Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments

Fengbin Wang[a,1], Diana P. Baquero[b,c,1], Leticia C. Beltran[a], Zhangli Su[a], Tomasz Osinski[a], Weili Zheng[a], David Prangishvili[b,d], Mart Krupovic[b,2], and Edward H. Egelman[a,2]

[a]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908; [b]Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, 75015 Paris, France; [c]Collège Doctoral, Sorbonne Universités, 75005 Paris, France; and [d]Academia Europaea Tbilisi Regional Knowledge Hub, Ivane Javakhishvili Tbilisi State University, 0179 Tbilisi, Georgia

Living organisms expend metabolic energy to repair and maintain their genomes, while viruses protect their genetic material by completely passive means. We have used cryo-electron microscopy (cryo-EM) to solve the atomic structures of two filamentous double-stranded DNA viruses that infect archaeal hosts living in nearly boiling acid: *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1), at 2.8-Å resolution, and *Sulfolobus islandicus* filamentous virus (SIFV), at 4.0-Å resolution. The SIFV nucleocapsid is formed by a heterodimer of two homologous proteins and is membrane enveloped, while SSRV1 has a nucleocapsid formed by a homodimer and is not enveloped. In both, the capsid proteins wrap around the DNA and maintain it in an A-form. We suggest that the A-form is due to both a nonspecific desolvation of the DNA by the protein, and a specific coordination of the DNA phosphate groups by positively charged residues. We extend these observations by comparisons with four other archaeal filamentous viruses whose structures we have previously determined, and show that all 10 capsid proteins (from four heterodimers and two homodimers) have obvious structural homology while sequence similarity can be nonexistent. This arises from most capsid residues not being under any strong selective pressure. The inability to detect homology at the sequence level arises from the sampling of viruses in this part of the biosphere being extremely sparse. Comparative structural and genomic analyses suggest that nonenveloped archaeal viruses have evolved from enveloped viruses by shedding the membrane, indicating that this trait may be relatively easily lost during virus evolution.

cryo-EM | extremophiles | hyperthermophilic archaea | filamentous viruses

Viruses may constitute the largest source of genetic diversity on Earth (1). They greatly impact human health but also have large and direct impacts on the ecosystem of the planet (2). Random sampling of ocean and lake water led to the surprising observation that the concentration of viruses was on the order of 10^8 per milliliter in these aquatic environments (3), and current estimates for the number of viruses in the Earth's oceans are on the order of 10^{30} (2), with estimates of 10^{31} viruses on the Earth in total (4), compared with $\sim 10^{23}$ stars in the universe. Studying viruses has thus been of great interest in many areas of biology and has had an enormous impact on structural biology due to the relative simplicity of virus particles compared to any cellular organisms (5).

Many pathways for DNA repair exist in all living organisms, with some conserved from bacteria to humans (6). These mechanisms are essential to preserve the integrity of genomes due to both exogenous as well as endogenous sources of DNA damage (7). In contrast, when outside the host, viruses must depend upon entirely passive means to protect their genomes using a very limited number of different structural proteins. In

some instances, there may only be a single protein (present in many copies) in a virion (8). Viruses that infect archaea have been of particular interest (9–12), as many of these have been found in the most extreme aquatic environments on earth: Temperatures of 80 to 90 °C with pH values of ~2 to 3. How DNA genomes can be passively maintained in such conditions is of interest not only in terms of evolutionary biology and the origin of life on Earth (1, 13), but has potential consequences for everything from biotechnology (14) to human therapeutics (15).

We have previously used cryo-electron microscopy (cryo-EM) to determine the atomic structures of four archaeal filamentous viruses that infect hyperthermophiles (16–19). One of these, SIRV2, had a protein coat composed of a homodimer of the major capsid protein (MCP) (19), while AFV1 (18), SFV1 (17), and PFV2 (16) had protein coats composed of heterodimers. AFV1, SFV1, and PFV2 are membrane enveloped, while SIRV2 is not. These viruses are classified into three different families, *Rudiviridae* (SIRV2), *Lipothrixviridae* (AFV1 and SFV1), and *Tristromaviridae* (PFV2), with families *Rudiviridae* and *Lipothrixviridae*

unified into the order *Ligamenvirales* (20). Furthermore, a class "*Tokiviricetes*" has been proposed for unification of the three families of structurally related viruses (16). Here, we extend these observations by solving the structure of the nonenveloped *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1) (21) at 2.8-Å resolution, the highest resolution yet achieved for these filamentous viruses, and solving the structure of the membrane-enveloped *Sulfolobus islandicus* filamentous virus (SIFV) (22) at 4.0-Å resolution. With six structures from this group of filamentous viruses, we can make a number of comparisons providing insights into what has been conserved and what has diverged. Although the sequence similarity of the protein subunits within this virus assemblage can be undetectable (17), all share a relatively conserved fold that encapsidates A-form DNA, demonstrating common ancestry. The high resolution of the SSRV1 structure allows us to look with greater confidence at how the A-form is maintained in the virion. While it has been reasonable to assume that enveloped and nonenveloped viruses must have diverged early due to entirely different mechanisms of entry and egress from cells (23), we can show striking similarities between enveloped and nonenveloped archaeal filamentous viruses. It is also clear that other families of archaeal filamentous viruses infecting hyperthermophiles are likely to exist, as the structure of APBV1 (24) shows no apparent homology with the six viruses we now describe.

Results

Two Major Capsid Protein Homologs in SSRV1.

All members of the proposed genus "*Hoswirudivirus*," including SSRV1, encode two proteins homologous to the MCP of SIRV2 (21). The two proteins, SSRV1 gp10 (108 aa; QJF12286) and gp11 (134 aa; QJF12287), are encoded next to each other and can be aligned in their C-terminal regions (22% identity over 74 aa; E = 2e-09), whereas the characteristic N-terminal region is missing in gp10. To investigate whether the virions of rudiviruses in the genus "*Hoswirudivirus*" are constructed from a heterodimer, similar to lipothrixviruses and tristromaviruses, or from a homodimer, like in all other known rudiviruses, SSRV1 virions were purified and subjected to SDS/PAGE (*SI Appendix*, Fig. S1) and mass spectrometry analyses, which showed that only gp11 protein is present in the virions. This result indicates that, if expressed, gp10 is not incorporated into the SSRV1 virions at detectable levels.

Cryo-EM of SSRV1.

From cryo-EM images of SSRV1 (Fig. 1*A*), 470,216 overlapping segments were extracted and used for a three-dimensional (3D) reconstruction. The possible helical symmetries were determined from an averaged power spectrum (Fig. 1*B*), which showed multiple orders of the 1-start helix ($n = 1, 2, 3, 4...$) as well as additional layer lines. Indexing the additional layer lines was done by trial and error, and only one symmetry (~14.67 subunits per turn of the 43.1-Å pitch 1-start helix) yielded a reconstruction with interpretable features,
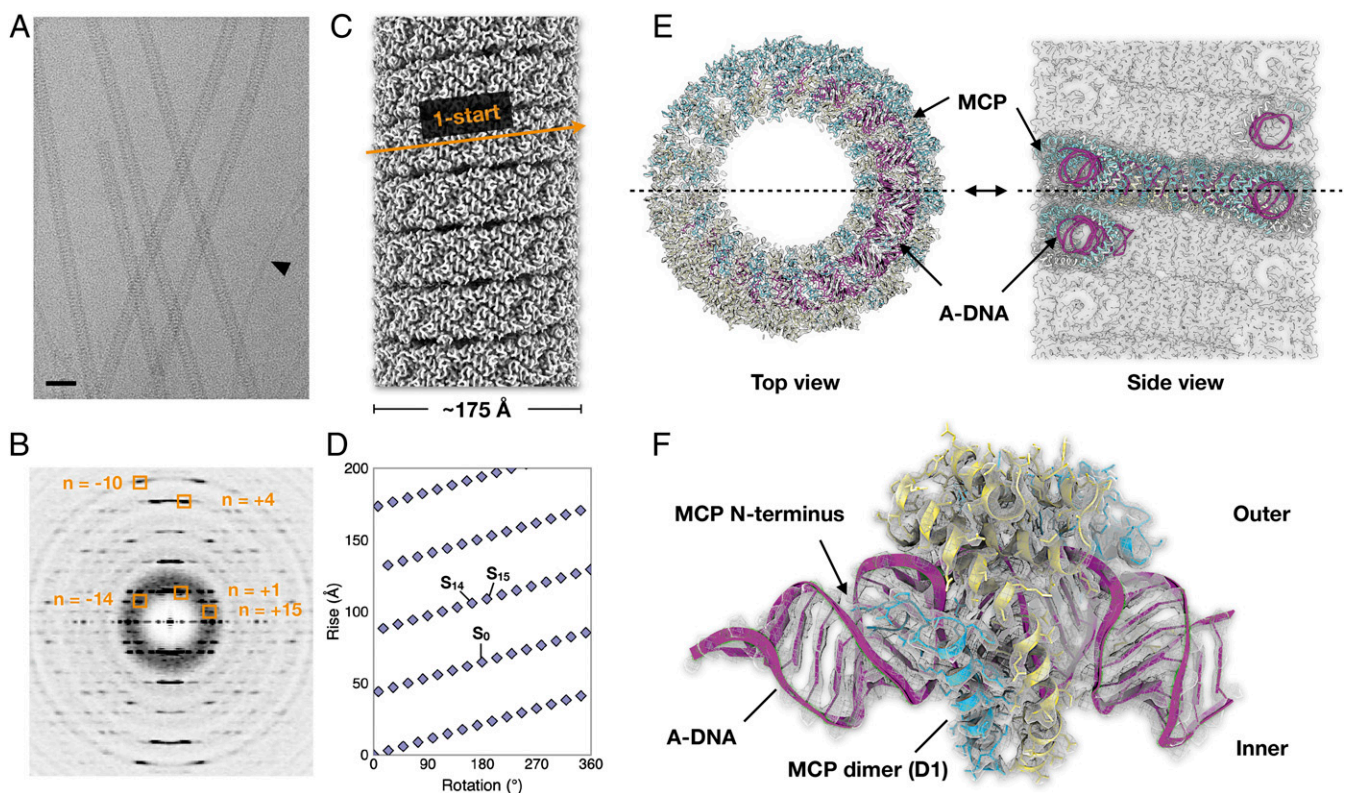

