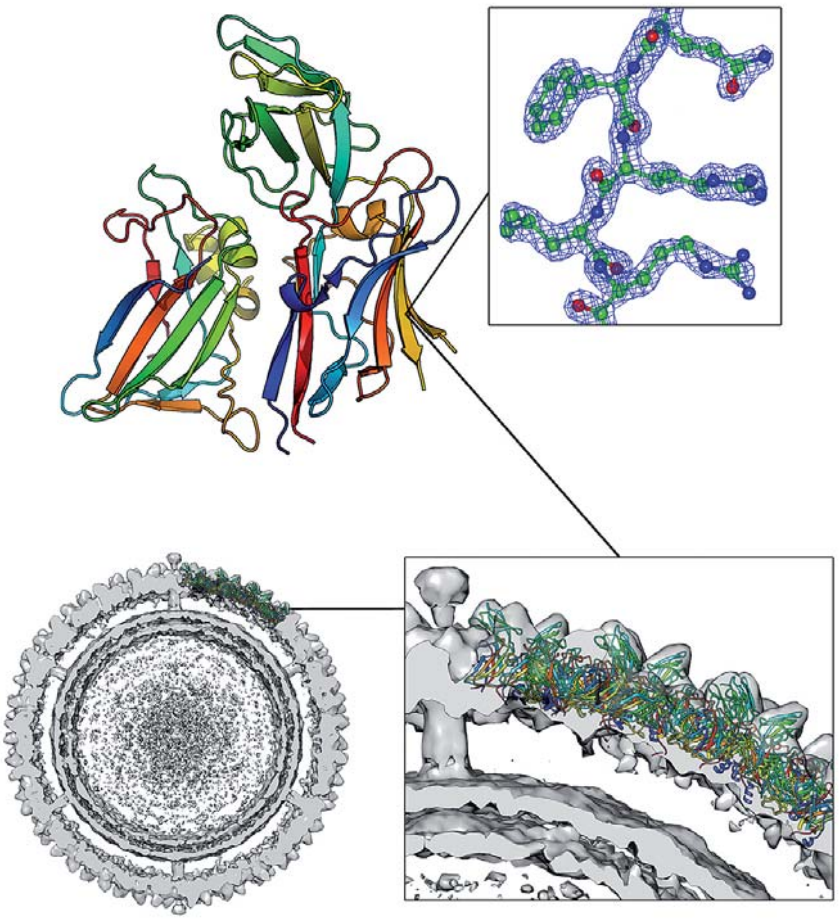
JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE
276

Ilona Rissanen

An Ancient Virus Type from Extreme Environments



JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 276

Ilona Rissanen

An Ancient Virus Type from Extreme Environments

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ylistönrinteellä, salissa YAA303
maaliskuun 28. päivänä 2014 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in Ylistönrinne, hall YAA303, on March 28, 2014 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

An Ancient Virus Type from Extreme Environments

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 276

Ilona Rissanen

An Ancient Virus Type from Extreme Environments

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

Editors
Varpu Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

Jyväskylä Studies in Biological and Environmental Science

URN:ISBN:978-951-39-5642-4
ISBN 978-951-39-5642-4 (PDF)

ISBN 978-951-39-5641-7 (nid.)
ISSN 1456-9701

Jyväskylä University Printing House, Jyväskylä 2014

ABSTRACT

Rissanen, Ilona
An ancient virus type from extreme environments
Jyväskylä: University of Jyväskylä, 2014, 60 p.
(Jyväskylä Studies in Biological and Environmental Science
ISSN 1456-9701; 276)
ISBN 978-951-39-5641-7 (nid.)
ISBN 978-951-39-5642-4 (PDF)
Yhteenveto: Muinainen virustyyppi ääriolosuhteista
Diss.

Rapid pace of genetic differentiation in viruses makes it difficult to trace ancient evolutionary relationships using prevalent sequence comparison methods. Recently, the increasing availability of structural data has led to a new way of identifying viral lineages: "viral self" elements, including certain conserved protein structures, allow the detection of deep evolutionary relationships between viral species. The most extensively characterized structural virus lineage, namely the double beta-barrel lineage, includes viruses that infect hosts from all three domains of life. Recent discovery of thermophilic bacteriophage P23-77 indicates that the lineage may be even more diverse than expected, containing an ancient branch of viruses that exhibit a single beta-barrel major capsid protein homologous to the fold observed in the double beta-barrel lineage. This thesis reports the crystal structures of the major capsid proteins of bacteriophage P23-77, which identify the virus as an archetype of an ancient virus lineage sharing common ancestry with the double beta-barrel lineage. Unique features of P23-77 major capsid proteins and capsid organization epitomize an archaic virus type that is currently found only in extreme environments. Further characterization revealed that the major capsid proteins are highly thermostable and interact poorly in solution, which suggests that the capsid assembly requires a trigger, possibly functioning in a fashion similar to minor capsid proteins in distantly related viruses STIV and PM2. Despite discoveries reporting structural conservation in viruses, current virus taxonomy is primarily based on genetic similarity. This thesis establishes a level where both structural and genetic evidence can be utilized to associate P23-77 with other putative single beta-barrel lineage members. P23-77 is proposed to belong to prospective viral family *Sphaerolipoviridae* and define a novel genus titled *Gammasphaerolipovirus*.

Keywords: Viral lineage; bacteriophage P23-77; virus capsid proteins; virus capsid assembly; virus evolution.

Ilona Rissanen, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

Author's address Ilona Rissanen
Department of Biological and Environmental Science
P.O. Box 35 (Survontie 9)
40014 University of Jyväskylä, Jyväskylä
Finland
ilona.a.rissanen@jyu.fi

Supervisors Professor Jaana Bamford
Department of Biological and Environmental Science
P.O. Box 35 (Survontie 9)
40014 University of Jyväskylä, Jyväskylä
Finland

Reviewers Reader Roman Tuma
Institute of Molecular and Cellular Biology
Faculty of Biological Sciences
University of Leeds, Leeds, LS2 9JT
United Kingdom

Docent Tomi Airenne
Department of Biosciences, Biochemistry
Åbo Akademi University
Tykistökatu 6 A, FI-20520 Turku
Finland

Opponent Research Professor Nicola Abrescia
Structural Biology Unit
CIC bioGUNE, CIBERehd
Bizkaia Technology Park, Bld 800
48160 Derio
Spain

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original articles, which will be referred to in the text by their Roman numerals I-IV.

My responsibilities and the contributions of the other authors in each of the articles are described in the table at the bottom of the page.

I Rissanen I., Pawlowski A., Harlos K., Grimes J.M., Stuart D.I. & Bamford J.K.H. 2012. Crystallization and preliminary crystallographic analysis of the major capsid proteins VP16 and VP17 of bacteriophage P23-77. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* 68: 580–583.

II Rissanen I., Grimes J.M., Pawlowski A., Mäntynen S., Harlos K., Bamford J.K.H & Stuart D.I. 2013. Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch from a major virus lineage. *Structure* 21: 718–726.

III Rissanen I., Määttä J.A., Hytönen V.P. & Bamford J.K.H. Capsid protein properties and the regulation of capsid assembly of thermophilic bacteriophage P23-77. Manuscript.

IV Pawlowski A., Rissanen I., Bamford J.K.H, Krupovic M. & Jalasvuori M. 2014. *Gammasphaerolipovirus*, a newly proposed bacteriophage genus, unifies viruses of halophilic archaea and thermophilic bacteria within the novel family *Sphaerolipoviridae*. *Archives of Virology*. In press.

Author contributions. Initials stand for the following authors: **IR,** Ilona Rissanen; AP, Alice Pawlowski; SM, Sari Mäntynen; JM, Juha Määttä; VH, Vesa Hytönen; MJ, Matti Jalasvuori; MK, Mart Krupovic; KH, Karl Harlos; JG, Jonathan Grimes; DS, David Stuart; JB, Jaana Bamford.

	I	**II**	**III**	**IV**
Original idea	JB, DS	JB, DS	**IR**, JB, VH	MJ, MK, AP
Molecular cloning	AP	-	**IR**	-
Protein production	**IR**, AP	**IR**	**IR**	-
Protein characterization	-	AP, **IR**, SM	**IR**, VH, JM	-
Protein crystallization	**IR**, AP, KH	**IR**, KH	-	-
Crystallography	KH	KH, **IR**, JG	-	-
Computational methods	-	**IR**, JG, DS	**IR**	MJ, AP
Writing	**IR**, AP, DS, KH	**IR**, DS, JG, KH, AP	**IR**, VH, JM	AP, MJ, MK, **IR**
Molecular graphics	**IR**, JG, KH	**IR**, DS, JG	**IR**	**IR**, AP, MJ

ABBREVIATIONS

ATPase	adenosinetriphosphatase
BTV	bluetongue virus
bp	base pair
cryo-EM	cryo-electron microscopy
ds	double-stranded
EMDB	The Electron Microscopy Data Bank
HK97	Honk Kong 97 virus
ICTV	International Committee on Taxonomy of Viruses
id	identifier
λ (lambda)	wavelength
LUCA	last universal common ancestor
ORF	open reading frame
PDB	Protein Data Bank
PEG	polyethylene glycol
SHP	Structure Homology Program
ss	single-stranded
STIV	*Sulfolobus* turreted icosahedral virus
T_m	melting temperature
TMV	tobacco mosaic virus
T, T-number	triangulation number
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
VEDA	Visual Environment for Docking Algorithms

1 INTRODUCTION

It was not until the middle 19th century when the theory of spontaneous generation – the belief that living organisms could arise from non-living matter – was ultimately debunked by Louis Pasteur's findings that revealed the role of microbial organisms in biological processes. Nascent branch of science, microbiology, soon established bacteria as the causative agent in many diseases. While this revolutionized medical practice of the time, focus on bacteriology also deterred researchers from recognizing the existence of viruses for decades (Artenstein 2012). At the turn of the 19th century, extensive research culminated in the conception of viruses as a novel class of infectious agents. Discovery of viruses led to significant developments in the medical field, and for a long time, most of the research on viruses focused on their medical aspects. However, during the last decades it has become evident that viruses play a much larger role in development of life and biosphere than previously understood (Dolja and Krupovic 2013, Koonin and Dolja 2013).

Early paradigms define viruses as sub-microscopic filterable infectious agents, consisting of genomic material encapsulated in a protective protein shell. Novel findings such as giant viruses the size of smallest cells (Claverie et al. 2006), polydnaviruses with symbiotic lifecycle (Strand and Burke 2012), and viral plasmoids without a protein shell (Solorzano et al. 2000), challenge these notions and call attention to the changing concept of the viral universe. Viruses have significant impact on life on Earth: high prevalence of viral fragments in the genomes of cellular organisms indicates an interchange between viruses and cells that has shaped life since primordial times (Koonin et al. 2006, Krupovic and Bamford 2008). Recent evidence suggests that viruses may even influence global phenomena including geochemical cycles and climate change (Suttle 2007).

The origin and evolutionary history of viruses, as well as many of their functions and the extent of their influence on the biosphere remain unresolved. The advance of structural biology has engendered new perspectives in virology and the study of viral evolution. One of the major discoveries was that folds of certain viral proteins are highly conserved and consecutively, allow

identification of deep evolutionary relationships (Benson et al. 1999, Abrescia et al. 2012). Virus lineages based on conserved structural features could provide means to establish new higher-level taxa, thus bringing order to the viral universe.

This thesis discusses the concept of viral lineages and presents structural studies of thermophilic bacteriophage P23-77, discovered during the course of the study to represent an ancient branch from a major virus lineage. The results describe five novel crystal structures of the major capsid proteins of P23-77, which were used to infer the evolutionary history of P23-77 and construct a model of the P23-77 capsid. Capsid assembly process and the implications of P23-77 phylogeny on current virus taxonomy are also major points of interest in this study. Unique features of P23-77 are postulated to represent a previously unknown, ancient viral type that has been retained under negative selection caused by the extreme environment inhabited by the host bacterium *Thermus thermophilus*. Prospective group of viruses belonging to the ancient type are found in high temperature or salinity environments, and may form one of the oldest viral groups in existence today. Next chapters will provide the reader with an overview of the relevant literature covering viruses, their impact on life as we know it, dilemmas of viral classification and viral lineages as a prospective solution.

2 REVIEW OF THE LITERATURE

2.1 Virology

Viruses are generally defined as small obligate parasites of cellular organisms, consisting of a nucleic acid genome packaged into a protein shell. Viruses are unable to reproduce on their own; instead they infect cells and exploit the cellular machinery of the host to propagate. Despite their relative simplicity compared to cellular organisms, the variety that viruses exhibit in their genetic and structural components is so extensive that viruses may represent the greatest source of diversity in the natural gene pool (Hendrix et al. 1999, Pedulla et al. 2003). Many viral genes have no known homologues in cellular organisms and the origin of viruses is still debated (Forterre and Prangishvili 2009, Holmes 2011). Closer look to what constitutes a virus reveals the range of viral diversity.

2.1.1 Constituents of a virus

In sharp contrast to cellular organisms, which always have double-stranded (ds) DNA genomes, viruses utilize a wide array of different genetic systems. Viral genome may consist of DNA or RNA, be double- or single-stranded (ss), and exhibit circular, linear or segmented configuration. Complexity of virus genomes varies greatly, from simple viruses such as tobacco mosaic virus (TMV) that has 6400 base ssRNA genome encoding four proteins (Hirth and Richards 1981), to the giant *Pandoravirus salinus* with DNA genome up to 2.5 million base pairs (bp) long containing 2556 putative protein-coding sequences (Philippe et al. 2013). Viral genomes are subject to rapid differentiation caused by the short generation time, high progeny number and, especially in case of RNA viruses, high mutation rates (Drake et al. 1998).

The viral genome requires a carrier to be transported between hosts, and in most cases this is provided by an intricately arranged proteinaceous shell that may contain internal or external lipid membrane layers (Rossmann and Rao 2012). Membrane component found in many viruses is host-derived,

though the virus may incorporate lipids selectively during virion assembly. In addition to the structural proteins and a potential membrane constituent, virions typically encompass replication, regulatory and accessory proteins. Virus shell, *i.e.* the capsid, bears host-recognition structures that the virus utilizes to attach to and enter the host. Capsid architecture follows either icosahedral or helical symmetry, or in absence of both is classified as complex (Fig. 1). These architectures produce a multitude of shapes, including rods, flexible filaments, polyhedrons, ellipsoids, spheres and pleomorphs.

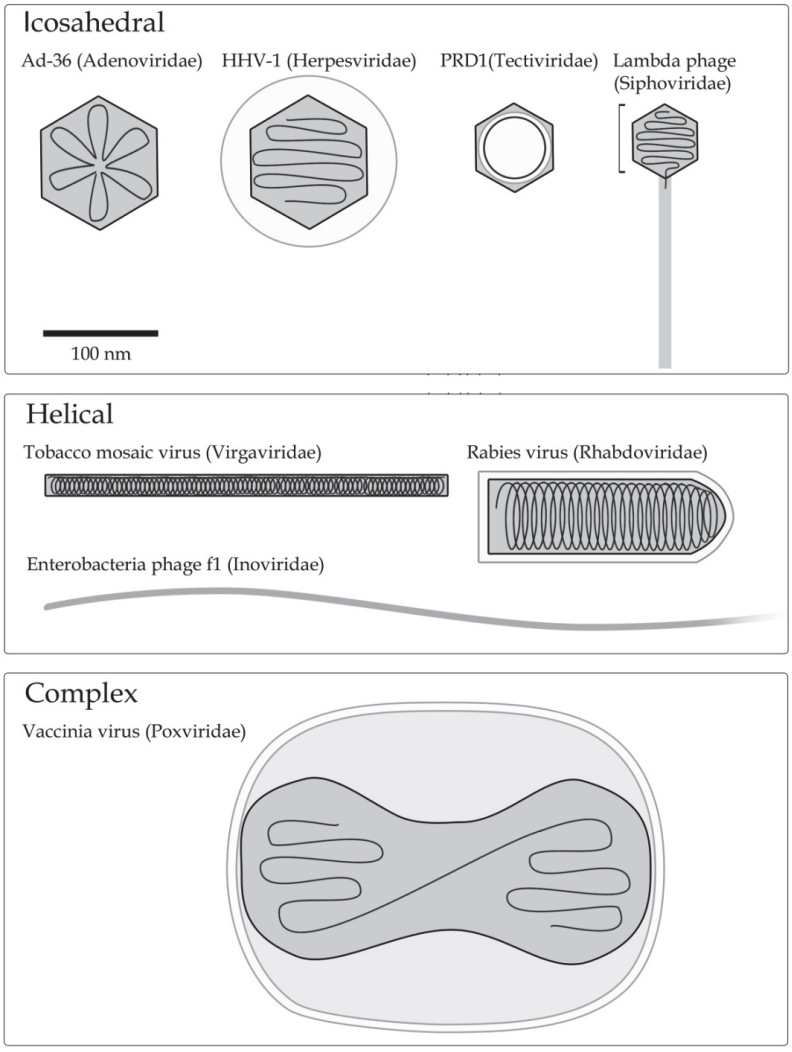


FIGURE 1 Three types of capsid architecture: icosahedral, helical and complex. Virion consists of a protein shell enclosing a genome (including Ad-36 and Tobacco mosaic virus) and may contain membrane components either internal (PRD1) or external to the capsid shell (*e.g.* HHV-1). Ad-36: human adenovirus 36, HHV-1: human herpesvirus 1. Illustrations are based on models by ViralZone, Swiss Institute of Bioinformatics (http://viralzone.expasy.org/).

Icosahedral architecture is found in various enveloped and non-enveloped viruses that prevalently feature DNA genomes (Rossmann and Rao 2012). In addition to clearly polyhedral viruses, most capsids that appear spherical follow icosahedral symmetry. Helical viruses commonly have RNA genomes and their number includes tobacco mosaic virus, rabies virus and certain types of influenza viruses. Complex viruses, exemplified by poxviruses, are the third and rarest class of capsid architecture. Mature poxviruses apparently lack the symmetry features common to other viruses, and are instead composed of large 'brick-shaped' virion with a complicated internal structure including a walled core encased in lipid layers (Condit et al. 2006). Particle morphology may vary greatly during virion maturation. For example, poxviruses go through a spherical maturation step involving the formation of an icosahedrally ordered protein lattice on lipid vesicle surface (Heuser 2005, Szajner et al. 2005). Structure of a capsid is a compromise between stability and metastability: the virion needs to be stable enough to withstand extracellular conditions, yet allow disassembly upon infection (Rossmann and Rao 2012).

