

Constituents of SH1, a Novel Lipid-Containing Virus Infecting the Halophilic Euryarchaeon *Haloarcula hispanica*

Dennis H. Bamford,^{1,2} Janne J. Ravantti,¹ Gunilla Rönholm,² Simonas Laurinavičius,^{1,3}
Petra Kukkaro,¹ Mike Dyll-Smith,⁴ Pentti Somerharju,³ Nisse Kalkkinen,²
and Jaana K. H. Bamford^{1*}

Department of Biological and Environmental Sciences and Institute of Biotechnology, University of Helsinki, Viikki Biocenter, P.O. Box 56, FIN-00014 Helsinki, Finland¹; Institute of Biotechnology, University of Helsinki, Viikki Biocenter, P.O. Box 65, FIN-00014 Helsinki, Finland²; Department of Biochemistry, Institute of Biomedicine, P.O. Box 63, University of Helsinki, Biomedicum, FIN-00014 Helsinki, Finland³; and Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia⁴

Received 20 January 2005/Accepted 15 March 2005

Recent studies have indicated that a number of bacterial and eukaryotic viruses that share a common architectural principle are related, leading to the proposal of an early common ancestor. A prediction of this model would be the discovery of similar viruses that infect archaeal hosts. Our main interest lies in icosahedral double-stranded DNA (dsDNA) viruses with an internal membrane, and we now extend our studies to include viruses infecting archaeal hosts. While the number of sequenced archaeal viruses is increasing, very little sequence similarity has been detected between bacterial and eukaryotic viruses. In this investigation we rigorously show that SH1, an icosahedral dsDNA virus infecting *Haloarcula hispanica*, possesses lipid structural components that are selectively acquired from the host pool. We also determined the sequence of the 31-kb SH1 genome and positively identified genes for 11 structural proteins, with putative identification of three additional proteins. The SH1 genome is unique and, except for a few open reading frames, shows no detectable similarity to other published sequences, but the overall structure of the SH1 virion and its linear genome with inverted terminal repeats is reminiscent of lipid-containing dsDNA bacteriophages like PRD1.

Viruses infecting hosts in the domain *Archaea* represent diverse morphotypes (25, 57, 58). However, only some dozen genome sequences have been deposited into sequence databases to date (December 2004). Within the *Crenarchaeota*, viruses have been isolated and sequenced that infect members of the classes *Sulfolobales* and *Thermoproteales*. By morphology, the viruses of *Sulfolobales* represent spindle-shaped (*Sulfolobus* spindle-shaped viruses [72]), rod-shaped (*Sulfolobus islandicus* filamentous virus [3]; *S. islandicus* rod-shaped viruses 1 and 2 [53]; *Acidianus* filamentous virus 1 [15]), and icosahedral (*Sulfolobus* turreted icosahedral virus [STIV] [60]) forms. For *Thermoproteales* the only sequenced isolate is the *Pyrobaculum* spherical virus, which has been proposed to possess a lipid envelope surrounding a helical nucleoprotein core (30). The kingdom *Euryarchaeota* contains physiologically diverse members of methanogenic, halophilic, and thermoacidophilic organisms, and viruses of this group include the *Methanothermobacter* viruses psiM1 and psiM2 (47, 54), the lytic HF1 that infects cells of the *Haloferax* and *Haloterrigena* genera (68), the lytic HF2 that infects *Halorubrum caricense* (67), and the temperate *Natrialba magadii* virus ϕ Ch1 (40). All have head-tail morphology reminiscent of bacteriophages belonging to the families *Myoviridae* (such as T4) and *Siphoviridae* (such as λ).

There is a growing body of sequence information accumulating for these viruses. The tailed double-stranded DNA (dsDNA) bacteriophages, the viruses most commonly isolated from the environment, have strongly influenced the current view of viral genomics (18, 52). Such viruses probably infect most if not all bacterial and some archaeal species. The genomes of tailed phages are mosaics that have likely experienced global genetic exchange (44, 67, 68). On the other hand, tailless bacterial viruses with an internal membrane belonging to the family *Tectiviridae* do not seem to obey this rule. The family consists of six isolates that infect gram-negative bacteria (PRD1, PR3, PR4, PR5, L17, and PR772) and three that infect gram-positive bacteria (Bam35, AP50, and GIL01) (5; www.ncbi.nlm.nih.gov/ICTV). The genome sequences of isolates that infect gram-negative bacteria are highly similar, with overall identities between 91.9% and 99.8% (A.-M. Saren, J. J. Ravantti, S. D. Benson, R. M. Burnett, L. Paulin, D. H. Bamford, and J. K. H. Bamford, unpublished data). The two sequenced *Tectiviridae* isolates that infect gram-positive bacteria, Bam35 and GIL01 (59, 71), have only a few nucleotide differences.

Although the two groups of tectiviruses (infecting gram-negative or gram-positive bacteria) show no sequence similarity, their morphologies are nearly identical, and their genome organization and length are similar (59). Much is known of the structural organization of PRD1, with the atomic resolution structure determined down to 4 Å (1). The structure revealed a widely utilized assembly principle for the protein capsid and showed the icosahedrally ordered internal virus membrane (1,

* Corresponding author. Mailing address: Department of Biological and Environmental Sciences, P.O. Box 56 (Viikinkaari 5), University of Helsinki, FIN-00014 Helsinki, Finland. Phone: 358 9 191 59101. Fax: 358 9 191 59098. E-mail: jaana.bamford@helsinki.fi.

20). When the sequence of the Bam35 coat protein was threaded to the X-ray structure of the PRD1 coat protein, the result indicated that these viruses most probably share a related coat protein fold (14). Most interestingly, the archaeal virus STIV has been proposed to have a capsid architecture and coat protein fold similar to that of PRD1 and to possibly have a membrane (14, 60).

The detailed characterization of archaeal viruses is in its infancy because the *Archaea* are the most recently described domain of life, and many of the members live in extreme environmental conditions (16, 28). The sequencing of some of the archaeal virus genomes has yielded comparable results to those obtained for the tailed bacteriophages; e.g., the sequences of haloviruses HF1 and HF2 suggest the presence of mosaic genomes produced by genetic exchange (68). However, DNA sequences have not revealed much information about the structures or functions of individual viral proteins since matches to known sequences in the databases are scarce. The same observation has been made for sequences of viruses in the family *Tectiviridae* (8, 59, 71).

Our interests reside in icosahedral dsDNA viruses with an internal membrane. We have proposed that a number of these viruses that infect bacterial and eukaryotic host, such as PRD1, Bam35, *Paramecium bursaria* *Chlorella* virus-1 (PBCV-1), Chilo iridescent virus, and Mimivirus, are structurally related, suggesting a common ancestor that predates the separation of the three domains of life (6, 7, 14). Strong support for such a proposal would come from data that viruses with such architecture also infect archaeal hosts. The first such proposal has been made based on structural studies of STIV (31, 60). However, rigorous evidence of a lipid component in the STIV virion has yet to be provided.

To further explore the relationships between archaeal, bacterial, and eukaryotic viruses, we analyzed SH1, a recently identified virus isolate that infects the halophilic euryarchaeon *Haloarcula hispanica* (25, 55). We sequenced the viral genome, analyzed the lipid components, and identified genes for most of the structural proteins. We found SH1 to be an icosahedral dsDNA virus with morphology that suggests the presence of a membrane underneath the protein capsid, similar to PRD1 and related bacterial and eukaryotic viruses (14, 55). Thus, SH1 is a prime candidate to be included in the PRD1-type virus lineage.

MATERIALS AND METHODS

Growth and purification of SH1. SH1 virus was grown on *H. hispanica* (ATCC 33960) (34) according to the method of Dyall-Smith (24). The virus purification has previously been described (55). In short, the produced virus was concentrated from the virus stock (5×10^{10} to 1×10^{11} PFU/ml) with 10% (wt/vol) polyethylene glycol (PEG 8000) (Sorvall SLA3000 rotor; 10,000 rpm for 30 min at 4°C). The virus was then resuspended in purification buffer used throughout the procedure (40 mM Tris-HCl, pH 7.2, 40 mM MgCl₂, and 1 M NaCl), followed by purification in a linear 5% to 20% (wt/vol) sucrose gradient (Sorvall AH626 rotor; 22,000 rpm for 55 min at 20°C) and an equilibrium CsCl gradient (Sorvall AH627 rotor; 24,000 rpm for 22 h at 20°C). The equilibrated virus band was collected and diluted 1:2 in 40 mM Tris-HCl (pH 7.2)–40 mM MgCl₂ and collected by differential centrifugation (Sorvall T647.5 rotor; 32,000 rpm for 3 h at 20°C). After the pellet was resuspended in purification buffer, the protein concentration was measured by the Coomassie blue method of Bradford (17) using bovine serum albumin as a standard. The number of infective particles was then determined. The specific infectivity was $\sim 2 \times 10^{12}$ PFU/mg of protein.