Fig. 1. Cryo-EM of the *S. solfataricus* rod-shaped virus (SSRV1). (*A*) Representative cryo-electron micrograph of SSRV1 virions. A *Saccharolobus solfataricus* Type IV pilus, indicated by black arrowhead, is the cellular receptor for SSRV1. (Scale bar, 50 nm.) (*B*) Averaged power spectrum of the segments used in the IHRSR reconstruction. The layer lines that were used to initially estimate the helical symmetry are indicated, along with the Bessel orders for the correct symmetry found. (*C*) Surface of the SSRV1 cryo-EM reconstruction at 2.8-Å resolution. The right-handed ~43-Å pitch 1-start helix passing through every asymmetric unit is shown. (*D*) Helical net of SSRV1, using the convention that the surface is unrolled and we are viewing it from the outside. Since there are 14.67 subunits per turn of the right-handed 1-start helix, subunits S_{14} and S_{15} will be above the reference subunit S_0 indicated. (*E*) Top and side views of the SSRV1 atomic model fit into the EM map. The side view is from a central slice of the map/model, indicated by the dashed line. A-form DNA is colored magenta, and MCP dimers are cyan and yellow. (*F*) Ribbon models for the SSRV1 protein dimer and 36 bp of A-DNA, fit into the EM map. The clear separation of DNA base pairs is seen in the map, despite averaging over the entire genomic sequence with the imposition of helical symmetry.

www.manaraa.com

including side chains (Fig. 1 C–F). It was apparent that the asymmetric unit contained a symmetrical dimer of coat proteins wrapped around 12 bp of A-form DNA, so a dihedral symmetry was imposed on the overall reconstruction. This dihedral symmetry relates the 5′-to-3′ strand of the DNA to the 3′-to-5′ strand. Although the helical symmetry imposed will average together 12 bp of DNA throughout the genome, the separation of these base pairs is still preserved very well in the reconstructed volume (Fig. 1F) even though all sequence information has been lost. The DNA supercoils in the virion with one right-handed supercoil for every 43.1-Å turn of the protein helix (Fig. 1 C and E). Since there are 14.67 asymmetric units per turn of this 1-start helix, there are 44 asymmetric units in three turns containing 528 bp. The native twist of the DNA is therefore 528 bp in 47 turns (44 local turns plus three right-handed supercoil turns), which is 11.2 bp/turn, similar to what was found for SIRV2 (19). The helical radius of the DNA solenoidal supercoil is ~60 Å.

Since the EM reconstruction reached 2.8-Å resolution (*SI Appendix*, Figs. S2 and S3), it was possible to do an unambiguous C_α trace through the density map resulting in a chain containing 131 residues, confirming that the 108-aa-long gp10 does not function as an MCP. The MCP of SSRV1, gp11, shares 80% sequence identity with the MCP of SIRV2. A SIRV2-based homology model was then fit and refined into the map (Table 1 and Fig. 1 E and F). The overall fold of the SSRV1 model is very similar with that of SIRV2, but with a more accurate model due to the much higher resolution (2.8 versus 4 Å).

Cryo-EM of SIFV. The membrane-enveloped SIFV had previously been demembranated, imaged by negative staining and reconstructed, but no high-resolution information was obtained (22). Here, we reexamined SIFV using cryo-EM, and it was reconstructed by a similar approach as described above. From cryo-electron micrographs (Fig. 2A), 167,541 overlapping segments were extracted and processed. The averaged power spectrum

(Fig. 2B) showed that the pitch of the 1-start helix was 51.3 Å, significantly larger than the 43.1-Å pitch of SSRV1. Again, the symmetry was determined by trial and error, and the correct symmetry was 9.35 subunits per turn of this 51.3-Å pitch helix (Fig. 2 C–E). In contrast to the symmetrical protein homodimer in SSRV1, there was a pseudosymmetrical heterodimer formed by MCP1 and MCP2 in SIFV (Figs. 1F, 2F, and *SI Appendix*, Fig. S2), and as a consequence the virions lacked the overall dihedral symmetry in SSRV1 and were polar (Fig. 2F). The final EM map had ~4.0-Å resolution (*SI Appendix*, Figs. S2 and S3), and an unambiguous C_α trace was made through the density map for both MCPs within one asymmetric unit. Each asymmetric unit also contained 12 bp of A-form DNA. The DNA supercoils in SIFV with one right-handed supercoil for every 51.3-Å turn of the protein helix (Fig. 2 C and E). Since there are 9.35 asymmetric units per turn of this 1-start helix, there are ~28 asymmetric units in three turns containing 336 bp. The native twist of the DNA is therefore 336 bp in 31 turns (28 local turns plus three right-handed supercoil turns), which is 10.8 bp/turn, similar to what was found for AFV1 (18). The values of 10.8 for SIFV and 11.2 for SSRV1 bracket 11.0 bp per turn, the canonical value for A-form DNA (25). The helical radius of the DNA solenoidal supercoil is ~40 Å.

The membrane can occasionally be seen to be absent in SIFV (Fig. 2A, white arrowhead), and we suggest that this is an artifact of cryo-EM sample preparation, which results from the fluid flow and large shear forces that can be present just prior to vitrification (26). Virions lacking the membrane appear to be more flexible, as was also observed for AFV1 (18). Since the membrane is fluid, we suggest that the greater rigidity of the virions that are enveloped is not due to the rigidity of the membrane, per se, but rather that the presence of the membrane constrains the protein. Although helical symmetry has been imposed on the membrane during the 3D reconstruction, no features appeared in the membrane that might be due to perturbations of the membrane by the protein subunits, or integral membrane proteins that were aligned with the capsid proteins. However, missing residues 122 to 132 appear to form a loop that extends out from the nucleoprotein core to the membrane (Fig. 2G), which has not been seen in the three other membrane-enveloped viruses: AFV1, SFV2, and PFV2. This loop does not appear to generate ordered density in the membrane at this location, and the resolution of the density connecting the nucleoprotein capsid to the membrane is too poor to allow for an unambiguous chain trace in this region. We therefore cannot say whether this loop inserts into the membrane, interacts with the membrane headgroups, or is associated with an integral membrane protein that is disordered. The membrane thus appears in all respects as a fluid, and the radial density profile of the membrane (Fig. 2C) is shown after cylindrical averaging and low-pass filtering. As with AFV1 (18), the membrane is anomalously thin (~20 Å).

