This is to certify that the
dissertation entitled

OVER-EXPRESSION, PURIFICATION, CHARACTERIZATION AND CRYSTALLIZATION OF RECOMBINANT SNAPc VARIANTS

presented by

Andrej Hanzlowsky

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Chemistry

Major Professor's Signature

11/13/06

Date

OVER-EXPRESSION, PURIFICATION, CHARACTERIZATION AND CRYSTALLIZATION OF RECOMBINANT SNAPc VARIANTS

By

Andrej Hanzlowsky

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements
For the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2006

ABSTRACT

OVER-EXPRESSION, PURIFICATION, CHARACTERIZATION AND CRYSTALLIZATION OF RECOMBINANT SNAPc VARIANTS

By

Andrej Hanzlowsky

In humans, the small nuclear RNA (snRNA) genes are transcribed by either RNA polymerase II or III. Different protein-protein and protein-DNA interactions are responsible for polymerase specificity and a large number of transcription factors are involved in this process. The small nuclear RNA activating protein complex (SNAPc) is required for transcription initiation for both classes of snRNA genes. SNAPc is a five subunit complex responsible for recognition of the proximal sequence element (PSE) on the DNA and interacts with multiple transcription factors that guide the transcription process. Co-expression of SNAPc subunits in baculovirus resulted in small quantities of material that is active for both the DNA binding and transcription. On the other hand, expression of the subunits in bacteria and reconstitution of the complex resulted in material capable of DNA binding, but not capable of transcription initiation.

Here we present a method for efficient expression and purification of a mini SNAPc (mSNAPc) complex, composed of SNAP190(1-505), SNAP50, SNAP43, and SNAP19 in bacteria. The complex was co-expressed in *E. coli* and was active for both DNA binding and transcription initiation from snRNA genes. Furthermore, the amount of recovered material is greatly increased compared to any other previously described method.

The larger amounts and increased purity of the complex obtained by this new approach enabled structural studies and analysis of zinc content in mSNAPc. We determined that the SNAP50 subunit of the complex contains a zinc finger domain that is involved in DNA binding of the complex. Crystallization conditions were identified, yielding protein crystals that diffracted both electrons in a transmission electron microscope and x-rays.

Increased production of this complex is a step forward in understanding how the SNAPc recognizes DNA, and how it interacts with a multitude of transcription factors.

ACKNOWLEDGMENTS

I sincerely thank my advisor, Prof. James H. Geiger for his guidance and support throughout the research. I would like to thank our collaborator Prof. Bill Henry for providing DNA materials and for his countless discussions regarding this research. I would like to thank my co-workers Blanka, Justin, Suzan, Xiaofei and Fang for their direct involvement in this project. I would also like to thank all present and past members of Dr. Geiger's and Dr. Henry's research groups for their help and advice. I would like to thank Dr. Kathleen Foley for introducing me into the world of cloning. I must sincerely thank Gauri W. Jawdekar for performing all DNA binding and transcription initiation assays, Dr. Alicia Pastor-Leicha for her help with the transmission electron microscopy, Dr. Kathryn Severin for her help with flame atomic absorbtion, David Szymanski for assistance with the ICP-MS, Dr. Aizhuo Liu for his assistance with NMR, and Janet Haun for her advice on non-research related subjects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	alanine
A/ADE	adenine
Amp	ampicillin
BOG	n-octyl-beta-D-glucopyranoside
bp	base pair
C	cysteine
Cam	chloramphenicol
C/CYT	cytosine
C terminal	carboxy terminal
D	aspartic acid
Da	Dalton
DNA	deoxyribonucleic acid
DNase	deoxyribonucleosidase
DSE	distal sequence element
dsDNA	double stranded DNA
DTT	dithiothreitol
E	glutamic acid
EDTA	Diaminoethanetetraacetic acid

EMSA	electrophoretic mobility shift assay
F	phenylalanine
FAA(S)	flame atomic absorption (spectroscopy)
G	glycine
G/GUA	Guanine
GST	Glutathione S-Transferase
GSH	glutathione
H	histidine
HCl	hydrochloric acid
Hepes	N-[2-hydroxyethyl] piperazine-N'-[ethane sulfonic acid]
HSQC	heteronuclear single quantum coherence transfer
I	isoleucine
ICP-MS	inductively coupled plasma mass spectroscopy
IPTG	Isopropyl-B-D-Thiogalactopyranoside
K	lysine
Kan	kanamycin
KCl	potassium chloride
kDa	kilo Dalton

L	leucine
M	methionine
MCS	multiple cloning site
mg	milligram
µg	microgram
mL	milliliter
MW	molecular weight
mRNA	messenger RNA
mSNAPc	mini small nuclear RNA activating protein complex
mrSNAPc	mini recombinant small nuclear RNA activating protein complex
N	asparagine
NaCl	sodium chloride
ng	nanogram
NMR	nuclear magnetic resonance
N-terminal	amino terminal
OIR	Oct-1 interacting region
P	proline
PCR	polymerase chain reaction

pol	polymerase
PEG	polyethylene glycerol
PMSF	phenylmethylsulfonylfluoride
ppb	parts per billion
ppm	parts per million
PSE	proximal sequence element
Q	glutamine
R	arginine
Rb	retinoblastoma protein
RNA	Ribonucleic Acid
RNAP	Ribonucleic Acid Polymerase
rRNA	Ribosomal RNA
S	serine
siRNA	small interfering RNA
SNAP	small nuclear RNA activating protein
SNAPc	small nuclear RNA activating protein complex
snRNA	small nuclear RNA
SDS-PAGE	sodium dodecyl sulfate – polyacrylamide gel electrophoresis
Str	Streptomycin

T	threonine
T/THY	Thymine
TBP	TATA (box) Binding Protein
TEM	transmission electron microscopy
TFI	Transcription Factor I
TFII	Transcription Factor II
TFIII	Transcription Factor III
Tris	2-Amino-2-(hydroxymethyl)-1,3-propendiol
tRNA	transfer RNA
U	uracil
UV	ultraviolet
V	valine
VIS	visual
W	tryptophan
WCE	whole cell extract
Y	tyrosine

1. Introduction

Genetic information of all living organisms, with the exception of retro viruses, is stored on the DNA. Transcription, a process in which the information on the DNA is transferred to the RNA, is achieved by accurate synthesis of RNA from the DNA template by different RNA polymerases. One RNA polymerase exists in prokaryotes and it is solely responsible for the transcription of the genomic information of the organism. In contrast, eukaryotes contain four classes of polymerases. The four classes differ in localization, template specificity, and susceptibility to activators and inhibitors. RNA polymerase I, located in the nucleoli, is responsible for transcription of ribosomal RNA and transcribes the genes for 18S, 5.8S and 28S ribosomal RNA (rRNA). RNA polymerase II, located in the nucleoplasm, is responsible for transcription of all messenger RNA (mRNA) and some small nuclear RNA (snRNA), such as U1 and U2 snRNA. RNA polymerase III, also located in the nucleoplasm, is responsible for transcription of 5S ribosomal RNA, the transfer RNA (tRNA) and also the U6 snRNA. RNA polymerase IV was recently identified in plants and is responsible for transcription of short interfering RNAs (siRNAs) which were implicated in gene silencing and also mitochondrial RNA[1,2,3].

Strict control of transcription is required for healthy cell existence and is achieved by targeting regions of DNA called promoters, located upstream or downstream from transcription start sites. A promoter is a region on the DNA that is capable of recruiting transcription factors and eventually the RNA polymerase. An enhancer is a region on the DNA that is responsible for recruitment of transcription factors that influence transcription, but do not directly recruit the RNA polymerase. Different transcription

factors bind to the promoter region and can initiate transcription by helping the polymerase to specifically bind to the DNA and start the transcription process. Other transcription factors function as repressors and inhibit the binding of the polymerase to the DNA, and thus reduce or abolish transcription. Some factors, called basal transcription factors, are always present in the cell[4]. Some are incapable of DNA binding and can, only by specific external stimuli, overcome this inability. When the inability to bind the DNA is eliminated, these factors become active and can then activate or repress the transcription process. The organism uses such mechanisms to adapt to a change in external conditions or to a different developmental stage, by expressing the proteins required to adapt to the new environment. A different type of transcription factor is expressed only in response to external stimuli and can then regulate transcription. In the absence of the stimuli these transcription factors are present in only small concentrations or are not present at all[1,3,4].

1.1. RNA polymerase and promoters

1.1.1. RNA polymerase I

RNA polymerase I (RNAP I) is unique among the nuclear RNAs in that it is responsible for transcription of only one set of genes, the ribosomal RNA (rRNA) and has to be able to recognize only one promoter structure. There are essentially two different genes for all the rRNAs in a tandem repeated form. A long precursor for rRNA is transcribed by RNA polymerase I and it is subsequently cleaved and processed into the individual ribosomal RNAs. Two control regions make up the promoter for this polymerase; the core promoter (CP) extending from -45 to +20 and the upstream control

element (UCE) extending from -180 to -107 which increases the efficiency of the core promoter. Both promoter regions are unusually GC-rich. SL1 (*H. sapiens*), TIF-IB (*M. muscullus*) or CF (*S. cerevisiae*) are TBP associated multi-subunit complexes responsible for binding to the core promoter and are essential for RNAP I transcription. The upstream control element recruits another multi-subunit factor called the upstream binding factor or UBF (UBF-1, *H. sapiens*) and this factor greatly enhances SL1 recruitment and its specificity[1,3].

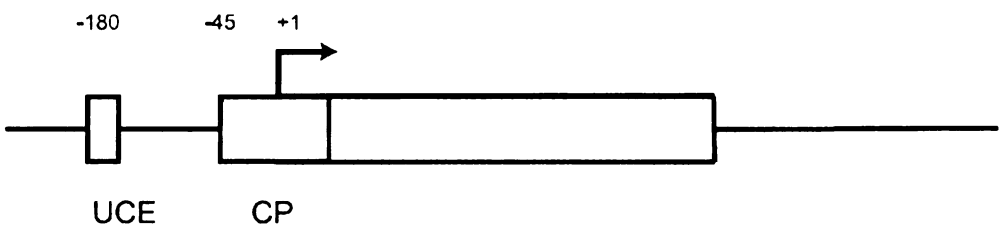


Figure 1: The upstream control element (UCE, -180 to -107) and the control promoter (CP, -45 to +20) are the regulatory regions required for efficient targeting by RNA polymerase I.

1.1.2. RNA polymerase II

RNA polymerase II (RNAP II) is responsible for transcription of genes coding for all messenger RNAs (mRNAs) and most small nuclear RNA (snRNA). The RNAP II promoter region is the most diverse compared to other polymerases recognized by other promoters. Many different factors bind to this promoter region and influence the transcription of the genes. This is understandable because of the need for a living cell to be able to quickly increase or decrease the expression of particular proteins in response to changes in the environment or progress in the developmental stage of the cell. A variety

of promoter elements have been identified including the TATA box (TATAAAA), upstream elements such as the CCAAT box, the GC box (GGGCGG) and downstream elements. RNAP II promoters have one or more of these elements in different numbers and combinations (Figure 2)[3].

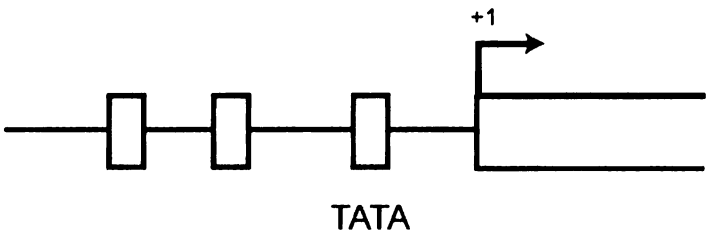


Figure 2: mRNA promoters are often characterized by the presence of a TATA box. Other promoter regions are often present upstream or downstream of the transcription start site.

Human U1 small nuclear RNA promoter is an example of a TATA box-less promoter and differs from the U6 snRNA promoter by the absence of the TATA box. The two promoters serve as an interesting model in the understanding of RNAP specificity for transcription of genes with similar promoter regions. The U1 promoter contains two regions; a distal sequence element (DSE) and the proximal sequence element (PSE, Figure 3)[4.5.6].

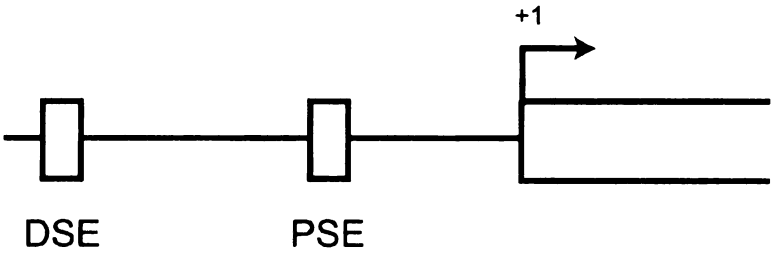


Figure 3: *H. sapiens* U1 snRNA promoter region determined by the presence of both the DSE and the PSE, but by the absence of the TATA box.

The upstream sequence element (USE) in *A. thaliana* (Figure 4) and the PSEA in *D. melanogaster* (Figure 5) are similar to the human PSE. Factors that are recruited to these promoter regions are not well characterized, but it is possible that the polymerase specificity in these systems is determined by different conformations of the same factor, which are defined by the USE or PSEA sequence. Different conformations of a factor can than recruit other factors that can than bind to the promoter and regulate transcription[7].

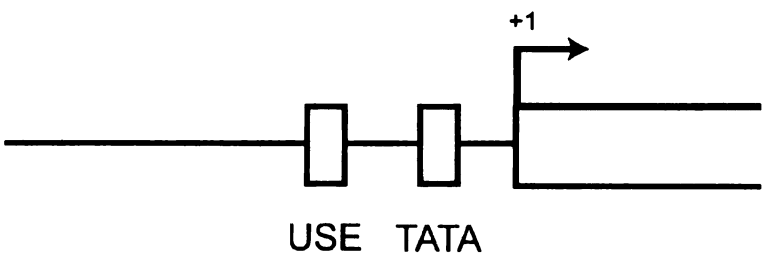


Figure 4: *A. thaliana* snRNA recognized by RNAP II contains the upstream sequence element (USE) and a TBP recruiting TATA box[7].

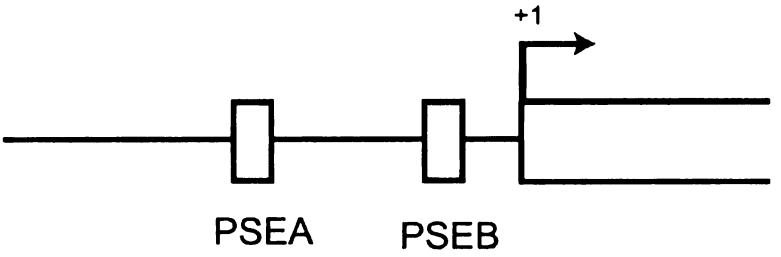


Figure 5: snRNA promoter region from *D. melanogaster* with two promoter elements recognized by RNA polymerase II[7].

1.1.3. RNA polymerase III

RNA polymerase III (RNAP III) transcribes genes that encode structural or catalytic RNA and are in general shorter than 400 base pairs (bp), which is also consistent with the elongation properties of RNA polymerase III. The RNA molecules transcribed by RNA polymerase III are involved in fundamental metabolic processes. These RNAs are components of the protein synthesis, mRNA splicing and tRNA processing apparatus.

RNA polymerase III uses three main types of promoters. The type 1 RNA polIII promoter has only one example; the *X. laevis* 5S promoter, which contains an internal control region (ICR). This promoter region consists of three elements; the A box (+50 to +60), the intermediate element (IE; +67 to +72) and the C box (+80 to +90; Figure 6). In *S. cerevisiae* 5S genes, only C box is required for efficient transcription[3,4,8,9].

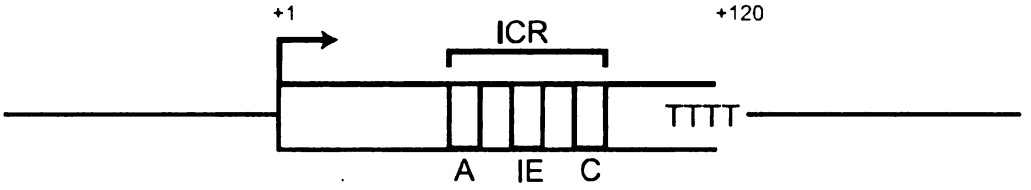


Figure 6: *X. laevis* 5S RNA gene contains three elements in the internal control region (ICR) which can be divided into box A (+50 to +60), an intermediate element (IE, +67 to +72) and box C (+80 to +90). A simple run of T residues determines the end of the gene[8].

Type 2 promoters are represented by the genes of Ad2 VA1 and most tRNAs. The main hallmark of this promoter type is the presence of gene internal A and B boxes. These regions are conserved in tRNA genes between various species. The A-boxes of type 1 and 2 are interchangeable in *X. laevis* (Figure 7)[10]. This reflects the similarity in sequence rather than a conserved function, since the A-boxes of 5S RNA and tRNA genes bind different transcription factors[11].

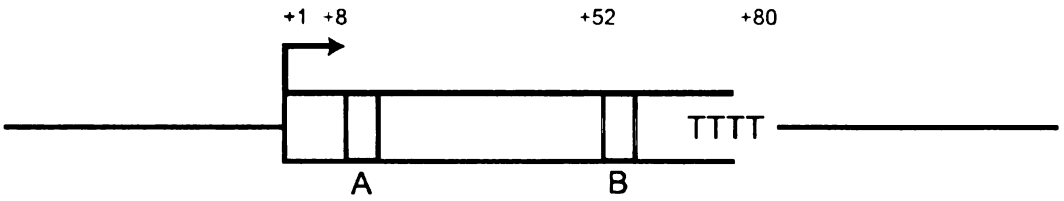


Figure 7: *X. laevis* $tRNA^{leu}$ internal promoter region consisted of boxes A (+8 to +19) and B (+52 to +62)[8].

The *H. sapiens* U6 snRNA gene contains a type 3 promoter and is characterized by the presence of gene external promoter regions only. The distal sequence element

(DSE), proximal sequence element (PSE), and the TATA box are located upstream from the transcription start site. Representative members of this promoter type are the U6 snRNA genes, coding for the U6 snRNA component of the spliceosome[12,13,14], the 7SK gene whose RNA product was implicated in the regulation of the CDK9/cyclin T complex[15], and the HI RNA gene, which codes the RNA component of hRNaseP and the RNA component of hRNase MRP[16,17]. Interestingly in the vertebrate snRNA promoters, RNA specificity can be switched between RNAPII and RNAPIII and vice versa by deletion or generation of the TATA box element (Figure 8)[18].

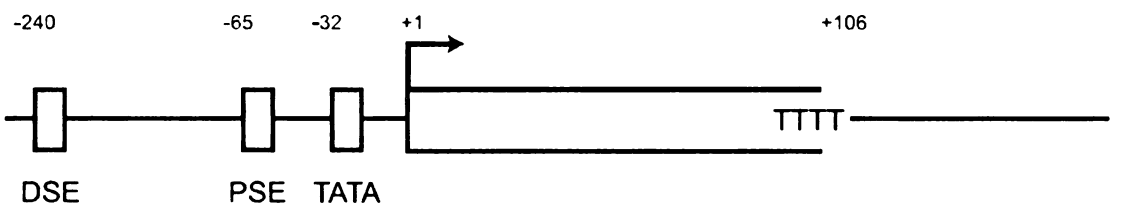


Figure 8: *H. sapiens* U6 snRNA with gene external promoter regions. The DSE (-240 to -215), the PSE (-65 to -48) and the TATA box (-32 to -25) are all located upstream of the transcription start site[8].

The *S. cerevisiae* U6 snRNA gene is a type 4 RNAPIII promoter and is a hybrid consisted of gene internal and gene external regions. The TATA box is located upstream of the transcription start site and the B-box is located downstream from the coding

region. The A-box is located gene internally (+21 TO +31, Figure 9). All three of these promoter elements are required for transcription in vivo[19].

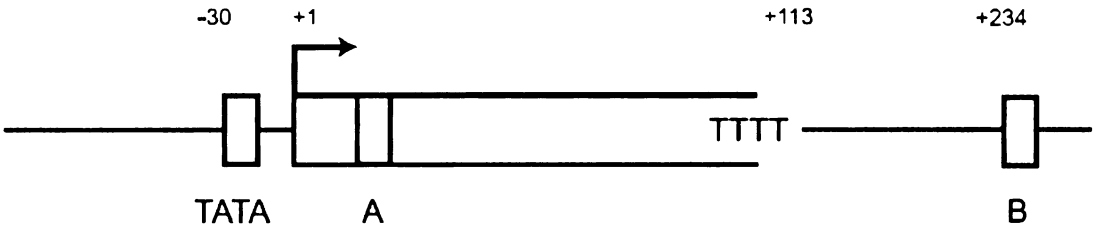


Figure 9: The *S. cerevisiae* U6 gene with a combination of gene internal A-box (+21 to +31) and gene external promoter elements; The TATA box (-30 to -23) upstream from the transcription start site and B-box downstream (+234 to +244) from the U6 coding region[8].

1.2. Composition of human RNAP transcription initiation complexes

1.2.1. RNA polymerase I transcription initiation complex

The TATA box binding protein (TBP) is the only known transcription factor that seems to be almost universally required for transcription regardless of the promoter type and polymerase used to transcribe a gene. Interestingly TBP is required for transcription whether or not the TATA box element is present in the promoter. In humans and in the case of RNA polymerase I the necessary transcription factor SL1 (selectivity factor) is a TBP containing multi-subunit complex and is responsible for the recognition of the USE region of the promoter, while the upstream binding factor (UBF1) is responsible for recognition of the CP region (Figure 10) [1,3].

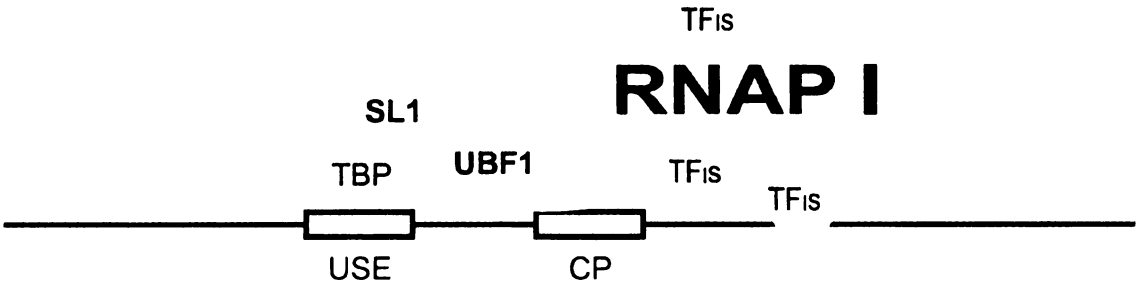


Figure 10: *H. sapiens* RNA polymerase I transcription from an rRNA promoter with SL1, a TBP containing complex and UBF1, a multi-subunit transcription factor targeting the USE and the CP regions respectively. Placement of TFIs on the figure is arbitrary.

1.2.2. RNA polymerase II transcription initiation complexes

mRNA type promoter, which is transcribed by RNA polymerase II and contains the TATA box, can be recognized by TFIID or TBP. Once TFIID or TBP are successfully recruited to the TATA box, other TAF_{II}s are then recruited to the promoter. First to follow TFIID is TFIIB, then the TFIIF-RNAP II complex, TFIIE and TFIIH. TFIIA can enter the initiation complex at any stage of the assembly and its main function is counteracting repressors that associate with TBP and reduce or prevent its association with the DNA (Figure 11)[20,21,22,23,24].

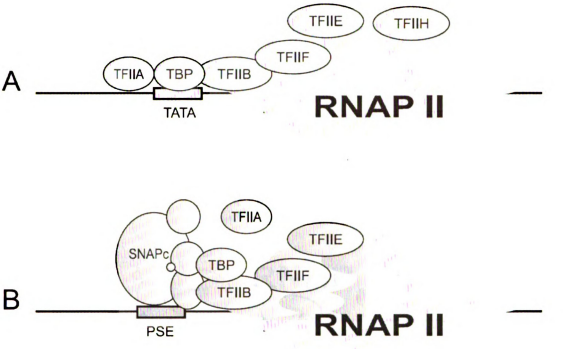


Figure 11: *H. sapiens* polymerase II transcription initiation complex: (A) an example of a mRNA promoter with a TATA box, recognized by the TBP or by the TBP containing TFIID complex and (B) TATA-less U1 snRNA type of promoter, with the PSE sequence recognized by the SNAP complex.

