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

2011

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

_____

_____

_____

_____

_____

Chair

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

LIST OF FIGURES

General Introduction

Chapter 1

Chapter 2

Chapter 3

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)

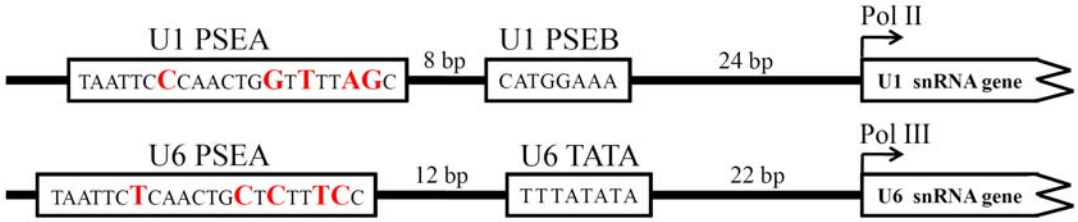


Figure G.1. Conserved structure of *Drosophila* snRNA gene promoters transcribed by Pol II and Pol III. The promoter sequences of the specific U1 and U6 genes used in our studies are shown. The U1 and U6 PSEA sequences differ at only 5 of 21 nucleotide positions (in red).

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.

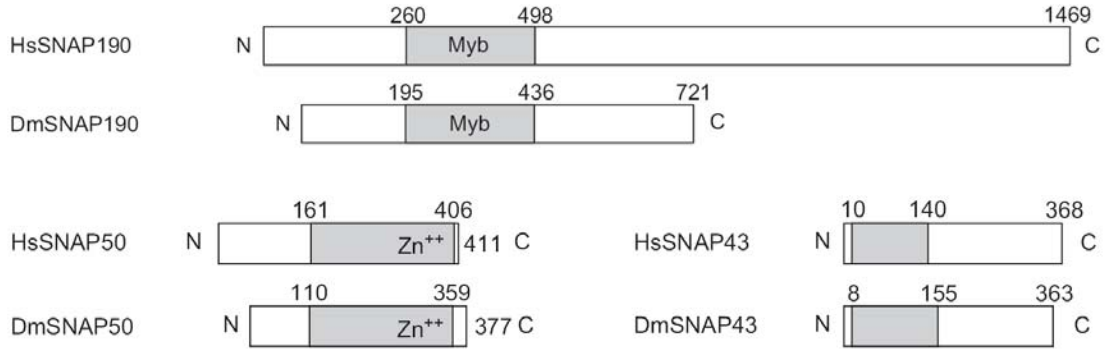


Figure G.2. Comparison of *D. melanogaster* (Dm) and human (Hs) SNAP subunits. The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous protein pairs. The numbers above the rectangles designate the amino acid positions 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^{++} 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.

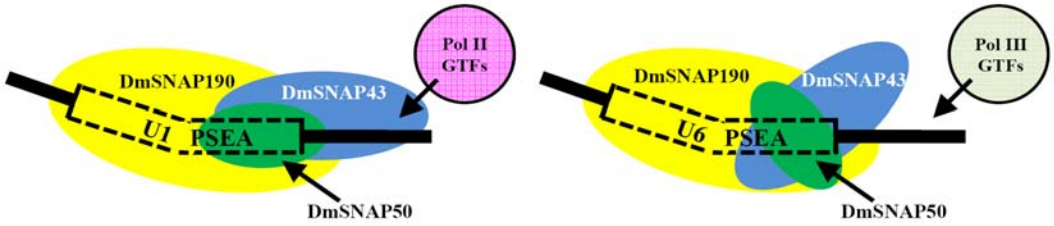


Figure G.3. Working model for RNA polymerase specificity at snRNA promoters. Each subunit of DmSNAPc is drawn schematically to indicate the region of DNA with which it interacts based upon photo-cross-linking data. We propose that the different conformations of DmSNAPc recruit different sets of general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

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 34, pp. 22568–22579, August 21, 2009

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[‡§], **Mitchell Titus**[‡¶1], **Shu-Chi Chiang**[‡§], **and William E. Stumph**[‡¶2]

From the [‡]*Molecular Biology Institute and the Departments of* [§]*Biology and* [¶]*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).[4] 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.

* This work was supported by National Science Foundation Grants MCB-0641350 and MCB-0842770 and the California Metabolic Research Foundation.

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

[2] To whom correspondence should be addressed. Tel.: 619-594-5575; Fax: 619-594-4634; E-mail: wstumph@sciences.sdsu.edu.

[3] 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*.

[4] K.-H. Hung, M. Titus, S.-C. Chiang, and W. E. Stumph, unpublished observations.

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 MluI 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^+, 5.5 mM $MgCl_2$, 5.0 mM dithiothreitol, 10% glycerol, and 0.1 mM EDTA (pH 7.6)). For high-salt immunopurifications, 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× FLAG peptide (Sigma catalog number F4799) at a concentration of 200 μg/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 $MgCl_2$, 10 μM $ZnCl_2$, 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′-GTGTGGCATACCTTATAGGGGTGCT-3′) and reverse primer (5′-GCTTTTCGATGCTCGGCAGCAG-3′) amplify the promoter region of the

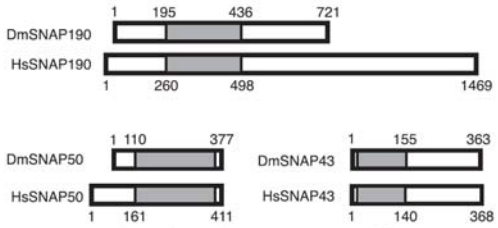


FIGURE 1. **Comparison of *D. melanogaster* (*Dm*) and human (*Hs*) SNAP subunits.** The *rectangles* indicate the relative lengths of the proteins, and the *shaded areas* indicate the evolutionarily conserved regions of the orthologous protein pairs. The *numbers above* and *below* the *rectangles* 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 HsSNAP190 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. Full-length and N- and C-terminal DmSNAP190 truncation constructs under control of the copper-inducible metallothionein promoter were prepared with FLAG-His_6 tags at the C terminus or with His_6-FLAG tags at the N terminus (Fig. 2*A*). 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. 2*B* shows that each of the DmSNAP190 truncation mutants was expressed in the corresponding stably transfected cells (*lanes 1–7*, *top panels*). Fig. 2*B* 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. 2*C*) or high-salt (Fig. 2*D*) 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. 3*A*, *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-

[5] K.-H. Hung, M. Titus, S.-C. Chiang, and W. E. Stumph, unpublished data.

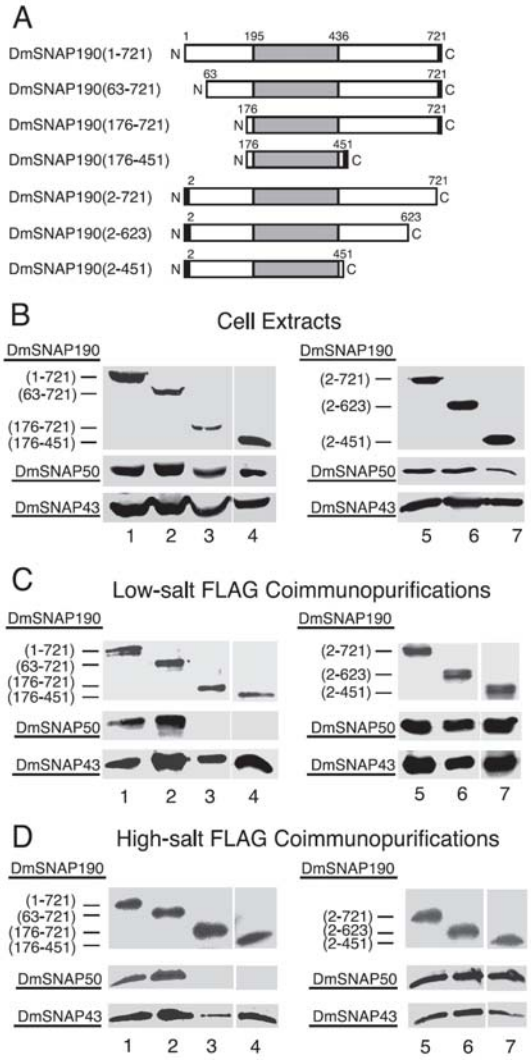


FIGURE 2. **Domains of DmSNAP190 involved in assembly *in vivo* with DmSNAP50 and DmSNAP43.** *A,* schematic representation of full-length and truncated DmSNAP190 constructs co-transfected into S2 cells together with constructs expressing full-length untagged DmSNAP50 and DmSNAP43. The *shaded* area represents the 4.5-Myb-repeat domain conserved between humans and fruit flies. 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 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 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 *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).

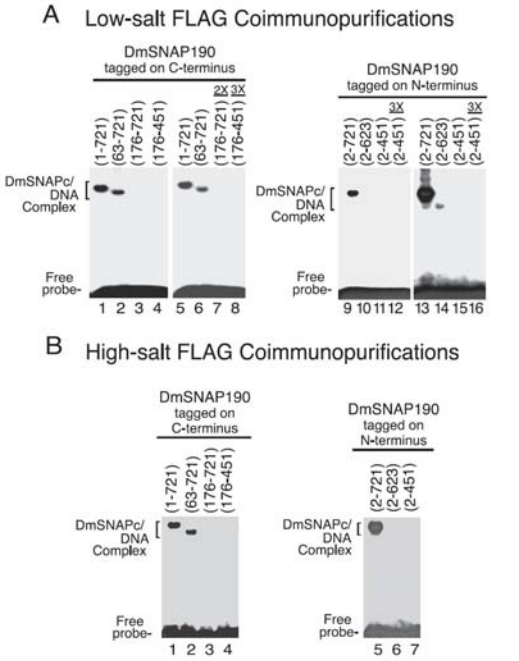


FIGURE 3. **Domains of DmSNAP190 required for sequence-specific DNA binding by DmSNAPc.** *A,* DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 2*C*) were used for DNA mobility shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP190 constructs 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. 2*C*, 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. 2*D*.