Protein–DNA Interactions. Similar to what was previously observed for SFV1 (17), it is clear from the EM density that the MCPs in both SSRV1 and SIFV contain N-terminal portions that project into a DNA groove (Fig. 3A). While some of the residues in these regions are disordered, we can clearly see the side-chain densities for a few polar N-terminal residues in SSRV1, such as Lys3, Arg5, and Arg8, as they are making contacts with the DNA phosphate backbone (Fig. 3 A and B). Such side-chain densities are not observed in SIFV, even though both MCPs of SIFV are also arginine and lysine rich in their N terminus (Fig. 3A). This is presumably due to greater flexibility and the resulting lower resolution in that region. Similarly, since SIRV2 is structurally very similar to SSRV1 and AFV1 is similar to SIFV, it is very likely their MCPs also insert into a DNA groove, but this was not observed previously due to limited resolution. Interestingly, in the tristromavirus PFV2, a virus evolutionarily divergent from

Table 1. Cryo-EM and refinement statistics of the SSRV1 and SIFV filaments

| Parameter	SSRV1	SIFV
Data collection and processing		
Voltage, kV	300	300
Electron exposure, e⁻·Å⁻²	50	44
Pixel size, Å	1.08	1.4
Particle images, n	470,216	167,541
Helical symmetry		
Point group	D1	C1
Helical rise, Å	2.94	5.48
Helical twist, °	24.53	38.49
Map resolution, Å		
Map:map FSC (0.143)	2.7	3.9
Model:map FSC (0.38)	2.8	4.0
d₉₉	2.8	4.3
Refinement and model validation		
Map-sharpening B-factor, Å²	−122	−185
Bond lengths rmsd, Å	0.007	0.005
Bond angles rmsd, °	0.692	0.704
Clashscore	3.71	18.8
Poor rotamers, %	0	0
Ramachandran favored, %	98.4	91.9
Ramachandran outlier, %	0	0
MolProbity score	1.16	2.26
Deposition ID		
PDB (model)	6WQ0	6WQ2
EMDB (map)	EMD-21867	EMD-21868

Wang et al.
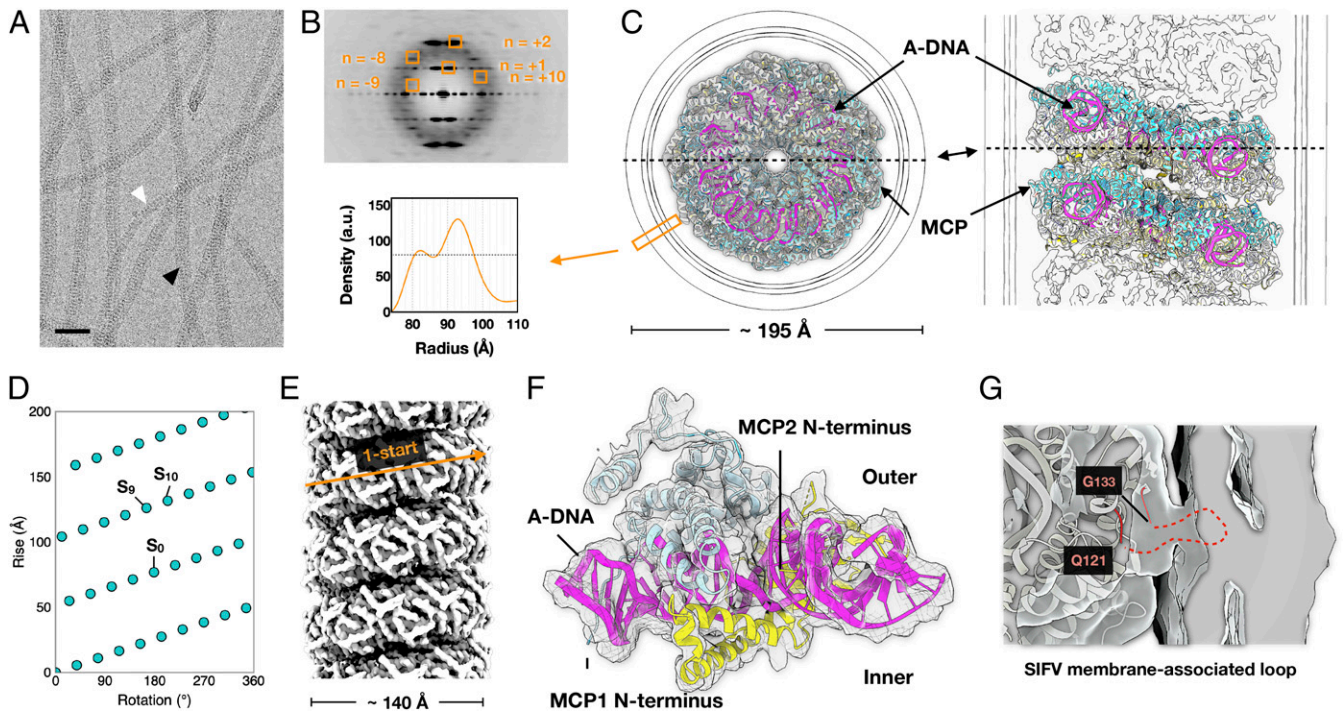
Fig. 2. Cryo-EM of the *S. islandicus* filamentous virus (SIFV). (*A*) Representative cryo-electron micrograph of SIFV. Most of the virions are enveloped by a membrane, indicated by a black arrowhead. Some virions have lost their membrane, and one is indicated by a white arrowhead. This results in a narrower diameter, and these virions appear to have an increased flexibility. (Scale bar, 50 nm.) (*B*) Averaged power spectrum of the segments used in the IHRSR reconstruction. The layer lines that were used to initially estimate the helical symmetry are indicated, along with the Bessel orders for the correct symmetry found. (*C*) Top and side views of the SIFV atomic model fit into the EM map. The side view is from a central slice of the map/model, indicated by the dashed line. A-form DNA is colored magenta, and MCP dimers are in blue and yellow. The membrane has been filtered to 7-Å resolution and cylindrically averaged in this figure. The radial density profile of the cylindrically averaged and filtered membrane is shown. (*D*) Helical net of SIFV, using the convention that the surface is unrolled and we are viewing from the outside. Since there are 9.35 subunits per turn of the right-handed 1-start helix, subunits S_9 and S_{10} will be above the reference subunit S_0 indicated. (*E*) The surface of the 3D reconstruction of SIFV at 4.0-Å resolution, with the membrane removed. (*F*) Ribbon models of the SIFV MCP dimer (cyan and yellow) and 36 bp of A-DNA (magenta), fit into the EM map. (*G*) An 11-aa loop (residues 122 to 132) of SIFV-1 associated with the membrane. An unambiguous backbone trace in this region was not possible due to the low resolution resulting from structural flexibility. The membrane shown here is not cylindrically averaged.

members of the *Rudiviridae* and *Lipothrixviridae* but sharing a similar fold for the MCPs, such N-terminal projections into DNA were not observed at a fairly high resolution of 3.4 Å (16).

A characteristic feature of filamentous archaeal viruses is that their genomic double-stranded DNA (dsDNA) is complexed in A-form (*SI Appendix*, Fig. S4), which is considered to be more resilient to various forms of stress, compared to the physiologically more common B-form (16–19). In both SSRV1 (Fig. 3*B*) and SIFV (Fig. 3*C*), there is an extensive coordination of the phosphate groups in the DNA backbone by positively charged protein sidechains, primarily arginines and lysines. This provides a high degree of specificity in maintaining the DNA in an A-form conformation. In contrast, a nonspecific aspect of the protein–DNA interactions may be a desolvation of the DNA by the extensive wrapping of the protein around the DNA. For SSRV1, we have a resolution (2.8 Å) where one begins to see highly ordered bound water molecules (Fig. 3*D*). There are ~20 ordered water molecules seen around 12-bp A-DNA within the asymmetric unit of SSRV1, with most of them coordinating the phosphate groups. The phosphate groups in the SSRV1 DNA are therefore presumably making hydrogen bonds with either these waters or adjacent protein side chains. In a 0.83-Å resolution crystal structure of A-DNA, most of the bound waters were found surrounding the phosphate groups (27). The water molecules forming a "spine of hydration" in the minor groove of B-DNA (27–29), hypothesized to stabilize B-DNA, do not exist in the SSRV1 structure.

