell Titus and Shu-Chi Kiki Chiang for their contributions to the publication that constitutes Chapter 1 of this thesis. I am also grateful to Kathleen McNamara-Schroeder and Cheryn Lee for their assistance in preparation of plasmid constructs used in Chapter 3.

I would like to thank my parents for their enormous support and constant faith in me during this journey. I am also deeply grateful to my wife, Hsiao-Chun Janet Huang, for her endless support and encouragement to allow me to focus on my research and motivate me to overcome all difficulties encountered through this process.

Chapter one, in full, is a reprint of the material as it appears in The Journal of Biological Chemistry, 2009. A map of *Drosophila melanogaster* small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and

DNA binding. Hung, K. H.; Titus, M.; Chiang, S. C.; Stumph, W. E. The dissertation author was the primary researcher and author of this paper.

Chapter two, in full, is a reprint of the material as it appears in Critical Reviews in Biochemistry and Molecular Biology, 2010. Regulation of snRNA genes expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc). Hung, K. H.; Stumph, W. E. The dissertation author was the primary author of this paper.

The material in chapter three, in full, will be submitted for publication immediately following approval of this thesis. The dissertation author was the primary researcher and author of this paper.

# **VITA**

1995-1999	B.S., Biology
	National Cheng Kung University
	Tainan, Taiwan
1999-2001	Military service in Air Force
	Taiwan
2002-2003	Research Assistant
	Institute of Biochemistry, National Defense Medical Center
	Taipei, Taiwan
2003-2004	Research Assistant
	Institute of Clinical Medicine, National Taiwan University
	Taipei, Taiwan
2004-2011	Teaching Associate
	Department of Biology, San Diego State University
	San Diego, California
2009-2011	Pre-doctoral Candidate
	Department of Biology (Cell and Molecular Biology)
	University of California, San Diego and
	San Diego State University
	San Diego, California
2011	Ph.D. in Biology (Cell and Molecular Biology)
	University of California, San Diego and
	San Diego State University
	San Diego, California

#### **PUBLICATIONS**

**Hung KH** and Stumph WE (In preparation) Localization of residues in a novel DNA-binding domain of DmSNAP43 required for DmSNAPc DNA-binding activity.

**Hung KH** and Stumph WE (2010) Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc). Critical Reviews in Biochemistry and Molecular Biology 46:11-26.

**Hung KH**, Titus M, Chiang SC, and Stumph WE (2009) A map of Drosophila melanogaster small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. The Journal of Biological Chemistry 284:22568-22579.

Su WC, Liu WL, Cheng CW, Chou YB, **Hung KH**, Huang WH, Wu CL, Li YT, Shiau AL, and Lai MY (2009) Ribavirin enhances interferon signaling via stimulation of mTOR and p53 activities. FEBS Letters 583:2793-2798.

#### **CONFERENCE ABSTRACTS**

**Hung KH** and Stumph WE (2011) DmSNAP43 residues required for the sequence-specific DNA-binding activity of DmSNAPc. Cold Spring Harbor Laboratory meeting: Mechanisms of Eukaryotic Transcription, Cold Spring Harbor, NY.

**Hung KH**, Titus M, Chiang SC, and Stumph WE (2009) Identification of functional domains involved in subunit assembly and DNA binding with D. melanogaster small nuclear RNA-activating protein complex (DmSNAPc). Cold Spring Harbor Laboratory meeting: Mechanisms of Eukaryotic Transcription, Cold Spring Harbor, NY.

**Hung KH**, Titus M, Chiang SC, and Stumph WE (2008) DmSNAPc domains for subunit assembly and DNA binding. ASBMB annual meeting, San Diego, CA.

**Hung KH**, Titus M, Chiang SC, and Stumph WE (2008) DmSNAPc domains for subunit assembly and DNA binding. Meeting on Transcription Regulation by Chromatin and RNA Polymerase II, Tahoe City, CA.

# FIELDS OF STUDY

Major Field: Cell and Molecular Biology

Studies in Transcriptional Regulation of Eukaryotic Gene Expression Professor William E. Stumph

#### ABSTRACT OF THE DISSERTATION

Structural and functional studies of the *Drosophila melanogaster* snRNA activating protein complex (DmSNAPc)

by

## Ko-Hsuan Hung

Doctor of Philosophy in Biology
University of California, San Diego, 2011
San Diego State University, 2011

Professor William E. Stumph, Chair

The goal of this study is to better understand the structure-function relationships of the small nuclear RNA activating protein complex (SNAPc), and how this complex is involved in transcription activation and RNA polymerase specificity of small nuclear RNA (snRNA) genes. The SNAP complex is the major component uniquely required for transcription of snRNA genes, some of which are transcribed by RNA polymerase II (Pol II) and some by RNA polymerase III (Pol III). In the fruit fly, SNAPc contains three distinct subunits (DmSNAP190, DmSNAP50, and DmSNAP43) that form a complex prior to binding to DNA; moreover, all three subunits are required for the sequence-specific DNA binding activity of DmSNAPc and each makes direct contact with DNA.

Chapter 1 describes truncational analysis that mapped domains within each subunit of DmSNAPc that are involved in complex formation and DNA binding. Our results indicated that the most evolutionarily conserved regions of the subunits were involved in complex assembly. However, domains outside the conserved regions were also important for the DNA binding activity of DmSNAPc, even though they were not required for subunit assembly.

Chapter 2 summarizes our present understanding of how snRNA transcription is regulated in the fruit fly, and further compares this knowledge with information obtained from other systems. The structure of snRNA promoters and the contribution of these promoter sequences to RNA polymerase selection were reviewed followed by a discussion of structure-function features of DmSNAPc in comparison to the homologous proteins from other organisms. Evidence that snRNA promoter sequences act as differential allosteric effectors of DmSNAPc conformation was discussed, and how these conformational differences of the DmSNAPc-DNA complex may lead to distinct RNA polymerase specificities of Pol II and Pol III snRNA genes were proposed.

Chapter 3 describes studies that investigated the contribution made to DmSNAPc DNA-binding activity by amino acid residues within a novel DNA-binding domain of DmSNAP43. My results revealed that some of the most evolutionarily conserved residues within this domain were essential for DNA binding, whereas other residues made little or no contribution to the DNA binding activity of DmSNAPc.

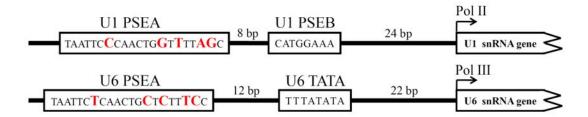
# **GENERAL INTRODUCTION**

### Drosophila snRNA gene promoter structure and RNA polymerase specificity

The small nuclear RNAs (snRNAs) are a highly abundant class of metabolically stable non-polyadenylated RNA molecules first identified over 30 years ago (Busch et al., 1982). The snRNAs are best characterized for their roles in many RNA processing events such as pre-messenger RNA splicing, ribosomal RNA processing, and histone messenger RNA 3' end formation (Steitz et al., 1988; Kass et al., 1990; Bond et al., 1991; Guthrie, 1991; Peculis and Steitz, 1993; Sharp, 1994). However, it was appreciated only recently that some snRNAs are also involved in other essential gene regulatory events such as transcription initiation and transcription elongation (Kwek et al., 2002; Nguyen et al., 2001; Yang et al., 2001). Thus, it is now clear that the snRNAs play important roles at many stages of gene expression, and the accurate and efficient control of the expression of snRNA genes is consequently critical for cell survival.

In animals, most of the snRNA genes (e.g. U1, U2, U3, U4, U5 and U7 genes) are transcribed by RNA polymerase II (Pol II), but others are transcribed by RNA polymerase III (Pol III) (e.g. U6 and 7SK genes) (Zieve et al., 1977; Dahlberg and Lund, 1988; Parry et al., 1989; Hernandez, 1992; Lobo and Hernandez, 1994). Interestingly, despite transcription by different RNA polymerases, the promoter structures of both classes of snRNA genes are remarkably similar. Transcription of animal snRNA genes by either RNA polymerase is dependent upon a proximal sequence element (PSE) located about 40-65 base pairs upstream of the transcription start site (Das et al., 1987; Dahlberg and Lund, 1988; Zamrod et al., 1993; Lobo and Hernandez, 1994). In the fruit fly *Drosophila melanogaster*, the PSE is referred to

more specifically as the PSEA, a unique 21 bp promoter element well-conserved in all D. melanogaster snRNA genes (Fig. G.1) (Jensen et al., 1998; Hernandez et al., 2007). In the promoter of Pol II transcribed snRNA genes (e.g. U1 genes in the figure), the PSEA is located 8 bp upstream of a conserved PSEB, but in the promoter of Pol III transcribed snRNA genes (e.g. U6 genes in the figure), the PSEA is located 12 bp upstream of a TATA box. Interestingly, the RNA polymerase specificity of Drosophila snRNA genes is largely determined by a few nucleotide differences between the PSEAs of the two classes of snRNA genes, not by the PSEB vs. TATA box sequence nor by the spacing difference between the conserved elements (Jensen et al., 1998; McNamara-Schroeder et al., 2001; Lai et al., 2005). The PSEB and TATA elements, as well as the inter-element spacing affect transcription efficiency but do not directly affect the choice of RNA polymerase. Indeed, our previous in vitro and in vivo experiments indicated that the *Drosophila* U1 and U6 PSEAs are not interchangeable: the U1 PSEA cannot function for Pol III transcription, and the U6 PSEA cannot function for Pol II transcription, despite the fact that they bind the same transcription factor DmSNAPc (McNamara-Schroeder et al., 2001; Barakat and Stumph, 2008). Interestingly, our bioinformatic studies suggested that this mechanism is conserved among insects in which PSEA sequence differences are likely responsible for determining RNA polymerase specificity (Hernandez et al., 2007)



**Figure G.1. Conserved structure of** *Drosophila* **snRNA gene promoters transcribed by Pol II and Pol III.** The promoter sequences of the specific U1 and U6 genes used in our studies are shown. The U1 and U6 PSEA sequences differ at only 5 of 21 nucleotide positions (in red).

# Characterization of the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc)

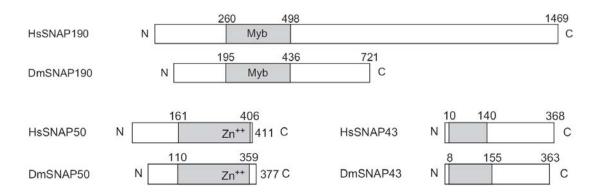
The small nuclear RNA activating protein complex (SNAPc) is a multi-subunit protein which binds to the PSE and is required for transcription of snRNA genes (Waldschmidt et al., 1991; Sadowski et al., 1993; Goomer et al., 1994; Henry et al., 1995; Yoon et al., 1995). In humans, SNAPc contains five distinct polypeptide chains (HsSNAP190, HsSNAP50, HsSNAP45, HsSNAP43, and HsSNAP19) named based upon the apparent molecular weights of these subunits (Henry et al., 1995; Henry et al., 1998; Yoon and Roeder, 1996). However, a complex composed of only the three subunits HsSNAP190, HsSNAP50 and HsSNAP43 was sufficient for PSE binding and activation of snRNA transcription (Mittal et al., 1999; Ma and Hernandez, 2001; Hinkley et al., 2003; Jawdekar et al., 2006). Indeed, evidence indicated that the other two human subunits HsSNAP45 and HsSNAP19 may only play minor regulatory roles in human snRNA transcription (Henry et al., 1998; Mittal et al., 1999; Ma and Hernandez, 2001). Thus, the human SNAP subunits HsSNAP190, HsSNAP50, and

HsSNAP43 represent the "core SNAP subunits" required for human SNAPc function in sequence-specific DNA binding and basal transcription activity.

In the fruit fly, DmSNAPc contains only three polypeptide chains DmSNAP190, DmSNAP50, and DmSNAP43 that are orthologous to the human core SNAP subunits (Wang and Stumph, 1998; Li et al., 2004). No genes encoding proteins homologous to HsSNAP45 and HsSNAP19 could be identified in the *Drosophila* genome. The three fly SNAP subunits tightly associate with each other in solution even when the complex is not bound to DNA (Su et al., 1997). Moreover, each of the three subunits is essential for sequence-specific binding to the PSEA as none can bind individually or in any pair-wise combination (our unpublished observations). Furthermore, site-specific protein-DNA photo-cross-linking experiments demonstrated that each of the SNAP subunits made direct contact with the DNA when DmSNAPc was bound to the PSEA (Wang and Stumph, 1998; Li et al., 2004). However, it was not yet clear how these subunits interact with each other to form the complex or exactly how the individual subunits contribute to the DNA-binding activity of DmSNAPc. These questions are addressed in studies described in Chapter 1.

A comparison of the structural features of the homologous fly and human SNAP subunits are shown in **Fig. G.2.** The most conserved region of SNAP190 contains a unique Myb domain that consists of 4.5 Myb repeats (Wong et al., 1998; Li et al., 2004). The Myb repeats were first identified in the Myb oncoprotein and are involved in DNA binding (Klempnauer and Sippel, 1987; Biedenkapp et al., 1988). Indeed, studies in human indicated that the Myb repeats in HsSNAP190 are essential for the DNA binding activity of human SNAPc (Wong et al., 1998; Ma and

Hernandez, 2002; Hinkley et al., 2003). SNAP50 is the most conserved subunit between human and fly SNAP proteins (Li et al., 2004; Jawdekar et al., 2006). The C-terminal conserved region of SNAP50 contains an unorthodox zinc-binding motif termed the "SNAP finger", which was shown to be crucial for DNA binding in the human system (Jawdekar et al., 2006). SNAP43 is the least-characterized SNAP subunit with no clear homology to any other proteins in existing databases. The most evolutionarily conserved region of SNAP43 resides in the N terminus. Interestingly, SNAP43 does not contain any well-characterized DNA-binding domain, despite the fact that fly SNAP43 directly contacted the DNA when DmSNAPc bound to the PSEA (Wang and Stumph, 1998; Li et al., 2004). The partial characterization of a novel domain of DmSNAP43 required for DNA-binding will be described in Chapter 3.

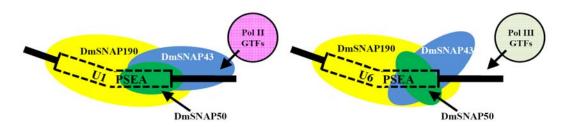


**Figure G.2. Comparison of** *D. melanogaster* (**Dm**) **and human** (**Hs**) **SNAP subunits.** The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous protein pairs. The numbers above the rectangles designate the amino acid positions in the proteins. The most conserved region of DmSNAP190 contains a Myb domain, which consists of 4.5 tandem Myb repeats in fly and human SNAP190. This region of DmSNAP190 shares 27% identity and 44% similarity with HsSNAP190. For DmSNAP50, the most conserved region is located at the C-terminus (residues 110 to 359), which includes an unorthodox zinc-binding domain (noted as Zn<sup>++</sup> in the figure) termed the "SNAP" finger. This region shares 33% identity and 51% similarity with HsSNAP50. The most conserved region of DmSNAP43 resides in the N-terminus (residues 8 to 155), which shares 31% identity and 48% similarity with HsSNAP43.

## Interaction of DmSNAPc with snRNA promoters and RNA polymerase selection

Our lab showed that DmSNAPc binds to the U1 and U6 PSEAs and can activate transcription of the *Drosophila* U1 and U6 snRNA genes in vitro (Su et al., 1997). Moreover, site-specific protein-DNA photo-cross-linking data indicated that the cross-linking pattern of the DmSNAP subunits to the DNA was different depending upon whether DmSNAPc was bound to a U1 or a U6 PSEA (Wang and Stumph, 1998; Li et al., 2004). This result suggested that the conformation of DmSNAPc is different depending upon the DNA sequence bound. Furthermore, our in vitro and in vivo transcription data indicated that these conformational differences are differentially interpreted by the transcription machinery to confer promoter specificity for either Pol II or Pol III (Jensen et al., 1998; McNamara-Schroeder et al., 2001; Lai et al., 2005; Barakat and Stumph, 2008). Based on these findings, we believe that such conformational differences of DmSNAPc lead to the subsequent recruitment of distinct sets of general transcription factors (GTFs) to different classes of snRNA promoters, and eventually recruit different RNA polymerases to start transcription (Fig. G.3). This hypothesis is further supported by our recent finding that the recruitment of the TATA-binding protein (TBP) to the U1 promoter in vivo is disrupted by replacement of the U1 PSEA with a U6 PSEA (Barakat and Stumph, 2008). Chromatin immunoprecipitation (ChIP) assays revealed that substituting a U6 PSEA for the U1 PSEA did not affect the binding of DmSNAPc to the U1 promoter. Instead, it interfered with the recruitment of TBP. This is consistent with our model in

which DmSNAPc binds to the U6 PSEA in a conformation that makes it unable to recruit Pol II GTFs to the U1 promoter.



**Figure G.3. Working model for RNA polymerase specificity at snRNA promoters.** Each subunit of DmSNAPc is drawn schematically to indicate the region of DNA with which it interacts based upon photo-cross-linking data. We propose that the different conformations of DmSNAPc recruit different sets of general transcription factors (GTFs).

# Potential role of DmSNAP43 in Pol II GTFs recruitment and RNA polymerase selection

From the site-specific protein-DNA photo-cross-linking data summarized schematically in **Fig. G.3**, it is apparent that among the three DmSNAP subunits, DmSNAP43 showed the most striking difference in cross-linking patterns when DmSNAPc bound to the different PSEAs: it cross-linked to DNA up to 20 bp downstream of a U1 PSEA, but only 5 bp downstream of a U6 PSEA (Li et al., 2004). The distinct DNA binding pattern of DmSNAP43 on different PSEAs raised the possibility that DmSNAP43 might play a major role in the differential recruitment of transcription machineries to Pol II or Pol III snRNA promoters. The locations where DmSNAP43 contacted the DNA further suggested the possibility of interaction of DmSNAP43 and Pol II GTFs on U1 snRNA promoters. These observations suggested that DmSNAP43 could potentially play a role in recruiting the Pol II general

transcription machinery to U1 snRNA promoters to establish Pol II preinitiation complex.

## Questions investigated by the work described in this dissertation

To understand the structure-function relationships of DmSNAPc, it is necessary to know how each of the subunits interacts with the others and with the snRNA promoter DNA. Chapter 1 describes work toward resolving those questions by utilizing a truncational analysis to map domains of each of the three DmSNAPc subunits that are required for 1) DmSNAPc assembly and 2) DNA binding (Hung et al., 2009). A series of N-terminal and C-terminal truncations of each of the DmSNAP subunits were made and co-expressed with the other two complementary full length subunits in *Drosophila* S2 cells. Because we strongly suspected that any deletion of the most conserved regions would be highly detrimental to DmSNAPc activity, our truncations mainly focused on the non-conserved regions of each subunit, though some constructs truncated in conserved regions were also prepared. The ability of the truncated constructs to form complexes with the complementary subunits was assayed by co-immunoprecipitation assays. In vitro DNA binding activity of the immunopurified truncated DmSNAPc was examined by electrophoretic mobility shift assays (EMSA). Chromatin immunoprecipitation (ChIP) experiments with anti-FLAG antibodies were employed to determine whether the truncated DmSNAPc could be recruited to the endogenous U1:95Ca gene promoter in vivo. Results from this first chapter have been published in the Journal of Biological Chemistry (2009) and the dissertation author is the primary investigator of this research.

Chapter 2, in full, is a reprint of a review article published in Critical Reviews in Biochemistry and Molecular Biology (2011) (Hung and Stumph, 2011). The dissertation author is the primary author of this paper. This article reviewed studies of the regulation of snRNA gene transcription by using the fruit fly as a model organism, and further compared these findings to knowledge gained from studies in other systems. The following subjects were covered and discussed: 1) the structure of Drosophila snRNA promoters and contributions of these promoter sequences to determining RNA polymerase specificity; 2) structure-function features of DmSNAP subunits and a comparison with their orthologs in other organisms; 3) the structure of the DmSNAPc-DNA complex and 4) a current model of how structural change of the DmSNAPc-DNA complex induced by snRNA promoter sequences are interpreted by transcription machineries to establish RNA polymerase specificity. Since this review article was written after the work described in Chapter 1 was performed but before the work described in Chapter 3, the article included and discussed results from Chapter 1 but did not cover the work described in Chapter 3.

Chapter 3 describes work that partially characterized a novel DNA-binding domain of DmSNAP43. DmSNAP43 has no recognizable canonical DNA binding domains despite the fact that it makes direct contact with the PSEA. Truncational analysis described in Chapter 1 revealed that a region in the non-conserved C-terminal domain is required for DmSNAPc DNA binding activity (Hung et al., 2009). Subsequent site-specific protein-DNA photo-cross-linking combined with site-specific chemical digestion indicated that this same region makes direct contact to the DNA (Kim et al., 2010). In order to better characterize this domain, the contribution of

amino acid residues within this region to DNA binding activity was evaluated by alanine substitution analysis. Blocks of three to six amino acids at a time were mutated to alanines throughout this region. Individual mutant DmSNAP43 constructs were coexpressed with the other two complementary subunits in S2 cells. The *in vitro* and *in vivo* DNA binding activities of DmSNAPc containing each mutant construct were assessed by EMSA and ChIP assays. Results from this chapter have been submitted for publication and the dissertation author is the primary researcher of this study.

## REFERENCES

- Barakat NH, Stumph WE. (2008). TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proximal sequence element A (PSEA) for the U1 PSEA. FEBS Letters, 582, 2413-16.
- Biedenkapp H, Borgmeyer U, Sippel AE, Klempnauer KH. (1988). Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature, 335, 835-7.
- Bond UM, Yario TA, Steitz JA. (1991). Multiple processing-defective mutations in a mammalian histone pre-mRNA are supressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev, 5, 1709-22.
- Busch H, Reddy R, Rothblum L, Choi YC. (1982). SnRNAs, SnRNPs, and RNA processing. Annu Rev Biochem, 51, 617-54.
- Dahlberg JE, Lund E. (1988). The genes and transcription of the major small nuclear RNAs. In: Birnstiel ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag KG, 38-70.
- Das G, Henning D, Reddy R. (1987). Structure, organization, and transcription of Drosophila U6 small nuclear RNA genes. J Biol Chem, 262, 1187-93.
- Goomer RS, Urso O, Kunkel GR. (1994). A complex that contains proteins binding to the PSE and TATA sites in a human U6 small nuclear RNA promoter. Gene, 148, 269-75.

- Guthrie C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science, 253, 157-63.
- Henry RW, Mittal V, Ma BC, Kobayashi R, Hernandez N. (1998). SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. Genes Dev, 12, 2664-72.
- Henry RW, Sadowski CL, Kobayashi R, Hernandez N. (1995). A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. Nature, 374, 653-56.
- Hernandez G, Valafar F, Stumph WE. (2007). Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. Nucleic Acids Res, 35, 21-34.
- Hernandez N. (1992). Transcription of vertebrate snRNA genes and related genes. In: McKnight SL, Yamamoto KR, eds. Transcriptional Regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 281-313.
- Hinkley CS, Hirsch HA, Gu LP, LaMere B, Henry RW. (2003). The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem, 278, 18649-57.
- Hung KH, Stumph WE. (2011). Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc). Crit Rev Biochem Mol Biol, 46, 11-26.
- Hung KH, Titus M, Chiang SC, Stumph WE. (2009). A map of Drosophila melanogaster small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. J Biol Chem, 284, 22568-79.
- Jawdekar GW, Hanzlowsky A, Hovde SL, Jelencic B, Feig M, Geiger JH, Henry RW. (2006). The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem, 281, 31050-60.
- Jensen RC, Wang Y, Hardin SB, Stumph WE. (1998). The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of Drosophila U-snRNA genes. Nucleic Acids Res, 26, 616-22.

- Kass S, Tyc K, Steitz JA, Sollner-Webb B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell, 60, 897-908.
- Kim MK, Kang YS, Lai HT, Barakat NH, Magante D, Stumph WE. (2010). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol, 30, 2411-23.
- Klempnauer KH, Sippel AE. (1987). The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. EMBO J, 6, 2719-25.
- Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol, 9, 800-5.
- Lai HT, Chen H, Li C, McNamara-Schroeder KJ, Stumph WE. (2005). The PSEA promoter element of the Drosophila U1 snRNA gene is sufficient to bring DmSNAPc into contact with 20 base pairs of downstream DNA. Nucleic Acids Res, 33, 6579-86.
- Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE. (2004). Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol, 24, 1897-906.
- Lobo SM, Hernandez NT. (1994). Transcription of snRNA genes by RNA polymerases II and III. In: Conaway RC, Conaway JW, eds. Transcription: Mechanisms and Regulation. New York, NY: Raven Press, 127-59.
- Ma B, Hernandez N. (2001). A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc. J Biol Chem, 276, 5027-35.
- Ma BC, Hernandez N. (2002). Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol, 22, 8067-78.
- McNamara-Schroeder KJ, Hennessey RF, Harding GA, Jensen RC, Stumph WE. (2001). The Drosophila U1 and U6 gene proximal sequence elements act as important determinants of the RNA polymerase specificity of snRNA gene promoters in vitro and in vivo. J Biol Chem, 276, 31786-92.
- Mittal V, Ma BC, Hernandez N. (1999). SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes Dev, 13, 1807-21.

- Nguyen VT, Kiss T, Michels AA, Bensaude O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature, 414, 322-5.
- Parry HD, Scherly D, Mattaj IW. (1989). Snurpogenesis: the transcription and assembly of U snRNP components. Trends Biochem Sci, 14, 15-19.
- Peculis BA, Steitz JA. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell, 73, 1233-45.
- Sadowski CL, Henry RW, Lobo SM, Hernandez N. (1993). Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. Genes Dev, 7, 1535-48.
- Sharp PA. (1994). Split genes and RNA splicing. Cell, 77, 805-16.
- Steitz JA, Black DL, Gerke V, Parker KA, Kramer A, Frendewey D, Keller W. (1988). Functions of the abundant U-snRNPs. In: Birnsteil ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag KG, 115-54.
- Su Y, Song Y, Wang Y, Jessop L, Zhan LC, Stumph WE. (1997). Characterization of a Drosophila proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. Eur J Biochem, 248, 231-37.
- Waldschmidt R, Wanandi I, Seifart KH. (1991). Identification of transcription factors required for the expression of mammalian U6 genes in vitro. EMBO J, 10, 2595-603.
- Wang Y, Stumph WE. (1998). Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 snRNA genes. Mol Cell Biol, 18, 1570-79.
- Wong MW, Henry RW, Ma BC, Kobayashi R, Klages N, Matthias P, Strubin M, Hernandez N. (1998). The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1. Mol Cell Biol, 18, 368-77.
- Yang ZY, Zhu QW, Luo KX, Zhou Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature, 414, 317-22.
- Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG. (1995). Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Mol Cell Biol, 15, 2019-27.

- Yoon JB, Roeder RG. (1996). Cloning of two proximal sequence element-binding transcription factor subunits (g and d) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein. Mol Cell Biol, 16, 1-9.
- Zamrod Z, Tyree CM, Song Y, Stumph WE. (1993). In vitro transcription of a Drosophila U1 small nuclear RNA gene requires TATA box-binding protein and two proximal cis-acting elements with stringent spacing requirements. Mol Cell Biol, 13, 5918-27.
- Zieve G, Benecke BJ, Penman S. (1977). Synthesis of two classes of small RNA species in vivo and in vitro. Biochemistry, 16, 4520-5.

# **CHAPTER 1**

A map of *Drosophila melanogaster* small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding

# A Map of *Drosophila melanogaster* Small Nuclear RNA-activating Protein Complex (DmSNAPc) Domains Involved in Subunit Assembly and DNA Binding\*

Received for publication, June 8, 2009, and in revised form, June 23, 2009 Published, JBC Papers in Press, June 25, 2009, DOI 10.1074/jbc.M109.027961

Ko-Hsuan Hung<sup>‡§</sup>, Mitchell Titus<sup>‡¶1</sup>, Shu-Chi Chiang<sup>‡§</sup>, and William E. Stumph<sup>‡¶2</sup>

From the <sup>‡</sup>Molecular Biology Institute and the Departments of <sup>§</sup>Biology and <sup>§</sup>Chemistry and Biochemistry, San Diego State University, San Diego, California 92182-1030

Transcription of genes coding for the small nuclear RNAs (snRNAs) is dependent upon a unique transcription factor known as the small nuclear RNA-activating protein complex (SNAPc). SNAPc binds to an essential proximal sequence element located about 40-65 base pairs upstream of the snRNA transcription start site. In the fruit fly Drosophila melanogaster, DmSNAPc contains three distinct polypeptides (DmSNAP190, DmSNAP50, and DmSNAP43) that are stably associated with each other and bind to the DNA as a complex. We have used mutational analysis to identify domains within each subunit that are involved in complex formation with the other two subunits in vivo. We have also identified domains in each subunit required for sequence-specific DNA binding. With one exception, domains required for subunit-subunit interactions lie in the most evolutionarily conserved regions of the proteins. However, DNA binding by DmSNAPc is dependent not only upon the conserved regions but is also highly dependent upon domains outside the conserved regions. Comparison with functional domains identified in human SNAPc indicates many parallels but also reveals significant differences in this ancient yet rapidly evolving system.

The small nuclear RNA (snRNA)³-activating protein complex (SNAPc) is a multisubunit protein required for transcription of genes that code for the spliceosomal (and certain other) snRNAs (1–4). SNAPc recognizes and binds specifically to a proximal sequence element (PSE) located about 40–65 base pairs upstream of the transcription start site. SNAPc has also variously been called PSE-binding protein (5, 6) and PSE-binding transcription factor (1, 3, 7). In humans, SNAPc contains five distinct polypeptide chains (SNAP190, SNAP50, SNAP45, SNAP43, and SNAP19) named based upon the apparent molecular weights of these subunits (4, 7–12). For the remainder of this article, the human protein and its subunits will be indicated by the prefix "Hs."

In the fruit fly *Drosophila melanogaster*, DmSNAPc contains three distinct polypeptide chains that are orthologous to HsSNAP190, HsSNAP50, and HsSNAP43 (13, 14). The three fly subunits, DmSNAP190, DmSNAP50, and DmSNAP43, are each present in a single copy in native DmSNAPc (15) and have calculated molecular masses of 84, 43, and 42 kDa, respectively. Interestingly, a homologous complex (tSNAPc) is required for transcription of the spliced leader snRNA in trypanosomes (16–18). This indicates that a SNAP-like complex arose very early in eukaryotic evolution and continues to be essential for snRNA transcription in widely divergent contemporary eukaryotes. However, even within insects, snRNA gene promoter sequences recognized by SNAPc have diverged fairly rapidly (19).

The subunits of eukaryotic SNAPc tightly associate with each other in solution even when the complex is not bound to DNA. The subunits co-purify through numerous chromatography columns (1–3, 16–18, 20). Moreover, each of the three metazoan core subunits is essential for sequence-specific binding to the PSE as none can bind individually or in pairwise combinations (21–23). However, an isolated region of HsSNAP190 that contains two Myb repeats binds weakly and with little sequence specificity to DNA (11, 22, 24).

We now report a mutational analysis of the three subunits of DmSNAPc from the fruit fly D. melanogaster to identify functional domains within each of the subunits. Truncated tagged versions of each of the proteins were expressed in D. melanogaster S2 cells together with the other two untagged subunits. The ability of the truncated subunits to participate in the assembly of DmSNAPc in the homologous in vivo system was monitored by coimmunoprecipitation assays from cellular extracts. In each case, we have identified regions within the polypeptides that are necessary or sufficient for in vivo assembly with the other two subunits. We have also identified domains within each DmSNAP polypeptide that are essential for sequence-specific binding to the PSEA (the fly equivalent of the PSE) but are dispensable for assembly with the other two subunits. Our mutational analysis of DmSNAPc expands on the knowledge obtained in the human system and in some cases reveals surprising differences in the localization of functional domains for complex assembly and DNA binding of the different metazoan SNAP complexes.

<sup>\*</sup>This work was supported by National Science Foundation Grants MCB-0641350 and MCB-0842770 and the California Metabolic Research Foundation

Recipient of an Arne N. Wick Pre-doctoral Research Fellowship from the California Metabolic Research Foundation.

<sup>&</sup>lt;sup>2</sup>To whom correspondence should be addressed. Tel.: 619-594-5575; Fax: 619-594-4634; E-mail: wstumph@sciences.sdsu.edu.

<sup>&</sup>lt;sup>3</sup> The abbreviations used are: snRNA, small nuclear RNA; SNAPc, small nuclear RNA-activating protein complex; PSE, proximal sequence element; PSEA, proximal sequence element A; ChIP, chromatin immunoprecipitation; Hs, human; Dm, D. melanogaster.

<sup>&</sup>lt;sup>4</sup> K.-H. Hung, M. Titus, S.-C. Chiang, and W.E. Stumph, unpublished observations.

#### DmSNAPc Domains for Subunit Assembly and DNA Binding

#### **EXPERIMENTAL PROCEDURES**

DmSNAP Expression Constructs—Constructs for expression of full-length untagged or V5 epitope-tagged DmSNAPs in D. melanogaster S2 cells have been previously described (14). Briefly, open reading frames encoding each of the DmSNAP proteins had been cloned into the pMT/V5-His-TOPO vector, a component of Invitrogen's Drosophila Expression System. To convert those to FLAG-tagged constructs, the appropriate plasmids were digested with BstBI and MIuI to remove the V5-coding sequence; this was then replaced with synthetic DNA sequence encoding the FLAG epitope.

Truncation constructs were prepared by using the full-length constructs as templates for the PCR together with appropriate primers. For N-terminal truncations, forward primers were designed to contain the natural start codon of the protein together with about 7 nucleotides 5′ of the start codon as the first one-third of the forward primer (~10 nucleotides). The second two-thirds of the forward primer (~20 nucleotides) was designed to match the sequence at the starting point of the desired truncation. PCR products were then cloned initially into the pMT/V5-His-TOPO expression vector. Fragments were removed from these constructs and re-cloned into the FLAG-modified expression vector.

To prepare C-terminal truncations tagged at the N terminus, PCR was used to introduce a C-terminal stop codon at each of the desired positions within the DmSNAP genes. The amplified DNA was initially cloned into pMT-V5-His-TOPO. These constructs were used as templates with forward primers that contained an SpeI site just upstream of the codon for the second amino acid of the DmSNAP190, DmSNAP50, or DmSNAP43 genes. Following digestion with SpeI and NotI, the PCR fragments were cloned into a pMT-V5-His-TOPO-based expression vector that contained an N-terminal FLAG sequence. This N-terminal FLAG vector was prepared as follows: synthetic DNA containing a translation initiation codon and the FLAG encoding sequence was cloned between the KpnI and SpeI sites of pMT-V5-His-TOPO. This FLAG-containing vector was then used for cloning C-terminal-truncated DmSNAPs after digestion with SpeI and NotI.

Stably Transfected S2 Cells—Expression plasmids for each of the DmSNAPs were used to co-transfect S2 cells together with pCoBlast according to conditions recommended by Invitrogen. Stably transfected cell lines were established by selection in blasticidin-containing medium. Expression of DmSNAPs was induced from the metallothionein promoter by treatment of the cells for 24 h with 0.5 mm copper sulfate.

FLAG Purification/Coimmunoprecipitation—Following induction, cells were washed in phosphate-buffered saline and lysed in CelLytic M lysis buffer (Sigma catalog number C2978) containing protease inhibitor mixture (Sigma catalog number P8340). For low-salt FLAG immunopurifications, cell lysates were used directly in overnight incubations with anti-FLAG M2-agarose beads (Sigma catalog number A2220). The beads were then washed four times in 0.05 m Tris-HCl (pH 7.4), 0.15 m NaCl and then twice in HEMG wash buffer (81 mm KCl, 32.5 mm Hepes K<sup>+</sup>, 5.5 mm MgCl<sub>2</sub>, 5.0 mm dithiothreitol, 10% glycerol, and 0.1 mm EDTA (pH 7.6)). For high-salt immunopurifi-

cations, lysates were first adjusted to a NaCl concentration of 0.35 M prior to incubating overnight with the anti-FLAG beads. The beads were then washed twice in 0.01 M Tris-HCl (pH 7.4), 0.35 M NaCl and then three times with HEMG wash buffer. In both cases, the DmSNAP complexes were eluted from the affinity gel by competition with  $3\times$  FLAG peptide (Sigma catalog number F4799) at a concentration of 200  $\mu g/ml$  in HEMG wash buffer.

Immunoblots—Whole cell lysate and FLAG-purified samples were run on 10-14% denaturing polyacrylamide gels and the proteins were transferred to polyvinylidene difluoride membranes. FLAG-tagged DmSNAP subunits were detected using anti-FLAG M2 monoclonal antibodies (Sigma) conjugated to either alkaline phosphatase or horseradish peroxidase. Untagged DmSNAP subunits were detected as previously described using primary antibodies generated in rabbits against amino acid sequences at or near the C termini of the DmSNAP polypeptides (14). Data shown in the figures are from gels in which the amount of FLAG-tagged protein loaded in each lane was normalized by running gels repetitively and each time adjusting the amounts of protein until the intensity of the signals from the tagged subunit in the different samples was as similar as reasonably possible. For all immunoblots shown, the lanes in each horizontal panel (top, middle, or bottom) are from the same gel blot developed for the same length of time. Thus the intensities of the bands within a horizontal panel reflect the relative amounts of protein present in the various cell extracts or FLAG affinity-purified fractions. This permitted an easy visual assessment of the relative amounts of the untagged subunits in each sample in comparison to a constant amount of the tagged subunit.

Electrophoretic Mobility Shift Assays-DNA mobility shift assays were carried out in 18-µl volumes in a final concentration of ~80 mm KCl, 25 mm Hepes (pH 7.6), 5 mm MgCl<sub>2</sub>, 10 μm ZnCl2, 0.1 mm EDTA, 9 mm dithiothreitol, 9% glycerol. The radioactive DNA probe contained the PSEA sequence of the D. melanogaster U1:95Ca gene (formerly called the U1-95.1 gene (25)). FLAG affinity-purified DmSNAP complexes, normalized for the content of the FLAG-tagged subunit (as determined by immunoblotting, described above), were added to each reaction. Incubation was carried out for 30 min at 20 °C. For complexes exhibiting weak or no binding, 2-6 times as much protein (constrained by the maximum volume of the reaction) was sometimes used in additional lanes to maximize the possibility of observing binding. Complexes were run on 5% nondenaturing polyacrylamide gels and the bands were detected by autoradiography.

Chromatin Immunoprecipitations (ChIPs)—ChIP assays were carried out as recently described (26). Affinity-purified rabbit polyclonal antibodies produced against the FLAG peptide (Sigma catalog number F7425) were used for FLAG ChIPs. The anti-DmSNAP43 antibodies used as positive controls in the ChIP assays were produced in a rabbit immunized with bacterially expressed recombinant DmSNAP43 (26). The preimmune serum was from the same rabbit prior to immunization. The ChIP PCR forward primer (5'-GTGTGGCATACCTTAT-AGGGGTGCT-3') and reverse primer (5'-GCTTTTCGATG-CTCGGCAGCAG-3') amplify the promoter region of the

#### DmSNAPc Domains for Subunit Assembly and DNA Binding

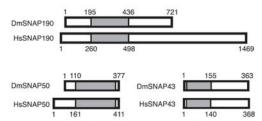


FIGURE 1. Comparison of *D. melanogaster (Dm)* and human (*Hs*) SNAP subunits. The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous protein pairs. The numbers above and below the rectangles designate the amino acid positions in the proteins.

*U1:95Ca* gene (26). In all experiments, the amounts of DNA samples and PCR conditions were chosen such that the resultant signals were within a semi-quantitative range.

#### RESULTS

Functional Domains of DmSNAP190-An unusual feature of the amino acid sequence of DmSNAP190 is the existence of 4.5 tandem Myb repeats located between residues 195 and 436 (14). This region (Fig. 1) is conserved between fly and human SNAP190 with an identity of 27% and overall similarity of 44% (11, 14). In human SNAPc, the Myb repeat domain of Hs-SNAP190 is involved in DNA binding (11, 22, 24). Because it seemed likely that deletion of this most conserved region of DmSNAP190 would be highly detrimental to DmSNAPc activity, we focused our studies on the non-conserved regions. Fulllength and N- and C-terminal DmSNAP190 truncation constructs under control of the copper-inducible metallothionein promoter were prepared with FLAG-His6 tags at the C terminus or with His-FLAG tags at the N terminus (Fig. 2A). Each tagged DmSNAP190 construct was used separately to transfect Drosophila S2 cells together with pCoBlast (Invitrogen) and the two complementary constructs to express the full-length DmSNAP50 and DmSNAP43 subunits without tags. Stably transfected cells were selected by growth on blasticidin.

Cellular extracts were prepared following copper induction and immunoblots were carried out to examine the expression of the DmSNAP190, DmSNAP50, and DmSNAP43 proteins in each of the cell lines. Detection of the respective proteins employed monoclonal antibodies against the FLAG epitope (for the DmSNAP190 constructs) or polyclonal antibodies specific for C-terminal peptides of DmSNAP50 or DmSNAP43 (14). Fig. 2B shows that each of the DmSNAP190 truncation mutants was expressed in the corresponding stably transfected cells (*lanes 1–7, top panels*). Fig. 2B also shows that both DmSNAP50 and DmSNAP43 were expressed among the different cell lines at relatively similar levels (*middle* and *bottom panels*). Expression of these proteins from the endogenous genes in normal S2 cells was undetectable under these conditions (data not shown, but see Ref. 14).

To determine whether the truncated DmSNAP190 proteins were able to assemble *in vivo* with the co-expressed DmSNAP50 and DmSNAP43 subunits, the extracts were applied to FLAG antibody affinity resins under either low-salt (Fig. 2C) or high-salt (Fig. 2D) conditions (10 and 350 mm NaCl,

respectively). Bound protein was eluted by competition with FLAG peptide. The *top panels* in Fig. 2, *C* and *D*, show that all seven full-length or truncated DmSNAP190 proteins could be detected in the elution fractions by immunoblot analysis using the FLAG antibody. As revealed in the *bottom most panels* of Fig. 2, *C* and *D*, DmSNAP43 co-purified with each of the DmSNAP190 constructs, including construct DmSNAP190-(176–451), which contained little more than the 4.5 Myb repeats. This indicates that the Myb domain of DmSNAP190 is sufficient for association with DmSNAP43.

The result with DmSNAP50, however, was different. As expected, DmSNAP50 co-purified with full-length DmSNAP190 tagged at either the C or N terminus (Fig. 2, C and D, lanes 1 and 5, middle panels). Moreover, DmSNAP50 associated with the truncation construct that lacked the first 62 amino acid residues of DmSNAP190 (lane 2 in Fig. 2, C and D); likewise, DmSNAP50 associated with the two constructs that lacked amino acids C-terminal to the Myb domain (lanes 6 and 7). However, DmSNAP50 failed to co-purify with the two different DmSNAP190 constructs that were missing 175 residues from the N terminus (Fig. 2, C and D, lanes 3 and 4, middle panels). These results indicate that residues located between positions 63 and 175 of DmSNAP190 are essential for its assembly with DmSNAP50.

Next, to investigate which regions of DmSNAP190 contribute to DNA binding by DmSNAPc, the FLAG affinity-purified complexes were subjected to electrophoretic mobility shift analysis (Fig. 3). In Fig. 3, A, lanes 1-4, and B, lanes 1-4, protein amounts in each lane were normalized as determined from the immunoblots shown in Fig. 2, C and D, respectively. The complexes that contained DmSNAP190 that was either full-length or missing 62 amino acids from its N terminus bound efficiently to the PSEA (Fig. 3, A, lanes 1 and 2, and B, lanes 1 and 2). However, the complex that contained DmSNAP190 lacking 175 N-terminal residues was not able to bind DNA (Fig. 3, A, lane 3, and B, lane 3). DmSNAP190 that contained only the Myb domain gave a similar result (lane 4). Neither the use of 2 or 3 times the amount of normalized protein (Fig. 3A, lanes 7 and 8), nor longer exposure of the film (data not shown), was able to detect any DNA binding activity for these latter two truncation constructs. However, this was expected because these complexes lacked DmSNAP50 (Fig. 2, C and D), and previous results have indicated that all three subunits of DmSNAPc contact the DNA (13, 14) and are required for the sequence-specific DNA-binding activity of DmSNAPc.5

DmSNAPc that contained full-length DmSNAP190 tagged at the N terminus also bound efficiently to DNA (Fig. 3, *A*, *lane* 9, and *B*, *lane* 5). However, DNA binding was severely compromised by truncations from the C terminus following residues 623 or 451 (Fig. 3, *A*, *lanes* 10–12, and *B*, *lanes* 6 and 7). Fig. 3*A*, *lanes* 13–16, show a much longer exposure of film to the same gel shown in *lanes* 9–12. This longer exposure revealed that DmSNAPc, which contained DmSNAP190 truncated following position 623 (*lane* 14), retained an extremely low level of DNA-

<sup>&</sup>lt;sup>5</sup> K.-H. Hung, M. Titus, S.-C. Chiang, and W. E. Stumph, unpublished data.

#### DmSNAPc Domains for Subunit Assembly and DNA Binding

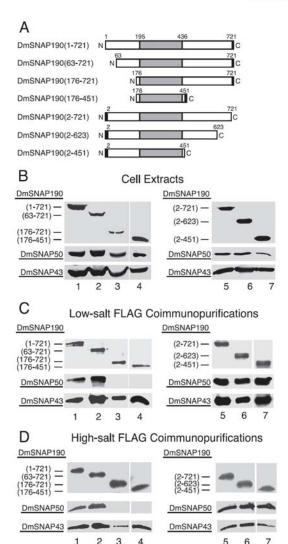
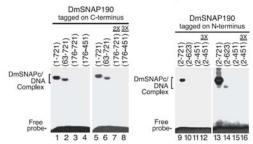


FIGURE 2. Domains of DmSNAP190 involved in assembly in vivo with DmSNAP50 and DmSNAP43. A, schematic representation of full-length and truncated DmSNAP190 constructs co-transfected into \$2 cells together with constructs expressing full-length untagged DmSNAP50 and DmSNAP43. The shaded area represents the 4.5-Myb-repeat domain conserved between humans and fruit files. The black rectangles represent FLAG tags at the C or N termini of DmSNAP190. The names of the constructs in the column at the left indicate the extent of the wild type amino acid residues present in the expressed constructs. B, co-overexpression of tagged DmSNAP190 with DmSNAP50 and DmSNAP43 in stably transfected \$2 cells. Whole cell extracts from stably transfected cells co-overexpressing all three DmSNAP subunits were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of extract loaded in each lane was normalized so that the intensity of the signal obtained from the tagged DmSNAP190 construct was similar in each lane. The top panels show full-length or truncated tagged DmSNAP190 detected using monoclonal antibodies against the FLAG epitope. The middle and bottom panels show detection of DmSNAP50 or DmSNAP43, respectively, by using polyclonal anti-peptide antibodies prepared against C-terminal peptides of the respective proteins (14). For all immunoblots shown in this and subsequent figures, the lanes in each

### A Low-salt FLAG Coimmunopurifications



#### B High-salt FLAG Coimmunopurifications

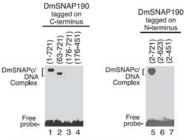


FIGURE 3. Domains of DmSNAP190 required for sequence-specific DNA binding by DmSNAPc. A, DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 2C) were used for DNA mobility shift analysis with a DNA probe containing a UT PSEA sequence. Complexes containing DmSNAP190 constructs tagged at the C terminus (N-terminal truncations) are shown in lanes 1–8, and constructs tagged at the N terminus (C-terminal truncations) are shown in lanes 9–16. Lanes 1–6 were carried out with complexes containing a normalized amount of DmSNAP190 as determined in Fig. 2C, whereas lanes 7 and 8 contained 2 or 3 times the normalized amount of truncated protein. (The maximum excess of protein that could be added was limited by the final volume of the reaction and the relative concentrations of each sample.) Lanes 9–11 also contained normalized amounts of protein, whereas lane 12 contained 3 times this amount of protein. Lanes 13–16 show a longer exposure of the same gel shown in lanes 9–12. B, similar to A except DmSNAP complexes were affinity purified under high-salt conditions and protein amounts were normalized based upon the immunoblots shown in Fig. 2D.

binding activity, whereas truncation following position 451 completely abrogated detectable DNA binding (*lane 15*) even when 3 times the normalized amount of protein was used (*lane 16*).

horizontal panel (top, middle, or bottom) are from the same gel blot developed for the same length of time. Thus the intensities of the bands within a horizontal panel reflect the relative amounts of protein present in the various transfected cell lines. C, co-purification of DmSNAP50 and DmSNAP43 with full-length and truncated DmSNAP190 constructs following FLAG affinity purification under low-salt conditions. Complexes containing full-length or truncated tagged DmSNAP190 were purified using FLAG antibody beads, and the presence of the individual subunits in the elution fractions was evaluated by immunoblotting. The volume of elution fraction loaded in each lane was normalized so that the intensity of the signal for the DmSNAP190 construct was similar in each lane when detected with FLAG antibodies. DmSNAP50 and DmSNAP43 subunits that co-purified with the FLAG-tagged DmSNAP190 constructs were detected with antibodies prepared against C-terminal peptides of the respective proteins. D, same as C except FLAG affinity purification was carried out under high-salt conditions (0.35 m NaCl).

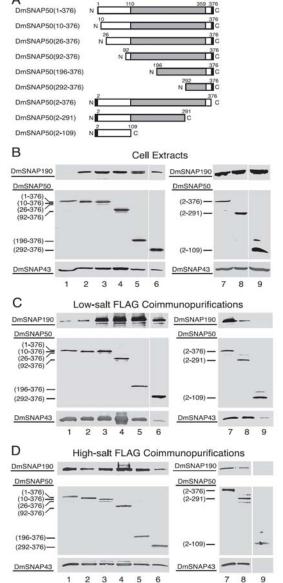


FIGURE 4. Domains of DmSNAP50 involved in assembly in vivo with DmSNAP190 and DmSNAP43. A, schematic representation of full-length and truncated DmSNAP50 constructs co-transfected into \$2 cells together with constructs expressing full-length untagged DmSNAP190 and DmSNAP43. The shaded area represents the region most conserved between humans and fruit flies. The black rectangles represent FLAG tags at the C or N termini of DmSNAP50. The names of the constructs in the column at the left indicate the extent of the wild type amino acid residues present in the expressed constructs. B, co-overexpression of tagged DmSNAP50 with DmSNAP190 and DmSNAP43 in stably transfected \$2 cells. Whole cell extracts from stably transfected cells co-overexpressing all three DmSNAP subunits were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of extract loaded in each lane was

These results indicated that the C-terminal domain of DmSNAP190 is essential for DmSNAPc to bind efficiently to the PSEA, even though it is not required for assembly with either DmSNAP43 or DmSNAP50 (Fig. 2, C and D). This result was surprising because human SNAP190 truncated immediately following the Myb domain can associate with HsSNAP43 and HsSNAP50 into a complex known as miniSNAPc that binds very efficiently to DNA (21, 22, 24). Thus, the C-terminal domain of DmSNAP190 is important for DNA binding by fly SNAPc despite the fact that it is not required, and is even inhibitory (21), for human SNAPc binding to DNA.

Functional Domains of DmSNAP50—The SNAP50 orthologs are the most evolutionarily conserved of the SNAP proteins (14, 16). The most conserved region of DmSNAP50 is the C-terminal region where residues 110 to 359 have 33% identity and 51% similarity to the C-terminal region of Hs-SNAP50 (Fig. 1) (14). We concentrated most of our efforts on truncations from the N terminus, but two C-terminal truncations were also prepared (Fig. 4A). N-terminal truncations had FLAG-His<sub>6</sub> tags at the C terminus, whereas the C-terminal truncations had His<sub>6</sub>-FLAG tags at the N terminus. Each DmSNAP50 construct was co-overexpressed in S2 cells together with full-length untagged DmSNAP190 and DmSNAP43.

Successful expression of each of the tagged DmSNAP50 constructs was indicated by immunoblots of total cellular lysates prepared from each of these cell lines (Fig. 4B, lanes 1–9, middle panels). DmSNAP43 and DmSNAP190 were also readily detected in all cell lines with one exception: DmSNAP190 was difficult to detect in lysates from cells expressing full-length C-terminal tagged DmSNAP50 under conditions where it was readily detectable in the remaining cell lines (Fig. 4B, upper panel, lane 1 versus lanes 2–9). However, upon FLAG purification of DmSNAPc from those cells, DmSNAP190 became detectable in the FLAG-purified fraction (see Fig. 4, C and D, below).

The abilities of DmSNAP190 and DmSNAP43 to associate with the DmSNAP50 truncations were assayed by immunoblotting following FLAG affinity purification under either low-salt (Fig. 4C) or high-salt (Fig. 4D) conditions. Lanes 1–9 in the middle panels of Fig. 4, C and D, show that each of the FLAG-tagged DmSNAP50 variants was immunopurified in the elution fractions from the FLAG affinity resins. The top panels in Fig. 4,

normalized so that the intensity of the signal obtained from the tagged DmSNAP50 construct was similar in each lane. The middle panels show full length or truncated tagged DmSNAP50 detected using monoclonal antibodies against the FLAG epitope. The top and bottom panels show detection of DmSNAP190 or DmSNAP43, respectively, by using polyclonal anti-peptide antibodies prepared against C-terminal peptides of the respective proteins (14). C, co-purification of DmSNAP190 and DmSNAP43 with full-length and truncated DmSNAP50 constructs following FLAG affinity purification under low-salt conditions. Complexes containing full-length or truncated tagged DmSNAP50 were purified using FLAG antibody beads, and the presence of the individual subunits in the elution fractions was evaluated by immuno-blotting. The volume of elution fraction loaded in each lane was normalized so that the intensity of the signal for the DmSNAP50 construct was similar in each lane when detected with FLAG antibodies. DmSNAP190 and DmSNAP43 subunits that co-purified with the FLAG-tagged DmSNAP50 constructs were detected with antibodies prepared against C-terminal peptides of the respective proteins. D, same as C except FLAG affinity purification was carried out under high-salt conditions.

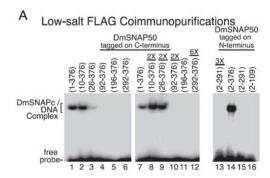
C and D, reveal the ability of DmSNAP190 to co-purify with the various full-length and truncated DmSNAP50 constructs. Under low-salt conditions of purification, DmSNAP50 truncation mutants that had 25 or more amino acid residues deleted from the N terminus co-precipitated greater amounts of DmSNAP190 compared with that precipitating with the full-length or nearly full-length constructs (Fig. 4C, top panels, lanes 3–6 versus lanes 1 and 2). However, when the co-immunopurifications were carried out in higher salt conditions (Fig. 4D), the amount of co-immunopurified DmSNAP190 exhibited much less variability. Thus, under certain conditions the N terminus of DmSNAP50 may play a role in preventing an artifactual association of extra copies of DmSNAP190 with DmSNAP50.

The last 85 residues of DmSNAP50 (residues 292–376) contain an evolutionarily conserved (14, 16) yet unorthodox zinc-binding domain termed the "SNAP finger" (23). Interestingly, these last 85 residues were sufficient to co-immunoprecipitate both DmSNAP190 and DmSNAP43 under either low- or high-salt conditions, although possibly at reduced levels in certain cases (Fig. 4, C and D, lane 6, top and bottom panels). Thus the SNAP finger region of DmSNAP50 appeared to interact with both DmSNAP190 and DmSNAP43; however, from those data alone it remained possible that the interaction with one of these subunits could be indirect as a result of mutual interactions between DmSNAP190 and DmSNAP43.

To further examine whether DmSNAP43 and DmSNAP190 can independently interact with the SNAP finger of DmSNAP50, we expressed the SNAP finger domain, DmSNAP50-(292–376), in pairwise combinations either with DmSNAP190 alone or with DmSNAP43 alone. In each case, we observed that DmSNAP190 and DmSNAP43 copurified with FLAG-tagged DmSNAP50-(292–376) (data not shown). Thus both DmSNAP190 and DmSNAP43 can interact independently with the SNAP finger domain of DmSNAP50.

Results from the C-terminal truncations of DmSNAP50 are shown in Fig. 4, C and D, lanes 7-9. Truncation of DmSNAP50 beyond residue 291 partially impaired the co-purification of DmSNAP190 (compare lanes 7 and 8 in the top panels of Fig. 4, C and D) although having no effect on the co-purification of DmSNAP43 (compare lanes 7 and 8 in the bottom panels of Fig. 4, C and D). These data suggest that residues of DmSNAP50 between 110 and 291 comprise a very significant site of interaction with DmSNAP43, even though the SNAP finger region undoubtedly also contributes to the interaction. Deletion of DmSNAP50 amino acids C-terminal to position 109 completely abrogated the binding of DmSNAP190 and severely compromised or eliminated the binding of DmSNAP43 (Fig. 4, C and D, top and bottom panels, lanes 9). Together with the results from the other truncations, this suggests that the non-conserved region of DmSNAP50 (residues 1-109) makes little if any contribution toward forming a complex with DmSNAP190 and DmSNAP43.

To map regions of DmSNAP50 important for the binding of DmSNAPc to the PSEA, the FLAG-purified DmSNAP complexes that contained full-length or truncated DmSNAP50 were used in electrophoretic mobility shift assays (Fig. 5). DmSNAPc that contained full-length DmSNAP50 tagged on





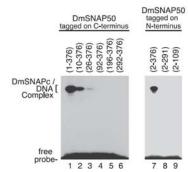


FIGURE 5. Domains of DmSNAP50 required for sequence-specific DNA binding by DmSNAPc. A, DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 4C) were used for DNA mobility shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP50 constructs tagged at the C terminus (N-terminal truncations) are shown in lanes 1–12, and constructs tagged at the N terminus (C-terminal truncations) are shown in lanes 13–16. Lanes 1–7 and 11 were carried out with complexes containing a normalized amount of DmSNAP50 as determined in Fig. 4C, whereas lanes 8–10 contained two times the normalized amount of truncated protein, and lane 12 contained six times the normalized amount of truncated protein, and lane 12 contained normalized amount of protein, whereas lane 13 contained 3 times the normalized amount of protein, (The maximum amount of protein that could be used was limited by the volume of the reaction and the relative concentration of each sample.) B, similar to A except DmSNAP complexes were affinity purified under high-salt conditions and protein amounts were normalized based upon the immunoblots shown in Fig. 4D.

the C terminus bound DNA efficiently (Fig. 5, A, lanes 1 and 7, and B, lane 1). SNAP complexes that contained DmSNAP50 lacking either the first 9 or 25 amino acid residues retained partial DNA-binding activity (Fig. 5, A, lanes 2, 3, 8, and 9, and B, lanes 2 and 3). However, deletion of the first 91 amino acid residues of DmSNAP50 resulted in the complete loss of detectable DNA-binding activity (Fig. 5, A, lanes 4 and 10, and B, lane 4) even though this DmSNAP50 construct associated very well with both DmSNAP190 and DmSNAP43 (Fig. 4, C and D, lane 4). Thus, amino acids between positions 26 to 91 of DmSNAP50 are critical for DNA binding even though these lie outside of the evolutionarily conserved region and are not essential for complex formation with DmSNAP190 and DmSNAP43.

Although DmSNAPc that contained full-length DmSNAP50 tagged at the N terminus bound efficiently to the PSEA (Fig. 5,

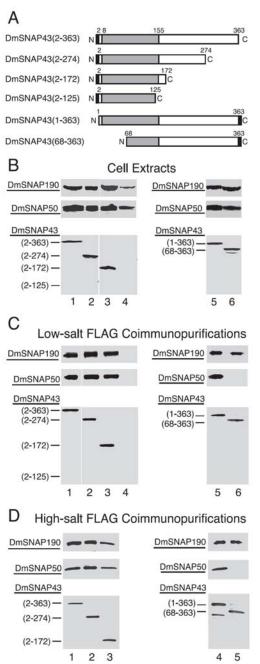


FIGURE 6. Domains of DmSNAP43 involved in assembly in vivo with DmSNAP190 and DmSNAP50. A, schematic representation of full-length and truncated DmSNAP43 constructs co-transfected into \$2 cells together with constructs expressing full-length untagged DmSNAP190 and DmSNAP50. The shaded area represents the region most conserved between humans and fruit flies. The black rectangles represent FLAG tags at the C or N termini of DmSNAP43. The names of the constructs in the column at the left

A, lane 14, and B, lane 7), deletion of the SNAP finger of DmSNAP50 eliminated DNA-binding activity (Fig. 5, A, lanes 13 and 15, and B, lane 8). These results were to be expected because the SNAP finger was clearly required for DNA binding in the human system (23).

Functional Domains of DmSNAP43—DmSNAP43 is 363 amino acid residues in length, and the evolutionarily most conserved region (residues 8 to 155, shaded in Fig. 1) has 31% identity and 48% similarity to the human ortholog (14). We concentrated our analysis on truncations that removed portions of the non-conserved C terminus; however, one N-terminal truncation was also prepared (Fig. 6A). Each construct was co-overexpressed in S2 cells with untagged DmSNAP190 and DmSNAP50

Expression of each of the full-length and truncated proteins was examined by immunoblot analysis of cellular extracts prepared from the individual stably transfected cell lines. The bottom panels in Fig. 6B show that expression of the two tagged full-length and three of the truncated DmSNAP43 constructs were readily detected. However, expression of the construct truncated following position 125 was nearly undetectable (lane 4), suggesting that this shortest construct was unstable. The upper panels in Fig. 6B illustrate that DmSNAP190 and DmSNAP50 were expressed in all six cell lines.

The DmSNAP complexes were then immunoprecipitated by FLAG affinity chromatography under either low-salt or high-salt conditions, and the ability of DmSNAP190 or DmSNAP50 to co-precipitate with the tagged DmSNAP43 was assayed by immunoblotting (Fig. 6, *C* and *D*). As expected, there was no telestable purification of the poorly expressed DmSNAP43-(2–125) construct (Fig. 6*C*, lane 4, bottom panel); furthermore, no co-eluting DmSNAP190 nor DmSNAP50 could be detected in the elution fractions from that cell line (Fig. 6*C*, lane 4, top and middle panels).

However, all other truncated and full-length DmSNAP43 constructs were readily detectable in the FLAG elution fractions (Fig. 6, C, lanes 1–3, 5–6, and D, lanes 1–5, in the bottom panels). The top most panels in Fig. 6, C and D, show that

indicate the extent of the wild type amino acid residues present in the expressed constructs. B, co-overexpression of tagged DmSNAP43 with DmSNAP190 and DmSNAP50 in stably transfected S2 cells. Whole cell extracts from stably transfected cells co-overexpressing all three DmSNAP subunits were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of extract loaded in each lane was normalized so that the intensity of the signal obtained from the tagged DmSNAP43 construct was similar in each lane. The bottom panels show full-length or truncated tagged DmSNAP43 detected using monoclonal antibodies against the FLAG epitope. The top and middle panels show detection of DmSNAP190 or DmSNAP50, respectively, by using polyclonal anti-peptide antibodies prepared against C-terminal peptides of the respective proteins (14), C, co-purification of DmSNAP190 and DmSNAP50 with full-length and truncated DmSNAP43 constructs following FLAG affinity purification under low-salt conditions. Complexes containing full-length or truncated tagged DmSNAP43 were purified using FLAG antibody beads, and the presence of the individual subunits in the elution fractions was evaluated by immuno-blotting. The volume of elution fraction loaded in each lane was normalized so that the intensity of the signal for the DmSNAP43 construct was similar in each lane when detected with FLAG antibodies. DmSNAP190 and DmSNAP50 subunits that co-purified with the FLAG-tagged DmSNAP43 constructs were detected with antibodies prepared against C-terminal peptides of the respective proteins. D, same as C except FLAG affinity purification was carried out under high-salt conditions.

DmSNAP190 co-purified with each of the stably expressed DmSNAP43 constructs. This indicates that neither the last 191 residues (those beyond position 172) nor the first 67 amino acid residues of DmSNAP43 are essential for recruitment of DmSNAP190 into the complex. By inference, the data suggest that DmSNAP43 residues between 68 and 172 are likely involved in interaction with DmSNAP190.

Interestingly, DmSNAP50 co-purified with the two stably expressed DmSNAP43 C-terminal truncation constructs (Fig. 6, C and D, middle panels, lanes 2 and 3) but failed to co-purify with the N-terminal truncation construct DmSNAP43-(68 – 363) (Fig. 6, C, lane 6, and D, lane 5, middle panels). This latter result indicates that the first 67 amino acid residues of DmSNAP43 are required for stable complex formation with DmSNAP50.

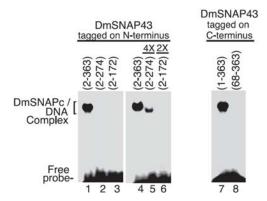
Finally, to investigate which domains of DmSNAP43 contribute to DNA binding by DmSNAPc, the FLAG affinity-purified complexes were subjected to electrophoretic mobility shift analysis (Fig. 7). Complexes that contained N-terminal tagged full-length DmSNAP43 bound efficiently to the PSEA (Fig. 7, A, lanes 1 and 4, and B, lane 1). However, DNA binding by the construct that was truncated following residue 274 was weakly detectable only when 4 times the normalized amount of protein was used (Fig. 7, A, lanes 2 and 5, and B, lane 2). No DNAbinding activity was detectable for the construct that was truncated following position 172 (Fig. 7, A, lanes 3 and 6, and B, lane 3). These results indicate that the non-conserved C-terminal domain of DmSNAP43 plays a critical role in DmSNAPc binding to the PSEA, even though this region is not required for DmSNAP43 to assemble into a complex with both DmSNAP190 and DmSNAP50.

DmSNAPc that contained full-length C-terminal tagged DmSNAP43 bound efficiently to DNA (Fig. 7, A, lane 7, and B, lane 4). However, no DNA-binding activity was detectable following the deletion of the N-terminal 67 amino acid residues of DmSNAP43 (Fig. 7, A, lane 8, and B, lane 5). This was to be expected because DmSNAP50 was not complexed with this N-terminal-truncated DmSNAP43 (Fig. 6, C, lane 6, and D, lane 5).

Requirements for the Binding of DmSNAPc to the U1 Gene Promoter in Vivo—Electrophoretic mobility shift experiments described above (Figs. 3, 5, and 7) examined the *in vitro* sequence-specific DNA-binding activity of DmSNAPc that contained tagged full-length or truncated DmSNAP subunits. To examine the *in vivo* DNA-binding activity of these mutant DmSNAP complexes, chromatin immunoprecipitation (ChIP) assays were conducted by using the same stably transfected cell lines prepared to make extracts for the immunoaffinity purification experiments.

ChIPs were carried out by using antibodies against the FLAG epitope to examine the *in vivo* occupancy of the well characterized endogenous *U1:95Ca* gene promoter by the FLAG-tagged constructs. The promoter of the *U1:95Ca* gene is active *in vivo* and has been previously employed for ChIP analysis (26). The PCR primers specific for this promoter amplify a 107-base pair DNA fragment, and their genomic locations in the 5'-flanking DNA of the *U1:95Ca* gene are illustrated in Fig. 8A. Polyclonal antibodies prepared against full-length DmSNAP43 were used

## A Low-salt FLAG Coimmunopurifications



## B High-salt FLAG Coimmunopurifications

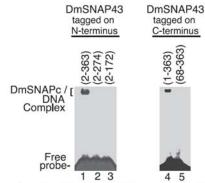


FIGURE 7. Domains of DmSNAP43 required for sequence-specific DNA binding by DmSNAPc. A, DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 6C) were used for DNA mobility shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP43 constructs tagged at the N terminus (C-terminal truncations) are shown in lanes 1–6, and constructs tagged at the C terminus (N-terminal truncations) are shown in lanes 7 and 8. Lanes 1–4 and 7 and 8 were carried out with complexes containing a normalized amount of DmSNAP43 as determined in Fig. 6C, whereas lanes 5 and 6 contained 4 or 2 times the normalized amount of truncated protein as indicated above each lane. (The maximum amount of protein that could be used was limited by the volume of the reaction and the relative concentration of each sample.) 8, similar to 4 except DmSNAP complexes were affinity purified under high-salt conditions and protein amounts were normalized based upon the immunoblots shown in Fig. 6.D.

as positive ChIP controls because DmSNAP43 should consistently be present at this promoter either as a component of the overexpressed FLAG-tagged DmSNAPc, or in its absence, as a component of endogenous DmSNAPc. Preimmune antibodies from the same rabbits were used as negative controls in the ChIP assays.

In the transfected cell lines that overexpressed tagged fulllength DmSNAP190, the anti-FLAG antibodies efficiently pre-

cipitated the U1 promoter. This was true whether the tag was at the C or N terminus of the protein (Fig. 8*B*, *lanes 3* and *19*). Truncation of amino acid residues from either terminus pro-

gressively decreased the intensity of the PCR signal (Fig. 8B, lanes 7, 11, 15, 23, and 27). This indicates that residues both N- and C-terminal to the conserved Myb domain of DmSNAP190 are important for the efficient binding of DmSNAPc to the U1 promoter in vivo. Moreover, these results were similar to those obtained with the in vitro binding assays of the full-length and truncated DmSNAP190 constructs shown in Fig. 3.

ChIP results from cells expressing FLAG-tagged DmSNAP50 and DmSNAP43 truncation constructs were also quite consistent with the band shift data. DmSNAP complexes that contained tagged full-length DmSNAP50 and DmSNAP43 were readily detected on the U1 promoter by ChIP (Fig. 8, C, lanes 3 and 19, and D, lanes 3 and 15). Removal of the first nine amino acids from DmSNAP50 had little effect on DmSNAPc-binding activity in vivo (Fig. 8C, lane 7), whereas U1 promoter occupancy was greatly reduced or no longer detectable as a result of additional truncations from the N or C terminus of DmSNAP50 (Fig. 8C, lanes 11, 15, and 23). Similarly, each of the tested truncations of DmSNAP43 from the N or C terminus severely damaged the in vivo DNA-binding activity of DmSNAPc (Fig. 8D, lanes 7, 11, and 19).

### DISCUSSION

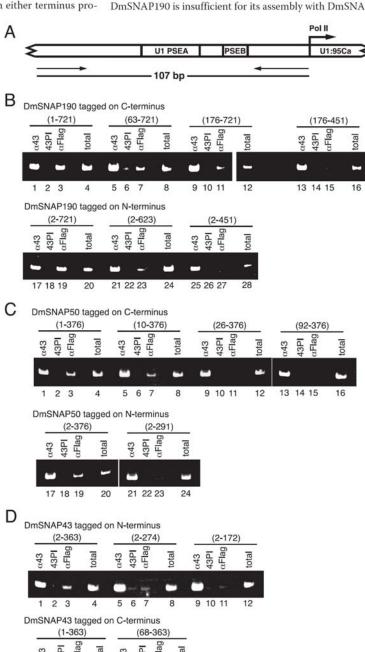
Functional Domains of DmSNAPc Subunits-Fig. 9 summarizes the results of our truncation analyses to identify protein domains involved in DmSNAPc assembly and in binding to the PSEA. With one exception, the evolutionarily most conserved region of each DmSNAP subunit is sufficient for its assembly with the other two subunits. For example, DmSNAP43-(2-172) can assemble with both DmSNAP190 and DmSNAP50. DmSNAP50-(92-376) can assemble with both DmSNAP190 and DmSNAP43. Furthermore, the conserved Myb

13 14 15

16

17 18 19

domain of DmSNAP190-(176-451) can associate wi DmSNAP43. On the other hand, this conserved region DmSNAP190 is insufficient for its assembly with DmSNAP5



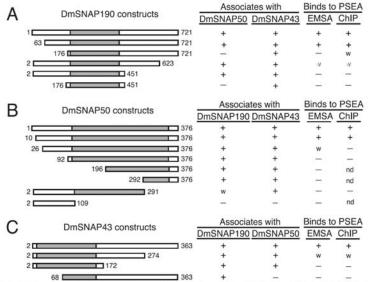


FIGURE 9. Summary of the effects of DmSNAP truncations on DmSNAPc assembly and DNA-binding activity. The various DmSNAP truncations analyzed are shown on the left side of the figure. The columns on the right indicate whether the specific truncation construct can associate with the other two DmSNAP subunits and whether the DmSNAP complex that contains the truncation construct can bind sequence specifically to the PSEA as assayed either by electrophoretic mobility shift assay or ChIP. A w indicates weak association or binding. nd indicates "not done," reflecting the fact that these represent more extensive truncations beyond constructs that already lacked DNA-binding activity.

Additional amino acid residues located just N-terminal to the Myb domain of DmSNAP190 (between positions 63 and 176) are required for DmSNAP50 assembly into the complex. In all other cases, however, the regions of the proteins that are the most evolutionarily conserved mediate assembly of the DmSNAPc subunits.

Furthermore, all DmSNAP mutations that impaired complex assembly also impaired DNA-binding activity both *in vitro* and *in vivo* (Fig. 9). This was to be expected because all three subunits are required for sequence-specific DNA binding by metazoan SNAPc (21–23).<sup>4</sup>

Somewhat surprisingly, mutations in the non-conserved regions of the DmSNAP subunits often resulted in the loss of DNA-binding activity, even when subunit assembly was not affected (Fig. 9). For example, deletion of DmSNAP190 residues beyond position 623 severely compromised DNA binding by the complex, and deletion of residues beyond position 451

destroyed the ability of the complex to bind to DNA. Similarly, deletion of non-conserved DmSNAP50 residues between positions 10 and 92 resulted in a loss of DNA-binding activity. Finally, deletion of amino acid residues following position 274 of DmSNAP43 significantly weakened the DNA binding activity of the complex, and deletion beyond position 172 of DmSNAP43 eliminated DNA binding. Thus, the evolutionarily non-conserved regions of the DmSNAP subunits are in many cases essential for the DNA binding activity of DmSNAPc. Whether these domains directly contact DNA, or whether they are required for DmSNAPc to adopt a conformation required for DNAbinding activity, remains an open question. However, considering that DmSNAPc contacts over 40 base pairs of DNA on the U1 promoter (14) and that all three subunits contact DNA (13), it is entirely possible that the less-conserved domains are involved in making direct contacts to the DNA. This suggests that the requirements for

subunit-subunit interactions within metazoan SNAPc are more constrained evolutionarily than the protein-DNA interactions. This agrees with our recent observations that the DNA sequences of the 3′ half of the PSEAs have changed fairly rapidly during insect evolution (19), suggesting that interactions between the protein and the DNA are more free to co-evolve.

Comparison to Findings with Human SNAPc—A number of studies have been published that address the roles of domains within the human SNAP subunits (21–24, 27). Sometimes our results with fly SNAPc have lead to conclusions similar to those reached in the human system; in other cases the fly data expand our knowledge and provide new insights; at still other times unexpected differences have been revealed. Some of the main similarities and differences are summarized in the following paragraphs.

FIGURE 8. **Effect of DmSNAP mutations on the** *in vivo* **occupancy of a U1 snRNA gene promoter by DmSNAPc.** *A*, schematic diagram of the endogenous *D. melanogaster U1:95Ca* promoter. Positions of primers utilized for ChIP are indicated by *arrows. B*, ChIP from S2 cell lines that co-express full-length or truncated FLAG-tagged DmSNAP190 constructs (as shown in Fig. 2A) with full-length untagged DmSNAP50 and DmSNAP43. Antibody against the FLAG epitope (*acFLAG*) was used to detect the presence of tagged DmSNAPc at the *U1:95Ca* promoter. Antibody against DmSNAP43 (*ac43*), which recognizes both endogenous and tagged DmSNAPc was used as a positive control. DmSNAP43 preimmune serum (*43P*) was used as a negative control. Positive PCR controls that utilized 20% of the unselected total input DNA (total) were also included. The *upper panel* shows ChIP results from S2 cell lines expressing DmSNAP190 constructs tagged at the C terminus (N-terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP190 constructs tagged at the N terminus (C-terminal truncations). C, ChIP from S2 cell lines that co-express full-length or truncated FLAG-tagged DmSNAP50 constructs (as shown in Fig. 4A) with full-length untagged DmSNAP30 and DmSNAP43. Antibodies and other conditions were the same as in *B*. The *upper panel* shows ChIP results from S2 cell lines expressing DmSNAP50 constructs tagged at the C terminus (N-terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs (as shown in Fig. 6A) with full-length untagged DmSNAP43 constructs tagged at the N terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs tagged at the N terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs tagged at the N terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs tagged at the N terminal truncations)

Fly SNAP190 Versus Human SNAP190—Figs. 2 and 9 show that the conserved Myb domain of DmSNAP190 interacts with DmSNAP43. In contrast, no interaction was observed between HsSNAP43 and the Myb domain of HsSNAP190 (27). In the human system, an additional subunit, HsSNAP19, is required for stable association between HsSNAP190 and HsSNAP43, and this interaction requires a small region (residues 84–138) in the non-conserved N terminus of HsSNAP190 (27). Our finding that DmSNAP43 interacts with the Myb domain of DmSNAP190 likely reveals an actual difference between the fly and human systems as no ortholog of HsSNAP19 has been identified in the fruit fly. The lack of strong interactions of HsSNAP43 with the Myb domain of HsSNAP190 may be compensated in the human system by interactions involving HsSNAP19 (and HsSNAP43) with residues 84–138 of HsSNAP190 (27).

We have determined in the fly that amino acids of DmSNAP190 located between positions 63 and 176, just upstream of the Myb domain, are required for association with DmSNAP50 (Figs. 2 and 9). We are not aware of any comparable experiments reported in the human system that localize a region of HsSNAP190 required for interaction with HsSNAP50. In fact available evidence in the human system indicates that HsSNAP190 and HsSNAP50 do not directly interact (11, 21, 23, 27).

Even more surprisingly, we found that the non-conserved C-terminal domain of DmSNAP190 is essential for the binding of DmSNAPc to the PSEA (Figs. 3 and 8). This was unexpected because the C-terminal domain of HsSNAP190 is completely dispensable for sequence-specific DNA binding in the human system (21, 22, 24, 27, 28). In fact, the C-terminal domain of HsSNAP190 serves to dampen DNA binding by HsSNAPc (21). In contrast, in flies it appears that the C-terminal domain of DmSNAP190 is required for the stable binding of DmSNAPc to the DNA.

Fly SNAP50 Versus Human SNAP50-Domains of Hs-SNAP50 involved in interacting with HsSNAP190 have not been reported in the human system. In fact, as mentioned above, there is no published data supporting a direct interaction between HsSNAP190 and HsSNAP50. However, our results (Figs. 4 and 9 and data not shown) indicate that sequences at the C terminus of DmSNAP50 (residues 292-376 that form part of the highly conserved SNAP finger domain) are sufficient for interaction with DmSNAP190. Jawdekar et al. (23) did not report any interaction of this region of HsSNAP50 with Hs-SNAP190. Instead, their work emphasized the DNA-binding activity of the SNAP finger domain. More than likely, the SNAP finger domain of DmSNAP50 is involved both in DNA binding and in interaction with DmSNAP190. The SNAP finger of DmSNAP50 also interacts with DmSNAP43 (Figs. 4 and 9 and data not shown). However, the central region of DmSNAP50 (residues 110-291) also participates in complex assembly with DmSNAP43 (Figs. 4 and 9). This is the same region of SNAP50 that was found to interact with SNAP43 in the human system (23).

It was not surprising that the C-terminal truncation DmSNAP50-(2–291) that deleted a portion of the SNAP finger domain eliminated DNA-binding activity. However, DNA-

binding activity was also ablated by a DmSNAP50 N-terminal truncation that deleted amino acids 1–91 even though this construct assembled efficiently with DmSNAP190 and DmSNAP43 (Fig. 9). Thus a portion of the non-conserved N terminus (as well as the C terminus) of DmSNAP50 is required for DNA binding. Comparable studies targeting the N-terminal function of HsSNAP50 have to our knowledge not been reported.

Fly SNAP43 Versus Human SNAP43—Results shown in Figs. 6 and 9 implicate DmSNAP43 residues located between positions 68 and 172 as those most likely involved in interaction with DmSNAP190. On the other hand, Ma and Hernandez (27) found that residues 164–268 of HsSNAP43 (which are equivalent to residues 179–283 of DmSNAP43) were sufficient for association with HsSNAP190 together with HsSNAP19. As mentioned above, HsSNAP43 does not interact strongly with HsSNAP190 in the absence of HsSNAP19. Therefore, the mapping of different regions in fly and human SNAP43 involved in association with SNAP190 could reflect the additional requirement and role for HsSNAP19 for stable association in the human system.

Our results indicate that the first 172 amino acids of DmSNAP43 are sufficient to recruit DmSNAP50 into the complex and that the first 68 amino acids of DmSNAP43 are required for this interaction. Ma and Hernandez (27), who employed very similar truncations of HsSNAP43, obtained nearly identical results in the human system. On the other hand, Hinkley *et al.* (22), also working in the human system, found that the C-terminal as well as the N-terminal half of HsSNAP43 could interact with HsSNAP50 in a glutathione S-transferase pulldown assay. It is possible that the conditions utilized by us and by Ma and Hernandez (27) are more stringent and detect only stronger interactions among the SNAP subunits.

Deletion of DmSNAP43 amino acids beyond position 274 reduced but did not completely eliminate sequence-specific binding of DmSNAPc to the PSEA. However, DNA-binding activity was undetectable following further truncation of DmSNAP43 to position 172 (Figs. 7–9). Results essentially identical to these were obtained by Ma and Hernandez (27) when they used truncations that were analogous to our truncations of DmSNAP43. However, Hinkley et al. (22) found that either the N-terminal half or the C-terminal half of HsSNAP43 could be incorporated into mini-SNAPc complexes that were capable of binding to DNA. The reason for the different findings within the human system is not clear. However, our data with fly SNAP43 parallel more closely the findings of Ma and Hernandez (27).

Conclusion—We have co-expressed full-length and truncated forms of all three DmSNAPc subunits in homologous S2 cells to determine the domains required for in vivo assembly of DmSNAPc and for DNA binding both in vitro and in vivo. With one exception, the regions evolutionarily conserved among animals are sufficient for complex assembly. However, non-conserved regions of the proteins are required for sequence-specific DNA binding. The ability to compare results in the human and fly systems has in several cases allowed us to identify domains that function similarly in both systems. However, in other cases novel insights into metazoan SNAPc functional

domains have been obtained, and differences between the two systems suggest possible alternative mechanisms for achieving complex assembly and stable DNA binding.

Acknowledgments—We thank Anna D'Agostini, Kathleen McNamara-Schroeder, and David Almeida for assistance in plasmid construction and preparation.

#### REFERENCES

- Murphy, S., Yoon, J. B., Gerster, T., and Roeder, R. G. (1992) Mol. Cell. Biol. 12, 3247–3261
- Sadowski, C. L., Henry, R. W., Lobo, S. M., and Hernandez, N. (1993) Genes Dev. 7, 1535–1548
- Yoon, J. B., Murphy, S., Bai, L., Wang, Z., and Roeder, R. G. (1995) Mol. Cell. Biol. 15, 2019 – 2027
- Henry, R. W., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1995) Nature 374, 653–656
- Waldschmidt, R., Wanandi, I., and Seifart, K. H. (1991) EMBO J. 10, 2595–2603
- Wanandi, I., Waldschmidt, R., and Seifart, K. H. (1993) J. Biol. Chem. 268, 6629 – 6640
- 7. Yoon, J. B., and Roeder, R. G. (1996) Mol. Cell. Biol. 16, 1-9
- Sadowski, C. L., Henry, R. W., Kobayashi, R., and Hernandez, N. (1996) Proc. Natl. Acad. Sci. U.S.A. 93, 4289 – 4293
- Henry, R. W., Ma, B., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1996) EMBO J. 15, 7129 –7136
- Bai, L., Wang, Z., Yoon, J. B., and Roeder, R. G. (1996) Mol. Cell. Biol. 16, 5419 –5426

- Wong, M. W., Henry, R. W., Ma, B., Kobayashi, R., Klages, N., Matthias, P., Strubin, M., and Hernandez, N. (1998) Mol. Cell. Biol. 18, 368–377
- Henry, R. W., Mittal, V., Ma, B., Kobayashi, R., and Hernandez, N. (1998) Genes Dev. 12, 2664–2672
- 13. Wang, Y., and Stumph, W. E. (1998) Mol. Cell. Biol. 18, 1570-1579
- Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (2004) Mol. Cell. Biol. 24, 1897–1906
- Lai, H. T., Kang, Y. S., and Stumph, W. E. (2008) FEBS Lett. 582, 3734–3738
- 16. Das, A., and Bellofatto, V. (2003) Proc. Natl. Acad. Sci. U.S.A. 100, 80 85
- Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A., and Bellofatto. V. (2005) Mol. Cell. Biol. 25, 7314

  –7322
- Schimanski, B., Nguyen, T. N., and Günzl, A. (2005) Mol. Cell. Biol. 25, 7303–7313
- Hernandez, G., Jr., Valafar, F., and Stumph, W. E. (2007) Nucleic Acids Res. 35, 21–34
- Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L., and Stumph, W. E. (1997) Eur. J. Biochem. 248, 231–237
- 21. Mittal, V., Ma, B., and Hernandez, N. (1999) Genes Dev. 13, 1807-1821
- Hinkley, C. S., Hirsch, H. A., Gu, L., LaMere, B., and Henry, R. W. (2003)
   J. Biol. Chem. 278, 18649 18657
- Jawdekar, G. W., Hanzlowsky, A., Hovde, S. L., Jelencic, B., Feig, M., Geiger, J. H., and Henry, R. W. (2006) J. Biol. Chem. 281, 31050 –31060
- 24. Ma, B., and Hernandez, N. (2002) Mol. Cell. Biol. 22, 8067-8078
- Zamrod, Z., Tyree, C. M., Song, Y., and Stumph, W. E. (1993) Mol. Cell. Biol. 13, 5918 – 5927
- 26. Barakat, N. H., and Stumph, W. E. (2008) FEBS Lett. 582, 2413-2416
- 27. Ma, B., and Hernandez, N. (2001) J. Biol. Chem. 276, 5027-5035
- Hanzlowsky, A., Jelencic, B., Jawdekar, G., Hinkley, C. S., Geiger, J. H., and Henry, R. W. (2006) Protein Expr. Purif. 48, 215–223

### **ACKNOWLEDGEMENTS:**

This Chapter, in full, is a reprint of the material as it appears in The Journal of Biological Chemistry, 2009. A map of *Drosophila melanogaster* small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. Hung, K. H.; Titus, M.; Chiang, S. C.; Stumph, W. E. The dissertation author was the primary researcher and author of this paper. Plasmid construction, low-salt protein purification and subsequent immunoblotting and EMSA analysis of DmSNAP50 truncations were carried out by Shu-Chi Chiang, whereas plasmid construction and analysis of low-salt purified complexes containing DmSNAP190 truncations were done by Mitchell Titus. All other data in this paper were produced by this dissertation author.

### **CHAPTER 2**

Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc)



#### **REVIEW ARTICLE**

### Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc)

Ko-Hsuan Hung1, and William E. Stumph2

<sup>1</sup>Department of Biology, and <sup>2</sup>Department of Chemistry and Biochemistry and Molecular Biology Institute, San Diego State University, San Diego, CA, USA

#### **Abstract**

The small nuclear RNAs (snRNAs) are an essential class of non-coding RNAs first identified over 30 years ago. Many of the well-characterized snRNAs are involved in RNA processing events. However, it is now evident that other small RNAs, synthesized using similar mechanisms, play important roles at many stages of gene expression. The accurate and efficient control of the expression of snRNA (and related) genes is thus critical for cell survival. All snRNA genes share a very similar promoter structure, and their transcription is dependent upon the same multi-subunit transcription factor, termed the snRNA activating protein complex (SNAPc). Despite those similarities, some snRNA genes are transcribed by RNA polymerase II (Pol II), but others are transcribed by RNA polymerase III (Pol III). Thus snRNA genes provide a unique opportunity to understand how RNA polymerase specificity is determined and how distinct transcription machineries can interact with a common factor. This review will describe efforts taken toward solving those questions by using the fruit fly as a model organism. Drosophila melanogaster SNAPc (DmSNAPc) binds to a proximal sequence element (PSEA) present in both Pol II and Pol III snRNA promoters. Just a few differences in nucleotide sequence in the Pol II and Pol III PSEAs play a major role in determining RNA polymerase specificity. Furthermore, these same nucleotide differences result in alternative conformations of DmSNAPc on Pol II and Pol III snRNA gene promoters. It seems likely that these DNA-induced alternative DmSNAPc conformations are responsible for the differential recruitment of the distinct transcriptional machineries.

**Keywords:** protein–DNA interaction; transcription initiation; pre-initiation complex; RNA polymerase specificity; promoter sequences; site-specific protein–DNA photo-cross-linking

### Introduction

The small nuclear RNA activating protein complex (SNAPc) is a unique multi-subunit protein complex required for the synthesis of small nuclear RNAs (snRNAs) (Murphy et al., 1992; Sadowski et al., 1993; Yoon et al., 1995; Henry et al., 1995; Su et al., 1997; Li et al., 2004). The snRNAs are non-coding RNA molecules highly expressed in eukaryotic cells, and each is a product of its own independent transcription unit. The snRNAs are involved in many essential cellular functions such as pre-mRNA splicing, rRNA processing and histone mRNA 3' end-formation (Steitz et al., 1988; Kass et al., 1990; Bond et al., 1991; Guthrie, 1991; Peculis and Steitz,

1993; Sharp, 1994). In animals, most snRNAs (e.g. U1, U2, U3, U4, U5 and U7) are synthesized by RNA polymerase II (Pol II), but other small RNAs (e.g. U6 snRNA, 7SK RNA, tRNA<sup>sec</sup>, H1 RNA, and MRP RNA) are synthesized by RNA polymerase III (Pol III) (Zieve *et al.*, 1977; Dahlberg and Lund, 1988; Lee *et al.*, 1989; Parry *et al.*, 1989; Baer *et al.*, 1990; Yuan and Reddy, 1991; Hernandez, 1992; Lobo and Hernandez, 1994).

Interestingly, despite this differential requirement of RNA polymerase, the genes coding for all these small RNAs share a very similar promoter structure. In animals, snRNA promoters are characterized by a unique and essential upstream promoter element termed the proximal sequence element (PSE) (more specifically called

Address for Correspondence: William E. Stumph, Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA 92182-1030, USA. Tel: +1 619 594 5575; E-mail: wstumph@sciences.sdsu.edu

the PSEA in Drosophila melanogaster) (Das et al., 1987; Dahlberg and Lund, 1988; Zamrod et al., 1993; Lobo and Hernandez, 1994). Despite this commonality, evidence indicates that the mechanisms involved in determining RNA polymerase specificity can be different in evolutionarily divergent organisms. In vertebrates, a TATA box present downstream of the PSE acts as a dominant determinant of Pol III specificity, whereas the absence of the TATA box results in snRNA gene transcription by Pol II (Mattaj et al., 1988; Lobo and Hernandez, 1989). In flies, on the other hand, the exact sequence of the PSEA itself acts as the primary determinant of RNA polymerase specificity (Jensen et al., 1998; McNamara-Schroeder et al., 2001; Lai et al., 2005). As described in greater detail in the following section, plant snRNA genes utilize a still different mechanism to determine polymerase specificity.

The PSEs of all snRNA genes (whether Pol II or Pol III-specific) are recognized and bound by the same evolutionarily conserved transcription factor, SNAPc. The interaction between the PSE and SNAPc initiates the recruitment of Pol II- or Pol III-specific factors for snRNA transcription (Sadowski et al., 1993; Kuhlman et al., 1999; Schramm et al., 2000; Teichmann et al., 2000; Cabart and Murphy, 2001; 2002; Das et al., 2005; Schimanski et al., 2005; Lee et al., 2007; Barakat and Stumph, 2008). However, as mentioned above, the pathway for achieving RNA polymerase specificity can vary among distantly related organisms. This unusual scenario makes the snRNA genes an intriguing system for investigating how RNA polymerase specificity is determined and how a common factor (SNAPc in this case) is able to recruit different transcription machineries.

The transcription of animal snRNA genes has been studied most thoroughly in vertebrates (particularly the human system) and in the fruit fly D. melanogaster. Interestingly, fruit fly snRNA gene promoters exhibit a higher degree of conservation with regard to both sequence and location of promoter elements than generally observed in other animals, particularly vertebrates (Dahlberg and Lund, 1988; Lo and Mount, 1990; Hernandez et al., 2007; Jawdekar and Henry, 2008). As a result, the specific molecular interactions that govern snRNA gene activation and RNA polymerase specificity may be more apparent and more accessible for study in fruit flies in comparison to other systems. Reviews on the transcriptional regulation of human snRNA genes have recently been published by Jawdekar and Henry (2008) and by Egloff et al. (2008). This article will therefore concentrate on mechanisms of snRNA transcription in the fruit fly while attempting to place this information into the context of the knowledge available from other

First, the structure of snRNA (and snRNA-like) promoters and the contribution of these promoter sequences to

determining RNA polymerase specificity will be reviewed with an emphasis on the fruit fly. This will be followed by a discussion of the evolutionarily conserved and nonconserved structural features of DmSNAPc in comparison to the homologous proteins from other organisms. Next the mapping and localization of functional domains within each of the DmSNAPc subunits will be described and compared to results published in the human system. We will then review evidence that DmSNAPc adopts different conformational states on Pol II and Pol III snRNA promoters as an allosteric effect of DNA sequence recognized. Finally, we will speculate on how these conformational differences of the DmSNAPc-DNA complex may lead to RNA polymerase specificity on Pol II and Pol III snRNA gene promoters.

### Structure and RNA polymerase specificity of Drosophila snRNA gene promoters

Early comparisons of DNA sequences identified conserved blocks of sequence upstream of cloned *D. melanogaster* snRNA genes as putative promoter elements (Beck *et al.*, 1984; Saba *et al.*, 1986; Das *et al.*, 1987; Lo and Mount, 1990). The functionality of these conserved elements was subsequently demonstrated by *in vitro* and *in vivo* transcription assays of mutated versus wild-type templates (Zamrod *et al.*, 1993; Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). Figure 1A schematically shows the promoter structure of fly snRNA genes transcribed either by Pol II or by Pol III. For purposes of comparison, the positions of the analogous Pol II and Pol III snRNA gene promoter elements in vertebrates and plants are illustrated in the lower parts of Figure 1A.

The PSEA is a more specific name for the insect PSE and is a unique ~21 base pair (bp) element located within a region 40-65 nucleotides upstream of the transcription start site. It was named the PSEA to distinguish it from the PSEB, a moderately conserved 8bp promoter element located downstream of the PSEA in the insect Pol II-transcribed snRNA genes (Figure 1A). In contrast, the Pol III-transcribed snRNA genes possess a strongly conserved 8 bp TATA box instead of the PSEB downstream of the PSEA (Figure 1A). Interestingly, there is an 8 bp separation of the PSEB from the PSEA but a 12 bp separation of the TATA sequence from the PSEA, and these distinctive separations are strictly conserved among the Pol II and Pol III fly snRNA genes (Figure 1B). Like the vertebrate PSE (Dahlberg and Lund, 1988; Parry et al., 1989; Lobo and Hernandez, 1994), the fly PSEA is the dominant element for specifying the transcription start site and is essential for snRNA promoter activity (Zamrod and Stumph, 1990; Jensen et al., 1998; McNamara-Schroeder et al., 2001; Lai et al., 2005).

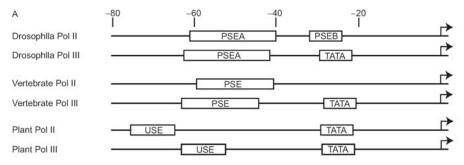


Figure 1. Promoter structure of snRNA genes. (A) Schematic representation of conserved promoter elements in the 5\( \textit{Z}\$-flanking DNA of snRNA genes transcribed either by Pol II or by Pol III in Drosophila, vertebrates, and plants. The numbers at the top indicate the distance in base pairs upstream of the transcription start site. Abbreviations: PSE, proximal sequence element; PSEA, proximal sequence element A; PSEB, proximal sequence element B; USE, upstream sequence element; TATA, TATA box. (B) Alignment and analysis of snRNA and snRNA-like gene promoters of D. melanogaster. The sequences of the promoter regions of 23 Pol II-transcribed D. melanogaster snRNA genes are shown in the upper section. The schematic diagram above the sequences indicates the boundaries of the Pol II PSEA and of the PSEB that were originally identified as 21 bp and 8 bp elements respectively. The shaded areas represent possible extensions of the PSEA based upon the more comprehensive set of sequences shown in the figure. The numbers below the schematic diagram (-10, 1, 21, 30, 37) indicate the base positions relative to the traditional first nucleotide of the PSEA. The numbers above the schematic (-61, -41, -32, -25) indicate the approximate distances upstream of the transcription start site, but these can vary by up to 2 bp depending upon the particular gene. The lower section shows the sequences of the promoter regions of four snRNA and three snRNA-like genes transcribed by Pol III. The schematic drawing at the bottom indicates the boundaries of the Pol III PSEA and TATA box. The shading and numbering are similar to that described above for the Pol II schematic drawing. Nucleotides shown in boldface type at PSEA positions 19 and 20 indicate nucleotides never found at the corresponding position in the promoter sequences of genes transcribed by the opposite RNA polymerase (Pol II versus Pol III). Immediately below the Pol II gene sequences and above the Pol III gene sequences are shown consensus promoter sequences for the Pol II- and Pol III-transcribed snRNA (and snRNA-like genes), respectively. The nucleotides overlined in the Pol II consensus sequence and those underlined in the Pol III consensus sequence indicate nucleotides that are 100% conserved in all aligned Pol II and Pol III promoter sequences, respectively. Between the Pol II and Pol III consensus promoter sequences, a Pol II/III consensus PSEA is shown. A Pol II/III consensus nucleotide is shown only if it is present in both individual consensus promoter sequences. The bold Xs indicate positions where different nucleotides are always (or nearly always) preferred in the Pol III versus Pol II gene PSEAs.

Figure 1. continued on next page

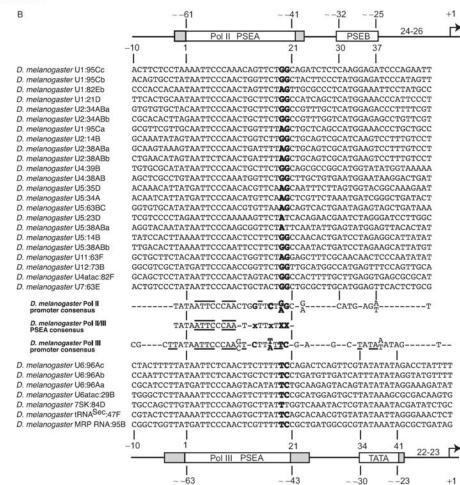
The extent of conservation of these elements is evident from the comparison of the sequences of a large set of snRNA and snRNA-like gene promoters from *D. melanogaster*, as shown in Figure 1B. A detailed sequence analysis of snRNA genes from six different insects revealed that this promoter structure is well conserved throughout the insect sub-phylum (Hernandez *et al.*, 2007). Generally, the insect PSEA is longer than the vertebrate PSE and exhibits a higher degree of sequence conservation.

Comparison between the sequences of the Pol II PSEAs and the Pol III PSEAs of *D. melanogaster* revealed significant differences as well as the obvious similarities. (Promoter consensus sequences, including the PSEA, are shown in the central area of Figure 1B.) It is notable that the 5' half of the PSEA is very highly conserved between the Pol II and Pol III promoters. (Seven nucleotides in the 5' half of the PSEAs are 100% conserved among the 30 genes listed.) In contrast, the Pol II and Pol III PSEAs diverge significantly from each other in their 3' halves especially at certain specific nucleotide positions, most prominently at positions 19 and 20 denoted by Xs in the Pol II/III consensus PSEA (Figure 1B). For example, position 19 is always an A or G in the Pol II PSEAs versus always a T in the Pol III PSEAs, and position 20 is

nearly always a G in the Pol II PSEAs but a C in the Pol III PSEAs (Figure 1B). These nucleotide positions have been said to be "conserved to be different" in the Pol II and Pol III PSEAs (Hernandez *et al.*, 2007). Other positions where there are often "conserved nucleotide differences" between the fly Pol II and the Pol III PSEAs are at positions 14 and 17 (denoted by lower case Xs in the Pol II/ III consensus PSEA, Figure 1B).

Remarkably, the divergent 3' half of the PSEA was found to play a key role in determining the RNA polymerase specificity of Drosophila snRNA genes. Altering as few as three nucleotide pairs in a U1 PSEA (including positions 19 and 20) to those found in a U6 PSEA completely switched the RNA polymerase specificity of the U1 promoter in vitro from Pol II to Pol III even in the absence of a TATA sequence (Jensen et al., 1998). Moreover, the PSEAs from Pol II and Pol III promoters are not interchangeable in vivo. Substituting a U6 PSEA into a U1 promoter (a total of five nucleotide changes) resulted in the complete loss of U1 gene promoter activity in vivo (McNamara-Schroeder et al., 2001; Lai et al., 2005; Barakat and Stumph, 2008). Similarly, a reciprocal substitution of the U1 PSEA into the U6 promoter completely inactivated the U6 promoter in living cells (McNamara-Schroeder et al., 2001; Lai et al., 2005). Thus, the U6 PSEA cannot function

Figure 1. Continued.



for Pol II transcription, and the U1 PSEA cannot function for Pol III transcription, even though they both bind DmSNAPc. Conversely, other experiments showed that swapping the PSEB and TATA box of fly U1 and U6 snRNA promoters affected primarily transcription efficiency, but not the RNA polymerase selectivity of these promoters (Jensen *et al.*, 1998; Lai *et al.*, 2005).

The results described in the paragraph above were very surprising because, in vertebrates, snRNA gene Pol II and Pol III PSEs had been reported to be interchangeable (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989). In vertebrates, the presence or absence of the TATA box (Figure 1A) was found to be the primary and dominant determinant of RNA polymerase specificity. For example, mutation of the vertebrate U6 TATA box to an unrelated sequence changed the promoter specificity to Pol II, while

introduction of a TATA sequence into the vertebrate U1 or U2 promoters altered their specificity to Pol III (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989).

In plants, the RNA polymerase specificity of snRNA genes is determined by a still different mechanism. In that case, both classes of promoters contain interchangeable USEs as well as interchangeable TATA boxes (Figure 1A). In this instance, RNA polymerase specificity is determined by the distance between the USE and the TATA box (Waibel and Filipowicz, 1990; Goodall *et al.*, 1991; Kiss *et al.*, 1991).

After the above differences in Pol II and Pol III snRNA gene promoters were discovered, the next question of general interest was the following: How are these differences in promoter structure read out by the transcription machinery to effect the recruitment of the

required polymerase? To seek answers to this question, it is necessary to have some knowledge of the structure and function of the small nuclear RNA activating protein complex, SNAPc.

### SNAPc: the key regulator of snRNA transcription

The PSEs of both Pol II- and Pol III-transcribed snRNA genes are recognized by the same multi-subunit transcription factor, SNAPc. SNAPc, also known as PSEbinding transcription factor (PTF) and PSE-binding protein (PBP), was first identified in the human system (Waldschmidt et al., 1991; Murphy et al., 1992; Sadowski et al., 1993). Transcription of both Pol II and Pol IIItranscribed snRNA genes was shown to be dependent upon SNAPc. Human SNAPc (HsSNAPc) contains five distinct polypeptide chains (HsSNAP190 or PTFα, HsSNAP50 or PTFB, HsSNAP45 or PTFb, HsSNAP43 or PTFy, and SNAP19) for which the HsSNAP nomenclature reflects the apparent molecular weights of these subunits (Henry et al., 1995; 1998; Yoon and Roeder, 1996). Significantly, a complex containing only the three subunits HsSNAP190, HsSNAP50, and HsSNAP43 was sufficient to reconstitute sequence-specific DNA binding as well as the basal transcription activity of human SNAPc (Mittal et al., 1999; Ma and Hernandez, 2001; 2002; Hinkley et al., 2003; Jawdekar et al., 2006). Thus, these three subunits represent the "core subunits" of HsSNAPc required for pre-initiation complex assembly on snRNA genes. The other two subunits, HsSNAP45 and HsSNAP19, may play roles in regulation of SNAPc activity and complex stability (Henry et al., 1998; Mittal et al., 1999; Ma and Hernandez, 2001).

In the *Drosophila* system, DmSNAPc was first identified in a soluble nuclear fraction prepared from fly embryos. Partially purified DmSNAPc (originally termed DmPBP) exhibited sequence-specific PSEA-binding activity and was capable of stimulating PSEA-dependent activation of U1 and U6 snRNA gene transcription (Su *et al.*, 1997). The fly genome contains genes that code for proteins homologous to the HsSNAP190, HsSNAP50, and HsSNAP43 subunits, but no recognizable genes capable of encoding fly homologs of HsSNAP45 or HsSNAP19 have been detected (Li *et al.*, 2004). The evolutionary conservation of the SNAP190, SNAP50, and SNAP43 subunits further argues that these three subunits comprise the essential core of SNAPc.

In fact, orthologs of these three subunits have even been characterized in the anciently diverged trypanosomes, where tSNAPc is required for Pol II transcription of the spliced leader RNA, a small nuclear RNA that is trans-spliced onto the 5' end of trypanosomal mRNAs (Huie *et al.*, 1997; Luo *et al.*, 1999; Das and Bellofatto, 2003; Schimanski *et al.*, 2004; 2005; Das *et al.*, 2005). The

existence of tSNAPc indicates that the SNAP complex appeared very early in eukaryotic evolution and continues to be essential for snRNA transcription in diverse contemporary eukaryotes. A comparison of the structural features of the three orthologous fly, human, and trypanosomal (*T. brucei*) SNAPc subunits is presented in Figure 2. The shading indicates the regions most evolutionarily conserved among the three organisms shown.

In flies and humans, the most conserved region of the largest subunit, SNAP190, is a unique domain that consists of 4.5 tandem Myb repeats, termed respectively Rh, Ra, Rb, Rc, and Rd (Wong et al., 1998; Li et al., 2004). Myb repeats were first identified in the Myb oncoprotein and are involved in DNA binding (Klempnauer and Sippel, 1987; Biedenkapp et al., 1988). In contrast to the 4.5 Myb repeats found in animal SNAP190, all other known Myb-domain proteins (to our knowledge) contain only one to three Myb repeats (reviewed in Rosinski and Atchley1998). Thus, the binding of the Myb repeats of SNAP190 to DNA is likely to be more complicated than that of proteins that contain fewer Myb repeats. The overall length of the fly protein is only about half the length of the human protein. This most likely reflects the presence of additional functional domains in the C-terminal region of the human protein such as those important for interaction with the HsSNAP45 subunit and with the enhancerbinding protein Oct-1 (Ford et al., 1998; Mittal et al., 1999). Interestingly, the trypanosomal SNAP190 protein is still shorter in overall length (about half the length of DmSNAP190 and quarter the length of HsSNAP190); furthermore, it contains only 2.5 identifiable Myb repeats that align best with the human and fly Ra (C-terminal half), Rb, and Rc repeats (Schimanski et al., 2005).

The SNAP50 orthologs are the most evolutionarily conserved of the SNAPc subunits (Das and Bellofatto, 2003; Li et al., 2004; Jawdekar et al., 2006). Perhaps the most unusual feature of the SNAP50 protein is that the C-terminal domain consists of a unique non-canonical zinc finger (specifically named the "SNAP finger") (Bai et al., 1996; Henry et al., 1996; Das and Bellofatto, 2003; Li et al., 2004; Jawdekar et al., 2006). This domain contains seven conserved cysteine or histidine residues that when changed to alanine significantly reduced both zinc-binding and DNA-binding by human SNAPc (Jawdekar et al., 2006). However, it should be emphasized that the SNAP finger is unique and has no detectable homology to other well-characterized zinc finger DNA-binding domains.

SNAP43 is probably the least characterized of the SNAPc subunits. The most evolutionarily conserved region lies toward the N terminus of SNAP43. Neither this region nor the non-conserved region of SNAP43 has any clear homology to other proteins in existing databases. However, as described further below, SNAP43 may play a critical role in the determination of RNA polymerase specificity at snRNA promoters.

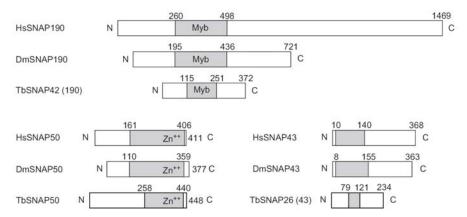


Figure 2. Comparison of *D. melanogaster* (Dm), human (Hs), and *Trypanosoma brucei* (Tb) SNAPc subunits. The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous proteins. The numbers above and below the rectangles designate the amino acid positions in the proteins. The most conserved region in SNAP190 contains a Myb domain, which consists of 4.5 tandem Myb repeats in fly and human SNAP190 (Wong *et al.*, 1998; Li *et al.*, 2004), but just 2.5 Myb repeats in the trypanosome ortholog, TbSNAP42 (Schimanski *et al.*, 2005). This region of DmSNAP190 shares 27% identity and 44% similarity with HsSNAP190, but 15% identity and 57% similarity with TbSNAP42. For DmSNAP50, the most conserved region is located at the C terminus (residues 110 to 359), which includes an unorthodox zinc-binding domain (noted as Zn\*\* in the figure) termed the "SNAP finger". This region shares 33% identity and 51% similarity with HsSNAP50, but 24% identity and 35% similarity with TbSNAP50. The most conserved region of DmSNAP43 resides in the N terminus (residues 8 to 155), which shares 31% identity and 48% similarity with HsSNAP43, but has 32% identity and 50% similarity to a shorter region (residues 79 to 121) of its trypanosome ortholog TbSNAP36.

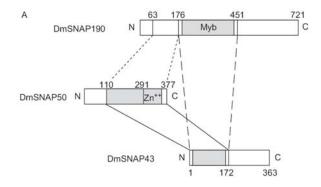
Although it was known for quite some time that HsSNAPc and DmSNAPc were heteromeric complexes that contained a number of distinct polypeptides, the stoichiometry of the three core subunits in the complex was not absolutely clear (although often portrayed as 1:1:1). By using a mixture of tags on individual subunits and employing mobility supershift analysis with monoclonal antibodies against the tags, convincing evidence was recently obtained that the DmSNAP190, DmSNAP50, and DmSNAP43 subunits are each present in a single copy in native DmSNAPc bound to DNA (Lai *et al.*, 2008).

The SNAPc subunits co-purify with each other in solution, thus indicating a tight association with each other even when the complex is not associated with DNA (Murphy et al., 1992; Sadowski et al., 1993; Yoon et al., 1995; Su et al., 1997; Das and Bellofatto, 2003; Li et al., 2004; Das et al., 2005; Schimanski et al., 2005). Although the isolated Rc and Rd repeats of HsSNAP190 can bind weakly but apparently without sequence specificity to DNA (Wong et al., 1998; Ma and Hernandez, 2002; Hinkley et al., 2003), all three core subunits of both human and fly SNAPc are essential for sequence-specific binding to the PSE(A). None of the three subunits can bind to the PSE(A) either individually or in any pair-wise combinations (Mittal et al., 1999; Jawdekar et al., 2006 and our unpublished observations). It is also clear from protein-DNA photo-cross-linking studies that each of the three core subunits, at least in flies, makes direct contact with the DNA and thereby contributes to the DNA-binding activity of the complex (Wang and Stumph, 1998; Li et al., 2004; Kim *et al.*, 2010). The photo-cross-linking studies are described in a later section of this review.

### DmSNAPc subunit domains involved in DmSNAPc assembly

Mutational analyses have been used to identify domains within each of the core subunits that are required for complex formation with each of the other two subunits. Such studies have been carried out in both the fly (Hung et al., 2009) and human systems (Mittal et al., 1999; Ma and Hernandez, 2001; 2002; Hinkley et al., 2003; Jawdekar et al., 2006). Although there is a good deal of similarity in the findings from the two organisms, there also appear to be several significant differences.

Figure 3A schematically indicates the mapped domains within the fly proteins that are involved in subunit-subunit interactions. One of the main conclusions from those findings was that, with one exception, the evolutionarily most-conserved region of each DmSNAPc subunit was sufficient for its association with the other two subunits (Hung *et al.*, 2009). For example, the conserved Myb domain of DmSNAP190 was sufficient for its interaction with DmSNAP43. Furthermore, two adjacent regions within the conserved C-terminal region of DmSNAP50 (residues 110–291 and residues 292–377 respectively, which includes the SNAP finger) each interacted with both DmSNAP190 and DmSNAP43 (Hung *et al.*, 2009). That is, DmSNAP50 residues 292–377,



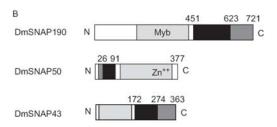


Figure 3. DmSNAPc subunit domains required for complex assembly and DNA binding. The lightly shaded area of each subunit represents the evolutionarily conserved region as shown in Figure 2. The numbers above and below the rectangles indicate the amino acid positions in the subunits where truncations were made to produce mutant proteins. (A) Map of DmSNAPc subunit-subunit interactions. The finely dashed lines indicate domains involved in interaction between DmSNAP190 and DmSNAP190 and DmSNAP190. The longer dashed lines indicate the domains sufficient for interaction between DmSNAP190 and DmSNAP43, and the solid lines indicate regions of interaction between DmSNAP50 and DmSNAP43 (Hung et al., 2009). (B) Map of subunit domains required for DmSNAPc DNA-binding activity. The lightly shaded areas represent the most conserved region in each subunit; these regions are required for subunit assembly and thus necessarily for DNA binding by DmSNAPc. The moderately shaded areas represent subunit domains not required for complex assembly but, when deleted, reduce (but do not eliminate) the DNA binding activity of DmSNAPc. The darkly shaded areas, when deleted, eliminate detectable DNA binding by DmSNAPc while not interfering with subunit assembly (Hung et al., 2009).

which included the SNAP finger, interacted strongly with both DmSNAP43 and DmSNAP190. On the other hand, DmSNAP50 residues 110–291 interacted strongly with DmSNAP43 but more weakly with DmSNAP190. Finally, the conserved domain at the N terminus of DmSNAP43 (residues 1–172) was sufficient to associate with both DmSNAP190 and DmSNAP50. The only exception to the "conserved-domain rule" was that an evolutionarily nonconserved region of DmSNAP190 (residues 63–175) was required for interaction with DmSNAP50.

Comparison of these findings with the results of studies on human SNAPc revealed some important similarities and a number of surprising differences. As an example of similarity, in both organisms the conserved C-terminal region of SNAP50 interacts with the conserved N-terminal region of SNAP43 (Jawdekar *et al.*, 2006; Hung *et al.*, 2009). A significant difference, on the other hand, is that the strong interaction between DmSNAP190 and DmSNAP50 in flies has not been observed between the two homologous subunits of the human system. In

fact available evidence suggests that there is no direct interaction between HsSNAP190 and HsSNAP50 (Wong *et al.*, 1998; Mittal *et al.*, 1999; Ma and Hernandez, 2001; Jawdekar *et al.*, 2006). The reason for this difference between the two systems is not clear.

Another significant difference is that Hung et al. (2009), working in the fly system, observed a strong interaction between the conserved N-terminal domain of DmSNAP43 and the Myb domain of DmSNAP190, but no interactions were observed between the analogous regions of the orthologous human subunits (Ma and Hernandez, 2001). Instead, an interaction was mapped between the nonconserved N-terminal region of HsSNAP190 and a nonconserved centrally-located region of HsSNAP43 (Ma and Hernandez, 2001). Interestingly, in the human system, stable association between HsSNAP190 and HsSNAP43 requires the additional subunit, HsSNAP19, which is not present in the fruit fly (Ma and Hernandez, 2001). Thus, it is quite possible that the lack of a strong direct interaction between the evolutionarily conserved regions

of HsSNAP190 and HsSNAP43 may be compensated in humans by the presence of HsSNAP19, which may act via non-analogous domains of the human subunits.

### DmSNAPc subunit domains required for DNA binding to the PSEA

Because the evolutionarily conserved regions of DmSNAPc are required for DmSNAPc assembly, the DNA-binding activity of DmSNAPc is also dependent upon those conserved regions of the subunits. But deletions within the non-conserved regions can also lead to reductions in DmSNAPc DNA binding activity while still allowing the assembly of all three subunits. In the case of fly SNAPc, even relatively short truncations of the DmSNAPc subunits often had a noticeable effect on DNA binding activity (Hung et al., 2009).

Figure 3B indicates subunit domains required for effective DNA binding by DmSNAPc but not necessary for assembly of the three-subunit complex. As in previous figures, the lightest shading indicates the evolutionarily conserved region of each protein. The intermediate shading in Figure 3B indicates a region in each subunit that, when deleted, significantly decreased but did not completely eliminate DNA binding. However, the subsequent deletion of the darkly shaded region within any one of the subunits eliminated the DNA binding activity of the truncated DmSNAPc.

In the case of DmSNAP190, a C-terminal truncation following residue 623 greatly weakened but did not completely eliminate DNA-binding activity; however, a truncation following residue 451 of DmSNAP190 completely eliminated detectable DNA binding by DmSNAPc (Hung et al., 2009). This was a very unexpected result when compared to findings in the human system because a human "mini-SNAPc", that completely lacked the non-conserved C-terminal region of HsSNAP190, bound very efficiently and with high specificity to PSE sequences (Mittal et al., 1999; Ma and Hernandez, 2001; 2002; Hinkley et al., 2003; Hanzlowsky et al., 2006). In fact, the non-conserved region C-terminal to the Myb domain of HsSNAP190 was inhibitory to binding by HsSNAPc (Mittal et al., 1999). It is not clear why the C-terminal domains of human and fly SNAP190 should have such differential effects on the DNA-binding activity of the complete SNAPc.

Figure 3B also indicates domains of DmSNAP50 and DmSNAP43 that are required for the DNA-binding activity of DmSNAPc (but not required for complex formation). Deletion of DmSNAP50 amino acids between residues 10 and 26 significantly reduced the DNA-binding activity of DmSNAPc, and deletion through residue 91 eliminated detectable DNA-binding activity (Hung *et al.*, 2009). We are not aware of any comparable studies that targeted the N-terminal function of human SNAP50.

In the case of DmSNAP43, deletion of residues following position 274 greatly weakened the DNAbinding activity of DmSNAPc, and deletion of nearly the entire non-conserved C terminus (that follows residue 172) resulted in a total loss of DmSNAPc DNA-binding activity (Figure 3B) (Hung et al., 2009). These results are very similar to those obtained by Ma and Hernandez (2001) in the human system. In conclusion, it is clear from the truncation experiments described that domains of DmSNAP190, DmSNAP50, and DmSNAP43 that are not evolutionarily conserved contribute to the DNA-binding activity of DmSNAPc. Whether this contribution arises from direct contacts between the protein and the DNA is not addressed through the truncation experiments. It is possible that the non-conserved domains that are required for the DNA-binding activity may be necessary for DmSNAPc to adopt a conformation compatible with efficient DNA binding.

### The structure of the DmSNAPc-DNA complex

An atomic structure of SNAPc (or its subunits) or of the SNAPc-DNA complex is not yet available. However, site-specific protein-DNA photo-cross-linking studies carried out in the *Drosophila* system have provided a wealth of information regarding the architecture of the protein-DNA complex (Wang and Stumph, 1998; Li *et al.*, 2004; Lai *et al.*, 2005; Kim *et al.*, 2010). These studies have revealed the position of each DmSNAPc subunit along the length of the PSEA as well as their rotational positions relative to the DNA sequence and to each other. Most interestingly, these studies have provided considerable evidence that DmSNAPc assumes different conformations depending upon whether the protein is bound to a U1 or a U6 PSEA.

Figure 4 shows a summary of the results of these sitespecific protein-DNA photo-cross-linking studies. The diagram at the top of the figure shows the position of the PSEA that is aligned in all the DNA diagrams below. The phosphate positions that cross-linked to each of the three subunits are indicated by colored spheres (yellow for DmSNAP190 cross-links; green and red for DmSNAP50, and blue and red for DmSNAP43). Oddnumbered phosphates indicate positions mapped on the non-template strand of the DNA, and even-numbered phosphates indicate positions mapped on the template strand. The colored areas thus represent the positions where each of the individual subunits closely approach the DNA when DmSNAPc binds either to a U1 PSEA or to a U6 PSEA (upper and lower diagrams respectively of each pair). DmSNAP190 cross-linked to phosphate positions that extended over the entire length of either a U1 or a U6 PSEA (Wang and Stumph, 1998). DmSNAP50 cross-linked to phosphate positions

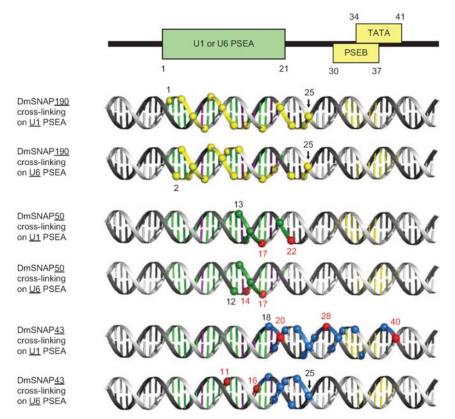


Figure 4. Contact points between snRNA gene promoter sequences and DmSNAPc. Site-specific protein-DNA photo-cross-linking data of DmSNAPc to DNA probes that contained either U1 or U6 PSEAs are shown projected onto B-form DNA (Wang and Stumph, 1998; Li et al., 2004). (For experimental details, see below.) The colored spheres on the DNA structures indicate phosphate positions that specifically cross-linked to DmSNAP190 (yellow), DmSNAP50 (green and red), or DmSNAP43 (blue and red). (The red spheres indicate the phosphate positions specifically used for the DmSNAP50 and DmSNAP43 domain-mapping experiments summarized in Figure 5.) On each pair of DNA structures, the upper illustration shows the cross-linking pattern to DNA probes containing a U1 PSEA, and the lower illustration shows cross-linking to probes containing a U6 PSEA. The schematic drawing at the top of the figure shows the location of the PSEA (either U1 or U6) and the relative locations of the PSEB or TATA sequence in snRNA gene Pol II and Pol III promoters respectively. The locations of these promoter elements are also indicated by colored bases within the DNA structures below (green for PSEA and yellow for PSEB or TATA box). The five bases that differed in the DNA probes are indicated by magenta coloring (U1 versus U6 base pairs at PSEA positions 7, 14, 16, 19, and 20). The DNA sequence flanking the U1 or U6 PSEA was maintained identical in all probes (and contained a PSEB but not TATA sequence) in order to ensure that any differences in the crosslinking patterns were due entirely to the five different nucleotides within the U1 and U6 PSEAs. Experimental details: A series of 86 individual DNA probes were prepared each of which contained a photo-cross-linker at an individual phosphate position (with an adjacent 32P radiolabel) within probes containing either a U1 or U6 PSEA. Cross-linker was placed at every second phosphate position on either the non-template strand (odd-numbered positions) or on the template strand (even-numbered positions). After binding and covalent cross-linking of DmSNAPc, all but two to three nucleotides of DNA were removed by nuclease digestion, and the radiolabeled proteins that cross-linked to each phosphate position were identified by gel electrophoresis and autoradiography. For further details see Wang et al. (1998) and Li et al. (2004).

extending from position 13 through position 22 of a U1 PSEA, but extending from phosphates 12 through 17 of a U6 PSEA (Wang and Stumph, 1998). DmSNAP43 exhibited the greatest differences in cross-linking patterns depending upon the source of the PSEA. This subunit cross-linked to phosphate positions extending from positions 18 to 40 of a U1 PSEA, but to a shorter and more upstream region of a U6 PSEA (phosphates

extending from positions 11 to 25) (Wang and Stumph, 1998; Li et al., 2004).

The experiments summarized in Figure 4 were carried out using DNA photo-cross-linking probes that were identical except at five nucleotide positions within the PSEA (U1 versus U6 bases at PSEA positions 7, 14, 16, 19, and 20). Therefore, it is certain that the observed differences in protein-DNA contacts arose from the five base

differences within the U1 and U6 PSEA sequences and not from the sequences flanking the PSEAs. It is particularly notable that DmSNAP43-DNA interactions occurred up to 20 bp (two turns of the DNA helix) downstream of the U1 PSEA. In contrast, DmSNAP43-DNA interactions were limited to only a distance of 4 bp downstream of a U6 PSEA (Figure 4) (Li et al., 2004).

The cross-linking studies, besides providing information regarding the nucleotides contacted along the longitudinal axis of the PSEA, furthermore revealed the particular face of the DNA contacted by each of the three DmSNAPc subunits (Wang and Stumph, 1998). The cross-linking patterns indicated that, when the DNA is oriented as shown in Figure 4, DmSNAP190 interacted with the front face (and to some extent with the upper and lower faces) of the DNA in the 5' half of the PSEA (either U1 or U6), but it contacted primarily the lower face of the DNA duplex in the 3' half of the PSEA. On the other hand, DmSNAP50 occupied the front face of the DNA in the 3' half of the PSEA. Finally, DmSNAP43 resided primarily on the upper face of the DNA in the 3' portion of the PSEA (as well as far downstream of the U1 PSEA). It is worth noting that although this modeling is done on B-form DNA in Figure 4, it is certainly possible that the binding of DmSNAPc may distort the DNA. Indeed, a study that made use of circular permutation, mini-circle binding, and ligase-catalyzed circularization assays suggested that the DNA of both the U1 and U6 PSEAs was modestly but similarly bent by DmSNAPc toward the face of the DNA helix contacted by DmSNAP43 (Hardin et al., 2000).

### Mapping protein domains within DmSNAP50 and DmSNAP43 that contact specific nucleotides of the U1 and the U6 PSEAs

Recently performed work has localized domains within the DmSNAPc subunits that cross-link strongly to certain individual phosphate positions (indicated by the positions of the red spheres in Figure 4) within the U1 and U6 PSEAs (Kim *et al.*, 2010). This was accomplished by combining the site-specific protein–DNA photo-cross-linking technique with site-specific chemical digestion of the protein. The cross-linked protein fragments were then identified by gel electrophoresis. For DmSNAP190, the comprehensive mapping of domains that cross-link to the U1 and U6 PSEAs is still in progress, but the domain-mapping experiments for DmSNAP50 and DmSNAP43 have been completed (Kim *et al.*, 2010) and the findings for these two smaller subunits are summarized in Figure 5.

There were four phosphate positions that cross-linked most strongly to DmSNAP50 (indicated by the red spheres in Figures 4 and 5). Those were phosphate positions 17

and 22 in the U1 PSEA, but phosphates 14 and 17 in the U6 PSEA (Wang and Stumph, 1998). Therefore, experiments were carried out to localize domains of DmSNAP50 associated with those four strongly-cross-linking phosphate positions (Kim *et al.*, 2010).

As indicated in Figure 5A, phosphate position 17 of both the U1 and U6 PSEAs cross-linked strongly to a polypeptide fragment of DmSNAP50 that spans amino acid residues 103-179. This same phosphate position (#17) of both PSEAs cross-linked even more strongly to the C-terminal fragment of DmSNAP50 comprised of residues 238-377 (which contains the zinc-binding SNAP finger domain). On the other hand, phosphate position 22 in the U1 PSEA cross-linked only to the DmSNAP50 fragment that contains residues 103-179. In a contrasting fashion, phosphate position 14 in the U6 PSEA cross-linked only to DmSNAP50 residues 238-377. Due to the limitations of the assay, it was not possible to know whether or not the region of DmSNAP50 between residues 180-237 interacts with phosphate position 17 of the U1 and U6 PSEAs. It is nonetheless apparent from the results that at least two distinct regions of DmSNAP50 (residues 103-179 and residues 238-377) are involved in contacting the DNA at different positions in the U1 and U6 PSEAs (phosphates 22 and 14 respectively).

An illustration of the mapped domains of DmSNAP50 interacting with the U1 and U6 PSEAs (represented on B-form DNA) is diagrammed in Figure 5B. When DmSNAPc binds to a U1 PSEA, a domain encompassing DmSNAP50 residues 103–179 closely approaches phosphates 17 and 22, and the SNAP finger domain at the C terminus (residues 238–377) is also close to phosphate 17. On the other hand, when DmSNAPc binds to a U6 PSEA, the SNAP finger domain (residues 238–377) closely approaches phosphate 14 as well as phosphate 17, and the region from residues 103–179 closely approaches phosphate 17 (but not 22) of the U6 PSEA.

An unanswered question from the results described above is whether the most N-terminal domain of DmSNAP50 (residues 1-102) is involved in contacting the DNA. The earlier truncation experiments indicated that amino acids in this region were required for DmSNAPc to bind to the DNA (Figure 3B). It is possible that the N-terminal domain of DmSNAP50 may contact nucleotides other than the four utilized in the photocross-linking/domain-mapping experiments described above, or it may be that this N-terminal domain is required for DmSNAPc to adopt a proper DNA-binding conformation.

Experiments were also carried out to map domains of the smallest subunit, DmSNAP43, that contact specific phosphates in the U1 and U6 PSEAs (Kim *et al.*, 2010). Because DmSNAP43 cross-linked strongly to phosphate positions 11 and 16 of the U6 PSEA,

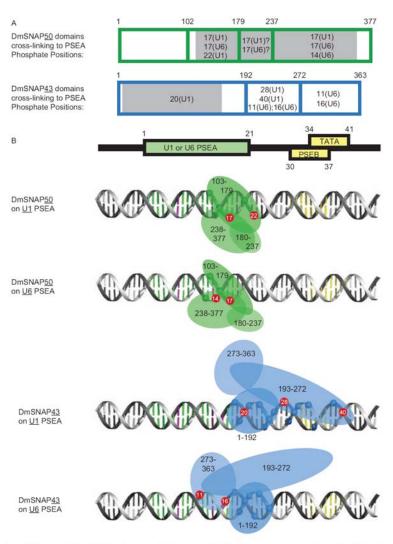


Figure 5. Domains of DmSNAP50 and DmSNAP43 that cross-link to specific phosphate positions on U1 or U6 DNA probes. (A) The rectangles at the top represent the linear amino acid sequences of the DmSNAP50 (green) and DmSNAP43 (blue) subunits. The shaded areas represent the evolutionarily conserved region of each subunit. The numbers above the rectangles indicate the amino acid positions at which the polypeptides were specifically cleaved with hydroxylamine to identify domains that cross-linked to specific nucleotide positions within the U1 and U6 PSEA probes (Kim et al., 2010). Phosphate positions that cross-linked to the indicated domains of each subunit are indicated within the rectangles. (B) The results summarized in (A) are projected onto B-form DNA. The color-coding of the schematic drawing at the top, and of the DNA helices, is the same as described in the legend to Figure 4. The phosphate positions used in these domain-mapping experiments are specifically labeled in red with white lettering. The mapped domains of the DmSNAP50 and DmSNAP43 subunits that cross-linked to the red-colored phosphate positions are represented as individual ellipses. The cross-linking patterns for each subunit are shown separately. It should be noted that there are no data to reveal which domains contact the phosphate positions represented as the green or blue spheres. For further details, see Kim et al. (2010).

but strongly to positions 20 and 28 of the U1 PSEA (Figure 4), domains of DmSNAP43 were mapped that interact closely with those four positions. In addition, phosphate 40 of the U1 PSEA was employed in the mapping experiments since this is the phosphate

furthest downstream of the PSEA that cross-linked to DmSNAP43. Those results are summarized in the lower part of Figure 5A.

First, phosphate 20 of the U1 PSEA cross-linked to the N-terminal fragment (residues 1-192) of DmSNAP43.

Phosphates 28 and 40, also of the U1 PSEA, cross-linked exclusively to the fragment of DmSNAP43 that spans residues 193–272. Finally, phosphates 11 and 16 of the U6 PSEA cross-linked strongly to the most C-terminal fragment of DmSNAP43 (residues 273–363) and much less strongly to a fragment encompassing residues 193–272.

A pictorial representation of these findings is presented in the lower part of Figure 5B. The N-terminal half of DmSNAP43 is in close proximity to position 20 within the U1 PSEA (and possibly relatively close to position 20 within the U6 PSEA, although no data was obtained for U6 position 20 due to the weakness of the cross-linking to that site; see Wang and Stumph, 1998). The DmSNAP43 domain encompassing residues 193-272 closely approaches phosphates 28 and 40 on the U1 PSEA, but not on a U6 PSEA. Residues within this same domain are within weak cross-linking distance to phosphates 11 and 16 when DmSNAPc binds to a U6 PSEA, but not when DmSNAPc binds to a U1 PSEA. Finally, the C-terminal domain of DmSNAP43 (residues 273-363) closely approaches phosphates 11 and 16 when DmSNAPc binds to a U6 PSEA, but not when DmSNAPc binds to a U1 PSEA. Thus, at least three separable regions of DmSNAP43 can be involved in interactions with the DNA, and these protein-DNA interactions differ depending upon the origin of the PSEA and its sequence (U1 or U6).

The locations of the mapped protein domains as drawn in Figure 5B are also consistent with the data that mapped the subunit domains involved in proteinprotein interactions (illustrated in Figure 3A). Those studies indicated that the evolutionarily conserved region of DmSNAP50 is involved in protein-protein interactions with the evolutionarily conserved region of DmSNAP43 (Hung et al., 2009). Interestingly, the protein-DNA photo-cross-linking studies described above place these same conserved domains of DmSNAP50 (DmSNAP50 residues 110 to 377) and DmSNAP43 (DmSNAP43 residues 1 to 155) into close proximity on the DNA (Figure 5B). This would seem to be a necessary consequence of the fact that the conserved domains of these two subunits participate in protein-protein interactions with each other.

### Role of DmSNAPc in the establishment of RNA polymerase specificity

Relatively minor sequence differences in the 3' halves of the PSEAs of Pol II- and Pol III-transcribed snRNA genes are responsible for determining the differential RNA polymerase specificity of these promoters (Figure 1B) (Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). Interestingly, the 3' half of the PSEA is the region contacted by all three subunits of DmSNAPc (Figure 4). How then are the signals encoded in the DNA sequence of the PSEAs transmitted to recruit distinct transcriptional machineries to the promoter? The protein–DNA photo-cross-linking experiments clearly indicate that the conformations of the DmSNAPc/DNA complexes are different on Pol II and Pol III PSEAs (Figures 4 and 5). We have hypothesized that the different PSEA sequences act as differential allosteric effectors of DmSNAPc conformation, and that the conformational differences of DmSNAPc on Pol II and Pol III snRNA promoters then lead to the differential recruitment of distinct sets of general transcription factors (GTFs) and subsequently different RNA polymerases to the different classes of snRNA genes.

A working model of how this might occur is shown in Figure 6. In this model, DmSNAPc adopts different conformations induced by the U1 and U6 PSEAs. The exposure of different surfaces of the DmSNAPc subunits, as well as differences in protein–DNA interactions, then results in the recruitment of Pol II GTFs to the U1 promoter but Pol III GTFs to the U6 promoter. This has the effect of recruiting Pol II to transcribe the U1 gene but Pol III to transcribe the U6 gene (Figure 6).

One of the Pol II GTFs required at the fly U1 promoter is the TATA binding protein (TBP). This was shown first by in vitro transcription assays and more recently by chromatin immunoprecipitation (ChIP) assays (Zamrod et al., 1993; Barakat and Stumph, 2008). Importantly, when the U1 PSEA was changed to a U6 PSEA (i.e. five base pair changes in only the PSEA), TBP could not be detected on the U1 promoter in vivo as determined by ChIP assays, in spite of the fact that DmSNAPc was able to bind in vivo to the mutant U1 promoter that contained the U6 PSEA (Barakat and Stumph, 2008). When DmSNAPc was bound to a U6 PSEA in the context of the U1 promoter, it appears that DmSNAPc either failed to recruit TBP or alternatively inhibited the binding of TBP to the DNA (Barakat and Stumph, 2008). However, the negative ChIP results cannot definitively eliminate the possibility that TBP might bind the mutant U1 promoter but could not be cross-linked to DNA or recognized by the antibodies. In any case, when the U6 PSEA was substituted for the U1 PSEA in the U1 promoter, it is apparent that a structural change occurred that prevented the recruitment or detection of TBP and also prevented the formation of an active transcription complex (Barakat and Stumph, 2008). Any of these interpretations is consistent with the generalized model presented in Figure 6.

It is tempting to believe that the conserved PSEB, although having little resemblance to a TATA sequence, might be a site of interaction of TBP with the DNA of the U1 promoter. The PSEB, like a TATA sequence, is 8 bp in length, and it is located at precisely the expected distance

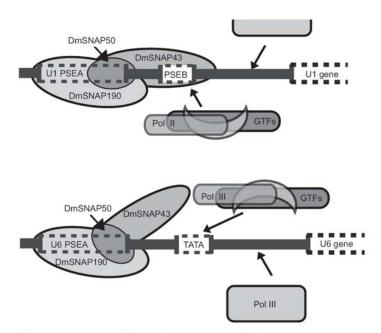


Figure 6. A working model illustrating how differential interactions of DmSNAPc with U1 or U6 PSEAs may lead to RNA polymerase specificity at *D. melanogaster* snRNA gene promoters. Minor sequence differences between U1 and U6 PSEAs are believed to act as allosteric effectors of DmSNAPc conformation, resulting in differential protein-DNA interactions as well as the potential exposure of different surfaces of the DmSNAPc subunits. The conformational differences of DmSNAPc are postulated to recruit Pol II GTFs to the U1 promoter (probably centered on the PSEB) but Pol III GTFs to the U6 promoter (probably centered on the TATA box). Due to the 12 bp versus 8 bp spacing difference between the PSEAs and either the TATA box or PSEB, a rotational as well as a longitudinal difference in the orientation and spacing of the Pol II and Pol III GTFs would be expected. The binding of the distinct GTFs would then recruit Pol II to the U1 promoter but Pol III to the U6 promoter to initiate transcription.

for a TATA box upstream of the Pol II transcription start site. However, there is not yet direct experimental evidence that TBP interacts with the PSEB. It is alternatively possible that TBP, rather than contacting the PSEB, could be tethered indirectly to the promoter by DmSNAPc or by other undefined components of the system. Additional work will be required to distinguish between these various possibilities. In any case, the fly U1 promoter may serve as an excellent system to study the role of TBP at TATA-less Pol II promoters.

The DmSNAP43 subunit in particular contacts the DNA very differently when DmSNAPc binds to the U1 and U6 PSEAs, and this subunit may be a particularly good candidate to play a role in recruitment of the Pol II GTFs. In the human system, transcription of snRNA genes by Pol II requires the GTFs TBP, TFIIA, TFIIB, TFIIE, and TFIIF (Sadowski *et al.*, 1993; Henry *et al.*, 1995; Kuhlman *et al.*, 1999). It is not clear whether TFIIH is involved, but a trypanosomal version of TFIIH was found to be essential for Pol II-dependent SL RNA transcription in those highly diverged organisms (Lee *et al.*, 2007). Because DmSNAP43 contacts the DNA at nucleotides within and flanking the PSEB on U1 promoters (Figures 4, 5, and 6),

it seems possible that DmSNAP43 could be involved in recruiting TFIIA, TFIIB, and/or TBP by means of direct protein-protein contacts. Although human SNAP43 is capable of interacting directly with TBP in the absence of DNA (Hinkley *et al.*, 2003), it is not yet known if this is true in flies. It will be important in future experiments to examine the possible role of DmSNAP43 in the recruitment of the Pol II GTFs.

Pol III transcription of U6 (and 75K) snRNA genes has been studied much more extensively in the human system than in flies (see Schramm and Hernandez, 2002; Jawdekar and Henry, 2008 and references therein). In terms of Pol III GTF recruitment, HsSNAPc was able to recruit both TBP and Brf2 to the human U6 promoter, and the recruitment of TBP occurred through direct interactions with HsSNAP190 (Ma and Hernandez, 2002; Hinkley *et al.*, 2003). Interestingly, Pol III transcription in fruit flies was found to employ the insect-specific TBP-related factor TRF1 rather than TBP itself (Takada *et al.*, 2000). It will be interesting to examine whether the DmSNAP190 subunit in flies plays an essential role in recruiting any of the Pol III GTFs to fly U6 snRNA gene promoters. Might it be possible that DmSNAP43

plays the most direct role in Pol II GTF recruitment, but that DmSNAP190 plays the most prominent role in Pol III GTF recruitment? Or are both of these subunits directly involved in GTF recruitment to both types of promoters? And what is the role of DmSNAP50? Further investigations will be required to identify all the players engaged in this process and to elucidate the molecular mechanisms by which distinct preinitiation complexes are assembled on either Pol II or Pol III snRNA gene promoters.

### Acknowledgements

We thank Cheryn Lee for substantial assistance in the preparation of figures involving graphical illustrations of DNA structure.

#### Declaration of interest

This work was supported by a grant from the National Science Foundation (MCB-0842770) and in part by a grant from the California Metabolic Research Foundation. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

#### References

- Baer M, Nilsen TW, Costigan C and Altman S. 1990. Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. Nucleic Acids Res 18:97–103.
- Bai L, Wang ZX, Yoon JB and Roeder RG. 1996. Cloning and characterization of the β subunit of human proximal sequence element-binding transcription factor and its involvement in transcription of small nuclear RNA genes by RNA polymerases II and III. Mol Cell Biol 16:5419-5426.
- Barakat NH and Stumph WE. 2008. TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proximal sequence element A (PSEA) for the U1 PSEA. FEBS Letters 582:2413-2416.
- Beck E, Jorcano JL and Alonso A. 1984. Drosophila melanogaster U1 and U2 small nuclear RNA genes contain common flanking sequences. J Mol Biol 173:539-542.
- Biedenkapp H, Borgmeyer U, Sippel AE and Klempnauer KH. 1988. Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature 335:835-837.
- Bond UM, Yario TA and Steitz JA. 1991. Multiple processing-defective mutations in a mammalian histone pre-mRNA are supressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev 5:1709-1722.
- Cabart P and Murphy S. 2001. BRFU, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with TATA-binding protein. J Biol Chem 276:43056–43064.
- Cabart P and Murphy S. 2002. Assembly of human small nuclear RNA gene-specific transcription factor IIIB complex de novo on and off promoter. J Biol Chem 277:26831–26838.
- Dahlberg JE and Lund E. 1988. The genes and transcription of the major small nuclear RNAs, pp. 38-70. In: Birnstiel ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein

- Particles. Heidelberg, Federal Republic of Germany: Springer Verlag.
- Das A and Bellofatto V. 2003. RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. Proc Natl Acad Sci USA 100:80-85.
- Das A, Zhang Q, Palenchar JB, Chatterjee B, Cross GA and Bellofatto V. 2005. Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol Cell Biol 25:7314-7322.
- Das G, Henning D and Reddy R. 1987. Structure, organization, and transcription of Drosophila U6 small nuclear RNA genes. J Biol Chem 262:1187-1193.
- Egloff S, O'Reilly D and Murphy S. 2008. Expression of human snRNA genes from beginning to end. Biochem Soc Trans 36: 590-594.
- Ford E, Strubin M and Hernandez N. 1998. The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. Genes Dev 12:3528–3540.
- Goodall GJ, Kiss T and Filipowicz W. 1991. Nuclear RNA splicing and small nuclear RNAs and their genes in higher plants. Oxford Surv Plant Mol Cell Biol 7:255-296.
- Guthrie C. 1991. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science 253:157-163.
- Hanzlowsky A, Jelencic B, Jawdekar G, Hinkley CS, Geiger JH and Henry RW. 2006. Co-expression of multiple subunits enables recombinant SNAPc assembly and function for transcription by human RNA polymerases II and III. Protein Expr Purif 48:215-223.
- Hardin SB, Ortler CJ, McNamara-Schroeder KJ and Stumph WE. 2000. Similarities and differences in the conformation of protein-DNA complexes at the U1 and U6 snRNA gene promoters. Nucleic Acids Res 28:2771-2278.
- Henry RW, Sadowski CL, Kobayashi R and Hernandez N. 1995. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. Nature 374:653–656.
- Henry RW, Ma BC, Sadowski CL, Kobayashi R and Hernandez N. 1996. Cloning and characterization of SNAP50, a subunit of the snRNAactivating protein complex SNAPc. EMBO J 15:7129-7136.
- Henry RW, Mittal V, Ma BC, Kobayashi R and Hernandez N. 1998. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. Genes Dev 12:2664-2672.
- Hernandez G, Valafar F and Stumph WE. 2007. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. Nucleic Acids Res 35:21-34.
- Hernandez N. 1992. Transcription of vertebrate snRNA genes and related genes, pp. 281–313. In: McKnight SL, Yamamoto KR, eds. Transcriptional Regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Hinkley CS, Hirsch HA, Gu LP, LaMere B and Henry RW. 2003. The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem 278:18649-18657.
- Huie JL, He P and Bellofatto V. 1997. In vitro transcription of the Leptomonas seymouri SL RNA and U2 snRNA genes using homologous cell extracts. Mol Biochem Parasitol 90:183-192.
- Hung KH, Titus M, Chiang SC and Stumph WE. 2009. A map of Drosophila melanogaster small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. J Biol Chem 284:22568-22579.
- Jawdekar GW and Henry RW. 2008. Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta 1779: 295-305.
- Jawdekar GW, Hanzlowsky A, Hovde SL, Jelencic B, Feig M, Geiger JH and Henry RW. 2006. The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem 281:31050-31060.
- Jensen RC, Wang Y, Hardin SB and Stumph WE. 1998. The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of Drosophila U-snRNA genes. Nucleic Acids Res 26:616-622.