1.2.3. RNA polymerase III transcription initiation complexes

The key player in polymerase III transcription initiation is the TFIIIB complex, because it contacts the polymerase directly. *H. sapiens* Brf1-TFIIIB is composed of three subunits, the TATA-box binding protein, Brf1 and Bdp1. An example of Brf1-TFIIIB dependent initiation is transcription from the tRNA type of promoters, where the TBP containing TFIIIB is recruited to the DNA by positioning of TFIIIC to boxes A and B.

TBP is required even though the promoter is TATA-less. In the case of U6 snRNA promoters, the Brf2-TFIIIB complex, is assembled from TBP, Brf2 and Bdp1. This complex is recruited by binding of the SNAPc complex to the proximal sequence element upstream of the TATA box (Figure 12)[1,3,25,26].

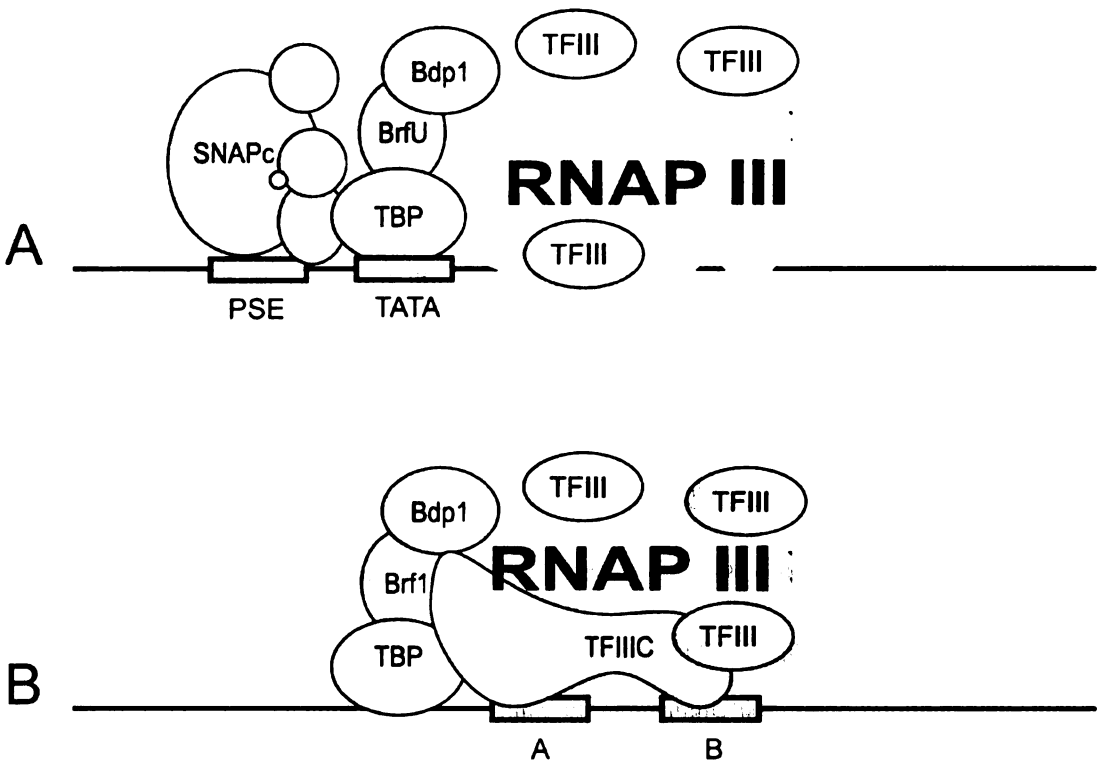


Figure 12: *H. sapiens* polymerase III transcription initiation complex: a U6 snRNA promoter (A) with the PSE sequence, targeted by the SNAP complex, and the TATA box, targeted by the TFIIIB like complex, assembled from TBP, Brf2 (BrfU) and Bdp1. (B) TATA-less tRNA type of promoter with boxes A and B targeted by TFIIIC. TFIIIB is required for transcription initiation.

1.3. SNAPc

1.3.1. The composition of SNAPc

The small nuclear RNA activating protein complex (SNAPc) is a basal transcription factor responsible for transcription initiation for polymerase II and III snRNA systems. This transcription factor is at least a five subunit complex, composed of SNAP190, SNAP50, SNAP43, SNAP45 and SNAP19[6,27,28]. TBP was found to often co-purify with the SNAP complex purified from HeLa cells[29].

The largest subunit of SNAPc, SNAP190, is the backbone of the complex. This subunit interacts with all other subunits of the complex. The N-terminal part of SNAP190 is the most important region and interacts with SNAP50, SNAP43 and SNAP19 (Figure 13). The N-terminus is responsible for the DNA binding activity of the protein and also interacts with TBP. The unusual DNA binding domain of SNAP190 resembles the Myb DNA binding motif and is composed of four full (Ra, Rb, Rc and Rd) and one half (Rh) repeats. An exceptionally serine rich area following an arginine rich region in the N-terminal part of SNAP190 was shown to be a target for phosphorylation. This modification had an important impact on the ability of the complex to bind DNA and to initiate transcription by changing the electrostatic properties of this region of SNAP190[30,31]. Oct-1 is recruited to the complex by interaction with SNAP190 in the area encompassing amino acids 869-912[32,33,34,35].SNAP43 recruits SNAP50 to the largest subunit of the complex and was shown to also interact with SNAP19 and TBP (Figure 13)[6]. SNAP50 is, like SNAP190, responsible for DNA binding of this transcription factor judged from the DNA cross-linking experiments. An unusual zinc binding domain was predicted in the C-terminal part of SNAP50. This zinc-containing domain is believed to

be involved in the DNA binding properties of SNAP50[6]. The smallest subunit of the SNAP complex is SNAP19. An exceptionally charged C-terminal tail and a leucine zipper motif characterize this subunit. SNAP19 is not required for the formation of the complex between SNAP190, SNAP43 and SNA50 at higher concentrations, but it seems to be needed for efficient complex formation at lower, physiologically relevant concentrations[36]. SNAP45 interacts with the C-termini of SNAP190 and TBP. A SNAP complex missing SNAP45 (SNAPc-SNAP45) showed a diffusive band in the EMSA, which migrated slower than complete SNAPc, suggesting that the C-terminal domain of SNAP190 in a SNAP complex missing SNAP45 may assume multiple conformations. Addition of recombinant SNAP45 to the SNAPc-SNAP45 resulted in a band that migrated at the same position as the complete SNAPc. This suggests that SNAP45 was incorporated into the complex and due to direct protein-protein interactions, the C-terminal domain of SNAP190 assumed a more rigid conformation[32].

The stoichiometry of the subunits in the complex is unknown and so is the size of the complex[37].

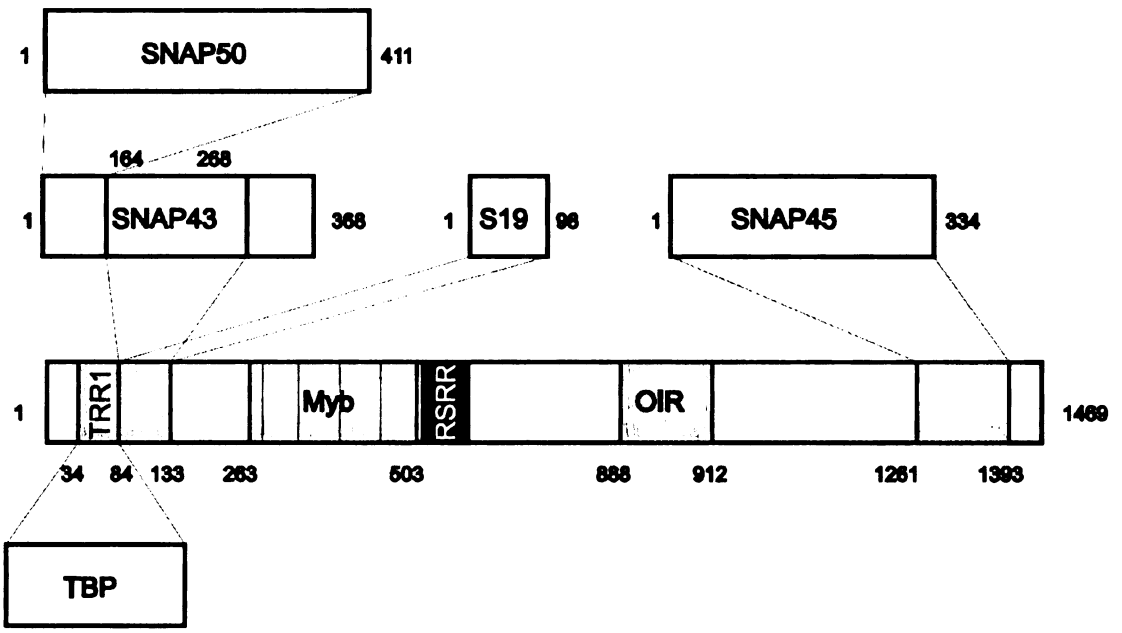


Figure 13: Protein-protein/DNA contact map of SNAPc. The complex is composed of five subunits SNAP190 interacts with TBP using the TBP recruitment region (TRR1) from amino-acids 34 to 84 and with Oct-1 using the OIR (Oct-1 interacting region) from amino-acids 888 to 912. RSRR is the arginine and serine rich area of SNAP190, a target for phosphorylation. The unusual Myb domain is responsible for DNA binding of SNAP190 (adopted from Hernandez et al.[28]).

1.3.2. The SNAP complex and chromatin

The two critical promoter elements on the DNA, the proximal sequence element (PSE) and the distal sequence element (DSE) are separated by 150 base pairs. The separation of the two promoter elements is conserved in several organisms. SNAPc binds the PSE and the transcription activator Oct-1 binds to the DSE, resulting in transcription activation. It was also shown that SNAPc dependent Oct-1 activation is not observed

when naked DNA with separated DSE and PSE is used in in-vitro assays[39]. The Oct-1 POU domain is capable of increasing SNAPc recruitment to the promoter if DNA between the DSE and PSE is spaced closer together. It was also observed that reconstruction of the nucleosome, by addition of purified histones, leads to increased activation of the Oct-1 induced SNAPc recruitment and transcription activation. DNA foot printing revealed that a nucleosome core is positioned between the two regions. The localization of the DNA wrapped around the histones is defined by the presence of the PSE and the DSE and the proteins that bind to the two regions, SNAPc and Oct-1. Interestingly, the two proteins come in close contact with the nucleosome core, since 146 base pairs of the DNA are needed to form the nucleosome[38,39,40]. It is thus possible that SNAPc or Oct-1 come into direct contact with the nucleosome, not only with the PSE and the DSE regions, hence stabilizing the formation of the pre-initiation complex (Figure 14).

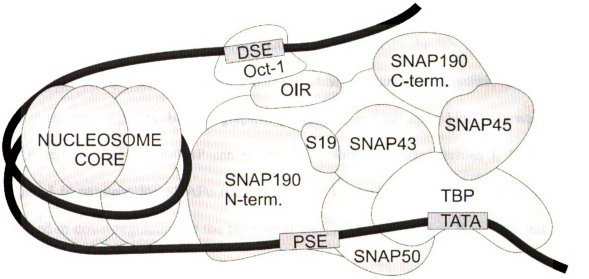


Figure 14: Proposed model of SNAPc recruitment to the U6 promoter. SNAPc is recruited to the promoter by interaction with Oct-1 and TBP is recruited to the TATA box

by interaction with SNAPc. The N-terminal domain of SNAP190, SNAP43 and SNAP45 are involved in the interaction with TBP. SNAP190 contains three main domains; the N-terminal region, the Oct-1 interacting region (OIR) and the C-terminal region responsible for the interaction with SNAP45. SNAP190 and SNAP50 interact with the PSE region of the DNA. Positioning of the nucleosome is crucial for Oct-1 dependent activation.

1.4. Proteins that interact directly with the SNAP complex

1.4.1. TBP is associated with SNAPc through multiple interactions

Multiple subunits of the SNAP complex interact with TBP and the protein was co-immuno purified with antibodies to different SNAP subunits. SNAPc was shown to improve TBP binding to the DNA by specific interactions with the N-terminal region of TBP[42]. TBP contains an extremely conserved C-terminal domain and a nonconserved N-terminal domain. The C-terminal domain is capable of performing all TBP functions like DNA binding and interaction with transcription factors. In EMSA the high affinity mouse U6 DNA was used because the human U6 DNA binds with the SNAP complex only weakly. Interestingly when full length h-TBP and SNAPc are used together, the two can bind to the human U6 probe strongly, though SNAPc or TBP alone bind only weakly[41]. The N-terminal region of TBP can be divided into three parts; regions I, II and III. Region I is required for SNAPc recruitment and U6 transcription. This region is also involved in down-regulation of the binding to the TATA box. Region II is made out of a run of glutamine residues and the number of glutamines varies drastically between different organisms. This region appears to be involved in SNAPc recruitment. Region III

has no apparent function, since a deletion mutant without this region is fully active and capable of SNAPc dependent DNA binding and U6 transcription (Figure 15)[42].

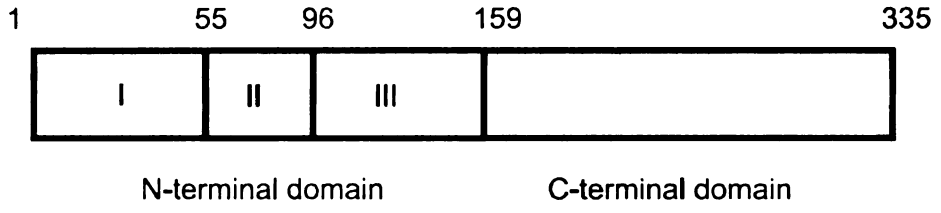


Figure 15: TBP contains a highly conserved C-terminal domain and an N-terminal domain that can be divided into three regions. Region II is a stretch of glutamines and the number of glutamine residues varies between organisms.

h-TBP was shown to interact strongly with SNAP190, SNAP43 and SNAP45, weakly with SNAP50, but not with SNAP19. Interestingly, when the U6 probe with mutant TATA box was used, no change in migration was observed in EMSA by addition of TBP. Thus, although TBP interacts with SNAPc subunits, the intact U6 promoter with both the PSE and the TATA box is needed for cooperative binding of SNAPc and TBP[41,42,43].

1.4.2. Oct-1 POU domain relieves self inhibitory properties of SNAPc towards DNA binding

Complete SNAPc does not bind efficiently to the human PSE region of the U6 DNA, but binding is enhanced drastically in the presence of Oct1-1 bound to the DSE. mSNAPc, composed of only SNAP190(1-505), SNAP50 and SNAP43, binds efficiently to the PSE, even though it does not interact with Oct-1. SNAP190 interacts with Oct-1

POU domain with its OIR (869-912), which is located in the C-terminal half of the largest subunit of SNAPc, SNAP190. This suggests that the C-terminal domain of SNAP190 has a built-in DNA-binding damper, which is deactivated by Oct-1 POU domain[32,33,44].

1.4.3. p53 associates with SNAPc and TBP

p53 is a tumor suppressor protein that plays a critical role in preventing uncontrolled cell proliferation and was shown to be responsible for the selection of cellular outcome, such as apoptosis or cell cycle arrest[45,46,47]. p53 is capable of repressing RNAPI and RNAPIII transcription. RNAP III transcription is elevated in several cancer-derived cell lines that lack active p53[48]. Ultimately, the function of this protein is to prevent the passage of mutations to daughter cells by inducing apoptosis or cell cycle arrest as a response to DNA damage.

Furthermore, p53 was also shown to repress RNAP II transcription, along with RNAP III transcription of human snRNA genes. p53 was shown to interact with SNAP190 and SNAP43 subunits of SNAPc and also with TBP, but no significant interaction was observed for SNAP45, SNAP50 and SNAP19[49].

1.4.4. Rb associates with SNAP50 and SNAP43[50]

The retinoblastoma (Rb) protein is also a tumor suppressor that controls cell proliferation by its influence on cell cycle progression, differentiation, apoptosis and growth. Mutations of the Rb coding gene are found in several human cancers, as are mutations in the promoter regions that are targeted by Rb. The function of Rb as a tumor

suppressor is linked to its ability to down-regulate transcription and therefore gene expression[51,52,53,54,55].

Rb is capable of inhibiting transcription of polymerase I, II and III and regulates rRNA, mRNA, tRNA and snRNA expression. Clearly the activity of Rb thru the repression of different RNAs plays an important role in cell cycle progression and apoptosis, since these RNAs are essential for cell viability.

In the context of U6 snRNA expression Rb was shown to be capable of repressing transcription by targeting the SNAP complex and TFIIIB. RB was shown to strongly interact with SNAPc through SNAP43 and also through SNAP50, as determined by co-immunoprecipitation and Western analysis[56].

2. Results and discussion

2.1. SNAP subunits - overview

The SNAPc is made out of five subunits: SNAP190, SNAP50, SNAP43, SNAP45, and SNAP19. The main focus of my research was the crystallization of SNAP subunits. SNAP19 being the smallest subunit was a perfect candidate for crystallization and the expression levels were significant. A crystal form was also identified. SNAP190(1-505) that is capable of forming an active mSNAPc, was also successfully over-expressed and purified and a few crystal forms were identified. SNAP43 was expressed in significant levels, but the recovery of the material was notoriously inefficient. SNAP50 was over-expressed, partially purified and low quality crystals were obtained[57]. SNAP45 was over-expressed and partially purified, but the cleavage of the GST tag resulted in protein over-digestion.

2.1.1. SNAP19

SNAP19 is the smallest subunit of the complex and is responsible for enhancing the interaction between SNAP190 and SNAP43. The protein contains a leucine zipper motif at its N-terminus and a glutamic acid rich C-terminus. The only known function of the protein lies in its ability to bind to the N-terminal domain of SNAP190 and recruit SNAP43 into the complex.

2.1.1.1. Over-expression, purification and crystallization of SNAP19

The protein was over-expressed in *E. coli* as a GST fusion peptide and purified over GSH (glutathione) agarose (Figure 16, see Methods and materials for expression and purification protocols). The protein material was further purified on Source-Q ion-exchange column and the resulting highly purified material was estimated to be at least 95% pure by SDS-PAGE and Commassie stain (Figure 17).

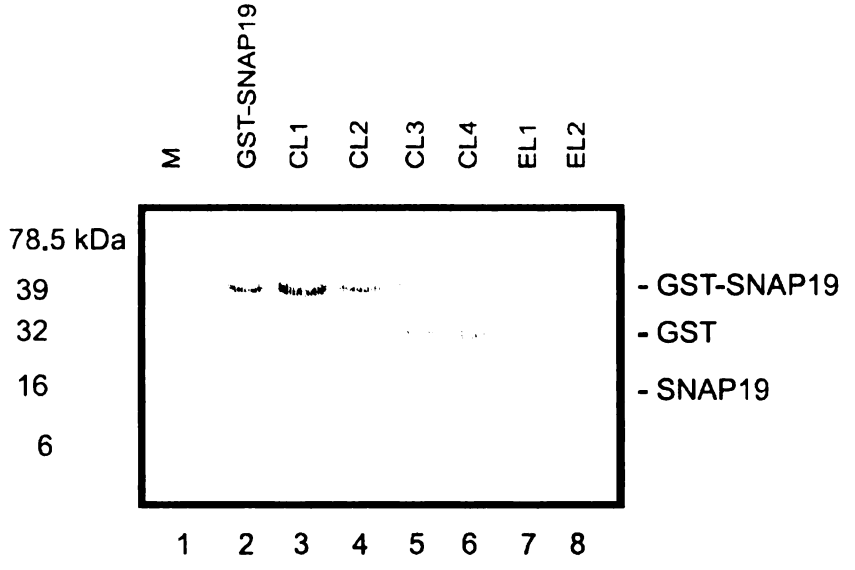


Figure 16: SNAP19 purified over GSH resin. Protein and beads suspension after 10 min (CL1), 20 min (CL2), 2 h (CL3) and 6 h (CL4) thrombin cleavage. Cleaved SNAP19 (EL1 and EL2, two successive fractions of the eluted material).

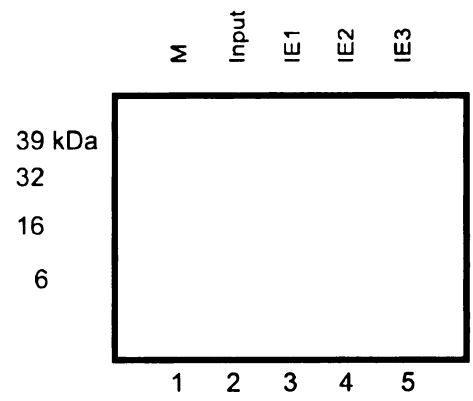


Figure 17: Ion-exchange purification of SNAP19. Consecutive fractions containing highly purified SNAP19 (IE1, IE2 and IE3). 15 mg of SNAP19 purified by affinity chromatography were diluted 3 fold, to bring the salt concentration below 50 mM and loaded onto Source-Q ion-exchange resin (15 mL). The protein material was eluted from the column by a salt gradient (0-70%, 1 M KCl). 8-10 mg of highly purified SNAP19 were recovered in two fractions (IE2, IE3).

The high expression levels and complete recovery of SNAP19 enabled crystallization studies, and a crystal form was identified in a PEG containing condition. Crystals were of extremely poor quality and optimization of crystallization did not improve the crystal quality or morphology. The crystals obtained had an appearance of a quasi-crystal, multi crystal cluster. Seeding experiments resulted in reduced time needed for crystal formation and growth, but the morphology remained poor. Using X-ray radiation, no diffraction pattern was observed for these crystals (data not shown).

2.1.1.2. Cloning, over-expression, purification and crystallization of SNAP19(1-85)

Since SNAP19 contains a glutamic acid rich region on its C-terminus with 10 successive glutamic acid residues, we hypothesized that this region might be responsible for the low crystal quality due to this glutamic acid stretch, which is likely to be unstructured. The C-terminal domain was removed by quick change PCR and the resulting truncated version of SNAP19 was over-expressed and purified under identical conditions to the wild-type (see Methods and materials) with similar recovery levels (Figure 18).

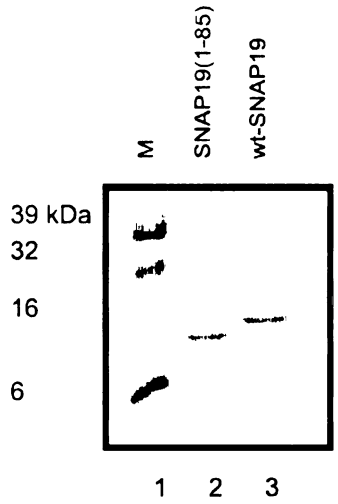


Figure 18: Commassie blue stained SDS-PAGE of GST purified SNAP19(1-85) (lane 2) and wild type SNAP19 (lane 3).

Crystallization of SNAP19(1-85) in sparse matrix screens (CS1, CS2, Mystic72, Peg/pH) yielded crystals similar in appearance to the wild-type SNAP19. Further optimization, by varying crystallization buffer ingredients and concentrations, of the

crystallization conditions including additive screening failed to improve the quality of crystals (data not shown).

2.1.1.3. Nuclear magnetic resonance analysis of SNAP19

Crystallization of this protein was proven to be problematic as no crystals of diffraction quality were obtained. SNAP19 with the molecular weight of 11 kDa is sufficiently small to be a suitable candidate for NMR analysis. The high solubility and high expression levels make SNAP19 an even better candidate for such studies. Preliminary studies on this protein using dynamic light scattering (DLS) methods suggested the protein to be a tetramer and therefore at the limit of practical structural analysis by NMR (data not shown). In order to test whether the NMR spectra of SNAP19 would be of sufficient quality ^{1}H spectra of this protein were collected. The width of the peaks was determined to be sufficiently narrow to allow for NMR structure determination (Figure 19). Next, the cells expressing GST-SNAP19 were grown in M9 minimal media supplemented with ^{15}N ammonium chloride, to facilitate the expression of ^{15}N labeled protein. Although the expression levels of the protein in M9 media were severely reduced, compared to the growth in LB medium, enough purified ^{15}N labeled material was obtained for NMR analysis.

