

# Characterization of the Archaeal Thermophile *Sulfolobus* Turreted Icosahedral Virus Validates an Evolutionary Link among Double-Stranded DNA Viruses from All Domains of Life

Walid S. A. Maaty,<sup>1</sup> Alice C. Ortmann,<sup>2</sup> Mensur Dlakić,<sup>3</sup> Katie Schulstad,<sup>1</sup> Jonathan K. Hilmer,<sup>1</sup>  
Lars Liepold,<sup>1</sup> Blake Weidenheft,<sup>2</sup> Reza Khayat,<sup>4</sup> Trevor Douglas,<sup>1</sup>  
Mark J. Young,<sup>2</sup> and Brian Bothner<sup>1\*</sup>

Department of Chemistry and Biochemistry, Montana State University, Bozeman,<sup>1</sup> Departments of Microbiology and Plant Sciences, Montana State University, Bozeman,<sup>2</sup> and Department of Microbiology, Montana State University, Bozeman,<sup>3</sup> Montana 59717, and Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037<sup>4</sup>

Received 13 March 2006/Accepted 17 May 2006

**Icosahedral nontailed double-stranded DNA (dsDNA) viruses are present in all three domains of life, leading to speculation about a common viral ancestor that predates the divergence of *Eukarya*, *Bacteria*, and *Archaea*. This suggestion is supported by the shared general architecture of this group of viruses and the common fold of their major capsid protein. However, limited information on the diversity and replication of archaeal viruses, in general, has hampered further analysis. *Sulfolobus* turreted icosahedral virus (STIV), isolated from a hot spring in Yellowstone National Park, was the first icosahedral virus with an archaeal host to be described. Here we present a detailed characterization of the components forming this unusual virus. Using a proteomics-based approach, we identified nine viral and two host proteins from purified STIV particles. Interestingly, one of the viral proteins originates from a reading frame lacking a consensus start site. The major capsid protein (B345) was found to be glycosylated, implying a strong similarity to proteins from other dsDNA viruses. Sequence analysis and structural predication of virion-associated viral proteins suggest that they may have roles in DNA packaging, penton formation, and protein-protein interaction. The presence of an internal lipid layer containing acidic tetraether lipids has also been confirmed. The previously presented structural models in conjunction with the protein, lipid, and carbohydrate information reported here reveal that STIV is strikingly similar to viruses associated with the *Bacteria* and *Eukarya* domains of life, further strengthening the hypothesis for a common ancestor of this group of dsDNA viruses from all domains of life.**

In comparison to viruses with eukaryotic and bacterial hosts, little is known about the viruses that infect *Archaea*. This is due, in part, to the relatively recent delineation of the archaeal domain of life but, more significantly, to the challenges of isolating and culturing the host organisms (42). The extreme environments favored by many archaeal species and limited knowledge about their biochemistry and biology exacerbate this problem. Often, it is through the study of host-virus interactions that insights to the biology of the host are elucidated. The recent discovery of *Sulfolobus* turreted icosahedral virus (STIV) presents an opportunity to expand our knowledge of virology, study host biology, and investigate the evolutionary relationship of viruses from all three domains of life. Studies on the structure of STIV have revealed similarities with prokaryotic and eukaryotic viruses that suggest a common ancestry for icosahedral double-stranded DNA (dsDNA) viruses (30, 38).

STIV was isolated from *Sulfolobus* enrichment cultures that were established from a high-temperature acidic hot spring (~80°C, pH ~3) in Yellowstone National Park (38). The virus was subsequently shown to infect virus-free isolates of *Sulfolobus solfataricus* strain P2, for which the complete genome has been

sequenced. The electron cryomicroscopy (cryo-EM) reconstruction revealed a capsid with pseudo-T=31 icosahedral symmetry that is composed primarily of a 37-kDa major capsid protein, plus at least three additional capsid proteins at the fivefold vertices (38). Two features in particular stand out: the turrets, which are proposed to function in host recognition and DNA translocation, and two electron-dense layers sandwiched between the protein capsid and the packaged genome that may be composed of lipids (30) (Fig. 1). While not common, internal lipid layers are present in a number of dsDNA viruses. *Paramecium bursaria Chlorella* virus 1 (PBCV-1), which replicates in unicellular *Chlorella*-like green algae (41, 46), and the bacteriophage PRD1 (13) have a lipid layer beneath the glycoprotein capsid shell. The spherical halophilic euryarchaeon *Haloarcula hispanica* virus (SH1) also has an internal lipid layer that is selectively acquired from the host (4).

STIV's dsDNA genome has 17,663 bp and 36 predicted open reading frames (ORFs) (38). To date, only three proteins have been characterized, and the remaining 33 ORFs represent hypothetical proteins. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) coupled with protein N-terminal sequencing and peptide mass mapping identified the major capsid protein, B345, from preparations of purified virus (38). The jelly roll fold of B345 is nearly identical to the major capsid protein of adenovirus, PBCV-1, and PRD1 (7, 30, 34). Structural models based on X-ray crystallography have been determined for two other STIV proteins, F93 and A197. F93

\* Corresponding author. Mailing address: 308 Gaines Hall, Chemistry and Biochemistry Department, Bozeman, MT 59715. Phone: (406) 994-5270. Fax: (406) 994-5407. E-mail: bbothner@chemistry.montana.edu.

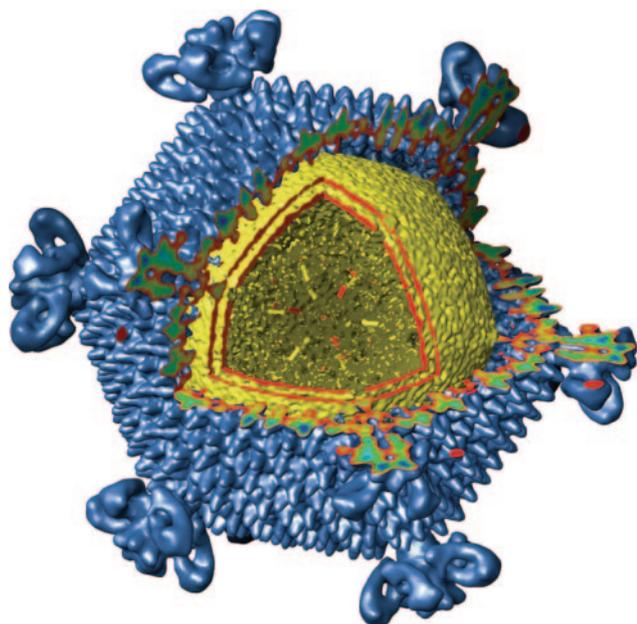


FIG. 1. Model of an STIV particle. Shown is a cutaway view of the T=31 icosahedral capsid of STIV based on cryo-EM reconstruction (30). Extending from each of the fivefold vertices are turret-like projections. The protein shell is blue and has been removed from one quarter of the particle to reveal the inner lipid layers (yellow). The capsid diameter is ~70 nm, and the turrets extend 13 nm above the surface.

is a winged-helix DNA binding protein (M. Lawrence, personal communication), and A197 is a glycosyltransferase-like protein (31).

Characterization of a virus necessarily involves the identification and analysis of the components that assemble to form the particle. Proteomics-based approaches are a powerful tool for dissecting macromolecular complexes such as virus particles (44). In this report STIV particle composition was characterized using mass spectrometry. Nine viral and two host proteins in purified preparations of STIV were identified by mass spectrometry. One of the viral proteins originates from a noncanonical reading frame, confirming that standard translation rules are not sufficient to generate the entire proteome of archaeal organisms. Even though primary sequence-based searches failed to find homologous proteins, fold recognition-based searches suggested potential roles for many of the virion-associated proteins. Structural prediction indicates that four of the proteins are likely to be part of the turrets, including an ATPase involved in DNA packaging. We confirm the presence of an internal lipid layer and show that the virus selectively incorporates a subset of total host lipids. We also demonstrate that the major capsid protein (B345) is glycosylated.

#### MATERIALS AND METHODS

**Virus purification.** A single-colony isolate of a *Sulfolobus* sp. producing an icosahedral virus-like particle was established by previously described methods (37). This isolate was designated YNPRC179, while the particle was described as STIV (38). STIV was purified from infected cultures using a modification from the original method. Viruses were purified from a culture of YNPRC179 and inoculated into virus-free *Sulfolobus solfataricus* P2 cells. Production of the virus

was carried out by growing the *S. solfataricus* P2 cells in 1 liter of medium 182 ([www.dsmz.de/microorganisms/html/media/medium000182.html](http://www.dsmz.de/microorganisms/html/media/medium000182.html)) at pH 2.5 until the cultures reached early log phase. Previously purified STIV was used to inoculate the cultures, and the inoculated cultures were monitored for virus production using an immuno-dot blot assay for the major capsid protein. Cultures were harvested between 72 to 96 h postinfection. Cells were removed by centrifugation (~6,000 × g for 10 min), and the supernatant was filtered through a 0.22- $\mu$ m Seritop filter (Millipore). The filtrate was concentrated over Amicon membranes (Millipore) with a molecular mass cutoff of 100 kDa until the retained volume was ~15 ml. The filter retentate was subsequently loaded onto Cs<sub>2</sub>SO<sub>4</sub> (39%, wt/vol) and centrifuged at 247,600 × g for 16 h. Virus bands were removed and dialyzed against 50 mM citrate buffer (pH 3.2). Purity was gauged by banding characteristics on Cs<sub>2</sub>SO<sub>4</sub> gradients, negative-stain transmission electron microscopy, and UV-visual spectroscopy.

