MORPHOLOGICAL ADAPTATIONS FACILITATING ATTACHMENT FOR ARCHAEAL VIRUSES

by

Ross Alan Hartman

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

In

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2019

DEDICATION

This work is dedicated to my daughters Roslyn and Mariah; hope is a little child yet it carries everything.

ACKNOWLEDGEMENTS

I thank all the members of my committee: Mark Young, Martin Lawrence, Valerie Copie, Brian Bothner, and John Peters. I especially thank my advisor Mark for his child-like fascination with the natural world and gentle but persistent effort to understand it.

Thank you to all the members of the Young lab past and present especially: Becky Hochstein, Jennifer Wirth, Jamie Snyder, Pilar Manrique, Sue Brumfield, Jonathan Sholey, Peter Wilson, and Lieuwe Biewenga. I especially thank Jonathan and Peter for their hard work and dedication despite consistent and often inexplicable failure.

I thank my family for helping me through the darkness by reminding me that sanctification is through suffering, that honor is worthless, and that humility is a hard won virtue.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

ABSTRACT

Little is known regarding the attachment and entry process for any archaeal virus. The virus capsid serves multiple biological functions including: to protect the viral genome during transit between host cells, and to facilitate attachment and entry of the viral genome to a new host cell. Virus attachment is conducted without expenditure of stored chemical energy i.e. ATP hydrolysis. Instead, virus particles depend on diffusion for transportation and attachment from one host cell to another. This thesis examines the attachment process for two archaeal viruses.

Sulfolobus turreted icosahedral virus (STIV) is well characterized for an archaeal virus. Still, no information is available concerning STIV attachment or entry. The research presented here shows that STIV attaches to a host cell pilus. Furthermore, combining the previously determined atomic model for the virus, with cryo-electron tomography, a pseudo-atomic model of the interaction between the host pilus and virus was determined. Based on this data, a model is proposed for the maturation of the virus capsid from a noninfectious to an infectious form, by dissociation of accessory proteins. Finally, a locus of genes is identified in the host cell, encoding proteins essential for viral infection, that are likely components of the pili structure recognized by STIV.

The isolation of a new archaeal virus, Thermoproteus Piliferous Virus 1 (TSPV1), is also presented here. The TSPV1 virion has numerous fibrous extensions from the capsid, of varying length, that are the first observed for any virus. The capsid 2-3nm fibers likely serve to extend the effective surface area of the virus, facilitating attachment to host cells. Characterization of this new virus was conducted, including genome sequencing and determination of the protein identity for the capsid fibers.

The research presented here provides a substantial advancement in our knowledge of the attachment process for archaeal viruses. Attachment to host pili is now emerging as a common theme for archaeal viruses. Furthermore, the isolation of the new archaeal virus TSPV1 demonstrates a novel strategy to increase the probability of interaction between a virus and host cell.

CHAPTER ONE

INTRODUCTION AND RESEARCH OBJECTIVES

Archaea The 1st, 2nd, and 3rd Domains of Life

Cellular life is divided into 3 domains: Bacteria Eukarya, and Archaea (1, 2). Although originally classified as a subtype of bacteria given their lack of the nucleus, the Archaea eventually earned recognition of separate domain with the realization that their replication, transcription, translation machineries are more closely related to eukaryotes than to bacteria (3-6). Recently, the entire 3 domain system has been again challenged with the discovery of the TACK, Asgard, and DPANN superphyla (7, 8). Extensive phylogenetic analysis suggests that eukaryotes either branch within or constitute sister phylum to the Asgard archaea (7, 9). A large cohort of proteins once considered to be signatures of eukaryotic cells have also been discovered in various members of the Asgard group, including cytoskeleton proteins (7).

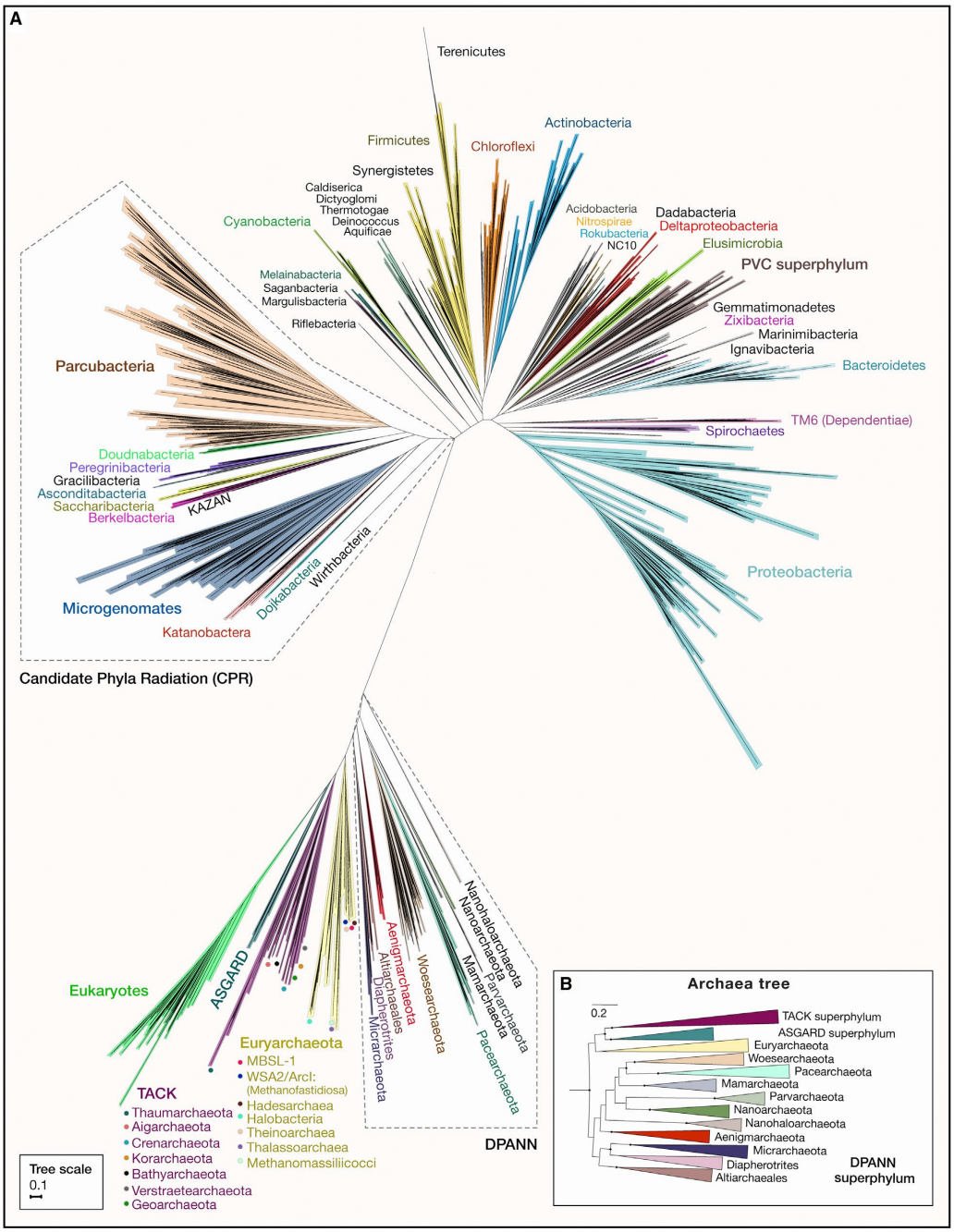


Figure 1.1. Universal Tree of Life Showing Eukaryotes as a Clade within the Archaea A) Universal Tree of Life based on phylogeny of 14 ribosomal proteins. B) Insert shows phylogeny of sequences from archaeal phyla. Figure is adapted from Castelle et al. (10)

Still, the lipid divide between archaea and all other organisms remains a perplexing issue (11). Archaeal lipid membranes differ from bacterial and eukaryotic membranes by four distinguishing features. First, rather than being composed at fatty acid chains, archaea use highly branched, saturated, isoprenoid based lipid tails (12). Cyclopentane rings are often incorporated into these lipid tails to adjust membrane fluidity (13). Second, the glycerol group is joined by ether rather than ester bonds to the lipid tail (12). Thirdly, the glycerol moiety has sn-2,3 stereochemistry which is opposite in chirality to the sn-1,2 glycerol groups of bacteria and eukarya (12). Finally, the diether building block of archaeal membranes (archaeol) is frequently observed as a fused tetra-ether species (14). These tetra-ethers form a monolayer as opposed to the familiar bilayers of other organisms (14).

Together these features are believed to give archaeal cells enhanced stability and low solute permeability enabling them to survive energetically demanding situations including extremes in pH, temperature, and salt concentration (15). Some of these membrane features are observed as lipid species in thermophilic bacteria, such as ether bonds and cyclopentane rings; still many are unique to archaea (13, 16). Particularly the unique stereochemistry of the glycerol group in archaeal lipids requires a suite of stereo specific enzymes not found in bacteria or eukaryotes (17, 18). Combined with the close phylogenetic relationship between the Asgard archaea and all living eukaryotes, this presents a puzzle for how eukaryotes acquired bacteria type lipid membranes if they are descended from archaea. Clearly, an understanding of archaea is essential for research into the origin of life and development of the eukaryotic cell.

Origin and Evolution of Viruses

As with archaea, the origin and evolution of viruses in particular, and archaeal viruses specifically is hotly debated. Three main theories exist to explain the emergence of viruses. The first, which we might term the progressive model, posits that viruses or virus-like particles preceded the emergence of cellular life (19). This is supported by the use of ssRNA by many virus families (19). RNA can serve both the catalytic and information storage role making it an attractive candidate for primordial life (19, 20). Furthermore, ssRNA viruses have small genomes and encode error-prone RNA - dependent RNA polymerases (19). In contrast, dsDNA viruses are frequently large, complex entities with low error rates during genome replication (19). It is plausible that the first polymerases had high error rates that would be lethal in organisms with large complex genomes, yet could suffice for a small virus-like self-replicating entity (19).

The 2nd model for the origin of viruses speculates a regressive direction of evolution. In this theory viruses originate from parasitic cells that have undergone successive rounds of genome reduction until they rely upon host protein machinery to perform their basic replication (21). This theory has enjoyed recent popularity with the discovery of giant algal viruses and nano -sized bacteria and archaea. Pandora virus encodes 2500 proteins and is 1000 nm in diameter, while Nanoarchaea equitans is 400 nm in diameter and encodes only 500 proteins (22, 23).

Finally, the escape model postulates that viruses originated on multiple independent occasions from selfish genetic elements that repurposed cellular proteins to

construct viral capsids and facilitate horizontal transfer between infected cells (21). Such escape events likely occurred at different stages of cellular evolution, with some predating the Last Universal Cellular Ancestor (LUCA), while others appear to be relatively recent events (21). This last model is very compelling given the amount of structure-based evidence suggesting evolutionary relationships between viral capsid proteins and cellular proteins (21). Still it is difficult to determine with certainty the direction of gene flow, as frequently viral proteins have been co-opted cellular organisms and repurposed for a new function (24).

With regard to archaeal viruses, understanding their evolutionary origin and trajectory is particularly challenging given the large number of unusual capsid architectures, and the sparsity of high-resolution structural data. Some archaeal virus families appear to be members of cosmopolitan viral lineages spanning the 3 domains of life and predating LUCA (25). This includes STIV, part of the Turriviridae virus family, whose capsid architecture and major coat protein fold (double jellyroll or DJR fold) unequivocally link this virus to relatives infecting bacteria and eukaryotes (26).

Still, many archaeal viruses have coat protein folds and/or capsid architectures that indicate they have no relationship to virus families that infect cells from the other domains of life. High-resolution structural data indicates that the Pleolipoviridae, Lipothrixviridae, Rudiviridae, Claviviridae are unrelated to other virus families (21, 27). Finally, several archaeal virus families lack structurally determined capsid folds, yet there overall morphology suggests these viruses are unique to archaea. This includes the

Globuloviridae, Fuselloviridae, Ampullaviridae, Spiraviridae, Guttaviridae, and Tristromaviridae (28).

Influence of Hot Spring Environment on Viral Evolution

The harsh conditions of hot spring environments may be responsible for the panoply of unique virus structures observed for archaea. Figure 1.2 depicts the 18 recognized or proposed archaeal virus families and emphasizes that virion structures unique to archaea are predominantly from thermophilic habitats. These high temperature and frequently low pH systems are dominated by archaea and their morphologically unusual viruses (29). Replication in such an environment presents several challenges, particularly to the attachment and entry phase.

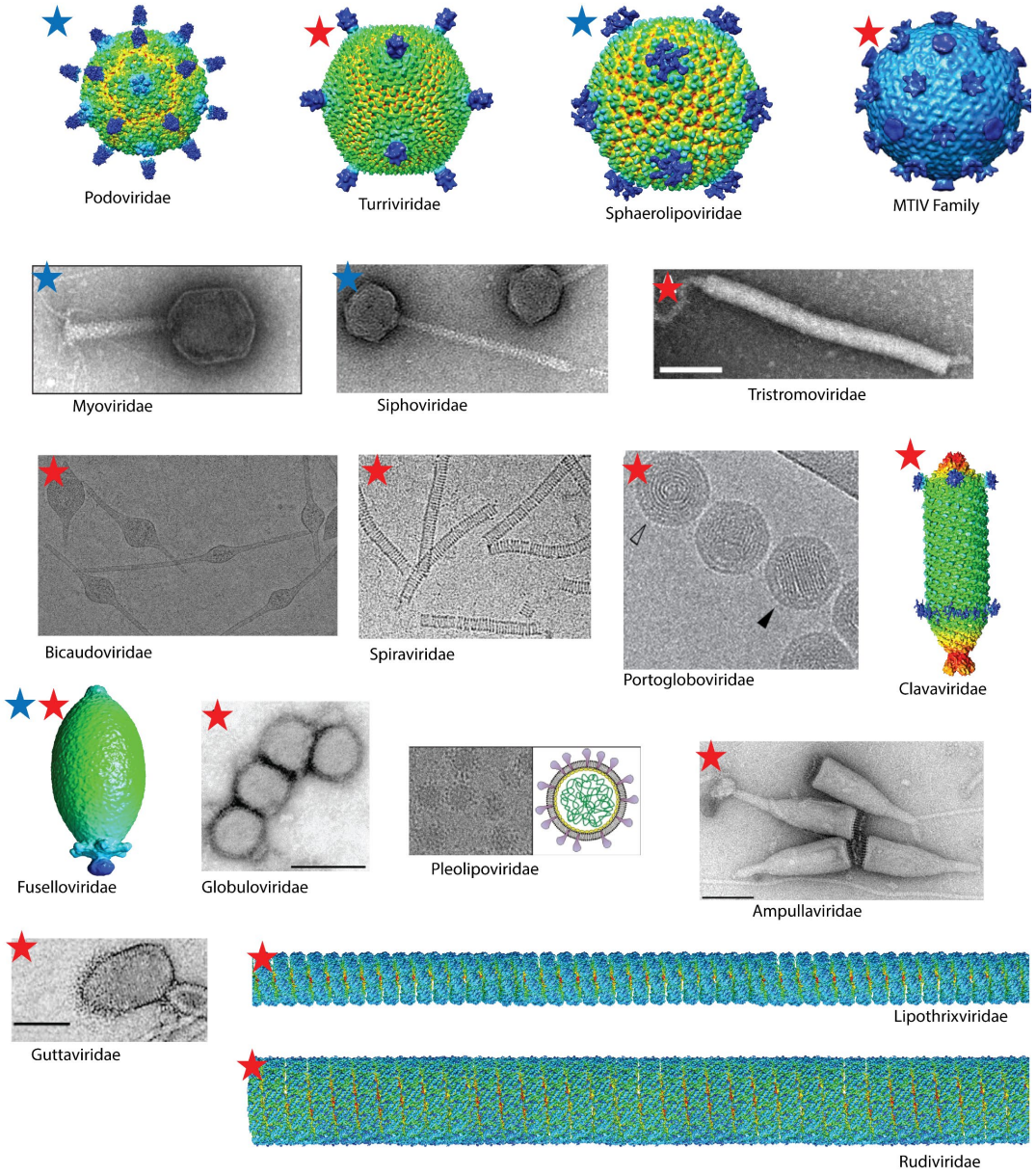


Figure 1.2. Mesophilic and Thermophilic Archaeal Virus Families. Red stars indicate viruses isolated from thermophilic environments, while blue stars indicate viruses from mesophilic environments.

While the interior of cell has pH values near neutral and regulated ion concentrations, once released into the environment thermophilic archaeal viruses must cope with low pH, bioreactive metals, and high temperature (30). In the acidic hot springs

of Yellowstone National Park, for instance, temperatures range up to boiling (94°C), with pH values typically pH 0 to pH 4, and dissolved Fe (II) up to 100 mg/L (31). In addition, these habitats have relatively low cell densities, from 1×10^6 cells/ml to only 100 cells/ml (32-36). Comparing this to ocean sediment (1×10^8), terrestrial soil (1×10^9), and the human gut (1×10^11) (37).

Viral particles are passively transported from one host cell to another by diffusion and hot spring currents. Their binding probability is directly proportional to the surface area of the virus and host cell. Several strategies may be employed by viruses to increase the likelihood of interaction with a permissive host. One of these is to produce large quantities of virus particles, yielding high burst sizes and increasing the odds that at least one particle will establish a new, productive infection. Another strategy is to construct highly stable capsids providing extended environmental lifetime, and thus also increased chance of contacting a permissive cell. Finally, viruses can maximize binding probability by targeting surface extensions from archaeal cells (pili and flagella) or by constructing capsids which maximize effective surface area.

Attachment and Entry for Bacteriophages and Eukaryotic Viruses

Viral replication cycle consists of 5 stages: attachment, entry, replication, assembly, and egress. Before discussing the current state of knowledge on attachment for archaeal viruses, we will examine this stage of replication in bacteriophages and eukaryotic viruses as a point of comparison.

One of the best studied enveloped viruses is HIV-1, including research into its attachment and entry mechanism. The envelope of HIV-1 is decorated with 10 to 100 copies of a trimeric protein assembly termed the envelope glycoprotein or Env (38). The core of this assembly is three gp41 proteins, which are anchored into the viral membrane by one C-terminal transmembrane helix for each protein subunit (39). The N-terminal domain of gp41 forms a three helix, metastable, coiled-coil structure(40). A fusion peptide region is then located at the extreme N-terminus of the protein (38).

Initially, gp41 is totally occluded, being buried inside a trimeric assembly of gp120 Env proteins (40). Contact with the host cell is initiated by interaction between gp120 of the virus and the CD4 receptor on the host cell (40). This 1st binding event leads to conformational changes in gp120 that expose another face of the gp120 protein to engage with a coreceptor on the cell (either CXCR4 or CCR5) (38, 40). This causes even further conformational changes to the entire Env assembly, leading to extension of the core gp41 trimer and insertion of its fusion peptide loops into the target host membrane (38).

In the extended state, gp41 is now connected to both the viral membrane by a C-terminal transmembrane domain, and to the host membrane by its N-terminal fusion peptide. This extended state collapses back into a final, stable 6 helix bundle structure, and in doing so draws the two membranes into close proximity (38). From this point, a hemifusion intermediate forms leading ultimately to full membrane fusion (38).

Another common strategy for eukaryotic viruses is represented by the non-enveloped Adenovirus. This virus is of particular interest to our research given its

structural and evolutionary relatedness to STIV, and its well-characterized attachment mechanism (41). Like STIV, Adenovirus is an icosahedral virus with spike proteins protruding from its fivefold symmetry axes (42). The knob domain of the spike protein is a jellyroll protein fold oriented in the same manner with respect to the viral capsid as the 3rd domain of the STIV C381 turret protein (43). Like HIV, binding of Adenovirus is a multi-step process. First, the knob domain of the spike interacts with the CAR receptor protein (42). This results in release of the spike and binding of the penton base to the secondary integrin receptor (42). Clustering of integrin proteins then promotes clathrin mediated endocytosis (42). As the pH of the endosome drops the penton base undergoes a structural rearrangement that exposes hydrophobic regions which insert into the endosome bilayer and disrupt it (42).

The small (T=3) icosahedral bacteriophage MS2 binds to the F-pili during infection, using its maturation or Mat protein to establish a network of hydrophobic and electrostatic interactions with the pilus (44). F-pili are Type IV retractile pili, frequently occurring in bacteria, with related non-retractile structures being present in archaea (45). Although the F-pilus is hollow and facilitates DNA exchange between bacterial cells, it appears that MS2 does not inject its RNA into this structure (44, 46). MS2 may use the base of the F-pilis as an entry route through the bacterial cell wall (44).

Infection of E. coli with T4 phage is perhaps the best characterized attachment and entry process for a bacteriophage, and undoubtedly one of the most complex. T4 is a complex multi-component molecular machine built to store energy in the form of unstable protein confirmations (47). Receptor binding is initiated by interaction of the

long tail fibers to a specific polymannose moiety in the host lipopolysaccharide (48). This leads to unfolding of short tail fibers which bind to a hexose lipopolysaccharide moiety (48). The next step is a massive structural rearrangement of the bacteriophage tail, causing its contraction, and driving a needlelike protein assembly through the outer membrane (47). At this stage a viral protein with lysozyme activity chews through the peptidoglycan layer (47).

Some general themes of virus infection can be drawn by analyzing these four unrelated viruses. First, attachment is often a multi-step process involving more than one receptor pair. The use of multiple receptors imparts host specificity and serves to choreograph the series of structural rearrangements necessary for membrane fusion, endocytosis, or genome injection. Second, given the cell wall present in bacteria, bacteriophages tend to use genome injection strategies for entry. Eukaryotic viruses, presented with an accessible lipid envelope, opt for either receptor mediated endocytosis or direct membrane fusion. Finally, bacteriophages frequently target cell surface structures like pili and flagella, to provide host specificity, and possible entry routes to the cell.

Current Knowledge of Attachment and Entry for Archaeal Viruses

The process of the attachment has been studied in only 2 archaeal viruses: SMV1 and SIRV2 (Figure 1.3). In 2013 Quemin et al. published the first results for attachment of an archaeal virus based on their study of the linear, Rudivirus SIRV2. They found that SIRV2 attaches to a specific host cell pilus using fiber structures present on the ends of

the virion then translocates to the cell body (Figure 1.3A). The authors presume SIRV2 uses DNA injection to enter the cell based on the observation of disassembled virus particles at the cell surface. Recently, the pilus structure to which SIRV2 binds has been identified using helical reconstruction and de novo modeling of its structure (49).

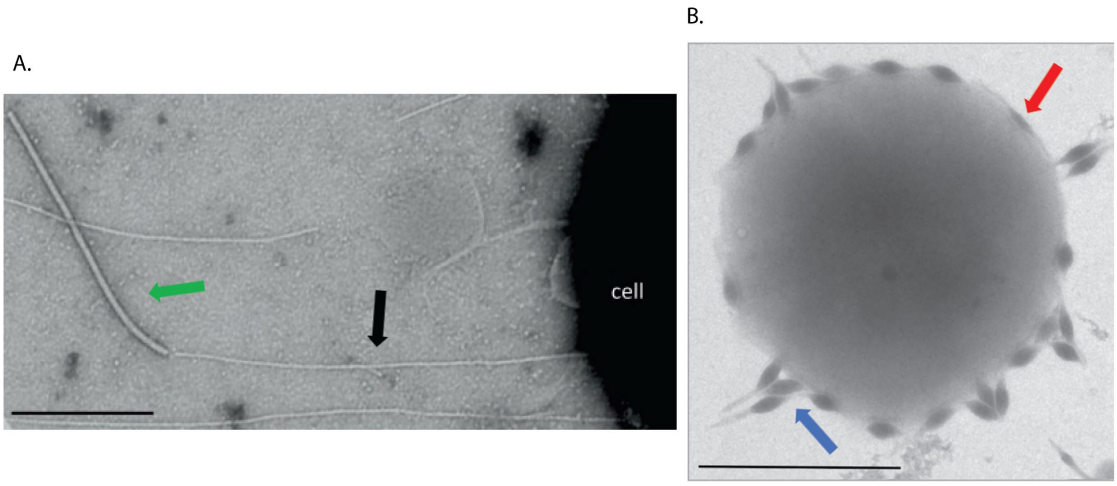


Figure 1.3. Attachment for archaeal viruses SMV1 and SIRV2. A) SIRV2 is shown (green arrow) attached to host cell pili. There is evidence that SIRV2 binds initially to the tip of the pilus and then works its way toward the cell body. B) SMV1 is shown to attach to the host cell body, the authors propose that the virus transitions from a spindle-shape (blue arrow) into an elongated flat structure (red arrow) upon host cell binding. Images are adapted from (50) and from (51).

SMV1 represents the only other archaeal virus to have published data regarding attachment. SMV1 is a spindle-shaped virus from the Bicaudaviridae virus family. Rather than attaching to a pili structure, SMV1 was observed to bind directly to the host cell (Figure 1.3B). Upon binding, the virus may undergo a morphologic transition from a short-tailed, spindle shaped virus to a long, flat, cylindrical, morphology. This morphologic transition may be a result of membrane fusion between virus and host.

In general, SIRV2 and SMV1 demonstrate two basic modes of archaeal virus attachment, with one following a bacteriophage model and the other a eukaryotic model. By targeting of host pili, SIRV2 initiates attachment in a manner similar to that used by many bacteriophages. In addition, the proposal that SIRV2 injects its DNA also follows a bacteriophage model of infection. In contrast, SMV1 appears to be an enveloped virus, and like enveloped eukaryotic viruses it interacts directly with the cell body. Membrane fusion may then follow this initial interaction, adding another layer of similarity between SMV1 infection and infection of eukaryotic envelope viruses.

Neither of these studies were able to provide molecular details concerning the attachment process. Although an atomic resolution structure is available for SIRV2, the structure of the terminal virion fibers that mediate attachment is unresolved. SMV1 lacks any high-resolution data concerning its capsid architecture or even the location of the proposed lipid envelope. Furthermore, it is clear that in comparison to bacteriophages and eukaryotic viruses, incredibly little is known of attachment and entry for archaeal viruses.

Structural Features of Archaeal Viruses and their Functional Roles

Advances in environmental sequencing have revealed a large number of previously unknown viruses that represent a deep pool of genetic diversity (52). This ‘dark matter’ is a major component of the virosphere, the entire collection of viruses on Earth (52). Classification of this collection of unknown viruses and characterization of their encoded genes is a challenging task, complicated by their limited homology to genes with known function (52-54). Virion structure, both at the level of individual protein

subunits (capsid proteins) and at the level of their three-dimensional arrangement (architecture), provides a compelling approach to organizing the diverse virosphere (55).

In contrast to ever increasing number of viral genomes being discovered, there is a countertrending decrease in discovery of new protein folds. At present there are 1393 known folds (56). Of these, only 20 folds are known for the major capsid proteins (MCPs) of viruses (57, 58). Despite the genetic diversity of viruses, their structural proteins, and especially their MCPs, fit within a limited number of groups. This limited number of common structural themes unite viruses across all three domains of life (Eukarya, Bacteria, and Archaea) and evolutionary history (59). Virion structure not only provides a useful tool for viral classification, it also allows for insights into viral assembly, release, attachment, and entry by comparing well-characterized members of a structural lineage with less characterized members (60-62).

Although viruses infect hosts in all domains of life, our understanding of viruses that infect Archaea (archaeal viruses), lags far behind those infecting Bacteria and Eukaryotes. While Archaeal viruses encompass virion morphologies also observed in bacteriophages such as head-tail, icosahedral, and filamentous morphologies, unique morphologies are also observed including spindle, bottle, spiral, and droplet shapes (25, 63-66). There are several excellent reviews on archaeal viruses which discuss archaeal virus biology and taxonomy (25, 28, 67). Here we focus on structural features unique to archaeal viruses and their role in virion architecture and dynamics within the extreme environments, in which these viruses replicate.

Out of the 125 known archaeal viruses representing 18 recognized or proposed taxonomic families, only 15 have high-resolution structural information available (28, 68). Table 1.1 lists the archaeal viruses with high-resolution structural data, while Figure 1.4 provides images of their MCPs and overall virion architecture . All of these structurally characterized archaeal viruses replicate in either high temperature (> 80^{o} C) acidic (pH < 4) or high salt (> 2.0 M) environments. Therefore, their virion morphologies and dynamics should be placed in the context of these unusual environments.

Virus Family	Virus Name	PDB/EMDB Accession #	Resolution	Major Discoveries	Ref
Bicaudaviridae	ATSV	5EQW	1.679 Å	New capsid architectural principle	ATV: ATSV: (69)
Caudovirales	HSTV1 HVTV1 HSTV2	EMD2279 EMD2234 EMD2235	8.9 Å 10.5 Å 9.8 Å	Linking of these halophilic archaeal viruses to Caudovirales	HSTV1: (70) HVTV 1&2:
Claviviridae	APBV1	EMD3857 5OXE	3.7 Å 3.7 Å	Unique capsid architecture MCP fold not used by any other known viruses	APBV1: (71)
Fuselloviridae	SSV1 His1	Not Avail. EMD6222	32 Å 20 Å	Possible fullerene cone-based architecture Capsid transformation from spindle to tube during DNA ejection	SSV1: (72) His1: (73)
Rudiviridae	SIRV-YNP SIRV2	3F2E EMD6310 3J9X	1.85 Å 3.8 Å 3.8 Å	A-form DNA	SIRV: (74, 75)
Lipothrixviridae	AFV1 SFV1	3FBZ & 3FBL EMD8780 5W7G EMD7797 6D5F	2.3 Å & 1.95 Å 4.5Å 4.5 Å 3.7 Å 3.7 Å	Rudiviridae and Lipothrixviridae share common MCP fold Membrane with monolayer of lipids	AFV1: (76, 77) SFV1: (78)
Sphaerolipoviridae	HHIV2 P23-77 (phage)	EMD3109 EMD1353 3ZN6	13 Å 9.6 Å 1.53 Å	Insight into evolution of Double Jelly Roll Fold	HHIV2: (79) P23-77: (80)
Turriviridae	STIV-1	4IND 2BBD & 6BO3 EMD5584 3J31 EMD1679	1.8 Å 2.04 Å & 1.83 Å 4.5 Å 4.5 Å 20 Å	The PRD1 viral lineage extends to archaeal viruses	STIV-1: (81)
Pleolipoviridae	HRPV2 HRPV6	6QGI 6QGL EMD9779	2.46 Å 2.69 Å 16 Å	The VP5 spike protein: likely new class of fusion proteins The V-shaped fold may represent a new protein fold	HRPV 2&6: (82)

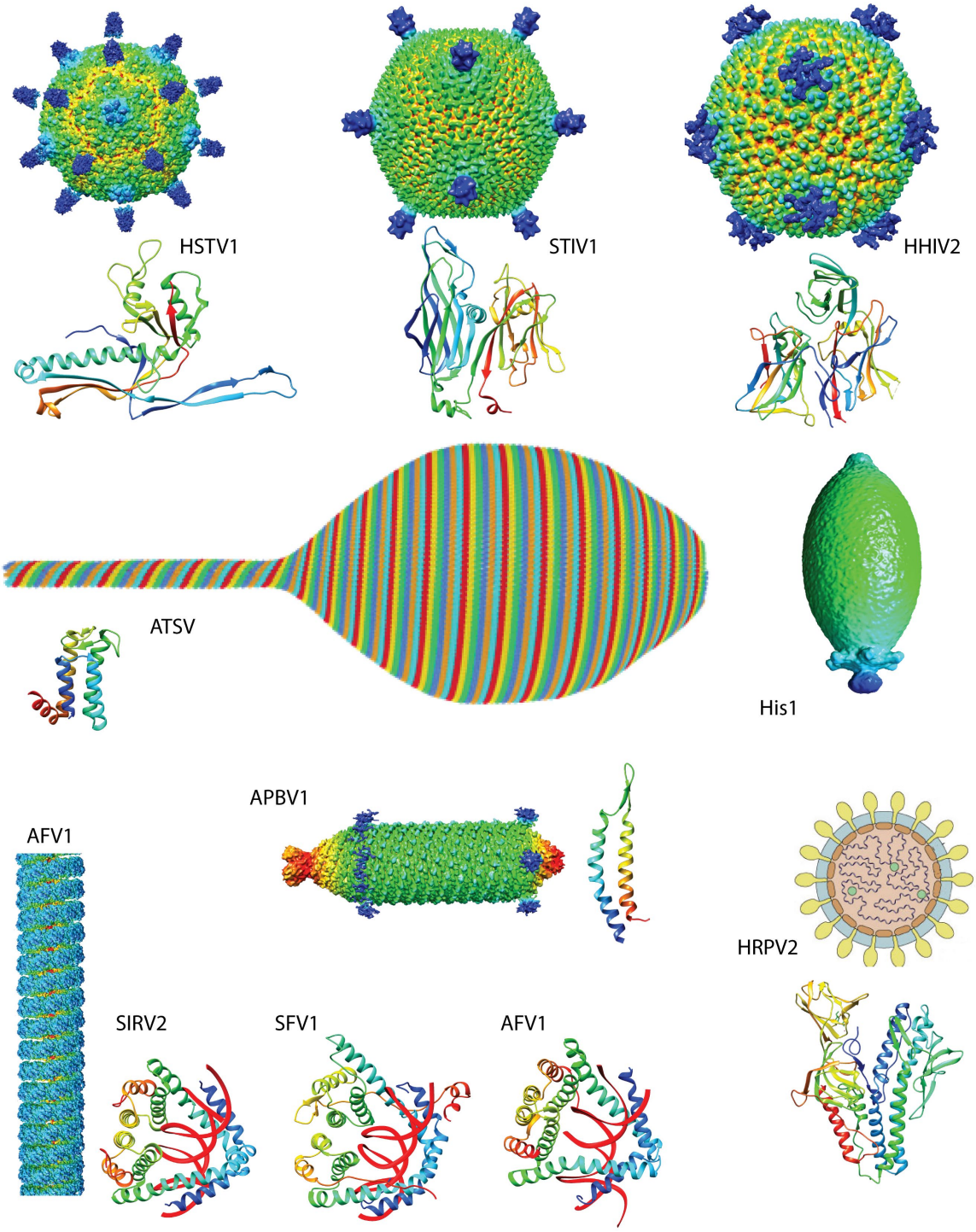

HSTV1
STIV1
HHIV2
ATSV
His1
APBV1
AFV1
HRPV2
SIRV2
SFV1
AFV1

Figure 1.4. Coat protein folds and capsid structures are shown for representative members of the Bicaudaviridae, Caudovirales, Claviviridae, Ligamenvirales, Fuselloviridae, Sphaerolipoviridae, Pleolipoviridae and Turriviridae. All capsids are at the same relative scale, except SFV1 and APBV1 which are at 2× scale and have been truncated in length. Images were generated with UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) except for the Pleolipoviridae capsid whose image is from ViralZone (reproduced with permission).

