UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Subunit interactions and protein-DNA interactions of the Drosophila melanogaster small nuclear RNA activating protein complex

Permalink

https://escholarship.org/uc/item/4xx4957r

Author

Titus, Mitchell

Publication Date

2007

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Subunit Interactions and Protein-DNA Interactions of the *Drosophila melanogaster* Small Nuclear RNA Activating Protein Complex

A Dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Chemistry

by

Mitchell Titus

Committee in charge:

University of California, San Diego

Professor John E. Crowell
Professor Elizabeth A. Komives
Professor Amy Pasquinelli

San Diego State University

Professor William E. Stumph, Chair
Professor Greg L. Harris
Professor John J. Love
Professor Shelli McAlpine

2007

The Dissertation of Mitchell Titus is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

San Diego State University

2007

DEDICATION

I would like to dedicate this dissertation to my mother Simone Titus and my father Dr. John Titus who stood by me unconditionally during my long years of difficulty following my war service in Vietnam.

TABLE OF CONTENTS

Signature Page....iii

Dedication....iv

Table of Contents....v

List of Figures....vii

List of Tables....ix

Vita....x

Abstract....xii

General Introduction....1

Chapter 1....12

Subunit Interactions and Protein-DNA Interactions of the *Drosophila melanogaster* Small Nuclear RNA Activating Protein Complex

Abstract....13

Introduction....14

Experimental Procedure....16

Results....20

Discussion....37

References....45

Chapter 2....47

Mutagenesis of *Drosophila melanogaster* to Alter RNA Polymerase Specificity at an snRNA Gene Promoter

Abstract....48

Introduction.....50

Experimental Plan.....51

Results.....54

Discussion.....62

References.....63

Chapter 3.....65

Can the RNA Polymerase Specificity of an snRNA Promoter Be Altered by the Binding of a Synthetic Polyamide Compound?

Abstract.....66

Introduction.....67

Experimental Plan.....68

Results.....71

Discussion.....76

References.....77

Concluding Remarks.....78

Appendix.....84

LIST OF FIGURES

General Introduction

Fig. 1. **Comparison of core promoter elements in the 5'-flanking DNA of several snRNA genes**..... 4

Fig. 2. **Comparison of *Drosohila* U1 and U6 gene promoters**.....5

Fig. 3. **Schematic model of the recruitment of Pol II or Pol III to the U1 and U6 snRNA promoters**.....8

Chapter 1

Fig. 4. **Comparison of *D. melanogaster* (Dm) and human (Hs) SNAP subunits**.....21

Fig. 5. **Co-expression of tagged DmSNAP190 with DmSNAP50 and DmSNAP43 in stably transfected S2 cells.**.....22

Fig. 6. **Domains of DmSNAP190 required for assembly with DmSNAP50 and DmSNAP43 in vivo**.....23

Fig. 7. **Domains of DmSNAP190 required for sequence-specific DNA binding by DmSNAPc.**.....25

Fig. 8. **Co-expression of tagged DmSNAP50 with DmSNAP190 and DmSNAP43 in stably transfected S2 cells.**.....27

Fig. 9. **Domains of DmSNAP50 required for assembly with DmSNAP190 and DmSNAP43 in vivo.**.....28

Fig. 10. **Domains of DmSNAP50 required for sequence-specific DNA binding by DmSNAPc.**.....32

Fig. 11. **Co-expression of tagged DmSNAP43 with DmSNAP190 and DmSNAP50 in stably transfected S2 cells.**.....33

Fig. 12. **Domains of DmSNAP43 required for assembly with DmSNAP190 and DmSNAP50 *in vivo***.....35

Fig. 13. **Domains of DmSNAP43 required for sequence-specific DNA binding by DmSNAPc.**.....36

Fig. 14. **Summary of the effects of DmSNAP truncations on DmSNAPc assembly and DNA binding activity.**.....38

Chapter 2

Fig. 15. **Schematic model of the recruitment of Pol II or Pol III to the U1 and U6 snRNA promoters**.....51

Fig. 16. **Variations of the U1 PSEA containing U6 characteristics**.....54

Fig. 17. **Relevant portion of the P-element vector used to generate transgenic flies for the selection of mutants with altered RNA polymerase specificity.**.....55

Fig. 18. **Survival of flies on different concentrations of G418.**.....58

Fig. 19. **Survival of flies on different concentrations of G418.**.....59

Fig. 20. **Flow chart for one round of mutations.**.....61

Chapter 3

Fig. 21. **The HIV-1 polyamide bound to DNA**.....70

Fig. 22. **Wild type U1 and U6 templates transcribed in the same reaction.**.....73

Fig. 23. **Transcription with added HIV-1 polyamide.**.....74

Fig. 24. **Effect of HIV-1 polyamide on transcription of U1-19HIV1-30PSEB, U1wt and U6wt templates.**.....75

LIST OF TABLES

Chapter 3

Table 1. **DNA sequences of U1 promoters containing HIV-1 recognition sites between the PSEA and PSEB.**.....72

VITA

1966-1970	U.S. Air Force
1969	Decorated for valor in action—Tuy Hoa, Republic of Vietnam
1972	Associate of Arts, Grossmont College, El Cajon, California
1974	Bachelor of Arts, Political Science, University of California, Berkeley
1975-1976	Survey Researcher, City of Berkeley, Health Department.
1977-1978	Firefighter EMT, Alpine Fire Department, Alpine California
1978-1981	Crash Rescue Engineer, Kwajalein Fire Department, Marshall Islands
1982-1986	Captain, various Sailing Vessels, San Diego to Aukland New Zealand
1986-1990	Commercial Diver, San Diego California
1993-1996	Teaching Assistant, Dept. of Politcal Science, Grossmont College
1996	Associate of Science, Chemistry, Grossmont College, El Cajon California
1998-1999	Lecturer, Department of Chemistry, San Diego State University
1999	Bachelor of Science, Biochemistry, San Diego State University
2000	Teaching Associate, Dept. of Chemistry and Biochemistry, University of California, San Diego
2000-2007	Teaching Associate, Dept. of Chemistry, San Diego State University
2007	Doctor of Philosophy, University of California, San Diego, San Diego State University

HONORS

Arne N. Wick Predoctoral Research Fellowship (2004-2005)

Pi Sigma Alpha, the National Political Science Honor Society (1972)

PUBLICATIONS

Titus, M., Chiang, S.-C., Hung, K.-H., Stumph. W. E. **Subunit Interactions and Protein-DNA Interactions of the *Drosophila melanogaster* Small Nuclear RNA Activating Protein Complex,** submitted.

FIELD OF STUDY

Major Field: Biochemistry (Molecular Genetics)

Experimental Studies in Mechanisms of Gene Expression
Professor William E. Stumph

ABSTRACT OF THE DISSERTATION

Subunit Interactions and Protein-DNA Interactions of the *Drosophila melanogaster* Small Nuclear RNA Activating Protein Complex

by

Mitchell Titus

Doctor of Philosophy in Chemistry

University of California, San Diego, 2007

San Diego State University, 2007

Professor William E. Stumph, Chair

Small nuclear RNAs (snRNAs) are important in eukaryotic organisms because they are an integral part of the spliceosome, an RNA-protein complex that removes introns from the primary transcript of messenger RNAs. The focus

of the work in our lab is to understand the transcriptional regulation of the genes that code for the snRNAs. In particular, we are interested in the molecular mechanisms and underlying structural differences that contribute to RNA polymerase specificity at the promoters of small nuclear RNA genes transcribed by RNA polymerase II (Pol II) or RNA polymerase III (Pol III) in *Drosophila melanogaster*. Transcription of genes coding for the small nuclear RNAs depends upon a unique transcription factor known as the small nuclear RNA activating protein complex SNAPc. In *D. melanogaster*, DmSNAPc consists of three distinct subunits (DmSNAP190, DmSNAP50, and DmSNAP43) so named to correspond with the three homologous human SNAPc subunits.

Chapter 1 of this dissertation describes the identification of domains within each subunit that are 1) involved in complex formation with the other two subunits or 2) required for the DNA binding activity of the complex.

Chapter 2 describes work I performed in an attempt to mutate live *D. melanogaster* flies to induce altered RNA polymerase specificity *in vivo* at a hybrid U1 snRNA promoter transgene inserted into the fly genome.

In Chapter 3, I describe an attempt to alter RNA polymerase specificity *in vitro* using a small, synthetic polyamide that binds in a sequence-specific manner to various versions of the U1 promoter that contain a recognition site for the polyamide.

GENERAL INTRODUCTION

The overall goal of the work in our lab is to understand the molecular mechanisms and underlying structural differences that contribute to RNA polymerase specificity at the promoters of small nuclear RNA genes transcribed by RNA polymerase II (Pol II) or RNA polymerase III (Pol III) in *Drosophila melanogaster.* In eukaryotes, small nuclear RNAs (snRNAs) are required for pre-mRNA splicing.

Characteristics of "U" small nuclear RNAs - Small nuclear RNAs (snRNAs) are a metabolically stable class of RNA molecules that reside in eukaryotic nuclei. U-snRNAs are so named because many U-snRNAs are rich in uridylic acid. The importance of the U-snRNAs is reflected in the fact that the predicted secondary structures among U-snRNAs from distant species are virtually identical (1).

The five snRNAs U1, U2, U4, U5, and U6 are involved in splicing the precursors of mRNAs (2-4). *In vivo*, one U-snRNA is usually associated with six to ten polypeptides and which all together comprise a small nuclear ribonucleoprotein particle (snRNP) (5-7). The U4, U5, and U6 snRNPs interact to form a single particle called the U4-U5-U6 tri-snRNP complex (8,9).

The roles of small nuclear ribonucleoprotein particles in pre-mRNA splicing - The U1, U2, U4/U5/U6 snRNPs and a number of additional proteins assemble into a large multicomponent ribonucleoprotein complex, or spliceosome (10,11). The spliceosome recognizes and binds to the highly conserved boundary sequences at the 5’ and 3’ splice sites at exon/intron junctions in the mRNA precursor. Spliceosome assembly is initiated when a nucleotide sequence at the 5' end of the U1 snRNA forms

base pairs complementary to the nine-nucleotide consensus sequence at the intron 5' splice site. The U2 snRNP then recognizes and binds to the branch point adenosine, which is located about 30 nucleotides from the 3' end of the intron (12,13). Then, with the help of the U4-U5-U6 tri-snRNP, and following an ATP-dependant conformational change, the branch point "A" nucleotide attacks and cleaves the 5' splice site (14). The cut 5' end of the intron sequence becomes covalently bonded to the branch point "A" nucleotide by nucleophilic substitution by the 2' hydroxyl ribose group of the adenosine forming a branched nucleotide closed loop or "lariat" (9). With the help of the U5 snRNP the 3' splice site of the intron is then cleaved freeing the looped intron from the pre-mRNA (14). The two free ends of the exon fragments are joined by nucleophilic substitution of the 3' OH end of the 5' exon fragment with the 5' end of the 3' exon fragment (11). A single vertebrate gene can contain as many as 50 or more introns which are removed by an orderly process, thus forming a functional mRNA (15).

Structure and expression of genes coding for the snRNAs - Eukaryotic genomes contain multiple copies of the genes that encode the snRNAs. In *Drosophila melanogaster*, there are two to seven copies of each of the snRNA genes (16). In humans, there are around thirty copies of the U1 snRNA gene, around ten copies of the U2 gene (17-19), eight copies of the U4 gene, twelve copies of the U5 gene and five copies of the U6 gene (20).

RNA polymerases I, II, and III (Pol I, Pol II, and Pol III) direct transcription of different classes of genes dependant on the genes' distinct promoter structures. Each

RNA polymerase recognizes its target promoters via protein-protein interactions with a distinct set of basal transcription factors bound to the core promoter elements. The genes that code for the snRNAs are interesting in that they have unusual properties and unique core promoter structures (Fig. 1). The genes coding for the U1, U2, U4, and U5 snRNA are transcribed by Pol II, but the U6 genes are transcribed by Pol III (21-25).

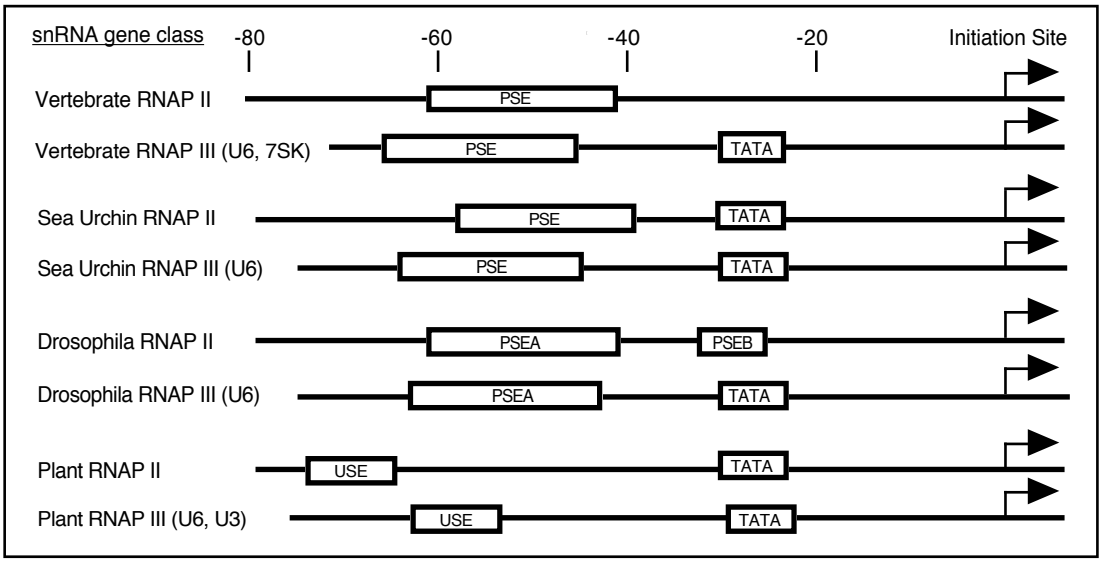


FIG 1. **Comparison of core promoter elements in the 5'-flanking DNA of several snRNA genes.** All the snRNA genes shown contain an essential element (designated PSE, PSEA, or USE) upstream of position-40. Most snRNA genes also contain a regulatory element near position -30 or -25, which probably represents a site of interaction with TBP or a TBP-related factor.

Most snRNA promoters contain two important cis-acting elements: a proximal sequence element (PSE), about 40-65 base pairs upstream of the transcription start site (26,27), and a TATA box or similar conserved sequence at the normal location of the TATA box about 25-30 base pairs upstream of the transcription start site (16,28). Vertebrate Pol II-transcribed snRNA genes are a notable exception (Fig. 1). The

presence or absence of a TATA box in vertebrate snRNA gene promoters determines RNA polymerase specificity of the gene. In vertebrates, the promoters of snRNA genes transcribed by Pol II lack TATA boxes or any other conserved sequence at the location of the TATA box, whereas in the U6 snRNA gene, transcribed by Pol III, the promoter contains a TATA box. The proximal sequence elements (PSEs) in vertebrate U1 and U6 gene promoters are functionally interchangeable and do not appear to contribute to RNA polymerase specificity (29).

In plants, both Pol II- and Pol III-transcribed genes contain a TATA box and a PSE (called a USE or upstream sequence element) but RNA polymerase specificity is determined by the spacing between the two (30). The plant USEs are interchangeable and do not contribute to polymerase specificity (30-32).

In *Drosophila*, the PSE is referred to as the PSEA to distinguish it from a second conserved core promoter element called the PSEB, present in the promoters of all *Drosophila* snRNA genes that are transcribed by Pol II (16,28).

There is a strictly conserved 8 base pair spacing between the PSEA and the PSEB of the Pol II-transcribed snRNA genes, and a strictly conserved 12 base pair spacing between the U6 (Pol III) PSEA and the TATA box (Fig. 2).

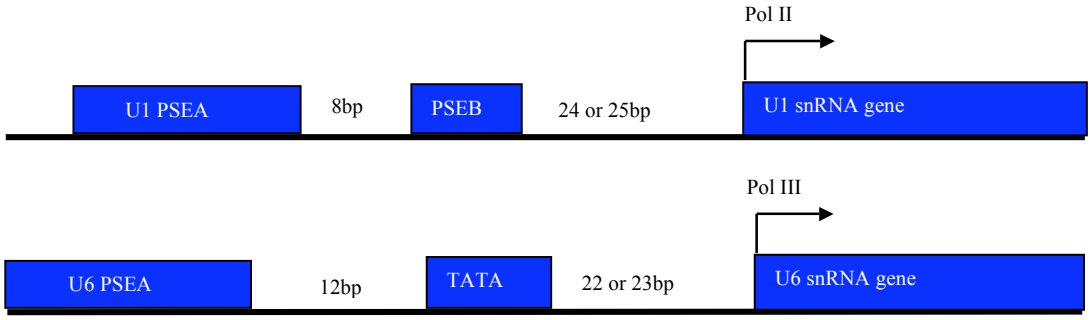


FIG 2. **Comparison of *Drosophila* U1 and U6 snRNA gene promoters.**

In Drosophila, the PSEA is the dominant element for determining the RNA polymerase specificity of snRNA genes - The *Drosophila* U1 and U6 PSEAs are very similar (16). Out of 21 base pairs, the PSEAs of the U1 and U6 gene on which we have focused our studies vary at only 5 positions: 7, 14, 16, 19, and 20 (see Chapter 2, Figure 2). Previous *in vitro* transcription experiments from our lab have shown that when the PSEA in a U1 promoter with a PSEB was changed from a U1 PSEA to a U6 PSEA, the U1 gene was transcribed by Pol III instead of the normal Pol II. Similar experiments were carried out switching the PSEB in a U1 promoter to a TATA box. The results showed no change in RNA polymerase specificity. Similar results were obtained when the spacing between the PSEA and PSEB was increased from the normal 8 base pairs to 12 base pairs. Even though there was a decrease in transcription efficiency, there was no change in polymerase specificity. These results led to the important conclusion that the sequence of the PSEA determined the RNA polymerase specificity, and not the PSEB, TATA, nor the spacing differences (33,34).

*The Drosophila snRNA activating protein complex (*Dm*SNAPc) binds to the U1 PSEA in a different conformation than it does to the U6 PSEA* - Although the PSEAs of all *D. melanogaster* snRNA genes share extensive sequence similarity, the PSEAs of the U6 genes present in the fly genome consistently vary at a few nucleotide positions from the PSEAs of the RNA polymerase II-transcribed snRNA genes (16). Nonetheless, both PSEAs are recognized by the same PSE-binding protein the *Drosophila melanogaster* snRNA activating protein complex (DmSNAPc).

DmSNAPc contains three distinct subunits, DmSNAP43, DmSNAP50, and DmSNAP190, so named to correspond with the three homologous human SNAPc subunits (35,36).

Previous photo-cross-linking studies from our lab have shown that the binding of DmSNAPc to the promoter is different depending upon whether DmSNAPc is bound to a U1 or a U6 PSEA (35-37). Those results, together with functional polymerase specificity studies, led us to propose a model in which the U1 and U6 PSEAs act as differential allosteric effectors of the conformation of DmSNAPc, subsequently leading to the specific downstream recruitment of either RNA polymerase II or RNA polymerase III. For the normal expression of the U1 gene, wild type DmSNAPc recognizes the U1 PSEA sequence and binds to it in a certain conformation. This in turn recruits Pol II transcription factors and Pol II (Fig. 3).

In much the same way, for the normal expression of the U6 gene, the exact same DmSNAPc recognizes the U6 PSEA sequence and binds in a different conformation than the DmSNAPc bound to the U1 PSEA. This in turn recruits Pol III transcription factors and Pol III (Fig. 3). Our working model to explain the RNA polymerase specificity of *Drosophila* snRNA gene promoters based upon conformational differences of DmSNAPc is schematically diagrammed in Fig. 3.

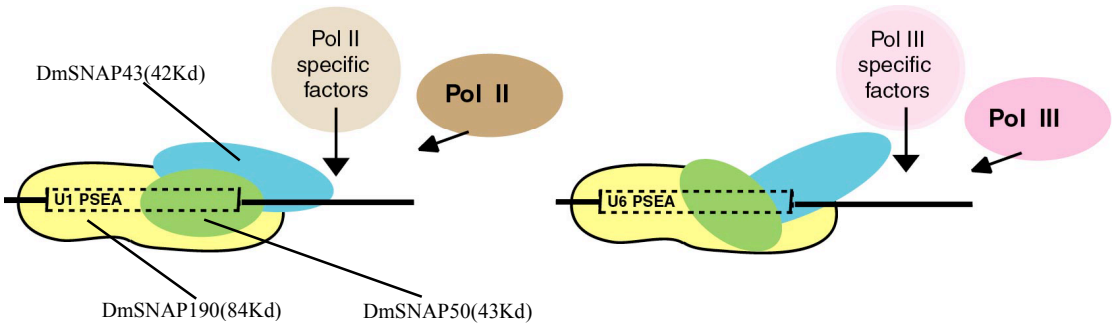


FIG 3. **Schematic model of the recruitment of Pol II or Pol III to the U1 and U6 snRNA promoters.** The U1 and U6 PSEAs are each 21 base pairs in lingth. The yellow, green, and blue ovals represent the three DmSNAP subunits. The length of DNA covered by each oval corresponds to the region of DNA each subunit interacts with based on photo-cross-linking data.

Subject matter of this dissertation - In order to understand the activation of Pol II and Pol III transcription at snRNA promoters at the molecular level, it is essential to obtain knowledge of the structure-function relationships of DmSNAPc and to understand how DmSNAP interacts differentially with U1 and U6 snRNA promoters. In Chapter 1 of this dissertation, I describe experiments to identify functional domains of the DmSNAP subunits that are required for DmSNAPc assembly and DNA binding. To do this, tagged full length and tagged truncated versions of each of the DmSNAP subunits were over-expressed in S2 cells along with untagged, full length versions of the other two subunits.

From this work I was able to identify a region of the DmSNAP190 subunit necessary for binding to the DmSNAP50 subunit, a region sufficient for binding to the DmSNAP43 subunit, and a region necessary for binding of the DmSNAPc to the U1 promoter. I also describe work by other students in the lab that identify similar functional domains within the DmSNAP50 and DmSNAP43 subunits.

Chapter 2 describes a project in which I attempted to randomly mutate live transgenic *Drosophila melanogaster* in order to induce a change in RNA polymerase specificity *in vivo* at a hybrid U1 promoter inserted into the fly genome.

Chapter 3 describes work I performed in an attempt to alter RNA polymerase specificity *in vitro* at various versions of the U1 promoter that contain a recognition sequence that binds a small synthetic polyamide at various positions in the DNA separating the PSEA and the PSEB.

REFERENCES

1. Reddy, R. and Busch, H. (1988) Small nuclear RNAs: RNA sequences, structure, and modifications. In: Birnstiel, M. L. (ed). Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Springer Verlag KG, Heidelberg, Federal Republic of Germany

2. Sharp, P. A. (1994) *Cell* **77**, 805-816

3. Guthrie, C. (1991) *Science* **253**, 157-163

4. Steitz, T. A. (1998) *Nature* **391**(6664), 231-232

5. Berget, S. M., and Robberson, B. L. (1986) *Cell* **46**(5), 691-696

6. Chabot, B., Black, D. L., LeMaster, D. M., and Steitz, J. A. (1985) *Science* **230**(4732), 1344-1349

7. Krainer, A. R., and Maniatis, T. (1985) *Cell* **42**(3), 725-736

8. Maniatis, T., and Reed, R. (1987) *Nature* **325**(6106), 673-678

9. Wassarman, D. A., and Steitz, J. A. (1992) *Science* **257**(5078), 1918-1925

10. Balvay, L., Libri, D., and Fiszman, M. Y. (1993) *Bioessays* **15**(3), 165-169

11. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119-1150

12. Wu, J., and Manley, J. L. (1989) *Genes Dev.* **3**(10), 1553-1561

13. Zhuang, Y., and Weiner, A. M. (1989) *Genes Dev.* **3**(10), 1545-1552

14. Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. (1992) *Genes Dev.* **6**(12B), 2542-2553

15. Ruby, S. W., and Abelson, J. (1991) *Trends Genet.* **7**(3), 79-85

16. Hernandez, G., Valafar, F., and Stumph, W. E. (2007) *Nucleic Acids Res.* **35**(1), 21-34

17. Hammarstrom, K., Westin, G., Bark, C., Zabielski, J., and Petterson, U. (1984) *J. Mol. Biol.* **179**(2), 157-169

18. Lindgren, V., Bernstein, L. B., Weiner, A. M., and Francke, U. (1985) *Mol. Cell. Biol.* **5**(9), 2172-2180

19. Pavelitz, T., Rusché, L., Matera, A. G., Scharf, J. M., and Weiner, A. M. (1995) *EMBO J.* **14**, 169-177

20. Domitrovich, A. M., and Kunkel, G. R. (2003) *Nucleic Acids Res.* **31**(9), 2344-2352

21. Dahlberg, J. E., and Lund, E. (1988) The genes and transcription of the major small nuclear RNAs. In: Birnstiel, M. L. (ed). Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Springer Verlag KG, Heidelberg, Federal Republic of Germany

22. Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988) *Cell* 55(3), 435-442

23. Hernandez, N. (1992) Transcription of vertebrate snRNA genes and related genes. In: McKnight, S. L., and Yamamoto, K. R. (eds). Transcriptional Regulation, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

24. Lobo, S. M., and Hernandez, N. (1989) *Cell* **58**, 55-67

25. Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L. C., and Stumph, W. E. (1997) *Eur.J.Biochem.* **248**, 231-237

26. Lo, P. C. H., and Mount, S. M. (1990) *Nucleic Acids Res.* **18**, 6971-6979

27. Das, G., Henning, D., and Reddy, R. (1987) *J.Biol.Chem.* **262**, 1187-1193

28. Zamrod, Z., Tyree, C. M., Song, Y., and Stumph, W. E. (1993) *Mol.Cell.Biol.* **13**, 5918-5927

29. Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988) *Cell* **55**, 435-442

30. Waibel, F., and Filipowicz, W. (1990) *Nature* **346**, 199-202

31. Kiss, T., Marshallsay, C., and Filipowicz, W. (1991) *Cell* **65**, 517-526

32. Goodall, G. J., Kiss, T., and Filipowicz, W. (1991) *Oxford Surveys of Plant Molecular & Cell Biology* **7**, 255-296

33. Jensen, R. C., Wang, Y., Hardin, S. B., and Stumph, W. E. (1998) *Nucleic Acids Res.* **26**, 616-622

34. McNamara-Schroeder, K. J., Hennessey, R. F., Harding, G. A., Jensen, R. C., and Stumph, W. E. (2001) *J.Biol.Chem.* 276(34), 31786-31792

35. Wang, Y., and Stumph, W. E. (1998) *Mol.Cell.Biol.* **18**, 1570-1579

36. Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (2004) *Mol.Cell.Biol.* **24**(5), 1897-1906

37. Hardin, S. B., Ortler, C. J., McNamara-Schroeder, K. J., and Stumph, W. E. (2000) *Nucleic Acids Res.* **28**(14), 2771-2778

CHAPTER 1

Subunit Interactions and Protein-DNA Interactions of the *Drosophila melanogaster* Small Nuclear RNA Activating Protein Complex

ABSTRACT

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. This is consistent with an earlier finding that all three polypeptides are closely associated with the DNA and are required for DNA binding activity. With one exception, domains required for subunit-subunit interactions lie in the most evolutionarily conserved regions of the proteins; however, DNA binding is dependent not only upon the conserved regions, but is also highly dependent upon domains outside the conserved regions. A comparison with functional domains identified in human SNAPc indicates parallels but also reveals significant differences.

INTRODUCTION

The small nuclear RNA (snRNA) activating protein complex (SNAPc) is a multi-subunit 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 (PBP) (5,6) and Proximal Transcription Factor (PTF) (1,3). 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 chapter, 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 have calculated molecular weights of 84, 43, and 42 kDa respectively. Interestingly, a homologous complex (tSNAPc) is required for transcription of the spliced leader snRNA in trypanosomes (15-17). 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.

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 (1-3,15-18). Moreover, each of the three metazoan core subunits is essential for sequence-specific binding to the PSE as none can bind individually or in pair-wise combinations (19-21, our unpublished observations). However, an isolated region of HsSNAP190 that contains two Myb repeats binds weakly and non-sequence specifically to DNA (11,20,22).

We now report a mutational analysis of the three subunits of DmSNAPc from the fruit fly *Drosophila melanogaster* designed 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 *in vivo* assembly of DmSNAPc in a homologous system was monitored by co-immunoprecipitation assays. 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 within different metazoan SNAP complexes.

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 PCR templates 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 (approximately 10 nucleotides). The second two thirds of the forward primer (approximately 20 nucleotides) was designed to exactly 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 Spe1 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-terminally 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 hours with 0.5mM copper sulfate.

FLAG purification/co-immunoprecipitation - Following induction, cells were washed in phosphate buffered saline and lysed in CelLytic M lysis buffer (Sigma cat. # C2978) containing protease inhibitor cocktail (Sigma cat. # P8340). The cell lysate was incubated overnight with anti-FLAG M2 agarose beads (Sigma cat. # A2220). The beads were then washed four times in 0.05M Tris HCl, 0.15M NaCl and then twice in HEMG wash buffer (81 mM KCl, 32.5 mM Hepes K^+, 5.5 mM $MgCl_2$, 5.0 mM DTT, 10% glycerol and 0.1mM EDTA [pH 7.6]). DmSNAP complexes were

eluted from the affinity gel by competition with 3X FLAG peptide (Sigma cat. # F4799) at a concentration of 200 μg/ml in HEMG wash buffer.

Immunoblots - Whole cell lysate and FLAG-purified samples were run on 10% polyacrylamide gels and the proteins were transferred to PVDF membranes. FLAG-tagged DmSNAP subunits were detected using either anti-FLAG M2® monoclonal antibody-alkaline phosphatase conjugate from Sigma (product no. F 3165) and Western Blue Stabilized Substrate for alkaline phosphatase from Promega (product no. S3841) or ANTI-FLAG M2® monoclonal antibody-horseradish peroxidase (HRP) conjugate from Sigma (product no. A8592) and chemoluminescent peroxidase substrate (CPS) from Sigma (product no. CPS-1-60). Untagged DmSNAP subunits were detected as previously described by 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 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. This permitted a quick 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 ul volumes in a final concentration of approximately 80 mM KCl, 25 mm HEPES [pH 7.6], 5 mM $MgCl_2$, 10 uM $ZnCl_2$, 0.1 mM EDTA, 9 mM dithiothreitol,

9% glycerol. The radioactive DNA probe contained the PSEA and PSEB sequences of the *D. melanogaster* U1PSEA/B-22/-68LO with the following sequence:

```
                    PSEA                    PSEB
5'  GTTCGTTGCAATTCCCAACTGGTTTTAGCTGCTCAGCCATGGAAACC  3'
   3'AAGCAACGTTAAGGGTTGACCAAAATCGACGAGTCGGTACCTTTGGG  5'
```

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 used in additional lanes to maximize the possibility of observing binding. Complexes were run on 5% non-denaturing polyacrylamide gels and the bands were detected by autoradiography.

In reactions using less than 12 ul of DmSNAPc, volume was made up to 12 ul with HEMG (81 mM KCl and 10% glycerol). Three ul of BCZ (150mM KCl, 10% glycerol) and 3ul of a premix containing 1ul of 0.1M dithiothreitol, 2 μg of poly d(I-C), and 1ul of ^{32}P radiolabeled PSEA/PSEB probe (50,000 cpm/ul) were added to each reaction for a total volume of 18 ul. The reaction mixtures were incubated at 20°C for 30 minutes and then run on a 5% non-denaturing polyacrylamide gel at 70 volts for 5.5 hours using buffer containing 25mM tris, 190mM glycine and 1mM EDTA and the bands detected by autoradiography.

RESULTS

Funtional 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. 4) is conserved between Drosophila 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 (11,20,22). Because it seemed likely that deletion of this most conserved region of *Dm*SNAP190 would be highly detrimental to *Dm*SNAPc activity, we focused our studies on the non-conserved regions. Full-length and N-terminal and C-terminal DmSNAP190 truncation constructs under control of the copper-inducible metallothionein promoter were prepared with FLAG-6His tags at the C-terminus or with 6His-FLAG tags at the N-terminus (Fig. 5A). Each tagged DmSNAP190 construct was used separately to transfect Drosophila S2 cells together with pCoBlast (Invitrogen) and 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

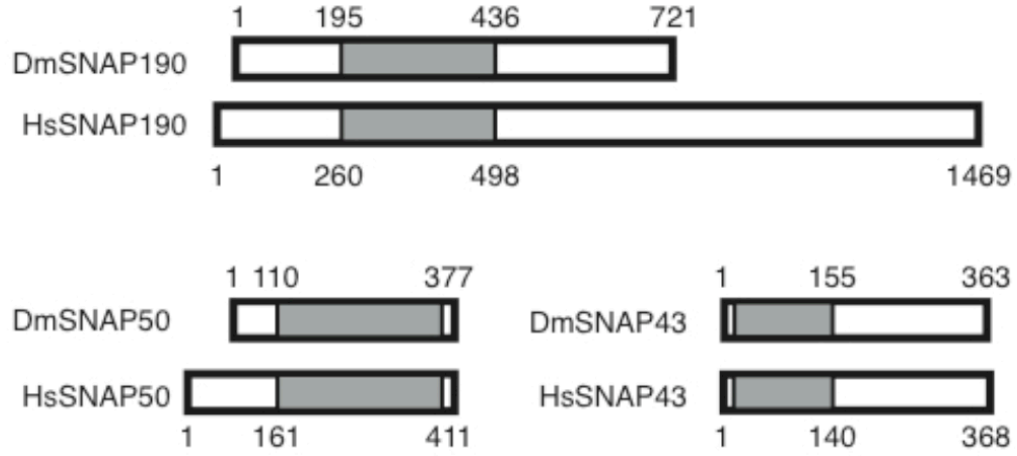


FIG. 4. **Comparison of *D. melanogaster* (Dm) and human (Hs) SNAP subunits.** The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous protein pairs. The numbers above and below the rectangles designate the amino acid positions in the proteins.

DmSNAP43 (14). Fig. 5B shows that each of the DmSNAP190 truncation mutants was expressed in the corresponding stably transfected cells (lanes 1-7, top panels). Fig. 5B 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 reference (14)].

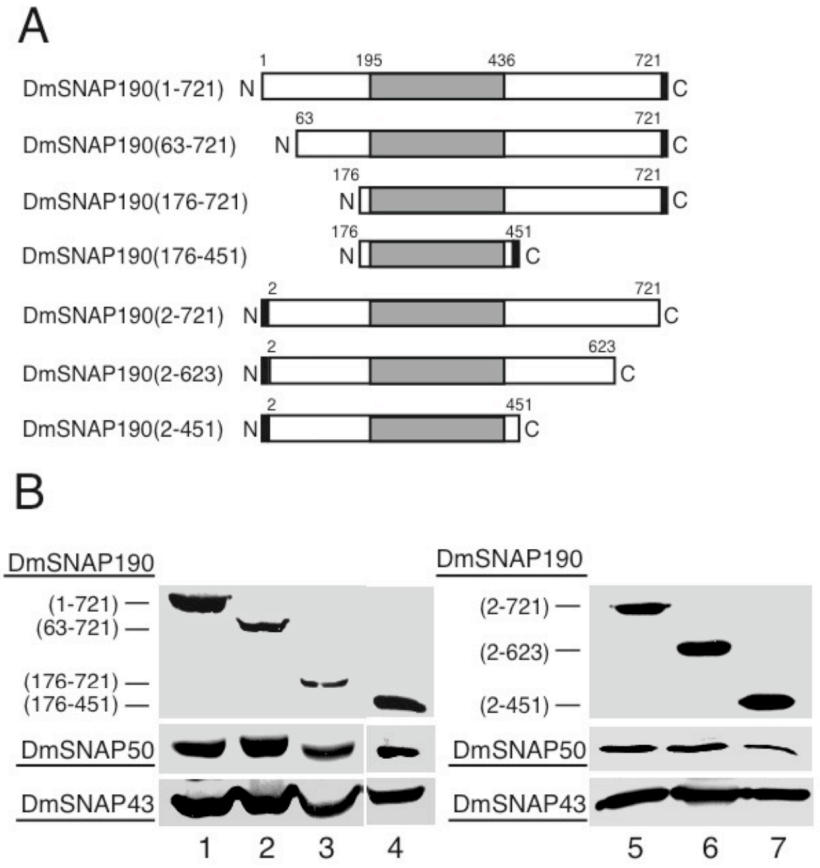


FIG. 5. **Co-expression of tagged DmSNAP190 with DmSNAP50 and DmSNAP43 in stably transfected S2 cells.** A) Schematic representation of the full length and truncated DmSNAP190 constructs co-transfected into S2 cells together with constructs expressing full length untagged DmSNAP43 and DmSNAP50. The shaded 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) Detection of DmSNAP subunits co-expressed in permanently transfected S2 cells following copper sulfate induction. Whole cell lysates were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of lysate loaded in each lane was normalized so that the intensity of the signal obtained from the tagged DmSNAP190 construct was similar in each lane (lanes 1,2,5: 30 μg; lane 3: 75 μg; lane 4: 15 μg; lane 6: 90 μg; lane 7: 10 μg). 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). In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

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. Bound protein was eluted by competition with FLAG peptide. All seven full-length or truncated DmSNAP190 proteins could be detected in the elution fractions by immunoblot analysis using the FLAG antibody (Fig. 6, top panels). As revealed in the bottom-most panels of Fig. 6, 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.

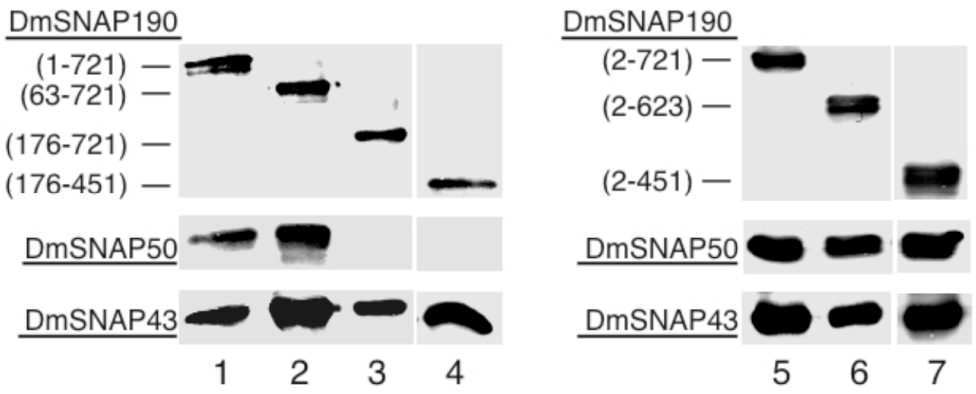


FIG. 6. **Domains of DmSNAP190 required for assembly with DmSNAP50 and DmSNAP43 in vivo.** Complexes containing full length or truncated DmSNAP190 were purified using FLAG antibody beads, and the presence of individual subunits in the elution fractions was assayed by immunoblotting. The volume of elution 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. In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

However, the result with DmSNAP50 was different. As expected, DmSNAP50 co-purified with full-length DmSNAP190 tagged at either the C- or N-terminus (Fig. 6, lanes 1 and 5, middle panels). Moreover, DmSNAP50 associated with the truncation construct that lacked the first 62 amino acid residues of DmSNAP190 (Fig. 6, lane 2); 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. 6, 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. 7). 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. 7, lanes 1, 2, 5, and 6). However, the complex that contained DmSNAP190 lacking 175 N-terminal residues was not able to bind DNA (Fig. 7, lane 3). DmSNAP190 that contained only the Myb domain gave a similar result (lane 4). Neither the use of 2 or 3 times as much protein (Fig. 7, lanes 7 and 8), nor longer exposure of the film (data not shown), was able to detect any DNA binding activity for these two truncation constructs. However, this was expected because these complexes lack DmSNAP50 (Fig. 6), 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 (our unpublished data).

