

Analysis of the First Genome of a Hyperthermophilic Marine Virus-Like Particle, PAV1, Isolated from *Pyrococcus abyssi*^{∇†}

C. Geslin,^{1‡*} M. Gaillard,^{1‡} D. Flament,^{1*} K. Rouault,^{1,2} M. Le Romancer,¹
D. Prieur,¹ and G. Erauso¹

Laboratoire de Microbiologie des Environnements Extrêmes, IUEM, UBO, CNRS, IFREMER, UMR 6197, Technopôle Brest-Iroise, Place Nicolas Copernic, 29280 Plouzané, France,¹ and Laboratoire de Génétique Moléculaire et Génétique Épidémiologique, INSERM U613, Brest 29220, France²

Received 15 December 2006/Accepted 12 April 2007

Only one virus-like particle (VLP) has been reported from hyperthermophilic *Euryarchaeotes*. This VLP, named PAV1, is shaped like a lemon and was isolated from a strain of “*Pyrococcus abyssi*,” a deep-sea isolate. Its genome consists of a double-stranded circular DNA of 18 kb which is also present at a high copy number (60 per chromosome) free within the host cytoplasm but is not integrated into the host chromosome. Here, we report the results of complete analysis of the PAV1 genome. All the 25 predicted genes, except 3, are located on one DNA strand. A transcription map has been made by using a reverse transcription-PCR assay. All the identified open reading frames (ORFs) are transcribed. The most significant similarities relate to four ORFs. ORF 180a shows 31% identity with ORF 181 of the pRT1 plasmid isolated from *Pyrococcus* sp. strain JT1. ORFs 676 and 678 present similarities with a concanavalin A-like lectin/glucanase domain, which could be involved in the process of host-virus recognition, and ORF 59 presents similarities with the transcriptional regulator CopG. The genome of PAV1 displays unique features at the nucleic and proteinic level, indicating that PAV1 should be attached at least to a novel genus or virus family.

Recent studies on hyperthermophilic members of the domain *Archaea* from terrestrial or oceanic hydrothermal environments suggest the existence of an impressive morphological and genomic viral diversity (19, 39, 44, 45, 47).

The vast majority of the hyperthermophilic viruses were isolated from the *Crenarchaeota* phylum and found to infect members of the genera *Sulfolobus*, *Thermoproteus*, *Acidianus*, and *Pyrobaculum*. The unusual morphological and genomic properties of these viruses have led to the definition of seven new families (*Fuselloviridae*, *Lipothrixviridae*, *Rudiviridae*, *Guttaviridae*, *Globuloviridae*, *Bicaudaviridae*, and *Ampullaviridae*) encompassing the 20 representatives identified to date (3, 4, 8, 22–25, 28, 34, 43, 56, 59). They all carry double-stranded DNA (dsDNA) genomes, either circular or linear ranging from 15 to 75 kbp in size. Sequence similarities between genes of the different crenarchaeal viral families are generally limited, and most predicted genes have homologues only in other members of the same family (46).

In the euryarchaeal phylum, which mainly includes extreme halophiles, methanogens, and hyperthermophilic sulfur reducers represented by the *Thermococcales* order, almost all the viruses described were isolated from mesophilic hosts. Most of

the euryarchaeal viruses possess a linear dsDNA genome of 14.9 to 230 kbp (47, 51) with two exceptions: the A3 virus-like particle (VLP), isolated from *Methanococcus voltae*, has a circular dsDNA genome (23 kb) and can integrate into the host chromosome (like fuselloviruses) (61) and ϕ Ch1 isolated from *Natrialba magadii* which strikingly contains RNA in addition to DNA and may also integrate into the host chromosome (60). The majority of the characterized viruses display the “classical” head and tail morphology and are distributed between the two well-known families of *Myoviridae* and *Siphoviridae* found mostly represented in the *Bacteria* domain. Four viruses, His1, His2, and SH1 isolated from *Haloarcula hispanica* and the particle A3 VLP, however, are exceptions. SH1 is a polyhedral virus, A3 VLP is oblate, and His1 and His2 are lemon-shaped viruses (6, 15, 42, 61). Direct electron microscopy observations of hypersaline waters have shown that lemon-shaped and round VLPs are not an exception but the predominant morphotypes in this particular environment, while head and tail particles are less common but represent the majority of reported haloviruses (15, 38).

Lemon-shaped viruses are widespread and have been isolated from a broad host range in both archaeal phyla. The best-studied lemon-shaped viruses were isolated from the hyperthermophilic genus *Sulfolobus* of the *Crenarchaeota* phylum. *Sulfolobus* spindle-shaped viruses all have a circular dsDNA genome of about 15 kbp with approximately 34 open reading frames (ORFs). Their genomes integrate into the host tRNA genes. These *Fusellovirus* ORFs show little or no similarity to genes in the public databases. In contrast, 18 ORFs are common to all and may represent the minimal set defining this viral group (59).

Only one VLP has been described so far from hyperthermophilic euryarchaeotes. PAV1, a lemon-shaped VLP (120 nm ×

* Corresponding author. Mailing address for Claire Geslin: Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197 (CNRS, UBO, IFREMER), Place Nicolas Copernic, Technopôle Brest-Iroise, 29280 Plouzané, France. Phone: 33 (0)2 98 49 88 58. Fax: 33 (0)2 98 49 87 05. E-mail: claire.geslin@univ-brest.fr. Phone for Didier Flament (sequence analysis questions): 33 (0)2 98 22 45 27. E-mail: Didier.Flament@ifremer.fr.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

‡ Claire Geslin and Mélusine Gaillard contributed equally to this work.

[∇] Published ahead of print on 20 April 2007.

80 nm), was isolated from *Pyrococcus abyssi* strain GE23, a deep-sea isolate previously described in our laboratory (18). We found that host cells spontaneously release few PAV1 particles without lysis in the growth cycle with a maximum reached in stationary phase. PAV1 contains a circular dsDNA of 18 kb which is present at a high copy number and in a "plasmidic" form within the host cytoplasm (no chromosomal integration detected) (20).

The resemblance of PAV1 and His1 viruses to spindle-shaped virus SSV1 in morphology and genome size first led to the proposal that His1 and PAV1 be included in the *Fuselloviridae* family. The isolation and characteristics of a second spindle-shaped halovirus, His2, showed that His1 and His2 are distantly related to each other but are not related to members of the *Fuselloviridae*. Analysis showed that these viruses have a lytic life cycle; their linear genomes replicate by a protein-primed DNA synthesis and encode a DNA polymerase. All these features differ fundamentally from those of the fuselloviruses, which led to the classification of His1 and His2 into the genus *Salterprovirus* (6). How should PAV1 be classified? Here, we present the results of analysis of the complete double-stranded sequence of PAV1 genome with its transcriptional map. We show that the PAV1 genome notably shares little similarity with those of other archaeal viruses. We discuss the putative functions of proteins encoded by its genome, and finally we propose that PAV1 be the first member of a new virus genus or family.

MATERIALS AND METHODS

DNA manipulation. Total DNA from the *Pyrococcus abyssi* GE23 host strain was prepared as previously described (20). The viral covalently closed circular DNA (cccDNA) was isolated by the alkaline lysis method as previously described (19) and purified by isopycnic centrifugation in a cesium chloride gradient (0.81 g/ml of CsCl) in the presence of ethidium bromide (0.52 mg/ml) (48).