Turriviridae and Sphaerolipoviridae

Virion structure has served to unify disparate segments of the viral world and shed light on possible deep evolutionary relationships between viruses (55). Characterization of STIV-1, the founding member of the Turriviridae family, proved a landmark case. The major structural protein fold of the 38Kda monomer consists of two domains, with each containing 8 anti-parallel beta strands to form a flattened beta-barrel (41, 81, 83). This same fold, termed the double jellyroll or DJR fold had been previously observed for the capsid proteins of bacteriophage PRD1 and the Eukaryotic Adenovirus (84). Later, the Sphaerolipoviridae, a family of cosmopolitan viruses infecting both bacteria and archaea, were found to use paralogous single jellyroll capsid proteins in much the same way as their DJR brethren (79, 80, 85, 86). For all these viruses, the beta strands lie perpendicular to the capsid surface. Each jellyroll domain comprises 1 of the 6 subunits in the pseudo-hexamer.

The presence of this MCP fold orientated in the same way suggests that these viruses are evolutionarily related. Further support is gained by common features shared across this diverse group including: internal lipid membranes, homologous packaging enzymes, and receptor spikes at 5-fold vertices (26). This has prompted creation of the

PRD1 viral lineage which presently comprises 11 viral families infecting hosts from all 3 domains of life (26, 79).

Jellyroll domains are located not only in the viral hexamers making the major surface lattice. The pentameric spike proteins of Adenovirus, PRD1, and STIV-1 are also composed of jellyroll domains (59, 81). While these vary from the canonical 8 stranded DJR, their flat anti-parallel folds suggest common ancestry. We propose that the founding member of the PRD1 lineage used a single jellyroll domain protein to occupy both sixfold and fivefold positions of the capsid. Gene duplication then allowed specialization and divergence of the penton jellyrolls to produce either the stock and knob structures for PRD1/Adenovirus or the 7 jellyroll multitiered structure of STIV-1. Meanwhile, further gene duplication created the ubiquitous fused DJR and the separate paralogous single jellyroll's of the Sphaerolipoviridae.

The success of the jellyroll fold is likely do to its stability and relatively small size. The assembled pseudo-hexamers of Adenovirus, PRD1, and STIV-1 can resist boiling temperature, corrosive pH, and chemical denaturants (41, 87, 88). In addition, the core beta-barrel fold is encoded by a relatively small protein: facilitating the space restrictions of viral genomes. Finally, the basic fold is easily modified with extended loops to provide additional functionality, like immune evasion for Adenovirus (84).

Rudiviridae and Lipothrixviridae

Even before high-resolution structures were obtained, the rod-shaped Rudiviridae (SIRV1, and SIRV2) and Lipothrixviridae (AFV1, SFV1) were thought to share a common ancestor given the abundance of orthologous genes (89). Crystallization of the

MCP from each family revealed nearly identical folds, prompting creation of the Ligamenvirales order (76). The canonical fold is a left handed 4-helix bundle with 3 helices of nearly identical length, while the 4th helix at the N terminus of the bundle is much longer (74-78). This MCP is highly compact with short loops connecting each helix and extensive hydrophobic contacts is well-suited for the high temperature environment inhabited by Ligamenvirales.

Cryo- EM helical reconstructions of the Rudivirus SIRV2, the Lipothrixviridae SFV1 and AFV1 confirm that their MCP homology extends to their capsid architecture (74, 77, 78). In each virion, the DNA is packaged with homodimers of capsid protein in which the extended N-terminal helices wrap around and completely encircle the DNA, resulting in a cylinder shaped, right-handed super helical nucleoprotein complex with a hollow interior (74, 77, 78). In contrast to SIRV2, AFV1 and SFV1 employ a heterodimer of paralogous proteins (74, 77, 78). The SFV1 structure is further elaborated by a C-terminal extension, encompassing a 2-stranded antiparallel β-finger that packs against helices 3 and 4 on the outside of the nucleoprotein complex. This is followed by an extended random coil and two short α-helices that are collectively involved in subunit interactions with an adjacent super-helical turn. Strikingly, all three viruses condense and desolvate their DNA into A-form, protecting it from hydrolysis in the acidic hot spring environment (74, 77, 78).

Perhaps more surprising than their relatedness, is the use of the same protein fold to build either a rigid (SIRV) or flexible virion (SFV1, AFV2). The non-enveloped SIRV2 has the most tightly packed nucleoprotein arrangement giving it a rigid structure.

This includes extensive interactions across super-helical turns and complete occlusion of DNA by anti-parallel packing of the long N-terminal helices of its MCP (74). By contrast, SFV1 and AFV2 can suffice with less rigid capsids, presumably since their lipid envelope forms a proton barrier around the nucleoprotein core.

Some members of Lipothrixviridae have higher genetic homology to Rudiviruses than to members of their own family, suggesting either high horizontal gene transfer or lack of monophylogeny (89, 90). Additionally, the heterodimers of SFV1 are structurally more similar to the SIRV2 homodimer than to the AFV1 heterodimer (78). The extended N-terminal α-helices show high structural conservation between the VP4 and VP5 paralogs of SFV1, and with the SIRV capsid protein as well. Thus, the main feature distinguishing these 2 viral families is the presence or absence of an external lipid envelope. The AFV1 envelope is composed of cyclic glycerol dialkyl glycerol tetraether (GDGT) lipids, primarily the GDGT-0. In contrast, the SFV1 envelope is dominated by archaeol, a diether lipid (77, 78). Therefore, both viruses specifically recruit minor host lipids; each building a membrane different from that of their host and from each other (77, 78). Given these differences, it remains to be determined if the Lipothrixviridae are actually a polyphyletic group.

Archaeal cyclic tetraether lipids (C_{40}) such as GDGT-0 have long been known to span the entire membrane, contributing to the thermal stability of the host (91, 92). In this context, the high-resolution helical reconstruction of AFV1 revealed an additional surprise, the envelope is 2 nm, half the thickness of the host membrane (77). In the viral envelope, molecular dynamics simulations indicated that the individual tetraether GDGT-

0 lipids are likely folded back on themselves into a horseshoe-shaped configuration (77). Cyclic tetraether lipids have previously been observed to adopt this configuration at air water interfaces, however this represents the first indication of a biological membrane composed of this thinner horseshoe monolayer (77). Critically, the tetraether GDGT-0 lipids lack cyclopentane rings present in GDGT-4, allowing the flexibility needed to adopt this unique configuration and explaining the preferential recruitment of GDGT-0 into the viral envelope. Similarly, SFV1 shows an external lipid envelope thinner than expected. Interestingly, during budding the fusellovirus SSV1 also incorporates a GDGT-0 envelope with reduced thickness suggesting a related membrane structure, at least during this stage of the viral life-cycle (93, 94). Thus, lipid monolayers of single-leaflet dimensions may play a critical role in the life cycles of thermophilic viruses.

Claviviridae

APBV1 is the sole representative of the Claviviridae family (95). Although it's rod -shaped morphology is reminiscent of the Ligamenvirales, the 3 Å single particle reconstruction by cryo-EM and de novo modeling of the MCP distinguish it as an unrelated virus family (71). The VP1 MCP uses a helix-turn-helix motif with a type-1 beta hairpin comprising the turn (71). An abundance of hydrophobic residues facilitate intramolecular packing of each helix as well as extensive contacts with six adjacent subunits (71). Overall the virion is highly compact with extensive helix-helix packing between each VP1 monomer and the neighboring 6 proteins (71). The axis of each VP1 helix lies along the capsid surface with the helix curvature dictating the overall curvature of the virion.

The fold of VP1 has not been identified for any other viral MCPs, and its architectural organization is unlike that of other rod-shaped viruses (71). Nearly all archaeal viruses have been isolated from either thermo-acidic or hypersaline environments where archaea dominate (25). Less research has been conducted on circumneutral hot springs in which bacteria are more abundant than archaea (96-98). The discovery of the Claviviridae from a circumneutral hot spring highlights the importance of expanding archaeal virus isolation and study beyond hypersaline and thermo-acidic realms.

The absence of Claviviridae members from acid hot springs indicates that although this morphotype is suitable for maintaining thermal integrity it cannot withstand low pH. Isolated thermo-acidophilic archaeal viruses are frequently enveloped. Archaeal lipids composed of fused diethers or tetraether components show extremely low proton permeability (15, 99-101). Thus, they provide effective shields for packaged DNA within the virions from the acidophilic Turriviridae, Globuloviridae, Lipothrixviridae, and Fuselloviridae. The lack of a membrane envelope for APBV1 precludes this mechanism of pH tolerance. Although APBV1 shows gross morphological similarities to Rudiviridae it uses the standard hydrated B-form DNA packaged in the center of the virion with only the outermost strands making partial contact with the capsid protein (71). Together these results suggest that the Claviviridae are specifically adapted to circumneutral environments and are unable to withstand acidic hot springs.

Caudovirales Archaeal Viruses

Head-tail virus morphologies similar to bacteriophages are abundant among euryarchaeal viruses (67). Positive validation of their relatedness occurred through single particle reconstruction of HSTV1 and rigid body fitting of both the crystallographically determined HK97 fold and the modeled MCP fold for HSTV1 (70). The HK97 fold is a complex mixed α-helix β-sheet structure with two discontigious domains and a large anti-parallel β-hairpin loop (102). The signature HK97 covalent cross-links that weld the assembled capsid together do not appear conserved HSTV1 ; compensatory mechanisms for achieving capsid stability have not been proposed but are likely present in this virus (70).

Spindle Viruses: Bicaudaviridae and Fuselloviridae

At present, spindle-shaped viruses are only known to infect archaea. They are divided into 2 families, the Bicaudaviridae (large tailed spindles, 50-100 kb genome) and the Fuselloviridae (smaller tailless spindles, 15-20 kb). The size of the spindle-shape is likely related to the difference in their packaged dsDNA genomes, ~75Kb and 15Kb respectively. The coat protein structures have been determined for two members of the Bicaudaviridae: ATV and ATSV (69, 103). The central fold consists of a classic right-handed antiparallel 4-helix bundle capped end on by a small ß-sheet, and a short 5th helix at the C-terminus. Less structural information is available for the Fuselloviridae. Their MCPs are predominately hydrophobic making them insoluble and so far, challenging to characterize (64).

Despite appearing evolutionarily unrelated, both Bicaudaviridae and Fuselloviridae are built upon a spindle-shaped capsid that transitions into an elongated cylindrical form, giving the appearance of tail growth. Two Bicaudaviridae members are known to undergo this morphologic change: ATV in response to temperature and SMV1 upon host cell contact (51, 104). For ATV this reduces the central spindle volume by 50% and could serve to prime it for genome ejection (104). It is possible that other members like STSV1 and ATSV also undergo this transition, given their heterogeneous lengths ranging from 50 nm spindles to 500 nm cylinders (105, 106).

The Fusellovirus His1 can also be induced to undergo a related morphological transition, changing from a spindle-shaped body to a cylindrical tube when subjected to elevated temperatures, alkaline pH, or detergent (73). In addition, the transition also results in partial genome ejection, where the internal capsid pressure of 10 atm is the driving force (107, 108).

Proposed mechanisms to explain these gross structural rearrangements in spindle-shaped viruses have been lacking until recently (69). The capsid protein of ATSV crystalizes as a four-start or quadruple super-helical assembly, exhibiting structural similarity to the virus tail (69). Combined with radially symmetric striations on the surface of the spindle-shaped capsid in cryo-EM micrographs; this led to a proposed architecture in which a multi-start helix of varying width forms the spindle-shaped capsid as well as the tail (69). Interactions in the crystal reveal extensive intrastrand contacts, but fewer interstrand contacts at the subunit intrefaces (69). This combination of strong intrastrand and weak interstrand forces could allow the capsid coils to slip past one

another, smoothly altering the capsid diameter as it transitions to the lower energy cylindrical state (69). It is likely that the dynamic nature of multi-start super helical assemblies will be a unifying feature of spindle-shaped virions, and that the transition from the metastable spindle-shape to the lower energy, tailed and cylinder-like structures seen in ATV, SMV1 and His1 are energetically responsible for genome delivery into the host cell. We also note that the ability to encapsidate a complete viral genome and then mature to a smaller pressurized particle might provide the large-tailed spindle-shaped viruses with flexibility in genome packaging; making it easier to tolerate changes in genome size, while still arriving at a fully pressurized capsid through tail growth.

In essence, these spindle-shaped capsids are proposed to exist in a liquid crystalline, or smectic state (109). And like His1 (108), they can also be pressurized. Dharmavaram et al have shown that the equilibrium shape of a liquid shell, enclosing a uniformly pressurized volume, has to be that of a *surface of constant mean curvature* (109). Only a limited number of such surfaces exist (110), and of these, only the unduloid approximates the shape of a lemon-shaped capsid (109). Interestingly, two additional surfaces of constant mean curvature are the sphere and cylinder, which happen to approximate the shapes of spherical and filamentous shaped capsids, respectively. While the unduloid differs from the sphere and cylinder in containing regions of both positive and negative curvature, as a surface of constant mean curvature it may suggest an underlying reason for the occurrence of lemon-shaped viruses in the archaea. From this perspective, rather than ask why lemon-shaped capsids are found in the archaea, we might instead ask why they have not yet been found in bacterial or eukaryotic viruses.

Biophysical Adaptation to Extreme Environments

Current archaeal virus isolates are predominantly either halophilic or thermophilic, and among the thermophiles, frequently acidophilic. All archaeal viruses characterized to date have enveloped capsids and/or glycosylated MCPs (Table 1.2). Glycosylated capsid proteins are found in 8 archaeal virus families (68, 71, 77, 78, 94, 111-114). Additionally, viral encoded glycosyltransferases are a common feature and have been identified in an additional 6 families (63, 106, 113, 115-117). Glycosylation often promotes increased protein stability by creating a network of glycan-glycan and glycan-protein non-covalent interactions and by blocking proteolysis (118).

Virus Family Representatives (Habitat)	MCP Glycosylation	Enveloped	References
Ampullaviridae ABV (87-93° pH 1.5)	Putative Glycosyltransferase	Likely Enveloped	ABV: (66, 117)
Bicaudaviridae ATV, ATSV, STSV1&2 (80-85° pH 2-3)	Putative Glycosyltransferase	STSV1&2: Yes ATV,ATSV: ?	STSV1: (119) STSV2: (105)
Caudovirales HSTV1 (Hypersaline)	No Data Reported	No	HSTV1: (70)
Claviviridae APBV1 (90° pH 7)	Yes	No	APBV1: (71)
Fuselloviridae SSV1 (80° pH 3) Salterproviridae His1 (Hypersaline)	SSV1: Yes His1: No	SSV1: Yes His 1: Lipid Modified	SSV1: (93, 94) His1: (120)
Globuloviridae PSV (85° pH 6)	Putative Glycosyltransferase	Yes	PSV: (115)
Guttaviridae APOV1 (90° pH 7) SNDV (80° pH 3)	No Data Reported	Yes	SNDV: (65) APOV1: (65)
Lipothrixviridae AFV1, SFV1 (80-85° pH 2-3)	Yes	Yes	AFV1: (77) SFV1: (78)
Pleolipoviridae HRPV1 (Hypersaline)	HRPV1: Yes HRPV2&6: No	Yes	HRPV1: (112) HRPV2&6: (82)
Portogloboviridae SPV1 (80° pH 3.7)	Putative Glycosyltransferase	Yes	SPV1: (116)
Rudiviridae SIRV2 (87-93° pH 1.5-2)	Yes	No	SIRV2: (114)
Sphaerolipoviridae HHIV2, SH1, HCIV (Hypersaline)	No	Yes	SH1: (121) HHIV2: (79) HCIV1: (122)
Spiraviridae ACV (90° pH 7)	Putative Glycosyltransferase	No	ACV:(63)
Tristromaviridae PFV1 (88° pH 6)	Yes	Yes	PFV1: (113)
Turriviridae STIV-1 (80° pH 3)	Yes	Yes	STIV-1: (81, 111)
Unclassified MTIV (75-82° pH 2.1)	Yes	Yes	MTIV: (68)

Likewise, lipid envelopes are found in 11 of the 18 archaeal virus families (65, 68, 113, 116, 123). Archaeal lipids composed of fused diethers or tetraether components show extremely low proton permeability, indicating that one function in the virion could be the protection of the packaged genome (100, 101). Intriguingly, a number of these viruses selectively acquire a subset of host lipids by an unknown mechanism (113, 116, 123). The Fuselloviridae and Lipothrixviridae envelopes are composed primarily of GDGT-0 (SSV1 and AFV1) or archaeol (SFV1), which have much higher flexibility than the GDGT-4 (GDGT with 4 cyclopentane groups) that predominates in the membranes of their hosts (77, 78, 93, 94). These flexible lipids are likely required to facilitate the high membrane curvature of Fuselloviridae and Lipothrixviridae capsids (77, 78). Increased environmental stability is another rationale for selected lipid acquisition. SH1 enriches its membrane with PGP-Me which is known to increase membrane stability in the high salt environments inhabited by this virus (121, 124).

While the unique properties of these envelopes solve one problem, they raise new ones. Specifically, SSV1 has been shown to bud from the host cell in a manner reminiscent of eukaryotic viruses (93), and it seems likely that of the many enveloped archaeal viruses, at least some of them will fuse with host cell membranes to deliver their genomes. How do these viruses bud from or fuse with membranes composed of membrane spanning cyclic tetraether lipids that are frequently found in thermophilic archaea? The presence of such membrane spanning lipids would seem to prohibit the separation of the membrane into the individual leaflets, characteristic of the hemifusion intermediate proposed for eukaryotic viruses (125). This suggests unique strategies and

protein machinery might be needed to drive the process. One possibility is that the horseshoe-shaped lipid monolayer could serve as an intermediate equivalent to a single membrane leaflet in the hemi-fusion model.

The ESCRT complex is repurposed by most enveloped eukaryotic viruses to facilitate budding (126). Interestingly, Archaea also possess ESCRT machinery, and an ESCRT-III paralog is necessary for STIV replication, where it has been shown to associate with STIV lysis structures (127). However, STIV does not bud, but instead, like SIRV2, utilizes a virus associated pyramid-like (VAP) cellular lysis structure (128). The ESCRT machinery has not been directly linked to SSV1 budding (93), and whether it is involved in budding or fusion for archaeal viruses in general remains an open question.

While the mechanics of membrane remodeling may not be as unique, the Pleolipoviridae, have recently been shown to utilize spike proteins in their envelope to catalyze membrane fusion with the host cell (82). The VP5 spike structure lacks homology to other known fusion proteins, but given the lipid bilayer of the halophilic host the reaction likely precedes by a hemifusion intermediate (82).

Research Objectives

Archaeal viruses isolated from acidic hot springs replicate in low cell density environments. Low host availability, combined with the harsh chemical conditions, indicate that these viruses are under strong selective pressure to rapidly find and infect a host. Still these viruses have no means of active transport, and are carried by diffusion and hot spring currents from one host to another. Given these environmental constraints,

imposed upon archaeal viruses, we expect them to use strategies that increase their effective surface area to facilitate attachment to a host cell . This can be accomplished in 3 ways. One method is to increase virion aspect ratio. High aspect ratio (cylindrical) particles undergo a tumbling motion, when exposed to shear forces, which increases their effective surface area. For example, the tenfold change in the length of ATV (60-80 nm to 600-800 nm) increases by 100× the effective surface area (104). A second strategy is to target host cell extensions (pili and flagella). This strategy is common in bacteriophages and has also been observed for the archaeal virus SIRV2 (50). Finally, a virus could decorate its capsid with long extensions that mimic cellular pili; essentially a pilated virus.

The research presented here represents two largely independent projects, united under the common theme of: Structural adaptations in archaeal viruses for promoting virus attachment. In one project, the goal being to characterize the attachment of STIV onto its *Sulfolobus solfataricus* host. In another project we discovered a totally new archaeal virus (TSPV1) with hair-like extensions decorating its capsid. We sought to identify these unusual fibrous extensions, which we suspect serve an attachment function for TSPV1.

Research Questions

The 1st question of this research: How does STIV achieve specific, efficient attachment to its host cell? To this end, we developed a viral binding assay to measure the kinetics of STIV-1 attachment. Then, employing cryo-electron tomography we visualized STIV attached to host cell pili. Finally, using the previously determined crystal

structure of the viral turret protein, we produced a pseudo-atomic model the virus-host interaction. This represents the 1st data collected on attachment for STIV, or for any other icosahedral archaeal virus.

The 2nd question of this research: What receptors does STIV target to facilitate attachment and/or entry? To this end, we successfully isolated a new STIV-type virus (STIV-NG05), and a naïve, susceptible host cell. Resistant mutants of the host cell were generated, sequenced, and compared against the original sensitive strain. A cluster of genes was identified, containing mutations from multiple, independently isolated, resistant strains. Since mutation of genes in this locus provides complete resistance to STIV infection, this region is likely responsible for production of an STIV receptor protein/structure. Bioinformatic analysis allowed identification of this region as a new pili locus, likely responsible for production of the pilus structure to which STIV binds.

The 3rd question of this research: What is the composition of the fiber structures observed on the newly discovered TSPV1 archaeal virus? TSPV1 contains unusual fiber extensions from the capsid that are likely involved with attachment. Characterization of the structures was challenging given their incredible physical, chemical, and biologic stability. We were successful in isolating separated fibers, purified away from the rest of the viral capsid. Denaturation of these fibers was achieved using high temperature and saturated levels of urea. This enabled proteolysis and LC-MS based sequencing of peptides, which mapped back to a single viral gene product.

CHAPTER TWO

THE MOLECULAR MECHANISM OF CELLULAR ATTACHMENT FOR AN ARCHAEAL VIRUS

Introduction

Once thought to occur only in extreme environments, Archaea are now estimated to constitute 20% of the total biomass on earth (129). In oceans, for example, Archaea constitute 30% of prokaryotic diversity in the pelagic zone, where members of the archaeal phylums Crenarchaeota and Thaumarchaeota represent the oceans most abundant phyla (130, 131). Further, more than 80% of the biomass in marine subsurface sediments is also archaeal (132). Thus, it is not surprising that Archaea play significant roles in global nitrogen and carbon cycles (133, 134), including a predominant role in ammonia oxidation in soils (135, 136) and an exclusive role in methanogenesis *(137)*, which is thought to be a strictly archaeal metabolism. In addition, Archaea are also an important, though poorly understood component of the human microbiome (134, 138). For these reasons, there is now significant interest in Archaea, and by extension, archaeal viruses (54, 131).

Sulfolobus turreted icosahedral virus (STIV-1) is among the best studied archaeal viruses, where it serves as a model system for viruses in this 3rd domain of life (41, 81, 83, 111, 127, 128, 139-150). STIV-1 packages a 17.5 kbp circular double stranded DNA genome, encoding 37 open reading frames (ORFs), within a 75 MDa icosahedral shell built upon a pseudo T=31d icosahedral lattice (41, 111), with an internal lipid envelope

present between the nucleic acid core and the outer protein shell. The major capsid protein adopts the double jelly roll fold, which places STIV within the Adenovirus-PRD1 viral lineage, and suggests the possibility that these viruses derive from a common ancestor existing more than 3 billion years ago, before the divergence of Bacteria, Archaea, and Eukarya (41, 81, 83).

Its namesake turret-like assemblies are present at the 12 five-fold vertices. Pentameric assemblies composed of three viral proteins, A55, A223 and C381, constitute each turret (81, 111, 144). A223 adopts a non-canonical double jellyroll fold and positions the N-terminal jellyroll domain within the capsid shell at the penton base. N-terminal extensions on the A223 pentamer project into the capsid, where they interdigitate with the C-terminal end of the A55 pentamer to form a 10-stranded, hemolysin-like, anti-parallel β-barrel, with the N-terminus of A55 firmly anchored in the internal lipid layer (81). In contrast, the A223 C-terminal jellyroll domain projects outward from the exterior surface of the capsid where it interacts with C381 to collectively form the exterior turret complex, which protrudes 13 nm above the capsid surface (81). At times a fourth protein, thought to be C557, is also present as "petals" that decorate the STIV turrets (41, 111, 144). The function of C557 remains an enigma (41, 143). Interestingly, in STIV-2 the A223 and C381 homologs are present as a fusion into a single turret protein (151).

Whole cell cryo-electron tomography (CET) identified mature virions, procapsids without genome cores, and partially assembled particles in the cytoplasm of the infected *Sulfolobus* host (142). This suggests the capsid and inner membrane of STIV-1 co-

assemble in the cytoplasm to form a procapsid, that is then packaged with DNA (142), presumably through a unique vertex that includes B204, the presumptive packaging ATPase (152). This assembly model also requires that the lipids constituting the viral membrane are recruited directly in the cytoplasm, rather than by a budding mechanism from the host membrane during virion release.

Following assembly, STIV-1 then utilizes a remarkable, seven-sided pyramid-like structure formed by a single viral protein (C92) to lyse the cell (128, 139). The base of this 7-sided pyramid is embedded within the cyclic tetra-ether lipid membrane of the host cell and extends 20-150 nm above the extracellular proteinaceous surface-layer (S-layer). During lysis, the seven leaflets of the pyramid open, producing ~100 nm holes through which the newly formed virus particles exit the cell (127, 128, 139, 142, 149).

In contrast to assembly, genome packaging, and release, our understanding of STIV attachment and entry are limited, and speculative (41, 81, 144, 151). Still, we suspect that the STIV turret structures are involved in host cell recognition based upon the fact that evolutionarily related bacteriophage PRD1 and the eukaryotic adenovirus have similar spike protein structures built from jellyroll domains, that are known to mediate host receptor binding (153-155). For this reason, we undertook studies on the initial attachment of STIV to *S. solfataricus* and show here that initial attachment is to host pili. Using the previously determined structure of STIV-1 by high resolution cryo-EM (81), a high resolution crystallographic structure of C381 described here, and cryo-electron tomography (CET) with subtomographic averaging, the interaction is localized

to conserved residues lining a cleft formed by the second and third jelly-roll domains of C381.

Materials and Methods

Quantification of Viral Attachment

Sulfolobus solfactaricus strain P2^3 was grown to early log phase at 80°C and pH 2.5 as previously described (148). STIV was added to cells at a viral genome to cell ratio of 2 in a 500 µl microfuge tube. The virus and cells were incubated at 80 °C with cells for times ranging from 0 – 40 minutes. Cells were then separated from unbound virus by three rounds of pelleting the cells at 5,000 × g for 5 minutes, discarding unbound virus in the supernatant, and resuspending the cells in media. After the final resuspension, cells were cultured for 24 hours, which represents one cycle of viral replication. To estimate adsorption rate, The adsorption rate constant k was calculated using $k = 2.3/Bt \times \log10 (Pmax /Pt)$, where B = concentration of host cells as number of cells per milliliter, Pmax = average maximum viral yield obtained, and Pt = viral yield at time t (156).

Phylogenetic Tree and Identification of Conserved Residues in C381

Using the C381 sequence of STIV-1 or the analogous region of B631 from STIV-2, Blast searches were performed on a collection of metagenomic sequence data from Yellowstone National Park thermal features (157), cellular isolates and publicly available sequences using both tBLASTn and psi-blast programs (158, 159). Extracted protein sequences were aligned with ClustalW (Blosum Matrix, gap open 15, gap extend 1) and realigned with MUSCLE (default settings) (160, 161). Sequences greater than 75% of

full length C381 were then used to build a phylogenetic tree with MEGA X using a Maximum Likelihood method and JTT matrix-based model over 2000 bootstrap iterations (162). The sequence alignment used to construct the phylogenetic tree is shown in Supplemental Figure 7. All sequences belonging to the STIV-1 clade, including those less than 75% of full length, were subsequently realigned with ClustalW and used to identify strictly conserved residues (Supplemental Figure 8). The accession numbers for all sequences are indicated in Supplemental Figures 3, 7 and 8.

Virus Binding Assay and Imaging

S. solfactaricus cells (1 ml) at approximately 5×10^8 cells/ml was combined with 100 µl of 1.88 mg/ml virus and incubated for 20 minutes at 80 ºC. Unbound virus was removed by centrifugation (5,000 × g for 5 min). Final resuspension was in 15 µl of 5 mM Citrate at pH 2.5. Cells with bound virus were stained with 2% uranyl acetate on carbon coated grids prior to transmission electron microscopy (TEM) using a LEO 912AB with 2Kx2K CCD camera.

Cryo-Electron Tomography

Cryo- electron tomography was conducted at the Max Planck Institute for biochemistry. Cells with bound virus in 5 mM citrate were obtained as described above and frozen on Quantifoil R2/1 holey carbon grids (200 mesh copper) in liquid propane/ethane using a Vitrobot Mark III (FEI, Hillsboro, OR). Tilt series were taken using a Titan Krios (FEI, Hillsboro, OR) operated at 300 kV and equipped with a Volta phase plate (163, 164), a Quantum post-column energy filter (Gatan, Pleasanton, CA),

and a Summit K2 camera (Gatan, Pleasanton, CA) operated in electron counting mode. Tilt-series images were collected using SerialEM (165). Individual frames of images acquired with the K2 camera were aligned in Digital Micrograph (Gatan, Pleasanton, CA). Acquisition parameters were: magnification 35,700 X; tilt range +/-60°; tilt increment 2°; total dose ~60 $e^-/Å^2$; pixel size 0.14 nm/px; defocus with phase plate -0.25 μm. The Volta phase plate was operated as previously described (166, 167).

Tomograms were reconstructed using IMOD 4.7 (168). Contrast transfer function (CTF) corrections were not performed. Gold nanoparticles were used as fiducial markers for alignment of tilt-series projection images. Local alignment solutions were used for rotation, magnification, and tilt angle, while a global solution was employed for distortion. Gold particles were erased from the final aligned stack. Images were binned 2x, radially filtered with a cutoff of 0.4 and falloff of 0.05, and run through 10 SIRT iterations. Alternatively, images were binned 2x but left unfiltered and weighted backprojection used for the reconstruction. These unfiltered weighted back projections were used for subtomographic and icosahedral averaging of STIV.

Subvolume Alignment and Averaging for STIV

Virus particles were aligned from the filtered tomograms using Particle Estimation for Electron Tomography (PEET) (169, 170), with the 3.9 Å structure previously determined by single-particle analysis used as an initial reference (81). The motive list was then used to align and extract virion sub-volumes from the unfiltered tomograms. Excised particles were translated using MAPMAN in the RAVE averaging package (171, 172) such that the center of the volume/particle was at the origin. Particles

were then subjected to 60-fold averaging with strict icosahedral symmetry using AVE (171, 172). Finally, the density for 6 icosahedrally averaged particles were summed within Chimera using VOP (173).

Subvolume Alignment and Averaging for Pili

Model points were manually placed along host pili with additional points then being added every 5 nm with addModPts. This resulted in 1774 sub-volumes that were then aligned and averaged with PEET (169, 170). Particle model points were used to specify the Y-axis and random axial rotations were used for the initial motive list. For a reference, a cylindrical volume with a diameter of 6 nm was generated with the Shape command of Chimera (173). To prevent misalignment near viral particles, density associated with the capsid was set to zero.

