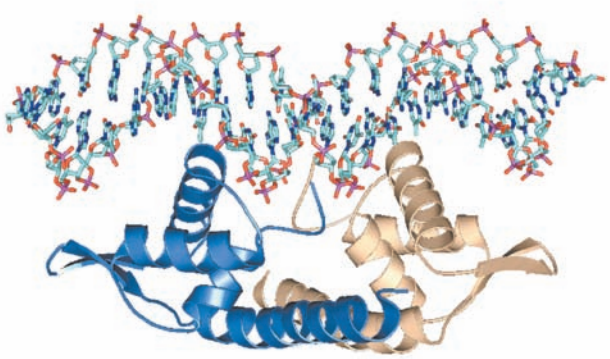
Structural Studies of Crenarchaeal Viral Proteins: Structure Suggests Function

Paul Kraft[1,2] | George H. Gauss[2] | Mark Young[1,3] | C. Martin Lawrence[1,2*]

[1] Thermal Biology Institute, Montana State University, Bozeman
[2] Department of Chemistry and Biochemistry, Montana State University, Bozeman
[3] Department of Microbiology, Montana State University Bozeman

*Corresponding Author:
Department of Chemistry and Biochemistry
108 Gaines Hall
Montana State University
Bozeman, MT 59717

Phone: 406.994.5382 **Fax:** 406.994.5407 **E-mail:** lawrence@chemistry.montana.edu



ABSTRACT

Sulfolobus spindle viruses, or Fuselloviridae, are ubiquitous crenarchaeal viruses found in high temperature acidic hot springs around the world (pH ≤ 4.0, T ≥ 70°C), and are especially prevalent in Yellowstone National Park. Because they are relatively easy to isolate, they represent the best studied of the crenarchaeal viruses. This is particularly true for the type virus, *Sulfolobus* spindle virus one (SSV1). SSV1 contains a double-stranded DNA genome of 15.5 kb, encoding 34 putative open reading frames (ORFs). Interestingly, the genome shows little sequence similarity to organisms other than its SSV homologues. Previous to our work, the limited sequence similarity combined with biochemical analysis had suggested functions for only five of the 34 putative ORFs. Thus, despite its position as the best-studied crenarchaeal virus, functions for most (29) of the SSV1 ORFs remain unknown. In an effort to assign functions to these proteins, we have undertaken biochemical and structural studies of the SSV1 proteome. We have completed structures for two of these proteins. In both cases, the structures reveal well-known folds. This suggests that the lack of sequence similarity in the SSV1 genome is not generally indicative of novel protein folds or functions. Rather, it is indicative of significant evolutionary distance between members of the SSV1 proteome and their homologues in the protein databases. Further, the use of structural homology and the analysis of conserved surface properties suggest functions for both of these SSV1 proteins. This demonstrates the general utility of structural studies on proteins of unknown function; specifically, that structure can predict function.

Key Words

archaeal viruses
crenarchaeal viruses
SSV
SSV1
structural genomics
structural proteomics

1.0 INTRODUCTION

More than 5,000 eukaryotic viruses and bacterial phage have been described. In stark contrast, fewer than 50 archaeal viruses are now known, and many of them are only poorly characterized (Rice et al. 2004). Of these, *Sulfolobus* spindle-shaped virus 1 (SSV1) is the most extensively studied, and thus serves as a paradigm for crenarchaeal viruses. Further, studies of SSV1 were instrumental in highlighting transcriptional similarities between archaeal and eukaryotic organisms (Reiter et al. 1988) and in the acceptance of *Archaea* as a separate domain of life (Zillig et al. 1985; Woese et al. 1990).

SSV1 was the first crenarchaeal virus to be described in detail. Originally isolated from *Sulfolobus shibatae* (Martin et al. 1984; Grogan et al. 1990), it has also been shown to be lysogenic in other species of *Sulfolobus*. SSV1 is characterized by a 60 x 100 nm lemon-shaped virion with tail fibers emanating from one end (Zillig et al. 1996). Packaged within the virion is a 15.5 kb circular double-stranded DNA genome (Zillig et al. 1996). The complete nucleotide sequence for SSV1 has been determined, and 34 open reading frames (ORFs) have been identified (Palm et al. 1991).

Characterization of SSV1 necessitated the creation of a new viral family, the Fuselloviridae. Fuselloviridae members are commonly found in high temperature (≥70°C), acidic (pH ≤ 4) hot springs around the world, and are particularly prevalent in Yellowstone National Park. Five additional viruses are now tentatively assigned to this family (Zillig et al. 1998; Stedman et al. 2003; Wiedenheft et al. 2004), and genomic sequencing for three of these—SSV2 (Stedman et al. 2003), SSV RH (Wiedenheft et al. 2004), and SSV KM (Wiedenheft et al. 2004)—is now complete. Stedman et al. (2003) found that the SSV1 and SSV2 genomes share 55% identity at the nucleotide level; and that each has 34 putative ORFs, with 27 ORFs common to both genomes. Sequence identity at the amino acid level for these 27 ORFs varies from 75.9% to 16.3%. Subsequent analysis of the SSV RH and SSV KM genomes by Wiedenheft et al. (2004) served to identify 18 ORFs that are conserved across all four SSV genomes. Further, the relative organization of the conserved ORFs within all four genomes was found to be similar, and nucleotide sequences of many of the putative promoters are also conserved (Stedman et al. 2003; Wiedenheft et al. 2004). Analysis of the viral transcripts by Reiter et al. (1987b), and subsequently by Palm et al. (1991), identified eleven distinct transcripts (T_1 - T_{11}). Transcripts T_1 through T_9 are abundant in lag-phase cultures infected with SSV1, while transcripts T_{10} and T_{11} require UV induction. All 34 ORFs are encoded in transcripts T_1 through T_9, suggesting that the viral genes are translated in a polycistronic manner.