2.1.2 Icosahedral capsid architecture

Icosahedral capsid architecture is prevalent in viruses discussed in this thesis. Surface architecture of icosahedral capsids is often described in terms of triangulation number (T), derived from the theory of quasi-symmetry (Caspar and Klug 1962). The T-number describes the number of different environments occupied by each subunit, arising from the relationship between hexagonal and pentagonal units in the capsid. T is defined by $T=h^2+hk+k^2$, where h and k describe the distance between successive pentagons on each edge (Fig. 2).

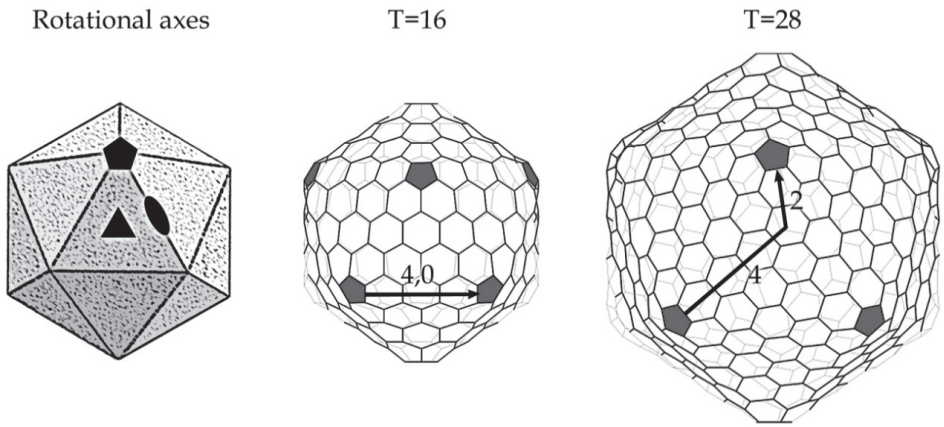


FIGURE 2 Icosahedral capsid architecture. Rotational axes on an icosahedron on the left, followed by illustrations of T=16 (middle) and T=28 (right) architecture. 5-fold (pentagon), 3-fold (triangle) and 2-fold (ellipse) rotational axes illustrated. Capsid architecture illustrations include h and k values shown on the capsid surface. Lattices were created with Icosahedron Generator (http://johnsonlab.scripps.edu/research/icosahedron-generator.php).

Simplest T=1 lattice contains 60 identical subunits organized in 12 pentameric capsomers (Caspar and Klug 1962, Johnson and Speir 1997). When T>1, polyhedral architecture requires the presence of both hexagonal and pentameric capsomers, resulting in quasi-equivalent subunit environment. This is achieved if chemically identical subunits change their oligomerization mode or if more than one type of polypeptide building block is incorporated into the capsid.

Common terminology, when discussing icosahedral capsid organization, includes 2-fold, 3-fold, and 5-fold axes of rotational symmetry (Fig. 2). Accordingly, subunits comprising each axis are titled disymmetron, trisymmetron and pentasymmetron. The term asymmetric unit is used for the smallest repeating unit in a symmetrical whole; in this thesis either in the context of a crystal lattice or an icosahedral capsid model. Asymmetric unit of the capsid model represents the smallest biological ensemble that can be used to generate the icosahedral capsid shell via symmetry operations, in other words, 60 copies of the asymmetric unit comprise the capsid.

2.1.3 Capsid assembly

During virus assembly, structural proteins, additional components and the genome oligomerize in an ordered fashion to form virions (Mateu 2013). Assembly is initiated by a nucleation event where viral components begin to interact, inducing conformational arrangement of the building blocks. Resulting newly formed surfaces trigger subsequent assembly steps (Tuma et al. 2001). Self-assembly proceeds through a series of usually transient assembly intermediates, often directed by interaction with scaffolding proteins or the viral nucleic acid (Rossmann and Rao 2012). In many cases the assembled capsid will undergo further maturation steps involving covalent modification and conformational rearrangement before final virion state is achieved. Assembly mechanisms often fall into one of three categories which are distinguished by genome incorporation method: i) co-assembly of genome and capsid proteins, ii) genome packaging into pre-formed procapsids and iii) capsid assembly around a pre-condensed genome. Principles of the mechanisms are briefly described below.

One of the simplest assembly pathways is found in helical ssRNA TMV (Hirth and Richards 1981). TMV capsids are rigid rods built of single structural protein enclosing the ssRNA genome (Fig. 1, page 12). The TMV assembly begins when a structural protein oligomer binds at a packaging recognition site in the genome. Assembly proceeds by addition of more structural protein oligomers that lock the genome in helical configuration inside the building capsid. The process terminates when the entire genome is encapsulated. The TMV assembly occurs if isolated components are mixed *in vitro*, which signifies that the process is independent of any host or virally encoded accessory factors. Other ssRNA viruses successfully utilize this co-assembly technique to build polyhedral capsids, as is seen in satellite tobacco mosaic virus and bean pod mottle virus (Stockley and Twarock 2010).

By contrast, the more complex assembly of the icosahedral dsDNA bacteriophage PRD1 is considered to involve the host cytoplasmic membrane: major capsid protein P3 expression in the host cell leads to the formation of P3 trimers that along with minor capsid proteins congregate on the membrane (Abrescia et al. 2004, Butcher et al. 2012). Protein P3 trimers interact with the minor capsid proteins, most notably P30, building facets that upon formation extract a lipid vesicle from the host membrane. Interactions between P3, minor capsid proteins and the internal lipid membrane coordinate the assembly. PRD1 forms empty procapsids into which the linear DNA genome is consequently packaged by a specialized packaging motor.

Yet another assembly pathway is found in dsDNA viruses with circular genomes. This method is the least well understood among the three described types. Presumably, the circular genome is pre-condensed and provides a nucleation site for the capsid assembly. This assembly mechanism is found in polyomavirus Simian virus 40 and suggested for bacteriophage PM2 (Gordon-Shaag et al. 2002, Abrescia et al. 2008). Reported features of these assembly pathways include the use of DNA-associating host proteins and calcium ions to promote genome pre-condensation and protein interactions, respectively.

Genome composition and configuration influences the assembly strategies adapted by each viral species (Mateu 2013). Overall, most DNA viruses have double-stranded genomes packaged in icosahedral capsids, while most RNA viruses have helically ordered capsids co-assembled with the genome. Configuration of the genome has an effect: linear DNA genomes are typically packaged in pre-formed procapsids while circular genomes may also be encapsidated via assembly around a pre-condensed genome. Capsid assembly involves numerous interesting research prospects: knowledge of assembly pathways may provide opportunities for targeted drug-design and biotechnology applications, such as engineering nanoscale containers and gene therapy vectors (Douglas and Young 2006).

2.1.4 Exceptions and rules

Viruses exhibit considerable diversity in almost every aspect, which means that generalizations always have exceptions. This chapter briefly illustrates a few examples of the more unusual forms and functions. Firstly, regarding the concept of a virus as genomic material encased in a protein capsid, we must consider that many RNA viruses do not require a capsid and pleolipoviruses utilize a lipoprotein vesicle instead (Dawe and Nuss 2001, Pietila et al. 2012). The traditional definition of viruses as submicroscopic filterable agents is challenged by Mimivirus (Raoult et al. 2004), Megavirus (Arslan et al. 2011) and pandoraviruses (Philippe et al. 2013), which rival smallest cells in size and genetic complexity. In addition, viral functions may deviate from the common principles in both extracellular and intracellular phases. It was recently found that the extracellular phase is not exclusively dormant in crenarchaeal virus ATV which completes its morphological development outside of the host (Haring et al. 2005). Infection does not always include host-recognition as

exemplified by certain plant viruses that can spread from cell to cell via plasmodesmata (Benitez-Alfonso et al. 2010). Even the paradigm of viruses as parasites of cellular organisms has exceptions: life cycle of polydnaviruses suggests symbiotic rather than parasitic relationship (Strand and Burke 2012), while virophages such as Sputnik parasitize other viruses (Desnues et al. 2012).

It could be said that the viral world involves more exceptions that rules. However, certain general principles may be extracted from the sea of diversity. These rules are defined by the boundaries posed by physicochemical properties: for example, genomes larger than ~30 kilobase always consist of dsDNA, most likely to allow sufficient stability and topological malleability (Koonin and Dolja 2013). Genome packaging is a crucial part of the viral viability and the packaging mechanisms seem to follow general patterns: for dsDNA viruses, common method is packaging via a packaging motor into pre-formed procapsids while ssRNA viruses incorporate their genome simultaneously with the capsid assembly (Mateu 2013). One important aspect in bringing order to the viral universe concerns viral taxonomy (Krupovic and Bamford 2010). The current taxonomy of viruses reflects viral genetic diversity: taxa higher than order are not recognized and most viruses are only assigned to a family. Recent decades have seen the advent of a concept that certain viral elements are highly conserved at the structural level and could allow the identification of viral lineages beyond the current genome sequence-based means of classification (Benson et al. 1999, Krupovic and Bamford 2010).

2.2 Virosphere and biosphere

Earth's biosphere consists of all terrestrial cellular life and the habitats supporting that life. Life on Earth is currently classified, based on cellular type, into domains Bacteria, Archaea and Eukarya, all of which are descended from a primordial common ancestor, LUCA (last universal common ancestor) (Woese and Fox 1977, Woese et al. 1990). Molecular clock studies suggest that the time that the ancestors of Bacteria and Archaea diverged from LUCA was about four billion years ago, followed by the divergence of the Eukaryal ancestor from cells related to Archaea about a billion years later (Hedges et al. 2001, Battistuzzi et al. 2004). Features that distinguish the domains include differences in genetic machinery, membrane constitution and the presence of nuclear envelope; features common to all define cells as biological units with dsDNA genomes surrounded by cellular membrane. This three-branched 'tree of life' has given rise to all modern cellular organisms.

Virosphere surrounds the cellular tree of life and viruses have been evolving with their hosts for billions of years in all imaginable environments (Villarreal and Witzany 2010). All cellular life is constantly under pressure from viruses which are estimated to outnumber cells by at least a factor of ten (Bergh et al. 1989, Wommack and Colwell 2000). Viruses infecting domain Eukarya are referred to as eukaryal viruses, while viruses of domains Archaea and Bacteria

are known as archaeal viruses and bacteriophages, or by the common moniker prokaryal viruses. This thesis reports studies of an extremophile bacteriophage, albeit with a significant connection to viruses infecting hosts from other domains of life.

Bacteriophages (phages) are the largest group of viruses described, comprising over 5500 examined species (Ackermann 2007). Phage morphology can be classified into four prevalent types: icosahedrally ordered tailed, icosahedral, filamentous and pleomorphic. Tailed phages of order *Caudovirales* are by far the most abundant, comprising approximated 96% of examined bacterial viruses. Bacteriophages are relatively simple viruses and easy to obtain, making them widely appreciated model systems. Since their discovery in early 20th century, phages have played an important role in scientific research. Studies of bacteriophages have led to a number of significant breakthroughs in the field of molecular biology, including the discovery of the roles of DNA and mRNA (Hershey and Chase 1952, Brenner et al. 1961), restriction enzymes (Smith H.O. and Wilcox 1970) and the use of phages as antimicrobials and tools in protein engineering and diagnostics (Smith G.P. 1985, Smith H.W. et al. 1987, Schofield et al. 2012).

2.2.1 Origin and evolution of viruses

The origin of viruses remains shrouded in mystery, yet it is commonly accepted that their origin is very ancient, presumably dating back to the same period as the origin of cellular life. Most theories of virus origin fall into one of three categories: i) viruses are seen as predecessors of cellular life, ii) viruses are considered to originate from cells, being in essence reduced cells, and iii) viruses are seen as 'break away' elements from the early cells, potentially genetic transport vehicles that at some point became self-replicating parasites (Hendrix et al. 2000, Koonin et al. 2006, Forterre and Prangishvili 2009). However, conclusive evidence to affirm how viruses originated is difficult to come by and each current model struggles to explain certain viral features. The virus-first model seems to contradict with the fact that all current viruses are fundamentally reliant on cellular hosts for reproduction. The reduction model implies that intermediates between cells and viruses should have existed; yet none have been found when the effects of horizontal gene transfer are ruled out. In addition, current evidence that certain genes are conserved and prevalent in viruses yet absent in cellular organisms strongly suggests that viruses are at least as old as cells instead of being their derivatives (Koonin et al. 2006). Likewise, the 'escaped gene' model faces the dilemma that viral hallmark genes, shared by diverse groups of viruses, are unique to viruses and have no parallels in cellular organisms. If viruses are escaped cellular elements, how is it possible that they are built from structures with no known cellular counterparts? In any case, viruses are closely connected with the origins of life and have affected the development of cellular life since the primordial times. Features of viral evolution are further elaborated in the next chapter.

Self-evident evolution proceeds via natural selection that favors certain features, present in the natural diversity of the gene pool and constantly emerging via mutations during genome replication (Drake et al. 1998). Characteristics of virus propagation, namely short generation time, great number of progeny and high mutation frequency, when combined with natural selection, promote rapid adaptation and evolutionary pace. In addition to self-evident evolution, horizontal gene transfer is among most important evolutionary mechanisms (Hendrix et al. 1999). Horizontal gene transfer refers to a situation where a gene is not inherited from the parental generation (*i.e.* vertical gene transfer) but acquired from an external source. During propagation, certain viruses can recombine their genetic material with host DNA or viral DNA present in the host, resulting in new genomic arrangement and novel genetic modules (Juhala et al. 2000, Hendrix 2003, Pedulla et al. 2003). Highest variability in the mosaic genomes of the T4-type phage family was found in regions that are suggested to facilitate environmental adaptation, implying that mosaicism conveys an evolutionary edge against stressors (Desplats and Krisch 2003). Bacteriophages are most likely to engage in recombination when they share the same host range, in other words, phylogenetically distant hosts harbor increasingly dissimilar phage populations. However, it is possible for a genetic element to traverse multiple organisms, thus enabling horizontal gene transfer even between distant species (Dunning Hotopp et al. 2007, Redrejo-Rodriguez et al. 2012). Viral evolution is most commonly considered in the context of epidemiology and viral diseases. As important as that is, it should be remembered that viral evolution is also a fundamental process closely linked to the origin and development of life.

2.2.2 The impact of viruses on cellular life

Most obvious effect of viruses on cellular life is the stress that they cause to their host organisms, influencing survival and potentially inducing selection pressure towards certain traits. Selection pressure directly caused by viruses has been extensively studied with bacteria and bacteriophages, revealing that phages have major effect on the evolution of their host. Emergence of novel traits in bacterial population is affected by the impact it has on infection susceptibility, and phage-induced pressure may accelerate host diversification and life strategy (Laanto et al. 2012, Hall et al. 2013, Williams 2013). Experimental evolution studies are easiest to conduct in closed systems using microbial organisms, but the same principles apply to all virus-host interactions. The arms race between viruses and hosts has shaped the evolution of all life forms.

Viral community, especially the bacteriophage gene pool, has been suggested to constitute the largest source of natural gene diversity. Molecular interactions between viruses and their hosts generate new genes, possibly constituting the principal source of novel genes in the biosphere (Hendrix 2002, Daubin and Ochman 2004). In addition, many viruses are able to integrate into the host genome, becoming elements that the host treats as parts of its own

genome. This mechanism makes viruses important agents of horizontal gene transfer (Canchaya et al. 2003, Redrejo-Rodriguez et al. 2012). Viruses can transfer a variety of elements in single evolutionary event and may mediate horizontal gene transfer between phylogenetically distant species. The impact of virus-mediated horizontal gene transfer is evident in the genomes of cellular organisms which are littered with inserts of viral DNA. In the human genome, the amount of DNA of viral origin is roughly equal to the entire coding region (Feschotte 2010). However, viruses are not the only benefactors from the arrangement: hosts have also been demonstrated to recruit viral genes to promote their own functions, such as horizontal transfer of their own genes (Lang et al. 2012).

Previous examples of the impact of viruses on cellular life were, though major, still considered at the level of populations. Recent studies indicate that viruses influence the entire planetary ecosystem (Suttle 2007, Rohwer and Thurber 2009). Abundance of viruses in marine environments, at least one to two orders of magnitude larger than the number of cells, indicates the fundamental position viruses occupy. Viruses were recently shown to have major influence on the composition of marine biota and, by extension, the formation of marine sediments. Marine viruses directly and indirectly influence biogeochemical cycles, carbon sequestration capacity of the oceans and the gas exchange between the ocean surface and the atmosphere (Suttle 2007). It follows that viruses are associated with the planetary ecosystem and global phenomena, including climate change.

2.2.3 Classification of viruses

Over the years, many systems have been created for virus classification and taxonomy. Earliest attempts were based on host range, *e.g.* the 1948 Holmes classification describes three classes based on whether the virus infects bacteria, plants or animals (at that time, Archaea were not considered separate from other prokaryotes). Two most influential current classification schemes are the Baltimore system (Baltimore 1971b, a) and classification by the International Committee on Taxonomy of Viruses (ICTV, http://www.ictvonline.org/). Former is the most commonly used classification method for animal viruses while the latter is an ongoing effort to develop, refine and maintain internationally agreed upon taxonomy for all viral species.

Baltimore classification defines seven major classes of viruses based on distinct nucleic acid composition and replication systems (Baltimore 1971b). Classes include i) dsDNA, ii) ssDNA, iii) dsRNA, iv) ssRNA (+) sense, v) ssRNA (-) sense, vi) reverse transcribing ssRNA (+) sense *e.g.* retroviruses, and vii) reverse transcribing dsDNA utilizing viruses. While the genetic system affects the replication and life strategies applied by each species and may create similarities between members of the same class, Baltimore classification is not based on phylogenetic analysis and as such, does not reflect evolutionary background of the classified viruses.