Subvolume Alignment and Averaging of the Turret/Pilus Interaction

Viral turrets located near host pili were chosen as candidates for pilus/turret junction averaging. Identification of these sites was further aided by using the clonevolume program of PEET (169, 170) to place averaged virus density into the tomogram. Model points were placed at the center of each virion and at the center of each candidate turret. The stalkInit program was then used to calculate the initial motive list and rotation axis. The entire capsid was utilized for alignment of the chosen turret along the y-axis with PEET (169, 170). At this stage Theta and Psi were determined, and due to the five-fold symmetry of the turret, only 5 possible solutions remained for Phi.

To build an initial reference, three interactions were chosen with long straight sections of pilus traversing them. These were skewed into a common orientation using the appropriate rotation matrixes and the average command of AVE from Uppsala Software (171, 172). After generating the initial reference, candidate turrets were rotated about the y-axes in 72° increments (exploiting the 5-fold symmetry of the turret) to search for correct alignment to the reference. A spherical mask was used to score only density near the turret pilus junction. Background noise levels were determined using 30 dummy particles with no pilus.

Fitting the Atomic Structure of STIV into the Subvolume Average

Subvolume averages were determined to be larger than the atomic model necessitating scale correction. Using Alterheader from IMOD, a series of resized capsids were generated in 0.01 Å increments by modifying the pixel size. These resized capsids were then resampled onto the original volume using Chimera (173). Cross correlation values were calculated between each resized sub-tomogram and the 4.5 Å single particle reconstruction using PEET with search angles and distance set to zero (169, 170). Alternatively, Fourier shell correlations were calculated using algorithms from EMAN2 and Imagic implemented in the PDBe Fourier Shell Correlation server (174).

Isolation of a New STIV-subtype Virus-Host System

The host of STIV-1 (S. solfataricus P^3) is an unstable genome with a high proportion of mobile genetic elements that frustrate attempts to perform site directed mutagenesis on these cells. In contrast, *Sulfolobus acidocaldarius* has a stable genome

and proven techniques for introducing mutations (175). Furthermore, *Sulfolobus acidocaldarius* strain NG05A.C05.01 contains an integrated STIV-type viral genome (STIV-NG05) (176). The following procedure outlines our successful strategy for induction of this integrated virus and infection of a closely related naïve strain. *Sulfolobus acidocaldarius* strain NG05A.C05.01 was cultured as previously described in pH 3.5 DT media aerobically in glass flasks in a shaking oil bath at 75° C (175). To induce the integrated virus, at a cell density of 1E9 cells/ml pH of the media was lowered to 2.5 by direct addition of 0.1M HCL, and cultures were transferred to a dry incubator at 80° C. Virus production was monitored by qPCR of the culture supernatant using the primers 6forward59-62 CATGTAGCATTCTTGCAGCCTGGACA and 6reverse54-59 TTGAGCTGGTAACGCTGCGAGT. At peak viral production (generally 20 hours) cells were removed by centrifugation (5000 × g for 5 min), and virus in the resulting supernatant collected as a inoculation source.

Harvested virus was used to inoculate naïve cells of the closely related *S. acidocaldarius* NG05B.C03.01 strain (176). Infection of naïve cells again required a combination of temperature and pH shock. Cells were cultured as described above, then the pH was reduced to 2.5 and virus added to cells at a viral genome to cell ratio of 2. Finally, cultures were incubated at 80° C for 20 hours at which point cells were removed by centrifugation. This process was repeated over several iterations to let STIV-NG05 evolve to the new host strain.

Generation of STIV Resistant Host Mutants

To generate strains of S. *acidocaldarius* strain NG05B.C03.01 cellular resistant to viral infection we used an iterative process of infecting initially naïve cells, isolating resistant survivors, and then re-infecting the surviving cells. Briefly naïve S. *acidocaldarius* NG05B.C03.01 cells were cultured and infected as described above. At 24 hours post infection dilutions of the culture were plated onto DT media plates. Colonies produced were inoculated back into liquid media reinfected and tested for viral production. This process was repeated through 3 cycles of infection and isolation of resistant cells. At each stage isolates were screened by both qPCR and electron microscopy for the presence of STIV to verify their resistance to viral infection.

Sequencing Resistant Mutants and
Identification of Mutation Sites

Four independent STIV resistant strains and one wild type STIV sensitive parental isolate (never exposed to virus) were selected for genome sequencing. DNA was extracted from the cells using the MoBio PowerWater kit and DNA was then submitted to the sequencing center of Illinois State University, Urbana Champaign, IL. Sequencing was performed using Illumina NovaSeq 6000 2×150 paired reads. Quality control and cellular contig assembly was conducted using SPAdes version 3.9.0 (177). Reads from the resistant strains were subsequently mapped on to the sensitive parental strain using Breseq (178). In addition, contigs from resistant mutants were compared to the sensitive genome. Mutated genes and neighboring genes were analyzed by eggNOG-mapper, HHpred, and FlaFind to assign functional annotations (179-181).

Results

The C381 Crystal Structure

The 1.9 Å C381 crystal structure was determined by Dr. Brian Eilers. Here we present a detailed analysis of its structure. Within the context of the viral turret, the N-terminal jelly-roll domain of C381 is closest to the capsid surface, interacting with the C-terminal domain of A223, which forms the penton base in the STIV capsid. In contrast, the C-terminal jelly-roll is distant from the capsid surface, and is present at the tips of the turrets, with the second jelly-roll domain intermediate between the N- and C-terminal jelly-roll domains (Figure 2.1).

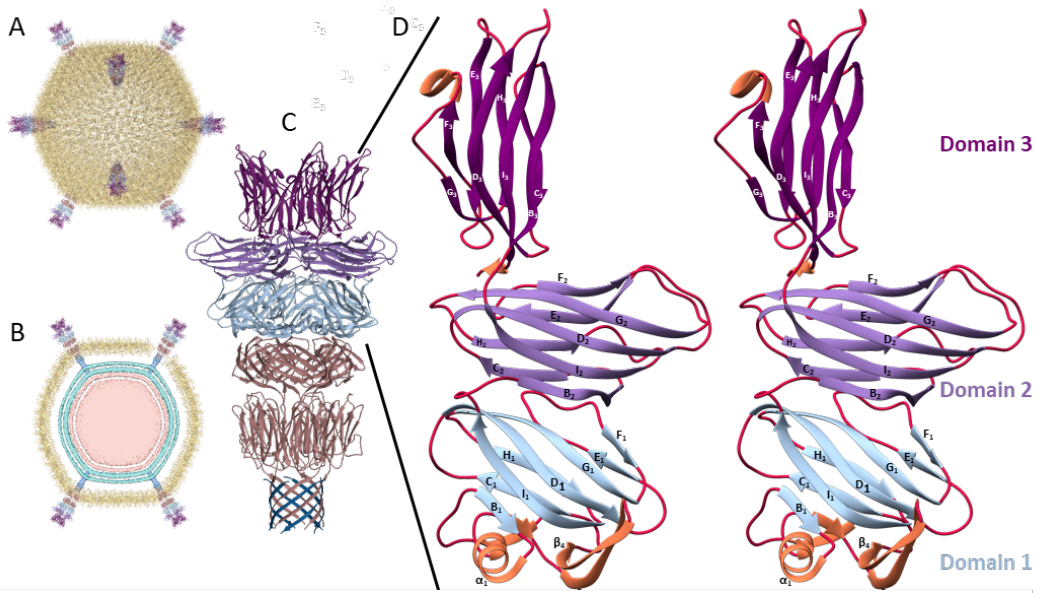


Figure 2.1. Structure of STIV C381. A) The composite atomic structure of STIV-1 from high resolution single particle analysis and crystallographic studies of the turret proteins. The major coat protein, which exhibits a double jellyroll fold and forms a

pseudo T=31L lattice is in wheat, while proteins composing the turret in are various shades of blue and purple. B) A cross section through the atomic mode and internal structure from single particle analysis reveal the internal lipid membrane in light blue, and density for the dsDNA genome (pink). Anchoring of basement layers of the turret structure are anchored in the internal membrane. C) The structure of a single turret. Elements of the A55 membrane anchor are at the bottom in blue where the interdigitate with the N-terminus of A223 to form a 10-stranded hemolysin-like beta-barrel. The 2-domain structure of A223 is shown above that in pink, followed by the 3-domain structure of C381 in light blue, light purple and dark purple at the top. D) Stereo pair of the C381 subunit with the BIDG and CHEF sheets of each domain labeled. The relative position of each domain is such that the BIDG sheets of each domain are toward and the CHEF sheet away from the viewer. This domain organization results in solvent exposed BC, DE, FG and HI loops, while the CD, EF, and GH loops in domains 2 and 3 face the turret axis, and for domain 3 are at the interface with domain 2. Domains 2 and 3 are minimal jellyroll folds, while domain 1 harbors several embellishments involved in subunit interactions.

The canonical viral capsid jelly-roll fold is a β-sandwich, with two 4-stranded anti-parallel β-sheets juxtaposed against each other (182-184). Historically, the first β-sheet is composed of strands B, I, D and G (BIDG), and the second of strands C, H, E and F (CHEF) (182, 184). Relative to this canonical fold, the second and third domains of C381 are largely devoid of additional secondary structural elements (Figure 2.1), though a 2 residue β-bulge is present in strand B of the second domain (B2), giving the appearance of a bipartite strand that is analogous to the bipartite strand B in domain 1 of the STIV coat protein (83).

In contrast, the first jelly roll domain houses additional secondary structural elements. An N-terminal β-strand (β1) precedes strand B1, and an α-helix (α1) and three successive β-strands (β2, β3, β4) are inserted between stands B1 and C1. Strands β3 and β4 form an extended 2-stranded antiparallel β-finger. The short β2 strand runs antiparallel to β-strand C1 strand along the outside edge of the C1H1E1F1 sheet.

Similarly, the C-terminus of β4 briefly runs antiparallel to the N-terminal end of B1. Collectively, these N-terminal embellishments facilitate interactions with the C-terminal domain of A223, present at the penton base (81), and augment the inter-subunit domain1-domain1 contacts within the C381 pentamer (Figure 2.1), where strand β3 in the β-finger runs antiparallel to β1 in the neighboring subunit, to form an intersubunit 3-stranded antiparallel β-sheet.

Within a single subunit, the first two jelly-roll domains are oriented with the CD, EF and GH loops projecting in towards the 5-fold axis, and the BC, DE, FG and HI loops facing outward. And, overall, the relative orientation of the two domains with respect to each other is similar to that in the major capsid proteins for STIV, PRD-1 and adenovirus. In these classic double jellyroll folds, the BC, DE, FG and HI loops all face the exterior surface. In addition, the two domains are aligned parallel to each other, and include a few H-bond interactions between strand F of the first domain and strand B of the second domain. Thus, the two-domain fold can be visualized as a central, 8-stranded β--sheet (CHEFB'I'D'G') that is flanked on each face by the 4-stranded BIDG and C'H'E'F sheets, respectively. Indeed, the arrangement is similar in C381, however, the β-bulge present in strand B2 of the second domain results in an antiparallel rather than parallel interaction between strands F1 and B2 (Figure 2.1 and Supplemental Figure 1).

In contrast, the third domain is rotated 90 degrees relative to domain 2. But like the preceding domains its B3C3, D3E3, F3G3 and H3I3 loops are solvent exposed, projecting up and outward at the tips of the turret, while the 3 loops at the opposite end of

the sandwich (C3D3, E3F3, G3H3) pack against the exposed edges of strands G2 and H2 (Figure 2.1).

The C5 rotational symmetry of the pentamer results in three pentameric rings composed of domains 1, 2 and 3 respectively. Within the pentameric domain 3 assembly, we were fascinated to find that adjacent subunits recapitulate the inter-jellyroll interactions described above for the STIV, PRD-1 and adenovirus MCPs; specifically, the F3 strand in each subunit runs antiparallel to the B3 strand of the adjacent subunit, resulting in an 8-stranded inter-subunit β-sheet that spans each subunit interface (Figure 2.1 and Supplemental Figure 1). The overall, the organizational similarity of domains one and two to the classic double jellyroll MCPs in STIV, PRD-1 and adenovirus suggest the first two domains of C381 may have evolved from a double jellyroll MCP. Similarly, the conserved intersubunit interactions of the 3rd domain may suggest that, while a subsequent addition, it too is derived from a double jelly-roll capsid protein.

Structural Similarity

Each of the individual domains was submitted for a DALI search to identify structural similarity with other proteins (185). Domain 1 shows greatest similarity to the carbohydrate binding module (CBM) of endo-1,4-beta-xylanase and the ligand binding domains of other hydrolases [PDBID 2WYS, Z=9.2 (186)]. In this non-catalytic carbohydrate binding domain, a polysaccharide binding groove runs across the solvent exposed face of the CHEF-sheet, perpendicular to the direction of the β-strands (Supplemental Figure 2). When super-positioned on the C381 pentamer, this "carbohydrate binding groove" is occupied by the β-finger of a neighboring C381

subunit. The structural equivalent of the "carbohydrate binding groove" in C381 domain 1 is thus utilized in constructing the C381 pentamer, where it is an important element of the subunit interface (Figure 2.2).

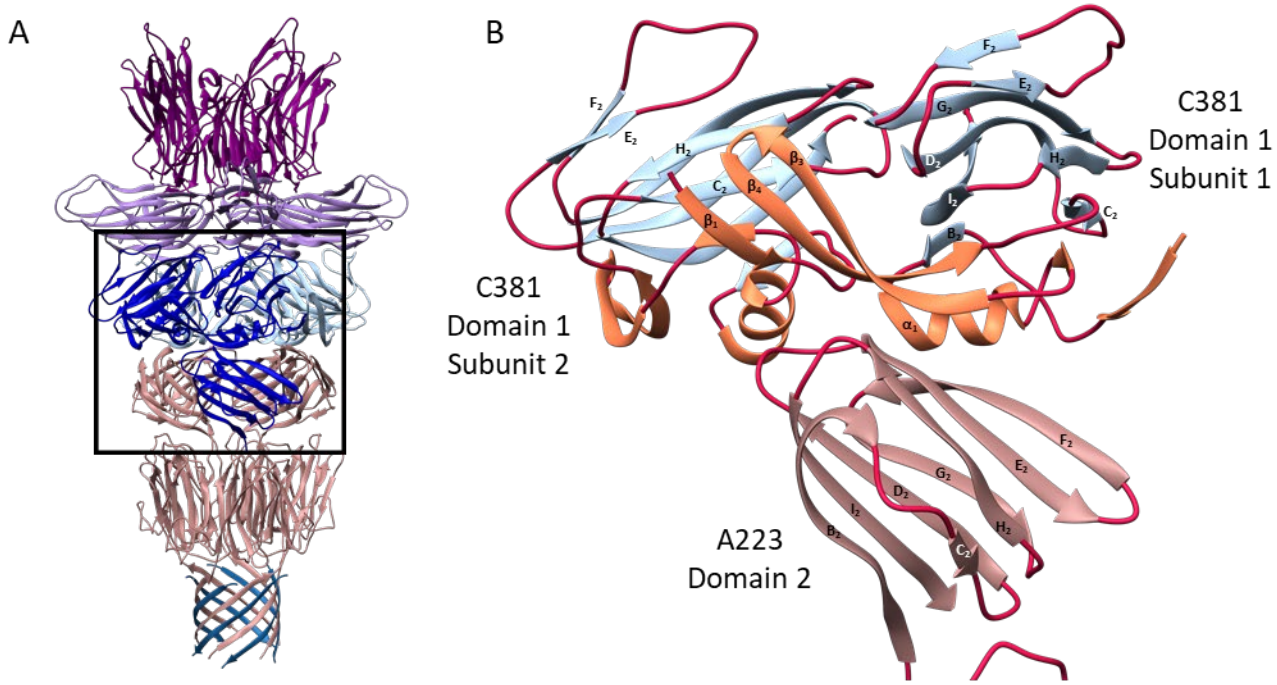


Figure 2.2. C381 Domain 1 Interactions A) Structure of the pentameric STIV turret with specific C381 domain 1 inter-subunit interactions highlighted in dark blue (black box). B) Enlargement of the three highlighted domains in the boxed region of panel A. The β3/β4 beta-finger of C381 domain 1 subunit 1 packs across a shallow grove of the domain 1 CHEF sheet in the adjacent (second) subunit. β4 runs parallel to β1 of the adjacent subunit to give a mixed 3-stranded β-sheet at the subunit interface while helix α1 of C381 interacts with strands E2 and F2 (pink) of A223 (a double jelly roll) to anchor the C381 pentamer to the A223 penton base.

In contrast, domains 2 and 3 are most similar to other viral proteins. For domain 2, the greatest similarity is found to the knob domain of the tail needle tip in podovirus HS1 [(Supplemental Figure 2), PDB ID 4K6B, Z=8.5, (187)]. Like other Podoviridae,

this phage has a long thin shaft for a tail needle that serves as a DNA exit channel plug. The knob domain is also distal to the viral capsid, in a position that would allow first contact with the outer bacterial membrane during adsorption. However, genetic alteration of the knob domain does not affect host specificity (presumably controlled by the 6 tail-spike proteins) and instead suggests a role in controlling DNA ejection into the host. Still, P22-like knob domains are structurally similar to the fiber knobs of bacteriophage PRD1 and Adenovirus, which do mediate primary cell attachment (187, 188). Overall, then, these structural similarities suggest potential roles for C381 domain 2 in viral attachment and/or genome release.

Greatest structural similarity to domain 3 is found with the protrusion or "P" domain of the infectious bursal disease virus (IBDV) VP2 subviral particle [(Supplemental Figure 2), PDB-ID 2DF7, Z=8.2, (189)]. The P domain of IBDV VP2, also presumed to be involved in interaction with cellular receptors (189), is rotated by 90 degrees relative to the shell or "S" domain of VP2, and sits end-on atop the S domain jellyroll fold, utilizing its CD, EF and GH loops to mediate the interaction . In this context, the S and P domains of the IBDV VP2 trimer appear analogous to the second and third domains of the C381 pentamer, again suggesting a role for C381 in viral attachment.

Conserved Residues

To identify C381 residues serving conserved functions, we queried available hot spring metagenomes and genomic data for relevant archaeal organisms and viruses with STIV C381 and its homolog from STIV-2 (151). The search returned a total of 65 C381-

like sequences. Sequences at 75% full length or greater were used to construct a phylogenetic tree that showed the presence of three distinct clades; one representing sequences more closely related to STIV-1, another more closely related to STIV-2, and a third clade distinct from either of these (Supplemental Figure 3). Intriguingly, sequences belonging to both the STIV-1 and STIV-2 clades were found to coexist in the same thermal feature. A multiple sequence alignment with all sequences from each of the three clades revealed a paucity of strictly conserved residues. However, when the alignment was confined to the 55 STIV-1-like sequences from clade 1, 143 of the 381 residues (37.5%) were strictly conserved (Supplemental Figures 4).

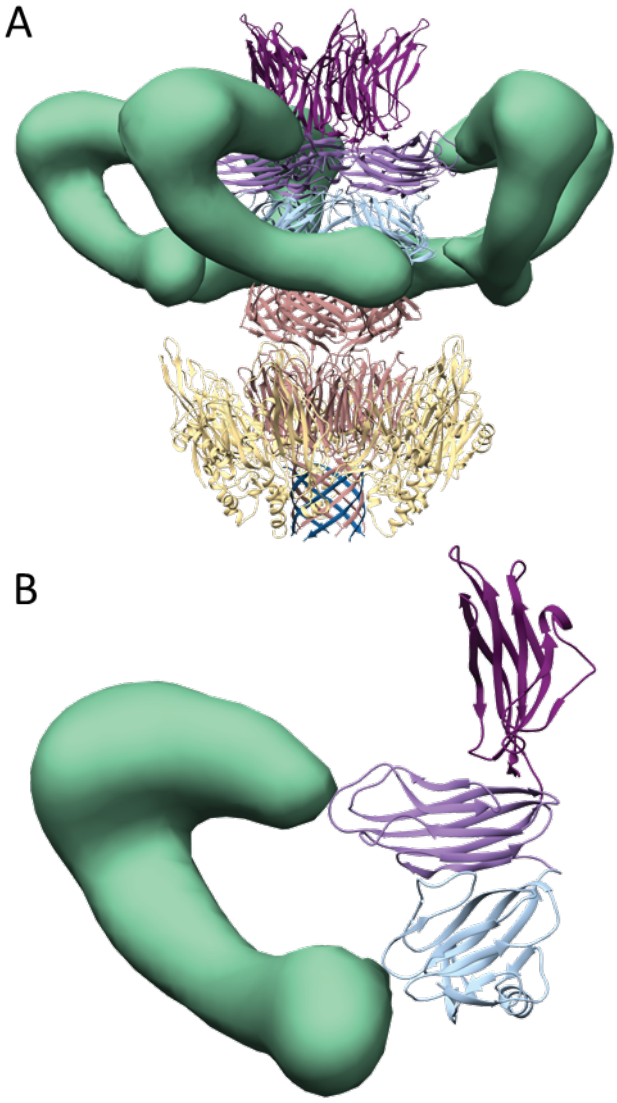


Figure 2.3. C381 – Petal Protein Interactions. A) A difference map for the turret petals, thought to be formed by the C557 protein, is shown in green juxtaposed to the atomic structure of the C381 turret protein. The view of C381 is "opposite" that in Figure 2.1, with the CHEF sheet of each domain toward the reader (instead of BIDG) . The coloring scheme is the same as in Figure 2.1C except the B345 major capsid protein subunits immediately surrounding the turret are shown in wheat, to illustrate the location of the capsid shell. B) Density from the petal protein intersects the D_1-E_1 and D_2-E_2 loops of C381.

The C381 Subunit Interface

As expected, a majority of the strictly conserved side chains are buried within the hydrophobic cores of the jelly roll domains, presumably serving a largely structural role. However, 54 strictly conserved residues were found at the subunit interfaces. Of these, 12 were positively charged (9 Lys, 3 Arg), 16 were polar (5 Tyr, 4 Asn, 4 Gly, 2 Thr, 1 Ser) and 16 negatively charged (10 Glu, 6 Asp), at least at the intracellular pH of *Sulfolobus* (pH 6.5) where STIV-1 assembles (142, 190). Only 9 were hydrophobic (2 Pro, 2 Trp, 1 Ala, 1 Phe, 1 Val 1 Leu, 1 Ile).

In this context, the total surface area per subunit is ~17,000 Å2 with ~2,400 Å2 buried at each of the two contact surfaces, giving a total buried surface area of 4,800 Å2 (38%) per subunit. Consistent with the large buried surface area, PISA (191) gives a complex formation significance score of 0.994. PISA also identifies 46 H-bond and 20 salt bridge interactions at each of the interacting surfaces, again indicating the hydrophilic nature of the interface. The subunit interface surfaces of the first and second domains are largely devoid of conserved hydrophobic side chains, leaving 6 of 9 conserved hydrophobics at the domain3/domain3 interface. These residues form a small hydrophobic interface between strictly conserved Val330, Trp331, Ala335, Pro341 and Trp344 on one subunit with strictly conserved Phe271 (and conserved hydrophobics at 370 and 372) on the neighboring subunit. Notably, Phe271, Val330 and Trp331 lie within strands F3 and B3 at the inter-strand interface between adjacent subunits (Supplemental Figure 1B).

The A223 Interface

In contrast to the subunit interfaces within the C381 pentamer, the C381 interface with A223 shows only 4 strictly conserved residues (Domain 1: Asn33, Tyr40, Leu42, Thr89), and thus is not well conserved. However, like the inter-subunit C381 interface, it is largely hydrophilic in nature. Overall, the hydrophilic C381 subunit interfaces and corresponding lack of solvent exposed hydrophobic residues likely plays an important role in stabilizing the monomeric form of the recombinant protein.

The Petal Interface

A prominent feature of the initial 27 Å resolution single particle reconstruction of STIV (EMDB 5584) are the petal-like decorations on the viral turrets (41). The only capsid protein with a molecular weight near the predicted 63 kDa mass of these decorations is C557 at 59.4 kDa (111, 144). Later reconstructions showed either absence or partial occupancy of the petals at each turret, inhibiting further structural analysis of the petals (81, 144). To identify interactions between C381 and the petals, a difference map was calculated by subtracting the more recent high-resolution reconstruction from the earlier 27 Å reconstruction. The difference map indicates the petals recognize the 1st and 2nd jellyroll domains within a single C381 subunit (Figure 2.3). Specifically, petal density intersects the B1 C1 and D2 E2 loops of C381. The interaction appears to be mediated by non-conserved hydrophilic residues, although some conserved residues present in the vicinity may also serve as contacts. These include the aromatics Y186, F/Y194, and Y/F245 in addition to the polar N188 residue. Given the small contact

surface area between C381 and the petals, it is not surprising that these structures are lost during virion purification or shed as a result of virus maturation.

Virion Adsorption Kinetics

We previously suggested that the STIV turrets are involved in host cell recognition and attachment (144) . The structural similarity of C381 domains 2 and 3 to proteins involved in host cell attachment and genome release such as HS1 and Sf6 of the P22-like phage (187, 192) and VP2 of IBDV (189), as well as the roll of the knob-like proteins of Adenovirus (154, 187, 188, 193-195) and PRD1 (187) strengthen this hypothesis. For this reason, we undertook kinetic and structural studies of initial attachment of STIV to *Sulfolobus solfataricus* P2^3, a preferred hyperthermophilic, acidophilic host derived from *S. solfataricus* P2 by successive rounds of isolating infection permissive cells (148).

Typically, plaque assays are used to determine virus absorption rates to host cells. In our experiments plaque assays proved unreliable. Therefore, we developed a viral adsorption assay where the number of viral genomes serves as a proxy for the number of infected cells. In this assay, virions were allowed to attach/enter the cells for a defined length of time followed by removal of unbound virus through centrifugation and washing. The remaining tightly bound virus was permitted to complete one replication cycle before the number of viral genomes were measured by qPCR. Assuming virus production correlates directly to the number of infected cells (the same assumption is needed for plaque assay based determination), this allows relative comparison of the number of infected cells as a function of incubation time. Saturated adsorption was

found to occur at 10 minutes (Figure 2.4), and the adsorption rate for STIV was 2 × 10 -9 ml min -1 (Supplemental Table 1). The process was temperature dependent, as binding at room temperature was not observed, even with incubation times up to 1 hour. While slower than that reported for other hyperthermophilic archaeal viruses like SMV1 and SIRV2, it is faster than rates observed for halophilic archaeal viruses (50, 51, 196), and in line with values for bacteriophage like T2 (156).

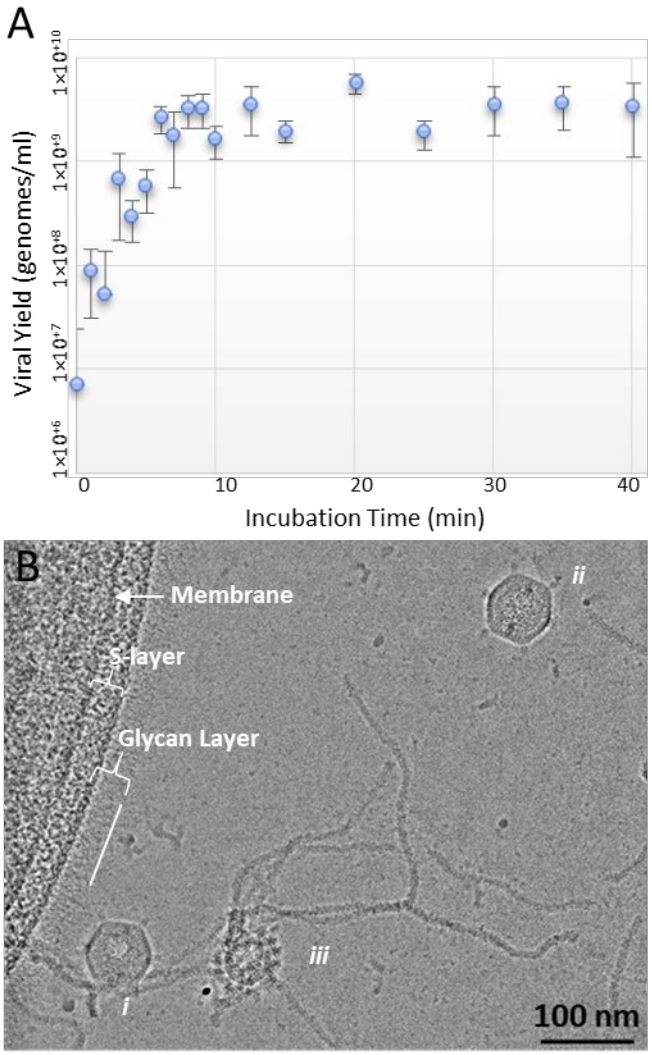


Figure 2.4. Initial STIV Attachment. A) Kinetics of viral attachment were determined by incubating virus and cells at 80°C for the specified time, then pelleting and washing cells, and quantitating viral infection by qPCR. Initial binding was complete within 10

minutes. B) Cryo-EM micrograph taken at 0.2 µm underfocus with the Volta phase plate revealed STIV bound to host pili (particles i and ii). The lipid membrane and outer surface-layer of the *Sulfolobus* host are also apparent (white arrow and small bracket respectively). The surface-layer is heavily decorated with a lawn or network of very fine filaments that likely represent an extensive glycan layer (white bracket and line). A third particle that may represent degraded or disassembled virus is also present (iii).

Initial STIV Attachment to Pili

Following incubation of host cells with STIV, virus was observed bound to host pili in both negative stain and cryo-electron micrographs (Figure 2.3, Supplemental Figure 5). The pili were highly flexible and showed an average diameter of 7 nm measured directly from micrographs. These pili resemble the 'thread' structures observed on the surface of *Sulfolobus acidocaldarius* (197).

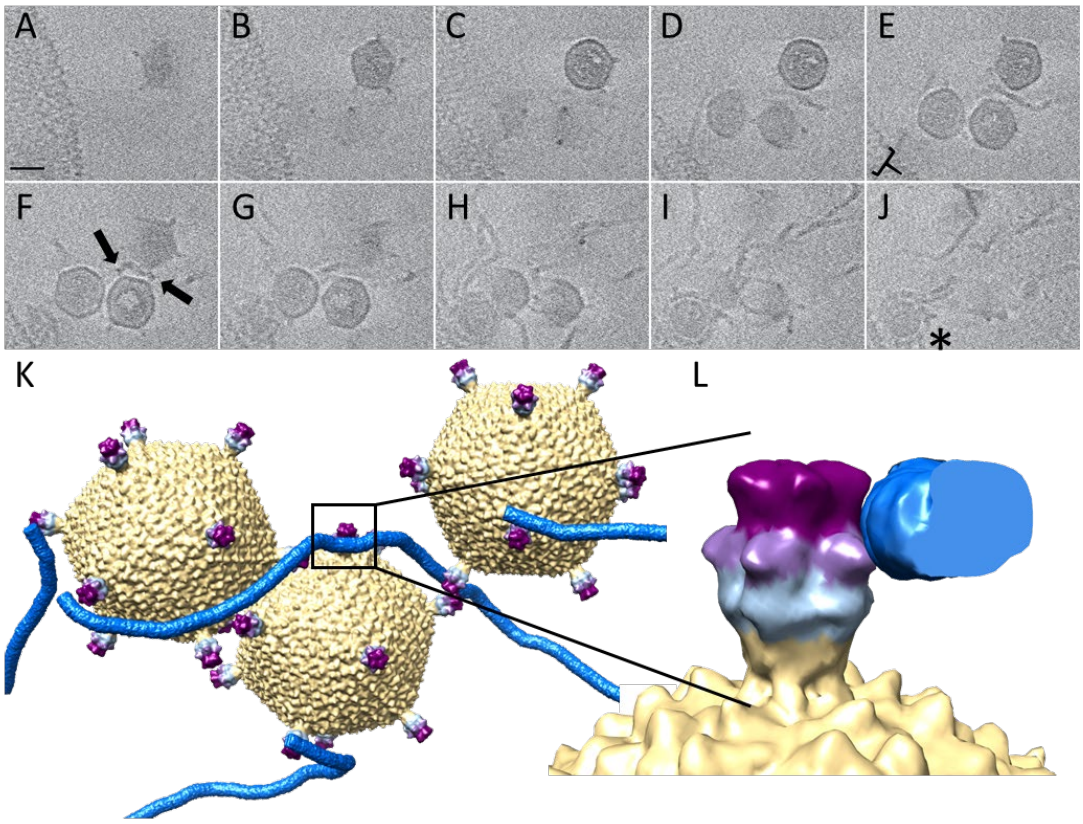


Figure 2.5. Tomograpic structure of the STIV pilus interaction (A—J) 2.25 nm thick sections at 10 nm spacings through a representative tomogram of STIV bound to host pili. Black arrows indicate binding sites between virus and pili. Three intact virions are visible along with a fourth degraded or disassembled particle (asterisk). The cell edge is visible along the left and lower left of sections; the surface-layer is marked by a bracket in E. (K) The subvolume averages for the 3 virions and the pili are shown in their relative locations and orientations within a tomogram. The point of interaction between the virion and pilus is always at the viral turret. (L) An enlarged view of the turret highlighted in panel A rotated 90^{o} about the vertical axis and looking into a cross section through the pilus. The C381 turret protein is colored light, medium, and dark purple to indicate the position of the 1st, 2nd, and 3rd jelly roll domains of the protein, respectively. Pilus recognition is mediated by 2nd and 3rd domains of C381.

