JOURNAL OF VIROLOGY, Nov. 2004, p. 11544–11550 Vol. 78, No. 21
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.21.11544–11550.2004

Crystal Structure of F-93 from *Sulfolobus* Spindle-Shaped Virus 1, a Winged-Helix DNA Binding Protein

Paul Kraft,[1,2] Andrea Oeckinghaus,[1,3] Daniel Kümmel,[1,3] George H. Gauss,[2] John Gilmore,[2] Blake Wiedenheft,[1,4] Mark Young,[1,4] and C. Martin Lawrence[1,2]*

Thermal Biology Institute,[1] Department of Chemistry and Biochemistry,[2] and Department of Microbiology,[4] Montana State University, Bozeman, Montana, and Physiologisch-Chemisches Institut der Universität, Tübingen, Germany[3]

Received 26 March 2004/Accepted 23 June 2004

***Sulfolobus* spindle-shaped viruses (SSVs), or *Fuselloviridae*, are ubiquitous crenarchaeal viruses found in high-temperature acidic hot springs around the world (pH ≤4.0; temperature of ≥70°C). Because they are relatively easy to isolate, they represent the best studied of the crenarchaeal viruses. This is particularly true for the type virus, SSV1, which contains a double-stranded DNA genome of 15.5 kilobases, encoding 34 putative open reading frames. Interestingly, the genome shows little sequence similarity to organisms other than its SSV homologues. Together, sequence similarity and biochemical analyses have suggested functions for only 6 of the 34 open reading frames. Thus, even though SSV1 is the best-studied crenarchaeal virus, functions for most (28) of its open reading frames remain unknown. We have undertaken biochemical and structural studies for the gene product of open reading frame F-93. We find that F-93 exists as a homodimer in solution and that a tight dimer is also present in the 2.7-Å crystal structure. Further, the crystal structure reveals a fold that is homologous to the SlyA and MarR subfamilies of winged-helix DNA binding proteins. This strongly suggests that F-93 functions as a transcription factor that recognizes a (pseudo-)palindromic DNA target sequence.**

More than 5,000 eukaryotic viruses and bacterial phage are now known. In stark contrast, fewer than 35 archaeal viruses have been described. Of these, *Sulfolobus* spindle-shaped virus type 1 (SSV1) is by far the best studied and serves as a paradigm for the archaeal viruses. Further, studies of SSV1 have been instrumental in the acceptance of Archaea as a separate domain of life (28, 33) and in revealing similarities in the transcriptional machinery of archaeal and eukaryotic organisms (22).

SSV1 was the first crenarchaeal virus to be described in detail. Originally isolated from *Sulfolobus shibatae* (9, 15), it has also been shown to be lysogenic in other species of *Sulfolobus*. SSV1 is characterized by a 60- by 100-nm lemon-shaped virion with tail fibers emanating from one end (32). Packaged within the virion is a 15.5-kb circular double-stranded DNA genome (32). The complete nucleotide sequence for SSV1 has been determined, and subsequent analysis identified 34 open reading frames (ORFs) (20).

Identification of SSV1 necessitated the creation of a new viral family, the *Fuselloviridae*. The *Fuselloviridae* are ubiquitous in high-temperature (≥70°C), acidic (pH ≤4) hot springs around the world, and five additional viruses are now tentatively assigned to this family (24, 26, 31). Sequencing of three additional fusellovirus genomes, SSV2 (24), SSV RH (26), and SSV KM (26), has now been completed. Comparison of the SSV1 and SSV2 genomes shows 55% identity at the nucleotide level, with 27 of 34 putative ORFs common to both genomes (24). 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 shows that 18 of the 34 ORFs found in SSV1 are conserved across all four SSV genomes (26). In addition, the overall organization of ORFs within SSV2, SSV RH, and SSV KM are very similar to that found in SSV1, while nucleotide sequences corresponding to many of the putative promoters are also conserved (24, 26).

