ATTEMPTS TO FIND THE FUNCTION OF *Pf* -1806301 USING ANTIBODIES

by

GAURI C. WANNERE

(Under the direction of Dr. Bi-Cheng Wang)

ABSTRACT

Pf -180301 (*Pf*-1955, a putative archaeal Lsm protein, shows sequence similarity to both archaeal Lsm and bacteria Hfq proteins. Sm/Lsm proteins are a large family of RNA-binding proteins that have been identified in archaea, bacteria and eukarya. The X-ray structure revealed that *Pf* -1955 exists as dimer of octamer (16 units). The three-dimensional structure of *Pf* -1955 has revealed structural similarity with Sm/Lsm proteins, but also shows absence of key RNA-binding residues. This thesis attempts to recognize if *Pf*1955 binds to DNA or RNA.

The gene encoding *Pf* -1955 was recloned into pDESTC1 vector, which provides a Tev cleavage site for the removal of the His-tag after protein purification. The antibodies were generated using purified recombinant protein in solution. However, antibodies failed to recognize *Pf* -1955 from the native *Pyrococcus furiosus* cell extract. Any of the several reasons may be attributed to this failure: (a) the concentration of the *Pf* -1955 in native cell extract is not known (b) the protein should have been submitted as a gel-band to generate specific antibodies and (c) the 100 ^{o}C of boiling temperature is not enough to denature *Pf* -1955.

INDEX WORDS: *Pyrococcus furiosus,* Archaea, Sm/Lsm, Hfq

ATTEMPTS TO FIND THE FUNCTION OF *Pf* -1806301 USING ANTIBODIES

by

GAURI C. WANNERE

M.S., University, India, 1999

B.S., University, India, 1997

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2006

ATTEMPTS TO FIND THE FUNCTION OF *Pf* -1806301

USING ANTIBODIES

by

Gauri C. Wannere

Major Professor: Bi-Cheng Wang

Committee: Robert S. Phillips
John P. Rose

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2006

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. B. C. Wang for providing me an opportunity to work in his group and also for his consistent guidance, encouragement and support. I appreciate the knowledge and experience that I have gained.

I would like to thank Dr. Ira for helping with the initial PFU-1803601 cloning. I am grateful to Dr. Peter Horanyi for guiding me with PFU cloning and also the initial Western blot experiments. Dr. Praveen Alamuri's initial discussions and his valuable suggestions on immunoprecipiation were very essential for my understanding on this topic. I am also thankful to Dr. James Liu for assigning me this interesting project and to Lily for providing me with her opinion on general matters. Dr Hao's help on GP6 purification protocol is appreciated. I thank and appreciate Wang's Structural genomics division group for their lovely support and friendship. I am very thankful to Caryn for working with me through the Western blot and immunoprecipiation experiments and to Kevin for supplying his reagents for some of the experiments.

I acknowledge Dr. Rebecca and Dr. Michael Turns for allowing me to work in their lab for the immunoprecipiation experiments and also use their chemicals. Discussions and suggestions with you were not only extremely helpful but also served as a great learning experience.

Mom and Dad for their moral support and showing confidence in my abilities. Sisters for always being there for me. Chait for his constant support and encouragement.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Hyperthermophiles and *Pyrococcus furiosus*

Structural genomics (1) involves multidisciplinary research from bioinformatics, molecular biology, biochemistry, and biophysical methods such as NMR spectroscopy, mass spectroscopy, and x-ray crystallography. To determine the unique three-dimensional structures of proteins throughout the genome in order to better understand the relationship between protein sequence, structure and function is the primary goal of structural genomics.

The Southeast Collaboratory for Structural Genomics (SECSG) is one of the original seven NIH-funded Pilot Centers for high throughput (HTP) structure determination. SECSG consists of five collaboration institutes: The University of Georgia, the University of Alabama at Birmingham, the University of Alabama at Huntsville, Georgia State University and Duke University Medical Center (2).

Pyrococcus furiosus is one of the major model organisms selected by SECSG. *Pyrococcus*, a Latin name, stands for "fireball", indicates the ability of this organism to withstand and grow at temperatures above 100°C. The "extremophile" is also resistant to radiation. Due to its tendency to grow favorably at higher temperature, *Pyrococcus* belongs to a group of organisms called hyperthermophiles. In 1982, Stetter and coworkers isolated the first hyperthermophile from shallow marine volcanic vents (3). More than 20 genera of hyperthermophiles have been discovered ever since (Table 1.1) (4a). Hyperthermophilic organisms are abundant in geothermally-heated ecosystems. These organisms are mostly obligate

anaerobes, and depend on electron donors like elementary sulfur (S^0) instead of O_2 for their energy conservation (4b).

The hyperthermophiles categorization offers new insight into the origin of life. Based on 16S-rRNA analyses of known hyperthermophilic microbes, except for two bacterial genera, all of the hyperthermophile genera are classified as Archaea (formerly known as Archaeabacteria) (5). In 1990, Woese *et al*. proposed a universal phylogenetic tree for all living organisms. Woese divided the living world into three domains: the Bacteria, the Archaea, and the Eukarya (5). The Archaea is further subdivided into two kingdoms: the Crenarchaeota and the Euryarchaeota. While the Crenarchaeota contains a few thermophiles and all the methanogens, Euryarchaeota comprises of homogenous tight clustering of extreme thermophiles including *Pyrococcus* (Figure 1.1).

The phylogenetic tree give indications that the archaea and eukarya are most likely evolved from a common ancestor, which is not shared by bacteria. Within both the archaea and the bacteria domain, the hyperthermophilic species appear to be the most slowly evolved organisms (6). Hence the hyperthermophilic organisms are believed to be most closely related to the ancestral form of life, which was first evolved under the high temperature on earth. Phylogenic studies show that the rest of the life forms result from evolutionary pressure to adapt to lower temperatures (5).

Table 1.1 The hyperthermophilic genera: organisms that grow optimally at 90 °C

Genus	T_{max} (°C) [(a)]	Metabolism [(b)]	Substrate [(c)]	Acceptors
S^0-depentdent archaea				
Thermofilum	100	Hetero	Pep	S^0, H^+
Staphylothermus	98	Hetero	Pep	S^0, H^+
Thermodiscus	98	Hetero	Pep	S^0, H^+
Desulfurococcus	90	Hetero	Pep	S^0, H^+
Thermoproteus	92	Hetero (auto)	Pep, CBH (H_2)	S^0, H^+
Pyrodictum	110	Hetero (auto)	Pep, CBH (H_2)	S^0, H^+
Pyrococcus	105	Hetero	Pep	S^0, H^+
Thermococcus	97	Hetero	Pep, CBH	S^0, H^+
Hyperthermus	110	Hetero	Pep (H_2)	S^0, H^+
Stetteria	103	Hetero	Pep + H_2	S^0, $S_2O_3^{2-}$
Pyrobaculum	102	Hetero (auto)	Pep (H_2)	S^0, mO_2, NO_3^-
Acidianus	96	Auto	S^0, H_2	S^0, O_2
S^0-indepentdent archaea				
Sulfobococcus	95	Hetero	Pep	-
Aeropyrum	100	Hetero	Pep	O_2
Pyrolobus	113	Auto	H_2	$S_2O_3^{2-}$ NO_3^- mO_2
Sulfate-reducing archaea				
Archaeoglobus	95	Hetero (auto)	CBH (H_2)	SO_4^{2-} $S_2O_3^{2-}$
Iron-oxidizing archaea				
Ferroglobus	95	Auto	Fe^{2+}, H_2, S_2^-	$S_2O_3^{2-}$ NO_3^-
Methanogenic archaea				
Methanococcus	91	Auto	H_2	CO_2
Methanothermus	97	Auto	H_2	CO_2
Methanopyrus	110	Auto	H_2	CO_2
Bacteria				
Thermotoga	90	Hetero	Pep, CBH	S^0, H^+
Aquifex	95	Auto	S^0 (H_2)	NO_3^- mO_2

From Roopali, R. Ph. D dissertation, University of Georgia, 2001
(a) Maxium growth temperature
(b) Heterotrophic (hetero) of autotrophic (auto) growth mode
(c) Growth substrates: peptides (Pep), carbohydrates (CBH), hydrogen (H_2) and elemental sulfur (S^0) as electron receptor

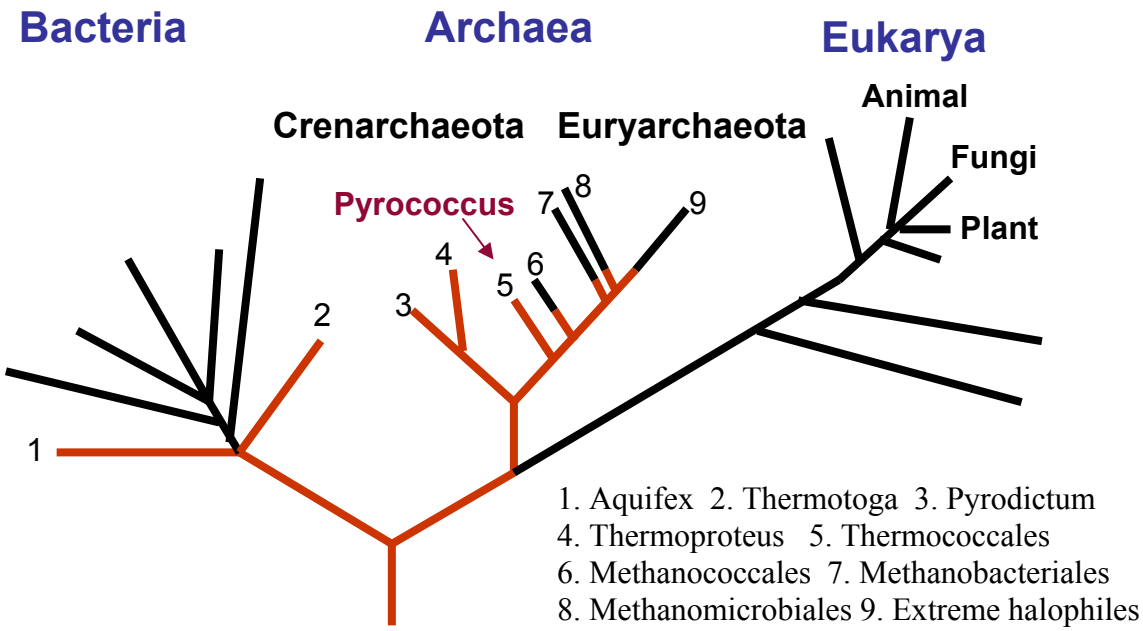


Figure 1.1 The rooted universal phylogenetic tree based on 16 rRNA sequence analysis. Showing three domains: bacteria, archaea and eukarya. Hyperthermophilic genera are colored red. Position of *Pyrococcus* is pointed by arrow.

Modified from Woese et al., *Proc Natl Acad Sci* U S A. 1990 87(12):4576-9.

***Pf* -1955 is a putative Sm-like protein**

Archaeal genomes share a common ancestor with Eukarya based on ribosomal RNA sequence comparisons (7). In addition these genomes encode many protein complexes characteristic of eukaryotic small nuclear ribonucleoproteins (snRNPs). SnRNPs are small RNA-protein complexes found in the eukaryotic nucleoplasm (8) and are considered crucial components in several RNA processing events, including pre-mRNA splicing (9, 10), rRNA processing (9, 11), mRNA degradation (12, 13), telomere synthesis (14) and histone pre-mRNA processing (15-17).

In general, snRNP consists of a small nuclear RNA (snRNA), associated with several snRNP specific proteins and a set of seven Sm or Sm-like (Lsm) proteins (18). snRNP's are tight complexes of one or more proteins with a short RNA molecule (usually 60 to 300 nucleotides) inhabitating every compartment of eukaryotic cells. Antibodies against the protein or RNA moiety of snRNPs are often generated for unknown reason during the autoimmune diseases (19). Autoantibodies were found to recognize an antigen called Sm in patients suffering from systematic lupus erythematosus (20). The Sm antigen was found present in U7 snRNP, spliceosomal snRNP U1, U2, U4 and U5, and some snRNPs with unknown functions. The Sm antigen was identified as seven small homologous proteins B/ B´, D1, D2, D3, E, F and G, which are also referred as the canonical Sm proteins (21).

Analyses based on amino acid sequence reveal that Sm proteins form a distinct family that share an approximately 70-residue signature sequence called the Sm domain. There are two conserved regions in this domain termed the Sm1 and Sm2 motif, which are linked by a variable length loop region (22- 24). Database search of eukaryotes indicates that a large number of proteins incorporate the Sm domain. Some of them are highly homologous to human Sm

proteins, hence they might be the Sm counterpart in other eukaryotic species. However, there are yet other proteins that share no clear sequence similarity with Sm proteins. These proteins were named Sm-like (Lsm) proteins (23, 24). Based on the relationship between the canonical Sm proteins and Sm-like proteins, nine subtypes of this Sm/Lsm protein family could be divided (Table 1.2). Among the nine Sm-like proteins identified in yeast, the Lsm2-Lsm8 complex directly associates with spliceosomal U6 RNA and pre-RNase P RNA, indicating that both Sm and Sm-like proteins are involved in RNA processing events (25).

The formation of Sm/Lsm complex involves the interaction between Sm/Lsm proteins and the specific recognition of cognate RNAs. The seven canonical Sm proteins B/B´, D1, D2, D3, E, F and G assemble around snRNA at a conserved uridine-rich site called Sm site to form the Sm core domain in eukaryotic spliceosomal snRNPs, (26). It is hypothesized that the formation of the Sm core domain plays an important role in snRNP biogenesis. Sm core domain triggers the hypermethylation of the m^7G cap of snRNA (27), and also acts as one of the bipartite nuclear localization signal for the transportation of the newly assembled snRNPs back to the nucleus (28). A remarkable similarity is found between human canonical Sm core domains and purified human Lsm complex based on Electron micrographs; Sm/Lsm complexes are doughnut-shaped ring structures with a diameter of about 70-80Å (29, 30). However, due to the complex composition of these eukaryotic Sm/Lsm complexes, the three-dimensional structures of the complexes have not been determined yet.

Although the Sm/Lsm proteins in eukaryotic spliceosomal snRNPs have been well studied for many years, their archaeal counterparts had not been identified until 1999, when database searches were conducted on genomes of *Methanobacterium thermoautotrophicum* and *Archaeoglovus fulgidus* (25).

Table 1.2 Classification of the Sm/Lsm family in human, yeast, archaea and bacteria

Classification	1	2	3	4	5	6	7	8	9
Human	SmB	SmD1	SmD2	SmD3	SmE	SmF	SmG		
		Lsm2	Lsm3	Lsm4	Lsm5	Lsm6		Lsm1	
Yeast	SmB	SmD1	SmD2	SmD3	SmE	SmF	SmG		
	Lsm8	Lsm2	Lsm3	Lsm4	Lsm5	Lsm6	Lsm7	Lsm1	Lsm9
Archaea	Lsm1 (AF, PA, PH, MT, SS, AP, TA, HN, PyA, TV)								
	Lsm2 (AF, AP, MT, SS, TA, TV)								
Bacteria	Hfq (~30 species)								

AF, *Archaeoglobus fulgidus*; PA, *Pyrococcus abyssi*; PH, *Pyrococcus horikoshi*;
MT, *Methanobacterium thermoautotrophicum*; SS, *Sulfolovus solfataricus*;
AP, *Aeropyrum pernix*; TA, *Thermoplasma acidophilum*; HN, *Halobacterium* sp. NRC-1;
PyA, *Pyrobaculum aerophilum*; TV, *Thermoplasma volcanicum*

In contrast to eukaryotes, archaeal genomes were found to encode at most two subtypes of Sm-like proteins, termed as Lsm1 and Lsm2 (25). Until now Sm-like proteins have been identified in more than ten archaeal species. Some species encode only Lsm1 type proteins, while other species encode both Lsm1 and Lsm2 type proteins (31) (Table 1.2). The sequence comparison of these Lsm1 proteins reveals a high level of sequence homology, suggesting that Lsm1 from different archaeal genomes are orthologues. On the other hand, Lsm2 type proteins are more diverse in their sequences, having only 30% or less sequences identity within Sm2 type and with Sm1 type proteins. Another class, known as Hfq proteins, are a widespread and well-conserved family of bacterial factors involved in RNA modulation. Based on the crystal structures, these proteins were recently found to share the characteristic Sm fold (32, 33) (Figure 1.2). Although the overall sequence similarity between the eukaryotic Sm proteins, archaeal Lsm proteins and bacterial Hfq proteins are low, their core Sm-folds are strikingly conserved, with an average root mean square deviation of 0.91± 0.28Å (32). This gives another good example that for proteins with similar biochemical function, their 3-D structures are more conserved than the sequences during the evolution (Figure 1.3).