Extraction and analysis of the lipids. (i) **Lipid extraction.** Lipids from cell and virus suspensions were extracted according to the method of Folch et al. (27), as modified for halophiles (37), and stored in chloroform-methanol (9:1, vol/vol) at -20°C until analyzed.

(ii) **TLC.** Lipids of total extracts were separated by thin-layer chromatography (TLC) on silica gel 60 plates (Merck) using chloroform-methanol-90% acetic acid (65:4:35, vol/vol/vol) as the solvent (21). The lipids were visualized by iodine vapor. Molybdate (23) and 0.1% orcinol-sulfuric acid reagent were used to detect phospholipids and glycolipids, respectively. The phosphorus content of lipid spots was determined as described by Kahma et al. (35). Lipid-containing areas were scraped from a preparative TLC and reextracted for mass spectrometry (MS) analysis. Semiquantitative TLC was performed as described previously (42).

(iii) **MS analyses.** Samples were dissolved in chloroform-methanol (1:2, vol/vol) plus 1% ammonia and infused (7 $\mu\text{l}/\text{min}$) into a Micromass Quattro Micro mass spectrometer (Micromass, Manchester, United Kingdom) operated in the negative ion mode. Nitrogen was used as a desolvation (flow, 500 liters/h) and cone (flow, 50 liters/h) gas. The potentials of the capillary, cone, extractor, and RF lens were -3800 V, -40 V, -2 , and -0.5 , respectively. The spectra were acquired over the m/z range from 200 to 1,999 (5 s/scan), and 30 to 50 scans were averaged for each spectrum. Lipids were identified based on (i) their m/z value, (ii) product ion analysis, and (iii) precursor ion scans. Thus, for example, the presence of phosphate, sulfate, and glycerophosphate moieties was detected by scanning for precursors of m/z 79, 97, and 153, respectively.

Genome sequencing. Genomic DNA was isolated from the purified virus particles by protease K and sodium dodecyl sulfate (SDS) treatment. For this, the virus preparation was diluted 1:5 with water to lower the salt concentration of the buffer. Protease K was added to the final concentration of 150 $\mu\text{g}/\text{ml}$, and the mixture was incubated for 1 h at 37°C . SDS was added (final concentration, 2%), and the incubation was continued for an additional 40 min, after which the mixture was treated with multiple phenol-chloroform and ether extractions. DNA was recovered from the aqueous phase by ethanol precipitation and used for cloning or as template for sequencing.

The DNA was digested either with AccI, RsaI, or HincII restriction enzymes, treated with Klenow polymerase to create blunt ends, and the fragments were ligated into the SmaI-digested pSU18 vector (10). The resulting clones were sequenced, and the sequences were further used to design specific oligonucleotides for primer walking using the virus genome as a template. Sequencing reactions were performed by the dideoxy chain termination method using a DYEnamic ET terminator Cycle sequencing Kit with ThermoSequenase II DNA polymerase (Amersham Biosciences) on an ABI Prism 377 automated sequencer (Applied Biosystems) at the sequencing service facility of the Institute of Biotechnology, University of Helsinki. The genomic sequence was assembled using the assembly program of the Vector NTI^R Suite8 software package (Informax Inc.). Difficult regions, which did not yield high-quality sequence with direct genomic walking, were amplified by PCR, and the sequence was determined from the fragments using specific primers. As a result, each nucleotide was determined more than twice except for the region 3050 to 4690, where only one strand could be determined (several times, however).

Protein analysis. (i) **Protein identification.** The protein composition of the virion was analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) using either 8% or 17% polyacrylamide gels (51) or 14% polyacrylamide-tricine-SDS gels (63).

(ii) **N-terminal sequence analysis.** SDS-PAGE-separated proteins were transferred by electroblotting (46) onto a polyvinylidene difluoride membrane (ProBlott; Perkin Elmer Applied Biosystems Division, Foster City, Calif.). After being stained with Coomassie brilliant blue, the protein bands of interest were removed and subjected to N-terminal sequence analysis in a Procise 494A protein sequencer (Perkin Elmer Applied Biosystems Division, Foster City, CA).

(iii) **Protein digestion and mass spectrometric analysis of fragments.** SDS-PAGE-separated proteins were stained with Coomassie brilliant blue, and protein bands of interest were cut from the gel. The proteins were “in-gel” digested essentially as described by Shevchenko et al. (64). Proteins were reduced with dithiothreitol and alkylated with iodoacetamide before digestion with trypsin (Sequencing Grade Modified Trypsin, V5111; Promega). Peptides generated by enzymatic cleavage were analyzed by matrix-assisted laser desorption-ionization time of flight (MALDI-TOF) MS for mass fingerprinting and electrospray tandem mass spectrometry for determining partial amino acid sequences. MALDI-TOF mass spectra of peptide fragments were acquired using an Ultraflex TOF/TOF instrument (Bruker-Daltonik GmbH, Bremen, Germany) and electrospray ionization quadrupole time of flight tandem mass spectra for de novo sequencing using a Q-TOF instrument (Micromass, Manchester, United Kingdom) as described previously (56). The obtained peptide masses and partial sequences were