Despite extensive study of SSV1, functions have been suggested for only 6 of the 34 ORFs. Sequence analysis revealed similarity between D-335 and the integrase genes belonging to the type I tyrosine recombinase family (2), and D-335 has been expressed, purified, and characterized, confirming this activity (17). B-251 shows limited sequence similarity to the nucleotide binding protein DnaA (12). And three structural proteins (VP1, VP2, and VP3) have been isolated from purified virus particles and identified by N-terminal sequencing (21). Finally, structural studies of D-63 suggest that it may function as an adaptor protein (13). The functions of the remaining 28 ORFs remain unknown, and direct genetic and biochemical examination will be required to elucidate their functions. Towards this goal, we have undertaken a structural analysis of the SSV1 proteome. As in the case of D-63, structural analysis of SSV1 gene products is likely to make specific suggestions regarding their function. Here we report the biochemical characterization and crystal structure of a second SSV1 protein, F-93.

MATERIALS AND METHODS

Cloning. ORF F-93 was amplified by nested PCR from SSV1 genomic DNA prepared as described previously (23, 24, 26, 30). The PCR primers added a Shine-Delgarno sequence, a C-terminal His_6 tag, and *attB* sites to facilitate ligase-free cloning by using the Gateway cloning system (Invitrogen). The internal forward and reverse primers were CATCACCATCACCATCACAAAAGT ACACCATTCTT and CAAGAAAGCTGGGTCCTACCTCTTTTTGATTT GTT, respectively, while the external forward and reverse primers were GGGG ACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACC and GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGTGATGGTGATG

* Corresponding author. Mailing address: 108 Gaines Hall, Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717. Phone: (406) 994-5382. Fax: (406) 994-5407. E-mail: lawrence@chemistry.montana.edu.

GTGATG, respectively. The sequence of the entry clone was verified by sequencing. The His-tagged F-93 construct was then inserted into the *Escherichia coli* destination vector pDEST14 (Invitrogen), yielding the expression vector pDest14/F-93.

Expression and purification. Typically, pDEST14/F-93 was freshly transformed into BL21(DE3)-RIL *E. coli* (Novagen), and a single colony was used to inoculate a 5-ml overnight Luria broth culture, with subsequent serial expansion to 5 liters of fermentor batch culture (New Brunswick BIOFLO 2000). The medium for batch fermentation was as previously described (18). All batch cultures were supplemented with 50% oxygen. Cells were grown to an approximate optical density at 600 nm of 5 and then induced with 1 mM isopropyl-β-D-thiogalactopyranoside for 4 h and harvested by centrifugation at 6,000 × *g* for 20 min. Cell pellets were stored frozen at −80°C until needed. Cell pellets were then thawed and resuspended in lysis buffer (1.0 M NaCl, 0.02% NaN_3, 20 mM Tris-Cl, and 10 mM imidazole adjusted to pH 8.0 with freshly added phenylmethylsulfonyl fluoride at a final concentration of 0.1 mM). The concentration of lysis buffer volume to cell pellet mass was a minimum of 5 ml/g. The cell suspension was then passed once through a microfluidizer (Microfluidics Corporation, Newton, Mass.) and spun at 15,000 × *g* for 20 min. The resulting supernatant was applied to a column containing a 3-ml bed volume of Ni-nitrilotriacetic acid resin (QIAGEN), washed with lysis buffer, and eluted with approximately 10 ml of 400 mM NaCl, 20 mM Tris-Cl, 0.02% NaN_3, and 400 mM imidazole at pH 8.0. Eluted F-93 was then diluted to 50 ml with distilled H_2O. Solid NaCl was added incrementally to a final concentration of 4 M, and the sample was loaded onto a Pharmacia butyl hydrophobic interaction column (HIC) with a 50-ml super-loop. After loading, a NaCl gradient (0 to 5 min, 4 M; 5 to 20 min, 4 to 2 M; 20 to 50 min, 2 M; 50 to 70 min, 2 to 0 M; 70 to 90 min, 0 M) was used to elute F-93 at approximately 2 M NaCl. Solid NaCl was again added (final concentration, 4 M) to the pooled F-93 fractions, and the protein was reapplied to the HIC and eluted as described above. The doubly HIC-purified F-93 was then concentrated with Millipore spin concentrators (5,000-molecular-weight cutoff), loaded onto a Superdex S-75 gel filtration column, and exchanged into 200 mM NaCl–20 mM Tris-Cl (pH 8.0). F-93 eluted at ~12.5 ml, corresponding to a calculated molecular mass of 25,000 Da. Superdex S-75 fractions were pooled and concentrated to 5 to 10 mg/ml as measured by Bradford assay (3) by using standardized Bradford reagent (Bio-Rad) and bovine serum albumin as a protein standard. Protein yield was typically 0.5 to 1.0 mg/5 g of wet cell paste. Molecular weight and purity were assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (16.5% acrylamide; Tris-tricine buffer system) and matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS). MALDI-TOF MS was performed with a Bruker Biflex III instrument in the matrix alpha-cyano-4-hydroxycinnamic acid.

