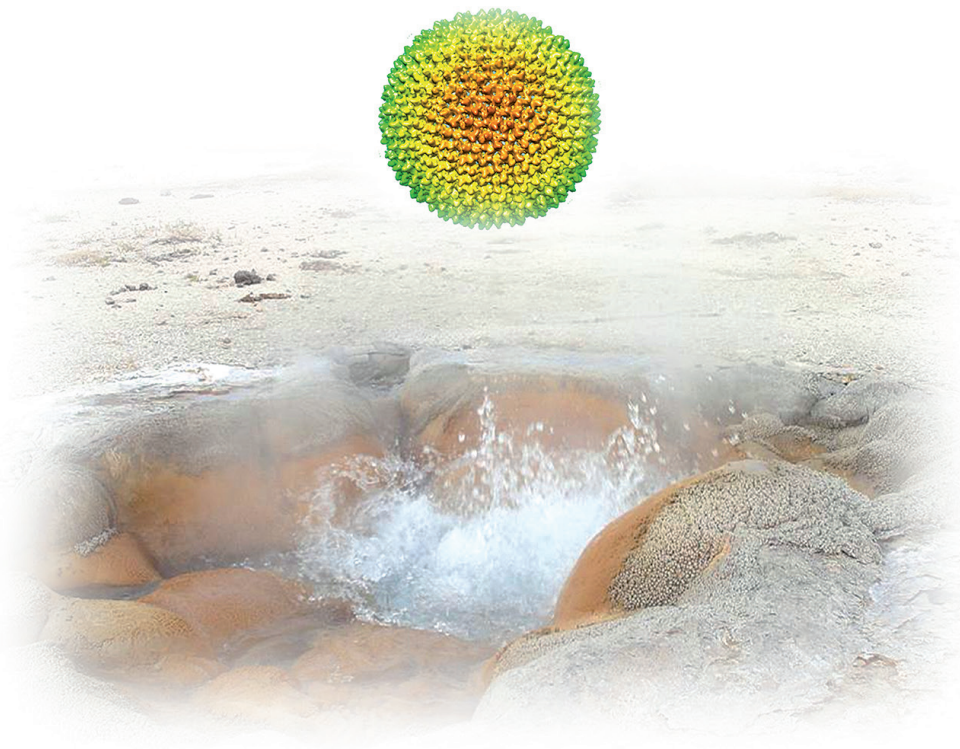


Alice Pawlowski

Thermus Bacteriophage P23-77:
Key Member of a Novel, but
Ancient Family of Viruses from
Extreme Environments



Alice Pawlowski

Thermus Bacteriophage P23-77:
Key Member of a Novel, but Ancient
Family of Viruses from Extreme
Environments

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen auditoriossa 3,
huhtikuun 17. päivänä 2015 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Agora, auditorium 3, on April 17, 2015 at 12 o'clock noon.



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Thermus Bacteriophage P23-77:
Key Member of a Novel, but Ancient
Family of Viruses from Extreme
Environments

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 300

Alice Pawlowski

Thermus Bacteriophage P23-77:
Key Member of a Novel, but Ancient
Family of Viruses from Extreme
Environments



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Editors

Varpu Marjomäki

Department of Biological and Environmental Science, University of Jyväskylä

Pekka Olsbo, Ville Korhokangas

Publishing Unit, University Library of Jyväskylä

Jyväskylä Studies in Biological and Environmental Science

Editorial Board

Jari Haimi, Anssi Lensu, Timo Marjomäki, Varpu Marjomäki

Department of Biological and Environmental Science, University of Jyväskylä

Cover picture: *Thermus* phage P23-77 (EM data bank entry 1525) above
geysers steam boiling Yellowstone by Jon Sullivan / Public Domain.

URN:ISBN:978-951-39-6154-1

ISBN 978-951-39-6154-1 (PDF)

ISBN 978-951-39-6153-4 (nid.)

ISSN 1456-9701

Copyright © 2015, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2015

Für Jan

ABSTRACT

Pawlowski, Alice

Thermus bacteriophage P23-77: key member of a novel, but ancient family of viruses from extreme environments

Jyväskylä: University of Jyväskylä, 2015, 70 p.

(Jyväskylä Studies in Biological and Environmental Science

ISSN 1456-9701; 300)

ISBN 978-951-39-6153-4 (nid.)

ISBN 978-951-39-6154-1 (PDF)

Yhteenveto: *Thermus*-bakteriofagi P23-77: muinaisen ääriolosuhteiden virus-suvun merkittävä jäsen

Diss.

Hot springs harbor a remarkable diversity of viruses, making them a valuable source for the discovery of novel viral types, yet only a few viruses have been characterized in detail. One of those, bacteriophage P23-77, exhibits an unusual capsomer architecture and -arrangement. The capsid is built of two major capsid proteins in contrast to other tailless, icosahedral dsDNA viruses with an inner membrane that have one trimeric capsid protein as the major capsid component. P23-77 shares similarities with a group of haloarchaeal viruses, indicating that those extremophilic viruses belong to a novel lineage of ancient viruses.

In this thesis, the large and the small major capsid protein were successfully crystallized for subsequent structure analysis. A third capsid-associated component, the minor capsid protein VP11, was shown to be a highly heat-stable dimer with predominantly α -helical secondary structure, thus excluding a function as a penton protein. Instead, it interacts with lipids and the large major capsid protein, indicating localization between the outer capsid shell and the inner membrane. VP11 most likely reinforces capsid stability and facilitates correct incorporation of the two different major capsid proteins during capsid assembly. Comparative genomics was used to detect proviruses related to P23-77 in the genomes of *Thermus*- and *Meiothermus* species isolated from all over the world, indicating a global distribution of this virus type in thermal environments. Moreover, the phylogenetic relationship to a group of viruses infecting archaeal halophiles was determined. P23-77-like bacteriophages and archaeal viruses were merged in the novel family *Sphaerolipoviridae*. It is the first assigned bacteriophage family in 30 years and the only family comprising viruses from two domains of life besides the families of head-tail viruses.

Keywords: Capsid-associated proteins; extreme environments; viral lineages; viral self; virion assembly; virus taxonomy.

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

Author's address Alice Pawlowski
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
alice.pawlowski@jyu.fi

Supervisor Professor Jaana Bamford
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

Reviewers Docent Sarah Butcher
Institute of Biotechnology
P.O. Box 56
FI-00014 University of Helsinki
Finland

Docent Petri Auvinen
Institute of Biotechnology
DNA Sequencing and Genomics laboratory
P.O. Box 56
FI-00014 University of Helsinki
Finland

Opponent Associate Professor Kenneth Stedman
Department of Biology
Center for Life in Extreme Environments
Portland State University
P.O. Box 751
Portland, OR 97207-0751
United States

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

ABBREVIATIONS

1	INTRODUCTION	11
2	REVIEW OF THE LITERATURE	13
2.1	Viruses from extreme environments	13
2.1.1	Viral diversity in hot and hypersaline environments	14
2.1.2	Potentials of extremophilic viruses	16
2.1.3	<i>Thermus</i> phage P23-77	18
2.2	Classification of viruses	20
2.2.1	Baltimore classification	20
2.2.2	Classification according to the International Committee on Taxonomy of Viruses (ICTV)	21
2.3	Detection of viral phylogenetic relationships	22
2.3.1	Viral genes and genomes	22
2.3.2	Capsid protein structure defines viral lineages	24
2.4	Assembly of prokaryotic icosahedral dsDNA viruses	28
3	AIMS OF THE STUDY	30
4	OVERVIEW OF THE METHODS	31
5	RESULTS	32
5.1	Analysis of the P23-77 capsid-associated proteins VP11, VP16 and VP17 (II, IV)	32
5.1.1	Protein expression and purification	32
5.1.2	Crystallization of major capsid proteins VP16 and VP17, and virus particles	33
5.1.3	Characterization of the minor capsid protein VP11	35
5.2	Analysis of the phylogenetic relationship of P23-77 to other extremophilic viruses and detection of related proviruses (I, III)	37
5.2.1	P23-like phages form a new genus, gammasphaerolipovirus, within the novel family sphaerolipoviridae	37
5.2.2	Detection of virus-like elements in <i>Thermaceae</i> and <i>Halobacteriaceae</i>	40
6	DISCUSSION	42
6.1	A group of thermophilic bacteriophages and haloarchaeal viruses constitute an ancient viral family 42	
6.1.1	<i>Thermus</i> phage P23-77 represents an ancient virus type	42
6.1.2	Sphaerolipoviruses are globally distributed	44
6.1.3	Relation to other viruses from extreme environments	46
6.1.4	" <i>Sphaerolipoviridae</i> " in the framework of the ICTV	46
6.2	The role of the P23-77 minor capsid protein VP11 in capsid assembly	49

6.2.1	VP11 is excluded as a penton protein.....	49
6.2.2	VP11 links the outer capsid shell to the internal membrane.....	49
6.2.3	VP11 facilitates the capsid assembly.....	50
6.2.4	The P23-77 capsid assembly.....	51
7	CONCLUSION.....	53
	ACKNOWLEDGEMENTS.....	55
	YHTEENVETO (RÉSUMÉ IN FINNISH).....	56
	REFERENCES.....	58

LIST OF ORIGINAL PUBLICATIONS

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

- I Jalasvuori M., Pawlowski A. & Bamford J.K.H. 2010. A unique group of virus-related, genome-integrating elements found solely in the bacterial family *Thermaceae* and the archaeal family *Halobacteriaceae*. *Journal of Bacteriology* 192: 3231–3234.
- II Rissanen I., Pawlowski A., Harlos K., Grimes J.M., Bamford J.K.H. & Stuart D.I. 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.
- III 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* 159: 1541-1554.
- IV Pawlowski A., Moilanen A.M., Rissanen I., Määttä J.A., Hytönen V.P., Ihalainen J.A. & Bamford J.K.H. 2015. The minor capsid protein of extreme thermophilic bacteriophage P23-77 facilitates capsid assembly using lipid-protein interactions. Submitted manuscript.

Article I was part of the thesis of Matti Jalasvuori (2010). Articles II and III were part of the thesis of Ilona Rissanen (2014). My contribution to the original publications was as follows:

- I I identified provirus sequences in the genomes of *Meiothermus* species, performed initial analysis of the MeioSilP1 and MeioRubP1 sequences and took part in writing. Matti Jalasvuori analyzed and annotated the genomes, prepared the figures and wrote most of the article.
- II I planned and supervised most of the experiments, constructed plasmids for protein expression, optimized VP17 protein purification, crystallized VP17 and virus material and took part in writing. Ilona Rissanen purified VP16, conducted initial purification of VP17, crystallized VP16 and the VP16/VP17-complex, prepared parts of the figures and wrote most of the article.
- III I had the original idea for the article, analyzed and annotated novel proviral sequences TP1 and MeioRubP2, prepared the tables and most of the figures and wrote most of the article. Ilona Rissanen prepared molecular graphics and took part in writing.
- IV I had the idea for the publication, performed most of the experiments, prepared the figures and wrote most of the article.

ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CD	Circular dichroism
cryo-EM	Cryo-electron microscopy
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
ds	Double-stranded
HHIV	<i>Haloarcula hispanica</i> icosahedral virus
ICTV	International Committee on Taxonomy of Viruses
kb	Kilo bases
kDa	Kilo Dalton
LUCA	Last universal cellular ancestor
M	Molarity
MCP	Major capsid protein
mRNA	messenger ribonucleic acid
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
POGs	Phage orthologous groups
RNA	Ribonucleic acid
RT	Reverse transcriptase
SDS	Sodium dodecyl sulphate
ss	Single-stranded
STIV	<i>Sulfolobus</i> turreted icosahedral virus
T- LUV	<i>Thermus</i> -derived large unilamellar vesicle
VP	Virion protein

1 INTRODUCTION

Virus research focuses mainly on pathogenic viruses of humans, animals and agricultural crops. However, eukaryotes represent just a small fraction of the biosphere that is dominated by prokaryotes, namely bacteria and archaea. About 10^{30} prokaryotes are estimated to dwell in the oceans and similar numbers are predicted for other environments. There are around 50 times more bacteria found in one gram of human feces than people living in the world (Clokie *et al.* 2011). Prokaryotes and their viruses reside in the most inhospitable environments where eukaryotes have no chance to survive. These viruses outnumber their hosts by at least a factor of 10, making prokaryotic viruses the most abundant biological entity on Earth. Thus, it has become evident that viruses have a much higher impact on the evolution of life and the global ecosystem than initially thought (Koonin and Dolja, 2013). As predators of prokaryotes, viruses shape the marine microbial community composition and population dynamics, thereby affecting global carbon and energy cycles (Suttle *et al.* 2007). Viruses are also ancient. They co-evolve with their hosts in a constant arms race, driving the evolution of cellular life for billions of years. Their genomes are highly diverse and the majority of the protein sequences encoded in the genomes show no similarities to any other sequences in public data bases. Moreover, viral genomes lack a universal marker gene that can be used to create a viral tree of life analogous to the tree of life of cellular organisms based on the small subunit ribosomal RNA gene. Sequence comparison can identify phylogenetic relationships of closely-related viruses (Genus level) but it fails to detect higher order relationships. However, a Linnean-like approach of structure comparison has proved successful in identifying some of the phylogenetic relations between viruses stretching across hosts from all three domains of life (Abrescia *et al.* 2012).

The goal of this thesis was to elucidate the capsid structure and the assembly of *Thermus thermophilus* phage P23-77 and its evolutionary relationship to other icosahedral viruses with inner membrane. The two major capsid proteins of P23-77 were successfully crystallized (II). The resolved structures (Rissanen *et al.* 2013) identified P23-77 as the most ancient member

known of the diverse double-beta barrel structural viral lineage, thus providing insights into the very early evolution of viruses before the occurrence of the last universal ancestor of cells (LUCA). Genetic as well as structure-based methods were used to identify members of an ancient viral lineage (I, III). Bacteria-infecting members of the lineage, represented by *Thermus thermophilus* phage P23-77, were proposed to form a new genus "*Gammasphaerolipovirus*" within the novel family "*Sphaerolipoviridae*" comprising archaeal viruses from hypersaline environments. The family was very recently approved by the International Committee on Taxonomy of Viruses (ICTV). With the addition of a bacteriophage genus, *Sphaerolipoviridae* represents the first family of tailless viruses that spans two domains of life, namely archaea and bacteria (III). Hot environments seem to support the conservation of ancient virus types. They also force their inhabitants to adapt to harsh conditions. How are virus particles stabilized in such surroundings? How does the complex capsid of P23-77 assemble? The results obtained by the analysis of the minor capsid protein VP11 suggest ways in which this type of membrane-containing viruses stabilize capsid components and orchestrate the capsid assembly (IV).

2 REVIEW OF THE LITERATURE

2.1 Viruses from extreme environments

Wherever there is water on Earth, there is life and wherever there is life, there are viruses. Even the most hostile environments such as hot acid springs, hydrothermal vents, deep subsurface sediments, hypersaline and soda lakes, or Antarctic sea ice harbour an astonishing variety of organisms, so called extremophiles (Rothschild and Mancinelli 2001). The record holder in terms of tolerance to high temperature is *Methanopyrus kandleri* of the archaeal phylum Euryarchaeota. This methane-producing organism was isolated from a black smoker at a depth of 2000 m depth in the Gulf of California. It is able to grow and reproduce at 122 °C (Takai *et al.* 2008). Viruses infecting extremophiles have been found in all of the above mentioned environments (LeRomancer *et al.* 2007) but the vast majority were isolated from thermo- and halophilic environments. Habitats characterized by high temperature (> 65 °C) or high salinity (15 - 30% = 2.5 - 5.2 M NaCl) are relatively uniform in terms of species diversity and dominated by only a few organisms (Skirnisdottir *et al.* 2000, Hacene *et al.* 2004, Sharp *et al.* 2014). In such a surrounding, viruses remain the only predators of prokaryotes and the major driving force for evolution. The number of viruses in moderate aquatic environments ranges from 10⁴ to 10⁸ ml⁻¹ and exceeds the abundance of prokaryotic cells by at least one order of magnitude (Wommack and Colwell 2000, Suttle 2007). The population sizes of viral communities in extreme environments could be even greater, reaching particle concentration of 10⁹ ml⁻¹ in solar salterns and alkaline lakes (Guixa-Boixareu *et al.* 1996, Jiang *et al.* 2004). With a few exceptions of viruses with a ssDNA genome (Pederson *et al.* 2001, Pietilä *et al.* 2009, Mochizuki *et al.* 2012), all viruses isolated from extreme environments so far have a dsDNA genome of moderate complexity.

2.1.1 Viral diversity in hot and hypersaline environments

Over 6000 prokaryotic viruses have been studied by electron microscopy (Ackermann and Prangishvili 2012). The vast majority (98.5%) of them are bacteriophages due to the fact that they are relatively easy to cultivate. Bacteriophages have been intensively studied for over a century. On the other hand, the first archaeal virus was described in 1974 (Torsvik and Dundas 1974), 16 years before the formal establishment of *Archaea* as the third domain of life by Woese *et al.* (1990). Although viruses infecting archaea have gained increasing interest during the last years, only about 100 of those have been studied so far (Pina *et al.* 2011, Pietilä *et al.* 2014). All of them were isolated from hot or hypersaline environments. The relatively small number reflects the difficulties in virus isolation and culturability. Most microorganisms dwelling in extreme habitats are impossible to cultivate in the laboratory for now, preventing the study of their associated viruses. Archaea reside also in moderate environments but no archaeal virus has been isolated from such habitat so far.

The majority of the known bacteriophages (97 %) have an icosahedral or nearly icosahedral head to which a linear tail is attached and no lipid moiety in their virions (in the following sections described as a “head-tail” morphology) (Ackermann and Prangishvili 2012). They are divided into three families, *Siphoviridae* (long contractile tail), *Myoviridae* (long, non-contactile tail) and *Podoviridae* (short tail) (Fig. 1). All other bacterial viruses are tailless, with icosahedral (either with or without lipid membrane), linear or pleomorphic morphology. Despite the extensive examination of bacteriophages, novel morphotypes have not been detected for 40 years. This is in contrast to the amazing diversity of hyperthermophilic archaeal viruses: Viruses obtained from hot springs in Japan, Iceland, New Zealand, Italy, and Yellowstone National Park revealed various novel virus types including filamentous, rod-, spindle-, droplet- and bottle-shaped viruses. These novel virus types infect exclusively members of the phylum *Crenarchaeota*, mainly of the genera *Sulfolobus* and *Acidianus* (Rice *et al.* 2001, Pina *et al.* 2011). Remarkably, head-tail viruses, the most abundant virus type among bacteriophages and probably in the whole world (Hendrix 2002, Ackermann 2007), were so far not reported for hyperthermophilic *Crenarchaeota*. Among those, spindle shaped virus types are dominating (Fig. 1). The only known thermophilic archaeal head-tail viruses, siphoviruses Ψ M1 and Ψ M2, infect *Methanobacterium*, a member of the Euryarchaeota (Meile *et al.* 1989, Pfister *et al.* 1998). The phylum Euryarchaeota comprises both, halophilic and thermophilic viruses whereas all crenarchaeal viruses are hyperthermophilic.