Attachment is mediated by the 2nd and 3rd Domains of C381–Cryo-Electron Tomography (CET) was employed to achieve a 3D reconstruction of the STIV-pili interaction. The tomograms localized the point of contact between STIV and the pili at the viral turrets. Interestingly, a single virion was frequently observed to bind multiple pili (Figure 2.5, Supplemental Movie 1).

Subtomographic averages of the STIV/pilus interaction were pursued in order to realize a higher resolution structure that would allow us to leverage the existing atomic structures of C381 and the entire STIV virion to obtain a pseudo-atomic model for the STIV/pilus complex. Two approaches to subtomographic averaging were taken. In the first approach, subvolume averages were separately calculated for the pili, and for the virus particles, which were also averaged over their icosahedral symmetry. The viral and pili subvolume averages were then redistributed across the tomogram in their refined locations and orientation to give a composite model (Figure 2.5B). The 6 averaged STIV particles yielded a 2.0 nm resolution STIV structure as judged by a 0.143 FSC cutoff when compared to the 3.9 Å STIV single particle structure (81) (Supplemental Figure 6).

Averaging of host pili resulted in a cylindrical density 6 nm in diameter. Consistent with the lack of structural detail for the helices in single exposure micrographs, we were unable to resolve the helical sense and pitch of the pili, resulting in a low-resolution envelope for the pili. However, this approach did clearly allow localization of the pili with respect to the viral turrets. In all virus turret interactions in the tomograms, the viral turret utilizes the cleft between the second and third domains of C381 to recognize the host pili (Figure 2.5, Supplemental Movie 1).

In a second, complementary strategy, candidate turret-pilus interactions identified in the composite model described above were aligned and averaged. The entire capsid was used to align the turret of interest along the y-axis. This effectively solved for theta and psi and left only five possibilities for Phi. Rotation in fixed 72° increments then explored the five possibilities for each candidate interaction. The final average contained 10 turret-pilus interactions and displayed strong density on one side of the viral turret indicating the region of pilus binding (Figure 2.6A). As expected, density reinforces only near the turret, since the pili appear highly flexible and their paths quickly diverge, both before and after the point of contact.

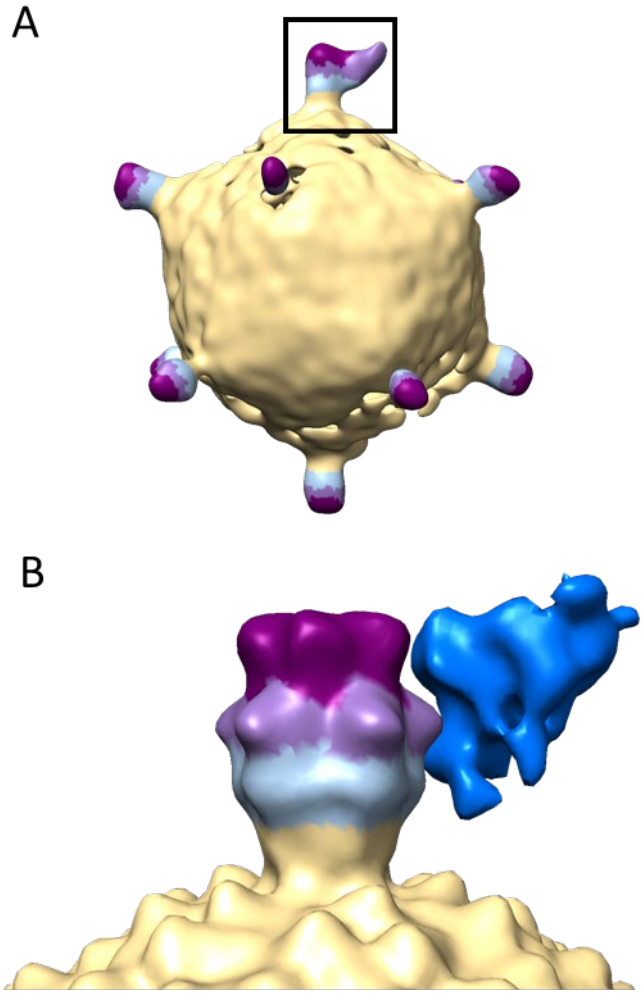


Figure 2.6. Subtomogram average of 10 STIV-Pili interactions. A) Each Turret-Pilus interactions was rotated to place the virus in a standard orientation with the identified interacting turret running along the Y axis (top turret, square). Each subvolume was then rotated about Phi (Y) in 72o increments to orient all pili bound sites. Unlike strategy 1, in which the virus and pili were treated independently, the virus particle has not been averaged over the icosahedral symmetry. Asymmetric, duck bill shaped density for the pilus at the pilus/turret contact is seen along the right side of the top of the turret (black square). B) A difference map (STIV/Pilus – STIV) was then calculated to highlight the new density due to the STIV pilus interaction (dark blue density, right side of panel B). This difference density is juxtaposed against the 6-particle icosahedral average of the tomographic virions, in which the STIV turret is colored by domain (light blue, light purple, dark purple/ domains 1, 2 and 3). A 10 SD Gaussian filter was applied to generate the map in panel A, whereas a 5 SD Gaussian filter was applied to the same map in generating the difference density depicted in panel B, thus the additional features. The averaged virion density in panel B is filtered with a 5 SD Gaussian filter.

To highlight the density contribution of the pilus, a difference map was calculated by subtracting the STIV density from the 10-particle turret-pilus average. The location of the resultant difference density relative to the single particle reconstruction of STIV and the hybrid molecular model are shown in Figure 2.6B. In agreement with the composite model described above, the strongest pilus density from the difference map lies between the 2nd and 3rd domains of a single subunit of the C381 pentamer. Regardless of the strategy utilized for subtomgraphic averaging, when the STIV atomic model is fit within the averaged density, specific regions of C381 are implicated in pilus recognition (Figure 2.7, Supplemental Movie 1). Both models suggest domain 2 utilizes strand F2, the F2G2 edge of the ß sandwich and the F2G2 loop, and that domain 3 utilizes strand C3 and the B3C3 edge of the beta sandwich to bind host pili. Surface exposed, strictly conserved residues lie within these regions and potentially mediate recognition of the pili. These include Asn196 and Ser215 in domain 2, and Glu285 and Asn289 in domain 3. Highly conserved residues Gln274 and Gln291 may also be involved. Notably, the pilus binding site appears to occlude the petal binding site in domain 2, suggesting that binding of petal protein C557 and pilus are mutually exclusive, and that the petal bound form of STIV would thus be non-infectious.

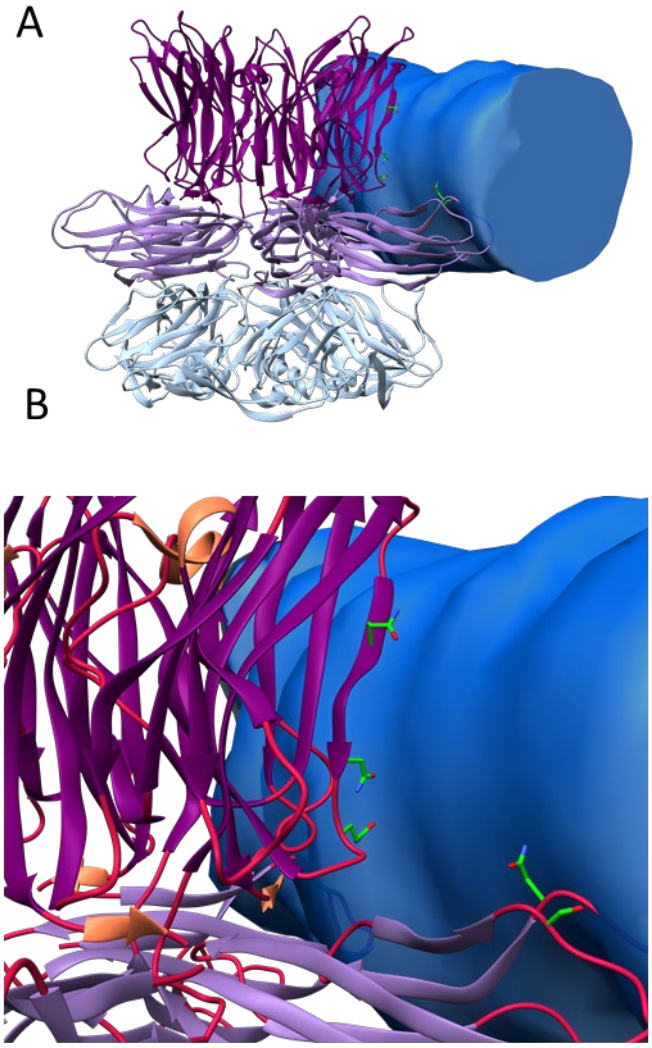


Figure 2.7. Pilus Recognition by C381 A) Domain 2 utilizes strand F_2, the F_2G_2 edge of the β sandwich and the F_2G_2 and H_2I_2 loops, while domain 3 utilizes strand C_3 and the B_3C_3 edge of the beta sandwich to bind host pili. In this model the F_2G_2 loop intrudes into the pilus density. Modest conformational changes in C381 or surface features of the filament not visible at this resolution likely accommodate the interaction. B) Surface exposed, strictly conserved residues lie within these regions and potentially mediate recognition of the pili. These include Asn196 and Ser215 in domain 2, and Glu285 and Asn289 in domain 3. Highly conserved Gln274 may also be involved.

Identification of Receptor Pili Candidates Using Virus Resistant Mutants

The host for STIV-1 (S. Solfactaricus P2) possesses high-level genetic instability from an abundance of mobile insertion elements (198). This makes generation and screening of virus resistant cells challenging, because the selected resistance mutation is likely from transient disruption of the gene by an insertion element. To circumvent this, we developed a new virus-host system based upon isolation of an STIV strain that infects *S. acidocaldarius* strain NG 05, which has both a stable genome and proven ability for genetic manipulation (175).

By using iterative rounds of infecting *S. acidocaldarius* strain NG 05 with STIV, then plating and re-screening survivors of viral infection, we produced 5 independent isolates unable to support viral infection. The genomes of these 5 isolates (along with the parental, sensitive strain) were sequenced. Cross comparison of the resistant isolates to the sensitive genome using the read mapping, SNP identification tool Breseq, revealed surprisingly few mutations (Table 2.1). We were able to identify a localized set of genes that became mutated in three of the resistant isolates NG05-2L, NG05-1N, and NG05-4Y (Figure 2.8). Homologs to the genes in this region and be found in multiple *Sulfolobus* species including *S. islandicus*, and *S. Solfataricus,* and show a high degree of gene synteny.

Using a combination of bioinformatic techniques we were able to annotate this region (Table 2.2). Three proteins with pre-pilin peptidase signals were identified by FlaFind. These are hallmarks of secreted protein components of the Type IV pili system, and a conserved feature of proteins required for assembly of filaments, such as the

archaellum and pilin proteins, on the external surface of *Sulfolobus* cells (199). Most of the proteins in this region have predicted transmembrane helices, which also supports their role as components of a Type IV pilus system (181, 199). Finally, genes SaNG 214 and SaNG 216 show distant homology to major pilin proteins, particularly the archaellum by HHpred analysis. Still, the region lacks an identifiable assembly ATPase, a necessary (and generally well-conserved/easily identified) component for Type IV pilus system (199). Still, pili systems have been identified where the ATPase is located at a distant position in the genome (199). In addition, this pili system might rely upon an ATPase encoded by another pili operon. Together this suggests that this region encodes one or more cell surface structures, that while related to Type IV pili, are distinct from the currently described structures.

Table 2.1 Genomic Coordinates for Resistant *S. acidocaldarius* strains.

Name	Type	Notes	Start	End	Length
NG05-1A	SNP CtoT		94,058	94,058	1
NG05-4Y	Deletion	Pili locus	140,940	147,765	6,826
NG05-1N	SNP CtoA	Pili locus	149,912	149,912	1
NG05-2L	Deletion	Pili locus	153,082	153,544	463
NG05-9Z	SNP AtoT		398,990	398,990	1
NG05-9Z	SNP TtoC		546,549	546,549	1
NG05-1N	SNP CtoA		547,396	547,396	1
NG05-2L	Deletion		637,354	637,354	1
NG05-1N	Insertion +C		637,714	637,714	1
NG05-9Z	SNP AtoT		687,413	687,413	1
NG05-1N	Insertion +AT	repeat region	713,247	713,247	2
NG05-9Z	SNP GtoA		749,422	749,422	1
NG05-9Z	SNP GtoC		752,097	752,097	1
NG05-9Z	Insertion +T	repeat region	996,712	996,712	1
NG05-9Z	SNP TtoC		1,038,575	1,038,575	1

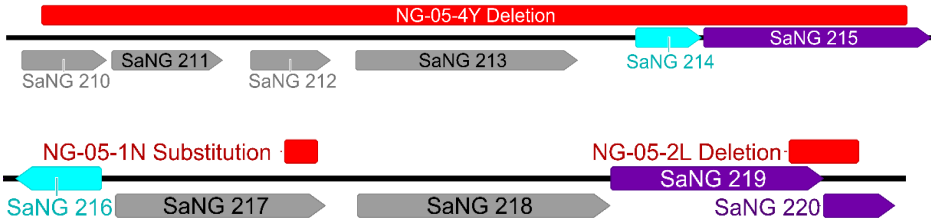


Figure 2.8. Mutated region in STIV resistant cells. Mutations identified in strain NG05-2L, NG05-1N, and NG05-4Y are shown in grey together with genes occurring in this region. Region 1 has not been identified as a pilus locus in previous studies, yet it contains hallmarks of Type IV pili systems including genes with homology to major pilins (light blue) and genes containing pre-pillin peptidase signal sequences (purple).

Table 2.2: Mutated and Neighboring Genes in STIV Resistant Cells

Gene	Mutation Type	DSM639 Homologue	arCOG group	HHpred Homology	FlaFind Signal Peptide	# TM helices
SaNG 210	NG05-4Y Deletion	gi|70607409	07227			
SaNG 211	NG05-4Y Deletion	gi|70607410	00967			
SaNG 212	NG05-4Y Deletion	gi|70607411	07227			
SaNG 213	NG05-4Y Deletion	gi|70607412	01646			
SaNG 214	NG05-4Y Deletion	gi|70607413	03854	Low homology to archaellum		1
SaNG 215	NG05-4Y Deletion	gi|70607414	14673		Yes	1
SaNG 216		gi|70607415	03855	Low homology to archaellum		1
SaNG 217	NG05-1N Substitution	gi|70607416	03854			1
SaNG 218		gi|568164228 (SUSAZ)	06002			
SaNG 219	NG05-2L Deletion	gi|70607417	06002		Yes	2
SaNG 220	NG05-2L Deletion	gi|70607418	06003		Yes	1

Discussion

C557 Petal Protein as a Maturation Factor

It is clear that STIV turrets fully decorated with petals would be unable to bind host pili. This suggests the petal protein, tentatively identified as C557 (41, 111, 144) functions as a maturation protein. In this role, newly released virions would contain petal protein at near full occupancy, preventing re-binding to pili on the infected (lysed) cell. Later these petals would dissociate, priming the virus for attachment to pili on a new host cell.

An S-layer Glycan Barrier

Most studied archaea are wrapped in a proteinaceous, para-crystalline S-layer (200, 201). An unanticipated outcome of this study is the observation in cryo-EM micrographs of a massive lawn, or network of ultra-thin filaments extending 40 μm (400 Å) or more from the proteinaceous S-layer. The S-layers in general, and *Sulfolobus* species specifically, are known to be heavily glycosylated (200-202). However, visualized in the context of STIV infection, the magnitude of the glycan barrier that must be overcome by an invading virus becomes much clearer. Not only must the virion traverse or circumvent the proteinaceous component of the S-layer and subsequent membrane, it must also first traverse (or circumvent) this substantial glycan exo-layer as well. Indeed, one of the virions (particle i) in Figure 2.4B appears immediately adjacent to this external glycan barrier.

Pilus Identity

Type IV pili (TP4) are the most common bacterial and archaeal surface structure (45). TP4 pili are both abundant and diverse on the surface of *Sulfolobus* and other Crenarchaeal cells (45). In its simplest form, the TP4 system consists of three proteins: an assembly ATPase of the VirB11 family, a major pilin protein, and a platform protein of the TadC family (199). These proteins are frequently organized into gene clusters or operons, together with minor pilin proteins (199). Additionally, major and minor pilin proteins contain N-terminal signal peptides that target these proteins for secretion (45, 199).

STIV-1 is closely related to our newly isolated STIV-NG05. Furthermore, the pilus structure recognized by STIV-1 resembles thread-like structures observed on the surface of its *S. acidocaldarius* cells by both size and degree of flexibility (197, 203). Threads are frequently observed in growing cultures of *S. acidocaldarius* (Young Lab unpublished, S. Albers and R. Whittaker personal communication). The genes encoding these threadlike structures have not been determined. We propose that the threadlike filaments recognized by STIV-1 are encoded in the novel pilin locus we have identified through screening of infection resistant mutants (Table 2.2, Figure 2.8). This region contains a number of genes predicted to have distant homology to major pilin proteins and/or to contain signal peptide sequences. No genes could be identified with homology to an assembly ATPase. These VirB11 type proteins are generally the highly conserved proteins in a TP4 system (199). It is possible however, that the suspected threadlike pili system in Region 1 utilizes an assembly ATPase from another pili operon.

Host Discrimination

The structural analysis presented here clearly implicates the prominent turrets of STIV-1, and the second and third domains of C381 in particular in host receptor binding, and host recognition. In this light, sequence divergence between the STIV-1 and STIV-2 clades of C381, and the lack of conservation in surface residues implicated in receptor recognition between these clades could indicate use of different type IV pili types, resulting in host discrimination. This is supported by examination of a micrograph in which STIV-2 co-purified with host filaments resembling aaP type pili (appearing thicker and more rigid than the pili recognized by STIV-1) (151). Pili and flagella are frequently targeted by bacteriophages to allow host discrimination (44, 204-207). Considering this, we suspect STIV-1 and STIV-2 target different surface structures. Still, given the close phylogenetic relationship between STIV-1 and STIV-NG05, especially in their C381 protein sequence (Supplemental Figure 3), we anticipate these viruses to target the same surface structure.

Genome Delivery

We currently lack any information on the mechanism employed by any archaeal virus to delivers it genome into the host cell. However, we do know that initial binding to cellular appendages such as type IV pili and flagella are an emerging theme for archaeal viruses. For example, the rudivirus SIRV2 binds to the tip of an unknown type IV pili-like structure (50, 208). At later time points SIRV2 is found in partial states of

disassembly localized on the cell surface. How it migrates from the tip of the pili to the cell surface is unknown.

We face a similar question for STIV-1. Were the unknown pili able to retract (209), this could certainly explain the presence of particle in Figure 2.4B. Archaea lack identified homologues to the disassembly ATPase (PilT) that drives TP4 pili retraction in bacteria (45). Still, a number of bacteria lacking PilT can retract their pili, although inefficiently, probably by dissociation of the unstable terminal pilin protein which is embeded in the cell membrane (210). Even if the pilus recognized by STIV cannot retract, alternative mechanisms exist to explain translocation of the virus. One possibility is a directional walk along the pili until it encounters the cell surface, i.e., the classic "nut on a bolt" hypothesis in which phage move along a pilus to the cell body, presumably directed in their motion by the 'threads', or twist and rise of the pilus (211, 212). However, in this work STIV was most frequently observed bound to multiple pili, as its icosahedral symmetry makes it a multivalent particle. In this context, the "nut on a blot" model seems less viable for this system. A third alternative is that the virus might bind become entangled in multiple pili, and like a burr in a dog's coat that stochastically work its way through the fur until it's entangled and trapped against the skin. In a similar manner, STIV might be drawn toward the cell surface through entanglement with multiple pili.

Clearly, additional data at both earlier and later time points will be needed to distinguish among these and other possibilities. Still given our experimental data the entanglement hypothesis appears most viable. In other viral systems, attachment to

multiple receptor sites is generally required to initiate conformational changes that prepare the viral capsid for entry (refer to "Attachment and entry of bacterial and eukaryotic viruses" thesis Introduction). Consistent with this in each of our tomograms we detect, what appear to be disassembled viral particles, occurring most frequently near the cell surface (Figure 2.5J). These disassembled particles are always heavily entangled in a multitude of pili. Therefore, we propose that multivalent binding of host pili serves both to draw the virus closer to the cell surface and to initiate destabilization of the capsid, priming it for cellular entry.

Isolated archaeal viruses frequently possess lipid envelopes, and of the 18 archaeal virus families, 11 are enveloped (65, 68, 113, 116, 123).In the case of STIV, it is an internal lipid envelope (Figure 2.1). This raises the question of how the genome transits both the host and the viral membranes. As a member of the PRD1 viral lineage, we consider the possibility that genome delivery might be similar to other members of this lineage utilizing an internal lipid membrane within an icosahedral shell. Remarkably, following initial receptor attachment in PRD1, the internal membrane "vesicle" in this phage is observed to develop a tube-like structure that extends through the unique vertex of PRD1 and into the host cell membrane, providing a conduit for transfer of the viral genome into the host cell (61, 213). For STIV-1, we can consider that following localization to the cell surface, secondary receptor binding sites on STIV-1 could be engaged (perhaps unmasked by conformational changes upon recognition of the type IV pilus), catalyzing a similar transformation for STIV-1 with subsequent genome insertion.

Alternatively, PM2, also a member of the PRD1 viral lineage housing an internal membrane is thought to bind to a cell surface receptor, upon which the protein capsid dissociates, revealing the viral envelope and promoting fusion with the host cell membrane (213, 214). This model requires disassembly of presumably meta-stable viral capsid. Interestingly, disassembled states of STIV-1 have been observed (144), and an entangled virus in the process of disassembly may be visible in Figure 2.4 (particle iii), although this might just as well represent a damaged particle present in the purified virus used for these experiments, or an unsuccessful dead end in the infection process due to entrapment in multiple pili. Regardless, from the structural perspective, the C381 subunit interfaces are surprisingly hydrophilic, and recombinantly expressed C381 is a well-behaved monomer prior to crystallization, as are the major capsid protein (B345) and the penton base (A223). From the thermodynamic view, this is consistent with the idea of a metastable icosahedral shell that might spontaneously dissociate upon interaction with cellular receptors. Indeed, in addition to serving as a maturation protein that prevents re-adsorption to an infected cell, the C557 petal protein may also serve to pre-organize C381 for assembly into the virion, and then dissociate once the virus emerges into the acidified hot spring.

However, if a disassembly mechanism is utilized, this raises additional questions. Does the membrane enclosed genome enter the cell by endocytosis, direct injection through the A55/A223 α-hemolysin-like structure, or by a fusion event such as that suggested for *Sulfolobus monocaudavirus* (51). Critically, with regard to endocytosis or membrane fusion, the hyperthermophilic membranes in *S. solfataricus* and STIV are

composed of cyclic tetraether lipids that span the entire membrane (15, 91, 111). These membranes lack the inner and outer leaflets found in mesophilic organisms. Thus, if membrane fusion is involved, we anticipate a fundamentally different process than current models for fusion of lipid bilayers (125). Regardless of the mechanism of genome delivery, we are greatly interested to study these later points in STIV attachment and entry. Such studies have clear potential to make fundamental contributions to our understanding of membrane dynamics, archaeal biology, and life on earth.

CHAPTER THREE

DISCOVERY AND CHARACTERIZATION OF TSPV1: A PILATED ARCHAEAL VIRUS

Introduction

High temperature acidic environments, such as those found in the hot springs of Yellowstone National Park, USA (YNP), provide a rich source of archaeal viruses (157). Archaeal viruses isolated from these environments show a range of virion morphologies from typical rod-shaped helical viruses and spherical icosahedral viruses to unusual spindle and bottle-shaped virions (25, 63-66). This virion diversity is also apparent at the genome level, where most genes lack detectable homology to known functions (27). Of the 18 recognized or proposed archaeal virus families, 11 were isolated from high temperature environments (28, 68). The factors driving archaeal virus diversity in acidic hot springs are unknown, but likely reflect adaptations to the physical and geochemical nature of these environments, as well as the cell biology and biochemistry of their archaeal hosts.

Of the 65 known thermophilic archaeal viruses, 47 infect hosts from the order Sulfolobales (25, 28, 29, 68, 115, 215, 216). The remaining 15 viruses have been isolated from members of the Thermoproteales, Desulfurococcales, Methanobacteriales, and Thermacoccales (63, 95, 217-220). This bias results from the ease with which many Sulfolobales hosts are isolated and cultured under predominantly aerobic, heterotrophic

conditions. It is therefore reasonable to suspect that a substantial level of archaeal virus diversity remains unexplored given the recent expansion of archaeal phyla (221, 222).

In order to expand the diversity of cultured archaeal viruses, we sought to isolate new viruses infecting the order Thermoproteales. A common genus of this order, *Thermoproteus*, is found in sulfur sediments of YNP hot springs (223, 224). These anaerobic sulfur reducing organisms have a global distribution in acidic and near-neutral thermal features (115, 216, 223, 225, 226). They are capable of growing chemolithotropically using H_2 and CO_2, or organoheterotropically using a wide variety of sugars, organic acids, and alcohols (227, 228). The cells exist in different morphologic states including rod -shaped, branched, and cocci (226, 227). Like other archaea, members of this genera possess a lipid monolayer, formed by ether linked isoprenoid chains (15). This membrane is then surrounded with a para-crystalline proteinaceous layer or S-layer (229).

To date, only 6 Thermoproteales viruses have been described from 3 archaea specific virus families (115, 216, 220, 230). TTV1, TTV2, and TTV3 are from the Lipothrixviridae, a family of flexible linear viruses, while the TTV4 part of the Rudiviridae, a related family of stiff rod like viruses (220). The remaining 2 viruses are Pyrobaculum spherical virus (PyrSV) and Thermoproteus tenax spherical virus 1 (TTSV1), the only members of the Globuloviridae archaeal virus family (115, 216). We describe here a seventh Thermoproteales virus with unusual hair-like fibers extending from its virion surface. We propose that these heavily glycosylated fibers facilitate viral attachment to host cells.

Materials and Methods

Host and Virus Source

Thermoproteus CP80 strain is an anaerobic, sulfur reducing organism originally isolated from the Cinder Pool hot spring (44° 43' 56.9", 110° 42' 32.4") in Yellowstone National Park (224). This thermal feature is a relatively stable, high temperature (80-89°C) and low pH (4-4.4) hot spring. *Thermoproteus CP80* was isolated from hot spring sediment, and metagenomic sequencing of the cellular community identified it as the most abundant species in the sediment of this spring (224). A pure culture of this strain was obtained by dilution to extinction and confirmed by 16S rRNA sequencing (224).

Pure cultures of *Thermoproteus CP80* were screened for the presence of virus-like particles (VLPS) by negative stain electron microscopy. Samples were stained with 2% uranyl acetate on carbon coated grids prior to transmission electron microscopy (TEM) using a LEO 912AB with 2Kx2K CCD camera (Zeiss, Germany).

Culturing Conditions

Media was prepared as in Boyd et al. 2007 with the following modifications (231). In place of peptone, 3 g/L tryptone and 0.2 g/L yeast extract was used for a carbon source. Anaerobic bottles (0.5 L) were filled with 150 ml of media, heated to 80˚C followed by purging for 30 minutes with nitrogen gas and then inoculation with 5 ml of cell culture. Inoculated cultures were incubated for 10 days at 80°C or until peak viral production was observed. Viral production was by monitored by qPCR amplification of a genome segment using qPCR primers orf14_F (TGGCCGGCCGCTTCCAGATA) and

orf14_R (GCGGTCGATAGGTTAGCGGGC). Standards for qPCR consisted of a 255 bp fragment from the predicted major capsid protein GP11 cloned into a TOPO TA pCR2.1 vector (ThermoFisher, Waltham, MA). SSoAdvanced Syper Green Supermix was used for qPCR reactions according to manufacturer directions (Biorad, Hercules, CA). A Rotor-Gene series Q (Qiagen, Hilden, Germany) was used for the qPCR machine with an initial denaturation of 98°C for 30 sec followed by 22 cycles of 98°C for 5 sec, 65°C for 30 sec.

Virus Isolation and Purification of Virion Fibers

Tangential flow filtration (TFF) was used to concentrate 12 L of cell culture to 100 mL using a 10K MWCO PES column (Amersham, Westborough, MA), followed by centrifugation for 15 minutes at 1200×g in a swinging bucket rotor to pellet cells and media debris. Virus present in the supernatant was further concentrated to 50 mL by TFF, followed by in-line filtration through 0.8 µm Acrodisc syringe filter (PALL, New York, NY) to remove any remaining cells. The final step in virus purification was accomplished by banding virus on buoyant density gradients created by addition of CsCl to a final density of 1.286 g/ml, followed by ultracentrifugation for 24 hours at 37,000 rpm (175,000 x g) in an SW41 rotor. Gradients were hand fractionated, dialyzed against 5 mM citrate acid pH 4, and analyzed by both qPCR and by negative stain electron microscopy. The protein concentration in fractions were measured using Cubit fluorometric protein quantification (Invitrogen, Eugene, Oregon). Fractions containing separated fibers were further purified by 10 cycles of diafiltration with a 0.5 ml Amicon Ultra spin column (Millipore, Burlington, MA).

Cryo-Electron Microscopy

For cryo-electron microscopy, purified virus was frozen on Quantifoil R2/1 holey carbon grids (200 mesh copper) in liquid propane/ethane using a Vitrobot Mark III (FEI, Hillsboro, OR). Tilt series were taken using a Titan Krios (FEI, Hillsboro, OR) operated at 300 kV and equipped with a Volta phase plate (163, 164), a Quantum post-column energy filter (Gatan, Pleasanton, CA), and a Summit K2 camera (Gatan, Pleasanton, CA) operated in electron counting mode at … name the facility). Tilt-series images were collected using SerialEM (165). Individual frames of images acquired with the K2 camera were aligned in Digital Micrograph (Gatan, Pleasanton, CA). Acquisition parameters were: magnification 35,700 X; tilt range +/-60°; tilt increment 2°; total dose ~60 $e^-/Å^2$; pixel size 0.14 nm/px; defocus with phase plate -0.25 μm. The Volta phase plate was operated as previously described (166, 167).

Tomograms were reconstructed using IMOD 4.7 (168). Contrast transfer function (CTF) corrections were not performed. Gold nanoparticles were used as fiducial markers for alignment of tilt-series projection images. Local alignment solutions were used for rotation, magnification, and tilt angle, while a global solution was employed for distortion. Gold particles were erased from the final aligned stack. Images were binned 2x, radially filtered with a cutoff of 0.4 and falloff of 0.05 and run through 10 SIRT-like filter iterations.

Genome Sequencing and Assembly

Viral nucleic acid was extracted from CsCl purified virus using a PureLink viral RNA/DNA extraction kit (Thermo Fisher, Waltham, MA); and sequenced at the University of Illinois Sequencing Center (Urbana-Champaign, IL) using Illumina HiSeq technology with 2x250bp paired-end reads (Illumina, San Diego, CA). Genome assembly was performed using the MIRA assembly program (version 4.0.4). Initial assembly required subsampling 30,000 reads and assembling these in order to avoid assembly failure from high sequence coverage. Read recruitment to verify the inverted terminal repeats (ITR) was performed using Geneious assembler (Biomatters, Newark, NJ). For this process, the ITR region was trimmed from each end of the genome and then paired reads recruited to the trimmed sequence. Genome alignments between TSPV1, TTSV1, and PyrSV were conducted using MAUVE v. 20159226 (232).

Gene Prediction and Annotation

Glimmer and Prodigal were used to predict viral genes (233, 234). Predicted gene products were compared to the NCBI nr database using BLAST (159). In addition, proteins from both PYRSV and TTSV were queried against the TSPV1 genome to identify homologs using BLAST (159). Protein alignments were generated using MUSCLE (160). Predicted proteins were further analyzed using the Phyre2 structure prediction server and HHpred to detect distant homologues (180, 235). PSIPRED, FlaFind, and SignalP 5.0 were used to identify potential signal peptides in the predicted proteins of TSPV1 (181, 236, 237). PSIPRED was also used for protein secondary structure prediction (237). PROSPER was used to predict protein cleavage sites (238).

