Structure of D-63 from Sulfolobus Spindle-Shaped Virus 1: Surface Properties of the Dimeric Four-Helix Bundle Suggest an Adaptor Protein Function

Paul Kraft,1,2 Daniel Kümmel,1,3 Andrea Oeckinghaus,1,3 George H. Gauss,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 59717, and Physiologisch-Chemisches Institut der Universität, D-72076 Tübingen, Germany3

Received 8 January 2004/Accepted 12 March 2004

Sulfolobus spindle-shaped virus 1 (SSV1) and its fusellovirus homologues can be found in many acidic (pH ≤ 4.0) hot springs (≥70°C) around the world. SSV1 contains a 15.5-kb double-stranded DNA genome that encodes 34 proteins with greater than 50 amino acids. A site-specific integrase and a DnaA-like protein have been previously identified by sequence homology, and three structural proteins have been isolated from purified virus and identified by N-terminal sequencing (VP1, VP2, and VP3). The functions of the remaining 29 proteins are currently unknown. To assign functions to these proteins, we have initiated biochemical and structural studies on the SSV1 proteome. Here we report the structure of SSV1 D-63. The structure reveals a helix-turn-helix motif that dimerizes to form an antiparallel four-helix bundle. Mapping residues conserved among three fusellovirus isolates onto the structure shows that one face of the rod-shaped molecule is highly conserved. This conserved surface spans the dimer axis and thus exhibits 2-fold symmetry. Two smaller conserved patches, also related by 2-fold symmetry, are found on the opposite face of the molecule. All of these conserved surfaces are devoid of clefts or pockets typically used to bind small molecules, suggesting that D-63 may function as an adaptor protein in macromolecular assembly.

* Corresponding author. Mailing address: Gaines Hall 108, Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717. Phone: (406) 994-5392. Fax: (406) 994-5407. E-mail: lawrence@chemistry.montana.edu.
undertaken structural analysis of the SSV1 proteome. Here we report the initial biochemical characterization and crystal structure of SSV1 D-63.

MATERIALS AND METHODS

Cloning. The D-63 ORF was amplified by nested PCR from SSV1 genomic DNA prepared as described previously (21, 23, 26, 28). The PCR primers added a Shine-Dalgarno sequence, a C-terminal His6 tag, and attB sites to facilitate ligation-free cloning with the Gateway cloning system (Invitrogen). The internal forward and reverse primers were TACAAGAAAGCTGGGTCCTAGTGATGGTGATGGTGATG3' and AAAGCAGGCTTCGAAGGAGATAGAACC3', respectively. The overnight culture was added to 1 liter of an identical medium except that methionine was added to a final concentration of 50 μM. The initial phases were improved with density modification, including fourfold NCs averaging over residues 8 to 28 and 40 to 61 (2, 5). The resulting electron density map was of excellent quality and was used to build the initial model with the program O (8). Iterative rounds of refinement with Refmac5 (2, 14) and manual rebuilding with O yielded a final model with an R factor of 22.5% (R_free = 27.0%). The model has good stereochemistry, with no residues in the disallowed regions of the Ramachandran plot (10). Structural comparisons were performed with the DALI server (http://www.ebi.ac.uk/dali) (7). Figures were generated with PYMOL (http://www.pymol.org). Atomic coordinates and structure factors have been deposited into the Protein Data Bank under accession code 1SKV.

RESULTS AND DISCUSSION

The SSV1 D-63 construct used in this work codes for the 63 amino acids of native D-63, plus an additional C-terminal His6 tag, for a total of 69 residues. MALDI-TOF indicates that the N-terminal methionine has been cleaved. Purified D-63 mutant was 200 mM NaCl–50 mM ammonium acetate, pH 5.0, the same pH used for crystal growth.

Crystalization and data collection. D-63 was crystallized by hanging-drop vapor diffusion. Drops were assembled with 2 μl of D-63 mixed with 2 μl of well solution: 0.2 M (NH4)2SO4, 0.2 M ammonium acetate, and 25% polyethylene glycol 4000, pH 4.7 to 5.0. Plates were incubated at 18°C, and crystals appeared in 4 to 6 weeks. Single crystals were placed in dialysis bags and moved at 2-h intervals through a series of well solutions supplemented with increasing glycerol concentration in steps of 5% to a final concentration of 25% glycerol. Crystals were then plunged-frozen in liquid nitrogen. A three-wavelength multianomalalous deposition data set to 2.9 Å was collected at the selenium K edge (edge, peak, and remote wavelengths) at BioCARS beamline 14-BM-D at the Advanced Photon Source (Tables 1 and 2). Data were integrated and reduced in space group P62 with the HKL software package (16).