The number of described bacteriophages from extreme environments is far less than the number of archaeal viruses, especially in the case of haloviruses. A comprehensive study of hypersaline environments carried out by Atanasova *et al.* (2011) describes 49 viruses, from which only 6 % infect bacteria. Sabet (2012) lists 70 halophilic viruses, 9 of those bacteriophages. The majority of haloviruses

adopt the head-tail morphology (86 % of bacterial, 85% of archaeal viruses) (Sabet 2012).

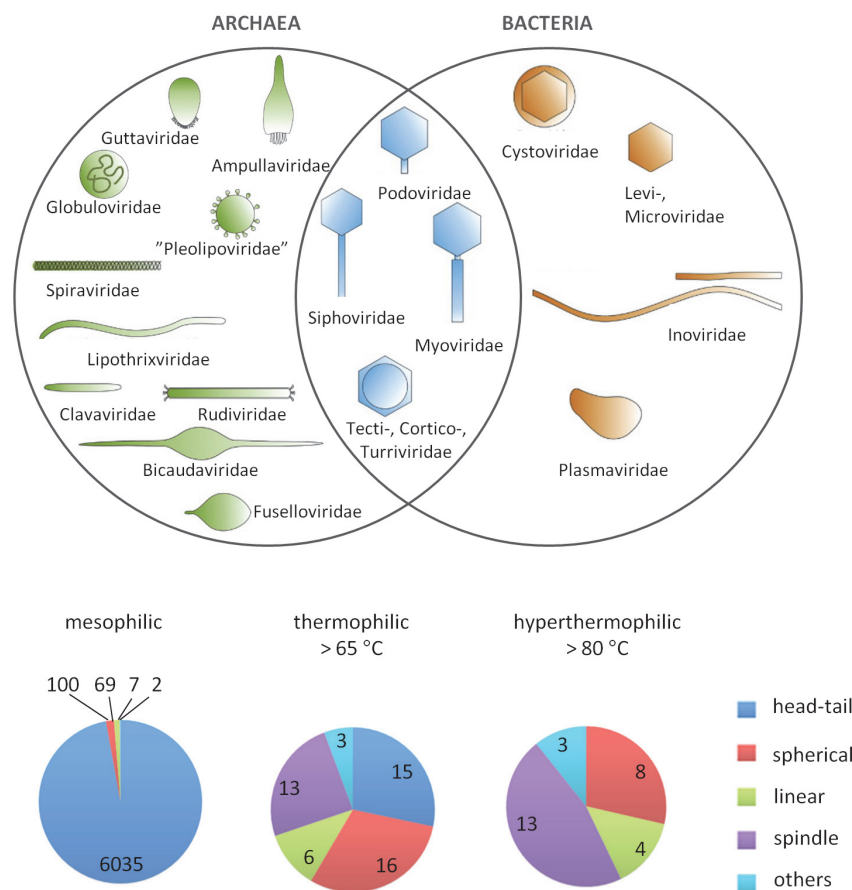


FIGURE 1 Archaeal and bacterial virion morphotypes. Morphotypes are reproduced from Pietilä *et al.* (2014) with permission Elsevier. Virus types depicted in blue share a common ancestry. Family names are according to nomenclature of the International Committee on Taxonomy of Viruses (ICTV). Diagrams show the distribution of virus types in various temperatures. Numbers of viruses according to Ackermann and Prangishvili (2012) and Uldahl and Peng (2013).

Is the astonishing diversity of crenarchaeal viruses a unique characteristic of this group? Direct electron-microscopic observation of viruses and virus-like particles (VLPs) from hypersaline environments revealed a high abundance of unusual morphotypes typical for crenarchaeal viruses. Only small amounts of head-tail viruses were observed, indicating that halophilic euryarchaeal viruses are as diverse as hyperthermophilic crenarchaeal viruses (Oren *et al.* 1997, Sime-Ngando *et al.* 2011). Moreover, a group of pleomorphic archaeal viruses ("Pleolipoviridae") with lipid envelope and spike-like structures on the virion

surface was exclusively found in hypersaline environments (Pietilä *et al.* 2014). The by far most extensive survey on thermophilic bacteriophages revealed that the majority (51%) of isolated *Thermus* phages were spherical without a tail (Yu *et al.* 2006). The result contradicts the previous observed dominance of head-tail viruses among bacteriophages. Furthermore, *Thermus* siphoviruses mainly have an unusual morphology with exceptionally long tails (Minakhin *et al.* 2008, Lin *et al.* 2010).

From the observations reported so far, one can draw the conclusion that viruses from extreme environments are more heterogenous than moderate-temperature viruses. However, virus research is strongly biased by focussing on viruses infecting easily isolated and cultivated laboratory strains, which may not be the dominating species in the natural environment. Lytic viruses producing clear plaques on agar plates are easier to detect than lysogenic viruses or viruses that exit the host cell without lysis, a common egress mechanism among archaeal viruses (Pina *et al.* 2011). Unfortunately, there are so far no viruses reported of moderate temperature crenarchaea or hyperthermophilic bacteria of the genera *Thermatoga* and *Aquifex*. It remains an open question if bacteriophages of extreme hot environments (~ 90 °C) show morphologies so far restricted to archaeal viruses or if head-tail viruses can be detected from crenarchaea thriving in moderate habitats.

2.1.2 Potentials of extremophilic viruses

Extreme environments, especially the hot ones, harbour a plethora of novel virus types (Rice *et al.* 2001, Pina *et al.* 2011). Some of them revealed features that have not been observed before and challenge our current concept of what viruses are. Viruses are generally defined as parasites strictly depending on the host cell machinery for replication and propagation and unable to generate energy by themselves. The extracellular form of the virus, the virion, is regarded as an inert particle, a vehicle protecting the genome and enabling transmission from one host cell to the other. However, the virion of *Acidianus* two-tailed virus (ATV) undergoes major morphological changes independent from the host and external energy sources. Virions are released as tailless lemon-shaped particles and develop two long tails at each end in a temperature dependent manner (Häring *et al.* 2005). The unrelated *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) and *Sulfolobus* turreted icosahedral virus (STIV) share a unique virion release mechanism involving the formation of pyramid-like structures on the cell surface called virus-associated pyramids (VAPs) (Prangishvili and Quax 2011). VAPs are formed by a single protein and exhibit a unique geometry not yet found among other viruses or cellular organisms.

The hosts of thermophilic viruses root deepest in the phylogenetic tree of life (Pace 1997). Surroundings like volcanic springs might mimic best the conditions on the early Earth when the three domains of life have formed around 4 billion years ago (Battistuzzi *et al.* 2004). Neutral mutations, well-tolerated at standard temperatures, become highly deleterious at high temperatures. Thus, thermophilic organisms are under strong negative selection

and evolve more slowly than mesophilic organisms (Friedmann *et al.* 2004, Drake 2009). The same might apply to other extremophiles, where the harsh conditions of the environment favor negative selection. Every improved feature acquired in the adaptation process to a specific niche will be highly conserved. Extant species might have preserved many characteristics of their ancestors. Extremophilic viruses have adapted to the harsh conditions of their environment same as their hosts. Consequently, examination of viruses from extreme environments provides valuable insights into ancient virus types and early viral evolution. Indeed, the analysis of major capsid protein folds received from protein crystals produced in this thesis work (II) identified *Thermus* phage P23-77 as the most ancient member of a viral lineage dating back in the primordial world before the occurrence of LUCA (Rissanen *et al.* 2013).

The high thermostability of proteins from organisms dwelling in hot environments makes them attractive for an application in some industrial processes. One of the milestones in molecular biology, the development of the polymerase chain reaction (PCR) by Kary Mullis of Cetus Corporation in 1983, would not have been successful at the industrial scale without the usage of a thermostable DNA-polymerase such as *taq* from *Thermus aquaticus*. Virions are exposed to the same extreme conditions as their hosts and replicate very effectively. Therefore, mainly viral enzymes involved in nucleic acid metabolism sparked the interest of biotechnology companies. Indeed, a thermostable viral DNA polymerase with novel features was identified from a viral metagenome from Octopus hot spring in Yellowstone National Park (Moser *et al.* 2012). The enzyme has a dual function as DNA-polymerase and reverse transcriptase (RT). Conventional RT-PCR systems require a two-step reaction, with initial reverse transcription at low temperature followed by DNA amplification at high temperature. The viral RT-polymerase allows reverse transcription at high temperatures, thereby increasing fidelity of the reaction. The RT-polymerase is commercialized by Lucigen Corporation, WI, US (PyroPhage® 3173 DNA Polymerase). Other enzymes of thermophilic viral origin already commercialized include ssDNA/RNA ligase I (CircLigase™ ssDNA ligase, Epicentre, WI, US) and polynucleotide ligase I (ThermoPhage™ Polynucleotide kinase, Pokazyme, Reykjavik, Iceland) (Blondal *et al.* 2005a, 2005b). Enzymes - yet not commercialized but with potential for biotechnological application - include a thermostable protein-primed DNA polymerase with potential for amplification of long PCR fragments (~70 kb) (Peng *et al.* 2007) and a nonspecific nuclease capable of degrading various nucleic acid types at temperatures up to 80 °C (Song and Zhang 2008). In addition, the small size and self-assembly capability of some viruses hold promise for their usage as nanoparticles. Viral nanoparticles retrieved from extreme environments may provide resistance to inorganic solvents or high temperature required for nanotechnological applications (Uldahl and Peng 2013).

2.1.3 *Thermus* phage P23-77

The search for novel heat-stable DNA-polymerases applicable in biotechnology was the driving force for the most extensive survey on *Thermus*-viruses performed by Promega Corporation (Yu *et al.* 2006). During this study, 115 bacteriophages including head-tail (44%), spherical (51%) and filamentous viruses (5%) were obtained from alkaline hot springs in the US, Russia and New Zealand. Seven tailless, icosahedral dsDNA viruses from the Promega collection were analyzed at the University of Helsinki for their capability to propagate in *Thermus thermophilus* ATC33923 (Jaatinen *et al.* 2008). One of them, P23-77, was characterized in detail due to its high stability and capability to grow to high titers. P23-77 is a strictly lytic phage with a life cycle of approximately 70 min at 70 °C. Virus particles are spherical with an average diameter of 78 nm and thin, 15 nm long spikes emerging from the vertices. A lipid membrane is located between the capsid shell and the dsDNA genome. The membrane lipids are selectively acquired from the host (Jalasvuori *et al.* 2009) as typical for icosahedral dsDNA viruses with inner membrane (Laurinavicius *et al.* 2004, Bamford *et al.* 2005b). The overall morphology strongly resembles the morphology of tectiviruses, explaining why P23-77 was initially regarded as a member of the *Tectiviridae* family (Yu *et al.* 2006). However, the genomic and structural characterization of P23-77 revealed clear differences to tectiviruses (Jaatinen *et al.* 2008, Jalasvuori *et al.* 2009): Firstly, the genome has no sequence similarity to genomes of tectiviruses and is circular instead of linear. Secondly, the capsid is built of two major capsid proteins (MCPs) instead of one.

The 17 kb circular genome contains 37 ORFs, which are arranged in the same direction (Jalasvuori *et al.* 2009). The majority of the predicted gene products as well as the genome type and gene arrangement are shared with *Thermus aquaticus* bacteriophage IN93 (Matsushita and Yanase 2009), another tailless, spherical phage isolated from hot spring soil in Japan. The vast majority of the gene products (78 %) lack similarity to any other protein sequences in public databases. Remarkably, P23-77 and IN93 do not encode their own DNA-polymerase. Cryo-electron microscopy (cryo-EM) and image reconstruction of the P23-77 virion revealed an unusual capsid architecture. The geometrical organization of capsomers in an icosahedral virus can be defined by the triangulation (T) number (Casper and Klug 1962). The T-number describes the number of proteins per icosahedral asymmetrical unit and thus the size of the virus. The capsid of P23-77 is arranged in an unusual T = 28, dextro lattice. Capsomers, the building blocks of the virion protein shell, have tower-like crenellations emerging from the capsomer base (Jaatinen *et al.* 2008). The exploration of the biotechnological potential of P23-77 has not yielded a commercially viable product yet. However, the genome encodes several predicted lytic enzymes, including an endolysine, an amidase and a thermophilic lysozyme (Matsushita and Yanase 2008), offering a putative commercial exploitation.

All virus types identified for thermophilic bacteriophages (head-tail, linear and spherical) have counterparts in the archaeal domain. Bacterial viruses with linear morphology (*Inoviridae*) resemble their archaeal counterparts (*Lipothrixviridae*, *Rudiviridae*, *Clavaviridae*, *Spiraviridae*) in shape. Yet, inoviruses have ssDNA genomes whereas archaeal viruses have dsDNA genomes. Furthermore, no significant sequence similarity could be detected at the amino acid level. Other factors such as presence of lipids, life cycle and capsid protein fold also distinguish bacterial and archaeal helical viruses, rendering a phylogenetic relatedness unlikely (Pietilä *et al.* 2014). On the contrary, head-tail viruses share a common ancestry as shown by particular similarities in their genomes and a similar structure of the major capsid protein (Pietilä *et al.* 2013, Sencilo *et al.* 2013) (Fig. 1). P23-77 belongs to the third virus group stretching across the domains of bacteria and archaea, namely tailless icosahedral viruses with internal membrane and dsDNA genome. Viruses with this morphology have been isolated from hyperthermophilic-, hypersaline- and moderate environments. Besides general similarities in the virion morphology, hyperthermophilic viruses of the *Turriviridae* family, STIV and STIV2 (Rice *et al.* 2004, Happonen *et al.* 2010), are linked to viruses from moderate environments such as corticovirus PM2 (Abrescia *et al.* 2008) and tectivirus PRD1 (Abrescia *et al.* 2004) by a similar double beta-barrel core fold of their MCP, indicating a common ancestry of these viruses (described in more detail in chapter 2.3.2). In contrast, thermophilic bacteriophages P23-77 and IN93 use two MCPs for capsid assembly. Notably, the same applies to a group of tailless icosahedral viruses from hypersaline environments including archaeal *Haloarcula hispanica* viruses SH1 (Bamford *et al.* 2005b, Porter *et al.* 2005, Kivelä *et al.* 2006, Porter and Dyall-Smith 2008, Porter *et al.* 2008, Jääliñoja *et al.* 2008), PH1 (Porter *et al.* 2013), and HHIV-2 (Jaakola *et al.* 2012), and *Natrinema* virus SNJ1 (Zhang Z. *et al.* 2012) as well as bacterial *Salisaeta* phage SSIP-1 (Aalto *et al.* 2012). The three lytic *Haloarcula* viruses have linear genomes of around 30 kb in size that are highly similar with 72 % nucleotide sequence identity between SH1 and PH1, 59 % between SH1 and HHIV-2, and 54 % between PH1 and HHIV-2. The level of gene synteny is high between the three virus genomes (Porter *et al.* 2013, Dyall-Smith *et al.* 2013a). In contrast, SNJ1 is a temperate virus with a circular dsDNA genome of much smaller size (~16 kb). Remarkably, the virus was originally described as plasmid pHH205 (Ye *et al.* 2003). The genome of SNJ1 shares little similarity with the genomes of the *Haloarcula* viruses. Yet, a similarity is observed in the amino acid sequence of some proteins and in the arrangement of viral core genes encoding a putative packaging ATPase and the two MCPs (Porter *et al.* 2013). The best characterized virus of this group is SH1. The capsid is icosahedral, tailless, with horn-like spikes on the five-fold vertices and an inner lipid membrane located between the capsid and the 31 kb, linear dsDNA genome (Bamford *et al.* 2005b, Jääliñoja *et al.* 2008). The lipids are selectively acquired from the host during particle assembly (Bamford *et al.* 2005b). SH1 is the only other virus so far with a T = 28, dextro capsid symmetry and a similar crenellated capsomer architecture as P23-77 (Jääliñoja *et al.* 2008), suggesting an

evolutionary relatedness between the two viruses (Jaatinen *et al.* 2008, Jalasvuori *et al.* 2009).

In this thesis, key elements were defined to demonstrate the common ancestry of those thermophilic bacteriophages and haloarchaeal viruses. In addition, the relationship to other tailless, icosahedral viruses with an inner membrane was elucidated (I, III).

2.2 Classification of viruses

Viruses lack a universal gene such as the 16S ribosomal RNA genes that allow the classification of all living organisms into one universal tree of life (Woese and Fox 1977, Woese *et al.* 1990). For this reason, there is no single universal taxonomy for viruses. However, it is convenient for the daily work of a scientist to have a system that brings order to the vast continuum of viruses. Two commonly used current classification systems are the Baltimore system and the taxonomic system established by the International Committee on Taxonomy of Viruses (ICTV).

2.2.1 Baltimore classification

The system, developed by the American biologist David Baltimore in 1971, accommodates the importance of the production of mRNA for virus propagation since all virus proteins are synthesized by the cellular translational machinery (Baltimore 1971). The system was developed for animal viruses, which can be grouped in seven classes according to the nature and polarity of the genome.

Class I viruses have a dsDNA genome that can be transcribed into mRNA from the negative (-) DNA strand. *Class II* comprises viruses with (+) ssDNA genome. Viruses of this class have to produce a dsDNA intermediate for mRNA transcription during their life cycle. *Class III* contains viruses with a dsRNA genome. Here, the (-) RNA strand is used as a template for transcription into mRNA. *Class IV* viruses have a (+) ssRNA genome that can be directly translated into protein. *Class V* viruses contain a (-) ssRNA genome. The (-) ssRNA strand serves as a template for the production of mRNA. *Class VI* are (+) ssRNA viruses. The genome does not serve directly as mRNA for protein production as in *Class IV* viruses. Instead, a dsDNA intermediate is produced for replication. RNA is converted to DNA by reverse transcriptase. *Class VI* viruses are therefore named Retroviruses. *Class VII* comprises a small group of viruses with a dsDNA genome. The genome serves as template for mRNA synthesis as well as for the production of a genomic RNA intermediate, which is the template for reverse transcriptase during genome replication.

Vertebrates are the only hosts infected by all seven classes of the Baltimore system. RNA viruses are as yet unknown for Archaea, whereas dsDNA viruses

are missing in multicellular plants. Class VII viruses are restricted to animals and higher plants. Class V, VI and VII viruses are absent in bacteria.

2.2.2 Classification according to the International Committee on Taxonomy of Viruses (ICTV)

The ICTV was created as a committee of the Virology Division of the International Union of Microbiological Societies (IUMS) in the early 1970s. It aims to order the growing variety of identified viruses under a single, universal taxonomy of viruses of animals, plants, fungi, bacteria and archaea, an on-going effort. The ICTV tries to detect phylogenetic relationships between viruses by comparing several factors such as genome type (as in the Baltimore system), sequence similarities, host range and virion morphology. Both, taxonomic issues and new proposals are examined and validated by international Study Groups. Results are published in the form of ICTV reports (King *et al.* 2012). An up-to-date taxonomy is available online (ICTV 2014). The lowest taxonomic level in the Linnean-like hierarchy approved by the ICTV is the virus species, defined as “a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria” (Adams *et al.* 2013). Criteria may include host range, pathogenicity, cell- and tissue tropism, physicochemical properties of the virion and the relatedness of genomes. They are established by the appropriate Study Group to which taxonomic proposals are submitted. Higher taxon levels are genus (suffix *-virus*), subfamily (*-virinae*), family (*-viridae*) and order (*-virales*). The current taxonomy release lists 2827 virus species distributed amongst 455 genera, 22 subfamilies, 103 families and 7 orders (ICTV 2014).