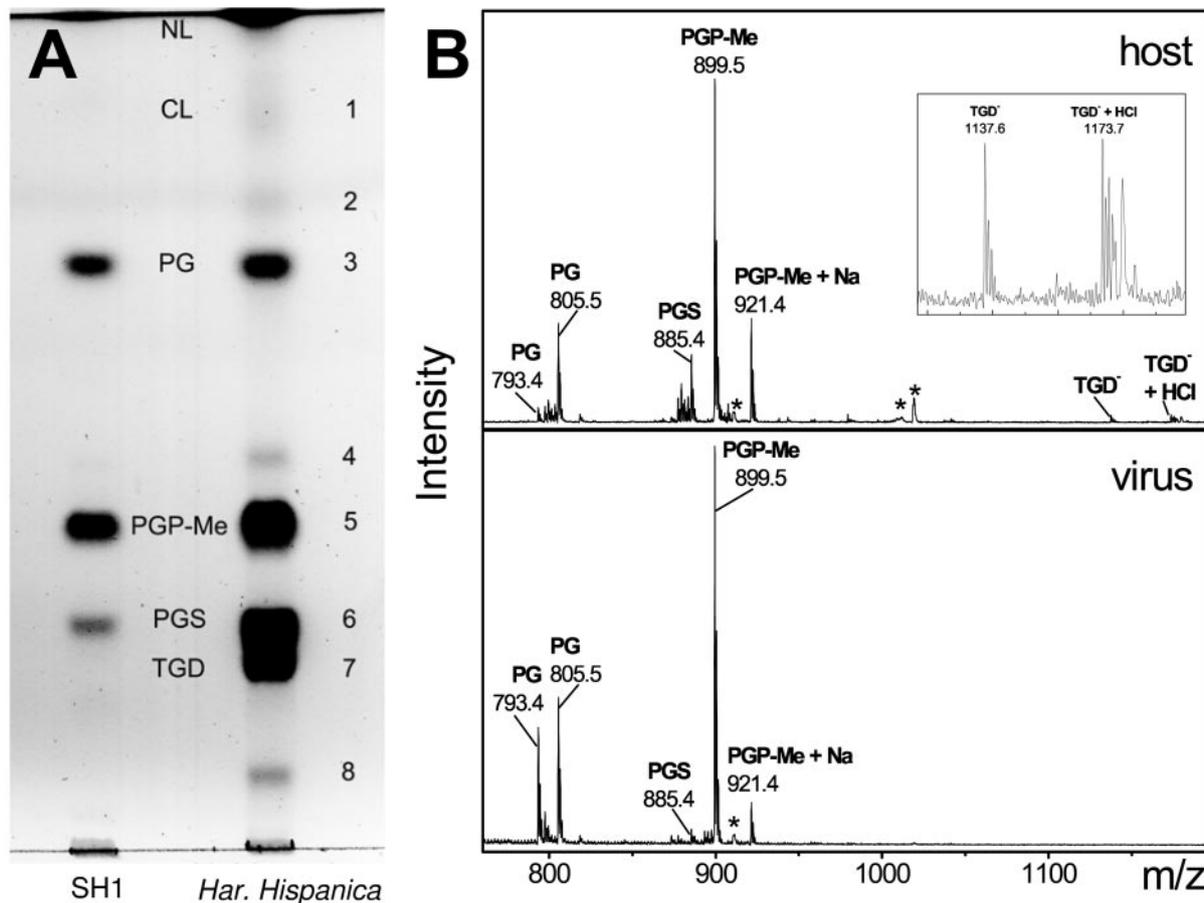


FIG. 1. (A) Thin-layer chromatogram of the lipids extracted from *H. hispanica* and SH1. Lipids were identified on the basis of their relative retention times (which were very similar to previously reported ones [21]), staining behavior, and their molecular mass as determined by MS. Some lipids present in the host lipid extract could not be identified. Due to the limited amount of virus available, the amount of total lipids on the SH1 lane is somewhat less than the amount on the host lane. (B) Negative ion mass spectrum of the lipids extracted from *H. hispanica* (upper panel) and SH1 (lower panel). Inset in the upper panel is a magnification of the 1,120 to 1,200 m/z region from the spectrum of the host lipid extract showing TGD and its adduct. Peaks representing doubly charged noncovalent lipid heterodimers are marked with an asterisk (*). For example, m/z 910.9 is a doubly charged dimer of PGP-Me (m/z 899.5) and PGP-Me plus Na (m/z 921.4), whereas m/z 1019.2 is a doubly charged dimer of PGP-Me (m/z 899.5) and TGD (m/z 1137.6). NL, neutral lipids; CL, cardiolipin.

then compared with the sequences of the structural proteins translated from the determined gene nucleotide sequences.

Computational methods. Programs from the EMBOSS software package (61) were used with locally made shell scripts to calculate several of the results. The GC percentage and length were calculated using the program Infseq. Inverted DNA repeats were identified using the Einverted program. The predicted genes were translated from open reading frames (ORFs) to proteins with Transeq. Statistics for each ORF were calculated with the program Pepstats and Freak. The secondary structure predictions were made using the mfold program (73). Several web-based services were used to further characterize the SH1 genome. The putative ORFs were found using GeneMark.hmm and GeneMark programs (45; <http://opal.biology.gatech.edu/GeneMark/>). Several different Markov models were used to ensure that all reasonable ORFs were found. Possible homologies to known proteins were searched with PSI-BLAST (2; <http://www.ncbi.nlm.nih.gov/BLAST/>). PSI-BLAST was executed for each ORF until no new hits were reported. Prediction of transmembrane helices was performed using the TMHMM program (41; <http://www.cbs.dtu.dk/services/TMHMM/>). Protein fold recognition was performed using the 3D-PSSM method (39; <http://www.sbg.bio.ic.ac.uk/~3dpssm/>).

Nucleotide sequence accession number. Nucleotide sequence data for SH1 have been deposited in the GenBank database under accession number AY950802.

RESULTS

SH1 is a lipid-containing virus. The total lipid extracts of *H. hispanica* and highly purified SH1 were examined by thin-layer chromatography (Fig. 1A). The individual lipid components were identified based on their relative positions and specific staining of the lipid spot on the TLC plate, the mass of the intact lipid, and its fragmentation products as determined by electrospray ionization mass spectrometry (Fig. 1B).

TLC analysis showed that the lipid composition of *H. hispanica* was quite complex, containing four major and some minor components (Fig. 1A, spots designated 1 to 8). The R_f of spot 1 and a minor, doubly charged peak ($m/z = 760.2$) in the negative ion mass spectrum (not shown) suggested the presence of cardiolipin. Spots 3, 5, and 6 most probably represent archaeal phosphatidylglycerol (PG), phosphatidylglycerophosphate methyl ester (PGP-Me), and phosphatidylglycerosulfate (PGS), respectively, as indicated by their relative R_f values (21)

TABLE 1. Phospholipid class composition of SH1 and its host *H. hispanica*^a

Lipid	Phospholipid composition (mol%)	
	SH1	<i>H. hispanica</i>
PG	16.5 ± 1.9	13.6 ± 2.8
PGP-Me	81.7 ± 2.0	56.9 ± 4.1
PGS	1.4 ± 0.7	24.7 ± 2.0
Other ^b	~0.5	~4

^a Compositions are expressed as molar percentages of total phospholipids. Values represent the means ($n = 6$) ± standard deviation.

^b In the case of *H. hispanica*, this includes cardiolipin and spots 2, 4, and 8, while in the case of SH1, only spot 4 contained traces of phosphorus.

and specific staining (data not shown). The masses of these lipids (805.4, 899.4, and 885.4, respectively) (Fig. 1B), as well as precursor and product ion fragmentation analyses (see Materials and Methods), strongly support this identification. These lipids are common in halophiles (37), and their masses are identical to those of the saturated C₂₀/C₂₀ molecular species that are typical for *Haloarcula* species (36). However, small amounts of lower-mass species of PG (m/z range, 793.5 to 805.5), PGS (m/z range, 873.4 to 885.4), and PGP-Me (m/z range, 887.4 to 899.5) were also detected, indicating that molecular species with unsaturated and/or cyclopentyl-containing alkyl chains could also be present. Spot 7 was carbohydrate stain positive, suggesting the presence of a glycolipid in the host. The mass spectrum of the material extracted from this spot showed two peaks with m/z of 1137.65 and 1173.65 (Fig. 1B). These masses (and their isotopic patterns) correspond to those predicted for triglycosyl glycerodiether (TGD) ([M-H]⁻ = 1137.82), found in *Haloarcula* species (26, 38), and its hydrogen chloride adduct ([M-H]⁻ + HCl = 1173.82). The minor lipid components (i.e., spots 2, 4, and 8) could not be positively identified.

Figure 1A and B indicate that the lipid composition of SH1 is simpler than that of the host; i.e., it contained only three major lipids (PG, PGP-Me, and PGS). Phosphate analysis revealed that SH1 contained a higher proportion of PGP-Me and a lower proportion of PGS compared to the host. The proportion of PG was similar in the host and the virus (Table 1). SH1 did not contain detectable amounts of glycolipids, which constituted approximately 20% of total host lipids, as estimated from semiquantitative TLC analysis. The TLC data also indicated that neutral lipids represented only about 10% to 15% in SH1, while they comprised 20% to 30% of the total lipids of the host.

Genome sequence. The double-stranded linear genomic DNA (55) was isolated from highly purified virus particles, and its sequence was determined. The linear genome of SH1 is 30,898 bp long and contains 309-bp inverted terminal repeats (ITRs). The genome has an overall G+C content of 68.4%, which is comparable to the corresponding values in halophilic archaeal organisms, e.g., the sequenced regions of the host of SH1 *H. hispanica* (63.5%), the entire genome of *Halobacterium* sp. NRC-1 (66.9% [50]), and the largest chromosome (62%) of the recently sequenced genome of *Haloarcula marismortui* (4). Since the linear SH1 genome has long ITRs, it was computationally folded to determine whether specific hairpin structures occur in the single-stranded form, as in some linear virus

genomes (A.-M. Saren, J. J. Ravantti, S. D. Benson, R. M. Burnett, L. Paulin, D. H. Bamford, and J. K. H. Bamford, unpublished data). The minimum folding energy of SH1 single-stranded DNA (-4884.0 kcal/mol) did not differ from five randomly generated sequences with statistics (ITRs, genome length, and G+C content) similar to the SH1 genome, suggesting the absence of such structures in the SH1 genome.

The potential ORFs were assigned using the Gene Mark.hmm 2.1 program (<http://opal.biology.gatech.edu/GeneMark/>) and the *Halobacterium* sp. NRC-1 codon usage table (<http://www.kazusa.or.jp/codon/>). The analysis yielded 48 potential ORFs corresponding to proteins longer than 40 amino acid residues. We also analyzed the regions upstream of the potential ORFs to determine putative (bacterial type) ribosome binding sequences, and additional potential ORFs were found. By combining these observations, we concluded that the SH1 genome might contain 56 ORFs (Table 2). This number agrees well with the number of annotated ORF frequencies in other sequenced archaeal virus genomes (1.7 ORFs/kb on average for all previously known archaeal virus genomes versus 1.8 ORFs/kb in SH1). Interestingly, using these criteria, two of the ORFs (ORFs 1 and 56) are assigned to the ITR regions. It will be intriguing to see whether they are functional. SH1 ORFs begin with either ATG (35) or GTG (20). Many of the ORFs are translationally coupled by overlapping start and stop codons (TAATG or TAGTG). Based on the transcriptional direction of the ORF clusters in the SH1 genome, there appears to a minimum of four operons (Fig. 2). BLAST searches provided only a few hits, the most significant being an ORF17 hit to ATPases (for details, see the description of structural proteins below). ORF48 has high similarity (52%) to an unassigned ORF71 of halovirus ϕ Ch1 (accession number AAM88745).