By contrast, ICTV is an official body of numerous experts in the field of virology, charged with the massive task of developing universal virus taxonomy – an effort continuing since the 1960s (Lwoff et al. 1962, Pringle 1998). ICTV is divided in subcommittees and groups that discuss emerging taxonomic issues in the field, process proposals for new taxonomy and revise the publicly available ICTV taxonomy. The ICTV classification is generally based on the genetic features, replication system and host range of each virus. The latest released ICTV taxonomy from year 2012 defines 7 orders, 96 families, 22 subfamilies, 420 genera and 2618 species. Currently, most families are not assigned to an order and taxa higher than order are not recognized.

Establishing viral taxonomy is, even with ever more powerful analytical methods, a truly demanding task since the rapid pace of genetic differentiation and effects of horizontal gene transfer (mosaicism) typical of viral genomes effectively make it impossible to infer a single phylogeny for viral genes (Abrescia et al. 2012). Two solutions to this have been proposed: evolution of viruses could be represented as a network illustrating the dynamics between viral genomes (Lima-Mendez et al. 2008), or the phylogeny could be derived from a selected set of genes and gene products considered to be most crucial to the virus (the 'viral self') and vertically inherited (Bamford 2003, Krupovic and Bamford 2010). The former suggestion relies on sequence comparisons which typically fail to recognize relationships between viruses diverged further back in time. By contrast, deeper phylogenetic relationships are the province of the 'viral self', which may be used to reveal higher level viral taxons discussed here as viral lineages.

2.3 Viral lineages

The difficulty in observing deep evolutionary relationships in viruses may be resolved by analysis focused on protein structure rather than the sequences (Krupovic and Bamford 2010). Last decades have seen considerable increase in published atomic coordinates, among them numerous high-resolution crystal structures of viral proteins. This led to the significant discovery that many viruses that infect phylogenetically distant hosts, and were previously considered unrelated, exhibit a highly similar fold in certain structural proteins (Benson et al. 1999, Bamford 2003). Bamford et al. (2003) define a 'viral self,' a set of essential proteins so vital for the virus that they will be retained through a remarkable timespan, at the structural level if not in sequence. It follows that conserved structures can be used for the higher-level classification of viruses even at a level where genetic signal of a shared origin is lost due to differentiation. The structural approach may be the only practical way to glean new insight into the origins of viruses in the face of tremendous variety in the primary sequence data. Indeed, conserved protein structures identify viral lineages that reach further back in evolutionary time than results obtained by any other method.

Presumably, only a limited subset among possible protein folds have potential to form any kind of a functional capsid (Abrescia et al. 2012). Genesis of such folds via natural selection would thus be a rare occurrence, explaining why there seem to be a limited number of observed capsid architectures and major capsid protein folds. Accordingly, the capsid protein fold is one of the most important and strongly conserved features of a virus. The trait that ultimately distinguishes a virus from other self-replicating genetic elements is that a virus can produce a virion capable of enclosing the genome and delivering it to a new host. Thus, the capsid, and by extension the capsid protein fold, is one of the most important hallmarks of a virus. In addition, capsid proteins are often among the easier virus components to isolate and crystallize. Comparative structural analyses of currently determined capsid protein structures have identified four prevalent structural lineages with ancient origins (Krupovic and Bamford 2010, Abrescia et al. 2012). These four lineages (picorna-like, double beta-barrel, HK97-like and BTV-like) encompass approximately half of the virus families currently classified by ICTV (Fig. 3). A small number of polyhedral viruses do not fall into any of the lineages, and viruses with helical and pleomorphic architectures are likely to belong to their own structural lineages, identified by folds with tendency toward alpha helical bundles.

The concept that structures could signify common origins was first suggested in 1980s, when the structures of major capsid proteins of southern bean mosaic virus and two animal picornaviruses revealed an unexpectedly similar eight-stranded beta-barrel fold (Abad-Zapatero et al. 1980, Rossmann et al. 1985). This fold is today considered the characteristic of the picorna-like viral lineage that consists of 15 families of small ssRNA viruses that infect eukaryotes (Abrescia et al. 2012). Second structural lineage was identified at the turn of the millennium when bacteriophage PRD1 capsid protein fold was found to be highly similar to the major capsid protein of human adenovirus (Benson et al. 1999). In addition to the capsid protein fold, a double beta-barrel aligned radially (upright) in the capsid, lineage members share similar use of a penton protein, spike proteins and a putative genome packaging adenosinetriphosphatase (ATPase) (Bamford 2003, Merckel et al. 2005). Soon afterwards, the archaeal *Sulfolobus* turreted icosahedral virus (STIV) was also found to exhibit the double beta-barrel capsid protein (Khayat et al. 2005). The double beta-barrel lineage is currently the most diverse and best characterized of all structural lineages. In 2011, Krupovic and Bamford attributed the double beta-barrel major capsid protein fold to 10 out of 28 dsDNA virus families classified by ICTV at the time (Krupovic and Bamford 2011).

Yet another structural lineage is exemplified by Hong Kong 97 (HK97) virus, a tailed bacteriophage with a dsDNA genome. HK97 major capsid protein gp5 is the archetype of one of the four prevalent capsid protein folds, the HK97-type (Helgstrand et al. 2003, Krupovic and Bamford 2011). Protein gp5 assembles to form icosahedral particles which go through several distinct procapsid intermediates during virion maturation (Duda et al. 1995). The maturation induces major changes in the capsid protein, including autocatalytic cleavage, significant cross-linking between capsid protein subunits, rigid body rotations

and refolding of some portions of the capsid protein structure (Duda 1998). The organization of the mature virion is famously described as viral 'chain mail'. The HK97-like structural lineage is among the most successful viral lineages and includes tailed phages and, perhaps surprisingly, herpesviruses (Baker M.L. et al. 2005).

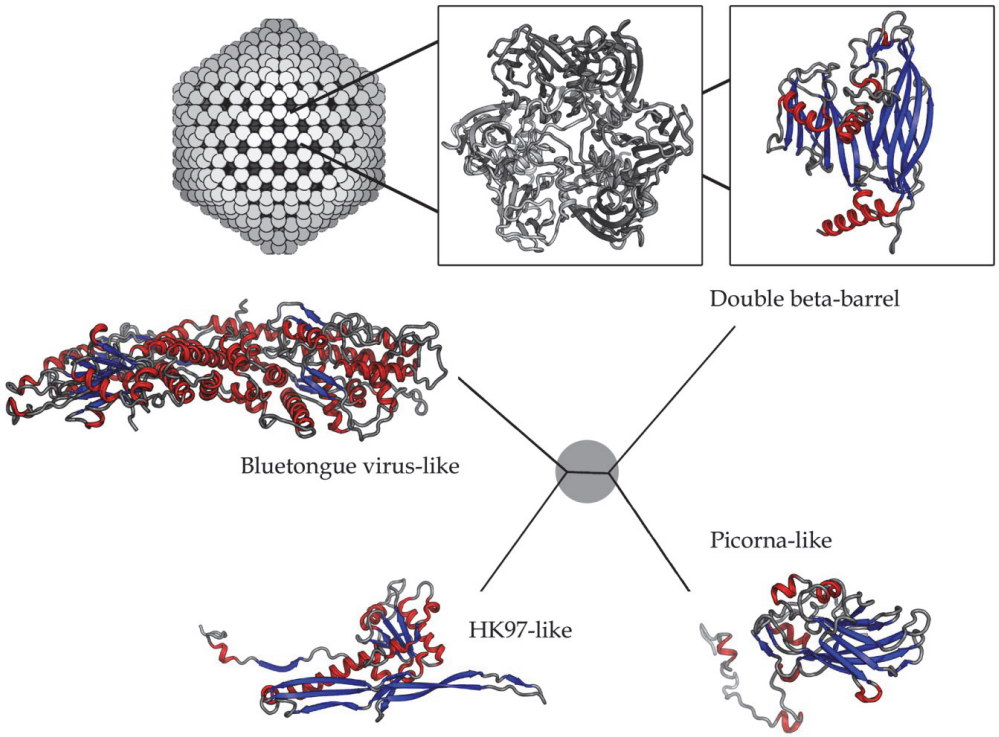


FIGURE 3 Four prevalent folds in icosahedral viruses. Clockwise from the top: the double beta-barrel fold from bacteriophage PRD1 (Protein Data Bank identifier, PDB id: 1GW7), the picorna-like fold from Cricket paralysis virus (PDB id: 1B35), the HK97-like fold from Hong Kong 97 phage (PDB id: 1OHG) and the BTV-like fold from bluetongue virus (PDB ID BTV2, VP3). Structures are presented in the orientation they assume in the capsid (the double beta-barrel fold sits upright in the shell while the major capsid proteins from other lineages assume a more level alignment). Grey circle at the middle signifies that the viral lineages do not share a monophyletic origin. Capsid illustration was created with Icosahedron Generator (http://johnsonlab.scripps.edu /research/icosahedron-generator.php).

The fourth structural lineage is characterized by intricate capsid morphology wherein the conserved fold is found in an inner capsid protein layer. Archetype of the lineage is the dsRNA bluetongue virus (BTV) that exhibits two major capsid proteins, namely VP7 and VP3, organized in layers. VP3 (the lineage archetype fold) forms an inner layer organized in a quaternary arrangement of 120 copies. This internal shell is cloaked by an outer layer of VP7 trimers (Grimes et al. 1998). Although features of the surface layer vary, the internal

structure of 120 capsid protein copies is found in other members of *Reoviridae* (McClain et al. 2010) as well as *Pseudomonas* phage phi6 (Cystoviridae) (Huiskonen et al. 2006), and suggested for a *Penicillium chrysogenum* virus and totivirus L-A (Naitow et al. 2002).

Problems involved in the use of comparative structural analysis to infer evolutionary relationships revolve around issues of coincidence and analogy (Abrescia et al. 2012). Physicochemical properties of peptide chains favor certain folding outcomes, which leads to the issue that all protein structures are bound to have certain similarities and makes it difficult to determine a point where similarities become coincidental. In evolutionary terms, the controversy lays in whether structural similarity is strong enough evidence to assume divergent evolution rather than convergent evolution. During convergent evolution, similar (*i.e.* analogous) features arise in different species due to environmental factors and do not signify common ancestry. Intriguingly, cellular protein sTALL-1 from the tumor necrosis factor cytokine family has a single beta-barrel fold reminiscent of the major capsid protein of tobacco necrosis satellite virus and sTALL-1 has also been reported to form virus-like icosahedral particles (Liu et al. 2002, Lane et al. 2011). This instance could be seen as either a case of ancient horizontal gene transfer from viruses to cells, evidence toward the escaped gene theory of virus origin, or as a case of convergent evolution where demand for certain characteristics has coincidentally brought about a similar fold in both sTALL-1 and small ssRNA viruses. Given the simplicity of the fold, lack of other similarities between cytokines and ssRNA viruses, and differences in the oligomerization pattern, convergent evolution seems most likely.

By contrast, homologous features derived from divergent evolution have been vertically inherited from parental generation. It follows that if the evolutionary path was divergent, more than one common feature inherited from the ancestors is likely to be found in descendant species. In case of the double beta-barrel lineage of viruses, common features include the major capsid protein fold, similar use of a penton protein, spike proteins and a putative ATPase sequence containing the P9-type motif (Benson et al. 1999, Bamford 2003, Abrescia et al. 2012). Lineage members display similar virion morphology yet have different assembly and disassembly pathways. These shared features in addition to the diverse host range are indicative of divergent evolution. Furthermore, while beta-barrels are certainly prevalent in a variety of biomolecules other than double beta-barrel viral capsid proteins, the alignment of the two beta-barrels in regards to each other and the prevalence of similar extensions such as the helical loop between strands F and G also point towards a common origin rather than coincidental emergence. Thus, strong structural similarity based on careful analysis of structural characteristics makes a compelling argument for homology arising from divergent evolution. In many cases, further evidence can be found in other features of the virus such as genome arrangement, genome packaging machinery and overall virion morphology.

2.3.1 The double beta-barrel lineage

The double beta-barrel lineage is the best characterized structural lineage and includes members infecting hosts from all domains of life (Krupovic and Bamford 2008). 'Viral self' characteristics of the lineage include a highly conserved major capsid protein fold consisting of two jelly rolls oriented upright in the capsid shell, and a putative genome packaging ATPase. In addition, most members share the same overall morphology consisting of an icosahedral capsid with an internal lipid membrane (Fig. 4). A related branch of viruses utilizing a similar major capsid protein based on a single beta-barrel rather than a duplicate barrel has been postulated based on morphological evidence. One of the putative members of the single beta-barrel lineage is bacteriophage P23-77, discussed more closely in Chapter 2.4. Together the two lineages are suggested to comprise a viral super lineage, characterized by a conserved jelly roll fold aligned upright in the capsid, thereinafter called the beta-barrel super lineage.

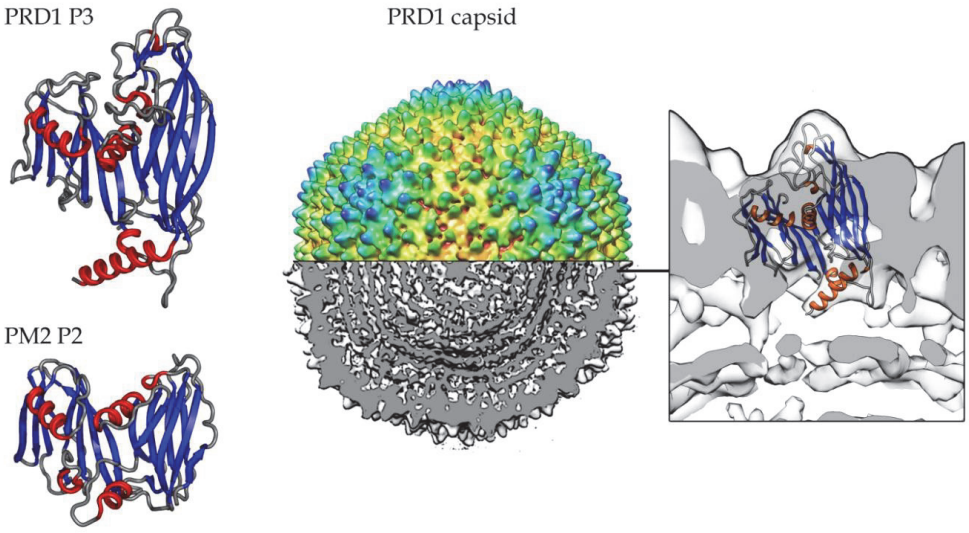


FIGURE 4 Characteristics of the double beta-barrel-lineage. Lineage members, bacteriophages PRD1 and PM2, feature the titular double beta-barrel major capsid protein fold (PDB ids: 1GW7 for PRD1 P3 and 2VVF for PM2 P2). Major capsid protein sits upright in the capsid shell, as exemplified by PRD1 major capsid protein fitted into the capsid electron density (The Electron Microscopy Data Bank identifier, EMDB id: 1011).

Members of the double beta-barrel lineage form the bulk of their capsids from single major capsid protein, organized in trimers, each trimer comprising one pseudohexagonal capsomer in the capsid lattice (Krupovic and Bamford 2008). The major capsid protein fold includes two jelly roll domains (V1 and V2), each consisting of eight antiparallel beta-strands arranged into two sheets, strands following the B-I-D-G and C-H-E-F naming scheme for viral jelly rolls (Fig 4).

Currently identified lineage members include eukaryal adenoviruses (Roberts et al. 1986, Athappilly et al. 1994), *Paramecium bursaria chlorella* virus-1 (Nandhagopal et al. 2002) and vaccinia virus (Bahar et al. 2011), bacteriophages PRD1 (Benson et al. 1999, Abrescia et al. 2004) and PM2 (Abrescia et al. 2008), archaeal virus STIV (Khayat et al. 2005) and virophage Sputnik (Zhang et al. 2012). Crystal structures have been determined for both PRD1 and PM2 virions, allowing detailed analysis. Furthest removed members of the lineage, such as adenoviruses, often exhibit additional domains or extensions on top of the double beta-barrel fold; nevertheless, the core fold is retained in high detail.

Comparative genomic studies have revealed that a number of large nuclear and cytoplasmic DNA viruses have monophyletic origin (Iyer et al. 2001, Iyer et al. 2006). Interestingly, families revealed to share common ancestry are poxviruses, asfarviruses, iridoviruses, phycodnaviruses and mimiviruses, all of which have been connected to the double beta-barrel lineage either by morphological and genetic evidence or, when atomic coordinates are available, by conserved structural protein folds (Krupovic and Bamford 2010, Bahar et al. 2011). This illustrates that subgroups of the lineage may be identified by genomic methods. Based on the fact that each family infects a vastly different eukaryotic host, divergence of the previously mentioned five families of DNA viruses is predicted to have taken place at an early stage of evolution, prior to the establishment of the major eukaryotic lineages. The entire double beta-barrel lineage can be traced back from the subgroups, outlined by Krupovic and Bamford, towards the root of the lineage (Krupovic and Bamford 2008). Current members of the lineage infect hosts across all domains of life, leading to the conclusion that the ancestor of the entire lineage may have existed before the divergence of the three domains, over three billion years ago.

One of the most interesting outcomes of tracing a lineage based on such deep evolutionary relationships is that it allows us to examine lineage members that exhibit varying levels of complexity, and infer the features that the common ancestor may have possessed. These would be difficult to determine from a more closely related group where the variety of specialized traits may hamper the identification of the 'viral self'. Current understanding is that the common ancestor of the double beta-barrel lineage was an icosahedral DNA virus that utilized an upright beta-barrel major capsid protein and a specific ATPase in genome packaging (Krupovic and Bamford 2008, Jalasvuori et al. 2009). This ancestor may have then diverged to form two main lineages: single beta-barrel lineage, the main focus of this thesis, and the double beta-barrel lineage.

2.4 Bacteriophage P23-77

P23-77 is an icosahedral, tailless dsDNA bacteriophage that infects themophilic Gram-negative bacterium *Thermus thermophilus* (ATCC 33923). It was isolated in 2005 from alkaline hot springs on the North Island of New Zealand during a study directed by the Promega Corporation (Yu et al. 2006). Of the 115

bacteriophage strains isolated and characterized during the study, P23-77 was among the most stable. Further characterization showed that the roughly spherical capsid shell consists mainly of two protein species and is 78 nm in diameter (facet-to-facet), complemented by 15 nm long spikes extending from the vertices (Jaatinen et al. 2008). The capsid is arranged in an unusual T=28 lattice and encloses an internal membrane consisting mainly of phosphoglycolipids and neutral lipids that are obtained selectively from the host. The lipid membrane is located between the capsid shell and the dsDNA genome.