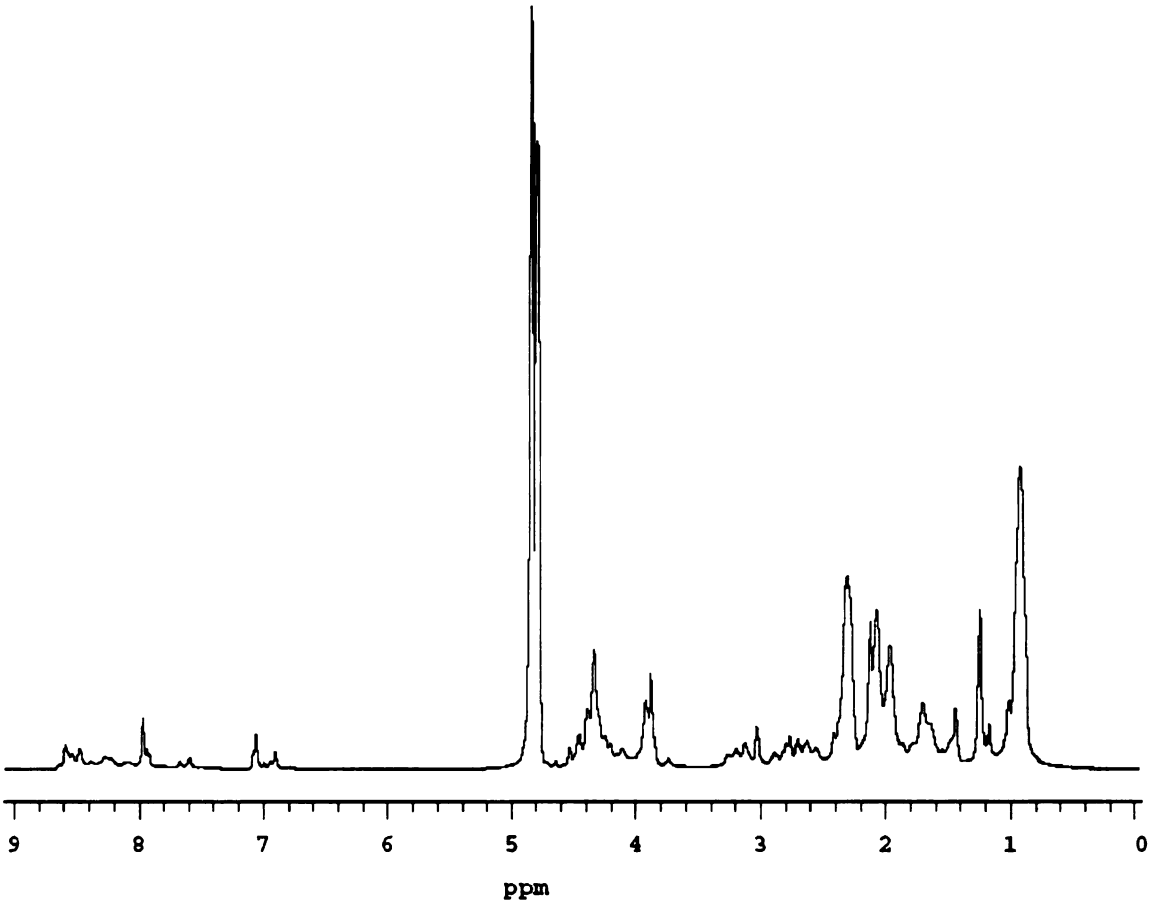


Figure 19: ^{1}H NMR spectra of SNAP19.

^{15}N Labeled SNAP19 was subjected to a ^{15}N-detected ^{1}H-^{15}N correlation experiment; HSQC (heteronuclear single quantum coherence transfer) to assess for the viability of the structural analysis of this protein using NMR techniques, since ^{15}N ammonium chloride is relatively inexpensive compared to ^{13}C labeled carbon sources. Although the peaks were relatively sharp, the dispersity of the amide proton (8-8.5 ppm) peaks was low, suggesting that the protein is not very well folded. Folded proteins have a high level of dispersity (6.5-10 ppm) of the amide protons due to different environments of the proton. Unfolded proteins or peptides have low dispersity (8-8.5 ppm) due to similar environments of the amide protons. According to these results it seems that

SNAP19 has very little tertiary structure and structure determination of such a protein is at best complicated due to poor separation of the signals (Figure 20).

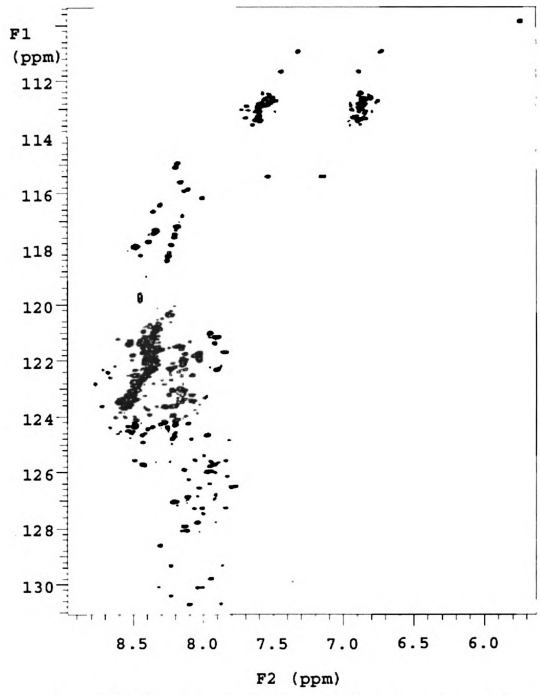


Figure 20: HSQC (heteronuclear single quantum coherence transfer) spectra of SNAP19.

We hypothesized that SNAP190 is required for efficient folding of the short SNAP19 and that SNAP19 should, when mixed with the partnering region of SNAP190, assume a more ordered tertiary structure. To test this, an excess of SNAP190 (84-133) peptide was mixed with SNAP19. SNAP19, the SNAP190 peptide, and a mixture of the two were all analyzed by NMR. Efficient binding and presumed folding of SNAP19 should be evident from the dispersity of the amide proton peaks in the ^{1}H spectra. Unfortunately, the spectra obtained for the mixture had similar dispersion of peaks to those of SNAP19 and SNAP190 peptide alone, showing that SNAP19 does not have much ternary structure even in the presence of SNAP190(84-133), based solely on the dispersion of the proton signal (Figure 21).

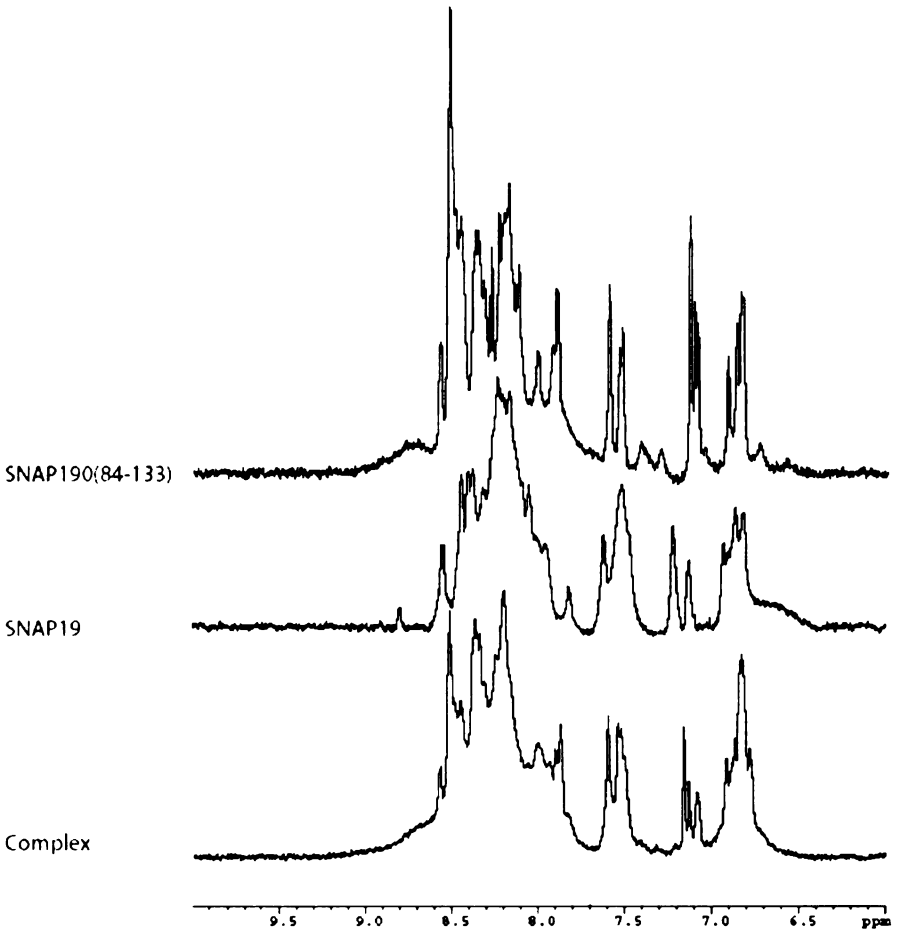


Figure 21: ^{1}H NMR spectra of SNAP190(84-133) peptide, SNAP19 and the complex obtained by mixing the two proteins.

2.1.2. SNAP190(1-505)

SNAP190(1-505) was a second candidate for structural analysis from the SNAPc family of proteins. The protein was expressed in sufficient quantity and efficiently purified to high homogeneity. Crystallization trials were performed and a few crystallization conditions were identified.

2.1.2.1. Over-expression and purification of SNAP190(1-505)

The pGST190(1-505) plasmid was transformed in to *E. coli* BL21(DE3) Codon Plus RIL cells and expressed at 16°C for 12-16 h. The protein was purified using the GST purification protocol and the tag was removed by thrombin cleavage (Figure 22). The protein was further purified using ion-exchange chromatography (Source Q). Approximately 1.5-2 mg of protein was recovered in the last purification step per 1 L culture. Greater than 95% purity was achieved as determined by SDS-PAGE and Commassie staining.

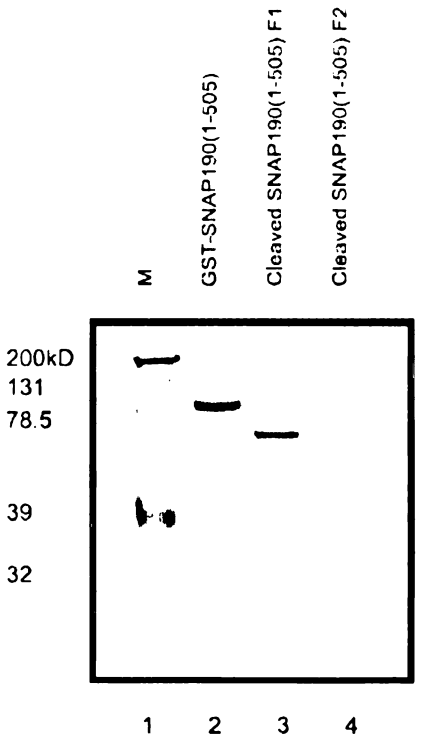


Figure 22: SNAP190(1-505) purification. GST-SNAP190(1-505) immobilized on the GSH agarose (lane 2) and cleaved and eluted material (lanes 3 and 4)

2.1.2.2. Crystallization of SNAP190(1-505)

The protein was buffer exchanged in to a buffer (10 mM Hepes pH 7.5, 100 mM KCl, 3 mM DTT and 10% glycerol), concentrated to 4-5 mg/mL and sparse matrix crystallization solutions were used to find a crystallization condition (30% 2-propanol, 0.1 M Hepes, 0.2 M $MgCl_2$, pH 7.5) that resulted in formation of crystals with needle cluster appearance (data not shown). The crystals passed the poke test and were labeled as protein crystals. However, extensive screening of the condition did not result in an improved crystal form.

2.1.2.3. Crystallization of SNAP190(1-505)/SNAP19 complex

SNAP190(1-505) was mixed with purified SNAP19 in a 1 to 1.2 molar ratio, buffer exchanged in to a buffer (10 mM Hepes pH 7.5, 100 mM KCl, 3 mM DTT and 10% glycerol) and concentrated to 3 mg/mL. Sparse matrix screening (CS1, CS2, Mystic72, PEG/pH) was performed in search of a crystallization condition and a few conditions yielding potential protein crystals were identified (30% 2-propanol, 0.1 M Hepes, 0.2 M $MgCl_2$, pH 7.5 and 30% PEG4000, 0.1 M Tris HCl, 0.2M $MgCl_2$ pH 8.5). The crystals obtained had the appearance of needle clusters and were highly fragmented. Extensive optimization screening failed to find a condition with improved crystal size and morphology (data not shown).

2.2. Over-expression and purification of SNAPc subunits

To facilitate crystallization and X-ray structural studies, individual SNAPc subunits (SNAP190(1-505), SNAP19, SNAP43 and SNAP50) were over-expressed as GST fusion peptides in *E. coli* and purified on an affinity column using glutathione resin. Although the expression of the proteins was satisfactory, the recovery of the protein material from the affinity beads was often troublesome at best. Most SNAPc subunits were inherently prone to aggregation, over-digestion by thrombin and precipitation. SNAP19 was the only subunit that was straightforwardly recovered from the affinity resin and further purified (Figure 24). Full length SNAP190 can not be expressed in *E. coli*, but the truncated version SNAP190(1-505) can. After successful immobilization of this protein on the glutathione affinity resin, great care must be taken when thrombin is used to cleave the protein material off the beads. SNAP190(1-505) is extremely prone to thrombin over-digestion, but was successfully recovered from the beads and further purified using ion-exchange chromatography (Figure 23). In *E. coli* SNAP43 was over-expressed in large quantities and immobilized on the GSH resin. The tag was successfully removed by thrombin cleavage. Only minuscule amounts of this protein were recovered from the beads and the recovered protein precipitated and adhered to purification beads or reaction vessels (Figure 23). SNAP50 was also expressed in large quantities and successfully immobilized to the resin. SDS-PAGE analysis of bound material revealed a strong presence of another protein that appears to be about 60 kDa in size (Figure 24). Removal of the tag and elution of the SNAP50 resulted in non-homogenous material with the strong presence of the 60kDa protein. Further purification methods were unsuccessful in removal of this contamination. Great care had to be taken when the tag was being

removed with thrombin cleavage as SNAP50 is also prone to ever-digestion and the apparent single sized GST-fusion protein often resulted in several products smaller in size compared to SNAP50, which were probably degradation products of SNAP50.

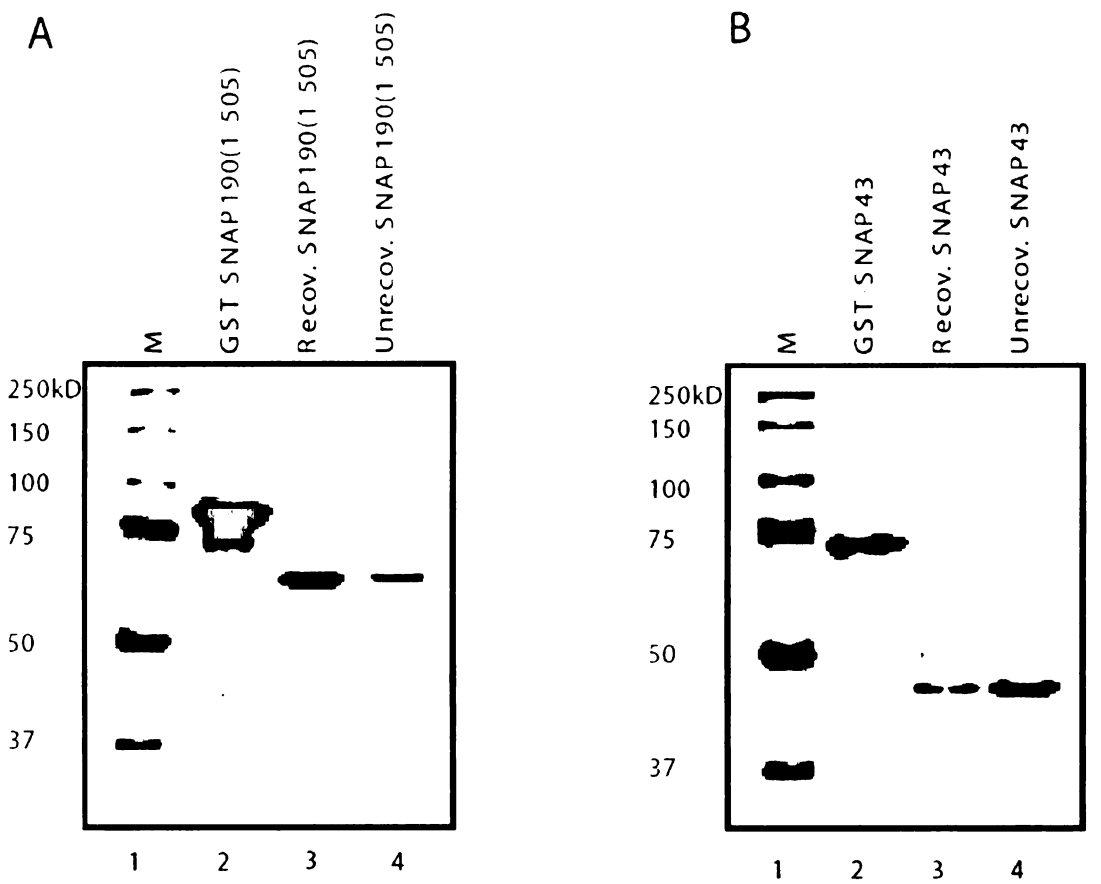


Figure 23: Purification of (A)SNAP190(1-505) and (B)SNAP43 using GST affinity chromatography. The proteins were purified from equal culture volumes under identical conditions. The majority of SNAP43 can not be recovered and the material that eluted off the beads sticks to the reaction vessel in a short time. The material can be recovered by boiling in the sample buffer for SDS-PAGE.

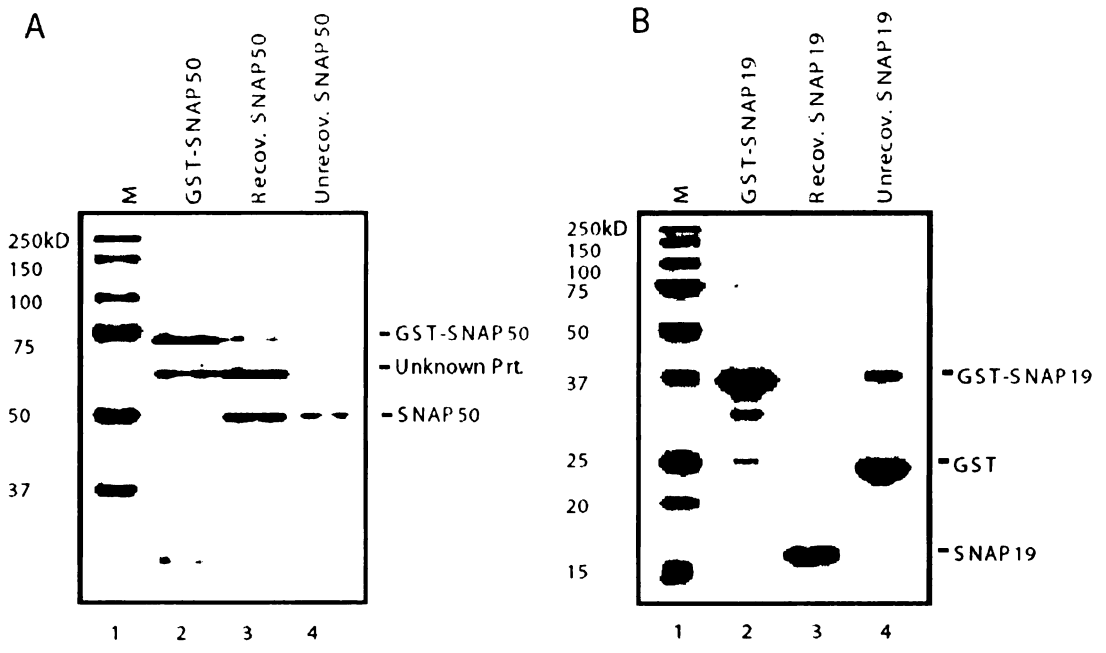


Figure 24: Purification of (A) SNAP50 showing the strong presence of an unidentified protein that can not be removed by subsequent purification and (B) SNAP19 using GST affinity chromatography. The proteins were purified from equal culture volumes under identical conditions.

It is apparent that the isolated subunits share a few common characteristics. They are prone to thrombin degradation, which is most likely due to structurally unstable regions like exposed loops. Secondly, some subunits like SNAP43 and SNAP50 and to some extent SNAP190(1-505), were hard to recover after thrombin cleavage. SNAP43 is virtually impossible to recover from the glutathione resin and different approaches were tested to cope with this problem, but were unsuccessful. Only small amounts of SNAP43 can be successfully eluted from the beads and the quantity of the protein is only sufficient

for biological, but not for structural studies and the majority of the cleaved material remained immobilized on the resin or adhered to the reaction vessel.

The mini SNAP complex, made from SNAP190(1-505), SNAP50 and SNAP43 (mSNAPc) can be assembled from individually expressed and purified subunits. Although the production of such material is at best inefficient, the amounts obtained and the quality of the material were sufficient to perform DNA binding studies using U1 and U6 DNA probes. The complex was capable of binding to the PSE region of the U1 or U6 DNA, as can be judged by the retardation of the mobility of radiolabeled DNA as was the endogenous SNAPc isolated from HeLa cells. DNA binding was specific since it was abolished when mutant PSE was used (Figure 25)[58].

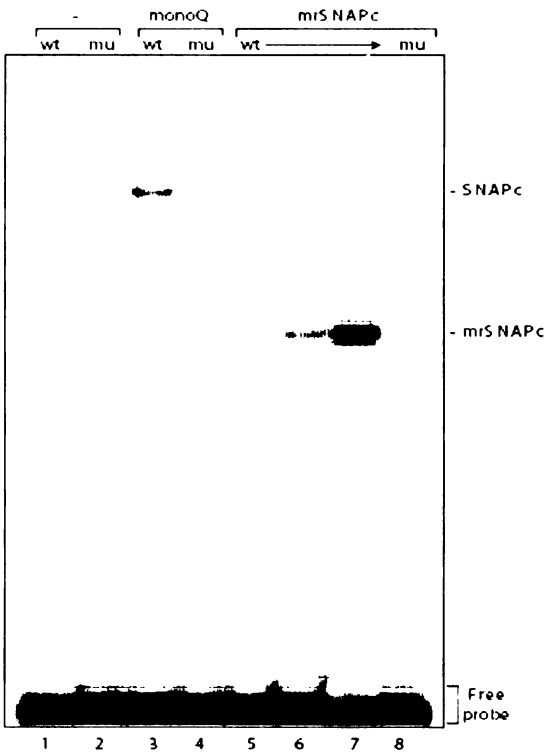


Figure 25: Electro mobility shift assay (EMSA) of the recombinant mSNAPc (mrSNAPc) assembled from individually expressed and purified subunits, and

endogenous SNAPc obtained from *HeLa* cells (monoQ fractions). Increasing amounts of mSNAPc that was assembled from subunits individually expressed in *E. coli* (approximately 1, 2.5, and 7.5 ng; lanes 5–7, respectively) were added to EMSA reactions containing dsDNA probes containing a wild-type (wt) or mutant (mu) PSE, as indicated. Reactions loaded in lanes 3 and 4 contained 7.5 μL of partially purified endogenous SNAPc (approximately 0.3 ng SNAPc/mL). Reactions containing only the probe DNAs are shown in lanes 1 and 2 (experiment performed by Craig Hinkley) [58].

This complex was also tested for the ability to restore transcription from SNAPc depleted nuclear extract. Surprisingly, the complex was unable to restore transcription significantly for either U1 or U6 types of promoters (Figure 26). Endogenous SNAPc isolated from *HeLa* cells (monoQ fractions) was in contrast capable of both DNA binding and transcription initiation. The amounts of the material obtained from this expression system are significantly lower than the amounts that can be obtained form bacterial expression[58].