Despite extensive study of SSV1, functions have been suggested for only five of the 34 ORFs. Sequence analysis revealed similarity between D-335 and the integrase genes belonging to the type I tyrosine recombinase family (Argos et al. 1986). The names of the ORFs and corresponding proteins, such as D-335, derive from the combination of the six possible reading frames (A-F) and the number of encoded amino acids. Thus ORF D-335 is in reading frame D and codes for 335 amino acids. D-335 has been expressed, purified, and characterized, confirming tyrosine integrase activity (Muskhelishvili et al. 1993). B-251 shows limited sequence similarity to the nucleotide binding protein DnaA (Koonin 1992). And three structural proteins (VP1, VP2, and VP3) have been isolated from purified virus particles and identified by N-terminal sequencing (Reiter et al. 1987a). Interestingly, a VP2 homologue was not found in SSV2, SSV RH or SSV KM. The functions of the remaining 29 ORFs remain unknown, and direct genetic and biochemical examination will be required to elucidate their functions.

Does the lack of significant sequence similarity indicate novel protein function? While this may be true for select SSV1 proteins, we believe that most of the SSV1 proteome is homologous at some level to proteins of known function. We believe that the lack of observed sequence similarity is a function of evolutionary distance and the temperature and pH extremes of the environment in which these viruses exist. In this regard, it is now well established that tertiary structural similarities persist far longer than similarities in either the nucleotide or amino acid sequence. Thus, structural similarities can suggest distant evolutionary relationships between proteins. Further, just as in primary

sequence analyses, the homology between structurally related proteins may indicate a related function. It follows then, that structural analysis of the SSV1 proteome could be used to identify structural and functional homologues that are not evident from primary sequence analysis, and thus provide important insights into the biochemistry of these viral proteins. Towards this goal, we have undertaken structural characterization of the SSV1 proteome. In cases where there is no obvious structural homology, the structures may still provide important clues to the functions of these proteins. For example, analysis of the surface properties of these proteins might reveal putative binding sites for macromolecules, or active sites that can accommodate small molecules. Further, docking exercises that screen virtual databases of biologically relevant small molecules might suggest possible substrates or ligands. Here we examine the initial structural results of our efforts, and their implications for the SSV1 proteome and for crenarchaeal viruses in general.

2.0 MATERIALS AND METHODS

A brief description of the materials and methods is presented here. For greater detail, please consult the work by Kraft et al. (2004a, 2004b).

2.1 Cloning

ORFs D-63 and F-93 were amplified by nested PCR from SSV1 genomic DNA prepared as described (Yeats et al. 1982; Schleper et al. 1992; Stedman et al. 2003; Wiedenheft et al. 2004). The polymerase chain reaction (PCR) primers added a Shine-Delgarno sequence, a C-terminal His_6-tag, and attB sites to facilitate ligase-free cloning using Invitrogen's Gateway™ Cloning Technology. The His-tagged constructs were then inserted into *Escherichia coli* destination vector pDEST14 (Invitrogen), yielding the expression vectors pDest14/D-63 and pDest14/F-93.

2.2 Expression and Purification

Native protein was expressed in BL21(DE3)-RIL *E. coli* (Novagen) and selenomethionine labeled protein in Novagen's B834(DE3). The selenomethionine incorporated proteins were used in the subsequent structure determinations and exist with selenomethionine in place of the native methionine residues. Protein expression was induced with 1 mM IPTG. D-63 was purified utilizing immobilized metal affinity chromatography (Ni-NTA, Qiagen), followed by anion exchange chromatography (Q anion exchange resin, Amersham). The F-93 purification also utilized immobilized metal affinity chromatography (Ni-NTA, Qiagen), followed by hydrophobic interaction chromatography (butyl-Sepharose, Amersham) and size exclusion chromatography (Superdex S-75, Amersham). Protein concentrations were measured by Bradford assay (Bradford 1976) with BSA as protein standard. Molecular weight and purity were assessed with SDS-PAGE and MALDI-TOF MS. MALDI-TOF MS was performed with a Bruker Biflex III in the matrix alpha-cyano-4-hydroxycinnamic acid.

2.3 Crystallization

D-63 and F-93 were both crystallized by hanging drop vapor diffusion. Drops were assembled with 2 µL of protein mixed with 2 µL of well solution. For D-63, the well solution was composed of 0.2 M $(NH_4)_2SO_4$, 0.2 M NH_4^-, 25% PEG 4000, pH 4.7 to pH 5.0, and drops were incubated at 18°C. Single crystals were stepped through a series of well solutions supplemented with increasing glycerol concentration to a final concentration of 25%. Crystals were then plunge-frozen in liquid nitrogen. The well solution for F-93 consisted of 3-3.5 M NaCl, 120 mM HEPES pH 7.4-7.8, and drops were incubated at 4°C. Single crystals of F-93 were isolated and dipped (10-30 seconds) in well solution supplemented with 25% glycerol and plunge-frozen in liquid nitrogen.