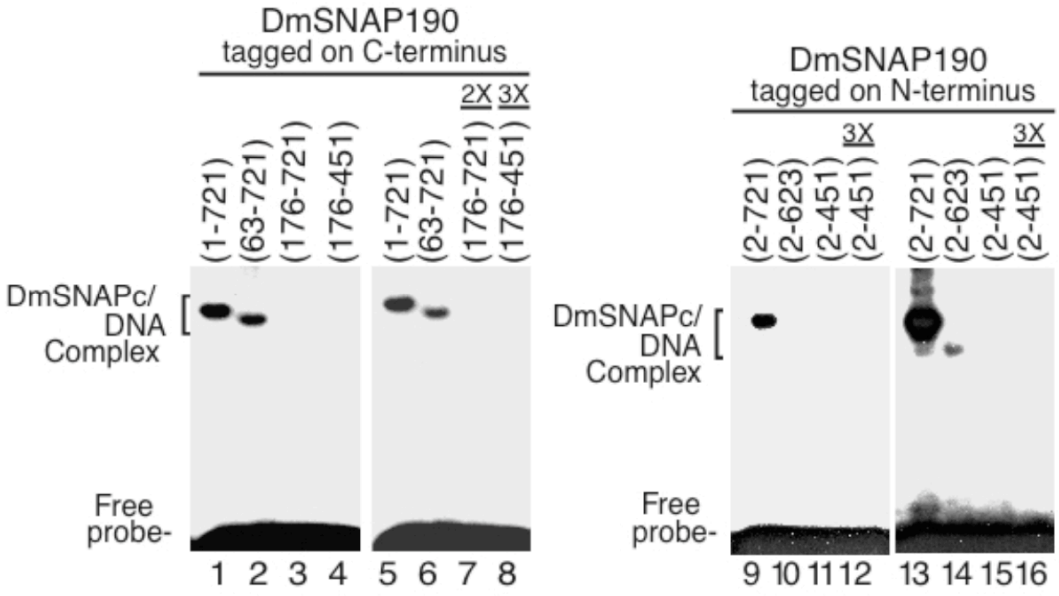


FIG. 7. **Domains of DmSNAP190 required for sequence-specific DNA binding by DmSNAPc.** FLAG-affinity-purified DmSNAP complexes (Fig. 3) were used for DNA mobility-shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP190 constructs tagged at the C-terminus (N-terminal 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. 3, whereas lanes 7 and 8 contained 2 or 3 times the normalized amount of truncated protein. 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.

DmSNAPc that contained full-length DmSNAP190 tagged at the N-terminus also bound efficiently to DNA (Fig. 7, lane 9). However, DNA binding was severely compromised by truncations from the C-terminus following residues 623 or 451 (lanes 10-12). A longer exposure of film to the same gel (Fig. 7, lanes 13-16) revealed that

DmSNAPc containing DmSNAP190 truncated at position 623 (lane 14) retained a very low level of DNA binding activity, but truncation following position 451 completely abrogated detectable DNA binding (lane 15) even when 3 times as much protein was used (lane 16). These results indicate 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. 6). This result was surprising because HsSNAP190 truncated immediately following the Myb domain can associate with HsSNAP43 and HsSNAP50 into a complex known as miniSNAPc that binds very efficiently to DNA (19,20,22). Thus, the C-terminal domain of SNAP190 plays an important role in DNA binding by DmSNAPc even though it is not required, and is even inhibitory (19), for HsSNAP190 binding to DNA.

Funtional domains of DmSNAP50 - The SNAP50 orthologs are the most evolutionarily conserved of the SNAP proteins (14,15). 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. 4) (14). We concentrated most of our effort on truncations from the N-terminus, but two C-terminal truncations were also prepared (Fig. 8A). N-terminal truncations had FLAG-6His tags at the C-terminus, whereas the C-terminal truncations had 6His-FLAG tags at the N-terminus. Each DmSNAP50 construct was co-overexpressed in S2 cells together with full-length untagged DmSNAP190 and DmSNAP43.

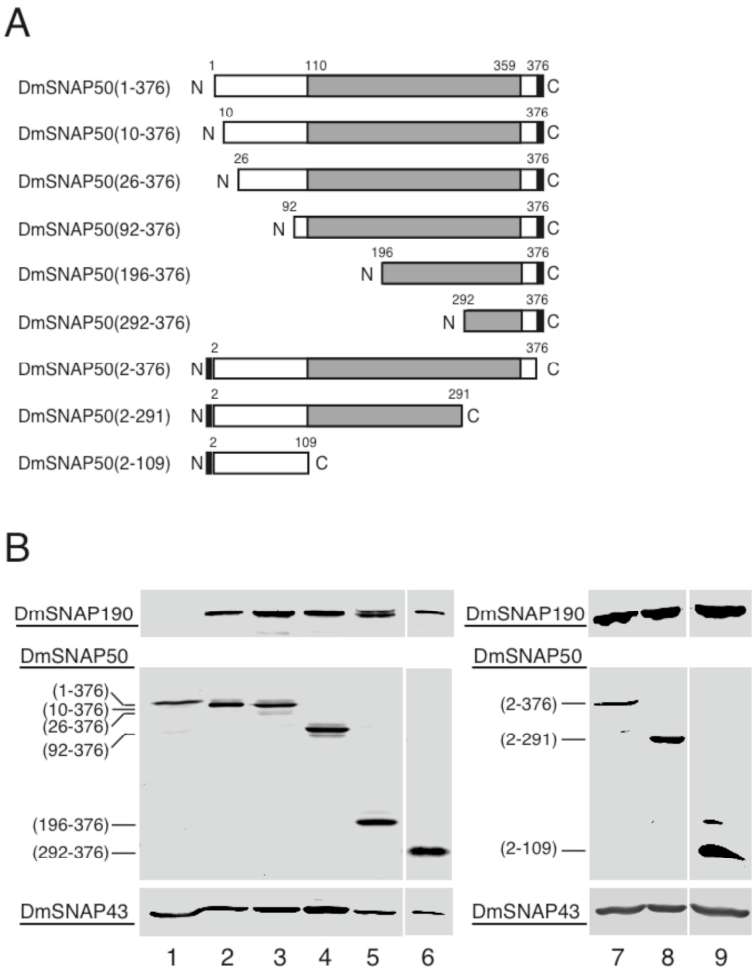


FIG. 8. **Co-expression of tagged DmSNAP50 with DmSNAP190 and DmSNAP43 in stably transfected S2 cells.** A) Schematic representation of the full length and truncated DmSNAP50 constructs co-transfected into S2 cells together with constructs expressing full length untagged DmSNAP190 and DmSNAP43. The shaded 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) Detection of DmSNAP subunits co-expressed in permanently transfected S2 cells following copper sulfate induction. Whole cell lysates were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of lysate loaded in each lane was normalized so that the intensity of the signal obtained from tagged DmSNAP50 was similar in each lane (lane 1: 61μg; lane 2: 3.7μg; lane 3:10μg; lane 4: 4.9μg; lane 5: 14μg; lane 6: 1.3μg; lane 7: 40μg; lane 8: 20μg; and lane 9: 60μg). 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. In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

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. 8B, 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-terminally tagged DmSNAP50 under conditions where it was readily detectable in the remaining cell lines (Fig. 8, upper panel, lane 1 vs. lanes 2-9). However, upon FLAG purification of DmSNAPc from those cells, DmSNAP190 was found to be present in the FLAG-purified fraction although at a reduced level (see Fig. 9 below).

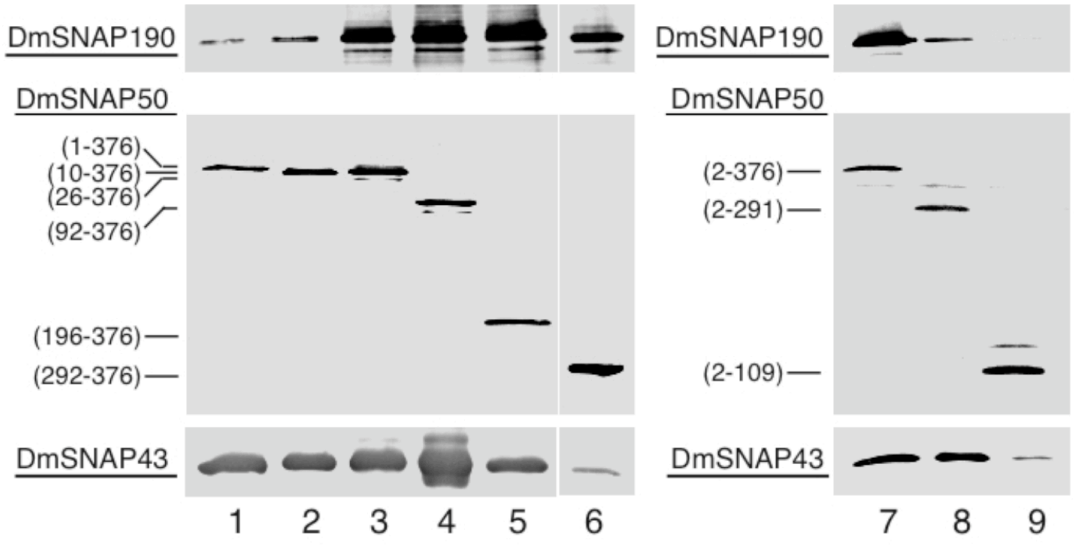


FIG. 9. **Domains of DmSNAP50 required for assembly with DmSNAP190 and DmSNAP43 in vivo.** Complexes containing full length or truncated DmSNAP50 were purified using FLAG antibody beads, and the presence of individual subunits in the elution fractions was assayed by immunoblotting. The volume of elution 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. In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

The abilities of DmSNAP190 and DmSNAP43 to associate with the DmSNAP50 constructs were assayed by immunoblotting following FLAG affinity purification. Lanes 1-9 in the middle panels of Fig. 9 show that each of the FLAG-tagged DmSNAP50 variants was obtained in the elution fractions from the FLAG affinity resins. The top panels in Fig. 9 reveal the ability of DmSNAP190 to co-purify with the various full-length and truncated DmSNAP50 constructs. Surprisingly, DmSNAP50 truncation mutants that had 25 or more amino acid residues deleted from the N-terminus co-precipitated greater amounts of DmSNAP190 compared to that precipitating with the full-length or nearly full-length constructs (Fig. 9 top panels, lanes 3-6 vs. lanes 1 and 2). The low amount of DmSNAP190 that co-precipitated with FLAG-tagged full-length DmSNAP50 is likely related to the low expression levels of DmSNAP190 in that particular stably transfected cell line (Fig. 8B, top panel, lane 1). However, that reasoning cannot account for the relatively low amount of DmSNAP190 that co-precipitated with the DmSNAP50 construct that lacked just nine N-terminal amino acid residues (compare Fig. 9 lane 2 and Fig. 8B lane 2 vs. lanes 3-6 in each figure). This raises the possibility that the first 25 amino acid residues of DmSNAP50 may inhibit DmSNAP50 association with DmSNAP190.

The last 85 residues of DmSNAP50 (residues 292-376) contain an evolutionarily conserved (14,15) yet unorthodox zinc-binding domain termed the "SNAP finger" (21). Interestingly, these last 85 residues were sufficient for interaction with DmSNAP190 (Fig. 9, lane 6, top panel). This was surprising because direct interactions between HsSNAP190 and HsSNAP50 had not been reported in the human

system (11,19,21,24). Results from the C-terminal truncations of DmSNAP50 (Fig. 9, lanes 7-9) were also consistent with the importance of the SNAP finger for interaction with DmSNAP190, as truncation of DmSNAP50 beyond residue 291 significantly impaired the co-purification of DmSNAP190 (compare lanes 7 and 8 in the top panels of Fig. 9). Moreover, deletion of DmSNAP50 amino acids C-terminal to position 109 completely abrogated the ability of DmSNAP190 to associate with DmSNAP50 (Fig. 9 top panel, lane 9), suggesting that residues between 110 and 291 may stabilize the interaction of DmSNAP50 with DmSNAP190.

DmSNAP43 co-purified efficiently with the majority of the DmSNAP50 truncation constructs (Fig. 9, bottom panels, lanes 1-5 and 7-8). However, deletion of DmSNAP50 residues between 196 and 292 impaired the association of DmSNAP43 with the truncated DmSNAP50 (lane 6, bottom panel), suggesting that this central region of DmSNAP50 is important for its interaction with DmSNAP43. This interpretation is supported by the data from the C-terminal truncations shown in lanes 7-9. DmSNAP43 associated well with DmSNAP50 that was either full length or truncated beyond residue 291 (Fig. 9, lanes 7 and 8) but associated only weakly with the DmSNAP50 construct that lacked residues beyond 109 (lane 9, bottom panel). Taken together, these data suggest that the central region of DmSNAP50 (most likely residues between 196-292) are probably most important for efficient assembly with DmSNAP43. However, even the shortest N-terminal and C-terminal DmSNAP50 constructs exhibited weak association with DmSNAP43, suggesting that nearly the

entire length of DmSNAP50 could be involved in forming some sort of interface with 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. DmSNAPc that contained full-length DmSNAP50 tagged on the C-terminus bound efficiently to DNA (Fig. 10, lanes 1 and 7). SNAP complexes that contained DmSNAP50 lacking either the first 9 or 25 amino acid residues also bound to the PSEA (Fig. 10, lanes 2, 3, 8, and 9). However, deletion of the first 91 amino acid residues of DmSNAP50 resulted in the complete loss of detectable DNA binding activity (Fig. 10, lanes 4 and 10) even though this DmSNAP50 construct associated very well with both DmSNAP190 and DmSNAP43 (Fig. 9, 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 (lane 14), both C-terminal truncations of DmSNAP50 resulted in the loss of DNA binding activity (lanes 13, 15, and 16). These results were to be expected based upon the data from Fig. 9 showing that DmSNAP190 did not efficiently associate with these DmSNAP50 truncation constructs. Furthermore, in the human system the SNAP finger is important for DNA binding (21).

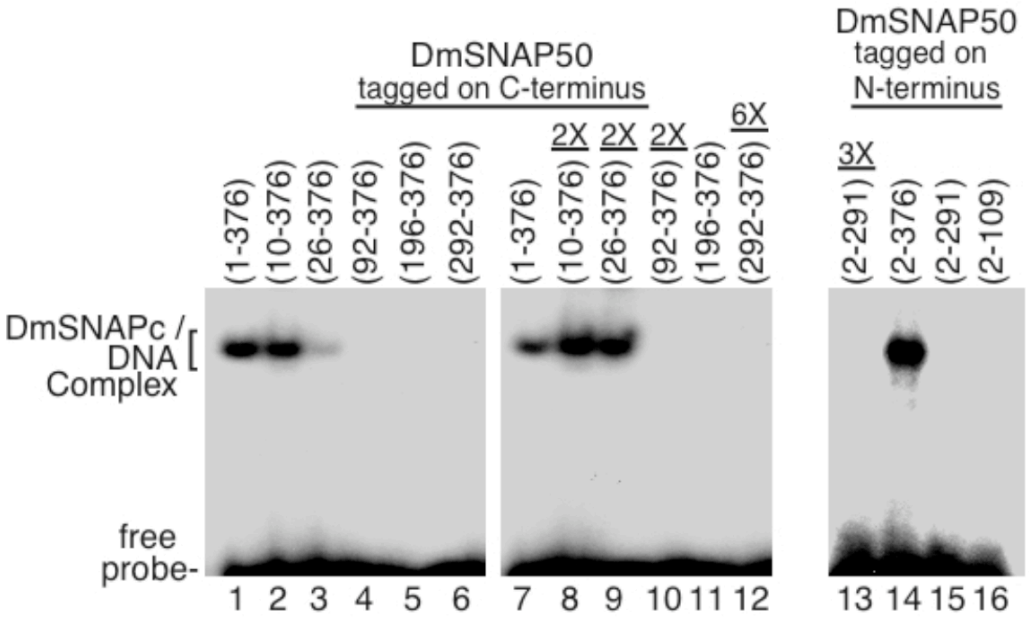


FIG 10. **Domains of DmSNAP50 required for sequence-specific DNA binding by DmSNAPc.** FLAG-affinity-purified DmSNAP complexes (Fig. 6) were used for DNA mobility-shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP50 constructs tagged at the C-terminus (N-terminal 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 DmSNAP190 as determined in Fig. 6, 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.)

Funtional 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. 11A). Each construct was co-overexpressed in S2 cells with untagged DmSNAP190 and DmSNAP50.

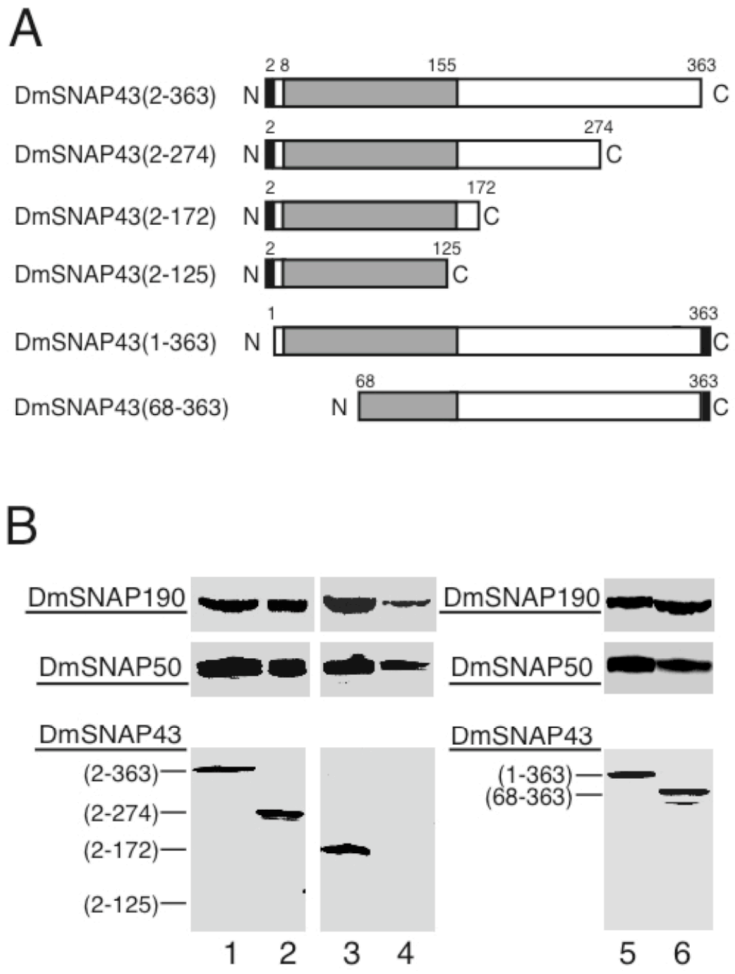


FIG. 11. **Co-expression of tagged DmSNAP43 with DmSNAP190 and DmSNAP50 in stably transfected S2 cells.** A) Schematic representation of the full length and truncated DmSNAP43 constructs co-transfected into S2 cells together with constructs expressing full length untagged DmSNAP190 and DmSNAP50. The 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) Detection of DmSNAP subunits co-expressed in permanently transfected S2 cells following copper sulfate induction. Whole cell lysates were run on denaturing gels and DmSNAP subunits were detected by immunoblot analysis. The amount of lysate loaded in each lane was normalized so that the intensity of the signal obtained from the tagged DmSNAP43 construct was similar in each lane (lane 1: 70μg; lanes 2 and 5: 20μg; lane 3: 100μg; lane 4: 74μg; and lane 6: 60μg). 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. In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

Expression of each of the full-length and truncated proteins was examined by immunoblot analysis of cellular extracts prepared from the individual stably transfected cells lines. The bottom panels in Fig. 11B show that expression of the two tagged full-length and three of the truncated DmSNAP43 constructs was 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. 11B illustrate that DmSNAP190 and DmSNAP50 were expressed in all six cell lines.

The DmSNAP complexes were then immunoprecipitated by FLAG affinity chromatography, and the ability of DmSNAP190 or DmSNAP50 to co-precipitate with the tagged DmSNAP43 was assayed by immunoblotting (Fig. 12). As expected, there was no detectable purification of the poorly expressed DmSNAP43(2-125) construct (Fig. 12, lane 4, bottom panel); furthermore, no co-eluting DmSNAP190 nor DmSNAP50 could be detected in the elution fractions from that cell line (lane 4, top and middle panels).

However, all other truncated and full-length DmSNAP43 constructs were readily detectable in the FLAG elution fractions (Fig. 12, lanes 1-3 and 5-6 in the bottom panels). The top-most panels 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 interaction with DmSNAP190. By inference, the data

suggest that DmSNAP43 residues between 68 and 172 are likely involved in interaction with DmSNAP190.

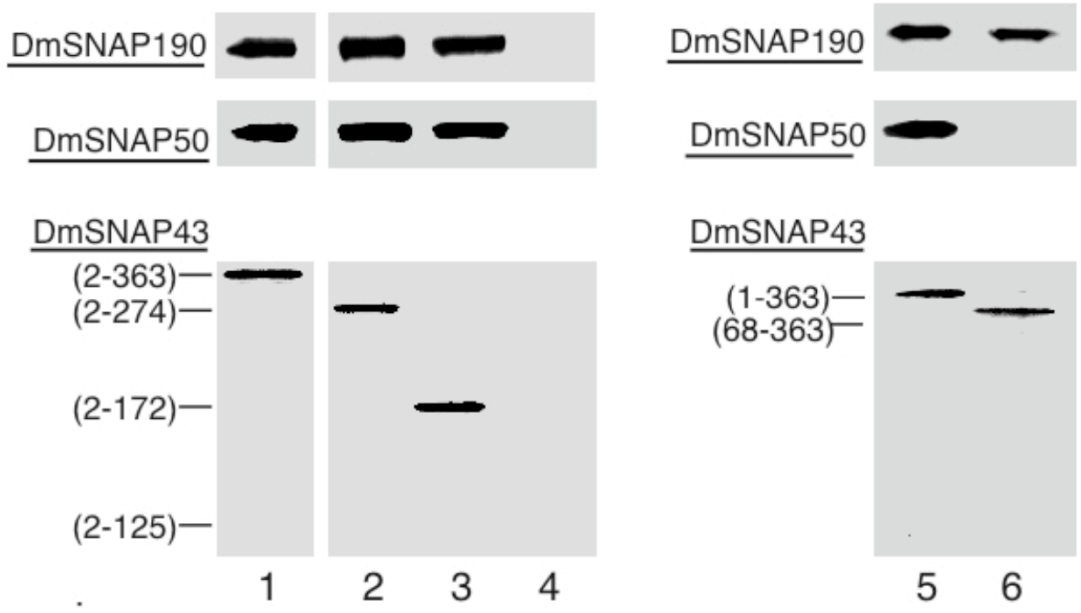


FIG. 12. **Domains of DmSNAP43 required for assembly with DmSNAP190 and DmSNAP50 in vivo.** Complexes containing full length or truncated DmSNAP43 were purified using FLAG antibody beads, and the presence of individual subunits in the elution fractions was assayed by immunoblotting. The volume of elution 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. In all cases, the lanes in each panel (top, middle, or bottom) are from the same gel blot developed for the same length of time so the intensities are directly comparable.

Interestingly, DmSNAP50 co-purified with the two stably-expressed DmSNAP43 C-terminal truncation constructs (Fig. 12, middle panels, lanes 2 and 3) but failed to co-purify with the N-terminal truncation construct DmSNAP43(68-363)

(Fig. 12, middle panels, lane 6). This 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. 13).

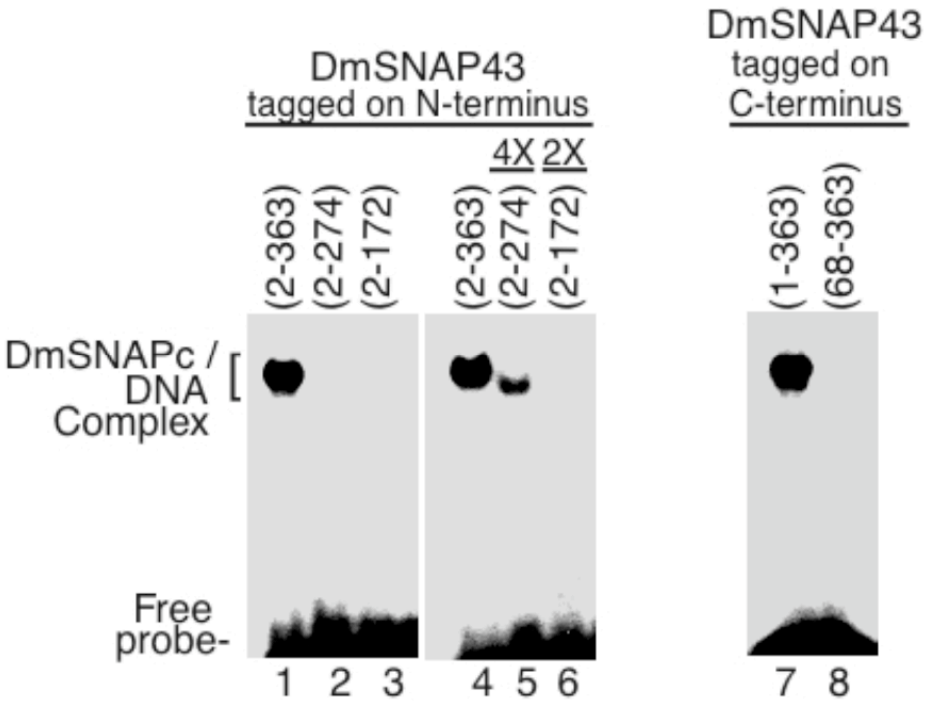


FIG. 13. **Domains of DmSNAP43 required for sequence-specific DNA binding by DmSNAPc.** FLAG-affinity-purified DmSNAP complexes (Fig. 9) were used for DNA mobility-shift analysis with a DNA probe containing a U1 PSEA sequence. Complexes containing DmSNAP190 constructs tagged at the N-terminus (C-terminal 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 DmSNAP190 as determined in Fig. 3, 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.)

N-terminal-tagged full-length DmSNAP43 bound efficiently to the PSEA (lanes 1 and 4), but the construct that was truncated following residue 274 bound DNA only very weakly (lanes 2 and 5). Furthermore, the construct that was truncated following position 172 had no detectable DNA binding activity (Fig. 13, lanes 3 and 6). 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-terminally tagged DmSNAP43 bound efficiently to DNA (Fig 10, lane 7). However, no DNA-binding activity was detectable following the deletion of the N-terminal 67 amino acid residues of DmSNAP43 (Fig. 13, lane 8). This was to be expected since DmSNAP50 was not complexed with this N-terminally truncated DmSNAP43 (Fig. 12, lane 6).

DISCUSSION

Funtional domains of DmSNAPc subunits - Fig. 14 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.

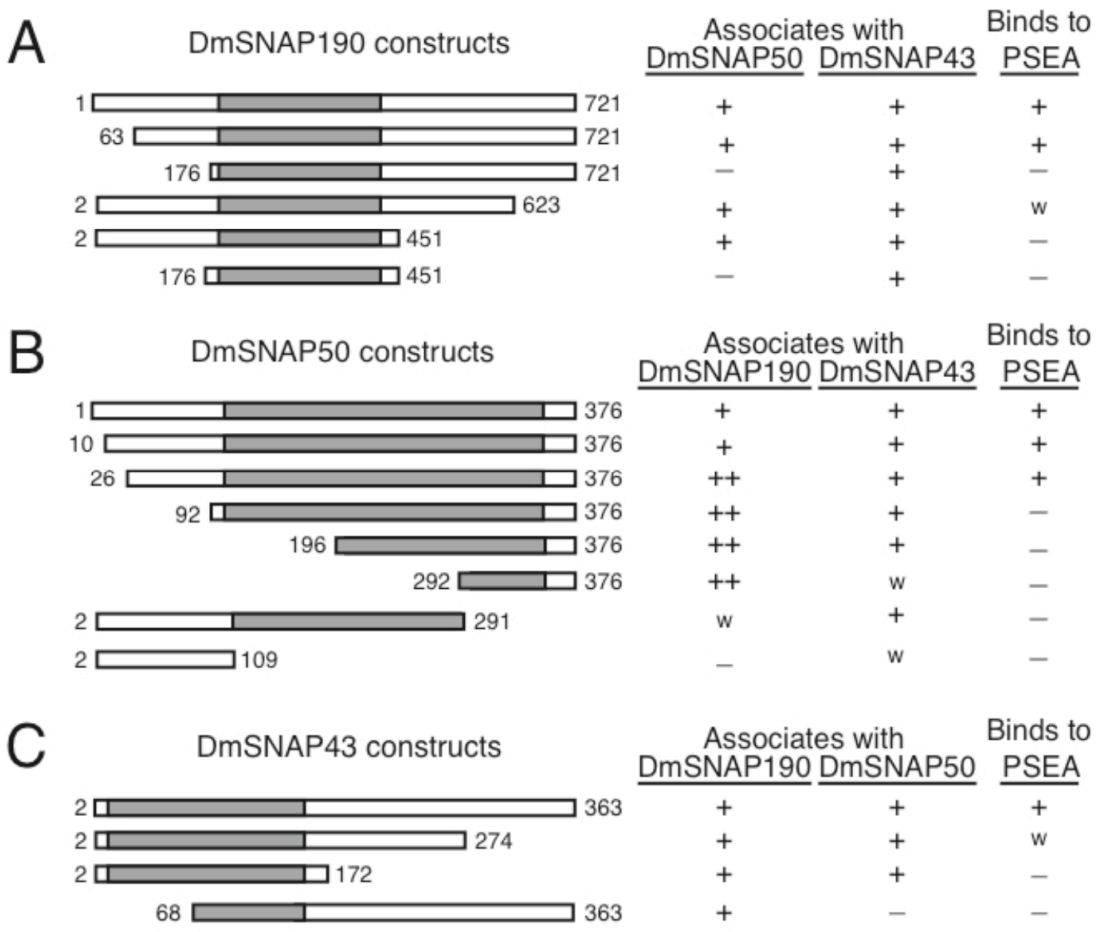


FIG. 14. **Summary of the effects of DmSNAP truncations on DmSNAPc assembly and DNA binding activity.** The various DmSNAP truncations analyzed are shown on the left side of the figure. The columns on the right indicate whether the specific truncation construct can associate with the other two DmSNAP subunits and whether the DmSNAP complex that contains the truncation construct can bind sequence-specifically to the PSEA. A "w" indicates weak association or binding.

Furthermore, DmSNAP190(176-451), which contains the conserved Myb domain, can associate with DmSNAP43. However, the conserved Myb domain of DmSNAP190 is insufficient for its assembly with DmSNAP50. 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 most cases, however, assembly of the DmSNAPc subunits into the complex is mediated primarily by the regions of the proteins that are most evolutionarily conserved.

Furthermore, all DmSNAP mutations that impaired complex assembly also destroyed DNA binding activity. This was to be expected since all three subunits are required for sequence-specific DNA binding by metazoan SNAPc (19-21, our unpublished observations).

Interestingly, 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. 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 26 and 92 resulted in the 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 non-evolutionarily 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(25), suggesting that interactions between the protein and the DNA are more free to 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 (19-22,24). 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.

Fly SNAP190 vs. human SNAP190 - Figs. 6 and 14 show that the conserved Myb domain of DmSNAP190 interacts strongly with DmSNAP43. In contrast, no interaction was observed between HsSNAP43 and the conserved Myb domain of HsSNAP190 (24). 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 (24). These findings may represent real differences between the fly and human systems as no ortholog of HsSNAP19 has been identified in the fruit fly. The

strong interactions of DmSNAP43 with the Myb domain of DmSNAP190 may be replaced in the human system by interactions involving HsSNAP19 (and HsSNAP43) with residues 84-138 of HsSNAP190 (24).

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. 6 and 14). 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 (11,19,21,24).

Even more surprisingly, we found that the non-conserved C-terminal domain of DmSNAP190 is essential for the binding of DmSNAPc to the PSEA (Fig. 7). This was unexpected because the C-terminal domain of HsSNAP190 is completely dispensable for sequence-specific DNA binding in the human system (19,20,22,24,26). In fact, the C-terminal domain of HsSNAP190 serves to dampen DNA binding by HsSNAPc (19). 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 vs. 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. 9 and 14) 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 strong interaction with DmSNAP190. Jawdekar et al. (21) 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 appeared to interact, albeit more weakly, with DmSNAP43. To examine the possibility that either DmSNAP43 or DmSNAP190 was recruited into the complex by mutual interactions with each other rather than by interactions with the SNAP finger domain of DmSNAP50, we expressed truncated DmSNAP50(292-376) in pair-wise combinations with either DmSNAP190 or DmSNAP43 alone. In each case, we observed that DmSNAP190 and DmSNAP43 co-precipitated with FLAG-tagged DmSNAP50(292-376) (data not shown). Thus the SNAP finger domain of DmSNAP50 can interact independently with both DmSNAP190 and DmSNAP43. The SNAP finger appears to be the major domain contributing to the association of DmSNAP50 with DmSNAP190. On the other hand, the central region of SNAP50 (residues 196-291) appears to be the more dominant region contributing to SNAP50-SNAP43 association in both humans (21) and flies (Figs. 9 and 14).

It was anticipated that the C-terminal truncation [DmSNAP50(2-291)] that deleted a portion of the SNAP finger domain would eliminate DNA-binding activity since the association of DmSNAP190 with this complex was impaired. However, DNA binding activity was also ablated by a DmSNAP50 N-terminal truncation that

deleted amino acids 26-91 even though this construct assembled efficiently with DmSNAP190 and DmSNAP43 (Figs. 10 and 14). 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 human SNAP50 have to our knowledge not been reported.

Fly SNAP43 vs. human SNAP43 - Results shown in Figs. 12 and 14 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 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 for interaction with DmSNAP50 and that the first 68 amino acids of DmSNAP43 are required for this interaction. Nearly identical results were obtained in the human system by Ma and Hernandez (24) who employed very similar truncations of HsSNAP43. On the other hand, Hinkley et al. (20), 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 GST pulldown assay. It is possible that the conditions utilized by

us and by Ma and Hernandez are more stringent and detect only stronger interactions among the SNAP subunits.

Deletion of DmSNAP43 amino acids beyond position 274 reduced but did not eliminate sequence-specific binding of DmSNAPc to the PSEA. However, DNA binding activity was undetectable following further truncation of DmSNAP43 to position 172 (Figs. 13 and 14). Results essentially identical to these were obtained by Ma and Hernandez (24) when they used truncations that were analogous to our truncations of DmSNAP43. However, Hinkley et al. (20) 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 Herrnandez.

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. 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 the 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 Luy, 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.USA* **93**, 4289-4293

9. Henry, R. W., Ma, B. C., Sadowski, C. L., Kobayashi, R., and Hernandez, N. (1996) *EMBO J.* **15**, 7129-7136

10. Bai, L., Wang, Z. X., Yoon, J. B., and Roeder, R. G. (1996) *Mol.Cell.Biol.* **16**, 5419-5426

11. Wong, M. W., Henry, R. W., Ma, B. C., 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. C., 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**(5), 1897-1906

15. Das, A., and Bellofatto, V. (2003) *Proc.Natl.Acad.Sci.USA* **100**(1), 80-85

16. Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A., and Bellofatto, V. (2005) *Mol.Cell.Biol.***25**(16), 7314-7322

17. Schimanski, B., Nguyen, T. N., and Gunzl, A. (2005) *Mol.Cell.Biol.***25**(16), 7303-7313

18. Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L. C., and Stumph, W. E. (1997) *Eur.J.Biochem.***248**, 231-237

19. Mittal, V., Ma, B. C., and Hernandez, N. (1999) *Genes Dev.* **13**(14), 1807-1821

20. Hinkley, C. S., Hirsch, H. A., Gu, L. P., LaMere, B., and Henry, R. W. (2003) *J.Biol.Chem.***278**(20), 18649-18657

21. Jawdekar, G. W., Hanzlowsky, A., Hovde, S. L., Jelencic, B., Feig, M., Geiger, J. H., and Henry, R. W. (2006) *J.Biol.Chem.***281**(41), 31050-31060

22. Ma, B. C., and Hernandez, N. (2002) *Mol.Cell.Biol.* **22**(22), 8067-8078

23. Zamrod, Z., Tyree, C. M., Song, Y., and Stumph, W. E. (1993) *Mol.Cell.Biol.* **13**, 5918-5927

24. Ma, B., and Hernandez, N. (2001) *J.Biol.Chem.* **276**(7), 5027-5035

25. Hernandez, G., Valafar, F., and Stumph, W. E. (2007) *Nucleic Acids Res.***35**(1), 21-34

26. Hanzlowsky, A., Jelencic, B., Jawdekar, G., Hinkley, C. S., Geiger, J. H., and Henry, R. W. (2006) *Protein Expr.Purif.*

CHAPTER 2

Mutagenesis of *Drosophila melanogaster* to Alter RNA Polymerase Specificity at an snRNA Gene Promoter

ABSTRACT

Our lab is interested in obtaining a molecular understanding of how RNA polymerase specificity is determined at snRNA gene promoters. While the U1 and U6 snRNA genes have similar promoter structures, the U1 gene is transcribed by RNA polymerase II but the U6 gene is transcribed by RNA polymerase III.

In *Drosophila melanogaster*, the RNA polymerase specificity of snRNA genes is determined by the proximal sequence element "A" (PSEA), a promoter element 40-65 base pairs upstream of the transcription start site. Both the U1 and U6 PSEAs are 21 base pairs in length and vary at only a few positions. Moreover, the PSEAs of the U1 and U6 snRNA promoters bind the exact same basal transcription factor, DmSNAPc. However, DmSNAPc binds to the U1 PSEA in a different conformation for Pol II transcription than it does to the U6 PSEA for Pol III transcription.

We have introduced different versions of the U1 promoter, driving a neomycin resistance gene, into the genome of live fruit flies. These transgenes contained U6-like PSEAs that inactivated the U1 promoter. We then attempted

to induce mutations in DmSNAPc (or other transcription factors) using ethyl methane sulfonate to bring about a change in RNA polymerase specificity at the U1 promoter as evidenced by the flies' new-found resistance to G418, a neomycin analog. One fly line that exhibited unusual developmental behavior had increased neomycin resistance, but we were unable to show that it was due to the expression of the neomycin resistance gene.

INTRODUCTION

The overall goal of the work done in our lab is to understand the molecular mechanisms and underlying structural differences that contribute to RNA polymerase specificity at the promoters of small nuclear RNA genes transcribed by RNA polymerase II (Pol II) or RNA polymerase III (Pol III) in *Drosophila melanogaster*. The products of these genes, the small nuclear RNAs, are an integral part of the spliceosome, which is an RNA-protein complex that removes introns from the primary transcript of messenger RNA in eukaryotes and splices the exons together to form a continuous sequence specifying a functional polypeptide(1,2).

While the U1 and U6 snRNA genes have similar proximal sequence elements (PSE) in their promoters, and use exactly the same proximal sequence binding protein (SNAPc) as a basal transcription factor, the U1 gene is transcribed by Pol II but the U6 gene is transcribed by Pol III (3,4). Previous work from our lab has shown that in *D. melanogaster* the RNA polymerase specificity of the promoter *in vitro* is clearly determined by the PSEA (5,6). The proximal sequence binding protein DmSNAPc consists of three subunits and binds to the U1 PSEA in a different conformation than it does to the U6 PSEA (7-9). For the normal expression of the U1 gene, wild type

DmSNAPc recognizes the U1 PSEA sequence and binds to it in a certain conformation. This in turn recruits Pol II transcription factors and Pol II. In much the same way, for the normal expression of the U6 gene, the exact same DmSNAPc recognizes the U6 PSEA sequence and binds in a different conformation than the DmSNAPc bound to the U1 PSEA. This in turn recruits Pol III transcription factors and Pol III. Our working model to explain the RNA polymerase specificity of *Drosophila* snRNA gene promoters based upon conformational differences of DmSNAPc is schematically diagrammed in Fig. 15.