- Kass S, Tyc K, Steitz JA and Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell 60:897–908.
- Kim MK, Kang YS, Lai HT, Barakat NH, Magante D and Stumph WE. 2010. Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol 30:2411-2423.
- Kiss T, Marshallsay C and Filipowicz W. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro. Cell 65:517-526.
- Klempnauer KH and Sippel AE. 1987. The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. EMBO J 6: 2219-2725.
- Kuhlman TC, Cho H, Reinberg D and Hernandez N. 1999. The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Mol Cell Biol 19:2130-2141.
- Lai HT, Chen H, Li C, McNamara-Schroeder KJ and Stumph WE. 2005. The PSEA promoter element of the Drosophila U1 snRNA gene is sufficient to bring DmSNAPc into contact with 20 base pairs of downstream DNA. Nucleic Acids Res 33:6579-6586.
- Lai HT, Kang YS and Stumph WE. 2008. Subunit stoichiometry of the Drosophila melanogaster small nuclear RNA activating protein complex (SNAPc), FEBS Lett 582:3734-3738.
- Lee BJ, Kang SG and Hatfield D. 1989. Transcription of Xenopus selenocysteine tRNA Ser (formerly designated opal suppressor phosphoserine tRNA) gene is directed by multiple 5'-extragenic regulatory elements. J Biol Chem 264:9696-9702.
- Lee JH, Nguyen TN, Schimanski B and Gunzl A. 2007. Spliced leader RNA gene transcription in Trypanosoma brucei requires transcription factor TFIIH. Eukaryot Cell 6:641–649.
- Li C, Harding GA, Parise J, McNamara-Schroeder KJ and Stumph WE. 2004. Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol 24:1897–1906.
- Lo PCH and Mount SM. 1990. Drosophila melanogaster genes for U1 snRNA variants and their expression during development. Nucleic Acids Res 18:6971-6979.
- Lobo SM and Hernandez N. 1989. A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell 58:55-67.
- Lobo SM and Hernandez NT. 1994. Transcription of snRNA genes by RNA polymerases II and III, pp. 127-159. In: Conaway RC, Conaway JW, eds. Transcription: Mechanisms and Regulation. New York, NY: Raven Press.
- Luo H, Gilinger G, Mukherjee D and Bellofatto V. 1999. Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. J Biol Chem 274: 31947-31954.
- Ma B and Hernandez N. 2001. A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc. J Biol Chem 276:5027-5035.
- Ma BC and Hernandez N. 2002. Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol 22:8067–8078.
- Mattaj IW, Dathan NA, Parry HD, Carbon P and Krol A. 1988. Changing the RNA polymerase specificity of U snRNA gene promoters. Cell 55:435-442
- McNamara-Schroeder KJ, Hennessey RF, Harding GA, Jensen RC and Stumph WE. 2001. The Drosophila U1 and U6 gene proximal sequence elements act as important determinants of the RNA polymerase specificity of snRNA gene promoters in vitro and in vivo. J Biol Chem 276:31786-31792.
- Mittal V, Ma BC and Hernandez N. 1999. SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes Dev 13:1807–1821.
- Murphy S, Yoon JB, Gerster T and Roeder RG. 1992. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. Mol Cell Biol 12:3247-3261.

- Parry HD, Scherly D and Mattaj IW. 1989. Snurpogenesis: the transcription and assembly of U snRNP components. Trends Biochem Sci 14:15-19.
- Peculis BA and Steitz JA. 1993. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell 73:1233-1245.
- Rosinski JA and Atchley WR. 1998. Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. I Mol Evol 46:74-83.
- Saba JA, Busch H, Wright D and Reddy R. 1986. Isolation and characterization of two putative full-length Drosophila U4 small nuclear RNA genes. J Biol Chem 261:8750-8753.
- Sadowski CL, Henry RW, Lobo SM and Hernandez N. 1993. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. Genes Dev 7:1535-1548.
- Schimanski B, Laufer G, Gontcharova L and Gunzl A. 2004. The Trypanosoma brucei spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. Nucleic Acids Res 32:700-709.
- Schimanski B, Nguyen TN and Gunzl A. 2005. Characterization of a multisubunit transcription factor complex essential for splicedleader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol 25:7303-7313.
- Schramm L and Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev: 16 2593–2620.
- Schramm L, Pendergrast PS, Sun YL and Hernandez N. 2000. Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. Genes Dev 14:2650–2663.
- Sharp PA. 1994. Split genes and RNA splicing. Cell 77:805-816.
- Steitz JA, Black DL, Gerke V, Parker KA, Kramer A, Frendewey D and Keller W. 1988. Functions of the abundant U-snRNPs, pp. 115-154. In: Birnsteil ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag.
   Su Y, Song Y, Wang Y, Jessop L, Zhan LC and Stumph WE. 1997.
- Su Y, Song Y, Wang Y, Jessop L, Zhan LC and Stumph WE. 1997. Characterization of a Drosophila proximal-sequence-elementbinding protein involved in transcription of small nuclear RNA genes. Eur J Biochem 248:231–237.
- Takada S, Lis JT, Zhou S and Tjian R. 2000. A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. Cell 101:459-469.
- Teichmann M, Wang ZX and Roeder RG. 2000. A stable complex of a novel transcription factor IIB-related factor, human TFIIIB50, and associated proteins mediate selective transcription by RNA polymerase III of genes with upstream promoter elements. Proc Natl Acad Sci USA 97:14200-14205.
- Waibel F and Filipowicz W. 1990. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. Nature 346:199-202.
- Waldschmidt R, Wanandi I and Seifart KH. 1991. Identification of transcription factors required for the expression of mammalian U6 genes in vitro. EMBO J 10:2595–2603.
- Wang Y and Stumph WE. 1998. Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 snRNA genes. Mol Cell Biol 18:1570–1579.
- Wong MW, Henry RW, Ma BC, Kobayashi R, Klages N, Matthias P, Strubin M and Hernandez N. 1998. The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1. Mol Cell Biol 18:368-377.
- Yoon JB and Roeder RG. 1996. Cloning of two proximal sequence element-binding transcription factor subunits (γ and δ) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein. Mol Cell Biol 16:1–9.
- Yoon JB, Murphy S, Bai L, Wang Z and Roeder RG. 1995. Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Mol Cell Biol 15:2019-2027.

Yuan Y and Reddy R. 1991. 5' flanking sequences of human MRP/7-2 RNA gene are required and sufficient for the transcription by RNA polymerase III. Biochim Biophys Acta 1089:33–39.

polymerase III. Biochim Biophys Acta 1089:33–39. Zamrod Z and Stumph WE. 1990. U4B snRNA gene enhancer activity requires functional octamer and SPH motifs. Nucleic Acids Res 18:7323–7330.

Zamrod Z, Tyree CM, Song Y and Stumph WE. 1993. In vitro transcription of a Drosophila U1 small nuclear RNA gene requires

TATA box-binding protein and two proximal cis-acting elements with stringent spacing requirements. Mol Cell Biol 13: 5918–5927.

Zieve G, Benecke BJ and Penman S. 1977. Synthesis of two classes of small RNA species in vivo and in vitro. Biochemistry 16:4520-4525.

Editor: Michael M. Cox

### **ACKNOWLEDGEMENTS:**

This Chapter, in full, is a reprint of the material as it appears in Critical Reviews in Biochemistry and Molecular Biology, 2010. Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc). Hung, K. H.; Stumph, W. E. The dissertation author was the primary author of this paper.

### **CHAPTER 3**

Localization of residues in a novel DNA-binding domain of DmSNAP43 required for DmSNAPc DNA-binding activity

### **ABSTRACT**

Transcription of snRNA genes depends upon the recognition of the proximal sequence element (PSE) by the snRNA activating protein complex SNAPc. In *Drosophila melanogaster*, all subunits of DmSNAPc (DmSNAP43, DmSNAP50, and DmSNAP190) are required for PSE-binding activity. Previous work demonstrated that a non-canonical DmSNAP43 domain bounded by residues 193-272 is essential for DmSNAPc to bind to the PSE. In this study, the contribution of amino acid residues within this domain to DNA binding by DmSNAPc was investigated by alanine scanning mutagenesis. The results have identified two clusters of residues within this domain required for the sequence-specific DNA-binding activity of DmSNAPc.

### **INTRODUCTION**

The small nuclear RNA (snRNA)-activating protein complex (SNAPc) is a multi-subunit transcription factor required for transcription of snRNA genes (Waldschmidt et al., 1991; Sadowski et al., 1993; Goomer et al., 1994; Henry et al., 1995; Yoon et al., 1995). The snRNAs are small non-coding RNAs involved in many essential cellular functions including RNA processing (e.g. pre-messenger RNA splicing and ribosomal RNA processing) and other gene regulatory events (e.g. transcription initiation and elongation) (Steitz et al., 1988; Kass et al., 1990; Guthrie, 1991; Peculis and Steitz, 1993; Sharp, 1994; Kwek et al., 2002; Nguyen et al., 2001; Yang et al., 2001). In animals, the transcription of snRNA genes depends upon a promoter element termed the proximal sequence element (PSE) located about 40-65 bp upstream of the transcription start site. The binding of SNAPc to the PSE is an essential step for preinitiation complex (PIC) assembly on snRNA promoters and the subsequent recruitment of RNA polymerase to initiate transcription (Sadowski et al., 1993; Kuhlman et al., 1999; Schramm et al., 2000; Teichmann et al., 2000; Cabart and Murphy, 2001; Cabart and Murphy, 2002; Schimanski et al., 2005; Das et al., 2005; Lee et al., 2007; Barakat and Stumph, 2008). Interestingly, despite the universal utilization of SNAPc and the PSE, different animal snRNA genes have distinct RNA polymerase requirements: one class of snRNA genes (e.g. U1, U2, U3, U4, U5, and U7 snRNA genes) are transcribed by RNA polymerase II (Pol II), but another class of snRNA genes (e.g. U6 snRNA and 7SK RNA genes) are transcribed by RNA

polymerase III (Pol III) (Zieve et al., 1977; Dahlberg and Lund, 1988; Parry et al., 1989; Hernandez, 1992; Lobo and Hernandez, 1994).

In the fruit fly *Drosophila melanogaster*, DmSNAPc contains three distinct subunits (DmSNAP43, DmSNAP50, and DmSNAP190) that form a stable heterotrimeric complex prior to binding to the PSEA (the Drosophila PSE) (Su et al., 1997; Wang and Stumph, 1998; Li et al., 2004). These three subunits are evolutionarily conserved from trypanosomes to humans and represent the "core SNAP subunits". All three subunits are essential for sequence-specific PSEA binding as none of the three subunits can bind to the PSEA either individually or in any pair-wise combination (our unpublished observations). Earlier studies that investigated the binding of DmSNAPc to the PSEA sequence by using site-specific protein-DNA photo-cross-linking revealed that each of the three subunits directly contacted the DNA when DmSNAPc was bound to either a U1 or a U6 PSEA (Wang and Stumph, 1998; Li et al., 2004). SNAP190 contains an evolutionarily conserved Myb domain with multiple Myb repeats that are involved in binding to the DNA, and SNAP50 contains a zinc-binding "SNAP finger" that is required for SNAPc PSE binding activity (Wong et al., 1998; Li et al., 2004; Jawdekar et al., 2006). In contrast, SNAP43 does not possess any DNA-binding domain with known resemblance to any other canonical DNA-binding protein. However, truncation analysis revealed that a Cterminal region of DmSNAP43 encompassing residues 172-274 was required for the U1 PSEA binding activity of DmSNAPc, even though this region was not required for DmSNAP43 assembly with either DmSNAP50 or DmSNAP190 (Hung et al., 2009). However, those studies did not address whether this region (residues 172-274) directly

contacted the DNA or simply assisted DmSNAPc to adopt a conformation compatible with efficient DNA binding.

This question was addressed by subsequent experiments that used site-specific protein-DNA photo-cross-linking combined with site-specific chemical cleavage of the protein to map domains of DmSNAP43 that interact with the U1 PSEA (Kim et al., 2010). The results, summarized in **Fig. 3.1A**, showed that a domain of DmSNAP43 that spans residues 193-272 was in close proximity to phosphates 28 and 40 of the U1 PSEA. Together, those results suggested that the DmSNAP43 domain bounded by residues 193 and 272 is involved in contacting the DNA downstream of the U1 PSEA. Since this domain appears not to share any sequence similarity to any characterized DNA-binding domains, we decided to investigate the contribution of amino acid residues within this domain to DmSNAPc DNA-binding activity.

To accomplish this, blocks of three to six amino acids at a time were mutated to alanine throughout this region, and the DNA binding activity of DmSNAPc containing each mutant construct was assessed. Our findings have identified groups of DmSNAP43 amino acid residues that are essential for DmSNAPc to bind sequence-specifically to DNA.

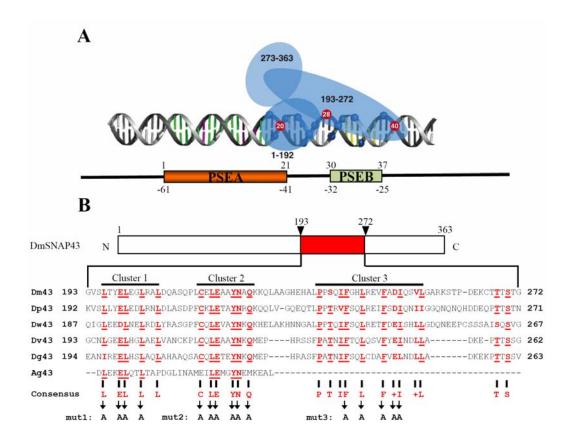


Figure 3.1. Identification of evolutionarily conserved amino acid residues within a DNA-binding domain of DmSNAP43. (A) Schematic representation of domains of DmSNAP43 that contact phosphate positions 20, 28, and 40 when DmSNAPc binds to a U1 PSEA. From previous work, a domain bounded by residues 193 and 272 cross-linked to phosphate positions 28 and 40 (Kim et al., 2010) and was required for the DNA-binding activity of DmSNAPc (Hung et al., 2009). This region of DmSNAP43 is the subject of the current work. A domain bounded by residues 1-192 contacted position 20 (Kim et al., 2010) but was also required for the association of DmSNAP43 with DmSNAP50 and DmSNAP190 (Hung et al., 2009). Other phosphate positions colored in blue cross-linked to DmSNAP43 but have not been localized to a specific region of the protein. The linear diagram below the double helical DNA shows the positions of the U1 PSEA and PSEB as projected onto the double helix above. The numbers above the linear diagram indicate the nucleotide positions relative to the first nucleotide of the PSEA, and the numbers below indicate the positions relative to the transcription start site. (B) Alignment and analysis of insect SNAP43 sequences within the domain bounded by residues 193 and 272. Five Drosophila sequences were used in this comparison: Dm, melanogaster; Dp: pseudoobscura; Dw: willistoni; Dv: virilis; Dg: grimshawi. A sequence from the mosquito Anopheles gambiae was also included as an example from a more distant insect. The residues in this region that are most strongly conserved are shown in red and are underlined if they are identical to the consensus amino acids listed below. They group into three clusters as indicated by the horizontal lines above the sequences. An initial round of alanine substitutions was carried out for each cluster and the three DmSNAP43 constructs were named mut1, mut2, and mut3 as shown at the very bottom of the figure.

### MATERIALS AND METHODS

### **DmSNAPc** constructs and expression

The preparation of untagged constructs encoding wild type DmSNAP50 and DmSNAP190 and N-terminal 6xHis-FLAG-tagged DmSNAP43 constructs under the control of the copper-inducible metallothionein promoter has been previously described (Hung et al., 2009). N-terminal 6xHis-FLAG-tagged DmSNAP43 constructs with alanine substitutions were prepared by using the QuickChange II site-directed mutagenesis kit (Stratagene).

Tagged wild type or mutant DmSNAP43 constructs were each co-expressed with untagged DmSNAP50 and DmSNAP190 wild type constructs in stably transfected Drosophila S2 cells as previously described (Hung et al., 2009). Subunit co-expression was induced with copper sulfate and confirmed by immunoblotting. DmSNAP43 constructs were detected by using anti-FLAG M2 monoclonal antibodies (Sigma). The detection of untagged DmSNAP50 or DmSNAP190 was performed by using antibodies prepared against synthetic peptides whose sequences correspond to C-terminal amino acid sequences of either DmSNAP50 or DmSNAP190 (Li et al., 2004).

### **Protein purification**

Following copper sulfate induction, cells were washed in phosphate-buffered saline and lysed in CelLytic M lysis buffer (Sigma) containing 1% protease inhibitor cocktail (Sigma). Lysates were then adjusted to a NaCl concentration of 0.5 M prior to incubating with ProBond resin (Invitrogen) for 2 hours to allow the capture of

complexes containing 6xHis-FLAG-tagged DmSNAP43 together with associated DmSNAP50 and DmSNAP190. The resins were then washed three times in 50 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 8.0), 0.5 M NaCl, 20 mM imidazole and then once in HEMG-100 buffer (100 mM KCl, 25 mM HEPES K<sup>+</sup> (pH 7.6), 12.5 mM MgCl<sub>2</sub>, 10 μM ZnCl<sub>2</sub>, 0.1 mM EDTA (pH 8.0), 10% glycerol, 3 mM dithiothreitol, and 0.5 mM phenylmethylsulfonyl fluoride) with 20 mM imidazole. The complexes were than eluted from the resin with 750 mM imidazole in HEMG-100 buffer followed by dialysis against HEMG-100 buffer to remove the imidazole.

### **Electrophoretic mobility shift assays (EMSA)**

Protein-DNA incubations for DNA mobility shift assays were carried out in 21-μl volumes in a final concentration of ~80 mM KCl, 20 mM HEPES K<sup>+</sup> (pH 7.6), 10 mM MgCl<sub>2</sub>, 8 μM ZnCl<sub>2</sub>, 80 μM EDTA (pH 8.0), 8% glycerol, 2 mM dithiothreitol and 0.4 mM phenylmethylsulfonyl fluoride. The radioactive DNA probe contained the wild type promoter sequence of the D. melanogaster U1:95Ca gene from -73 to -5 relative to the transcription start site. Incubation of the DNA probe with DmSNAPc containing wild type or alanine-scanning mutants of DmSNAP43 (purified as described above) was carried out for 30 min at 20 °C. Complexes were then run on 5% non-denaturing polyacrylamide gels in 1x non-circulation buffer (25 mM Tris, 190 mM glycine, 1 mM EDTA, pH 8.3) and then detected by autoradiography.

### **Chromatin immunoprecipitations (ChIPs)**

ChIP assays were carried out as previously described (Hung et al., 2009).

Affinity-purified polyclonal anti-FLAG antibodies (Sigma) were used to

immunoprecipitate protein-DNA complexes containing 6xHis-FLAG tagged DmSNAP43 constructs. Anti-DmSNAP43 antibodies produced in a rabbit immunized with bacterially expressed recombinant DmSNAP43 were used as a positive control in the ChIP assays. The preimmune serum from the same rabbit prior to immunization was used as a negative ChIP control. The ChIP PCR forward primer (5'-GTGTGGCATACTTATAGGGGTGCT-3') and reverse primer (5'-GCTTTTCGATGCTCGGCAGCAG-3') amplify the promoter region of the U1:95Ca gene from -1 to -107 relative to the transcription start site.

### RESULTS

# Identification of three clusters of evolutionarily conserved residues within the DmSNAP43 DNA-binding domain bounded by residues 193 and 272

Initially to localize amino acid residues potentially required for DNA-binding activity within the DmSNAP43 DNA-binding domain (193-272), we first retrieved from Flybase (<a href="http://flybase.org">http://flybase.org</a>) SNAP43 protein sequences from five different Drosophila species and aligned these amino acid sequences by utilizing constraint based protein multiple alignment tool (COBALT) (<a href="http://www.ncbi.nlm.nih.gov/tools/cobalt/">http://www.ncbi.nlm.nih.gov/tools/cobalt/</a>). The result for amino acid residues 193-272 of DmSNAP43 is shown in **Fig. 3.1B**. The most conserved amino acid residues in the fruit fly SNAP43 sequences fell primarily into three clusters (1, 2, and 3, as denoted by horizontal lines above the Dm43 sequence). Most of the conserved residues in clusters 1 and 2 were also conserved in mosquito SNAP43 (line labeled

Ag43), but the residues in cluster 3 were not well-conserved between fruit flies and mosquitoes (results not shown). Most of the consensus residues in clusters 1 and 3 were hydrophobic amino acids, whereas most of the consensus residues in cluster 2 were uncharged polar amino acids.

To examine the possibility that these conserved residues might be involved in binding to DNA, we used site-directed mutagenesis to convert some of the most conserved amino acids within each cluster to alanines. The resultant mutant constructs were named mut1, mut2, and mut3 in reference to the conserved cluster of residues where the alanine substitutions were introduced. Wild type or individual mutant DmSNAP43 constructs were co-expressed in stably-transfected *D. melanogaster* S2 cells together with untagged DmSNAP50 and DmSNAP190. The overexpression of these constructs under the control of the metallothionein promoter was confirmed by immunoblotting (data not shown). DmSNAP complexes containing tagged wild type or mutant DmSNAP43 were purified from the stably transfected cells by nickel-chelate chromatography.

# Mutation of DmSNAP43 amino acid residues in two of three conserved clusters compromised the DNA binding activity of DmSNAPc

To determine whether the mutant DmSNAP43 proteins were able to assemble in vivo with the co-expressed DmSNAP50 and DmSNAP190 subunits, immunoblotting was used to examine the co-purification of DmSNAP50 and DmSNAP190 with the tagged wild type or mutant DmSNAP43 constructs purified by nickel-chelate chromatography. The top panel in **Fig. 3.2A** shows that both wild type (lane 1) and DmSNAP43 mutant proteins (lanes 2, 3, and 4) could be detected in the

nickel column elution fractions by using anti-FLAG antibody. As revealed by the immunoblots shown in the middle and bottom panels, both DmSNAP50 and DmSNAP190 co-purified with the wild type and each of the mutant DmSNAP43 constructs. This indicates that the alanine substitutions introduced into DmSNAP43 (mut1, mut2 or mut3) did not affect the ability of any of the mutant proteins to associate with DmSNAP50 and DmSNAP190 to form the DmSNAP complex. This result is in agreement with earlier truncation studies that indicated this domain (193-272) was not required for DmSNAP43 to associate with either DmSNAP50 or DmSNAP190 (Hung et al., 2009).

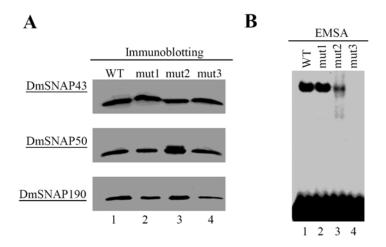


Figure 3.2. Mutations within two clusters of DmSNAP43 conserved residues have no affect on DmSNAP complex formation but result in reduced DmSNAPc DNA-binding activity. (A) Immunoblots demonstrating that DmSNAP50 and DmSNAP190 co-purified with tagged DmSNAP43 wild type and mutant constructs following nickel chelate column chromatography. The amount of protein in each lane was normalized to provide nearly identical band intensities for the tagged DmSNAP43 detected with anti-FLAG antibodies (upper panel). The same amounts of protein were then analyzed with antibodies against DmSNAP50 or DmSNAP190 (middle and lower panels). (B) Electrophoretic mobility shift analysis (EMSA) of nickel column-purified DmSNAP complexes with a DNA fragment containing a U1 PSEA sequence. The amounts of protein used were normalized based upon the results of the immunoblots shown in (A).

Next, to investigate the effect of these alanine substitutions on DmSNAPc DNA binding activity, the wild type and mutant complexes were subjected to EMSA analysis (**Fig. 3.2B**). Protein amounts in each lane were normalized as determined from the immunoblot data shown in **Fig. 3.2A**. The complexes that contained either wild type DmSNAP43 or DmSNAP43 mut1 bound efficiently to the U1 PSEA (**Fig. 3.2B**, lanes 1 and 2). However, the DNA-binding activity of the complex that contained DmSNAP43 mut2 was severely compromised (**Fig. 3.2B**, lane 3). Finally, DmSNAPc that contained DmSNAP43 cluster 3 mutations lacked detectable DNA binding activity (**lane 4**). These results indicated that amino acid residues substituted with alanine in cluster 2 or cluster 3 are required for DmSNAPc to bind efficiently to the U1 PSEA.

## Alanine substitutions outside of the conserved clusters had little or no effect on DmSNAP complex formation or DNA binding activity

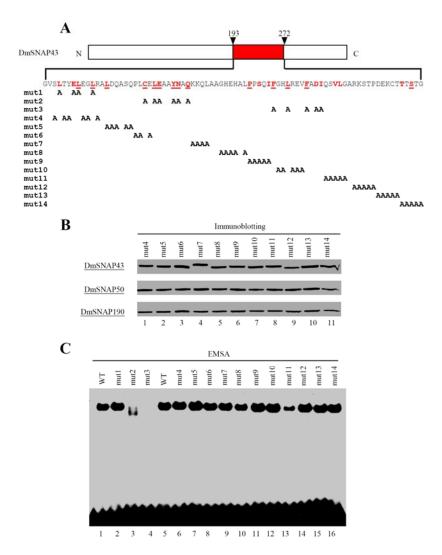
To more comprehensively evaluate the roles of the amino acid residues within the DmSNAP43 DNA-binding domain (193-272), eleven additional alanine substitution mutants of DmSNAP43 were prepared that scanned throughout the entire region comprising residues 193-272. These eleven additional DmSNAP43 mutants are illustrated in **Fig. 3.3A**. Although most have five alanine substitutions, some have fewer (mut6 and mut7) and one (mut4) has six alanine substitutions. Furthermore, although most of these alanine substitutions occur at residues that are not evolutionarily conserved, some of the new mutations are at well-conserved positions that were not included in the first round of alanine substitutions.

Each tagged DmSNAP43 mutant was co-overexpressed in S2 cells together with untagged DmSNAP50 and DmSNAP190. Following purification by nickel-chelate chromatography, complex formation was assessed by immunoblotting to

measure co-purification of the DmSNAP50 and DmSNAP190 subunits with the tagged DmSNAP43 mutant constructs (**Fig. 3.3B**). Lanes 1-11 in the upper panel of **Fig. 3.3B** show that each of the tagged DmSNAP43 variants was detected in the elution fractions. The middle and bottom panels show that, as expected, both DmSNAP50 and DmSNAP190 co-purified with each of the DmSNAP43 constructs. These results, together with those of **Fig. 3.2A**, indicate that none of the amino acid residues in the DmSNAP43 (193-272) domain are essential for DmSNAP43 association with DmSNAP50 and DmSNAP190.

Next, the DNA-binding activities of DmSNAP complexes containing DmSNAP43 mut4 to mut14 constructs were analyzed by EMSA (Fig. 3.3C, lanes 6-16). For comparative purposes, reactions were also carried out with wild type and mut1, mut2, and mut3 complexes analyzed on the same gel (Fig. 3.3C, lanes 1-5). Consistent with the results shown in Fig. 3.2B, wild type and mut1 complexes efficiently bound to the U1 PSEA (Fig. 3.3C, lanes 1, 2 and 5), but mut2 and mut3 complexes exhibited little or no PSEA-binding activity (Fig. 3.3C, lanes 3 and 4). On the other hand, complexes that contained the remaining DmSNAP43 mutants retained DNA-binding activity comparable to the wild type (Fig. 3.3C, lanes 6-16). An exception was mut11 which appeared to have a reduced DNA-binding activity (Fig. **3.3C, lane 13**). It is notable that these mut11 alterations are in proximity to the mut3 mutations that eliminated DmSNAPc DNA-binding activity. In fact, mut11 contains mutations that are part of conserved cluster 3 which is rich in amino acids with bulky hydrophobic side chains (Fig. 3.1). Interestingly, in mut11, well-conserved hydrophobic amino acids with branched side chains (valine, leucine, and isoleucine in

**Fig. 3.1B**) at positions 254 and 255 were mutated (**Fig. 3.3A**). Thus, it is possible that the valine and leucine residues at these positions contribute to the DNA-binding activity of DmSNAPc.



**Figure 3.3.** Alanine scanning mutagenesis within the DNA-binding domain of DmSNAP43 between residues 193 and 272. (A) Locations of alanine substitutions in fourteen mutant DmSNAP43 constructs are shown. Residues in red in the upper sequence were very strongly conserved in the five *Drosophila* species analyzed in Fig. 3.1 (only highly conservative substitutions were observed); underlined residues indicate positions of 100% identity among the five fruit fly species. (B) Immunoblots demonstrating that DmSNAP50 and DmSNAP190 co-purified with tagged DmSNAP43 wild type and mutant constructs following nickel chelate column chromatography. Analysis was carried as described in the legend to Fig. 3.2. (C) EMSA of nickel column-purified DmSNAP complexes. The amounts of protein used were normalized based upon the results of the immunoblots in (B).

#### In vivo DNA-binding activity correlates with that observed in vitro

EMSA experiments described above examined the effect of alanine scanning mutations on the in vitro sequence-specific DNA binding activity of DmSNAPc. To examine the DNA binding activity of these mutants in vivo, chromatin immunoprecipitation (ChIP) assays were conducted by using the same stably transfected cell lines employed to make extracts for protein purification. ChIPs were carried out by using antibodies against the FLAG epitope of the tagged DmSNAP43 constructs to examine their in vivo occupancy of the well-characterized endogenous U1:95Ca gene promoter. The PCR primers specifically amplified a 107-base pair DNA fragment from this promoter (Barakat and Stumph, 2008; Hung et al., 2009). Polyclonal antibodies prepared against full-length DmSNAP43 were used as positive ChIP controls for their ability to detect endogenous DmSNAP43 as well as the overexpressed tagged DmSNAP43 constructs on the U1 promoter. Preimmune antibodies from the same rabbit were used as negative controls in the ChIP assays.

In the transfected cell line that overexpressed tagged wild type DmSNAP43, the anti-FLAG antibodies efficiently precipitated the U1 promoter (**Fig. 3.4, lane 3**), although with less efficiency than the polycolonal antibodies (**lane 1**). The mut1 DmSNAPc, that contains alanine substitutions in DmSNAP43 conserved residues of cluster 1, likewise was able to bind to the U1 promoter in vivo (**Fig. 3.4, lane 7**). In contrast, DmSNAP43 with mutations in cluster 2 or cluster 3 (mut2 or mut3) could not be detected at the U1 promoter (**Fig. 3.4, lanes 11 and 15**). These findings were consistent with the results of the in vitro DNA-binding assays shown in **Figs. 3.2B and 3.3C**.

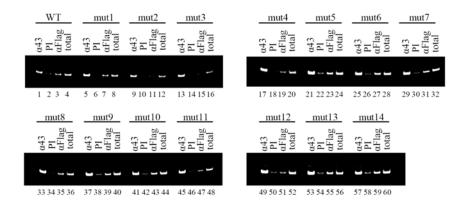


Figure 3.4. ChIPs from stably-transfected S2 cells that over-express FLAG-tagged WT or mutant DmSNAP43 constructs that contain alanine substitutions. Anti-FLAG antibodies were used for ChIP to examine the relative binding of the mutant constructs to the U1 promoter *in vivo* (lanes labeled  $\alpha$ FLAG). Positive controls utilized anti-DmSNAP43 antibodies prepared against full-length DmSNAP43 ( $\alpha$ 43). Pre-immune antibodies (PI) were used as negative controls. Lanes labeled "total" were positive PCR controls.

ChIP results from cells expressing DmSNAP43 mut4 to mut14 are shown in lanes 17-60 of **Fig. 3.4**. In every case, a signal was obtained with the FLAG-antibodies that exceeded that obtained with the pre-immune antibodies. Furthermore, the signal was generally slightly weaker than that obtained with the total input DNA. Although these results are more qualitative than quantitative, the results indicate that all of the mut4-mut14 constructs retain a significant level of DNA-binding activity.

#### **DISCUSSION**

Two clusters of conserved amino acid residues within DmSNAP43 domain (193-272) are essential for the sequence-specific DNA binding activity of DmSNAPc

In accord with previous findings (Hung et al., 2009), the alanine scanning data presented here showed that none of the amino acid residues within DmSNAP43 (193-272) was essential for DmSNAP43 to associate with either DmSNAP50 or with DmSNAP190 (Figs. 3.2A and 3.3B). Because complex formation was unaffected, this provided us with an opportunity to evaluate the contribution of residues throughout this entire region to the DNA binding activity of DmSNAPc. Results shown in Figs. **3.3** and **3.4** revealed that two clusters of evolutionarily conserved residues within this domain were required for the DNA binding activity of DmSNAPc in vitro and in vivo, whereas residues outside of these conserved clusters had little or no effect on DNA binding activity. For example, alanine substitutions of DmSNAP43 residues conserved in cluster 2 severely compromised DNA binding by the complex, and a combination of alanine substitutions of residues conserved in cluster 3 totally eliminated the ability of the complex to bind to DNA. Also, the partial reduction in DNA binding observed with DmSNAP43 mut11 might be attributed to the mutation of two hydrophobic branched side chain residues conserved at the C-terminal end of cluster 3.

On the other hand, alanine substitutions in residues conserved in cluster 1 had no effect on DNA binding by DmSNAPc even though those residues were well-conserved between mosquito and flies. In fact, some of these residues are conserved even in vertebrates (unpublished observation). Taking into consideration that the

DmSNAP43 (193-272) domain exhibits very little overall sequence similarity with vertebrate SNAP43, it is plausible that these conserved residues might be important for SNAP43 functions other than DNA binding. For example, photo-cross-linking studies have placed this domain of DmSNAP43 closer to the transcription start site than other components of DmSNAPc and near to sites normally occupied by components of the transcription pre-initiation complex (Lai et al., 2005; Kim et al., 2010). Thus this DmSNAP43 domain may also play a role in recruitment of the Pol II and/or Pol III general transcription factors into pre-initiation complexes established on snRNA genes. Further experiments will be required to evaluate such possibilities.

#### ACKNOWLEDGEMENTS:

The material in this chapter, in full, will be submitted for publication immediately following approval of this thesis. The dissertation author was the primary researcher and author of this paper.

#### REFERENCES

- Barakat NH, Stumph WE. (2008). TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proximal sequence element A (PSEA) for the U1 PSEA. FEBS Letters, 582, 2413-16.
- Cabart P, Murphy S. (2001). BRFU, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with TATA-binding protein. J Biol Chem, 276, 43056-64.
- Cabart P, Murphy S. (2002). Assembly of human small nuclear RNA gene-specific transcription factor IIIB complex de novo on and off promoter. J Biol Chem, 277, 26831-38.

- Dahlberg JE, Lund E. (1988). The genes and transcription of the major small nuclear RNAs. In: Birnstiel ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag KG, 38-70.
- Das A, Zhang Q, Palenchar JB, Chatterjee B, Cross GA, Bellofatto V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol Cell Biol, 25, 7314-22.
- Goomer RS, Urso O, Kunkel GR. (1994). A complex that contains proteins binding to the PSE and TATA sites in a human U6 small nuclear RNA promoter. Gene, 148, 269-75.
- Guthrie C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science, 253, 157-63.
- Henry RW, Sadowski CL, Kobayashi R, Hernandez N. (1995). A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. Nature, 374, 653-56.
- Hernandez N. (1992). Transcription of vertebrate snRNA genes and related genes. In: McKnight SL, Yamamoto KR, eds. Transcriptional Regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 281-313.
- Hung KH, Titus M, Chiang SC, Stumph WE. (2009). A map of Drosophila melanogaster small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. J Biol Chem, 284, 22568-79.
- Jawdekar GW, Hanzlowsky A, Hovde SL, Jelencic B, Feig M, Geiger JH, Henry RW. (2006). The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem, 281, 31050-60.
- Kass S, Tyc K, Steitz JA, Sollner-Webb B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell, 60, 897-908.
- Kim MK, Kang YS, Lai HT, Barakat NH, Magante D, Stumph WE. (2010). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol, 30, 2411-23.

- Kuhlman TC, Cho H, Reinberg D, Hernandez N. (1999). The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Mol Cell Biol, 19, 2130-41.
- Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol, 9, 800-5.
- Lai HT, Chen H, Li C, McNamara-Schroeder KJ, Stumph WE. (2005). The PSEA promoter element of the Drosophila U1 snRNA gene is sufficient to bring DmSNAPc into contact with 20 base pairs of downstream DNA. Nucleic Acids Res, 33, 6579-86.
- Lee JH, Nguyen TN, Schimanski B, Gunzl A. (2007). Spliced leader RNA gene transcription in Trypanosoma brucei requires transcription factor TFIIH. Eukaryot Cell, 6, 641-9.
- Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE. (2004). Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol, 24, 1897-906.
- Lobo SM, Hernandez NT. (1994). Transcription of snRNA genes by RNA polymerases II and III. In: Conaway RC, Conaway JW, eds. Transcription: Mechanisms and Regulation. New York, NY: Raven Press, 127-59.
- Nguyen VT, Kiss T, Michels AA, Bensaude O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature, 414, 322-5.
- Parry HD, Scherly D, Mattaj IW. (1989). Snurpogenesis: the transcription and assembly of U snRNP components. Trends Biochem Sci, 14, 15-19.
- Peculis BA, Steitz JA. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell, 73, 1233-45.
- Sadowski CL, Henry RW, Lobo SM, Hernandez N. (1993). Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. Genes Dev, 7, 1535-48.
- Schimanski B, Nguyen TN, Gunzl A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol, 25, 7303-13.

- Schramm L, Pendergrast PS, Sun YL, Hernandez N. (2000). Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. Genes Dev, 14, 2650-63.
- Sharp PA. (1994). Split genes and RNA splicing. Cell, 77, 805-16.
- Steitz JA, Black DL, Gerke V, Parker KA, Kramer A, Frendewey D, Keller W. (1988). Functions of the abundant U-snRNPs. In: Birnsteil ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag KG, 115-54.
- Su Y, Song Y, Wang Y, Jessop L, Zhan LC, Stumph WE. (1997). Characterization of a Drosophila proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. Eur J Biochem, 248, 231-37.
- Teichmann M, Wang ZX, Roeder RG. (2000). A stable complex of a novel transcription factor IIB-related factor, human TFIIIB50, and associated proteins mediate selective transcription by RNA polymerase III of genes with upstream promoter elements. Proc Natl Acad Sci USA, 97, 14200-05.
- Waldschmidt R, Wanandi I, Seifart KH. (1991). Identification of transcription factors required for the expression of mammalian U6 genes in vitro. EMBO J, 10, 2595-603.
- Wang Y, Stumph WE. (1998). Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 snRNA genes. Mol Cell Biol, 18, 1570-79.
- Wong MW, Henry RW, Ma BC, Kobayashi R, Klages N, Matthias P, Strubin M, Hernandez N. (1998). The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1. Mol Cell Biol, 18, 368-77.
- Yang ZY, Zhu QW, Luo KX, Zhou Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature, 414, 317-22.
- Yoon JB, Murphy S, Bai L, Wang Z, Roeder RG. (1995). Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Mol Cell Biol, 15, 2019-27.
- Zieve G, Benecke BJ, Penman S. (1977). Synthesis of two classes of small RNA species in vivo and in vitro. Biochemistry, 16, 4520-5.

## **CONCLUDING REMARKS**

#### DmSNAP domains for subunit assembly and DNA binding

Work described in Chapter 1 identified domains within each subunit of DmSNAPc required for complex assembly and DNA binding. Our findings show that with one exception, the evolutionarily most-conserved region of each DmSNAP subunit is sufficient for its assembly with the other two subunits. For example, the conserved N-terminal region of DmSNAP43 was sufficient to associate with both DmSNAP50 and DmSNAP190, and the conserved C-terminal region of DmSNAP50 was sufficient for its association with both DmSNAP190 and DmSNAP43. Furthermore, the conserved Myb domain of DmSNAP190 was sufficient for its interaction with DmSNAP43. The only exception to this "conserved-domain rule" was that a non-conserved N-terminal region of DmSNAP190 was required for its interaction with DmSNAP50.

On the other hand, DNA binding by DmSNAPc is dependent not only upon the conserved regions but is also highly dependent upon domains outside the conserved regions. For example, the non-conserved C-terminal region of DmSNAP190 and DmSNAP43, and the non-conserved N-terminal region of DmSNAP50 were all required for the DNA-binding activity of DmSNAPc even though these regions were not required for complex formation. This suggests that the requirements for subunit-subunit interactions within metazoan SNAPc are more constrained evolutionarily than the protein-DNA interactions. This agrees with our recent observations that the DNA sequences of the 3' half of the PSEAs have change fairly rapidly during insect evolution (Hernandez et al., 2007), suggesting that interactions between the protein and DNA are more free to co-evolve.

A number of studies have been published that address the role of domains within the human SNAP subunits (Mittal et al., 1999; Ma and Hernandez, 2001; Ma and Hernandez, 2002; Hinkley et al., 2003; Jawdekar et al., 2006). Comparing findings in Chapter 1 with human studies revealed some important similarities and several surprising differences. As an example of human-fly similarity, the central region of DmSNAP50 and the N-terminal conserved region of DmSNAP43 required for the interaction between these fly subunits also contribute to association between human SNAP50 and SNAP43 (Ma and Hernandez, 2001; Jawdekar et al., 2006). Another example is that the region of DmSNAP43 involved in DNA binding was also required for DNA binding in human system (Ma and Hernandez, 2001).

On the other hand, many significant differences between human and fly systems were observed: our data showed strong interactions between DmSNAP190 and DmSNAP50, whereas the homologous subunits in the human system have never been reported to directly interact. In fact available evidence in the human system indicates that human SNAP190 and SNAP50 do not directly interact (Wong et al., 1998; Mittal et al., 1999; Ma and Hernandez, 2001; Jawdekar et al., 2006). We also observed a strong interaction between DmSNAP190 and DmSNAP43. However, in human system, stable association between SNAP190 and SNAP43 requires an additional subunit, SNAP19 (Ma and Hernandez, 2001), which is not present in the fruit fly. Thus, this finding may reveal that different strategies are employed in the fly and human systems for stable assembly of the SNAPc subunits. Moreover, we found that the C-terminal non-conserved domain of DmSNAP190 is essential for DNA binding by DmSNAPc. Interestingly, this region of human SNAP190 is dispensable

for DNA binding (actually, this region even down-regulates human SNAPc binding to DNA) (Mittal et al., 1999; Ma and Hernandez, 2001; Ma and Hernandez, 2002; Hinkley et al., 2003; Hanzlowsky et al., 2006). Lastly, our results indicate that the N-terminal non-conserved region of DmSNAP50 is required for DmSNAPc binding to PSEA. To our knowledge, comparable studies targeting the N-terminal function of human SNAP50 have not been reported.

The ability to compare results in the human and fly systems has in several cases allowed us to identify domains that function similarly in both systems. However, in other cases novel insights into metazoan SNAPc functional domains have been obtained, and differences between the two systems suggest possible alternative mechanisms for achieving complex assembly and stable DNA binding.

#### Characterization of a novel DNA-binding domain of DmSNAP43

Truncational analysis described in Chapter 1 revealed that a region in the non-conserved C-terminal domain of DmSNAP43 was required for the DNA binding activity of DmSNAPc even though this region is not required for complex assembly. Subsequent experiments that combined site-specific protein-DNA photo-cross-linking with site-specific chemical digestion of the protein confirmed that this same region (bounded by residues 193 to 272) was capable of making direct contact to the DNA (Kim et al., 2010). Interestingly, this domain does not resemble any canonical DNA-binding domain available in the existing database. Work described in Chapter 3 partially characterized this novel DNA-binding domain by evaluating the contribution of amino acid residues within this domain to the DNA binding activity of DmSNAPc by alanine substitution analysis.

Bioinformatic analysis that compared six insect SNAP43 protein sequences from this domain revealed that most of the evolutionarily conserved residues can be grouped into three clusters (named as cluster 1, cluster 2, and cluster 3). We suspected that alanine substitutions of those conserved residues might affect the DNA-binding activity of DmSNAPc. Indeed, my data indicated that the DNA-binding activity of DmSNAPc was severely compromised or eliminated from mutant constructs with alanine-substituted residues in cluster 2 or cluster 3. On the other hand, alanine substitutions of residues throughout this domain outside of the conserved clusters had little or no effect on the DNA-binding activity of DmSNAPc. The above findings indicated that the conserved residues in cluster 2 and cluster 3 within DmSNAP43 domain (193-272) are required for DmSNAPc binding to the DNA, whereas all other residues within this same domain make little or no contribution to the DNA-binding activity of DmSNAPc.

The result that alanine substitution of residues conserved in cluster 1 had no effect on the DNA-binding activity of DmSNAPc was perhaps surprising because these hydrophobic residues are well-conserved from flies to mosquitoes and some of them are even conserved in vertebrate SNAP43 (our unpublished observation). This suggests that residues conserved in this first cluster might play an important role in SNAPc functions other than DNA binding. Previous site-specific protein-DNA photocross-linking experiments localized DmSNAP43 to the 3' end of the PSEA and downstream DNA sequences (Wang and Stumph, 1998; Li et al., 2004; Lai et al., 2005). This suggests that DmSNAP43 might potentially interact with GTFs and lead to the recruitment of Pol II or Pol III to U1 and U6 snRNA promoters, respectively. In

such a case, those residues conserved in cluster 1 within the DmSNAP43 (193-272) may be important for such interactions with GTFs or other components of the Pol II or Pol III transcription pre-initiation complexes. Future experiments that analyze protein-protein or protein-DNA interactions will be required to examine those possibilities.

#### REFERENCES

- Hanzlowsky A, Jelencic B, Jawdekar G, Hinkley CS, Geiger JH, Henry RW. (2006). Co-expression of multiple subunits enables recombinant SNAPc assembly and function for transcription by human RNA polymerases II and III. Protein Expr Purif, 48, 215-23.
- Hernandez G, Valafar F, Stumph WE. (2007). Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. Nucleic Acids Res, 35, 21-34.
- Hinkley CS, Hirsch HA, Gu LP, LaMere B, Henry RW. (2003). The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem, 278, 18649-57.
- Jawdekar GW, Hanzlowsky A, Hovde SL, Jelencic B, Feig M, Geiger JH, Henry RW. (2006). The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem, 281, 31050-60.
- Kim MK, Kang YS, Lai HT, Barakat NH, Magante D, Stumph WE. (2010). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol, 30, 2411-23.
- Lai HT, Chen H, Li C, McNamara-Schroeder KJ, Stumph WE. (2005). The PSEA promoter element of the Drosophila U1 snRNA gene is sufficient to bring DmSNAPc into contact with 20 base pairs of downstream DNA. Nucleic Acids Res, 33, 6579-86.
- Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE. (2004). Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol, 24, 1897-906.

- Ma B, Hernandez N. (2001). A map of protein-protein contacts within the small nuclear RNA-activating protein complex SNAPc. J Biol Chem, 276, 5027-35.
- Ma BC, Hernandez N. (2002). Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol, 22, 8067-78.
- Mittal V, Ma BC, Hernandez N. (1999). SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes Dev, 13, 1807-21.
- Wang Y, Stumph WE. (1998). Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 snRNA genes. Mol Cell Biol, 18, 1570-79.
- Wong MW, Henry RW, Ma BC, Kobayashi R, Klages N, Matthias P, Strubin M, Hernandez N. (1998). The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1. Mol Cell Biol, 18, 368-77.

#### **APPENDIX**

- A. Plasmid constructs for expressing various forms of the DmSNAP subunits in S2 cells
- B. List of stably transfected S2 cell lines
- C. High-salt FLAG purification of FLAG-tagged DmSNAPc from S2 cells
- D. Purification of His-tagged DmSNAPc from S2 cells by nickelchelate chromatography
- E. Detailed protocol for electrophoretic mobility shift assay (EMSA)
- F. Detailed protocol for chromatin immunoprecipitation assay (ChIP)

# Appendix A: Plasmid Constructs for expressing various forms of the DmSNAP subunits in S2 cells

## A. Plasmid constructs made for truncational analysis described in Chapter 1

Sequences of plasmid constructs for expressing tagged truncated DmSNAP43 proteins used in Chapter 1 are included in this appendix. The truncated genes were prepared by PCR and then cloned initially into the vector PMT/V5-His-TOPO (Invitrogen). Fragments were removed from these constructs and re-cloned into the FLAG-modified expression vector. To minimize possible tag interference, all N-terminal truncations had the FLAG-Myc-6xHis tag at the C-terminus, and all C-terminal truncations had the 6xHis-FLAG tag at the N-terminus.

#### List of constructs:

pMT-HisFlag-DmSNAP43 full STOP pMT-HisFlag-DmSNAP43 (2-274) STOP pMT-HisFlag-DmSNAP43 (2-172) STOP pMT-HisFlag-DmSNAP43 (2-125) STOP pMT-DmSNAP43 full noSTOP-FlagMycHis pMT-DmSNAP43 (68-363) noSTOP-FlagMycHis

## **B.** Plasmid constructs made for alanine scanning described in Chapter 3

Sequences of plasmid constructs for expressing N-terminal 6xHis-FLAG-tagged alanine-substituted DmSNAP43 proteins used in Chapter 3 are included in this appendix. These DmSNAP43 mutants were prepared by using the QuickChange II

site-directed mutagenesis kit (Stratagene). Wild-type 6xHis-FLAG-tagged DmSNAP43 construct was used as the template for performing the mutagenesis.

#### List of constructs:

pMT-HisFlag-DmSNAP43 full STOP
pMT-HisFlag-DmSNAP43 mut#1 STOP
pMT-HisFlag-DmSNAP43 mut#2 STOP
pMT-HisFlag-DmSNAP43 mut#3 STOP
pMT-HisFlag-DmSNAP43 mut#4 STOP
pMT-HisFlag-DmSNAP43 mut#5 STOP
pMT-HisFlag-DmSNAP43 mut#6 STOP
pMT-HisFlag-DmSNAP43 mut#7 STOP
pMT-HisFlag-DmSNAP43 mut#8 STOP
pMT-HisFlag-DmSNAP43 mut#9 STOP
pMT-HisFlag-DmSNAP43 mut#10 STOP
pMT-HisFlag-DmSNAP43 mut#11 STOP
pMT-HisFlag-DmSNAP43 mut#12 STOP
pMT-HisFlag-DmSNAP43 mut#13 STOP
pMT-HisFlag-DmSNAP43 mut#13 STOP

## pMT-5'HisFlag-Dm43 full STOP

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
201 CCGCACAGAT GCGTAAGGAG AAAATACCGC	ATCAGGCGCC	ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT
301 TACGCCAGCT GGCGAAAGGG GGATGTGCTG	CAAGGCGATT	AAGTTGGGTA	ACGCCAGGGT	TTTCCCAGTC	ACGACGTTGT	AAAACGACGG	CCAGTGCCAG
401 TGAATTAATT CGTTGCAGGA CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCAC CGCCCACCGC CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTGGAGAA CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAACGAAAG ACCCGTGTGT .	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AATCATCTCA GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
				> M H	н н н	H $H$ $T$ $D$	Y K D
901 GACGATGACA AGGGCACTAG TGAGCTGAAT .	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
> D D D K G T S E L N	I F D	D C W E	L V Q	R F Q	R L V N	D G E	N C E
1001 TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
	L Q L Q		T A Q	T N H $T$	E V I	A T T	L A A L
1101 GCATGTGGCC AAGCGACTGT CGTGCTCCCG	ACGCACCACC				AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
> H V A K R L S C S R		G D V	F P A S		R I G	G F F L	L Y V
1201 ATCTACTACA AGCAGCCCAC GCACAACTTT				CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
> I Y Y K Q P T H N F		E V S P			T D Y A	L D L	R K D
1301 GTCCGGAGCG GAAGGACACT CATCAGATCG							
	A Y M L	WRL	T Q E	Q A F R		L D Y	C Q G L
1401 GGACAATCTG GTGGACTACG ACCGTGTGGA							
		G A K	E Q R Q		M Q K	QQRA	N G V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA				GAACTGGAAG	CGGCATACAA	TGCCCAAAAG	AAGCAATTGG
> S L T Y E L E G L R		Q A S Q			A A Y N	A Q K	K Q L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT						CTAGGAGCTA	GAAAGAGTAC
	D &	G H L		F A D I	~	L G A	R K S T
1701 TCCAGATGAG AAATGCACCA CAACATCTAC							
> P D E K C T T T S T			R Q R V		A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG							
		N E T Y		M S S	ATVF	QRE	L P E
1901 ACGTGCAGCA AGAGTATGAG ATGATTGAGT							
~ ~	F S D D		E V G	E S E E	V T E	E E L	KAIL
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
γ π σ <							
2101 20020200000 0000000000 200000200	ааааааааатт	GG3 3 GGT3 3 G	GGM3 MGGGM3	* aaamamaam	ааататаа т	mama aaaama	aaaamaa maa
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG							
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG	ATCAGCCTCG	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TGAGTTTGGA	CAAACCACAA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT	ATCAGCCTCG GTGAAATTTG	ACTGTGCCTT TGATGCTATT	CTAAGATCCA GCTTTATTTG	GACATGATAA TAACCATTAT	GATACATTGA AAGCTGCAAT	TGAGTTTGGA AAACAAGTTA	CAAACCACAA ACAACAACAA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA	GACATGATAA TAACCATTAT AAACCTCTAC	GATACATTGA AAGCTGCAAT AAATGTGGTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA	CAAACCACAA ACAACAACAA TGATCAGTCG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAACCTT GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCCGGGGGAG 2801 CGGCGAGCGG TATCAGCTCA CTCCAAAGGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGA 2801 CGGCGAGGGG TATCAGCTCA CTCAAAGGCG 2901 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTT 2301 CTAGAATACA TGCATTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCATGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGAG 2801 CGGCGAGCG TATCAGCTCA CTCAAAGGCG 2901 CCAGGAACGG TATCAAAGAGT GCCTTGCTGG 3001 CCCGACAGGA CTATAAAGAT ACCAGGCGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC ACACACATGTG AAATCGACGC CTGCCGCTTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCGGCCAAC GCCGGGGGAC 2801 CGGCGAGCGG TATCAAGCTCA CTCAAAGGCG 2901 CCCGACAGGA CTATAAAAGGC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGT TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGG 2801 CGGCGAGCGG TATCAGCTCA CTCAAAGGCC 2901 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGCGTT 3101 CTCCCTTCGG GAACCGTGG GCCTTTCTCAT 3201 CCGTTCAGCC CGACCGCTG GCCTTATCCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ACTGACTCGC ACGACATGGCACGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGGCT ATCGCCACTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCGGCCAAC GCCGGGGGAC 2801 CGGCGAGCGG TATCAAGCTCA CTCAAAGGCG 2901 CCCGACAGGA CTATAAAAGGC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGGTCC TCTTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC CACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATGCCACTG AGAAGGACAC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCCGCCAAC GCCCGGGGAC 2801 CGGCGAGCGG TATCAAGCTCA CTCAAAGGCG 2901 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAAGCGTGG GCTTTCTCAT 3201 CGGTTCAGCC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCGAGGTATG TAGGCGGTGC 3401 CTGAAGCCAG TTACCTTCCG AAAAAGAGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTAACTATCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTCACAGA TAGCCTCCACGA AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTTGAGTGG GATCCGCAA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTCTCC CAGTTCGGTG AACCCGGTAA ACCCCGTAA ACAAACACAC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGGTC AACGCAGGAA TGTTCCGACC TAGGTCGTTC GACACGACTA GGCATCACAC GCTGGTACCC GCTGGTACCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG GAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG GAGAAGACGACA GTGGTTTTTT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTGTGTG GCAGCACCA TATTTGGTAT TGTTTGCAAG	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGCCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAACTTA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA TGGAAAAAA TGGTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGGGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGAC 2801 CGGGCAACGG TATCAACGCTCA CTCAAAGGCG 2901 CCCGACAGGA CTATAAAGAT ACCAGGCGTT 3101 CCCGTCAGC GAACGGTGC GCTTTCTCAT 3201 CCGTTCAGC GAACGCTGC GCCTTATCCG 3301 GATTAGCAGA GCGAGGTATG TAGCCGTGC	ATCAGCCTCG GTGAAATTTG GGAGGTTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TAACTATCG TAACTATCG TACAGAGTTC GGTAGCTCT CTTTGATCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA TTCTACGGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC	GACATGATAA TAACCATTAT AAACCTTCTAC TATTCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT CGCGTACACG AGTGGAACGA AGTGGAACGA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG ACTGACTGG ACTGACTGGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAACAGTGAAGCT ATCGCCACTG AGAAGACATGTG AGAAGGACAG GTGGTTTTTTT AAACTCACGT	TGAGTTTGGA AAACAGTTA TGGCTGATTA CGACTGATTA CGACATCGG GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGAATACCT CAGGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGCCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCACACTACACACTCT CAGCAACACTCT CAGCAACACTCT CAGCACTCTAG TGCTCTATGAG TGGTCATGAG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGAG 2801 CCGGACAGGG TATCAAGCCG CTCAAAGGCG 2901 CCCGACAGGA CTATAAAGAT ACCAGGCGTT 3101 CCCGTCAGC GAGCGTGC GCTTTCTCAT 3201 CCGTTCAGC GAAGCGTGC GCTTTCTCAT 3201 CCGTTCAGC GAAGCGTGC GCCTTATCCG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3501 ATTATCAAAAA AGGATCTTCA CCTAGATCCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCTATAGC TGGCCTACT GGAGTTTG GTAATACGGT CCGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATC TACAGAGTTC GGTAGCTCTT CTTTGAACTTT TTTTAAATTAA	ACTGTGCCTT TGATGCTATT GGAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TTCTACGGGA AATGAAGTT	CTAAGATCCA GCTTTATTTG ANAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA ACAAACCACC TCTGACGCTC TTAAATCAAT	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACT GCTGGTACGC CTAGATATA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AAATCGACGC TTGCCGGCTTA GCTCCAAGGT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT ATTCACGTAAA	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTTG GGCGTGTTG GCACGACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGCGAAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCCGCCAAC GCCGCGGGAC 2801 CGGCGAGCGG TATCAAGCTCA CTCAAAAGGCG 2901 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAAACGTGGC GCTTTCTCAT 3201 CGGTTCAGCC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCGAGGTATG TAGGCGGTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT 3701 TGCTTAATCA GTGAGGCACC TATCTCAGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGCCTATAGC TGAGCTACG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGCTTC TTCAGAGTTC CTTTGATCTT TTTTAAATTAA ATCTGTCTAT	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGACTCCGC GTAGGTATCT TCTTGAGTCC TTTGAAGTGC TTTGAAGTGGT TATCACGGGA TTCTACGGGG AAATGAAGTT TTCGTTCATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC TCTGACGCTC TCTAACTA CATAGTTTGCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC TACCCGGTC AACGCAGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT GCTGGTACCA GCTGGTACCA AGTGGAACGA CTAAAGTATA TGACTCCCCC	GATACATTGA AAGCTGCAAT AAATGTGTAC CAATTCCAGTGG ACTGACTGC AGAACATGTG AAATCGACGC CTGCCGCTTA ACTCGCACTG ACTCCCACTG AGAACATTTTTAAACTCACGT TATGACTACTTTTTAAACTCACGT TATGAGTAAT TATGAGTAAA	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CTGCGCTTT CAGCAACCCC CTGGTAACAG CAGCAGATTA CGGCACATTA CGGCACACACACACACACACACACACACACACACACAC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGAG 2801 CCGGACAGGG TATCAAGCCG CTCAAAGGCG 2901 CCCGACAGGA CTATAAAGAT ACCAGGCGTT 3101 CCCGTCAGC GAGCGTGC GCTTTCTCAT 3201 CCGTTCAGC GAAGCGTGC GCTTTCTCAT 3201 CCGTTCAGC GAAGCGTGC GCCTTATCCG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3501 ATTATCAAAAA AGGATCTTCA CCTAGATCCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG TTACAGAGTTC CTTTGATCTT TTTTAATTTA ATCAGACTC CGGAGACCC CGGAGACCC	ACTGTGCCTT TGATGCTATT GGAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGGCTCCCTCG GTAAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TTCTACGGGG AAATGAAGTT TTCGTTCATC ACGCTCACCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCTTGCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	GACATGATAA TAACCATTAT AAACCTCTAT TATCCGCTCA CACTGCCCGC TACCTGGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACAGCACACACAC GCTACACTT GACACGACTACACT GCTGGTACACG AGTGGAACGA CTAAAGTATA TGACTCCCCCG TATCAGCAT	GATACATTGA AAGCTGCAAT AAGCTGCATA CAATTCCAGTG CAATTCCAGCA TTTCCAGTGG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACAG ATCGCCACTG AGAAGGACAG TTCGCCACTG AGAAGGACAG TTCGCCACTG AGAAGACATAGA TAGACTAAAA TAACCACGT TATGAGTAAAA AAACCAGCCA	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA GCCGGAAGGG	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAC CTGCGCTCTG CAGCAACTAC CAGCAGTTACCAC CAGCAGTTACCAC CAGCAGTTACCAA CGGGAGGGCT CCGACCGCAG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGG 2801 CGGGCAGCGG TATCAAGCTCA CTCAAAGGCC 2901 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGCCGTT 3101 CTCCCTTCGG GAACCGTGGC GCCTTTCTCAT 3201 CGGTCAGCC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGG TTACCTTCGG AAAAAGAGTT 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCTATAGC TGGCCTACT GTAATACGT CGTTATTTCCA TCCCCCTGGA AGCTACGT TCCCCCTGGA AGCTACT TTACAGAGTT TTACAGAGTTC TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGCATTAATT CGTATTGGG TATCCACAGA TAGGCTCCGG GTAGGTACT TCTTGAGTCC TTGAAGTAGT TCTTGAGTGG AATCGAGAA TTCTACGGGA ATTCTACGGG AAATGAAGTT TCGTTCATC ACGCTCACCG AATTGTTGCC AATTGTTCACC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCCGGTAA AACCAAC AACCAACCA TCTGAACCA TCTGAACCAC TCTGAACTA CAAATCAAC GCTCAGAGT GCTCAGAGTT GGGAACCTAG	GACATGATAA TAACCATTAAT AAACCTCTTAC TATCCGCTCA CACTGCCGC AACGCAGGAA AGCATCACAA AGCATCACAA GCACTCACAC TAGTCCGACC TAGGTCGTC GACACGACTT GACACACT GCTGGTACGC CTAAAGTATT TGACTCCCCG TATAAGTATT TGACTCCCCG TATACACAT TATCAGCAT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGACCACTG ACTGCCAGCTTA GCTCCAAGGT ATCGCCACTG AGAAGAGAGACAG AGAAGGACAG TGGTTTTTT AAACTCACGT ATATGAGGAAA TCGTGTAGAT AAACCAGCCA TAGGTAAA	TGAGTTTGGA AAACAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGTG GCAGCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTGCA	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGATCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCAAAA GGTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGTCATCACAACAG CTGCGCTCTG CAGCACATTA CGGAGGCTCTG CAGCACATTA CCGAGCGCAG CAACCTTGTTCCACACACACACACACACACACACACACAC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAA 2701 GCATTAATGA ATCGCCAAC GCCGGGGAC 2801 CGGCGAGCGG TATCAGCTCA CTCAAAGGCC 2901 CCAGCAAGCA CTATAAAGAT ACCAGCGGTT 3101 CCCCTTCGG GAACCGTGC GCCTTCTCAT 3201 CCGTTCAGC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCGAGGTATG TAGCGGTGC 3401 CTGAAGCCA TACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATT CAAGAAGATC 3601 ATTATCAAAAA AGGATCTTCA CTTAGATCCT 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG 3801 TACCATCTGG GAACTTTAT CCGCCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCC GTAACTACC GTAACTATC GGTAGCTCT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTTTT CCCCGTTTTT TCGTCCTTTT TCGTCCTTTT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGG AAATGAAGTT TTCGTTCATC ACGCTCACCG AATTGTTCATC ACGCTCACCG GTATGGCTTC GTATGGCTCC GTATGGCTTC	CTAAGATCCA GCTTATTTG GTGAAATTGT GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACCCCGTAACTA ACACACCAC TCTGACGTC TCTGACGTC CTAACTA CATAGTTGCC CCTCAGATT CGCCCAGATT CGCAGATTGCC CCTCAGATT CATAGTTGCC CCTCCAGATT CGGGAAGCTAC ATTCAGCTCC ATTCAGCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGGTCGTC GACACGACT GGCTACACAC GCTGGTACCA AGTAGACGA TAGACGACT TAGACCACA TAGACCAC AGTGGAACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTAGT AGTAAGTAGT GGTTCCCAAC	GATACATTGA AAGCTGCAAT AAAGCTGCACA TTTCCAGTCG AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGCTA GCTCCAAGCT ATCGCCACTG AGAAGACATGT ACCCACTG AGAAGACAGCT ATCGCCACTG AGAAGACAGCT ATCACCACTG AGAAGACAGCA TTGGTTTTTT AAACTCACGT TATGAGTAAA TACGAGCCA TCGCCAGTTA GATCAAGGCG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATCGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA CTTGGTATT CTTGGTATT CTTGGTATT CTTGCAGG AACTACGATA CCGGAAGGG ATGTTTGCA	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGGCAGCT CGTCGGCTG CAGCAAAAGG GGTGGCAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCGAGGCGCAG CAACGTTGTT TCCCCCATGT
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGCCCAAC GCCGGGGGA 2801 CGGCGACGGG TATCAAGCC GCGTGGGGGA 2901 CCCGACAGGA CTATAAAAGGCC GCTTTCTCGT 3101 CTCCCTTCGG GAAGCGTGG GCTTTCTCAT 3201 CGGTTCAGC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCGAGGTATG TAGCGGTGT 3401 CTGAAGCCAG TTACCTTCGG AAAAAGGATT 3501 GCCGCAGAAA AAAAGGATCT CAAGAAGATC 3501 ATTATCAAAA AGGATCTTCA CCTAGATCCT 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG 3801 TACCATCTGG CCCAGTGCT GCAATGATAC 3901 AAGTGGTCCT GCAACTATTA CCGCCTCCAT 4001 GCCATTGCTA CGAGCACCT GGTGTCCAC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCTGGA AGCTCATCG TTACAGACTTC CTTAGACTTT CTTTGATCTT TTTAAATTAA ATCAGACTCC CCAGTCTATT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCAGATCTGT CTCCGATCGT CTCCGATC	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA AGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TTCAAGTGGG AAATGAGTT TTCATCATC ACGCTCACC ACTGTTCATCCCC ACTGTTCAGCCC TTGTAGGAGT TTCGTTCATC TCGTAGAGTT TCGTTCATC TGTTAGGCTTC TGTCAGAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTACT TTAAATCAAT TATAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTCC AACTTGGCCC AACTTGGCCC	GACATGATAA TAACCATTAT AAACCCTTAT TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGATAACGA TGTGTCGTC GCTGGTACACT AGTGGACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAAGTATA GGTAACGA CAGTGTTATC	GATACATTGA AAGCTGCAAT AAGCTGCATA CAATTCCAGCA CTCCAGCA GAGACATGTG AAATCGACGC CTGCCGCTTA AGAGGACATGTG AGAGGACATGTG AGAGGACATGTG AGAAGATTTTTT AAACTCACGT TATGAGTAAA TCGTGAGAGAAAAAACCAGCCA TCGCCAGTTA GATCAAGGCA GATCAAGGCA GATCAAGGCA AACCAGCCAGTTA GATCAAGGCA CACCAGTTA GATCAAGGCA CACTCATGGTT	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTCGA ATTACATCA ATGACAGAA ATGACAGAA ATGACAGAA ATGACAGAA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCCATGT TCCCCCATGT TGCATAATTC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGG 2801 CGGGGAGCGG TATCAAGCTCA CTCAAAGGCC 2901 CCAGGAACCG TAAAAAGGC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGCCGTT 3101 CTCCCTTCGG GAACCGTGC GCCTTTCTCAT 3201 CGGTCAGCC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA CTACTAAAAGAT AGGAGTTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3501 ATTATCAAAA AGGATCTTCA CCTAGATCCT 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT 4001 GCCATTGCTA CAGGCATCGT GCTACTCAGCG 4101 TGTGCAAAAA AGCGGTTTAGC TCCTTCGGCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGCCTACT GTAATACGGT CGTTATTTCCA AGCCGGTTTG GTAATACGGT TCCCCCTGGA AGCTACT TTACAGAGTTC GTACTACT TTTAAATTAA ATCTGTCTAT TCGCAGACCC CCAGTCTATT TCGTCGTTTG CTTCGTTTG TTCGTCGTTTG CTCCGATCGT TTCTGTGACT TTCTTGGATCT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGCATTAATT CGTATTGGG TATCCACAGA TAGGCTCCGG GTAGGTACT TCTTGAGTCC TTGAAGTACT TTGAAGTGGT ATCCACCAGA ATTCACCGGCAA TTCTACCGGCAA TTCTACCGGCAA TTCTACCGCAA TTGTCACC AATTGTTCAC GTATGCTTCACC GTATGGCTTC TGTCAGAGTT TGTCAGAGT GTAGGCTTC TGTCAGAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCCGGTAA AACCAAC TCTGAACCA TCTGAACCA CTAAGTTCGCG GCTCAACTA ACAAACCAC CTCAAGTT GGGAAGCTAC ATCAGGTCC AAGTTCGCCC AAGTTGCCC AAGTTGCCC AAGTTGCCC AAGTTGCCC AAGTTGCCC AAGTTGCCC AACCAAGTC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA GCATCACAC TGTTCCGACC TAGGTCGTC GACACGACT GCTGGTACGC CTGGTACGC CTAAAGTAT TGACTCCCCG TATACACAC TATCCAGCAT TGTCCCAC TATCCAGCAT TATCAGCAT TATCAGCAT CGTTATCCAAC CAGTTATC CATTCTGAGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGACCACTG CTGCCGGCTTA GCTCCAAGGT ATCGCCACTG AGAAGAGAGACAG AGAAGAACACACCA TGGGTTTTTT AAACTCACGT ATACAGCTA AAACCACCA AAACCAGCA TGGGTAGATAA TCGGCAGTTA GATCAAGGCG ACTCAATGGT TAGATATATG	TGAGTTTGGA AAACAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCGTGTGTG GCAGCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTAT CCTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTGCA AGTTACATA GCCGGAAGGG ATAGTTTGCA AGTTACATA AGCAGCAC AGTTACATGA AGGCACCCAC GCGCACCGAG	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGATCG GCCGGAAGCA CGTGCCAGCT CGTCCGCTG CAGCAAAAGG GTCGCCATT CACGAACCC CTGGTAACAG CTGCGCTCT CAGCACATTA CAGCAGATTA CGGAGGGCT CCGAGCGCAA CGGTCTGCT CCAGCAGACTT CCCACTTT CCCCATGT TCCCCCATGT TCCTCTCCCCTTCC TTGCTCTTCC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAA 2701 GCATTAATGA ATCGCCAAC GCCGGGGAC 2801 CGGCGAGCGG TATCAGCTCA CTCAGAGGCC 2901 CCAGCAAGCA CTATAAAGAT ACCAGCGGTT 3101 CTCCCTTCGG GAAGCGTGGC GCCTTCTCTG 3201 CCGTTCAGCC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCCAGCGTTG TAGAGAGCT 3401 CTGAAGCCAG TACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATT CAAGAAGAGT 3501 TGCATAATCA GTGAGGCACC TATCTCAGCC 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCAT 4001 GCCATTGCTA CAGGCATCT GGGTGCCAT 4001 GCCATTGCTA CAGGCATCT GGGTGCCAT 4001 TGTGCAAAAA AGCGGTTGCT GGTGTCACGC 4101 TGTGCAAAAA AGCGGTTACC TCCTTCCGTC 4201 TCTTACTGTC ATGCCATCCC TAAGATCTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCT GTAACTATC GTAACTATC GTACTACT TCTTGACTT TTTTAAATTAA ATCTGTCTAT CCGGAGACCC CCAGTCTATT TCGTCGTTTG CTCGGATCGT TTCTCGTCGATCGT TTCTCGTGGACT CTCATACGAGAGA CC CATAGCAGAGA CC CATAGCAGAA	ACTGTGCCTT TGATGCTATT GGAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTACGGGG AAATGAAGTT TTCGTTCATC ACGCTCACCG GTATGCTTCACGGG GATTGCTCACCG GTATGGCTTC TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGTACT TGTCAGAAGT CTTTAAAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG TGCGCTCTCC CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC TTAAATCAACT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGGCC CACTTGGCCC CCCAAGTT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGACA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGACC TGGTCGACC TGGCTACACC TGGCTACACT GCTGGTAGCG ACTGGAACGA TTAAGGTATA TGACTCCCCG TATCAGCAAT AGTAAGTAGT GGTTCCCAAC CAGTGTTATC AGTAGTAGT AGTAGT AGT	GATACATTGA AAGCTGCAAT AAGCTGCACA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGACGC ATGCCCATA GCTCCAAGCT ATCGCCACTG AGAAGACACACACACACACACACACACACACACACACA	TGAGTTTGGA AAACAGTTA CAACATACGA GGAAAACTGA TGCGCTGATTA CACATACGA GGAAAACGGT TCAAGTCAGA CCGGATACCG GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAACTACGA AACTACGAA AACTACGAA AACTACGAA AACTACGAACACACA AGGATTTG GCAGCAGCCA AGGAACGA AGGATTACATGA ATGCAGCACCAC AGGCACCAC AGACCACA AAAACTCTCA	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGGCAGCT CGTCGGCTG CAGCAAAAGG GGTGGCAAAA GGTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCGAGCGCAG CAACCGTTGTT TCCCCCATGT TGCATAATTC TTGCTCTTGC AGGATTACC AGGATTATC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTG 2401 TTGCATTCAT TTTATGTTTC AGGTTCATGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG 2701 GCATTAATGA ATCGCGCAAC GCCGGGGAG 2801 CGGCGAGCGG TATCAAGCC GCGCGGGGAG 3001 CCCGCACAGGA CTATCAAGAT ACCAGCGGT 3101 CTCCCTTCGG GAAGCGTGC GCCTTATCCG 3201 CCGTTCAGC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGA GCAGGGTTG TACCTATCG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGAT 3501 ATTACCAAAT AGGACTCTAC CCTAAGACCT 3701 TGCTTAATCA GTGAGGCAC TATCTCAGCG 3801 TACCATCTGG CCCCAGTGCT GCAATGATCA 4001 GCCATTGCTA CAGGCATCT GCATTGATCAC 4001 GCCATTGCTA CAGGCATCT CAAGAACCT 4001 GCCATTGCTA CAGGCATCT GCATCTCAT 4001 GCCATTGCTA CAGGCATCA TCCTCCGT 4101 TGTGCAAAAA AGGGGTTTGC TCCTTCGGTC 4201 TCTTACTGC ATGCCATCCG TAAGAACCT 4301 CCGGCGCTCAA TACCGGCATAA TACCGCGCCCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCAGCT GTAACATCG TACAGGTTC GTTACAGGTTC GTTACAGGTTC CTTTGATCTT TTTAAATTAA ATCAGTCTAT CGCGAGACCC CCAGTCTATT CCTCGTCTT CTCCGATCGT TCCTCGTTTC CTTCGTCGTTT CTTCGTCGATCGT TCCTCGTCACC CATAGCAGAC CTCATAGCAGAA CTCGTGCACC	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TTCAACGGGA AATGCACCACA TTCTACGGGCA AATGTATT CCGCTCACCG AATTGTTGCC GTATGGCTC TGTCAGAAGT TGTCAGAAGT TTCATAAAAGT CTTTAAAAGT CTAACTGATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTTCTCC CAGTTCGGTG AACACGGTAACTA ACAAACCACC TCTGACGTC TTAAATCAAT CATAGTTGCC GCTCAGATTGCC GCTCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCCC CAACCAAGTC CTCACCATCT CTACACTACTAT	GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACCA TGTTCCGACC TAGGTCGTTC GACACGACT CGGCTACACT GCTGGTACCT TGACTCCCCG TATCAGCAA AGTAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA CAGTCTCAAC CAGTGTTATC ATTCTGACAA CAGTGTTATC ATTCTGACAAA CTTAACTTTTCTACAT	GATACATTGA AAGCTGCAAT AAGCTGCATA CAATTCCAGCA CTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCT ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTTGCA ATGGATTTTGCA ATGGACAC ATGGACACAC ATGGACACAC AGGCACCAG AAACTCTCA AAAACTCTCA AAAACTCCAC AAAACTCTCA AAAACTCCAC AAAACTCCAC AAAACTCCAC AAAACTCCAC AAAACTCCAC AAAACTCCAC AAAACTCCAC AAAACTCCAC ACGCGACCACA AAAACTCCAC AGGCGACCACA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCCTCTG CAGCACATTA TGGTCATGAC CAGCTCTG CAGCACATTA TGGTCATGAG CAGCTCTG CCAGCGCAG CAACCTCT CCAGCACTT TCCCCCATGT TCCCCCATGT TTGCTCTTGC AGGATCTTCC AGGATCTTAC AAACAGAAG
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATCAA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCAC 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCGGCCAAC GCGGGGGGG 2801 CGGCGAGCGG TATCAAGCTCA CTCAAAGGCC 2901 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG 3001 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAAGCGTGGC GCCTTATCCAC 3301 GATTAGCAGC GGCAGCTGC GCCTTATCCAC 3301 GATTAGCAGA GCGAGGTATG TAGGCGGTGC 3401 CTGAAAGCAG AAAAGGATCT CAAGAAAGAT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAAGAT 3501 ATTATCAAAA AGGATCTTCA CCTAGATCCT 3701 TGCTTAATCA GTGAGGCAC TATCTCAGCG 3801 TACCATCTGG GAAAACTTTAT CCGCCTCCAT 4001 GCGCATGCAT CAGGCATCT GCGATCACC 4101 TGTGCAAAAA AGCGTTAGC TGCTCACCC 4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC 4201 TCTTACTGTC ATGCCATCCG TAAAGATCTT 4301 CCGCGCTCAA TACCGGGATAA TACCGCCCCAA 4401 CCGCGTCTAA TACCGGGATAA TACCGCCCCACAT 4401 CCGCGTCTAA TACCGGGATAA TACCGCCCCACAT 4401 CCGCGTCTAA TACCGGGATAA TACCGCCCCCAT 4401 CCGCGTCTAA TACCGGGATAA TACCGCCCCACAT CCCCCCCCATCATCCC TAAAGATCCTACCCCACCACATCCCCCCCCCATCACCCCCCCC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTACT GTAATACGGT CGTTTTTCCA AGCCGGTTGG TAACACGCT GTAACTACC GTACTACC GTACTACC GTACTACC GTACTACC GTACTACC GTACTACC GTACTACT TCTTAAATTA ATCTGTCTAT TCGCAGACC CCAGTCTATT TCGCGTCGT TTCCGTTGG TTCCGTTGG TTCCGTCGACC CATCCTACT CATAGCAGAA CCCGATCCTACC CATCCTACC CATCCTCC CATCCACC CATCCTACC CATCCACC CATCCTACC CATCCACC CATCCACC CATCCACC CATCCACC CACCC CACCCACC	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGG GTAGGTACT TCTTGAGTCC TTGAAGTAGT TTGAGTGGT AAATGAAGT TTCATCACGG AAATGATACT TCGTCACC GAATTGTTGC GTAGAGTT TCGTCACC GTAGGCTTC TGTTGAAGT TTTTAAAAGT TGTTGAAGT CTTTAAAAGT TTTAAAAGT CTTTAAAAGT TGTTGAATAC TGTTGAATAC TGTTGAATAC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCCGTAA ACAAACCAC TTGAAACCAC CTTAAATCAAC CTCCAGATT GGGAAGCTAC AAGTCAGCTC AAGTGGCC AAGTTCAGCTC CAGATT GGCAAGCTAC ATCAGCTCC AAGTTGCCC CAGATT CTCACACTT CTCACATT CCACACTT CCACACTT TCACACTCT TCAGACTCT TCAGCTCT TCAGACTCT TCAGCTCT TCAGACTCT TCAGACTCT TCAGACTCT TCAGACTCT TCAGACTCT TCAGACTCT TCAGACTCTT	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCAGCAGTT GACACAGCTT GACACAGCTT TGACTACAC CTAGATAGCA CTAAAGTATAT GACTCCCCG TATCAGCAA TATCAGCAAT AGTAAGTAGT GGTTACCAAC CAGTGTTACC CAGTTATCACAC GGAAAACGTT TACATTCCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC CTGCCGGTTA GCTCCAAGCT ATCGCCACT ATCGCCACT ATCGCCACT ATCGCCACT ATCGCACT AAACTCACGC ATCGTATATT AAACTCACGT ATCACGACT AAACTCACGC ACTCATAGACT AAACTCAGCT CACCACTTA CACCACTT TACTAGTATATC CTTCGGGGCG CACGCTTTCT TATTATTGAA	TGAGTTTGGA AAACAAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCGTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA CCTTGGTCTGA AACTACGATA GCCGGAAGGG ATACTTTGCA AGTACATA CCCGGAAGGA AAACTCTCA GGCGACCGA AAACTCTCA GGGGACCAA GCATTTATA	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGATCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAAA GGTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA CGGAGGCGCAG CAGCTTACCAA CGGGAGGGCT CCGAGCGCAG CAACCTTGTT TCCCCCATGT TCCACCATGT TTCCTCTTGC AGGATCTTAC AAACAGAAAG GGGTTTATTGT
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG 2301 CTAGAATACA GTGAAAAAAA TGCTTTATTT 2401 TTGCATTCAT TTTATGTTT AGGTTCAGGG 2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCAC 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC 2701 GCATTAATGA ATCCGCCAAC GCCGGGGGAC 2801 CCGGCAGCGG TATCAGCTCA CTCCAGGGGGT 3001 CCCGACAGGA CTATAAAGAT ACCAGGCGTT 3101 CTCCCTTCGG GAACCGTGG GCTTTCTCAT 3201 CCGTTCAGC CGACCGCTGC GCCTTATCCG 3301 GATTAGCAGC GCAGCGTTGT TAGCCGTGC 3401 CTGAAGCCAG TACCTTCGG AAAAAAGAGTT 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCC 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC 3901 AAGTGGTCCT GCAATGATAC 4001 GCCATTGCTA CAGCATCCT GGTGTCACCT 4101 TGTGCAAAAA AGCGTTTGT CGTGTCCACT 4201 TCTTACTGTC ATGCCATCCG TAAGAGTCT 4301 CGGGGTCAA TACGGGATAA TACCGGCCCA 4401 CGCTGTTGAG ATCCAGTCCT TACTCAGCC 4401 CGCTGTTGAG ATCCAGTTCG TAAGAGTCT 4301 CGGGGTCAA TACGGGATAA TACCGGCCCA 4401 CGCTGTTGAG ATCCAGTTCG TAGAAGAGCCC 4401 CGCTGTTGAG ATCCAGTTCG TAGAAGCCCA 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCT GTAACTATC GTAACTATC GTACTACT TCTTGACTT TTTTAAATTA ATCTGTCTAT CCCGATCGT TCCGATCGT TTCCGTCGATCGT TTCCGTCGATCGT TTCCGTCGATCGT TCCTTGGACCT CATAGCAGAA CTCGTGCACC GACACCGACA TAGCAGAAATA	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTACGGG GAATGATCT TCGTTCATC ACGCTCACCG GTATGCTTCAC GTAGGTACT TCTTCAAGAGG TCTTCAAGAGG CGTATGCTTC TGTCAGAAGT TGTTCAGAAGT CACTGATACT TGTTGAAAAGT CAACTGATCT TGTTGAATAC AACAAATAGG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG TGCGCTCTCC CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC TAAATCAACTAC CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGGCC GCTCAGATT TCAGCTCC GCTCCAAGTT TCAGCTCT GCTTCCTCT GGTTCCGCGC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCAGCAGTT GACACAGCTT GACACAGCTT TGACTACAC CTAGATAGCA CTAAAGTATAT GACTCCCCG TATCAGCAA TATCAGCAAT AGTAAGTAGT GGTTACCAAC CAGTGTTACC CAGTTATCACAC GGAAAACGTT TACATTCCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC CTGCCGGTTA GCTCCAAGCT ATCGCCACT ATCGCCACT ATCGCCACT ATCGCCACT ATCGCACT AAACTCACGC ATCGTATATT AAACTCACGT ATCACGACT AAACTCACGC ACTCATAGACT AAACTCAGCT CACCACTTA CACCACTT TACTAGTATATC CTTCGGGGCG CACGCTTTCT TATTATTGAA	TGAGTTTGGA AAACAAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCGTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA CCTTGGTCTGA AACTACGATA GCCGGAAGGG ATACTTTGCA AGTACATA CCCGGAAGGA AAACTCTCA GGCGACCGA AAACTCTCA GGGGACCAA GCATTTATA	CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGATCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAAA GGTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCATACAG CAGCACATAA CGGGAGGGCT CCGAGCGCAA CAGCACATT TCCCCCATGT TCCCCCATGT TCCACCATGT TCCACCAACACACACACACACACACACACACACACACA

## pMT-5'HisFlag-Dm43 (2-274) STOP

1 TCGCGCGTTT C									
101 TCAGGGCGCG T									
201 CCGCACAGAT G									
301 TACGCCAGCT G									
401 TGAATTAATT C									
501 GGCCCCCCAC C	CGCCCACCGC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA A	AGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT C	GAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG A	ATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
						> M H	н н н	ннтр	Y K D
901 GACGATGACA A	GGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
	GTS	E L N		D C W E		R F O	R L V N	DGE	N C E
1001 TCGAGGTGTT C	TGCCGGTGC	TGGCGAGAAC			-	-	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
>F E V F	CRC		LOLO		T A O	TNHT	E V I	АТТ	LAAL
1101 GCATGTGGCC A									
> H V A	KRL	SCSR	RTT	G D V	F P A S	R A O	RIG	GFFL	T. Y V
1201 ATCTACTACA A	GCAGCCCAC	GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
	OPT	H N F	I K I	E V S P	R T W	O E L	T D Y A	L D L	R K D
1301 GTCCGGAGCG G	~								
>S P E R			A Y M L		T O E	O A F R		L D Y	C O G L
1401 GGACAATCTG G									
		D R V E		G A K		S A L	M O K	O O R A	N G V
1501 AGTCTCACAT A					~ ~		~	~ ~	
	E L E	G L R		O A S O			A A Y N	A O K	K O L
1601 CGGCTGGTCA T								~	~
		L P P					O S V		
>A A G H 1701 TCCAGATGAG A			S Q I F			F A D I			R K S T
				TAGAAGGGCA	ATTCTGCAGA	TATCCAGCAC	AGIGGCGGCC	GCTCGAGTCT	AGAGGGCCCG
		T T S T	G N Q	maa	aaama aaaam	a. ma. ma. aa	1 ma1 aa1 mma		aaama, ma, a
1801 CGGTTCGAAG G									
1901 CCTCGACTGT G									
2001 ATTTGTGATG C									
2101 TGTGGGAGGT T									
2201 ATAGCTGTTT C									
2301 TAACTCACAT T									
2401 GTTTGCGTAT T									
2501 ACGGTTATCC A									
2601 TTCCATAGGC T									
2701 CTGGAAGCTC C									
2801 ACGCTGTAGG T									
2901 TATCGTCTTG A									
3001 AGTTCTTGAA G									
3101 CTCTTGATCC G									
3201 ATCTTTTCTA C									
3301 ATTAAAAATG A									
3401 TCTATTTCGT T									
3501 GACCCACGCT C									
3601 CTATTAATTG T									
3701 GTTTGGTATG G									
3801 ATCGTTGTCA G									
3901 TGACTGGTGA G									
4001 CAGAACTTTA A									
4101 GCACCCAACT G	SATCTTCAGC	ATCTTTTACT	TTCACCAGCG	$\tt TTTCTGGGTG$	AGCAAAAACA	GGAAGGCAAA	ATGCCGCAAA	${\tt AAAGGGAATA}$	AGGGCGACAC
4201 GGAAATGTTG A	ATACTCATA	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA
4301 AAATAAACAA A	TAGGGGTTC	CGCGCACATT	TCCCCGAAAA	GTGCCACCTG	ACGTCTAAGA	AACCATTATT	ATCATGACAT	TAACCTATAA	AAATAGGCGT
4401 ATCACGAGGC C	CCTTTCGT								

## pMT-5'HisFlag-Dm43 (2-172) STOP

	1 T	CGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
	101 T	CAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
- :	201 C	CCGCACAGAT	GCGTAAGGAG	AAAATACCGC	ATCAGGCGCC	ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT
- 1	301 I	CACGCCAGCT	GGCGAAAGGG	GGATGTGCTG	CAAGGCGATT	AAGTTGGGTA	ACGCCAGGGT	TTTCCCAGTC	ACGACGTTGT	AAAACGACGG	CCAGTGCCAG
4	401 T	GAATTAATT	CGTTGCAGGA	CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
	501 G	GCCCCCCAC	CGCCCACCGC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
6	501 C	CAAGTCCCCA	AAGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
,	701 A	AGAGGTGAAT	CGAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
8	301 A	AATCAAGTG	AATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
								> M H	н н н	${\tt H}  {\tt H}  {\tt T}  {\tt D}$	Y K D
9	901 G	BACGATGACA	AGGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	${\tt GATTGGTTAA}$	TGATGGCGAA	AACTGCGAGT
	>	D D D	K G T S	E L N	I F D	D C W E	L V Q	R F Q	${\tt R}  {\tt L}  {\tt V}  {\tt N}$	D G E	N C E
10	001 T	CGAGGTGTT	CTGCCGGTGC	TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
	>F	EVF	C R C	W R E	L Q L Q	H L F	T A Q	${\tt T}  {\tt N}  {\tt H}  {\tt T}$	E V I	A T T	L A A L
1:	101 G	CATGTGGCC	AAGCGACTGT	CGTGCTCCCG	ACGCACCACC	GGGGACGTTT	TCCCGGCATC	TCGCGCTCAA	AGGATCGGAG	$\tt GTTTCTTTCT$	GCTCTACGTA
	>	H V A	K R L	S C S R	R T T	G D V	F P A S	R A Q	R I G	G F F L	L Y V
12	201 A	ATCTACTACA	AGCAGCCCAC	GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
	>	I Y Y	K Q P T	H N F	I K I	E V S P	R T W	QEL	T D Y A	L D L	R K D
13	301 G	STCCGGAGCG	GAAGGACACT	CATCAGATCG	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
	>S	PER	K D T	H Q I	${\tt A}  {\tt Y}  {\tt M}  {\tt L}$	WRL	T Q E	Q A F R	F T A	L D Y	C Q G L
14	401 G	GACAATCTG	GTGGACTACG	ACCGTGTGGA	GACCTAGCAA	GGGCAATTCT	GCAGATATCC	AGCACAGTGG	CGGCCGCTCG	AGTCTAGAGG	GCCCGCGGTT
	>	D N L	V D Y	D R V E	T						
15	501 C	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT	TCTACGCGTA	CCGGTCATCA	TCACCATCAC	${\tt CATTGAGTTT}$	AAACCCGCTG	ATCAGCCTCG
16	501 A	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TGAGTTTGGA	CAAACCACAA	CTAGAATGCA	GTGAAAAAAA	TGCTTTATTT	GTGAAATTTG
1'	701 I	GATGCTATT	GCTTTATTTG	TAACCATTAT	AAGCTGCAAT	AAACAAGTTA	ACAACAACAA	TTGCATTCAT	${\tt TTTATGTTTC}$	AGGTTCAGGG	GGAGGTGTGG
18	301 G	SAGGTTTTTT	AAAGCAAGTA	AAACCTCTAC	AAATGTGGTA	TGGCTGATTA	TGATCAGTCG	ACCTGCAGGC	ATGCAAGCTT	GGCGTAATCA	TGGTCATAGC
19	901 I	GTTTCCTGT	GTGAAATTGT	TATCCGCTCA	CAATTCCACA	CAACATACGA	GCCGGAAGCA	TAAAGTGTAA	AGCCTGGGGT	GCCTAATGAG	TGAGCTAACT
20	001 C	CACATTAATT	GCGTTGCGCT	CACTGCCCGC	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	GCATTAATGA	ATCGGCCAAC	GCGCGGGGAG	AGGCGGTTTG
2:	101 C	CGTATTGGGC	GCTCTTCCGC	TTCCTCGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA	CTCAAAGGCG	GTAATACGGT
22	201 I	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA
23	301 I	TAGGCTCCGC	CCCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA	CCCGACAGGA	${\tt CTATAAAGAT}$	ACCAGGCGTT	TCCCCCTGGA
24	401 A	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC	GCTTTCTCAT	AGCTCACGCT
25	501 G	STAGGTATCT	CAGTTCGGTG	TAGGTCGTTC	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACTATCG
26	501 I	CTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC
2	701 I	TGAAGTGGT	GGCCTAACTA	CGGCTACACT	AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT
						TGTTTGCAAG					
						TAAGGGATTT					
						CTTGGTCTGA					
						AACTACGATA					
						GCCGGAAGGG					
						ATAGTTTGCG					
						AGTTACATGA					
						ATGGCAGCAC					
						GGCGACCGAG					
						AAAACTCTCA					
						GGGTGAGCAA					
						GCATTTATCA					
				ACATTTCCCC	GAAAAGTGCC	ACCTGACGTC	TAAGAAACCA	TTATTATCAT	GACATTAACC	TATAAAAATA	GGCGTATCAC
4:	101 G	BAGGCCCTTT	CGT								

### pMT-5'HisFlag-Dm43 (2-125) STOP

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
201	CCGCACAGAT	GCGTAAGGAG	AAAATACCGC	ATCAGGCGCC	ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT
301	TACGCCAGCT	GGCGAAAGGG	GGATGTGCTG	CAAGGCGATT	AAGTTGGGTA	ACGCCAGGGT	TTTCCCAGTC	ACGACGTTGT	AAAACGACGG	CCAGTGCCAG
401	TGAATTAATT	CGTTGCAGGA	CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501	GGCCCCCCAC	CGCCCACCGC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601	CAAGTCCCCA	AAGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701	AGAGGTGAAT	CGAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801	AAATCAAGTG	AATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
							> M H	н н н	H H T D	Y K D
901	GACGATGACA	AGGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
3	> D D D	K G T S	E L N	I F D	D C W E	L V Q	R F Q	${\tt R}  {\tt L}  {\tt V}  {\tt N}$	D G E	N C E
1001	TCGAGGTGTT	CTGCCGGTGC	TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
3	>F E V F	C R C	W R E	L Q L Q	H L F	T A Q	T N H T	E V I	A T T	L A A L
1101	GCATGTGGCC	AAGCGACTGT	CGTGCTCCCG	ACGCACCACC	GGGGACGTTT	TCCCGGCATC	TCGCGCTCAA	AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
3	> H V A	K R L	S C S R	R T T	G D V	F P A S	R A Q	R I G	G F F L	L Y V
1201	ATCTACTACA	AGCAGCCCAC	GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCTAACAAG
3	> I Y Y	K Q P T	H N F	I K I	E V S P	R T W	QEL	T D Y A	L D L	R
1301	GGCAATTCTG	CAGATATCCA	GCACAGTGGC	GGCCGCTCGA	GTCTAGAGGG	CCCGCGGTTC	GAAGGTAAGC	${\tt CTATCCCTAA}$	CCCTCTCCTC	GGTCTCGATT
1401	CTACGCGTAC	CGGTCATCAT	CACCATCACC	ATTGAGTTTA	AACCCGCTGA	TCAGCCTCGA	CTGTGCCTTC	TAAGATCCAG	ACATGATAAG	ATACATTGAT
	GAGTTTGGAC									
1601	AACAAGTTAA	CAACAACAAT	TGCATTCATT	TTATGTTTCA	GGTTCAGGGG	GAGGTGTGGG	AGGTTTTTTA	AAGCAAGTAA	AACCTCTACA	AATGTGGTAT
1701	GGCTGATTAT	GATCAGTCGA	CCTGCAGGCA	TGCAAGCTTG	GCGTAATCAT	GGTCATAGCT	GTTTCCTGTG	TGAAATTGTT	ATCCGCTCAC	AATTCCACAC
1801	AACATACGAG	CCGGAAGCAT	AAAGTGTAAA	GCCTGGGGTG	CCTAATGAGT	GAGCTAACTC	ACATTAATTG	CGTTGCGCTC	ACTGCCCGCT	TTCCAGTCGG
1901	GAAACCTGTC	GTGCCAGCTG	CATTAATGAA	TCGGCCAACG	CGCGGGGAGA	GGCGGTTTGC	GTATTGGGCG	CTCTTCCGCT	TCCTCGCTCA	CTGACTCGCT
2001	GCGCTCGGTC	GTTCGGCTGC	GGCGAGCGGT	ATCAGCTCAC	TCAAAGGCGG	TAATACGGTT	ATCCACAGAA	TCAGGGGATA	ACGCAGGAAA	GAACATGTGA
	GCAAAAGGCC									
	CAAGTCAGAG									
2301	CGGATACCTG	TCCGCCTTTC	TCCCTTCGGG	AAGCGTGGCG	CTTTCTCATA	GCTCACGCTG	TAGGTATCTC	AGTTCGGTGT	AGGTCGTTCG	CTCCAAGCTG
	GGCTGTGTGC									
	CAGCAGCCAC									
	ATTTGGTATC									
	GTTTGCAAGC									
	AAGGGATTTT									
	TTGGTCTGAC									
	ACTACGATAC									
	CCGGAAGGGC									
	TAGTTTGCGC									
	GTTACATGAT									
	TGGCAGCACT									
3501	GCGACCGAGT	TGCTCTTGCC	CGGCGTCAAT	ACGGGATAAT	ACCGCGCCAC	ATAGCAGAAC	TTTAAAAGTG	CTCATCATTG	GAAAACGTTC	TTCGGGGCGA
	AAACTCTCAA									
	GGTGAGCAAA									
	CATTTATCAG								CATTTCCCCG	AAAAGTGCCA
3901	CCTGACGTCT	AAGAAACCAT	TATTATCATG	ACATTAACCT	ATAAAAATAG	GCGTATCACG	AGGCCCTTTC	GT		

## pMT-Dm43 full noSTOP-FlagMycHis

1 TCGCGCGTTT CGG	TGATGAC GGT	TGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAG									
201 CCGCACAGAT GCG	TAAGGAG AA	AATACCGC	ATCAGGCGCC	ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT
301 TACGCCAGCT GGC									
401 TGAATTAATT CGT									
501 GGCCCCCCAC CGCC									
601 CAAGTCCCCA AAG									
701 AGAGGTGAAT CGAA									
801 AAATCAAGTG AATC									
901 AGAATGGAGC TGAA									
		D D C		V O R F	O R L		G E N C	E F E	V F C
				~					
1001 GGTGCTGGCG AGAZ							T I A		V A K R
	~	~	L F T A	Q T N		V I A T			
1101 ACTGTCGTGC TCCC									
		T G D	VFP		AQRI	G G F	FLL	YVIY	Y K Q
1201 CCCACGCACA ACT									
		I E V		T W Q E	L T D		D L R K	D S P	ERK
1301 ACACTCATCA GATO									
			R L T Q			TALD	Y C Q	G L D	N L V D
1401 CTACGACCGT GTG									
> Y D R V		A G A			A L M Q	K Q Q	R A N	G V S L	T Y E
1501 CTGGAGGGTC TGC						TACAATGCCC		ATTGGCGGCT	GGTCATGAGC
		D Q A		L C E L	E A A	Y N A	Q K K Q	L A A	G H E
1601 ACGCTTTACC GCCC	CTCTCAA AT	'ATTCGGCC	ATTTGCGAGA	AGTCTTTGCC	GATATCCAAA	GTGTTCTAGG	AGCTAGAAAG	AGTACTCCAG	ATGAGAAATG
>H A L P P	S Q I	F G	H L R E	V F A	D I Q	S V L G	A R K	S T P	D E K C
1701 CACCACAACA TCTA	ACAGGCA ACC	CAGTTGGA .	AGTGCGCCAG	AGGGTGCGGA	ACAAGGCCAT	GTACGGCGTC	GAGGAGCGGG	AGCCGCAACA	CCAGACGGAT
> T T T S	T G N	Q L E	V R Q	R V R	N K A M	Y G V	E E R	E P Q H	Q T D
1801 GAACTAGAAG TGCA	AGCTGGA GG	TCAACGAG .	ACTTATCAAC	GCCGCATGTC	CTCGGCCACC	GTTTTCCAGA	GGGAACTTCC	AGAAGACGTG	CAGCAAGAGT
> E L E V (	LEI	V N E	T Y Q	R R M S	SAT	V F Q	R E L P	E D V	QQE
1901 ATGAGATGAT TGAG	GTTTAGT GAG	.CGACGAGG	AAATGGAAGT	GGGTGAAAGC	GAGGAGGTCA	CGGAAGAAGA	ACTCAAAGCT	ATTTTGGATA	CCAAGGGCAA
>Y E M I E	F S D	DE	E M E V	G E S	E E V	T E E E	L K A	I L D	T K G N
2001 TTCTGCAGAT ATC	CAGCACA GTO	GGCGGCCG	CTCGAGTCTA	GAGGGCCCGC	GGTTCGACTA	CAAGGATGAC	GATGACAAGG	GCGAGCAGAA	GCTGATCTCC
2001 TTCTGCAGAT ATCC > S A D I	CAGCACA GTO Q H S	GGCGGCCG G G R	CTCGAGTCTA S S L	GAGGGCCCGC E G P	GGTTCGACTA R F D Y	CAAGGATGAC K D D	GATGACAAGG D D K	GCGAGCAGAA G E Q K	GCTGATCTCC L I S
2001 TTCTGCAGAT ATCC > S A D I 2101 GAGGAAGATC TGAC	CAGCACA GTO Q H S CGCGTAC CGO	GGCGGCCG G G R GTCATCAT	CTCGAGTCTA S S L CACCATCACC	GAGGGCCCGC E G P ATTGAGTTTA	GGTTCGACTA R F D Y	CAAGGATGAC K D D	GATGACAAGG D D K	GCGAGCAGAA G E Q K	GCTGATCTCC L I S
2001 TTCTGCAGAT ATCC  > S A D I  2101 GAGGAAGATC TGAC  > E E D L	CAGCACA GTO Q H S CGCGTAC CGO F R T O	GGCGGCCG G G R GTCATCAT G H H	CTCGAGTCTA S S L CACCATCACC H H H	GAGGGCCCGC E G P ATTGAGTTTA H	GGTTCGACTA R F D Y AACCCGCTGA	CAAGGATGAC K D D TCAGCCTCGA	GATGACAAGG D D K CTGTGCCTTC	GCGAGCAGAA G E Q K TAAGATCCAG	GCTGATCTCC L I S ACATGATAAG
2001 TTCTGCAGAT ATCC  > S A D I  2101 GAGGAAGATC TGAC  > E E D L C  2201 ATACATTGAT GAGC	CAGCACA GTO Q H S CGCGTAC CGO F R T O FTTGGAC AAA	GGCGGCCG G G R GTCATCAT G H H ACCACAAC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG	CAAGGATGAC K D D TCAGCCTCGA TGAAATTTGT	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT	GCTGATCTCC L I S ACATGATAAG AACCATTATA
2001 TTCTGCAGAT ATCC  > S A D I 2101 GAGGAAGATC TGAG > E E D L 2201 ATACATTGAT GAG 2301 AGCTGCAATA AACA	CAGCACA GTO Q H S CGCGTAC CGO F R T O FTTGGAC AAA AAGTTAA CAA	GGCGGCCG G G R GTCATCAT G H H ACCACAAC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG AGGTTTTTTA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCTCTACA
2001 TTCTGCAGAT ATCC  S A D I  2101 GAGGAAGATC TGAC  E E D L 2201 ATACATTGAT GAC: 2301 AGCTGCAATA AAC2 2401 AATGTGGTAT GGC:	CAGCACA GTO Q H S CGCGTAC CGO F R T O FTTGGAC AAA AAGTTAA CAA	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA TGCAAGCTTG	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG AGGTTTTTTA GTTTCCTGTG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCTCTACA ATCCGCTCAC
2001 TTCTGCAGAT ATCC  S A D I  2101 GAGGAAGATC TGAC  E E D L  2201 ATACATTGAT GAG: 2301 AGCTGCAATA AACZ  2401 AATGTGGTAT GGC: 2501 AATTCCACAC AACZ	CAGCACA GTO Q H S CGCGTAC CGC F R T C FTTGGAC AAA AAGTTAA CAA FGATTAT GAT ATACGAG CCC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC	GCTGATCTCC L I S ACATGATAAG  AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT
2001 TTCTGCAGAT ATCC  > S A D I 2101 GAGGAAGATC TGAC  > E E D L . 2201 ATACATTGAT GAG: 2301 AGCTGCAATA AACC 2401 AATGTGGTAT GGC: 2501 AATTCCACAC AACC 2601 TTCCAGTCGG GAAC	CAGCACA GTO Q H S CGCGTAC CGO F R T ( TTTGGAC AAA AAGTTAA CAA TGATTAT GAC ATACGAG CCC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA GGAAGCAT	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAA TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGGAGA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGCCG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCT	GCTGATCTCC L I S ACATGATAAG  AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCAC
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC I R T ( ITTGGAC AAA AAGTTAA CAA TGATTAT GAT ATACGAG CCC ACCTGTC GTO	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA GGAAGCAT GCCAGCTG	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAGCGGT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA TGCAAGCTTCA GCCTGGGGTG TCGGCCAACC ATCAGCTCAC	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC TAATACGGTT	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCT TCAGGGGATA	GCTGATCTCC L I S ACATGATAAG  AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCA ACGCAGGAAA
2001 TTCTGCAGAT ATCC  S A D I 2101 GAGGAAGATC TGAC  E E D L 2201 ATACATTGAT GAG: 2301 AGCTGCAATA AAC; 2401 AATGTGGTAT GGC: 2501 AATTCCACAC AAC; 2601 TTCCAGTCGG GAG: 2701 CTGACTGGG GAG: 2801 GAACATGTGA GCA; 2801 GAACATGTGA GCA;	CAGCACA GTO Q H S CGCGTAC CGG I R T C ITTGGAC AAA AAGTTAA CAA TGATTAT GAT AACGAG CCC CTCCGTC GTC AAAGGCC AGC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACAAT TCAGTCGA GGAAGCAT GCCACGTG TCGGCTGC CCAAAAGGC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAGCGGT CAGGAACCGT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCAC AAAAAGGCCG	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CGTTGCTGGC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC TAATACGGTT GTTTTTCCAT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGGT TCAGGGGATA CCCCTGACGA	GCTGATCTCC L I S ACATGATAAG  AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S COCGTAC COC F R T C FTTGGAC AAI AAGTTAA CAI FGATTAT GAT ATACGAG CCC ACCTGT GTC AAAAGGCC AGC GTCAGAG GTC AAAAGGCC AGC GTCAGAG GTC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACACAAC ACACAAC GGAAGCAT TCAGTCGA TGCCAGCTG TCGGCTGC GCAAAAGGC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAGCGGT CAGGAACCGT CAGGAACCGT CCGACAGGAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA GCCTGGGGTG GCCTGGGCTG TCGGCCAACG ATCAGCTCAC AAAAAGGCCG	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG GGTTGCTGGC CCAGGCGTTT	CAAGGATGAC K D D TCAGCCTCGA TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC TAATACGGTT TATTTTCCAT CCCCCTGGAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCT TCAGGGGATA CCCCTGACGA GCGCTCTCCT	GCTGATCTCC L I S ACATGATAAG  AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GTTCCGACCC
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC I R T ( ITTGGAC AAA AAGTTAA CAA TAACGAG CCC ACCTGTC GTO CTCGGTC GTO AAAGGCC AGC TAAAAGGCA GTO AAAAGGCA GTO AAAAGGCA GTO ATAACCAG GTO ATAACCAG GTO ATAACCTG TCC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA GGAAGCAT GCCAGCTG TCGGCTGC CAAAAGGC GGCGAAAC CGCCTTTC	CTCGAGTCTA S S L CACCATCACC H H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAACCGT CAGGAACCGT CCGCACAGGAC TCCCTTCGGG	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCAC AAAAAGCCCT ATATAAAGATA AAGCCTGGCG	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CGTTGCTGGC CCAGGCGTTT CTTTCTCATA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC TAATACGGTT GCTTTTCCAT GCCCCTGGAA GCTCACCCTG	GATGACAAGG D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG GTATTGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT TAGGTATTCT	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCT TCAGGGGATA CCCCTGACGA GCGCTCTCTCA	GCTGATTTCC L I S ACATGATAATA AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GTTCCGACCC AGGTCGTTCG
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC F R T ( TTTGGAC AAA AGTTAA CAA TGATTAT GAA ATACGAG CCC ACCTGTC GTC AAAGGCC GTC AAAGGCC AGC ATTCAGGCC GTC AAAGGCC AGC ATTCAGAG GTC ATACCTG TCC TGTGTGT GTC	GGCGGCCG G G R GTCATCAT G H H AACCACAAC ACAACAAT TCAGTCGA GGCAGCTG TCGGCTGC CCACAAC GGCGAAAC GGCGCATTC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAGGGGT CAGGAACCGT CCGCACAGGAC CCTTCAGCC CCTTCAGCC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCCAGGTT TCGGCCACC AACAGCCCACC AAAAGGCCC TATAAAGATA AAGCGTGCG GACCGCTGCG	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTG GGGTCAGGGG GCGTAATCAT CCTAATGAGT CCTGGGGAGA TCAAAGGCGG CGTTGCTGGC CCAGGCGTTT CCTTTCTCATA CCTTATCACT CCTTTCTCATA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGCTAACTC GAGCTTACT GTAATACGTT GTTTTTCAT CCCCCTGGAA GCTCAGCCT TAACTACGTT GAACTATCGT TAACTATCGT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTTCGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT TAGGTATCC CTTGAGTTCC	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCGCTCTCCT AGTTCGGTGT ACCCGGTTAAG	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCTCTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GTTCCGACCC ACGCTCTTCC ACGCACCACAC ACACGACTTA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC F R T ( FTTGGAC AAA AAGTTAA CAA ATACGAG CCC CTCCGGTC GTC CTCGGTC GTC AAAAGGCC AG CTCAGAG GTC ATACCTA TCC CTCGGTC CTC ATACCTA TCC CTCGGTC CTC CTCAGAG GTC ATACCTA TCC CAGCCAC TGC CAGCCAC TGC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACAACAAT TCAGTGGA GGAAGCAT TCGGCTGC CCAAAAGGC GGCGAAAC CGCCTTTC GGAACCCCC GGAACAGG	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GCGAACGGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAGAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAGCTTG GCCTGGGGTG ATCAGCTAAC ATCAGCTCAAC ATAAAGCCC AAAAAGCCCG ATCAGCTGCG GACCGCTGCG GACCGCTGCG	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT TCAAAGGCG CCTTGCTGGC CCAGGCGTTT CTTTCTCATA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGTGGG GGTCATAGCT GAGCTTAACTC TAATACGGTT GTTTTTCCAT CCCCCTGGAA GCTCACGCTG TAACTACCGT TAACTACGTT AACAGAGTTCT	GATGACAAGG D D K CTGTGCCTTC GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCCCT GCTCCTTGAGTATCCC CTTGAGTATCTC CTTGAGTGCGATCGATCAGTAGTGCCATGAGTGCTCGTTGAGTGCTGTTGAGTGTGTGT	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGGCGTC TCAGGGGATA CCCTGACGA GCGCTCTTCCA AGTTCGGTGT ACCCGGTAAG GCCTAACTAA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTGCCCGCT TCCTCGGTCA ACGCAGGAAA GCATCACAAA GCTTCACAAC AGGTCGACC AGGTCGTTCG ACACGACTTA GGCTACACTA GGCTACACTA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGGATAC CGC F R T ( TTTGGAC AAA AGATTAA CAA TGATTAT GAA TATACGAC CGT TCTCGGTC GTT AAAGGCC AG TCTCAGGC GTC TCTCGGTC GTT AAAGCT TCC TCTCGGTC GTT ATACCTT TCC TCTGGTG ACC TCTCAGCAC TCC TCGGTGTATACTT TCC TCTGGTTATACTT TCC TCGGTGTT TCGGTGT TCGGT	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACAAC TCAGTCGA GGCAAGCAT TCGGCTGC TCGGCTGC CGAAAC CGCCTTTC GAACCCCC GGTAACCCCC GGTAACGCCCC GGTAACGCCCC GGTAACGCCCC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCCC ATTAGCAGGAC TCATTAGCAGAGAC TCATTAGCAGAGAC TCATTAGCAGAGAC TCATTAGCAGAGAC TGAAGCCAGT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG ATCAGCTCAC ATAAAGGCCG ATCAGCTCAC AAAAAGGCCG GACGGTGGC GACGGTGGC GACGGTTGC TACCTCGGATAGT	GGTTCGACTA R F D Y AACCCGCTGA GGTTCAATGGG GGTTCAAGGG GCGTAATCAT CCTAAAGAGT CCCAGGCGGAGA TCAAAGCCG CCTTCCTATCCTG CCTTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGACTTG	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGTCATAGCT GAGCTAACTC TAATACGGTT GTTTTTCCAT CCCCCTGGAA GCTCACGCTT TAACTATCGT TAACTATCGT ACAGAGTTCT GTAGCTCTTG	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGGTTCCGCC GGTCCCTCGT TAGGTATCTC CTGAGTCCA ATGAGTGGTG ATCCGCAAA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCT TCAGGGGATA CCCCTGACGA GCGCTCTCTC AGTTCGGTGT ACCGGGTAAC CCAAACCACC	GCTGATCTCC L I S ACATGATAAA AACCATTATA AACCATCTACA ATCCGCTCAC ACTGCCTCAC ACTGCCTCAC ACGCAGGAAA GCATCACAAA GCTTCCGACC AGGTCGTTCC ACACGACTTCC ACACGACTTAC ACACGACTTAC ACACGACTTAC ACACGACTTAC ACACGACTTAC ACACGACTTAC ACACGACTTAC
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC F R T ( TTTGGAC AAA AAGTTAA CAA TGAATTAT GAA AACTGAC CCC TCTGGGC GTC AAAGGCC AGC TTCAGAC GTC ATAACCTG TCC TCTGGTC GTC AAAGGCC AGC TCCAGAC GTC TCTGGTAC TCC TGTGTGC ACC TGGTATAC TGC TGGTATAC TGC TGGTATAC TGC TGGTATAC TGC TGGTATAC TGC TGGTATAC TGC TGGTAAAA AGC TGCAAAG AGC TGCAAAG AGC TGCAAAG AGC TGCAAAG AGC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACACAAT TCAGTCGA GGAAGCAT GCCAGCTG CCAAAAGGC GGCGAAAC GGAACCCC GGAACCCC GGAACCCC GGAACCCC GGAACCCC GCCTTCC GCACTTC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA CATTAATGAA GGCGAGGGGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAG TGAAGCAGG GTGAGCAGG	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCCAGGCTG TCGCCACGTCA AAAAGGCTCA AAAAAGGCCG TATAAAAAT AAGCGTGGCG GACCGCTGCG CGAGGTATTTCTCGA AAAAGGATATGT TACCTTCCGA AAAAGGATATGT	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CGTTGCTGGC CCAGGCGTTGCTTCATCATA CCTTATCCATA CCTTATCCAG AGGCGGTGCT AAAAGAGATCA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGTCATAGCT GAGCTAACTC GGCGGTTTGC TAATACGGTT GCTCCTGGAA GCTCACGCTG TAACTATCGT ACAGAGTTCT GTAGCTCTTG TTTGATCTTTT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCTGGCG ACATTAATTG GTATTGGCCG GCTCCCTCGT TAGGTATCCC CTTGAGTACCA TAGGTATCC TGAAGTGGTG ATCCCGCGCAA TCTACGGCAA TCTACGGGTA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCGCTCTCCT AGTTCGGGTGA ACCCGGTAAG GCCTAACTAC CAAACCACCG CTGACGCCC CTGACGCCC CTGACGCCC CTGACGCCC CTGACGCTCC CTGCTCC CTGCTCC CTCC	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCTCTACA ACTGCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GCTCCCGACCC ACGCGCTTCGACCC ACGCACTACACAAA GCTCCAACCACACACGCTTCCGACCC ACGCACTTCCGACCC ACACGACTTA GCTACACTA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGCATAC CGC F R T ( FTTGGAC AAA GATTAT GA: ATACGAG CCC ACCTGTC GTO CTCGGTC GTO AAAAGGCA GTO ATACCAG GTO ATACCAG TCC AGCACA TCC CAGCAC TGC AGGCAC TGC AGGCAC TGC AGGAGAC AGG AGGAGAC AGG AGGAGAC AGG AGGAGAC AGGAC AGGAC AGGAGAC AGGAGAC AGGAC AGG	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACACA TCAGTGA GGCAGCTG CCAAAAGC CGCAAAAC CGCCTTTC GGACCCCC GGACCCCC GGACCCCC GGAACCCC CGCTAACAGG CGCATTAC CCAGATTAC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GCGAACCGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAGA TTAGCAGAAAAA TTATCAAAAA	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG ATCAGCTAAC ATCAGCTCAAC ATAAAAGCCC ATAAAAGCCC GAACGTTATAAACATA AAGCGTGGCC GCAGGTATGT TACCTTCGGA AAAAGACATTCAC GGATCTTCAC	GGTTCGACTA R F D Y AACCCGCTGA GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CCTTGCTGCG CCAGGCGTTT CTTTCTCATA CCTTATCCGG AGGCGTGCT AAAAGAGTTG AAGAAGATTC CTAGATCCTT	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG  GGCGATTACCT  GAGCTTTTCCAT  CCCCCTGGAA  GCTCACGCT  TAACTATCGT  TAACTATCTT  TTAAATTTAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCCTCGT TAGGTATCTC CTTGAGTCCTC TGAAGTGGTG ATCGGCAAA TCTACGGCAAA TCTACGGCAAA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA GCGCTCTTCCAGTTAGGTTC ACTTCGGTTA ACCCGGTAAG GCCTAACTAC CAAACCACCG CTGACGACCAC TAAATCAATCA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTCCCGCT TCCTCCGCTCA ACGCAGGAAA GCATCACAAA GCATCACAAA ACACGACTTACA ACACGACTTA GCCTACACTA GCGTACACTA GCGTACACTA GTGGAACGA TAAAGTTATAT
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGGATAC CGC F R T ( FITTGGAC AAA FGATTAT GA: AAAGTTAA CAA FGATTAT GA: ATACCAG CGC TCTCGGTC GTC AAAGGCC AGG TCTACAGC GTC ATACCTG TCC ACGCCAC TGC TGCGAAC TGC TGCGAACA AG: TTCTGAC AG:	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACAAC TCAGTCGA GGCAAGCAT TCGGCTGC TCGGCTTC GGAACC CGCTTTC GGAACCCC GGAACC CGCATTC GAACCCCC GTACCAG CCGCTTCC CAGAACCCC CGCATTC TCATGAGA TCATGAGA TCATGAGA TTACCAAT	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCCGACAGGAC TCCCTTCGGG CGTTCAGCCC ATTAGCAGAG TGAAGCCAGT TGAGCCAGT TGAGCCAGT TGAGCCAGT TGAGCCAGT TGAGCCAGT TGAGCCAGT GGCCAGAAAA TTATCAAAAA GCTTAATCAG	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCAC AAAAAGGCCG ATCAGCTCAC AAAAAGGCG GACGGTGGC GACGGTGGC GACGGTTGC AAAAGATTC TACCTTCGGA AAAGGATCTC GGACGACCT	GGTTCGACTA R F D Y AACCCGCTGA GGTTCATTTG GGTTCATGGG GCGTAATCAT CCTAATGAGT CCCAGGCGGAGA TCAAAGGCG CGTTGCTGGC CCAGGCGTT CTTTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGACTTC AAAAGATCC CTAGATCCT ATCTCACACA	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGCGTTTGC TAATACGGTT GTTTTTCCAT CCCCCTGGAA GCTCACGCT TAACTATCGT TAACTATCGT TAACTATCT TAACTATCT TTGAACTCTT TTGAACTCTT TTTGAATTTAA TCTGTCTATT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGCTCCCCC GTTAGGTACTCCC TGAGGTACTCC AGGAGTACCA ATCAACGGGT AATCAACGGGT AATCAACGTAT TCCGTCAATC	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA ACCCCGGTTAAG GCCTAACTAC CAAACCAACCC CTGACGCTCA TAAATCAATC ATAGTTCGCT TAAATTAGT	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTATA AACCATTACA ACTGCCCGCT TCCTCGGTCAC ACTGCCGCAC ACTGCACAC ACTGCACAC ACTGCACC AGGTCGTTCC ACACGACTTCC ACACGACTTCC ACACGACTTCC ACACGACTTAC CTGGTACCGG GTGGAACGAA TAAAGTATAT TAAAGTATAT
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC F R T C FTTGGAC AAA AGTTAA CAAI TGATTAT GA: ACCTGTC GTO AAAGGCC AGC TCAGAGG GTO ATTACCAG TCC TCGGTC TCC TCGGTC TCC TCGGTC TCC TCGGTAC TCC TCGGTAC TCC TCGGTAC TCC TCGGTAC TCC TCGGTACT GGTATC TCGGTATC TCC TCGGTACT GGTATC TCGGTACT GGTACC TCGGTACT GGTACC TCGGTACC TCGGTACC TCGGTACC TCGGTACC TCGCACAC TCCCACAC T	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACACAAT TCAGTCGA GGCAGCTG CGCCGCTTCC GGAACC GAACC GGAACC GGACC GGAACC GGAA	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGA GCGCAGAAAA TTATCAAAAA GCCTAAAAAAA ACCATCTGGC ACCATCTGGC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAAT TTATGTTTCA TCGCCAACGTG ATCAGCTAC AAAAGGCCA AAAAGGCCG TATAAACATA AAGCGTGGCG GACCGCTGCG CGAGGTATGT TACCTTCGGA AAAAGGTCTC GGATCTTCCG GGATCTCCC CCCAGTGCTCC CCCAGTGCTCC	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CGTTGCTGGC CCAGGCGTTTTTTCCATA CCTTATCCAG AGGCGGTGCT AAAAGAGTTG AAAAGAGTTC CTAGATCCTT ATCTCAGCGA ATCTACAGCGA CAATGATACC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGTCATAACT GAGCTAACT GGCGTTTTCCAT CCCCCTGGAA GCTCACGTG TAACTATCGT ACAGAGTTCT ACAGAGTTCT TTAAATTAAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG GATGCTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCCACGAAAA AGGCTCCGCC GCTCCCTCGT TAGGTATTCT CTTGAGTATCT CTTGAGTATCT TGAGTATCT TGAGTATCT TGAGTATCT TGAGTATCT TGAGTATCT TGAGTATCT ATCCGCAAA TCTACGGGGT AATGAAGTTT TCGTTCATCC CGCTCACCGC	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCGCTCTCCT ACTTCACTACTAC CAAACCACC CTGACGCTCA TAAATCAATC TAAATAATC AATACTACC CTCACGATTT	GCTGATCTCC L I S ACATGATAGA AACCATTATA AACCTCTACA ACTGCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GCATCACAAA GCATCACACA ACGACGTTCGACCC ACGCGCTTCGACCC ACGCGCTTCGACCC ACGCGCTTACGACCA ACGCGCTTACGCGACCACTA GCTGCTACCGACCA TGGATACCGAC TGAAGTATAT GACTCCCCGT ATCACCAATA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGGATAC CGC F R T ( FITTGGAC AAA RGATTAT GA: ATACAGA CCC TCTCGGTC GTC TCTCGGTC GTC TCAGAGG CCC TCTCGGTC GTC TCAGAGG CCC TCGTGTC TCC TCGTGTC TCC TCGTGTC TCC TCGTGTC TCC TCGTGTC TCC TCGTGTTC TCC TCGTGTTC TCC TCGTGTTC ACC TGGTATC TCC TGGTATC TCC TGCAAGC AC TGCAAGC TGCAAGC TGCAAGC TGCAAC TGCAAGC TGCAACC TGCAAGC TGCAAGC TGCAAGC TGCAAGC TGCAAGC TGCAAGC TGCAAGC TGCAACC TGCAAGC TGCAAGC TGCAACC TGCAA	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACAACAAT TCAGTCGA GGAACCAT TCGGCTGC CCAAAAGCC GGCGAAAC GGCCTTTC GGAACCCCC GGAACCCCC GTAACAGG CGCACTCTC CAGAACAGG CGCACTCTC CAGAACAC CGCCACAT CAGACCAC CACACCAC CAGACCAC CAGACCAC CAGACCAC CAGACCAC CAGACCAC CAGACCAC CACACCAC CAGACCAC CACACCAC CACACCAC CACACCAC CACACCAC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GCGAGAGCGGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAGA TTATCAAAAA GCTTAATCAG GCGCAGAAAA TTATCAAAAA GCTTAATCAG ACCATCTGGC AGGGAGACCT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTACA AAAAGCCCG ATCAGCTGCG GACCGTGCG GACCGTTCG GACCTTCGGAACG AAAAGGCCTGCG GACCTTCGC GACCTTCGC CAACTTCAC TGAGGCACCT CCCAGTTCTC	GGTTCGACTA R F D Y AACCCGCTGA ACCCGCTGA GGTTCAGGGG GCGTAATCAT CCTAATGAGT CCGCGGGGAGA TCAAAGGCGG CCTGCGCGCGAGC CCTACCTCTCCTCCCAGCCGTTT CTTTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGAGTTC AAGAAAGACTC CTAGATCCTT ATCTCAGCGA CAATGATACC	CAAGGATGAC  K D D  TCAGCCTCGA  GAGATTTGT GAGATTGGT GAGCTGAGCT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCCTCGT TAGGTATCTC CTTGAGTCCA TGAAGTGGTG ATCTAGGGGT ATCTAGGGGT ATCTAGGGGT ATCTACGGCAAA TCTACGGCAAA TCTACGGCAAA TCTACGGCAAA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGGCCTC CTCTTCCGCT TCAGGGGATA CCCCTGACGA GCGCTCTTCA AGTTCGGTGT ACCCGGTAAG GCCTAACTAC CAAACCACCG CTGACGGAC CTGACGACACCA CTGACGACACCA CTGACGACACCA CTGACGACACCA CTGACGCTCA ATAGTTGCCT CTCCAGAGTTT GGAAGCTAGA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATCTACA ATCCGCTCAC ACTGCCGCT TCCTCGGTCA ACGCAGGAAA GCATCACAAA GCTTCACAACA ACGCACGACCA AGGTCGTCA ACGCACCA ACGCACTTA ACACGACTTA ACACGACTTA ACACGACTTA ACACGACTA TAAAGTATAT GACTCCCCGT ATCACGACATA ATAAAGTATATG ATCAGCAATA GTAAGTAGTT
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGGATAC CGC F R T ( FTTGGAC AAI AAGTTAA CAA FGATTAT GA: ATACCAG CGC TCTCGGTC GTC AAAGGCC AG TCTAGGCC AG TCTAGGTC TCC ACGCAC TCC AGGATAC GGC AGGATTTT GG: GGATTTT GG: GTCTGAA AG AGGATAC AG AGGATAC GGC AGGATTTTGCGC AAC AGGATAC AGG AGGATTTTGCGC AAC AGGATAC GGC AGGATTTTGCGC AAC AGGATAC GGC AGGATAC GGC AGGATTTTGCGC AAC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACAAC TCAGTCGA GGCAAGCAT TCAGTCGA GGCAACC CGCTTTC GGAACC CGCTTTC GAACCCC GGAACC CGCTTTC CAGATTAC TCATGAGA TTACCAAT GAGGGCTT CAGAGTT CAGATTAC TCATGAGA TTACCAAT GAGGGCTT AGCGCATT CAGCGCTTT CAGCGCTT CAGCGCTT CAGCGCTT CAGCGCT CAGCTT CAGCGCT CAGCT CAGCT CAGCT CAGCT CAGCT CAGCT CAGCC CAGCT CAGCT CAGCC CAGC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCCTCCGGC CGTTCAGCC ATTAGCAGAC TCATTAGAA GCGAGAACAGAC TCATTAGAAAAA TTATCAAAAA GCTTAATCAAAAA GCTTAATCAAAAA GCTTAATCAG ACCATTGGC ACGTTCAGCC CCTTCGCC CCTTCGCC CCTTCAGCC CCTTCAGCC CCTTCAGCC CCTTCAGCC CCTTCAGCC CCATTGCTAC CCATTGCTAC CCATTGCTAC CCATTGCTAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCA AAAAAGCCCG ATCAGCTCAC AAAAAGGCCG TATAAAGTATA AACGTTGGC GACGGTGCG GACGGTGCG GACGGTTCC GCAGGTATCT CCCAGTGCTC CCCAGTGCTC CCAGTGCTC CCAGTGCTC CCAGTGCTC CCAGTGCTC CCAGTGCTC CCAGTGCTC AACTTTTATC AGGCATCGTG	GGTTCGACTA R F D Y AACCCGCTGA GGTTTATTTG GGTTCAGGG GCGTAATCAT CCTAATGAGT CCGCGGGGAGA TCAAAGGCG CGTTGCTGGC CCAGGCGTT CTTTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGATCC CTAGATCCT ATCTCAGCA CATGATACCG CAGGCGTCCT CTAGATCCC CTCCATC CGCCTCCATC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGTCATAGCT GAGCTAACTC TAATACGGTT GTTTTTCCAT CCCCCTGGAA GCTCACGCT TAACTATCGT TAACTATCT TAACTATCT TTGAACTCTT TTGAACTTTT TTTGAATTTAT TCGGAGCCCA CCGCAGCCCA CCGCCTTTGC	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGCTCCCCC GTTAGGTACTCCC TGAGGTACTCC AGGGTATCTC ATCCAGGGGT AATCAAGGGGT AATCAAGGGGT AATCAAGTTT CCGTCAACCG ATTGTCCC TATGGCTTA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA GCGCTCCTC AGTTCGGTT ACCGGGTAAC CCAAACCAC CCAAACCAC CTAACTAC CAAACCAC CTAACTAC CAAACCAC CTGACGTTAG TAAATCAATC ATAATTCACT CTCAGATTT GGAAGCTAGA TTCAGCTTCCAG	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATCTACA ATCCGCTCAC ACTGCCCGCT ACCGCAGAAA GCATCACAAA GCATCACAAA GCATCACACA ACGCAGGAAA GCATCACACA ACGCAGGACTA CGCAGACCT ACAGGACTTA CTGGTAGCGG GTGGAACGAA TAAAGTATAT GACTCCCCGT ATCAGCAATA GTTAGCAATA GTTAGCAATA GTTAGCAATA GTTACCAACG
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CGCGTAC CGC F R T C FTTGGAC AAA AGTTAA CAA TGATTAT GA: ATACAGA CCC ACCAGAG GTC ATACAGAG GTC AGGATAC GGG AGGATAC GGG AGGATAC GGG AGAGAGGC CGG ATTAGAGGC AGG ACGATAC GGC ACAGAGAG CGC ATTAGAGAG CGC ACAGAGAG CGC ATTAGAGAG CGC ACAGATAC GGC ACAGAGAG CCC ATTAGAGAG CCC ATTAGAGAG CCC ATTAGAGAG CACAGAGAA CCATGAT CCC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACACAT TCAGTCGA GGCAGCTG CGCCTTCC GGCGCTTC GGAACCCC GGAACCCC GGAACCCC GGAACCCC GGAACCCC GTAACAGG CCGCTTCC CGAGATTAC CGCTTGC CAGATTAC CAGACCCC CGCACTTCC CAGATTAC CAGACCCC CCAGATTAC CCCCCC CCCCCC CCCCCCC CCCCCCC CCCCCCC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAG TGAAGCCAGT GCGCAGAAAA TTATCAAAAA GCCTAATCAAAAA ACCATCTGG ACTGGTCCAGCAGACCAGC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TCGCCAGCTTG GCCTGGGTG TCGGCCAACG ATAAGATA AAGCCTCG ATAAGATA AAGCGTGGC GACCGTGCG CGAGTATGT TACCTTCGGA AAAGGATCTC GGATCTTCAC CCCAGTGCT CCAGTGCT CAACTTTATC AAGCATCTG GAGCCACT	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG CCTTGCTGGC CCAGGGGTTTCTCATA CCTTATCCGG AGGCGGTT AAAAGAGTTG AAAAGATCAT AAAAGATCAT ATCTCAGCGA CCTAGTCCTCATA CCTCATCCGCCTCCATC CGCCTCCATC CGCCTCCATC CCTTCCGCTCC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGGGG GGTCATAACT GAGCTAACTC GAGCTAACTC GTCATCATC GTCATCATC GTCATCATC TTAAATACGT TAACTATCGT ACAGAGTTCT TTAAATTAAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG GATGCTATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT TAGGTATCTC CTTGAGTACTC TGAAGTGGTG ATCCACGAAA TCTACGGGGT AATGAAGTTAT CCGTCAACCA ATCTTCATCC CGCTCACCGG ATTGTTGCCTCA TCAGGTCTCA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AGGCAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCCTTAACTAC CAAACCACG CTGACGCTCA TAAATCAATC ATACTACT TCAGGGGTTA CTGACGCTCA TAAATCATC CATACTACT CAAACTACT CAAACTACAC AGTTGCCCG AGTTGCCCG	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATCTACA ATCCGCTCACA ACTGCCGCGT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GCATCACAAA GCTTCCGACCC ACGCGCTTTCG ACACGACTTA GGCTACACTA ACGCAGGATA GAGTACGTACGACTA GGCACCTA ACGCAGACTA GACTACCTA GACTACCTA GACTACCTA GACTACCTA GACTACCACACA ACGCACTACACACACACACACACACACACACACACACACA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGGATAC CGC F R T ( FTTGGAC AAA AGTTAA CAA FGATTAT GAT ATACGAG CCC TCTCGGTC GTT AAAGGCC AG TCTTGGTC AC CAGCAC TGC AGACACT TGC AGACACT TGC AGACACT TGC AGACACAC AGACACAC AGACACAC AGACACAC AGACACAC CAGCACC CAGCACC CGC CAGCACC CGC CG	GGCGGCCG G G R GGCACACAC G H H ACCACACAC ACACACAC GGAAGCAT TCAGTCGA GGCAGCTG TCGGCTGC CAAAAGCC GGCCATTC GGAACCCC GGAACCCCC GTACAGG CCAGATTAC TCATGAGA TCATGAGA TCATGAGA CGCCATTC CAGACCCCC CAGATTAC CAGACTTC CAGACCCCC CAGATTAC CAGACTTC CAGACCCCC CAGATTAC CACATTAC CAC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GCGAGAGCGT CAGGAACCGT CCGACAGGAC TCCCTTCGGG CGTCCAGCAGAAA TTATCAAAAA GCTAAAAAA GCTTAATCAG ACCATTGCTC CCATTGCTC CCTTCGCC ATTATCAAAAA CTTACTTCAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTAC AAAAAGCCCG GACCGTGCG GACCGTGCG GACCTTCGGAATT TACTTCAGA AAAGGATCT CGAGGTATGT TACCTTCAGA AAAGGATCTC CCAGTTCAC TGAGCACCT CCAGTTCTAC CAACTTTATC AGGCATCGTG CAGCTTAGCT CAGCTTAGCT CGCGTTAGCT TGCCGTTACCT	GGTTCGACTA R F D Y AACCCGCTGA GGTTCATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CCGCGGGGAGA TCAAAGGCGG CGTTGCTGGC CCAGGCGTTT CTTCTCATA CCTAAAGAGTTC AAAAGAGTTC AAAAGACTTC AACAGATCCTT ATCTCAGCGA CAATGATACC CAGCGCTCCATC GTGTCACGC	CAAGGATGAC  K D D  TCAGCCTCGA  GAGATTTGT  GAGATTAGT  GAGGTGGG  GGCGTTTGC  TAATACGGTT  GTTTTTCCAT  CCCCCTGGAA  GCTCACGCTG  TAACTACGT  TAACTATCGT  TAACTATCGT  TAACTATCGT  TTTGATCTT  TTTAAATTAA  TCTGTCTATT  GCGAGGACCA  CAGTCTATTA  CGTCGTTTGG  TCCGATTGG  TCCGATCGTT  TCTGTGACTCTT  TCTGTGACTCTT	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC CTGAGTCCCCC CTGAGTCACCAGA ATCAACGGGT ATCACGGCAAA TCTACGGGCAAA TCTACCGGCAAA TCTACCGGCACACACACACACACACACACACACACACACA	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA GCGCTCTTCAGTACCACCG CCCTAACTACCACCG ACCCAGTAAG ACCACCG CTAACGACCACCC CTGACGTAAC ACCAGACTCA ACTACCACCACCACCACCACCACCACCACCACCTCACACTACCACC	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTGCCGCT TCCTCGGTCA ACGCAGGAAA GCATCACAAA GCTTACCA ACGCAGGACTA ACGCATCTCG ACACGACTTA CTGGTAGCGG GTGGAACGAA TAAAGTATAT GACTCCCCGT ATCAGCAATA GATCACCAACA ATCAGCAATA GATCTCCCAACG AGTAGTAGTT GTTCCCAACG AGTGTATATCA TGTTATCA
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGCATCA Q H S CAGCATCA F R T ( FTTGGAC AAI AAAGTTAA CAI FGATTAT GA: ATACCAGC GTO ATACCAGCACA ATACCTG TCC AAAGGCC AG ATACCTG TCC AGCACAC TCC GGATTT GG: GGATTT GG: GTCTGACAC AGCACACAC AGCACAC ACACACAC	GGCGGCCG G G R GGCACACAC G H H ACCACACAC ACACACAC GGCAGCTG TCGGCTGC CCACACAC CGCCTTTC GGAACCCC GGAACCCC CGCACTCC CCACACAC TCACCCC CCACACAC CCCCTTCC CCACACAC CCCCTTCC CCCCTTCCC CCCCTTCC CCCCTTCCC CCCCTTCCC CCCCTTCCC CCCCTTCCC CCCCTTCC CCCCTTCCC CCCCTTCCC CCCCTTCC CCCCTTCCC CCCCTTCCC CCCCTTCCC CCCCCC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAAACGGT CCGCACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAG TGAAGCCAGT TGAAGCCAGT GCGCAGAAAAA GCTTAATCAG ACCATCTGCC CGGTCGCGCGCGTCACCC CGGCGTCAAT	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCAC AAAAAGCCCG ATCAGCTCAC AAAAGCCCG GACGGTGCG GACGGTGCG GACGTTCAC TGAGGACTTCAC TGAGGCACCT CCCAGTGCTG CCCAGTGCTG CCCAGTTTATC AGGCATCGTG GCGGTTAGCT TGCCATCCGT ACGGGATAGT	GGTTCGACTA R F D Y AACCCGCTGA GGTTCATTTG GGTTCAGGG GCGTAATCAT CCTAAATGAGT CCGCGGGGAGA TCAAAGGCGG TCTTCTCATA CCTTATCCG AGGCGGTGCT AAAAGAGTTC AAAAGAGTTC ATCTCAGCG CAAGCCGTCCTAGATCCTT ATCTCAGCG CAAGAGTCCT ATCTCAGCG CAAGAGTCCT ATCTCAGCG AAAGATCCT ATCTCAGCCAC CGCCTCCATC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGGG GGTCATAGCT GAGCTAACTC TAATACGGTT GTTTTTCCAT TAATATACGTT GAGCTAACTC GAGCTAACTC GTAGCTCTG TAACTATCGT TTAACTATCT TTTGAATTAAT TCTGTCTATT GCGAGACCCA CAGTCTATTA CGTCGTTTGG TCCGATGGAT TCCGTCGATGACTC TCCGTTTGG TCCGATGACTC ATAGCACACA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGGTTCCCCC GTTAGGTCA ATCCACAGAA ATCCACGGGT ATCCAGTCA ATCCACGGCT ATCGTCACCG ATCCACGG ATCCACGG ATCTCACCG GTTCACCG GTTCACCG GTTCACCG TATGGTTCAC GGTTCACCG TATGGCTCACGG TATGGTTCACGG TATGGCTTCA GTCAGAAGTA GTCAGAAGTA TCTTCAAAGTA TTTAAAAGTG TTTAAAAGTG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT ANGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA ACCCAGCTGACGA ACCCAGCTCA AGTTCGGTGT ACCCGGTAG CCCTAACTAC CAAACCACCG CTGACGACTCA ATAGTTGCCTC CTCACGATTTG TCAGGTTAG TTCAGCTCCA AGTTTGGCCCC AGTTTGGCCGC AGTTTGGCCCC AGTTTGGCCCC ACTTGACCAAGTCA ACCAAGTCA CTCATCATTTG	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTGCCCGCT TCCTCGCTCAC ACTGCACACA GCATCACAAA GCATCACAAA GCATCACAAA GCATCACAAC ACTGCACCA ACGCACTAC ACACGACTTA CTGGTAGCCG GTGGAACGAA TAAAGGTATAT GACTCCCCGT ATCAGCAATA GTAGCGATTATC GTTCCCAACC AGTGTTATCA ACTGCTTACTAC ACTGCTACAC ACTGCTTACTAC ACTGCTTACTAC ACTGCTACAC ACTGCTTACTAC ACTGCTACTAC ACTGCTACAC ACTGCTTACTAC ACTGCTACAC ACTGCTACAC ACTGCTACAC ACTGCTTACTAC ACTGCTACAC ACTGCTAC
2001 TTCTGCAGAT ATCC	CAGCACA GTC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACAAC TCAGTCGA GGCAGCTG CGCATGC CGCATTC GGAACCCC GGAAAC CGGCATTC GGAACCCC GTAACAGG CGCTTTC CGAGATTAC TCAGTCGC CGCATTAC TCAGTCGC CGCTTTC CGAGATTAC TCACAGA TTACCAAT CGAGGCTAC CGCCTTTC CAGAGTTAC TCATCAAT CGAGGCTT CAGAGTTAC TCATCATC CCCATGTT CATCATTC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAG TGAAGCCAGT GCGCAGAAAA TTATCAAAAA GCTTAATCAA ACCATCTGGC ACTGGCAGACCAGT CCCATTGGC CCATTGGCAGACCAGT CCCTTCGGG CCCATTGCAGACCAGT CCCATTGCAGACCAGT CCCATTGCTACACAGACCAGT CCATTGCTACACAGACCAGTCTACACGCGCTCATAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCCAGCTTG GCTGGGGTG TCGGCCAACG ATAAGATA AAGCGTCAC AAAAAGCCG TATAAAGATA AAGCGTGCG GACCGCTGCG GACCGTGCG CGAGTATGT TACCTTCGAA AAAGGATCTC CCAGTGCTC CCAGTGCT CAACTTTAC TCCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAA TCCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAACTTCAA TCCAA	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG TCTTCCTCATA CCTTATCCGG AGGCGGTT CTTTCTCATA CCTTATCCGG AGGCGTGCT AAAAGATCCT CTAGATCCT AAGAAGATCC CTAGATCCT GTGCTCCATC GCCTCCATC GCCTCCATC CCTTCGGTCC AAGATGCTTT AACGCGCCAC TGTAACCCAC TGTAACCCAC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGGG GGTCATAACT GAGCTAACTC GAGCTAACTC GTTTTTCCAT CCCCCTGGAA GCTCACGCTG TAACTATCGT ACAGAGTTCT TTAAATTAAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG GATGCTATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT TAGGTATCCC TGAAGTGCTC TGAAGTGCTC TGAAGTGCTC ATCCGCCAACA ATCCTACCGGAA TCCTACCGGAA TCGTTCATCC CGCTCACCGG ATTGTTGCCT GTCAGCAGTA TCTACGGAGTA TCTACGGAGTA TCTACGGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TTTAAAAGTC TTTAAAAGTC AACTGATCTT	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AGGCAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCGCTCTCCT ACTTCACTAC CAAACCACC CTGACGCTCA TAAATCAATC CATCAGAGCTCA TAAATCAATC CTCCAGATTT GGAAGCTGAC ATCAGCCC AGTTGCCCC AGTTGCCCC AGTTGCCCC AGTTGCCCC ACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCAC CACCAAGTCAC CACCACACTCA CACCACTCTC CAGCATCTTC CAGCATCTTT CAGCATCTT CAGCATCTTT CAGCATCTTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCT CAGCATC CAGCATCT CAGCATCT CAGCATCT CAGCATCT CAGCATCT CAGCATC CAGCATCT CAGCATC CAGCATCT CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAG	GCTGATCTCC L I S ACATGATAGA AACCATTATA AACCATCTACA ATCCGCTCACA ACTGCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GCTTCCACCA ACGCAGCATTA GGCTACACTA ACGCAGGATA GGTCCTCGACCC ACGCAGCTTA GGCTACCTTA GGCTACCTTA GGCTACCTTA GGTACCTTA GCTACCCGT ATCAGCAATA GTACGCAGA AGTGTTATCA TTCTGAGAAT TTCTGAGAAT GAAAACGTTC TACTTTCACC
2001 TTCTGCAGAT ATCC	CAGCACA GTC	GGCGGCCG G G R GGTCATCAT G H H ACCACAAC ACAAC TCAGTCGA GGCAGCTG CGCATGC CGCATTC GGAACCCC GGAAAC CGGCATTC GGAACCCC GTAACAGG CGCTTTC CGAGATTAC TCAGTCGC CGCATTAC TCAGTCGC CGCTTTC CGAGATTAC TCACAGA TTACCAAT CGAGGCTAC CGCCTTTC CAGAGTTAC TCATCAAT CGAGGCTT CAGAGTTAC TCATCATC CCCATGTT CATCATTC	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAACCGT CCCGACAGGAC TCCCTTCGGG CGTTCAGCC ATTAGCAGAG TGAAGCCAGT GCGCAGAAAA TTATCAAAAA GCTTAATCAA ACCATCTGGC ACTGGCAGACCAGT CCCATTGGC CCATTGGCAGACCAGT CCCTTCGGG CCCATTGCAGACCAGT CCCATTGCAGACCAGT CCCATTGCTACACAGACCAGT CCATTGCTACACAGACCAGTCTACACGCGCTCATAC	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCCAGCTTG GCTGGGGTG TCGGCCAACG ATAAGATA AAGCGTCAC AAAAAGCCG TATAAAGATA AAGCGTGCG GACCGCTGCG GACCGTGCG CGAGTATGT TACCTTCGAA AAAGGATCTC CCAGTGCTC CCAGTGCT CAACTTTAC TCCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAA TCCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAACTTCAA TCCAACTTCAACTTCAA TCCAA	GGTTCGACTA R F D Y AACCCGCTGA  GCTTTATTTG GGTTCAGGGG GCGTAATCAT CCTAATGAGT CGCGGGGAGA TCAAAGGCGG TCTTCCTCATA CCTTATCCGG AGGCGGTT CTTTCTCATA CCTTATCCGG AGGCGTGCT AAAAGATCCT CTAGATCCT AAGAAGATCC CTAGATCCT GTGCTCCATC GCCTCCATC GCCTCCATC CCTTCGGTCC AAGATGCTTT AACGCGCCAC TGTAACCCAC TGTAACCCAC	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT GAGGTGGG GGTCATAACT GAGCTAACTC GAGCTAACTC GTTTTTCCAT CCCCCTGGAA GCTCACGCTG TAACTATCGT ACAGAGTTCT TTAAATTAAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG GATGCTATTG GTATTGGGCG ATCCACAGAA AGGCTCCGCC GCTCCCTCGT TAGGTATCCC TGAAGTGCTC TGAAGTGCTC TGAAGTGCTC ATCCGCCAACA ATCCTACCGGAA TCCTACCGGAA TCGTTCATCC CGCTCACCGG ATTGTTGCCT GTCAGCAGTA TCTACGGAGTA TCTACGGAGTA TCTACGGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TCTACAGAGTA TTTAAAAGTC TTTAAAAGTC AACTGATCTT	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AGGCAGTAA TGAAATTGTT CGTTGCGCTC CTCTTCCGCTT TCAGGGGATA CCCCTGACGA GCGCTCTCCT ACTTCACTAC CAAACCACC CTGACGCTCA TAAATCAATC CATCAGAGCTCA TAAATCAATC CTCCAGATTT GGAAGCTGAC ATCAGCCC AGTTGCCCC AGTTGCCCC AGTTGCCCC AGTTGCCCC ACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCA CACCAAGTCAC CACCAAGTCAC CACCACACTCA CACCACTCTC CAGCATCTTC CAGCATCTTT CAGCATCTT CAGCATCTTT CAGCATCTTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCTT CAGCATCT CAGCATC CAGCATCT CAGCATCT CAGCATCT CAGCATCT CAGCATCT CAGCATC CAGCATCT CAGCATC CAGCATCT CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAGCATC CAG	GCTGATCTCC L I S ACATGATAGA AACCATTATA AACCATCTACA ATCCGCTCACA ACTGCCGCT TCCTCGCTCA ACGCAGGAAA GCATCACAAA GCTTCCACCA ACGCAGCATTA GGCTACACTA ACGCAGGATA GGTCCTCGACCC ACGCAGCTTA GGCTACCTTA GGCTACCTTA GGCTACCTTA GGTACCTTA GCTACCCGT ATCAGCAATA GTACGCAGA AGTGTTATCA TTCTGAGAAT TTCTGAGAAT GAAAACGTTC TACTTTCACC
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGCATCA Q H S CAGCATCA CAGCATCA F R T ( FTTGGAC AAI AAGTTAA CAI FGATTAT GA: ATACCAGC CTCCGGTC GTC ATACCAGCACA ATACCAGCACA TCAGCACAC TCGGTATC TCG TCGTATC TCG TCGTATC TCG TCGTATC TCG TCGCAACA ACA TCGACACAC TCGCACACACACACACACACACACACACACACACACACAC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACACA GGAAGCAT TCAGTCGA GGCAGCTG TCGGCTGC CCACAAAC CGCCTTTC GGAACCCC GGAACCCC GGAACCCC CAGAAT TCAGTCGA CGCTTTC GAACCCCC GTACCAG CGCTTTC CAGACCCC CAGATTAC TCATCACAT GAGGGCTT TCATCACT TCATCACT TCATCACT TCCTTTCC CAGAACC CCCATGTT CCTCTTTCC CAGAACG	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAAACGGT CCGCTCGGG CGTTCAGCC TCCTTCGGG CGTTCAGCC TTATCAGCA TTATCAGAAA GCCAGTACTACACA GCCATTCGTC CCTTCGGC CGTTCAACC CCGCTCAAC CCCTTCGCC CCGTTCGCC CCATTGCTAC CCGCTCCAC CCGCTCCAAC CCAACATCCCC CCACTCCAAC CCACTCCAAC CCACTCCAAC CCACTCCAAC CCAACATCCCC CCACTCCAAC CCACTCCAAC CCAACATCCCC CCAACACAC CCAACATCCCC CCAACACAC CCAACACC CCAACACC CCAACACC CCAACACC CCACACC CCAACACC CCAACACC CCAACACC CCAACACC CCAACACC CCACACC CCAACACC CCACACC C	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCA AAAAGGCCC ATCAGCTCAC AAAAGGCCC GACCGCTGCG GACGGTGCG GACGTATGT TACCTTCGGA AAAGATTC CCAGTGCTG CGAGTTATC CCAGTGCTG CGAGTTATC CCAGTGCTG CGAGTTAGCT CCCAGTGCTG CGAGTTAGCT TCCAGTTCTC CCAGTGCTG CGGGTTAGCT TGCCATCCGT ACGGGATATT TCCAGTTCGA CAAAAAAAAAGGC AAAAAAAAAGG	GGTTCGACTA R F D Y AACCCGCTGA GGTTCATGG GGTTCAGGG GCGTAATCAT CCTAAATGAGT CCTAAAAGGCGG CGTTGCTGGC CCAGGCGTTCTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGAGTTC CTTACTATCCGG CAAGAGATCCT ATCTCACGG CAAGAGATCCT ATCTCACGCA CAAGATGATCC CTTCGTCC CCTTCGGTCC CCTTCGGTCC CCTTCGGTCC CCTTCGACCCAC TGTAACCCAC TGTAACCCAC AATAAGGCG GAATGTATTT	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGTCATAGCT GAGCTAACTC CACCCTGGAA GCTCACGCTG TAATACGGTT GATTATCAGT TAATATACGTT TAATATATCGT TAATATATCGT TAACTATTGT TTTAAATTTAA TCTGTCTATT GCGAGACCCA CAGTCGTTGCACGCT CCGCATGGTT CCGCATGGTT CCGCATGGTT CCGCATGGTT ACGCATGGTT ACGCATGGTT ACGCATGGTT ACGCATGGTT ACGCATGGTT ACGCAGAAT ACACAGGAAAT ACACAGGAAAT AAAAAAATAA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGGTTCCCCC GTCAGTCAGTCCCC TAGGTATCTC CTGAGTCCA ATCCACAGAA ATCACGGGT AATCACGGGT AATCACGGGT ATCGTCACCGG TAGGTCACCGG TATGGTCACCGG TATGGTTCATCC GGTCACCGG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGAAGTA ACTGAAAGTA AACTGAAATACG GTTGAATACT GTTGAAATACT GACAAAATACGG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA ACCACGGTAG GCCTAACTAC CAAACCACCG CTGACGGTAG ACCAGGTAC ACCAGGTAG TAAATCAATC CTACAGATTT CTCAGGTCCG AGTTGGCGCC AGTTGGCCGC ACCAAGTCA ACCAAGTCA ACCAAGTCA CCAAGTCTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCATCTTT CAGCACTTT CATACTTTC CATTCCGCGCA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTCCCGCTCA ACGCAGGAAA GCATCACAAA GCATCACAAA GCATCACAAA ACGCATGACCA ACGCAGCACTA ACGCAGCACTA ACACGACTTA ACACACACAA ATCCCCACGT ATCCCACCGT ATCCCCACGT ATCCCAACTA GTAGTAGTT GTTCCCAACG ATCTCCCAACG ATTCTCAACA ATCTCCAACG ATCTTATCAA TACATTCAACT TACTTCAACT TACTTCAACT TACTTCAACT TACTTCAAC TTCTTCAACT TACTTCAACT TACTTTCAACT TACTTCAACT TACTTTCAACT TACTTTCAACT TACTTTCAACT TACTTCAACT TACTTTCAACT TACTTTC
2001 TTCTGCAGAT ATCC	CAGCACA GTO Q H S CAGCATCA Q H S CAGCATCA CAGCATCA F R T ( FTTGGAC AAI AAGTTAA CAI FGATTAT GA: ATACCAGC CTCCGGTC GTC ATACCAGCACA ATACCAGCACA TCAGCACAC TCGGTATC TCG TCGTATC TCG TCGTATC TCG TCGTATC TCG TCGCAACA ACA TCGACACAC TCGCACACACACACACACACACACACACACACACACACAC	GGCGGCCG G G R GTCATCAT G H H ACCACAAC ACACACA GGAAGCAT TCAGTCGA GGCAGCTG TCGGCTGC CCACAAAC CGCCTTTC GGAACCCC GGAACCCC GGAACCCC CAGAAT TCAGTCGA CGCTTTC GAACCCCC GTACCAG CGCTTTC CAGACCCC CAGATTAC TCATCACAT GAGGGCTT TCATCACT TCATCACT TCATCACT TCCTTTCC CAGAACC CCCATGTT CCTCTTTCC CAGAACG	CTCGAGTCTA S S L CACCATCACC H H H TAGAATGCAG TGCATTCATT CCTGCAGGCA AAAGTGTAAA GGCGAGCGGT CAGGAAACGGT CCGCTCGGG CGTTCAGCC TCCTTCGGG CGTTCAGCC TTATCAGCA TTATCAGAAA GCCAGTACTACACA GCCATTCGTC CCTTCGGC CGTTCAACC CCGCTCAAC CCCTTCGCC CCGTTCGCC CCATTGCTAC CCGCTCCAC CCGCTCCAAC CCAACATCCCC CCACTCCAAC CCACTCCAAC CCACTCCAAC CCACTCCAAC CCAACATCCCC CCACTCCAAC CCACTCCAAC CCAACATCCCC CCAACACAC CCAACATCCCC CCAACACAC CCAACACC CCAACACC CCAACACC CCAACACC CCACACC CCAACACC CCAACACC CCAACACC CCAACACC CCAACACC CCACACC CCAACACC CCACACC C	GAGGGCCCGC E G P ATTGAGTTTA H TGAAAAAAT TTATGTTTCA TGCAAGCTTG GCCTGGGGTG TCGGCCAACG ATCAGCTCA AAAAGGCCC ATCAGCTCAC AAAAGGCCC GACCGCTGCG GACGGTGCG GACGTATGT TACCTTCGGA AAAGATTC CCAGTGCTG CGAGTTATC CCAGTGCTG CGAGTTATC CCAGTGCTG CGAGTTAGCT CCCAGTGCTG CGAGTTAGCT TCCAGTTCTC CCAGTGCTG CGGGTTAGCT TGCCATCCGT ACGGGATATT TCCAGTTCGA CAAAAAAAAAGGC AAAAAAAAAGG	GGTTCGACTA R F D Y AACCCGCTGA GGTTCATGG GGTTCAGGG GCGTAATCAT CCTAAATGAGT CCTAAAAGGCGG CGTTGCTGGC CCAGGCGTTCTCTCATA CCTTATCCGG AGGCGGTGCT AAAAGAGTTC CTTACTATCCGG CAAGAGATCCT ATCTCACGG CAAGAGATCCT ATCTCACGCA CAAGATGATCC CTTCGTCC CCTTCGGTCC CCTTCGGTCC CCTTCGGTCC CCTTCGACCCAC TGTAACCCAC TGTAACCCAC AATAAGGCG GAATGTATTT	CAAGGATGAC  K D D  TCAGCCTCGA  TGAAATTTGT  GAGGTGTGGG GGCGTTTGC GAGCTAACTC CCCCCTGGAA GCTCACGCTG TAATACGGT TAATATACGT TAATATATATATATATATATATATATATATATATATA	GATGACAAGG D D K CTGTGCCTTC  GATGCTATTG AGGTTTTTTA GTTTCCTGTG ACATTAATTG GTATTGGGG ATCCACAGAA AGGGTTCCCCC GTCAGTCAGTCCCC TAGGTATCTC CTGAGTCCA ATCCACAGAA ATCACGGGT AATCACGGGT AATCACGGGT ATCGTCACCGG TAGGTCACCGG TATGGTCACCGG TATGGTTCATCC GGTCACCGG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGTTCACCG TATGGAAGTA ACTGAAAGTA AACTGAAATACG GTTGAATACT GTTGAAATACT GACAAAATACGG	GCGAGCAGAA G E Q K TAAGATCCAG CTTTATTTGT AAGCAAGTAA TGAAATTGTT CGTTGCGCTC TCAGGGGATA CCCCTGACGA ACCACGGTAG GCCTAACTAC CAAACCACCG CTGACGGTAG ACCAGGTAC ACCAGGTAG TAAATCAATC CTACAGATTT CTCAGGTCCG AGTTGGCGCC AGTTGGCCGC ACCAAGTCA ACCAAGTCA ACCAAGTCA CCAAGTCTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCACTTT CAGCATCTTT CAGCACTTT CATACTTTC CATTCCGCGCA	GCTGATCTCC L I S ACATGATAAG AACCATTATA AACCATTACA ATCCGCTCAC ACTCCCGCTCA ACGCAGGAAA GCATCACAAA GCATCACAAA GCATCACAAA ACGCATGACCA ACGCAGCACTA ACGCAGCACTA ACACGACTTA ACACACACAA ATCCCCACGT ATCCCACCGT ATCCCCACGT ATCCCAACTA GTAGTAGTT GTTCCCAACG ATCTCCCAACG ATTCTCAACA ATCTCCAACG ATCTTATCAA TACATTCAACT TACTTCAACT TACTTCAACT TACTTCAACT TACTTCAAC TTCTTCAACT TACTTCAACT TACTTTCAACT TACTTCAACT TACTTTCAACT TACTTTCAACT TACTTTCAACT TACTTCAACT TACTTTCAACT TACTTTC