To determine the cccDNA copy number of PAV1 virus, total DNA from exponentially growing cultures of the *Pyrococcus abyssi* GE23 host strain was cleaved successively with HindIII and NaeI. Appropriate dilutions of the digested DNA were run on a 0.8% agarose gel and transferred to a nylon membrane (Hybond N+; Amersham). Southern hybridization was performed with an equimolar mixture of two fluorescein-labeled probes, a 1.7-kbp HindIII fragment of the PAV1 genome and a 0.9-kbp NaeI cloned fragment of the 16S rRNA gene of strain GE23. The probes were labeled using ECF random-prime labeling kit (Amersham). Hybridization and detection were carried out following the ECF system procedure. Hybridization signals were recorded and quantified using a Typhoon scan imager (Amersham). PAV1 cccDNA copy number was calculated by measuring the average ratio of the PAV1-specific signal to the 16S rRNA gene signal corrected for the fragment length difference.

Detection of putative single-stranded DNA (ssDNA) intermediates of replication was carried out as previously described (17). The different forms present in the native PAV1 cccDNA preparation were separated by electrophoresis on 0.8% agarose gels containing 0.5 µg/ml of ethidium bromide. To detect ssDNA only, DNA in the gel was directly transferred in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) to a nylon membrane (Hybond N; Amersham), omitting the denaturation and neutralization steps of the standard protocol (49). Other gels were transferred under standard denaturing conditions. After transfer, DNA was immobilized on the membranes by UV cross-linking. RNA probes specific to each of the PAV1 DNA strands were prepared. For that purpose, a PCR product of 1.1 kb of the PAV1 genome was cloned into the pGEM-T vector (Promega); the cloning site is flanked by a T7 RNA polymerase promoter on one side and an SP6 RNA polymerase promoter on the other side. The plasmid linearized by cutting either with SalI (for the T7 promoter) or SphI (for the SP6 promoter) was used as a template for runoff transcription using either the T7 or SP6 RNA polymerase and the digoxigenin-labeled UTP ribonucleotide mix (Roche) for probe labeling. Hybridization and detection were carried out following the digoxigenin-labeled UTP procedure according to the manufacturer's instructions (Roche).

Determination of the nucleotide sequence. Purified PAV1 cccDNA was completely digested with HindIII and BamHI, and all the fragments obtained were cloned in the corresponding sites of pUC28 to obtain an overlapping clone library of PAV1 genome. Sequencing reactions were carried out with the BigDye terminator kit (Applied Biosystems) and analyzed on an ABI PRISM 377 automatic sequencer (Applied Biosystems). Each insert was sequenced from both ends using the M13 forward and M13 reverse universal primers. Gaps in the sequence were filled by using specific primers either directly for sequencing on library clones or to sequence PCR amplicons obtained with PAV1 cccDNA as the template. The sequences were trimmed and assembled using the SeqMan II program (Lasergene, Inc., Madison, WI) with both strands completely sequenced and with a minimum threefold coverage.

Sequence analysis and annotation. GLIMMER (14) and RBS finder (53) was used to find ORFs. Each ORF was submitted to sequence similarity searches (BLASTN, BLASTP, and BLASTX) against the NCBI nonredundant protein and nucleic acid databases (October 2006). ORFs were also analyzed with the library of hidden Markov model pfam version 14.0 with the HMMER package (16) and by the fold recognition method GenTHREADER (36). Membrane-spanning region in ORFs were predicted by the TMHMM program (30), and protein membrane topology was predicted using the TMPred program (26). Cumulative GC skew analysis was performed using GC Skew Tool (<http://bioinformatics.upmc.edu/SKEW/index.html>). It was calculated according to the following formula: $\Sigma(G - C)/(G + C)$, using a sliding window of 20 nucleotides (nt).

Analysis of the helical stability is computed using the nearest-neighbor thermodynamics algorithm (27) using the web-based program WEB-THERMODYN (<http://wings.buffalo.edu/gsa/dna/dk/WEBTHERMODYN/>).

Finally, the similarity between sequences was calculated using the GAP program (Wisconsin package version 10.3, accelrys).

RT-PCR. Total RNA isolation was prepared using the TRIzol reagent and the procedure described by Invitrogen from 100-ml batch cultures of *P. abyssi* strain GE23 arrested in the late exponential phase (3×10^8 cells ml⁻¹). RNA pellets (50 to 80 µg) were resuspended in 50 µl of 5 mM Tris-HCl (pH 8.0). The concentration and purity of the RNA were determined by measuring the absorbance at 260 and 280 nm, and the RNA quality was estimated by electrophoresis of an aliquot on a 1.2% agarose gel under denaturing conditions (48). Contaminant DNA was removed by treatment with RNase-free DNase (Promega) according to the manufacturer's instructions. After deproteinization by phenol-chloroform extraction, RNA in aqueous phase was ethanol precipitated, washed with 75% ethanol, air dried, and finally resuspended in 10 mM Tris-HCl (pH 8.0). Reverse transcription-PCR (RT-PCR) was performed by using the Superscript RT-PCR kit (Invitrogen). Aliquots of 1 µg RNA were used for each RT-PCR in a 50-µl reaction volume containing 0.5 U of RNase inhibitor (Promega) and the appropriate primers at a final concentration of 0.2 µM. The following program was run on a PCR machine: cDNA synthesis at 48°C for 25 min, predenaturation at 94°C for 3 min, and then 30 cycles consisting of denaturation at 94°C for 30 s, annealing at 48°C for 30 s, and elongation at 70°C for 1 min. This was followed by one final extension step at 70°C for 7 min. Negative controls were obtained by replacing the reverse transcriptase/*Taq* mixture by the same unit amount of *Taq*. The products were separated on a 1.5% agarose gel.

A list of primers used for mapping transcripts and agarose gels showing the results of RT-PCR assays are given in the supplementary material.

Isolation and identification of a major protein of PAV1. PAV1 VLPs were purified as previously described (20). *P. abyssi* strains GE23 and GE9 (closely related to GE23 but VLP-free) were used to distinguish residual cellular proteins from viral proteins. The two strains were centrifuged at 6°C for 15 min at 3,000 × *g*. The supernatant was filtered through Ascrodisc PF 0.8/0.2-µm filters (Pall Gelman Laboratory) and concentrated by another centrifugation at 6°C for 3 h at 100,000 × *g*. The pellet was resuspended in 50 µl of TE buffer (10 mM Tris-HCl, 0.1 mM EDTA [pH 8]). Equal volumes of sample and loading buffer (187.5 mM Tris-HCl [pH 6.8], 30% glycerol, 6% sodium dodecyl sulfate [SDS], 15% β-mercaptoethanol, 0.15% bromophenol blue) were mixed in microtubes and boiled at 95°C for 4 min before loading on 15% acrylamide-SDS gels. Proteins in the sample were analyzed by the method of Sambrook et al. (48). Gels were stained by Coomassie blue. The major band was cut off, and the N-terminal region of the protein in the band was sequenced by the Edman technique using the facilities of the genomic platform at INRA, Jouy-en-Josas, France.

Nucleotide accession number. The complete PAV1 genome sequence was deposited in GenBank under accession number EF071488.