The first crystal structures of archaeal Sm-like proteins were determined in 2001. The structures revealed a donut-shaped heptameric ring structures comprised of seven identical Lsm subunits (34-36). Each subunit consists of an N-terminal α-helix followed by five-twisted β-sheets, which can be divided into two sequence motifs: Sm1 (β strand 1,2 and 3) and Sm2 (β strand 4 and 5) (Figure 1.2). These two parts are linked together by a loop region (L4) of variable length and composition. The archaeal Lsm fold was found to be highly conserved with human canonical Sm proteins and bacterial Hfq proteins, indicating only slight differences at the N-terminal and L4 region (32). The whole complex is held together by hydrogen bonding from the

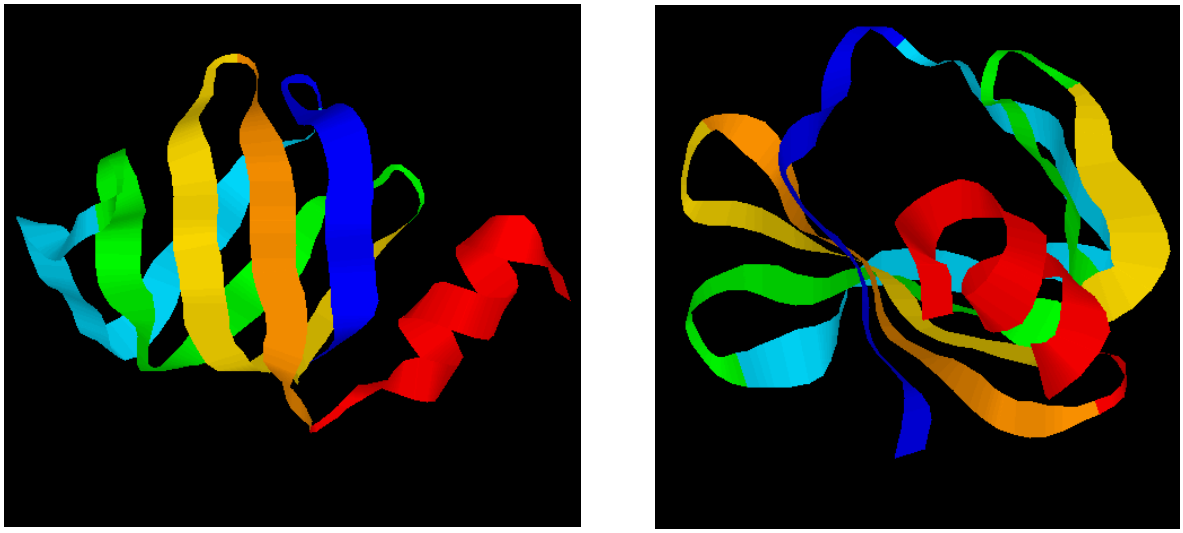
backbone between β4 of one subunit and β5 of the adjacent subunit, as well as the hydrophobic and other side-chain interactions.

Oligomerization and RNA-binding of Sm/Lsm proteins

Sm and Sm-like proteins from eukarya, archaea and bacteria are organized in a ring structure. This organization can be easily explained by the structure of the monomer, where the curved β strands form a compact hydrophobic core. By interacting and pairing with the adjacent subunits, the hydrophobic core is shielded from the hydrophilic environment. The formation of the ring structure thus allows the hydrophobic cores of all component subunits to be enclosed (34).

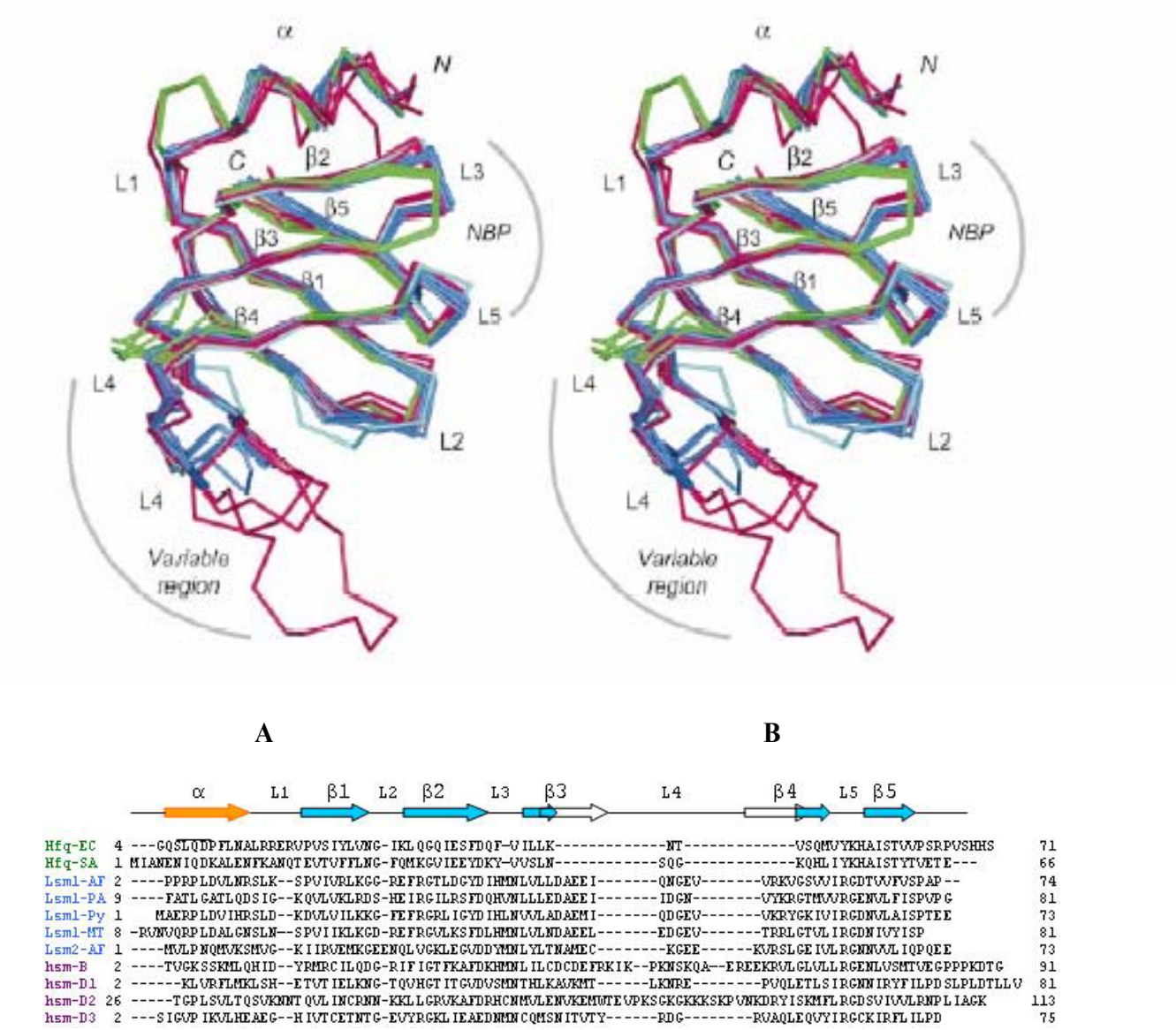
Sm/Lsm, Sm and Sm-like protein tend to form cyclic oligomers (Figure 1.4). The formation of a eukaryotic hetero-heptamer Sm core structure of snRNPs is confirmed by biochemical and electron microscopic studies (37, 38). Five crystallographic structures of archaeal Lsm proteins have also been determined. Except for the only Lsm2 protein that is in a homo-hexameric formation, all four Lsm1 proteins arrange in a homo-heptameric organization. On the other hand Hfq proteins in bacterial organisms seem to exclusively favor the homo- hexameric formation (32). Comparison of the sequences and structures of the above Sm/Lsm proteins could not give an indication to what drives the preference for heptamer or hexamer formation. More structural data are needed to help answer this question.

Crystallographic structures of some archaeal Lsm proteins show that these prefer a higher degree of polymerization. For example, Lsm1 complex from *Pyrococcos abyssi* is packed as two heptamers in a head-to-head orientation (head refers to the face containing the N-terminal α-helix).



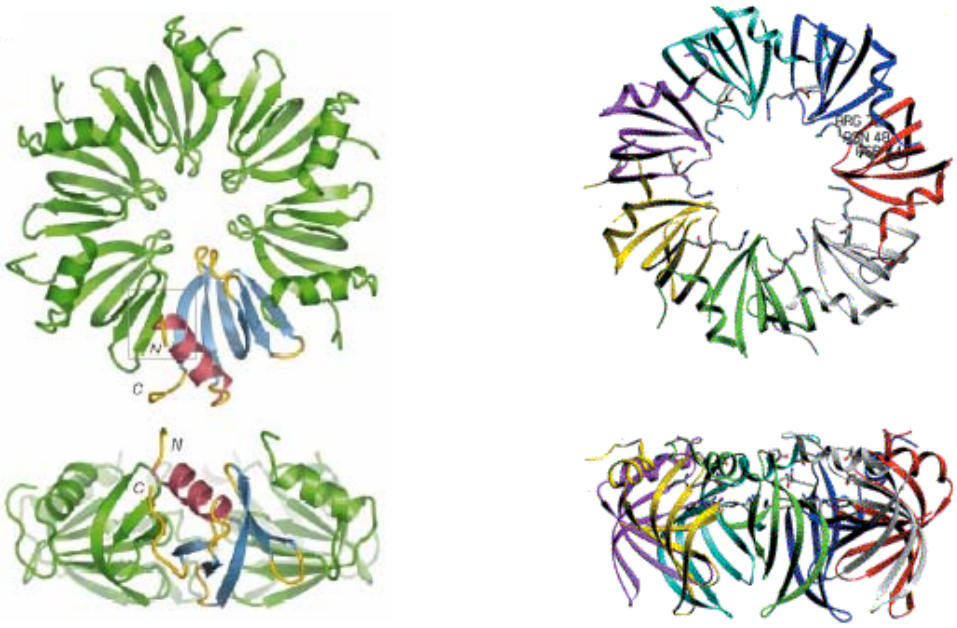
A **B**

Figure 1.2 Crystal structure of Lsm1 monomer from *P. abyssi* (1H64). Figure A and B show the same structure viewed from different orientations. This structure represents a typical Sm-fold, consisting of a N-terminal α-helix followed by five β-strand, which are colored by red, orange, yellow, green, cyan and blue respectively.



From Sauter et al., (2003) Nucleic acids research 14, 4091-4098.

Figure 1.3 Sequence and structural comparison of the Sm-fold in Sm, Lsm and Hfq proteins from human, archaea and bacteria, respectively. (A) Superposition of the Sm-fold in Hfq monomers from bacteria (green), Lsm monomers from archaea (blue) and Sm proteins from human (red) (B) The sequences of the Sm, Lsm and Hfq proteins whose structures are shown above.



A. Hexameric bacterial Hfq protein

Sauter *et al. Nucleic Acid Research*, 2003

B. Heptameric archaeal Lsm protein

Collins *et al. J. Mol. Biol*, 2001

Figure 1.4 Oligomeric organizations of Sm/Lsm and Hfq complexes. (A) The front and side views of the hexameric *E.Coli* Hfq protein complex. (B) The front and side views of the heptameric Lsm1 protein complex from *Methanobacteriaum thermoautotophicum*.

This dimerization of the heptamer was further confirmed by EM and gel filtration studies. It was suggested that the dimerization might imply the real function (39). Other Lsm1 proteins from *Pyribaculum aerophilum* and *Methanobacterium thermautotrophicum* also showed similar head-to-head organization of heptamers (40).

Sm and Sm-like proteins, in addition to their ability to form oligomers, are characterized by being able to bind RNA at a uridine-rich site called the Sm site. EM studies suggest that the eukaryotic Sm core proteins assemble around the RNA in a ring structure. Furthermore, this RNA is channeled through the central cavity of the ring (37). In sharp contrast, archaeal Lsm and bacterial Hfq complexes were shown to act only as a docking site where the cognate RNA molecules bind at the surface of the oligomers instead of threading through the central cavity. These differences were further verified by the facts that eukaryotic Sm proteins assemble into the heptameric core only in the presence of its target RNA, while archaeal Lsm and Hfq proteins exist as stable oligomers without binding to the RNA (32).

It was proposed that archaeal Lsm proteins have a generic single-stranded RNA binding activity. *P. abyssi*, *A. fulgidus* and *M.thermoautotrophicum* are several of the Lsm-RNA complexes, whose structures have been determined, (31, 39, 40). The length of the binding ligand RNAs range from a single UMP to a uridine heptamer (U7). However, it has been noted that the residues involved in uridine interaction are strictly conserved. The uridine-binding pocket includes histidine and asparagine in loop3 and arginine in loop5. The uracil ring is stacked firmly between the side chains of arginine and histidine, and form hydrogen bonds with the conserved asparagine (Figure 1.5).

The structure of *Pf* -1955 was identified in 2004. *Pf* -1955, a putative Lsm protein in archaea *P. furiosus,* was identified to be an Lsm2 type protein by sequence comparison. Earlier

attempts in crystallization of recombinant *Pf* -1955 with His-tag produced the crystals that diffracted to only ~8Å. It is known that a His-tag can sometimes interfere with protein interaction and lead to low diffraction quality crystals. The cDNA of *Pf* -1955 was recloned into vector pET-28, a thrombin cleavage site downstream of the N-terminal His-tag. The recombinant protein was expressed and purified from *E.Coli* BL21 (DE3) and further cleaved from His-tag to produce native crystals that diffracted to 2.7Å at the Advance Photon Source (APS), Argonne National Laboratory. The best crystal diffracted to 2.7Å using the home-source X-ray system. The overall structure of *Pf* -1955 was predicted to be a dimer of octamers.

Since *Pf* -1955 is a dimer of an octamer, it is expected that the core of this protein may bind to a variety of DNA and/or RNA substrates. In principle *Pf* -1955 and more than several thousand combinations of given nucleic acid sequence are possible, but such in vitro experiments may not necessarily reveal the true function of 1955. However the function of native *Pf* -1955 can only be identified by performing an immuno pull down assay on the native cell extract, which contains *Pf* -1955 bound to a specific RNA or DNA sequence. There are several steps involved in immuno pull down assay (discussed in materials and methods section), starting with getting antigen specific antibodies.

In this thesis we attempted to understand the function of *Pf* -1955 by performing the antibodies study. The production of antibodies, which are host glycoproteins, is an immune response to the presence of a foreign body. They are synthesized primarily by plasma cells, and circulate throughout the blood and lymph, where they bind to foreign antigens. Functionally antibodies bind not only to the antigens but also to specific cells or proteins, which are a part of immune system.