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*).

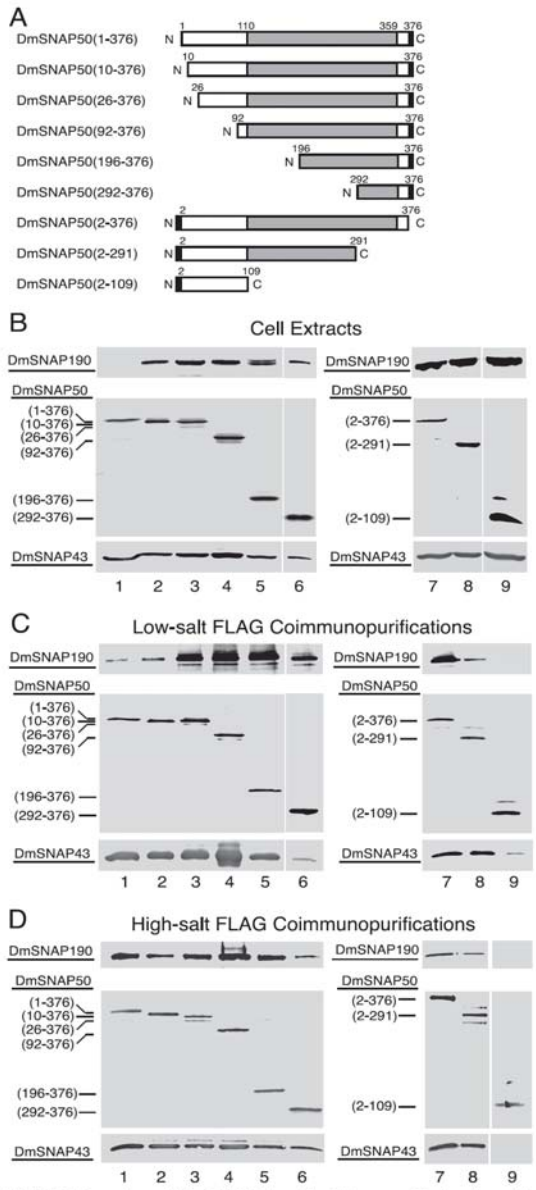


FIGURE 4. **Domains of DmSNAP50 involved in assembly *in vivo* with DmSNAP190 and DmSNAP43.** *A*, schematic representation of full-length and truncated DmSNAP50 constructs co-transfected into S2 cells together with constructs expressing full-length untagged DmSNAP190 and DmSNAP43. The *shaded* area 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 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 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 immunoblotting. 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.

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 HsSNAP50 (Fig. 1) (14). We concentrated most of our efforts on truncations from the N terminus, but two C-terminal truncations were also prepared (Fig. 4*A*). N-terminal truncations had FLAG-His_6 tags at the C terminus, whereas the C-terminal truncations had His_6-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. 4*B*, *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. 4*B*, *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. 4*C*) or high-salt (Fig. 4*D*) 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,

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. 4*C*, *top panels*, *lanes 3–6 versus lanes 1* and *2*). However, when the co-immunopurifications were carried out in higher salt conditions (Fig. 4*D*), 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

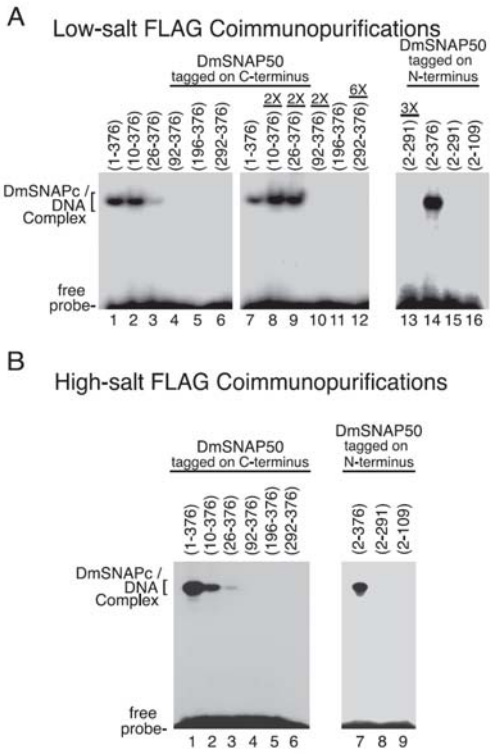


FIGURE 5. **Domains of DmSNAP50 required for sequence-specific DNA binding by DmSNAPc.** *A*, DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 4*C*) were used for DNA mobility shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP50 constructs 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. 4*C*, 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. *Lanes 14–16* also contained normalized amounts 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. 4*D*.

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,

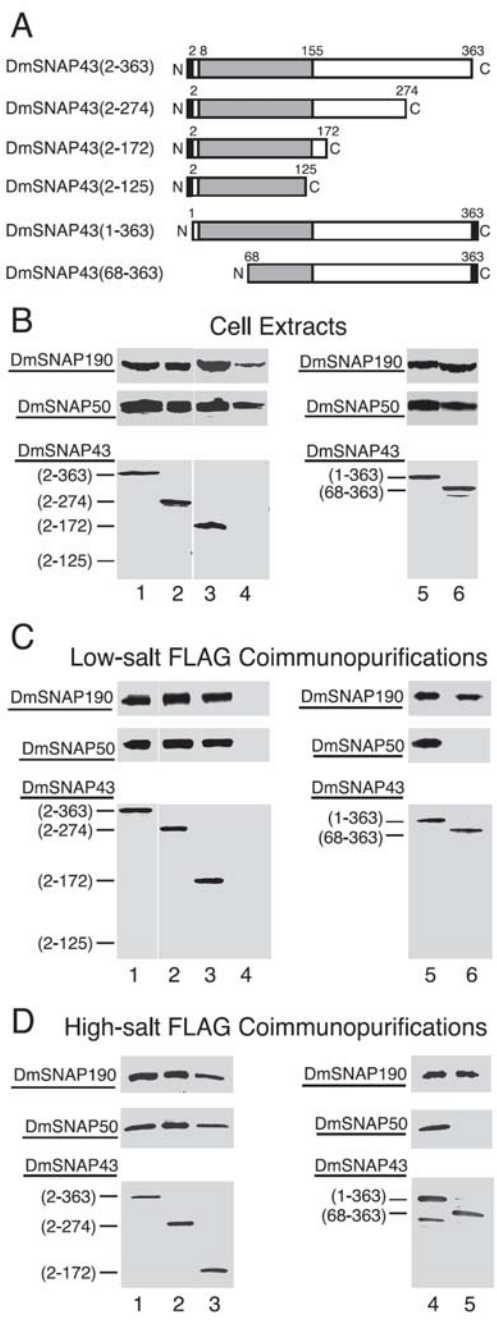


FIGURE 6. **Domains of DmSNAP43 involved in assembly *in vivo* with DmSNAP190 and DmSNAP50.** *A*, schematic representation of full-length and truncated DmSNAP43 constructs co-transfected into S2 cells together with constructs expressing full-length untagged DmSNAP190 and DmSNAP50. The *shaded* area 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* 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 immunoblotting. 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.

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. 6*A*). 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. 6*B* 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. 6*B* 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 detectable 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

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 DNA-binding 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. 8*A*. Polyclonal antibodies prepared against full-length DmSNAP43 were used

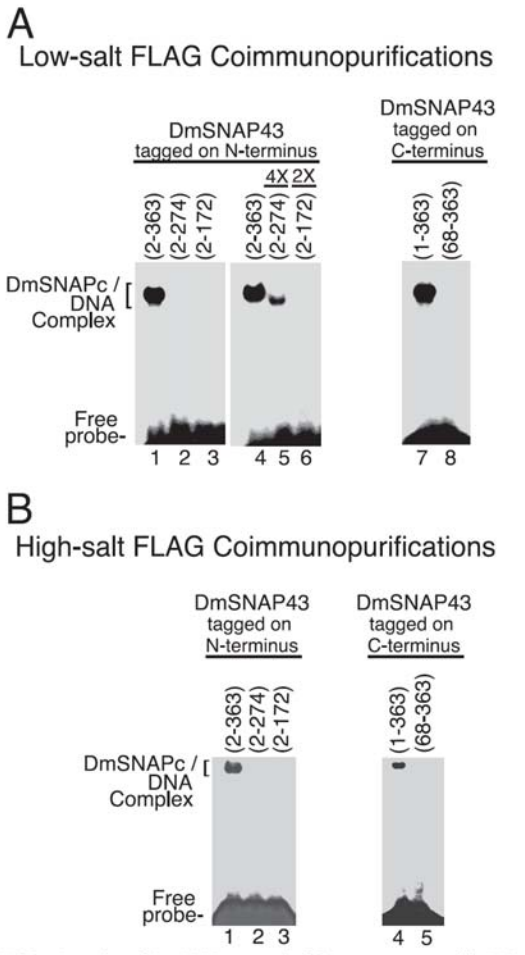


FIGURE 7. **Domains of DmSNAP43 required for sequence-specific DNA binding by DmSNAPc.** *A*, DmSNAP complexes FLAG affinity purified under low-salt conditions (Fig. 6*C*) were used for DNA mobility shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP43 constructs 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. 6*C*, 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.) *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. 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.

