DNA BINDING PROTEINS OF ARCHAEAL VIRUSES

by

Joella Suzanne Geary

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2008

APPROVAL

of a thesis submitted by

Joella Suzanne Geary

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Mensur Dlakic

Approved for the Department Microbiology

Dr. Mike Fraklin

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Joella Suzanne Geary

November 2008

ACKNOWLEDGEMENTS

My sincere thanks to Mensur Dlakic, my professor and mentor, for his dedication to my success and for his invaluable help during this work; to my parents, Charles and Kathleen Geary, for their support and encouragement; to Tim, my husband, for pushing me forward and keeping the rest of life intact while I was working; to Cathy Castle, my lab mate, for being an example of patience and kindness and for all the pastries; and to Kari Cargill, my friend and advisor, for always sharing her time to answer my questions. I am indebted to each of you for the vital roll you have played.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

Archaea are single-celled organisms comprising the third domain of life. The Achaeal species *Sulfolobus* are infected by the *Fuselloviridae* virus family: SSV1, SSV2, SSV-RH, and SSV-K. The genomes of these viruses have been annotated and contain putative DNA-binding proteins. The purpose of this work is to identify DNA sequences bound by the SSV1 putative DNA-binding protein C43. C43 protein was cloned, expressed, purified, and assayed at various temperatures for interaction with three SSV1 DNA sequences. C43 binds the T5-promoter, T6-promoter, and C43-promoter sequentially and consistently. Additionally, C43 protein is functional at temperatures of 50ºC and 65ºC. Thus, C43 appears to be an important regulator of the *Fuselloviridae* SSV1 viral genome.

INTRODUCTION TO ARCHAEA

Archaea

Until the 1990's life was classified into six kingdoms. However, as technological advances enabled the exploration of life at a molecular and biochemical level it became apparent that life was more reasonably divided into three domains which encompassed the former six kingdoms (48). Distinctions between the domains of Bacteria, Eukarya, and Archaea are based on differences in genetic and molecular characteristics. On the cytological level, Archaea are prokaryotes but on the molecular level they are neither Bacteria nor Eukarya (48). Differences between Archaea and the other two domains can be seen in the 16S and 18S rRNA sequences (16), the hairpin loop of the small rRNA subunit (47), their membrane lipids (9), and the components of the translation machinery (16). Additionally, preference in codon usage appears to be correlated with domain (26).

Although Archaea arguably have a long and complex history, our understanding of this domain, in comparison to that of Eukarya and Bacteria, is much less developed. This circumstance is partially due to Archaea's relatively recent discovery (12), yielding the field time poor. Also, the difficulty in culturing extremophiles has hampered efforts to study these organisms in the laboratory. However, a strong beginning has been made in understanding the third domain of life. The Archaeal domain is comprised of prokaryotic, single-cell organisms much like Bacteria. They are inhabitants of almost every environment on Earth, including hydrothermal vents, the rumen, ocean waters, sewage, salt lakes, soda lakes, soil, and hot springs, to name a few (6). Although found

in all environments, these organisms dominate in hot acidic springs (28) and other extreme environments (7). Additionally, Archaea play important roles in numerous nutrient cycles such as the nitrogen cycle and ammonia oxidization (4, 5, 18, 19). The genomes of Archaea species appear to be relaxed or positively supercoiled; in contrast to Eukarya and Bacteria whose genomes are negatively supercoiled (11, 25). Also, many Archaea do not have a high G + C content as might be expected of organisms living in hyperthermal environments (14).

Many studies have probed the transcriptional and translational machinery utilized by the Archaea. These studies have found that in general the transcription machinery of Archaea tends to be homologous to that of eukarya (1, 2, 33, 45). The replication machinery also tends to be homologous to that of eukarya (10, 21). However, the translation machinery of Archaea tends to be homologous to that of bacteria (1, 45).

The Archaeal domain is currently divided into two phyla: Crenarchaeota and Euryarchaeota; although there are also two proposed phyla: Korarchaeota and Nanoarchaeota. The phyla of Archaea are based mainly on differences in rRNA sequences. The Crenarchaeota members are mostly thermophilic or hyperthermophilic organisms while the Euryarchaeota are generally methanogens. The Korarchaeota represent organisms that are different from the Crenarchaeota and Euryarchaeota in the 16S rRNA sequences. Finally, the Nanoarchaeota are nano-sized hyperthermophilic symbiotes (15).

The Crenarchaeota, although not obligate extremophiles, do well in environments characterized by temperatures above 75°C and pH of 4.0 or less. These organisms hold

particular interest for the biotechnical community because their bio-machinery maintains functionality at non-physiological conditions. Archaea possess great potential as reservoirs of novel technological advancements as well as holding significant academic interest.

The unique and global distribution of Archaea alludes to their enormous impact and influence on other organisms and in the transformation of the environment. Specifics of this impact are poorly understood and are undoubtedly as complex and diverse as the better understood influence of bacteria.

Sulfolobus Species

Within the Crenarchaeota, the *Sulfolobus* genus has immerged as a model organism because it is culturable in either liquid or solid media as single colonies or lawns. *Sulfolobus spp.* are aerobic and heterotrophic (7, 41), oxidize sulfide to elementary sulfur or hydrogen sulfide (3, 20, 41), while other strains oxidize ferrous iron (3).

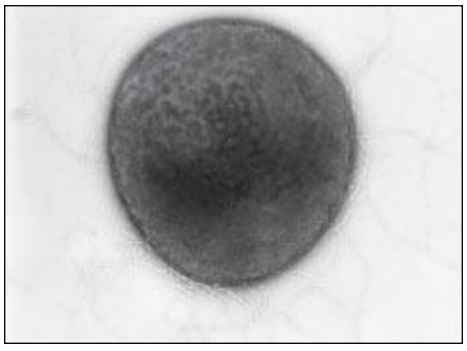


Figure 1: TEM of *S. solfataricus* (from reference 44)

Sulfolobus spp. can grow at temperatures of 60°C to 95°C and inhabit environments with a pH of 1 to 5. They are gram-negative, have pili and pseudopodia, can possess flagella, and appear as misshapen spheres (7, 27, 44) (Figure 1).

Sulfolobus was first described by Brock in 1978 and sequenced in 2001 (40). The genomes are about 3 x 10^6 base-pairs (bp) and contain about 2900 open-reading-frames (ORFs). They also contain many insertions and repetitive elements and short tandem repeats that are clustered into long stretches (7). *Sulfolobus spp.* also have promoters which contain TATA elements (35), and some species contain putative Shine-Dalgarno (SD) motifs (45). Gene transfer occurs via conjugation, transduction, and transformation (7). In addition to the circular genome, other genetic elements include cryptic and conjugative plasmids as well as viruses (41).

Crenarchaeota Viruses

In 1982 the first reports of viruses infecting Archaea were published (50). Since then, less than 100 Archaeal viruses have been identified. This number is in sharp contrast to the thousands of known bacterial and eukaryotic viruses. The Crenarchaeota are infected by a variety of viruses and to date approximately twenty-five have been identified. These viruses are summarized in Table 1 and have been divided into five families and two proposed families. Virus families are determined by virion morphotype, which is highly diverse (30), and the properties of their genomes. To date all virus genomes have been double-stranded DNA (dsDNA). Crenarchaeota viruses have been found in *Sulfolobus*, *Acidianus*, *Pyrobaculum*, and *Thermoproteus* (30, 31).

Table 1: Properties of Crenarchaeota Viruses. Adapted from (31).

Family	Known Species	Host	Morpho-Type	Lipids	Genome form	Genome Size (bp)	G+C Content %
Fuselloviridae	SSV1	*Sulfolobus*		nd		15465	40
	SSV2	*Sulfolobus*	Single-tailed spindles	nd	Circular	14796	39
	SSVK1	*Sulfolobus*		nd		17385	39
	SSVRH	*Sulfolobus*		nd		18473	39
Rudiviridae	SIRV1	*Sulfolobus*		-		32308	25
	SIRV2	*Sulfolobus*	Rigid rods	-	Linear	35450	25
	ARV1	*Acidianus*		-		24655	39
Lipothrixviridae Alpha	TTV1	*Thermoproteus*	Flexible filaments, diverse termini	+	Linear	16000[c]	
Lipothrixviridae Beta	SIFV	*Sulfolobus*		+		40852	33
Lipothrixviridae Beta	TTV2	*Thermoproteus*		nd		16000[c]	nd
Lipothrixviridae Beta	TTV3	*Thermoproteus*		nd		27000[c]	nd
Lipothrixviridae Gamma	AFV1	*Acidianus*		+		21000	37
Lipothrixviridae Delta	AFV2	*Acidianus*		-		31787	36
Globuloviridae[d]	PSV	*Pyrobaculum*	Spherical	+	Linear	28337	48
	TTSV1[b]	*Thermoproteus*				20933	50
Bicaudaviridae[d]	ATV	*Acidianus*	Two-tailed Spindle	nd	Circular	62730	41
Guttaviridae	SDNV	*Sulfolobus*	Droplet	nd	Unknown	nd	nd
Ampullaviridae[d]	ABV	*Acidianus*	Bottle	nd	Linear	nd	nd
Unclassified	STIV	*Sulfolobus*	Icosahe-dral with turrets	nd	Circular	17663	36
Unclassified	STSV1	*Sulfolobus*	Large Spindle	+	Circular	75294	35

aAbbreviations: J, Japan; I, Iceland; R, Kamchatka, Russia; Y, Yellowstone National Park, USA; P, Pozzuoli, Italy; K,
Korea; C, China;. K, negative; C, positive; nd, not determined.
bTTSV1 Sequence Accession Number AY722806.
cApproximate genome size.
dVirus families proposed but not yet approved by the International Committee for Taxonomy of Viruses.

Fuselloviridae: *Sulfolobus* Spindle-Shaped Viruses (SSVs)

Figure 2: Characteristic Spindle Shaped Virus Morphology (ref. 28)

Four non-lytic viruses isolated from the *Sulfolobus spp.* belong to the Fuselloviridae family (31). These viruses are SSV1 from Bappu, Japan, SSV2 from Reykjavik, Iceland, SSV-K1 from Kamchatka, Russia, and SSV-RH from Yellowstone National Park, USA (Figure 3). SSVs are 60 x 90 nm in size and spindle shaped with short tail fibers at one pole (31, 46) (Figure 2). These tail fibers are used to attach to the host membrane (31). They have circular dsDNA and exist as a prophage (viral genome) in the *Sulfolobus* host (31, 51). The prophage can be integrated into the host genome at different sites, depending on the strain of virus (39, 46) or exist episomally as either positively or negatively supercoiled or relaxed (25). The genome can also exist outside of the cell packaged in the viral particle (36, 46). Viral infection occurs most efficiently at temperatures of 80°C, indicating that these viruses are also hyperthermophiles.

Interestingly, the ORFs on half of the genome appear to be conserved among the SSVs while ORFs on the other half are not (46). Of the annotated ORFs, only four were predicted functionally. This is due to the fact that the ORFs of Archaea and their viruses

are not significantly similar to those in the public database (31). These four ORFs include a type 1 tyrosine recombinase (integrase) (24) and three structural proteins VP1, VP2, and VP3 (41).

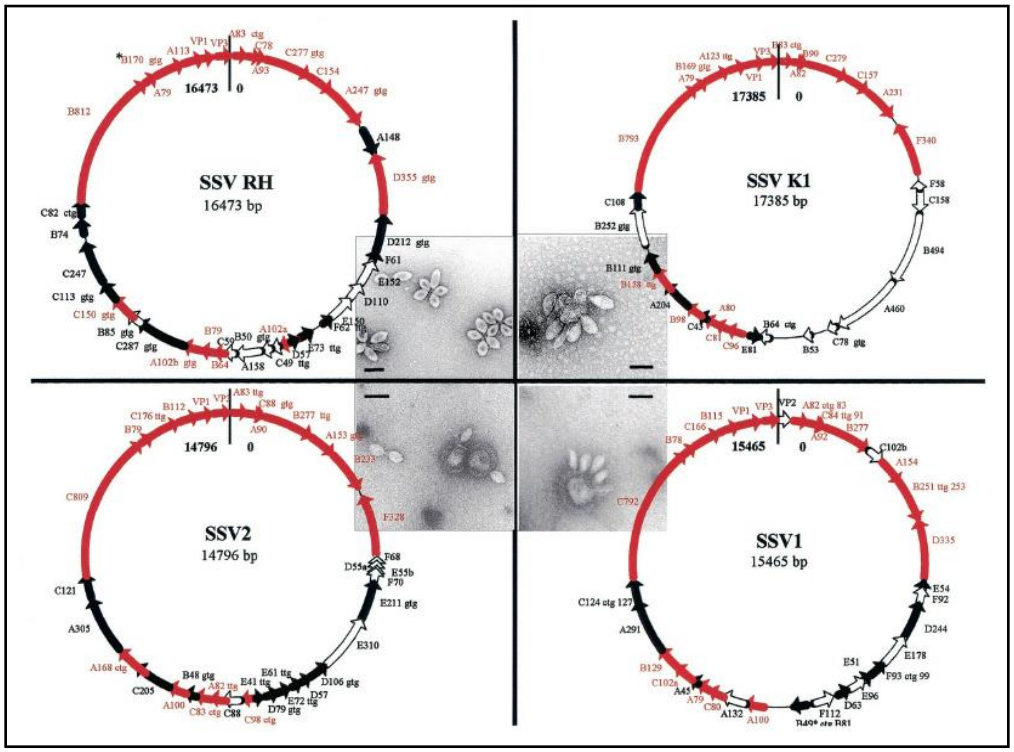


Figure 3: Genome Maps of the Four Sequenced SSVs. Conserved ORFs shared by all genomes are shown as red arrows. ORFs shared between two or three of the SSV genomes are shown as solid black arrows, and ORFs unique to each isolate are shown as open arrows. TEM images of each virus are positioned next to maps of their respective genome. The alternative initiation codons (asterisks) are indicated directly following the name of each ORF in which they are identified. (46)

Sulfolobus Spindle Virus 1: SSV1 was the first Chrenarchaeal virus to be isolated (23) and later completely sequenced (29). SSV1 is integrated into the arginine tRNA gene of *Sulfolobus* (36) and viral particle production can be induced by UV-light irradiation (37, 50) or mitomycin C, a DNA damaging agent (45). The SSV1 genome is a circular dsDNA of 15.5kb and like its host, has a low G + C content of 39.7%. The

genome contains 9 transcripts and 34 originally annotated ORFs (13, 29) (Figure 4). When *Sulfolobus* is UV irradiated, integrated SSV1 genome replication is induced without excision from the host genome (37). Additionally, the transcript t_{ind} is produced and peaks within four hours followed by the T5 and T6 transcripts within three hours (32, 39, 51). T_{ind} is between the T5 and T6 promoter and has been proposed to be the trigger for SSV1 replication (32). Structural proteins VP1 and VP3 are located in the host cell membrane while VP2 is attached to viral DNA. The viral DNA appears to be packaged in the host cell membrane and is released without lysis of the host cell (28, 51).