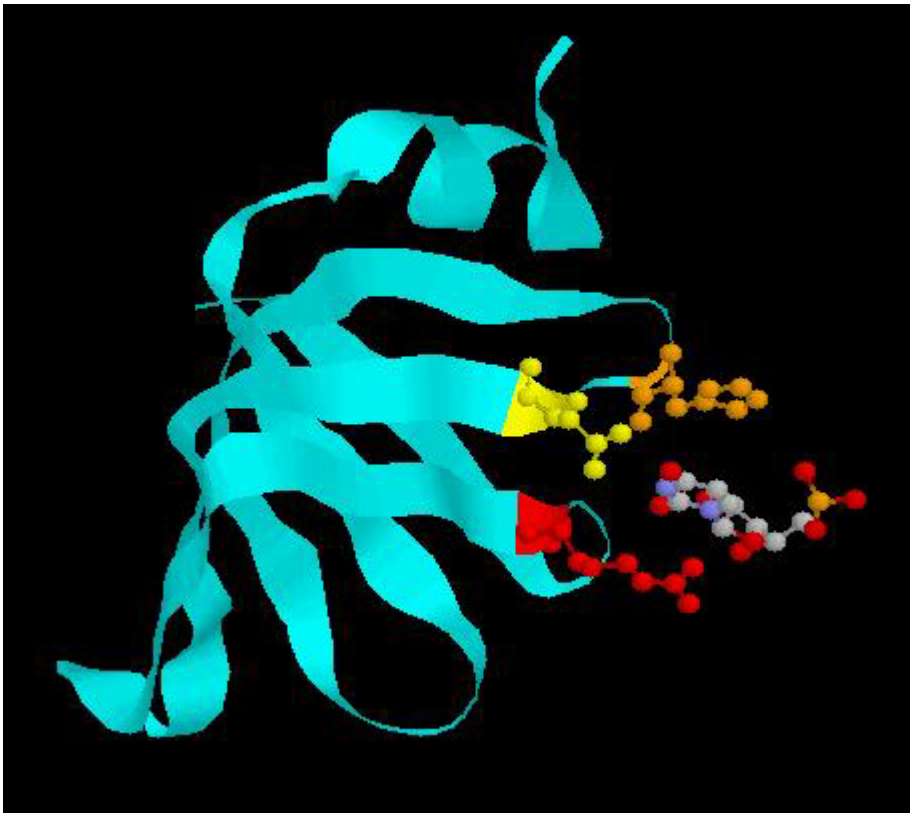


Figure 1.5 The internal uridine-binding pocket of *P. abyssi* Lsm1 complex (1M8V)
The uracil is stacked between His37 (orange) and Arg63 (red), and also forms hydrogen bonds with Asn39 (yellow).

Antibodies are comprised of one or more copies of Y shaped unit containing four polypeptides; two identical copies of heavy chain and two of light chain (41). The two heavy polypeptide chains are approximately 55000 daltons while the light chains are 25000 daltons (Figure 1.6). One light chain associates with the amino terminal region of one heavy chain to form an antigen binding domain. The carboxy terminal regions of the two heavy chains fold together to make the FC domain. The four polypeptide chains are held together by the disulphide bridge and non covalent interactions. The lateral and rotational movement of the Y shaped segment gives mobility to the antigen binding domain of the antibody. Thus antibodies can interact with a large number of different conformations of antigens (Figure 1.7).

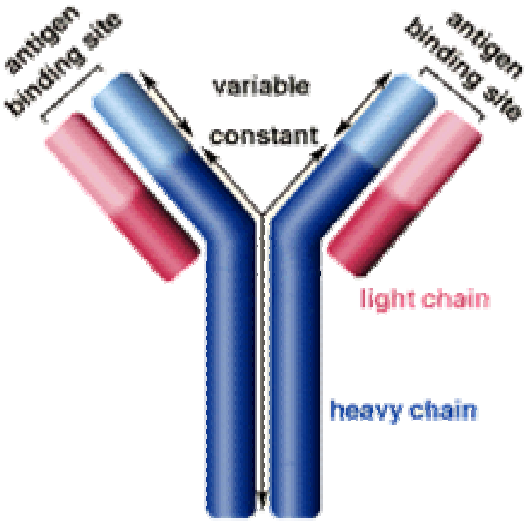


Figure 1.6 Basic structural domain of antibody (from reference 41).

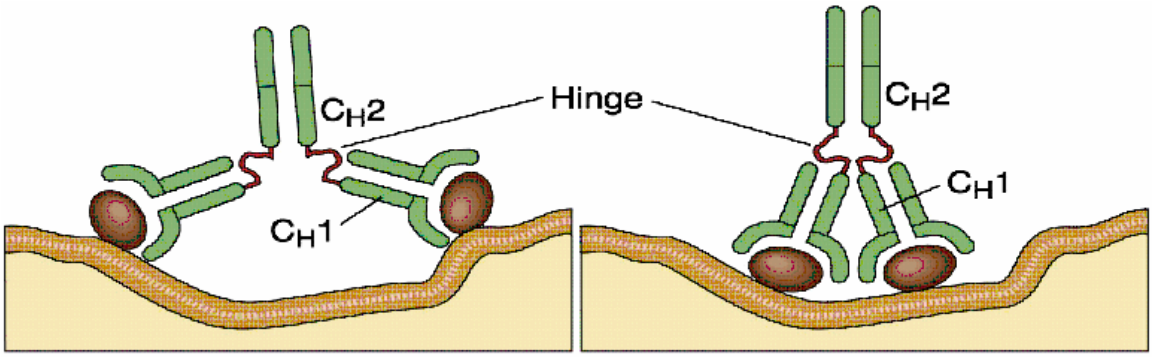


Figure 1.7 Flexibility of antibody hinge to bind at variable regions (from reference 41).

CHAPTER 2

MATERIALS AND METHODS

Materials

Bacto Tryptone and Bacto Yeast Extract were purchased from BD Biosciences; SDS-PAGE molecular weight marker was purchased from Bio-Rad; Restriction enzymes (*Nde I* and *Xho I*), Bovine Serum Albumin (BSA), DNA ladder (200bp and 1kbp), 100mM dNTP mix and Isopropyl-beta-D-thiogalactopyranoside (IPTG) were purchased from Promega; T4 DNA Ligase was purchased from Fisher Scientific; Clonase from Gateway; the pET-28 expression vector was from Novagen; pDEST-527, pDEST-544, pDEST-565, pDEST-566 expression vector from National Cancer Institute, Federick (MD); pDESTC1 from Dr. Peter Horanyi, pDor entry clone from from National Cancer Institute, Federick (MD), Bacteriological Agar was purchased from Oxoid; the Qiaquick PCR purification kit and the gel purification kit were purchased from Qiagen; Kanamycin and high pure plasmid isolation kit were purchased from Roche; *Pfu* Turbo DNA polymerase, Top 10 competent cells and BL21(DE3) GOLD® competent cells were from Stratagene.

Media were prepared as follows:

ZYP(925mL):10g tryptone,5g yeast extract

LB (1L): 10g tryptone, 5g yeast extract, 10g NaCl.

SOC (1L): 20g Tryptone, 5g Yeast Extract, 0.5g NaCl, 5g $MgSO_4.7H_2O$, 20 mM glucose.

PA-0.5G (50ml): 46.13 ml H_2O, 50 µl 1M $MgSO_4$, 5 µl 1000x metal mix, 0.63 ml 40% glucose, 2.5 ml 20x NPS, 0.2 ml Methionine 25 mg/ml, 0.5 ml 17 amino acids mix 10mg/ml, 1ml

100mg/ml Ampicillin.

1000x metal mix: 0.1 M $FeCl_3$-$6H_2O$, 1 M $CaCl_2$, 1 M $MnCl_2$-$4H_2O$, 1 M $ZnSO_4$-$7H_2O$, 0.2 M $CoCl_2$-$6H_2O$, 0.1 M $CuCl_2$-$2H_2O$, 0.2 M $NiCl_2$-$6H_2O$, 0.1 M Na_2MoO_4-$5H_2O$, 0.1 M Na_2SeO_3-$5H_2O$ and 0.1 M H_3BO_3. 20x NPS: 0.5 M $(NH_4)_2SO_4$, 1 M KH_2PO_4 and 1 M Na_2HPO_4

17 amino acids mix: 10mg/ml 17 commonly used amino acids except Cys, Tyr and Met.

PASM-5052 (1L): 900ml H_2O, 1.25 ml 1M $MgSO_4$, l.25ml 1000x Metal Mix, 20ml 50x 5052, 50ml 20x NPS, 1ml 100 μM vitamin B12, 5ml 10mg/ml 17 amino acids mix, 400μl Methionine 25mg/ml, 5ml 25mg/ml Se-Methionine, 1ml 100mg/ml Ampicillin.

50x5052: 0.5% glycerol, 0.05% glucose and 0.2% α-lactose

Zyp-5052(50mL):46.4 mLZYP,0.05 mL 1M $MgSO_4$, 0.05mL 1000x Metal Mix,1ml 50x 5052, 2.5ml 20x NPS,0.05 mL streptomycin 50mg/mL.

ZYP-5052 (1L): 928mLZYP,1mL 1M $MgSO_4$, lmL 1000x Metal Mix,20ml 50x 5052, 50ml 20x NPS,1mL streptomycin 50mg/mL

Pyrococcus furiosus genomic DNA was provided by Dr. Michael Adams, University of Georgia. Antibodies: Polyclonal antibodies from two different rabbits (subject 85 and 86) provided by the animal facility in Life Science Building, University of Georgia. Goat antirabbits and Nop 56/58 antibodies was provided by Dr. Michael Terns, University of Georgia. Horseradish peroxidase substrate from Amersham Biosciences was provided by Dr. Michael Terns, University of Georgia.

Methods

Sequence Analysis

Using *Pf* -1955 sequence as a query, PSI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/) homology searches

were carried out. The multiple sequence alignment of *Pf* -1955, Hfq and Sm/Lsm sequences was manually adjusted based on the alignment previously presented (32). Also the sequence identities of *Pf* -1955 with Hfq and Sm/Lsm proteins were calculated by program 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/) using *Pf* -1955 sequence as a query.

Cloning and Expression of Recombinant *Pf* -1955

Primers were designed for the *Pf* -1955 gene in *P. furiosus*. A *Nde I* site (underlined) was placed in the forward primer (5′-GTTCTT*CATATG*GGTATGGAAAG CCTACTTG – 3′), and a *Xho I* site (underlined) was placed in the reverse primer (5′- GTTCTT*CTCGAG*CTACT CTAGTAGCATTATCCAGT-3′). Primers were synthesized and purified by Integrated DNA Technologies (IDT). PCR reaction was carried out using a Peltier Thermal Cycler (MJ Research) to amplify the 216-bp gene from the *P. furiosus* genomic DNA. The amplification protocol was as follows: incubation for 4 minutes at 94°C, followed by 35 cycles of the following; 30 seconds at 94°C, 30 seconds at 48°C and 30 seconds at 72°C, and finished by 10 minutes at 72°C. The PCR product was visualized by electrophoresis on a 1% agarose gel using TAE (Tris-Acetate-EDTA) buffer and ethidium bromide, followed by purification using a Qiagen PCR purification kit. The pET-28 expression vector (Novagen) contains the following elements: a regular promoter/operator, consisting of the *E. coli* phage T7 promoter and one *lac* operator sequence; a ribosome binding site, a translation start codon; a 6xHis affinity tag coding sequence; a thrombin recognition site coding sequence; a T7 tag coding sequence; a multi-cloning site; a second 6xHis affinity tag coding sequence; translation stop codon and a transcriptional terminator. The PCR product and pET-28 expression vector were then digested by another endonuclease restriction enzyme *Nde I* (Promega) at 37°C for 48 hours, followed by *Xho I* (Promega) for 4 hours. In

another trial the PCR product and pET-28 vector was digested with *Nde I* (Promega) at 37°C and *Xho I* (Promega) for 2 hours. The resulting fragments were loaded onto 1% agarose gel stained by ethidium bromide, and purified using Qiagen gel purification kit. The ligation reaction of the digested pET-28 vector and *Pf*-1955 DNA fragment was carried out at room temperature and kept over night using T4 DNA Ligase (Fisher). The whole ligation product was then transformed into BL21 (DE3)® Competent cells (Statagene). Several trials of above procedure did not yield positive transformants.

Since the above protocol did not succeed, gateway cloning procedure was attempted. Primers were designed for the *Pf* -1955 gene in *P. furiosus* by placing the forward primer (5-GAAAACCTGTACTTCCAAGGCGGCTCTGGCATGGGTATGGA AAGCC TACTT-3) and the reverse primer (5-GGGGACCACTTTGTACAAGAAAGCT GGGTATTATCTAGTAGCATTATCCAGTTTA-3). PCR reaction was carried out using a Peltier Thermal Cycler (MJ Research) to amplify the 216-bp gene from the *P. furiosus* genomic DNA. The amplification protocol was as follows: incubation for 5 minutes at 95°C, 5 cycles followed by the 25 cycles of the following; 15 seconds at 95°C, 30 seconds at 60°C and 30 seconds at 68°C. The PCR product was visualized by electrophoresis on a 1% agarose gel using TAE buffer and ethidium bromide, followed by purification using a Qiagen kit. The LR reaction was carried out using pDNOR vector and the PCR product for 48 hours (entry clone). The whole ligation product was then transformed using one shot top 10 Component cells. The transformation was plated onto an LB agar plate containing 50μg/ml Kanamycin. Three transformants were selected, and the insertion was confirmed by PCR amplification. Plasmid from one transformant was extracted using the High Pure Plasmid Isolation kit, and sent to the Sequencing and Synthesis Facility. The plasmid with the correct insertion was ligated with

pDEST527, pSUMO, pDEST544, pDEST565, pDEST566 expression vectors (Figure 2.1). The transformation was performed in *E.Coli* BL21 (DE3)® Competent cells (Statagene). The transformation was plated onto an LB agar plate containing 50μg/ml ampicillin, and the plate was incubated at 37°C for overnight. Transformants from each of the above vectors were grown in LB media with IPTG induction to test for the expression and solubility of recombinant protein.

The recombinant protein *Pf* -1955 obtained from the four different expression vectors was found to be completely insoluble; however, the protein procured from pDEST566 was found to be partially soluble. The partially soluble clone was further tried for purification but the protein precipitated after the removal of tag.

The entry clone from above was ligated using clonase enzyme into the pDEST-C1 expression vector and then transformed in to EB5Alpha cells (Edge Biosystems, Gaithesburg, MD). The plasmid was isolated by mini-prep. The pDEST-C1-*Pf*1955 plasmid was then transformed into BL21 (DE3) cells (Edge Biosystems). This was the expression strain used for protein synthesis. The map of the pDEST-C1 plasmid is given below

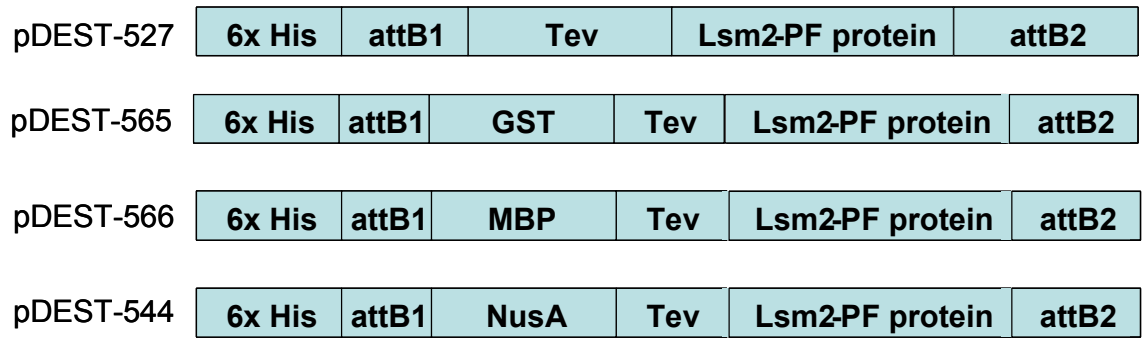


Figure 2.1 Parallel cloning of *Pf*1955 using different expression vectors pDEST-527, pDEST-565, pDEST566, and pDEST-544.