**SDS-PAGE analysis and in-gel digestion for protein identification.** Purified virus was denatured and separated on 4 to 20% SDS-PAGE gels and stained with Coomassie brilliant blue R250 (Bio-Rad), and all visible protein bands were excised. The gel slices were destained using 50% acetonitrile in 50 mM ammonium bicarbonate (pH 7.9) and vacuum dried. Samples were rehydrated with 1.5 mg/ml dithiothreitol (DTT) in 25 mM ammonium bicarbonate (pH 8.5) at 56°C for 1 h, subsequently alkylated with 10 mg/ml iodoacetamide (IAA) in 25 mM ammonium bicarbonate (pH 8.5), and stored in the dark at room temperature for 1 h. The pieces were subsequently washed with 100 mM ammonium bicarbonate (pH 8.5) for 15 min, washed twice with 50% acetonitrile in 50 mM ammonium bicarbonate (pH 8.5) for 15 min each, vacuum dried, and rehydrated with 4  $\mu$ l of proteomics grade modified trypsin (100  $\mu$ g/ml; Sigma) in 25 mM ammonium bicarbonate (pH 8.5). The pieces were covered in a solution of 10 mM ammonium bicarbonate with 10% acetonitrile (pH 8.5) and incubated at 37°C for 16 h. Solution digestion of viral capsid proteins using thermolysin and LysC followed by trypsin was also performed. The purified capsid buffer consisted of 15  $\mu$ l 50 mM Tris (pH 8.8) and 10  $\mu$ l 1.5 M NaOH. This buffer was added to 200  $\mu$ l of virus in citrate buffer for a final pH of 8.4, the optimum alkalinity for LysC. A volume of 1  $\mu$ l of LysC (proteomics grade; Sigma) was added to 10  $\mu$ l of virus solution in the presence or absence of 8 M urea, DTT, and IAA; the reaction proceeded at 37°C for 12 h. A threefold volume of 50 mM Tris (pH 8.8) was added, followed by the addition of 2  $\mu$ l of trypsin and 1.4  $\mu$ l of 0.1 M CaCl<sub>2</sub>. The samples were incubated at 37°C for 2 h.

**MALDI analysis.** The matrix for matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry was prepared by mixing twice-recrystallized  $\alpha$ -cyano-4-hydroxy cinnamic acid in a solution of 50% acetonitrile with 0.1% trifluoroacetic acid. Volumes of 1  $\mu$ l sample and 0.7  $\mu$ l matrix were spotted onto an aluminum MALDI-TOF plate and allowed to air dry. Calibration was performed using bradykinin and insulin peptide standards. Spectra were collected by using a Biflex III MALDI-TOF mass spectrophotometer (Bruker Daltonics, Billerica, MA) and internally calibrated by using known tryptic digest peptides. Results were compared with theoretical trypsin digests of viral ORFs by using a protein analysis work sheet from ProteoMetrics Inc. (<http://65.219.84.5/paws.html>).

**Nanospray LC-MS/MS.** Peptides eluted from the in-gel digests were analyzed on an integrated Agilent 1100 liquid chromatograph (LC) (XCT-plus) mass-selective detector (MSD) controlled by ChemStation LC 3D (Rev. A.10.02). The Agilent XCT-plus ion trap mass spectrometer is fitted with an Agilent 1100 CapLC and nano-LC sprayer under the control of MSD trap control version 5.2 Build no. 63.8 (Bruker Daltonics GmbH). Injected samples were first trapped and desalted isocratically on a Zorbax 300 SB-C<sub>18</sub> precolumn (5  $\mu$ m, 5 × 300- $\mu$ m inside diameter; Agilent) for 5 min with 0.1% formic acid delivered by the auxiliary pump at 3  $\mu$ l/min. The peptides were then reverse eluted from a precolumn and loaded onto an analytical C<sub>18</sub> capillary column (15-cm by 75- $\mu$ m inside diameter, packed with 3.5- $\mu$ m Zorbax 300, SB C<sub>18</sub> particles; Agilent) connected in-line to the mass spectrometer with a flow of 300 nl/min. Peptides were eluted with a 15 to 50% acetonitrile gradient over 50 min. Data-dependent acquisition of collision induced dissociation tandem mass spectrometry (MS/MS) was utilized. Parent ion scans were run over the  $m/z$  range of 400 to 2,000 at 8,100  $m/z$ -s. MGF compound list files were used to query an in-house database using Biotools software version 2.2 (Bruker Daltonics) with 0.2-Da MS/MS ion mass tolerance. Samples were run in parallel on a Waters Q-TOF Premier mass spectrometer equipped with a nano-Acquity LC system and MassLynx 4.0 software under similar conditions.

**Homology searches for the STIV structural proteins.** The amino acid sequences of the identified proteins were analyzed using BLAST and PSI-BLAST (40) homology searches against the nonredundant protein database at NCBI. An automated procedure implemented in SAM (28) was used to align all identified sequences and to generate profile hidden Markov models. Sensitive profile-

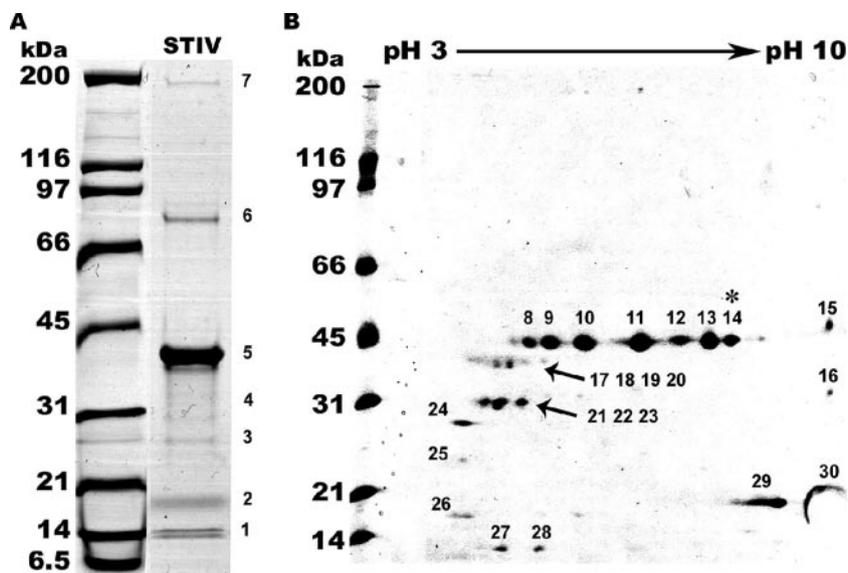


FIG. 2. Analysis of STIV capsid proteins by SDS-PAGE. Purified virus was denatured and separated (A) on a 4 to 20% SDS-PAGE gel or (B) on isoelectric focusing (3 to 10 nonlinear) gel strips and then separated in the second dimension on an 8 to 18% SDS-PAGE gel. Gels stained with Coomassie brilliant blue and individual protein spots and bands were then excised and digested with thermolysin and/or trypsin, and the peptides were analyzed using mass spectrometry. The unmodified major capsid protein expressed in *E. coli* runs similarly to 14 (\*).

profile comparisons (43) were used for alignments with known structures. Putative transmembrane domains in each structural protein sequence were predicted by TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) and DNASTAR (Lasergene Inc., Madison, WI). Signal peptide prediction for each protein was performed using SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) and Psort 2.0 (<http://bioweb.pasteur.fr/seqanal/interfaces/psort2-simple.html>) against the SwissProt database, with each search submission consisting of less than 80 amino acids taken from the N terminus of the protein. Sequence repeats were analyzed using EMBL-EBI rapid automatic detection and alignment of repeats (<http://www.ebi.ac.uk/Radar/>). Protein fold recognition was done using one-dimensional (1D) and three-dimensional (3D) sequence profiles coupled with secondary structure information as implemented in the 3D position-specific scoring matrix (PSSM) (29) (<http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html>).

**Protein modeling and docking into the cryo-EM map.** Sequence profiles generated as described above were used for profile-profile comparisons (43) with various databases of protein families, including the collection of protein structures in PDB (15). Alignments obtained in this fashion were used for 3D modeling (39). Based on model evaluations (39), the alignments were adjusted manually and used for additional cycles of model building until no further improvements were observed. Structural models of B164 were docked into a reconstructed cryo-EM map using 3SOM (11).