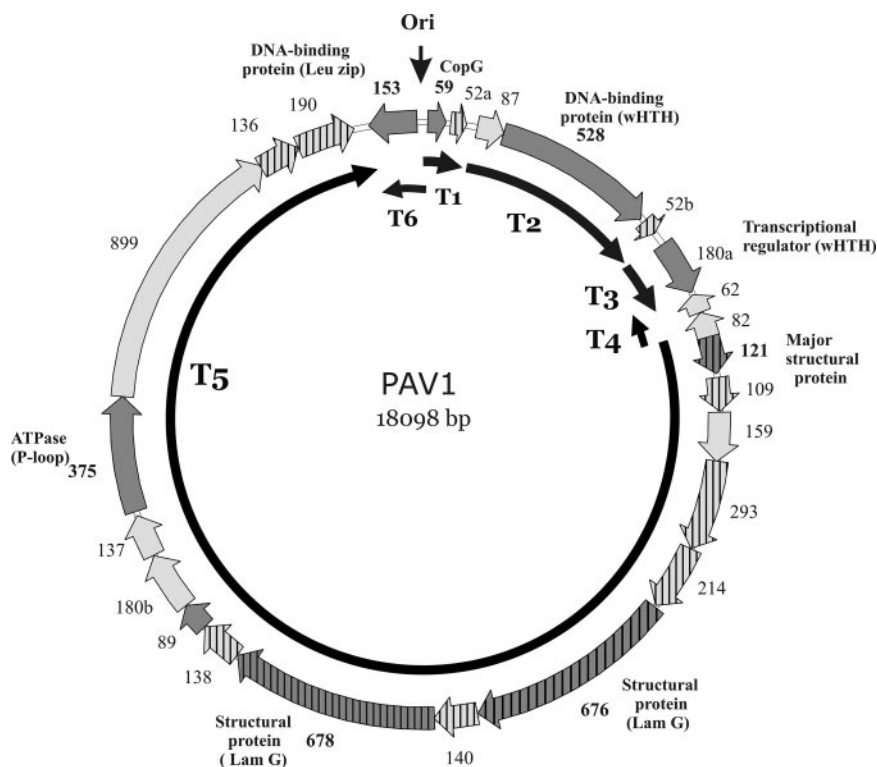


FIG. 1. PAV1 genome map. Predicted genes are represented by thick arrows; light gray shading indicates ORFs with no similarity and no assigned function, and dark gray shading indicates either conserved hypothetical ORFs or ORFs with a hypothetical function; hatching indicates ORFs encoding a putative membrane-associated protein. Transcript locations mapped by using RT-PCR (T1 to T6) are shown by arrows inside the circular genome map. The approximate location of the origin (Ori) of replication as predicted by cumulative GC skew is also indicated. Protein motif or domain names: Lam G, laminin G; Leu zip, leucine zipper; P-loop, nucleoside triphosphate binding site or Walker motif A; wHTH, winged helix or winged helix-turn-helix.

RESULTS

Properties of the nucleotide sequence of PAV1. The complete nucleotide sequence of the PAV1 genome was determined on both strands as outlined in Materials and Methods. The restriction map determined previously using purified ccdDNA was found to be in strict agreement with that predicted from the sequence, and the estimated genome size determined in this study ($17,500 \text{ bp} \pm 500 \text{ bp}$) (20) was very close to that determined by sequencing (18,098 bp).

Sequence analysis in the six possible frames allowed us to identify 25 putative ORFs encoding at least 50 amino acids which cover 95% of the total sequence. The positions of the ORFs are indicated in Fig. 1, and their main features are listed in Table 1. The majority of the putative genes (22 ORFs) have the same orientation, and only three are present on the complementary strand. The average G+C content of the coding regions is 46%, which is similar to the overall G+C contents of the PAV1 genome (47.15%) and host genome (42.8%).

All ORFs on the forward strand of PAV1 virion were preceded by a potential ribosome binding site with a consensus sequence GG(A/G)GGTGA, similar to that computed for the *P. abyssi* genome and located at positions -13 to -5 upstream of the putative start codon as previously shown for *P. abyssi* genes (11). The distribution of the PAV1 ORF start codon (76.2% ATG, 14.3% GTG, and 9.5% TTG) is similar to that

observed for *P. abyssi* genes (83%, 11%, and 6%, respectively), indicating that the prediction of the initiation codon is globally correct for the PAV1 genome. The mean deviation in codon usage between the coding regions of PAV1 and its host is 20%. This suggests that PAV1 genes fit well with its host translational apparatus. Accordingly, the genome of PAV1 does not encode any tRNA.

Hypothetical origin of replication. In an attempt to identify the origin of replication, a cumulative GC skew analysis has been performed on the PAV1 genome (Fig. 2A). Although the diagram reveals a strand asymmetry between the leading and lagging strands, it does not allow us to identify a clear origin of replication. Yet, two inflections can be observed, indicating a change in base composition bias. The inflection between positions 16000 and 17386 bp corresponds to the end of the putative membrane protein operon (see below), the genes in this region are tightly packed, and no repeated domains could be detected. In contrast, the DNA sequence between positions 17386 and 200 bp contains the two largest intergenic regions, 140 bp and 111 bp, respectively. Furthermore, helical stability analysis indicates that positions 17987 to 1 correspond to the lowest helical stability region, suggesting that this region contains a DNA unwinding element (DUE). It has been shown that this element corresponds to the DNA sequence that first unwinds during the initiation of genome replication for differ-

TABLE 1. General features of the predicted genes (ORFs and operons) of PAV1 virus

ORF ^a	Nucleotide positions (start-stop)	S-D motif ^b	Promoter and terminator motifs ^c	Similar sequence ^d	Domain(s) ^e
59	1-180		TATAATAATTA (-31)		[9-48] RHH protein of the CopG family (pfam01402, E value of 3e-05)
52a	222-380		TTTTTTT (+50)		TM (2)
87	484-747	GGTGGG (-5)	AAACTTTATA (-35)		[352-414] "winged helix" DNA binding domain (SSF46785, E value of 4.7e-07)
528	744-2330	GGGTGA (-7)			
52b	2364-2522	GGGGAG (-8)	GGGGGGG (+6)		TM (2)
180a	2652-3194	GGTGG (-6)	AAAACATTTAAA (-36); (+24) TCTTCTTT	ORF 181 of plasmid pRT1 (<i>Pyrococcus</i> sp. strain JT1) [4e-09]	
62 (-)	3214-3402	GGGGGTGA (-5)	CCCTCCCC		
82 (-)	3402-3650				
121	3649-4014	GGAGGT (-7)	AAATGCTTAAA (-78)		TM (2)
109	4062-4391	GGAGGTGA (-8)			TM (3)
158	4388-4867	AGGAGG (-6)			TM (2)
293	4864-5745	AGGAGG-GA (-5)			TM (4)
214	5751-6395	GGGTGA (-6)			TM (3); [204-281] concanavalin A-like lectin/glucanase (SSF49899, E value of 4.6e-04)
676	6411-8441	AGGGGTG (-6)			TM (3)
140	8448-8870	AGGGGTGA (-7)			Hypothetical protein ORF 175 of bacteriophage S-PM2 [4e-04]
678	8867-10903	GGGGGTG (-6)	TM (4)		
138	10909-11325	GAGGTGA (-6)	Hypothetical protein PAB2035 (<i>Pyrococcus abyssi</i>) [2e-04]		
89	11325-11594	GGGGGTGA (-5)			
180b	11599-12141	AGGAGGTGA (-5)		ORF pNG4027 of plasmid pNG4000 (<i>Haloarcula marismortui</i> ATCC 43049) [3e-05]	9.5e-05)
137	12148-12561	GAGGTGA (-7)			
375	12603-13730	GAGAGGTG (-6)			
898	13730-16426	GGAGG (-5)	CCGCGGTGGTTGCCCTTCTATTTTT (+41)		TM (3)
136	16404-16814	GGAGTG (-7)			TM (4)
190	16814-17386	GGAGGTG (-9)			
153 (-)	17526-17987		AATTATTATA (-65); (+9) TTTCTTATTT		[13-42] leucine zipper; [86-148] "winged helix" DNA-binding domain (SSF46785, E value of 6.2e-06)

^a The ORFs are named according to the size (in amino acids) of the protein they encode. (-) indicates that the ORFs are located on the reverse strand. The ORFs are grouped in putative operons by the vertical lines (e.g., ORF 59 and ORF 52a are one putative operon).