N-terminal Sequencing and LC-MS Sequencing

N-terminal Edman sequencing was performed using purified fibers (5 µg) that were applied to a Prosorb cartridge followed by 8 sequencing cycles conducted using a Shimadzo PPSQ-53A sequencer at the Iowa State University Sequencing Center.

LC-MS analysis was performed on in solution digests of both purified viral capsids and purified fibers as follows. Samples were dialyzed into 12.5mM NH_4HCO_3 pH 8 followed by addition mass-spectrometry grade urea to a final concentration of 8M. Following this, samples were reduced in 5 mM DTT, alkylated with 14 mM Iodoacetic acid (IAA), followed by digested for 4 hrs (37° C) at a 1:20 protease/protein ratio using Trypsin/Lys-C mix (Promega, Madison, WI) with ProteaseMax (Promega, Madison, WI) addition to 0.03%. Samples were diluted to a urea concentration of 1M followed by overnight digestion (37° C).

For in-gel digestion, 9 µg of purified capsids or 7 µg of purified fibers were dialyzed into 12.5mM NH_4HCO_3 pH 8 then mass-spectrometry grade urea was added directly to a final concentration of 8M together with SDS-PAGE loading buffer (50 mM Tric-Cl, 2% w/v SDS, 0.1% w/v bromophenol blue, 10% glycerol, 100 mM β-mercaptaethanol, pH 6.8). Samples were denatured at 98°C for 40 min or at RT for 2 hours. Electrophoresis was performed using a 4-12% Bis-Tris NuPage gel (Thermo Fisher, Waltham, MA) for separation. Gels were stained with Colloidal Coomassie and imaged with a Typhoon Trio laser scanner. Gel slices were excised and proteolyzed as follows. Gel slices were destained in 50 mM ammonium bicarbonate until clear, then reduced with 100 mM DTT followed by alkylation with 55 mM IAA. Digestion was

performed overnight in 25 mM NH_4HCO_3 using 0.3 µg mass spectrometry grade trypsin (Promega). Samples were subject to reverse phase chromatography using a Dionex nano-UHPLC, then analyzed with a Bruker MaXis Impact mass spectrometer (Billerica, MA).

Data analysis was performed using MetaMorpheus v. 0.0.300 with precursor and product tolerance of 35 ppm, 3 max missed cleavages, and PTMs of: common fixed, common variable, glycosylation, common biological, less common, common artifact, metal, trypsin digested, UniProt, and Unimod modifications. Searches were performed against a database of viral proteins, cellular proteins, and/or common contaminants. Alternatively, searches were performed with SearchGui v. 3.3.13 with precursor and product tolerance of 35 ppm, 3 max missed cleavages, oxidation of M and carbamilation of K as variable modifications using X! tandem, MS-GF+, and Tide search engines.

Capsid Protein Glycosylation Analysis

The NetNGlyc server was used to predicted N-linked glycosylation sites (239). For glycan staining, electrophoresis was performed with 9 µg of purified capsids or 7 µg of purified fibers using a 4-12% Bis-Tris NuPage SDS-PAGE gel (ThermoFisher, Waltham, MA) for separation with 4 µl CandyCane ladder (Invitrogen, Eugene, OR) and 4 µl of Precision Plus Protein Dual Xtra ladder (Biorad, Hercules, CA) as standards. The gel was stained with the Pro-Q Emerald 488 Glycoprotein Gel and Blot Stain Kit (Invitrogen, Eugene, OR) following manufacturer recommendations and visualized using an AlphaImager 2200 imager (ProteinSimple, San Jose, California).

Diethyl Ether Sensitivity Assay

To test for the presence of a viral envelope the method developed by Andrewes was employed, briefly diethyl ether was added to viral particles in 5 mM citric acid buffer pH 4 (240). The final concentration of diethyl ether was 20% v/v; samples were incubated at room temperature for 24 hrs and then analyzed by negative stain TEM.

Results

Actively growing *Thermoproteus CP80* cultures were found to be chronically infected with a 75-85 nm spherical virus (Fig 1A&B.) . There was no evidence for host cell lysis, and cultures remained viable after multiple rounds of passaging over a two-year time period, with no significant change in either virus production or host cell viability. This supports the conclusion that the virus has established a chronic mode of infection, that and was likely present during the original *Thermoproteus CP80* isolation from the YNP Cinder Pool hot springs.

Morphology of TSPV1

Cesium chloride density gradient centrifugation was utilized to purify viral capsids, which appeared by TEM visualization to be intact and purified at a density of 1.29 g/ml, a density typical for enveloped virions (Figure 3.1C). Using both negative stain and frozen hydrated samples, electron micrographs and tomograms were acquired of the virus. These images reveal a spherical capsid morphology with an average diameter of 83 nm (Figure 3.1C,D,E and Supplemental Video 2). Often observed were the

presence of 10-20 nm protrusions from the virion surface (green arrows Figure 3.1C). A lipid envelope also appears to be present in the electron micrographs (blue arrow Figure 3.1E). The most striking, and morphologically unique feature of the virion is numerous 3 nm diameter fibers, of various lengths, extending from the capsid surface (black arrows Figure 3.1C,D,E). Each virion contains an average of 7 fibers with the typical range varying between 0 and 20 fibers per virion. Some of these fibers extend up to 500 nm from the capsid surface and are highly flexible. Figure 3.2 provides a histogram of fiber lengths, that shows a relatively even distribution up to 200 nm; it is important to note that the lengths are biased against long fibers, since these frequently extend beyond the dimensions of a given micrograph. No unique terminal structures appear present at the ends of fibers; which suggests that the entire fiber, rather than just the ends, serves a biologic purpose for the virus.

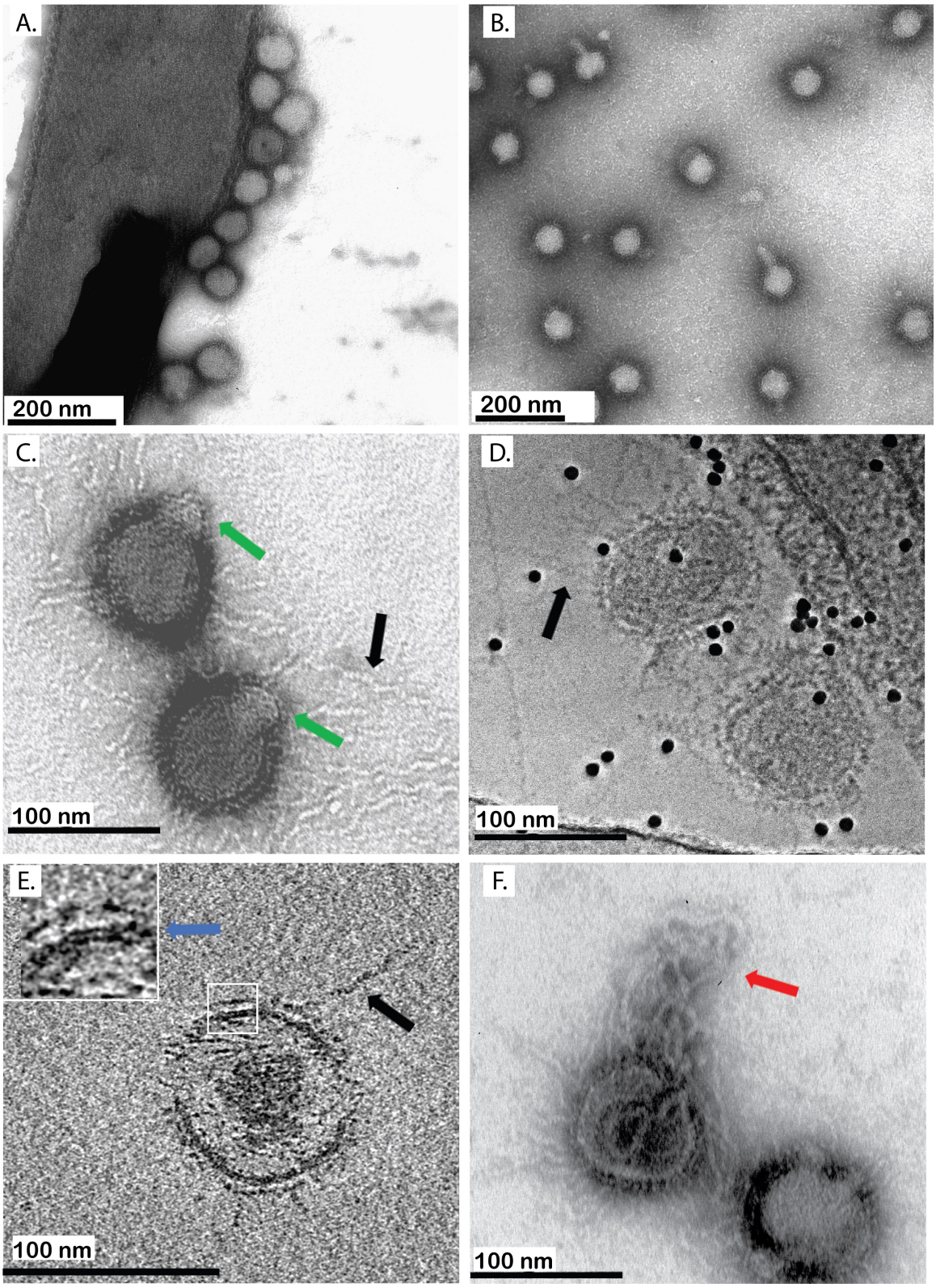

A.
200 nm
B.
200 nm
C.
100 nm
D.
100 nm
E.
100 nm
F.
100 nm

Figure 3.1. Morphology of TSPV1. Black arrows indicate viral fibers, green arrows—capsid protrusions, blue arrow—the viral envelope, and red arrow—nucleoprotein. A) VLPs in association with a *Thermoproteus CP80* host cell. B) Concentrated VLPs prior to density gradient purification. C) Negative stain TEM image of CsCl purified virions. D) Cryo-EM micrographs TSPV1 virions against a host cell. Note; black dots are nano-gold used during tomography. E) Tomographic slice of TSPV1 (2.5 nm thick). F) Negative stain TEM image of ruptured TSPV1 virion showing exposed nucleoprotein.

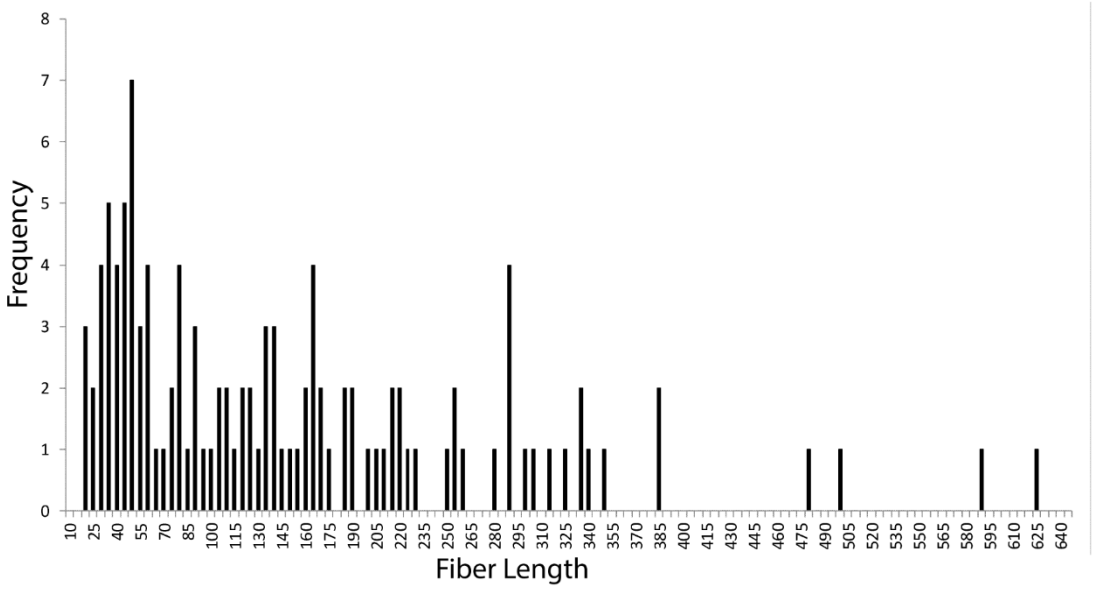


Figure 3.2. Histogram of TSPV1 fiber lengths. Conntinuous variability in fiber length is observed from 20 nm up to approximately 200 nm with some fibers exceeding 400 nm in length.

The morphology of TSPV resembles Pyrobaculum Spherical Virus (PyrSV) and Thermoproteus tenax spherical virus 1 (TTSV1) both members of the Globuloviridae (115, 216). This includes the presence of protrusions from the capsid surface, the diameter of the virion, and the appearance of a lipid bilayer in micrographs. However, the fibers present in TSPV were not observed for PyrSV or TTSV1. Subjecting TSPV1 to increased pH ruptured the particles, releasing a nucleoprotein fiber that matches the nucleoprotein observed in ruptured PyrSV particles (Figure 3.1F) (115).

TSPV1 likely contains an external lipid envelope, like the other Globuloviridae members (115). Enveloped viruses are sensitive to diethyl ether treatment, which solubilizes their lipid membrane (240). Addition of 20% diethyl ether to a sample of TSPV1 resulted in disruption of approximately 70% of the virions, with most of the remaining virus displaying obvious damage to their capsids, by TEM analysis. In addition to this, TSPV1 has a buoyant density of 1.29g/cm3 in CsCl density gradients, which is typical for spherical enveloped viruses. Finally, electron micrographs of TSPV1 reveal a characteristic lipid envelope surrounding the virus (Figure 3.1E (Insert)).

TSPV1 Genome Assembly and Annotation

Assembly of the TSPV1 genome yielded a 18,655 bp linear ds DNA genome with 1200x average coverage (Figure 3.3A). The TSPV1 genome was found to have inverted terminal repeats (ITRs) of 102 bp in length at each end. PyrSV also has ITRs of similar length (120 bp). However, no sequence homology could be detected between the repeats of each virus. At the nucleotide level, TSPV1 is distinct from the other two viruses of the Globuloviridae, since only small segments of the genome were able to be aligned using MAUVE whole genome alignment (Figure 3.3B).

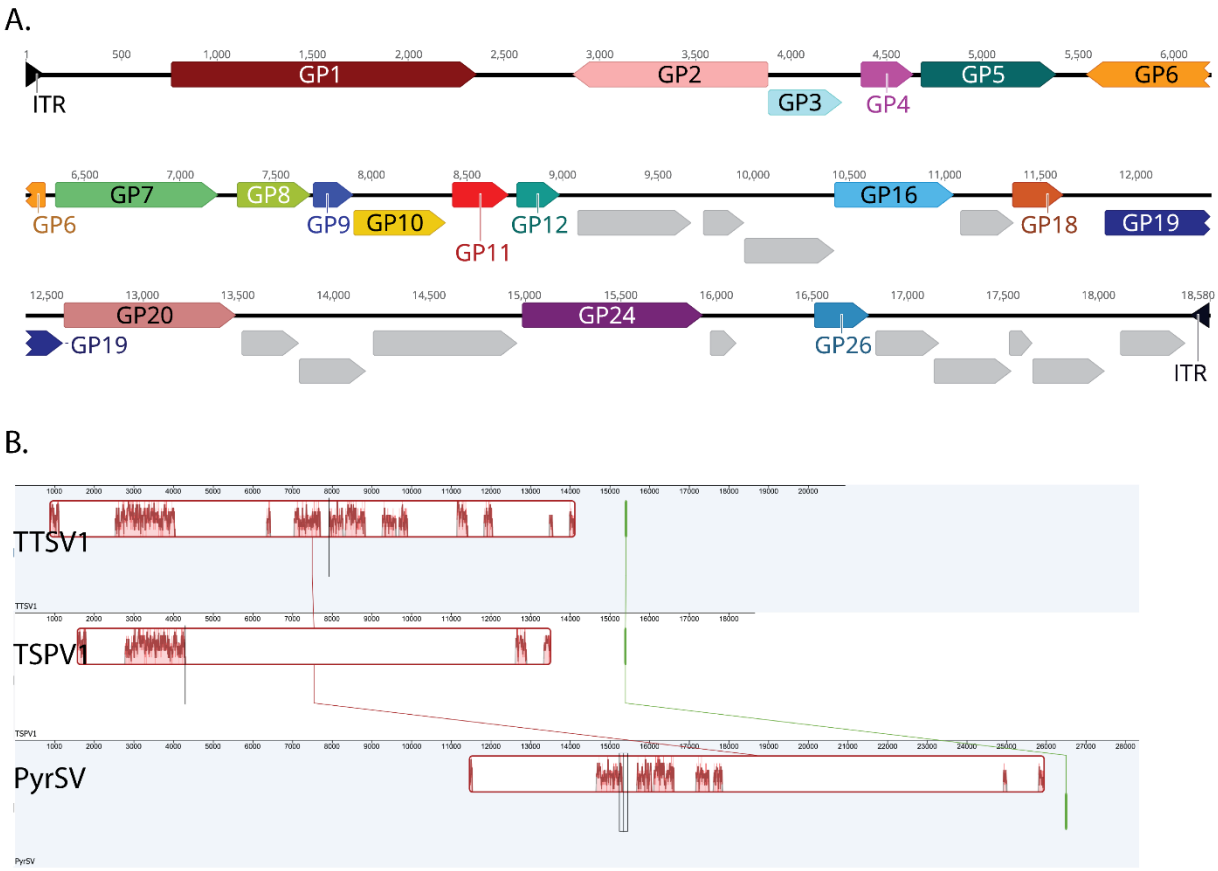


Figure 3.3. Map of the TSPV1 genome. A) The genome of TSPV 1 is 18,655 bp in length and contains 31 predicted gene products (GPs). Like PyrSV, this linear genome has inverted terminal repeats of approximately 100 bp in length. Colored genes are conserved across the Globuloviridae. B) MAUVE alignment of the 3 Globuloviridae genomes shows only limited regions of sequence similarity at the nucleotide level, however TSPV 1 appears more closely related to TTSV1 than to PyrSV.

The TSPV1 genome codes for 31 gene products (GPs) with an 85% coding density. Most of these GPs occur on a single strand of the genome, an emerging theme for the Globuloviridae, with only GP2 and GP6 found on the complimentary strand. TSPV1 has the smallest genome of the 3 viruses, encoding only 31 proteins, with the majority, 60%, being conserved across the Globuloviridae (Figure 3.4). One block of genes shows a high degree of synteny in all three viruses; given that this block contains the MCP the other proteins in this gene block could play a role in virion assembly (Figure

3.4 Boxed Region). Based on both gene content and morphology, this virus likely represents a new member of the Globuloviridae family of archaeal viruses.

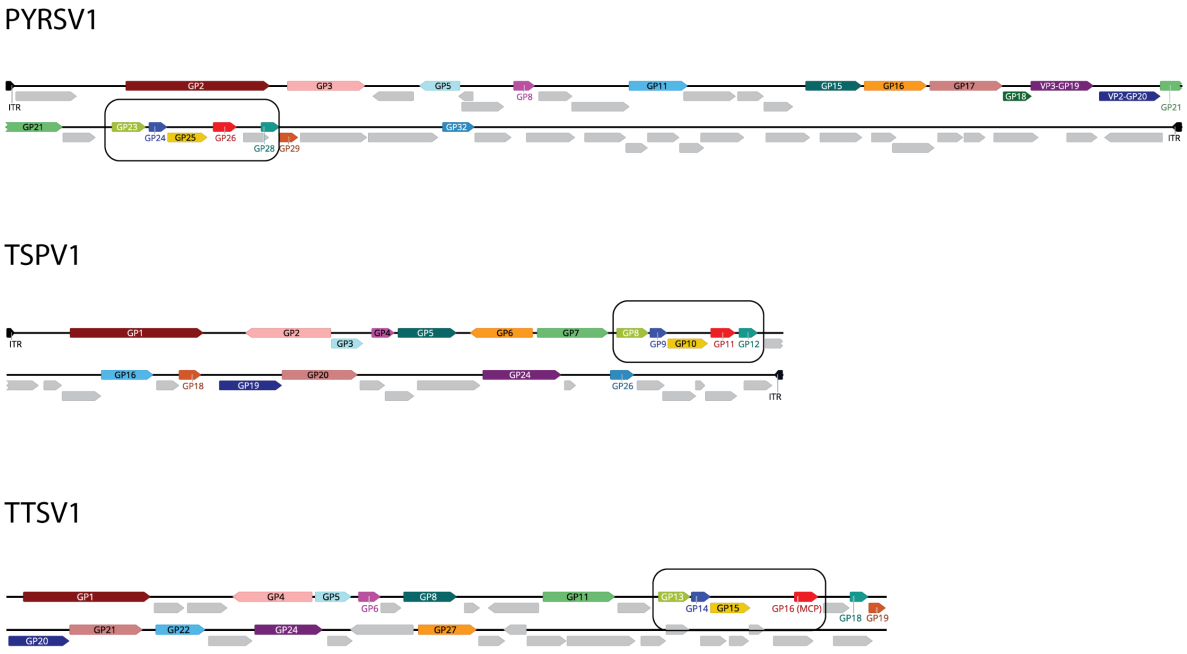


Figure 3.4. Location and identity of Globuloviridae genes, whose presence is conserved in all three viruses. Genome maps for the three Globuloviridae are shown with conserved genes colored the same in each map. The TSPV1 18.6 kb genome is the smallest of the three. Furthermore, 60% (18/31 ORFs) of the TSPV 1 genes are shared with the other two Globuloviridae. A block of genes (boxed regions) is highly conserved among the 3 viruses in both gene content and senteny. The predicted major capsid protein (GP11) resides in this gene block.

Not surprisingly, the highest similarity of predicated TPV proteins was to proteins found in the other two Globuloviridae. Of the 31 predicted TPV genes, 15 have BLAST scores < 1×10^-5 and these are exclusively to proteins from TTSV1 and PyrSV. Based on gene length, order, secondary structure profiles, and localized segments of homology we were also able to link GP3, GP8, and GP24 to proteins from PyrSV and

TTSV1. In total then, 18 proteins were identified whose presence is conserved in all 3 viruses.

In an effort to annotate the proteins of TSPV1, we employed a combination of bioinformatic tools including BLAST, HHpred, and Phyre2 yet were unable to identify structural or functional homologs for most of the TSPV1 proteins. Assignment of structure and/or function was achieved for only 9 proteins (Table 3.1). Some proteins from PyrSV have had their structure previously determined by X-ray crystallography, and two of these proteins have homologs in TSPV1 (241). PyrSV GP11 (PDB: 2X3M) is homologous to TSPV1 GP16, and PyrSV GP32 (PDB: 2X5C) is homologous to TSPV1 GP26. Despite having solved crystal structures, functions for these two proteins have yet to be determined.

Table 3.1 Homologues of TSPV1 Proteins and their Characteristics

TSPV Gene	TTSV1 homologue	PyrSV homologue	Weight (kDa)	Glycan Sites Predicted	Predicted Signal Peptide	Annotations
GP1	GP1	GP2	59.2	1 N-Linked		AAA+ Helicase
GP2	GP4	GP3	33.4			Na+/Ca2+ exchanger
GP3	GP5	GP5	16.0	1 N-Linked		Thioredoxin fold
GP4	GP6	GP8	10.2			
GP5	GP8	GP15	26.2			
GP6	GP27	GP16	26.5		Yes (FlaFind)	Minor virion protein (TSPV1)
GP7	GP11	GP21	31.4			
GP8	GP13	GP23	13.3	1 N-Linked	Yes (FlaFind & SignalP)	
GP9	GP14	GP24	8.0	4 N-Linked		
GP10	GP15	GP25	18.3		Yes (FlaFind)	
GP11	GP16	GP26	11.1			Major Capsid Protein
GP12	GP18	GP28	8.3			
GP13			21.0	7 N-Linked	Yes (SignalP)	
GP14			7.4			
GP15			18.0			
GP16	GP22	GP11	24.0			PyrSV PDB ID 2X3M
GP17			11.0	1 N-Linked		
GP18	GP19	GP29	10.0			
GP19	GP20	vp2	26.0			Minor virion protein (PyrSV)
GP20	GP21	GP17	31.7			
GP21			11.3			
GP22			13.4	1 N-Linked		
GP23	GP27		27.0	3 N-Linked		
GP24	GP24	VP3	33.5	6 N-Linked	Yes (SignalP)	Fiber protein (TSPV1)
GP25			4.7			
GP26		GP32	10.4			PyrSV PDB ID 2X5C
GP27			12.7			
GP28			15.6			
GP29						
GP30			14.3			
GP31		GP18	12.5			

HHpred analysis of GP1 revealed it to contain an AAA+ ATPase domain from residue 146 to residue 374. The best match being to the helicase of a transposable element in *Staphylococcus. aureus* (E-value 3.6e-6) and to the E1 helicase of Papillomavirus (E-value 5.4e-6) both of which form hexameric assemblies, to facilitate

DNA unwinding (242, 243). Viruses frequently encode helicase enzymes to facilitate replication of their genome (244).

Both Phyre2 and HHpred revealed GP2 to likely be (79% of residues model that 93% confidence for Phyre2) a sodium calcium exchanger protein (NCX). Eukaryotic viruses frequently perturb calcium homeostasis during infection, although the precise reasons for modulating cellular calcium vary greatly (245). To our knowledge, TSPV1 and the other Globuloviridae would be the first viruses known to encode their own NCX proteins.

The function of GP3 was also predicted, using both Phyre2 and HHpred. With this protein, 81% of residues were modeled, at 96% conference, against Thioredoxin 2 from *Pseudomonas aeruginosa* (PDB 2LRC). Thioredoxin fold enzymes are involved in multiple cellular processes including: disulfide bond formation and reduction, protecting cells from oxidative damage, and maintaining the cytoplasm in a reduced state (246, 247). The enzymatic activity and function of the thioredoxin class of proteins is modulated primarily by the XX dipeptide, sandwiched between two enzymatic cysteine residues (CXXC), and by a single residue occurring just before a strictly conserved cis-proline (XP) (247). Together, these residues set the redox potential for a given thioredoxin fold protein, and so dictate its specific function (247). Both GP3 and Thioredoxin 2 from P. aeruginosa have the same residues present at these sites (CPAC for CXXC & TP for the XP site). Unfortunately, the activity Thioredoxin 2 has not been determined, preventing us from using this homologue to know the function of GP3.

Due to the enveloped and non-lytic nature of the Globuloviridae, we suspect that TSPV1 buds from the host cell surface. Given this we predict that envelope proteins are secreted into the host membrane prior to budding through either the Sec or Tat translocase pathway, both of which are present in archaea (248). Consistent with this, 5 proteins in the TSPV genome are predicted to contain N-terminal secretion signal peptides (Table 3.1). Using the SignalP prediction software, GP8, GP13, and GP24 returned prediction values of 0.84, 0.87, and 0.88 respectively to Sec/SPI type signal peptides. In addition, GP6, GP8, and GP10 were predicted to contain Sec/SPIII signal peptides using FlaFind. As a comparison, in the *Thermoproteus CP80* host, out of the 1600 predicted proteins, only 8 were found to contain signal peptides by FlaFind, and only 42 by SignalP.

Identification of Major Coat Protein

SDS-PAGE of the purified viral capsids revealed an intense band migrating with an 10 kDa estimated mass, a less intense band at with an estimated mass of 15 kDa, and a protein smear stretching from top of the gel to 200 kDa (Figure 3.5A). In- gel tryptic digestion of the 15kDa band produced LC-MS spectra matching to four peptides from GP6 (Figure 3.5B&D). The predicted molecular weight of GP6 is 26 kDa which suggests that this protein identified in our SDS-PAGE gel represents a post-translational cleaved product. In support of this, the four identified peptides map to the N-terminus with a predicted molecular weight of 15.6 kDa (Figure 3.5D). The location of protein cleavage, and lack of Serine protease sites in the region, suggest that this protein, although containing a Type III secretion signal, is not processed by a Type III signal peptidase at

this site. Rather we suspect a metalloprotease since the region contains several predicted metalloprotease cleavage sites. Still, without determining the precise site of truncation this is speculative at best.

In-gel tryptic digestion identified GP11 as the protein product for the 10 kDa band (Figure 3.5C). Four peptide sequences belonging to GP11 were detected, resulting in 71% coverage of this protein (Figure 3.5E). Based on band intensity, we will refer to GP11 as the major capsid protein (MCP). This agrees with the previous results that found a homologous protein from TTSV1 (GP16) as the MCP (216). GP11 is predicted by PSIPRED to contain 4 α-helices. Four-helix bundle proteins are frequently used as major capsid proteins (27).

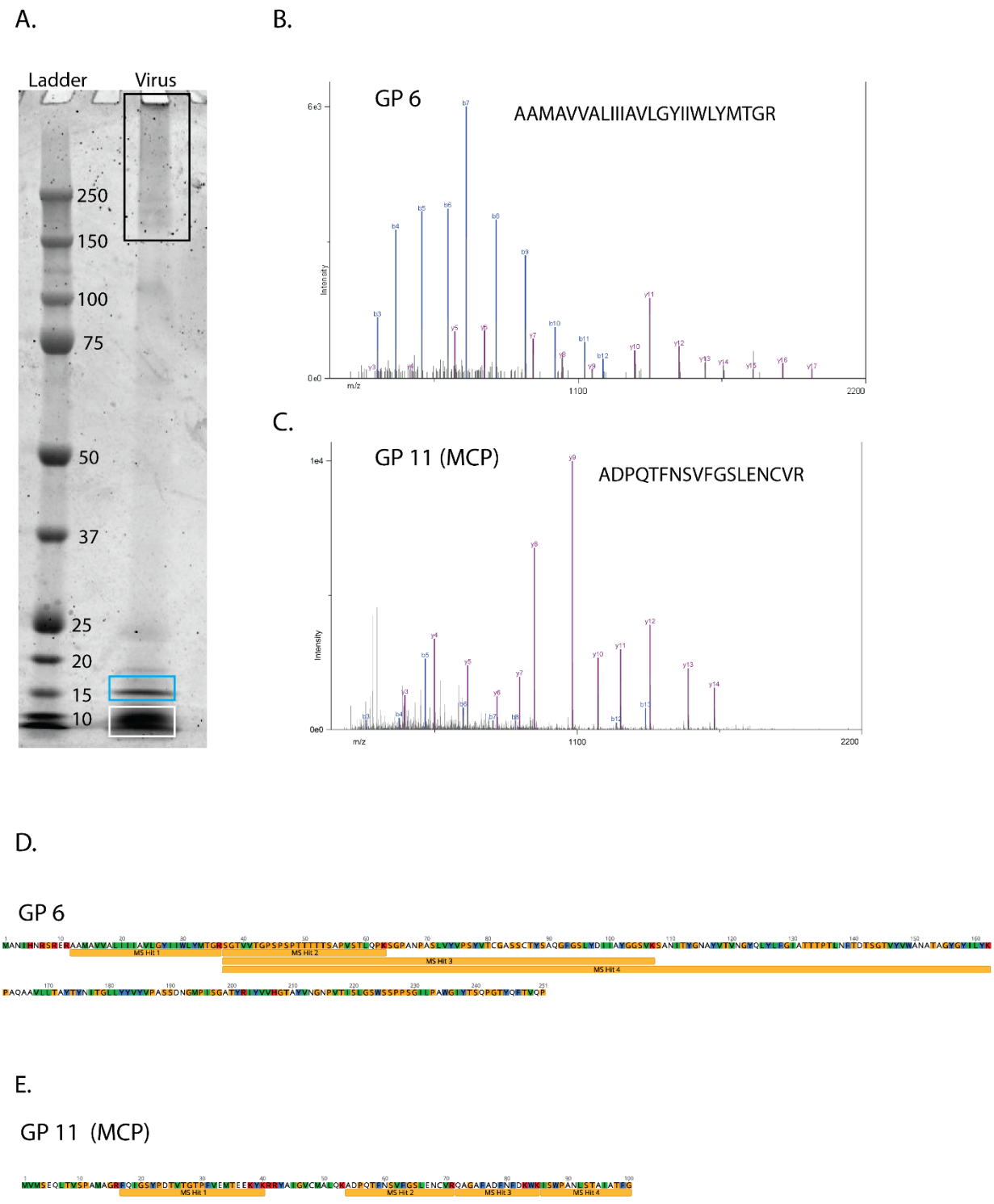


Figure 3.5. Identification of GP11 and GP6 as major and minor capsid proteins. A) An SDS-PAGE gel of purified TSPV1 virions, showing protein bands at 10 kDa (white box) and 15 kDa (blue box). In addition, a smear of protein stretches from 200 kDa to the top

of the gel (black box). Boxed regions were excised and proteolyzed prior to LC MS analysis. B) The 15 kDa band produced spectra by LC-MS matching 4 peptides from GP6. Shown is a representative spectrum for one of peptide matches with the identified sequence shown above. C) The 10 kDa band produced spectra by LC-MS matching 4 peptides from GP11. GP 11 is homologous to the predicted major capsid protein for TTSV1. D&E) MS Hits (peptides) identified for GP6 and GP11.