### pMT-Dm43 (68-363) noSTOP-FlagMycHis

1 TCGCGCTTT CGGTGATGAI 101 TCAGGGCGC TCAGCGGGT 201 CCGCACAGAT GCCTAAGGAI 301 TAGGCCAGCT GCGGAAAGG 401 TGAATTAATT CGTTGCAGG, 501 GGCCCCCAC CGCCCACCG 601 CAAGTCCCA AACTGCAAC, 701 AGAGGTGAAT CGAACGAAAI 801 AAATCAAGTG AATCATCTC	TTGGCGGGTG TCGC AAAATACCGC ATCF GGATGTGCTG CAAC CCGCCCATA CATT A CCGAACCAAT TCTT A CCCGATGTGT AAAC ACCGTGTGT AAAC	EGGGCTGG CTTAACTATG PAGGCGATT AAGTTGGCTATGGCCATT AGGCGATT GACTAGCTCT PATGTGGT ACGCAAGTAA CTCGCGGG CAGAACAAAAAAAAAAAAAAAAAAAAAAAAA	CGGCATCAGA CAGGCTGCGC ACGCCAGGGT TTGCTGCAGG GAGTGCCTGC GCTTCTGCAC TATAAAACCG	GCAGATTGTA AACTGTTGGG TTTCCCAGTC CCGTCCTATC GCATGCCCCA ACGTCTCCAC AGAGCATCTG	CTGAGAGTGC AAGGGCGATC ACGACGTTGT CTCTGGTTCC TGTGCCCCAC TCGAATTTGG GCCAATGTGC	ACCATATGCG GGTGCGGGCC AAAACGACGG GATAAGAGAC CAAGAGTTTT AGCCGGCCGG ATCAGTTGTG	GTGTGAAATA TCTTCGCTAT CCAGTGCCAG CCAGAACTCC GCATCCCATA CGTGTGCAAA GTCAGCAGCA
901 CCCGACGCAC CACCGGGGA						>	M S C
>S R R T T G D	V F P A S	S R A Q R I	G G F	F L L Y	V I Y	Y K Q	P T H N
1001 CTTTATTAAG ATCGAGGTC	CACCGCGCAC TTGG	GCAAGAA CTAACAGACT	ACGCTCTAGA	TCTGCGCAAG	${\tt GATAGTCCGG}$	${\tt AGCGGAAGGA}$	CACTCATCAG
> F I K I E V			Y A L D		D S P	E R K D	T H Q
1101 ATCGCCTACA TGCTGTGGC							
> I A Y M L W 1 1201 TGGAGACCGT AGCGGGTGC		Q A F R F I			G L D N	L V D	Y D R
>V E T V A G A		O S A L M O			V S L	T Y E	L E G L
1301 GCGAGCACTG GACCAGGCA	~ ^	~ ~	~ ~				
> R A L D O A			Y N A O		L A A		A L P
1401 CCCTCTCAAA TATTCGGCC	A TTTGCGAGAA GTC1	TTTGCCG ATATCCAAAG	TGTTCTAGGA	GCTAGAAAGA	GTACTCCAGA	TGAGAAATGC	ACCACAACAT
> P S Q I F G I			V L G		${\tt S}  {\tt T}  {\tt P}  {\tt D}$	E K C	T T T
1501 CTACAGGCAA CCAGTTGGA							
>S T G N Q L E	-	V R N K A M	Y G V	E E R E	P Q H	QTD	E L E V
1601 GCAGCTGGAG GTCAACGAG							
> Q L E V N E 1701 GAGTTTAGTG ACGACGAGG			V F Q R		E D V	Q Q E Y	E M I
> E F S D D E 1		E S E E V I			I L D T	K G N	S A D
1801 TCCAGCACAG TGGCGGCCG							
		G P R F D Y					
>I O H S G G R	SSLEG	GPKFDI	K D D	D D K G	E O K	LIS	EEDL
>1 Q H S G G R 1901 GACGCGTACC GGTCATCATC > T R T G H H							
1901 GACGCGTACC GGTCATCATC  TRTGHH  2001 AGTTTGGACA AACCACAACC	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA	BAGTTTAA ACCCGCTGAT	CAGCCTCGAC GAAATTTGTG	TGTGCCTTCT ATGCTATTGC	AAGATCCAGA TTTATTTGTA	CATGATAAGA ACCATTATAA	TACATTGATG GCTGCAATAA
1901 GACGCGTACC GGTCATCATC  TRTGHH  2001 AGTTTGGACA AACCACAACC 2101 ACAAGTTAAC AACAACAAT	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA F GCATTCATTT TATG	AAAAAATG CTTTATTTGT	CAGCCTCGAC GAAATTTGTG AGGTGTGGGA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA	AAGATCCAGA TTTATTTGTA AGCAAGTAAA	CATGATAAGA ACCATTATAA ACCTCTACAA	TACATTGATG GCTGCAATAA ATGTGGTATG
1901 GACGCGTACC GGTCATCATC  TRTGHH  2001 AGTTTGGACA AACCACACAC  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATG ATCAGTGAC	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA F GCATTCATTT TATG C CTGCAGGCAT GCAA	AGTTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT CTTTCAG GTTCAGGGGG AGCTTGG CGTAATCATG	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA
1901 GACGCGTACC GGTCATCAT'  TRTGHH 2001 AGTTTGGACA AACCACAC' 2101 ACAAGTTAAC AACAACAT' 2201 GCTGATTATG ATCAGTGGA' 2301 ACATACGAGC CGGAAGCAT'	C ACCATCACCA TTGA H H H H C AGAATGCAGT GAAA C GCATTCATTT TATG C CTGCAGGCAT GCAA A AAGTGTAAAG CCTG	AGTTTAA ACCGCTGAT AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CGTAATCATG	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT	TACATTGATG  GCTGCAATAA  ATGTGGTATG  ATTCCACACA  TCCAGTCGGG
1901 GACGCGTACC GGTCATCATC  TRTGHH  2001 AGTTTGGACA AACCACACAC  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATG ATCAGTCGAC  2301 ACAATACGAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGGTG	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA C CTGCAGGCAT GCAA A AAGTGTAAAG CCTC C ATTAATGAAT CGGC	AGTTTAA ACCGCTGAT AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CGTAATCATC GGGGTGC CTAATGAGTG GCCAACGC GCGGGGAGAG	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTCGCTG
1901 GACGCGTACC GGTCATCATC  TRTGHH  TRTGHH  2001 AGTTTGGACA AACCACAACC  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATG ATCAGTCGAC  2301 ACATACGAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCCGGCTGCC	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA F GCATTCATTT TATG C CTGCAGGCAT GCAA A AAGTGTAAAG CCTG ATTAATGAAT CGG G GCGAGCGGTA TCAG	AGTTTAA ACCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CGTAATCATC GGGGTGC CTAATGAGTG CCCAACGC GCGGGGAGAG GCTCACT CAAAGGCGG	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC TCCACAGAAT	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGGATAA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG	TACATTGATG GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTCGCTG AACATGTGAG
1901 GACGCGTACC GGTCATCATC  TRTGHH  TRTGGHH  2001 AGTTTGGACA AACCACAACT  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATGA TCAGTCGAC  2301 ACATACGAGC CGGAAGCAT  2401 AAACCTGTCG TGCCAGCTG  2501 CGCTCGGTCG TTCGGCTGC  2601 CAAAAGGCCA GCAAAAGGCC	C ACCATCACCA TTG# H H H H T AGAATGCAGT GAA# GCATTCATTT TATTC C CTGCAGGCAT GCA# A AAGTGTAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAGC AGGAACCGTA AAA#	AGTTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTCAGGGGG AGGCTTGG CTAATGAGTG CCAAACGC GCGGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGGCTGGCC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTCA AGCTAACTCA GCGGTTTGCG AATACGGTTA TTTTTCCATA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC TCCACAGAAT GGCTCCGCCC	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTTA TCTTCCGCTT CAGGGGATAA CCCTGACGAG	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA	TACATTGATG GCTGCAATAA ATGTGGTATG ATTCCACAACA TCCAGTCGGG TGACTCGCTG AACATGTGAG ATCGACGCTC
1901 GACGCGTACC GGTCATCATC  TRTGHH  TRTGHH  2001 AGTTTGGACA AACCACAACC  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATG ATCAGTCGAC  2301 ACATACGAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCCGGCTGCC	C ACCATCACCA TTGA H H H H H T AGAATGCAGT GAAA G GCATTCATTT TATC C CTGCAGGCAT GCAA A AAGTGTAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAC C AGGAACCGTA AAAA C CGACAGGACT ATAA	AGATTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATG CCCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGCTGCCC AAAGATAC CAGGCGTTTC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA TTTTTCCATA CCCCTGGAAG	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC TCCACAGAAT GGCTCCGCCC CTCCCTCGTG	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG CGCTCTCCTG	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA TTCCGACCCT	TACATTGATG GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTCGCTG AACATGTGAG ATCGACGCTC GCCGCTTACC
1901 GACGCGTACC GGTCATCATC  TRTGHHA  2001 AGTTTGGACA AACCACAAC  2101 ACAAGTTAAC AACAACAAT  2201 GCTGATTATG ATCAGTCGAC  2301 ACATACCAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCGGCTGCC  2501 CAAAAGGCCA GCAAAAAGGC  2701 AAGTCAGAGG TGGCGAAACC  2701 AAGTCAGAGG TGGCGAAACC	C ACCATCACCA TTGA H H H H F AGAATGCAGT GAAA GCATTCATTT TATC C TGCAGGCAT GCAA A AAGTGTAAAG CCTC ATTAATGAAT CGGG GCGAGCGGTA TCAC AGGAACCGTA AAAA C CGACAGGACT ATAA C CCCTTCGGGA AGGC	AGATTAA ACCCGCTGAT  AAAAATG CTTTATTTGI GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC GCGGGTGC CTAATCATC CCCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGI AAGGCCGC GTTGCTGCGC CTGGTGCGCCCC TTCCTCATAC GTGGCCGC TTTCTCATAC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA TTTTTCCATA CCCCTGGAAG CTCACGCTGT	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC TCCACAGAAT GGCTCCGCCC CTCCCTCGTG AGGTATCTCA	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG CGCTCTCCTG GTTCGGTGTA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA TTCCGACCCT GGTCGTTCGC	TACATTGATG GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTCGCTG AACATGTGAG ATCGACGCTC GCCGCTTACC TCCAAGCTGG
1901 GACGCGTACC GGTCATCATC  TRTGHHA 2001 AGATTGGACA AACCACACC 2101 ACAAGTTAAC AACAACAAT 2201 GCTGATTATC ATCAGTCGAC 2301 ACAACCTGC TGGCAGCAC 2401 AAACCTGTCG TGCCAGCTG 2501 CGCTCGGTCG TTCGGCTGC 2501 CAAAAGGCAC GCAAAAAGGC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTC 2901 GCTGTGTGCA CGAAACCCCC 3001 AGCAGCCACT GGTAACAGG	C ACCATCACCA TTGA H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAC C AGGAACCGTA AAAA C CGACAGGACT ATTA C CCCTTCGGGA AGCC G TTTAGCCCG ACCC ATTAGCAGACG GACC	AGATTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATG CCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGCTGCGC AAAGATAC CAGGCGTTTC CGTGGCGC TTTCTCATAC GCTGCGC CTTATCCGGT GGTTATGTA GGCGGTGCTA	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CCCCTGGAAG CTCACGCTGT AACTATCGTC CACAGGTGTT	TGTGCCTTCT  ATGCTATTGC GGTTTTTAA TTTCCTGTGT CATTAATTGC TCCACAGAAT GGCTCCGCCC CTCCCTCGTG AGGTATCTCA AGGTATCTCA GAAGTGGTGG	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG GGTCTCCTG GTTCGGTGTA CCCGGTAAGA CCTAACTACG	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC GCAGGAAAG CATCACAAAA TTCCGACCCT GGTCGTTCGC GGCAGGATTAT GCTACACTAG	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTCGCTG AACATGTGAG ATCAGCGTTAC TCCAAGCTGC CGCCTTACC TCCAAGCTGG CGCACTGGC AAGGACAGTA
1901 GACGCGTACC GGTCATCATC  TRTGHHA 2001 AGTTTGGACA AACCACAACC 2101 ACAAGTTAAC AACAACAAT 2201 GCTGATTATG ATCAGTGGAC 2301 ACAATACGAGC CGGAAAGCAT 2401 AAACCTGTCG TGCCAGCTGC 2501 CGCTCGGTCG TTCGGCTGCC 2601 CAAAAGGCCA GCAAAAGGCC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTCC 2901 GCTGTGTGCA CGAACCCCC 3001 AGCAGCCACT GGTAACAGG 3101 TTTGGTATCT GCGCTCTGCC	C ACCATCACCA TTGA H H H H GAGATGCAGT GAAF GCATTCATTT TATC CTGCAGGCAT GCAA AAGTGTAAAG CCTC ATTAATGAAT CGGC GCGAGCGGTA TCAC CGACAGGACT ATAA CCCTTCGGGA AGCC GTTCAGCCCC ACCC ATTAGCAGACCGTA ACCC ATTAGCAGAGC GAGC GTCAGCACCAGT	AGATTAA ACCGCTGAT  AAAAATG CTTTATTTGT  GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC GCGGGGC CTAATCATC CCCACC CGGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTCCTGCGC CTTATCCGGC CTTATCCGGT GCTGGCGC CTTATCCGGT CTTCGGAA AAAGAGTTGC CTTCGGAA AAAGAGTTGC	CAGCCTCGAC GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA AGCTATACGGTTA TTTTTCCATA TTTTTCCATA CCCCTGGAAG CTCACGGTGT AACTATCGTC CAGAGTTCTT TAGCTCTTGA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGGG TCCACAGAAT GGCTCCGCC CTCCCTCGTG AGGTATCCA TTGAGTCCAA TGAGTCCAA GAACTGGTGG TCCGGCAAAC	AAGATCCAGA TTTATTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTC CAGGGGATAA CCCTGACGAG GGTTCCGTG GTTCGGTGTA CCCGGTAAGA CCCTAACTACA AAACCACCG	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA TTCCGACCCT GGTCGTTCGC CACGACTTAT TGCTACACTAG TGGTAGCGGT	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTGGGT AACATGTGAG ATGAGAGGCTC TCCAAGCTGG CGCGCTTACC TCCAAGCTGG CGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC GGCACTGGC
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACAAGTTACC AACCACACC 2101 ACAAGTTACC AACACACT 2201 GCTGATTATCA TCAGTCGAC 2301 ACATACGAGC CGGAAGCAT. 2401 AAACCTGTCG TCCCACCTGG 2501 CGCTCGGTCG TTCGGCTGC 2601 CAAAAGGCCA GCAAAAGGC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTCC 2901 GCTGTGTGCA CGAACCCCC 3001 AGCAGCCACT GGTAACAGG. 3101 TTTGGTATCT GCGCTCTCGC 3201 TTTGCAAGCA GCAGATTACC	C ACCATCACCA TTGA H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAC AGGAACCGTA AAAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CCCTTCGGGA AGCC G TTCAGCCCG ACCC A TTAGCACAGC GACC G TTAGCACAGC G TGAAGCCAGTT ACCT G GAAGCCAGTT ACCT G CGCAGAAAAA AAGG	AGATTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATGATCATC GCGAGGC CTAATGAGTG GCCAACGC GCGGGGAGAG AGCCCG GTTGCTGGCC AAAGATAC CAGGGGTTTC GGTGGCG TTTCTCATAG GCTCACC CTTATCCGGT GGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTA	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCAATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CCCCTGGAAG CTCACGCTGT AACTATCGTC CAGAGTTCTT TAGCTCTTGA	TGTGCCTTCT  ATGCTATTGC GGTTTTTAA TTTCCTGTGT CATTAATTGC TCACAGAAT GGCTCCGCC CTCCCTCGTG AGGTATCCA AGGTATCCA TGAGTCCA GAAGTGCTGG TCCGCAAAC CTACGGGTAC CTACGGGTAC	AAGATCCAGA TTTATTGTA AGCAAGTAAA GAAATTGTTA GTTACGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG GCTCTCCTG GTTCGGTGTA CCCGGTAAGA CCTAACTACG AAACCACCGC TGACGCTCA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTGGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC GGCACGACTTAT GCTACACTAG TGGTAGCGGT TGGGAACGAAA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGCGG TGACTCGCT AACATCTGAG TCCAGCGCT CCCAGCTTACC TCCAAGCTG CGCCACTGGC AAGAACAGTA GGTTTTTTTG ACTCAGCTTA
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGHR  2001 ACAGGTTACC AACCACACC  2101 ACAGGTTACC AACCACACC  2201 GCTGATTATC ATCAGTCGAC  2301 ACATACCAGC CGGAAGCAT  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCGGCTGCC  2601 CAAAAGGCCA GCAAAAGGCC  2701 AAGTCAGAGG TGGCGAAACC  2801 GGATACCTGT CCGCCTTTCC  2901 GCTGTGTGCA CGAACCCCCC  3001 AGCAGCCACT GGTAACAGG  3101 TTTGGTATCT GCGCTCTGCC  3201 TTTGCAAGCA GCAGATTACC  3301 AGGGATTTTC GTCATGAGAC	C ACCATCACCA TTGA H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGCTTCAGGCA AGCC GTTCAGCCCG ACCC A TTAGCAGAGC GAAGCCAGTT ACCT C CGCAGAAAAA AAGC T TATCAAAAAAG GATC	AGATTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC GCGACGC CTAATCATC CCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGT AAGGCGGC GTTGCTGGCC AAAGATAC CAGGCGTTTC GCTGGCGC TTTCTCATAG GCTTGCGC CTTATCCGGT GGTATGTA GGCGGTGCTA TTTCAGAA AAAGAGTTGC GGATCTCA AGAAGATTCC TCTTCACC TAGATCCTT	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTGA GCGGTTTGCG AATACGGTTA CCCCTGGAAG CTCACGCTGT AACTATCGTC CAGAGTTCTT TAGCTCTTGA TTGATCTTTT TAAATTAAAA	TGTGCCTTCT  ATGCTATTGC GGTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TGGCTCCGCC CTCCCTCGTG AGGTATCTCA TGAGTCCAA GAAGTGGTCCA GAAGTGGTGC TCCGGCAAAC CTACGGGGTA CTACGGGGTT	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAAA CCCTGACGAG CCCTGACTAGCAG CCCGGTAAGA CCCTGACTACC AAACCACCGC TGACGCTCAG AAATCAATCT	CATGATAAGA  ACCATTATAA  ACCTCTACAA  TCCGCTCACA  CGCAGGAAGA  TTCCGACCCT  GGTGGTTCGC  ACGACTTAT  GCTACACTAA  TGGTAGCGGT  TGGAACGAA  AAAGTATATA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGCTG AACATGTGAG TGACTCGCTG AACATGTGAG GCCGCTTACC TCCAAGCTGG CGCCACTGGC AAGGACAGTA GGTTTTTTTT GGTTTTTTTTTT
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGHR  2001 AGTTTGGACA AACCACAAC  2101 ACAAGTTAAC AACCACACAC  2201 GCTGATTATG ATCAGTGGAC  2301 ACATACGAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCGGCTGCC  2601 CAAAAGGCCA GCAAAAGGGC  2701 AAGTCAGAGG TGCCGAAACC  2801 GGATACCTGT CCGCCTTTCC  2901 GCTGTGTGCA CGGAACCCCC  3001 AGCAGCCACC GGTAACAGG  3101 TTTGGTATCT GCGCTCTGCC  3201 TTTGGTAAGCA GCAGATTACC  3301 AGGGATTTTG GTCATGAGAC  3401 TGGTCTGACA GTTACCAATC	C ACCATCACCA TTGA H H H H I GAAATGCAGT GAAA GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAA CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAC C AGGAACCGTA AAAA C CGACAGGACT ATTA C CCCTTCGGGA AGCC ATTAGCAGGC ACCC ATTAGCAGGC GACC T TATCAGAAAAA AAGC G GCAGCAGAAAAA AAGC T TATCAAAAAG GATC G CTTAATCAATAGG GATC C CTTAATCAGT GAGC C CTTAATCAGT GAGC	AGTTTAA ACCCGCTGAT  AAAAATG CTTTATTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC CCCAACGC CCGGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGCTGCGC AAGGATAC CAGGGGTTTC CGTGGCGC CTTATCCGGT SGTATGTA GGCGGTGCTA CTTCAGCA AAAGATCCT AGACACCT AAAGACTCT CGTCACC TAGATCCCTT CGCACCT AAAGACTCT CGCACCT AAGACATCCT CGCACCT ATCACGGAT	CAGCCTCGAC  GAAATTTGTC AGGTGTGGGA GTCATAGCTG AGCTAACTCA CCGTTTTCCGTA TTTTTCCATA TTTTCCATA ACTCATCGTC AACTCATCTCT TAGCTCTTTAGCTCTTT TAGCTCTTGA TTGATCTTT TAGCTCTTGA TTGATCTTT TAGATTATAAA CTGTCTATTT	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGGC TCCACAGAAT GGCTCCGCC CTCCCTCGTG AGGTATCCA TTGAGTCCAA GAAGTGGTGG TCCGGCAAAC CTACGGGGTC ATGAAGTTTT CGTTCATCCA	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTC CCTGACGAG CCTTCCTG GTTCGGTGTA CCCGGTAGA CCTAACTACG AAACCACCGC TGACGCCCGC TGACGCCCAC AAACCACCGC TGACGCTCAG AAACCACCGC TGACGCTCAG AAATCAATCT TAGTTGCCTG	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA TTCCGACCCT GGTCGTTCGC CACGACTAAT TGGTAACCAA TGGTAGCGGT TGGAACGAAA AAGTAATATA ACTCCCGGTC	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTGGGG ACATGTGAG ACCACTGGACGCTC TCCAAGCTGG CGCGCTTACC TCCAAGCTGG CGCACTGGC AACGACAGTAG AGGACAGTA GGTTTTTTG ACTCACGTTA
1901 GACGCGTACC GGTCATCATC  TRTGHT  ACACACA  2001 ACAGGTTACA AACCACAAC  2101 ACAAGTTACA AACACAAC  2101 ACAAGTTACA AACACAAC  2201 GCTGATTATCA TCAGTCGAC  2301 ACATACGAGC CGGAAGCAT  2401 AAACCTGTCG TGCCAGCTG  2501 CGCTCGGTCG TTCGGCTGC  2601 CAAAAGGCCA GCAAAAGGC  2701 AAGTCAGAGG TGGCGAAAC  2801 GGATACCTGT CCGCCTTTC  2901 GCTGTGTGCA CGAACCCCC  3001 ACCAGCCACT GGTAACAGG  3101 TTTGGTATCT GCGCTCTGC  3201 TTTGCAAGCA GCAGATTAC  3301 AGGGATTTTG GTCATCAGAC  3401 TGGTCTGACA GTTACCAATC  3501 CTACGATACC GGAGGGCTT.  3501 CTACGATACC GGAGGGCTT.	C ACCATCACCA TTGA H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTCCAGGCAT GCAA AACTGTAAAG CCTC ATTAATGAAT CGCC AGCAGCGGTA AAAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ACC G TTTAGCACAGC G TTAGCACAGC G TGAAGCCAGTT ACCT G CAGAGCCAGTT ACCT G CAGAGAAAAA AAGG T TATCAAAAAG GATC C CTTAATCAGT GAGC A CCATCTGGCC CACC C CTTCAGCT GAGC	AGATTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATGATGT GCGAGGG CTAATGAGT GCAACGC GCGGGGAGAG AGCTCAC CAAAGGCGGT AAGGCCGG GTTGCTGGCC AAAGATAC CAGGGGTTTC GTGGGC CTTATCAGTA GGTAGTAGT GTTTCCAGTAGTAGTAGT AAAAAGAGTTGC GGAACTCA AGAAGATCCT CTTCACC TAGATCCTT AGCACTA AGTACTCT AGGACCTA TCTCAGCGAT AGTACTCT AGTACTCA AGTACTCT AGGACCTA TCTCAGCGAT AGTACTCT AGTACTCA AGTACTCT AGGACCTA TCTCAGCGAT AGTACTCC AATGATACCC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCAATAGCTG AGCTAACTCA GCGGTTTCGG AATACGGTTA CCCTGGAAG CTCACGCTGT AACTATCGTC CAGAGTTCTT TAGATCTTTA TAGATCTTTT TAAATTAAAA CTGTCTATTT CGAGACCCAC	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TTGGGTCCCC CTCCCTCGTG AGGTATCCAA GAAGTGCTGG TCCAGCAAAC CTACGGGAAAC CTACGGGGTC ATGAAGTTTT CGTTCATCAC GCTCACCGGC	AAGATCCAGA  TTTATTTGTA AGCAAGTANA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CCCGGGATAA CCCCGGCTAAGA CCCAACACCAC AAACCACCG AAACCACCG TGAGGCTCA AATCAATCT TAGTTGCCTG TCCAGATTTA	CATGATAAGA ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCGCTT CCTCGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC ACGACTTAT GCTACACTAG TGGTAGCGTT TGGTAGCGGT TGGGACGAAA AAAGTATATA AACTCCCCGTC TCAGCAATAA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGTGGG TGACTCGCTG AACATCTGAG TCGAGTGGCT GCCGCTTACC TCCAAGCTGG CGCCATTGGC AAGACAGTA GGTTTTTTT GGTTTATTTT TGAGTAAACT TGGAGTAAACT TGGAGTAAACT TGGAGTAAACT TGAGTAAACT TGAGTAAACT ACCAGCCAGC
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACARGTTAGAC AACCACAAC  2101 ACAAGTTAAC AACACAAC  2201 GCTGATTATCA TCAGTCGAC  2301 ACATACCAGC CGGAAGCAT  2401 AAACCTGTCG TGCCAGCTGG  2501 CGCTCGGTCG TTCGGCTGCC  2601 CAAAAGGCCA GCAAAAGGC  2701 AAGTCAGAGG TGGCGAAAC  2801 GGATACCTGT CCGCCTTTC  2901 GCTGTGTGCA CGAACCCCC  3001 ACCAGCCACT GGTAACAGG  3101 TTTGGTATCT GCGCTGTCC  3201 TTTGCAAGCA GCAGATTACC  3301 AGGGATTATC GCGCTGTGC  3401 TGGTCTGACA GTTACCAATC  3501 CTACGATACG GGAGGGCTT  3601 CGGAAGGCC GAACCCCAGA	C ACCATCACCA TTGA H H H H H I GAGANTGCAGT GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAAG CCTC ATTAATGAAT CGGC GGCAGCGGTA TAAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ACCC ATTAGCAGAG CGAC GTTAGCAGAG CGAC GGAAGCAGTT ACCT C CTTAGCAGAG GAGC GAAGCAGATA ACCC TATCAGAGAG GAGC GAAGCAGATA ACCC TATCAGAGAG GACC ATTAGCAGAG GACC CACATTAATCAGT GACC ACCATCTGGC CACC A CGACTCCTGC CACC	AGATTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC GCGACGC CTAATCATC CCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGT AAGAATAC CAGGCGTTC AAGAATAC CAGGCGTTTC CGTGGGGC CTTATCCGGA GGTATGTA GGCGGTGCTA TCTCAGAA AAGAATCC CCTTCACC TAGATCCTT AGCACCTA TCTCAGCGAT AGTGCTCC AAGAATCCTT AGCACCTA TCTCAGCGAT AGTGCTCC AATGATACCT AGTGTGCTCC AATGATACCC TTTTATCC GCCTCATCC TTTTATCC GCCTCATCC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CCCCTGGAAG CTCACGCTGT AACTATCGTC CAGAGTTCTT TAGCTCTTGA TTGATCTTTTT TAAATTAAAA CTGTCTATTT CGAGACCCAC AGTCTATTAA	TGTGCCTTCT  ATGCTATTGC GGTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TGGGTCCGCC CTCCCTCGTG AGGTATCCAA GAAGTGCTGG TCCAGGAAC CTACGGGAAAC CTACAGGGTC CTACGGGTC CTACGGCAAC CTACAGGGTTC CTACACGCC CTCCTCTCTC CGTTCATCCA GAAGTCTTCATCCA CTCACCGCC TTGTTGCCCGC TTGTTGCCCG	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG CCCTACTAGA CCCGATAGA CCCGATAGA CCTGACACG CTACTACC AAACCACCGC TGACGCTCCC AAACCACCT TAGTTGCCTG TCCAGATTCA GAAGCTAGA	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCGGTTCAC CGCAGGAAGA TTCCGACCTT GGTCGTTCGC ACGACTTAT GCTACACTAA AAGTATTATA ACTCCCGGTC TCACGACTAT TCAGCAATAA TAAGTACTTC	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGCTG AACATGTGAG TGACTCGCT AACATGTGAG GCCACTTACC TCCAAGCTGG ACGACTGGC AAGGACAGTA GGTTTTTTTG ACTACAGTTA TGAGTAAACT GTGTAGATAA ACCAGCCAGC GCCACTTAAT
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACAAGTTACC AACCACACC 2101 ACAAGTTACC AACCACACC 2101 ACAAGTTACC ACAACCACC 2201 GCTGATTATCA TCAGTCGAC 2301 ACATACCAGC CGGAAGCAT. 2401 AAACCTGTCG TGCCAGCTG 2501 CGCTCGGTCG TTCGGCTGCC 2601 CAAAAGGCCA GCAAAAGGCC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTC 2901 GCTGTGTGCA CGAACCCCC 3001 ACCAGCCACT GGTAACAGG. 3101 TTTGGTATCT GCGCTCTGC 3201 TTTGCAAGCA GCAGATTACC 3301 AGGGATTTTG GTCATCAGAC 3401 TGGTCTGACA GTTACCAATC 3501 CTACGATACC GGAGGGCTT.	C ACCATCACCA TTGA H H H H H GAAATGCAGT GAAA GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAG CCTC ATTAATGAAT CGGG G GCGAGCGGTA TCAC CGACAGGACT ATAA CCGTTCAGGCA AGCC ATTAGCAGGACT ATAA CGCAGAGCAGTA ACCC ATTAGCAGGC GACC ATTAGCAGGC GACC GTAGAGCCAGT ACCC TTAGCAAAAA AAGC CGACAGAAAAA AAGC CTATAGCAGC GAGC GAAGCCAGTT ACCT CGCAGAAAAA AAGC CTATACTAGCC CAAC CCTTAATCAGT GAGC ACTAGTCAGCC CAAC CCATTCGCC CAAC CGTGGTCCTGC AACT	AGTTTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC CCAACGC CCGGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGCTGCCC AAGGAGAT GCTGCGC CTTATCCGGT GTTGCGGA AAAGATCC TTTCGGAA AAAGATCCT AGGCGTCTA AGACATCCAT AGACCTTA AGACGTCCTA AGACGTCCTA AGTGCTGC ATCTCACC ATCTCAC ATCTCACC ATCTCACC ATCTCAC ATCTCAC ATCTCAC ATCTCAC ATCTCAC ATCTCAC ATCTCAC AT	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA ACTCA CCCCTGGAAG CCCCTGGAAG CTCACGCTGT AACTATCGTC AGATTTTTT TAGCTCTTGA TTGATTATCATC TAGATCTTT TAGATTAAAA CTGTCTATTT CGAGACCCAC AGTCTATTAA GTCCTTTGAT	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGGCGC TCCACAGAAT GGCTCCGTCG AGGTATCCA TTGAGTCCAA CAAGGTCCAA CTACGGGGTC TCAGCGGGT CCACCGGC ATGAACTTTCATCA TCACCGGC TCACCGGC ATGTACTCACCGC TTGTTGCCGG ATGCCTTCAT	AAGATCCAGA TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTC CCTGACGAG CCCTGACGAG CCTACCTCCTG GTTCGGTGTA CCCGGTAAGA CCTAACTACG AAACCACCGC TGACGCTCAG AAATCAATCT TAGTTGCCTG TCCAGATTTA CAAGCTAAGA CAACCTAAGA TCAAGCTCAGA TCAAGCTCAG	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA TCCGCTCACA CTGCCCGCTT CCTCGCTCAC CGCAGGAAAG CATCACAAAA TCCGCACCT GGTGGTTCGC CACGACTTAT TGGTAGCGGT TGGAACGAAA AAAGTATATA ACTCCCGTC TCACCAATAA TCACCAATAA TCACCAATAA TAAGTAGTTC TTCCCAACGA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTGGGG ATCGAGCGTC ACCATGGGC TCCAGCTTACC TCCAAGCTGG GGCCACTGGC GCCACTGGC ACCACGGG ACGACAGTAC GTTTTTTTG ACTCAGGTTAA ACCAGCCAG GCCAGTAAC TCAAGGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC TCAAGCCAGC
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGHR  2001 AGTTTGGACA AACCACAAC  2101 ACAAGTTAAC AACAACAAC  2201 GCTGATTATC ATCAGTCGAC  2301 ACAATACGAC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGCTGC  2501 CGCTCGGTCG TTCGGCTGCC  2501 CAAAAGGCCA GCAAAAGGCC  2701 AAGTCAGAGG TGCCGAAACC  2801 GGATACCTGT CCGCCTTTCC  2901 GCTGTGTGCA CGAACCCCC  3001 AGCAGCCACT GGTAACAGG  3101 TTTGGTATCT GCGCTCTGCC  3201 TTTGGAAGCA GCAGATTACC  3301 AGGGATTTGC GCAGATTACC  3401 TGGTCTGACA GTTACCAATC  3501 CTACGATACC GGAGGGCTT.  3601 CGGAAGGGCC GAGCGCCAGA  3701 AGTTTGCCA ACCTTCTTGCC  3701 AGTTTTCGCA ACCTTCTTGTGCCA	C ACCATCACCA TTGA H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTCCAGGCAT GCAA AACTGTAAAG CCTC ATTAATGAAT CGCC ATTAATGAAT CGCC AGCAGCGGTA AAAA C CGACAGGACT ATAA C CGACAGGACT ACCT G TATACACAGG T TATCAAAAAA GACT C TATCAAAAAAA GACG A CTGGTCCTGC CCAA A GTGGTCCTGC ACCT C CATTGCTACA GGCA C TATCGTACA GGCA C TGCAAAAAAA	AGATTAA ACCCGCTGAT  AAAAATG CTTTATTGT GTTTCAG GTCAGGGGG AGCTTGG CTAATGATGT GCGAGGG CTAATGATGT GCAACGC GCGGGGAGAG AGCTCACT CAAAGGCGGT AAGGCCGC GTTGCTGGCC AAAGATAC CAGGGGTCTATCGGG GCTAGCGC CTTATCCGGT GGTAGTATA AAAGAGTCT TTCCACA AAAAGATCT CCTCACA AAAAACTCT CCTCACA TAGACATCT AGCACTA TCTCAGCAT AGTAGTCCTA TCTCAGCAT AGTAGTCCTA TCTCAGCAT AGTAGTCCTA TCTCAGCAT CGCACTA TCTCAGCAT CGTACTCA TCTCAGCAT CTTTAGCC AATGATACCC CTTTATCC CTTCCACTC CTTCAGCTCT TTTAGCTCT TTTCAGCTCT	CAGCCTCGAC  GAAATTTGTG AGGTGGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CCCTGGAAG CTCACGCTGT TAGCTCTTGA ACTATCGTC CAGAGTCTTTAACTTTTTAAATAAA CTGTCTATTT CAGAGCCAC AGTCTATTAA GCCGATCGTTTGGA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TGGGTCCCC TCCCTCGTG AGGTATCCA AAGTGCTCA CAAGGAAT CCAGCAAAC CTACGGGAAAC CTACGGGTC ATGAAGTTTT CGTTCATCAC GCTCACCGGC TTGTTGCCGG ATGCTTCATCA ATGACTTCAT TCAGAAGTTAT	AAGATCCAGA TTTATTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTTA CCCTGACGAG CGCTCCCTG GTTCGGTGTA CCCGGTAAGA CCTAACTACG AAACCACCGC AAATCAATCT TAGTTGCCTG TAGTTGCCTG TCCAGATTTA GAAGCTACAG GTTCAGCTCCGG GTTTGCCCGG	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCGCTT CCTCGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC ACGACTTAT GCTACACTAG TGGTACGCTT TGGTACGACTAA AAATATATAA AAAGTATATA TAAGTAGTAT TCAGCAATAA TAAGTAGTTC TTCCCAACGA GTGTTTCTC	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACGA TCCAGTGGG TGACTCGCTG AACATCTGAG TCGCGCTTACC TCCAAGCTGG CGCCATTGGC AAGACAGTG GATTTTTTG ACTCAGTTATTTTTT TGAGTAAACT TCGAGTAAACT TCGAGTAAACT TCGAGTAAACT TCGAGTAAACT TCAAGGGGG TCAAGGTAAT TCAAGGCAG TCAATGATAA TCAAGGCAG TCAATGATAAT TCAAGGCAG TCAATGATTAT
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGHR  2001 AGTTTGGACA AACCACAAC  2101 ACAAGTTAAC AACAACAAC  2201 GCTGATTATG ATCAGTGGAC  2301 ACAATACGAGC CGGAAGCAT.  2401 AAACCTGTCG TGCCAGGTGC  2501 CGCTCGGTCG TTCGGCTGCC  2501 CAAAAGGCCA GCAAAAGGCC  2701 AAGTCAGAGG TGCCGCTTTC  2901 GCTGTTGGA CGAACCCCC  3001 AGCAGCCACT GGTAACAGG  3101 TTTGGTATCT GCGCTTGC  3201 TTTGGAAGG GCAGATTACC  3301 AGGATTTTG GTCATGAGA  3401 TGGTCTGACA GTTACCAAT  3501 CTACGATACG GGAGGCTT  3601 CGGAAGGCC GAGCGCAGA  3701 AGTTTGCGCA ACGTTGTG  3801 TTACATGATC CCCCATGTTG  3801 TTACATGATC CCCCATGTTG  3901 GGCAGCACT CATAATTCTC  4001 CGACCGAGTT GCTCTTGCCC  4001 CGACCGAGTT GCTCTTGCCC	C ACCATCACCA TTGA H H H H H I AGAATGCAGT GAAA GCATTCATTT TATC CTGCAGGCAT GCAA AAGTGTAAAG CCTC ATTAATGAAT CGGC GCGAGCGGTA TCAC CGACAGGACT ATAA CCGCTCAGGAACCGTA AAAA CCGACAGGACT ATAA CCGACAGGACT ATAA CCGCAGAAAAA AAGC GAAGCCAGTA ACCC ATTAGCAGAGC GAGC GAAGCCAGT ACCC ATTAGCAGAGC GAGC GAAGCCAGTT ACCT CCGCAGAAAAA AAGC CTATACTACAGC GAGC CCACATCTGGCC CACA ACTAGTCAGCC CACA CCATTCGCC CACA CCATTCGCC CACA CCATTCGCC CACA CCATTCTACA GGCA TCCAAAAAAA GGCT TTACTACAA GGCA TCCAAAAAAA GGCT TTACTTACAT GCCAC CGCGTCAATA CGCCAC CGCGTCAATA CGCCAC CGCGTCAATA CGCCAC	AGATTTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC CCCAACGC CCGGGGAGAG AGCTCACT CAAAGGCGT AAGGCCGC GTTGCTGCCC AAAGATACCATC AGTTGCAGC CTTATCCGGT CTTATCCGGT CTTCACCAACACCACCACCACCACCACCACCACCACCACC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA ACTCA CCCCTGGAAG CCCCTGGAAG CTCAGGCTT TAGCTCTTGA TGATCTTT TAGCTCTTGA TGAATCTTT TAAATTAAAA CTGTCTATT CGAGACCAC AGTCTATTAG GTCATTTGGT CCGATCGTTG TCGATCGTTG TCGATCGTTG TCGATCGTTG TCGATCGTTG TCGATCGTTG TCGATCGTTG TCGATCGTTG TCGACTGG TAGCAGAACT	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT GGCTCCCTCGTG AGGTATCCA AGGTATCCA CTACGGGCT CCACGGGAAAC CTACGGGGT CCACCGGC ATGACTTTCACCA GCTCACCGG ATGACTTTCATCCA TGTGACCACGG ATGCTCAT TCAGAAGTAA TCAGAAGTAA TGAGTACTCA TTAAAAGTGC	AAGATCCAGA  TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGGTT CAGGGGATAA CCCTGACGAG GCTCTCCTG GTTCGGTGTA CCCGGTAAGA CCTAACTACG AAACCACCGC TGAGGCTCAG AAATCAATCT TAGTTGCCTG TCCAGATTTA CAGCTCAG TCAGCTCAG TCATCATTAGT TCATCATTATTG	CATGATAAGA  ACCATTATAA  ACCTCTACAA  TCCGCTCACA  CTGCCCGCTT  CTTCGCTCAC  CGCAGGAAAG  CATCACAAAA  TTCCGACCCT  GGTGGTTGGC  CACGACTAAT  TGGTAGCGGT  TGGAACGAAA  AAAGTATATA  ACTCCCGTC  TCACGAATAA  TAAGTAGTTC  TTCCCAACGA  GTGTTATCAC  TCTGAGAATAA  AAAACGTTCT	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA TCCAGTCGGG TGACTGGGG TGACTGGGG ATCGACGCTG ACCATGGG CGCCACTGGC TCCAGCTGG ACCACTGG CGCACTGGC TGAGGAGATA AGGACAGTA GGTTTTTTG ACTCAGGTTA ACTCAGGTTA ACCAGCTAG CCCAGGTAAA TCAAGGCAGG TCATGGTTAT TCAAGGCAG TCATGGTTAT TCAAGGCAG TCATGGTTAT TCAAGGCAG TCATGGTTAT TCAGGCGAG TCATGGTTAT TCAGGCGAG TCATGGTTAT TCAGGCGAG TCATGGCGAC
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACAAGTTACC AACCACACC 2101 ACAAGTTACC AACACACAC 2101 ACAAGTTACC ACAACCACC 2201 GCTGATTATCA TCAGTCGAC 2301 ACATACGAGC CGGAAGCAT. 2401 AAACCTGTCG TGCCAGCTG 2501 CGCTCGGTCG TTCGGCTGCC 2601 CAAAAGGCCA GCAAAAGGCC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTC 2901 GCTGTGTGCA CGAACCCCC 3001 ACCAGCCACT GGTAACAGG. 3101 TTTGGTATCT GCGCTCTGC 3201 TTTGCAAGCA GCAGATTACC 3301 TGGCTGTACCAGC 3401 TGGTCTGACA GTTACCAATC 3601 CGGAAGGCCC GACGCCAGA. 3701 AGTTTGCCAC ACGTTGTTG 3801 TTACATGATC CCCCATGTT 3901 GGCAGCACTG CATAATTCTT 4001 CGACCGAGTT GCTCTTGCC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACCC	C ACCATCACCA TTGA H H H H H H G AGANTGCAGT GCATTCATTT TATC C CTCCAGGCAT GCAA ACTGTATAAG CTC ATTAATGAAT CGCC ATTAATGAAT CGCC ATTAGCAGAGCAT AAAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ACC A TTAGCAGAGC GACC G TTAGCAGAGC GACC C GTTCAGCCC ACCC C GACAGCCAGT ACCT G CAGAGCCAGTT ACCT G CAGAGCCAGTT ACCT C CATTAATCAGT GAGC A CTGGTCTCGC CAAC C CTATATCAGT GAGC C TATCAAAAAG GATC C CATTGCTACA GGCA G TGGTCCTGC CAACT C CATTGCTACA GGCA G TGGTCTCTGC GCCC G TGCAAAAAAG CGCT TTACTGTTCAT GCCC G GGCGTCAATA CGCG G GGCGTCAATA CGCG G GGCGTCAATA CGCG G CTCTTGAGAT CCAC G TTGAGAT CCAC	AGATTTAA ACCCGCTGAT  AAAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATG GGGGTGG CTAATGAGT GCAACGC GCGGGGAGAG AGCTCACT CAAAGGCGGT ATGTGGCCGC GTTGCTGGCC AAAGATAC CAGGCGTCTT CGTGGCC CTTATCAGGT GGTATGTA AAAGAGTCT TCTCACC TAGAACTCT CGTACCA TAGAACTCT AGCACTA TCTCAGCAT AGTATACC GCTCCATCC ATTATCC GCTCCATCC ATTATCC GCTCCATCC ATTAGCT TGTCAGCGT ATTACCT TGTCAGCTT AGTATCAC TGTCAGCTT AGCACTT AGTATCAC CTTCAGCAT AGTATCAC ATTAGCT ATGTCAGCTCT ATTAGCT ATGTCAGCTT ATTAGCT CTTCAGCTCT ATTAGCT CTTCAGCCAT AGATAATA CCGCCACAG AGTTCAGT GTAACCCACT	CAGCCTCGAC  GAAATTTGTG AGGTGGGGA GTCAATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CTCCTGGAAG CTCACGCTGT TAGCTCTTTA ACTATCGTC AGAGATTCTT TAAATTAAAA CTGACTATTT TAAATTAAAA CTGTCTATTT CGAGACCCAC AGTCTATTAA GCCGATCGTTTGGT CCGATCGTTTGGT CTGTGACTGG CTGTGACTGG CGTGCACCCC CGTGCACCCCA CGTGCACCCC	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TGGGTCCCC TCCCTCGTG AGGTATCCAA GAAGTGCTGG TCGGCAAAC CTACGGGAAAC CGTCACGGGT ATGAAGTTTT CGTTCATCAC GCTCACCGGC TTGTTGCCGG ATGGCTTCAT TCAGAAGTTAT TCAGAAGTAT TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC TCAGAAGTAC ACTGATCTTC	AAGATCCAGA  TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTTA CCCTGACGAG CGCTCCCTG GTTCGGTGTA ACTACCACAC AAACCACCGC AAATCAACT TAGTTGCCTG TCCAGATTTA GAAGCTACAG GTCAGCTCCAG GTTCCAGATTTA GAACCTACAG TCCAGATTTA GAACTACAG TCAGCATCTCAG ACCAAGTCAT TCATCATCATTAG AGCATCTTTT	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCGCTT CCTCGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC GGTCGTTCGC GGTCGTTCGC TGGACCATAA AAGTATATA AACCCCATC TCAGCAATAA TAAGTAGTAT TACACAACAA TAAGTAGTAT TACACAACAA TAAGTAGTAT TACACAACAA TAAGTAGTAT TACACAACAA TACAGTATAA AAAACGTTCT CCAGCAACTAA AAAACGTTCT CCATCACCAACAA AAAACGTTCT CTCAGCAATAA AAAACGTTCT CTCAACCAACAA CCTTTCACCA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCAGCGG TGACTCGGGG TGACTCGCTG AACATCTGAG TCCAGTGGG TGCCGCTTACC TCCAAGCTGG AAGACTGG AAGACTGG AAGACAGTA GGTTTTTTT TGAGTAAAACT GTGTAGATAA ACCAGCCAGC GCCAGTTAAT TCAAGGGGG TCAAGGGGG TCAAGGCGGG TCAAGGCGGG TCAAGGCGGG TCAAGGCGGG TCATGGTTAT GTGTATATCCGG TCGGGGCGAA GCGTTTCTGG
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACAAGTTAAC AACACAAC 2101 ACAAGTTAAC AACACAAC 2201 GCTGATTATCA TCAGTCGAC 2301 ACAACACAC 2301 ACAACACAC 2401 AAACCTGTCG TGCCAGCTGG 2501 CGCTCGGTGC TTCGGCTGC 2501 CGAACGGCA GCAAAAGGC 2701 AAGTCAGAGG TGGCGAACC 2801 GGATACCTGT CCGCCTTTC 2901 GCTGTGTGCA CGAACCCCC 3001 AGCAGCCACT GGTAACAGG 3101 TTTGGTATCT GCGCTGTGC 3201 TTTGCAAGAG GCAGATTACC 3301 AGGGATTTTG GTCATGAGA 3401 TGGTCTGACA GTTACCAAT 3501 CGGAAGGGCC GAGCGCTGGC 3601 CGGAAGGGCC GAGCGCTGTGC 3601 CGGAAGGGCC GAGCGCAGA 3701 AGTTTGCACA TTTACCATGT 3901 GGCAGCACTG CATAATTCT 4001 CGACCGAGTT GCTCTTGCC 4101 AACTCTCAAG GATCTTCCC 4201 GTGAGCAAAA ACAGGAAGG	C ACCATCACCA TTGA H H H H H H G AGAATGCAGT GAAA G GCATTCATTT TATC C CTGCAGGCAT GCAA AAGTGTAAAAG CTC ATTAATGAAT CGGC G GCGAGCGGTA TAAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ATAA C CGACAGGACT ACAC A TTAGCAGAGC GAGC A TTAGCAGAGC GAGC TATCAAAAAA GATC C CTTTAATCAGT GAGC A CCAATCTGCC CACA C GTGGTCACAC G TGTAAAAAA GGTC C TATCAAAAAA GGTC C TTATCATAAAA GGCT C TATCAAAAAA GGCT C TTATTCATCC C GGCGCAAAAA AACC C TTGTCATCC C GGCGCCAAAAA CGCC C TTGTCATCC C GGCGCCAAAAA CGCC C TTGTCTCAT GCCC C GGCGTCAATA CGGC C TGTTGAGAT CGCC C CTGTTGAGTA CCAC C TTGTGAGT CCAC C AAAATGCCGC AAAA	AGATTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC GCGACGC CTAATCATC CCAACGC GCGGGAGAG AGCTCACT CAAAGGCGGT AAGATAC CAGGCGTC AAGATAC CAGGCGTTC CTTCACC GTTGCTCATA AGAGATCC CTTCACC AGAAGTCCT CTTCACC AGAAGTCCT CTTCACC TAGATCCTT AGCACCTA TCTCAGCGAT AGTGCGGC CATCCCATCC CTTCACC TAGATCCTT AGCACCTA TCTCAGCGAT CTTCACC TAGATCCTT AGCACCTA TCTCAGCGAT CTTCACC TAGATCCTT AGCACCTA TCTCAGCGAT CTTCACC TAGATCCTT AGTTCAGC CTTCAGCCACC CTTTATCC GCCTCCATCC CATCGGACACCT CTTCACC TAGATCCTC CATCGGACACCACCACCACCACCACCACCACCACCACCACCAC	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA GCGGTTTGCG AATACGGTTA CCCCTGGAAG CTCACGCTGT TAGCTCTTT TAGCTCTTT TAGACTCTTT TAGATTATT TAAATTAAAA CTGTCTTTAT CAGAGCCCAC AGTCTATTA GTCGTTTGGT CCGATCGTTGGT TGCGTGACCCAC CGGACACCC CACGGAAACT CGTCGCACCC CACGGAAAATG	TGTGCCTTCT  ATGCTATTGC GGTTTTTAA TTTCCTGTGT CATTAATTGC TATTGGCGC TCCACAGAAT TGGGTCCGCC CTCCCTCGTG AGGTATCCAA GAAGTGGTGG TCAGGGAAAC CTAGAGGGTC ATGAAGTTTT CGTTCATCCA ATGACTCAC TGTTGCCGG ATGCCTCCGG ATGCCTCCTGT TATAAAAGTGAT TGAATACTCA TTAAAAAGTGC ACTGATCATCT TTGAATACTC TTGAATACTC	AAGATCCAGA  TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTCA TCTTCCGCTT CAGGGGATAA CCCTGACGAG CCTCACTG GTTCGGTGTA ACACACCGC CAAACTACC AAACCACCGC TAGAGCTCAG AAATCAATCT TAGTTGCCTG TCCAGAATTA GAAGCTAGAG TCAGCTCAGA TCACACTCGC ACCAAGTCAT TCATCATTGG AGCATCATTTA TCATCATTTG AGCATCTTTCT ATACTCTTCT	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTTC CCTCGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC CACGACTTAT GCTACACTAG TGGACCGAT TGGAACGAAA AAGTATTATA ACTCCCGGTC TCACCAATAA TAAGTAGTTCT TCCCCAACGA TTCCCAACGA TTCTCCAACGA TAAGTATTTT TCCCAACGA TTCTCAACAATAA AAAGTTTCT TCCCAACGA TTTTTCACCA TTTTTCAACTA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGCTG AACATGTGGG TGACTCGCTG AACATGTGAG GCGCTTACC TCCAAGCTGG AACGACGTG AACGACTGG AACGACTGG AACGACTGG CGCACTGGC AAGGACAGTA GGTTTTTTTG TGTAGATAA ACCAGCCAGG GCCAGTTAAT TCAAGGCGAG GCCAGTTAAT TCAAGGGAA GTCATGGTAT TCATGGTAAT GTGTATCGGT TCATGGTAAT GTGTATCGGT TCATGGTAAT GTGTATTCAG TCATGAGAAG GCGTTTCTGG TCAGGGCGAA GCGTTTCTGG TTATTCAAGC
1901 GACGCGTACC GGTCATCATC  TRTGHR  TRTGGHR  2001 ACAAGTTACC AACCACACC 2101 ACAAGTTACC AACACACAC 2101 ACAAGTTACC ACAACCACC 2201 GCTGATTATCA TCAGTCGAC 2301 ACATACGAGC CGGAAGCAT. 2401 AAACCTGTCG TGCCAGCTG 2501 CGCTCGGTCG TTCGGCTGCC 2601 CAAAAGGCCA GCAAAAGGCC 2701 AAGTCAGAGG TGGCGAAACC 2801 GGATACCTGT CCGCCTTTC 2901 GCTGTGTGCA CGAACCCCC 3001 ACCAGCCACT GGTAACAGG. 3101 TTTGGTATCT GCGCTCTGC 3201 TTTGCAAGCA GCAGATTACC 3301 TGGCTGTACCAGC 3401 TGGTCTGACA GTTACCAATC 3601 CGGAAGGCCC GACGCCAGA. 3701 AGTTTGCCAC ACGTTGTTG 3801 TTACATGATC CCCCATGTT 3901 GGCAGCACTG CATAATTCTT 4001 CGACCGAGTT GCTCTTGCC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACC 4101 AACTCTCAAG GATCTTACCC	C ACCATCACCA TTGA H H H H H H G AGAATGCAGT GCATTCATTT TATC C CTGCAGGCAT GCAA AGCTGTAAAG CCTC ATTAATGAAT CGGC G GCGAGCGGTA TCAC CGACAGGACT ATAA C CCCTCCGGGA AGCC G GTTCAGCCCG ACCC A TTAGCAGACCGT ACCC TATCAGCCCG ACCC TATCAGCCCG ACCC TATCAGCCCG ACCC TATCAGCCCG ACCC CGACAGGAAAAAAAG CCTTAATCAGT GAGC CACTCTGGCC CCAC CACTCTGGCC CCAC CATTGCTCAGC CGCC CGTTCAATCAG GGCA TACCAAAAAAG GATC TTACTGATCAG GGCA CCATTGCTCAG AGCC CTTACTCTACA GGCA TTACTCATCAG CGGC TTACTGATCAG CGGC TTACTGATCAG CGGC TTACTGATCAG CGGC TTACTGATCAG CGGC TTACTGATCAG CGGC CTTTGAGAT CCAC AAAATGCCGC AAAAC CATGGCGCA TACCC CATGGCGCA TACCC CATGGCGCA TACCC CATGGCGCA TACCC CATGGCGCA TACCC	AGTTTAA ACCCGCTGAT  AAAAATG CTTTATTTGT GTTTCAG GTTCAGGGGG AGCTTGG CTAATCATC CCCAACGC CCGGGGAGAG AGCTCACT CAAAGGCGT AAGGCCGC GTTGCTGCCC AAAGATAC CAGGGGTTTC AGTGGGC TTTCTCATAG AGTAGTAC CTTACTCACA AAAGATTCC AGTTCCAA AAAAGATCCT AGTTCCAA AAAAGATCCT AGTTCCAA AAAAGATCCT AGTTCCAA AATAAACCACT AATCGTAA AGAAGATCCT ATTCACGTAA AATCGTCTTCACCT CCTTCACCC CATCGTCGC CTTATCCGTCCT CTTTACCC CATCGGTCCT CTTCACCCACACACACACACACACACACACACACA	CAGCCTCGAC  GAAATTTGTG AGGTGTGGGA GTCATAGCTG AGCTAACTCA ACTCA CCCCTGGAAG CTCACGCTGT AACTCATCACCA CCAGGTTTCCTTACCTTA	TGTGCCTTCT  ATGCTATTGC GGTTTTTTAA TTTCCTGTGT TATTGGCGC TCCACAGAAT GGCTCCCTCGTG AGGTATCTCA TTGAGTCCA AGGATCTCA TTGAGTCCA ATGAGTCCA TTGAGTCCA TTGAGTCCA TCCACGGGT TCCACCGGC ATGCATCAC TTGTTCATCCA GCTCACCGGC ATGCTTCAT TCACACGGT TCACACGGT TTGATCCAC TTGATGCCA TTGATACACA TTAAAAGTGC ACTAATATC CAAATATC CAAATACTC CAAATACGG	AAGATCCAGA  TTTATTTGTA AGCAAGTAAA GAAATTGTTA GTTGCGCTC TCTCCGGTT CAGGGGATAA CCCTGACGAG GCTCTCCTG GTTGGGTGTA CCCGGTAGA CCTAACTAC CAGACTAC AAACCACCGC TGACGCTCAG TGACGTCAG AAATCAATCT TAGTTGCCTG TCCAGATTTA GAACTACAG TCAGCTCAG TCAGCTCCGG GTTGGCCGCA TCATCATTAG TCAGCTCAG TCATCATTG AGCATCTTT AGCATCTTC TCCCGCGCAC	CATGATAAGA  ACCATTATAA ACCTCTACAA TCCGCTCACA CTGCCCGCTTC CCTCGCTCAC CGCAGGAAAA TTCCGACCCT GGTCGTTCGC CACGACTTAT GCTACACTAG TGGACCGAT TGGAACGAAA AAGTATTATA ACTCCCGGTC TCACCAATAA TAAGTAGTTCT TCCCCAACGA TTCCCAACGA TTCTCCAACGA TAAGTATTTT TCCCAACGA TTCTCAACAATAA AAAGTTTCT TCCCAACGA TTTTTCACCA TTTTTCAACTA	TACATTGATG  GCTGCAATAA ATGTGGTATG ATTCCACACA ATCCAGCTG AACATGTGGG TGACTCGCTG AACATGTGAG GCGCTTACC TCCAAGCTGG AACGACGTG AACGACTGG AACGACTGG AACGACTGG CGCACTGGC AAGGACAGTA GGTTTTTTTG TGTAGATAA ACCAGCCAGG GCCAGTTAAT TCAAGGCGAG GCCAGTTAAT TCAAGGGAA GTCATGGTAT TCATGGTAAT GTGTATCGGT TCATGGTAAT GTGTATCGGT TCATGGTAAT GTGTATTCAG TCATGAGAAG GCGTTTCTGG TCAGGGCGAA GCGTTTCTGG TTATTCAAGC

## pMT-5'HisFlag-Dm43 full STOP

1 TCGCGCGTTT C	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG T	CAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
201 CCGCACAGAT G	GCGTAAGGAG	AAAATACCGC	ATCAGGCGCC	ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT
301 TACGCCAGCT G	GCGAAAGGG	GGATGTGCTG	CAAGGCGATT	AAGTTGGGTA	ACGCCAGGGT	TTTCCCAGTC	ACGACGTTGT	AAAACGACGG	CCAGTGCCAG
401 TGAATTAATT C	CGTTGCAGGA	CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC C	CGCCCACCGC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA A	AAGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT C	CGAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG A	AATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
						> M H	ннн	ннтр	Y K D
901 GACGATGACA A	AGGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
	GTS	E L N		D C W E			R L V N	DGE	N C E
1001 TCGAGGTGTT C	CTGCCGGTGC	TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
>F E V F	C R C	WRE	LOLO	H L F	T A O	TNHT	E V I	ATT	LAAL
1101 GCATGTGGCC A	AAGCGACTGT	CGTGCTCCCG	ACGCACCACC	GGGGACGTTT	TCCCGGCATC	TCGCGCTCAA	AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
> H V A	KRL	SCSR	RTT	G D V	F P A S	R A O	RIG	G F F L	L Y V
1201 ATCTACTACA A	AGCAGCCCAC	GCACAACTTT	ATTAAGATCG		GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
> I Y Y K	COPT	H N F	I K I	E V S P	R T W	O E L	T D Y A	LDL	R K D
1301 GTCCGGAGCG G									
>S P E R	KDT		A Y M L		T O E	Q A F R		L D Y	COGL
1401 GGACAATCTG G									
		D R V E		G A K			M O K	OORA	N G V
1501 AGTCTCACAT A									AAGCAATTGG
	ELE	G L R	A L D	O A S O		E L E	A A Y N	A O K	K O L
1601 CGGCTGGTCA T									~
>A A G H	E H A		S O I F			F A D I		L G A	R K S T
1701 TCCAGATGAG A									
		T T S T		L E V				G V E E	R E P
1801 CAACACCAGA C									
	r D E L		L E V				A T V F	O R E	L P E
1901 ACGTGCAGCA A									
>D V O O							V T E		K A I L
>D V Q Q 2001 GGATACTTGA G	E Y E	M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
	E Y E	M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA G > D T	E Y E GTTAGTTTAT	M I E AAAACTTTTA	F S D D CATAATTAAA	E E M TAACTAGCAT	E V G TTTTGCGCGA	E S E E TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
2001 GGATACTTGA G > D T 2101 AGCACAGTGG C	E Y E GTTAGTTTAT	M I E AAAACTTTTA AGTCTAGAGG	F S D D CATAATTAAA GCCCGCGGTT	E E M TAACTAGCAT CGAAGGTAAG	E V G TTTTGCGCGA CCTATCCCTA	E S E E TGTGATCTTG	V T E TTTATCTGAA CGGTCTCGAT	E E L GGGCAATTCT TCTACGCGTA	K A I L GCAGATATCC
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C  2201 TCACCATCAC C	E Y E GTTAGTTTAT CGGCCGCTCG CATTGAGTTT	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C  2201 TCACCATCAC C  2301 CTAGAATGCA G	E Y E GTTAGTTTAT CGGCCGCTCG CATTGAGTTT GTGAAAAAAA	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T	E Y E GTTAGTTTAT CGGCCGCTCG CATTGAGTTT GTGAAAAAA FTTATGTTTC	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C  2201 TCACCATCAC C  2301 CTAGAATGCA G	E Y E GTTAGTTTAT CGGCCGCTCG CATTGAGTTT GTGAAAAAA PTTATGTTTC ATGCAAGCTT	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A	E Y E  GTTAGTTTAT  CGGCCGCTCG  CATTGAGTTT  GTGAAAAAA  PTTATGTTTC  ATGCAAGCTT  AGCCTGGGGT	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A	E Y E GTTAGTTTAT  CGGCCGCTCG CATTGAGTTT GTTAGAAAAA PTTATGTTTC ATGCAAGCTT AGCCTGGGGT ATCGGCCAAC	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA G  D T  1010 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGACCGG T	E Y E ETTAGTTTAT  CGGCCGCTCG CATTGAGTTT ETGAAAAAA FTTATGTTTC ATGCAAGCTT ACCCTGGGGT ATCGGCCAAC FATCAGCTCA	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGAG CTCAAAGGCG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TAGGCGGTTTG GTAATACGGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T	E Y E ETTAGTTTAT  CGGCCGCTCG CATTGAGTTT  FTGAAAAAAA  PTTATGTTC AGCCTGGGGT AGCCTGGGGAC  ATCGGCCAAC  TATCAGCTCA  TAAAAAGGCC	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CTCAAAGGCG GCGTTGCTGG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT GTAATACGGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC ACGCAGGAA AGCATCACAA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CGACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAAAAGGC TCAAGTCAGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAAA
2001 GGATACTTGA G  D T  101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT STGAAAAAA STTATGCAAGCTTA AGCCTGGGT ATCGGCCAAC TATCAGCTCA TAAAAAGGCC CTATAAAAGAT	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG ACCAGGGGTT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCATAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA	TAACTAGCAT  CGAAGGTTAT  GAGGTTTTTT  GAGATTAAT  CGTATTCCTGT  CACATTAATT  CGTATCCCACAT  TATCCCACAG  AGCTCCTCCC  AGCTCCTCCC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCCTGACG TGCGCTCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTCA ACGCAGGAA TGTTCCGACCA TGTTCCGACCA	T T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATCCACA CTTCCACA CTGCACTCGC ACTACACTCGC ACTACACTCGC ACTACCCCCC ACTACCCCCCCCCC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCA	K A I L GCAGATATCC  CCGGTCATCA CAAAACAACA ACAACAACA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAG GGTGGCGAAA
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA 3101 CTCCCTTCGG G	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT ETGAAAAAA FITATGTTTC AGCCTAGGGT ATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTCA FATCAGCTGC FATCAGGTGGC	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTATT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTAGC TGAGCTAACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCCGCTGGA AGCTCACGCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TAGCCCCCG GTAGGTATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GCGATAGCG GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAC CACTGCCCGC CACACGGCAGA AGCATCACAA AGCATCACAA TGTTCCGACC TAGGTCGTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACA TTTCCAGTCG AGAACATGGTA AAATGGTGGAACATGGC CTGCCGGCTGAAGCT GCTCCAGCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CAGCAAAAGG GGTGGCGAAA GCCGCCTTT CACGACCCC
2001 GGATACTTGA G  D T  2101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2701 GCAGGAGCG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG C 3201 CCGTTCAGC C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAAA  ATCGCCAAC TATCAAGCTCA  TAAAAAGGCC TATAAAAGT TAAACGTGGGTGGC TATAAAAGT TAAACGTGGCTGGCT TAAAAAGGT TAAAAAGGT TAAAAAGGT TAAAAAGGT TAAAAAGGT TAAAAAGGT TAACCGTGGCT TATAAAAGGT	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCCGGGGAG CTCAAAGGCG CCTTAGTCTGCGG ACCAGGGGTT GCCTTTCTCAT GCTTTCTCAT	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCAATACG TGACCTAACT AGCGGGTTTG GTAAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG	E E M TAACTAGCAT  CGAAGGTAAC ACTGTGCCTT TGATGCTATT GAGGTTATT TGTTTCCTGT CACATTAATT CGTATTCACAG TATCCACAG AGGTCCCTCG GTAGGTATCTTTTTTTTTT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTACGC CAGTTCCGTAAC AACCCGGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGACC AACGCAGCAA TGTTCCGACC TAGGTCGTCT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ACTACATCGC AGAACATGTC AGACACTCGC AGACACTCAAGCT ATCCCACACTCAAGCT ATCCCCACTCAACCT ATCCCCACTCAACCT ATCCCCACTCAACCT	E L GGGCAATTCT TCTAGGCGTA TGAGTTTTGGA ANACAAGTTA TGGCTGATTA CAACATACGA AGGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTTGTG GGGCTCGTTGTG GCACCAGCCA	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA 3101 CTCCCTTCGG G	E Y E STTAGTTTAT  CGGCCGCTCG LATTGAGTTT  TTGAAAAAA  FTTATGTTTC ATGCGAAGCTT  ATCGGCCAAC FATCAGCTCA FATAAAAGGCC CTATAAAAGAT  LAGGCTGGC CGACCGCTGC CGCACGCTGC CGCACGCTGC CGCACGCTGC CGCACGACTATC	M I E AAAACTTTA AGTCTAGAGG AAACCCCCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGTTGCTG ACCAGGGGTG GCCTTTCTCAT GCCTTATCCTG GCCTTATCCAT GCCTTATCCAT GCCTTATCCAT GCCTTATCCG TAGGCGGTG CCTAGGCGGTG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGGCTAATC AGGCGGTTTG GTAATACGGT TCGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTACG	$\begin{tabular}{ll} E & E & M \\ TAACTAGCAT \\ CGAAGGTAAG \\ ACTGTGCCTT \\ TGATGCTATT \\ GAGGTTTTT \\ GAGGTTTTT \\ GACATTAATT \\ CGCATTAATT \\ CGTATTGGC \\ AGACTCCTCG \\ GAGGTCCTCG \\ GTAGGTATCT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGGT \\ CTGTAGTGT \\ CTGTAG$	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGGCTAACCAGAA TGTTCCGACC TAGGTCGTTC GACACGACCTTC GACACGACCTTC GACACGACCTTC GGCTACACTTC	V T E TTTATCTGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATGACATGA ACATGACTGACAGACATGTG ACATGACATG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA GGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	K A I L GCAGATATCC  CCGGTCATCA CAAAACAACA AACAACAA TGATCAGTCG GCCGGAAGCA CGTGCGAGCT CGGTCGGCTG CAGCAAAAGG GCTCGGCTTT CACGAACCC CTGGTAACAG CTGCGCTTG
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA 3101 CTCCCTTCGG G 3201 CGTTCAGCC C 3301 GATTAGCAGA G 3401 CTGAAGCAG T	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  ETGAAAAAA  FTTATGTTTC ATGCAAGCTT  ATCGGCCAAC FATCAGCTCA FATAAAAAGCC TATAAAAGGC CTATAAAAGT CAACCCTGC CGCACCCTGC CGCACCTTCC CGCACCTTC C	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCCAATGAG GCCCTAATGAG CTCAAAGGCG GCGTTGCTGG ACCAGGGGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAAGATT	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTAACT AGGCGGTTTTG GTAATACGGT CGTTTTTTCCA AGCTCGTGAACTACC GTAACTATCG GTAACTACG TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAGCTCTT GGTAGCTCTT	E E M TAACTAGCAT  CGAAGGTTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT CGTATTGGGC TATCGCCTCG GTAGGTTCTTCT TCTTGATCCTTCG GTAGGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAACCAAGTA GCGATCGGCT GCTCTCCGC ATCAGGGGAT CCCCCTGACG ACCCGGTAA GGCCTAACTA ACAAACCACC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATCAC CACTGCCCGC CACACGACTAC AAGCATCACAA AGCATCACAA AGCATCACAA AGCATCACAA CACACGACTT CGACACGACT CGCTACACCT CGCTACACCT CGCTTGCTACCCT CGCTTGCTACCCT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACA TTTCCAGTCG AAATCGCCACA TTTCCAGTCG AAATCGCCGCTTA GCTCCAGCGCTTA GCTCCAGCGCTTA GCTCCAGGCAGACACGGCTTA GCTCCAGGCGCTTA AGAAGGACAGGCAGAGGACAG GTGGTTTTTT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCGCTGTGTG GCACAGCCA TATTTGGTATT TGTTTGCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTTG
2001 GGATACTTGA G  D T  101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3301 GATTAGCAGA C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TGGAAAAAA  TTATGTTC AGCCTGGGGT ATGGAGCTCA  TATCAGCCAAC  TATAAAAGC  TATAAAAGC  TATAAAAGC  TATAAAAGT  TAACACCTGC  CGACCCTGC  CGACCCTCC  CGAACGTTCG  AAAAAGGATTC	M I E AAAACTTTA AGTCTAGAGG AAACCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCCGGGGAG CTCAAAGGCG CCTTAGTCTGCTG ACCAGGGGTT GCTTTCTCAT GCTTTCTCAT GCTTATCCG AAAAAGAGTT	F S D D CATAATTAAA  GCCGCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGACCTAACT AGCGGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT TACAGAGTTC GTAACTATCG TACAGAGTTC GGTAGCTCT TCCAGAGTTC TCTTGATCT CTTTTGATCTT CTTTTGATCTT	E E M TAACTAGCAT  CGAAGGTAAC ACTGTGCCTT TGATGCTATT GAGGTTATT TGTTTCCTGT CACATTAATT CGTATTCCACAGA TAGGCTCCGC AGCTCCTCG GTAGGTATCTTTCTTGAGTAGTCC TCTGAAGTGGT GATCCGCAGAGTTCTTGAGTCC TTGAAGTGGT GATCCGGCAA TTCTTACGGGG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTACGT AACCCGCTAACG GCCTAACG GCCTAACA GCCTAACA GCCTAACA CCCCGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGACC TACCACGACC TAGGTCGTCCACC CAGGAAC GGACACGACT CGGCTACACT AGGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG	V T E TTTATCTGAA  CGGTCTCGAT  GATACATTGA  AAAGCTGCAAT  AAATTGGTA  CAATTCCACA  ACTACTCGC  ACTACTCGC  AGAACATGTC  AGACACTGC  AGACACTGC  AGACACTGC  AGACACTGC  AGACCACTG  AGAGCACACTG  AGAGCACACTG  AGAGCACACTG  AGAGCACACG  AGACTCACACGT  AGAACTCACACGT  AGAACTCACACGT  AGAACTCACACGT  AGAACTCACACGT	E L GGGCAATTCT TCTAGGCGTA TGAGTTTTGGA ANACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGAATACCT GGGCTCGGT GGGCTCGGT GGGCTCGGT GGACAGCA TATTTGGTAT TGTTTGCAAG TAAAGGGATTC	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTCGCAGCT CAGCAAAAG GTCCGCCTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGAGCAGATTA TGGTCATGAG
2001 GGATACTTGA G  D T  101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3201 CGTTCAGCC C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3501 CGCGCAGAAA A 3601 ATTATCAAAA A	E Y E STTAGTTTAT  CGGCCGCTCG LATTGAGTTT  TTGGAAAAAA  ITTATGTTC AGCCTGGGGT ATCGGCCAAC  TATCAGGTCA  TATAAGGCC TATAAAGGC  CTATAAAGGC  CTATAAAGGT  CGACCGCTGC  CGACCGCTGC  CGACGGTATG  TTACTTCGG  AAAAGGATCTTCA	M I E AAAACTTTA AGTCTAGAGG AAACCCCCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGTGGGGA CTCAAAGGCG ACCAGGCGTT GCTTTCTCAT GCCTTATCCA GCCTTATCCA TGCCTGTATCA TGCCTGTTCTCAT GCCTTATCCA TAGGCGGTG AAAAGAGTT CAAGAAGATC CCTAGATCCT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTATCT TTTAAATTAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCTTT TGATGCTATT GAGGTTTTT TGATTCCACAGA TAGCACACAGA TAGCCACAGA TAGGCTCCTCG GTAGGTATTCTTTCTTGAGTCCTCTG GTAGTTGTGAGTCGT AGTCGCAAATTCAGAGAAATGAAGTT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTATTTG AAAGCAGTA GCGAAATTGT GCGTTGCGC TCCCCCTGACG TCCGCTTCCCC AGTTCGGTA ACCAGGAT ACAAACCAC TCTGACG TCTGACG TCTTCGCTACT TCTGACACT TCTGACCTC TTTAAATCAAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCGCGCAA AGCATCACAA TGTTCCGACC TCGGCTACACAC GACAGCACTC CTAAAGTATA	V T E TTTATCTGAT GATACATTGA AAAGCTGCAT AAATGTGGTA TTTCCAGTG ACTGACATGA ACATGACTGACTG ACTGACATGA ACATGACTGACAG AGAAGAGTGTTTT ACACCACAG ACAGAGAGACAGTGACAGAAGAGACATTTT ACACCACAGAGAGACAGACAGTGACAGAGAGACAGACAGTGACAGAGAGACAGAC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTTGCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAAACAACA AACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTCGGCTG CAGCAAAAGG GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA
2001 GGATACTTGA G  D T  1010 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2701 GCATTAATGA A 2801 CGGCGACGGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CGGTTCAGCC C 3301 GATTAGCAGA G 3401 CTGAAGCCAG T 3501 CGCGCAGAAA A	E Y E STTAGTTTAT  CGGCCGCTCG LATTGAGTTT  ETGAAAAAA  ITTATGTTTC ATGCAAGCTT ATCGGCCAAC LATCAGCTCA LATCAGCTCA CAGAGCTGC CAGACGCTGC CAGACGCTGC CAGACGCTGC LACAGCTCAC AGAGCATCT CAGAGCATCT CAGAGCACC CGGAGGTATG CTACAGCACC CTACAGCACC CTACAGCACC CAGACCACC CACACCAC CAGACCACC CACACCAC CACACCAC CACACCAC CACACCAC	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCGGGGGAG CTCAAAGGCG GCGTTECTGG GCCTTATCTG GCCTTATCCAT GCCTTATCCAT GCCTTATCCAT GCCTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGATC CCTAGATCT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTAACT AGGCGGTTTTTCCA AGGCGGTTTTTCCA AGGCGGTTTTTCCA AGGCTACCT GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAGGTTC GGTAGCTCT TTTTAAATTAA ATCTGTCTAT	$\begin{tabular}{llll} E & E & M \\ TAACTAGCAT \\ CGAAGGTTAGCTT \\ TGATGCTATT \\ GAGGTTTTTT \\ GAGATAAATT \\ GAGGTTCTGT \\ CACATAATT \\ GAGTCCCTGG \\ TAGGCTCCTGG \\ GAGGTTCTTGAAGGT \\ TCTTGAAGTGGT \\ GATCCGGGAAATTACGGGGAAATTCATTCATCCACAAGAGTACTTCATCATCTCATTCCATCC \\ AAATGAAGTAGAAGTTCCATCATTCATCATCATTCATTCA$	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGAAAGTA GCGTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGC AACCCGGTAA ACCACC CTCAACCA TCAAAACCACC TCTGACGCT TTAAAATCAAT CATAGTTGCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATCAC CACTGCCCGC CACACGAGAA AGCATCACCA AGCATCACCA TGTTCCGGTC GACACGACTT GACACGACTT GCTGTACGAC GCTGGTAGCG AGTGGAACGA CTAAAGTATA TGACTCCCCC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACA TTTCCAGTCG ACTGACATTGA AAATGTGAAACATGGC AGAACATGTG AAATCGACGC TTA GCTCCAAGCT ATCGCCATG AGAAGACATGTT ATGACGC TTAAACTCACGT TAAACTCACGT TATGAGTAAGAT TATGAGTAAGAT TATGAGTAAGAT TATGAGTAAGAT TATGAGTAAGAT TATGAGTAAGAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGCCTG TCAAGTCAGA CCGGATAACT GCAGCAGCCA TCAAGTCAGA TATTTGGTAT TGTTTGCAAG TAGTGGTATG CTTTGGTAT CTTTGGTAT CTTTGGTAT CTTTGGTCTGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAASCA CGTGCCAGCT CGGTCGGCGAAA GCCCCCTGGCCCTTT CACGAACCCC CTGGTAACAG CTGCCGCTTT CACGAACCCC CTGGTAACAG CTGCCGCTTT CACGAACCCC CTGGTAACAG CTGCCGCTTT CACGAACCCC CTGGTAACAG CAGCTTACCAA CGGGAGGGCCT CGGGGCAATTA CGGGAGGGCCT
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2801 CGGCGACAGGA C 2901 CCAGGAACCG T 3001 CCCGTCAGCC C 3201 CCGTCAGCC C 3201 GATTAGCAGA A 3401 CTGAAGCAG T 3501 CGGGCAGAAA A 3601 ATTATCAAAA A 3701 TGGTTAATCA C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TGGAAAAAAA  TTATGTTTC AGCCTGGGGT ATGGAAGCTT AACAGCTGAC TATAAAGAT EAAAAAGGC CTATAAAGAT EAAAAAGGT GGACGCTGC GGAGGTATG FTACCTTCG AAAAGGATCTT AGGATCTTCA STGAGGCACC CCCCACTGCT CCCCACTGCT	M I E AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCCGGGGAG CTCAAAGGCG CTCTAATGAGG ACCAGGGGTT GCTTTCTCAT GCTTTCTCAT CCTAGAGAGATC CAAGAAGAGTC CAAGAAGATC CTAAGATATAC	F S D D CATAATTAAA  GCCCGCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGTCATACT AGCGGGTTTC GTAATACGGT TCGTTTTTCCA TCCCCCTGGA AGCTCACGCT TACAGAGTTC GGTAACTATCG TTAACTATCG TTAACTATCG TTTAAATTAA ATCTGTCTTT TTAAATTAA TCCGCGAGACCC	E E M TAACTAGCAT  CGAAGGTTAG ACTGTGCCTT TGATGCTATT GAGGTTATT TGTTTCCTGT CACATTAATT CGTATTCCCGC AGCTCCCGC GTAGGTACTCTT TCTTGAGTCCTT TGAAGTAGT TCTTGAGTCC TCTGAAGTAGT GATCCGGCAA AATGAAGTT TTCTTCGGGG AAATGAAGTT TTCTACGGGG AAATGAAGTT TTCGTCACCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCGGC ATCAGGGGAT GCGCTTACGT AACCAGGTAAC GCCTAACT GCGCTAACT GCGCTAACT TGCAGTACACAC TCTGACGTAACACACC TCTGACGGTAA CAAACCACC TCTGACGGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCAACCACC AACGCAGCAA TGTTCCGACC TAGGTCGTTCCGACC TAGGTCGTTC GGCTACACT GGCTACACT AGGTGGAACG AGTGGAACG AGTGGAACG CTAAAGTATA TGATCCCCCG TATCCGCA	V T E TTTATCTGAA  CGGTCTCGAT  GATACATTGA  AAAGCTGCAAT  AAATTGGTA  CAATTCCACA  ACTACTCGC  ACTACTCGC  AGAACATGTC  AGAACATGTC  AGACACTGC  AGACACACTG  AGACACACTG  AGAACTCACACTT  AGAAGGACAC  GTGGTTTTTC  AGAACTCACGT  TATGAGTAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAAAA  TCGTGTAGAGCAC	E L GGGCAATTCT TCTAGGCGTA TGAGTTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGGT GGACCAGCA TATTTGCTAG TGTTTGCAAG TAAAGGGATTT CTTGGTCTGA	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTG CAGCAAAAG GTCCGCCTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGCAGATTA CGGGAGGCC CCGGAGGGCCT CGGAGGGCCC CCGAGAGAGAC CGGGAGGGCCT CGGAGGCGCAGAC CCGAGCGCAGAC CCGAGACCCC
2001 GGATACTTGA G  D T  1010 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2701 GCATTAATGA A 2801 CGGCGACGG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3301 GATTAGCAGA C 3501 GATTACCAGA A 3601 ATTATCAAAA A 3601 ATTATCAAAA A 3701 TGCTTAATCA G 3801 TACCATCTGG G	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAA  FTTATGTTTC ATGGGGGT ATCGGCCAAC FATCAGCTCA FAAAAAGGC CTATAAAAGGC CTATAAAAGGT CGCACGCTGC CGCAGGTATG FTACCTTCGG AAAAGGATCTTCA STGAGGCACC CCCAGGTGCT CCCCAGTGCT CCCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCAACTTTAT	M I E AAAACTTTA AGTCTAGAGG AAACCCCCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCGCGGGAG CTCAAAGGCG TCCAAAGGCG TGCTTTCTCAT GCTTTCTCAT GCTTTTCTCAT CAGAGAGATC CAAGAAGATC CAAGAAGATC CAAGAAGATC CATAGAGATC TATCTCAGCG GCATTCAGCG CCAAGAAGAC CCGCCTCCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGGCTAATC AGGCGGTTTG GTAATACGGT TGTATATCGA TCCCCCTGGA AGCTCACGCT TACAGAGTTC TACAGAGTTC TACAGAGTTC TTTTAAATTAA ATCTGTCTTAT CGCGAGACCC CCAGTCTATT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTATT GATGCTATT GAGGTTTTT TGATGCTATT GATGCTATT GAGGTTCGGC TATCCACAGA TAGGGTCCGCAG TAGGGTCCGCAG TAGGGTATT TCTTGAGTGGT GATCCGCAA TTCTAGAGTGT TCTTGAAGTTGT TCTTCAGGG AAATGAAGTT TTCGTTCACG AAATGAAGTT TCGTTCACC AAATGAAGTT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTGACG TCCCCCTGACG TCCCCCTGACG TACCGGTAA GCCTAACTA ACAAACCAC TCTGACGCTA CTAAATCAAC CTAAGTTGCG GCTCTCCC CGTTACTACTA CAAACCAC TCTGACGCTC CTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAA TGTTCCGACC TAGGTAGACA TGTTCCGACC GACAGGACTT CGGCTACACT GCTGGTACGC CTAAAGTAATT TGACTCCCCC TAGGACGA TGTAAGTAAGTAATT TGACTCCCCC TAGTAGCAT	V T E TTTATCTGAT GATACATTGA AAAGCTGCAT AAATGTGGTA TTTCCACT ACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA GGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GGAGCAGCCA TATTTGGTAT TGTTTTGCAAG TAAGGGATTC TTTGCAAG TAAGGATTACGATA GCCGGAAGCCA AATTACGATA GCCGGAAGCA AATTACGATA GCCGGAAGGA AATTACGATA	K A I L GCAGATATCC  CCGGTCATCA CAAAACAACAA ACAACAACA TGATCAGTCG GCCGGAAACG CGTCGGCTG CAGCAAAAGG CTCGGCTTT CACGAACCC CTGGTAACAG CTGCGCTCT CAGCACATTA CAGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA CGGAGTCATG CAGCAGATTA CCGAGCGCAA CGGGAGGGCT CCGAGCAGCAA CGGGAGGCCCAG CAACGTTGTT
2001 GGATACTTGA G  D T  101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2801 CGGCGAGCGG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3401 CTGATAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 TAGCTAAAA A 3701 TGCTTAATCA G 3801 TACCATCTGG C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  ETGAAAAAA  TTTATGTTTC ATGCAAGCTT ATCGGCCAAC TATAAGGCT CATATAAGGT CATATAAGGT CATATAAGGT CATATAAGGT CATATAAGGT CAGAGGTATG FTACCTTCG AAAAGGATCT AGGATCTTCA STGAGGCACC CCCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CCCAGTGCT CAGACTTTAT CAGGATCTT CAGCATCGT CAGACTTTAT CAGGATCTT CAGCATCGT	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCGGGGGAG CTCAAAGGCG GCTTATCAT GCTTTATCAT GCTTATCAT GCTTATCCAT GCTTATCCT TAGGCGGTGC AAAAGAGAT CAAAAGAGTT CAAGAAGATC CCTAGATCAT CAATGATCA GCTCTCAAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTAACT AGGCGGTTTTCCA AGGCGGTTTTCCA AGGCGGTTTC GTAATACCGC GTAACTATCC GTAACTATCC GTAGCTACT TTCAGAGGTTC GTAGCTCT TTTTAAATTAA ATCTGTCTAT CCCAGTCTTTT CCAGGAGACCC CCAGTCTTTT TCGTCGTTTT TCGTCGTTTT TCGTCGTTTT	E E M TAACTAGCAT  CGAAGGTTAAG ACTGTGCCTT TGATGCTATTTT TGTTTCCTGT CACATAATT GAGGCTCCTCG GTAGGTACTCACAGA TAGGCTCCTCG GTAGGTACTTGATCCT GTAGGTGGT GATCCGCAA TTCTACGGC AAATGAAGTT TCGTTCTCTCTCTCTCTTCACCCGCAAATGAAGTT TCGTTCAAGTTCCCCTCCG AAATGAAGTT TCGTTCCCCTCCCGTACGCTCACCCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GAAACTAG GCGTTCGCGT GCTCTCCGC ATCAGGGAT CCCCCTGACG GCCCCCTGACG ACCCCGCTAA ACCACC TCTGACGTC ACCACCC TCTGACGTC TCTAACTAACTAACAA CATAGTTGCC GCTCCAGATT GGGAAGCTAC GCGCAATT GGGAAGCTAC GCGCAATT GGGAAGCTAC ATTCAGCTCC ACTTCCAGATT GGGAAGCTAC ACTAGTTGCC GCTCCAGATT GGGAAGCTAC ACTAGTCCCCACATT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAC CACTGCCCGC CACCGGTCA ACGCAGGAA AGCATCACAC AGGATCACAC AGGATCACAC CAGGTACCAC GCTGGTACC GCTGGTACCC TAGGTACCC TAGATATAT TGACTCCCCG TATCAGCAAT TGACTCCCCG TATCAGCAAT TAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACGACCTCGCCGCTTA GCTCCAGCTTTT GCTCCAGCTTTTTTTAAACTCACGC GTGGTTTTTTTTAAACTCACGCT TAGAGAGACAGCAGTAAA TCGTGTAGAT AAACCAGCCA TCGCCGCTTA GATCACGCCA TTGCTAGATAAA TCGTGTAGAT AAACCAGCCA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCCTGATTA CAACATACGA GGAAAACGTC TCAAGTCAGA GCCAAAAAGGC TCAAGTCAGA TACTTGGTATT GCTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTATGAAGAACTGGT AACTACGAAAAACAACAAACAAAACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTCAGCT CAGCAAAAGG GTCGCAAAA GTCCGCTTT CACGAACCC CTGGTAACAG CTGCGCATT CAGCAAACG CTGCGCTTT CAGGAACCC CTGGTAACAG CTGCGCTTT CAGGAACCC CTGGTAACAG CTGCGCTTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGCCT CCGAGCGCAG CAACGTTGTT TCCCCCATGT
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2801 CGGCAGAGCG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CGGTCAGGC G 3401 CTGCAGCAGA C 3401 CTGAAGCCAG T 3501 CGGGCAGAAA A 3701 TGCTTAAACA A 3701 TGCTTAAACA A 3701 TGCTTAAACA C 3801 TACCATCTGG C 3901 AAGTGGTCCT C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAAA FTTATGTTTC AGCCTGGGGT ATCGGCCAAC TATCAGCTCA FAAAAGGCC CTATAAAGAT SAAGCGTGCC CGAGCGCTGC CGAGCGTTCGG AAAAGGATCTTCA STGAGCACC CCCCACTGCT CCCACTGCT CAGCATCTTCA AGATCTTCA AGATCTTCA AGATCTTCA AGATCTTCA AGACCTTCA AGACCTTTCA AGACCTTTCA CCCCACTGCT AGACCTTTAT AGACCTTTAT CAGGCATCTTAT AGGACTTTAT CAGGCATCTTAT AGGACTTTAT	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGGTAATCA GCGCGGGGAG CTCAAAGGCG CCTTATCTG ACCAGGCGTT GCTTTCTCAT GCTTTCTCAT CCTAGAGAGAGT CAAGAAGAGT CCTAGATACCT TATCTCAGCG CCAAGAAGAGT CCTAGATACCT TATCTCAGCG CCAATGATACCT TATCTCAGCG CCAATGATACCT CCGCTCCAT GGTGTCACGC TCCTTCAGGT CCCTTCAGGT CCCTTCAGT CCCTTCAGGT CCCTTCAGT CCCTTCAGGT CCCTTCAGT CCCTTCAGGT CCCTTCAGT CCCTTCA	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGACTAACT AGCGGGTTTC GTAATACGGT TCGTCATAGC TCGTCATACT TCACAGGTTC GTAACTATCG TAACAGAGTTC GGTAGCTCT TTTAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT CTGCGGTTCTTT CTCGTCGTTTC CTCCGATCGTTC	E E M TAACTAGCAT  CGAAGGTTAG ACTGTGCCTT  GAGGTTTTTT GATGCTATT GATTCCTGT CACATTAATT CGTATTCCCGC AGCTCCCCG AGGTTCCTCG GTAGGTATCTTCTTGAGTCCTTCAGTGCTACT TCTTGAGTTCCTCGGCAAGTTCATCAGGCTACGGCAAGTTCATCACGGGAAGTTCATCACGGCTACGCTACGGCTACGGCTACGGCTTCACGGCTACGGCTTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGGCTCACGATTGTTCCCTGTCAGAAGTT TGTCAGAAGTCTCTGTCAGAAGTCTCTGTCAGAAGTCACGAAGTCTCAGAAGTCTCTGTCAGAAGTCACGAATTGTTCACGTCAGAAGTCTTCATGCACGAAGTCACACGACTACGAACTTCATGACGACTACGACTACACACAC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT ACCCCCTAACG TGCGCTTACGT AACCACGTAACA GCCTAACTA GCCTAACTA GCCTAACTA GCCTAACTA CACAGTTCGTC TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AACTTGCCC AACTTGGCCC AACTTGGCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGATCACAA GGTTCCGACC TAGGTGAACC TAGGTGAACC TAGGTCGTTC GGCTACACT GGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG TAGACTCCCCG TATCACACA TGATCCCCAC CTAGACCATATATA TGATCCCCAC CAGTGTTATCC CACC CA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCACA ACTGACTGG ACTGACTGG AGAACATGTG ACTGACTGG AGAACACTGTA GAACACACTTTT AAACCACGCA TAGAGGC TACGCCAGTTA AAACCACGCA TAGAACGCA TCGCCAGTTA AAACTCACGTTA AAACTCACGCA TCGCCAGTTA AAACTCATGGTTA AAACTCATGGTTA AAACTCATGGTTA AAACTCATGGTTA AAACTCATGGTTA AACTCATGGTTA AACTCATGGTTA AACTCATGGTT	E L GGGCAATTCT TCTAGGCGTA TGAGTTTTGGA AAACAAGTTA GGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTCGGT AGCAACCCA TCTATGGTATT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA AACTACGATA AACTACGATA ATGCCAGCAA ATGCCAGCAA	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACA TGATCAGTCG GCCGGAAGCC CGTGCCAGCT CGTTCGGCTG CAGCAAAAG GTCGCCTT CACGAACCC CTGGTAACAG CTGCGCTTT CACGAACCCC CTGGTAACAG CTGCGCTTT CACGAGATTA CGGAGATTA CGGGAGGCCT CAGGAGGCCAG CAACGTTGTT TCCCCCCTGTTTTT TCCCCCATGT TCCCCCATGT TGCATAAATTC
2001 GGATACTTGA G  D T  1011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GCATTAATGA A 2701 GCATTAATGA A 2801 CGCGGACGGG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3301 GATTAGCAGA A 3601 ATTATCAAAA A 3601 ATTATCAAAA A 3601 ATTACTAAAA C 3801 TACCATCTGG C 3901 AAGTGGTCTT G 4001 GCCATTGCTA C 4101 TGTGCAAAAA A	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAAA PTTATGTTTC ATGCCAAC TATCAGCTCA ATCAGCTCA ATCAGCTCA ATCAGCTCA ATCAGCTCGC CTATAAAGAT ATAAAAGGC CTATAAAGAT AAAAAGGATCT AGGATCTTCA ATGATTCA ATGA	M I E AAAACTTTA AGTCTAGAGG AAACCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCGCGGGAG CTCAAAGGCG ACCAGGCGTTCCGC TAGGCGGTTCCCTAGGCGTTCCCTAGACAGATC CCTAGATCCC TATCTCAGCG GCATGATACC TATCTCAGCG GCATGATACC TATCTCAGCG GCATGATACC TATCTCAGCG TAGGTCCT TATCTCAGCG TAGGTCCTTCAGCG TAGGTCCTTCAGCG TAGGTCCTCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGCTATAGC TGGCTATAGC TGAGCTAATCG GTAATACGGT GTAATACGGT TGACTAATCG TACACAGAGTTC TACAGAGTTC TACAGAGTTC TTTAAATTAA ATCTGTCTAT TCGCCAGACCC CCAGTCTATT TCGTCGTTTG	E E M TAACTAGCAT  CGAAGGTTAT GAGGCTATT GAGGCTATT GAGGCTATT GAGGCTATT GATGCTATT GAGGTCGA TAGCATCAGA AGAGTAGGTACGG AGAGTAGT TCTTGAAGTCC TCTGAAGTCGGGA AAATGAGTCACGGGA AAATGAGTTCACGGGG AAATGAGTTCTCCGCAAGT TTCGTCAGGGGAAGT TGCAGAAGT GGTGAAGTACT GGAAGT GGTGAAGTACT GGAAGT GGTGAAGTACT GGAGGGAAGT GGTGAAGTACT GGAGGGAAGT GGTGAAGTACT TGGTGAGAAGT GGTGAAGTACT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTACGG ATCAGGGGAT ACCCGGTAA GGCCTAACTA GCCAACCAGCT CTGAACGCT TCTGACGCT TTAAATCAAT GCAACCAGCT GCTCAGCGCT AAATCAGT GCTCAGAGT ACCAGGTA AGCAGCT AAATCAAC ACAAGTTGCC AACTAGATTGCC CAACCAAGTC AAGTTGCCC CAACCAAGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAA TGTTCCGACC TAGCTGGTCT ACCCAGCAA TGTTCCGACC TAGGTGTACCG CTAAAGTATA TGACTCCCCAC TTGTTCCCAAC CAGTGTAGCA AGTAGTAGTAGT AGTAGTAGT AGTTAGTAGT AGTTAGTAGT	V T E TTTATCTGAT GATACATTGA AAGCTGCAT AAATTCCACA ATTTCCACA ATTTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACATGA AAATCACACA ATTTCCACA ACTGAC ACTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA GGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GGAGCAGCA TATTTGCAAG TAAGGGATT CTTGGTCTGA AACTACGATA CATGTTTGCAAG AACTACGATA CATGTTTGCAAG AACTACGATA AATTTGCAAGAGC AATTTGCAAGAGC AATTTGCAAGAGG AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGACAC AGGCACCGAC GGCGACCGAC	K A I L GCAGATATCC  CCGGTCATCA CAAAACAAAA TGATCAGTCG GCCGGAAGCA CGTGCGAGCT CGTCGGCTG CAGCAAAAGG GCTGCGAAA GTCGGCTTT CACGAACCC CTGGTAAACAG CTGCGCTCTG CAGCAGATTA CAGAACCCC CTGGTAACAG CTGCGCTCTT CAGGAGCCAG CAGCAGATTA CGGAGCAGT CCGAGCGCAG CAACGTTGTT TCCCCCATGT TCCCCCATGT TGCATAATTC TTGCTCTTGC
2001 GGATACTTGA G  D T  101 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA C 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2601 CAGAGACCG T 2901 CCAGGAACCG T 3001 CCCGACAGGA C 3101 CTCCCTTCGG G 3201 CCGTTCAGCC C 3201 CGTTCAGCC C 3101 GATTAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAGCAGA C 3401 CTGAAAAA A 3701 TGCTTAATCA G 3801 TACCATCTGG C 4001 GCCATTGCTC C 4001 GCCATTGCTC C 4101 TGTGCAAAAA A 4201 TCTTACTGTC A	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAAAAAA  FTTATGTTTC ATGCCTGGGT ATGCGCAAC  TATCAGGCAAC  TATAAAAGGC  TATAAAAGGC  TATAAAAGTT  TAAGCTGGC  CGACCGCTGC  CGACCGCTGC  CGACCGCTGC  CGACCGTTACTTCAG  TAACTTCAG  TAAGCATCTTCAG  TAGAGCACC  CCCAGTGCT  CAGGCATCGC  TAGGCATCGC  TAGGCATCGC  TAGGCATCGC  TACCGGATACCT  TAGGCATCGC  TACCGGATACCG  TACCGGATACCG  TACCGGATACCG  TACCGGATACCG  TACCGGATACCG  TACCGGATACCG  TACCGGATACCG  TACCGGATACACCG  TACCGGATACACCG  TACCGGATAA	M I E AAAACTTTAA AGTCTTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTATCAG GCCTAATCAG GCCGGGGGAG CTCAAAGGCG GCTTATCTG ACCAGGGGTTC TAGGCTTATCCG TAGGCGGTGTC TAGGAGCT TATCTCAG GCCTTATCCG TAGAAAGAGT CCTAAGAACT CCTAGATCCT TATCTCAGC GCAATGATCA CCTCCTTCAGC TCCTTCAGCT TCTTCAGCT TAAGATAGTT TACCCGCCCA	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGACCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCAGAGTTTC TACAGAGTTTC TTAAATTAA ATCAGAGTTC CTTTTGATCTT TTTAAATTAA ATCAGAGTTC CCAGAGACCC CCAGTCTAT TCCTCGTTTT CTCGTCGTTTT CTCGTCGTTT CTCGTCGTTTG CTCGATCGT CTCATAGCAGA	E E M TAACTAGCAT  CGAAGGTTAGCAT  TGATGCTATT  GAGGTTTTTT  TGTTTCCTGT  TATCCACAGA  TATCCACAGA  TAGCCTCCGC  AGCTCCCTCG  GAGCTCCCTCG  AGATGGCTATT  TCTTGAGTCC  TTGAAGTGGT  TTCTTCATCACAGA  AATGGCTCACCA  AATGTGTCATCACACAGA  TTCTACACACACACACACACACACACACACACACACA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCCGGTAA GGCCTAACT ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGGCTC AAGTTGCCC CACCAAGTC CAACCAAGTC CACCACAAGTC CCTCACACATT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGGTCA ACCAGGAA AGCATCACAA AGCATCACACA TGTTCCGACC GACGGACACACACACACACACACACACACACAC	V T E TTTATCTGAA  CGGTTCCAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA TGTCCACTGC ACTACTGC CTTCCGGGGCC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTTGGA AAACAAGTTA TGAGTTATA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT TGTTTGCAAG TGTTTGCAAG TAGTGTCAAG TAGTGTCAAG TAGTGTTGAAG TAGGGAAGGC ATATTTGCTAT ACTGTTTGAAT TAGTGTTTGAAG TAGGGAAGGC ATATTTGCTAT ACCGGAAGGG ATACTTTCGTAT GCCGGAAGGG ATAGTTACAGAT ACGCACCAAG AAAACTTCCAA	K A I L GCAGATATCC  CCGGGTCATCA CAAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGCTTGCAGCTT TCCCCCTTT TCCCCCATGT TCCCCCATGT TCCACAAATTC AGGATATATTC CTGCATAATTC CTGCATAATTC CAGGATCTTTCC AGGATCTTAC
2001 GGATACTTGA G  D T  2011 AGCACAGTGG C 2201 TCACCATCAC C 2301 CTAGAATGCA G 2401 TTGCATTCAT T 2501 ACCTGCAGGC A 2601 TAAAGTGTAA A 2701 GGATTAATGA A 2801 CGGACAGGA G 3101 CTCCCTTCGG G 3201 CCGTCAGGC G 3101 CTCCCTTCAGC C 3201 CGGTCAGACA G 3401 CTGAAGCCAG G 3501 GATTAGCAGA G 3501 CGGGCAGAAA A 3701 TGCTTAATCA G 3801 TACCATCTGG C 3801 AAGTGGTCCT G 4001 GCCATTCCTA C 4101 TGTGCAAAAA A 4201 TCTTACTCTA C 4301 CCGGCGCAAA A 4201 TCTTACTCTA C	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAAA PTTATGTTTC AGCCTGGGGT ATCGGCCAAC TAAAAAGAC TAAAAAGGC CTATAAAGAT SAAGCGTGC CGAGCGCTGC CGAGCGCTCC CGAGCGCTCC CGAGCGCTCC CCCACTGC AAAAGGCATCT CAGGATCTTCA TGAGCATCTTAT CAGGATCTTAT CAGGCTTAGC AGCGCTTAGC AGCACTTAT CAGGATCTTAGC AGCGCTTAGC AGCCCACTCC AGCACCACTCC AGCACCACCAC AGCACCACCACCAC AGCACCACCAC AGCACCACCACCAC AGCACCACCACCAC AGCACCACCAC AGCACCACCACCAC AGCACCACCACCAC AGCACCACCACCACCAC AGCACCACCACCACCAC AGCACCACCACCACCACCAC AGCACCACCACCACCACCACCAC AGCACCACCACCACCACCACCACCACCACCACCACCACCA	M I E AAAACTTTTA AGTCTAGAGG AAACCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCGCGGGGAG CTCAAAGGCG GCCTTATCA GCTTTCTCAT GCTTTCTCAT TAGGTTCTCAGG ACAAGAGATC CCTAGAAGACT CCTAGATCCT TATCTCAGCG CCATTATCCT TATCTCAGCG CCATTATCAGC CCTCTCAT GGTGTCACGC TCAGAAGAGTC TATCTCAGCG TCAGAGATCCT TACCCGGCTCAT TACCCGGCTCAT TACCCGGCCT TACCCGGCCT TACCCGGCCT TACCCGGCCT TACCCGCCCA TCACCCCCCCA TCACCCCCCCCA TCACCCCCCCC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTGGGCTCG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGGTCATAGC TGGTCATAGC TGGTCATAGC TGGTCATACT TGAGCTAACT AGGCGGTTTC GTAATACGGT TCGTCATATCG TACAGAGTTC TACAGAGTTC TTTAAATTAA ATCTGTCTAT TCGTCAGTTTT CCCGAGACCC CCAGTCTATT CTCCGATCGTTT CTCCGATCGT TTCGTCGTTTC CTCCGATCGT TCTTGTGACT CATAGCAGA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCTTT GATGCTATT GAGGTTTTTT GGGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCGC AGCTCCGC AGATGATTCATC TCTTGAGTCCTC TCAAGTAGTC TTCTTGAGTCC TTCAAGTCC TTCAAGTCC GATCCGCA AATTGAGCTC TCTGAAGTACT TCGTAGGCTC TCTGAAGTACT TCGTAGGCTC TCTGAAGT TCCTTCACGGC TTCAACTACC TCTCAGAAGT TCTTCATTCC TCTCAGAAGT TCTTAAAAGT TCTTTAAAAGT TCTTAAAAGT TCTTTAAAAGT TCTTAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAGT TCTTAAAAAGT TCTTAAAAGT TCTTAAAAAGT TCTTAAAAT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAAGT TCTTAAAAT TCTTAAAAT TCTTAAAT TCTTAAAAAT TCTTAAAAAT TCTTAAAAT TCTTAAAAT TCTTAAAAT TCTTAAAAAT TCTTAAAAAAT TCTTAAAAAT TCTTAAAAAAT TCTTAAAAAAAA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT AACCACGTAACA GCCTAACTA GCCTAACT CCTGACG TCTGACG TCTGACG TCTAACCAC TCTAACTTC TCACCAC TCTAACTC TCACCAC TCAACCAAC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGATCACAA GGTTCCGAC TAGGTCGTC AAGGACTACAC GGTGGAACG CTAGGTAGCA CTAGGTCGTACCA GCTGGTAGCG CTAGAGCACTAAGTATAT AGAAGTATAA GGAACGAAT AGTAAGTAAT AGTAAGTAAT AGTACAGAA CTAGTTTATC ATCTGGACA ATGGAACGAA CTAGTTTATC ATCTGGAAA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ATTTCCACA ACTGACTGC ACTGCC ACTGC ACTGCC ACTCC ACTGCC ACTCC ACTGCC ACTGCC ACTCC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTCGGT AGTTAGCAAAAGGC AAATGCA CCGGATACCT GGGCTCGGT TGTTGCAG TAATGGAATCCT CTGGTCTGA TAAGGGATTT CTTGGTCTGA AACTACGATA AACTACGATA AACTACGATA ATGCCAGAAGGC ATAGTTTACATGA ATGGCAGCAC GGCGACCGAG AAAACTCTCA GAGGCAACCAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAG GTCGCCTT CACGAACAC CTGGTAACAG CTGCGCTTT CACGAACTC CAGGAACCC CTGGTAACAG CTGCGCTTT CACGAACTT CCGCATTT CCCCATTT TCCCCCATTT TCCCCCATTT TCCCCCATTT TTGCTCTTGC AGAACTT CTGCTCTTGC AGAACTT CTGCTCTTGC AGAACTT CTGCTCTTGC AGAACTT CTGCTCTTGC AGAACTTAC AAACAGGAAG
2001 GGATACTTGA G	E Y E STTAGTTTAT  CGGCCGCTCG CATTGAGTTT  TTGAAAAAAA PTTATGTTTC ATGCCAAC ATGCGAAC ATGCGGCAAC TATAAAGGC CTATAAAGGC CTATAAAGGT SAACGCTGC CGACGCTGC CGACGCTGC CGACGCTCC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CTACGGGATCT CAGCGCTTCA CGCACCCC CTACGGGATCA CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTGC CCCAGTCC CCCAGTGC CCCAGTCC CCCAGTC CCCAGTCC CCCAGTC CCCAG	M I E AAAACTTTA AGTCTAGAGG AAACCCCCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCGCGGGAG CTCAAAGGCG ACCAGGCGTTCCGC TAGGCGTTCCCTAGGCGTTCCCTAGAGAGATC CCTAGATCCC TATCTCAGCG GCATGATACCC TAGGCCTCCAT GGTTCACGC TCCTCGGTC TAGGATCCT TACGCGCCCA TAGGATCCT TACGCGCCCA TAGGATCACT TACGCGCCCA TAGGATCCT TACGCGCCCA TAGGATCCT TACGCGCCCA TAGGATCACCC TAACACCC TAGGATCACCC TAGGATCACC TAGGATC TAGGATC TAGGATC TAGGATC TAGGATC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGGCTAATC AGGCGGTTT GTAATACGGT TGTAATACGGT TGTAATACGGT TACAGAGTTC TACAGAGTTC TACAGAGTTC TACAGAGTTC TTTAAATTAA ATCTGTCTAT TCGCAGAGACC CCAGTCTATT TCGTCGTTTG TTCTGTGACT CATAGCAGAA CTCCGATGGACC CAGTCTATT TCGTGGACT CTCGATGGACC CAGTCTATT TCGTGACC CAGTCGACC CAGTCGACC CAGTCGACC CAGGCAAA CTCGTGACC CAGCACGGAAA CCCGACACGGAAA CCCGACACGAAA CTCGTGCACC CAGACCGAAA	E E M TAACTAGCAT  CGAAGGTTAT GAGGCTATT GAGGCTATT GATGCTATT GAGGCTATT GATGCTATT GATGCTATT GAGGTCGA TAGCTCGA AGGTCCGCA AGGTATTGGGA AGTCGGAAGT GATCACA AATGATCACA AATTGTTCAC GGAAGT GGTGAAGT GGTGAAGT CGTGAAGT CGTGAAGT GGTGAAGT CGTGAAGT CGTGAAGT CGTGAAGT CGTGAAGT CGTGAAGT CGTGAAGT CACTGAAGT CGTGAAGT CACTGAAGT CACTGAAGT CACTGAAGT CACTGAAGT CACTGAAGT CACTGAAGT CACTGAATACT CATTGAATACT CACTGAATACT CACTGAATACT CTTTGAATTACT CTTTTGAATTACT CTTTTTTTTTT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTATTTG AAAGCAGTA GCGATTCGCTG GCTCTTCCGC ATCAGGGGAT ACCCGGTAA GGCCTAACTA GCCAACCAGT CTGAACCAGT TCTGACGCT TTAAATCAAT GCAACCAGCT CTTAAATCAAT GGCAGCT CTTAAATCAAT GGCAGCT ACCAGATT GGGAACCTAG ATTCAGCTC CACCAGATT GGCACCAGTT TCAACACTT TCAGCACT TCAACACTCT TCAACACTC TCAACACTCT TCAACACTCT TCAACACTCT TCAACACTCT TCAACACTCT TCAACACTC TCAACACT TCAACACTC TCAACACTC TC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGGAC TGCTCGCTC AACCCAGGAC TGGTCGTCC GACAGGACT AGGATCGTTC GGCTACACT GGTACAG TGTACAGCAGATAT TGACTCCCAG CTAAAGTATA TGACTCCCAG GGTTACCAAT TGTTCCAAC CAGTGTTACC CAGTGTTACC CAGTGTTATC ATTCTGAGAA GGAAAACGTT TTACTTCTAGAA	V T E TTTATCTGAT GATACATTGA AAGCTGCAT AAATTCCACA ATTTCCACA ACTGACTGCACT ACACTGACTGCACT ACACTGACTGAC ACTGACTGAC ACTGACTGAC ACTGACTGAC ACTGAC AC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA GGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACGAGCA TATTTGCAAG TAAGGGATT CTTGGTCTGA AACTACGATA AACTACGATA AATTTACATACATA AAGTACAGAAGGC ATAGTTTACATGA AGGCAACGAC AGTTTACATGAAGAAGGC AGGCGACCACACAC GGGGCGACCACACAC GGGGCGACCGAC	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA CGAGAGCT CGGTTACCAA CGGGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCCATGT TCCCCCATGT TTGCTCTTGC AGGAATTA TCTTGCTCTTGC AGGAATTA TCTTGCTCTTGC AGGAATTATTGT
2001 GGATACTTGA G	E Y E STTAGTTTAT  CGCCCGCTCG CATTGAGATT  STGAAAAAA  TTATGTTC ATGCCTGGGC AAC TATGAGCTTA AGCCTGGCC AAC TAAAAAGGC  TAAAAAGGC  TAAAAAGGC  TAAAAAGGC  CGACGGTGGC AAAAGGATCT AGGATCTCA AGGATCTTCA AGGATCTTCA AGGATCTTAA AGGATATCG ATACGAGATAA ATCCAGTTCG AAAAAAAGG  SCAAAAAAAAG	M I E AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTATCAG GCCTAATCA GCCGGGGGAG CTCAAAGGC GCTTTCTCAG GCTTTATCT AGGTTTATCG TAGGCGGTAGC AAAAAGAGC GCTTATCAG GCTTATCAG GCTTATCAG GCTTATCAG GCTTATCAG GCTTATCAG GCTTTATCCAG GCTTATCAG TAAAAAGATT CAAGAACT CCTAGATCCT TATCTCAGC GCATGATCCT TATCTCAGC GCATGATCCT TATCTCAGC GCATGATCCT TATCTCAGC GCATGATCCT TAAGATCCT TAACGCC TAAGATCTT TACCGCGCCA ATGTAACCCA GAATAAAGGCT	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTTTG TGGTCATAGC TGGTCATACT AGCGGGTTTG GTAACTACGGT TCCCCCTGGA AGCTCAGGAGTTCG TAACTATCG TAACTATCG TAACTATCG TAACTATCG TAACTATCG TAACTATCG TCCCCTGGA ACTCTTTTTAATTAA ATCTGTCTTT TTTAAATTAA ATCTGTCTTT CGCGAGACCC CCAGTCTATT CTCGTCGTTTG CTCGGATCGT TCGTGGTTTG CTCGGATCGT TCTTGACCT CATAGCAGAA CTCGTGCACC GACACGGAA TCGTGCACC GACACGGAAA TAGGAAAAATA	TAACTAGCAT  CGAAGGTTATT GAGGTTATT GAGGTTATT GAGGTTATT GAGGTTATT GTATTCCTGT TATCACAGA TAGCCTCCGC AGCTCCCCGC AGCTCCCTCG AGCTCCTCC TGGAAGTGGT TTCATCATC ACAGAAGTGGT TTCATCATC ACAGAAGTGGT TTCATCATC ACAGAAGTGGT TTCATCATC ACAGAAGT GTTTAATT TGCT TGTAGGCTTC TGTGAGAAGT GGTAGGTATCT TGTTGAGATACT TGTTGAGATACT TGTTGAGATACT ACACAAATTAGG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCAGTTCGGTA ACCAGCTAACT ACAAACCAA CTCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCC CACCAGTTCGCC CACCAGTT GGGAACCTAG TTAAATCAAT TTAATCATT TCAGCATCT TCAGCATCT TCAGCACTCT TCATACTCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCATACCTCC GGTTCCCGCG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGGAC TGCTCGCTC AACCCAGGAC TGGTCGTCC GACAGGACT AGGATCGTTC GGCTACACT GGTACAG TGTACAGCAGATAT TGACTCCCAG CTAAAGTATA TGACTCCCAG GGTTACCAAT TGTTCCAAC CAGTGTTACC CAGTGTTACC CAGTGTTATC ATTCTGAGAA GGAAAACGTT TTACTTCTAGAA	V T E TTTATCTGAT GATACATTGA AAGCTGCAT AAATTCCACA ATTTCCACA ATTTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACATGA AAATCACACA ATTTCCACA ACTGACACTGAC ACTGACTGAC ACTGACTAC ACTGACT ACTGA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA GGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACGAGCA TATTTGCAAG TAAGGGATT CTTGGTCTGA AACTACGATA AACTACGATA AATTTACATACATA AAGTACAGAAGGC ATAGTTTACATGA AGGCAACGAC AGTTTACATGAAGAAGGC AGGCGACCACACAC GGGGCGACCACACAC GGGGCGACCGAC	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA CGAGAGCT CGGTTACCAA CGGGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCCATGT TCCCCCATGT TTGCTCTTGC AGGAATTA TCTTGCTCTTGC AGGAATTA TCTTGCTCTTGC AGGAATTATTGT