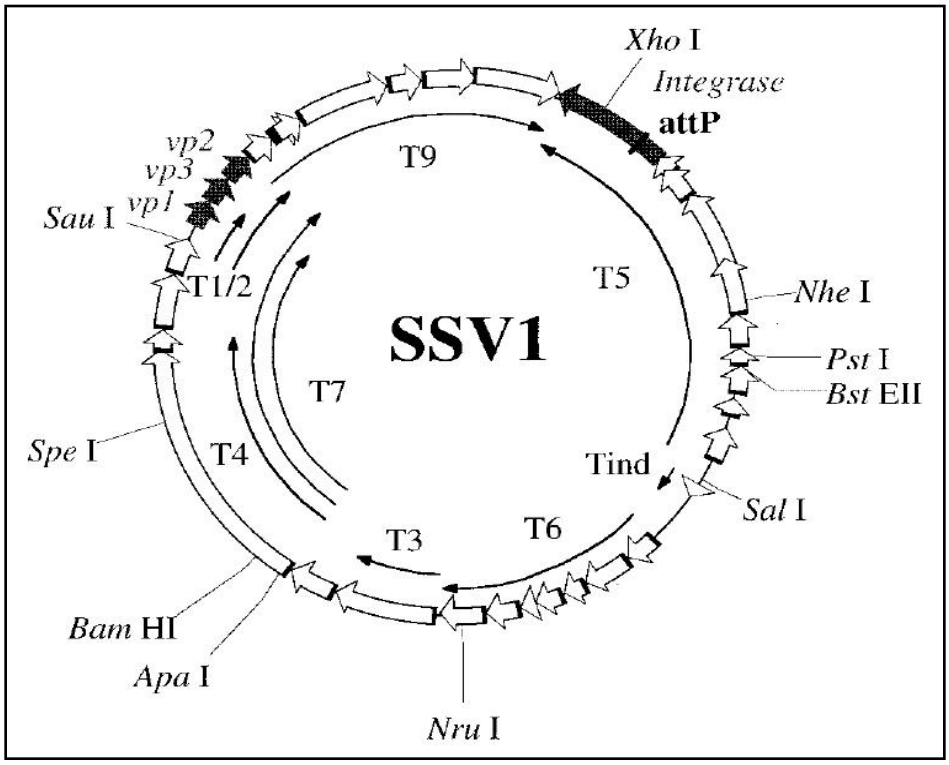


Figure 4: Genome Map of SSV1. Transcripts are marked with thin arrows. Open reading frames are marked with wide arrows. The four identified genes, *vp1*, *vp3*, *vp2*, and the viral Integrase are labeled and highlighted in gray (ref. 42)

Importance of SSV Study

Archaeal viruses, like other viruses, depend on their hosts biochemical machinery and are therefore an attractive tool for determining the biochemical and genetic intricacies of their Archaeal hosts. Additionally, understanding of the biochemical machinery of Archaea opens the door for the development of molecular model systems and techniques specific to Archaea such as shuttle vectors, complementation, and expression studies. Further, the gene products of Archaea are attractive as potential reagents for biotechnology due to their inherent ability to withstand non-physiological conditions.

DNA BINDING PROTEINS OF ARCHAEAL VIRUSES

DNA binding proteins have an affinity for DNA sequences due to the presence of DNA-binding domains. DNA-binding proteins can function in a number of ways such as controlling DNA-replication and transcription, DNA packaging, cutting, splicing, and changing the shape of DNA. There are 54 structural families of DNA-binding proteins (22); three of these are known to be represented in SSVs. These three families are the ribbon-helix-helix (RHH), winged-helix-turn-helix (wHTH), and zinc-finger proteins (Figure 5).

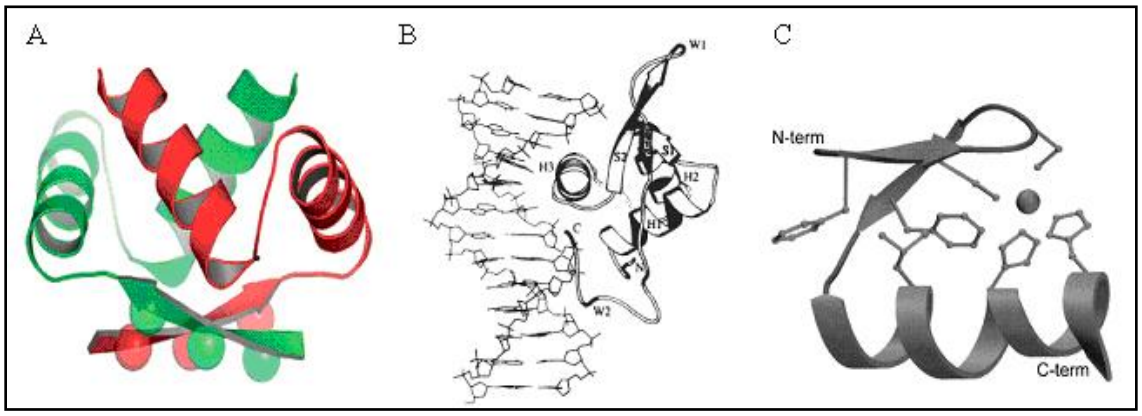


Figure 5: Models of DNA Binding Proteins. (A) Ribbon-helix-helix (ref. 38), (B) Winged-helix-turn-helix (ref. 8) and (C) Zinc-Finger (ref. 49).

RHH proteins are composed of an N-terminal beta-strand containing the DNA binding motif and two alpha-helices. RHH proteins bind DNA as dimers with the beta-strand of each monomer anti-parallel to the other and binding to the major grove of DNA (34). SSV1 contains three (C43, B55, E51) predicted RHH DNA-binding proteins (Figure 6 and Table 2).

The wHTH DNA-binding proteins are composed of a “wing” made up of a string of amino acids which interact with DNA to stabilize binding. The rest of the protein has

a few beta strands followed by a bundle of helices, usually three, one of which interacts with the major grove of DNA (8). SSV1 contains two (B115, E178) predicted and two (F112, F93) solved wHTH DNA-binding proteins.

Finally, the zinc-finger proteins are a family of DNA binding proteins which utilize a zinc ion for protein folding. A single zinc-finger is usually 30 to 35 amino acids in length and is typically composed of two anti-parallel beta-strands followed by an alpha-helix. The protein utilizes cysteine and histidine residues to sandwich itself around a single zinc ion. Zinc-finger proteins do not bind DNA well as monomers but a string of zinc-fingers can wrap around DNA and bind it quite stably (49). SSV1 contains one (A79/A126) predicted zinc-finger protein and one solved (B129) zinc-finger protein. Zinc-finger proteins are found in eukaryotes but not bacteria.

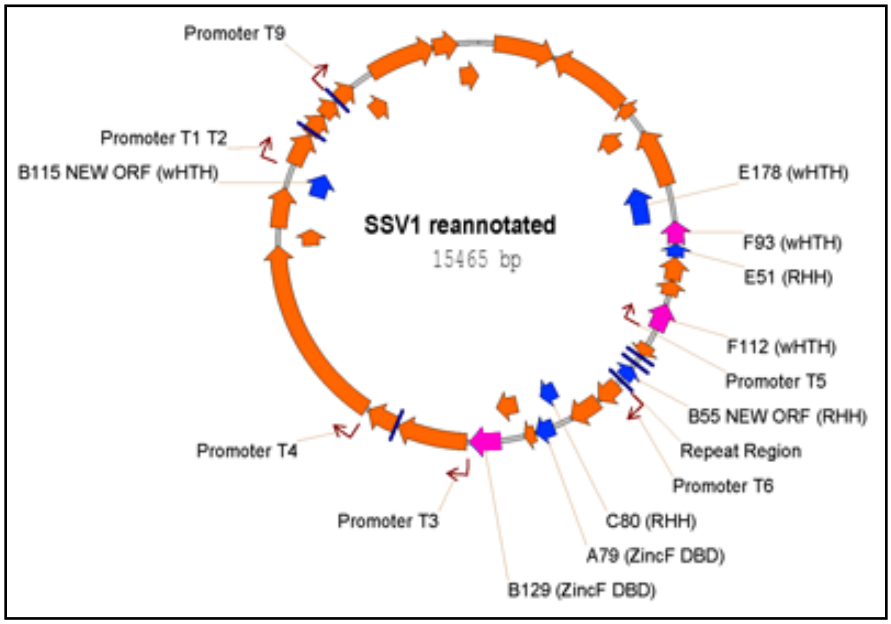


Figure 6: The SSV1 Genome Reannotated by M. Dlakic. ORFs marked in blue or pink indicate predicted DNA-binding proteins. Provided by M. Dlakic.

Table 2: DNA-Binding Proteins of SSVs. M. Dlakic

	SSV1	SSV2	SSV-K	SSV-RH
Ribbon-helix-helix	C80*(C43) E51 B55	A82	C81 D79 E61	B64(B43) E73 F62
Winged helix-turn-helix	B115 **F93** E178* **F112**	B112 D106	A123 E81	A113 E150
Zinc-finger	A79/A126 **B129**	C83 A168	A80 B158	B79* C150

Bold indicates proteins with solved structures
Blue indicates cloned and expressed by Joella Foust
Green indicates cloned and expressed by Mensur Dlakic
Red box indicates annotated by Mensur Dlakic PhD
Black box indicates cloned by Martin Lawrence PhD
* indicates insoluble

The original annotation of SSV1 yielded 34 ORFs. Of these ORFs, seven were predicted to be DNA-binding proteins. However, upon further scrutiny of the genome by our lab two additional ORFs, also predicted to be DNA-binding proteins, were identified. The first of these is B115, a wHTH protein 115 amino-acids in length. B115 is located just upstream of the T1/T2 promoter. The second is B55, a RHH protein 55 amino-acids in length. B55 is located between the T6 and T5 promoters and coincides with the UV-inducible transcript T_{ind} mentioned by Palm (29) and Stedman (42).

Within the four SSVs, twelve DNA-binding proteins have been cloned, three have solved structures (Martin Lawrence, personal communication), two have been truncated from the original annotation (B43, C43) (M. Dlakic, personal communication), and one has been expanded (A126) (M. Dlakic, personal communication). A126 was originally

annotated as A79, a predicted two zinc-finger protein which is followed by a stop codon and a single zinc-finger protein A45. The original annotation of 79 amino-acids was predicted because of the stop codon but the author did not consider that Archaea can use selenocysteine and pyrrolysine during translation to read through stop-codons (29). Additionally, there are no known examples of a zinc-finger existing as a monomer. When these considerations are taken into account we suggest extending A79 to A126, encompassing the single zinc-finger after the stop-codon (M. Dlakic, personal communication) (Figure 7).

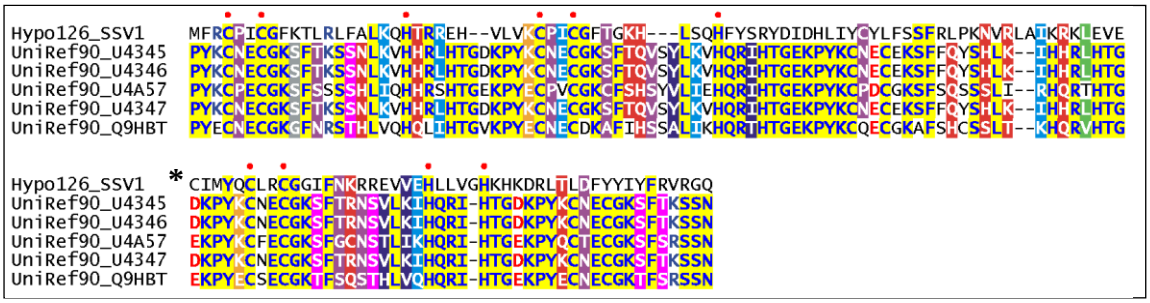


Figure 7: New Protein Sequence of A126 (top row) Compared to Other Known Three-Zinc-Finger Proteins. The position of the stop-codon in the original (A79) annotation replaced by selenocystein is denoted by an asterisk. Provided by M. Dlakic.

The truncations of C80 (SSV1) to C43 and its homolog B64 (SSV-RH) to B43 were a result of several observations. First, RHH proteins tend to begin with a ribbon at their N-terminus. This ribbon is not known to be preceded by additional coding sequence. The preceding 37 amino-acids originally annotated within C80 would be the first example of coding sequence preceding the N-terminal of an RHH. Second, 43 amino-acids is a typical size for RHH proteins; the truncation does not create a novel property of RHH proteins. Third, the truncated 37 amino-acids, when aligned with the other three homologues within the SSVs do not align well. There is a short alignment at

the five prime portion of the original annotation but this may function as a non-coding DNA regulatory sequence (Figure 8). Finally, Dr. Martin Lawrence's lab cloned and expressed C80 and found that it is insoluble (M. Dlakic, personal communication).

Figure 8: Alignment of C80 Homologues Found in SSVs. Provided by M. Dlakic.

Possible DNA Binding Sequences of SSV1 DNA-Binding Proteins

The DNA-binding proteins of SSVs most likely bind either the DNA of their own genomes or that of the Archaeal host. Several regions in the SSV1 genome do not appear to be coding sequences (Figure 9) and therefore are likely candidates for interaction with DNA-binding proteins. It is also possible that these non-coding sequences are target sequences for DNA binding proteins from the host. However, such a scenario would render SSV1 genome regulation dependent in part on the host creating a precarious evolutionary situation for the virus.

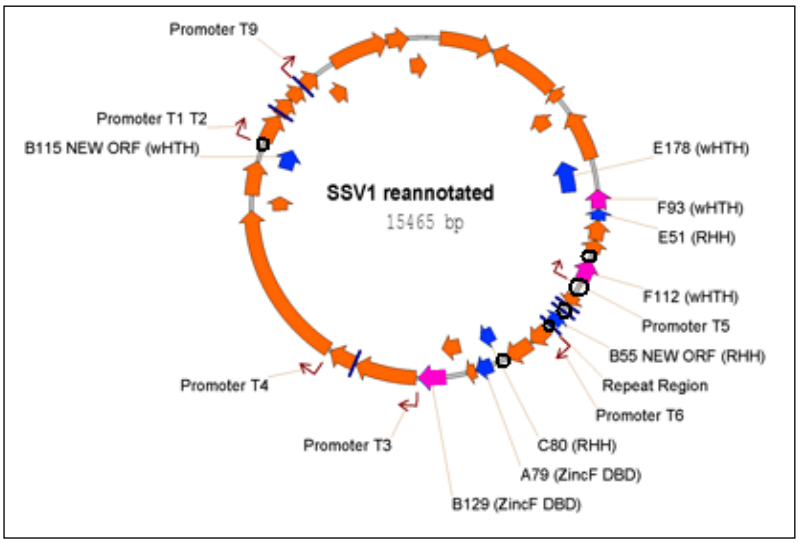


Figure 9: Reannotated Genome of SSV1. Non-coding sequences are noted by black circles. Provided by M. Dlakic.

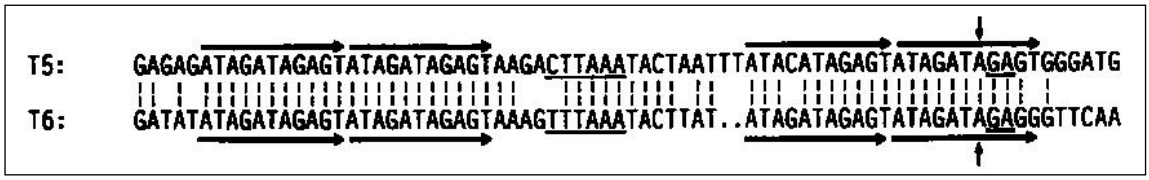


Figure 10: Promoter Sequences of the T5 and T6 Transcripts Contain All but Two of the SSV1 DNA Binding Proteins. (ref 29).

Interestingly, the promoter sequences of T5 and T6 transcripts contain all but two of the DNA-binding proteins (B55 and B115) (Figure 9). Additionally, these two promoters appear to be almost identical (Figure 10). It is therefore likely that one of the SSV1 DNA-binding proteins interacts with these sequences thereby regulating most of the genome control system of the virus. We expected interesting insights from screening for such interactions.

STATEMENT OF PURPOSE

The purpose of the following work is to identify C43 protein DNA binding sequences in the SSV1 genome. Elucidating the interaction of C43 protein with specific SSV1 sequences is undoubtedly an important step towards understanding the molecular interactions involved in SSV1 genome regulation.

MATERIALS AND METHODS

Cloning the C43 Coding Sequence into the pDS-56 Vector

ATGCAAGTAATATCCTTTAAACTACCC
CCGGCGTTAAATGCAAAACTAGAACA
AATTGCGATCAAAGAAAAGAAAAGCA
AGAGTGAAATTATTCGAATAGCGTTA
GCGAGGTATGTAGAAAATGTTTAG
Figure 11: DNA sequence of native C43

Figure11: DNA Sequence of Native C43.