2.4 Data Collection

The structures were determined by multiwavelength anomalous diffraction. Data sets to 3.0 Å resolution were collected at three wavelengths for both crystals (F-93 and D-63) at BioCARS beamline 14-BM-D at the U.S. Department of Energy's Advanced Photon Source. In addition, a single wedge of higher resolution data was collected to 2.7 Å at the remote wavelength. Data were integrated and reduced in space group $P4_332$ and $P6_2$ respectively using the HKL software package (Otwinowski and Minor 1997).

2.5 Structure Determination and Refinement

In each case, SOLVE (Terwilliger and Berendzen 1999) was used to identify positions of the selenium substructure, and for calculation of the initial phases. The initial phases were improved by density modification including averaging over the non-crystallographic symmetry followed by phase extension to highest resolution (Bailey 1994; Cowtan 1994). The resulting electron density maps were of excellent quality and were used to build the initial models with the program O (Jones et al. 1991). Iterative rounds of refinement with REFMAC5 (Bailey 1994; Murshudov et al. 1997) and manual rebuilding with O yielded a final model for F-93 with an R_{cryst} of 19.4% (R_{free} of 25.9%), and a final model for D-63 with an R_{cryst} of 22.5% (R_{free} of 27.0%). Both models exhibit good stereochemistry, with an absence of residues in the disallowed regions of the Ramachandran plot (Laskowski et al. 1993). Structural comparisons were performed using the DALI (Holm and Sander 1993) server (**http://www.ebi.ac.uk/dali**) and VAST server (**http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html**). Figures were generated with PyMOL (**http://www.pymol.sourceforge.net**).

2.6 Coordinates

Atomic coordinates and structure factors have been deposited into the Protein Data Bank (PDB) under accession code 1TBX for F-93 and 1SKV for D-63.

3.0 RESULTS AND DISCUSSION

We have determined structures for two members of the SSV1 proteome—D-63 and F-93. D-63 encodes a 63 amino acid polypeptide in frame D, while F-93 codes for a 93-residue polypeptide in frame F of the SSV1 genome.

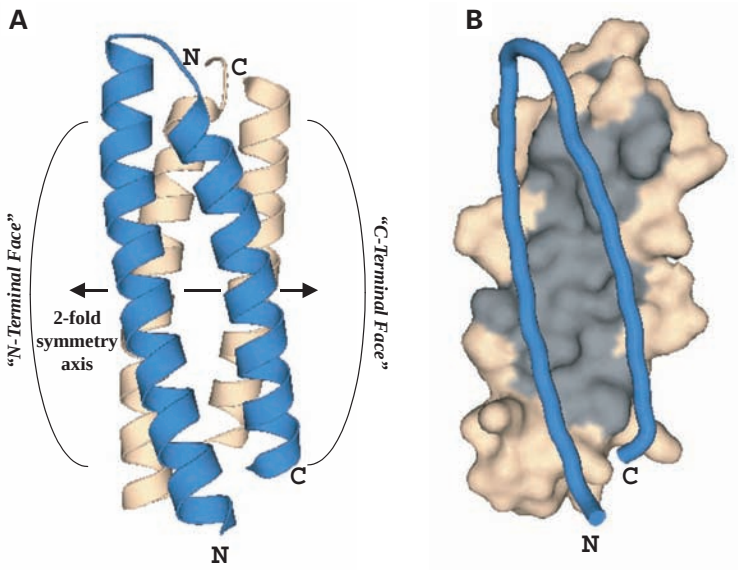


Figure 1. A. The D-63 homodimer forms an antiparallel four-helix bundle. Monomers A (blue) and B (beige) are each shown as a ribbon diagram depicting secondary structure elements. The two-fold symmetry axis, represented as a black double-headed arrow, runs horizontally between the two monomers relating the N- and C-termini of the blue monomer at the bottom of the figure to the N- and C-termini of the beige monomer at the top. **B.** The dimer interface is hydrophobic. The beige C_α trace of monomer A is depicted in blue, while B is shown in a beige and gray colored surface representation. All surface residues in B are colored beige with the exception of residues Phe^{11}, Leu^{14}, Val^{18}, Leu^{21}, Ile^{25}, Ile^{28}, Leu^{42}, Ala^{45}, Val^{49}, Ile^{52}, Leu^{56}, and Leu^{59}, which are depicted in gray, and illustrate the extent of the hydrophobic surface found at the dimer interface (Kraft et al. 2004a).

3.1 Structure and Function of SSV1 D-63

The crystal structure of D-63 (Kraft et al. 2004a) revealed a pair of long α-helices that run anti-parallel to each other. The connecting loop is composed of residues 35 through 39 (**Figure 1B**). This helix-turn-helix motif is quite common; however, it usually occurs within the context of a larger protein. In the case of D-63, two such monomers are tightly packed against each other to form a dimer, resulting in an anti-parallel four-helix bundle. While dimerization of the helix-turn-helix motif is common, the two-fold axis usually runs parallel to the long axis of the helical bundle. In this case, however, the two-fold axis is unusual in that it runs perpendicular to the long axis of the four-helix bundle. Thus, as one looks down the dimer axis, the four-helix bundle will present one face composed

of the two N-terminal helices (N-terminal face), or the opposite face composed of two C-terminal helices (C-terminal face). The dimer interface is largely hydrophobic, and the presence of the hydrophobic core is discernible in the primary sequence as a clear heptad repeat (**Figure 1A**). The dimer is also stabilized by intermolecular salt bridges and hydrogen bonds. Thus, packing of D-63 within the crystal strongly suggests that D-63 is a homodimer, a finding that is consistent with the behavior of the protein on size exclusion chromatography columns.