Purification of Viral Fibers

In the same cesium chloride density gradients utilized to purify viral capsids, a second protein band at 1.35 g/cm3 density was observed. Inspection of this band by TEM revealed abundant fibers and some virus particles (Figure 3.6A). Treatment of this material with 50 mM NH_4HCO_3 4M urea pH 8 followed by concentration using a 100 K MWCO spin column, was effective for removal of virus particles, and other cell debris: leaving exclusively viral fibers when observed by EM (Figure 3.6B&C). SDS-PAGE analysis of these purified fibers resulted in a series of protein bands ranging in estimated masses from 35 kDa to >250 kDa (Figure 3.6D). The prominent MCP bands at 10 kDa for GP11 and at 15 kDa for GP6 were absent from the sample. The series of protein bands observed in both intact capsids and purified fiber samples indicates that the fibers are composed of multimers formed from protein(s) with a monomeric mass of ~35 kDa. GP2 and GP24 have predicted molecular weights closest to this value (Table 3.1). Since GP2 appears to be an NCX type transport protein this leaves only GP24 as a likely fiber constituent.

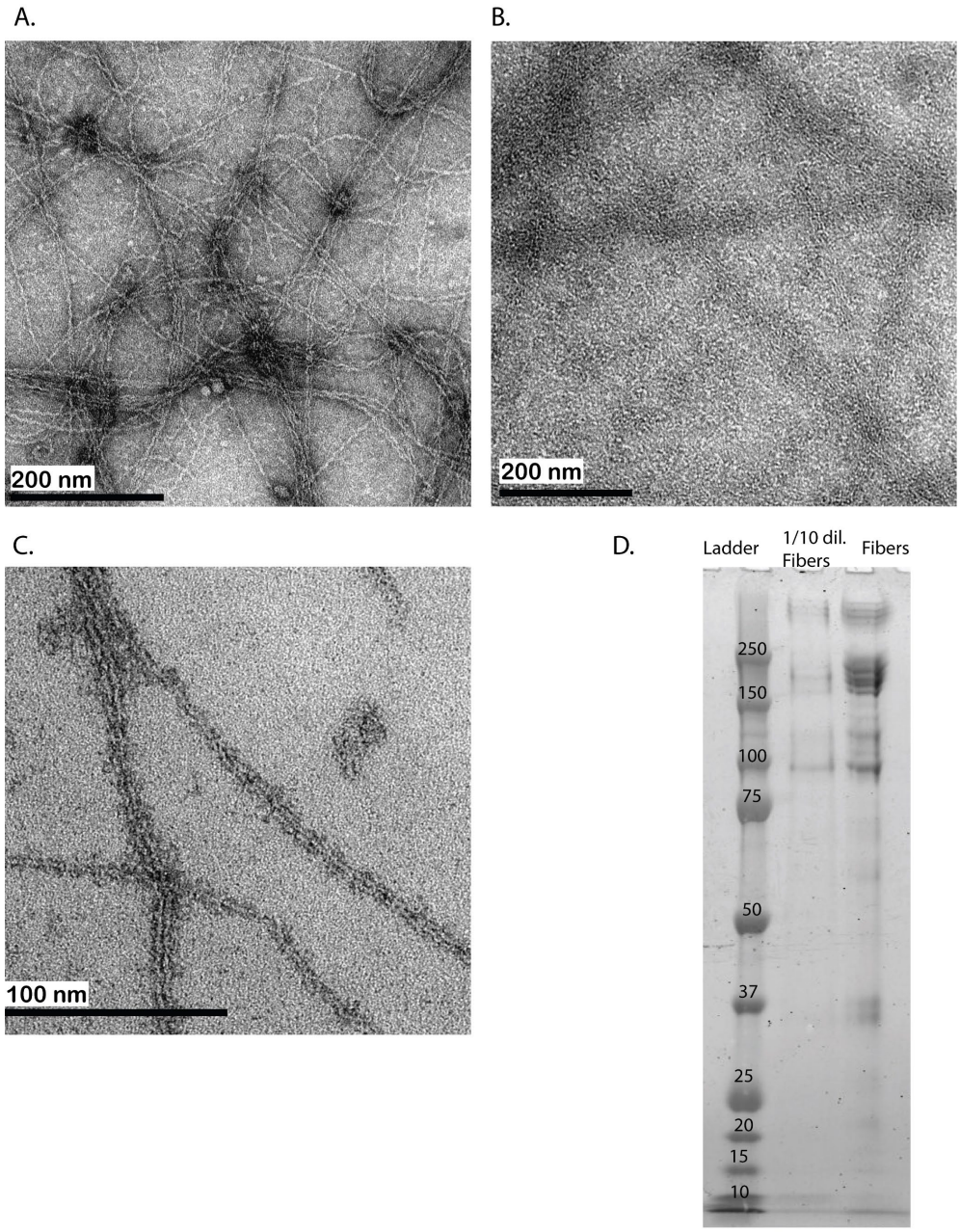


Figure 3.6. Purification of TSPV1 fibers. A) A CsCl band corresponding to the density of pure protein contained an abundance of viral fibers along with some virus particles and cell debris. B&C) Further purification was obtained by diafiltration of the sample at elevated pH in 4M urea, conditions which dissociate virus particles, cell debris, and vesicles, yet have no effect on pili structure. D) SDS-PAGE gel of purified fibers lacks the protein bands at 10 and 15 kDa corresponding to GP 11 and GP 6. A series of high molecular weight bands stretching from ~35 kDa to > 250 kDa are present and presumed to be multimers of a ~35 kDa fiber protein. Note: protein denaturation for electrophoresis

of capsid and fiber proteins was accomplished here by incubation in SDS buffer with 8M urea at 98° C for 40 min.

Characterization of the Viral Fibers

N-terminal sequencing of purified fibers identified one peptide sequence; AVFLVAVAIYITYT. This sequence was present in the N-terminal sequence of TSPV1 GP24. A nearly identical N-terminal sequence is present in the homologous PYRSV protein GP16 (Figure 3.7D). The sequenced N-terminus is recessed 13 amino acids from the initiator methionine of PYRSV and 11 amino acids for TSPV1. Although the TSPV1 and PyrSV proteins have low overall homology (16% identity), the sequence immediately preceding the identified N-terminal match is well conserved in all 3 viruses. This suggests that it may serve as a protein cleavage signal, cellular export signal, or other related function. Supporting this, SignalP predicts an N-terminal signal peptide for this protein targeting, it to the Sec/SPI export pathway (Table 3.1). Frequently signal peptides are cleaved from the protein after export to the membrane, which would provide an explanation for the missing 11 amino acids in TSPV1 and 13 amino acids in PyrSV (236, 248). Consistent with this, PROSPER predicts multiple Serine protease sites at amino acids 8,9,and 10 immediately preceding the N-terminal sequence obtained.

To achieve further validation that TSPV1 GP24 corresponds to the viral fibers, a variety proteases and digestion conditions were explored prior to LC MS analysis. Treatment with protease K, pepsin, and thermolysin produced no visible change in fiber structure by EM, or detectable peptide signatures by LC-MS. Similarly, pH conditions ranging from 1 to 10 had no observable effect on the fiber structure. However, exposure

to either 6 M GuHCL or 8 M Urea, combined with incubation at 98°C for 40 min proved to completely eliminate pili structures from the urea treated sample, and substantially reduced the number of pili structures in the GuHCL treated samples, as observed by TEM visualization. Use of saturated urea and extended thermal denaturation were therefore incorporated into both in-solution and in-gel digestion of the purified viral fibers. LC-MS based sequencing of these digests yielded matches to peptides from only GP24 (Figure 3.7A,B,&C). Identified peptides were VVDGKSLTTVSQLGSR, SLTTVSQLGS, and SLTTVSQLGSRLATIAVNNTAYIPWIIVSESK; giving a total protein coverage of 19% for GP24. This low coverage level is not unexpected since GP24 contains only 9 tryptic sites for the entire 33 kDa protein. Furthermore, urea in combination with high temperature is known to carbamylate Lys and Arg residues, blocking the few tryptic sites present. This is especially evident with the frequent observation of a VVDGKSLTTVSQLGSR peptide containing a missed cleavage site at the internal Lys residue, and a mass shift corresponding to a carbamyl group on this residue. The three identified tryptic peptides, combined with the N-terminal sequencing data, and the apparent monomeric mass of near 35 kDa all support GP24 as the fiber protein.

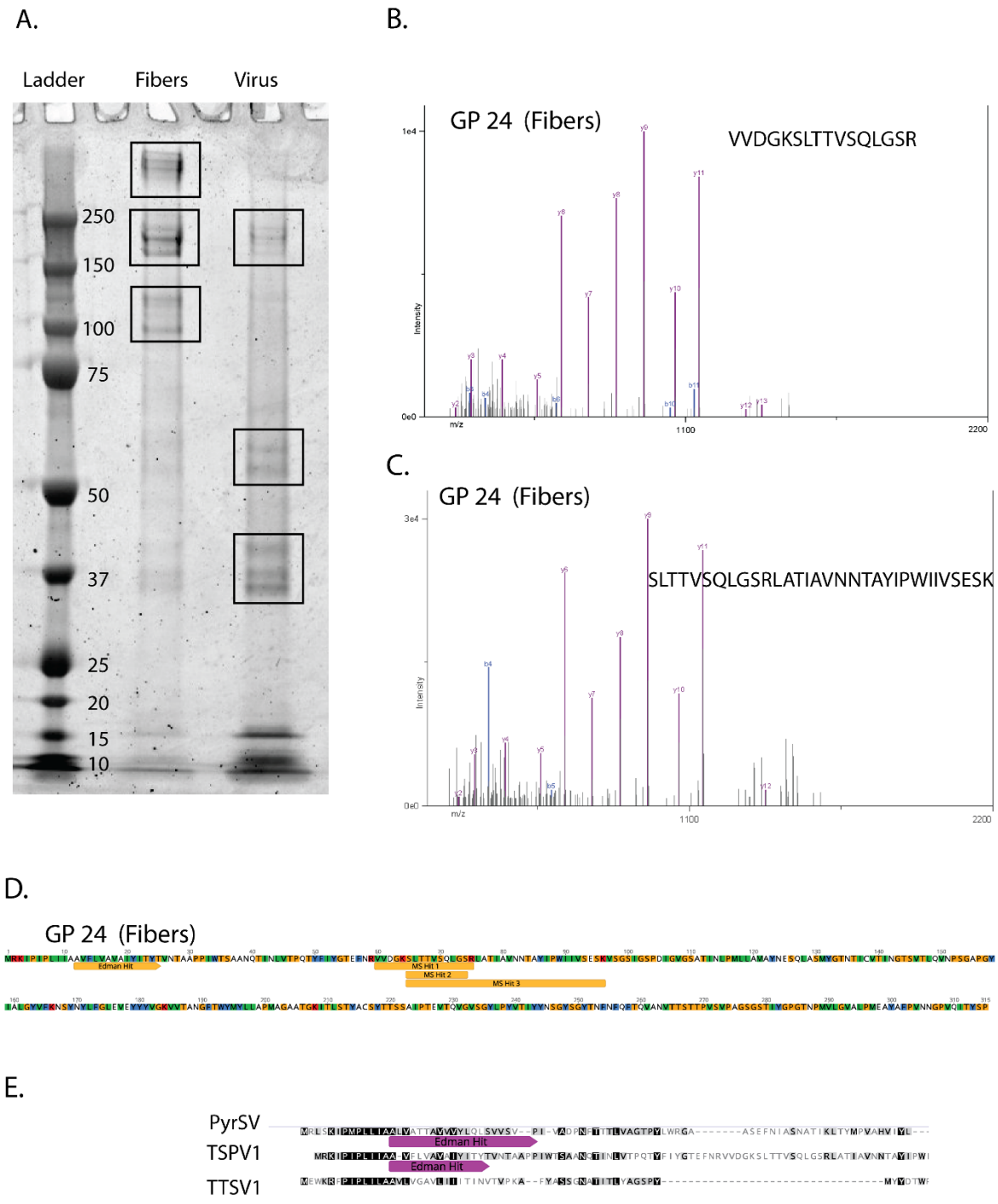


Figure 3.7. Identification of TSPV1 Fibers A) SDS-PAGE gel of proteins from both purified fibers and viral capsids. Boxed regions corresponding to multimers and monomers of the fiber protein were excised and proteolyzed. B&C) Representative spectra obtained from LC MS analysis of digests. D) Gene map showing locations of N-terminal sequencing and LC MS identified peptides from purified fiber samples

corresponding to GP24. MS Hit 1 and MS Hit 3 contain missed cleavage sites due to carbamylation of Lys or Arg residues resulting from high temperature urea denaturation. E) The N-terminal sequence for TSPV1 is recessed 11 amino acids from the start of the protein. This corresponds to the identified N-terminal sequence for PyrSV which begins at an equivalent starting residue. TSPV1 is predicted to contain an N-terminal secretion signal; frequently these signal sequences are cleaved from the protein after export, providing an explanation for the missing 11 amino acids. Note: protein denaturation for electrophoresis of capsid and fiber proteins was accomplished here by incubation in SDS buffer with 8M urea at 98° C for 40 min.

Glycosylation Status of Capsid Proteins

GP9, GP13, and GP24 are all predicted to contain multiple N-glycosylation sites (Table 3.1). To validate the glycosylation status of GP6, GP11, and GP24 intact capsids and fibers were analyzed by SDS-PAGE and glycan gel staining, two different protein ladders were used to serve as negative and positive controls for glycosylation. The results show glycosylation of all virion proteins (Figure 3.8). Notably the bands associated with the viral fibers appear to be highly glycosylated relative to the major coat protein band.

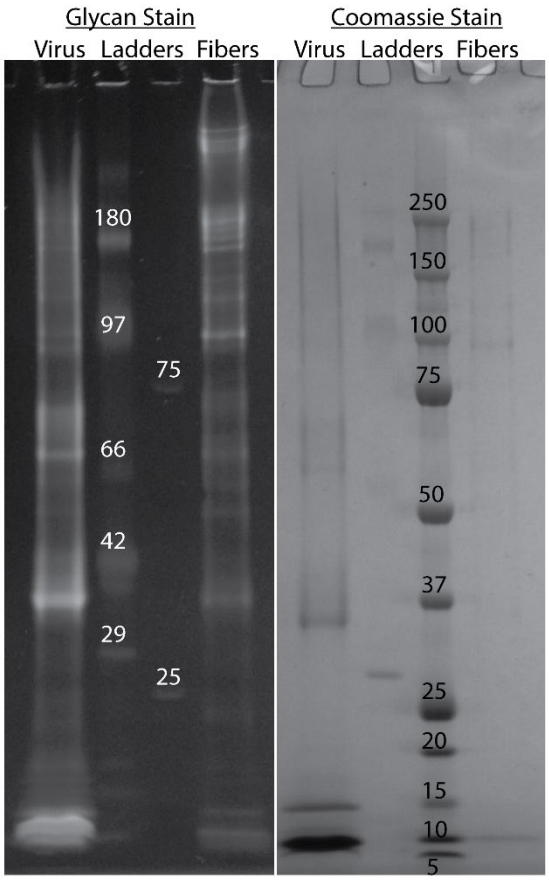


Figure 3.8. Glycosylation status of TSPV1 fiber and capsid proteins. By Coomassie stain, protein bands corresponding to the purified fibers are barely visible, yet when fluorescently stained for glycans these bands show a much higher change in intensity compared to the 10 kDa capsid protein band. This indicates that the GP 24 fiber protein is highly glycosylated, as predicted by our bioinformatic analysis.

Discussion

To our knowledge, the hair-like fibers that project from the TSPV1 virion have not been observed in any other virus. The fibers differ substantially from the spike proteins frequently encountered at the vertices of icosahedral capsids. In contrast to the fibers of TSPV1, these spike proteins have a defined length (generally much shorter than

the capsid diameter) and a defined number of spikes per virion, corresponding to their presence at specific axes of symmetry.

Surprisingly, similar fibers were not observed for the two closely related viruses, PyrSV and TTSV1, even though these viruses contain distant homologues to the GP24 fiber gene (Table 3.1). Protein gels for PyrSV and TTSV1 show a similar banding pattern to TSPV1, with a range of protein bands extending from ~30 kDa to the top of the gel. However, TSPV1 required extended, high temperature, denaturation in saturated urea for the fiber protein to even enter the gel. In contrast, PyrSV and TTSV1 produced this banding pattern under standard denaturation conditions. This could indicate that while GP24 forms a long, extended structure in TSPV1, its homologs create much shorter structures in PyrSV and TTSV1, which are not easily detected by electron microscopy.

Surface exposed proteins in archaea, particularly S-layer proteins, pili, and the archaealum are often heavily glycosylated (249, 250). The determined structures of archaeal glycans show a high level of carbohydrate diversity (202, 249, 251). These post-translational modifications have been shown to impart protease resistance, structural integrity, and salt tolerance to archaeal surface exposed proteins (251). Archaeal viruses frequently contain glycosylated capsid proteins and encode their own glycosyltransferases (27). One role of glycosylation is to mediate virus-host interactions during attachment and entry; examples of this are found in the archaeal virus HRPV1 and the eukaryotic virus HIV-1 (112, 252). Another function described for glycosylation of virus capsid proteins is to mediate hydration of the virion surface, which has been described for the plant virus PVX (253). We have shown the fiber protein is heavily

glycosylated, yet the purpose of this modification remains unknown. Still it likely serves to increase stability of the fibers and/or to mediate recognition of host receptor proteins.

Based on band intensity, GP11 is the MCP for the virus, which agrees with results from Ahn et al. for TTSV1 (216). Enveloped viruses frequently contain a nucleoprotein complex ie. a condensed, superhelical, protein bound form of the DNA (254, 255). In both PyrSV and TSPV1, a nucleoprotein complex is released from ruptured virions. Therefore, we propose that GP11, the MCP, serves to encapsulate the viral DNA into this nucleoprotein complex. Furthermore, GP11 is predicted to have 4 α-helices in its secondary structure, and so we speculate that the MCP forms a 4 helix bundle, a common capsid protein topology to facilitate DNA condensation.

This leaves GP6 with no assigned function. Given its predicted secretion signal, and presence in purified TSPV capsids, we suspect that GP6 is an envelope protein. Aside from this it is difficult to speculate on its particular role in the virus capsid. One possibility is that it could serve as an adapter protein to anchor GP24 into the lipid membrane. GP6 and GP24 show several points of similarity to each other including overall gene length, location of transmembrane helices, predicted signal peptides, and a pairwise identity of 17.8%. Together this suggests that these proteins are paralogs of each other, and perhaps serve related functions.

Major and minor pilin proteins in Archaea contain signal peptides to target them for secretion (199). Makarova et al. used a combination of Flafind and SignalP to predict pilin proteins from 168 archaeal genomes resulting in 5000 pili components (199). Borrowing from their strategy we used the same programs to analyze the genome of

TSPV1 for signal peptides. Using this approach we discovered 5 proteins are predicted to contain signal peptides, two of which (GP6 & GP24) we have identified as capsid proteins. Given this, we suspect the virus may be repurposing the cellular pili secretion pathway to enable assembly of the capsid scaffold and structural proteins.

With regard to virion assembly, we have no direct knowledge of how TPV1 assembles and exits the cell. In general, knowledge of assembly and egress for archaeal viruses is minimal. A unique lysis system used by both the lytic Turriviridae and Rudiviridae creates holes in the host cell, facilitating egress of these viruses (149). In contrast, the non-lytic SSV1 buds from the host cell in a manner reminiscent of eukaryotic viruses (93).

Given the suspected non-lytic lifestyle of TSPV1, and presence of a lipid envelope, we hypothesize that TSPV1 also buds from the host cell. Under this model, the fiber proteins would be secreted and assembled at the host surface prior to virus budding. Still, we have yet to observe either virus budding or the presence of fibers on the host cell surface. It is possible that the non-synchronous chronic infection of *Thermoproteus CP80* by TSPV1 makes observation of a budding event by EM unlikely.

GP8 and GP10 occur in close proximity to the MCP and also contain predicted secretion signals. Furthermore, the synteny of GP8-11 is conserved across the Globuloviridae (Figure 3.4). Since these proteins have not been identified in the capsid of TSPV1, TTSV1, or PyrSV they could instead be secreted to the host membrane and serve to regulate/promote budding of the virus.

It is worthwhile to speculate on the role the virion fibers play in the TSPV1 replication cycle. High temperature, acidic, hot spring environments are typically low total cell density habitats. Combined with the harsh chemical conditions, this creates an evolutionary impetus for viruses to develop strategies to increase the probability of interacting with a host cell. One solution is for the virus to target cellular surface fibers, as is the case for SIRV and STIV(50, 256). However, many fibers are not constitutively expressed on the cell surface, nor necessary for survival (197, 257). In contrast, it is conceivable that TSPV1 may decorate its virion with a structural analog to host pili, and thereby increase its effective surface area and binding probability. Presence of thin fiber extensions on the viral capsid would serve to increase the effective surface area, at a low cost of additional protein subunits in the assembled virion. Despite its theoretical advantages, such a structural solution has not been observed in viruses other than TSPV1. Still, we might expect to see additional isolates of “pilated” viruses emerge from further analysis of viruses infecting Thermoproteales.

CHAPTER FOUR

CONCLUSIONS AND FUTURE DIRECTIONS

Archaeal viruses from thermophilic environments have a strict evolutionary constraint imposed upon them to protect, transport and deliver the viral genome in a metastable capsid within a high temperature, low pH, and low cell density environment. This unique set of circumstances may be a primary cause for the amazing structural diversity observed in archaeal viruses. Three strategies are conceivable to increase the probability of virus binding to a host cell. One is to use a cylindrical high aspect ratio particle; this is employed by rod-shaped viruses like SIRV2 and ATV. Another strategy is to target host cell surface extensions (pili and flagella). Finally, the 3nd strategy is to utilize a capsid morphology with extensions to yield a high surface area.

This research has yielded the 1st pseudo-atomic model for an archaeal virus during attachment. The binding interaction occurs in the upper cleft of the viral turret, a region enriched with conserved residues. Additionally, we now have a working hypothesis for the function of the viral petal proteins. We suspect that these serve as maturation proteins, being present when the virus is first released to prevent it from binding to pili of the cell from which it is leaving. Finally, we have identified a novel locus of pilin like genes that we suspect encode the pilus structure to which STIV binds. This represents only the 3rd virus whose attachment phase has been characterized to any degree. Our study of STIV advances the hypothesis that archaeal viruses target cellular pili in order to achieve specific, efficient attachment.

In addition, we have isolated a novel archaeal virus with morphologic characteristics unlike any previously known. To characterize this new virus and its unique fiber structures we employed an array of genomic, bioinformatic, and proteomic techniques. The results of these methods, in combination with cryo- electron tomography, has enabled us to posit a model for the architecture of the virus. We predict that the abundant GP11 MCP is used to encapsulate the viral genome, while GP24 forms the fiber extensions. Polysaccharides are frequently involved with molecular recognition, for example the heavily glycosylated Env trimer of HIV-1, and glycosylated CD4 of the host. We predict the carbohydrate groups on the viral fibers mediate recognition of glycosylated surface proteins in the host.

STIV Research Conclusions and Future Directions

The targeting of host pili is emerging as a common theme for archaeal viruses. Initially Happonen et al. identified SIRV2 attachment to host pili. We have also observed both the Fusellovirus SSV9 and the unclassified MTIV co-purify with host pili in CsCl gradients, indicating that these of viruses may also use pili structures for attachment (Young Lab unpublished). In all, four independent archaeal virus families show evidence of attachment to host pili.

Frequently we observed dissociated STIV particles bound to multiple host pili. Often these disrupted viral capsids were against the cell surface. From previous studies we know that STIV is susceptible to dissociation, perhaps by sequential shedding of the viral turrets, and then disruption of the major capsid protein shell (144). We suspect that

pili binding induces conformational changes in the viral turret, leading to its instability and dissociation. This then destabilizes the entire capsid shell. Such a conformational change could require simultaneous binding of pili to multiple turret proteins, since multivalent binding is frequently a prerequisite to conformational changes during virus attachment and entry.

Future experiments for this project should include: Collection of a time lapsed series of cryo- EM tomograms, and directed mutagenesis of the turret-pilus binding interface. With regard to directed mutagenesis, our collaborators are generating viral mutants based upon our list of conserved turret interface residues. We are also in the process of characterizing the novel pilus locus, again in collaboration with another laboratory. Collection of additional time points will help elucidate the entire process of attachment and entry for STIV. After binding to host cell pili, the virus must still traverse the S-layer and promote fusion between the tetraether lipids membranes in the virus and host. Neither of these basic biologic processes is understood for archaea; undoubtedly archaeal virus research will be essential for producing answers to fundamental biology questions like these.

Final Remarks and Future Directions for TSPV1

Pili occur frequently on bacterial and archaeal cells, where they mediate DNA exchange, movement, and adhesion. Although it is theoretically possible for virus to use an analogous structure to facilitate host cell attachment, this has never been observed. In this thesis we report on the discovery and initial characterization of the first “pilated”

virus: TSPV1. The thin, fiberous extensions from the TSPV1 capsid do indeed resemble cellular pili.

We were able to acquire negative stain and cryo- electron micrographs of both the virus and fibers. Furthermore, we identified the protein constituent of these structures. Our ability to isolate large amounts of separated viral fibers makes the structure suitable for single particle Cryo-EM. In addition, future studies should attempt cloning and expression of the MCP GP15 and the GP24 fiber protein. This would facilitate crystallographic studies, and potentially structure determination of these proteins which show no detectable sequence homology to known protein folds. To this end we have already produced gene blocks for both GP15 and GP28 containing flanking attB1 and attB2 sites for use in the Gateway cloning system. Expression of GP24 may enable *invitro* synthesis of assembled fiber structures. Synder et al. was able to demonstrate production of STIV viral lysis pyramids in *E. coli* by expression of a single STIV protein. Likewise, it is possible that expression of GP24 will result in assembled fiber structures.

We hypothesized that the fibers of TSPV1 serve to increase the effective virion surface area and facilitate binding to the host cell. Still we have no direct evidence that fibers function in this way. To determine this, the affinity of purified fibers for either whole cells (using surface plasmon resonance) or for cell surface proteins (using a virus overlay protein binding assay) could be examined.

In conclusion, this work has contributed to a greater understanding of the STIV replication cycle. Although one of the best studied archaeal viruses, nothing was known of the attachment/entry phase for STIV. We now realize that one function of the

namesake turrets is attachment to cellular pili. Additionally, we have a cluster of cell surface proteins presumably encoding a novel pilus machinery that is linked to STIV infection. Finally, we have added new archaeal virus (TSPV1) to the small cohort of archaeal viruses in culture. What's more, this virus possesses extensions from the capsid surface unlike anything previously observed in the world of virology. These extensions likely serve an attachment/entry function being morphologically and perhaps functioning analogous to cellular pili.

REFERENCES CITED

1. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090 (1977).
2. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**, 4576-4579 (1990).
3. T. Fouqueau *et al.*, The cutting edge of archaeal transcription. *Emerging Topics in Life Sciences* (2018).
4. L. M. Kelman, Z. Kelman, Archaeal DNA replication. *Annu Rev Genet* **48**, 71-97 (2014).
5. S. D. Bell, S. P. Jackson, Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**, 222-228 (1998).
6. D. Benelli, P. Londei, Translation initiation in Archaea: conserved and domain-specific features. *Biochem Soc Trans* **39**, 89-93 (2011).
7. L. Eme, A. Spang, J. Lombard, C. W. Stairs, T. J. G. Ettema, Archaea and the origin of eukaryotes. *Nat Rev Microbiol* **15**, 711-723 (2017).
8. A. Spang *et al.*, Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet* **14**, e1007080 (2018).
9. K. Zaremba-Niedzwiedzka *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358 (2017).
10. C. J. Castelle, J. F. Banfield, Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* **172**, 1181-1197 (2018).
11. J. Lombard, P. Lopez-Garcia, D. Moreira, The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol* **10**, 507-515 (2012).
12. A. Gambacorta, A. Gliozzi, M. De Rosa, Archaeal lipids and their biotechnological applications. *World J Microbiol Biotechnol* **11**, 115-131 (1995).
13. Y. Koga, Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea* **2012**, 789652 (2012).
14. A. Jacquemet, J. Barbeau, L. Lemiegre, T. Benvegnu, Archaeal tetraether bipolar lipids: Structures, functions and applications. *Biochimie* **91**, 711-717 (2009).
15. J. L. van de Vossenberg, A. J. Driessen, W. N. Konings, The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2**, 163-170 (1998).
16. S. Schouten *et al.*, Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in hot springs of Yellowstone National Park. *Appl Environ Microb* **73**, 6181-6191 (2007).
17. Y. Boucher, Lipids: Biosynthesis, Function, and Evolution. *Archaea: Molecular and Cellular Biology*, 341-353 (2007).
18. S. Jain, A. Caforio, A. J. Driessen, Biosynthesis of archaeal membrane ether lipids. *Front Microbiol* **5**, 641 (2014).
19. E. C. Holmes, What Does Virus Evolution Tell Us about Virus Origins? *Journal of Virology* **85**, 5247-5251 (2011).
20. G. F. Joyce, Building the RNA world. Ribozymes. *Curr Biol* **6**, 965-967 (1996).
21. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *P Natl Acad Sci USA* **114**, E2401-E2410 (2017).

22. L. J. Ghuneim, D. L. Jones, P. N. Golyshin, O. V. Golyshina, Nano-Sized and Filterable Bacteria and Archaea: Biodiversity and Function. *Front Microbiol* **9**, 1971 (2018).
23. N. Philippe *et al.*, Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science* **341**, 281-286 (2013).
24. G. Cornelis *et al.*, Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *P Natl Acad Sci USA* **112**, E487-E496 (2015).
25. 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).
26. M. Krupovic, D. H. Bamford, Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol* **6**, 941-948 (2008).
27. R. Hartman, J. Munson-McGee, M. J. Young, C. M. Lawrence, Survey of high-resolution archaeal virus structures. *Current Opinion in Virology* **36**, 74-83 (2019).
28. D. Prangishvili *et al.*, The enigmatic archaeal virosphere. *Nat Rev Microbiol* **15**, 724-739 (2017).
29. J. H. Munson-McGee, J. C. Snyder, M. J. Young, Archaeal Viruses from High-Temperature Environments. *Genes (Basel)* **9** (2018).
30. J. L. Slonczewski, M. Fujisawa, M. Dopson, T. A. Krulwich, Cytoplasmic pH Measurement and Homeostasis in Bacteria and Archaea. *Adv Microb Physiol* **55**, 1-79 (2009).
31. J. W. Ball, R. B. McMleskey, D. K. Nordstrom (2010) Water-chemistry data for selected springs, geysers, and streams in Yellowstone National Park, Wyoming, 2006-2008. in *Open-File Report* (Reston, VA).
32. N. L. Belkova, K. Tazaki, J. R. Zakharova, V. V. Parfenova, Activity of bacteria in water of hot springs from Southern and Central Kamchatskaya geothermal provinces, Kamchatka Peninsula, Russia. *Microbiol Res* **162**, 99-107 (2007).
33. D. G. Ellis, R. L. W. Bizzoco, Y. Maezato, J. N. Baggett, S. T. Kelley, Microscopic examination of acidic hot springs of Waiotapu, North Island, New Zealand. *New Zeal J Mar Fresh* **39**, 1001-1011 (2005).
34. F. Inagaki *et al.*, Distribution and phylogenetic diversity of the subsurface microbial community in a Japanese epithermal gold mine. *Extremophiles* **7**, 307-317 (2003).
35. D. R. Meyer-Dombard, E. L. Shock, J. P. Amend, Archaeal and bacterial communities in geochemically diverse hot springs of Yellowstone National Park, USA. *Geobiology* **3**, 211-227 (2005).
36. T. Schoenfeld *et al.*, Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol* **74**, 4164-4174 (2008).
37. W. B. Whitman, D. C. Coleman, W. J. Wiebe, Prokaryotes: The unseen majority. *P Natl Acad Sci USA* **95**, 6578-6583 (1998).
38. R. Blumenthal, S. Durell, M. Viard, HIV entry and envelope glycoprotein-mediated fusion. *J Biol Chem* **287**, 40841-40849 (2012).