### pMT-HisFlag-DmSNAP43 mut#1 STOP

1 TCGCGCGTTT CGGTGATG		mamaa aa aa m	aas aamaaaa	a. a. aaama.	as aammamam	ama a a a a a a a a	00000000000	a. a aaaaa
101 TCAGGGCGCG TCAGCGGG								
201 CCGCACAGAT GCGTAAGG								
301 TACGCCAGCT GGCGAAAG								
401 TGAATTAATT CGTTGCAG	GA CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC CGCCCACC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTGGAG	AA CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAACGAA								
801 AAATCAAGTG AATCATCT								
001 111110111010 1111011101	010011101111	110000001110	1110111100000	11101111101111	> M H	н н н	н н т р	Y K D
901 GACGATGACA AGGGCACT	C TCACCTCAAT	A TOTTO ACC	летестесса	сстсстссл л				
> D D D K G T			D C W E			R L V N	D G E	
1001 TCGAGGTGTT CTGCCGGT								
>F E V F C R		L Q L Q			T N H T		ATT	L A A L
1101 GCATGTGGCC AAGCGACT								
> H V A K R L	SCSR			F P A S	~	R I G	G F F L	L Y V
1201 ATCTACTACA AGCAGCCC	AC GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
> I Y Y K Q P	T H N F	I K I	E V S P	R T W	Q E L	T D Y A	L D L	R K D
1301 GTCCGGAGCG GAAGGACA	CT CATCAGATCG	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
>S P E R K D	r h Q i	AYML	WRL	T Q E	QAFR	F T A	L D Y	CQGL
1401 GGACAATCTG GTGGACTA	CG ACCGTGTGGA	GACCGTAGCG	GGTGCCAAGG	AACAGAGGCA	GAGTGCCTTG	ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
> D N L V D Y	DRVE	T V A	G A K	E O R O	SAL	M O K	OORA	N G V
1501 AGTGCCACAT ACGCAGCG								
> S A T Y A A			O A S O			A A Y N		K O L
1601 CGGCTGGTCA TGAGCACG								~
>A A G H E H		S O I F			F A D I		L G A	R K S T
						Q S V		
1701 TCCAGATGAG AAATGCAC								
> P D E K C T	TTST			R Q R V		A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGAA								
> Q H Q T D E	. ~		N E T Y	~		A T V F	QRE	L P E
1901 ACGTGCAGCA AGAGTATG								AAGCTATTTT
>D V Q Q E Y	E MIE	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTT	AT AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	CCACATATCC
2001 GGMIMCIIGM GIIMGIII								OCHOMINICC
> D T								GCHGHIHICC
	CG AGTCTAGAGG	GCCCGCGGTT	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT		
> D T 2101 AGCACAGTGG CGGCCGCT							TCTACGCGTA	CCGGTCATCA
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT	TT AAACCCGCTG	ATCAGCCTCG	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TCTACGCGTA TGAGTTTGGA	CCGGTCATCA CAAACCACAA
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA	TT AAACCCGCTG AA TGCTTTATTT	ATCAGCCTCG GTGAAATTTG	ACTGTGCCTT TGATGCTATT	CTAAGATCCA GCTTTATTTG	GACATGATAA TAACCATTAT	GATACATTGA AAGCTGCAAT	TCTACGCGTA TGAGTTTGGA AAACAAGTTA	CCGGTCATCA CAAACCACAA ACAACAACAA
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT	TT AAACCCGCTG AA TGCTTTATTT FC AGGTTCAGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA	GACATGATAA TAACCATTAT AAACCTCTAC	GATACATTGA AAGCTGCAAT AAATGTGGTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC	FT AAACCCGCTG AA TGCTTTATTT FC AGGTTCAGGG FT GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG	FT AAACCCGCTG AA TGCTTTATTT FC AGGTTCAGGG FT GGCGTAATCA GT GCCTAATGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA	FT AAACCCGCTG AA TGCTTTATTT FC AGGTTCAGGG FT GGCGTAATCA GT GCCTAATGAG AC GCGCGGGGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT	TT AAACCGCTG AA TGCTTTATTT FC AGGTTCAGGG FT GGCGTAATCA GT GCCTAATGAG AC GCGCGGGGAG CA CTCAAAGGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT CGTTCGGCTG CAGCAAAAGG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTCCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG	TT AAACCGCTG AA TGCTTTATTT C AGGTTCAGGG TT GGCGTAATCA GCCTAATGAG AC GCGCGGGAG CA CTCAAAGGCC CC GCGTTGCTGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGGGAGCGG TATCAGGT 2901 CCCGACAGGA CTATAAAG 3001 CCCGACAGGA CTATAAAG	TT AAACCGCTG AA TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCA GT GCCTAATGAG AC GCGCGGGAG CA CTCAAAGGCG CC GCGTTGCTGC AT ACCAGGCGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC ACACACATGTG AAATCGACGC CTGCCGCTTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAAGCGTG	TT AAACCGGTG AA TGCTTTATTT TGCTTATTT TGCTTATTG TGCTTATTG TGCGTAATGAG AC GCGCGGGAG AC CTCAAAGGCG AC GCGTTGCTGG AT ACCAGGCGTT ACCAGGCGT ACCAGGCGGGGT ACCAGGCGGGGGT ACCAGGCGGGGGT ACCAGGCGGGGGGT ACCAGGCGGGGGGT ACCAGGCGGGGGGGGT ACCAGGCGGGGGGGGT ACCAGGGGGGGGT ACCAGGGGGGGGGT ACCAGGGGGGGGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACCT GGGCTGTGTG	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAACGT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATTAAAG 3101 CTCCCTTCGG GAAGCGGTG 3201 CGGTTCAGCC CGACCGGTG 3201 CGGTTCAGCC CGACCGGTG	TT AAACCGCTG AA TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCA GC GCCGGGGAG CA CTCAAAGGCG CC GCGTTGCTGG AT ACCAGGCGTT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT	ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCCA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAAGCGTG	TT AAACCGCTG AA TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCA GC GCCGGGGAG CA CTCAAAGGCG CC GCGTTGCTGG AT ACCAGGCGTT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT	ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCCA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAACGT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATTAAAG 3101 CTCCCTTCGG GAAGCGGTG 3201 CGGTTCAGCC CGACCGGTG 3201 CGGTTCAGCC CGACCGGTG	AAACCCGCTG AA TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCAG GC GCCGGGGAA CC CCAAAGGCG CC GCCTTGCTG ACCAGGCGTT CC GCCTTATCAG TG GCCTTATCAG TG GCTTTCTCAG TAGGCGTGC	ATCAGCCTCG GTGAAATTTG GGGAGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGC AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGGACAC	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAC CTGCGCTCTCG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CCGTTCAGCC CGACCGCT 3301 GATTAGCAGA GCGAGGTA	TT AAACCGCTG AT TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCA GT GCCTAATCAG CG CGCGGGGAG AC CTCAAAGGCG CC GCGTTGCTGG GT ACCAGGCGTT GC GCTTTCTCAT GC GCCTTATCCG GT TAGCGGTGC GC TAGCGGTGC GC TAGCGGTGC GC AAAAGAGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCTT GGTAGCTCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTGAGTCC TCTTGAGTCG TGATCGGCAA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCTGACG TGCGCTCCC CAGTTCGGTG AACCCGGTAA ACCAGTAACTA ACAAACCACC	GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACAC GCTGGTACCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGACGACG GTGGTTTTTT	TCTACGCGTA TGACTTTGGA AACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CCAGGATACCA GCCAGTAGCA GCAGATACGA TCAGGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCACTTC CAGCACTTC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGCG ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAACCGTG 3201 CAGTTAGCAGC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCCAG TTACCTTC	TA AACCCGCTG AT TGCTTTATTT TC AGGTCAGGG TG GGCGTAATCAG GGCGTAATCAG CGCGGGGGAG AC CTCAAAGGCG AT ACCAGGCGTT GGCTTTCTCAT GGCTTTCTCAT GGCTTTCTCAT GCGCTTTCTCAT GCGCTTATCCG TAGGCGGTGC GAAAAAGAGTT TC CAGGAGATT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTACCTCT TCACAGGTTC CTTTGATCTT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TTCTACGGGA TTCTACGGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACACGACTTCCGTCT GACACGACTTCCGCTCC GCTGCTACACC AGTGGAACGA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTGG AGAACATGTG AAATGACTGC CTGCCGCTTA ACTCACCTG AGAAGACATG ACTCACCTG AGAAGACATTG AGAAGACATG ACTCCACAGG ACTCCACAGG ACTCCACAGGT ACAGGACATTTTTT AAACTCACGT	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTGT GGCCTGTGT GGCAGCAGCCA TATTTGCTAT TGTTTGCAAG TAAGGGATTT	CCGGTCATCA CAAACCACAA ACAACAACAA GCACCAGCT GCTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGCGCATA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGGCTTCG CAGCAGATTA CAGCAACTTG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGACGGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGGT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAAGCAG TTACCCTTC 3501 CGGTCAGCA GCGAGGTA 3401 CTGAAGCCAG TTACCCTTC 3501 CGCGCAGAAA AAAAGGAT	TT AAACCGCTE A TGCTTTATTT TC AGGTTCAGGG TT GGCGTAATCAG GC GCCGGGGAG AC CTCAAAGGCG AT ACCAGGCGTT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT GC GCTTTCTCAT TC TAGGCGTGG AAAAAGAGTT TC CAAGAAGATC CT CTGATCCT CTAGATCCT CTAGATCCT CTAGATCCT CTAGATCCT CTAGATCCT CAGATCCT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT CTAGATCCT CTAGATCAT C	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT GGAGTTTG GTAATACGGT CCCCCTGGA AGCTCACCGT GTAACTATCG TACAGAGTTC GTAACTATCG TACAGAGTTC TTTGACTTT TTTAAATTAA	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTCTG TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT ATCCGCAA ATCACGGGA ATCACGGGA ATCACGGGA AATCAAAGTT	CTAAGATCCA GCTTATTTG ANAGCAAGTA GTGAAATTGT GCGTTGGGCT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA ACAAACCAC TCTGACGCT TCTGACGCT TTAAATCAAT	GACATGATAA TAACCATTAT AAACCTTAC TATCCGCTCA CACTGCCGC ACCGGCTAACGCAGAA AGCATCACAA AGCATCACAA AGCATCACAC TGTTCCCAGC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGACCACGC ATCCCAAGGT ATCCCACGT ATCCCACG AGAAGACAGC AGAGGACAG AGAAGACACA TGGTTTTTT AAACTCACGT	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCAAG TGTTTGCAAG TTTTGCAAG TTTTGCTAGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAC CTGCTGCTACACAC CTGCGCTTT CAGCAGATTA TGGTCATCACAC CAGCACATTA TGGTCATCACAC CAGTTACCAC CAGTTACCAC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAACCGTG 3201 GATTAGCAGC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCAGA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA	TAAACCGCTG TGCTTTATTT TGCTTTATTT TGCAGGTTATCAGG TGGCGTAATCAG TGGCGGGAG CACCGCGGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGGAG CACCGCGGAG CACCGCGGAG CACCGCGGAG CACCGCGGAG CACCGCGGAG CACCGCGGAG CACCGCGGAG CACCGCGAGACCCT CCCGCGCGCGCG CCCGCGAGACCCT CCCCGCGCGCGCG CCCGCGCGCGCGCGCGCGCGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAATA AGCCGGTTTG GTAATACGGT CGTTTTTCCGA AGCTCACGCT GTAACTATCG GTAACTACTC GGTACCTCTTCATTGATCTT TTTAAATTAA ATCTGTCTAT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTC TCTTGAGTCC TTGAAGTGC TTGAAGTGT TTTTACGGGG AATCAGAGTT TTCCTTCATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCGCG ATCAGGGGAT CCCCTGACG TGCGGTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCTC TCTGACGCTC TTAAATCAAT CATAGTTGCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC TACCCGGCC AACGAGAA AGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA TGTTCCGAC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACCA AGTGGAACGA CTAAAGTATA TGACTCCCCG	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG AGACATTGC AGACATGGC AGAACATGTG AAATCGAGGC TTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACATGT ATCGCACTG AGAAGACAGC ATCACCACTG AGAAGACAGC ATCACCACTG AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC ATCACCACTT AAACTCACGT TATGAGATAAA	TCTACGCGTA TGAGTTTGGA AACAGTTA TGGCTGATTA CACCTGT TGCGCTCGGT TGCGCTCGGT TCAGTACGA GCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG GTCCGCTCT CACGAAACC CTGGTAACAC CTGCGCTCT CACGAACCC CTGGTAACAC CTGCGCTCT CAGCACATTA TGGTCATGAG CAGTTACCAA CAGGAGGCC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGACGGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CGGTTCAGCC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCCAG TTACCTTC 3501 CGCGCAGAAA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTGG	TAAACCCGTTG TATTATTT TG AGGTTAATTATT TG AGGTTAATCAGG TG GGCGTAATCAG TG GCCTAATGAG AC GCGCGGGGAG AC CCCAAAGGCG TG GCGTTGCTGG TG GCGTTTCTCAT TG GCCTTATCCG TG TAGGCGTTG TG AAAAAGAGTT TC CAAGAAGATC TC TATCTCAGCG TC TACTCAGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTACTATCG TCACAGGTTC CTTTGATCTT TTTAAATTTA ATCTGCTACTAC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAAGTT TCTTACGGGA TTCTACGGGA AATGAAGTT TCGTTCATCACACAC ACGCTCACCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTACTA TCTGACGCT TTAAATCAAT CATAGTTGGC GCTCCAGATT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGACTACACT GCTGCTACACT GCTGCTACACT AGTGCACCT AGTGCACCA AGTGAACGA TGAACCGA TGACTCCCCCG TATCAGCAT	GATACATTGA AAGCTGCAAT AAGCTGCAGTA CAATTCCAGCA TTTCCAGCA CTGCCACTGC AGAACATGTG AAATCGACGC CTGCCCGCTTA ACTCGCCACTG AGAAGACATGTG AGAAGACATGTG AGAAGACAGC TCGCCACTG AGAAGACAGC TATCAGCAACT ATCAGCAACT ATCAGCTAAACT AAACCAACGT AAACCAACGCA AAACCAACGCA	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGAATACCT ATTTGGTATT GGTCTGA TAAGGGATT CTTGGTCTGA AACTACGATA ACTACGATA ACCACGAAAGGC	CCGGTCATCA CAAACCACAA ACAACAACAA GCACCAGCT CGTTCGGCTG CAGCCAAAAGG GGTGCGCATA GTCGCCTTT CACGAACCC CTGGTAACAG CTGGCTACAC CTGGCTTT CAGCAACTT TGGTCATGAC CAGCAGATTA TGGTCATGAC CAGGAGGGCT CCGAGCGCAC CCGAGCGCAC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACGG TAATAAAG 3101 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3301 GATTAGCAGA GCGAGGGTA 3401 CTGAAGAGCAG TTACCTTC 3501 CGCGCAGAAA AAAAGGAT 3501 CGCGCAGAAA AAAAGGAT 3601 ATTATCAAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCCAGTG 3901 AAGTGGTCCT GCAACTTT	TAAACCGCTG TAATCAGG TG AGGTTATTT TG AGGTTCAGGG TG GGCTAATCAG TG GCCTAATCAG CG CGCGGGAA CA CTCAAAGGCG CG GCGTTGCTG CG GCTTTCTCAT CG GCCTTATCCG TAGGCGTTG TG TAGGCGTTG CG AAAAAGAGTT CAAGAAGATC CC CTTACACGC TTACTCAGGC TTACTCAGT TCAGATTAC TCCGCCTCCAT	ATCAGCTTG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGTTTG GTAATACGGT CGTTTTTTCCA AGCTGACTTTTCCA TACAGGTTC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTTCT TTTAAATTAA ATCTGTCTAT TTTAAATTAA ATCGCGAGACCC CCAGTCTATT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TTGAGTCG GATCGGCAA TTCTACGGG AATCAAGAT TTCGTCATC AAGTCACCACA AATTGATTCC AATTGTTCACC AATTGTTCACC AATTGTTCACC AATTGTTCACC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGC ATCAGGGGAT CCCCTGACG ATCAGGGGAT CCCCTGACG AACCCGTAA ACAAACCAC TCTGACGT TCTGACGT TTAAATCAAT CATAGTTGCC GCTCAGACT GCTCAGACT GGGAAGCTAG	GACATGATAA TAACCATTAAT AAACCTCTAC TATCCGCTCA CACTGCCGC TCCCTGGTCAACGCAGCA AGCATCACAA AGCATCACAA AGCATCACAC CAGCAGCAGCAC CAGCAGCACT CAGCACCACC CAGCACCACC CTAGACCACC CTAGACCACC CTAGACCACC CTAGACCAC CTAGACCACAC CTAGACCACAC TAGCACCACACACT TATCAGCAACAT AGTAAGTAAGTAGT	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG ACTGCACTCGACACA CTGCCGCTTA GCTCCAACGT ATCGCCACTG AAATCACGC AGAACATTTT CACCACTG ATCGCCACTG AGAAGGACAG AGAGGACAG AGAGGACAG AACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACCACCA AAACCAGCA AAACCAGCCA TCGCCAGTTA AAACCAGCCA TCGCCAGTTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCAAG TAAGGGATTC TCTTGGTCTGA AACTACGATA GCCGGAAGGC	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTTT CACGAGATTA TGGTCATCAG CAGCAGATTA CGGCAGGTTA CGGCAGGTTA CGGCAGGTTA CGGCAGGCCAG CAGTTACCAA CAGGAGGCCAG CAAAGTTGTT
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAACCGTG 3201 GATTAGCAGC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCAGA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CGAGCCTT	TAAACCGCTG AT TGCTTTATTT TG AGGTTCAGGG TG GGCTAATCAG GC GCGGGGGAG AC CTCAAAGGCG CC GCGTTCTCAT GC GCTTATCAG GCTTATCAG GCTTATCAG GCTTATCCAG GCTTATCCAG GCTTATCCAG GCTAAGAAGATC CTAAGAAGATC CC CAAGAAGATC CC TATCTCAGCG CT GCAATGATAC CC TATCTCAGCG CT GCAATGATAC CC TATCTCACGG CT GCAATGATAC CC TATCTCACGG CT GCAATGATAC CT GCATTGATCAGCG CT GCAATGATAC CC GCATCCAT CC GCGCTCCAT CT GGTGTCACGCG CT GGTTCACGCG CT GTTCACGCG CT GGTTCACGCG CT GTTCACGCG CT GTTTCACGCG CT GTTCACGCG CT GTTTCACGCG CT GTTTTCACGCG CT GTTTTCACGCG CT GTTTTCACGCG CT GTTTTCACGCG CT GTTTTTTTTTTTTTTTTTTTTTTTTTTTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGGTTTG GTAATACGGT CGTTTTTCCA AGCTCACGCT GTAACTACC GTAACTACC GTAACTACC GTAACTACT TTTCAATCTA ATCTGATCT TTTCAATCT TTTCAATCTA ATCTGTCTTT TCGCGAGACCC CCAGTCTCTT TCGTCGTTTT TCGTCGTTTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTACTC GTAGGTACT TCTTGAGTCC TTGAAGTGG TATCCGCAA TTCTACGGG AAATGAAGTT TTCGTTCATC ACGCTCACCG GATGTCACCG GATGTCATC ACGCTCACCG GATGTCATCC GTATGGCTTC GTATGGCTTC GTATGGCTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCAACTA ACAAACCAC TCTGACGTC TCTGACGCTC CTTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCAC GCTCCAGATT CGGGAAGCTA ATTCAGCTCC ATTCAGCTCC ATTCAGCTCC ATTCAGCTCC ACTTGCCAGATT GGGAAGCTAG ATTCAGCTCC	GACATGATAA TAACCATTAT AAACCTCTTAC TATCCGCTCA CACTGCCGCC TACCAGGAA AGCATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA TGGTCGTACACT GGCTACACT GCTGGTACCA AGTGAACGA TTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTATA GGATTCCCAAC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG AGAACATGTC AAATCGAGCC AGAACATGTG AAATCGAGCC AGACACTGC AGACACTGC AGACACTGT ACCCCACTG AGAGGCACTA ACCCCACTG AGAAGACAGCT ATCCCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA GATCAAGGCG	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAGCATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CCGGATACCA GGCAGTCAGA CCGGATACCA TATTTGGTAT TGTTTGCAAG TAGGGTTTCGTCTG ACTTACGATA CCGGAAGGG ATGTTTGCAAG ACTACGATA CCGGAAGGG ATGTTTTCG	CCGGTCATCA CAAACCACAA ACAACAACAA CGACCACGC CGTCCGCTCGCTCGCTCGCTCGCTCTCGCTCTC CACGAAACCC CTGGTAACAG CTCCGCTCTC CAGCACATTA TGGTCATCACAC CAGCACATTA TGGTCATCACAC CAGCACGTTCT CAGCACGTTCC CAGCACGTTCACCAC CAGCACGTTCT CCGCAGCGCAG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGACGGG TATCAGCT 2901 CCAGGAAGCA CTATAAAG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CGGTTCAGCC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCCAG TTACCTTC 3501 CGCCCAGGAA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTGG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCT AGAGCATCT 4001 AGAGCATTAC 4001 GCCATTGCT AGAGCATCT 4001 AGAGCATTAC 4001 AGAGCATTAC 4001 AGAGCATCT 4001 AGAGCATCT 4001 AGAGCATTAC 4001 AGAGCATCT	TAAACCCGCTG TATTATTT TG AGGTTAATTAA TG GGCTAATCAA TG GCCTAATCAA TG GCCTGAAGGCG CG GCGTGGGAG TACAAGGCGT TG GCCTTCCAT TG GCCTTCCAT TG GCCTTATCCG TG TAGGCGGTG TAGGCGGTC TAGGCGGTC TAGGCGTT TC TAATCAGGG TC TAATCAGGG TC TAATCAGGG TT CAAGAAGATC TC TAATCAGGG TT GCAATGATCAT TG CGCCTTCAT TG GCAATGATAC TG TGTTCAGGG TG TGTTCAGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGGG TG TGTTCAGG TG TGTTCAGGG TG TGTTCAGG TG TG TGTTCAGG TG T	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTACTATCC TTAACTATC CTTTGATCTT TTTAAATTTA ATCTGCTATC CCCAGTCTAT CGCGAGACCC CCAGTCTATT CGTCGGTTTG CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CGCGATCGT CTCGATCGT CTCGATCGT CTCGATCGT CTCGATCGT CTCGATCGT CTCGATCGT TCGTCGTTGT CTCCGATCGT CTCGATCGT TCGTCGTTGT CTCCGATCGT TCGTCGTTGT CTCCGATCGT TCGTCGTTGT CTCGATCGT TCGTCGATCGT TCGCAATCGT TCGTCGATCGT TCGTCATCGT TCGTCGATCGT TCGTCGATCGT TCGTCGATCGT TCGTCATCGT TCGTCGATCGT TCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TACTCCGCAA TTCTACGGGAA TTCTACGGGAA TTCTACGGGAA TTCTACGGGAA TTCACTCACC AATTGTTCCC TGTAAGGCTCC TGTAAGAGTT TGTCAGAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTACT TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AACTTGCCC AAGTTGCCC AAGTTGCCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGACTACCAC TGTGCTGTC GCTGGTAGCG AGTGGACGA GTAGACGA TGACTCCCCG TATCAGCAAT AGTAAGTATA AGTAAGTATA GGTACCCAC TATCAGCACA CAGTGTTATC	GATACATTGA AAGCTGCAAT AAGCTGCATCA CAATTCCACAC TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACAGT ATCGCACTG AGAAGATTTTTT AAACTCACGT TATCAGTAAGATAAA AAACCAGCCA TCGCCAGTTA GAAGACAG ATCAGCAGTAAAACCAGCCA ATCAGCAGTAAAACCAGCCA ACTCAATGGTTAAAAGCAG ACTCAATGGTTAAAAGCAG ACTCAATGGTTA	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT AGTTTGGTATT GGTCTGA TAAGGGATT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA CCGGAAGGG ATAGTTTGCA ATGCCAGCA ATGCCAGCA	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GGTCACCACC CGTTCCGCTT CAGCAAAAGG GGTGCGCATA GTCCGCTTT CAGCAACCC CTGGTAACAG CTGGCTATACAA GTCTGCTTT CGGCTGT CAGCAAGATTA GGGAGGGC CCGAGCGCAC CAGGGGCAC CAGAGCCC CCGAGCCCAC CAGAGCCC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCGTCAG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCCAG TAACCTTC 3501 GGCGGAGAA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTGG 3901 AAGTGGTCCT CAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTGTC ATGCCATC	TAACCCGCTG TAGCTTATTT TG AGGTTCAGGG TG GGCTAATCAG GC GCCGGGGAG AC CCCAGGGGGAG AC ACCAGGCGTT CCCAGGGGAG ACAGGCGTT CCCAGGGGAG ACAGGCGTT CCCAGGGGAG ACAGGCGTT CCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	ATCAGCTTG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGTTTG GTAATACGGT CGTTTTTCCA AGCTCACCT GTAACATATC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTTTT TCTTGATCTT TTTAAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT TCGTCGTTTG TCTCTGGATCGT TTCTGTGATTT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAGTGGT ATCCGCAA TTCTACGGGA TTCTACGGGA TTCTACGGG AATTGATCT TCGTCATC AATTGATCC GTAGGCTACCG CGTAGGCTTC TCTGAGAGT TTCGTCACAGG GGTAGGTTCC GTAGGAGTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT GCTTTCCGC ATCAGGGGAT CCCCCTGACG GCCCTGACG ACCAGGTAA ACCACGTAA ACAAACCAC TCTGACGCT TTAAATCAAC CATAGTTCGCG GCTCAGCTC CTAAGTCAGCTC CTAAGTCAGTT GCCAGATT GGGAAGCTAG ATCAGCTTC AAGTTGCCC AAGTTGCCC AAGTTGCCC CAACCAAGTC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGACC TAGGTCGTC GACACGACTT GACACACT GCTGGTACGA CTAAAGTATT TGACTCCCG TATCAGCAT TATCAGCAT TATCAGCAT CATTATCAGCA CATGTTATC CAGTTTATC CAGTTTATC CAGTTTATC CAGTTTATC CAGTTTATC CATTCTAGAAA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA GCTCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCAGCTATA AACTCACGT ATACAGGTAAA TCGGTAGATAA TCGGCAGTTA GATCAAGGCG ACTCAATGGT AATCAAGGCG ACTCAATGGT TAGATATATG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTT GGGCTCTGTG GCACCACCA TATTTGCAAG TAATTGCATA CTTAGCATA CTTGGTCTGA AACTACGATA AACTACGATA AACTACGATA AACTACAATA AACTACACACCAC GCGCACCCAC GGCGACCCAC	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGAT CAGCAAAAGG GCGGCAAAAGG CTGGCTCT CACGAACCC CTGGTAACAC CTGCTAACAC CTGCTACTAC CAGCAGATTA CAGCAGATTA CCGAGCGCGC CAACGTTGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCTCTTGC TTGCTCTTGC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAACCGTG 3201 CGGTCAGGC CGACCGCT 3301 GATTAGCAGA GCGAGGTA 3401 CTGAAGCAGA AAAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTTCC ATGCCGTC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTTCT ATGCGGATC 4101 CCGGCGTCAA ATGCCGATC	TAAACCGCTG TATTTATTT TG AGGTTCAGGG TG GGCTAATCAG TG GGCTAATCAG TG GCCTAATCAG TG CCCTACTGCTGG TG GCCTTACCGG TG GCCTTACCGG TG AGAAGAGATC TC CAAGAAGATC TC TAACAAGACT TC CAAGAAGATC TC TATCTCACGG TG CAATGATCACGC TG CAATGATCACGCG TG CAATGATCACGCG TG CAATGATCACGCG TG CAATGATCACGCG TG CAATGATCACGCG TG CAATGATCACGCG TG TACTCGGCGC TATCTCGGCT TATCTCGCGCCCAA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGTTTG GTAATACGGT CGTTTTTCCGA AGCTCACCCT GTAACTATCC GTAACTATCC GTAGCTACT TTCAACTTT TTTAACTTT TTTAAATTTA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCGATCGT TTCTGTGTCT CTTCGATCGT CTCGATCGT CTCGATCGT CATAGCAGAAC CCAGTCGTTTG CTCGATCGT CTCGATCGT CATAGCAGAA CCC CAGTCGTTTG CTCGATCGT CATAGCAGAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TAGAGTGGT GATCCGCAA TTCTACGGGG AAATGAAGT TTCGTTCATC ACGCTCACCG GTATGCTCACCG GTATGGCTTC TGTCAGAAGT CTTTAAAAGT CTTTAAAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG ACCCGTTAA ACCAACTA ACAAACCAC TCTGACGCT CTCAACTA CATAGTTGCC GCTCCAGATT CGCACTCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC AGTTGCCC CACCAGATT ATCAGCTCC AAGTTGCCC CACCAAGTC CACCAAGTC CACCAAGTC CCTCATCATT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGACA AGCATCACAA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGTC GACACGACT GCTGGTACCAC GCTGGTACCA CTACAGACT TGACTCCCCG TATCAGCAAT TGACTCCCCAAC CAGTGTATAG GGTTCCCAAC CAGTGTTATC AGTAGTATAT GACTCCCCG TATCAGCAAT TGACTACTAGCAAC CAGTGTTATC AGTAGTATAT GACTCCCAAC CAGTGTTATC CATCTGAGAA GGAAAACGTT	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTCG AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGAGCG AGACATGTG AAATCGAGCG AGACATGTG AAATCGAGCT ATCGCCAGTTA ACCCACTGG TGGTTTTTT AAACTCACGT ATACTCAGATA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA GATCAAGGCG ACTCATGGTT TTGGTAGGT TAGGTTATGC CTTCGGGGCG	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAGCATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGCAGACCA TCAGTCAGA CCGGATACCA TATTTGGTAT TGTTTGCAAG TAGTCCGAA ACTACGATA CCGGAAGGG ATACTTTGCTAT ACTGCGTAGAC ACTACCGTA ACTACCGTA ACGCACCAC AGTTACATGA ATGCCACCAC AGGCACCAC AGGCACCAC AAAACTCTCA	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GGAGCCAGCT CGTTCGGCTG CGTCGGCTAG GGTCGGCAAAG GGTCGCGCTTT CACGAACCC CTGGTAACAG CTGCGCTCT CAGCAACAT TGGTCATACAG CAGTAACAG CAGTAACAG CAGTAACAG CAGTAACAG CAGTACACAG CAGCACATT TCGCCCATGT TCCCCCATGT TCCCCCATGT TCCCTCCTTCC CAGGACTTACT TTCCTCTTCC CAGGACTTATC CAGGACTTATC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAGCT 2901 CCAGGAAGCA CTATAAAG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CGGTCAGCA GCGAGGTA 3401 CTGAAGCCAG TTACCCTTC 3501 CGCGCAGCAG TTACCCTTC 3501 ATTAGCAGA AGCAGTA 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTA 4201 CCTTATCTC ATGCCATC 4301 CCGGCGTCAA TACCGGGTA 4401 CGCTGTTCAGA ATCCAGGTA 4401 CGCTGTTCAGA ATCCAGGTA 4401 CGCTGTTCAGA ATCCAGGTA 4401 CGCTGTTCAGA ATCCAGGTA	TA AACCCGCTG AT TGCTTTATTT TG AGGTCAGGG TG GGCGTAATCAG TG GCCTAATGAG CG CGCGGGGAG CA CTCAAAGGCG CG GCGTTGCTGG CG GCCTTTCTCAT CG GCCTTATCCG CG AAAAAGAGTCT CT CAAGAAGATCT CT CAAGAAGATCT CT CAAGAAGATCCT CTACTCAGCG CT GCAATGATACC CT GCAATGATACC CT GCAATGATACC CGCCTTCCAT CGCCTTCCAT CGCTTCAGCG CT GAAGAAGATC CT TATCTCAGCG CT TATCTCAGCG CT TATCAGCG CT TAAGATCCT CATTCAGCG CT TAAGATCCT CATTCAGCG CT TAAGATCCT CATTCAGCG CT TAAGATCCT CATTCAGCG CT TCCTTCAGCG CT TCCTTCAGCG CT TCCTTCAGCG CT TCCTTCAGCC CATTCAGCCC CATTCAGCCC CATTCAGCCC CATTCAGCCC CATTCAGCCC CATTCAGCCC CATTCAGCCC CATTCAGCC CATTCAGCCC CATTCAGCC CATTCAGC CATTCAGCC CATTCAGCC CATTCAGCC CATTCAGC CATTCAGCC CATTCAGCC CATTCAGCC CATTCAGCC CATTCAGC CATTCAGCC CATTCAGC CATTCAGC CATTCAGC CATTCAGCC CATTCAGC CATTCAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTATCG GTAATACGGT GTAATACGGT GTAATACGGT GTAACAGGT TCCCCTGGA AGCTCACCCT GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCAGTCTAT CGCGAGACCC CCAGTCTATT CGTCGTTTG CTCCGATCGT TCCTGTGACC CATCGTGACCACACACACACACACACACACACACACACAC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGGTCCGC AGCTCCCTCG GATAGGTATC TCTTGAGTCC TTGAAGTGGT TCTACGGGA AATGAAGTT TTCCTTCATC ACGCTCACCG AATTGTTGCC GTATGGCTTC TGTCAGAAGT TTCTCATC ACGCTCACCG AATTGTTGCC TGTCAGAAGT TTTCATAAAAGT CTTTAAAAAGT CAACTGATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCTGACG TGGGCTTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGTC TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG GTCCAGATT GCGAAGCTAC ATTCAGCTC AAGTTGCCC CACCCAGTT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTACCT GCTACACT CGCTACACT TGACTCCCCC TATCAGCAAT AGTAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA CAGTCTCAAC CAGTGTTATC ATTCTGACAA CAGTGTTATC ATTCTGACAAA CTTAACTTTCTACATT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG ACTGACTGGC AGAACATGTG AAATGGAGGC CTGCCGGCTTA AGAGGACACTG AGAAGATTTT AAACTCACGT TATGAGTAAA TCGTGAGTAAA TCGTGAGTAAA AACCAGCCA TCGCCAGTTA CATCAGGTAGAT AAACCAGCCA TCGCCAGTTA CATCAGGGAGA CATCATGGTTATGGTATAAA CATCAAGGCG ACTCATGGTT TAGGTATAGC CATCATGGTT TAGGTGTATAGC CATCATGGTT TAGGTGTATGC CTTCGGGGGG CACGCTTTCT	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGAATACCT AGTTTGGTTGT TGTTTGCAAG TAAGGGATTCT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATACTTTCGTAT ACGGAAGGG ATACTTTCAGA ATGCAGCA ATGCAGCA ATGCAGCA ATGCAGCA AGGACCAG AAACTCTCA GGGGACCAG GAAACCTCCA GGGTGACCAC	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GCAGCAGCT GCTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGCAACCC CTGGTAACAG CTGGCATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGGAGGCT CCGAGCGCT CCGAGCGCAT TCCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCCTTTGC AGGATCTTACCAA AGGATCTTTC AGGATCTTTC
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCCGCCA 2801 CGGCGACGGG TATCAAGCT 2901 CCAGCAAGGA CTATCAAGG 3101 CCCCACAGGA CTATCAAGG 3201 CCGTTCAGC CGACCGCT 3201 CGGTTCAGCC CGACCGCT 3301 GATTAGCAGA GCAGGTTA 3401 CTGCAGCAGAAA AAAAGGAT 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AGTGGTCTC CAGCCATG 3901 AGTGGTCTC CAGCGTT 4001 GCCATTGCTA CAGCGATC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTGTC ATGCCATC 4301 CCGCGTCAA TACGGGAT 4401 CGCTGTTGAG ATCCAGTT 4401 CGCTGTTGAG ATCCAGTT 4401 CGCAGTTAGA	TAAACCGCTG TAATCAGG TG AGGTTATTT TG AGGTTCAGGG TG GGCTAATCAG GC GCGGGGAG AC CCCAGGGGGAG AC CCCAGGCGTT TCCAGGGCTT TCCAGGCGTT TCCAGGCGTT TCCAGGCGTT TCCAGGCGTT TCCAGGAGT TCCAGGAGT TCCAGGAGT TCCAGGAGT TCCAGGAT TCCAGGAT TCCAGGAT TCCAGGAT TCCAGGAT TCCAGGT TCCAGGT TCCAGGT TCCAGGT TCCAGGT TCCTCCGGT TCCTCCGGT TCCTCCGGT TAACGCCCA TAACAGAGCT TAACGCGCCA TAACAGCGCCA TAACAGCGCCA TAACAGCGCCA TAACACCCA TCCATCAGGT TAACCCCAC TCCATCAGGT TAACCCCAC TCCATCAGGT TAACACCCA TCCATCAGGT TCCACGCCCA TAACACCCA TCCACGCCCA TCCACGCCA TCCACCA TCCACCAC	ATCAGCTTG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGTTTG GTAATACGGT CGTTTTTTCCA AGCTGACTACC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTTTT TCTTGATCTT TTTAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCTGTGACT CTCTGGATCGT CTCTGGATCGT CATAGCAGAA CCGGACAGAA CTCGTGCACC GACACGGAAA	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGGA ATTCTACGGCA ATTCTACGGCA ATTCTACGGC AATTGTTCAC GGAGTACT CGTAGAGGT TCTCAGAGGT TTCAGAGGT TCTCAGAGGT CGTAGAGGT CGTAGAGGT CGTAGAGGT CTTTAAAAGT CAACTGATCT TGTTGAAAGT CAACTGATCT TGTTGAAATAC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG GCCCTGACG ACCAGGTAA ACCACGTAA ACAAACCAC CTTGAACTA CTAAATCAAC CTAAGTTCGCG GCTCACAGTT GCGAACCTAC ATCAGGTCC AATTCAGCTCC AAGTTGCCG CTCCAGATT GCGAACCTAC ATCAGCTTC CTACATCT CTACACTTC CTACATCT CTCATCATT CCACATCT TCACATCTT TCACATCTT TCATCTCT	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCACAGGCG ACTCATGATT TAGTGTATAGC CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GCAGCATT CACGAACCC CTGGTAACAG CTGCGCTTC CAGCAGATT TGGTCATGAG CAGTTACCAA CGGGAGGCCAG CAGTTACCAA CGGGAGGCCAG TCGCATTGTT TCCCCCATGT TCCCCCATGT TTGGTCTTGC AGAACTTTAC AAACAGGAAG GGGTTATTGT AAACAGGAAG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGG ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA AGCCTGGG 2701 GCATTAATGA ATGCGCA 2801 CGGCGAGCGG TATCAAGCT 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAACCGTG 3201 CGGTCAGCA GCGAGGTA 3401 CTGAAGCCAG TTACCTTC 3501 CGCGCAGAAA AAAGGAT 3601 ATTATCAAAA AGGATCTT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTGC ATGCCATC 4101 CCGGCGTCAA TACCGGCT 4401 CGCGTTTGAG ATCCAGTT 4401 CGCGGTTCAA TACCGGAT 4401 CGCGGTTCAA TACCGGAT 4401 GCCATTGCTCA CGCATC 4101 CCGGCGTCAA TACCGGTT 4501 GCAAAATGCC GCAAAAAA 4601 CTCATTGAGCG GCAAAAAA	TAAACCGCTG TATTTTTTTTTTTTTTTTTTTTTTTTTTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTACTG GTAATACGGT CGTTTTTCCG AGCTCACCGT TACACGATC GTAACTATCG GTAGCTCTC TTCACTCTT TTTGATCTT TTTAATTTA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCGATCGT TTCTGTGTTT CTCGATCGT CATAGCAGAA CTCGTGCACC GACACGAAA TAGAAAAATA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGG TAGCTCCGC GTAGGTACTCT CTTGAGTCC TTGAAGTGC TTGAAGTGT TTGATTCC TAGAGTGT TTCGTTCACGG AATCAAGAGT TTCGTCACC GTAGGTATCT TTCGTTCAT ACGCTCACCG GTATGGCTTC TGTCAGAAGT TGTTCAGAAGT CACTGACTAC TGTTGAATAC AAACAAATAGG	CTAAGATCCA GCTTATTTG GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCAACTA ACACACCAC ACTACGTC CAGTTCGCT CTGACGCT CTCAGACGTC CTCAGACGTC CTCAGACGTC ACTAGTTGCC GCTCCAGATT GGGAAGCTAC ATTCAGCTC AAGTTGCCC CACCAGATT CGCACAGTT CAGCACTC CACCAAGTC CACCAAGTC CACCAAGTC CCTCATCATT TCAGCATCTT TCAGCATCTT TCATACTCTT TCATACCTCC GCTTCCCCGC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCACAGGCG ACTCATGATT TAGTGTATAGC CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GCACCTT CACGAACCC CTGGTAACAG CTGCGCTTC CAGCAGATT TGGTCATGAG CAGTTACCAA CGGGAGGCCAG CAGTTACCAA CGGGAGGCCAG TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AAACAGGAAG GGGTTATTGT AAACAGGAAG
> D T 2101 AGCACAGTGG CGGCCGCT 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCCGCCA 2801 CGGCGACGGG TATCAAGCT 2901 CCAGCAAGGA CTATCAAGG 3101 CCCCACAGGA CTATCAAGG 3201 CCGTTCAGC CGACCGCT 3201 CGGTTCAGCC CGACCGCT 3301 GATTAGCAGA GCAGGTTA 3401 CTGCAGCAGAAA AAAAGGAT 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGCA 3801 TACCATCTGG CCCCAGTG 3901 AGTGGTCTC CAGCCATG 3901 AGTGGTCTC CAGCGTT 4001 GCCATTGCTA CAGCGATC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTGTC ATGCCATC 4301 CCGCGTCAA TACGGGAT 4401 CGCTGTTGAG ATCCAGTT 4401 CGCTGTTGAG ATCCAGTT 4401 CGCAGTTAGA	TAAACCGCTG TATTTTTTTTTTTTTTTTTTTTTTTTTTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTACTG GTAATACGGT CGTTTTTCCG AGCTCACCGT TACACGATC GTAACTATCG GTAGCTCTC TTCACTCTT TTTGATCTT TTTAATTTA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCGATCGT TTCTGTGTTT CTCGATCGT CATAGCAGAA CTCGTGCACC GACACGAAA TAGAAAAATA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGG TAGCTCCGC GTAGGTACTCT CTTGAGTCC TTGAAGTGC TTGAAGTGT TTGATTCC TAGAGTGT TTCGTTCACGG AATCAAGAGT TTCGTCACC GTAGGTATCT TTCGTTCAT ACGCTCACCG GTATGGCTTC TGTCAGAAGT TGTTCAGAAGT CACTGACTAC TGTTGAATAC AAACAAATAGG	CTAAGATCCA GCTTATTTG GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCAACTA ACACACCAC ACTACGTC CAGTTCGCT CTGACGCT CTCAGACGTC CTCAGACGTC CTCAGACGTC ACTAGTTGCC GCTCCAGATT GGGAAGCTAC ATTCAGCTC AAGTTGCCC CACCAGATT CGCACAGTT CAGCACTC CACCAAGTC CACCAAGTC CACCAAGTC CCTCATCATT TCAGCATCTT TCAGCATCTT TCATACTCTT TCATACCTCC GCTTCCCCGC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCACAGGCG ACTCATGATT TAGTGTATAGC CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GCACCTT CACGAACCC CTGGTAACAG CTGCGCTTC CAGCAGATT TGGTCATGAG CAGTTACCAA CGGGAGGCCAG CAGTTACCAA CGGGAGGCCAG TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AAACAGGAAG GGGTTATTGT AAACAGGAAG

### pMT-HisFlag-DmSNAP43 mut#2 STOP

1 TCGCGCGTTT CGGTC	GATGAC GGTGAAAAC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGG								
201 CCGCACAGAT GCGTA								
301 TACGCCAGCT GGCGA								
401 TGAATTAATT CGTTC								
501 GGCCCCCAC CGCC								
601 CAAGTCCCCA AAGTC								
701 AGAGGTGAAT CGAAG								
801 AAATCAAGTG AATCA								
801 AAAICAAGIG AAICA	AICICA GIGCAACIAA	AGGGGGGAIC	TAGATCGGGG	TACTAAGATT	> M H	H H H	H H T D	Y K D
001 0200200202 2000	33.0M3.0 M03.00M03.30	3 mamma 3 aa	3 CMCCMCCC3	aamaamaa a				
901 GACGATGACA AGGGG							D G E	N C E
1001 TCGAGGTGTT CTGC		I F D				R L V N		
			H L F		T N H T	E V I		L A A L
1101 GCATGTGGCC AAGCC								
> H V A K F			G D V	F P A S		RIG	G F F L	L Y V
1201 ATCTACTACA AGCAC								
~	PTHNF	IKI		R T W	~	T D Y A		R K D
1301 GTCCGGAGCG GAAGG								
	D T H Q I	AYML		T Q E	Q A F R		L D Y	C Q G L
1401 GGACAATCTG GTGGA								
> D N L V I			G A K			M Q K		N G V
1501 AGTCTCACAT ACGA								
	L E G L R	ALD				A A A A		K Q L
1601 CGGCTGGTCA TGAGO								
	H A L P P	S Q I F		R E V		Q S V	L G A	R K S T
1701 TCCAGATGAG AAATG								
> P D E K 0		~	L E V	~		A M Y	G V E E	R E P
1801 CAACACCAGA CGGAT								
	E L E V Q	L E V				A T V F	Q R E	L P E
1901 ACGTGCAGCA AGAGT	TATGAG ATGATTGAGT	TTACTCACCA	CCACCAAATC	CAACTCCCTC	AAAGCGAGGA	CCTCACCCAA	GAAGAACTCA	አአርርጥ አጥጥጥጥ
1701 1100100110011 110110.					HHHOCOHOOH	GGICHCGGH	0.1.0.1.0.	MMGCIMIIII
>D V Q Q E	Y E M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
	Y E M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
>D V Q Q E	Y E M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
>D V Q Q E 2001 GGATACTTGA GTTAG	Y E M I E GTTTAT AAAACTTTTA	F S D D CATAATTAAA	E E M TAACTAGCAT	E V G TTTTGCGCGA	E S E E TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
>D V Q Q E 2001 GGATACTTGA GTTAC > D T	Y E M I E GTTTAT AAAACTTTA CGCTCG AGTCTAGAGG	F S D D CATAATTAAA GCCCGCGGTT	E E M TAACTAGCAT CGAAGGTAAG	E V G TTTTGCGCGA CCTATCCCTA	E S E E TGTGATCTTG	V T E TTTATCTGAA CGGTCTCGAT	E E L GGGCAATTCT TCTACGCGTA	K A I L GCAGATATCC
>D V Q Q E 2001 GGATACTTGA GTTAC > D T 2101 AGCACAGTGG CGGCC	Y E M I E GTTTAT AAAACTTTTA CGCTCG AGTCTAGAGC GAGTTT AAACCCGCTC	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA
>D V Q Q E 2001 GGATACTTGA GTTAC > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTC	Y E M I E GTTTAT AAAACTTTA CGCTCG AGTCTAGAGG GAGTTT AAACCCGCTC AAAAAA TGCTTTATT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
>D V Q Q E 2001 GGATACTTGA GTTAG > D T 2101 AGCACAGTGG CGGCG 2201 TCACCATCAC CATTG 2301 CTAGAATGCA GTGAA	Y E M I E GTTTAT AAAACTTTA CGCTCG AGTCTAGAGG GAGTTT AAACCCGCTC AAAAAA TGCTTTATTI GTTTC AGGTTCAGGG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
>D V Q Q E 2001 GGATACTTGA GTTAC > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTY 2301 CTAGAATGCA GTGAJ 2401 TTGCATTCAT TTTA:	Y E M I E GTTTAT AAAACTTTI CGCTCG AGTCTAGAGG GAGTTT AAACCCGCTC AAAAAA TGCTTTATTI TGTTTC AGGTTCAGG AAGCTT GGCGTAATCI	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGGAGGTGTGG TGGTCATAGC	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
>D V Q Q E 2001 GGATACTTGA GTTAC > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTC 2301 CTAGAATGCA GTGAJ 2401 TTGCATTCAT TTTAC 2501 ACCTGCAGGC ATGCJ	Y E M I E GTTTAT AAAACTTTA CGCTCG AGTCTAGAGG GAGTTT AAACCCGCTC AAAAAA TGCTTTATT TGTTTC AGGTCAGGC AAGCTT GGGGTAATGAC TGGGGT GCCTAATGAC	F S D D CATAATTAAA GGCCCGCGGTT ATCAGCCTCG GGAGAATTTG GGAGGTGTGG ATGGTCATAGC TGAGCTAACT	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
DVQQQETACTTAC DT 2001 GGATACTTGA GTTAC DT 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTC 2301 CTAGAATGCA GTGAA 2401 TTGCATTCAT TTTAC 2501 ACCTGCAGGC ATGCA 2601 TAAACTGTAA AGCC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCGCTT AAAAAAA TAGCTTTATTI TGTTTC AGGTTAATCI TGGGGT GCCTAATGAG GCGCAGGAG GCCAAG GCGCGGGGA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TAGCCTATTG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCGCTT AAAAAAA TGCTTTATTI TGTTTC AGGTTCAGG AAGCTT GGCGTAATCI TGGGGT GCCTAATGA GCCAAC GCCGGGGAA AGCTCA CTCAAAGGCC	F S D D CATAATTAAA GGCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTGAAACT GGAGGTTTG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TCTTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CGTTCGGCTG CAGCAAAAGG
>D V Q Q E 2001 GGATACTTGA GTTAC > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTAC 2301 CTAGAATGCA GTGA 2401 TTGCATTCAT TTTAC 2501 ACCTGCAGGC ATGC2 2601 TAAACTGTAA AGCCC 2701 GCATTAATGA ATCGC 2801 CGGCGAGCGG TATCC	Y E M I E SITTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCCGCTC AAAAAA TGCTTTATTI TGTTC AGGTTCAGG AAGCTT GCCTAATGA GCCAAC GCGCGGGAC AGCTCA CTCAAAGGCC AAGGC GCGTTGCTG GCGAAGGC GAGGC GGGTTGCTG GCGAAGGC GGGTTGCTG GCGAAGGC GGGTTGCTG GGGGGAC GGGTTGCTG GGGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGAC GGGGAC GGGGGAC GGGGCGAC GGGGGAC GGGGGAC GGGGGAC GGGGGAC GGGGGAC GGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGAC GGGGGAC GGGGCGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGAC GGGGGGAC GGGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGGAC GGGGGCGGCGC GGGGCGCGGCGCGC GGGGCGCGGCC GGGGCGCGCGGCC GGGGCGCGGCC GGGGCC GGGGCGCGCGCGCC GGGGCC GGGGCGCGCGCGCC GGGGCC GGGGCC GGGGCC GGGGCGCGCGCC GGGCGCGCGCC GGGCGCGCGCC GGGCGCC GGGCGCGCC GGGCCC GGCCC G	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGGAGTTGG TGGTCATAGC TGAGCTAACT AGGCGTTTG AGGCGTTTG GTAATACGGT GCTATACGGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAAAAGGC TCAAGTCAGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA
>D V Q Q E E 2001 GGATACTGA GTTAC   > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCA CATTC 2301 CTAGAATGCA GTGA 2401 TTGCATTCAT TTTA 2501 ACCTGCAGGC ATGC 2701 GCATTAATGA ATGC 2801 CGGCGAGCGG TATC 2901 CCAGGAACGC TAAA 2901 CCAGGAACGC TAAA 2501 CCAGGAACG TAAA 2501 CCAGGAACGC TAAA 2501 CCAGGAACG TAACC TAACC TAACA 2501 CCAGGAACG TAA	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCCGCTT AAAAAAA TGCTTTATTI TGTTTC AGGTTAATCI TGGGGT GCCTAATGAG GCCAAC GCGCGGAA AGCTCA CTCAAAGGC AAAGAC ACGTCA CTCAAAGGC AAAGAA ACCAACACGCCGT AAAAGAA ACCAACACGCAAAAAAAAAAAAAAAAAAAAA	F S D D A CATAATTAAA G GCCCGCGGTT ATCAGCCTCG G GGAAATTG G GGAGGTGTG TGGTCATAGC TGAGCTAACT G AGGCGGTTTG G GTAATACGGT C CGTTTTTCCA TCCCCTGGA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT CGTATTCGTG CACATTAATT CGTATTGGC TATCCACAG AGCTCCCTCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCCTGACG TGCGCTCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCGC TCCCTCGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGAGTCGA ACTGACTCGC AGAACATGTG AGAACATGTG AAATTCGACGC CTGCCGCTTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAA GTCCGCCTTT
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCGCTT AAAAAAA TGCTTTATTT TGTTTC AGGTTCAGGG AAGCTT GGCGTAATGA GGCAAC GGCGGGGGA AGCTCA CTCAAAGGC AAGGCC GCGTTCTCA AAAGAT ACCAGGCGT CGTGG GCTTCTCTCA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GGAAATTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGAGCTACG GGAGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCG AGCTCCCCG GTAGGTATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCCGC CAGTTCGGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCGGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGACC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACAC TTTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGGCTA AATCGACGC CTGCCGGCTA GCTCCAACCT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAAGG GGTGGGCAAA GGTCGGCCATT CACGAACCCC
>D V Q Q E EACTOR STATE	Y E M I E STITTAT AAAACTTTI CCCTCC AGTCTAGACC SAGTTT AAACCCGCTC AAAAAA TGCTTTATTI TGTTC AGGTTCAGC AAGCTT GCCTAATGA GCCAAC GCGCGGGAC AGCTCA CTCAAAGGCC AAGGCC GCGTTGCTG CAAAGGC GCGTTCCTG CAAAGGC GCGTTCCTG CGTGGC GCTTTCCTC	F S D D CATAATTAAA GCCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTCATAGC TGGTCATAGC TGAGCTAACT GGTAAATACGGT GGTAATACGGT CGTATTCCA TCCCCTGGA AGCTCACGCT GTAACTACG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGGTTTTT TGTTTCCTGT CACATTAATT CGTATTCACAGA TAGCCTCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGGTCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTC TAGGTCGTC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATCGATCA CAATTCCACA ACTACTCGC ACTACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTGTG GCACCAGCCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCC CTGGTAACAG
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCCGCTT AAAAAAA TGCTTTATT TGTTTC AGGTTAATC! TGGGGT GCCTAATGAG GCCAAC GCGCGGGAA AGCTC CTCAAAGGC AAGGCC GCGTTGCTG AAAGAT ACCAGGCGTI CGTGG GCTTTCTCAT CGCTGC GCCTTATCA GCGTTGC GCCTTATCCC GGTTATG TAGGCGGTG GGTTATG TAGGCGGTG GGTTATG TAGGCGGTGC GGTTATG TAGGCGGTGC	F S D D CATAATTAAA GCCCCGGGTT ATCAGCCTCG GGAAATTG GGAGGTGGT TGAGCATCG TGAGTAACT AGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT TCACCACTCGCT TACACAGTTC TACACAGTTC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACACTAATT CGTATTGGG TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA AGCCGGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACGTGCAAT AAATGTGGTA CAATTCCACA TTTCCACAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC TGCCGCTTA GCTCCAAGCT ATCGCCACTG ACAAGCACTACCACCACTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTG CAGCAAAAGG GGTGCGAAAA GTCCGCCTTT CACGAACCCC CTTGGTAACAG CTGCGCTCTG
>D V Q Q E EACTOR STATE	Y E M I E STTTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCGGCTT AAAAAAA TGCTTTATTI TGTTTC AGGTTCAGGG AAGCTT GGGGTAATCA TGGGGT GCCTAATGAG GCCAAC GGGGGGGAA AGGCTC CTCAAAGGCC AAGGCC GCGTTGCTG AAAAGAT ACCAGGGGT CGTGGC GCTTTCTCAT CGCTGC GCCTTATCCC GGGTAGC TAGGCGTGC CTTCGG AAAAACAGTC CTTCGG AAAAACAGTC CTTCGG AAAAACAGTC	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GGAGAATTG GGAGGTGTGG GGAGGTGTGG GGAGGTGTGG GGAGGTTTG GTAATACCTACT GTAATACGT GCTAATACGT GCTAATACGGT GCTAATACGGT GCTAATACGGT GCTAACTACC GGTAACTACCG GGTAACTACCG GGTACCTCGC GGTACCTCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGCTCCGC GTAGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC GATCCGCGCA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCGTTCGGCT ATCAGGGGAT CCCCTGACG ATCAGGGGAT ACCAGGTAACTA ACCACACACAACAA ACAAACCACC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCGCTCA CACTGCCGC CACGCAGA AGCATCACAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT CGGCTACACT CGGCTACACC CCTGGTACCC CCTGGTACCC CCTGGTACCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG ACTGACACTGC AGAACATGTG AAATCGACGC TTGCCGCTTA GCTCCAGCTTA GCTCCAGCT ATCGCCACTG AGAAGACAGCT ATCGCCACTG AGAAGACAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCCA TATTTGGTAT TGTTTCCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG CCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGCGAA GGTCGCGCTT CACGAACCC CTGGTAACAG CTGCGCTTCACGTTACAG CTGCGCTCTC
>D V Q Q E 2001 GGATACTTGA GTTAC  > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTCL 2301 CTAGAATGCA GTGAI 2401 TTGCATTCAT TTTA: 2501 ACCTGCAGGC ATGC: 2601 TAAAGTGTAA AGCC: 2701 GCATTAATGA ATCGC 2801 CGGCGAGGGG TAC: 2901 CCAGCAACG TAAA: 3101 CTCCCTTCGG GAAGC 3201 CAGTTAGCAG CGACC 3401 CAGTAGCAG TGAGA 3401 CTGAAGCCAG TTACC 3401 CTGAAGCCAG TTACC	Y E M I E STITAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTIT AAACCCGCTC AAAAAAA TGCTTTATI TGTTTC AGGTTCAGG AAGCTT GCCTAATGAC GCCAAC GCCCGGGAC AAGGCC GCTACTGC AAAGAT ACCAGGCGT CGTGAG GCCTTTCTCAT CGCTGG GCCTTTCTCAT CGGGT GCTTTCTCAT CGGTAG GCCTTATCCC GGTATG TAGGCGGTG CTTCGG AAAAAAAAAGAGT CCTTCGG AAAAAAAAAGT CGTTCCATACCC GGTATG TAGGCGGTGC CTTCGG AAAAAAAAAGT GGATCT CAAGAAGATT	F S D D CATAATTAAA GCCCCGGGTT ATCAGCCTCG GTGAAATTG GGAGGTGGA TGGTCATAGC TGAGCTAACT AGGCGGTTTG GGTAAATACGGT CGTATACGGT CGTATACGGT TCCCCTGGA AGCTCACGCT TAACTATCG TAACTATCG TAACTATCG TAACTATCG TAACTATCG TAACTATCG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTAATTAGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAAGTGGT GATCCGCCA TTCTTCGGGG TATCCGCCA TTCTTCGGGG TTCTTCGGGGGATTCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTGCCC CAGTTCGGT AACCCGCTAAC GCCTAACA GCCTAACA GCCTAACA CACCGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGGCC TCCCTCGGCC AACGCAGGAA TGTTCCGACC TAGGTCGTC GACACGACT TCGGCTACACT GGCTACACT GGTGGTACGAC AGTTGGAACG AGTTGGAACG AGTTGGAACG AGTTGGAACG	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA TTTCCAGTG ACTGACTGG AGAACATGTG AGAACATGTG GCTCCAAGCT ATCGCCAGTTA AGAAGAGACAG GCTCCAAGCT ATCGCCACTT AGAAGGACAG GTGGTTTTT AAACTCACAGTT AAACTCACAGTT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CGACTACGA AGGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGATACCT GGGCTCGTGTG GGGCTCGTTGTG GCACCAGCCA TATTTGCTAT TGTTTGCAAG TAAAGGGATTT	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGCTAACAC CTGGGCTCTG CAGCAGATTA
>D V Q Q E EACTOR   2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCGCTC AAAAAAA TGCTTTATT IGTTTC AGGTTCAGG AAGCTT GGCGTAATC; GGGGT GCCTAATGAG GCCAAC GCGCGGGAA AGCTC CTCAAAGGC AAAGGC GCGTTCCTGC GAAAGGC GCTTCTCAA CGCTGC GCCTTATCCC GGTTAT TAGGCGGTGC TTCGG AAAAAGAGT GTTCG AAAAAGAGT GGATCT CAAGAACGT TCTTCA CCTAGATCCT CTTCA CCTAGATCCT CTAGATCCT CTAGATCT CTAGATCCT CTAGATCT CTAGATC CTAGATCT CT	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GGAAATTG GGAGGTGG TGAGATTG GGAGGTGTG GGAGTTTG GGAGGTTTG GGAGTTTTTCCA TCCCCCTGGA TCCCCCTGGA TCCCCCTGGA TCCCCCTGGA TCCCCTGGA TCCCCTGGA TCCCCTTGGA TCCCCTTGA TCCCCTTGA TTCCCCTTGT TTTAAATTAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACCACAGA TAGCGTCAGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAGTGGG ACTCCGCAA TTCTACGGG ATCCGCCAA TTCTACGGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTA ACCACGGTAA ACCACGCTAA CTAAACCAC TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TTAAATCAAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCACAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC TGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATCACGT AAACTCACGT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGC GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTCCTTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA GTCCGCCTTT CAGCAACCC CTTGGTAACAG CTGGTAACAG CTGGGCTCTG CAGCAGATTA TGGTCATGATCA CAGCAGATTA TGGTCATGAGC CAGTTATCAG CAGTTACCAA
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCGGCTT AAAACAA AGAAA TGCTTTATTI TGTTTC AGGTTCAGGG AAGCTT GGGGTAATCA TGGGGT GCCTAATGAG GCCAAC GCGCGGGAA AGGCC CTCAAAGGCC AAAGGAT CCCAGGGGT CGTGGC GCCTTATCCC GGTGG GCTTTCTCAT TGGGGTG GCCTTATCCC GGTAGT CAAGAGAT TCTCGG AAAAAGAT TCTTCAG GGACC TATCTCAGCCTCC TCTCAGATCC TCTCAGATCCC TCTCACACC TCTCAGATCCC TCTCACACC TCT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GGGAAATTG GGGGGTAGC GGAGGTGTG GGAGCTAGC TGAGCTAGC TGAGCTAGC AGCCCTGGA AGCCCTGGA AGCTCACGCT GTAACACACGCT GTAACACACC TACAGAGTC GGTAGCTCT CTTTGATCTT TTTAAATTAA ATCTGTCTAT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGGCTCCTC GTAGGTACT TCTTGAGTCC TTGAAGTC TTTAAGGGG AATCAGAGA TTTACGGGAACT TCTTACGGGG AATCAGAGT TTCCTTCATC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGT AACCCGGTAA ACCACACTA ACAAACCAC TCTGACGCT CTGACGCT TTAAATCAAT CATAGTTGCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC ACCGGCACA AGCATCACCA AGCATCACCA CGCTGCACCACCACCACCACCACCACCACCACCACCACCACCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACATTGA AAATGGTAA CAATTCCACA TTTCCAGTCG ACTGACACTGC ACTGCCGCTTA GCTCCAGCGTTA GCTCCAGCT ATCCCCACTG AGAACATTTT ATCAGAGC TTGCTCAAGCT ATCCCACTG TTTTT AAACTCACGT TATGAGTAA TATGAGGAC TTGGTTTTTT AAACTCACGT TATGAGTAAA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA GGGCTAATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGT GCACCACCA TATTTGGTT TGTTTGCAAG TAAGGGATTT CTTTGCTCTGAAG TAAGGGATTT CTTTGCTCTGAAG TAAGGGATTT CTTTGCTCTGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTCCCAGCT CGTCCGCTT CAGCAAAAGG GCTCGGCTT CACGAACCC CTGGTAACAG CTGCGCTT CAGCAGATTA TGGTCATGAG CAGTTACCAG CAGTTACCAG CAGCAGATTA TGGTCATGAG CAGTTACCAG
>D V Q Q E 2001 GGATACTTGA GTTAC  > D T 2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTY 2301 CTAGAATGCA GTGAJ 2401 TTGCATTCAT TTTAC 2501 ACCTGCAGGC ATGCJ 2601 TAAAGTGTAA AGCCC 2701 GCATTAATGA ATCGC 2801 CGGCGAGGCG TAACJ 3001 CCCGACAGGA CTATJ 3101 CTCCCTTCGG GAAGC 3201 CGGTTCAGCC CGACC 3301 GATTAGCAGA CGGAC 3401 CTGAAGCCAG TTACC 3501 CGCGCAGAAA AAAA 3601 ATTATCAAAA AGGAC 3701 TGCTTAATCA GTGAC	Y E M I E STITAT AAAACTTTI CGCTCG AGTCTAGAGG 3AGTTT AAACCCGCTC AAAAAAA TGCTTTATI TGTTTC AGGTTCAGGG AAGCTT GCGTAATGA GCCAAC GCCCAGGGAC AAAGAC GCCTAATGAC CGTAGC GCTTTCTCA CGTGGG GCTTTCTCA CGCTGG GCTTTCTCA CGCTGG GCTTTCTCA CGTAGC GCTTTTCTCA TAGGCGGTT CCTTCG AAAAAGAGT TCTTCA CAAAAGAGT TCTTCA CCTAGATCCT GGCACC TATCTCAC AGGCC TATCTCAC GGCACC TATCTCAC AGGCACC TATCTCAC AGGCACC TATCTCAC CGCATGTC GCAATGATAC	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTGG GTGAAATTG GGAGGTGGA TGGTCATAGC TGAGCATTG GGAAACTTG GGAAACTTG GTAAATACGGT CGTATACGT GTAATACGGT TCCCCTGGA AGCTCAGCG TAACTATCG TAACAGAGTTC GGTAACTATCG TTAAATTACA TCCCCTGGA AGCTCACGCT TTTAAATTAA ACTGTCTTT TTTAAATTAA ACTGTCTCTAT CCCCAGAGACCC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT TGATTCCTGT TATCACAGA TAGCTCCCC AGCTCCCCG GTAGGTATC TCTTGAGTCC TCTTGAGTCC TCTGAGTAGT TCTTGAGTC TCTGAGTC TCTGAGTC TCTGAGTC TCTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTCATC ACGCTACACG AATGAAGTT TCCTTCACG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTAC ACCAGGTAACACCA CCTAACACAC CTTAAACCAC TTAAATCAAT CATAGTTGCC GTTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGTTGCGTC CATAGTTGCC GCTCCAGATT	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGGTCATCAC TAGGTCATC GGCTACACT GGCTACACT GGCTACACT GCTGTAGCG CTAGAGTATAT TGGCTACCCGC TAGGTCACCCCCC TAGGTCCCCCC TAGGTCCCCCC TAGGTCCCCCC TAGGTCCACC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTCCACA CTTCCACTCACCACACACACACAC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CGACATACGA GGAAACGTT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTCGGT GGGCTCGGT GGGCTCGTG GGACTCAGT TATTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA AACTACGATA ACCACAGAAAGGC	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGTACCAC CAGCAGATTA
>D V Q Q E E 2011 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCGCTC AAAAAAA TGCTTTATT IGTTTC AGGTTCAGG AAGCTT GGCGTAATCA GGGGT GCCTAATGA GCCAAC GCGCGGGAA AACACT CTCAAAGGC AAAGGC GCGTTGCTGC CGCTGC GCCTTATCCA CGCTGC GCCTTATCC GGATCT CAAGAAGAT CTTCCA CAAGAAGAT CTTCCA CTAAGACC TCTTCCA CTAAGACC GGCACC TATCTCAGCC GGCACC TATCTCAGCC GGCACC TATCTCAGCC TTTTAT CCGCCTCCAT	F S D D CATAATTAAA GCCCCGGGTT ATCAGCCTCG GGAAATTG GGAGGTGGG TGAGCTAGC TGAGCAACT AGGCGGTTG AGGCGTTG AGGCGTTG AGGCGTTG AGGCGTTG TTAGACAACT CGTTATACGA TCCCCTGGA AGCTCACGCT TACAGAGTTC GGTAACTATCG TACAGAGTTC TTACAGAGTTC TTTTAAATTAA ATCTGTCTAT CCCGGAGACCC CCAGTCTATT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTGGT ATCCGCAA TTCTACGGG AAATGAAGTT TCGTTCATC AAATGAAGTT TCGTTCATC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGGTA AACCAGCT CTGACACC TCTGACG TTTCGGTA TACAAACCAC TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATG GCTCCAGATG TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATT GGGAAGCTAG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAA TGTTCCGACC TAGCTCGTC GACACAGCAC TGGTCGTC GACACGACT GGGTACACT GCTGGTACGC CTGGTACGC TAGATTAT TGACTCCCCC TAGATTAT TGACTCCCCCT TATCAGCATT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATCTG ACTGACTCGC AGACACTTG GCTCCAAGCT ATCGCCACTG AGAAGGACA GTGGTTTTTT AAACTCACGT TATCGAGTAAA TCGTGTAGAT AACCAGCCA TGGTTAGAT AACCAGCCA TGGTGTAGAT ACGCCAGTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACCAAGTTA TGGCTGATTA CACATACGA GGAAACGTT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTAGGGATTC TTTGCTAT CTTGGTCTGA ACTTACGATA GCCGGAAGGC ATATTTCGTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAAA GTCCGCTTT CAGCAACCC CTTGGTAACAG TGGTCAGCT CTGGTAACAG CTGGGCTCT CAGCAGATTA CTGGGCAGAT TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTAT
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCGCTC AAAAAA TGCTTTATTI TGTTTC AGGTTCAGGG AAGCTT GGGGTAATCA TGGGGT GCCTAATGAG GGCAAC GCGCGGGAA AGGTC CTCAAAGGC AAAGAT ACCAGGCGTI CGTGG GCTTTCTCAI CGCTGG GCTTTCTCAI TGTTCG AAAAAAGAT CTTCGG AAAAAGAT CTTCGG AAAAAGAT CTTCAG CCTAGATCCI GGCACC TATCTCAGC GGCACC TATCTCAGC GGCACC TATCTCAGC TGTTAGT CGCCTCCAG CATTGT GCGCTCCACA CATTGT GGCGTTCAGATCCI CATTGT GGCGCTCCAC CATTGT GGCGCTCCAC CATTGT GGCGTTCAGC CATTGT GGCGTTCAGC CATTGT GGCGCTCCAC CATTGT GGTGTCAGC	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GGAAATTG GGAGGTGTG GGAGGTGTG GGAGGTAGC TGAGCTACCT AGCCTAGC TCCCCTGGA TCCCCGGAGCCC CCGAGACCC CCGAGACCC CCCAGTCTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TATCACAGA AATGAAGTT TTCGTCACC GTAGCTCGCCAC GTAGCTCACCG GAACTCACCG GAACTCACCG GAACTCACCG GAATGTTCACCGGG AATGATTCTCCC GTAGGTACCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GCGAAATTG GCGTTCCGC ATCAGGGGAT CCCCCTGACG CCCCCTGACG ACCAGCACT ACCACACC CCCCGTAA ACCACC TCTGACGTC TCTGACGTC CCCCCGTAA CAAACCACC CCCCCAGAT ACAAACCACC CCCCCCAGAT CCCCCCCAGAT CCCCCCCAGAT CGCACCTCCCCCCCCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCATCAT TTCCTCGCTC ACCGCGC AGCAGGAA AGCATCACA AGCATCACA GCTGCACC TGGTCACCAC TGGTCACCAC GCTGGTACCAC GCTGGTACCA AGTAGACAT TGACTCCCC TATCAGACAT TGACTCCCC TATCAGCAAT TGACTCCCAAC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAACATGGAA AAATGGTAA CAATTCCACA TTTCCAGTCG ACTGACTGAACATGGAACATGGAACATGGAACATGGAGAACATGGAGAACAGCTTA AACCCACTGAACAGCTAAAACCAGCAAAACCAGCAAAACCAGCAAAACCAGCTAAAACCAGCTAAAACCAGCTAAAACCAGCTAAAACCAGCTAAAACCAGCTAAAACCAGCTAAAACCAGCAACAACCAAC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TCACGTCGGT AGCAAAACGC TCAAGTCAGA GCAGCCA CCGGATACCT GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGATTT CTTGGTAT GCCGAAAGGA AACTACGAACA AACTACGAACACAACAACAACAACAACAACAACAACAACAACAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGCTC CACCAAAAGG GGTGGCGAAA GTCCGCCTT CACCAACACC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGAGGCCAG CAAGGTTGTT TCCCCCATGT
>D V Q Q E E   2001 GGATACTTGA GTTAC   > D T   2101 AGCACAGTGG CGGCC   2201 TCACCATCAC CATTC   2301 CTAGAATGCA GTGAA   2401 TTGCATTCAT TTTAT   2501 ACCTGCAGGC ATGC   2601 TAAAGTGTAA AGCC   2701 GCATTAATGA ATGCC   2801 CCGGCAGGCG TATC   2901 CCAGGAACGC TAAAA   3001 CCCGACAGGA CTATT   3101 CTCCCTTCGG GAAG   3301 GATTAGCAGA CGGAA   3401 CTGAAGCCAG TTACC   3501 CGGCCAGAAA AAAAC   3601 ATTATCAAAA AGGGA   3701 TGCTTAATCA GTGAC   3901 AAGTGGTCCT   CCAC   3901 AAGTGGTCCT   4001 GCCATTGCTA CAGGC   4001 GCCATTGCTA CAGGC   4001 TGTGCAAAAA AGCGC   4001 TGTGCAAAAAA AGCGCC   4001 TGTGCAAAAAA AGCGCC   4001 TGTGCAAAAAA AGCGCC   4001 TGTGCCAAAAAA AGCCCC   4001 TGTGCCAAAAA AGCCCC   4001 TGTGCCAAAAAA AGCCCC   4001 TGTGCCAAAAAA AGCCC   4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAA AGCCC    4001 TGTGCCAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCCC    4001 TGTGCCAAAAAA AGCCC    4001 TGTGCCAAAAA AGCCC    4001 TGTGCCAAAAAA AGCCC    4	Y E M I E STITTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTIT AAACCCGCTC AAAAAAA TGCTTTATI TGTTTC AGGTTAGACG AAGCTT GCGTAATGA GCCAAC GCGCGGGAC AAAGAT ACCAGGGGAC AAAGAT ACCAGGGGT CGTTGCTG CGTTATG TAGGCGGTGCTAGTGC GGTATG TAGGCGGTGCTACTGC GGTATG TAGGCGGTGCTACTGC GGTATG TAGGCGGTG CTTCCG AAAAAGAGTT CTTCA CCTAGATCCT GGCACC TATCTCAGC AGTGCT GCATCTAGACC AGTGCT GCATCTAGACC AGTGCT GCATCTCAGT CTTTAG CCACTCCAT CTTTAG CGCCTCCAT CATCTTCAGTC AGTGCT GCATCTCAGT CTTTAG CGCCTCCAT CATCTTCAGT CTTTAGC GGTGTCACC STTAGC TCTTCAGTC TTTAGC TCTTCAGTC TCTTTAGC TCTTCAGTC TCTTTAGC TCTTCAGTC TCTTTAGC TCTTCAGTC TCTTTAGC TCTTCAGTC TCTTTCAGTC TCTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTCAGTC TCTTTCAGTC TCTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCAGTC TCTTTCA	F S D D C CATAATTAAA GCCCCGGGTT ATCAGCCTCG GTGAAATTG GGAGGTGGG TGGTCATAGC TGAGCTAACT AGGCGGTTG GTAAATACGGT CGTAAATACGGT CGTAAATACGGT TCCCCTGGA AGCTCACGCT TAAATATGGT TTAAATTAC TCCCCTGGA AGCTCACGCT TTAAATTAC TCCCCTGGA AGCTCACCCT CTTTGATCTT TTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TGGTCGTTTG TGGTGGTTTG TGGTGGTTTG TGGTAGTTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT GAGGTTTTT TGTTTCCTGT TATCCACAGA TAGCCTCCACG AGCTCCCTCG GTAGGTACT TCTTGAGTCC TCTGAAGTGGT GATCCGCA AATTGAGTC AATTGAGTC ACGCTACCG AATTGTTGCC GTATGGCTTC TGTCAGGAAGT TGTCAGGAAGT TGTCAGGAAGT TGTCAGCAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTACGT AACCAGCTACC CAGTTCGGTA ACCAGCTACT TGAAACCACC TCTGACGTAACCACC TCTAACGTAC TCTAACGATA GCCTAACT ACAACCACC TCTAACGTTA GGCAACT TTAAATCAAT CATAGTTGCC CCTCCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC AAGTTGGCCC	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGGAA GGTTCCGACC TACGTCACAC GACACGACAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTCCACA ATTTCCACT ACTTCCACTC ACTACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTCCAAGCT ACTCCAAGCT ACTCCACTC AGAAGGACAG GTGGTTTTT AGAAGTAAA TCGTGAAGT AAACTCACGT AAACCACGCA TCGCCAGTTA CGCCAGTTA CGCCAGTTA CGACCCAGTTA CAACCAGCCAGTTA CAACCAGCCAGTTA CAACCAGCCAGTTA CAACCAGCCAGTTA CACCAGCTA CACCAGTTA CACCAGTT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA AGGAAAACGT TCAAGTCAGA AGCAATACGA AGCAATACGA AGCAATACGA AGCAATACGT GGGCTCGGT GGGCTCGGT GGGCTCTGTG GGACACCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA ATGCAACGCA ATGTTTACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GCTGGCAAAAGG CTGGCAAAAGG CTGGCAAACG TGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTACACA CTGGTCATG CAGCAGATTA TGGTCATGACA CAGGAGGCT CCGAGGGCCC CCGAGGCCAC CAGGTGTT TCCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCGCTC AAAAAAA TGCTTTATT IGTTTC AGGTTCAGG AAGCTT GGCGTAATCA GGGGT GCCTAATGA GCCAAC GCGCGGGAA AACACT CTCAAAGGC AAAGGC GCGTTGCTGC CGCTAG GCGTTCCTCA CGCTGC GCCTTATCCA GGATCT CAAGAACGT TCTTCA AAAAAAAA ACCAGGCGT CTTCGG AAAAAGAGT GGATCT CAAGAACGT CCTTCGG TAACAACAC TCTTCA CCTAGATCC TCTTCA CCTAGATCC TCTTCA GCAATGATA CCTTCAGC CATCGT GCAATGATA CTTTAT CGGCCTCCAA CTTTAT CGGCCTCCAC CATCGT GGTGTCACG CTTAGT CTTCACTCGT CATCGT CTCTTCGGTC CATCGT TAAGATGCT CATCCT TAAGATCCT CATCCT TAAGATCCT CATCCT TAAGATGCT CATCCT TAAGATGCT CATCCT TAAGATCCT CATCCT	F S D D CATAATTAAA GCCCCGGGTT ATCAGCCTCG GGAAATTG GGAGGTGG TGAGTCATAGC TGAGTAACT AGGCGTTG AGGCGTTG AGGCGTTG AGGCGTTTG AGGCGTTG TTAATACGCT TCCCCTGGA AGCTCACGCT TACAGAGTTC GTAACTATCG TACAGAGTTC TTTAAATTAA ATCTGTCTAT TCTTAATTAT ATCTGTCTAT TCGCGAGACCC CCAGTCTATT TCGGGTTTG CCTCGGATCGT TCTCGGATCGT TTCTGGATCTT TTCTGACTTT CCTCGATCGT TTCTGACTTT CCTCGATCGT TTCTGTGATCTT TCGGATCGT TTCTGTGACT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTAGT TCTTGAGTAC ACTCTACGGG AAATGAAGT TTCTTCACC AAATGATTT TCGTTCACC GAATTGTTCAC GTAGGTACT TCGTTCACC GAATTGTTCAC GTAGGGTACT TTCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA CAAAACCAC TCTGACGCT TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGAT GCCCAGAT ACCAGATC ACCAGATC ACCAGATC ACCAGATC ATCAGCTCC AAGTTGCCC CACCAAGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGCTCGCTC AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGACT GGCTACACT GACAGACT GTTCGTACG CTAGATTAT TGACTCCCAC TAGATTAT TGACTCCCAC TAGTAGGACT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTTAGC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTATC AGTTGTTACC AGTGGTTACC ATTCTGAGAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA TTTCCAGTG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGGTAGAT AACCAGCCA TCGCCAGTTA GATCAAGGCG ACTCATGATT TCGCCAGTTA GATCAAGGCG ACTCATGATT TAGTGTATATC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACGTT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTCTGTG GCACCACA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTAAGGATA CCTGGAAGGC AATTTCGTAT CTTGGTCTGA AACTACGATA GCCGGAAGGC ATATTTGCTAG ACTTACGATA GCCGGAAGGC ATACTTTGC AGTTACATGA AGTTACATGA AGTTACATGA AGTTACATGA AGTTACATGA AGTTACATGA AGTTACATGA AGTGCAGCAC GGCGACCGAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAAA GTCCGCTTT CACGAACCA CTGGTAACAG TGGTCATCA CTGGTAACAG CTGGGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTACTAG CAGAGTTTT TGCTCTTGTTTGCTCTTTGC TGCAGATTTT TGCATAATTC TTGCTCTTTGC
>D V Q Q EACTOR   2001 GGATACTTGA GTTAC	Y E M I E STITTAT AAAACTTTI CGGTCG AGTCTAGAGG SAGTTT AAACCCGCTC AAAAAA TGGTTTATTI TGTTC AGGTCAGG AAGCTG GCCTAATGA GCCAAC GCGCGGGAC AAGGTC CTCAAAGGC AAGGTC GCGTTACTG CGCTGC GCTTTCTG CGTTGCT GCCTTACTG CGTTCC GCTTTCCC GGTATG TAGGCGGT CTTCGG AAAAAGAGT CCTTCA CAAAAGAGT CCTTCA CAAAAGAGT CCTTCA CAAAAGAGT CCTTCA CAAAAGAGT CTTCA CAAAAGAGT CCTTCA CAAAAGAGT CCTTCA GCATGATAC CGCCCC CATGCT GCATGATAC CGCCTCCAT CATCCT GGTTCACGG CTTAGC TAAGATGCT CATCCT GGTTCACGGTC CATCCT GGTTCACGG CTTAGC TAAGATGCT CATCCT GGTTCACGG CTTAGC TAAGATGCT CATCCT GGTTCACGG CTTAGC TAAGATGCT CATCCT TAAGATGCT CATCCT TAAGATGCT CATCCG TAAGATGCT CATCCGTTAAAAAAAAAAAAAAAAAAAAAA	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTCATAGC TGGTCATAGC TGAGCATTG GGAACGTTG GGAACTTCCA TCCCCTGGA AGCTCACGCT GTAACTACG TTAACTATCC TTAACTATC CGTAGCTCT CGTAGCTCT CTTTAATTAA ATCTGTCTT CCCGAGCACC CCCAGTCTT CCCGATCGTT CCCCATCGT TTCGTCGTTTC CCCCATCGT TTCGTCGTTTG CCCCATCGT TTCTGTCTTT TTCGTCGTTTG CCCCATCGT TTCTGTACCT CATACCAGAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATAATT CGTATTAGGC TATCCACAGA ATTGCCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTGGT TCTTCACGGA AATGAGTT TCTTCACGGA AATGAGTT TCCTTCACACAC AATTGTACC ACGCCAACGC AATTCACACGC AATTGTTCCT GTCAGAAGT TGTCAGAGAGT TGTAGGCTC TGTAGAGTACT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCCGCTAAC ACCAGTCGGTA ACCAGTAC ACCAGTACAT ACAAACCAA TTAAATCAAT TAAATCAAT CATAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CAGTTCGCC CAGTTGCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCCC CAACCAAGTC CACCAAGTC CCTCATCATT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AACCTCTAC TATCCGCTCA CACTGCCGC TTCCTCGCTC AACCAGGAA AGCATACACA TGTTCCGACC TGCTTCCGACC GACGCACACC GCTGCTACACC TGTGCTCCGCC TAGGTCGTC CTGGTACCG TGTGGTACCG TGTGGTACCG TGTGGACCG TACACCAAT TGACTCCCC TACACCAAT AGTGTTATC ATCTGGTTTCGAAC CAGTGTTATC ATTCTGAGAA GGAAACCTT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTGG ACTGACTGG ACTGACTGG AAATCGCGCTTA GAAAGCACTGG AGAAGAAGTTTTT AAACTCACGT TATGAGTAAGT TATGAGTAAACT ACCACCATTA AAACCAGCA TCGCCAGTTA AAACCAGCA TCGCCAGTTA CAGCAGCA TCGCCAGTTA CAGCAGTAAGCT ACCACGT TATGAGTAAAA TCGCCAGTTA CACCAGTTA TAGTAGTATGC CTTCGGGGGGGCG CTTCCGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGGCG CTTCGGGGGGCG CTTCGGGGGCG CTTCGGGGGCG CTTCGGGGCG CTTCGGGCG CTTCGGGGCG CTTCGGGGCG CTTCGGGCG CTTCGGGCC CTTCGGGCC CTTCGGGCC CTTCGGGCC CTTCGGGCC CTTCGGGCC CTTCGGCC CTTCGCGCC CTTCGCGCC CTTCGCGCC CTTCGCGCC CTTCCGCC CTTCCGCC CTTCCGCC CTTCCGCC CTTCCGCC CTTCCGCC CTCCC CTCCC CTTCCCC CTTCCCC CTTCCCC CTTCCCC CTCCC CTCC CTCC CTCCC CTCC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT TGTTTGCAAG TAGTTGGTAT TGTTTGCAAG TAGTGTCGTA ACTACGATA GCGGAAGGG ATACTTTGCTAT ACCGAAGGG ATACTTTCAAG ATGCAACGAA AACTTACGATA ACGGAAGGG ATACTTTGCAAG ATGCAACAGA AAAACTTCTCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTCGGCGTAACAG GTCGGCTTT CAGGAACCC CTGGTAACAG CTGGCCTTT CAGGAACCC CTGGTAACAG CTGGCTTTACCAA GCGGAGGCTCTG CGGAGCGCTCT CCGAGCCATGT TCCCCCCATGT TCCCCCATGT TCCCCCCATGT TCCATCATTC AGGATTATTC TTCCTCTTCC AGGATTATTC TTGCTCTTTCC AGGATTATTC TTGCTCTTTCC AGGATTTTC
>D V Q Q E EACTOR   2001 GGATACTTGA GTTAC	Y E M I E STITTAT AAAACTTTI CCCTCG AGTCTAGAGG SAGTIT AAACCCGCTC AAAAAAA TGCTTTATI TGTTTC AGGTTCAGGG AAGCTT GCCTAAAGGC AAGCTA CTCAAAGGC AAAGAT ACCAGGGGA AAGGC GCTTTCTCA CCCTGC GCTTTCTCA CCCTGC GCCTTATCCC CCTTCG AAAAAAAAAT TCTTCA CCTAGATCC SGAATC TCAAAGGCT TAGGCGGTG CCTTCGG AAAAAGAGT CCTTCGG TATCTCAGC GGAATCT CAAGAGATC CATAGATC GGCATC CATAGATC GGCATC CCTAGATC CCTAGATC CCTTCGG CCATAGATC CCTTCGGT CATCCC CATAGATC CCTTCGGT CATCCC CATAGATC CATCC CATAGATC CATCC CATC	F S D D C CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GGGAGTTGGGGGGGGGG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT GGAGTTTTT TGTTTCCTGT TATCCACAGA TAGCTCCCC AGCTCCCCG GTAGGTACT TCTTGAGTCC TCTTGAGTCC TCTTGAGTC TTCTTGAGTC TTCTTCGCG AATTGAGTC TTCATCACGA AATTGATCC CGTATGGCTTC TCTCACGC TTCGTCACCG TTCGTCACCG TTCATC TCTCACGC TTCTTCATC TCTCACGC TTCTCACCG TTCTATCC TCTCAGAAGT TCTCACCG TGTCAGAAGT TCTTAAAAGT TTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CAACTGATCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GTGAAATTGT GCGTTGGCGT GCTCTTCCGC ATCAGGGGAT GCGCTACCT CCCCTGACG TGCGCTACCT ACCCGGTAA GGCCTAACTA GCCCTAACTA GCCTAACTA GCCTAACTA ACCACGTA TTAAATCAAT CATAGTTGCC CTCCAGAGTT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCG CAACCAAGTC CACCAAGTT TCAGCATCT TTAACATCT CAGCATCT TTAACATCT CAGCATCT CACCAACTT TCAGCATCT TCAGCATCT TTAACCAACTCT TTAACCAACTCT TTAACACTCT TTAACACTCT TTAACCAACTCT TTAACACTCT TTAACACTCT TTAACACTCT TTAACACTCT TTAACACTCT TTAACACTCT TTAACACTCT TTAACACTCT TCAGCATCTT TCACCATCTT TCACCATCTT TCACCATCTT TTAACCACTCT TTAACACACTTT TCACCATCTT TTAACCACTCT TTAACCACT TTAACCACT TTAACCACTCT TTAACCACTCT TTAACCACTCT TTAACCACTCT TTAACCAC	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGCTACACA GGTTCCGACC TAGGTCATAT GGCTACACT GGCTACACT GGCTACACT GCTGGTACCG CTAGGTACCG CTAGGTACCG CTAGATATA TGACTCCCCG TATCCACAC TAGGTTATC TATCACACAT TGTTCCCAAC TATCACACAT TGTTCCCAAC TATCACACAT TATCACACAT TATCACACAT TATCACACAC CAGTGTTATC ATTCTGACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTCCACA TTTCCAGTCACT ACTTCCACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACTACACAC ACTACACACAC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CACCATACGA AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTCGGT AGCAAAAGGC GCACACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TGGTCTGA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ATTACATGA ATGCCAGCA ATTGCTTGCA GGGTGACCAG ATACTTCCA ACTACCATCA ACTACCATCA ACTACCACCA ACTACCACCA ACTACCACCA ACTACCACCA GCACCCAG ACACCCAG ACACCCAC ACCCAC ACCCA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAAACCA TGATCAGTCG CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCAGATTA TGGTCATGAC CAGCAGATTA TCGTCATGAT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCTCATGT TTCCTCTTGC AGGATCTTTC TGGTCTTTC TGGTCTTTC TGGTCATAATTC TTGCTCTTTGC AGGATCTTTC AGGATCTTTAC AAGCAGGAAC AAACAGGAAC
>D V Q Q E 2001 GGATACTTGA GTTAC	Y E M I E STTTAT AAAACTTTY CGCTCG AGTCTAGAGG SAGTTT AAACCGCTCA AAAAAA TGCTTTATT TGTTTC AGGTTCAGG AAGCTT GGCGTAATGA GCCAAC GCGCGGGA ACCTCA CTCAAAGGC AAAGGC GCGTTGCTGC AAAGGC GCGTTGCTGC CGTTGTGC GCTTATGCC GGTATG TAGGCGTGC GCTATGCC GGTATG TAGGCGTGC GCTTTCCA TGCGTC GCATAGATCT TCTTCA CCTAGATCCT GGCACC TATCTCAGC AAAACGC TTTTAT CCGCCTCAG CATCGT GGTGCTACG CATCGT GAGTGCT CATCGT GAGTGCT CATCGT GAGTGCT CATCGT GAGTGCT CATCGT TAGGTCCT CATCGT TAGGTCT CATCGT CATCGT TAGGTCT CATCGT	F S D D CATAATTAAA  G GCCCGCGTT ATCAGCCTCG GGAAATTG GGAGGTGG TGGTCATAGC TGAGTAACT AGGCGTTG AGGCGTTTG AGGCGTTTG AGGCGTTTG AGGCGTTTG AGGCGTTTG TACAGAGTTC TACAGAGTTC TACAGAGTTC TACAGAGTTC TTTAAATTAA ATCTGTCTT TTTAAATTAA ATCTGTCTT TTTAAATTAT TCGCGATCAT CCCAGTCTATT CCGCGAGCAC CCCAGTCTATT CTGTGGTTTG CTTCGTTTG CTTCGTTGACT CTTTGTGACT CATAGCAGAA CTCGTGCACC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCAGC AGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCG AAATGAGTATT TCCTTCAGC AAATGAGTATT TCCTTCACC AAATGAGTACT TTCGTTCATC GTATGGCTTC TTTGAGTACC TTTCAGAGTACT TTCTTCAGT TCTTCAGC TCTTCAGC TCTTCAGC TCTTTAGAGTTC TTTTAGAGTTC TTTTAAAAGT CACTGAGTACT CTTTAAAAGT CAACTGATCT TGTTGAAATAC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTACCG ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCAGCT TCTGACGCT TTAAATCAAT GGCAACCAC GCTCAGCG TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGAT GGCAAGTC CACCAGATT CGCAACTAC TCCACACT CCCAACTC CACCAACT CCCAACTT CCCAACTT TCACACTCT TCAACACTT TCACATCTT TCACATCTT TCACATCTT TCACATCTT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATTCCACA ATTTCCACAC ATTTCCACAC ATTTCCACAC ACTGACTGC ACACACCC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACTACGATA ACTACGATA CCGGAAGGG ATACTTGCA AGTTACAGTA ACTTACATA AGCGGAAGG ATACTTCC AGTTACATGA AGGCGACCAG AGAACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTATCA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAAACCA CTGCTACTAC CAGCAGATTA TGGTCATGACAC CAGTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACTT TCCCCCATGT TGCATATTC TTGATATTC TTGATATTC AAACAGGAAG CAGCTTATCA AACAGGAAG CAGCTTATTTC AGCATTTTAC AAACAGGAAG CAGCTTATTTC CAGCAGATTTAC CAGCAGATTAC CAGCAGATTAC CAGCAGATTAC CAGCAGAGATTTAC CAGCAGAGAGAGTT CCCCCATGT TGCATATTCT CAGCAGCAGAGAGAGAGGAGAG
>D V Q Q E EACTOR   2001 GGATACTTGA GTTAC	Y E M I E STITTAT AAAACTTTI CGCTCG AGTCTAGAGG SAGTTT AAACCCGCTCAAAAAAA TGCTTTATTI TGTTC AGGTTCAGGG AAGCTT GCCTAATGAC GCCAAC GCGCGGGAC AAGCTC CTCAAAGGCC AAAGAT ACCAGGCGT CGCTGC GCTTTCTCA CGCTGC GCTTTCTCA CGCTTCC GCTTTCTCA CGCTTCC GCTTTCCC GGTATC TAGGCGGTG CGCTC GCAATGATC CGCTCC GCAATGATC CGCTCCC CATCTC GCAATGATC CATCTC GCAATGATC CATCTC GCAATGATC CATCTC GCAATGATC CATCTC GAATGATC CATCTC CATCTC CATCTC CATCTC CATCTC CATCTC CATCTC CATCTC CATCT CATCTC CATCT CATCTC CATCTC CATCT	F S D D CATAATTAAA  GCCCCGGGTT ATCAGCCTCG GTGAAATTG GGAGGTCATAGC TGGTCATAGC TGAGCTTG GTAAACTTACGGT GTAAACTACGGT GTAACTACGGT GTAACTATCG TCCCCTGGA AGCTCACGCT CGGAGCTCTT CGGAGCTCTT CGGAGCTCTT CCCGAGACCC CCAGTCATT CGCGATCGTT CGCGATCGTT CGTCGTTTG CTCCGATCGT TTCGTGCTTT CATACGAGTAC CATACGAGTAC CCAGTCATAT CGCGAGACCC CAGCTCATT CGCGATCGTT TCGTGCTTT CGTCGATCGT TCGTGCACC CATACCAGAA CTCCTGTGACC GACACGGAAA TAGAAAAAAAAAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGCTATT GAGGTTTTT TGTTTCCTGT TATCCAGAG TATCCACAGA TAGGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTTCCTTT TCTTGAGTCC TCAAGTAGT TCTTCAGTCA TCTACGGG AAATGAGTT ACGCTCACCG AATTGTTGCC GTAGGGTTACT TCTTCAGAGTACT TCTTCAGAGTACT TCTTCAGAGTACT TCTTCAGAAGT TTCTTCAGAAGT TTGTTGAGTAC TTTTAAAAGT CAACTGATCT TGTTGAATAC AACAAATAGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCATTATTTG AAAGCAAGTA GCGTTGACTTTCGG ATCAGGGAT CCCCCTGACG GCTCTTCCGC ATCAGGGGAT ACAGCACTA ACAGCACTA ACAGCACTA GGCTAACT ACAAGCACT TTAAATCAAT ACAAGCACC CTCGAGGTTGC GCTCAGGT ACAGCACT ACAGCACT ATTAGCC CCCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCC CACCAAGTC CCCACAAGTC CCCCACAACT CGCCACAACT CGCTCACACT TCAAGCTCT TCAACCATT TCAGCATCTT TCATACCTT TCATACCTCT TCATACCTCT GGTTCCCGCG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATTCCACA ATTTCCACAC ATTTCCACAC ATTTCCACAC ACTGACTGC ACACACCC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC ACACCC ACACC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACTACGATA ACTACGATA CCGGAAGGG ATACTTGCA AGTTACAGTA ACTTACATA AGCGGAAGG ATACTTCC AGTTACATGA AGGCGACCAG AGAACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTATCA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAAACCA CTGCTACTAC CAGCAGATTA TGGTCATGACAC CAGTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACTT TCCCCCATGT TGCATATTC TTGATATTC TTGATATTC AAACAGGAAG CAGCTTATCA AACAGGAAG CAGCTTATTTC AGCATTTTAC AAACAGGAAG CAGCTTATTTC CAGCAGATTTAC CAGCAGATTAC CAGCAGATTAC CAGCAGATTAC CAGCAGAGATTTAC CAGCAGAGAGAGTT CCCCCATGT TGCATATTCT CAGCAGCAGAGAGAGAGGAGAG

### pMT-HisFlag-DmSNAP43 mut#3 STOP

1 TCGCGCGTTT CGGTGATGA	C GGTGAAAACC	TCTCACACAT	GCAGCTCCCG	GAGACGGTCA	CACCTTCTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGGGG								
201 CCGCACAGAT GCGTAAGGA								
301 TACGCCAGCT GCGTAAGG								
401 TGAATTAATT CGTTGCAGG								
501 GGCCCCCAC CGCCCACC								
601 CAAGTCCCCA AAGTGGAGA								
701 AGAGGTGAAT CGAACGAA								
801 AAATCAAGTG AATCATCTC	A GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT				CTACAAGGAT
					> M H	н н н	H H T D	Y K D
901 GACGATGACA AGGGCACTA	G TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
> D D D K G T	S E L N	I F D	D C W E	L V Q	R F Q	R L V N	D G E	N C E
1001 TCGAGGTGTT CTGCCGGTC	C TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
>FEVFCRO	WRE	LOLO	H L F	T A O	TNHT	E V I	ATT	LAAL
1101 GCATGTGGCC AAGCGACTC	T CGTGCTCCCG				TCGCGCTCAA	AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
> H V A K R L	S C S R		G D V	F P A S		RIG	G F F L	L Y V
1201 ATCTACTACA AGCAGCCCA								
> I Y Y K O P			E V S P	R T W		T D Y A		R K D
1301 GTCCGGAGCG GAAGGACAC					~			
				T O E	O A F R			
>S P E R K D 1		A Y M L					L D Y	
1401 GGACAATCTG GTGGACTAG								
> D N L V D Y	D R V E		G A K			M Q K	Q Q R A	
1501 AGTCTCACAT ACGAACTG								
> S L T Y E L		A L D						K Q L
1601 CGGCTGGTCA TGAGCACGC								
>A A G H E H A	L P P	S Q I A	G H A	R E V	A A A A	Q S V	L G A	R K S T
1701 TCCAGATGAG AAATGCAC	A CAACATCTAC	AGGCAACCAG	TTGGAAGTGC	GCCAGAGGGT	GCGGAACAAG	GCCATGTACG	GCGTCGAGGA	GCGGGAGCCG
> P D E K C T	TTST	G N Q	L E V	R Q R V	R N K	A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGAAG	T AGAAGTGCAG	CTGGAGGTCA	ACGAGACTTA	TCAACGCCGC	ATGTCCTCGG	CCACCGTTTT	CCAGAGGGAA	CTTCCAGAAG
> O H O T D E	L E V O	L E V	NETY	ORR	M S S	ATVF	O R E	L P E
1901 ACGTGCAGCA AGAGTATGA	G ATGATTGAGT	TTAGTGACGA	CGAGGAAATG	GAAGTGGGTG	AAAGCGAGGA	GGTCACGGAA	GAAGAACTCA	AAGCTATTT
>D V O O E V I	M T E	FSDD	E E M		ESEE			
>D V Q Q E Y I		F S D D		E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTA				E V G		V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTA > D T	T AAAACTTTTA	CATAATTAAA	TAACTAGCAT	E V G TTTTGCGCGA	TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
2001 GGATACTTGA GTTAGTTTA > D T 2101 AGCACAGTGG CGGCCGCTG	T AAAACTTTTA G AGTCTAGAGG	CATAATTAAA GCCCGCGGTT	TAACTAGCAT CGAAGGTAAG	E V G TTTTGCGCGA CCTATCCCTA	TGTGATCTTG ACCCTCTCCT	V T E TTTATCTGAA CGGTCTCGAT	E E L GGGCAATTCT TCTACGCGTA	K A I L GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTTTA > D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT	T AAAACTTTTA G AGTCTAGAGG T AAACCCGCTG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA	TGTGATCTTG  ACCCTCTCCT GACATGATAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA GTTAGTTT/ > D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT: 2301 CTAGAATGCA GTGAAAAA	T AAAACTTTTA G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA GTTAGTTT/  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA; 2401 TTGCATTCAT TTTATGTT.	T AAAACTTTTA G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAAGC.	T AAAACTTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCA	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT: 2301 CTAGAATGCA GTGAAAAAI 2401 TTGCATTCAT TTTATGTT: 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC	T AAAACTTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAAGC.	T AAAACTTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT: 2301 CTAGAATGCA GTGAAAAAI 2401 TTGCATTCAT TTTATGTT: 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC	T AAAACTTTA G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT GGGTTCAGGG T GGCGTAATCA T GCCTAATGAG C GCGCGGGAG	GCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT 2301 CTAGAATGCA GTGAAAAAI 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGCCCAI	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT C AGGTTCAGGG T GCCGTAATCA T GCCTAATGAG C GCGCGGGAG A CTCAAAGGCG C TCAAAGGCG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAJ 2801 CGGCGAGCGG TATCACCTC.	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTATTT C AGGTTCAGG T GCGTAATGA T GCCTAATGAG C GCGCGGGGAG A CTCAAAGGC C GCGTTGCTGG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTATTTCCA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA; 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA ACCCTGGGC 2701 GCATTAATGA ATCGGCCA; 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAAGC 2901 CCAGGAACCG TAAAAAAGC	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG C GCGCGGGGAG A CTCAAAGGCG C GCGTTGCTGG T ACCAGGCGTT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCGTTTTCCA TCCCCTGGA	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGATTTCCTGT  CACATTAATT  CGTATTAGGC  TATCCACAGA  TAGGCTCCGC  AGCTCCCTCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCCTGACG TGCGCTCCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA TTTCCAGTCG ACTGACTCGC AGAACATCTG AGAACATCTG CTGCCGCTTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCCAAA GTCCGCCTTT
2001 GGATACTTGA GTTAGTTTY  D T  2011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAAGT 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAGG	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCAG G GCGCGGGGAG A CTCAAAGGCG C GCGCGGTGCTGCTGG T ACCAGGCGTT C GCTTTCTCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA AGCTCTGGA AGCTCACGCT	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGTTTCCTGT  CGCATTAATT  CGTATTGGC  TATCCACAGA  TAGGCTCCGC  GTAGGTATCT  GTAGGTATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCCGC CAGTTCGGTC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCGCC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTCT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACCA TTTCCAGTCG ACTGACTGGC AGAACATGTG AAATCGACGC CTGCCGCTTGCCGCTTGCCGCTTGCCGCTTGCCACCT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATACGA GGAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTG CAGCAAAAAGG GGTGGGCAAA GGTCGGCCTT CACGAACCCT CACGAACCCT CACGAACCCCT CACGAACCCCT CACGAACCCC CACGAACCC CACGAACCCC CACGAACCCC CACGAACCC CACGAACCC CACGAACCCC CACGAACCC CACGAACCAC CACCACACAC CACCACACCAC CACCACACAC CACCAC
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAAI 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGG ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAI 2801 CGGGGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAGG 3101 CTCCCTTCGG GAAGCGTGC 3201 CCGTTCAGCC CGACCGCTC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GCCGAAATCA T GCCTAAATGAG C GCGCGGGGGG C GCGTTGCTGG T ACCAGGGGTT C GCTTTCTCAT C GCCTTATCCG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGCCGGTTTG GTAATACCGT CCCCTCGA AGCTCACCCCTGGA AGCTCACCGCT GTAACTATCG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CACATTAATT CGTATTGGG TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CATGCCGCC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACT TAGGTCGTTC GACACGACTT GACACGACTT GACACGACTT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATCGACTA CAATTCCACA ACTACTCGC ACTACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTGTG GCACCAGCCA	K A I L GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GCTGCGAAAGGA GTCGGCCTTT CACGAACCCC CTGGTAACAG
2001 GGATACTTGA GTTAGTTTY  D T  2010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGCATTGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCA 2801 CGGCGAGCGG TATCAAGT 2901 CCAGGAACGG TATCAAGT 3001 CCCGACAGGA CTATAAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CCGTTCAGCC CGACCGCTC 3301 GATTAGCAGGA GCGAGGTGC 3301 GATTAGCAGGA GCGAGGTGC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T AGGTTCAGGG T GGCGTAATCA G CCCTAATCAG G CGCGGGGAG A CTCAAAGGCG T ACCAGGGGTT C GCCTTATCCG G TAGGCGGTTG G TAGGCGGTG G TAGGCGGTG T GCCTTATCCG G TAGGCGTGG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT GGAGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCC TACAGAGTTC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT GTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA AGCCGGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT AAACCTTAC TATCCGCTCA CACTGCCGCG ACCGAGAA AGCATCACAA TGTTCCAGCT TAGTCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCACAC ACTGACTCGC AGAACATCTG AGAACATCTG ACTGCCGCTTA GCTCCAAGCT ATCGCCACTC ACGAGGACACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGACACC ACGAGGACACC ACGACC ACGAGGACACC ACGAGGACACC ACGACC ACGAGGACACC ACGACC ACCAC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTGG GCGCTGTTG GCACCAC TATTTGGTAT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTGGCGAACA GTCCGCCTTT CAGCAACCC CTTGGTAACAG CTGGCTCTCG CTGGTAACAG
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT. 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGCCCCA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAGG 3001 CCCGTCAGCC CGACCGCTC 3201 CAGTTAGCCC CGACCGCTC 3301 GATTAGCAGA GCGAGGTAC. 3401 CTGAAGCCAG TTACCTTCC 3401 CTGAAGCCAG TTACCTTCC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T CAGGTTAGGG T GGCGGAATCA G CGCTAATGAG C CGCGGGGAG A CTCAAAGGCG C CGGTTGCTGG T ACCAGGCGTT C GCTTTCTCAT C GCCTTATCCG G TAGGCGGTGC AAAAGAGTT AAAACAGTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTATAGC TGAGCTATAGC GTTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCTT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT GGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCGC GGTAGGTACT TCTTGAGTCC TTGAAGTGGC GATCCGCAAG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCGTTCGGCT ATCAGGGGAT CCCCTGACG ATCAGGGGAT ACCAGGTAACTA ACCACACACAACAA ACAAACCACC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCTCGGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACATC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTA CAATTCCAGTCG ACTGACATGG ACTGACTGG ACTGACTGG ACTGCCGCTTA GCTCCAGCT ATCGCCACTG AGAGGACAG GTGGTTTTTT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GCAGCTGCGG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAAA TGATCAGTCG CCGGAAGCA CGTTCCGCTG CAGCAAAAGG GGTGGCGAAA GGTCGCCTTT CACGAACCC CTGGTAACAG CTGCCGCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA
2001 GGATACTTGA GTTAGTTTY  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAGGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCCGGCCA 2801 CGGGGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTCAGGA GCGAGGTAC 3301 GATTAGCAGA GCGAGGTAC 3401 CTGAAGCCAG TTACCTTCC 3501 CGCGCAGAAA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T CAGGTCAGGG T GGCGTAATCAG G CGCGGGGAG C GCGTGCTGG T ACCAGGGGT T ACCAGGGGT C GCTTTCTCAT C GCTTTCTCAT C GCTTATCCG TAGGCGTTG C TAGGCGTTG C AAAAAGAGTT T CAAGAAGAT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGCGGTTTT CCA TCCCCCTGGA AGCTCACGCT GTAACTATCG GTACTATCG TACAGAGTTC GTACTATCG TACAGAGTTC GTTACTCT TTTACATCT CTTTGATCTT	CAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGATTCCTGT  CACATTAATT  CGTATTGGGC  AGCTCCCTCG  GTAGGTATCT  TCTTGAGTCC  TTGAAGTGGT  GATCCGCAA  GATCCGCAA  TTCTACGGGG  TTCTACGGGG  TTCTACGGGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTGCCC CAGTTCGGT AACCCGCTAAC GCCTAACA GCCTAACA GCCTAACA CACCGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CATGCCGC TTCCTGCTC AACGCAGGAA AGCATCACAA TGTTCCGACT GACACGGTTCCGTCT GACACGACT GGTGGTACACT GGTGGTACACT GGTTGGTACGC AGTGGAACGA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA TTTCCACGTC ACTGACTCGC AGAACATGTG ACTGACTCGC AGAACATGTG GCTCCAAGCT ATCGCCACTTA GCTCCAAGCT ATCGCCACTTA AGAAGGACAG GTGGTTTTT AAACTCACGT	E L GGGCAATTCT TCTACGCGTA TGACTTTIGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGAATACCT GGGCTCGGT GGGCTCGTGTG GGACTCGTGTG GCACCAGCCA TATTTGCTAT TGTTTGCAAG TAAAGGGATTT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCC CTGGTAACAC CTGGCTAACAC CTGGCTACT CAGCAACTTA
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGCATTGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA ATCGGCCAA 2801 CGGCAGCGG TATCACGTC 2901 CCAGGAACGG TATCACGTC 2901 CCCGACAGGA CTATAAAAG 3101 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTC 3201 CCGTTCAGC CGACCGCTC 3301 GATTAGCAGA GCAGGTAC 3401 CTGAAGCCAG TTACCTTCC 3501 CGCGCAGAAA AAAAGGATC 3501 CGCGCAGAAA AAAAAGGATC 3501 ATTATCAAAA AGGATCTTC	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT T AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG G GCGGGGGAG A CTCAAAGGGG T ACCAGGGGTT C GCCTTATCAT G GCTTTCTCAT G GCTTTCTCAT T GCCTTATCAT T GAAGAGAGT T CAAGAAGAGT T CAAGAAGATC A CTTAGATCC A CTTAGATCC T CATAGATCC T CAAGATCAT C CCTAGATCCT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTATAGC TGGCCTATAGC TGACCTAACT TCCCCCTGGA AGCTCACCGT GTAACTATCC GTAACTATCC GTAACTATCC GTAACTATCC GTAACTATCC GTAACTCT TTCAACAGTT TTTTAAATTAA	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTCCTGT CGAGTTTTTT GGTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT CTTTGAGTCC TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TCTCACGGGA TATCCACAGA	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTA ACCACGGTAA ACCACGCTAA CTAAACCAC TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TTAAATCAAT	ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAC TATCCGCTCA CACTGCCGC TCCCTCGCTCA ACCCAGCAA AGCATCACAA AGCATCACAA GCATCACAC TGTTCCGCTC GACACGACAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATCTG AAATCGACGC TGCCGCTTA GCTCCAAGCT ATCGCCACTG ACGAGGACAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATCACGT AAACTCACGT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAGGTCAGT AGCAAAAGGC CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTCT TGTTTGCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTC CGCTCGGCTT CAGCAAAAGG GTGCCAAAAGG CTGCTCTC CTGGTAACAG CTGGTAACAG CTGGGCTTCT CAGCAGATTA TGGTCATATACAG CAGTTATACAG CAGTTATACAG CAGTTACAGA CAGTTACACA
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT. 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGCCCAC. 2801 CGGCGAGCGG TATCAGCT. 2901 CCAGCAACGG CTATAAAG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTTCAGCC CGACCGCTC 3201 GATTAGCAGA GCGAGGTA: 3401 CTGAAGCCAG TTACCTTCC 3501 GCGCCAGAAA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGCCAC 3701 TGCTTAATCA GTGAGCCAC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T CAGGTTAGGG T GGCGTAATCAG G CGCTAATGAG C CGCTGGGGAG A CTCAAAGGCG C GCGTTGCTGG G CTTTCTCAT C GCCTTATCCG G TAGGCGGTGC G TAGGCGGTGC G AAAAAGAGTT T CAAGAAGATC C CCTGAATCCT C TATCTCAGCG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTATAGC GTTATTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GTAACTATCG TTACAGAGTTC TTTGATCTT TTTTAAATTAA ATCTGCTATA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTACTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC AGCTCCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGCT TTGAAGTGCT ATTCACGGGA TCTTACGGGG AATCAGAGTT TTCCTTCATC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGT AACCCGGTAA ACCACACTA ACAAACCAC TCTGACGCT CTGACGCT TTAAATCAAT CATAGTTGCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGAC TGTCTCGAC TGTCCGAC TGGCTCTC GACACGACT GCTGGTACCG GCTGGTACCG AGTGGAACGA CTAAAGTATA TGACTCCCCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT TTTCCAGTCG ACTGACTGGA AATTGACAC TTTCCAGTCG ACTGACCGTTA GCTCCAGCG TACCCCACTG GAGACATTTT ATGAGCC TGCGGGTTA GCTCCAGCGTTA TCCCAGTG TATGACACT TATGAGACA TTGGACACT TATGAGACA TTGGACACT TATGAGACA TTGGATAAA TCGTGTAGAT TATGAGTAAA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCACCAGCCA TATTTGGTA TATTTGGTA TATTTGCTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTTACAG CTGCGCTTC CAGCAGATTA TGGTCATGAG CAGTTACCAG CAGTTACCAG CAGGAGGCT
2001 GGATACTTGA GTTAGTTTY  D T  2010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGC ATGCAGGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAI 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAGG 3101 CTCCCTTCGG GAAGCGTGC 3301 GATTAGCAGA GCGAGGTAC 3301 CTGCAAGCAG ATACCTTCC 3501 CGGCCAGAAA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCACTGC 3801 TACCATCTGG CCCCACTGCC 3801 TACCATCTGG CCCCCACTGCC	T AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTATTT CAGGTTCAGGG GGGGAAACA CGCTAAAAGGCG CGCGTGAGGGAG ACCAGAGGGTT CGCTTATCTG CGCTTTCTCAT CGCTTATCCG AAAAAGAGTT CAAGAGATCT CAGAGAGATC CTAGATCAGGGT CTATCTCAGGG TACCAGGGT CAAAAAGAGTT CAAGAGATCAT CTAGCGGT GGAATGATACC	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG GGAGGTAACT AGGCGTATAC GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCC GTAACTATCC GTAACTATCC TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCTGCTTCTT CCCCGGAGCCC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TCTTCAGGGGA AATGAAGTT TCTTACGGGGA TTCTACGGGA TTCTACGGGA AATGAAGTT TCTTCACTCACCACACACCACCACCACCACCACCACCACC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTAC ACCAGGTAACACCA CCTAACACAC CTTAAACCAC TTAAATCAAT CATAGTTGCC GTTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGTTGCGTC CATAGTTGCC GCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTCTAC CATGCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC GACAGGACT GCTGCTTC GACACGACT GCTGCTACACT GCTGGTACACT GCTGGTACACG AGTGGAACGA TAAAGTATA TGACTCCCCC TATCAGCACT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA TTTCCACGTC ACTACTGCACTGC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAAGTCAGA AGCAAAAGGC CCGGATACCT GGCTCGGT AGCAAAAGGC CCGGATACCT GGCTCTGTG GGACTCGGT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGTACCAC CAGCAGATTA CAGCAGATTA
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGCATTGA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA ACTCGGCC 2701 GCATTAATGA ATCGGCCAC 2801 CGGCAGCGG TATCAAGGC 2901 CCCGACAGGA CTATAAAG 3001 CCCGACAGGA CTATAAAGG 3101 CCCGTTCAGC GAACCGTC 3201 CCGTTCAGC GAACCGTC 3201 CGTTCAGC GAACCGTC 3401 CTGAAGCCAG TTACCTTC 3501 CGCGCAGAAA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTC 3701 AGGTGGTCCT GCAACTTCT	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT T AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG G GCGGGGGAG A CTCAAAGGGG T ACCAGGGGTT C GCCTTACTAT G GCTTTCTCAT G GCTTTCTCAT T GAAGAGAGT T CAAGAGAGT T CAAGATGAT T CAAGATGAT T CACCTCCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCATAGC TGGCCTATAGC TGACCTAACT GTAATACGGT CGTTTTTCCA AGCCGGTTTG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTAAATTAA ATCTGTCTAT ACCCAGTCTATT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT CTTGAGTCC TCTTGAGTCC TCTTGAGTCG TATCCACCAGA TTCTACGGGA TTCTACGGGA TTCTACGGGA TTCTACGGGA TTCTACGGCA ATTGTTCACC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGGTA AACCAGCT CTGACACC TCTGACG TTTCGGTA TACAAACCAC TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATG GCTCCAGATG TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATT GGGAAGCTAG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAAT TATCCGCTCA CACTGCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GCTGGTACACT GCTGGTACGC AGTGGAACGA TAAAGTATAT TGACTCCCCG TATCAGCAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTTT ACTCCAACCT ACTGACACT ACTGACACT ACTGACACT ACTGACACT ATCGCCACTG AGAAGGACA GTGGTTTTTT AAACTCACGT TATTGAGTAAA TCGTGTAGAT AACCCAGCCA ACGCCACTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAAG TAAGGGATTC TTTGCAAG ACTACGAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GTGCGAAAAG GTCCGCTTT CACGAACCAC CTGGTAACAG CTGGTAACAG CTGGGCTCTC CAGCAGATTA TGGTCATGACAG CTGGGAGACCCC CTGGTAACAG CTGGGAGACCCC CCGAGATTA
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT. 2501 ACCTGCAGGC ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGCCCAC. 2801 CGGCGAGGGG TATCAGCTC 2901 CCAGCAGGG CTATAAAGG 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTCAGCC CGACCGCTC 3301 GATTAGCCAG GCCAGCTGC 3401 CTGAAGCCAG AAAAAGGAT 3401 CTGAAGCCAG TTACCTTCC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA AAAGGATTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGC CCCACTGCG 3901 AAGTGGTCCT CACACTTTY 4001 GCCATTGCTA CAGGCATCT 4001 GCCATTGCTA CAGGCATCCTT	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T CAGGTTAGGG T GGCGTAATCA T GCCTAATGAG C GCGTGGGGA A CTCAAAGGCG C GCGTTGCTGG T ACCAGGGTT C GCTTTCTCAT C GCCTTATCCAT C GCCTTATCCG T TAGAGAAGATC C CTAGATCCT C TATCTCAGCG T TATCTCAGCG T CCAATGATATC C CCCCCCAT T GGTGTCACGC T CCGCCCCCAT T GGTGTCACGC	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTATAGC GTTATACCGT GGATTTTCCA AGCCGGTTTTCCA AGCTACCT GTAACTATCG GTAACTATCG GTAGCTCTT CTTTGATCTT TTTAAATTAT ATCTGTCTAT CGCAGAGCCC CCAGTCCTTT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TATCACAGA TTCTACAGGG AATCAAGAT TTTACAGGG AATCATACAGT TTCGTTCAT ACGCTCACCG AATGATTATC ACGCTCACCG AATTGATTCCC GTATGGCTTCC GTATGGCTTCC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GCGAAATTG GCGTTCCGC ATCAGGGGAT CCCCCTGACG CCCCCTGACG ACCAGCACT ACCACACC CTCTACCACTC CAGTTCGCG TCTACTACTA ACAAACCACC TCTGACGTC TCTAACTAACTAACCACC CCCCCAGAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATCAGCTCC ACTAGTTGCC ACTAGTTCAGCTCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCTGCGCCG ACCAGGAA AGCATCACAA AGCATCACAC TGTTCCCAAC TGTTCCAAC TGTTCCAAC TGGCTACTC GACACGACT GCTGGTACCG ACTGGAACGA TAAAGGATAT TGACTCCCG TATCAGCAAT TGACTCCCCG TATCAGCAAT GGTTCCCCAAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACATTGAA AAATGGCAA TTTCCAGTCG ACTGACCTCGAC AGAACATGG AAACATGGCAA ATCGCCGCTTA GCTCCAAGCT ATCGCCACTG GTGGTTTTTT AAACTCAGCG TTGGTGGTTTTT AAACTCAGCG AGAGGACAA TCGCGCAGTA AACCAGCCA TTGCTAGATAA TCGTGTAGAT AAACCAGCCA TCGCCAGCTA AAACCAGCCA TCGCCAGTTA TCAGCAGCTA TCGCCAGTTA TC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA GCAGCAACAT GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTGGAACACTGGAAGGA TAACGGAATGGAACACACACAACAAACATAACAAAAACAAAAACAAAAAACAAAAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTCGGCTT CAGCAAAAGG GTCCGCCTT CACCAACCC CTGGTAACAG CTGGCTCTG CAGCAGATTA TGGTCATGAG CAGCAGATTA CGGAGGCCAG CAAGGTTGTT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTY  D T  2010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGT. 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTT. 2501 ACCTGCAGGG ATGCAAGC. 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCCGGCCA 2801 CGGGCAGCGG TATAAAAGG 3001 CCCGACAGGA CTATAAAGG 3001 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCCTG 3201 CGGTCAGCA GCACGCTC 3301 GATTAGCAGA GCAAGGTAC 3501 CGCGCAGAAA AAAAGGAT 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTI 4001 GCCATTGCTA CAGGCATCC 4101 TGTGCAAAAAA AGCGTTTAC 4101 TGTGCAAAAAA AGCGGTTACCTTCC 4101 TGTGCAAAAAA AGCGGTTACCTTCC 4101 TGTGCAAAAAA AGCGGTTACCTTCC 4101 TGTGCAAAAAA AGCGGTTACCTTCC 4101 TGTGCAAAAAA AGCGGTTACCTTCCCAACTTTI 4001 GCCATTGCTA CAGGCATCCC 4101 TGTGCAAAAAA AGCGGTTACCTTCCCAACTTTI 4001 GCCATTGCTA CAGGCATCCC 4101 TGTGCAAAAAA AGCGGTTACCAACTTTI 4011 TGTGCAAAAAA AGCGGTTACCAACAAAAAAAAAAAAAA	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATGAG C GCGCGGGAG A CTCAAAGGGG C GCGTTGCTGG G TAGCGGGTT C GCTTTCTCAT C GCTTATCCG TAGGCGTTC T CAAGAGAGTT T CAAGAGAGT T CAAGAGATCT T CAAGAGATCT T TATCTCAGCG T TATCTCAGCG T GCAATGATAC T CCGCCTCCAT C GCTTCAGCG C TCCTTCGGTC	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG GTGATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCT GTAACTATCG GTAACTATCG TAACTATCG TAACTATCG TAACTATCG TAAATTAAAT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATGTTTCCTGT CGCATTAGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TCTACGGGA ATCTACGGGA AATGAAGTGT TTCCTTCACCGAATTATTCACCGCAATTCACTCACGAATTCATCACGGGAATTCTACGGGAATTCTACGGGAATTCTACGGGAATTCTCATCACGCGAATTCTCATCACGCGAATTCTCATCACGCGAATTCTACGCTCACGGAATTCTTCTTCATCACGCGAATTCTTCTTCATCACAGAGTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTGTCAGAAGTTTTTGTCAGAAGTTTTTGTCACTTTGTCACAGAGTTTTGTCACTTTGTCACAATTTTTGCTCAGAAGTTTTGTTCATCATTTTGTCACAATTTTTGTCACAAGTTTTTGTCACAATTTTTGTCAGAAGTTTTTGTCACAATTTTTTTGTCACAATTTTTTGTCACAATTTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTACGT AACCAGCTACC CAGTTCGGTA ACCAGCTACT TGAAACCACC TCTGACGTAACCACC TCTAACGTAC TCTAACGATA GCCTAACT ACAACCACC TCTAACGTTA GGCAACT TTAAATCAAT CATAGTTGCC CCTCCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC AAGTTGGCCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTCTAC CATTGCCGCC TTCCTCGCTC ACCTGCCCC TTCCTCGACC TGGTCCTCACC TGGTCCTTCCACC CAGGAAAGCATCACAA TGTTCCGACC CAGGACAAGCAAAGCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA ATTTCCACA ACTTCCACA ACTGACTGG ACTGACTGG ACTGACTGG ACTGACTGG ACTGACTGG ACTCCAAGCT ATCGCCACTTA AGAAGACAG GTGGTTTTT AGAAGTAAA TCGTGAGTAAA TCGTGAGTAAA TCGTGAGTAAA TCGTGAGTAAA AACCAGCT AGAAGCAGCA AAACCAGCT AGAAGCAGCA AAACTAAGGCA AAACTAAGGCA ACTCATGGTTA	E L GGGCAATTCT TCTAGGCGTA TGAGTTTIGGA AAACAAGTTA TGGCTGATTA CAACATACGA AGGAAAACGC TCAAGTCAGA AGCAATACGA AGCAATACGA AGCAATACGA AGCAATACGT GGGCTCGGT GGGCTCGGT GGGCTCTGTG GGACACCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA ATGCCAGCAA ATGCCAGCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCCAAAAGG CTGGCAAAAGG CTGGCTAACAG CTGGTAACAG CTGGTAACAG CTGGTAACAG CTGGTACACA CAGCAGATTA TGGTCATGAG CAGGAGGCT CCGAGGCCCAG CAGGTGTT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATGGCCAA 2801 CGGCAAGCG TATAAAGG 3001 CCCGAACGGA CTATAAAG 3101 CCCGCACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTG 3201 CCGTTCAGC CGACCGCT 3301 GATTAGCCAG GCAGGTAT 3401 CTGAAGCCAG TTACCTCC 3501 CGCGCAGAAA AAAAGGATC 3501 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCACTGG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCT CAGCACTCT 4001 GCCATTGCTAAAAAA AGCGGTTCC 4101 TGTGCAAAAA AGCGGTTCC 4101 TGTGCAAAAAA AGCGGTTCC 4101 TGTGCAAAAAAAGCGGTTCC 4101 TGTGCAAAAAAACGCGTTCC 4101 TGTGCAAAAAAACGCGTTCC 4101 TGTGCAAAAAAACGCGTTCC 4101 TGTGCAAAAAAACGCGGTTCCCACTGCC 4101 TGTGCAAAAAAACGCGGTTCC 4101 TGTGCAAAAAAACGCGGTTCCCACTGCCACTCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCACTGCCCCCCACTGCCCCCCACTGCCCCCCACTGCCCCCACTGCCCCCCACTGCCCCCCCC	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG A TGCTTTATTT T AGGTTCAGG T GGCGTAATCA T GCCTAATGAG G CGCGGGAG A CTCAAAGGCG T ACCAGGCGT T CAGGCGTTC G TAGGCGTGC G TAGGCGTGC G TAGGCGTGC G TAGGCGTGC T CAAAAAGAGTT T CAAGAAGATT T CAAGAAGATT T CAAGAAGATT T CAGATCCT T CAGGTCCT T CCTTCAGCC T GCGTTCAGC T GCGTTCAGC T GCGTTCAGC T GCGTTCAGC T GCGTTCAGC T CCTTCGGTC T TAGGTTCAGC T TAGGTTCAGC T TCCTTCGGTC T TAGGTTCAGC T TAGGTTC T TAGGTTCAGC T T TAGGTTCAGC T T TAGGTTCAGC T T T T T T T T T T T T T T T T T T T	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAACT GTAATACGGT CGTTTTTCCA AGCCGGTTTG GTAACACGT TACAGAGGTTC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTTTT TCTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT TCGTCGTTTG TTCTGTCTTT TCGTCGATTGT TTCTTCGATCTT TTCTTCGATCTT TCTTCGATCGT TTCTTTTAACTTT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGAA TTCTACGGGA TTCTACGGGA ATTCTACGGGA ATTCTACTCACCG AATTGTTCAC CGAATTGTTCCC GTAGGCTTCCCC GTAGGGTACT TCTGAGAGT TTCGTCACAGG GGTAGGTTCCC GTATGGCTTC CGAAGAGT TGTCAGAAGT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA CAAAACCAC TCTGACGCT TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGAT GCCCAGAT ACCAGATC ACCAGATC ACCAGATC ACCAGATC ATCAGCTCC AAGTTGCCC CACCAAGTC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAC TATCCGCTCA CACTGCCGC TACCTCGCTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGACACAC TAGACACAC TAGACACAC TATACAGCAT TATCAGCAAC TATCAGCAAC CAGTTATCCAAC CAGTTATCCAAC CAGTTTATC CAGTTATCAAC CAGTTTATC CAGTTATCAACA TATTCTAGAAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTTT ACTCCAACCT ACTGACACT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT ACCGCCAGTTA GATCAAGGCG ACTCATGATTATACACTT TAGTGTAATT TAGTGTATATC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAGAGAACGCC AATTTGGTAT CTTGGTCTGAACTACGATA ACTTACGATA GCCGGAAGGC AATACTTTGCAAG AACTACGATA ACTTACGATA ACTTACGATA ACTTACGATA AACTACGATA ACTACATACATA AACTACATACATA AACTACATACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GTGCGAAAAG GTCCGCTTT CACGAACCAC CTGGTAACAG CTGGTAACAG CTGGGTTCT CAGCAGATTA TGGTCATGAC CTGGGAGCT CAGCAGATTA TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTACTAG CAGTTACTAG CAGAGTTT TCCCCATGTT TCCCCCATGT TGCATATATT TCCCCCATGT TGCATATATT TCCCTTTGCTCTTTGC
2001 GGATACTTGA GTTAGTTTY  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAC 2801 CGGCGAGGGG TATCAAGCT 2901 CCAGCAAGCA CTATAAAGG 3001 CCCGCACAGGA CTATAAAGG 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTTCAGCC CGACCGCTC 3301 GATTAGCAGCA GCGAGGTAC 3401 CTGAAGCCAG TTACCTTCC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT CCAACTTTY 4001 GCCATTGCTA CAGGCATCC 4101 TGTGCAAAAA AGCGGTTAC 4101 TGTGCAAAAAA AGCGGTTAC 4101 TGTGCAAAAAA AGCGGTTAC 4101 TGTGCCAACTTC ATGCCGTCC 4101 TGTGCCAACTTC ATGCCGTCC 4101 TGTGCCAACAAAAA AGCGGTTAC 4101 CCGGCCGTCAA TACCGGCATCC 4101 CCGGCCGTCAA TACCGGCATCC 4101 CCGGCCGTCAA TACCGGCATCC	T AAAACTTTA  G AGTCTAGAGG T AAACCCGCTG A TGCTTTATTT T CAGGTTCAGGG T GGCGTAATCA T GCCTAATGAG C GCGTGCGGAG A CTCAAAGGCG C GCGTTCTCAT C GCCTTATCCAT C GCCTTATCCAT C AAAAAGAGTT T CAAGAAGATC C CTAGATCCT C TATCTCAGCG TAGATCCT C TACTCAGCG TAGATCCT C TACTCAGCG T GCTCTCAT T GGTGTCACGC C TCCTTCGGTC C TAAGATCTT A TACCGCCCA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTATAGC TGAGCTATAGC TGAGCTATAGC GTATATACGT CGTTATACAGT CGTCATAGC GTAATACGGT TCCCCCTGGA AGCTCACCC GTAACTATCG TACAGAGTTC GGTAGCTCTT TTTAAATTAA ATCTGTCTAT TCGTCGATCTAT TCGTCGTTTG CTCGGATGCT CTCGATCGTTTG CTCGATGGTTTG CTCGATGGTTTG CTCGATGGTTTC CATAGCAGAA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTTGAAGTCC TTGAAGTGC TTTGAGTCC TTGAGTACT TTCGTTCATC ACGCTCACCG AATTGTTACC GTAGGAGTT TTCGTTCATC ACGCTCACCG TATGGCTT TGTCAGAAGT CTTTAAAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCCGCTAAC ACCAGTCGGTA ACCAGTAC ACCAGTACAT ACAAACCAA TTAAATCAAT TAAATCAAT CATAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CAGTTCGCC CAGTTGCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCCC CAACCAAGTC CACCAAGTC CCTCATCATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCGGCAGA AGCATCACAA AGCATCACA TGTTCCGACC TAGGTCGTC GACACGACT GGCTACACT GCTGGAAC AGTGAACA AGTATACAA TGACTCCCCG TATCAGCAAC TAGACTACACA TAGACTACACA TAGACTACACA TAGACTACACAC TAGACACACA TAGACACACA TACTAGACAAC AGTGTTACA CAGTGTTATAC AGTAACACAC GGTTCCCAAC CAGTGTTATAC AGTAACACT ATCTAGACAA GGAAAACCTT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACGA ATCGCCACTG AGAAGGACAGT ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGCACGCA TCGCCAGTTA AACCACCAC TCGCCAGTTA AAACCAGCA TCGCCAGTTA CACCAGTTA CACCA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCTA TGTTTGCAA TGTTTTGCAA TATTGGTTGA TATGGTCTA ACTACGATA CCGGAAGGC ATACTT CTTGGTCTGA ACTACGATA GCCGAAGGC ATATTTGCTA GCAGCAGCAG ATAGTTACATA ACGACAGAA ACTACCATA ATGCAACACAA AAACTCTCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCAC CTGGTAACAG CTGGCTTACAA GCGCAGTTACCAA GCGGAGGCT CCGAGCGCATT TGGTCATGAG CAGTTACCAA CAGTTACCAA CAGTTACCAA TCCGCAGTTT CCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCATCATCATC AGGATCATTC TTCCTCTTCC AGGATCTTTAC
2001 GGATACTTGA GTTAGTTTY  D T  2011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATCAC GATGAGATA 2401 TTGCATTCAT TTTATGTTT. 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGGCAGCGG TATCACGTC 2901 CCAGGAACCG TAAAAAGG 3001 CCCGACAGGA CTATAAAGG 3101 CTCCCTTCGG GAAGCTGC 3201 CGGTTCAGCC CGACCGCTC 3201 CAGTTCAGCC GCACCGTC 3701 GATTAGCAGA GCGAGGTA 3601 ATTATCAAAA AGGATCTC 3701 TGCTTAATCA GTGAGGCA 3801 ATCACATCTGC GCCAACTGCT 4001 GCCATTGCC GCCAACTTT 4001 GCCATTGCT AGGCATCC 4101 TGTGCAAAAA AGCGGTTA 4201 TCTTACTTCT ATGCCATCA 4301 CCGCGCGCAA TACCGGTAC 4301 CCGCGCGCAA TACCGGTAC 4401 CGCGCTCAA TACCGGTAC 4401 CGCTGTTGAG ATCCAACTCT 4401 CGCTGTTGAG ATCCAACTCT 4401 CGCTGTTGAG ATCCAACTCATCA CGCGGTCAA TACCGGGTAC CGCGCTCAA TACCGGGTTAC ATCCGATTGAG ATCCAACTTCT ATCCGATTGAG ATCCAACTCT ATCCGATTGAG ATCCAACTCT ATCCGATTGAG ATCCAACTCT ATCCGATTGAG ATCCAACTCT ATCCGATGAG ATCCAACTTCT ATCCGATTGAG ATCCAACTCT ATCCGATTGAGA ATCCAACTCT ATCCGATTGAGATCAACTCT ATCCGATTGAGATCAACTCT ATCCAACTCT ATCCAACTC ATCCAACTCT ATCCAACTCT ATCCAACTC ATCCAACT ATCCAACTC ATCCAACTC ATCCAACTC ATCCAACTC ATCCAACTC ATCCAACTC ATC	T AAAACTTTA  G AGTCTAGAGG T AAACCGCTG T AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG C GCGCGGGAG A CTCAAAGGCG T ACCAGGCGTT C GCTTATCCAG T ACCAGGCGTT C GCTTATCCAG T TACTCAGCG T AAAAAGAGT T CAAGAAGATC T CATCTAACG T TACTCAGCG T GCATTAGTAC T CAGGCCTCAT T GGGCTCCAT T GGGCTCCAT T GGGTTGCTAGCG T CAAGAAGATC T CTTCAGCG T CAAGAGTCCT T CAAGAAGATC T CAGGCCTCCAT T GGGTTCAGCG T TAAGATGCT T AAAGATGCT T TACCGCCCCA T TACCCACCA ATGTAACCCA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTC GGAGGTGTGG TGGTCATAGC TAGCCTAACT AGCCGTTTTTCCA TCCCCCTGGA AGCTCACCT TACAGAGTTC GTAATATCGT TACAGAGTTC GTAACTATC TACAGAGTTC TTTTAAATTAA ATCTGTCTAT TCGCAGAGCC CCAGTCTATT TCGTCTATT TCGTCTATT TCGTCTATT TCGTCGTTTC TCTCGTCATT TCTCGTCACC CAGACCC CAGACCAT TCTTGTGACT CATAGCAGAA CTCTGTCACC CATCTATC TCTTGTGACT CATAGCAGAA CTCCTGCACC	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTC GTAGGTACT CTTGAAGTCGT TCTTGAAGTCG TCTTGAAGTCG AATGAGTACT TTCATCACGGA AATGAGTT TCTTCACC AATGATCC CTACC AATGATCT TTCTCACC AATGATCT TGTCAGAAGT GGTACC CTACAAGATC CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CAACTGATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGAT GCGCTTACGT GCCCTGACG TGCGCTACC ACCAGGTAA GCCCAACACA GCCTAACTA GCAACCAC TCTAAACCAC TCTAAACCAC TCTAAACCAC TAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGGCC CACCAACT CACCAACT CACCAACT CACCAACT CACCAACT CACCAACT CACCAACT TCACCACT TTAAACCAACT CACCAACT CACCAACT CACCAACT TCACCACT TCACACT TCACCACT TCACCAC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCCGC TAGCTGACA TGTTCCTGGTC AACCGAGGAA AGCATCACAA TGTTCCGACC TAGGTGTTCC GACAGGACTT CGCTGATACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT TGACTCCCCAG CTATAAGTATA TGACTCCCAG CTATAAGTATA TGACTTCCAAC CAGTGTTATC ATTCTGAGAA AGTAAACGTT TTACTTTCAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTCCACA TTTCCAGTG ACTACTGCA ACACTGCA ACACTGCA ACACTGCA ACACTGCA ACACTGCA ACACTGCA ACACTGCA ACACTCCA ACACTCCA ACACTCCA ACACTCCA ACACTCCA ACACTCCA ACACTCCA ACACTCCA ACACTCA ACCCA ACCC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CACCATACGA AGCAAAACGC TCAAGTCAGC AGCAAAACGC CGGCTCGTT GCGCTCGGT AGCAAAACGC ATTTGGTATA TGTTTGCAAG AACTACGAA AACTACGAA AACTACGAAAACGC AATTTGGTAT TGTTTGAAGGATTT CTTGGTCTGA AACTACGATA AACTACGATA AACTACGATA ATGCCAGCACGAAAACGC ATGCCAGCACCAG ATGCCAGCACCAG ATGCCAGCACCAG AACACCCAG AACACCCAG AACACCCAG AAAACTCTCA GGGTGACCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCGAAA GTCGGCTTT CACGAAACCA TGATTAACAG CTGGCTATT CAGCAGATTA TGGTCATGAC CAGCAGATTA TGGTCATGAC CAGCAGTTGT CCGCAGCCCAG CAAGGTGTT TCCCCCATGT TCCCCCATGT TTCCCTCTTC TGCATAATTC TTGCTCTTGC AGGAATTAC AAACAGGAAC AAACAGGAAC AAGGAACTTACCAAC TTCCCCCATGT TCCCCCATGT TCCCTCTTGC AGGATCTTAC AGGATCTTTAC AGGATCTTTAC AGGATCTTTAC
2001 GGATACTTGA GTTAGTTTY  D T  2011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATGGCCAA 2801 CGGCGAGGG TATCAAGGC 2901 CCAGGAACGG TATAAAAG 3001 CCCGACAGGA CTATAAAG 3101 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTC 3201 CGGTTCAGC CGACCGCTC 3301 GATTAGCAGA GCGAGGTAC 3401 CTGAAGCCAG TTACCTTC 3501 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCACTGG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTC 4101 TGTGCAAAAA AGCGGTTC 4301 CCGGCGTCAA TACGGGATC 4301 CCGGCGTCAA TACGGGATC 4401 CGTGTTGAG ATCCAGTTC 4401 CGCTGTTGAG ATCCAGTTC 4501 GCAAAATGCC GCAAAAAAC 4501 GCAAAAATGCC GCAAAAAAC 4501 GCAAAAAACCC GCAAAAAAC 4501 GCAAAAAACCC GCAAAAAACCC	AAAACTTTA   AAACCCGCTG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAAC TCCCCCTGGA AGCTCACCT GTAATACCGT GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGATTG CTCCGATCGT CTTCGATCGT CTTCTGCATCGT CTTCTGTCACT CTTCGTCTAT CGCAGACCC CAGTCTATT CGTCGATCGT CTCCGATCGT CATAGCAGAA CTCGTGCACC GACACGGAAA	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGAA TTCTACGGGA ATTCTACGGGA ATTCTACGGGA ATTCTACGGGA ATTGTTCAC CGAATTGTTCC CTATGGCTTC CTATGGGTTC CTTGAGAGTC TTCTAGAGTC CTTTAGAGTC CTTTAGAGTC TTCTAGAGTC TTTCACTC TTTCAGAGT TTCTTCACAGGG GGTGAGTACT CTTTAAAAGT CAACTGATCT TGTTGAAATAC CATGATCAT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTACCG ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCAGCT TCTGACGCT TTAAATCAAT GGCAACCAC GCTCAGCG TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGAT GGCAAGTC CACCAGATT CGCAACTAC TCCACACT CCCAACTC CACCAACT CCCAACTT CCCAACTT TCACACTCT TCAACACTT TCACATCTT TCACATCTT TCACATCTT TCACATCTT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGCACCAC CAGCAGCAT TACACACAC TATCAGCAAC TATCAGCAAC CAGTAGTAGC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTTA GCTCCAAGCT ACTGACTTA AACCACACT AACCACACT TATGAGTAAA TCGTGTAGAT TATGACTAA AACCACACT TATGACTAA AACCACACT CACCACTTA GATCAAGCCC TCGCCAGTTA GATCAAGCCC ACTCATGTT TATGATTATGC CTTCGGGGCG CACGTTTT TATGATTATGC CTTCGGGGCG CACGTTTTTATTTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTTGC GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAAACCA CTGGTAACAC CTGCTATCA CAGAACCAC AGGTAACAC CTGCGAACCC CTGGTAACAC CTGCGATTACAC CAGGTACCA CAGTTACCA CAGGTACCA CAGTTACCA CGGAGGGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGAGTCTTAC AGAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG
2001 GGATACTTGA GTTAGTTTY  D T  2010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTGTT 2501 ACCTGCAGGC ATGCAAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATCGGCCAC 2801 CGGCGAGGGG TATCAAGCT 2901 CCAGGAACCG TAAAAAGGA 3001 CCCGCACAGGA CTATAAAGG 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTTCAGCC CGACCGCTC 3301 GATTAGCAGA GCGAGGTAC 3401 CTGAAGCCAG TTACCTTCC 3501 CGCGCAGAAA AAAAGGATT 3601 ATTATCAAAA AGGATCTTC 3501 AGCGCAGAAAA AGGAGCTTC 3601 ATTATCAAAA AGGATCTTC 3601 AGTGGTCTC GCAACTTTC 4001 GCCATTGCTA CAGGCATCC 4101 TGTGCAAAAAA AGCGGTTAC 4201 TCTTACTTC ATGCGCATC 4201 TCTTACTTCA ATGCCACTC 4201 TCTTACTTCA ATGCCACTC 4201 TCTTACTTCA ATGCCACTC 4201 TCTTACTTCA ATGCGATC 4401 CGCGCTCAA TACGGGATT 4401 CGCGCTCAA TACGGGATC 4501 GCAAAATGCC GCAAAAAAAC 4501 CTCATGAGCCG GCAAAAAAAC 4501 CTCATGAGCCG GCAAAAAAAC	AAAACTTTA   AAACCCGCTG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTATAGC TGAGCTATAGC TGAGCTATAGC GTATATACGT CGTTATACGGT AGCCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACCC TAAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT TCGTCGATCCT CCAGTCTTTT CTCGTCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTC CATAGCAGAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCACGGAAAATA	CAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TAGACTCACCG AAATGAAGTT TTCGTTCATC ACGCTCACCG GTAGGTATCT TCGTCAGAAGT TCTTCATCAGC GTATGGCTT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CGTCACCG TATTAGATAC TGTTGAATAC CACAGATAGG CAACTGATCT TGTTGAATAC AACAAATAGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCATTATTTG AAAGCAAGTA GCGTTGACTTTCGG ATCAGGGAT CCCCCTGACG GCTCTTCCGC ATCAGGGGAT ACAGCACTA ACAGCACTA ACAGCACTA GGCTAACT ACAAGCACT TTAAATCAAT ACAAGCACC CTCGAGGTTGC GCTCAGGT ACAGCACT ACAGCACT ATTAGCC CCCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCC CACCAAGTC CCCACAAGTC CCCCACAACT CGCCACAACT CGCTCACACT TCAAGCTCT TCAACCATT TCAGCATCTT TCATACCTT TCATACCTCT TCATACCTCT GGTTCCCGCG	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGCACCAC CAGCAGCAT TACACACAC TATCAGCAAC TATCAGCAAC CAGTAGTAGC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTTA GCTCCAAGCT ACTGACTTA AACCACACT AACCACACT TATGAGTAAA TCGTGTAGAT TATGACTAA AACCACACT TATGACTAA AACCACACT CACCACTTA GATCAAGCCC TCGCCAGTTA GATCAAGCCC ACTCATGTT TATGATTATGC CTTCGGGGCG CACGTTTT TATGATTATGC CTTCGGGGCG CACGTTTTTATTTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTTGC GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAAACCA CTGGTAACAC CTGCTATCA CAGAACCAC AGGTAACAC CTGCGAACCC CTGGTAACAC CTGCGATTACAC CAGGTACCA CAGTTACCA CAGGTACCA CAGTTACCA CGGAGGGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGAGTCTTAC AGAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG
2001 GGATACTTGA GTTAGTTTY  D T  2011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGGGC 2701 GCATTAATGA ATGGCCAA 2801 CGGCGAGGG TATCAAGGC 2901 CCAGGAACGG TATAAAAG 3001 CCCGACAGGA CTATAAAG 3101 CCCGACAGGA CTATAAAG 3101 CTCCCTTCGG GAAGCGTC 3201 CGGTTCAGC CGACCGCTC 3301 GATTAGCAGA GCGAGGTAC 3401 CTGAAGCCAG TTACCTTC 3501 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCACTGG 3901 AAGTGGTCCT GCAACTTT 4001 GCCATTGCTA CAGGCATC 4101 TGTGCAAAAA AGCGGTTC 4101 TGTGCAAAAA AGCGGTTC 4301 CCGGCGTCAA TACGGGATC 4301 CCGGCGTCAA TACGGGATC 4401 CGTGTTGAG ATCCAGTTC 4401 CGCTGTTGAG ATCCAGTTC 4501 GCAAAATGCC GCAAAAAAC 4501 GCAAAAATGCC GCAAAAAAC 4501 GCAAAAAACCC GCAAAAAAC 4501 GCAAAAAACCC GCAAAAAACCC	AAAACTTTA   AAACCCGCTG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTATAGC TGAGCTATAGC TGAGCTATAGC GTATATACGT CGTTATACGGT AGCCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACCC TAAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT TCGTCGATCCT CCAGTCTTTT CTCGTCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTG CTCGATCGTTTC CATAGCAGAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACCGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCGTGCACC CACACGGAAA CTCACGGAAAATA	CAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TAGACTCACCG AAATGAAGTT TTCGTTCATC ACGCTCACCG GTAGGTATCT TCGTCAGAAGT TCTTCATCAGC GTATGGCTT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CGTCACCG TATTAGATAC TGTTGAATAC CACAGATAGG CAACTGATCT TGTTGAATAC AACAAATAGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCATTATTTG AAAGCAAGTA GCGTTGACTTTCGG ATCAGGGAT CCCCCTGACG GCTCTTCCGC ATCAGGGGAT ACAGCACTA GGCCTAACT ACAGCACTA GGCCTAACT ACAAACAAT ACAAACAAT ACAAACAAT ACAAACAA	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGCACCAC CAGCAGCAT TACACACAC TATCAGCAAC TATCAGCAAC CAGTAGTAGC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTTA GCTCCAAGCT ACTGACTTA AACCACACT AACCACACT TATGAGTAAA TCGTGTAGAT TATGACTAA AACCACACT TATGACTAA AACCACACT CACCACTTA GATCAAGCCC TCGCCAGTTA GATCAAGCCC ACTCATGTT TATGATTATGC CTTCGGGGCG CACGTTTT TATGATTATGC CTTCGGGGCG CACGTTTTTATTTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTTGC GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAAACCA CTGGTAACAC CTGCTATCA CAGAACCAC AGGTAACAC CTGCGAACCC CTGGTAACAC CTGCGATTACAC CAGGTACCA CAGTTACCA CAGGTACCA CAGTTACCA CGGAGGGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGAGTCTTAC AGAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG

### pMT-HisFlag-DmSNAP43 mut#4 STOP

1 TCGCGCGTTT CGGTGATGA	CCTCAAAACC	TCTCACACAT	CCACCTCCCC	CACACCCTCA	CACCTTCTCT	CTAACCCCAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGGGT								
201 CCGCACAGAT GCGTAAGGA								
301 TACGCCAGCT GGCGAAAGG								
401 TGAATTAATT CGTTGCAGG								
501 GGCCCCCCAC CGCCCACCG								
601 CAAGTCCCCA AAGTGGAGA								
701 AGAGGTGAAT CGAACGAAA								
801 AAATCAAGTG AATCATCTC	A GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT				CTACAAGGAT
					> M H	н н н	H H T D	Y K D
901 GACGATGACA AGGGCACTA	G TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
> D D D K G T	BELN	I F D	D C W E	L V Q	R F Q	R L V N	D G E	N C E
1001 TCGAGGTGTT CTGCCGGTG	C TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
>FEVFCRC	WRE	L O L O	H L F	T A O	T N H T	E V I	ATT	L A A L
1101 GCATGTGGCC AAGCGACTG	CGTGCTCCCG	ACGCACCACC	GGGGACGTTT	TCCCGGCATC	TCGCGCTCAA	AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
> H V A K R L	SCSR	RTT	G D V	F P A S	R A O	R I G	G F F L	LYV
1201 ATCTACTACA AGCAGCCCA	CGCACAACTTT							CGCAAGGATA
> I Y Y K O P			E V S P	R T W		T D Y A		R K D
1301 GTCCGGAGCG GAAGGACAC					~			
>S P E R K D T	H O I	A Y M L		T O E	O A F R		L D Y	C O G L
1401 GGACAATCTG GTGGACTAC								
> D N L V D Y	D R V E		G A K			M O K		
						~	~ ~	
1501 GCTCTCGCAG CCGAACTGG								
> <b>A</b> L <b>A</b> A E L		ALD		P L C		AAYN	A Q K	K Q L
1601 CGGCTGGTCA TGAGCACGC								
>A A G H E H A		S Q I F		R E V		Q S V	L G A	R K S T
1701 TCCAGATGAG AAATGCACC								
, , , , , , , , , , , , , , , , , , , ,	TTST	~	L E V	~		A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGAAC	r agaagtgcag	CTGGAGGTCA	ACGAGACTTA	TCAACGCCGC	ATGTCCTCGG	CCACCGTTTT	CCAGAGGGAA	CTTCCAGAAG
> Q H Q T D E	E V Q	L E V	N E T Y	Q R R	M S S	${\tt A}  {\tt T}  {\tt V}  {\tt F}$	Q R E	L P E
1901 ACGTGCAGCA AGAGTATGA	G ATGATTGAGT	TTAGTGACGA	CGAGGAAATG	GAAGTGGGTG	AAAGCGAGGA	GGTCACGGAA	GAAGAACTCA	AAGCTATTTT
>D V Q Q E Y E	M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
>D V Q Q E Y E 2001 GGATACTTGA GTTAGTTTA								
2001 GGATACTTGA GTTAGTTTA	r aaaactttta	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
2001 GGATACTTGA GTTAGTTTA  > D T	F AAAACTTTTA G AGTCTAGAGG	CATAATTAAA GCCCGCGGTT	TAACTAGCAT CGAAGGTAAG	TTTTGCGCGA CCTATCCCTA	TGTGATCTTG ACCCTCTCCT	TTTATCTGAA CGGTCTCGAT	GGGCAATTCT TCTACGCGTA	GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTTTA  > D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT	AAAACTTTTA G AGTCTAGAGG F AAACCCGCTG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA	TGTGATCTTG  ACCCTCTCCT GACATGATAA	TTTATCTGAA CGGTCTCGAT GATACATTGA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA	GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA GTTAGTTTA  > D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA	F AAAACTTTTA G AGTCTAGAGG F AAACCCGCTG A TGCTTTATTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA GTTAGTTTA  D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT CAGGTTCAGGG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	TTTTGCGCGA  CCTATCCCTA  CTAAGATCCA GCTTTATTTG AAAGCAAGTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA GTTAGTTTA  D T  1010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT	AAAACTTTTA  GATCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA GTTAGTTTA  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG	AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTTTA  > D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTC AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCCGGGGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC	AAAACTTTA  AGTCTAGAGG AAACCGGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGGAG ACTCAAAGGCG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	TTTTGCGCGA  CCTATCCCTA  CTAAGATCCA GCTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG
2001 GGATACTTGA GTTAGTTTA  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAAGGC	AAAACTTTA AGTCTAGAGG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGGAGA ACTCAAAGGCG GCGTTGCTGG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	TTTTGCGCGA  CCTATCCCTA  CTAAGATCCA  GCTTTATTTG  AAAGCAAGTA  GTGAAATTGT  GCGTTGCGCT  GCTCTTCCGC  ATCAGGGGAT  CCCCTGACG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAAA	AAAACTTTA  AGTCTAGAGG AAACCGCTG ATGCTTATTT AGGTTCAGGG GGCGTAATGAG GCCCTAATGAG CGCGCGGGAG ACTCAAAGGCG CGCGTGCTGCTGG ACCAGGCGTT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCGTTTTTCCA TCCCCTGGA	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGATTTCCTGT  CACATTAATT  CGTATTAGGC  TATCCACAGA  TAGGCTCCGC  AGCTCCCTCG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT TCCCCCTGACG TGCGCTCTCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGATCAGA CCGGATACCT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CGGCAAAAGG GCTGGCGAAA GTCCGCCTTT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGAGA ACTCAAAGGCG GCGTTCTGC ACCTTTCTCAT GCTTTCTCAT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA AGCTCTGGA AGCTCACGCT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTACT	TTTTGCGCGA  CCTATCCCTA CTAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCCC CAGTTCGGTG CAGTTCGGTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCGCC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTCT	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTCGC ACGACACTGC ACACACTGC CTGCCGCTTA GCTCCAAGCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2001 GGATACTTGA GTTAGTTTA  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ATCCGCCAA 2801 CGGCGACGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAAGG 3001 CCCGCTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG CCTAAATGAG CCTCAAAGGCG CCCGCGGGAGAA ACCAGCGTT ACCAGCGGT ACCAGCGGT CGCTTCTCAT CGCTTTCTCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGCCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CACATTAATT CGTATTGGG TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CATGCCGCC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACT TAGGTCGTTC GACACGACTT GACACGACTT GACACGACTT	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTGG ACTGACTGG AGAACATGTG AAAACGCGC CTGCCGCTTA ACTCGACGA ACTCGACGACG ACTCCACACT ACCCCACTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GGCCTGTGTG GCAGCAGCCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCAGCCA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 2901 CCCGACAGGA CTATTAAAGA 3101 CTCCCTTCGG GAACCGCTG 3201 CGGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT	AAAACTTTA AGTCTAGAGG AAACCCGCTG AGCTTATTT AGGTTCAGGG GGCGTAATCA CGCTAATGAG CGCGGGGAG ACTCAAAGGCG GGCTTCTGG ACCAGGCGTT CGCTTATCGG TACCAGGCGTT TACCGGGTACCGG TACCAGGCGTT TACCGGGTGCG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAACT TGAGCTAACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCGT TACAGAGTTC TACAGAGTTC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT GTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGAAATTG GCGTTCGGCT ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA AGCCGGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT AAACCTTAC TATCCGCTCA CACTGCCGC ACCGCGACA AGCATCACAA TGTTCCAGCT TAGTCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACAT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGAT CAATTCCACA TTTCCAGTCG ACTGACATCGACA ACTGACTGAC ACTGACTGAC ACTGCCGCTTA GCTCCAACGT ATCGCCACT ACGAGGACACAC ACGAGCACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGCACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACGAGGACACAC ACCACACACA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATAC GGAAAACTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGTG GGCTGTGTG GGACCAGCCA TATTTGGTAT	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGAGA ACTCAAAGGCG GCGTTCTCG ACCTTTCTCAT GCCTTATCAG GTTGCTGG GTAGCGGTGC AAAAAGAGTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGTCATAGC TGAGCTAACT GTAATACGGT CGTTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCTT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT GGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCGC GGTAGGTACT TCTTGAGTCC TTGAAGTGGC GATCCGCAAG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTG GCGTTGCGCT ACCCTGACG ATCAGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCAGTAACTA ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC ACCAACCAC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCTCGGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACATC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG ACTGACTGCA AGACATGTG AAATCGACGC TTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAGACATGTG AGAGGACAG GTGGTTTTTT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CGACTGATTA CAACATCGA GGAAAACCTGT TGCGCTCCGGT AGCAAAACGC TCAAGTCAGA CCGGATAACCTG GCAGTGTTG GCACCAGCACCA TATTTGGTAT TGTTTGCAAG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTCGCAGC GGTCGGCTG CAGCAAAAAGG GGTGGCGACA TCAGCACATTA CAGCACCTCTG CAGCACTTC CAGCACTTC CAGCACTTC CAGCACTTC CAGCACTTC CAGCACATTA
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATC	AAAACTTTA AGCCGCTG AAACCCGCTG ATGCTTATTT AGGTTATATT GGCTAATGAG GCCTAATGAG CTCAAAGGCG CTCAAAGGCG CTCAAAGGCGT ACCAGCGTT CGCTTCTCAT CGCTTATCCG TAGGCGGTGC AAAAAGAGTT CAAAAAGATT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTAACTATCG TACAGAGTTC GGTAGCTCT TTTGATCT CTTTGATCT	CAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGATTCCTGT  CACATTAATT  CGTATTGGGC  AGCTCCCTCG  GTAGGTATCT  TCTTGAGTCC  TTGAAGTGGT  GATCCGCAA  GATCCGCAA  TTCTACGGGG  TTCTACGGGG  TTCTACGGGG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCCTTACGC TCGCGCTTCCGC ACCCCTGACG TGCGCTTCCGTG AACCCGCTAAC GGCCTAACTA GGCCTAACTA ACAAACCACC TCTGACGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CATGCCGC TTCCTGCTC AACGCAGGAA AGCATCACAA TGTTCCGACT GACACGGTTCCGTCT GACACGACT GGTGGTACACT GGTGGTACACT GGTTGGTACGC AGTGGAACGA	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTGG ACTGACTGG ACTGACTGG CTGCCGCTTA AAATCGACGC ATCGCCACTA ACACCACT ACAAGCACAG ATCGCCACTT AGAAGGACAG GTGGTTTTTT AAACTCACGT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCTAC TGTTTGCAAG TAAGGGATTC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGCTCG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGGCTTCG CAGCAGATTA CAGCAGATTA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2401 TAGAGTTAAA AGCCTGCGGG 2701 GCATTAATGA ATCCGCCCA 2801 CGGCGAGCGG TATCAACTC 2901 CCAGGAACCG TAAAAAGGCT 2901 CCCGACAGGA CTATAAAAA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3201 CGGTTCAGCC GGACCGCTTG 3401 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCAGA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCTAAATGAG GCGGGGAG ACTCAAAGGCG ACCTTATCTG ACCAGGCGTT GCCTTATCCG TAGCGCGTT GCCTTATCCG TAGGCGGTGC AAAAACAGTT CAAGAAGAGT ACCAGAACCCT ACCACACCCT ACCAGACCCT ACCACACCCT ACCACACCCT ACCACACCCT ACCACACCCT ACCACCCT ACCACACCCT ACCACACCT ACCACACCCT ACCACACCCT ACCACACCCT ACCACACCCT ACCACACCCT ACCACACCT ACCACACCCT ACCACACCC ACCACACC ACCACACCC ACCACACCAC	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTAAGC TGAGCTAACT TGACCTAACT GTAATACGGT CGTTTTTCCA AGCTCACCT GTAACATATC GTAACTATC GTAACTATC GTAACTATC GGTAGCTCT TTCAACAGTT TTTTAAATTAA	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTCCTGT CGAGTTTTTT GGTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT CTTTGAGTCC TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TCTCACGGGA TATCCACAGA	CTATCCCTA CTAAGATCCA CCTATCCTA CTAAGATCCA CCTTATTTG GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA ACAAACCAC TCTGACGCT TCTGACGCT TTAAATCAAT	ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAAT TATCCGCTCA CACTGCCGC TCCCTCGCTCAACGCAGCA AGCATCACAA AGCATCACAA GCATCACAC TGTTCCCGCTC AACGCAGGAA AGCATCACAA GCATCACAC TGGCTACACT GACACGACTT CGGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGACACTGC ATCGCCGCTTA GCTCCAAGGT ATCGCCACTG AGAAGAGACA GTGGTTTTTT AAACTCACGT TATCAGGT ATATGAGTAAA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACA GGAAAACTGT TGCCGTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGT GCAGCCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTC CAGCAAAAGG GGTGGGCAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CAGCAGATTA CTGCGCTCTT CAGCAGATTA CAGCAGATTA CAGCAGATTA TGGTCATACAG CAGTTACCAA
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACT 2601 TAAAGTGTA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGAC 3001 CCCGCACAGGA CTATAAACA 3101 CTCCCTTCGG GAAGCGTGG 3201 CAGTAACAGA GCACACGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTC 3701 TGCTTTAATCA GTGAGGCAC	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGAG ACTCAAAGGCG GCTTCTCTCAT CGCTTTTCTCAT CGCCTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGATC CCTAAGAGATC CCTAAGACCT TCAAGAAGATC CCTAAGACCT TTATCAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGTCATAGC TGATACGGT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTAA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTACTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC AGCTCCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGCT TTGAAGTGCT ATTCACGGGA TCTTACGGGG AATCAGAGTT TTCCTTCATC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGT AACCCAGTAACTA ACCACCTAACTA GCACTAACTA ACAAACCACC TCTGACGCT TCTGACGCT TTAAATCAAT CATAGTTTGCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC AACGCAGGAA AGCATCACAA AGCATCACAA CTGTCCGAC TGTCCCACC TAGGTCGTC GACACGACTT CGGCTACACT GCTGGTACCG CTGGTACCG AGTGGAACGA CTAAAGTATA TGACTCCCCC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG ACTGACTGCA AAATCGACGC AGAACATGTG AAATCGACGC AGACATGTT ACTCCAGCTG ACTCCCAAGCT ATCGCCACTG AGAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGAGTAA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CGACTACGA GGAAAACTGA TGCCGCTCGGT AGCAAAACGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTTGCAAG TAGGGATTC AACTACGAA AACTACGAA TAGGGATTT CTTTGCTAT AACTACGATA AACTACGATA AACTACGATA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCTC CAGCAAAAAGC GGTGGCGTC CAGCAAAAGC GTTCGCTTC CAGCACATTA TGGTTCATGAG CAGTTACCAA CAGTTACCAA CAGTTACCAA CAGTTACCAA CAGGAGGGCT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGACGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCAGGTAT 401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATCTC 3701 TGCTTAATCA GTGAGGCCC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC	AAAACTTTA AGCCGCTG AAACCCGCTG ATGCTTATTT AGGTTCATGGG GGCGTAATCA CCCAAAAGGCG CGCGGGGAGA ACCAGGCGTT ACCAGGGGT ACCAGGGTT CGCTTATCCG TAGGCGGTGC TAGCAGGTTC CAAAAAGGTT CAAGAAGATT CAAGAAGATT TATCTCAGC GCATGCTGC GCATGCTGC	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC CTTAGACTATC TTTAAATTAA ATCTGTTCTAT CCGCGGAGCCC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TCTTCAGGGGA AATGAAGTT TCTTACGGGGA TTCTACGGGA TTCTACGGGA AATGAAGTT TCTTCACTCACCACACACCACCACCACCACCACCACCACC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCCTGACG TGCGCTCTCCGC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTAACTA CACAACCACT TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTCTAC CATGCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC GACAGGACT GCTGCTTC GACACGACT GCTGCTACACT GCTGGTACACT GCTGGTACACG AGTGGAACGA TAAAGTATA TGACTCCCCC TATCAGCACT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA ACGATTCCACA CTGCACTCGACTGC ACTGACTCGC CTGCCGCTTA AGACGCC ATCGCCACTG AGACGAGGC ATCGCCACTG AGACGAGGC ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGTAAA TCGTGTAAAT AAACCACGCA AAACCAGCCA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TGTTGTCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGCG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTCGGCTG CAGCAAAAGG GTGGCGAAA GTCCGCCTT CAGCAACCCC CTGGTAACAG CTGGCTCTG CAGCACATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCGAGCGCAC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACT 2601 TAAAGTGTA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGAC 3001 CCCGCACAGGA CTATAAACA 3101 CTCCCTTCGG GAAGCGTGG 3201 CAGTAACAGA GCACACGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTC 3701 TGCTTTAATCA GTGAGGCAC	AAAACTTTA AGCCGCTG AAACCCGCTG ATGCTTATTT AGGTTCATGGG GGCGTAATCA CCCAAAAGGCG CGCGGGGAGA ACCAGGCGTT ACCAGGGGT ACCAGGGTT CGCTTATCCG TAGGCGGTGC TAGCAGGTTC CAAAAAGGTT CAAGAAGATT CAAGAAGATT TATCTCAGC GCATGCTGC GCATGCTGC	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC CTTAGACTATC TTTAAATTAA ATCTGTTCTAT CCGCGGAGCCC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TCTTCAGGGGA AATGAAGTT TCTTACGGGGA TTCTACGGGA TTCTACGGGA AATGAAGTT TCTTCACTCACCACACACCACCACCACCACCACCACCACC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCCTGACG TGCGCTCTCCGC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCAGCTAACTA CACAACCACT TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTCTAC CATGCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC GACAGGACT GCTGCTTC GACACGACT GCTGCTACACT GCTGGTACACT GCTGGTACACG AGTGGAACGA TAAAGTATA TGACTCCCCC TATCAGCACT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA ACGATTCCACA CTGCACTCGACTGC ACTGACTCGC CTGCCGCTTA AGACGCC ATCGCCACTG AGACGAGGC ATCGCCACTG AGACGAGGC ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGTAAA TCGTGTAAAT AAACCACGCA AAACCAGCCA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TGTTGTCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGCG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTCGGCTG CAGCAAAAGG GTGGCGAAA GTCCGCCTT CAGCAACCCC CTGGTAACAG CTGGCTCTG CAGCACATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCGAGCGCAC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGACGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCAGGTAT 401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATCTC 3701 TGCTTAATCA GTGAGGCCC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC	AAAACTTTA AGTCTAGAGG AAACCCGCTG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCTAATGAG GCGGGAGA ACTCAAAGGCG ACTCAAAGGCG TGCTTATCTC TACCTTATCCG TAGGCGTTCCAG TAGGCGGTG AAAAACAGTT CAAGAAGAGT ACTAAAACAGTT CAAGAAGAGT ACTAAAACAGTT TAACTCAGCG TATCTCAGCG TATCTCAGCG TATCTCAGCG TATCTCAGCG TATCTCAGCG TATCTCAGCG CGCATCATA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTATAGC TGGCCTATAGC TGACCTAACT GTAATACCGT CGTTTTTCCA AGCTCACCT GTAACATATC GTAACTATC GTAACTATC GTAACTATC GGTAACTCT TTTTAAATTAA ATCTGTCTAT ACCGCAGACCC CCAGTCTATT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT CTTGAGTCC TCTTGAGTCC TCTTGAGTCG TATCCACCAGA TTCTACGGGA TTCTACGGGA TTCTACGGGA TTCTACGGGA TTCTACGGCA ATTGTTCACC	CTATCCCTA CTAAGATCCA CCTATCCTA CTAAGATCCA CCTTATTTG GCATGAAATTGT GCGTTGCGCT ACCCCCTGACG ACCAGGGAT CCCCCTGACG ACCAGGTAA ACCACACTAACTA ACAAACCAC CTCTAACTAACT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAAT TATCCGCTCA CACTGCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GCTGGTACACT GCTGGTACGC AGTGGAACGA TAAAGTATAT TGACTCCCCG TATCAGCAT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAGCTA ATCGCCACTG ATCGCCACTG AGAAGACAGCT ATCGCCACTG ATCACCACTT AAACTCACGT ATACACGT ATACACGT AAACTCACGT AAACTCACGT ATGGGATAAA TCGTGTAGAT AAACCAGCCA TCGGCAGTTA AACCAGCCA TCGCCAGTTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTAC GGAACTTG TGCCGTATTA CAACATACGA TGCCTCGGT TGCCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGGTAT TATTACGATA ACTACGATA ACTACGATA GCCGGAAGGG ATAGTTTTGC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTC CAGCAAAAGG GGTCGGCATT CACGAACCCC CTGGTAACAG CTGCGCTTCT CACGAGATTAACAG CTGCGCTCTC CACCAGATTAACAG CAGATTAACAG CAGATTAACAG CAGATTACAGA CAGTTACCAA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2401 TAGAATGTAA ACCTGCGGG 2701 GCATTAATGA ATCCGCCCA 2801 CGGCGAGCGG TATCAACTC 2901 CCAGGAACCG TAAAAAGGCT 2901 CCCGACCAGGA CTATAAAAA 3101 CTCCCTTCGG GAACCGCTG 3201 CCGTTCAGCC CGACCGCTG 3201 CGTTCAGCC GGACCGCTTG 3401 CTGAAGCAGA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3501 TACCATCTGG 3801 TACCATCTGG 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTA	AAAACTTTA AGTCTAGAGG AAACCCGCTG AGTCTCAGGG GGCGTAATCA GGCTAATGAG GGCGGAGA ACCGGGGAG ACCGGGGAG ACCGGGGGGGG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAAGC TGAGCTAAGC TGACTATAGC GTTATTCGA AGCTGACTTCG TACAGCAGTTC GTAACTATCG TACAGCAGTTC GTTAGCTTTT CTTTGATCTT TTTTAATCTT TTTAATCTT TTTAATCTT TTTAATCTT TTTAATCTT TCGCGAGACCC CCAGTCCTTTT TCGTCGTTTT TCGTCGTTTT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TATCACAGA TTCTACAGGG AATCAAGAT TTTACAGGG AATCATACAGT TTCGTTCAT ACGCTCACCG AATGATTATC ACGCTCACCG AATTGATTCCC GTATGGCTTCC GTATGGCTTCC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACCACCGTTAA ACCACACTA ACAACCAC TCTGACGTC TCTGACGTC TTAAATCAA CATAGTTGCC GCTCCAGATT GCGCATACTA ACAACCAC TCTGACGTC TCTGACGTC CTCTGACGTC GCTCCAGATT GGGAAGCAT GGGAAGCTAC ATTCAGCTCC ATTCAGC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCTGCGCCG ACCAGGAA AGCATCACAA AGCATCACAC TGTTCCCAAC TGTTCCAAC TGTTCCAAC TGGCTACTC GACACGACT GCTGGTACCG ACTGGAACGA TAAAGGATAT TGACTCCCG TATCAGCAAT TGACTCCCCG TATCAGCAAT GGTTCCCCAAC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATCTGATC CAATTCCAGTCG AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGAGCG AGACATTTT GCTCCAAGCT ATCGCCACTG AGAACACTGT AAACTCACCATTA AAACTCACGTA AAACTCACGTA AAACTCACGTA AAACCAGCCA TCGTGAGAT AAACCAGCCA TTCGTGAGAT AAACCAGCCA TCGCAGCTA TCGCCAGCTA TCACAGCCG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTAC GGCAGAACCTGT TGCCGTCGGT AGCAAAACGGC TCAAGTCAGA CCGGATACCGG GGCACGGATAC GGCTGTGTG GCACCAGCCA TATTTGGCAT TGTTTGCAAG TATTGGTTTG AACTACGAT CCGGAAGGGATTA GCCGGAAGGG ATAGTTTGCA	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAG GGTGGCGAA GTCGCCTT CACGAACCC CTGGTAACAG CTGGCTCTC CAGCAGATTA TGGTCAGAG CAGTACACA CGGGAGGCCAG CAGTACCAA CGGGAGGCCT CCAGCACTT CCCCATGT
2001 GGATACTTGA GTTAGTTTA	AAAACTTTA AGCCGCTG AAACCCGCTG AGCTTATTT AGGTTCATGA GGCGTAATCA GCCTAATGAG CGCGGGGGAG ACCAGCGTT CGCTTCTCTC GCTTTCTCAT CGCTTATCCG TAGGCGGTGC AAAAAGGTT CAAGAAGGTT CAAGAAGGTT CAAGAAGTT CAAGAAGTT CAAGAAGTC CCTAATCAC CCGCCTCAT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCT GTAACTATC GTAACTATC GTAACTATC TTAAATTAA ATCTAA ATCTGTCTT CGCGAGACCC CCAGTCTATT CGCGATCGT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATGTTTCCTGT CGCATTAGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TCTACGGGA ATCTACGGGA AATGAAGTGT TTCCTTCACCGAATTATTCACCGCAATTCACTCACGAATTCATCACGGGAATTCTACGGGAATTCTACGGGAATTCTACGGGAATTCTCATCACGCGAATTCTCATCACGCGAATTCTCATCACGCGAATTCTACGCTCACGGAATTCTTCTTCATCACGCGAATTCTTCTTCATCACAGAGTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTCGTCAGAAGTTTTTGTCAGAAGTTTTTGTCAGAAGTTTTTGTCACTTTGTCACAGAGTTTTGTCACTTTGTCACAATTTTTGCTCAGAAGTTTTGTTCATCATTTTGTCACAATTTTTGTCACAAGTTTTTGTCACAATTTTTGTCAGAAGTTTTTGTCACAATTTTTTTGTCACAATTTTTTGTCACAATTTTTTTT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAT TTAAATCAAT TATAATCAAT CATAGTTGCC GCTCAGATTGCC GCTCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTCTAC CATTGCCGCC TTCCTCGCTC ACCTGCCCC TTCCTCGACC TGGTCCTCACC TGGTCCTTCCACC CAGGAAAGCATCACAA TGTTCCGACC CAGGACAAGCAAAGCA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA AAATCACCACA TTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACATG AGACATTTT AAACCACCT ACAGCACTT ACAGCACTT ACAGCACT ACACCACT ACACC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTGG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTTCGA ATGGCAGCAG ATGGTAGAA ATGCCAGCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTC CAGCAAAAGG GCTGGCAAA GTCCGCCTTT CAGCAACCCC CTGGTAACAG CTGGCTCTG CAGCAGATTA TGGTCATGAG CAGCAGTTCAT CGGAGGGCT CCGAGCGCAG CAGCGGTTGT TCCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGCA CGACCGCTG 3301 GATTAGCAGA GCAGGTAT 401 CGGAGGAAA AAAAGGATCTTC 3501 CGCGCAGAAA AAAAGGACTTC 3701 TGCTTAATCAA AGGACTTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA 401 GCCATTGCTA CAGCCATTGG 4101 TGTGCAAAAA AGCGGTTAG	GAAAACTTTA GAGTCTAGAGG AAACCCGCTG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCGTTACTGG ACCAGAGGCG ACCAGAGGCGT GCCTTATCGG ACCAGGCGTT CGCTTATCGG ACAGGCGTT CGCTTATCG ACAGAGAGAT CAGAAAGAGT ACAGAAGAGAT ACCAGAAGAGAT CAGAAGAGAT CAGAAGAGAT CCGCCTCAT CGGCTTCAGC GGCAATGATAC CCGCCTCAT CGGGTTCAGG CCTCCTCGGT CCTCCGGTC TAAGATGATC	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAACT GTAATACCGT CGTTTTTCCA AGCTCACTATC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG TTCTGTCGTTTG TTCTGTCACT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGAA TTCTACGGGA TTCTACGGGA ATTCTACGGGA ATTCTACTCACCG AATTGTTCAC CGAATTGTTCCC GTAGGCTTCCCC GTAGGGTACT TCTGAGAGT TTCGTCACAGG GGTAGGTTCCC GTATGGCTTC CGAAGAGT TGTCAGAAGT TGTCAGAAGT	CCTATCCCTA CTAAGATCCA CCTATCCTA CTAAGATCCA CCTTTATTTG GTGAAATTG GCGTTGCGCT GCCCTGACG ATCAGGGAT CCCCCTGACG ACCAGGTAA AACCAACC CCTGACG TCGCCTTACT CCCCGTAA CAAACCAC TCTGACGTC TCTGACGTC TTAAATCAAT CATAGTTCGC GCTCCAGATT GGGAACTAC ATCAGCTCC AGTTCAGCTC ACTAGCTCC AGTTCAGCTC CCCCAGATT GCGAACCTAC ATCAGCTTC AACTAGCTCC AACTAGCTCC AACTAGCTCC AACTAGCTCC CAACCTAGATT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAC TATCCGCTCA CACTGCCGC TACCTCGCTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGACACAC TAGACACAC TAGACACAC TATACAGCAT TATCAGCAAC TATCAGCAAC CAGTTATCCAAC CAGTTATCCAAC CAGTTTATC CAGTTATCAAC CAGTTTATC CAGTTATCAACA TATTCTAGAAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA GCTCCAAGGT ATCGCCAGTT AAACTCACGT TATGAGGACA GTGGTTTTT AAACTCACGT AAACCAGCA AAACACGCCA ACGCAGTTA GATCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGCG ACTCAAGGTT TAGAGTATATG	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTAGCAACAGTA AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGCGGATACCT GGCGTGTGG CCAGCAACA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT AACTACGATA GCCGGAAGGG ACCGAGCAA AATTACGATA GCCGGAAGGG AATACTTTGCA AATTACGATA GCCGGAAGGG ATACTTTGCA AGTTACATAA AATGCAATA GCCGAACGG AGTTACATGA ATGCAACAC GGCGACCCAC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAAGG GGTCGGCATT CACGAACACCA CTGGTTAACAG CTGGCTTCT CAGCACATTA CAGCAGATTA TGGTCATGACAG CAGATTACAGA CAGATTACAGA CAGATTACAGA CAGATTACATACCAA CGGGAGGGCT CCCGAGCGCAG CAACGTTGTT TCCCCCATGT TGCATATATC TTGCTCTTTGC
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TAAAAAGGCT 2901 CCCGACCAGGA CTATAAAAA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCAGA TAACCTTCG 3501 CGCGCAGAAA AAAAGGATC 3501 TACCATCTGG CCCCAGTGC 3601 ATTATCAAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGCCAC 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGATTCG 4101 TGTGCAAAAA AGCGATTCG 4101 TGTGCAAAAA AGCGGTTAG	AAAACTTTA AGTCTAGAGG AAACCCGCTG AGTCTAGTATATT GCTTTATTT GCTTAGGG GGCGGAGA CTCAAAGGCG CGCGGGAGA ACCGGGGAGA ACCGGGGAGGAGACACCTTACCG GCTTTCTCAT CGCTTATCGG AAAAAGAGTT CAAGAAGATC TATCTCAGCGGAGAGACCCTCAT GGTGTCAGATCT CGCCTCCAT GGTGTCAGGCCTCCAT GGTGTCAGGCCTCCAT GGTGTCAGGCCTCCAT TGCTTCAGGCCTCCAT TGCTTCAGGCC TCCTTCGGTC TAACAAGATCTT ATACCGCCCCC TAACAAGATCTT ATACCGCCCCC TAACAAGATCTT ATACCGCGCCCCCT TAACATGCTT ATACCGCCCCA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAAGC TGAGCTAAGC GGTTTTTCCA TCCCCCTGGA AGCTACCT GTAACTATCG TACAGAGTTC GGTAGCTCT TTTAAATTAA ATCTATCT TTTAAATTAA ATCTGTCTAT TCGTCGAGACCC CCAGTCTTT CTCGGTTTG CTCGGATGCT CTCGGTTTG CTCGGATGCT CTCGATGCT CTCGATCGTTTC CTCGATCGTTTC CATAGCAGAA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTTGAAGTCC TTGAAGTCC TTGAAGTCC TAGACTCCCC GAATCCCCACA TCTTACGGG AATCACCC AATTGTTCATC ACGCTCACCG TATGGCTTC TGTCAGAAGT CGTGAGAGTAC CTTTAAAAGT	TTTTGCGCGA  CCTATCCCTA CTAAGATCA GCTTTATTTG GTGAAATTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACCCGGTAA ACCAAGTA ACAAACCAC TCTGACGCT CTGACGCT CTGACGCT GCGCATTGCGCT AGTTGCCG ATTGCGCG ACTAGTTGCC CAGTTTGCC CAGTTTGCC CCAGATTCC CCCAGATT CCAGATTCC CCCAGATT CCAGATTCC CCCAGATT AATTCACCC CACCAAGTC CAACCAAGTC CCTCATCATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCGGCAGA AGCATCACAA AGCATCACA TGTTCCGACC TAGGTCGTC GACACGACT GGCTACACT GCTGGAAC AGTGAACA AGTATACAA TGACTCCCCG TATCAGCAAC TAGACTACACA TAGACTACACA TAGACTACACA TAGACTACACAC TAGACACACA TAGACACACA TACTAGACAAC AGTGTTACA CAGTGTTATAC AGTAACACAC GGTTCCCAAC CAGTGTTATAC AGTAACACT ATCTAGACAA GGAAAACCTT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATC CAATTCCAGTGG AGACATGCA AATCGAGCG AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGACGC AGCACATGTG AAATCGACGCTA GCTCCAAGCTT ACGCCACTG AGAAGACAGCTA ACACCACCAGTA AACCAGCCA ATCGCAGCTA ACCAGCAGTA AACCAGCCA ACTCAAGGTT AGATCAAGGCG ACTCATGGTT TAGTGATATC CTTCGGGGCG	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTAC GGAAAACCTGT TGCGCTCGGT AGCAAAACGGC AGCAAAACGGC TCAAGTCAGA CCGGATACCGG GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TATGGTATT TGTGTGTGTGT TGTGTGTGTGTGTG	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAA GCACAAAAGG GGTGGCGAC CTGGTAACAG CTGGCTCTG CAGCAACATA TGGTTCATGAG CAGTTACCAA CGGGAGGGCT CCGGGAGGCT CCGGAGGGCT CCGAGCGATT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCATCATTC TTGCTCTTCC AGGATTATC TTGCTCTTTCC AGGATTATC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGGAATGCA GTGAARAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGGC 3001 CCCGGCACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTTAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 ATTATCAAAA AGGATCTC 3601 ATTATCAAAA AGGATCTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCCTCAA TACCGGGATA	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTATTT AGGTTCATGAGG GGCGTAATCA GCCTAATGAG CGCGGGGGAG ACCAGAGGGT CGCTTCTCTGT CGCTTATCCG AAAAGAGTT CAAGAGGTT CAAGAGGT CAAGAGGTT CAAGAGGT CAAGAGG CAAGAG CAAGAGG CAAGGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAGG CAAGAG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCTGGA AGCTCACCCT GTAACTATCG GTAACTATCG GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCTGATCTT TCGTCGTCTAT CGCGAGACCC CCAGTCTATT CGCGATCGT TCCTGGTTGG CATAGCAGAC CATGGTTGC CATAGCAGAC CCTCGTCACC CCTCGTCACC CCTCGTCACC CCTCGTCACC CCTCGTCACC CCTCGTCACC CCTCTCACC CCTCGTCACC CCTCGTCACC CCTCTCACC CCTCTCACC CCTCTCACC CCTCTCACC CCCTCTCACC CCTCTCACC CCCTCTCACC CCCTCTCTCACC CCCCTCTCACC CCCTCTCACC CCCTCTCACC CCCTCTCACC CCCCTCTCACC CCCTCTCACC CCCCTCTCACC CCCCTCTCACC CCCCTCTCACC CCCTCTCACC CCCCTCTCACC CCCTCTCTCT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATGTTCCTGT CGAATTAATT CGTATTCGGC TATCCACAGA TAGGCTCCTCG GTAGGTATT TCTTAAGTC TTGAGTCCTTC GATCCCTCG GATCGCGCA TTCTACGGG AATTAATT ACGCTCACC GATCAGGG AATTCATC ACGCTCACCG CGTAGGGTATCT TCGTTCATC ACGCTCACCG CTAGGGTACT TCTCATCATC ACGCTCACCG CTATGGCTTC TGTCAGAAGT TTCTAAAAAGT CTTTAAAAAGT CAACTGATCT CTTTAAAAAGT CAACTGATCT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCAGCTAACTA ACCAGCTAACTA ACAAACCACC TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCAGATTGCCC GCTCAGATTGCCC AAGTTGCCC AAGTTGCCC CACCAAGTT CGCAACTA CTGACCTC CACCAAGTT CGCAACTACTT TCACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTATC CATTGCCGC TTCCTCGCTC AACGGAA AGCATCACAA TGTTCCGACC TGTCTCTTC GACACGGAA AGCATCACAA TGTTCCGACC TGTGTACAC TGTGTACAC TGTGACACAT TGACTCCCCC TATCAGCAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAGTA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA AAGCTGCAAT AAATTCAACA TTTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAGGACATGT ACTCACC AGAAGATTTTT AAACTCACGT TATGAGTAAA TCGTCAAGCT ACCGCAGTTA AACCACCA CGCCAGTTA AACCAGCCA TCGCCAGTTA CATCAGCAGTA AAACCAGCCA TCGCAGTTA CATCAGGCG ACTCATGGTT TAGGTAAAG CATCAGGCG ACTCATGGTT TAGGTGTAGTC CTTCGGGGCG CAGCGTTTCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCA GGCTGTTTG GCACCAGCA TATTTGGTAT TGTTTGCAAG TAAGGCATT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTGCA ATGGTCTGA ATGGCACCAG ATGGTCAGA ATGCACACAG ATGGCACCAG AGTTACATGA ATGGCACCAG AAAACTCCCA GAGGTGACCAC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTCCGCCTTT CAGCAAACCC CTGGTAACAG CTGCGCTTT TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGCAGGTTCT CCGAGCGCAT TCCCCCATGT TCCCCCATGT TCCCCCATGT TGCATAATTC TTGCTCTTGC AGGAGTCTTTC TGGTCTTTGC AGGAGTCTTTC AGGAGACAAAACAGGAACAAAACAGGAACAA
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGTCAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACACTC 3501 CGCGCAGAAA AAAAGGATCTTC 3701 TGCTTAATCA GTGAGGACTCTG 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GAACTTTA 401 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCGTCAA TACGGGATA 4401 CGCGCGTCAA TACGGGATCA	GAAAACTTTA GAGTCTAGAGG AAACCCGCTG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCTTAATGAG GCTTAATGAG GCTTACTGG ACCAGAGCGGT GCTTCTCGG ACCAGAGCGGT CGCTTATCCG AAAAACAGATC ACCTAAAGAGT CACAAAAAACT ACCAGATCAT GGCTTCAGT GGCATCAGT CTATCCAGCG TACAGATCCT TATCTCAGCG TACCTCAGTC TATCTCAGCG TCCTTCAGT CGCCTCAT TGGGTCTCAGT TGAGTCACC TCCTTCGGTC TAAAAACGTT ATCCTCAGGCC TACTTCAGCG TCCTTCCGGTC TAAAAGAGTT ATCCTCAGGCC TAAAGATGCTT ATCCGGCCCA ATCATCACCCA GAATAAAGGGC	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAAC TGCCCCTGGA AGCTCACCT GTAATACCGT GTAATATCGT TACAGAGTTC GGTACTATC TTACAGAGTTC TTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTTCGGTCGATCGT CTTCGGTCGATCGT CACAGCAGAAC CCAGCCTATT CGTCGATCGT CTCCGATCGT CATAGCAGAA CTCGTGCACC GACACGGAAA CTCGTGCACC GACACGGAAA	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGAA TTCTACGGGA ATTCTACGGGA ATTCTACGGGA ATTCTACGGGA ATTGTTCAC CGAATTGTTCC CTATGGCTTC CTATGGGTTC CTTGAGAGTC TTCTAGAGTC CTTTAGAGTC CTTTAGAGTC TTCTAGAGTC TTTCACTC TTTCAGAGT TTCTTCACAGGG GGTGAGTACT CTTTAAAAGT CAACTGATCT TGTTGAAATAC CATGATCAT	CCTATCCCTA CTAAGATCCA CCTATCCCTA CTAAGATCCA CCTTTATTTG GTGAAATTGA GTGAAATTGA GTGAAATTGA GTGAAATTGA GCCTTGCGCT ACCAGGGGAT CCCCCTGACG GCCCTGACG AACCAGCTC CAGTTCGGTG AACCAGCTC CTGACGCTC CTGACGTCA ACAAGCACC CTCAGACTT GGGAAGCTAG ATCAGCTCC AAGTTGGCCG AAGTTGGCCG CAGCTAGCAGTC CTCAGATTT CGCAACTAC CTCACATCT CTCACACTT CCACACTT CCACACTT TCACACTCTT TCACATCTT TCACACTCTT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGCACCAC CAGCAGCAT TACACACAC TATCAGCAAC TATCAGCAAC CAGTATACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACTCCAAGGT ATCGCCAGTTA AAACTCAGGT ATACAGGCA ATGGTATATT AAACTCAGGT AAACTCAGT AAACTCAGT AAACTCAGGT ATCGCCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CTGCTAGGT CTGGGGGCG CAGCGTTTCT TAGTGTATGGC CAGCGTTTCT TAGTGTATGGC CAGCGTTTCT TATTATTGAA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTTA GCACTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT GCCGGAAGGG ACCCAGCAAAAACTCTCA GGCGACCGAA AAAACTCTCA GGGCGACCGAG AAAACTCTCA GGGTAGCACA GCATTTATCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAAGG GGTCGGCATT CACAACACCA CTGGCTAACAG CTGGCTCTC CAGCACATTA CAGCAGATTA TGGTCATGACAG CAGTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTTCTTCCTCTTCC AGGACACTTTCT CAGCAGATCTTAC AGAACAGGAAG GGGTTATTGT
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACA 2601 TAAAGTGTAA AGCCTGGGGG 2701 GCATTAATGA ATCAGCCCA 2801 CCGGCAGCAGG TATCAACTC 2901 CCAGGAACCG TAAAAAGGC 2901 CCCGACCAGGA CTATAAAAA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3401 CATTAAGCAA AGCACTCT 3401 CTGAAAAAGAGATC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGCCAC 3801 TACCATCTG 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGATTCG 4101 TGTGCAAAAA AGCGGTTAG 4101 CGCGCGTCAA TACCGGATCA 4401 CGCGCGTCAA TACCGGATCA 4401 CGCGCGTCAA TACCGGATCA 4401 CGCTGTTGAG ATCCAGTTC 4501 GCAAAATGCC GCAAAAAAAG	AAAACTTTA AGTCTAGAGG AAACCCGCTG AGTCTAGTG GCGTAATCA GCCTAATGAG GCGGGGAG ACCGGGGAG ACCGGGGAG ACCGGGGTCTCTCTCT GCTTTCTCAT GCCTTATCT GCTTTCTCAT GCCTTATCG AAAAAGAGTT CAAGAAGATC ACTAGATCCT TATCTCAGCG CCTCAT GCTTCCAGT CCCTCCAT CGCCTCCAT CGGCTCCAT CGGCTCCAT CGGTGCCAGAGATCCT TACCTGGGC TCCTTCGGTC TAAGATGCT ATACAGAGCT TAACAGAGCT TAACAGGCCA ATGTAACCCA ATGTAACCCA GAATAAAGGGC TGAATAATTATT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTATAGC TGACTATAGC TGACTATAGC TGACTATAGC GTAATACGGT AGCCGGTTTT GCACTAGA AGCTCACCC GTAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT CGCAGAGACCC CCAGTCTTT CTCGTCGTTTG CTCGGATGGT CTCGTTTG CTCGATCGTTTC CATAGCAGA CTCGTGCACC CACACCGAAA CTCGTGCACC CACACCGACAC CACACCGAAA ATGAAAAATA	CAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TTGAAGTCC TAGACTCACCG AAATGAAGTT TTCGTTCATC ACGCTCACCG GTAGGTATCT TCGTCAGAAGT TCTTCATCAGC GTATGGCTT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CGTCACCG TATTAGATAC TGTTGAATAC CACAGATAGG CAACTGATCT TGTTGAATAC AACAAATAGG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACCCCGTTAA ACCACACTA ACCACACTA ACAAACCAC TCTGACGTC TTAAATCAAT ACTAGCGTC GGGAATCA ACTAGTTGCC CACTTGCGCC CACTACTA CATAGTTGCC CCCACATT CCCACATT CCCACATT CCCACATT CCCACATT CCCCCCACATT CCCCCCC CACCTCACT CCCCCCCC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCAT CAGCAGCAT CAGCAGCAT CAGCAGCAT TAGCACCAC CAGCAGCAT TACACACAC TATCAGCAAC TATCAGCAAC CAGTATACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CAGTTACC CCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC CCCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACTCCAAGGT ATCGCCAGTTA AAACTCAGGT ATACAGGCA ATGGTATATT AAACTCAGGT AAACTCAGT AAACTCAGT AAACTCAGGT ATCGCCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CTGCTAGGT CTGGGGGCG CAGCGTTTCT TAGTGTATGGC CAGCGTTTCT TAGTGTATGGC CAGCGTTTCT TATTATTGAA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTTA GCACTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT GCCGGAAGGG ACCCAGCAAAAACTCTCA GGCGACCGAA AAAACTCTCA GGGCGACCGAG AAAACTCTCA GGGTAGCACA GCATTTATCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAAGG GGTCGGCATT CACAACACCA CTGGCTAACAG CTGGCTCTC CAGCACATTA CAGCAGATTA TGGTCATGACAG CAGTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACCAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTACAG CAGCTTACTTCTTCCTCTTCC AGGACACTTTCT CAGCAGATCTTAC AGAACAGGAAG GGGTTATTGT

## pMT-HisFlag-DmSNAP43 mut#5 STOP

1 TEGESCETT COGREGARIC GOTGARGABACC TOTACACAT CACACTT CAGACTESCT TAGACGGAT GOCGAGGAC ACCTATIONS TAGACGGAT COGRAGAGACA TOTACACTAT CAGACTAGA TOTACACTAGA TAGACTAGA TAGACTAGAC	1	maaaaaaammm	aaama, ma, a	aamaaa	mama1 a1 a1 m	aa3 aamaaaa	a. a. aaama.	a. a.ammamam	am aaaaa. m	~~~~~~	a. a
201 COCCACAGAT GOSTAAGGAG AAATACOCC ATCAGGCCC ATCAGCATT CAGGCTGCC AAGGCCTC GETGCGGCC TCTCCCTAT 301 TACACCAGCT GOSTAAGGAG GAATTGCTC CAGGCCGATT AAGCCCAGGT TTCTCCCAGT CAGGCCCATT CATACAGGAG 401 TACACCACCAC CGCCACCCCCATC CAGGCCAGT TAGATGGGGAT AGCCCAGGT TTCTCCCAGT CAGGCCCATT CATACAGGAG CAGGCCCA 501 GACCCCCCA CGCCCCCCACCCCCCCCCCCCCCCCCCC											
301 TACGECCACT GGGGAAGGG GGGATGGGTC CAAGGCGAT AAGTGGGTA ACCCAGGGT TITCCCATC ACACGTTCT AAAACGGG CCAFGTCCC 501 GGCCCCCCC GGCCACCCCCC ACAGGCTTCT CATTAGTGGG CCGCCCACCCCC TITCCCCCCC GGCCCCCCCC CAGGGCTCCCCC CAAGGCGCCCCC CAAGGCCCCCCCCCC											
401 TGAATTAATT GSTTGCAGGA CAGGATGGG TGCCCGATTG TACTCTCTCAC CGTTGCCCCC CAGATAGGAC CAGAGACCATA 501 GACGCCCCCCC CAGCACCACC CACACACCATA CATATATGGATA AGGATGATCA GATGATCACCA 501 CAAGTCCCCCA AGTGGAGAA CCGGACCACAT TCTTCCAGATAT TCTTCCACATA 501 CAAGTCCCCCA AGTGGAGAA AGCCGGTGT AAAGCCCGCCT TTCCCAAATTG AGGATGATCAC 501 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGATCTCCCAAATTGA GAGCACATCTC 501 GACGATGACAA AGGGCACCATA TACTTTCACCA GATGATCGGG TACTAAGATT CAGATATCACATACCGA CTGCCACACA 501 AAATCAAGACA AGGGCACCATA GAGATGATAA AGGGGGACCACCA CACTTCACACA CACTATCACGA CATTCACACATA 501 GACGATGACAA AGGGCACCATA GAGATGATAA TACTTTCACCA GACCCTTCTC 501 GACGATGACACA AGGGCACCACA GACCACCACA CAGAGCACCA CAGAGCACCAC AGAGCACCAC AGGGGACTAT 501 GCCAGGACCTT CTCCCAGATCAC CACACACCAC CGGGACCATT 501 GCCAGGACCAC CACACACTTT ATTAAGATCA GAGCTCACCACA CGGACACACAC AGAGCTGATCACACACAC AGGCACCACAC TGGCGGACCAT 501 ATCTACACAA AGCCAGCCCAC GACCACACTT TATTAAGATCA GAGCTCACACAC AGGCACCACAC AGAGCACACAC AGAGCACACAC AGAGCACACAC AGAGCACACAC AGACCACACAC AGACCACACAC AGACCACACAC AGACCACACACA											
501 GGCCCCCCC GGCCCACCGC CACCACTA CATTATGGGG ACGCCCATA 601 CAAGTCCCCC AGGGGCTCC CACCACATA CATTATGGGG CAGAGCAAN A COTTOCCCA ACGCTCCCCA CACCACTTC GCCAAATGG ACCACCACCACACCA											
601 CARGTCCCCA ARGTGGGAA CACGAT TCTCCGGGG CAGACCAMA CCTTCCCAC TCCAANTITGA ARCCGCCGG GGTGCCAAA 801 AAACCAAGTG AATCATCTCA GTGCACCTAA AGGGGGGATC TAGAACCAGTG CAGACACTTCG CAGACACTTCTGC 801 AAACCAAGTG AATCATCTCA GTGCACCTAA AGGGGGGATC TAGATCGGGG TACTAAGGAT CAGACACTCTC ACCATACCGA TCACAGCGA 801 AAACCAAGTG AATCATCTCA GTGCACCTAA AGGGGGGATC TAGATCGGGG TACTAAGGAT CAGATCACCAC ACCATACCGA CACATCCCAG CACACCACAC											
701   AGAGGTGAATG CGAACGAAGA ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCC AGAGGTCTG GCCAATGCG ATCACGACTA CACATACCAG TACAAGGTA SID ACCCAAAGGTA TAGAAGGTGA CGAATTCAGAGGTA TACAAGGTGAT GAAGGTGATG AGAGGTGATG GCCAATGCGC ATCACAGGTATA TACAAGGTGAT GAAGGTGATA AGAGGTGATA AGAGGTGATA CACATACCAG CATCAAGGTATA TACAAGGTGAT AGAGGTGATA CGAATTCAGAGGTATA TACAAGGTGATA CGAATTCAGAGGTATA TACAAGGTGATA CGAATTCAGAGGTGATA CACACTACGAC TACAAGGTGATA CGACCACACAC TGGGGGCCTT TITL ACTACACTACAGGATATCACACACACACACACACACACAC											
801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGG TACTACCATA CAGTTTCCAG CTACCCAG CTACCCAGA CGATTCGGCGATT CTGCGCGAGAC TGGCGGACTTT CTGCGCGAGA CGATTCGGCGAA CGATTCGGCGAA CGATTCGGCGATT CTGCGCGACTTT CTGCCGCAGA CGATCGGCGAC CGAGGGGGACTT CTGCGCCAGA CGATCGGCGAC CGAGGGGGATT CTGCGGCGACCT CTGCGCCAGA CGAGGTGGGGA TTTCTTTCTTCT CTCTCAGTTA 1201 ACTACTACA AGCAGCCAC AGCAGACTT ATTAGAGGGA AACCACCACC CGGGGGGGTGT CGGGGCCCCC CGGGAGGGGGAGACT CATCAGAGTTG CGGGACCACCC CGGGGGGGTTT CTCCGGCGCCT TCGCCGCCG CTCTCAGATCTG CGCAGGGGTT CTGCGCGCGCCCC CGGGAGGGGGAGACACCACCACGGGGGGGG	601	CAAGTCCCCA	AAGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
901 GACGATGACA AGGGCACTAG TGAGCTGAAT ATCTITGACG ACTGCTGGG CGCTGGTGCACA CGAGTGCACA CGAGGGCACTA AGGGGGCCT 1101 GCATGTGGCA AGGGACTGT GGTGCTCCCG AGGCACCACC GGGACCTTT CTCGGCGCCCA CAGGGGGGGGGTT 1101 GCATGTGGCA AGGGACTGT GGTGCTCCCG AGGCACCACC GGGACCTTT ATTAGGGGGGGCT 1101 GCATGTGGCACCAC AGGCACCTT ATTAGGGAGGTG AGGGCGCCCC CGCACACCACCACCACCACCACCACCACCACCACCACCAC	701	AGAGGTGAAT	CGAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
1010 CGAGGGGGT TGGGGGAGAAC TGGGGGGCCCC GGAGACACTE ACTGCCCAGA GGAGCACACAC AGAGGTGATA GCCACCACAC TGGGGGCCCT T1010 GCATGGGGC AGAGGACTAT AGAGGGGGGGGGGGAGACTATGCCCACAC GGAGGACTAT ATAAGATCG AGGGGCGCT GCCGGGGTT TCCGGGTCAA AGCAGCCCAC GCACACTT ATAAGATCG AGGGGCGCT GCCGGGGGTT CACCAGGGGC AGAGGACACAC AGAGGACACAC CACACACTT ATAAGACTG AGGGGGGGGGG	801	AAATCAAGTG	AATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
1101 GCATGTGGCC AGCCACTGT CATGCACCAC GGGACACTG GGGAGGTTT TCCCGGCACT AGCCACCACTA AGGATCGGAG GTTCTTCTT GCAGGAGGAGACTACTA ATTACACCACTAC CTCACACTTGC CAGGAGGCACTGGC CTCAGACTGG CCCAGGGGTT AGGACTACTA CAGACTACCC CTCAGACTGG CAGGGGTT CGCGAGGGATA AGGACTACCC CTCAGACTGG CAGGGGTT CGCGAGGGATA CGCGAGGGATA CGCGAGGATA CGCGAGGAGGACA CGCGAGGAGGAGAGAGAGAGAGAG	901	GACGATGACA	AGGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
1201 ACTECTRICA AGGAGCCCAG GGCAGACTT ATTAGANCS AGGECTACC GGCAGACTTGG CAAGAACTTAG CAAGACTTGG CAGGAGTATC  N L W R L T Q E Q A F R F T A L D Y C Q G L  1401 GGACATCTG GTGGACTACG ACCGTGTGGA GACGCGTGAGC CAGGAGCAGC AGGGCTGCCTG ATCACGGGG CTGACACTGTGC  N D N L V D Y D R V E T V A G A K E Q R Q S S A L M Q K Q Q R A N G V  1501 AGGTCTACACT AGGACTGGA GACGGTGTGGAG GAGGGGGCGG GAGGGGGGGAG GAGGGGGGGAG GAGGGGGTGAGCAGGTGTGAGCAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1001	TCGAGGTGTT	CTGCCGGTGC	TGGCGAGAAC	TGCAGCTGCA	GCACCTTTTC	ACTGCCCAGA	CGAACCACAC	AGAGGTGATA	GCCACCACAC	TGGCGGCCCT
1301 GTCCGGAGGG GAAGGACAC CATCAGATCG CCTACATGCT GTGGCCCTG ACCCAGGAGG AGGCCTTCG CTTCACCGCG CTCGACTATT GCCAGGGGT $\times$ M L $\times$ L T Q E Q A R F F T A L D V C Q G C 1501 AGGCCAGGGG AGGCAATCG GAGCCAGGGG AGACGAGCGC GAGCGCCTG ATCCAGGAGC AACAGGGTC GAAGGGCTC $\times$ D N L V D Y D R V E T V A G A K E Q R Q S L M Q Q Q R A N G Q L 1501 AGGCCAGGAGCA ACCAGCGGC GCGCAGCGGC GCGCAGCGC TACCGGAGGCC TACCGGCGC TTACCGCCCC TCAAATATT CGGCCATTG CGGAGGGCT TGGCAGAAG AACTGCCA CAACTCTAC AGGCACCCA TTGGAAGCGC ATCGAGGGG GCGGAGCC $\times$ P D E K C T T S T R N Q L E V R Q R V R N K A M Y G V E R R E P T T S T R N Q L E V R Q R V R N K A M Y G V E R R E P T T S T R N Q L E V N R T Y Q R R N S A T V F Q R R L P E T Y Q R R N S A T V F Q R R L P E T Y Q R R N S A T V F Q R R L P E T Y D T S C N R C T T S T S N D R T Y Q R R N S A T V F Q R R L P E T Y D T S T T S D R N G T Y C R R T T S T T S T T S T T S T T T T T T	1101	GCATGTGGCC	AAGCGACTGT	CGTGCTCCCG	ACGCACCACC	GGGGACGTTT	TCCCGGCATC	TCGCGCTCAA	AGGATCGGAG	GTTTCTTTCT	GCTCTACGTA
M	1201	ATCTACTACA	AGCAGCCCAC	GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
1401 GGACAATOTG GTGGACTACG ACCGTOTGGA GACCGTACCG GTTCCCAAGG AACAGAGGC AGAGGCCTTG ATCCAGAGGC AACAGGGTC V V N G N K V O Y D Y D Y D Y D Y D Y D Y D Y D Y D Y	1301	GTCCGGAGCG	GAAGGACACT	CATCAGATCG	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
1401 GGACANTCTG GTGGACTACC ACCETOTIGGA GACCETACCG GTTCCCAAGG AACAGAGGCA GAGTGCCTTG ATGCAGAAGC AACAGGTCC COUNTY											
S D N L V D Y D R V E T V A G A K E Q R Q S A L M Q K Q Q R A N G V S S L T Y E L E G L R A A A R A R A R A R A R A R A R A R	1401	GGACAATCTG	GTGGACTACG	ACCGTGTGGA	GACCGTAGCG	GGTGCCAAGG			ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
1501 AGTOTCACAT AGGACTGGG GGLCAGCGGG GCCAGCGGG GCCATTGT GAACTGGAAG GGCATACAA GCCCAAAAG AGCAATTGG S											
S L T Y E L E G L R A A A A A P L C E L E A A Y N A O K K Q L 1610 CGGCTGGTCA TGAGGACGCT TTACCGCCTT CTAAATATT CGGCCATT CGGGATATC CGGAAATGTT TGGAGACTA GAAAAGATTAC A A G H E H A L P P S Q I F G H L C GCGAAATGT TTACCGCATT CTAGGAGCT GAAAAGATTAC CAAAAGATTAC CACAAAGATTAC CGGCAAATGTT TGGAGCCCA CACACATCTAC AGCACACAG TTGGAAGTT GCGAGAGGGGGGGGGG											
1610 CGGCTGGTCA TGAGCAGGCT TTACCGCCCT CTCAAATTATT CGGCCATTTG CGAAAGTCT TTGCCGATAT CCAAAGTGT TAGCGCCTA CAAAGTCAC  A A G H E H A L P P S Q I F G H L R E V F A D I Q S V L G A R K S T  1701 TCCAGATGGA AAATGCACC CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGGA GCGGGAGCCG  P D E K C T T T S T G N Q L E V R Q R V R N K A M Y G V E E R P P  1801 CAACACCAGA CGAGTGAACT AGAAGTGCAG CTGGAGGTCA ACCAGAGCTTA TCAACGCCCC ATGTCCTCGG CCACCGTTTT CCCAGAGGGAA CTCCAGAGAACAACAACAACAACAACAACAACAACAACAACA											
**N A G H E H A L P P S Q I F G H L R E V F A D I Q S V L G A R K S T 1701 TCCAGATGAG AAATGCACCA CACACATCACA AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAG GCCATGTACG GCGTCAGAGG GCGGGACCAG P D E K C T T T S T G N Q L E V R Q R V R N K A M Y G V E E R E P 1801 CAACACCAGA CGGATGACAC AGACATCACA ACACACACACACACACACACACACACACACA											~
1701 TCCAGATGGG AAATGCACCA CACATCTAC AGGCAACCAG TTGGAAGTGG GCGCAGAGGGG GCGGAACCAG S P D D E K C T T T S T G N Q L E V R Q R V R N K A M Y G V E E R E P 1801 CAACCACCAGA GGGATGACCT AGGAGTGCAG CTGGAGGTGC ACGAGGTTA TCAACGCCCG ATGTCACCGG CCACCGTTTT CCAGAGGAG CTCCAGAGAG P 1901 ACGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG											
P D E K C T T T S T G N Q L E V R Q R V R N K A M Y G V E E R E P 1801 CARCACCAGA COGATGACCA CAGAGGACTA ACAGACCAGA ACGACCAGA TOTCACCAGAG S V D R D E L E V Q L E V N R T Y Q R R M S S A T V F Q R E L F E P 1901 ACGTGCAGA ACAGACCAGA ACGACCAGAGACTA ACAGACCAGA ACGACCAGAGACTAGA ACAGACAGAGA ACGACCAGA ACCAGAGAGACTAGA ACAGACAGAGAGAGACAGAGAGAGACAGAGAGAGAGAGACAGAGAGAGAGAGACAG					~				~		
1801 CAACACCAGA CGGATGAACT AGAAGTCAG CTGAAGGTCA ACAGACTTA TCAACGCCGC ATTOTCCTCGG CACCGTTTT CCAGAGGGAA CTTCCAGAAG											
> Q H Q T D E L E V Q L E V N E T Y Q R R M S S A T V F Q R E L F E 1901 ACGTGCAGCA AGAGTAGGA ATGASTAGGA TTAGGAGGA CAGAGGAAGT GAGGGAGGA AGAGGAGGA GAGAGCACCA AAGCTATTT > D V Q Q E Y E M I E F S D D E E M E V G E S E E V T E E E L K A I L L 2001 GGATACTTGA GTAGATTTA AAACCTATTA CATAATAAA TAACTAGCAT TTTTGCGCGA TGTGATCTG TTTATCTGAA GGGCAATTCT GCAGATATCC > D T 2101 AGCACAGTGG GCGCCCCCG AGTCTGAGGG GCCCCGCGGTT CGAGGGAGA CCTATCCCTA ACCCTTCCT CGGTCTCCAT TCTAGGCCTA CCACACAA 2201 TCACCATCAC CATTGAGGTT AAACCGCTG ATCACCACCA ATCACCACCAC ACCCTTCCT CGGTCTCCAT CGGTCAATAAAA TGATTGATTTT AAACCGCTG ATCACCACA ACCCACACA 2201 TCACCATCAC CATTGAGGTT AAACCACCACA ATCACCACCAC ACCCTTCCT AGGCTCCA ACCCACCAC ACCCACCAC ACCCACCAC ACCCACCA											
1901 AGGTGCACA AGAGTATGAG ATCATTGAT TRATGACGA CGAGGAAATG GAĀGTGGGT AAAGCGAGA GAĀGACTCA AAGCTATTT   D V Q Q E Y E M I E F D D E E M E V G E S E V T E E L K A I L   2011 GGATACTTGA GTTAGTTTAT AAAACTTTA CATAATTAAA TAACTAGCAT TITTGCGGA TGTGATCTG TITATCTGAA GGGCAATTCT GCAGAATACC   D T   1010 AGCCAGTGG CGGCCCCCG AGTCTGAGGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCCTC TGGTCTCAT TCTACGCGTA CCGGTCACCA   2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCTT CTAAGATCCA GACATTATA GATACATTGA TGAGTTTGGA CAAACCACA   2301 CTAGAATGCA GTGAAAAAAAA TGCTTTATTT GTGAAATTT TGAGAACTAT TAACCATTAT AAGCTGCAAT AAACCACAA   2301 CTAGAATGCA GTGAAAAAAAA TGCTTTATTT GTGAAATTT TGAGAACTAT TAACCATTAT AAGCTGCAAT AAACCACAA   2301 CTAGAATGCA GTGAAAAAAAA TGCTTTATTT GTGAAATTT TGTGAAATTT TAACCATTATT AGCTGCAAA AAACCACAA   2401 TTGCATCAT TTTATGTTTC AGGTTCAAGG GGAGGTGGTGG GAAGGTTATTT TAACCATTAT AAGCTGCAAT AAACCACTAC AAATCGACTA AAACCACAA   2501 ACCTGCAGGC ATCCAAGCT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TAACCATTAA AAACCTCAA CAACATACGA GCCGGAAGCA   2601 TAAAAGTGTAA AGCCTGAGGT GCTAAATGA TGAGCTAAC TGTTCCTGT GTGAAATTGT TACCGCCCA CAATTCCACA CAACATACGA GCCGGAGCA   2601 CCGGCAAGGG TATCAGCTCA CTCAAAGGCG GTAATTACGTT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGG GGAAACCGG CAGCAAAAGG   2701 GCATTAAAACAT ACCAGGCGTT CGTCTGG CGTTTTTCCA TAGGCTCCAC TGCGCCCTCC TCCGCCTCC TCCGCCTC TCCGCCCTTT   2801 CCCGACAGGA CTATAAACAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTC TGGCTCCC TTCCGCCTC TCCGCCTTT   2801 CCCGTACAGG GAAGCGTGG GCTTTATCCA TAGCCCACCA GACCCCCTGAC ACCACAA AAATCGACCC TCAAGCACCA GCCCTCCCC GTACCTCACT AGACCACA CAACACACAC CAACACACAC CAACCACAC CAACCACAC CAACCACAC CAACCACACA AAACCACC											
DI VI Q Q E Y E MI E F S D D E E M E V G E S E E V T E E E L K A I L  2011 GGATACTTGA GTTAGTTTAT AAAACTTTA CATAATATAA TAACTAGCAT TTTTGCGGG TTGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC  > D T  2101 AGCACAGTGG GGGCGGCTG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTTATCCCTA ACCCTTCCTCT CGGTCTCGAT TCTACCGGTA CCGGTACTCA  2201 TCACCATCAC CATTGAGTTT AAACCCCTG ATCACCCTG ACTGTGCCTT CTAACAGACCA ACCCTTCCTC CGGTCTCGAT TCTACCGGTA ACCACACA  2301 CTAGAATGCA GTGAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTAT GCTTATTTG TAACCACTAA GATCACTAA TAGGTTTTGGA CAAACCACA  2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTTTGG GAGGTTTTTT AAACCCACTAA AAACCTCTAC CAAATTCACCAC CAACACACA  2501 AACTGCAGGC AGCAGAGCT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAAATTCCACAC CAACATACACA  2501 AACTGCAGGC AGCGAGGAT GCCTAATGAC TGTCATGCT TTTCCTGT GTGAAATTGT TATCCGCTCA CAAATTCCACAC CAACATACCACA  2501 CGGCGAGCGG TATCAGCTCA CTCAAAGGG GTAATCAGC TATCCACCAG ATCCACCAC TTCCCCCCT CACTGCCCCC TTCCAGCCCC  2501 CGGCGAGCGG TATCAGCTCA CTCAAAGGG GTAATCAGGT TATCCACCAGA ATCCAGCAC ACTAACCACAC ACCGCAGGAA GAGCGGTTG CGTTCGGCTG  2501 CCGGCAAGGG TATCAGCTCA CTCAAAGGG GTATTCCAC TAGGCTCCCC CCCCCTGACC ACCTACCCCC TTCCACCCCC TCCAAGCCC CTCAAAGGC  2501 CCGGCAAGCG TAAAAAGGCC CTCAAAGGG GTATTCCAC TAGGCCCCC CCCCCTGAC ACCACCACAC AAATCCACAC CAACATACCACAC  2501 CCGCACAGGA CAACCGCCGC GCCTTCCCG GTACCACCC GTAGGCCCCCCCGGTCC GCCCCTGAACACC CAACACCCC  2501 CCGCACAGGA CAACCACCCCC GCCCCTCGG CCCTTCCCCC TCCCCCCTCCCC TCCCCCCTCCCCC TCCCCCC											
2001 GGATACTTGA GTTAGTTTAT AAACCTTTTA CATAATTAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC > D T											
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA 2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTC ACTGGTCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGA CAAACCACAA 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTAT GCTTTATTTG TAACCATTAA AACCACTAA AAACCACAA 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTT AAAGCAAGTA AAACCACTAA AAACCACAA 2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTT AAAGCAAGTA AAACCACTAC AAATTGGTAT TGCTGCATTA TGCTGCATTA TGCTGCATTA TGCTGCATTA TGCTGCATTA TGCTGCATTA TGCTGCATCA CAAATTCACA CAAACCACACAA 2401 TTGCATTCAT TTTATGTTTC AGGTTAATCA TGGTCATAGG TGTTTCCGTGT TAACCACTACA AAAACCACACAA 2401 TTGCATTCAT TTTATGTTTC AGGTTCATGC TGTTTCCTGTG TATTCCGGCTC CAATTCCACCA CAACCTACCAC CAACCTACCAC CGCGAGAGCA 2501 ACCTGCAGGG TGCCCACAC GCCCGGGGGA AGCGCGTTTC CGTTTCCTGTG GGTTTCCTGT TATCCGGCTC CAATTCCACCA AAACCACCACACACACCACCACCACCACCACCA											
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCT CGGTCTCGAT TCTACGGTA CCGGTCACA 2201 TCACCATCAC CATTGAGTTT AACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAGA ATACATTGA TGAGTTTGAG CAAACCACA 2401 TGAGATTCAC GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGAGATTTGA TAACCATTATA AACCTGCACAT AAACAACACAA 2401 TTGCATTCAT TTTATGTTTC AGGTCAGGG GGAGGTGTGG GAGGTTTTT AAAGCAAGTA AAACCCCTAC CAATTCCACA CAACACACAA 2401 TTGCATTCAT TTTATGTTTC AGGTCAGGG GGAGGTAGTG GAGGTTTTT AAACCAAGTA AAACCTCTAC AAATGTGGTA TGGCCTAATCA 2501 TAAAGTGTAA AGCCTGGGGT GCCTAATCACA TGGTCATACC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACACACACA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATCACA TGGTCATACCA TGGTCATACCA CAACATACGA CCCGGAAGCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAC TGGTCATACCA CAACATACAT GCGTCACCTC 2701 GCATTAAAGA ATCGGCCAAC GCGCGGGGA AGCGGTTTT CGCTTATGGC GTTCCCGCT CATCCACC CAACCCACC CAACCACCAC 2801 CCGGCAGGG TATCAGCTCA CTCAAAGGC GTAATACCGT TATCCACAGA ATCCACCG TTCCTAGTCC GTGCCTCGCT 2801 CCAGGAACCG TAAAAAGGC GCGTTGTTGC GTTATCCACAGA ATCCACAGA AAACCACACA AAATCCACC 2801 CCCGCAAGGA CTATAAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCCCCTGACG AGCATACCAA AAATCCACC 2801 CCCTTCAGG GAAGCGTGGC GCTTCTCAT AGCTCACGCT GTAGGTACTC TGGGCTCCC TGTCCGGCTC CTCCCCGCGTTA CCGGATACCT 3101 CCCTTCAGG GAAGCAGGGG GCTTCTCAT AGCTCACGCT GTAGGTATCT CAGGTTCGGT TGCGCCTTC 3201 CCGTTCAGCC CGACCCCTGC GCCTTATCCG GTAACTATCG TCTTGAGTC ACCCCGGTAA GACACACAC AGCACACCACA 3301 GATTAGCAGA GCGAGGTATC TAGGCGGTG TACAGAGTT TTGAGTCC ACCCCGGTAA GACACACAC TACCCACGT GCAGCACCC CTGGTAACCA 3301 GATTAGCAGA AAAAGGATC CAACACACCC GCCCCCCCCGCGTAACCT AGGGCAGACA TATTTGGTAT CTGCGCCTCT 3401 CTGCACCTCG GAACAGATTA TAGGCGGTG TACACATACACCC GAACACACCC GCGGCAGCCA CTGGTAACAC TACACCCAGAATAA 3501 GCGCCAGAAA AAAAGGATC TACACTCAG AAAAAGACT CTTGAACCAC GCGGTAACAC AGAACACACC GCGGCAGCCA CTGCGAACCC 3301 CACGACACAC TACCTCACC AAAAAGACT CCAACACACCC GCTGGTAACCT TAGAGGCACA TACCTCAGCA TACCACCAGATTA AAACACACC GCGGCAGCAC TACCTCACCG GAACACCACC TACCCCCCCCCC			GITAGTTAT	AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCA GACATGATAA GATACATTGA TGAGTTTGA CAAACCACA 2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTTT GTGAAAATTG TGATGCTATT GCTTTATTTT TAAGCCATTAT AAGCCAGCAT AAACCAGCTTA AAACCAGTTTA CAGCTGCAAT AAACCAGTTTA CAGCTGCAAT AAACCAGTTTA CAGCTGCAGT TATTGCAGTCTA TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGGTTTA TAAGCCAGCA AAACCCACCA CAACATACCA CAACATACCA CACCTAGCA ATGCCAGCA ATGCCAGCA ATGCCAGCA ATGCCAGCA CACCTAGCAGCA AGGGTTTATA AAACCAGTTA TAACCAGTCA CAATTCCACA CAACATACCA CACCTACCAC CACCTAGCA ATGCCAGCA AGCCTACCAC CACCTACCAC ACCCTACCAC CACCTACCAC ACCCTACCAC CACCTACCAC CACCTACCAC ACCCTACCAC CACCTACCAC ACCCTACCAC ACCCTACCAC ACCCCACCACCAC CACCTACCAC ACCCCACCACCAC ACCCCACCACCACCACCACC											
2301 CTAGAATGCA GTGAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCAGCAT AAACAGCTA AAACAGCTA AAACAGACA 2401 TTGCATTCAT TTTATTTTTT CAGGTCAGGG GGAGGTGTGG GAGGTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGCCTGATTA TGATCAGTCA 2501 ACCTGCAGG ATGCAAGCTT GGCTGATACA TGGTCATCATCA TATTCAGTCA CAATTCACA CAATTCACA CAATTCACA CAACATCACA 2601 TAAAGTGTAA AGCCTGGGGG GCCTAATCA TGGCTCATCA CAATTCACA CAATTCACA CAACATCACA 2601 TAAAGTGTAA AGCCTGGGGG GCCTAATCA TGGCTGATCA CACATTCACA CAATTCACA CAACATCACA CAACATCACA CACCTACATAATTA AGCCTGCAGC TTCCAGCCG TTCCAGCCG CTTCCAGCCG GGAAACCTG CGCGCAGCT CACTGCAGCG TACCAGCAGA AGCCAACAGC GCGCGAGCA AGCCAACAGCA ACAACACCACA AAATCGACC TCCAAAAGGC CACCAAAAGGC CACCAACACCA ACCACCACCA ACAATCACA AAATCGACC TCCAAAGGC GTAATACGGT TCCCCCTGGA ACCACCCACCA ACCACCACCA ACCACCACCA CCCCCC											
2401 TTGCATTCAT TTTATGTTTC AGGTCAGGG GGAGGTGTG GAGGTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG 2501 ACCTGCAGGC ATGCAAGCTT GGCCTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA 2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT GTATCCACA CACTACCACC TTCCCAGTC GGAAACCTGT GGGCAGCTC 2701 GCATTAATGA ATCGGCCAAC GCGCGGGAG AGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTGCTC ACTGACCGC TTCCAGTC GGGAAACGT 2801 CGGCGAGCGG TATCACCTC CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AAACCGAGGAA AGAACATGT ACCAAAAAGG 2901 CCGGCAAGGA CTATAAAGAT ACCAGGCGTT TCCCCTGGA AGCCTCCTC TGCCCCTGCAC ACTACAAA AAACCACCC TCAAAAGGC 2901 CCGCACAGGA CTATAAAAAATA ACCAGGCGTT TCCCCCTGGA ACCCCCTGACA ACCACACAA AAACCACCC TCAAACGC TCAAACGC 2901 CCCGACAGGA CTATAAAAAAA ACCAGGCGT TCCCCCTGGA ACCCCCTGACA ACCACACAA AAACCAACCC TCAAACGC TCAAGCAAAGG 2901 CCCGACAGGA CTATAAAAAATA ACCAGGCGTT TCCCCTGGA ACCCCCTGCAC ACCCCTGACC CTGCCGCTTT 3101 CTCCCTTCGG GAAGCTGC GCTTTCCAT AGCTCACGC TTGAGTATCT CAGTCCCTCT TGTTCCGAC CTCCCGCTTT CCGGATACCT ACCGCACCAC GCCCCCTGACCA ACCCCCTGACCA ACCCCGCTACCT ACCGCACCAC GCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCTGACCA ACCCCCCTGACCA ACCACCACCACCACCACCACCACCACCACCACCAC											
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATCGA GCCGGAAGCA 2601 TAAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACTA GCCTTGCGCT CACTGCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT 2701 GCATTAATGA ATCGCGCAAA AACCGCGCTGT GCGCAGCT CGTATTGGGC TGCCTGCTGC TTCCCGCTC ACTGCCTGC TTTCCAGCT CGGAAACCTGT CGTCCAGCT 2701 GCATTAATGA ATCGGCCAAA ACCGCAGGAA ACGCGGGAAAAGGC GCGTTGCGCTG CGTTTCTCCA TATCCACAGA ATCAGGGGAA AGAACATGT AGCAAAAGGC CAGCAAAAGG CACCACAAA AAATCGACG TAAAAAAGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCCG CCCCTGACG AGCATCACAA AAATCGACG TCAACTCAGA GGTGCCAAA ACCCCCTGC GCCCCTGACG AGCATCACAA AAATCGACG TCAACTCAGA GGTGCCAAA ACCCCCCTGC GAACCCCCC CGCCCTGACG AGCATCACAA AAATCGACG TCAACTCAGA GGTGCCAAA ACCCCCCCGC CCCCCTGACG AGCATCACAA AAATCGACG TCAACTCAGA GGTGCCAAA ACCCCCCCCCC											
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTACTC CACATTAATT GCGTTGCGCT CACTGCCCG TTTCCAGTC GGAAACCTGT CGTGCCAGCT 2701 GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGCCGGTTTC CGTATTGGGC GCTCTCCGCT TTCCTCGCTC ACTGACTCGC TGCGCTCGGCTG CGTTCGCTC 2801 CGGCGAGCGG TATCACCTCA ATCCACAGA ATCACGAGAA ACACACAGA AGAACATGT ACCCACAGAAGAG ACACACAGA AGAACATGT ACCCACAGAAGAG CACACAGAGAC TAAAAAGAG CACACAGAGAC TAACACAGGAA ACACACAGAGA ACACACAGAGA ACACACAGAGA ACACACAGAGA ACACACAGAGA ACACACAGAGA ACACACAC											
2701 GCATTAATGA ATCGGCCAC GCGCGGGGA AGGCGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCT ACTGACTCG TGCGCTGGT CGTTCGCTG 2801 CGGCGACGGG TATCACCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGAT AAACGAGCAA AGAACATGTC ACCAAAAAGGC 2901 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTCCA TAGGCTCGC CCCCCTGACC ACCACACAA AAAACGAGCA CTAAAGGCC CAAGCACAAAAGG 3001 CCCGACAGGA CTATAAAAGT ACCAGGCGTT TCCCCCTGCA AGCCTCCCTCG TGCGCTCC TGTTCCGAC CTGCCGCTT CCGGATACCT GTCCGCCTTT 3101 CTCCCTTCGG GAAGCATGG GCTTTCCAT AGCTCACGCT GTAGGTATCT CAGTTCGGT TAGGTCGTC GTCCAACGT GGGCGAAA 301 CCTCATTCGG GAAGCATGG GCTTATCCG GTACCTCG GTACATACC AACCCGGTAA GACACGACT ATCGCCACTG GGGCGACA 3201 CCGTTCAGC CGACCGTGC GCCTTACCG GTACATATCG TCTGAAGTTC TTGAAGTCT ACCCCGGTAA GACACGACTT ATCGCCACTG GCACCACCC CTGGTAACGA 3301 GATTAGCAGA GCGAGGTATG TAGGCGGTC TACAGAGGTTC TTGAAGTGCT GGCCTAACTA CGGCTACCT AGAAGGACAC TATTTGGTAT CTGCGCTCTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGACT CTTTGATCTT TTCAAAGGGAA ACAAACCAC GCTGGTACCA AGAAGGACAC TATTTGGTAT CTGCGCTCTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGACT CTTTGATCTT TTCACAGGG TCTGACCAC ACAAACCAC GCTGGTACCA AAACCACCT TAGGGATTT TGTTTGAAG 3501 ATTATCAAAA AAAAGGATC CAGAAGAAC CTTTTAAATTAA AAATCAAAT CTAAAGAACAA AAACCCAC TTGGTCTGA CAGAAGACAC ACAACCAC GCTGGAACGA AAACCACC AAACCACC AAACCACC AAACCAC AAACCACC											
2801 CGGCGAGGG TATCAGCTCA CTCAAAGGC GTAATACGT TATCCACAGA ATCAGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGG CAGCAAAAGG 2901 CCAGGAAAGC TAAAAAAGGC GCGTTGTCG CGTTTTTCCA TAGGCTCCGC CCCCTGAGC AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCAAA 3001 CCCGCACAGGA CTATAAAAGAT ACCAGGCGTT TCCCCCTGGA AGCCTCCTCC TGGGCTCTCC TGTTCCGACC CTGCCGCCTTA CCGGATACCTA CAGCATCCTTA CCGCACAGGA CTCACACAA AAATCGACGC TCAACCTTA CCGCCACAGGA CTCACCACAGCT GCGCTCTCC TGTTCCGACC CTGCCGCCTTA CCGGATACCT GCGCCTTCT 3101 CTCCCTTCGG GAACCCCCCCC GCCTTATCCG GTAACTATCC TCTTGAGTCC AGCTCACGT TAGGCCACCT GCGCCCCCTG GCACCACCC GTAACACACCC GTAACACACCC CGACCCCC GCCCTTACCCG GTAACTATCG TCTTGAGTCC ACCCGGTAA GACACGACTT ATCGCCACCT GCACCACCC CTGGCAACACC 3201 CGTACACACC TACCCTCCG AAAAAGAGT GTAACTATCG TCTTGAGTCT TAGAGTGGT GCCTAACTA CGGCTCACACT AGAAGGACA TATTTGGTAT TGGCCCTCT 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGT GTTGACTCT TTAAAGTGAT CAGCACCC GCTGGTAGCA GAAACCACC GTGGTTTTT TGTTTGGAA 3501 CCCCCAGAAA AAAAGGATC CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCCC ACGGACCACA AAACTCACCT TAAGGGATTT TGGTCATGAG 3601 ATTATCAAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGG TCTGACCGC ACGCGAACAAAAACCACC GCGGAAGAAA AAAACCACCC GCGCAAGAAC AAACTCACGT TAAGGGATTT TGGTCATGAG 3701 TGCTTAATCA GTGAGGCACC TACCTCACCG ACTCTATT AAAACAAACACCC GCGCGAAAGAAAAACACACC GCGGAAGGCCAC AAACTCACCT TAAGAGAACACACC ACGTTACCAACCTATAACACTATAACACTATAACACACCACCA											
2901 CCAGGAACG TAAAAAGGC GCGTTGCTGG CGTTTTTCA TAGGCTCCG CCCCTGACG AGCATCACA AAATCGACG TCAAGTCAGA GGTGGCGAA 3001 CCCGACAGGA CTATAAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCT TGTCTCCGACC CTGCCGCTTA CCGCATCACT GTCCGCCTTT 3101 CTCCCTTCGG GAAGCGTGGC GCTTTCCTA AGCTCACGT GTGGGTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT 3201 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTACTC TAGGTGTTC TAGGTGCTT ATGGCGACT ATCGCCACT GCAGACCCC 3201 CAGTACACA GAAGCACGC GCCTTATCCG GTAACTACCG GTAACTACTC GAGTCACT AGAAGCACAC GCAACACCAC GCAGACCCC 3201 CAGTACACA TAACCTCCG AAAAAGAATT GGTAGCTCT TGAAGTCA ACAACCACC GCTGCTAACT AGAAGCACA TATTTGTTAT CTGCGCTCTC 3201 CAGCAGAAAA AAAAGGATC CAAGAAGAT CGTTGACTCT TTCAAGTGC ACCAGACCCA GTGGTAACAT AGAACCACC GTGGTACACT AGAAGCACA TATTTGTTAT CTGCGCTCTC 3201 ATTTCAAAA AGGATCTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAC CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTACACAA 3201 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCATCACCG GCTCCACCAC AGACCACAC AAACCACCC TAGGCAGATTA 3201 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCATCACCAC GCTCCACCT CAGACCCA CAGACCACACACACACACACACACACACAC											
3001 CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT 3101 CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACCAGACCC 3201 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG CTCTGAGTCC ACCCGGTAA GACAGGACTA TAGGCACTG GCACGACCAC 3301 GATTAGCAGA GCGAGGATAT TAGGCGGTC TACAGAGTTC TTGAAGTGT GGCCTAACTA AGAAGGACGA TATTTGGTAT CTGGTACAG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGACT GGTAGCTCTT GATCCGGCAA ACAAACCAC GCTGGTACACT AGAAGGACG TATTTGGTAT CTGCGCTCTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGACT CTTTGAATCT TTCACGGG TAGACGA ACAAACCAC GCTGGTTTTT TTTTTTGCAAG CAGCAGATTA 3501 ATTATCAAAA AGGATCTTCA CCTAGAGTCC TTTAAAATTAA AAATGAAGT TTAAAATCAAT CTAAAGGACAA AAACTCACGT TAGGGATTT TGGTCATGAG 3601 ATTATCAAAAA AGGATCTTCA CCTAGAGTCC TTTAAAATTAA AAATGAAGT TTAAAATCAAT CTAAAGGAATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATTGGTCCC TCGTGTAGAT AACCAGCCA GCCGGAAGGG CCGAGAGGCCA 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAAACCA CAGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTA ATAGAATAT AACCAAGCCA GCCGGAAGGG CCAGCCGAG 4001 TGTGCAAAAAA AGCGGTTACC TCCTCGGTC TCGTCGTTTG GTATGGCTC ATTCAGCAC AGTTCACACA GATCAAGGC AGTTACATA 4001 TGTGCAAAAA AGCGGTTACC TCCTCCGTC TCTCGTTTG GTATGAGAT AACTAGCACC AGTCTTACCAC AGCCCAAGTGC ATTCACAGA AACCAGCCA ACAACCAC CAGTCTTAT ATTCAGCTC ATTCAAGAAA AACCAGCCA ACAACCAC CAGCCAAAAACACCA ACAAGCC AGCCAAAAACACACCC AGCCAAAAACACACCC AGCCAAAAACACACAC	2801	CGGCGAGCGG	TATCAGCTCA	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAAGGC	CAGCAAAAGG
3101 CTCCCTTCGG GAAGCGTGG GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTG CACGAACCCC 3201 CCGTTCAGCC CGACCCCTG GCCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG 3301 GATTAGCAGA GCAGGATTG TAGGCGGTGC TACAGAGTTC TGAAGTGGT GGCCTAACTA CGGCTAACTA CAGCACCT GAAGCAGCT TATTGGTAC TGGCCTCTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTT TGTTTGGAAG 3501 ACTATACCAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTAGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG 3601 ATTATCAAAA AAGAGATCTC CCTAAGACCC TCTTCAACTAT TTCATCAGCG TCTAACCACCT TAAAGATAAA TATGACTAAA CTAGAAA AAACCACCC GCTGGTAGAC AAACTCACGT TAAGGGATTT TGGTCATGAG 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTCACC GCTCCAGACT TATCACAATA AAACCACCA GCCGGAAGGAC 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCACCA GCCGGAAGGGC CGAGGCCGA 3901 AAGTGGTCCT GCAACTTATA CGCGCTCCCAT CCAGTCATTA AATTGTTAGC GGGAAGCTAG AGTAGGAAGT TCCCCAGTTA TATCATTTGCG CAACGTTATT 4001 GCCATTGCTA CAGGCATCGT GGTGTCACCG TCGTCGTTT GTACAGAAGT AATTGGCCC CAGGTTCACCA GGTCCAGAAAA AGCGGGTTAA CTCCATCGGT CTCCCAGTCTTT AATTGGTCC TACAGCAGAT TATCAGCAAT CACCCAGG AGTTACATGA TCCCCCAGGT 4101 TGTTCACATCA CAGGCATCGT TCTCTGGTC CTCCGATCCT TGTCAGAAGT AACTGCCCC GGTTCCCAC GATCCAGGC AGTTCATCAGATT CACCCATCGT ATAGCAATCC TCCCCATCGT TGTCAGAAGT AACTGCCCC CAGGTTTATC ACTCCATCGT ATAGCAATCC TACCCATCGT TGTCAGAAGT AACTGCCCC CAGGTTATATC ACTCCATCGT AGCACATCG TACGATCTT TTCCTAGACT CAACCAAGTC ATTCTACTCT ATCCCATCGT AGCACATCG TACGATCTT TTCCTCAGCC CAGTGTTATC ACTCCATCGT AGCACATCG TACGATCTT TTCCTCTAGCC CAGCTACTT TACCATCGT CACCCAAGTTTC CACCCAAGTTCT AGCACACCA GCGCAAAAAACACACACAGCC AGCGAATAAACCAC CACCAAGATCT ATTCAACAAACCAC AGCGCAAAAACACACAAACCACAAGCACAAACACAAACCACAAACCACAAACCACAAACCACAAACCACA											
3201 CGETTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG 3301 GATTAGCAGA GCGAGGTATA TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACACT AGAAGGACAG TATTTGGTTAT CTGCGCTCTG 3401 CTGAAGCCAG TTACCTTCGG AAAAAGAAGTT GGTAGCTCTT GACCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTATCTG TGTTGTTGAAG CAGCAGATTA 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGACT CTTTGAACTTT TCTACGGGG TCTGACCCT AGTGGACGA AAACTCACGT TAAGGGATTT TGGTCATGA 3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAC CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCA 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCATT TCTTCTCATC CATAGTTGAC TGACCCCG TCGTGTAGAT AACTACAGTA CAGGAGGGCAC 3701 AACCACTGG CCCAGTGCT GCAACTTTAT CCGCCTCCAT CAGGACCAC CAGCCACCG GCCCACAGTT TATCAGCAAT AAACCACCA GCCGGAAGGA CAGCCACGAGACCA 3701 TGCTAAATAA ACAGGGCTCT CCGCTCCAT CAGTCTATT AATTGTTCCC GGGAAGCCTA GAGTAGAGAAAAAAAAAA	3001	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT
3301 GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TIGAAGTGGT GCCCTAACTA CGGCTACACT AGAAGGACA TATTTGGTAT CTGCGCTCTG 3401 CTGAAGCCAG TIACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTT TGTTTGCAAG CAGCAGATTA 3501 CGCCCAGAAA AAAAGGATCT CAAGAAGACC TITTAGACTT TCTACGGGG TCTGACCCTC AGTGGAACGA AAAACCAACCG TAGGGACCA ACAACCACCAC GCTGGTAGCA AAAACCAACCAG CAGCAAGTTA 3601 ATTATCAAAA AGGATCTTCA CCTAGAGTCCT TITAAAATAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 3701 TGCTTAAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGC TGACACCT CAGCGTAGATA AACCAGCCA GCCGGAAGGG CCGAGGGCT 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCAAGCTCA 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAC AGTAAGTAGT TCGCCAAGTT ATAGTTTTGC CAACCTTATT 4001 GCCATTGCTA CAGCGATCGT GCTTCTGGTC CTCCGATCGT TGTCAGAAGT AACTTGCCCC GGTTCCCAAC GATCAAGGCA AGTTACATCA TCCCCAAGTT 4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AACTTGCCCC GGTTCACAC GATCAAGGCA AGTAACTAC TCCCCAAGTT 4201 TCTTACTGTC ATGCCAACCCG TAGAATGCTT TTCGTGACT GGTAGAGAT AACTAGGAT ATCCATGGTT ATGCAGCACC GCGAACCAGAGT TGTCTCTGGTC 4301 CCGCCGTCAA TACGGGATAA TACCGCCCA CATAGCAGAA CATTTAAAAGT GCTCATCATT GGAAAACGT CTCTCGGGCC AAAACTCC AGCGAACTAC TGCCCCAACTATC 4401 CCGCTGTTAGA ACCAGTTCG ATGTAACACCA CTCGTGCACC CAACTGACT TAAAAGT CAACCAAGTC TTCAGCAACT TTCAGGGGC AAAACCCACAGT TTCTCTCTTCC 4501 CCACAGAGTC TAGCAATACC CATAGCAGAA CTTTAAAAGT GCTCATCATT TACATTTTCAA GCGTTCTCAACCAGAGT TAGCAACCAC CAACAGAGC AAAACCCACAGATTCT AACCAGATCT TAGCAACCCA CAACAGAGC AAAACCCACAGAGT TTCAGCAACCCA CAACAGAGC AAAACCCACAGAGT TTCACACCAGAGT TTCAGCAACCCA CAACAGAGC AAAACCAGAACCAAAGTCC TAGCAAATACC CAACAGAGC AAAACCAGCAAAATCC CAACAGAGC AAAACCGCAAAATACC CAAAAATCAC TCATACATTT TAGCAAACCCA CAACAGAGC AAAACCAGCAAAATCC CAACAAGAGC AAAACCGCAAAAACCCA CAACAGAGC AAAACCAGCAAAATCC CAAAAATCC CAACAAGAGC AAAACGCCA AAAACCAGAAATCC TAAACAACCACAAAATCC CAAAAATCC CAAAAATCC CAAAAATCC CAACAAGTCC ACCTTACATTTCAA CATCTATTACA CACCTACACTCC CAAAA	3101	CTCCCTTCGG	GAAGCGTGGC	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTC	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC
3401 CTGAAGCCAG TTACCTTCGG AAAAAGGTT GGTAGCTCTT GATCCGGCAA ACAAACCAC GCTGGTAGCG GTGGTTTTT TGTTTGCAAG CAGCAGATTA 3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG 3601 ATTATCAAAAA AGGACTTCTA CCTGAGATCCT TTTAAATTAA AAATGAAATT TTAAAATCAAT CTAAAGTATAA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 3701 TGCTTAATCA GTGAGGCAC TATCTCAGCG ATCTGTCTAT TCGTTCATC CATAGTTGCC TGAATCTATA TATCAGATAA AACCAGCCA GCCGGAAGGGC 3801 TACCATCTGG CCCCAGTGCT GCAATGTATA CGGCGCCCAC CGGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCCGA 3901 AAGTGGTCCT GCAACTTATA CGGCCTCCAT CCAGTCTATA AATTGTTGCC GGGAAGCTAG AGTAGGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT 4001 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTT GTACAGAAGT AATTGCAGCC GGTTCCCAAC GATCAAGGC AGTTACATGA 4001 TGTGCAAAAA AGCGGTTACC TCCTCTGGTC CTCCGATCGT TGTCAGAAGT AACTGGCCC CAGTGTTATA CATCATGGTT ATGCCAGCAC TGCATAATTC 4201 TCTTACCTGT ATGCCATCCG TAAGAAGCTT TCTGGACAG CTCACAAGTC AACCAAGTC ATCCATGGTT ATGCCAGCAC TGCATATTC 4201 TCTTACCTGT ATGCCATCCG TAAGAAGCTT TCTGGACAGA CTTTAAAAGT GCCCACAAGTC ATCCATGGTT ATGCCATCCA AAACCAGTCT 4301 CCGCCGTCAA TACGGGCTCA TACGCACCCC CATGCAGAAC CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGCC AAAACTCCA AGGAACCAAGTCT TCGCCCATCGT 4401 CCGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCAATCT TTACATATTCA CACCAGTTCC AAACCAGTCT TTAAAAGT GCCCACCGTTTCT GGGTGAGCCAA AAACCGCAAAAAAACGACCAAAAAAAAAA	3201	CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACTATCG	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG
3501 CGCGCAGAA AAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGG TCTGACCTC AGTGGACGA AAACTCACGT TAAGGGATTT TGGTCATGAG 3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCG TCGTGTAGAT AACTACGAT AGGGAGGGCT 3801 TACCATCTGG CCCCAGTGCT GCAAGTACC CGCGAGACCC AGCCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAGAGG CCGAGCGCG 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA AATTGTTGC 4001 GCCATTGCTA CAGGCATCCT GGTGCACCG CTCTCATT AATTGTTCC GGGAAGCTAG AGTACAAGGCA GATCAAGGCA ATAGTTTGCC 4001 GCCATTGCTA CAGGCATCCT GGTGCACCC TCCCGATCGT TGTCAGAAGT AATCTAGGCAC CAGTCTAAACACCA GATCAAGGCA ATCCATGGTT ATCCCCCATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTCCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCC CAGTGTTATA CACCAAGGC AGTCAAGGCA TGCCATCTC 4201 TCTTACTGTC ATGCCACCC TAAGATGCTT TCTCTGGACT TGTCAGAAGT AAGTTGGCCC CAGTGTTATC ACTCCATGGTT ATGCCAGCAA TAGGCCACCA TGCCACCAGAGC ATCCATGCTATC 4201 TCTTACTGTC ATGCCACCCA CATAGCAGAAC ATCTTGAACTCA TCTCGGGGC AAAACTCTCA AGGATCTTC 4201 TCTTACTGTC ATGCCACCAC CATAGCAGAA CTTTAAAAGT GCCCACCTATCATT GGAAAACCTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC 4401 CCGCGCTCAA TACCGGGCAA TACCGGCCA CATGGCAGAA TATGATGATCA TCTCTCAGACCT TTTCCTGGGGCG AAAACTCTCA AGGATCTTAC 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGACTT TCAGCATCCT TTACTTCACCC CAACTGAACAAATAC TCTAACTCT TCTTCTCAAAATCC CACCAAAATAC CTCATACTCT TCATACTCT TCTTTTCAA CCCTATCTT TGACAAACCAAGTCT TCAACAACCAAGCC ACTCTACATT TTACAAAACCA GCATTAACCCA CTCGTACATT TGAAAAATAA ACAAATAG GGTTCTCCCC GAAAATGCC ACCTGACCTC TAAGAAACCA 4601 CTCATGAGCC GAAAAAAGG GAATAAAGTATT TGAAAAAATAA AACAAATAG GGTTCTCCCC GAAAATGCC ACCTGACCTC TAAGAAACCA	3301	GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA	CGGCTACACT	AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA 3701 TGCTTAAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTTGC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT 4001 GCCATTGCTA CAGGCATCGG GGTGCACG TCGTCGTTTG GTATGGCTC ATTCAGCAC GATCACAGGC AGTTACATGA TCCCCAATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTCCGTC CTCCGATCGT TGTCAGAAGT AAGTTGCCCG CAGTGTTATC ACTCATGGTT ATGGCAACCA TCCATAATC 4201 TCTTACTGTC ATGCCATCCG TAAGAAGTGTT TTCTGTGACT GGTGAGAGTAC TATCAGGAA TAGTGTATC GGCGACCGAG TTGCTCTTGC 4301 CCGCGGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGT CTTCGGGGCG AAAACTCCA AGGATCTTCAGAAATGCACCACCACGAG TTGCTCTTGC 4401 CGCTGTTGAG ATCCACTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCAATCT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 CCAAAATGCC GCAAAAAAGG GAATAAGGG GACACGGAAA TATGTAACACA TATTATTGAA GCATTATTCA GGGTTATCTCACACAAGTCT TCAGCAAATTCC CCAAAATGCA GAATAAGGG GAATAAGGG GAATAAGGG GATTAAGGC GACTAAACCACACAGAATCC TCCTTTTTCAC CACCGGTTTCT GGGTGAGCAA AAACAGGAAG 4601 CTCATGAGCG GAAAAAGGG GAATAAGGG GATTAAGGG GATTAACGCAGAAATAC CAACTATCCCC GAAAAGTGCC ACCTGACCACT TAAGAAAACACAAATAA AACAAATAAG GGTTCCACCC GAAAAAGGC CACAAGAACCCACAGAATACCCA CACTGACCACCCC CAACAAGACCACACCAC	3401	CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT	TGTTTGCAAG	CAGCAGATTA
3701 TGCTTAATCA GTGAGGCAC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT 3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAGAGG CCGAGCGCGAGGAGG CAGGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	3501	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	TCTGACGCTC	AGTGGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
3801 TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG 3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT 4001 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAGAG AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC 4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC 4301 CCGGCGTCAA TACCGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAAACGT CTTCGGGGCG AAAACTCTCA AGGATCTTAC 4401 CCCTGTTGGA ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 GCAAAAATGC GCAAAAAAGG GAATAAAGGG GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATGT 4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCCC ACATTTCCC GAAAAGTGCC ACCTGACCT TAAGAAACCA	3601	ATTATCAAAA	AGGATCTTCA	CCTAGATCCT	TTTAAATTAA	AAATGAAGTT	TTAAATCAAT	CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTACCAA
3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGT 4001 GCCATTGCTA CAGGCATCAG GGTGCACGG TCGTCGTTTG GTATGGCTTC ATTCAGGTCC GGTTCCCAAC GATCAAGGCG AGTTAACATGA TCCCCCATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTCCGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATC 4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC 4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGGC AAAACTCTCA AGGATCTTAC 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCA CTCGTGCACC CAACTGATCT TCAGCAATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA 4501 GCAAAATGCC GCAAAAAAGG GAATAAAGG GACACGGAAA TGTTGAATAC TCATACTCTT TCCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATCT 4601 CTCATGAGCG GATACAATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCCG ACATTTCCCC GAAAAGTGCC ACCTGACCGT TAAGAAACCA	3701	TGCTTAATCA	GTGAGGCACC	TATCTCAGCG	ATCTGTCTAT	TTCGTTCATC	CATAGTTGCC	TGACTCCCCG	TCGTGTAGAT	AACTACGATA	CGGGAGGGCT
3901 AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGT 4001 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGGTCC GGTTCCCAAC GATCAAGGCC AGTTAACATGA TCCCCCATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTCCGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATC 4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC 4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGGC AAAACTCTCA AGGATCTTAC 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCA CTCGTGCACC CAACTGATCT TCAGCAATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA 4501 GCAAAATGCC GCAAAAAGG GAATAAGGG GACCGGAAA TGTTGAATAC TCATACTCTT TCCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATCT 4601 CTCATGAGCG GATACAATTT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCC CACTTTCCCC GAAAAGTGCC ACCTGACCGT TAAGAAACCA											
4001 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT 4101 TGTGCAAAAA AGCGGTTAGC TCCTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC 4201 TCTTACCTGTC ATGGCATCGG TAAGATGGTT TGTCGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAAA TAGTGATTGC GGGGACCAGA TTGGCTCTGC 4301 CCGGCGTCAA TACGGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGT CTTCGGGGCG AAAACTCCA AGGATCTTAC 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 GCAAAAATGCC GCAAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACCTCT CCTTTTTCAA TATTATTGAA GCATTTATCAC 4601 CTCATGAGCG GATACATATT TGAAATGTATT TAGAAAAAATA AACAAATAG GGTTCCGCC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAAACAC											
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGCCAGCAC TGCATAATTC 4201 TCTTACTGTC ATGCCATCGG TAAGATGGTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGG GGCGACCGAG TTGCTCTTGC 4301 CGCGCGTCAA TACGGGATAA TACCGCGCCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC 4401 CGCTGTTGGA ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAA CACGGTTCT GGGTGAGCAA AAACAGGAAG 4501 CGCAAAAAAGG GAATAAAGG GAATAAAGGG GACACGAAA AGCAAAATAG GGTTCACTC CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT 4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCCG ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA											
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC 4301 CCGGCGTCAA TACGGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAAACGT CTTCGGGGCG AAAACTCTCA AGGATCTTAC 4401 CCCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 GCAAAAATGC GCAAAAAAGG GAATAAGGG GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA 4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA											
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC 4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACCAC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 GCAAAATGCC GCAAAAAGG GAATAAGGG GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT 4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAG GGTTCCGCC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA											
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG 4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT 4601 CTCATGAGGG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA											
4501 GCAAAAAGG GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT 4601 CTCATGAGGG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAAACCA											
4601 CTCATGAGGG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA											
7/01 ITALIATORI GACATIAACC IATAMAMATA GGCGTATCAC GAGGCCCTIT CGT								ACMITICULU	GWWWWGIGCC	ACCIGNOGIC	INMOMMUCA
	- / U I	TIMITMICAL	GACALIANCC	INIMMMMAIA	GGCGIMICAC	GAGGCCC111	CGI				

### pMT-HisFlag-DmSNAP43 mut#6 STOP

1 TCGCGCGTTT CGG 101 TCAGGGCGCG TCA									
101 TCAGGGCGCG TCA									
201 CCGCACAGAT GCG									
301 TACGCCAGCT GGC									
401 TGAATTAATT CGI	TTGCAGGA CAG	GGATGTGG '	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC CGC	CCCACCGC CAC	CCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAG	GTGGAGAA CCG	GAACCAAT '	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGA	AACGAAAG ACC	CCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AAT									
001 188110181010 1811	1011101011 010		110000001110	1110111100000	11101111101111	> M H	н н н	H H T D	Y K D
901 GACGATGACA AGG	CCCACTAC TCA	лестеллт	ATCTTTCACC	A CTCCTCCCA	COTCOTCOAA				
				D C W E	L V O		R L V N	D G E	
1001 TCGAGGTGTT CTG									
			L Q L Q			T N H T	E V I		L A A L
1101 GCATGTGGCC AAG									
		C S R	RTT		F P A S	~	RIG	G F F L	L Y V
1201 ATCTACTACA AGO	CAGCCCAC GCA	ACAACTTT .	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
> I Y Y K	Q P T H	H N F	I K I	E V S P	R T W	Q E L	T D Y A	L D L	R K D
1301 GTCCGGAGCG GAA	AGGACACT CAT	CAGATCG (	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
>S P E R K	K D T H	Q I	AYML	WRL	T Q E	QAFR	FTA	L D Y	C Q G L
1401 GGACAATCTG GTG	GGACTACG ACC	CGTGTGGA	GACCGTAGCG	GGTGCCAAGG	AACAGAGGCA	GAGTGCCTTG	ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
> D N L V	D Y D	R V E	T V A	G A K	E O R O	SAL	M O K	OORA	N G V
1501 AGTCTCACAT ACG									
				0 A S 0	A A C		A A Y N		K O L
1601 CGGCTGGTCA TGA									
			S O I F			F A D I			R K S T
							Q S V		
1701 TCCAGATGAG AAA									
		T S T			R Q R V		A M Y	G V E E	REP
1801 CAACACCAGA CGG									
~ ~		~		N E T Y	~		A T V F	Q R E	L P E
1901 ACGTGCAGCA AGA									AAGCTATTTT
>D V Q Q E	E Y E M	I E	F S D D	E E M	E V G	E S E E	V T E	E E L	KAIL
2001 GGATACTTGA GTT	TAGTTTAT AAA	AACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	${\tt TTTATCTGAA}$	GGGCAATTCT	GCAGATATCC
> D T									
2101 AGCACAGTGG CGG	GCCGCTCG AGT	CTAGAGG (	GCCCGCGGTT	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT	TCTACGCGTA	CCGGTCATCA
2201 TCACCATCAC CAT		ACCCGCTG	3 ma3 aaamaa	л ототооотт		CACATCATA	GATACATTGA	TONGTOTOGA	CAAACCACAA
	I'I'GAG'I'I'I' AAA		AICAGCCICG		CTAAGATCCA				
2301 CTAGAATGCA GTG	GAAAAAAA TGO	CTTTATTT (	GTGAAATTTG	TGATGCTATT	$\tt GCTTTATTTG$	TAACCATTAT	AAGCTGCAAT	AAACAAGTTA	ACAACAACAA
2301 CTAGAATGCA GTG 2401 TTGCATTCAT TTT	GAAAAAAA TGO TATGTTTC AGO	CTTTATTT (	GTGAAATTTG GGAGGTGTGG	TGATGCTATT GAGGTTTTTT	GCTTTATTTG AAAGCAAGTA	TAACCATTAT AAACCTCTAC	AAGCTGCAAT AAATGTGGTA	AAACAAGTTA TGGCTGATTA	ACAACAACAA TGATCAGTCG
2301 CTAGAATGCA GTG 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATG	GAAAAAAA TGO FATGTTTC AGO GCAAGCTT GGO	CTTTATTT ( GTTCAGGG ( CGTAATCA '	GTGAAATTTG GGAGGTGTGG TGGTCATAGC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT	GCTTTATTTG AAAGCAAGTA GTGAAATTGT	TAACCATTAT AAACCTCTAC TATCCGCTCA	AAGCTGCAAT AAATGTGGTA CAATTCCACA	AAACAAGTTA TGGCTGATTA CAACATACGA	ACAACAACAA TGATCAGTCG GCCGGAAGCA
2301 CTAGAATGCA GTG 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGG	GAAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC	CTTTATTT ( GTTCAGGG ( CGTAATCA ' CTAATGAG '	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2301 CTAGAATGCA GTG 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGG 2701 GCATTAATGA ATG	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC CGGCCAAC GCG	CTTTATTT ( GTTCAGGG ( CGTAATCA ( CTAATGAG ( GCGGGGAG (	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2301 CTAGAATGCA GTG 2401 TTGCATTCAT TTI 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC CGGCCAAC GCC TCAGCTCA CTC	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  GCGGGGAG  CAAAGGCG	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGGAT	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT 2901 CCAGGAACCG TAA	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC CGGCCAAC GCG TCAGCTCA CTC AAAAAGGCC GCG	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  GCGGGGAG  CAAAGGCG  GTTGCTGG	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAA 2901 CCCGACAGGA CTA 3001 CCCGACAGGA CTA	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC GCGCCAAC GCC TCAGCTCA CTC AAAAAGGCC GCC ATAAAAGAT ACC	CTTTATTT ( GTTCAGGG ( CGTAATCA ( CTAATGAG ( GCGGGGAG ( CAAAGGCG ( GTTGCTGG ( CAGGCGTT ( CAGGCGTT ( CTTTCTGT ( CAGGCGTT ( CTTTCTGT ( CAGGCGTT ( CTTTCTGT ( CTTTCTGT ( CAGGCGTT ( CTTTCTGT ( CTTTCTTGT ( CTTTTTTT ( CTTTTTTTT ( CTTTTTTTT ( CTTTTTTTT ( CTTTTTTTT ( CTTTTTTTT ( CTTTTTTTT ( CTTTTTTT ( CTTTTTT ( CTTTTTT ( CTTTTTT ( CTTTTTT ( CTTTTT ( CTTTT ( CTTTTT ( CTTTTT ( CTTTTT ( CTTTTT ( CTTTT (	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTP 3101 CTCCCTTCGG GAA	GAAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGCTAC GCC TCAGCTCA CTC AAAAAGGCC GCC ATAAAAGAT ACC AGCGTGGC GCT	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  GCGGGGAG  CAAAGGCG  GTTGCTGG  CAGGCGTT  TTTCTCAT	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGG AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGGG TAI 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CGA	GAAAAAAA TGC TATTTTC AGG GCAAGCTT GGC CCGGGCAAC GCC TCAGCTCA CTC AAAAGGCC GCC AGCATGGC GCC AGCATGGC GCC AGCCTGGC GCC	CTTTATTT GTTCAGGG CGTAATCA CTAATGAG GCGGGGAG CAAAGGCG GTTGCTGG CAGGCGTT CTTCTCAT CTTATCCG	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	TGATGCTATT GAGGTTTTCTT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TIT 2501 ACCTGCAGGC ATC 2601 TAAAAGTGTAA ACC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCC	GAAAAAAA TGC TATGTTC AGC GCAAGCTT GGC CCGGGCAAC GCC TCAGCTCA CTC AAAAGGCC GCC AGCGTGGC TAC	CTTTATTT   CTTCAGGG   CGTAATCA   CTAATGAG   CCGGGGAG   CAAAGGCG   CAGGCGTT   CTTCTCAT   CTTATCCG   CGCGGTGC   CGCGGTGC	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCCC GGTAGGTACCT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGCT AACCCGGTAA GGCCTAACTA	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT	AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAC	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGGG TAI 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CGA	GAAAAAAA TGC TATGTTC AGC GCAAGCTT GGC CCGGGCAAC GCC TCAGCTCA CTC AAAAGGCC GCC AGCGTGGC TAC	CTTTATTT   CTTCAGGG   CGTAATCA   CTAATGAG   CCGGGGAG   CAAAGGCG   CAGGCGTT   CTTCTCAT   CTTATCCG   CGCGGTGC   CGCGGTGC	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCCC GGTAGGTACCT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGCT AACCCGGTAA GGCCTAACTA	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT	AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAC	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TIT 2501 ACCTGCAGGC ATC 2601 TAAAAGTGTAA ACC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCC	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGC CCTGGGGT GCC CCGGCCAAC GCC ATAAAGGCC GCC ATAAAGAT ACC AGCGTGGC GCT AGCGTGGC GCT AGCGTGGC GCT AGCGTTGG AAA	CTTTATTT  GTTCAGGG  CTTAATCAG  GCGGGGAG  CAAAGGCG  GTTGCTGG  AGGCGTT  CTTATCCG  GGCGGGGC  AAAGAGTT	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTACTG TACAGAGTTC GGTAGCTTCT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTTGAGTGGT TTGAGTGGT GATCCGCAA	GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT GCTGCTACACT GCTGGTACCC	AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACAG GTGGTTTTTT	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCATAT CACGAACCCC CTGGTACAG CTGCGCTCTC CAGCAGATTA
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CAA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA	GAAAAAA TGC TATGTTTC AG GCAAGCTT GGC CCTGGGGT GCC CGGCCAAC GCC AAAAAGGC GCC AAAAAGAT ACC ACCGCTGC GCC GAGGTATG TAA AACGATCT CAA	CTTTATTT  GTTCAGGG  CGTAATCAG  GCGGGGAG  CAAAAGGCG  CAGGCGTT  CTTCTCAT  CTTCTCAT  CTTCTCAT  CGGCGTGC  AAAAGAGTT  AGAAGATC	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GGTAGCTTCT GGTAGCTCTT CTTTGATCTT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCACAGA TAGGCTCCCTC GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGCT TTGAAGTGGT TTTAAGTGGT TTTAAGTGGG	GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCT	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGCTTACACT GCTGGTACCG AGTGGAACGA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCACGCC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTTT AAACTCACGT	AAACAAGTTA TGGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAG GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGAG
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TIT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA	GAAAAAA TGC GAAGATTTC AGG GCAGACTT GGC CCTGGGGT GCC CCGCCAAC GCC TCAGCTCA CTC AAAAAGGC GCC ATAAAAGAT ACC AACGCTGC GCC GAGGTATG TAC ACCTTCG AAA AAGGATCT CAA	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  CCGGGGAG  CAAAGGCG  CAAAGGCG  CTTCCTGG  CTTCTCTCAT  CTTATCCG  GGCGGTGC  CAGAGAGAGATT  AGAAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGAGACC  LAGACC  LAGACC	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGGT TACAGGTTACGGT TACAGGTTTCC TACAGGTTTC TACAGAGTTC TTTGAATTAC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TTCTAGAGTGC TTCAGAGGGAAATGAAGGT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCACC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC TTAAATCAAT	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA TGTTCCGACC TAGGTCGTC GCACACCTCGCTC GCTGGTACAC GCTGGTACAC GCTGGTACAC GCTGGTACAC CTAAAGTATA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGG AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGT GGCACACCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TAAGGGATTT CTTGGTCTGA	ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCAGAATACAG CAGTTACCAA
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TIT 2501 ACCTGCAGGC ATC 2601 TAAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGG	GAAAAAA TGC TATGTTTC AGG GCAAGCTT GGG CCTGGGGT GCC CCAGGCTA CTC AAAAGGCC GCC ATAAAAGGC GCC ACCGCTGC GCC ACCGCTGC GCC ACCGCTGC GCC ACCGCTGC GCC ACCGCTGC AAAAGGAT CAC ACCTTCG AAAAAGGAT CTAC ACCGTTCG AAAAAGGATCT CAAAAGGATCT CAAAAGGATCT CAAAAGGATCT CAAAAGGACCTTCA CCT SAAGGCACC TAT	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  GCGGGGAG  CAAAGGCG  CAAAGCGT  CTTTCTCAT  CTTATCGG  GGCGGTGC  AAAGAGTT  AGAAGATC  ICCCAGCGC  ICCCAGCC  ICCCAGCGC  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGC  ICCCAGCG  ICCCAGC  ICCCAGCG  ICCCAGCG  ICCCAGCG  ICCCAGC  ICCCAGCG  ICCCAGC  ICCCAG	GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTATTCCA TCCCCCTGGA AGGTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGGT ATTCTACGGGG AAATGAAGTT TTCGTTCATC	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGTC TTAAATCAAT CATAGTTTGCC	TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAACGACTT CGGCTACACT GCTGGTAGCG AGTGGAACGA CTAAAGTATA TGACTCCCCG	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGACATGTG AAATCGACGC CTGCCGCTTA AGCTCCAGCAGCT ATCGCCACTG AGAAGAAGCAGG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGATA	AAACAAGTTA TGGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGAAGCC CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTCAAC CTTGGTCTAC CTTGGTCTCA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTCCCAGCT CGTTCGCTTC CGTTCGCTTC CAGCAAAAGG GGTGGCAAAA GTCCGCCTTT CAGCAACCC CTGGTAACAG CTCCGCTCTC CAGCAAGATTA TGGTCATCAG CAGTTACCAA CAGGGAGGCT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC	GAAAAAA TGC GAAGATTTC AGG GCAAGGTT GGC CCTGGGGT GCC CCGCCAAC GCC AAAAGGC GCC ATAAAGAT ACC AGCGTGGG GCC AGCGTATG GCC ACCGCTGC GCC GAGGTATG TAC ACCTTCGG AAA AAGGATCT CAA GATCTTCA CCT GAGGAGCACC TAT CCAGTGC GCC	CTTTATTT  GTTCAGGG  CGTAATCA  CTTAATGAG  GCGGGGAG  AAAGGCG  CAGGCGTT  CTTCTCAT  CGCGGTGC  AAAGAGTC  AGAAGATC  TAGAAGATC  TACTCAGCG  AATGATAC	GTGAAATTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG CGTATTCCA TCCCCTGGA AGCTACGCT GTAACTATCG TACAGAGTTC GGTAGCTCT TTTAAATTAA ATCTGCTACT CCGCAGACCC CGGAGACCC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAAGTGT TTCTTCACTCACACACACACACACACACACACACACAC	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGGAT CCCCTGACG TGCGCTTACG AACCCGCTAA GGCCTAACTA ACAAACCACC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	TAACCATTAT AAACCTCTAC TATCCGCCCG TTCCTCGCTC ACGCAGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACT GGCGTACACT GGCGTACACT GTGGACGACGA TGAGTCCCCCCG TATCACCACT TGACCCACT	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAACAGCT ACCACCTG AGAAGACATTT AGAAGGACAG TTCGCCACTG AGAAGACACT ATCACCTT AGAAGTAAA TCGTGTAAAA TCGTGTAAAA AAACCACGCA AAACCACGCA	AAACAAGTTA TGGGTGATTA CAACGTACTGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTGTTGTCAAG AACTACGATA GCCGGAAGGG	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGGAACCC CTGGTAACAG CTGGGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CAGGAGGCCT CCGAGGGCCCC CCGAGCGCCC
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TTT 2501 ACCTGCAGGC ATC 2601 TAAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CAA 3201 CGGTTCAGCA GCC 3401 CTGAAGCCAG TTA 5501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCC 3901 AAGTGGTCCT GCA	GAAAAAA TGC TATGTTTC AGG CCTGGGGT GCC CCTGGGGT GCC CCGCCAAC GCC TCAGCTCA CTC AAAAAGGC GCC ATAAAAGAT ACC AGCGTGG GCT ACCGCTGG GCC GAGGTATG TAC ACCTTCGG AAA GAGGATCT CAA GATCTTCA CCT GAGGCACC TAT CCCAGTGCT GCC AACCTTTCA CCT AACCTTTCA CCT AACCTTTCA CCT AACCTTTCA CCAACCTTTCA CCAACCTTTCA CCAACCTTCA CCAACCTTCA CCAACCTTCA CCAACCTTCA CCAACCTTCA CCCAACCTTCA CCCAACCTCAACAA	CTTTATTT  GTTCAGGG  CGTAATCA  CTAATGAG  CAAAGGCG  GTTGCTGG  CAGGCGTT  CTTTCTCAT  CTTATCGG  GGCGGTGC  AAAGAGTT  AGAAGATC  TCTCAGCG  AATGATACC  GCCTCCAT	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGGT TACAGGTT GTAACTATCG TACAGGTT TTAAATTAC TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCCAGTCTATT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCTCG GTAGGTATCT TCTTGAAGTGC TTGAAGTGCT TTGAAGTGCT TTCACGGGA AATGAAGTT TTCTTCACGGCA AATGATTCT ACGCTCACCG AATTGTTCACC AATTGTTCACC AATTGTTCAC	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGAT CCCCCTGACG TGGGCTCCC CAGTTCGGTG AACCCAGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGATTA GGCAACCTAC	TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGCA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACTTCCGACC TAGGTCGTTC GCTGTACCG GCTGGTACCG CTGAGACGACTT TGACTCCCCG TAAAGTATA TGACTCCCCGA AGTAAGCAAT AGTAAGTAGT AGTAAGTAGT	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATGCACACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGGTAAGAT ACGCCAGTG ACACGCCA TCGCCAGTT	AAACAAGTTA TGGCTGATTA TGGCTGATTA GAAAACGC TCAAGTCAGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGT GGCAGCAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA CCGAAGGATACT CATTTGCAAG AACTACGATA CATTTGCAAG AACTACGATA CCCGAAAGGC ATAGTTTTCCA	ACAACAACAA TGATCAGTCG GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAC CAGTTACCAA CAGTTACCAA CAGGAGGGCT CCGAGCGCAC CAACTTTACCAA CAGCTTGTACAC CAGCTTGTACAC CAGCTTACCAA CAGCTTGTT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGGG TAT 3001 CCCGACAGGA CTA 3001 CTCCCTTCGG GAA 3201 CTCTCTCGG GAA 3201 CGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCA 3901 AAGTGGTCCT GCA 4001 GCCATTGCTA CAC	GAAAAAA TGC GAAGATTTC AGG GCAGCTT GGG CCTGGGGT GCC CCGCCAAC GCC AAAAGGCC GCC ATAAAAGGC GCC ACCGCTGC GCC ACCGCTGC GCC ACCGCTGC GCC GAGATCT CAC GAGATCT GCC GAGATCT GCC GCCATCGT GCC GCCATCGT GCG	CTTTATTT  GTTCAGGG  CTAATGAG  CTAGAGAT  CTTATCCAG  AAAGAGT  TAGAATCCT  TAGAATCCT  TAGAATCCT  TAGATCAT  CTCCAGGG  AATGATAC  GCCTCCAT  CTTCAGGC  CTCCAGCC  CTTCAGGC  CTCCAGCC  CTTCAGGC  CTCCAT  CTCCAGCC  CTTCAGGC  CTCCAT  CTCCAGCC  CTCCAT  CTCCAGCC  CTCCACCC  CTTCAGCC  CTCCAT  CTCCACCC  CTCTCACCC  CTCTCACCC  CTCCACCC  CTCTCACCC  CTCTCACC  CTCTCACCC  CTCTCACC  CTCTCACCC  CTCTCACCC  CTCTCACCC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTCTCACC  CTC	GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTATTCCA TCCCCTGGA AGCTCACGCT GTAACTACGT TACAGAGTTC GGTAGCTCT TTTCAACTATC TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTT TCGTCGTTTT TCGTCGTTTT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGG AAATGAAGTT TTGGTTCATC ACGCTCACCG GAATTGTTCATC ACGCTCACCG GTATGGCTTC	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGTTGCGTT	TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAAGACTT CGGCTACACT GCTGGTAGCG AGTGGAACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTAGT GGTTCCCAAC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGACAGACT ACGCCACTG AGAAGATTTT AAACTCACGT TATGAGACAG TATGAGAAGAAGAAAAACCAGCCA TCCGCGAGTAAA TCGCCAGTTA AAACCAGCCA TCGCCAGTTA GATCAAGGCG	AAACAAGTTA TGGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGCAGAGCCAGATACCT CAGCAGAGCCATACTT TGTTTGGTAT TGTTTTGCAAG TAAGGATTCATACTTTGCTAT CTTGGTAT CCCGGAAGGG ATAGTTTTCCG AGTTACATGA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGCAACCCC CTGGTAACAG CTGCGCTCTG CAGCAACATT TGGTCATACAG CAGCAGATTA CAGGAAGGCCC CCGGAGGGCCAG CAACGTTGTT CCCCCATGT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCAG GAA 3201 CGGTTCAGCC CGA 3301 GATTAGCAG GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCC 3901 AAGTGGTCCT GCA 4001 GCCATTGCTA CAC 4101 TGTGCAAAAA AGC	GAAAAAA TGC GAAGATTTC AGG GCAAGGTT GGC CCTGGGGT GCC CCTGGGGT GCC CCGCCAAC GCC AAAAAGGC GCC ATAAAAGAT ACC ACCGCTGG GCC ACCGCTGC GCC AACCTTCG ACCTTCGA AACCTTCGA CTC AACCTTCGA CTC AACCTTCTCA CTC AACTTTCA CCT AACTTTCA CCT AACTTTCA CCT AACTTTCA CCT CCAGTGCT GCC AACCTTTCA CCT CCAGTGCT GCC CCGTTAGC TCCC CCGTTAGC TCCC CCGTTAGC TCCC CCGTTAGC TCCC	CTTTATTT  GTTCAGGG  CTAATGAG  CTAATGAG  CTAATGAG  CCGGGAGG  CAGGGGTG  CTTATCCG  GGCGGTGC  AGAAGATCT  AGAAGATCAT  AGAAGATCT  AGAAGATC  AGAAGATC	GTGAAATTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG CGTATTCCA TCCCCTGGA AGCTACGCT GTAACTATCG GTACTATCC GTACTATC TTAAATTA TCTTGATCT TTTAAATTA TCGCGAGACCC CCAGTCTATT TCGTCGTTTC TCGTCGTTTC TCGTCGTTTC TCGTCGTTTC TCGTCGTTTC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAGTT TTCTTCACTCGCA AATTGTTGCC GTATGGTTCATC GGTATGGTTCATC CGTATGGTTCCT TGTCACAGAGT TTCGTTCACCG TTGAGTACCC TGTAGGAGT TTCGTTCACCG TGTATGGTTCC TGTCAGAAGT TTCGTTAGC TGTCAGAAGT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGTTGCGCT GCGCTTACGC TGGGCTTACC AACCCGCTAA GGCCTAACTA AACAACACA TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTCC AACTTGCCCA	TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC ACGCACAA AGCATCACAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTACAC GCTGCTACACT GCTGGTACACG AGTGGAACGA CTAAAGTATA GTAACCACA AGTAAGTAGT GGTTCCCAC TATCCAGCAT AGTATCCCAAC CACTGTTATC	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAACAGCT ACCACCTG AGAAGACATTT AGAAGCACGT ATCGCCACTG AGAAGACACGT ATCGCCACTG AGAAGTAAA TCGTGTAAAA TCGTGTAAAA AAACCACGCA TCGCCAGTTA GAATCAAGGCC ACTCAAGGCC ACTCATGGTT	AAACAAGTTA TGGCTGATTA CAACCATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT GGACTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTGCG AGTTACATGA ATGGCACCAC	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGGAACCC CTGGTAACAG CTGGGCTATA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGGAGGCT CCGAGGGCAC CAGGGGCAG CAAGGTTGT TCCCCCATGT TCCCCCATGT TCCCCAATTT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCC 3901 AAGTGGTCCT CGA 4001 GCCATTGCTA CAG 4101 TGTGCAAAAA AGC	GAAAAAA TGC GAAGATTTC AGG CCTGGGGT GCC CCGCCAAC GCC TCAGCCAC GCC AAAAAGGCC GCC AAAAAGAC GCC AAAGGATC TAAAAAGAT ACC AAGCGTGG GCC GAGGTATG TAC GAGCTCC GCC GAGGTATC CAC GAGCACC TAAAAGAT GCC AACTTTAT CCC GCATACT GCT GCCATACT GCT GCCATACCT GCT GCCATACCT GCT GCCATACCT GCT GCCATACCT GCT GCCATACCT GCT GCCATACCT TAAC GCCATACCT	CTTTATTT  GTTCAGGG  CTAATGAG  CAAGAGGTT  CTTACTCAG  CTAGATCAT  CTCAGCG  AATGATAC  CCTCCAT  CGTCACGC  CTTCACGC  CTTCACGC  CTTCACGC  CTTCACGC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCACCC  CTTCCACCC  CTTCCACC  CTTCACC  CTTCCACC	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGCT GTAACTATCG TTAACTATCG TTAACTATCG TTAACTATCG TTAACTATCG TTAACTATCG TTTGATCTT TTTAAATTAA ATCTGTCTAT TCGGAGACC CCAGTCTATT TCGTCGTTTG TCCTCGTTGACT TTCGTGACT TTCTGTGACT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACGA ATGGCTCCGC AGCTCCCTCG GTAGGTACT TCTTGAGTACT TCTTGAGTACT TTTGAGTGC TTCTAGAGTGT TTCTACGGGA AATGAAGTT TTCTTCACC AATTGTTCACC AATTGTTCACC GTATGGCTTC TGTCAGAGTT TGTCAGAAGT GGTGAGTACT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGCGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCCAGTAAACAAC CCC TTAAAACAACT TCTGACGGTC TTAAATCAAT CATAGTTGCC GCTCCAGATTA GGCAGATTA GGCAGATTA GGAAGCTAC ATTCAGCTCC AAGTTGCCC	TAACCATTAT AAACCTCTAC TATCCGCCCGC TTCCTCCGCTC AACGCAGGAA AGCATCACAA GGTTCCGACC TAGGTCGTCC GACACGACTT CGGCTACACT GCTGGTACCG CTGAGGCA CTAAAGTATA TGACTCCCGG GGTACCAC TATCAGCAAT AGTAAGTATA AGTAAGTATA AGTAAGTACCAC CATGTTACCA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGACAGTG ATGGTTTTT AGACTACAGT TATGAGTAAA TCGTGTAAAT TCGTGTAAAT TCGCCAGTTA GATCAAGGCA TCGCCAGTTA GATCAAGGCG ACTCATGAGTTA TATGATTAAT TATGATTATGT TATGATGTATGT	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT GGGCTGTGTG GCACCAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATT CTTGGTCTGA AACTACGATA AACTACGATA AACTACGATA AATTTCCA GCCGAAAGGC ATAGTTTCCA GCTAACGTAA AACTACATAA AACTACATAA AACTACATACATAA AACTACACATA AACTACACATA AACTACACATA AACTACACATA AACTACACACAC	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GCTGGCAAAA GTCCGCCTTT CACGAACCC CTGGTAACAC CTGCTAACAC CAGCAGATTA GGTCAGTACAC CAGCAGATTA CCGAGCGCAC CAACGTTGT TCCCCCATGT TCCCCCATGT TGCATAATTC TTGCTCTTGC
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCA 4001 GCCATTGCTA CAC 4101 TGTGCAAAAA AGC 4201 TCTTTACTGTA CAC 4201 TCTTTACTGTA ATC 4301 CCGGCGTCAA TAC	GAAAAAA TGC GAAGATTTC AGG GCAAGGTT GGG CCTGGGGT GCC CCGCCAAC GCC ATAAAAGGC GCC ATAAAAGGT ACC ACCGCTGC GCC ACCGCTGC GCC ACCGCTGC GAA AAGGTTC CAC GAACTTCG AAA AAGGATCT CAC GAACTTCA CTC GAAGATCT CAC GAACTTCA CTC GAAGGCACC TAT CCAGTGCT GCC GAACTTTAG CCC GCCATCAC GCC GCCATCAC TCC GCCATCAC TCC GCCATCAC TCC GCCATCAC TCC CCGCATAA TAC CCGGCATAA TAC CCGGCATAA TAC	CTTTATTT  GTTCAGGG  CTAATGAG  CTAGACGT  CTTACCAG  AAAGAGT  TCTCAGGG  AATGATAC  CTTCAGGC  CTTCAGGC  CTTCAGGC  CTTCAGGC  CTTCAGGC  CTTCAGGC  CTTCAGCC  CTTCAGCC  CTCAGCCCAC  CTCCAGCC  CTCCAGCC  CTCCAGCC  CCCCCCCCC  CCCCCCCCC  CCCCCCCCCC	GTGAAATTG GGAGGTGTGG TGAGCTAACT TGAGCGATTGG GTAATACGGT CGTTATTCCA TCCCCTGGA AGCTCACGT GTAACAGCT GTACAGTTC GTTACAGGTTC GTTACAGAGTTC TTTTAAATTAA ATCTGTCTT TCTAGACTT TCCAGTCTAT TCGCAGACCC CCAGTCTAT TCCTCGATCGT TTCTGTGACT TTCTGTCTTT TCGTCGATCGT TTCTTTGACCC CATAGCAGAC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAATGC GATCCGCAA TTCTACGGG AAATGAAGTT TTCGTTCATC ACGCTCACCG GATAGCTTCATC GTTAGAGTT TTGTTCATC GGTAGGCTTC CGTTAGAAGT CGTTAGAGAGT CGTTAGAAGT CGTGAGAAGT CGTGAGAAGT CGTGAGAAGT CGTGAGAAGT CGTGAGAAGT CTTTAAAAGT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGAC TGGGCTCTCCC CAGTTCGGTG AACCCGGTAA GGCTAACTA ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGGCC CACCAAGTC CAACCAAGTC CAACCAAGTC CACCAAGTC CCTCATCATT	TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAAGACTT CGGCTACACT GCTGGTAGCG AGTGGAACGA CTAAAGTATA AGTAAGTATA AGTAAGTATA AGTAAGTA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGACAGACT ACGCCACTG AGAAGATTTT AAACTCACGT TATGAGACAG TATGCCACTG AGAAGACAGCA ACCAGCAA TCGCTAGATAA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA GATCAAGGCG ACTCATGGTT TAGTGATATT TAGTTATGC CTTCGGGGCG	AAACAAGTTA TGGGTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGCAGAGCAG	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCAAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAACACC CTGCGATTA TGGTCATGAG CAGCAGATTA TCGTCAGCAG CAGCAGTTTA CCACAGTTTACCAA CAGTTGTT CCCCCATGT TCCCCCATGT TCCACCATGT TGCATCATTC TTGCTCTTCC AGGAGCATTATC
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC TAT 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGGGG TAT 2901 CCAGGAACCG TAA 3001 CCCGACAGGA CTA 3101 CTCCCTTCAG GAA 3201 CGGTTCAGCC CGA 3301 GATTAGCAGA GCC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3901 AAGTGGTCCT GCA 4001 GCCATTGCTA CAC 4101 TGTGCAAAAA AGC 4201 TCTTACTGTC ATC 4301 CCGGCGTCAA TAC 4301 CCGGCGTCAA TAC 4401 CGCGCGTCAA TAC	GAAAAAA TGC GAAGTTTC AGG GCAAGGTT GGC CCTGGGGT GCC CCTGGGGT GCC CCTGAAAAGGC GCC AAAAAGGC GCC AACCGCTGC GCC AACCGCTGC GCC AACCTCG ACCTCG AACTTCA CTA CCAGTGC GCC AACTTTCA CTA CCAGTGC GCC AACTTTAT CCC GCCATCC GCC GCCATCC TAACCCC CCAGTTAG TCC CCAGTTAG TCC CCAGTTAG TCC CCAGTTAG TCC CCAGTTAG TAC CCAGTTAG TCC CCAGTTAG TCC CCAGTTAG TCC CCAGTTAG TAC CCAGTTAG TAC CCAGTTAG TAC CCAGTTAG TAC CCAGTTAG TCC	CTTTATTT  GTTCAGGG  CTAATGAG  CTAATGAG  CTAATGAG  CCAGGGAG  CTACTGG  CAGGGGTG  AGAAGATCT  CCCCAGCAC  CCCCCGCC  GAATGCTT  CCCCCCCCCC	GTGAAATTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGC CGTTTTCCA TCCCCCTGGA AGCTAACT GTAACTATCG GTAACTATCG TTACAGAGTTC CTTTGATCTT CTTTGATCTT TTTAAATTAA TCCGCAGACCC CCAGTCTATT CCTCGTTTCCTTTCC	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTTGGGC TATCCACGA TAGGCTCCCC GTAGGTACT TCTTGAGTC TTTAAAGTGC AATTCTACGGA ATTCTACGGGA ATTCTACGGGA AATTCATCACGGA AATTCATCT ACGCTCACC AATTGTTGCC GGTAGGCTC TGTCAGAAGT GGTAGGCTC TGTCAGAAGT CGTTCAGAAGT CGTTAGAAGT CGTTAAAAGT CACTTTAAAAGT CAACTGATCT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCTGACG TGCGCTTACCG AACCCGGTAA GGCCTAACTA AACAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCCTC AACTTGCCC CAACCAGTC CAACCAGTT TCAGCATT	TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC ACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACT GCTGGTACAC AGTGGAACGA CTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTATACTAGT ATCTGAGCAC CAGTGTTATC ATCTGAGAA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGACGACTG ACTCAGCTG ACTCAGCT ACTCAGCT ACTCACCT ACTCAGCT ACTCACCT ACTCACC ACTCACCT ACTCACCT ACTCACCT ACTCACCT ACTCACC ACTCACCT ACTCACC ACTCACT ACTCACC AC	AAACAAGTTA TGGGTGATTA CAACGTACTA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAAGTCAGA CCGGATACCT TGTTTGCAAG TAATTGGTAT CTTGGTTGCAAG TAAGGGATTCT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTACGATA GCTGAAGGA ATGGCACGACGAG ATGGCACGAC GGCGACCGAG GAAAACTCTCA GGGTGACCAA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGGAACCC CTGGTAACAG CTGCGCTTT TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGGAGGCT CCGAGCGCAT TCCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCCTCTTGC AGGATCTTTC TGGTCTTTGC AGGATCTTTC AGGATCTTTAC AAACAGGAAC
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TTA 2901 CCAGGAACGG TTA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CGA 3301 GATTAGCAGA CGC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCC 3901 AAGTGGTCCT GCA 4001 GCCATTGCTA CAG 4201 TCTTACTGTC ATG 4301 CCGGCGTCAAA AAC 4201 TCTTACTGTC ATG 4401 CGCGCTGTTAA TAC 4401 CGCGCTGTTAA TAC 4401 CGCGCTGTTAA TAC 4401 CGCGCGTCTAA TAC 4401 CGCGCTGTTGAA AAC 4401 CGCGCGTCTAA TAC 4401 GCCATTTGAA	GAAAAAA TGC GAAGATTTC AGG CCTGGGGT GCC CCGCCAAC GCC CCGCCAAC GCC AAAAAGGCC GCC ATAAAAGAT ACC AACCTTCG ACC AACCTTCG ACC CCAGGCT GCC AACTTCA CCC GCCATCA TGC CCAGGCT GCC AACTTTAT CCC GCCATCCT GCC CCGCTCGT GCC CCGCTTAGC TCC CCCGCTCGT GCC CCCGTCTCG TACC CCCGCTTAGC TCC CCCGCTTAGC TCC CCCAGTTCG TCC CCCAGTTCG TACC CCAGTTCG TACC CCAGTTCC CCAGTTCC CCAGTTCC CCAGTTCC CCAGTTCC CCAGTTCC CCA	CTTTATTT  STTCAGGG  CTTAATGAG  CTAATGAG  CTAATGAG  CTAATGAG  CAAAGGCG  STTGCTGG  CAGGGGTG  CAGGGGTC  CTTACTCAT  CTTACTCAG  AAAGAGATT  CTCAGCG  AATGATAC  CCTCCAT  CCTCAGCG  AGATGCT  CCTCAGCG  AGATGCT  CCTCCACC  CTTCGTCCACC  CTTCGTCCACC  CTTCGTCCACC  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CTTCGCCCA  CGATGCTT  CCCCCGCCA  CGATGCTT  CCCCGCCCA  CGTAACCCA  CTTAAGGCC  CTTAAGGCC  CGTAACCCA	GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GTAACTATCG TACAGAGTTC GTAACTATCG TACAGAGTTC GTAACTATCG CTCGTGAACTATC CTTTGATCTT TTTAAATTAA ATCTGTTAT TCGCGAGAACC CCAGTCTATT TCGTCATTT TCTCGTTTGT TTCTGTGACT CTCCGTAGCT CTCCGTAGC CATAGCAGAA	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACGGA AGCTCCCTCG GGTAGGTACT TCTTGAGTGC TTCTAGAGTGGT TTCTACGGGA AATGAGTT TTCTTCACGGG AATGTTCAT ACGCTCACG AATTGTTGAC GTATGGCTTC TGTAGAAGT GGTGAGTACT CTTTAAAAGT CTAACTGATCT TGTTGAATAC TGATTGATCT TGTTGAATT	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGAT CCCCCTGACG TGCGCTTACCC CAGTTCGGTG AACCAGCTAA GCCTAACTA ACAAACCACC TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAC ATTCAGCTCC AACTAGTCC CAACCAAGTC GCTCAACATT TCAACATCTT TCAACATCTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTTT TCAATACTCTT	TAACCATTAT AAACCTCTAC TATCCGCCCG TACCTCGCCCG TTCCTCGCTC AACCCAGGAA AGCATCACAA GGTACACA TGTTCCGACC TAGGTCGTC GACCACACA GCTGGTACCA GCTGGTACCA TGTTCCCAAC TAAGGTACAA AGTAAAGTATA AGTAAGTATA AGTAAGTATA AGTAAGTA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGC AGAACATGTC AGAACACGC CTGCCGCTTA GCCACACT ATCCACACT ATCCACACT ATCACACT ATCACACT ATCACACT AGAAGACAC TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACGCA TCGCCAGTTA GATCAAGGCC ACTCCACGCT TAGTATGC TAGTGTATGC TAGTGTATGC TAGTGTATGC TTCGGGGCC CTCGCGGGTCC TATTATTGAA	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCACCA TATTTGCAAG TATTGCAGA TAAGGGATTT CTTGTCTGA AACTACGATA AACTACGATA ATTACATGA ATTACATGA ATGGCAGAAGGC AGTATACTCA GGGGACCGAG AAAACTCTCA GGGTAGCAAA GCATTTATCA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAACCC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCAGATTA CAGCAGATTA CGGAGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCATGT TCCCCCATGT TCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGAACAGGAG GGTTATTGT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TI 2501 ACCTGCAGGC ATC 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCAGGGG TAT 2901 CCAGGAACCG TAA 3001 CCGCACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGCC CGA 3301 GATTAGCAGA GCA 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGG 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCA 4101 GCCATTGCTA CAC 4101 TGTGCAAAAAA AGG 4201 TCTTAATGTA ATC 4401 CGCTGTTCAA ATC 4401 CGCTGTTCAA ATC 4401 CGCTGTTCAA ATC 4401 GCCAAAATGCC GCA 4401 GCCAAAATGCC GCA 4401 GCCAAAATGCC GCA 4401 GCCAAAATGCC GCA 4401 CTCAATGAGGG GAA	GAAAAAA TGC GAAGATTTTC AGG GCAGCTTGGGGT GCC CCGCCAAC GCC CCGCCAAC GCC AAAAGGCC GCC ATAAAAGC GCC ACCGCTGC GCC ACCGCTGC GCC AGCAGTATC TAC ACCTTCGG AAA AAGGATCT CAA GAGCATCA CTC GAGCACC TAT CCAGTGCT GCC AAACTTTAT CCC GGCATAA TAC CCAGTTCG ATC AACTTTAG TCC CCGGATAA TAC CCAGTTCG ATC AAAAAAAAGG AT TACATATT TGT	CTTTATTT  STTCAGGG  CGTAATGAG  CTAATGAG  CTACTCAG  CTACTCAG  AAAGAGT  AGAAGATC  TAGATCCT  TCTCAGCG  AATGATAC  CTTCAGCG  CTTCAGGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGC	GTGAAATTG GGAGGTTTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTACTATCG GTACTATCG GTACTATCG TTTTGATCTT TTTTAAATTAA ATCTGTCTTT TCGGAGACCC CCAGTCTATT TCGTCGATCGT TTCCTGTGACT TTCTTGACTT TTCTTTTCTT	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTTGGGC TATCCACGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT AATTGAGTCC AATTGAGTCC TGAAGTGGT AATTGATCATC ACGCTCACCG AATTGTTGCT GTTAGAGAGT TGTTGAGAAGT TGTTGAGAAGT TGTTGAGAAGT TGTTGAAAAGT CAACTGATCT TGTTGAAAAGT CAACTGATCT TGTTGAATACT ACCACTAATAGA	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGACG TGCGCTTCCGC TGCGCTTCCGC ACGTCGGTA ACCCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCCC CAACAGTC CAACCAAGTC TCAACAAGTC CACCAAGTT TCAGCACTT TCAGCACTT TCATACTTT TCATACTTT	TAACCATTAT AAACCTCTAC TATCCGCCCG TACCTCGCCCG TTCCTCGCTC AACCCAGGAA AGCATCACAA GGTACACA TGTTCCGACC TAGGTCGTC GACCACACA GCTGGTACCA GCTGGTACCA TGTTCCCAAC TAAGGTACAA AGTAAAGTATA AGTAAGTATA AGTAAGTATA AGTAAGTA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGC AGAACATGTC AGAACACGC CTGCCGCTTA GCCACACT ATCCACACT ATCCACACT ATCACACT ATCACACT ATCACACT AGAAGACAC TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACGCA TCGCCAGTTA GATCAAGGCC ACTCCACGCT TAGTATGC TAGTGTATGC TAGTGTATGC TAGTGTATGC TTCGGGGCC CTCGCGGGTCC TATTATTGAA	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCACCA TATTTGCAAG TATTGCAGA TAAGGGATTT CTTGTCTGA AACTACGATA AACTACGATA ATTACATGA ATTACATGA ATGGCAGAAGGC AGTATACTCA GGGGACCGAG AAAACTCTCA GGGTAGCAAA GCATTTATCA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAACCC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCAGATTA CAGCAGATTA CGGAGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCATGT TCCCCCATGT TCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGAACAGGAG GGTTATTGT
2301 CTAGAATGCA GTC 2401 TTGCATTCAT TT 2501 ACCTGCAGGC ATG 2601 TAAAGTGTAA AGC 2701 GCATTAATGA ATC 2801 CGGCGAGCGG TTA 2901 CCAGGAACGG TTA 3001 CCCGACAGGA CTA 3101 CTCCCTTCGG GAA 3201 CCGTTCAGC CGA 3301 GATTAGCAGA CGC 3401 CTGAAGCCAG TTA 3501 CGCGCAGAAA AAA 3601 ATTATCAAAA AGC 3701 TGCTTAATCA GTC 3801 TACCATCTGG CCC 3901 AAGTGGTCCT GCA 4001 GCCATTGCTA CAG 4201 TCTTACTGTC ATG 4301 CCGGCGTCAAA AAC 4201 TCTTACTGTC ATG 4401 CGCGCTGTTAA TAC 4401 CGCGCTGTTAA TAC 4401 CGCGCTGTTAA TAC 4401 CGCGCGTCTAA TAC 4401 CGCGCTGTTGAA AAC 4401 CGCGCGTCTAA TAC 4401 GCCATTTGAA	GAAAAAA TGC GAAGATTTTC AGG GCAGCTTGGGGT GCC CCGCCAAC GCC CCGCCAAC GCC AAAAGGCC GCC ATAAAAGC GCC ACCGCTGC GCC ACCGCTGC GCC AGCAGTATC TAC ACCTTCGG AAA AAGGATCT CAA GAGCATCA CTC GAGCACC TAT CCAGTGCT GCC AAACTTTAT CCC GGCATAA TAC CCAGTTCG ATC AACTTTAG TCC CCGGATAA TAC CCAGTTCG ATC AAAAAAAAGG AT TACATATT TGT	CTTTATTT  STTCAGGG  CGTAATGAG  CTAATGAG  CTACTCAG  CTACTCAG  AAAGAGT  AGAAGATC  TAGATCCT  TCTCAGCG  AATGATAC  CTTCAGCG  CTTCAGGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGCC  CTTCAGC  CTTCAGCC  CTTCAGC  CTTCAGC	GTGAAATTG GGAGGTTTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTACTATCG GTACTATCG GTACTATCG TTTTGATCTT TTTTAAATTAA ATCTGTCTTT TCCGGAGACCC CCAGTCTTTT CCTCGATCGT TTCCTCGATCGT CTCAGTGACC CTCAGTCAGACAC CTCGATCGT CTTCGATCGT CTCAGTGACAC CTCGATCGT CATAGCAGAA CTCGTGCACC GACACGGAAA TAGAAAAATA	TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTTGGGC TATCCACGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT AATTGAGTCC AATTGAGTCC TGAAGTGGT AATTGATCATC ACGCTCACCG AATTGTTGCT GTTAGAGAGT TGTTGAGAAGT TGTTGAGAAGT TGTTGAGAAGT TGTTGAAAAGT CAACTGATCT TGTTGAAAAGT CAACTGATCT TGTTGAATACT ACCACTAATAGA	GCTTTATTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGCGACG TGCGCTTCCGC TGCGCTTCCGC ACGTCGGTA ACCCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTC AAGTTGCCC CAACAGTC CAACCAAGTC TCAACAAGTC CACCAAGTT TCAGCACTT TCAGCACTT TCATACTTT TCATACTTT	TAACCATTAT AAACCTCTAC TATCCGCCCG TACCTCGCCCG TTCCTCGCTC AACCCAGGAA AGCATCACAA GGTACACA TGTTCCGACC TAGGTCGTC GACCACACA GCTGGTACCA GCTGGTACCA TGTTCCCAAC TAAGGTACAA AGTAAAGTATA AGTAAGTATA AGTAAGTATA AGTAAGTA	AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTGC AGAACATGTC AGAACACGC CTGCCGCTTA GCCACACT ATCCACACT ATCCACACT ATCACACT ATCACACT ATCACACT AGAAGACAC TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACACT TATCACGCA TCGCCAGTTA GATCAAGGCC ACTCCACGCT TAGTATGC TAGTGTATGC TAGTGTATGC TAGTGTATGC TTCGGGGCC CTCGCGGGTCC TATTATTGAA	AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCACCA TATTTGCAAG TATTGCAGA TAAGGGATTT CTTGTCTGA AACTACGATA AACTACGATA ATTACATGA ATTACATGA ATGGCAGAAGGC AGTATACTCA GGGGACCGAG AAAACTCTCA GGGTAGCAAA GCATTTATCA	ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAACCC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCAGATTA CAGCAGATTA CGGAGAGGCT CCGAGCGCAG CAACGTTGTT TCCCCATGT TCCCCCATGT TCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGAACAGGAG GGTTATTGT