The putative ribosome binding sequence (RBS) and its distance from the translation initiation codon (ATG or GTG) in each ORF is presented in Table 2. Analogous to prokaryotic viral genes, a sequence complementary to the 3' end of the host *H. hispanica* 16S rRNA was found in most of the SH1 ORFs. However, seven of the ORFs did not have obvious RBSs. This is not surprising since archaeal transcription and translation machineries are mosaics of eukaryotic and prokaryotic ones (11). Typical bacterial features of archaeal transcripts are that they are polycistronic and the genes usually have upstream RBSs (22). However, the first gene of the transcript (or an individually transcribed gene) often lacks such signals, suggesting a different initiation mechanism (62, 69). If both mechanisms are used in SH1, there could be six to eight operons (Table 2).

Most of the predicted SH1 gene products had a low calculated isoelectric point ($pI < 5$) (Fig. 2), similar to haloarchaeal organisms (4, 67). However, based on the calculated pI , some of the SH1 gene products were extremely basic in character ($pI > 10$). One of the highly basic gene products, VP12 encoded by ORF23 (see below), was confirmed to be a protein component of the virion, suggesting that the acidic character of a predicted protein cannot alone be used as a criterion to determine the putative gene products of halophilic viruses.

SH1 virion is composed of ~15 structural protein species. To identify the genes encoding the structural proteins of the virion, the highly purified virus material was subjected to SDS-

TABLE 2. Predicted SH1 ORFs and their characteristics

ORF	Position ^a		Direction ^b	No. of residues	Molecular mass (kDa)	Calculated isoelectric point	TM ^c	GC%	Functional assignments	RBS sequence ^d	RBS distance ^e
	Start	Stop									
ORF 1	164	421	+	85	8.89	11.62		64.73			
ORF 2	418	573	+	51	5.68	12.57		73.72		AGG	-9
ORF 3	570	848	+	92	10.00	12.04		64.87			
ORF 4	849	1106	+	85	9.42	3.74		69.38		AGGAGG	-8
ORF 5	1103	1330	+	75	8.52	3.49		67.54		GGTG	-5
ORF 6	1323	1733	+	136	14.25	4.97	2	69.59		GGTG	-6
ORF 7	1730	1870	+	46	4.77	4.52		71.63		(GGGG)	-7
ORF 8	1863	2177	+	104	11.67	5.13		68.57		GGAGGT	-8
ORF 9	2170	2430	+	86	9.32	3.94		68.97		GAGGTG	-5
ORF 10	2427	2585	+	52	5.35	12.12		66.04		GGAGG	-7
ORF 11	2582	2920	+	112	12.63	4.43		68.14		GGAGG	-5
ORF 12	3047	3208	+	53	5.71	11.74		61.11			
ORF 13	3321	7493	+	1,390	152.33	3.96		68.58	VP1		
ORF 14	7498	7770	+	90	9.50	4.46	2	70.33		AGGAGGTG	-7
ORF 15	7772	8125	+	117	12.27	4.09	3	67.51		GGAGG	-8
ORF 16	8118	8348	+	76	8.48	3.55		67.97		AGGAGG	-9
ORF 17	8400	9122	-	240	26.89	5.37		64.45	ATPase?	AGGAGG	-5
ORF 18	9123	9326	-	67	6.93	4.13		67.16		AGGAGG	-3
ORF 19	9319	9750	-	143	15.53	3.91		67.13		GGT	-7
ORF 20	9747	10397	-	216	23.07	3.16		67.28		GGTGA	-9
ORF 21	10397	10804	-	135	14.37	3.69	1	67.65		GGAGG	-9
ORF 22	10809	11480	-	223	24.18	3.95		66.07		AGGA	-4
ORF 23	11545	11829	+	94	9.83	11.03	2	61.05	VP12	GAGG	-15
ORF 24	11845	12402	+	185	20.01	4.15		64.7	VP7	GGAGGT	-8
ORF 25	12404	13102	+	232	25.70	3.88		65.67	VP4	(GGGG)	-8
ORF 26	13118	13339	+	73	7.88	3.64		63.96		GGAGGT	-8
ORF 27	13343	13585	+	80	8.81	4.88		65.43	VP13	GGA	-14
ORF 28	13589	16063	+	824	81.06	3.64		71.84	VP2	GGAG	-8
ORF 29	16060	16971	+	303	31.84	4.41		69.74	VP5	GGAGGTG	-7
ORF 30	16974	17459	+	161	16.88	4.31		68.52	VP10	GGAGGT	-8
ORF 31	17459	17908	+	149	16.54	3.97		62.67	VP9	(GGGG)	-10
ORF 32	17921	18934	+	337	37.50	4.02		63.51	VP3		
ORF 33	18941	19633	+	230	24.97	4.09		65.95	VP6	GGAG	-9
ORF 34	19636	19950	+	104	11.40	5.63	1	67.62		GTGA	-2
ORF 35	20047	20232	-	61	6.69	8.5		71.51		GGAGG	-9
ORF 36	20229	20408	-	59	6.20	3.9		70		GGAGG	-7
ORF 37	20405	20521	-	38	4.23	7.36		73.5		GAGGTG	-5
ORF 38	20518	20664	-	48	5.37	7.81		65.99		GGT	-5
ORF 39	20661	20930	-	89	9.73	4.51		71.85		GGAGGT	-6
ORF 40	20927	21088	-	53	6.39	12.17		66.67		GGAGG	-5
ORF 41	21085	21576	-	163	19.15	5.32		67.07		GGAGG	-5
ORF 42	21573	22772	-	399	44.38	4.93		69.25		GGAGG	-8
ORF 43	22776	23135	-	119	13.12	3.83		69.17		GAGG	-3
ORF 44	23119	24357	-	412	46.30	4.04		67.88		GGAGGTG	-7
ORF 45	24354	24530	-	58	6.20	8.81		66.1			
ORF 46	24533	24790	-	85	10.03	4.12		64.34		GGT	-5
ORF 47	24794	24970	-	58	6.55	8.67		68.36		GAGGTGA	-6
ORF 48	24967	25407	-	146	16.34	4.64		70.52		GGTGA	-6
ORF 49	25404	25940	-	178	19.79	4.5		70.95		AGGTG	-5
ORF 50	25937	26602	-	221	24.38	4.5		67.72		GGT	-4
ORF 51	26708	27085	-	125	13.43	4.23		72.49		GGAGG	-7
ORF 52	27082	27330	-	82	9.04	4.02		67.87		GGAGG	-10
ORF 53	27327	27677	-	116	12.90	3.84		71.23		GAGG	-12
ORF 54	27770	27883	-	37	3.97	7.12		71.05		GAGG	-7
ORF 55	27880	30477	-	865	94.89	4.02		72.4			
ORF 56	30514	30726	-	70	7.67	13.06		68.54			

^a Numbers refer to the SH1 genome sequence (accession number AY950802).

^b The ORF is transcribed from left to right (+) or from right to left (-) in the linear genome (Fig 2.).

^c Putative transmembrane helices predicted using the program TMHMM (41; <http://www.cbs.dtu.dk/services/TMHMM/>).

^d Putative 5' upstream ribosome binding sequence (RBS) complementary to the 3' end of the *H. hispanica* 16S rRNA (exact match AGGAGGTGA. . .).

^e The distance of the RBS from the initiation codon calculated from the last nucleotide of the determined RBS motif.