^b S-D motif, Shine-Dalgarno motif. The putative 5' upstream ribosome binding sequence complementary to the 3' end of the *P. abyssi* 16S rRNA (GG[A/G]GGTGA). The numbers in parentheses indicate the position of the last nucleotide of the ribosome binding sequence relative to the predicted start codon.

^c The putative TATA-like promoter and pyrimidine-rich terminator sequences are listed. The numbers in parentheses indicate the position of the last nucleotide of the promoter the first nucleotide of the terminator sequence relative to the predicted start or stop codon.

^d E values of BLAST hits are given in brackets.

^e Predicted domains revealed by InterProScan and TMHMM programs. The numbers in brackets indicate the positions of the domains in the polypeptide sequence and are given before the domain. The presence and number of transmembrane segments are indicated [e.g., TM (2) for two transmembrane segments]. In parentheses, the accession numbers given for the motifs and domains found refer to Pfam (pfam) and Superfamily (SSF) databases.

ent species as well as for certain viruses and phages (27). In particular, two DUEs have also been identified within the *oriC* region that corresponds to the origin of replication of *P. abyssi* (35). In addition, this DNA sequence portion also contains two large inverted repeats, as well as six copies of an irregular 12- to 16-bp repeat that has the ability to form a stem-loop structure (Fig. 2B).

To obtain further insight into the mechanism of PAV1 DNA replication, detection of a putative ssDNA intermediate of replication, generated by a rolling circle (RC) mechanism, was

carried out as previously described (17). Although ssDNA was detected, it hybridized equally with either the plus- or minus-strand-specific RNA probe, indicating that this ssDNA was not a RC replication intermediate but rather an unspecific denatured form of cccDNA produced during the preparation of PAV1 lysate.

Sequence similarities of predicted proteins from PAV1. Half of the ORFs (12 of 25) have been predicted to have transmembrane regions.

To infer hypothetical functions to the predicted proteins of

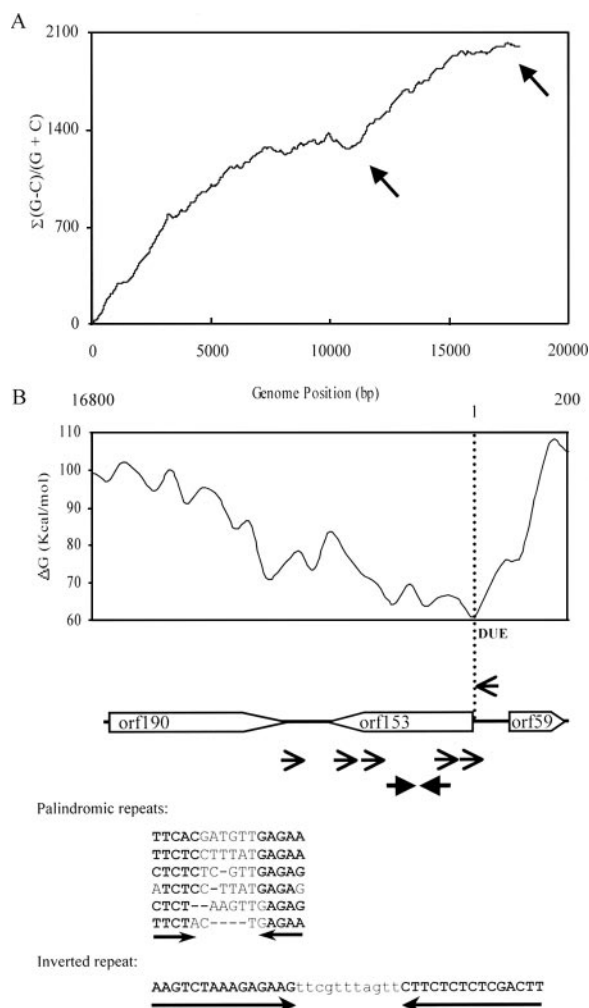


FIG. 2. Structure of the putative DNA replication origin region. (A) Cumulative GC skew (window size of 20 bp). (B) Identification of a DNA unwinding element (DUE) and nucleotide repeats (indicated by arrows). The longer inverted repeats are indicated by the thick black arrows.

PAV1, their amino acid sequences were compared to those in the public sequence databases (see Materials and Methods). Sixty-five percent of the predicted proteins had no significant similarity with any protein sequences stored in the public databases. Among them nine of the ORFs have been predicted to have membrane-spanning regions.

ORF 59 encodes a small protein of 59 amino acids. Multiple sequence alignments with a domain of 41 amino acids of ORF 59 with various members of the CopG family, as well as secondary structure prediction, suggest that the ORF 59 gene product has a ribbon-helix-helix (RHH) arrangement (E value of $3.6e-05$). The RHH domain proteins are known to be transcription regulators.

ORF 180a encodes a putative protein of 22.1 kDa that is 31% identical to ORF 181 of plasmid pRT1 from *Pyrococcus* sp. strain JT1 (E value of $4e-17$) (58). It was found that ORF 181 has a level of identity of approximately 50%, on a 35-amino-acid region at the C terminus, with ORF 80 of the *Sulfolobus* plasmid family pRN (32).

The overall organization of the gene products ORF 676 and ORF 678 is similar. The predicted peptides are composed of a typical signal peptide and two transmembrane segments at the carboxy-terminal regions. The prediction of the membrane topology suggests that both peptides are exposed to the surface of the enveloped virus and anchored to the membrane by the C-terminal transmembrane regions.

After two iterations of PSI-BLAST search, ORFs 676 and 678 show significant similarities with very large proteins that have not been assigned a clear function but seem to be involved in adhesion. This set mainly includes VCBS protein (IPR 010221) and laminin G-like jellyroll fold (LamGL) containing proteins (IPR 006558). This result is in accordance with the fact that ORF 676 and ORF 678 contain one and two occurrences, respectively, of the LamGL domain (Fig. 3). ORF 678 also presents similarities with ORF 175 of the bacteriophage S-PM2, a T4-type bacteriophage that infects the marine photosynthetic bacteria *Synechococcus* spp. (33). ORF 175 of S-PM2 also contains two occurrences of sequence similar to the LamGL domain. This domain is a member of the concanavalin-like lectin/glucanase superfamily. ORF 175 has been assigned a putative function in host recognition on the basis of both the primary structure of the protein and the genomic environment of the encoding gene (33).

