Available online at www.sciencedirect.com
ScienceDirect
Virology 368 (2007) 249–261
VIROLOGY
www.elsevier.com/locate/yviro

A winged-helix protein from *sulfolobus* turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus

Eric T. Larson [a,b], Brian Eilers [a,b], Smita Menon [a,b], Dirk Reiter [a,b,c], Alice Ortmann [a,d], Mark J. Young [a,d,e], C. Martin Lawrence [a,b,*]

[a] *Thermal Biology Institute, Montana State University, Bozeman, MT 59717, USA*
[b] *Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA*
[c] *Physiologisch-Chemisches Institut der Universität Tübingen, Hoppe-Seyler-Strasse 4, D-72076 Tübingen, Germany*
[d] *Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717, USA*
[e] *Department of Microbiology, Montana State University, Bozeman, MT 59717, USA*

Received 20 December 2006; returned to author for revision 1 June 2007; accepted 28 June 2007
Available online 31 July 2007

Abstract

Sulfolobus turreted icosahedral virus (STIV) was the first non-tailed icosahedral virus to be isolated from an archaeal host. Like other archaeal viruses, its 37 open reading frames generally lack sequence similarity to genes with known function. The roles of the gene products in this and other archaeal viruses are thus largely unknown. However, a protein's three-dimensional structure may provide functional and evolutionary insight in cases of minimal sequence similarity. In this vein, the structure of STIV F93 reveals a homodimer with strong similarity to the winged-helix family of DNA-binding proteins. Importantly, an interchain disulfide bond is found at the dimer interface, prompting analysis of the cysteine distribution in the putative intracellular proteins of the viral proteome. The analysis suggests that intracellular disulfide bonds are common in cellular STIV proteins, where they enhance the thermostability of the viral proteome.

Keywords: Sulfolobus turreted icosahedral virus; STIV; F93; B116; Disulfide; Intracellular disulfide; Intermolecular disulfide; Thermostable; Thermostability; Archaea

Introduction

While the number of viruses known to infect hyperthermophilic organisms is growing rapidly, the study of these viruses is still in its infancy when compared to similar studies on eukaryotic viruses and bacteriophage. Characterization of hyperthermophilic viruses inhabiting the hot springs of Yellowstone National Park and other thermal features around the world reveals fascinating diversity, apparent in both viral morphology and genomic sequence. This diversity has spawned the creation of at least five new viral families reflecting the unique characteristics of these viruses (Prangishvili and Garrett, 2005). In general, little is known regarding the life cycles of these viruses, their virus–host relationships, genetics, or biochemistry, and further study of these unique viruses is clearly warranted. Importantly, while such studies will lead to a greater understanding of the viruses themselves, they are also expected to provide genetic, biochemical, and evolutionary insight into their crenarchaeal hosts and the requirements for life in the harsh environments in which these organisms thrive.

Sulfolobus turreted icosahedral virus (STIV) infects *Sulfolobus* species resident in the acidic hot springs (pH 2.9–3.9 and 72–92 °C) of Yellowstone National Park and was the first hyperthermophilic virus described with icosahedral capsid architecture (Rice et al., 2004). Structural analysis of the STIV particle and its major capsid protein suggest that it belongs to a lineage of double-stranded DNA viruses that predates the fundamental evolutionary events giving rise to the Archaea,

* Corresponding author. Thermal Biology Institute, Montana State University, Bozeman, MT 59717, USA. Fax: +1 406 994 5407.
E-mail address: lawrence@chemistry.montana.edu (C.M. Lawrence).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.06.040

Bacteria, and Eukarya (Bamford et al., 2005; Benson et al., 2004; Khayat et al., 2005; Larson et al., 2006; Maaty et al., 2006; Rice et al., 2004). The STIV genome encodes 37 putative open reading frames (ORFs). Similar to other crenarchaeal viruses, functional predictions for these hypothetical proteins are hindered by their lack of similarity to sequences of known function. However, sequence analysis by our laboratory and others suggests the presence of several classes of transcriptional regulators [M. Dlakic, personal communication (Prangishvili et al., 2006)] and an ATPase (Iyer et al., 2004; Maaty et al., 2006). In addition, analysis of the purified virus has identified the major capsid protein (Khayat et al., 2005; Rice et al., 2004) and eight other proteins, including two from the host (Maaty et al., 2006). Finally, structural studies strongly suggest that the A197 gene product is a glycosyltransferase (Larson et al., 2006) and the B116 gene product is a unique nucleic acid binding protein (Larson et al., 2007). The functions of the remaining gene products, however, remain a mystery.