For expression and purification of selenomethionine-incorporated F-93, pDEST14/F-93 was transformed into *E. coli* strain B834(DE3) (Novagen), a methionine auxotroph. Initial cell growth was in LB medium with expansion to 1 liter. Cells were then pelleted by centrifugation, washed with distilled H_2O, and resuspended in 4 liters of minimal M9 medium with a glucose carbon source supplemented with biotin and thiamine at a concentration of 1 mg/liter. Methionine was added to a final concentration of 50 μg/ml. After growth to an optical density at 600 nm of 0.7, protein expression was induced, and the protein was purified as described for native protein (above). Selenium incorporation was estimated to be >90% as measured by MALDI-TOF MS.

Chemical cross-linking. After buffer exchange into 200 mM NaCl–20 mM HEPES (pH 8.0) with a Superdex S-75 gel filtration column, an aliquot of F-93 was diluted to give 10 ml at a final concentration of 500 nM. To this, 1 mg of the chemical cross-linking reagent BS3 [bis(sulfosuccinimidyl) suberate] (Pierce) was added, mixed by inversion, and incubated at room temperature for 30 min. The reaction was terminated by freezing at −80°C followed by lyophilization. The residue was then dissolved in 50 μl of sample buffer and analyzed by SDS-PAGE.

Crystallization and data collection. F-93 was crystallized at 4°C by hanging-drop vapor diffusion. Drops were assembled with 2 μl of F-93 mixed with 2 μl of well solution (3 to 3.5 M NaCl, 120 mM HEPES [pH 7.4 to 7.8]). Crystals typically appeared in 3 to 10 days. Selenomethionine-incorporated crystals were grown under identical conditions. Single crystals were isolated and dipped (10 to 30 s) in well solution supplemented with 25% glycerol. Crystals were then plunge-frozen in liquid nitrogen.

A three-wavelength anomalous diffraction data set was collected to a resolution of 3.0 Å at the selenium K edge (edge, peak, and remote wavelengths) at BioCARS beamline 14-BM-D at the Advanced Photon Source (Table 1). 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 $P4_332$, using the HKL software package (19).

Structure determination and refinement. SOLVE (25) was used for identifying positions of the selenium substructure and for calculation of the initial phases. Two selenomethionine sites were found corresponding to two monomers per asymmetric unit. The initial phases were improved by density modification including twofold noncrystallographic symmetry averaging followed by phase extension to 2.7 Å (5, 6). The resulting electron density map was of excellent quality and was used to build the initial model with the program O (11). Iterative rounds of refinement with Refmac 5 (5, 16) and manual rebuilding with O yielded a final model with a crystallographic R factor (R_{cryst}) of 19.4% (free R factor [R_{free}] of 25.9%). The refinement included the use of TLS parameters in which each monomer was defined as a rigid body. TLS refinement is often useful in the case of noncrystallographic symmetry, as different copies of a molecule in the asymmetric unit may have different overall displacements. TLS refinement resulted in a large drop in the mean B-factor, from 54.3 $Å^2$ to 29.5 $Å^2$, and meaningful decreases for both R_{cryst} and R_{free}. The final model has good stereochemistry, with no residues in the disallowed regions of the Ramachandran plot (14). Structural comparisons were performed by using the DALI (10) (http://www.ebi.ac.uk/dali) and VAST (http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html) servers. Figures were generated with PYMOL (http://www.pymol.org).

Coordinates. Atomic coordinates and structure factors have been deposited into the Protein Data Bank under accession code 1TBX.