Primers flanking the C43 coding region of SSV1 (Figure 11) were designed: 5' primer CAT CAC ATC GAT GGT CGT ATG CAA GTA ATA TCC TTT AAA CTA CCC and 3' primer ATT AAG CTT GGA TCC CTA AAC ATT TTC TAC ATA CCT CGC TAA. These primers were used in a PCR reaction to amplify the C43 coding region and introduce 5' *ClaI* and 3' *BamHI* restriction sites. Additionally a 3' 6-His tag was introduced. The PCR reaction contained: 1µl SSV1-pUC19 DNA, 1mg of each primer, 2-3 units Vent DNA polymerase, 1x Thermo-pol buffer (NEB), 400µM dNTP mix, and diH_2O for a total volume of 50µl. The melting temperature was 95°C (20 sec), the annealing temperature was 60°C (30 sec), and the extension temperature was 72°C (60 sec). The PCR was run for 25 cycles with an initial cycle at 98°C (240 sec).

The PCR reaction was then digested with the restriction enzymes *ClaI* and *BamHI*. The pDS-56 vector was also digested with these restriction enzymes but the reaction included alkaline phosphatase (CIP). The restriction enzyme digest reaction included: 1µl *BamHI*, 1µl *ClaI*, 0 or 1µl CIP, 50µl PCR reaction, and 5µl NEB-Buffer #4. The digestion mix was incubated at 37°C overnight. Subsequent gel-electrophoresis

and gel-purification yielded an insert (C43) and vector (pDS-56) which was ligated (3/1 insert to vector). Ligations were incubated at room temperature overnight. 30µl NEB-5 alpha *Escherichia coli* competent cells were transformed with 2µl of the ligation reaction and plated on LB-agar plates containing ampicillin (LB_a+amp). After overnight incubation at 37°C colonies were screened via PCR. Colonies that were positive were sent for sequencing to The University of Nevada Reno Genomics Center. Glycerol stocks were made from those colonies that carrying the plasmid with the correct sequence. This general strategy was used for all SSV1 DNA binding proteins cloned into the pDS-56 vector (Appendix Table 1).

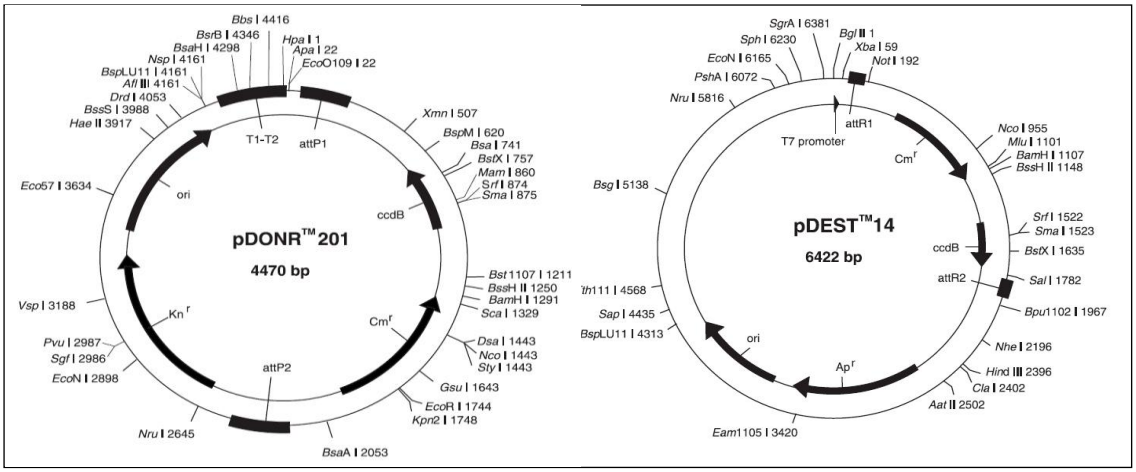


Figure 12: Plasmid Maps of pDONR201 Recombination Vector and pDEST14 Expression Vector. (from www.invitrogen.com).

Cloning the C43 Coding Sequence from pDS-56 Vector into pDONR201 Vector

Cloning of DNA-sequences into the pDONR201 vector relies on Gateway™ Cloning Technology (Invitrogen). This method of cloning utilizes the attR1 and attR2 recombination sites and a recombination enzyme mix to move segments of DNA from

PCR product into a vector or from one vector to another. Generally, a pDONR201 vector is recombined with a PCR product to produce a pDONR201/gene construct. This construct is then recombined with a pDEST14 vector to yield a pDEST14/gene expression construct. The pDEST14 expression vector is T7 RNA-polymerase driven.

Two rounds of PCR amplification are required to produce a PCR product which may be recombined with the pDONR201 vector. The first PCR uses primers which amplified the C43 gene from the pDS-56 vector using the 5' primer TCC GAA GGA GAT AGA ACC ATG CAA GTA ATA TCC TTT AAA CTA and the 3' primer GTG ATG GTG ATG GTG ATG AAC ATT TTC TAC ATA CCT CGC. The PCR reaction contained 1mg of each primer, 2-3 units PFU-DNA polymerase, 1x PFU-DNA polymerase 10x buffer, 400µM dNTP mix, 1µl plasmid DNA pDS-56/C43 (mini-prep), and diH_2O for a total volume of 50µl. The thermo-cycler was run at a melting temperature of 95°C (30 sec), an annealing temperature of 60°C (60 sec), and an extension temperature of 72°C (180 sec). The PCR was run for 30 cycles with an initial cycle at 95°C (15 sec). The second PCR was performed on the first PCR product and introduced the attR1 and attR2 recombination sites necessary for recombination with the pDONR201 vector. The primers used were the 5' primer GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CGA AGG AGA TAG AAC C and the 3' primer GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CCT AGT GAT GGT GAT GGT GAT G. The PCR reaction contained 1mg of each primer, 2-3 units PFU-DNA polymerase, 1x PFU-DNA polymerase 10x buffer, 400µM dNTP mix, 2µl PCR product DNA from the previous PCR, and diH_2O for a total volume of 50µl. The thermo-cycler settings were identical to the previous PCR.

The PCR product from the second PCR was gel purified and recombined with the pDONR201 vector. The recombination reaction contained 9ng of PCR product, 60ng pDONR201 vector, 1µl BP-clonase-enzyme-mix, and was incubated at 25°C overnight. 1µl Proteinase K was then added and the reaction incubated at 37°C for one hour.

100µl NEB-5 alpha *E. coli* competent cells were transformed with 2.5µl recombination reaction and plated on LB-agar plates containing kanamycin (LB_a+kan). Plates were incubated overnight at 37°C. Colonies were randomly selected from the plate and grown overnight in 5ml of LB broth containing kanamycin (LB_b+kan) at 37°C in a shaking incubator. Plasmids were extracted from cells using the QIAprep spin Miniprep Kit (Qiagen) and screened via PCR. Primers used for screening were the 5' primer TAA CGC TAG CAT GGA TCT C and the 3' primer GTA ACA TCA GAG ATT TTG AGA CAC. This general strategy was used for all SSV1 DNA binding proteins cloned into the pDONR201 vector (Appendix Table 1).

Recombination of pDONR201/C43 Vector with pDEST14 Vector

The pDEST14 vector was digested with *HindIII* restriction enzyme (New England Biolabs) and recombined with the C43/pDONR201 construct. The recombination reaction contained 150ng *HindIII* digested pDEST14, 50ng pDONR201/C43, and 1µl LR enzyme mix. The reaction was incubated overnight at 25°C. 1µl Proteinase K was added and the reaction incubated for 10 minutes at 37°C. 50µl NEB-5 alpha *E. coli* competent cells were transformed with 1µl LR recombination reaction and plated on LB_a+amp plates. Plates were incubated overnight at 37°C. Colonies were randomly selected from

the plate and grown overnight in 5ml LB_b+amp at 37°C in a shaking incubator. Plasmids were extracted from cells using the QIAprep spin Miniprep Kit (Qiagen) and screened by PCR. Primers used for screening were the 5' primer AGA TCT CGA TCC CGC GA and the 3' primer AAG ACC CGT TTA GAG GCC. Colonies that were positive were sent for sequencing to The University of Nevada Reno Genomics Center. One sample that sequenced correctly was chosen for a glycerol stock. This general strategy was used for all SSV1 DNA binding proteins cloned into the pDEST14 vector (Appendix Table 1).

Transformation of BL21-DE3-pLysS E. coli Cells with pDEST14/C43 Construct

pDEST14 is a T7-RNA-polymerase (T7-RNA-pol) driven vector and genes cloned into the restriction sites of this vector can only be expressed by T7-RNA-pol. NEB-5 alpha cells do not carry a T7-RNA-pol. However, BL21-DE3-pLysS cells carry the T7-RNA-pol under the control of a *lac*-promoter. Additionally, this strain carries a pLysS plasmid whose gene product (LysS) will inactivate T7-RNA-pol in the event of *lac*-promoter leakage. Thus, the induction of expression is tightly controlled.

The transformation reaction contained 30µl of BL21-DE3-pLysS cells and 1µl of pDEST14/C43 plasmid and was plated on LB_a+amp and chloramphenicol plates (LB_a+amp+chl). Plates were incubated overnight at 37°C. A colony was chosen and grown in 5ml LB_b+amp+chl overnight at 37°C in a shaking incubator. Plasmids were extracted from cells using the QIAprep spin Miniprep Kit (Qiagen) and screened by PCR for the pDEST14/C43 plasmid. Primers used for screening were the 5' primer AGA TCT CGA TCC CGC GA and the 3' primer AAG ACC CGT TTA GAG GCC. One positive sample

was used to make a glycerol stock. This general strategy was used for all SSV1 DNA binding proteins that were moved into the BL21-DE3-pLysS expression strain (Appendix Table 1).

Expression of C43 from pDEST14 Vector

After the C43 coding sequence was transformed into an appropriate expression vector (pDEST14) and appropriate expression strain (BL21-DE3-pLysS), the C43 protein could be expressed. An auto-induction system was employed for expression (43). The auto-induction system relies on *E. coli's* preferential use of glucose as an energy source if placed in a media containing both glucose and lactose. However, when the supply of glucose becomes exhausted, lactose is utilized. If the concentrations of glucose to lactose in the expression media are carefully calculated, the glucose supply becomes exhausted when the culture becomes substantially overgrown. At this point the culture begins to utilize lactose which activates the *lac*-promoter and results in the expression of T7 RNA-pol. T7 RNA polymerase drives the expression of C43 from the T7-RNA-pol driven pDEST14 vector.

The pDEST14/C43 containing BL21-DE3-pLysS cells were grown in 5ml of non-expression culture median (ZY media, 1mM $MgSO_4$, 0.8% glucose, 1x NPS, 50µg/ml ampicillin, and 25µg/ml chloramphenicol) for 8 hours at 37°C. When the 5ml culture reached a cloudy appearance it was used to inoculate 50ml of non-expression culture media and grown overnight at 37°C. Expression media (500ml, ZY media, 1mM $MgSO_4$, 1x trace metals, 1x 5052, 1X NPS, 25µg/ml chloramphenicol, and 50µg/ml

ampicillin) was inoculated with 10ml from the 50ml overnight culture and grown 24 hours at 37°C in a shaking incubator. After 24 hours the expression culture was split into two 250ml hard walled ultra-centrifuge tubes and spun at 2,400g for 15min. The supernatant was discarded and the remaining cell pellet resuspended in 10ml native lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole, 1mM beta-mercaptoethanol, pH 8.0). The resuspended cell pellet was transferred into a 50ml ultracentrifuge round bottomed tube, sonicated for 30 seconds and centrifuged at 20,400g for 30 minutes. The supernatant was poured over a 200ml Ni-NTA Agarose binding column which had been previously equilibrated with native lysis buffer. The supernatant was allowed to pass through the column and the column then washed with 2ml native wash buffer (50mM NaH_2PO_4, 300mM NaCl, 20mM imidazole, 1mM beta-ME, pH 8.0). After this, 500µl native elution buffer (50mM NaH_2PO_4, 300mM NaCl, 250mM imidazole, 1mM beta-ME, pH 8.0) was used to elute the C43 protein.

It is known that the presence of imidazole can interfere with protein binding assays, so the elution buffer was exchanged for a buffer containing no imidazole (50mM NaH_2PO_4, 300mM NaCl, 1mM beta-ME, and pH 8.0). Buffer exchange was accomplished by placing the elution in an YM-3 MICROCON Millipore filter column and spinning at 14,000g for 60min. The buffer was exchanged twice by adding exchange buffer to the spin column and spinning again at 14,000g for 60min. Finally the volume in the column was brought up to 500µl and transferred into a clean 1.5ml tube.

Additionally, a denaturing purification was performed on the remaining cell pellet. The cell pellet was resuspended in 10ml of denaturing lysis buffer (6M guanidine-

HCL, 0.1M NaH_2PO_4, 0.01M Tris-Cl, pH 8.0). After sonication the tube was centrifuged at 20,400g for 30 minutes. The supernatant was poured over a 200ml Ni-NTA Agarose binding column which had been previously equilibrated with denaturing lysis buffer. The supernatant was allowed to pass through the column and the column was then washed with 2ml denaturing wash buffer (6M guanidine-HCL, 0.1M NaH_2PO_4, 0.01M Tris-Cl, pH 6.3). After this step, 500µl denaturing elution buffer (6M guanidine-HCL, 0.1M NaH_2PO_4, 0.01M Tris-Cl, pH 5.75) was used to elute the C43 protein.

Because a denaturing purification unfolds the purified protein, it is critical to exchange the elution buffer with a buffer containing no guanidine. Thus, a buffer containing no guanidine (50mM NaH_2PO_4, 300mM NaCl, 1mM beta-ME, and pH 8.0) was used to exchange buffers as described for the native purification. Purifications were performed at 26°C because thermal proteins should not be affected by temperatures which normally damage non-thermal proteins.

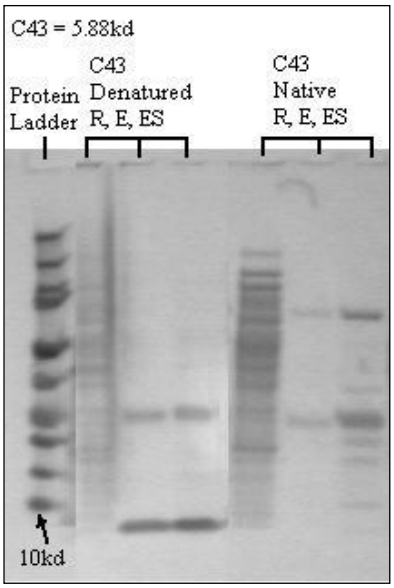


Figure 13: C43 Expression. C43 is being expressed but can only be purified via denaturing purification because the protein is located in inclusion bodies.

Bradford assays were performed on all samples and concentrations determined via a pre-calibrated curve; raw extract (R), flow thorough (FT), wash (W), elution (E), spun elution (ES). A SDS-polyaclyamide gel was used to determine the presence of C43 and to determine purity (see Figure 13). Expression of other SSV1 DNA binding proteins was carried out in a similar manner with varying results (Appendix Table 2).

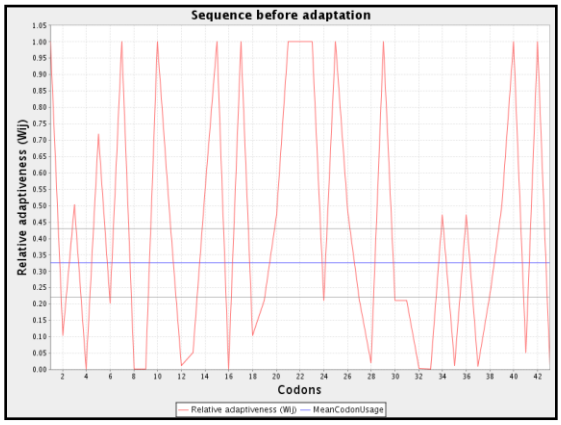
Construction of Synthetic C43 ($C43_s$) DNA Coding Sequence

The native purification did not yield pure protein. However, a denaturing purification did yield purified C43 protein. Native purification is the preferred method when purifying proteins because the protein does not need to be refolded after purification, increasing the likelihood of obtaining a functional protein. Because

expressing C43 under native conditions failed consistently, a synthetic C43 construct was made that would be identical to C43 at the amino-acid level but be codon optimized for *E. coli* at the nucleotide level.