Flexibility of the Virions.
It is clear from looking at cryo-EM images that some of the archaeal filamentous viruses are more rigid than others. A number of formalisms exist for deducing the flexural rigidity, *a*, of an object from the thermodynamic fluctuations in curvature that occur in solution. These can be described in terms of a persistence length, *p*, where $p = a/kT$. The persistence length defines a characteristic distance over which the orientation of a filament persists. It does not mean that a filament can be approximated as a rigid rod over such a length. When the length of a filament, L, is much smaller than *p*, such a rigid rod approximation is reasonable. When L >> *p*, one has a random coil and there is no correlation between the orientation of segments separated by a distance L. We have used cryo-EM images to estimate *p* for the six filamentous viruses, but there is an important caveat that must be made. The formalism being used assumes thermodynamic equilibrium, which means in the absence of forces and fluxes. However, we know that filaments being prepared for vitrification experience large fluid flow forces, and compression from the thinning film (26, 30), and thus are being captured far from thermodynamic equilibrium. Nevertheless, we show estimates of *p* in Fig. 4*A* where the relative (as opposed to absolute) rigidity may be inferred. The strongest correlation is between the nucleocapsid diameter and the persistence length, which is consistent with the fact that for a homogeneous rod the rigidity will scale as the fourth power of the diameter. However, it can be seen that SIFV is significantly more flexible than AFV1, even though both have about the same nucleocapsid diameters. This is due to the significantly greater pitch
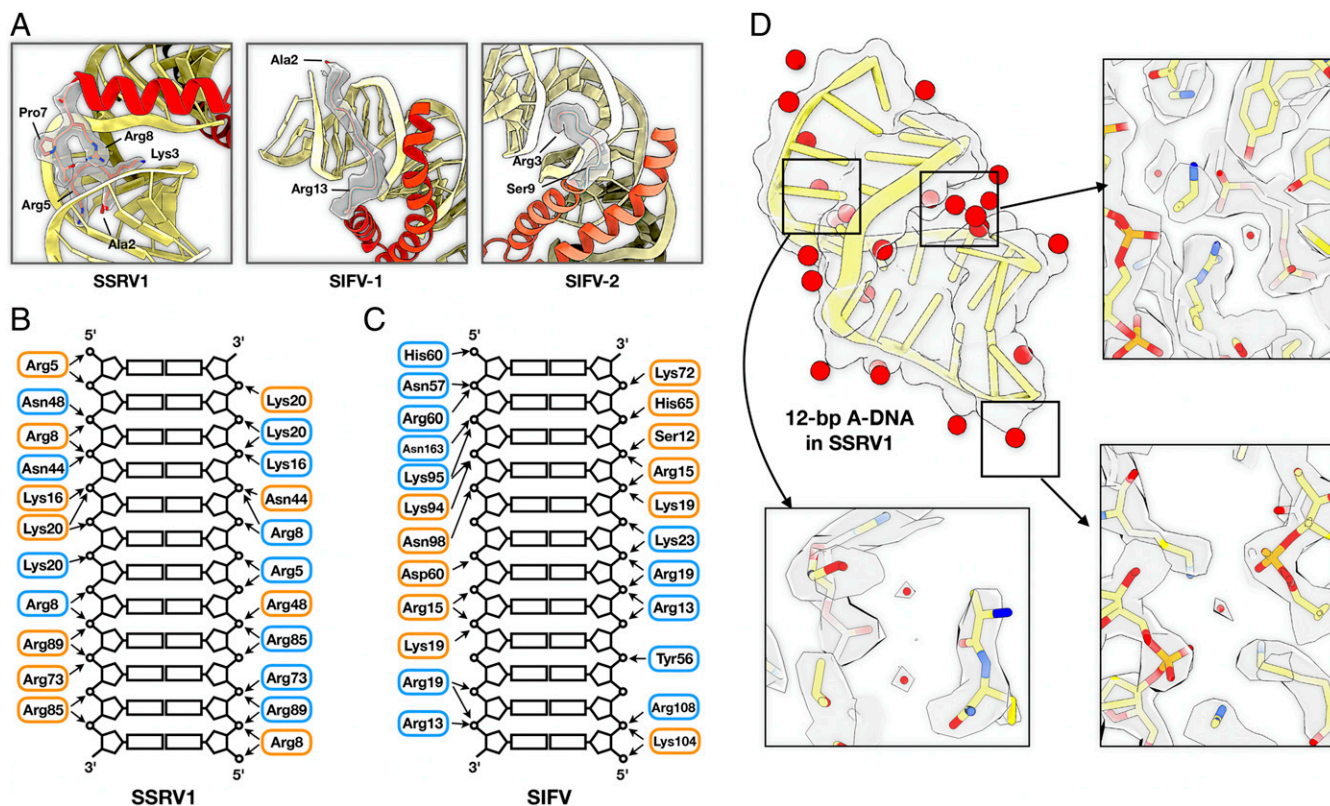
Wang et al.

www.manaraa.com

Fig. 3. Extensive MCP–DNA interactions in SSRV1 and SIFV. (*A*) N-terminal residues of SSRV1 MCP, SIFV MCP1 (SIFV-1), and SIFV MCP2 (SIFV-2) inserting into a groove of the A-DNA. The corresponding EM density in this region is shown. For SIFV, these residues are poorly ordered and therefore only a backbone trace is shown. Protein is red, and DNA is yellow. (*B*) Schematic indicating all of the polar protein–DNA contacts in SSRV1. The MCPs form a symmetrical dimer, so residues colored blue are related by the dihedral symmetry to those colored orange. (*C*) Schematic indicating all of the polar protein–DNA contacts in SIFV. Residues from SIFV-1 are colored blue, and those from SIFV-2 are colored orange. (*D*) Water molecules around A-DNA. The 12-bp A-DNA (wrapped by the SSRV1 homodimer) is shown in a yellow ribbon representation, and the water molecules are shown as red spheres. Several close-up views of water and nearby amino acids are shown with the EM density map.

in SIFV (51 vs. 43 Å), which allows more compression of the helical groove during bending (Fig. 4 *B* and *C*). Consistent with this, we expect that the greater flexibility of SFV1 compared to SIRV2 and SSRV1, despite its larger diameter, is due to a larger helical pitch. As with the flexible filamentous plant viruses (31), the SIFV structure is held together by a flexible linker in the protein, which crosses the helical groove and makes contacts with subunits in the next turn of the helix.

Another important factor that contributes to the filament rigidity is the interface between the MCPs. To investigate this, we did PISA (Protein Interfaces, Surfaces, and Assemblies) analysis (32) on all six archaeal virions on three types of interfaces (*SI Appendix*, Fig. S5): 1) the MCP dimer interface; 2) the interface between MCP dimers; and 3) the interface between MCP dimers across the groove of the helix. Within the dimer, as previously reported, protein–protein interfaces are extensive for SIRV2, SSRV1, SFV1, and PFV2 (Fig. 5 and *SI Appendix*, Fig. S5). The interfaces of AFV1 and SIFV are a bit smaller, due to the turn in the N-terminal helix (Fig. 5*C*). The protein–protein interface between adjacent dimers are similar among all six virions, with ~23% total accessible surface area buried. Interactions between dimers across the groove of the helix are weaker compared to the interface between dimers. When excluding the C-terminal hook, nonenveloped virions SSRV1 and SIRV2 have almost three times larger buried surface area than enveloped virions SFV1, PFV2, AFV1, and SIFV, consistent with the greater rigidity observed in SIRV2 and SSRV1 (Fig. 4*A* and *SI Appendix*, Fig. S5). When including the C-terminal hook in SFV1 and SIFV, the

interfaces between MCP dimers across the helical groove are much larger, which probably stabilize the virions.

Next, we looked at the surface electrostatic potential of the six filamentous viruses without A-DNA (Fig. 4*D*). As expected, positively charged tunnels were observed for all six viruses to accommodate their dsDNA genome. The nonenveloped virions SIRV2 and SSRV1 have very similar surface electrostatic potential with a moderate number of charged residues on their surfaces. The electrostatic potential surfaces of the protein cores within enveloped virions are more diverse. For example, SFV1 has a large number of negatively charged residues on its surface. The differences in their surface electrostatic potential may explain differences in their selectivity of lipids acquired from the hosts.