RESULTS AND DISCUSSION

The F-93 construct used in this study codes for the 93 amino acids of native F-93, plus an additional C-terminal His_6 tag, for a total of 99 residues, with a calculated mass of 11,789 Da. Purified F-93 migrates at a mass of 25,000 Da (results not shown) on a calibrated Superdex S-75 size exclusion column, suggesting that the protein is actually present as a homodimer in solution. Chemical cross-linking experiments analyzed by SDS-PAGE demonstrate conversion of monomer to cross-linked dimer (Fig. 1), lending further support to the homodimer as the relevant biological assembly. F-93 crystallizes in space group $P4_332$, and the crystals diffract to a resolution of 2.7 Å. Two copies of the F-93 polypeptide are found in the asymmetric unit, corresponding to the dimer that is seen in solution. Details of the data collection, phasing, and refinement are presented in Table 1.

The structure of the F-93 monomer is shown in Fig. 2B. It reveals a common fold that places F-93 into the helix-turn-helix superfamily of DNA binding proteins. A search for nearest structural neighbors by using DALI or VAST indicates that F-93 is most similar to the SlyA (29) and MarR (1) subfamilies of winged-helix DNA binding proteins (1, 8, 29). F-93 clearly shares a common structural core with both of these proteins. For F-93 and *Enterococcus faecalis* SlyA (1LJ9) (29), VAST identifies 69 structurally equivalent C_α atoms that can be superimposed with a root mean square deviation of 1.5 Å. The match between F-93 and *E. coli* MarR (1JGS) gives 66 equivalent C_α positions with a root mean square deviation of 1.5 Å. This common fold includes the first three helices of F-93, followed by the two antiparallel β-strands. 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 that follow 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 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

TABLE 1. Data collection and refinement statistics

Type of data	Space group $P4_332$		
Crystallographic			
Unit cell parameters			
a = b = c (Å)	120.49		
α = β = γ (°)	90.0		
Data collection (SeMet MAD)	Inflection	Peak	Remote
Wavelength	0.9790	0.9787	0.9562
Resolution range (Å)[a]	20–3.00 (3.11–3.00)	20–3.00 (3.11–3.00)	20–2.70 (2.81–2.70)
No. of observations	84,316	84,241	139,831
No. of unique reflections observed/no. of possible unique reflections	11,008/11,290	11,002/11,290	15,117/15,472
Redundancy	7.65	7.65	9.24
Data coverage (%)	97.5 (99.4)	97.3 (99.4)	97.7 (98.9)
Average I/sigma I	33.2 (5.8)	31.9 (5.74)	41.0 (6.9)
R_{sym} (%)[b]	5.6 (25.4)	5.7 (28.3)	5.0 (16.7)
Refinement			
Resolution range (Å)[a]			14–2.7
R_{cryst} (%)/R_{free} (%)[c]			0.192/0.259
Coordinate error[d]			0.263
Real space CC[e]			0.920
RMSD from ideality			
Bond lengths (Å)			0.019
Bond angles (°)			1.924
Ramachandran plot (% amino acids in region)[f]			
Most favored			91.6
Additionally allowed			7.2
Generously allowed			1.2
Average *B* value (A^2)			29.46

[a] Numbers in parentheses refer to the highest resolution shell.
[b] $R_{sym} = 100*\Sigma_h\Sigma_i|I_i(h) - <I(h)>|/\Sigma_h I(h)$ where $I_i(h)$ is the *i*th measurement of the *h* reflection and $<I(h)>$ is the average value of the reflection intensity.
[c] $R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$ where F_o and F_c are the structure factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} with 5% of test set structure factors.
[d] Based on maximum likelihood.
[e] Correlation coefficient (CC) is between the model and $2mF_o$-DF_c density map.
[f] Calculated by using PROCHECK (14).

formation of the dimer interface. Thus, it is not surprising to find a C-terminal extension for F-93 that gives rise to a fourth helix. This helix, along with the N terminus and the first half of helix 1, form much of the F-93 dimer interface (Fig. 3A). The dimer interface is substantial and primarily hydrophobic. Formation of the dimer buries 1,998 $Å^2$, or 32.7%, of the total monomer surface area. Substantial dimer interfaces are also seen for members of the SlyA and MarR subfamilies.