In order to achieve synthetic C43 gene construction the online tool JCAT, a codon adaptation tool, was used to determine the oligos necessary for synthetic gene construction. This online tool can be found at http://www.jcat.de. The native C43 sequence was entered into the tool and the target strain *E. coli* K12 selected. In Figure 14A it is shown that many of the codons used by the C43 native gene are not well represented in *E. coli* (fall below the blue threshold line). Also, the optimization value for the codons in the sequence is well under 1.0 (1.0 representing perfect codon representation for a given organism). Therefore, *E. coli* may not be able to express the native sequence of C43 effectively. The JCAT tool provided an alternative sequence that is optimized for codons well represented in *E. coli*. The improved sequence has an optimization value of 1.0 and, as can be seen by the corresponding graph, is much more suited to the codons utilized by *E. coli* (Figure 14 B).

A:



B:

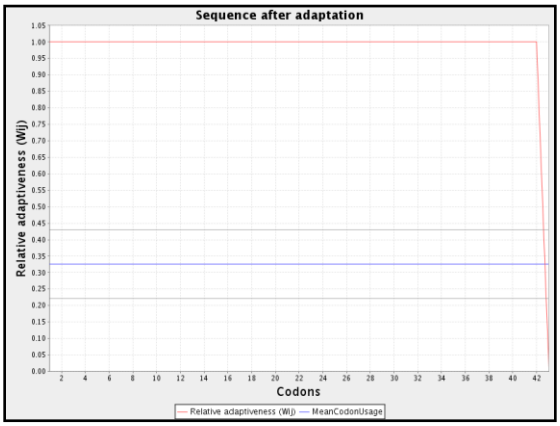


Figure 14: Codon Optimization Analysis of C43 in E. coli. http://www.jcat.de
A: Codon usage analysis of the native C43 sequence in *E. coli* strain K12. Native C43 sequence with 6-His tag:
ATGCAAGTAATATCCTTTAAACTACCCCCGGCGTTAAATGCAAAACTAGAACAAATTGCGATC
AAAGAAAAGAAAAGCAAGAGTGAAATTATTCGAATAGCGTTAGCGAGGTATGTAGAAAATGT
TCATCACCATCACCATCACTAG. Optimization value: 0.169
B: Codon usage analysis of the synthetic C43 DNA sequence in *E. coli* strain K12. Synthetic C43 sequence with 6-His tag:
ATGCAGGTTATCTCTTTCAAACTGCCGCCGGCTCTGAACGCTAAACTGGAACAGATCGCTATC
AAAGAAAAAAAATCTAAATCTGAAATCATCCGTATCGCTCTGGCTCGTTACGTAGAGAACGTC
CATCACCATCACCATCACTAG Optimization value: 1.0

In order to construct the synthetic gene a series of oligonucleotides were used:

oligo 1:

ATGCAGGTTATCTCTTTCAAACTGCCGCCGGCTCTGAACGCTAAA

oligo 2:

TTTCAGATTTAGATTTTTTTTCTTTGATAGCGATCTGTTCCAGTTTAGCGTTCAGAGCCGG

oligo 3:

CGCTATCAAAGAAAAAAAATCTAAATCTGAAATCATCCGTATCGCTCTGGCTCGTTACGTAG

oligo 4:

CTAGTGATGGTGATGGTGATGGACGTTCTCTACGTAACGAGCCAGAGC.

These oligonucleotides anneal to from the following C43 dsDNA.

```
5'-ATGCAGGTTATCTCTTTCAAACTGCCGCCGGCTCTGAACGCTAAA      CGCTATCAAAGAAAAAAAATCTAAATCTGAAATCATCCGTATCGCTCTGGCTCGTTACGTAG-3'
3'-                GGCCGAGACTTGCGATTTGACCTTGTCTAGCGATAGTTTCTTTTTTTTAGATTTAGACTTT      CGAGACCGAGCAATGCATCTCTTGCAGGTAGTGGTAGTGGTAGTGATC-5'
```

The four oligonucleotides were combined into one PCR reaction (dNTP's, 1x Thermo pol buffer, 1µl Vent polymerase, diH_2O for a total volume of 100µl). The thermo-cycler was run at a melting temperature of 94°C (90 sec), an annealing temperature of 54°C (120 sec), and an extension temperature of 72°C (120 sec). The PCR was run for 8 cycles with an initial cycle at 94°C (390 sec). 1µl of the PCR product was then amplified in another PCR reaction using the 5' primer TTC GAA GGA GAT AGA ACC ATG CAG GTT ATC TCT TTC AAA CTG C and the 3' primer CTA GTG ATG GTG ATG GTG ATG. The thermo-cycler was run at a melting temperature of 94°C (30 sec), an annealing temperature of 54°C (120 sec), and an extension temperature of 72°C (90 sec). The PCR was run for 25 cycles with an initial cycle at 94°C (300 sec). The PCR product was then gel purified and another PCR was performed on this product to introduce the attR1 and attR2 recombination sites. The primers used were the 5' primer GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CGA AGG AGA TAG AAC C and the 3' primer GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CCT AGT GAT GGT GAT GGT GAT G.

This PCR product was also gel purified and then recombined with the pDONR201 vector. The recombination reaction contained 150ng of PCR product, 150ng pDONR201 vector, and 1µl BP clonase enzyme mix and was incubated at 25°C overnight. 1µl Proteinase K was then added and the reaction incubated at 37°C for 10 minutes. The transformation reaction was plated on LB_a+kan plates and contained 30µl NEB-5 alpha *E. coli* competent cells and 2µl recombination reaction. Plates were incubated overnight at 37°C. Colonies were randomly selected from the plate and grown overnight in 5ml LB_b+kan at 37°C in a shaking incubator. Plasmids were extracted from cells using the

QIAprep spin Miniprep Kit (Qiagen) and screened by PCR. Primers used for screening were the 5' primer TAA CGC TAG CAT GGA TCT C and the 3' primer GTA ACA TCA GAG ATT TTG AGA CAC. A positive screened sample was used to make a glycerol stock. This general strategy was also used for the C43 homolog of SSV1-RH, B43.

pDONR201/C43$_s$ was recombined with the pDEST14 vector. The recombination reaction contained 150ng pDEST14, 50ng pDONR201/C43$_s$, and 1µl LR enzyme mix and was incubated overnight at 25°C; 1µl Proteinase K was then added and the reaction incubated for 10 minutes at 37°C. The transformation reaction was plated on LB$_a$+amp+chl plates and contained 30µl NEB-5 alpha *E. coli* competent cells and 2µl of the recombination reaction. Plates were incubated overnight at 37°C. Colonies were randomly selected from the plate and grown overnight in 5ml LB$_b$+amp+chl at 37°C in a shaking incubator. Plasmids were extracted from cells using the QIAprep spin Miniprep Kit (Qiagen) and screened by PCR. The primers used for screening were the 5' primer AGA TCT CGA TCC CGC GA and the 3' primer AAG ACC CGT TTA GAG GCC. Positive samples were sent for sequencing to The University of Nevada Reno Genomics Center. One sample that had a correct sequence was chosen for a glycerol stock. This general strategy was also used for the C43 homolog of SSV1-RH B43.

The transformation reaction was plated on LB$_a$+amp+chl plates and contained 30µl NEB-5 alpha *E. coli* competent cells and 1µl of pDEST14/C43$_s$ plasmid. Plates were incubated overnight at 37°C. A colony was chosen and grown in 5ml LB$_b$+amp+chl overnight at 37°C in a shaking incubator. Plasmids were extracted from cells using the QIAprep spin Miniprep Kit (Qiagen) and screened by PCR for the pDEST14/C43$_s$

plasmid. The primers used for screening were the 5' primer AGA TCT CGA TCC CGC GA and the 3' primer AAG ACC CGT TTA GAG GCC. One sample that had a correct sequence was chosen for a glycerol stock.

Expression of $C43_s$ from Synthetic C43 Sequence

The auto-induction expression system was again employed for the expression of the $C43_s$. Expression, native purification, and buffer exchange procedures were carried out as described above. In Figure 15 it is shown that the exchange of buffers resulted in a pure prep of $C43_s$. The purified and buffer exchanged $C43_s$ protein was then snap frozen in 30µl aliquots in 0.5ml tubes using 100% ethanol and dry ice. The frozen protein aliquots were stored at -80°C.

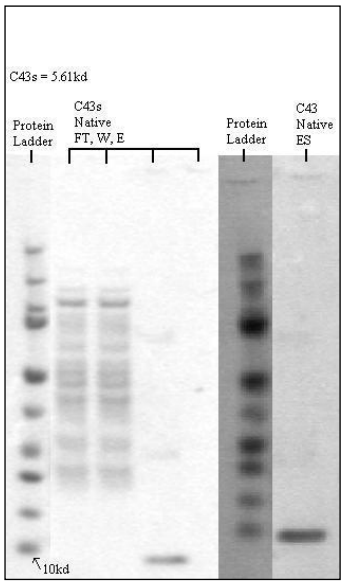


Figure 15: $C43_s$ Expression. $C43_s$ is expressed and can be purified natively.

DNA Fragments Used in Binding Assays

As stated previously, C43 is predicted to have a ribbon-helix-helix structure and therefore most likely is a DNA binding protein. It is possible that if C43 binds DNA, its binding site could be positioned in the genome of the host (*Sulfolobus*), in the SSV1 viral genome, or both. There are several non-coding regions in the SSV1 viral genome (Figure 9 and Table 3). It is possible that these short segments of sequence serve as a binding site for C43.

Table 3: Selected Noncoding Sequence Regions of the SSV1 Genome.

Region of SSV1 genome	DNA Fragment of Interest	Nucleotide Number
T6 – B55 – T5	3'B55 to T5	470bp
T6 – B55 – T5	T6 to T5	843bp
T6 – B55 – T5	B55 to T5	720bp
T6 – B55 – T5	T6 to B55	291bp
T6 – B55 – T5	T5 promoter	101bp
T6 – B55 – T5	T6 promoter	99bp
Other	Non-coding region immediately following F112	182bp
Other	Truncated portion of C80	119bp
Other	Promoter region immediately following B115	109bp

Of particular interest are the T5 and T6 promoters. The T5 and T6 promoters are homologous and are comprised of two sets of direct repeats separated by a region containing a TATA box. These putative promoters are directly 5-prime of the T5 and T6 transcripts that contain all but two of the annotated DNA-binding proteins in the SSV1 genome (Figure 10).

The T5-promoter and T6-promoter were amplified from the SSV1 genomic DNA via PCR using the following primers.

Fragment one contains the T5-promoter and is 101bp in length.

5’ primer TCC TCT AGA ACC CAA ACA CTG TGT ATA TAG AGA G

3’ primer TGC GTC GAC ATA TTC CCA TCC CAC TCT ATC

Fragment two contains the T6-promoter and is 99bp in length.

5’ primer TCC TCT AGA GAT AAT ATT AAA TGA TTC ACG ATA T

3’ primer TGC GTC GAC CCA TTT TTT GAA CCC TCT AT

PCR products were gel purified from a 2% agarose gel, and their relative concentrations determined by running 2µl samples of the gel purified fragment on an 2% agarose gel.

T5 & T6 Promoters

In order to determine if the T5 & T6 promoter fragments were indeed binding sites of C43, DNA/protein binding assays were performed. In principal, a fragment of DNA will have a particular rate of migration through a polyacrylamide gel. If this fragment is bound by a protein, such as C43, the bulk of the fragment/protein complex causes a slower rate of migration through the same gel. Thus, binding of a DNA fragment by a protein can be detected by a shift in the migration of the DNA fragment, revealed by staining of the DNA. As protein concentrations increase more DNA is bound and a decrease in the amount of free DNA can be detected.

A binding assay to test whether C43 binds to the T5 promoter fragment was performed. Binding assays were performed at four temperatures: 26°C, 37°C, 50°C, and 65°C. These temperatures were chosen because of the hyperthermophilic nature of SSV1

and its host, *Sulfolobus*. The binding reaction contained 10ng poly-dIdC DNA competitor, 1x 50% glycerol, 2x Taq ligase buffer, a consistent amount of T5 promoter DNA, and protein amounts ranging from 50ng to 400ng. Binding reactions were incubated at the appropriate temperatures for 30 minutes and then loaded directly onto an 8% polyacrlyamide gel (8% acylamide 19:1bis solution, 10% APS, 1x Tris-glycine buffer, water, and 35% TEMED). Gels were pre-run at 150 volts for 15 minutes. Samples were then loaded on the gel and run at 250 volts. GelPilot Loading Dye (Qiagen) was used as a loading dye and also as a marker for gel run time. Gels were run until the Xylene cyanol band had migrated two-thirds of the way through the gel. Gels were then stained in 100ml 1x Tris-glycine and 1µl SyberGreen by shaking at 60rpm for 3 hours at 26°C. Gels were imaged with a Computar H6Z0812 camera and UV illuminator.

Truncated Sequence of C80

As discussed previously, C43 was originally annotated as C80 but was truncated by M. Dlakic to C43. Of interest was the possibility that C43 also binds the putative C43 promoter region. The truncated portion of C43 was amplified via PCR from SSV1 genomic DNA and purified in the same manner as the T5 and T6 promoters. Primers used for amplification were the 5' primer TGA AGG CTA GGG TTG AAT AC and 3' primer CCT GCA TTG TTT CCT CTA AC. The putative C43 promoter was used in a binding assay with $C43_s$. The procedure for the binding assay was identical to the one previously described for the T5 promoter with the exception that the 37°C and 65°C temperatures were used for the 30 min reaction incubation.

Phasing Constructs

In addition to determining $C43_s$ binding interactions between specific DNA sequences such as the T5 and T6 promoter regions, binding assays may be performed which can detail the angle and degree to which DNA is bent by $C43_s$ binding. Such a binding assay is termed phasing. In general phasing is based on a set of short specific DNA fragments that contain an inherent bend at a specific location and all vary by 2bp in length. There are six constructs ranging from no base additions to 10 base additions. Since DNA makes one helical turn about every 10 bases, the addition of 2bp introduces a certain amount of orientation change into the total DNA segment (17, 52) (Figure 16, 17, 18 and Table 4).

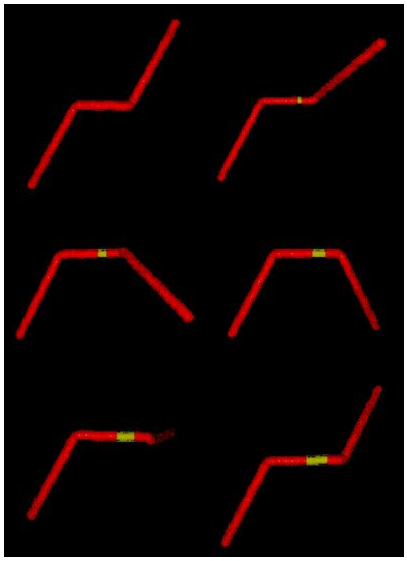


Figure 16: Phasing Construct Models. Red tubes represent the phasing DNA segment. Yellow segments represent the sequential addition of bases. Note the change in rotation as bases are added. M. Dlakic

Restriction enzyme sites are present at specific regions within the phasing constructs and allow for the insertion of any DNA sequence of interest. In this work, one direct repeat of the T5 or T6 promoter (denoted by T5/T6 double) was cloned into each phasing construct containing sequential 2bp additions 3' of the *SalI* site. Other sequences of interest include the T5-promoter, the T6-promoter, and a single repeat of the T5 or T6 promoter (denoted T5/T6 single) (Table 5). These constructs were then used in a binding assay with C43. The procedure for performing the binding assay is identical to the one previously described for the T5-promoter. The only change occurs in the DNA that is added to each binding reaction.