Bacteriophage P23-77 has 17036 bp long circular dsDNA genome containing 37 assigned open reading frames (ORFs), among which approximately ten ORFs encode structural proteins (Jalasvuori et al. 2009). Three structural proteins are associated with the capsid shell: major capsid proteins VP16 (19.1 kDa, coded by ORF16) and VP17 (31.8 kDa, coded by ORF17) and minor capsid protein VP11 (22.1 kDa, coded by ORF11). In addition, assigned ORFs encode five membrane-associated proteins, a putative DNA replication initiation protein, enzymes for bacterial cell wall degradation and an ATPase that is predicted to function in DNA packaging (ORF13). Viral polymerases were not identified with sequence-based methods, suggesting that the virus may utilize host polymerases.

One of the most notable characteristics of the P23-77 capsid is its crenellated topology. Each hexagonal capsomer bears two turrets that occupy different positions depending on the symmetron in which the capsomer belongs (Jaatinen et al. 2008). Presence of two towers overrules the possibility of an underlying three-fold symmetry, observed in the homotrimer capsomers of double beta-barrel lineage members. Triangulation number 28 together with the arrangement of the turrets suggests an organization where, since the capsomers forming the disymmetron are themselves 2-fold symmetric, the capsomers may not be formed from a trimer of double beta-barrels but rather of six single beta-barrels. Similar capsid architecture is found in extremophile viruses SH1, HHIV-2 and SSIP-1, all putative members of a proposed single beta-barrel branch of the beta-barrel super lineage (Jaalinoja et al. 2008, Aalto et al. 2012, Jaakkola et al. 2012). Prior to this study, existence of the single beta-barrel lineage has not been confirmed at the level of protein structure.

2.4.1 Relation to the double beta-barrel lineage

P23-77 exhibits the overall morphology typical of the double beta-barrel lineage: an icosahedrally ordered capsid with an internal lipid membrane (Jaatinen et al. 2008). Lipid constitution of P23-77 is similar to PRD1 and the spacing of both the internal membrane and packaged DNA is relatively close to corresponding values in PRD1, STIV and PM2 (Abrescia et al. 2004, Abrescia et al. 2008, Khayat et al. 2010). In addition, P23-77 major capsid protein VP17 was predicted to have 45% structural similarity to the major capsid protein of STIV (Jalasvuori et al. 2009). Sequence analysis revealed that P23-77 genome codes for a putative ATPase containing the P9-type motif typical of the double beta-barrel lineage. Given the morphological similarities and the presence of lineage-typical ATPase (considered a part of the set of features comprising the 'viral self'), association of

P23-77 to the lineage is well founded. However, the hypothesis that P23-77 utilizes capsid proteins with single beta-barrel core fold as opposed to the highly conserved double beta-barrel clearly sets the virus apart from the adeno/PRD1-type.

In addition to double beta-barrel lineage members, P23-77 is related to certain viruses, plasmids and integrated genetic elements of halophilic archaea and thermophilic bacteria (Jalasvuori et al. 2009, Jalasvuori et al. 2010). The highest sequence similarity was found between P23-77 and *T. thermophilus* (previously reported as *T. aquaticus*) phage IN93, especially considering the major capsid proteins VP16 and VP17 and the putative genome packaging ATPase with sequence identities of 79, 74 and 79 %, respectively (Jalasvuori 2009). Related genetic elements were discovered in *Haloarcula hispanica* virus SH1 genome, in *Halobacterium salinarium* plasmid pHH205 and integrated in the genomes of bacterial *Meiothermus* species and archaeal *Halomicrobium*- and *Haloarcula* species. Interestingly, the genetic evidence designates viruses infecting hosts from both *Thermaceae* (Bacteria) and *Halobacteriaceae* (Archaea) as relatives to P23-77, establishing unique link between viruses of extremophiles belonging to the two domains. All the genomes and the plasmid are formed of double-stranded DNA and contain three common viral genes, two coding for major capsid proteins and one coding for a putative genome-packaging ATPase, arranged similarly within the genome. This ORF trinity appears to be a genetic hallmark of the putative single beta-barrel branch of the beta-barrel super lineage.

Given the host range and morphological features pointing to single beta-barrel capsid protein composition, P23-77 could represent an ancient branch of the beta-barrel super lineage. Many neutral mutations that are well tolerated at standard temperatures become strongly deleterious at higher temperatures, inducing negative selection and a decreased evolutionary pace (Friedman et al. 2004, Drake 2009). In a constant environment, these conditions may allow the survival of ancient viral forms that have been outcompeted in moderate environments. Thus, P23-77 is an interesting target for structural studies as these could shed light on the origins of the double beta-barrel viral lineage as well as reveal ancient viral features only preserved in extreme conditions.

2.5 From protein structure to assemblies and phylogeny

2.5.1 Viral capsid modelling

Atomic level structures have been determined for a vast array of viral components, but major challenges remain in single crystal X-ray diffraction studies of entire virions. Still in 2014, relatively few virion crystal structures are available in the Protein Data Bank (PDB) (Berman et al. 2000). Main obstacle lays in the virion crystallization: the larger the target, the more difficult it will be to grow well-diffracting crystals. In certain cases, methodologies may be combined to produce results similar to virion crystal structure without requiring the

extensive crystallization process (Stockley and Twarock 2010). Cryo-electron microscopy (cryo-EM) provides a feasible venue to determine the topology of viral capsids, and numerous cryo-EM density maps are available from The Electron Microscopy Data Bank (EMDB). Cryogenic freezing of unstained samples and techniques for data collection that allow 3-D reconstructions of the sample can construe virus structures that rival those obtained by X-ray crystallography, and even at lower resolution, capsid electron density from the cryo-EM combined with the crystal structures of viral capsid proteins may yield a realistic presentation of the capsid. In most cases, data from cryo-EM and X-ray crystallography has proven consistent. Accordingly, X-ray structures can often be fitted into cryo-EM density, which dramatically increases the amount of interpretable information yielded by cryo-EM while providing a certain validation for crystal structure reliability in an *in vivo* framework. In a number of situations, the intermediate-resolution cryo-EM data and high-resolution X-ray diffraction data have been combined to achieve biological insights impossible to either method alone. In this study, the P23-77 major capsid protein structures are refined into the P23-77 virion cryo-EM density to discern capsid organization and postulate the capsid assembly process.

2.5.2 Comparative structural analysis

Sequence-based comparative methods have been used to establish the phylogeny of cellular organisms, and are sometimes applicable to closely related viruses (*i.e.* at family level). However, deeper evolutionary relationships between viruses are generally not discernible from sequences due to loss of signal at genetic level. Methods have been developed for structural comparison using selected sets of atomic coordinates (Stuart et al. 1979, Holm and Sander 1993, Krissinel and Henrick 2004), and recently these comparisons have been used to infer evolutionary relationships (Bamford et al. 2002, Stockley and Twarock 2010). Homology detection is based on attaining suitable superposition and calculating, based on the similarities, objective evolutionary distance. In this study we utilized the Structure Homology Program (SHP) (Stuart *et al.* 1979), the principles of which were recently discussed by Abrescia et al. (2012). SHP compares two structures by exploring all relative orientations where each residue is compared to its counterpart in the other structure. Local shape and C-alpha atom distance are used to discern the likelihood of equivalency for the residues and a matrix containing the equivalency probabilities for all residue pairs is analyzed to find a best set of matching residues. Sum of the probabilities in the best set provides a measure of structural similarity in the relative orientation and can be converted into an indication of evolutionary distance via metrics analogous to those used in sequence comparison. Nevertheless, calculating correct superimposition and inferring from this the objective evolutionary distance remains a recursive and challenging process. As it becomes increasingly evident that structures contain conserved features that can be used as evolutionary markers, comparative structural analysis is gaining importance as a phylogenetic tool.

3 AIMS OF THE PRESENT STUDY

This thesis focuses on conserved structures in viruses, more specifically the major capsid proteins of extremophile bacteriophage P23-77. Based on genomic structure and morphological features, P23-77 is suggested to be related to a diverse viral lineage identified by a highly conserved double beta-barrel major capsid protein fold. However, surface architecture of P23-77 implies that the bacteriophage may exhibit features not previously found in the lineage. The extremophilic nature of P23-77 and the connection the virus has to the double beta-barrel lineage indicate potential for completely novel viral protein structures and capsid organization. In addition, phylogeny of the double beta-barrel lineage reaches far back in evolution, and identification of novel members from extreme environments may give insight into the nature of early viruses on primordial Earth. The detailed aims of the present study were as follows:

I To set up efficient methods for expression and purification of the two major capsid proteins of bacteriophage P23-77 and to optimize crystallization conditions to yield single well-diffracting crystals of each protein separately and in complex with each other;

II To determine high-resolution structures of the major capsid proteins, namely VP16, VP17 and VP16/VP17 complex, resolve relationship to the double beta-barrel lineage and infer the organization of the P23-77 capsid;

III To characterize the properties of the major capsid proteins and their function in capsid assembly in more detail;

IV To relate current findings to existing viral taxonomy.

4 OVERVIEW OF THE METHODS

Detailed description of materials and methods used can be found in the original publications indicated by Roman numerals. Methods used in this study are summarized in Table 1.

TABLE 1 Methods used in the original publications included in the thesis

Method	Publication
DNA cloning and sequencing	I, III
QuikChange mutagenesis	III
Protein expression and purification	I, II, III
Virus production and purification	I, II
Crystallization	I, II
X-ray diffraction experiments	I, II
Screening for heavy-metal derivatives and data processing	II
Structure determination and refinement	II
Comparative structural analysis	II
Computational capsid model generation and refinement	II
Protein oligomeric state assays	II, III
Differential scanning calorimetry	III
Dynamic light scattering spectroscopy	III
Comparative sequence analysis	IV
Molecular graphics and analysis	II, III, IV

5 RESULTS AND DISCUSSION

5.1 P23-77 major capsid protein expression and purification (I)

Genes coding for VP16 and VP17, the major capsid proteins of bacteriophage P23-77, were identified by Jalasvuori et al. (2009). Further analysis required the development of expression and purification protocols for the full-length native major capsid proteins (I). Genes ORF16 and ORF17 were PCR amplified from the P23-77 genome and ligated into pET22b(+) vector and the resulting plasmids were used to transform competent cells for recombinant protein expression. Efficient expression was achieved in *Escherichia coli* HMS174 (DE3) cells grown at +28 °C, 230 rev min^{-1} and expression induced with 1 mM IPTG. Cells were collected after 22 hours of cultivation and disrupted.

Majority of each protein was found in the soluble fraction after disruption and ultracentrifugation. Given the termophilic nature of the P23-77 host, major capsid proteins were screened for heat stability by incubation and centrifugation cycles. P23-77 virion is most stable at moderate temperatures and most infective around +70 °C (Jaatinen et al. 2008); however, the major capsid proteins were discovered to be much more robust and withstand temperatures of over +80 °C (I). To my knowledge, this level of thermostability has not been previously reported for recombinant viral capsid proteins and is high even compared to certain specifically engineered proteins (Asial et al. 2013). The thermal stability of P23-77 major capsid proteins was further characterized in 2013 (III). In the current study, high thermostability was exploited during the purification process of each major capsid protein: soluble fraction from the ultracentrifuged crude lysate was incubated at +90 °C for 10 minutes, which aggregated host-derived impurities while retaining most of the target protein (I).

Chromatography conditions were screened for efficient protein purification. For VP16, purification was achieved with anion exchange chromatography in 20 mM ethanolamine pH 9.5, where VP16 flows through the column (5 ml Q HP HiTrap column, GE Healthcare) immediately while majority of host-derived protein binds. This reversal of the typical execution of anion exchange

chromatography yielded the highest purity level for VP16. VP17 begins to bind to the column in pH >9 and was accordingly purified with an anion exchange chromatography in 20 mM ethanolamine pH 9.5 where VP17 was eluted with 50 mM NaCl. Lastly, remaining impurities were removed with size exclusion chromatography. Achieved protein samples were stored at +8 °C in 20 mM Tris pH 7.5, and were very stable with little to no aggregation over multiple months.

5.2 Crystallization and preliminary crystallographic analysis (I)

Previously described protocols can be used to produce native full-length VP16 and VP17 at purity level suitable for crystallization trials. Both major capsid proteins and P23-77 virus material (purified according to the protocol by Jaatinen et al. 2008) were screened for crystallization conditions at room temperature using hanging- and sitting-drop vapor diffusion methods (I). Experiments yielded a number of different crystal types.

Promising conditions for VP16 crystallization (VP16-type-1) were identified during the very first screen when 10 % (weight/volume) polyethylene glycol (PEG) 8000, 10 % PEG 1000 buffer conditions (Crystal Screen HT, Hampton Research) yielded plate- and needle-like crystal forms. Conditions were manually optimized until single crystals with cuboid morphology were attained in 5 % PEG 1000, 5 % PEG 8000. VP17 was found to form bipyramid crystals in solutions containing sodium formate. The conditions for VP17 crystallization were further optimized to produce sizable single crystals in 1.9 M sodium formate, 0.1 M bis-Tris pH 7.0. Both VP16-type-1 and VP17 crystals were robust and grew to full size within two weeks. During optimization, various protein concentrations were screened to enhance crystallization. Improved crystallization for both VP16 and VP17 was achieved when protein concentration was approximately 1 mg/ml in the crystallization drop, prepared 1:1 with the well solution. Accordingly, subsequent screening and final crystallization trials were conducted in similar concentrations. Second type of VP16 crystal (VP16-type-2), putative VP16/VP17 complex and virion derived crystal types were discovered during screening at the Oxford Protein Production Facility UK. Cuboid VP16-type-2 crystals grew within days in 20% PEG 6000, 0.1 M citrate pH 4. By contrast, the single putative complex crystal took three months to grow from a mixture of VP16 and VP17 (~1:1 molar ratio) in 1.1 M di-ammonium tartrate pH 7. Crystallization trials for purified P23-77 virion yielded needle shaped crystals in conditions containing 0.1 M citric acid pH 3.5, 20 mM Tris–HCl pH 7.5, 5 mM MgCl2, 150 mM NaCl and 25% PEG 3350.

All five crystal types, VP16-type-1, VP16-type-2, virion derived, VP17, and putative VP16/VP17 complex, were tested for diffraction. Diffraction was confirmed with *in house* experiments, followed by native data set collection at the Diamond Light Source synchrotron, Didcot, England. Native data for VP16-type-1 were collected at beamline I03 at wavelength (λ) 0.979 Å; for VP16-type-2, at beamline I04 (λ = 1.000 Å); for VP16-virus-derived, at beamline I24 (λ = 0.969 Å);

for VP17, at beamline I04 (λ = 1.071 Å); and for the putative complex, at beamline I24 (λ = 1.071 Å). The data were automatically processed with xia2/XDS (see I for crystallography table) (Kabsch 1993, Winter 2010). All crystal types diffracted to higher than 3 Å resolution, with multiple crystal types diffracting up to 1.26-1.8 Å. Unit cell dimensions and crystal solvent content analysis indicated that all crystal types, excepting the virion derived, likely contain 1 or 2 proteins in their asymmetric unit. Unit cell for the virion derived crystal was too small to contain an entire virion; thus, crystallized material was most likely VP16 and/or VP17, as these comprised most of the protein material in the sample and have been shown in this study to crystallize readily. Data sets collected in the current study were used for structure determination after suitable heavy atom derivatives were achieved (II).

5.3 The major capsid protein structures and phylogeny of P23-77 (II)

Previous experiments had produced five well-diffracting crystal types: VP16-type-1, VP16-type-2, virion derived, VP17, and putative VP16/VP17 complex (I). According to the original hypothesis, both VP16 and VP17 contain a novel single beta-barrel fold, rendering molecular replacement ineffective for phasing. However, native protein crystals were stable enough to allow heavy atom soaking, which after comprehensive screening yielded isomorphous derivatives for VP16 and VP17 (II).). Derivatives of VP16-type-1 and VP17 used for phasing were prepared by soaking a crystal in lead acetate and mercury acetate, respectively. Derivative data were collected at the Diamond Light Source synchrotron beamline I03 at λ = 0.940 Å (VP16-type-1) and beamline I02 at λ = 1.009 Å (VP17). Crystal structures were solved by isomorphous replacement. Heavy atom binding sites were determined with SHELXD (Schneider and Sheldrick 2002), phases calculated with autoSHARP and initial models built by ARP/wARP within autoSHARP (Perrakis et al. 1999, Vonrhein et al. 2007). Resulting models contained 95% and >57% of the residues of VP16 and VP17, respectively. Structures of VP16-type-2, virion derived crystal and putative VP16/VP17 complex were solved by molecular replacement with PHASER (McCoy et al. 2007) using the structures of VP16-type-1 and VP17 as search models. Initial models of all structures were further refined by cycles of manual model building in Coot (Emsley and Cowtan 2004), followed by refinement by PHENIX (Adams et al. 2010) and/or Buster (Bricogne et al. 2011). Refined structural coordinates were deposited to PDB (Table 2).

Structures revealed that both VP16 and VP17 exhibit nearly identical, anti-parallel eight-stranded single beta-barrel core fold consistent with the B-I-D-G and C-H-E-F naming scheme for viral jelly rolls (Fig. 5), as expected based on the association with the double beta-barrel lineage (II). Surprisingly however, VP16 was revealed to form strand-swapped dimers, and VP17 bears an additional

domain composed of a seven-stranded beta-barrel insertion. As suggested by preliminary crystallographic analysis (I), crystals derived from purified P23-77 virus did not contain entire virion but were composed of VP16 (VP16-virus-derived) (II). To our satisfaction, the putative VP16/VP17 complex crystal was revealed to contain both VP16 and VP17 in an oligomeric arrangement.

TABLE 2 Summary of the P23-77 major capsid protein structures determined in this study. Values in parentheses are for the highest resolution shell. For detailed crystallography table, see Rissanen et al. 2013 (II).