How might F-93 bind DNA? A superposition of F-93 upon a SlyA homologue complexed to DNA would represent a good starting point for addressing this question. Unfortunately, such a complex is not present in the Protein Data Bank. However, the diphtheria toxin repressor (DtxR) is also a close structural neighbor, and the structure of the DtxR homodimer-DNA complex has been determined (Protein Data Bank accession number 1F5T) (4). DtxR shares the core winged-helix structure seen in F-93, with the addition of a C-terminal extension that gives rise to the dimer interface (Fig. 3C). 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. A similar interaction is easily imagined for the F-93 dimer shown in Fig. 3A, though the DNA would need a slight rotation about its pseudo-twofold axis (vertical axis) in order to optimize contacts with the recognition helix and the wing. In addition, there are four positive charges clustered about the N terminus of the dimer axis. Two charges come from the N termini themselves and two more from Lys[2]. This cluster of positive charges is in position to interact with the ribose phosphate backbone of the DNA. 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 bind-

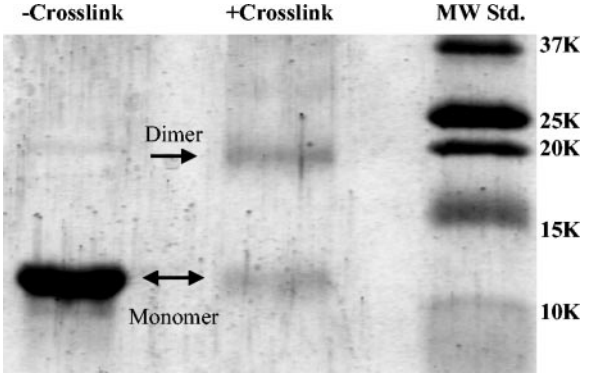

FIG. 1. Chemical cross-linking of F-93. The chemical cross-linking reagent BS3 was used to probe the oligomeric state of F-93. Incubation of F-93 with and without BS3 was followed by SDS-PAGE to determine the molecular weights of the cross-linked species as described in Materials and Methods. The left lane shows monomeric F-93 in the absence of the cross-linker. The middle lane shows residual monomer and the formation of cross-linked dimer due to the presence of BS3. Molecular weight standards (MW Std) are shown in the right lane.