Structural Conservation and Diversity of the Filamentous Virus MCPs. Except for the closely related MCPs of SIRV2 and SSRV1, no other homology is detected at the sequence level. However, a flexible secondary structure alignment indicates obvious homology among the 10 MCPs (Fig. 5*A*). From the structure-based alignment, it is clear that all MCPs have two or three N-terminal α-helices wrapping the DNA in the lumen of the virus (Fig. 5 *A* and *B*), but the details are quite different: SSRV1, SIRV2, and SFV1 have long helices with a kink in the middle to continuously wrap around the DNA; PFV2 is similar to SSRV1, SIRV2, and SFV1, but one of its MCPs, PFV2-2, has domain swapping (Fig. 5*C*); AFV1 and SIFV, on the other hand, form helix–turn–helix motifs, which fold back to cover the DNA on both sides (Fig. 5 *B* and *C*). In the C-terminal domain, all MCPs
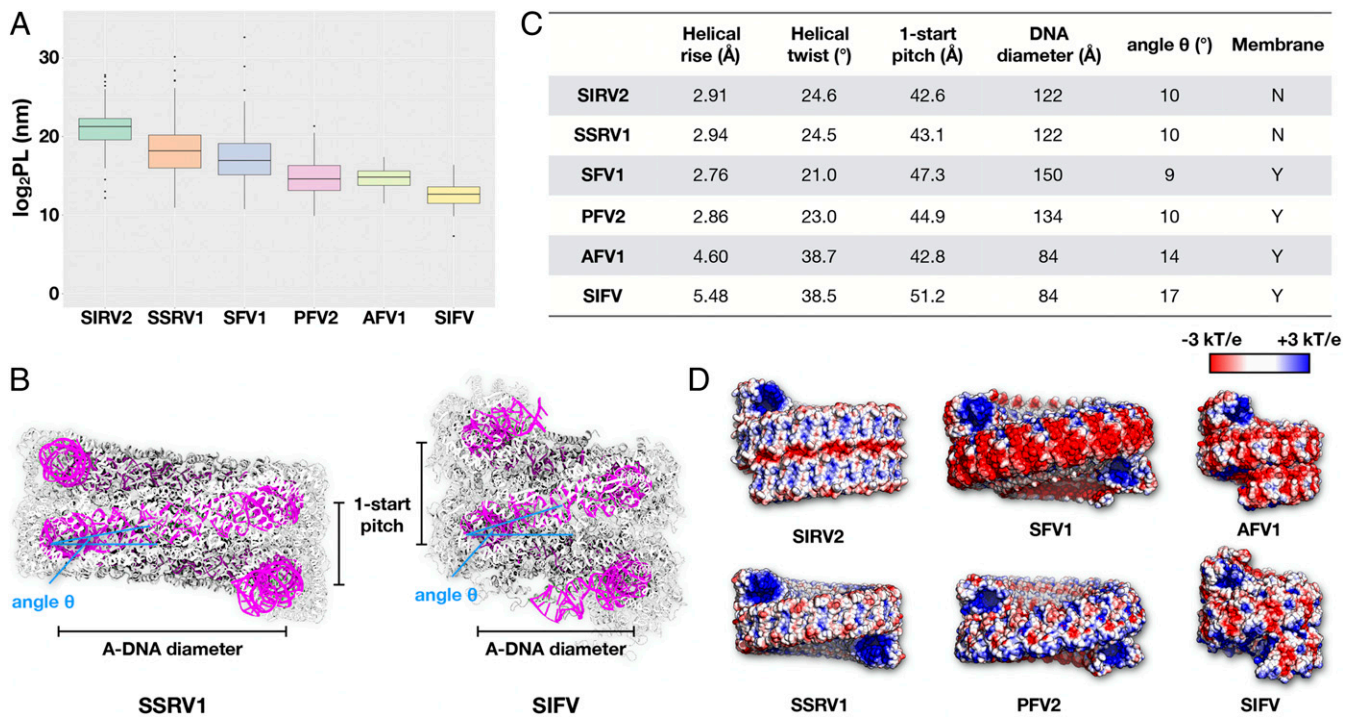
Wang et al.

Fig. 4. The flexibility of archaeal filamentous virions. (*A*) Estimates of the persistence length of the six archaeal filamentous virions. The measurements for each virus were from 100 filaments randomly selected from cryo-EMs. The measurements are shown in box-and-whisker plots that display five summary statistics (the median, two hinges, and two whiskers), and all "outlying" points individually. (*B*) The filamentous protein–DNA models of SSRV1 and SIFV, with DNA in magenta and protein in gray. (*C*) The parameters of all six archaeal filamentous virus structures, including helical rise and twist, 1-start pitch, A-DNA diameter, pitch angle θ, and the presence/absence of membrane. The DNA diameter is taken as the distance from the axis of the DNA on one side to the other, as this is more precisely defined than something like the outer diameter. (*D*) The electrostatic potential surface of all six archaeal filamentous virus structures, calculated by APBS (63).

have a more conserved four-helix bundle (*SI Appendix*, Fig. S6 *A and B*), where the MCPs have three shorter helices (Cα1–3) that combine with the longer helix that wraps the DNA to form the bundle. In the PFV2 MCPs, there is a short additional helix in the C terminus that is not part of the bundle (Fig. 5*C*). Previously, the C-terminal four-helix bundle domain of the major capsid protein from an uncharacterized member of the *Rudiviridae* family (which also includes SIRV1, SIRV2, and SSRV1) was solved by a combination of X-ray crystallography and NMR (33), and its structure is very similar to the C-terminal four-helix bundle domain of SSRV1 and SIRV2 determined by cryo-EM (*SI Appendix*, Fig. S6*C*). Like MCP1 of SFV1, MCP1 of SIFV also has a C-terminal extension reaching out to make extensive contacts with the subunits in the adjacent helical turn (Fig. 5*C*).

A TM-score (34) matrix analysis of the six MCP dimers is shown in Fig. 5*D*. TM score is a metric for measuring the similarity of two protein structures, and it has a value in the range from 0 to 1, where 1 indicates a perfect match between two structures. It has been shown that TM scores below 0.17 correspond to randomly chosen unrelated proteins, while a TM score higher than 0.5 is indicative of the same general protein fold (34). Pairwise comparison of the dimeric MCP structures from the six filamentous viruses followed by single-linkage clustering, which is suited to detect outliers, revealed three major clusters: SIRV2, SSRV1, and SFV1 clustered together with a TM score of >0.75, AFV1 and SIFV dimers formed a smaller cluster with a TM score of 0.72, whereas PFV2 was an outlier (Fig. 5*D*). This clustering is generally consistent with the current classification of the corresponding viruses into three families. Notably, however, SFV1, which is an enveloped virus (family *Lipothrixviridae*), clustered with the nonenveloped rudiviruses rather than with other lipothrixviruses.

To gain further insights into the evolutionary relationship between filamentous archaeal viruses, we performed phylogenomic analysis of all available rudivirus, lipothrixvirus, and tristromavirus genomes using the Genome-BLAST Distance Phylogeny (GBDP) method implemented in VICTOR (35). In the phylogenomic tree based on the shared gene content and rooted with tristromaviruses, all four genera of lipothrixviruses are retrieved with maximal support values. However, family *Lipothrixviridae* is paraphyletic with respect to rudiviruses, which form a sister clade to genus *Gammalipothrixvirus* (Fig. 6). Indeed, 10 of the gammalipothrixvirus AFV1 genes are shared with rudivirus SIRV2 (16). This result suggests that rudiviruses have evolved from within lipothrixvirus diversity.