	VP16 type-1	VP16 type-2	VP16 virus-derived	VP17	VP16/VP17 complex
PDB ID	3ZMO	3ZN4	3ZN5	3ZMN	3ZN6
Resolution (Å)	1.8	1.26	2.36	2.26	1.53
Space group	$P6_222$	C2	$P2_12_12_1$	$P6_122$	C2
Rmerge	0.075 (1.045)	0.053 (0.626)	0.114 (1.137)	0.082 (0.917)	0.064 (1.015)
Rwork/Rfree	0.189/0.215	0.156/0.184	0.193/0.234	0.186/0.213	0.172/0.197
Residue range*	2-173	19-167	2-169	41-273	21-165/46-271
Chains	1	1	8	4	2

* The determined residue range for the most complete polypeptide chain for each crystal type; native full-length VP16 comprises 173 amino acids and VP17 291 amino acids.

VP16 structures type 1 and type 2 both contain one VP16 in the asymmetric unit; the other strand-swapped subunit is found in the symmetry cell. Virus derived VP16 crystal contains 8 VP16 proteins arranged in 4 strand-swapped dimers. A chloride ion is found at the strand swap site in all determined VP16 structures and may contribute to the strand swap, reminiscent of CD47 (Hatherley et al. 2008). To my knowledge, strand-swaps have not been identified in the capsid proteins of any of the described viral lineages. The feature seems unique to P23-77 and raises interest in understanding how it factors into the capsid organization and assembly. Comparative analysis of all determined crystal structures revealed that the folds are highly similar in different crystal types of the same protein; the C-alpha atoms of VP16 from different crystal types superimposed with root-mean-square deviation ranging from 0.2 to 1.2 Å. Strand-swap was present in all virion-derived VP16 structures which strongly implies that the strand-swap was not an artefact of recombinant expression but rather the native form of VP16. VP16-type-1 was the most complete of the determined VP16 structures, lacking only the N-terminal methionine residue. Besides the prominent strand-swap dimerization, other biologically relevant interaction surfaces were not evident in the subunit and lattice organization of any of the VP16 crystals.

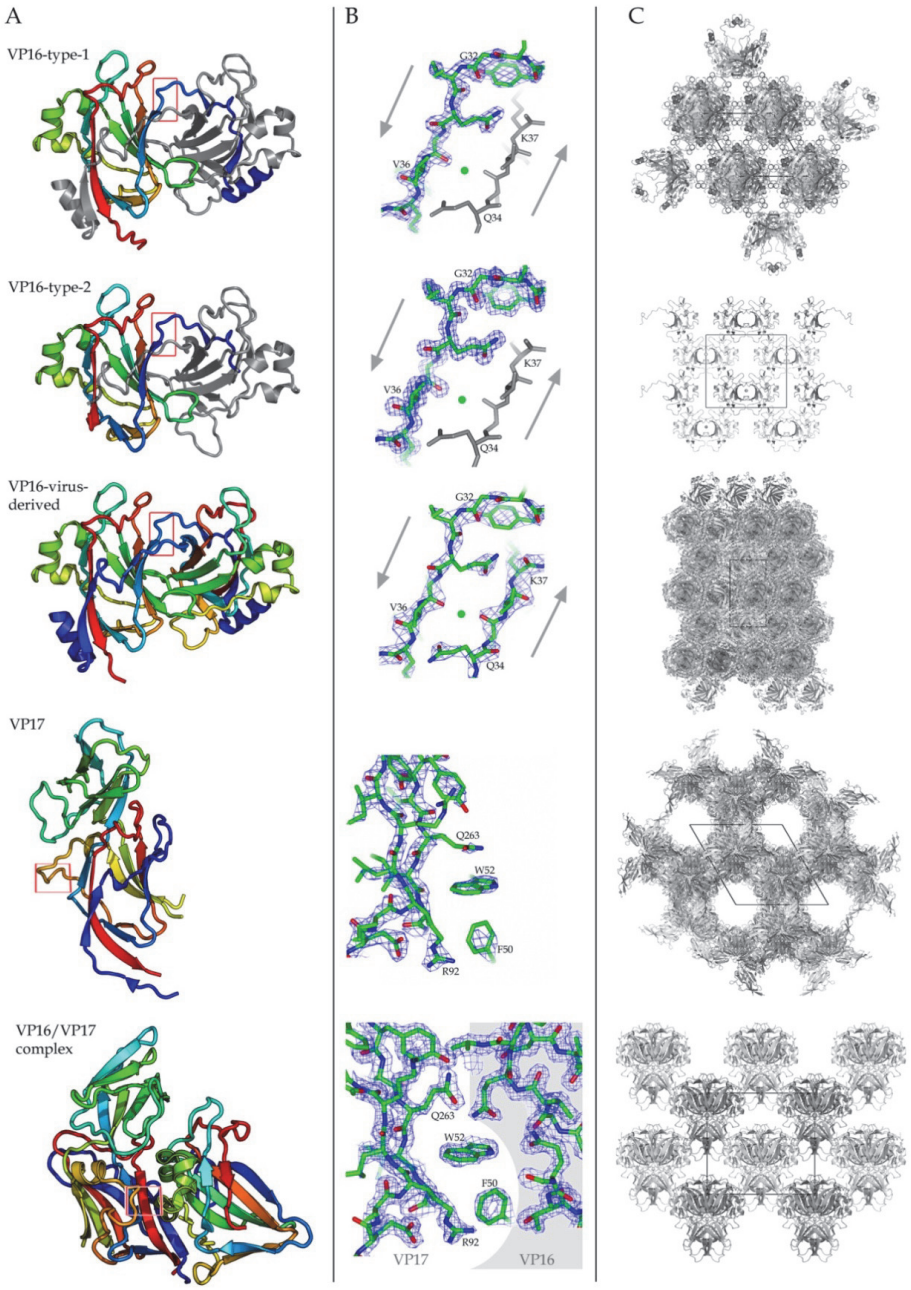


FIGURE 5 P23-77 major capsid protein structures from five different crystal types (II). Panel A illustrates the protein folds, colored blue to red from the N- to the C-terminus; symmetry molecules are shown in gray for VP16-type-1 and VP16-type-2 and red frames indicate the location of the residues presented in panel B. Panel B displays electron density at the strand swap site in VP16 structures (from above, green dot signifies the chlorine ion) and at the oligomerization surface for VP17 and the complex. Panel C shows crystal packing (not to scale). Content of one unit cell is indicated; view from the ab face.

VP17 crystal has two VP17 proteins in the asymmetric unit. The N- and C-termini could not be determined (42 and 18 undetermined residues, respectively), indicating flexibility typical of viral capsid proteins. The lower domain of VP17 has a full set of eight antiparallel strands; the upper domain is an insertion at the middle of the barrel between strands D-E and comprises seven strands. Function of the upper domain is not known. The domain could enhance stability in high-temperature environment or function in capsid assembly by blocking inappropriate oligomerization. *e.g.* with other VP17. Like the VP16 strand-swap, the upper domain of VP17 is a feature uncommon to the major capsid protein structures found in the double beta-barrel lineage of viruses. Packing of VP17 in the crystal lattice creates channels devoid of organized structure (Fig. 5, panel C). Biologically relevant homotypic interaction surfaces were not evident in the lattice, but the VP17 termini were always aligned towards the channels, which could indicate that the channels are occupied by the unorganized termini.

The single well-diffracting crystal of the VP16/VP17 complex contains one VP17 and one VP16 dimerized in the crystal asymmetric unit. Consistent with the other VP16 structures, the VP16 subunit displays the strand-swap also in the complex form; the complementary VP16 subunit is found in the symmetry related molecule. Considering the symmetry, VP16/VP17 complex is a heterotetramer where each subunit of a VP16 dimer binds one VP17. Lateral arrangement of the VP16 and VP17 subunits in the complex carries important implications for the capsid organization, as discussed in Chapter 5.4.

Structural comparison analysis by SHP (Stuart et al. 1979) revealed that the core fold of both P23-77 major capsid proteins is highly similar to individual V1 and V2 domains from the characteristic fold of the double beta-barrel lineage (II). In other words, each single beta-barrel in VP16 and VP17 is similar to one half of the double beta-barrel observed in the lineage. Similarity was most significant between P23-77 major capsid proteins and the major capsid protein P2 of marine bacteriophage PM2, previously suggested as the most ancestral member of the lineage (Abrescia et al. 2008). The PM2 major capsid protein has two upright beta-barrels (V1 and V2) fairly similar in shape, contrasting with other related viruses where the domains are distinctly different. Superimposition of P23-77 VP17 lower domain and PM2 P2 domain V1 illustrates the level of similarity in the fold (Fig. 6). In addition to the barrel shape, orientation and dimensions, all major capsid proteins from beta-barrel super lineage members, including P23-77, share a helical loop between strands F and G (II). The site bearing the upper domain in VP17 is also found to carry major insertion in vaccinia virus scaffolding protein D13 (Bahar et al. 2011). Overall, structural similarity of the core fold confirms that the P23-77 major capsid proteins exhibit the conserved upright beta-barrel, but the virion is formed of two single beta-barrel major capsid protein species instead of a one capsid protein bearing two barrels (II). Since the major capsid protein is of critical importance to the virus, part of the 'viral self', it is very unlikely to be acquired via horizontal gene transfer, which signifies that the fold is a marker for common ancestry. Seeing that the upright beta-barrel core fold is present in lineage members infecting hosts across the

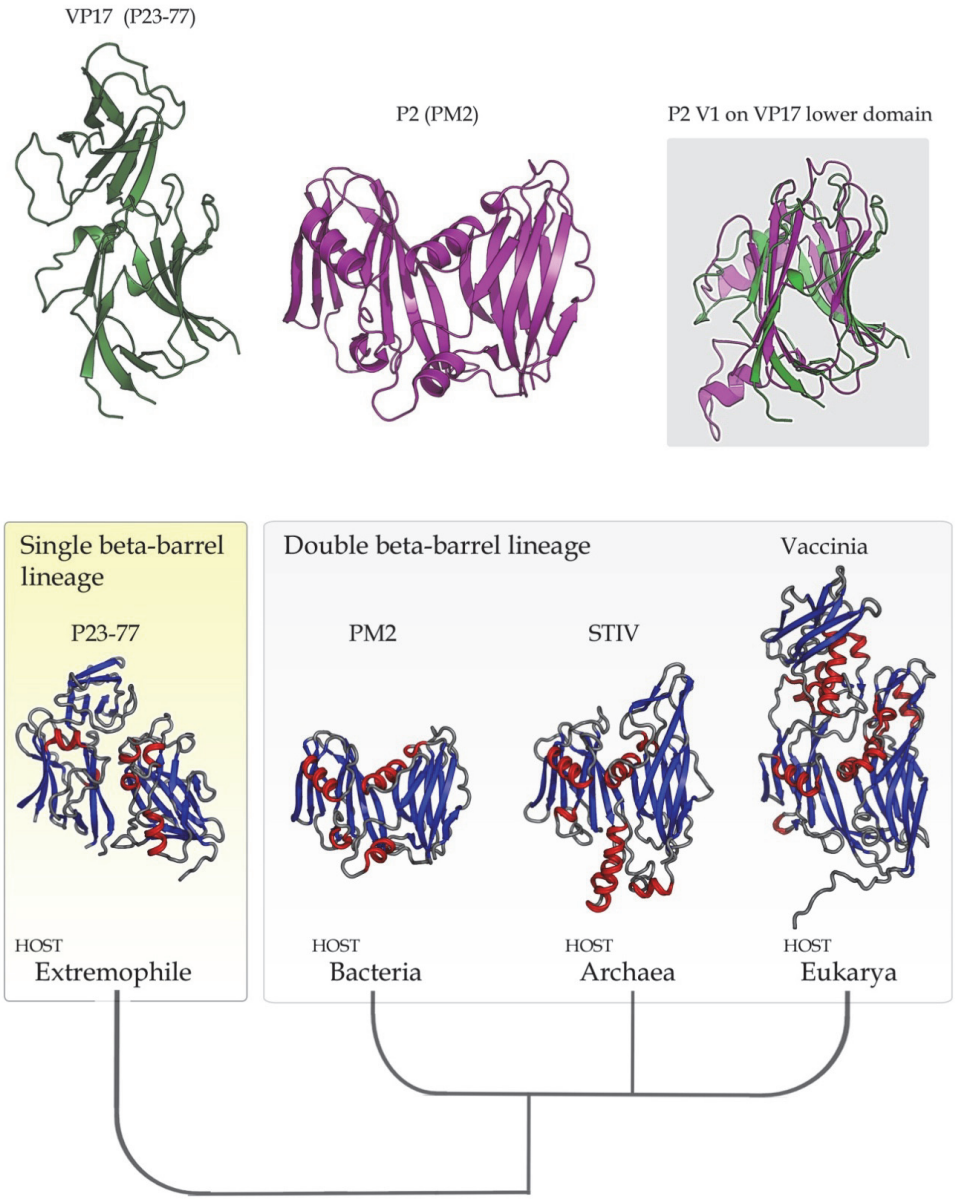


FIGURE 6 Structural similarity associates P23-77 to the double beta-barrel lineage. P23-77 protein VP17 (green) and PM2 protein P2 (domains V1 and V2, purple) contain a homologous beta-barrel core fold. Superimposition of VP17 lower domain and PM2 domain V1 highlights structural similarity (upper right corner). Below lies an illustration of the concept of the beta-barrel super lineage: single and double beta-barrel viral lineages have diverged from each other before the divergence of the three domains. The single beta-barrel lineage is suggested to include viruses of extremophilic archaea and bacteria. PDB codes for coordinates used in creating the figure were 3ZN6 (P23-77), 2VVF (PM2), 3J31 (STIV) and 2YGB (vaccinia).

domains of life, the common ancestor of the lineage may be traced back to the time before the divergence of the domains of life. A phylogenetic tree, based on the major capsid protein folds of the super lineage members, shows the divergence of the single and double beta-barrel branches (Fig. 5 in publication II). The analysis is based on complete capsid protein structures instead of individual domains, which accounts for the distance between the single beta-barrel P23-77 and the double beta-barrel lineage members. By contrast, comparison of individual domains signifies the closeness of P23-77 to PM2 (Fig. 2 in publication II). Both phylogenetic trees reported by Rissanen et al (II) were generated using SHP (see page 28 for details), except in case of archaeal virus SH1 where crystal structures of the major capsid proteins were not available. However, SH1 is included in the phylogeny as a putative member of the single beta-barrel branch based on similar morphology and genomic organization (Fig. 5 in publication II). Divergence of the single- and double beta-barrel lineages is likely an early event, followed by further diversification within each branch.

Thermus thermophilus, the host of P23-77, thrives in conditions where the temperature is around +65 °C (up to +85 °C for certain strains) (Oshima 1974, Henne et al. 2004). Such extreme environments are typically uniform in terms of species diversity and are dominated by only a few organisms (Skirnisdottir et al. 2000, Oren 2002, Hacene et al. 2004). However, studies of crenarchaeal hyperthermophilic viruses have shown that viral communities in such environments can exhibit exceptionally diverse genomes and morphotypes (Rice et al. 2001, Prangishvili 2003, Prangishvili and Garrett 2005). One possible explanation for this avenue of observations is that extreme environments may harbor viral forms extinct elsewhere. Many neutral mutations that are well tolerated at standard temperatures become strongly deleterious at higher temperatures, causing thermophilic organisms to evolve more slowly than mesophiles (Friedman et al. 2004, Drake 2009). Accordingly, every improved feature acquired for a specific niche will be highly conserved and if the ecological niche does not alter for millions of years, contemporary species might retain many of the features from their ancestors. *Thermus* bacteria belong to the Deinococcus-Thermus phylum, which is considered one of most ancient prokaryote groups. Thus, both P23-77 host range and the host's preferred environment can be considered to represent a stable niche for P23-77 that may have stayed highly similar since primordial times (II). Of the two types of major capsid protein fold found in the modern lineage members, the common ancestor of the beta-barrel super lineage is more likely to have utilized a single rather than double beta-barrel major capsid protein, given that is more probable for the double beta-barrel structure to have arisen through duplication than for it to have been split into distinct single beta-barrels. Thus, given that all putative members of the single beta-barrel lineage are found in extreme environments, we propose that the archetype of the branch, P23-77, represents an ancient virus type closest to the common ancestor of the super lineage.

5.4 P23-77 capsid organization (II)

Structures of VP16 and VP17, and most importantly the VP16/VP17 complex, allow insight into how the major capsid proteins are organized in the capsid (II). Superimposition of VP16 and VP17 with the major capsid proteins from other lineage members makes it apparent that VP16 and VP17 jelly rolls sit orthogonal to the capsid plane, with termini pointing towards the internal lipid membrane. VP16/VP17 complex clearly demonstrates that the two protein species align laterally so that the jelly rolls are at the same level, with VP17 upper domain spanning above the oligomerization site (Fig. 5, page 35).

In 2008, P23-77 capsid cryo-EM reconstruction had revealed crenellated topology, explained at the time by a two-protein configuration where one major capsid protein would form the capsomer base and another would be positioned on top, comprising the crenellations (Jaatinen et al. 2008). However, the crystal structures for VP16 and VP17 challenge this model, implying that the capsomer base is formed from VP16 and VP17 lower domain while VP17 upper domain comprises the crenellations (II). To test this hypothesis, the novel VP16 and VP17 structures were fitted into the 15 Å cryo-EM reconstruction of the capsid, published by Jaatinen et al. (2008). An asymmetric unit for the icosahedral capsid, consisting of individual VP16 and VP17 subunits, was constructed manually in Coot (Emsley and Cowtan 2004) and then refined into the density with Visual Environment for Docking Algorithms (VEDA), a novel tool for refining X-ray structures into EM electron density. VEDA compares the EM map with a calculated model-based electron density within a selected volume and optimizes the variables that specify the positions of the protein coordinates (rigid bodies) within the macromolecular assembly. Highest quality fit was achieved with a cryo-EM map where the lattice hand had been inverted in MAPMAN (Kleywegt and Jones 1996) which implies that P23-77 follows T=28 *laevo* symmetry rather than *dextro* symmetry. Refinement produced a fit that, in the confines of resolution, shows a high level of accordance with the capsid density (Fig. 7). Atomic models thus confirm that P23-77 capsid is built from single beta-barrel capsid proteins, positioned upright in the shell to form pseudohexagonal capsomers. Capsomer base consists of both proteins and the crenellations are formed from an additional domain on VP17. Capsomer structure of P23-77 is reminiscent of the pseudohexameric capsomers of the double beta-barrel lineage. However, the major difference that P23-77 relies on two single beta-barrel capsid protein species instead of one double beta-barrel, leads to a unique oligomerization pattern.