The PDB was searched for structural similarities using the DALI server. The search returned a multitude of similar structures that generally fell into one of two categories. The first was a group of larger proteins in which the helix-turn-helix motif was merely a super-secondary structural element. The second was a group of helical bundles. The latter group provided an excellent example of the

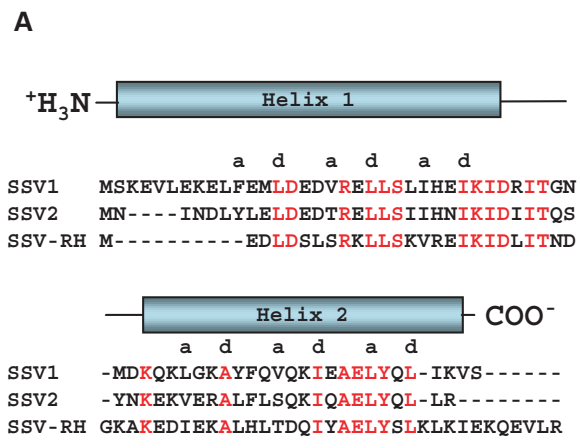


Figure 2A. Sequence alignments of D-63 fusellovirus homologues. Strictly conserved residues are colored red, non-conserved residues are in black. Secondary structure assignments are indicated above the sequence. The hydrophobic residues occur within the alpha helices in a clear heptad repeat, $(\mathbf{a}bc\mathbf{d}efg)_n$, with hydrophobic residues at positions **a** and **d**. The three-way alignment shows that 20 of 63 residues, ~32%, are strictly conserved. Additional similarity is seen between the N-terminal end of SSV1 and the C-terminal end of SSV RH. The N-terminus of SSV1 contains three glutamate residues, two basic residues (both Lys) and a Leu/Val pair; while the C-terminus of SSV RH contains two glutamate residues (plus one glutamine), two basic residues (Lys and Arg) and a Val/Leu pair. While these residues are at opposite ends of the linear sequence, the hairpin fold of the helix-turn-helix motif places these residues at similar positions within the three dimensional structure of D-63.

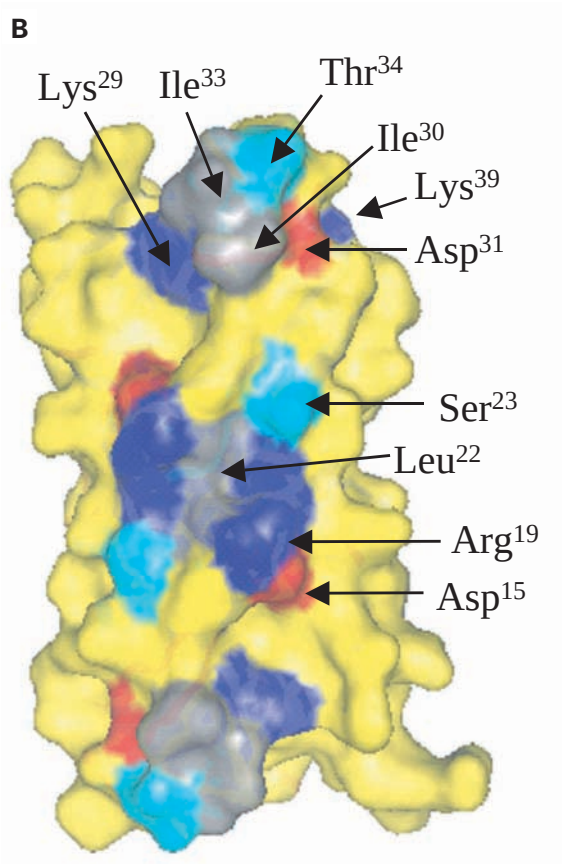


Figure 2B. Surface residues conserved among D-63 homologues. Conserved basic, acidic, polar, and nonpolar residues are shown in blue, red, cyan and gray, respectively. Non-conserved residues are colored in yellow. Strong conservation of SSV1, SSV2, and SSV-RH surface residues is seen for the "N-terminal" face (Figure 1A) of the four-helix bundle. The two-fold symmetry axis runs perpendicular to this face through the center of the four-helix bundle (perpendicular to the face of the page). Thus the conserved surface exhibits two-fold symmetry. Conserved residues contributing to this surface are labeled for one of the two monomers. The second monomer is related to the first by the two-fold symmetry axis. This putative ligand binding surface is further enlarged when conservative substitutions are also mapped to the surface (not shown). The surface is a mixture of both polar and nonpolar residues (Kraft et al. 2004a).

multitude of functions to which helical bundles have been adapted. However, upon close inspection of each of these, we were not able to identify any function with clear relevance to the SSV1 life cycle.