39. B. Chen, J. J. Chou, Structure of the transmembrane domain of HIV-1 envelope glycoprotein. *FEBS J* **284**, 1171-1177 (2017).
40. P. W. Keller, O. Morrison, R. Vassell, C. D. Weiss, HIV-1 gp41 Residues Modulate CD4-Induced Conformational Changes in the Envelope Glycoprotein and Evolution of a Relaxed Conformation of gp120. *J Virol* **92** (2018).
41. G. Rice *et al.*, The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *PNAS* **101**, 7716-7720 (2004).
42. L. K. Medina-Kauwe, Endocytosis of adenovirus and adenovirus capsid proteins. *Adv Drug Deliv Rev* **55**, 1485-1496 (2003).
43. D. Xia, L. J. Henry, R. D. Gerard, J. Deisenhofer, Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution. *Structure* **2**, 1259-1270 (1994).
44. R. Meng *et al.*, Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nat Commun* **10**, 3130 (2019).
45. P. Chaudhury, T. E. F. Quax, S. V. Albers, Versatile cell surface structures of archaea. *Mol Microbiol* **107**, 298-311 (2018).
46. T. R. D. Costa *et al.*, Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* **166**, 1436-1444 e1410 (2016).
47. M. L. Yap, M. G. Rossmann, Structure and function of bacteriophage T4. *Future Microbiol* **9**, 1319-1327 (2014).
48. J. Bertozzi Silva, Z. Storms, D. Sauvageau, Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* **363** (2016).
49. F. Wang *et al.*, An extensively glycosylated archaeal pilus survives extreme conditions. *Nat Microbiol* 10.1038/s41564-019-0458-x (2019).
50. E. R. Quemin *et al.*, First insights into the entry process of hyperthermophilic archaeal viruses. *J Virol* **87**, 13379-13385 (2013).
51. K. B. Uldahl *et al.*, Life Cycle Characterization of Sulfolobus Monocaudavirus 1, an Extremophilic Spindle-Shaped Virus with Extracellular Tail Development. *J Virol* **90**, 5693-5699 (2016).
52. S. R. Krishnamurthy, D. Wang, Origins and challenges of viral dark matter. *Virus Res* **239**, 136-142 (2017).
53. G. F. Hatfull, R. W. Hendrix, Bacteriophages and their genomes. *Current Opinion in Virology* **1**, 298-303 (2011).
54. C. M. Lawrence *et al.*, Structural and functional studies of archaeal viruses. *J Biol Chem* **284**, 12599-12603 (2009).
55. N. G. Abrescia, D. H. Bamford, J. M. Grimes, D. I. Stuart, Structure unifies the viral universe. *Annu Rev Biochem* **81**, 795-822 (2012).
56. N. L. Dawson *et al.*, CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* **45**, D289-D295 (2017).
57. 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).

58. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, Scop - a Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *Journal of Molecular Biology* **247**, 536-540 (1995).
59. M. C. Merckel, J. T. Huiskonen, D. H. Bamford, A. Goldman, R. Tuma, The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* **18**, 161-170 (2005).
60. B. Gowen, J. K. Bamford, D. H. Bamford, S. D. Fuller, The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. *J Virol* **77**, 7863-7871 (2003).
61. B. Peralta *et al.*, Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biol* **11**, e1001667 (2013).
62. N. J. Stromsten, D. H. Bamford, J. K. Bamford, The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J Virol* **77**, 6314-6321 (2003).
63. T. Mochizuki *et al.*, Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc Natl Acad Sci U S A* **109**, 13386-13391 (2012).
64. M. Krupovic, E. R. Quemin, D. H. Bamford, P. Forterre, D. Prangishvili, Unification of the globally distributed spindle-shaped viruses of the Archaea. *J Virol* **88**, 2354-2358 (2014).
65. D. Prangishvili, T. Mochizuki, M. Krupovic, C. Ictv Report, ICTV Virus Taxonomy Profile: Guttaviridae. *J Gen Virol* **99**, 290-291 (2018).
66. M. Haring, R. Rachel, X. Peng, R. A. Garrett, D. Prangishvili, 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 (2005).
67. N. S. Atanasova, D. H. Bamford, H. M. Oksanen, Haloarchaeal virus morphotypes. *Biochimie* **118**, 333-343 (2015).
68. C. Wagner *et al.*, Isolation and Characterization of Metallosphaera turreted icosahedral virus (MTIV), a founding member of a new family of archaeal viruses. *J Virol* 10.1128/JVI.00925-17 (2017).
69. R. Hochstein *et al.*, Structural studies of Acidianus tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. *Proc Natl Acad Sci U S A* **115**, 2120-2125 (2018).
70. M. K. Pietila *et al.*, Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proc Natl Acad Sci U S A* **110**, 10604-10609 (2013).
71. D. Ptchelkine *et al.*, Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nat Commun* **8**, 1436 (2017).
72. K. M. Stedman, M. DeYoung, M. Saha, M. B. Sherman, M. C. Morais, Structural insights into the architecture of the hyperthermophilic Fusellovirus SSV1. *Virology* **474**, 105-109 (2015).
73. C. Hong *et al.*, Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. *Proc Natl Acad Sci U S A* **112**, 2449-2454 (2015).

74. F. DiMaio *et al.*, Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914-917 (2015).
75. B. R. Szymczyna *et al.*, Synergy of NMR, computation, and X-ray crystallography for structural biology. *Structure* **17**, 499-507 (2009).
76. A. Goulet *et al.*, Acidianus filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. *Proc Natl Acad Sci U S A* **106**, 21155-21160 (2009).
77. P. Kasson *et al.*, Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *Elife* **6** (2017).
78. Y. Liu *et al.*, Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nat Commun* **9**, 3360 (2018).
79. D. Gil-Carton *et al.*, Insight into the Assembly of Viruses with Vertical Single beta-barrel Major Capsid Proteins. *Structure* **23**, 1866-1877 (2015).
80. S. T. Jaatinen, L. J. Happonen, P. Laurinmaki, S. J. Butcher, D. H. Bamford, Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic Thermus thermophilus. *Virology* **379**, 10-19 (2008).
81. 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).
82. K. El Omari *et al.*, The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nat Commun* **10**, 846 (2019).
83. R. Khayat *et al.*, Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *PNAS* **102**, 18944-18949 (2005).
84. S. D. Benson, J. K. Bamford, D. H. Bamford, R. M. Burnett, Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* **98**, 825-833 (1999).
85. A. P. Aalto *et al.*, Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing Salisaeta icosahedral phage 1. *Proc Natl Acad Sci U S A* **109**, 7079-7084 (2012).
86. H. T. Jaalinoja *et al.*, Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proc Natl Acad Sci U S A* **105**, 8008-8013 (2008).
87. K. F. Shortridge, F. Biddle, The proteins of adenovirus type 5. *Arch Gesamte Virusforsch* **29**, 1-24 (1970).
88. L. Mindich, D. Bamford, T. McGraw, G. Mackenzie, Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J Virol* **44**, 1021-1030 (1982).
89. D. Prangishvili, M. Krupovic, A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". *Arch Virol* **157**, 791-795 (2012).
90. X. Peng *et al.*, Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology* **291**, 226-234 (2001).

91. E. S. Boyd, T. L. Hamilton, J. Wang, L. He, C. L. Zhang, The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. *Front Microbiol* **4**, 62 (2013).
92. M. Kates, Archaebacterial lipids: structure, biosynthesis and function. *Biochem Soc Symp* **58**, 51-72 (1992).
93. E. R. Quemin *et al.*, Eukaryotic-Like Virus Budding in Archaea. *MBio* **7** (2016).
94. 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).
95. T. Mochizuki *et al.*, Diversity of viruses of the hyperthermophilic archaeal genus Aeropyrum, and isolation of the Aeropyrum pernix bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. *Virology* **402**, 347-354 (2010).
96. C. S. Chan *et al.*, Effects of Physiochemical Factors on Prokaryotic Biodiversity in Malaysian Circumneutral Hot Springs. *Front Microbiol* **8**, 1252 (2017).
97. A. Y. Merkel *et al.*, Microbial diversity and autotrophic activity in Kamchatka hot springs. *Extremophiles* **21**, 307-317 (2017).
98. S. Wang *et al.*, Control of temperature on microbial community structure in hot springs of the Tibetan Plateau. *PLoS One* **8**, e62901 (2013).
99. M. F. Siliakus, J. van der Oost, S. W. M. Kengen, Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* **21**, 651-670 (2017).
100. D. L. Valentine, Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat Rev Microbiol* **5**, 316-323 (2007).
101. J. L. van de Vossenberg, T. Ubbink-Kok, M. G. Elferink, A. J. Driessen, W. N. Konings, Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea. *Mol Microbiol* **18**, 925-932 (1995).
102. W. R. Wikoff *et al.*, Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129-2133 (2000).
103. A. Goulet *et al.*, Getting the best out of long-wavelength X-rays: de novo chlorine/sulfur SAD phasing of a structural protein from ATV. *Acta Crystallogr D Biol Crystallogr* **66**, 304-308 (2010).
104. D. Prangishvili *et al.*, Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol* **359**, 1203-1216 (2006).
105. S. Erdmann *et al.*, A novel single-tailed fusiform Sulfolobus virus STSV2 infecting model Sulfolobus species. *Extremophiles* **18**, 51-60 (2014).
106. R. A. Hochstein, M. J. Amenabar, J. H. Munson-McGee, E. S. Boyd, M. J. Young, Acidianus Tailed Spindle Virus: a New Archaeal Large Tailed Spindle Virus Discovered by Culture-Independent Methods. *J Virol* **90**, 3458-3468 (2016).
107. K. J. Hanhijarvi, G. Ziedaite, E. Haeggstrom, D. H. Bamford, Temperature and pH dependence of DNA ejection from archaeal lemon-shaped virus His1. *Eur Biophys J* **45**, 435-442 (2016).

108. K. J. Hanhijarvi, G. Ziedaite, M. K. Pietila, E. Haeggstrom, D. H. Bamford, DNA ejection from an archaeal virus--a single-molecule approach. *Biophys J* **104**, 2264-2272 (2013).
109. S. Dharmavaram, J. Rudnick, C. M. Lawrence, R. F. Bruinsma, Smectic viral capsids and the aneurysm instability. *J Phys Condens Matter* **30**, 204004 (2018).
110. C. Delaunay, Sur la surface de révolution dont la courbure moyenne est constante. *J. Math. Pures Appl.* **6**, 309-320 (1841).
111. W. S. Maaty *et al.*, Characterization of the archaeal thermophile Sulfolobus turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. *J Virol* **80**, 7625-7635 (2006).
112. L. Kandiba *et al.*, Diversity in prokaryotic glycosylation: an archaeal-derived N-linked glycan contains legionaminic acid. *Mol Microbiol* **84**, 578-593 (2012).
113. 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).
114. N. F. Steinmetz *et al.*, Site-specific and Spatially Controlled Addressability of a New Viral Nanobuilding Block: Sulfolobus islandicus Rod-shaped Virus 2. *Adv Funct Mater* **18**, 3478-3486 (2008).
115. M. Haring *et al.*, 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 (2004).
116. Y. Liu *et al.*, A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea. *J Virol* **91** (2017).
117. X. Peng, T. Basta, M. Haring, R. A. Garrett, D. Prangishvili, Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology* **364**, 237-243 (2007).
118. N. G. Jayaprakash, A. Surolia, Role of glycosylation in nucleating protein folding and stability. *Biochem J* **474**, 2333-2347 (2017).
119. X. Xiang *et al.*, Sulfolobus tengchongensis spindle-shaped virus STSV1: virus-host interactions and genomic features. *J Virol* **79**, 8677-8686 (2005).
120. M. K. Pietila, N. S. Atanasova, H. M. Oksanen, D. H. Bamford, Modified coat protein forms the flexible spindle-shaped virion of haloarchaeal virus His1. *Environ Microbiol* **15**, 1674-1686 (2013).
121. D. H. Bamford *et al.*, Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon Haloarcula hispanica. *J Virol* **79**, 9097-9107 (2005).
122. T. A. Demina *et al.*, HCIV-1 and Other Tailless Icosahedral Internal Membrane-Containing Viruses of the Family Sphaerolipoviridae. *Viruses* **9** (2017).
123. N. S. Atanasova *et al.*, Comparison of Lipid-Containing Bacterial and Archaeal Viruses. *Adv Virus Res* **92**, 1-61 (2015).
124. B. Tenchov, E. M. Vescio, G. D. Sprott, M. L. Zeidel, J. C. Mathai, Salt tolerance of archaeal extremely halophilic lipid membranes. *Journal of Biological Chemistry* **281**, 10016-10023 (2006).
125. S. C. Harrison, Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698 (2008).

126. J. H. Hurley, P. I. Hanson, Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat Rev Mol Cell Biol* **11**, 556-566 (2010).
127. J. C. Snyder, R. Y. Samson, S. K. Brumfield, S. D. Bell, M. J. Young, Functional interplay between a virus and the ESCRT machinery in archaea. *Proc Natl Acad Sci U S A* **110**, 10783-10787 (2013).
128. J. C. Snyder, S. K. Brumfield, N. Peng, Q. She, M. J. Young, Sulfolobus turreted icosahedral virus c92 protein responsible for the formation of pyramid-like cellular lysis structures. *J Virol* **85**, 6287-6292 (2011).
129. E. F. DeLong, N. R. Pace, Environmental diversity of bacteria and archaea. *Syst Biol* **50**, 470-478 (2001).
130. M. B. Karner, E. F. DeLong, D. M. Karl, Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507-510 (2001).
131. M. Pina, A. Bize, P. Forterre, D. Prangishvili, The archeoviruses. *FEMS Microbiol Rev* (2011).
132. J. S. Lipp, Y. Morono, F. Inagaki, K. U. Hinrichs, Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991-994 (2008).
133. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034-1039 (2008).
134. P. S. Adam, G. Borrel, C. Brochier-Armanet, S. Gribaldo, The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**, 2407-2425 (2017).
135. S. Leininger *et al.*, Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**, 806-809 (2006).
136. C. Wuchter *et al.*, Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A* **103**, 12317-12322 (2006).
137. R. K. Thauer, A. K. Kaster, H. Seedorf, W. Buckel, R. Hedderich, Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**, 579-591 (2008).
138. J. Lloyd-Price, G. Abu-Ali, C. Huttenhower, The healthy human microbiome. *Genome Med* **8**, 51 (2016).
139. S. K. Brumfield *et al.*, Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus sulfolobus turreted icosahedral virus. *J Virol* **83**, 5964-5970 (2009).
140. N. Dellas *et al.*, Structure-Based Mutagenesis of Sulfolobus Turreted Icosahedral Virus B204 Reveals Essential Residues in the Virion-Associated DNA-Packaging ATPase. *J Virol* **90**, 2729-2739 (2015).
141. C. Y. Fu, J. E. Johnson, Structure and cell biology of archaeal virus STIV. *Curr Opin Virol* **2**, 122-127 (2012).
142. C. Y. Fu *et al.*, In vivo assembly of an archaeal virus studied with whole-cell electron cryotomography. *Structure* **18**, 1579-1586 (2010).
143. J. Fulton *et al.*, Genetics, biochemistry and structure of the archaeal virus STIV. *Biochem Soc Trans* **37**, 114-117 (2009).

144. R. Khayat, C. Y. Fu, A. C. Ortmann, M. J. Young, J. E. Johnson, The architecture and chemical stability of the archaeal Sulfolobus turreted icosahedral virus. *J Virol* **84**, 9575-9583 (2010).
145. E. T. Larson *et al.*, A winged-helix protein from Sulfolobus turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus. *Virology* **368**, 249-261 (2007).
146. E. T. Larson, D. Reiter, M. Young, C. M. Lawrence, Structure of A197 from Sulfolobus turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. *J Virol* **80**, 7636-7644 (2006).
147. W. S. Maaty *et al.*, Proteomic analysis of Sulfolobus solfataricus during Sulfolobus Turreted Icosahedral Virus infection. *J Proteome Res* **11**, 1420-1432 (2012).
148. A. C. Ortmann *et al.*, Transcriptome analysis of infection of the archaeon Sulfolobus solfataricus with Sulfolobus turreted icosahedral virus. *J Virol* **82**, 4874-4883 (2008).
149. J. C. Snyder *et al.*, Insights into a viral lytic pathway from an archaeal virus-host system. *J Virol* **87**, 2186-2192 (2013).
150. J. F. Wirth *et al.*, Development of a genetic system for the archaeal virus Sulfolobus turreted icosahedral virus (STIV). *Virology* **415**, 6-11 (2011).
151. L. J. Happonen *et al.*, Familial relationships in hyperthermo- and acidophilic archaeal viruses. *J Virol* **84**, 4747-4754 (2010).
152. N. Dellas *et al.*, Structure-Based Mutagenesis of Sulfolobus Turreted Icosahedral Virus B204 Reveals Essential Residues in the Virion-Associated DNA-Packaging ATPase. *J Virol* **90**, 2729-2739 (2016).
153. J. Caldentey, R. Tuma, D. H. Bamford, Assembly of bacteriophage PRD1 spike complex: Role of the multidomain protein P5. *Biochemistry-Us* **39**, 10566-10573 (2000).
154. M. C. Bewley, K. Springer, Y. B. Zhang, P. Freimuth, J. M. Flanagan, Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**, 1579-1583 (1999).
155. E. Seiradake, H. Lortat-Jacob, O. Billet, E. J. Kremer, S. Cusack, Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in complex with the D1 domain of coxsackie and adenovirus receptor. *J Biol Chem* **281**, 33704-33716 (2006).
156. T. T. Puck, A. Garen, J. Cline, The mechanism of virus attachment to host cells. I. The role of ions in the primary reaction. *J Exp Med* **93**, 65-88 (1951).
157. B. Bolduc, Wirth, J., Mazurie, A., Young, M., Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *The ISME Journal* **In press** (2015).
158. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
159. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

160. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
161. M. A. Larkin *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
162. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**, 1547-1549 (2018).
163. M. Khoshouei, S. Pfeffer, W. Baumeister, F. Forster, R. Danev, Subtomogram analysis using the Volta phase plate. *J Struct Biol* **197**, 94-101 (2017).
164. R. Danev, W. Baumeister, Expanding the boundaries of cryo-EM with phase plates. *Curr Opin Struct Biol* **46**, 87-94 (2017).
165. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
166. R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko, W. Baumeister, Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc Natl Acad Sci U S A* **111**, 15635-15640 (2014).
167. Y. Fukuda, U. Laugks, V. Lucic, W. Baumeister, R. Danev, Electron cryotomography of vitrified cells with a Volta phase plate. *J Struct Biol* **190**, 143-154 (2015).
168. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71-76 (1996).
169. J. M. Heumann, A. Hoenger, D. N. Mastronarde, Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J Struct Biol* **175**, 288-299 (2011).
170. D. Nicastro *et al.*, The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313**, 944-948 (2006).
171. T. A. Jones, "A, yaap, asap, @#*? A set of averaging programs" in Molecular Replacement*,* E. J. Dodson, S. Gover, W. Wolf, Eds. (SERC Daresbury Laboratory, Warrington, 1992), pp. 91-105.
172. G. J. Kleywegt, R. J. Read, Not your average density. *Structure* **5**, 1557-1569 (1997).
173. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
174. S. Abbott, A. Iudin, P. K. Korir, S. Somasundharam, A. Patwardhan, EMDB Web Resources. *Curr Protoc Bioinformatics* **61**, 5 10 11-15 10 12 (2018).
175. C. Zhang, T. E. Cooper, D. J. Krause, R. J. Whitaker, Augmenting the genetic toolbox for Sulfolobus islandicus with a stringent positive selectable marker for agmatine prototrophy. *Appl Environ Microbiol* **79**, 5539-5549 (2013).
176. R. E. Anderson, A. Kouris, C. H. Seward, K. M. Campbell, R. J. Whitaker, Structured Populations of Sulfolobus acidocaldarius with Susceptibility to Mobile Genetic Elements. *Genome Biol Evol* **9**, 1699-1710 (2017).
177. A. Bankevich *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477 (2012).

178. D. E. Deatherage, J. E. Barrick, Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* **1151**, 165-188 (2014).
179. J. Huerta-Cepas *et al.*, Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115-2122 (2017).
180. J. Soding, A. Biegert, A. N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244-248 (2005).
181. Z. Szabo *et al.*, Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J Bacteriol* **189**, 772-778 (2007).
182. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 A resolution. *Nature* **276**, 368-373 (1978).
183. J. S. Richardson, The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167-339 (1981).
184. M. G. Rossmann *et al.*, Structural comparisons of some small spherical plant viruses. *J Mol Biol* **165**, 711-736 (1983).
185. L. Holm, P. Rosenstrom, Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549 (2010).
186. S. Najmudin *et al.*, Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: crystal structure of the CBM22-1-GH10 modules complexed with xylohexaose. *J Struct Biol* **172**, 353-362 (2010).
187. J. C. Leavitt *et al.*, The tip of the tail needle affects the rate of DNA delivery by bacteriophage P22. *PLoS One* **8**, e70936 (2013).
188. E. Wu, G. R. Nemerow, Virus yoga: the role of flexibility in virus host cell recognition. *Trends Microbiol* **12**, 162-169 (2004).
189. C. C. Lee *et al.*, Crystal structure of infectious bursal disease virus VP2 subviral particle at 2.6A resolution: implications in virion assembly and immunogenicity. *J Struct Biol* **155**, 74-86 (2006).
190. R. Moll, G. Schäfer, Chemiosmotic H^+ cycling across the plasma membrane of the thermoacidophilic archaebacterium Sulfolobus acidocaldarius. *FEBS Lett* **232**, 359-363 (1988).
191. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007).
192. A. Bhardwaj, I. J. Molineux, S. R. Casjens, G. Cingolani, Atomic structure of bacteriophage Sf6 tail needle knob. *J Biol Chem* **286**, 30867-30877 (2011).
193. J. M. Bergelson *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320-1323 (1997).
194. P. W. Roelvink *et al.*, The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. *J Virol* **72**, 7909-7915 (1998).

195. P. W. Roelvink, G. Mi Lee, D. A. Einfeld, I. Kovesdi, T. J. Wickham, Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. *Science* **286**, 1568-1571 (1999).
196. J. Svirskaite, H. M. Oksanen, R. Daugelavicius, D. H. Bamford, Monitoring Physiological Changes in Haloarchaeal Cell during Virus Release. *Viruses* **8**, 59 (2016).
197. A. L. Henche *et al.*, Structure and function of the adhesive type IV pilus of Sulfolobus acidocaldarius. *Environ Microbiol* **14**, 3188-3202 (2012).
198. Q. She *et al.*, The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A* **98**, 7835-7840 (2001).
199. K. S. Makarova, E. V. Koonin, S. V. Albers, Diversity and Evolution of Type IV pili Systems in Archaea. *Front Microbiol* **7**, 667 (2016).
200. S. V. Albers, B. H. Meyer, The archaeal cell envelope. *Nat Rev Microbiol* **9**, 414-426 (2011).
201. A. Klingl, S-layer and cytoplasmic membrane - exceptions from the typical archaeal cell wall with a focus on double membranes. *Front Microbiol* **5**, 624 (2014).
202. K. F. Jarrell *et al.*, N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* **78**, 304-341 (2014).
203. K. F. Jarrell, Y. Ding, D. B. Nair, S. Siu, Surface appendages of archaea: structure, function, genetics and assembly. *Life (Basel)* **3**, 86-117 (2013).
204. A. C. Malmborg, E. Soderlind, L. Frost, C. A. Borrebaeck, Selective phage infection mediated by epitope expression on F pilus. *J Mol Biol* **273**, 544-551 (1997).
205. J. Manchak, K. G. Anthony, L. S. Frost, Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. *Mol Microbiol* **43**, 195-205 (2002).
206. J. L. Pate, S. J. Petzold, T. H. Umbreit, Two flagellotropic phages and one pilus-specific phage active against Asticcacaulis biprosthecum. *Virology* **94**, 24-37 (1979).
207. T. B. Shea, E. Seaman, SP3: a flagellotropic bacteriophage of Bacillus subtilis. *J Gen Virol* **65 (Pt 11)**, 2073-2076 (1984).
208. L. Deng *et al.*, Unveiling cell surface and type IV secretion proteins responsible for archaeal rudivirus entry. *J Virol* **88**, 10264-10268 (2014).
209. M. M. Poranen, R. Daugelavicius, D. H. Bamford, Common principles in viral entry. *Annu Rev Microbiol* **56**, 521-538 (2002).
210. L. Craig, K. T. Forest, B. Maier, Type IV pili: dynamics, biophysics and functional consequences. *Nat Rev Microbiol* **17**, 429-440 (2019).
211. H. C. Berg, R. A. Anderson, Bacteria Swim by Rotating Their Flagellar Filaments. *Nature* **245**, 380-382 (1973).
212. A. D. T. Samuel *et al.*, Flagellar determinants of bacterial sensitivity to chi-phage. *P Natl Acad Sci USA* **96**, 9863-9866 (1999).

213. S. Mantynen, L. R. Sundberg, H. M. Oksanen, M. M. Poranen, Half a Century of Research on Membrane-Containing Bacteriophages: Bringing New Concepts to Modern Virology. *Viruses* **11** (2019).
214. H. M. Kivela, R. Daugelavicius, R. H. Hankkio, J. K. Bamford, D. H. Bamford, Penetration of membrane-containing double-stranded-DNA bacteriophage PM2 into Pseudoalteromonas hosts. *J Bacteriol* **186**, 5342-5354 (2004).
215. E. W. Sayers *et al.*, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **47**, D23-D28 (2019).
216. D. G. Ahn *et al.*, TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote Thermoproteus tenax. *Virology* **351**, 280-290 (2006).
217. C. Geslin *et al.*, PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "Pyrococcus abyssi". *J Bacteriol* **185**, 3888-3894 (2003).
218. A. Gorlas, E. V. Koonin, N. Bienvenu, D. Prieur, C. Geslin, TPV1, the first virus isolated from the hyperthermophilic genus Thermococcus. *Environ Microbiol* **14**, 503-516 (2012).
219. T. Mochizuki, Y. Sako, D. Prangishvili, Provirus Induction in Hyperthermophilic Archaea: Characterization of Aeropyrum pernix Spindle-Shaped Virus 1 and Aeropyrum pernix Ovoid Virus 1. *Journal of Bacteriology* **193**, 5412-5419 (2011).
220. W. Zillig *et al.*, Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. *Fems Microbiology Reviews* **18**, 225-236 (1996).
221. L. A. Hug *et al.*, A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).
222. D. H. Parks *et al.*, Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542 (2017).
223. Z. J. Jay *et al.*, The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. *Environmental Microbiology* **18**, 4755-4769 (2016).
224. M. R. Urschel, T. L. Hamilton, E. E. Roden, E. S. Boyd, Substrate preference, uptake kinetics and bioenergetics in a facultatively autotrophic, thermoacidophilic crenarchaeote. *FEMS Microbiol Ecol* **92**, fiw069 (2016).
225. K. J. Yim, H. S. Song, J. S. Choi, S. W. Roh, Thermoproteus thermophilus sp nov., a hyperthermophilic crenarchaeon isolated from solfataric soil. *Int J Syst Evol Micr* **65**, 2507-2510 (2015).
226. W. Zillig *et al.*, Thermoproteales - a Novel Type of Extremely Thermoacidophilic Anaerobic Archaebacteria Isolated from Icelandic Solfataras. *Zbl Bakt Mik Hyg I C* **2**, 205-227 (1981).
227. E. A. Bonchosmolovskaya, M. L. Miroshnichenko, N. A. Kostrikina, N. A. Chernych, G. A. Zavarzin, Thermoproteus-Uzoniensis Sp-Nov, a New Extremely Thermophilic Archaebacterium from Kamchatka Continental Hot-Springs. *Arch Microbiol* **154**, 556-559 (1990).

228. B. Siebers *et al.*, The complete genome sequence of Thermoproteus tenax: a physiologically versatile member of the Crenarchaeota. *PLoS One* **6**, e24222 (2011).
229. T. Rodrigues-Oliveira, A. Belmok, D. Vasconcellos, B. Schuster, C. M. Kyaw, Archaeal S-Layers: Overview and Current State of the Art. *Front Microbiol* **8**, 2597 (2017).
230. D. Janekovic *et al.*, Ttv1, Ttv2 and Ttv3, a Family of Viruses of the Extremely Thermophilic, Anaerobic, Sulfur Reducing Archaebacterium Thermoproteus-Tenax. *Mol Gen Genet* **192**, 39-45 (1983).
231. E. S. Boyd *et al.*, Isolation, characterization, and ecology of sulfur-respiring crenarchaea inhabiting acid-sulfate-chloride-containing geothermal springs in Yellowstone National Park. *Appl Environ Microbiol* **73**, 6669-6677 (2007).
232. A. C. Darling, B. Mau, F. R. Blattner, N. T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394-1403 (2004).
233. A. L. Delcher, D. Harmon, S. Kasif, O. White, S. L. Salzberg, Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636-4641 (1999).
234. D. Hyatt *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
235. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).
236. J. J. A. Armenteros *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-+ (2019).
237. D. W. A. Buchan, D. T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* **47**, W402-W407 (2019).
238. J. Song *et al.*, PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One* **7**, e50300 (2012).
239. C. Steentoft *et al.*, Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *Embo J* **32**, 1478-1488 (2013).
240. C. H. Andrewes, D. M. Horstmann, The susceptibility of viruses to ethyl ether. *J Gen Microbiol* **3**, 290-297 (1949).
241. M. Oke *et al.*, The Scottish Structural Proteomics Facility: targets, methods and outputs. *J Struct Funct Genomics* **11**, 167-180 (2010).
242. I. Mir-Sanchis *et al.*, Staphylococcal SCCmec elements encode an active MCM-like helicase and thus may be replicative. *Nature Structural & Molecular Biology* **23**, 891-898 (2016).
243. E. A. Abbate, J. M. Berger, M. R. Botchan, The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2. *Genes Dev* **18**, 1981-1996 (2004).
244. D. N. Frick, A. M. I. Lam, Understanding helicases as a means of virus control. *Curr Pharm Design* **12**, 1315-1338 (2006).

245. R. H. Kretsinger, V. N. Uversky, E. A. Permi̇akov, *Encyclopedia of metalloproteins* (Springer Reference, New York, 2013), pp. 4 volumes (lvii, 2574 pages).
246. J. L. Pan, J. C. Bardwell, The origami of thioredoxin-like folds. *Protein Sci* **15**, 2217-2227 (2006).
247. G. Ren *et al.*, Properties of the thioredoxin fold superfamily are modulated by a single amino acid residue. *J Biol Chem* **284**, 10150-10159 (2009).
248. A. F. Ellen, B. Zolghadr, A. M. Driessen, S. V. Albers, Shaping the archaeal cell envelope. *Archaea* **2010**, 608243 (2010).
249. L. Kandiba, J. Eichler, Archaeal S-layer glycoproteins: post-translational modification in the face of extremes. *Front Microbiol* **5**, 661 (2014).
250. C. Schaffer, P. Messner, Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* **41**, 49-91 (2017).
251. J. Eichler, M. Koomey, Sweet New Roles for Protein Glycosylation in Prokaryotes. *Trends in Microbiology* **25**, 662-672 (2017).
252. D. C. Montefiori, W. E. Robinson, Jr., W. M. Mitchell, Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A* **85**, 9248-9252 (1988).
253. L. A. Baratova *et al.*, N-Terminal segment of potato virus X coat protein subunits is glycosylated and mediates formation of a bound water shell on the virion surface. *Eur J Biochem* **271**, 3136-3145 (2004).
254. A. Ariza *et al.*, Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic Acids Res* **41**, 5912-5926 (2013).
255. S. Chenavas *et al.*, Monomeric nucleoprotein of influenza A virus. *PLoS Pathog* **9**, e1003275 (2013).
256. R. Hartman, The Molecular Mechanism of Cellular Attachment for an Archaeal Virus. *Structure* **In Review** (2019).
257. L. Hoffmann *et al.*, Structure and interactions of the archaeal motility repression module ArnA-ArnB that modulates archaellum gene expression in Sulfolobus acidocaldarius. *Journal of Biological Chemistry* **294**, 7460-7471 (2019).