Discussion

Filamentous helical viruses are common in all three domains of life (36–40), but despite overall similar morphology, virions of bacterial, eukaryotic, and archaeal filamentous viruses are built from unrelated capsid proteins with different structural folds (41). Furthermore, the ways the nucleic acids are protected by the corresponding capsid proteins are radically different in viruses from the different domains of life (19, 31, 42–44). With the two virion structures presented herein, archaeal filamentous viruses of the proposed class "*Tokiviricetes*" (families *Rudiviridae*, *Lipothrixviridae*, and *Tristromaviridae*) emerge as valuable models for understanding the evolution of virus structure and adaptation to extreme environments. High-resolution structures are now available for six evolutionarily related viruses with different degrees of flexibility, providing insights into structural changes underlying the differences in mechanical properties of the viral particles. Archaeal viruses are unique in this respect, because in eukaryotes, rod-shaped viruses, such as tobacco mosaic virus, the
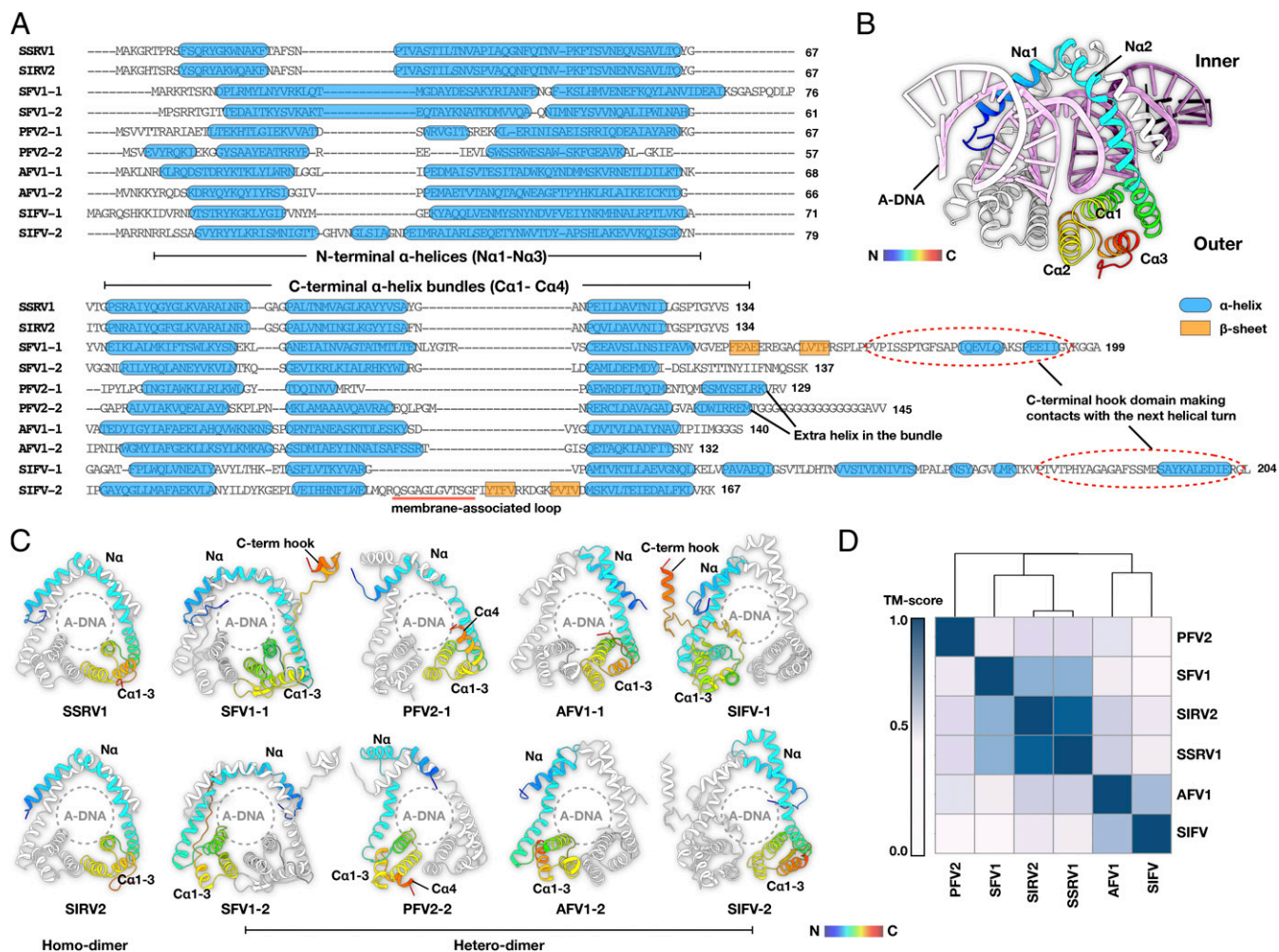
Wang et al.

Fig. 5. Structural conservation and diversity of the filamentous virus MCPs. (*A*) Structure-based sequence alignments of 10 MCP sequences from six archaeal filamentous virus structures. The α-helices are indicated by blue rounded rectangles, and β-sheets are indicated by orange rectangles. (*B*) Representative domain architecture of the MCP in known filamentous virus. An SSRV1 dimer with 36-bp A-DNA is shown: One MCP is shown in rainbow coloring and the other in white; DNA is colored magenta. N-terminal α-helices wrap DNA on the luminal side and C-terminal α-helix bundles wrap DNA facing the outer solvent or the membrane. (*C*) MCP comparison of SSRV1, SIRV2, SFV1, PFV2, AFV1, and SIFV. SSRV1 and SIRV2 are homodimers, while the other four MCP dimers are heterodimers. (*D*) All-against-all comparison of the six MCP dimers. The matrix is based on the pairwise TM score calculated from the MM-align server.

first virus ever isolated (42), on the one hand, and flexible viruses, such as potyviruses (45) and alphaflexiviruses (31), on the other hand, are built from nonhomologous capsid proteins. Notably, we find that some of the adaptations in virion structure of archaeal viruses are shared with virions of flexible plant viruses. In particular, the flexibility in SIFV, the most flexible among the compared viruses, can be, in part, attributed to the C-terminal tail, which extends into the adjacent helical turn, where it makes contacts with other capsid protein subunits. As a result, all adjacent helical turns in the virion are tied together, allowing deformations such as compression, extension, and torsion, while still maintaining the structural integrity of the virion. A similar solution has previously been shown for plant alphaflexiviruses (31). More generally, comparison of the persistence lengths of the archaeal viruses showed that nucleocapsid diameter and helical rise are the best predictors for virion rigidity, where the smaller diameter and larger rise are associated with a more flexible virion.

In lipothrixviruses, the nucleocapsid is enveloped by a thin monolayer membrane (17, 18), which is half the thickness of the host membrane from which it is derived, whereas rudiviruses are not enveloped. Due to its importance for various aspects of

virus–host interactions, such as during virus entry and egress, the presence or absence of an envelope is generally considered an inherent property of a given virus group, which is not frequently acquired or lost during evolution. Nevertheless, transitions from enveloped to nonenveloped virions have occurred among certain groups of eukaryotic RNA viruses, apparently as an adaptation to plant hosts (46). Comparison of the relatively closely related rudiviruses and lipothrixviruses provides insights into virion envelopment. The DNA is more tightly complexed by the capsid protein in rudiviruses SIRV2 and SSRV1 than in lipothrixviruses SIFV and AFV1, suggesting that the envelope in the latter group of viruses is an adaptation providing an additional layer of protection for the genome against the acidic extracellular environment in which these viruses thrive. Comparison of the (nucleo) capsid surfaces of lipothrixviruses and rudiviruses did not reveal major differences in their hydrophobicity (*SI Appendix*, Fig. S7). This is not entirely surprising, as the protein surface would be facing polar headgroups when a membrane is present. Notably, the major difference in capsid surfaces is actually electrostatic potential, suggesting that acquisition/loss of the membrane has likely been accompanied by extensive sequence changes of the capsid surface charged residues. SIFV contains a unique
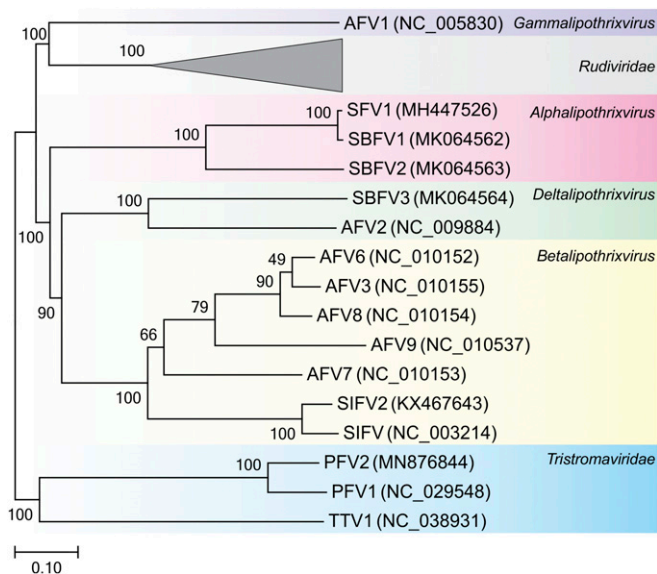
Wang et al.