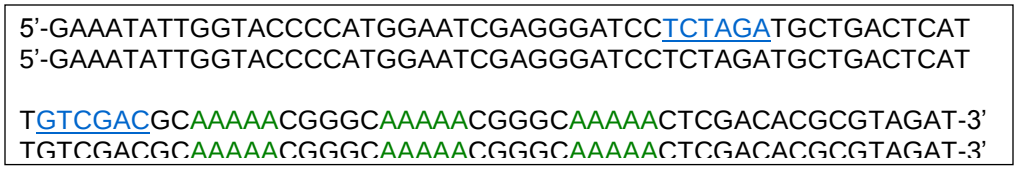


Figure 17: Sequence of Phasing Construct With no bp Additions. Notice the *XbaI* and *SalI* cleavage sites where any desired segment of DNA may be cloned in. The A-tract repeat is the site of inherent bend.

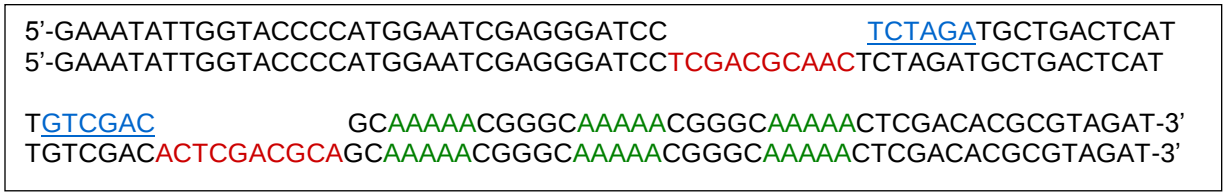


Figure 18: Sequence of Phasing Construct With 10 bp Added 5' of the *XbaI* Site and 3' of the *SalI* site. Note each 2bp addition will twist the helix a certain amount in relation to the "inherent-kink" located at the A-tract repeats.

Table 4: Parental Phasing Constructs with sequential 2bp additions.

	SalI+0	*SalI+2*	*SalI+4*	*SalI+4*	*SalI+8*	*SalI+10*
XbaI+0	X	X	X	X	X	X
XbaI+2	X	X	X	X	X	X
XbaI+4	X	X	X	X	X	X
XbaI+6	X	X	X	X	X	X
XbaI+8	X	X	X	X	X	X
XbaI+10	X	X	X	X	X	X

X denotes sequence confirmation of construct

Table 5: Phasing Constructs with Various Cloned Sequences.

Sequence	T5 promoter	T6 promoter	T5/T6 single	**T5/T6 double**
Phasing +0bp	**+0**	**+0**	+0	**+0**
Phasing +2bp	+2	+2	**+2**	**+2**
Phasing +4bp	+4	**+4**	+4	**+4**
Phasing +6bp	+6	+6	**+6**	**+6**
Phasing +8bp	**+8**	**+8**	**+8**	**+8**
Phasing +10bp	**+10**	**+10**	**+10**	**+10**

Sequences were cloned between the *Xba1* site and the *SalI* site of the parental phasing constructs. **Bold** indicates sequence verified constructs. Gray indicates constructs yet to be made.

RESULTS

The SSV1 RHH protein $C43_s$ was chosen for DNA-protein binding experiments. This choice was made because screening binding assays indicated $C43_s$ binding to select DNA fragments. Recall that C43 is a truncated version of C80, a truncation that is supported by the factors listed previously (Figure 8).

The C43 coding sequence was extracted from the SSV1 genomic DNA via PCR and cloned into a vector. This accomplished isolation of C43 from the SSV1 genome and made working with the C43 sequence more efficient. C43 was cloned into the pDS-56 vector and subsequently into the vectors pDONR201 and pDEST14. Additionally, a codon optimized C43 ($C43_s$) was constructed and cloned into the vectors pDONR201 and pDEST14. $C43_s$ was expressed, purified, and refolded via buffer exchange. This yielded a pure native protein which was used in binding assays. Additionally, non-coding regions of SSV1 were amplified: The T5-promoter, T6-promoter, and the putative C43 promoter. Binding assays were performed to determine if $C43_s$ interacts with the various DNA sequences. Refer to Materials & Methods for specific protocols.

T5 Promoter and $C43_s$ Protein Binding at 26°C and 37°C

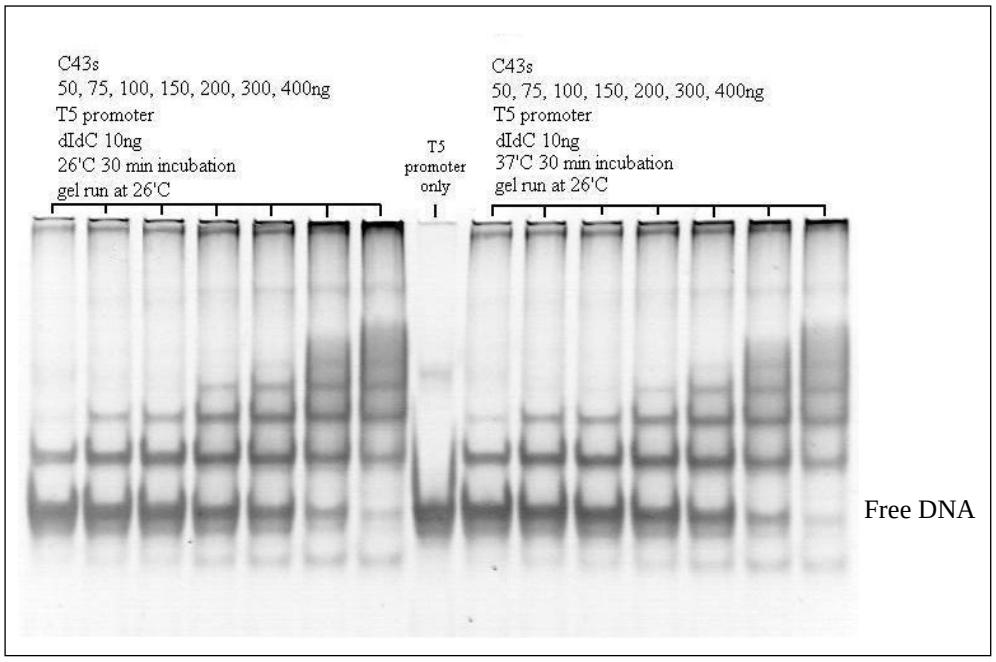


Figure 19: Binding Gel #1: $C43_s$ and T5 promoter binding assays at 26°C and 37°C

In Figure 19 binding gel #1 illustrates the binding of $C43_s$ to the T5-promoter sequence at incubation temperatures of 26°C and 37°C. The gel was cast as described in Materials & Methods. As $C43_s$ concentration increases (left to right) the amount of free DNA decreases and the number of complexes increases. Because the T5 promoter is a tandem direct repeat, it appears that C43 binds these repeats in succession as $C43_s$ concentration increases. At a concentration of 50ng-$C43_s$, one complex band can be seen clearly with a second perhaps beginning. At a concentration of 200ng-$C43_s$, three complex bands can be seen clearly with a potential fourth forming.

Because $C43_s$ is a RHH and because RHH proteins are known to bind DNA non-specifically, it is not surprising that at a concentration of 400ng-$C43_s$ there appears to be five if not six complexes forming even though the binding site consist of four sites. It is

likely that as the binding site becomes saturated with $C43_s$, that $C43_s$ begins to bind non-specifically. Non-specific binding could account for the smearing seen at higher concentrations (300ng and 400ng $C43_s$). Additional evidence that the T5 binding site is becoming saturated is the first complex bands disappearance at 400ng-$C43_s$. The entire first repeat has been bound and more of the second repeat is being bound. A high band can also be seen developing as early as 150ng-$C43_s$ and increasing in intensity as concentrations increase to 400ng. This high band could represent the beginning and increase in non-specific binding as $C43_s$ concentrations increase.

An interesting development is the difference in non-specific binding seen between the two incubation temperatures; 26°C and 37°C. Although the non-specific binding appears to occur at 37°C, the intensity of the high band is much lower. Additionally, there is less smearing of the 300ng and 400ng lanes. These observations indicate increased binding specificity at higher temperatures. Thus, it appears that $C43_s$ binds the T5-promoter sites double tandem repeats sequentially as $C43_s$ concentration increases. Also, after T5-promoter saturation is reached, non-specific binding begins to occur. As temperature increases, binding appears to be more specific.

T5 Promoter and $C43_s$ Protein Binding at 50°C and 65°C

Because of the effect of temperature on the non-specific binding of C43 to the T5 promoter sequence, two additional temperatures were tested: 50°C and 65°C.

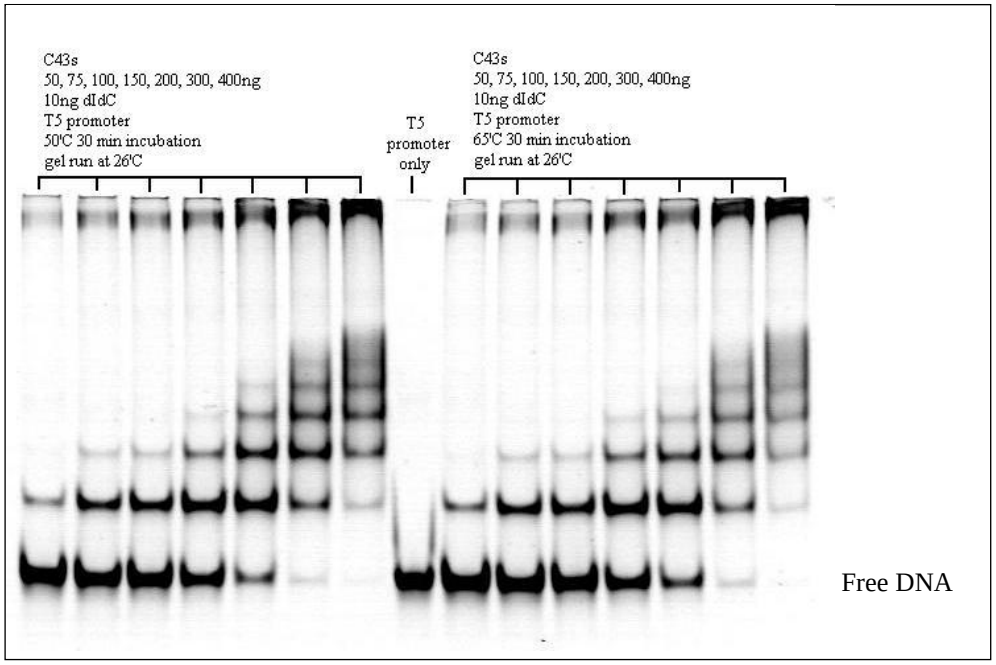


Figure 20: Binding Gel #2: C43 and T5 promoter binding assays at 50°C and 65°C.

In Figure 20 binding gel #2 illustrates the binding of $C43_s$ to the T5-promoter sequence at incubation temperatures of 50°C and 65°C. The gel was cast as described in Materials & Methods. Similar to the observations made at temperatures of 26°C and 37°C, as $C43_s$ concentration increases (left to right) the amount of free DNA decreases and the number of complexes increases: The effect of the double-tandem repeat in the T5-promoter can again be seen as well as one complex band formation at a concentration of 50ng-$C43_s$. However at 200ng-$C43_s$ four complex bands may be seen clearly and at 300ng-$C43_s$ a fifth complex band is appearing. The appearance of more than four complex bands is again most likely due to $C43_s$ saturation of the T5-promoter double tandem repeats and the beginning of non-specific binding. Non-specific binding does not seem to differ greatly when incubation occurs at 50°C or 65°C. What is most remarkable is that $C43_s$ is still quite stable and functional at these high temperatures. These

temperatures would normally destroy a nonthermal purified protein. However, $C43_s$ appears to bind well at high temperatures and in fact, bind more specifically.

T6 Promoter and $C43_s$ Protein Binding at 26°C and 37°C

The T5-promoter differs from the T6-promoter by one base located within the tandem repeats. Therefore, it is of interest whether the T6-promoter is also a binding site of $C43_s$. A binding assay was carried out in an identical manor to the T5-promoter binding assay.

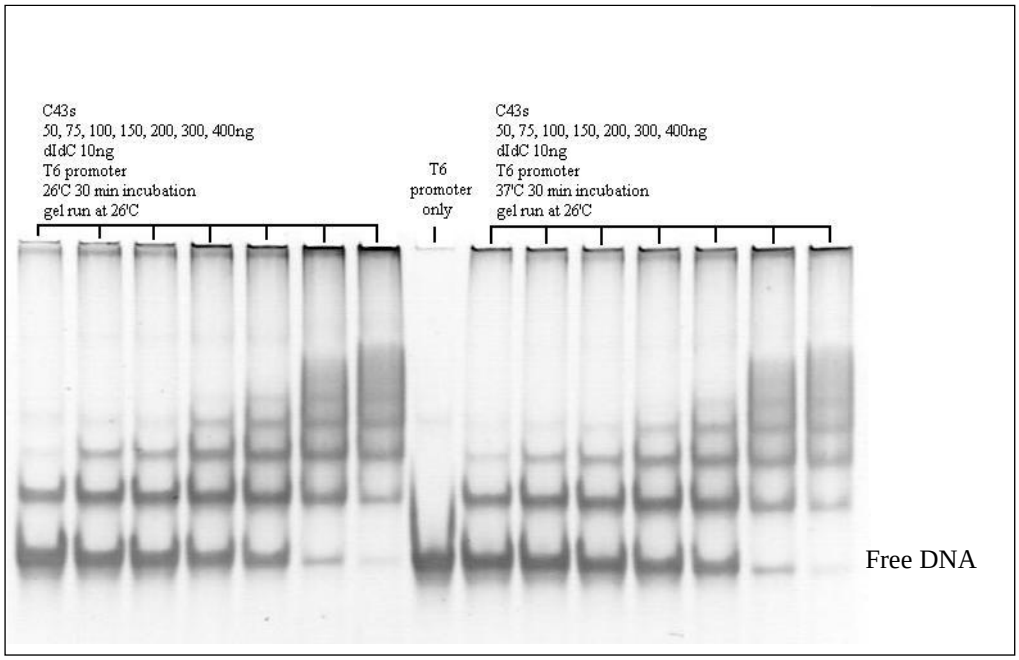


Figure 21: Binding Gel #3: C43 and T6 promoter binding assays at 26°C and 37°C.

Binding gel #3 in Figure 21 shows the binding of $C43_s$ to the T6-promoter sequence at incubation temperatures of 26°C and 37°C. The gel was cast as described in Materials & Methods. Similar to the T5-promoter, $C43_s$ binds the T6-promoter double tandem repeats sequentially as $C43_s$ concentration increases. Additionally, when T6-

promoter saturation is reached, non-specific binding appears to develop. Finally, as seen with the T5-promoter, binding becomes more specific as temperatures increase: note the decrease in smearing at 37°C and the decreased amount of non-specific complex in the well of the 400ng-$C43_s$ lane. Because of the effect of temperature on the non-specific binding of $C43_s$ to the T6-promoter sequence, 50°C and 65°C were also tested.