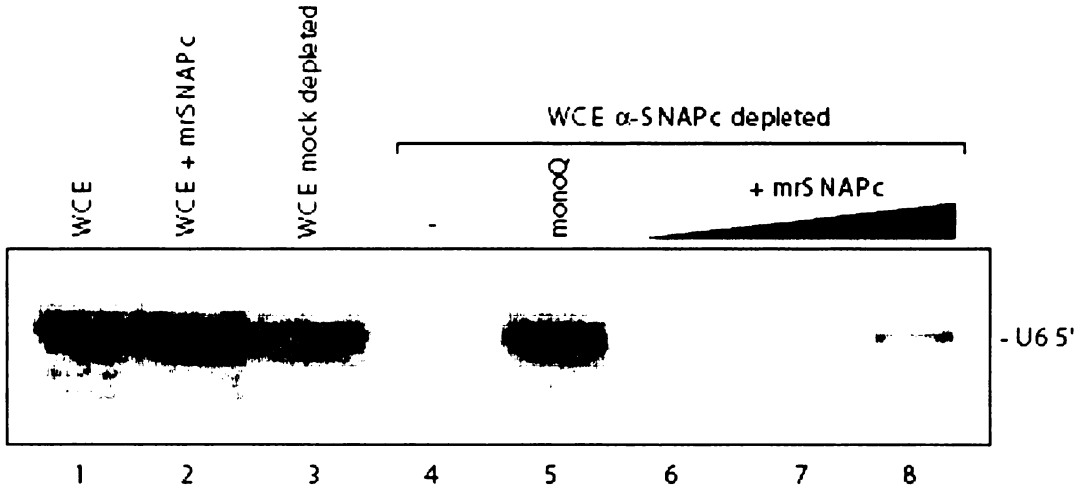


Figure 26: In-vitro U6 transcription initiation assay of the recombinant mSNAPc. Increasing amounts of mSNAPc were added to human U6 in vitro transcription reactions for which the HeLa whole cell extract (WCE) was treated with α-SNAP43 antisera to remove endogenous SNAPc, as shown in lanes 6–8. Lane 4 shows the decreased signal for the correctly initiated U6 transcription upon removal of endogenous SNAPc. Lane 5 shows the U6 signal dependent upon addition of endogenous SNAPc obtained from biochemical fractionation of a *HeLa* cell nuclear extract. Note, this signal is comparable to the mock depleted WCE (lane 3) or WCE alone (lane 1)[58].

We hypothesized that the SNAPc components are unable to fold properly when expressed separately and have exposed surfaces that render the material unstable and sticky. Also, since co-expression using the baculo virus expression system yielded fully active mSNAPc, we thought that co-expressing the subunits in *E. coli* could potentially

yield quantities sufficient for biological and structural studies. We also hypothesized that the material obtained from a co-expression system would, unless specific secondary modifications are needed, yield fully active complex similar to the baculovirus system, but with increased expression levels.

2.3. Cloning of plasmids for co-expression

The goal of simultaneously expressing 3 or 4 subunits in the same cell can be achieved by either using 3 or 4 plasmids with different antibiotic resistance, or by using a polycistronic plasmid with several expression cassettes. We decided to use a combination of the two approaches, since the number of compatible plasmids with different antibiotic resistance is limited. Also, the strategy to ligate several genes in to the same plasmid, targeting many different multiple cloning sites (MCS) is complicated. First, a theoretical purification plan was devised that would enable rapid and efficient extraction and purification of the co-expressed complex. Since all subunits were previously cloned in a GST expression plasmid we decided to keep the existing plasmid with SNAP190(1-505) in the cloning site following the GST and thrombin recognition sequences (Figure 27). Other subunits would be placed on different plasmids with different antibiotic resistance. First, SNAP43 was cloned into pCDFDuet-1 to form pCDF43-1 and SNAP50 was cloned into pRSFDuet-1 to form pRSF50-1. Since pCDF43-1 and pRSF50-1 carry the information for streptomycin and kanamycin resistance. respectively, it is possible to transform these two plasmids into the same *E. coli* cell with the pGST-190 plasmid carrying the ampicillin resistance already present. pCDFDuet-1 was also used to insert the SNAP50 gene in to the multiple cloning site I (MCSI) and following that the SNAP43

gene was ligated into the second multiple cloning site (MCSII), resulting in the pCDF43R/50-1 expression plasmid (Figure 28). The SNAP19 gene was cloned in to pRSFDuet-1 to yield pRSF19-1 (Figure 29).

Plasmids were sequentially transformed into *E. coli*, following calcium treatment to make the cells competent (to make mSNAPcγ3 pGST190(1-505) was transformed first, then pCDF43R/50-1, and finally, to make mSNAPcγ4 pRSF19-1 was transformed). After successful insertion of the plasmid in to the bacterial host, the expression was performed at 16°C and the proteins were extracted, purified using the GST affinity column and analyzed (see Methods and materials for details).

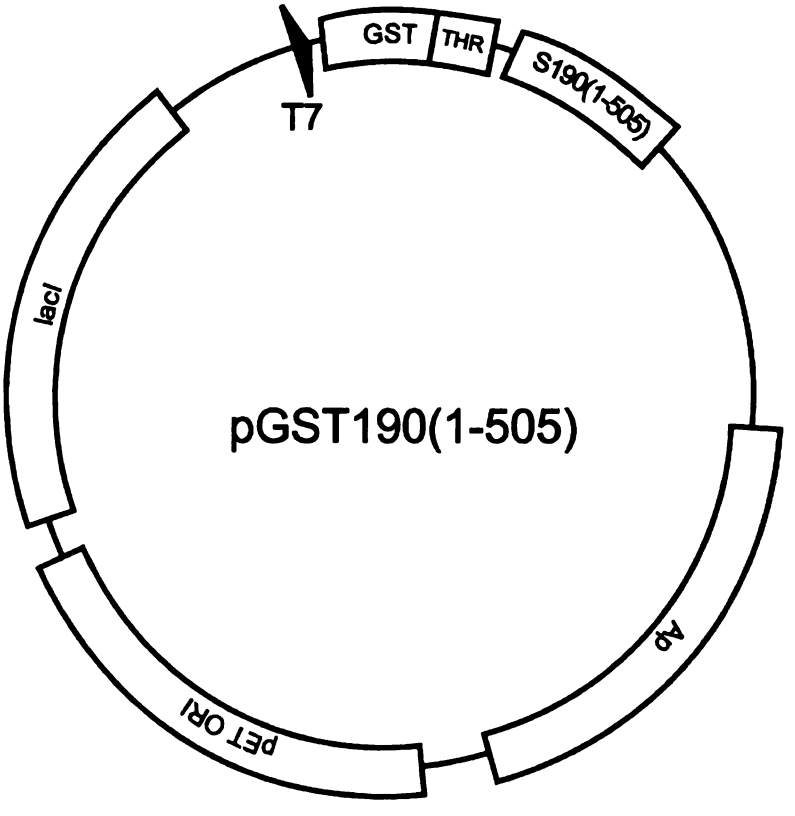


Figure 27: The pGST190(1-505) plasmid with pET origin of replication and ampicillin resistance. The SNAP190(1-505) sequence follows the sequence for the GST tag and the thrombin recognition sequence.

Different combinations of which gene is present in which cloning site and the type of plasmid used were used and tested to check for the maximum efficiency of the co-expression system. And although all subunits were expressed well in almost every case, the different co-expression systems varied in the apparent stoichiometry of the subunits.

Good expression and recovery levels were obtained using the pGST190(1-505) and pCDF43/50-1, with ampicillin and streptomycin resistances, respectively. The complex (SNAP190(1-505), SNAP50, and SNAP43) obtained using this co-expression system is referred as mSNAPcγ3. Better expression and recovery levels were obtained using the pGST190(1-505), pCDF43/50-1 and pRSF19-1 with ampicillin, streptomycin and kanamycin resistances, respectively. The complex (SNAP190(1-505), SNAP50, SNAP43, and SNAP19) obtained using this co-expression system is referred to as mSNAPcγ4.

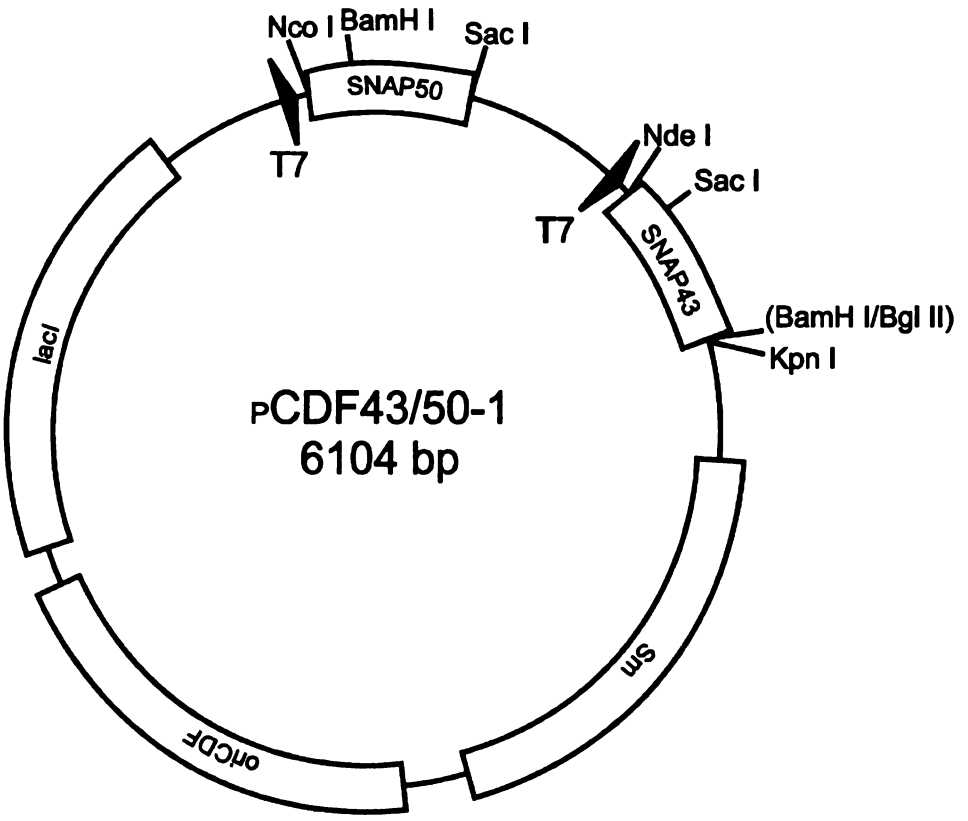


Figure 28: pCDF43/50-1 plasmid with CDF origin of replication and streptomycin resistance. SNAP50 gene is in MCSI and SNAP43 is in MCSII.

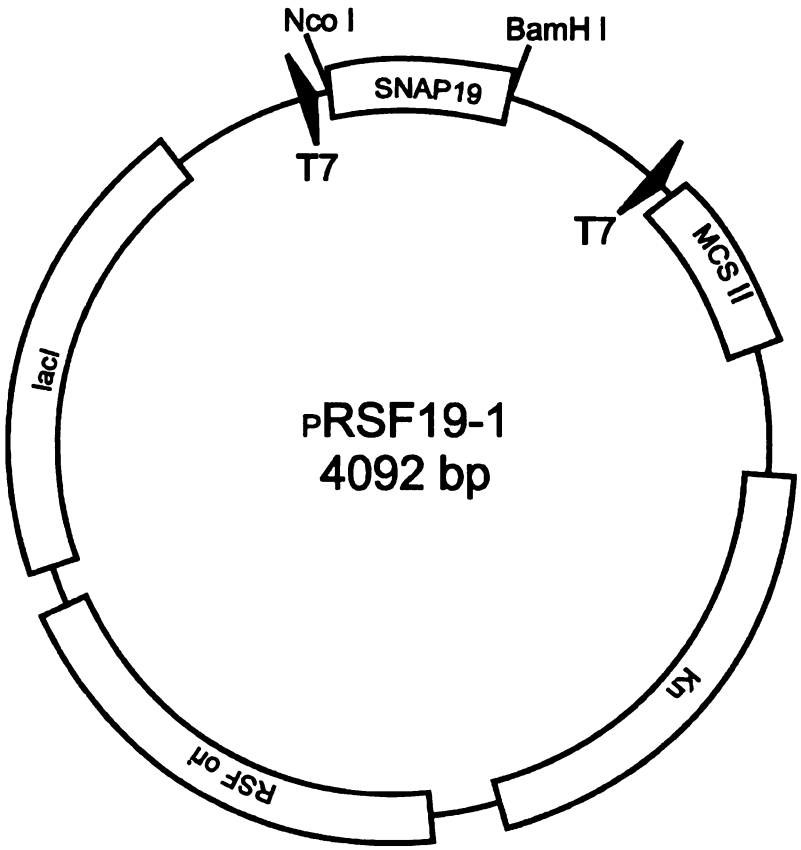


Figure 29: pRSF19-1 plasmid with RSF origin of replication and kanamycin resistance with SNAP19 gene in MCSI and un-utilized MCSII.

2.4. Over-expression of mSNAPc variants

Once the plasmids were sequentially inserted in to the *E. coli* BL21-CodonPlus(DE3)-RIL competent cells, the cultures were grown to the desired density and over-expression was induced at different temperatures to determine the optimal expression conditions. 16°C was determined to be the temperature that yielded the most material. The expression at increased temperatures resulted in a drastic reduction of the target protein material. Both the mSNAPcγ3 and mSNAPcγ4 systems expressed all the desired proteins as was determined by SDS-PAGE and Western blot analysis.

Different strains of *E. coli* were tested in search of the best co-expression host, like BL21(DE3), CodonPlus-RIL, CodonPlus-RP, and pLysS (Stratagene). The BL21(DE3) competent cells (genotype: *E. coli* B F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3)) are an all-purpose strain for high-level protein expression. The BL21(DE3) pLysS competent cells (genotype: *E. coli* B F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3) [pLysS Cam^r]) contain a pLysS plasmid with chloramphenicol resistance and provide tighter control of protein expression of toxic proteins. The BL21-CodonPlus(DE3)-RIL competent cells (genotype: *E. coli* B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tetr *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^r]) contain a plasmid encoding rare codons for arginine, isoleucine and leucine (*argU* (AGA, AGG), *ileY* (AUA) and *leuW* (CUA)) and provides chloramphenicol resistance. The BL21-CodonPlus(DE3)-RP competent cells (genotype: *E. coli* B F^- *ompT hsdS*(r_B^- m_B^-) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte [*argU proL* Cam^r]) contain a plasmid encoding rare codons for arginine and proline (*argU* (AGA, AGG), *proL* (CCC)) and provides chloramphenicol resistance.

mSNAPcγ4 was successfully expressed in all types of *E. coli* with similar amounts of the protein obtained, but with the highest amount of the mSNAPc being expressed in the CodonPlus-RIL cells, as judged by immobilization and recovery of the complex using the glutathione resin. It was observed from the amino-acid sequence and the codon analysis that several subunits, especially SNAP190(1-505) and SNAP50, contain rare codons and could potentially lead to lower expression levels due to codon bias.

Next, different cell culture growth media were tested in order to further optimize the expression levels of partial SNAP complexes. LB and TB media, as well as M9

minimal media were tested by immobilization and recovery of the protein from the glutathione resin. M9 minimal media resulted in hardly detectable levels of the complex, as judged by Commassie blue staining of proteins separated by SDS-PAGE. LB and TB yielded the highest levels of expressed material. The over-expression in TB had a two fold increase compared to the expression levels for the cells grown in LB medium. The density of the cells grown in TB media was about twice that of the cells grown in LB media. Overall, the expression levels per cell are comparable for the two growth medias (data not shown).

The time of IPTG induction was also studied to further improve the expression levels of the mSNAP complexes. Different times of induction, defined by the cell density in the growth culture, were tested in the range of 0.1 to 1.4 optical density (OD_{600}). After the induction the cell culture was grown for 12-16 h at 16°C and the amount of the expressed material was evaluated using the GST purification scheme. It was determined that the expression levels of the protein are in direct proportion to the cell density at the time of induction. When the cell culture was induced with IPTG at an OD_{600} of 1.4, almost double amount of the mSNAPc was obtained as compared to the cell culture induced at an OD_{600} of 0.7. The amount of cells obtained, when the culture was induced at an OD_{600} of 1.4 was also doubled (data not shown). It seems that the mSNAPc components are non-toxic to *E. coli* and the expression level per cell is constant when grown at identical conditions and independent of the induction point. The only consequence of the different induction point is the final cell density and therefore the final amount of the produced material and not the actual amount of protein per cell as is usually the case when the proteins are even slightly toxic to the cell.

The temperature of induction was studied. And the complex over-expressed only at 16°C. Expression at 24°C or 38 °C resulted in diminished expression levels of mSNAPc.

2.5. Comparison of the individual expression of SNAPc subunits to the co-expression system

Cell cultures expressing the mSNAPcγ4 or individual subunits were grown under identical conditions and the expressed material was analyzed by Western blotting method using the antibody directed against the specific subunit of interest. The SNAP50 and SNAP43 subunits expressed well in both the individual and co-expression systems. Interestingly the amount of leaky expression for these two proteins is relatively high in the co-expression system. The individual expression systems yielded no expressed protein before induction (Figure 30, panels A and B). The level of the protein in the insoluble or soluble fractions is similar for the two systems. The relatively high amount of the material in the insoluble fraction is most likely due to insufficient rinsing of the material and may not be a measure of the total material actually in the insoluble form. The amount of the target proteins is also comparable between the two expression systems, suggesting that expression was not drastically improved. Nevertheless, the expression levels for both SNAP43 and SNAP50 are elevated in the co-expression system compared to the individual expression.

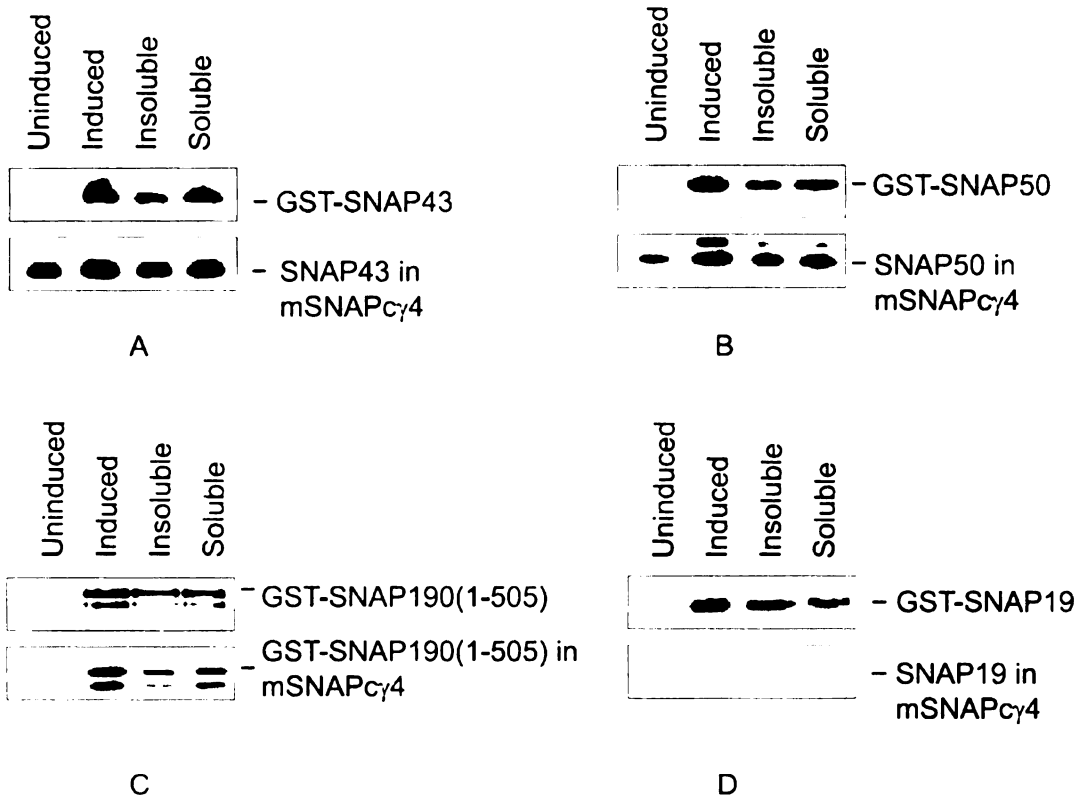


Figure 30: Analysis of expression for SNAPc subunits in the individual expression and co-expression approach by Western blot. GST labeled protein were visualized using α-GST antibodies and the SNAPc subunits were visualized using antibodies directed against the subunit of interest.

The expression levels of GST-SNAP190(1-505) are to some extent elevated in the co-expression system compared to individual expression. Leaky expression was not observed for this protein and the material is mostly in the soluble fraction of the cell extract (Figure 30, panel C). In contrast the amounts of SNAP19 in the co-expression system were not detectable in the cell extract using Western blot. The levels of this protein are significantly reduced compared to the individual expression system (Figure 30). SNAP19 is the smallest of the subunits and to achieve complex formation where the

subunits are equimolar, a smaller amount of SNAP19 is needed compared to the other subunits, so a complete mSNAPcγ4 complex could still be assembled.

2.6. Purification of mSNAPc

Since the SNAP190(1-505) contains a cleavable linker and a GST tag, the soluble culture extract harboring mSNAPCγ3 or mSNAPcγ4 was clarified and allowed to bind to the glutathione resin to immobilize the expressed complex. Other subunits were not tagged, yet they were efficiently immobilized thru the GST-190(1-505). Different purification conditions were tested to explore different approaches to achieve highest purity of the material in the first step of purification. Parameters like salt, detergent and glycerol concentration were varied and the purity of the immobilized material was evaluated with SDS-PAGE and silver staining. It was determined that the detergent (Tween-20) is absolutely critical for the extraction of the complex from *E. coli* since virtually no material was immobilized when Tween-20 was not present as judged by SDS-PAGE and Commassie blue staining. The Tween-20 could be replaced with n-octyl-beta-D-glucopyranoside (BOG) with slightly reduced extraction levels in this purification step. The salt concentration was varied from 150 to 500 mM with the highest purity achieved when higher salt is used. Lower salt concentration leads to increased presence of impurities. The bound material can be efficiently removed from the beads by glutathione elution or by thrombin cleavage of the linker connecting the SNAP190(1-505) and the GST tag. In comparison to the GST purification step of some SNAPc subunits, like SNAP43 or SNAP50, the material was almost fully recovered from the resin (Figure 31). Furthermore, the relative intensities of the bands observed on the

Coomassie stained SDS-PAGE suggest that the subunits are equimolar, although in some cases the intensity of SNAP50 was weaker than the intensities of SNAP190(1-505) or SNAP43. The co-expressed material is also more stable during thrombin cleavage as relatively high concentration of thrombin can be used for extended periods of time without ill effects on the complex composition. When the optimal cleavage conditions were investigated, no truncated products were observed even after three-day cleavage with 200 units of thrombin at 4°C (6 L cell culture). On the other hand, the individually expressed subunits were extremely sensitive during thrombin cleavage and prone to over digestion. More importantly, the co-expressed material was successfully recovered from the resin, which was not the case for SNAP43 where relatively small amounts of the protein were recovered (Figure 32).

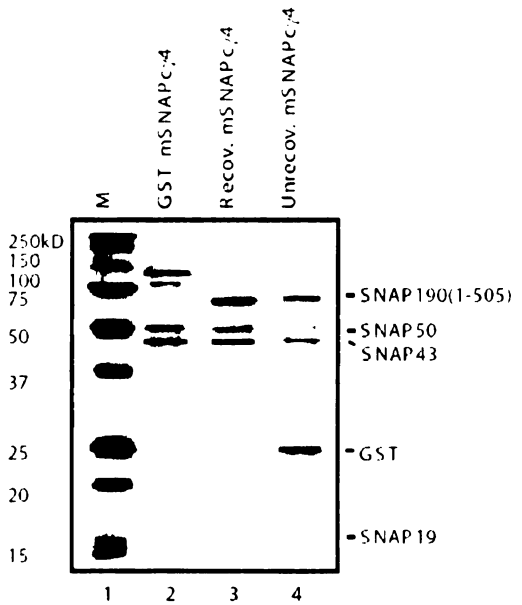


Figure 31: Purification of co-expressed mSNAPcγ4 using the GST affinity resin and thrombin cleavage. The mSNAPcγ4 was purified under the same conditions as the individual subunits.