Structure determination and refinement. SOLVE (24) was used to locate six expected selenomethionine sites corresponding to four monomers per asymmetric unit; the two additional selenomethionine residues were disordered. The initial phases were improved with density modification, including fourfold NCs averaging over residues 8 to 28 and 40 to 61 (2, 5). The resulting electron density map was of excellent quality and was used to build the initial model with the program O (8). Iterative rounds of refinement with Refmac5 (2, 14) and manual rebuilding with O yielded a final model with an R factor of 22.5% (R_free = 27.0%). The model has good stereochemistry, with no residues in the disallowed regions of the Ramachandran plot (10). Structural comparisons were performed with the DALI server (http://www.ebi.ac.uk/dali) (7). Figures were generated with PYMOL (http://www.pymol.org). Atomic coordinates and structure factors have been deposited into the Protein Data Bank under accession code 1SKV.
grates at a calculated molecular mass of 17,100 Da (results not shown) on a calibrated Superdex S-75 size exclusion column, suggesting that D-63 assembles to form a homodimer at pH 5.0, the same pH at which crystals are grown. D-63 crystallizes in space group P6_2, and the crystals diffract to 2.6-Å resolution. Details of crystal growth, data collection, phasing, and refinement are summarized in Tables 1 and 2 and detailed in Materials and Methods. Four copies of the D-63 polypeptide, designated chains A through D, are present in the asymmetric unit, providing the opportunity to observe multiple conformations of D-63. Each of the four identical chains folds into a helix-turn-helix motif with the two helices running antiparallel to each other. The connecting loop is composed of residues 35 through 39.

The amino- and carboxy-terminal ends of the structure show various amounts of well defined electron density, and within chain D, density at the C-terminal end of the first helix and the connecting loop is also poorly defined. Well-ordered density is seen for chain A, residues 2 through 65; chain B, residues 6 through 66; chain C, residues 7 through 64; and chain D, residues 2 through 29 and 39 through 66 (residue numbers greater than 63 correspond to the C-terminal His6 tag).

Structural comparison of the four identical chains shows that the conformations are generally similar, though there are significant conformational differences for the C terminus of the first helix and the residues comprising the loop (not shown). The disorder present through this region in chain D further highlights the conformational flexibility within this part of the structure.

The A and B monomers are related to each other by a 2-fold symmetry axis such that the chains pack against each other to form an antiparallel four-helix bundle (Fig. 1A). An identical relationship is seen for chains C and D. In each case, the 2-fold axes run perpendicular to the long axis of the four-helix bundle. Thus, as one looks down the dimer axis, the four-helix bundle presents 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). There is extensive contact between the monomers at the dimer interface. The calculated surface area for the D-63 monomer is 5,346 Å^2 (2, 11). Formation of the AB dimer buries 1,300 Å^2, or 24%, of the solvent-accessible surface area per monomer. The dimer interface is largely hydrophobic and is composed of residues Phe^{11}, Leu^{14}, Val^{18}, Leu^{21}, Ile^{25}, Ile^{28}, Leu^{32}, Ala^{45}, Val^{49}, Ile^{52}, Leu^{56}, and Leu^{59} (Fig. 1B). The presence of the hydrophobic core is discernible in the primary sequence as a clear heptad repeat (Fig. 2A). The dimer is also stabilized by intermolecular salt bridges and hydrogen bonds. These include His^{A26} to Asp^{B15}, Arg^{A32} to Glu^{B37}, Tyr^{A46} to Glu^{B55}, Glu^{A55} to Glu^{B50}, Ser^{A63} to Arg^{B32}, and their five symmetry-related equivalents between chain B and chain A. Thus the 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 columns.

Meaningful sequence similarity between D-63 and nonfusellovirus proteins has not been found. However, D-63 can be aligned with its fusellovirus homologues from SSV2 (D-57) and SSV RH (F-61), though a D-63 homologue is not found in SSV KM. Alignment of the D-63 homologues with ClustalW (25) shows 51% sequence identity between the SSV1 and SSV2 proteins and 37% sequence identity between the SSV1 and SSV RH proteins (Fig. 2A). The three-way alignment shows that ~32% of the residues, 20 of 63, are strictly conserved. We also note additional similarity 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.