In sharp contrast to the search for function by structural homology, analysis of the surface properties of D-63 was quite enlightening. Generally speaking, residues on the surface of a protein are less likely to be conserved than residues buried within the core of the protein. However, exceptions occur when surface residues are required for the activity of the protein (Moult and Melamud 2000). In such cases, conservation of surface residues suggests the location of ligand binding sites, or in the case of an enzyme, the active site. Sequence alignments of D-63 with its homologues from SSV2 (D-57) and SSV RH (F-61) are shown in **Figure 2A**. When the strictly conserved residues are mapped to the surface of D-63, a substantial area of conserved surface is found on the N-terminal face of the four-helix bundle (**Figure 2B**). The strictly conserved residues are found primarily within the first (N-terminal) helix. Because the dimer axis passes through the N-terminal face of the four-helix bundle, this conserved surface patch is necessarily two-fold symmetric. The conserved surface area is extensive and devoid of obvious clefts or pockets that might accommodate a small molecule. These properties are suggestive of a macromolecular binding site. Thus, D-63 may function in macromolecular assembly.

Because the four-helix bundle of D-63 is two-fold symmetric, the conserved surface might recognize a macromolecule that itself possesses two-fold symmetry, or it could function to bind two copies of a monomeric entity, effectively serving to dimerize the

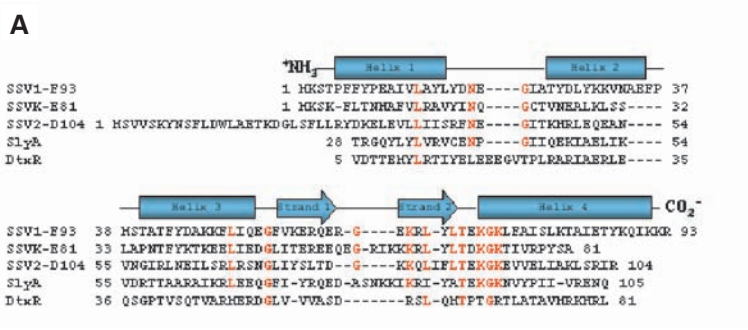


Figure 3A. Multiple sequence alignment for structural homologues of SSV1 F-93. The SSV1 F-93 sequence is aligned with its fuselloviral homologues from SSV2 (D-104/D-106) and SSV KM (E-81). Also included are structure-based alignments between SSV1 F-93, SlyA, and DtxR. Secondary structural elements identified in the SSV1 F-93 structure are mapped above the sequence. Strictly conserved residues in the fuselloviral homologues are shown in red. The cluster of conserved residues in the connecting loop between strand 2 and helix 4 are predominantly surface residues giving rise to a patch of conserved surface at the N-terminal end of helix 4. The 28 residue extension at the N-terminus of SSV2 D-104 is noteworthy. The D-104 extension is the same length as the N-terminal extension in SlyA, suggesting that the SSV2 D-104 dimer interface is more complex, possibly containing an additional N-terminal α-helix at the dimer interface, as is seen for SlyA.

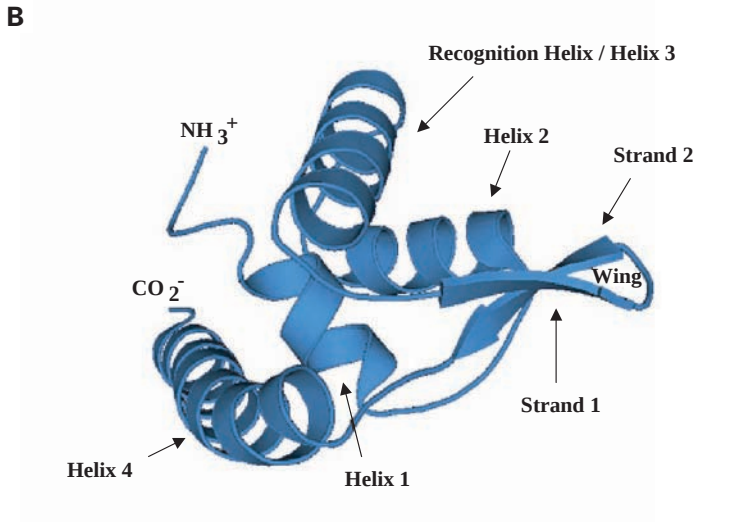


Figure 3B. Structure of the F-93 monomer. A ribbon diagram depicts the secondary structure elements of the F-93 monomer. The polypeptide chain threads its way through α-helices 1 through 3, followed by β-strands 1 and 2, and finally helix 4. The helix-loop-helix motif corresponds to α-helices 2 and 3, with helix 3 acting as the recognition helix that binds within the major grove of target DNA. The "wing" of the winged-helix motif is formed by β-strands 1 and 2, and is also usually involved in DNA recognition (Kraft et al. 2004b).

molecule. Presumably, formation of the complex with D-63 will serve to regulate or modify the behavior of these unidentified ligands, thus assisting in maintenance of the viral life cycle. From this perspective, experiments aimed at identification of potential binding partners are indicated. Possible approaches include two-hybrid screens and pull-down assays using coimmunoprecipitation or immobilization of D-63 via the C-terminal His_6-tag.

3.2 Structure and Function of SSV1 F-93

Purified F-93, with a monomer molecular weight of 11,780, migrates at a mass of 25,000 Da (results not shown) on a calibrated Superdex S-75 size exclusion chromatography column, suggesting that F-93 is present as a homodimer. In addition, chemical cross-linking experiments analyzed by SDS-PAGE show a strong conversion of monomer to cross-linked dimer, lending further support to the homodimer as the relevant biological assembly.