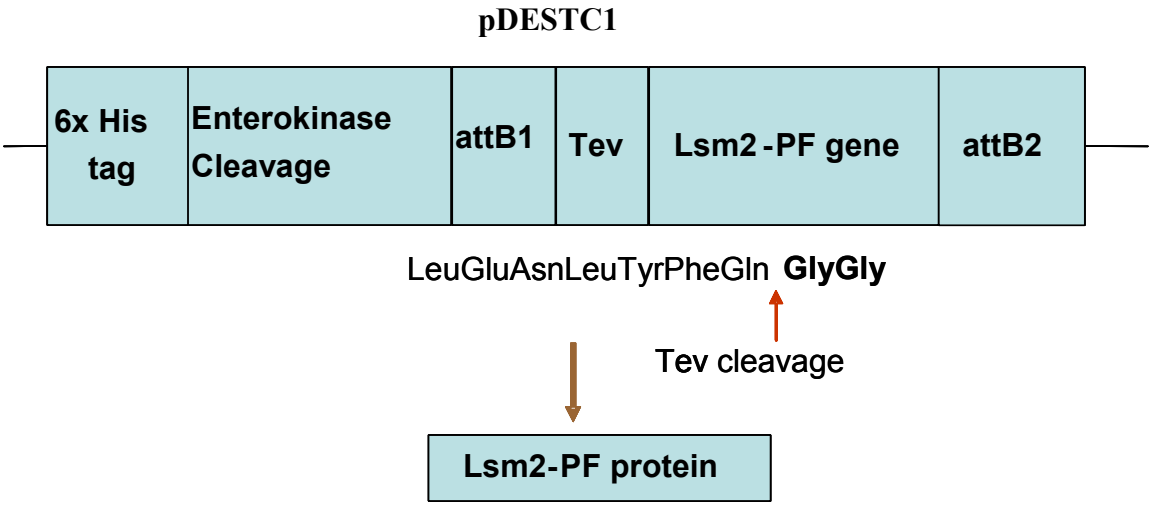


Figure 2.2 Purification scheme. Lsm2-PF gene was cloned into the multiple cloning sites pDEST C1 vector. Lsm2-PF was expressed in *E.Coli* BL21 (DE3). The His-tag was removed by Tev cleavage after Ni-affinity purification. The final purified Lsm-PF protein contains two glycine residues at the N-terminal.

.

Purification of Recombinant *Pf* -1955

Cells containing the recombinant *Pf* -1955 construct were grown in 5 ml LB medium containing 50µg/ml Streptomycin. After 8 hours of incubation at 37°C with shaking at 200 rpm, the culture was used to inoculate 500 mL of ZYP5052 medium. The 500 mL culture was incubated at 37°C on a shaker for 18 hours. Cells were harvested by centrifugation at 7,500 rpm for 15 min. The pellet (~ 2.5 grams) was resuspended thoroughly in 25 ml working buffer (100mM 100mM Hepes pH 7.6). Also 25 µL of 14 mM beta-mercapto ethanol and 250 µl of 1 mM PMSF (phenyl methyl sulfonyl fluoride) of 0.1 M stock solution were added to the cell solution to decrease the protease activity. The solution was then sonicated on ice using a Branson Sonifier Cell Disruptor 450 at power level 7 with 6 bursts of 30 seconds on and 30 seconds off. The resulting lysate was centrifuged at 12,000 rpm for 30 min. The supernatant was transferred to another centrifuge tube and incubated in a Fisher ISOTEMP 228 water bath at 95°C for 1 hour. The supernatant was again collected by centrifugation at 12,000 rpm for 30 min, and then was put on ice ready for affinity chromatography.

A 5 ml Hitrap affinity column from Amersham Biosciences was charged with Ni^{2+} by first washing the matrix with 2 bed volumes of 1 M NaOH followed by 10 bed volumes of water, then washing with 2 bed volumes of 0.5 M EDTA followed by 10 bed volumes of water, and charging the matrix with 2 bed volumes of 0.1 M $NiCl_2$ followed by 10 bed volumes of water wash, and finally equilibrated with 5 bed volumes of sample loading buffer, containing 100 mM Hepes pH 7.6. Using the AKTA Prime chromatography system from Amersham Biosciences, the supernatant was loaded onto the 5 ml Ni - NTA Agarose column previously equilibrated. The column was eluted with a linear gradient from 0 mM to 400 mM Imidazole. Fractions were collected and analyzed by SDS-PAGE. The recombinant *Pf*-1955with His-tag was sent to the

UGA Chemical and Biological Science Mass Spectrometry Facility to confirm the amino acid sequence.

Preparation of Sample for MALDI Analysis

Band Preparation and Destaining

The band from the above gel run was cut in to 2 x 2 mm pieces and rehydrated in 100 µl 50mM ammonium bicarbonate at 37 C for 10 mins with continuous shaking. The destaining solution from the tube was removed and discarded. The solution was dehydrated in 200 µl 50% acetonitrile in 50mM ammonium bicarbonate for 15 mins. at 37 C with shaking. The acetonitrile was removed and the steps repeated 2 to 4 two additional times.

Tripsin Digestion

The gel pieces from above were dried in a vacuum centrifuge for 15 mins. in absence of heat.The dried pieces were resuspend in 15 µl of 10 ng/µl tripsin followed by incubation of the sample at 37 C overnight with shaking. The peptides were extracted with 3 x 15 mins. washing (1 x with 15 µl 50mM ammonium bicarbonate, 2 x with 15µl 75% acetonitrile, 0.5% TFA). The supernatant from each wash was saved. All the supernatant from the wash was removed and dried in a vacuum centrifuge for 10 mins. with out heat until the volume was reduced to 4-5µl. The sample was further submitted for the MALDI analysis.

TEV Cleavage of His-Tagged Protein

After the Ni-NTA purification, the recombinant protein was dialyzed against buffer containing 100 mM Hepes with pH 7.6 overnight to decrease imidazole concentration. 100 units

of TEV protease were added to every 20 mg of protein solution to the final volume of 2~3 ml. The TEV digestion reaction was carried out at room temperature for 24 hours. The digestion process was monitored by taking 10 µl aliquots, running the sample on an SDS-PAGE gel, and looking for the band shift. The undigested protein with His-tag was separated from the digested protein by running the whole mixture though another Ni-NTA column under the same condition as the first time, and the flow through was collected.

The recombinant protein obtained from the second NTA was loaded onto Hiload 16/60 Superdex 75 gel filtration column from Amersham Biosciences. Elution was carried out using 100mM Hepes, pH 7.6, at 1min/ml. The protein was collected and concentrated for future use.

SDS Polyacylamide Gel Electrophoresis

The 10-well precast 4-20% iGel was purchased from Bio-rad and the 26-well 10-20% Criterion® Precast gel was purchased from Bio-Rad. Protein samples were mixed with Laemmli sample loading buffer containing 62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol and 0.01% Bromophenol Blue in a 1:1.5 ratio, and boiled in water for 5 min. Samples were then loaded onto the gel of choice, and the SDS Polyacrylamide electrophoresis was run in buffer containing 25 mM Tris at pH 8.3, 192 mM glycine and 0.1% SDS at 150 Volt for 60 minutes. The gel was stained with Coomassie G-250 for about 20 min, and then destained in buffer containing 5% ethanol and 7% acetic acid until the background was clear.

Purification of Recombinant *Pf* -1955 (pDEST527 vector) from Inclusion Bodies

Cells containing the recombinant *Pf* -1955 construct were grown in 5 ml LB medium containing 50µg/ml Ampicillin. After 8 hours of incubation at 37°C with shaking at 200 rpm, the

culture was used to inoculate 500 mL of ZYP5052 medium. The 500 mL culture was incubated at 37°C on a shaker for 18 hours. Cells were harvested by centrifugation at 7,500 rpm for 15 min. The pellet (~ 2.5 grams) was resuspended thoroughly in 25 ml working buffer (100mM 100mM Hepes pH 7.6). The pellet (~6 grams) was resuspended in 50 mL of cell lytic reagent containing DNAse and RNAse. The pellet was shaken for 25 mins at room temperature and centrifuged at 12,000 rpm for 30 mins. The supernatant solution was discarded and the pellet was redissolved 50 mL of cell lytic reagent containing DNAse and RNAse and the above procedure repeated. The pellet was dissolved in 10 mL of 6 molar guanidine chloride at pH 8.0 and the solution was shaken at room temperature overnight. Following day the solution was centrifuged at 12,000 rpm for 30 mins. and the supernatant was saved. The protein concentration was determined to be 16 milligrams/mL.

Purification from the Inclusion Bodies:

The Ni column was equilibrated with 5 volumes of 40mM Tris-HCl, 6M Gu-HCl, 0.5M NaCl pH 8.0. 1mL of supernatant *Pfu*1955 protein was loaded on the column. The column was spinned in cold room at 4 ^{o}C overnight. The column was washed with 5 volumes of 40mM Tris-HCl, 6M Gu-HCl, 0.5M NaCl and washed further with 10 volumes of 0.1% SB3-14, 40mM Tris-HCl, 0.5M NaCl at pH 7.5 and the washing continued with 5 volumes of 20mM Tris-HCl, 5mM betaCD, 0.1M NaCl at pH 7.5. The column was again washed with 5 volumes of 20 mM Tris-HCl, 0.1M NaCl at pH 7.5

and eluted with 10mL of 20 mM Tris-HCl, 0.1M NaCl, 600mM imidazole at pH 7.5.

The elute was collected and dialyzed against 20mM Tris-HCl buffer pH 7.5. The fractions were further concentrated to 3 ml and further purified using Superdex-75 using buffer (20 mM Hepes,

0.1 mM NaCl pH 7.5). The collected fractions were further dialyzed using 20 mM Hepes, 0.1 mM NaCl pH 7.5 and analyzed using SDS.

Serum Storage:

Pure *Pfu*1955 (5.0 mgs) was submitted to the University of Georgia animal facility for receiving polyclonal antibodies from two different subjects (rabbits). The preimmune, bleed-1, bleed-2, bleed-3, bleed-4, and bleed-5 sera were subsequently collected from the UGA animal facility. On receiving bleeds, 0.1% sodium azide was added to prevent infections. 100 µl aliquots of the preimmune bleed and 200 µl aliquots of the rest of the bleeds were stored at -80 ^{o}C.

Western Blotting:

A western blot (also known as immunoblot) is a method to detect protein in a given sample or extract. The protein samples were separated on SDS polyacrylamide gels (4-20% TRIS-HCl from Biorad) and electrotransferred to nitrocellulose membranes using Towbin buffer (25mM TRIS, 0.192M glycine, 1.45mM SDS in 2L water, 800ml methanol diluted to total volume of 4L, pH 8.1–8.5), which was prechilled to 4 ^{o}C, at 100 V with 1.47 amperes for 1 hour. The blot was washed two times with distilled water and stained with ponceae stain (to ascertain that all the protein was transferred to the membrane). The blot was blocked with 5% skim milk in TBS buffer (20 mM TRIS, 150 mM NaCl pH 7.5) at 4 ^{o}C overnight and the milk was decanted the following day. Subsequently, the blot was incubated with primary antibodies (1:100 dilution) in 5% skim milk in TTBS (20 mM TRIS, 150 mM NaCl, 0.05% Tween 20 pH 7.5) for 1 hour at room temperature. The blot was washed with TTBS buffer for 20 mins. The blot was incubated with secondary antibodies (goat anti-rabbit antibodies) in 1:2500 dilution in 5% skim milk in

TTBS buffer at room temperature for 1 hour. Subsequently the blot was washed with TTBS for 20 mins. The protein detection was carried out by treating the blots with Western blot detecting reagents from Amersham Biosciences (ECL) and exposing them to the X-ray film in dark room.

Immunoprecipitation

Immunoprecipitation is one of the most useful and common immunochemical techniques and can be used for many purposes. Immunoprecipitation is a procedure by which peptides or proteins that react specifically with an antibody are removed from solution and examined for quantity or physical characteristics (molecular weight, isoelectric point, etc.). As usually practiced, the name of the procedure is a misnomer since removal of the antigen from solution does not depend upon the formation of an insoluble antibody-antigen complex. Rather, antibody-antigen complexes are removed from solution by addition of an insoluble form of an antibody binding protein.

A number of important characteristics of antigen can be determined readily by immunoprecipation techniques. These assays can determine: (a) the presence and quantity of antigen, (b) relative molecular weight of the polypeptide chain, (c) rate of synthesis or degradation, (d) presence of certain post-translational modifications and (e) interaction with proteins, nucleic acid or other ligands.

Protocol for Immunoprecipation:

Antibodies used in these experiments included polyclonal antibodies directed against the *Pfu*1955 protein. Antibodies (10 µl of serum) were coupled to preswollen protein A-Sepharose (15 µl of 80% slurry) beads (Sigma) by end-over-end rotation in 0.5 ml of IPP-150 buffer (20 mM TRIS-HCl, 150 mM NaCl, 1 mM EDTA, pH 7.5 0.1% Igepal, 1mM PMSF, 5mM

Benzamidine) at 4 ^{o}C for 12-16 hours. The beads were spun at 3000 rpm for 4 mins. The supernatant (10% saved and the rest discarded) was separated and the beads were washed three times with the IPP-150 buffer. The native *Pfu* lysate cell extract (1 mg) was coupled to the beads by end-over-end rotation at 4 ^{o}C for 12-16 hours. The beads were spun at 3000 rpm for 4 mins and the 10% supernatant was saved and the rest discarded. The beads were washed five times with IPP-150 buffer. The supernatant was discarded and the immunocomplexes were subsequently eluted from the beads in 20 µl of SDS page loading buffer after boiling for 8 mins at 100 ^{o}C. The samples were separated on SDS polyacrylamide (4-20% Tris-HCl) gels and then were transferred to nitrocellulose and incubated with primary and secondary antibodies for the western blotting procedure described above.

Chapter 3
RESULTS AND DISCUSSION

Sequence Analysis

The PSI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) search using the *Pf* -1955 sequence as a query failed to give a significant hit, partly because of *Pf* -1955's small molecular size and the relatively low sequence similarity with Sm/Lsm sequences. Further homology search using Superfamily (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/) (42) homology searching revealed that *Pf* -1955 belongs to the Sm and Sm-like protein family, which includes eukaryotic Sm/Lsm, archaeal Lsm and bacterial Hfq proteins. This result was also consistent with TIGR annotation (www.tigr.org). The multiple alignments of *Pf* -1955, Hfq and Sm/Lsm sequences were shown in Figure 3.1, where *Pf* -1955 displayed sequence homology to both Hfq and Lsm proteins. *Pf* -1955 was therefore identified as an Lsm2 type protein, which is denoted as Lsm2-PF in Figure 3.1. In addition, using the threading program 3D-PSSM (43), the sequence identity of *Pf* -1955 with Hfq and Sm/Lsm sequences could be calculated (Table 3.1).