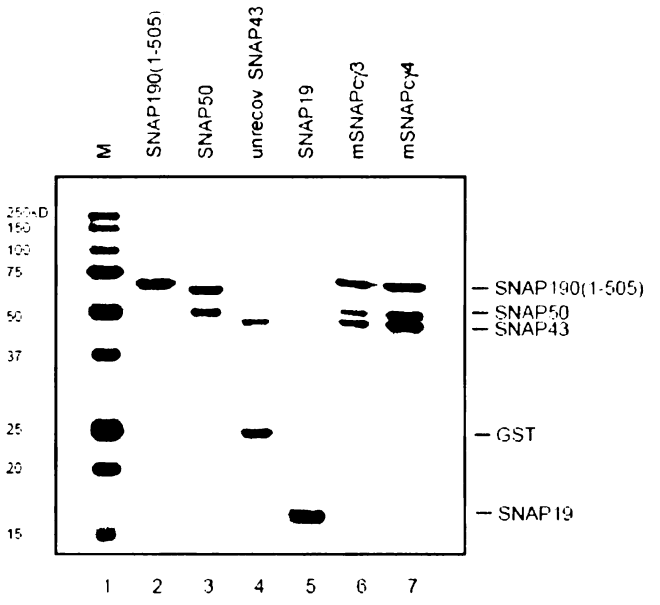


Figure 32: Recovery of the co-expressed material compared to individually expressed and purified subunits. SNAP19 improves complex stability and subunit recovery. Subunit recovery for partial SNAPC assembled without and with co-expressed SNAP19 (mSNAPcγ3 and mSNAPcγ4, respectively) was estimated by SDS–PAGE and Coomassie blue staining. Note the increased recovery of SNAP43 and SNAP50 in mSNAPcγ4 (lane 7) compared to mSNAPcγ3 (lane 6). Lanes 2–5 contain aliquots of the individually expressed SNAPc subunits for reference. Note that the unrecovered SNAP43 is used to show the migration of SNAP43.

The recovered material was further purified using ion-exchange chromatography. Cation and anion (SourceQ and SourceS, Pharmacia) exchange resins were tested in a wide array of purification conditions. The complex had the tendency to bind to the ion-exchange resin, but the recovery of the protein material was often accomplished only with severe losses (data not shown). Optimal purification conditions were eventually

determined and the detergent Tween-20 was again essential for this purification. Surprisingly, addition of magnesium ions to the purification buffer resulted in even better recovery rates. The protein does, however, elute off the column over a broad range of salt concentrations and the purity is not greatly improved compared to the GST purification step alone.

The co-expression and GST purification scheme using the GSH resin works well with both mSNAPcγ3 and the mSNAPcγ3 and is a major improvement compared to individually expressed and purified subunits or the baculo-virus co-expression system, which resulted in low amounts of material or material of poor purity.

2.7. DNA binding activity of mSNAPc variants

The mSNAPcγ4 was tested for DNA binding activity for both U1 (Figure 33 lane 11) and U6 (Figure 33 lanes 2, 3) promoters in an electrophoretic mobility shift assay (EMSA) and was similar to the assembled recombinant mSNAPc active for DNA binding to both types of promoters[58].

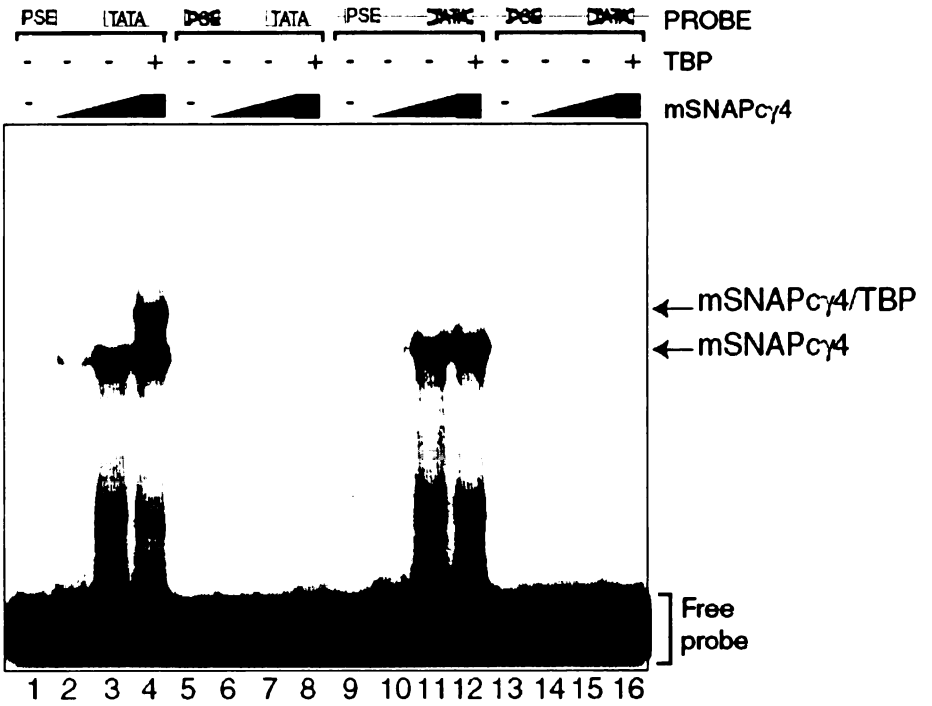


Figure 33: mSNAPcγ4 facilitates TBP promoter recruitment. Increasing amounts of mSNAPcγ4 (3 and 10 ng) were added to DNA binding reactions containing radiolabeled probes that harbor a wt PSE and wt TATA box (lanes 1–4), mu PSE and wt TATA box (lanes 5–8), wt PSE and mu TATA box (lanes 9–12), or mu PSE and mu TATA box (lanes 13–16). Lanes 4, 8, 12, and 16 contain 50 ng of recombinant full-length human TBP in addition to 10 ng of mSNAPcγ4. Reactions containing only the DNA probes are shown in lanes 1, 5, 9, and 13. Positions of the mSNAPcγ4 and mSNAPcγ4 plus TBP complexes are indicated (experiment was performed by Gauri W. Jawdekar)[58].

The complex was also tested for the ability to recruit TBP to the promoter and was capable of doing so in the case of the U6 promoter (Figure 33 lane 4) but was unable to recruit TBP in the absence of the TATA box, as is the case in a U1 type of snRNA

promoter (Figure 33 lane 12). TBP alone was not capable or binding to the DNA in the absence of the SNAP complex, under the experimental conditions, regardless of TATA box presence. mSNAPcγ4 binds to the promoter region of the DNA specifically and mutations in the PSE region abolish the interaction[58].

2.8. mSNAPcγ4 is capable of recruiting TFIIIB to the promoter

The co-expressed mSNAPcγ4 was tested for its ability to recruit Brf2-TFIIIB to the U6 promoter. Human TBP, Brf2 and Bdp1(1-470) were mixed with mSNAPc and the labeled U6 DNA probe. In the absence of mSNAPc, Brf2-TFIIIB did not bind to the DNA (Figure 34, lanes 2-7) or did so weakly (Figure 35, lanes 5 and 7). The Brf2-TFIIIB complex bound the probe in the presence of mSNAPc. Migration of mSNAPc was altered in the presence of TBP (Figure 34, lane 9), but not by addition of Brf2 or Bdp1 alone (Figure 34, lanes 10 and 11), suggesting that TBP is essential for subsequent recruitment of Brf2 and Bdp1 and that mSNAPc through direct protein-protein interactions stabilizes the TBP/DNA interaction. A supershift was also observed when Brf2 was added to the mSNAPc/TBP/DNA complex, but not when Bdp1 was added in the absence of Brf2 (Figure 34, lanes 12 and 13). The presence of both Brf2 and Bdp1(1-470) resulted in a band migrating slower than mSNAPc/TBP/DNA and mSNAPc/TBP/Brf2/DNA (Figure 34, lane 15, Figure 35, lane 12). Specific protein-protein contacts are involved in the formation of the mSNAPc/Brf2-TFIIIB/DNA complex and the complex is assembled in the order mSNAPc-DNA-TBP-Brf2-Bdp1[59].

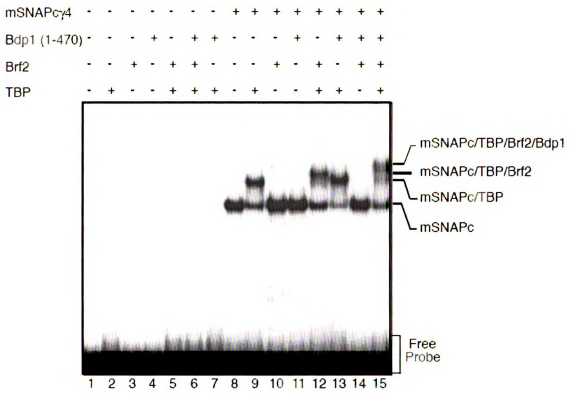


Figure 34: mSNAPcγ4 is capable of recruiting TFIIIB to the promoter. Coordinated DNA binding by SNAPc and TBP facilitates higher order complex assembly with Brf2 and Bdp1. DNA binding reactions were performed with the indicated combinations of SNAPCγ4 and Brf2-TFIIIB subunits. These results suggest that preinitiation complex assembly follows the order SNAPc>TBP>Brf2>Bdp1 (experiment was performed by Gauri W. Jawdekar)[59].

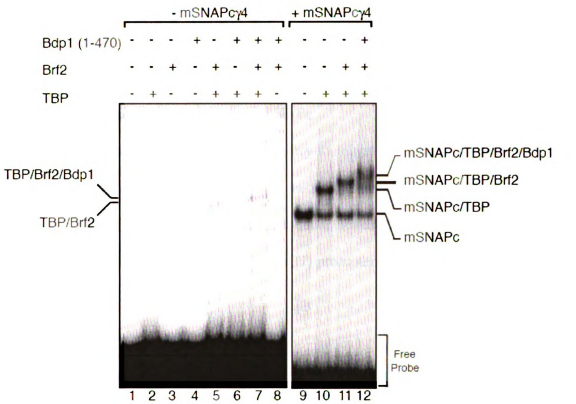


Figure 35: SNAPc stimulates DNA binding by Brf2-TFIIIB. Electrophoretic mobility shift assays were performed using a U6 probe containing a wild-type mouse U6 PSE and a wild-type TATA-box (AC probe). DNA binding was carried out in the absence (lanes 1-8) or presence (lanes 9-12) of wild type SNAPCγ4. Reactions containing individual TBP, Brf2, and Bdp1 (1-470) subunits are shown in lanes 2-4. Reactions containing pair wise combinations of TBP with Brf2, TBP with Bdp1 (1-470), and Brf2 with Bdp1 (1-470) are shown in lanes 4, 5, and 8. DNA binding by the complete Brf2-TFIIIB complex in the absence of SNAPc is shown in lane 7. Additional reactions were performed with SNAPc alone (lane 9) or in combination with Brf2-TFIIIB subunits (lanes 9-12), as indicated. Lane 1 shows migration of the probe alone. The relative positions of the various

SNAPcγ4/Brf2-TFIIIB complexes are shown on the right (experiment was performed by Gauri W. Jawdekar)[59].

2.9. Transcription initiation mediated by mSNAPc

mSNAPcγ4 was then tested for its ability to initiate transcription from both U1 and U6 promoters in an in-vitro transcription initiation assay and was, in contrast to the recombinant mSNAPc (assembled from individual subunits), active for transcription initiation from U1 (Figure 36) and U6 (Figure 37) promoters[58].

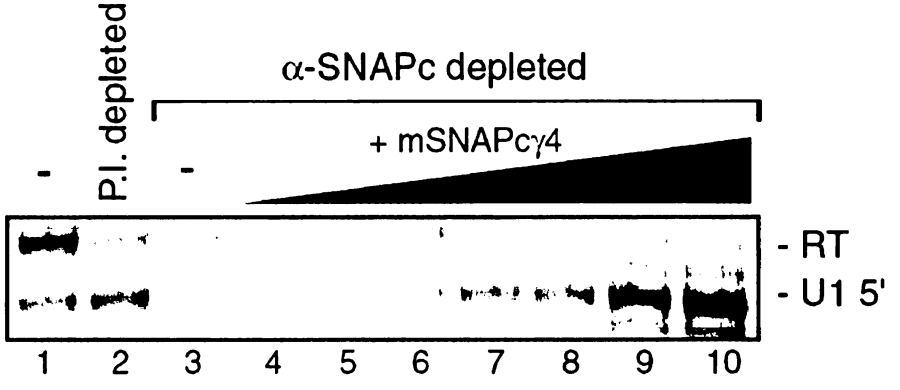


Figure 36: mSNAPcγ4 functions for U1 snRNA transcription by RNA polymerase II. HeLa cell nuclear extract was either mock depleted with a preimmune rabbit sera (lane 2) or α-SNAP43 antisera (lanes 3–10) to deplete endogenous SNAPc. Extracts were then used for human U1 in vitro transcription assays. The U1-specific signal was diminished upon removal of endogenous SNAPc, as shown in lane 3. Increasing amounts of mSNAPcγ4 (0.08, 0.25, 0.75, 2.5, 7.5, 25, and 75 ng) reconstituted the correctly initiated transcription from a human U1 reporter, as shown in lanes 4–10. Lanes 1 and 2 show the U1 signal obtained from either

untreated or mock depleted reactions (experiment was performed by Gauri W. Jawdekar)[58].

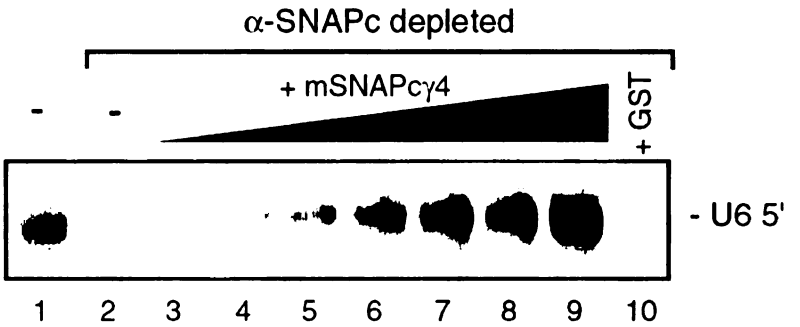


Figure 37: mSNAPcγ4 functions for U6 snRNA transcription by RNA polymerase III. In vitro transcription of human U6 snRNA was carried out using HeLa cell nuclear extract that was treated as in Figure 36. Lane 2 shows the reduced U6 signal upon removal of endogenous SNAPc. Increasing amounts of mSNAPcγ4 (0.08, 0.25, 0.75, 2.5, 7.5, 25, and 75 ng) reconstituted the correctly initiated transcription from a human U6 reporter as shown in lanes 3–9. Approximately 75 ng of GST was added to the transcription reaction shown in lane 10 (experiment was performed by Gauri W. Jawdekar)[58].

2.10. Zinc finger domain in SNAP50

The C-terminal region of SNAP50 is cysteine and histidine rich and is a potential Zn finger domain based on amino-acid sequence, which might be involved with protein-DNA or protein-protein contacts. DNA binding domains sometimes consist of multiple repeats of related zinc finger motifs. The SNAP50 zinc finger domain however contains an unusual arrangement of six histidine and nine cysteine residues that can be grouped

into region I that resembles a TFIIIA-like C2H2 zinc finger and region II that resembles a glucocorticoid-like C2C2 zinc finger.

Sequence alignment of the C-terminal region (301-411) of SNAP50 with the SNAP50 homologues from different mammals, fish, insects, worms, plants, slime mold and parasites, revealed two highly conserved potential zinc finger domains (Figure 38). Interestingly the residues from region I are highly conserved, with the exception of C312, which is conserved in mammals, but not in lower organisms. The fact that SNAP50 is highly conserved in lower organisms suggests that this subunit plays an important role in the transcription initiation mechanism. Region 2 is invariantly conserved in all SNAP50 homologues and the four conserved cysteine residues are similar to steroid receptor ($C_{X2}C_{X22}C_{X2}C$) type zinc fingers. Overall the consensus motif $L_{X4}G_{X6}H_{X3}C_XH_{X20-23}YP_{X11-12}C_{X2}C_{X18}P_{X3-4}C_{X2}CF_{X3}H_{X1-4}G$ is unlike any other family of zinc fingers described[59].

Figure 38: Sequence alignment of human SNAP50 C-terminal amino acids (301-411) with corresponding regions from SNAP50 homologues of other species.

Putative zinc fingers similar to TFIIIA ($H_{X5}H_{X10}H_{X3}C$) and steroid receptors ($C_{X2}C_{X22}C_{X2}C$) are indicated as region 1 and region 2, respectively. The sequence of highly conserved amino acids derived from this alignment corresponds to $L_{X4}G_{X6}H_{X3}C_XH_{X20-23}YP_{X11-12}C_{X2}C_{X18}P_{X3-4}C_{X2}CF_{X3}H_{X1-4}G$. This alignment was performed using the Clustal W program. *Homo sapiens* (NP003075), *Cannis familiaris* (XP853813), *Bos taurus* (AAX08912), *Mus musculus* (NP084225), *Rattus norvrgicus* (NP001013230), *Drosophila melanogaster* (NP724647), *Drosophila pseudaoobscura* (EAL25490), *Trypanosoma brucei* (XP827295), *Arabidopsis thaliana* (AAO30067), *Caenorhabditis elegans*-1 (NP500819), *Caenorhabditis elegans*-2 (NP497807), *Plasmodium falciparum* (), *Danio rerio* (XP694501), *Leishmania major* (XP843572), *Anopheles gambiae* (XP310411), *Dictyostelium discoideum* (XP644064), *Entamoeba histolytica* (XP653151). This comparison was prepared by Gauri W. Jawdekar[59].

2.10.1. ICP-MS and FAAS analysis of mSNAPc variants

SNAP50 contains a potential zinc finger domain. Since SNAP50 is the only subunit of SNAPc containing a Zn finger motif, the SNAP50-containing mSNAPcγ4 and mSNAPcγ3 were used in the metal analysis study to confirm the presence of a zinc atom. The complex was co-expressed in *E. coli* and extensively purified. The material was digested in 2% nitric acid for the inductively coupled plasma mass spectrometry (ICP-MS) or ashed for flame atomic absorption (FAAS) analysis. 1 zinc atom per molecule of the complex was determined with the ICP-MS method, where two isotopes ^{66}Zn and ^{68}Zn were quantified. The FAAS determination of zinc was in good agreement with the ICP-

MS study. Both mSNAPcγ4 and mSNAPcγ3 complexes were determined to contain 1 Zn atom per molecule of the complex, as they yielded 0.90 and 0.89 moles of zinc per mol of the complex respectively (Figure 39, Table 1).

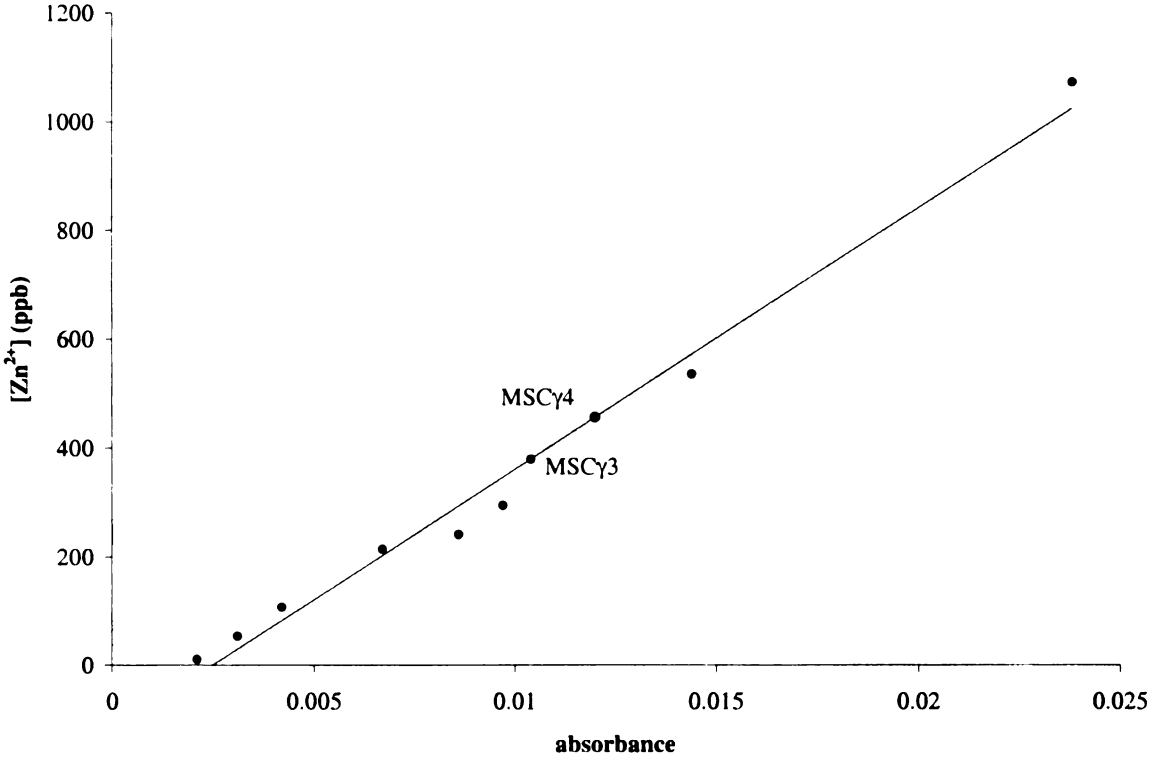


Figure 39: Flame atomic absorption spectroscopic analysis of mSNAPcγ3 and mSNAPcγ4. Unlabeled data points represent the absorption obtained for standard zinc solutions.

Two separately prepared samples were used for ICP-MS analysis and yielded 1.1 and 1.2 equivalents of zinc and are in agreement with previously determined zinc content that utilized the flame atomic absorption. Nickel was also measured during this experiment and the signal for this metal was on the level of the background noise and

insignificant. Nickel is therefore not present in the protein sample. Although two potential metal binding domains were identified in the C-terminal region of SNAP50 by ab-initio modeling, the presence of only one zinc atom was determined experimentally. It is possible that a second metal is still present and required when the complex interacts with the DNA (Table 1).

Table 1: Quantification of zinc and nickel using ICP-MS and FAAS for mSNAPcγ3 and mSNAPcγ4. Two isotopes ^{66}Zn ([a]) and ^{68}Zn ([b]) were measured for ICP-MS. Two independently prepared samples of mSNAPcγ4 were used in the analysis: (1) extensively purified complex used for crystallization and (2) void fraction from gel filtration purification of the complex. The protein concentration was determined by absorbance measurement in 6M urea and by the Bradford assay.

	Sample	[Zn] ppb	[Ni] ppb	[protein] mg/mL	n(Zn)/n(prt)	n(Ni)/n(prt)
ICP-MS	mSNAPcγ4 (1)	[a]1250.4	-	2.62	1.16	-
		[b]1268.5	-	2.62	1.14	-
		-	17.9	2.62	-	0.02
	mSNAPcγ4 (2)	[a]1415.5	-	2.90	1.18	-
		[b]1443.6	-	2.90	1.17	-
		-	35.5	2.90	-	0.03
FAAS	mSNAPcγ3	1519.7	-	3.86	0.89	-
	mSNAPcγ4	1827.3	-	4.95	0.90	-

2.10.2. Ab-initio modeling of SNAP50 C-terminus

Since an experimental structure of SNAP50 is not available, computational methods were employed by Dr. Michael Feig to predict the structure of the C-terminal domain of SNAP50 so that the mechanism of SNAP50 function could be better understood. Because of the unavailability of a structural homolog, an ab-initio approach using the predicted secondary structure and the amino acid sequence was taken. This method is possible because the SNAP50 finger domain region is sufficiently short. Interestingly, one of the structures produces by this ab-initio method had a convincing arrangement of cysteine and histidine residues for coordination of one zinc atom, however some of the conserved residues could, although unlikely, also coordinate another metal atom like zinc, nickel or iron[59].

2.10.3. Mutagenesis

In order to examine the function of the zinc finger domain in SNAP50 the cysteine and histidine residues were changed to alanine. The presumption that the zinc finger domain is responsible for the DNA interacting function of SNAP50, was tested by the DNA binding assay. HA-SNAP50 with selected cysteine or histidine single point mutation to alanine was co-expressed with GST-SNAP190(1-505), full length SNAP43 and SNAP19. The recombinant complexes were immobilized to the glutathione resin, washed extensively and released from the beads by the cleavage of the linker with thrombin. Since the presumption was that the mutation would only affect the binding of the complex to the DNA and not the actual formation of the complex, GST pulldown and co-immuno precipitation experiments were performed to confirm the existence of the

mSNAP complex. The complex was purified using the glutathione beads in the first step of the purification. Since only SNAP190(1-505) contains a GST tag, all the other subunits were obtained through this subunit. Next, the partially purified complex, containing SNAP190(1-505), SNAP43, HA-SNAP50 and SNAP19, was treated with anti-SNAP43 antibody cross-linked to agarose beads, thus allowing specific immobilization of the complex through SNAP43 only. The obtained material was tested for the presence of HA-SNAP50 by western blot targeting the HA tag on SNAP50. The presence of HA-SNAP50, proves that different point mutations of the C-terminus of SNAP50 did not change the consistency of the complex (Figure 40)[59].