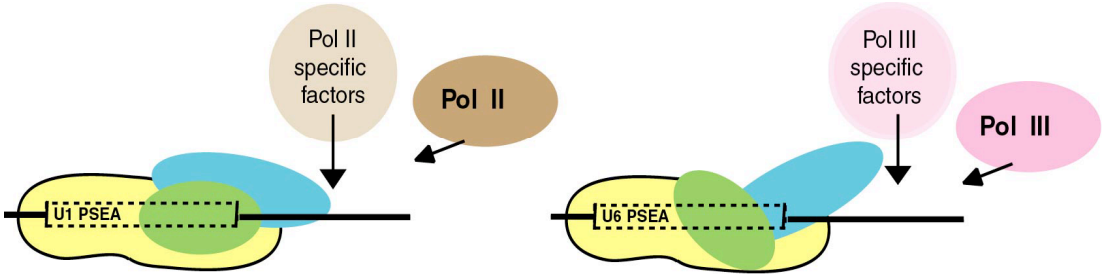


FIG. 15. **Schematic model of the recruitment of Pol II or Pol III to the U1 and U6 snRNA promoters.**

EXPERIMENTAL PLAN

The specific goal of the project described in this chapter was to generate mutations in DmSNAPc that alter the interaction of DmSNAPc with the PSEA so that the polymerase specificity of the snRNA promoter is switched from Pol III to Pol II and to characterize these mutations. The strategy that was chosen involves *in vivo* random mutagenesis of transgenic fly lines that contain a heterologous neomycin resistance (*neo)* gene under the control of various (inactive) forms of the U1 gene

promoter. A genetic selection for mutants displaying relaxed polymerase specificity would be carried out by using the U1 promoter to drive the expression of the neomycin resistance gene. A chimeric U1-*neo* gene was introduced into the germline of flies by P-element transformation. In the chimeric construct the wild-type PSEA of the U1 promoter was replaced by a hybrid U1/U6 PSEA sequence that fails to promote efficient Pol II transcription *in vivo*. Previous experiments from our lab had shown that when the transgene contained a wild-type U1 PSEA, the *neo* gene was efficiently expressed and the flies were resistant to G418, a neomycin analog. If the transgene contained the U6 PSEA, the *neo* gene was not expressed and the flies were killed by G418 at the larval stage. Flies containing genes with hybrid U1/U6 PSEAs exhibited either low or intermediate levels of resistance. Thus G418 potentially provided a selection tool to obtain mutant flies with an altered or relaxed RNA polymerase specificity.

My plan was to perform *in vivo* random mutagenesis of G418-sensitive transgenic fly lines using ethyl methane sulfonate (EMS). This would be followed by selection on food that contains G418. Mutated transgenic flies that survived doses of G418 in their food that were lethal to the parental line would potentially have a mutation we were interested in. The mutation could be *cis*-acting, meaning that a mutation had occurred in the U1-*neo* promoter, most likely the hybrid PSEA. This would be of marginal interest for further research. Or, the mutation could be *trans*-acting meaning that the mutation had possibly occurred in a gene that codes for one of the subunits of DmSNAPc. This is the type of mutation that we were most interested

in obtaining, because such a mutation would likely reveal information about amino acids in DmSNAPc that regulate DmSNAPc conformation and RNA polymerase specificity. Alternatively, the mutation could be in some other unknown or unanticipated factor, and this could inform us of unknown factors involved in specific RNA polymerase recruitment.

To prove that the increased resistance of the flies to G418 is due to increased expression of the *neo* gene, it is necessary to show that there are increased levels of *neo* mRNA in the resistant line. If the *neo* mRNA is present in the mutated flies, but not present in the unmutated parental, line we would have evidence that a mutation had occurred to allow the *neo* gene to be expressed. This would indicate that mutations have occurred most likely in either the DmSNAPc or in the promoter itself. On the other hand, if no *neo* mRNA was detected we could surmise that the cause of the resistance to G418 was due to some other factor (e.g., a mutation in a cell membrane transport protein) and not due to the presence of *neo* mRNA.

Once a bona-fide G418-resistant line was established and we had eliminated the possibility of a cis-acting mutation we would use PCR to clone the genes for DmSNAPc or any other known transcription factors from the DNA of the mutant fly line and compare the sequences to the genes of the non-resistant parental strain of flies.

RESULTS

Three different constructs were cloned into the standard P-element vector pP(caSpeR-4) that can be used to incorporate heterologous genes into the fruit fly genome (10). Each construct carries a different PSEA, a variation of the U1 PSEA with certain U6 nucleotide characteristics. These lines were named 11161, 11166, and 11666. The ones and sixes denote whether a U1 or a U6 nucleotide resides in the PSEA at positions 7, 14, 16, 19, and 20 (Fig. 16).

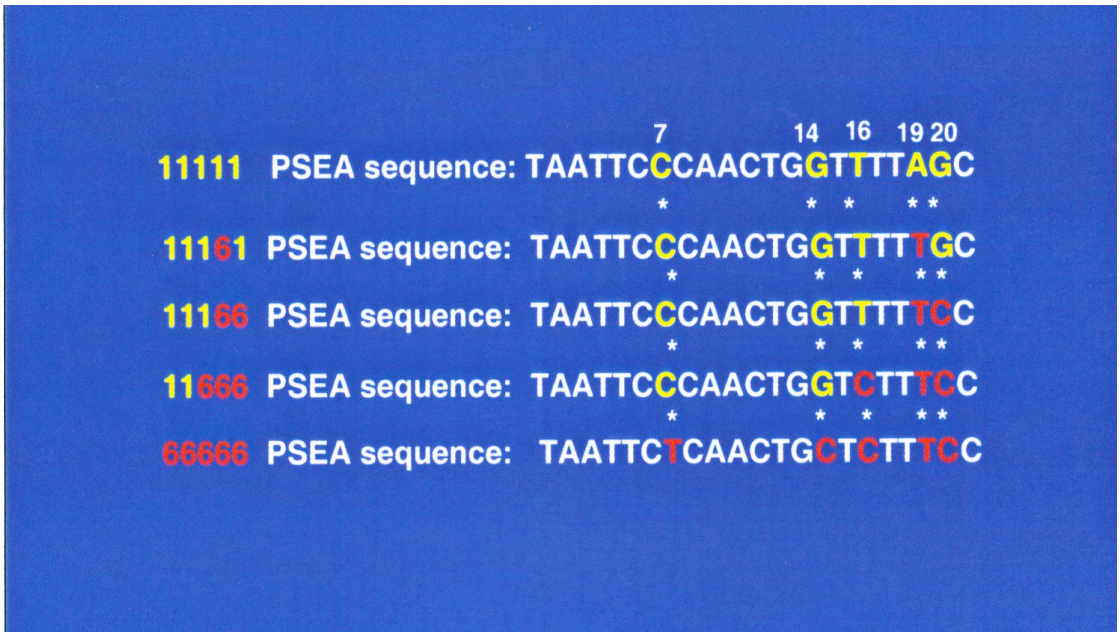


FIG. 16. **Variations of the U1 PSEA containing U6 characteristics.** The U1 and U6 PSEAs are both 21 base pairs in length and differ at only five positions: 7, 14, 16, 19, and 20. The yellow nucleotides indicate U1-specific nucleotides. The red colored nucleotides indicate U6 nucleotides substituted for U1 nucleotides in each construct.

Each construct also carries both a neomycin resistance gene and a white^{+} gene for eye color (Fig. 17). In all three constructs the *neo* gene is under the control of the modified U1 promoter. These constructs were then introduced into the genome of the yellow/white (yw) strain of *Drosophila* by microinjection into embryos. The newly laid embryos were collected and lined up on slides and placed under a microscope. The P-element vector plasmids containing the U1 constructs were then injected into the embryos in the area where the germ cells develop using a special needle attached to a micromanipulator for very precise movement. As the germ cells develop, the injected DNA is incorporated into the genome. Since the construct is contained in only the germ cells of the injected flies, those flies would not exhibit the orange-eye phenotype conferred by the white+ gene incorporated into the construct. They will all have white eyes. The offspring of these flies will have the construct including the *neo*-resistant gene and white+ gene in all of their cells. Fewer than ten percent of the fly embryos injected actually take the construct up into their genomes. So, the flies with the incorporated construct will have orange eyes. The others will have white eyes. In this way, we were able to select the few transgenic flies from the many that do not have the construct.



FIG. 17. **Relevant portion of the P-element vector used to generate transgenic flies for the selection of mutants with altered RNA polymerase specificity.** The vector contains a U1 promoter with a hybrid U1/U6 PSEA which drives a *neo* gene. The vector also contains a white+ gene for identification of transgenic flies.

From this process we were able to produce 21 stable lines of flies. These flies were then bred in a series of crosses to make them homozygous for the construct. With two copies of the white+ gene their eyes became darker orange. Experiments were then carried out to determine which chromosome, 2, 3, or X, had incorporated the construct in each line of flies. Of these 21 lines, nine lines had the insert on the second chromosome, seven had the insert on the third chromosome, and one line had the insert on the X chromosome. We were unable to determine the chromosome of the inset on four of the lines. Of the nine lines selected for further testing, one line, 11166/3-9X had the construct on the X chromosome. Three lines, 11166/21-6, 11666/17-10, and 11161/19-7 had the construct on the second chromosome, and four lines, 11166/26-14, 11666/17-3D, 11666/28-13, and 11161/19-5 had the transgene on the third chromosome. For one line of flies, 11161/27-2 we were unable to determine with certainty which chromosome carried the construct. Since the chromosome markers used to determine which chromosome had the insert are only able to show if the insert is on the large second or large third chromosome, it is possible that the construct had inserted onto the small fourth chromosome.

In order to effectively select mutated flies for relaxed RNA polymerase specificity, it was necessary to determine the minimal concentration of G418 that would be sufficient to kill all or nearly all of the offspring of each line of flies. So, if EMS-mutated flies survived a substantially higher concentration of G418 than that same line before it was mutated, we would have evidence that the RNA polymerase specificity had been relaxed. Therefore, we carried out titration curves with nine

different transgenic fly lines. Each fly line was subjected to four different concentrations of G418: 0, 200, 400, and 600 micrograms of G418 per milliliter of water in their food.

Fig. 18 illustrates the results for each fly line as a percentage of flies surviving compared to the number of flies that eclosed in the same time period that had no G418 in their food. In general it is evident that the lines of flies in which the inserted construct carried more U1 character (ie 11161) better survived higher concentrations of G418 than did the lines in which the construct had more U6 character (11666) as expected. In the one line for which this is not true (11161/27-2), perhaps the transgene had inserted into a region of chromatin where it could not be well expressed. Six promising lines were selected for further testing of G418 tolerance. Because of the generally high resistance of the 11161 lines, none of those lines were selected for further study.

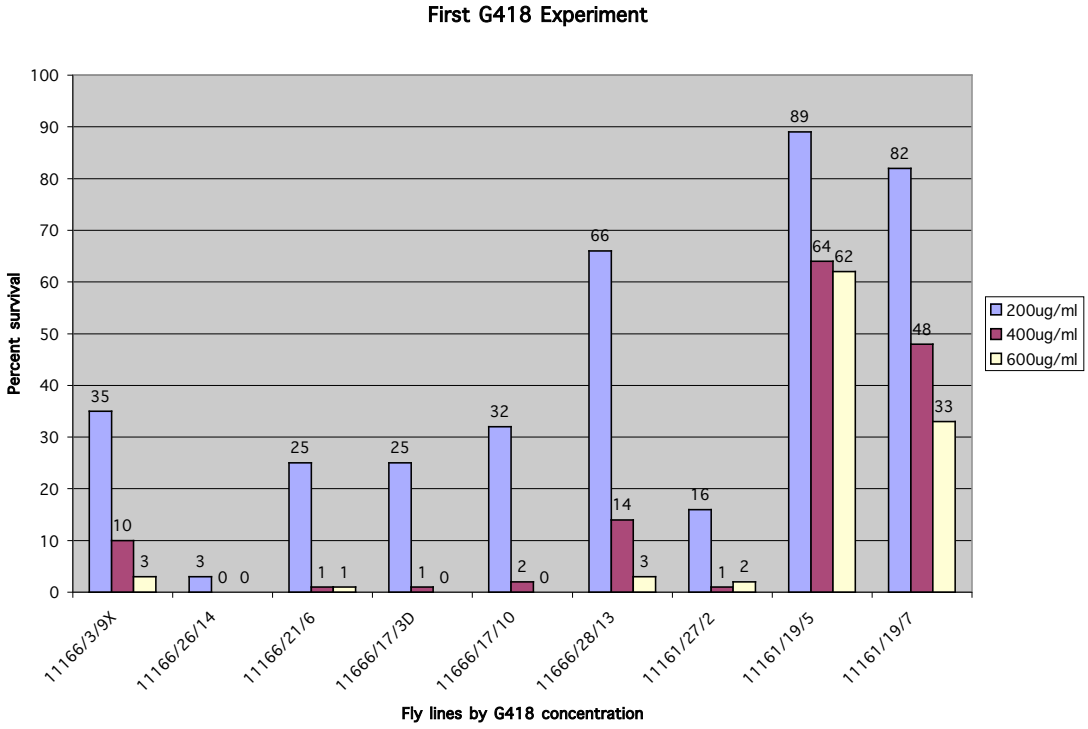


FIG. 18. **Survival of flies on different concentrations of G418.** The numbers at the top of each column indicate the percentage of flies surviving that particular concentration compared to the number of flies of the same line that eclosed on food containing no G418.

For the next stage of G418 resistance testing the selected fly lines were exposed to six concentrations of G418: 000, 200, 400 600, 800, and 1000 micrograms of G418 per milliliter of food. The idea was to choose concentrations that would barely kill all the flies, so even with a small amount of added resistance from a mutation some would survive. This helped to "zero in" on the best G418 concentration to use in selecting the mutated flies from each line. Fig. 19 shows the results for each of the six fly lines tested. Four lines of flies were chosen from these lines to participate in the mutagenesis experiment and the concentration of G418 best suited for selection of each line was determined. The selected fly lines were 11166/3-9X at 900ug/ml, 11166/21-6 at 800ug/ml, 11666/17-3D and 1166/17-10 at 600ug/ml of G418.

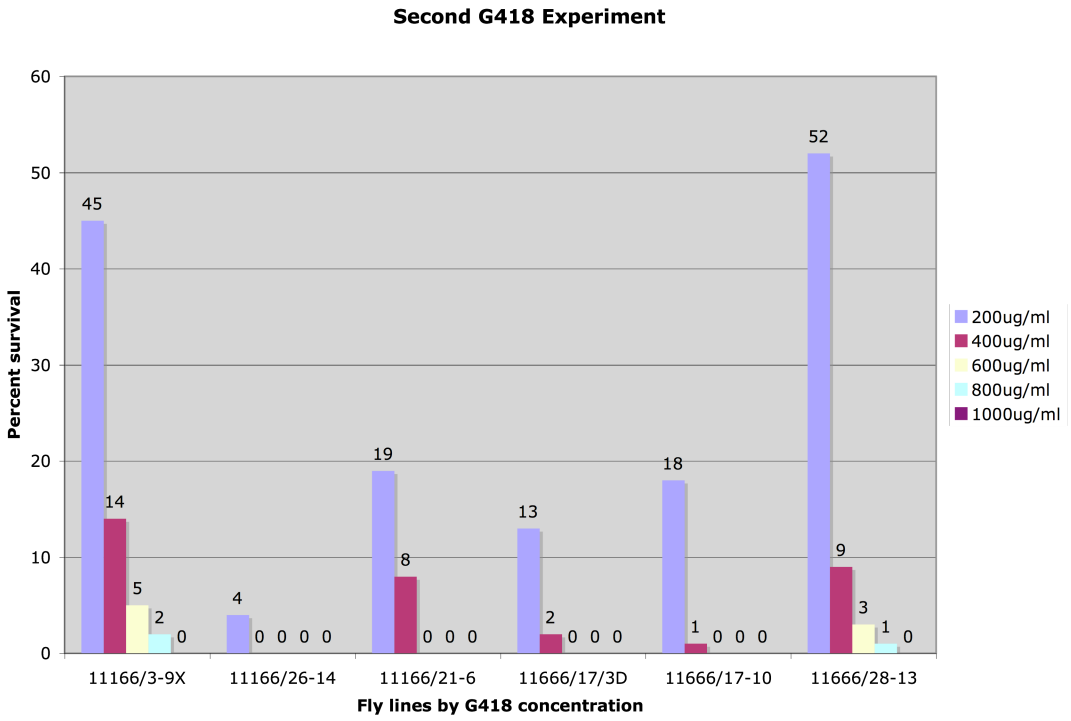


FIG. 19. **Survival of flies on different concentrations of G418.** The numbers at the top of each column indicate the percentage of flies surviving that particular concentration compared to the number of flies of the same line that eclosed on food containing no G418.

The transgenic flies containing these constructs were exposed to ethyl methane sulfonate. EMS is a strong mutagen which is known to cause G:C to A:T transition mutations and is thought to induce only point mutations in *Drosophila*. The optimum concentration of EMS was determined through previous experiments to be 37.5mM--not enough to kill too many flies, but enough to ensure many mutated flies. Our hope was that this would cause one or more point mutations in the genes coding for DmSNAPc that would cause the protein complex to bind to the U6-like PSE in the U1 promoter in a conformation that will recruit Pol II instead of Pol III thus driving effective expression of the neomycin resistance gene. Larval offspring of the mutagenized flies were then exposed to G418 in their food.

One hundred male flies from each line were exposed to EMS for 24 hours in each round of mutation. Male flies were used because they constantly produce sperm cells which are subject to mutation, so only their progeny would show the mutation phenotype. The mutated male flies were then mated to virgin female flies from the same line. Ten mutated males and 20 females were placed in bottles containing food without G418 for one generation. The second generation of flies was then placed in bottles containing food with the determined concentration of G418. Generally, a few flies in each line survived the first selection by G418. These flies were subjected to a second selection using food with a 20% higher concentration of G418. Rarely, a few flies survived the second selection, and they were subjected to a third selection with a still higher dose of G418. Fig. 20 shows a typical flow chart for one round of mutations.

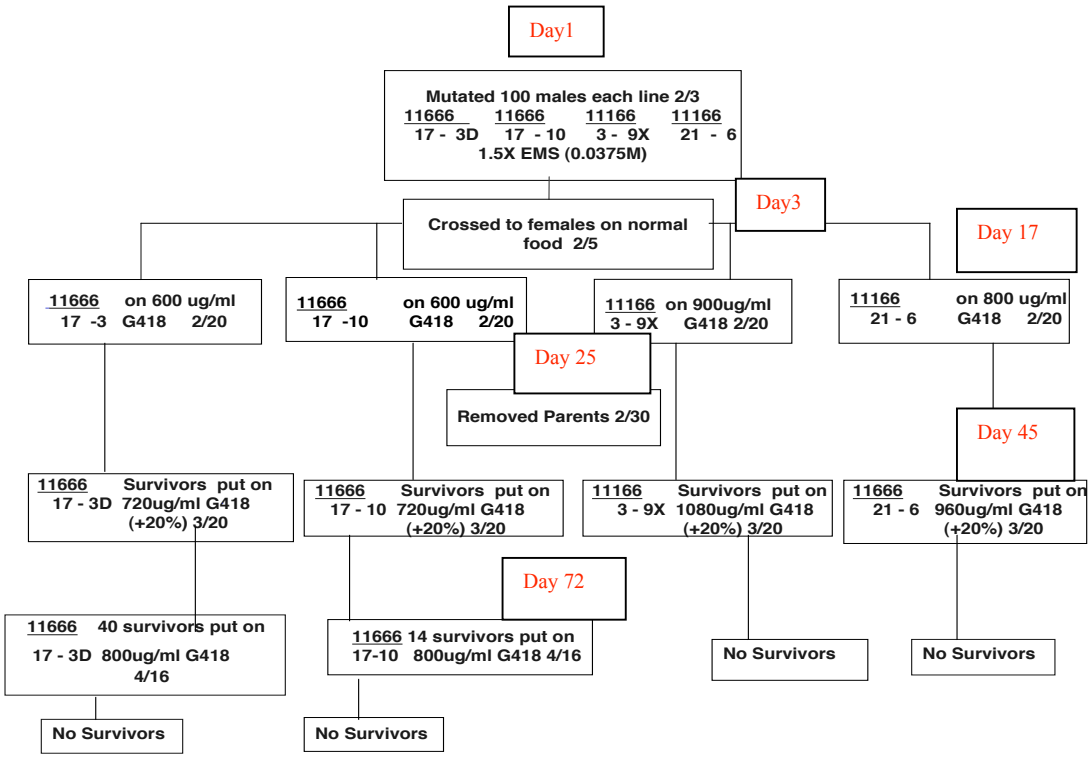


FIG. 20. **Flow chart for one round of mutations.** The flow chart illustrates the timeline for one complete round of mutations. The numbers on the left indicate the fly line involved. The middle numbers indicate the concentration of G418, and the date is indicated in the lower left hand corner of each box. A new round of mutations was started each week for about one year.

After over a year of exposing 100 flies per week to EMS, we obtained only one line of flies that was consistently and stably resistant to G418. However, this mutant line of flies, which arose from the parental line 11666/17-3D, had unusual developmental characteristics described further below.

To assess the appearance of *neo* mRNA in the resistant fly line, we first attempted to isolate poly (A) mRNA from fly larvae by oligo dT cellulose

chromatography. The *neo* mRNA would then be detected by northern blot or primer extension assay. The procedure was performed twice using control lines of flies, but we were unable to extract enough mRNA for a northern blot or primer extension assay. The next approach was to extract total RNA from the larvae of the four lines of flies to be tested for *neo* RNA. These included the yw and U1/68 as negative and positive controls respectively. The "wild type" yw line should make no *neo* RNA, whereas the U1/68 should make abundant *neo* RNA from the wild type U1 reporter. Also RNA was isolated from the unmutated 11666/17-3D parental line, and the 11666/17-3D mutated flies that were resistant to G418. A phenol, chloroform, isoamyl alcohol extraction was performed followed by precipitation in isopropanol. This process yielded enough RNA to proceed with the primer extension experiment. The smaller larvae from the mutated line however yielded only enough RNA for one primer extension experiment. No *neo* mRNA was detected in either the control lines or the mutated line in primer extension experiments. Moreover, for some reason the mutated flies stopped reproducing, so the project was abandoned.

DISCUSSION

The goal of this project was to induce mutations in DmSNAPc and/or other *trans*-acting factors that result in altered or relaxed RNA polymerase specificity at U1 and U6 snRNA gene promoters, and to characterize these mutations. Clearly, some sort of mutation was induced in the 11666/17-3D line by exposure to EMS making the flies resistant to G418, for a few generations anyway. This became apparent during the

collection of larvae for the total RNA extraction. For the control lines of flies the collection of larvae was routine, but the mutated line of flies appeared to have developed an unusual characteristic probably connected to the mutation that caused resistance to G418. The larvae did not develop as do normal fly larvae. They crawled out of the food upon hatching instead of staying in the food and developing to an advanced larval stage. This was baffling. There were normal pupal cases and normal adult flies in the bottle, but only very tiny (not normal) larvae clinging to the side of the bottle. After several generations, this line of flies stopped reproducing. Had this line of flies become resistant to G418 due to accidental cross breeding with a resistant line, they should not have exhibited the unusual developmental characteristics that they did.

It was decided that further attempts to EMS mutagenize live flies to obtain a mutation in DmSNAPc would be too time consuming to be of value, so the project was stopped.

REFERENCES

1. Sharp, P. A. (1994) *Cell* **77**, 805-816

2. Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendewey, D., and Keller, W. (1988) Functions of the abundant U-snRNPs. In: Birnsteil, M. L. (ed). Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, Springer Verlag KG, Heidelberg, Federal Republic of Germany

3. Dahlberg, J. E., and Lund, E. (1988) The genes and transcription of the major small nuclear RNAs. In: Birnstiel, M. L. (ed). Structure and Function of Major

and Minor Small Nuclear Ribonucleoprotein Particles, Springer Verlag KG, Heidelberg, Federal Republic of Germany

4. Lobo, S. M., and Hernandez, N. T. (1994) Transcription of snRNA genes by RNA polymerases II and III. In: Conaway, R. C., and Conaway, J. W. (eds). Transcription: Mechanisms and Regulation, Raven Press, New York, NY

5. Jensen, R. C., Wang, Y., Hardin, S. B., and Stumph, W. E. (1998) *Nucleic Acids Res.* **26**, 616-622

6. McNamara-Schroeder, K. J., Hennessey, R. F., Harding, G. A., Jensen, R. C., and Stumph, W. E. (2001) *J.Biol.Chem.* **276**(34), 31786-31792

7. Wang, Y., and Stumph, W. E. (1998) *Mol.Cell.Biol.* **18**, 1570-1579

8. Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (2004) *Mol.Cell.Biol.* **24**(5), 1897-1906

9. Hardin, S. B., Ortler, C. J., McNamara-Schroeder, K. J., and Stumph, W. E. (2000) *Nucleic Acids Res.* **28**(14), 2771-2778

10. Rubin, G. M., and Spradling, A. C. (1982) *Science* **218**, 348-353

CHAPTER 3

Can the RNA Polymerase Specificity of an snRNA Promoter Be Altered by the Binding of a Synthetic Polyamide Compound?

ABSTRACT

While the promoters of the U1 and U6 snRNA genes are similar and both are recognized by the same proximal sequence binding protein, DmSNAPc, the U1 gene is transcribed by RNA polymerase II, but the U6 gene is transcribed by RNA polymerase III. DmSNAPc, which consists of three distinct subunits, binds to the U1 PSEA in a different conformation for recruitment of Pol II than it does to the U6 PSEA for recruitment of Pol III. In the U1 conformation of DmSNAPc, one of the three subunits makes extensive contact with DNA downstream of the U1 PSEA whereas in the U6 conformation the subunit only contacts a few nucleotides downstream of the U6 PSEA. Moreover, there are distinct differences in the protein-DNA interactions of DmSNAPc with the 3' end of the U1 and U6 PSEAs.

Here I have used a synthetic polyamide compound designed to bind to a recognition site inserted within or downstream of the U1 PSEA in an attempt to change the polymerase specificity of the U1 promoter by artificially changing the conformation of DmSNAPc bound to the DNA.

INTRODUCTION

The overall goal of the work in our lab is to understand the molecular mechanisms and underlying structural differences that contribute to RNA polymerase specificity at the promoters of small nuclear RNA genes transcribed by RNA polymerase II (Pol II) or RNA polymerase III (Pol III) in *Drosophila melanogaster*. While the promoters of the U1 and U6 snRNA genes are similar and both are recognized by the same proximal sequence binding protein, DmSNAPc, the U1 gene is transcribed by Pol II, but the U6 gene is transcribed by Pol III (1,2). As stated earlier, the PSEAs of the U1 and U6 snRNA gene promoters are both 21 nucleotides long and, in the two genes we work with, differ at only five positions (3). These five differences in the PSEA cause the DmSNAPc to bind to the U1 PSEA in a different conformation than it does to the U6 PSEA (4,5).

The goal of the project described in this chapter was to induce altered RNA polymerase specificity *in vitro* at the U1 promoter by use of a small synthetic polyamide compound that can be designed to bind to specific DNA sequences seven to eight base pairs in length (6). The idea was to force a change in the conformation of DmSNAPc bound to the U1 promoter by binding a small synthetic, sequence specific polyamide compound to the DNA separating the U1 PSEA from the PSEB thus altering the RNA polymerase specificity from Pol II to Pol III.

Pyrrole-imidazole polyamides are a novel class of small molecules that bind predetermined DNA sequences in the minor groove. Sequence-specific DNA

recognition depends on side-by-side pairing of pyrrole (Py) and imidazole (Im) amino acids. A Py opposite an Im targets a C-G base pair, and an Im opposite a Py targets a G-C base pair (7). Py/Py, Py/β-alanine (β) and β/β bind both A-T and T-A base pairs. Replacing one pyrrole of the Py/Py combination with a 3-hydroxypyrrole (Hp) results in the Hp/Py compound preferring T-A over A-T base pairs (8).

Sequence-specific pyrrole-imidazole polyamides can be designed to interfere with transcription factor binding and to regulate gene expression. Polyamides bound adjacent to the recognition sites for TBP, Ets-1, and LEF-1 in the human immunodeficiency virus type 1 (HIV-1) long terminal repeat inhibited transcription in cell free-assays and inhibited viral replication in human blood lymphocytes (9).

For this project we attempted to use a small synthetic polyamide in *in vitro* transcription experiments to attempt to alter RNA polymerase specificity. Previous work in our lab has shown that nucleotide differences at the 3' end of the PSEA are critical for determining RNA polymerase specificity (2). Moreover, previous photo-cross-linking experiments in our lab have shown that DmSNAPc contacts the DNA differently downstream of and at the 3' end of the U1 and U6 PSEAs. In particular, DmSNAP43 makes extensive protein-DNA contacts downstream of the U1 PSEA but not downstream of the U6 PSEA (4,10)). The DmSNAP 190 and the DmSNAP50 subunits also bind differently to the 3'ends of the U1 and U6 PSEAs (5)) (see Fig. 3 of the General Introduction). By placing a polyamide binding site downstream of or overlapping the 3' end of the U1 PSEA, we thought that it might be possible to alter binding of one or more of the DmSNAP subunits forcing them into a conformation

more resembling their U6 conformation and thus recruiting Pol III to the promoter rather than Pol II.

EXPERIMENTAL PLAN

Since the polyamide compounds are difficult to design and synthesize, we decided to use an existing compound and alter the usual sequence in the *in vitro* transcription template between the PSEA and the PSEB to one that is recognized by the HIV-1 inhibiting polyamide. Previous studies in our lab have shown that changes in this sequence have no effect on RNAP specificity and little effect on transcription efficiency (2,11). Because of the sequence of the recognition site for the polyamide, it was also possible to partially overlap the PSEA and PSEB sequences with the polyamide binding site.

The HIV-1 compound, ImPy-β-ImPy-γ-ImPy-β-ImPy-β binds to the DNA sequence WGCWGCW, where W represents either a T or an A, as illustrated in Fig. 21 (9). Dr. Joel Gottesfeld of the Scripps Research Institute was kind enough to provide us with an adequate supply of the HIV-1 polyamide compound.

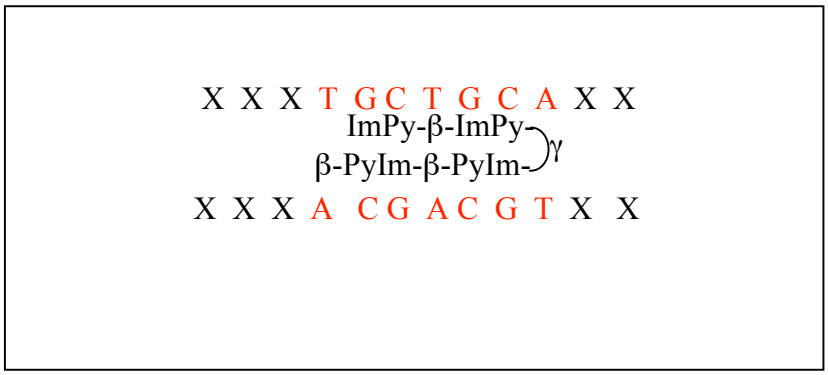


FIG. 21. **The HIV-1 polyamide bound to DNA.** The HIV-1 polyamide binds DNA in a hairpin shape. The γ represents γ-aminobutyric acid which acts as a flexible linker to keep the DNA recognition amide pairs properly lined up.

To introduce the polyamide binding site at various positions in the U1 promoter we initially digested the plasmid that contained the U1 promoter with EcoRI and Sty1 to cut out the PSEA and part of the PSEB. I then ligated in synthetic DNA that reconstituted the PSEA and PSEB with the HIV-1 compound recognition site at various positions.

In vitro transcription reactions were carried out using the pseudo-wild type template and its derivatives that contained the HIV-1 polyamide recognition site at various positions. Reactions also contained various amounts of the HIV-1 compound and sometimes α-amanitin (a Pol II inhibitor) or tagetitoxin (a Pol III inhibitor) to examine RNA polymerase specificity. A detailed protocol for *in vitro* transcription is included in the Appendix.

RESULTS

Six different forms of the U1 promoter sequence containing the HIV-1 recognition site were cloned into pUC118 plasmids to be used as *in vitro* transcription templates. Five of these constructs had the wild type U1 eight base pair spacing between the PSEA and PSEB, and one of them had a twelve-base-pair U6 spacing. Five of these constructs had a wild type U1 PSEA, and one of them had a modified U1 PSEA containing U6 nucleotides at positions 19 and 20. Table 1 shows the sequences of these constructs. In the table, purple font indicates the PSEA, black represents the 8-bp spacer between the PSEA and PSEB, and green is the PSEB sequence. The HIV-1 polyamide binding site is shown in red.

The nomenclature of these constructs reflects the nature of the PSEA, the position of the HIV-1 recognition site and the spacing between the PSEA and PSEB. The first part of the name reflects the nature of the PSEA (U1 or 11166). The second part shows the number of base pairs between the beginning of the PSEA and the beginning of the HIV-1 recognition site (e.g., 21HIV1). The third part of the name shows the number of base pairs between the beginning of the PSEA and the beginning of the PSEB (e.g., 30PSEB).

Initial *in vitro* transcription experiments were carried out using these templates and wild type U1 and U6 templates to determine whether: 1) multiple templates could be

transcribed in one reaction, and 2) the templates with the HIV-1 recognition site between the PSEA and PSEB could support transcription.

Table 1. **DNA sequences of U1 promoters containing HIV-1 recognition sites between the PSEA and PSEB.**

U1-8-PSEB (pseudo-wild-type)
5'TAATTCCCAACTGGTTTTAGC GGTACCGC CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCG CCATGGCG GTACCTTT 5'

U1-25HIV1-30PSEB
5'TAATTCCCAACTGGTTTTAGC TGATGCTG CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCG ACTACGAC GTACCTTT 5'

U1-23HIV1-30PSEB
5'TAATTCCCAACTGGTTTTAGC TTGCTGCA CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCG AACGACGT GTACCTTT 5'

U1-21HIV1-30PSEB
5'TAATTCCCAACTGGTTTTAGT GCTGCAGC CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCA CGACGTCG GTACCTTT 5'

U1-19HIV1-30PSEB
5'TAATTCCCAACTGGTTTTAGC TGCACCGC CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCG ACGTCGTC GTACCTTT 5'

U1-26HIV1-34PSEB
5'TAATTCCCAACTGGTTTTAGC TGATTGCTGCAC CATGGAAA 3'
3'ATTAAGGGTTGACCAAAATCG ACTAACGACGTG GTACCTTT 5'

11166-25HIV1-30PSEB
5'TAATTCCCAACTGGTTTTTCC TGATGCTG CATGGAAA 3'
3'ATTAAGGGTTGACCAAAAAGG ACTACGAC GTACCTTT 5'

Multiple templates can be transcribed in a single in vitro reaction - It is important that reaction conditions are able to support transcription of multiple templates in a single reaction because this allows comparison of results for each template under the exact same conditions. Fig. 22 shows that the wild type U1 and U6 can both be transcribed in the same reaction. Lanes 8-10 show that α-amanitin specifically inhibits U1 transcription by Pol II and that tagetitoxin specifically inhibits

U6 transcription by Pol III. Initial *in vitro* transcription experiments also showed that each of the template constructs containing the HIV-1 recognition site could support transcription (data not shown).

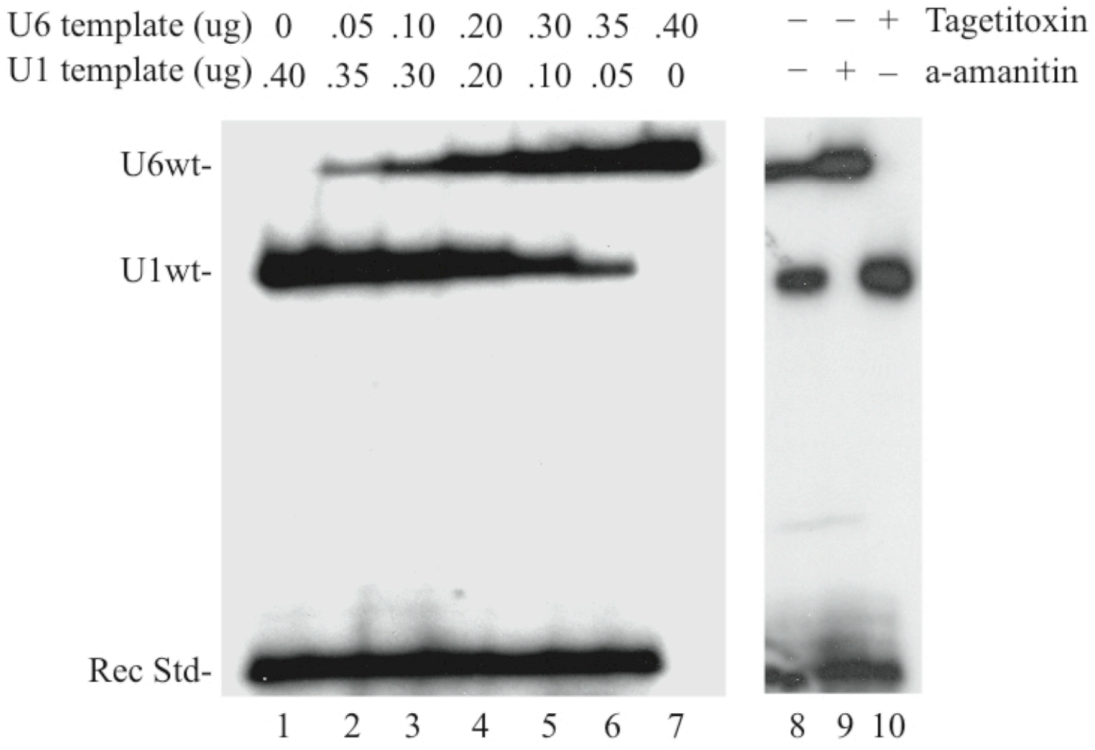


FIG. 22. **Wild type U1 and U6 templates transcribed in the same reaction.** The left panel shows a titration of the amounts of each template in transcription reactions. The right panel shows 0.4 μg of each template in each reaction. In lane 9 transcription of the wt U1(Pol II) template was inhibited by α-amanitin, but transcription of the wt U6 template (Pol III) was not. Lane 10 snows that transcription of the wt U6 template was inhibited by tagetitoxin, but the wt U1 was not.

Effects of the HIV-1 polyamide on in vitro transcription. - Fig. 23 shows results of *in vitro* transcriptions using the pseudo-wild-type template U1-8-PSEB and two templates that contained the HIV-1 polyamide binding site beginning at position

25 or 23. The first lane in each panel shows transcription efficiency of each template in the absence of polyamide, and the subsequent two lanes show the effects of adding 1 or 2 μM polyamide. All three templates were inhibited comparably. Since the U1-8-PSEB template has no identifiable HIV-1 polyamide binding site, this inhibition is probably non-specific. It appears that an HIV-1 polyamide binding site beginning at either position 23 or 25 has no specific effect on transcription in the presence of polyamide.

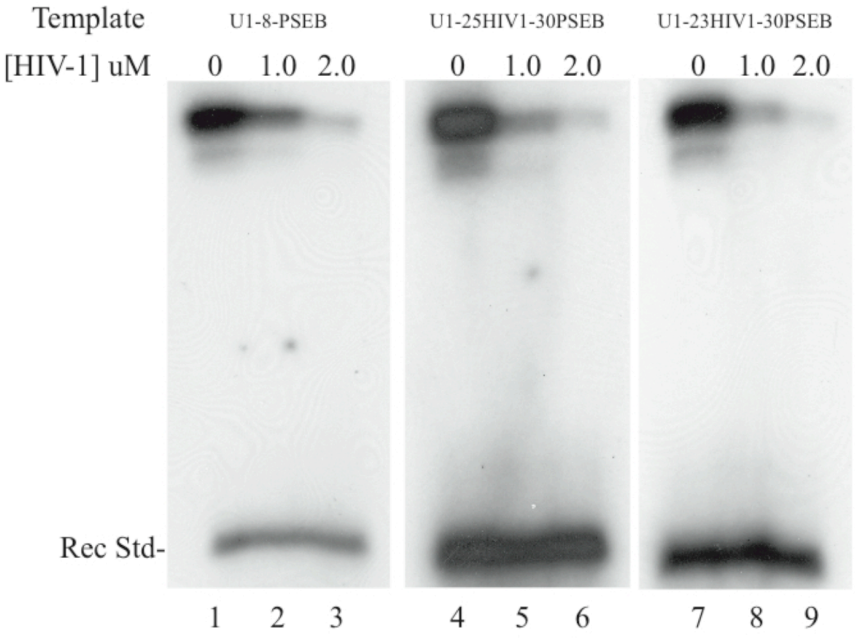


FIG. 23. **Transcription with added HIV-1 polyamide.** Reactions were carried out in the presence of 0, 1.0, and 2.0 μM HIV-1 polyamide compound.