**Glycoprotein analysis.** The Pro-Q Emerald 488 glycoprotein gel and blot stain kit (P21875; Molecular Probes) were used to visualize glycosylated proteins in 1D and two-dimensional (2D) gels. Briefly, purified virus solutions were adjusted with 1/4 volume of 4X Pro-Q buffer. The samples were then boiled for 2 min at 100°C before being loaded onto a 4 to 20% SDS-PAGE gel. Molecular weight markers, the CandyCane glycoprotein molecular weight marker (Molecular Probes; C21852), and SDS-PAGE standard broad range (Bio-Rad) were used as standards. The gel was run for 44 min at 120 V, rinsed in distilled H<sub>2</sub>O (twice for 10 min), and fixed with a solution of 50% methanol and 5% glacial acetic acid (twice for 1 h). The gel was then washed twice with 3% glacial acetic acid and incubated in oxidizing solution (component C, Pro-Q Emerald glycoprotein gel and blot stain kit) for 40 min. The gel was washed again (three times for 15 min). The gel was stained in the dark using 25 ml of Pro-Q Emerald 488 staining solution for 4 h, followed by two more 15-min washes. Glycoprotein gels were analyzed using a Typhoon scanner (blue laser, 488 nm) and stained using Sypro Ruby (Bio-Rad) and Coomassie brilliant blue R250 (Bio-Rad).

For an analysis of potential N glycosylation 20  $\mu$ l of STIV and *S. solfataricus* P2 was solubilized in 7 M urea–2 M thiourea–4% CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}–50 mM Tris-HCl (pH 8.8) buffer. Solubilized capsid proteins were acetone precipitated and resuspended into 40  $\mu$ l phosphate-buffered saline (PBS) buffer, pH 7.4. An equal volume of glycopro-

tein-denaturing buffer (BioLabs) was added. The solution was boiled at 100°C for 10 min. A 1/10 volume of NEBuffer for G7 (BioLabs) was added, followed by a 1/10 volume of 10% NP-40 (BioLabs). Of this solution, 40  $\mu$ l was kept as the enzyme control. The remaining solution received 2  $\mu$ l of the enzyme PNGase F (BioLabs). This digestion solution was then split, with one half being incubated at 37°C for 1 h. To test for the presence of O glycosylation, STIV and *S. solfataricus* P2 proteins were resuspended in 50 mM PBS, pH 7.4, mixed with 2  $\mu$ l of O-glycosidase (Roche Applied Science), and incubated at 37°C overnight.

**Lipid analysis.** Viral and host cell lipids were extracted and analyzed as described previously (18). Briefly, each sample was acetone precipitated and the air-dried pellet was solubilized with chloroform-methanol (1:3). Virus samples were diluted 100-fold and analyzed by direct infusion electrospray in both positive and negative modes on a micro-TOF (Bruker Daltonics). Instrument settings were as follows: capillary, 4,000 V; capillary exit, –240 V; dry gas (N<sub>2</sub>), 3 liters/min; temperature, 200°C. The negative mode had a more intense and consistent signal in the expected *m/z* range for tetraether lipids (900 to 2,000), as previously reported (33).

## RESULTS

**SDS-PAGE profile of purified viral capsid proteins.** Highly purified samples of STIV reproducibly contained seven prominent bands when analyzed under denaturing conditions (Fig. 2A). 2D gel analysis of the same sample revealed that a number of the proteins were present in multiple forms (Fig. 2B). Twenty-three protein spots were resolved between 14 to 60 kDa within the pI range of 3 to 10. The higher-molecular-weight proteins present on the 1D gel were not visualized in the 2D analysis, suggesting that they may have extreme pIs or be hydrophobic. The protein spots at ~40 and 45 kDa on the 2D gel (spots 17 to 20 and 8 to 15, respectively) represent multiple forms of two viral proteins (Fig. 2B). The isoforms differ in pI, but their molecular weights are the same or nearly the same. Posttranslational modifications which alter the pI but do not significantly change the molecular weight offer a possible explanation for the observed pattern. Nomenclature for STIV proteins is based on the reading frame (A, B, or C)

TABLE 1. Virus-associated proteins

Identified protein	Mass (Da)	pI	Sample origin			Predicted structure or function
			Band no.	2D spot no.	In solution	
<b>STIV</b>						
C557	58,574	5.5	6		+	Protein interaction
C381	41,655	5.56	4, 7	17, 18, 19, 20	+	PRD1 P5 vertex protein
B345 (coat protein)	37,810	6.17	5	8, 9, 10, 11, 12, 13, 14, 15, 16	+	Major capsid protein
A223	24,410	4.72	7	24	+	PRD1 P5 vertex protein
B164	19,025	9.29	3		+	Poxvirus ATPase
B130	13,768	4.97	1	27, 28	+	NSM <sup>b</sup>
B109	11,969	5.04	2, 5	17, 18, 25	+	NSM
*A78	9,610	10.79	1		+	NSM
A55	6,338	4.49	1		+	NSM
<b>Host</b>						
7 DNA-binding protein-kDa (SSO7D)	7,735	9.52			+	DNA binding
Conserved hypothetical protein SSO0881	25,020	5.86	4		+	VPS24 vacuolar sorting protein
<b>Others</b>						
PIC/E <sup>a</sup>	31,540	5.07		17, 18	-	ASP <sup>c</sup>
PIC/E	≈17,000	≈8.5-9.5		29, 30	-	ASP
PIC/E	≈17,000	≈4.7		26	-	ASP
PIC/E	≈26,000	≈4-5		21, 22, 23	-	ASP

<sup>a</sup> PIC/E, protease inhibitor cocktail and endonucleases added during the sample preparation for analysis by 2D electrophoresis.

<sup>b</sup> NSM, no significant match.

<sup>c</sup> ASP, added during sample preparation.

and predicted number of amino acids. To investigate this further, the major capsid protein of STIV, B345, was expressed in *Escherichia coli* and this protein was analyzed with a 2D gel (30). The *E. coli*-expressed protein comigrated with the most basic isoform of B345 found in infectious virus particles (spot 14). Analysis of the spot pattern using the protein modification screening tool ProMost (23) (<http://prometheus.brc.mcw.edu/promost/>) indicated that differential phosphorylation and/or modifications to basic side chains were likely low-molecular-weight posttranslational modifications that could account for the observed gel shifts. Alternative explanations include heterogeneous glycosylation with sulfate or acidic carbohydrates and isoelectric focusing artifacts.

**Identification of STIV proteins by mass spectrometry.** In-gel digestion and mass analysis were conducted to determine the identity of the protein bands and spots (Fig. 2). Seven bands and 23 spots were excised from the gels for analysis. Protein identifications were based on peptide mass and fragmentation pattern matches in the NCBI nr database and an in-house database that contained the STIV genome translated in all six reading frames and the complete genome of *S. solfataricus* P2. The in-house database contained all ORFs greater than 24 nucleotides in the STIV genome and also included a pyrrolysine codon in place of the amber stop codon (12, 24, 26). In-gel digestion and nanospray MS/MS analysis using both an ion trap and Q-TOF instruments led to the identification of nine STIV-encoded proteins (Table 1). The amino acid sequence coverage ranged from 20 to 90%, and confidence in the identifications was based on scores from MASCOT (Matrix Science Inc.) and MassLynx v 4.0 (Waters Inc.) (see supplemental material at [www.chemistry.montana.bothner/jvi\\_stiv2006](http://www.chemistry.montana.bothner/jvi_stiv2006)). MASCOT protein scores were highly significant for each of the reported proteins and ranged from 52 to 727 using  $\pm 0.5$ -Da error

tolerance (Table 1). Only peptides with individual scores of greater than 40 were included in the protein scoring. These results were confirmed using MALDI-TOF analysis of in-gel digestion samples. Additional LC-MS experiments used direct in-solution digestion of purified particles. In addition to the virus proteins, two host proteins were identified. Three independent viral preparations were used in these experiments, and the same two host proteins were identified each time. One is small basic protein SSO7D (7 kDa), of unknown function. The other virally associated host protein, SSO0881 (25 kDa), is similar to the vacuolar sorting VPS24 domain that is found in both *Saccharomyces cerevisiae* and mammals (48, 51) and is conserved across the *Sulfolobus* family.

Of the nine STIV proteins identified here, only the major capsid protein (B345) had previously been confirmed in virus particles (38). Seven of the nine proteins are ordered sequentially within the genome, suggesting coordinated transcription and translation (Fig. 3). Interestingly, the last protein in this series (A78) was not included in the original annotation. The presence of this protein is based on a MASCOT score of 108 from MS/MS sequence data and peptide mass mapping from MALDI-TOF MS. Moreover, MassLynx reported a probability of 99.5% for two peptides with an average mass error of 4.9115 ppm. The peptide closest to the N terminus begins with a valine whose codon is nine nucleotides downstream of the A55 stop codon (see supplemental material at [www.chemistry.montana.bothner/jvi\\_stiv2006](http://www.chemistry.montana.bothner/jvi_stiv2006)). A78 may be a read-through product from the A55 ORF or represent the use of an alternative start codon.