Both, Baltimore and ICTV classification appreciate the importance of the genome for virus classification. Knowledge on the nature and polarity of the genome provides the researcher with immediate information about the replication and propagation mode of the virus of interest. Yet, it is a very broad classification that does not infer phylogenetic relatedness and viruses belonging to the same group do not need to share a common ancestor. In contrast, the ICTV classification is based on the presumption that members of a given taxon share a common ancestry. This is a challenge to prove for higher order taxa (above family level) given the rapid pace of genetic differentiation, genome mosaicism as a result of extensive horizontal gene transfer and – quite often – lack of sufficient structural information on viruses. As a result, the majority of families (77) are not assigned to an order. Taxonomic levels higher than order are not established at all.

2.3 Detection of viral phylogenetic relationships

The parasitic life style of viruses contributes to rapid replication- and mutation-rates and massive gene exchange - between viruses themselves and between viruses and their hosts. As a consequence, viruses are the most diverse biological entity on Earth, covering an enormous sequence space. The following section describes possibilities and obstacles in the detection of viral phylogenetic relationships by the analysis of viral genes and genomes, and how high resolution structures of major capsid proteins could be used to infer a common ancestry of very distantly related viruses.

2.3.1 Viral genes and genomes

Comparative genomics is the major tool for detecting phylogenetic relationships between viruses. Whole genome sequences, genes, and gene order are analyzed to detect orthologous sequences which indicate a common ancestry. Genes are compared on the nucleic acid level when they are so closely related that the corresponding protein sequences are identical. Otherwise, a comparison of the amino acid sequences of the gene products is used to identify homologies. Sequences are aligned to detect similarities using specific sequence alignment algorithms. Common algorithms include ClustalW (Thompson *et al.* 1997) for general sequence alignments and BLAST (Altschul *et al.* 1997) for public database searches. Both were used in this study (I, III, IV). Sequence alignments are used to construct phylogenetic trees reflecting evolutionary relationships between homologous genes present in the genomes of divergent species. The degree of differences between aligned sequences is translated into an evolutionary distance. In other words: high sequence identity suggests comparatively recent divergence while low identity indicates a more ancient divergence from a common ancestor.

The small genome size of viruses compared to cellular organisms simplifies the comparison of whole genome sequences. Indeed, one of the first comparative genomic studies was performed on genomes of small ssRNA viruses infecting animals and a plant, revealing relatedness between these viruses (Argos *et al.* 1984). Recent development of advanced sequencing techniques (Hall 2007) in combination with effective methods for virus isolation (Thurber *et al.* 2009, Hurwitz *et al.* 2013) has allowed the massive acquisition of viral genome sequences. Culture-independent metagenomics has tremendously increased the number of novel genotypes identified in viromes (all virus genomes present in a particular environment) from various environments, for example up to 1300–1400 for hot springs (Schoenfeld *et al.* 2008). The majority of those virome-derived sequences lack significant similarity to any sequences in public databases, and there seems to be no end to this variety in the continuously accumulating sequence data (Koonin and Dolja, 2013). Viral genomes contain numerous ORFans - ORFs existing only in one particular genome or a group of closely related genomes without homologs in current

databases (Siew *et al.* 2004). A significant number of ORFans in bacterial and archaeal genomes are considered to be of viral origin (Yin and Fischer 2006, Cortez *et al.* 2009, Makarova *et al.* 2014). Hence, viruses are not only the most abundant biological entity; they also represent the highest genetic diversity in the world. The majority of distinct genes are considered to reside in viral genomes. Metagenomic approaches allow us to explore the full viral diversity in different environments and to expand the enormous viral sequence space. This might help to construct more accurate phylogenetic trees, but fail when it comes to ORFans: Without having any reference sequence and information about the function of a gene, a sequence remains just “dark matter of the viral biosphere” (Comeau *et al.* 2008, Youle *et al.* 2012). In this case, a detailed characterization of the virus is inevitable, a complex task that requires successful isolation and culturability of the virus-host system.

However, there is some order in the diversity of viral genes and genomes: Comprehensive comparative studies on prokaryotic genomes revealed the presence of several phage-specific gene families, so called phage orthologous groups (POGs). These provide essential tools for tracking down viral phylogenetic relationships (Kristensen *et al.* 2011, 2013). POGs include protein encoding genes with function in replication (e.g., DNA-polymerase, - helicase), genome packaging (e.g., terminase) or structure (e.g., major capsid protein, portal protein). Some of them may serve as signature genes for the detection of viruses from particular taxa in metagenomic samples. The usage of POGs in virus phylogeny is of special importance in the case of dsDNA viruses with head-tail morphology, whose genomes are characterized by modular organization and mosaicism (Hendrix *et al.* 1999, Brüssow and Hendrix 2002, Krupovic *et al.* 2011). Extensive horizontal gene transfer, a central feature of tailed bacteriophages, led to the formation of mosaic genomes. Single genes, gene domains or even entire functional modules (gene clusters encoding proteins involved in the same process) are exchanged between different viruses, hampering the reconstruction of evolutionary relationships. The position of functional modules in the genome might differ between virus groups. Genes within a module might have no detectable sequence similarity in distantly related viruses but the gene order within a given module is usually well conserved (Hendrix 2002).

Genes could be divided in core- and non-core genes (Krupovic *et al.* 2011). Non-core genes comprise genes of unknown function or functions linked to specific interactions with the host cell such as host recognition, cell entry and release of virus progeny. They are extensively swapped horizontally either by gene exchange between viruses infecting the same cell or, most likely, between a virus and a resident provirus. They are generally not common to all members in a virus group and evolve rapidly, covering the tracks of evolutionary relationships.

In contrast, core genes are shared by all members of a related virus group and comprise essential features of a virus. They are therefore regarded as the innate “self” of the virus (Bamford D.H. *et al.* 2002, Bamford 2003). Elements of the viral self comprise fundamental structural and functional principles of the

virion such as capsid architecture and genome packaging. They are not transferred horizontally (Filée *et al.* 2006) and highly conserved between viruses with common ancestry.

The order of viral self genes is well conserved in the genomes of tailless, icosahedral dsDNA viruses (Krupovic and Bamford 2008b). Comparison of conserved gene arrangements in those viruses led to the discovery of related proviruses in the genomes of several bacterial and archaeal species (Krupovic and Bamford 2007, 2008a, Jalasvuori *et al.* 2009) and during the course of this study (I, III). One of the self genes encodes an ATPase required for genome packaging. The packaging of the genome into the capsid is a crucial step during virus assembly and therefore highly conserved. The ATPase of icosahedral viruses with inner membrane contains – besides the classical Walker A and Walker B motifs – an additional signature motif, which was originally identified in the P9 packaging ATPase of PRD1 (Strömsten *et al.* 2005). ATPases of this type belong to the FtsK-HerA superfamily of P-loop ATPases (Iyer *et al.* 2004, Happonen *et al.* 2013). Within this family, the viral DNA packaging ATPases form a distinct clade (Iyer *et al.* 2004).

2.3.2 Capsid protein structure defines viral lineages

The rapid pace of viral evolution restricts the identification of evolutionary relationships based on comparison of gene- or protein sequences to closely related, recently diverged groups of viruses. In terms of longer evolutionary time scale, sequences evolve to the point beyond any recognizable similarity rendering it difficult to track down phylogenetic relationships between distantly related viruses, for example between viruses infecting hosts from different domains of life. On the other hand, protein folds could be highly conserved. The viral sequence space is immense, yet the number of functional protein folds is apparently limited (Abrescia *et al.* 2012). With regard to viral coat proteins, only a small portion of protein folds have the potential to form a functional capsid, resulting in approximately 50 unique viral capsid protein folds (Oksanen *et al.* 2012). The capsid is what makes a virus a virus. It enables the formation of a virion, the extracellular stage in the life cycle of a virus during which the viral genetic information is transferred from one host cell to another. The production of a protective protein shell surrounding the genome uncouples the transfer of genetic information from the necessity of cell-to-cell contact. This clearly distinguishes a virus from other mobile genetic elements such as plasmids or transposons. This is strikingly exemplified by the comparison of porcine circovirus 1 (PCV1) and *Helicobacter pylori* plasmid pAL236-5: both are minimalistic genetic systems with just one (pAL236-5) or two (PCV1) genes, respectively (Krupovic and Bamford 2010). Both carry a gene required for replication, but only PCV1 carries in addition a gene encoding a capsid protein, enabling virion-formation. Hence, the main component building the capsid, the major capsid protein (MCP), represents the key self-element of the virus (Bamford D.H. *et al.* 2002, Bamford 2003). The comparison of the three-dimensional atomic structures of major capsid proteins (MCPs) of virions have

revealed insights into the evolutionary relationships between viruses (Rossmann *et al.* 1984, Benson *et al.* 1999, Tate *et al.* 1999, Bamford *et al.* 2002, Benson *et al.* 2004, Bamford *et al.* 2005a, Baker *et al.* 2005, Abrescia *et al.* 2010, Abrescia *et al.* 2012). It has become obvious that MCPs are highly conserved between viruses with common ancestry and provide a useful tool for higher-order classification (above family level) of viruses. Based on the similarity of high resolution structures of MCPs, the enormous diversity of viruses is so far ordered into just four different major lineages (Abrescia *et al.* 2012).

The most successful lineage is the *HK97-like lineage*, exemplified by the MCP structure of bacteriophage Hong Kong 97 (Wikoff *et al.* 2000). All dsDNA bacteriophages with head-tail morphology as well as archaeal members of the *Caudovirales* and herpesviruses belong to this lineage (Baker *et al.* 2005, Bamford *et al.* 2005a, Pietilä *et al.* 2013). Therefore, the lineage includes members of all three domains of life.

The *BTV-like viral lineage*, exemplified by bluetongue virus (BTV) (Grimes *et al.* 1998), comprises icosahedral dsRNA viruses from the domains of Eukarya and Bacteria. Yeast L-A virus, a totivirus (Naitow *et al.* 2002), and bacterial cystovirus phi6 (Huiskonen *et al.* 2006) and phi8 (El Omari *et al.* 2013) share the inner capsid architecture and MCP fold with BTV and other rheoviruses (Nakagawa *et al.* 2003, McClain *et al.* 2010). Chryso-, partiti- and picobirnaviruses are also suggested to belong to this lineage (Abrescia *et al.* 2012).

The *picorna-like lineage* encompasses 16 families of icosahedral viruses, in majority small ssRNA viruses (Abrescia *et al.* 2010). The core fold of the MCPs of this lineage is an eight-stranded antiparallel beta barrel – the so-called jellyroll – orientated planar to the virion surface. The single jellyroll fold is predominantly found in eukaryotic plant and animal viruses with positive ssRNA genome, but is also present in the capsid proteins of ssDNA bacteriophage phiX174 (Dokland *et al.* 1997) as well as dsDNA papillomaviruses and polyomaviruses (Stehle *et al.* 1996, Chen *et al.* 2000). Those are consequentially placed in the same lineage although the nature of their genomes is different (Abrescia *et al.* 2010). The lineage comprises viruses from the two domains Eukarya and Bacteria.

The *double beta-barrel lineage* is the fourth and best characterized major structural lineage. It was established when the crystal structure of the MCP of enterobacteriophage PRD1 revealed a similar architecture as the hexon of human Adenovirus, namely an eight-stranded, antiparallel, double beta-barrel (Athappily *et al.* 1994, Benson *et al.* 1999) (Fig. 2). The major capsid protein of both viruses is a trimer. The six individual beta barrels of the trimer create the pseudo-hexameric base of the capsomer, the major hexagonal building block of the icosahedral capsid. In contrast to the single jellyroll capsid proteins (see above) the double jellyroll proteins of dsDNA viruses are orientated perpendicular to the surface of the capsid. Over the years, MCP structures with similar fold were identified in green algae *Paramecium bursaria* chloroella virus-1 (PBCV-1) (Nandhagopal *et al.* 2002), archaeal virus *Sulfolobus* turreted icosahedral virus (STIV) (Khayat *et al.* 2005) and marine bacteriophage PM2

(Abrescia *et al.* 2008), covering viruses from all three domains of life. Nucleocytoplasmatic large DNA viruses (NCLDVs) (Iyer *et al.* 2001, 2006) were proposed as members of the double beta-barrel lineage based on homology modelling of the MCPs of several large eukaryotic dsDNA viruses (Benson *et al.* 2004, Yan *et al.* 2009). Interestingly, brick-shaped vaccinia virus is assigned as member of the lineage by the structure of a scaffolding protein present in immature virions despite its otherwise deviating virion morphology (Bahar *et al.* 2011). The discovery that Sputnik, a virophage, also adopts the double beta-barrel MCP fold (Zhang X. *et al.* 2012), added another fascinating piece to the variety of hosts infected by members of the lineage.

The two jellyrolls in the double beta-barrel are named V1 and V2 according to their position in the sequence. They share the same topology but have no apparent sequence similarity (Benson *et al.* 1999). The lack of sequence similarity indicates an ancient origin of the double beta-barrel fold thought to be a result of gene duplication followed by gene fusion (Abrescia *et al.* 2008, Benson *et al.* 2004, Krupovic and Bamford 2008b). The two beta-barrels in the MCP of marine bacteriophage PM2 are more similar to each other than to any other capsid protein and lack elaborate extensions on the top of the jellyrolls (Fig. 2). Hence, the capsid protein of PM2 is regarded as a minimal double beta-barrel protein and most ancient member of the lineage (Abrescia *et al.* 2008). A precursor virus using six single beta-barrel proteins instead of one trimeric double beta-barrel protein for the formation of hexameric capsomers had not been discovered prior to this thesis work. Yet, archaeal virus SH1 (Jääliñoja *et al.* 2008) and bacteriophage P23-77 (Jaatinen *et al.* 2008) were suggested to be such a “missing link”. Both viruses use two major capsid proteins for building up their capsids. Based on protein analysis, protein mass estimates and cryo-EM image reconstructions it was suggested that the capsomer base is composed of six similarly sized beta-barrels of the small MCP while the large MCP sits on top of it (Jääliñoja *et al.* 2008, Jaatinen *et al.* 2008). Subsequently, SH1 and P23-77 were proposed to form the earliest branch in a so called “vertical beta-barrel superlineage” of dsDNA viruses, which use hexagonal capsomers for capsid assembly, formed by proteins with the canonical jelly-roll fold orientated vertical to the virion surface (Krupovic and Bamford 2008b). However, the structural information gained from cryo-EM image reconstructions (Jääliñoja *et al.* 2008, Jaatinen *et al.* 2008) did not provide evidence for the existence of such single beta-barrel capsid proteins. Thus, we aimed to crystallize the MCPs of P23-77 for the determination of their three-dimensional structure at the atomic level by X-ray crystallography (II).

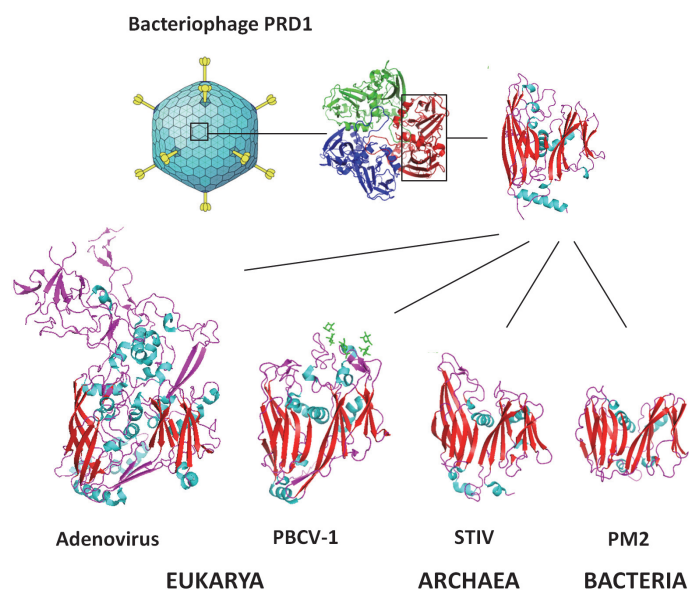


FIGURE 2 Major capsid proteins (MCPs) of representatives of the double beta-barrel lineage. Facets of the icosahedral capsid are assembled from pseudo-hexagonal capsomers built by the trimeric coat protein. The characteristic double jelly-roll core fold is conserved in MCPs of viruses from all three domains of life. PDB-codes for MCPs were 1CJD (PRD1), 1P30 (Adenovirus), 1M3Y (*Paramecium bursaria* chlorella virus-1 (PBCV-1)), 2BBB (*Sulfolobus turreted* icosahedral virus (STIV)) and 2VVF (PM2). PRD1 virion presentation adapted with permission from the Swiss Institute of Bioinformatics (ViralZone: www.expasy.org/viralzone, Swiss Institute of Bioinformatics).

Nearly half of the 103 viral families assigned by the ICTV to date (February 2015) can be placed into one of the above mentioned four major lineages and the majority of the lineage members have an icosahedral or – in case of head-tail viruses – a nearly icosahedral morphology. Hence, it seems possible to classify the vast majority of icosahedral viruses with this structural based approach but how about other virus types? The lack of structural data hampers the identification of lineages for helical and enveloped viruses. The diverse virion morphology of enveloped viruses complicates the definition of viral self-elements (Abrescia *et al.* 2012). Helical archaeal dsDNA viruses of the *Rudoviridae* and *Lipothrixviridae* share a similar MCP fold consisting of a four-helix bundle (Goulet *et al.* 2009). The structural similarity of the capsid proteins in combination with genome analysis was used to classify the two families within a new order *Ligamenvirales*, despite otherwise remarkable differences in virion morphology and genome replication mode (Prangishvili and Krupovic, 2012). A four-helix bundle fold is also found in the capsid of rod-shaped ssRNA viruses, yet the topology is different (Goulet *et al.* 2009). Thus, structure comparison of coat proteins seems to be useful to detect phylogenetic relationships between helical viruses.

Another obstacle of the structure-based phylogeny is the distinction between convergence and divergence. As outlined above, physicochemical constraints limit the number of protein folds required for the formation of functional viral capsids. Thus, the invention of certain capsid folds might rather reflect analogy (convergence) than homology (divergence). The ubiquitous viral jellyroll-fold is present in the coat proteins of small icosahedral ssRNA viruses (Rossmann and Johnson 1989) as well as dsDNA viruses and some nonviral proteins such as tumor necrosis factor (TNF) (Liu *et al.* 2002). The double beta-barrel fold is also present in the capsids of ssRNA cowpea mosaic virus (CPMV) (Lin *et al.* 1999). A comparison of structural details such as length of chains and loops, loop insertion sides and topological structure similarity may provide arguments towards homology or analogy. There are four different ways to fold an eight-stranded jellyroll (Stirk *et al.* 1992) but only one is seen in spherical viruses, suggesting common ancestry. The orientation of the beta-barrel in the capsid (parallel to the surface versus perpendicular) clearly distinguishes the Picorna-like lineage from the double beta-barrel lineage. The orientation of the beta-sheets in the double beta-barrel of CPMV indicates relatedness to the single beta-barrel of ssRNA viruses (Abrescia *et al.* 2010). However, comparison of structural folds alone is not sufficient to differentiate convergence and divergence. Instead, additional components of the viral self have to be taken into account. These include for example nature of the genome, virion architecture, gene synteny and genome packaging mechanism. Thus, the more similarities are shared between members of a viral lineage, the less probable it gets that they are a result of convergent evolution.