T6 Promoter and $C43_s$ Protein Binding at 50°C and 65°C

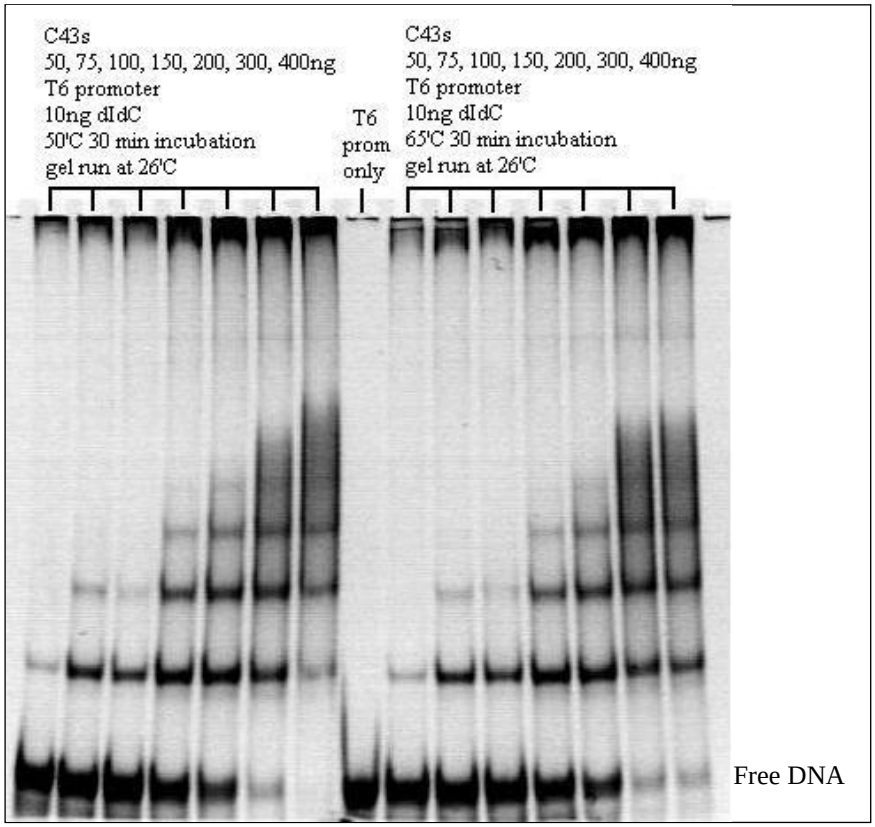


Figure 22: Binding Gel #4: C43 and T6 promoter binding assays at 50°C and 65°C.

In Figure 22 binding gel #4 is shown which illustrates the binding of $C43_s$ to the T6-promoter sequence at incubation temperatures of 50°C and 65°C. The gel was cast as described in Materials & Methods. It is remarkable to note the clear and specific binding

exhibited by $C43_s$ at these elevated temperatures, temperatures which would normally destroy nonthermal proteins.

T5 and T6 Promoters and $C43_s$ Protein Binding at 37°C

Both the T5-promoter and the T6-promoter sequences appear to be specific binding sites of $C43_s$. Of interest was a potential difference in protein/DNA complex formation between these two promoters. A binding assay comparing the binding of $C43_s$ to the T5 and T6 promoters was performed using the protocol outlined in Materials & Methods. However, samples were incubated at 37°C only.

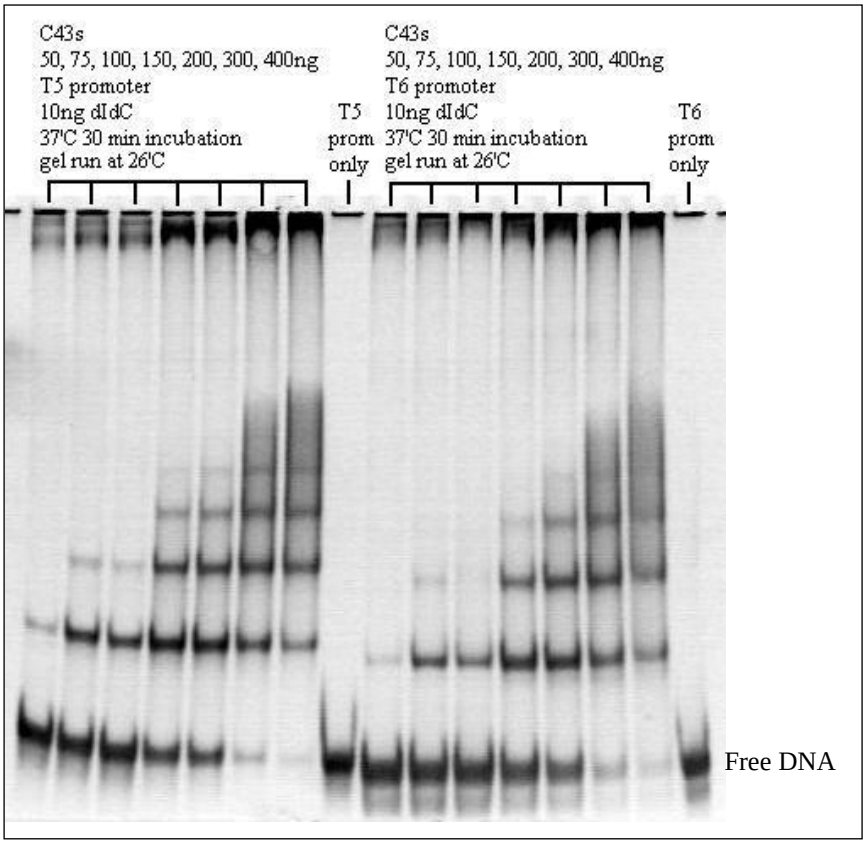


Figure 23: Binding Gel #5: Comparison of C43 & T5 and C43 & T6 promoters. Binding assays performed at 37°C..

In Figure 23 binding gel #5 illustrates the binding of $C43_s$ to the T5 and T6 promoter sequences at an incubation temperature of 37°C. The gel was cast as described in Materials & Methods. In comparing the formation of sequential complex formation, the disappearance of free DNA, and the initiation of non-specific complexes, there appears to be no difference in promoter/$C43_s$ complex formation between the T5 and T6 promoters. However, this may be due to a procedural practice of incubating the binding reactions at a specified temperature (26°C, 37°C, 50°C, or 65°C) and running all gels at 26°C. The transition from incubation temperature to gel run temperature may mask any subtle binding differences between the T5 and T6 promoter DNA complexes. To investigate this possibility, binding gel #5 was redone with sample incubation and gel run temperatures of 50°C.

T5 and T6 Promoters and $C43_s$ Binding at 50°C

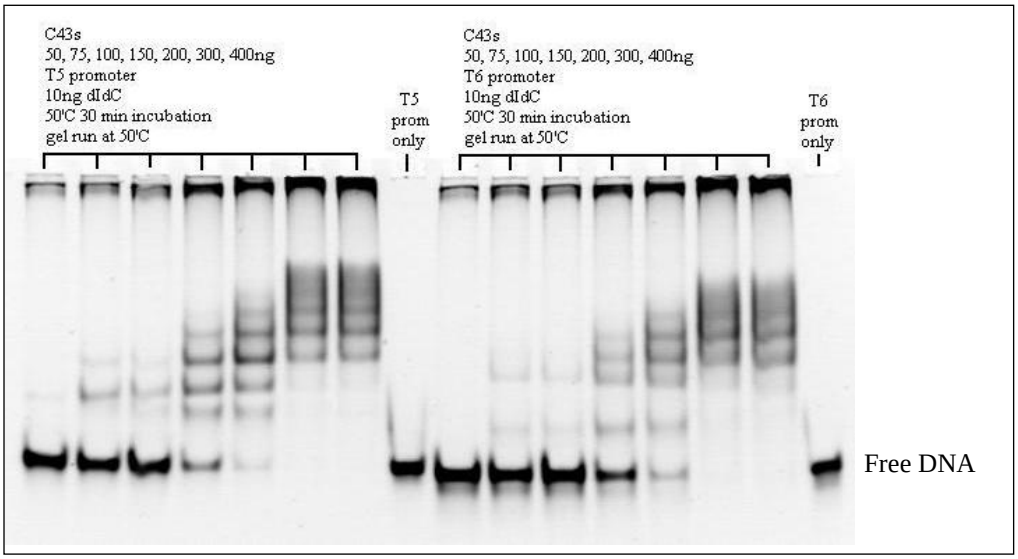


Figure 24: Binding Gel #6: Comparison of $C43_s$ & T5 and $C43_s$ & T6 promoters. Binding Assays performed at 50°C and gel run at 50°C.

Binding gel #6 shown in Figure 24 illustrates the binding of $C43_s$ to the T5 and T6 promoter sequence at an incubation temperature of 50°C and a gel run temperature of 50°C. The gel was cast as described in Materials & Methods. As seen in binding gel #5, $C43_s$ appears to bind the double direct repeats of the T5 and T6 promoters in a sequential manner. Also the initiation of non-specific binding upon double direct repeat saturation can be seen.

In addition to these observations the T5-promoter sequence appears to be the preferred binding site of $C43_s$ as evidenced by the rapid disappearance of free DNA (left portion of gel). However, very interesting is the observation that the amount of DNA shifting into the multiple complex bands is proportional to the amount disappearing from the free DNA band for both the T5 and T6 promoters. This indicates that the difference in DNA shifts is due to differences in complex formation at comparable $C43_s$ concentrations and not to an increase in non-specific T5-promoter binding.

C43 Promoter and $C43_s$ Protein Binding at 37°C

The truncation of C80 to C43 results in a previously unrealized portion of non-coding sequence in the SSV1 genome. Of interest is the potential interaction of C43 with this putative C43-promoter. A binding assay to test the interaction of $C43_s$ with the C43-promoter followed the binding reaction protocol set forth in Materials & Methods.

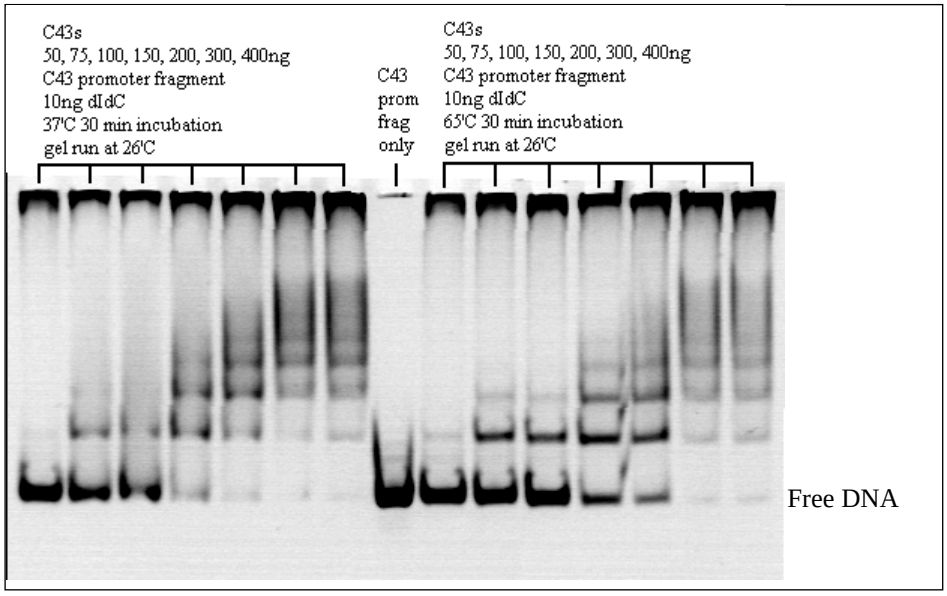


Figure 25: Binding Gel #7: $C43_s$ and C43 putative promoter binding assays at 37°C and 65°C.

In Figure 25 binding gel #7 is shown which illustrates the binding of $C43_s$ to the putative C43-promoter at an incubation temperature of 37°C and a gel run temperature of 26°C. The gel was cast as described in Materials & Methods. As $C43_s$ concentration increases (left to right) the amount of free DNA decreases and the number of complexes increases. The putative C43-promoter contains homology to the T5 and T6 promoters in that it contains six repeats of a five nucleotide sequence (Figure 26). Therefore, this sequence may be a binding site of $C43_s$.

GAAGGCTAGGGTTGA**ATACA**TCAAATTAC**CTAGA**TGTTACACAAAAACT
T**ATAGA**AAAATCGAAGCGAA**AAAGA**ACAACGACGGTACA**ATAGA**ATTA
ACG**TTAGA**GGAAACAATGCAAGT

Figure 26: Putative C43 Promoter Sequence of SSV1.

At first glance it appears that $C43_s$ binding of the putative C43 promoter is preferred at 37°C than at 65°C. At 37°C the free DNA does disappear more quickly than at 65°C. However, the amount of DNA shift into complex formation is not proportional

to the amount of disappearing free DNA. Additionally, complex formation at 37°C appears more smeared. In contrast, at 65°C complex formation appears more discrete. The free DNA does not disappear as quickly, but the amount of DNA shifted into complex formation is proportional to the amount of disappearing free DNA. This indicate that at 37°C there is more non-specific binding of the putative C43-promoter while at 65°C binding is more specific.

T5 and C43 Promoters and $C43_s$ Protein Binding at 50°C

Our observations of T5-promoter binding and putative C43-promoter binding by $C43_s$ indicate multiple rolls of $C43_s$ in SSV1 genome regulation. A binding assay was performed to compare $C43_s$ binding of the T5 and putative C43 promoters. The binding assay was carried out as outlined in Materials & Methods.

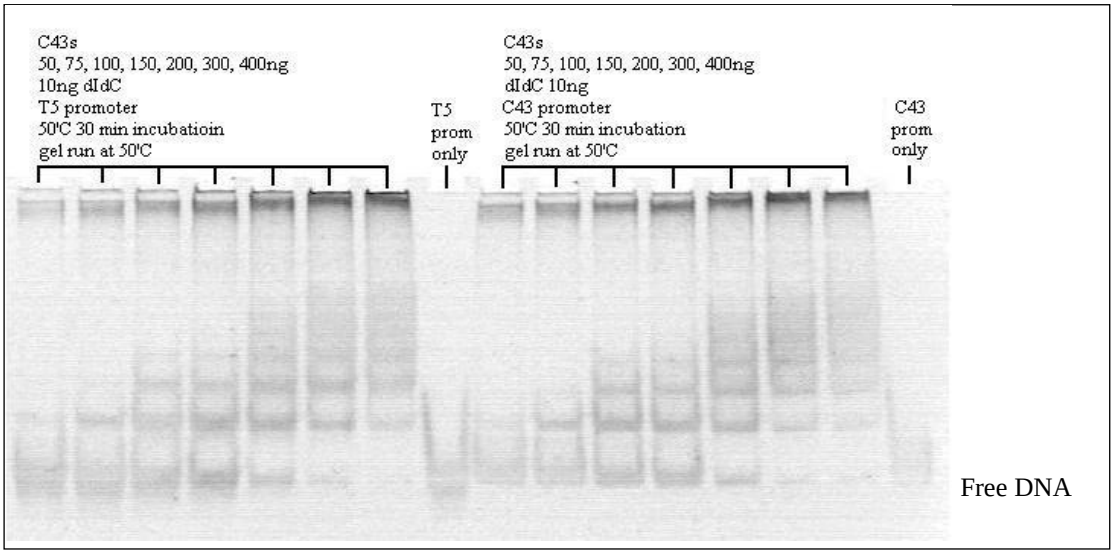


Figure 27: Binding Gel #8: $C43_s$ and T5-promoter and C43-promoter binding assays at 50°C.

In Figure 27 binding gel #8 is shown which illustrates the binding of $C43_s$ to the T5 and putative C43 promoters with an incubation and gel run temperature of 50°C. The

gel was cast as described in Materials & Methods. Multiple complexes of promoter/$C43_s$ form sequentially as $C43_s$ concentrations increase, while free DNA gradually disappears. The shift of DNA into complexes is comparable to the disappearance of free DNA for both promoters. Thus, there appears to be no difference in promoter/$C43_s$ complex formation between the two promoters.

However, comparison of the T5 and putative C43 promoter sequences does not support comparable binding of these promoters by $C43_s$. The T5-promoter contains double direct repeats (Figure 10) while the putative C43 promoter contains at best six sites homologous to those found in the T5 or T6 promoters (Figure 26). Thus, $C43_s$ binding of the putative C43 promoter should not be identical to the direct repeats of the T5-promoter.

The T5-promoter (101bp) and the putative C43 promoter (120bp) are comparable in length. The initial binding of $C43_s$ to the first binding site may trigger subsequent aggregative binding, resulting in multiple and smeared complexes seen at higher $C43_s$ concentrations. This aggrigative binding may account for the similarity in DNA shifts seen for the T5 and putative C43 promoters bound by $C43_s$.

However, binding of $C43_s$ to the T5, T6, or putative C43 promoters may be non-specific because RHH proteins have an inherent affinity for DNA and as we have shown the C43 RHH protein is stable at high temperatures. A binding assay to test the interaction of $C43_s$ with six non-protein-coding DNA fragments from the SSV1 genome was performed as outlined in Materials & Methods. The Assay was performed at 50°C and the gel run at 26°C. The gel was cast as outlined in Materials & Methods.