The structure of the F-93 monomer (Kraft et al. 2004b) is shown in **Figure 3B.** It reveals a common fold that places F-93 into the helix-turn-helix (HTH) super-family of DNA-binding proteins. This HTH motif is distinct from the helix-turn-helix motif described above for D-63. A search for structural nearest neighbors using DALI or VAST indicates that F-93 is most similar to the SlyA (Wu et al. 2003) and MarR (Alekshun et al. 2001) subfamilies of "winged helix" DNA-binding proteins (Gajiwala and Burley 2000; Alekshun et al. 2001; Wu et al. 2003). F-93 clearly shares a common structural core with both of these proteins. This common fold includes the first three helices of F-93, followed by

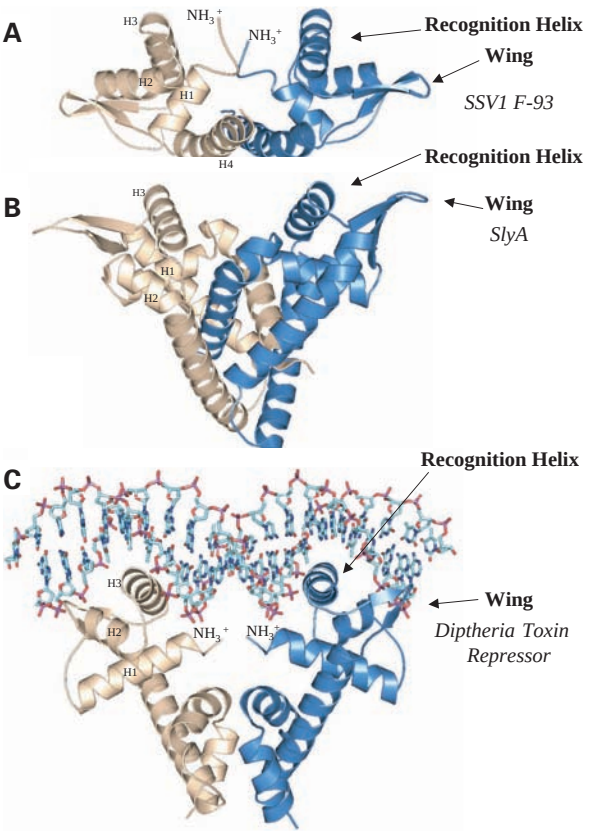


Figure 4. A. Structure of the F-93 dimer. The F-93 homodimer is depicted with chain A in blue and chain B in beige. The orientation of the A chain is identical to that of Figure 3B. The dimer axis runs vertically through the dimer interface (in the plane of the page), which is formed by the N-terminus—the N-terminal end of helix 1—and helix 4. The connecting loop between β-strands 1 and 2 is well ordered for the wing of chain A; however, the density is poor for this region in chain B, indicating conformational flexibility for the tips of the wings. The same is true for the extreme N-terminus of both chains. The conformational flexibility may be related to DNA recognition. **B.** Structure of the SlyA homodimer. The SlyA structure shows a similar winged-helix motif but with a more extensive and elaborate dimer interface. The positions of the recognition helix and the wing are indicated, and are found in orientations similar to those seen in F-93 (Figures 3B and 4A). The dimer axis runs vertically, coincident with that of F-93 in Figure 4A. The relative position of the winged-helix motif is determined by the dimer interface. Thus, in comparing the F-93 and SlyA homdimers, one sees a slight difference in the orientations of the winged-helix motifs. An optimal superposition of the monomers would require a slight rotation about the horizontal axis. **C.** Structure of the DtxR/DNA complex. The DtxR dimer axis runs vertically, passing through the pseudo-twofold axis of the DNA, and is coincident with the twofold axis of F-93 and SlyA. The N-terminus, the recognition helix, and the wing of DtxR mediate DNA recognition. Figures 4A through 4C also illustrate diversity in the relative orientation of the winged-helix motif that is used to tune the fit between homodimeric transcription factors and their target DNA (Kraft et al. 2004b).

the two anti-parallel β-strands (**Figure 3A**). The second and third helices constitute the helix-turn-helix motif, with the third helix functioning as the "recognition helix" that binds the major groove of target DNA. The two β-strands—C-terminal to the recognition helix—are connected by a reverse turn, and form the flanking "wing" of this structural motif.

This minimal fold is common to all members of the SlyA and MarR families of winged helix proteins and serves to define the motif. However, this motif is usually found within a larger polypeptide, with extensions at the N- or C-terminus. In the SlyA and MarR subfamilies, these extensions generally contribute to formation of the dimer interface. This is true for F-93 as well, where a C-terminal extension gives rise to a fourth helix. This helix, along with the N-terminus and the first half of helix one, form much of the F-93 dimer interface (**Figure 4B**). The dimer interface is substantial, and primarily hydrophobic.

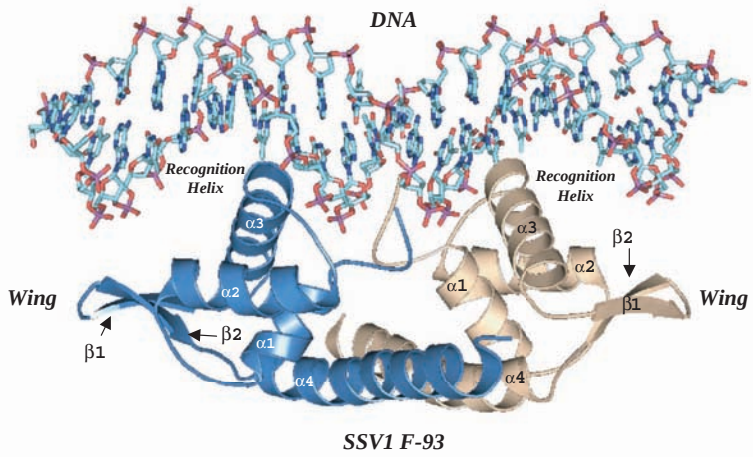


Figure 5. Modeling F-93/DNA interactions. The F-93 dimer is depicted in dark blue and beige with docked B-form DNA depicted in "stick" form. The interactions shown are analogous to those seen in the DtxR/DNA cocrystal structure. While DtxR binds to linear DNA, other winged helix proteins are known to promote bending of the DNA. We cannot rule this out for F-93. In either event, the recognition helix and wings would be expected to participate in DNA recognition.