With one possible exception, similar results were obtained for all of the template constructs containing the HIV-1 recognition site as well as another version of the wild type U1 template, and the wild type U6 template (data not shown).

One template that contained an HIV-1 polyamide binding site appeared to be partially resistant to inhibition - A potentially interesting result was obtained with the U1-19HIV1-30PSEB template. Fig. 24 shows the results of co-transcribing this template together with versions of the wild type U1 and U6 promoters that give different lengths of primer extension products. Interpretation of the data is somewhat confused by the fact that the overall levels of transcription among different reactions varied significantly.

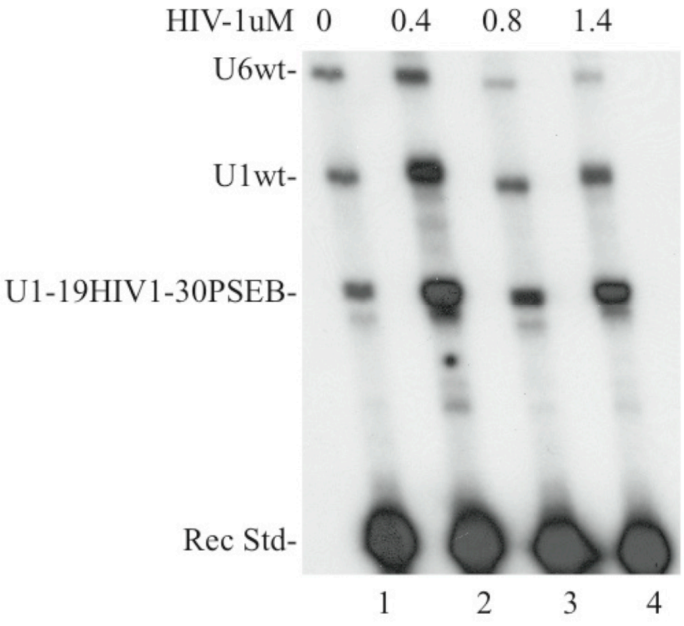


FIG. 24. **Effect of HIV-1 polyamide on transcription of U1-19HIV1-30PSEB, U1wt and U6wt templates.** *In vitro* transcription reactions carried out with 3 templates per reaction for comparison of relative transcription efficiency.

However, the relative levels of transcription of the three templates within each lane should be directly comparable. Without polyamide (lane 1) the relative transcription levels of the three templates were approximately 1:1:1. With 1.4 μM HIV-1 polyamide (lane 4) the relative signals were approximately 0.2:0.4:1 for the U6 wt, U1 wt, and the U1-19HIV1- 30PSEB templates respectively. Intermediate levels of polyamide resulted in intermediate effects on relative transcription levels (lanes 2 and 3). These results suggest the possibility that the polyamide, by binding to a site that overlaps the last three positions of the PSEA, may actually be stabilizing the binding of DmSNAPc or doing something else to stimulate transcription relative to the other two templates.

DISCUSSION

The goal of this project was to examine the effects on transcription of polyamide binding to sites within the U1 promoter and to determine whether the polyamide could be used to alter the RNA polymerase specificity of the promoter. The binding specificity of the HIV-1 polyamide we used proved to be problematic. While it did effect transcription at the promoters containing the HIV-1 recognition site, it similarly inhibited transcription at other promoters that did not contain the HIV-1 recognition site. Also, results tended to be inconsistent and not always reproducible, so it was decided not to put any more effort into the project.

REFERENCES

1. Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L. C., and Stumph, W. E. (1997) *Eur.J.Biochem.* **248**, 231-237

2. Jensen, R. C., Wang, Y., Hardin, S. B., and Stumph, W. E. (1998) *Nucleic Acids Res.* **26**, 616-622

3. Hernandez, G., Valafar, F., and Stumph, W. E. (2007) *Nucleic Acids Res.* **35**(1), 21-34

4. Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (2004) *Mol.Cell.Biol.* **24**(5), 1897-1906

5. Wang, Y., and Stumph, W. E. (1998) *Mol.Cell.Biol.* **18**, 1570-1579

6. Dervan, P. B. (2001) *Bioorg. Med. Chem.* **9**, 2215-2235

7. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature* **382**(6591), 559-561

8. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature* **391**(6666), 468-471

9. Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J., and Gottesfeld, J. M. (1999) *J.Biol.Chem* **274**(18), 12765-12773

10. Lai, H. T., Chen, H., Li, C., McNamara-Schroeder, K. J., and Stumph, W. E. (2005) *Nucleic Acids Res.* **33**(20), 6579-6586

11. Zamrod, Z., Tyree, C. M., Song, Y., and Stumph, W. E. (1993) *Mol.Cell.Biol.* **13**, 5918-5927

CONCLUDING REMARKS

DmSNAPc is an intriguing transcription factor in that it is required for both the transcription of the U1-U5 genes by Pol II and the transcription of the U6 genes by Pol III. In work described in Chapter 1 of this dissertation, I and two other students have determined domains in all three of the DmSNAP subunits required for *in vivo* assembly of the complex. We have also mapped domains required for DNA binding.

With one exception, the regions of the proteins that are evolutionarily conserved among animals are sufficient for complex assembly. This knowledge can be used to gain insight into the function of similar proteins in other organisms. The ability to compare results in the human and fly systems, for example, has allowed us to identify domains that function similarly in both systems. More importantly, however, differences between the two systems have provided novel insights into metazoan SNAPc function and suggest possible alternative mechanisms for complex assembly and stable DNA binding.

The work that is described in Chapter 1 will lead the way to further research into the nature and function of the DmSNAP complex. The next step would be to identify domains involved in transcriptional activity, especially domains that are specific for Pol II or Pol III transcription. First, I would test the ability of the DmSNAP190(63-721) N-terminal truncated subunit complexed with the full length versions of DmSNAP43 and DmSNAP50 to support transcription *in vitro*. This version of DmSNAP190 binds to the promoter almost as well as the full length version (Chapter 1, Fig. 5, lanes 1, 2, 5, and 6). It would also be worthwhile to test the

transcriptional ability of the DmSNAP190(2-623) C-terminal truncated version of the subunit along with the full length versions of the other two subunits, although it binds only weakly to the promoter (Chapter 1, Fig. 7, lane 14).

Each of these *in vitro* transcription experiments would be carried out with both U1 and U6 templates in the presence or absence of α-amanitin, which inhibits transcription by pol II, or tagitetoxin, which inhibits transcription by Pol III. In this way we could potentially identify domains in DmSNAP190 that are specifically required for either Pol II or Pol III transcription. For instance, if the N-terminally truncated DmSNAP190(63-721) version supported transcription of U6 by Pol III but not U1 by Pol II, we would know that residues located between positions 1 and 62 in the N-terminal end of DmSNAP190 are involved in Pol II, but not Pol III transcription.

Similar experiments would be carried out with the truncated versions of the other two subunits that are able to bind DNA. Two of the truncated versions of the DmSNAP50 subunit were able to bind to DNA. The version with 10 residues removed from the N-terminal end bound the U1 promoter as well as the full length version (Chapter 1, Fig. 10, lanes 1 and 2). The DmSNAP50(26-376) also bound to the DNA, but perhaps more weakly (Chapter 1, Fig. 10, lanes 3 and 9). These two N-terminally truncated versions of DmSNAP50 would be used in *in vitro* transcription experiments as described above. They provide an opportunity to determine if the N-terminus of DmSNAP50 is involved in activating Pol II and/or Pol III transcription. Only one of the truncated versions of DmSNAP43, the C-terminally truncated (2-274)

version was able to bind to DNA and this was weak (Chapter 1, Fig. 13, lanes 2 and 5). Nonetheless, it may still be worthwhile to investigate the transcriptional activity of DmSNAPc with this version of DmSNAP43. Ultimately I would like to find the smallest version of DmSNAPc that would support transcription using combinations of all the truncated subunits that bind to DNA in *in vitro* transcription experiments.

In order to perform these experiments, it is necessary to have an *in vitro* transcription system for snRNA genes that is dependent upon the addition of exogenous DmSNAPc. I have already done preliminary work in our lab to immunodeplete *Drosophila* nuclear extract of endogenous SNAPc so that it could be used with added truncated SNAPc for these experiments. In these experiments, I linked rabbit anti-DmSNAP antibodies against each of the three DmSNAPc subunits to an insoluble resin, and then used this antibody resin to remove endogenous DmSNAPc from the soluble *Drosophila* nuclear extract. I have had some success with this, but so far, the results have been inconsistent.

In Chapter 2, I describe my unsuccessful attempts to randomly mutate live fruit flies in such a manner as to change the RNA polymerase specificity of a U6-like PSEA in the U1 promoter so that an inactive neomycin resistance gene will be transcribed by RNA polymerase II. This was a difficult project without guaranteed success, but the great value of successful results in this project made us believe it to be a worthwhile endeavor. Perhaps It may have been wiser to do a preliminary statistical analysis of the chances of success before starting; however, the project was reviewed and funded by the National Science Foundation.

Chapter 3 describes a project using a synthetic polyamide compound designed to recognize and bind to a specific DNA sequence in the minor groove of the DNA helix. The idea was to alter RNA polymerase specificity at the U1 promoter by binding the polyamide at various positions between the PSEA and PSEB to block contact with the DNA by DmSNAPc. Conceivably this would force a conformational change leading to the recruitment of Pol III instead of Pol II. The DNA specificity of the polyamide compound used proved to be problematic, and the results were inconsistent. In all cases except one, the presence of the polyamide compound inhibited transcription almost uniformly from all templates. In one case however, the results could be interesting. In the U1-19HIV1-30PSEB template, the polyamide binding site overlaps the last three base pairs of the PSEA. This could certainly stabilize DmSNAPc or alter its conformation, and indeed, very preliminary results using this template showed enhanced transcription in the presence of the polyamide compound. Those results could be followed up to determine if the effect is reproducible. If it is, the reaction should be carried out in the presence of α-amanitin and tagetitoxin to determine the RNA polymerase specificity of the transcription because it is conceivable that polyamide binding in this region could change the polymerase specificity of the promoter.

Understanding the structure and function of the small nuclear RNA (snRNA) genes in eukaryotes is important because they serve as a good model for the study of the structure and function of a much wider variety of genes that have a major impact on medical issues, including oncogenes, viral genes, and other pathogenic genes.

Knowledge of the nature of snRNA transcription factors could give insight into the nature of transcription factors controlling the expression of many other genes.

APPENDIX

A. Flag purification protocol.

B. Detection of protein by chemoluminescence protocol.

C. *In Vitro* transcription protocol.

D. Composition of a basic 25ul *In-Vitro* transcription reaction

E. Immunodepletion protocol.

F. DNA sequences of DmSNAP constructs.

A. Flag Purification protocol:

6-23-06

Protocol For FLAG Tagged Protein Purification Kit (cellmm2)

I. Storage:
A. Check each component of the kit individually for proper storage. Do not freeze the resin antibody beads or they will be ruined.

II. Preparation of solutions:
A. Prepare a 5mg/ml 3x FLAG peptide solution.
1. 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-Llys-C) is acidic. In order to dissolve it properly add 160 ul of 10x Wash Buffer to 4mg of 3x FLAG peptide.
2. When the peptide is completely dissolved, add 640ul of distilled water to the sample.
3. Mix well and store aliquots of 55 ul at -20° C. This will prepare enough of the 3X FLAG peptide stock solution to do several purifications.

B. Thaw all kit solutions as needed and mix to ensure homogeneity.

C. Prepare 12ml of 1x Wash Buffer by adding 1.2 ml of 10x Wash buffer to 10.8ml sterile deionized water and mixing well.

D. ~~Dilute a portion (~2ul) of the FLAG-BAP fusion protein stock solution to a concentration of 50 ng/ul with 1x wash buffer.~~
1. ~~The diluted solution is stable at -20° C for about two months~~.

III. Procedure:

A. Perform all lysis, binding, and elution steps at 2-8°C (in the cold room).
1. Use precooled lysis and wash buffers and equipment.
2. All centrifugations are done at 2-8°C with precooled rotors.

B. FLAG resin preparation. (Steps B1-7 which are used to prepare the beads may be carried out at room temp.)

1. Thoroughly suspend the ANTI-FLAG M2 affinity gel in the vial in order to make a uniform suspension of the resin.
    a. Ratio of volume of suspension to packed gel=2 to 1.
2. Using a P 1000 with ~2mm cut off the end of the tip **immediately** transfer 320ul of suspended resin (160ul packed resin) into a new precooled 1.5ml tube
3. Centrifuge the resin for 4 minutes at 2000 x g.
4. Remove the supernatant being careful not to remove any of the beads.
5. Wash the beads twice in 1xWash Buffer.
    a. Resuspend the beads in 0.5ml of 1x wash buffer
    b. Centrifuge resin 30 seconds at 2000 x g
    c. Remove the supernatant
    d. Repeat steps a-c
6. Wash the resin twice with "elution" buffer from the kit (0.1M glycine, pH 3.5)
    a. Resuspend the resin in 0.5 ml "elution buffer".
    b. Centrifuge at 2000 x g for 30 sec.
    c. Immediately remove the supernatant. Do not leave the resin in "elution" buffer for more than 2 minutes.
    d. Repeat steps 6a-c
7. Wash the beads four times in 1X wash buffer using 0.5ml for each wash.
    a. see step 5.
    b. leave the resin suspended in 1X wash buffer.

C. Cell Lysis: for 100% confluent 100mm dish ($6.7x10^7$ cells)

1. Add 70ul of protease inhibitor cocktail to 7ml CelLytic M lysis buffer.
2. Wash cells:
    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 15ml centrifuge tube.
    c. Spin for 5 minutes in the RT-6000 at 1600 RPM (420 x g) (~3 on the dial).
    d. Decant the supernatant and discard.
    e. Wash the cells by re-suspending the pellet in 10ml of PBS and centrifuge for 5 minutes at 1600 RPM in the RT-6000 centrifuge.
    f. Decant the supernatant and discard, then pipet liquid off

thoroughly, touching the mouth of the tube with a Kimwipe.

g. Resuspend the cell pellet in 840ul of CelLytic M lysis buffer.

h. Remove cells and buffer to a 1.5 ml conical screw-cap Eppendorf tube.

3. Incubate the cells 15 minutes end over end in the cold room.

4. Centrifuge the cell lysate for 10 minutes at 12,000 x g (11,400RPM) in the microfuge in the cold room.

5. Pool the supernatants from the 8 tubes (~7ml total) into a new chilled15 ml tube.

a. 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 10ml syringe.

b. Suck the lysate into the syringe and expel it back into the tube slowly four times.

7. Remove 200ul of the lysate into a screw-cap tube and store frozen at -80°C or in liquid N_2.

a. Keep the remainder (~7ml) on ice.

D. Binding the beads

1. Add the washed resin to the cell lysate (7ml)

a. Remove the 1X wash buffer from above the settled resin beads.

b Resuspend the beads very gently in 500ul of the lysate

c transfer the resuspended beads to the 15ml tube containing the lysate.

d. Rinse the 1.5ml tube again with 500ul of lysate and transfer that to the tube containing lysate.

2. Incubate the lysate and beads for overnight in cold room on rocker or with <u>end over end rotation</u>.

3. On the day before eluting the Flag fusion protein, prepare 20ml of HEMG wash buffer and store it in the refrigerator so it will be cold before use.

a) The HEMG wash buffer has the following concentrations:

KCl	81mM

Hepes K+ pH 7.6	32.5mM
$MgCl_2$	5.5mM
DTT	5.0mM
Glycerol	10%
EDTA	0.1mM

b) Make 20ml

To 12ml of H_2O add in order:
1.62ml of 1M KCl
650ul of 1M HEPES K+ pH 7.6
110ul of 1M $MgCl_2$
4ul 0.5M EDTA
100ul of 1M DTT
2ml Glycerol

Add H_2O to make 20ml.

E. Elution of the FLAG-fusion protein with 3X FLAG peptide.

1. On the second day of the purification add a portion of the 3X FLAG peptide from step IIA to cold HEMG wash buffer to a final concentration of 200ug/ml. To do this, add 50ul of 3X FLAG peptide (5mg/ml, step IIA) to 1200ul of HEMG wash buffer (5mg/ml----->200ug/ml.)

2. Centrifuge the resin for 2 minutes at 1600 RPM in the RT-6000 , then remove the supernatant and save it in a 15 ml tube.
3. Wash the resin four times with 1.0 ml of 1X Flag wash buffer.
   a Add 1ml 1X Flag wash buffer and resuspend the beads.
   b. Remove beads and wash buffer to a new 1.5ml tube and c. Centrifuge for 30 seconds at 2000g
   d. Remove the 1X Flag wash buffer supernatant from each wash and save in a 1.5ml tube.
   e. Wash three more times with 1X Flag wash buffer.

4. Wash the resin twice with 1ml. of HEMG wash buffer **NOT containing 3X Flag peptide.**
   a Add 1ml HEMG wash buffer and resuspend the beads.
   b Centrifuge for 30 seconds at 2000g
   c. Remove the HEMG wash buffer supernatant from each wash and save in a 1.5ml tube.
   d. repeat .

5. Elute the bound Flag Fusion protein with five 230ul volumes of HEMG elution buffer containing, 3 X Flag peptide, 200ug/ml, from step E1.
    a. Add 230 ul of the 3 X FLAG peptide solution to the resin.
    b. Resuspend the resin and let the sample sit on ice in the cold room for 4-5 minutes.
    c. Centrifuge at 2000 X g for 30 seconds.
    d. Remove and save the supernatant into a chilled eppendorf tube. This is elution fraction 1.
    e. Repeat the above steps to collect elution fractions 2,3 and 4.
    f. Collect elution fraction 5 by adding 230ul of the FLAG peptide solution and incubating for 10 minutes at room temp.
    g. Keep each fraction separate.
    h. Divide each elution fraction into 5X50ul aliquats. Carefully label each aliquat and store them in liquid nitrogen.
    i. The elution fractions will be assayed by Western Blot, gel shifts and transcription.

IV. Recycle and store the resin immediately.

A. Recycle the resin.

1. Add 500ul of elution buffer (0.1M glycine pH 3.5) to the resin and centrifuge at 2000 X g for 30 seconds
2. Repeat the previous step two more times.
3. Immediately wash the resin with 1ml of 1 X wash buffer and remove the supernatant.
4. Repeat previous step four more times

B. Storing the resin.

1. Add 1ml of 1X wash buffer containing 50% glycerol and 0.02% sodium azide.
3. Store at 4° C.

B. Chemoluminescent protein detection protocol.

1-30-07
M.Titus

Chemiluminescent Detection of FLAG Epitope-tagged Proteins on Western Blots

I. Run the protein samples on a minigel and transfer to a blotting membrane in the normal manner.

A. Run 1-10ng of N-FLAG-BAP Control Protein as a positive control

1. Skip lanes between control and experimental protein lanes.
2. Dilute the control protein to 1ng/ul from 10ng\ul stock.

II. Block the unoccupied protein binding sites on the membrane.

A. Make the blocking solution by dissolving the contents of one packet of TBS with 3%(w/v) Nonfat Milk, pH 8.0 in 100 ml ddH2O.

B. Place the membrane protein-side up in 50 ml of blocking solution and incubate at room temperature for 60 min. with gentle rocking.

C. Save the remaining blocking solution for antibody dilution.

III. Incubate the membrane in anti-Flag antibody.

A. Pour off and discard the blocking solution and immediately add 50 ml of a 1/10,000 dilution of ANTI-FLAG M2 Monoclonal Antibody-Peroxidase Conjugate diluted in blocking solution from step II.A. (5ul of antibody in 50ml of blocking solution)

B. Incubate for 60 min. with gentle rocking at room temperature.

IV. Wash the membrane.

A. Make the wash solution by dissolving the contents of one packet of tris-Buffered Saline with 0.05% Tween 20, pH 8.0 (TBST) in 500ml of ddH2O.

B. Pour off and discard the diluted antibody solution from the blot and wash the membrane at least 3 times for 5 minutes with at least 50ml of TBST per wash.

V. Incubate the membrane in Chemiluminescent Peroxidase Substrate solution.

A. Pour off the last wash and drain well. Keep the blot damp; do not let the membrane dry.

B. Immediately prepare 5ml of CPS-1 Chemiluminescent Peroxidase Substrate solution by adding 1.67ml of Component 1 to 3.33ml Component 2.

C. Place the membrane in a small tray and add the substrate solution by gently pipetting it onto the surface of the membrane to cover it completely.

D. Incubate the membrane in the substrate solution for 5 minutes at room temperature with gentle rocking.

VI. Make an image of the membrane on X-ray film.

A. Drain excess substrate solution from the blot leaving the surface with a wet covering of substrate solution.

B. Using a flat surface gently and quickly place the wet membrane with the protein side down on a sheet of plastic wrap (large enough to completely wrap the membrane) and wrap the membrane in plastic.

C. Expose the X-ray film in the dark room by placing an 8X11 inch piece of film in a film cassette and placing the membrane protein side down on the upper left quadrant of the film,

D. Immediately close the cassette making sure that the membrane is tight against the film and expose the film for 5 seconds.

E. Quickly open cassette and picking the membrane up off the film (do not slide it) place the membrane again protein side down on the upper right quadrant of the film.

F. Immediately close the cassette making sure that the membrane is tight against the film and expose the film for 20 seconds.

G. Repeat the procedure by exposing the film for 60 and 180 seconds using the lower left and lower right quadrants of the film.

H. Develop the film in the usual way.

C. *In Vitro* transcription protocol.

Protocol for in Vitro Transcription

I. Solutions

Stock Solutions:

When making solutions, always wear gloves, and always use RNase-free H_2O that you have personally ordered. All glassware, metal spatulas, and stir bars should be autoclaved to destroy RNase.

1 M Hepes, K^+, pH 7.6 (100 ml)
Add 23.83 g of HEPES to about 60 ml of dH_2O in a 150 ml beaker. Add a stir bar, place on a stir plate, and stir. Slowly add solid KOH until the pH is 7.6. Pour the solution into a 100 ml graduated cylinder and add dH_2O to the 100 ml mark. Store in freezer.

HEMG, pH 7.6 (10 ml)	Final concentration
8.6 ml dH_2O (RNase-free)	
250 ul 1 M HEPES, K^+, pH 7.6	25 mM HEPES
125 ul 1 M $MgCl_2$	12.5 mM $MgCl_2$
2 ul 0.5 M EDTA, pH 8.0	0.1 mM EDTA
1 ml 100% Glycerol	10% Glycerol
Store in freezer	

Transcription Stop (50 ml)	Final concentration
40.5 ml dH_2O	
2 ml 0.5 M EDTA, pH 8.0	20 mM EDTA
2.5 ml 4 M NaCl	200 mM NaCl
5 ml 10% (w/v) Sodium Dodecyl Sulfate	1% SDS
Store at room temperature	

Dithiothreitol (1 M; 1 ml)
154.3 mg DTT
1 ml dH_2O
Store in freezer

Sodium Acetate-NaOAc (3M; 50 ml)

24.4 g NaOAc
Make up to 50 ml with dH_2O
Store in refrigerator

Sodium Acetate-NaOAc (0.3M; 50 ml)
5 ml 3 M NaOAc
45 ml dH_2O
Store in refrigerator

Ethanol, 100% (50 ml each)
Uncap a fresh, unused 500ml bottle of ethanol and pour into ten 50ml conical polypropylene tubes. Store in refrigerator.

Sevag (Chloroform/Isoamyl Alcohol 19:1)
Pipet 5 ml of Isoamyl Alcohol into a brown screw-capped 100 ml glass bottle. Add 95 ml of Chloroform into a 100 ml graduated cylinder, and pour into the brown glass bottle. Store at room temperature.

5x Annealing Buffer (10 ml)

	Final 5x Concentration
6.75 ml dH_20 (RNase-free)	
100 ul 1 M Tris-HCl, pH 7.8	10 mM Tris-HCl, pH 7.8
20 ul 0.5 M EDTA, pH 8.0	1 mM EDTA, pH 8.0
3.13 ml 4 M KCl	1.25 M KCl

Store in freezer

2x Denaturing Loading Buffer (10 ml)
10 ml 100% deionized Formamide
20 mg Xylene Cyanol
20 mg Bromophenol Blue
Store at room temperature

□-amanitin (50 ug/ml; 50 ul each)---highly toxic!
Put used pipet tips into a biohazard waste container.
Inside the hood, with a dust mask on, remove 30 ul of the concentrated 1 mg/ml □-amanitin stock solution and pipet into a pre-labeled tube. Return the concentrated stock to the freezer. Then pipet 2.5 ul of this smaller stock solution into 10 1.5 ml screw-capped tubes, each containing 47.5 ul dH_2O (Rnase-free) for a final volume of 50 ul each. Store the labeled tubes (□ amanitin, 50 ug/ml) in the freezer.

Glycogen (1 ml)
Add 20 mg powdered glycogen to 1 ml dH_2O
Store in freezer

- -

Labile solutions (make fresh every two months):

Ribonucleotide Triphosphates (5 mM each; 200 ul)
100 mM NTPs are purchased from Boehringer Mannheim
10 ul each of 100 mM rATP, rCTP, rGTP, rTTP
160 ul of dH_2O (RNase-free)
Store in freezer

Extension Mix (5.4 ml; 140 reactions)---toxic!
Put all waste (needles, tubes, pipet tips, etc.) into a biohazard waste container.

First, with a 16 gauge needle, inject 1 ml of 100% Ethanol into the 2 mg bottle of Actinomycin D (highly toxic) inside the hood. Mix by swirling the bottle side to side. With another 16 gauge needle, remove about 400 ul (0.4 ml) of the yellow Actinomycin D from the bottle, and "inject" carefully into a 1.5 ml screw-capped tube. Then pipet 62.5 ul of Actinomycin D each into six 1.5 ml screw-capped tubes. Put the caps on the tubes, transfer them to the speedvac, remove the caps, and spin the tubes to dryness (30-60 min). During this time, prepare the rest of the extension mix ingredients in six <u>different</u> screw-capped tubes than the ones spinning:

Important: add components in the following order

777 ul dH_2O (RNase-free)
62.5 ul 1 M Tris-HCl, pH 8.3
12.5 ul 100 mM $MnCl_2$
Keep on ice

Then add 25 ul each of dATP, dCTP, dGTP, and dTTP (100 mM each, Boehringer Mannheim) into a new screw-capped tube, and dilute with 150 ul RNase-free dH_2O.

Add 35.6 ul of this dNTP solution into each of the six tubes already containing Tris and $MnCl_2$.Vortex briefly. Then add 12.5 ul 1 M Dithiothreitol (DTT) to each tube. Vortex again.

After the tubes containing Actinomycin D have evaporated to dryness, put the caps on the tubes and transfer them to the hood. With a P1000, pipet the

contents of the six tubes (Tris, $MnCl_2$, dNTP's, DTT) into the Actinomycin D-containing tubes. Vortex vigorously for at least 1 minute, or until the orange pellet at the bottom is gone. Store in freezer, wrapped in aluminum foil (Actinomycin D is light-sensitive). If you do not expect to be doing a lot of transcription reactions, it may be better to prepare fewer than 6 tubes. Day before the experiment: be sure to write out the templates to be used and any other variables in the experiment in your notebook. Also read the MSDS for □-amanitin, Actinomycin D, Phenol, Chloroform, and Acrylamide. Always wear gloves for protection and to avoid RNase contamination.

II. Transcription reactions

1 Get two buckets of ice. Fill each to the top. Check to make sure that the two water baths are at 37°C and 45°C. Remove polyacrylamide from the refrigerator. Check the oil in the speed vac pump.

2. Thaw the following solutions (taken from the -20 ^{o}C freezer) by rolling between one's fingers.

   1 M Dithiothreitol (DTT)
   1 M HEPES, K^+,pH 7.6
   5 mM NTP solution (pre-mixed, contains 5 mM ATP, GTP, CTP, UTP)
   0 M HEMG*
   Each DNA template, 0.2 ug/ul (prepared using Qiagen maxi-preps).
   Each should have been previously sequenced and also examined for integrity by agarose gel electrophoresis.

   * The "zero M" means no KCl has been added to this buffer. Different concentrations of KCl may be used according to the type of promoter being used.

3. When the above solutions are thawed, vortex them for 5s. Spin tubes for 3s. Place all tubes on ice.

4. Go to the enzyme freezer with an ice bucket and obtain a tube of RNasin (Promega, 40 U/ul). Put in the ice bucket and return to bench.

5. Place one 1.5 ml screw-capped eppendorf tube on ice. Using only autoclaved (unbroken seal) pipet tips, prepare a transcription pre-mix which contains components common to all reactions. If you are going to do 8 transcription reactions, prepare enough pre-mix for 10 reactions as follows:

Each rxn.:	(Ex.: for 10 rxns.:)
2.5 ul rNTPs	25 ul rNTPs
0.5 ul HEPES	5 ul HEPES

0.5 ul O M HEMG	5 ul 0 M HEMG
0.1 ul 1 M DTT	1 ul 1 M DTT
0.7 ul RNasin	7 ul Rnasin
(4.3 ul Fin. Vol.)	(43 ul Fin. Vol.)

6. Again using Fisher screw-capped eppendorf tubes, place the appropriate number of reaction tubes into the rack, and number them 1, 2, 3, etc.

7. Fill a bucket with tap water, put in a thermometer, and adjust temp. of H_2O to 25 ^oC (adjust with either ice or warm water taken from a water bath).

8. Pipet the required amount of Rnase-free H_20 into each of the tubes as required (check experimental plan table).

Uninhibited tubes	Tubes with a amanitin	Tubes with Tagetitoxin	Tubes with Both inhibitors
1.7 ul	0 ul	0 ul	0 ul

9. Pipet 4.3 ul of the reaction mixture made in step 5 into each tube.

10. Pipet 2.0 ul of diluted DNA (0.2 ug/ul) of each template (U1 and U6) into each appropriately labeled tube. Put all buffer and DNA tubes back in the freezer.

11. Take the 20 U/ul tube of tagetitoxin out of the refrigerator. Vortex 5s, spin 3s. Pipet 2 ul of tagetitoxin (40 U) into the appropriate tubes (check notebook). Put the tagetitoxin back in the refrigerator.

12. Take out a 50 ug/ml stock of Rnase free □ amanitin from the freezer. Thaw the □ amanitin by itself in the hood. Pipet 2 ul of the □ amanitin into each appropriate tube in the hood. Bring the rack of tubes back to the bench. Return the □-amanitin stock to the freezer.

13. Remove one of the appropriate tubes of SNF from the liquid nitrogen tank (after experimentally choosing an SNF date for further work, they should be aliquotted into screw-capped tubes of 200 ul each). Thaw the tube by hand, then place on ice. Mix the thawed SNF thoroughly by finger flicking. Spin 3s. Pipet 15 ul of the SNF into each reaction tube. Before changing tips, mix the SNF with the rest of the tube's contents by pipetting 3 times.

14. Close the tubes, put the SNF back in the liquid nitrogen (if there is a sufficient volume). Further mix the tubes by finger-flicking. Spin 3s.

III. Termination of Transcription Reactions

15. Place the tubes in a yellow 16-tube holder, and place the holder in the 25 ^oC H_2O bucket. Incubate tubes for 60 min.

16. During this time, thaw the labeled recovery standard and glycogen (20 ug/ml) solutions in the 37 ^{o}C water bath. Vortex 5s, spin 3s.

17. Weigh out 1.5-3.0 mg of proteinase K and place in an empty tube. Add an appropriate amount of RNase-free H_2O such that the final concentration of proteinase K is 12.5 mg/ml (ex. 2.5 mg would have 200 ul H_2O added).

18. Add the following volume of solutions to an appropriate tube, depending upon the total number of transcription reactions being done:

Add Transcription Stop, then Glycogen, vortex, then add Proteinase K and Recovery Standard, then mix by finger flicking.

Total # Rxns	Transcription Stop	Glycogen	Prot K	Recovery Standard*
2-4	500 ul	6.25 ul	10ul	1500 cpm eq.
4-8	1000 ul	12.50 ul	20ul	3000 cpm eq.
9-12	1500 ul	18.75 ul	30ul	4500 cpm eq.
13-16	2000 ul	25.00 ul	40ul	6000 cpm eq.

* The counts per minute (cpm) of recovery standard per reaction tube is 300 cpm in this example. This number will vary according to the length of time of exposure for each experiment.

Test a 105ul sample of this transcription stop mix in the scintillation counter.

Prepare ethanol/dry ice bath.

19. Prepare for phenol / choloroform extraction and ethanol precipitation of the transcription reactions: place two sets of screw-capped tubes in new rack(s), and number them accordingly, two at a time (ex. 1, 1, 2, 2, 3, 3, etc.). The racks have 5 holes vertically; put these tubes in the 3rd and 5th holes. Skip a hole horizontally when placing the empty tubes. Unscrew the tubes, but leave the caps on.

20. Take the rack of tubes to the hood together with phenol, chloroform, and a 12 ml round bottom polypropylene 2059 tube (in a styrofoam holder). Into the 12 ml tube pipet 250 ul chloroform and 250 ul phenol per transcription reaction (i.e., for eight reactions, pipet 2 ml phenol and 2 ml chloroform). Put the cap on tightly and vortex to mix.

21. After the 60 min. transcription incubation, remove the tubes from the 25 ^{o}C water bath and place them in order into the first row of holes in the rack(s). Using one tip, pipet 105 ul of the transcription stop mix into each tube (do not touch the tubes with the pipet tip). Close the tubes, mix by gentle finger-flicking and wait for 5-10 min, keeping the tubes at room temp.

22. During this time, pipet 400 ul of chloroform (not phenol/chloroform) into the middle set of tubes in the rack.

The samples are to be extracted once with phenol/chloroform and once with chloroform.

23. After the proteinase K digestion (step 21) has gone to completion (5-10 min.), pipet 200 ul of 0.3M NaOAc into each tube (same pipet tip), and then pipet 400 ul of the phenol / chloroform mix into (up to) 8 reaction tubes (same pipet tips), or one rack of tubes.

24. Screw these tubes shut, place them in a completely empty rack, and put them on your bench. Vortex each tube (up to five tubes at a time) for 30s, then spin all tubes for 3-6 min. (<9 tubes = 3 min.; >8 tubes = up to 6 min.). During this time pipet 400 ul of the phenol / chloroform mix into any remaining tubes in the hood, vortex 30 s. At the beginning of the last vortex, stop the microfuge. The microfuge should take at least 30s to stop, allowing for a full 30s vortex of the last set of tubes.

25. Take the newly spun tubes out of the microfuge, and (if more than 8 total tubes) put in the remaining tubes. Spin those 6 min.

26. Take the newly spun tubes into the hood, place back into the rack in their 1st row position, and loosen their caps. Open the □ amanitin / actinomycin D waste container. Set the P200 pipet to "200".

27. Take the first tube's cap off (i.e., tube #1), and lay it down close to the tube. With a P200 pipet, extract the top (aqueous) layer out of the tube and into the middle row #1 tube (that already contains chloroform). It will take 2 pipettings to remove the 332 ul of upper aqueous layer. Tilt the tube 45°, so that the aqueous layer facing away from you rolls around to face you (pipet out that, too). Expel used pipet tips, tubes, and caps in the waste container.

28. Put the organic phase tube and its cap into the waste container, and repeat step 27 for all of the tubes in the rack.

29. Screw the caps on to the middle set of tubes, put them in the empty rack, and put them back on your bench.

30. Vortex 30sand spin 3 min , and make sure the microfuge is spinning until the 2nd vortex, then press stop. Vortex the final set of tubes, then replace them with the tubes in the microfuge. If you have 8 or less tubes, this is the final spin before the ethanol precipitation. If you have 9 or more tubes, repeat steps 27-30 for the second set of tubes.

31. Remove the upper aqueous phase (about 332 ul) from the choloform phase in each tube, pipetting into the appropriately labeled empty tubes in the 5th row of holes in the rack. Be careful not to pipet any of the chloroform phase in with the sample. Close the waste container. Put the rack of tubes on your bench.

32. With one pipet tip, add 1 ml of cold 100% ethanol from the freezer to each tube. Close caps and vortex (up to 5 at a time) for 30s. Incubate samples in

ethanol/dry ice bath for 15 min. Spin in cold room for 30 min. The numbers on the tubes should be facing out in order to locate the pellet when the spin is done.

V. Annealing the Primer and Primer Extension Reactions

33. Put on a new pair of gloves, safety goggles, a lab coat, and radioactive body detector and finger ring. With a plastic shield and beta-blok (4-hole plastic bar), take out the radiolabelled primer and 5x annealing buffer. Thaw them in the 37 ^{o}C water bath. Remove promptly upon thawing.

34. Calculate the volume of primer needed to give each reaction about 600,000 cpm of primer (ex. 8 reaction tubes = 4,800,000 cpm). Perform all subsequent steps behind a plastic shield, preferably the large plastic bench shield. When the primer and annealing buffer have thawed, vortex each 5s and spin 3s. Add the appropriate amounts of primer and annealing buffer to give about 12 ul of volume (600,000 cpm / reaction tube). For example, if you calculated 15 ul of primer (at 335,000 cpm/ul) to add for 8 reactions, then add 20 ul 5X annealing buffer and 65 ul H_2O for a final volume of 100 ul. Put the primer and annealing buffer tubes back in the freezer (put the primer in a large beta-blok tube holder).

35. Put 400 ml of dH_2O into a 2L beaker, put the beaker on a heat/stir plate, and turn on the heat to about 2/3 maximum. Put a thermometer in the beaker.

36. Put 10 ul of the newly mixed primer/annealing buffer into a <u>non</u> screw-capped eppendorf tube, and count in the scintillation counter. If >600,000 cpm or <400,000 cpm, then adjust accordingly.

37. When the 30 minute spin in the microfuge is over immediately take the tubes back to your bench and pipet off about 2/3 of the ethanol with your p-1000 set at 1000 and then with a P200 pipet set at 200, carefully take the rest of the liquid out from the other side of the pellet (with the pipet tip on the bottom). Always check to see if the pellet is there afterwards, and check each tip for signs of a pellet. Leave the caps off as you go.

38. When you have removed all of the ethanol you can from the last tube, go back and starting with the first tube carefully pipet out the rest of the liquid in each tube. Avoid pellet!

39. With one pipet tip, place 10ul of primer/annealing buffer onto the middle side of each tube and screw the caps on. Tap the tubes so the solution falls to the bottom. Discard the tip and any remaining primer solution in the radioactive trash.

40. With your face and body shielded by the large bench shield, vortex the tubes (3-5 at a time) for 20s. Spin 3s.

41. While the tubes are spinning down, check the temperature of the water. Adjust the heat control so that the water will be between 70-80 ^{o}C (you may have to take the beaker off the heat plate if over 85 ^{o}C).

42. Put the newly spun tubes in the yellow tube holder and put the holder in the water bath for 2 min at 70-80 ^{o}C. Then put the holder in the 45 ^{o}C water bath for 60 min, and put the plastic shield on top of the bath to let others know there is radioactivity in the bath.

IV. Assembling the Gel Apparatus

43. If running the gel today, while the tubes are incubating in the 45°C water bath, assemble the gel: Wearing gloves, get two large plates (long and short, both thick glass). Take one at a time to the sink and scrub thoroughly with dish soap and a paper towel. Rinse thoroughly. Take two thin spacers and wash them thoroughly with dish soap and water, and rinse thoroughly. Wipe the water off the plates with Kimwipes. Then squirt room temperature 100% ethanol on both plates, and wipe with Kimwipes. Squirt ethanol on the sides of the both plates that will be inside facing the gel, and scrape it off with a razor blade Then squirt about 2 square inches of Gelslick on the middle of the small plate. Spread the Gelslick over all of the small plate with Kimwipes. After spreading Gelslick on the small plate, spread residual Gelslick from the Kimwipe on the large plate. Wait about 5 min.