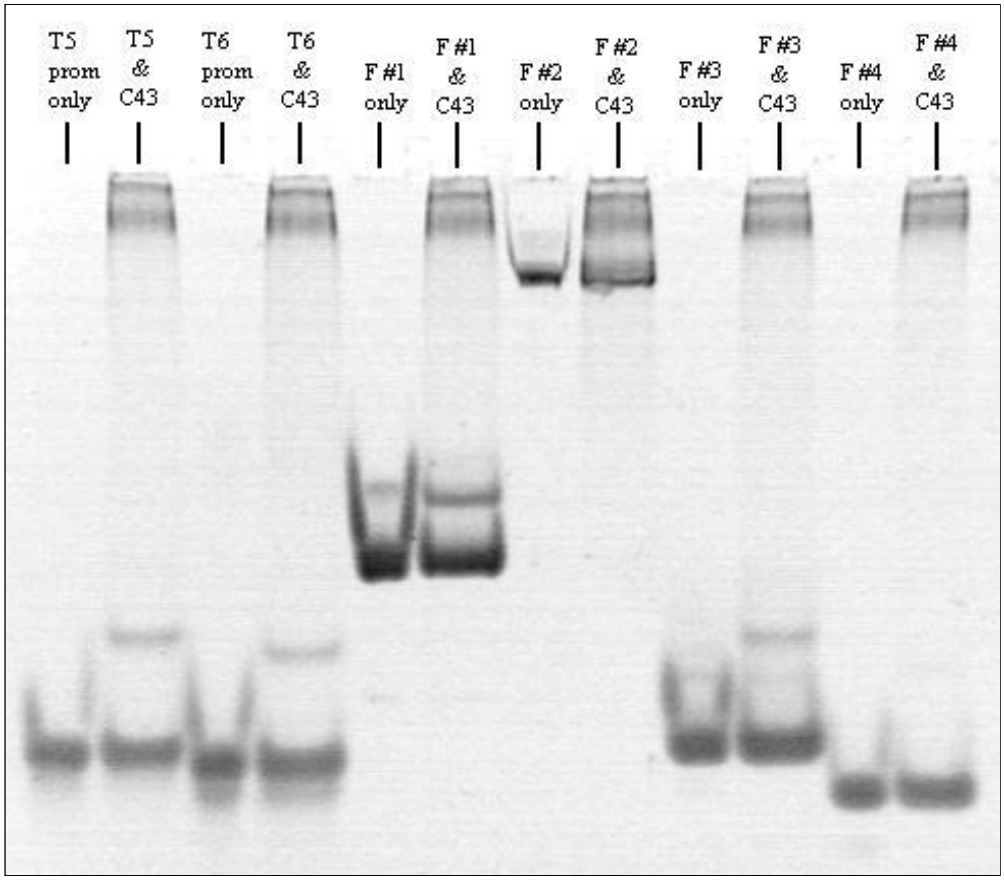


Figure 28: Binding Gel #9: C43 and six all non-coding sequences of the SSV1 genome.

Binding gel #9 (Figure 28) shows the interaction of $C43_s$ and six non-coding DNA fragments from the SSV1 genome. The T5 and T6 promoter as well as the putative C43 promoters all indicate binding by $C43_s$ by the shift of free DNA. The other three fragments show no binding by $C43_s$. Thus $C43_s$ appears to bind the T5, T6, and putative C43 promoters specifically. In addition, the high temperature at which the assay was performed argues for the thermo-stability and binding specificity of $C43_s$ to these promoters.

Phasing T5 and T6 Promotor Constructs and $C43_s$ Binding

The reproducible, sequential, and specific binding of $C43_s$ to the T5 and T6 promoters argues for the binding of $C43_s$ to the double direct repeats of these promoters. A single direct repeat of the T5 or T6 promoter was cloned into the six phasing constructs (Table 5) and binding assays performed with $C43_s$. As discussed in Materials & Methods, phasing constructs differ in length by 2bp. These 2bp additions increase the separation and shift the relative orientation between the inherent bend in the phasing construct and the bend induced by protein binding to the binding site; in this case T5/T6-double. Upon protein binding to a phasing construct a shift in DNA migration is expected with the amount of shift dependent upon the number of 2bp addition in a particular phasing construct. Figure 28 A illustrates the expected DNA shifts associated with protein binding to the six phasing constructs. Measurements of these shifts can be used to calculate the angle and amplitude of bend introduced into DNA as a result of a protein binding to its binding site.

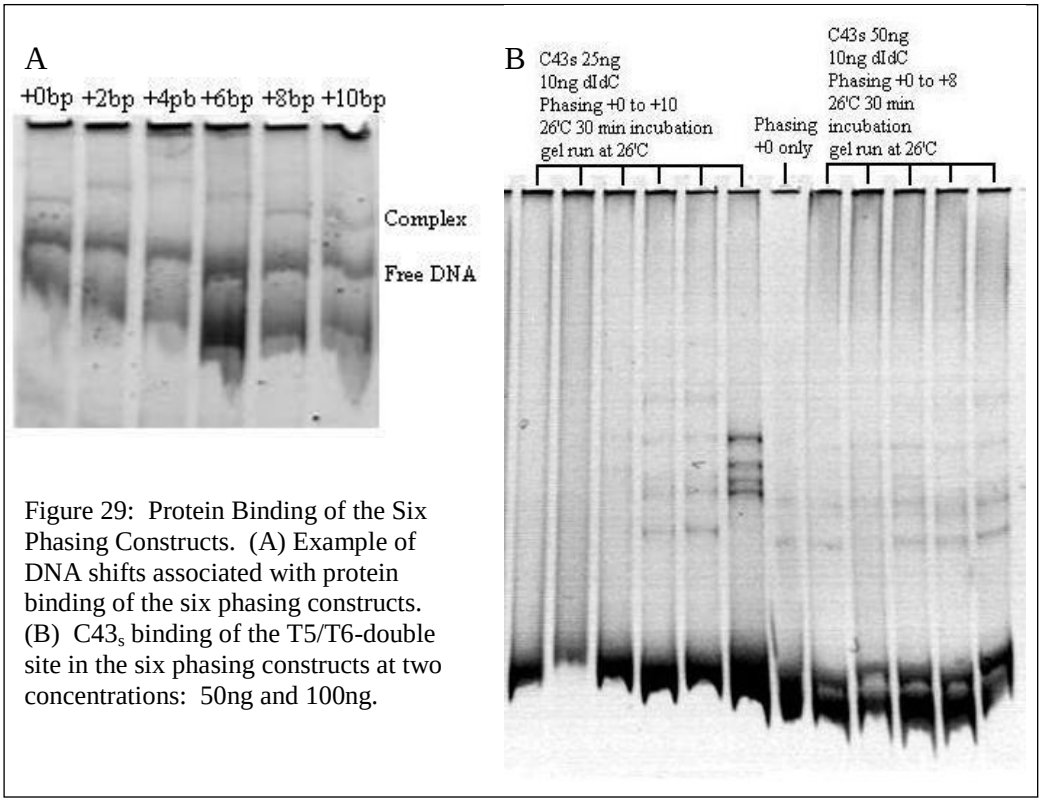


Figure 29: Protein Binding of the Six Phasing Constructs. (A) Example of DNA shifts associated with protein binding of the six phasing constructs. (B) $C43_s$ binding of the T5/T6-double site in the six phasing constructs at two concentrations: 50ng and 100ng.

Figure 29 B shows the binding of $C43_s$ to the T5/T6-double phasing constructs at concentrations of 50ng and 100ng. The gel was cast as described in Materials & Methods. Binding reactions were incubated at 26°C for 30 minutes and the gel was run at 26°C. Gel #10 lacks a distinguishable phasing binding pattern which may be due to the absence of phasing construct binding by $C43_s$. Conversely, the formation of numerous complexes may inhibit complex resolution. Re-examination of the phasing construct sequence as well as binding assay parameter manipulation is necessitated in order to resolve phasing data for $C43_s$. However, determining the actual binding sequence of $C43_s$ prior to performing further phasing assays would determine the binding usefulness of the cloned repeat within the phasing vectors.

CONCLUSIONS AND DISCUSSION

The Archaean *Sulfolobus spp.* are infected by the lysogenic viruses of the *Fuselloviridae* family: the SSVs. The genomes of four SSVs have been sequenced and contain about 34 ORFs. These ORFs and many of the host ORFs possess little sequence similarity to those in the public data bases. Of special interest to the biotechnology community are *Sulfolobus spp*. and their viruses because of their hyperthermophilic and acidophilic life styles.

C43, a RHH DNA binding protein of SSV1, was truncated from the original annotation of C80. This truncation was due to several observations made by M. Dlakic. First, the preceding 37 amino acids originally annotated in C80 would be the first example of coding sequence preceding the N-terminal ribbon of a RHH protein. Second, RHH proteins are typically 43 amino-acids in length. Third, the first 37 amino-acids of C80 do not align well with homologues from other sequenced SSV1 viruses. Finally, Martin Lawrence found that C80 is insoluble (M. Dlakic personal communication).

In addition to these observations, this work supports the truncation of C80. C43 was found to be soluble and active in that it binds multiple DNA sequences specifically. Additionally, we have shown that the 37 amino acids in front of C43 (former C80) are a non-coding sequence that is specifically bound by C43. Thus, C43 is the correct annotation of the previously annotated C80 ORF.

SSV1 proteins present challenges when attempting their manipulation in the laboratory. Limited methods for successful cloning of these proteins, their expression, and purification are available and have yet to be improved or established. C43 in

particular necessitated codon-optimization for the *E. coli* expression system before the native protein could be produced and purified. Expression of $C43_s$ is most productive at 37°C for 24 hours. Even so, a substantial portion of $C43_s$ is stored in inclusion bodies. Perhaps $C43_s$ is stored in inclusion bodies because it is a foreign protein within the *E. coli* system. However, codon optimization and high levels of expression yield enough $C43_s$ protein that some protein escapes inclusion-body aggregation and is located in the cytoplasm from where it can be purified.

In this work, $C43_s$ was cloned into an expression vector, expressed, purified, and its ability to bind three non-coding regions of the SSV1 genome assessed. These regions are the T5-promoter, the T6-promoter, and the putative C43-promoter.

It appears that $C43_s$ binds the T5 and T6 promoter double-tandem-repeats sequentially as $C43_s$ concentration increases. In addition, it appears that after several binding sites on the T5 or T6 promoter complex are occupied that $C43_s$ begins to aggregate on the promoter complex. The T5-promoter appears to be the better binding site of $C43_s$ (Figure 24). As temperatures increase, binding appears to become more specific for both T5 and T6 promoters and unlike non-thermophile proteins, $C43_s$ appears to maintain functionality at temperatures of 50ºC and 65ºC.

The putative C43-promoter region also appears to be a binding site of $C43_s$. Like the T5 and T6 promoter regions, as $C43_s$ concentration increases the number of complexes formed also increases. $C43_s$ also appears to bind the putative promoter region with higher specificity at 65°C, attesting again to the thermo-stability of the $C43_s$ protein. The T5-promoter, T6-promoter, and putative C43-promoter all contain multiple copies of

a homologous five nucleotide sequence, this sequence is a likely candidate for the preferred binding site of $C43_s$. Comparison of the T5 and putative C43 promoters indicated comparable binding by $C43_s$. However, further consideration of the promoter sequences suggest the initiation of aggregative $C43_s$ binding after binding site saturation and increased $C43_s$ concentration.

$C43_s$ appears to bind in a sequential manner. Because increased $C43_s$ concentration appears to trigger $C43_s$ aggregation it is likely that $C43_s$ dimerizes preferentially: a dimer of dimers. Once $C43_s$ binds to a site on the promoter complex, further C43s binding occurs and is enhanced by the dimer of dimer complex formation.

DNA binding proteins regulate the use of genetic information within a cell. C43 may act as a repressor or activator of the T5 and T6 transcripts and may also act in one of these capacities on the C43-promoter. One of the possible scenarios is that C43 production is somehow induced after host genome damage. Perhaps B55, the UV induced t_{ind} transcript, plays a role in activating transcription from the C43 coding sequence. C43 may go on to activate transcription of the T5 and T6 transcripts seen three hours after UV-irradiation. These transcripts contain all but two of the DNA binding proteins of SSV1 and therefore their transcription may induce events associated with replication and packaging. Activation of transcription from the T6-transcript would result in additional production of C43. At elevated concentrations, C43 may act as its own transcription repressor by binding to the C43-promoter region (perhaps negative feedback). Or perhaps C43 may act as a transcription activator of its own sequence (perhaps a positive feedback), resulting in further transcription of the T5 and T6

transcripts. The elucidation of the SSV1 transcription and translation control mechanism is sure to further our understanding of the *Sulfolobus spp.* and provide valuable insights which may be used in further developing the Archaea/virus system for molecular biological purposes.

REFERENCES CITED

1. Bell, S. D., and S. P. Jackson. 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol 6:222-8.

2. Bell, S. D., C. Jaxel, M. Nadal, P. F. Kosa, and S. P. Jackson. 1998. Temperature, template topology, and factor requirements of archaeal transcription. Proc Natl Acad Sci U S A 95:15218-22.

3. Brock, T. 1978. Thermophilic Microorganisms and Life at High Temperatures. Springer-Verlag, New York.

4. Cabello, P., M. D. Roldan, and C. Moreno-Vivian. 2004. Nitrate reduction and the nitrogen cycle in archaea. Microbiology 150:3527-46.

5. Cavicchioli, R., M. Z. Demaere, and T. Thomas. 2007. Metagenomic studies reveal the critical and wide-ranging ecological importance of uncultivated Archaea: the role of ammonia oxidizers. Bioessays 29:11-4.

6. Chaban, B., S. Y. Ng, and K. F. Jarrell. 2006. Archaeal habitats--from the extreme to the ordinary. Can J Microbiol 52:73-116.

7. Ciaramella, M., F. M. Pisani, and M. Rossi. 2002. Molecular biology of extremophiles: recent progress on the hyperthermophilic archaeon Sulfolobus. Antonie Van Leeuwenhoek 81:85-97.

8. Clark, K. L., E. D. Halay, E. Lai, and S. K. Burley. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 364:412-20.

9. de Rosa, M., GAMBACORTA, A. 1976. THE CALDARIELLA GROUP OF EXTREME THERMOACIDOPHILE BACTERIA: DIRECT COMPARISON OF LIPIDS IN SULFOLOBUS, THERMOPLASMA, AND THE MT STRAINS. Phytochemistry 15:143-145.

10. Duggin, I. G., and S. D. Bell. 2006. The chromosome replication machinery of the archaeon Sulfolobus solfataricus. J Biol Chem 281:15029-32.

11. Forterre, P., A. Bergerat, and P. Lopez-Garcia. 1996. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. FEMS Microbiol Rev 18:237-48.

12. Fox, G. E., L. J. Magrum, W. E. Balch, R. S. Wolfe, and C. R. Woese. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc Natl Acad Sci U S A 74:4537-4541.

13. Grogan, D., P. Palm, and W. Zillig. 1990. Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus Sulfolobus, Sulfolobus shibatae, sp. nov. Arch Microbiol 154:594-9.

14. Grogan, D. W. 1998. Hyperthermophiles and the problem of DNA instability. Mol Microbiol 28:1043-9.

15. Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417:63-7.

16. Huet, J., R. Schnabel, A. Sentenac, and W. Zillig. 1983. Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. Embo J 2:1291-4.

17. Kim, J., C. Zwieb, C. Wu, and S. Adhya. 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85:15-23.

18. Konneke, M., A. E. Bernhard, J. R. de la Torre, C. B. Walker, J. B. Waterbury, and D. A. Stahl. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437:543-6.

19. Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G. W. Nicol, J. I. Prosser, S. C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806-9.

20. Lipps, G. 2006. Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. Extremophiles 10:17-28.

21. Lundgren, M., A. Andersson, L. Chen, P. Nilsson, and R. Bernander. 2004. Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci U S A 101:7046-51.

22. Luscombe, N. M., S. E. Austin, H. M. Berman, and J. M. Thornton. 2000. An overview of the structures of protein-DNA complexes. Genome Biol 1:REVIEWS001.

23. Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher, and W. Zillig. 1984. SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. Embo J 3:2165-2168.

24. Muskhelishvili, G., P. Palm, and W. Zillig. 1993. SSV1-encoded site-specific recombination system in Sulfolobus shibatae. Mol Gen Genet 237:334-42.

25. Nadal, M., G. Mirambeau, P. Forterre, W. D. Reiter, and M. Duguet. 1986. Positively Supercoiled DNA in a Virus-Like Particle of an Archaebacterium. Nature 321:256-258.

26. Neme, A., and P. Miramontes. 2005. Biological domain identification based in codon usage by means of rule and tree induction. Computational Methods in Systems Biology 3082:221-224.

27. Ng, S. Y., B. Chaban, and K. F. Jarrell. 2006. Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications. J Mol Microbiol Biotechnol 11:167-91.

28. Ortmann, A. C., B. Wiedenheft, T. Douglas, and M. Young. 2006. Hot crenarchaeal viruses reveal deep evolutionary connections. Nat Rev Microbiol 4:520-8.

29. Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W. D. Reiter, and W. Zillig. 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. Virology 185:242-50.

30. Prangishvili, D., and R. A. Garrett. 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. Biochem Soc Trans 32:204-8.

31. Prangishvili, D., and R. A. Garrett. 2005. Viruses of hyperthermophilic Crenarchaea. Trends Microbiol 13:535-42.

32. Qureshi, S. A. 2007. Protein-DNA interactions at the Sulfolobus spindle-shaped virus-1 (SSV1) T5 and T6 gene promoters. Can J Microbiol 53:1076-83.

33. Qureshi, S. A., S. D. Bell, and S. P. Jackson. 1997. Factor requirements for transcription in the Archaeon Sulfolobus shibatae. Embo J 16:2927-36.

34. Raumann, B. E., B. M. Brown, and R. T. Sauer. 1994. Major Groove DNA Recognition by Beta-Sheets - the Ribbon-Helix-Helix Family of Gene Regulatory Proteins. Current Opinion in Structural Biology 4:36-43.

35. Reiter, W. D., U. Hudepohl, and W. Zillig. 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. Proc Natl Acad Sci U S A 87:9509-13.

36. Reiter, W. D., P. Palm, and S. Yeats. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res 17:1907-14.

37. Schleper, C., K. Kubo, and W. Zillig. 1992. The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. Proc Natl Acad Sci U S A 89:7645-9.

38. Schreiter, E. R., and C. L. Drennan. 2007. Ribbon-helix-helix transcription factors: variations on a theme. Nat Rev Microbiol 5:710-20.

39. She, Q., B. Shen, and L. Chen. 2004. Archaeal integrases and mechanisms of gene capture. Biochem Soc Trans 32:222-6.

40. She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost. 2001. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A 98:7835-40.

41. Snyder, J. C., K. Stedman, G. Rice, B. Wiedenheft, J. Spuhler, and M. J. Young. 2003. Viruses of hyperthermophilic Archaea. Res Microbiol 154:474-82.

42. Stedman, K. M., C. Schleper, E. Rumpf, and W. Zillig. 1999. Genetic requirements for the function of the archaeal virus SSV1 in Sulfolobus solfataricus: construction and testing of viral shuttle vectors. Genetics 152:1397-405.

43. Studier, F. W. 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207-34.

44. Szabo, Z., M. Sani, M. Groeneveld, B. Zolghadr, J. Schelert, S. V. Albers, P. Blum, E. J. Boekema, and A. J. Driessen. 2007. Flagellar motility and structure in the hyperthermoacidophilic archaeon Sulfolobus solfataricus. J Bacteriol 189:4305-9.

45. van der Oost, J., M. Ciaramella, M. Moracci, F. M. Pisani, M. Rossi, and W. M. de Vos. 1998. Molecular biology of hyperthermophilic Archaea. Adv Biochem Eng Biotechnol 61:87-115.

46. Wiedenheft, B., K. Stedman, F. Roberto, D. Willits, A. K. Gleske, L. Zoeller, J. Snyder, T. Douglas, and M. Young. 2004. Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol 78:1954-61.

47. Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev 47:621-69.

48. Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87:4576-9.

49. Wolfe, S. A., L. Nekludova, and C. O. Pabo. 2000. DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct 29:183-212.

50. Yeats, S., P. McWilliam, and W. Zillig. 1982. A plasmid in the archaebacterium Sulfolobus acidocaldarius. Embo J 1:1035-1038.

51. Zillig, W., D. Prangishvilli, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic, and D. Gotz. 1996. Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. FEMS Microbiol Rev 18:225-36.

52. Zinkel, S. S., and D. M. Crothers. 1987. DNA bend direction by phase sensitive detection. Nature 328:178-81.

APPENDIX A

SSV1 SEQUENCE CLONES AND DNA FRAGMENTS

The SSV1 coding sequences A79, A126, B115, B55, & B129 were cloned into various vectors.

Table 6: SSV Coding Sequence Constructs

Species	Backbone	Coding Seq.	Construct	Cell Line	Antibiotic
SSV1	pDS-56	A79+A45	pDS-56/A79	NEB 5	Amp
SSV1	pDS-56	A79+A45	pDS-56/A79	M15	Amp/Kan
SSV1	Pet 28b	A79+A45	Pet28b/A79	NEB 5	Kan
SSV1	Pet 28b	A79+A45	Pet28b/A79	BL21 DE3 pLysS	Kan/Chlor
SSV1	pDONR201	A79	pDONR201/A79	NEB 5	Kan
SSV1	pDEST14	A79	pDEST14/A79	NEB 5	Amp
SSV1	pDEST14	A79	pDEST14/A79	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDS-56	A126	pDS-56/A126	NEB 5	Amp
SSV1	pDS-56	A126	pDS-56/A126	M15	Amp/Kan
SSV1	Pet 28b	A126	Pet28b/A126	NEB 5	Kan
SSV1	pDS-56	A126	pDS-56/A126	BL21 DE3 pLysS	Amp/Chlor
SSV1	Pet 28b	A126	Pet28b/A126	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDONR201	A126	pDONR201/A126	NEB 5	Kan
SSV1	pDEST14	A126	pDEST14/A126	NEB 5	Amp
SSV1	pDEST14	A126	pDEST14/A126	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDEST14	A126 thr	pDEST14thr/A126	NEB 5	Amp
SSV1	pDS-56	B115	pDS-56/B115	NEB 5	Amp
SSV1	pDS-56	B115	pDS-56/B115	M15	Amp/Kan
SSV1	Pet 15b	B115	Pet15/B115	NEB 5	Amp
SSV1	pDONR201	B115	pDONR201/B115	NEB 5	Kan
SSV1	pDEST-14	B115	pDEST14/B115	NEB 5	Amp
SSV1	pDEST-14	B115	pDEST14/B115	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDS-56	C43	pDS-56/C43	M15	Amp/Kan
SSV1	pDONR201	C43	pDONR201/C43	NEB 5	Kan
SSV1	pDEST14	C43	pDEST14/C43	NEB 5	Amp
SSV1	pDEST14	C43	pDEST14/C43	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDEST14	C43 thr	pDEST14thr/C43	NEB 5	Amp
SSV1	pDONR201	C43 synth	pDONR201/C43s	NEB 5	Kan
SSV1	pDEST14	C43 synth	pDEST14/C43s	NEB 5	Amp
SSV1	pDEST14	C43 synth	pDEST14/C43s	BL21 DE3 pLysS	Amp/Chlor
SSV1	pDS-56	B55	pDS-56/B55	M15	Amp/Kan
SSV-RH	pDS-56	B43	pDS-56/B43	NEB 5	Amp
SSV-RH	pDS-56	B43	pDS-56/B43	M15	Amp/Kan
SSV-RH	pDONR201	B43	pDONR201/B43	NEB 5	Kan
SSV-RH	pDEST14	B43	pDEST14/B43	NEB 5	Amp
SSV-RH	pDEST14	B43	pDEST14/B43	BL21 DE3 pLysS	Amp/Chlor
SSV-RH	pDEST14	B43 thr	pDEST14thr/B43	NEB 5	Amp
SSV-RH	pDEST14	B43 thr	pDEST14thr/B43	BL21 DE3 pLysS	Amp/Chlor
SSV-RH	pDONR201	B43 synth	pDONR201/B43s	NEB 5	Kan
SSV-RH	pDEST14	B43 synth	pDEST14/B43s	NEB 5	Amp
SSV-RH	pDEST14	B43 synth	pDEST14/B43s	BL21 DE3 pLysS	Amp/Chlor

Due to the difficulty encountered when purifying some of the SSV1 coding sequences, GST tagged constructs were made for each coding sequences with engineered thrombin and factor-Xa cleavage sites.

Table 7: Expression Conditions for Cloned SSV Sequences

Coding Sequence	Expression System	Temperature	Time	Purification Conditions	Exchange/ Folding
A79	Auto-induction	37°C	24h	Denaturing	Dilute & Dialysis
A126	Auto-induction	37°C	24h	Denaturing	Dilute & Dialysis
B115	Auto-induction	37°C	24h	Denaturing	Dilute & Dialysis
$B43_s$	Auto-induction	37°C	24h	Native	Spin
$C43_s$	Auto-induction	37°C	24h	Native	Spin

Table 8: GST Clones

SSV1 coding Sequence	Backbone	Cleavage Site	Resulting Plasmid	Cleavage Site	Resulting Plasmid
A79	pDEST15-GST	Xa	pDEST15-GST-Xa/A79	Thrombin	pDEST15-GST-Thr/A79
A126	pDEST15-GST	Xa	pDEST15-GST-Xa/A126	**Thrombin**	pDEST15-GST-Thr/A126
B43	pDEST15-GST	Xa	pDEST15-GST-Xa/B43	**Thrombin**	pDEST15-GST-Thr/B43
$B43_s$	pDEST15-GST	Xa	pDEST15-GST-Xa/$B43_s$	Thrombin	pDEST15-GST-Thr/$B43_s$
B115	pDEST15-GST	Xa	pDEST15-GST-Xa/B115	Thrombin	pDEST15-GST-Thr/B115
C43	pDEST15-GST	Xa	pDEST15-GST-Xa/C43	Thrombin	pDEST15-GST-Thr/C43
$C43_s$	pDEST15-GST	Xa	pDEST15-GST-Xa/$C43_s$	**Thrombin**	pDEST15-GST-Thr/$C43_s$

Bold indicated sequence confirmed constructs. Gray indicated unmade constructs.

Table 9: SSV1 Binding Fragments

DNA Fragment	Sequence
T5 promoter 101bp	ATATTCCCATCCCACTCTATCTATACTCTATGTATAAATTAGTATTTA AGTCTTACTCTATCTATACTCTATCTATCTCTCTATATACACAGTGTTT GGGT
T6 promoter 99bp	GATAATATTAAATGATTCACGATATATAGATAGAGTATAGATAGAGT AAAGTTTAAATACTTATATAGATAGAGTATAGATAGAGGGTTCAAAA AATGG
Fragment #2 470bp	AATTCTGTCTGACTGCTGTCTGACAAGAGTTTACTCTATCTCTCTATA TCTATATACACAAACAGAGTTAGTCGACTCTGTGTATCTTATGTATCT TATACAAAAAATATGGGATGTGCAAAATCTGAGCTACTAATACTGCT TGAATATATAGATAGAGAGTGTAAGGACTACGAGAGTTGTAAAAGA ATAATAGTAGAGCTAGAAGAGAGAGTGAAGAAAATAGCTTTCGTAG AAGCAATAAATGATTTGTTCTAAACTACTTTTTTCTCTCTATCTCTATA TCTATATATATACATAACTAAAACTAAAAGAATAAACAAAAAACTAA CAAAATCAACTCACCATTATACAAACTCAGAAAAACTATTTTTTTGTT ATACTCTTACCCCATATATATATAGATATATAGATAGAGAGAGATAG AGTATAGTAGGGCATTTAAGATTTTAGAAGTTCTTCAATGCGTCT
Fragment #4 109bp	GATAATGCGAAACAAATTCAGATTCTGAATTCAGAACTGGAGGGGTT TAAAAACGTAAGCGGGAAGCCGATATTGACCAAGGATGAGACTCCT AAGGAGTTTGGCAAGAA

Table 10: List of Primers

Primer Name	Primer Sequence
SSV1 C43 5’	CAT CAC ATC GAT GGT CGT ATG CAA GTA ATA TCC TTT AAA CTA CCC
SSV1 C43 3’	ATT AAG CTT GGA TCC CTA AAC ATT TTC TAC ATA CCT CGC TAA
pDS56 C43 5’	TCC GAA GGA GAT AGA ACC ATG CAA GTA ATA TCC TTT AAA CTA
pDS56 C43 3’	GTG ATG GTG ATG GTG ATG AAC ATT TTC TAC ATA CCT CGC
attR1 & 2 C43 5’	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CGA AGG AGA TAG AAC C
attR1 & 2 C43 3’	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CCT AGT GAT GGT GAT GGT GAT G
pDONR screen C43 5’	TAA CGC TAG CAT GGA TCT C
pDONR screen C43 3’	GTA ACA TCA GAG ATT TTG AGA CAC
pDEST14 screen C43 5’	AGA TCT CGA TCC CGC GA
pDEST14 screen C43 3’	AAG ACC CGT TTA GAG GCC
C43s oligo 1	ATGCAGGTTATCTCTTTCAAACTGCCGCCGGCTCTGAACGCTAAA
C43s oligo 2	TTTCAGATTTAGATTTTTTTTCTTTGATAGCGATCTGTTCCAGTTTA GCGTTCAGAGCCGG
C43s oligo 3	CGCTATCAAAGAAAAAAAATCTAAATCTGAAATCATCCGTATCGC TCTGGCTCGTTACGTAG
C43s oligo 4	CTAGTGATGGTGATGGTGATGGACGTTCTCTACGTAACGAGCCAG AGC
C43s amplification 5’	TTC GAA GGA GAT AGA ACC ATG CAG GTT ATC TCT TTC AAA CTG C
C43s amplification 3’	CTA GTG ATG GTG ATG GTG ATG

Table 10 Continued: List of Primers Continued

attR1 & 2 C43s 5’	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CGA AGG AGA TAG AAC C
attR1 & 2 C43s 3’	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CCT AGT GAT GGT GAT GGT GAT G
T5 5’	TCC TCT AGA ACC CAA ACA CTG TGT ATA TAG AGA G
T5 3’	TGC GTC GAC ATA TTC CCA TCC CAC TCT ATC
T6 5’	TCC TCT AGA GAT AAT ATT AAA TGA TTC ACG ATA T
T6 3’	TGC GTC GAC CCA TTT TTT GAA CCC TCT AT
C80 promoter 5’	TGA AGG CTA GGG TTG AAT AC
C80 promoter 3’	CCT GCA TTG TTT CCT CTA AC

Binding assays were performed with B115, an SSV1 DNA binding protein. The binding assay protocol and gel casting methods are outlined in Materials & Methods.

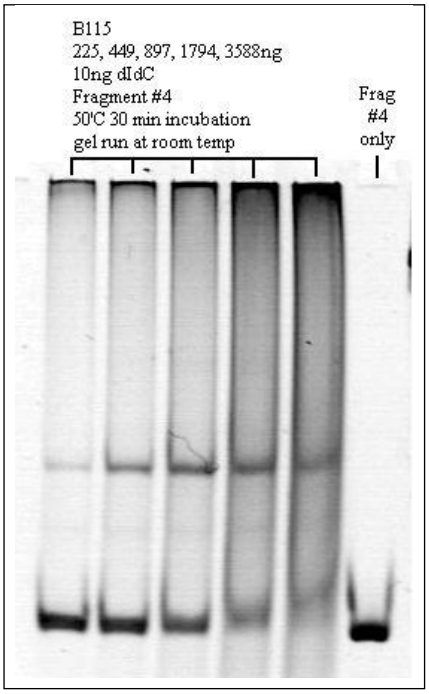


Figure 30: Binding Gel of B115 and Fragment #4