How might F-93 bind DNA? The structure of a SlyA/DNA or MarR/DNA complex has yet to be determined. However, the complex has been solved for diphtheria toxin repressor (DtxR, PDB identifier 1F5T), which is also a close structural neighbor (Chen et al. 2000). DtxR shares the winged-helix core structure seen in F-93, with the addition of a C-terminal extension that gives rise to the dimer interface (**Figure 4C**). As expected, DtxR places the recognition helix in the major groove of the DNA, while the wings were found to interact with the ribose-phosphate backbone.

Based on the DtxR/DNA structure, we have modeled the putative F-93/DNA structure (**Figure 5**). With the exception of the recognition helix, which is presumably involved in base specific interactions, the electrostatic surface calculations show significant positive potential for the putative DNA-binding surface. This is particularly true for the wings, lending further support for a role in binding target DNA. In addition, the model includes four positive charges clustered about the dimer axis, in position to interact with the phosphodiester backbone. Two charges come from the N-termini themselves, and two more from the lysine side chain present as the second residue in the polypeptide chain. Similar to DtxR and MarR, this binding model suggests that F-93 will recognize a DNA palindrome, or pseudo-palindrome. The lack of a suitable palindrome within the SSV1 genome supports the pseudo-palindrome as the *in-vivo* target sequence. Alternatively, F-93 might be active *in vivo* as a heterodimeric complex with other winged helix proteins from SSV1 or its *Sulfolobus* host.

In comparing the structures of F-93, SlyA, and DtxR (**Figure 4A-C**), the SlyA and DtxR dimers clearly show much more extensive dimer interfaces. We are unaware of any other transcriptional regulators belonging to the winged helix family that are as small as F-93. In the case of DtxR, the complex dimer interface serves the added

function of a regulatory iron binding site, which modulates the activity of DtxR. Thus, the apparently simpler dimer interface observed in SSV1 F-93 could indicate a lack of regulation, suggesting that F-93 is constitutively active. Another possibility is that regulatory elements could be provided *in trans* by another viral or host protein, or regulation could occur through binding of a small molecule that would occlude the DNA-binding surface, as is seen with binding of salicylate to MarR (Alekshun et al. 2001).

The winged-helix fold has been adapted to other functions. Thus, even though F-93 shows greatest similarity to the DNA-binding domain of known transcriptional activators or repressors (SlyA, MarR, DtxR), alternative roles played by a few members of the winged helix family should at least be considered as an explanation for the small dimer interface. For example, F-93 shares structural similarities to the replication terminator protein of *Bacillus subtilis* (Wilce et al. 2001), where the dimer interface is also composed of a single C-terminal α-helix, suggesting that F-93 might participate in replication of the viral genome. This activity could be relevant, particularly since F-93 is encoded by viral transcript T_5. T_5 also encodes a DnaA-like protein, which could participate in initiation of DNA replication. However, rolling circle replication is a likely mechanism for propagation of the viral DNA (Koonin 1992). If that is indeed the case, it is not clear that SSV1 would have need for a replication terminator protein.

It is natural to ask whether there is structural similarity to other viral proteins. In this regard, there is a resemblance to the activation domain of MotA, a transcription factor from phage T4 (Finnin et al. 1997). However, the DALI match for MotA is weaker than those seen for transcription factors of the SlyA and MarR families. Although we cannot exclude an F-93 function in replication termination, or as an activation domain, the most likely function would seem to be that of a transcriptional regulator.

The structure determination and analysis presented here strongly suggest that F-93 binds viral or host DNA during the course of the SSV1 viral life cycle, possibly in the role of a transcription factor. Further, it suggests specific experiments to identify the putative target DNA. Possible approaches include electrophoretic mobility shift assays with viral DNA followed by DNA foot-printing, *in vitro* selection to identify an optimal DNA binding sequence, or the use of microarrays to examine host and viral gene expression in the presence and absence of F-93.

4.0 FURTHER CONCLUSIONS

It is noteworthy that SSV1 transcript T_5 encodes 10 of the 34 ORFs (Palm et al. 1991) found in SSV1. This includes D-63 and F-93, as well as the viral integrase and the DnaA-like protein. Because DnaA and integrase activities are typically early events in a viral life cycle, Palm et al. (1991) have suggested that all of the T_5 gene products will serve such early functions. Our results, which suggest adaptor protein function for D-63, and DNA binding activity for F-93, are consistent with this hypothesis.

The structural analysis presented here suggests specific avenues for future investigation. As this work is pursued, important details of the genetics and biochemistry of D-63 and F-93 will certainly emerge. The current study also highlights the utility of structural studies for proteins of unknown function. It clearly demonstrates that structure can suggest function.