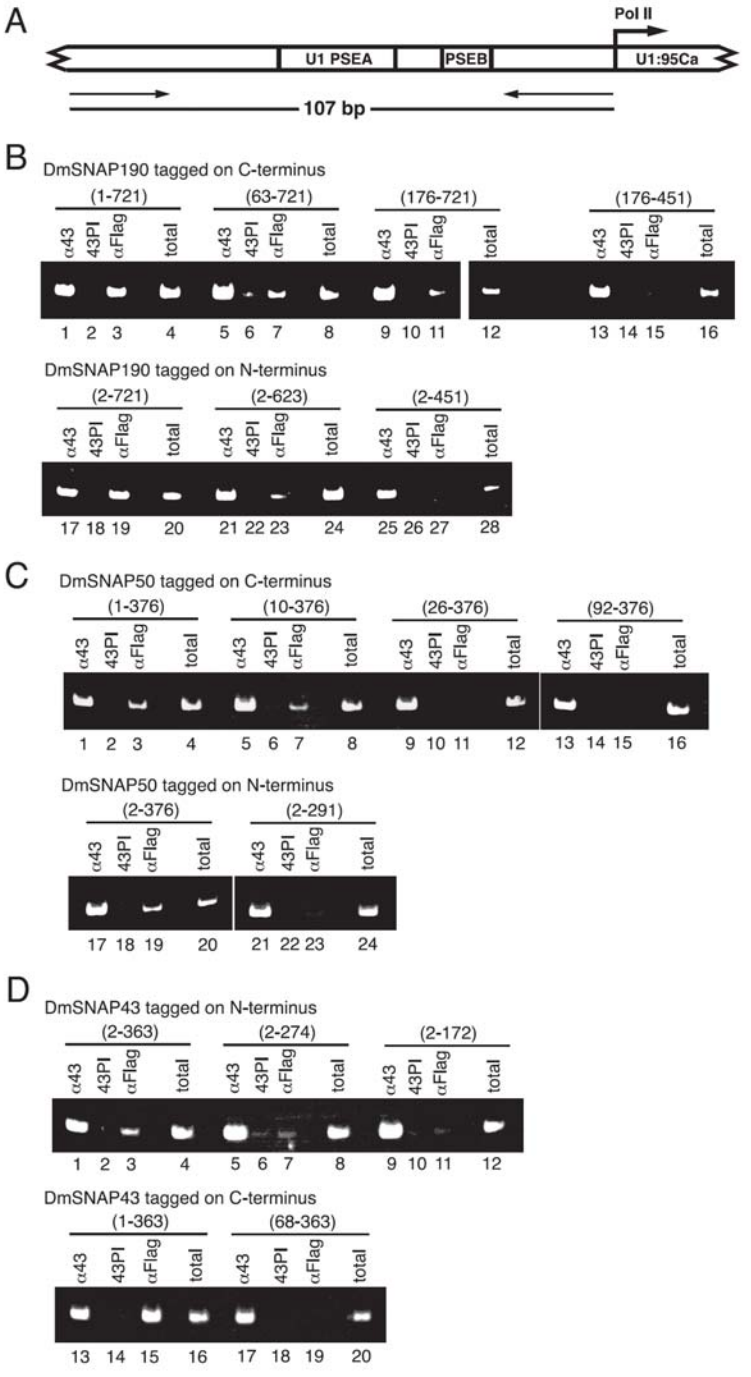
In the transfected cell lines that overexpressed tagged full-length DmSNAP190, the anti-FLAG antibodies efficiently pre-

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 progressively decreased the intensity of the PCR signal (Fig. 8*B*, *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. 8*C*, *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. 8*C*, *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. 8*D*, *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 domain of DmSNAP190-(176–451) can associate wi DmSNAP43. On the other hand, this conserved region DmSNAP190 is insufficient for its assembly with DmSNAP5



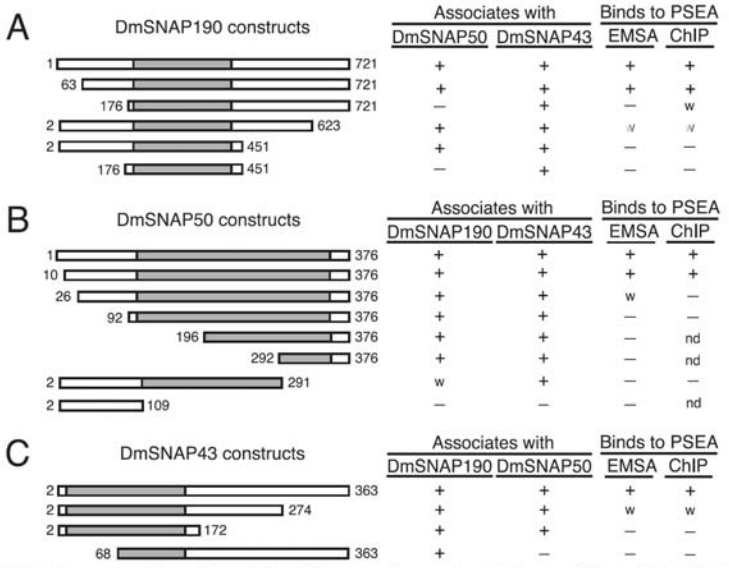


FIGURE 9. **Summary of the effects of DmSNAP truncations on DmSNAPc assembly and DNA-binding activity.** The various DmSNAP truncations analyzed are shown on the *left side* of the figure. The *columns on the right* indicate whether the specific truncation construct can associate with the other two DmSNAP 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).[4]

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 DNA-binding 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. 2*A*) with full-length untagged DmSNAP50 and DmSNAP43. Antibody against the FLAG epitope (*αFLAG*) was used to detect the presence of tagged DmSNAPc at the *U1:95Ca* promoter. Antibody against DmSNAP43 (*α43*), which recognizes both endogenous and tagged DmSNAPc was used as a positive control. DmSNAP43 preimmune serum (*43PI*) 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. 4*A*) with full-length untagged DmSNAP190 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 DmSNAP50 constructs tagged at the N terminus (C-terminal truncations). *D*, ChIP from S2 cell lines that co-express full-length or truncated FLAG-tagged DmSNAP43 constructs (as shown in Fig. 6*A*) with full-length untagged DmSNAP190 and DmSNAP50. Antibodies and other conditions were the same as in *B*. The *upper panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs tagged at the N terminus (C-terminal truncations). The *lower panel* shows ChIP results from S2 cell lines expressing DmSNAP43 constructs tagged at the C terminus (N-terminal truncations).

DmSNAPc Domains for Subunit Assembly and DNA Binding

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 HsSNAP50 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 HsSNAP190. 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

1. Murphy, S., Yoon, J. B., Gerster, T., and Roeder, R. G. (1992) *Mol. Cell. Biol.* **12,** 3247–3261
2. Sadowski, C. L., Henry, R. W., Lobo, S. M., and Hernandez, N. (1993) *Genes Dev.* **7,** 1535–1548
3. Yoon, J. B., Murphy, S., Bai, L., Wang, Z., and Roeder, R. G. (1995) *Mol. Cell. Biol.* **15,** 2019–2027
4. Henry, R. W., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1995) *Nature* **374,** 653–656
5. Waldschmidt, R., Wanandi, I., and Seifart, K. H. (1991) *EMBO J.* **10,** 2595–2603
6. 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
8. Sadowski, C. L., Henry, R. W., Kobayashi, R., and Hernandez, N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 4289–4293
9. Henry, R. W., Ma, B., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1996) *EMBO J.* **15,** 7129–7136
10. Bai, L., Wang, Z., Yoon, J. B., and Roeder, R. G. (1996) *Mol. Cell. Biol.* **16,** 5419–5426
11. 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
12. 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
14. Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (2004) *Mol. Cell. Biol.* **24,** 1897–1906
15. 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
17. Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A., and Bellofatto, V. (2005) *Mol. Cell. Biol.* **25,** 7314–7322
18. Schimanski, B., Nguyen, T. N., and Günzl, A. (2005) *Mol. Cell. Biol.* **25,** 7303–7313
19. Hernandez, G., Jr., Valafar, F., and Stumph, W. E. (2007) *Nucleic Acids Res.* **35,** 21–34
20. 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
22. Hinkley, C. S., Hirsch, H. A., Gu, L., LaMere, B., and Henry, R. W. (2003) *J. Biol. Chem.* **278,** 18649–18657
23. 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
25. 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
28. 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)

Critical Reviews in Biochemistry and Molecular Biology, 2011; 46(1): 11–26

informa
healthcare

REVIEW ARTICLE

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

Ko-Hsuan Hung[1], and William E. Stumph[2]

[1]*Department of Biology, and* [2]*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, $tRNA^{Sec}$, 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

(Received 01 July 2010; revised 19 August 2010; accepted 20 August 2010)

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 systems.

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 non-conserved 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 8 bp 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).

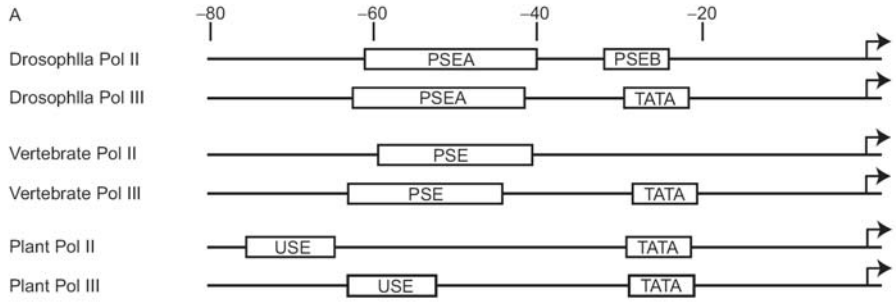


Figure 1. Promoter structure of snRNA genes. (A) Schematic representation of conserved promoter elements in the 5′-flanking DNA of snRNA genes transcribed either by Pol II or by Pol III in *Drosophila*, vertebrates, and plants. The numbers at the top indicate the distance in base pairs upstream of the 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.