Two significant observations arose during the capsid modelling process. Firstly, the VP16/VP17 complex fits perfectly well into the density (Fig. 7). Even if VP16 and VP17 chains were submitted to VEDA refinement randomly oriented in the capsid cryo-EM density, the refinement resulted in VP16/VP17 alignment identical to their organization in the complex. This strongly indicates

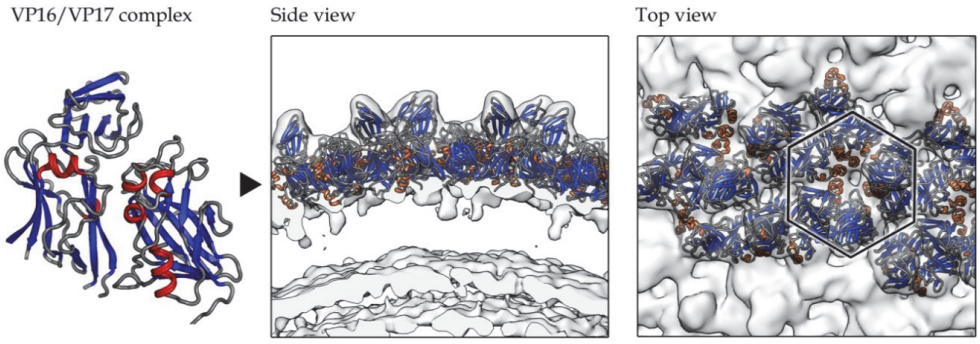


FIGURE 7 P23-77 capsid model generation. Crystal structures of the major capsid proteins were fitted into the cryo-EM density of the P23-77 virion (EMDB id 1525, hand inverted). The VP16/VP17 complex (PDB id 3ZN6), a key building block and refined among other P23-77 major capsid protein crystal structures to the capsid shell density, is shown from the side and from above. In the 'Top view' frame on the right, one hexagonal capsomer is outlined in black.

that the alignment of the two capsid proteins in the complex crystal structure is biologically relevant and represents the heterotypic oligomerization mode in the virion (II). Second, even though each P23-77 capsomer does contain six single beta-barrels, VP16 dimers always span between adjacent capsomers instead of contributing to compact hexameric capsomers. The six jelly rolls in each P23-77 capsomer are oriented similarly to the jelly rolls in the capsomers of double beta-barrel lineage members, yet strong dimerization across capsomer borders is a surprising and unique feature. Considering the oligomers found in the five crystal types reported in this study, it is evident that P23-77 capsid is formed from multiple oligomeric building blocks instead of the trimers seen in the double beta-barrel lineage (Fig. 8). Similar arrangement is likely to be present in other putative members of the single beta-barrel lineage, including archaeal virus SH1.

Capsid building blocks for P23-77, based on subunit alignments detected in the crystal structures, are the VP16/VP17 complex, VP16/VP17 trimer and VP16 dimer. Capsid organization is suggested to reflect an assembly strategy where the assembly, beginning at the 5-fold, would proceed via heterotypic interactions and terminate when assembly frontiers meet at the 3-fold. VP17 termini are proposed to play a role coordinating the assembly, possibly in concert with the internal lipid membrane. This unique interaction network is unusually intricate and merits further analysis (III). To conclude, the P23-77 capsid model provides a unique view of the capsid of a thermostable virus with numerous unique and likely ancient features.

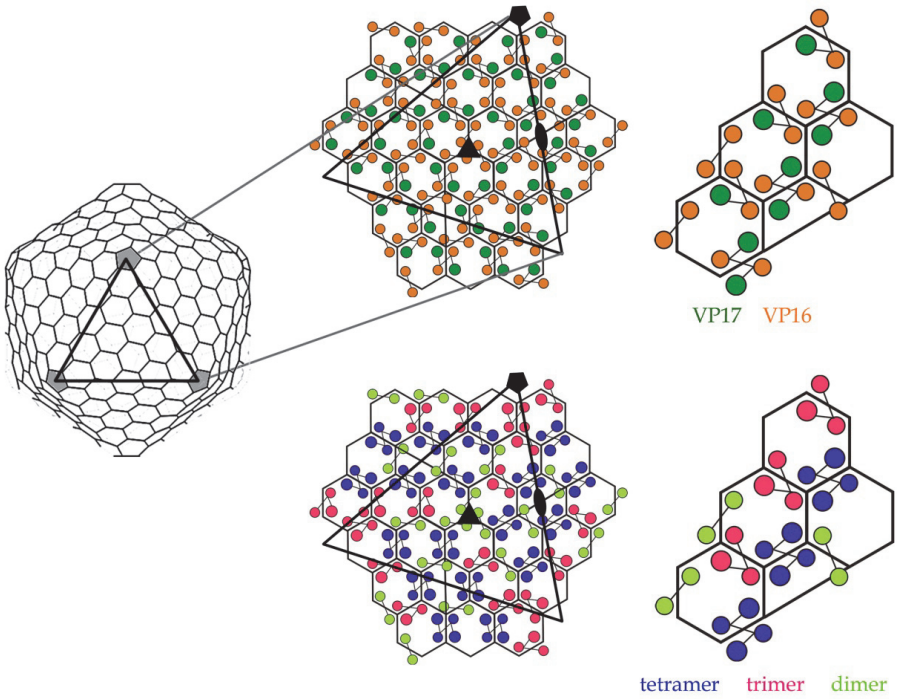


FIGURE 8 P23-77 capsid organization. Icosahedral facets (left and middle) are comprised of asymmetric unit (right). Major capsid proteins are drawn as circles connected by lines to represent oligomeric arrangement. Color scheme illustrates the distribution of the two protein species (top) and the oligomeric building blocks based on crystal structures (bottom). Lattice was created with Icosahedron Generator (http://johnsonlab.scripps.edu/research/icosahedron-generator.php).

5.5 Capsid protein properties and regulation of capsid assembly (III)

The high thermal tolerance (I) and intricate capsid organization (II) illustrate the unique characteristics of the structural components of the P23-77 capsid. Characterization of the melting temperatures (T_m) of VP16 and VP17, their oligomeric state in solution and structural features that influence the stability and interaction properties may provide insight into the dilemma of the thermostable virus assembly. Most viral proteins are flexible and loosely packed, contrasting with the typical thermostable protein; however, this does not include capsid proteins that comprise a unique class lacking parallels in cellular organisms (Tokuriki et al. 2009). By the nature of their function, capsid proteins are relatively rigid. Nevertheless, a level of interaction potential, typically followed by a level of flexibility, is required for capsid assembly. In addition to the experimental methods, the connection between P23-77 and the double beta-barrel viral lineage was exploited to screen for corresponding

features and mechanisms among the well-characterized lineage members. Assembly principles of many lineage members have been analyzed in detail and atomic or near-atomic level virion structures are available for PRD1, PM2 and STIV (Abrescia et al. 2004, Abrescia et al. 2008, Veesler et al. 2013).

The thermodynamic stability of proteins in general depends on the energy gap between the correct fold (native state) and the unfolded and misfolded conformations. Native state can be enforced by attraction forces, *e.g.* van der Waals contacts between hydrophobic residues in the hydrophobic core, and formation of other states discouraged by repulsion forces (Berezovsky et al. 2007). Both P23-77 major capsid proteins are rich in proline and possess well-organized hydrophobic cores, in line with their thermostable nature. The major capsid protein folds are rigid to the point that different crystal types contain nearly identical conformations, including the VP16/VP17 co-crystal. The melting temperatures, determined with differential scanning calorimetry, are +85.9 °C for VP16 and +81.5 °C for VP17 (III). In addition, VP16 was shown to refold when the temperature is lowered after unfolding. Oligomeric states of the two capsid proteins in solution are ~40 kDa dimer (VP16) and ~32 kDa monomer (VP17), based on analytic size exclusion chromatography (II), gradient centrifugation and dynamic light scattering experiments (III).

Electrostatic potentials on the surface of the P23-77 major capsid proteins were modelled with APBS (Baker N.A. et al. 2001). The strand-swap in VP16 dimers is enforced by an electrostatic potential complement. Overall charge distribution of the P23-77 capsomers, based on the published capsid model (II), follows the principle seen in double beta-barrel lineage members: negative charges are oriented distal to membrane, positive towards the negatively charged internal membrane (III). Lateral contacts of VP16 and VP17 subunits in the P23-77 capsomers are enhanced by a charge complement. The characterization confirms that the electrostatic potentials facilitate the correct organization of the P23-77 protein shell: positive charges face the negatively charged internal lipid membrane, and complementary charges drive the formation of stable VP16 dimers and later pseudohexagonal capsomers consisting of both VP16 and VP17.

The P23-77 capsid assembly was previously proposed to be concerted by major capsid protein interactions with each other and the internal lipid membrane; yet various experiments designed to identify major capsid protein interactions failed to identify significant oligomerization between protein species. Since VP17 termini are disordered in the crystal structure and, analogous to the major capsid protein termini of double beta-barrel lineage members, likely to participate in and control major capsid protein oligomerization, it was postulated that they might have autoinhibitory effect on the VP16/VP17 interaction. To study this, truncates of VP17 were prepared where either 18 C-terminal or 20 N-terminal residues were removed. A series of gradient centrifugation and dynamic light scattering experiments were conducted with purified native VP16 and VP17, VP17 truncates and lipid vesicles produced from *T. thermophilus* derived lipids. Even in these conditions, oligomerization was not observed, which led to a search for similar features in related viruses (III).

Major factors involved in capsid assembly in the lineage are the internal lipid membrane, mediating minor capsid proteins and assembly factors and the major capsid proteins. The lack of multimerization between the purified P23-77 major capsid proteins and with the viral membrane strongly indicates that additional assembly factors are required to mediate interactions; in this case, minor capsid protein VP11 seems the most likely candidate. VP11 is the third protein species identified as capsid associated in P23-77; besides this, little is known of the protein. To discern potential mechanisms for a minor capsid protein to facilitate assembly, P23-77 was compared to other lineage members.

Among related viruses from the double beta-barrel lineage, P23-77 shares significant features with archaeal virus STIV: thermophilic nature, lack of major capsid protein multimerization *in vitro* and significant density connecting the capsid and internal lipid membrane at the vertices (Khayat et al. 2005, Veesler et al. 2013). Another candidate is *Pseudoalteromonas* phage PM2, previously considered the most ancient member of the lineage and comprising of major capsid protein P2 which displays highest level of similarity to the folds of P23-77 capsid proteins (Abrescia et al. 2008). For both PM2 and P23-77, empty virions are rarely observed during virion purification, which could indicate very efficient DNA packaging, high instability of the empty particles, or that the virion assembly process does not include an empty procapsid intermediate (Jaatinen et al. 2008). The latter explanation is further discussed in the context of PM2 capsid assembly. Both STIV (T=31) and PM2 (T=21) are icosahedral viruses with internal lipid membrane and capsid shell formed from a double jelly roll major capsid protein arranged in trimeric pseudohexagonal capsomers. Current understanding of STIV and PM2 capsid assembly is largely based on high-resolution X-ray and cryo-EM reconstructions of the virions.

Similar to P23-77, the major capsid proteins of STIV do not readily oligomerize in purified form, and both P23-77 and STIV capsid cryo-EM reconstructions feature prominent density linking the internal membrane to the capsid shell at the icosahedral 5-folds (Khayat et al. 2005). Components of the STIV vertex channel have been determined to high resolution, revealing a ten-stranded beta-pore formed from the C-terminal domain of penton protein A223 together with membrane protein A55 (Veesler et al. 2013). Since the vertex channel is the only density feature discernible between internal membrane and the capsid shell in STIV, it has been proposed to be the main orchestrator of capsid assembly. The penton protein of STIV fits well into the density at the vertices of the P23-77 cryo-EM (Fig. 9), allowing for a hypothesis that the P23-77 could exhibit a similar structure with an assembly coordinating function.

Another prospective assembly pathway for P23-77 is based on PM2. Crystal structure of the PM2 virion has revealed the structures of two helical minor capsid proteins, P3 and P6, located between the capsid shell and the internal lipid membrane (Abrescia et al. 2008). Assembly of PM2 is suggested to begin when P3 dimers and P6 monomers anchor into the host membrane and form a scaffold which defines the dimensions of the forming capsid and extracts

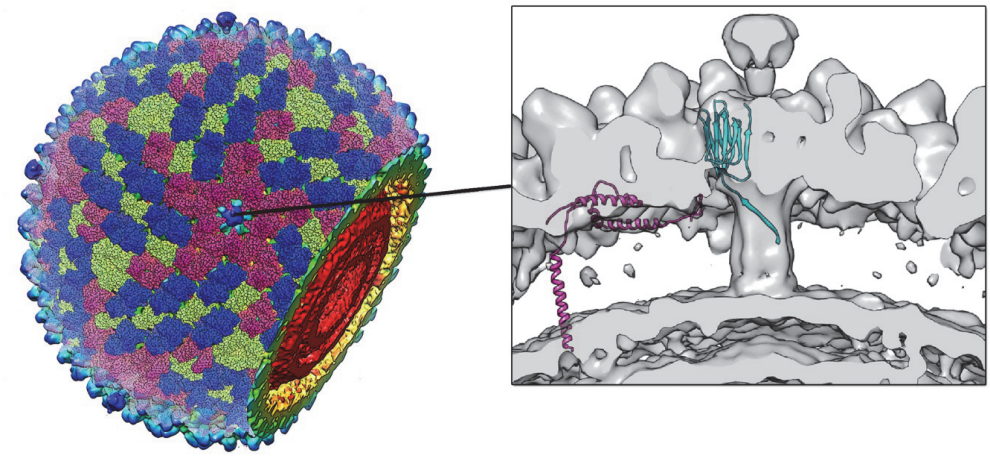


FIGURE 9 Capsid model of P23-77 with hypothetical assembly factors. The model illustrates the nearly spherical capsid, capsid shell built from major capsid proteins colored according to oligomeric building block (II and Fig. 8), and shows an intersection of the capsid density (EMDB id 1525, hand inverted) (left). Minor capsid protein P3 from PM2 (purple, PDB id 2W0C) and STIV penton A223 (cyan, PDB id 3J31) are illustrated on the right, representing the positions of hypothetical assembly factors in P23-77 capsid. For clarity, the capsid density where the assembly factors are fitted is shown in grey without the major capsid proteins.

the vesicle from host membrane during P6 association to pre-condensed viral DNA. Capsid assembly is completed by the major capsid protein P2 shell assembly on top of vesicle which bears both the genome and the P3-P6 scaffold. P2 assembly relies on calcium ion assisted interaction between planar triangles of P3 and trimers of P2. The described assembly mechanism omits the procapsid stage, explaining the rarity of empty virus particles in preparations.

Certain P23-77 features seem contrary to both assembly mechanisms, including the fact that PM2-like strategy produces a lipid core vesicle much more robust than what is observed in P23-77, and STIV penton has high prevalence of beta-sheets, which is not predicted for VP11. It remains to be seen whether P23-77 assembly process resembles more the PM2-like minor capsid protein mediated assembly or vertex-orchestrated STIV-like assembly. Both possibilities require minor capsid proteins that facilitates and coordinate interactions (Fig. 9). Further studies of P23-77 minor capsid protein VP11, among other potential assembly factors, are required to resolve the issue. Characteristics of the thermostable (T_m >+80°C) major capsid proteins and the capsid model refine the understanding of the archetype of the single beta-barrel lineage: besides the jelly roll fold, members are likely to contain negatively charged internal lipid membrane, capsid electrostatic potentials distributed so that negative charges are oriented distal to membrane, and possibly lack heterotypic oligomerization *in vitro* and require a trigger to initiate capsid assembly.

5.6 P23-77 and the current virus taxonomy (IV)

Previously described studies confirm the connection between P23-77 and the double beta-barrel lineage (II). However, prevalent taxonomy efforts are based on relationships detected at the sequence level; therefore, structural virus lineages are currently not included in ICTV taxonomy. In this study, genomic features were identified that associate bacteriophage P23-77 and related viruses to recently proposed viral family *Sphaerolipoviridae* (IV) (Table 3). The family includes most of the prospective single beta-barrel lineage members, indicating a level where structural and genomic evidence can be combined to establish novel taxonomy.

Viruses of Archaea and Bacteria are so far removed from each other that the current taxonomy only recognizes one viral group (order *Caudovirales*) that includes members infecting hosts from both domains. Recently, another proposal was submitted to ICTV designating a novel group of archaeal viruses with a potential connection to bacteriophages: family *Sphaerolipoviridae* (proposal by Dyall-Smith, Porter and Tang, 2013, http://talk.ictvonline.org/files/proposals/taxonomy_proposals_prokaryote1/m/bact01/4633.aspx). *Sphaerolipoviridae* comprise icosahedral, tailles haloarchaeal viruses originally assigned to two genera, *Alphasphaerolipovirus* and *Betasphaerolipovirus*. Members of the family have icosahedrally ordered, near-spherical capsids encompassing an internal lipid membrane and a dsDNA genome. Intriguingly, the type virus of genus *Alphasphaerolipovirus* is *H. hispanica* virus SH1, previously associated with bacteriophage P23-77 by strong morphological similarity (Jaalinoja et al. 2008, Jaatinen et al. 2008). Both viruses utilize two major capsid protein species and exhibit T=28 capsid organization with crenellated topology indicative of single beta-barrel capsid protein fold. In addition to SH1, *Alphasphaerolipovirus* includes two other *H. hispanica* viruses, HHIV-2 and PH1, which have highly similar genomes regarding sequence homology and gene synteny.

Recently isolated SH1-like membrane containing halovirus, *Natrinema* sp. J7-1 virus SNJ1, is currently designated the sole member of the genus *Betasphaerolipovirus*. SNJ1 shares several genes with the members of *Alphasphaerolipovirus*, including putative ORFs for the two major capsid proteins and the predicted genome packaging ATPase. As a result of the lysogenic lifecycle, SNJ1 genome was originally reported as plasmid pHH205 (Ye et al. 2003). Plasmid pHH205 was found, by sequence comparison, to be related to P23-77 (Jalasvuori et al. 2009). Besides the viruses described above, proviruses and genetic elements related to the sphaerolipoviruses have been found integrated into the genomes of diverse haloarchaea, where they are typically found adjacent to tRNA genes (Jalasvuori et al. 2009, Jalasvuori et al. 2010).