**Analysis of identified proteins.** Thirty-six ORFs were predicted in the original annotation of the STIV genome (38). Standard BLAST searches based on the predicted amino sequence encoded by the ORFs failed to identify homologous

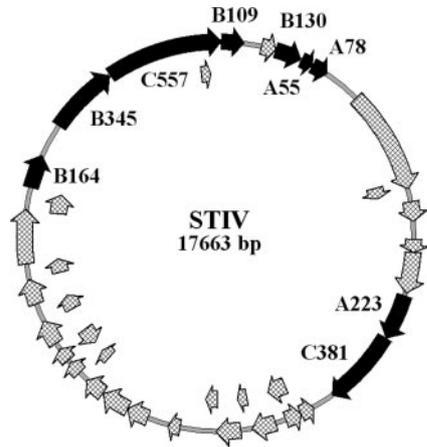


FIG. 3. Location of capsid proteins on the STIV genome map. ORFs are named according to frame (A to F) and number of predicted amino acids. Proteins identified by protease mass mapping (solid black arrows), in general, cluster together. There are eight proteins from annotated ORFs and one additional protein, A78. This protein lacks a start methionine and was not included in the original annotation. The map was created using Vector NTI Advance 10.1.1.

proteins in the NCBI nonredundant database. In evolutionarily distant proteins, structural homology is likely to be retained even when the primary sequences have diverged beyond detection. Therefore, as an alternative approach, structural prediction was used to identify potential functional roles for these proteins (Table 1).

Structural predictions of the first 347 N-terminal residues of C557 show high similarity to an ankyrin repeat (3D PSSM confidence > 95%) that mediates protein-protein interaction (16) and the presence of an ~30-amino-acid repeat. The sequence (VEKVIQKDITHPPKFYLPVHLPNIHQIEAG IGH) is repeated three times between residues 32 and 130 (see supplemental material at [www.chemistry.montana.bothner/jvi\\_stiv2006](http://www.chemistry.montana.bothner/jvi_stiv2006)). Each repeat contains a conserved YLP motif (underlined), which is involved in protein-protein interaction in many organisms. For example, this motif has been found in

receptor tyrosine kinases and members of the ErbB family of human epidermal growth factor receptors (20, 50). The C-terminal region is marked by a proline- and serine-rich motif, the functional significance of which is currently unknown. In an attempt to determine relative capsid composition a densitometry analysis of the 1D gel using ImageJ was performed. This shows a 15:1 ratio between the major capsid protein (B345) and C557. However, this ratio may not be accurate due to posttranslational modifications (see below).

PSI-BLAST (40) searches with B164 identified significant matches to a large class of P-loop ATPases. Profile-profile comparisons (43) showed the most significant similarity ( $E = 1 \times 10^{-33}$ ) to poxvirus A32 proteins (Pfam accession number PF04665; Fig. 4), which are thought to be involved in viral DNA packaging (10). This group of proteins, however, has no close homologue of known structure. For that reason, several less-related ATPases (with  $E$  values between  $10^{-7}$  and  $10^{-13}$ ) were used as templates for B164 model building in order of their match significance (39). After several cycles of manual realignment and building, we generated models that were of good quality regardless of the starting template. It should be pointed out, however, that even though these models are likely to have the correct overall fold, they are not expected to be atomic-quality models because the starting templates were considerably different at sequence level from B164. Each of these models was fitted into the cryo-EM reconstruction of STIV (30) by surface overlap maximization (11). This procedure is specifically designed to place partial structures or models into low-resolution density maps by maximizing surface overlap. All models docked into the same general area at the base of the turret on the fivefold axis of symmetry. Depending on the template used for modeling, the surface overlap between models and the map was between 0.6 and 0.78, with the corresponding correlation coefficients in the 0.5 to 0.6 range. Importantly, when we attempted to dock more than 12 models at a time, the new solutions overlapped with the existing ones, thus indicating that the position at the base of turrets is the only one compatible with B164 models.

Two virion-associated proteins share predicted structural

<p>B164                  Q9Q8V8_RPOXV                  Q6TUP8_YMTV                  Q8V3H8_SWPV                  Q98306_MCV1                  O57235_POXVV                  Q80FJ6_9POXV                  Q9J536_FOWPV                  Consensus 90%</p>	<p>MNPDDIVVITCRKRS<del>GS</del>SYLIKHYIPVLKAHKISYIIDHNLRS<del>GSEYSK</del>FGYNVVTLS<del>DVSKO</del>VVYVDREKNDVFFEKLW                  LMSPFRIALV<del>CGSGSGKTAYLLSL</del>ENTLVCKYKHFIFLFTP--VYNSAYDSYVWPDHVNKVTTP<del>EELD</del>ALITTKQKIE-----K                  LKHPFRIALV<del>CGSGSGKTAYLLSL</del>ENTFVEKYRHIFLFTP--VYNSAYDSYVWPDHVNKVTTP<del>EELD</del>SLITTKQKIE-----K                  NKSPFRIALV<del>CGSGSGKTAYLLSL</del>ENTLIDKYKHFIFLFTP--VYNEAYDSYVWPDHVNKVTTP<del>EELD</del>SLVTTKQKIE-----K                  LRAPFRMALV<del>CGSGSGKTLYLLSL</del>ERTLVGAYRHIFLFTP--VLNSAYDGYVWPDHIQKVSSHEELEYTL<del>SVAKKIE</del>-----R                  LKMPFRMVL<del>CGSGSGKTIYLLSL</del>STLVKKYKHVFLFTP--VYNPDYDGYVWPNHIN<del>FVSSQESLE</del>YNLIRTKSNIE-----K                  LSDYFRMVL<del>CGSGSGKTIYLLSL</del>KTFTVTKYKHFIFLFTP--LINPSYNYVWPDHYK<del>IADFSELE</del>YALMKMRDDL-----N                  LNDYFRMVL<del>CGSGSGKTIYLLSL</del>KTFTVTKYKHFIFLFTP--LINPSYNYVWPDHYK<del>IADFSELE</del>YSLSKMKQDLV-----N                  .....bhslhg...sgkshh1b.hf.sh1..ach.alhss..lh.....as..l..lss...b.Ys1...c.....</p>
<p>B164                  Q9Q8V8_RPOXV                  Q6TUP8_YMTV                  Q8V3H8_SWPV                  Q98306_MCV1                  O57235_POXVV                  Q80FJ6_9POXV                  Q9J536_FOWPV                  Consensus 90%</p>	<p>NGAKLHAKKW<del>TSV</del>ITIDEAYYHFKYRQKVTPAIDEALHANRHAGIGLILST<del>RVYDLMP</del>IVYKQADLIIM--FIPESLMN                  YIE-CKGTTKADMFLLITLDD---MGDKQTKSSCLLDLFLNHGRHLN<del>TSV</del>VLLC<del>Q</del>TYKHPVNGRTSITHFCCCNVSDSDIEN                  YVKKQGNDKTTEMFLLITLDD---MGDKQIRSSCLLDLFLNHGRHLN<del>TSV</del>IILL<del>C</del>QTYKHPVNGRTSITHFCCCNVSDSDIEN                  YIE-SKGTKNADMFLLITLDD---MGDKQTRSSCLLDLFLNHGRHLN<del>TSV</del>IILL<del>C</del>QTYKHPVNGRTSITHFCCCNVSDSDIEN                  YVA---GSRGGERFLLITLDD---LGMQMLRSRTLLGLLNYGRHLN<del>TSV</del>LAML<del>C</del>QTYKHPVNGRASITHLCCCNVSESVDEN                  CIAVAQNHHKSAHFLLITLDD---VGDKLSKCNLTIEFLNFRHGRHLN<del>TSV</del>IILL<del>C</del>QTYRHVPILGRANITHFCFNISISDAEN                  L--EKKGSS--NHNFLITLDD---LGAQLKSKILPWL<del>VNTGR</del>HKMSIVMIC<del>Q</del>TYRHVPTNCRSSITHLCCCNVSDADVEN                  LG---KRGSSINHKFLVITLDD---LGMQMLKSKILSWL<del>VNTGR</del>HIRMSIVML<del>C</del>QTYRHVPSNCRSSITHLCCCNVSDADVEN                  .....s..hLI.cs...h...b..s.h1..h1.hsRHh.hs1hh.sq.hbc1...sb...hshhh...h...shbN</p>

FIG. 4. B164 alignment with several proteins from the family Poxviridae. SwissProt accession numbers are separated by an underscore from species abbreviations (FOWPV, *Fowlpox virus*; 9POXV, *Vultur gryphus poxvirus*; POXVV, *Vaccinia virus*; MCV1, *Molluscum contagiosum virus* subtype 1; SWPV, *Swinepox virus*; YMTV, *Yaba monkey tumor virus*; RPOXV, *Rabbit fibroma virus*). Residues are shaded according to 90% consensus, with white letters on black background signifying perfect conservation. Abbreviations on the consensus line are as follows: h, hydrophobic; s, small; l, aliphatic; b, big; a, aromatic; c, charged.