2.4 Assembly of prokaryotic icosahedral dsDNA viruses

In prokaryotic icosahedral dsDNA viruses, the genome is either packaged into preformed procapsids or the capsid assembles around a pre-condensed genome. The first strategy is employed by head-tail viruses and members of the double beta-barrel lineage (Mindich *et al.* 1982, Fu *et al.* 2010). The second one is proposed for another lineage member, marine bacteriophage PM2 (Huiskonen *et al.* 2004, Abrescia *et al.* 2008).

The procapsid, or prohead, of tailed phages is assembled with the help of scaffolding proteins. They act as chaperones to ensure the correct assembly of the capsid (Prevelige and Fane 2012). Scaffolding proteins could be either separate proteins or N-terminal domains of the major capsid proteins (King *et al.* 1973, Hendrix and Duda 1998, Morais *et al.* 2003, Huet *et al.* 2010). They are removed after the formation of the prohead to make space for the packaging of the genome. In many phages, scaffolding proteins are degraded by proteases present in the procapsid (Leiman *et al.* 2003, Thomas and Black 2013). The digestion products exit the prohead before the genome is packaged. In some viruses, such as *Salmonella typhimurium* phage P22 and in *Bacillus subtilis* phage ϕ 29, the scaffolding proteins exit the prohead intact during genome packaging

and can be reused in subsequent assembly rounds (King and Casjens 1974, Björnsti *et al.* 1983). DNA is packaged through the portal vertex by a DNA translocation motor powered by ATP-hydrolysis (Rao and Feiss 2008). During this process, the capsid is expanded to form the mature capsid. After the genome is filled into the mature capsid, the translocation motor is dislodged and the portal gate sealed by head completion proteins. They provide a platform for the tail-assembly of podoviruses or the attachment of the preassembled tails of siphoviruses and myoviruses, the final step of the virus assembly (Tavares *et al.* 2012).

Capsids of icosahedral prokaryotic dsDNA with inner membrane do not undergo the large structural rearrangements visible in head-tail viruses and empty procapsids resemble DNA-filled mature capsids. Instead, the internal membrane expands during the packaging process analogously to the capsid expansion observed in head-tail viruses (Butcher *et al.* 1995), thereby bringing the membrane in closer contact to the protein shell. Hence, the membrane acts like a scaffold in capsid assembly (Butcher *et al.* 2012). The DNA is packaged through a special vertex by a specific packaging ATPase (Gowen *et al.* 2003, Strömsten *et al.* 2005).

Assembly of bacteriophage PM2 does not involve an empty procapsid stage. The newly synthesized circular dsDNA genome in the cytoplasm is supercoiled as in the capsid (Ostrander and Grey 1974). There is no evidence for the presence of a special packaging vertex in the virion. The majority of virion proteins are associated with the membrane. Lipids and membrane proteins form stable DNA-filled vesicles, the so-called lipid core (Kivelä *et al.* 2002). Therefore, PM2 virion formation presumably begins with the assembly of the membrane around the pre-condensed genome analogous to the capsid assembly of Simian virus 40 (SV40), which starts with the polymerization of capsomers around the circular supercoiled dsDNA genome (Gordon-Shaag *et al.* 2002, Huiskonen *et al.* 2004, Abrescia *et al.* 2008).

In addition to major coat proteins, minor capsid proteins are found in the virions of prokaryotic icosahedral dsDNA viruses that reinforce capsid stability (“cementing”-proteins) (Butcher *et al.* 2012, Fokine and Rossmann, 2014). In some membrane-containing viruses, the capsid shell is linked to the internal membrane through interactions with the N- or C-terminus of the major capsid protein (Abrescia *et al.* 2004, Cockburn *et al.* 2004, Khayat *et al.* 2010). In marine bacteriophage PM2, which lacks the terminal extensions on the MCP, the capsid is linked to the membrane by two integral-membrane proteins (Abrescia *et al.* 2008). Both proteins are largely helical with domains reaching out from the inner and the outer membrane leaflet, thereby linking the condensed genome and the outer capsid protein shell to the membrane. They are supposed to form a scaffold, on which first the major capsid protein assembles followed by the receptor binding protein plugging in at the vertices as final step (Abrescia *et al.* 2008). In addition, capsids are connected to the membrane at the vertices (Jaatinen *et al.* 2004, Abrescia *et al.* 2004, Jääliñoja *et al.* 2008, Veessler *et al.* 2013).

3 AIMS OF THE STUDY

Thermus phage P23-77 was proposed to be the most ancient member of the double beta-barrel lineage with a capsid built of proteins with a single beta-barrel core fold. However, the structural evidence confirming the existence of such single beta-barrel capsid proteins was missing. Do the major capsid proteins VP16 and VP17 exhibit such fold? The third capsid-associated protein, the minor capsid protein VP11, has no homologs in other viruses. What is its localization and function in the virion? P23-77 shares a conserved set of core genes and similar overall architecture with a group of haloarchaeal viruses. Is there enough evidence to postulate a phylogentic relationship between these viruses spanning two domains of life? In order to answer these questions, the research performed on *Thermus thermophilus* phage P23-77 during this thesis had the following aims:

1. To establish methods for expression and purification of P23-77 capsid components for subsequent structural and functional analyses;
2. To produce single well-diffracting crystals of the two major capsid proteins VP16 and VP17 and P23-77 virus particles for high resolution structure analysis;
3. To characterize the minor capsid-associated protein VP11 and elucidate its localization in the virus particle and function in virion assembly;
4. To compare P23-77 genetically and structurally to other viruses and viral elements to detect phylogenetic relationships and classify related viruses according to the criteria of the ICTV.

4 OVERVIEW OF THE METHODS

The table below provides a summary of the methods used in this thesis. Detailed descriptions of the methods are found in the original publications and the manuscript indicated by Roman numerals.

TABLE 1 Summary of the methods used in this thesis

Method	Publication
Molecular cloning	II, IV
DNA sequencing	I, II, IV
DNA-shift assay	IV
Recombinant protein expression	II, IV
Protein purification	II, IV
Virus propagation	II, IV
Virus particle purification	II, IV
Crystallization and X-Ray diffraction experiments	II
SDS-PAGE	II, IV
Immunodetection of proteins	IV
Circular dichroism spectroscopy (CD)	IV
Dynamic light scattering (DLS)	IV
Differential scanning calorimetry (DSC)	IV
Preparation of large unilamellar lipid vesicles	IV
Comparative genomics	I, III, IV

5 RESULTS

5.1 Analysis of the P23-77 capsid-associated proteins VP11, VP16 and VP17 (II, IV)

5.1.1 Protein expression and purification

Dissociation experiments with P23-77 virus particles identified three capsid-associated proteins: The two major capsid proteins VP16 and VP17, which build the bulk of the capsid, and the minor capsid protein VP11 (Jalasvuori *et al.* 2009). Two hypotheses constituted the starting point of our experiments: 1) the hexagonal capsomers of the P23-77 capsid are formed by six copies of the small MCP with a single beta-barrel core fold while the large MCP forms the tower-like protrusions emerging from the capsomer base, and 2) the minor capsid protein is the penton protein located at the vertices of the virion (Jaatinen *et al.* 2008). With the aim to verify those, we developed expression- and purification protocols for the three capsid-associated proteins in order to get the starting material for the crystallization of the MCPs (5.1.2) and the analysis of the minor capsid protein (5.1.3).

We used an efficient system for protein expression in *Escherichia coli*, the pET-expression system (Novagen), for the production of all three proteins. The system allows a stringent control of protein expression and the production of high amounts of target protein by using the highly efficient RNA-polymerase of bacteriophage T7. Full length gene sequences were amplified from P23-77 genomic DNA by PCR with primers allowing the insertion into expression vector pET22b(+) at an optimal distance to the strong ribosomal binding site. High yields of recombinant soluble protein were obtained by culturing the expression strains for approximately 20 h at 28 °C after induction of protein production at an optical cell density of 0.5–0.7 at 550 nm with 1 mM isopropyl-beta-D-thiogalactopyranoside. Best yields were obtained with *E. coli* strain HMS174(DE3) as expression host for the MCPs (II) and strain BL21(DE3) for the expression of VP11 (IV).

The thermophilic origin of P23-77 prompted us to use heat treatment as part of the purification procedure (II). This worked well for the MCPs. Crude extracts obtained from ultracentrifugation were heated for 10 min at 90 °C causing rapid denaturation of the majority of host proteins while retaining the capsid proteins in the soluble fraction. In contrast, minor capsid protein VP11 precipitated in crude extracts incubated at 70 °C, although the recombinant protein is highly heat stable ($T_m = 84.4$ °C, IV). Most likely, VP11 co-precipitated with denatured heat-sensitive *E. coli* proteins.

Proteins were purified to homogeneity by ion exchange chromatography, followed by size exclusion chromatography. The two MCPs were purified with anion exchange chromatography in 20 mM Ethanolamine pH 9.5 (II). In this condition, VP16 (pI = 8.14) run through the column while the majority of the host proteins bound, yielding highly purified protein. VP17 (pI = 5.76) bound to the column at pH 9.5 and was eluted with 50 mM NaCl. The high content of basic amino acids in VP11 (pI = 10.71) enabled the purification with cation exchange chromatography (IV). The protein bound to the column in 50 mM Bicine pH 8.5. A washing step with 600 mM NaCl removed the majority of host derived proteins from the column. Subsequently, VP11 was eluted by increasing the NaCl concentration to 750 mM. The remaining impurities in protein samples of VP16, VP17 and VP11 were removed by size exclusion chromatography. A high amount of salt (1M NaCl) had to be added in the case of VP11 to prevent protein aggregation on the column. The MCPs eluted in one single peak whereas the minor coat protein eluted in two peaks, a minor one containing a 50 kDa protein besides the 22 kDa VP11 protein and a major one containing mainly VP11. The 50 kDa protein was identified as VP11-dimer (Fig. 2 in IV). Proteins could be stored for months at 8 °C without showing significant degradation or aggregation.

In conclusion, high yields of soluble recombinant protein were obtained from the expression in *E. coli* for all three capsid-associated proteins. Proteins were purified to the required purity grade and concentration to perform subsequent analyses.

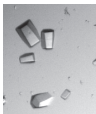
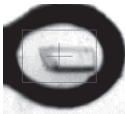
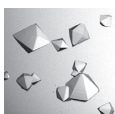
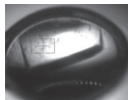
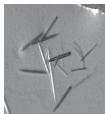
5.1.2 Crystallization of major capsid proteins VP16 and VP17, and virus particles

The preparation of well-diffracting protein single crystals is indispensable for the analysis of the high resolution structure by X-ray crystallography. To achieve this goal, protein samples have to be as pure, homogenous and monodisperse as possible. The purified protein samples of the major capsid proteins VP16 and VP17 matched those requirements and were used for crystallization trials (II). In addition, we aimed for the crystallization of whole virus particles. Virus particles were obtained by a two-step purification procedure including rate zonal centrifugation in a linear 5–20% (w/v) sucrose gradient, followed by equilibrium centrifugation in 1.30 mg/ml CsCl₂ (Jaatinen *et al.* 2008). Virions were highly stable and homogenous as checked with

dynamic light scattering (DLS) measurement, transmission electron microscopy, and infectivity assays.

Crystallization conditions were screened at room temperature by hanging-drop vapour diffusion using a crystallization robot and 96-well format commercial crystallization screens. Proteins (VP16 and VP17 in 20 mM Tris pH7.4, virus particles in 20 mM Tris-HCl pH7.5, 5 mM MgCl₂ and 150 mM NaCl) were mixed in 1:1 (v/v) ratio with the screen solution. Conditions yielding crystals were manually optimized by varying protein- and crystallization reagent concentrations. This way, three single crystal types were obtained, VP16-type 1, VP17, and virion-derived crystals (Table 2). Additional crystal types (VP16-type 2, VP16/VP17-complex) were obtained by sitting drop vapour diffusion at the Oxford Protein Production Facility, U.K., either by setting up conditions manually or by using the nanolitre high-throughput facility. All together, 576 crystallization conditions were tested for VP16, 288 for VP17 and 480 for the virion material. Most conditions (984) were checked to achieve a crystal of a putative VP16/VP17-complex, which took also longest to grow (3 months). PEG has proved to be a suitable precipitant for the crystallization of inner membrane-containing icosahedral viruses and capsid proteins (Bamford J.K. *et al.* 2002, Abrescia *et al.* 2008). It was also the best precipitant for the crystallization of VP16 and virion-derived material (Table 2). The optimal protein concentration was approximately 1 mg ml⁻¹ in the crystallization drop. All five crystal types diffracted to resolutions higher than 3 Å. Preliminary data analysis based on crystal unit cell dimension and solvent content implied 1–2 proteins per asymmetric unit, with the exception of the virion-derived crystals. Here, the unit cell was far too small to contain an 800 Å complete virus particle. Later it turned out that the virion-derived crystals actually contained VP16 (Rissanen *et al.* 2013). The only crystals obtained with virus material grew in solutions containing 0.1 M citric acid pH 3.5 and PEG3350. Most likely, virions dissociated in low pH, thus releasing VP16 that crystallizes well in PEG as shown above. At higher pH, only spherulites were observed, indicating improper or too fast crystallization of virus particles.

TABLE 2 Crystallization of P23-77 major capsid proteins and virion-derived material. Crystal types taken for diffraction were received from drops containing a mix of 1 μ l protein solution (concentration as indicated) and 1 μ l crystallization solvent. Images of VP16-type 2 crystals and the VP16/VP17 complex are modified from II.

	VP16-type 1	VP16-type 2	VP17	VP16/VP17-complex	virion-derived crystals
crystal					
protein concentration	2-3 mg ml ⁻¹	2-3 mg ml ⁻¹	2-3 mg ml ⁻¹	1.7 mg ml ⁻¹ VP16, 2.0 mg ml ⁻¹ VP17	2.4 mg ml ⁻¹
crystallization reagent	5% (w/v) PEG1000 5% (w/v) PEG8000	20% (w/v) PEG6000, 0.1 M citrate pH4.0	1.9 M sodium formate, 0.1 M bis-Tris pH7.0	1.1 M diammonium tartrate pH7.0	0.1 M citric acid pH3.5, 20 mM Tris-HCl pH7.5, 5 mM MgCl ₂ , 150 mM NaCl, 25% PEG3350
crystal growth	1-2 weeks	few days	2 weeks	3 months	12 h-8 d
diffraction resolution	1.8 Å	1.26 Å	2.26 Å	1.53 Å	2.92 Å

5.1.3 Characterization of the minor capsid protein VP11

Dissociation experiments (Jalasvuori *et al.* 2009) have shown that the major capsid proteins VP16 and VP17 could be released from the virion in 3 M urea and 3 M guanidine hydrochloride. By lowering the pH to 6.0, one other protein, the 22 kDa protein VP11, was also released while the membrane aggregated with the DNA. The function and localization of this minor capsid protein was unknown, prompting us to characterize it in detail (IV).

The VP11 gene is an ORFan. It has no homologues in related virus and proviruses (Fig. 1 in IV). BLAST-searches with the VP11 amino acid sequence produced one significant hit to a protein (WP_018467834.1) derived from *Meiothermus timidus*, suggesting the presence of a P23-77-like provirus in the corresponding genome. Protein sequence-based secondary structure analysis predicted a lack of transmembrane helices, short disordered stretches at The N- and C-termini and a high content of α -helices. The latter was confirmed experimentally by circular dichroism (CD) measurements in the far UV-region (Fig. 4 in IV).

The estimated melting point of VP11 in differential scanning calorimetry measurements was 84.4 °C. However, CD-spectra revealed a continuous decrease in the helicity of the protein when temperatures were raised from 20 °C to 85 °C, indicating partial unfolding. This process was completely reversed when protein samples were cooled down again from 85°C to 4 °C under the same conditions (Fig. 4 in IV), suggesting a partially loose α -helical structure. The hydrodynamic diameter of VP11 in dynamic light scattering (DLS) measurements was 8.4 ± 1.1 nm. The value refers to a molecular weight of 27.4 kDa in case of a linear protein or 96.7 kDa in case of a globular protein, respectively, pointing to a rather linear shape of VP11.

VP11 was shown to form dimers (Fig. 2 in IV). Dimer formation occurs through a disulfide bridge between single cysteine residues located at the N-terminus of each VP11 monomer. The intermolecular disulfide-bond of VP11 is extremely strong and cannot be disrupted by boiling in buffer containing SDS and 1% reducing agent β -mercaptoethanol. The dimer is most likely the native form of the protein in the virion since it was the predominant form detected by immunolabeling of freshly prepared virus particles with a VP11 polyclonal antibody. Disulfide-bond formation is a common mechanism for protein stabilization at high temperatures in cells (Beeby *et al.* 2005, Jorda and Yeates 2011) and viruses (Larson *et al.* 2007, Menon *et al.* 2008). The copy number of VP11 in the capsid was estimated by measuring the intensity of Coomassie stained protein bands of viral particles in SDS-PAGE. The theoretical copy numbers of the MCPs in the capsid are 1080 (VP16) and 540 (VP17) in total. By setting the VP16 protein band intensity as 1080, the estimated copy number of VP17 was 527, matching the theoretical value quite closely. The capsid consists of 60 asymmetric units. The estimated copy number for VP11 was 147, suggesting either 120 or 180 absolute copy number per capsid.

We analyzed the capability of VP11 to interact with various macromolecules (DNA, lipids, protein). VP11 strongly associated with DNA in an unspecific manner (Fig. 3 in IV). The positive charge of the protein (pI = 10.71) neutralized the negative charge of DNA at a protein:DNA mass ratio of 1:1 resulting in precipitation of the protein/DNA-complex. The DNA binding was particularly strong. Even 2.5 M NaCl was insufficient to dissociate VP11 and DNA completely. As a comparison: 0.3 M NaCl releases lysozyme from DNA (Steinrauf *et al.* 1999) and 1 M NaCl prevents the positively-charged domain of the scaffolding protein of bacteriophage P22 from binding to the coat protein (Parker and Prevelige 1998).

We also tested the interaction of VP11 with lipids. Large unilamellar lipid vesicles were produced from the virus host *T. thermophilus* ATCC33923 (in the following termed T-LUVs) to functionally mimic the viral membrane as close as possible. The zeta potential of those *Thermus*-derived lipid vesicles was -59.5 ± 3.96 mV, indicating a negative surface charge. The addition of VP11 to the T-LUVs immediately changed the zeta potential of the vesicles $+ 34.1 \pm 1.43$ mV, indicating an adsorption of VP11 to the lipids. Next, we analyzed the interaction of all three capsid-associated proteins with the T-LUVs by

estimating the apparent mean hydrodynamic diameter with the dynamic light scattering technique (DLS) (Fig. 3). VP16 and VP17 did not change the size of the T-LUVs, regardless of whether they were added to the vesicles alone or in combination. Addition of VP11 in turn led to a twofold increase in size of the T-LUVs. An incubation of T-LUVs with VP11 and VP16 increased the size same way as VP11 alone. On the other hand, addition of VP11 together with VP17 immediately caused a rapid increase in size, followed by aggregation and sedimentation of the lipid/protein-complex.