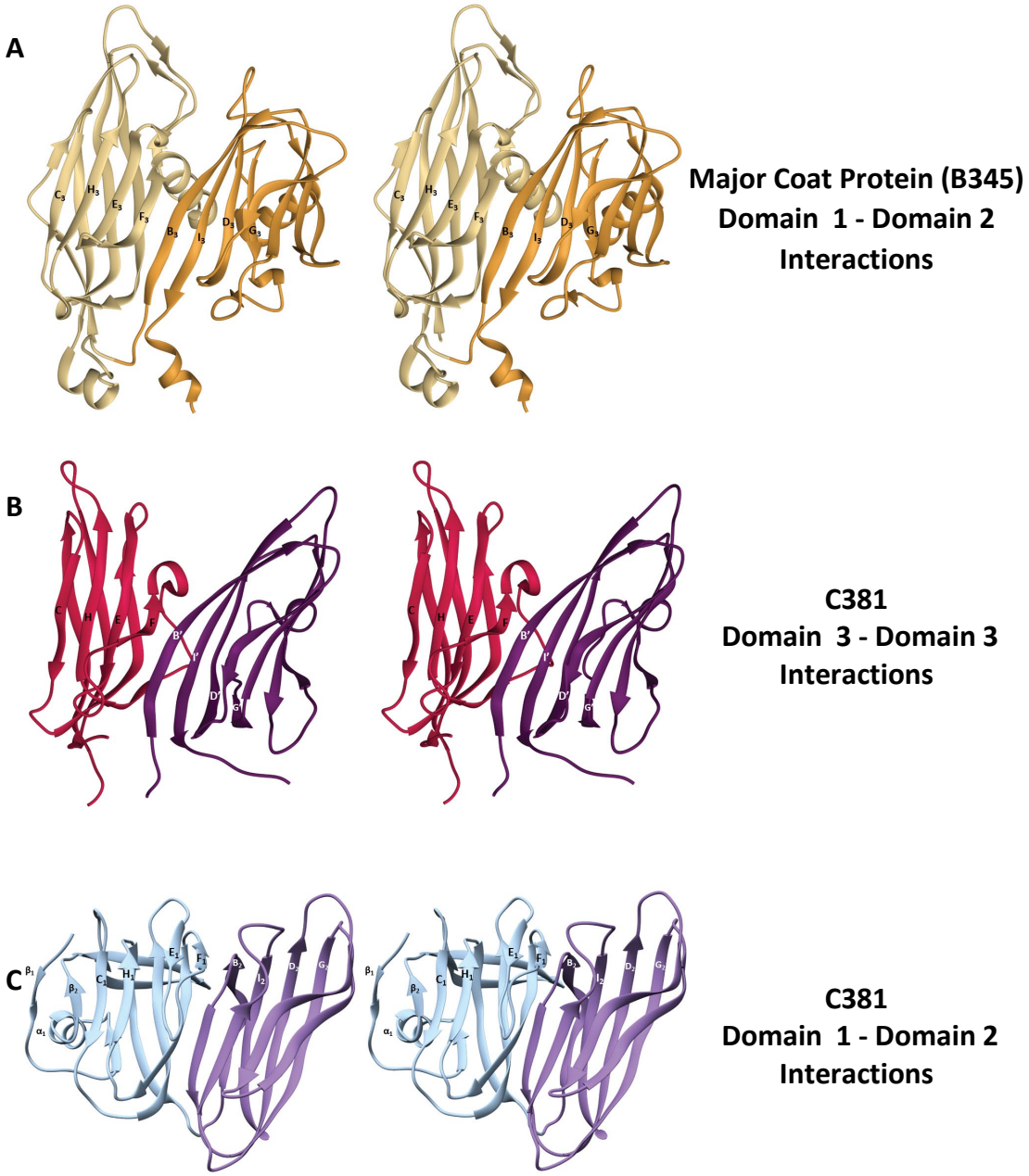
APPENDICES

APPENDIX A

SUPPLEMENTAL TABLES AND FIGURES FROM CHAPTER TWO

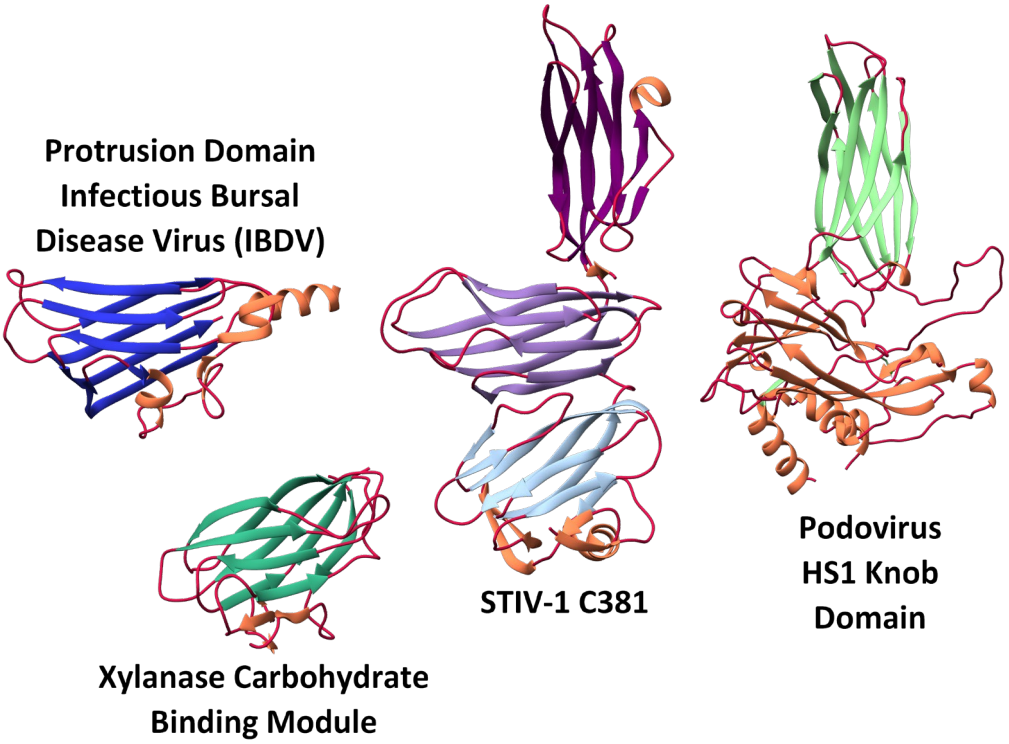
Supplemental Table 1 Viral Adsorption Rate Constants

Virus	Adsorbtion Rate Constant k (ml/min)
STIV-1	2×10^{-9}
SMV1	7×10^{-9}
SIRV2	2×10^{-8}
SSV9	8.39×10^{-11}
T1	3.1×10^{-9}
T2	2.1×10^{-9}
T3	3.0×10^{-9}
T4	3.1×10^{-9}
φ6	3.3×10^{-10}
SH1	1.1×10^{-11}
HHTV1	2.9×10^{-13}
His1	1.9×10^{-12}
His2	5.0×10^{-12}
PBCV-1	5×10^{-9}

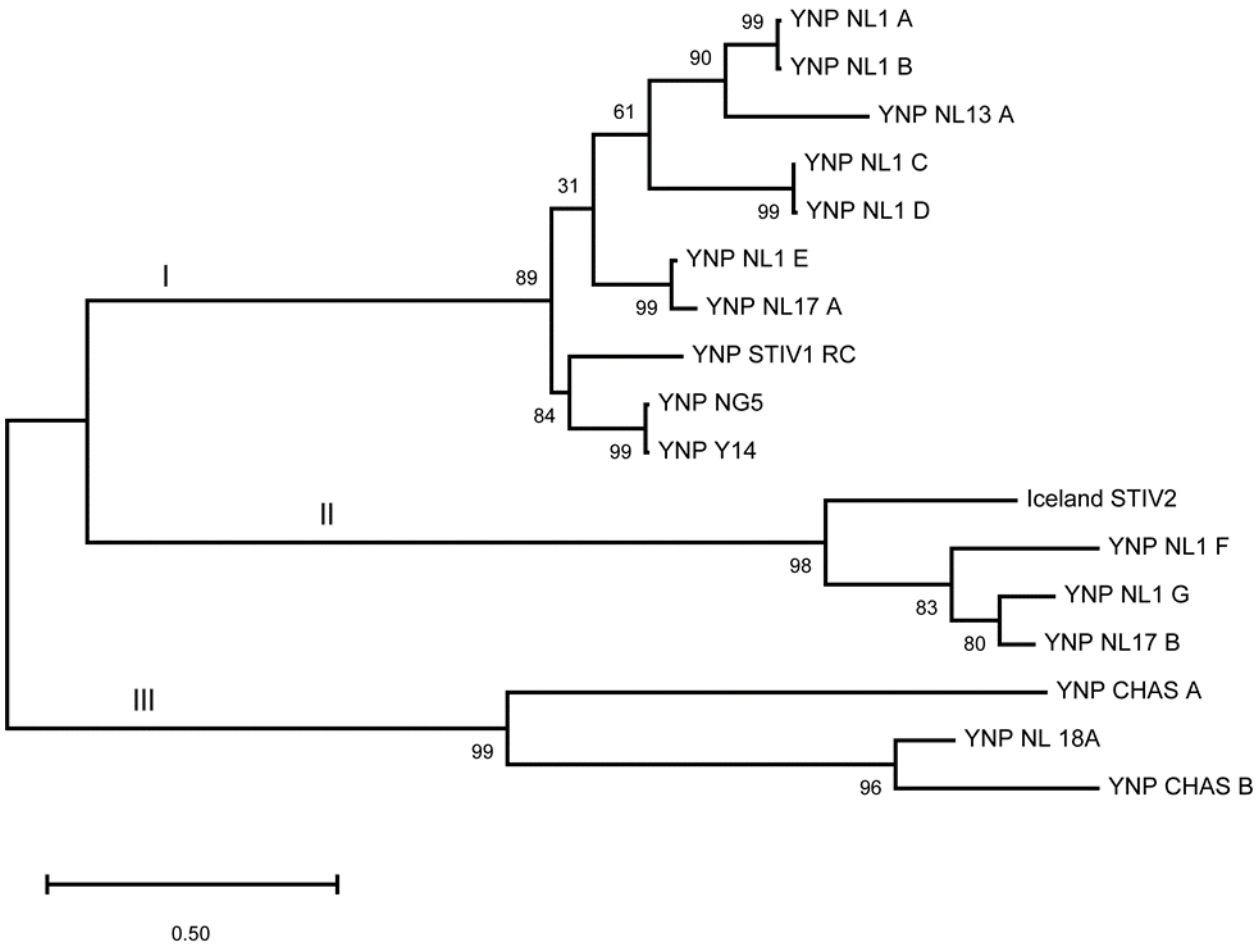


Supplemental Figure 1. Domain Organization in STIV Double Jellyroll Proteins (Stereo Pairs). A) The major coat protein (B345) that forms the STIV capsid surface is a classic double jellyroll. Its structure determination extended observation of the PRD1 viral lineage into the archaeal domain of life. Note the relative orientation of the two domains giving rise to a central 8-stranded β-sheet composed of strands CHEFB'I'D'G'. B) Two copies of the domain 3 C381 turret protein in adjacent subunits. The relative domain orientation gives rise to an 8-stranded β-sheet (C3H3E3F3B3I3D3G3) extending across

both subunits. C) A similar two domain arrangement between domains 1 and 2 of the C381 turret protein (C1H1E1F1B2I2D2G2)

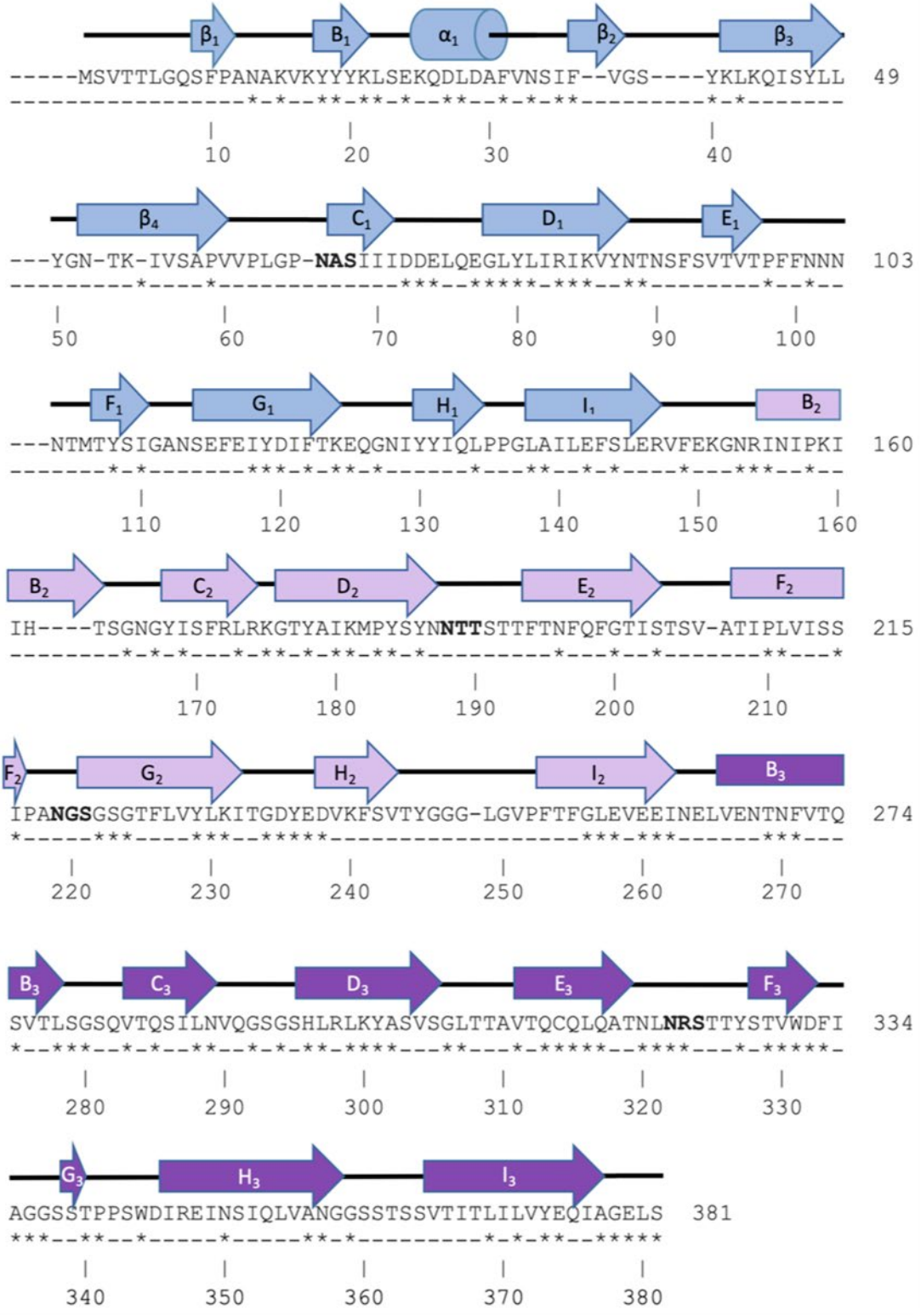


Supplemental Figure 2. Structural Similarity to First, Second and Third Domains of C381. (Holm and Rosenstrom, 2010). Domain 1 (light blue/orange/red) shows greatest similarity to the carbohydrate binding module (CBM) of endo-1,4-beta-xylanase (bottom left – green/orange/red, PDBID 2WYS, Z=9.2) and the ligand binding domains of other hydrolases. Domain 2 exhibits greatest similarity to the knob domain of the tail needle tip in podovirus HS1 (top left, blue/orange/red, PDB ID 4K6B, Z=8.5). Domain 3 is most similar to the protrusion or "P" domain of the infectious bursal disease virus (IBDV) VP2 subviral particle (right, green/orange/red, PDB-ID 2DF7, Z=8.2). For the P-domain of IPDV and the xylanase carbohydrate binding module, other domains in the structure have been omitted for the sake of clarity. However, for the HS1 structure, the knob domain (top domain, predominantly green) is shown relative to the second domain, as the relative domain organization is similar to that in C381. However, the lower domain in this structure, while a beta-sandwich, does not adopt a jellyroll fold.

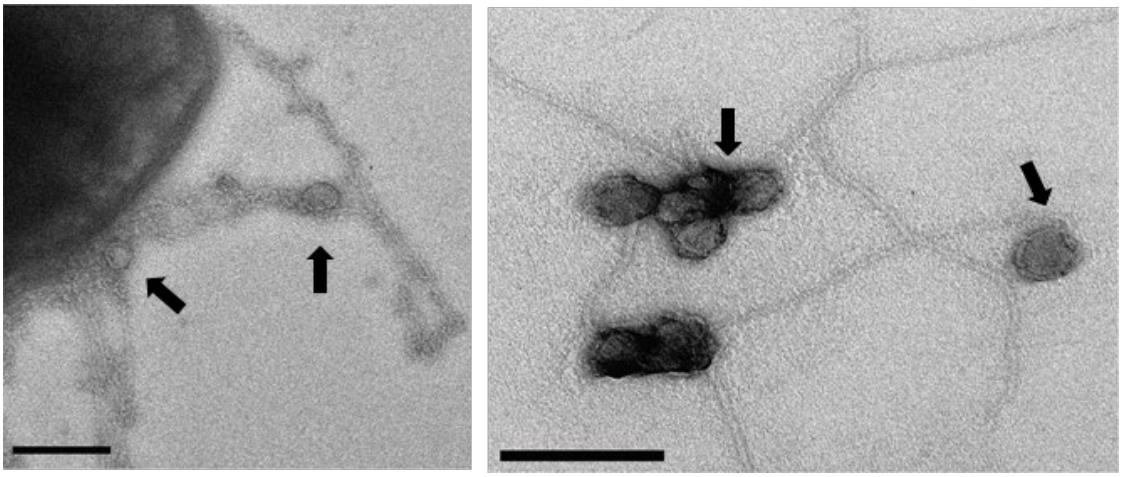


Supplemental Figure 3. C381 Phylogenetic Tree. Sequences related to C381 were extracted from metagenomic datasets across multiple sampling time points. Near full-length sequences were used to generate this maximum likelihood (ML) tree with the percentage of trees in which the associated taxa clustered together shown next to each branch. The phylogenetic tree reveals three C381 clades. Clade 1 represented by STIV-1, Clade 2 by STIV-2, and an uncharacterized Clade 3. As sequence conservation between clades was minimal, only sequences grouping within the STIV-1 clade (I) were used to identify strictly conserved surface residues (Supplemental Figure 4). Accession number for the respective sequences are as follows: YNP NL1 A; MN091910, YNP NL1 B, MN091911; YNP NL13 A, MN091917; YNP NL1 C, MN091912; YNP NL1 D, MN091913; YNP NL1 E, MN091914; YNP NL17 A, MN091918- MN091919; YNP STIV-1 RC, NC_005892; YNP NG5, CP020360.1; YNP Y14, CP013695.1; Iceland STIV-2, ; YNP NL1 F, MN091915; YNP NL1 G, MN091916; YNP NL17 B, MN091920- MN091922; YNP CHAS A, MN091908; YNP NL18 A, MN091923; YNP CHAS B, MN091909

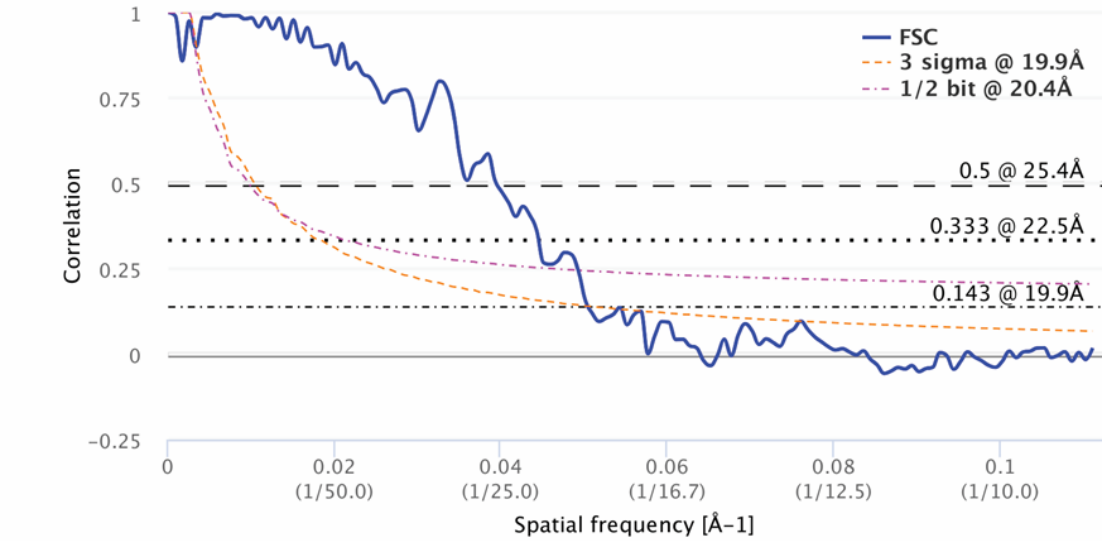
C381 Secondary Structure Elements and Strictly Conserved Residues



Supplemental Figure 4. Secondary Structural Elements and Strictly Conserved Residues. Secondary structural elements are indicated above the sequence, and conserved residues below the sequence by an asterisk (*). Potential N-linked glycosylation sites (NRS/T) are in bold.



Supplemental Figure 5. Initial Attachment of STIV to Host Pili by Negative Stain TEM. Panels A and B - Following incubation with host cells and removal of unbound virus, STIV (black arrows) was observed bound to pili emanating from the cells using negative stain microscopy. Note: Scale bars are 200 μM in length.

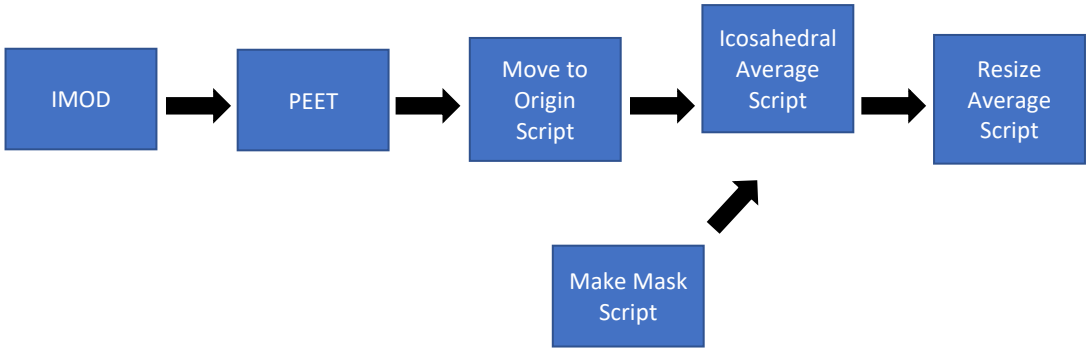


Supplemental Figure 6. Fourier shell correlation between the subtomographic average of 6 pilus bound STIV particles (averaged over their icosahedral symmetry) and the 3.9 Å single particle structure of STIV determined by Veesler et al., 2013. The correlation falls below 0.143 at ~2 nm resolution. The relatively high resolution correlation is likely a factor of: i) the quality of the images from the Titan Krios microscope equipped with Gatan K3 camera and Volta phase plate used with minimal defocus, ii) highly accurate determination of virion orientation using the high resolution structure from the single particle analysis to prime the search, and iii) imposition of perfect icosahedral symmetry when averaging the STIV-1 particles (it is probable that this symmetry is broken at the pilus bound turret, especially for the interacting subunit. Indeed, density suggesting this asymmetry is apparent in panel B of supplemental Figure 7, where density unexplained by the pilus interaction is seen adjacent to domains 1 and 2.

APPENDIX B

CUSTOM SCRIPTS FOR SUB-TOMOGRAM AVERAGING

The flow chart below illustrates the sequence of data processing from tomogram generation to final sub-tomogram average.



Make Mask File

```
#!/bin/csh
source /home/user/Programs/CCP4/ccp4-7.0/bin/ccp4.setup-csh

date > make_mask.out
cat make_mask.com >> make_mask.out

setenv MAPSIZE 100000000
setenv MASKSIZE 100000000

#  To find the grid, extent, origin etc of your current map, can run mapman (lx_mapman) and read in the map; grid, extent, origin, etc of mask should match values of map being masked

/home/user/Programs/Uppsala_Software/rave_linux/lx_mama <<EOF>>
make_mask_trans_all.out
new grid 376 376 376
new cell 1692 1692 1692 90 90 90
new origin 0 0 0
new extent 376 376 376
new spacing 4.5
new ball mask_1 0 0 0 500
#500 will be radius of spherical mask in pixels
write mask_1 spherical_500.msk
quit
EOF
```

Move Aligned Sub-Tomogram to Origin

```
#!/bin/csh
source /home/user/Programs/CCP4/ccp4-7.0/bin/ccp4.setup-csh
setenv MAPSIZE 100000000


date > gt_translate.out
cat gt_translate.com >> gt_translate.out

/home/user/Programs/Uppsala_Software/rave_linux/lx_mapman << EOF >>
gt_translate.out
read map1 aligned_tom021_P0001.mrc ccp4
multiply map1 -1
gt map1
-188
-188
-188
write map1 aligned_mpmntrns_tom021_P0001.mrc ccp4
quit
EOF
```

Icosahedral Average Aligned Sub-Tomogram

```
#!/bin/csh
source /home/user/Programs/CCP4/ccp4-7.0/bin/ccp4.setup-csh

date > avg_icos_trans_all.out
cat avg_icos_trans_all.com >> avg_icos_trans_all.out

setenv MAPSIZE 100000000
setenv MASKSIZE 100000000

rm -rf avg_3x_combined_masked.ccp4
rm -rf avg_3x5x_combined_masked.ccp4
rm -rf avg_3x5x2x_combined_masked.ccp4
rm -rf avg_3x5x2x2x_combined_masked.ccp4

/home/user/Programs/Uppsala_Software/rave_linux/lx_ave <<EOF>>
avg_icos_trans_all.out
Average
aligned_mpmntrns_tom021_P0001.mrc
spherical_500.msk
```

```
avg_3x_combined_masked.ccp4
/home/user/Programs/Uppsala_Software/symm/p1.sym
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/1x_1_a.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/3x_1_a_trans.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/3x_1_b_trans.o

EOF

/home/user/Programs/Uppsala_Software/rave_linux/lx_ave <<EOF>>
avg_icos_trans_all.out
Average
avg_3x_combined_masked.ccp4
spherical_550_trans_all.msk
avg_3x5x_combined_masked.ccp4
/home/user/Programs/Uppsala_Software/symm/p1.sym
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/1x_1_a.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/5x_1_a_trans.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/5x_1_b_trans.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/5x_1_c_trans.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/5x_1_d_trans.o

EOF

/home/user/Programs/Uppsala_Software/rave_linux/lx_ave <<EOF>>
avg_icos_trans_all.out
Average
avg_3x5x_combined_masked.ccp4
spherical_550_trans_all.msk
avg_3x5x2x_combined_masked.ccp4
/home/user/Programs/Uppsala_Software/symm/p1.sym
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/1x_1_a.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/2x_1_a.o

EOF

/home/user/Programs/Uppsala_Software/rave_linux/lx_ave <<EOF>>
avg_icos_trans_all.out
Average
avg_3x5x2x_combined_masked.ccp4
spherical_550_trans_all.msk
avg_3x5x2x2x_combined_masked.ccp4
/home/user/Programs/Uppsala_Software/symm/p1.sym
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/1x_1_a.o
/mnt/data0/lawrence/MSUV/Lawrence/STIV_PDBs/Matrices_2235/2x_2_a.o
```

```
EOF

cp avg_3x5x2x2x_combined_masked.ccp4 icos_tom021_P0001.mrc
```

Resize Averaged Volume to Fit Single-Particle Reconstruction

```
#!/bin/csh

date > change_pixel_size.out
cat change_pixel_size.com >> change_pixel_size.out

# This will change the pixel size to 4.31; generally want to explore multiple resize values in steps of 0.1 or 0.01

cp icos_tom021_P0001.mrc icos_tom021_P0001_4_31.mrc
alterheader  icos_tom021_P0001_4_31.mrc << EOF >> change_pixel_size.out
del
4.31 4.31 4.31
done
EOF

# Then in Chimera resample the resized volume onto the original to obtain a smaller averaged virus with the same grid spacing.
# vop resample #0 onGrid #1
# icos_tom021_P0001_4_31.mrc is # 0
# icos_tom021_P0001.mrc  is #1
```

APPENDIX C

CUSTOM SCRIPTS FOR MOVIE GENERATION IN CHIMERA

Following scripts will re-create Supplemental Movie 1 from Chapter 2. These 4 scripts should be executed sequentially.

Movie Part 1 Open Volumes and Setup

```
close #0-10; wait 1
reset

open #0 virus1_3_12_18allvirus031cloned_4_21_bounded.mrc
open #1 virus2_3_12_18allvirus031cloned_4_21_bounded.mrc
open #2 virus3_3_12_18allvirus031cloned_4_21_bounded.mrc

volume #0 level 61 style surface step 1 color #ecd292
#showPlane false
volume #0 expandSinglePlane true
volume #1 level 61 style surface step 1 color #ecd292
volume #2 level 61 style surface step 1 color #ecd292

open #3 entire_stiv_relative_virus1.pdb; wait 1

color #ecd292 #3; wait 1
color #6e016b #3:264-400.P
color #8c6bb1 #3:150-263.P
color #9ebcda #3:1-149.P

scolor #0 zone #3:.P range 9

close #3; wait 1
open #4 entire_stiv_relative_virus2.pdb; wait 1

color #ecd292 #4
color #6e016b #4:264-400.P
color #8c6bb1 #4:150-263.P
color #9ebcda #4:1-149.P

scolor #1 zone #4:.P range 9

close #4; wait 1
open #5 entire_stiv_relative_virus3.pdb; wait 1
color #ecd292 #5
color #6e016b #5:264-400.P
color #8c6bb1 #5:150-263.P
color #9ebcda #5:1-149.P
```

```
scolor #2 zone #5:.P range 9; wait 1

close #5; wait 1

open #6 6_22_18clonedfibervol_even_fiber25_4_21_tight_bounded.mrc

volume #6 expandSinglePlane true; wait 1
volume #6 level 3.35 style surface step 1 color #000069c9e8b9

background solid white

#change 4 to 5 ect in final


open #7 tomo_03_fxali_resized_for_movie_tight_bounded.mrc; wait 1

open #8 6_22_18clonedfibervol_single_fiber.mrc
volume #8 expandSinglePlane true; wait 1
volume #8 level 3.35 style surface step 1 color #000069c9e8b9

open #9 entire_stiv_relative_virus2.pdb; wait 1

#color only ribbon:  color #6e016b,r #9:263-387.P
#color turret domains
color #6e016b #9:263-387.P
color #8c6bb1 #9:151-262.P
color #9ebcda #9:1-150.P
#color capsid
color #ecd292,r #9:.A-O
#color a223
color #a67573 #9:.Q
#color A55
color #175487,r #9:.R

#colors remain even after atomic model closes so now close to save memory

~ribbon #9; wait 1

#resize/align tomogram to fit viruses not needed for this corrected volume
#volume #7 voxelSize 4.21
#move y -110 model #5

#volume #0,1,2,4,5 planes z,205,210
```

volume #7 style solid step 1
volume #7 colorMode rgb8
#make solid planes for #7 non-tranparent, so you dont see the virus behind
volume #7 level -2500,0 color #79e779e779e7 level 1177,1 color white level 5000,1 color white

volume #7 planes z,1

Movie Part 2 Save Positions

vol #0-7 show; wait 1

reset

vol #6 region 0,0,0,850,850,400

turn x 180
turn y 180
savepos moviestart

reset moviestart
turn z -18
turn x 30
savepos figurepos

reset figurepos
scale 4
move x 200
savepos zoomedfigure

vol #0,2,7,8 hide
vol #6 region 460,0,0,520,500,400; wait 1

reset zoomedfigure
turn y 40
turn z 20
move y 120
scale 2
turn y 20
move x 60
turn z 4
turn x -14

```
move y -40
turn y 20
turn z 3

savepos zoomed_single_virus

reset zoomed_single_virus
move y 100
move x 100
scale 1.2
scale 1.2
scale 1.2
turn z -1
savepos zoomed_3

#reset moviestart

vol #6 region 460,0,0,520,500,400

reset figurepos
ribbon #9
savepos atomic_fly_in2
move x 2800 model #9
savepos atomic_fly_in1
~ribbon #9


reset moviestart
reset moviestart
transparency 0 #1
vol #6 region 0,0,0,850,850,400
vol #0-8 show
```

Movie Part 3 Run and Save Movie

```
#movie reset
reset moviestart
transparency 0 #0,1,2,6
#vol #6 region 0,0,0,850,850,400
vol #0-6 hide; wait 1
#need to scale to avoid white border
scale 1.3
```

```
volume #0-2,3 hide
volume #6,8 hide
~ribbon #9; wait 1

set  maxframerate 25

volume #7 planes z,273,1,1
movie record
wait 276

volume #7 planes z,1,350,1
wait 353

volume #0-2 show
volume #6,8 show; wait 1

volume #7 planes z,350,1,1
wait 360

volume #7 hide
turn y 1 360
#turn y 6 60 center #1
#turn y 6 60 center view
wait 360
fly  20 start figurepos
wait 20

fly 50 start zoomedfigure
wait 100
fly 50 zoomedfigure figurepos
wait 100

fly atomic_fly_in1
ribbon #9; wait 1
move x -20 140 model #9
wait 180

transparency 50 #0,2,6 frame 70
wait 70
transparency 100 #0,2,6 frame 70
wait 70

volume #0,2,6 hide; wait 1
```

```
fly 100 figurepos zoomed_3
wait 120

transparency 100 #1 frame 70
wait 100

wait 10

#volume #0,2 hide

#vol #6 region 460,0,0,520,500,400

freeze
movie stop
movie encode framerate 25 bitrate 60000  output
/home/k26s361@msu.montana.edu/C381/C381_Figures/mymovie.mov
#  Go back to faster Chimera display frame rate.
#
#set maxframerate 60
```