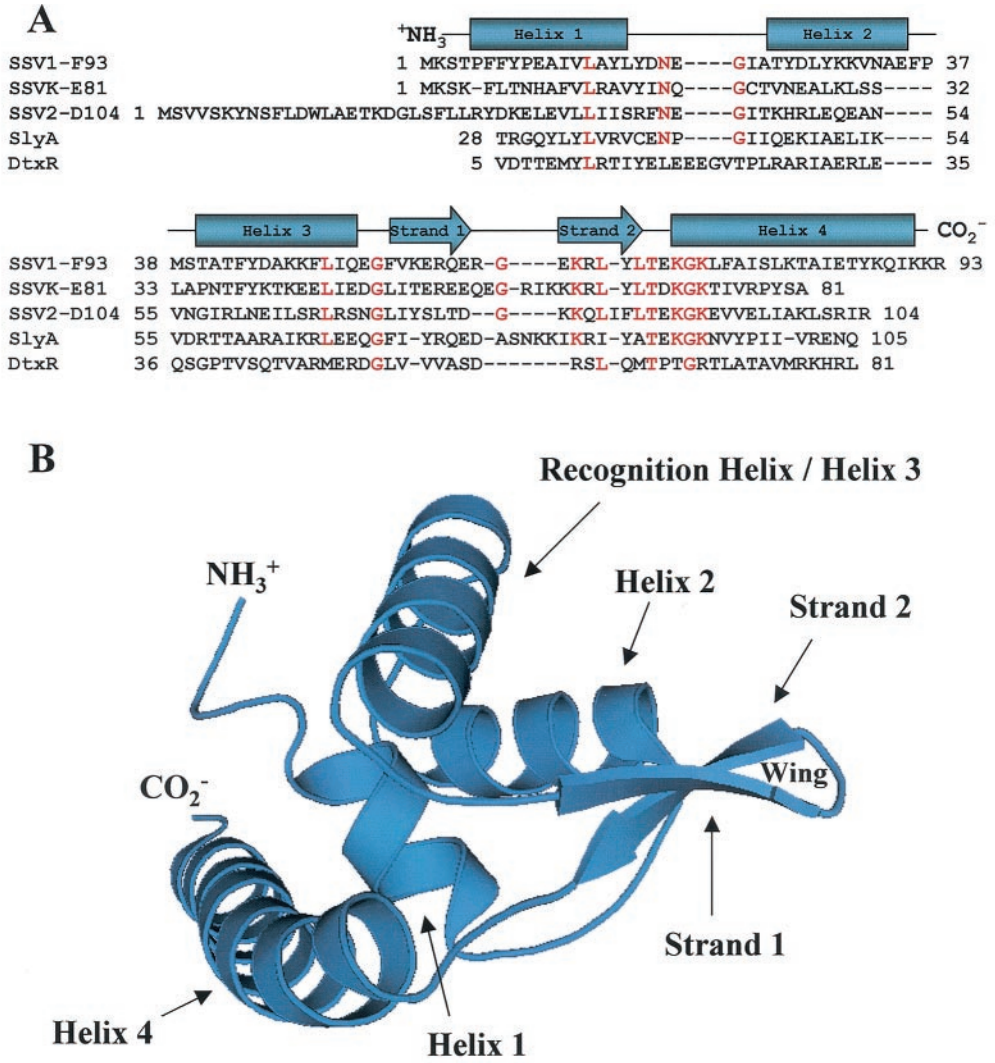

FIG. 2. (A) Multiple sequence alignment for structural homologues of SSV1 F-93. The SSV1 F-93 sequence is aligned with its fuselloviral homologues from SSV2 (D104/D106) and SSV K (E-81). Also included are structure based alignments between SSV1 F-93, SlyA, and DtxR. Secondary structural elements identified in the SSV1-F93 structure are mapped above the sequence. Strictly conserved residues in the fusellovirus 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. This feature is also found in SlyA. The 28-residue extension at the N terminus of SSV2 D-104 is noteworthy. The 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. (B) Structure of the F-93 monomer. A ribbon diagram depicts the secondary structural 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.

ing surface. This is particularly true for the wings, thus supporting their putative role in binding target DNA. Similar to DtxR and MarR, this binding model suggests that F-93 will recognize a DNA palindrome (or pseudo-palindrome). In this case, we have not been able to identify a palindrome with appropriate spacing in the SSV1 genome, supporting the pseudo-palindrome as the in vivo target sequence. There are, however, alternative explanations for the lack of a suitable palindrome in the SSV1 genome. For example, F-93 could be active in vivo in a form other than the homodimer, or it might target the DNA of its *Sulfolobus* host.

The multiple sequence alignment depicted in Fig. 2 shows residues that are strictly conserved among the F-93 fusellovirus homologues. Most of these residues appear to be structurally important and are scattered throughout the structure. For example, the conserved leucine side chains are found buried in the hydrophobic core of the molecule, while three of the four conserved glycines are found in connecting loops at the surface of the protein. In addition to these structurally important residues, a cluster of conserved residues is found as the polypeptide leaves strand 2 and enters helix 4. When these are mapped to the protein structure, along with conserved Gly^{54}, they are found to form a patch of conserved surface area at the N terminus of helix 4 (Fig. 2B and 3A), a surface that is not