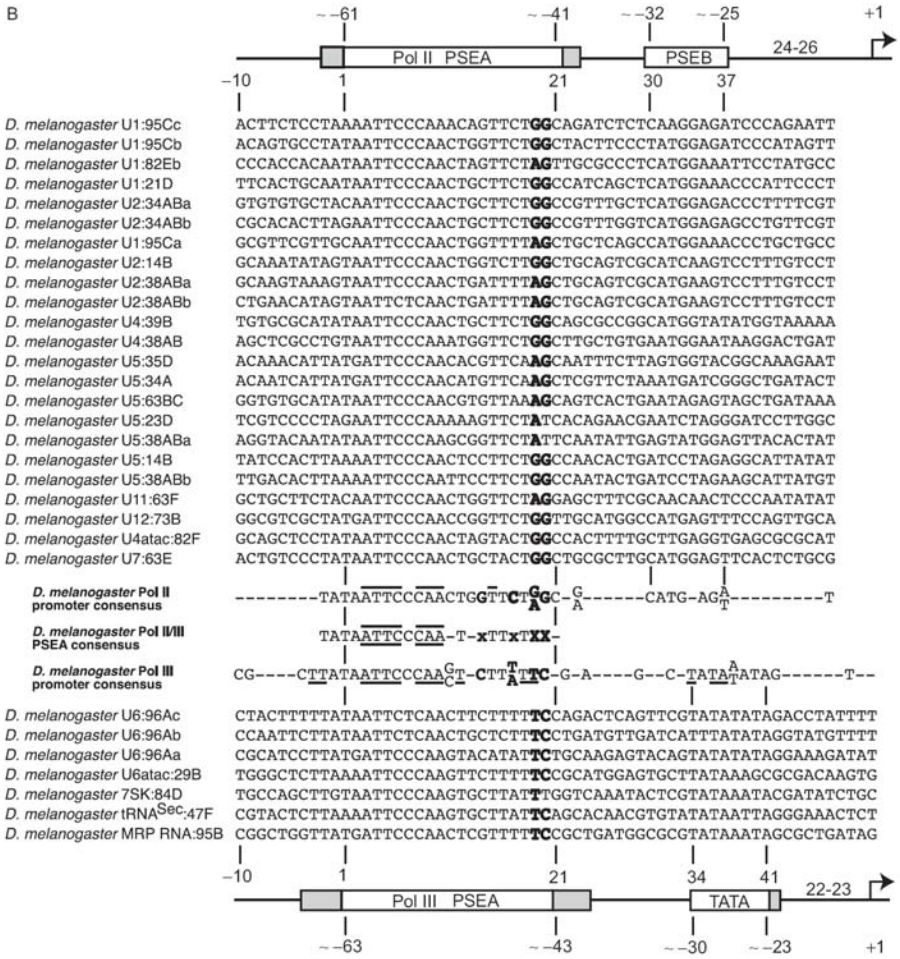
Figure 1. continued on next page

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 PSE-binding 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 III-transcribed snRNA genes was shown to be dependent upon SNAPc. Human SNAPc (HsSNAPc) contains five distinct polypeptide chains (HsSNAP190 or PTFα, HsSNAP50 or PTFβ, HsSNAP45 or PTFδ, HsSNAP43 or PTFγ, 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 enhancer-binding 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.

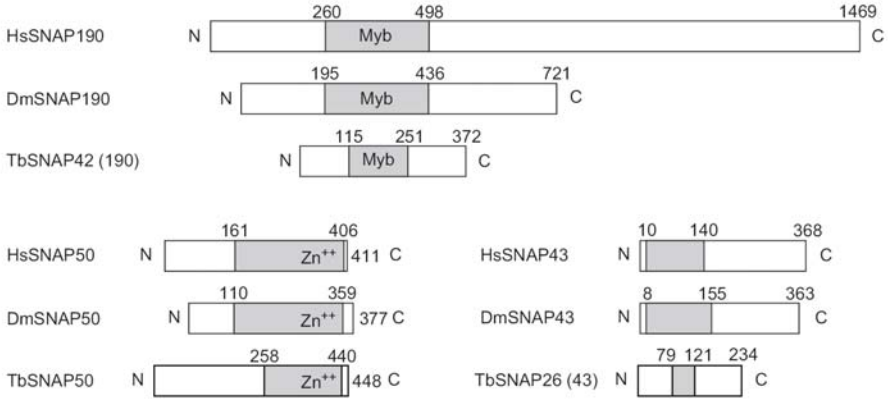


Figure 2. Comparison of *D. melanogaster* (Dm), human (Hs), and *Trypanosoma brucei* (Tb) SNAPc subunits. The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous proteins. The numbers above and below the rectangles 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 TbSNAP26.

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,

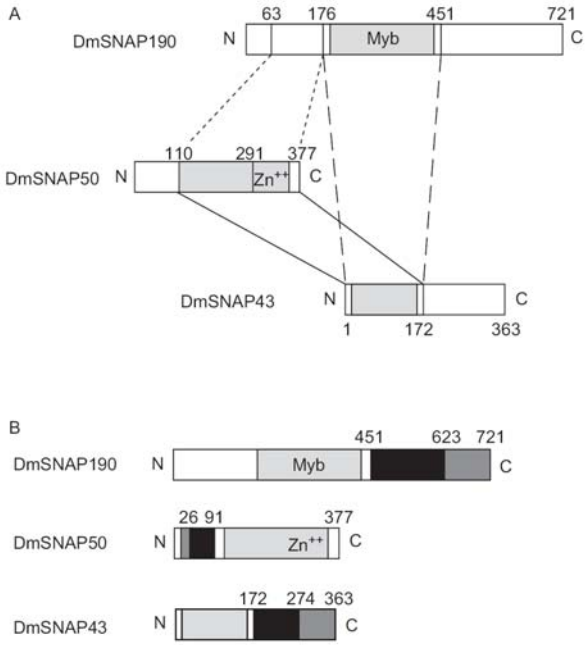


Figure 3. DmSNAPc subunit domains required for complex assembly and DNA binding. The lightly shaded area of each subunit represents the evolutionarily conserved region as shown in Figure 2. The numbers above and below the rectangles indicate the amino acid positions in the subunits where truncations 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 DmSNAP50. 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 non-conserved 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 non-conserved N-terminal region of HsSNAP190 and a non-conserved 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 DNA-binding 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 site-specific 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). Odd-numbered 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

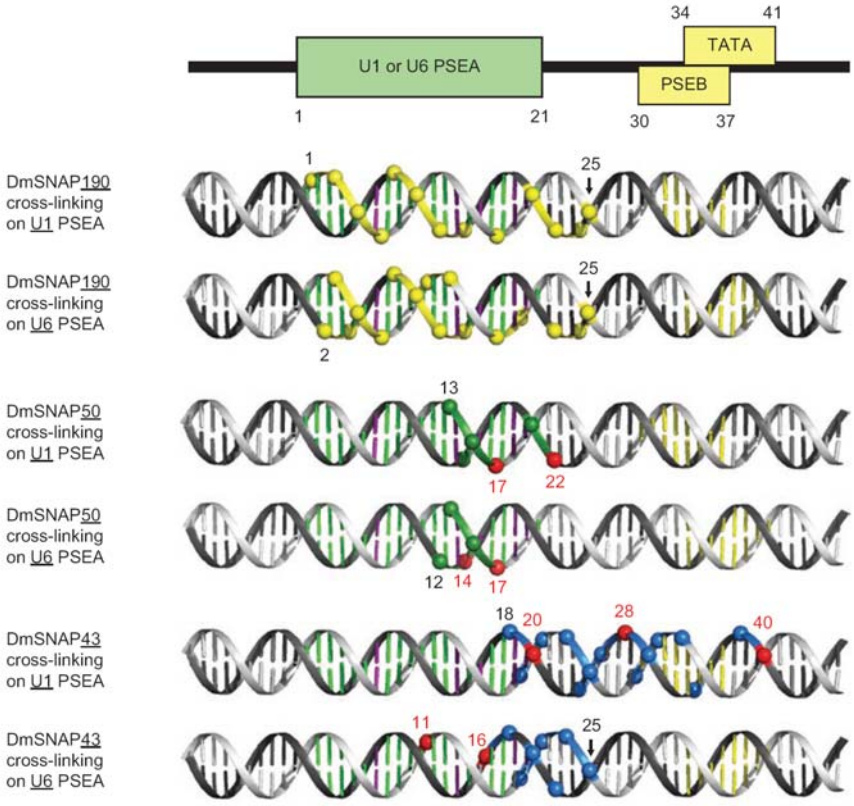


Figure 4. Contact points between snRNA gene promoter sequences and DmSNAPc. Site-specific protein–DNA photo-cross-linking data of DmSNAPc to DNA probes that contained either U1 or U6 PSEAs are shown projected onto B-form DNA (Wang and Stumph, 1998; Li *et al.*, 2004). (For experimental details, see below.) 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 cross-linking 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 ^{32}P 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 photo-cross-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,