The ORF 89 gene product is the sole protein of PAV1 that has sequence similarity with archaeal proteins only. After five PSI-BLAST iterations, ORF 89 shows significant similarities (E value from $5e-13$ to $3e-24$) with proteins from *Pyrococcus abyssi* (gi 33356692), *Thermococcus kodakarensis* (gi 57160005 and gi 57159608), and *Aeropyrum pernix* (gi 5104530). Those proteins are 90 to 125 amino acids in length and are annotated as hypothetical proteins.

The putative gene product of ORF 375 presents a predicted ATP binding site of the canonical form ($GX_4GK[T/S]$) and is thought to be an ATP/GTP binding protein.

ORF 153 had no sequence similarity to sequences in public databases; however, it displays a conspicuous domain organization and location. It is indeed located at the putative origin and contains a coiled-coil domain (positions 13 to 42) followed by a winged helix DNA binding domain (positions 86 to 148; Table 1). The alpha-helical representation of the N-terminal coiled-coil domain shows a clear distribution of hydrophobic residues (L, I, and V) in a leucine zipper region (positions 13 to 42). This domain organization might indicate that the leucine zipper region could pack into a coiled coil to form a homodimer and that the C-terminal region could have DNA binding properties via the winged helix DNA binding domain.

ORF 528 also contains a "winged helix" DNA binding domain (positions 352 to 414).

Polycistronic mRNA analysis and potential transcription signals. A transcription map was made using RT-PCR assays. Primers were designed on two colinear genes, so that cotranscribed ORFs could be amplified. This procedure was repeated step by step all along the genome, and the resulting transcript map is shown in Fig. 1.

By this approach, we found that all predicted genes of PAV1 were transcribed in six mRNAs whose approximate sizes (T1 [380 nt], T2 [2,530 nt], T3 [580 nt], T4 [440 nt], T5

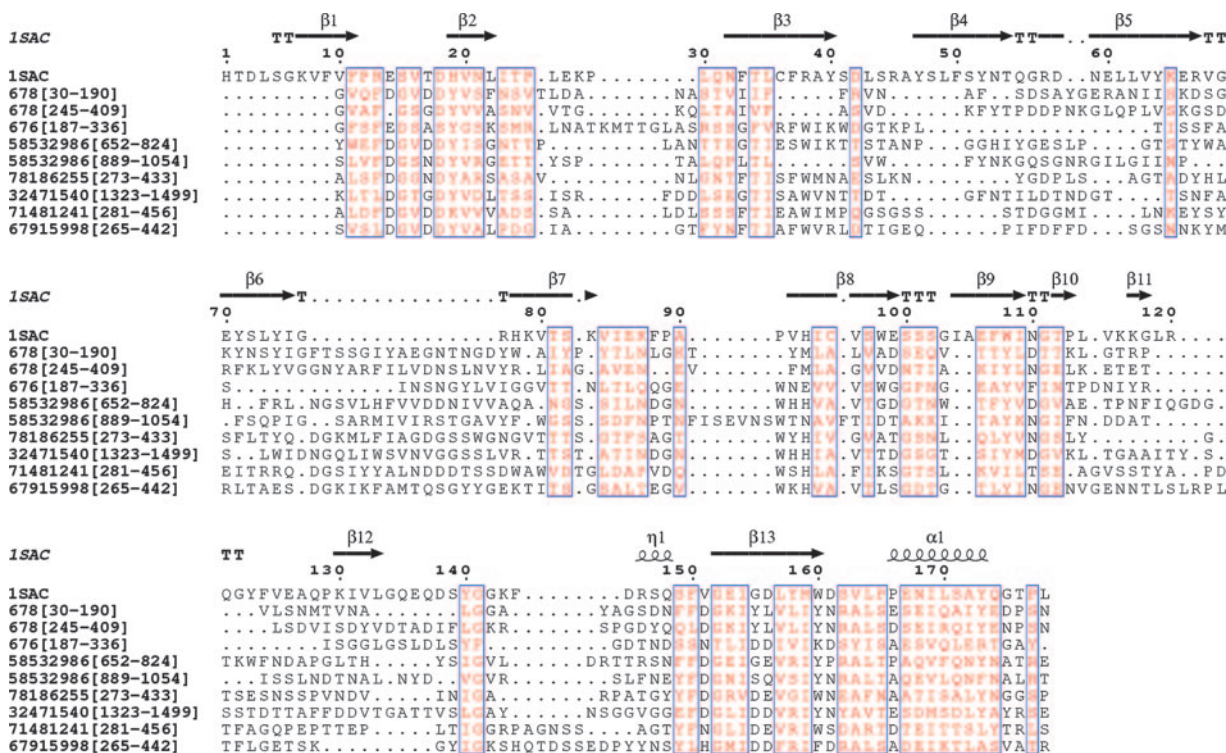


FIG. 3. ORF 676 and ORF 678 possess domains related to laminin G-like jellyroll fold. Amino acid sequence alignment of the internal repeats of ORFs 676 and 678 with various members of the concanavalin A/glucanase structural family (human serum amyloid P component [1sac_A], hypothetical protein from bacteriophage S-PM2 [gi 58532986], VCBS from *Pelodyction luteolum* [gi 78186255], hypothetical protein from *Rhodopirellula baltica* [gi 32471540], VCBS from *Prosthecochloris vibrioformis* [gi 71481241], and S-layer protein from *Clostridium thermocellum* [gi 67915998]) are shown. The numbers in brackets indicate the positions in the sequences. The secondary elements of the human serum amyloid P component are displayed above the alignment.

[13,860 nt], and T6 [470 nt]) were estimated by the identification of potential transcriptional signals upstream and downstream of the putative transcription unit (see below) (Table 1). The largest one, T5, is a polycistronic messenger which covers about 75% of the entire genome and 16 ORFs. Most of these ORFs have been predicted to have membrane-spanning regions, suggesting that this large operon might encode polypeptides inserted into the membrane of the enveloped particle. T1 corresponds to ORF 59 which carries the RHH motif (CopG) and to ORF 52a located just downstream of the putative replication origin. T2 is composed of three ORFs: 87, 52b, and 528, which contains a “winged helix” DNA binding domain. T3 contains only ORF 180a. T4 and T6 are in the opposite direction compared to other transcripts. T4 overlaps ORFs 82 and 62. T6 mRNA corresponds to ORF 153, which is located at the putative origin and contains a leucine zipper motif and “winged helix” DNA binding domain. Sequences resembling the consensus promoter signal of *Pyrococcus* genes (A/G)AAA—T (A/T)(A/T)(A/T)A were found in front of all transcripts, although the two putative promoters corresponding to the transcripts on the complementary strand (T4 and T6) loosely fit the consensus. Most of these putative promoters are centered 35 to 40 nt upstream of the start codon of the first gene in the operon, as previously observed for *Pyrococcus* genes. CT-rich sequence stretches, typical of archaeal

terminators, were also found at the end of most of the transcripts (Table 1).