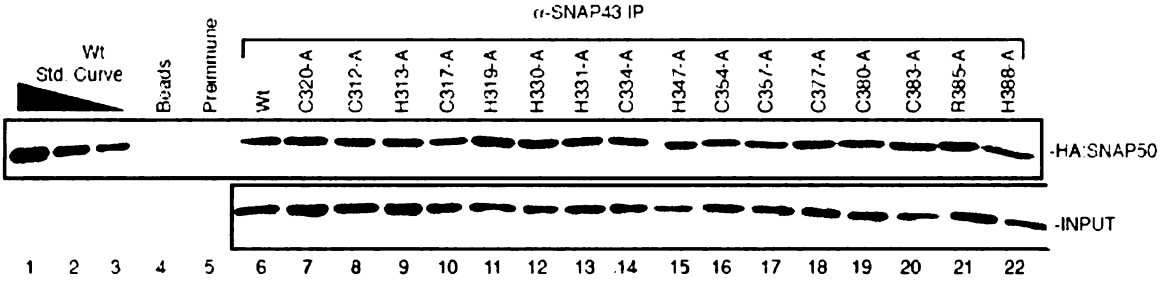


Figure 40: Mutations in the SNAP50 zinc finger domain do not disrupt SNAPc assembly. Approximately 20 ng of each of the SNAPcγ4 complex containing substitution mutations in HA-SNAP50 were affinity purified first using glutathione agarose to pull down GST-SNAP190 (1-505) followed by immunoprecipitation with α-SNAP43 antibodies. Associated wild type or mutant HA-SNAP50 was detected by α-HA Western analysis (lanes 6-22). A titration of wild type SNAPcγ4 using 8, 4 and 2% of the input material is shown in lanes 1-3, respectively. Lanes 4 and 5 contain wild type SNAPcγ4 recovered with the protein-G agarose beads alone or with pre-immune serum,

respectively. The bottom panel represents 4% of the input material directly analyzed by α-HA Western analysis (experiment was performed by Gauri W. Jawdekar)[59].

2.10.4. DNA binding activity of SNAP50 mutants

The mSNAPcγ4 harboring HA-SNAP50 with different point mutations was tested for DNA binding activity. A PSE-containing DNA probe was used for the assay. Point mutations in the evolutionarily conserved $H_{X3}C_XH$ motif, H313-A, C317-A and H319-A resulted in reduced DNA binding as compared to the wild type. Interestingly point mutations C354-A, C357-A, C380-A and C383-A in the conserved $C_{X2}C_{X23}C_{X2}C$ domain exhibited complete abolition of the DNA binding activity of the complex. Furthermore the mutation of the last evolutionarily conserved residue in the C-terminal area of SNAP50 the H388-A also resulted in reduced DNA binding activity (Figure 41)[59].

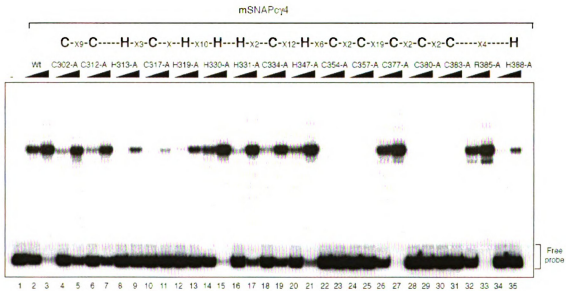


Figure 41: Mutations in the SNAP50 zinc finger domain disrupt DNA binding by SNAPc. Increasing amounts (3 and 10 ng) of SNAPcγ4 with wild type (lanes 2 and 3) or

mutant HA-SNAP50 (lanes 4-35) containing the indicated substitution mutations was tested in an EMSA for binding to a radiolabeled DNA probe containing a high affinity PSE and TATA box (AC probe). Lane 1 shows the probe alone with no added proteins. Mutations H313-A, C317-A, and H319-A resulted in reduced DNA binding, whereas mutations C354-A, C357-A, C380-A, and C383-A completely abolished DNA binding activity. Mutation H388-A also exhibits weakened DNA binding activity (experiment was performed by Gauri W. Jawdekar)[59].

2.10.5. Transcription initiation assay

Next, the wild type or mutant HA-SNAP50 containing mSNAPcγ4 was tested in an in-vitro transcription initiation assay. HeLa cell nuclear extract was first depleted of mSNAPc with anti-SNAP43 antibodies and then transcription was tested in the presence of purified mSNAPCγ4 containing either wild type or mutant HA-SNAP50. As expected the mutations that showed abolished DNA binding activity also resulted in abolished transcription initiation for both U1 and U6 types of promoters. The addition of the mSNAPc with SNAP50 point mutants C354-A, C357-A, C380-A and C383-A in the conserved $C_{X2}C_{X23}C_{X2}C$ domain were completely unable to restore transcription for both types of promoters (Figure 42; lanes 13, 14, 16 and 17). Interestingly, the mutations on the $H_{X3}C_XH$ motif, H313-A, C317-A and H319-A, resulted in complete inability to restore transcription from the U1 snRNA promoter. Transcription, albeit weakened, was still evident for the U6 system (Figure 42; lanes 6, 7 and 8). The SNAP50 H388-A mutant was capable of wild type transcription from the U6 promoter, but its activity was severely reduced in the U1 system (Figure 42; lane 19; experiment was performed by Gauri W. Jawdekar)[59].

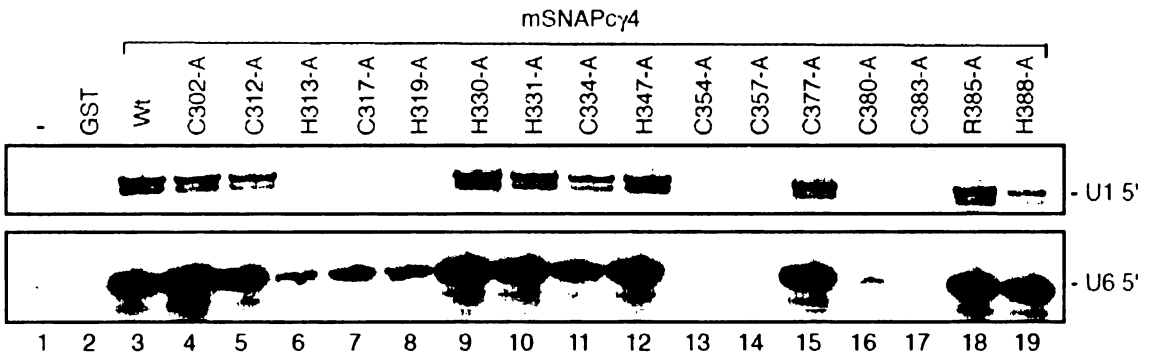


Figure 42: U1 and U6 in-vitro transcription initiation assay of wt or mutant HA-SNAP50 containing mSNAPcγ4. Mutations that resulted in reduced DNA binding also showed reduced transcription initiation activity for both types of promoters, with the exception of the H388-A mutation, which resulted in weakened U1 activity, but wild type U6 activity of transcription initiation59. HeLa cell nuclear extract was depleted with α-SNAP43 antibodies to immunodeplete endogenous SNAPc. In vitro U1 and U6 transcription was then tested in the absence (lanes 1) or presence of purified SNAPcγ4 (5 ng) containing wild type SNAP50 (lane 3) or mutant SNAP50 with the indicated alanine substitutions (lanes 4-19). Addition of GST alone did not reconstitute either U1 or U6 transcription as shown in lane 2 (experiment was performed by Gauri W. Jawdekar)[59].

2.11. Crystallization of mSNAPc variants

The mSNAPcγ3 and mSNAPcγ4 complexes obtained from the GST and ion exchange purification schemes was each concentrated to approximately 4.5 mg/ml and screened for crystallization using various crystallization screens, namely the Crystal Screens I and II (Hampton Research) and the PEG/pH grid screen. At room temperature, needle clusters were obtained in a variety of crystallization conditions. Most conditions

had magnesium chloride present as a salt additive and a PEG as a precipitating agent. The crystals grew in a wide range of pH; from pH 7 to pH 10. The crystals did not appear to be single under the light microscope and the size was insufficient for X-ray diffraction studies (Figure 43).

Figure 43: Crystals of mSNAPcγ4 as observed under the light microscope grown in 9% PEG5000MME, 100 mM Tris pH 8.5, 100 mM NaCl and 100 mM $MgCl_2$.

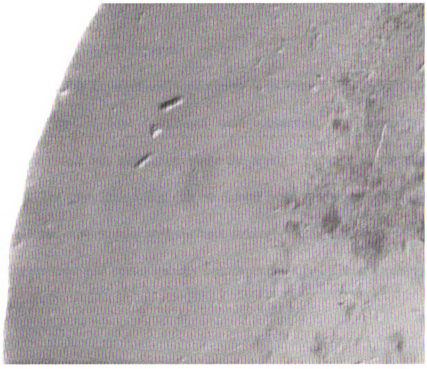


Figure 44: Optimized crystals of mSNAPcγ4 as observed under the light microscope grown in 8% PEG5000MME, 100 mM Tris pH 8.4, 100 mM NaCl and 100 mM $MgCl_2$.

Optimizations of the crystallization condition eventually lead to a condition where the crystals were larger, but the quality was not sufficient for X-ray crystallographic studies due to small size and morphology (Figure 44). Additive screening and replacement of all components of the crystallization condition with their analogs were tried. The effect of temperature on the crystallization was also investigated, by attempting to grow the crystals at 4°C. The crystals failed to appear at lower temperature even after extended periods of time. Further crystallization optimizations were unsuccessful, therefore micro and macro seeding experiments were performed, again without success.

2.12. X-ray diffraction of mSNAPc crystals

A needle cluster was cryogenically protected and tested for diffraction at the synchrotron facility. Rings at low resolution were observed, similar to a powder diffraction pattern. The result was surprising, since the crystals forming the needle cluster were relatively small. The conclusion that can be drawn from this experiment is that the crystals, albeit small and non-single, still diffract strongly (data not shown). A single crystal of larger dimensions and better quality could potentially diffract strongly and to high resolution.

2.13. Evaluation of mSNAPc crystals using transmission electron microscopy (TEM)

Since all crystallization optimization efforts did not result in improved crystal size and quality, and neither did the seeding experiments, we decided to observe the crystals under the TEM and investigate the morphology of crystals. The crystal structure of a protein of this size can not be determined using electron diffraction, but some interesting crystal parameters, like the unit cell dimensions, could.

The crystals were loaded on the EM grid and observed under TEM. Some crystal like formations were observed, but they were destroyed by the powerful electron beam. In order to be able to observe the crystals under the TEM, they were first cross linked with a formaldehyde and glutaraldehyde mixture, washed with water and then evaluated with the electron microscope.

The crystals observed were surprisingly not needle like, but appeared rectangular in nature and needle formations were built out of the small rectangles. What appeared to be needle clusters under the light microscope were actually large aggregates of small crystals (Figure 45, Figure 46).

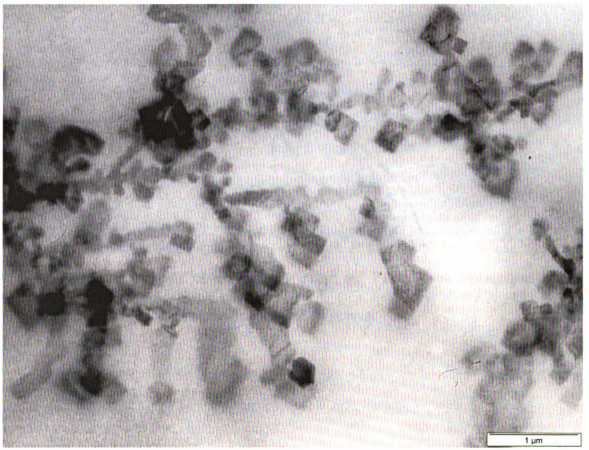


Figure 45: Needle clustered crystals of mSNAPcγ4 under the TEM have a heavily branched appearance and the branches are even further sectioned.

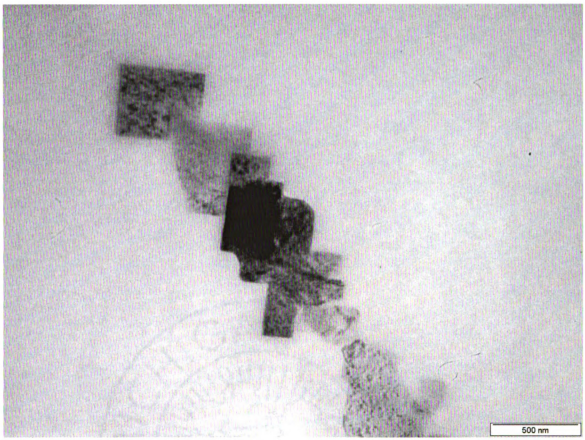


Figure 46: Crystals of mSNAPcγ4 under the TEM. Close-up on one of the branches showing a heavily fragmented formation.

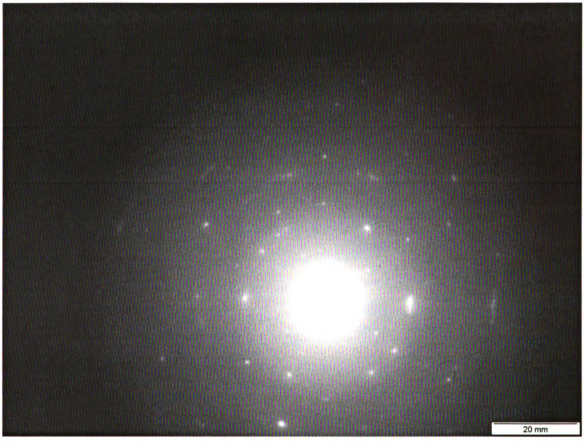


Figure 47: Circular diffraction pattern typical for a powdered material.

When observed in diffraction mode, these crystal clusters diffracted with powder diffraction like pattern as can be seen by the appearance of the circular pattern (Figure 47).

2.14. Diffraction of mSNAPc crystals using TEM

A few small single crystals were found under the TEM (Figure 48) and the diffraction aperture was set to select only one crystal. These crystals diffracted like a

single crystal and a different diffraction pattern was observed when the grid with the crystal was rotated for approximately 0.5 degrees (Figure 49).

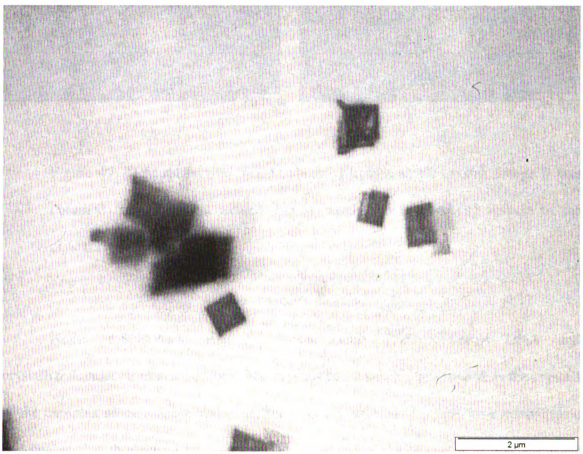
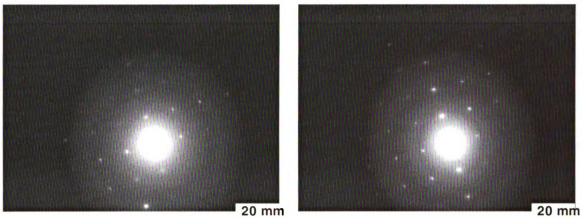


Figure 48: TEM image of mSNAPcγ4 single crystals as observed under TEM.



A B

Figure 49: TEM diffraction images obtained from a single crystal. Image B was obtained from a crystal rotated approximately 0.5 degree with respect to the crystal from image A.

Next, mSNAPcγ4 was mixed with PSE-containing DNA and crystallized under similar conditions. The crystals had a similar appearance as the crystals of the complex alone as observed under the light microscope. Crystals were cross-linked and analyzed with electron microscopy. The appearance of these crystals was improved compared to the crystals of mSNAPcγ4 alone, but the crystals were still fragmented and not single. Branch like single crystals were observed, emerging from a stalk (Figure 50). When the electron beam was focused on a single branch, a diffraction pattern, typical for a single crystal, was obtained (Figure 51).

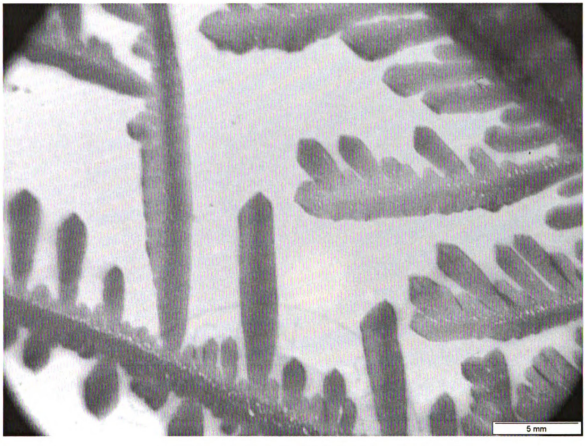


Figure 50: TEM image of mSNAPcγ4/DNA crystals.

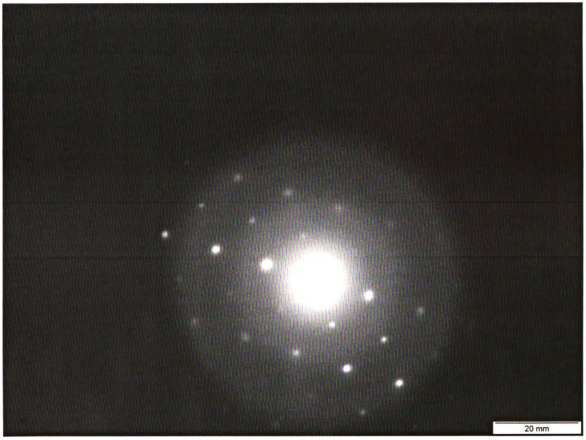


Figure 51: Diffraction pattern from a single branch obtained under TEM.

2.15. Purification optimization of mSNAPcγ4

Although the material that was used for crystallization and activity assays is relatively pure according to the SDS-PAGE stained with Commassie, the inability to grow high quality crystals suggests that further purification of the complex is required. Both anion and cation exchange resins were used in the attempts to prepare a more homogeneous material. The success of such purification was rather disappointing, since the protein eluted in an extremely broad range of the salt gradient. Modifications to the buffer used in this purification, with variations in pH of the buffer and additives, were

made, without much improvement in the purity of the complex. Due to the fact that the complex is eluted from the column in such a broad range, the efficiency of the ion-exchange purification is low. The complex is eluted in approximately 50% of the fractions, or more specifically in a salt gradient from 150 mM to 650 mM (Figure 52).

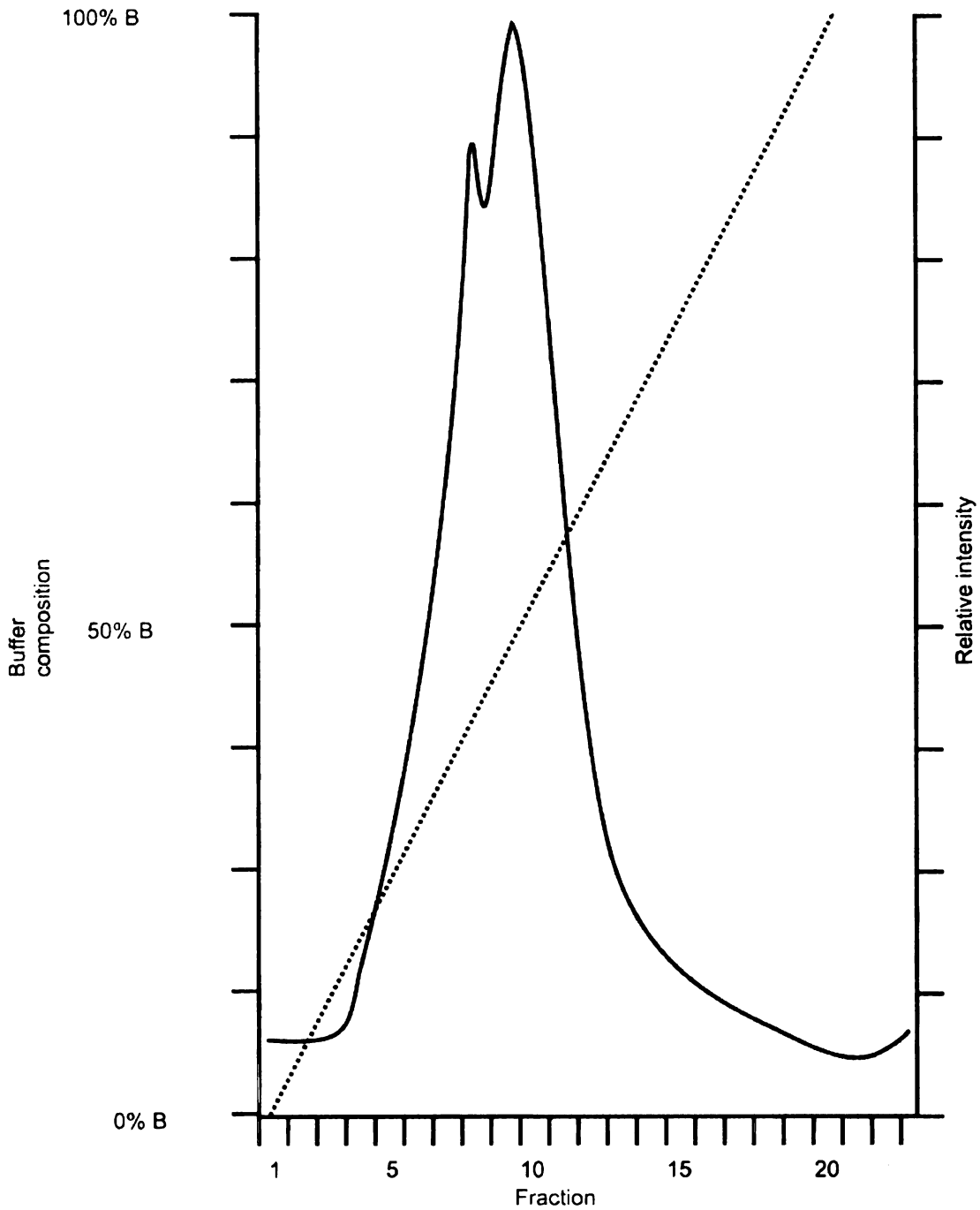


Figure 52: A representative elution profile for ion-exchange purification (Source-S) of mSNAPcγ4. The solid line represents the relative absorption intensity of the eluted protein material and the dotted line represents the salt gradient.

Since ion exchange chromatography using SourceQ or SourceS resin (Pharmacia) failed to accomplish an improvement in the purity of the material, gel filtration chromatography was employed next. Partially purified mSNAPcγ4 obtained from the affinity purification was concentrated either by centrifugation or by ammonium sulfate precipitation to enable efficient detection during the purification run. The complex was purified in the absence of the detergent using the size exclusion gel filtration and two peaks were observed in the chromatogram. The first peak that eluted in the void contained mSNAPcγ4. Interestingly, the second peak with the apparent molecular weight of 150 kDa contained the complex as well. The majority of the protein material submitted to gel filtration eluted in the first peak and only a fraction (approximately 5%) of the total material was eluted in the second peak. SDS-PAGE analysis of all fractions confirmed the presence of the complex, with some impurities being equally represented in both peaks (data not shown).

The presence of two peaks confirmed that the complex is not homogenous and an experiment was designed to determine how the removal of components from one peak would affect the behavior of the other. The significant difference in the apparent size of the two peaks and the fact that the first peak eluted close to the void volume suggested that the complex might be forming aggregates. Several gel filtration runs were performed and fractions were combined from the runs and concentrated to approximately 5 mg/mL. Crystallization experiments were set up with these fractions to determine which of the complexes is required for successful crystallization. Interestingly crystals formed first in the material from fraction 12, which corresponds to the second peak with the material of the apparent size of 150 kDa. mSNAPcγ4 obtained only with the use of affinity

chromatography was used as a positive control and also yielded crystals, but slightly later than fraction 12. More interestingly fractions 9 and 10 never yielded crystals and fraction 13 yielded crystals that were of better quality than those obtained from fraction 12 or the control. It is obvious that the material with the apparent size of 150 kDa (second peak) is required for crystal formation. Unfortunately this material is also a minor part of the total material obtained from the affinity chromatography, which consisted mostly of the aggregated material. Furthermore the separation of the two peaks is poor, since the second peak is only a shoulder of the first peak on the gel filtration run (Figure 53; A).