Fig. 6. Inferred phylogenomic tree of archaeal filamentous viruses. The tree is based on whole-genome VICTOR analysis at the amino acid level. The tree is rooted with tristromaviruses, and the branch length is scaled in terms of the Genome BLAST Distance Phylogeny (GBDP) distance formula D6. The numbers above branches are GBDP pseudobootstrap support values from 100 replications. For each genome, the abbreviated virus name and GenBank accession number are indicated. The tree is divided into colored blocks according to the taxonomy of the compared viruses.

surface-exposed insertion in one of the capsid proteins, which might be involved in membrane binding. However, equivalent regions are absent in other lipothrixviruses, suggesting that the determinants underlying envelopment are present in the core regions of the major capsid protein. It remains unclear whether the membrane interacts directly with the MCPs or through an intermediate matrix protein, as in most eukaryotic enveloped viruses (47). Abundant structural proteins other than the MCPs were indeed identified in SFV1 (17) and PFV1 (48), but not in other lipothrixviruses, including SIFV (*SI Appendix*, Fig. S2).

The considerations above beg the question: Was the membrane acquired by the ancestor of rudiviruses to produce the first lipothrixvirus or was it lost in a particular lineage of lipothrixviruses, giving rise to rudiviruses? The existence of a subgroup of rudiviruses encoding relics of a second MCP is consistent with the latter scenario. Notably, the inactivated and functional MCPs are adjacently encoded, as are the two functional MCPs in lipothrixviruses and tristromaviruses. The inactivated MCPs might have been exapted for a different function, a common process in virus evolution (49), thus explaining why they have not been lost. Interestingly, structural comparison of the capsid proteins of filamentous archaeal viruses showed that enveloped filamentous viruses form three clusters, one of which includes SFV1 and nonenveloped rudiviruses. The similarity between the capsid proteins of the three viruses extends also to the overall virion characteristics, including relatively high rigidity and presence of a central cavity, which is not present in SIFV and AFV1. Furthermore, whereas the MCPs of rudiviruses are conserved at the sequence level, those of lipothrixviruses are highly divergent, to the extent that homology between the two MCPs of the same virus or of viruses in the same family but from different genera are not recognizable. This suggests that lipothrixviruses (and their MCPs) have diverged in a more distant past than rudiviruses. Note that there is no reason to assume that evolutionary rate would be different between the two virus families. Based on shared gene content, we have previously proposed that

tristromaviruses have diverged from lipothrixviruses and rudiviruses concomitantly with the divergence of their respective hosts, Thermoproteales and Sulfolobales (16), suggesting that the virion organization of the postulated viral ancestor resembled that of lipothrixviruses and tristromaviruses. Finally, phylogenomic analysis of filamentous archaeal viruses also suggests that rudiviruses are a derived rather than ancestral group of viruses. Consequently, currently available data suggest that the most recent common ancestor of extant archaeal filamentous viruses was an enveloped virus with heterodimeric MCPs that encapsidated A-form DNA. Prior to this most recent ancestor, it appears reasonable to imagine a simpler ancestral virus with a homodimeric MCP, resembling the contemporary rudiviruses but not directly related to them. The envelope might have been an ancestral feature, which was shed only in rudiviruses, presumably following the emergence of the MCP capable of efficient DNA protection in harsh environmental conditions. It is almost certain that further exploration of archaeal virus diversity and their structural and biochemical characterization is bound to shed more light on the evolution of this remarkable group of viruses.

The high-resolution structure achieved for SSRV1 allowed us to better understand how B-form DNA is converted into A-form, which is a characteristic feature in this group of viruses (16–19). Notably, storage of DNA in A-form has been convergently evolved by bacterial spores (50, 51) and by an icosahedral hyperthermophilic archaeal virus SPV1 (52), suggesting that A-form is a preferred conformation for DNA storage under extreme environmental conditions. However, A-form DNA might not be the only mechanism of genome protection. For instance, the genome of spindle-shaped virus SSV1 is positively supercoiled in virions (53) and is complexed with the host chromatin protein of the Sul7d family (54). In contrast, the circular genomes of hyperthermophilic bacilliform clavavirus APBV1 (24) and icosahedral turrivirus STIV (55) are encapsided as naked dsDNA, and it remains unclear whether there are dedicated mechanisms for the DNA protection. Elucidation of these mechanisms should uncover adaptations to life in extreme environments and open doors for biomedical and nanotechnological applications.

Methods

Virus Production and Purification. Exponentially growing cultures of *Saccharolobus solfataricus* POZ149 (21) and *Sulfolobus islandicus* LAL14/1 (56) cells were infected with fresh preparations of SSRV1 and SIFV, respectively. The infected cultures were incubated at 75 °C under agitation for 2 d. After the removal of cells (7,000 rpm, 20 min; Sorvall 1500 rotor), viruses were collected and concentrated by ultracentrifugation (40,000 rpm, 2 h, 10 °C; Beckman 126 SW41 rotor). For cryo-EM analysis, the concentrated particles were resuspended in buffer A (54): 20 mM KH_2PO_4, 250 mM NaCl, 2.14 mM $MgCl_2$, 0.43 mM $Ca(NO_3)_2$, and <0.001% trace elements of Sulfolobales medium, pH 6. For SDS/PAGE and mass spectrometry analyses, virus particles were further purified by ultracentrifugation in a CsCl buoyant density gradient (0.45 g·mL⁻¹) with a Beckman SW41 rotor at 39,000 rpm for 20 h at 10 °C. The opalescent bands were collected with a needle and a syringe, and dialyzed against buffer A.

Analysis of SIFV and SSRV1 Structural Proteins. The purified virions were analyzed by SDS/PAGE, and proteins were stained with InstantBlue (Expedeon). The stained protein bands of SIFV virions were excised from the gel and in-gel digested with trypsin. For SSRV1, the CsCl-purified virions were trypsinized in solution. The generated peptides were separated and identified by nano-LC-MS/MS (Proteomics Platform, Institut Pasteur) using an Ultimate 3000 system (Dionex) coupled to an LTQ-Orbitrap Velos system (Thermo Fisher Scientific). Peptide masses were searched against annotated SSRV1 and SIFV proteomes using Andromeda with MaxQuant software, version 1.3.0.5 (57).

Phylogenomic Analysis. All pairwise comparisons of the amino acid sequences of rudivirus, lipothrixvirus, and tristromavirus genomes were conducted using the GBDP method implemented in VICTOR, under settings recommended for prokaryotic viruses (35). The resulting intergenomic distances derived from

www.manaraa.com

pairwise matches (local alignments) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing for D6 formula, which corresponds to the sum of all identities found in high-scoring segment pairs divided by total genome length. Branch support was inferred from 100 pseudobootstrap replicates each. The tree was rooted with members of the family *Tristromaviridae*.

Cryo-EM Image Analysis. The virus sample (~4.0 μL) was applied to discharged lacey carbon grids and plunge frozen using a Vitrobot Mark IV (FEI). Frozen grids were imaged in a Titan Krios at 300 keV and recorded with a K3 camera at 1.08 Å/pixel (SSRV1) or a Falcon III camera at 1.4 Å/pixel (SIFV). The SIFV data were collected at the University of Virginia core facility, and the SSRV1 data were collected at the National Cryo-EM Facility of the National Cancer Institute. Micrographs were collected using a defocus range of 1 to 2.5 μm, with a total dose of ~50 electrons/Å^2 distributed into ~25 fractions. To get a preliminary helical reconstruction volume in SPIDER, all of the micrographs were first motion corrected and dose weighted by MotionCorr v2, and then CTF-multiplied by the theoretical CTF. Filament images corresponding to ~20 electrons /Å^2 were extracted using EMAN2 (58). Small subsets containing ~30,000 overlapping 384-pixel-long segments were used to search for the correct helical symmetry. The helical symmetry was determined in SPIDER (59) using IHRSR (60) after searching through all possible symmetries by trial and error, until recognizable secondary structural features were seen. A ~6-Å initial reconstruction was generated from this small subset, and this volume was subsequently filtered to 7 Å as the starting reference used in RELION (61). After using the full dataset in RELION, doing CTF-refinement and Bayesian polishing, the final volume was estimated to have a resolution of 2.8 Å for the SSRV1 and 4.0 Å for SIFV, based on the map:map FSC, model:map FSC and d_{99} (62). The final volumes were then sharpened with a negative B-factor automatically estimated in RELION, and the statistics are listed in Table 1.