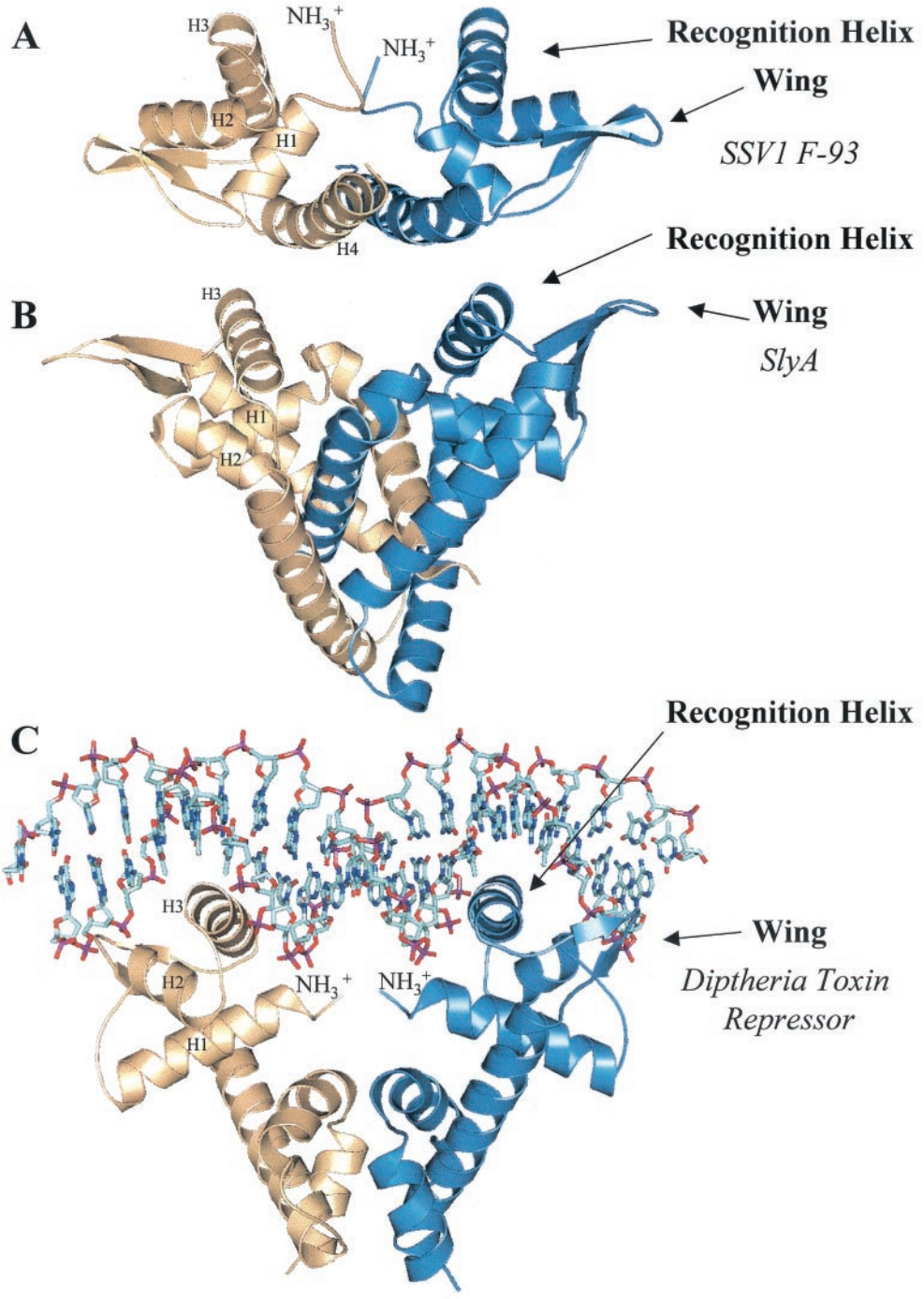

FIG. 3. (A) Structure of the F-93 dimer. The F-93 homodimer is depicted with chain A in blue and chain B in amber. The orientation of the A chain is identical to that shown in Fig. 2B. Successive helices in chain B are labeled H1 through H4. Helices H1 through H3 are common to the winged-helix fold. The dimer axis runs vertically through the dimer interface, 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 three helices common to the winged-helix motif are marked for the amber monomer (H1 through H3), and the wings and recognition helix (H3) are indicated for the blue monomer. These elements are found in orientations similar to those seen in F-93 (Fig. 2B and 3A). Obviously, the winged-helix motif found in SlyA is accompanied by a more extensive and elaborate dimer interface. The dimer axis runs vertically, coincident with that of F-93 in Fig. 3A. The relative position of the winged-helix motif is determined by the dimer interface. Thus, in comparing the F-93 and SlyA homodimers, 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. As also depicted in Fig. 3B, the elements of the winged-helix motif are indicated in both the amber and blue subunits. 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. DNA recognition is mediated by the N terminus, the recognition helix, and the wing of DtxR. Similar interactions can be imagined for the interaction of F-93 or SlyA with DNA. In each case, however, the fit could be improved by a small rotation of the DNA about its pseudo-twofold axis. Figures 3A through 3C 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. The relative pitch of the recognition helix and the spacing between these helices can be adjusted to recognize (pseudo-)palindromic sequences of various lengths.

predicted to interact with DNA. Thus, this surface might represent a potential binding site for partners in transcriptional regulation. This would most likely be a host protein, as the motif is conserved not only in the SSV1 F-93 homologues but also in the *E. faecalis* SlyA sequence (Fig. 2A) and other select members of the MarR and SlyA subfamilies of winged-helix DNA binding proteins.