Recent isolation and characterization of a group of bacteriophages infecting bacterial genus *Thermus* has revealed that these viruses exhibit a number of features highly reminiscent of archaeal sphaerolipoviruses. This group includes *T. thermophilus* phages P23-77, P23-72 and P23-65H that were isolated from hot

springs in New Zealand (Yu et al. 2006) and *T. aquaticus* phage IN93, isolated from hot-spring soil in Japan (Matsushita and Yanase 2009). Since P23-77 has been extensively characterized, it was chosen as the type member of the group. In addition to previously described morphological similarities with SH1, P23-77 genome carries a block of core genes that encode a putative genome-packaging ATPase and the two major capsid proteins ('viral self' features). This conserved block of genes and their organization in the genome is common to all sphaerolipoviruses. Moreover, all sphaerolipoviruses infect hosts from extreme environments and have relatively narrow host range. Similar capsid architecture found across the family implies that many members could belong to the structural single beta-barrel lineage, an ancient branch of viruses suggested to prevail in extreme environments. These findings suggest that P23-77-like viruses should be designated to family *Sphaerolipoviridae* and establish within the family a novel genus titled *Gammasphaerolipovirus* (Table 3).

TABLE 3 Putative genera of virus family *Sphaerolipoviridae*, not including proviruses. Type virus of each genus is underlined.

Genus	Members	Host (host domain)	Type
Alphasphaerolipovirus	SH1, HHIV-2, PH1	*Haloarcula hispanica* (Archaea)	Halophile
Betasphaerolipovirus	SNJ1	*Natrinema* sp J7-1 (Archaea)	Halophile
Gammasphaerolipovirus	P23-77, IN93	*Thermus thermophilus* (Bacteria)	Thermophile

To conclude, genomes of P23-77 and *Sphaerolipoviridae* members all share a conserved block of three ORFs that encode predicted major capsid proteins and genome-packaging ATPase, arranged similarly within their genomes. Besides structural features, this arrangement comprises one of the features considered a hallmark of the single beta-barrel lineage. Incorporating the recent P23-77 findings to the current virus taxonomy has lead us to propose that the so far unclassified P23-77 belongs to *Sphaerolipoviridae* and together with a few thermophilic bacteriophages forms a novel genus, *Gammasphaerolipovirus*. Consequently, the family includes viruses infecting both prokaryotic domains and establishes a new link between bacterial and archaeal virospheres.

6 CONCLUSIVE REMARKS

The main conclusions in this thesis are as follows:

I The major capsid proteins of thermophilic bacteriophage P23-77, namely VP16 and VP17, are effectively expressed in *E. coli* HMS174DE(3) cells and purification protocols have been established for full-length, native proteins. VP16 and VP17 were found to have high thermal tolerance. Both proteins and purified P23-77 virus samples were screened for crystallization and yielded five crystal types: two types of VP16, one type of VP17, one putative co-crystal comprising VP16 and VP17, and one crystal type derived from purified P23-77 virion material. Initial diffraction experiments established that all crystal types were well-diffracting with maximum resolution between 1.26 - 2.92 Å.

II The structures of all five crystal types described in the previous study were determined. Consistent with the proposed hypothesis, VP16 and VP17 exhibit a single beta-barrel core fold that is homologous to individual beta-barrels found in the major capsid proteins of the double beta-barrel lineage members. Thus, P23-77 represents an archetype of a novel viral lineage, the single beta-barrel lineage, which shares a common ancestor with the double beta-barrel lineage. In addition to the conserved beta-barrel fold, the P23-77 major capsid proteins display certain unique features: VP16 form strand-swapped dimers and VP17 bears an additional seven-stranded beta-barrel domain. A model of the P23-77 capsid shell, generated by fitting the major capsid protein structures into the virion electron density from cryo-EM, revealed that these features affect capsid organization. Pseudohexagonal capsomers are built from VP16 and lower domains of VP17 while the VP17 upper domains comprise the characteristic turrets observed in the cryo-EM. Capsid organization seems more intricate than is typical for the double beta-barrel lineage; instead of compact trimers, the building blocks for P23-77 capsid are various oligomers with strong interactions spanning across capsomer borders.

Differences in the distribution of the two major capsid proteins imply that the assembly may begin at the 5-fold. Unique features of P23-77 may be explained by the intriguing notion that, since extreme environments have reported to convey reduced mutation rates, the extremophile P23-77 may be closest in form to the ancient common ancestor of the single and double beta-barrel lineages.

III The extent of VP16 and VP17 thermostability was characterized and T_m-values were determined to be +85.9 °C and +81.5 °C, respectively. Purified native proteins as well as VP17 terminal truncates were screened for solution oligomeric state and interaction potential; in accordance with the crystal structures, results from various analysis confirmed that purified recombinant VP16 is dimeric and VP17 exists as monomers. Despite extensive screening, larger multimers or heterotypic interactions were not observed, which suggests that P23-77 capsid assembly requires an additional agent to initialize and possibly mediate the assembly process. The trigger is proposed to be a minor capsid protein with positioning and function possibly similar to the STIV penton or the PM2 minor capsid proteins P3 and P6.

IV P23-77 genome includes three specific genes, namely two major capsid proteins and a putative viral ATPase, which have been identified in similar organization in the genomes of viruses recently proposed to form a novel virus family, *Sphaerolipoviridae*. Morphological and genetic similarities were analyzed in detail and the results indicate that P23-77 and certain related viruses belong to sphaerolipoviruses and comprise their own novel genus, *Gammasphaerolipovirus*. Gammasphaerolipoviruses infect bacteria of genus *Thermus*, and like all sphaerolipoviruses, inhabit extreme environments. Sphaerolipoviruses assemble putative single beta-barrel lineage members into a group established by genetic evidence, creating an association between structural lineages and current virus taxonomy.

Acknowledgements

This work was carried out during 2010-2014 at the University of Jyväskylä, at the Department of Biological and Environmental Science and Nanoscience Center, and funded by the Finnish Centre of Excellence (CoE) Program of the Academy of Finland 2006–2011 CoE in Virus Research (#1129648) and CoE in Biological Interactions 2012-2017 (#252411), and an Academy of Finland personal grant to J.K.H.B. (#251106).

I am utmostly grateful to my supervisor Professor Jaana Bamford for all the great opportunities and trust she has given me. Under her guidance I have gained the skills that I hope will allow me to make 'good science'. I also wish to express my deepest gratitude to Dr Nicola Abrescia for kindly agreeing to act as my opponent. His work in the field of virology has provided me with great inspiration. I would like to thank the reviewers of this thesis, Dr Roman Tuma and Dr Tomi Airenne, for their insightful and constructive comments. My thesis committee members, Professor Jari Ylänne and Dr Vesa Hytönen are greatly acknowledged for guidance and advice over the years. Also, this thesis would not have been possible without the expert contributions of Professor David Stuart, Dr Jonathan Grimes and Dr Karl Harlos, who are warmly acknowledged.

I want to thank everybody in and affiliated with the JB group: first of all Alice, who taught me to handle proteins in the lab and supervised my master's thesis project that subsequently evolved into this dissertation; Petri for highly appreciated help with protein purification; the adjunct professors/post docs Matti, Lotta-Riina and Nadine, who always give such interesting talks and have tons of knowledge about everything; and off course my fellow PhD students Jenni, Elina, Sari, Sari, Reetta, Anni, Ville, Hanna, and everybody else in our diverging group for a nice working environment. Thanks for the floorball Fridays and remembering to invite me to group outings even when I've been cloistered in my room far away at the other end of the corridor. Also, I want to thank all my friends, old and new, for good times and refreshing conversations, especially Saana for insights and just the right kind of humor. I have been lucky to belong to an excellent graduate school and thank the National Doctoral Program in Informational and Structural Biology (ISB) and the ISB people for great times and interesting science.

Undoubtedly, the reason that me and all my siblings are drawn to science are my parents, who receive my warmest thanks for everything. I think the atmosphere at home has always been mental(ly stimulating), morphing through the times from my mother pointing out different bird species to pre-school kids to now having dinners turn into full-fledged interdisciplinary debates. Finally, I thank my dear boyfriend and teammate against the world Ville who has had the pleasure of following the progress of my work from close distance and has been very gracious about it.

YHTEENVETO (RÉSUMÉ IN FINNISH)

Muinainen virustyyppi ääriolosuhteista

Virukset ovat biologisia järjestelmiä, jotka hyödyntävät solullisia eliöitä lisääntyäkseen. Tyypillinen viruspartikkeli koostuu perimästä (genomista) ja sitä ympäröivästä proteiinikuoresta, johon voi kuulua myös lipidikalvoja. Virus ei kykene aineenvaihduntaan ja sitä voikin ajatella perimänä, joka siirtyy isäntäsoluun, monistuu solun mekanismeja hyväksikäyttäen, paketoi itsensä partikkelin sisään ja siirtyy seuraavaan isäntään aloittaakseen uuden monistumiskierron. Virusten alkuperää ei tunneta, joskin vallitsevien teorioiden mukaan niiden synty ajoittuu samalle muinaiselle aikakaudelle kuin solujen. Solulliset eliöt ovat aina eläneet jatkuvassa virusten aiheuttamassa paineessa ja virukset ovat yksi merkittävimmistä elämän kehitykseen vaikuttaneista tekijöistä. Vaikka viruksia ei aineenvaihdunnan puuttuessa luokitella elollisiksi, myös virukset kehittyvät evoluution myötä: monistuviin virusgenomeihin syntyy mutaatioita, joiden mahdolliset hyödylliset tai haitalliset vaikutukset vaikuttavat mutaation leviämiseen populaatiossa. Toista merkittävää evolutiivista mekanismia viruksissa kutsutaan mosaikismiksi, joka viittaa virusten kykyyn yhdistellä perimäelementtejä ja näin tuottaa uusia geenejä. Näistä mekanismeista aiheutuvat virusten huomattava sopeutumiskyky sekä solullisia organismeja nopeampi perimän erilaistuminen.

Solullisen elämän sukupuu on kyetty määrittämään geneettisiä sekvenssejä vertailemalla, mutta virusten tapauksessa nopea perimän erilaistuminen vähentää sekvenssianalyysien tehoa kaukaisempia sukulaisuussuhteita määritettäessä. Ongelma voidaan ratkaista tutkimalla virusten genomien sijaan tiettyjä voimakkaasti evoluutiossa säilyneitä (konservoituneita) proteiinirakenteita. Mitä merkittävämpi rakenne on viruksen selviytymiselle, sitä todennäköisemmin sen laskostuminen on konservoitunut: vaikka rakennetta koodaava geneettinen sekvenssi mutatoituu, on sen tuottaman proteiinin yhä kyettävä laskostumaan toiminnalliseen muotoon. Tärkeimpiin virusta määritteleviin ominaisuuksiin kuuluu sen proteiinikuori eli kapsidi. Viime vuosikymmeninä on havaittu, että tiettyjen konservoituneiden kuoriproteiinirakenteiden perusteella voidaan määrittää virussukulinjoja. Näistä tutkituimpiin kuuluu kaksois-beta-tynnyrilinja (double beta-barrel lineage). Sukulinjaan kuuluvat muun muassa bakteriofagi PRD1, arkkivirus STIV ja aitotumallisia infektoivat adenovirukset. Nämä virukset ovat geneettisiltä sekvensseiltään erilaisia ja lisääntyvät täysin erilaisissa isäntäorganismeissa, mutta niiden kuoriproteiineissa on havaittavissa lähes identtinen kaksois-beta-tynnyrirakenne. Koska kaksois-beta-tynnyrirakenne löytyy eri eliökunnan domeeneja infektoivista viruksista, on esitetty, että sukulinjan esi-isä on esiintynyt ajalla ennen eliökunnan domeenien eroamista toisistaan. Näin ollen sukulinjan historia ylettyy miljardien vuosien päähän, mikä tekee siitä mielenkiintoisen tutkimuskohteen elämän ja virusten synnyn tutkimuksen kannalta.

Vuonna 2005 Uuden-Seelannin kuumista lähteistä eristetty virus, bakteriofagi P23-77, ilmentää tiettyjä kaksois-beta-tynnyrisukulinjalle ominaisia morfologisia ja geneettisiä piirteitä, joiden perusteella viruksen on arvioitu olevan suvun aiemmin tuntematon jäsen. Tässä väitöskirjassa raportoidaan bakteriofagin P23-77 kahden kuoriproteiinin rakenteet, joiden perusteella määritetään viruksen sukulaisuussuhteet ja kapsidin rakentuminen. Lisäksi karakterisoidaan kuoriproteiinien ominaisuudet ja käsitellään mekanismeja, jotka johtavat viruspartikkelin kokoamiseen solussa. Väitoskirjassa esitellään edellä mainittuihin löydöksiin ja geneettisiin piirteisiin perustuva, virusten luokittelusta vastaavalle kansainväliselle komitealle toimitettu ehdotus uudesta virussuvusta nimeltä *Gammasphaerolipovirus.* Seuraava kappale sisältää yhteenvedon merkittävimmistä tuloksista.

Bakteriofagin P23-77 kapsidi muodostuu kahdesta kuoriproteiinista, VP16 ja VP17, joille kehitettiin tuotto-, puhdistus- ja kiteytyprotokollat ja joiden rakenteet ratkaistiin röntgensädekristallografialla. Ratkaistut kuoriproteiinirakenteet edustavat tyyppiä, jota ei ole aiemmin havaittu: ydinrakenteena molemmissa kuoriproteiineissa on yksittäinen beta-tynnyri, jonka rakennehomologia kaksois-beta-tynnyrivirusten kuoriproteiineihin on huomattava. Lisäksi esitetystä kapsidimallista havaitaan, että P23-77:n proteiinikuori muodostuu monimutkaisesta multimeerien yhdistelmästä, joka poikkeaa kaksois-beta-tynnyrivirusten suhteellisen yksinkertaisesta trimeereihin pohjautuvasta mallista. Rakennetutkimus osoittaa, että P23-77 edustaa muinaista haaraa vertikaalisten beta-tynnyrivirusten suvussa ja on säilyttänyt piirteitä, jotka ovat todennäköisesti lähimpänä sukulinjan muinaista esi-isää. Kuoriproteiinien karakterisointi paljasti niiden olevan erittäin lämpöstabiileja, mutta vuorovaikuttavan keskenään yllättävän heikosti. Tämä viittaa siihen, että kapsidin rakentumisen aloittaminen vaatii ylimääräisen tekijän, todennäköisesti avustavan proteiinin. Tiettyjen arkkeja infektoivien ääriolosuhteiden virusten esitetään kuuluvat samaan virustyyppiin kuin P23-77. Morfologiset ja geneettiset piirteet, erityisesti kolme ydingeeniä (kaksi kuoriproteiinia ja ATPaasi), jotka esiintyvät näiden virusten genomeissa, yhdistävät bakteriofaagi P23-77:n esitettyyn heimoon *Sphaerolipoviridae*, jonka sisällä P23-77:n piirteet määrittelevät uuden suvun, *Gammasphaerolipovirus*. P23-77 on ensimmäinen näistä ääriolosuhteiden viruksista, jota on tutkittu atomitason rakenteiden tarkkuudella ja jonka kuoren rakentuminen tunnetaan. P23-77 tarjoaa harvinaisen mallin ääriolosuhteiden viruksesta ja ikkunan muinaisten virusten maailmaan.

REFERENCES

Aalto A.P., Bitto D., Ravantti J.J., Bamford D.H., Huiskonen J.T. & Oksanen H.M. 2012. 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.

Abad-Zapatero C., Abdel-Meguid S.S., Johnson J.E., Leslie A.G., Rayment I., Rossmann M.G., Suck D. & Tsukihara T. 1980. Structure of southern bean mosaic virus at 2.8 A resolution. *Nature* 286: 33-39.

Abrescia N.G., Bamford D.H., Grimes J.M. & Stuart D.I. 2012. Structure unifies the viral universe. *Annu Rev Biochem* 81: 795-822.

Abrescia N.G., Grimes J.M., Kivela H.M., Assenberg R., Sutton G.C., Butcher S.J., Bamford J.K., Bamford D.H. & Stuart D.I. 2008. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol Cell* 31: 749-761.

Abrescia N.G., Cockburn J.J., Grimes J.M., Sutton G.C., Diprose J.M., Butcher S.J., Fuller S.D., San Martin C., Burnett R.M., Stuart D.I., Bamford D.H. & Bamford J.K. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432: 68-74.

Ackermann H.W. 2007. 5500 Phages examined in the electron microscope. *Arch Virol* 152: 227-243.

Adams P.D., Afonine P.V., Bunkoczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C. & Zwart P.H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-221.

Arslan D., Legendre M., Seltzer V., Abergel C. & Claverie J.M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci U S A* 108: 17486-17491.

Artenstein A.W. 2012. The discovery of viruses: advancing science and medicine by challenging dogma. *Int J Infect Dis* 16: e470-473.

Asial I., Cheng Y.X., Engman H., Dollhopf M., Wu B., Nordlund P. & Cornvik T. 2013. Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. *Nat Commun* 4: 2901.

Athappilly F.K., Murali R., Rux J.J., Cai Z. & Burnett R.M. 1994. The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9 A resolution. *J Mol Biol* 242: 430-455.

Bahar M.W., Graham S.C., Stuart D.I. & Grimes J.M. 2011. Insights into the evolution of a complex virus from the crystal structure of vaccinia virus D13. *Structure* 19: 1011-1020.

Baker M.L., Jiang W., Rixon F.J. & Chiu W. 2005. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol* 79: 14967-14970.

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

Baltimore D. 1971a. Viral genetic systems. *Trans N Y Acad Sci* 33: 327-332.

Baltimore D. 1971b. Expression of animal virus genomes. *Bacteriol Rev* 35: 235-241.

Bamford D.H. 2003. Do viruses form lineages across different domains of life? *Res Microbiol* 154: 231-236.