PAGE. The resulting 15 separable Coomassie brilliant blue-stained bands were subjected to identification by protein chemical methods. The proteins, corresponding to apparent molecular masses from 185 kDa to 4 kDa (Fig. 3), were as-

signed as VP1 to VP15 (where VP is virion protein). For gene start identification, the proteins were subjected to N-terminal sequence analysis, with the results shown in Table 3 (additional data may be accessed at www.DBLAB.HELSINKI.FI/SUP-

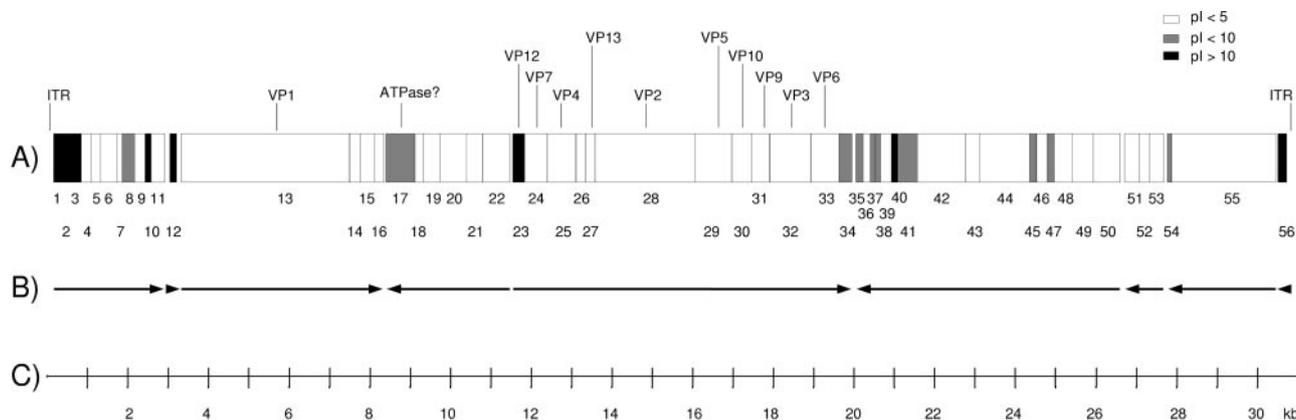


FIG. 2. Genome organization of SH1. (A) Predicted ORFs are numbered 1 to 56 from left to right and shaded according to the calculated isoelectric point of the predicted gene products. Structural virion components (VP1 to VP14) determined by protein chemistry methods are marked. (B) Direction of transcription is depicted by arrows. (C) Scale bar.

PLEMENTS). Eight N-terminal amino acid sequences were identified, all of which correlated with the translated gene products from the genome sequence. In all the determined protein N termini, the initiating methionine had been removed. The seven protein bands yielding no detectable signals in N-terminal sequencing (VP5, VP6, VP8, VP11, VP13, VP14, and VP15) were all minor bands in the Coomassie-stained SDS-polyacrylamide gel (Fig. 3) and were hardly visible on the polyvinylidene difluoride membrane after electroblotting and staining. Repeated attempts to obtain N-terminal sequences from these protein bands failed, even though our sequencer is routinely used for sequencing proteins in the 2 to 5 pmol range.

All proteins (VP1 to VP15) were further subjected to mass spectrometric identification by peptide mass fingerprint analysis. For this the protein bands were digested in gel with trypsin, and the resulting peptides were analyzed. Peptide mass fingerprint analysis by MALDI-TOF MS yielded peptide masses that were compared with the theoretical tryptic peptide masses from the deduced virion proteins. The number of peptide masses from the mass fingerprint corresponding to the masses of theoretical tryptic peptides (mass error within 100 ppm) for each of the virion proteins is shown in Table 3. Analysis of tryptic peptides by liquid chromatography-electrospray tandem mass spectrometry yielded partial sequences from a number of virion proteins. These data (Table 3) were further used to confirm the genes and to determine nucleotide sequences. The results from mass fingerprint analyses and determined partial amino acid sequences are provided elsewhere (www.DB.LAB.HELSINKI.FI/SUPPLEMENTS). Protein sequencing with Edman degradation and MALDI-TOF mass fingerprint analysis together with tandem mass spectrometry confirmed the identification of genes for the virion proteins VP1 to VP7, VP9, VP10, VP12, and VP13. No sequence or peptide mass information could be obtained from the suggested viral protein bands VP8, VP11, VP14, and VP15.

Most of the genes for the structural proteins are located in the middle of the genome. Of the 15 potential structural proteins, 11 were confirmed using protein chemistry methods. Comparison of the amino acid sequences to the DNA sequences revealed that the proteins are coded by the following ORFs: ORF13 (VP1), ORF23 (VP12), ORF24 (VP7), ORF25

(VP4), ORF27 (VP13), ORF28 (VP2), ORF29 (VP5), ORF30 (VP10), ORF31 (VP9), ORF32 (VP3), and ORF33 (VP6). All genes encoding structural proteins, except ORF13, are located in the middle of the genome and are most probably transcribed from a single operon from left to right in the genome map (Fig. 2). The most abundant virion proteins are VP3, VP4, VP7, and VP12 (Fig. 3). The calculated masses of VP3 (37.5 kDa), VP4 (25.7 kDa), and VP7 (20.0 kDa) suggest that they might represent the coat-associated proteins of the SH1 virion (see also protein complexes below). VP12 (9.8 kDa) is highly abundant and has two membrane-spanning helices according to the TMHMM prediction (Table 2). These characteristics suggest a role for VP12 as a major structural membrane protein of SH1.

The calculated molecular mass of VP2 is 81.1 kDa. VP2 is rich in glycine (15%) and serine (7%) residues. There are two linker-type regions in the protein: 11 glycine residues at positions 114 to 124 and a GS-rich repeat at positions 282 to 302 that separate the N-terminal, middle, and C-terminal domains. A stretch of 19 amino acid residues is repeated in the N-terminal domain (Fig. 4). The C-terminal domain contains a heptameric sequence DXAARGY (where X is D/E and Y is A/S/T) that is repeated five times (amino acids 640 to 674). If more variation was allowed in the sequence alignments, the heptad repeat could cover 178 residues starting at residue 490 and ending at residue 688 (Fig. 4). The heptapeptide repeat pattern is characteristic for alpha-helical coiled-coil proteins, and secondary structure prediction proposes high helicity (data not shown). These characteristics suggest that VP2 could be an elongated fiber-like protein suitable for forming an extended spike. Similar to bacteriophage PRD1 spike protein P5, VP2 has a domain structure. In the PRD1 spike, a collagen-like repeated region and eight glycine residues function as linkers to separate the protein domains and to make the spike flexible (9, 19, 32).

No peptide sequence or mass spectrometric data could be obtained for resolving the genes of the minor virion proteins VP8, VP11, VP14, and VP15. The database search with ORF17 (coding capacity of 26.9 kDa) gave hits to ATPases including PRD1 DNA packaging ATPase. In PRD1 the packaging ATPase is a structural component of the virion, unlike the packaging terminases of tailed phages (8, 48, 65). When

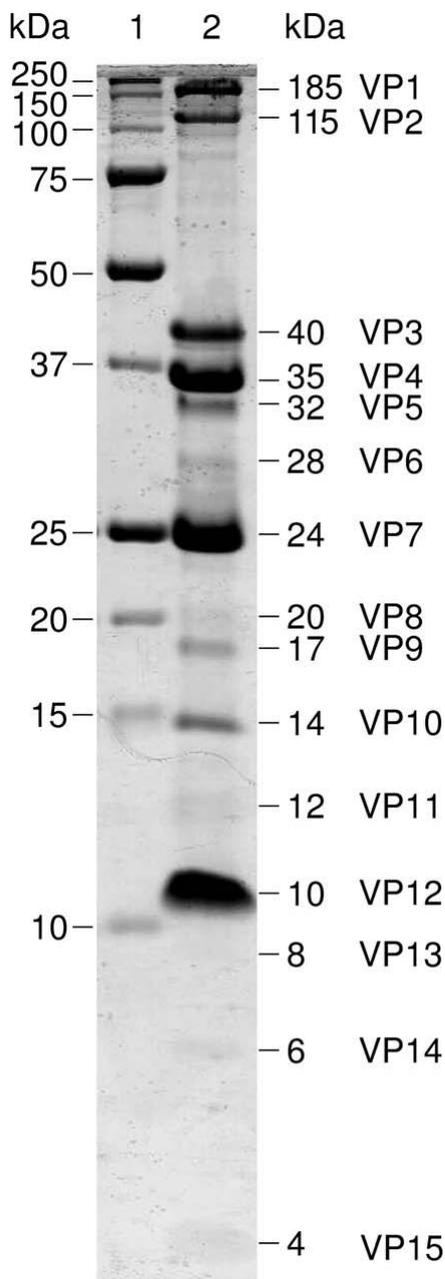


FIG. 3. SDS-PAGE (17% acrylamide) of the purified SH1 depicting the structural proteins (lane 2). Molecular mass marker are shown in lane 1.

dsDNA virus ATPase sequences were aligned, the predicted gene product of SH1 ORF17 appeared to have the same conserved motifs as the packaging ATPases of PRD1-type bacterial viruses. These motifs include the classical Walker A and B motifs as well as a third conserved motif common to membrane-containing dsDNA viruses (such as PBCV-1 and vaccinia virus) but not found in other dsDNA viruses (such as tailed phages or herpesvirus [66]). SH1 has a linear dsDNA genome and an internal membrane reminiscent of PRD1. The packaging system might also resemble that of PRD1, with the

TABLE 3. Determined biochemical data of the SH1 virion proteins

Protein	N-terminal sequence ^a	No. of masses ^b	No. of peptides/ no. of residues ^c
VP1	ADSTNTDMPL...	17	9/69
VP2	GFFSDLKG...	15	7/72
VP3	TTIGPKTDNL...	2	2/17
VP4	ADQTQEYTISH...	6	6/64
VP5	No result	5	No result
VP6	No result	No result	1/12
VP7	GNIGNLSAEK...	6	6/56
VP8	No result	No result	No result
VP9	VPGLXDNE...	5	No result
VP10	GKFALAS...	4	7/71
VP11	No result	No result	No result
VP12	ASINVSR...	No result	1/10
VP13	No result	1	No result
VP14	No result	No result	No result
VP15	No result	No result	No result

^a N-terminal protein sequences determined by Edman degradation.
^b Number of tryptic peptide masses corresponding (100 ppm mass tolerance) to the tryptic peptide masses calculated from the translated gene sequences.
^c Number of peptides analyzed by electrospray tandem mass spectrometry and residue coverage of peptide sequences determined.