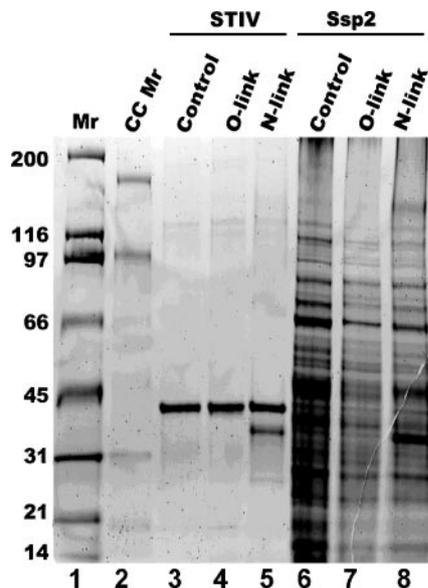


FIG. 5. Glycosylation of STIV proteins. Purified virus and whole-cell lysates of *S. solfataricus* P2 were digested with *O*- or *N*-deglycosidase enzymes and loaded onto a 4 to 20% SDS-PAGE gel. The proteins were analyzed for glycoproteins and visualized using a Pro-Q glycoprotein stain. CC Mr, CandyCane glycoprotein marker. The glycosylated major capsid protein ran at ~41 kDa (lanes 3, 4, and 5). The band at ~35 kDa is the *N*-link-specific glycosidase and not a deglycosylated product (lane 5). In contrast to the viral proteins, total Ssp2 proteins show a significant reduction in glycosylation after treatment (lanes 7 and 8).

similarity to the P5 protein of PRD1. Bands 4 and 7 and spots 17, 18, 19, and 20 (Fig. 2) contained STIV protein C381. Though this protein shows no primary sequence similarity to characterized proteins, fold recognition algorithms (29) predict similarity to the vertex protein P5 of PRD1 (32). Interestingly, the fourth identified STIV protein, A223, also has significant structural similarity to the P5 protein of PRD1. Collectively, these two predictions suggest that A223 and C381 may have similar functions, which is consistent with the formation of heteromers (discussed below).

Docking of the X-ray model of B345 into the cryo-EM reconstruction revealed that the C terminus of this protein interacts with the proposed lipid layer (30). In addition, a search using Tmpred indicates that STIV proteins B130 (residues 110 to 127), C381 (354 to 373), C557 (365 to 383 and 538 to 556), and A55 (37 to 54) also possess possible transmembrane domains. However, known sequence motifs for signal peptides in bacteria and eukaryotes were not detected by either Psort version II or SignalP v3.0. This does not appear to be due to the archaeal origin of these proteins as these algorithms have proven valuable in accurately identifying both transmembrane regions and signaling domains in large-scale analysis of archaeal species (6).

**Glycosylation of STIV proteins.** The description of a glycosyltransferase in the genome of STIV (31) prompted an investigation into the glycosylation status of virion-associated proteins. We tested for glycosylation by staining 1D and 2D gels with the carbohydrate-specific fluorescent stain Pro-Q (Fig. 5). The band corresponding to the major capsid protein was

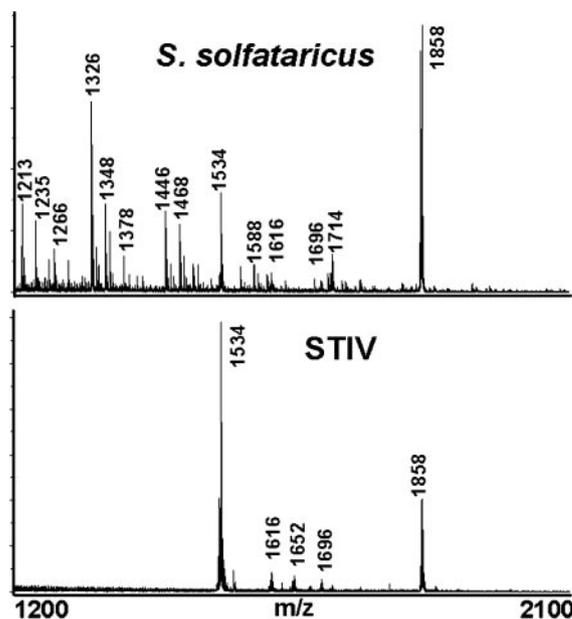


FIG. 6. Total lipid extracts from host and viral membranes. Lipids isolated from host cells (upper) and STIV (lower) were analyzed by negative-mode electrospray mass spectrometry. Viral lipids are a subset of those found in the host.

clearly positive for glycosylation. In an attempt to determine the nature of the protein-polysaccharide linkages, samples were treated with enzymes specific for *O*- and *N*-link glycosidic bonds. A variety of reaction conditions were used, without significant deglycosylation (lanes 4 and 5 in Fig. 5). Control reactions using total isolated SP2 protein indicate that the enzymes were active under the reaction conditions used, and both *O*- and *N*-link treatments led to dramatic reductions in the overall level of glycosylation (lanes 7 and 8 in Fig. 5). At the present time, the nature of the viral carbohydrates is not known. It is possible that, due to the highly stable nature of these thermophilic proteins, attempts at denaturation were not successful or that these virus-associated glycans are not recognized by standard enzymes. Interestingly, glycosylation has been shown to increase protein thermal stability (47) and modulate recognition and is present on Vp54 from PBCV-1.

**STIV is a lipid-containing virus.** The initial cryo-EM reconstructions of STIV suggested that the particle contained an internal lipid layer (38), and the recently improved model (30) strengthened this proposal. However, direct biochemical evidence for the presence of lipids in STIV has not been reported. To this end, we initiated a set of experiments to determine if lipids are present and to identify the nature of these molecules. Total lipids were extracted from *S. solfataricus* P2 host cells and purified virus using a chloroform-methanol (1:3) solution. This solvent system has been shown to enhance the variety of lipids that can be detected and allows a distinction between neutral and acidic lipids (33). Electrospray ionization-time of flight (ESI-TOF) analysis was conducted in both positive- and negative-ionization modes. The analysis of tetraether lipids in positive-ionization mode is less representative of sample composition as spectra tend to be dominated by neutral species (33). Extracted lipids were diluted to equal concentrations with

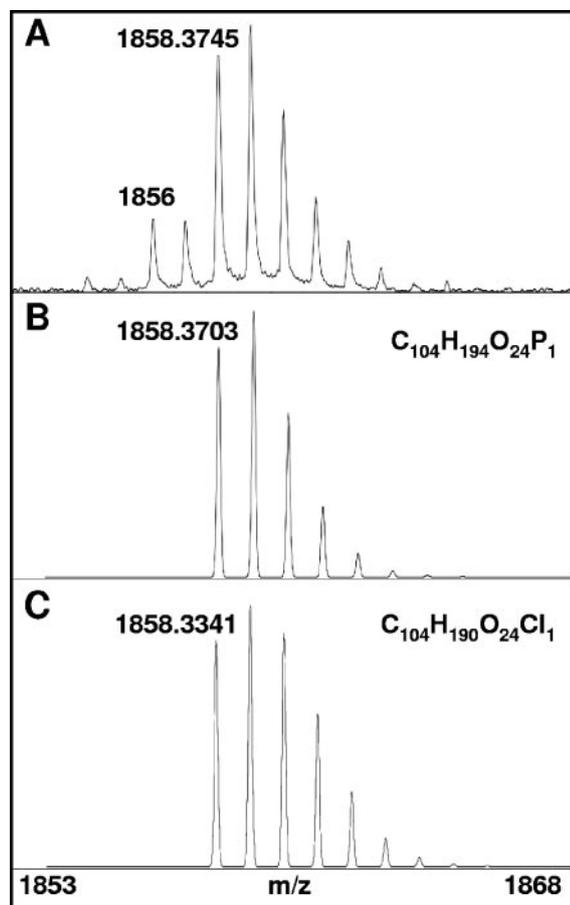


FIG. 7. Experimental and theoretical isotope distributions for the lipid at  $m/z \sim 1,858$ . A) Close-up of peak at  $m/z$  1,858 in the STIV sample. The minor component present at  $m/z$  1,856 has an additional cyclopentane group in the iospranyl chain. Based on the unit molecular weight, this could be PGL1 or PGL2 with four cyclopentane groups (B). However, a closely related neutral molecule that has become ionized by chloride differs by only 0.0362 amu (C). Theoretical isotope distributions for the acidic (B) and neutral (C) lipids clearly show the effect chloride has on the pattern. The influence is largest 2 mass units above the monoisotopic peak. A comparison of the STIV sample and the two theoretical spectra reveals that STIV has the acidic lipid. The measured  $m/z$  of  $1,858.3745 \pm 0.009$  also matches that for the acidic species ( $1,858.3703m/z$ ) but not that for the neutral molecule plus  $\text{Cl}^-$  ( $1,858.3341m/z$ ).

the same solvent system and analyzed directly by ESI-TOF. The negative-ionization mode mass spectrum of total lipids from *S. solfataricus* P2 revealed a complex mixture of lipid species in the  $m/z$  range 1,100 to 1,900 (Fig. 6). In contrast, the lipids isolated from virus particles were highly enriched for a subpopulation of those present in the host. The ions detected represent  $[\text{M}-\text{H}]^-$  species for acidic lipids and  $[\text{M}+\text{Cl}]^-$  for neutral molecules (33).