Bamford D.H., Burnett R.M. & Stuart D.I. 2002. Evolution of viral structure. *Theor Popul Biol* 61: 461-470.

Battistuzzi F.U., Feijao A. & Hedges S.B. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* 4: 44.

Benitez-Alfonso Y., Faulkner C., Ritzenthaler C. & Maule A.J. 2010. Plasmodesmata: gateways to local and systemic virus infection. *Mol Plant Microbe Interact* 23: 1403-1412.

Benson S.D., Bamford J.K., Bamford D.H. & Burnett R.M. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 98: 825-833.

Berezovsky I.N., Zeldovich K.B. & Shakhnovich E.I. 2007. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput Biol* 3: e52.

Bergh O., Borsheim K.Y., Bratbak G. & Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature* 340: 467-468.

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.

Brenner S., Jacob F. & Meselson M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190: 576-581.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P., Sharff A., Smart O.S., Vonrhein C. & Womack T.O. 2011. *BUSTER*. Global Phasing Ltd., Cambridge, UK.

Butcher S.J., Manole V. & Karhu N.J. 2012. Lipid-containing viruses: bacteriophage PRD1 assembly. *Adv Exp Med Biol* 726: 365-377.

Canchaya C., Fournous G., Chibani-Chennoufi S., Dillmann M.L. & Brussow H. 2003. Phage as agents of lateral gene transfer. *Curr Opin Microbiol* 6: 417-424.

Caspar D.L. & Klug A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27: 1-24.

Claverie J.M., Ogata H., Audic S., Abergel C., Suhre K. & Fournier P.E. 2006. Mimivirus and the emerging concept of "giant" virus. *Virus Res* 117: 133-144.

Condit R.C., Moussatche N. & Traktman P. 2006. In a nutshell: structure and assembly of the vaccinia virion. *Adv Virus Res* 66: 31-124.

Daubin V. & Ochman H. 2004. Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli. *Genome Res* 14: 1036-1042.

Dawe A.L. & Nuss D.L. 2001. Hypoviruses and chestnut blight: exploiting viruses to understand and modulate fungal pathogenesis. *Annu Rev Genet* 35: 1-29.

Desnues C., Boyer M. & Raoult D. 2012. Sputnik, a virophage infecting the viral domain of life. *Adv Virus Res* 82: 63-89.

Desplats C. & Krisch H.M. 2003. The diversity and evolution of the T4-type bacteriophages. *Res Microbiol* 154: 259-267.

Dolja V.V. & Krupovic M. 2013. Accelerating expansion of the viral universe. *Curr Opin Virol* 3: 542-545.

Douglas T. & Young M. 2006. Viruses: making friends with old foes. *Science* 312: 873-875.

Drake J.W. 2009. Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS Genet* 5: e1000520.

Drake J.W., Charlesworth B., Charlesworth D. & Crow J.F. 1998. Rates of spontaneous mutation. *Genetics* 148: 1667-1686.

Duda R.L. 1998. Protein chainmail: catenated protein in viral capsids. *Cell* 94: 55-60.

Duda R.L., Hempel J., Michel H., Shabanowitz J., Hunt D. & Hendrix R.W. 1995. Structural transitions during bacteriophage HK97 head assembly. *J Mol Biol* 247: 618-635.

Dunning Hotopp J.C., Clark M.E., Oliveira D.C., Foster J.M., Fischer P., Munoz Torres M.C., Giebel J.D., Kumar N., Ishmael N., Wang S., Ingram J., Nene R.V., Shepard J., Tomkins J., Richards S., Spiro D.J., Ghedin E., Slatko B.E., Tettelin H. & Werren J.H. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317: 1753-1756.

Emsley P. & Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-2132.

Feschotte C. 2010. Virology: Bornavirus enters the genome. *Nature* 463: 39-40.

Forterre P. & Prangishvili D. 2009. The origin of viruses. *Res Microbiol* 160: 466-472.

Friedman R., Drake J.W. & Hughes A.L. 2004. Genome-wide patterns of nucleotide substitution reveal stringent functional constraints on the protein sequences of thermophiles. *Genetics* 167: 1507-1512.

Gordon-Shaag A., Ben-Nun-Shaul O., Roitman V., Yosef Y. & Oppenheim A. 2002. Cellular transcription factor Sp1 recruits simian virus 40 capsid proteins to the viral packaging signal, ses. *J Virol* 76: 5915-5924.

Grimes J.M., Burroughs J.N., Gouet P., Diprose J.M., Malby R., Zientara S., Mertens P.P. & Stuart D.I. 1998. The atomic structure of the bluetongue virus core. *Nature* 395: 470-478.

Hacene H., Rafa F., Chebhouni N., Boutaiba S., Bhatnagar T., Barratti J.C. & Ollivier B. 2004. Biodiversity of prokaryotic microflora in E1 Golea Salt lake, Algerian Sahara. *Journal of Arid Environments* 58: 273-284.

Hall J.P., Harrison E. & Brockhurst M.A. 2013. Viral host-adaptation: insights from evolution experiments with phages. *Curr Opin Virol* 3: 572-577.

Haring M., Vestergaard G., Rachel R., Chen L., Garrett R.A. & Prangishvili D. 2005. Virology: independent virus development outside a host. *Nature* 436: 1101-1102.

Hatherley D., Graham S.C., Turner J., Harlos K., Stuart D.I. & Barclay A.N. 2008. Paired receptor specificity explained by structures of signal regulatory proteins alone and complexed with CD47. *Mol Cell* 31: 266-277.

Hedges S.B., Chen H., Kumar S., Wang D.Y., Thompson A.S. & Watanabe H. 2001. A genomic timescale for the origin of eukaryotes. *BMC Evol Biol* 1: 4.

Helgstrand C., Wikoff W.R., Duda R.L., Hendrix R.W., Johnson J.E. & Liljas L. 2003. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. *J Mol Biol* 334: 885-899.

Hendrix R.W. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol* 61: 471-480.

Hendrix R.W. 2003. Bacteriophage genomics. *Curr Opin Microbiol* 6: 506-511.

Hendrix R.W., Lawrence J.G., Hatfull G.F. & Casjens S. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol* 8: 504-508.

Hendrix R.W., Smith M.C., Burns R.N., Ford M.E. & Hatfull G.F. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* 96: 2192-2197.

Henne A., Bruggemann H., Raasch C., Wiezer A., Hartsch T., Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R., Jacobi C., Starkuviene V., Schlenczeck S., Dencker S., Huber R., Klenk H.P., Kramer W., Merkl R., Gottschalk G. & Fritz H.J. 2004. The genome sequence of the extreme thermophile Thermus thermophilus. *Nat Biotechnol* 22: 547-553.

Hershey A.D. & Chase M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 36: 39-56.

Heuser J. 2005. Deep-etch EM reveals that the early poxvirus envelope is a single membrane bilayer stabilized by a geodetic "honeycomb" surface coat. *J Cell Biol* 169: 269-283.

Hirth L. & Richards K.E. 1981. Tobacco mosaic virus: model for structure and function of a simple virus. *Adv Virus Res* 26: 145-199.

Holm L. & Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233: 123-138.

Holmes E.C. 2011. What does virus evolution tell us about virus origins? *J Virol* 85: 5247-5251.

Huiskonen J.T., de Haas F., Bubeck D., Bamford D.H., Fuller S.D. & Butcher S.J. 2006. Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* 14: 1039-1048.

Iyer L.M., Aravind L. & Koonin E.V. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* 75: 11720-11734.

Iyer L.M., Balaji S., Koonin E.V. & Aravind L. 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res* 117: 156-184.

Jaakkola S.T., Penttinen R.K., Vilen S.T., Jalasvuori M., Ronnholm G., Bamford J.K., Bamford D.H. & Oksanen H.M. 2012. Closely related archaeal Haloarcula hispanica icosahedral viruses HHIV-2 and SH1 have nonhomologous genes encoding host recognition functions. *J Virol* 86: 4734-4742.

Jaalinoja H.T., Roine E., Laurinmaki P., Kivela H.M., Bamford D.H. & Butcher S.J. 2008. Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus. *Proc Natl Acad Sci U S A* 105: 8008-8013.

Jaatinen S.T., Happonen L.J., Laurinmaki P., Butcher S.J. & Bamford D.H. 2008. Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic Thermus thermophilus. *Virology* 379: 10-19.

Jalasvuori M., Pawlowski A. & Bamford J.K. 2010. A unique group of virus-related, genome-integrating elements found solely in the bacterial family Thermaceae and the archaeal family Halobacteriaceae. *J Bacteriol* 192: 3231-3234.

Jalasvuori M., Jaatinen S.T., Laurinavicius S., Ahola-Iivarinen E., Kalkkinen N., Bamford D.H. & Bamford J.K. 2009. The closest relatives of icosahedral viruses of thermophilic bacteria are among viruses and plasmids of the halophilic archaea. *J Virol* 83: 9388-9397.

Johnson J.E. & Speir J.A. 1997. Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol* 269: 665-675.

Juhala R.J., Ford M.E., Duda R.L., Youlton A., Hatfull G.F. & Hendrix R.W. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* 299: 27-51.

Kabsch W. 1993. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *Journal of Applied Crystallography* 26: 795-800.

Khayat R., Fu C.Y., Ortmann A.C., Young M.J. & Johnson J.E. 2010. The architecture and chemical stability of the archaeal Sulfolobus turreted icosahedral virus. *J Virol* 84: 9575-9583.

Khayat R., Tang L., Larson E.T., Lawrence C.M., Young M. & Johnson J.E. 2005. Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc Natl Acad Sci U S A* 102: 18944-18949.

Kleywegt G.J. & Jones T.A. 1996. xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr D Biol Crystallogr* 52: 826-828.

Koonin E.V. & Dolja V.V. 2013. A virocentric perspective on the evolution of life. *Curr Opin Virol* 3: 546-557.

Koonin E.V., Senkevich T.G. & Dolja V.V. 2006. The ancient Virus World and evolution of cells. *Biol Direct* 1: 29.

Krissinel E. & Henrick K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60: 2256-2268.

Krupovic M. & Bamford D.H. 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol* 6: 941-948.

Krupovic M. & Bamford D.H. 2010. Order to the viral universe. *J Virol* 84: 12476-12479.

Krupovic M. & Bamford D.H. 2011. Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. *Curr Opin Virol* 1: 118-124.

Laanto E., Bamford J.K., Laakso J. & Sundberg L.R. 2012. Phage-driven loss of virulence in a fish pathogenic bacterium. *PLoS One* 7: e53157.

Lane S.W., Dennis C.A., Lane C.L., Trinh C.H., Rizkallah P.J., Stockley P.G. & Phillips S.E. 2011. Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413: 41-50.

Lang A.S., Zhaxybayeva O. & Beatty J.T. 2012. Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Microbiol* 10: 472-482.

Lima-Mendez G., Van Helden J., Toussaint A. & Leplae R. 2008. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol Biol Evol* 25: 762-777.

Liu Y., Xu L., Opalka N., Kappler J., Shu H.B. & Zhang G. 2002. Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* 108: 383-394.

Lwoff A., Horne R. & Tournier P. 1962. A system of viruses. *Cold Spring Harb Symp Quant Biol* 27: 51-55.

Mateu M.G. 2013. Assembly, stability and dynamics of virus capsids. *Arch Biochem Biophys* 531: 65-79.

Matsushita I. & Yanase H. 2009. The genomic structure of thermus bacteriophage {phi}IN93. *J Biochem* 146: 775-785.

McClain B., Settembre E., Temple B.R., Bellamy A.R. & Harrison S.C. 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 A resolution. *J Mol Biol* 397: 587-599.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C. & Read R.J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* 40: 658-674.

Merckel M.C., Huiskonen J.T., Bamford D.H., Goldman A. & Tuma R. 2005. The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* 18: 161-170.

Naitow H., Tang J., Canady M., Wickner R.B. & Johnson J.E. 2002. L-A virus at 3.4 A resolution reveals particle architecture and mRNA decapping mechanism. *Nat Struct Biol* 9: 725-728.

Nandhagopal N., Simpson A.A., Gurnon J.R., Yan X., Baker T.S., Graves M.V., Van Etten J.L. & Rossmann M.G. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci U S A* 99: 14758-14763.

Oren A. 2002. Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol Ecol* 39: 1-7.

Oshima T. 1974. [Comparative studies on biochemical properties of an extreme thermophile, Thermus thermophilus HB 8 (author's transl)]. *Seikagaku* 46: 887-907.

Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C., Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R., Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S., Kriakov J., Lawrence J.G., Jacobs W.R., Jr., Hendrix R.W. & Hatfull G.F. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* 113: 171-182.

Perrakis A., Morris R. & Lamzin V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6: 458-463.

Philippe N., Legendre M., Doutre G., Coute Y., Poirot O., Lescot M., Arslan D., Seltzer V., Bertaux L., Bruley C., Garin J., Claverie J.M. & Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341: 281-286.

Pietila M.K., Atanasova N.S., Manole V., Liljeroos L., Butcher S.J., Oksanen H.M. & Bamford D.H. 2012. Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *J Virol* 86: 5067-5079.

Prangishvili D. 2003. Evolutionary insights from studies on viruses of hyperthermophilic archaea. *Res Microbiol* 154: 289-294.

Prangishvili D. & Garrett R.A. 2005. Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol* 13: 535-542.

Pringle C.R. 1998. The universal system of virus taxonomy of the International Committee on Virus Taxonomy (ICTV), including new proposals ratified since publication of the Sixth ICTV Report in 1995. *Arch Virol* 143: 203-210.

Raoult D., Audic S., Robert C., Abergel C., Renesto P., Ogata H., La Scola B., Suzan M. & Claverie J.M. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344-1350.

Redrejo-Rodriguez M., Munoz-Espin D., Holguera I., Mencia M. & Salas M. 2012. Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc Natl Acad Sci U S A* 109: 18482-18487.

Rice G., Stedman K., Snyder J., Wiedenheft B., Willits D., Brumfield S., McDermott T. & Young M.J. 2001. Viruses from extreme thermal environments. *Proc Natl Acad Sci U S A* 98: 13341-13345.

Roberts M.M., White J.L., Grutter M.G. & Burnett R.M. 1986. Three-dimensional structure of the adenovirus major coat protein hexon. *Science* 232: 1148-1151.

Rohwer F. & Thurber R.V. 2009. Viruses manipulate the marine environment. *Nature* 459: 207-212.

Rossmann M.G. & Rao V.B. 2012. *Viral molecular machines*. Advances in experimental medicine and biology,, Springer, New York.

Rossmann M.G., Arnold E., Erickson J.W., Frankenberger E.A., Griffith J.P., Hecht H.J., Johnson J.E., Kamer G., Luo M., Mosser A.G. & et al. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317: 145-153.

Schneider T.R. & Sheldrick G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58: 1772-1779.

Schofield D.A., Sharp N.J. & Westwater C. 2012. Phage-based platforms for the clinical detection of human bacterial pathogens. *Bacteriophage* 2: 105-283.

Skirnisdottir S., Hreggvidsson G.O., Hjorleifsdottir S., Marteinsson V.T., Petursdottir S.K., Holst O. & Kristjansson J.K. 2000. Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Appl Environ Microbiol* 66: 2835-2841.

Smith G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228: 1315-1317.

Smith H.O. & Wilcox K.W. 1970. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *J Mol Biol* 51: 379-391.

Smith H.W., Huggins M.B. & Shaw K.M. 1987. The control of experimental Escherichia coli diarrhoea in calves by means of bacteriophages. *J Gen Microbiol* 133: 1111-1126.

Solorzano A., Rodriguez-Cousino N., Esteban R. & Fujimura T. 2000. Persistent yeast single-stranded RNA viruses exist in vivo as genomic RNA.RNA polymerase complexes in 1:1 stoichiometry. *J Biol Chem* 275: 26428-26435.

Stockley P.G. & Twarock R. 2010. *Emerging topics in physical virology*. Imperial College Press, London.

Strand M.R. & Burke G.R. 2012. Polydnaviruses as symbionts and gene delivery systems. *PLoS Pathog* 8: e1002757.

Stuart D.I., Levine M., Muirhead H. & Stammers D.K. 1979. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J Mol Biol* 134: 109-142.

Suttle C.A. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 5: 801-812.

Szajner P., Weisberg A.S., Lebowitz J., Heuser J. & Moss B. 2005. External scaffold of spherical immature poxvirus particles is made of protein trimers, forming a honeycomb lattice. *J Cell Biol* 170: 971-981.

Tokuriki N., Oldfield C.J., Uversky V.N., Berezovsky I.N. & Tawfik D.S. 2009. Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 34: 53-59.

Tuma R., Tsuruta H., Benevides J.M., Prevelige P.E., Jr. & Thomas G.J., Jr. 2001. Characterization of subunit structural changes accompanying assembly of the bacteriophage P22 procapsid. *Biochemistry* 40: 665-674.

Veesler D., Ng T.S., Sendamarai A.K., Eilers B.J., Lawrence C.M., Lok S.M., Young M.J., Johnson J.E. & Fu C.Y. 2013. 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.

Villarreal L.P. & Witzany G. 2010. Viruses are essential agents within the roots and stem of the tree of life. *J Theor Biol* 262: 698-710.

Vonrhein C., Blanc E., Roversi P. & Bricogne G. 2007. Automated structure solution with autoSHARP. *Methods Mol Biol* 364: 215-230.

Williams H.T. 2013. Phage-induced diversification improves host evolvability. *BMC Evol Biol* 13: 17.

Winter G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43: 186-190.

Woese C.R. & Fox G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74: 5088-5090.

Woese C.R., Kandler O. & Wheelis M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87: 4576-4579.

Wommack K.E. & Colwell R.R. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* 64: 69-114.

Ye X., Ou J., Ni L., Shi W. & Shen P. 2003. Characterization of a novel plasmid from extremely halophilic Archaea: nucleotide sequence and function analysis. *FEMS Microbiol Lett* 221: 53-57.

Yu M.X., Slater M.R. & Ackermann H.W. 2006. Isolation and characterization of Thermus bacteriophages. *Arch Virol* 151: 663-679.

Zhang X., Sun S., Xiang Y., Wong J., Klose T., Raoult D. & Rossmann M.G. 2012. Structure of Sputnik, a virophage, at 3.5-A resolution. *Proc Natl Acad Sci U S A* 109: 18431-18436.