SH1 packaging ATPase a structural component of the virion, possibly VP8 (apparent molecular mass, 20 kDa).

Virion proteins VP1, VP4, and VP7 form stable complexes. Electrophoresis of a highly purified virion sample in nonreducing conditions revealed additional major bands representing protein complexes (55). The information on SH1 gene sequences and structural proteins that we obtained was used to analyze these complexes. Figure 5 depicts a nonreducing low-percentage polyacrylamide (8%) gel displaying proteins and protein complexes from 50 kDa to several hundred kDa. The identity of indicated bands was confirmed by peptide mass fingerprinting with a significant number of peptide hits. Monomeric proteins VP1 (calculated molecular mass, 152.3 kDa) and VP2 (calculated molecular mass, 81.1 kDa) were identified. In addition, five major bands designated C1 to C5 (where C is complex) were observed. C1 and C2 yielded tryptic peptide masses only corresponding to VP1. Although the apparent complex masses from SDS-PAGE are somewhat lower than those calculated from the amino acid sequence for a VP1 trimer and dimer, respectively, they most probably represent such homomultimers. Band C3 (apparent mass, 74 kDa) yielded tryptic peptide masses from VP4 only. Bands C4 and C5 yielded tryptic peptide masses from both VP4 and VP7. It should be noted, however, that the accuracy of determining molecular masses for multimers using electrophoresis analysis is relatively low, especially under nonreducing conditions.

DISCUSSION

Here we directly demonstrated the presence of lipids in the highly purified SH1 virion. Several phospholipids, at least one glycolipid (TGD), and ~25% neutral lipids were present in *H. hispanica*, whereas SH1 contained only three major phospholipids (PG, PGP-Me, and PGS) and ~12% neutral lipids. In addition to qualitative differences, there were also quantitative ones; for example, the level of PGP-Me was higher and that of PGS was lower in the virus than in the host (Table 1). Such

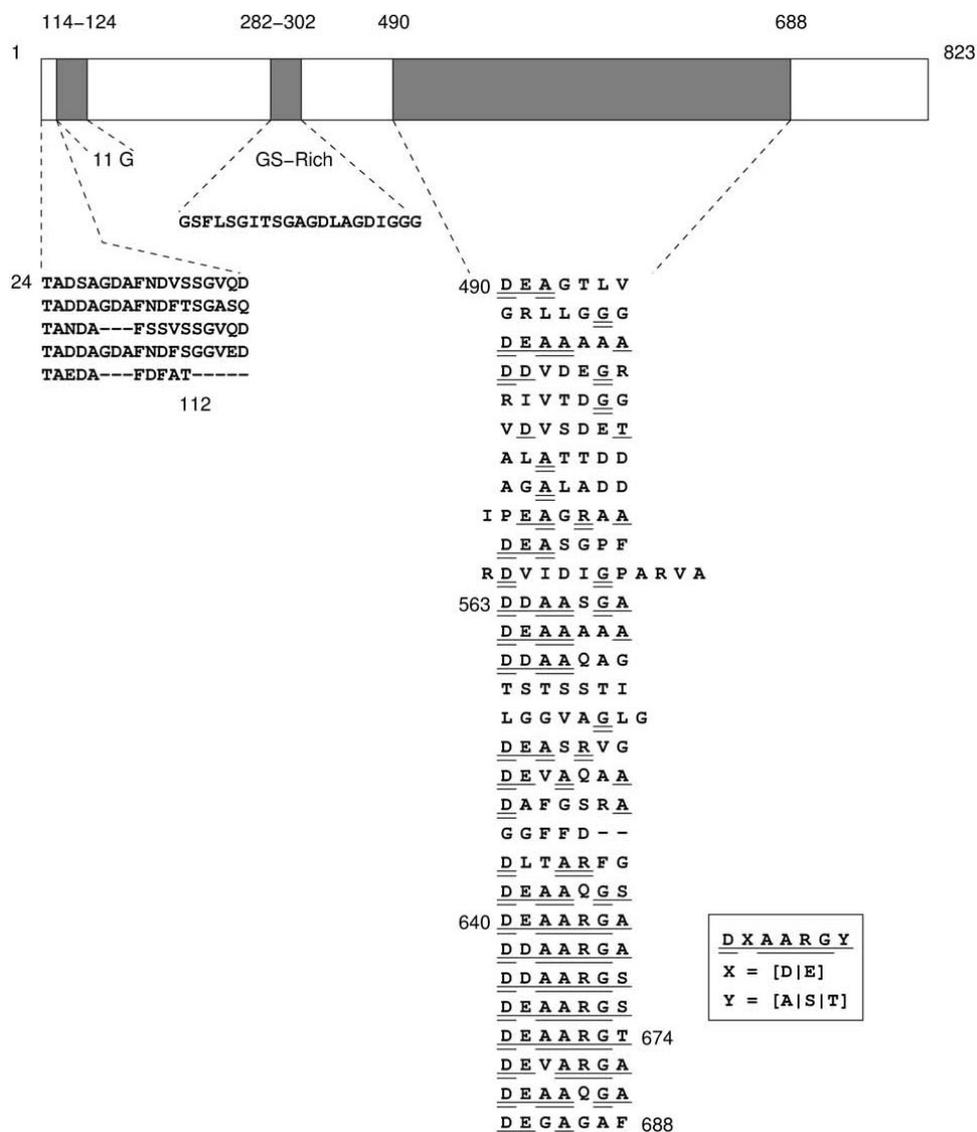


FIG. 4. Schematic presentation of VP2. The two linker-type regions (11G residues and GS-rich sequence) divide the protein in three domains. The repeated sequences in the N- and C-terminal domains are shown below.

qualitative and quantitative differences indicate that the virus selectively acquires lipids from the host lipid pool. Several mechanisms could explain this selectivity. For instance, the virus could derive its lipids from a putative host membrane domain (lateral or transversal) that lacks glycolipids. Alternatively, selection could take place due to selective lipid-protein and/or lipid-DNA interactions, as suggested for the lipid-containing bacteriophage PRD1 (20, 42), or could be due to coulombic repulsion and the shape of lipid molecules, as discussed elsewhere (43). Interestingly, MS analysis showed that in addition to lipid head group selectivity, SH1 is enriched in certain PG species (e.g., $m/z = 793.4$) that probably contain unsaturated and/or cyclopentyl-containing alkyl chains. Such enrichment supports the idea that the molecular shape of such lipids favors their incorporation in the viral membrane (see references 42 and 43).

Database searches using the SH1 genome sequence did not

reveal clear similarities to previously identified sequences. This reinforces the observation that only a limited number of archaeal genomes have been sequenced and that archaeal genes differ considerably from those of bacterial and eukaryotic origin. However, the ORFs seem to be arranged in multiple polycistronic operons, and most of them possess a potential 5' ribosome binding site, a genome organization that resembles that of bacterial viruses (Table 2).