44. After the five minutes, squirt ddH_2O on the both plates and wipe off thoroughly with Kimwipes. Put the side spacers on the large plate. Wrap the bottom and sides with tape.

45. About 10 min from the end of the incubation, prepare the tubes for primer extension: thaw a 1 ml tube of extension mix in the 37 ^{o}C water bath. Remove promptly after thawing and place on ice. Take the ice bucket to the enzyme freezer and put the reverse transcriptase (Superscript II, BRL) and RNasin (Promega) into the ice bucket.

46. Invert the extension mix tube several times to mix. A small brown pellet stuck to the bottom is ok, but if there a significant amount of solution at the bottom of the tube that is darker than the rest of the solution, the extension mix has gone bad.

47. Pipet the following solutions into an empty screw-cap eppendorf tube:

# Rxns	Extension Mix	Rev. Trans.	RNasin
2-4	200 ul	3.3 ul	2.5 ul
5-8	400 ul	6.7 ul	5.0 ul
9-12	600 ul	10.0 ul	7.5 ul
13-16	800 ul	13.3 ul	10.0 ul

With the P200 set at "200", thoroughly mix the glycerol-laden proteins (Reverse Transcriptase; RNase inhibitor) into the extension mix by

pipetting slowly at the bottom of the tube several times. Keep the prepared extension mix / RT / RNasin on ice. Put the RT and RNasin back into the enzyme freezer, and the rest of the extension mix back in your freezer.

48. After the 60 min annealing period is over, take the tubes out of the 45 ^{o}C water bath, and spin 3s. Place the tubes in beta-bloks behind the hand-held shield. Unscrew the caps (leave them on the tubes).

49. With a different tip each time, pipet 41.5 ul of extension mix / RT / RNasin into each tube, and mix by pipetting slowly 3 times. Expel pipet tips into the waste container. Close the caps.

50. Put the tubes in the yellow holder, and incubate in the 45 ^{o}C water bath for 60 min (put shield on bath again). Start the speed-vac.

51. During this time, take out a 60 cc syringe, spacer for sharktooth comb, 200 ul and 1000 ul pipet tips and their pipetters, and put them next to the two plates. Also obtain a 100 ml graduated cylinder and a 250 ml beaker and put them next to the plates.

52. Measure out 94 ml of 10% denaturing acrylamide (7M urea, 2X TEB) into a 100 ml graduated cylinder. Then pour into a 250 ml beaker. Return the acrylamide to the refrigerator. Take the TEMED and 10% APS (1g ammonium persulfate+ 10ml RNase-free water) solutions out of the refrigerator. Pipet 500 ul of the 10% APS solution into the beaker, then pipet 51 ul of the TEMED into the beaker.

53. Mix the above solution by swirling with the 60cc syringe, then draw up the solution into the syringe. Pick up one side of the gel at a 45° angle and squirt the acrylamide into one side. When the acrylamide is out of the syringe, put the plates face down on an empty tip box, and fill the plates to the top. Then put the square spacer for sharktooth comb in, first from one side, then the other. Do it carefully, or the acrylamide will splatter several inches. Then immediately clamp the gel with 5 clamps on the top (all the way in), and 3 clamps on each side (halfway in). Continue to add drops of acrylamide to the gaps beside the spacer as needed to maintain a bead, until the acrylamide is completely polymerized (about 15 min.) Clean up everything carefully.

VI. Termination of Primer Extension Reactions; Preparation of Samples for Gel Electrophoresis

54. About 10 minutes before the incubation time is over, prepare the tubes for phenol / chloroform extraction and ethanol precipitation: put a single line of screw-cap tubes in the rack(s), number them, and uncap them. Mix 63 ul of 0.3 M NaOAc with 7.5 ul of 3 M NaOAc / reaction tube (ex. 8 rxns. = 500 ul 0.3 M NaOAc and 60 ul 3 M NaOAc ex 9rxns= 630ul .3MNaOac, 75ul 3M NaOac). Mix by vortexing.

55. Put the rack(s) of tubes, 0.3 / 3 M NaOAc tube, phenol, chloroform, P200, and pipet tips into the hood.

56. When the 60 min incubation is over, take the tubes out of the 45 ^{o}C water bath, spin down, and put them in the beta-bloks behind the small shield. Bring the tubes into the hood.Loosen the caps, and with one pipet tip add 200 ul of the phenol / chloroform mix to each tube. Then add 56 ul of the 0.3 / 3 M NaOAc mix into each tube. Tighten the caps.

57. Vortex up to eight tubes for 30s. Spin 3 min. If 9 or more tubes, vortex the remainder of the tubes not in the microfuge.

58. After the 1st group of tubes is finished spinning (if more than 8 tubes total), exchange tubes and spin 3 min. If not more than 8 tubes, take the tubes out of the microfuge and put them in the hood.

59. Put the tubes in the 3rd spaces in the racks and loosen the caps. Remove and save the aqueous phase of each tube by tilting the tube 45° and pipetting the top layer with a P200 Pipetman. Wait a few seconds after doing this (keep tube angle at 45°) to see the rest of the aqueous phase "comes around" to face you. Remove this also; being careful not to pick up any of the chloroform phase, and pipet into the new numbered but empty tubes. Remember to dispose of all caps, tubes, and pipet tips in the waste container.

60. Bring the tubes back to your bench, and add 300 ul cold 100% EtOH from your freezer. Close caps, vortex 30s, incubate in dry ice/EtOH bath for 15 min., then spin in cold room 15 min. Put the beta-bloks and shield just outside the cold room.

61. If running gel today, begin heating 400 ml of dH_2O in a 2L beaker at full heat. Clean up everything, including in the hood. Run the Geiger counter over all objects before putting them back in your drawers. > 300 cpm = contaminated (put in beaker of dH_2O with some Ivory soap and let sit for 1 day).

62. Carefully peel off the tape from the gel plates. Even more carefully take the square sharktooth spacer out: loosen the spacer at first by inserting a metal spatula between the spacer and the plate at several places. Then, loosen the spacer away from the plate slowly by turning the spatula on each edge of the spacer so that it pushes out the spacer. Try to push each edge as evenly as possible. wipe the liquid acrylamide off the plates and then remove the excess solidified acrylamide from the plates with a razor blade. Rinse the spacer with dH_2O after removing it.

63. Obtain a large polyacrylamide gel apparatus, a plastic horizontal bar, and two blue rubbery spacers. Put the plates on the apparatus (large plate facing out). Make sure the plates are straight up and down (perpendicular to the bench) and then put the plastic bar over them. Put the blue spacers above the small plate (behind the long plastic spacers). Push them down tight with a metal spatula. Tighten the horizontal bar.

64. Mix 400 ml of 10X TEB with 1600 ml ddH_2O in a 2 liter graduated cylinder (final = 2X TEB). Pour the TEB in the top and bottom reservoirs. With a plastic pipet, blow the excess acrylamide out of the spacer well. Using a 5ml syringe with a curved needle, remove any bubbles from the bottom edge of the gel.

65. Take a curved large shield from a bench (the tall, kinked-at-the-top shield), and put it in front and under the gel apparatus by carefully tilting the apparatus and sliding the base of the shield under the apparatus. Put a temperature gauge in the middle of the front gel plate. Hook the electrodes to the power supply. Prewarm the gel at 1800 volts (65 watts). Do not warm past 50 ^oC or the plates may crack and ruin your gel. When the gel temp gets to 49 ^oC, turn the watts down (it will take practice to know how long and how many watts to warm the gel).

66. Put a 50 ml conical tube in a 250ml beaker or styrofoam holder.

67. When the 15 min spin is complete, bring the tubes back to your bench and place behind shield.

68. Carefully pipet the ethanol out of the tube (twice) with the P200, the second time touching the bottom of the tube on the side opposite the pellet. Make sure you did not pick up the pellet on the pipet tip. Put the ethanol and pipet tips in the 50 ml conical tube. Put all waste in the radioactive trash.

69. Dry the samples by spinning in the Savant Speed-Vac for 2-3 min.

70. With a P200, pipet out about 100 ul of 2x denaturing dye (loading buffer). With one pipet tip, pipet 4 ul of loading buffer onto the side of each tube. Close the tubes, and tap them to put the dye at the bottom.

71. Vortex the tubes normally 30s each, then individually vigorously vortex (on the side of the vortexer) for 5-30s, in order to remove the pellets from being stuck on the bottom of the tubes. Spin 5s.

72.. If running gel today, skip to step 74. If not, put the tubes in the yellow holder, and put the holder and the small shield in your freezer overnight. Tomorrow go to step 74.

VII. Gel Electrophoresis and Autoradiography

73. If you did the experiment the night before, take out the small shield and tubes from the freezer. Thaw them briefly in the 37 ^oC water bath, and wipe off the condensation from the shield. Vortex tubes 30s, then individually 5-30s each on the side of the vortex. Spin 5s. Prepare an ice-water bath in an insulated bucket.

74. When the H_2O is boiling, put the reaction tubes in the yellow holder and put the holder in the boiling water for 2 min.

75. During this time, check the gel temp. and adjust accordingly (the temp. should be about 50 ^{o}C when you load the samples). After 2 min. of boiling, take the yellow holder out of the beaker and immediately put it in the ice-water bucket for 90s. Take the tubes out of the ice and place in numerical order in beta-bloks behind the small shield.

80. Stop the gel, and blow out the urea from the well, insert the sharkstooth comb just into the gel, about even with the brown gasket (hopefully the temp. is at 50 ^{o}C; don't worry when it falls as you are loading your samples).

81. Place a large Kimwipe next to the gel apparatus, along with a P10 pipet and pipet tips. Load one lane with 4ul of loading dye only, and wait a few seconds to make sure the samples won't leak out of the lanes

82..Pick a good spot to load the first sample (middle-left). Set the P20 to "7.0" and suck up the sample from the tube. Then turn the knob on the pipetman down until the sample comes to the end of the tube. Load the sample into the well without loading any air bubble if possible.

83. After all samples are loaded, Wait a few seconds to let all the samples settle equally to the bottoms of the wells. Turn on gel again, this time to 65 watts. Clean up radioactive waste, and remember timer, ice bucket, yellow tube holder, and hot beaker.

84. Check the temperature of the gel every 10-20 min, maintaining it at 50 ^{o}C. The gel should be run until the fast dye reaches 27 cm from the bottom of the wells .

86. About 5 min before the stopping the gel, cut out a 6" by 8" piece of filter paper. Also bring from your bench a pair of scissors, a spatula, squirt bottle with water and the radioactive waste container. Have Saran Wrap near.

87. Turn off the power to the gel, remove wires and put the power supply away. Take off the temperature gauge from the plate. Empty the top reservoir of most of its buffer with a 250 ml beaker. Unscrew the horizontal plastic bar, and lay the gel small plate up.

88. With a metal spatula, pry the smaller glass plate off of one corner. Lift carefully. If the gel sticks to the small plate, turn both plates over and then remove the large plate.

89. Put the filter paper on the gel over the reaction lanes. The bottom of the paper should be about half way between the fast and slow dyes. With a razor blade cut the gel around the edge of the paper. Squirt water around the paper where the edges of the gel are, and make a big X across the middle of the paper with the water. Press down only on the sides and bottom of the gel (1/2 inch perimeter). With the top (lanes) facing toward you, push down on the top (inward 1 inch). Then slowly peel of the filter paper and gel, from the top to the bottom.

90. Cover the gel with a piece of Saran Wrap. Then trim the saran wrap, leaving 1/2 inch space on both sides up to the top of the gel. Very carefully fold the saran wrap under the edges of the gel/paper.

91. Take the Saran Wrap covered gel to the gel dryer and dry 20-40 min at 80 ^{o}C. You can tell when the gel is dry when the filter paper is firm and doesn't sag when held up. Meanwhile, clean up everything else (put radioactive waste in lab's large waste container, and use Geiger counter on all objects, including bench).

92. When the gel is dry, put it into a labeled large metal cassette that has an intensifying screen. Put on film in the dark room, and place in the -80 ^{o}C freezer for at least 12 hours.

D. Composition of a basic 25ul *In-Vitro* transcription reaction using 15ul of SNF, 4.0ul of template and no inhibitors

I. Composition of HEMG (20) 160mM KCl and Buffer AB used in SNF prep

A. HEMG (20) 160mM KCl

KCl	160mM
HEPES K^+ pH7.6	25mM
$MgCl_2$	12.5mM
EDTA	0.1mM
Glycerol	20%
DTT	1.5mM
Benzamidine	1.0mM
PMSF	0.1mM
Na Metabisulfite	1.0mM

B. Buffer AB

KCl	110mM
HEPES K^+ pH7.6	15mM
$MgCl_2$	5.0mM
EDTA	0.1mM
DTT	2.0mM
Benzamidine	1.0mM
PMSF	0.2mM
Na Metabisulfite	1.0mM

II. Composition of SNF

The nuclear pellet during the SNF preparation is in Buffer AB.
0.5 ml of HEMG (20) 160mM KCl is added per gram of nuclear pellet.
If we assume that the nuclear pellet contains 50% solids and 50% Buffer AB, after adding 0.5 ml of HEMG (20) 160mM KCl the composition in the SNF will be 50% HEMG (20) 160mM KCl and 50% Buffer AB.

KCl	135mM
HEPES K^+ pH7.6	20.mM
$MgCl_2$	8.75mM
EDTA	0.1mM
Glycerol	10%
DTT	1.75mM
Benzamidine	1.0mM
PMSF	0.15mM
Na Metabisulfite	1.0mM

In the basic 25ul *In-Vitro* transcription reaction, 15ul of SNF are used.

III. **Volumes of reagents:** **Contribution to 25ul Transcription reaction**

rNTP	**2.5ul**	
20mM rNTP (lithium salt)		1/10=2.0mM rNTP (lithium salt)
1M HEPES K^+ pH7.6	**0.5ul**	1/50=20.0mM HEPES K^+ pH7.6
HEMG (10) 0 KCl	**0.5ul**	
25mM HEPES K^+ pH7.6		1/50=0.5mM HEPES K^+ pH7.6
12.5M $MgCl_2$		0.25mM $MgCl_2$
0.1mM EDTA		0.002mM EDTA
10% glycerol		0.2% glycerol
1M DTT	**0.1ul**	1/250=4.0mM DTT
40U/ul RNasin	**0.7ul**	28U RNasin
H_2O	**1.7ul**	
Template DNA	**4.0ul**	
0.4g U1 in water		
0.4g U6 in water		
SNF	**15.0ul**	
135mM KCl		3/5=81.0mM KCl
20mM HEPES K^+ pH 7.6		12.0mM HEPES K^+ pH 7.6
8.75mM $MgCl_2$		5.25mM $MgCl_2$
0.1M EDTA		0.06mM EDTA
10% glycerol		6.0% glycerol
1.75mM DTT		1.05mM DTT
1.0mM Benzamidine		0.6mM Benzamidine
0.15mM PMSF		0.09mM PMSF
1.0mM Na Metabisulfite		0.6mM Na Metabisulfite

Total: 25.0ul

IV. Concentrations of all components in a basic 25ul *In-Vitro* transcription reaction using 15ul of SNF, 4.0ul of template and no inhibitors.

KCl	81mM
HEPES K^+ pH7.6	32.5mM
$MgCl_2$	5.5.mM
DTT	5.05mM
rNTP (Lithium Salt)	2.0mM
Glycerol	6.2%
RNasin	28U/rxn
U1 Template	0.4g/rxn
U6 Template	0.4g/rxn
EDTA	0.062mM
Benzamidine	0.6mM
Na Metabisulfite	0.6mM
PMSF	0.09mM

E. Immunodepletion protocol

7/31/06
Prepare an affinity immunoadsorbent to remove DmPBP from SNF.

This protocol will make use of the Pierce Immunopure Immobilized Protein A (not plus) and Protein G Immunoprecipitation kit and is based upon the instructions that come with that kit.
Immunopure Immobilized Protein A, Pierce catalogue # 20333
Seize X Protein G Immunoprecipitation Kit, Pierce catalog # 45210.

The rationale is to link rabbit anti-DmPBP antibodies to an insoluble resin, and then to use this antibody resin to remove native DmPBP from the SNF. Removal of DmPBP from the SNF should result in the loss of transcriptional activity of the SNF when the U1 and U6 promoters are used as templates. Following the removal of the DmPBP from the SNF, wild type or mutant forms of DmPBP can be added back in order to determine if they can restore activity or not. Ideally, we would like to eventually discover mutant forms of DmPBP that can support transcription by one RNA polymerase, but not the other. The current protocol addresses only the portion described above in black type, not the restoration of activity by adding back DmPBP.

To optimize our chances of success, we will use a mixture of both the Protein G and Protein A resins. Also, we will attach to the resin antibodies made against all three of the subunits: DmSNAP43, DmSNAP50, and DmSNAP190. We will use the protein A and G resins to bind IgG fractions from the purified antibodies. As a negative control, we will also prepare resins to which we attach IgG from pre-immune sera.

Pre-preparations

2% sodium azide (This is a 100x stock to be added as a preservative in buffers to give a final concentration of 0.02%.

10 ml ddi water
200 mg sodium azide

1 M ethanolamine, pH'd to 7.4 with HCl. This will be added to an aliquot of the Binding/Wash Buffer to make DSS Quenching Buffer. The purpose of the ethanolamine is to inactivate the DSS crosslinker. It will be added to a final concentration of ~0.1 M, so this is a 10x stock.

6 ml ethanolamine
DDI water to 80 ml
pH to 7.4 with concentrated HCl
DDI water to 100 ml

Binding/Wash Buffer

Add 500 ml ddi water to the dry-blend buffer from the Pierce kit. The reconstituted Binding/Wash Buffer contains 0.14 M NaCl, 0.008 M sodium phosphate, 0.002 M potassium phosphate, and 0.01 M KCl. To store excess buffer, add sodium azide to a final concentration of 0.02% and store at 4 C.

DSS Quenching Buffer:

5 ml Binding/Wash Buffer (from Pierce kit)
0.5 ml 1 M ethanolamine, pH 7.4

Resins

Resin. The standard protocol from Pierce calls for using 0.4 ml of a 50% slurry of protein A or G resin (0.2 ml of settled resin). We will use 0.2 ml of protein A slurry and 0.2 ml of protein G slurry, combining them to give 0.2 ml of settled resin for each reaction.This mixture of the protein A and G resin will be prepared in duplicate.

Crosslinking of Affinity-Purified Antibodies to Protein A and Protein G beads (to be used later for immunodepletion of DmSNAPc from SNF).

Two duplicate crosslinking reactions will be performed of antibody IgG.

For each reaction, 200 ul protein A 50% slurry and 200 ul protein G 50% slurry will be used. Protein A beads: ImmunoPure Immobilized Protein A (not plus); arrived July 2006; lot # ……..
Protein G beads: ImmunoPure Immobilized Protein G Plus; from Seize X kit arrived August 2006, lot # ……..

For each reaction, 100 ul of antibodies will be used from each of the same three rabbits from the pre-immune serum.

Antibody mixture for attachment to protein A/G beads

100ul Binding/wash buffer (first)

AntiDmSNAP43: From Research Genetics
Rabbit 65459 prepared against and purified with CESEEVTEEELKAILDT
0.53mg/ml, 100ul

AntiDmSNAP50: From Sigma Genosys
Rabbit 12141 prepared against and purified with CEGEEEDIKQQNLPS
1.0mg/ml, 100ul

AntiDmSNAP190: From Sigma Genosys
Rabbit 4128 prepared against and purified with CNTVKQELEMEYTVP
1.2mg/ml, 100ul

Remove and save a 15ul aloquat for a Bradford assay (600ul sample + 150ul dye)

Notes

Perform all steps that involve the binding and crosslinking of the IgG to the resin *at room temperature.*

Perform all gel centrifugation steps for 1 minute at 3000-5000 x g (setting 5, 6318 rpm, in the Marathon Micro A table top microcentrifuge.) Centrifuging at greater speeds may cause the gel to clump and make resuspending the gel difficult.

A. Binding of IgG to Immobilized Protein A/G resin.

1. Equilibrate the Immobilized Protein A and Protein G resins and reagents to room temperature.

2. Thaw the antibodies and prepare the mixture of antibodies and buffer as described above**.**

3. Gently swirl the bottle of ImmunoPure Immobilized Protein A to obtain an even suspension. Using a cut-off P-1000 tip add 0.2 ml of the protein A resin (50% slurry) into each of two Handee Spin Cup Columns. Then gently swirl the bottle of ImmunoPure Immobilized Protein G to obtain an even suspension. Using a cut-off P-1000 tip add 0.2 ml of the protein G (50% slurry) into the same two Handee spin cup columns. Place each of the columns inside a Handee Microcentrifuge Tube and centrifuge at 3500 x g for 1 min (setting 5 on the Marathon microcentrifuge).

4. Discard the flow-through and replace spin cups into the tubes.

5. Wash gel by adding 0.4 ml of Binding/Wash Buffer to each spin cup containing the gel. Cap tube and resuspend gel by inverting with gentle shaking. Centrifuge the tube and discard flow-through. **Repeat this step two more times.**

6. Place spin cups into new microcentrifuge tubes and label with "A" for antibodies. Add the appropriate mixtures of antibodies prepared above to the tubes.

7. Cap the tubes and rotate them end over end for 1 hour at room temp to let the IgG bind to the gel beads.

8. After the 1hour incubation, centrifuge the tubes. Save the flowthrough (unbound antibodies). If anything went wrong with the binding the very valuable antibodies would be in this flowthrough. Do a Bradford on this fraction when time permits.

9. Place the spin cups into new microcentrifuge tubes and add 0.4 ml of Binding/Wash buffer to each. Invert tubes enough times to resuspend the gel. Centrifuge the tubes. **Repeat this step five additional times using the same collection tube, for a total of six washes.**

10. Transfer the spin cups into new microcentrifuge tubes and add 0.4 ml of Binding/Wash buffer to each.

B. Crosslinking the Bound Antibody.

Note: The disuccinimidyl suberate (DSS) cross-linker is moisture-sensitive. Dissolve DSS in DMSO or DMF immediately before use. The DSS is not stable and any unused DSS must be discarded. DSS is not compatible with amine-containing buffers (e.g., Tris, glycine).

1. Puncture the foil covering of a single tube of DSS with a pipette tip and add 80 ul of DMSO. Use the pipette to thoroughly mix the solution (i.e., draw up and expel the solution) until the DSS is dissolved.

2. Add 25 ul of the DSS solution to each spin cup containing the IgG bound gel beads.

Note: Because DSS is a hydrophobic molecule, a microprecipitate may form when it is added to the aqueous medium, which results in a cloudy appearance. Nevertheless, the reaction will proceed efficiently and the microprecipitate may disappear during conjugation.

3. Place caps on the microcentrifuge tubes and gently mix end over end for 30—60 minutes. Centrifuge the tubes and discard the flow-throughs.

4. To each tube add 0.4 ml of Binding/Wash Buffer containing 0.1 M ethanolamine. Cap tubes and invert 10 times. Centrifuge the tubes and discard flow-throughs. Place the spin cups back into the microcentrifuge tubes.

5. Repeat step 4 three additional times.

6. Place the spin cups in a new microcentrifuge tube and wash each three times with 400 ul of Binding/Wash Buffer (lacking ethanolamine).

Storage of the antibody resins: If the beads are to be incubated with SNF within a few days, add 0.4ml of Binding/Wash buffer to each spin cup and gently resuspend the beads. Using a cut-off P-1000 tip, combine the beads from each cup into one, new, labeled 1.5ml screw-cap tube. Centrifuge the tube at 3600 x g (setting 5 on the microfuge, 6318rpm) for 1 minute. Pipet off and discard the supernatant being very careful not to remove any of the beads. Rinse out each spin cup with an additional 0.4ml of Binding/Wash buffer and again using a cut-off P-1000 tip add these rinses to the new 1.5ml tube containing the rest of the beads. Centrifuge for 1 minute at 3600 x g and remove the supernatant being careful not to remove any of the beads Add 1.0 ml of Binding/Wash Buffer to the beads. Cap the tube, and wrap the cap with Parafilm to prevent gel from drying. Store at 4 C. (DO NOT FREEZE as that would destroy the beads!)

The antibody resins can be stored for months as follows. Centrifuge for 1 minute at 3600 x g and remove the supernatant being careful not to remove any of the beads. Add 1.0 ml of Binding/Wash Buffer that contains 0.02% sodium azide to the beads. Cap the tube, and wrap the cap with Parafilm to prevent gel from drying. Store at 4 C. (DO NOT FREEZE as that would destroy the beads!)

8-1-06
Prepare Pre-immune beads as a negative control for immunoprecipitaation and *In Vitro* transcription experiments.

This protocol will make use of the Pierce Immunopure Immobilized Protein A (not plus) and Protein G Immunoprecipitation kit and is based upon the instructions that come with that kit.
Immunopure Immobilized Protein A, Pierce catalogue # 20333
Seize X Protein G Immunoprecipitation Kit, Pierce catalog # 45210.

The rationale is to link rabbit anti-DmPBP antibodies to an insoluble resin, and then to use this antibody resin to remove native DmPBP from the SNF. Removal of DmPBP from the SNF should result in the loss of transcriptional activity of the SNF when the U1 and U6 promoters are used as templates. Following the removal of the DmPBP from the SNF, wild type or mutant forms of DmPBP can be added back in order to determine if they can restore activity or not. Ideally, we would like to eventually discover mutant forms of DmPBP that can support transcription by one RNA polymerase, but not the other. The current protocol addresses only the portion described above in black type, not the restoration of activity by adding back DmPBP.

To optimize our chances of success, we will use a mixture of both the Protein G and Protein A resins. Also, we will attach to the resin antibodies made against all three of the subunits: DmSNAP43, DmSNAP50, and DmSNAP190. We will use the protein A and G resins to bind IgG fractions from the purified antibodies. As a negative control, we will also prepare resins to which we attach IgG from pre-immune sera.

Pre-preparations

2% sodium azide (This is a 100x stock to be added as a preservative in buffers to give a final concentration of 0.02%.

10 ml ddi water
200 mg sodium azide

1 M ethanolamine, pH'd to 7.4 with HCl. This will be added to an aliquot of the Binding/Wash Buffer to make DSS Quenching Buffer. The purpose of the ethanolamine is to inactivate the DSS crosslinker. It will be added to a final concentration of ~0.1 M, so this is a 10x stock.

6 ml ethanolamine
DDI water to 80 ml

pH to 7.4 with concentrated HCl
DDI water to 100 ml

Binding/Wash Buffer

Add 500 ml ddi water to the dry-blend buffer from the Pierce kit. The reconstituted Binding/Wash Buffer contains 0.14 M NaCl, 0.008 M sodium phosphate, 0.002 M potassium phosphate, and 0.01 M KCl. To store excess buffer, add sodium azide to a final concentration of 0.02% and store at 4 C.

DSS Quenching Buffer:

5 ml Binding/Wash Buffer (from Pierce kit)
0.5 ml 1 M ethanolamine, pH 7.4

Resins

Resin. The standard protocol from Pierce calls for using 0.4 ml of a 50% slurry of protein A or G resin (0.2 ml of settled resin). We will use 0.2 ml of protein A slurry and 0.2 ml of protein G slurry, combining them to give 0.2 ml of settled resin for each reaction. This mixture of the protein A and G resin will be prepared in duplicate

Crosslinking of Pre-immune IgG to Protein A and Protein G beads (to be used later as a control for immunodepletion of DmSNAPc from SNF).

Two duplicate crosslinking reactions will be performed of pre-immune IgG.

For each reaction, 200 ul protein A 50% slurry and 200 ul protein G 50% slurry will be used. Protein A beads: ImmunoPure Immobilized Protein A (not plus); arrived July 2006; lot # ……..
Protein G beads: ImmunoPure Immobilized Protein G Plus; from Seize X kit arrived August 2006, lot # ……..

For each reaction, 67 ul of preimmune antiserum will be used from each of three rabbits (rabbits that were later used to make antibodies against the three DmSNAP subunits.)

Pre-immune sera mixture for attachment to protein A/G beads

200ul Binding/wash buffer (first)

67ul Pre-immune serum Rabbit, Rabbit 65459
67ul Pre-immune serum Rabbit, Rabbit 12142, 11/17/03
67ul Pre-immune serum Rabbit, Rabbit 4126, 8/1/02

200ul pre-immune 50 rabbit #12141

Notes
Perform all steps that involve the binding and crosslinking of the IgG to the resin *at room temperature.*

Perform all gel centrifugation steps for 1 minute at 3000-5000 x g (setting 5, 6318 rpm, in the Marathon Micro A table top microcentrifuge.) Centrifuging at greater speeds may cause the gel to clump and make resuspending the gel difficult.

A. Binding of IgG to Immobilized Protein A/G resin.

1. Equilibrate the Immobilized Protein A and Protein G resins and reagents to room temperature.

2. Thaw the preimmune sera and prepare the mixture of sera and buffer as described above.

3. Gently swirl the bottle of ImmunoPure Immobilized Protein A to obtain an even suspension. Using a cut-off P-1000 tip add 0.2 ml of the protein A resin (50% slurry) into each of two Handee Spin Cup Columns. Then gently swirl the bottle of ImmunoPure Immobilized Protein G to obtain an even suspension. Using a cut-off P-1000 tip add 0.2 ml of the protein G (50% slurry) into the same two Handee spin cup columns. Place each of the columns inside a Handee Microcentrifuge Tube and centrifuge at 3500 x g for 1 min (setting 5 on the Marathon microcentrifuge).

4. Discard the flow-through and replace spin cups into the tubes.

5. Wash gel by adding 0.4 ml of Binding/Wash Buffer to each spin cup containing the gel. Cap tube and resuspend gel by inverting with gentle shaking. Centrifuge the tube and discard flow-through. **Repeat this step two more times.**

6. Place spin cups into new microcentrifuge tubes and label with "P" for preimmune serum. Add the appropriate mixtures of serum prepared above to the tubes.

7. Cap the tubes and rotate them end over end for 1 hour at room temp to let the IgG bind to the gel beads.

8. After the 1hour incubation, centrifuge the tubes.

9. Place the spin cups into new microcentriguge tubes and add 0.4 ml of Binding/Wash buffer to each. Invert tubes enough times to resuspend the gel. Centrifuge the tubes. **Repeat this step five additional times using the same collection tube, for a total of six washes.**

10. Transfer the spin cups into new microcentrifuge tubes and add 0.4 ml of Binding/Wash buffer to each.

B. Crosslinking the Bound Antibody.

Note: The disuccinimidyl suberate (DSS) cross-linker is moisture-sensitive. Dissolve DSS in DMSO or DMF immediately before use. The DSS is not stable and any unused DSS must be discarded. DSS is not compatible with amine-containing buffers (e.g., Tris, glycine).

1. Puncture the foil covering of a single tube of DSS with a pipette tip and add 80 ul of DMSO. Use the pipette to thoroughly mix the solution (i.e., draw up and expel the solution) until the DSS is dissolved.

2. Add 25 ul of the DSS solution to each spin cup containing the IgG bound gel beads.

Note: Because DSS is a hydrophobic molecule, a microprecipitate may form when it is added to the aqueous medium, which results in a cloudy appearance. Nevertheless, the reaction will proceed efficiently and the microprecipitate may disappear during conjugation.

3. Place caps on the microcentrifuge tubes and gently mix end over end for 30—60 minutes. Centrifuge the tubes and discard the flow-throughs.

4. To each tube add 0.4 ml of Binding/Wash Buffer containing 0.1 M ethanolamine. Cap tubes and invert 10 times. Centrifuge the tubes and discard flow-throughs. Place the spin cups back into the microcentrifuge tubes.

5. Repeat step 4 three additional times.

6. Place the spin cups in a new microcentrifuge tube and wash each three times with 400 ul of Binding/Wash Buffer (lacking ethanolamine).

Storage of the pre-immune resins: If the beads are to be incubated with SNF within a few days, add 0.4ml of Binding/Wash buffer to each spin cup and gently resuspend the beads. Using a cut-off P-1000 tip, combine the beads from each cup into one, new, labeled 1.5ml screw-cap tube. Centrifuge the tube at 3600 x g (setting 5 on the microfuge, 6318rpm) for 1 minute. Pipet off and discard the supernatant being very careful not to remove any of the beads. Rinse out each spin cup with an additional 0.4ml of Binding/Wash buffer and again using a cut-off P-1000 tip add these rinses to the new 1.5ml tube containing the rest of the beads. Centrifuge for 1 minute at 3600 x g and remove the supernatant being careful not to remove any of the beads Add 1.0 ml of Binding/Wash Buffer to the beads. Cap the tube, and wrap the cap with Parafilm to prevent gel from drying. Store at 4 C. (DO NOT FREEZE as that would destroy the beads!)

The preimmune resins can be stored for months as follows. Centrifuge for 1 minute at 3600 x g and remove the supernatant being careful not to remove any of the beads. Add 1.0 ml of Binding/Wash Buffer that contains 0.02% sodium azide to the beads. Cap the tube, and wrap the cap with Parafilm to prevent gel from drying. Store at 4 C. (DO NOT FREEZE as that would destroy the beads!)

**Removing DmSNAPc from SNF by Immunoadsorption,
and Testing by In Vitro Transcription**

1. Resuspend the Antibody and Preimmune resin beads and transfer 400ul of this suspension to a spin cup using a cut-off P-1000 tip. Centrifuge at 3000-5000 x g (setting 5, 6318 rpm, in the Marathon MicroA table top microcentrifuge) for 1 minute. Discard the flow-through.

   Then transfer the remaining suspension to the spin cup and centrifuge again for one minute at 3000-5000 rpm. Discard the flow-through. Wash the tube the beads came in with an additional 400 ul of binding/wash buffer to get all the beads into the cup and centrifuge again for one minute at 3000-5000 rpm. Discard the flow-through.

2. Wash the prepared Antibody and Preimmune gel resins each twice with 400 ul of HEMG-135. Do this in the spin cups, centrifuging at 1000 x g (setting 2, 3427 rpm, in the Marathon MicroA table top microcentrifuge) for 1 minute. I am suggesting a slower spin than before so that the increased viscosity of the HEMG from the glycerol does not crush the beads as the liquid is centrifuged through. However, we need to observe and check whether this is sufficient to spin through all of the solution.

3. Add 400 ul HEMG-135 to each spin cup, and resuspend the resins. Using a cut-off P-1000 tip, transfer each resin to a new labeled 1.5ml snap-cap microcentrifuge tube (labeled A or P).

4. Rinse out each spin cup with an additional 400 ul of HEMG-135, and add these rinses to the respective resins previously transferred to the microcentrifuge tubes. Mix the contents of each tube by inversion.

5. Centrifuge each tube at 1000 x g for 1 minute. Pipet off and discard the supernatants, leaving the beads in their respective tubes. Place tubes on ice for a few minutes.

6. Thaw a 1ml aliquat of 3-8-04 SNF by hand. Mix the SNF thoroughly by gentle finger flicking.

7. To the 200 ul of settled Antibody resin beads, add 500 ul of 3-8-04 SNF preparation. Add the rest of the 1ml aliquat of 3-8-04 SNF to the 200ul of settled Preimmune beads.

8. Make sure that the caps of the snap-cap tubes are fully and tightly closed and tumble these mixtures slowly end over end in the cold room for five hours.

9. After five hours of tumbling, centrifuge the tubes in the cold room at 1000g (3300RPM) for one minute to settle the resin; then being careful not to remove any beads remove the SNF from each tube in 150ul aliquats and place these samples into new labeled 1.5ml screw-cap tubes. Place these tubes in liquid nitrogen (unless they are used the same day for in vitro transcription).

Regeneration of the antibody resins

10. Add 400 ul of Immunopure Elution Buffer (pH 2.8, contains primary amine, from Pierce kit) to each of (the spin cups containing) the resins. Invert 10 times. Centrifuge. Save flow-through to analyze by western blot for the DmPBP subunits. **Repeat this wash three more times.**

11. Wash the resin with 0.5 ml of Binding/Wash buffer. Repeat this wash three more times.

12. Add 0.5 ml of Binding/Wash Buffer that contains 0.02% sodium azide. Wrap the tube with Parafilm to prevent gel from drying. Store at 4 C.