Experimental identification of the most abundant protein of PAV1. In a previous study, three major proteins (6, 13, and 36 kDa) were observed by SDS-polyacrylamide gel electrophoresis of purified VLPs (20). These proteins were visible by silver staining but in quantities too low to permit identification by matrix-assisted laser desorption ionization–time of flight analysis. Novel preparations were performed and have shown only the 13-kDa band after staining with Coomassie blue (not shown). The 13-kDa band was cut off the gel for microsequencing. Its N-terminal sequence (MMDALEDV) was found to correspond to part of the N-terminal region of the deduced amino acid sequence of ORF 121 of the PAV1 genome. The calculated size of ORF 121, 13.4 kDa, fits well with that of the major protein observed by SDS-polyacrylamide gel electrophoresis. However, the predicted peptide extends 26 amino acid residues upstream of the N terminus determined experimentally on the 13-kDa protein.

DISCUSSION

The genome organization of PAV1 is unusual in that almost all predicted genes are located on the same strand (Fig. 1). Only three small ORFs are found on the opposite strand. This organization differs from that generally reported in other ar-

chaeal viruses (notably in members of the *Fuselloviridae*) except for the linear genome of the crenarchaeal virus PSV, isolated from an anaerobic hyperthermophilic archaeon of the genus *Pyrobaculum*, which encodes all genes except four on one strand (22). Such a "compact" organization, often observed for viral genomes, is also found for PAV1, since 12 ORFs are partially overlapping.

All together, the structure of the largest intergenic regions, the identification of a DUE, and the presence of an ORF similar to an ORF of a member of the CopG family downstream of the DUE supports the hypothesis that a replication origin is located between ORF 153 and ORF 59. Despite the presence of a CopG homologue, a copy number regulator frequently found in RC replicon, and the general trend of the GC skew graph, we showed that PAV1 replication most likely does not progress via the RC mechanism. Moreover, as most of the ORFs are oriented in the same direction, it is tempting to hypothesize that replication is probably unidirectional and proceeds by strand displacement or by a theta mode mechanism.

Half of the ORFs (12 of 25) have been predicted to have transmembrane regions. It is comparable to the content of membrane proteins predicted for the *Sulfolobus shibatae* SSV1 virus, where 11 out of the 32 ORFs are potentially membrane proteins (40). Strikingly, the genomic distributions of the putative membrane proteins are similar for the PAV1 and SSV1 genomes. In both genomes, most of the predicted genes encoding membrane proteins are clustered on one half of the genome. In the case of PAV1, all but two of the predicted genes are part of the larger transcript T5, including ORF 121, encoding the major VLP protein of 13 kDa, which is the first gene of the operon. Such organization suggests that the T5 transcript encodes all the structural proteins of PAV1.

As found in other sequenced genomes of archaeal viruses, a majority of the ORFs in the PAV1 genome do not have a known function. Only six of the predicted ORFs have been assigned a putative function.

A small protein (ORF 59) contains a domain similar to the RHH motif. This domain is found in CopG proteins (21). CopG proteins have been reported to be responsible for the regulation of RC plasmid copy number. It binds to the *cop-rep* promoter and controls synthesis of the plasmid replication initiator protein RepB (1). However, the search of sequence similarities failed to identify any ORFs homologous to the RepB protein in the genome of PAV1. The RHH domain was also found in Arc repressor from *Salmonella* bacteriophage P22 (12) and in the methionine repressor MetJ (52). A recent study showed that the most common gene products in crenarchaeal viruses are small proteins containing the RHH domain (47). The authors of this study wondered about the fact that no RHH domain proteins were detected in the available genomes of euryarchaeal viruses infecting mesophilic or moderately thermophilic hosts, suggesting that these small and compact proteins are particularly proficient for transcription regulation in hyperthermophiles. Our study seems to confirm this hypothesis, since PAV1 was isolated from a hyperthermophilic euryarchaeote.

The genome contains two other proteins with DNA binding domains, ORF 153 and ORF 528. These proteins display both a winged helix DNA binding domain. Many different proteins with diverse biological functions possess this domain, including

hypothetical transcriptional factors, such as ORF 93 of the *Fusellovirus* SSV1 (29). In addition, the domain organization of ORF 153 is similar to the basic region-leucine zipper (b/Zip) family of eukaryotic transcription factors (57). Therefore, it is tempting to hypothesize that the products of these genes could be involved in transcription regulation.

A putative protein of 22.1 kDa (ORF 180a) is 31% identical to ORF 181 of plasmid pRT1 from *Pyrococcus* sp. strain JT (58). It has been shown that ORF 181 is 50% identical to ORF 80 of the *Sulfolobus* plasmid family pRN in a 35-amino-acid region at the C terminus. The ubiquity and high degree of sequence conservation of ORF 80 proteins in *Sulfolobales* plasmids suggest they have an important function, although their precise physiological role remain obscure (31, 32). Yet, it has been shown that these proteins are sequence-specific dsDNA binding proteins and that the basic C-terminal portions of these proteins are involved in DNA binding. The binding sites of ORF 80 members has been defined and consist of two palindromic sequences separated by 65 bp (31). Whereas the canonical TTAAN₇TTAA motif was not identified upstream of the ORF 180a gene, it is worth noting that two 14-bp inverted repeats of TATAACCAAAATTG with about the same spacing (68 bp) were present in the region upstream of the gene (positions 2554 to 2567 and 2625 to 2638), suggesting that this domain on ORF 180a could participate in the binding of this protein to DNA to achieve structural or regulatory functions.

ORF 676 and ORF 678 contain one and two occurrences, respectively, of a LamGL domain that belongs to the structural superfamily of concanavalin A-like lectin/glucanase. In addition, they display similarities to large proteins likely involved in adhesion that also contain several occurrences of this LamGL domain.

The taxonomy report of the BLAST search displays a curious distribution that is similar to the species distribution of the LamGL domain, as already reported in the SMART database (accession number SM00560). Indeed, the significantly similar sequences of ORFs 676 and 678 are found in prokaryotes and metazoans. In addition, the vast majority of the prokaryotic species have been isolated from marine or aquatic environments.

The LamGL domain displays binding activity to complex carbohydrates, either in the context of storage and transport of carbohydrates (50), catalysis of glucans (37), or cell recognition and adhesion (13, 55). In particular, this domain is present, as a pair or in a single module, in numerous extracellular matrix proteins in eukaryotes, where it has been shown to be involved in interaction with extracellular sulfate ligands like heparin (54). In a comparative study of sulfated polysaccharides from marine angiosperms, Aquino et al. (2) suggest that a convergent adaptation due to environmental pressure may explain the occurrence of high concentrations of sulfated polysaccharides in many marine organisms, suggesting that the occurrence of sulfated polysaccharides seems to be an adaptation to marine life. It is thus tempting to hypothesize that the LamGL domain of ORFs 676 and 678 could also be involved in interaction with sulfated carbohydrates in the aquatic environment. In this context, the species distribution of sequences similar to ORFs 676 and 678 could reflect an adaptation of this domain to the binding of sulfated ligands that are abundant both in the marine environment and in the extracellular matrix of metazoans.

Up to now, very little is known concerning the composition and modification of the surface layer proteins from *Thermococcales* species; however, it has been reported that S-layer proteins of hyperthermophilic archaea have more charged residues than their mesophilic counterparts and that S-layer glycoproteins in archaeal halobacteria contain sulfated glucuronic acid residues (10). Altogether, this suggests that ORFs 676 and 678 are exposed to the surface of the enveloped virion and might be implicated in host recognition and attachment via the binding capacity of the LamGL domain to sulfated glycoproteins exposed to the surface of *P. abyssi*.