We performed a detailed protein chemistry analysis of the SH1 structural proteins. This allowed us to positively identify 11 genes encoding structural proteins. Comparison of SH1 protein composition to that of PRD1, a bacterial dsDNA virus with an internal membrane, highlighted a clear difference: PRD1 has at least nine integral membrane proteins, whereas SH1 has one very abundant putative integral membrane protein (VP12) and possibly only a few others. As the PRD1 membrane proteins are mostly involved in DNA delivery (29),

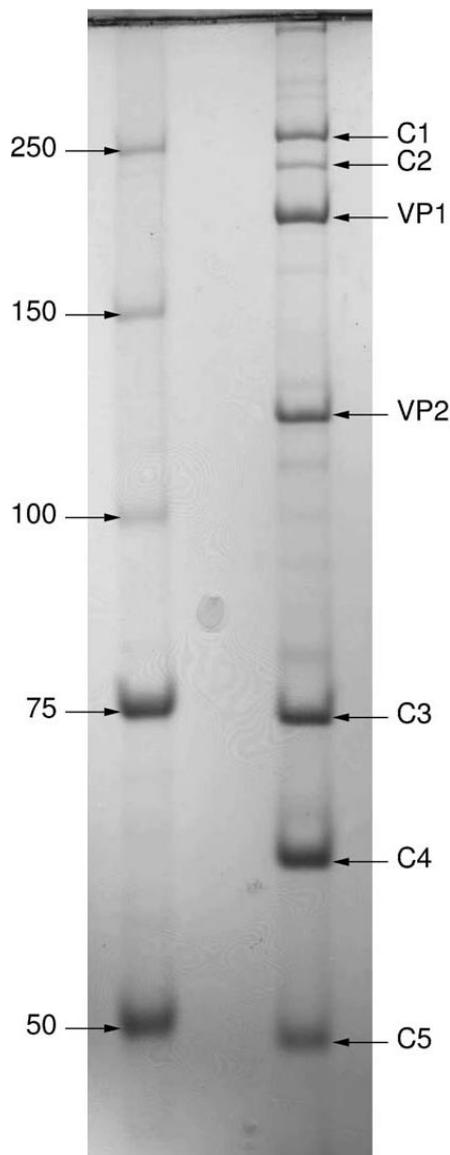


FIG. 5. Nonreducing SDS-PAGE (8% acrylamide) of SH1 showing proteins VP1 and VP2 and protein complexes from 50 kDa to several hundred kDa (C1 to C5). Molecular mass markers (kDa) are shown on the left. The bands were further subjected to peptide mass fingerprinting to identify their protein composition (see text for details).

this suggests that the infection mechanism for SH1 deviates from that of PRD1. Interestingly, one of the SH1 ORFs (ORF 17) encodes a protein with ATPase motifs similar to those observed in internal membrane-containing viruses of bacterial and eukaryotic hosts. A similar protein has been identified as the packaging ATPase in PRD1 (66). The presence of a putative packaging ATPase in SH1 predicts an assembly pathway in which preformed empty viral particles (procapsids) are packaged. This prediction is in agreement with the observed empty progeny particles in thin sections of infected cells (55). The abundance of VP1, VP2, VP3, VP4, and VP7 in the virus particle and their sequence features and multimericity (except VP2 and VP3) suggest that they are components of the virion coat (Fig. 4 and 5). The most obvious explanation for the

multimericity under nonreducing conditions would be the presence of S-S bridges. However, VP4, VP7, and VP1 all have only one cysteine residue. The nature of these multimers requires a more detailed structural analysis.

In addition to the 11 genes encoding structural proteins, there appears to be a large number of ORFs (approximately 40) residing in several operons. Surprisingly, we were not able to identify any ORF with canonical DNA polymerase motifs. Obviously, there are many putative genes available for study that may be involved in replication, transcription, and regulatory functions as well as for cell entry and exit, creating a challenge for future studies.

Can SH1 be related to other viruses? The closest candidate obviously is STIV. SH1 and STIV infect hosts belonging to different archaeal kingdoms. Both viruses are icosahedral with no tail and most likely have an internal membrane component. It has been proposed that the STIV coat architecture and coat protein fold resemble those of bacterial virus PRD1 (60). PRD1, again, has a coat architecture and capsid protein fold that are similar to adenovirus, PBCV-1, and possibly several other viruses infecting bacterial and eukaryotic hosts (12, 13, 14, 33, 49, 70). Based on such observations, it has been proposed that all these viruses originated from a common ancestor (6, 7). This lineage hypothesis predicts that architecturally similar viruses should also be found to infect archaeal hosts. More detailed structural studies of SH1 will reveal whether this lineage hypothesis has predictive power.

ACKNOWLEDGMENTS

Terhi Kemppinen is acknowledged for skillful technical assistance. This investigation was supported by grant 1201964 (to J.K.H.B.), grant 55081 to P.S., and grants 1202855 and 1202108 (to D.H.B.), Finnish Centre of Excellence Program, 2002 to 2005) from the Academy of Finland. The laboratory of M.D.-S. is supported by an MRGS grant from the University of Melbourne.

REFERENCES

- Abrescia, N. G. A., 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. H. Bamford. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**:68–74.
- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Arnold, H. P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J. F. Weidmann, J. K. Kristjanson, H. P. Klenk, K. E. Nelson, and C. M. Fraser. 2000. A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virology* **267**:252–266.
- Baliga, N. S., R. Bonneau, M. T. Facciotti, M. Pan, G. Glusman, E. W. Deutsch, P. Shannon, Y. Chiu, R. S. Weng, R. R. Gan, P. Hung, S. V. Date, E. Marcotte, L. Hood, and W. V. Ng. 2004. Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res.* **14**:2221–2234.
- Bamford, D. H., and H.-W. Ackermann. 2000. Family *Tectiviridae*, p. 111–116. In M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (ed.), *Virus taxonomy: classification and nomenclature of viruses*. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.
- Bamford, D. H., R. M. Burnett, and D. I. Stuart. 2002. Evolution of viral structure. *Theor. Popul. Biol.* **61**:461–470.
- Bamford, D. H. 2003. Do viruses form lineages across different domains of life? *Res. Microbiol.* **154**:231–236.
- Bamford, J. K. H., A.-L. Hänninen, T. M. Pakula, P. M. Ojala, N. Kalkkinen, M. Frilander, and D. H. Bamford. 1991. Genome organization of PRD1, a membrane-containing bacteriophage infecting *E. coli*. *Virology* **183**:658–676.
- Bamford, J. K. H., and D. H. Bamford. 2000. A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to the vertex. *J. Virol.* **74**:7781–7786.