F. DNA sequences of DmSNAP constructs

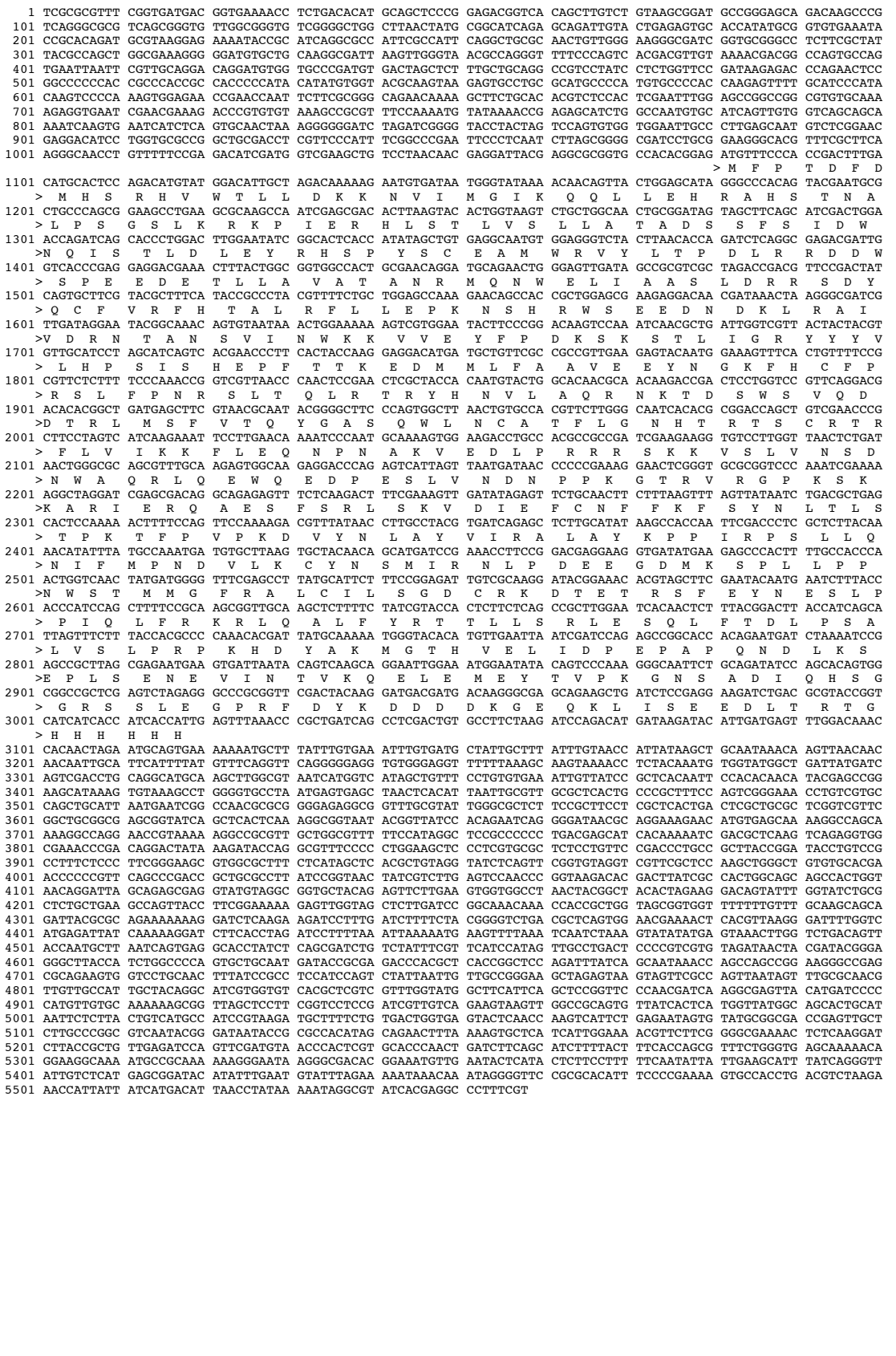
pMTDmSNAP190FlagMycHis

```
   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 CTTCAACCTG AGCAATGGAC
                                                                                                          > M  D
 901 GATGCCGATG ATCTGCGCAT AAACCAGCAG CAGCTTATGG ACGAGATTTC GGCCCTCTTG CAGAACACAG CGAAGCCGAG CGCCGCCAAC GCACTGCAGC
    > D  A  D   D  L  R  I   N  Q  Q   Q  L  M   D  E  I  S   A  L  L   Q  N  T   A  K  P  S   A  A  N   A  L  Q
1001 TGAATCAGGA GCTCCAGCGC CGGCTCATGC AAGTCCGGAC CAAAATTCTG GCCATGTTAC AAGTAGTAAG GGCCCGCTTC TCTCGGAACG AGGACATCCT
    >L  N  Q  E   L  Q  R   R  L  M   Q  V  R  T   K  I  L   A  M  L   Q  V  V  R   A  R  F   S  R  N   E  D  I  L
1101 GGTGCGCCGG CTGCGACCTC GTTCCCATTT CGGCCCGAAT TCCCTCAATC TTAGCGGGGC GATCCTGCGG AAGGGCACGT TTCGCTTCAA GGGCAACCTG
    > V  R  R   L  R  P   R  S  H  F   G  P  N   S  L  N   L  S  G  A   I  L  R   K  G  T   F  R  F  K   G  N  L
1201 TTTTTCCGAG ACATCGATGG TCGAAGCTGT CCTAACAACG AGGATTACGA GGCGCGGTGC CACACGGAGA TGTTTCCCAC CGACTTTGAC ATGCACTCCA
    > F  F  R   D  I  D  G   R  S  C   P  N  N   E  D  Y  E   A  R  C   H  T  E   M  F  P  T   D  F  D   M  H  S
1301 GACATGTATG GACATTGCTA GACAAAAAGA ATGTGATAAT GGGTATAAAA CAACAGTTAC TGGAGCATAG GGCCCACAGT ACGAATGCGC TGCCCAGCGG
    >R  H  V  W   T  L  L   D  K  K   N  V  I  M   G  I  K   Q  Q  L   L  E  H  R   A  H  S   T  N  A   L  P  S  G
1401 AAGCCTGAAG CGCAAGCCAA TCGAGCGACA CTTAAGTACA CTGGTAAGTC TGCTGGCAAC TGCGGATAGT AGCTTCAGCA TCGACTGGAA CCAGATCAGC
    > S  L  K   R  K  P   I  E  R  H   L  S  T   L  V  S   L  L  A  T   A  D  S   S  F  S   I  D  W  N   Q  I  S
1501 ACCCTGGACT TGGAATATCG GCACTCACCA TATAGCTGTG AGGCAATGTG GAGGGTCTAC TTAACACCAG ATCTCAGGCG AGACGATTGG TCACCCGAGG
    > T  L  D   L  E  Y  R   H  S  P   Y  S  C   E  A  M  W   R  V  Y   L  T  P   D  L  R  R   D  D  W   S  P  E
1601 AGGACGAAAC TTTACTGGCG GTGGCCACTG CGAACAGGAT GCAGAACTGG GAGTTGATAG CCGCGTCGCT AGACCGACGT TCCGACTATC AGTGCTTCGT
    >E  D  E  T   L  L  A   V  A  T   A  N  R  M   Q  N  W   E  L  I   A  A  S  L   D  R  R   S  D  Y   Q  C  F  V
1701 ACGCTTTCAT ACCGCCCTAC GTTTTCTGCT GGAGCCAAAG AACAGCCACC GCTGGAGCGA AGAGGACAAC GATAAACTAA GGGCGATCGT TGATAGGAAT
    > R  F  H   T  A  L   R  F  L  L   E  P  K   N  S  H   R  W  S  E   E  D  N   D  K  L   R  A  I  V   D  R  N
1801 ACGGCAAACA GTGTAATAAA CTGGAAAAAA GTCGTGGAAT ACTTCCCGGA CAAGTCCAAA TCAACGCTGA TTGGTCGTTA CTACTACGTG TTGCATCCTA
    > T  A  N   S  V  I  N   W  K  K   V  V  E   Y  F  P  D   K  S  K   S  T  L   I  G  R  Y   Y  Y  V   L  H  P
1901 GCATCAGTCA CGAACCCTTC ACTACCAAGG AGGACATGAT GCTGTTCGCC GCCGTTGAAG AGTACAATGG AAAGTTTCAC TGTTTTCCGC GTTCTCTTTT
    >S  I  S  H   E  P  F   T  T  K   E  D  M  M   L  F  A   A  V  E   E  Y  N  G   K  F  H   C  F  P   R  S  L  F
2001 CCCAAACCGG TCGTTAACCC AACTCCGAAC TCGCTACCAC AATGTACTGG CACAACGCAA CAAGACCGAC TCCTGGTCCG TTCAGGACGA CACACGGCTG
    > P  N  R   S  L  T   Q  L  R  T   R  Y  H   N  V  L   A  Q  R  N   K  T  D   S  W  S   V  Q  D  D   T  R  L
2101 ATGAGCTTCG TAACGCAATA CGGGGCTTCC CAGTGGCTTA ACTGTGCCAC GTTCTTGGGC AATCACACGC GGACCAGCTG TCGAACCCGC TTCCTAGTCA
    > M  S  F   V  T  Q  Y   G  A  S   Q  W  L   N  C  A  T   F  L  G   N  H  T   R  T  S  C   R  T  R   F  L  V
2201 TCAAGAAATT CCTTGAACAA AATCCCAATG CAAAAGTGGA AGACCTGCCA CGCCGCCGAT CGAAGAAGGT GTCCTTGGTT AACTCTGATA ACTGGGCGCA
    >I  K  K  F   L  E  Q   N  P  N   A  K  V  E   D  L  P   R  R  R   S  K  K  V   S  L  V   N  S  D   N  W  A  Q
2301 GCGTTTGCAA GAGTGGCAAG AGGACCCAGA GTCATTAGTT AATGATAACC CCCCGAAAGG AACTCGGGTG CGCGGTCCCA AATCGAAAAA GGCTAGGATC
    > R  L  Q   E  W  Q   E  D  P   E  S  L  V   N  D  N   P  P  K  G   T  R  V   R  G  P   K  S  K  K   A  R  I
2401 GAGCGACAGG CAGAGAGTTT CTCAAGACTT TCGAAAGTTG ATATAGAGTT CTGCAACTTC TTTAAGTTTA GTTATAATCT GACGCTGAGC ACTCCAAAAA
    > E  R  Q   A  E  S  F   S  R  L   S  K  V   D  I  E  F   C  N  F   F  K  F   S  Y  N  L   T  L  S   T  P  K
2501 CTTTTCCAGT TCCAAAAGAC GTTTATAACC TTGCCTACGT GATCAGAGCT CTTGCATATA AGCCACCAAT TCGACCCTCG CTCTTACAAA ACATATTTAT
    >T  F  P  V   P  K  D   V  Y  N   L  A  Y  V   I  R  A   L  A  Y   K  P  P  I   R  P  S   L  L  Q   N  I  F  M
2601 GCCAAATGAT GTGCTTAAGT GCTACAACAG CATGATCCGA AACCTTCCGG ACGAGGAAGG TGATATGAAG AGCCCACTTT TGCCACCCAA CTGGTCAACT
    > P  N  D   V  L  K   C  Y  N  S   M  I  R   N  L  P   D  E  E  G   D  M  K   S  P  L   L  P  P  N   W  S  T
2701 ATGATGGGGT TTCGAGCCTT ATGCATTCTT TCCGGAGATT GTCGCAAGGA TACGGAAACA CGTAGCTTCG AATACAATGA ATCTTTACCA CCCATCCAGC
    > M  M  G   F  R  A  L   C  I  L   S  G  D   C  R  K  D   T  E  T   R  S  F   E  Y  N  E   S  L  P   P  I  Q
2801 TTTTCCGCAA GCGGTTGCAA GCTCTTTTCT ATCGTACCAC TCTTCTCAGC CGCTTGGAAT CACAACTCTT TACGGACTTA CCATCAGCAT TAGTTTCTTT
    >L  F  R  K   R  L  Q   A  L  F   Y  R  T  T   L  L  S   R  L  E   S  Q  L  F   T  D  L   P  S  A   L  V  S  L
2901 ACCACGCCCC AAACACGATT ATGCAAAAAT GGGTACACAT GTTGAATTAA TCGATCCAGA GCCGGCACCA CAGAATGATC TAAAATCCGA GCCGCTTAGC
    > P  R  P   K  H  D   Y  A  K  M   G  T  H   V  E  L   I  D  P  E   P  A  P   Q  N  D   L  K  S  E   P  L  S
3001 GAGAATGAAG TGATTAATAC AGTCAAGCAG GAATTGGAAA TGGAATATAC AGTCCCAAAG GGCAATTCTG CAGATATCCA GCACAGTGGC GGCCGCTCGA
    > E  N  E   V  I  N  T   V  K  Q   E  L  E   M  E  Y  T   V  P  K   G  N  S   A  D  I  Q   H  S  G   G  R  S
3101 GTCTAGAGGG CCCGCGGTTC GACTACAAGG ATGACGATGA CAAGGGCGAG CAGAAGCTGA TCTCCGAGGA AGATCTGACG CGTACCGGTC ATCATCACCA
    >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  T   R  T  G   H  H  H  H
3201 TCACCATTGA GTTTAAACCC GCTGATCAGC CTCGACTGTG CCTTCTAAGA TCCAGACATG ATAAGATACA TTGATGAGTT TGGACAAACC ACAACTAGAA
    > H  H
3301 TGCAGTGAAA AAAATGCTTT ATTTGTGAAA TTTGTGATGC TATTGCTTTA TTTGTAACCA TTATAAGCTG CAATAAACAA GTTAACAACA ACAATTGCAT
3401 TCATTTTATG TTTCAGGTTC AGGGGGAGGT GTGGGAGGTT TTTTAAAGCA AGTAAAACCT CTACAAATGT GGTATGGCTG ATTATGATCA GTCGACCTGC
3501 AGGCATGCAA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT
3601 GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA
3701 ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA
3801 GCGGTATCAG CTCACTCAAA GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA GGAAAGAACA TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA
3901 ACCGTAAAAA GGCCGCGTTG CTGGCGTTTT TCCATAGGCT CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT CAGAGGTGGC GAAACCCGAC
4001 AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT
4101 TCGGGAAGCG TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA AGCTGGGCTG TGTGCACGAA CCCCCGTTC
4201 AGCCCGACCG CTGCGCCTTA TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA GCCACTGGTA ACAGGATTAG
4301 CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC TCTGCTGAAG
4401 CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGTGGTT TTTTTGTTTG CAAGCAGCAG ATTACGCGCA
4501 GAAAAAAAGG ATCTCAAGAA GATCCTTTGA TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA TGAGATTATC
4601 AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA
4701 ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT
4801 CTGGCCCCAG TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG
4901 TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTTGCCATT
5001 GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA
5101 AAAAAGCGGT TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA ATTCTCTTAC
5201 TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG
5301 TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT
5401 TGAGATCCAG TTCGATGTAA CCCACTCGTG CACCCAACTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA
5501 TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG
5601 AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT CCCCGAAAAG TGCCACCTGA CGTCTAAGAA ACCATTATTA
5701 TCATGACATT AACCTATAAA AATAGGCGTA TCACGAGGCC CTTTCG
```

pMTHisFlagDmSNAP190STOP

```
   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 TGACGATGCC GATGATCTGC GCATAAACCA GCAGCAGCTT ATGGACGAGA TTTCGGCCCT CTTGCAGAAC ACAGCGAAGC
   > D D D   K G T S   D D A   D D L   R I N Q   Q Q L   M D E   I S A L   L Q N   T A K
1001 CGAGCGCCGC CAACGCACTG CAGCTGAATC AGGAGCTCCA GCGCCGGCTC ATGCAAGTCC GGACCAAAAT TCTGGCCATG TTACAAGTAG TAAGGGCCCG
   >P S A A   N A L   Q L N   Q E L Q   R R L   M Q V   R T K I   L A M   L Q V   V R A R
1101 CTTCTCTCGG AACGAGGACA TCCTGGTGCG CCGGCTGCGA CCTCGTTCCC ATTTCGGCCC GAATTCCCTC AATCTTAGCG GGGCGATCCT GCGGAAGGGC
   > F S R   N E D   I L V R   R L R   P R S   H F G P   N S L   N L S   G A I L   R K G
1201 ACGTTTCGCT TCAAGGGCAA CCTGTTTTTC CGAGACATCG ATGGTCGAAG CTGTCCTAAC AACGAGGATT ACGAGGCGCG GTGCCACACG GAGATGTTTC
   > T F R   F K G N   L F F   R D I   D G R S   C P N   N E D   Y E A R   C H T   E M F
1301 CCACCGACTT TGACATGCAC TCCAGACATG TATGGACATT GCTAGACAAA AAGAATGTGA TAATGGGTAT AAAACAACAG TTACTGGAGC ATAGGGCCCA
   >P T D F   D M H   S R H   V W T L   L D K   K N V   I M G I   K Q Q   L L E   H R A H
1401 CAGTACGAAT GCGCTGCCCA GCGGAAGCCT GAAGCGCAAG CCAATCGAGC GACACTTAAG TACACTGGTA AGTCTGCTGG CAACTGCGGA TAGTAGCTTC
   > S T N   A L P   S G S L   K R K   P I E   R H L S   T L V   S L L   A T A D   S S F
1501 AGCATCGACT GGAACCAGAT CAGCACCCTG GACTTGGAAT ATCGGCACTC ACCATATAGC TGTGAGGCAA TGTGGAGGGT CTACTTAACA CCAGATCTCA
   > S I D   W N Q I   S T L   D L E   Y R H S   P Y S   C E A   M W R V   Y L T   P D L
1601 GGCGAGACGA TTGGTCACCC GAGGAGGACG AAACTTTACT GGCGGTGGCC ACTGCGAACA GGATGCAGAA CTGGGAGTTG ATAGCCGCGT CGCTAGACCG
   >R R D D   W S P   E E D   E T L L   A V A   T A N   R M Q N   W E L   I A A   S L D R
1701 ACGTTCCGAC TATCAGTGCT TCGTACGCTT TCATACCGCC CTACGTTTTC TGCTGGAGCC AAAGAACAGC CACCGCTGGA GCGAAGAGGA CAACGATAAA
   > R S D   Y Q C   F V R F   H T A   L R F   L L E P   K N S   H R W   S E E D   N D K
1801 CTAAGGGCGA TCGTTGATAG GAATACGGCA AACAGTGTAA TAAACTGGAA AAAAGTCGTG GAATACTTCC CGGACAAGTC CAAATCAACG CTGATTGGTC
   > L R A   I V D R   N T A   N S V   I N W K   K V V   E Y F   P D K S   K S T   L I G
1901 GTTACTACTA CGTGTTGCAT CCTAGCATCA GTCACGAACC CTTCACTACC AAGGAGGACA TGATGCTGTT CGCCGCCGTT GAAGAGTACA ATGGAAAGTT
   >R Y Y Y   V L H   P S I   S H E P   F T T   K E D   M M L F   A A V   E E Y   N G K F
2001 TCACTGTTTT CCGCGTTCTC TTTTCCCAAA CCGGTCGTTA ACCCAACTCC GAACTCGCTA CCACAATGTA CTGGCACAAC GCAACAAGAC CGACTCCTGG
   > H C F   P R S   L F P N   R S L   T Q L   R T R Y   H N V   L A Q   R N K T   D S W
2101 TCCGTTCAGG ACGACACACG GCTGATGAGC TTCGTAACGC AATACGGGGC TTCCCAGTGG CTTAACTGTG CCACGTTCTT GGGCAATCAC ACGCGGACCA
   > S V Q   D D T R   L M S   F V T   Q Y G A   S Q W   L N C   A T F L   G N H   T R T
2201 GCTGTCGAAC CCGCTTCCTA GTCATCAAGA AATTCCTTGA ACAAAATCCC AATGCAAAAG TGGAAGACCT GCCACGCCGC CGATCGAAGA AGGTGTCCTT
   >S C R T   R F L   V I K   K F L E   Q N P   N A K   V E D L   P R R   R S K   K V S L
2301 GGTTAACTCT GATAACTGGG CGCAGCGTTT GCAAGAGTGG CAAGAGGACC CAGAGTCATT AGTTAATGAT AACCCCCCGA AAGGAACTCG GGTGCGCGGT
   > V N S   D N W   A Q R L   Q E W   Q E D   P E S L   V N D   N P P   K G T R   V R G
2401 CCCAAATCGA AAAAGGCTAG GATCGAGCGA CAGGCAGAGA GTTTCTCAAG ACTTTCGAAA GTTGATATAG AGTTCTGCAA CTTCTTTAAG TTTAGTTATA
   > P K S   K K A R   I E R   Q A E   S F S R   L S K   V D I   E F C N   F F K   F S Y
2501 ATCTGACGCT GAGCACTCCA AAAACTTTTC CAGTTCCAAA AGACGTTTAT AACCTTGCCT ACGTGATCAG AGCTCTTGCA TATAAGCCAC CAATTCGACC
   >N L T L   S T P   K T F   P V P K   D V Y   N L A   Y V I R   A L A   Y K P   P I R P
2601 CTCGCTCTTA CAAAACATAT TTATGCCAAA TGATGTGCTT AAGTGCTACA ACAGCATGAT CCGAAACCTT CCGGACGAGG AAGGTGATAT GAAGAGCCCA
   > S L L   Q N I   F M P N   D V L   K C Y   N S M I   R N L   P D E   E G D M   K S P
2701 CTTTTGCCAC CCAACTGGTC AACTATGATG GGGTTTCGAG CCTTATGCAT TCTTTCCGGA GATTGTCGCA AGGATACGGA AACACGTAGC TTCGAATACA
   > L L P   P N W S   T M M   G F R   A L C I   L S G   D C R   K D T E   T R S   F E Y
2801 ATGAATCTTT ACCACCCATC CAGCTTTTCC GCAAGCGGTT GCAAGCTCTT TTCTATCGTA CCACTCTTCT CAGCCGCTTG GAATCACAAC TCTTTACGGA
   >N E S L   P P I   Q L F   R K R L   Q A L   F Y R   T T L L   S R L   E S Q   L F T D
2901 CTTACCATCA GCATTAGTTT CTTTACCACG CCCCAAACAC GATTATGCAA AAATGGGTAC ACATGTTGAA TTAATCGATC CAGAGCCGGC ACCACAGAAT
   > L P S   A L V   S L P R   P K H   D Y A   K M G T   H V E   L I D   P E P A   P Q N
3001 GATCTAAAAT CCGAGCCGCT TAGCGAGAAT GAAGTGATTA ATACAGTCAA GCAGGAATTG GAAATGGAAT ATACAGTCCC CTAGAAGGGC AATTCTGCAG
   > D L K   S E P L   S E N   E V I   N T V K   Q E L   E M E   Y T V P
3101 ATATCCAGCA CAGTGGCGGC CGCTCGAGTC TAGAGGGCCC GCGGTTCGAA GGTAAGCCTA TCCCTAACCC TCTCCTCGGT CTCGATTCTA CGCGTACCGG
3201 TCATCATCAC CATCACCATT GAGTTTAAAC CCGCTGATCA GCCTCGACTG TGCCTTCTAA GATCCAGACA TGATAAGATA CATTGATGAG TTTGGACAAA
3301 CCACAACTAG AATGCAGTGA AAAAAATGCT TTATTTGTGA AATTTGTGAT GCTATTGCTT TATTTGTAAC CATTATAAGC TGCAATAAAC AAGTTAACAA
3401 CAACAATTGC ATTCATTTTA TGTTTCAGGT TCAGGGGGAG GTGTGGGAGG TTTTTTAAAG CAAGTAAAAC CTCTACAAAT GTGGTATGGC TGATTATGAT
3501 CAGTCGACCT GCAGGCATGC AAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC ATACGAGCCG
3601 GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG
3701 CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG CTCGGTCGTT
3801 CGGCTGCGGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC
3901 AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG
4001 GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC
4101 GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG
4201 AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG
4301 TAACAGGATT AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC
4401 GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC
4501 AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT
4601 CATGAGATTA TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA AGTATATATG AGTAAACTTG GTCTGACAGT
4701 TACCAATGCT TAATCAGTGA GGCACCTATC TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT ACGATACGGG
4801 AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA
4901 GCGCAGAAGT GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG TTTGCGCAAC
5001 GTTGTTGCCA TTGCTACAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC
5101 CCATGTTGTG CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA
5201 TAATTCTCTT ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC TGAGAATAGT GTATGCGGCG ACCGAGTTGC
5301 TCTTGCCCGG CGTCAATACG GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCGAAAA CTCTCAAGGA
5401 TCTTACCGCT GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTTCTGGGT GAGCAAAAAC
5501 AGGAAGGCAA AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT TTTCAATATT ATTGAAGCAT TTATCAGGGT
5601 TATTGTCTCA TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT GACGTCTAAG
5701 AAACCATTAT TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG CCCTTTCGT
```

pMTDmSNAP190(63-721)FlagMycHis



pMTDmSNAP190(176-721)FlagMycHis

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTCAACCTG AGCAATGGAG
                                                                                                > M  E
 901 CGACACTTAA GTACACTGGT AAGTCTGCTG GCAACTGCGG ATAGTAGCTT CAGCATCGAC TGGAACCAGA TCAGCACCCT GGACTTGGAA
   > R  H  L   S  T  L  V   S  L  L   A  T  A   D  S  S  F   S  I  D   W  N  Q   I  S  T  L   D  L  E
 991 TATCGGCACT CACCATATAG CTGTGAGGCA ATGTGGAGGG TCTACTTAAC ACCAGATCTC AGGCGAGACG ATTGGTCACC CGAGGAGGAC
   > Y  R  H   S  P  Y  S   C  E  A   M  W  R   V  Y  L  T   P  D  L   R  R  D   D  W  S  P   E  E  D
1081 GAAACTTTAC TGGCGGTGGC CACTGCGAAC AGGATGCAGA ACTGGGAGTT GATAGCCGCG TCGCTAGACC GACGTTCCGA CTATCAGTGC
   > E  T  L   L  A  V  A   T  A  N   R  M  Q   N  W  E  L   I  A  A   S  L  D   R  R  S  D   Y  Q  C
1171 TTCGTACGCT TTCATACCGC CCTACGTTTT CTGCTGGAGC CAAAGAACAG CCACCGCTGG AGCGAAGAGG ACAACGATAA ACTAAGGGCG
   > F  V  R   F  H  T  A   L  R  F   L  L  E   P  K  N  S   H  R  W   S  E  E   D  N  D  K   L  R  A
1261 ATCGTTGATA GGAATACGGC AAACAGTGTA ATAAACTGGA AAAAAGTCGT GGAATACTTC CCGGACAAGT CCAAATCAAC GCTGATTGGT
   > I  V  D   R  N  T  A   N  S  V   I  N  W   K  K  V  V   E  Y  F   P  D  K   S  K  S  T   L  I  G
1351 CGTTACTACT ACGTGTTGCA TCCTAGCATC AGTCACGAAC CCTTCACTAC CAAGGAGGAC ATGATGCTGT TCGCCGCCGT TGAAGAGTAC
   > R  Y  Y   Y  V  L  H   P  S  I   S  H  E   P  F  T  T   K  E  D   M  M  L   F  A  A  V   E  E  Y
1441 AATGGAAAGT TTCACTGTTT TCCGCGTTCT CTTTTCCCAA ACCGGTCGTT AACCCAACTC CGAACTCGCT ACCACAATGT ACTGGCACAA
   > N  G  K   F  H  C  F   P  R  S   L  F  P   N  R  S  L   T  Q  L   R  T  R   Y  H  N  V   L  A  Q
1531 CGCAACAAGA CCGACTCCTG GTCCGTTCAG GACGACACAC GGCTGATGAG CTTCGTAACG CAATACGGGG CTTCCCAGTG GCTTAACTGT
   > R  N  K   T  D  S  W   S  V  Q   D  D  T   R  L  M  S   F  V  T   Q  Y  G   A  S  Q  W   L  N  C
1621 GCCACGTTCT TGGGCAATCA CACGCGGACC AGCTGTCGAA CCCGCTTCCT AGTCATCAAG AAATTCCTTG AACAAAATCC CAATGCAAAA
   > A  T  F   L  G  N  H   T  R  T   S  C  R   T  R  F  L   V  I  K   K  F  L   E  Q  N  P   N  A  K
1711 GTGGAAGACC TGCCACGCCG CCGATCGAAG AAGGTGTCCT TGGTTAACTC TGATAACTGG GCGCAGCGTT TGCAAGAGTG GCAAGAGGAC
   > V  E  D   L  P  R  R   R  S  K   K  V  S   L  V  N  S   D  N  W   A  Q  R   L  Q  E  W   Q  E  D
1801 CCAGAGTCAT TAGTTAATGA TAACCCCCCG AAAGGAACTC GGGTGCGCGG TCCCAAATCG AAAAAGGCTA GGATCGAGCG ACAGGCAGAG
   > P  E  S   L  V  N  D   N  P  P   K  G  T   R  V  R  G   P  K  S   K  K  A   R  I  E  R   Q  A  E
1891 AGTTTCTCAA GACTTTCGAA AGTTGATATA GAGTTCTGCA ACTTCTTTAA GTTTAGTTAT AATCTGACGC TGAGCACTCC AAAAACTTTT
   > S  F  S   R  L  S  K   V  D  I   E  F  C   N  F  F  K   F  S  Y   N  L  T   L  S  T  P   K  T  F
1981 CCAGTTCCAA AAGACGTTTA TAACCTTGCC TACGTGATCA GAGCTCTTGC ATATAAGCCA CCAATTCGAC CCTCGCTCTT ACAAAACATA
   > P  V  P   K  D  V  Y   N  L  A   Y  V  I   R  A  L  A   Y  K  P   P  I  R   P  S  L  L   Q  N  I
2071 TTTATGCCAA ATGATGTGCT TAAGTGCTAC AACAGCATGA TCCGAAACCT TCCGGACGAG GAAGGTGATA TGAAGAGCCC ACTTTTGCCA
   > F  M  P   N  D  V  L   K  C  Y   N  S  M   I  R  N  L   P  D  E   E  G  D   M  K  S  P   L  L  P
2161 CCCAACTGGT CAACTATGAT GGGGTTTCGA GCCTTATGCA TTCTTTCCGG AGATTGTCGC AAGGATACGG AAACACGTAG CTTCGAATAC
   > P  N  W   S  T  M  M   G  F  R   A  L  C   I  L  S  G   D  C  R   K  D  T   E  T  R  S   F  E  Y
2251 AATGAATCTT TACCACCCAT CCAGCTTTTC CGCAAGCGGT TGCAAGCTCT TTTCTATCGT ACCACTCTTC TCAGCCGCTT GGAATCACAA
   > N  E  S   L  P  P  I   Q  L  F   R  K  R   L  Q  A  L   F  Y  R   T  T  L   L  S  R  L   E  S  Q
2341 CTCTTTACGG ACTTACCATC AGCATTAGTT TCTTTACCAC GCCCCAAACA CGATTATGCA AAAATGGGTA CACATGTTGA ATTAATCGAT
   > L  F  T   D  L  P  S   A  L  V   S  L  P   R  P  K  H   D  Y  A   K  M  G   T  H  V  E   L  I  D
2431 CCAGAGCCGG CACCACAGAA TGATCTAAAA TCCGAGCCGC TTAGCGAGAA TGAAGTGATT AATACAGTCA AGCAGGAATT GGAAATGGAA
   > P  E  P   A  P  Q  N   D  L  K   S  E  P   L  S  E  N   E  V  I   N  T  V   K  Q  E  L   E  M  E
                                                                  junction marker        Flag Tag
2521 TATACAGTCC CAAAGGGCAA TTCTGCAGAT ATCCAGCACA GTGGCGGCCG CTCGAGTCTA GAGGGCCCGC GGTTCGACTA CAAGGATGAC
   > Y  T  V   P  K  G  N   S  A  D   I  Q  H   S  G  G  R   S  S  L   E  G  P   R  F  D  Y   K  D  D
                 Myc tag                                                 6X His tag
2611 GATGACAAGG GCGAGCAGAA GCTGATCTCC GAGGAAGATC TGACGCGTAC CGGTCATCAT CACCATCACC ATTGAGTTTA AACCCGCTGA
   > D  D  K   G  E  Q  K   L  I  S   E  E  D   L  T  R  T   G  H  H   H  H  H   H
2701 TCAGCCTCGA CTGTGCCTTC TAAGATCCAG ACATGATAAG ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT
2791 GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTAA CAACAACAAT TGCATTCATT
2881 TTATGTTTCA GGTTCAGGGG GAGGTGTGGG AGGTTTTTTA AAGCAAGTAA AACCTCTACA AATGTGGTAT GGCTGATTAT GATCAGTCGA
2971 CCTGCAGGCA TGCAAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG
3061 CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG
3151 GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT TCCTCGCTCA
3241 CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA
3331 ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC
3421 CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA
3511 GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA
3601 GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG
3691 CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG
3781 CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC
3871 TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC
3961 AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT
4051 AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT
4141 ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT
4231 GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG
4321 CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA
4411 ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT
4501 CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT
4591 CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTC
4681 TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC
4771 CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA AAACTCTCAA
4861 GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG
4951 GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT
5041 ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA
5131 CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC
5221 GT
```

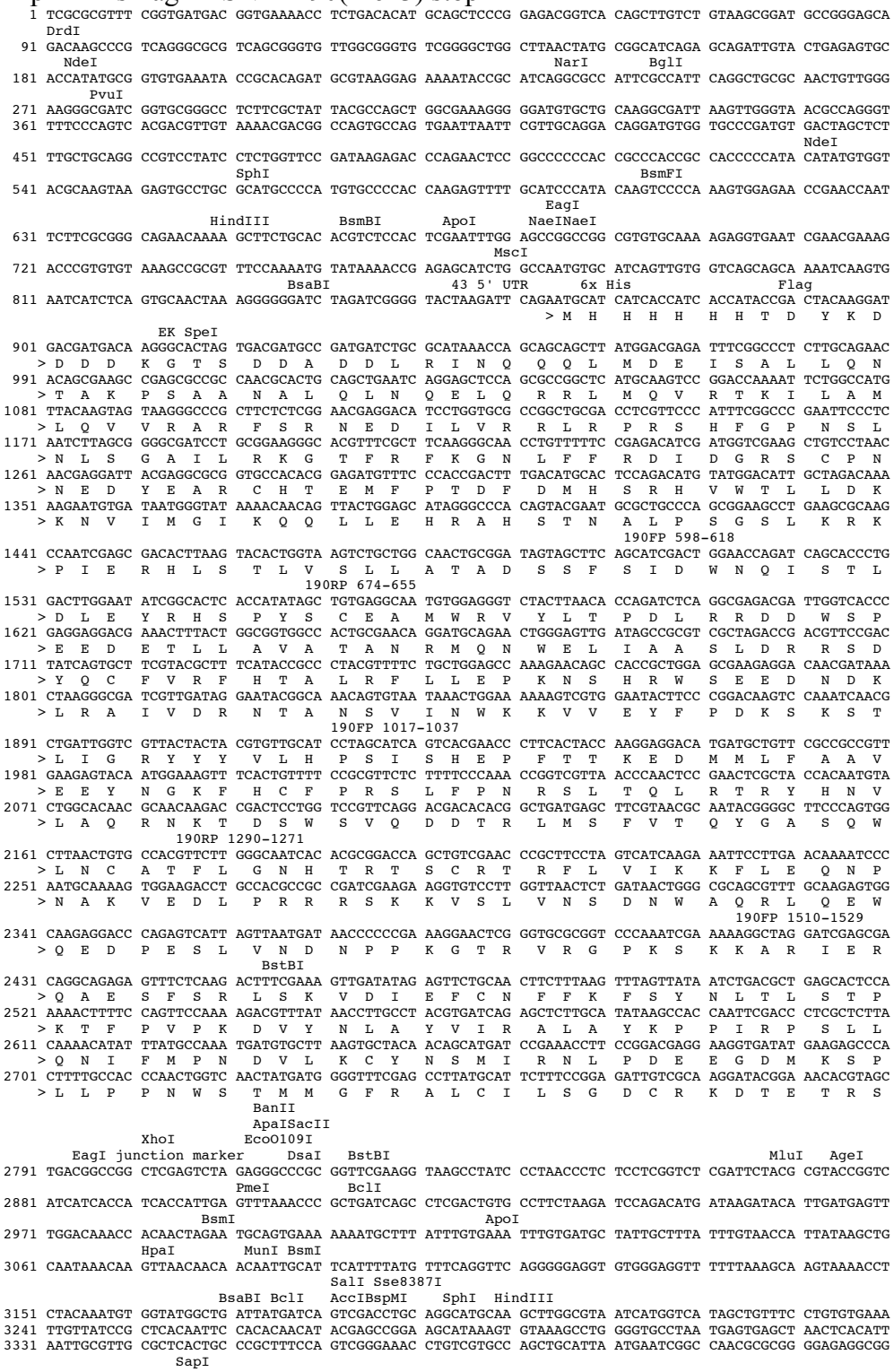
pMTHisFlagDmSNAP190(2-451)stop

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
                                              43 5' UTR      6x His                Flag
 811 AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACTAAGATT CAGAATGCAT CATCACCATC ACCATACCGA CTACAAGGAT
                                                       > M H   H H H   H H T D   Y K D
                          junction marker
 901 GACGATGACA AGGGCACTAG TGACGATGCC GATGATCTGC GCATAAACCA GCAGCAGCTT ATGGACGAGA TTTCGGCCCT CTTGCAGAAC
   > D D D   K G T S   D D A   D D L   R I N Q   Q Q L   M D E   I S A L   L Q N
 991 ACAGCGAAGC CGAGCGCCGC CAACGCACTG CAGCTGAATC AGGAGCTCCA GCGCCGGCTC ATGCAAGTCC GGACCAAAAT TCTGGCCATG
   > T A K   P S A A   N A L   Q L N   Q E L Q   R R L   M Q V   R T K I   L A M
1081 TTACAAGTAG TAAGGGCCCG CTTCTCTCGG AACGAGGACA TCCTGGTGCG CCGGCTGCGA CCTCGTTCCC ATTTCGGCCC GAATTCCCTC
   > L Q V   V R A R   F S R   N E D   I L V R   R L R   P R S   H F G P   N S L
1171 AATCTTAGCG GGGCGATCCT GCGGAAGGGC ACGTTTCGCT TCAAGGGCAA CCTGTTTTTC CGAGACATCG ATGGTCGAAG CTGTCCTAAC
   > N L S   G A I L   R K G   T F R   F K G N   L F F   R D I   D G R S   C P N
1261 AACGAGGATT ACGAGGCGCG GTGCCACACG GAGATGTTTC CCACCGACTT TGACATGCAC TCCAGACATG TATGGACATT GCTAGACAAA
   > N E D   Y E A R   C H T   E M F   P T D F   D M H   S R H   V W T L   L D K
1351 AAGAATGTGA TAATGGGTAT AAAACAACAG TTACTGGAGC ATAGGGCCCA CAGTACGAAT GCGCTGCCCA GCGGAAGCCT GAAGCGCAAG
   > K N V   I M G I   K Q Q   L L E   H R A H   S T N   A L P   S G S L   K R K
1441 CCAATCGAGC GACACTTAAG TACACTGGTA AGTCTGCTGG CAACTGCGGA TAGTAGCTTC AGCATCGACT GGAACCAGAT CAGCACCCTG
   > P I E   R H L S   T L V   S L L   A T A D   S S F   S I D   W N Q I   S T L
1531 GACTTGGAAT ATCGGCACTC ACCATATAGC TGTGAGGCAA TGTGGAGGGT CTACTTAACA CCAGATCTCA GGCGAGACGA TTGGTCACCC
   > D L E   Y R H S   P Y S   C E A   M W R V   Y L T   P D L   R R D D   W S P
1621 GAGGAGGACG AAACTTTACT GGCGGTGGCC ACTGCGAACA GGATGCAGAA CTGGGAGTTG ATAGCCGCGT CGCTAGACCG ACGTTCCGAC
   > E E D   E T L L   A V A   T A N   R M Q N   W E L   I A A   S L D R   R S D
1711 TATCAGTGCT TCGTACGCTT TCATACCGCC CTACGTTTTC TGCTGGAGCC AAAGAACAGC CACCGCTGGA GCGAAGAGGA CAACGATAAA
   > Y Q C   F V R F   H T A   L R F   L L E P   K N S   H R W   S E E D   N D K
1801 CTAAGGGCGA TCGTTGATAG GAATACGGCA AACAGTGTAA TAAACTGGAA AAAAGTCGTG GAATACTTCC CGGACAAGTC CAAATCAACG
   > L R A   I V D R   N T A   N S V   I N W K   K V V   E Y F   P D K S   K S T
1891 CTGATTGGTC GTTACTACTA CGTGTTGCAT CCTAGCATCA GTCACGAACC CTTCACTACC AAGGAGGACA TGATGCTGTT CGCCGCCGTT
   > L I G   R Y Y Y   V L H   P S I   S H E P   F T T   K E D   M M L F   A A V
1981 GAAGAGTACA ATGGAAAGTT TCACTGTTTT CCGCGTTCTC TTTTCCCAAA CCGGTCGTTA ACCCAACTCC GAACTCGCTA CCACAATGTA
   > E E Y   N G K F   H C F   P R S   L F P N   R S L   T Q L   R T R Y   H N V
2071 CTGGCACAAC GCAACAAGAC CGACTCCTGG TCCGTTCAGG ACGACACACG GCTGATGAGC TTCGTAACGC AATACGGGGC TTCCCAGTGG
   > L A Q   R N K T   D S W   S V Q   D D T R   L M S   F V T   Q Y G A   S Q W
2161 CTTAACTGTG CCACGTTCTT GGGCAATCAC ACGCGGACCA GCTGTCGAAC CCGCTTCCTA GTCATCAAGA AATTCCTTGA ACAAAATCCC
   > L N C   A T F L   G N H   T R T   S C R T   R F L   V I K   K F L E   Q N P
                          junction marker
2251 AATGCAAAAG TGGAAGACCT GTAGAAGGGC AATTCTGCAG ATATCCAGCA CAGTGGCGGC CGCTCGAGTC TAGAGGGCCC GCGGTTCGAA
   > N A K   V E D L
2341 GGTAAGCCTA TCCCTAACCC TCTCCTCGGT CTCGATTCTA CGCGTACCGG TCATCATCAC CATCACCATT GAGTTTAAAC CCGCTGATCA
2431 GCCTCGACTG TGCCTTCTAA GATCCAGACA TGATAAGATA CATTGATGAG TTTGGACAAA CCACAACTAG AATGCAGTGA AAAAAATGCT
2521 TTATTTGTGA AATTTGTGAT GCTATTGCTT TATTTGTAAC CATTATAAGC TGCAATAAAC AAGTTAACAA CAACAATTGC ATTCATTTTA
2611 TGTTTCAGGT TCAGGGGGAG GTGTGGGAGG TTTTTTAAAG CAAGTAAAAC CTCTACAAAT GTGGTATGGC TGATTATGAT CAGTCGACCT
2701 GCAGGCATGC AAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC ATACGAGCCG
2791 GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA
2881 ACCTGTCGTG CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC TCGCTCACTG
2971 ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG
3061 CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC
3151 CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT
3241 CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT TCTCATAGCT
3331 CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT
3421 TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA
3511 GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA
3601 AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC
3691 AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG
3781 GGATTTTGGT CATGAGATTA TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA AGTATATATG
3871 AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC
3961 TCCCCGTCGT GTAGATAACT ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC
4051 CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT
4141 GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTACAGG CATCGTGGTG TCACGCTCGT
4231 CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC CCATGTTGTG CAAAAAAGCG GTTAGCTCCT
4321 TCGGTCCTCC GATCGTTGTC AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT ACTGTCATGC
4411 CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG
4501 CGTCAATACG GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCGAAAA CTCTCAAGGA
4591 TCTTACCGCT GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTTCTGGGT
4681 GAGCAAAAAC AGGAAGGCAA AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT TTTCAATATT
4771 ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT
4861 TTCCCCGAAA AGTGCCACCT GACGTCTAAG AAACCATTAT TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG CCCTTTCGT
```

pMTDmSNAP190(176-451)FlagMycHis

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTCAACCTG AGCAATGGAG
 901 CGACACTTAA GTACACTGGT AAGTCTGCTG GCAACTGCGG ATAGTAGCTT CAGCATCGAC TGGAACCAGA TCAGCACCCT GGACTTGGAA
 991 TATCGGCACT CACCATATAG CTGTGAGGCA ATGTGGAGGG TCTACTTAAC ACCAGATCTC AGGCGAGACG ATTGGTCACC CGAGGAGGAC
                                    > M  W  R   V  Y  L  T   P  D  L   R  R  D   D  W  S  P   E  E  D
1081 GAAACTTTAC TGGCGGTGGC CACTGCGAAC AGGATGCAGA ACTGGGAGTT GATAGCCGCG TCGCTAGACC GACGTTCCGA CTATCAGTGC
   > E  T  L   L  A  V  A   T  A  N   R  M  Q   N  W  E  L   I  A  A   S  L  D   R  R  S  D   Y  Q  C
1171 TTCGTACGCT TTCATACCGC CCTACGTTTT CTGCTGGAGC CAAAGAACAG CCACCGCTGG AGCGAAGAGG ACAACGATAA ACTAAGGGCG
   > F  V  R   F  H  T  A   L  R  F   L  L  E   P  K  N  S   H  R  W   S  E  E   D  N  D  K   L  R  A
1261 ATCGTTGATA GGAATACGGC AAACAGTGTA ATAAACTGGA AAAAAGTCGT GGAATACTTC CCGGACAAGT CCAAATCAAC GCTGATTGGT
   > I  V  D   R  N  T  A   N  S  V   I  N  W   K  K  V  V   E  Y  F   P  D  K   S  K  S  T   L  I  G
1351 CGTTACTACT ACGTGTTGCA TCCTAGCATC AGTCACGAAC CCTTCACTAC CAAGGAGGAC ATGATGCTGT TCGCCGCCGT TGAAGAGTAC
   > R  Y  Y   Y  V  L  H   P  S  I   S  H  E   P  F  T  T   K  E  D   M  M  L   F  A  A  V   E  E  Y
1441 AATGGAAAGT TTCACTGTTT TCCGCGTTCT CTTTTCCCAA ACCGGTCGTT AACCCAACTC CGAACTCGCT ACCACAATGT ACTGGCACAA
   > N  G  K   F  H  C  F   P  R  S   L  F  P   N  R  S  L   T  Q  L   R  T  R   Y  H  N  V   L  A  Q
1531 CGCAACAAGA CCGACTCCTG GTCCGTTCAG GACGACACAC GGCTGATGAG CTTCGTAACG CAATACGGGG CTTCCCAGTG GCTTAACTGT
   > R  N  K   T  D  S  W   S  V  Q   D  D  T   R  L  M  S   F  V  T   Q  Y  G   A  S  Q  W   L  N  C
1621 GCCACGTTCT TGGGCAATCA CACGCGGACC AGCTGTCGAA CCCGCTTCCT AGTCATCAAG AAATTCCTTG AACAAAATCC CAATGCAAAA
   > A  T  F   L  G  N  H   T  R  T   S  C  R   T  R  F  L   V  I  K   K  F  L   E  Q  N  P   N  A  K
                                                                                  Flag Tag
1711 GTGGAAGACC TGCCAAAGGG CAATTCTGCA GATATCCAGC ACAGTGGCGG CCGCTCGAGT CTAGAGGGCC CGCGGTTCGA CTACAAGGAT
   > V  E  D   L  P
                 Myc tag                                                 6X His tag
1801 GACGATGACA AGGGCGAGCA GAAGCTGATC TCCGAGGAAG ATCTGACGCG TACCGGTCAT CATCACCATC ACCATTGAGT TTAAACCCGC
1891 TGATCAGCCT CGACTGTGCC TTCTAAGATC CAGACATGAT AAGATACATT GATGAGTTTG GACAAACCAC AACTAGAATG CAGTGAAAAA
1981 AATGCTTTAT TTGTGAAATT TGTGATGCTA TTGCTTTATT TGTAACCATT ATAAGCTGCA ATAAACAAGT TAACAACAAC AATTGCATTC
2071 ATTTTATGTT TCAGGTTCAG GGGGAGGTGT GGGAGGTTTT TTAAAGCAAG TAAAACCTCT ACAAATGTGG TATGGCTGAT TATGATCAGT
2161 CGACCTGCAG GCATGCAAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT GTTATCCGCT CACAATTCCA CACAACATAC
2251 GAGCCGGAAG CATAAAGTGT AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCCAGT
2341 CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC
2431 TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG
2521 ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC
2611 GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG
2701 GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC
2791 ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT
2881 GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA
2971 GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC
3061 TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA
3151 AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC
3241 GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA
3331 TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGTTG
3421 CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC
3511 CGGCTCCAGA TTTATCAGCA ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA
3601 TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGCCATTGC TACAGGCATC GTGGTGTCAC
3691 GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA
3781 GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG
3871 TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT GCGGCGACCG AGTTGCTCTT
3961 GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAAACTCT
4051 CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC ACCAGCGTTT
4141 CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC
4231 AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC
4321 GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT
4411 TTCGT
```

pMTHisFlagDmSNAP190(2-623) stop