In order to obtain homogeneous material for crystallization, large quantities of mSNAPcγ4 would need to be purified and then the two major components would have to be separated on the gel filtration column. The separation would also be further complicated due to the fact that the two peaks are not well resolved and several sequential gel filtration experiments, each enriching the complex in the second peak, would need to be performed. Such a purification scheme would be relatively complicated and would yield little material suitable for crystallization from a large amount of the mixture. A pilot experiment was performed to determine the efficiency of the separation of the two peaks and also to check whether the two forms of the complex are in equilibrium. Fraction 10, which corresponds to the peak in the void and is presumably an aggregated form of the complex was concentrated and analyzed on the gel filtration column. The material eluted in the void volume as before, without any apparent presence of the second peak. Fractions 12 and 13, which correspond to the second peak that is presumed to be a hetero tetramer, were pooled and analyzed in the same manner. This material yielded a chromatogram where the first peak is under-represented compared to

the second peak resulting in material which is highly enriched compared to the previous gel filtration separations (Figure 53). Removal of the fractions containing the aggregate again would most likely yield a highly homogenous hetero tetramer mSNAPcγ4.

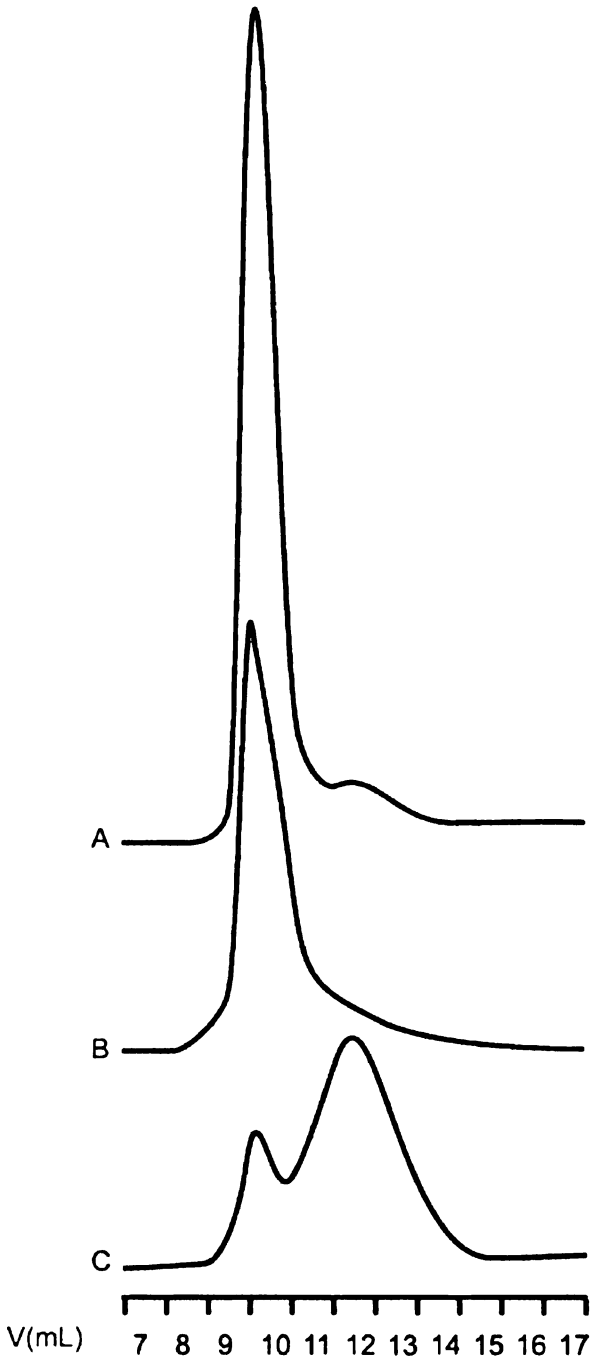


Figure 53: Gel filtration chromatograms of mSNAPcγ4 purified with affinity chromatography (A); mSNAPcγ4 rerun fraction 10 (B); mSNAPcγ4 rerun fractions 12 and 13 (C). The first peak in fractions 9 and 10 corresponds to 650 kDa and is also the

void. The second peak corresponds to a protein with an apparent molecular weight of 150 kDa according to the gel filtration standards (Biorad, data not shown).

Due to the fact that the un-aggregated mSNAPcγ4 is under-represented (approximately 5% of injected protein material) in the starting mixture and the laborious purification scheme yielding low amounts of homogenous material (approximately 5% of the total injected protein material), the amount of un-aggregated mSNAPc must be increased.

In order to change the ratio of the two peaks a few approaches were taken. First, since the ICP-MS metal analysis of the gel filtration fraction 9, which corresponds to the aggregate peak, yielded 0.5 mol of iron per 1 mol of protein for fraction 9 and 0.3 equivalents of iron for the mixture (data not shown), it seems plausible that the iron is responsible for aggregate formation. The aggregate fraction is enriched in iron and it is possible that the iron binds to the putative second metal binding site in SNAP50. Since the mixture contains less iron, it is also possible that the un-aggregated material contains less iron. The ab-initio modeling suggested that zinc, nickel or possibly iron could occupy the tentative metal binding site in SNAP50. Therefore, supplementing iron into the mixture of the two forms of the complex could result in complete aggregation of the complex and in contrast supplementation with zinc or nickel could result in a reduction of the aggregated material. Chelating the iron with EDTA could also potentially increase the amounts of the desired form of the complex. To test this assumption, mSNAPcγ4 samples obtained from affinity chromatography were incubated with different metal ions or with EDTA and then analyzed with size exclusion gel filtration. Interestingly, none of the

additives tested had any relevant effect of the ratio of the two peaks. The effect of the nickel supplementation could not be tested, since nickel was reduced upon mixing with the sample. The resulting elementary nickel could potentially damage the gel filtration column and was therefore not tested. 10 mM and 100 mM metal ions or EDTA were tested, but again resulted in no visible effect on the ratio of the two peaks suggesting that the formation of the aggregate is not related to metal content (Figure 54).

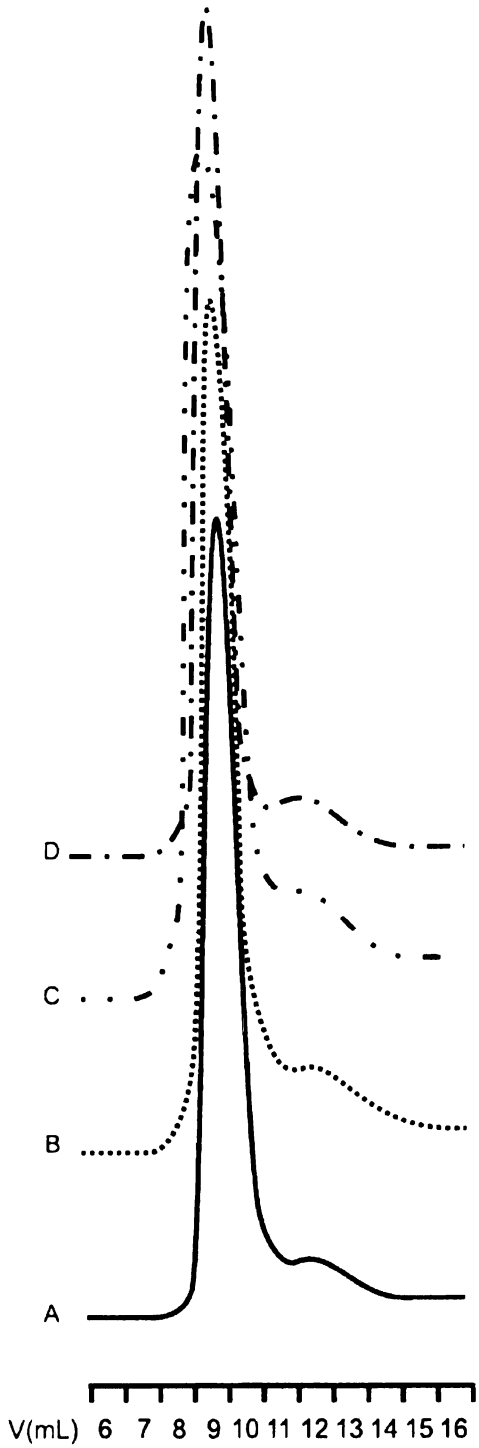


Figure 54: Gel filtration chromatogram of mSNAPcγ4 (A) supplemented with zinc (B), iron (C) or EDTA (D).

None of the additives used in the gel filtration separation had any effect on the relative ratio between the aggregate and the hetero-tetramer. In order to efficiently separate the two peaks, a gel filtration matrix with larger pores (Sephacryl-300) was used. All aspects of the purification remained the same as with Sephadex-200 gel filtration media used previously. The pores on this gel filtration matrix are approximately 50% larger compared to Sephadex-200 and the hypothesis was that the complex should migrate slower and therefore separate from the void peak. The hetero-tetramer peak actually separated and the purification yielded almost baseline separation. Fractions containing the hetero-tetramer were pooled, concentrated and injected to the same gel filtration column again. This time the two peaks were separated completely and the aggregate peak was present as a minor component of the binary mixture. It should be noted that both gel filtration runs were performed in a buffer without the detergent Tween-20 and the material could be efficiently concentrated using the Centriprep spin concentrators. The peak that corresponds to the void could only be concentrated in the presence of the detergent (Figure 55).

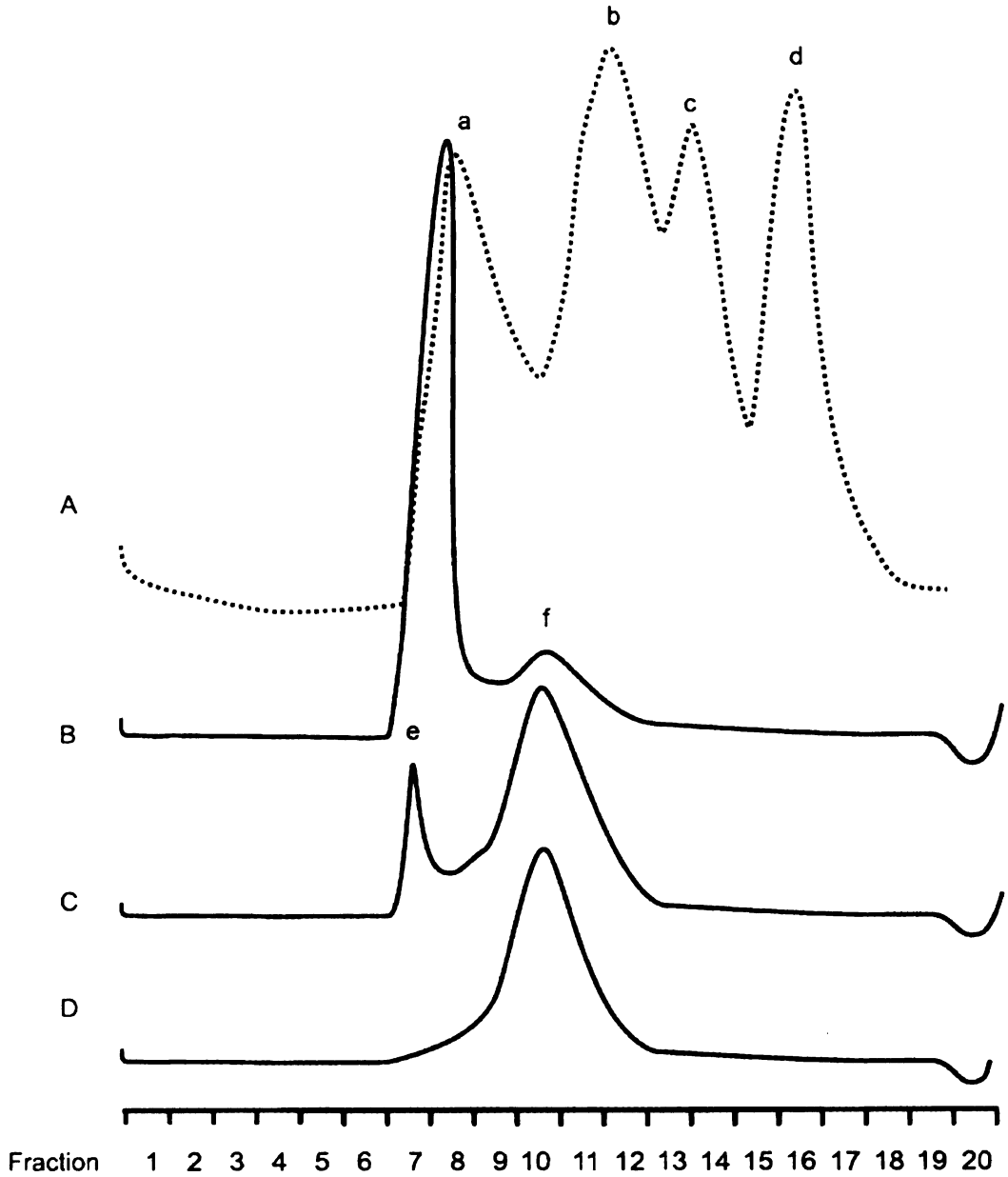


Figure 55: Gel filtration of mSNAPcγ4 using Sephacryl-300 size exclusion beads. (A) Biorad molecular weight standards (a=670 kDa, b=158 kDa, c=44 kDa, d=17 kDa), (B) GST purified mSNAPcγ4, (C) fraction 9-11 from the first gel filtration rerun using the same column, (D) fractions 9-11 from the second gel filtration rerun (e= void peak, f= mSNAPcγ4 heterotetramer). Fractions were 5 mL.

Highly purified and homogenous material that was obtained by two successive runs on Sephacryl-300 gel filtration matrix and was concentrated in the absence of the detergent with high recovery rates. Crystallization trials were started using the sparse matrix formulations and many conditions were found where the protein crystallized. Compared to previous preparations of the complex, the new approach resulted in protein material that gave crystals of improved quality. Although the crystals were not significantly larger the quality was greatly improved. Crystals no longer formed needle clusters but resembled single needle-like or orthogonal crystals (Figure 56). The number of successful crystallization hits was also increased drastically compared to the previous preparations of mSNAPCγ4.

Furthermore, the purification protocol was modified and Tween-20 is not required in any step of the purification. Although the amount of protein material that is immobilized on the GSH resin is reduced compared to the purification scheme where Tween-20 is used, the homogeneity of the complex is improved. The relative amounts of heterotetramer are comparable in the two purification approaches, but the amount of the aggregated material is somewhat reduced when no Tween-20 is used. Due to improved separation of the two peaks in the gel filtration step when Sephacryl-300 is used the recovery of the heterotetramer is also improved.

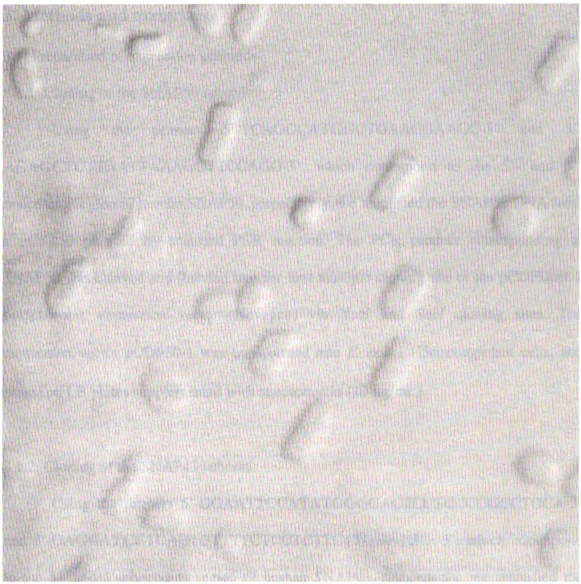


Figure 56: Crystals of mSNAPcγ4 grown in 100 mM Tris pH 8.5, 10% PEG4000, 100 mM NaCl and 100 mM MgCl2. Protein material used to obtain the crystals was prepared by successive gel filtration purification steps.

3. Methods and materials

3.1. Preparation of expression plasmids

3.1.1. Cloning of the SNAP50 subunit

Using the primers 5'-TCAG**CCATGG**CTGAAGGAAGC-3' and 5'-A**GAGCTC**TTAATTAAAGGTTCCAGG-3', which correspond to the 5'- and 3'-noncoding region of human SNAP50, respectively, we amplified the SNAP50 DNA from a pGST50 plasmid by standard PCR reaction. The PCR product corresponding to SNAP50 was cleaved and inserted into the first multiple cloning site of the pCDFDuet-1 polycistronic expression vector (Novagen) via *NcoI* and *SacI* cloning sites. The expression vector pCDF50-1 was transformed into *E. coli* DH5α competent cells, and plated on LB plates supplemented with streptomycin (20 μg/mL).

3.1.2. Cloning of the SNAP43 subunit

Using the primers 5'-GGAATTC**CATATG**GGGACTCCTCCCGGCCTGCA-3' and 5'-GA**GGATCC**TCAGTGTTTTCTCCTCTTCTTGGATGC -3', which correspond to the 5'- and 3'-noncoding region of human SNAP43, respectively, we amplified the SNAP43 DNA from a pGST43 plasmid by standard PCR reaction. The resulting PCR product corresponding to SNAP43 was digested with *NdeI* and *BamHI* and inserted into the pRSF50-1 expression vector via *NdeI* and *BglII* cloning sites. The expression vector pCDF43/50-1 was transformed into *E. coli* DH5α competent cells, and plated on LB plates supplemented with streptomycin (20 μg/mL).

3.1.3. Cloning of the SNAP19 subunit

Using the primers 5'-CTCA**CCATGG**TGAGCCGGCTTC-3' and 5'-AA**GGATCC**TTAGGAATCTGATTCTTC-3' which correspond to the 5'- and 3'-noncoding region of human SNAP19, respectively, we amplified the SNAP19 DNA from a pGST19 plasmid by standard PCR reaction. The resulting PCR product corresponding to SNAP19 was inserted into pRSFDuet-1 (Novagen) via *NcoI* and *BamHI* cloning sites. The expression vector pRSF19-1 was transformed into *E. coli* DH5α competent cells, and plated on LB plates supplemented with kanamycin (50 μg/mL).

3.2. Preparation of expression host

3.2.1. Preparation of competent cells for heat shock

The *E. coli* BL21-CodonPlus(DE3)-RIL were chosen for the over-expression of SNAP subunits and for co-expression. The cells were made competent by calcium treatment and the procedure is done on ice; 50 mL culture was grown until the OD_{600} (optical density) was 0.3 to 0.4 and the culture was cooled on ice for 5 minutes. The cell pellet was collected by centrifugation, resuspended in 4 mL ice-cold FSB (Frozen Storage Buffer; 10 mM potassium acetate, 100 mM KCl, 50 mM CaCl2 and 10% glycerol; pH 6.2) and kept on ice for 20 minutes. The cell pellet was again collected by centrifugation and resuspended in 3 mL FSB. Aliquots of the desired volume were prepared, frozen on dry ice and stored at -80°C.

3.2.2. Preparation of competent cells for electroporation

The procedure for preparation of electro competent cells was done on ice. Cells were spun at 3000×g. 50 mL cell culture was grown until the OD_{600} of 0.5. The cell pellet was collected by centrifugation and resuspended in 25 mL of milliQ water twice. The cells were harvested by centrifugation and resuspended in 10 mL 10% glycerol in milliQ. The cells were spun again and resuspended in 0.25 mL of 10% glycerol in milliQ. The resulting suspension of cells was stored in aliquots and frozen at -80°C.

3.2.3. Preparation and use of a glycerol stock

Cell colonies harvested from freshly grown plates were used to inoculate 50 mL of LB media supplemented with the appropriate antibiotic. The cells were grown for 12-16 h at 37°C. Glycerol was added to the culture to bring the concentration of glycerol to 20%. Cells were incubated for 1 h at 37°C, aliquoted and stored at -80°C for long term storage.

To use a glycerol stock, an aliquot was thawed at room temperature and 100 µL to 1mL of the cells were used to start a 50 mL LB culture, which was then grown for 12-16 h for further use.

Glycerol stocks prepared this way are viable up to at least 2 years.

3.2.4. Transformation into the cloning bacterial host

DH5α competent cells were thawed on ice and 1-5 µL of the DNA was added, mixed gently and the cells were placed on ice for 10 minutes. The cells were then placed

in a heated bath at 37°C for 40 seconds and then immediately placed on ice for 2 minutes. 900 µL of LB was added and the cells were incubated at 37°C for 1 h while shaking. 100 µL of this mixture was plated on agar plates supplemented with the appropriate antibiotic.

3.2.5. Transformation into the expression bacterial host Codon Plus RIL

BL21(DE3) Codon Plus RIL competent cells were thawed on ice and 1-2 µL of the DNA were added, mixed gently and the cells were placed on ice for 20 minutes. The cells were then placed in a heated bath at 42°C for 20 seconds and then placed immediately on ice for 2 minutes. 900 µL of LB was added and the cells were incubated at 37°C for 1 h while shaking. 100 µL of this mixture was plated on the agar plates supplemented with the appropriate antibiotic. Chloramphenicol at 50 µg/mL is required for the cells to retain the Codon Plus RIL plasmid.

3.2.6. Transformation into the co-expression bacterial host

The cells that were used for co-expression contained a single plasmid after the first transformation and were made competent by calcium treatment. Next, the cells were thawed on ice, mixed with 1-2 µL of plasmid DNA and placed on ice for 15 minutes. The cells were placed on a heated plate at 42°C for 45 seconds and then immediately returned on ice for 2 minutes. 900 µL of LB was added and the cells were incubated at 37°C for 1 hour and then plated on agar plates supplemented with the appropriate antibiotic mixture. The order in which the plasmids are transformed into the cells is irrelevant and expression levels are not dependant on the order in which the plasmids were transformed

(data not shown). Usually the pGST190(1-505) was transformed first, followed by pCDF43R/50-1 (to make mSNAPcγ3 producing cells) and finaly pRST19-1 (to make mSNAPcγ4 producing cells).

3.2.7. Electroporation

Electro-competent cells were thawed on ice and 20 μL of the cells were mixed with 1-2 μL of the DNA. The mixture was transferred into a sterile electroporation cuvete. The cells were then pulsed with 2000 V using a pulse electroporator (Biorad), 480 μL of LB was added and the cells were incubated at 37°C for one hour and plated on agar plates supplemented with the appropriate antibiotic mixture.

3.3. Growth media

3.3.1. Preparation of culture plates

To make 1 L of agar media for plates, 15 g of Tryptone, 5 g of yeast extract, 5 g of NaCl and 15 g of agarose were mixed and dissolved in 1 L of water. The solution was autoclaved. When the autoclaved solution was sufficiently cooled (60°C) the appropriate antibiotic was added and the media was poured into Petri-dishes and allowed to cool and harden. Prepared culture plates were stored at 4°C until use.

Table 2: Combinations of antibiotics used for preparation of agar plates. AMP, KAN, STR and CAM represent ampicillin, kanamycin, streptomycin and chloramphenicol respectively.

Concentration of antibiotics used for preparation of agar plates [µg/mL]				
Antibiotic used	Ampicillin	Chloramphenicol	Streptomycin	Kanamycin
AMP	100	-	-	-
KAN	-	-	-	100
STR	-	-	20	-
AMP/CAM	100	50	-	-
AMP/CAM/ STR	50	50	15	-
AMP/CAM/ KAN	50	50	-	15
AMP/CAM/ KAN/STR	50	50	10	10

3.3.2. LB growth media

To make 1 L of LB media, 15 g of Tryptone, 5 g of yeast extract and 5 g of NaCl were mixed and dissolved in 1 L of water. The solution was autoclaved. Before use appropriate antibiotic was added (Table 2).

3.3.3. TB growth media

To make 1 L of TB media, 12 g Tryptone, 24 g yeast extract and 4 mL glycerol was mixed and dissolved in 900 mL of water and autoclaved. Before use 100 mL of salt solution (2.31 g KH2PO4, 12.54 g K2HPO4) was added to the broth.