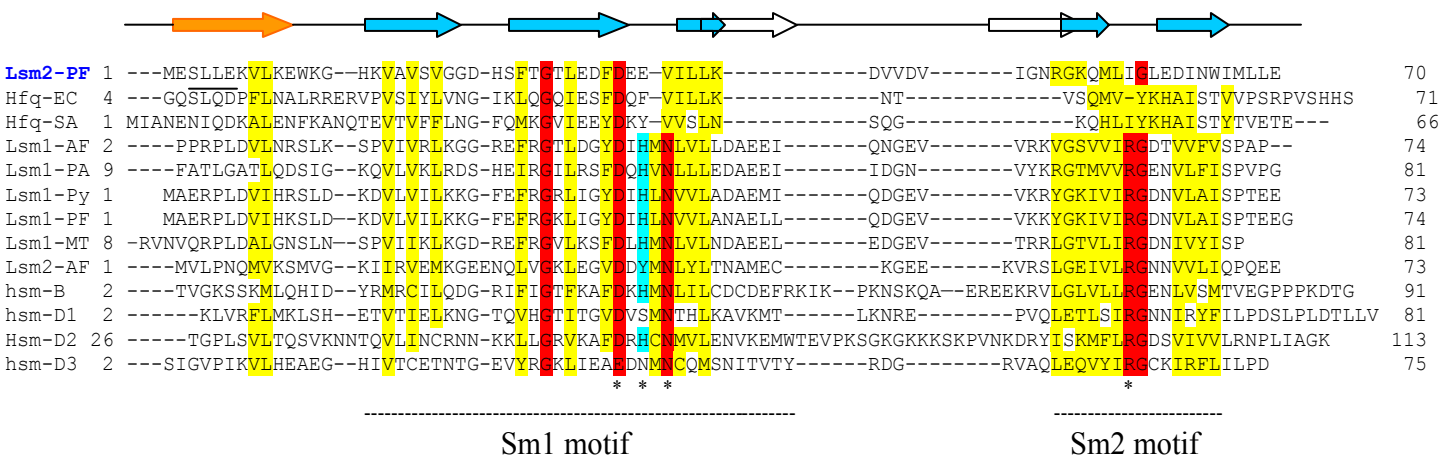


Figure 3.1 Sequence alignments of archaeal Sm-like proteins with human canonical Sm (hsm) proteins and bacterial Hfq proteins. Shown on top is the secondary structure assignment determined by X-ray analysis. Orange color arrow denotes α-helix, and blue color arrows denote β-strands. β1,2 and 3 strands constitute the first part of the Sm domain, and β4 and 5 strands form the second part. These two parts are linked together by a loop region (L4) with variable length and composition. Residues that almost fully conserved in all Sm/Lsm proteins are labeled in red, and residues highly conserved are in yellow. Residues in loop3 and 5 that involved in uridine binding are indicated by star. *Pyrococcus furiosus* Lsm2, also named as *Pf* -1955, was colored in blue. This figure was adapted from Sauter *et al*., *Nucleic Acids Research, 2003*, 31, 4091-4098.

Table 3.1 Statistics of current Sm, Lsm and Hfq crystal structures at PDB

	PDB ID	Oligomeric Formation	Dimension a/b/c (Å)	Angle α/β/γ (°)	Reso-lution	Space group	Sequence Identity*
Lsm1-AF	1I4K	heptamer	110.40/ 64.56/ 129.86	90.00/ 92.09/ 90.00	2.5	$P2_1$	17%
Lsm1-Py	1H64	heptamer	69.33/ 70.16/ 116.01	90.21/90.70/ 107.48	1.9	P1	17%
Lsm1-MT	1JBM	heptamer	45.07/ 54.08/ 62.35	87.58/ 72.86/ 81.45	1.85	P1	19%
Lsm1-PA	1I8F	heptamer	100.26/95.74/ 62.16	90.00/ 92.09/ 90.00	1.75	C2	30%
Lsm2-AF	1LJO	hexamer	58.42/ 58.42/ 32.08	90.00/90.00/ 120.00	1.95	P6	22%
Hfq-EC	1HK9	hexamer	61.35/ 61.35/ 166.10	90.00/90.00/ 120.00	2.15	$P6_1$	28%
Hfq-SA	1KQ1	hexamer	67.03/ 89.95/ 67.77	90.00/ 97.95/ 90.00	1.55	$P2_1$	27%
hsm-D3	1D3B	hexamer	107.35/108.45/ 110.42	90.00/ 90.00/ 90.00	2.00	$P2_12_12_1$	17%
hsm-B	1D3B	hexamer	107.35/108.45/ 110.42	90.00/ 90.00/ 90.00	2.00	$P2_12_12_1$	14%
hsm-D1	1B34	dimer	75.30/ 75.30/ 91.99	90.00/90.00/ 120.00	2.50	$P6_2$	22%
hsm-D2	1B34	dimer	75.30/ 75.30/ 91.99	90.00/90.00/ 120.00	2.50	$P6_2$	19%
Lsm2-PF	N/A	octamer	69.57/ 69.64/ 69.80	80.55/80.64/ 71.05	2.8	P1	
		N/A	82.96/82.96/ 189.785	90.00/90.00/ 120.00	2.7	$P6_222$/ $P6_422$	

*Sequence identity was calculated by the program 3D-PSSM using the Lsm2-PF sequence as a query.

Cloning and Expression of Recombinant *Pf*-1955

We were able to amplify the 216 bp gene from *Pyrococcus furiosus* genomic DNA using primers designed from the *Pyrococcus furiosus* ORF #1955DNA sequence. Attempts using the PCR product digested by *Nde I* and *Xho I*, and ligated into the pET-28 expression vector by T4 DNA ligase were unsuccessful. Hence we used the gateway cloning method. The PCR product (Figure 3.2) was recombined with pDOR vector followed by ligating with pDEST527, pSUMO, pDEST544, pDEST565, pDEST566 expression vectors (Figure 2.1). Of all the above vectors, *Pf*-1955 protein expressed using the pDEST566 was found to be only partially soluble (Figure 3.3). However, the protein was found to be insoluble (precipitate) after removing the maltose-binding-protein tag.

Ligating the entry clone in pDEST-C1 expression vector (Figure 2.2) the protein was found to be soluble. The further advantage of choosing pDEST-C1 vector as an expression vector is that it includes a TEV cleavage site followed by the N-terminal 6xHis tag coding sequence. Therefore, we were able to remove the His-tag once the protein was purified.

Purification of Recombinant *Pf* -1955

The results of the Ni-NTA column purification are shown in Figure 3.4. From the purification profile, it is evident that the protein height was 1800 mAu appeared during the elution using a 0-400 mM imidazole gradient. The 4-20% gradient SDS-PAGE showed that the recombinant protein was relatively pure, and the majority of the contamination was removed after the heat treatment. Also, a special protein band with the molecular weight of 60~70 kDa was shown always purified together with *Pf* -1955. Since the appearance of the band was dependent on both the concentrations of the pure *Pf* -1955 sample and the completeness of the

denaturation of the protein during SDS-PADE, the 60-70kDa band was considered to be a polymeric band of *Pf* -1955. We ran the Maldi-MS on the protein to confirm the sequence of the amino-acids. The molecular weight together with the PSI-BLAST (http://www.expassy.com) confirmed that protein purified using new vector, pDEST-C1, shared similar 56 % sequence coverage with that of *Pf* -1955.

TEV Cleavage of His-Tag Protein and Ni-affinity Purification

A high concentration of imidazole inhibits TEV protease, therefore buffer exchange was necessary to remove excess imidazole after Ni-affinity purification. For a more complete cleavage result, the protein was incubated at 25°C for 24 hrs, and no multiple digestions occurred (Figure 3.6). The *Pf* -1955, without the His-tag, does not have the ability to bind to a Ni-affinity column. Thus it was separated from the undigested protein by passage through the column.

Purification of *Pf* -1955 using Inclusion Bodies

Purification of *Pf*-1955 from inclusion bodies involves denaturation of the protein using Gu-HCl. The protein was washed with detergent to wash away Gu-HCl to retain its denatured state. The subsequent washing with cyclodextrin gets rid of the detergent buffer. The results of the Ni-NTA column purification showed a protein peak of ~ 1200 mAu appeared during the elution using a 20-600 mM imidazole gradient. The 4-20% gradient SDS-PAGE showed that the recombinant protein was impure, and hence passing through Superdex-75 column resulted in significantly pure protein (Figure 3.4). The subsequent attempts to remove the tag were unsuccessful and hence we ran the CD spectrum with the tag uncleaved. The attached CD spectrum clearly shows that both the proteins purified from inclusion bodies and obtained using pDEST-C1 have similar

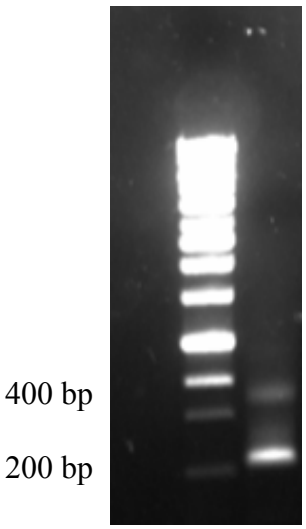


Figure 3.2 PCR-product of *Pf* -1955.

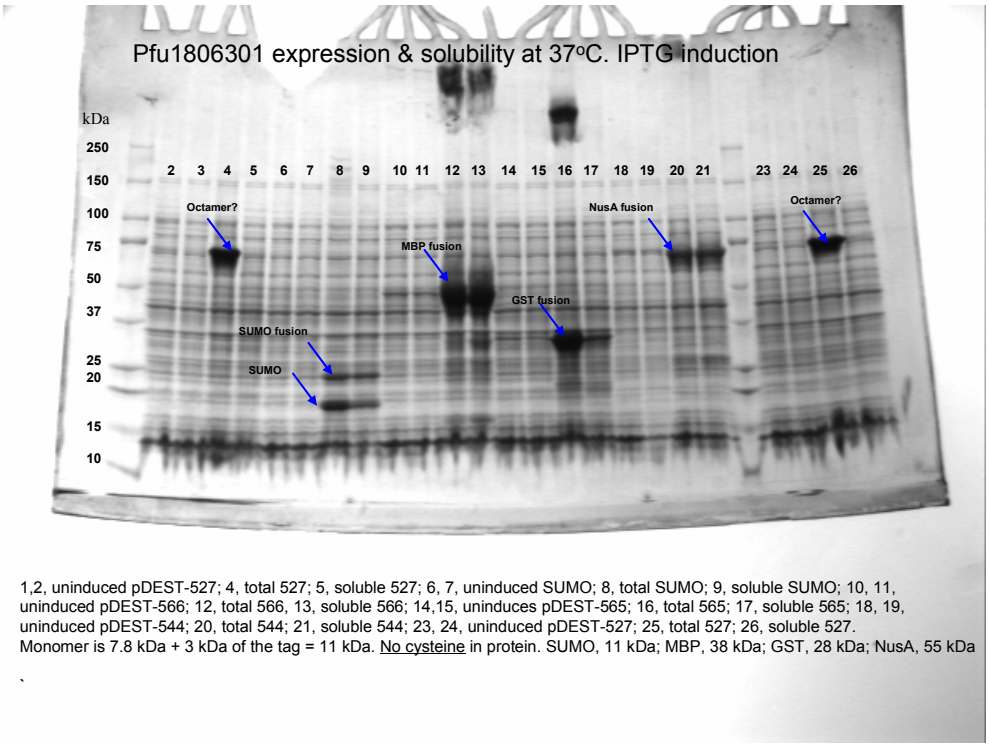


Figure 3.3 Expression of *PF*1955 using different expression vectors.

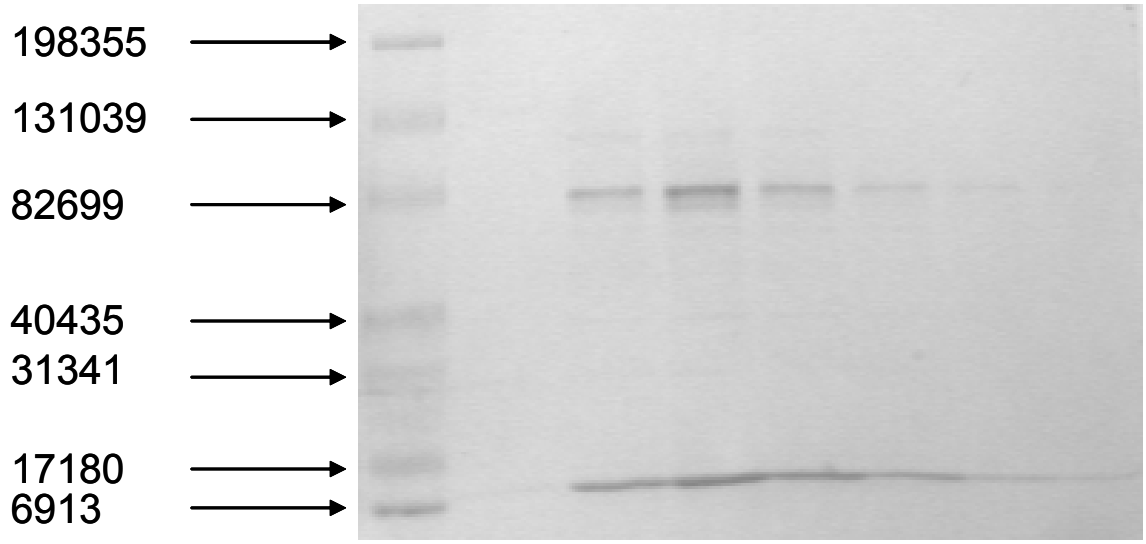


Figure 3.4 Superdex-75 gel filtration purification of *Pf* -1955 from the inclusion bodies with His-tag.

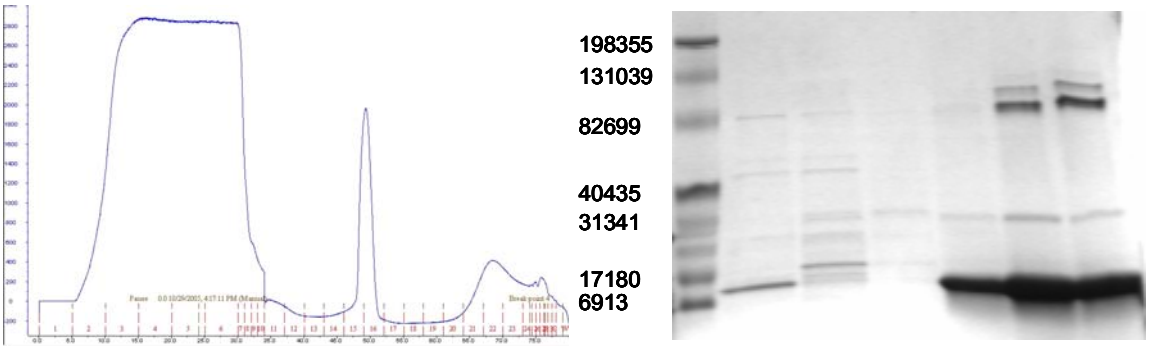


Figure 3.5 *Pf* -1955 was purified over Ni-NTA agarose. (A) The purification profile of the imidazole concentration is colored green and UV_{280} is colored blue. (B) SDS-PAGE analysis of selected fractions in the peak area.

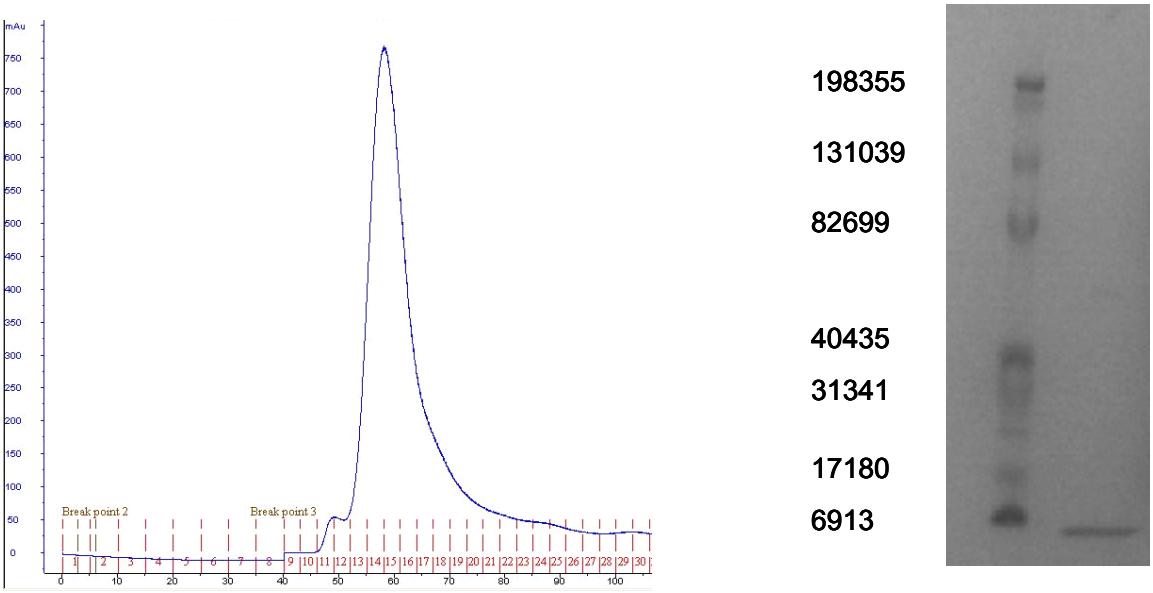


Figure 3.6 Superdex-75 gel filtration purification of *Pf* -1955 after the His-tag was removed. (A) Purification profile (B) SDS-PAGE shows the final purified protein without a His-tag, which has the MW of 6.8 kDa.

folding. The far-UV CD spectrum of the recovered protein shows the presence of helical structure with negative signals at 222 and 208 nm, and positive signal at 195 nm. Both these peaks (negative at 225 nm and positive at 190-200 nm) also are present in the soluble protein. Furthermore the recorded near-UV spectrum shows that both these proteins are stable up to high temperatures, 85 ^{o}C, a characteristic feature of hyperthermophiles. Also, a special protein band with the molecular weight of 60~70 kDa was shown always purified together with *Pf* -1955. Since the appearance of the band was dependent on both the concentrations of the pure *Pf* -1955 sample and the completeness of the denaturation of the protein during SDS-PADE, the 60-70kDa band was considered to be a polymeric band of *Pf* -1955.