It is unlikely that most of the STIV proteome consists of unique protein folds serving novel functions. Rather, the limited sequence similarity is a likely function of the unique environment in which these viruses replicate and the evolutionary distance between them and their bacterial and eukaryotic homologues. Because similarity between related protein structures persists longer on an evolutionary time scale than similarities in their corresponding amino acid sequences, protein structure may uncover functional and evolutionary relationships not apparent from the primary sequence (Aravind et al., 2005; Benson et al., 1999; Benson et al., 2004; Buehner et al., 1973; Khayat et al., 2005; Kraft et al., 2004a,b; Larson et al., 2006; Moult and Melamud, 2000; Rossmann et al., 1981; Unligil and Rini, 2000; Unligil et al., 2000). To this end, we have initiated crystallographic studies of the STIV proteome. Like similar studies on *Sulfolobus* Spindle-shaped Virus 1 (SSV1) (Kraft et al., 2004a,b) and our recent work on STIV (Larson et al., 2007), structural analysis of the STIV gene products is expected to provide insights into the possible functions of these proteins and their roles in the STIV life cycle.

We report on the structure determination of F93, a 93-residue protein in reading frame F of the viral genome. Structural and biochemical characterization of F93 reveals a homodimeric winged-helix protein that is likely to function as a transcriptional regulator. Notably, we find an interchain disulfide bond that spans the dimer interface, and biochemical studies show that it enhances protein thermostability. The observation of a disulfide bond in a presumptive intracellular protein prompted further analysis of the cysteine distribution in the STIV proteome. The analysis reveals a pattern consistent with the frequent use of disulfide bonds in small intracellular proteins encoded by STIV, strongly suggesting that intracellular disulfide bonds make a significant contribution to the thermostability of the viral proteome. While stable disulfide bonds are rarely found in the cytoplasm of most organisms (Kadokura et al., 2003), our conclusion is consistent with the observations of Yeates et al. on the occurrence of stabilizing disulfide bonds in cytoplasmic proteins of certain Archaea (Beeby et al., 2005; Mallick et al., 2002; O'Connor and Yeates, 2004). Our work strongly supports these conclusions and extends the observed occurrence of stabilizing disulfide bonds in cytoplasmic proteins to the genome of a hyperthermophilic virus.

Results

The F93 construct used in this study codes for the 93 amino acids of the native protein plus an additional C-terminal His-tag, resulting in 99 residues with a calculated mass of 11,970 Da. Purified F93 elutes from a Superdex™ 75 size exclusion column as a single peak with an apparent molecular weight of approximately 25 kDa, suggesting that it is present as a homodimer in solution. The protein crystallizes in space group $C222_1$ with two copies of the F93 polypeptide per asymmetric unit. The structure was initially solved at 2.4 Å resolution by single-wavelength anomalous diffraction (SAD) using selenomethionyl-incorporated protein, with subsequent refinement at 2.2 Å resolution using native data. Details on data collection and model refinement are presented in Tables 1 and 2. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (www.pdb.org) under accession code 2CO5.

Structure of F93

The structure of F93 (Fig. 1) reveals a winged-helix (or winged-helix-turn-helix, wHTH) fold, a subclass of the helix-turn-helix (HTH) protein superfamily (Aravind et al., 2005). Members of the wHTH family share a bundle of three α-helices, which is comprised of helices α1, α2, and α3 in F93. Helices α2, α3, and the intervening turn comprise the HTH motif, in which α3 serves as the recognition helix. Following α3, two β-strands connect through a reverse turn to form a small flanking antiparallel β-sheet that constitutes the "wing" of this structural motif. These features are important to the DNA-binding function of these proteins; the recognition helix generally makes base-specific contacts within the DNA major groove,

Table 1
Data collection [a]

Parameter	Result [b] for data set	
	Se peak	Native
Wavelength (Å)	0.97908	0.98789
Space group	$C222_1$	
Cell constants (a, b, c; Å) $\alpha=\beta=\gamma=90°$	41.99, 102.55, 92.76	42.05, 102.59, 92.40
Resolution range (Å)	50–2.4 (2.49–2.40)	50–2.2 (2.28–2.20)
Unique reflections	8000 (791)	10466 (1005)
Average redundancy	6.3 (5.8)	7.1 (7.3)
I/σ	28.8 (5.5)	33.0 (4.2)
Completeness (%)	97.8 (99.4)	99.8 (100)
R_{sym} [c]	0.068 (0.22)	0.058 (0.34)

[a] Data were integrated, scaled, and reduced using the HKL-2000 software package (Otwinowski and Minor, 1997).
[b] Numbers in parenthesis refer to the highest resolution shell.
[c] $R_{sym}=100*\sum_h\sum_i|I_i(h)-<I(h)>|/\sum_h I(h)$ where $I_i(h)$ is the ith measurement of reflection h and $<I(h)>$ is the average value of the reflection intensity.