3.3.4. M9 minimal media

To make 1 L of M9 minimal media 4 g of $(NH_4)_2SO_4$ or 1 g NH_4Cl, 13.4 g of Na_2HPO_4, 6 g KH_2PO_4, 4 g NaCl and $MgSO_4$ were dissolved in water. 200 µL of the

trace elements solution (10 mM $FeCl_3$, 10 mM $ZnCl_2$ and 100 mM $CaCl_2$) was added and the mixture was autoclaved. Before use 20 mL of 20% autoclaved glucose, 20 mL of the Basal Eagle vitamin medium and the appropriate antibiotic were added.

3.4. Over-expression

3.4.1. Over-expression of individual SNAP subunits

Cells harboring the plasmid with the sequence for the desired protein were grown in LB medium supplemented with the appropriate antibiotic (100 μg/mL ampicillin). The 50 mL cultures were started from a freshly plated plate or from a glycerol stock and were allowed to grow at 37°C for 12-18 h. The cell culture was then transferred to 1 L LB containing flasks and allowed to grow at 37°C until 0.6 to 0.8 optical density. The cultures were induced by the addition of 1 mM IPTG and grown at the appropriate temperature (16°C for SNAP190(1-505), SNAP50 and SNAP43, and 23°C for SNAP19). The cells were grown for 12-18 h. The cells were collected by centrifugation and stored at -20°C until needed.

3.4.2. Co-expression of mSNAPcγ3 and mSNAPcγ4

Co-expression was performed virtually the same way as the expression of individual subunits. The only difference was the amount and the type of antibiotic used (Table 3).

Table 3: Antibiotic concentrations used in individual and co-expression systems

Concentrations of antibiotics used for growth in LB or TB medium [μg/mL]				
Protein construct in BL21(DE3) Codon Plus RIL	Ampicillin	Chloramphenicol	Streptomycin	Kanamycin
GST-SNAP190(1-505)	100	-	-	-
GST-SNAP50	100	-	-	-
GST-SNAP43	100	-	-	-
GST-SNAP19	100	-	-	-
MSCγ3	50	50	20	-
MSCγ4	50	50	20	20

3.5. Purification

3.5.1. GST purification

3L of cells harboring the over-expressed protein material were thawed on ice, resuspended in 80 mL TEMGT-250 (25 mM Tris pH 7.9, 2 mM EDTA, 12.5 mM $MgCl_2$, 10% glycerol, 0.1% Tween-20, 250 mM KCl and 3 mM DTT) and sonicated (3x 1 minute pulse, 1 minute pause). The cell extract was clarified by centrifugation, mixed with the glutathione (GSH) agarose beads (Amersham Pharmacia) and allowed to shake for 12 h at 4°C. GSH beads were washed with TEMGT-250 by centrifugation, transferred to a glass column equipped with a frit and washed with TEMGT-250 until the flow-thru showed no detectable protein matter by Bradford assay. 20 mL of TEMGT-250 were mixed with the resin and 50 u of thrombin were added to remove the GST tag. Cleavage was allowed to proceed to completion (12-20 h) and the protein material was eluted off the beads by several consecutive TEMGT-250 washes. This method was used to purify the individual subunits and the co-expressed material.

3.5.2. GST purification in the absence of a detergent

3L of cells harboring the over-expressed protein material were thawed on ice, resuspended in 80 mL TEG-250 (25 mM Tris pH 7.9, 2 mM EDTA, 10% glycerol, 250 mM KCl and 3 mM DTT) and sonicated (3x 1 minute pulse, 1 minute pause). The cell extract was clarified by centrifugation, mixed with the glutathione (GSH) agarose beads (GE) and allowed to shake for 12 h at 4°C. GSH beads were washed with TEG-250 by centrifugation, transferred to a glass column equipped with a frit and washed with TEG-250 until the flow-thru showed no detectable protein matter by Bradford assay. 20 mL of TEG-250 were mixed with the resin and 50 u of thrombin were added to remove the GST tag. Digestion was allowed to proceed for 6-8 h, while monitoring the cleavage progress by SDS-PAGE. Once the protein was completely digested 1 mM PMSF was added to stop the digest and the protein material was eluted off the beads by several consecutive TEG-250 washes.

3.5.3. Ion-exchange purification

The protein material obtained after the GST purification step was diluted with water, until the salt concentration was below 50 mM (5 fold) and loaded to ion exchange resin (Source-Q or Source-S, GE). The protein was then eluted with a 50 mM to 1 M NaCl salt gradient (50 mM or 1 M NaCl, 20 mM Tris pH 7.5, 10% glycerol, 2 mM EDTA, 12.5 mM $MgCl_2$, 0.1 % Tween-20 and 3 mM DTT). The purification was performed with 2 mL of resin, 5-8 mg of protein material, flow rate was 4-6 mL/min, and the FPLC instrument used was by Amersham Pharmacia.

3.5.4. Gel filtration purification

The partially purified protein material was concentrated to a desired concentration (0.5-5 mg/mL) and injected onto the gel filtration matrix. A variety of buffers were tested for this purification with relatively successful results. The formulation of a gel filtration buffer that works well is 10 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 0.05% Tween-20 and 3 mM DTT. For the Superdex-200 (GE) and Sephacryl-300 (GE) the flow rates were 1 mL/min and 0.5 mL/min respectively.

3.6. SDS-PAGE

Sample solution was mixed with an equal volume of 2xSPRS buffer (30 mM Tris HCl, 2.5% glycerol, 6% SDS, 0.005% Bromophenol Blue, 0.025% β-mercaptoethanol) and placed into a bath with boiling water for 1-2 minutes to denature the proteins. The samples were loaded in the well portion of the stacking gel (4% acrylamide) and separated at 220 V for 50-60 minutes. The resolving portion of the gel was 9-15 % acrylamide, 1% SDS in Tris HCl pH 7.5.

Proteins on the gel were stained with Commassie (4% methanol, 10% acetic acid and 1% Commassie blue in water; Commassie blue must be dissolved in methanol first, then mixed with acid and finally diluted with water due to the low solubility of the dye in water).

The gel was destained to reveal the positions of the protein material with DESTAIN (5% acetic acid, 20% methanol in water).

Destained gels were soaked in water for 30 minutes, then soaked in 5% glycerol for 30 minutes and dried between cellophane sheets for 24-48 h, to preserve the gels for long term storage.

3.7. Western blotting

Protein samples were separated using SDS-PAGE. Two filter paper sheets were placed on the bottom of the transfer cell (Biorad). Semi-dry buffer (50 mM Tris, 40 mM glycine, 0.037% SDS and 20% w/v methanol) was added and any air was removed by rolling with a cylinder. The gels were then placed on the prepared filter paper and more semi-dry buffer was added. The membrane was cut to fit the gel, soaked in the semi-dry buffer and placed on top of the gel carefully. 6 sheets of filter paper were soaked in semi-dry buffer and placed on the membrane. Any air was removed by rolling the cylinder. The transfer was allowed to continue for 50 minutes at 12 V. The membrane was removed and blocked with 4% milk solution in TBST (20 mM Tris pH 7.6, 0.15 M NaCl and 0.2 % v/v Tween-20) buffer. The membrane was then transferred to the primary antibody solution (antibody for the protein of interest in 4% milk solution in TBST) for 1 h. The membrane was rinsed with TBST buffer twice by soaking the membrane in the buffer for 10 minutes. The membrane was then transferred to the secondary antibody solution (secondary antibody in 4% milk solution in TBST) for 1 h and again rinsed with TBST as before. Solutions 1 and 2 were mixed in equal ratio and the membrane was soaked in this solution for 1 minute. The membrane was then placed on a filter paper, wrapped in Saran wrap and the bands were visualized on a photographic paper.

Primary antibody solution was prepared by a 1000 fold dilution of the antibody stock in 4% milk solution in TBST. The secondary antibody solution was prepared by a 3000 fold dilution of the stock antibody.

Rabbit and mouse antibodies were used for SNAPc subunits and GST, respectively, as primary antibodies. Horse α-rabbit and α-mouse were used as secondary antibodies. Antibodies α-SNAP190 #234, α-SNAP50 CSH303, α-SNAP43 CS48, α-SNAP19 CS543, α-mouse and α-rabbit were a generous gift from Dr. R. W. Henry.

3.8. DNA purification

DNA was separated on 1% agarose gel, stained with ethydium bromide and observed under UV light. Bands corresponding to the desired DNA product were excised with a spatula and the DNA was extracted from the gel slice with QIAGEN gel extraction kit, following the manufacturer's instructions.

3.9. TEM analysis of mSNAPc crystals

3.9.1. Cross-linking of mSNAPc crystals

A 1:1 mixture of paraformaldehyde and glutaraldehyde was added directly to the hanging drop containing protein crystals to a final concentration of 1% aldehyde and 1% glutaraldehyde. The drop was returned on top of the well for 1 h. Cross-linked crystals were then washed with water by addition of water and removal of the solvent with a pipette, to remove buffer, salt and precipitant components of the crystallization reagent.

3.9.2. Preparation of the TEM grid

Glass strips were thoroughly washed with a detergent and wiped dry. A few drops of 25% Formvar solution were added to the glass strip to form a thin film, without creating thick areas. The prepared film was allowed to dry completely (2-5 min). A razor blade was used to mark and cut the edges on the film. The glass strip with film was then carefully placed in a bowl with water and if successful the film was lifted from the glass and floated on the surface. 300 mesh copper grids were then carefully placed on the floating film and a paper was placed on top to sandwich the grids between the film and the paper. The prepared grids on the paper were removed from the water and dried.

3.9.3. TEM analysis

Cross-linked crystals were crushed with small crystallization tools to create a suspension with a large number of small crystals. A drop (1-5 μl) of the crushed and cross-linked crystals was placed on a Formvar grid (300 mesh copper) and excess liquid was removed with a filter paper. The grid was then placed in the sample holder and inserted in the instrument.

Samples were analyzed with JEOL 100CX transmission electron microscope. Voltage on the electron gun was 100,000 V. Condenser aperture #2 and objective aperture #2 was used. In diffraction mode diffraction aperture #2 or #3 was used.

The grid was viewed in low magnification mode (magnification 5,000) to identify an area on the grid with desired properties such as intact membrane, absence of precipitate, absence of dark areas, and presence of crystalline material. The crystals were

then observed at increasingly higher magnification and focus was adjusted at every magnification increase until the desired images were obtained. Images were recorded with a digital camera installed in the electron microscope and controlled with a MegaViewII digital camera controller. Images were viewed and manipulated with the AnalySIS software package.

3.10. Quantification of zinc in SNAP50

3.10.1. Determination of protein concentration

3.10.1.1. UV absorbance in 6 M urea

The approximate protein concentration was first determined with Bradford assay as per manufacturer's directions.. The sample was then mixed with urea in such manner that the measured absorbance was in the range of 0.1 to 0.6 where the assay is linear. In the protein concentration range of approximately 2 to 4 mg/mL (for mSNAPc) the expected absorbance was in the linear range if 50 µL of the sample was mixed in 450 µL of 6.5 M urea.

3.10.1.2. Standardized Bradford assay

Working Bradford solution (20%) was prepared fresh before use, by diluting the Bradford Assay stock in water. To measure the concentration, the protein (1-20 µL) is mixed in 1 mL of the working Bradford solution and absorbance at 600 nm was measured. A linear plot was obtained with the BSA standard solution (Biorad) in the range of concentrations from 0.05 to 0.8 mg/mL. The protein concentration was then determined from the plot. It is important that the absorbance of the sample in the

Bradford solution is within 0.1 and 0.6 to remain in the linear range. The plot points were prepared in duplicate and the protein was measured in triplicate.

3.10.2. Flame atomic absorption

The concentration of the protein sample was determined by UV absorbance measurement or by the standardized Bradford assay. The sample was transferred to a crucible, dried and ashed at 260°C in an oven and concentrated nitric acid was added to the brown powder and heated until all liquids evaporated. The process was repeated until only white powder remained in the crucible. Solution was reconstituted by the addition of a known volume of 5% nitric acid. To prepare the zinc standards, the mass of zinc metal (Spectrum, 99.9 %) was measured carefully and the zinc was dissolved in concentrated nitric acid and brought to the desired volume. Standards in the concentration range from 10 to 1000 ppb were made by serial dilution of the stock zinc solution with 5% nitric acid. The samples and standards were analyzed on the flame atomic absorption instrument (Varian SpectrAA-200) equipped with a zinc hollow cathode lamp operating at 213.9 nm. The flow rate was 1 mL/min.

3.10.3. ICP-MS

The protein concentration was determined with the standardized Bradford Assay using BSA as a standard. The sample solutions (protein, buffer) were transferred to a Teflon vial and brought to dryness on a hot plate. Concentrated nitric acid was added and the sample was placed on the hot plate and allowed to hydrolyze for 30 min. After cooling the sample was diluted with water to bring the acid concentration to 2% and 2%

nitric acid was added to bring the solution to the desired volume. Standards containing nickel and zinc in the concentration range 0-1000 ppb were prepared in 2% nitric acid from a 1000 ppm stock solution (Trace grade). The samples and standards were mixed with indium and bismuth solution as internal standards. The samples were run on an ICP-MS instrument (Micromass) with flow rate 0.5 mL/min. ^{66}Zn, ^{68}Zn and ^{60}Ni isotopes were measured and quantified. The responses of zinc and nickel were corrected according to indium and bismuth response.

4. References

1. Paule M. R., White R. J., Survey and summary: transcription by RNA polymerases I and III, *Nucleic Acids Research*, **28**, 1283 (2000).

2. Herr A. J., Jensen M. B., Dalmay T., Baulcombel D. C., RNA Polymerase IV Directs Silencing of Endogenous DNA, *Science*, **308**, 118 (2005).

3. Lee T. I., Young R. A., Regulation of gene expression by TBP-associated proteins, *Genes Dev.*, **12**, 1398 (1998).

4. Lobo S., Hernandez N., in Transcription, Mechanisms and Regulation (Conway R. C., and Conway J. W., eds) pp. 127-159, Raven Press, Ltd., New York (1994).

5. Bai L., Wang Z., Yoon J.-B., R. G. Roeder., Cloning and characterization of the beta subunit of human proximal sequence element-binding transcription factor and its involvement in transcription of small nuclear RNA genes by RNA polymerases II and III, *Mol. Cell. Biol.*, **16**, 5419 (1996).

6. Henry R. W., Ma B., Sadowski C. L., Kobayashi R., Hernandez N., Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc, *EMBO J.*, **15**, 7129 (1996).

7. Wang Y., Stumph W. E., RNA Polymerase II/III Transcription Specificity Determined by TATA Box Orientation, *Mol. Cell. Biol.*, **18**, 1570 (1998).

8. Schramm L., Hernandez N., Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters, *Genes Dev.*, **16**, 2593 (2002).

9. Challice J. M., Segall J., Transcription of the 5 S rRNA gene of *Saccharomyces cerevisiae* requires a promoter element at +1 and a 14-base pair internal control region, *J. Biol. Chem.*, **264**, 20060 (1989).

10. Ciliberto G., Raugei G., Costanzo F., Dente L., Cortese R., Common and interchangeable elements in the promoters of genes transcribed by RNA polymerase III, *Cell*, **32**, 725 (1983).

11. Braun B. R., Kassavetis G. A., Geiduschek E. P., Bending of the *Saccharomyces cerevisiae* 5S rRNA gene in transcription factor complexes, *J. Biol. Chem.*, **267**, 22562 (1992).

12. Krol A., Carbon P., Ebel J. P., Appel B., *Xenopus tropicalis* U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes, *Nucleic Acids Res.*, **15**, 2463 (1987).

13. Das G., Henning D., Wright D., Reddy R., Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III, *EMBO J.*, **17**, 503 (1988).

14. Kunkel G. R., Pederson T., Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used, *Genes Dev.*, **2**, 196 (1988).

15. Murphy S., Tripodi M., Melli M., A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes, *Nucleic Acids Res.*, **14**, 9243 (1986).

16. Baer M., Nilsen T. W., Costigan C., Altman S., Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P, *Nucleic Acids Res.*, **18**, 97 (1989).

17. Topper J. N., Clayton D. A., Characterization of human MRP/Th RNA and its nuclear gene: full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP, *Nucleic Acids Res.*, **18**, 793 (1990).

18. Lobo S., Hernandez N., A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter, *Cell*, **58**, 55 (1989).

19. Echenlauer J. B., Kaiser M. W., Gerlach V. L., Brow D. A., Architecture of a yeast U6 RNA gene promoter, *Mol. Cell Biol.*, **13**, 3015 (1993).

20. Orphanides G., Lagrange T., Reinberg D., The general transcription factors of RNA polymerase II, *Genes Dev.*, **10**, 2657 (1996).

21. Kuhlman T. C., Cho H., Reinberg D., Hernandez N., The General Transcription Factors IIA, IIB, IIF, and IIE Are Required for RNA Polymerase II Transcription from the Human U1 Small Nuclear RNA Promoter, *Mol. Cell. Biol.*, **19**, 2130 (1999).

22. Hampsey M., Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery, *Microbiol. Mol. Biol. Rev.*, **62**, 465 (1998).

23. Smale S. T., Kadonaga J. T., The RNA polymerase II core promoter, *Annu. Rev. Biochem.*, **72**, 449 (2003).

24. Wu S. –Y., Kershnar E., Chiang C. –M., TAFII-independent activation mediated by human TBP in the presence of the positive cofactor PC4, *EMBO J.*, **17**, 4478 (1998).

25. Nikitina T. V., Tishchenko L. I., RNA polymerase III transcription machinery: Structure and transcription regulation, *Mol. Biol.*, **39**, 161 (2004).

26. Ghavidel A., Schultz M. C., Casein kinase II regulation of yeast TFIIIB is mediated by the TATA-binding protein, *Genes Dev.*, **11**, 2780 (1997).

27. Yoon J. B., Roeder R. G., Cloning of two proximal sequence element-binding transcription factor subunits (gamma and delta) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein, *Mol. Cell. Biol.*, **16**, 1 (1996).

28. Hernandez N., Small Nuclear RNA Genes: a Model System to Study Fundamental Mechanisms of Transcription, *J. Biol. Chem.*, **276**, 26733 (2001).

29. Henry R. W., Sadowski C. L., Kobayashi R., Hernandez N., A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerase II and III, *Nature*, **374**, 653 (1995).

30. Wong M. W., Henry R. W., Ma B., Kobayashi R., Klages N., Matthias P., Strubin M., Hernandez N., The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1, *Mol. Cell. Biol.*, **18**, 368 (1998).

31. Gu L., Esselman J., Henry R. W., Cooperation between Small Nuclear RNA-activating Protein Complex ($SNAP_C$) and TATA-box-binding Protein Antagonizes Protein Kinase CK2 Inhibition of DNA Binding by $SNAP_C$, *J. Biol. Chem.*, **280**, 27697 (2005).

32. Mittal V., Ma B., Hernandez N., SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain, *Genes & Dev.*, **13**, 1807 (1999).

33. Ford E., Strubin M., Hernandez N., The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1, *Genes & Dev.*, **12**, 3528 (1998).

34. Hovde S., Hinkley C. S., Strong K., Brooks A., Gu L., Henry R. W., Geiger J., Activator recruitment by the general transcription machinery: X-ray structural analysis of the Oct-1 POU domain/human U1 octamer/SNAP190 peptide ternary complex, *Genes & Dev.*, **16**, 2772, (2002).

35. Sytina E. V., Pankratova E. V., Transcription Factor Oct-1: Plasticity and Multiplicity of Functions, *Mol. Biol.*, **37**, 637 (2003).

36. Henry R. W., Mittal V., Ma B., Kobayashi R., Hernandez N., SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III, *Genes & Dev.*, **12**, 2664 (1998).

37. Sadowski C. L., Henry R. W., Kobayashi R., Hernandez N., The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein, *Proc. Natl. Acad. Sci.*, **93**, 4289 (1996).

38. Kunkel G. R., Hixsom J. D, The distal elements, OCT and SPH, stimulate the formation of preinitiation complexes on a human U6 snRNA gene promoter in vitro, *Nucl. Acids Res.*, **26**, 1536 (1998).

39. Zhao X., Pendergrast P. S., Hernandez N., A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain, *Mol. Cell.*, **7**, 539 (2001).

40. Spector D. L., The dynamics of chromosome organization and gene regulation, *Annu. Rev. Biochem.*, **72**, 573 (2003).

41. Hinkley C. S., Hirsch H. A. Gu L., LaMere B., Henry R. W., The Small Nuclear RNA-activating Protein 190 Myb DNA Binding Domain Stimulates TATA Box-binding Protein-TATA Box Recognition, *J. Biol. Chem.*, **278**, 18649 (2003).

42. Mittal V., Hernandez N., Role for the amino-terminal region of human TBP in U6 snRNA transcription, *Science*, **275**, 1136 (1997).

43. Das D., Scovell W. M., The Binding Interaction of HMG-1 with the TATA-binding Protein/TATA Complex, *J. Biol. Chem.*, **276**, 32597 (2001).

44. Ford E., Hernandez N., Characterization of a Trimeric Complex Containing Oct-1, SNAPc, and DNA, *J. Biol. Chem.*, **272**, 16048 (1997).

45. Jin S., Levine A. J., The p53 functional circuit, *J. Cell Sci.*, **114**, 4139 (2001).

46. Ko L. J., Prives C., p53: puzzle and paradigm, *Genes Dev.*, **10**, 1054 (1996).

47. Levine A. J., p53, the Cellular Gatekeeper for Growth and Division, *Cell*, **88**, 323 (1997).

48. Stein T., Crighton D., Boyle J. M., Varley J. M., White R. J., RNA polymerase III transcription can be derepressed by oncogenes or mutations that compromise p53 function in tumours and Li-Fraumeni syndrome, *Oncogene*, **21**, 2961 (2002).

49. Gridasova A. A., Henry R. W., The p53 Tumor Suppressor Protein Represses Human snRNA Gene Transcription by RNA Polymerases II and III Independently of Sequence-Specific DNA Binding, *Mol. Cell Biol.*, **25**, 3247 (2005).

50. Hirsch H. A., Gu L., Henry R. W., The Retinoblastoma Tumor Suppressor Protein Targets Distinct General Transcription Factors To Regulate RNA Polymerase III Gene Expression, *Mol. Cell Biol.*, **20**, 9182 (2000).

51. Weinberg R. A., The retinoblastoma protein and cell cycle control, *Cell*, **81**, 323 (1995).

52. Horowitz J. M., Park S. H., Bogenmann E., Cheng J.C., Yandell D.W., Kaye D. J., Minna J.D., Dryja T. P., Weinberg R. A., Frequent Inactivation of the Retinoblastoma Anti-Oncogene is Restricted to a Subset of Human Tumor Cells, *Proc. Natl. Acad. USA*, **87**, 2775 (1990).

53. Horowitz J. M., Yandell D.W., Park S. H.,Canning S., Whyte P., Buchkovich K., Harlow E., Weinberg R. A.,Dryja T. P., Point mutational inactivation of the retinoblastoma antioncogene, *Science*, **243**, 937 (1989).

54. Novitch B. G., Mulligan G. J., Jacks T., Lassar A. B., Skeletal muscle cells lacking the retinoblastoma protein display defects in muscle gene expression and accumulate in S and G2 phases of the cell cycle, *J. Cell Biol.*, **135**, 441 (1996).

55. Onadim Z., Hogg A., Baird P. N., Cowell J. K., Oncogenic Point Mutations in Exon 20 of the RB1 Gene in Families Showing Incomplete Penetrance and Mild Expression of the Retinoblastoma Phenotype, *Proc. Natl. Acad. USA*, **89**, 6177 (1992).

56. Hirsch H. A., Gu L., Henry R. W., The Retinoblastoma Tumor Suppressor Protein Targets Distinct General Transcription Factors To Regulate RNA Polymerase III Gene Expression, *Mol. Cell Biol.*, **20**, 9182 (2000).

57. Hovde S. L., personal communication.

58. Hanzlowsky A., Jelencic B., Jawdekar G. W., Hinkley C. S., Geiger J. H., and Henry R. W., Co-expression of multiple subunits enables recombinant SNAPC assembly and function for transcription by human RNA polymerases II and III, *Protein Expr. Purif.*, **48**, 215 (2006).

59. Jawdekar G. W., Hanzlowsky A., Hovde S. L., Jelencic B., Feig M., Geiger J. H., Henry R.W., The unorthodox SNAP50 zinc finger domain contributes to co-operative promoter recognition by human SNAPc, *J. Biol. Chem.*, **281**, 31050 (2006).