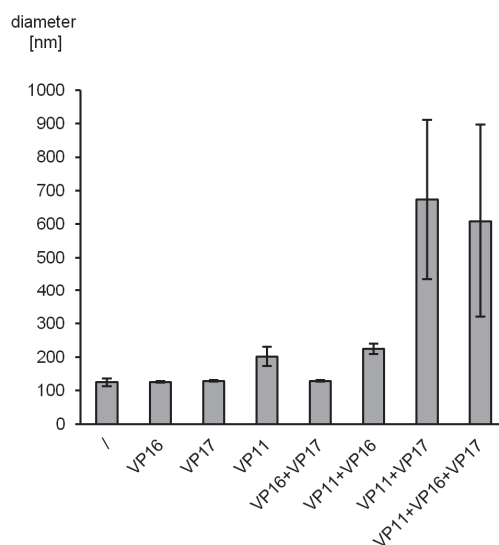


FIGURE 3 Incubation of *Thermus*-derived lipid vesicles with capsid-associated proteins show the effect of major capsid proteins VP16 and VP17, and minor capsid protein VP11 on the size of the T-LUVs. Reproduced from IV.

5.2 Analysis of the phylogenetic relationship of P23-77 to other extremophilic viruses and detection of related proviruses (I, III)

5.2.1 P23-like phages form a new genus, gammasphaerolipovirus, within the novel family sphaerolipoviridae

The overall morphology and capsid architecture of P23-77 resembles *Haloarcula hispanica* virus SH1 (see 2.1.3). SH1 in turn is similar to a set of viruses that have been identified during the course of this thesis and infect also members of the family *Halobacteriaceae*: *Haloarcula hispanica* viruses PH1 (Porter *et al.* 2013) and HHIV-2 (Jaakola *et al.* 2012), and *Natrinema* virus SNJ1 (Zhang Z. *et al.* 2012). This group of haloarchaeal viruses is clearly distinguished from other described viruses by the host spectrum, an unusual capsid geometry (T=28 dextro,

determined for SH1), the use of two MCPs for capsid assembly, a conserved arrangement of core genes, and the lack of sequence similarity to other published sequences (Dyall-Smith *et al.* 2013a). Hence, they were proposed to constitute a novel family of viruses termed "*Sphaerolipoviridae*", from the Latin *sphaero*, for "sphere", and the Greek *lipos*, for "fat" by Dyall-Smith *et al.* (2013a). The family included two genera, "*Alphasphaerolipovirus*" comprising the closely-related viruses SH1, PH1 and HHIV-2, and "*Betasphaerolipovirus*" comprising more distantly related virus SJ1 as the sole member. The similarity between P23-77 and SH1 (Jaatinen *et al.* 2008, Jalasvuori *et al.* 2009) prompted us to define features shared by P23-77- and SH1-like viruses with the aim to unify these two virus groups within one family (Table 3). P23-77 is related to *Thermus* phage IN93. From the 37 predicted ORFs of P23-77, 32 have homologs in IN93 as shown by the comparison of the corresponding gene products. The arrangement of the putative genes is highly conserved across the aligned genomes (Fig. 3 in III). Bacteriophages P23-77 and IN93 are similar to the haloarchaeal sphaerolipoviruses on the genome-level with respect to the nature of the genome (dsDNA), a conserved arrangement of viral core-genes (more detailed described in 5.2.2) and the lack of an own DNA-polymerase encoding gene. More similarities are found on the morphological and structural level (see below). Consequently, we proposed to include the two bacteriophages into a third genus, "*Gammasphaerolipovirus*", within the novel family (III). A proposal combining the three genera was submitted to the ICTV (Dyall-Smith *et al.* 2013b) and replaced the initial proposal, which included only the haloarchaeal viruses.

The capsids of all sphaerolipoviruses are built from two MCPs. In cases where more detailed structural information is available (SH1, P23-77), capsids are formed by two types of crenellated capsomers (Jääliñoja *et al.* 2008, Jaatinen *et al.* 2008). P23-77 has two turrets per capsomer arranged either on the same side or on opposite corners of the pseudo-hexameric capsomer. In SH1, one capsomer type has two turrets arranged on opposite corners like in P23-77 while the other type has three crenellations (III, Fig. 5). The structures of the P23-77 major capsid proteins derived from the crystals produced in this study superimpose well with the cryo-EM density map of SH1 capsomers (Rissanen *et al.* 2013), indicating a similar arrangement of large and small MCPs in the capsomers of SH1. Notably, in both viruses the strongest protein-protein interactions are observed between dimers spanning adjacent capsomers. The structural similarities clearly point to a common ancestry of the two viruses.

Apart from features common to all members of the family, some features are different. The most obvious difference refers to the hosts, which reside in different domains of life: Alpha- and betasphaerolipoviruses infect archaea of the family *Halobacteriaceae* while gammasphaerolipoviruses infect bacteria of the family *Thermaceae*. Alphasphaerolipoviruses have linear genomes with terminal repeats and covalently attached terminal proteins whereas all other family members have circular genomes. Members exhibit different life cycles: P23-77 and alphasphaerolipoviruses are lytic whereas IN93 and SNJ1 are temperate viruses. The temperate life style of IN93 compared to lytic phage P23-77 is

reflected in the otherwise highly similar genomes (III, Fig. 3). The genome of IN93 carries a block of genes required for lysogeny, the only genes located on the opposite strand of the genome (Matsushita and Yanase 2009). P23-77 lacks such an integration cassette explaining the smaller genome size (17 kb versus 19 kb) and the strictly lytic life-style.

TABLE 3 Main characteristics of the novel family sphaerolipoviridae (adapted from III, table 1)

genus	alphasphaerolipovirus	betasphaerolipovirus	gammasphaerolipovirus
virion morphology	icosahedral, nearly spherical capsid, tailless, spikes on vertices, ~10 structural proteins, Ø 50 – 80 nm		
membrane	internal lipid membrane, lipids acquired selectively from host cell during viral assembly*		
capsid	two major coat proteins (MCP)		
	triangulation number T=28*	n.d.	triangulation number T=28*
	two capsomer types with either two or three crenellations*	n.d.	two capsomer types with two crenellations in different orientation*
	n.d.	n.d.	MCPs with single β -barrel core fold, various MCP building blocks for capsid assembly*
genome	dsDNA, P9-type packaging ATPase encoding gene, similar arrangement of core genes (ATPase, MCP), no DNA-polymerase encoding gene		
	size ~ 30 kb	size ~ 16 kb	size ~ 17–19 kb
	linear, with inverted terminal repeats and terminal proteins	Circular	circular
life cycle	lytic	lysogenic (plasmid stage)	lytic or lysogenic (genome integrated)
host		halophilic archaea, family <i>Halobacteriaceae</i>	thermophilic bacteria, family <i>Thermaceae</i>
type virus	<u>SH1</u> , HHIV-2, PH1	<u>SNJ1</u>	<u>P23-77</u> , IN93

n.d., not determined

type-virus of each genus underlined

*determined for the type-virus of the genus

5.2.2 Detection of virus-like elements in *Thermaceae* and *Halobacteriaceae*

All members of “*Sphaerolipoviridae*” share a conserved set of genes encoding essential core proteins constituting the innate self of the virus, namely an ATPase required for packaging the genome (see 2.3.1) and the two major capsid proteins. The arrangement of those three genes is highly conserved in all members of the “*Sphaerolipoviridae*” family, with the ATPase gene located in close proximity to the in tandem arranged MCP genes (Fig. 4). The P9-specific motif, consisting of a conserved arginine (usually preceded by a glycine) and a conserved glutamine, separated by 10 amino acid residues (Strömsten *et al.* 2005) is present in the ATPases of all sphaerolipoviruses (suppl. 2 of III). The overall sequences similarity of the core proteins is low between genera, yet distinctive conserved residues could be detected (suppl. 2 of III). By using the amino acid sequence of the three viral self elements – the ATPase and the two MCPs – as seeds in BLAST searches, we identified two putative proviral sequences in the genome of *Halomicrobium mukohataei* DSM12286: HaloMukP1 and HaloMukP2 (I). The detection of proviral sequences in the genomes of *Haladaptatus paucihalophilus* DX253 and *Halobiforma lacisalsi* AJ5 by Porter *et al.* (2013) completed the picture of proviruses of *Halobacteriaceae*. We named them HalaPauP1 and HaloLacP1 to be in line with the designations of the other proviruses (III, Table 2). On the bacterial side, we detected proviruses in the genomes of *Thermus* sp. RLM (TP1, III), *Meiothermus silvanus* DSM9946 (MeioSilP1, I) and *Meiothermus ruber* DSM1279 (MeioRubP1, MeioRubP2, I, III) (Table 2 in III). All proviruses integrate into tRNA genes, a common integration site for bacterial and archaeal viruses (Reiter *et al.* 1989, Canchaya *et al.* 2004, Wiedenheft *et al.* 2004), allowing multiple integrations into a single host genome. None of the predicted proviral sequences seem to constitute a complete virus genome with the exception of TP1. It remains to be elucidated if TP1 or the other proviruses are capable of forming virions in order to assign them officially as sphaerolipoviruses. The proviral genomes are mosaics containing homologues of different members of a given genus. HalaPauP1 and HaloLacP1 are related to alphasphaerolipoviruses SH1, PH1 and HHIV-2 while IHP, HaloMukP1 and HaloMukP2 are related to betasphaerolipovirus SNJ1. The proviral sequences provided valuable data for the phylogenetic analysis of the family. Using the viral self elements (ATPase, large and small MCP) for a phylogenetic reconstruction produced congruent trees with all three genera falling into distinct clades (III, Fig. 6). Moreover, the phylogenetic grouping correlates with the arrangement of the self elements in the corresponding genomes. In alphasphaerolipoviruses and related proviruses, the ATPase gene is located 5–7 ORFs apart from the MCP genes on the opposite strand. In the other genera and their proviruses the ATPase gene is located 1–2 ORFs upstream of The MCP genes on the same strand (Fig. 4).

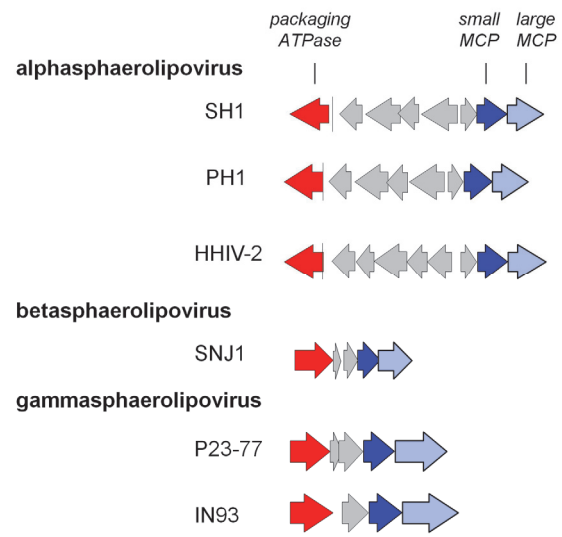


FIGURE 4 Arrangement of core genes in members of the "*Sphaerolipoviridae*" family

6 DISCUSSION

6.1 A group of thermophilic bacteriophages and haloarchaeal viruses constitute an ancient viral family

6.1.1 *Thermus* phage P23-77 represents an ancient virus type

The high quality of the five crystal types obtained in this study paved the way for the determination of the high resolution structures of the two MCPs (Rissanen *et al.* 2013) (Fig. 5). The small major capsid protein VP16 was shown to be a strand-swapped homodimer. Each subunit of the dimer is completed by a strand from the other subunit, resulting in an eight-stranded single beta-barrel fold per subunit. The strand-swap was also present in the dimers of the virion-derived crystals, indicating that the strand-swapped dimer is the native form of VP16 in the virion. The large major capsid protein VP17 has two domains, a lower domain with an eight-stranded beta-barrel fold and an upper domain with a six-banded beta-barrel fold.

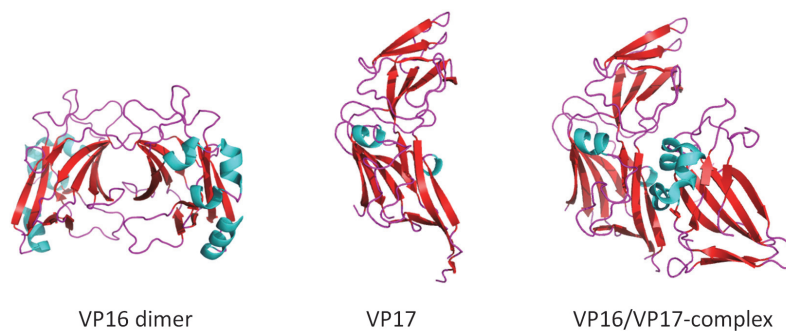


FIGURE 5 Crystal-derived P23-77 major capsid protein structures. PDB-codes for creating the structures were 3ZN5 (VP16 dimer obtained from virion-derived crystal), 3ZMN (VP17) and 3ZN6 (VP16/VP17 complex).

The eight-stranded single beta-barrels present in VP16 and the lower domain of VP17 are nearly identical (Rissanen *et al.* 2013). Moreover, structure homology comparison revealed that they are most closely related to the two individual beta-barrels of the major capsid protein of marine bacteriophage PM2 and – more distantly – to the V2 barrels of other double beta-barrel lineage members (Rissanen *et al.* 2013, Fig. 4 in III). PM2 is regarded as the most ancient member of the double beta-barrel lineage (Abrescia *et al.* 2008). The single beta-barrels of the P23-77 MCPs sit upright in the capsid as do the double jellyrolls (Rissanen *et al.* 2013). The upper domain of VP17 is inserted in the jellyroll of the lower domain between D and E strand (Rissanen *et al.* 2013). This is also a major insertion site in double beta-barrel proteins (Bahar *et al.* 2011). Thus, P23-77 most likely forms indeed the most ancient branch in the so called vertical beta-barrel superlineage of single- and double beta-barrel viruses (see 2.3.2). Viruses of that lineage are mainly icosahedral without tail and possess an internal lipid bilayer between the capsid shell and the dsDNA genome. In addition, pentameric proteins with a single beta-barrel fold and spike proteins are found at the vertices.

However, the capsomer arrangement of P23-77 differs from the proposed architecture of an ancestral virus of the lineage, in which six copies of one single beta-barrel protein form the capsomer base. Instead, the capsomer base is formed of four copies of the small major capsid protein VP16 and two copies of the large major capsid protein VP17, creating once more a pseudo-hexameric capsomer (Rissanen *et al.* 2013). The crystal structures of the two MCPs fit well in the cryo-EM density map of haloarchaeal virus SH1 (Jääliñoja *et al.* 2008), indicating that SH1 uses a similar capsomer arrangement (Rissanen *et al.* 2013). In both viruses, the strongest protein-protein interactions are observed between dimers spanning adjacent capsomers (Rissanen *et al.* 2013, Fig. 5 in III). This is in contrast to the pseudo-hexameric capsomers of the double beta-barrel viruses, where the strongest interactions are found between the subunits of the trimeric MCP within the capsomer. This finding strongly indicates that hexagonal “capsomers” are not the building blocks in the P23-77 and SH1 capsid assembly, revealing a novel assembly mode not observed in other viruses before (Rissanen *et al.* 2013). So far, structural information of the capsomer architecture is missing for the other proposed sphaerolipoviruses. However, there is a high level of amino acid sequence similarity and gene synteny between members of the same genus (see 5.2.1). This finding in combination with several other features shared by all members of the family (Table 3) strongly suggests a similar assembly principle for all sphaerolipoviruses.

Single beta-barrel members – represented by the “*Sphaerolipoviridae*” family – and double beta-barrel members of the superlineage diverged from a common ancestor that utilized hexameric capsomers built from a single beta-barrel MCP for capsid assembly (Jalasvuori *et al.* 2009, Rissanen *et al.* 2013) (Fig. 6). Gene duplication led to the formation of a virus with two separate single beta-barrel capsid proteins in one case and in the other case to two single beta-barrel proteins that diverged and fused during time to produce the double jellyroll protein present in the members of the today’s double beta-barrel

lineage (Rissanen *et al.* 2013). The example of the beta-barrel superlineage strikingly demonstrates how the structural based approach of virus phylogeny provides insights into the very early evolution of viruses that has been going on since long before the establishment of the three cellular domains of life.

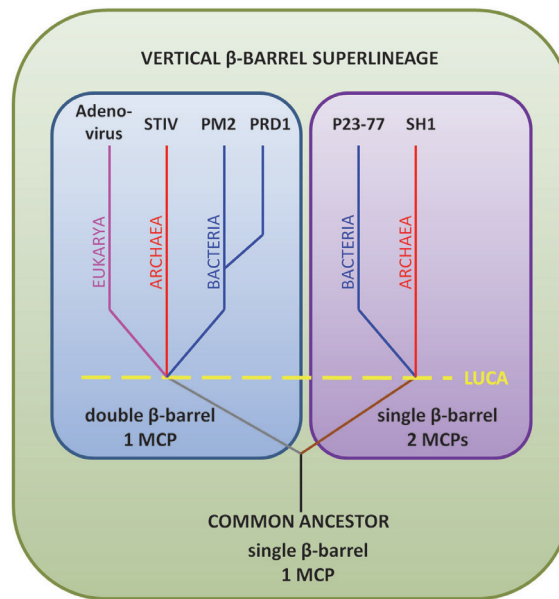


FIGURE 6 Evolution of the double beta-barrel superlineage based on structure of major capsid proteins (MCPs). Single and double beta-barrel branches diverged before the occurrence of the Last Universal Cellular Ancestor (LUCA).

6.1.2 Sphaerolipoviruses are globally distributed

The number of viruses, which are now grouped together in the “*Sphaerolipoviridae*” family, doubled during the course of this thesis and include currently HHIV-2 (Jaakola *et al.* 2012), SNJ1 (Zhang Z. *et al.* 2012) and PH1 (Porter *et al.* 2013) in addition to the initially described SH1 (Porter *et al.* 2005, Bamford *et al.* 2005b, Jääliñoja *et al.* 2008), P23-77 (Jaatinen *et al.* 2008) and IN93 (Matsushita and Yanase 2009). Eight related putative provirus sequences were detected in addition to the one initially described by Jalasvuori *et al.* (2009) (see 5.2.2, Table 2 in III). Thus, sphaerolipoviruses seem to represent a common virus-type in extreme environments and more viruses of this type are expected to be detected in the future. A conserved set of viral self elements – the genome packaging ATPase and the two MCPs – may serve as markers to detect putative new members of the three sphaerolipoviridae genera in newly published sequence data.