```
3421 TTTGCGTATT GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA GCGGTATCAG CTCACTCAAA
3511 GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA GGAAAGAACA TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA
                                                                      DrdI
3601 GGCCGCGTTG CTGGCGTTTT TCCATAGGCT CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT CAGAGGTGGC GAAACCCGAC
3691 AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC
3781 CTTTCTCCCT TCGGGAAGCG TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA AGCTGGGCTG
3871 TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC
           AlwNI
3961 ACTGGCAGCA GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA
                                        Eco57I
4051 CACTAGAAGG ACAGTATTTG GTATCTGCGC TCTGCTGAAG CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC
4141 CACCGCTGGT AGCGGTGGTT TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA GATCCTTTGA TCTTTTCTAC
4231 GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA
4321 TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG CACCTATCTC
                                          Eam1105I
4411 AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG
       BsrDI                                                                 BglI
4501 TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG
                                                                                                  Psp1406I
4591 TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT
           BsrDI
4681 TGTTGCCATT GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG CTCCGGTTCC CAACGATCAA GGCGAGTTAC
                                                  PvuI
4771 ATGATCCCCC ATGTTGTGCA AAAAAGCGGT TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT
                                                                        ScaI
4861 GGTTATGGCA GCACTGCATA ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA AGTCATTCTG
                     BcgI-2
                     BcgI-1
4951 AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT
           Psp1406I
5041 CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG TTCGATGTAA CCCACTCGTG CACCCAACTG
       Eco57I
5131 ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG
                                    SspI
5221 GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG AGCGGATACA TATTTGAATG
                                                                    AatII
5311 TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT CCCCGAAAAG TGCCACCTGA CGTCTAAGAA ACCATTATTA TCATGACATT
                                   EcoO109I
5401 AACCTATAAA AATAGGCGTA TCACGAGGCC CTTTCGTtaa
```

pMT 5'HisFlagDmSNAP50(new start)Stop

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
                                                                                 > M  H   H  H  H   H  H  T  D   Y  K  D
 101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
   > D  D  D   K  G  T  S   D  T  T   Y  I  T   T  N  P  A   L  S  L   R  D  F   L  A  D  Y   Q  K  K   I  S  G
 201 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT
   >T  A  D  E   V  P  F   F  L  Q   K  N  P  V   L  V  D   V  S  E   S  C  S  L   D  L  I   E  S  P   D  D  S  S
 301 TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG
   > I  A  V   F  Q  P   A  A  D  K   C  R  P   T  F  S   P  P  Q  E   N  Q  H   V  P  S   T  F  T  A   L  S  Q
 401 TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC
   > H  K  D   K  S  R  K   C  P  F   G  R  T   Q  Y  S  H   K  L  N   P  T  P   T  D  S  P   N  L  H   I  D  A
 501 GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA
   >C  G  E  L   E  L  T   V  R  L   Y  R  P  P   R  A  Y   H  R  G   F  K  V  E   I  P  V   F  A  E   E  F  V  C
 601 CAAGTCCCCA AAGTGGAGAA CCGAACCAAT TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA
   > L  G  S   N  Y  L   T  E  L  R   D  K  I   S  C  V   C  N  G  K   R  F  V   D  I  S   D  D  P  D   A  P  L
 701 AGAGGTGAAT CGAACGAAAG ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA
   > P  S  I   D  T  N  P   G  Y  F   F  I  N   D  T  F  Y   N  D  Q   R  N  P   D  N  P  D   Y  S  K   T  V  L
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACTAAGATT CAGAATGCAT CATCACCATC ACCATACCGA CTACAAGGAT
   >Q  W  A  A   R  A  N   G  V  N   G  E  T  L   K  V  E   S  M  E   G  K  R  F   I  D  L   T  V  S   P  G  S  P
 901 GACGATGACA AGGGCACTAG TGATACCACC TATATTACAA CGAATCCAGC TTTAAGCCTG CGAGACTTTC TAGCCGACTA TCAAAAGAAA ATATCCGGTA
   > L  H  Y   L  H  H   G  N  C  E   H  L  F   V  I  S   Q  V  E  V   L  T  P   L  S  K   R  P  D  R   S  L  Y
1001 CTGCAGACGA AGTGCCCTTC TTTCTCCAAA AAAACCCAGT ACTTGTCGAC GTGTCTGAAA GCTGTTCCCT GGACTTGATC GAGTCCCCAG ACGACAGCTC
   > P  S  I   D  T  N  P   G  Y  F   F  I  N   D  T  F  Y   N  D  Q   R  N  P   D  N  P  D   Y  S  K   T  V  L
1101 GATTGCAGTC TTCCAGCCCG CCGCAGACAA ATGCCGCCCA ACCTTTTCTC CGCCGCAGGA GAACCAGCAT GTGCCCTCCA CTTTTACTGC TCTAAGCCAG
   > H  K  D   K  S  R  K   C  P  F   G  R  T   Q  Y  S  H   K  L  N   P  T  P   T  D  S  P   N  L  H   I  D  A
1201 CACAAGGACA AGTCCCGTAA GTGTCCCTTT GGCCGCACCC AGTACAGCCA CAAACTGAAT CCGACGCCGA CAGACTCGCC TAACCTGCAC ATCGATGCTT
   >C  G  E  L   E  L  T   V  R  L   Y  R  P  P   R  A  Y   H  R  G   F  K  V  E   I  P  V   F  A  E   E  F  V  C
1301 GCGGCGAACT GGAGCTAACC GTGCGTCTAT ATCGACCACC GCGGGCCTAC CATCGCGGGT TTAAGGTGGA GATTCCCGTT TTCGCCGAGG AGTTCGTATG
   > L  G  S   N  Y  L   T  E  L  R   D  K  I   S  C  V   C  N  G  K   R  F  V   D  I  S   D  D  P  D   A  P  L
1401 TCTGGGCAGC AATTATCTGA CCGAGCTGCG GGACAAGATC AGCTGCGTTT GCAATGGAAA ACGGTTTGTC GACATCAGCG ACGATCCAGA TGCACCGTTG
   > P  S  I   D  T  N  P   G  Y  F   F  I  N   D  T  F  Y   N  D  Q   R  N  P   D  N  P  D   Y  S  K   T  V  L
1501 CCATCCATTG ACACAAATCC CGGTTACTTC TTCATCAACG ACACATTCTA CAACGACCAG CGCAATCCGG ACAATCCCGA CTATTCCAAG ACCGTCTTGC
   >Q  W  A  A   R  A  N   G  V  N   G  E  T  L   K  V  E   S  M  E   G  K  R  F   I  D  L   T  V  S   P  G  S  P
1601 AGTGGGCTGC CAGAGCGAAC GGAGTGAACG GAGAAACGCT TAAGGTGGAG AGTATGGAGG GCAAAAGATT CATCGATCTC ACTGTCAGCC CCGGATCACC
   > L  H  Y   L  H  H   G  N  C  E   H  L  F   V  I  S   Q  V  E  V   L  T  P   L  S  K   R  P  D  R   S  L  Y
1701 GCTACACTAC CTGCACCACG GCAATTGCGA ACACCTGTTT GTGATCTCCC AGGTAGAGGT GCTAACGCCA CTTAGTAAAC GGCCAGATCG CAGCCTGTAT
   > P  Y  P   H  A  F  S   T  F  N   R  R  T   C  Y  M  C   G  I  R   S  Y  S   F  I  V  N   Q  S  R   R  Q  L
1801 CCCTATCCTC ACGCCTTCAG TACGTTTAAT CGCAGGACTT GCTATATGTG CGGCATTCGC AGCTATAGCT TCATCGTGAA CCAGTCCCGG CGACAGCTGC
   >H  D  P  S   Y  L  C   R  R  C   F  L  S  F   F  Y  V   D  G  V   K  L  G  Q   F  K  A   Y  R  M   Y  D  H  V
1901 ACGATCCCTC CTACCTGTGC CGCAGATGTT TTCTCAGCTT CTTTTACGTG GATGGTGTCA AGCTGGGCCA GTTCAAAGCC TATCGCATGT ATGACCATGT
   > E  L  E   G  E  E   E  D  I  K   Q  Q  N   L  P  S
2001 AGAGCTGGAG GGCGAAGAGG AGGACATTAA ACAGCAAAAT CTCCCCTCCT GAATAAAGGG CAATTCTGCA GATATCCAGC ACAGTGGCGG CCGCTCGAGT
2101 CTAGAGGGCC CGCGGTTCGA AGGTAAGCCT ATCCCTAACC CTCTCCTCGG TCTCGATTCT ACGCGTACCG GTCATCATCA CCATCACCAT TGAGTTTAAA
2201 CCCGCTGATC AGCCTCGACT GTGCCTTCTA AGATCCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA GAATGCAGTG AAAAAAATGC
2301 TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG CTGCAATAAA CAAGTTAACA ACAACAATTG CATTCATTTT ATGTTTCAGG
2401 TTCAGGGGGA GGTGTGGGAG GTTTTTTAAA GCAAGTAAAA CCTCTACAAA TGTGGTATGG CTGATTATGA TCAGTCGACC TGCAGGCATG CAAGCTTGGC
2501 GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC GGAAGCATAA AGTGTAAAGC CTGGGGTGCC
2601 TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC GGCCAACGCG
2701 CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC
2801 AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG
2901 TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC
3001 AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT
3101 TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC
3201 TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG
3301 GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA
3401 AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA
3501 GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT
3601 AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT
3701 CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA
3801 ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA ACTTTATCCG
3901 CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG GCATCGTGGT
4001 GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC
4101 TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA
4201 GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC GGGATAATAC
4301 CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG
4401 TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA
4501 TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA
4601 ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT
4701 AAAAATAGGC GTATCACGAG GCCCTTTCGT
```

pMT 5'HisFlagDmSNAP50(2-291S)

```
   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 tGATACCACC TATATTACAA CGAATCCAGC TTTAAGCCTG CGAGACTTTC TAGCCGACTA TCAAAAGAAA ATATCCGGTA
   > D D D   K G T S   D T T   Y I T   T N P A   L S L   R D F   L A D Y   Q K K   I S G
1001 CTGCAGACGA AGTGCCCTTC TTTCTCCAAA AAAACCCAGT ACTTGTCGAC GTGTCTGAAA GCTGTTCCCT GGACTTGATC GAGTCCCCAG ACGACAGCTC
   >T A D E   V P F   F L Q   K N P V   L V D   V S E   S C S L   D L I   E S P   D D S S
1101 GATTGCAGTC TTCCAGCCCG CCGCAGACAA ATGCCGCCCA ACCTTTTCTC CGCCGCAGGA GAACCAGCAT GTGCCCTCCA CTTTTACTGC TCTAAGCCAG
   > I A V   F Q P   A A D K   C R P   T F S   P P Q E   N Q H   V P S   T F T A   L S Q
1201 CACAAGGACA AGTCCCGTAA GTGTCCCTTT GGCCGCACCC AGTACAGCCA CAAACTGAAT CCGACGCCGA CAGACTCGCC TAACCTGCAC ATCGATGCTT
   > H K D   K S R K   C P F   G R T   Q Y S H   K L N   P T P   T D S P   N L H   I D A
1301 GCGGCGAACT GGAGCTAACC GTGCGTCTAT ATCGACCACC GCGGGCCTAC CATCGCGGGT TTAAGGTGGA GATTCCCGTT TTCGCCGAGG AGTTCGTATG
   >C G E L   E L T   V R L   Y R P P   R A Y   H R G   F K V E   I P V   F A E   E F V C
1401 TCTGGGCAGC AATTATCTGA CCGAGCTGCG GGACAAGATC AGCTGCGTTT GCAATGGAAA ACGGTTTGTC GACATCAGCG ACGATCCAGA TGCACCGTTG
   > L G S   N Y L   T E L R   D K I   S C V   C N G K   R F V   D I S   D D P D   A P L
1501 CCATCCATTG ACACAAATCC CGGTTACTTC TTCATCAACG ACACATTCTA CAACGACCAG CGCAATCCGG ACAATCCCGA CTATTCCAAG ACCGTCTTGC
   > P S I   D T N P   G Y F   F I N   D T F Y   N D Q   R N P   D N P D   Y S K   T V L
1601 AGTGGGCTGC CAGAGCGAAC GGAGTGAACG GAGAAACGCT TAAGGTGGAG AGTATGGAGG GCAAAAGATT CATCGATCTC ACTGTCAGCC CCGGATCACC
   >Q W A A   R A N   G V N   G E T L   K V E   S M E   G K R F   I D L   T V S   P G S P
1701 GCTACACTAC CTGCACCACG GCAATTGCGA ACACCTGTTT GTGATCTCCC AGGTAGAGGT GCTAACGCCA CTTAGTAAAC GGCCAGATCG CTAACAAGGG
   > L H Y   L H H   G N C E   H L F   V I S   Q V E V   L T P   L S K   R P D R
1801 CAATTCTGCA GATATCCAGC ACAGTGGCGG CCGCTCGAGT CTAGAGGGCC CGCGGTTCGA AGGTAAGCCT ATCCCTAACC CTCTCCTCGG TCTCGATTCT
1901 ACGCGTACCG GTCATCATCA CCATCACCAT TGAGTTTAAA CCCGCTGATC AGCCTCGACT GTGCCTTCTA AGATCCAGAC ATGATAAGAT ACATTGATGA
2001 GTTTGGACAA ACCACAACTA GAATGCAGTG AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG CTGCAATAAA
2101 CAAGTTAACA ACAACAATTG CATTCATTTT ATGTTTCAGG TTCAGGGGGA GGTGTGGGAG GTTTTTTAAA GCAAGTAAAA CCTCTACAAA TGTGGTATGG
2201 CTGATTATGA TCAGTCGACC TGCAGGCATG CAAGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA
2301 CATACGAGCC GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA
2401 AACCTGTCGT GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC
2501 GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC
2601 AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA
2701 AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG
2801 GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG
2901 CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA
3001 GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT
3101 TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT
3201 TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA
3301 GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT
3401 GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC
3501 TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC
3601 GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA
3701 GTTTGCGCAA CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT
3801 TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG
3901 GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC
4001 GACCGAGTTG CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA
4101 ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG
4201 TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA
4301 TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC
4401 TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAAATAGGC GTATCACGAG GCCCTTTCGT
```

pMT 5’HisFlagDmSNAP50(2-109S)

```
   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 tGATACCACC TATATTACAA CGAATCCAGC TTTAAGCCTG CGAGACTTTC TAGCCGACTA TCAAAAGAAA ATATCCGGTA
   > D  D  D   K  G  T  S   D  T  T   Y  I  T   T  N  P  A   L  S  L   R  D  F   L  A  D  Y   Q  K  K   I  S  G
1001 CTGCAGACGA AGTGCCCTTC TTTCTCCAAA AAAACCCAGT ACTTGTCGAC GTGTCTGAAA GCTGTTCCCT GGACTTGATC GAGTCCCCAG ACGACAGCTC
   >T  A  D  E   V  P  F   F  L  Q   K  N  P  V   L  V  D   V  S  E   S  C  S  L   D  L  I   E  S  P   D  D  S  S
1101 GATTGCAGTC TTCCAGCCCG CCGCAGACAA ATGCCGCCCA ACCTTTTCTC CGCCGCAGGA GAACCAGCAT GTGCCCTCCA CTTTTACTGC TCTAAGCCAG
   > I  A  V   F  Q  P   A  A  D  K   C  R  P   T  F  S   P  P  Q  E   N  Q  H   V  P  S   T  F  T  A   L  S  Q
1201 CACAAGGACA AGTCCCGTAA GTGTCCCTTT GGCCGCACCC AGTACTAACA AGGGCAATTC TGCAGATATC CAGCACAGTG GCGGCCGCTC GAGTCTAGAG
   > H  K  D   K  S  R  K   C  P  F   G  R  T   Q  Y
1301 GGCCCGCGGT TCGAAGGTAA GCCTATCCCT AACCCTCTCC TCGGTCTCGA TTCTACGCGT ACCGGTCATC ATCACCATCA CCATTGAGTT TAAACCCGCT
1401 GATCAGCCTC GACTGTGCCT TCTAAGATCC AGACATGATA AGATACATTG ATGAGTTTGG ACAAACCACA ACTAGAATGC AGTGAAAAAA ATGCTTTATT
1501 TGTGAAATTT GTGATGCTAT TGCTTTATTT GTAACCATTA TAAGCTGCAA TAAACAAGTT AACAACAACA ATTGCATTCA TTTTATGTTT CAGGTTCAGG
1601 GGGAGGTGTG GGAGGTTTTT TAAAGCAAGT AAAACCTCTA CAAATGTGGT ATGGCTGATT ATGATCAGTC GACCTGCAGG CATGCAAGCT TGGCGTAATC
1701 ATGGTCATAG CTGTTTCCTG TGTGAAATTG TTATCCGCTC ACAATTCCAC ACAACATACG AGCCGGAAGC ATAAAGTGTA AAGCCTGGGG TGCCTAATGA
1801 GTGAGCTAAC TCACATTAAT TGCGTTGCGC TCACTGCCCG CTTTCCAGTC GGGAAACCTG TCGTGCCAGC TGCATTAATG AATCGGCCAA CGCGCGGGGA
1901 GAGGCGGTTT GCGTATTGGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC
2001 GGTAATACGG TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG
2101 GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT
2201 TTCCCCCTGG AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA
2301 TAGCTCACGC TGTAGGTATC TCAGTTCGGT GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG CGCCTTATCC
2401 GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG
2501 CTACAGAGTT CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGGACA GTATTTGGTA TCTGCGCTCT GCTGAAGCCA GTTACCTTCG GAAAAAGAGT
2601 TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT
2701 CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC
2801 TTTTAAATTA AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA ATGCTTAATC AGTGAGGCAC CTATCTCAGC
2901 GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC GTCGTGTAGA TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA
3001 CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA
3101 TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT TGCCATTGCT ACAGGCATCG TGGTGTCACG
3201 CTCGTCGTTT GGTATGGCTT CATTCAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT
3301 CCTCCGATCG TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT
3401 TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ATACGGGATA ATACCGCGCC
3501 ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC
3601 ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAAG GGAATAAGGG
3701 CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT
3801 TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
3901 AGGCGTATCA CGAGGCCCTT TCGT
```

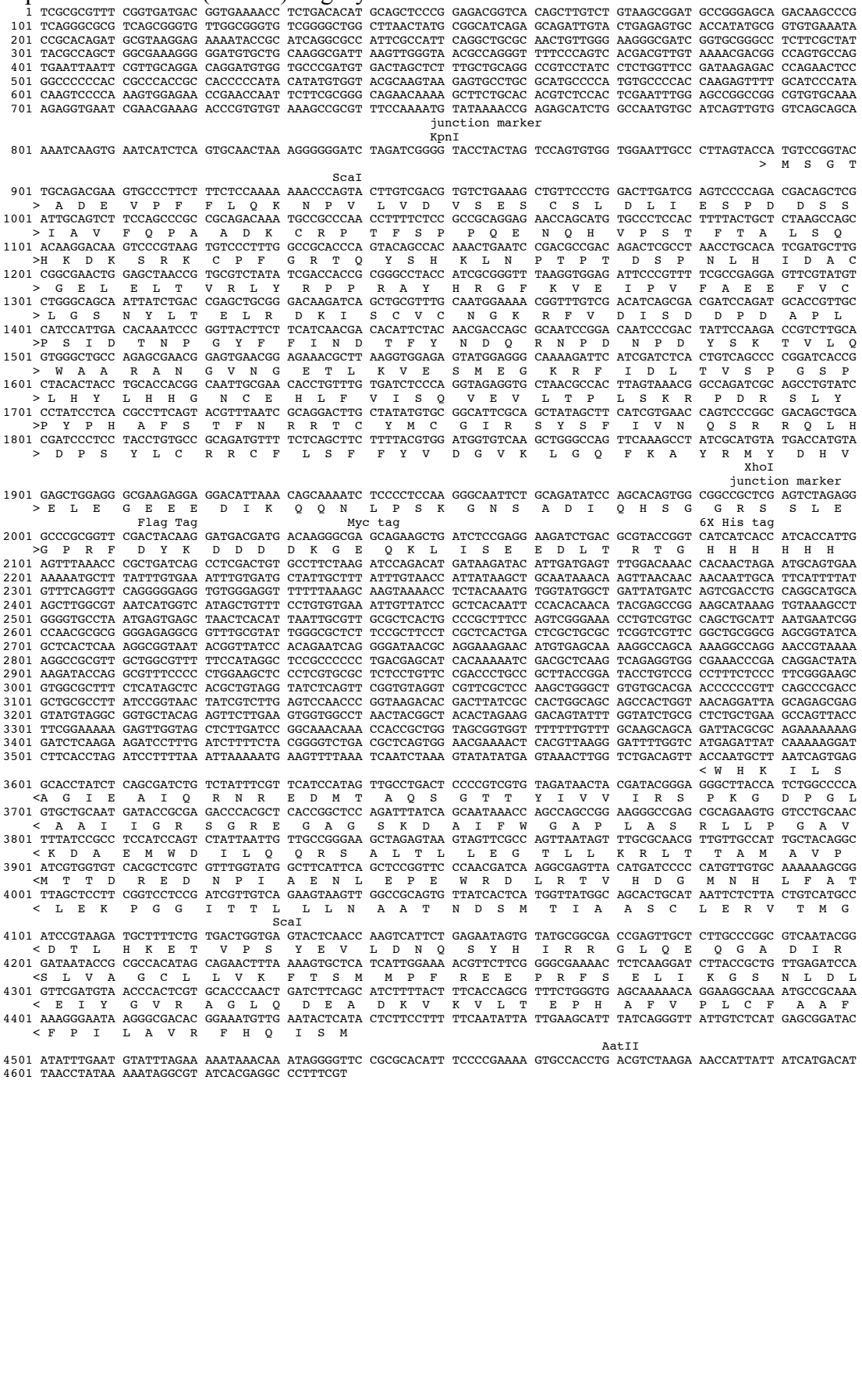
pMTDmSNAP50FlagMycHis

```
   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 CTTGAGAACA TGAACGATTC
 901 CTCCGATTCG GAGTACGAGG ACACAGAGTA CCTGGTATTT GCGGATTTCA AGAACCACAT TCATCCGCAC CAGCTGAAGC ACGAGGACGC GGCCATTAAG
1001 ATCATCGGAA TCGAGAGTGA CACGCCCATG GCCGAGGTAA ATGGAAGCAT TTATCGGGGA CGCTACGAAC ATTCCGTGGG AACCAACGTG TTCTTTGAGA
1101 AGGAAAAAGA CAATCTGGCA AGTGATCCGC TTTTTGAGTC AGTGTGTCGC CAGCGGTACC AGTACGTGGA CAAATCCACC AAGGTGATAT CTTTTGAAAG
1201 GGTCTACGTG GACAATTTGC ACCAGGAGGT GGAGAAGTCG GCTGAGACTG AAGAGGACGA TCAGACCGAG CCAAAGCCAG AATCTCTGAA GCTAAATATA
1301 ACTTACAAGG AGGCTATTAA CAAGTTTGGA GAAGATCATT GAGTGTGGAG AACTGAGTCT AATCCTGAAC AAAAAACACT CTTTCAAGAC ATACCTCTTA
1401 GTACCATGGA TACCACCTAT ATTACAACGA ATCCAGCTTT AAGCCTGCGA GACTTTCTAG CCGACTATCA AAAGAAAATA TCCGGTACTG CAGACGAAGT
        > M D   T T Y   I T T   N P A L   S L R   D F L   A D Y Q   K K I   S G T   A D E V
1501 GCCCTTCTTT CTCCAAAAAA ACCCAGTACT TGTCGACGTG TCTGAAAGCT GTTCCCTGGA CTTGATCGAG TCCCCAGACG ACAGCTCGAT TGCAGTCTTC
    > P F F   L Q K   N P V L   V D V   S E S   C S L D   L I E   S P D   D S S I   A V F
1601 CAGCCCGCCG CAGACAAATG CCGCCCAACC TTTTCTCCGC CGCAGGAGAA CCAGCATGTG CCCTCCACTT TTACTGCTCT AAGCCAGCAC AAGGACAAGT
    > Q P A   A D K C   R P T   F S P   P Q E N   Q H V   P S T   F T A L   S Q H   K D K
1701 CCCGTAAGTG TCCCTTTGGC CGCACCCAGT ACAGCCACAA ACTGAATCCG ACGCCGACAG ACTCGCCTAA CCTGCACATC GATGCTTGCG GCGAACTGGA
    >S R K C   P F G   R T Q   Y S H K   L N P   T P T   D S P N   L H I   D A C   G E L E
1801 GCTAACCGTG CGTCTATATC GACCACCGCG GGCCTACCAT CGCGGGTTTA AGGTGGAGAT TCCCGTTTTC GCCGAGGAGT TCGTATGTCT GGGCAGCAAT
    > L T V   R L Y   R P P R   A Y H   R G F   K V E I   P V F   A E E   F V C L   G S N
1901 TATCTGACCG AGCTGCGGGA CAAGATCAGC TGCGTTTGCA ATGGAAAACG GTTTGTCGAC ATCAGCGACG ATCCAGATGC ACCGTTGCCA TCCATTGACA
    > Y L T   E L R D   K I S   C V C   N G K R   F V D   I S D   D P D A   P L P   S I D
2001 CAAATCCCGG TTACTTCTTC ATCAACGACA CATTCTACAA CGACCAGCGC AATCCGGACA ATCCCGACTA TTCCAAGACC GTCTTGCAGT GGGCTGCCAG
    >T N P G   Y F F   I N D   T F Y N   D Q R   N P D   N P D Y   S K T   V L Q   W A A R
2101 AGCGAACGGA GTGAACGGAG AAACGCTTAA GGTGGAGAGT ATGGAGGGCA AAAGATTCAT CGATCTCACT GTCAGCCCCG GATCACCGCT ACACTACCTG
    > A N G   V N G   E T L K   V E S   M E G   K R F I   D L T   V S P   G S P L   H Y L
2201 CACCACGGCA ATTGCGAACA CCTGTTTGTG ATCTCCCAGG TAGAGGTGCT AACGCCACTT AGTAAACGGC CAGATCGCAG CCTGTATCCC TATCCTCACG
    > H H G   N C E H   L F V   I S Q   V E V L   T P L   S K R   P D R S   L Y P   Y P H
2301 CCTTCAGTAC GTTTAATCGC AGGACTTGCT ATATGTGCGG CATTCGCAGC TATAGCTTCA TCGTGAACCA GTCCCGGCGA CAGCTGCACG ATCCCTCCTA
    >A F S T   F N R   R T C   Y M C G   I R S   Y S F   I V N Q   S R R   Q L H   D P S Y
2401 CCTGTGCCGC AGATGTTTTC TCAGCTTCTT TTACGTGGAT GGTGTCAAGC TGGGCCAGTT CAAAGCCTAT CGCATGTATG ACCATGTAGA GCTGGAGGGC
    > L C R   R C F   L S F F   Y V D   G V K   L G Q F   K A Y   R M Y   D H V E   L E G
2501 GAAGAGGAGG ACATTAAACA GCAAAATCTC CCCTCCAAGG GCAATTCTGC AGATATCCAG CACAGTGGCG GCCGCTCGAG TCTAGAGGGC CCGCGGTTCG
    > E E E   D I K Q   Q N L   P S K   G N S A   D I Q   H S G   G R S S   L E G   P R F
2601 ACTACAAGGA TGACGATGAC AAGGGCGAGC AGAAGCTGAT CTCCGAGGAA GATCTGACGC GTACCGGTCA TCATCACCAT CACCATTGAG TTTAAACCCG
    >D Y K D   D D D   K G E   Q K L I   S E E   D L T   R T G H   H H H   H H
2701 CTGATCAGCC TCGACTGTGC CTTCTAAGAT CCAGACATGA TAAGATACAT TGATGAGTTT GGACAAACCA CAACTAGAAT GCAGTGAAAA AAATGCTTTA
2801 TTTGTGAAAT TTGTGATGCT ATTGCTTTAT TTGTAACCAT TATAAGCTGC AATAAACAAG TTAACAACAA CAATTGCATT CATTTTATGT TTCAGGTTCA
2901 GGGGGAGGTG TGGGAGGTTT TTTAAAGCAA GTAAAACCTC TACAAATGTG GTATGGCTGA TTATGATCAG TCGACCTGCA GGCATGCAAG CTTGGCGTAA
3001 TCATGGTCAT AGCTGTTTCC TGTGTGAAAT TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG TAAAGCCTGG GGTGCCTAAT
3101 GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGCGGG
3201 GAGAGGCGGT TTGCGTATTG GGCGCTCTTC CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTCGTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG
3301 GCGGTAATAC GGTTATCCAC AGAATCAGGG GATAACGCAG GAAAGAACAT GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC
3401 TGGCGTTTTT CCATAGGCTC CGCCCCCCTG ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC
3501 GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT
3601 CATAGCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT
3701 CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTAGGCGG
3801 TGCTACAGAG TTCTTGAAGT GGTGGCCTAA CTACGGCTAC ACTAGAAGGA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA
3901 GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA TTACGCGCAG AAAAAAAGGA TCTCAAGAAG
4001 ATCCTTTGAT CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT
4101 CCTTTTAAAT TAAAAATGAA GTTTTAAATC AATCTAAAGT ATATATGAGT AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA
4201 GCGATCTGTC TATTTCGTTC ATCCATAGTT GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA
4301 TACCGCGAGA CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT CCTGCAACTT TATCCGCCTC
4401 CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT GTTGCCATTG CTACAGGCAT CGTGGTGTCA
4501 CGCTCGTCGT TTGGTATGGC TTCATTCAGC TCCGGTTCCC AACGATCAAG GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG
4601 GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT GTCATGCCAT CCGTAAGATG
4701 CTTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGTA TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA TAATACCGCG
4801 CCACATAGCA GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC
4901 CCACTCGTGC ACCCAACTGA TCTTCAGCAT CTTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA AGGGAATAAG
5001 GGCGACACGG AAATGTTGAA TACTCATACT CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT
5101 ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA
5201 ATAGGCGTAT CACGAGGCCC TTTCGT
```

pMTDmSNAP50(10-376)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
                                                  junction marker
                                               KpnI
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTAGTACCA TGGCTTTAAG
                                                                                                 >  M  A  L  S
                                                                                              ScaI
 901 CCTGCGAGAC TTTCTAGCCG ACTATCAAAA GAAAATATCC GGTACTGCAG ACGAAGTGCC CTTCTTTCTC CAAAAAAACC CAGTACTTGT CGACGTGTCT
    >  L  R  D   F  L  A   D  Y  Q  K   K  I  S   G  T  A   D  E  V  P   F  F  L   Q  K  N   P  V  L  V   D  V  S
1001 GAAAGCTGTT CCCTGGACTT GATCGAGTCC CCAGACGACA GCTCGATTGC AGTCTTCCAG CCCGCCGCAG ACAAATGCCG CCCAACCTTT TCTCCGCCGC
    >  E  S  C   S  L  D  L   I  E  S   P  D  D   S  S  I  A   V  F  Q   P  A  A   D  K  C  R   P  T  F   S  P  P
1101 AGGAGAACCA GCATGTGCCC TCCACTTTTA CTGCTCTAAG CCAGCACAAG GACAAGTCCC GTAAGTGTCC CTTTGGCCGC ACCCAGTACA GCCACAAACT
    >Q  E  N  Q   H  V  P   S  T  F   T  A  L  S   Q  H  K   D  K  S   R  K  C  P   F  G  R   T  Q  Y   S  H  K  L
1201 GAATCCGACG CCGACAGACT CGCCTAACCT GCACATCGAT GCTTGCGGCG AACTGGAGCT AACCGTGCGT CTATATCGAC CACCGCGGGC CTACCATCGC
    >  N  P  T   P  T  D   S  P  N  L   H  I  D   A  C  G   E  L  E  L   T  V  R   L  Y  R   P  P  R  A   Y  H  R
1301 GGGTTTAAGG TGGAGATTCC CGTTTTCGCC GAGGAGTTCG TATGTCTGGG CAGCAATTAT CTGACCGAGC TGCGGGACAA GATCAGCTGC GTTTGCAATG
    >  G  F  K   V  E  I  P   V  F  A   E  E  F   V  C  L  G   S  N  Y   L  T  E   L  R  D  K   I  S  C   V  C  N
1401 GAAAACGGTT TGTCGACATC AGCGACGATC CAGATGCACC GTTGCCATCC ATTGACACAA ATCCCGGTTA CTTCTTCATC AACGACACAT TCTACAACGA
    >G  K  R  F   V  D  I   S  D  D   P  D  A  P   L  P  S   I  D  T   N  P  G  Y   F  F  I   N  D  T   F  Y  N  D
1501 CCAGCGCAAT CCGGACAATC CCGACTATTC CAAGACCGTC TTGCAGTGGG CTGCCAGAGC GAACGGAGTG AACGGAGAAA CGCTTAAGGT GGAGAGTATG
    >  Q  R  N   P  D  N   P  D  Y  S   K  T  V   L  Q  W   A  A  R  A   N  G  V   N  G  E   T  L  K  V   E  S  M
1601 GAGGGCAAAA GATTCATCGA TCTCACTGTC AGCCCCGGAT CACCGCTACA CTACCTGCAC CACGGCAATT GCGAACACCT GTTTGTGATC TCCCAGGTAG
    >  E  G  K   R  F  I  D   L  T  V   S  P  G   S  P  L  H   Y  L  H   H  G  N   C  E  H  L   F  V  I   S  Q  V
1701 AGGTGCTAAC GCCACTTAGT AAACGGCCAG ATCGCAGCCT GTATCCCTAT CCTCACGCCT TCAGTACGTT TAATCGCAGG ACTTGCTATA TGTGCGGCAT
    >E  V  L  T   P  L  S   K  R  P   D  R  S  L   Y  P  Y   P  H  A   F  S  T  F   N  R  R   T  C  Y   M  C  G  I
1801 TCGCAGCTAT AGCTTCATCG TGAACCAGTC CCGGCGACAG CTGCACGATC CCTCCTACCT GTGCCGCAGA TGTTTTCTCA GCTTCTTTTA CGTGGATGGT
    >  R  S  Y   S  F  I   V  N  Q  S   R  R  Q   L  H  D   P  S  Y  L   C  R  R   C  F  L   S  F  F  Y   V  D  G
1901 GTCAAGCTGG GCCAGTTCAA AGCCTATCGC ATGTATGACC ATGTAGAGCT GGAGGGCGAA GAGGAGGACA TTAAACAGCA AAATCTCCCC TCCAAGGGCA
    >  V  K  L   G  Q  F  K   A  Y  R   M  Y  D   H  V  E  L   E  G  E   E  E  D   I  K  Q  Q   N  L  P   S  K  G
                                        XhoI
                                junction marker                    Flag tag                        Myc tag
2001 ATTCTGCAGA TATCCAGCAC AGTGGCGGCC GCTCGAGTCT AGAGGGCCCG CGGTTCGACT ACAAGGATGA CGATGACAAG GGCGAGCAGA AGCTGATCTC
    >N  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
                                     6X His tag
2101 CGAGGAAGAT CTGACGCGTA CCGGTCATCA TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA GACATGATAA
    >  E  E  D   L  T  R   T  G  H  H   H  H  H   H
2201 GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT
2301 AAGCTGCAAT AAACAAGTTA ACAACAACAA TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC
2401 AAATGTGGTA TGGCTGATTA TGATCAGTCG ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA
2501 CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC
2601 TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC
2701 ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA
2801 AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA
2901 AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC
3001 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC
3101 GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA GACACGACTT
3201 ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT
3301 AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG
3401 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA
3501 AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA
3601 TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCCG
3701 TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT
3801 AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT
3901 TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC
4001 GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTTATC
                                                                                          ScaI
4101 ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA
4201 TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT
4301 CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC
4401 CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA
4501 TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC
4601 GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMTDmSNAP50(26-376)FlagMycHis