The flow chart in Figure 3.9 shows the intended set of experiments proposed to verify if PF1955 binds to DNA, RNA or to neither of these. In the Western blot technique, it is verified that the antibodies are correct. Subsequently, in the column the antibodies are bound to the native *P. furiosus* cell extract. We expect that our specific (pfu) protein binds to our antibody with its substrate, be it nucleic acid or a protein forming a complex. The above acquired column complexed with antibodies to the natively purified PF1955 is split into two parts: **1.** All the proteins sequence will be verified by MS-analysis **2.** Isolate nucleic acids to perform RT-PCR and cloning to reveal the RNA/DNA sequence through sequencing at IBL here in campus.

The polyclonal antibodies directed against *Pf*-1955 are utilized for the flow-chart experiments. The Western blot is performed using the serums from the preimmune bleed and bleed 1 through 4 samples to confirm that the antibodies recognizes the recombinant *Pf*-1955. The *Pyrococcus furiosus* native cell extract was also used to ascertain if the antibodies recognize the protein *Pf*-1955. The absence of signals in the preimmune serum as shown in Figure 3.10

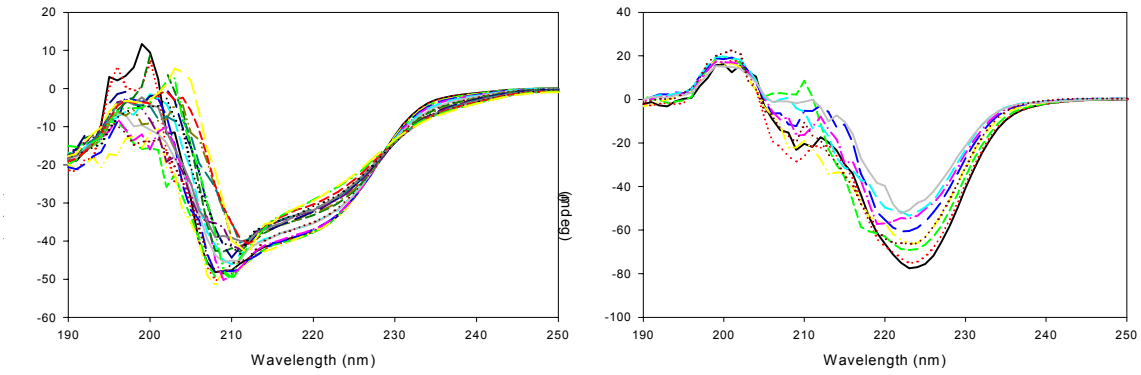


Figure 3.7 Far-UV CD spectra of *Pf*-1955. The left hand side spectrum is recorded for the protein purified from inclusion bodies while the one on right is recorded for the soluble protein purified from pDEST-C1.

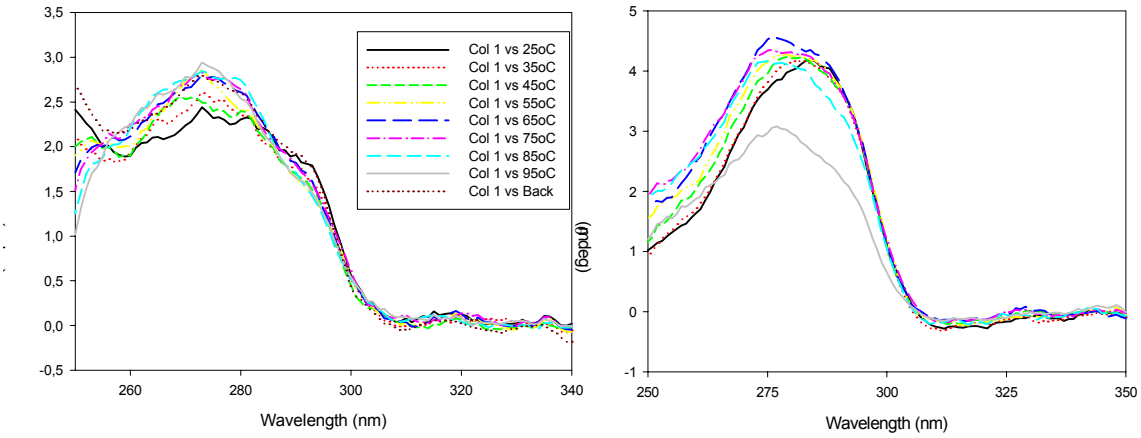


Figure 3.8 Near-UV CD spectra of *Pf*-1955. The left hand side spectrum is recorded for the protein purified from inclusion bodies while the one on right is recorded for the soluble protein purified from pDEST-C1.

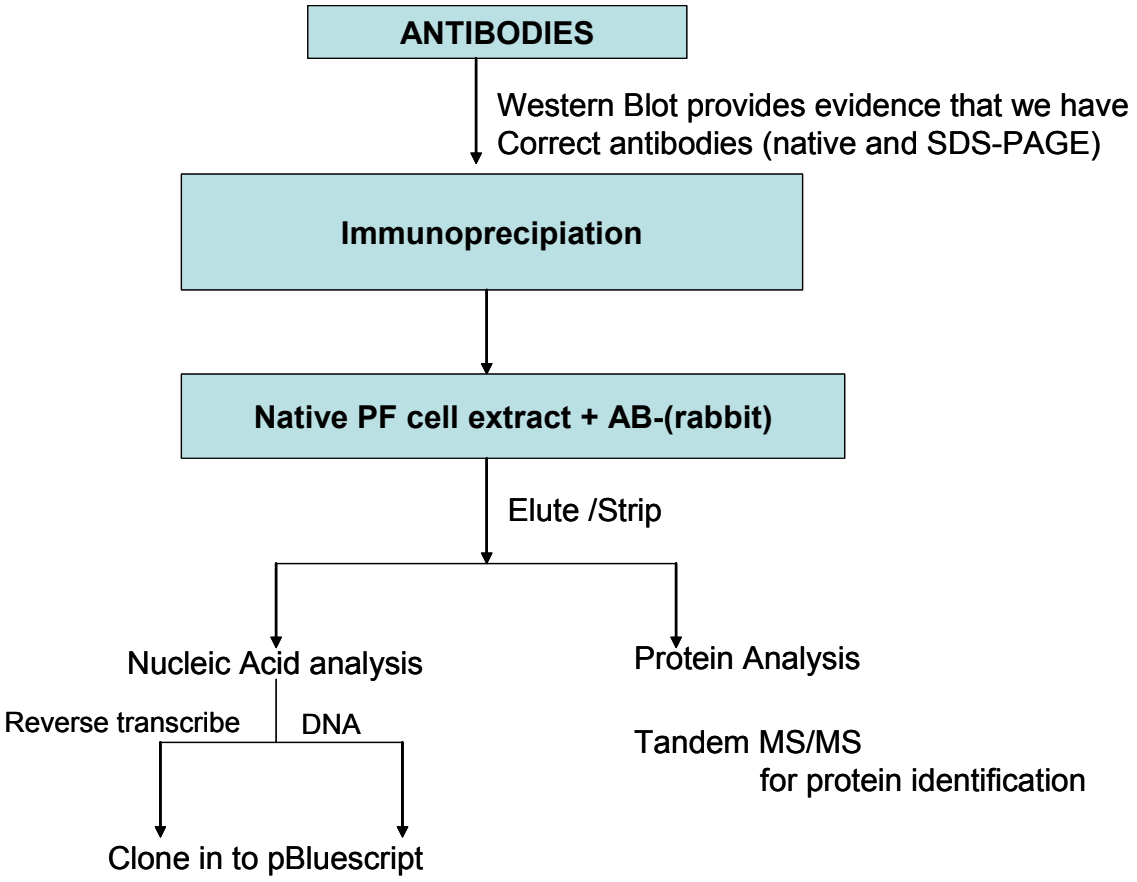


Figure 3.9 Flow chart indicates the proposed set of experiments to verify that *Pf*-1955 is DNA or RNA binding protein.

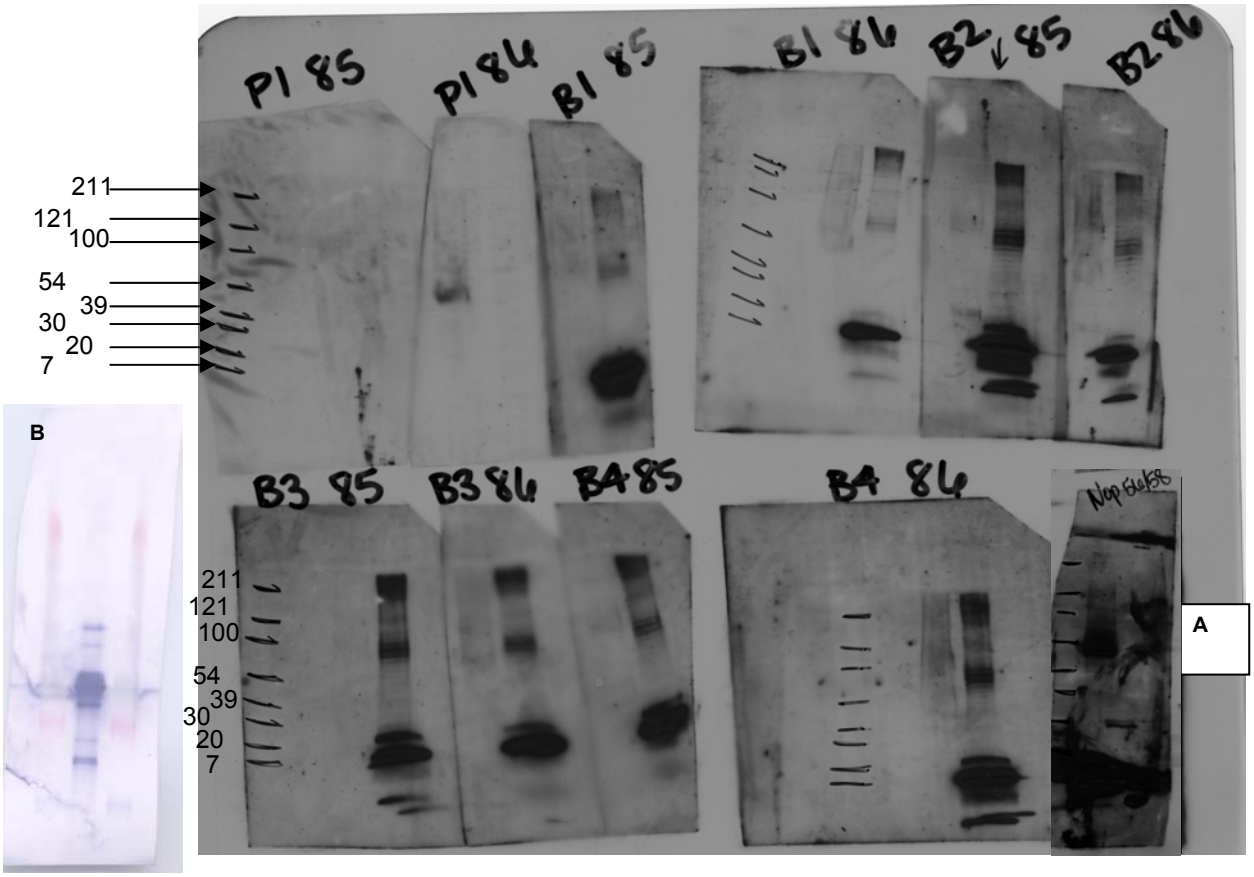


Figure 3.10 Western Blot (SDS page) of recombinant *Pf*-1955 (right) and the native *Pyrococcus furiosus* cell extract (central lane) probed with antibodies from preimmune serum (PI) and using the antibodies (against *Pf*-1955) from the blood serum 1 through 4 (B1, B2, B3, and B4) from two different rabbits (85 and 86). A is the control (Nop 56/58). Western Blot (of native gel) of recombinant *Pf*-1955 probed with B2 serum shown as **B**.

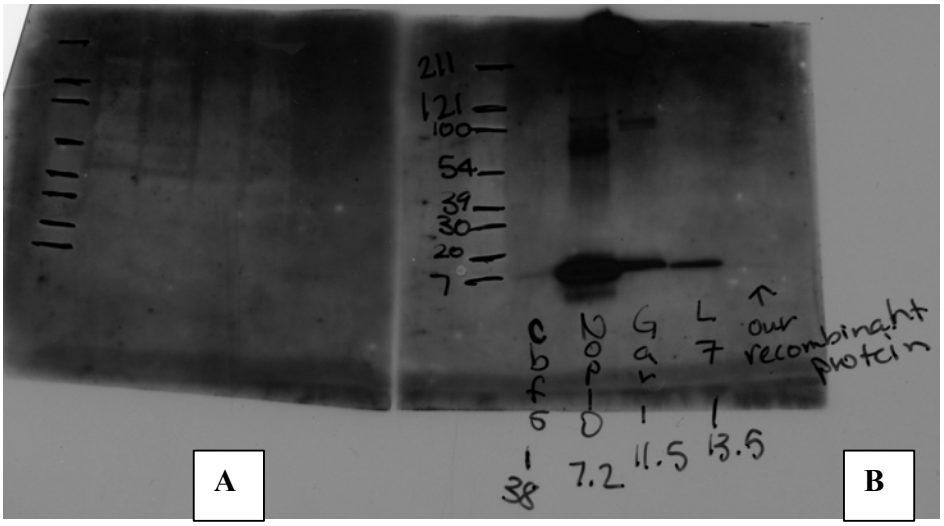


Figure 3.11 **A**. Western blot of Native cell extract at different concentrations **B**. Western blot with four different recombinant *Pyrococcus furiosus* proteins (Cbf-5, Nop-10, Gar1, and L7)

indicate that the "specific" antibodies were absent initially and were generated as an immune response after the rabbit was induced with the protein. In this context "specific" refers to antibodies that bind to *Pf*-1955. The presence of signals in bleed 1 through 4 (Figure 3.10) demonstrate that the recombinant *Pf*-1955 is recognized by the antibodies. The authenticity of the native cell extract is verified by running a positive control (A in Figure 3.10). Figure 3.10 indicates either a weak or no signal for the native cell-extract. This demonstrates that the *Pf*-1955 from the native cell extract is not recognized by the antibodies.