### pMT-HisFlag-DmSNAP43 mut#7 STOP

1 TCGCGCGTTT CGGTGAT	CAC CCTCAAAACC	TCTCACACAT	GCAGCTCCCG	GAGACGGTCA	CACCTTCTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGG								
201 CCGCACAGAT GCGTAAG								
301 TACGCCAGCT GGCGAAA								
401 TGAATTAATT CGTTGCA	GGA CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCAC CGCCCAC	CGC CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTGGA	GAA CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAACGA	AAG ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AATCATC								
					> M H	н н н	ннтр	A K D
901 GACGATGACA AGGGCAC	דאם דמאממדמאזד	ATCTTCACC	летестесса	сстсстссал				- 10 -
> D D D K G T			D C W E			R L V N		N C E
1001 TCGAGGTGTT CTGCCGG								
>F E V F C R		L Q L Q			T N H T	E V I	ATT	L A A L
1101 GCATGTGGCC AAGCGAC								
> H V A K R			G D V	F P A S		RIG	G F F L	L Y V
1201 ATCTACTACA AGCAGCO								
> I Y Y K Q P	T H N F	I K I	E V S P	R T W	QEL	T D Y A	L D L	R K D
1301 GTCCGGAGCG GAAGGAC	ACT CATCAGATCG	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
>S P E R K D	T H Q I	A Y M L	WRL	T Q E	QAFR	FTA	L D Y	CQGL
1401 GGACAATCTG GTGGACT	ACG ACCGTGTGGA	GACCGTAGCG	GGTGCCAAGG	AACAGAGGCA	GAGTGCCTTG	ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
> D N L V D			G A K			M O K		N G V
1501 AGTCTCACAT ACGAACT	GGA GGGTCTGCGA					CGGCATACAA	TGCCCAAGCG	GCGGCAGCGG
> S L T Y E L		A L D				A A Y N		A A A
1601 CGGCTGGTCA TGAGCAC							-	
>A A G H E H		S O I F		R E V		O S V	L G A	R K S T
1701 TCCAGATGAG AAATGCA								
> P D E K C			L E V			A M Y	G V E E	R E P
		~		~				
1801 CAACACCAGA CGGATGA								
> Q H Q T D E			N E T Y			A T V F	QRE	L P E
1901 ACGTGCAGCA AGAGTAT	GAG ATGATTGAGT	TTAGTGACGA	CGAGGAAATG	CAACTCCCTC	AAAGCGAGGA	GGTCACGGAA	GAAGAACTCA	λλ <i>СС</i> ТλТТТТ
>D V Q Q E Y		F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT		F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT > D T	TAT AAAACTTTTA	F S D D CATAATTAAA	E E M TAACTAGCAT	E V G TTTTGCGCGA	E S E E TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA	F S D D CATAATTAAA	E E M TAACTAGCAT	E V G TTTTGCGCGA	E S E E TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT > D T	TAT AAAACTTTTA TCG AGTCTAGAGG	F S D D CATAATTAAA GCCCGCGGTT	E E M TAACTAGCAT CGAAGGTAAG	E V G TTTTGCGCGA CCTATCCCTA	E S E E TGTGATCTTG	V T E TTTATCTGAA CGGTCTCGAT	E E L GGGCAATTCT TCTACGCGTA	K A I L GCAGATATCC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT > D T 2101 AGCACAGTGG CGGCCGC	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATCT	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TTC AGGTTCAGGG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TTC AGGTTCAGGG CTT GGCGTAATCA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACASTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGG	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATT TTC AGGTTCAGGG CTT GGCGTAATCA GGT GCCTAATGAG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2701 GCATTAATGA ATCGGCC	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCGGTG AAA TGCTTTATTT TTC AGGTTCAGGG CTT GGCGTAATCA GGT GCCTAATGAG AAC GCGCGGGGAG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGCGG TATCAGC	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TTC AGGTTCAGGG CTT GGCGTAATCA GGCT GCCTAATGAG AAC GCGCGGGGAG TCA CTCAAAGGCG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TCTTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CGTTCGGCTG CAGCAAAAGG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCG 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA ACGCTGG 2701 GCATTAATGA ATGGGC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAACAAG	TCA AAAACTTTAA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT AGGTTCAGGG CTT GGCGTAATCA GGT GCCTAATGAG AAC GCGCGGGGAA TCA CTCAAAGGCG GCC GCGTTGCTGG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCCGGTTTG GTAATACGGT CGTTATTCCA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGAC GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCAGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAG 3001 CCCGACAGGA CTATAAAA	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTAGGG CTT GGCGTAATCA GGT GCCTAATGAG AAC GCGCGGGGAG TCA CTCAAAGGCG GCC GCGTTGCTGG GAT ACCAGGCGTT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAACT TAGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT CGTATTCGTG CACATTAATT CGTATTGGC TATCCACAG AGCTCCCTCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCCTGACG TGCGCTCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCGC TCCCTCGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGGGTA ACTGACTCGC ACTGACTCGC AGAACATGTG AGAACATGTG AAATTCGACGC CTGCCGCTTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	K A I L GCAGATATCC  CCGGTCATCA CAAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAA GTCCGCCTTT
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAAG 3001 CCCGGCAGGG CTATAAAA 3101 CTCCCTTCGG GAAGCGT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTAGGG CTT GGCGTAATCA GGT GCCTAATGAG AAC GCGCGGGGAG TCA CTCAAAGGCG GCC GCGTTGCTGG GAT ACCAGGCGTTG GGC GCTTTCTCAT GGC GCTTTCTCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAAGC TGAGCTAAGT GTAATACGGT CGTTTTTTCCA TCCCCCTCGA AGCTCACGCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCG AGCTCCCCG GTAGGTATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCCGC CAGTTCGGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCGGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGACC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGCG ACTGACTGGA ACTGACTGGAACATGTG AAATCGACGC CTGCCGGCTTA GCTCCAAGCT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATACGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAAGG GGTGGGCAAA GGTCGGCCATT CACGAACCCC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAGG 2701 GCATTAATGA ATCGGCC 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAA 3001 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTTCAGCC CGACCGC	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG CTT GGCGTAATCA GGT GCCTAATCAG TCA TCCAAAGGCG GCC GCGTTGCTGG GAT ACCAGGCGTT TGC GCCTTATCCAG GC GCTTTCTCAT TGC GCCTTATCCAG TCAAAGGCG GCC GCGTTGCTGG GAT ACCAGGCGTT TGC GCCTTATCCG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGA TGGTCATAGC TGAGCTAACT AGCGGTTTG GTAATACGGT TCCTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGGTTTTT TGTTTCCTGT CACATTAATT CGTATTCACAGA TAGCCTCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGGTCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTC TAGGTCGTC TAGGTCGTC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ACTACATCGC AGAACATGTG AAATCGACTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTGTG GCACCAGCCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCC CTGGTAACAG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCAGC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTCAGCC GGACCGC 3301 GATTAGCCAGA GCGAGGT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTAGGG CTT GGCGTAATCA GAAC ACC CCCCGGGGAG ACC CCGTTGCTG GAT ACCAGGCGTT GGC GCTTTCTCAT TGC GCCTTTTCCAT TGC GCCTTTTCCAT TGC GCCTTATCCG ATG TAGGCGTTG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TAGGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTT TACAGAGTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACACTAATT CGTATTGGG TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA AGCCGGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATTCCACA TTTCCACAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACT ACGAGAGCACT ACGAGAGCACACT ACGAGAGCACACT ACGAGAGAGACACACT ACGAGAGACACACT ACGAGAGACACACACT ACGAGAGACACACACACACACACACT ACGAGAGACACACACACACACACACACACACACACACAC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTGG GCGCTGTTGG GCAGCAGCCA TATTTGGTAT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTG CAGCAAAAGG GGTGCGAAAA GTCCGCCTTT CACGAACCCC CTTGGTAACAG CTGCGCTCTG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGGGG TATCAGC 2901 CCAGGAACCG TAAAAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTTCAGCC CGACCGC 3301 GATTAGCAGA GCGAGGT 3401 CTGAAGCCAG TTACCTT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCC AGGTTCAGGG CTT GGCGTAATCA GGT GCCTAATCAG ACC GCGGGGAG TCA CTCAAAGGCG GCC GCGTTGCTGG GCC GCTTTCTCAT GGC GCTTTATCGG ACCTTATCGG ACCTTATCGG ACCGGTGC CGG AAAAAGACTT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TCCCCCTGGA AGCTCGTTTTTCCA AGCTCACGCT GTAACTATCG TACAGAGTTC GTACAGAGTTC GGTAGCTCTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGCTCCGC GTAGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC GATCCGCGCA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCGTTCGGCT ATCAGGGGAT CCCCTGACG ATCAGGGGAT ACCAGGTAACTA ACCACACACA ACAAACCACC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCGCTCA CACTGCCGC CACGCAGA AGCATCACAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT CGGCTACACT CGGCTACACC CCTGGTACCC CCTGGTACCC CCTGGTACCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG ACTGACATGG AAATCGACGC TGACACGTTA GCTCCAAGCT ATCCCACTG ATCCCACTG AGAAGATTTTTT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GCAGCTGCGG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG CCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGCGAA GGTCGCGCTT CACGAACCC CTGGTAACAG CTGCGCTTCACGTTACAG CTGCGCTCTC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA ACGCCTGC 2701 GCATTAATGA ACGCTGC 2701 CCAGGCAGCG TATCAGC 2901 CCAGGCAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTTCAGCC CGACCGC 3301 GATTAGCAGA GCGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG CTT GCGCTAATGAG AAC GCCGCGGGAGA TCA TCCAAAGGCG GCC GCGTTGCTGG GAT ACCAGGCGTT TGC GCCTTATCCG ATG TAGGCGGTGC GCG TAGGCGGTAT TGC GCCTTATCCG GCG AAAAAGAGTT TCT CAAGAAGATT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGAGCTAACT TGAGCTAACT AGCCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGCT TAACTATCG TAACAGGTTC GCAGAGTTACGGTAACTATCG TCACAGCT TCACAGAGTTC TCACAGAGTTC GGTAGCTCTT TCTTGATCTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTAATTAGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAAGTGGT GATCCGCCA TCTTGAGTCC TCTGAAGTGGT TCTTCAGGGG TATCCGCCAA	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTGCCC CAGTTCGGT AACCCGCTAAC GCCTAACA GCCTAACA GCCTAACA CACCGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGGCC TCCCTCGGCC AACGCAGGAA TGTTCCGACC TAGGTCGTC GACACGACT TCGGCTACACT GGCTACACT GGTGGTACGAC AGTTGGAACG AGTTGGAACG AGTTGGAACG AGTTGGAACG	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATGTGGTA CAATTCCACA ATTTCCACA ACTGACTGG ACTACACTGG AGAACATGTG AGAACATGTG GCTCCAAGCT ATCGCCACTTA GGAAGGACAG GTGGTTTTTT AGAAGGACAG GTGGTTTTTT AAACTCACAGT	E L GGGCAATTCT TCTACGCGTA TGACTTTIGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGAATACCT GGGCTCGGT GGGCTCGTGTG GGACTCGTGTG GCACCAGCCA TATTTGCTAT TGTTTGCAAG TAAAGGGATTT	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGCTAACAC CTGGGCTCTG CAGCAGATTA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTTGAAG 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCAGC 2801 CGGCGAGAGCG TATCAGC 2901 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTCAGCA GCAGGGT 3401 CTGAAGCCAG 3301 GATTAGCCAG GCGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3601 ATTATCAAAA AGGATCT	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTTC AGGTTACTATTC AGGTAACAGG GCT GCCGAGGAGAC ACCCGCGGGGAGACACACAGCGT GCC GCGTTGCTGG GAT ACCAGGCGTT GCC GCTTTCTCAT TCC CCCTTATCCG ATG TAGCGGTGC GGAAAAAGAGTT TCT CAAGAAGTC TCA CATAGATCC TCA CATAGATCC TCA CATAGATCC TCA CATAGATCC TCA CATAGATCC TCA ACTAGATCC TCA ACTAGATCCT TCA ACTAGATCC TCA ACTAGATCCT TCA ACTAGATCCT TCA ACTAGATCC TCA ACTAGATCCT TCA TCTAGATCCT TCA TCTAGATCCT TCA TCTAGATCT TCAGATCT TCAGATCT TC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT GTTTTTGCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TAACATATCG TAACATATCG TAACAGGTTC TCTACAGGTTC TCTTGATCTT CTTTGATCTT TTTAAATTAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACCACAGA TAGCGTCAGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAGTGGG ACTCCGCAA TTCTACGGG ATCCGCCAA TTCTACGGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGAAATTGT GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTA ACCACGGTAA ACCACACC TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TTAAATCAAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATTCCACA TTTCCACAGTCG ACTGACTCGC AGAACATGTG CTGCCGCTTA GCTCCAAGCT ATCCCACT GCTCCAAGCT ATCGCCACTG AGAAGGACA GTGGTTTTTT AAACTCACGT TATGAGTAAA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAGGTCAGT AGCAAAAGGC CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTCT TGTTTGCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA GTCCGCCTTT CAGCAACCC CTTGGTAACAG CTGGTAACAG CTGGGCTCTG CAGCAGATTA TGGTCATGATCAG CAGCTTATACAG CAGTTACAG CAGTTACAG CAGTTACAG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGGGG TATCAGC 2901 CCAGGAACCG TAAAAAC 3001 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTCAGGC CGACCGC 3301 GATTAGCAGA GCCAGGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTTAATCA GTGAGGCC	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGGG CTT GGCGTAATCA GGT GCCTAATCAG ACC GCGCGGGAG TCA CTCAAAGGCG GCC GCGTTGCTGG GCC GCTTTCTCAT TGC GCCTTATCCG AGAAAGAGATT TCT CAAGAAGATC TCA CCTAGACCCT ACC TATCTCAC TCA CCTAGACCCT ACC TATCTCAC TCA CCTAGACCT TCA CCTAGACCT TCA CCTAGACCT ACC TATCTCAGCG TATCTCAGCG TTATCTCACCT TCA CCTAGATCCT ACC TATCTCAGCG TATCTCAGCG TATCTCAGCG TTATCTCACCT TTATCT TTATCTCACCT TTA	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TCCCCCTGGA AGCTCACTGG AGCTACT TCCCCCTGGA AGCTACCT TACACAGTTC GTAACTATCG TACACAGTTC TTTGATCTT TTTAAATTAA ATCTGTCTAT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGGCTCCTC GTAGGTACT TCTTGAGTCC TTGAAGTC TTTAAGGGG AATCAGAGA TTTACGGGAACT TCTTACGGGG AATCAGAGT TTCCTTCATC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGT AACCCGGTAA ACCACACTA ACAAACCAC TCTGACGCT CTGACGCT TTAAATCAAT CATAGTTGCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC ACCGGCACA AGCATCACCA AGCATCACCA AGCATCACCA CGCTGCACC CGCTCACCACC CGCTCACACC CGCTGCACCACC AGTGGAACGA CTAAGGTATA TGACTCCCCC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGGTA CAATTCCACA TTTCCAGTCG ACTGACATGG AAACATGG AAACATGG AAACATGG AAACATGG AAACACGCTTA GCTCCAAGCT ATCGCCACTG AGAAGAGACAG GTGGTTTTT AAACTCACGT TAGAGTAAA TCGTGTAGATT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCACCAGCCA TATTTGGTAT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTCTGA AACTACGATA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTCCCAGCT CGTCCGCTT CAGCAAAAGG GCTCGGCTT CACGAACCC CTGGTAACAG CTGCGCTT CAGCAGATTA TGGTCATGAG CAGTTACCAG CAGTTACCAG CAGCAGATTA TGGTCATGAG CAGTTACCAG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTTGAAG 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCAGC 2801 CGGCGAGAGCG TATCAGC 2901 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTCAGCA GCAGGGT 3401 CTGAAGCCAG 3301 GATTAGCCAG GCGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3601 ATTATCAAAA AGGATCT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGGG CTT GGCGTAATCA GGT GCCTAATCAG ACC GCGCGGGAG TCA CTCAAAGGCG GCC GCGTTGCTGG GCC GCTTTCTCAT TGC GCCTTATCCG AGAAAGAGATT TCT CAAGAAGATC TCA CCTAGACCCT ACC TATCTCAC TCA CCTAGACCCT ACC TATCTCAC TCA CCTAGACCT TCA CCTAGACCT TCA CCTAGACCT ACC TATCTCAGCG TATCTCAGCG TTATCTCACCT TCA CCTAGATCCT ACC TATCTCAGCG TATCTCAGCG TATCTCAGCG TTATCTCACCT TTATCT TTATCTCACCT TTA	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TCCCCCTGGA AGCTCACTGG AGCTACT TCCCCCTGGA AGCTACCT TACACAGTTC GTAACTATCG TACACAGTTC TTTGATCTT TTTAAATTAA ATCTGTCTAT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGGCTCCTC GTAGGTACT TCTTGAGTCC TTGAAGTC TTTAAGGGG AATCAGAGA TTTACGGGAACT TCTTACGGGG AATCAGAGT TTCCTTCATC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGT AACCCGGTAA ACCACACTA ACAAACCAC TCTGACGCT CTGACGCT TTAAATCAAT CATAGTTGCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC ACCGGCACA AGCATCACCA AGCATCACCA AGCATCACCA CGCTGCACC CGCTCACCACC CGCTCACACC CGCTGCACCACC AGTGGAACGA CTAAGGTATA TGACTCCCCC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGGTA CAATTCCACA TTTCCAGTCG ACTGACATGG AAACATGG AAACATGG AAACATGG AAACATGG AAACACGCTTA GCTCCAAGCT ATCGCCACTG AGAAGAGACAG GTGGTTTTT AAACTCACGT TAGAGTAAA TCGTGTAGATT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCACCAGCCA TATTTGGTAT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTCTGA AACTACGATA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTCCCAGCT CGTCCGCTT CAGCAAAAGG GCTCGGCTT CACGAACCC CTGGTAACAG CTGCGCTT CAGCAGATTA TGGTCATGAG CAGTTACCAG CAGTTACCAG CAGCAGATTA TGGTCATGAG CAGTTACCAG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGTT 2501 ACCTGCAGGC ATGCAGC 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGGGG TATCAGC 2901 CCAGGAACCG TAAAAAC 3001 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTCAGGC CGACCGC 3301 GATTAGCAGA GCCAGGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTAATCA GTGAGGCC 3701 TGCTTTAATCA GTGAGGCC	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG CTT GCCTAATGAG AAC CCGCGGGGAG TCA CCGAAGGGG GCC GCGTTGCTG GAT ACCAGGCTTCAT TGC GCCTTTCTAT TGC GCCTTATCCG ATG TAGGCGGTGC GG AAAAAGAGTT TCT CAAGAAGAT TCA CCTAGATCGT ACC TATCTCAGCG GGT GCAATGATTAT	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAAACTACGTT TCCACCCTGGA AGCTCACGCT TAACTATCG TAACAGAGTTC GGTAGCTATTC TTTAAATTATA ATCTGTTCTTT TTAAATTAA ATCTGTCTTT CGCGAGACCC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT TGATTCCTGT TATCACAGA TAGCTCCCC AGCTCCCCG GTAGGTATC TCTTGAGTCC TCTTGAGTCC TCTGAGTAC TCTTGAGTC TCTTGAGTC TCTGAGTC TCTTGAGTC TCTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTCATC ACCCCACCA AATGAAGTT TCCTTCATC ACGCTACCC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTAC ACCAGGTAACACCA CCTAACACAC CTTAAACCAC TTAAATCAAT CATAGTTGCC GTTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGTTGCGTC CATAGTTGCC GCTCCAGATT	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGGTCATCAC TAGGTCATC GGCTACACT GGCTACACT GGCTACACT GCTGTAGCG CTAGAGTATAT TGGCTACCCGC TAGGTCACCCCCC TAGGTCCCCCC TAGGTCCCCCC TAGGTCCCCCC TAGGTCCACC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGAA AAATGTGGTA CAATTCCACA ATTTCCACAC ACTACACAC ACTACACAC AGAACATGTG CTCCAAGCT ACTCCAAGCT ACCACCACTG AGAAGGACAG GTGGTTTTT AGAAGGACA GTGGTTTTT TAGAGTAAA TCGTGAAGAA AAACCAAGGA AAACCAAGCA AAACCAAGCA AAACCAAGCA AAACCAAGCA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAAGTCAGA AGCAAAAGGC CCGGATACCT GGCTCGGT AGCAAAAGGC CCGGATACCT GGCTCTGTG GGACTCGGT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGTACCAC CAGCAGATTA
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCGGCC 2701 CCAGGCAGCG TATCAGC 2701 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTTCAGCC CGACCGC 3301 GATTAGCAGC GGAGGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTGG CCCCACTGC	TAT AAAACTTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGTG TTC AGGTTAGTG GCT GCCTAATGAG ACC GCCGGGGGAG ACC GCGTTGGG GAT ACCAGGCGTT TGC GCCTTTCTCAT TGC GCCTTATCCG ATG TAGGGGTG CGAAAAAGAGTT TCT CAAGAAGAT TCT CAAGAAGTCT TCA CTAGATCCT ACC TATCTCAGG CGT GCAATGATAC TAT CCGCCTCAT TCA	F S D D CATAATTAAA  GCCCGGGGTT ATCAGCCTCG GTGAAACTTG GGAGGTGTG TGGTCATAGC TGGGCTAACT AGGCGGTTTG GTAATACGGT TGTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TAACATATCG TAACATATCG TTAACATATCA TTAACATATC GGTAGCTCT TTTTGATCTT TTTTAACTTA ATCTGTCTTAT CCCAGTCTATT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTGGT ATCCGCAA TTCTACGGG AAATGAAGTT TCGTTCATC AAATGAAGTT TCGTTCATC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGGTA AACCAGCT CTGACACC TCTGACG TTTCGGTA TACAAACCAC TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATG GCTCCAGATG TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATT GGGAAGCTAG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAA TGTTCCGACC TAGCTCGTC GACACAGCAC TGGTCGTC GACACGACT GGGTACACT GCTGGTACGC CTGGTACGC TAGATTAT TGACTCCCCC TAGATTAT TGACTCCCCCT TATCAGCATT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATTCCACA TTTCCAGTG ACTGACTCGC AGAACATCTG ACTGACTCGC AGAACATCTG ACTGCCACTTA GCTCCAAGCT ATCGCCACTG AGAAGGACA GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACCAGCCA TAGCTAAA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAAG AACTACGAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAAA GTCCGCTTT CAGCAACCC CTTGGTAACAG TGGTCAGCT CTGGTAACAG CTGGGCTCT CAGCAGATTA CTGGGCAGAT TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTAT
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCAGC 2801 CGGCGAGCGG TATCAGC 2801 CCGGCAGAGCG TATAAAC 3001 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3401 CGGTTCAGCC CGACCGC 3301 GATTAGCAGA GCGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTGG CCCCAGT 3901 AAGTGGTCCT GCAACTT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGTG GCTAATGAG AC GGCGGGAA TCA CTCAAAGGCG GC GCGTTGCTGG GC GCTTTCTCAT TGG GCTTATCCAG GC GCTTATCCAG TCA CCTAAGAGATC TCA CCTAAGAGATC TCA CCTAGATCAT ACC TATCTACAG GCT GCAATGATAC TCA CCTAGATCAT ACC TATCTACAGCG TCAGATCAT ACC TCAGAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TGACCTAACT TCCCCCTGGA AGCTCACGCT GTAACTATC GTAATACGCT TCCCCCTGGA AGCTCACGCT TCACAGGTTC GTAGCTACT TTCAACTATC GTAGCTCTT TTTTAATCTT TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGCTCATTT TCGTCGTTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TATCACAGA AATGAAGTT TTCGTCACC GTAGCTCGCCAC GTAGCTCACCG GAACTCACCG GAACTCACCG GAACTCACCG GAATGTTCACCGGG AATGATTCTCCC GTAGGTACCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GCGAAATTG GCGTTCCGC ATCAGGGGAT CCCCCTGACG CCCCCTGACG ACCAGCACT ACCACACC CCCCGTAA ACCACC TCTGACGTC TCTGACGTC CCCCCGTAA CAAACCACC CCCCCAGAT ACAAACCACC CCCCCCAGAT CCCCCCCAGAT CCCCCCCAGAT CGCACCTCCCCCCCCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCATCAT TTCCTCGCTC ACCGCGC AGCAGGAA AGCATCACA AGCATCACA GCTGCACC TGGTCACCAC TGGTCACCAC GCTGGTACCAC GCTGGTACCA AGTAGACAT TGACTCCCC TATCAGACAT TGACTCCCC TATCAGCAAT TGACTCCCAAC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGAT CAATTCCACA TTTCCAGTCG ACTGACTCGACACTTGA AAATCGACGC TGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGATTTTTTAAATCACGC GTGGTTTTTT AAACTCACGT TATGAGGACAG TGGTTTTTT AAACTCACGT AAACCAGCCA TACGCCACTTA CGCCAGTTA ACCAGCCA TTGCAGAGTAAA TCGTGAGAT AAACCAGCCA TCGCCAGTTA GATCAAGGCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA GCAGCAACAT GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTGGAACACTGGAAGGA TAACGGAATGGAACACACACAACAAACATAACAAAAACAAAAACAAAAAACAAAAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGCTC CACCAAAAGG GGTGGCGAAA GTCCGCCTT CACCAACACC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGAGGCCAG CAAGGTTGTT TCCCCCATGT
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA ACGCCTGC 2701 GCATTAATGA ATCGGCC 2701 CCAGGCAGCG TATCAGC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGCACGG CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTTCAGCC CGACCGC 3301 GATTAGCAG GCAGGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTGG CCCCAGT 3901 AAGTGGTCCT GCAACTT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 TGTGCAAAAA AGGGGTT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG CTT GCCGTAATGAG GCT ACCAGGGGGAG TCA CCCAAAGGGG GCC GCGTGGTGCTG GCC GCTTTCTCAT TGC GCCTTTCTCAT TGC GCCTTATCCG AAAAAGAGTT TCT CAAGAAGATC TCT CAAGAAGATC TCA CTATCAGGG GCT GCAATGATCA TCTCAGCG GCT GCAATGATCA TCT CCGCTCCAT TCT GCGTGCTCACC GCT GCAATGATAC TCC TACTCAGCG GCT GCAATGATAC TCT GCTGCTCCCAC TCT GCTCCTCCGT CCT GCTTCAGCG ACC TCCTTCGGTC	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAAACTACGTT CCCCCTGGA AGCTCACGCT TAACTATCG TAACAGGTTC TACAGAGTTC GGTAGCTATT TTAAATTAA ATCTGTCTAT TGCGCAGACCC CCAGTCTATT TCGTCAGTTT TCGTCAGTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT GAGGTTTTT TGTTTCCTGT TATCCACAGA TAGCCTCCACG AGCTCCCTCG GTAGGTACT TCTTGAGTCC TCTGAAGTGGT GATCCGCA AATTGAGTC AATTGAGTC ACGCTACCG AATTGTTCCC GATTGTTCCC TGTCAGCGCTA CGTATGGCTTC TGTCAGAAGT TGTCAGCGCTT CTGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTACA GCCTAACT GCGCTAACT TGCAGTACACA GCCTAACT TGCAGTACT ACAACCACC TCTGACGTAA GCCTAACT ACAACCACC TCTAAACCACC CTTAAACCACC CTCTAACGTT GAAACCACC CTCAAGGTT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC AAGTTGGCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCA AGCATCACA GGTTCCGACC TAGCTCACA GGTTCCGACC TAGGTCGTC GACACGACT GGCTACACT GGTGGTACCC GCTGAACG CTAAAGTATA TGACTCCCCG TATCACACA GGTTCCAAC CAGGTGAACT CAGCTACACT CGCTACTACC CTATCACCAC CTATCACCAC CTATCACCAC CAGTGTAACT CACCCCC CATCCCCCC CATCCCCC CATCCCCCC CATCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCCC CATCCCCC CATCCCC CATCCCC CATCCCC CATCCCC CATCCC CATCCCC CATCCC CATCC CATCCC CATCC CATCCC CATCCC CATCCC CATCCC CATCCC CATCCC CATCCC CATCCC CATCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATGTGGTA CAATTCCACA ACTGACTGG ACTGACTGG ACTGACTGG ACTGACTGG AGAACATGTG GCTCCAAGCT ATCGCCACTTA GGAAGGACAG GTGGTTTTT AGAAGGACAG TGGTTTTT TAGAGTAAA TCGTGAAGT AAACTCACGT TATGAGTAAA TCGTGTAAGT AAACCAGCA TCGCCAGTTA GAAACGCA CGCCAGTTA GAAACGCA ACTCATGGTT	E L GGGCAATTCT TCTAGGCGTA TGAGTTTIGGA AAACAAGTTA TGGCTGATTA CAACATACGA AGGAAAACGC TCAAGTCAGA AGCAATACGA AGCAATACGA AGCAATACGA AGCAATACGT GGGCTCGGT GGGCTCGGT GGGCTCTGTG GGACACCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA ATGCCAGCAA ATGCCAGCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GCTGGCAAAAGG CTGGCAAAAGG CTGGCAAACG TGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTACACA CTGGTCATG CAGCAGATTA TGGTCATGACA CAGGAGGCT CCGAGGGCCC CCGAGGCCAC CAGGTGTT TCCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCCGCC 2701 GCATTAATGA ATCCGCC 2801 CGGCGAGCGG TATCAAC 3001 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3401 CGGTCAGCA CGACCGC 3301 GATTAGCCAG GCGACGC 3301 GATTAGCCAG AAAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTGG CCCCCGT 3901 AAGTGGTCCT GCAACTT 4001 GCCATTGCTA CAGGCAT 4101 TGTGAAAAA AGCGGTT 4001 GCCATTGCTA CAGGCAT 4101 TGTGCAAAAA AGCGGTT 4201 TCTTACTGTC ATGCCAT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG CTT GGCGTAATCAG GGT GCCTAATCAG AAC CGCGGGGAG TCA CTCAAAGGG GGAT ACCAGGCGTT TGC GCCTTACTCG ATT TAGCGGTTC TCC AAAAAGAGTT TCT CAAGAAGAT TCT CAAGAAGT TCT CAAGAAGT TCT CAGATCCT TCT CAGTCCACT TCCTTCGGTC TCCTTCGTC TCCTTCGGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCGTC TCCTTCCGTC TCCTTCCGTC TCCTTCCGTC TCCTTCCGTC TCCTTCCT	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT GGTTTTGCA TCCCCCTGGA AGCTCACCCT GTAACTATCG TAACATATCG TAACATATCG TAACATATCG TTAACATATCG GTAACTTCT TTTAAATTAA ATCTGTCTTA TCTTTGATCTT TCTTAACTAT CCCAGTCTATT TCGTCGTTTG CTCCGATCGT TCTCTGATCTT CTTCGGTTGG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTAGT TCTTGAGTAC ACTCTACGGG AAATGAAGT TTCTTCACC AAATGATTT TCGTTCACC GAATTGTTCAC GTAGGTACT TTCGTTCACC GAATTGTTCAC GTATGGCTTC TTTCAGAGTTC TTTCACTACAC GAATTGTTCAC GAATTGTTCAC GTATGGCTTC TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA CAAAACCAC TCTGACGCT TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGAT GCCCAGAT ACCAGATC ACCAGATC ACCAGATC ACCAGATC ATCAGCTCC AAGTTGCCC CACCAAGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGCTCGCTC AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGACT GGCTACACT GACAGACT GTTCGTACG CTAGATTAT TGACTCCCAC TAGATTAT TGACTCCCAC TAGTAGGACT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTTAGC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTATC AGTTGTTACC AGTTGTTACC ATTCTGAGAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATTCCACA TATTCCACAC ATTTCCACAC ATTTCCACAC ATTCCACAC ACTGACTCGC AGAACATCTG ACTGACCACT ACTGCCACTC AGAAGGACA TGGTTATTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACCAGCCA TGGTGTAGAT AAACCAGCCA TGATAAA TCGGCAGTTA GATCAAGGCG ACTCATGATT TATGAGTAAT TCGCCAGTTA TATGAGTAAT TCGCAGTTA TATGAGTAAT TCACTGTAATGT TATGAGTATATC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAGAGAACGCC AATTTGGTAT CTTGGTCTGAACTACGATA ACTTACGATA GCCGGAAGGC AATACTTTGCAAG AACTACGATA ACTTACGATA ACTTACGATA ACTTACGATA AACTACGATA ACTACATACATA AACTACATACATA AACTACATACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCAAAA GTCCGCTTT CACGAACCA CTGGTAACAG TGGTCATCA CTGGTAACAG CTGGGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTACTAG CAGAGTTTT TGCTCTTGTTTGCTCTTTGC TGCAGATTTT TGCATAATTC TTGCTCTTTGC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAG 2701 GCATTAATGA ATGGGCC 2701 GCATTAATGA ATGGGCC 2701 CCAGGCAGGG TATCAGC 2701 CCCGACAGGA CTATAAA 3001 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGG 3201 CGGTTCAGCC 3301 GATTAGCAGA GCGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CGCGCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTAG CCCAGGT 3901 AAGTGGTCCT GCAACTT 4001 GCCATTGCTA CAGGCAT 4101 TGTGCAAAAA AGGGGTT 4201 TCTTACTGTA ATGCCAT 4301 CCGGCCCCAA TTACCGTA 4301 CCGGCCTCAA TACCGCAT 4301 CCGGCCTCAA TACCGCAT 4301 CCGGCCTCAA TACCGCAT	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGTG GCTAATGAG ACCGCGGGAA TCA CTCAAAGGCG GC GCGTTACTCAG GC GCTTTCTCAT TGG GCTTATCAG GC GCTTATCCAG AAAAGAGAT TCT CAAGAAGAT TCT CACCTAGATCAT ACC TATCTCAGCG GCT GCAATGATAC CCT GCTTCGATC CCT GGTTCAGCG AGC TCCTTCGGTC CCG TAAGATGCTT TAA TACCCGCCAT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAACT AGGCGGTTTG GTAAACTACGGTTTTCCA TCCCCCTGGA AGCTCACCGGAACTACT TACAGAGTTC TTTAAATTAA ATCTGATCTT TTTAAATTAA TCTGATCTT TTTAATTAT CGCAGAGACCC CCAGTCTATT CGCAGTCTTT CTCGTCGTTTG CTCCGATCGTTC CTCGATCGTTTC CTCGATCGTTC CATAGCAGA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATAATT CGTATTAGGC TATCCACAGA ATTGCCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTGGT TCTTCACGGA AATGAAGTT TCTTCACGGA AATGAAGTT TCCTTCACACA AATGAAGTT TCCTTCACACA TCTACACACA AATGAAGTT CACCCCACACA AATGAAGTT CACCCCAACAC CGTAGGCTACCC TGTAGAGTACT TCTCACACAC CTATGAGAAGT GGTCAGAAGT GGTCAGAAGT CGTCAGAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCCGCTAAC ACCAGTCGGTA ACCAGTAC ACCAGTACAT ACAAACCAA TTAAATCAAT TAAATCAAT CATAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTAC CCTCAGAGTTAC ACTTGACCCT ACTTGACCCT ACTTGACCCT ACTTGACCCC AACTTGCCC AACTTGCCC CAACCAAGTC CAACCAAGTC CCTCATCATT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AACCTCTAC TATCCGCTCA CACTGCCGC TTCCTCGCTC AACCAGGAA AGCATACACA TGTTCCGACC TGCTTCCGACC GCTGCTACACC GCTGCTACACC TGTGCTACCC TAGGTCGTACC TGTGGACCG TCTAGCTCC TGTGACCC TAGACCAT TGACTCCCC TACACCAAT AGACTACAC AGTGTAACT GGTTCCCAAC AGTGTTATC ATTCTGAGAA GGAAACGTT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA ACTACTCGC ACTACACTGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACACTG AGAAGGACAGT ATCGCCACTG AGAAGGACAGT ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGCAGCA TCGCCAGTTA AACCAGCCA TCGCCAGTTA CACCAGTTA CACCAGGGGGCG CTTCGGGGGGG	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCTA TGTTTGCAA TGTTTTGCAA TATTGGTTGA TATGGTCTA ACTACGATA CCGGAAGGC ATACTT CTTGGTCTGA ACTACGATA GCCGAAGGC ATATTTGCTA GCAGCAGCAG ATAGTTACATA ACGACAGAA ACTACCATA ATGCAACACAA AAACTCTCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTCGGCGTAACAG GTCGGCTTT CAGGAACCC CTGGTAACAG CTGGCCTTT CAGGAACCC CTGGTAACAG CTGGCTTTACCAA GCGGAGGCTCTG CGGAGCGCTCT CCGAGCCATGT TCCCCCCATGT TCCCCCATGT TCCCCCCATGT TCCATCATTC AGGATTATTC TTCCTCTTCC AGGATTATTC TTGCTCTTTCC AGGATTATTC TTGCTCTTTCC AGGATTTTC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA ACGCGTGC 2701 GCATTAATGA ATCGGCC 2701 CCAGGCAGCG TATCAGC 2701 CCAGGCAGCG TATCAGC 2701 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTTCAGCC CGACCGC 3301 GATTAGCAGA GCAGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CCGCCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3901 AAGTGGTCCT GCAACTT 4001 GCCATTGCTC ACCACCT 4001 TGTGCAAAAA AGGGGTT 4201 TCTTACTGTC ACGCCAT 4001 CCGGCTCAA AACGGGTT 4201 TCTTACTGTC ATGCCAT 4301 CCGGCGTCAA TACGGGAT 4301 CCGGCGTCAA TACGGGAT 4401 CGCTGTTGAG	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TC AGGTTCAGGG TTT GCCGTAATGAG GCT GCCGTAATGAG ACC GCGCGGGGAG ACC GCGTGCTGCTG GCC GCGTTCTCAT GCC GCTTCTCAT TCG GCCTTATCAG GCT GCCTTATCAG GCT GCCTATCAG ACC AAAAGAGTCT TCC CAAAAAGAGT TCC CAACAAGAGT TCC CAAGAAGAT CCT GCTCTCAGC GCT GCAATGATAC TAT CCGCCTCCAT TAT CCG GCTTCAGC AGC TCCTTCGGTC CCG TAAGATGCT TAT ACCGCGCA ACC TCCTTCGGTC CCG TAAGATGCT TAT ACCGCCCA TCG ATGTAACCCA TCG ATGTAACCA TCG ATGTAACA TCG ATGTAACCA TCG ATGTAAC	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGCCGTTTG GTAAATACGGT TCGTTTTTCCA TCCCCCTGGA AGCTCACGCT TAAACTATCG TAAACAGTT TAAAATATA TCTTTGATATTA ATCTGTATT TTAAATTAA ATCTGTCTAT TCGCCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCGTCGTTTG CTCCGATCGT TTCTTGTGACT CTTCTGTGACT CTTCTGTGACT CTTCTGTGACT CTTCTGCACC CATCATCT CTTCTGACC CATCATCT CTCTGCACC CATCATCT CTCTGCACC CATCATCT CTCTGCACC CATCATCT CTCTGCACC CATCATCT CTCTGCACC CATCATCT CTCTGCACC CTCTTCCACC CCCTCTCCACC CCCTCTCCACC CCTCTCTCCACC CTCTTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCTCTCCACC CTCCTCCACC CTCTCTCCACC CTCTCTCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT GGAGTTTTT TGTTTCCTGT TATCCACAGA TAGCTCCCC AGCTCCCCG GTAGGTACT TCTTGAGTCC TCTTGAGTCC TCTTGAGTC TTCTTGAGTC TTCTTCGCG AATTGAGTC TTCATCACGC AATTGTTCCC GTATGGCTC CGTAGGAAGT TCTTCATCACC TGTCAGAAGT CTTTAAAAGT TTTCTACACGC TGTCAGAAGT CGTAGGATACT CTTTAAAAGT CAACTGATCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GTGAAATTGT GCGTTGGCGT GCTCTTCCGC ATCAGGGGAT GCGCTACCT ACCCGGTAA GCCTAACTA GCCTAACTA GCCTAACTA GCCTAACTA GCCTAACTA GCCTAACTA GCCTAACTA CACCAGGTA TTAAATCAAT CATAGTTGCC CTCCAGAGTT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCG CAACCAAGTC CACCAAGTT TCAGCATCT TTAACATCT CAGCATCT TTAAGCATCT CACCAACTT TCAGCATCT TTAAGCATCT TTAAGCATCT TTAAGCATCT TTAAGCATCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TTAAGCATCTT	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGCTACACA GGTTCCGACC TAGGTCATAT GGCTACACT GGCTACACT GGCTACACT GCTGGTACCG CTAGGTACCG CTAGGTACCG CTAGATATA TGACTCCCCG TATCCACAC TAGGTTATC TATCACACAT TGTTCCCAAC TATCACACAT TGTTCCCAAC TATCACACAT TATCACACAT TATCACACAT TATCACACAC CAGTGTTATC ATTCTGACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATGTGGTA CAATTCCACA TTTCCACACT ACTTCCACA ACTACTACA TATACACACT TATACACACT TATACACT TATACACACT TATACACACC TCGCCAGTTA AAACTCATGGTT TAGTGTATGC CTTCGGGGCC CTTCCGGGCC CTTCCGCACTTTCT CACGCTTTCT CACCACTTTCT CACCACTT CACCACTTTCT CACCACTT CACCACTT CACCACTT CACCACTT CACCACTT CACCACTT CACCACT CACCACTT CACCACT CAC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CACCATACGA GGAAACCTG TCAGGTCAGT AGCAAAAGGC CCGGATACCT GGCTCGGT AGCAAAAGGC GCACCAGCA TATTTGGTAT TGTTTGCAAG TAACGGATACGT AACTACGATA AACTACGATA AACTACGATA AACTACGATA ATGCCAGCA ATGCCAGCA ATGCCAGCA ATGCCAGCA ATGCCAGCA AACTTCCA ACGCCAGA AAAACTCTCA AAAACTCTCA AAGGCAGCAACAAAACTCTCA AAGGCAGCACAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAAACCA TGATCAGTCG CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCAGATTA TGGTCATGAC CAGCAGATTA TCGTCATGAT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCTCATGT TTCCTCTTGC AGGATCTTTC TGGTCTTTC TGGTCTTTC TGGTCATAATTC TTGCTCTTTGC AGGATCTTTC AGGATCTTTAC AAGCAGGAAC AAACAGGAAC
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAAGTGTAA AGCCTGC 2701 GCATTAATGA ATCAGCC 2801 CGGCGAGAGGG TATCAGC 2801 CCGGCAGAGG CTATAAAA 3101 CTCCCTTCGG GAAGCGT 3401 CTGCATCAGC CGACCGC 3301 GATTAGCAG GCGAGGGT 3401 CTGAAGCCAG TTACCTT 3701 TGCTTAATCA GTGAGGC 3801 TACCATCTAG CCCCAGCT 3901 AAGTGGTCT GCACACT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 CGCGCGTAAAA AGCGGTT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 GCCATTGCTA CAGGCAT 4001 GCCGCGTCAA TACCGGAT 4001 CGCGCGTCAA TACCGGAT 4001 CGCGCGTCAA TACCGGAT 4001 CGCGCGTCAA TACCGGCAT 4401 CGCAGATATGCC GCAAAAAG 4501 GCAAAATGCC GCAAAAAG	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTCAGGG CTT GGCGTAATCAG GGT GCCTAATCAG GAC GCCGGGGAG TCA CTCAAAGCGG GAT ACCAGGCGTT TCC GCCTTACTCT GGC GCTTTCTCAT TCC CACAGAGCT TCC AAGAAGACT TCT CAAGAAGAT TCT CAAGAAGAT TCT CAAGAAGT TCT CAAGAAGT TCT CAGATCCT TCT CCTTCGGTC TCCTTCGGTC TCCTTCGTC TCCTTCGGTC TCCTTCGGTC TCCTTCGTC TCCTTCGTC TCTTCGTC T	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTAG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT GGTTTTCCA TCCCCCTGGA AGCTCACCT TAACATACT GGTAGCTCT TTTAAATTAA ATCTGTCTAT TCTTGATCT TCTTGATCTT TCTTGATCTT TCGTGTTTG CCCAGTCTATT TCGTCGTTTG CTCCGATCGT CTCCGATCGT CATAGCAGAA CTCCGGAACA CACAGGAAA CTCGTGCACC CACAGGAAA CTCGTGCACC CACAGGAAA CTCGTGCACC CACAGGAAA CTCGGACAG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCAGC AGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCG AAATGAGTATT TCCTTCAGC AAATGAGTATT TCCTTCACC AAATGAGTACT TTCGTTCATC GTATGGCTTC TTTGAGTACC TTTCAGAGTACT TTCTTCAGT TCTTCAGC TCTTCAGC TCTTCAGC TCTTTAGAGTTC TTTTAGAGTTC TTTTAAAAGT CACTGAGTACT CTTTAAAAGT CAACTGATCT TGTTGAAATAC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTACCG ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCAGCT TCTGACGCT TTAAATCAAT GGCAACCAC GCTCAGCG TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGAT GGCAAGTC CACCAGATT CGCAACTAC TCCACACT CCCAACTC CACCAACT CCCAACTT CCCAACTT TCACACTCT TCAACACT TCACACTT TCACACTCT TCACACTCT TCACACTCT TCACACTCT TCACATCTT TCACACTCTT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAACTGCAA AAATTCCACA AAATTCCACA ATTTCCACAC ATTTCCACACT ACACTCGACACTCCACACCTCACACCTCACACCTCACCCTTA GCACCACCTCACACCTCACCACCTCACCACCTCACCACCTCACCCTCACCCTTACACCAC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAAACCA CTGCTACTAC CAGCAGATTA TGGTCATGACAC CAGTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACTT TCCCCCATGT TGCATATTC TTGATATTC TTGATATTC AAACAGGAAG CAGCTTATCA AACAGGAAG CAGCTTATTTC AGCATTTTAC AAACAGGAAG CAGCTTATTTC CAGCAGATTTAC CAGCAGATTAC CAGCAGATTAC CAGCAGATTAC CAGCAGAGATTTAC CAGCAGAGAGAGTT CCCCCATGT TGCATATTCT CAGCAGCAGAGAGAGAGGAGAG
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT  >D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAC 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAC 2601 TAAAGTGTAA ACGCGTGC 2701 GCATTAATGA ATCGGCC 2701 CCAGGCAGCG TATCAGC 2701 CCAGGCAGCG TATCAGC 2701 CCCGCACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTTCAGCC CGACCGC 3301 GATTAGCAGA GCAGAGGT 3401 CTGAAGCCAG TTACCTT 3501 CCGCCAGAAA AAAAGGA 3601 ATTATCAAAA AGGATCT 3701 TGCTTAATCA GTGAGGC 3901 AAGTGGTCCT GCAACTT 4001 GCCATTGCTC ACCACCT 4001 TGTGCAAAAA AGGGGTT 4201 TCTTACTGTC ACGCCAT 4001 CCGGCTCAA AACGGGTT 4201 TCTTACTGTC ATGCCAT 4301 CCGGCGTCAA TACGGGAT 4301 CCGGCGTCAA TACGGGAT 4401 CGCTGTTGAG	TAT AAAACTTTA TCG AGTCTAGAGG TTT AAACCCGCTG AAA TGCTTTATTT TCT AGGTTAGTG GCTAATGAG GCT GCCTAATGAG ACC GCGCGGGAA TCA CTCAAAGGCG GC GCTTTCTCAT TGG GCTTATCTG GC GCTTATCCAG AATG TAGGCGGTA TCA CCTAGACGT TCA CCTAGACGT TCA CCTAGACGT TCA CCTAGATGAT ACC TATCTCAGCG GCT GCAATGATAC TCA CCTAGATCAT ACC TATCTCAGCG TCA CTAGATGATCA TCC TATCTCAGCG TCA CTAGATGATCA TCC TATCTCAGCG TCA TAGATGATCA TCC TATCTCAGCG TCC TAAGATGATCA TCC TATCTCAGCG TCC TAAGATGATCA TCC TAGATGATCA TCC TAAGATGATCA TCC AAATAAGGGC ATT TGAATGATTT	F S D D CATAATTAAA GCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC GTAAACTTG GTAAACTACGGT TCCCCCTGGA AGCTCACGAA AGCTCACGGA AGCTCATCT TTAAATTAA ATCTAAATTAA ATCTGTCTT TTAAATTAA ATCTGTCTT TGGTGGTTTG CCAGTCTTT CGTGGTTTG CTCCGATCGT TCTTGTGACT TCTTGTGACT TCGTGGTTTG CTCCGATCGT TCTTGTGACC CAGCGACAC GACACGGAAA TCCGTGCACC GACACGGAAAATTA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGCTATT GAGGTTTTT TGTTTCCTGT TATCCAGAG TATCCACAGA TAGGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTGGT TCTTGAGTCC TCAAGTGGT AATCAGCGC AATTGTACT ACGCTCACCG AATTGTTGCC GTAGGGTTACT TCTCAGAAGT TCTCAGAAGT TCTCAGAAGT TCTCAGAAGT TCTCAGAAGT TTCTTCAGTCAT TCGTCAGAAGT TGTGAGAAGT TGTGAGATACT TGTTGAATAC AACAAATAGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCATTATTTG AAAGCAAGTA GCGTTGACTTTCGG ATCAGGGAT CCCCCTGACG GCTCTTCCGC ATCAGGGGAT ACAGCACTA GGCCTAACT ACAGCACTA GGCCTAACT ACAAACAAT ACAAACAAT ACAAACAAT ACAAACAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAACTGCAA AAATTCCACA AAATTCCACA ATTTCCACAC ATTTCCACACT ACACTCGACACTCCACACCTCACACCTCACACCTCACCCTTA GCACCACCTCACACCTCACCACCTCACCACCTCACCACCTCACCCTCACCCTTACACCAC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAAACCA CTGCTACTAC CAGCAGATTA TGGTCATGACAC CAGTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACCA CAGCTTACTT TCCCCCATGT TGCATATTC TTGATATTC TTGATATTC AAACAGGAAG CAGCTTATCA AACAGGAAG CAGCTTATTTC AGCATTTTAC AAACAGGAAG CAGCTTATTTC CAGCAGATTTAC CAGCAGATTAC CAGCAGATTAC CAGCAGATTAC CAGCAGAGATTTAC CAGCAGAGAGAGTT CCCCCATGT TGCATATTCT CAGCAGCAGAGAGAGAGGAGAG

### pMT-HisFlag-DmSNAP43 mut#8 STOP

1 TCGCGCGTTT CGGTGATGAG								
101 TCAGGGCGCG TCAGCGGGTC								
201 CCGCACAGAT GCGTAAGGA								
301 TACGCCAGCT GGCGAAAGGG								
401 TGAATTAATT CGTTGCAGG	CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC CGCCCACCGC	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTGGAGAA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAACGAAAG	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AATCATCTC								
001 111110111010 11110111010	. 0100111011111	110000001110	1110111 00000	1110 111110111 1	> M H	н н н	H H T D	Y K D
901 GACGATGACA AGGGCACTAG	TGAGCTGAAT	ΔΤΟΤΤΤΩΔΟΩ	ACTCCTCCCA	GCTGGTGCA A				
> D D D K G T S		I F D				R L V N		N C E
1001 TCGAGGTGTT CTGCCGGTGC								
					T N H T		ATT	L A A L
1101 GCATGTGGCC AAGCGACTG								
> H V A K R L	SCSR			F P A S		RIG	G F F L	LYV
1201 ATCTACTACA AGCAGCCCA								
> I Y Y K Q P 5					~	T D Y A		R K D
1301 GTCCGGAGCG GAAGGACAC								
>S P E R K D T	H Q I	A Y M L	WRL	T Q E	Q A F R	F T A	L D Y	C Q G L
1401 GGACAATCTG GTGGACTAC	ACCGTGTGGA	GACCGTAGCG	GGTGCCAAGG	AACAGAGGCA	GAGTGCCTTG	ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
> D N L V D Y	D R V E	T V A	G A K	E Q R Q	SAL	M Q K	QQRA	N G V
1501 AGTCTCACAT ACGAACTGGA	GGGTCTGCGA	GCACTGGACC	AGGCAAGCCA	GCCATTGTGT	GAACTGGAAG	CGGCATACAA	TGCCCAAAAG	AAGCAATTGG
SLT YELI	GLR	ALD	0 A S 0	PLC	E L E	AAYN	A O K	K O L
1601 CGGCTGCTGC TGCGGCCGC								GAAAGAGTAC
>A A A A A A A	A P P		G H L			o s v	L G A	RKST
1701 TCCAGATGAG AAATGCACCA		-				_		
> P D E K C T	T T S T		L E V	R O R V		A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGAAC								
> 0 H 0 T D E 1		L E V				A T V F	O R E	L P E
~ ~	~			~			~	
1901 ACGTGCAGCA AGAGTATGAG								
>D V Q Q E Y E	MIE	F S D D		E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTA								
2001 GGATACTTGA GTTAGTTTA: > D T	' AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
2001 GGATACTTGA GTTAGTTTA:  D T  2101 AGCACAGTGG CGGCCGCTCG	AAAACTTTTA AGTCTAGAGG	CATAATTAAA GCCCGCGGTT	TAACTAGCAT	TTTTGCGCGA CCTATCCCTA	TGTGATCTTG ACCCTCTCCT	TTTATCTGAA CGGTCTCGAT	GGGCAATTCT TCTACGCGTA	GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTTTA: > D T	AAAACTTTTA AGTCTAGAGG	CATAATTAAA GCCCGCGGTT	TAACTAGCAT	TTTTGCGCGA CCTATCCCTA	TGTGATCTTG ACCCTCTCCT	TTTATCTGAA CGGTCTCGAT	GGGCAATTCT TCTACGCGTA	GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTTTA:  D T  2101 AGCACAGTGG CGGCCGCTCG	AAAACTTTTA AGTCTAGAGG AAACCCGCTG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA	TGTGATCTTG  ACCCTCTCCT GACATGATAA	TTTATCTGAA CGGTCTCGAT GATACATTGA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA	GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA GTTAGTTTA:  > D T  2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT.	AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA GTTAGTTTA:  > D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA	AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT	TTTTGCGCGA  CCTATCCCTA  CTAAGATCCA  GCTTTATTTG  AAAGCAAGTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATGCA GTGAAAAAA; 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT:	AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA GTTAGTTTA:  D T  2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATGCA GTGAAAAAA; 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAACCT: 2601 TAAAGTGTAA AGCCTGGGG:	AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTTTA:  D T  2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATGCA GTGAAAAAA; 2401 TTGCATTCAT TTTATGTTTX 2501 ACCTGCAGGC ATGCAAGCT; 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGCGGGAG	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGCC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAAAAAA: 2401 TTGCATTCAC TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCCT. 2601 TAAAGTGTAA AGCCTGGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC.	AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGGGGAG CTCAAAGGCG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA	TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	GGGCAATTCT  TCTACGCGTA  TGAGTTTGGA  AAACAAGTTA  TGGCTGATTA  CAACATACGA  GGAAACCTGT  TGGCCTCGGT  AGCAAAAGGC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAGG
2001 GGATACTTGA GTTAGTTTA:  D T  2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT: 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGACTC 2901 CCAGGAACCG TAAAAAGGCC	AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGTTTCCTGT  CACATTAATT  CGTATTGGGC  TATCCACAGA  TAGGCTCCGC	TTTTGCGCGA  CCTATCCCTA  CTAAGATCCA  GCTTTATTTG  AAAGCAAGTA  GTGAAATTGT  GCGTTGCGCT  ATCAGGGGAT  CCCCTGACG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC ACACACATGTG AAATCGACACGC	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTC: 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATCAC 2401 TTGCATTCAT TTTATGTTT: 2501 ACCTGCAGGC ATGCAGCCT: 2601 TAAAGTGTAA AGCCTGCGGG: 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC: 2901 CCAGGAACCG TAAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGA	AAAACTTTA AGTCTAGAGG AAACCGCTG TGCTTATTT AGGTTCAGGG GGCGTAATGAG CCTAAATGAG CCTAAAGGCG GCGTTGCTGG ACCAGGCGTT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCGTTTTTCCA TCCCCTGGA	TAACTAGCAT  CGAAGGTAAG  ACTGTGCCTT  TGATGCTATT  GAGGTTTTTT  TGTTTCCTGT  CACATTAATT  CGTATTGGC  TATCCACAGA  TAGGCTCCGC  AGCTCCTCG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT TCCCCCTGACG TGCGCTCTCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC ACTGACTCGC ACTGACTCGC ACTGCCGCTTA	GGGCAATTCT TCTACGCGTA TCAGTTTGGA AAACAAGTTA TGGCTGATTA GGAAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAAATGCA GTGAAAAAAG 2401 TTGCATTCAC TTTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCCT. 2601 TAAAGTGTAA AGCCTGGGGC 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGGCC 3101 CTCCCTTCGG GAAGCGTGGG	AAAACTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA AGCTCTGGA AGCTCACGCT	TAACTAGCAT  CGAAGGTAAG ACTGTGCTATT GATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTACT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCCC CAGTTCGGTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCACGTCG ACTGACTCGC AGAACATCTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGGT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2001 GGATACTTGA GTTAGTTTA:  D T  2001 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCCAAGCT. 2601 TAAAGTGTAA AGCCTGGGG. 2701 GCATTAATGA TCGGCCAAC 2801 CGGCGAGCGG TATCAGGTCA 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA: 3101 CTCCCTTCGG GAAGCGTGGG. 3201 CCGTTCAGCC CGACCGCTGG	AAAACTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCTGATGAGG CTCAAAGGCG ACAGGCGTT ACGGCTTTCTCAT GCTTTTCTCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGCCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CACATTAATT CGTATTAGGG TATCCACAGA TAGGCTCCCTCG AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	TTTTGCGCGA  CCTATCCCTA CTAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCTGACG TGCGCTTCCGCC CAGTTCGGTG AACCCGGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTCC GACACGACT TAGGTCGTCC GACACGACTT	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGAGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCCACACT ATCCACACT	GGGCAATTCT TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAC CTGGTAACAC
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATCAC 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAGCCT 2601 TAAAGTGTAA AGCCTGGGG: 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC: 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGA 3101 CTCCCTTCGG GAAGCGTGGG 3201 CCGTTCAGCC CGACCGCTGG 3301 GATTAGCAGA GCGAGGTATC	AAAACTTTA AGTCTAGAGG AAACCGCTG TGCTTTATTT AGGTCAGGG GGCGTAATCAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTCTTCTG GCTTTTCTCAT GCTTTTTCAT TGCCGGTGC TAGGCGGTGC	CATAATTAAA GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGACCTAACT TGAGCTAACT GTAATACGGT GTATTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TAACAGAGTTC	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GGTGTTTTT TGTTTCCTGT CGAATTAGTC CACATTAATT TAGGCTCCGC TAGGCTCCCGC GTAGGTATCT TCTTGAGTCC TCTTGAGTCG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTATTTG AAAGCAAGTA GCGAAAATTG GCGTTCGGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTA AACCCGGTAA GGCCTAACTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCCTCGGTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTTC GACACGACTT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATCGC AGAACATCGC CTGCCGCTTA GCTCCAAGCT ATCGCACCT AGAAGCACAC AGAAGACACAC ATCGCACCT AGAAGACACAC ATCGCACCACT AGAAGACACAC AGAAGACACAC ATCGCACCACT AGAAGACACAC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACATTTA CAACATACGA GGAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTGTG GCGCTGTGTG GCACCAGCCA TATTTGGTAT	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTTGGTAACAG CTGCGCTCTG
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAAATGCA GTGAAAAAAG. 2401 TTGCATTCAC TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCC. 2601 TAAAGTGTAA AGCCTGGGGG. 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGGGG TATCAGCTC. 2901 CCAGGAAACCG TAAAAAAGGCC. 3001 CCCGCACAGGA CTATAAAAAG 3101 CTCCCTTCGG GAAGCGTGGG 3201 CGGTTAAGCA GCGAGGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3401 CTGAAGCCAG TTACCTTCGG 3401 CTGAAGCCAG TTACCTTCGG	AAAACTTTTA AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG CGCGGGGAG CTCAAAGGCG CGCGTTGCTGG ACCAGGCGTTGCTGG ACCAGGCGTTGCTGG TGCTTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAAGAGTT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACTATAGC GTTTTTTCCA TCCCCCTGGA AGCTCACCG GTAACTATCG TACAGGAGTTC GGTAGCTCTT GGTAGCTCTC GGTAGCTCTT	TAACTAGCAT  CGAAGGTAAG ACTGTGCTAT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTGAAGTGCT TGTAAGTGC GATCCGCCAA	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTA GTGAAATTTG GCGTTGCGCT GCTCTTCGCG ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTA AACCCAGTAACTA ACCAACCACC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTCAT TATCCGCTCA CACTGCCGCC AACGCAGGAA AGCATCACAA TGTTCCGACT TGTTCCGAC TAGGTCGTT GACACGACTT GCGCTACACT GCTGCTACCT GCTGCTACCC GCTGCTACCC GCTGCTACCC GCTGCTACCC GCTGCTACCC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ATCCAGCCA AGACATGTGA AAATGAGGCA CTGCCGGCTTA GCTCCAAGCT ATCGCACTGA AGAGGACAG GTGGTTTTTT	GGGCAATTCT TCTACGCGTA TGAGTTTGA AAACAGTTA CGACTATA CAACATTA CAACATCGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAAGG GGTGGCGAAA GCAGCACTT CACGAACCCC CTGGTAACAG CTGCGCTTC CAGCAACTC CAGCACATTA
2001 GGATACTTGA GTTAGTTTA:  D T  1011 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAA; 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT. 2601 TAAAGTGTAA AGCCTGGGGC 2701 GCATTAATGA ATCGGCCAAG 2801 CGGCGAGGGG TATCAGGTC; 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGGC 3101 CTCCCTTCGG GAAGCGTGGG 3201 CGGTTCAGCC CGACCGCTGG 3301 GATTAGCAGA GCGAGGTTAT 401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATC	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGTTTAGAGG GGCGTAATCA GCCTAATGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAAAGAGTC CAAGAAGATC	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCT TCCAGTTCCC TCTGCAGCTC TCTTGATCCC TTTGATCCC TTTTGATCT CTTTGATCT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCGGCAA TTCTTCGGGGA TTCTTCGGGGA	TTTTGCGCGA  CCTATCCCTA CTAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC TCCGCTGACG TGCGCTCTCCGC ACCCCTGACG TGCGCTCTCCGTG AACCCGCTAA GGCCTAACTA ACCACCC TCTGACGCTCTCCC TCTGACGTCGACC TCTGACCGTCACC TCTGACCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTTAC CATGCCGC TTCCTCGCTC ACCGGGAA AGCATCACAA TGTTCCGACT TAGGTCGTTC GACACGACT GCTGCTTC GACACGACT GCTGCTACACT GCTGGTACACT AGTGGTACACT AGTGGAACGA AGTGGAACGA AGTGGAACGA AGTGGAACGA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGAGTA AACTGCACA CTCCACACT ACTCCACACT AAATCGACGC CTGCCGCTTA AATCGACGC ACTCAAGCT ATCCCACT AGAAGACACT ACCCACTT AGAAGACACT ACCCACTT AGAAGACACAG GTGGTTTTTT AAACTCACGT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGCCTT CAGCAAAAGG GTCGCCTTT CACGACCCC CTGGTAACAG CTCGCCTTT TGGTCATGAT TGGTCATGAC
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT: 2301 CTAGAATCAC 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCACCCC 2601 TAAAGTGTAA AGCCTGCGG: 2701 GCATTAATGA ATCCGCCCAC 2801 CGGCGAGCGG TATCAAGCTC 2901 CCAGGAACCG TAAAAAAGGCTC 3001 CCCGACAGGA CTATAAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATC 3401 CTGAAGCCAG TTACCTTCGC 3501 CGCGCAGAAA AAAAGGATCTCC 3601 ATTATCAAAA AGGATCTTCC 3601 ATTATCAAAA AGGATCTTCC	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTTA AGGTTCAGGG GGCGTAATCA GCCTAATCAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTTCTTG GCTTTCTCAT GCCTTATCCAT TGCCGGTGC TAGGCGGTGC TAGGCGGTGC CAAAAGAGTT CAAAAAGACTT CCTAGATCCT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTAACC TGACCTAACC TGACCTAACC TCGCTTTTTCCA TCCCCCTGGA AGCTCACCGT GTAACTATCG TACAGAGTTC GTAACTATCC GTAACTATCC GTAACTATCC TTCAGAGTTC TCTTGAACTT CTTTGATCTT TTTTAAATTAA	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTATT TGTTTCCTGT CGAATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCTC GTAGGTATCT CTTTGAGTCC TCTTGAGTCC TTTGAGTCG GATCCGCCAA TCCCCCAG AATCAACGGAA TCCACAGA	CTATCCCTA CTAAGATCCA CCTATATTG CAAGACTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG TGCGCTCTCCC CAGTTCGGTA AACCCACGGTAA GCCCTAACTA ACAAACCAC TCTGACGCT TCTGACGCT TCTGACGCT TCTAAATCAAT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT AAACCTTAC TATCCGCTCA CACTGCCGC ACCGGCTAACCAGAA TGTTCCAGCT TAGTTCCAGC TAGGTCGTTC GACACGACTAC GACACGACTAC GGCTACACT GGCTACACT GCTGGTACGC CTAGATGTAC CTAGATGTAC CTAGATGTATA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTA CAATTCCAGTCG AGAACATCGCA AGAACATCGC AGAACATCGC CTGCCGCTTA GCTCCAAGGT ATCGCAGCT AGAAGAGACAGC AGGAGAGACAGC ATCGCTTA ACGCCACTG AGAAGACATTTTT AAACTCACGT TATGAGTAAA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACATTTG CGACTATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACCT GGCTGTGTG GCACCAGCCA ATATTTGGTAT TGTTTGCAAG TAAGGGATTT	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGCAACCC CTTGGTAACAG CTGGGCTCTC CAGCAGATTA TGGTCATGACAG CTGGGCTCTG CAGCAGATTA TGGTCATGACAC CAGTTACCAA
2001 GGATACTTGA GTTAGTTTA:  D T  101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAAG. 2401 TTGCATTCAC TTTATGTTC 2501 ACCTGCAGGC ATGCAAGCC. 2601 TAAAGTGTAA AGCCTGGGGC 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGGGG TATCAGCTC. 2901 CCAGGAAACCG TAAAAAAGGCC. 3001 CCCGCACAGGA CTATAAAAGGA 3101 CTCCCTTCGG GAAGCGTGGG 3201 CGGTCAGGC GGAGCGCTGG 3301 GATTAGCAGA GCGAGGGTAC 3401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCG 3701 TGCTTAATCA GTGAGGCACC	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG CGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTG GCTTTCTCAT GCCTTATCCG AAAAAGAGTT CAAGAAGATC CCCTAGATCACT TATCTCACG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAAGC TGAGCTATAGC TGATATACGT CGTTTTTCCA TCCCCCTGGA AGCTACCT GTAACTATCG TACAGAGTTC GGTAGCTCT CTTTGATCTT TTTAAATTAA	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGC TTGAAGTGCT TTGAGGCCAT TCTTACGGGG AATCACAGATT TTCCTTCACTCC TTCTACGGGG TTCTACGGGG TTCTACGGGGT TTCTACTCATC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGCCCCGCA ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGGTA ACCACCAC ACCAACCA ACCAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC CATTGCTCC CAGTTCGTT ACCACCAC TCTGACGCTC TCTGACGCTC TTAAATCAAT CATAGTTGCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGCTC AACGCAGGAA AGCATCACAA TGTTCCGAC TGTGTCGAC TGGCTGC GCTGGTACGC GCTGGTACGC GCTGGTACGC CTAGGTACGA TGAGTACCA TGGACCGC TGGAACGA TTAAGGTATA TGACTCCCCC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCG AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGT ACTGCCACTTA GCTCCAGCGTA ACTGCCACTG AGAAGACATGT ATCGCACTG AGAAGACATGT ATCACCACTG TATGACTACTT TATGAGTAAA TCGTGTAGAT TCGTGTAGAT	GGGCAATTCT TCTACGCGTA TGAGTTTGA AAACAGTTA CAACATTA CAACATAC TGCCGGATA CACATACGA GGAAAACTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTCCAAG TAGGGATTT CTTGGTATT CTTGGTATT CTTGGTATT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAGG GGTGGCGAA GCACAAAAGG GCTGCGCCTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGAG CAGTTACCAG CAGTTACCAG CAGTTACCAG CGGGAGGGCT
2001 GGATACTTGA GTTAGTTTA:  D T  1011 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAA; 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT. 2601 TAAAGTGTAA AGCCTGGGGC 2701 GCATTAATGA ATCGGCCAAG 2801 CGGCGAGGGG TATCAGGTC; 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGGC; 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGGC CGACCGCTG; 3301 GATTAGCAGA GCGAGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT; 3601 ATTATCAAAA AGGATCTTC; 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCC 3801 TACCATCTGG CCCCAGTGCC 3801 TACCATCTGG CCCCAGTGCC	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT TGGTTTCTCAT GCCTTATCCG TAGGCGGTGC CAAAAAAGACT CAAGAAGATC CCTAGATCCT TATCTCAGC GCATGTTCAGCG GCATGTTCAGCG GCATGTTA	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC CTTAGACTATC TTTAAATTAA ATCTGTTCTT CCCCAGAACCC CCCAGAACCC CCCCAGAACCC CCCCAGAACCC CCCCAGAACCC CCCCAGAACCC CCCCAGAACCC CCCCAGAACCC CCCCAGAACTTC CCCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACCC CCCAGAACC CCCAGAACCC CCCAGAACC CCCAGAACCC CCCAGAACC CCCAGAACCC CCCACAACCC CCCACAACCC CCCACAACCC CCCACAACCC CCCACAACCC CCCACAACC CCCACAACCC CC	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TCTTCAGGGCA ATTCACGGGA AATGAAGTT TCTTACGGGGA TTCTACGGGG AATGAAGTT TCTTCACTCACG	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCTGACG TGCGCTCTCGTG AACCCGGTAA GGCCTAACTA ACCAGCTACT TCTGACGCT TTAAATCAAT CATAGTTGCC CTCTGACGCT CTGACGCTC CTCTGACGCT CTCTGACGCTC CTCTGACGCTC CTCTGACGCTC CTCAAGTTGCC CCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTATC CATTGCCCGC TTCCTCGCTC ACCTGCCCGC TTCCTCGCTC TGCTTCCGACC TGGTTCCGACC GACAGGAA AGCATCACAA TGTTTCCGACC GACTACACT GCTGCTACACT GCTGCTACACG AGTGGAACGA CTAAAGTATA TGACTCCCCC TATCAGCATT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGAGTA AACTGCACA CTGCACTGC ACTGACTGC ACTGACTGC CTGCCGCTTA AAATCGACGC ACTGACCGCTTA AGAAGACATTTTTT AAACTCACGT TATGAGTAAA CCACGTT ATCGCGTTA AGAGGACAG ATCGCCACTG AGAAGACAGT ATCGCCACTG AGAAGACAGCA ATCGCACTG AGAAGACAGCA AAACCACGCA AAACCACGCA AAACCACGCA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGCAGACCA TCAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA ACTACGATA GCCGGAAGGG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGCTTC CAGCAAAAGG GCTGGCAAA GTCCGCCTTT CAGGAACCCC CTGGTAACAG CTGGCTATACCAA TGGTCATGAG CAGGAGATTA TGGTCATGAG CAGGAGGGCT CCGAGGCGCCCC CCGAGCGCAC
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCTTACTGG ACCAGGGGAG CTCAAAGGCG CTCTATCTG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG AAAAGAGATC CATAAAGAGATC TATCTCAGCG GCATTCACGG CCTAGATCCT TATCTCAGCG CCTAGATCCT CATAACACCT CATAACACCT CATAACACCT CATAACACCT CATAACACCT CATAACACCT CATAACACCT CATAACAATGATAC CCGCCTCCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGCCTATAGC TGACCTAAC TGACCTAAC TCCCCCTGGA AGCTCACCT GTAACACTT GTAACACTT CGTACACTATC GTACACTATC GTACCTCT TTTTAAATTAA ATCTGTCTAT TTTTAAATTAA ATCTGTCTAT	TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTACTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGGTCCGC GTAGGTACT TCTTGAGTCC TTGAAGTGGT GATCCGGCAA TTCTACGGGG AATTGAATT TCCGTCACCACACACACACACACACACACACACACACACA	CCTATCCCTA CTAAGATCCA CCTATACTCA CTAAGATCA CCTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTCCCCTGACG ATCAGGGGAT CCCCCTGACG AACCACC CAGTTCGGTC AACCAACTA ACAAACCAC TCTGACGTC TCTGACGTC TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTTAC TATCCGCTC CACTGCCGC TCCTCGGTC AACGCAGAA AGCATCACAA GCATCACAC TGTTCCGACC TGGTCGACC GCTGGTACGCAC GCTGGTACGC AGTGGAACGA AGTAGAACT TGACCCCCG TAAGTATT TGACTCCCCCC TATCAGCAT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTA CAATTCCACA TTTCCAGTCG ACTGCACTCGACA CTGCCGCTTA GCTCCAAGCT ATCCCAACG AGAACATCTT ACTCCAACG TTCCAACGT ATCCCACT ATCCCACT TTCAGTCA TTCAGTCA TTCAGTCA TTCAGTCA TTCAGAGT ATCAGAGT ATCAGAGT AAACTCACGT TATCAGATAA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACA GGAAAACCTGT TGCCGTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGCGGATACCT GGCGTGTGG CCAGCACCA TATTTGGTAT TGTTTGCAAG TAAGGATTCT TGTGTCTGA AACTACGATA ACTACGATA ACTACGATA GCCGGAAGGG ATAGTTTTGCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCAGCT CAGCAAAAAGG GGTGGGCAAA GCACCCC CTGGTAACAG CTGCGCTT CACCAACACCC CTGGTAACAG CTGGCTCTC CAGCACATTA TGGTCATGACA CGGAGAGTACAGA CAGATTACAA CGGAAGGCCAC CCGAGCGCCAC CAGTTACCAA CAGGAGGCCAC CAGATGCACAC CAAACGTTGTT
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG GCGTGCTGCTG ACCAGGGGAG CTCAAAGGCG TGCTTTCTCAT GCCTTATCCG AAAAAGAGTT CAAGAGATC CCCTAGATCAG GCAATGATTAC CCTAGATCAC GCAATGATTAC CCGCCCCCCAT GCGTCTCAGG GCAATGATAC GCGTCTCAGG GGAATGATAC CCGCCCTCAGG	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAAGC TGAGCTAAGC GGTTTTTCCA TCCCCCCGGAA AGCTACCT GTAACTATCG TACAGAGTTC GGTAGCTCT TTTAAATTAA ATCTATCT TTTAAATTA ATCTATCT	CGAAGGTAGA ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TTGAAGTGC TAGACTGCGCA TTCTACGGGG AATCACAGA TTCTACGGGG AATCACCG TTCTACTCATC ACGCTCACCG GATGTTCATC CGGTCACCG GATGTTTCGTCCC GTATGGCTCC	TTTTGCGCGA  CCTATCCCTA CTAAGATCA GCTTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAATTGCCG ATCAGGGGAT CCCCTGAACG ACCAGTTCGGTC AACCAGCTAACTA ACAAACCAC TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCAGAGGTC GCTCAGAGGTC GCTCAGAGGTA CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGGTCC ATTCAGCTCC ATTCAGATTGCC GCGCAGATT GGGAAGCTAG ATTCAGGTCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC CACTGCCGCC AACGCAGAA AGCATCACAA TGTTCCGACT TGTTCCGAC TGGTCGCT GACACGACT CGCTTACCAC CTGGTACCA CTGGTACCA AGTAGACGA TTAAGTATA TGACTCCCCC TATCAGCAT TGACTCCCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA AATCCAGCG AGAACATGTG AAATCGACGCA AGACATGTG AAATCGACGT ACTGCCACTG GCTCCAGCTT GCTCCAGCTT ACTGCACTT AAACCAGCCA TCGCCACTTA GATCAAGCCG	GGGCAATTCT TCTACGCGTA TGAGTTTGA AAACAGTTA CAACATTA CAACATAC TGCCGGATTA CAACACTAC AGCAAAACTGA TGCCGCTCGGT AGCAAAACGG TCAAGTCAGA CCGGATACCG GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTCCAAG TAGGGATTT CTTGGTAT CTTGGTAT CTTGGTAT CTTGGTAT CTTGGTAT CTTGGTAT CTTGGTAT CTGGGAAGGG AACTACGATA CCGGAAGGG ATACTTTCCA AGTTACATGA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAGG GGTGGCGAA GCACAAAAGG GCTGGCCTT CACGAACCCC CTGGTAACAG CTGCGCTCT CAGCAACTT CAGCAGATTA TGGTCATGAG CAGCTCT CAGGAGGCCAC CCGGAGGGCC CCGGAGGCCAC CCAGTGTT CCCCCCATGT
2001 GGATACTTGA GTTAGTTTA:  D T  1011 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT. 2601 TAAAGTGTAA AGCCTGGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGGTC. 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGAG. 3011 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGCA CGACCGCTGC 3301 GATTAGCAGA CGAGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCTC 3601 ATTATCAAAA AGGATCTTCC 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCT 3901 AAGTGGTCCT GCAACTTTAC 4001 GCCATTGCTA CAGGCATCG 4011 TGTGCAAAAA AGCGGTTAGC	AAAACTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGGGGAG CTCAAAGGCG CTCAAAGGCG CGTTTCTCAT GCCTTATCCG TAGGCGTTGC AAAAAGAGTT CAAGAAGAT CCTAAGATCCT TATCTCAGGG GCAATGATAC CCGCCTCCAT GGGTTCAGGC TCCTTCAGGC TCCTTCAGGC TCCTTCAGGC TCCTTCAGGC TCCTTCAGGC TCCTTCAGT	CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCAGA ATCCGCTCTATT TCGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTCCCATCGT CTCCCATCGT CTCCCATCGT CTCCCATCGT CTCCCATCGT CTCCCATCCT CTCCCATCCT CTCCCATCCT CTCCCATCCT CTCCCATCT CTCCCATCCT CTCCATCCT CTCCCATCCT CTCCCATCCT CTCCCATCCT CTCCCATCCT CTCCCATCCT CTC	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGATTCCTGT CGCATTAGGC TATCCACAGA TAGGCTCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGCT TCTACGGCA ATTCACGCAAATTATT TCTTAGGTC TTGAGTCC TTGAAGTGGT TCTTCACGCAATTCACTCACCAATTCACTCACCACAATTCATCACCCACAATTCATCA	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCCTGACG TCGGCTCTCCGC ACCGTTACGGT AACCCGGTAA GGCCTAACT ACCAGTTCGGTG AACCAGTACAT ACAAACCAAT TCTGACGCT TTAAATCAAT TAAATCAAT CATGTTGCGCG GCTCCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC AAGTTGGCCC AAGTTGGCCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTATC CATTGCCGC TTCCTCGCTC ACCTGCCCG TTCCTCGCTC ACCTGCACA AGCATCACAA GCATCACAA TGTTCCGACC GCTGCACACT GCTGGTACACT GCTGGTACAC AGTGGACGA AGTAAGTATA TGACTCCCCC TATCAGCAAT AGTAAGTATA GGTAACGA CAGTGTTATC CACTTTATC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTGTA AAATCACTCGAC ACTGACTGC AGAACATTCACCGC AGAACATGTG AAATCGACGC ATCGCCACTG AGAAGACATGT ACTCACCGTTA AGAAGACAGC ATCGCCACTG AGAAGACACT ATCGCCACTG AGAAGACACT ATCGCACTG AGAAGACACG ATCAGAGTAAA AAACCAGCA TCGCCAGTTA GATCAAGCC ATCAGCAGTTA GATCAAAGCC ACTCATGGTT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTCGA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAAA GTCCGCCTTT CAGCACACCC CTGGTAACAG CTGGCTCTG CAGCACATTA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGCAGGTTGT CCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG GCTTACTTG ACCAGGGGAG CTCAAAGGCG TCCTTATCTG TAGGCGGTC TAGGCGGTC TAGGCGGTC TAAGAAGAGTT CAAAAAGAGT CCTAACATCC TTATCCAG GCAATGATAC CCTACATCC TATCTCAGCG TCAGATGATC TTATCTCAGCG GCAATGATAC CCGCCTCCAT GGTGTCAGCG TCCTTCGGTC TAAGATGCTT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAACT GTAATACCGT CGTTTTTCCA AGCTCACTATC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG TTCTGTCGTTTG TTCTGTCACT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT GAGGTTTTTT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCAACAGA TCCTACGGCAA TTCTACGGCAA TTCTACGGCA ATTCTTACT ACGCTCACCG AATTGATTCC CGAATTGTTCCC GTATGGCTTCCTCGCAGAGT TTCGAGAGT TTCGAGAGT TGTCAGAGGT GGTGAGTACT	CCTATCCCTA CTAAGATCCA CCTTATTTG CTAAGATCA GCTTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTCCCCCTGACG ATCAGGGGAT CCCCCTGACG GGCCTAACTA ACAAACCAC TCTGACGTC TCTGACGTC TTAAATCAAT CATAGTTCGC GCTCAGATT GGGAAGCTAG ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC CAGATT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGGTCAACGCAGCA AGCATCACAA AGCATCACAA AGCATCACAC TGTTCCCAGC TAGGTCGTC GACACGACTT GACACGACTT GACACACT CTAGATACC CTAGATACC CTAGATACC CTAGATACC CTAGATACC CTAGACACT TATCAGCAC TATCAGCACT TATCAGCAC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTATC CATTCTAGAAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATCTG AAATCTGAGCG AGAACATCTG AAATCTGAGCG AGAACATCTG AAATCAGCT ATCGCAGCTTA GCTCCAAGGT ATCGCCACTG TATCAGCTAAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGGT AAACTCAGGT AAACTCAGGT ACGCCAGTTA GATCAAGGGG ACTCATGAGT TAGTGTAGGT TAGTGTAGT TAGTGTATATG	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACA GGAAAACTGT TGCCGTCGGT AGCAAAAAGC TCAAGTCAGA CCGGATAACT GGCGGATACT GGCGCTGGTG GCAGCAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA CTTTGGTAT GCTACGATA GCCGAAGGG AACTTGCATA GCCGAAGGG ATACTTTGCAA AACTACATA GCCGAAGGG ATACTTTGCAA AACTACATA GCCGAAGGG ATACTTTGCC AGTTACATA ACGCACCAC GGCCACCCAG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAAGG GGTCGGCATT CAGCAAAAAGG GCTCGGCTT CAGCAACACC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGACAG CAGTTACAGA CAGTTACAGA CAGTTACAGA CAGTTACTAGA CAGTTACCAA CGGGAGGCT TCCCGACTGT TCCCCCATGT TTCCCCCATGT TTGCTCTTTGC
2001 GGATACTTGA GTTAGTTTA:  D T  1011 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT. 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT. 2601 TAAAGTGTAA AGCCTGGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGGTC. 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGAG. 3011 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGCA CGACCGCTGC 3301 GATTAGCAGA CGAGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCTC 3601 ATTATCAAAA AGGATCTTCC 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCT 3901 AAGTGGTCCT GCAACTTTAC 4001 GCCATTGCTA CAGGCATCG 4011 TGTGCAAAAA AGCGGTTAGC	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG GCTTACTTG ACCAGGGGAG CTCAAAGGCG TCCTTATCTG TAGGCGGTC TAGGCGGTC TAGGCGGTC TAAGAAGAGTT CAAAAAGAGT CCTAACATCC TTATCCAG GCAATGATAC CCTACATCC TATCTCAGCG TCAGATGATC TTATCTCAGCG GCAATGATAC CCGCCTCCAT GGTGTCAGCG TCCTTCGGTC TAAGATGCTT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAACT GTAATACCGT CGTTTTTCCA AGCTCACTATC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG TTCTGTCGTTTG TTCTGTCACT	CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT GAGGTTTTTT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCAACAGA TCCTACGGCAA TTCTACGGCAA TTCTACGGCA ATTCTTACT ACGCTCACCG AATTGATTCC CGAATTGTTCCC GTATGGCTTCCTCGCAGAGT TTCGAGAGT TTCGAGAGT TGTCAGAGGT GGTGAGTACT	CCTATCCCTA CTAAGATCCA CCTTATTTG CTAAGATCA GCTTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTCCCCCTGACG ATCAGGGGAT CCCCCTGACG GGCCTAACTA ACAAACCAC TCTGACGTC TCTGACGTC TTAAATCAAT CATAGTTCGC GCTCAGATT GGGAAGCTAG ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC ATCAGCTTC CAGATT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA ACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCCGAC TAGGTCGTC GACACGACTT GACACGACTT GACACGACTT CGGCTACACT GCTGGTACGC CTAAAGTAT TGACTCCCCG TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT CATTCCCAAC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTATC CAGTTTATC CAGTTATC CAGTTTATC CATTCTAGAAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATCTG AAATCTGAGCG AGAACATCTG AAATCTGAGCG AGAACATCTG AAATCAGCT ATCGCAGCTTA GCTCCAAGGT ATCGCCACTG TATCAGCTAAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGGT AAACTCAGGT AAACTCAGGT ACGCCAGTTA GATCAAGGGG ACTCATGAGT TAGTGTAGGT TAGTGTAGT TAGTGTATATG	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACA GGAAAACTGT TGCCGTCGGT AGCAAAAAGC TCAAGTCAGA CCGGATAACT GGCGGATACT GGCGCTGGTG GCAGCAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA CTTTGGTAT GCTACGATA GCCGAAGGG AACTTGCATA GCCGAAGGG ATACTTTGCAA AACTACATA GCCGAAGGG ATACTTTGCAA AACTACATA GCCGAAGGG ATACTTTGCC AGTTACATA ACGCACCAC GGCCACCCAG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAAGG GGTCGGCATT CACGAACACCA CTGGTAACAG CTGCGCTCT CAGCAAAATA TGGTCATACAG CAGCATTACAG CAGCATTACAG CAGCATTACAG CAGCTTACAGA CAGCTTACCAG CAAGTTACAGA CAGTTACCAA CGGGAGGCT TCCCGACTGT TCCCCCATGT TCCCCCATGT TGCATATATC TTGCTCTTTGC
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG GCTTACTGT GCTTTCTCAT GCCTTATCTCAT GCCTTATCCAT GCCTTATCCAT GCCTTATCCAT GCCTTATCCAG CGCAGAGAGATC TACTCAGAGAGATC TACTCAGCG GCAATGATAC GCGCTCCAT GGTGTCAGGC TCCTTCGGTC TACCTCAGGT TACCTCAGGT TACCTCAGGT TACCTCAGGT TACCTCAGGT TACCTCAGT TACCTCAGGT TACAGAGATC TACCTCAGT TACCTCAGT TACCCAGGT TAAGATAGTT TACCGCCCAT	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAATC GTAATACGT CGTTATACCGT GGTCATAGC GTTATTCCGA AGCTCACT GTAACTACG TACAGAGTTC GTAACTATCG TACAGAGTTC TTTGATCTT TTTGATCTT TTTAATTTA ATCTGTCTAT CGCAGAGCCC CCAGCCTATT TCGTCGTTTG CTCGGATCGT TTCTGTGCTT CTTCGATCGT CTCGATCGTTTC CTCGATCGTTTC CATAGCAGAA	CGAAGGTAGA ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTTACGGGG AATCATAATT TTCGTTCATC ACGCTCACCG GTAGGTATCT TCGTTCATC ACGCTCACCG GATTGTTCATC ACGCTCACCG GTAGGTATCT TCTCAGAGAGT TCTTCAGAGAGT TCTTCAGAGAGT CTTTGAAAGT TGTCAGAAGT TCTTTAAAAGT CTTTTAAAAGT CTTTTAAAAGT CTTTTAAAAGT	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGAACG ACCAGTAACTA ACCACACTA ACCACACTA ACCACACTA ACAACCAC TCTGACGTC CTCAACTA CATAGTTGCC GCTCCAGATT GGGAACTA CATAGTTGCC AGTTGGCGT ATTCAGCTC AGTTGCCGCAATT ACCACACTC AGTTGCCC CACCACTC CCCACATT ACCACACTC CCCACACTC CCCACACTC CACCACTC CACCACTC CACCACTC CACCACTC CACCACTC CACCACTC CACCACTC CACCACTC CCTCATCATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAAT TATCCGCTCA CACTGCCGCC ACCGGCAGAA AGCATCACAA AGCATCACAC TGTTCCCAGC TGGTCGCTC GACACGACTT GACACGACTT GACACGACTT TGGCTACACT TGGCTACACT AGTAAGTAAT TGACTCCCCG TATCAGCAAT TGACTCCCCG TATCAGCAAT AGTAAGTAAT GGTTCCCAAC CAGTGTTATC AGTATGTATAC AGTATGTATAC AGTATGTATAC AGTATGTAAC GATACTACAAC GATTCTACAA GGAAAACGTT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATCTGGTA AATCCAGCG AGAACATGC AAATCCAGCG AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGACGC AGACATTTC GCTCCAGCGT ATCGCCACTG GCTGCTTTTTT AAACTCAGCGA TATGGCACTA AACCAGCCA TATGGTAGAT TCGTGTAGAT ACGCAGTTA TCGTGTAGAT AAACCAGCCA ATCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCAGTTA TCGCCATGTATGC CTTCGGGCCG CTTCGGGGCG	GGGCAATTCT  TCTACGCGTA TGAGTTTGA AAACAGTTA CAACATTA CAACATTA CAACATCTA CAACACTA TGCCGTACTA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACGAG CCGGATACGTA TATTTGGTTT TGTTTGCAG TATTTGCTA AACTACGAT ACTACGAACGACTA AACTACGATA CCCGGAAGGG AAAACTCTCA ATGCACCAC AGGCACCAC AGGCACCAC AGGCACCAC AAAACTCTCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAGG GGTGGCGAA GCACACACACACACACACACACACACACACACA
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATCAG GCTAATCAG GCGGGGAG CTCAAAGGCG ACCAGGGGTT GCTTTCTCAT GCTTATCCG TAGGCGTTG CCTAATCAG GCTTATCCG TAGGCGTTG CCTAATCAG CCTAATCAG CCTAATCAG CCTAATCAG TATCAGATCCT TATCTCAGCG GCAATGATAC CCGCTCCAT TATCTCAGCG TCCTTCGGTC TAAGATGCT TAACAGCGT TACGGCCAA ATGTAACCCA ATGTAACCCA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCTGGA AGCTCACCCT GTAACTATCG GTAACTATCG GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCTGATCTT TCGTCGTTTG CCCAGTCTATT CGTCGTTTG CTCCGATCGT TTCTGTGACT CATAGCAGAC CATCGTTTC CTTCGTGACC CATCGATCGT TCTTGTGACT CATAGCAGAC	CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT GATGTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGGCTCCTCG GAAGTAATT TCTTGAGTCC TTGAGTCCTTC GAATGAGT TTCTACGGGA AATGAAGTGT TTCTTCATCACCACA TTCTACGGG AAATGAAGTT TCTTCATCACC AATTGTTCCC GTATGGCTCACCG CAATTAATAGT TCTCATCACCC CGAATTAATAGT CGTAGGAGTACT TGTCAGAAGT CGTAGGAGTACT CTTAAAAAGT CTTTAAAAAGT CAACTGATCT	CCTATCCCTA CTAAGATCCA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCAGATTGGCAACTACAT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC CACCCAACTC CCCACACTC CCCACACTC CCCACACTC CTCACACTC CACCCACTC CACCCAACTC CCTCACACTC TTACACATCT CACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTTATC CATTGCCCC CATCCCCCC AACTGCCCGC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGACT CGCTACACT GCTGGTAACG AGTGGAACGA CTAAAGTATA GGATCACCCC TATCAGCAAT AGTAAGTATA AGTAAAGTATA AGTAAAGTATA AGTATAC CAGTGTTACCA CAGTGTTATC ATCTTGACAA CAGTGTTATC ATCTTGACAA CAGTGTTATC ATCTTGACAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTGTA AAATCACTGC ACTGCCACTGC AGAACATTGACGCC CTGCCGCTTA AATCGACGC ATCGCCACTG AGAAGACATGTG AGAAGACATGTG AGAAGACATGTG AGAAGACATGTG AGAAGGACGG TTATGAGTAAA AAACCACCA TCGCCAGTTA AAACCAGCA CGCCAGTTA CATCAGGCA ACTCATGGTAATA CATCAGGCA CTCAGGGTTAGAT AAACCAGCCA TCGCCAGTTA CATCAGGTAAA CATCAGGCG ACTCATGGTT TAGTGTATGC CTTCGGGGCG CAGCGTTTCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CGACATACGA GGAAACCTGT TCGCGTCGGT AGCAAAAGCTC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCACGAGCA TATTTGCAAG TAAGGGAATTC TTAGGTCTGA AACTACGATA CCGGAATTC TAAGGGAATT CTTTGCAAG ACTACGATA CATTTGCAAG AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACAATA AACTACAACA AACTACCACA AACTACCACA AACTACCACA GGCGACCCAG AAAACTCTCA GGGTGACCAC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAAA GTCCGCCTTT CAGCAACCC CTGGTAACAG CTGGCTATACAA CAGCACATTACCAA CGGGAGGGCT CCGAGCGCT CCGAGCGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TGCATAATTC TTGCTCTTGC AGGATCTTTC AGGAGACACA AAACAGGAAC
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG ACCAGGGGAG CTCAAAGGCG TGCTTGCTGG ACCAGGCGTTCTTGTG CCTTATCCAT CCTACAAGGCGT TAGGCGGTC TAGGCGGTC TAAGAAGAGTT CAAGAAGAGTC CTAAAAGAGT CCTACATCAT CCTACATCAT TATCTCAGCG GCAATGATAC CCGCCTCCAT TAGCCGCCA TAAGATGCTT TACCCGCCCA ATGTAACCCA GAATAAGGGC	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGCCTATAGC TGACCTAAC TGCCCCTGGA AGCTCACCT GTAATACCGT GTAATATCGT TACAGAGTTC GGTACTATC TTACAGAGTTC TTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTTCGGTCGATCGT CTTCGGTCGATCGT CACAGCAGAAC CCAGCCTATT CGTCGATCGT CTCCGATCGT CATAGCAGAA CTCGTGCACC GACACGGAAA	CAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATTGCTATT TGTTTCCTGT CGAATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTACT TCTTGAGTCC TTGAAGTGGT ATCCGCAA AATGAAGTT TTCGTCATC AAATGAAGT TTCGTCACC GAATTGTTC GTAGGGGT TTCTGAGTCC CGTATGGCTCACC GAATTGTTCAC CGTATGGCTTC CTTTGAGAGT TTCGTCACC GTATGGCTTC CTTTAGAAGT TGTCAGAAGT TGTTGAAAGT CTTTAAAAGT CAACTGATCT	CCTATCCCTA CTAAGATCCA CCTATCCCTA CTAAGATCCA CCTTTATTTG GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTA CCCCCTGACG ATCAGGGGAT CCCCCTGACG GGCCTAACTA ACAAACCAC TCTGACGTTC TTAAATCAAT CATAGTTCGC GCTCCAGATT GGGAAGCTAG ATCAGCTTC CAGATTT GCGAACTAG CTCCAGATT CTGACCTC CACAACTA CTCACACTC CACCAACTC CACCAACTT CCCACATT TCACACTCT TCATACTCTT	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCTT CAGCAGCAGT CAGCAGCAT CAGCAGCAT TAGACACC CAGCAGTAT TGACTCCCCG TATAAGTAT TGACTCCCAC TATCAGCAAT AGTAAGTAGT GGTTCCCAAC CAGTTTTCCAAC CAGTTTTCAA GAAAACGTT TTACTTTCAC CCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AATCCACAC TTTCCAGTCG ACTGCACT ACTCCACAC TTCCAGTCG AGAACATCTG AAACATCGC AGAACATCTG AAACATCGACG AGAGCACAC ATCGCCACTTA ACTCCAACG TGGTTTTT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT ATCGCAGTTA AACTCAGGT ATCGCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CTCGGGGCG CACGCTTCT TAGTGTATATC CTTCGGGGCG CACGCTTTC TATTATTGAA	TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATAC GGAAAACTGT TGCCGTCGGT AGCAAAAAGC TCAAGTCAGA CCGGATAACT GGCGGATACT GGCGCTGGT GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTA GCTACGATA GCCGAAGGA AACTCTGA AGTACATA GCCGAAGGA AAAACTCTCA GGCGACCAA AAAACTCTCA GGGGAGCCAA GCATTATCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTCGGCAAT CAGCAAAAGG GCTCGCCTTT CACGACACTT CAGCAACACCA CTGGTTACAG CAGCTTACAG CAGCTTACAG CAGCTTACAG CAGCTTACAGA CAGCTTACAGA CAGCTTACAGA CAGCTTACTAGA CAGCTACACAG CAACTTACTAGA CAGCAGATTAT TCCCCCATGT TGCATCATGT TGCATCATTC TTGCTCTTTG AGAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG CAGCATTTGC
2001 GGATACTTGA GTTAGTTTA:	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG TGCTTTATTT AGGTCAGGG GGCGTAATCAG GCCTAATCAG GCCTAATCAG GCTTACTGT GCCTTGCTGG ACCAGGGGAG CTCAAAGGCG TGCTTTCTCAT GCCTTATCCG AAAAAGAGAT CCTAGAGAGAT CCTAGATCAT TATCTCAGG GCAATGATAC GCGTTCCGC TAGATGAT CTAGTCCT TATCTCAGG CCAATGATAC CCGCCTCCAT TACCCAGGC TCCTTGGGTC TAAGAGATCT TACCGCCA ATGTAACCCA AGATAAACCCA GGATAAAGGCT TCAAAAAGGCT TAAAAAGGCT TAAAAAAGGCT TAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAGGCT TAAAAAAGGCT TAAAAAGGCT TAAAAAGGCT TAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAAGGCT TAAAAAGGCT TAAAAAAAAAA	CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGAGCTAAGC TGAGCTAACA AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGAACTACC GTAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT TCGTCGATCTAT TCGTCGTTTG CTCGATCGTTTC CTCGATCGTTTC CTCGATCGTTC CATAGCAGA CTCGTGCACC CACCGGACACC CACACCGACAC ACACACGAAA TCAGAAAATA	CAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTAGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC GTAGTAGT GATCCGCAA TTCTACGGG AATCGATAAT TTCGTTCATC ACGCTCACCG GTAGGTATCT TCTGAGTACC GTAGGTGCT TCTTACGGGG AATCGTCACCG GATGTCATCAC GTAGGCTTC TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CAACTGATCT TGTTGAATAC CAACTGATCT TGTTGAATAC AACAAATAGG	CCTATCCCTA CTAAGATCCA GCTTATTTG GCGTTAGTTG GTGAAATTATTG GCGTTGCGCT ATCAGGGGAT CCCCTGAACG ATCAGGGGAT CCCCTGAACG ACCAGTACACA ACAAACACA ACAAACACA ACAACCAC TCTGACGCT CTTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAACCAC CCCAGATT GCGAACTA ATTCAGCTC AAGTTGCCC CAACTACTA ATTCAGCTC CAACTACTTC CGCCAAGTC CAACCACT CTGAACTAT TCAGCATCTT TCAGCATCTT TCATACTCTT TCATACTCTT GGTTCCCCCC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGGTCA ACGCAGGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CAGCAGGAA AGCATCACAA GCATCACAC CAGCAGCTT CAGCAGCAGT CAGCAGCAT CAGCAGCAT TAGACACC CAGCAGTAT TGACTCCCCG TATAAGTAT TGACTCCCAC TATCAGCAAT AGTAAGTAGT GGTTCCCAAC CAGTTTTCCAAC CAGTTTTCAA GAAAACGTT TTACTTTCAC CCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AATCCACAC TTTCCAGTCG ACTGCACT ACTCCACAC TTCCAGTCG AGAACATCTG AAACATCGC AGAACATCTG AAACATCGACG AGAGCACAC ATCGCCACTTA ACTCCAACG TGGTTTTT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT ATCGCAGTTA AACTCAGGT ATCGCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CGCCAGTTA CTCGGGGCG CACGCTTCT TAGTGTATATC CTTCGGGGCG CACGCTTTC TATTATTGAA	TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATAC GGAAAACTGT TGCCGTCGGT AGCAAAAAGC TCAAGTCAGA CCGGATAACT GGCGGATACT GGCGCTGGT GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTA GCTACGATA GCCGAAGGA AACTCTGA AGTACATA GCCGAAGGA AAAACTCTCA GGCGACCAA AAAACTCTCA GGGGAGCCAA GCATTATCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTCGGCAAT CAGCAAAAGG GCTCGCCTTT CACGACACTT CAGCAACACCA CTGGTTACAG CAGCTTACAG CAGCTTACAG CAGCTTACAG CAGCTTACAGA CAGCTTACAGA CAGCTTACAGA CAGCTTACTAGA CAGCTACACAG CAACTTACTAGA CAGCAGATTAT TCCCCCATGT TGCATCATGT TGCATCATTC TTGCTCTTTG AGAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG AACAGAAG AAACAGAAG AAACAGAAG AAACAGAAG CAGCATTTGC

## pMT-HisFlag-DmSNAP43 mut#9 STOP

1 magaaaaammm aaama		mama, a, a, m	aas aamaaaa	a. a. aaama.	as aammamam	ama a a a a a a a a	00000000000	a. a aaaaa
1 TCGCGCGTTT CGGTG.								
101 TCAGGGCGCG TCAGC								
201 CCGCACAGAT GCGTA								
301 TACGCCAGCT GGCGA								
401 TGAATTAATT CGTTG								
501 GGCCCCCCAC CGCCC	ACCGC CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTG	GAGAA CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAAC	GAAAG ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AATCA	TCTCA GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT	CAGAATGCAT	CATCACCATC	ACCATACCGA	CTACAAGGAT
901 GACGATGACA AGGGC	ACTAG TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA	CGATTTCAGC	GATTGGTTAA	TGATGGCGAA	AACTGCGAGT
1001 TCGAGGTGTT CTGCC								
1101 GCATGTGGCC AAGCG								
1201 ATCTACTACA AGCAG								
1301 GTCCGGAGCG GAAGG								
1301 GICCGGAGCG GAAGG.	ACACI CAICAGAICG	> M L		T O E	O A F R		L D Y	C O G L
1401 003033 0000	ama aa aaamamaaa							~
1401 GGACAATCTG GTGGA								
> D N L V D			G A K	E Q R Q		M Q K	QQRA	N G V
1501 AGTCTCACAT ACGAA								
> S L T Y E			Q A S Q			AAYN		K Q L
1601 CGGCTGGTCA TGAGC								
		A A A F			F A D I	Q S V	L G A	R K S T
1701 TCCAGATGAG AAATG								
> P D E K C		~	L E V	R Q R V		A M Y	G V E E	R E P
1801 CAACACCAGA CGGAT								
~ ~	E L E V Q		N E T Y	~		A T V F	Q R E	L P E
1901 ACGTGCAGCA AGAGT								
	Y E M I E	F S D D			E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAG	TTTAT AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
> D T								
2101 AGCACAGTGG CGGCC	GCTCG AGTCTAGAGG	GCCCGCGGTT	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT	TCTACGCGTA	CCGGTCATCA
2101 AGCACAGTGG CGGCC 2201 TCACCATCAC CATTG								
	AGTTT AAACCCGCTG	ATCAGCCTCG	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TGAGTTTGGA	CAAACCACAA
2201 TCACCATCAC CATTG	AGTTT AAACCCGCTG AAAAA TGCTTTATTT	ATCAGCCTCG GTGAAATTTG	ACTGTGCCTT TGATGCTATT	CTAAGATCCA GCTTTATTTG	GACATGATAA TAACCATTAT	GATACATTGA AAGCTGCAAT	TGAGTTTGGA AAACAAGTTA	CAAACCACAA ACAACAACAA
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA	GACATGATAA TAACCATTAT AAACCTCTAC	GATACATTGA AAGCTGCAAT AAATGTGGTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA	CAAACCACAA ACAACAACAA TGATCAGTCG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTATC 2501 ACCTGCAGGC ATGCA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGGG AGCTT GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA 2401 TTGCATTCAT TTTAT' 2501 ACCTGCAGGC ATGCA 2601 TAAAGTGTAA AGCCT'	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGGG AGCTT GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITATT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGGG AGCTT GGCGTAATCAGGGGT GCCTAATGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATCA	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGGG AGCTT GGCGTAATCAG GCGGT GCCTAATGAG CCAAC GCGCGGGAG GCTCA CTCAAAGGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTATT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT. 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATCA. 2901 CCAGGAACCG TAAAA.	AGTTT AAACCGCTG AAAAA TGCTTTATTI GTTTC AGGTTCAGGG AGCTT GGCGTAATCAG GGGGT GCCTAATGAG CCAAC GCGCGGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTCCA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA 2401 TTGCATTCAT TTTATT 2501 ACCTGCAGGC ATGCA 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG 2801 CGGGGAGCGG TATCA 2901 CCAGGAACCG TAAAA 3001 CCCGACAGGA CTATA	AGTTT AAACCCGCTG AAAAA TGCTTTATTI GTTTC AGGTTCAGGG AGCTT GCCTAATGAG CCAAC GCGCGGGAGA GCTCA CTCAAAGGCG AGGCC GCGTTGCTGG AAGAT ACCAGGCGT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCCCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC ACAGACATGTG AAATCGACGC CTGCCGCTTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG 2801 CGGCGAGCGG TATCA. 2901 CCCGGACAGGA CTATAA. 3001 CCCGACAGGA CTATAA. 3101 CTCCCTTCGG GAAGCC	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GGGGT GCCTAATGAG CCAAC GGCGGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTGG AAGAT ACCAGGCGTT GTGGC GCTTTCTCAT	ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCTGGA AGCTCACGCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGG CCGGATACCT GGGCTGTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTATT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGAA AGCCT. 2701 GCATTAATGA ATCGG. 2801 CGGGCAGCGG TATCA. 2901 CCAGGAACCG TAAAA. 3001 CCCGACAGGA CTATA. 3101 CTCCCTTCGG GAAGC. 3201 CGGTTCAGCC CGACC.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGG AGGTT GCCTAATCAG CCGAC GCCCGGGGAG GCTCA CTCAAAGGCC AGGCC GCGTTGCTGG AAGAT ACCAGGCGTTG GTGGC GCCTTATCCG	ATCAGCCTCG GTGAAATTG GGAGGTGTGG TGGGCTAACC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTACCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTATT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT 2701 GCATTAATGA ATCGG 2801 CGGCGAGGG TATCA. 2901 CCCGACAGGA CTATAA. 3101 CCCGACAGGA CTATA. 3101 CCCGTCAGCC GAACC. 3201 CCGTTCAGC GAACG. 3201 CGGTTCAGCC GACC. 3301 GATTAGCAGA GCGAGG	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GCCTAATGAG CCAAC GCGCGGGGAG AGCCC CCTCAACGCGTT GTGGC GCTTCCTG AAGAT ACCAGGCGTT GTGGC GCTTCTCAT GCTGC GCCTTATCCG GTATG TAGGCGGTGC TAGGCGTTATG TAGGCGGTGC	ATCAGCCTCG GTGAAATTTG GGGAGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTGAGTCC TTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC CACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCCGCTCGGT TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG 2801 CGGCGAGCGG TATCA. 2901 CCCGACAGGA CTATAA 3001 CCCGACAGGA CTATA 3101 CTCCCTTCGG GAAGCC 3201 CGTTCAGCC CGACC. 3301 GATTAGCAGA GCGAGG 3401 CTGAAGCCAG TTACC	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GCGGCGGGAGA GCTCA CTCAAAGGCG AAGACT ACCAGCGCTT GTGGC GCTTTCTCAT GCTGC GCTTTTCTCAT GCTGC TAGGCGGTG TTCGC AAAAAGAGTT	ATCAGCCTCG GTGAAATTTG GGAAGTTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAACTATCG GTAACTATCG GGTAGCTCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTCTCC CAGTTCGGTG AACCCGGTAA ACCACGGTAA ACAAACCAC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TCCCTCGGCTC AACGCAGGAA TGTTCCGACC TAGGTCGTTC GACACGACTA GGCATCACAC GCTGGTACCC GCTGGTACCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGGT ATCGCCACTG AGAAGAGACAGC GTGGTTTTTT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TACAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT. 2701 GCATTAATGA ATCGG. 2801 CGGGCAGCGG TATCA. 2901 CCGACAGGA CTATA. 3101 CTCCCTTCGG GAAGC. 3201 CGGTCAGC GAAGC. 3301 GATTAGCAGA GCGAG. 3401 CTGAAAGCAGA TTACC. 3501 CGCGCAGAAA AAAAG.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGG AGGTT GCCTAATCAG CCAAC GCGCGGGAG GCTCA CTCAAAGGCC AGGCC GCGTTGCTGG AAGAT ACCAGGCGTT GCTGC GCCTTATCCG GTATG TAGGCGGTG GTATG TAGGCGGTG GTATG TAGGCGGTG GAAAAAGAGTT GAATC CAAGAAGATT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGAGCTAACT GTAATACGGT CCCCTTGGA AGCTCACGGT GTAACTATCG TACACAGCTC GTAACTATCG TACACAGCTC CGTAGCTCTT CTTTGATCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TGATCGGCAA TTCTACGGGA TTCTACGGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCTC	GACATGATAA TAACCATTAT AAACCTTCTAC TATTCCGCTCA CATGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACT GACACGACTT GACACGACT GCTGCTACACG AGTGGAACGA AGTGGAACGA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG ACTGACTGG AGAACATGTG AAATCGACGC CTGCCGCTTA ACTCACCAG ACTCACCTG AGAAGACATGTG AGAAGACATGTG AGAAGACATGTG AGAAGACATGTG AGAAGACATGTA AAACTCACAGT	TGAGTTTGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCGAAA GTCCGCCTTT CAGGAACCCC CTGGTAACAG CTGGCTCTT TGGTCATGAT TGGTCATGAC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTATT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT 2701 GCATTAATGA ATCGG 2801 CGGCAGAGGG TATCA. 2901 CCCGACAGGA CTATAA. 3101 CCCGACAGGA CTATA. 3101 CCCGTCAGCC GAACC. 3201 CCGTTCAGC GAACG. 3201 CGGTTCAGC TACC. 3301 GATTAGCAGA GCAGG. 3401 CTGAAGCCAG TTACC. 3501 CGCGCAGAAA AAAAG. 3501 ATTATCAAAA AGGAT.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GCCTAATCAG GCGTCA CCCAAC AGCCC CCTAATCAG AGCCC CCTCATCAG AAGAC ACCAGCGTT GTGGC GCTTTCTCAT GCTGC GCCTTATCCG GTATG TAGGCGGTGC TTCGG AAAAAGAGTT CATCAGATCCT CTTCA CCTAGATCCT CTTCA CTAGATCCT CTTCA CTAGATCT CTTCA CTAGATCCT CTTCA CTAGATCCT CTTCA CTAGATCT CTTCATT CTTCATT CTTCATT CTTCATT CTTCATT CTTCATT CT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCGTTTG GTAATACGGT CCCCCTGGA AGCTCACCGT GTAACTATCG TACAGAGTTC GTAACTATCG GTAACTTT TTTAAATTTA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TGAAGTGGT GATCCGCAA TTCTACGGGA ATCACGGGA ATCACAGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA ACAAACCACC TCTGACGCTC TCTGACGCTC TTAAATCAAT	GACATGATAA TAACCATTAT AAACCTCTTAC TATCCGCTCA CACTGCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA GCATCACAC TGTTCCGACC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACGC CTAGATATA	GATACATTGA AAGCTGCAATA AAGCTGCAGTA CAATTCCAGTCG AGAACATCGC AGAACATCGC AGAACATCGC AGAACATCGC ACTCCCAAGCT ATCCCCAACGT ATCCCCACGT ATCGCCACTG AGAAGACACGT ATCACCACTT ATCACTACT ATCACTACT ATCACTACT ATCACTACT ATCACTACT ATCACTACT	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTCGGCGATC CTGGTAACAG CTGGGTAACAG CTGGGCTCT CAGCAGATTAACAG CTGCGCTCTC CAGCAGATTAA TGGTCAGTATACAG CAGTTACAGA
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGACAGGA CTATAA. 3001 CCCGACAGGA CTATAA. 3101 CTCCCTTCGG GAAGCC 3201 CAGTTACAGC CGACCC 3301 GATTAGCAGA GCGAGG. 3401 CTGAAGCCAG TTACC 3501 CGCGCAGAAA AAAAG. 3601 ATTATCAAAA AGGAT. 3701 TGCTTAATCA GTGAGG.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATGA GCCAAC GCGCGGGGA GCTCA CTCAAAGGCG AAGACT ACCAGCGCTT GTGGC GCTTTCTCAT GCTGC GCCTTATCCG GTATG TAGGCGGTGC TTCGG AAAAAGATT GATCT CAAGAACGT GATCT CAAGAACCT CCTCAA CCTAGATCCT GCACC TATCTCAGTCCG GCACC TATCTCAGTCCG GCACC TATCTCAGATCCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTACGT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAAGGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GAGCTCCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGGCAA TTCTACGGGG AAATGAAGTT TTCGTTCATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCTTGCGCT ATCAGGGGAT CCCCCTGAACT AACCCGGTAA AGCCCAACTA ACAAACCAC TCTGAACGCT CTGAACGAT CATAGTTGCC CATTGTTGCCT CTGAACGAT CATAGTTGCC CATTGTTGCC CATTGTTGCC CATTGTTGCC CATTGTTTGCC CATTGTTTTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC AACGCAGAA AGCATCACAA AGCATCACAA GCATCACAA GCTGGTACGA AGTGGAACGA CTAAAGGATAT TGACTCCCCG	GATACATTGA AAGCTGCAAT AAATGTGATCACACA TTTCCAGTCG AGAACATGTC AAATCGACGC AGAACATGTC AAATCGACGC AGACACTTA GCTCCAAGCT ATCGCCACTA AGAAGGACAC GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT TCGTGTAGAT	TGAGTTTGAA AAACATTA TGACTGATTA CAACATACGA GGAAAACTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT TGTTTGCAAG AACTACGATA AACTACGATA AACTACGATA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTG CAGCAAAAAGG GGTGGGCAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTTC CAGCAGATTA TGGTCATGAG CAGCAGATTA CAGCAGATTA CAGGAAGGCT CAGGAGGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT CAGGAGGCT
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT. 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATAA. 3001 CCCGACAGGA CTATAA. 3101 CTCCCTTCGG GAAGCC. 3201 CGGTCAGCA GCGAG. 3201 GATTAGCAGA GCGAG. 3401 CTGAAGCCAG ATACC. 3501 GGCGCAGAAA AAAAG. 3601 ATTATCAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGG AGCTT GCCTAATCA CCAAC GCGCGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTG AAGAT ACGTGTTCTCAA GCTGC GTATG TAGGCCGTG GTATG TAGGCCGTG CTTCGA AAAAAGATT GATCT CAAGAAGAT CCTTCA CTAGAGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTAACTATCG TCACGAGGTTC CTTTGATCTT TTTAAATTTAA ATCTGCTACTAC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAAGTT TCTTACGGGA AATGAAGTT TCGTTCACCACA ACCCTCACCACACACACACACACACACACA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	GACATGATAA TAACCATTAT AAACCTTCTAC TATTCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACAGCACACCTCGTCC GACAGCACCTCGTCC GACAGCACCTCGTCC GACAGCACCTCGTCCGCCCCCCCCCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGACATG AGAACATTG AGAAGACATG ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA AAACCAGCA AAACCAGCA	TGAGTTTGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAAGG	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGGACCCC CTGGTAACAG CTGGCTACGCTT TGGTCATGAC CTGGTAACAG CTGGTAACAG CTGGTAACAG CTGGTATACCAA TGGTCATGAG CAGCAGATTA TGGTCATGAG CAGGAGGCCC CCGAGGGCCCC CCGAGGGCCC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT. 2701 GCATTAATGA ATCGG. 2801 CGGCAGACGG TATCA. 2901 CCAGCAAGGA CTATA. 3101 CTCCCTTGG GAACC. 3201 CCGTTCAGCC CGACC. 3201 CGTTCAGCC GACC. 3201 CGGCAGAGAA AAAAG. 3401 CTGAAGCCAG TTACC. 3501 CGCCCAAAA AAAAG. 3601 ATTATCAAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 3901 AAGTGGTCCT GCAAC.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATGAG CCAAC GCGCGGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTGG AAGAT GTGGC GCTTTCTCAT GTGGC GCTTTCTCAT GTGGC GCTTTCTCAT GTGGC GCTTTCTCAT GTGGC CTTATCCG GTATG TAGGCGGTGC TTCGG AAAAAGAT CCTTCA CCTAGATCCT CCTAGCTCG GCACC TATCTCAGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTAACT GGATTTTG GTAATACGGT CGCTTTTTCCA AGCCGCTTGA AGCTCACCGT TACAGAGTTC GTAACTATCG GTAACTATCG GTAACTTCT TTTAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTCTG GTAGGTACT TCTTGAGTCC TTGAGTCG TATCACGGCAA TTCTACGGGA TTCTACGGCAA TTCTACGGCAA TTCTACTCACCAC AATTGATTCCC AATTGTTCACC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTC AACCAGCTAA ACAAACCAC TCTGAACCA TCTGAACCT TTAAATCAAT CATAGTTGCC GCTCAGACTT GGGAACCTAG	GACATGATAA TAACCATTAT AAACCTCTTAC TATCCGCTCA CACTGCCGC TACCTGGTCA AACGATACACA AGCATCACAA AGCATCACAA GCAACACACT TGTTCCGACC TAGGTCGTTC GACACGACTT GCTGGTACGA CTAAAGTATT TGACTCCCCG TATAAGTATT TGACTCCCCG TATACACAT TATCAGCAT	GATACATTGA AAGCTGCAAT AAAGCTGGATA CAATTCCAGTGG ACTGACTGGC AGAACATGTG AAATCGAGGC TTGCCGGTTA GCTCCAAGGT ATCGCCACTG AGAAGATTTTT AAACTCACAGT ATGGAGTATATTTT AAACTCACGT ATGGAGTAAA TCGGGTAGATAA	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATCGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCGTGTGTG GCAGCAACAT TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTAGA ACTACGATA ACTACGATA ACTACGATA GCCGGAAGGG ATAGTTTTGCA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGGCAAA GTCCGCCTTT CACCAACCC CTGGTAACAG CTGGCTAACAG CTGGCAATAA TGGTCATAACAG CAGCATTA CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCCGAGCGCAG CAACGTTGTT
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCAGAGCG TATAA. 3001 CCCGACAGGA CTATAA. 3101 CTCCCTTCGG GAAGCC 3201 CAGTTAAGCA CGACC. 3301 GATTAGCAGA CGAGC. 3401 CTGAAGCCAG TTACC. 3501 GCGCCAGAAA AAAAG. 3601 ATATCAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCAC. 3901 AAGTGGTCCT GCAAC. 4001 GCCATTGCTA CAGGA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GGGGT GCCTAATGAG CCAAC GGCGGGGAG AGCC CCTCAAAGGCG AAGACT ACCAGGCGTT GCTGC GCTTTTCTCAT GCTGC GCTTTTCTCAT GATTC CAAGAAGATT CCTTCA CCTAGATCCT GCACC TATCTCAGCG GTGCT GCAATGATAC ATTGAT ATCGT GGCGCTCCAT ATCGCTCAAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGAGCTACGT GTAATACGGT GCTATTCCA AGCTCACGCT GTAACTATCG TACAGACTTC GGTAGCTCTT CTTGATCTT TTTGATCTT TTTAATTAA ATCAGTCTCTT TCGCGGAGCCC CCAGTCTCTTT TCGTCGTTTT TCGTCGTTTT TCGTCGTTTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGTCCCG GTAGGTACTG GTAGGTACT TCTTGAGTCC TTGAAGTGC TTGAAGTGC TTGAAGTGT TTTACGGGG AATCAGAGTT TTCCTTCATC ACGCTCACCG AATTGTTGCC GTATGGCTT GTTTGCC GTATGGCTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT GCCTTGCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTA AACCCGGTAA ACCACGGTAA CACAACCAC TCTGACGCT TCTGACGCT CTGACGTC CTTAATTAC CATAGTTGCC GCTCCAGATT GGGAAGCTAC ACTAGCTCC	GACATGATAA TAACCATTAT AAACCCTCAT TATCCGCTCA CACTGCCGCC AACGCAGGAA AGCATCACAA GCATCACAA GCATCACAA GCATCACAA GCATCACAA GCATCACAA GCATCACAA TGTTCCGACC TAGGTCGTTC GACAGACTT CGGCTACACT GCTGGTAACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTAGTA GGTTCCCAAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG AGAACATGTG AAATCGACGC TTCCAGTCG AGAACATGTG AAATCGACGC TTGCCGCTTA GCTCCAAGCT ACGCCACTG AGAAGCATGT AGAAGCAGCA TTGGTTTTT AAACTCACGT TATGAGTAAA AAACCAGCCA TCGCCAGTTA CAGCCAGTTA GATCAAGGCG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAGGGATTT CTTGGTATT CCGGAAGGG ATAGTTTTCCG AGTTACATGA	CAAACCACAA ACAACAACAA TGAATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGGCAAA GCTCGGCCTT CACGAACCCC CTGGTAACAG CTGCGCTAT CAGCAACTA CGGAAGCTAC CAGCAGATTA CGGGAGGCT CCGAGCGCAG CAACGTTGTT CCCCCATGT CCCCCATGT TCCCCCATGT
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT. 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATAA. 3001 CCCGACAGGA CTATAA. 3101 CTCCCTTCGG GAAGC. 3201 CCGTTCAGCC CGACC. 3301 GATTAGCAGA GCGAG. 3401 CTGAAGCCAG ATACC. 3501 CGCGCAGAAA AAAAG. 3601 ATTATCAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 3901 AAGTGGTCCT GCAAC. 4001 GCCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGG.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGG AGCTT GCCTAATCAG CCAAC GCCCGGGGAG GCTCA CTCAAAGGCC AGGCC GCGTTGCTG AAGAT GCTGC GCTTATCCG GTATG TAGGCCGTGC TTCGG AAAAAGATT GATCT CAAGAAGATC CTTCA CTAAGAGCC GTGCT GCAACCT TTTAT CCGCCTCCAT ATCGCCTCCAT ATCGCTGCT TTTAGC TAGGCCGTTC TTTAGC TAGCCCGTTC TTTAGC TCAAGAAGATC TTTAT CCGCCTCCAT ATCGCTCCAT ATCGCTCCAT TCTTCGG TCTTCGGT TTTAGC TCATCCGCT TTTAGC TCCTTCCGT TTTAGC TCCTTCCGT TCTTCGT TCTT TCTTCGT TCTTCGT TCTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGCTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG TCACAGGTTC CTTAGATCTT TTTAAATTAA ATCAGATTC CCCAGGAGACC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCGAAATTCA CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCGAATCGT CTCGAATCGT CTCGAATCGT CTCGAATCGT CGCAAACT CTCGAATCGT CTCCAATCGT CTCGAATCGT CTCGAATCGT CTCGAATCGT CTCCAATCGT CTCGAATCGT CTCCAATCGT CT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TCTACGGGAA TTCTACGGGA ATCTACGGGA AATGAAGTGT TCGGTCATC ACGCTCACCG AATTGTTGCC GTATGGCTTC TGTCAGAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAT TTAAATCAAT TATAATCAAT CATAGTTGCC GCTCAGATTGCCAATTGCC AATTGGCCC AAGTTGGCCC AAGTTGGCCC AAGTTGGCCC AAGTTGGCCC AAGTTGGCCC AAGTTGGCCC AAAATCAGCCC AAGTTGGCCC AAGTTGGCCC	GACATGATAA TAACCATTAT AAACCTTCTAC TATTCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACAGCACACAC GATGACCAC TAGGTCGTTC GACACGACT GCTGGTACAC AGTGGAACGA CTAAAGTATA TGACTCCCCC TATCAGCAAT AGTAAGTAGT GGTTCCCAAC CAGTGTTATC	GATACATTGA AAGCTGCAAT AAAGTGCAAT CAATTCCACA TTTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGACATG AGACACTG AGAAGACATG ATCGCCACTG AGAAGACATG ATCGCCACTG AGAAGACACT ATCGCCACTG AGAAGACAC ATCACTCAAGTAAA AAACCAGCCA TCGCCAGTTA GATCAAGCC ATCAAGGCC ATCAAGGCC ATCAAGGCC ACTCATGGTTAT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTCGA ATTACATCA ATGACAGAA ATGACAGAA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCT CAGCAAAAGG GGTGGCAAAA GTCCGCCTTT CAGCACCCC CTGGTAACAG CTGGCTATACAA CAGCACATTACCAA CAGCAGCACTTACCAA CAGCAGCGCCT CCGAGCGCCT CCGAGCGCCT CCGAGCGCT CCCACAGTTT TCCCCCCATGT TCCCCCCATGT TCCCCCATGT TCCCCCCATGT TCCCCCATGT TCCCCCATGT TCGCATAATTC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG 2801 CGGCGAGCGG TATCA. 2901 CCCGACAGGA CTATA. 3101 CTCCCTTCGG GAAGC. 3201 CCGTTCAGC GACGG. 3401 CTGCACAGGA GCGAG. 3401 CTGAGAGCAG TTACC. 3501 CGCCCAGAAA AAAAG. 3601 ATTATCAAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 3901 AAGTGGTCCT GCAAC. 4001 GCCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGGT. 4201 TCTTACTGTC ATGCC.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GGGGT GCCTAATGAG CCAAC GCGCGGGGAG AGGCC GCGTTGCTG AAGAT GTGC GCTTTCTCAT GCTGC GCTTTCTCAT GCTGC GCTTTCTCAT GATCT CAAGAAGAT CTTCAG CAAAAAAAAAA GATCTTCAAC GCTGC TCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTATAGC TGAGCTAACT GGAATACGGT CGTTTTTCCA AGCCGGTTTG GTAATACGGT TCACAGGTT GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG TCTCCGATCGT TTCTGTGATTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTT TCTTGAGTCC GTAGGTACT TTCAAGGG GATCCGGCAA TTCTACGGG AATGATACT TCCGTCATC ACGCTCACCG GTAGGTTCT TCGTCATC CGTAGGCTTC TTCAAGGGG GTAGGTACT GTCAGAAGT TTCGTCAAGG GTAGGATACT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACCAGTCCGTAAA ACCAACCA TCTGACGTC TTAAATCAACT CTGACGTC GCTCAAGTT GGGAAGCTAG ATCAGGTCC AACCAAGTT ATCAGCTCC AACTAGATTGCC AACTAGACTAC ATCAGCTTC CAACTAGATTGCCC CAACCAAGTT CAACCAAGTC	GACATGATAA TAACCATTAT TAACCCTTAC TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCATCACAA GCATCACAA GCATCACAA GCATCACAA GCATCACAA GCAGACTT CGGCTACACT GCTGGTACGA CTAAAGTATA TGACTCCCG TATCAGCAA TATCAGCAAT AGTAAGTAGT GGTTACCAAC CAGTTTACC CAGTTTACC CAGTTACAAC CAGTTTATC CAGTTACAAC CATTCTAGAAA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG AGAACATGGC AGAACATGTG AAATCGAGGC TTCCCAAGCT ATCGCAGCTTA GCTCCAAGCT ATCGCCACTG AGAACATTTT AAACTCACGC TGGGGGTTA TATCAGGT TATGAGGACAG TGGTTTTTT AAACTCACGT AAACTCACGT AAACTCAGTAAA TCGGCAGTTA GATCAAGGCG ACTCAATGAGT TAGAGTATATT TAGAGTATATG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA GCTAAGGATACT GCCGAAAGG ATACTTTGCA AACTACGATA GCCGAAAGG ATACTTTGCA AGTTACATA GCCGAAAGG ATACTTTTGCA AGTTACATA CGCGCACCAC AGTTACATGA ATGCCAGCAC GGCGACCCAG	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTC CAGCAAAAGG GGTGGCGAAA GCTCCCCTTTC CACGAACCCC CTGGTAACAG CTGCGCTTACAG CAGCATAACAG CAGCACATTA TGGTCATACAG CAGATTACAGA CAGATTACATAC CAGAAGCCC CCGAGCGCT CCGAGCTCT CCCAACTTT TCCCCCATGT TGCATATATTC TTGCTCTTTGC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATCA. 2901 CCCGACAGGA CTATA. 3101 CTCCCTTCGG GAAGC. 3201 CGGTTAAGCA GCGAG. 3401 CTGAAGCCAG TTACC. 3501 GATTAGCAGA GCGAG. 3401 CTGAAGCCAG AAAAAA 3601 ATTATCAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 4001 GCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGGT 4201 TCTTACTTC ATGCG. 4201 TCTTACTTC ATGCG. 4201 CCGGCGTCAA TACCG.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGGG AGCTT GGCGTAATGAG CCAAC GGCGGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTG AAGAT ACCAGGCGTT GTGGC GCTTTCTCAT GCTGC GCCTTATCCG GCTAT CAAGAAGAT CCTCAA GAAAAAAAAAAAT GATCT CAAGAACCT TTTAT CCGCCTCCAT ATCGT GGTGCCTCAAT ATCGT GGTGTCACGG TTAGTTCAGGCG TAGTTCAGTCCT TAGATCCT TTAAT CCGCCTCCAT ATCGT GGTGTCACGC TAGCTCAGTCCT TAGATCCT TAGATCC TAGATC TAGATCC TAGATCC TAGATC TAGATCC TAGATC T	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGAGCTACGT GTAATACGGT CGTTTTTCCA ACCCCTCGGAGCTCT TTCAACTATC GTAGCTCTT TTTGATCTT TTTAATTTA ATCTGTCTAT CGCGAGCCC CCAGTCTTT TCGTCGTTTG CTCGATCGT TCCGTTGGTTTC CTTCGATCGT CATAGCAGAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTGAAGTGGT GATCCGCAA TTCTACGGGA TTCTACGGCAA TTCTTACGGGC AATTGTTCCC GTAGGTTATCT TCGTCACC GTAGGCTTC TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CGTGAGAAGT CGTGAGAAGT CTTTAAAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGAAGT AACCCGTTAA ACCACGGTAA ACCAACTA ACAAACCAC TCTGAAGTA CATAGTTGCG GCTCCAGATT GGGAAGTAC ACTAGTTGCC AGTTGCCAGATT ACTAGATTGCC ACTTGCCAGATT ACTAGTTGCC CACCAGATT ACTAGTTGCC CACCAAGTC CAACCAAGTC CAACCAAGTC CCTCATCATT	GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC TACCTGGCTCA AACGCAGGAA AGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGATACCAAA TGACTCCCCG TATCAGCAAT AGTAAGTATA GGTTCCCAAC CAGTGTTACA CAGTGTTATCA GGAAAACGTT	GATACATTGA AAGCTGCAAT AAAGCTGCACA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGC AGCCCTA GCTCCAAGCT ATCGCCACTG AGAACACACAC ATCGCACTA ACACCACACAC ATCGCCACTA AAACCACCAC TATCGTGAGAT ACCAGCCA TCGCAGTTA AAACCAGCCA ACCAGTTA GATCAAGGCT ACTCAAGGTT TAGTGTAAGT ACTCAAGGTT TAGTGTAAGT CTTCGGGGCG CTTCGGGGCG	TGAGTTTGAA AAACAGTTA TGAGTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG AACTACGTA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ACGACAGCA AGGACCAGC AGGCACCAG AAAACTCTCA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GCTCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCT CAGCAGATTA TGGTCATGAG CAGTACCAA CGGGAGGGCT CCGAGCCAGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCTCATGT TCCTCTCTCC AGGATTATC
2201 TCACCATCAC CATTG. 2301 CTAGAATACA GTGAA. 2401 TTGCATTCAT TTTAT. 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCTT. 2701 GCATTAATGA ATCGG. 2801 CGGGCAGCGG TATAA. 3001 CCCGACAGGA CTATA. 3101 CTCCCTTCGG GAAGC. 3201 GATTAGCAGA GCGAG. 3201 GATTAGCAGA GCGAG. 3401 CTGAAGCCAG TTACC. 3501 GATTAGCAGA GCGAG. 3401 ATTATCAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTG GCACC. 3901 AAGTGGTCCT GCAAC. 4001 GCGCATGCTA CAGGC. 4101 TGTGCAAAAA AGGGG. 4201 TCTTATCTTA ATGCG. 4301 CCGGCGCTCAA TACGG. 4401 CGCTGTTTGAG ATCCA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGTTCAGG AGCTT GCCTAATCA GGGGT GCCTAATGA CCAAC GCGCGGGAG GCTCA CTCAAAGGCG AGGCC GCGTTGCTG AAGAAT GCTGC GCTTTCTCAA TTCGG AAAAAGAGT CTTCA CTAAGAGT GTAGT TAGGCGTGC TTTGG AAAAAGAGT CTTCA CCTAGATCCT GCACC TATCTCAGT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACCCT GTAACTATCG GTAACTACGGT TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCAGTCTC CCAGTCTATT CGCGAGCCC CCAGTCTATT CGTCGTTTG CTCCGATCGT TCCTGTGACC CATCGTTCT CTCGTCGATCGT TCCTGTGACC CATCGTCACC CATCGTCTCT CTCTGTGACC CATCGATCGT TCCTGTGACC CATCGATCGACC CATCGATCGT TCCTGTGACC CATCGACCACA CTCCTCCACC CCACGACCACC CATCGATCGT TCCTGTGACC CATCGACCACA CTCCTCCACCAC CTCCTCCACCAC CTCCTCCACCAC CTCCTCCACCAC CTCCTCCACCAC CTCCTCCACCAC CATCCACCAC CTCCTCCACCAC CTCC	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTACGGGA AATGAAGTT TTCGTTCATC ACGCTCACC AATTGTTGCC TGTAGGAGTACT TGTCAGAAGT TCTCATCACC TTCAACAGA TCTCACCG TTCAACAGA TCTCACCG TTCAACAGAGT CGTATGGCTC TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CGTAGAGTACT CTTTAAAAGT CAACTGATCT	CTAAGATCCA GCTTTATTTG AAAGCAACTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGGGCTTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCC TCTAACTA TAAATCAAT CATAGTTGCC GCTCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGCCC CAACCAACTC CTCACCATCT CACACTC CACCAACTT CACACTC CACCAACTT TCACACTCT CACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT	GACATGATAA TAACCATTAT AAACCTTATCA CATTGCCGCC CATTGCCCGCC TATCTCGGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGATACACT CGGCTACACT GCTGGTAGCG AGTGGAACGA CTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTATAACTATA CAGTCTCAAC CAGTGTTATC ATTCTGACAA CAGTGTTATC ATTCTGACAA	GATACATTGA AAGCTGCAAT AAAGTGGTA CAATTCCACA TTTCCAGTGG ACTGACTGC AGAACATGTG AAATGGAGC CTGCCGCTTA ACTGCAACT ACCACACT ACCACACT ACACACT AC	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CCTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTTGCA ATGGATTTTCGA ATGGACACCA ATGGCACCAC ATGGCACCAC ATGGCACCAC AGGCACCAG AAAACTCTCA GAGGACCAG GAAACCAC GGGGACCAG GAAAACTCTCA GGGTGACCAC	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCT CAGCAAAAGG GGTGGCAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCACATTACCAA CAGCACATTACCAA CGGAGGGCT CCGAGCGCAT TCCCCCCATGT TCCCCCCATGT TCCCCCCATGT TCCCCCCATGT TTGCTCTTTGC AGGAATCATCA AAACAGGAAC
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATCA. 2901 CCAGCAAGCA CTATA. 3101 CCCCTTCGG GAAGC. 3201 CCGTTCAGC GAAGC. 3201 CGGTTCAGC GCAGG. 3401 CTGCATCAGC GCAGG. 3401 CTGCAGAAA AAAAG. 3601 ATTATCAAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 3901 AAGTGGTCCT GCAAC. 4001 GCCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGT. 4201 TCTTACTGTC ATGCC. 4301 CCGGCGTCAA TACGG. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCCA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GCGCTCA CCAAC AGGCC GCGTGGGA AGGCC GCGTTGCTG AGGCT GCCTTATCCG GCTGC GCCTTATCCG GCTGC GCTTTCAG GATCT CAAGAAGAT CCTTCA GCACC TACTCAGATCCT TCGG AAAAAGAT CCTTCA CCTAGATCCT TTAGC GCACC TTTAGC GCACC TTTAGC GCACC TTTAGC TGCATCCT TTAGC TGCATCCT TTAGC TCCTTCAGC GTTCC TAAGATCCT TTAGC TAAGATGCTT ATCGT TAAGATGCT TTAGC TAAGATGCT TAAGATGCT ATCGT TAAGATGCT GATAA TACCGCGCA ATGTAACCCA AAAGG GAATAAAGGC	ATCAGCCTCG GTGAAATTTG GGAGGTGAGG TGGTCATAGC TGGGCTATAGC TGAGCTAACT GGATATCGGT CGTTTTTCCA ACCTCAGCA ACCTCACCT TACAGAGTTC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCA CCCAGTCTAT TCTTGATCTT TCTTGATCTT TCTTGATCTT TCGGAGACCC CCAGTCTATT TCGCGATCGT TCTCGGTTGG TCTCGGTCGATCGT CTCTGGATCGT CTCTGGATCGT CATAGCAGGA CTCGTGCACC GACACGGAAACCC GACACGGAAA CTCGTGCACC GACACGGAAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTCTG GTAGTAGTCC TCTTGAGTCC TTGAGTCC TTGAGTGC AATCACGGCAA TTCTACGGG AATCATCACC AATTGTTCAC GTAGGTACT CTTCAGAGT TCTCACAGG GGTAGGTTCC CTTTAGAAGT TTCTTCAGAAGT CTTTAAAAGT CTTTAAAAGT CAACTGATCT CTTTAAAAGT CAACTGATCAT CTGTTAAAAGT CAACTGATCAT TGTTGAATAC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCGTAACTA ACAAACCAC TCTGAACTA CTAAGTGCG CTCTAACTA ACAAACCAC CTCAGACT CGCAGACTACT GCGCAGACTAC GCTCCAGATT GCGAACTAC ATCAGCTCC AAGTTGCCC CACCAGATT CTGACCTCC CACCAGTT CTGACCTCC CACCAGTT CTCACCTT CTCACCTT CTCACCTT CCACCATCT CCACCATCT TCATCATCTT TCATCATCTT	GACATGATAA TAACCATTAT TAACCCTTAC TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCATCACTC GCTGGTACGAC CAGGTCGTC GCTGGTACGC CTAAAGTATT GACTCCCCG TATCAGCAT TATCAGCAT AGTAAGTAGT GGTTCCCAAC CAGTGTTACC CAGTTTTCCAAC CAGTTTTCAC CCCTTTTTCAC CCCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG AGAACATGGC AGAACATGGC AGAACATGTG AAATCGAGGC TTCCCAAGCT ATCGCCACTG AGAAGACTTTT ACTCCAAGCT ATCGCCACTG AGAAGACAGCA ATCGCACTA AAACTCACGCA ATCACAGCT AAACTCACGT AAACTCAGGT AAACTCAGGT AAACTCAGGT AAACTGAGGC ACTCATGATT TAGTGTAATGC CTTCGGGGCGC CACGTTTCT TATTATTGAA	TGAGTTTGGA AAACAGTTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTG GCACTGTTG GCACTGTTG TATTTGGTAT TGTTTGCAAG TAAGGATTAGAA ACTACGAA ACTACGAA ACTATACGATA GCCGGAAGGA GATACTTTGCA AGTACATA GCCGGAAGGA AAAACTCTCA GGCGACCGAA AAAACTCTCA GGGTAGCAA GCATTATACA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGCGAAA GCTCGCCTTT CACCAACCCC CTGGTAACAG CTGGCTAACAG CAGCATTA CAGCAGATTA TGGTCATGAC CAGAGATTA TGGTCATGAC CAGATTACAGA CAACGTTGT TCCCCATGT TGCATAATTC TTGCTCTTTG AGGAGATCTTAC AGAGATCTTAC AGAGATCTTAC AAACAGGAAG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TTTAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGAGGGG TATCA. 2901 CCGGCAGGGG CTATAA. 3101 CTCCCTTCGG GAAGC. 3201 CGGTTCAGC CGACCG. 3201 CATTAGCAG GCGAG. 3401 CTGAAGCCAG TTACC. 3501 GATTAGCAGA GCGAG. 3401 TGGTAACCA GTGAG. 3601 ATTATCAAAA AAGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 4001 GCCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGT 4201 TCTTACTTCA TGCG. 4201 CCGGCGTCAA TACGG. 4401 CGCGTGTTGAG ATCCA. 4501 GCAAAATGCC GCAAA. 4601 CTCATGAGCC GATAC.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GGGGT GCCTAATGAG CCAAC GCGCGGGAGA GCTCA CTCAAAGGCG AGGCC GCGTTGCTG AAGAA AGGAT GCTGC GCCTTATCCG GCTAC CCTCAACGCGTT CTCG AAAAAGAT CCTTCA CCTAGATCCT GCACC TATCTCAGCG GTGCT GCAATGATA ATCGT GGTGTCAGG TTTAT CCGCCTCCAA ATCGT GGTGTCAGG TTTAGT TACCGGGT TTAGT TACCGGGT AACAAGAGAT TTACGT GGTGTCAGG TTTAGT TACCGGCCCA GTTCG TAAGATGCT GATAA TACCGCGCCA GTTCG TAAGATACGT GATAA TACCGCGCCA GTTCG TAGATACGCA AAAGG GAATAAAGGG ATATT TGAATAGTATT	ATCAGCCTCG GTGAAATTTG GGAGGTGAGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGAGCTATCG GTATACCGT GTAATACGGT CGTTTTTCCGA AGCTCACCGT TACACAGTTC GGTAGCTCTC TTCAGCTTT TTTAATCTT TTTAATCTA ATCTGTCTAT TCGCGAGCCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCTGTGGTT CTTCGTCGTTC CATAGCAGAA CTCGTGCACC CAGCACCAC CAGCACCAC CAGCACCAC CAGCACACA CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC TAGCACGAAA TAGAAAAATA	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTGAAGTGG GATCCGCAA TTCTACGGG AAATGAAGT TTCGTTCATC ACGCTCACCG GTATGGCTCC GTATGGCTTC TGTCAGAAGT TCTTCAGAAGT TGTTCAGAAGT TGTTCAGAAGT TGTTGAATAC CACTGATCT TGTTGAATAC CACACAATAAGG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGC ATCAGGGGAT CCCCCTGAAG TGCGCTCTCC CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC ATAGTTGCC GCTCCAGATT GGGAACTAC ACTAGTTGCC GCTCCAGATT TCAGACTCC CACCTCAGATT CGCAACTCC CCTCAAGTT	GACATGATAA TAACCATTAT TAACCCTTAC TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCATCACTC GCTGGTACGAC CAGGTCGTC GCTGGTACGC CTAAAGTATT GACTCCCCG TATCAGCAT TATCAGCAT AGTAAGTAGT GGTTCCCAAC CAGTGTTACC CAGTTTTCCAAC CAGTTTTCAC CCCTTTTTCAC CCCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG AGAACATGGC AGAACATGGC AGAACATGTG AAATCGAGGC TTCCCAAGCT ATCGCCACTG AGAAGACTTTT ACTCCAAGCT ATCGCCACTG AGAAGACAGCA ATCGCACTA AAACTCACGCA ATCACAGCT AAACTCACGT AAACTCAGGT AAACTCAGGT AAACTCAGGT AAACTGAGGC ACTCATGATT TAGTGTAATGC CTTCGGGGCGC CACGTTTCT TATTATTGAA	TGAGTTTGGA AAACAGTTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTG GCACTGTTG GCACTGTTG TATTTGGTAT TGTTTGCAAG TAAGGATTAGAA ACTACGAA ACTACGAA ACTATACGATA GCCGGAAGGA GATACTTTGCA AGTACATA GCCGGAAGGA AAAACTCTCA GGCGACCGAA AAAACTCTCA GGGTAGCAA GCATTATACA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGCGAAA GCTCGCCTTT CACCAACCCC CTGGTAACAG CTGGCTAACAG CAGCATTA CAGCAGATTA TGGTCATGAC CAGAGATTA TGGTCATGAC CAGATTACAGA CAACGTTGT TCCCCATGT TGCATAATTC TTGCTCTTTG AGGAGATCTTAC AGAGATCTTAC AGAGATCTTAC AAACAGGAAG
2201 TCACCATCAC CATTG. 2301 CTAGAATGCA GTGAA. 2401 TTGCATTCAT TITAT 2501 ACCTGCAGGC ATGCA. 2601 TAAAGTGTAA AGCCT 2701 GCATTAATGA ATCGG. 2801 CGGCGAGCGG TATCA. 2901 CCAGCAAGCA CTATA. 3101 CCCCTTCGG GAAGC. 3201 CCGTTCAGC GAAGC. 3201 CGGTTCAGC GCAGG. 3401 CTGCATCAGC GCAGG. 3401 CTGCAGAAA AAAAG. 3601 ATTATCAAAAA AGGAT. 3701 TGCTTAATCA GTGAG. 3801 TACCATCTGG CCCCA. 3901 AAGTGGTCCT GCAAC. 4001 GCCATTGCTA CAGGC. 4101 TGTGCAAAAA AGGGT. 4201 TCTTACTGTC ATGCC. 4301 CCGGCGTCAA TACGG. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCGC. 4401 CGCTGTTAGC ATCCA.	AGTTT AAACCCGCTG AAAAA TGCTTTATTT GTTTC AGGGTTCAGG AGCTT GGCGTAATCA GGGGT GCCTAATGAG CCAAC GCGCGGGAGA GCTCA CTCAAAGGCG AGGCC GCGTTGCTG AAGAA AGGAT GCTGC GCCTTATCCG GCTAC CCTCAACGCGTT CTCG AAAAAGAT CCTTCA CCTAGATCCT GCACC TATCTCAGCG GTGCT GCAATGATA ATCGT GGTGTCAGG TTTAT CCGCCTCCAA ATCGT GGTGTCAGG TTTAGT TACCGGGT TTAGT TACCGGGT AACAAGAGAT TTACGT GGTGTCAGG TTTAGT TACCGGCCCA GTTCG TAAGATGCT GATAA TACCGCGCCA GTTCG TAAGATACGT GATAA TACCGCGCCA GTTCG TAGATACGCA AAAGG GAATAAAGGG ATATT TGAATAGTATT	ATCAGCCTCG GTGAAATTTG GGAGGTGAGG TGGTCATAGC TGGTCATAGC TGGTCATAGC TGAGCTATCG GTATACCGT GTAATACGGT CGTTTTTCCGA AGCTCACCGT TACACAGTTC GGTAGCTCTC TTCAGCTTT TTTAATCTT TTTAATCTA ATCTGTCTAT TCGCGAGCCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCTGTGGTT CTTCGTCGTTC CATAGCAGAA CTCGTGCACC CAGCACCAC CAGCACCAC CAGCACCAC CAGCACACA CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC CAGCACACAC TAGCACGAAA TAGAAAAATA	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTGAAGTGG GATCCGCAA TTCTACGGG AAATGAAGT TTCGTTCATC ACGCTCACCG GTATGGCTCC GTATGGCTTC TGTCAGAAGT TCTTCAGAAGT TGTTCAGAAGT TGTTCAGAAGT TGTTGAATAC CACTGATCT TGTTGAATAC CACACAATAAGG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGC ATCAGGGGAT CCCCCTGAAG TGCGCTCTCC CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC ATAGTTGCC GCTCCAGATT GGGAACTAC ACTAGTTGCC GCTCCAGATT TCAGACTCC CACCTCAGATT CGCAACTCC CCTCAAGTT	GACATGATAA TAACCATTAT TAACCCTTAC TATCCGCTCA CACTGCCGC TACCTGGTC AACGCAGGAA AGCATCACAA GCAATCACAA GCAATCACAA GCAATCACAA GCATCACTC GCTGGTACGAC CAGGTCGTC GCTGGTACGC CTAAAGTATT GACTCCCCG TATCAGCAT TATCAGCAT AGTAAGTAGT GGTTCCCAAC CAGTGTTACC CAGTTTTCCAAC CAGTTTTCAC CCCTTTTTCAC CCCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTGG AGAACATGGC AGAACATGGC AGAACATGTG AAATCGAGGC TTCCCAAGCT ATCGCCACTG AGAAGACTTTT ACTCCAAGCT ATCGCCACTG AGAAGACAGCA ATCGCACTA AAACTCACGCA ATCACAGCT AAACTCACGT AAACTCAGGT AAACTCAGGT AAACTCAGGT AAACTGAGGC ACTCATGATT TAGTGTAATGC CTTCGGGGCGC CACGTTTCT TATTATTGAA	TGAGTTTGGA AAACAGTTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTG GCACTGTTG GCACTGTTG TATTTGGTAT TGTTTGCAAG TAAGGATTAGAA ACTACGAA ACTACGAA ACTATACGATA GCCGGAAGGA GATACTTTGCA AGTACATA GCCGGAAGGA AAAACTCTCA GGCGACCGAA AAAACTCTCA GGGTAGCAA GCATTATACA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTG CAGCAAAAGG GGTGGCGAAA GCTCGCCTTT CACCAACCCC CTGGTAACAG CTGGCTAACAG CAGCATTA CAGCAGATTA TGGTCATGAC CAGAGATTA TGGTCATGAC CAGATTACAGA CAACGTTGT TCCCCATGT TGCATAATTC TTGCTCTTTG AGGAGATCTTAC AGAGATCTTAC AGAGATCTTAC AAACAGGAAG