*Sulfolobus* species contain a wide variety of cyclic tetraether lipids (25, 33, 35). Eight different types of tetra-*O*-di(biphytanyl) diglycerol derivatives have been described for *Sulfolobus*, and each of these can vary depending on the number of cyclopentane rings incorporated in the biphytanyl chains. Each chain can have up to four cyclic groups, and this varies depending on growth conditions (14). Because of this diversity

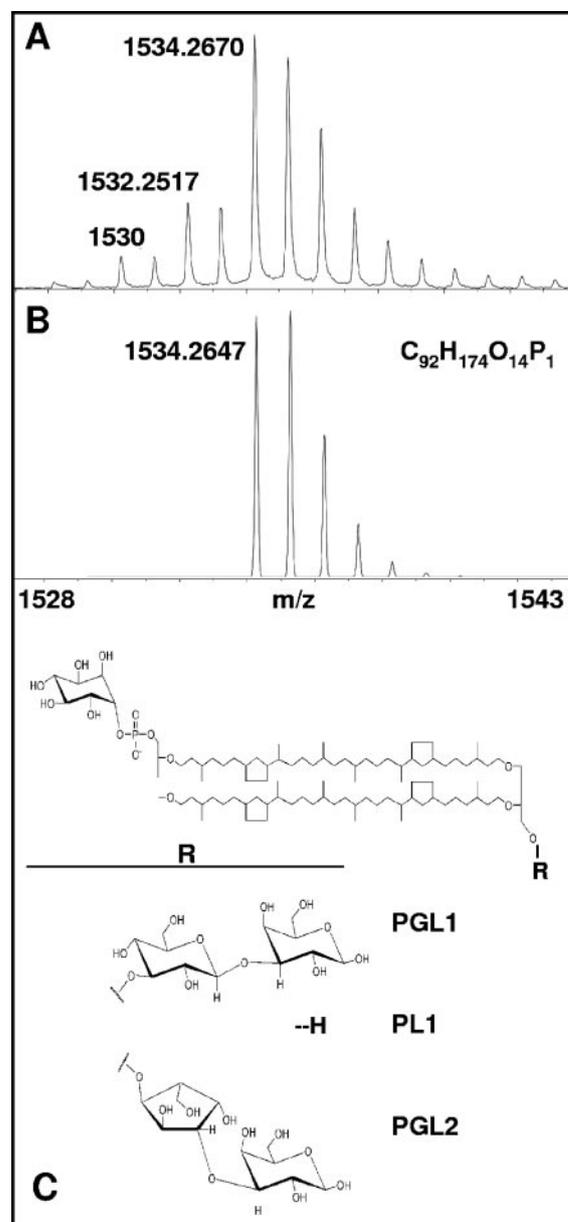


FIG. 8. Isotope pattern for the ion at  $m/z$  1,534. Shown are the experimental (A) and theoretical (B) patterns for the STIV predominant lipid molecule. The acidic lipid has a measured mass that is consistent with PL1. Approximately 30% of this lipid is present with an additional cyclopentane unit ( $m/z$  1,532), and a small amount has a second cyclopentane unit ( $m/z$  1,530). (C) Lipids found in STIV. The cyclic tetraether lipids found in *Sulfolobus* have the general backbone shown above. The number of cyclopentane groups per chain is variable.

nominal mass measurements alone cannot provide a definitive identification. However, isotopic distribution and accurate mass measurements allow a facile distinction between neutral and acidic molecules. Chlorine atoms are a mixture of two isotopes,  $^{35}\text{Cl}$  and  $^{37}\text{Cl}$ , with abundances of 75.8 and 25.2%, respectively. The isotopic pattern of lipids is dominated by  $^{12}\text{C}$  and  $^{13}\text{C}$ , with abundances of 98.9 and 1.1%, respectively. The presence of a chlorine atom alters the isotopic distribution, and

therefore the observed ratio can be used to distinguish between neutral and acidic lipids (33).

A close-up view of the peak at  $m/z$  1,858 reveals the isotope pattern found in the STIV lipid (Fig. 7A). The monoisotopic peak for the major component was at  $m/z$  1,858, and there was a minor component at  $m/z$  1,856. The unit molecular weight suggested that this is either PGL1 or PGL2 with four cyclopentane groups ( $C_{104}H_{194}O_{24}P_1$ ). However a closely related neutral molecule with a chloride molecule attached ( $C_{104}H_{190}O_{24}Cl_1$ ) differs by only 0.0362 amu. Theoretical isotope distributions for these molecules clearly show the effect chloride has on a molecule (Fig. 7B and C). The influence is largest 2 mass units above the monoisotopic peak. A comparison of the STIV sample and the two theoretical spectra reveals that this is an acidic lipid. The measured  $m/z$  of  $1,858.3745 \pm 0.009$  also matches that for the acidic species ( $m/z$  1,858.3703) but not that for the neutral lipid ( $m/z$  1,858.3341). The second major component of the STIV lipid layer,  $m/z$  1,534.2670  $\pm$  0.0077, is consistent with PL1 having six cyclopentane groups ( $C_{92}H_{170}O_{14}P_1$ ;  $m/z$  1,534.2647). The expected isotope ratio also matches that for the acidic species (Fig. 8). The structures of these interesting cyclic lipids are shown in Fig. 8C. Small numbers of both PGL1/2 and PL1 molecules have higher degree of cyclopentane incorporation. Each cyclized component reduces the mass by two, and these minor species are seen to the left of the monoisotopic peak in Fig. 7 and 8.

## DISCUSSION

Archaeal viruses isolated from extreme environments are powerful tools for elucidating the biology of their archeal hosts and for examining the effects of extreme physical conditions on supramolecular complexes. Structural similarities with the major capsid protein to prokaryotic (PRD1) and eukaryotic viruses (PBCV-1 and adenovirus) strongly suggest that a common ancestor gave rise to this group of dsDNA viruses in all three domains of life (9, 30). The recent structural model for STIV based on cryo-EM data revealed a novel pseudo-T=31 icosahedral capsid with large turrets at the fivefold vertices. This structure also revealed two layers of density below the protein shell, believed to represent an internal lipid layer. In spite of the impressive structural models, the identity of capsid proteins, except for B345, the composition of the particles, and the nature of the internal layer were not known. With the use of general proteomics and mass spectrometry techniques, we have characterized the protein composition of STIV particles, detected carbohydrates, and confirmed the presence of a lipid layer.

Proteomic analysis of gradient-purified virus particles in solution and after 1D and 2D gel analysis led to the identification nine virally encoded proteins. One of these proteins, A78, lacks a canonical initiation signal and was not listed in the original genome annotation (38) (Fig. 3). Recoding of genetic information is well documented in archaea, including the use of nonstandard amino acids, frameshifts, read-through of amber codons, and alternative sites for the initiation of transcription (12, 26). To assure that virally expressed proteins were not overlooked, we reformatted our protein sequence database such that ORFs began immediately after stop codons and allowed for read-through products of the amber codon. The

identification of A78 illustrates the utility of a proteomics-based approach. Due to the short lengths of the A78 ORF (234 bp) and the preceding A55 ORF (165 bp), these two ORFs may encode a single protein in which the C terminus is a read-through product. It is noteworthy that these ORFs are separated by an amber stop codon, a common site of recoding in archaea (26). The combined ORF has been tentatively annotated A137b, and this is under further investigation.

The genes coding for seven of the nine virally encoded proteins detected in the particle (Table 1) are clustered together on the viral genome map (Fig. 3). The sequential order of these proteins is indicative of a shared transcriptional regulatory mechanism, likely in an operon organization. The use of operons to express and regulate related clusters of genes is common in *Archaea* and has also been observed in other viruses that replicate in *S. solfataricus* (36).

A223 and C381 had apparent molecular masses of  $\sim$ 200 kDa on 1D SDS-PAGE reducing gels (Fig. 2A); however, their predicted masses are 24 kDa and 42 kDa, respectively. These proteins have significant hydrophobic and aromatic compositions: A223, 42% and 10%; C381, 24% and 12%. Their anomalously high molecular masses in 1D gel analysis suggest that they may reside in a complex that resists disruption. This behavior has been described for viral structural proteins from mesophilic organisms, such as PBCV-1 (34). Under the conditions used for 2D gel analysis, the proteins were detected close to their expected masses (Fig. 2B), lending support to the presence of highly stable noncovalent multimers.

It is tempting to speculate that a pentameric arrangement of C381 and A223 would be  $\sim$ 320 kDa and could be involved with formation of the turret structures present at the fivefold vertices. The turret structures are the most striking feature of STIV (Fig. 1). It has been suggested that these are used for cell entry and DNA translocation (30, 38). As described above, proteins A223 and C381 form highly stable oligomers. Results from the fold recognition algorithm PSSM (29) indicate that both A223 and C381 are structurally similar to the vertex protein P5 of PRD1 (32). Since the vertex of PRD1 is responsible for host cell binding and dsDNA delivery (21), we propose that putative turret components A223 and C381 have similar functions. Because icosahedral averaging was used during refinement of the cryo-EM data, there is currently no evidence to suggest that a unique vertex is present, as has been shown for PRD1 (21).