Since Figure 3.10 provided evidence that the *Pf*-1955 from the native cell extract is not recognized by the antibodies, it was necessary to ascertain if the antibodies were specific enough for *Pf*-1955. Hence subsequently we performed two set of experiments: **1**. Western blot using 0.3 μg of four different recombinant *Pyrococcus furiosus* proteins (Cbf-5, Nop-10, Gar1, and L7, **C** in Figure 3.11) and **2**. Western blot with different concentrations (25, 50, 75, and 100 μg) of native cell extract of *Pyrococcus furiosus* (**B** in Figure 3.11). The strong signals in Figure 3.11 **B** indicate that in addition to recognizing recombinant *Pf*-1955, antibodies also bind to three other recombinant proteins Nop-10, Gar1, and L7. However, the absence of signals for Cbf-5, show that this recombinant protein antibodies do not bind to this protein.

Figure 3.12 incorporates five different set of experiments of immunoprecipitation, **A** through **E**. The immunoprecipiation experiments were performed to pull-down a specific antigen from the native cell extract taking advantage of the antibody-antigen complexation characteristic. In experiment **A** antibodies from the preimmune bleed and native cell extract are bound to the protein A-Sepharose beads. The complex was denatured and separated on SDS/polyacrylamide and probed with antibodies from bleed-3. A Western blot with the preimmune bleed serves as a

negative control and should have been devoid of signals. Since this blot was developed using antibodies from bleed-3, the signals are prominently seen in this negative control. Experiment **B** is repetition of experiment **A** except that Bleed-3 was used instead of the preimmune bleed to bind to the native cell extract. Prominent signals are seen when Bleed-3 was used. However, absence of signals around 7000 Da molecular weight range confirms that the antibodies do not recognize the monomer unit of *Pf*-1955. Note that *Pf*-1955 is a dimer of octamers. Bleed-3 was utilized in Experiment **C** to couple to the recombinant antigen protein *Pf*-1955, instead of the native cell extract, in the protein A-Sepharose beads. A signal in the range of ca. 7000 Da molecular weight in Figure 3.12 **C**, demonstrates that the antibodies recognize the monomer unit from the recombinant *Pf*-1955 but does not recognize *Pf*-1955 protein present in the native cell extract. In experiment **D** antibodies specific to Nop-56/58 are bound to the native cell extract on the protein A-Sepharose beads. The complex were denatured and separated on SDS/polyacrylamide and probed with antibodies specific to Nop-56/58. The signals developed on blot, Figure 3.12 **D**, show that the antibodies specific to Nop-56/58 recognize Nop-56/58 protein in the native cell extract. The experiment **D** serves as a positive control. In experiment **E** antibodies from the Bleed-5 and the native cell extract are complexed on protein A-Sepharose beads and this complex was further denatured and separated on SDS/polyacrylamide and probed with antibodies from bleed-5. An absence of signal around 7 kDa indicates that the antibodies do not recognize the monomer *Pf*-1955 unit in the native cell extract. Note that the distinct signals for other molecular weights advocates that antibodies are nonspecific and bind to several *Pyrococcus furiosus* proteins present in the native cell extract.

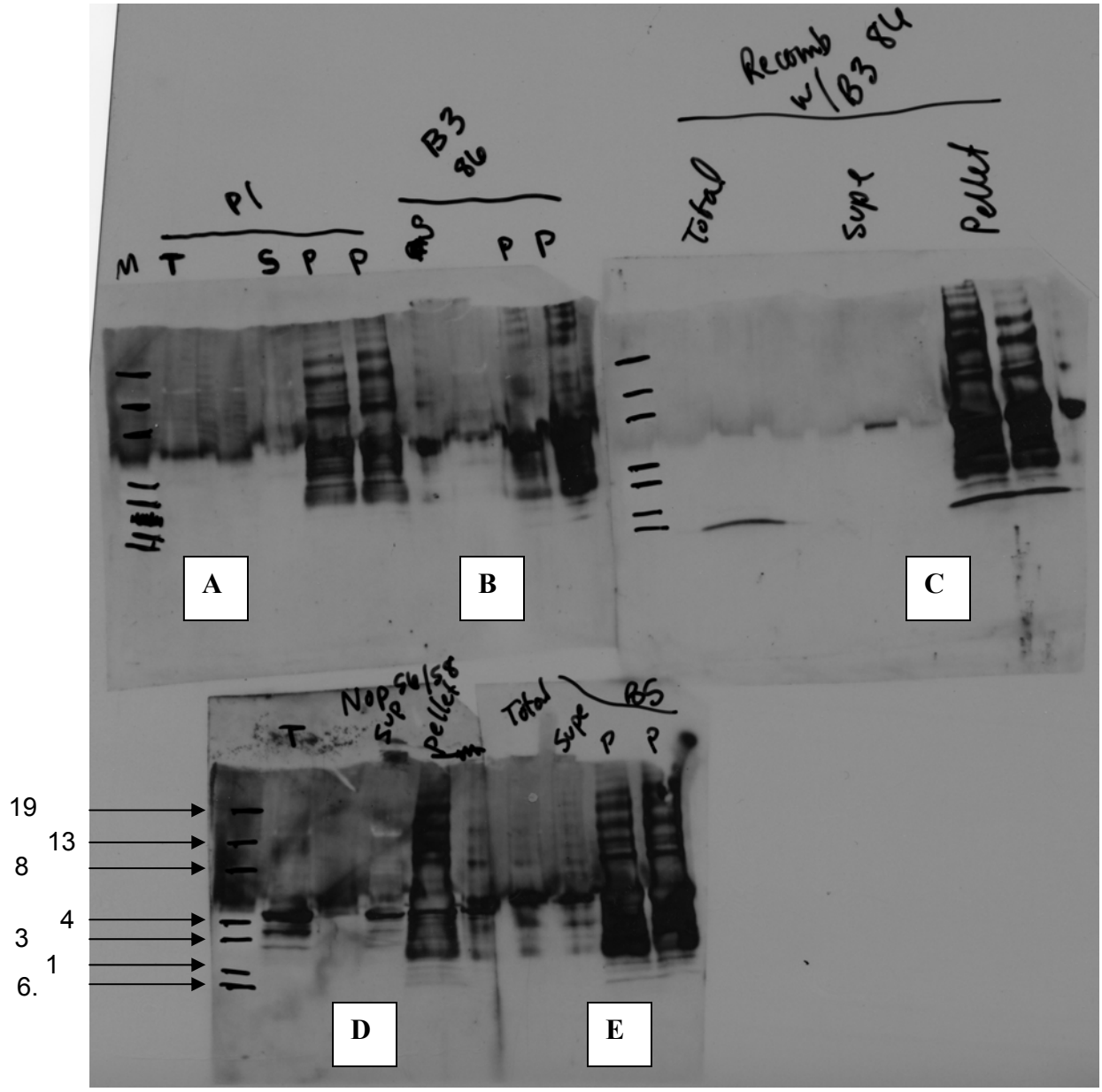


Figure 3.12 Western blots of immunoprecipiation (IP) of native *Pyrococcus furiosus* (PF) cell extract (Figure A, B, D, and E) and recombinant *Pf*-1955 (Figure C). (A) IP using native PF cell extract and preimmune blood serum and probed with bleed-3 serum. (B) IP using native PF cell extract and bleed-3 serum and probed with bleed-3 serum. (C) IP using recombinant *Pf*-1955 protein and bleed-3 serum and probed with bleed-3 serum. (D) IP using native PF cell extract and antibodies specific to Nop56/58 and probed with bleed-3 serum. (D) IP using native PF cell extract and antibodies from serum bleed-5 and probed with bleed-5 serum.

Discussion:

Sequence comparison and homology modeling showed that *Pf*-1955 shares with archaeal Lsm proteins the conserved pattern of hydrophobic residues in β1, β2 and β4, as well as highly conserved Gly29 and Asp35 residues. The analysis of the secondary structure of *Pf* -1955 also revealed sequence motifs consistent with the Sm-fold. However, *Pf* -1955 lacks the highly conserved residues His37, Asn39 and Arg63 that are crucial in uridine binding, but shares the sequence of VILLK at β3 with Hfq proteins. Nevertheless, *Pf* -1955 could not be characterized as an Hfq protein, since it lacks the highly conserved YKHAI sequence at the loop 5 region of all the known Hfq proteins. Notably only two types of Lsm proteins are present in each archaeal genome, and since Lsm1 type protein was already identified in *P. furiosus*, it is appropriate to consider *Pf* -1955 an Lsm2 type protein.

The crystal structure of recombinant *Pf* -1955 provides evidence that this protein is a dimer of octamers. The central cavity in the octamer is sufficiently large to accommodate a host RNA/DNA. Archaeal Lsm proteins show a clear preference for uridine-rich RNA. Furthermore, the 3-D structures of several archaeal Lsm proteins demonstrate the existence of a conserved uridine-binding pocket (31, 39, 40). However, *Pf* -1955 lacks the three conserved residues (His37, Asn39 and Arg63) that form the uridine-binding pocket. The protein itself may not have the ability to bind uridine-riched RNA, and it is also possible that the protein may bind other type of sequences. Hence, to determine the specific sequence to which *Pf* -1955 binds, we performed the immunoprecipiation by pull down assay.

In this thesis we have used a different procedure to clone and purify recombinant *Pf* -1955. Tev cleavage results in the retention of only two amino acids at the N-terminus of the

target protein. While the initial indications showed that the antibodies were specific, more detailed analysis of the antibodies demonstrated that these are not specific for *Pf* -1955 protein. Hence it is recommended to prepare the protein by a different method to immunize another set of rabbits. A simple gel purified protein following affinity purification may be sufficient since it would eliminate the protease digestion, the second Ni chromatography step and the gel filtration. The antibodies recognize a protein of the same size in samples that do not contain the protein (e.g. purified preps of other recombinant proteins) and also do not recognize any specific protein in Pf extract. The failure to succeed with the immunoprecipitation pull down assay might be due to several factors. It is possible that the antibodies are specific but the concentration of *Pf* -1955 in PF cell extract might be insignificant. The concentration may be beyond the detection levels of Western Blot experiments. Secondly, the SDS PAGE is transferred to the Western blot for detection with antibodies. However, SDS requires a boiling of the native PF cell extract. We had performed the boiling at 100 C for 8 minutes, while the PF proteins are stable to beyond 107 C. It is possible that boiling the cell extract at this temperature and for such a short time may not be sufficient to form monomers and denature the protein. Provided that the dimer of octamers *Pf* -1955 from the PF cell extract is not denatured, it may not bind to antibodies and consequently cannot be detected by the Western analysis.

CHAPTER 3

CRYSTALLIZATION TRIALS OF GP6

Introduction:

Bacillus subtilis phage Ø29 has a linear, double stranded genome of 19285 base-pairs (bp) with a terminal protein covalently linked to the 5` ends. Proteins involved in DNA replication (p1, p2, p3, p5 and p6) are transcribed by the host RNA polymerase from the left early operon together with proteins involved in the early late transcriptional switch (p4 and p6) (44). Therefore, protein p6 is involved both in DNA replication and transcriptional control (45). This pleiotropic effect is due to a more general feature: p6 is an architectural protein that organizes and compacts the viral genome, forming a nucleoprotein complex in which is the DNA is right handedly wrapped around a multimeric protein core.

The formation of the regular multimeric nucleoprotein complex in which the protein p6 repeated motif is found by protein dimer bound to a 24 bp DNA segment, with the centers of the two monomers binding sites located 12 bp apart. A protein monomer would contact and bend the DNA every ca. 12 bp, suggesting a model in which the DNA would wrap around a multimeric core of protein p6. The helical repeat of the DNA wrapped on the protein surface is 12 and therefore, larger than absolute helical repeat of the DNA a right handed super helix would be generated (46, 47).

Determination of the nature of the DNA signals recognized by the protein p6 is one of the key questions in addressing the formation of the nucleoprotein complex. Although recognition regions are multiple, the main ones were mapped between positions 62 and 125 from the Ø29

DNA right end, and between positions 46 and 68 from the left end (48). A computer search for nucleotide sequence homology between both Ø29 DNA terminal sequences did not indicate the existence of even a degenerate consensus sequence, suggesting that protein p6 does not recognize directly a specific sequence, but rather a sequence-dependent DNA structural feature. Protein p6-DNA complex formation involves strong DNA bending (49). The main protein p6 recognition regions contain sequences predicted to have bendable properties that favor a complex formation with DNA, which can bend every 12 bp (48).

Dimeric protein p6 binds with high affinity to DNA fragments containing tandem repeats of 24 bp-long sequence (protein p6 high-affinity binding unit) of the Ø29 DNA left terminus (50). As stated earlier, DNA wrapping, protein p6-DNA complexes show a considerable reduction in length with respect to the naked DNA.

An N-terminal deletion mutant composed of 5-amino acid deletions showed a reduced DNA binding affinity, and no activity was detected when the deletion was extended to 13 amino acids, suggesting that the N-terminal region of p6 could be involved in DNA binding (51). It also is speculated that protein p6 interacts with the DNA minor grove by hydrogen bonding and /or electrostatic interactions between amino acids of the polar side of the α-helix and the phosphate groups. Mutating the α-helix of p6 by replacing basic polar amino acids by alanine, to preserve the secondary structure, showed that both DNA binding and replication activation were impaired to different extents depending on the mutation. DNA binding was highly decreased for mutants p6K2A and p6K10A and completely prevented for mutant p6R6A. As this mutation does not affect the dimer formation, the arginine at position 6 plays a crucial role in DNA binding (52).

Protein p6 form dimer in solutions (45). At higher concentration, around 1 mM, p6 associates to oligomers. The shape of the oligomer species could not be rigorously determined

due to the complexity of the equilibrium mixture but demonstrated that p6 self-associates to elongated oligomers (53). The oligomers showed a helical structure and were found to interact head to tail either to form donut shaped structures or growing into right handed double helical filaments by a nucleation-dependent polymerization process (54).

As stated earlier, N-terminal deletion mutants of p6 lacking 5 or 13 amino acids were not able to interact with a Ø29 DNA terminal fragment (51). Since p6 binding to DNA is highly cooperative and strictly requires the formation of dimers, this lack of DNA binding could be due to a deficiency in protein self-association. p6NΔ5 and p6NΔ13 mutations of the protein not only lead to deficiency in dimer formation but also impaired the DNA binding. Mutant p6A44V, located at the center part of the protein, showed impaired dimer formation ability, and a reduced capacity to bind DNA and to activate the initiation of Ø29 DNA replication. More dramatic was the phenotype of mutant p6I8T, which had a self-association capacity reduced at least 10-fold and did not bind DNA nor activate Ø29 DNA initiation of replication. It was concluded that residues I8 and V44 are involved in protein dimer formation in vivo and that dimeric protein is the active form of the protein in vivo, required for viral DNA binding and replication (55).

X-ray structures of very few gene proteins, which bind to DNA, are known. Protein p6 initiates DNA replication and also functions as a transcriptional control protein (45). However, details of these processes are not known. In order to fully understand the function and structure of protein p6, it will be of great importance to determine its X-ray structure. The structure of p6 may also reveal several issues like amino acid residues involved in DNA binding and also the dimer formation.

Methods:

Purification of Protein p6

Cells containing the recombinant p6 construct were grown in 50ml LB medium containing 100μg/ml Ampicillin overnight. The culture was used to inoculate 1liter of LB medium and incubated at 37°C on a shaker until the OD_{600} was 0.6 - 0.8. The culture was then induced by adding IPTG to a final concentration of 1mM, and then incubated for 4 hours at 37°C with shaking. Cells were harvested by centrifugation at 7,000 rpm for 15 min. The pellet (~ 4 grams) was resuspended thoroughly in 30 ml working buffer (20mM Hepes, 20mM imidazole and 200 mM NaCl at pH 8.0). Also 14mM βME (β–mercapto ethanol) and 1mM of PMSF (phenyl methyl sulfonyl fluoride) of 0.1 M stock solution were added to the cell solution to increase the lysis efficiency and decrease the protease activity. The solution was then sonicated on ice using a Branson Sonifier Cell Disruptor 450 at power level 7 with 6 bursts of 30 seconds on and 30 seconds off. The resulting lysate was centrifuged at 12,000 rpm for 30 min.