# pMT-HisFlag-DmSNAP43 mut#10 STOP

1 TCGCGCGTTT CGGTGATGA								
101 TCAGGGCGCG TCAGCGGGT								
201 CCGCACAGAT GCGTAAGGA								
301 TACGCCAGCT GGCGAAAGG								
401 TGAATTAATT CGTTGCAGG	A CAGGATGTGG	TGCCCGATGT	GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC CGCCCACCG	CACCCCCATA	CATATGTGGT	ACGCAAGTAA	GAGTGCCTGC	GCATGCCCCA	TGTGCCCCAC	CAAGAGTTTT	GCATCCCATA
601 CAAGTCCCCA AAGTGGAGA	CCGAACCAAT	TCTTCGCGGG	CAGAACAAAA	GCTTCTGCAC	ACGTCTCCAC	TCGAATTTGG	AGCCGGCCGG	CGTGTGCAAA
701 AGAGGTGAAT CGAACGAAA	ACCCGTGTGT	AAAGCCGCGT	TTCCAAAATG	TATAAAACCG	AGAGCATCTG	GCCAATGTGC	ATCAGTTGTG	GTCAGCAGCA
801 AAATCAAGTG AATCATCTC								
001 1111110111010 11111011010	. 0100111011111	110000001110	1110111100000	11101111101111	> M H	н н н	н н т р	Y K D
901 GACGATGACA AGGGCACTA	י דכאככידכאאיד	አ ጥርጥጥጥር አ ርር	летестесса	сстсстссл х				
> D D D K G T			D C W E			R L V N	D G E	
1001 TCGAGGTGTT CTGCCGGTG								
>F E V F C R C		L Q L Q			T N H T		ATT	L A A L
1101 GCATGTGGCC AAGCGACTG								
> H V A K R L	S C S R			F P A S	~	R I G	G F F L	L Y V
1201 ATCTACTACA AGCAGCCCA	C GCACAACTTT	ATTAAGATCG	AGGTCTCACC	GCGCACTTGG	CAAGAACTAA	CAGACTACGC	TCTAGATCTG	CGCAAGGATA
> I Y Y K Q P '	HNF	I K I	E V S P	R T W	Q E L	T D Y A	L D L	R K D
1301 GTCCGGAGCG GAAGGACAC	CATCAGATCG	CCTACATGCT	GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
>S P E R K D T	H Q I	A Y M L	WRL	T Q E	QAFR	F T A	L D Y	C Q G L
1401 GGACAATCTG GTGGACTAC	ACCGTGTGGA	GACCGTAGCG	GGTGCCAAGG	AACAGAGGCA	GAGTGCCTTG	ATGCAGAAGC	AACAGCGTGC	GAACGGCGTC
> D N L V D Y	DRVE	T V A	G A K	E O R O	SAL	M O K	OORA	N G V
1501 AGTCTCACAT ACGAACTGG							~ ~	
> S L T Y E L					E L E			K O L
1601 CGGCTGGTCA TGAGCACGC								
>A A G H E H A	L P P	S O I F		A A A		0 S V	L G A	R K S T
		-				-		
1701 TCCAGATGAG AAATGCACC								
> P D E K C T	TTST			R Q R V		A M Y	G V E E	REP
1801 CAACACCAGA CGGATGAAC								
> Q H Q T D E	~	L E V		~		A T V F	QRE	L P E
1901 ACGTGCAGCA AGAGTATGA								AAGCTATTTT
>D V Q Q E Y E	MIE	F S D D	E E M	E V G	E S E E	V T E	E E L	KAIL
2001 GGATACTTGA GTTAGTTTA'	· AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	CCACATATCC
2001 GGMIMCIIGM GIIMGIIIM								OCHOMIMICC
> D T								GCHGHIHICC
	G AGTCTAGAGG	GCCCGCGGTT	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT		
> D T 2101 AGCACAGTGG CGGCCGCTC							TCTACGCGTA	CCGGTCATCA
> D T 2101 AGCACAGTGG CGGCCGCTCG 2201 TCACCATCAC CATTGAGTT	AAACCCGCTG	ATCAGCCTCG	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TCTACGCGTA TGAGTTTGGA	CCGGTCATCA CAAACCACAA
> D T 2101 AGCACAGTGG CGGCCGCTCG 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA	AAACCCGCTG TGCTTTATTT	ATCAGCCTCG GTGAAATTTG	ACTGTGCCTT TGATGCTATT	CTAAGATCCA GCTTTATTTG	GACATGATAA TAACCATTAT	GATACATTGA AAGCTGCAAT	TCTACGCGTA TGAGTTTGGA AAACAAGTTA	CCGGTCATCA CAAACCACAA ACAACAACAA
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA	GACATGATAA TAACCATTAT AAACCTCTAC	GATACATTGA AAGCTGCAAT AAATGTGGTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG	T AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG T GGCGTAATCA T GCCTAATGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA. 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC.	AAACCCGCTG TGCTTTATTT GGCGTTAATCA GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC GATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAAGGCC	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GCGTAATCA GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	CCGGTCATCA CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA
> D T 2101 AGCACATGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA. 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAAG	AAACCCGCTG TGCTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATGAG GCCGGGGAG ACTCAAAGGCG GGGTTGCTGG ACCAGGCGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCCGT ATCAGGGGAT CCCCTGACG TGCGCTCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG ACTGACTCGC ACACACACGCC AGAACATGTG AAATCGACGC CTGCCGCTTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGAG 3101 CTCCCTTCGG GAAGCGTGG	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCG AGCTCACGCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACCT GGGCTGTGTG	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCCAAA GTCCGCCTTT CACGAACCCC
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC GATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGCACGG CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGGG 3201 CCGTTCAGCC CGACCGCTGG	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGCGGGGAG ACTCAAAGGCG GCGTTGCTGG ACCAGGCGTT CGCTTTCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CCTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCCA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCCTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTATC	AAACCCGCTG ATGCTTTATTT AGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGAAAGGCG CTCAAAGGCG ACCAGGCGTT ACCAGGCGTT CCTTTCTCAT GCCTTATCCG TAGGCGGTG	ATCAGCCTCG GTGAAATTTG GGGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCGT TGTAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGGACAC	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTGCGAAAGCA GGTGGCGAAAAGG CTCGCCTTT CACGAACCCC CTGGTAACAC CTGGTAACAC CTGCGCTCTC
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC GATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGCACGG CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGGG 3201 CCGTTCAGCC CGACCGCTGG	AAACCCGCTG ATGCTTTATTT AGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGAAAGGCG CTCAAAGGCG ACCAGGCGTT ACCAGGCGTT CCTTTCTCAT GCCTTATCCG TAGGCGGTG	ATCAGCCTCG GTGAAATTTG GGGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCGT TGTAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGGACAC	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTGCGAAAGCA GGTGGCGAAAAGG CTCGCCTTT CACGAACCCC CTGGTAACAC CTGGTAACAC CTGCGCTCTC
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCCTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTATC	AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGGGGAG ACTCAAAGGCG ACCAGCGGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTG ACAGGCGTG ACAGGCGTG ACAGGCGTG ACAAAGAGGTG AAAAAGAGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG GGAGGTATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TTACAGAGTTC GGTAGCTCTT GGTAGCTCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTGAGTCC TCTTGAGTCG TGATCGGCAA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTCCC CAGTTCGGTG AACCCGGTAA ACCAGTAACTA ACAAACCACC	GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACAC GCTGGTACCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGACGACG GTGGTTTTTT	TCTACGCGTA TGACTTTGGA AACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CCAGGATACCA GCCAGATACGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCACATTA
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAACT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGGTC 2901 CCAGGAACGT TAAAAAGGC 3001 CCCGACAGGA CTATAAAGGC 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGCA GCGACGCTG 3301 GATTAGCAGA GCGAGGGTAG 3401 CTGAAGCCAG TTACCTTCGG 3401 CTGAAGCCAG TTACCTTCGG 3401 CTGAAGCCAG TTACCTTCGG	AAACCCGCTG A TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATCAG GCCTAATCAG CCCACAGGCGT CCCACAGCGTT CCCACAGCGTT CCCACAGCGTT CCCACAGCGTT CCCACAGCGTT CACAAAAAAAAGTC TAAGAAGAGTC CAAGAAGATT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TAACAGAGTTC TCAGAGGTTCT CTTTGATCTT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCCAA TTCTACGGGA TTCTACGGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGCTAA GGCCTAACTA AACCACC TCTGACGCT TCTGACGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACACGACTTCCGTCT GACACGACTTCCGCTCC GCTGCTACACC AGTGGAACGA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTGG AGAACATGTG AAATGACTGC CTGCCGCTTA ACTCACCTG AGAAGACATG ACTCACCTG AGAAGACATTG AGAAGACATG ACTCCACAGG ACTCCACAGG ACTCCACAGGT ACAGGACATTTTTT AAACTCACGT	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTGT GGCCTGTGT GGCAGCAGCCA TATTTGCTAT TGTTTGCAAG TAAGGGATTT	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGGCTACCC CTGGTAACAG TGGCCAGTTT CAGCAACTTG
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC GATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTT 2501 ACCTGCAGG ATGCAAGCT 2601 TAAAGTGTAA ATCGGCCAAC 2601 CAGCAGCAGG TATCAGCTC 2801 CCAGCAGGA CTATAAAGGC 3801 CCCGACAGGA CTATAAAGGC 3001 CCCGACAGGA CTATAAAGGC 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTATT 3401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATC	AAACCCGCTG ATGCTTTATTT AGGTTAATGA GGCGTAATGA GGCGTAATGA GGCGTAGTGG GGCGTGGTG ACCAGGGGA ACTCAAAGGCG ACCAGGGTTACTGG ACCAGGGTTACTGG GCTTTCTCAT GGCTTTATCCG TAGGCGGTG AAAAAGAGTT CAAAAAGAGTT CAAAAAGAGTC	ATCAGCCTGG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTAACT TGACCTACT GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTACTTTC CTTGATCTT CTTTGATCTT TTTTAAATTAA	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTCTG TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT ATCCGCAA ATCACGGGA ATCACGGGA ATCACGGGA AATCAAAGTT	CTAAGATCCA GCTTTATTTG ANAGCAAGTA GTGAAATTG GCGTTGGGCT GCCCCTGACG CCCCCTGACG CCCCCTGACG AACCGGTAA AACCACC CTGACACC CTGACGCCCCTTAACAAACCAC TCTGACGCCTTACACT TCTGACGCTT TTAAATCAAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC ACCGGCTAACGCAGAA AGCATCACAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AAATCGACGC TTGCCGGCTTA ACTCCAACGT ATCGCCACTG AGAAGACAGC AGGAGTTTTTT AAACTCACGT TATGAGTAAA	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCAAG TGTTTGCAAG TTTTGCAAG TTTTGCTAGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGACAG CAGCTATCA
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAAGGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTATC 3401 CTGAAGCCAA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3501 CGCCGCAAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC.	AAACCCGCTG ATCCTTTATTT AGGCTAATCAG GCGTAATCAG GCGTAATCAG CGCGGGGAG ACTCAAAGGCG ACCAGCGTTCTGG ACCAGCGTTTCTCAT GCCTTATCCAG TAGGCGGTG AAAAAGAGTT CAAGAAGATT CCTAAGATCT TATCTCAGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACTATCG TAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCT GTAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACTC TCTTGAGTCC TTGAAGTGC TTGAAGTGT TTTTACGGGG AATCAGAGTT TTCCTTCATT	CTAAGATCCA GCTITATITG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCGCG ATCAGGGGAT CCCCCTGACG CAGTTCGGTC CAGTTCGGTC AACCCGGTAA ACCACCACTAACTA ACAAACCAC TCTGACGCTC TCTGACGCTC TTAAATCAAT CATAGTTGCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC TACCCGGCC AACGAGAA AGCATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA TGGTCGTACACT GGCTACACT GCTGGTACCA AGTGGAACGA CTAAAGTATA TGACTCCCCG	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG AGACATTGC AGACATGGC AGAACATGTG AAATCGAGGC TTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACATGT ATCGCACTG AGAAGACAGC ATCACCACTG AGAAGACAGC ATCACCACTG AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC AGAAGACAGC ATCACCACTT AAACTCACGT TATGAGATAAA	TCTACGCGTA TGAGTTTGGA AACAGTTA TGGCTGATTA CAGCATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CCAGATCAGA CCGGATACCA GGCTGTTGT GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA	CCGGTCATCA CAAACCACAA ACAACAACAA GCACACACAC GCCGGAAGCA CGTGCCAGCT CGTCCGCTC CACCACAAAGG GCTGGCAAAA GTCCGCCTTT CACGAACCC CTGGTAACAG CTCCGCTCTC CACGAACCT CACGAACCC CTGGTAACAG CTCCGCACTT CACGAACT CACGAACT CGGGAGGGCT
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ATCGGCCAAC 2601 CAGCAGGAGGC TATCAGCTC. 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGGC 3101 CCCCTTCAGCC GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTATT 3401 CTGAAGCCAG TTACCTTCG 3501 CCGCCAGAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC. 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3801 TACCATCTGG CCCCAGTGCC 3801 TACCATCTGG CCCCAGTGCC	AAACCCGCTG ATGCTTTATTT AGGTTAATGA GCCTAATGAG GCGTAATGAG GCGCGGGGAG ACTCAAAGGCG GGGTTGCTGG ACTTCTCAT GCCTTATCCG TAGGCGTTG ATGCTTATCCG TAGAGCGTTG CAAGAAGATC CAAGAAGATC TATCTCAGCG TATCTCAGCGT CAAGAAGATC CAAGAAGATC TATCTCAGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG GTAACTATCG GTAACTATCG TTACAGAGTTC CTTTGATCTT TTTAAATTAA ATCTGTTCTAT CGCGAGACCC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAAGTT TCTTACGGGA TTCTACGGGA AATGAAGTT TCGTTCATCACACAC ACGCTCACCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGAC GCTCCCGCGCC CCTCAGATT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGACTACACT GCTGCTACACT GCTGCTACACT AGTGCACCG AGTGGACGA TGAACCGA TGACCCCCG TATCACCACT	GATACATTGA AAGCTGCAAT AAGCTGCAGTA CAATTCCAGCA TTTCCAGCA CTGCCACTGC AGAACATGTG AAATCGACGC CTGCCCGCTTA ACTCGCCACTG AGAAGACATGTG AGAAGACATGTG AGAAGACAGC TCGCCACTG AGAAGACACT ATCGCCACTG AGAAGACAG ATCGCCACTG AGAAGACAGA AAACCACGCA AAACCAGCA AAACCAGCCA	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT AGTTTGGTTGG CAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTGGTCTGA AACTACGATA ACTACGATA ACCACGAAAGGC	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGCGCATA GTCGCGCTTT CAGGAACCC CTGGTAACAG CTGGCTATCT TGGTCATGAC CAGCAGATTA TGGTCATGAG CAGTTACCAA CAGGAGGGCT CCGAGCGCCC CCGAGCGCCCC
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCAGA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC. 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTG	AAACCCGCTG ATGCTTTATTT AGGTTAATGA GCCTAATGAG GCGTAATGAG CGCGTGTTGTGG ACCAGGGGAG ACTCAAAGGCG CGCTTATCCAG ACCAGGGTAGTAG ACCAGGGTTACTAG ACCAGGGTAGTAG ACCAGGGTAGTAG ACCAGGGTAGTAG ACCAGAGAGATC CAAGAAGAGATC CTACAATGATAC CGCCTCCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTATTG GTAATACGGT CGTTTTTCCA AGCCGGTTTT GTAATACGGT TCCCCCTGGA AGCTCACCGT TACAGAGTTC GGTAACTATCG GTAACTATCG GTAACTCTT TTTTAAATTAA ATCTGTCTAT TCGCGAGACCC CCAGTCTATT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TTGAGTCG GATCGGCAA TTCTACGGG AATCAACAGA TTCTACGGC AATTGATCC AATTGATTCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGC ATCAGGGGAT CCCCTGACG ATCAGGGGAT CCCCTGACG AACCAGTAA AACAACCAC TCTGACGT TCAGACTC TCTGACGTC TTAAATCAAT CATAGTTCCAGT GCTCAGATT GGGAAGCTAG	GACATGATAA TAACCATTAAT AAACCTCTAC TATCCGCTCA CACTGCCGC TCCCTGGTCAACGCAGCA AGCATCACAA AGCATCACAA AGCATCACAC CAGCAGCATCACAA GCATCACAA GCATCACAC CAGCAGCATT CAGCACACACT CTGATACGA CTAAAGTATAT TGACTCCCCG TATCAGCAA TATCAGCAAT AGTAAGTAGT	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG ACTGCACTCGACACA CTGCCGCTTA GCTCCAACGT ATCGCCACTG AAATCACGC AGAACATTTT CACCACTG ATCGCCACTG AGAAGGACAG AGAGGACAG AGAGGACAG AACTCACGT AAACTCACGT AAACTCACGT AAACTCACGT AAACCACCA AAACCAGCA AAACCAGCCA TCGCCAGTTA AAACCAGCCA TCGCCAGTTA	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCAAG TAAGGGATTC TCTTGGTCTGA AACTACGATA GCCGGAAGGCA AACTACGATA GCCGGAAGGA TATTTGCAAG	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GCCGGAAGCA GGTGCCAGCT CAGCAAAAGG GGTGGCGAAACCC CTGGTAACAG CTGCGCTTT CAGCAAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAC CAGCAGATTA CGGAGAGGCCAG CAGTTACCAA CGGAGGGGCAA CAGCTACCAAC CAGTTACCAA CAGCAGATTAT CGGAGAGGCACAC CAACGTTGTT
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGGCAGGA CTATAAAGA' 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAG' 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATC' 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC' 3801 TACCATCTGG CCCCAGTGC' 3901 AAGTGGTCCT GCAACTTTC' 4001 GCCATTGCTA CAGGCACTGC'	AAACCCGCTG ATCCTTTATTT AGGCTAATCAG GCGTAATCAG GCGTAATCAG CGCGGGGAG ACTCAAAGGCG CGCGTTTCTCAT CGCTTTCTCAT CGCTTTTCCAT CGCTTTATCAG AAAAAGAGTT CAAGAAGATC CCTAGATCAT TTATCTCAGCG CGCAATGATTAC CCGCCTCCAT CGCGCTCCAT CGCGCTCACT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTAATTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCT GTAACTATCG TACAGAGTTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTTT CCGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTT CGTCGTTTGTTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTACTC GTAGGTACT TCTTGAGTCC TTGAAGTGG TATCCGCAA TTCTACGGG AAATGAAGTT TTCGTTCATC ACGCTCACCG GATGTCACCG GATGTCATC ACGCTCACCG GATGTCTTCCC GTATGGCTTC GTATGGCTTC GTATGCGTCT	CTAAGATCCA GCTITATITG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG TGCGGTCTCCC CAGTTCGGTC AACCCGGTAA ACCACCACTAACTA ACAAACCAC TCTGACGCTC TCTGACGCTC CTTAATTGCC GCTCCAGATT CGGGAAGCTAC TAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTCC	GACATGATAA TAACCATTAT AAACCTCTTAC TATCCGCTCA CACTGCCGCC TACCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGGTCGTC GACACGACT GGCTACACC TAGGTCGTAC GCTGGTACCA AGTAGACGA TTAAGGTACT TGACTCCCCG TATCAGCAAT AGTAAGTATA GGACTCCCCAC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCG AGAACATGTC AAATCGAGCG AGAACATGTG AAATCGAGCC AGCCCTA GCTCCAAGCT ATCGCCACTG AGAAGACAGCT ATCGCCACTG AGAAGACAGCT ATCACCACTG AGAAGACAGCA TAGCTACAGCT AAACCAGCCA TATCACGAGTAAA TCGTGTAGAT AAACCAGCCA TTCGCCAGTTA GATCAAGGCG	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAGCATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CCGGATACCA GGCAGTCAGA CCGGATACCA TATTTGGTAT TGTTTGCAAG TAGGGTTTCGTCTG ACTTACGATA CCGGAAGGG ATGTTTGCAAG ACTACGATA CCGGAAGGG ATGTTTTCGA	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GCAGCAGCT GGTCCGGCT CGTTCGGCTG GGTCGGCAAA GCCCCCTT CAGCAAAAGG CTCCGCCTTT TAGGCTACAGCACCC CTGGTAACAG CTCCGCTCTT TGGTCATACAG CAGCAGATTA TGGTCATACAA CAGTTACCAA CAGGAGGCCC CCGGAGGGCCC CCGGAGGCCCAG CAACGTTGTT CCCCCATGT
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTATGTTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ACCTGGGC 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTC. 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGGC 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 401 CGCGCAGAAA AAAAGGATCTC 3501 CGCGCAGAAA AAAAGGATCTC 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA' 4001 GCCATTGCTA CAGGCACCG 4101 TGTGCAAAAA AGCGGTTAGG	AAACCCGCTG ATGCTTTATTT AGGCTTATTG GCCTAATGAG GCGTAATGAG GCGTGGGGAG ACTCAAGGCG GGCTTGCTG ACTCTCTCTCT GCTTTCTCAT GCCTTATCAG TAGAAGAGT CAAAAAAGATT CAAGAAGAT CAAGAGT TCAAGATCCT TATCTCAGGG GCAATGATAC CCGCCTCCAT GGCATTCACGC TCATCTCAGCC TCCTTCAGCC TCCTTCAGC TCCTTCAGCC TCCTTCAGCC TCCTTCAGCC TCCTTCAGC TCCTTCAGCC TCCTTCAGC T	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCT GTAACTATCG GTAACTATCG GTAACTATCG TACAGAGTTC CTTTGATCTT TTTAAATTTAA ATCTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCCGATCGT CTCGATCGT CTCCGATCGT CTCGATCGT CTCCGATCGT CTCGATCGT CTCGATCGT CTCCGATCGT CTCGATCGT CTCCGATCGT CTCGATCGT CTCGATCT CTCGATCGT CTCTCT CTCGATCT CTCGATCT CTCT C	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TACTCCGCAA TTCTACGGGAA TTCTACGGGAA TTCTACGGGAA TTCTACGGGAA TTCACTCACC AATTGTTCCC TGTAAGGCTCC TGTAAGAGTT TCTTCATCACCC TTCAACAGAGT TTCATCATC TGTCAGAAGT TTCCTTCATC	CTAAGATCCA GCTITATITG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATTGCC AGTCCAGATT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACAGGACTACCAC TGTGCTGTC GCTGGTAGCG AGTGGACGA GTAGACGA TGACTCCCCG TATCAGCAAT AGTAAGTATA AGTAAGTATA GGTACCCAC TATCAGCACA CAGTGTTATC	GATACATTGA AAGCTGCAAT AAGCTGCATCA CAATTCCACAC TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACAGT ATCGCACTG AGAAGATTTTTT AAACTCACGT TATCAGTAAGATAAA AAACCAGCCA TCGCCAGTTA GAAGACAG ATCAGCAGTAAAACCAGCCA ATCAGCAGTAAAACCAGCCA ACTCAATGGTTAAAAGCAG ACTCAATGGTTAAAAGCAG ACTCAATGGTTA	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCT AGTTTGGTATT GGTCTGA TAAGGGATT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA CCGGAAGGG ATAGTTTGCA ATGCCAGCA ATGCCAGCA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CAGGAACCC CTGGTAACAG CTGGCATTA TGGTCATGAC CAGCAGATTA TGGTCATGAC CAGGAGGCC CCGAGCGCAC CAGGGGCTT CCCCCATGT TCCCCCATGT
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC GATGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATGGCCCAA 2801 CGGCGAGGA CTATAAAGA 2801 CCGGCAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGGTCAGCC CGACCGCTG 3301 GATTAGCAG GCAGGATAT 3401 CTGAAGCCAG TTACCTTCG 3501 ATTATCAAAA AGGATCTG 3701 TGCTTAATCA GTGAGGCACG 3801 TACCATCTGG CCCCAGTGC 3801 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4101 TCTTACTGTC ATGCCATCC	AAACCCGCTG ATGCTTTATTT AGG GCGTAATCAG GCGCTAATCAG GCGCGGGAG ACTCAAAGGCG GCGTTGCTGG ACCAGCGTTCTTG GCTTTCTCAT GCTTTCTCAT CAGAAAGAGTT CAAAAAGAGTT CAAAAAGAGTT CAAAAAGAGTT CAAAAAGAGTT CAAAAAGAGTT CAAGAAGATC TATCTCAGCG GCAATGATAC CCTCCATCCGTC CGCCTCCAT GGGTTCAGGC GCATTCCGTC TTCTTCGGTC TTCTTCGGTC TTCTTCGGTC TTCTTTCGGTC TTCTTTCGGTC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTATTG GTAATACGGT CGTTTTTCCA AGCTCACCT GTAACATATC GTAACATATC GTAACATATC GGTAGCTCT TTTAGAAGTTC TTTTAAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG TTCTGTCTTT	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAGTGGT ATCCGCAA TTCTACGGGA TTCTACGGGA TTCTACGGG AATTGATCT TCGTCATC AAGTTACTC CGTATGGCTTC TCTGAGTGGT TCGTCACCG GTAGGGTACT GTCAGAAGT TGTCAGAAGT TGTCAGAAGT GGTGAGTACT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG CGCCTACAGCT CAGCTCTCC CAGTTCGGCG AACCACC TCTGACGCTC TCTGACGCTC TCTGACGTC CGTCAACTA CATAGTTCGC GCTCCAGATT GGGAAGCTAG ATCAGCTTC AAGTTGCCC CAGATTGCCC CAGACCAACTC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGACC TAGGTCGTC GACACGACTT GACACACT GCTGGTACGA CTAAAGTATT TGACTCCCG TATCAGCAT TATCAGCAT TATCAGCAT CATTATCAGCA CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTATC CAGTTTATC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA GCTCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCAGCTATA AACTCACGT ATACAGGTAAA TCGGTAGATAA TCGGCAGTTA GATCAAGGCG ACTCAATGGT AATCAAGGCG ACTCAATGGT TAGATATATG	TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTT GGGCTCTGTG GCACCACCA TATTTGCAAG TAATTGCATA CTTAGCATA CTTAGCATA CATTACGATA ACTACGATA ACTACAATA CCGGAAGGC ATAGTTTCC AGTTACATGA AGTTACATGA AGTTACATGA AGTACACATA AGCACCAC GGCGACCCAC GGCGACCCAC CGCGCACCCAC	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT GGTTCGGCTT CACGAAAAGG GTGGCGAAACG CTGGTAACAG CTGCGCTTT CACGAGATCA CAGCAGATTA CGGTATGAG CAGTTACCAA CGGGAGGGCT CCGAGCGCAG CAACGTTGT TCCCCCATGT TCCCCCATGT TGCATAATTC TTGCTCTTGC
D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGGGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGGCAGGG CTATAAAGGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAGT 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA' 4001 GCCATTGCTA CAGGCATCG' 4101 TGTGCAAAAA AGGGTTAG' 4201 TCTTACTGTC ATGCCATCCC 4301 CCGGCGTCAA TACCGGGATAG'	AAACCCGCTG ATCCTTTATTT ACAGC GCGTAATCAG GCCTAATCAG GCGCGGGAG ACTCAAAGGCG CGCGTTTCTGG ACCAGCGTTCTGG ACCAGCGTTCTGG AAAAGAGATC CTAAAGAGATC TATCTCAGCG AAAAGAGATC CTATCCAGCG GCAATGATAC CGCCCTCCAT GGGTGCAGCG TACCGCGCCAT TACCTC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAATC GTAATACGT CGTTTTTCCA TCCCCCTGGA AGCTCACCT GTAACTACTC GTAACTACTC GTAACTACTC GTAGCTCT TCTTGATCTT TTTAATCTA ATCTGTCTAT CGCAGAGCCC CCAGCTCTAT TCGGCTCGTTTC CTCGGATCGT TCCGTTGGTCTT CTCGTCGTTTG CTCGATCGTTTC CATAGCAGAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TAGAGTGGT GATCCGCAA TTCTACGGGG AAATGAAGT TTCGTTCATC ACGCTCACCG GTATGCTCACCG GTATGGCTTC TGTCAGAAGT CTTTAAAAGT CTTTAAAAGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG TGCGCTCTCCC CAGTTCGGT AACCCGGTAA ACCAACTA ACAAACCAC TCTGAAGCTC CTCAACATC CCTCAACAT CATAGTTGCC GCTCCAGATT ATTCAGCTCC AAGTTGGCG CAACACAC CCCAAGTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGACA AGCATCACAA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGTC GACACGACT GCTGGTACCAC GCTGGTACCA CTACAGACT TGACTCCCCG TATCAGCAAT TGACTCCCCAAC CAGTGTTATC GCTCAGCAAT TGACTCCCAAC CAGTGTTATC AGTAGTATAT GACTCCCCAAC CAGTGTTATC CATCTGAGAA GGAAACGTT	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCAGTCG AGAACATGTG AAATCGAGCG AGAACATGTG AAATCGAGCG AGACACTGT ACCCCACTG ACTCCCAGCG ATCCCCAGCG ATCACCACTG ATCACCACTG AGAACACCAC ATCACCACTG ATCACCACTG AAACCACCAA TCGTTAGAT AAACCAGCCA TCGCCAGTTT GATCAAGGCG ACTCATGGTT TAGGTTATGC CTTCCGGGGCG CTTCGGGGCG	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAGCATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT CAGCAGACCA TCAGTCAGA CCGGATACCA TATTTGGTAT TGTTTGCAAG TAGTCCGAA ACTACGATA CCGGAAGGG ATACTTTGCTAT ACTGCGTAGAC ACTACCGTA ACTACCGTA ACGCACCAC AGTTACATGA ATGCCACCAC AGGCACCAC AGGCACCAC AAAACTCTCA	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GCACCACCACCACCACCACCACCACCACCACCACCACACAC
> D T 2101 AGCACAGTGG CGGCCGCTCC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTATGTTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGGC 3101 CCTCCTTCGG GAAGCGTGG 3201 CGGTTCAGCC CGACCGCTG 3301 GATTAGCAGA CGAAGGTAT 401 CGGAGGAGTAT 402 GCGAGGAAA AAAAGGACT 3601 ATTATCAAAA AGGACTCTCG 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTAC 401 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCCC 4301 CCGGCGTCAA TACGGGGTAG 4201 CCGGCGTCAA TACGGGGTTAG 4401 CGCTGTTAGA ATCCAGTTCC 4101 CGCGCTCCAA TACCGGGATAA 4401 CGCTGTTAGA ATCCAGTTCC	AAACCCGCTG ATGCTTATTT ATGCTTATTT GCCTAATGAG GCGTAATCAG GCGTAATCAG CGCGGGAG ACCAGGGGAG ACCAGGGGTACTCG ACCAGGGTTCTGG AACAGGCGTT CGCTTTCTCAT CGCTTTCTCAT CAAGAAGATC AAAAAAGATT CAAGAAGAT CCGCCTCCAT CGCGTTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGC TATCTCAGGCGTGA TATCTAGGGGTGC TCCTTCGGTC TATCTCAGGGCTAGAGAGATACCCAGCCAAAAAAAAAA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGGCTATAGC TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCCTGGA TACAGAGTTC TTAAATTAA ATCTAA TCTTTGATCTT TTTAAATTAA ATCTGTCTAT CGCGAGACCC CCACTCTATT CGTCGTTTG CTCCGATCGT TTCTGTGACT CATAGCAGAC CATCGTCTCTTCCTTGTGACT CTCTGTGACC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGGTCCGC AGCTCCCTCG GATAGGTATC TCTTGAGTCC TTGAAGTGGT TCTACGGGA AATGAAGTT TTCCTTCATC ACGCTCACCG AATTGTTGCC GTATGGCTTC TGTCAGAAGT TTCTCATC ACGCTCACCG AATTGTTGCC TGTCAGAAGT TTTCATAAAAGT CTTTAAAAAGT CAACTGATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCGCC ATCAGGGGAT CCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCC TCTAACTA TAAATCAAT TAAATCAAT AATAATCAAT AATTGCC GCTCAGATTGCCC CACCAGATC CAGCCAACTC CAGCCAACTC CACCAACTC CACCAACTC CACCAACTC CACCAACTC CTCACCAACTC CACCAACTC TCACAACTCT TCACCATCTT TCACCATCTT TCACCATCTT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TACCTGGTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTACCT GCTACACT CGCTACACT TGACTCCCCC TATCAGCAAA GGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA CAGTCTCAAC CAGTGTTATC ATTCTGACAA CAGTGTTATC ATTCTGACAAA CTTAACTTTCTACATT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTGG ACTGACTGGC AGAACATGTG AAATGGAGGC CTGCCGGCTTA AGAGGACACTG AGAAGATTTT AAACTCACGT TATGAGTAAA TCGTGAGTAAA TCGTGAGTAAA AACCAGCCA TCGCCAGTTA CATCAGGTAGAT AAACCAGCCA TCGCCAGTTA CATCAGGGAGA CATCATGGTTATGGTATAAA CATCAAGGCG ACTCATGGTT TAGGTATAGC CATCATGGTT TAGGTGTATAGC CATCATGGTT TAGGTGTATGC CTTCGGGGGG CACGCTTTCT	TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGAATACCT AGTTTGGTTGT TGTTTGCAAG TAAGGGATTCT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATACTTTCGTAT ACGGAAGGG ATACTTTCAGAG ATACTACGATA ACTACGATA ACGACGAGG ATACTTTCCG AGTTACAGA ATGCAGCAC GGCGACCGAG AAAACTCTCA GGGTGACGAC	CCGGTCATCA CAAACCACAA ACAACAACAA ACAACAACAA GCAGCAAGCA
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGGCCAA 2801 CGGCGACGGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGGTC 2901 CCCGCACAGGA CTATAAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTCAGGC CGACCGCTGG 3201 CGGTCAGGC GTACCACGCTGC 3501 CGCGCAGAAA AAAAGGGATC 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTG GCAACTTTA 4001 GCCATTGCT GCAACTTTA 4001 GCCATTGCT AGCGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCGCAAA AACGGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA	AAACCCGCTG ATGCTTTATTT AGGGGTATCAGG GGCGTAATCAG GCCTAATCAG GCGCGGGAG ACTCAAAGGCG GCGTTGCTG ACCAGCGTTCTGG ACCAGCGTTCTCAT CGCTTATCCAG ACAAAGAGATC CAAAAGAGATC CATACCAG CCGCCTCCAT GCAATGATCACCAG TCCTCCAT CGCGTTCCAGC GCAATGATACCA ATGCAACGCG AAAAAGAGTT ATACCCGCCAA ATGTAACCAG ATGTAACCAG AGGGTTAACCAG ATGTAACCAG AGGATTAACCAG AGGATAACGAG AGATAAAGGGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTATTG GTAATACGGT CGTTTTTCCA AGCTCACCT GTAACATATC GTAACATATC GTAACATATC GGTAGCTTC TTTAGATCTT TTTTAAATTAA ATCTGTCTAT TCGTCGATTG CCAGTCTATT TCGTCGTTTG CTCCGATCGT CTTCGATCGT CTTCTGTCATC CAGACCC CAGTCTATT CGTCGATCGT CTCCGATCGT CTCCGATCGT CATAGCAGAA CTCGTGCACC CACACCTGACC CAGACCGACACACGAAA CTCGTGCACC CACACGGAAA	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT AATCACAGGA ATTCTACGGCA ATTCTACGGCA ATTCTACGGC AATTGTTCAC GGAGTACT CGTAGAGGT TCTCAGAGGT TTCAGAGGT TCTCAGAGGT CGTAGGAGT CGTAGAGGT CGTAGAGGT CTTTAAAAGT CAACTGATCT TGTTGAAAGT CAACTGATCT TGTTGAAATAC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ACCCCCTGACG ATCAGGGAT CCCCCTGACG ACCAGCTC CAGTTCGGCG AACCAGCT TCTGACGCTC TCTGACGCTC TCTGACGCTC CGTCAACTA CAAACCAC CCCCCAGACT GCCCAGACT GCAACCTACA ATCAGCTCC AAGTTGCCC CACAAGTC CAACCAACTC CAACCAACTC CAACCAACTC CAACCAAC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCGCCAGTTA GATCAAGGCG ACTCATGATGT TAGTGTATATG CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTC CGGTCGGCTCT CACGAACCC CTGGTAACAG CTGCGCTTC CACGAACT CCGAGCACA CACGTTACACA CACGTTACT TCCCCATGT TCCCCATGT TCCCCATGT TTCCCCTTGC AGGATCTTAC AGGATCTTAC AGAACTTTAC AAACAGGAAG AGGTTATTGT
D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTT 2501 ACCTGCAGGC ATGCAAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA TTGCGCCAA' 2801 CGGCGAGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGGCAGGA CTATAAAGA' 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAG' 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATCT' 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC' 3801 TACCATCTGG CCCCAGTGC' 3901 AAGTGGTCCT GCAACTTTA' 4001 GCCATTGCTA CAGGCATCG' 4101 TGTGCAAAAA AGGGTTAG' 4201 TCTTACTGTC 4201 TCTTACTGTC 4201 TCTTACTGTC 4301 CCGGCGTCAA TACCGGGATAG' 4401 CGCTGTTGGA ATCCAGTTC' 4501 GCAAAATGCC GCAAAAAAGG 4601 CTCATTGGC GCAACATTC' 4501 GCAAAATGCC GCAAAAAAGGGTTC' 4501 GCAAAATGCC GCAAAAAAGGC	AAACCCGCTG A TGCTTTATTT A GAGGTTAATGA GCCTAATGAG GCGCGGGAG A CTCAAAGGCG CGCGTTTCTCAT CGCTTTCTCAT CGCTTTCTCAT CGCTTTCTCAT CAAAGAGATC ACAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCACATCATCACGC CTCAATGATACCAC CTCAATGATACCAC CTCAAAAGAGATC TAACCCACGCA ATGTAACCCA ATGTAACCCA ATGTAACCCA CGAATAATACCCA CGAATAAAAGGC TCTAAAAAGGCT TAACTCACCCC TTCAATTAAACCCA CGAATAAAAGGCT TAAAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATGATAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACCTAATCG GTAATACGGT CGTTTTTCCA AGCCGGTTTTTCCA TCACCACTGGTAACTACTC GTAACTACTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTTT TCGTCGTTTT CCCGATCCTT TCCTCGATCGTTT CTCGATCGTTTC CATAGCACAC CCCGACACCC CACACCGACAC CCGACACCC CACACCGACAC CCGACACCC CACACCGCACA TCCGTGCACC GACACGGAAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TTGAAGTGGT AATCACAGA TTCTTACGGG AATCATCACC GTAGGTTT TTCGTTCAT ACGCTCACCG GTATGGCTTC TGTCAGAAGT TGTCAGAAGT TGTTCAGAAAGT CACTGATCT TTTTAAAAGT CACTGATCT TGTTGAATAC AACAAATAAG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG TGCGCTCTCCC CAGTTCGGT AACCCGGTAA ACCAACTA ACAAACCAC TCTGACGTC CTCAACTA CATAGTTGCC GCTCCAGATT GGGAAGCTA ATTCAGCTC AAGTTGGCG CACCAAGTC CAACTACTTTGCCG CACCAAGTC CAACTACTTTGCCG CACCAAGTC CAACTACATT CAGCATCTT TCAGCATCTT TCAGCATCTT TCATACTCTT GGTTCCCGCG	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCACAGGCG ACTCATGATT TAGTGTATAGC CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTC CGGTCGGCTCT CACGAACCC CTGGTAACAG CTGCGCTTC CACGAACT CCGAGCACA CACGTTACACA CACGTTACT TCCCCATGT TCCCCATGT TCCCCATGT TTCCCCTTGC AGGATCTTAC AGGATCTTAC AGAACTTTAC AAACAGGAAG AGGTTATTGT
> D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGGCCAA 2801 CGGCGACGGG TATCAGGTC 2901 CCAGGAACCG TAAAAAGGGTC 2901 CCCGCACAGGA CTATAAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTCAGGC CGACCGCTGG 3201 CGGTCAGGC GTACCACGCTGC 3501 CGCGCAGAAA AAAAGGGATC 3601 ATTATCAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTG GCAACTTTA 4001 GCCATTGCT GCAACTTTA 4001 GCCATTGCT AGCGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCGCAAA AACGGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA TACCGGGTAA 4401 CGCGCTCAA	AAACCCGCTG A TGCTTTATTT A GAGGTTAATGA GCCTAATGAG GCGCGGGAG A CTCAAAGGCG CGCGTTTCTCAT CGCTTTCTCAT CGCTTTCTCAT CGCTTTCTCAT CAAAGAGATC ACAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCAAAAGAGATC CTCACATCATCACGC CTCAATGATACCAC CTCAATGATACCAC CTCAAAAGAGATC TAACCCACGCA ATGTAACCCA ATGTAACCCA ATGTAACCCA CGAATAATACCCA CGAATAAAAGGC TCTAAAAAGGCT TAACTCACCCC TTCAATTAAACCCA CGAATAAAAGGCT TAAAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATAAAGGCC TCAAATGATAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACCTAATCG GTAATACGGT CGTTTTTCCA AGCCGGTTTTTCCA TCACCACTGGTAACTACTC GTAACTACTC GGTAGCTCTT CTTTGATCTT TTTAAATTAA ATCTGTCTTT TCGTCGTTTT CCCGATCCTT TCCTCGATCGTTT CTCGATCGTTTC CATAGCACAC CCCGACACCC CACACCGACAC CCGACACCC CACACCGACAC CCGACACCC CACACCGCACA TCCGTGCACC GACACGGAAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT GGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGC TAGAGTGGT AATCAAGAGT TTCTACGGG AATCAATCAT TCGTTCAC GTAGGTATCT TCGTCACC GTAGGTTT TTCGTTCAT ACGCTCACCG GTATGGCTTC TGTCAGAAGT TGTTCAGAAGT CACTGATCT TTTTAAAAGT CACTGATCT TGTTGAATAC AACAAATAGG	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGCT ATCAGGGGAT CCCCCTGAAG TGCGCTCTCCC CAGTTCGGT AACCCGGTAA ACCAACTA ACAAACCAC TCTGACGTC CTCAACTA CATAGTTGCC GCTCCAGATT GGGAAGCTA ATTCAGCTC AAGTTGGCG CACCAAGTC CAACTACTTTGCCG CACCAAGTC CAACTACTTTGCCG CACCAAGTC CAACTACATT CAGCATCTT TCAGCATCTT TCAGCATCTT TCATACTCTT GGTTCCCGCG	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TAGTCCGAC TAGGTCGTC GACACGACTT GACACACACT CTAGACACAC TAGATACACA CATAAAGTATAT GACTCCCCG TATCAGCAT TACTAGCAT TACTAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCTAGCA CAGTGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTTCAC CCTTTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG ATCACCACTG AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGTT AAACTCAGGT AAACTCAGGT ATCACAGGCG ACTCATGATT TAGTGTATAGC CTTCGGGGCG CAGCGTTTCT ATTTATTGAA	TCTACGCGTA TGACTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACGA CCGGATACGA TATTTGGTT GCACCACCAC GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TATTGGTTA AACTACGATA CCCGGAAGGG ATTACATGA ATGCTACACAC AGTTACATGA AACTCTCA GGCGACCGAG AAAACTCTCA GGGTGACCAA GCATTTACA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCTC CGGTCGGCTCT CACGAACCC CTGGTAACAG CTGCGCTTC CACGAACT CCGAGCACA CACGTTACACA CACGTTACT TCCCCATGT TCCCCATGT TCCCCATGT TTCCCCTTGC AGGATCTTAC AGGATCTTAC AGAACTTTAC AAACAGGAAG AGGTTATTGT

# pMT-HisFlag-DmSNAP43 mut#11 STOP

1 TCGCGCGTTT CGGTGATGA	CCTCAAAACC	тстсасасат	GCAGCTCCCG	GAGACGGTCA	CACCTTCTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGGGT								
201 CCGCACAGAT GCGTAAGGA								
301 TACGCCAGCT GGCGAAAGG								
401 TGAATTAATT CGTTGCAGG								
501 GGCCCCCAC CGCCCACCG								
601 CAAGTCCCCA AAGTGGAGA								
701 AGAGGTGAAT CGAACGAAA								
801 AAATCAAGTG AATCATCTC	A GIGCAACIAA	AGGGGGGAIC	TAGATCGGGG	TACTAAGATT	> M H		H H T D	V K D
001 0200200202 20000200		3 mamma 3 aa	3 CMCCMCCC3	aamaamaa a		н н н		- 10 -
901 GACGATGACA AGGGCACTA > D D D K G T								N C E
1001 TCGAGGTGTT CTGCCGGTG			D C W E			R L V N		
>F E V F C R C 1101 GCATGTGGCC AAGCGACTG	W R E	L Q L Q			T N H T	E V I		L A A L
	S C S R							
> H V A K R L 1201 ATCTACTACA AGCAGCCCA			G D V	F P A S		R I G	G F F L	L Y V
> I Y Y K O P			E V S P	R T W		T D Y A		R K D
~					~			
1301 GTCCGGAGCG GAAGGACAC								
>S P E R K D T	H Q I	A Y M L		T Q E	Q A F R		L D Y	C Q G L
1401 GGACAATCTG GTGGACTAC  > D N L V D Y	D R V E						O O R A	
			G A K			M Q K	~ ~	N G V
1501 AGTCTCACAT ACGAACTGG								
> S L T Y E L		A L D						K Q L
1601 CGGCTGGTCA TGAGCACGC								
>A A G H E H A		S Q I F			F A D I		A A A	
1701 TCCAGATGAG AAATGCACC > P D E K C T								R E P
	TTST	~	L E V			A M Y		
1801 CAACACCAGA CGGATGAAC								L P E
> Q H Q T D E 1901 ACGTGCAGCA AGAGTATGA			N E T Y			A T V F	Q R E	
1901 ACGIGCAGCA AGAGIAIGA	J AIGAIIGAGI							
- D 11 O O B 11 B	м т п							
>D V Q Q E Y E	M I E	F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTA		F S D D	E E M	E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTA  > D T	r aaaactttta	F S D D CATAATTAAA	E E M TAACTAGCAT	E V G TTTTGCGCGA	E S E E TGTGATCTTG	V T E TTTATCTGAA	E E L GGGCAATTCT	K A I L GCAGATATCC
2001 GGATACTTGA GTTAGTTTA  > D T  2101 AGCACAGTGG CGGCCGCTC	r aaaactttta g agtctagagg	F S D D CATAATTAAA GCCCGCGGTT	E E M TAACTAGCAT CGAAGGTAAG	E V G TTTTGCGCGA CCTATCCCTA	E S E E TGTGATCTTG	V T E TTTATCTGAA CGGTCTCGAT	E E L GGGCAATTCT TCTACGCGTA	K A I L GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTTTA  > D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT	F AAAACTTTTA G AGTCTAGAGG F AAACCCGCTG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA GTTAGTTTA  > D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA	F AAAACTTTTA G AGTCTAGAGG F AAACCCGCTG A TGCTTTATTT	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	K A I L GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA GTTAGTTTA  D T 2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT	F AAAACTTTTA  G AGTCTAGAGG F AAACCCGCTG A TGCTTTATTT C AGGTTCAGGG	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA GTTAGTTTA  D T  1010 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT	AAAACTTTTA  G AGTCTAGAGG AAACCCGCTG ATGCTTTATTT C AGGTTCAGGG GGCGTAATCA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA GTTAGTTTA  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG	AAAACTTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	E E M TAACTAGCAT CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTTTA  > D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA	AAAACTTTA AGG AGCCGCTG ATGCTTTATT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGGAG ACTCAAAGGCG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TCTTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTTTCCGC ATCAGGGGAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACTA TTTCCAGTCG ACTGACTCGC AGAACATGTG	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2001 GGATACTTGA GTTAGTTTA  D T  2101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC	AAAACTTTA AGTCTAGAGG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCGGGGGAGA ACTCAAAGGGG CCCTAAAGGGG CCCGCGGGAGGAGACTCCTGCTGCCGGGGGGGGAGACACCCCGGGGGGGG	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCCGGTTTG GTAATACGGT CGTTATTCCA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAAA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAAGCTC 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAAA	F AAAACTTTA  G AGTCTAGAGG F AAACCGCTG A TGCTTTATTT C AGGTTCAGGG F GGCGTAATCA F GCCTAATGAG C GCGTGGGGAG A CTCAAAGGCG C GCGTTGCTGG F ACCAGGCGTT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCTTTTCCA TCCCCCTGGA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTTT CGTATTCGTG CACATTAATT CGTATTGGC TATCCACAG AGCTCCCTCG	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCCTGACG TGCGCTCCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCGC TCCCTCGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA TTTCCAGTCG ACTGACTCGC AGAACATCTG AGAACATCTG CTGCCGCTTA	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCCAAA GTCCGCCTTT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCAG GCCTAATGAG CGCGGGGAG ACTCAAAGGCG CGCGTTGCTGG ACCAGGCGTT CGCTTTCTCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAAGC TGAGCTAATTG GTAATACGGT CGTTTTTTCCA TCCCCCTGAA AGCTCACGCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCCG AGCTCCCCG GTAGGTATCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCCGC CAGTTCGGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC ACCGGCTCA ACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGACC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT CAATTCCACCA TTTCCAGTCG ACTGACTGGC AGAACATGTG AAATCGACGC CTGCCGCTTGCCGCTTGCCGCTTGCCGCTTGCCACCT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATACGA GGAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTG CAGCAAAAAGG GGTGGGCAAA GGTCGGCCTT CACGAACCCT CACGAACCCCTT CACGAACCCC
2001 GGATACTTGA GTTAGTTTA  D T  2001 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ATCCGCCAA 2801 CGGCGACGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAAGG 3001 CCCGCTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGGGTAATCAG GCCTAATGAG CCCGGGGGGAG ACCAGAGGGG ACCAGAGGGGT ACCAGGGGT ACCAGGGGTT CGCTTTCTCAT CGCTTTATCCG	F S D D CATAATTAAA  GCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGA TGGTCATAGC TGAGCTAACT AGCGGTTTG GTAATACGGT TCCTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATCG	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGGTTTTT TGTTTCCTGT CACATTAATT CGTATTCACAGA TAGCCTCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGGTCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTC TAGGTCGTC TAGGTCGTC TAGGTCGTC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATCGACTA CAATTCCACA ACTACTCGC ACTACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTCGTGTG GCACCAGCCA	K A I L GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GCTGCGAAAGGA GTCGGCCTTT CACGAACCCC CTGGTAACAG
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCAGCCA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 2901 CCCGACAGGA CTATTAAAGA 3101 CTCCCTTCGG GAACCGCTG 3201 CGGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT	AAAACTTTA AGTCTAGAGG ATGCTTTATTA AGGTTCATGG GGCGTAATCA CGCGTAGGG CGCGGGGGG ACTCAAAGGCG CGCTTCTCG ACCTTATCAG CTTATCAG CTTAGCGGTGC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TAACATATCG TAACATATCG TAACATATCG TAACATATCG TAACAGGTT TACAGAGTTC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACACTAATT CGTATTGGG TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TTGAAGTGGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA AGCCGGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCACAC ACTGACTCGC AGAACATCTG AGAACATCTG ACTGCCGCTTA GCTCCAAGCT ATCGCCACTC ACGAGGACACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGGACACC ACGAGACACC ACGAGGACACC ACGACC ACGAGGACACC ACGAGGACACC ACGACC ACGAGGACACC ACGACC ACCAC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCTGTTGG GCGCTGTTG GCACCAC TATTTGGTAT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGGCTC CAGCAAAAGG GGTGGCGAACA GTCCGCCTTT CAGCAACCC CTTGGTAACAG CTGGCTCTCG CTGGTAACAG
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGAG ACTCAAAGGCG CGCTGCTGCTGG ACCTTTCTCAT CGCTTTCTCAT CGCTTATCCAG CTTATCCAG CTTATCCAG AAAACAGTT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT GTAATACGGT CGTTTTTCCGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGCTCCGC GTAGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTCC TTGAAGTCC GATCCGCGCA	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCGTTCGGCT ATCAGGGGAT CCCCTGACG ATCAGGGGAT ACCAGGTAACTA ACCACACACA ACAAACCACC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCGCTCA CACTGCCGC CACGCAGA AGCATCACAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT CGGCTACACT CGGCTACACT CGGCTACACC CCTGGTACCC CCTGGTACCC CCTGGTACCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTA CAATTCCAGTCG ACTGCACTGA AATCGAGCC ACTGCCGCTTA GCTCCAGCT ACGCCACTG ACGCCACTG AGAGGACAGC ACGCCACTG ACCCCACTG ACCCCACTG ACCCCACTG ACCCCACTG ACCCCACTG ACCCACTG ACCCCACTG ACCCACTG ACCCCACTG ACCCACTG ACCCCACTG ACCCACTG ACCCCACTG ACCCACTG ACCCACTT ACCCACTG ACCCACTG ACCCACTG ACCCACTG ACCCACTG ACCCACTG ACCCACT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GCAGCTGCGG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAAA TGATCAGTCG CCGGAAGCA CGTTCCGCTG CAGCAAAAGG GGTGGCGAAA GGTCGCCTTT CACGAACCC CTGGTAACAG CTGCCGCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATC	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTATATT GCCTAATGAG GCGGGGGAG ACCAGAGGGGGGGGGG	F S D D CATAATTAAA GCCCGGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT TAACAGAGTTC GGAGAGTTTC GGAGAGTTC GGTAGTTTCCA TCCCCTTGAACTATCG TTAACTATCG TTAACTATCG TTTTGATCTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTAATTAGGC TATCCACAGA AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAAGTGGT GATCCGCCA TTCTTCGGGG TATCCGCCA TTCTTCGGGG TTCTTCGGGGGATTCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT TGCGCTTGCCC CAGTTCGGT AACCCGCTAAC GCCTAACA GCCTAACA GCCTAACA CACCGTAA	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGGCC TCCCTCGGCC AACGCAGGAA TGTTCCGACC TAGGTCGTC GACACGACT TCGGCTACACT GGCTACACT GGTGGTACGAC AGTTGGAACG AGTTGGAACG AGTTGGAACG AGTTGGAACG	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGCAA AAATGTGGTA CAATTCCAACA ACTTCCACA ACTGACTGG ACTGACTGG ACTGACTGG AGAACATGTG GTCCAAGCT ATCGCCACTTA GCTCCAAGCT ATCGCCACTTA AGAAGGACAG GTGGTTTTT AAACTCACGTTT AAACTCACAGTT	E L GGGCAATTCT TCTACGCGTA TGACTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGAATACCT GGGCTCGGT GGGCTCGTGTG GGACTCGTGTG GCACCAGCCA TATTTGCTAT TGTTTGCAAG TAAAGGGATTT	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCGGCTTT CACGAACCC CTGGTAACAC CTGGCTAACAC CTGGCTACT CAGCAACTTA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2401 TAGAGTTAAA AGCCTGCGGG 2701 GCATTAATGA ATCCGCCAA 2801 CGGCGAGCGG TATCAACTC 2901 CCAGGAACCG TAAAAAGGCT 2901 CCCGACAGGA CTATAAAAA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3201 CGGTTCAGCC GGACCGCTTG 3401 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCAGA AAAAGGATC 3601 ATTATCAAAA AGGATCTTC	AAAACTTTA  AGTCTAGAGG ATGCTTATTT AGGTTCAGG GGCGTAATCAG CGCTAATCAG CGCTTACTG ACCTTACTG ACCTTACTG ACCTTACTG ACCTTACTG ACCTTACTG ACCAGGGGT ACCAGAGGGGT ACCAGAGGGGT ACCAGAGGGGT ACCAGAGGGGT ACCTTACCG ACCAGAGGGGT ACCAGAGGGGT ACCAGAGGGGT ACCAGATCC ACCTAGATCCT ACCTTAGATCCT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGCTCT TCTGAACTT CTTGATCTT TTTTAAATTAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT TACCACAGA TAGCGTCAGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTGAGTGGG ACTCCGCAA TTCTACGGG ATCCGCCAA TTCTACGGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGAAATTGT GCGTTGCGCT GCCTCTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTA ACCACGGTAA ACCACACC TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TCTGACGCT TTAAATCAAT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCGCC TTCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATCTG AAATCGACGC TGCCGCTTA GCTCCAAGCT ATCGCCACTG ACGAGGACAGCT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATCACGT AAACTCACGT	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAGGTCAGT AGCAAAAGGC CCGGATACCT GGCTGTTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTCT TGTTTGCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTCT TGTTTTCCAAG TAAGGGATTC	K A I L GCAGATATCC  CCGGTCATCA CAAACCACAA TGATCAGTCG GCCGGAAGCA CGTTCCGCTC CGCTCGGCTT CAGCAAAAGG GGTGCGAACA GTCCGCCTTT CAGCAACCC CTGGTAACAG CTGGCTCTC CAGCAGATTA TGGTCATACAG CAGATTA
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACT 2601 TAAAGTGTA ATCCGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGAC 3001 CCCGCACAGGA CTATAAACA 3101 CTCCCTTCGG GAAGCGTGG 3201 CAGTAACAGA GCACACGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 CGGCAGAAA AAAAGGATC 3601 ATTATCAAAA AGGATCTC 3701 TGCTTTAATCA GTGAGGCAC	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTCAGGG GGCGGAGA CTCAAAGGCG CGCGGGAGA ACTCAAAGGCGT CGCTTTCTCAT CGCTTTTCTCAT CGCTTATCCAG CTTTCTCAT CGCTTATCCG AAAAAGAGTT CAAGAAGATC CAAGAAGATC CCTAAGACCT TCAAGAAGATC TCAAGAAGATC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TCCCCCTGGA AGCTCACTGGA AGCTACCT GTAACTATCG GTAACTATCG GTAGCTTC GTAGCTTC TTCTTTGATCTT TTTAAATTAA ATCTGTCTCT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATTCCTGT CACATTAATT GTATCACAGA TAGGCTCCTC GTAGGTACT TCTTGAGTCC TTGAAGTC TTTAAGGGG AATCAGAGA TTTACGGGAACT TCTTACGGGG AATCAGAGT TTCCTTCATC	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GTGAAATTGT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGGGCTCTCC CAGTTCGGT AACCCGGTAA ACCACACTA ACAAACCAC TCTGACGCT CTGACGCT TTAAATCAAT CATAGTTGCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC ACCGGCACA AGCATCACCA TGTTCCGCTC AACGACGACT CGCTCGACC TGGTCGACC TGGCTCACC GCTGGTACCA GCTGGTACCA AGTAGACGA TAAAGTATA TGACTCCCCC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT TTTCCAGTCG ACTGACTGGA AATTGACAC TTTCCAGTCG ACTGACCGTTA GCTCCAGCG TACCCCACTG GAGACATTTT ATGAGCC TGCGGGTTA GCTCCAGCGTTA TCCCAGTG TATGACACT TATGAGACA TTGGACACT TATGAGACA TTGGACACT TATGAGACA TTGGATAAA TCGTGTAGAT TATGAGTAAA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GCACCAGCCA TATTTGGTAT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTCTGA AACTACGATA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAA GTCCGCCTTT CACGAACCC CTGTAACAG CTGCGCTTAACAG CTGCGCTTC CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGACGGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCAGGTAT 401 CTGAAGCCAG TTACCTTCG 3501 CGCGCAGAAA AAAAGGATCTC 3701 TGCTTAATCA GTGAGGCCC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGGTTATATT GGCTAATGAG GCGGGGGAG ACCAGAGGGGGGGGGG	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTG TGGTCATAGC TGAGCTAACT AGCCGGTTTG GTAAATACGGT TCCCCCTGGA AGCTCACGCT TAACTATCG TAACAGAGTTC GGTAGCTATTC TTAAATTATA ATCTGTTCTT TTAAATTAA ATCTGTCTTT CGCGAGACCC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT TGATTCCTGT TATCACAGA TAGCTCCCC AGCTCCCCG GTAGGTATC TCTTGAGTCC TCTTGAGTCC TCTGAGTAC TCTTGAGTC TCTTGAGTC TCTGAGTC TCTTGAGTC TCTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTGAGTC TCTTCATC ACCCCACCA AATGAAGTT TCCTTCATC ACGCTACCC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTAC ACCAGGTAACACCA CCTAACACAC CTTAAACCAC TTAAATCAAT CATAGTTGCC GTTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGGTTACACACAC CCTCAAGTTGCGTC CATAGTTGCC GCTCCAGATT	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGGTCATCAC GGCTACACT GGCTACACT GGCTACACT GGCTACACT GGCTACACT GGCTACACT GGCTACACT GGCTACACT GGTGAGCG CTAAAGTATA TGACTCCCCG TATCAGCACT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA TTTCCACGTC ACTACTGCACTGC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTG TCAAGTCAGA AGCAAAAGGC CCGGATACCT GGCTCGGT AGCAAAAGGC CCGGATACCT GGCTCTGTG GGACTCGGT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCGAAA GTCGGCTTT CACGAACCCC CTGGTAACAC CTGGTACCAC CAGCAGATTA CAGCAGATTA
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2401 TAGAATGTAA ACCTGCGGG 2701 GCATTAATGA ATCCGCCCA 2801 CGGCGAGCGG TATCAACTC 2901 CCAGGAACCG TAAAAAGGCT 2901 CCCGACCAGGA CTATAAAAA 3101 CTCCCTTCGG GAACCGCTG 3201 CCGTTCAGCC CGACCGCTG 3201 CGTTCAGCC GGACCGCTTG 3401 CTGAAGCAGA AAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3501 TACCATCTGG 3801 TACCATCTGG 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTA	AAAACTTTA  AGTCTAGAGG ATGCTTATTT AGGTTCAGGG GGCGTAATCAG CGCTAATGAG CGCTTACTGG ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT ACCAGGGGT CCTTATCCG ACAGAGAGAG ACAGAGAGAG AAAAAAGAGT ACCAGATCAT ACCAGATCAT CCAGATCATAC CCCCCCCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTGG TGAGCTAAGC TGGGCTAAGC TGAGCTAACT AGGCGGTTTG GTAATACGCT GTATTCGCATCCCCTGGA AGCTCACCCT TCCCCTGGA AGCTCACCCT TACAGAGTTC GTAACTATCG TACAGAGTTC TTTAACTATAC TTTTGATCTT TTTTAACTTAA ATCTGTCTTAT CCCAGTCTATT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTGGT ATCCGCAA TTCTACGGG AAATGAAGTT TCGTTCATC AAATGAAGTT TCGTTCATC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGGTA AACCAGCT CTGACACC TCTGACG TTTCGGTA TACAAACCAC TCTGACGCTC TTAAATCAAT CATAAGTTGCC GCTCCAGATC GCTCCAGATC TCTGACGCTC TTAAATCAAT CATAATTGCC GCTCCAGATT GGGAAGCTAG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAA TGTTCCGACC TAGCTCGTC GACACAGCAC TGGTCGTC GACACGACT GGGTACACT GCTGGTACGC CTGGTACGC TAGATTAT TGACTCCCCC TAGATTAT TGACTCCCCCT TATCAGCATT	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTTT ACTGCCACTG ACTGACTTT ACTGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT ACCCACTG ACACGCCA TCGCCACTTA TCGCCAGTTA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAAG TAAGGGATTC TTTGCAAG ACTACGAAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GTGCGAAAAG GTCCGCTTT CACGAACCAC CTGGTAACAG CTGGTAACAG CTGGGCTCTC CAGCAGATTA TGGTCATGACAG CTGGGAGACCCC CTGGTAACAG CTGGGAGACCCC CCGAGATTA
2001 GGATACTTGA GTTAGTTTA	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGTTCAGGG GGGGTAATCA GGCTAATGAG CGCGGGAG ACTCAAAGGCG CGCGGGAG ACTCAAAGGCGT CGCTTTCTCAT CGCTTATCCAG CAGAGAGATC AAAAAGAGTT CAAGAAGATC CAAGAAGATC TATCTCAGCG CAATAAAGAGT TCAAGAAGATC TATCTCAGCG CCAATAAAAGAGT CCCTTATCAGCG CCAATAAAAGAGT CCCCTCAT	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT TGAGCTAACT TGACCTAACC TGACCTAACT TCCCCCTGGA AGCTCACGCT GTAACACTATCG GTAACTATCG GTAGCTCT TTCAACTATT CGTAGCTCT TTTTAACTT TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGCTCATT TCGTCGTTTT TCGTCGTTTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TATCACAGA AATGAAGTT TTCGTCACC GTAGCTCGCCAC GTAGCTCACCG GAACTCACCG GAACTCACCG GAACTCACCG GAATGTTCACCGGG AATGATTCTCCC GTAGGTACCT	E V G TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTTTATTTG AAACCAAGTA GCGAAATTG GCGTTCCGC ATCAGGGGAT CCCCCTGACG CCCCCTGACG ACCAGCACT ACCACACC CTCTACCACTC CAGTTCGCG TCTACTACTA ACAAACCACC TCTGACGTC TCTAACTAACTAACCACC CCCCCAGAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATCAGCTCC ACTAGTTGCC ACTAGTTGCC ACTAGTTGCC ACTAGTTGCC ACTAGTTGCC ACTAGTTGCC ACTAGTTAAATCAAT ATTCAGCTCC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCATCAT TTCCTCGCTC ACCGCGC AGCAGAA AGCATCACA AGCATCACA AGCATCACA AGCATCACA AGCATCACA AGCATCACA AGCATCACA AGCATCACA CTGTCCGAC TAGGTCGTC GCTGGTACCA AGTAGGAACGA AGTAGAACTA TGACTCCCG TATCAGCAAT AGTAAGTAGT AGTAAGTAGT GGTTCCCAAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACATTGAA AAATGGCAA TTTCCAGTCG ACTGACCTCGAC AGAACATGG AAACATGGCAA ATCGCCGCTTA GCTCCAAGCT ATCGCCACTG GTGGTTTTTT AAACTCAGCG TTGGTGGTTTTT AAACTCAGCG AGAGGACAA TCGCGCAGTA AACCAGCCA TTGCTAGATAA TCGTGTAGAT AAACCAGCCA TCGCCAGCTA AAACCAGCCA TCGCCAGTTA TCAGCAGCTA TCGCCAGTTA TC	E E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA GCAGCAACAT GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGATTGGAACACTGGAAGGA TAACGGAATGGAACTGGAAGGA AACTACGAACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTCGGCTT CAGCAAAAGG GTCCGCCTT CACCAACCC CTGGTAACAG CTGGCTCTG CAGCAGATTA TGGTCATGAG CAGCAGATTA CGGAGGCCAG CAAGGTTGTT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTCAGCA CGACCGCTG 3301 GATTAGCAGA GCAGGTAT 401 CGGAGGAAA AAAAGGATCTTC 3501 CGCGCAGAAA AAAAGGACTTC 3701 TGCTTAATCAA AGGACTTTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA 401 GCCATTGCTA CAGCCATTGG 4101 TGTGCAAAAA AGCGGTTAG	AAAACTTTA  AGTCTAGAGG AAACCGCTG AGGTTAATGT GGGGTAATCA GCCTAATGAG GGGGGGGGAG ACCAGAGGTT CGCTTATCTT CGCTTATCTT CGCTTATCGG TAGGGGGGAG ACAGGCGTT CGCTTATCCG TAGGCGGTGC AAAAAGATT CAAGAAGATT CAAGAAGAT TCAAGAGCT TCATCAGCG GCAATGATAC CGCCTCCAT CGCCTCCAT CGCCTCCAT CGCTTCCGCC	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAAACTACT CCCCCTGGA AGCTCACCGCT TAAACTATCG TAAACTATCG GTAACTATCG TAAACTATCG TAAACTATCG CTTTTGATCTT TTAAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT TGATGCTATT GAGGTTTTT TGTTTCCTGT TATCCACAGA TAGCCTCCAC AGCTCCCTCG GTAGGTACT TCTTGAGTCC TCTGAAGTGGT GATCCGCA AATTGAGTC TTCATCAGGA AAATGAAGTT TTCTTCACGAC AATTGATCC CGTATGGCTTC TGTCAGCAAGT TGTCAGCGTTC TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT GCGCTTACGT AACCAGCTACA GCCTAACT GCGCTAACT TGCAGTACACA GCCTAACT TGCAGTACT ACAACCACC TCTGACGTAA GCCTAACT ACAACCACC TCTAACGTTA GCAAACTAC CTCTGACGTTA GGAAGCTAC TTAAATCAAT CATAGTTGCC CTCCAAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCC AAGTTGGCCC	E S E E TGTGATCTTG ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGGAA GGTTCCGACC TAGCTCACAC GACAGGAA GGTTCCGACC TAGGTCGTTC GGCTACACT GGCTACACT GGTGGTAGCG CTAAAGTATA TGATCCCCGC TATCACGCAAT GGTTCCCAC CTATCACAC CTATCACAC CTATCACCAC CTATCACCAC CTATCACCAC CTATCACCAC CTATCACCAC CTATCACCAC CACTGTTATC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAAGTGGTA AAATTGCACA ATTTCCACA ACTTCCACA ACTGACTGG ACTGACTGG ACTGACTGG ACTGACTGG ACTGACTGG ACTCCAAGCT ATCGCCACTTA AGAAGACAG GTGGTTTTT AGAAGTAAA TCGTGAGTAAA TCGTGAGTAAA TCGTGAGTAAA TCGTGAGTAAA AACCAGCT AGAAGCAGCA AAACCAGCT AGAAGCAGCA AAACTAAGGCA AAACTAAGGCA ACTCATGGTTA	E L GGGCAATTCT TCTAGGCGTA TGAGTTTIGGA AAACAAGTTA TGGCTGATTA CAACATACGA AGGAAAACGC TCAAGTCAGA AGCAATACGA AGCAATACGA AGCAATACGA AGCAATACGT GGGCTCGGT GGGCTCGGT GGGCTCTGTG GGACACCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA ACTACGATA ATGCCAGCAA ATGCCAGCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCCAAAAGG CTGGCAAAAGG CTGGCTAACAG CTGGTAACAG CTGGTAACAG CTGGTAACAG CTGGTACACA CAGCAGATTA TGGTCATGAG CAGGAGGCT CCGAGGCCCAG CAGGTGTT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACAA 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TAAAAAGGCT 2901 CCCGACCAGGA CTATAAAAA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCAGA TAACCTTCG 3501 CGCGCAGAAA AAAAGGATC 3501 TACCATCTGG CCCCAGTGC 3601 ATTATCAAAAA AGGATCTTC 3701 TGCTTAATCA GTGAGCCAC 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGATTCG 4101 TGTGCAAAAA AGCGGTTAG	AAAACTTTA  AGTCTAGAGG ATGCTTATTT AGGTTCAGG GGCGTAATCAG GCCTAATCAG CGCGGGAG ACCAGGCGTT CGGCTTCTCAT CGTTTCTCAT CGCTTACTCAT CGCTTACTCAT CGCTTACTCAT CGCTTACTCAT CAAAAAAAAAA	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTAG GGAGGTGTG TGAGCTAACT AGGCGGTTTG GTAAACTCAC TCCCCCTGGA AGCTCACCCT TCACCCTTGA TACAGAGTTC TACAGAGTTC TACAGAGTTC TTACATATC GTAACTATC GGTAGCTCT TTTTGATCTT TTTTAATTAA ATCTGTCTAT TCGCAGAGCCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TCTCTGATCTT CTTCGGTTGGT TTCTGTGATTT CTTCCGTTTG TCTCCGATCGT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCC TCTAGAGTAGT TCTTGAGTAC ACTCTACGGG AAATGAAGT TTCTTCACC AAATGATTT TCGTTCACC GAATTGTTCAC GTAGGTACT TCGTTCACC GAATTGTTCAC GTAGGGTACT TTCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TCGTAGAGTACT TGTCAGAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCTCTTCCGC ATCAGGGGAT TCCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA CAAAACCAC TCTGACGCT TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGAT GCCCAGAT ACCAGATC ACCAGATC ACCAGATC ACCAGATC ATCAGCTCC AAGTTGCCC CACCAAGTC	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGCTCGCTC AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGACT GGCTACACT GACAGACT GTTCGTACG CTAGATTAT TGACTCCCAC TAGATTAT TGACTCCCAC TAGTAGGACT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTAGT AGTAGTTAGC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTACC AGTGGTTATC AGTTGTTACC AGTTGTTACC ATTCTGAGAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAAT AAATTCCACA TTTCCAGTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTCG ACTGACTTT ACTCCAACCT ACTGACACT ATCGCCACTG AGAAGGACAG GTGGTTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT ACCGCCAGTTA GATCAAGGCG ACTCATGATTATACACTT TAGTGTAATT TAGTGTATATC	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTTGCAGAGAACGCC AATTTGGTAT CTTGGTCTGAACTACGATA ACTTACGATA GCCGGAAGGC AATACTTTGCAAG AACTACGATA ACTTACGATA ACTTACGATA ACTTACGATA AACTACGATA ACTACATACATA AACTACATACATA AACTACATACA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GTGCGAAAAG GTCCGCTTT CACGAACCAC CTGGTAACAG CTGGTAACAG CTGGGTTCT CAGCAGATTA TGGTCATGAC CTGGGAGCT CAGCAGATTA TGGTCATGAG CAGTTACAG CAGTTACAG CAGTTACTAG CAGTATCCAA CGGAAGGCT CCCAACCTT TCCCCCATGT TCCCCCATGT TCCCCCATGT TGCATATATT TCCCTTTGCTCTTTGC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGGAATGCA GTGAARAAA 2401 TTGCATTCAC TTTATGTTT 2501 ACCTGCAGGC ATGCAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAGCGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGGC 3001 CCCGGCACAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CGGTTAGCC CGACCGCTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACCTTCG 3501 ATTATCAAAA AGGATCTC 3601 ATTATCAAAA AGGATCTC 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GCAACTTTA 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCCTCAA TACCGGGATA	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTTATTT AGTTCAGGG GGGGAGAACACACACACACACACACACACACACA	F S D D CATAATTAAA GCCCGCGGTT ATCAGCCTCG GTGAAATTTG GGAGGTCTGG TGACTAACT AGCCGGTTTG GTAAACTACGGT TCATACCG TCACCCCTGGA AGCTCACGG TTAACTATCC TACAGAGTTC CTTTGATCT TTTAAATTAA ATCTGTCTT TTAAATTAA ATCTGTCTT CGCAGGACCC CCAGTCTATT CGCAGTCTTT CTCGTCGTTTG CTCCGATCGTT CTCGTCGTTTG CTCCGATCGTT	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATAATT CGTATTAGGC TATCCACAGA ATTGCCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTGGT TCTTCACGGA AATGAGTT TCTTCACGGA AATGAGTT TCCTTCACACAC AATTGTACC ACGCCAACGC AATTCACACGC AATTGTTCCT GTCAGAAGT TGTCAGAGAGT TGTAGGCTC TGTAGAGTACT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT TGTCAGAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT CTTTAAAAGT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTTACG ACCCGCTAAC ACCAGTCGGTA ACCAGTAC ACCAGTACAT ACAAACCAA TTAAATCAAT TAAATCAAT CATAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTGCC CCTCAGAGTTAC CCTCAGAGTTAC ACTTGACCCT ACTTGACCCT ACTTGACCCT ACTTGACCCC AACTTGCCC AACTTGCCC CAACCAAGTC CAACCAAGTC CCTCATCATT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA AACCTCTAC TATCCGCTCA CACTGCCGC TTCCTCGCTC AACCAGGAA AGCATACACA TGTTCCGACC TGCTTCCGACC GACGCACACC GCTGCTACACC TGTGCTCCGCC TAGGTCGTC CTGGTACCG TGTGGTACCG TGTGGACGA TGATCCCC TAGACCAAT TGACTCCCC TACACCAAT AGTGTTACC ACTGTGTTACC ACTGTGTTACCAAC ACTGTTATC ATTCTGAGAA GGAAACGTT	V T E TTTATCTGAA  CGGTCTCGAT GATACATTGA AAAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGGACGA ATCGCCACTG AGAAGGACAGT ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGCACGCA TCGCCAGTTA AACCACCAC TCGCCAGTTA AAACCAGCA TCGCCAGTTA CACCAGTTA CACCA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGAATACCTA TGTTTGCAA TGTTTTGCAA TATTGGTTGA TATGTTGTAA TAGCGAGAGGA AACTACGATA CTGGTCTGA ACTACGATA GCGGAAGGG ATAGTTTTCGT ATGGTCTGA ACTACGATA CTGGAAGGA ATATTTCGAA ATGCAACAGA AAAACTTCTCA	K A I L GCAGATATCC CCGGTCATCA CAAACCACA ACAACACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCAC CTGGTAACAG CTGGCTTACAA GCGCAGTTACCAA GCGGAGGCT CCGAGCGCATT TGGTCATGAG CAGTTACCAA CAGTTACCAA CAGTTACCAA TTCCCCCCATGT TCCCCCCATGT TCCCCCCATGT TCCCCCCATGT TCCATCATCATC AGGATTATTC TTGCTCTTTCC AGGATTATTC TTGCTCTTTCC AGGATCTTTAC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGTCAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACACTTC 3501 CGCGCAGAAA AAAAGGATCTTC 3701 TGCTTAATCA GTGAGGACTCTG 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GAACTTTAA 401 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCGTCAA TACGGGATA 4401 CGCGCGTCAA TACGGGATCA	AAAACTTTA  AGTCTAGAGG AAACCCGCTG ATGCTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG CGCGGGGGAG ACCAAAGGCG CGCTTCTCTCT CGCTTATCCG AAAAGAGTT CAAGAGCGTT CAAGAGATCCT CAAGAAGATCCT CAAGAAGATCCT CACCCCTCAGT CGCCTCCAT CGCCTCCAT CGCTTCAGCG CCCTCCAT CGCCTCAGATCCT CGCCTCCAG CCCTCCAGATCCT CACCCCCCCCCAGATCCT CACCCCCCCCCAGATCCT CACCCCCCCAGATCCT CACCCCCCCCAGATCCT CACCCCCCCCAGATCCT CACCCCCCCCAGATCCT CTCCGCCTCAGAT AACACCCAAGATCCT ATACCCCACAGATCCT ATACCCCACAGA ATGTAACCCAAGATCCCAAGATCCCAAGATCCTA	F S D D CATAATTAAA GCCCGGGTT ATCAGCCTCG GTGAAATTGGGAGTGTGGTCATAGC TGAGCTAACT AGCCGCTTGTAACT CCCCCTGGA AGCTCACCCT TAAACAGTTTCCA TCCCCTGGA AGCTAACT TAAACAGTT TAAAATATA TCTTTGATATT TTTAAATTAA ATCTGTTATT CGCAGAGCC CCAGTCTATT TCGTCGTTTT CTCGTGTTTT CTCGTGTTTT CTCGTGTTTT CTCGTGTTTT CTCGTGTTTT CTCGTGACC CATCTATT TCGTCGATT CTCCGATCGT TCTTGTGACT CTTTGTGACT CTTTGTGACT CATAGCAGAA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGTTTTT TGTTTCCTGT CACATTAATT CGTAATTAGGC TATCCACAGA AGCTCCCTCG GTAGGTACT TCTTGAGTCC TCTGAGGTACT TCTTGAGTC TTCATCGGGAAAGT ACGCTCACCG AATTGTTCCC GTATGGCTC CGTAGGAAGT CGTAGGAAGT CTTTAAAAGT CAACTGATCT	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GTGAAATTGT GCGTTGGCGT GCTCTTCCGC ATCAGGGGAT GCGCTACCT CCCCTGACG TGCGCTACCT ACCCGGTAA GGCCTAACTA GCCCTAACTA GCCTAACTA GCCTAACTA ACCACGTA TTAAATCAAT CATAGTTGCC CTCCAGAGTT GGGAAGCTAG ATTCAGCTCC AAGTTGGCCG CAACCAAGTC CACCAAGTT TCAGCATCT TTAACATCT CAGCATCT TTAACATCT CAGCATCT CACCAACTT TCAGCATCT TCAGCATCT TTAACCAACTCT TTAACCAACTCT TTAACACTCT TTAACACTCT TTAACCAACTCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT	E S E E TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACGCAGGAA GGTTCCGACC TAGCTACACA GGTTCCGACC TAGGTCATAT GGCTACACT GGCTACACT GGCTACACT GCTGGTACCG CTAGGTACCG CTAGGTACGA CTAAAGTATA TGACTCCCCG TATCCACAC TAGGTCTTCCCAC CAGTGTAACT CACTCTCCCAC CACTGTTATC ACTCTCCAAC CACTGTTATC ATTCTGACAA	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTCCACA ATTTCCACA ATTTCCACA ACTTCGAC ACTACTGC ACTACTGC ACTACTGC ACTACTGC AGAACATGTG CTCCAAGCT ATCCCCACTT AGAAGGACAG GTGGTTTTT AAACTCACGT TATGAGTAAA TCGTGTAAAGCC ACTCCAGCTT AGAAGCAGCA TCGCCAGTTA CGTCAAGCT TCGCCAGTTA TATGAGTAAA TCGTGTAAAGCC ACTCATGGTT TAGGTAAAGCC CACTTACGGTC ACTCACGGC CTCCCCAGTTA CACCAGTTA CACCAGTT CACCAGTT CACCAGTT CACCAGTT CACCAGT CACCA	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CACCATACGA AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTCGGT AGCAAAAGGC GCACAGCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TTGGTCTGA AACTACGATA AACTACGATA AACTACGATA AACTACGATA AACTACGATA ATGCCAGCA ATGCCAGCA ATGTTTCGA AGGCACCAG AAAACTCTCA AAAACTCTCA AAAACTCTCA AAAACTCTCA AAAACTCTCA AAAACTCCAA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT GGTTCGGCTG CAGCAAAAGG GTGGCAAAAGG GTGGCAAAAGG CTGGCAAACG CTGGTAACAG CTGGTAACAG CTGGTAACAG CTGGTACAGA CTGGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CAGGAGGTGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCCTCTTGC AGGATCTTTC CAGGAGATTA TTGCTCTTTGC AGGATCTTTC CAGGAGATTACAA CAGGAACTTTAC
2001 GGATACTTGA GTTAGTTTA  D T  101 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAAATGCA GTGAAAAAA 401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAGACA 2601 TAAAGTGTAA ACCTGGGGG 2701 GCATTAATGA ATCAGCCCA 2801 CCGGCAGCAGG TATCAACTC 2901 CCCGACCAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTGG 3201 CCGTTCAGCC CGACCGCTGG 3201 CGGTTCAGCC CGACCGCTGG 3401 CTGAAAAGGATC 3401 CTGAAAAAGGATC 3501 CGCGCAGAAA AAAAGGATC 3701 TGCTTAATCA GTGAGCCAC 3801 TACCATCTGG 3801 TACCATCTGG 3801 TACCATCTGG 4001 GCCATTGCTA CAGGCATCG 4101 TGTGAAAAA AGCGATCTG 4101 TGTGAAAAA AGCGATTCG 4101 TGTGCAAAAA AGCGGTTAG 4101 CGCGCGTCAA TACCGGATCA 4401 CGCGCGTCAA TACCGGATCA 4401 CGCTGTTGAG ATCCAGTTC 4501 GCAAAATGCC GCAAAAAAAG	AAAACTTTA  AGTCTAGAGG ATGCTTATTT AGTTTAGG GGCGTAATCAG GGCGGAGA CTCAAAGGCG ACCTTACTG ACCTTACTG ACCTTACTG ACCTTACTG ACCTTACTG ACCAGAGGT CCTTACTG ACAGAGGT ACAGAGAGT CAGAGAGAT ACAGAGAT CAGAGAGAT CTCAAAGAGAG GGCATCAGAGAGAGAGT CTCAAGAGAGAGT CTCAGAGAGAGAT CTCAGAGAGAGAT CTCAGATCCT ATACTCAGCG GGCATCAGT GGGGTCAGT GGGGTCAGT ACCTCAGGT TAAGAGAGT TAAGAGAGAT ACCGCCCAG ATAAAGAGGT ATAGAGAGCCA ATACCCAG GAATAAAGGGC	F S D D CATAATTAAA  GCCCGCGGTT ATCAGCCTCG GTGAAACTTAGG TGGTCATAGC TGGGCTAACT AGGCGGTTTG GTAAATACGGT GTATACGATCG TCCCCTGGA AGCTCACCT TACAGAGTTC TACAGAGTTC TACAGAGTTC TTTAAATTAA ATCTGTCTAT TCTTCATCT TCTTGATCTTTT TCTTCGTTTG TCGCATCGT TCCGATCGT TCCGATCGT CTCCGATCGT CTCCGATCGATCGT CATAGCAGAA CTCCGGAACA CACAGGAAA CTCGTGCACC CACAGGAAA CTCGTGCACC CACAGGAAA CTCGTGCACC CACAGGAAA CTCGGACAGAA CTCGGACAC	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GATGCTATT GAGGTTTTT TGTTTCCTGT TAGCACAGA TAGGGTCAGC AGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTTGAGTCC TCTTGAGTCG AAATGAGTATT TCCTTCAGC AAATGAGTATT TCCTTCACC AAATGAGTACT TTCGTTCATC GTATGGCTTC TTTGAGTACC TTTCAGAGTACT TTCTTCAGT TCTTCAGC TCTTCAGC TCTTCAGC TCTTTAGAGTTC TTTTAGAGTTC TTTTAAAAGT CACTGAGTACT CTTTAAAAGT CAACTGATCT TGTTGAAATAC	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAGTA GCGATGCGCT GCCTCTACCG ATCAGGGGAT CCCCCTGACG TGCGCTTCCC CAGTTCGGTG AACCAGCT TCTGACGCT TTAAATCAAT GGCAACCAC GCTCAGCG TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGAT GGCAAGTC CACCAGATT CGCAACTAC TCCACACT CCCAACTC CACCAACT CCCAACTT CCCAACTT TCACACTCT TCAACACT TCACACTT TCACACTCT TCACACTCT TCACACTCT TCACACTCT TCACATCTT TCACACTCTT	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTTA GCTCCAAGCT ACTGACTCA ACTGACTTA ACCACACT AAACCACACT TATGAGTAAA TCGTGTAGAT TATGACTAA ACCACACT TATGACTAA ACCACACT TATGACTAA CCTTCGCCAGTTA GATCAAGCCC TCGCCAGTTA CATCAAGCCC ACTCATGTT TATGATTATGC CTTCGGGGGCC CACGTTTT TATGATTATGC CACGTTTTCT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTTGC GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA GCATTTATCA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAAACCA CTGGTAACAC CTGCTATCA CAGAACCAC AGGTAACAC CTGCGAACCC CTGGTAACAC CTGCGATTACAC CAGGTACCA CAGTTACCA CAGGTACCA CAGTTACCA CGGAGGGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGAGTCTTAC AGAACAGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC
2001 GGATACTTGA GTTAGTTTA  D T  1011 AGCACAGTGG CGGCCGCTC 2201 TCACCATCAC CATTGAGTT 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTT 2501 ACCTGCAGGC ATGCAAGCT 2601 TAAAGTGTAA AGCCTGGGG 2701 GCATTAATGA ATCGGCCAA 2801 CGGCGAACGG TATCAGCTC 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGTCAGGA CTATAAAGA 3101 CTCCCTTCGG GAAGCGTG 3301 GATTAGCAGA GCGAGGTAT 3401 CTGAAGCCAG TTACACTTC 3501 CGCGCAGAAA AAAAGGATCTTC 3701 TGCTTAATCA GTGAGGACTCTG 3701 TGCTTAATCA GTGAGGCAC 3801 TACCATCTGG CCCCAGTGC 3901 AAGTGGTCCT GAACTTTAA 401 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAG 4201 TCTTACTGTC ATGCCATCC 4301 CCGGCGTCAA TACGGGATA 4401 CGCGCGTCAA TACGGGATCA	AAAACTTTA AGTCTAGAGG AAACCCGCTG ATGCTTAATTA GAGTTCAGGG GGCGTAATCA GGCTTACTG GCCTTAATGAG ACCCGGGAG ACCCGGGAG ACCCGGGAG ACCCTAATCAG GCTTTCTCAT GCCTTATCCG AAAAAGAGTT AAAGAGATC AAGAAGATC CCCTCAT CCCTCCAT CGCCTCCAT CGCTCCAT CGTGTCAGGC TCCTTCGGTC TAACAGATCT TACCTGGTC TAACAGATCT TACCTCGGTC TAACAGATCT TACCTCGGTC TAACAGATCT TACCGCCCAAT TGGTGTCACGC TCCTTCGGTC TAACAGATCTT TACCCCAAT ATGTAACCCAA ATGTAACCCAA ATGTAACCCA AGAATAAAGGGC TGAATAATGT	F S D D CATAATTAAA GCCGCGGGTT ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC GTAAACTTCG GTAAACTACGGT TCCCCCTGGA AGCTCAGGA TCCCCTTGATCTT TTAAATTAA ATCAGAGTTC CTTTTGATCTT TTAAATTAA ATCAGAGTTC CCAGTCTAT CGCGAGACCC CCAGTCTATT CGTCGTTTTG CTCCGATCGT TCGTGGTTTG CTCCGATCGT TCTTGTGACC CATAGCAGA CTCCGTGCAC GACACGGAAA TCCGTGCACC GACACGGAAAATTA	E E M TAACTAGCAT  CGAAGGTAAG ACTGTGCCTT TGATGCTATT GAGGCTATT GAGGTTTTT TGTTTCCTGT TATCCAGAG TATCCACAGA TAGGCTCCTCG GTAGGTATCT TCTTGAGTCC TCTAAGTTCCTTT TCTTGAGTCC TCAAGTAGT TCTTCACGGG AAATGAGTT ACGCTCACCG AATTGTTGCC GTAGGGTTACT TCTCAGAAGT CTCTCAGAAGT TCTCAGAAGT TCTCAGAAGT TCTTCAGATACT TCTTCAGATACT TGTCAGAAGT TGTGAGATACT TGTTGAGATAC AACAAATAGG	E V G TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCATTATTTG AAAGCAAGTA GCGTTGACTTTCGG ATCAGGGAT CCCCCTGACG GCTCTTCCGC ATCAGGGGAT ACAGCACTA ACAGCACTA ACAGCACTA GGCTAACT ACAAGCACT TTAAATCAAT ACAAGCACC CTCGAGGTTGC GCTCAGGT ACAGCACT ACAGCACT ATTAGCCC CTCAGAGTT CGGAAGCTC CACCAGAGT GGCAACT ATTCAGCTC AAGTTGCCC CACCAAGTC CGCCACAAGTC CGCTCACATT CAGCATCTT TCAGCATCTT TCATACTCTT TCATACTCTT GGTTCCCGCG	E S E E TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA ACCCAGCAC TGTCCTCGCT AACCCAGCAC TGTTCCGACC TAGGTCGTC GACAGGACT TGGTCGTACC GCTGGTACCG CTAAAGTATA TGACTCCCAC TAGGTACAC AGTGAACGA AGTAAAGTATA TGACTCCCAC TATCAGCAC TATCAGCAC TATCAGCAT TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCAGCAC TATCTTCCAC CCTTTTTCAC CCTTTTTCAC	V T E TTTATCTGAA CGGTCTCGAT GATACATTGA AAACTGCAA AAATGTGGTA CAATTCCACA ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTGC ACTGACTTA GCTCCAAGCT ACTGACTCA ACTGACTTA ACCACACT AAACCACACT TATGAGTAAA TCGTGTAGAT TATGACTAA ACCACACT TATGACTAA ACCACACT TATGACTAA CCTTCGCCAGTTA GATCAAGCCC TCGCCAGTTA CATCAAGCCC ACTCATGTT TATGATTATGC CTTCGGGGGCC CACGTTTT TATGATTATGC CACGTTTTCT	E L GGGCAATTCT TCTACGCGTA TGAGTTTGGA AACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTTGC GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG ACTACGATA ACGCGAACGC AGTTACATGA ACTACGATA ACGCGAACGA AAAACTCTCA GGCGACCGAG AAAACTCTCA GGCTACTCAA GCATTTATCA	K A I L GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAAA TGATCAGTCG GCCGGAAGCA GGTGCCAAAAGG GGTGGCGAAA GTCCGCTTT CACGAAACCA CTGGTAACAC CTGCTATCA CAGAACCAC AGGTAACAC CTGCGAACCC CTGGTAACAC CTGCGATTACAC CAGGTACCA CAGTTACCA CAGGTACCA CAGTTACCA CGGAGGGCT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTCCTCTTGC AGGATCTTAC AGGATCTTAC AGGAGTCTTAC AGAACAGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC AAACAGGAAG AGGATTTTAC

# pMT-HisFlag-DmSNAP43 mut#12 STOP

Г									
1 TCGCGCGTTT	CCCTCATCAC	CCTCAAAACC	тстсасасат	GCAGCTCCCG	CACACCCTCA	СУССТАСТСТ	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG									
201 CCGCACAGAT									
301 TACGCCAGCT									
401 TGAATTAATT									
501 GGCCCCCCAC									
601 CAAGTCCCCA									
701 AGAGGTGAAT									
801 AAATCAAGTG									
OUI AMMICAMGIG	AAICAICICA	GIGCAACIAA	AGGGGGGAIC	IAGAICGGGG	IACIAAGAII	> M H	H H H	H H T D	Y K D
901 GACGATGACA	AGGGCACTAG	TGAGCTGAAT	ATCTTTGACG	ACTGCTGGGA	GCTGGTGCAA				
	K G T S		I F D			R F Q	${\tt R}  {\tt L}  {\tt V}  {\tt N}$	D G E	N C E
1001 TCGAGGTGTT									
>F E V F			L Q L Q		T A Q	T N H T		ATT	LAAL
1101 GCATGTGGCC									
> H V A	KRL	S C S R		G D V		~		G F F L	L Y V
1201 ATCTACTACA									
> I Y Y	K Q P T		I K I			QEL			R K D
1301 GTCCGGAGCG									
>S P E R 1401 GGACAATCTG			A Y M L		T Q E	Q A F R		L D Y	C Q G L
> D N L	V D Y	D R V E							
				G A K	E Q R Q		M Q K	Q Q R A	N G V
1501 AGTCTCACAT	ACGAACTGGA Y E I E			O A S O		E L E			
> S L T 1601 CGGCTGGTCA									K Q L
>A A G H		L P P	S O I F			F A D I		L G A	A A A A
1701 TGCAGATGAG									
> A D E		T T S T				R N K		G V E E	R E P
1801 CAACACCAGA									
> 0 H 0		E V O		N E T Y		M S S	A T V F	O R E	L P E
1901 ACGTGCAGCA									
>D V O O			F S D D		E V G	E S E E		E E L	K A I L
2001 GGATACTTGA									
> D T	011110111111			1111011100111	11110000011	10101110110	111111010111	00001111101	0011011111100
2101 AGCACAGTGG	CGGCCGCTCG	AGTCTAGAGG	GCCCGCGGTT	CGAAGGTAAG	CCTATCCCTA	ACCCTCTCCT	CGGTCTCGAT	TCTACGCGTA	CCGGTCATCA
2201 TCACCATCAC									
2301 CTAGAATGCA									
2401 TTGCATTCAT									
2501 ACCTGCAGGC									
2601 TAAAGTGTAA									
2701 GCATTAATGA	ATCGGCCAAC	GCGCGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCTCTTCCGC	TTCCTCGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG
2801 CGGCGAGCGG	TATCAGCTCA	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAAGGC	CAGCAAAAGG
2901 CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	CCCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA
3001 CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT
3101 CTCCCTTCGG	GAAGCGTGGC	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTC	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC
3201 CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACTATCG	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG
3301 GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA	CGGCTACACT	AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG
3401 CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	$\tt GTGGTTTTTT$	TGTTTGCAAG	CAGCAGATTA
3501 CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	TCTGACGCTC	AGTGGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
3601 ATTATCAAAA									
3701 TGCTTAATCA									
3801 TACCATCTGG									
3901 AAGTGGTCCT									
4001 GCCATTGCTA									
4101 TGTGCAAAAA									
4201 TCTTACTGTC									
4301 CCGGCGTCAA									
4401 CGCTGTTGAG									
4501 GCAAAATGCC									
4601 CTCATGAGCG						ACATTTCCCC	GAAAAGTGCC	ACCTGACGTC	TAAGAAACCA
4701 TTATTATCAT	GACATTAACC	TATAAAAATA	GGCGTATCAC	GAGGCCCTTT	CGT				

# pMT-HisFlag-DmSNAP43 mut#13 STOP

	GGEG33333GG	mamaa aa aa m	aas aamaaaa	a. a. aaama.	as a ammamam	am., aaaaa, m	~~~~~~	a. a aaaaa
1 TCGCGCGTTT CGGTGATGAC								
101 TCAGGGCGCG TCAGCGGGTG								
201 CCGCACAGAT GCGTAAGGAG								
301 TACGCCAGCT GGCGAAAGGG								
401 TGAATTAATT CGTTGCAGGA								
501 GGCCCCCCAC CGCCCACCGC								
601 CAAGTCCCCA AAGTGGAGAA								
701 AGAGGTGAAT CGAACGAAAG								
801 AAATCAAGTG AATCATCTCA	GTGCAACTAA	AGGGGGGATC	TAGATCGGGG	TACTAAGATT				
					> M H	н н н	H H T D	Y K D
901 GACGATGACA AGGGCACTAG								
> D D D K G T S			D C W E	. ~	~	R L V N	D G E	N C E
1001 TCGAGGTGTT CTGCCGGTGC								
>F E V F C R C			H L F			E V I	ATT	L A A L
1101 GCATGTGGCC AAGCGACTGT								
> H V A K R L	S C S R		G D V				G F F L	T A A
1201 ATCTACTACA AGCAGCCCAC								
> I Y Y K Q P T			E V S P			T D Y A		R K D
1301 GTCCGGAGCG GAAGGACACT								
>S P E R K D T	~	AYML		T Q E	Q A F R		L D Y	C Q G L
1401 GGACAATCTG GTGGACTACG								
> D N L V D Y	D R V E			E Q R Q		M Q K	Q Q R A	N G V
1501 AGTCTCACAT ACGAACTGGA								
> S L T Y E L E		ALD				A A Y N		K Q L
1601 CGGCTGGTCA TGAGCACGCT								
>A A G H E H A		S Q I F		R E V		Q S V		RKST
1701 TCCAGCTGCG GCAGCCGCCA								
> P A A A A A	TTST	~	L E V	~		A M Y	G V E E	REP
1801 CAACACCAGA CGGATGAACT								
> Q H Q T D E L			NETY			ATVF	QRE	L P E
1901 ACGTGCAGCA AGAGTATGAG								
>D V Q Q E Y E		F S D D		E V G	E S E E	V T E	E E L	K A I L
2001 GGATACTTGA GTTAGTTTAT  > D T	AAAACTTTTA	CATAATTAAA	TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
> 1) 'T'								
0101 2002020000		aaaaaaaamm	aaaama	aama maaama	3 aaamamaam	aaamamaa.m	mama aaaaama	aaaama, ma,
2101 AGCACAGTGG CGGCCGCTCG								
2201 TCACCATCAC CATTGAGTTT	AAACCCGCTG	ATCAGCCTCG	ACTGTGCCTT	CTAAGATCCA	GACATGATAA	GATACATTGA	TGAGTTTGGA	CAAACCACAA
2201 TCACCATCAC CATTGAGTTT 2301 CTAGAATGCA GTGAAAAAAA	AAACCCGCTG TGCTTTATTT	ATCAGCCTCG GTGAAATTTG	ACTGTGCCTT TGATGCTATT	CTAAGATCCA GCTTTATTTG	GACATGATAA TAACCATTAT	GATACATTGA AAGCTGCAAT	TGAGTTTGGA AAACAAGTTA	CAAACCACAA ACAACAACAA
2201 TCACCATCAC CATTGAGTTT 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC	AAACCCGCTG TGCTTTATTT AGGTTCAGGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA	GACATGATAA TAACCATTAT AAACCTCTAC	GATACATTGA AAGCTGCAAT AAATGTGGTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA	CAAACCACAA ACAACAACAA TGATCAGTCG
2201 TCACCATCAC CATTGAGTTT 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAC TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCACCTCA	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAACTCA 2901 CCAGGAACGG TAAAAAAGGCC	AAACCCGCTG TGCTTTATTT AGGTTCAGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAA 2401 TIGCATTCAT TITATGTTTC 2501 ACCTGCAGGC ATGCAGGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCAGCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGAT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACC AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG	CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGTCA ATTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCCT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCAGGAACCG TAAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC	AAACCCGCTG TGCTTATTT AGGTTCAGG GGCGTAATGA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTTG GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGTC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGC 3001 CCCGACAGGA CTATAAAGGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CGGTTCAGCC GGACCGCTGC	AAACCCGCTG TGCTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTTGGA AGCTCACGCT GTAACTATCC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GCTAGGTATCT TCTTGAGTCC TCTTGAGTCCC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ACTGACTCGC ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCCA	CAAACCACAA ACAACAACAA TGATCAGTGG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAGGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCAGTTAATGA ATCAGCCCAA 2801 CGGCGAGCG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGCT 3001 CCCGACAGGA CTATAAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CCGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATC	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGTGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TAACAACTATCG TACAGAGTTC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCCGCT AACCCGGTAA GGCCTAACTA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC CGACACGACTT CGGCTACACT	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGGACAG	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTAGAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCAGCAGCA TATTTGGTAT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCACTG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCCT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCAGGAACCG TAAAAAAGGCC 3001 CCCGACAGGA CTATAAAAGAI 3101 CTCCCTTCGG GAAGCGTGGC 3201 CGGTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATCA 3401 CTGAAGCCAG TTACCTTCCGG 3401 CTGAAGCCAG TTACCTTCCGG	AAACCCGCTG TGCTTTATTT TAGTTCAGGG GGCGTAATCA GCCTAATGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGTGC AAAAAGAGTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGAGCTAACT AGGCGGTTTG GTAATACGGT TCCCCCTGGA AGCTCACGCT GTAACTATCG TACAGAGTTC GGTAGTTCCC TCCCTGGA AGCTCACGCT GGAACTATCG GTAACTATCG GGTAGCTCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTCACAGA TAGCCTCCGC AGCTCCTCG GTAGGTATCT TCTGAGTCC TTGAGTGG GATCCGCA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGAACG TGCGCTCCC CAGTTCGGTG AACCCCGTAA ACCACACAACAA	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTT GGCTACACA GCTTGGTACC GCTGGTACCC	GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGAGACAGC GTGGTTTTTT	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC CCGGATACCT GGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG	CAAACCACAA ACAACAACAA TGATCAGCTTG GCCGGAAGCA CGTGGCAGCT CGTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CCGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATG 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT	AAACCCGCTG TGCTTTATTT AGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGATT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTAACT AGGCGGTTTG GTAATACGGT CCCCCTGGA AGCTCACCCTGGA AGCTCACGCT TACAGAGTTC TACAGAGTTC CTTTGATCTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGGCAA TTCTACGGGA TTCTACGGGA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGGGCTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACCACGCTAACTA ACCACGCTAACTA ACCACGCTACTTCCCC TCTGACGCCT	GACATGATAA TAACCATTAT AAACCTTCTAC TATTCGGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACACGACTTCCGTCT GACACGACTTCCGCTCC GCTGCTACACG AGTGGAACGA	GATACATTGA AAGGTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTGG AGAACATGTG AAATGACTGC CTGCCGCTTA ACTCACCTG AGAAGACATG ACTCACCTG AGAAGACATTG AGAAGACATG ATCGCCACTG AGAAGACATG AGAAGACAAGT ATCACCACTG AGAAGACAAG AGAAGACAAG AGAAGACAAG AGAACATCACACTT AAACTCACATT AAACTCACATT	TGAGTTTGA AAACAGTTA TGGCTGATTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTT	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTCCGGCT CACCAAAAAGG GCTGGCAAA GTCCCGCCTTT CACGAACCCC CTGGTAACAG CTCCGCTCTC CAGCAACATTA TGGTCATGAG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCAGCCCAAC 2801 CGGCGAGCGG TATCAGCTCA 2901 CCCGACAGGA CTATAAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CCGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GTACCTTCGG 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCA	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCAG GCCTAATGAG GCCTGAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCTTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGATC CCTAGATCCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTACT GGACTACT GTAATACCGT CGTTTTTCCA TCCCCCTGGA AGCTCACGCT GTAACTATC GTAACTATC GTAACTATC GTAACTATC GGTAGCTCT TCTTTGACTT TTTTAAATTAA	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACAGGA AATGAAGTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGGTAA ACAAACCAC TCTGACGCT TCTGACGCT TTAAATCAAT	GACATGATAA TAACCATTAT TAACCATTAC TATCCGCTCA CACTGCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCCAGC TAGGTCGTTC GACACGACTT GACACGACTT GGCTACACT GCTGGTACGC AGTGGAACGA CTAAAGTATA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AAATCGACGC TTGCCGGCTTA GCTCCAAGGT ATCGCCACTG AGAAGACACG GTGGTTTTTT AAACTCACGT ATCTCACGT	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACTGT TGCCCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGCCTGTTG GGCCTGTTG GCACGACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTTCGGCTC CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCAGGATTA TGGTCATGACAC CAGTTACAG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGC ATGCAAGCTI 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGAC 3001 CCCGCACAGGA CTATAAAGAI 3101 CTCCCTTCGG GAAGCGGTGGC 3201 CGGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCAG 3701 TGCTTAATCA GTGAGGCACC	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTG AAAAGAGTT CAAGAAGATC CCTAGATCCT TATCTCAGCG	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTACGT GTAATACGGT CGTTTTTCCA AGCCGGTTTG GTAATACGGT CGTTTTTCCA AGCTCACGCT GTAACTATCG GTACTCTCTTTGATCTT TTTTAAATTAA ATCTGTCTAT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTACCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGG AATCACAGAATT TTCTTCATC	CTAAGATCCA GCTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT ATCAGGGGAT CCCCTGAGG TGCGGTCTCCC CAGTTCGGT AACCCGGTAA GGCCTAACTA ACAAACCAC TCTGACGCT TTAAATCAAT CATAGTTGTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGC TACCGGCGC AACGAGAA AGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA GGCATCACAA TGTTCCGAC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACCA AGTGGAACGA CTAAAGTATA TGACTCCCCG	GATACATTGA AAGCTGCAAT AAAGCTGCACA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC TTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACATGT ACACCCACTG AGAAGACACTG ATCACCACTG AGAAGACACT TACACCACTG AGAAGACACT TACACCACTT TACACTT TATACAGTT TATGAGTAAA	TGAGTTTGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT TGTTGCAAG AACTACGATA AACTACGATA AACTACGATA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCTC CAGCAAAAGG GGTGGGCAAA GCTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCCTCTC CAGCACATTA TGGTCATGAC CAGTTACCAA CAGTTACCAA CAGGTACCAC
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGAA GEGAAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAACCTCA 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGACAGGA CTATAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CGGTTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATG 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCA 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCT 3801 TACCATCTGG CCCCAGTGCT	AAACCCGCTG TGCTTTATTT TAGTTCAGGG GGCGTAATCA GCCTAATGAG GCGTGGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGATT CCTAGATCCT TATCTCAGCG GCAATGATAC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTAACT AGGCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCTGGA AGCTCACGCT GTAACTATCG TAACTATCG TAACTATCG TTCACAGCTTC CTTTGATCTT TTTAAATTAA ATCAGACTCC CGCGAGACCC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGA AATGAAGTT TCTTACGGGA TTCTACGGGA AATGAAGTT TCGTTCATC ACGCTAACCA	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCAGCAC ACCAGTAAC GCCTAACTA ACCACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC GCTCCAGATT	GACATGATAA TAACCATTAT AAACCTTATC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC GACAGCATCACTAC GGCTACACT GCTGCTACACT GCTGGTACACG AGTGGACGA TGACTCCCCC TATCACCACT TGACTCCCCC TATCACCAT	GATACATTGA AAGCTGCAAT AAGTGTGATA CAATTCCACA TTTCCAGTCG AGAACATGTG AAATCGACGC CTGCCGCTTA AGAAGACATGTG AGAAGACTGTG AGAAGACATGTG AGAAGATTGAAGTCACCTG AGAAGACATGTG AGAAGACATGTG AGAAGACAGT ATCGCCACTG AGAAGACAGT ATCACTAGTAAAT TCGTGTAGATAAA TCGTGTAGATAAA AACCAGCCA	TGAGTTTGA AAACAGTTA TGACTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGTTGTGG GGACAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA ACTACGATA ACCACAGAAGGC	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CACCAAAAGG GGTGGGCAAA GTCCGCCTT CAGGAACCC CTGGTAACAG CTGGTAACAG CTGCGCTCT CAGCAACATT TGGTCATGAG CAGTATACCAA CGGGAGGGCT CCGGAGCGCCT CCGAGCGCACC CCGAGCGCCCC
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCG TATCAGCTCA 2901 CCCGACAGGA CTATAAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CCGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC GGACCGCTGC 3201 CGTTCAGCA AAAAGGATCT 3401 CTGAAAGCAA TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAAA AGGATCTTCA 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGGGGAG CTCAAAGGCG ACCAGGCGTT GCTTTCTCAT GCTTTATCCG TAGGCGGTGC AAAAAGAGTT CAAGAAGACT CCTAGATCCT TATCTCAGCG GCAATGATAC CCGCCTCCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTACT GTAATACCGT CGTTTTTCCA TCCCCCTGGA AGCTCACCT GTAACTATC TACAGAGTTC GTAACTATC GTAACTATC GTAACTATC TTTTAACTT TTTTAACTT TTTTAAATTAA ATCTGTCTAT GCGAGACCC CCAGTCTATT	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGG GTAGGTACT TCTTGAGTCC TCTTGAGTCC TTGAGTGG AATCACAGG AAATGAAGT TTCGTCACC AAATGAAGTT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCCCCTGACG TGCGCCTCTCC CAGTTCGGCG AACCAGGTA AACCAACC TCTGACGT TCTGACGCT TCTGACGCTC TTAAATCAAT CATAGTTCCG GCTCAGAGTT GCGCAGCTTC GCTCAGAGTT GGGAAGCTAG	GACATGATAA TAACCATTAAT AAACCATCAT TATCCGCTCA CACTGCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCCACC TAGGTCGTC GACACGACTAC GCTGGTACGC CTGGTACGC CTAAAGTATT TGACTCCCCG TATACACAC TATCAGCATT TGACTCCCCG TATACACAC TATCAGCAT	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGACATGACGC ATCGCCGCTTA GCTCCAAGGT ATCGCCACTG AGAAGGACAG AGAAGGACAG TGGTTTTTT AAACTCACGT ATTGAGTAAA TCGTGTAGAT AAACCAGCCA TCGGCAGTTA	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGG GCAGCAACAT TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTAGA ACTACGATA ACTACGATA GCCGGAAGGG ATAGTTTGCA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG TGCGCAGATTA TGGTCAGCATTA TGGTCATGACAG CAGATTACAG CAGATTACAG CAGATTACAA CGGGAGGGCC CCGAGCCCAC CAACCTTGTT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGC ATGCAAGCAT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGATC 3001 CCCGCACAGGA CTATAAAGAT 3101 CTCCCTTCGG GAAGCGGTGGC 3201 CGGTTCAGCC CGACCGCTGC 3201 CGGTCAGCA GCGAGCATTCAGC 3101 GATTAGCAGA AAAAGGATCT 3401 CTGAAGCCAG TTACCTTCGG 3501 GATTAGCAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCA 3701 TGCTTTAATCA GTGAGGCACC 3801 TACCATCTGC CCCCAGTGCT 3901 AAGTGGTCCT GCAACTTTAT 4001 GCCATTGCTA CAGGCATCCT	AAACCCGCTG TGCTTTATTT TAGTTCAGGG GGCGTAATCA GCCTAATGAG GCGCGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTCTCAT GCCTTATCCA TAGGCGGTGC AAAAGAGTT CAAGAAGATC CCTAGATCCT TATCTCAGCG GCAATGATCCT GCAATGATCCAGCGGCGCCTCAT GCCTCCAT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACCTACGT CGTTTTTCCA AGCCGGTTTG GTAATACGGT CCCCCTGGA AGCTCACGCT GTAACTATCG GTAGCTCTT CTTTGATCTT TTTTAAATTAA ATCTGTCTAT CGCGAGACCC CCAGTCTTTT TCGTCGTTTT TCGTCGTTTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGGT GATCCGCAA TTCTACGGGG AATTCATCACGGA TTCTTCATCACG AAATGAAGT TTCGTTCATC ACGCTCACCG GATTGTGCC GTATGGCTT	CTAAGATCCA GCTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGGT ATCAGGGGAT CCCCCTGAGG ATCAGGGGAT CCCCCTGAGG AACCAGTAACTA ACACACCAC TCTGACGTC TCTGACGTC CTTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAC GCTCCAGATT GGGAAGCTAC ATTCAGCTCC ATTCAGCTCC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGGAA AGCATCACAA GGATCACAA GGATCACAA GGATCACAA GGATCACAA TGTTCCGAC TAGGTCGTTC GACACGACTT GGCTACACT GCTGGTACCA AGTGGAACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTATA GGATTCCCAAC	GATACATTGA AAGCTGCAAT AAAGCTGCACA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGC AGCCCATA GCTCCAAGCT ATCGCCACTG AGAAGACAGCT ATCGCCACTG AGAAGACAGC ATCACCACTG AGAAGCACCA TAGCACTA AAACCAGCCA TACCACCACTA AAACCAGCCA TTCGCCAGTTA CAGCAGCTA CAGCAGCAGTTA GATCAAGGCG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT TGTTGCAAG AACTACGATA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ACTACGATA ACTACGATA AGTTTACATG	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCTC CAGCAAAAGG GGTGGGCAAA GCTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CACCAGATTA TGGTCATGAC CGGAGGGCAA CGGGAGGGCT CCGGGCTTT CCCAGCACTTT TCCCCCATGT TCCCCCATGT TCCCCCATGT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTAGAAGAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGC ATGCAAGCAT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAACCTCA 2901 CCAGGAACCG TAAAAAGGCC 3001 CCCGTCAGGA CTATAAAGAT 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTCAGCA CGACGGTGC 3301 GATTAGCAGA GGAGGTATC 3601 ATTAGCAGA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCAG 3501 TACCATCTGG CCCCAGTGCT 3901 AAGTGGTCCT GCAACCTTAT 4001 GCCATTGCTA CAGGCACCC 4101 TGTCACAAAA AGGACTCTT	AAACCCGCTG TGCTTTATTT TAGTTCAGGG GGCGTAATCA GCCTAATGAG GCGTGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCCTTATCCG TAGGCGTTG CAAAAAGAGTT CAAGAAGATC CCTAGATCCT TATCTCAGCG GCAATGATAC GCGCTCCAT GCCTCCAT GCCTCCAT GCTTCAGCG TAGTCCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTAACT AGCCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCT TACAGAGTTC TACAGAGTTC CTTTGATCTT TTTAAATTAA ATCAGTCTAT CGCGAGACCC CCAGTCTATT TCGTCGTTTT CTCGTCGTTTT CTCGTCGTTTT CTCGTCGTTTT CTCCGATCGTTT CTCCGATCGTTT CTCCGATCGTTT	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TAGTACGGCAA TTCTACGGGA AATGAAGTT TTCGTTCATC ACGCTCACCG AATTGTTGCC GTATGGCTTC TGTCAGAAGT	CTAAGATCCA GCTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCTGACG TGCGCTTCTCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA ACAAACCACC TCTGACGCC TCTGACGCCC GCTCAGGATTGCCC GCTCCAGATTGCC AACTAGCTACTA AACAAATCAAT AATAATCAAT AATAATCAAT AATAATCACAC AATAGCACC AAGTTGCCC AAGTTGCCC AAGTTGCCC AAGTTGCCCAAACTCAC AAGTTGCCCAAACTCAC AAGTTGCCCAAACTCAC AAGTTGCCCAAACTCAC AAGTTGCCCAAACTCAC AAGTTGCCCAAACTCACCAAACTCACCACAACTCACCACAACTCACCAC	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTACCAT GCTGGTAACCA TGTGACCGC TATCAGCAAT AGTAAGTATA AGTAAAGTATA AGTAAAGTATA GGTTCCCAC TATCCACC CAGTGTTATC	GATACATTGA AAGCTGCAAT AAAGCTGCAAT CAATTCCAGCA TTTCCAGCA GAACATGTG AAATCGACGC CTGCCGCTTA ACTGCCACTG AGAAGATTTT AAACTCACCT ACTGCACTG AGAAGACATTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA CAACCAGCAGTA AAACCAGCCA TCGCCAGTTA CATCAAGCGC ACTCAAGGCG ACTCAATGGTA	TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTTCGA ATTACATGA ATGACAGCA ATGACAGCA ATGACAGCA ATGACAGCA	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GGTCGGCATA GTCCGCCTT CACGAACCC CTGGTAACAG CTGCGCTCT CAGCAGAATT TGGTCATGAG CAGCAGGTT CCGCAGGGCC CAGGGGGCT CCGAGCGCAT TCCCCCATGT TCCCCCATGT TCCCCATATT CCCCCATGT TCCCCATATT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATGCA GTGAAAAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCTI 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCG TATCAGCTCA 2901 CCCGACAGGA CTATAAAAGAT 3101 CTCCCTTCGG GAAGCGTGGC 3201 CCGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC CGACCGCTGC 3201 CGTTCAGCC GGACCGCTGC 3401 CATTAGCAGA GCGAGGTATC 3401 CTGAAGCAC TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAAA AGGATCTTCA 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTAT 4001 GCCATTGCTA CAGGCATCGT 4101 TGTGCAAAAAA AGCGGTTAGC 4201 TCTTACTGTC ATGCCATCCCC 4201 TCTTACTGTC ATGCCATCCCC 4201 TCTTACTGTC ATGCCATCCCCCCATCCCCCCATCCT 4101 TGTGCAAAAAA AGCGGTTAGCACCTTCGTCACCTTCCTCCTCCTCCCCCCCC	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCGGGGAG CTCAAAGGCG ACCAGGCGTT GCTTTCTCAT GCTTTATCCG TAGGCGGTGC AAAAAGAGTT CCTAGATCCT TATCTCAGCG GCAATGATCA CCGCCTCCAT GGTGTCACG TCCTCAGTC TAGCCGTTCACC TAGCAGTC TATCTCAGCG TCCTTCAGTC TTATCTCAGCG TCCTTCAGTC TTATCTCACCT TATCTCAGCT TATCTCACCT TATCTCACCT TATCTCACCT TATCTCACCT TATCACCT TATCACCT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTACT GTAATACCGT CGTTTTTCCA AGCCGGTTTG GTAATACCGCT GTAACTATCC GTAACTATCC TACAGAGTTC GGTAACTATCG GTAACTATCG GTAACTATCT TTTTAAATTAA ATCTGTCTAT TCGCAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCTGTGACT	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCGC GTAGGTACT TCTTGAGTCC TTGAGTCG TATCAGGGA TATCACAGGA TATCACAGGA TATCACAGGA TATCACAGGA TATCACAGGA TATCACAGGA TATCATCACAG AATTGATCC GTATGACTTCC GTATGGCTTC TCTCAGAGT TGTCAGAGT GGTGAGTACT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG GGCCTACACTA ACAAACCAC TCTGACGTC TTAAATCAAC CTCAGAGTA CAAGTCAGACTAC GCTCAGAGTT GGGAAGCTAC ATCAGGTCC AATTCAGCTCC AAGTTCCAGATT GGGAAGCTAC ATCAGCTCC AAGTTGCCC CAACCAAGTTC	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCCACC TAGGTCGTC GACACGACT TAGGTCGTTC GACACGACT TAGGTACCAC TAGGTACCAC TAGATATT TGACTCCCCG TATACAGAATT TGACTCCCCAC TATCCAGCAT TATCAGCAT CGTTACCAC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTATT CATTCTCAGCA	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGACATGACGC ATCGCCGCTTA GCTCCAAGGT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTTA AACTCACGT ATACAGCTA AAACTCACGT ATCGCCAGTTA AACTCAGCT ATCGCCAGTTA CGCCAGTTA CATCAAGCGC ACTCAATGCT TAGTGTATATG	TGAGTTTGGA AAACAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCAA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGCTAGA ACTACGATA GCCGGAAGGG ATACTTTGCA AGTTACATA GCCGAAAGGC AGTTACATA AGCAGAAGGC AGGACCCAAGGCCAC AGGTACAGCACCAC GGCGACCCAC	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCC CTGGTAACAG TGGCTAACAG TGGCTACAG CAGCAGATTA TGGTCAGAGATTA TGGTCAGAGCC CAGTTACAG CAAGTTACAA CGGAAGGCT CCGAGCGCA CAACGTTGTT TCCCCCATGT TGCATAATTC TGCATTAATTC
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGC ATGCAAGCAT 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCG TATCAGCTCA 2901 CCAGGAACCG TAAAAAGGACC 3001 CCCGCACAGGA CTATAAAGAT 3101 CTCCCTTCGG GAAGCGGTGGC 3201 CGGTTCAGCC CGACCCGTGGC 3201 CGGTCAGCA GCGAGGTATC 3401 CTGAAGCCAG TTACCTTCGG 3501 CGCGCAGAAA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCA 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTCG CCCCAGTGCT 3901 AAGTGGTCCT GCAACTTTAI 4001 GCCATTGCTA CAGGCATCGT 4101 TGTGCAAAAA AGGGATTCG 4201 TCTTACTGTC ATGCCATCCCC 4301 CCGGCGTCAA TACCGGGATAGC	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCTGAGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAGAGTT CAAGAAGATCT TATCTCAGCG GCAATGATCA GCGTCCCAT GGTGTCACG TCCTTCGGTC TAGCGGGTC TAACAGATCT TACCCGCCCAT TACCTCGGTC TAACAGATCT TACCTCGGTC TAACAGATCT TACCTCGGTC TAACAGATCTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACCTACT GTAATACGGT CGTTTTTCCA AGCTCGTTTTCCA AGCTCACTGT TACAGAGTTC GTAACTATCG GTAACTATCG GTAGCTCT TTTTAACTTT TTTTAACTTT TTTTAACTTT TTTTAATTAA	ACTGTGCCTT TGATGCTATT GGAGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTACGGG GATCCGCAA TTCTACGGG GAATCATCT TCGTTCATC CGCGCAA TTCTACGGG GATCCTCACC GTAGGTTTCCC CTATGGATACC CTTTAAAAGT TCGTCAGAAGT TGTCAGAAAGT TGTCAGAAAGT TGTCAGAAAGT TGTCAGAAAGT TGTCAGAAAGT TGTCAGAAAGT CGTCAGGAAGT CGTCAGAAAGT CGTCAGAAAGT CTTTAAAAGT	CTAAGATCCA GCTTATTTG GCGTTGCGCT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT CCCCCTGACG CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC ACTAGTGCGT ACAAACCACC CCCCAGATT CGCACTCC CGCTCAAGTT CGCACTCC CATTGGTCC CACTTGCCC CACTCCAGATT CGCACAGTC ACTTGGCCC CACCAAGTC CAACCAAGTC CACCAAGTC CCTCATCATT	GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCGCC TACCAGACA AGCATCACAA AGCATCACAA GCATCACAC TGGTCGTC GACACGACT GGCTACACT GCTGGTACCAC GCTGGTACCA TAGACGACT TACACGACT AGTACGACAC GCTGGTACGA GCTACACT TACACACAC TACAGCAAT TGACTCCCCG TATCAGCAAT AGTAAGTAGT GGTTCCCAAC CAGTGTTACA CAGTGTTACA GGAAAACGTT	GATACATTGA AAGCTGCAAT AAGCTGCACA TTTCCAGTCG AGAACATGC AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGC AGACACTGT ACCCCACTG GCTCCAGCTTA GCTCCAAGCT ATCCCCACTG AGAACACCAC GTGGTTTTTT AAACTCACGT ATACCACCAA TCGTGAGAT ACCAGCCA TCGCAGTTA ACCAGCCA ACCAAGTTA CGCCAGTTA GATCAAGGCG ACTCAAGGTT TAGTGTAAGT CTTCGGGGCG CTTCGGGGCG	TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCG GGCTGTGTG GCACCAGCCA TATTTGGTAT TGTTTGCAAG TATTTGGTAT TGTTTGCAAG AACTACGAAT ACGGATTAC AACTACGAAT ACGGAATGG AACTACGAAC AGGATTACAT AGCGGAACGAC AGGACCAGC AGGCCACCAC AAAACTCTCA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCTC CAGCAAAAGG GGTGGGCAAA GCTCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTC CAGCAGATTA TGGTCATGAC CAGCAGATTA TGGTCATGAC CAGGAGGCCA CAGCAGTTT TCCCCATGT TCCCCATGT TTCCCCATGT TTCCTCTTGC AGGATTATC
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTAGAAGAAA 2401 TTAGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCAT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAACCTCA 2901 CCAGGAACCG TAAAAAGGCT 3001 CCCGTCAGGC CTATAAAGAT 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATC 3601 ATTAGCAGA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCAG 3501 CGCGCAGAAAA AAAAGGATCTTCAG 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCT 3901 AAGTGGTCCT GCAACCTTTAG 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAGC 4201 TCTTACTGT ATGCCATCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTAGC ATCCATTCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTAGA ATCCAGTTCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTGAG ATCCAGTTCCG	AAACCCGCTG TGCTTTATTT TAGTTCAGGG GGCGTAATCA GCCTAATGAG GCGTAGGGGAG CTCAAAGGCG ACCAGGCGTT GCTTTCTCAT GCCTTATCCG TAGGCGTTG CAAAAGAGTT CAAGAAGATC CCTAGATCCT TATCTCAGCG GCAATGATAC CCGCCTCCAT TGGTCTCAGCG TCCTTCGGTC TAAAAGAGTT TATCTCAGCG ACAATGATAC TCCTTCGGTC TAAACATGCTT TAACCCACACAA	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTAACT AGCCGGTTTG GTAATACGGT CGTTTTTCCA TCCCCCTGGA AGCTCACCTGGA AGCTCACTC TACAGAGTTC CTACAGAGTTC CTTTGATCTT CTTTGATCTT TTTAAATTAA ATCAGTCTAT CGCAGAGCCC CCAGTCTATT CTCGTCGTTTG CTCCGATCGT TCCTGTGACT CATAGCAGAT CCTCGTCGACC CATTCTTC CTCCGATCGT CTCTGTCACC CCTCGTCACC	ACTGTGCCTT TGATGCTATT GAGGTTTTTT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGGCTCCGC AGCTCCCTCG GTAGGTATCT TCTTGAGTCC TTGAAGTGGT TTGAAGTGGT TTCATCACGCA AATTGTACC AATTGTTGCC GTAGGGCTCACCG GATGGTAGGTCT TGCTCAGAAGT TCTCATCATC ACGCTCACCG AATTGTTGCC TGTCAGAAGT GGTGAGGTACT TGTCAGAAGT CGTAGAGTACT TTTTAAAAGT CAACTGATCT	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGGGCTTCTCC CAGTTCGGTG AACCAGCTTACTA ACCAGCTTACTA ACAAACCACC TCTGACGTC TTAAATCAAT CATAGTTGCC GCTCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGCCC CAACCAAGTC CTCACCATCT CTACCATCT CTACACTCT CACCATCT CTACACTCT CACCATCT CTACACTCT CACCATCTT TCACCATCTT TCACCATCTT TCACCATCTT	GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTACCT GCTGATACCT GCTGATACCT AGTGACCT AGTGACCAC AGTGAACGA TGACTCCCCC TATCAGCAAT AGTAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA TTACTTCGACAA CAGTGTTATC ATTCTGACAAA CAGTGTTATC ATTCTGACAAA	GATACATTGA AAGCTGCAAT AAGCTGCATA CAATTCCAGCA TTTCCAGCA GAGACATGTG AAATCGACGC CTGCCGCTTA AGAGGACATGTG AGAAGATTTTTT AAACTCACGT TATGAGTAAA TCGTGAGTAAA TCGTGAGTAAA AACCAGCCA TCGCCAGTTA GATCAAGGC ACTCATGGTAAT TAGAGTAAA CAACCAGCAGTTA CATCAAGGCG ACTCATGGTTATGGTATAGAT CATCAAGGCG ACTCATGGTT TAGTGTATAGC CTTCGGGGGCG CAGCGTTTCT	TGAGTTTGGA AAACAGTTA CGACTACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTCGA ATGGCACCA ATGGCACCAG ATGGCACCAG ATGGCACCAG AGGTACATCACA AGGCACCAG AAAACTCCAG GAGACCAG GAAACCACAG GGCGACCAG AAAACTCCAG	CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA GGTCGCGCTC CACCAAAAGG GGTGGCGAAT GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCT CACGAACCCC CTGGTAACAG CTGCGCTCT CACGAAGTTA TGGTCATGAG CACTATACCAA CAGCAGCATT TCGCCCATGT TCCCCCCATGT TCCCCCATGT TTCCATCATGA AGGATCTTAC AAACAGGAAG
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTAGAACAAA 2401 TTGCATTCAT TTTATGTTTC 2501 ACCTGCAGC ATGCAAGCAT 2601 TAAAGTGTAA AGCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CCGCGAGAGCG TATCACCTCA 2901 CCAGGAACCG TATCACCTCA 2901 CCCGCACAGGA CTATAAAGAT 3101 CTCCCTTCAGC GAAGCGGTGGC 3201 CCGTTCAGCC GAAGCGCTGCC 3301 GATTAGCAGA GCGAGGTATC 3401 CTGAAGCAA AAAAGGATCT 3601 ATTATCAAAAA AGGATCTTCA 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG 3901 AAGTGGTCCT GCAACTTTAT 4001 GCCATTGCTA CAGGCATCCG 4101 TGTGCAAAAA AGCGGTTAGC 4201 TCTTACTGTC ATGCCATCCC 4301 CCGCGCTCAA 4401 CCGCGCTCAA 4401 CGCTGTTGAA ATCCAGTTCAG 4501 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTGAG ATCCAGTTCC 4301 CCGCCTTAACC GCAACATCAG 4501 CCGCCTTAACCAGTTCC 4301 CCGCCTCTAAA	AAACCCGCTG TGCTTTATTT TAGGTTCAGG GGCGTAATCA GCCTAATGAG GCGGGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTT GCTTTCTCAT GCCTTATCAT CAGAAGAGTC TAAGAGAGTC TAATCAGCG GCAATGATCC TATCTCAGCG TCATCAGCT TATCTCAGCG TCATCAGCT TATCTCAGCG TCATCAGCT TATCTCAGCG TCATCAGCT TATCTCAGCG TCCTTCGGTC TAAGATGATC TAACATGATCCT TAACAGCGCCA ATGTAACCCA GAATAAAGGGC	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGACCTACCT GTAATACGGT CGTTTTTCCA AGCCGGTTTG GTAATACGGT TCCCCCTGGA AGCTACCT GTACTATCC GTAACTATCG GTAACTATCG GTAACTATCG GTAACTATCG CTTTGATCTT TTTTAAATTAA ATCTGTCTAT TCGGAGACCC CCAGTCTATT TCGTCGTTTG CTCCGATCGT TTCTGTGACT CTTCTGTGACT CTTCTGTGACT CTTCTGTGACT CTTCTGTGACT CTTCTGTGACT CATAGCAGAA CCCGACCAGAA	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGGC TATCCACAGA TAGGCTCCTG GTAGGTACT TCTTGAGTCC TTGAAGTGGT AATCGACAA TTCTACGGGA ATTCTACGGG AATCGTCACC AATTGTTCAC ACGCTCACC GTAGGTACT CTTGAAGGT CTGTCAAAGT TTCAGAAGT TTTCACAGAGG GGTGAGTACT CTTTAAAAGT CTTTAAAAGT CTACTGATCAC TGTTGAAAGT CTTTAAAAGT CTGTTGAATAC	CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGGAT CCCCCTGACG GCCCCCGTAAA ACCAACCA TCTGACGTC TAAATCAACT CTGACGTC GCTCCAGATT GGGAAGCTAC ATCAGGTCC AAGTCCAGATT GGGAAGCTAC ATCAGCTCC AAGTTGCCC CAGCTCC CAGCTCC CAGCTCC CAGCTCC CAGCTCC CTCCAGATT CGCAACACTC CTCCACATT TCAGCTCC TCAGCACTCT TCAGCTCT TCAGCATCT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCAGCATCTT	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TGTTCCGACC TAGGTCGTC GACACGACTT GCTGACACC CAGGTACACA CATAAAGTATA TGACTCCCCG TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAC TATTCTAGACA GGAAAACGTT TTACTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AAATCGACGC CTGCCGGCTTA GCTCCAAGCT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG AAACTCACGC ATCGCCACTTA AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGGC ACTCATGATTA CGCCAGTTA CGCCAGTTA CTCGGGGGGC CACGCTTTCT TAGTGTATGC CTCGGGGGCC CACGCTTTCT ATTTATTGAA	TGAGTTTGGA AAACAAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA GCTTGGTATACATA GCCGGAAGGG ATTACATACATA GCCGGAAGGG AGTTACATGA AGTACATACATA ACTACATA ACGCGCACCAC AGTTACATGA AGTACATACATA AGCCGAAGGA AAACTCTCA GGCGACCGAG AAACTCTCA GGGTGAGCAA GCATTTATCA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGGGAA GTCGCGCTTT CACGAACCCC CTGGTAACAG CTGCGCTCT CACGAGATTA TGGTCATACAG CAGATTAT TGGTCATGAG CAACTTATCAG CAACTTATCAG CAACTTATCAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATTT
2201 TCACCATCAC CATTGAGTTI 2301 CTAGAATCAC GTAGAAGAAA 2401 TTAGCATTCAT TTTATGTTTC 2501 ACCTGCAGGC ATGCAAGCAT 2601 TAAAGTGTAA ACCCTGGGGT 2701 GCATTAATGA ATCGGCCAAC 2801 CGGCGAGCGG TATCAACCTCA 2901 CCAGGAACCG TAAAAAGGCT 3001 CCCGTCAGGC CTATAAAGAT 3101 CTCCCTTCGG GAAGCGTGC 3201 CGGTCAGCC CGACCGCTGC 3301 GATTAGCAGA GCGAGGTATC 3601 ATTAGCAGA AAAAGGATCT 3601 ATTATCAAAA AGGATCTTCAG 3501 CGCGCAGAAAA AAAAGGATCTTCAG 3701 TGCTTAATCA GTGAGGCACC 3801 TACCATCTGG CCCCAGTGCT 3901 AAGTGGTCCT GCAACCTTTAG 4001 GCCATTGCTA CAGGCATCG 4101 TGTGCAAAAA AGCGGTTAGC 4201 TCTTACTGT ATGCCATCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTAGC ATCCATTCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTAGA ATCCAGTTCCG 4301 CCGGCGTCAA TACGGGATAA 4401 CGCTGTTGAG ATCCAGTTCCG	AAACCCGCTG TGCTTTATTT TAGGTTCAGGG GGCGTAATCA GCCTAATGAG GCCTGAGGGAG CTCAAAGGCG GCGTTGCTGG ACCAGGCGTTCTCAT GCCTTATCCG TAGGCGGTGC AAAAGAGTT CAAGAAGATCT TATCTCAGCG GCAATGATCA GGTGTCACGC TCCTCCAT GGTGTCACGC TCCTTCGGTC TAACAGATCT TACCCGCCCAT TACCGCGCCA TCTTCTCGTTC TACCGCGCCA ATGTAACCCA ATGTAACCCA AGAATAAGGGC TGAATAATGTTT	ATCAGCCTCG GTGAAATTTG GGAGGTGTGG TGGTCATAGC TGGTCATAGC TGACCTACG GTAATACGGT CGTTTTTCCA AGCTCGTTTTCCA AGCTCACCT GTAACTATC GTAGCTACCT GTACTACT TCTTGACCTT TTTTAACTTA ATCTGTCTAT TCGCAGAGCCC CCAGTCTAT TCGCGATCGT TTCCGTTGGCT TTCTGTGACTT TCTTCGTTGACT TCGTCGTTTG CAGTCGATCGT TCCGTTGGCT CATAGCAGAA CTCGTGCACC GACACCGACAC GACACGGAAA TCGGACAC GACACGGAAA TCGGACAC GACACGGAAA	ACTGTGCCTT TGATGCTATT TGATGCTATT TGTTTCCTGT CACATTAATT CGTATTGGC TATCCACAGA TAGGCTCCGC GTAGGTATCT TCTTGAGTCC TTGAAGTGG GATCCGCAA TTCTACGGG GAATCATCAT TCGTTCATC ACGCTCACCG GTATGCTCACC GTATGCTCACC GTATGCTCACC GTATGCTTC TGTCAGAAGT TGTTCAGAAGT TGTTCAGAAAGT TGTTCAGAAAGT CAACTGATCT TGTTGAATAC AACAAATAGG	CTAAGATCCA GCTTATTTG AAAGCAAGTA GTGAAATTG GCGTTGCGGT ATCAGGGGAT CCCCCTGAAG TGCGGTCTCC CAGTTCGGT AACCCGGTAA ACAAACCACC TCTGACGTC ATAGTTGCC GCTCCAGATT CATAGTTGCC GCTCCAGATT ATTCAGCTC AAGTTGGCG CACCAAGTC CAACTAGTTGCCG CACCAAGTC CACCAAGTC CACCAAGTC CACCAAGTC CTCACATT TCAGCACTTT TCAGCACTTT TCAGCACTTT GGTTCCCGCG	GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA AACGCAGGAA AGCATCACAA AGCATCACAA GCATCACAC TGTTCCGACC TAGGTCGTC GACACGACTT GCTGACACC CAGGTACACA CATAAAGTATA TGACTCCCCG TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAT TATCAGCAC TATTCTAGACA GGAAAACGTT TTACTTTCAC CCTTTTTCAC	GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACCA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AAATCGACGC CTGCCGGCTTA GCTCCAAGCT ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCGCCACTG ATCACCACTG AAACTCACGC ATCGCCACTTA AAACTCACGT AAACTCACGT AAACTCACGT AAACTCAGGC ACTCATGATTA CGCCAGTTA CGCCAGTTA CTCGGGGGGC CACGCTTTCT TAGTGTATGC CTCGGGGGCC CACGCTTTCT ATTTATTGAA	TGAGTTTGGA AAACAAGTTA CAACATACA TGGCTGATTA CAACATACGA GGAAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTGTG GCACCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTA GCTTGGTATACATA GCCGGAAGGG ATTACATACATA GCCGGAAGGG AGTTACATGA AGTACATACATA ACTACATA ACGCGCACCAC AGTTACATGA AGTACATACATA AGCCGAAGGA AAACTCTCA GGCGACCGAG AAACTCTCA GGGTGAGCAA GCATTTATCA	CAAACCACAA ACAACACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAGG GGTGGGGAA GTCGCGCTTT CACGAACCCC CTGGTAACAG CTGCGCTCT CACGAGATTA TGGTCATACAG CAGATTAT TGGTCATGAG CAACTTATCAG CAACTTATCAG CAACTTATCAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATCTAG CAACTTATTT