Gammasphaerolipovirus-related proviruses were mainly detected in the genomes of *Meiothermus* species suggesting a wide distribution of P23-77 like viruses in environments with temperatures below 70 °C. *Meiothermus* (meaning “less hot”) species thrive optimally at approximately 55 °C and are unable to grow above 60–70 °C, dependent on the strain (Nobre *et al.* 1996). Unfortunately, no virus has yet been reported for *Meiothermus* bacteria. Most likely, the preferred temperate life style among gammasphaerolipoviruses hampers the detection of such viruses. Hence, it remains to be elucidated if members of “*Sphaerolipoviridae*” could be detected from moderate environments or if they are yet another unique group of extremophilic viruses. Members of “*Sphaerolipoviridae*” and their putative proviruses are found all over the world, following the distribution pattern of their hosts (Fig. 7). In contrast to other *Thermus* genera, *T. thermophilus* is globally distributed (Hreggvidsson *et al.* 2012) explaining the isolation of similar virus types associated with it from New Zealand (P23-77) and from Japan (IN93) (Matsushita *et al.* 1995, Yu *et al.* 2006). Although extreme environments are confined areas, ecologically separated by large distances and physico-chemical boundaries, viruses form a global network. Similar virus sequences have been detected in metagenomes from hypersaline sources from Africa, the US and Spain (Sime-Ngando *et al.* 2011). HHIV-1 was isolated from a solar saltern in Italy (Atanasova *et al.* 2011). However, it revealed striking similarity to SH1 and PH1, isolated from salt lakes in Western Australia (Dyall-Smith *et al.* 2003, Porter *et al.* 2013). It seems that every newly-formed habitat is colonized by a similar community of organisms present in comparable habitats elsewhere in the world. Viral communities are dispersed within moderate distances by air (Snyder *et al.* 2007) but could such mechanism also account for a distribution on global scale? Microorganisms are spread globally by air for example by transpacific air plumes (Smith *et al.* 2013). The same may apply to viruses although virions could be regarded as highly sensitive to the pressure and desiccation present in the stratosphere. Temperate viruses are protected by the host cell during transport. Viruses from hot springs are suggested to travel across long distances covered in a protective silica-coat (Laidler *et al.* 2013). It remains to be elucidated how lytic viruses from hypersaline environments might be spread.

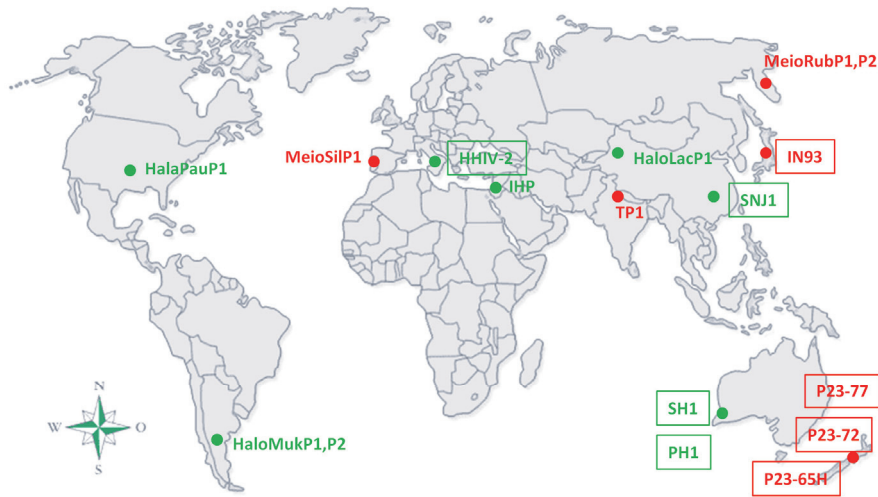


FIGURE 7 Isolation sites of sphaerolipoviruses (framed) and prokaryotic hosts with genome integrated proviruses. Archaeal members are shown in green, bacterial members in red.

6.1.3 Relation to other viruses from extreme environments

Three other icosahedral, tailless, membrane-containing dsDNA bacteriophages have been described that have capsids built from two major capsid proteins and might therefore be related to “*Sphaerolipoviridae*”. *Thermus* phages P23-72 and P23-65H (Yu *et al.* 2008) resemble P23-77 with respect to host spectrum, growth cycle, particle yields, and protein pattern in SDS-PAGE (Jaatinen *et al.* 2008). We assume that they belong to the same genus as P23-77. Yet, the lack of sequence information precludes a definite assignment. A more interesting case is *Salisaeta* icosahedral phage 1 (SSIP-1) (Aalto *et al.* 2012). The halophilic bacteriophage displays a similar arrangement of core proteins, namely a P9 type packaging ATPase in close proximity upstream of the sequentially arranged two MCPs. The conserved block of core genes links this bacteriophage from hypersaline environments to the thermophilic P23-77-like phages (Aalto *et al.* 2012). However, a lack of similarities in sequences, gene order, and genome size as well as differences in the capsid architecture prevented us from classifying the virus within “*Sphaerolipoviridae*” at this stage of research.

6.1.4 “*Sphaerolipoviridae*” in the framework of the ICTV

The 2013 taxonomy release of the ICTV (ICTV 2014) recognizes two orders of prokaryotic viruses, *Caudovirales*, comprising viruses of bacteria and archaea with head-tail morphology, and *Ligamenvirales*, comprising two families of archaeal viruses with linear morphology (Table 4). The number of assigned

viruses of the order *Caudovirales* (*Myo*-, *Sipho*- and *Podoviridae*) refers to the relatively small number of viruses, for which sufficient data on virion structure, genome organization, and replication properties are available to classify them in the ICTV-system. It does not reflect the actual quantity and dominance of the head-tail viruses among prokaryotic viruses: About 6000 head-tail viruses have been described (97% of all reported bacterial viruses) (Ackermann and Prangishvili 2012) and 96% of the 1506 prokaryotic viral genomes published in the GenBank database (Benson *et al.* 2012) belong to the order *Caudovirales* (viewed 10.12.2014). Only 66 complete genome sequences (4%) contribute to archaeal viruses. Members of *Caudovirales* infect bacterial and archaeal hosts isolated from hot and hypersaline environments (Sabet 2012, Uldahl and Peng 2013). The only viruses officially listed by the ICTV are - to my knowledge - haloarchaeal myovirus ϕ H (Schnabel *et al.* 1982) and thermophilic siphovirus Ψ M1 (Meile *et al.* 1989), and *Halomonas* myophage ϕ HAP-1 (Mobberley *et al.* 2008). All three types of head-tail viruses infect halophilic bacteria and archaea (Atanasova *et al.* 2011, Sabet 2012, Shen *et al.* 2012). With respect to thermophiles, archaeal hosts are so far infected by siphoviruses only and podoviruses are lacking amongst bacteria (Uldahl and Peng 2013). Outside the *Caudovirales*, all assigned archaeal families to date comprise crenarchaeal hyperthermophilic viruses. "*Pleolipoviridae*" is a newly proposed family of euryarchaeal viruses from hypersaline environments with pleomorphic morphology and either dsDNA or ssDNA genome (Pietilä *et al.* 2012). The proposal is currently pending at the ICTV. The acceptance of "*Pleolipoviridae*" as an officially assigned family would be - to my knowledge - a novum in the history of the ICTV, which considers the nature of the genome (as in the Baltimore system) as a major classification criterium. The bacterial *Tectiviridae* family of tailless icosahedral viruses includes one thermophilic member, *Thermus* phage P37-14. Neither genomic sequence, nor a structural characterization is published for P37-14. Hence, its classification as a tectivirus is arguable and it might rather be a sphaerolipovirus. *Thermus* phage PH75 is the only described thermophilic inovirus (Pederson *et al.* 2001). The virus is not listed in the ICTV record.

In conclusion, only a vanishingly small number of extremophilic viruses are known to infect bacteria while the opposite is true for archaeal viruses. Here, all described viruses have been isolated from extreme habitats. The huge diversity of novel virus forms detected in hyperthermophilic environments is reflected in the accelerating number of novel archaeal virus families assigned by the ICTV within the last decade. Ten new families have been added. On the contrary, the taxonomy of bacteriophage families is unaltered for 30 years and newly assigned genera all fall into the existing ten families, mostly into the three *Caudovirales* families.

TABLE 4 Families of prokaryotic viruses according to the 2013 release of the ICTV (ICTV 2014). Families *Sipho*-, *Myo*- and *Podoviridae* belong to the order *Caudovirales* spanning the domains of bacteria and archaea. *Lipothrix*- and *Rudiviridae* belong to the archaeal order *Ligamenvirales* (assigned 2012). Species numbers refer to species listed by the ICTV. Squares indicate the presence of thermophilic (red) or halophilic (green) members in a given family.

Bacteria	species	assigned	Archaea	species	assigned
<i>Siphoviridae</i>	■ ■	86	1984	<i>Siphoviridae</i>	■ ■ 1 1984
<i>Myoviridae</i>	■ ■	30	1981	<i>Myoviridae</i>	■ 1 1981
<i>Podoviridae</i>	■	31	1981	<i>Podoviridae</i>	■ / 1981
<i>Plasmaviridae</i>		1	1978	<i>Lipothrixviridae</i>	■ 9 1990
<i>Corticoviridae</i>		1	1978	<i>Rudiviridae</i>	■ 3 1997
<i>Tectiviridae</i>	■	4	1978	<i>Fuselloviridae</i>	■ 1 1993
<i>Inoviridae</i>	■	43	1978	<i>Guttaviridae</i>	■ 1 2004
<i>Microviridae</i>		12	1978	<i>Globuloviridae</i>	■ 2 2008
<i>Cystoviridae</i>		1	1978	<i>Ampullaviridae</i>	■ 1 2008
<i>Leviviridae</i>		4	1978	<i>Bicaudaviridae</i>	■ 1 2009
				<i>Clavaviridae</i>	■ 1 2011
				<i>Spiraviridae</i>	■ 1 2013
				<i>Turriviridae</i>	■ 2 2013
				" <i>Pleolipoviridae</i> "	■ 8 not assigned
<i>Sphaerolipoviridae</i>	■ ■	2	2014*	<i>Sphaerolipoviridae</i>	■ ■ 4 2014*

* assigned according to the latest taxonomy release (ICTV 2015)

The unification of bacterial and archaeal viruses in the family "*Sphaerolipoviridae*" (Dyall-Smith *et al.* 2013b) is based rather on similarities in structure and genome arrangement of specific core genes than on actual sequence similarities, an approach not well accepted by the ICTV in the past. However, the recent establishment of the order *Ligamenvirales* indicates that the structural-based approach for defining common ancestry is henceforth accepted by the ICTV: a major criterion to unify viruses of the families *Lipothrix*- and *Rudiviridae* within a new order is the structural similarity of their major capsid proteins (Prangishvili and Krupovic, 2012). Shortly before the publication of this thesis, *Sphaerolipoviridae* was officially approved as a family and listed in the latest ICTV taxonomy release (ICTV 2015). It has become clear that focussing on sequence similarities is insufficient to discover higher order phylogenetic relationships. A combined approach of structural and genome comparison will help to bring more order to the expanding viral universe. By including the bacteriophage genus *Gammasphaerolipovirus* within *Sphaerolipoviridae* (III), *Sphaerolipoviridae* is the first family comprising bacteriophages assigned by the ICTV since the classification of bacteriophages with long non-contractile tails into the *Siphoviridae* family in 1984 (Table 4). Secondly, it is the only other family besides the three families of the order *Caudovirales* that spans two domains of life.

6.2 The role of the P23-77 minor capsid protein VP11 in capsid assembly

6.2.1 VP11 is excluded as a penton protein

By fitting the high resolution structures of VP16, VP17 and the VP16/VP17-complex into the P23-77 cryo-EM density map, it was possible to create a reliable capsid model, in which the whole capsid could be assembled from three building blocks, namely a VP16 dimer, a VP16-VP17 heterotrimer and a VP16-VP17 heterotetramer, revealing a unique assembly mode not observed before in viruses (Rissanen *et al.* 2013). According to the model, the two major capsid proteins are unevenly distributed in the capsid. The 5-fold vertices are exclusively surrounded by large major capsid protein VP17, excluded from the 3-fold axes by assembly frontiers (Rissanen *et al.* 2013). The vertices of some icosahedral viruses are occupied by penton proteins that exhibit the canonical jelly-roll fold (Abrescia *et al.* 2004, Zubieta *et al.* 2005, Abrescia *et al.* 2008, Veesler *et al.* 2013). The three-dimensional reconstruction of the virion indicated that this should be also expected in P23-77 (Jaatinen *et al.* 2008). The results obtained in this study (IV) clearly demonstrate that VP11 – although the only identified capsid-associated protein besides VP16 and VP17 (Jalasvuori *et al.* 2009) – is not the penton in P23-77. Firstly, far UV-spectra received from circular dichroism (CD) measurements revealed a high content of α -helices contradicting the presence of an eight stranded beta-barrel. Secondly, the estimated copy number of 147 (calculated from comparison of band intensities in Coomassie stained polyacrylamid gels) was too high to fit for a penton protein (60 copies in an icosahedral capsid with 12 vertices). Thirdly, the only oligomeric state we observed for VP11 was a dimer.

6.2.2 VP11 links the outer capsid shell to the internal membrane

The results received from lipid-protein interactions in DLS measurements (5.1.3, Fig. 3) in combination with the dissociation behavior of VP11 (Jalasvuori *et al.* 2009) and the lack of predicted transmembrane helices strongly suggest that the minor capsid protein VP11 serves as a linker between the internal membrane and the outer protein shell. The binding of VP11 to the outer surface of the membrane is primarily based on electrostatic attraction of positively-charged protein surface domains with anionic phospholipids as shown for several protein species (Mulgrew-Nesbitt *et al.* 2006). The highly positive surface charge of VP11 (pI = 10.71) also accounts for its capability to interact with other negatively-charged molecules such as DNA, yet DNA-binding is most likely not a biological function of VP11 in the virion. VP11 has no homologs in related virus and proviruses. Yet, related virus IN93 and provirus TP1 carry genes, located at similar positions with respect to ORF11 in P23-77, encoding highly positively charged, α -helical proteins (Fig. 1, IV). Those showed similarity to structural proteins of thermophilic archaeal viruses such as SSV-1 protein VP2,

STIV2 protein B72 and *Archaeoglobus veneficius* SNP6 putative provirus (Reiter *et al.* 1987, Happonen *et al.* 2010, Makarova *et al.* 2014). VP2 is proposed to function in genome packaging based on its capability to bind DNA (Reiter *et al.* 1987). The results presented here suggest a localization of VP11 underneath the capsid shell (separated from the genome by the membrane). This implies alternative functions for positively-charged proteins found in the virions of other lipid containing viruses infecting thermophiles.

The minor capsid protein interacts with the internal viral membrane on the one hand and with VP17, but not with VP16, on the other hand. The membrane-association of VP11 seems to be a prerequisite for the binding of the large major capsid protein. Capsid protein binding might involve the N- or C-terminus of VP17, both disordered in the crystal structure (Rissanen *et al.* 2013). Hydrophobic residues are thought to anchor the N-terminus of VP17 to the internal membrane (Rissanen *et al.* 2013) as observed for the MCPs of other double beta-barrel viruses (Abrescia *et al.* 2004, Cockburn *et al.* 2004, Khayat *et al.* 2010, Veessler *et al.* 2013). The N-terminus in turn is connected by a proline-rich linker to the rest of the protein (Rissanen *et al.* 2013). This linker may provide an interaction site with VP11 since proline rich regions are often involved in protein-protein interactions (Williamson 1994).

6.2.3 VP11 facilitates the capsid assembly

The biophysical characteristics of VP11 received from CD- and DLS-measurements strongly resemble those of scaffolding proteins of head-tail viruses (Prevelige and Fane 2012, Mateu 2013): a predominately α -helical structure, elongated shape, structural flexibility, and protein binding via electrostatic interactions. The scaffolding proteins of podoviruses P22 and Phi29 have a linear shape and high content of α -helices (Fuller and King, 1982, Tuma *et al.* 1998, Morais *et al.* 2003). Positive charges on the surface of the C-terminal helix of the P22 scaffolding protein mediate the electrostatic interaction with the major capsid protein (Cortines *et al.* 2011). Experiments with a P22 scaffolding protein mutant, which formed disulfide cross-linked dimers, revealed a significantly faster assembly than with wild-type protein, implying an active role for dimers in capsid assembly (Parker *et al.* 1997). The same might apply for the role of VP11 dimers in the P23-77 assembly. Scaffolding proteins act as chaperones in capsid assembly and are removed from the procapsid before or during genome packaging (Prevelige and Fane 2012). In contrast to those, VP11 is not removed during capsid assembly and constitutes a significant component of the mature virion. Thus, VP11 is not a scaffolding protein *sensu stricto*. However, the similarities point to a function in assisting the correct assembly of the capsid by facilitating the binding of capsid components. In some double beta-barrel viruses, capsid and membrane are linked by the termini of the MCPs or integral membrane proteins (2.5). P23-77 viral particles are exposed to extremely high temperatures. Thus, additional factors are required to enhance the rigidity of the capsid. By tightly linking the capsid shell to the membrane, VP11 might represent such a factor.

6.2.4 The P23-77 capsid assembly

The mechanism of genome encapsidation is unknown. It is questionable if the capsid assembly of P23-77 includes a procapsid stage. Empty particles are rarely seen in virus preparations (Jaatinen *et al.* 2008) and there is so far no evidence for the presence of a special vertex for packaging the genome. Genome packaging into empty procapsids occurs predominantly in viruses with linear dsDNA genome (Earnshaw *et al.* 1980, Hong *et al.* 2014). The packaging mechanism might be different for viruses with circular genome. In marine bacteriophage PM2, the closest relative of P23-77 with respect to the MCP protein fold, the capsid is assumed to assemble around the pre-condensed circular dsDNA genome (Abrescia *et al.* 2008). P23-77 has a circular dsDNA genome as well. Thus, the assembly may also start around the already pre-condensed genome or around a partially condensed genome with further condensation taking place during the assembly process. However, the genome of P23-77 encodes the canonical P9-type genome packaging ATPase (Iyer *et al.* 2004, Strömsten *et al.* 2005), which is common to all membrane containing icosahedral dsDNA viruses and one of the elements we used to define the relatedness of sphaerolipoviruses (I, III). In PM2, the ATPase P9 is a structural component of the virion (Männistö *et al.* 1999, Kivelä *et al.* 2002). Consequentially, one should expect an ATPase-driven packaging of the genome into an empty procapsid as exemplified in enterophage PRD1 (Gowen *et al.* 2003, Strömsten *et al.* 2005). Recently, the ATPase of double beta-barrel virus STIV2 was characterized (Happonen *et al.* 2013). The enzyme can bind circular DNA, but only linear DNA stimulated the hydrolysis of ATP. Thus, further experiments are needed to elucidate the role of such ATPase in the packaging of circular genomes. The virus lipids are acquired selectively from the host membrane (Jalasvuori *et al.* 2009). As in hyperthermophilic virus STIV (Rice *et al.* 2004, Veesler *et al.* 2013), a transmembrane channel beneath the vertices is the only visible connection in cryo-EM images, suggesting the vertices as starting points of the capsid shell assembly (Veesler *et al.* 2013, Rissanen *et al.* 2013). Thus, the tight anchoring of a yet unknown vertex complex to the newly formed membrane might serve as a nucleation point for the capsid assembly, with VP11 functioning as a trigger. Electrostatic forces enable binding of minor capsid protein VP11 to the membrane. VP11 dimers form a mesh along the outer membrane surface on which major capsid protein VP17 assembles. The assembly starts at the vertices, which are exclusively surrounded by VP17 (Rissanen *et al.* 2013). Binding of VP17 in turn attracts dimers of small major capsid protein VP16 to bind to VP17. The correct assembly of the various capsid building blocks is facilitated by specific interaction of the minor capsid protein VP11 with VP17, but not with VP16. Capsid reinforcement is probably achieved by strong disulfide bond formation in VP11 dimers, binding of the VP17 N-terminus to the membrane, membrane connections at the vertices and electrostatic forces between the minor capsid protein and lipids as well as between the major capsid proteins in the protein shell. The role of the minor capsid protein VP11 in capsid assembly is regarded as an assisting one. It

facilitates the correct incorporation of the two major capsid proteins and enhances rigidity at high temperatures rather than forming a scaffold analogous to scaffolding proteins of tailed phages. This function is most likely fulfilled by the internal membrane and integrated proteins therein. The complex capsid assembly involving various building blocks of two proteins instead of one stable trimeric protein appears to be more error prone and subsequently not as efficient as the double beta-barrel way of assembly. This might explain why this form of capsid structure is so far exclusively found in viruses from extreme environments while the double beta-barrel viruses are widely spread throughout the biosphere.