A 5 ml Hitrap affinity column from Amersham Biosciences was charged with Ni^{2+} by first washing the matrix with 2 bed volumes of 1 M NaOH followed by 10 bed volumes of water, then washing with 2 bed volumes of 0.5 M EDTA followed by 10 bed volumes of water, and charging the matrix with 2 bed volumes of 0.1 M $NiCl_2$ followed by 10 bed volumes of water wash, and finally equilibrated with 5 bed volumes of sample loading buffer, containing 20mM Hepes, 20mM imidazole and 200 mM NaCl at pH 8.0. Using the AKTA Prime chromatography system from Amersham Biosciences, the supernatant was loaded onto the 5 ml Ni - NTA Agarose column previously equilibrated followed by elution with a linear gradient from 20 mM to 500 mM Imidazole in 100 ml. Fractions were collected and analyzed by SDS-PAGE.

Tev Cleavage of His-Tagged Protein

After the Ni-NTA purification, the recombinant protein was dialyzed against buffer containing 20 mM Hepes and 200 mM NaCl at pH 8.0 overnight to decrease imidazole concentration. 1ml of TEV protease were added to every 15 ml of protein p6 solution. The TEV digestion reaction was carried out at room temperature for 24 hours. The digestion process was monitored by taking 10 µl aliquots at various time intervals, running the sample on an SDS-PAGE gel, and looking for the band shift. The undigested protein with His-tag was separated from the digested protein by running the whole mixture though another Ni-NTA column under the same condition as the first time, and the flowthrough was collected.

The recombinant protein obtained from the second NTA was loaded onto Hiload 16/60 Superdex 75 gel filtration column from Amersham Biosciences. Elution was carried out using 20mM Hepes, 0.5 mM EDTA, 1mM DTT, 5% glycerol and 100mM NaCl, pH 8.0, at 1min/ml. The protein was collected and concentrated for future use.

SDS Polyacylamide Gel Electrophoresis

The 10-well precast 4-20% iGel was purchased from Gradipore and the 26-well 10-20% Criterion® Precast gel was purchased from Bio-Rad. Protein samples were mixed with Laemmli sample loading buffer containing 62.5 mM Tris-HCl, pH 6.8, 2% SDS, 25% glycerol and 0.01% Bromophenol Blue in a 1:2 ratio, and boiled in water for 10 min. Samples were then loaded onto the gel of choice, and the SDS Polyacrylamide electrophoresis was run in buffer containing 25 mM Tris at pH 8.3, 192 mM glycine and 0.1% SDS at 150 Volt for 60 minutes. After washing with H_2O for 5 min, the gel was stained with Coomassie G-250 for about 1 hour, and then destained in buffer containing 5% ethanol and 7% acetic acid until the background was clear.

Expression and Purification of Se-Met Protein

The plasmid containing the p6 gene was transformed into competent cells. The fresh transformants were inoculated into 50 ml of PA-0.5G minimal medium, and grown at 37°C for ~ 8 hrs. The culture was then transferred to 1 L PASM-5052 and incubated for ~ 20 hrs (A_{600} =1.37). After autoinduction, the cells were collected. The Se-Met derivatized protein was purified by a 3-step process as described before, Ni-affinity gradient purification followed by Tev cleavage of His-tag, and gel filtration purification.

Crystallization of Recombinant GP6

Initial crystallization screens were set up using modified microbatch methods by mixing 0.5 µl of protein solution with 0.5 µl of the crystallization reagent, covering the drop with mixed oil containing 70% of Paraffin and 30% Silicone oil, and incubating the crystallization tray at 18 °C. Using the Oryx 6 crystallization robot from Douglas Instruments, the native recombinant protein without the His-tag was first screened under six crystallization screens kits, using Hampton research's Crystal Screen 1 and 2, Crystal Screen Cryo (Cryo), Memfac, and PEG/Ion (PI). The screening was also tried using the Emerald Biosciences Wizard I and II and Molecular Dimension's Memsys. After several trials of optimization (done with modified microbatch PI-20, Cryo-6, Cryo-12, Cryo-29 and PI-7) under various conditions, the protein didn't crystallize.

Results and Discussion:

Purification of Recombinant p6

During the Ni-NTA column purification a protein peak of ~ 1200 mAu appeared during the elution using a 20-500 mM imidazole gradient. The 4-20% gradient SDS-PAGE showed that

the recombinant protein was relatively pure. A high concentration of imidazole inhibits Tev protease, therefore buffer exchange was necessary to remove excess imidazole after Ni-affinity purification. For a more complete cleavage result, protein was incubated at room temperature for 24 hrs, and no multiple digestions occurred. The p6, without the His-tag, does not have the ability to bind to a Ni-affinity column. Thus it was separated from the undigested protein by passage through the column followed by gel filtration. The molecular weight, shown in Figure 4.1, is approximately 12 kDa.

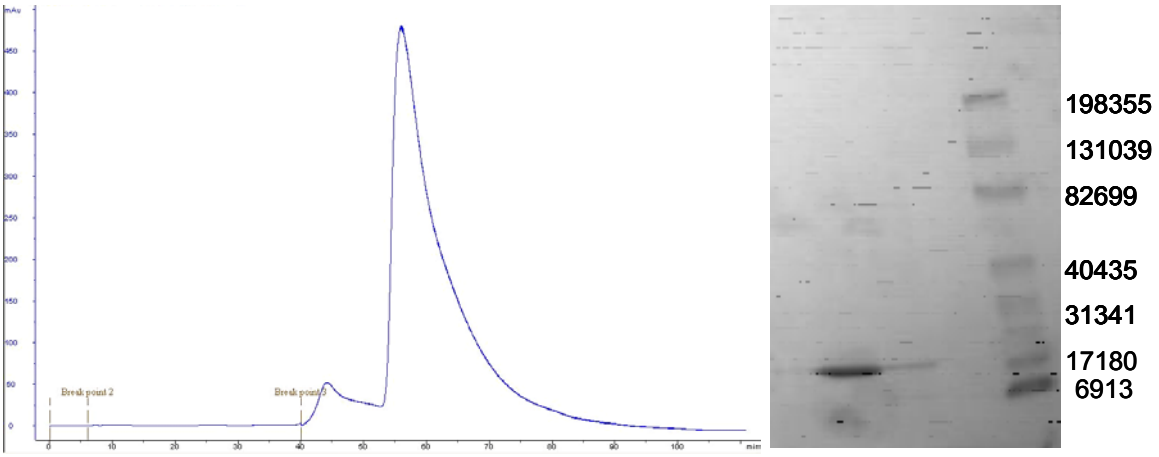


Figure 4.1 Superdex-75 gel filtration purification profile of protein p6 after the His-tag was removed. SDS-PAGE shows the final purified protein without a His-tag, which has the MW of 12 kDa.

CHAPTER 5

REFERENCES

(1) Kim, S. H. (1998) Shining the light on structural genomics. *Nature Struct Biol***. 5**, 643-645.

(2) Adams, M. W. W., Dailey, H. A., Delucas, L. J., Luo, M., Prestegard, J. H., Rose, J. P. and Wang, B. C. (2003) The southeast collaboratory for structural genomics: a high-throughput gene to structure factory. *Acc. Chem. Res*. **36**, 191-198

(3) Stetter, K. O. (1982) Ultrathin mycelia forming organisms from submarine volcanos areas having an optimum growth temperature of 105°C. *Nature*, **300**, 258-60.

(4a) Roopali, R. (2001) Tungsten-containing aldehyde oxidoreductases: a novel family of enzymes from hyperthermophilic archaea. Ph. D dissertation, University of Georgia

(4b) Adams, M. W. W. (1993) Enzymes and proteins from organisms that grow near and above 100 degrees C. *Annu Rev Microbiol*. **47**, 627-58.

(5) Woese, C.R., Kandler, O., Wheelis, M.L. (1990) Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87**, 4576-4579.

(6) Stetter, K. O. (1996). Hyperthermophilic prokaryotes. *FEMS Micro. Rev*. **18,** 149-158

(7) Magrum, L. J., Luehrsen, K. R., and Woese, C. R. (1978) Are extreme halophiles actually "bacteria"? *J. Mol. Evol*. **11**, 1-8.

(8) Burge, C. B., Tuschl, T. & Sharp, P. A. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), pp. 525-560, Cold Spring Harbor Press, Cold Spring Harbor, NY.

(9) Mattaj, I. W., Tollervey, D. and Seraphin,B. (1993) Small nuclear RNAs in messenger RNA and ribosomal RNA processing. *FESEB J*. **7**, 47-53.

(10) He,W. and Parker, R. (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Curr. Opin. Cell Biol*. **12**, 346-350.

(11) Kufel,J., Allmang,C., Petfalski,E., Beggs,J. and Tollervey,D. (2003) Lsm proteins are required for normal processing and stability of ribosomal RNAs. *J. Biol. Chem*. **278**, 1239-1247.

(12) Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. and Seraphin, B. (2000). A Sm-like protein complex that participate in mRNA degradation. *EMBO J*. **19**, 1661-1671.

(13)Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D. and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. *Nature*, **404**, 515-518.

(14) Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L. and Cech, T. R. (1999). Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle. *Science*, **401**, 177-180.

(15) Strub, K., Galli, G., Busslinger, M. and Birnstiel, M. L. (1984). The cDNA sequences of the sea urichin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. *EMBO J*. **3**, 2801-2807.

(16) Schaufele, F., Gilmartin, G. M., Bannwarth, W. and Birnstiel, M. L. (1986). Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of messenger-RNA. *Nature*, **323**, 777-781.

(17) Galli, G., Hofstetter, H., Stunnenberg, H. G. and Birnstiel, M. L. (1983). Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3' histone mRNA termini. *Cell*, **34**, 823-828.

(18) Will, C. L., Lührmann, R. (1997) snRNP structure and function. In *Eukaryotic mRNA Processing.* (Krainer, A. R.eds), pp.130-173, Oxford: IRL Press.

(19) Tan, E. M. (1989) Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv. Immunol*., **44**, 93-151.

(20) Lerner, M. R., and Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systematic lupus erythematosus. *Proc. Natl Acad. Sci. USA* **76**, 5495-5499.

(21) Lührmann, R., Kastner, B. and Bach, M. (1990) Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochem. Biophys. Acta*, **1087**,265-292.

(22) Cooper, M., Johnston, L. H. and Beggs, J. D. (1995). Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoprotiens. *EMBO J*., **14**, 2066-2075.

(23) Séraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO* J. **14**, 2089-2098.

(24) Hermann, H., Fabrzio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. and Lührmann, R. (1995) SnRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *EMBO J*. **14**, 2076-2088.

(25) Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-lewis., and Séraphin, B. (1999). Sm an Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J*. **18**, 3451-3462.

(26) Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. and Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4 and U5 RNAs. *EMBO J*. **1**, 1259-1265.

23. Kambach, C., Walke, S., Young, R., Avis, J. M., de laFortelle, E., Raker, V. A., Lührmann, R., Li, J., an Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* , **96**, 375-387.

(27) Mattaj, I. W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell,* **46**, 905-911.

(28) Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Lührmann, R. (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J*. **12**, 573-583.

(29) Kastner, B., Bach, M. and Lührmann, R. (1990). Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc. Natl Acad. Sci. USA*, **87**, 1710-1714.

(30) Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J*. **18**, 5789-5809.

(31) Törö, I., Basquin, J., Teo-Dreher, H. and Suck, D. (2002). Archaeal Sm proteins from heptameric and hexameric complex: crystal structures of the Sm1 and Sm2 proteins form the hyperthermophile *Archaeoglobus fulgidus*. *J. Mol. Biol.* **320,** 129-142

(32) Sauter, C., Basquin, J. and Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein form Escherichia coli. *Nucleic Acids Research*. **31**, 4091-4098.

(33) Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P. and Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and Hfq-RNA complex: a bacterial Sm-like protein. *The EMBO Journal*. **21**. 3546-3556.

(34) Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. w., Curmi, P. M. and Mabbutt, C. B. (2001). Crystal structure of a heptameric Sm-like protein complex from Archaea: implications for the structure and evolution of snRNPs. *J. Mol. Biol*. **309**, 915-923.

(35) Mura, C., Cascio, D., Sawaya, M. R. and Eisenberg, D. S. (2001). The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. *Proc. Natl Acad. Sci. USA*, **98,** 5532-5537.

(36) Törö, I., Thore, S., Mayer, C., Basquin, J., Séraphin, B. and Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J*. **20**, 2293-2303.

(37) Stark, H., Dube, P., Lührmann, R. and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particles. *Nature*, **409**, 539-542.

(38) Raker, V. A., Hartmuth, K., Kastner, B. and Lührmann, R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol*. **19**, 6554-6565.

(39) Thore, S., Mayer, C., Sauter, C., Weeks, S. and Suck, D. (2003) Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. *J. Biol. Chem.* **278**, 1239-1247.

(40) Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M. and Eisenberg, D. (2003) The oligomerization and lignad-binding properties of Sm-like archaeal proteins (SmAPs). *Protein Science*. **12**, 832-847.

(41) Harlow, E. and Lane, D. (1988) In *Antibodies* pp. 7-37, New York: Cold Spring Harbor Laboratory.

(42) Gough, J., Karplus, K., Hughey, R. and Clothia, C. (2001) Assignment of homology to genome sequences using a library of hidden markov models that represent all proteins of known structure. *J. Mol. Biol*., **313(4)**, 903-919.

(43) Kelley, L.A., MacCallum, R.M. and Sternberg M.J.E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol*. **299(2**), 499-520.

(44) Meijer, W.J.J., Horcajadas, J.A., and Salas M. (2001). Ø29 family of phages. *Microbiol. Mol. Biol. Rev.* **65**, 261–287.

(45) Pastrana, R., Lázaro, J.M., Blanco, L., García, J.A., Méndez, E. and Salas, M. (1985). Overproduction and purification of protein p6 of Bacillus subtilis phage Ø29: role in the initiation of DNA replication. *Nucl. Acids Res.* **13**, 3083–3100.

(46) White, J.H., Gallo, R.M. and Bauer, W.R. (1992). Closed circular DNA as a probe for protein-induced structural changes. *Trends Biochem. Sci.* **17**, 7–12.

(47) Serrano, M., Salas, M. and Hermoso, J.M. (1993). Multimeric complexes formed by DNA-binding proteins of low sequence-specificity. *Trends Biochem. Sci.* **18**, 202–206.

(48) Serrano, M., Gutiérrez, J., Prieto, I., Hermoso, J.M. and Salas M. (1989). Signals at the bacteriophage Ø29 DNA replication origins required for protein p6 binding and activity. *EMBO J.* **8**, 1879–1885.

(49) Satchwell, S.C., Drew, H.R. and Travers, A.A. (1986). Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659– 675.

(50) Serrano, M., Gutiérrez, C., Salas, M. and Hermoso, J.M. (1993). Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage Ø29 DNA replication. *J. Mol. Biol.* **230**, 248–259.

(51) Otero, M.J., Lázaro, J.M. and Salas, M. (1990). Deletions at the N-terminus of bacteriophage Ø29 protein p6: DNA binding and activity in Ø29 DNA replication. *Gene* **95**, 25–30.

(52) Freire, R., Salas, M. and Hermoso, J.M. (1994). A new protein domain for binding to DNA through the minor groove. *EMBO J.* **13**, 4353–4360.

(53) Abril, A.M., Salas, M., Andreu, J.M., Hermoso, J.M. and Rivas, G. (1997). Phage Ø29 protein p6 is in a monomer-dimer equilibrium that shifts to higher association states at the millimolar concentrations found in vivo. *Biochemistry* **36**, 11901– 11908.

(54) Abril, A.M., Marco, S., Carrascosa, J.L., Salas, M. and Hermoso, J.M. (1999). Oligomeric structures of the phage Ø29 histone-like protein p6. *J. Mol. Biol.* **292**, 581–588.

(55) Abril, A.M., Salas, M., and Hermoso, J.M. (2000). Identification of residues within two regions involved in self-association of viral histone- like protein p6 from phage Ø29. *J. Biol. Chem.* **275**, 26404–26410.