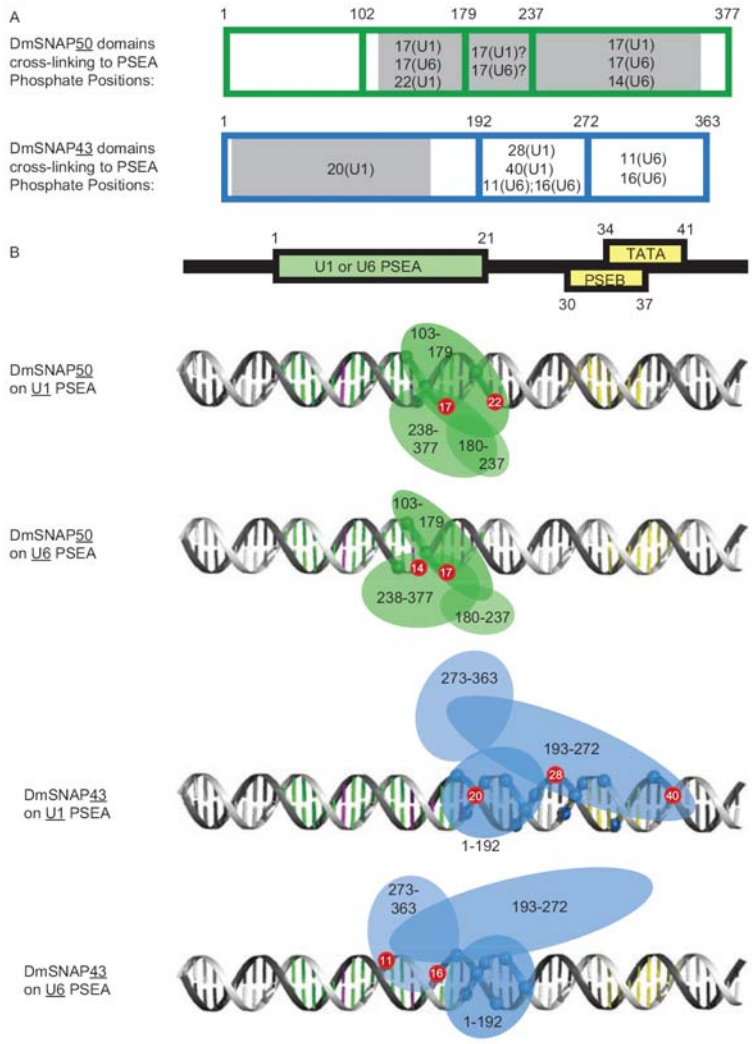


Figure 5. Domains of DmSNAP50 and DmSNAP43 that cross-link to specific phosphate positions on U1 or U6 DNA probes. (A) The rectangles at the top represent the linear amino acid sequences of the DmSNAP50 (green) and DmSNAP43 (blue) subunits. The shaded areas represent the evolutionarily conserved region 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 protein–protein 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

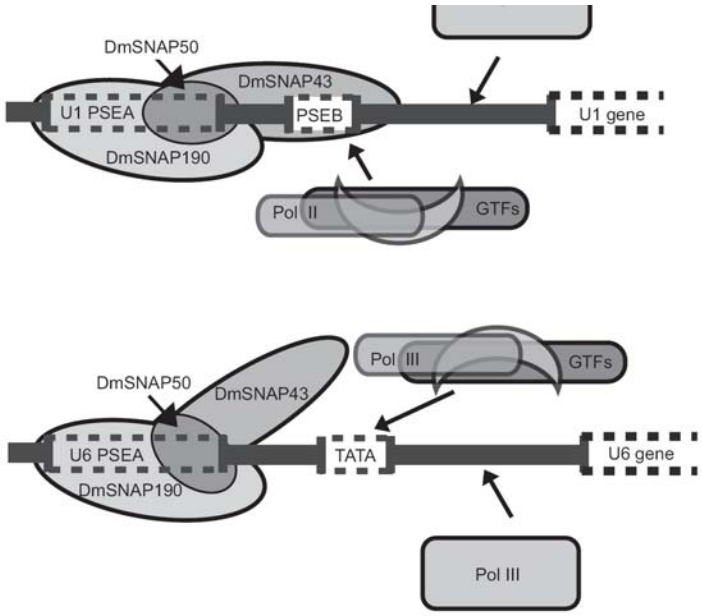


Figure 6. A working model illustrating how differential interactions of DmSNAPc with U1 or U6 PSEAs may lead to RNA polymerase specificity at *D. melanogaster* snRNA gene promoters. Minor sequence differences between U1 and U6 PSEAs are believed to act as allosteric effectors of DmSNAPc conformation, 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 7SK) 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 snRNA-activating 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: 2719–2725.

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. J 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 spliced-leader 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. Characterization of a Drosophila proximal-sequence-element-binding 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.
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 C-terminal 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.

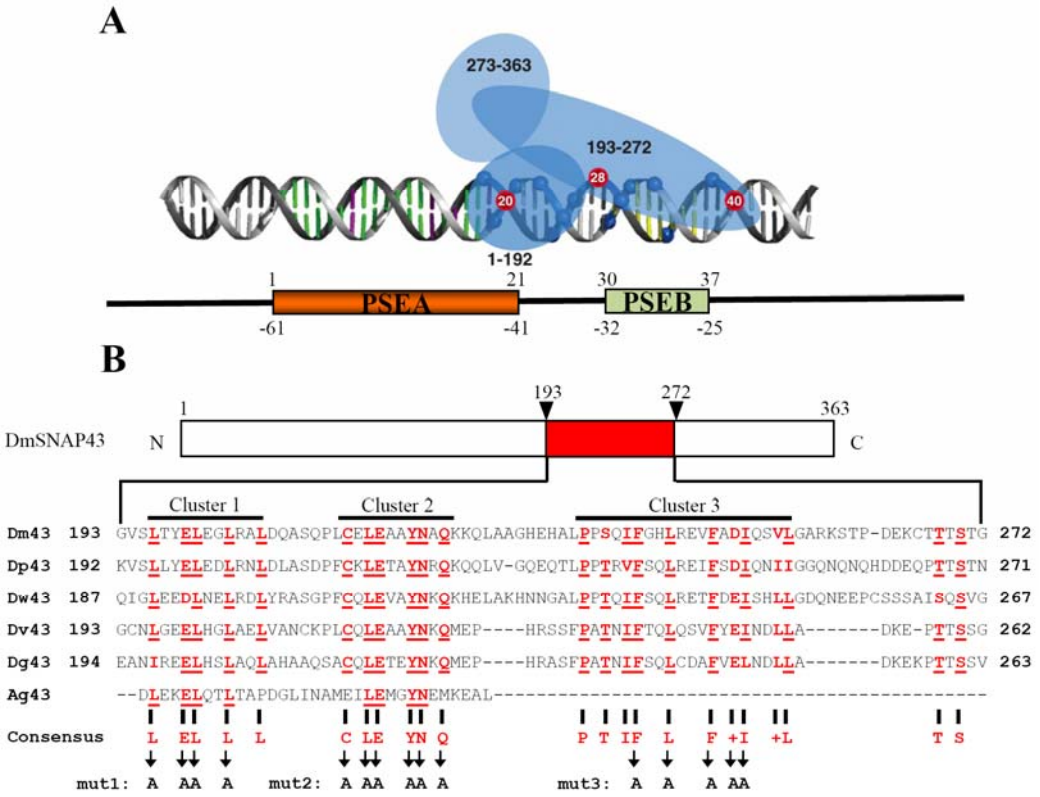


Figure 3.1. Identification of evolutionarily conserved amino acid residues within a DNA-binding domain of DmSNAP43. (A) Schematic representation of domains of DmSNAP43 that contact phosphate positions 20, 28, and 40 when DmSNAPc binds to a U1 PSEA. From previous work, a domain bounded by residues 193 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_2PO_4 (pH 8.0), 0.5 M NaCl, 20 mM imidazole and then once in HEMG-100 buffer (100 mM KCl, 25 mM HEPES K^+ (pH 7.6), 12.5 mM $MgCl_2$, 10 μM $ZnCl_2$, 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^+ (pH 7.6), 10 mM $MgCl_2$, 8 μM $ZnCl_2$, 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 ^{o}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 (http://flybase.org) SNAP43 protein sequences from five different Drosophila species and aligned these amino acid sequences by utilizing constraint based protein multiple alignment tool (COBALT) (http://www.ncbi.nlm.nih.gov/tools/cobalt/). 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).

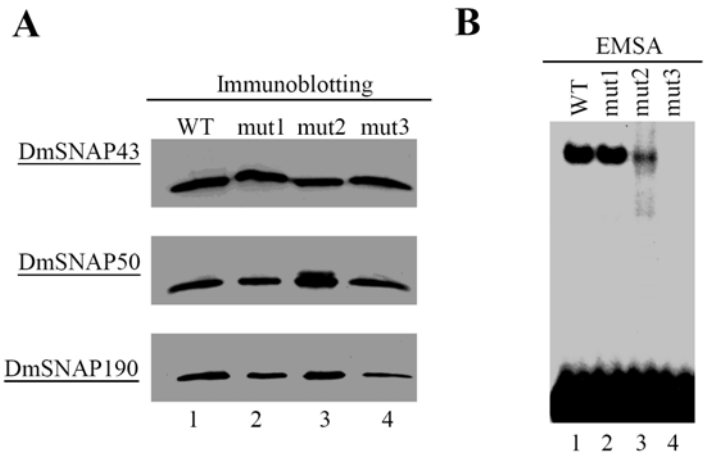


Figure 3.2. Mutations within two clusters of DmSNAP43 conserved residues have no affect on DmSNAP complex formation but result in reduced DmSNAPc DNA-binding activity. (A) Immunoblots demonstrating that DmSNAP50 and DmSNAP190 co-purified with tagged DmSNAP43 wild type and mutant constructs following 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.