# pMT-HisFlag-DmSNAP43 mut#14 STOP

	GAC GGTGAAAACC TCTGACA	CAT GCAGCTCCCC	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101 TCAGGGCGCG TCAGCGG							
201 CCGCACAGAT GCGTAAG							
301 TACGCCAGCT GGCGAAA							
401 TGAATTAATT CGTTGCA	GGA CAGGATGTGG TGCCCGA	TGT GACTAGCTCT	TTGCTGCAGG	CCGTCCTATC	CTCTGGTTCC	GATAAGAGAC	CCAGAACTCC
501 GGCCCCCCAC CGCCCAC							
601 CAAGTCCCCA AAGTGGA							
701 AGAGGTGAAT CGAACGA							
801 AAATCAAGTG AATCATC							
				> M H	н н н	ннтр	Y K D
901 GACGATGACA AGGGCAC	TAG TGAGCTGAAT ATCTTTG	ACG ACTGCTGGGA	GCTGGTGCAA				
> D D D K G T					R L V N	D G E	N C E
1001 TCGAGGTGTT CTGCCGG						GCCACCACAC	TGGCGGCCCT
>FEVFCR		OHLF				ATT	LAAL
1101 GCATGTGGCC AAGCGAC							
> H V A K R				R A O		G F F L	L Y V
1201 ATCTACTACA AGCAGCC							
		I E V S E			TDYA		R K D
1301 GTCCGGAGCG GAAGGAC	ACT CATCAGATCG CCTACAT	GCT GTGGCGCCTG	ACCCAGGAGC	AGGCCTTCCG	CTTCACCGCG	CTCGACTATT	GCCAGGGGTT
>S P E R K D		L W R L	T O E	OAFR		L D Y	COGL
1401 GGACAATCTG GTGGACT							
> D N L V D		A G A K				OORA	N G V
1501 AGTCTCACAT ACGAACT	GA GGGTCTGCGA GCACTGG	ACC AGGCAAGCCA	GCCATTGTGT	GAACTGGAAG	CGGCATACAA	TGCCCAAAAG	AAGCAATTGG
> S L T Y E L	E G L R A L	OASO	PLC	E L E	AAYN	A O K	K O L
1601 CGGCTGGTCA TGAGCAC							GAAAGAGTAC
>A A G H E H		F G H L				L G A	RKST
1701 TCCAGATGAG AAATGCA							GCGGGAGCCG
> P D E K C '	r A A A A A N	O L E V	R O R V	R N K	A M Y	G V E E	R E P
1801 CAACACCAGA CGGATGA	ACT AGAAGTGCAG CTGGAGG	CA ACGAGACTTA	TCAACGCCGC	ATGTCCTCGG	CCACCGTTTT	CCAGAGGGAA	CTTCCAGAAG
> Q H Q T D E	LEVQLE	V NETY	QRR	M S S	ATVF	QRE	L P E
1901 ACGTGCAGCA AGAGTAT	GAG ATGATTGAGT TTAGTGA	CGA CGAGGAAATO	GAAGTGGGTG	AAAGCGAGGA	GGTCACGGAA	GAAGAACTCA	AAGCTATTTT
	E M I E F S I	пеем	E 17 C	E S E E	VTE	E E L	KAIL
>D V Q Q E Y		D E E M	E V G		VIE		K A I L
>D V Q Q E Y 2001 GGATACTTGA GTTAGTT							
2001 GGATACTTGA GTTAGTT	FAT AAAACTTTTA CATAATT	AAA TAACTAGCAT	TTTTGCGCGA	TGTGATCTTG	TTTATCTGAA	GGGCAATTCT	GCAGATATCC
2001 GGATACTTGA GTTAGTT > D T	FAT AAAACTTTTA CATAATT	AAA TAACTAGCAT	TTTTGCGCGA CCTATCCCTA	TGTGATCTTG ACCCTCTCCT	TTTATCTGAA CGGTCTCGAT	GGGCAATTCT TCTACGCGTA	GCAGATATCC CCGGTCATCA
2001 GGATACTTGA GTTAGTT > D T 2101 AGCACAGTGG CGGCCGC	FAT AAAACTTTTA CATAATT FCG AGTCTAGAGG GCCCGCG FTT AAACCCGCTG ATCAGCC	AAA TAACTAGCAT GTT CGAAGGTAAG FCG ACTGTGCCTT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA	TGTGATCTTG ACCCTCTCCT GACATGATAA	TTTATCTGAA CGGTCTCGAT GATACATTGA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA	GCAGATATCC CCGGTCATCA CAAACCACAA
2001 GGATACTTGA GTTAGTT > D T 2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG	FAT AAAACTTTTA CATAATT FCG AGTCTAGAGG GCCCGCG FTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT	AAA TAACTAGCAT GTT CGAAGGTAAG CCG ACTGTGCCTT CTG TGATGCTATT	TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTTG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA.	FAT AAAACTTTTA CATAATT FCG AGTCTAGAGG GCCCGCC FTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT FTC AGGTTCAGGG GGAGGTC	AAA TAACTAGCAT GTT CGAAGGTAAG ICG ACTGTGCCTT ITG TGATGCTATTI IGG GAGGTTTTTT	CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA	GCAGATATCC CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TTT GGCGTAATCA TGGTCAI	AAA TAACTAGCAT GTT CGAAGGTAAG ICG ACTGTGCCTI ITG TGATGCTATI IGG GAGGTTTTTI AGC TGTTTCCTGT	CTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA	CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTTATGT  2501 ACCTGCAGGC ATGCAAGG	PAT AAAACTTTTA CATAATI PCG AGTCTAGAGG GCCCGCC PTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAI PTC AGGTTCAGGG GGAGGTC TCT GGCGTAATCA TGGCCAI GGT GCCTAATGAG TGAGCTA GGT GCCTAATGAG TGAGCTA	TAACTAGCAT TAACTAGCAT TCG ACTGTGCCTT TCG TGATGCTATT TCG GAGGTTTTTT AGC TGTTTCCTGT ACT CACATTAATT	CTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAACTGTAA AGCCTGG	FAT AAAACTTTTA CATAATI FCG AGTCTAGAGG GCCCGCG FTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT FTC AGGTTCAGGG GAAGGTC TT GGCGTAATCA TGGTCAT GGT GCCTAATGAG TGAGCTT AAC GCGCGGGGAG AGGCGGT	AAA TAACTAGCAT  CGAAGGTAAC  CGACGTTGCTT  CGGAGGTTTTTT  CGGAGGTTTTTT  ACC TGTTTCCTGT  ACT CACATTAATI  CGTATTGGGC	CCTATCCCTA CTAACATCCA GCTTTATTG AAAGCAAGTA GGGTTGCGCT GCGTTCCGC	TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC TATCGGCTCA CACTGCCCGC TTCCTCGCTC	TTTATCTGAA CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAG	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TC AGGTTCAGGG GGAGGTC TTT GGCGTAATCA TGGTCAT GGT GCCTAATGAG TGAGCTA AAC GCGCGGGGAA AGCCGGT ICA CTCAAAGGCC GTAATAC SCC GCGTTGCTGG CGTTTTT	TAACTAGCAT GAAGGTAAC TGG ACTGTGCCTI TGG TGATGGTATT AGC TGTTTCTG ACT CACATTAGT TTG CGTATTGG ACT ATCCACAGA TAGGCTCCGC TAGGCTCCGC TAGGCTCCGC	TTTTGCGCGA CCTATCCCTA CCTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGC ATCAGGGGAT CCCCCTGACG	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC TATCCGCTCA CACTGCCCGC TCCCTCGCTC AACGCAGGAA AGCATCACAA	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA TTTCCAGTCG ACTGACTCGC ACACACTGTG AGAACATGTG AAATCGACGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGGTCGGT AGCAAAAGGC TCAAGTCAGA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATCA GTGAAAA. 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCAGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAG 3001 CCCGACAGGA CTATAAAA	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTAAGGG GGAGGTG TTC GGCGTAATCA TGGTCAA GGC GCCGGGGGAG AGGCGGT TCA CTCAAAGGCG GTAATAG SCC GCGTTGCTGG CGTTTTT SAT ACCAGGCGTT TCCCCCT	TAACTAGCAT GAAGGTAAG GG ACTGTGCCTT TG TGATGCTATT GG GAGGTTTTT AGC TGTTTCCTGT TTT CGTATTCCTGT TTT CGTATTCATATT TTT CGTATTAGT TTT CACATAAT TAGCTCACAGA GAAGATCCCCCGG GAA AGCTCCCCCGG GAA AGCTCCCTCC	TTTTGCGCGA CCTATCCCTA CTAAGATCCA GCTTTATTG AAAGCAAGTA GTGAAATTGT GCGTTGCGCT GCTCTTCCGC ATCAGGGAT CCCCCTGACG TGCGCTCCCC	TGTGATCTTG  ACCCTCTCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCAGCA ATTCCAGTCG ACTGACTCGC AGAACATGTG AAATCGACGC CTGCCGCTTA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGATCAGA CCGGATACCT	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GGTGGCGAAA GTCCGCCTTT
2001 GGATACTTGA GTTAGTT  D T  2001 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTATAGT  2501 ACCTGCAGGC ATGCAAG  2601 TAAAGTGTAA ACCTGGC  2701 GCATTAATGA ATCGGCC.  2801 CGGCGAGCGG TATCAGC  2901 CCAGGAACCG TAAAAAAG  3001 CCCGACAGGA CTATAAAA  3101 CTCCCTTCGG GAAGCGT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATGA TGAGCTA ACC GCGCGGGGA AGGCGGT TCA CTCAAAGGCG GTAATAC GCC GCGTTGCTGG CGTTTTT GCC GCTTTCTCAT AGGTCAC GCG GCGCGGGGTATATAC GCG GCTTTCTCAT AGGTCAC GCG GCTTTCTCAT AGGTCAC GCG GCTTTCTCAT AGGTCAC	TAACTAGCAT TO GAAGGTAAG TCG ACTGTGCCTT TIG TGATGCTATT AGC TGTTTCCTGT AGC TGATTAGT TIG GGAGTTTTT TIG CGTATTGGG TTTCCACAG TAGCTCCCTCC TAGGTTCCTCC TGTAGGTATTCC TGTATTGGG TTTCCACAG TTTTCCACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACACAC TTTTCACACACAC TTTTCACACACA	TTTTGCGCA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCCC CAGTTCGGC CAGTTCGGC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGA CACTGCCGGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTCGC ACGACACTGC ACACACTGC CTGCCGCTTA GCTCCAAGCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATCA GTGAAAA. 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAAGTGTAA AGCCTGG 2701 GCATTAATGA ATCAGCC 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAG 3001 CCCGACAGGA CTATAAAA	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATGA TGAGCTA ACC GCGCGGGGA AGGCGGT TCA CTCAAAGGCG GTAATAC GCC GCGTTGCTGG CGTTTTT GCC GCTTTCTCAT AGGTCAC GCG GCGCGGGGTATATAC GCG GCTTTCTCAT AGGTCAC GCG GCTTTCTCAT AGGTCAC GCG GCTTTCTCAT AGGTCAC	TAACTAGCAT TO GAAGGTAAG TCG ACTGTGCCTT TIG TGATGCTATT AGC TGTTTCCTGT AGC TGATTAGT TIG GGAGTTTTT TIG CGTATTGGG TTTCCACAG TAGCTCCCTCC TAGGTTCCTCC TGTAGGTATTCC TGTATTGGG TTTCCACAG TTTTCCACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACAG TTTTCACACACAC TTTTCACACACAC TTTTCACACACA	TTTTGCGCA CCTATCCCTA CTAAGATCCA GCTTTATTTG AAAGCAAGTA GTGAAATTGT GCGTTGCGC ATCAGGGGAT CCCCCTGACG TGCGCTTCCCC CAGTTCGGC CAGTTCGGC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCGA CACTGCCGGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA CAATTCCACA ACTGACTCGC ACGACACTGC ACACACTGC CTGCCGCTTA GCTCCAAGCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGGCTGTGTG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCGAAA GTCCGCCTTT CACGAACCCC
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA. 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAGG 2601 TAAACTGTAA ATCAGCC. 2801 CGGCGAGCGG TATCAGC 2901 CCAGGAACCG TAAAAAG 3101 CCCGACAGGA CTATAAAA 3101 CTCCCTTCGG GAAGCGT 3201 CCGTTCAGCC CGACCGC 3301 GATTAGCCA GCGACGCC 3301 GATTAGCCA GCGACGCGC	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TTT GGCGTAATCA TGGTCAT GGT GCCTAATGAG TGAGCTF AAC GCGCGGGGAG AGGCGGT TCA CTCAAAGGCG GTAATAG SCC GCGTTGTGTG GCTTTT SAT ACCAGGCGTT TCCCCCT SGC GCTTTCTCAT AGCTCAC TGC GCCTTATCCG GTAACTF ATG TAGGCGGTGC TACAGAG	TAACTAGCAT GAAGGTAAG GG ACTGTGCCTT TIG TGATGCTAT TIG GAGGTTTT AGC TGTTTCCTG TIG CTATTCCTG TIG CTATTGGG GGT TATCCACAGA CA TAGGCTCCG GGA AGCTCCCTC GGT GTAGGTATC TIC CTTTGAGTGGT TITTGAGTGGT TITTGAGTGT TITTGAGTGT TITTGAGTGGT TITTGAGT T	TTTTGCGCGA CCTATCCCTA CTAGGATCCA CCTTATTTG AAAGCAGTA GGGAATTGCT GCGTTGCGCT ATCAGGGGAT CCCCCTGACG TCGCGTCTCCC CAGTTCGGTG AACCCGGTAA GGCCTAACTA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAAT TAACCATTAC TATCCGCTCA CACTGCCGC TCCCTCGCTC AACCCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACTTC GACACGACTT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATCGC AGAACATCGC CTGCCGCTTA GCTCCAACGT ATCGCACCT AGAAGCACACA ATCGACCACT ACGACGCACT ACGACGCACT ACGACGACACACACACACACACACACACACACACACACA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATAC GGAAAACTGT TGCCGTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGTG GGCTGTGTG GGACCAGCCA TATTTGGTAT	GCAGATATCC  CCGGTCATCA CAAACACACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCGAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTTG
2001 GGATACTTGA GTTAGTT  D T  2001 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTATAGT  2501 ACCTGCAGGC ATGCAAG  2601 TAAAGTGTAA ACCTGGC  2701 GCATTAATGA ATCGGCC.  2801 CGGCGAGCGG TATCAGC  2901 CCAGGAACCG TAAAAAAG  3001 CCCGACAGGA CTATAAAA  3101 CTCCCTTCGG GAAGCGT  3201 CAGTTAGCC CGACCGC  3301 GATTAGCAGA GCGAGGT.  3401 CTGAAGCCAG TTACATT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT AGC GCGCGGGGA GCGCGT TCA CTCAAAGGCG GTAATAC CCG GCGTTCTTG CCTTTTT AGC CGCGCGTA TCCCCCT AGC GCTTTTCTCAT AGCTCAC TCG GCTTTTTCAT AGCTCAC TCG AAAAGAGTT GGTAACTAC TCG AAAAAGAGTT GGTAACTAC TCG TCGCCGCTG TAACTAC TCG AAAAAGAGTT GGTAACTAC TCG AAAAAGAGTT GGTAACTAC TCG TCGCCGCTG TCCACAC TCGCCGCTG TCCACACAC TCGCCGCTG TCCACACAC TCGCCGCTG TCCACACAC TCGCCGCTG TCCACACAC TCGCCGCTG TCCACACAC TCGCCGCCGCCCCCCCCCCCCCCCCCCCCC	TAACTAGCAT TO GAAGGTAAG TCG ACTGTGCCTT TTG TGATGCTATT AGC TGTTTCCTGT AGC TACATTAGT TTG CGTATTGGG TTG CTATTGGG TTG TAGCTCCG TAGGTCCCTC TGAGTACT TTGAGTACT TTGAAGTGGT TTTGAAGTGGT TTTGAAGTGT TTTGAAGTGT TTTGAAGTGT TTTGAAGTGT TTTGAAGTGT TTTGAAGTGT TTTGAAGT TTTGAAGTGT TTTGAAGTGT TTTGAAGT TTTGAAGT TTTGAAGTGT TTTGAAGTGT TTTGAAGTGT TTTGAAGT TTTTGAAGT TTTTGAAGT TTTGAAGT TTTTGAAGT TTTTTTTT	TTTTGCGCGA CCTATCCCTA CTAGATCCA CCTAGATCCA CCTTATTTG AAAGCAACTA GTGAAATTGT GCGTTTGCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCC CAGTTCGCC AGCTCGCC AGCTCGCT AACCGGTAA	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TTCCCTCGCTCA AGCATCACAA GCATCACAA GCATCACCAC TGTTCCCAC TAGGTCGTC GACACGACTT CGGCTACACT GCTGCTACACT GCTGCTACACT GCTGCTACCC GCTGGTACCG	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATCACACA TTTCCAGTCG AGAACATGTG AAATTGAGCG CTGCCGCTTA GCTCCAAGCT ATCGCCACTG AGAAGACTGTCGCACTG AGAGGACAGG GTGGTTTTTTT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CGACTGATTA CAACATCGA GGAAAACCTGT TGCGCTCCGGT AGCAAAACGC TCAAGTCAGA CCGGATAACCTG GCAGTGTTG GCACCAGCACCA TATTTGGTAT TGTTTGCAAG	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCAAA GTCCGCCTTT CACGAACCC CTGGTAACAC CTGCTCTCT CAGCAACTCT CAGCAACTTA
2001 GGATACTTGA GTTAGTT  D T  2101 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTTATGT  2501 ACCTGCAGGC ATGCAAG  2601 TAAAGTGTAA ACCTGGC.  2701 GCATTAATGA ATCGGCC.  2801 CGGCAGGGG TATCAGC  2901 CCGGACAGGA CTATAAAA  3001 CCCGACAGGA CTATAAAA  3101 CTCCCTTCAG GAAGGGT  3201 CGGTTCAGCC CGACCGC  3301 GATTAGCAGA GCGAGGT.  3401 CTGAAGCCAG TTACCTTC  3501 CGCGCAGAAA AAAAGGA	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAA TTC AGGTCAGGG GGAGGTC TT GGCGTAATGA TGGTCAA GGT GCCTAATGAG TGAGCTA AAC GCGCGGGGGA AGGCGG TAA TCAAAAGGCG GTAATAC GCC GCGTTGCTGG CGTTTTI AACAGGCGTT TCCCCCT GGC GCTTTCTCAT AGGTCAC TGC GCCTTATCGG GTAACTA TGT TAGGCGGTGC TACAGAG TGG AAAAAGAGTT GGTAGCT TCT CAAGAAGATT CTTTGAT	TAACTAGCAT GAAGGTAAG TCG ACTGTGCCTT TTG TGATGCTATT AGC TGTTTCTG ACT CACATTAATT TTG CGTATTGGG GGT TATCCACAG CCA TAGGCTCCTCC TCG TGTAGTGTATT TCG TCTTGAGTCC TCT TTGAGTACT TTT TGTAGTGGC TTT TTGAGTGCT TTT TTTACGGGGAT TTTTACGGGGAT TTTTACGGGGGAT TTTTACGGGGAT TTTTACGGGGAT TTTTACGGGGAT TTTTACGGGGAT TTTTACGGGGAT TTTTACGGGGAT TTTTTACGGGGAT TTTTTACGGGAT TTTTTACGGGAT TTTTTACGGGAT TTTTTTACGGGAT TTTTTACGGGAT TTTTTTACGGGAT TTTTTTACGGGGAT TTTTTTACGGGAT TTTTTTACGGGAT TTTTTTTTTACGGGAT TTTTTTTTTT	CCTATCCCTA CTAAGATCCA CTAAGATCCA CTAAGATCCA CTTATTTG AAAGCAGTA CTGAAATTGT GCGTTGCGCT GCTCTTCCGC TCCGCGCTGCGT TCCGCTCACG AACCCCCTAACA GCCTAACTA GCCTAACCACC TCTGAGCGTC TCTGAGCGTC TCTGAGCGTCACC TCTGAGCGTCACC TCTGAGCGTCACC	ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CATGCCGGC TTCCTGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTTC GACACGACT GGTGGTACAC GGTGGAACG AGTGGAACG AGTGGAACG AGTGGAACG	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTCACACA TTCCAGTGG ACTGACTGG ACTGACTGG ACTGACTGG CTGCCGCTTA AAATCGACGC ATCGCCACTG ACTCAAGCT ATCGCCACTG AGAAGACACTG ACGCCACTTA AGAAGACACTG ACGCCACTTA AGAAGGACAG GTGGTTTTTT AAACTCACGT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGCTAC TGTTTGCAAG TAAGGGATTC	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CAGCAAAAAGG GGTGCCAAAA GTCCGCCTTT CACGAACCT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA. 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAACTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC. 2801 CCGGCAGAGCG TATCAGC 2901 CCAGGAACCG TAAAAAG 3101 CCCGACAGGA CTATAAAA 3101 CTCCCTTCGG GAAGGGT 3201 CCGTTCAGC CGACCGC 3301 GATTAGCAG GCGAGGT 3401 CTGAAGCAG TTACCTT 3501 CGCGCAGAAA AAAAGA 3601 ATTATCAAAA AGGATCT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCT AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GAC GCGGGGGAG TAGCCGC ACC ACCTTCTGGGGGGAGT TCA CTCAAAGGCG GTAATAA CCGCGGGGAG AGGCGGT TCA CTCAAAGGCG GTAATAA GCC GCGTTCTCAT AGCTCAC TCG GCTTTCTCAT AGCTCAC TCG CCCTTATCCG GTAACTA TGG TAGCGGGTGC TACAGAG TGC AAAAAGAGTT GGTAGCT TCT CAAGAAGATC CTTTGAT TCA CCTAGATCCT TTTAAAT TCA	TAACTAGCAT GAAGGTAAG GAAGGTAAG GAAGGTATAGT GAAGGTATAGAT AGC GAAGGTATAGAG GAAGGTATAGAG GAAGGTACAGAG GAAGGTACAGAG GAAGGTACAGAG GAAGGTACAGAGAG GAAGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	TTTTGCGCGA CCTATCCCTA CTAGGATCA CCTAGGATCA GCTTTGCTC GCTTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCTCCC CAGTTCGGTG AACCCGGTAA AACCAACCACC TTTGAATCAT TTTAAATCAAT	ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC GACACGACTAC AGTGGAACGA CTAAAGTATA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGGTA CAATTCCACA TTTCCAGTCG ACTGCACTCGACA CTGCCGCTTA GCTCCAACGT ATCCCACGT ATCCCACGT ATCCCACGT ATCGCCACT ATCGCCACT ATCGCCACT ATCGCACT ATCACT ATCACT TATCACT TATCACT TATCACT TATCACT TATCACT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACA GGAAAACTGT TGCCGTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGGCTGTGT GCAGCCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT	GCAGATATCC  CCGGTCATCA CAAACACAAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGTCGGCTG CAGCAAAAGG GTCGCGCTT CACGAACCCC CTGGTAACAG CTGCGCTTT CACGAACCCC CTGGTAACAG CTGCGCTTT CAGCACATTA TGGTCATGAG CAGTTATACCAA
2001 GGATACTTGA GTTAGTT  D T  2001 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTATAGT  2501 ACCTGCAGGC ATGCAAG  2601 TAAAGTGTAA ACCTGG  2701 GCATTAATGA ATCGGCC.  2801 CGGCGAGCGG TATCAGC  2901 CCAGGAACCG TAAAAAA  3001 CCCGTCAGGA CTATAAAA  3101 CTCCCTTCGG GAAGCGT  3201 CAGTTAGCA GACGGC  3201 CAGTTAGCAGA CTATAAAA  3601 GATTAGCAGA TTACCTT  3501 CGCGCAGAAA AAAAGGA'  3601 ATTATCAAAA AGGATCT  3701 TGCTTAATCA GTGAGGCC	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCT AGGTTAGGG GAGGTG TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT AAC CGGCGGGGA GGCGGT TCA CTCAAAGGCG GTAATAC GCG GCGTTCTTG AGT TCCCCCT AGC GCTTTCTCAT AGCTCAC TCG GCTTTTCCAT AGCTCAC TCG AAAAAGAGTT GGTAGCT TCT CAAGAAGATC CTTTGAT ACC TATCTCAGCG ATCTGTC TCA CTAGATCCT TTTAAAT ACC TATCTCAGCG ATCTGTC	TAACTAGCAT TO GAAGGTAAG TCG ACTGTGCCTT TTG TGATGCTATT AGC TGTTTCCTGT AGC TGATTAGT TTG CGTATTAGGC TGTTTCCACAG TCA TAGGTCCCC TGTAGGTACT TTGAGTACT TTGAGTGC TTTGAGTGC TTTGAGTGC TTTGAGTGC TTTGAGTGGT TTTTACAGGGA TCTTTTCACGGGA TCTTTTCATCGTTAAT TTCTTCATT	TTTTGCGCGA CCTATCCCTA CTAGATCCA CCTAGATCCA CCTTATTTG GTGAAATTGT GCGTTTCCGC ATCAGGGGAT CCCCCTGACG TGCGCTCCCC CAGTTCGGT AACCCGGTAA ACGAACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC CATTTTTACC CAGTTTGCTC CAGTTCGTG AACCACC TCTGACGCTC TCAACTA CATAGTTGCC CATACTA CATACTTACC CATACTTA CATACTTTCCC CATACTTA CATACTTTCCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT TATCCGCTCA CACTGCCCGC TACCCGGCTC AACGCAGGAA AGCATCACAA TGTTCCCACC TAGGTCGTT GACACGACTT GCACTACACT GCTGGTACACT GCTGGTACACT GCTGGTACACT CTGGTACACT TGGTACACA TAGATACA AGTATA TGACTCCCCC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATCTGACTGA AATCCAGTCG AGAACATGTG AAATCGACGC AGAACATGTG AAATCGACGC ATGCCCACTGCCGCTTGCCGCTTA GCTCCAAGCT ATCGCACTG AGAAGACATGTTA AAATCAAGTTA AAATCAAGTTA AAATCAAGTTA AACTCAAGTT TATGAGTAAA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CGACTACGA GGAAAACTGA TGCCGCTCGGT AGCAAAACGCT TCAGTCAGA CCGGATACCT GGGCTGTGTG GCACCACCA TATTTGGTAT TGTTTTGCAAG TAGGGATTT CTTGGTCTTA AACTACGAT	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GGTGGCAAA GCCGCCTT CACGAACCC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CTGGTAACAC CAGCACTTA CAGCACTTA CAGCACTTA CAGCACCC CAGCAGATTA CGGGAGGGCT CGGGAGGGCT CGGGAGGGCCT
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTTATGT  2501 ACCTGCAGGC ATGCAAG  2601 TAAAGTGTAA ACCCTGG  2701 GCATTAATGA ATCGGCC.  2801 CGGCAGGGG TATCAGC  2901 CCAGGAACCG TAAAAAG  3001 CCCGACAGGA CTATAAAA  3101 CTCCCTTCGG GAAGGGT  3201 CGGTTCAGCC CGACCGC  3301 GATTAGCAGA GCGAGGT.  3401 CTGAAGCCAG TACCTT  3501 CGCGCAGAAA AAAAGGA  3601 ATTATCAAAA AGGATT  3701 TGCTTAATCA GTGAGGC.  3801 TACCATCTGG CCCCAGTC	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAA TCTTTATTT GTGAAAA TCT GGCGTAATCA TGGTCAI GGT GCCTAATGA TGGCCAI AC GCCGCGGGGA AGCCGGT TCA CTCAAAGGCG GTAATAC GCC GCGTTCCTGG CGTTTT AACAGCGCGTT TCCCCCT AGC GCCTTATCCG GTAACTA TGC GCCTTATCCG GTAACTA TGC GCCTTATCCG GTAACTA TGC GCCTTATCCC GTACCTA TGC GCAAAAAGAGTT GCTACGC AAAAAGAGTT GCTACGTAC TCC CAAGAAGATC CTTTGAI TCA CCTAGATCCT TTTAAAI CC TACTCAGCG ATCTGCT GCAATGATACC CCCGGAGA	TAACTAGCAT GAAGGTAAG TCG ACTGTGCCTT TTG TGATGCTATT AGC TGTTTCCTGT ACT CACATTAATI TTG GGATTACT CACATTAATI TTG GGTATTGGG GGAT TATCCACAG GGAT TATCCACAG TAGGCTCCCTCC TCG TCTGAGTACT TCG TCTTGAGTCC TCT TTGACTGGC TTT TTCTACGGGA TAA TAATGAGTT TAT TTCGTTCACC CCC ACGCTCACCC TCC ACGCTCACC TCC ACGCTCACCC TCC ACGCTCACC TCC ACCCTCACC TCC ACCCTCAC T	CCTATCCCTA CTAAGATCCA CCTAAGATCCA CCTAAGATCCA CCTTATTTG AAAGCAAGTA CGGTTGCGCT ATCAGGGGAT CCCCTGACG TCCGCTCTCCGC AACCCGCTACA AACCCGGTAA GGCCTACCTA AACCACCAC TCTGACGCT TTAAATCAAT CATAGTTGCC CCTCCAGATT CATAGTTGCC GCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCTCTAC CACTGCCCC CACTGCCCGC TTCCTCGCTC ACCGGCAAAGCATCACAA AGCATCACAA AGCATCACAA TGTTCCGACC GACTACACC GCTACACT GCTGCTACCG AGTGGAACGA CTAAAGTATA TGACTCCCCC TATCACCATT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATTGAGTA CAATTCCACA TTTCCAGTGG ACTGACTGG ACTGACTGG CTGCCGCTTA AGACAGCC ATCGCCACT ACCCACT ACAGCT ATCGCCACT AGAAGGACAG TTCGCCACT AGAAGACAGC ATCGCCACT AGAAGACAGC ATCGCCACT AGAAGACAGC ATCGCTAAGCT ATCAGCTAAACT AAACCACGC	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT TGCGCTCGGT TGCAGTCAGA CCGAATACCT GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC TGTGTCCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGCG	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAAGG GTGCCAAAA GTCCGCCTTT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGAATAA TGGTCATGAG CAGCAGAATTA TGGTCATGAG CAGGAGGGCT CCGAGCGCAG CCGAGCGCAG
2001 GGATACTTGA GTTAGTT  D T  2011 AGCACAGTGG CGGCCGC 2201 TCACCATCAC CATTGAG 2301 CTAGAATGCA GTGAAAA. 2401 TTGCATTCAT TTTATGT 2501 ACCTGCAGGC ATGCAAG 2601 TAAACTGTAA AGCCTGG 2701 GCATTAATGA ATCGGCC. 2801 CCGGCAGGCG TATCAGC 2901 CCAGGAACGG TAAAAAGA 3101 CCCCGACAGGA CTATAAAA 3101 CCCGACAGGA CTATAAA 3101 CTCCCTTCGG GAAGCGT 3201 CGGTTCAGCC CGACCGC 3301 GATTAGCAGA GCGAGGT 3401 CTGAAGCAG AAAAGGA 3401 CTGAAGCAG AAAAGGA 3501 TACCTTAACA GTGAGGC 3801 TACCATCTGG CCCCCAGT 3901 AAGTGGTCCT GCAACTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCTTTATTT GTGAAAT TCT GGCGTAATCAG TGAGCTA GCC GCGTAATCAG TGAGCTA CAC GCGGGGGAG CC CCGTTGCTGG CGTTTT SAT ACCAGGCGTT TCCCCCT GGC GCTTTCTCAT AGCTCAG AGC GCTTTCTCAT AGCTCAG AGC AAAAAGAGTT GCTAGCT TCT CAAGAAGTC CTTTGAT TCA CCTAGATCCT TTTTAAT ACC TATCTCAGGG ATCTGT CAC GCAATGATCC TTTTAAAT ACC TATCTCAGGG ATCTGT CCC GCAATGATCC CCCGGGAG TAT CCGCTCCAT CCAGTCT CCT CAAGAATCC CCCGGGAG TAT CCGCCTCCAT CCAGTCT CCCCCCCTCCAT CCAGTCT CCCCCCCCCTCCAT CCAGTCT CCCCCCCCCCTCCAT CCAGTCT CCCCCCCCCCCTCCCAT CCAGTCT CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TAACTAGCAT GAAGGTAAG TGG ACTGTGCTTT TG TGATGCTATT AGC TGTTTCCTGT AGC TGATTAGTT TG GTATTGGG GGA TATCCACAG GGA AGCTCCCTC TGTAGGTCCT TTTTGAGTAGT TTTTTTTTTT	TTTTGCGCGA CCTATCCCTA CTAGGATCA GCTTTATTTG GGTGAAATTA GTGAAATTA GTGAAATTA GTGAAATTG GCGTTGCGC ATCAGGGGAT CCCCCTGACG TCCGCTCTCC CAGTTCGGC AACCCGTAA ACAACCACC TCTGACGT TCTGACGTC TCTGACGTC TCTGACGTC TCTGACGTC TTAAATCAAT CATAGTTGCC GCTCCAGATT	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTTAT TATCCGCTC CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGTTC GACACGACTT GACACGACTT GCTGTACACT GCTGTACGC CTAAAGTATAT TGACTCCCGC TATCACACT TATCAGCATT TGACTCCCGC TATAAGTATAT TGACTACTAGCAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGACATTTCCAACG TTGCCGAGGTTA GCTCCAAGGT ATCGCCACTG AGAAGGACA TCGGTTATTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACTCAGCT ATGCAGTAAA	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTAGCAACAGTA AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGCGGTATACT GGCGCTGTGG CCAGCACCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT AACTACGATA ACTACGATA CCCGGAAGGG ATAGTTTGCA	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACA TGATCAGTG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCAACA GTCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA CGGTCATACCAA CGGGAGGGCT CAGCTCATACCAA CGGGAGGGCT CCGAGCGCAG CAACGTTGTT
2001 GGATACTTGA GTTAGTT  D T  2001 AGACCAGGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTATAGT  2501 ACCTGCAGGC ATGCAGG  2601 TAAAGTGTAA AGCCTGG  2701 GGATTAATGA ATCGGCC.  2801 CGGCGAGGGG TATCAGC  2901 CCAGGAACCG TAAAAAG  3001 CCCGCACAGGA CTATAAAA  3101 CTCCCTTCGG GAAGCGT  3201 GATTAGCAG CGGACGGC  3401 GTATAGCAGA GCGAGGT  3401 CTGAAGCCAG TTACCTT  3501 GCGCAGAAAA AAAAGGA  3601 ATTATCAAAA AGGATCT  3701 TGCTTTAATCA GTGAGGC  3801 TACCATCGG CCCCAGTT  3901 AAGTGGTCCT GCAACTT  4001 GCCATTGCTA CAGGCAT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCT AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT ACC GCGCGGGGA GCGGGT TCA CTCAAAGGCG GTAATAC GCC GCGTTTCTTG AGC GCTTTCTCAT AGCTCAC TCG GCTTTTCAT AGCTCAC TCG AAAAAGAGTT GGTAGCT TCT CAAGAAGATC CTTTGAT ACC CTAGATCCT TTTAAAT ACC CTAGATCAT TTTAAAT ACC GCATTATCCG AACTCAC TCG CTAGATCAT TTTAAAT ACC TATCTCAGCG ATCTGTC TCG CGCATCACTC TCG CGCCTCCAT CCAGTCT TCT GCGCTCCACT CCAGTCT TCT GGTGTCACCC TCCTCGT TTTGTCATCCTCTCT TTTTAAAT TCCC TATCTCAGCG TCCTCTCT TCT GGGTTCACCC TCCTCTCT TCT GGGTTCACCC TCCTCTCT TCT GGTTCACCC TCCTCCT TCTCCTCTCTCTCT TCTCTCTCTCTCT	TAACTAGCAT TO GAAGGTAAG TG ACTGTGCCTT TG TGATGCTATT AGC TGTTTCCTGT AGC TGTTTCCTGT AGC TAGGTTGCG TG TATCCACAG TG TATCCACAG TG TAGCTCCG TG TAGGTTCCT TG TGAGTCC TC TGAAGTGCT TT GATCCGCGA TT TTCTACGGG TAG TT TTCTTCATC TAGTTCAC TT TTCTTCATC TAGTTCAC TAGTTCACCC TG TATTGCTTCACC TG TATTGCTTCACC TG TATTGCTTCACC TTG TATTGCTTCC TTG TATTGCTTC TTG TATTGCTTC TTG TATTGCTTC TTG TATTGCTC TTG TATTGCT TG TATTGCT TG TATTGCT TG TATTGCT TG TATTGCT TG TATTGCT TG TATTGC TTG	CCTATCCCTA CTAAGATCCA CCTAAGATCCA CCTAAGATCCA CCTTATTTG GCGTTGCCGT ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACACCGGTAA ACACACCAC ACACACCAC ACACCACCAC ACACCAC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGCC AACGCAGGAA AGCATCACAA TGTTCCGACC TAGGTCGTT GACACGACT TGGTCGAC TGGTACACAC TGGTACACAC TGGTACACAC TAGACGACT TGGCACC TACTCACAC TAGACGACT TGACTACCCCC TATCACCAAC TAGACTACCCCG TATCACCAAT AGTAAGTATA TGACTCCCCAAC	TTTATCTGAA  CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATC CAATTCCAGTCG AGACATGGA AACTGGCGCTA GACTCGACGTA GACTCGACGTA GCTCCAGCGTTA GCTCCAGCTT AGACCACTG AGAAGACATGT ATCGCACTG AGAAGACAGCT ATCGCACTG AGAAGACAGCT ATCGCACTG AGAAGCAGCA TTTTTT AAACTCACGT TATGAGTAAA TCGTGTAGAT AAACCAGCCA TCGCCAGTTA GATCAAGGCG ATCAAGGCG	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTAC GGAGAACTTA CAACATACGA GGAAAACTGT TGCGCTCGGT AGCAAAACGG TCAAGTCAGA CCGGATACCG GGGCTGTGTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAGGGATTT TGTGTGTGTA TGTGTGTGTG AACTACGATA AACTACGATA GCGGAAGGG ATGGTTACATGA	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGGCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAA GCAGCCTT CACGAACCC CTGGTAACAG CTGGCAGATTA TGGTCATGAG CAGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGGAGGGCT CCGAGCCCATGT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTT  D T  2001 AGCACAGTGG CGGCCGC  2201 TCACCATCAC CATTGAG  2301 CTAGAATGCA GTGAAAA.  2401 TTGCATTCAT TTTATGT  2501 ACCTGCAGCC ATGCAAG  2601 TAAAGTGTAA ACCCTGG  2701 GCATTAATGA ATCCGGCC.  2801 CGGCGAGCGG TATCAAGC  2901 CCAGGAACGG TATAAA  3001 CCCGACAGGA CTATAAA  3101 CTCCCTTCGG GAAGCGT  3201 CCGTTCAGCC CGACCGC  3301 GATTAGCAGA GCGAGGT.  3401 CTGAAGCCAG TTACCTGG  3601 ATTATCAAAA AGGATT  3701 TGCTTAATCA  3601 ATTATCAAAA AGGATT  4001 GCCATTGCTA CAGGCAT  4001 GCCATTGCTA CAGGCAT  4001 GCCATTGCTA CAGGCAT  4001 GCCATTGCTA CAGGCAT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TTT GGCGTAATCA TGGTCAT GGTTCAGGG GGAGGTC TTT GCCTAATGAG TGAGCTA ACC GCGCGGGGGA AGGCGGT TCA CCAAAGGCG GTAATAC GCC GCGTTGCTGG CGTTTTT GGC GCTTTCTCAT AGCTCAC TAGGCGGTGC TACCAGAC TCC CAAGAAGAGTC CTTTGAT TCA CAAGAAGAGTT GCTAGAGAC TCC CAAGAAGATC TTTAAAT ACC TATCTCAGC ATCTTTAAAT ACC TATCTCAGC ATCTGTC TCA CAAGAAGATC CTTTGAT TCA CCTACGATCCT TTTAAAT CCTACGTCACT CCAGTCT TCT GCAATGATAC CCCGGAGA TAT CCGCCTCCAT CCAGTCT TCT GGTGTCACGC TCCTCGTA ACC TATCTTCAGCC TCCTCGTAGCC TCCTTCGGTC CTCCGAT CCTCCGTC CTCCGAT CCTCCGTC CTCCCGAT CCTCCGTC CCCCCGAT CCCCCCCCAT CCCCCCCCCCCCCCCCCCCC	TAACTAGCAT GAAGGTAAG TCG ACTGTGCCTT TTG TGATGCTATT AGC TGTTTCCTGT AGT CACATTAATT TTG CGTATTGGGC GGT TATCCACAGA TCA TAGGTCCCTC TGAGGTATCT TCT TGAGGTACT TTT TTGAGTACT TTT TTTTACGGGCA TAT TAGGTCATA TAGTTCACAGA TAT TTCTTCACC TAGTTCATT TCTCACAGCA TAT AATTGTTGCC TGG TGTCAGAAGT TTGTCAGCACC TTGTCAGCAC	CCTATCCCTA CTAGATCCA CCTATACCTA CTAGATCCA CCTTATTTG AAAGCAACTA CTGAAATTGT GGTTGCGCT ATCAGGGAT CCCCCTGACG TGCGCTCTCC CAGTTCGGTG AACCCGCTAACTA ACCACCTACTA ACCACCTACTA CCTCTACCTC CTGACGCTC TCTAACCAC TCTGACGCTC TTAAATCAAT CATAGTTGCC GCTCCAGATT GGGAAGCTAG ATTCAGCTCC AAGTTGCCCC	TGTGATCTTG  ACCCTCTCCT GACATGATAA TAACCATTAT AAACCATTAT AAACCTTACC CATCCCGCC TTCCTCGCTC TTCCTCGCTC ACCTGGACC TAGGTCGTTC GACACGAATCACAA TGTTCCGACC GCTGCTACACT GCTGCTACACG AGTGGAACGA CTAAAGTATA TGACTCCCCC TATCAGCAAT AGTAAGTATA GGTACCACA AGTAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA AGTAAAGTATA CAGTTCTACAC CAGTGTTATC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGGTA AAATCACACG ATTCCAGTGG AAATCGACTGC AGAACATGTG AAATCGACGC CTGCCGCTTA AGAGGACAGT ATCGCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCGCCACTG AGAAGGACAG ATCACGGT TATGAGTAAA AAACCAGCCA TCGCCAGTTA CAGCCAGTTA AAACCAGCCA ATCAAGGCA ACTCAAGGCA ACTCAAGGCA ACTCAAGGCA ACTCATGGT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACAGTTA TGGCTGATTA CAACATACGA GGAAACCTGT TGCGCTCGGT AGCAAAAGGC TCAAGTCAGA CCGGATACCT GGCTGTTGG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTC CTTGGTCTGA AACTACGATA GCCGGAAGGG ATAGTTTTCGA ATGCAGCAGCA ATGCCAGCA ATGCCAGCA	GCAGATATCC  CCGGTCATCA CAAACCACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGCCTG CAGCAAAAAGG GTGCCAAAAAGG CTGCCATAT CACGAACCCC CTGGTAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA TGGTCATGAG CAGTTACCAA TCGGAGCGCT CCGAGCGCAG CAACGTTGTT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCTTTATTT GTGAAAT TCT AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT GAC GCGGGGGAG GCGTT CA CTCAAAGGCG GTAATAC CC GCGTTGCTGG CGTTTTT SAT ACCAGGCGTT TCCCCCT GGC GCTTTCTCAT AGCTCAC AGC AAAAAGAGTT GCTAGCT TCT CAAGAAGAC CTTTGAAT TCA CCTAGATCC TTTTGAAT TCA CCTAGATCC TTTTGAAT TCA CCTAGATCC TTTTAAAT TCA CCTAGATCC TTTTAAAT TCC GCATGATCC CCCGAGG TATCCGGT CCAGCTC TCGT CCACTCCCAT CCAGCTC TCGT GGTGTCACC TCCCGAT TCCT CAGATCCT TCCCGTC TCGT GGTGTCACC TCCCGAT TCCT CCGCTCCCAT CCAGCTC TCCT CCGTCCGTC TCCCTCCGTT CCCCGTT TCCTCCGTT TCCTCCGTT TTCTTTTTTTTT	TAACTAGCAT GAAGGTAAG TGA ACTGTGCCTT TG TGATGCTATT AGC TGTTTCCTGT AGC TGATTAGT TG GTATTGGG GGA TATCCACAG GGA AGCTCCCTC TG TAGGCTCCT TG TAGGTACT TT TCTACAGGC TT TTTGAGTGC TT TCTACAGGC TAAATGAAGT TACT TTCTACGGCAT TATTGTCACAGG TAAATGAAGT TACT TTCTACGGGCAT TATTGTCACAGG TAAATGAAGT TACT TTCTACGGCAT TATTGTCACC TGT TGTAGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCCT TGTTGAGTCGCT TGTTGAGTGCT TGTTGAGAGT TGTCAGAAGT	CTATCCCGA CCTATCCCTA CTAGACTCA CCTAGACCA CCTTATTTG CGGTTGCGCT CCCCTGACG TCCCCCTGACG TCCCCCTGACG TCCGCTACCA ACCCGGTAA ACCCGGTAA ACAACCACC TCTGACGCT TCTGACGCT TCTGACGT CGTCTCCCCAGATT TAAATCAAT CGATACTACAACT CGGTAGCTCC AGCTCCAGATT GGGAAGCTAG ATTCAGCTCC AACTAGCTCC CAACCAACTC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCCACC TAGGTCGACC TAGGTCGTC GACACGACTT GACACGACTT GACACGACTT TGACACACT GCTGGTACGC CTAAAGTATA TGACTCCCGC TATCAGCAAC TATCAGCAAC CAGTTATC CAGTTACC CAGTTATC CAGTTATC CAGTTATC CAGTTATC CAGTTTATC CAGTTTATC CAGTTATC CAGTTTATC CATTCTAGAAA	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACCCCACTG AGAAGGACA TGGGTTTTT AAACTCACGT TATGAGTAAA TCGGTAGATAA TCGTGTAGAT AACTCAGCT ATCGCCAGTTA AACTCAGCT ATCAGCTAGATTA AACTCAGCT ATCAGAGCAGTAAA TCGCCAGTTA CATCAAGCGC ACTCAAGCT TAGAGTAATATC	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTAGCAACAGTA AGCAAAAAGGC TCAAGTCAGA CCGGATAACT GGCGGATACCT GGCGTGTGG CCAGCAACA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT AACTACGATA GCCGGAAGGG ACCGAGCAA AATTACGATA GCCGGAAGGG AATACTTTGCA AATTACGATA AATTACGATA AACTACGATA AACTACGACACAC GGCGACCCAC	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACA TGATCAGTG GCCGGAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGA GCAGAAAAGA GCAGCACAC CTGGTAACAG CTGCGCTCT CAGCACACA CAGGAGATTA CGGTCATACCAA CGGGAGGGCT CCGAGCCCAA CGGAGCACAA CGGAGCACAA CGGAGCACAA CGGAGCACAA CGGACCCAA CCAACATTGTT CCCCCCATGT TGCATAATTC TTGCTCTTGC
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCT AGGTTCAGGG GGAGGTG TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT GGCGGAGTG TCA CTCAAAGGCG GTAATAC GCC GCGTTGCTGG CGTTTTT GGC GCTTTCTCAT AGCTCAC GCC GCTTTCTCAT AGCTCAC TCA CAGAGGGTT TCCCCCT GGC GAAAAGAGATT TCTAAAAT ACC GAATCAT TTTAAAAT ACC TATCTCAGCG ATCTGTC GCT GCAGTCACCTCCGTCT TTTAAAT ACC TATCTCAGCG TCCGCTCAT TCT GCAGTCTCCTCAGTCT TCT GAGATGCTT TCT GAGATGCTT TCT GAGATGCTT TCT GCAGTCT TCT GCAGTCT TCT GCAGTCT TCT GCAGTCT TCT GCAGTCT TCT GGTGTCACCC TCCCGCTCCCAT TCCTCGGTC TCCTCGTT TCCT TAAAATGCT TAAAATGCT TAAAATGCT TAAAATGCT TCTTCGTC TAAAATGCT TCTTCGTC TCCCCAT TCTTCGTC TCAAAGCT TCTTCGTC TCAAACCC TCTTCTTCGTC TCAAACCC TCTTCTTCTTC TCAACCC TCTTCTTCTTC TCTTCT	TAACTAGCAT TO GAAGGTAAG TO ACTGTGCCTT TO TGATGCTATT AGC TGTTTCCTGT AGC TGTTTCCTGT AGC TACATTAGT TG CGTATTGGG GGA AGGTCCCTC TG TAGGTCCTC TG TAGGTACT TTTAAGGGGA TTT TCTAAGGGGA TTT TCTAAGGGGA TTT TCTTCATC TGA TGTTCATC TGA TGTTGATCC TGT TGATGCC TTT TTTTAAGGGGA TTT TTCTTCATC TGT TGTTCATC TGT TGTTGATTGC TGT TGTTGAGTCC TGT TGTTGAGTGCT TGTTGAGTGCT TGTTGAGTGCT TGTTGAGTGCT TGTTGAGTGCT TGTTGAGTGCT TGTTGAGAGAT TTTTTAGGGGA TGTGGTGTAGTACT TGTTAAAAGT TGTTAAAAGT TGTTAAAAGT TGTTAAAAGT	CCTATCCCGA CCTATCCCTA CCTAGACTCA CCTAGACTCA CCTAGACTCA CCTCTACTTG CCCTCGACG ATCAGGGGAT CCCCCTGACG ATCAGGGGAT ACACCACTAACTA ACAAACCACC TCTGACGCTC TTAAATCAAT CATACTTGCC CGTCTCCAGATT CGGAAGCTAC CATACTACTA ACAAACCACC ATTACAA	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGCC CACTGCCGCC AACGCAGAA AGCATCACAA TGTTCCGACC TAGGTCGTC GCACGACC TAGGTCGTC GCACGACC TAGGTCGTACACT TGACTACCAC CTAGGAACGA CTAAAGTATA TGACTCCCCG TATCAGCAAT AGTAAGTATA GGTTCCCAAC CAGTGTTATC ATTCTGAGAA GGAAAACGTT	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATCTGATCA CAATTCCAGTCG AGACATGCA AAATCGAGCT AAATCGAGCT AAATCGACGC AGACATTG AAATCGACGC AGACATTT GCTCCAAGCT ATCGCACTG AGAAGACATGT AAACCAGCAA TAGCCACTG AGAACAAGCA AACCAGCAA AACCAGCAA AACCAGCAA CAGCAGTA AAACCAGCCA CTCAGGGTT AGATCAAGGCT ACTCATGGTT AGATCAAGGCT ACTCATGGTT TAGTGATATC CTTCGGGGCG	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTAC GGAAAACCTGT TGCGCTCGGT AGCAAAACGGC AGCAAAACGGC TCAAGTCAGA CCGGATACCGG GGCTGTGTG GCACCACCAC TATTTGGTAT TGTTTGCAAG TAAGGGATTT TGTGTGTAT TGTGTGTAT TGTGTGTAT AGCGGAAGGG AAACTCACA AGGAATTACGAA AGCAGCAAC AGGAACGAC AGGAACGAC AGGCAACCAC AGGCAACCAC AGGCAACCAC AAAACTTCTCA	GCAGATATCC  CCGGTCATCA CAAACCACA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGGCAGCT CGTTCGGCTG CAGCAAAAGG GCTGGCAAA GCAGCCTTT CACGAACCC CTGGTAACAG CTGGCAGATTA TGGTCATGACAG CAGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CGGAGCGCAG CAACGTTGT TCCCCCATGT TGCATAATTC TTGCTCTTGC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC AGGATCTTAC
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TTC AGGTTCAGGG GGAGGTG TTT GGCGTAATCA TGGTCAT AGC GCCTAATGAG TAGGCCT AGC GCGTTATCAG GTAATA ACC GCGCGGGAG AGGCGGT TCA CTCAAAGGCG TTACCAG AGC GCGTTATCAG GTAACTA ACG CACTTATCAG GTAACTA ACG CACTTATCAG GTAACTA ACG CACTTATCAG GTAACTA ACG CACTTATCAG GTAACTA ACG CACTTATCAGG TCA AAAAAGAGTT GTAACTA ACC TATCTCAGCG ATCTGTA TCA CACAAGAGATC CTTTGAAT ACC CCCCCCAT CCCAGTA ACG TCCTTCAGGTC TCCGCGTAG ACG TCCTTCAGGTC TCCCGTAG ACG TCCTTCAGGTC TCCCGTAG ACG TCCTTCAGGTC TCCCGTAG ACG TCCTTCAGGTC TCCCGTAG ACG TCCTTCGGTC TCCCGTAG ACC TCCTTCGGTC CTCCGGTC ACG ATGTAACCCA CTCGTCG ACG ATGTAACCCA CTCGTCG ACG ATGTAACCCA CTCGTCG ACC TCCTTCGTCC TCCCGTAG ACC TCCTTCGTCC CTCCGTAG ACC TCCTTCGGTC CTCCGGTC ACG ATGTAACCCA CTCGTCG ACC TCCTTCGTCC CTCCGTCC ACG ATGTAACCCA CTCGTCG ACC ACTAGCAC ACC TCCTTCGTCC CTCCGTCC ACC ACTAGCAC ACC TCCTTCGTCC CTCCGTCC ACC ACTTCAGCC CTCCGTCC ACC ACTTCGTCC CTCCGTCC ACC ACTTCGTCC CTCCGTCC ACC ACC ACC ACC ACC ACC ACC ACC ACC A	TAACTAGCAT CGAAGGTAAG CGA ACTGTGCCTT CGAGGTTTTT CGG GAGGTTTTT CGG CACATTAATT CGC CACATTAATT CGC CACATTAATT CGC CACATTAATT CGC CACATTCAGC CGC CACCACCC CCC CACCACC CCC CACCACCC CCC CACCACCC CCC CACCACCC CCC CACCACCC CCC CACCACC CCC CACCACCC CCC CACCACCC CCC CACCACCC CCC CACCACCC CCC CACCACC CCC CACCACCC CCC CACCACCC CCC CACCACCC CCC CCC CCC CCC CCC CCC CCC	TTTTGCGCGA  CCTATCCCTA CTAAGATCCA GCTATATTG AAAGCAGTA GGAAATTGT GGGTTTCCGC ATCAGGGGAT CCCCCTGACG TCGCCTTCCGC AACCGGTAA GGCTACTA ACAAACCACC TTTAATCATT CGAAACCACC TTTAATCATT CGAAACCACC ATCAGGGTA ACAGGCTA ACAGGCTA ACAGGCTA ACAGGCTA ACAGGCTA ACAGGCTA ACAGGCT AACCAGGTA ACAGGCT AACCAGGTA ACAGGCT AACCAGGTA ACAGGCT AACCAGGT AACCACGT ACAGATT CGGAAGCT AGCACAGATT CAGCACT CAACCAACT CCCCC CAACCAACT CCCCCC CAACCAAC	ACCCTCTCCT GACATGATAA TAACCATTAT TAACCATTAT TATCCGCTCA CACTGCCGC TCCTCGCTCA ACCCAGGAA AGCATCACAA TGTTCCCACC TAGGTCGTC GACAGGAC TAGGTCGTC GACAGGACT CGGCTACACT AGTTCCACC GTTGATACG CTAAAGTATA TGACTCCCAG TATCAGCAAT AGTAAGTATA AGTAAGTATA AGTAAGTATA TGTTCCCAAC CAGTGTTATC CAGTTTTTTACTTTCACT	CGGTCTCGAT GATACATTGA AAGCTGCAT AAATGTGAT CAATTCCACA TTTCCAGTCG AAATCGACC CTGCCGCTTA GCTCCAACCT ACGACAC ACGACCAC ACGACTTATCACAC TTGCAGCC TATGAGCAC TATGAGTAT AACCACAC TATGAGTAT TATGAGTAAA TCGTGTAGAT TATGAGTAAA TCGTGTAGGT TATGAGGTC TATGAGGTC TATGAGGTT TATGAGTATGC TCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CTCCGGGGCC CAGCGTTTCT	GGGCAATTCT TCTACGCGTA TGAGTTTGGA AAACATTGA TGGCTGATTA CAACATACGA GGAAACCTGT TCGCCTCGGT AGCAAAAGCT TCAGGTCAGA CCGGATACCT GGGCTGTGTG GCAGCAGCA TATTTGCAAG TAAGGGATT CTTGGTCTGA ACTACGATA CCCGGAAGGG AACTTTCGCA GGCGAAGGG ATACTTCGA AGTACATCGAC AGTACACTCCA AGGCCAACCAG GGCGACCAG GAAAACTCTCA GGGTGACCA GGGTGACCAG	GCAGATATCC  CCGGTCATCA CAAACACACA ACAACAACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGGCAGACA GGTGCCAGCT CAGCAAAAGG CAGCAAAAGG CTCGCCTTT CACGAACCCC CTGGTAAACAG CTGCGCTCTG CAGCAGATTA TGGTCATGAG CAGTTACCAA CCGGAGGGCT CCAGCGCAG CAACGTTGT TCCCCCATGT TGCCCCATGT TGCATAATTC TTGCTCTTGC AGACAGAG AAACAGGAAG
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCG TTT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCTTTATTT GTGAAAT TCT GGCGTAATCAG TGAGCTA GCC ACTAATGAG TGAGCTA CCC CCTTACTG GTAATCAG GCC GCGTTGCTG CGTTTT SAT ACCAGGCGTT TCCCCCT GCC GCTTTCTCAT AGCTCAG AAAAAGAGTT GCTAACTA ACC TAAGAGAC CTTTGAA TCC ACCAGAGACC TTTTGAA TCC ACAGAGACC TTTTGAA TCC ACAGAGACC TTTTGAA TCC GCATGATCC TTTTGAA TCC GCATGATCC TTTTGAA TCC GCATGATCC TTTTGAA TCC GCATGATCC TCCGGAG TAACAGC TCCTCCGT TCC GAAGAACCT TCCCGCT TCC GCATGATCC CCCGGAG TCC GCATGATCC CCCGGAG TCCT CCATCGGTC TCCT CCGCTCCAT CCAGTC TCCT TCCGTC TCCT TCCGGTC TCCT TCCGGTC TCCT TCCGGTC TCCT TCCGGTC TCCGTCGTC TCCGGTCATCCC TCCGGTC TAAGAACCCT CTCTGGTC TCCGGTCATCCGGTC TAAGAACCCT CTCTGGTC TCCGGTCATCCGGTC TAAGAACCCA CTCGGTC TAAGAACCCA CTCGGTG TCCGGGGGAATAACCCC TCCGGTG TCCGGTC TCCGGT	TAACTAGCAT GAAGGTAAG TGA ACTGTGCCTT TG TGATGCTATT AGC TGTTTCCTGT AGC TGTTTCCTGT AGC TAGGTCCG ACTTAACT TG CGTATTGGG AGATCCCTCC AGATCAGAG AGATCCCTCC TGAAGTGGT TTCTAAGGT TTCTAGGGGA TATCTGGGCAT TATTGTCAGGG TAAATGAAGT TACT TTCTACGGGAT TATTGTCAGT TTCTAGGGG TAAATGAAGT TATTTGTCAGGG TGAGTACC ACTTAGAAGT AATTGTTGAATAC ACGCTAGC ACTTAAAAGT ACTTAAAAGT AAAA TGTTGAATAC AAAA TGTTGAATAC	CTATCCCGA CTAAGATCA CTAAGATCA CTAAGATCA CTAAGATTGT GGGTTGCGCT CCCCTGACG TCCCCTGACG TCCGCTACAC AACCCGTAA AACCACC TCTAACTA ACAAACCAC TCTAACTA ACAAACCAC TCTAACTA ACAAACCAC TCTAACTTCGC CGCTCACAGT TCAGACTTC GGCAACCAACTA CATAGTTGCC CACTACAACT CTTAAATCAAT CATAGTTGCC CACCAACTA TCAGCTCC CAACCAACT TCACACTT TCACACTCT TCATACTTT TCACACTCT TCATACTCTT	ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGACC TAGGTCGTC GACACGACTT GACACACT GCTGTACACT GCTGTACCG CTAAAGTATA TGACTCCCG TATCAGCAC TATCAGCAT TATCATTCACAC CCTTTTTCAC CCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACTCCAAGGT ATCGCCAGTT AAACTCAGGT ATACAGGCA ACAGCAAACATGAAACACACCA ACACAGCAAACACACCA ACACAGCAACACACCACACACA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTTA GCACTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT GCCGGAAGGG ACCCAGCAAAAACTCTCA GGCGACCGAA AAAACTCTCA GGGCGACCGAG AAAACTCTCA GGGTAGCACA GCATTTATCA	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCAACA GTCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCACATA CGGTCATGC CAGCTCATG CAGCACATT CCCACATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTTTTT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTCTTTC AGGATCATTAC AGGAGATTTT TCCCCCATGT TGCATAATTC TTGCTCTTTC AGGATCTTTAC AGAACAGGAAG GGGTTATTTTT
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT GGCGGAGTG CA CTCAAAGGCG GTAATAC GCC GCGTTCTTG GCC GCTTTCTT AGC GCCTTATCCG GTAACTA TGC GCCTTTCTCAT AGCTCAC TCA AGAAAGAGTT GCTAGCT TCA CAAGAAGATC CTTTGAAT ACC GCATCACGC ACAAAGAGCT TCA CAAGAAGATC CTTTGAAT ACC TATCTCAGCG ATCTGTC GCT GCAATCATCC GCGAGCT TCA CCAAGACCCC TCCCGAT CCG TAACTTC TCG TAAGAAGATC TCCTCGTC TCG TAAGATCCT TTTAAAT TCA CTCTCAGCC TCCTCGTC TCCT TCCTCCTCT TCCT TCCTCCCAT CCCGATCT TCCT TCCTCCCAT CCCGATCT TCAAGAAGACC CTCCGCAT TCAAGAAGACC CTCCGCAT TCAAGAAGACC CATCGTCC TAAGATCCT TTCTGTC TCAAGAAGACC CATCGTCC TCCTAAAAGACCC CTCCGCATCT TCAATAACCCCCC CATCGTCC TCAAGCAACCC TTTTGATTT TAGAAAA	TAACTAGCAT TO GAAGGTAAG TOG ACTGTGCCTT TO TGATGCTATT AGC TGTTTCCTGT AGC TGTTTCCTGT AGC TACATTAGT TG CGTATTGGG GGA AGGTCCCTC GCT GTAGGTACT TTGAGTGC TTT TACAGGGA TTT TCTAGGGCA TTT TCTAGGGCA TACAGGAAGT TAT TTCGTTCATC CCC AGCTCACC TG GTAGGTACT TT TTCTTCATC CCC TGTAGGTACT TTTTTCTTCATC CCC AGCTCACC TG GTAGGTACT TTTTTTCTTCATC CCC AGCTCACC TG GTAGGTACT TGTTGAGTACT CCC AGCTTAGAGTACT CCC AGCTTAGAGACT TGTTAGAGACT AGC GTAGGTACT CCC AGCTTAGAACT AGC CAACTGATCA AGC GAGCTGACC TGTAGAACT AGC CAACTGATCA ATTGTTGAATAG ATTGTGAATAG ATTG	CCTATCCCTA CTAAGATCA CTAAGATCA CTAAGATCA CTAAGATCA CTTATTTG GTGTACCTA CTGAAATTGT CCCCTGACG ATCAGGGGAT ACACGGTAA ACACACCA ACAACCAC TCTGACGTC ACATACTA ACAAACCAC TCTGACGTC ATTACTACTA CATACTTGC CGTCCAGATT CGGAAGCTAC ATTCAGCTC AGTTGCCC CGTCACTA CTAACTACTT CAGCTCC AGTTGCCC CGTCACTA CTAACTACT CTAACTACT CTACACTC CGCTCATCATT CAGCACTCT CAACTACTT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCATACTCTT TCATACTCTC GGTTCCCGCC	ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGACC TAGGTCGTC GACACGACTT GACACACT GCTGTACACT GCTGTACCG CTAAAGTATA TGACTCCCG TATCAGCAC TATCAGCAT TATCATTCACAC CCTTTTTCAC CCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACTCCAAGGT ATCGCCAGTT AAACTCAGGT ATACAGGCA ACAGCAAACATGAAACACACCA ACACAGCAAACACACCA ACACAGCAACACACCACACACA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTTA GCACTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT GCCGGAAGGG ACCCAGCAAAAACTCTCA GGCGACCGAA AAAACTCTCA GGGCGACCGAG AAAACTCTCA GGGTAGCACA GCATTTATCA	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCAACA GTCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCACATA CGGTCATGC CAGCTCATG CAGCACATT CCCACATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTTTTT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTCTTTC AGGATCATTAC AGGAGATTTT TCCCCCATGT TGCATAATTC TTGCTCTTTC AGGATCTTTAC AGAACAGGAAG GGGTTATTTTT
2001 GGATACTTGA GTTAGTT	TAT AAAACTTTTA CATAATI TCG AGTCTAGAGG GCCCGCC TT AAACCCGCTG ATCAGCC AAA TGCTTTATTT GTGAAAT TCC AGGTTCAGGG GGAGGTC TT GGCGTAATCA TGGTCAT GGCGTAATCA TGGTCAT GGCGGAGTG CA CTCAAAGGCG GTAATAC GCC GCGTTCTTG GCC GCTTTCTT AGC GCCTTATCCG GTAACTA TGC GCCTTTCTCAT AGCTCAC TCA AGAAAGAGTT GCTAGCT TCA CAAGAAGATC CTTTGAAT ACC GCATCACGC ACAAAGAGCT TCA CAAGAAGATC CTTTGAAT ACC TATCTCAGCG ATCTGTC GCT GCAATCATCC GCGAGCT TCA CCAAGACCCC TCCCGAT CCG TAACTTC TCG TAAGAAGATC TCCTCGTC TCG TAAGATCCT TTTAAAT TCA CTCTCAGCC TCCTCGTC TCCT TCCTCCTCT TCCT TCCTCCCAT CCCGATCT TCCT TCCTCCCAT CCCGATCT TCAAGAAGACC CTCCGCAT TCAAGAAGACC CTCCGCAT TCAAGAAGACC CATCGTCC TAAGATCCT TTCTGTC TCAAGAAGACC CATCGTCC TCCTAAAAGACCC CTCCGCATCT TCAATAACCCCCC CATCGTCC TCAAGCAACCC TTTTGATTT TAGAAAA	TAACTAGCAT TO GAAGGTAAG TOG ACTGTGCCTT TO TGATGCTATT AGC TGTTTCCTGT AGC TGTTTCCTGT AGC TACATTAGT TG CGTATTGGG GGA AGGTCCCTC GCT GTAGGTACT TTGAGTGC TTT TACAGGGA TTT TCTAGGGCA TTT TCTAGGGCA TACAGGAAGT TAT TTCGTTCATC CCC AGCTCACC TG GTAGGTACT TT TTCTTCATC CCC TGTAGGTACT TTTTTCTTCATC CCC AGCTCACC TG GTAGGTACT TTTTTTCTTCATC CCC AGCTCACC TG GTAGGTACT TGTTGAGTACT CCC AGCTTAGAGTACT CCC AGCTTAGAGACT TGTTAGAGACT AGC GTAGGTACT CCC AGCTTAGAACT AGC CAACTGATCA AGC GAGCTGACC TGTAGAACT AGC CAACTGATCA ATTGTTGAATAG ATTGTGAATAG ATTG	CCTATCCCTA CTAAGATCA CTAAGATCA CTAAGATCA CTAAGATCA CTTATTTG GTGTACCTA CTGAAATTGT CCCCTGACG ATCAGGGGAT ACACGGTAA ACACACCA ACAACCAC TCTGACGTC ACATACTA ACAAACCAC TCTGACGTC ATTACTACTA CATACTTGC CGTCCAGATT CGGAAGCTAC ATTCAGCTC AGTTGCCC CGTCACTA CTAACTACTT CAGCTCC AGTTGCCC CGTCACTA CTAACTACT CTAACTTCC CGTCACATT CGGAACTAC CCACCAACTC CCACCACACTC CCACCACACTC CCACCACACTC CGCTCATCATT TCAGCACTCT TCAGCATCTT TCAGCATCTT TCAGCATCTT TCATACTCTT TCATACTCTT TCATACTCTT TCATACTCTC GGTTCCCCCC	ACCCTCTCCT GACATGATAA TAACCATTAT AAACCCTTAT TATCCGCTCA CACTGCCGC TTCCTCGCTC AACGCAGGAA AGCATCACAA AGCATCACAC TGTTCCGACC TAGGTCGACC TAGGTCGTC GACACGACTT GACACACT GCTGTACACT GCTGTACCG CTAAAGTATA TGACTCCCG TATCAGCAC TATCAGCAT TATCATTCACAC CCTTTTTCAC CCTTTTTCAC	CGGTCTCGAT GATACATTGA AAGCTGCAAT AAATGTGATA CAATTCCACA TTTCCAGTCG AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC AGAACATGGC ATCGCCAGTTA ACTCCAAGGT ATCGCCAGTT AAACTCAGGT ATACAGGCA ACAGCAAACATGAAACACACCA ACACAGCAAACACACCA ACACAGCAACACACCACACACA	GGGCAATTCT  TCTACGCGTA TGAGTTTGGA AAACAGTTA CAACATTA CAACATTA CAACATACGA GGAAAACCTGT TGCGCTGGTTA GCACTCGGT AGCAAAAAGGC TCAAGTCAGA CCGGATACCT GGCCTGTTG GCAGCAGCCA TATTTGGTAT TGTTTGCAAG TAAGGGATTT CTTTGGTAT GCCGGAAGGG ACCCAGCAAAAACTCTCA GGCGACCCAA AAAACTCTCA GGGCGACCGAG AAAACTCTCA GGGTAGCACA GCATTTATCA	GCAGATATCC  CCGGTCATCA CAAACACAA ACAACAACAA TGATCAGTCG GCCGGAAGCA CGTGCCAGCT CGTTCGGCTG CAGCAAAAGG GTGGCAACA GTCGCCTTT CACGAACCC CTGGTAACAG CTGCGCTCTG CAGCACATA CGGTCATGC CAGCTCATG CAGCACATT CCCACATGT TCCCCCATGT TCCCCCATGT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTTTTT TCCCCCATGT TTGCATCATC AGGATCATTAC AGGAGTCTTTC AGGATCATTAC AGGAGATTTT TCCCCCATGT TGCATAATTC TTGCTCTTTC AGGATCTTTAC AGAACAGGAAG GGGTTATTTTT

# **Appendix B: List of Stably Transfected S2 Cell Lines**

For establishment of all cell lines listed below, expression plasmids were cotransfected into S2 cells with pCoBlast for blasticidin selection according to conditions recommended by Invitrogen. For cell lines used in Chapter 1, those expressing tagged DmSNAP50 constructs were prepared by Shu-Chi Chiang, and those expressing tagged DmSNAP190 constructs were prepared by Mitchell Titus.

#### Cell lines used in Chapter 1 for truncational analysis:

#### **DmSNAP43** truncations DmSNAP43 C-terminal truncate (N-terminal tag) 302 HisFlag-DmSNAP43 full DmSNAP50 full DmSNAP190 full 303 HisFlag-DmSNAP43 (2-274) DmSNAP50 full DmSNAP190 full 304 HisFlag-DmSNAP43 (2-172) DmSNAP190 full DmSNAP50 full 306 HisFlag-DmSNAP43 (2-125) DmSNAP50 full DmSNAP190 full DmSNAP43 N-terminal truncate (C-terminal tag) 101 DmSNAP43 full-FlagMycHis DmSNAP50 full DmSNAP190 full 307 DmSNAP43 (68-363)-FlagMycHis DmSNAP50 full DmSNAP190 full **DmSNAP50** truncations DmSNAP50 C-terminal truncate (N-terminal tag) 412 HisFlag-DmSNAP50 full DmSNAP43 full DmSNAP190 full 410 HisFlag-DmSNAP50 (2-291) DmSNAP43 full DmSNAP190 full 411 HisFlag-DmSNAP50 (2-109) DmSNAP43 full DmSNAP190 full DmSNAP50 N-terminal truncate (C-terminal tag) 401/623 DmSNAP50 full-FlagMycHis DmSNAP43 full DmSNAP190 full 402 DmSNAP50 (10-376)-FlagMycHis DmSNAP43 full DmSNAP190 full 403 DmSNAP50 (26-376)-FlagMycHis DmSNAP43 full DmSNAP190 full 407 DmSNAP50 (92-376)-FlagMycHis DmSNAP43 full DmSNAP190 full 405 DmSNAP50 (196-376)-FlagMycHis DmSNAP43 full DmSNAP190 full 406 DmSNAP50 (292-376)-FlagMycHis DmSNAP43 full DmSNAP190 full

#### **DmSNAP190** truncations

DmSNA	P190 C-terminal truncate (N-terminal tag)		
505	HisFlag-DmSNAP190 full	DmSNAP43 full	DmSNAP50 full
506	HisFlag-DmSNAP190 (2-623)	DmSNAP43 full	DmSNAP50 full
508	HisFlag-DmSNAP190 (2-451)	DmSNAP43 full	DmSNAP50 full
DmSNA	P190 N-terminal truncate (C-terminal tag)		
504/626	DmSNAP190 full-FlagMycHis	DmSNAP43 full	DmSNAP50 full
501	DmSNAP190 (63-721)-FlagMycHis	DmSNAP43 full	DmSNAP50 full
502	DmSNAP190 (176-721)-FlagMycHis	DmSNAP43 full	DmSNAP50 full

DmSNAP190 N &C-terminal truncate (C-terminal tag)

509 DmSNAP190 (176-451)-FlagMycHis DmSNAP43 full DmSNAP50 full

## Cell lines used in Chapter 3 for alanine scanning:

DmSNAP43 alanine scanning mutants (N-terminal tag)

317	HisFlag-DmSNAP43-mut1	DmSNAP50	DmSNAP190
318	HisFlag-DmSNAP43-mut2	DmSNAP50	DmSNAP190
319	HisFlag-DmSNAP43-mut3	DmSNAP50	DmSNAP190
325	HisFlag-DmSNAP43-mut4	DmSNAP50	DmSNAP190
326	HisFlag-DmSNAP43-mut5	DmSNAP50	DmSNAP190
327	HisFlag-DmSNAP43-mut6	DmSNAP50	DmSNAP190
328	HisFlag-DmSNAP43-mut7	DmSNAP50	DmSNAP190
329	HisFlag-DmSNAP43-mut8	DmSNAP50	DmSNAP190
330	HisFlag-DmSNAP43-mut9	DmSNAP50	DmSNAP190
337	HisFlag-DmSNAP43-mut10	DmSNAP50	DmSNAP190
338	HisFlag-DmSNAP43-mut11	DmSNAP50	DmSNAP190
339	HisFlag-DmSNAP43-mut12	DmSNAP50	DmSNAP190
340	HisFlag-DmSNAP43-mut13	DmSNAP50	DmSNAP190
341	HisFlag-DmSNAP43-mut14	DmSNAP50	DmSNAP190

# Appendix C: High-Salt FLAG Purification of FLAG-tagged DmSNAPc from S2 cells

#### I. Solutions and materials

#### 5 mg/ml 3x FLAG peptide solution (SIGMA, product code F 4799)

The 3x FLAG peptide (N-Met-Asp-Tyr-Lys-Asp-His-Asp-Gly-Asp-Tyr-Lys-Asp-His-Asp-Ile-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-C) is acidic. In order to dissolve it properly, add 160  $\mu$ l of 10x Wash Buffer to 4 mg of 3x FLAG peptide.

After the peptide is completely dissolved, add 640  $\mu$ l of distilled water to the sample. Mix well and store aliquots of 75  $\mu$ l at -20°C.

#### 10x Wash Buffer

0.5M Tris HCl, pH7.4, with 1.5M NaCl

#### 1x Wash Buffer

Add 2 ml of 10x Wash Buffer to 18 ml sterile distilled water and mix well.

#### 1x Wash Buffer with 351 mM NaCl

Add 11.7 ml 10x wash buffer (with 1.5 M NaCl conc.) in a 50 ml Falcon tube. Fill to 50 ml marker with sterile distilled water to get a final concentration of 351 mM NaCl. Store at 4°C.

#### **HEMG Wash Buffer**

81 mM KCl, 32.5 mM HEPES K+ pH7.6, 5.5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 5.0 mM DTT, 10% Glycerol

Make 50 ml HEMG wash buffer by adding 4050 μl 1M KCl, 1625 μl 1M HEPES K+ pH 7.6, 275 μl 1M MgCl<sub>2</sub>, 10 μl 0.5M EDTA pH 8.0, 250 μl 1M DTT and 5ml Glycerol to 20 ml sterile distilled water in a 50 ml Falcon tube. Fill to 50 ml marker with sterile distilled water. Store at 4°C.

**Note:** To make 1M HEPES K+ pH 7.6, use 10 N KOH to titrate 100 ml 1M HEPES to pH 7.6.

#### **Elution Buffer (for column preparation only)**

0.1M Glycine, pH3.5

CelLytic M Lysis Buffer (SIGMA, product code C 2978)

ANTI-FLAG M2-Agarose Affinity Gel (SIGMA, product code A 2220)

Protease Inhibitor Cocktail (SIGMA, product code P 8340)

#### II. Expression of *Dm*SNAPs in S2 cells

- 1. Grow 8 plates (Corning 100 x 20 mm tissue culture plates) of cells in selective medium to 70-80% confluency.
- 2. Induce cells with copper sulfate (add to a final concentration of 0.5 mM).
  - a. Prespare 0.1 M sterile CuSO<sub>4</sub> (add 0.2497 g of CuSO<sub>4</sub> 5H<sub>2</sub>O into 9.91 ml sterile water and mix well, and then use the 0.2 μm pore size syringe filter and syringe to make the copper sulfate solution sterile).
  - b. Add 50 μl CuSO<sub>4</sub> into 10 ml cells. Swirl the plate to mix it well.
- 3. Incubate cells for ~24 hours at 22-25°C.

#### III. Preparation of FLAG resin

- 1. Thoroughly suspend the ANTI-FLAG M2 affinity gel in the vial in order to make a uniform suspension of the resin. Ratio of volume of suspension to packed gel is 2 to 1.
- 2. Using a P1000 with  $\sim$ 2 mm cut-off end of the tip. Immediately transfer 320  $\mu$ l of suspended resin (160  $\mu$ l packed resin) into a new pre-cooled 1.5 ml tube.
- 3. Centrifuge the resin for 4 min at 2000 g.
- 4. Remove the supernatant. Be careful not to remove any beads.
- 5. Wash the beads twice in 1x Wash Buffer.
  - a. Resuspend the beads in 0.5 ml of 1x Wash Buffer.
  - b. Centrifuge resin 30 seconds at 2000 g.
  - c. Remove the supernatant.
  - d. Repeat steps a-c.
- 6. Wash the resin twice with Elution Buffer.
  - a. Resuspend the resin in 0.5 ml Elution Buffer.
  - b. Centrifuge at 2000 g for 30 seconds.
  - c. Immediately remove the supernatant. Do not leave the resin in Elution Buffer for more than 2 min.
  - d. Repeat steps a-c.
- 7. Wash the beads four times in 0.5 ml of 1x Wash Buffer each wash.
  - a. a. Resuspend the beads in 0.5 ml of 1x Wash Buffer.
  - b. Centrifuge resin 30 seconds at 2000 g.
  - c. Remove the supernatant.
  - d. Repeat steps a-c 3 times.
  - e. Leave the resin suspended in 1x Wash Buffer.

# IV. Procedures of FLAG-tagged protein purification

#### A. Cell lysis

- 1. Add 70 μl protease inhibitor cocktail to 7 ml CelLytic M lysis buffer.
- 2. Wash cells (for each plate):
  - a. To remove cells adhering to the dish, pipet the medium over the cells gently several times.
  - b. Collect the cells and medium into a 15 ml centrifuge tube.
  - c. Spin for 5 min in the RT-6000 at 1600 RPM (420 g).
  - d. Decant the supernatant and discard.
  - e. Wash the cells by resuspending the pellet in 10 ml of PBS (Phosphate Buffered Saline). Centrifuge for 5 min at 1600 RPM in the RT-6000.

- f. Decant the supernatant and discard, then pipet liquid off thoroughly. Touch the mouth of the tube with a Kimwipe.
- g. Resuspend the cell pellet in 840 µl of CelLytic M lysis buffer.
- h. Remove cells and buffer to a 1.5 ml conical screw-cap Eppendorf tube.
- 3. Incubate the cells end over end for 15 min in the cold room.
- 4. Centrifuge the cell lysate for 10 min at 12,000 g (11,400 RPM) in the microfuge in the cold room.
- 5. Pool the supernatant from the 8 tubes (~7 ml total) into a new chilled 15 ml tube. Place the tube on ice.
- 6. If the lysate is viscous, shear the DNA by passing it through an 18-gauge needle four times.
  - a. Put an 18-gauge needle on a 10 ml syringe.
  - b. Suck the lysate into the syringe and expel it back into the tube slowly four times
- 7. Remove 200  $\mu$ l of the lysate into a screw-cap tube and store frozen in the liquid  $N_2$ . Keep the remainder on ice.
- 8. Measure the total amount of the lysate remaining. Add 4 M NaCl into lysate to give a final concentration of 350 mM.
  - a. add 92  $\mu$ l of 4M NaCl per ml. of lysate (giving a final concentration of 350mM NaCl.)
  - b. For example, if the volume of lysate is 7 ml, add 643.8 µl of 4M NaCl.

#### **B.** Beads binding

- 1. Add the washed resin to the cell extract ( $\sim$ 7 ml).
  - a. Remove the 1x wash buffer from the resin beads.
  - b. Resuspend the beads very gently in 500 μl of the lysate.
  - c. Transfer the resuspended beads to the 15 ml tube containing the lysate.
  - d. Rinse the 1.5 tube (which was containing the beads) with 500 μl of lysate and transfer that to the 15 ml tube containing lysate.
- 2. Incubate the lysate and beads O/N in cold room on rocker or with end over end rotation.

**Note:** The incubation time could be shorter, depends on what cell lines used.

- 3. Centrifuge the resin for 2min at 1600 RPM in the RT-6000. Remove the supernatant and save it in a 15 ml tube and store in -80°C. Also, remove 200 μl of the supernatant into a screw-cap tube and store frozen in the liquid nitrogen as "Flowthrough".
- 4. Wash the resin TWICE with wash buffer with 350 mM NaCl concentration.
  - a. Add 1 ml wash buffer into the tube containing the resin and then resuspend the beads.
  - b. Remove beads and wash buffer to a new 1.5 ml tube.
  - c. Centrifuge for 30 seconds at 2000 g.
  - d. Repeat steps a-c one more time.

- 5. Wash the resin THREE times with 1.0 ml HEMG wash buffer.
  - a. Add 1 ml 1x wash buffer into the tube containing the resin and then resuspend the beads.
  - b. Remove beads and wash buffer to a new 1.5 ml tube.
  - c. Centrifuge for 30 seconds at 2000 g.
  - d. Repeat steps a-c two more times.

#### C. Elution of the FLAG-fusion protein with 3x FLAG peptide

- 1. Add 50 μl of 3x FLAG peptide (5 mg/ml) to 1200 μl of HEMG wash buffer so that the final concentration of FLAG peptide would be 200 μg/ml.
- 2. Elute the bound FLAG-fusion protein with five 230μl volumes of 3x FLAG peptide (200 μg/ml).
  - a. Add 230 ul of the 3x FLAG elution buffer to the resin.
  - b. Resuspend the resin and let the sample sit on ice in the cold room for 4-5 min.
  - c. Centrifuge at 2000 g for 30 seconds.
  - d. Remove and save the supernatant into a chilled eppendorf tube. This is elution fraction 1.
  - e. Repeat steps a-d to collect fractions 2, 3, and 4.
  - f. Collect elution fraction 5 by adding 230 μl of the FLAG elution buffer and incubating for 10 min at room temperature.
  - g. Keep each fraction separate.
  - h. Remove 50  $\mu$ l from each fraction and store each as a separate aliquot in liquid nitrogen.
  - i. The elution fractions will be assayed by Immunoblotlot, Bandshifts, and transcription assays.

#### D. Recycle and store the resin immediately

- 1. Recycle the resin.
  - a. Add 500 μl of elution buffer (0.1M glycine, pH3.5) to the resin and centrifuge at 2000 g for 30 seconds.
  - b. Repeat previous step two more times.
  - c. Immediately wash the resin with 1 ml of 1x wash buffer and remove the supernatant.
  - d. Repeat step c four more times.
- 2. Storing the resin.
  - a. Add 1 ml of 1x wash buffer containing 50% glycerol and 0.02% sodium azide.

Store the resin at 4°C

# Appendix D. Purification of His-tagged DmSNAPc from S2 cells by Nickel-Chelate Chromatography

#### I. Solutions and materials

#### O.1M CuSO4

Add 0.25 g CuSO<sub>4</sub>· $5H_2$ O into 9.91 ml sterile d.d. water in a 15 ml Falcon tube. Votex to dissolve. Working in the cell culture hood, use 0.22  $\mu$ m pore-size syringe filter to filter/aliquate CuSO<sub>4</sub> solution into 10 of 1.5ml conical screw-cap tubes ( $\sim$ 1ml/tube). Store at 4°C.

#### Stock solution A (for making 5X Native purification buffer)

250 mM NaH<sub>2</sub>PO<sub>4</sub> (monobasic sodium phosphate), 2.5 M NaCl

Dissolve 7.8 g NaH<sub>2</sub>PO<sub>4</sub>·2H2O (MW 155.99) and 29.2 g NaCl in 200 ml d.d. water. Filter to sterilize. Store at 4°C.

#### Stock solution B (for making 5X Native purification buffer)

250 mM Na<sub>2</sub>HPO<sub>4</sub> (dibasic sodium phosphate), 2.5 M NaCl

Dissolve 7.1 g Na<sub>2</sub>HPO<sub>4</sub> (MW 141.96) and 29.2 g NaCl in 200 ml d.d. water. Filter to sterilize. Store at 4°C.

#### 3M Imidazole

Add 10.2 g imidazole into 30 ml sterile d.d. water in a 50 ml Falcon tube. Votex to dissolve. Fill with sterile d.d. water to 50 ml. Invert to mix. Store at 4°C.

#### **5X Native purification buffer**

Add 45 ml stock solution B in a 100 ml beaker with a stirring stir bar. Titrate with stock solution A (drop by drop, only very small amount is needed) until the pH reaches 8.0. Transfer the solution to a 50 ml Falcon tube and store at 4°C.

# 1X Native purification buffer (for making Native binding buffer and Native wash buffer) (100 ml/purification)

Add 20 ml 5X Native purification buffer and 75 ml sterile d.d. water in a 150 ml beaker with a stirring stir bar. Titrate with HCl or NaOH until the pH reaches 8.0 ( $\sim$ 11  $\mu$ l of 12M HCl). Bring the volume to 100 ml with sterile d.d. water. Store at 4°C.

#### Native binding buffer w/ 10 mM Imidazole (12 ml/purification)

Add 30 ml 1X Native purification buffer and 100  $\mu$ l 3M Imidazole in a 100 ml beaker with a stirring stir bar. Titrate with HCl or NaOH until the pH reaches 8.0 (~3  $\mu$ l of 12M HCl). Transfer the solution to a 50 ml Falcon tube and store at 4°C.

#### Native wash buffer w/ 20 mM Imidazole (12 ml/purification)

Add 50 ml 1X Native purification buffer and 335  $\mu$ l 3M Imidazole in a 100 ml beaker with a stirring stir bar. Titrate with HCl or NaOH until the pH reaches 8.0 (~15  $\mu$ l of 12M HCl). Transfer the solution to a 50 ml Falcon tube and store at 4°C.

#### **Chemicals for preparation of HEMG-100 buffer:**

#### $1M \text{ HEPES } K^+ \text{ (pH 7.6)}$

Dissolve 119.15 g HEPES (MW 238.3) in 500 ml d.d. water to make 1M HEPES. Titrate with 10N KOH until the pH reaches 7.6. Store at 4°C.

#### 1M MgCl<sub>2</sub>

Dissolve 101.65 g MgCl<sub>2</sub>·6H2O in 400 ml d.d. water. Adjust the volume to 500 ml with d.d. water. Filter or autoclave to sterilize. Store at 4°C.

**Note:** MgCl<sub>2</sub> is extreamly hygroscopic. Buy small bottles and do not store opened bottles for long periods of time. Once the crystals become saturated with water, dispose of the chemical properly.

#### 10mM ZnCl<sub>2</sub>

Dissolve 681.4 mg ZnCl<sub>2</sub> in 500 ml sterile d.d. water. Filter to sterilize. Store at room temperature.

#### 0.5M EDTA (pH 8.0)

Add 90.8 g of  $Na_2EDTA \cdot 2H_2O$  to about 400 ml of d.d. water. Stir and adjust the pH to 8.0 with NaOH (~20g of NaOH pellet). Bring the volume to 500 ml with d.d. water if necessary. Strerilize by autoclaving. Store at room temperature.

**Note:** Na<sub>2</sub>EDTA·2H<sub>2</sub>O will not be dissolved until the pH of the solution is adjusted to approximately 8.0 by the addition of NaOH.

#### 4M KCl

Dissolve 149.1 g KCl in 400 ml of d.d. water. Bring the volume to 500 ml with d.d. water. Strerilize by autoclaving. Store at room temperature.

#### 1 M DTT (dithiothreitol)

Add 1.54 g DTT into 10 ml d.d water in a 15 ml Falcon tube. Vortex to dissolve. Wrap with aluminum foil and store at -20°C.

#### 100 mM PMSF (phenylmethylsulfonyl fluoride)

Add 696 mg PMSF in a 50 ml Falcon tube. Add 30 ml 100% ethanol. Vortex to dissolve. Fill with 100% ethanol to 40ml. Invert to mix. Wrap the tube with aluminum foil and store at -20°C.

**Note:** PMSF is very toxic so handle with care. Aqueous solutions of PMSF are hydrolyzed very rapidly, so the stock solution needs to be made with absolute ethanol or 2-propanol and only add PMSF to aqueous solutions immediately before their use.

\_\_\_\_\_

**HEMG-100 buffer (w/ 20mM Imidazole) (4 ml/purification)** 

	orig	inal sol	ution		final so	lution	
Chemical	conc.	unit	add	unit	conc.	unit	
HEPES K <sup>+</sup> (pH7.6)	1.0	M	500.0	μl	25.00	mM	
MgCl2	1.0	M	250.0	μl	12.50	mM	
ZnCl2	10.0	mM	20.0	μl	0.01	mM	
EDTA (pH8.0)	0.5	M	4.0	μl	0.10	mM	
KCl	4.0	M	500.0	μl	100.00	mM	
Glycerol	100.0	%	2.0	ml	10.00	%	
Imidazole	3.0	M	133.2	μl	20.00	mM	
DTT	1.0	M	60.0	μl	3.00	mM	
PMSF	0.1	M	100.0	μl	0.50	mM	
Sterile d.d. water			16432.8	μl			
final volume	•	•	20	01	•	•	

final volume 20.0 ml

**Note:** Add everything except DTT and PMSF into a 50 ml Falcon tube. Vortex to mix. Store at 4°C. Add DTT and PMSF <u>right before use.</u>

# Elution buffer (HEMG-100 buffer w/ 750mM Imidazole) (3 ml/purification)

original solution					final so	final solution	
Chemical	conc.	unit	add	unit	conc.	unit	
HEPES K <sup>+</sup> (pH7.6)	1.0	M	500.0	μl	25.00	mM	
MgCl2	1.0	M	250.0	μl	12.50	mM	
ZnCl2	10.0	mM	20.0	μl	0.01	mM	
EDTA (pH8.0)	0.5	M	4.0	μl	0.10	mM	
KCl	4.0	M	500.0	μl	100.00	mM	
Glycerol	100.0	%	2.0	ml	10.00	%	
Imidazole	3.0	M	5.0	ml	20.00	mM	
DTT	1.0	M	60.0	μl	3.00	mM	
PMSF	0.1	M	100.0	μl	0.50	mM	
Sterile d.d. water			11566.0	μl			
final volume			20.	0 ml			

**Note:** Add everything except DTT and PMSF into a 50 ml Falcon tube. Vortex to mix. Store at 4°C. Add DTT and PMSF <u>right before use.</u>

					final so	olution
Chemical	conc.	unit	add	unit	conc.	unit
HEPES K <sup>+</sup> (pH7.6)	1.0	M	50.0	ml	25.00	mM
MgCl2	1.0	M	25.0	ml	12.50	mM
ZnCl2	10.0	mM	2.0	ml	0.01	mM
EDTA (pH8.0)	0.5	M	400.0	μl	0.10	mM
KCl	4.0	M	50.0	ml	100.00	mM
Glycerol	100.0	%	200.0	ml	10.00	%
DTT	1.0	M	6.0	ml	3.00	mM
PMSF	0.1	M	10.0	ml	0.50	mM
Sterile d.d. water			to 2L			
final valuma				<u>о</u> т		

final volume 2.0 L

**Note:** Add everything except DTT and PMSF into a 4L plastic beaker with 1.5 L d.d. water. Stir to mix. Bring the volume to 2L with d.d. water. Stir to mix. Store at 4°C. Add DTT and PMSF <u>right before use</u>.

#### DPBS (Invitrogen, product code 14190) (Dulbecco's PBS)

**Note:** You can use any other 1X PBS from other vendors.

Protease Inhibitor Cocktail (SIGMA, product code P8340)

CelLytic M Lysis buffer (SIGMA, product code C2978)

**ProBond Nickel-Chelating Resin (Invitrogen, product code 46-0019)** 

Poly-Prep Chromatography columns (BioRad, product code 731-1550)

Spectra/Por 2 dialysis tubing (MWCO 12-14000 Da, nominal flat width 10 mm, 0.32 ml/cm) (SpectrumLab, product code 132676)

**Dialysis tubing clamps** 

# II. Expression of DmSNAPs in S2 cells

- 1. Grow 4 big plates (Corning 100 x 20 mm tissue culture plates) of cells in 20 ml selective medium to 70-80% confluency.
- 2. Induce cells with copper sulfate. Add  $100~\mu l~0.1 M~CuSO_4$  into each plate (to final 0.5~mM).
- 3. Incubate cells for ~24 hours at 22-25°C.

### III. Preparation of ProBond column

- 1. Put a poly-prep chromatography column in a 50ml Falcon tube.
- 2. Thoroughly suspend the ProBond resin in the vial in order to make a uniform suspension of the resin. Ratio of volume of suspension to packed gel is 2 to 1.
- 3. Using a 10 ml serological pipet. Immediately transfer 2 ml of suspended resin (containing 1ml packed resin) into the chromatography column.
- 4. Cap the column. Centrifuge the resin for 1 min at 420 g.
- 5. Remove the supernatant. Be careful not to remove any resin.
- 6. Wash the resin once in sterile d.d. water.
  - a. Add 6 ml sterile d.d. water to the column. Resuspend the resin thoroughly by inverting/tapping the column.
  - b. Cap the column. Centrifuge at 420 g for 1min.
  - c. Remove the supernatant.
- 7. Wash the resin twice with Native binding buffer.
  - a. Add 6 ml Native binding buffer to the column. Resuspend the resin thoroughly by inverting/tapping the column.
  - b. Cap the column. Centrifuge at 420 g for 1min.
  - c. Remove the supernatant.
  - d. Repeat steps a-c.

**Note:** For the <u>last wash</u> with Native binding buffer, if the resin will not incubate with cell lysates immediately, stop at Step a, leave the resin suspended in Native binding buffer and keep the column in the cold room. Centrifuge to remove the supernatant right before use.

# IV. Preparation of dialysis tubing

- 1. Cut the dialysis tubing into 8cm pieces (the capacity is  $\sim 1$ ml/piece).
- 2. Put the tubing pieces one-by-one into a beaker containing d.d. water with magnetic stir bar stirring. Allow it to stir for 30 mins. Make sure all pieces completely immersed in the water during stirring.
- 3. Transfer the tubing pieces into another beaker containing d.d. water. Make sure all pieces completely immersed in the water. Cover the beaker with aluminum foil and keep it in the cold room until used.

# V. Procedures of His-tagged protein purification

#### A. Cell lysis

1. Add and mix 70 μl protease inhibitor cocktail (to final 1%) and 23.3 μl 3M imidazole (to final 10 mM) to 7 ml of chilled CelLytic M lysis buffer in a 15 ml Falcon tube. Keep the tube on ice.

- 2. Harvest cells: (from 4 big plates)
  - a. To remove cells adhering to the dish, pipet the medium over the cells to wash cells off the plate for several times.
  - b. Collect the cells and medium from every 2 plates to one 50 ml centrifuge tube. (So for each cell line you will need 2 tubes for the 4 big plates)
  - c. Centrifuge in the RT-Legend at 420 g for 5 min.
  - d. Suck out the supernatant and discard.
  - e. Wash the cells by resuspending the pellet in one 50 ml tube with 10 ml of DPBS. Transfer the suspended cells to another pellet-containing 50 ml tube. Pipet up and down to resuspend the pellet thoroughly. Now all our cells are in a single 50 ml tube.
  - f. Centrifuge in the RT-Legend at 420 g for 5 min (use balance tube if necessary).
  - g. Suck out the supernatant and discard.
- 3. Lyse cells:
  - a. Transfer all 7ml of CelLytic M lysis buffer supplemented with 1% protease inhibitor and 10mM imidazole (Step IV-1) to the 50 ml tube containing washed cell pellet (Step V-2-g). Pipet up and down until all cells are lysed.
  - b. Transfer the lysed cells back to the 15 ml tube used for CelLytic M lysis buffer storage. Keep the 15 ml tube on ice and move to the cold room.
- 4. In the cold room, rotate the 15 ml tube containing lysed cells end-over-end for 15 mins to ensure complete lysis.

**Note:** From now on, you need to handle the lysed cells in the cold room.

- 5. After the 15 min incubation, aliquot the lysed cells into 7 of chilled 1.5 ml conical screw-cap tube (~1ml/tube). Centrifuge the cell lysate 12,000g for 10 mins in the microfuge in the cold room (EPPENDORF, Centrifuge 5415D).
- 6. Pool the supernatant from the 7 tubes (~7 ml total) into a new chilled 15 ml Falcon tube. Place the tube on ice.
- 7. If the lysate is viscous, shear the DNA by passing it through an 18-gauge needle four times.
  - a. Put an 18-gauge needle on a 10 ml syringe.
  - b. Suck the lysate into the syringe and expel it back into the tube slowly four times.
- 8. Remove 100  $\mu$ l of the lysate into a chilled 1.5 ml conical screw-cap tube labeled as "lysates" and freeze in liquid  $N_2$ . Keep the remainder of lysates on ice.
- 9. Measure the total amount of the remaining lysate. Add calculated amount of 5M NaCl into lysate to give a final concentration of 500 mM NaCl.

**Note:** Add 105.1 µl of 5M NaCl per ml of lysate to get final concentration of 500mM NaCl. For example, if the volume of lysate is 7 ml, add 735.7 µl of 5M NaCl.

#### B. Resin binding

- 1. In the cold room, add the cell lysates (~7 ml) to the prepared column containing packed ProBond resin (Step III-7). Resuspend the resin thoroughly by inverting/tapping the column.
- 2. Incubate the lysates and resin for 2 hrs in the cold room on the rocker

**Note:** The incubation time may need to be optimized if proteins other than DmSNAPs are purified.

- 3. Centrifuge the resin for 1min at 420 g in the RT-Legend. Transfer the supernatant into a chilled 15 ml Falcon tube. Remove 100 µl of the supernatant from the tube into a chilled 1.5 ml conical screw-cap tube and store frozen in the liquid nitrogen as "Flowthrough". Store the remainder in -80°C.
- 4. Wash the resin three times with Native wash buffer.
  - a. Add 4 ml Native wash buffer to the column. Resuspend the resin thoroughly by inverting/tapping the column.
  - b. Cap the column. Centrifuge at 420 g for 1min.
  - c. Remove the supernatant.
  - a. Repeat steps a-c for two more times.
- 5. Wash the resin once with HEMG-100 buffer (w/ 20 mM Imidazole).
  - a. Add 4 ml HEMG-100 buffer (w/ 20mM Imidazole) to the column. Resuspend the resin thoroughly by inverting/tapping the column.
  - b. Cap the column. Centrifuge at 420 g for 1min.
  - c. Remove the supernatant.

#### C. Elution of the His-tagged protein

- 1. Elute the bound His-tagged protein with three 1ml volumes of Elution buffer (HEMG-100 buffer w/ 750 mM Imidazole).
  - a. Clamp the column in a vertical position and snap off the cap on the lower end. Allow the remainder of the buffer in the column flow out and discard (will be just few drops).
  - b. Place a chilled 15 ml Falcon tube under the column. Position the tube to ensure that in next step, every drop of the elution from the column will be collected by the 15 ml tube.
  - c. Add 1ml of the elution buffer to the column <u>drop by drop</u>. Allow the eluted proteins to come out into the 15 ml tube underneath the column until completely drained (about 5mins).

- d. Remove 50  $\mu$ l of the elution fraction into a chilled 1.5 ml conical screw-cap tube and store frozen in the liquid nitrogen as "Elution 1". Save the remainder in the tube (~950  $\mu$ l) on ice.
- e. Repeat steps b-d twice to collect fractions 2 and 3.
- 2. Now you should have three 15 ml Falcon tubes sitting on ice as Elution 1, 2, and 3.

#### VI. Dialysis to remove imidazole in elution fractions

- 1. Prepare a 2L beaker containing 1L of ice-cold dialysis buffer (HEMG-100 buffer W/O imidazole) with stirring. Add 3ml of 1M DTT and 5ml of 100mM PMSF. Stirring.
- 2. Take out a prepared dialysis tubing (Step IV-3). Remove all the water remained inside/outside of the tubing. Use a dialysis clamp to close one end of the tubing. Pipet the Elution 1 from Step V-C-2 into the tubing. Close the other end with another clamp. Put the clamped tubing into the beaker containing dialysis buffer with stirring. Make sure the tubing is completely submerged.
- 3. Repeat Step 1 and 2 twice to transfer Elution 2 and 3 into individual dialysis tubing. Allow samples to dialyze for 2hr.
- 4. Exchange the dialysis buffer in the beaker with another 1L of ice-cold dialysis buffer supplemented with DTT and PMSF. Dialyze for another 2 hr.
- 5. Take out the tubing containing Elution 1. Unclamp one end of the tubing. Transfer the solution into a chilled 1.5 ml conical screw-cap tube labeled with "Elution 1 dialyzed". Store frozen in the liquid nitrogen.
- 6. Repeat Step 5 twice to transfer Elution 2 and 3 into individual 1.5 ml conical screw-cap tubes as "Elution 2 dialyzed" and "Elution 3 dialyzed". Store frozen in the liquid nitrogen.

# Appendix E: Detailed Protocol for electrophoretic mobility shift assay (EMSA)

#### I. Solutions and materials

Materials for preparation of radioactive DNA oligo probes:

Annealed double stranded DNA oligos (PSEAs) (1 µg/5 µl)

T4 polynucleotide kinase (T4 PNK) (NEB, product code: M0201L)

10x PNK reaction buffer (reagent supplied with T4 PNK)

<sup>32</sup>P gamma-ATP (3000 Ci/mmol, 10m Ci/µl) (PerkinElmer)

Chloroform/isoamyl alcohol (24:1)

Saturated phenol

Quick Spin Columns for radiolabeled DNA purification Sephadex G-25, fine (ROCHE, product code: 11273949001)

#### **Materials for bandshift/supershift reactions:**

Radioactive oligo probes

**0.1M DTT (dithiothreitol)** 

Diluted from 1M DTT. Preparation of 1M DTT is described two pages ahead.

Poly(deoxyguanylic-deoxycytidylic) acid sodium salt [Poly (dG-dC)] (SIGMA, product code: P9389)

Dissolved in HEMG-100 buffer to obtain final concentration of 1µg/µl.

Poly(deoxyinosinic-deoxycytidylic) acid sodium salt [Poly (dI-dC)] (SIGMA, product code: P4929)

Dissolved in HEMG-100 biffer to obtain final concentration of 1µg/µl.

#### **Glycerol**

#### **Purified proteins (DmSNAPs)**

#### **Appropriate antibodies (for supershift reactions)**

#### **HEMG-100** buffer

	origin	al solu	tion		final so	olution
Chemical	conc.	unit	add	unit	conc.	Unit
HEPES K <sup>+</sup> (pH7.6)	1.0	M	500.0	μl	25.00	mM
MgCl2	1.0	M	250.0	μl	12.50	mM
ZnCl2	10.0	mM	20.0	μl	0.01	mM
EDTA (pH8.0)	0.5	M	4.0	μl	0.10	mM
KCl	4.0	M	500.0	μl	100.00	mM
Glycerol	100.0	%	2.0	ml	10.00	%
Sterile d.d. water			16726.0	μl		
<u>۳</u> 1 1			• •			

final volume 20.0 ml

**Note:** Add everything into a 50 ml Falcon tube. Vortex to mix. Store at 4°C

#### **Chemicals for preparation of HEMG-100 buffer:**

#### **1M HEPES K**<sup>+</sup> (**pH 7.6**)

Dissolve 119.15 g HEPES (MW 238.3) in 500 ml d.d. water to make 1M HEPES. Titrate with 10N KOH until the pH reaches 7.6. Store at 4°C.

#### 1M MgCl<sub>2</sub>

Dissolve 101.65 g MgCl<sub>2</sub>·6H2O in 400 ml d.d. water. Adjust the volume to 500 ml with d.d. water. Filter or autoclave to sterilize. Store at 4°C.

**Note:** MgCl<sub>2</sub> is extreamly hygroscopic. Buy small bottles and do not store opened bottles for long periods of time. Once the crystals become saturated with water, dispose of the chemical properly.

#### 10mM ZnCl<sub>2</sub>

Dissolve 681.4 mg ZnCl<sub>2</sub> in 500 ml sterile d.d. water. Filter to sterilize. Store at room temperature.

#### 0.5M EDTA (pH 8.0)

Add 90.8 g of Na<sub>2</sub>EDTA·2H<sub>2</sub>O to about 400 ml of d.d. water. Stir and adjusted the pH to 8.0 with NaOH (~20g of NaOH pellet). Bring the volume to 500 ml with d.d. water if necessary. Strerilize by autoclaving. Store at room temperature.

**Note:** Na<sub>2</sub>EDTA·2H<sub>2</sub>O will not be dissolved until the pH of the solution is adjusted to approximately 8.0 by the addition of NaOH.

#### 4M KCl

Dissolve 149.1 g KCl in 400 ml of d.d. water. Bring the volume to 500 ml with d.d. water. Strerilize by autoclaving. Store at room temperature.

#### 1 M DTT (dithiothreitol)

Add 1.54 g DTT into 10 ml d.d water in a 15 ml Falcon tube. Vortex to dissolve. Wrap with aluminum foil and store at -20°C.

#### 100 mM PMSF (phenylmethylsulfonyl fluoride)

Add 696 mg PMSF in a 50 ml Falcon tube. Add 30 ml 100% ethanol. Vortex to dissolve. Fill with 100% ethanol to 40ml. Invert to mix. Wrap the tube with aluminum foil and store at -20°C.

**Note:** PMSF is very toxic so handle with care. Aqueous solutions of PMSF are hydrolyzed very rapidly, so the stock solution needs to be made with absolute ethanol or 2-propanol and only add PMSF to aqueous solutions immediately before their use.

#### Materials for preparation of non-denaturing polyacrylamide gel:

#### 40% non-denaturing acrylamide stock solution (30:1)

Dissolve 38.71 g of electrophoresis-grade acrylamide, 1.29 g electrophoresis-grade bis-acrylamide in 100 ml d.d. water with stirring. Sterilize by passage through a 0.22-µm filter. Wrap the bottle with foil and store at 4°C. Discard the solution if the color turns yellow during storage.

#### 10x non-circulation buffer

Dissolve 60.58 g of Tris, 285.28 g of glycine, and 7.44 g of EDTA in 1.6 L d.d. water with stirring. Stir and adjust the pH to 8.3 with HCl. Bring the volume to 2 L with d.d. water. Store at 4°C (indefinitely). Discard the solution if the color turns yellow during storage.

#### 1x non-circulation buffer (gel-running buffer)

Dilute from 10x non-crculation buffer with d.d. water. Around 800 ml is required for each gel-running apparatus.

#### 10% Ammonia Persulfate (APS)

Add 1 g APS in a 15 ml Falcon tube. Add 10 ml sterile d.d. water. Vortex to dissolve. Aliquot the solution into microfuge tubes (1ml/tube). Store at -20°C (indefinitely).

**Note:** APS provides the free radicals that drive polymerization of acrylamide and bis-acrylamide. APS decomposes gradually (it will last only a week at 4°C). Thus, once leave the freezer and get thawed, the 10% APS must stay on ice all the time and put back to the freezer right after use.

#### **TEMED**

Store at 4°C. Keep on ice when in use.

**Note:** TEMED serves as the catalyst for the polymerization of acrylamide and bis-acrylamide.

Large and small glass plates, spacers, vacuum glue, metal clamps, 20well combs, non-stick reagent, pieces of sponge, gel-running apparatus, power supply and wires, food wrap, intensifier screen, film cassette

#### II. Preparation of Quick Spin columns

You will need to prepare two columns at the same time following the instruction below.

- 1. Place a collection tube with the end cut-off in a 15 ml Falcon tube. Do another set for the 2<sup>nd</sup> column.
- 2. Thoroughly suspend the G-25 resin in columns to make a uniform suspension of the resin by ticking/inverting the column.
- 3. Remove the cap and tip from columns. Place the two columns into each 15 ml Falcon tubes prepared in step III-1.
- 4. Centrifuge in the RT-Legend at 1000 g for 3 min, 4°C.
- 5. Use forceps to take out the column for a while. Remove the flowthrough and the end-cut-off collection tube from the 15 ml Falcon tube. Replace the column and the end-cut-off collection tube back to the 15 ml Falcon tube.
- 6. Centrifuge in the RT-Legend at 1000 g for 3 min, 4°C.
- 7. Use forceps to take out the column for a while. Remove the flowthrough and the end-cut-off collection tube from the 15 ml Falcon tube. Replace the column and an intact collection tube back to the 15 ml Falcon tube.

**Note:** label one intact collection tube with "1", and label the other with the name of your oligos and the date of radiolabeling. Put the column label with "1" into the 15 ml Falcon tube containing "1" collection tube, put the "2" column into the 15 ml Falcon tube containing the collection tube labeled with the detailed information of your oligos.

## III. Radiolabeling of annealed DNA oligos

1. Prepare the following reaction in a 1.5 ml conical screw-cap tube:

1 μg annealed oligos (PSEAs)	5 µl
10x T4 PNK reaction buffer	5 µl
T4 PNK (10 U/μl)	1 μl
<sup>32</sup> P gamma-ATP	8 µl
Sterile d.d. water	31 µl
final volume	50 μl

**Note:** use long P10 tips for transferring the gamma-ATP to avoid potential contamination.

- 2. Incubate the tube in a 37°C water bath for 30 mins.
- 3. While waiting, prepare the Quick Spin column as described in section II.
- 4. Take out the oligo tube from the water bath. Add 25  $\mu$ l of chloroform/isoamyl alcohol (24:1) into the tube.

- 5. Add 25 µl of saturated phenol (get the lower layer) into the tube.
- 6. Vigorously vortex for 30 sec.
- 7. Centrifuge at 12000 rpm or maximum speed for 3 min at room temperature.
- 8. Transfer the aqueous top layer (about 50  $\mu$ l) from the tube into the center of the prepared column labeled with "1" atop the "1" collection tube inside the 15 ml Falcon tube.
- 9. Centrifuge in the RT-Legend at 1000 g for 3 min, 4°C.
- 10. Use forceps to remove the column and transfer the collection tube containing radiolabeled oligos to a rack.
- 11. Transfer the radiolabeled oligos (about 50 µl) into the center of the prepared column labeled with "2" atop the collection tube labeled with the detailed oligo information inside the 15 ml Falcon tube.
- 12. Centrifuge in the RT-Legend at 1000 g for 3 min, 4°C.
- 13. Use forceps to remove the column and transfer the collection tube containing radiolabeled oligos to a rack. Insert the removable cap to the tube. Now this is your radiolabeled and purified DNA oligos. Store your purified oligos on ice if using immediately, otherwise store at -20°C.
- 14. Prepare two 1.5 ml conical screw-cap tubes each containing 2  $\mu$ l of purified radioactive oligos. Examine the radioactivity of oligos in these two tubes in the scintillation counter (use #6 for  $^{32}$ P). Store these two tubes in the -20°C freezer for future use.

**Note:** to calculate the radioactivity (cpm) of your probes, if the result from the counting is "A" cpm from tube 1 and "B" cpm from tube 2, then the cpm/ $\mu$ l of your probe is (A cpm + B cpm)/(2  $\mu$ l x 2) = (A+B)/4 cpm/ $\mu$ l

#### IV. Preparation of non-denaturing polyacrylamide gel

- 1. Apply non-stick reagents on a small glass plate. Use kimwipes to spread the reagent evenly. Assemble the treated small glass plate with a large glass plate and 3 spacers into a "sandwich" with vacuum glue applied at the joint of the spacers. Clamp the edge of the sandwich with metal clamps (2 on lower part of each right and left side, 3 on the bottom side. Total 7 clamps are used in this step). Lay the sandwich on a tip box on the bench so the sandwich is tilted with the bottom side touching the benchtop.
- 2. Prepare the 5% non-denaturing acrylamide solution as follows: a. Add 7.5 ml 40% non-denaturing acrylamide, 6 ml 10x non-circulation buffer, and 46.5 ml d.d. water into a 250 ml flask. Swirl to mix. Remove 5 ml from the gel solution and discard.
  - b. Add 400 μl of 10% APS and 40 μl of TEMED into the flask. Swirl to mix (avoid bubbles). Use a transfer pipet to remove bubbles if necessary.

**Note:** if you need to run supershift reactions on the gel, then you might need to prepare a 4% gel instead: mix together 5 ml of 40% non-denaturing acrylamide, 5ml of 10x non-circulation buffer, 39.56 ml of d.d. water, 400  $\mu$ l of 10% APS and 40  $\mu$ l TEMED in the flask. You don't need to discard any gel solution in this case.

- 3. Immediately (but slowly and steady) pour the gel solution into the middle space of the assembled sandwich. Avoid any bubbles that may occur. Insert a 20-well comb and immediately clamp two extra metal clamps (one for each upper part of right and left side) to fix the comb.
- 4. Wait for 30 min allowing the gel to completely polymerize.
- 5. Once the gel is solidified, hook up the gel sandwich onto a gel-running apparatus connected to a power supply with wires.
- 6. Remove the comb. Pour 1x non-circulation buffer into the upper tank and the lower tank of the gel-running apparatus so the wells of the gel are completely immersed in the buffer. Use a syringe with needle to remove unpolymerized acrylamide and bubbles inside wells, and to remove bubbles from the space in the bottom of the gel.
- 7. Run to warm up the gel at 100 V for 30 min.

**Note:** Do not start this warm-up step until your bandshift/supershift reactions are ready for the 30 min incubation.

#### V. Preparation of bandshift/supershift reaction

4. Calculate how much radioactive probe you need according to the number of your reactions and the radioactive strength of the probe. Each reaction requires 1  $\mu$ l of 50000 cpm/ $\mu$ l probe. For example, if you need 20 reactions and the radioactivity of your probes measured from step III-14 is 400000 cpm/ $\mu$ l, then:

$$(50000 \text{ cpm/}\mu \text{l x } 20 \text{ }\mu \text{l})/(400000 \text{ cpm/}\mu \text{l}) = 2.5 \text{ }\mu \text{l}$$

Thus, you need 2.5  $\mu$ l of the 400000 cpm/ $\mu$ l probe to dilute with (20-2.5=17.5)  $\mu$ l of d.d. water to make 20  $\mu$ l of 50000 cpm/ $\mu$ l probe for your bandshift reaction. You can also use 22  $\mu$ l instead of 20  $\mu$ l in the equation to make sure you have enough probe to use.

- 5. Prepare the probe-mix by mixing 1  $\mu$ l of the 50000 cpm/ $\mu$ l probe, 2  $\mu$ l of 1  $\mu$ g/ $\mu$ l poly (dI-dC) or poly (dG-dC), and 1  $\mu$ l of 0.1 M DTT in a 1.5 ml conical screw-cap tube for each reaction (so 4  $\mu$ l of probe-mix per reaction). Multiple by the number of your total reactions to see how much of each reagent you really need.
- 6. Prepare each bandshift reaction as follow:

HEMG-100	(15-X) μl
Sterile d.d. water	2 μl
Probe-mix	4 µl
Proteins (DmSNAPs)	Xμl
Final volume	21 µl

**Note:** the final salt concentration should be around 80 mM, and the final glycerol concentration should be around 8%.

7. Incubate the reactions in a 20°C water bath for 30 mins. If a supershift reaction is included, add antibodies in the middle of the incubation (15 min after incubation). Start this step with step IV-7 (gel warm-up) at the same time.

#### VI. Gel running and autoradiography

1. Load each well of the gel with each of your bandshift/supershift reactions. Load a empty well on the side with the non-denaturing dye. Run the gel at 100 V until the fast dye is approximately 3/4 through the gel (it will take around 3 hr and 20 min).

**Note:** you might need to run the gel longer to allow the dye close the bottom of the gel if you have supershift reactions. This will allow the protein-DNA bands to separate further then it will be much easier to observe supershift bands.

- 2. Detached the gel sandwich from the gel-running apparatus. Dissemble the sandwich to allow the gel to separate from the small glass plate but stay on the large plate.
- 3. Tilt the plate with the gel on it to allow the buffer remained on the gel to run away from the gel. Use kimwipes to absorb the buffer.
- 4. Immediately lay a piece of food wrap on the surface of the gel. That piece of food wrap needs to be large enough to cover the whole gel and the large glass plate to allow full wraping of the gel.
- 5. In a dark room, place your wrapped gel/glass plate in a film cassette. Place a film on top of the gel. Place an intensifier screen on top of the film. Close and tightly fasten the cassette. Put the cassette into a -80°C freezer to allow the exposure of the film up to 18 hrs.
- 6. Develop the film in the darkroom to see the result.

#### **Detailed** Protocol Chromatin **Appendix** F: for immunoprecipitation Assay (ChIP)

#### I. **Solutions and materials**

#### Formaldehyde cross-linking solution

		final so	lution				
Chemical	conc.	unit	add	unit	conc.	unit	
Tris-HCl (pH8.0)	1.0	M	1.400	ml	50.0	mM	
EDTA (pH8.0)	0.5	M	0.056	ml	1.0	mM	
EGTA	0.1	M	0.140	ml	0.5	mM	
NaCl	5.0	M	0.560	ml	100.0	mM	
Formaldehyde	37.0	%	0.757	ml	1.0	<b>%</b>	
final solution volume	28.0	ml	* final volume is from all chemicals+				
	25ml S2 call cultura						

25ml S2 cell culture.

**Note:** Add each individual chemical directly to the 25 ml S2 cells in the order shown.

#### 2M glycine

Dissolve 7.5g glycine in 50 ml sterile d.d. water. Store at 4°C.

#### **Sonication buffer**

		origi	final sol	lution		
Chemical	conc.	unit	add	unit	conc.	Unit
Tris-HCl (pH8.0)	1.0	M	0.50	ml	10.0	mM
EDTA (pH8.0)	0.5	M	0.10	ml	1.0	mM
EGTA	0.1	M	0.25	ml	0.5	mM
PMSF	0.2	M	0.125	ml	0.5	mM
final solution volume	50.0	ml	(add d.d. water to bring to the final			

Notes: PMSF is inactivated in aqueous solutions, so stock solution should be made in ethanol or isopropanol and only added to aqueous solutions immediately before use. Add 2.5 µl 0.2M PMSF per 1ml sonication buffer.

Also, add 10 µl protease inhibitor cocktail (SIGMA) per 1 ml sonication buffer right before use.

# Buffer for preparation of dialysis tubing

Buffer I	(	origina	final sol	ution		
Chemical	conc.	unit	add	unit	conc.	Unit
sodium bicarbonate (NaHCO3) (powder)	100.0	%	20.0	g	2.0	%
EDTA (pH 8.0)	0.5	M	2.0	ml	1.0	mM
final solution volume	1000.0	ml	(add d.d. volume.)	water to 1	oring to the f	inal

#### **Buffer II**: 1mM EDTA (pH 8.0)

2ml 0.5M EDTA (pH 8.0)/1000 ml d.d. water.

#### 6M Urea

Dissolve 18.02 g Urea in 30 ml d.d. water. Adjust the volume to 50 ml with d.d. water. Store at room temperature. Use within 1-2 weeks.

#### ChIP buffer

	o	riginal	final s	final solution		
chemical	conc.	unit	add	unit	conc.	unit
Tris-HCl (pH8.0)	1.0	M	10.0	ml	10.0	mM
EDTA (pH 8.0)	0.5	M	2.0	ml	1.0	mM
EGTA	0.1	M	5.0	ml	0.5	mM
PMSF	0.2	M	2.5	ml	0.5	mM
glycerol	100.0	%	100.0	ml	10.0	%
Triton-X100	100.0	%	10.0	ml	1.0	%
sodium deoxycholate (powder)	100.0	%	1.0	g	0.1	%
final solution volume	1000.0	ml	(add water to bring to the final volume. Store at 4°C.)			

**Notes:** PMSF is inactivated in aqueous solutions, so stock solution should be made in ethanol or isopropanol and only add to aqueous solutions immediately before use.

Usually only 200 ml ChIP buffer is needed per dialysis, so add 500  $\mu$ l 0.2M PMSF for 200 ml ChIP buffer.

#### Appropriate antiserum and pre-immune serum

Anti-DmSNAP43 Ab (DmSNAP43 (03978) antibody 12-6-04 Nermeen) and Anti-FLAG polyclonal Ab (SIGMA, product code: F7425) are used in this case.

#### Immobilized Protein A sepharose (PIERCE, product code: 20333)

**Note:** Binding specificities and affinities of different antibody-binding proteins (protein A, G, A/G, and L) differ between source species and antibody subclass. In this case, protein A is selected because of its high affinity to rabbit IgG (anti-DmSNAP43 Ab and anti-FLAG polyclonal Ab). You may need to use other antibody-binding proteins if other antibodies are used in your application.

#### TE buffer

Add 1 ml of 1M Tris-HCl pH 8.0 and 200 µl of 0.5 M EDTA pH 8.0 in a 100 ml cylinder. Fill d.d. water to 100 ml graduation. Filter to sterilize. Store at 4°C.

#### 10 mg/ml BSA (NEB, product code: B9001S)

Directly use the NEB 100X BSA (10 mg/ml) that comes with restriction enzymes.

#### Low-salt wash buffer

	0	riginal	final solution			
chemical	conc.	unit	add	unit	conc. unit	
Tris-HCl (pH8.1)	1.0	M	4.0	ml	20.0 mM	
EDTA (pH 8.0)	0.5	M	0.8	ml	2.0 mM	
NaCl	5.0	M	6.0	ml	150.0 mM	
SDS	10.0	%	2.0	ml	0.1 %	
Triton-X100	100.0	%	2.0	ml	1.0 %	
final solution volume	200.0	ml	(add d.d. water to bring to the final volume. Store at 4°C.)			

#### High-salt wash buffer

	original solution final solution					
Chemical	conc.	unit	add	unit	conc.	unit
Tris-HCl (pH8.1)	1.0	M	4.0	ml	20.0	mM
EDTA (pH 8.0)	0.5	M	0.8	ml	2.0	mM
NaCl	5.0	M	20.0	ml	500.0	mM
SDS	10.0	%	2.0	ml	0.1	%
Triton-X100	100.0	<b>%</b>	2.0	ml	1.0	%
final solution volume	200.0	ml	(add d.d. water to bring to the			
illiai solution volume	200.0	1111	final v	volume	. Store at 4°C	C.)

#### Lithium wash buffer

itiliaili wasii sallol						
	0:	final s	solution			
Chemical	conc.	unit	add	unit	conc.	unit
Tris-HCl (pH8.1)	1.0	M	1.0	ml	10.0	mM
EDTA (pH 8.0)	0.5	M	0.2	ml	1.0	mM
LiCl	10.0	M	2.5	ml	250.0	mM
NP-40	100.0	%	1.0	ml	1.0	<b>%</b>
sodium deoxycholate (powder)	100.0	%	1.0	g	1.0	%
final solution volume	100.0	ml	(add d.d. water to bring to the final volume. Store at 4°C.)			

#### **ChIP** elution buffer (freshly made)

	origii	nal sol	final s	olution		
Chemical	conc.	unit	add	unit	conc.	unit
sodium bicarbonate (NaHCO3)	1.0	M	10.0	ml	100.0	mM
SDS	10.0	%	10.0	ml	1.0	%
final colution volume	100.0	ml	(add water to brin			final
final solution volume		mı	volume	e. Do no	ot chill.)	

#### DPBS (Invitrogen, product code 14190) (Dulbecco's PBS)

**Note:** You can use any other 1X PBS from other vendors.

Protease Inhibitor Cocktail (SIGMA, product code P8340)

Sonicator (Branson sonifier 250 Analog, with microtip) (in Huxford lab)

Spectra/Por 2 dialysis tubing (MWCO 12-14000 Da, nominal flat width 10 mm, 0.32 ml/cm) (SpectrumLab, product code 132676)

**Dialysis tubing clamps** 

Tris-HCl (pH6.5)

Proteinase K (2mg/ml)

QIAquick PCR purification kit (QIAGEN, product code 28104)

Platinum PCR SuperMix (Invitrogen, product code: 11306-016)

Appropriate forward and reverse primers for PCR reactions (200 ng/reaction)

U1Forward (5'-GTGTGGCATACTTATAGGGGTGCT-3') and U1Backward (5'-GCTTTTCGATGCTCGGCAGCAG-3') primers that amplify the promoter region of the U1:95Ca gene from -1 to -107 relative to the transcription start site are used in this case.

#### PCR machine (BioRad iCycler)

#### 10X TBE

#### 10X ChIP loading dye

Add 2.1 ml 1% bromophenol blue to 2.5 ml glycerol in a 15 ml Falcon tube. Add water to bring to 5 ml. Vortex to mix. Store at room temperature.

### II. Preparation of dialysis tubing

- 8. Cut the dialysis tubing into 15 cm pieces (~2 ml capacity/piece).
- 9. Boil the tubing in 800 ml buffer I in a glass beaker for 10 minutes with stirring.
- 10. Rinse the tubing thoroughly in d.d. water.
- 11. Boil the tubing in 800 ml buffer II in a glass beaker for 10 minutes with stirring.
- 12. Allow the tubing to cool, and then store it in cold room overnight (cover the beaker with aluminum foil). Make sure the tubing is always submerged.

**Note:** From now on, always wear gloves to handle the tubing.

13. Before use, wash the tubing inside and out with d.d. water.

## III. Formaldehyde cross-linking, sonication, and dialysis

# Day 1

- 15. Grow 3 plates (Corning 100 x 20 mm tissue culture plate) of *Drosophila* S2 cells to 90 % confluency.
- 16. Harvest cells:
  - a. To remove cells adhering to the dish, pipet the medium over the cells gently several times.
  - b. Pool cells from all plates and transfer 25 ml of cells into a 50 ml Falcon tube.
- 17. Add chemicals of formaldehyde cross-linking solution into the 25 ml cells for cross-linking. Mix well. Incubate at room temperature for 10 minutes on a rotating wheel.
- 18. Add 3.8 ml 2M glycine to final 240 mM to quench the cross-linking reaction.
- 19. Spin the cells at 700 g, 4°C for 10 minutes (SORVALL, Legend RT). Discard the supernatant.

**Note:** The supernatant contains formaldehyde, which is a carcinogen. So it is important NOT to directly drain the supernatant into the sink or trash can. Instead, collect the supernatant into a 50 ml Falcon tube and toss it into the chemical hazard container.

- 20. Resuspend the cells with 10 ml ice-cold DPBS. Centrifuge at 700 g, 4°C for 10 minutes (SORVALL, Legend RT). Discard the supernatant.
- 21. Resuspend the cells with 1 ml sonication buffer with 2.5 µl 0.2M PMSF and 10 µl protease inhibitor cocktail. Transfer the suspension to a chilled 15 ml conical-bottom Falcon tube.
- 22. Sonicate the suspension on ice with the following condition: microtip, 60% duty cycle, 1.5 output, 30 second on/1 minute off, 10 cycles (14 minutes total).
- 23. Transfer sonicated solution to 2 chilled 1.5 ml screw-cap tubes (500 μl/tube). Centrifuge at 13,000 rpm, 4°C for 10 minutes (EPPENDORF, Centrifuge 5415D. In the cold room).
- 24. Pool supernatant from both tubes to a chilled 15 ml Falcon tube. Mix the solution with equal amount (~1ml) of 6M Urea.

**Note:** Mix well. Save a 50 µl aliquot separately in a 1.5 ml screw-cap tube at -20°C in case it is necessary to determine DNA size.

- 25. Working in the cold room, take out the prepared dialysis tubing (Step II-6). Use a dialysis clamp to close one end of the tubing. Pipet the solution from Step III-10 into the tubing. Close the other end with another clamp.
- 26. Prepare a beaker containing 200 ml of ice-cold ChIP buffer (prepared the day before). Add 500  $\mu$ l 0.2M PMSF per 200 ml ChIP buffer right before dialysis. Put tubing into the buffer and stir overnight in the cold room for dialysis.

# IV. Preparation of 50% protein A sepharose

# **Day 2**

1. Gently vortex to thoroughly suspend the immobilized protein A sepharose in the vial.

**Note:** Ratio of volume of suspension to packed gel is 2 to 1 in the vial.

2. Using a P-1000 with  $\sim$ 2 mm cut-off end of the tip, immediately transfer 200 µl suspended resin (100 µl packed resin) from the vial to a chilled 1.5 ml screwcap tube.

**Notes:** Always use tips cut off at the end to handle the resin. More than 200  $\mu$ l resin may need to be prepared depending upon the number of samples to be done.

- 3. Centrifuge at 2500 g, 4°C for 3 minutes (EPPENDORF, Centrifuge 5415D. In the cold room). Pipet off the supernatant.
- 4. Continue working with EPPENDORF Centrifuge 5415D. Wash resin twice with 1 ml sterile d.d. water.
  - a. Resuspend beads with 1 ml sterile d.d. water.
  - b. Centrifuge at 2500 g, 4°C for 3 minutes.
  - c. Pipet off the supernatant.
  - d. Repeat steps a-c.
- 5. Wash resin twice with 1 ml TE buffer.
  - a. Resuspend beads with 1 ml TE buffer.
  - b. Centrifuge at 2500 g, 4°C for 3 minute.
  - c. Pipet off the supernatant.
  - d. Repeat steps a-c.
- 6. Wash resin once with 1 ml TE+BSA (900 μl TE buffer +100 μl 10 mg/ml BSA).
  - a. Resuspend beads with 1 ml TE+BSA.
  - b. Centrifuge at 2500 g, 4°C for 3 minute.
  - c. Pipet off the supernatant.
- 7. Resuspend beads with 100 μl TE+BSA (90 μl TE buffer +10 μl 10 mg/ml BSA). Now you have 200 μl of prepared 50% protein A resin. Store at 4°C.

**Note:** Among 200 µl prepared 50% resin, 80 µl is for pre-clearing, 35 µl is for each pre-immune serum and each anti-serum treated samples.

# V. Pre-clearing and immunoprecipitation

- 1. Remove a clamp from one end of the dialysis tubing (Step III-12). Pipet out the solution into a chilled 1.5 ml conical screw-cap tube. Centrifuge at 13,000 rpm, 4°C for 10 minutes (EPPENDORF, Centrifuge 5415D. In the cold room).
- 2. Transfer the supernatant (~1 ml) into a chilled 1.5 ml screw-cap tube. This is the chromatin solution.
- 3. Add 80 µl of suspended protein A resin (Step IV-7) to the chromatin solution for pre-clearing. End-over-end rotate at 4°C for 30 minutes.

**Note:** This step (pre-clearing) is important to reduce the background signals.

- 4. Centrifuge at 13,000 rpm, 4°C for 5 minutes (EPPENDORF, Centrifuge 5415D. In the cold room).
- 5. Aliquot the supernatant into chilled 1.5 ml screw-cap tubes (200 μl for input; 150 μl for each pre-immune serum and each antiserum). Save the remainder in a chilled 1.5 ml screw-cap tube and store it along with the input tube at -80°C. Note: Input sample serves as positive PCR control. Pre-immune serum serves as negative immunoprecipitation control.

 Add 4 μl of antiserum or pre-immune serum into each corresponding tube containing pre-cleared chromatin solution. End-over-end rotate at 4°C overnight.

**Note:** The amount of antiserum added may need to be optimized according to what antiserum you use. In this case, the antiserum used are anti-DmSNAP43 Ab and anti-FLAG polyclonal Ab (SIGMA, product code: F7425).

## Day 3

- 7. Using a P-100 with the bottom of the tip cut off, add 35 µl prepared 50% protein A resin (Step IV-7) into each pre-immune tube and each antiserum tube. End-over-end rotate at 4°C for 2 hours.
- 8. Spin down the resin at 2500 g, 4°C for 3 minutes (EPPENDORF, Centrifuge 5415D. In the cold room). The resin is the important fraction, but save the supernatant (as "Flow through") in a 1.5 ml screw-cap tube at -80°C in case it should be needed.
- 9. Continue working in the cold room and with EPPENDORF Centrifuge 5415D. Wash resin 3 times with 1 ml ice-cold low-salt wash buffer.
  - a. Resuspend beads with 1 ml low-salt wash buffer.
  - b. End-over-end rotate at 4°C for 5-10 minutes.
  - c. Centrifuge at 2500 g, 4°C for 3 minute.
  - d. Pipet off the supernatant.
  - e. Repeat steps a-d for 5 more times.
- 10. Wash resin 3 times with 1 ml ice-cold high-salt wash buffer.
  - a. Resuspend beads with 1 ml high-salt wash buffer.
  - b. End-over-end rotate at 4°C for 5-10 minutes.
  - c. Centrifuge at 2500 g, 4°C for 3 minute.
  - d. Pipet off the supernatant.
  - e. Repeat steps a-d for 2 more times.
- 11. Wash resin twice with 1 ml ice-cold lithium wash buffer.
  - a. Resuspend beads with 1 ml lithium wash buffer.
  - b. End-over-end rotate at 4°C for 2 hours.
  - c. Centrifuge at 2500 g, 4°C for 3 minute.
  - d. Pipet off the supernatant.
  - e. Resuspend beads with 1 ml lithium wash buffer.
  - f. End-over-end rotate at 4°C overnight (the overnight wash is believed to be important).
  - g. Centrifuge at 2500 g, 4°C for 3 minute.
  - h. Pipet off the supernatant.

# Day 4

- 12. Wash resin with 1 ml ice-cold TE buffer.
  - a. Resuspend beads with 1 ml TE buffer.
  - b. End-over-end rotate at 4°C for 5 minutes.
  - c. Centrifuge at 2500 g, 4°C for 3 minute.
  - d. Pipet off the supernatant.
- 13. Resuspend resin with 1 ml TE buffer. Transfer the resin to another chilled 1.5 ml screw-cap tube to eliminate non-specific DNA bound on the tube wall.
- 14. Centrifuge at 2500 g, 4°C for 3 minute. Pipet off the supernatant.

### VI. Elution, reverse cross-linking, and DNA purification

- 1. Add 250 μl of freshly made elution buffer to the resin (Step V-14) to elute the immunoprecipitated protein-DNA complexes. Vortex briefly to mix well.
- 2. End-over-end rotate <u>at room temperature</u> for 15 minutes. Centrifuge at 2500 g, <u>room temperature</u> for 3 minute (EPPENDORF, Centrifuge 5424).
- 3. Transfer the supernatant (eluate) to a 1.5 ml screw-cap tube.
- 4. Add another 250  $\mu$ l of elution buffer to the resin to elute again. Repeat Step VI-2. Pool eluate from both elutions together in a 1.5 ml screw-cap tube (~500  $\mu$ l total). Also, prepare input DNA by adding 300 ul elution buffer to 200  $\mu$ l thawed input sample (Step V-5) to make final volume 500  $\mu$ l.
- 5. Add 20 μl of 5M NaCl to eluate and input DNA. Mix well and incubate at 65°C for 4 hours to reverse crosslinks.
- 6. Add 10 μl of 0.5M EDTA pH8.0, 20 μl of 1M Tris-HCl pH6.5, and 10 μl of 2 mg/ml proteinase K. Mix well and incubate at 45°C for 1 hour to digest proteins.
- 7. Using QIAquick PCR purification kit (QIAGEN), follow the manufacturer's instructions to purify the immunoprecipitated DNA and input DNA. Store the purified DNA at -20°C.

**Notes:** Use 2500  $\mu$ l of PB buffer (as 5 volumes PB: 1 volume sample) in the first step (binding step); use 30  $\mu$ l of TE buffer to elute the purified DNA in the last step. 30  $\mu$ l of purified DNA is enough for 15 PCR reactions.

Now you will have at least 3 purified DNA sample: 1 for input DNA; 1 (or more) for pre-immune serum precipitated DNA; 1 (or more) for antiserum precipitated DNA.

#### VII. PCR reaction

# **Day 5**

1. Prepare the PCR reaction by using purified pre-immune, anti-serum precipitated DNA, and input DNA (Step-VI-7) as instructed below:

Reagents	amount (µl)
Platinum PCR SuperMix	45
Forward primer (0.08 µg/µl)	2.5
Reverse primer (0.08 μg/μl)	2.5
DNA from ChIPs	2
Total	52

2. Use the following program to run the PCR reaction:

Cycle1: (1x)	Step 1	94°C	2:00 min	hot-start
Cycle2: (28x)	Step 1	94°C	0:30 min	denature
	Step 2	61°C	0:45 min	anneal
	Step 3	72°C	1:00 min	extend
Cycle3: (1x)	Step 1	72°C	10:00 min	final extension
Cycle4: (1x)	Step 1	4°C	forever	

**Note:** The PCR condition may need optimization. Change the anneal temperature according to the Tm of your primers. Change the extension time according to the length of your PCR products (1 min for 1 kb, but not less than 1 min).

3. Prepare the DNA loading sample as instructed below by using PCR products amplified from pre-immune or antiserum-precipitated or input DNA. Also prepare the DNA marker.

<PCR from pre-immune or antiserum>

reagents	amount (µl)
DNA	18
10X ChIP loading dye	2
Total	20

#### <PCR from input DNA>

reagents	amount (µl)
DNA	4
10X ChIP loading dye	2
1X TBE	14
total	20

#### <Invitrogen 1kb DNA marker>

	amount (µl)
DNA	2
10X ChIP loading dye	2
1X TBE	16
Total	20

4. Run the prepared DNA samples on a 8% native polyacrylamide gel (with 20 wells) in 1X TBE at 180 V until bromophenol blue migrates to the bottom of the gel (~2 hr and 40 min).

**Notes:** make an 8% polyacrylamide gel as instructed below: Add 16 ml 30% acrylamide (30:0.8 acrylamide: bisacrylamide), 37.6 ml d.d. water, 6 ml 10X TBE, 420  $\mu$ l 10% APS and 90  $\mu$ l TEMED in a 125 ml flask. Swirl to mix. Immediately pour the solution in a pre-assembled gel apparatus. Insert the 20 well comb. Wait for 20 minutes to solidify.

Acrylamide is neurotoxic. Always wear protection while handling it.

- 5. Stain the gel with  $0.5~\mu g/ml$  ethidium bromide (e.g. add  $100~\mu l$  5 mg/ml ethidium bromide in 1000~ml 1X TBE buffer) for 10~minutes. Destain the gel in deionized. water for 30~minutes.
- 6. Put the gel on a UV box. Turn on the UV and observe the DNA bands (wear protection). Take a photo of the gel for your record. Save the digital file.