A third STIV protein possibly involved in turret formation is C557, which has a predicted molecular mass of 58.6 kDa. As described above, this protein has three repeats of a 30-amino-acid sequence in the N terminus and a second proline-rich repeat in the C terminus along with two potential transmembrane domains. Based on the size of this protein and the structural model of the capsid, C557 would be expected to be part of the turret. Threading results indicate that an ankyrin-like fold is likely (PSSM score  $>$  95%) for the N-terminal repeat region. This fold is abundant in the genomes of a wide range of viruses (2, 3, 27); however, to our knowledge it has not been found in a structural model of archaeal proteins to date. It is of interest to note that the topology of ankyrin domains resembles a curved cylinder, like the petals attached to the turrets. The estimated mass of the turret structure, minus the cementing protein, is 637 kDa (30). We propose that each

pentamer subunit is composed of one copy of C557, C381, and A223. Together, these three proteins have an expected mass of 125 kDa, close to the 127 kDa estimated from the electron density present at each fivefold vertex. The ratio of C557 to the major capsid protein (B345), based on protein gel densitometry, was determined to be 1:15. This ratio is not consistent with the predicted ratio of  $\sim$ 1:30, based on the T=31 architecture of the STIV capsid. This underrepresentation of B345 is likely due to inefficient protein staining because of the glycosylation and the fact that on a molar basis large proteins stain more intensely than small ones (19).

Both sequence- and structure-based analyses suggested that B164 is a virally encoded ATPase. This 19-kDa protein is highly basic and has a nucleotide binding sequence motif found in all ATPases. This was confirmed by model building based on four known ATPase structures, each of which gave good models. Docking experiments consistently positioned a single copy of the protein at the base of the turret complex. It should be noted that we used homology models of B164, not actual B164 structures, and that the map was of relatively low resolution (30). This leaves some uncertainty about the fine details of B164 position within the capsid. This issue will be addressed when the B164 structure or better-homology models become available. Nevertheless, we are confident that the placement of B164 is largely accurate to the extent allowed by our homology models and the resolution of the map. There are several indirect lines of evidence that support this observation. First, this position of B164 is consistent with the expected location of a protein involved with nucleic acid packaging and/or release (1, 10). Second, profile-profile comparisons (43), arguably the most sensitive sequence-based approach to remote homology detection, identified putative viral packaging ATPases (10) as the closest relatives of B164 (Fig. 4), further strengthening the possibility that B164 may be involved in nucleic acid packaging. Finally, the central axis of the turret lacks substantial density except at the ends and its channel has a 3-nm opening (38) that could accommodate dsDNA, again consistent with the transport of nucleic acid (8, 21, 38).

Two host proteins were consistently found to copurify with STIV particles. SSO7D is a small (7-kDa), basic (pI 10.12) DNA-binding protein (22) that was identified with high confidence. We speculate that this protein plays a direct role in stabilizing and packaging the viral DNA within the virion. It has been suggested that the same 7-kDa protein plays the role of a histone-like protein in association with the *S. solfataricus* genome. A protein of similar size and charge, VP2, is found in purified preparations of the SSV1 (*S. solfataricus* spindle-shaped virus 1) particle (52). VP2 binds viral DNA in a cooperative manner and is suspected of playing a role in packaging the viral genome. The role of VP2 in packaging was very compelling until homologues were not found in other SSV genomes (49). An explanation for the absence of this protein is that the viruses have usurped a host protein to aid in the packaging of their genomes. Identification of proteins similar to SSO7D in SSV2, SSVKm, and SSVRH particles would provide compelling evidence for this hypothesis. A second host-derived protein, SSO0881, is a conserved hypothetical protein similar to yeast and mammalian VPS24, which has been implicated in cell sorting and trafficking (48, 51). It seems unlikely that the two detected host proteins in STIV prepara-

tions are merely contaminants because they consistently cosediment with particles, although this cannot be ruled out. Further experimentation will be required to determine if these host proteins play a functional role in the life cycle of STIV.

The composition of the suspected internal lipid layer of STIV has been unknown. Extraction of lipids from host cells and purified virus particles and subsequent analysis using negative-mode electrospray ionization showed that STIV does contain lipids and that this layer is enriched in a subpopulation of those present in the host (Fig. 6). A similar approach to lipid extraction was used to show that SH1 virus particles from *Haloarcula hispanica* selectively acquire a subset of lipids from the host pool (4). We were able to make identifications of the two major lipid components found in the STIV particles. Accurate mass measurements and isotope distribution were consistent with the acidic tetraether molecules PL1 and PGL1-like (Fig. 8) (33). The major capsid protein, B345, has a basic C-terminal helix that likely interacts with the inner lipid leaflet. The acidic nature of the lipids is therefore consistent, as this would stabilize the protein-lipid interface. This is also consistent with the lipid and protein charge complementation described for PRD1 (13).

The biological function of the internal lipid layer of STIV is not known. Liposomes composed of tetraether lipids such as those isolated from the particles are thermally stable (45), and this layer could help to protect the viral genome. Alternative roles for the lipid could be in particle assembly and or disassembly. The internal lipid layer of PRD1 is believed to function as a mediator of capsid assembly, recruiting structural proteins to the lipid surface (1). As described above, four of the viral proteins identified in this study contain potential transmembrane domains (C557, A55, C381, and B130), and B345 is known to interact with the lipid (30). Interactions between the protein domains and a preassembled lipid layer could initiate formation of the protein capsid, or, conversely, the protein domains could serve as nucleation sites for formation of the lipid layer.

The exterior surface of the virus particle formed by B345 is the likely region for glycosylation. Protein glycosylation has a number of functional roles, one of which is to increase thermal stability (47). Consequently it is not surprising that many of the proteins found in *S. solfataricus* are glycosylated (17). We have shown here that B345 is glycosylated (Fig. 5). Based on SDS-PAGE migration, B345 has a mass of  $\sim$ 41 kDa whereas the expected mass for the apoprotein is 37,790 Da. Even though the protein contains carbohydrate, it runs as a single band on a 1D gel, whereas 2D analysis clearly demonstrates that differentially charged forms are present. We cannot rule out the possibility that the multiple forms of B345 visualized on the 2D gels are an artifact of the technique. The exact sites of glycosylation are not known; however, predictions using standard rules map the potential sites to the surface of B345 that faces the exterior of the capsid. B345 and VP54, the major capsid protein of PBCV-1 (33), have strikingly similar tertiary structures (30), and the predicted sites of glycosylation on B345 overlay well with the identified positions on VP54. Interestingly, the sites of glycosylation on VP54 are noncanonical, and the PBCV-1 genome contains ORFs encoding glycosyltransferases. It has now been confirmed that STIV also encodes a glycosyltransferase (31). As reported here, deglycosy-

lation experiments failed to identify the nature of the carbohydrates on B345. Taken together, these results suggest that B345 of STIV may have a novel form of glycosylation. The purpose of encoding a glycosyltransferase in the viral genome is unclear. This may be due to protein/carbohydrate structure or may be a remnant of an ancestral pathway of glycosylation.

The detailed characterization of STIV particles involved the analysis of protein, lipid, and carbohydrate. Assembly of this data has revealed a picture of STIV akin to tail-less dsDNA viruses from the other domains of life. The evolutionary relationship of PBCV-1, PRD1, SH1, and STIV was not immediately apparent upon comparison of their genomes. Separation on an evolutionary time scale will lead to sequence divergence, but in the case of viruses this is exaggerated by necessary adaptations to unique viral lifestyles, unique environmental parameters (variable pH, temperatures, and salinity), variable host G/C ratios, and different mechanisms of transcriptional and translational regulation. It is therefore not surprising that the relationships between viral proteins from different domains of life can be reliably detected only at the level of structure (9, 30). While this makes the annotation of viruses much harder, it is deeply satisfying to uncover these shared architectural principles that have persisted both at the level of individual proteins and whole viral capsids over millions of years (5). The continued analysis of archaeal viruses, such as STIV and SH1, continues to strengthen the evidence for the existence of a common viral ancestor that predates the divergence of *Eukarya*, *Bacteria*, and *Archaea*.

#### ACKNOWLEDGMENTS

The Mass Spectrometry and Proteomic Laboratory at Montana State University is supported, in part, by a grant from the Murdock Foundation. B.B. is supported by Center for BioInspired Nanomaterials under ONR N00014-05-1-0386, and the Montana State University Thermal Biology Institute is funded in part by NASA Astrobiology Program NNG04GR46G and NIH grant number P20 RR-020185 from the COBRE program of the National Center for Research Resources. K.S. was supported in part by the NSF REU summer research program at MSU. M.D. is supported in part by NIH grant P20 RR16455-06 from the INBRE-BRIN program of the National Center for Research Resources. M.J.Y. is supported by NSF grant MCB 0132156.

We thank Rachel Neal for assistance with the mass spectrometry.