Model Building. The density corresponding to a single SSRV1 MCP was segmented from the experimental cryo-EM density using Chimera. Then a starting model was generated by homology modeling using the SIRV2 MCP as the reference, and then docked into the segmented map. Then this model was adjusted manually in Coot and real-space refined in PHENIX. The EM density corresponding to A-DNA was also segmented in Chimera, and the A-DNA was manually put in the map and refined in PHENIX. Finally, the refined single MCP and DNA model were used to generate a filamentous model using the determined helical symmetry, and this filament model was refined against the full cryo-EM map using in PHENIX real-space refinement. MolProbity was used to evaluate the quality of the filament model. The refinement statistics are shown in Table 1.

For SIFV, the density corresponding to a single MCP1 or MCP2 was segmented from the experimental filament density using Chimera. The full-length MCP1/MCP2 protein was built de novo into the segmented map using Rosetta-CM, then adjusted manually in Coot and real-space refined in PHENIX. The EM density corresponding to A-DNA was also segmented in Chimera, and the DNA model was manually built and refined in PHENIX. The filament model was refined in the same way as the SSRV1 model described above.

1. E. V. Koonin, V. V. Dolja, A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* **3**, 546–557 (2013).
2. C. A. Suttle, Viruses in the sea. *Nature* **437**, 356–361 (2005).
3. O. Bergh, K. Y. Børsheim, G. Bratbak, M. Heldal, High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
4. A. R. Mushegian, Are there 10^{31} virus particles on earth, or more, or fewer? *J. Bacteriol.* **202**, e00052-20 (2020).
5. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
6. Z. Li, A. H. Pearlman, P. Hsieh, DNA mismatch repair and the DNA damage response. *DNA Repair (Amst.)* **38**, 94–101 (2016).
7. J. Xia et al., Bacteria-to-human protein networks reveal origins of endogenous DNA damage. *Cell* **176**, 127–143.e24 (2019).
8. R. E. Franklin, K. C. Holmes, The helical arrangement of the protein subunits in tobacco mosaic virus. *Biochim. Biophys. Acta* **21**, 405–406 (1956).
9. M. Krupovic, V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili, E. V. Koonin, Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res.* **244**, 181–193 (2018).
10. D. Prangishvili et al., The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* **15**, 724–739 (2017).
11. D. Prangishvili, The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.* **67**, 565–585 (2013).
12. J. H. Munson-McGee, J. C. Snyder, M. J. Young, Archaeal viruses from high-temperature environments. *Genes (Basel)* **9**, 128 (2018).
13. A. J. Berliner, T. Mochizuki, K. M. Stedman, Astrovirology: Viruses at large in the universe. *Astrobiology* **18**, 207–223 (2018).
14. F. Pasin, W. Menzel, J. A. Daròs, Harnessed viruses in the age of metagenomics and synthetic biology: An update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnol. J.* **17**, 1010–1026 (2019).
15. M. Y. Chen, S. S. Butler, W. Chen, J. Suh, Physical, chemical, and synthetic virology: Reprogramming viruses as controllable nanodevices. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **11**, e1545 (2019).
16. F. Wang, D. P. Baquero, Z. Su, T. Osinski, D. Prangishvili, E. H. Egelman, M. Krupovic, Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. *Virus Evol.* **6**, veaa023 (2020).
17. Y. Liu et al., Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nat. Commun.* **9**, 3360 (2018).
18. P. Kasson et al., Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *eLife* **6**, e26268 (2017).
19. F. DiMaio et al., Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914–917 (2015).
20. D. Prangishvili, M. Krupovic, A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales." *Arch. Virol.* **157**, 791–795 (2012).
21. D. P. Baquero et al., New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J.* **14**, 1821–1833 (2020).
22. H. P. Arnold et al., A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virology* **267**, 252–266 (2000).
23. C. L. Moyer, G. R. Nemerow, Viral weapons of membrane destruction: Variable modes of membrane penetration by non-enveloped viruses. *Curr. Opin. Virol.* **1**, 44–49 (2011).
24. D. Ptchelkine et al., Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nat. Commun.* **8**, 1436 (2017).
25. J. M. Vargason, K. Henderson, P. S. Ho, A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7265–7270 (2001).
26. E. H. Egelman, Cryo-EM: Ice is nice, but good ice can be hard to find. *Biophys. J.* **118**, 1238–1239 (2020).
27. M. Egli et al., X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers* **48**, 234–252 (1998).
28. M. L. Kopka, A. V. Fratini, H. R. Drew, R. E. Dickerson, Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163**, 129–146 (1983).
29. H. R. Drew, R. E. Dickerson, Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535–556 (1981).
30. V. E. Galkin, A. Orlova, M. R. Vos, G. F. Schröder, E. H. Egelman, Near-atomic resolution for one state of F-actin. *Structure* **23**, 173–182 (2015).
31. F. DiMaio et al., The molecular basis for flexibility in the flexible filamentous plant viruses. *Nat. Struct. Mol. Biol.* **22**, 642–644 (2015).
32. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
33. B. R. Szymczyna et al., Synergy of NMR, computation, and X-ray crystallography for structural biology. *Structure* **17**, 499–507 (2009).
34. J. Xu, Y. Zhang, How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
35. J. P. Meier-Kolthoff, M. Göker, VICTOR: Genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**, 3396–3404 (2017).
36. S. J. Wylie et al.; ICTV Report Consortium, ICTV virus taxonomy profile: *Potyviridae*. *J. Gen. Virol.* **98**, 352–354 (2017).
37. M. Fuchs et al., ICTV virus taxonomy profile: *Closteroviridae*. *J. Gen. Virol.* **101**, 364–365 (2020).
38. M. J. Adams et al.; Ictv Report Consortium, ICTV virus taxonomy profile: *Virgaviridae*. *J. Gen. Virol.* **98**, 1999–2000 (2017).
39. S. Roux et al., Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).

Wang et al.

www.manaraa.com

40. G. Stubbs, A. Kendall, Helical viruses. *Adv. Exp. Med. Biol.* **726**, 631–658 (2012).

41. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2401–E2410 (2017).

42. A. Klug, The tobacco mosaic virus particle: Structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **354**, 531–535 (1999).

43. A. K. Tarafder *et al.*, Phage liquid crystalline droplets form occlusive sheaths that encapsulate and protect infectious rod-shaped bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4724–4731 (2020).

44. A. Grinzato *et al.*, Atomic structure of potato virus X, the prototype of the *Alphaflexiviridae* family. *Nat. Chem. Biol.* **16**, 564–569 (2020).

45. M. Zamora, E. Mendez-Lopez, X. Agirrezabala, R. Cuesta, J. L. Lavin, M. A. Sanchez-Pina, M. A. Aranda, M. Valle, Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Sci. Adv.* **3**, eaao2182 (2017).

46. R. Kormelink, M. L. Garcia, M. Goodin, T. Sasaya, A. L. Haenni, Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Res.* **162**, 184–202 (2011).

47. L. V. Kordyukova, E. V. Shtykova, L. A. Baratova, D. I. Svergun, O. V. Batishchev, Matrix proteins of enveloped viruses: A case study of influenza A virus M1 protein. *J. Biomol. Struct. Dyn.* **37**, 671–690 (2019).

48. E. I. Rensen *et al.*, A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2478–2483 (2016).

49. E. V. Koonin, M. Krupovic, The depths of virus exaptation. *Curr. Opin. Virol.* **31**, 1–8 (2018).

50. K. S. Lee, D. Bumbaca, J. Kosman, P. Setlow, M. J. Jedrzejas, Structure of a protein-DNA complex essential for DNA protection in spores of Bacillus species. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2806–2811 (2008).

51. P. Setlow, I will survive: DNA protection in bacterial spores. *Trends Microbiol.* **15**, 172–180 (2007).

52. F. Wang *et al.*, A packing for A-form DNA in an icosahedral virus. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22591–22597 (2019).

53. M. Nadal, G. Mirambeau, P. Forterre, W.-D. Reiter, M. Duguet, Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature* **321**, 256–258 (1986).

54. E. R. Quemin *et al.*, *Sulfolobus* spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J. Virol.* **89**, 11681–11691 (2015).

55. D. Veesler *et al.*, Atomic structure of the 75 MDa extremophile *Sulfolobus* turreted icosahedral virus determined by cryoEM and X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5504–5509 (2013).

56. C. Jaubert *et al.*, Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open Biol.* **3**, 130010 (2013).

57. J. Cox *et al.*, Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

58. G. Tang *et al.*, EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

59. J. Frank *et al.*, SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).

60. E. H. Egelman, A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).

61. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).

62. P. V. Afonine *et al.*, New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).

63. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).

www.manaraa.com