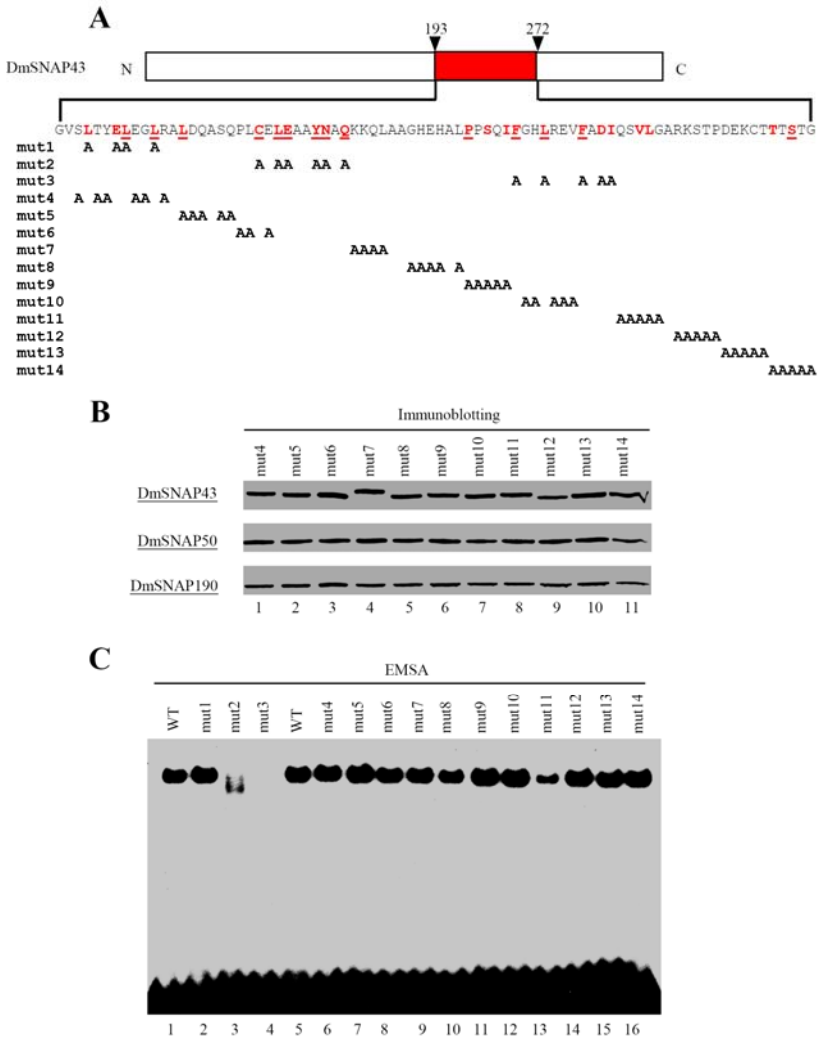


Figure 3.3. Alanine scanning mutagenesis within the DNA-binding domain of DmSNAP43 between residues 193 and 272. (A) Locations of alanine substitutions in fourteen mutant DmSNAP43 constructs are shown. Residues in red in the upper sequence were very strongly conserved in the five *Drosophila* species 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**.

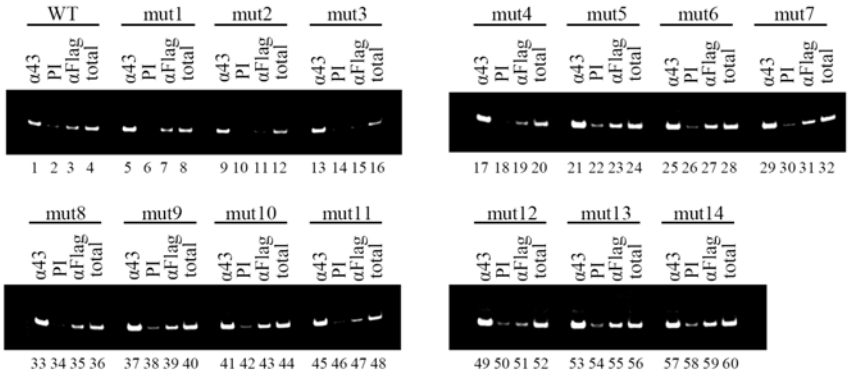


Figure 3.4. ChIPs from stably-transfected S2 cells that over-express FLAG-tagged WT or mutant DmSNAP43 constructs that contain alanine substitutions. Anti-FLAG antibodies were used for ChIP to examine the relative binding of the mutant constructs to the U1 promoter *in vivo* (lanes labeled αFLAG). Positive controls utilized anti-DmSNAP43 antibodies prepared against full-length DmSNAP43 (α43). 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 photo-cross-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 nickel-chelate 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#14 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 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   H  H  H   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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-5'HisFlag-Dm43 (2-274) 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   H  H  H   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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TAGAAGGGCA ATTCTGCAGA TATCCAGCAC AGTGGCGGCC GCTCGAGTCT AGAGGGCCCG
   > P  D  E   K  C  T   T  T  S  T   G  N  Q
1801 CGGTTCGAAG GTAAGCCTAT CCCTAACCCT CTCCTCGGTC TCGATTCTAC GCGTACCGGT CATCATCACC ATCACCATTG AGTTTAAACC CGCTGATCAG
1901 CCTCGACTGT GCCTTCTAAG ATCCAGACAT GATAAGATAC ATTGATGAGT TTGGACAAAC CACAACTAGA ATGCAGTGAA AAAAATGCTT TATTTGTGAA
2001 ATTTGTGATG CTATTGCTTT ATTTGTAACC ATTATAAGCT GCAATAAACA AGTTAACAAC AACAATTGCA TTCATTTTAT GTTTCAGGTT CAGGGGGAGG
2101 TGTGGGAGGT TTTTTAAAGC AAGTAAAACC TCTACAAATG TGGTATGGCT GATTATGATC AGTCGACCTG CAGGCATGCA AGCTTGGCGT AATCATGGTC
2201 ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA ATGAGTGAGC
2301 TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG
2401 GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT
2501 ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT
2601 TTCCATAGGC TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC
2701 CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC
2801 ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC
2901 TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG
3001 AGTTCTTGAA GTGGTGGCCT AACTACGGCT ACACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG
3101 CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG
3201 ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTCACCTAG ATCCTTTTAA
3301 ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG
3401 TCTATTTCGT TCATCCATAG TTGCCTGACT CCCCGTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA
3501 GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT
3601 CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC
3701 GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG
3801 ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG
3901 TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG
4001 CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT
4101 GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC
4201 GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA
4301 AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT
4401 ATCACGAGGC CCTTTCGT
```

pMT-5'HisFlag-Dm43 (2-172) 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   H  H  H   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
   >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
   >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG ACCGTGTGGA GACCTAGCAA GGGCAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT
   >  D  N  L   V  D  Y   D  R  V  E   T
1501 CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG
1601 ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG
1701 TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG
1801 GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC
1901 TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT
2001 CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG
2101 CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT
2201 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA
2301 TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA
2401 AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT
2501 GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG
2601 TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC
2701 TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT
2801 GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT
2901 TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA
3001 AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT
3101 TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC
3201 ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT
3301 AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG
3401 GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT
3501 TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT
3601 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA
3701 CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC
3801 CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA
3901 TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
4001 AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC
4101 GAGGCCCTTT 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   H  H  H   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
   >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
   >  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
   > 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
1301 GGCAATTCTG CAGATATCCA GCACAGTGGC GGCCGCTCGA GTCTAGAGGG CCCGCGGTTC GAAGGTAAGC CTATCCCTAA CCCTCTCCTC GGTCTCGATT
1401 CTACGCGTAC CGGTCATCAT CACCATCACC ATTGAGTTTA AACCCGCTGA TCAGCCTCGA CTGTGCCTTC TAAGATCCAG ACATGATAAG ATACATTGAT
1501 GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA
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
2101 GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT
2201 CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC
2301 CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG
2401 GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG
2501 CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT
2601 ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT
2701 GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT
2801 AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT ATGAGTAAAC
2901 TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA
3001 ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG
3101 CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA
3201 TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA
3301 GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA
3401 TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG
3501 GCGACCGAGT TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA
3601 AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG
3701 GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT ATTATTGAAG
3801 CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA
3901 CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC GT
```

pMT-Dm43 full noSTOP-FlagMycHis

```
   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 TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTTCGCGTA ATTAAGATTC
 901 AGAATGGAGC TGAATATCTT TGACGACTGC TGGGAGCTGG TGCAACGATT TCAGCGATTG GTTAATGATG GCGAAAACTG CGAGTTCGAG GTGTTCTGCC
       > M  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  F  E   V  F  C
1001 GGTGCTGGCG AGAACTGCAG CTGCAGCACC TTTTCACTGC CCAGACGAAC CACACAGAGG TGATAGCCAC CACACTGGCG GCCCTGCATG TGGCCAAGCG
    >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  H   V  A  K  R
1101 ACTGTCGTGC TCCCGACGCA CCACCGGGGA CGTTTTCCCG GCATCTCGCG CTCAAAGGAT CGGAGGTTTC TTTCTGCTCT ACGTAATCTA CTACAAGCAG
    >  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  I  Y   Y  K  Q
1201 CCCACGCACA ACTTTATTAA GATCGAGGTC TCACCGCGCA CTTGGCAAGA ACTAACAGAC TACGCTCTAG ATCTGCGCAA GGATAGTCCG GAGCGGAAGG
    > 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  S  P   E  R  K
1301 ACACTCATCA GATCGCCTAC ATGCTGTGGC GCCTGACCCA GGAGCAGGCC TTCCGCTTCA CCGCGCTCGA CTATTGCCAG GGGTTGGACA ATCTGGTGGA
    >D  T  H  Q   I  A  Y   M  L  W   R  L  T  Q   E  Q  A   F  R  F   T  A  L  D   Y  C  Q   G  L  D   N  L  V  D
1401 CTACGACCGT GTGGAGACCG TAGCGGGTGC CAAGGAACAG AGGCAGAGTG CCTTGATGCA GAAGCAACAG CGTGCGAACG GCGTCAGTCT CACATACGAA
    >  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  L   T  Y  E
1501 CTGGAGGGTC TGCGAGCACT GGACCAGGCA AGCCAGCCAT TGTGTGAACT GGAAGCGGCA TACAATGCCC AAAAGAAGCA ATTGGCGGCT GGTCATGAGC
    > L  E  G   L  R  A  L   D  Q  A   S  Q  P   L  C  E  L   E  A  A   Y  N  A   Q  K  K  Q   L  A  A   G  H  E
1601 ACGCTTTACC GCCCTCTCAA ATATTCGGCC 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 TCTACAGGCA ACCAGTTGGA 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 TGCAGCTGGA GGTCAACGAG ACTTATCAAC GCCGCATGTC CTCGGCCACC GTTTTCCAGA GGGAACTTCC AGAAGACGTG CAGCAAGAGT
    > 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  P   E  D  V   Q  Q  E
1901 ATGAGATGAT TGAGTTTAGT GACGACGAGG AAATGGAAGT GGGTGAAAGC GAGGAGGTCA CGGAAGAAGA ACTCAAAGCT ATTTTGGATA CCAAGGGCAA
    >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   T  K  G  N
2001 TTCTGCAGAT ATCCAGCACA GTGGCGGCCG CTCGAGTCTA GAGGGCCCGC GGTTCGACTA CAAGGATGAC GATGACAAGG GCGAGCAGAA GCTGATCTCC
    >  S  A  D   I  Q  H   S  G  G  R   S  S  L   E  G  P   R  F  D  Y   K  D  D   D  D  K   G  E  Q  K   L  I  S
2101 GAGGAAGATC TGACGCGTAC CGGTCATCAT CACCATCACC ATTGAGTTTA AACCCGCTGA TCAGCCTCGA CTGTGCCTTC TAAGATCCAG ACATGATAAG
    > E  E  D   L  T  R  T   G  H  H   H  H  H   H
2201 ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA
2301 AGCTGCAATA AACAAGTTAA CAACAACAAT TGCATTCATT TTATGTTTCA GGTTCAGGGG GAGGTGTGGG AGGTTTTTTA AAGCAAGTAA AACCTCTACA
2401 AATGTGGTAT GGCTGATTAT GATCAGTCGA CCTGCAGGCA TGCAAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC
2501 AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT
2601 TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT TCCTCGCTCA
2701 CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA
2801 GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA
2901 AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC
3001 TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG
3101 CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA
3201 TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA
3301 GAAGGACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG
3401 TGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA
3501 AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT
3601 ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT
3701 CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA
3801 AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT
3901 CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG
4001 ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA
4101 CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT
4201 AGTGTATGCG GCGACCGAGT TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC
4301 TTCGGGGCGA AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC
4401 AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT
4501 ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG
4601 AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC GT
```