#### REFERENCES

- Abrescia, N. G., J. J. Cockburn, J. M. Grimes, G. C. Sutton, J. M. Diprose, S. J. Butcher, S. D. Fuller, C. San Martin, R. M. Burnett, D. I. Stuart, D. H. Bamford, and J. K. Bamford. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**:68–74.
- Afonso, C. L., E. R. Tulman, Z. Lu, L. Zsak, F. A. Osorio, C. Balinsky, G. F. Kutish, and D. L. Rock. 2002. The genome of swinepox virus. *J. Virol.* **76**:783–790.
- Afonso, C. L., E. R. Tulman, Z. Lu, L. Zsak, G. F. Kutish, and D. L. Rock. 2000. The genome of fowlpox virus. *J. Virol.* **74**:3815–3831.
- Bamford, D. H., J. J. Ravantti, G. Ronnholm, S. Laurinavicius, P. Kukkaro, M. Dyal-Smith, P. Somerharju, N. Kalkkinen, and J. K. Bamford. 2005. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J. Virol.* **79**:9097–9107.
- Bamford, D. H., J. M. Grimes, and D. I. Stuart. 2005. What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.* **15**:655–663.
- Bardy, S. L., J. Eichler, and K. F. Jarrell. 2003. Archaeal signal peptides—a comparative survey at the genome level. *Protein Sci.* **12**:1833–1843.
- Benson, S. D., J. K. Bamford, D. H. Bamford, and R. M. Burnett. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **17**:825–833.
- Bothner, B., X. F. Dong, L. Bibbs, J. E. Johnson, and G. Siuzdak. 1998. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J. Biol. Chem.* **273**:673–676.
- Burnett, R. M. 2006. More barrels from the viral tree of life. *Proc. Natl. Acad. Sci. USA* **103**:3–4.
- Cassetti, M. C., M. Merchlinsky, E. J. Wolfe, A. S. Weisberg, and B. Moss. 1998. DNA packaging mutant: repression of the vaccinia virus A32 gene results in noninfectious, DNA-deficient, spherical, enveloped particles. *J. Virol.* **72**:5769–5780.
- Ceulemans, H., and R. B. Russell. 2004. Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization. *J. Mol. Biol.* **338**:783–793.
- Cobucci-Ponzano, B., M. Rossi, and M. Moracci. 2005. Recoding in archaea. *Mol. Microbiol.* **55**:339–348.
- Cockburn, J. J., N. G. Abrescia, J. M. Grimes, G. C. Sutton, J. M. Diprose, J. M. Benevides, G. J. Thomas, Jr., J. K. Bamford, D. H. Bamford, and D. I. Stuart. 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* **432**:122–125.
- De Rosa, M., and A. Gambacorta. 1988. The lipids of archaeobacteria. *Prog. Lipid Res.* **27**:153–175.
- Deshpande, N., K. J. Address, W. F. Bluhm, J. C. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, Z. Feng, R. K. Green, J. L. Flippen-Anderson, J. Westbrook, H. M. Berman, and P. E. Bourne. 2005. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.* **33**:D233–D237.
- Desrosiers, D. C., and Z. Y. Peng. 2005. A binding free energy hot spot in the ankyrin repeat protein GABP beta mediated protein-protein interaction. *J. Mol. Biol.* **354**:375–384.
- Eichler, J., and M. W. Adams. 2005. Posttranslational protein modification in *Archaea*. *Microbiol. Mol. Biol. Rev.* **69**:393–425.
- Estrada, R., and M. C. Yappert. 2004. Alternative approaches for the detection of various phospholipid classes by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **39**:412–422.
- Fountoulakis, M., J. F. Juranville, and M. Manneberg. 1992. Comparison of the Coomassie brilliant blue, bicinchoninic acid and Lowry quantitation assays, using non-glycosylated and glycosylated proteins. *J. Biochem. Biophys. Methods* **24**:265–274.
- Garnier, N., S. Crouzy, and M. Genest. 2003. Molecular dynamics simulations of the transmembrane domain of the oncogenic ErbB2 receptor dimer in a DMPC bilayer. *J. Biomol. Struct. Dyn.* **21**:179–200.
- Gowen, B., J. K. Bamford, D. H. Bamford, and S. D. Fuller. 2003. The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. *J. Virol.* **77**:7863–7871.
- Guagliardi, A., L. Cerchia, and M. Rossi. 2002. The Sso7d protein of *Sulfolobus solfataricus*: in vitro relationship among different activities. *Archaea* **1**:87–93.
- Halligan, B. D., V. Ruotti, W. Jin, S. Laffoon, S. N. Twigger, and E. A. Dratz. 2004. ProMoST (protein modification screening tool): a web-based tool for mapping protein modifications on two-dimensional gels. *Nucleic Acids Res.* **32**:638–644.
- Hao, B., W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, and M. K. Chan. 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**:1462–1466.
- Hopmans, E. C., S. Schouten, R. D. Pancost, M. T. van der Meer, and J. S. Damste. 2000. Analysis of intact tetraether lipids in archaeal cell material and sediments by high performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**:585–589.
- Ibbal, M., and D. Söll. 2002. Genetic code: introducing pyrrolysine. *Curr. Biol.* **12**:R464–R466.
- Johnston, J. B., G. Wang, J. W. Barrett, S. H. Nazarian, K. Colwill, M. Moran, and G. McFadden. 2005. Myxoma virus M-T5 protects infected cells from the stress of cell cycle arrest through its interaction with host cell cullin-1. *J. Virol.* **79**:10750–10763.
- Karplus, K., S. Katzman, G. Shackleford, M. Koeva, J. Draper, B. Barnes, M. Soriano, and R. Hughey. 2005. SAM-T04: what is new in protein-structure prediction for CASP6. *Proteins* **61**(Suppl. 7):135–142.
- Kelley, L. A., R. M. MacCallum, and M. J. Sternberg. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**:499–520.
- Khayat, R., L. Tang, M. J. Young, and J. E. Johnson. 2005. Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl. Acad. Sci. USA* **102**:18944–18949.
- Larson, E., D. Reiter, M. J. Young, and M. Lawrence. 2006. Structure of A197 from *Sulfolobus* turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. *J. Virol.* **80**:7636–7644.
- Merkel, M. C., J. T. Huiskonen, D. H. Bamford, A. Goldman, and R. Tuma. 2005. The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol. Cell* **18**:161–170.
- Murae, T., Y. Takamatsu, R. Muraoka, S. Endoh, and N. Yamauchi. 2002. Facile distinction of neutral and acidic tetraether lipids in archaea membrane by halogen atom adduct ions in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **37**:209–215.
- Nandhagopal, N., A. A. Simpson, J. R. Gurnon, X. Yan, T. S. Baker, M. V. Graves, J. L. Van Etten, and M. G. Rossmann. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl. Acad. Sci. USA* **99**:14758–14763.

35. Nicolaus, B., A. Trincone, E. Esposito, M. R. Vaccaro, A. Gambacorta, and M. De Rosa. 1990. Calditol tetraether lipids of the archaeobacterium *Sulfolobus solfataricus*. Biosynthetic studies. *Biochem. J.* **266**:785–791.
36. 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 archaeobacterium *Sulfolobus shibatae*. *Virology* **185**:242–250.
37. Rice, G., K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott, and J. M. Young. 2001. Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. USA* **98**:13341–13345.
38. Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J. E. Johnson, T. Douglas, and J. 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. USA* **101**:7716–7720.
39. Sanchez, R., and A. Sali. 1998. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci. USA* **95**:13597–13602.
40. Schaffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, and S. F. Altschul. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**:2994–3005.
41. Skrdla, M. P., D. E. Burbank, Y. Xia, R. H. Meints, and J. L. Van Etten. 1984. Structural proteins and lipids in a virus, PBCV-1, which replicates in chlorella-like alga. *Virology* **135**:308–315.
42. Snyder, J. C., J. Spuhler, B. Wiedenheft, F. F. Roberto, T. Douglas, and M. J. Young. 2004. Effects of culturing on the population structure of a hyperthermophilic virus. *Microb. Ecol.* **48**:561–566.
43. Soding, J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**:951–960.
44. Tsai, J. M., H. C. Wang, J. H. Leu, H. H. Hsiao, A. H. Wang, G. H. Kou, and C. F. Lo. 2004. Genomic and proteomic analysis of thirty-nine structural proteins of shrimp white spot syndrome virus. *Virology* **78**:11360–11370.
45. Vance, D. E., and J. E. Vance. 2002. *Biochemistry of lipids, lipoproteins and membranes*, 4th ed., p. 233–260. Elsevier, New York, N.Y.
46. Van Etten, J. L., and R. H. Meints. 1999. Giant viruses infecting algae. *Annu. Rev. Microbiol.* **53**:447–494.
47. Wang, C., M. Eufemi, C. Turano, and A. Giartosio. 1996. Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35**:7299–7307.
48. Whitley, P., B. J. Reaves, M. Hashimoto, A. M. Riley, B. V. Potter, and G. D. Holman. 2003. Identification of mammalian Vps24p as an effector of phosphatidylinositol 3,5-bisphosphate-dependent endosome compartmentalization. *J. Biol. Chem.* **278**:38786–38795.
49. Wiedenheft, B., K. Stedman, F. Roberto, D. Willits, A. K. Gleske, L. Zoeller, J. Snyder, T. Douglas, and M. Young. 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. *J. Virol.* **78**:1954–1961.
50. Williams, C. C., J. G. Allison, G. A. Vidal, M. E. Burrow, B. S. Beckman, L. Marrero, and F. E. Jones. 2004. The ERBB4/HER4 receptor tyrosine kinase regulates gene expression by functioning as a STAT5A nuclear chaperone. *J. Cell Biol.* **167**:469–478.
51. Yan, Q., P. R. Hunt, L. Frelin, T. A. Vida, J. Pevsner, and A. J. Bean. 2005. mVps24p functions in EGF receptor sorting/trafficking from the early endosome. *Exp. Cell Res.* **304**:265–273.
52. Zillig, W., D. Prangishvili, 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.