7 CONCLUSION

The main findings and contributions of the thesis are:

- The two major capsid proteins of P23-77 were successfully expressed in *E. coli* and purified to homogeneity using chromatographic methods. Optimization of crystallization trials yielded in the production of five well diffracting crystals (maximum resolution between 1.26–2.92 Å) derived from the small MCP (from recombinant protein and from virion material), the large MCP, and a complex of the two MCPs. The crystals provided the essential prerequisite for the determination of the MCP structures by X-ray crystallography (Rissanen *et al.* 2013), which defined P23-77 as an archetype of an ancient viral lineage.
- The comparative analysis of the genomes and the available structural information of members of the P23-77- like viral lineage was combined to create a new viral genus *Gammasphaerolipovirus* within the novel family of “*Sphaerolipoviridae*”. Comparative genomics of viral self elements revealed the presence of several sphaerolipovirus-related proviral sequences in the genomes of thermophilic bacteria and halophilic archaea. The detection of proviral sequences in the genomes of *Meiothermus* species expands the distribution of the family on the bacterial branch from extreme thermophilic (~75 °C) to moderate thermophilic environments (~55 °C).
- With the addition of a bacteriophage genus to the existing archaeal virus genera, *Sphaerolipoviridae* is now the only family comprising viruses from two domains of life besides the *Caudovirales* families of tailed icosahedral dsDNA viruses. Moreover, it is the first bacteriophage family proposed for over 30 years. The unification of the increasing number of so far unclassified P23-77-like viruses within a distinct taxon will help scientists dealing with such viruses. Novel viruses of the same type can be easily assigned to the family.

- The minor capsid protein VP11, the third capsid-associated component of the P23-77 virion, was characterized in detail. VP11 is a highly heat stable α -helical, dimeric protein with a proposed flexible structure. It interacts with *Thermus* derived lipids and the large major capsid protein VP17. The protein-lipid interaction is mainly based on electrostatic forces. VP11 is localized between the outer shell and the inner membrane, where it might stabilize the capsid and facilitate capsid assembly. The minor capsid protein presumably acts as an assembly factor, controlling the incorporation of the various capsid building blocks by the selective interaction with the large but not the small major capsid protein.

Acknowledgements

This work was carried out at the University of Jyväskylä at the Department of Biological and Environmental Science between the years 2008-2015. The work was funded by the Academy of Finland Centre of Excellence (CoE) Program in Virus Research (2006-2011, #1129648), CoE in Biological Interactions (2012-2017, #252411), and an Academy of Finland personal grant to Professor Jaana Bamford (#251106).

This thesis work would not exist without the support of many people. First of all, I would like to express my greatest gratitude to my supervisor Professor Jaana Bamford, who introduced me to the world of bacteriophages and always supported me throughout this project. Thanks for your trust. I like to express my gratitude to Dr. Kenneth Stedman for accepting the invitation to act as opponent of my dissertation. I also like to thank the pre-examiners of my thesis, Dr. Sarah Butcher and Dr. Petri Auvinen, for their careful and critical revision of the thesis, and their valuable comments. I am very glad that I had such great support group members as Prof. Janne Ihalainen and Dr. Vesa Hytönen. Thank you so much for your advice, help and patience during the VP11 project.

Work takes up a considerable part of life. I am happy that working in the JB-lab was not just doing research but also fun and inspiration, thanks to my wonderful colleagues. Ilona and Matti, together we brought the work on "Petey" to this stage, thank you for all the fruitful discussions and help. Petri, thanks for your invaluable advice in the lab and the insights into marathon training and the mysteries of geo-cashing. My dear office room-mates during these years, Sari and Sari, Nadine, Hanna, and Anni-Maria - thank you for sharing all the ups and downs in the lab, for lifting up my mind whenever needed and all the nice talks. Thanks to the other JB-group members Elina, Reeta, Jenni, Ville and Lotta-Reena. I will always remember the great atmosphere of our group-meetings in Konnevesi and Varjola. Thanks also to my students Jenni Viholainen, Joanna Lempiäinen, and Anni Moilanen for the help with my research.

My deepest gratitude goes to my family. Jan, you brought us to Finland, thus this thesis would not have been possible without you. There are not enough words to explain how your endless love, trust and encouragement has helped me always and especially during the last months. Even in my worst moments you know how to make me laugh. We had a great time in Finland, not least because our daughter was born here. Maija, when I look at you I am so glad that science cannot explain all the wonders in the world.

YHTEENVETO (RÉSUMÉ IN FINNISH)

***Thermus*-bakteriofagi P23-77: muinaisen ääriolosuhteiden virussuvun merkittävä jäsen**

Virustutkimus on perinteisesti keskittynyt taudinaiheuttajiin, jotka tartuttavat ihmisiä, karjaa ja viljelykasveja. Aitotumalliset eläimet, kasvit ja sienet edustavat kuitenkin vain pientä osaa biosfäärin eliöistä; suurin osa kaikista lajeista kuuluu esitumallisiin bakteereihin ja arkkeihin. Näiden eliöiden viruksia onkin valtavasti maapallolla, huomattavasti enemmän kuin solullisia organismeja. Viimeaikainen tutkimus on nostanut esiin esitumallisia infektoivien virusten merkityksen ekosysteemeille ja niiden vaikutuksen ilmastoon sekä globaaleihin energia- ja hiilisykleihin. Virukset ovat myös vaikuttaneet elämän kehitykseen elämän synnyn ajoista alkaen. Ne ovat kehittyneet yhdessä isäntiensä kanssa miljardien vuosien ajan ja olleet merkittävä solullisen elämän kehitystä ajava tekijä. Tämä tutkimus käsittelee muinaista tyyppiä edustavan, bakteereja infektoivan kuumien lähteiden viruksen rakenteita ja sen sukulaisuussuhteita muihin nykypäivän viruksiin.

Kuumien lähteiden lajikirjo on tyypillisesti muita vesistöjä kapeampi. Näistä ympäristöistä eristetyt virukset ilmentävät kuitenkin suurempaa muotojen eli morfologioiden kirjoa kuin missään muualla on tavattu. Tämä yllättävä löytö viittaa siihen, että kuumien lähteiden kaltaisissa ääriolosuhteissa voi selviytyä virustyyppiä, joita ei enää esiinny missään muualla. Ääriolosuhteiden suojaavan vaikutuksen esitetään perustuvan niiden samankaltaisuuteen muinaisten elinympäristöjen kanssa sekä niissä vallitsevaan negatiiviseen valintapaineeseen, joka voi hidastaa eliöiden erilaistumista. Ääriolosuhteiden virukset ovat suurelta osin tieteelle tuntemattomia tai heikosti karakterisoituja.

Uuden-Seelannin kuumista lähteistä eristetty virus, bakteriofagi P23-77, infektoi *Thermus*-suvun bakteereita. P23-77 koostuu ikosaedrin muotoisesta kuoresta eli kapsidista, joka sulkee sisäänsä lipidikalvon ja viruksen kaksijuosteisen DNA-genomin. P23-77:n kapsidi rakentuu pääosin kahdesta kapsidiproteiinista, joiden järjestäytyminen viruksen kuoressa poikkeaa aiemmin havaituista kapsidiarkkitehtuureista.

Tässä tutkimuksessa raportoidaan useiden ääriolosuhteiden eliöiden, mukaanlukien korkeassa lämpötilassa esiintyvien bakteerien ja korkeassa suolapitoisuudessa esiintyvien arkkien, kantavan genomissaan jäänteitä bakteriofagi P23-77:n kaltaisista viruksista. Näiden havaintojen perusteella esitetään bakteriofagin P23-77 ja genomisten viruselementtien kuuluvan ennen tuntemattomaan virussukulinjaan, jonka jäsenet kykenevät infektoimaan sekä arkkeja että bakteereja. Jatkotutkimuksessa havaittiin, että morfologisiin ja geneettisiin piirteisiin, erityisesti kolmeen ydingeeniin (kaksi kuoriproteiinia ja ATPaasi), perustuen bakteriofagi P23-77 lukeutuu virusheimoon *Sphaerolipoviridae*. Heimon sisällä P23-77:n piirteet määrittelevät uuden suvun *Gammasphaerolipovirus*. Ehdotus uudesta virussuvusta on toimitettu virusten luokittelusta vastaavalle elimelle, ICTV:lle (International Committee on Taxonomy of Viruses).

Väitöstutkimuksessa selvitettiin myös menetelmät P23-77:n kahden pääasiallisen kapsidiproteiinin (VP16 ja VP17) tuottamiseen, puhdistamiseen ja kiteyttämiseen röntgensädekristallografista rakennetutkimusta varten. Kiteytettyjen kapsidiproteiinien rakenteet osoittivat P23-77:n kuuluvan vertikaalisten beta-tynnyrivirusten sukulinjaan ja ilmentävän piirteitä, jotka todennäköisesti muistuttavat sukulinjan muinaista esi-isää. Kapsidiproteiinien VP16 ja VP17 lisäksi tutkimuksessa kehitettiin menetelmät kapsidiproteiinin VP11 puhdistamiseen, määritettiin proteiinin VP11 ominaisuudet ja selvitettiin sen merkittävä rooli viruksen kapsidin rakentumisessa. VP11 on erittäin lämpöstabiili, helikaalinen ja dimeerejä muodostava rakenneproteiini, joka kykenee sitoutumaan *Thermus*-suvun bakteereista eristettyihin lipideihin sekä kapsidiproteiiniin VP17. VP11 liittyy toisiinsa viruksen sisäisen lipidikalvon ja kapsidiproteiineista VP16 ja VP17 muodostuvan kuoren. Nämä ominaisuudet viittaavat siihen, että VP11 stabiloi kapsidin rakenteen ja ohjaa sen rakentumista viruksen monistumisen aikana. VP11 edustaa aiemmin tuntematonta proteiinityyppiä, jolla on merkittävä rooli ikosaedraalisten, lipidikalvon sisältävien virusten kapsidin rakentumisessa. Kokonaisuudessaan väitöstutkimus tarjoaa laajan geneettisen ja biokemiallisen analyysin uudesta ääriolosuhteiden virussuvusta ja sen tyyppilajista, bakteriofagista P23-77.

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.
- 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., Fry E.E., Ravantti J.J., Bamford D.H. & Stuart D.I. 2010. What does it take to make a virus: the concept of the viral “self”. In: Stockley P.G., Twarock R. (eds.), *Emerging topics in physical virology*, Imperial College Press, London, pp. 35–58.
- Abrescia N.G., Grimes J.M., Kivelä 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.
- Ackermann H.W. & Prangishvili D. 2012. Prokaryote viruses studied by electron microscopy. *Arch Virol* 157: 1843–1849.
- Adams M.J., Lefkowitz E.J., King A.M.Q. & Carstens E.B. 2013. Recently agreed changes to the International Code of Virus Classification and Nomenclature. *Arch Virol* 158: 2633–2639.
- Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Argos P., Kamer G., Nicklin M.J. & Wimmer E. 1984. Similarity in gene organization and homology between proteins of animal picomaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucleic Acids Res* 12: 7251–7267.
- Atanasova N.S., Roine E., Oren A., Bamford D.H. & Oksanen H.M. 2012. Global network of specific virus-host interactions in hypersaline environments. *Environ Microbiol* 14: 426–440
- Athappily 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 Å 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.
- Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol Rev* 353: 235.
- 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.
- Bamford D.H., Grimes J.M. & Stuart D.I. 2005a. What does structure tell us about virus evolution? *Curr Opin Struct Biol* 15: 655–663.
- Bamford D.H., Ravantti J.J., Rönnholm G., Laurinavicius S., Kukkaro P., Dyal-Smith M., Somerharju P., Kalkkinen N. & Bamford J.K. 2005b. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J Virol* 79: 9097–9107.
- Bamford J.K., Cockburn J.J., Diprose J., Grimes J.M., Sutton G., Stuart D.I., & Bamford D.H. 2002. Diffraction quality crystals of PRD1, a 66-MDa dsDNA virus with an internal membrane. *J Struct Biol* 139: 103–112.
- 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.
- Beeby M.D., O'Connor B., Ryttersgaard C., Boutz D.R., Perry L.J. & Yeates T.O. 2005. The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol* 3: e309.
- Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. Sayers E.W. 2012. GenBank. *Nucleic Acids Res* gks1195.
- 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.
- Benson S.D., Bamford J.K., Bamford D.H. & Burnett R.M. 2004. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol Cell* 16: 673–685.
- Blondal T., Hjorleifsdottir S., Ævarsson A., Fridjonsson O.H., Skirnisdottir S., Wheat J.O., Hermannsdottir A.G., Hreggvidsson G.O., Smith A.V. & Kristjansson J.K. 2005a. Characterization of a 5'-polynucleotide kinase/3'-phosphatase from bacteriophage RM378. *J Biol Chem* 280: 5188–5194.
- Blondal T., Thorisdottir A., Unnsteinsdottir U., Hjorleifsdottir S., Ævarsson A., Ernstsson S., Fridjonsson O.H., Skirnisdottir S., Wheat J.O., Hermannsdottir A.G., Sigurdsson S. Th., Hreggvidsson G.O., Smith A.V. & Kristjansson J.K. 2005b. Isolation and characterization of a thermostable RNA ligase 1 from a *Thermus scotoductus* bacteriophage TS2126 with good single-stranded DNA ligation properties. *Nucleic Acids Res* 33: 135–142.
- Brüssow H. & Hendrix R.W. 2002. Phage genomics: small is beautiful. *Cell* 108: 13–16.
- Butcher S.J., Bamford D.H. & Fuller, S.D. 1995. DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J* 14: 6078.

- Butcher S.J., Manole V. & Karhu N.J. 2012. Lipid-containing viruses: bacteriophage PRD1 assembly. In: *Viral Molecular Machines*. Rossmann M.G., Rao V.B. (eds.), Springer, US, pp. 365–377.
- Canchaya C., Fournous G. & Brüssow H. 2004. The impact of prophages on bacterial chromosomes. *Mol Microbiol* 53: 9–18
- Caspar D.L. & Klug A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27: 1–24.
- Chen X.S., Garcea R.L., Goldberg I., Casini G. & Harrison S.C. 2000. Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell* 5: 557–567.
- Clokic M.R., Millard A.D., Letarov A.V. & Heaphy S. 2011. Phages in nature. *Bacteriophage* 1: 31–45.
- Cockburn J.J., Abrescia N.G., Grimes J.M., Sutton G.C., Diprose J.M., Benevides, J.M., Thomas G.J. Jr., Bamford J.K., Bamford D.H. & Stuart D.I. 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* 432: 122–125.
- Comeau A.M., Hatfull G.F., Krisch H.M., Lindell D., Mann N.H. & Prangishvili D. 2008. Exploring the prokaryotic virosphere. *Res Microbiol* 159: 306–313.
- Cortez D., Forterre P. & Gribaldo S. 2009. A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol* 10: R65
- Cortines J.R., Weigele P.R., Gilcrease E.B., Casjens S.R. & Teschke C.M. 2011. Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. *Virology* 421: 1–11.
- Dokland T., McKenna R., Ilag L.L., Bowman B.R., Incardona N.L., Fane B.A., & Rossmann M.G. 1997. Structure of a viral procapsid with molecular scaffolding. *Nature* 389: 308–313.
- Drake J.W. 2009. Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS Genet* 5: e1000520.
- Dyall-Smith M., Porter K., Tang S.L. 2013a. 2013.001a-kB.N.v2.Sphaerolipoviridae. http://talk.ictvonline.org/files/proposals/taxonomy_proposals_prokaryote1/m/bact01/4633.aspx viewed 03.09.2013.
- Dyall-Smith M., Porter K., Tang S.L., Pawlowski A., Rissanen I., Bamford J.K.H., Krupovic M. & Jalasvuori M. 2013b. 2013.001a-oB.A.v5.Sphaerolipoviridae. http://talk.ictvonline.org/files/proposals/taxonomy_proposals_prokaryote1/m/bact04/5040.aspx. viewed 12.12.2014.
- Earnshaw W.C. & Casjens S.R. 1980. DNA packaging by the double-stranded DNA bacteriophages. *Cell* 21: 319–331.
- El Omari K., Sutton G., Ravantti J.J., Zhang H., Walter T.S., Grimes J.M., Bamford D.H., Stuart D.I. & Mancini, E. J. (2013). Plate tectonics of virus shell assembly and reorganization in phage $\phi 8$, a distant relative of mammalian reoviruses. *Structure* 21: 1384–1395.
- Filée J., Baptiste E., Susko E. & Krisch H.M. 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved

- a core genome with the viral replication and structural genes. *Mol Biol Evol* 23: 1688–1696.
- Fokine A. & Rossmann M.G. 2014. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4, e28281.
- 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.
- Fu C.Y., Wang K., Gan L., Lanman J., Khayat R., Young M. J., Grant J.J., Doerschuk P.C. & Johnson, J. E. 2010. In vivo assembly of an archaeal virus studied with whole-cell electron cryotomography. *Structure* 18: 1579–1586.
- Fuller M.T. & King J. 1982. Assembly *in vitro* of bacteriophage P22 procapsids from purified coat and scaffolding subunits. *J Mol Biol* 156: 633–665.
- 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, *in vivo*. *J Virol* 76: 5915–5924.
- Goulet A., Blangy S., Redder P., Prangishvili D., Felisberto-Rodrigues C., Forterre P., Campanacci V. & Cambillau C. 2009. *Acidianus* filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. *Proc Natl Acad Sci U S A* 106: 21155–21160.
- Gowen B., Bamford J.K., Bamford D.H. & Fuller S.D. 2003. The tailless icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. *J Virol* 77: 7863–7871.
- Grimes J.M., Burroughs J.N., Gouet P., Diprose J.M., Malby R., Zióntara S., Mertens P.P. & Stuart D.I. 1998. The atomic structure of the bluetongue virus core. *Nature* 395: 470–478.
- Guixa-Boixareu N., Calderón-Paz J.I., Heldal M., Bratbak G. & Pedrós-Alió C. 1996. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat Microb Ecol* 11: 215–227.
- Hacene H., Rafa F., Chebhouni N., Boutaiba S., Bhatnagar T., Baratti J.C. & Ollivier B. 2004. Biodiversity of prokaryotic microflora in El Golea salt lake, Algerian Sahara. *J Arid Environ* 58: 273–284.
- Hall N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* 210: 1518–1525.
- Happonen L.J., Oksanen E., Liljeroos L., Goldman A., Kajander T. & Butcher S.J. 2013. The structure of the NTPase that powers DNA packaging into *Sulfolobus* turreted icosahedral virus 2. *J Virol* 87: 8388–8398.
- Happonen L.J., Redder P., Peng X., Reigstad L.J., Prangishvili D. & Butcher S.J. 2010. Familial relationships in hyperthermo- and acidophilic archaeal viruses. *J Virol* 84: 4747–4754.
- Häring M., Vestergaard G., Rachel R., Chen L., Garrett R.A. & Prangishvili D. 2005. Virology: independent virus development outside a host. *Nature* 436: 1101–1102.
- Hendrix R.W. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol* 61: 471–480.