pMT-Dm43 (68-363) noSTOP-FlagMycHis

```
   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 TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTATTCAGA ATGTCGTGCT
                                                                                                    > M  S  C
 901 CCCGACGCAC CACCGGGGAC GTTTTCCCGG CATCTCGCGC TCAAAGGATC GGAGGTTTCT TTCTGCTCTA CGTAATCTAC TACAAGCAGC CCACGCACAA
    >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  I  Y   Y  K  Q   P  T  H  N
1001 CTTTATTAAG ATCGAGGTCT CACCGCGCAC TTGGCAAGAA CTAACAGACT ACGCTCTAGA TCTGCGCAAG GATAGTCCGG AGCGGAAGGA CACTCATCAG
    > F  I  K   I  E  V   S  P  R  T   W  Q  E   L  T  D   Y  A  L  D   L  R  K   D  S  P   E  R  K  D   T  H  Q
1101 ATCGCCTACA TGCTGTGGCG CCTGACCCAG GAGCAGGCCT TCCGCTTCAC CGCGCTCGAC TATTGCCAGG GGTTGGACAA TCTGGTGGAC TACGACCGTG
    > I  A  Y   M  L  W  R   L  T  Q   E  Q  A   F  R  F  T   A  L  D   Y  C  Q   G  L  D  N   L  V  D   Y  D  R
1201 TGGAGACCGT AGCGGGTGCC AAGGAACAGA GGCAGAGTGC CTTGATGCAG AAGCAACAGC GTGCGAACGG CGTCAGTCTC ACATACGAAC TGGAGGGTCT
    >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  L   T  Y  E   L  E  G  L
1301 GCGAGCACTG GACCAGGCAA GCCAGCCATT GTGTGAACTG GAAGCGGCAT ACAATGCCCA AAAGAAGCAA TTGGCGGCTG GTCATGAGCA CGCTTTACCG
    > R  A  L   D  Q  A   S  Q  P  L   C  E  L   E  A  A   Y  N  A  Q   K  K  Q   L  A  A   G  H  E  H   A  L  P
1401 CCCTCTCAAA TATTCGGCCA TTTGCGAGAA GTCTTTGCCG ATATCCAAAG TGTTCTAGGA GCTAGAAAGA GTACTCCAGA TGAGAAATGC ACCACAACAT
    > 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   T  T  T
1501 CTACAGGCAA CCAGTTGGAA GTGCGCCAGA GGGTGCGGAA CAAGGCCATG TACGGCGTCG AGGAGCGGGA GCCGCAACAC CAGACGGATG AACTAGAAGT
    >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   E  L  E  V
1601 GCAGCTGGAG GTCAACGAGA CTTATCAACG CCGCATGTCC TCGGCCACCG TTTTCCAGAG GGAACTTCCA GAAGACGTGC AGCAAGAGTA TGAGATGATT
    > Q  L  E   V  N  E   T  Y  Q  R   R  M  S   S  A  T   V  F  Q  R   E  L  P   E  D  V   Q  Q  E  Y   E  M  I
1701 GAGTTTAGTG ACGACGAGGA AATGGAAGTG GGTGAAAGCG AGGAGGTCAC GGAAGAAGAA CTCAAAGCTA TTTTGGATAC TAAGGGCAAT TCTGCAGATA
    > 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  T   K  G  N   S  A  D
1801 TCCAGCACAG TGGCGGCCGC TCGAGTCTAG AGGGCCCGCG GTTCGACTAC AAGGATGACG ATGACAAGGG CGAGCAGAAG CTGATCTCCG AGGAAGATCT
    >I  Q  H  S   G  G  R   S  S  L   E  G  P  R   F  D  Y   K  D  D   D  D  K  G   E  Q  K   L  I  S   E  E  D  L
1901 GACGCGTACC GGTCATCATC ACCATCACCA TTGAGTTTAA ACCCGCTGAT CAGCCTCGAC TGTGCCTTCT AAGATCCAGA CATGATAAGA TACATTGATG
    > T  R  T   G  H  H   H  H  H  H
2001 AGTTTGGACA AACCACAACT AGAATGCAGT GAAAAAAATG CTTTATTTGT GAAATTTGTG ATGCTATTGC TTTATTTGTA ACCATTATAA GCTGCAATAA
2101 ACAAGTTAAC AACAACAATT GCATTCATTT TATGTTTCAG GTTCAGGGGG AGGTGTGGGA GGTTTTTTAA AGCAAGTAAA ACCTCTACAA ATGTGGTATG
2201 GCTGATTATG ATCAGTCGAC CTGCAGGCAT GCAAGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA
2301 ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG
2401 AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG
2501 CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG
2601 CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC
2701 AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC
2801 GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG
2901 GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC
3001 AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA
3101 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG
3201 TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA
3301 AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
3401 TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA
3501 CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC
3601 CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT
3701 AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG
3801 TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT
3901 GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG
4001 CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA
4101 AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG
4201 GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC
4301 ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC
4401 CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG T
```