This hypothesis seems to be corroborated by the fact that ORF 678 presents similarities with ORF 175 of the bacteriophage S-PM2, which also contains two occurrences similar to the concanavalin A-like domain (33). The genome of S-PM2, a T4-type bacteriophage that infects marine photosynthetic bacteria *Synechococcus* spp., was analyzed. This analysis revealed that ORF 175 is part of a cluster potentially involved in recognition and attachment of virus to its host.

The adhesion activity of ORF 676 and ORF 678 and the composition of the S-layer of the host remain to be determined in order to better understand virus-host interactions at the cellular level.

Transcription analysis performed by a RT-PCR method suggests that all predicted genes are actually transcribed in five polycistronic mRNAs (T1 to T5) and a small monocistronic mRNA (T6). The longest transcript, T5, is remarkably long, as it covers about 75% of the genome. As generally observed in other archaeal viral genomes, the PAV1 promoters identified in the upstream sequence of each transcript resemble that of its host, *P. abyssi*, carrying a TATA-like box and a transcription factor B-responsive element. Most of the predicted PAV1 genes harbor a typical Shine-Dalgarno motif (in particular all the genes carried by the largest transcript T5) as previously observed in *Pyrococcus* genomes. However, at this point, several precautions must be taken concerning the interpretation of these results. The method we used (RT-PCR) did not allow estimation of the relative abundance of the different transcripts or detection of shorter transcripts that terminated before the predicted terminator signal or that started from internal promoters. Therefore, we cannot rule out the possibilities that additional transcripts may exist and that the longest version of T5 transcript represents only a minor fraction of the mRNA produced. Classical Northern experiments are commonly used to estimate both the size and relative abundance of major transcripts but are not reliable to detect transcripts of low abundance (7). Indeed, initial experiments to detect PAV1 transcripts by Northern analysis gave very weak and nonreproducible hybridization signals (not shown) probably because of the very low level of PAV1 transcripts under the culture conditions tested. This relative low abundance of PAV1 transcripts may be surprising when compared to the high copy number of "plasmidic" PAV1 DNA detected in the host cells (ca. 60 copies per host genome), but there is no indication that all these copies are actively transcribed at the same time.

Analysis of the protein composition of the PAV1 virion identified a major protein of 13 kDa. Its N-terminal sequence was compared with the whole viral sequence and found to correspond to ORF 121 of PAV1. The N-terminal sequence was not consistent with the theoretical initial methionine which

has been located 26 residues upstream, suggesting either that the ORF 121 gene was not properly annotated or that the protein could be produced as a preprotein further processed after being transported to the membrane. In support of the latter hypothesis, ORF 121 is predicted to have two membrane-spanning regions. In addition, the length of the putative peptide signal and the presence of an alanine at the -1 position are consistent with the features of archaeal signal peptides (5). Furthermore, some viral proteins were shown to be N-terminally processed as in the methanophage ψ M2 (41). Such proteolytic cleavage has already been shown for head proteins (9). In conclusion, we assume that the ORF 121 gene product constitutes the main protein of the PAV1 VLP coat to which it is specifically sent and integrated after N-terminal processing. The ORF 121 protein shows no sequence similarity to VP1 or VP2, the coat proteins of fuselloviruses.

In spite of its lemon shape, which could assign PAV1 to the *Fuselloviridae* family, the genomic properties of PAV1 most likely suggest instead that it represents the first archaeal member of a novel virus genus or family. Indeed, the genome of PAV1 displays unique features at the nucleic and proteinic levels compared to the genomes of archaeal viruses indexed in the public databases. Thorough comparison with other archaeal genomes, in particular, with the lemon-shaped viruses His1 and His2 from *Haloferax* and spindle-shaped viruses from *Sulfolobus* did not reveal any sequence similarities to the PAV1 genome. The uniqueness of PAV1 is undoubtedly a consequence of its individual evolutionary history. PAV1 was isolated from a deep-sea hydrothermal vent, whereas fuselloviruses commonly inhabit acidic hot terrestrial springs. Consequently, the unique features of the PAV1 genome could be the result of its geographic isolation and an adaptation to the particular features of the hydrothermal environments. Indeed, given the extreme nature of hydrothermal environments, namely, high temperature, high hydrostatic pressure, and the specific microbial diversity (mainly methanogens, sulfato reducers, and sulfur reducers) in deep-sea vents, we hypothesized that there would be a severe barrier to gene flow from organisms, reflecting a long evolution of host and virus in a relatively closed gene pool.

ACKNOWLEDGMENTS

We thank P. Forterre for critical reading of the manuscript and for helpful suggestions and Guest-Genopole for access to sequence analysis tools.

REFERENCES

1. Acebo, P., M. Garcia de Lacoba, G. Rivas, J. M. Andreu, M. Espinosa, and G. del Solar. 1998. Structural features of the plasmid pMV158-encoded transcriptional repressor CopG, a protein sharing similarities with both helix-turn-helix and beta-sheet DNA binding proteins. *Proteins* 32:248-261.
2. Aquino, R. S., A. M. Landeira-Fernandez, A. P. Valente, L. R. Andrade, and P. A. Mourao. 2005. Occurrence of sulfated galactans in marine angiosperms: evolutionary implications. *Glycobiology* 15:11-20.
3. Arnold, H. P., U. Ziese, and W. Zillig. 2000. SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* 272:409-416.
4. Arnold, H. P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J. F. Weidmann, J. K. Kristjansson, 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.
5. 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.
6. Bath, C., T. Cukalac, K. Porter, and M. L. Dyall-Smith. 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* 350:228-239.