A comparison of the structures of F-93, SlyA, and DtxR (Fig. 3) clearly shows that the SlyA and DtxR dimers have much more elaborate 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 that 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 (1).

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, MexR, and DtxR), alternative roles played by a few members of the winged-helix family should be considered as an explanation for the austere dimer interface. For example, F-93 shares structural similarities to the replication terminator protein of *Bacillus subtilis* (27), suggesting that F-93 might participate in replication of the viral genome. This is particularly noteworthy since F-93 is encoded by viral transcript T5, which also encodes a DnaA-like protein that might play a role in initiating DNA replication. Another noteworthy similarity is to the activation domain of MotA, a transcription factor from phage T4 (7). However, the matches for MotA and the replication terminator protein are weaker than those seen for transcription factors of the SlyA and MarR families. Thus, while we cannot exclude functions such as termination of replication, 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, the results of this study suggest specific experiments to identify the putative target DNA. Possible approaches include electrophoretic mobility shift assays followed by DNA footprinting, 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. Finally, this study 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 Philippe Benas and Ryan Todorovich for helpful discussions and able 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 number W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

REFERENCES

1. **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:**710–714.
2. **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:**433–440.
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:**248–254.
4. **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:**10397–10407.
5. **Collaborative Computational Project, Number 4.** 1994. The *CCP4* suite: programs for protein crystallography. Acta Crystallogr. Sect. D **50:**760–763.
6. **Cowtan, K.** 1994. DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsl. Protein Crystallogr. **31:**34–38.
7. **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.
8. **Gajiwala, K. S., and S. K. Burley.** 2000. Winged helix proteins. Curr. Opin. Struct. Biol. **10:**110–116.
9. **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:**594–599.
10. **Holm, L., and C. Sander.** 1993. Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233:**123–138.
11. **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 Crystallogr. Sect. A **47:**110–119.
12. **Koonin, E. V.** 1992. Archaebacterial virus SSV1 encodes a putative DnaA-like protein. Nucleic Acids Res. **20:**1143.
13. **Kraft, P., D. Kümmel, A. Oeckinghaus, G. H. Gauss, B. Wiedenheft, M. J. Young, and C. M. Lawrence.** 2004. Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four helix bundle suggest an adaptor protein function. Virology **78:**7438–7442.
14. **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:**283–291.
15. **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:**2165–2168.
16. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. Sect. D **53:**240–255.
17. **Muskhelishvili, G., P. Palm, and W. Zillig.** 1993. SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. Mol. Gen. Genet. **237:**334–342.
18. **New Brunswick Scientific R&D Laboratory.**1996. Fundamentals of fermentation: techniques for benchtop fermentors. Part I–*E. coli*. New Brunswick Scientific, Edison, N.J.
19. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode, p. 307–326. *In* C. Carter and R. Sweet (ed.), Macromolecular crystallography, part A, vol. 276. Academic Press, Inc., New York, N.Y.
20. **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:**242–250.
21. **Reiter, W. D., P. Palm, A. Henschen, F. Lottspeich, W. Zillig, and B. Grampp.** 1987. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. **206:**144–153.
22. **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. Nucleic Acids Res. **16:**1–19.
23. **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. USA **89:**7645–7649.
24. **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.

25. **Terwilliger, T. C., and J. Berendzen.** 1999. Automated MAD and MIR structure solution. Acta Crystallogr. Sect. D **55:**849–861.
26. **Wiedenheft, B., K. Stedman, F. Roberto, D. Willits, A.-K. Gleske, L. Zoeller, J. Snyder, T. Douglas, and M. J. Young.** 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. **78:**1954–1961.
27. **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:**206–210.
28. **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. USA **87:**4576–4579.
29. **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:**20240–20244.
30. **Yeats, S., P. McWilliam, and W. Zillig.** 1982. A plasmid in the archaebacterium *Sulfolobus solfataricus*. EMBO J. **1:**1035–1038.
31. **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:**131–140.
32. **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:**225–236.
33. **Zillig, W., R. Schnabel, and K. O. Stetter.** 1985. Archaebacteria and the origin of the eukaryotic cytoplasm. Curr. Top. Microbiol. Immunol. **114:**1–18.