ACKNOWLEDGMENTS

We thank the BioCARS staff for their excellent support during data collection at sector 14 of the Advanced Photon Source. We are grateful to Daniel Kümmel, Andrea Oeckinghaus, Philippe Benas, and Blake Wiedenheft for helpful discussions and technical assistance. This work was supported by grants from the National Science Foundation (MCB-0236344) and the National Aeronautics and Space Administration (NAG5-8807). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Use of BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

REFERENCES

Alekshun, M.N., S.B. Levy, T.R. Mealy, B.A. Seaton, and J.F. Head. 2001. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat Struct Biol* 8:710-14.

Argos, P., A. Landy, K. Abremski, J.B. Egan, E. Haggard-Ljungquist, R.H. Hoess, M.L. Kahn, B. Kalionis, S.V. Narayana, L.S. Pierson, III, et al. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* 5:433-40.

Bailey, S. 1994. The CCP4 suite - programs for protein crystallograpy. *Acta Cryst* D50:760-3.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-54.

Chen, C.S., A. White, J. Love, J.R. Murphy, and D. Ringe. 2000. Methyl groups of thymine bases are important for nucleic acid recognition by DtxR. *Biochemistry* 39:10397-407.

Cowtan, K. 1994. DM: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31:34-8.

Finnin, M.S., M.P. Cicero, C. Davies, S.J. Porter, S.W. White, and K.N. Kreuzer. 1997. The activation domain of the MotA transcription factor from bacteriophage T4. *Embo J* 16:1992-2003.

Gajiwala, K.S., and S.K. Burley. 2000. Winged helix proteins. *Curr Opin Struct Biol* 10:110-16.

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.

Holm, L., and C. Sander. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123-38.

Jones, T.A., J.Y. Zou, S.W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst* A47:110-19.

Koonin, E.V. 1992. Archaebacterial virus SSV1 encodes a putative DnaA-like protein. *Nucl Acids Res* 20:1143.

Kraft, P., D. Kummel, A. Oeckinghaus, G.H. Gauss, B. Wiedenheft, M. Young, and C.M. Lawrence. 2004a. Structure of D-63 from *sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. *J Virol* 78:7438-42.

Kraft, P., A. Oeckinghaus, D. Kummel, G.H. Gauss, J. Gilmore, B. Wiedenheft, M. Young, and C.M. Lawrence. 2004b. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. *J Virol* 78:11544-50.

Laskowski, R.A., M.W. MacArthur, D.S. Moss, and J.M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283-91.

Martin, A., S. Yeats, D. Janekovic, W.-D. Reiter, W. Aicher, and W. Zillig. 1984. SAV-1, a temperate UV inducible DNA virus like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J* 3:2165-8.

Moult, J., and E. Melamud, E. 2000. From fold to function. *Curr Opin Struct Biol* 10:384-9.

Murshudov, G.N., A.A. Vagin, and E.J. Dodson. 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst* D53:240-55.

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

Otwinowski, Z., and W. Minor. 1997. Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, Part A*, ed. C. Carter and R. Sweet, 307-26. New York: Academic Press.

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.

Reiter, W.D., P. Palm, A. Henschen, F. Lottspeich, W. Zillig, and B. Grampp. 1987a. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol Gen Genet* 206:144-53.

Reiter, W.D., P. Palm, S. Yeats, and W. Zillig. 1987b. Gene expression in archaebacteria: Physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol Gen Genet* 209:270-5.

Reiter, W.D., P. Palm, and W. Zillig. 1988. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. *Nucl Acids Res* 16:1-19.

Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J.E. Johnson, T. Douglas, and M. Young. 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci* 101:7716-20.

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* 89:7645-9.

Stedman, K.M., Q. She, H. Phan, H.P. Arnold, I. Holz, R.A. Garrett, and W. Zillig. 2003. Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res Microbiol* 154:295-302.

Terwilliger, T.C., and J. Berendzen. 1999. Automated MAD and MIR structure solution. *Acta Cryst* D55:849-61.

Wiedenheft, B., K. Stedman, F. Roberto, D. Willits, A.-K. Gleske, L. Zoeller, J. Snyder, T. Douglas, and M.J. Young. 2004.

Comparitive genomic analysis of hyperthermophilic archaeal fusselloviridae viruses. *J Virol* 78:1954-61.

Wilce, J.A., J.P. Vivian, A.F. Hastings, G. Otting, R.H. Folmer, I.G. Duggin, R.G. Wake, and M.C. Wilce. 2001. Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. *Nat Struct Biol* 8:206-10.

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* 87:4576-9.

Wu, R.Y., R.G. Zhang, O. Zagnitko, I. Dementieva, N. Maltzev, J.D. Watson, R. Laskowski, P. Gornicki, and A. Joachimiak. 2003. Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. *J Biol Chem* 278:20240-4.

Yeats, S., P. McWilliam, and W. Zillig. 1982. A plasmid in the archaebacterium *Sulfolobus solfataricus*. *EMBO J* 1:1035-8.

Zillig, W., H.P. Arnold, I. Holz, D. Prangishvili, A, Schweier, K. Stedman, Q. She, H. Phan, R. Garrett, and J.K. Kristjansson. 1998. Genetic elements in the extremely thermophilic archaeon *Sulfolobus. Extremophiles* 2:131-40.

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.

Zillig, W., R. Schnabel, and K.O. Stetter. 1985. Archaebacteria and the origin of the eukaryotic cytoplasm. *Curr Top Microbiol Immunol* 114:1-18.