7. Berkner, S., and G. Lipps. 2007. Characterization of the transcriptional activity of the cryptic plasmid pRN1 from *Sulfolobus islandicus* REN1H1 and regulation of its replication operon. *J. Bacteriol.* **189**:1711–1721.
8. Betstetter, 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.
9. Black, L. W. 1989. DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43**:267–292.
10. Claus, H., E. Akca, T. Debaerdemaeker, C. Evrard, J. Declercq, J. Harris, B. Schlott, and H. König. 2005. Molecular organization of selected prokaryotic S-layer proteins. *Can. J. Microbiol.* **51**:731–743.
11. Cohen, G. N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Querellou, R. Ripp, J. C. Thierry, J. Van der Oost, J. Weissenbach, Y. Zivanovic, and P. Forterre. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* **47**:1495–1512.
12. Cordes, M. H., and R. T. Sauer. 1999. Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Sci.* **8**:318–325.
13. Crennell, S., E. Garman, G. Laver, E. Vimr, and G. Taylor. 1994. Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* **2**:535–544.
14. Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
15. Dyall-Smith, M., S. L. Tang, and C. Bath. 2003. Haloarchaeal viruses: how diverse are they? *Res. Microbiol.* **154**:309–313.
16. Eddy, S. R. 1998. Profile hidden Markov models. *Bioinformatics* **14**:755–763.
17. Erauso, G., S. Marsin, N. Benbouzid-Rollet, M. F. Baucher, T. Barbeyron, Y. Zivanovic, D. Prieur, and P. Forterre. 1996. Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyssi*: evidence for rolling-circle replication in a hyperthermophile. *J. Bacteriol.* **178**:3232–3237.
18. Erauso, G., A. L. Reysenbach, A. Godfroy, J.-R. Meunier, B. Crump, F. Partensky, J. A. Baross, V. Marteinson, G. Barbier, N. R. Pace, and D. Prieur. 1993. *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.* **160**:338–349.
19. Geslin, C., M. Le Romancer, M. Gaillard, G. Erauso, and D. Prieur. 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. *Res. Microbiol.* **154**:303–307.
20. Geslin, C., M. Le Romancer, G. Erauso, M. Gaillard, G. Perrot, and D. Prieur. 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, “*Pyrococcus abyssi*.” *J. Bacteriol.* **185**:3888–3894.
21. Gomis-Ruth, F. X., M. Sola, R. Perez-Luque, P. Acebo, M. T. Alda, A. Gonzalez, M. Espinosa, G. del Solar, and M. Coll. 1998. Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the pMV158-encoded plasmid transcriptional repressor protein CopG. *FEBS Lett.* **425**:161–165.
22. Haring, M., X. Peng, K. Brugger, 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.
23. Haring, M., R. Rachel, X. Peng, R. A. Garrett, and D. Prangishvili. 2005. Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, *Acidianus* bottle-shaped virus, from a new family, the *Ampullaviridae*. *J. Virol.* **79**:9904–9911.
24. Haring, M., G. Vestergaard, K. Brügger, R. Rachel, R. A. Garrett, and D. Prangishvili. 2005. Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. *J. Bacteriol.* **187**:3855–3858.
25. Haring, M., G. Vestergaard, R. Rachel, L. Chen, R. A. Garrett, and D. Prangishvili. 2005. Virology: independent virus development outside a host. *Nature* **436**:1101–1102.
26. Hofmann, K., and W. Stoffel. 1993. TMbase—a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**:166.
27. Huang, Y., and D. Kowalski. 2003. WEB-THERMODYN: sequence analysis software for profiling DNA helical stability. *Nucleic Acids Res.* **31**:3819–3821.
28. Janekovic, D., S. Wunderl, I. Holz, W. Zillig, A. Gierl, and H. Neumann. 1983. TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaeobacterium *Thermoproteus tenax*. *Mol. Gen. Genet.* **192**:39–45.
29. Kraft, P., A. Oeckinghaus, D. Kummel, G. Gauss, J. Gilmore, B. Wiedenheft, M. Young, and C. Lawrence. 2004. Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. *J. Virol.* **78**:11544–11550.
30. Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**:567–580.
31. Lipps, G. 2006. Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*. *Extremophiles* **10**:17–28.
32. Lipps, G., P. Ibanez, T. Stroessenreuther, K. Hekimian, and G. Krauss. 2001. The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. *Nucleic Acids Res.* **29**:4973–4982.
33. Mann, N., M. Clokie, A. Millard, A. Cook, W. Wilson, P. Wheatley, A. Letarov, and H. Krisch. 2005. The genome of S-PM2, a “photosynthetic” T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* **187**:3188–3200.
34. Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher, and W. Zillig. 1984. SAV1, a temperature u. v.-inducible DNA virus-like particle from the archaeobacterium *Sulfolobus acidocaldarius*. *EMBO J.* **3**:2165–2168.
35. Matsunaga, F., C. Norais, P. Forterre, and H. Myllykallio. 2003. Identification of short ‘eukaryotic’ Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep.* **4**:154–158.
36. McGuffin, L. J., and D. T. Jones. 2003. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**:874–881.
37. Michel, G., L. Chantalat, E. Duee, T. Barbeyron, B. Henrissat, B. Kloareg, and O. Dideberg. 2001. The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. *Structure* **9**:513–525.
38. Oren, A., G. Bratbak, and M. Haldal. 1997. Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**:143–149.
39. Ortman, A. C., B. Wiedenheft, T. Douglas, and M. Young. 2006. Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat. Rev. Microbiol.* **4**:520–528.
40. Palm, P., C. Schleper, B. Grapp, 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.
41. Pfister, P., A. Wasserfallen, R. Stettler, and T. Leisinger. 1998. Molecular analysis of *Methanobacterium* phage psiM2. *Mol. Microbiol.* **30**:233–244.
42. Porter, K., P. Kukkaro, J. K. Bamford, C. Bath, H. M. Kivela, M. L. Dyall-Smith, and D. H. Bamford. 2005. SH1: a novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology* **335**:22–33.
43. Prangishvili, D., H. P. Arnold, D. Götz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig. 1999. A novel virus family, the Rudiviridae: structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. *Genetics* **152**:1387–1396.
44. Prangishvili, D., P. Forterre, and R. A. Garrett. 2006. Viruses of the Archaea: a unifying view. *Nat. Rev. Microbiol.* **4**:837–848.
45. Prangishvili, D., and R. A. Garrett. 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem. Soc. Trans.* **32**:204–208.
46. Prangishvili, D., and R. A. Garrett. 2005. Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol.* **13**:535–542.
47. Prangishvili, D., R. A. Garrett, and E. V. Koonin. 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.* **117**:52–67.
48. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning, a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
49. Sambrook, J., and D. W. Russell. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
50. Sanz-Aparicio, J., J. Hermoso, T. Grangeiro, J. Calvete, and B. Cavada. 1997. The crystal structure of Canavalia brasiliensis lectin suggests a correlation between its quaternary conformation and its distinct biological properties from Concanavalin A. *FEBS Lett.* **405**:114–118.
51. Snyder, J., K. Stedman, G. Rice, B. Wiedenheft, J. Spuhler, and M. Young. 2003. Viruses of hyperthermophilic Archaea. *Res. Microbiol.* **154**:474–482.
52. Somers, W. S., and S. E. Phillips. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by beta-strands. *Nature* **359**:387–393.
53. Suzek, B. E., M. D. Ermolaeva, M. Schreiber, and S. Salzberg. 2001. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* **17**:1123–1130.
54. Timpl, R., D. Tisi, J. F. Talts, Z. Andac, T. Sasaki, and E. Hohenester. 2000. Structure and function of laminin LG modules. *Matrix Biol.* **19**:309–317.
55. Tisi, D., J. F. Talts, R. Timpl, and E. Hohenester. 2000. Structure of the C-terminal laminin G-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin. *EMBO J.* **19**:1432–1440.
56. Vestergaard, G., M. Haring, X. Peng, R. Rachel, R. A. Garrett, and D. Prangishvili. 2005. A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. *Virology* **336**:83–92.
57. Vinson, C., M. Myakishev, A. Acharya, A. Mir, J. Moll, and M. Bonovich. 2002. Classification of human B-ZIP proteins based on dimerization properties. *Mol. Cell. Biol.* **22**:6321–6335.

58. Ward, D. E., I. M. Revet, R. Nandakumar, J. H. Tuttle, W. M. de Vos, J. van der Oost, and J. DiRuggiero. 2002. Characterization of plasmid pRT1 from *Pyrococcus* sp. strain JT1. *J. Bacteriol.* **184**:2561–2566.
59. 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.
60. Witte, A., U. Baranyi, R. Klein, M. Sulzner, C. Luo, G. Wanner, D. H. Kruger, and W. Lubitz. 1997. Characterization of *Natronobacterium magadii* phage Phi Ch1, a unique archaeal phage containing DNA and RNA. *Mol. Microbiol.* **23**:603–616.
61. Wood, A. G., W. B. Whitman, and J. Konisky. 1989. Isolation and characterization of an archaeobacterial viruslike particle from *Methanococcus voltae* A3. *J. Bacteriol.* **171**:93–98.