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 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   H  H  H   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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#1 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  H H H  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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S P E R  K D T  H Q I  A Y M L  W R L  T Q E  Q A F R  F T A  L D Y  C Q G L
1401 GGACAATCTG GTGGACTACG 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  S A L  M Q K  Q Q R A  N G V
1501 AGTGCCACAT ACGCAGCGGA GGGTGCGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S A T  Y A A E  G A R  A L D  Q A S Q  P L C  E L E  A A Y N  A Q K  K Q L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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 E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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 P E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#2 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGGCT GAAGCGGCAG CGGCAGCCGC TGCCGCAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  A   E  A  A   A  A  A  A   A  A  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#3 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATAGC CGGCCATGCG CGAGAAGTCG CTGCCGCTGC CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >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 AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#4 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  H H H  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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S P E R  K D T  H Q I  A Y M L  W R L  T Q E  Q A F R  F T A  L D Y  C Q G L
1401 GGACAATCTG GTGGACTACG 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  S A L  M Q K  Q Q R A  N G V
1501 GCTCTCGCAG CCGAACTGGC GGCTCTGGCA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > A L A  A E L A  A L A  A L D  Q A S Q  P L C  E L E  A A Y N  A Q K  K Q L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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 E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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 P E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#5 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
 901 GACGATGACA AGGGCACTAG TGAGCTGAAT ATCTTTGACG ACTGCTGGGA GCTGGTGCAA CGATTTCAGC GATTGGTTAA TGATGGCGAA AACTGCGAGT
1001 TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA GCCACCACAC TGGCGGCCCT
1101 GCATGTGGCC AAGCGACTGT CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA AGGATCGGAG GTTTCTTTCT GCTCTACGTA
1201 ATCTACTACA AGCAGCCCAC GCACAACTTT ATTAAGATCG AGGTCTCACC GCGCACTTGG CAAGAACTAA CAGACTACGC TCTAGATCTG CGCAAGGATA
1301 GTCCGGAGCG GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
                                          > M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCAGCGGCCG CGGCAGCCGC GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    > S  L  T   Y  E  L  E   G  L  R   A  A  A  A  A  A  A   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#6 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  H H H  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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S P E R  K D T  H Q I  A Y M L  W R L  T Q E  Q A F R  F T A  L D Y  C Q G L
1401 GGACAATCTG GTGGACTACG 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  S A L  M Q K  Q Q R A  N G V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GGCAGCCTGT GCCCTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S L T  Y E L E  G L R  A L D  Q A S Q  A A C  A L E  A A Y N  A Q K  K Q L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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 E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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 P E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#7 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  H H H  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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S P E R  K D T  H Q I  A Y M L  W R L  T Q E  Q A F R  F T A  L D Y  C Q G L
1401 GGACAATCTG GTGGACTACG 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  S A L  M Q K  Q Q R A  N G V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAGCG GCGGCAGCGG
   > S L T  Y E L E  G L R  A L D  Q A S Q  P L C  E L E  A A Y N  A Q  A  A A A
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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 E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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 P E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#8 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGCTGC TGCGGCCGCT GCACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >A  A  A  A   A  A  A   A  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 CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#9 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
 901 GACGATGACA AGGGCACTAG TGAGCTGAAT ATCTTTGACG ACTGCTGGGA GCTGGTGCAA CGATTTCAGC GATTGGTTAA TGATGGCGAA AACTGCGAGT
1001 TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA GCCACCACAC TGGCGGCCCT
1101 GCATGTGGCC AAGCGACTGT CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA AGGATCGGAG GTTTCTTTCT GCTCTACGTA
1201 ATCTACTACA AGCAGCCCAC GCACAACTTT ATTAAGATCG AGGTCTCACC GCGCACTTGG CAAGAACTAA CAGACTACGC TCTAGATCTG CGCAAGGATA
1301 GTCCGGAGCG GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
                                          > M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    > S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTAGGCGGCCG CTGCAGCATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >A  A  G  H   E  H  A   L  A  A  A   A  A  A  F   G  H  L   R  E  V   F  A  D  I   Q  S  V   L  G  A   R  K  S  T
1701 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#10 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGCCGCTTTG GCAGCAGCCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >A  A  G  H   E  H  A   L  P  P   S  Q  I  F   A  A  L   A  A  A   F  A  D  I   Q  S  V   L  G  A   R  K  S  T
1701 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#11 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  H H H  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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S P E R  K D T  H Q I  A Y M L  W R L  T Q E  Q A F R  F T A  L D Y  C Q G L
1401 GGACAATCTG GTGGACTACG 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  S A L  M Q K  Q Q R A  N G V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S L T  Y E L E  G L R  A L D  Q A S Q  P L C  E L E  A A Y N  A Q K  K Q L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CGCAGCTGCT GCAGCAGCTA GAAAGAGTAC
   >A A G H  E H A  L P P  S Q I F  G H L  R E V  F A D I  A A A  A A A  R K S T
1701 TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA 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 E P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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 P E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#12 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTG CAGCGGCTGC
    >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   A  A  A  A
1701 TGCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA GCGGGAGCCG
    >  A  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 CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#13 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   H  H  H   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
   >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
   > 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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
   >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
   > S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
   >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 TCCAGCTGCG GCAGCCGCCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA GCGGGAGCCG
   > P  A  A   A  A  A   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 CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
   > D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMT-HisFlag-DmSNAP43 mut#14 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   H  H  H   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
    >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
    >  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 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 GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT
    >S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E   Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L
1401 GGACAATCTG GTGGACTACG 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   S  A  L   M  Q  K   Q  Q  R  A   N  G  V
1501 AGTCTCACAT ACGAACTGGA GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  S  L  T   Y  E  L  E   G  L  R   A  L  D   Q  A  S  Q   P  L  C   E  L  E   A  A  Y  N   A  Q  K   K  Q  L
1601 CGGCTGGTCA TGAGCACGCT TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC
    >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 TCCAGATGAG AAATGCACCG CCGCAGCTGC AGCCAACCAG TTGGAAGTGC GCCAGAGGGT GCGGAACAAG GCCATGTACG GCGTCGAGGA GCGGGAGCCG
    >  P  D  E   K  C  T   A  A  A  A   N  Q   L  E  V   R  Q  R  V   R  N  K   A  M  Y   G  V  E  E   R  E  P
1801 CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT 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  P  E
1901 ACGTGCAGCA AGAGTATGAG 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
2001 GGATACTTGA GTTAGTTTAT AAAACTTTTA CATAATTAAA TAACTAGCAT TTTTGCGCGA TGTGATCTTG TTTATCTGAA GGGCAATTCT GCAGATATCC
    >  D  T
2101 AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA
2201 TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA
2301 CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG
2501 ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
2601 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT
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 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA
3501 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3701 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC 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 CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT
4101 TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC
4201 TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC
4301 CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG
4501 GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
4601 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA
4701 TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

Appendix B: List of Stably Transfected S2 Cell Lines

For establishment of all cell lines listed below, expression plasmids were co-transfected 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)	DmSNAP50 full	DmSNAP190 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

DmSNAP190 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

DmSNAP190 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 µl of 10x Wash Buffer to 4 mg of 3x FLAG peptide.

After the peptide is completely dissolved, add 640 µl of distilled water to the sample. Mix well and store aliquots of 75 µl 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 $MgCl_2$, 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 $MgCl_2$, 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 $CuSO_4$ (add 0.2497 g of $CuSO_4 \bullet 5H_2O$ 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 $CuSO_4$ 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 ~2 mm cut-off end of the tip. Immediately transfer 320 µl of suspended resin (160 µl 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 µl 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 µl 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 (~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 µl 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 µl 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 $CuSO_4 \cdot 5H_2O$ 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 μm pore-size syringe filter to filter/aliquate $CuSO_4$ solution into 10 of 1.5ml conical screw-cap tubes (~1ml/tube). Store at 4°C.

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

250 mM NaH_2PO_4 (monobasic sodium phosphate), 2.5 M NaCl

Dissolve 7.8 g $NaH_2PO_4 \cdot 2H2O$ (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_2HPO_4 (dibasic sodium phosphate), 2.5 M NaCl

Dissolve 7.1 g Na_2HPO_4 (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 (~11 µl 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 µl 3M Imidazole in a 100 ml beaker with a stirring stir bar. Titrate with HCl or NaOH until the pH reaches 8.0 (~3 µl 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 µl 3M Imidazole in a 100 ml beaker with a stirring stir bar. Titrate with HCl or NaOH until the pH reaches 8.0 (~15 µl 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 HEPES K^+ (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 $MgCl_2$

Dissolve 101.65 g $MgCl_2 \cdot 6H2O$ 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: $MgCl_2$ 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 $ZnCl_2$

Dissolve 681.4 mg $ZnCl_2$ 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: $Na_2EDTA{\cdot}2H_2O$ 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)

	original solution				final solution	
Chemical	conc.	unit	add	unit	conc.	unit
HEPES K^+ (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			20.0	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 right before use.

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

	original solution				final solution	
Chemical	conc.	unit	add	unit	conc.	unit
HEPES K^+ (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 right before use.

Dialysis buffer (HEMG-100 buffer W/O imidazole)

	original solution				final solution	
Chemical	conc.	unit	add	unit	conc.	unit
HEPES K^+ (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 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 right before use.

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 *Dm*SNAPs 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 μl 0.1M $CuSO_4$ 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 last wash 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 ~ 1ml/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 <u>in the cold room.</u>

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 μl 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 drop by drop. Allow the eluted proteins to come out into the 15 ml tube underneath the column until completely drained (about 5mins).

d. Remove 50 μl 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 μl) 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)

^{32}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

	original solution				final solution	
Chemical	conc.	unit	add	unit	conc.	Unit
HEPES K^+ (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		
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^+ (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 $MgCl_2$

Dissolve 101.65 g $MgCl_2 \cdot 6H2O$ 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: $MgCl_2$ 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 $ZnCl_2$

Dissolve 681.4 mg $ZnCl_2$ 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 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: $Na_2EDTA{\cdot}2H_2O$ 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, 20-well 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 2nd 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
^{32}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 μl 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 μl) 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 μl 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/μl of your probe is (A cpm + B cpm)/(2 μl x 2) = (A+B)/4 cpm/μl

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 µl of 10% APS and 40 µl 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 μl of 50000 cpm/μl probe. For example, if you need 20 reactions and the radioacitivity of your probes measured from step III-14 is 400000 cpm/μl, then:

   (50000 cpm/μl x 20 μl)/ (400000 cpm/μl) = 2.5 μl

   Thus, you need 2.5 μl of the 400000 cpm/μl probe to dilute with (20-2.5=17.5) μl of d.d. water to make 20 μl of 50000 cpm/μl probe for your bandshift reaction. You can also use 22 μl instead of 20 μl in the equation to make sure you have enough probe to use.
5. Prepare the probe-mix by mixing 1 μl of the 50000 cpm/μl probe, 2 μl of 1 μg/μl poly (dI-dC) or poly (dG-dC), and 1 μl of 0.1 M DTT in a 1.5 ml conical screw-cap tube for each reaction (so 4 μl 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)	X μl
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.

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

I. Solutions and materials

Formaldehyde cross-linking solution

			original solution		final solution	
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	%
final solution volume	28.0	ml	* final volume is from all chemicals+ 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

			original solution		final solution	
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 volume. Store at 4°C.)			

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

	original solution				final solution	
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. water to bring to the final volume.)			

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

	original solution				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 μl 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

	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	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	%	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.)			

Lithium wash buffer

	original solution				final 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	%
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)

	original solution				final solution	
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 solution volume	100.0	ml	(add water to bring to the final volume. Do not 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 µl 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 ~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 screw-cap tube.

   Notes: Always use tips cut off at the end to handle the resin.
   More than 200 µl 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.

6. 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 at room temperature for 15 minutes. Centrifuge at 2500 g, room temperature for 3 minute (EPPENDORF, Centrifuge 5424).
3. Transfer the supernatant (eluate) to a 1.5 ml screw-cap tube.
4. Add another 250 µl 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 µl total). Also, prepare input DNA by adding 300 ul elution buffer to 200 µl thawed input sample (Step V-5) to make final volume 500 µl.
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 µl of PB buffer (as 5 volumes PB: 1 volume sample) in the first step (binding step); use 30 µl of TE buffer to elute the purified DNA in the last step. 30 µl 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℃	2:00 min	hot-start
Cycle2: (28x)	Step 1	94℃	0:30 min	denature
	Step 2	61℃	0:45 min	anneal
	Step 3	72℃	1:00 min	extend
Cycle3: (1x)	Step 1	72℃	10:00 min	final extension
Cycle4: (1x)	Step 1	4℃	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 1min).

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 μl 10% APS and 90 μl 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 μg/ml ethidium bromide (e.g. add 100 μl 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.