pMTDmSNAP50(92-376)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
                                                 junction marker
                                                 KpnI
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTAGTACCA TGAGCCAGCA
                                                                                                  >  M  S  Q  H
 901 CAAGGACAAG TCCCGTAAGT GTCCCTTTGG CCGCACCCAG TACAGCCACA AACTGAATCC GACGCCGACA GACTCGCCTA ACCTGCACAT CGATGCTTGC
   >  K  D  K   S  R  K   C  P  F  G   R  T  Q   Y  S  H   K  L  N  P   T  P  T   D  S  P   N  L  H  I   D  A  C
1001 GGCGAACTGG AGCTAACCGT GCGTCTATAT CGACCACCGC GGGCCTACCA TCGCGGGTTT AAGGTGGAGA TTCCCGTTTT CGCCGAGGAG TTCGTATGTC
   > G  E  L   E  L  T  V   R  L  Y   R  P  P   R  A  Y  H   R  G  F   K  V  E   I  P  V  F   A  E  E   F  V  C
                                                                          SalI
1101 TGGGCAGCAA TTATCTGACC GAGCTGCGGG ACAAGATCAG CTGCGTTTGC AATGGAAAAC GGTTTGTCGA CATCAGCGAC GATCCAGATG CACCGTTGCC
   >L  G  S  N   Y  L  T   E  L  R   D  K  I  S   C  V  C   N  G  K   R  F  V  D   I  S  D   D  P  D   A  P  L  P
1201 ATCCATTGAC ACAAATCCCG GTTACTTCTT CATCAACGAC ACATTCTACA ACGACCAGCG CAATCCGGAC AATCCCGACT ATTCCAAGAC CGTCTTGCAG
   >  S  I  D   T  N  P   G  Y  F  F   I  N  D   T  F  Y   N  D  Q  R   N  P  D   N  P  D   Y  S  K  T   V  L  Q
1301 TGGGCTGCCA GAGCGAACGG AGTGAACGGA GAAACGCTTA AGGTGGAGAG TATGGAGGGC AAAAGATTCA TCGATCTCAC TGTCAGCCCC GGATCACCGC
   > W  A  A   R  A  N  G   V  N  G   E  T  L   K  V  E  S   M  E  G   K  R  F   I  D  L  T   V  S  P   G  S  P
1401 TACACTACCT GCACCACGGC AATTGCGAAC ACCTGTTTGT GATCTCCCAG GTAGAGGTGC TAACGCCACT TAGTAAACGG CCAGATCGCA GCCTGTATCC
   >L  H  Y  L   H  H  G   N  C  E   H  L  F  V   I  S  Q   V  E  V   L  T  P  L   S  K  R   P  D  R   S  L  Y  P
1501 CTATCCTCAC GCCTTCAGTA CGTTTAATCG CAGGACTTGC TATATGTGCG GCATTCGCAG CTATAGCTTC ATCGTGAACC AGTCCCGGCG ACAGCTGCAC
   >  Y  P  H   A  F  S   T  F  N  R   R  T  C   Y  M  C   G  I  R  S   Y  S  F   I  V  N   Q  S  R  R   Q  L  H
1601 GATCCCTCCT ACCTGTGCCG CAGATGTTTT CTCAGCTTCT TTTACGTGGA TGGTGTCAAG CTGGGCCAGT TCAAAGCCTA TCGCATGTAT GACCATGTAG
   > D  P  S   Y  L  C  R   R  C  F   L  S  F   F  Y  V  D   G  V  K   L  G  Q   F  K  A  Y   R  M  Y   D  H  V
                                                                                                  XhoI
                                                                                                  junction marker
1701 AGCTGGAGGG CGAAGAGGAG GACATTAAAC AGCAAAATCT CCCCTCCAAG GGCAATTCTG CAGATATCCA GCACAGTGGC GGCCGCTCGA GTCTAGAGGG
   >E  L  E  G   E  E  E   D  I  K   Q  Q  N  L   P  S  K   G  N  S   A  D  I  Q   H  S  G   G  R  S   S  L  E  G
             Flag Tag                  Myc Tag                                  6X His tag
1801 CCCGCGGTTC GACTACAAGG ATGACGATGA CAAGGGCGAG CAGAAGCTGA TCTCCGAGGA AGATCTGACG CGTACCGGTC ATCATCACCA TCACCATTGA
   >  P  R  F   D  Y  K   D  D  D  D   K  G  E   Q  K  L   I  S  E  E   D  L  T   R  T  G   H  H  H  H   H  H
1901 GTTTAAACCC GCTGATCAGC CTCGACTGTG CCTTCTAAGA TCCAGACATG ATAAGATACA TTGATGAGTT TGGACAAACC ACAACTAGAA TGCAGTGAAA
2001 AAAATGCTTT ATTTGTGAAA TTTGTGATGC TATTGCTTTA TTTGTAACCA TTATAAGCTG CAATAAACAA GTTAACAACA ACAATTGCAT TCATTTTATG
                                                                                                  SalI
2101 TTTCAGGTTC AGGGGGAGGT GTGGGAGGTT TTTTAAAGCA AGTAAAACCT CTACAAATGT GGTATGGCTG ATTATGATCA GTCGACCTGC AGGCATGCAA
2201 GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT GTAAAGCCTG
2301 GGGTGCCTAA TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA ATGAATCGGC
2401 CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA GCGGTATCAG
2501 CTCACTCAAA GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA GGAAAGAACA TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA
2601 GGCCGCGTTG CTGGCGTTTT TCCATAGGCT CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT CAGAGGTGGC GAAACCCGAC AGGACTATAA
2701 AGATACCAGG CGTTTCCCCC TGGAAGCTCC CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCGGGAAGCG
2801 TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA AGCTGGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG
2901 CTGCGCCTTA TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA GCCACTGGTA ACAGGATTAG CAGAGCGAGG
3001 TATGTAGGCG GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC TCTGCTGAAG CCAGTTACCT
3101 TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGTGGTT TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG
3201 ATCTCAAGAA GATCCTTTGA TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA TGAGATTATC AAAAAGGATC
3301 TTCACCTAGA TCCTTTTAAA TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG
3401 CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG
3501 TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG TCCTGCAACT
3601 TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTTGCCATT GCTACAGGCA
3701 TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT
3801 TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA ATTCTCTTAC TGTCATGCCA
3901 TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG TCAATACGGG
4001 ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG
4101 TTCGATGTAA CCCACTCGTG CACCCAACTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA
4201 AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG AGCGGATACA
4301 TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT CCCCGAAAAG TGCCACCTGA CGTCTAAGAA ACCATTATTA TCATGACATT
4401 AACCTATAAA AATAGGCGTA TCACGAGGCC CTTTCGT
```

pMTDmSNAP50(196-376)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
                                                 junction marker
                                                 KpnI    SpeI                                          junction
marker
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTAGTACCA TGTCCATTGA
                                                                                                    >  M  S  I  D
                                                          BspEI
 901 CACAAATCCC GGTTACTTCT TCATCAACGA CACATTCTAC AACGACCAGC GCAATCCGGA CAATCCCGAC TATTCCAAGA CCGTCTTGCA GTGGGCTGCC
    >  T  N  P   G  Y  F   F  I  N  D   T  F  Y   N  D  Q   R  N  P  D   N  P  D   Y  S  K   T  V  L  Q   W  A  A
                                  AflII                                  ClaI
1001 AGAGCGAACG GAGTGAACGG AGAAACGCTT AAGGTGGAGA GTATGGAGGG CAAAAGATTC ATCGATCTCA CTGTCAGCCC CGGATCACCG CTACACTACC
    > R  A  N   G  V  N  G   E  T  L   K  V  E   S  M  E  G   K  R  F   I  D  L   T  V  S  P   G  S  P   L  H  Y
1101 TGCACCACGG CAATTGCGAA CACCTGTTTG TGATCTCCCA GGTAGAGGTG CTAACGCCAC TTAGTAAACG GCCAGATCGC AGCCTGTATC CCTATCCTCA
    >L  H  H  G   N  C  E   H  L  F   V  I  S  Q   V  E  V   L  T  P   L  S  K  R   P  D  R   S  L  Y   P  Y  P  H
1201 CGCCTTCAGT ACGTTTAATC GCAGGACTTG CTATATGTGC GGCATTCGCA GCTATAGCTT CATCGTGAAC CAGTCCCGGC GACAGCTGCA CGATCCCTCC
    >  A  F  S   T  F  N   R  R  T  C   Y  M  C   G  I  R   S  Y  S  F   I  V  N   Q  S  R   R  Q  L  H   D  P  S
                                  BsaAI
1301 TACCTGTGCC GCAGATGTTT TCTCAGCTTC TTTTACGTGG ATGGTGTCAA GCTGGGCCAG TTCAAAGCCT ATCGCATGTA TGACCATGTA GAGCTGGAGG
    > Y  L  C   R  R  C  F   L  S  F   F  Y  V   D  G  V  K   L  G  Q   F  K  A   Y  R  M  Y   D  H  V   E  L  E
                                                                                        AvaI         ApaI
         BseRI                              StyI                EcoRV            NotI  junction marker      SacII
1401 GCGAAGAGGA GGACATTAAA CAGCAAAATC TCCCCTCCAA GGGCAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT
    >G  E  E  E   D  I  K   Q  Q  N   L  P  S  K   G  N  S   A  D  I   Q  H  S  G   G  R  S   S  L  E   G  P  R  F
       FLAG tag                    Myc tag                    BglII  MluI   AgeI      6X His tag          PmeI
1501 CGACTACAAG GATGACGATG ACAAGGGCGA GCAGAAGCTG ATCTCCGAGG AAGATCTGAC GCGTACCGGT CATCATCACC ATCACCATTG AGTTTAAACC
    >  D  Y  K   D  D  D   D  K  G  E   Q  K  L   I  S  E   E  D  L  T   R  T  G   H  H  H   H  H  H
1601 CGCTGATCAG CCTCGACTGT GCCTTCTAAG ATCCAGACAT GATAAGATAC ATTGATGAGT TTGGACAAAC CACAACTAGA ATGCAGTGAA AAAAATGCTT
                                                                                    HpaI
1701 TATTTGTGAA ATTTGTGATG CTATTGCTTT ATTTGTAACC ATTATAAGCT GCAATAAACA AGTTAACAAC AACAATTGCA TTCATTTTAT GTTTCAGGTT
                                                                                      AccI Sse8387I
1801 CAGGGGGAGG TGTGGGAGGT TTTTTAAAGC AAGTAAAACC TCTACAAATG TGGTATGGCT GATTATGATC AGTCGACCTG CAGGCATGCA AGCTTGGCGT
1901 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA
2001 ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG
                              SapI
2101 GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA
2201 AGGCGGTAAT ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT
2301 GCTGGCGTTT TTCCATAGGC TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG
2401 GCGTTTCCCC CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT
2501 CTCATAGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT
2601 ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC
2701 GGTGCTACAG AGTTCTTGAA GTGGTGGCCT AACTACGGCT ACACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA
2801 GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA
2901 AGATCCTTTG ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTCACCTAG
3001 ATCCTTTTAA ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT
                                    Eam1105I
3101 CAGCGATCTG TCTATTTCGT TCATCCATAG TTGCCTGACT CCCCGTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT
3201 GATACCGCGA GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC
3301 TCCATCCAGT CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT
3401 CACGCTCGTC GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT
3501 CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA
3601 TGCTTTTCTG TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTTGCCCGGC GTCAATACGG GATAATACCG
3701 CGCCACATAG CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA
3801 ACCCACTCGT GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA
3901 AGGGCGACAC GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT
4001 GTATTTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA
4101 AAATAGGCGT ATCACGAGGC CCTTTCGT
```

pMTDmSNAP50(292-376)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
                                                 junction marker
                                                 KpnI    SpeI                                          junction
marker
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTAGTACCA TGAGCCTGTA
                                                                                                     >  M  S  L  Y
 901 TCCCTATCCT CACGCCTTCA GTACGTTTAA TCGCAGGACT TGCTATATGT GCGGCATTCG CAGCTATAGC TTCATCGTGA ACCAGTCCCG GCGACAGCTG
   >  P  Y  P   H  A  F   S  T  F  N   R  R  T   C  Y  M   C  G  I  R   S  Y  S   F  I  V   N  Q  S  R   R  Q  L
                                                 BsaAI
1001 CACGATCCCT CCTACCTGTG CCGCAGATGT TTTCTCAGCT TCTTTTACGT GGATGGTGTC AAGCTGGGCC AGTTCAAAGC CTATCGCATG TATGACCATG
   > H  D  P   S  Y  L  C   R  R  C   F  L  S   F  F  Y  V   D  G  V   K  L  G   Q  F  K  A   Y  R  M   Y  D  H
                                                                                                 XhoI
                                                                                                 AvaI
                        BseRI                           StyI             EcoRV             NotI junction marker
1101 TAGAGCTGGA GGGCGAAGAG GAGGACATTA AACAGCAAAA TCTCCCCTCC AAGGGCAATT CTGCAGATAT CCAGCACAGT GGCGGCCGCT CGAGTCTAGA
   >V  E  L  E   G  E  E   E  D  I   K  Q  Q  N   L  P  S   K  G  N   S  A  D  I   Q  H  S   G  G  R   S  S  L  E
    BanII
    ApaISacII
        DsaI       FLAG tag                     Myc tag                  BglII   MluI  AgeI      6X His tag
1201 GGGCCCGCGG TTCGACTACA AGGATGACGA TGACAAGGGC GAGCAGAAGC TGATCTCCGA GGAAGATCTG ACGCGTACCG GTCATCATCA CCATCACCAT
   > G  P  R   F  D  Y   K  D  D  D   D  K  G   E  Q  K   L  I  S  E   E  D  L   T  R  T   G  H  H  H   H  H  H
      PmeI
1301 TGAGTTTAAA CCCGCTGATC AGCCTCGACT GTGCCTTCTA AGATCCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA GAATGCAGTG
                                                                                    HpaI        MunI
1401 AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG CTGCAATAAA CAAGTTAACA ACAACAATTG CATTCATTTT
                                                                                              SalI Sse8387I
                                                                                              AccIBspMI
1501 ATGTTTCAGG TTCAGGGGGA GGTGTGGGAG GTTTTTTAAA GCAAGTAAAA CCTCTACAAA TGTGGTATGG CTGATTATGA TCAGTCGACC TGCAGGCATG
1601 CAAGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC GGAAGCATAA AGTGTAAAGC
1701 CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC
                                          SapI
1801 GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT
1901 CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA
2001 AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA
2101 TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA
2201 GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA
2301 CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG
2401 AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA
2501 CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA
2601 AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG
2701 ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG
                                                 Eam1105I
2801 AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC TACGATACGG GAGGGCTTAC CATCTGGCCC
2901 CAGTGCTGCA ATGATACCGC GAGACCCACG CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG TGGTCCTGCA
3001 ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG
3101 GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC
3201 GGTTAGCTCC TTCGGTCCTC CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT TACTGTCATG
                                                                                 BcgI-2
                                       ScaI                                      BcgI-1
3301 CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC
3401 GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC
3501 CAGTTCGATG TAACCCACTC GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA AAATGCCGCA
                                                       SspI
3601 AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT
                                                                                    AatII
3701 ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGACGTCTAA GAAACCATTA TTATCATGAC
3801 ATTAACCTAT AAAAATAGGC GTATCACGAG GCCCTTTCGT
```

pMTDmSNAP43FlagMycHis

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTTCGCGTA ATTAAGATTC
 901 AGAATGGAGC TGAATATCTT TGACGACTGC TGGGAGCTGG TGCAACGATT TCAGCGATTG GTTAATGATG GCGAAAACTG CGAGTTCGAG
       > 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
 991 GTGTTCTGCC GGTGCTGGCG AGAACTGCAG CTGCAGCACC TTTTCACTGC CCAGACGAAC CACACAGAGG TGATAGCCAC CACACTGGCG
    > 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
1081 GCCCTGCATG TGGCCAAGCG ACTGTCGTGC TCCCGACGCA CCACCGGGGA CGTTTTCCCG GCATCTCGCG CTCAAAGGAT CGGAGGTTTC
    > A  L  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
1171 TTTCTGCTCT ACGTAATCTA CTACAAGCAG CCCACGCACA ACTTTATTAA GATCGAGGTC TCACCGCGCA CTTGGCAAGA ACTAACAGAC
    > F  L  L  Y  V  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
1261 TACGCTCTAG ATCTGCGCAA GGATAGTCCG GAGCGGAAGG ACACTCATCA GATCGCCTAC ATGCTGTGGC GCCTGACCCA GGAGCAGGCC
    > Y  A  L  D  L  R  K  D  S  P  E  R  K  D  T  H  Q  I  A  Y  M  L  W  R  L  T  Q  E  Q  A
          <  I  Q  A  L  I  T  R  L  P  L  V  S  M  L  D  G  V  H  Q  P  A  Q  G  L  L  L  G
1351 TTCCGCTTCA CCGCGCTCGA CTATTGCCAG GGGTTGGACA ATCTGGTGGA CTACGACCGT GTGGAGACCG TAGCGGGTGC CAAGGAACAG
    > F  R  F  T  A  L  D  Y  C  Q  G  L  D  N  L  V  D  Y  D  R  V  E  T  V  A  G  A  K  E  Q
    < E  A  E  G  R  E  V  I  A  L  P  Q  V  I  Q  H  V  V  V  T  H  L  G  Y  R  T  G  L  F  L
1441 AGGCAGAGTG CCTTGATGCA GAAGCAACAG CGTGCGAACG GCGTCAGTCT CACATACGAA CTGGAGGGTC TGCGAGCACT GGACCAGGCA
    > R  Q  S  A  L  M  Q  K  Q  Q  R  A  N  G  V  S  L  T  Y  E  L  E  G  L  R  A  L  D  Q  A
    < P  L  T  G  Q  H  L  L  L  L  T  R  V  A  D  T  E  C  V  F  Q  L  T  Q  S  C  Q  V  L  C
1531 AGCCAGCCAT TGTGTGAACT GGAAGCGGCA TACAATGCCC AAAAGAAGCA ATTGGCGGCT GGTCATGAGC ACGCTTTACC GCCCTCTCAA
    > 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  P  S  Q
    < A  L  W  Q  T  F  Q  F  R  C  V  I  G  L  L  L  L  Q  R  S  T  M
1621 ATATTCGGCC ATTTGCGAGA AGTCTTTGCC GATATCCAAA GTGTTCTAGG AGCTAGAAAG AGTACTCCAG ATGAGAAATG CACCACAACA
    > 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
1711 TCTACAGGCA ACCAGTTGGA AGTGCGCCAG AGGGTGCGGA ACAAGGCCAT GTACGGCGTC GAGGAGCGGG AGCCGCAACA CCAGACGGAT
    > 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
    > 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
1891 CAGCAAGAGT ATGAGATGAT TGAGTTTAGT GACGACGAGG AAATGGAAGT GGGTGAAAGC GAGGAGGTCA CGGAAGAAGA ACTCAAAGCT
    > 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
1981 ATTTTGGATA CCAAGGGCAA TTCTGCAGAT ATCCAGCACA GTGGCGGCCG CTCGAGTCTA GAGGGCCCGC GGTTCGACTA CAAGGATGAC
    > I  L  D  T  K  G  N  S  A  D  I  Q  H  S  G  G  R  S  S  L  E  G  P  R  F  D  Y  K  D  D
2071 GATGACAAGG GCGAGCAGAA GCTGATCTCC GAGGAAGATC TGACGCGTAC CGGTCATCAT CACCATCACC ATTGAGTTTA AACCCGCTGA
    > D  D  K  G  E  Q  K  L  I  S  E  E  D  L  T  R  T  G  H  H  H  H  H  H
2161 TCAGCCTCGA CTGTGCCTTC TAAGATCCAG ACATGATAAG ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT
2251 GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTAA CAACAACAAT TGCATTCATT
2341 TTATGTTTCA GGTTCAGGGG GAGGTGTGGG AGGTTTTTTA AAGCAAGTAA AACCTCTACA AATGTGGTAT GGCTGATTAT GATCAGTCGA
2431 CCTGCAGGCA TGCAAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG
2521 CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG
2611 GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT TCCTCGCTCA
2701 CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA
2791 ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC
2881 CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA
2971 GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA
3061 GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG
3151 CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG
3241 CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC
3331 TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC
3421 AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT
3511 AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT
3601 ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT
3691 GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCGG
3781 CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA
3871 ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT
3961 CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT
4051 CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT CTTACTGTCA
4141 TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC
4231 CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA AAACTCTCAA
4321 GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG
4411 GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC CTTTTTCAAT
4501 ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA
4591 CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC GT
```

pMTHisFlagDmSNAP43

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT
  81 GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA
 161 GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC
 241 ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG
 321 GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG
 401 TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT TTGCTGCAGG CCGTCCTATC CTCTGGTTCC
 481 GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT ACGCAAGTAA GAGTGCCTGC
 561 GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT TCTTCGCGGG
 641 CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA
 801 AAATCAAGTG AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACTAAGATT CAGAATGCAT CATCACCATC
                                                                          > M  H   H  H  H
 881 ACCATACCGA CTACAAGGAT GACGATGACA AGGGCACTAG TGAGCTGAAT ATCTTTGACG ACTGCTGGGA GCTGGTGCAA
    >H  H  T  D   Y  K  D   D  D  D   K  G  T  S   E  L  N   I  F  D   D  C  W  E   L  V  Q
 961 CGATTTCAGC GATTGGTTAA TGATGGCGAA AACTGCGAGT TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA
    > R  F  Q   R  L  V  N   D  G  E   N  C  E   F  E  V  F   C  R  C   W  R  E   L  Q  L  Q
1041 GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA GCCACCACAC TGGCGGCCCT GCATGTGGCC AAGCGACTGT
    >  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  L
1121 CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA AGGATCGGAG GTTTCTTTCT GCTCTACGTA
    >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
    > 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
1281 TCTAGATCTG CGCAAGGATA GTCCGGAGCG GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG ACCCAGGAGC
    >  L  D  L   R  K  D   S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L   T  Q  E
1361 AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT GGACAATCTG GTGGACTACG ACCGTGTGGA GACCGTAGCG
    >Q  A  F  R   F  T  A   L  D  Y   C  Q  G  L   D  N  L   V  D  Y   D  R  V  E   T  V  A
1441 GGTGCCAAGG AACAGAGGCA GAGTGCCTTG ATGCAGAAGC AACAGCGTGC GAACGGCGTC AGTCTCACAT ACGAACTGGA
    > 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
1521 GGGTCTGCGA GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG
    >  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
    >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
1681 CTAGGAGCTA GAAAGAGTAC TCCAGATGAG AAATGCACCA CAACATCTAC AGGCAACCAG TTGGAAGTGC GCCAGAGGGT
    > L  G  A   R  K  S  T   P  D  E   K  C  T   T  T  S  T   G  N  Q   L  E  V   R  Q  R  V
1761 GCGGAACAAG GCCATGTACG GCGTCGAGGA GCGGGAGCCG CAACACCAGA CGGATGAACT AGAAGTGCAG CTGGAGGTCA
    >  R  N  K   A  M  Y   G  V  E  E   R  E  P   Q  H  Q   T  D  E  L   E  V  Q   L  E  V
1841 ACGAGACTTA TCAACGCCGC ATGTCCTCGG CCACCGTTTT CCAGAGGGAA CTTCCAGAAG ACGTGCAGCA AGAGTATGAG
    >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
1921 ATGATTGAGT TTAGTGACGA CGAGGAAATG GAAGTGGGTG AAAGCGAGGA GGTCACGGAA GAAGAACTCA AAGCTATTTT
    > 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
    >  D  T
2081 GGGCAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA
2161 ACCCTCTCCT CGGTCTCGAT TCTACGCGTA CCGGTCATCA TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG
2241 ACTGTGCCTT CTAAGATCCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA
2321 TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA
2401 TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA
2481 TGGCTGATTA TGATCAGTCG ACCTGCAGGC ATGCAAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT
2561 TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT
2641 CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC
2721 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG
2801 CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG
2881 AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG
2961 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA
3041 AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC
3121 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC
3201 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG
3281 GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA
3361 CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT
3441 GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT
3521 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG
3601 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA
3681 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC
3761 TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC
3841 ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT
3921 CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT
4001 GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC GATCAAGGCG
4081 AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG
4161 CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT
4241 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA
4321 TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC
4401 CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT
4481 GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT
4561 CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
4641 AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC
4721 TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMTHisFlagDmSNAP43(2-274)

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 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
   > 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
 991 AACTGCGAGT TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA
   > N C E   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
1081 GCCACCACAC TGGCGGCCCT GCATGTGGCC AAGCGACTGT CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA
   > A T T   L A A L   H V A   K R L   S C S R   R T T   G D V   F P A S   R A Q
1171 AGGATCGGAG GTTTCTTTCT GCTCTACGTA ATCTACTACA AGCAGCCCAC GCACAACTTT ATTAAGATCG AGGTCTCACC GCGCACTTGG
   > R I G   G F F L   L Y V   I Y Y   K Q P T   H N F   I K I   E V S P   R T W
1261 CAAGAACTAA CAGACTACGC TCTAGATCTG CGCAAGGATA GTCCGGAGCG GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG
   > Q E L   T D Y A   L D L   R K D   S P E R   K D T   H Q I   A Y M L   W R L
1351 ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT GGACAATCTG GTGGACTACG ACCGTGTGGA GACCGTAGCG
   > 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 V E   T V A
1441 GGTGCCAAGG AACAGAGGCA GAGTGCCTTG ATGCAGAAGC AACAGCGTGC GAACGGCGTC AGTCTCACAT ACGAACTGGA GGGTCTGCGA
   > 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 R
1531 GCACTGGACC AGGCAAGCCA GCCATTGTGT GAACTGGAAG CGGCATACAA TGCCCAAAAG AAGCAATTGG CGGCTGGTCA TGAGCACGCT
   > 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
1621 TTACCGCCCT CTCAAATATT CGGCCATTTG CGAGAAGTCT TTGCCGATAT CCAAAGTGTT CTAGGAGCTA GAAAGAGTAC TCCAGATGAG
   > 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
1711 AAATGCACCA CAACATCTAC AGGCAACCAG TAGAAGGGCA ATTCTGCAGA TATCCAGCAC AGTGGCGGCC GCTCGAGTCT AGAGGGCCCG
   > K C T   T T S T   G N Q
1801 CGGTTCGAAG GTAAGCCTAT CCCTAACCCT CTCCTCGGTC TCGATTCTAC GCGTACCGGT CATCATCACC ATCACCATTG AGTTTAAACC
1891 CGCTGATCAG CCTCGACTGT GCCTTCTAAG ATCCAGACAT GATAAGATAC ATTGATGAGT TTGGACAAAC CACAACTAGA ATGCAGTGAA
1981 AAAAATGCTT TATTTGTGAA ATTTGTGATG CTATTGCTTT ATTTGTAACC ATTATAAGCT GCAATAAACA AGTTAACAAC AACAATTGCA
2071 TTCATTTTAT GTTTCAGGTT CAGGGGGAGG TGTGGGAGGT TTTTTAAAGC AAGTAAAACC TCTACAAATG TGGTATGGCT GATTATGATC
2161 AGTCGACCTG CAGGCATGCA AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA
2251 TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC
2341 AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT
2431 CGCTCACTGA CTCGCTGCGC TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG
2521 GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGGC
2611 TCCGCCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC
2701 CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT
2791 CTCATAGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC
2881 GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA
2971 GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA GTGGTGGCCT AACTACGGCT ACACTAGAAG GACAGTATTT GGTATCTGCG
3061 CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT
3151 GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAACT
3241 CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTCACCTAG ATCCTTTTAA ATTAAAAATG AAGTTTTAAA TCAATCTAAA
3331 GTATATATGA GTAAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTTCGT TCATCCATAG
3421 TTGCCTGACT CCCCGTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA GACCCACGCT
3511 CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT
3601 CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT
3691 CACGCTCGTC GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG
3781 TTAGCTCCTT CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT AATTCTCTTA
3871 CTGTCATGCC ATCCGTAAGA TGCTTTTCTG TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT
3961 CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC
4051 TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG
4141 TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC GGAAATGTTG AATACTCATA CTCTTCCTTT
4231 TTCAATATTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC
4321 CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCACGAGGC
4411 CCTTTCGT
```

pMTHisFlagDmSNAP43(2-172)

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 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
   > 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
 991 AACTGCGAGT TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA
   > N  C  E   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
1081 GCCACCACAC TGGCGGCCCT GCATGTGGCC AAGCGACTGT CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA
   > A  T  T   L  A  A  L   H  V  A   K  R  L   S  C  S  R   R  T  T   G  D  V   F  P  A  S   R  A  Q
1171 AGGATCGGAG GTTTCTTTCT GCTCTACGTA ATCTACTACA AGCAGCCCAC GCACAACTTT ATTAAGATCG AGGTCTCACC GCGCACTTGG
   > R  I  G   G  F  F  L   L  Y  V   I  Y  Y   K  Q  P  T   H  N  F   I  K  I   E  V  S  P   R  T  W
1261 CAAGAACTAA CAGACTACGC TCTAGATCTG CGCAAGGATA GTCCGGAGCG GAAGGACACT CATCAGATCG CCTACATGCT GTGGCGCCTG
   > Q  E  L   T  D  Y  A   L  D  L   R  K  D   S  P  E  R   K  D  T   H  Q  I   A  Y  M  L   W  R  L
1351 ACCCAGGAGC AGGCCTTCCG CTTCACCGCG CTCGACTATT GCCAGGGGTT GGACAATCTG GTGGACTACG ACCGTGTGGA GACCTAGCAA
   > 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  V  E   T
1441 GGGCAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG AGTCTAGAGG GCCCGCGGTT CGAAGGTAAG CCTATCCCTA ACCCTCTCCT
1531 CGGTCTCGAT CTTACGCGTA CCGGTCATCA TCACCATCAC CATTGAGTTT AAACCCGCTG ATCAGCCTCG ACTGTGCCTT CTAAGATCCA
1621 GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT
1711 GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTA ACAACAACAA TTGCATTCAT TTTATGTTTC AGGTTCAGGG GGAGGTGTGG
1801 GAGGTTTTTT AAAGCAAGTA AAACCTCTAC AAATGTGGTA TGGCTGATTA TGATCAGTCG ACCTGCAGGC ATGCAAGCTT GGCGTAATCA
1891 TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT
1981 GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA
2071 ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG
2161 CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
2251 CAGCAAAAGG CCAGGAACCG TAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC
2341 TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC
2431 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG
2521 TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC
2611 AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC
2701 TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT
2791 GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT
2881 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA
2971 AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA
3061 TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA
3151 CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA
3241 GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAGCTAG AGTAAGTAGT
3331 TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC
3421 GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT
3511 AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT
3601 GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA
3691 CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG
3781 ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC
3871 GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT
3961 CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
4051 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGT
```

pMTHisFlagDmSNAP43(2-125)

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 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
   > 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
 991 AACTGCGAGT TCGAGGTGTT CTGCCGGTGC TGGCGAGAAC TGCAGCTGCA GCACCTTTTC ACTGCCCAGA CGAACCACAC AGAGGTGATA
   > N  C  E   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
1081 GCCACCACAC TGGCGGCCCT GCATGTGGCC AAGCGACTGT CGTGCTCCCG ACGCACCACC GGGGACGTTT TCCCGGCATC TCGCGCTCAA
   > A  T  T   L  A  A  L   H  V  A   K  R  L   S  C  S  R   R  T  T   G  D  V   F  P  A  S   R  A  Q
1171 AGGATCGGAG GTTTCTTTCT GCTCTACGTA ATCTACTACA AGCAGCCCAC GCACAACTTT ATTAAGATCG AGGTCTCACC GCGCACTTGG
   > R  I  G   G  F  F  L   L  Y  V   I  Y  Y   K  Q  P  T   H  N  F   I  K  I   E  V  S  P   R  T  W
1261 CAAGAACTAA CAGACTACGC TCTAGATCTG CGCTAACAAG GGCAATTCTG CAGATATCCA GCACAGTGGC GGCCGCTCGA GTCTAGAGGG
   > Q  E  L   T  D  Y  A   L  D  L   R
1351 CCCGCGGTTC GAAGGTAAGC CTATCCCTAA CCCTCTCCTC GGTCTCGATT CTACGCGTAC CGGTCATCAT CACCATCACC ATTGAGTTTA
1441 AACCCGCTGA TCAGCCTCGA CTGTGCCTTC TAAGATCCAG ACATGATAAG ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG
1531 TGAAAAAAAT GCTTTATTTG TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTAA CAACAACAAT
1621 TGCATTCATT TTATGTTTCA GGTTCAGGGG GAGGTGTGGG AGGTTTTTTA AAGCAAGTAA AACCTCTACA AATGTGGTAT GGCTGATTAT
1711 GATCAGTCGA CCTGCAGGCA TGCAAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC
1801 AACATACGAG CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT
1891 TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTTCCGCT
1981 TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA
2071 TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT
2161 AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT
2251 CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG
2341 CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC
2431 GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG
2521 ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC
2611 TGCGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT
2701 GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA GTGGAACGAA
2791 AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC
2881 TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC
2971 ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA
3061 CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG CAACTTTATC CGCCTCCATC
3151 CAGTCTATTA ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTAC AGGCATCGTG
3241 GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA
3331 GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT
3421 CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT
3511 TGCTCTTGCC CGGCGTCAAT ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA
3601 AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC
3691 AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC
3781 CTTTTTCAAT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG
3871 GTTCCGCGCA CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG
3961 AGGCCCTTTC GT
```

pMTDmSNAP43(68-363)FlagMycHis

```
   1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
  91 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
 181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGGCTGCGC AACTGTTGGG
 271 AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
 361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAG TGAATTAATT CGTTGCAGGA CAGGATGTGG TGCCCGATGT GACTAGCTCT
 451 TTGCTGCAGG CCGTCCTATC CTCTGGTTCC GATAAGAGAC CCAGAACTCC GGCCCCCCAC CGCCCACCGC CACCCCCATA CATATGTGGT
 541 ACGCAAGTAA GAGTGCCTGC GCATGCCCCA TGTGCCCCAC CAAGAGTTTT GCATCCCATA CAAGTCCCCA AAGTGGAGAA CCGAACCAAT
 631 TCTTCGCGGG CAGAACAAAA GCTTCTGCAC ACGTCTCCAC TCGAATTTGG AGCCGGCCGG CGTGTGCAAA AGAGGTGAAT CGAACGAAAG
 721 ACCCGTGTGT AAAGCCGCGT TTCCAAAATG TATAAAACCG AGAGCATCTG GCCAATGTGC ATCAGTTGTG GTCAGCAGCA AAATCAAGTG
 811 AATCATCTCA GTGCAACTAA AGGGGGGATC TAGATCGGGG TACCTACTAG TCCAGTGTGG TGGAATTGCC CTTATTCAGA ATGTCGTGCT
                                                                                          > M  S  C
 901 CCCGACGCAC CACCGGGGAC GTTTTCCCGG CATCTCGCGC TCAAAGGATC GGAGGTTTCT TTCTGCTCTA CGTAATCTAC TACAAGCAGC
    >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
 991 CCACGCACAA CTTTATTAAG ATCGAGGTCT CACCGCGCAC TTGGCAAGAA CTAACAGACT ACGCTCTAGA TCTGCGCAAG GATAGTCCGG
    >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
1081 AGCGGAAGGA CACTCATCAG ATCGCCTACA TGCTGTGGCG CCTGACCCAG GAGCAGGCCT TCCGCTTCAC CGCGCTCGAC TATTGCCAGG
    >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
1171 GGTTGGACAA TCTGGTGGAC TACGACCGTG TGGAGACCGT AGCGGGTGCC AAGGAACAGA GGCAGAGTGC CTTGATGCAG AAGCAACAGC
    >G  L  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
1261 GTGCGAACGG CGTCAGTCTC ACATACGAAC TGGAGGGTCT GCGAGCACTG GACCAGGCAA GCCAGCCATT GTGTGAACTG GAAGCGGCAT
    >R  A  N  G   V  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
1351 ACAATGCCCA AAAGAAGCAA TTGGCGGCTG GTCATGAGCA CGCTTTACCG CCCTCTCAAA TATTCGGCCA TTTGCGAGAA GTCTTTGCCG
    >Y  N  A  Q   K  K  Q   L  A  A   G  H  E  H   A  L  P   P  S  Q   I  F  G  H   L  R  E   V  F  A
1441 ATATCCAAAG TGTTCTAGGA GCTAGAAAGA GTACTCCAGA TGAGAAATGC ACCACAACAT CTACAGGCAA CCAGTTGGAA GTGCGCCAGA
    >D  I  Q  S   V  L  G   A  R  K   S  T  P  D   E  K  C   T  T  T   S  T  G  N   Q  L  E   V  R  Q
1531 GGGTGCGGAA CAAGGCCATG TACGGCGTCG AGGAGCGGGA GCCGCAACAC CAGACGGATG AACTAGAAGT GCAGCTGGAG GTCAACGAGA
    >R  V  R  N   K  A  M   Y  G  V   E  E  R  E   P  Q  H   Q  T  D   E  L  E  V   Q  L  E   V  N  E
1621 CTTATCAACG CCGCATGTCC TCGGCCACCG TTTTCCAGAG GGAACTTCCA GAAGACGTGC AGCAAGAGTA TGAGATGATT GAGTTTAGTG
    >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   E  F  S
1711 ACGACGAGGA AATGGAAGTG GGTGAAAGCG AGGAGGTCAC GGAAGAAGAA CTCAAAGCTA TTTTGGATAC TAAGGGCAAT TCTGCAGATA
    >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
    >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
1891 AGGAAGATCT GACGCGTACC GGTCATCATC ACCATCACCA TTGAGTTTAA ACCCGCTGAT CAGCCTCGAC TGTGCCTTCT AAGATCCAGA
    >E  E  D  L   T  R  T   G  H  H   H  H  H  H
1981 CATGATAAGA TACATTGATG AGTTTGGACA AACCACAACT AGAATGCAGT GAAAAAAATG CTTTATTTGT GAAATTTGTG ATGCTATTGC
2071 TTTATTTGTA ACCATTATAA GCTGCAATAA ACAAGTTAAC AACAACAATT GCATTCATTT TATGTTTCAG GTTCAGGGGG AGGTGTGGGA
2161 GGTTTTTTAA AGCAAGTAAA ACCTCTACAA ATGTGGTATG GCTGATTATG ATCAGTCGAC CTGCAGGCAT GCAAGCTTGG CGTAATCATG
2251 GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC
2341 CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT
2431 CGGCCAACGC GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG
2521 GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA
2611 GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC
2701 AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT
2791 GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA
2881 GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA
2971 CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT
3061 GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG
3151 TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA
3241 AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG
3331 GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG
3421 CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG
3511 GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC
3601 CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC
3691 GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG
3781 TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA
3871 GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG
3961 TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA
4051 TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT
4141 GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAATGCCGC
4231 AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT
4321 CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA
4411 AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG T
```