**Biological implications.** It is well documented that tertiary (three-dimensional structural) similarities persist far longer on the evolutionary time scale than either primary (amino acid) or genomic sequence (DNA) similarities (4, 13, 20). Thus, prior to structure determination, we hypothesized that the lack of significant sequence similarity between D-63 and other proteins in the public database did not indicate a unique fold for this protein. Rather, we expected that the structure of D-63 would reveal a familiar fold and hoped that this fold would suggest a function for D-63. However, the D-63 helix-turn-helix motif is present in many larger proteins as a supersecondary structural element, while the four-helix bundle of dimeric D-63 is an extremely common fold that has been adapted to a multitude of functions. Thus, a search of the Protein Data Bank for nearest structural neighbors using the DALI structural biology server (7) returns a long list of proteins with statistically significant structural similarities and a
FIG. 2. (A) Sequence alignments of D-63 fusellovirus homologues. Strictly conserved residues are red, and nonconserved residues are black. Secondary structure assignments, i.e., the extent of the alpha helices, are indicated above the sequence. The hydrophobic residues are now numerous examples of structure suggesting function (13), it is clear that, for D-63, the commonality of the helix-turn-helix motif precludes any reliable assignment.

In sharp contrast to the search for function by structural homology, analysis of the surface properties of D-63 is 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 to this rule occur when surface residues are required for the activity of the protein (13). In such cases, conservation of these residues suggests the location of ligand binding sites, or, for an enzyme, the active site. Sequence alignments of D-63 with its homologues from SSV2 (D-57) and SSV RH (F61) are shown in Fig. 2A. When the strictly conserved residues are mapped to the surface of D-63, a large patch of conserved surface area is found on the N-terminal face of the four-helix bundle (Fig. 2B). The strictly conserved residues comprising this surface 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 2-fold symmetric. It is also devoid of obvious clefts or pockets that might accommodate a small molecule, suggesting that the conserved surface features are responsible for binding a macromolecule. Smaller patches of strictly conserved surface are seen on the opposite face (C-terminal face; not shown), with significant sequence similarity for the surrounding surface-exposed residues. These areas might also represent conserved ligand binding sites that, like the N-terminal face, are also 2-fold symmetric.

It is possible that the D-63 dimer might dissociate to monomers and that the monomer might bind to the target macromolecules. However, if D-63 is active as the dimeric four-helix bundle, the conserved surface features would serve to recognize macromolecules possessing 2-fold symmetry or would function to bind two copies of a monomeric entity, effectively serving to dimerize the molecule (D-63-induced dimerization). In either event, the conserved D-63 surface features strongly suggest that D-63 functions in macromolecular assembly and that the binding of an unknown ligand(s) by D-63 serves to regulate or modify the behavior of the ligand, thus assisting in long list of functions that might be considered for D-63. However, as each protein on the list is examined in detail, the match is found to be problematic in one way or another. As examples, consider the two best matches to the helix-turn-helix motif. The greatest similarity (highest Z score) is to a similar supersecondary structural element present at the dimer interface of valyl-tRNA synthetase, where, in addition to dimerization, the element also plays a critical role in the binding of tRNA to the synthetase. But a role for D-63 in RNA binding is easily discounted; its calculated pI is 5.07. The second best match is to a helix-turn-helix present as the "finger" of GreA (22), a bacterial transcription factor that suppresses elongation arrest. However, crenarchaeal RNA polymerases are more eukaryotic-like than prokaryotic, and one might expect the associated transcription factors to be eukaryotic in nature as well. In these cases, and for all the others on the list of similar structures, it is difficult to rationalize the suggested function based on the constraints of the D-63 structure, knowledge of the SSV1 life cycle, and crenarchaeal biochemistry. Thus, even though there are now numerous examples of structure suggesting function (13), it is clear that, for D-63, the commonality of the helix-turn-helix motif precludes any reliable assignment.
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 His6 tag.

From the structural analysis presented here, a preliminary understanding of the function of D-63 has emerged. These results will guide future biochemical and genetic studies, eventually leading to a more detailed understanding of D-63. As this occurs, the structure will only grow in relevance and will provide additional insight into the means through which D-63 exerts its action. Thus it is clear that the structural analysis of D-63 has been a worthwhile undertaking. Further, it demonstrates the utility of structure determination for proteins of unknown function.

ACKNOWLEDGMENTS

We thank the BioCARS staff for their excellent assistance during data collection at sector 14 of the Advanced Photon Source. We are grateful to Philippe Benas, Ryan Todorovich, Cinnamon Spear, and John Gilmore for their 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 no. 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