10. Bartolomé, B., Y. Jubete, E. Martínez, and F. de la Cruz. 1991. Construction and properties of a family of pACYC184-derived vectors compatible with pBR322 and its derivatives. *Gene* **102**:75–78.
11. Bell, S. D., and S. P. Jackson. 1998. Transcription and translation in *Archaea*: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* **6**:222–227.
12. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**:825–833.
13. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 2002. The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65Å resolution. *Acta Crystallogr. D* **58**:39–59.
14. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 2004. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* **16**:673–685.
15. Bettstetter, M., X. Peng, R. A. Garrett, and D. Prangishvili. 2003. AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. *Virology* **315**:68–79.
16. Boone, D. R., and R. W. Castenholz. 2001. The Archaea and the deeply branching and phototrophic Bacteria, p. 169–354. *In* G. M. Garrity and D. R. Boone (ed.), *Bergey's manual of systematic bacteriology*, vol. 1, 2nd ed. Springer-Verlag, New York, N.Y.
17. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
18. Brüsson, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. *Cell* **108**:13–16.
19. Caldentey, J., R. Tuma, and D. H. Bamford. 2000. Assembly of bacteriophage PRD1 spike complex: the role of the multidomain protein P5. *Biochemistry* **39**:10566–10573.
20. Cockburn, J. J. B., N. G. A. Abrescia, J. M. Grimes, G. C. Sutton, J. M. Diprose, J. Benevides, G. J. Thomas, Jr., J. K. H. Bamford, D. H. Bamford, and D. I. Stuart. 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* **432**:122–125.
21. Corcelli, A., M. Colella, G. Mascolo, F. P. Fanizzi, and M. Kates. 2000. A novel glycolipid and phospholipids in the purple membrane. *Biochemistry* **39**:3318–3326.
22. Dennis, P. P. 1997. Ancient ciphers: translation in *Archaea*. *Cell* **89**:1007–1010.
23. Dittmer, J. C., and R. L. Lester. 1964. A simple, specific spray for the detection of phospholipids on thin-layer chromatograms. *J. Lipid Res.* **15**:126–127.
24. Dyll-Smith, M. 2001. The Halo handbook: protocols for halobacterial genetics. [Online.] <http://www.microbiol.unimelb.edu.au/micro/staff/mds/>.
25. Dyll-Smith, M., S. L. Tang, and C. Bath. 2003. Haloarchaeal viruses: how diverse are they? *Res. Microbiol.* **154**:309–313.
26. Evans, R. W., S. C. Kushwaha, and M. Kates. 1980. The lipids of *Halobacterium marismortui*, an extremely halophilic bacterium in the Dead Sea. *Biochim. Biophys. Acta* **619**:533–544.
27. Folch, J. M., M. Lees, and G. H. Sloane-Stanley. 1957. A simple method for the isolation and purification of total lipids from animal tissue. *J. Biol. Chem.* **226**:497–509.
28. Graham, D. E., R. Overbeek, G. J. Olsen, and C. R. Woese. 2000. An archaeal genomic signature. *Proc. Natl. Acad. Sci. USA* **97**:3304–3308.
29. Grahn, A. M., R. Daugelavicius, and D. H. Bamford. 2002. Sequential model of phage PRD1 DNA delivery; active involvement of the viral membrane. *Mol. Microbiol.* **46**:1199–1209.
30. Häring, M., X. Peng, K. Brügger, R. Rachel, K. O. Stetter, R. A. Garrett, and D. Prangishvili. 2004. Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the *Globuloviridae*. *Virology* **323**:233–242.
31. Hendrix, R. W. 2004. Hot new virus, deep connections. *Proc. Natl. Acad. Sci. USA* **101**:7495–7496.
32. Huiskonen, J. H. T., L. Laakkonen, M. Toropainen, M. Sarvas, D. H. Bamford, and J. K. H. Bamford. 2003. Probing the ability of the coat and vertex protein of the membrane-containing bacteriophage PRD1 to display a meningococcal epitope. *Virology* **310**:267–279.
33. Iyer, L. M., L. Aravind, and E. V. Koonin. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **75**:11720–11734.
34. Juez, G., F. Rodriguez-Valera, A. Ventosa, and D. J. Kushner. 1986. *Haloarcula hispanica* spec. nov. and *Haloferax gibbonsii* spec. nov., two new species of extremely halophilic archaeobacteria. *Syst. Appl. Microbiol.* **8**:75–79.
35. Kahma, K., J. Brotherus, M. Haltia, and O. Renkonen. 1976. Low and moderate concentrations of lysobisphosphatidic acid in brain and liver of patients affected by some storage diseases. *Lipids* **11**:539–544.
36. Kamekura, M., and M. Kates. 1999. Structural diversity of membrane lipids in members of *Halobacteriaceae*. *Biosci. Biotechnol. Biochem.* **63**:969–972.
37. Kates, M. 1972. Techniques of lipidology: isolation, analysis and identification of lipids. North-Holland Publishing Co., Amsterdam, The Netherlands.
38. Kates, M. 1993. Membrane lipids of Archaea, p. 261–295. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), *The biochemistry of Archaea* (Archaeobacteria). Elsevier Science Publishers, Amsterdam, The Netherlands.
39. 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.
40. Klein, R., U. Baranyi, N. Rössler, B. Greineder, H. Scholz, and A. Witte. 2002. *Natrialba magadii* virus ϕ Ch1: first complete nucleotide sequence and functional organization of a virus infecting a haloalkaliphilic archaeon. *Mol. Microbiol.* **45**:851–863.
41. Krogh, B., G. Larsson, G. von Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
42. Laurinavicius, S., R. Käkelä, P. Somerharju, and D. H. Bamford. 2004. Phospholipid molecular species profiles of tectiviruses infecting gram-negative and gram-positive hosts. *Virology* **322**:328–336.
43. Laurinavicius, S., R. Käkelä, D. H. Bamford, and P. Somerharju. 2004. The origin of phospholipids of the enveloped bacteriophage ϕ h6. *Virology* **326**:182–190.
44. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglis of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol.* **184**:4891–4905.
45. Lukashin, A., and M. Borodovsky. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
46. Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblooded onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**:10035–10038.
47. Meile, L., U. Jenal, D. Studer, M. Jordan, and T. Leisinger. 1989. Characterization of ψ M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. *Arch. Microbiol.* **30**:105–110.
48. Mindich, L., D. H. Bamford, T. McGraw, and G. MacKenzie. 1982. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J. Virol.* **44**:1021–1030.
49. Nandhagopal, N., A. A. Simpson, J. R. Gurnon, X. Yan, T. S. Baker, M. V. Graves, J. L. Van Etten, and M. G. Rossmann. 2003. The structure and evolution of the major capsid protein of a large, lipid-containing, DNA virus. *Proc. Natl. Acad. Sci. USA* **99**:14758–14763.
50. Ng, W. V., S. P. Kennedy, G. C. Mahairas, B. Berquist, M. Pan, H. D. Shukla, S. R. Lasky, N. S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T. A. Dahl, R. Welti, Y. A. Goo, B. Leithauer, K. Keller, R. Cruz, M. J. Danson, D. W. Hough, D. G. Maddocks, P. E. Jablonski, M. P. Krebs, C. M. Angevine, H. Dale, T. A. Isenbarger, R. F. Peck, M. Pohlshroder, J. L. Spudich, K. W. Jung, M. Alam, T. Freitas, S. Hou, C. J. Daniels, P. P. Dennis, A. D. Omer, H. Ebbhardt, T. M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma. 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. USA* **97**:12176–12181.
51. Olkkonen, V. M., and D. H. Bamford. 1989. Quantitation of the adsorption and penetration stages of bacteriophage ϕ 6 infection. *Virology* **171**:229–238.
52. Pedulla, M. L., M. Ford, J. M. Houtz, T. Karthikeya, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, D., J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jr. Jacobs, R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**:171–182.
53. Peng, X., H. Blum, Q. She, S. Mallok, K. Brügger, R. A. Garrett, W. Zilling, and D. Prangishvili. 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology* **291**:226–234.
54. Pfister, P., A. Wasserfallen, R. Stettler, and T. Leisinger. 1998. Molecular analysis of *Methanobacterium* phage ψ M2. *Mol. Microbiol.* **30**:233–244.
55. Porter, K., P. Kukkaro, J. K. H. Bamford, C. Bath, H. M. Kivela, M. L. Dyll-Smith, and D. H. Bamford. 2005. SH1: a novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology* **335**:22–33.
56. Poutanen, M., L. Salusjärvi, L. Ruohonen, M. Penttilä, and N. Kalkkinen. 2001. Use of matrix-assisted laser desorption/ionization time-of-flight mass mapping and nanospray liquid chromatography/electrospray ionization tandem mass spectrometry sequence tag analysis for high sensitivity identification of yeast proteins separated by two-dimensional electrophoresis. *Rapid Commun. Mass Spectrom.* **15**:1685–1692.
57. Prangishvili, D., K. Stedman, and W. Zilling. 2001. Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol.* **9**:39–43.
58. Prangishvili, D., and R. A. Garrett. 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem. Soc. Trans.* **32**:204–208.
59. Ravanti, J. J., A. Gaidelyte, D. H. Bamford, and J. K. H. Bamford. 2003. Comparative analysis of the genomes of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. *Virology* **313**:401–414.
60. Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J. E. Johnson, T. Douglas, and M. Young. 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl. Acad. Sci. USA* **101**:7716–7720.
61. Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**:276–277.
62. Sartorius-Neef, S., and F. Pfeifer. 2004. In vivo studies on putative Shine-

- Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol. Microbiol.* **51**:579–588.
63. **Schägger, H., and G. von Jagow.** 1987. Tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**:368–379.
64. **Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass-spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**:850–858.
65. **Strömsten, N. J., D. H. Bamford, and J. K. H. Bamford.** 2003. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J. Virol.* **77**:6314–6321.
66. **Strömsten, N. J., D. H. Bamford, and J. K. H. Bamford.** 2005. In vitro DNA packaging of bacteriophage PRD1: a common mechanism for internal-membrane viruses. *J. Mol. Biol.* **348**:617–629.
67. **Tang, S.-L., S. Nuttall, K. Ngui, C. Fisher, P. Lopez, and M. Dyll-Smith.** 2002. HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol. Microbiol.* **44**:283–296.
68. **Tang, S.-L., S. Nuttall, and M. Dyll-Smith.** 2004. Haloviruses HF1 and HF2: evidence for a recent and large recombination event. *J. Bacteriol.* **186**:2810–2817.
69. **Tolstrup, N., C. W. Sensen, R. Garrett, and I. G. Clausen.** 2000. Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* **4**:175–179.
70. **Van Etten, J. L., M. V. Graves, D. G. Müller, W. Boland, and N. Delaroque.** 2002. *Phycodnaviridae*—large DNA algal viruses. *Arch. Virol.* **147**:1479–1516.
71. **Verheust, C., G. Jensen, and J. Mahillon.** 2003. pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar *israelensis*. *Microbiology* **149**:2083–2092.
72. **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.
73. **Zuker, M., D. H. Mathews, and D. H. Turner.** 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p. 11–43. *In* J. Barciszewski, and B. F. C. Clark (ed.), *RNA biochemistry and biotechnology*. NATO ASI Series. Kluwer Academic Publishers, Dordrecht, The Netherlands.