- Hendrix R.W. & Duda R.L. 1998. Bacteriophage HK97 head assembly: a protein ballet. *Adv Virus Res* 50: 235–288.
- 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.
- Hong C., Oksanen H.M., Liu X., Jakana J., Bamford D.H. & Chiu W. 2014. A Structural Model of the Genome Packaging Process in a Membrane-Containing Double Stranded DNA Virus. *PLoS Biol* 12: e1002024.
- Hreggvidsson G.O., Petursdottir S.K., Björnsdottir S.H. & Fridjonsson O.H. 2012. Microbial speciation in the geothermal ecosystem. In: Stan-Lotter H., Fendrihan S. (eds.), *Adaption of microbial life to environmental extremes*, Springer, Vienna, pp. 37–67.
- Huet A., Conway J.F., Letellier L. & Boulanger P. 2010. In vitro assembly of the T= 13 procapsid of bacteriophage T5 with its scaffolding domain. *J Virol* 84: 9350–9358.
- Huiskonen J.T., Kivelä H.M., Bamford D.H. & Butcher S.J. 2004. The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol* 11: 850–856.
- Huiskonen J.T., de Haas F., Bubeck D., Bamford D.H., Fuller S.D. & Butcher S.J. 2006. Structure of the bacteriophage $\phi 6$ nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* 14: 1039–1048.
- Hurwitz B.L., Deng L., Poulos B.T. & Sullivan M.B. 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* 15: 1428–1440.
- ICTV 2014. *Virus Taxonomy: 2013 Release*. <http://www.ictvonline.org/virusTaxonomy.asp>. viewed 04.11.2014.
- ICTV 2015. *Virus Taxonomy: 2014 Release*. <http://www.ictvonline.org/virusTaxonomy.asp>. viewed 14.03.2015.
- 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.
- Iyer L.M., Makarova K.S., Koonin E.V. & Aravind L. 2004. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res* 32: 5260–5279.
- Jaakkola S.T., Penttinen R.K., Vilén S.T., Jalasvuori M., Rönnholm 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.
- Jääliñoja H.T., Roine E., Laurinmäki P., Kivelä 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., Laurinmäki 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.
- Jaatinen S.T., Viitanen S.J., Bamford D.H. & Bamford J.K. 2004. Integral membrane protein P16 of bacteriophage PRD1 stabilizes the adsorption vertex structure. *J Virol* 78: 9790–9797.
- Jalasvuori M., Jaatinen S., Laurinavičius 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.
- Jiang S., Steward G., Jellison R., Chu W. & Choi S. 2004. Abundance, distribution, and diversity of viruses in alkaline, hypersaline Mono Lake, California. *Microb Ecol* 47: 9–17.
- Jorda J. & Yeates T.O. 2011. Widespread disulfide bonding in proteins from thermophilic archaea. *Archaea* 2011: 409156.
- 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.
- King A.M., Adams M.J., Lefkowitz E.J. & Carstens E.B. (Eds.) 2012. *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier, San Diego.
- King J., Lenk E.V. & Botstein D. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22: II. Morphogenetic pathway. *J Mol Biol* 80: 697–731.
- Kivelä H.M., Kalkkinen N. & Bamford D.H. 2002. Bacteriophage PM2 has a protein capsid surrounding a spherical proteinaceous lipid core. *J Virol* 76: 8169–8178.
- Kivelä H.M., Roine E., Kukkaro P., Laurinavičius S., Somerharju P. & Bamford D.H. 2006. Quantitative dissociation of archaeal virus SH1 reveals distinct capsid proteins and a lipid core. *Virology* 356: 4–11.
- Koonin E.V. & Dolja V.V. 2013. A virocentric perspective on the evolution of life. *Curr Opin Virol* 3: 546–557.
- Kristensen D.M., Cai X. & Mushegian A. 2011. Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. *J Bacteriol* 193: 1806–1814.
- Kristensen D.M., Waller A.S., Yamada T., Bork P., Mushegian A.R. & Koonin E.V. 2013. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J Bacteriol* 195: 941–950.
- Krupovic M. & Bamford D.H. 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics* 8: 236.

- Krupovic M. & Bamford D.H. 2008a. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum *Euryarchaeota*. *Virology* 375: 292–300.
- Krupovic M. & Bamford D.H. 2008b. 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., Prangishvili D., Hendrix R.W. & Bamford D.H. 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* 75: 610–635.
- Laidler J.R., Shugart J.A., Cady S.L., Bahjat K.S. & Stedman K.M. 2013. Reversible inactivation and desiccation tolerance of silicified viruses. *J Virol* 87: 13927–13929.
- Larson E.T., Eilers B., Menon S., Reiter D., Ortmann A., Young M.J., & Lawrence C.M. 2007. A winged-helix protein from *Sulfolobus* turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus. *Virology* 368: 249–261.
- Laurinavicius S., Käkälä R., Somerharju P. & Bamford D.H. 2004. Phospholipid molecular species profiles of tectiviruses infecting Gram-negative and Gram-positive hosts. *Virology* 322: 328–336.
- Leiman P.G., Kanamaru S., Mesyanzhinov V.V., Arisaka F. & Rossmann M.G. 2003. Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* 60: 2356–2370.
- Le Romancer M., Gaillard M., Geslin C. & Prieur D. 2007. Viruses in extreme environments. *Rev Environ Sci Biotechnol* 6: 17–31
- Lin T., Chen Z., Usha R., Stauffacher C.V., Dai J.B., Schmidt T. & Johnson J.E. 1999. The refined crystal structure of cowpea mosaic virus at 2.8 Å resolution. *Virology* 265: 20–34.
- Lin L., Hong W., Ji X., Han J., Huang L. & Wei Y. 2010. Isolation and characterization of an extremely long tail *Thermus* bacteriophage from Tengchong hot springs in China. *J Basic Microbiol* 50: 452–456.
- 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.
- Makarova K.S., Wolf Y.I., Forterre P., Prangishvili D., Krupovic M., & Koonin E.V. 2014. Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. *Extremophiles* 18: 877–893.
- Männistö R.H., Kivelä H.M., Paulin L., Bamford D.H. & Bamford J.K. 1999. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. *Virology* 262: 355–363.
- Mateu M.G. 2013. Assembly, stability and dynamics of virus capsids. *Arch Biochem Biophys* 531: 65–79.
- Matsushita I. & Yanase H. 2008. A novel thermophilic lysozyme from bacteriophage φIN93. *Biochem Biophys Res Commun* 377(1), 89–92.

- Matsushita I. & Yanase H. 2009. The genomic structure of *Thermus* bacteriophage ϕ IN93. *J Biochem* 146: 775–785.
- Matsushita I., Yamashita N. & Yokota A. 1995. Isolation and characterization of bacteriophage induced from a new isolate of *Thermus aquaticus*. *Microbiol Cult Collect* 11:133–138.
- 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 Å resolution. *J Mol Biol* 397: 587–599.
- Meile L., Jenal U., Studer D., Jordan M. & Leisinger T. 1989. Characterization of ψ M1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. *Arch Microbiol* 152: 105–110.
- Menon S.K., Maaty W.S., Corn G.J., Kwok S.C., Eilers B.J., Kraft P., Gillitzer E., Young M.J., Bothner B. & Lawrence C.M. 2008. Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology* 376: 270–278.
- Minakhin L., Goel M., Berdygulova Z., Ramanculov E., Florens L., Glazko G., Karamychev V.N., Slesarev A.I., Kozyavkin S.A., Khromov I., Ackermann H.W., Washburn M., Mushegian A. & Severinov K. 2008. Genome comparison and proteomic characterization of *Thermus thermophilus* bacteriophages P23-45 and P74-26: siphoviruses with triplex-forming sequences and the longest known tails. *J Mol Biol* 378:468–480.
- Mindich L., Bamford D., McGraw T. & Mackenzie G. 1982. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J Virol* 44: 1021–1030.
- Mobberley J.M., Authement R.N., Segall A.M. & Paul J.H. 2008. The temperate marine phage Φ HAP-1 of *Halomonas aquamarina* possesses a linear plasmid-like prophage genome. *J Virol* 82: 6618–6630.
- Mochizuki T., Krupovic M., Pehau-Arnaudet G., Sako Y., Forterre P. & Prangishvili D. 2012. Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc Natl Acad Sci U S A* 109: 13386–13391.
- Morais M.C., Kanamaru S., Badasso M.O., Koti J.S., Owen B.A., McMurray C.T., Anderson D.L. & Rossmann M.G. 2003. Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nat Struct Biol* 10: 572–576.
- Moser M.J., DiFrancesco R.A., Gowda K., Klingele A.J., Sugar D.R., Stocki S., Mead D.A. & Schoenfeld T.W. 2012. Thermostable DNA polymerase from a viral metagenome is a potent rt-PCR enzyme. *PloS one* 7: e38371.
- Mulgrew-Nesbitt A., Diraviyam K., Wang J., Singh S., Murray P., Li Z., Rogers L., Mirkovic N. & Murray D. 2006. The role of electrostatics in protein-membrane interactions. *Biochim Biophys Acta* 1761: 812–826.
- Naitow H., Tang J., Canady M., Wickner R.B. & Johnson J.E. 2002. LA virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nat Struct Biol* 9: 725–728.
- Nakagawa A., Miyazaki N., Taka J., Naitow H., Ogawa A., Fujimoto Z., Mizuno H., Higashi T., Watanabe Y., Omura T., Cheng H.R. & Tsukihara, T. 2003.

- The atomic structure of Rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* 11: 1227–1238.
- 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.
- Nobre M.F., Trüper H.G & Da Costa M.S. 1996. Transfer of *Thermus ruber* (Loginova *et al.* 1984), *Thermus silvanus* (Tenreiro *et al.* 1999), and *Thermus chliarophilus* (Tenreiro *et al.* 1995) to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., respectively, and emendation of the genus *Thermus*. *Int J Sys Bacteriol* 46: 604–606
- Oksanen H., Pietilä M., Sencilo A., Atanasova N., Roine E. & Bamford D. 2012. Virus universe: can it be constructed from a limited number of viral architectures. In: Witzany G. (ed.), *Viruses: essential agents of life*, Springer, Berlin, pp. 83–105.
- Oren A., Bratbak G. & Haldal M. 1997. Occurrence of virus-like particles in the Dead Sea. *Extremophiles* 1: 143–149.
- Ostrander D.A. & Grey H.B. Jr. 1974. Superhelix density heterogeneity in closed circular intracellular PM2 DNA. *Biopolymers* 13: 955–975.
- Pace N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276: 734–740.
- Parker M.H. & Prevelige Jr. P.E. 1998. Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. *Virology* 250: 337–349.
- Parker M.H., Stafford W.F. 3rd & Prevelige Jr. P.E. 1997. Bacteriophage P22 scaffolding protein forms oligomers in solution. *J Mol Biol* 268: 655–665.
- Pederson D.M., Welsh L.C., Marvin D.A., Sampson M., Perham R.N., Yu M. & Slater M.R. 2001. The protein capsid of filamentous bacteriophage PH75 from *Thermus thermophilus*. *J Mol Biol* 309: 401–421.
- Peng X., Basta T., Häring M., Garrett R. A. & Prangishvili D. 2007. Genome of the *Acidianus* bottle-shaped virus and insights into the replication and packaging mechanisms. *Virology* 364: 237–243.
- Pfister P., Wasserfallen A., Stettler R. & Leisinger T. 1998. Molecular analysis of *Methanobacterium* phage Ψ M2. *Mol Microbiol* 30: 233–244.
- Pietilä M.K., Demina T.A., Atanasova N.S., Oksanen H.M. & Bamford D.H. 2014. Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* 22: 334–344
- Pietilä M.K., Roine E., Paulin L., Kalkkinen N. & Bamford D.H. 2009. An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol Microbiol* 72: 307–319.
- Pietilä 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.

- Pietilä M.K., Laurinmäki P., Russell D.A., Ko C.C., Jacobs-Sera D., Hendrix R.W., Bamford D.H. & Butcher S.J. 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proc Natl Acad Sci U S A* 110: 10604–10609.
- Pina M., Bize A., Forterre P. & Prangishvili D. 2011. The archeoviruses. *FEMS Microbiol Rev* 35: 1035–1054.
- Porter, K. & Dyall-Smith, M.L. 2008. Transfection of haloarchaea by the DNAs of spindle and round haloviruses and the use of transposon mutagenesis to identify non-essential regions. *Mol Microbiol* 70: 1236–1245.
- Porter K., Kukkaro P., Bamford J.K., Bath C., Kivelä H.M., Dyall-Smith M.L. & Bamford D.H. 2005. SH1: a novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology* 335: 22–33.
- Porter K., Russ B.E., Yang J. & Dyall-Smith M.L. 2008. The transcription programme of the protein-primed halovirus SH1. *Microbiology* 154: 3599–3608.
- Porter K., Tang S.L., Chen C.P., Chiang P.W., Hong M.J. & Dyall-Smith M. 2013. PH1: An archaeovirus of *Haloarcula hispanica* related to SH1 and HHIV-2. *Archaea* 2013: 456318.
- Prangishvili D. & Krupovic M. 2012. A new proposed taxon for double-stranded DNA viruses, the order “Ligamenvirales”. *Arch Virol*: 791–795.
- Prangishvili D. & Quax T.E. 2011. Exceptional virion release mechanism: one more surprise from archaeal viruses. *Curr Opin Microbiol* 14: 315–320.
- Prevelige P.E. & Fane B.A. 2012. Building the machines: scaffolding protein functions during bacteriophage morphogenesis. In: Rossmann M.G., Rao V.B. (eds.), *Viral Molecular Machines*, Springer, US, pp. 325–350.
- Rao V.B. & Feiss M. 2008. The bacteriophage DNA packaging motor. *Annu Rev Genet* 42: 647–681.
- Reiter W.D., Palm P. & Yeats S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* 17:1907–1914
- Reiter W.D., Palm P., Henschen A., Lottspeich F., Zillig W. & Grampp B. 1987. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol Gen Genet* 206: 144–153.
- 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.
- Rice G., Tang L., Stedman K., Roberto F., Spuhler J., Gillitzer E., Johnson J.E., Douglas T. & Young M. 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci U S A* 101: 7716–7720.
- Rissanen I., Grimes J.M., Pawlowski A., Mäntynen S., Harlos K., Bamford J.K., & 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.

- 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. 1984. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317: 145–153.
- Rossmann M.G. & Johnson J.E. 1989. Icosahedral RNA virus structure. *Annu Rev Biochem* 58: 533–569.
- Rothschild L.J. & Mancinelli R.L. 2001. Life in extreme environments. *Nature* 409: 1092–1101.
- Sabet S. 2012. Halophilic viruses. In: Vreeland R.H. (ed.), *Advances in Understanding the Biology of Halophilic Microorganisms*, Springer, Netherlands, pp. 81–116.
- Schnabel H., Zillig W., Pfäffle M., Schnabel R., Michel H. & Delius H. 1982. *Halobacterium halobium* phage ϕ H. *The EMBO J* 1: 87.
- Schoenfeld T., Patterson M., Richardson P.M., Wommack K.E., Young M. & Mead D. 2008. Assembly of viral metagenomes from yellowstone hot springs. *Appl Environ Microbiol* 74: 4164–4174.
- Sencilo A., Jacobs-Sera D., Russell D.A., Ko C.C., Bowman C.A., Atanasova N.S., Österlund E., Oksanen H.M., Bamford D.H., Hatfull G.F., Roine E. & Hendrix, R.W. 2013. Snapshot of haloarchaeal tailed virus genomes. *RNA Biol* 10: 803–816.
- Sharp C.E., Brady A.L., Sharp G.H., Grasby S.E., Stott M.B. & Dunfield P.F. 2014. Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J* 8: 1166–1174.
- Shen P.S., Domek M.J., Sanz-García E., Makaju A., Taylor R.M., Hoggan R., Culumber M.D., Oberg C.J., Breakwell D.P., Prince J.T. & Belnap D.M. 2012. Sequence and structural characterization of great salt lake bacteriophage CW02, a member of the T7-like supergroup. *J Virol* 86: 7907–7917.
- Siew N., Azaria Y. & Fischer D. 2004. The ORFanage: an ORFan database. *Nucleic Acids Res* 32: D281–D283
- Sime-Ngando T., Lucas S., Robin A., Tucker K.P., Colombet J., Bettarel Y., Desmond E., Gribaldo S., Forterre P., Breitbart M. & Prangishvili D. 2011. Diversity of virus-host systems in hypersaline Lake Retba, Senegal. *Environ Microbiol* 13: 1956–1972.
- Skirnisdottir S., Hreggvidsson G.O., Hjörleifsdottir S., Marteinson V.T., Petursdottir S.K., Holst O. & Kristjansso 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 D.J., Timonen H.J., Jaffe D.A., Griffin D.W., Birmele M.N., Perry K.D., Ward P.D. & Roberts M.S. 2013. Intercontinental dispersal of bacteria and archaea by transpacific winds. *Appl Environ Microbiol* 79: 1134–1139.
- Snyder J.C., Wiedenheft B., Lavin M., Roberto F.F., Spuhler J., Ortmann A.C., Douglas T. & Young M. 2007. Virus movement maintains local virus population diversity. *Proc Natl Acad Sci U S A* 104: 19102–19107.
- Song Q. & Zhang X. 2008. Characterization of a novel non-specific nuclease from thermophilic bacteriophage GBSV1. *BMC Biotechnol* 8: 43.

- Stehle T., Gamblin S.J., Yan Y. & Harrison S.C. 1996. The structure of simian virus 40 refined at 3.1 Å resolution. *Structure* 4: 165–182.
- Steinrauf L.K., Shiuan D., Yang W.J. & Chiang M.Y. 1999. Lysozyme association with nucleic acids. *Biochem Biophys Res Commun* 266: 366–370.
- Stirk H.J., Woolfson D.N., Hutchinson E.G. & Thornton J.M. 1992. Depicting topology and handedness in jellyroll structures. *FEBS Lett* 308: 1–3.
- Strömsten N.J., Bamford D.H., Bamford J.K. 2005. In vitro DNA packaging of PRD1: A common mechanism for internal-membrane viruses. *J Mol Biol* 348: 617–629.
- Suttle C.A. 2007. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* 5: 801–812.
- Takai K., Nakamura K., Toki T., Tsunogai U., Miyazaki M., Miyazaki J., Hirayama H., Nakagawa S., Nunoura T. & Horikoshi K. 2008. Cell proliferation at 122 C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc Natl Acad Sci U S A* 105: 10949–10954.
- Tate J., Liljas L., Scotti P., Christian P., Lin T. & Johnson J.E. 1999. The crystal structure of cricket paralysis virus: the first view of a new virus family. *Nat Struct Biol* 6: 765–774.
- Tavares P., Zinn-Justin S. & Orlova E.V. 2012. Genome gating in tailed bacteriophage capsids. In: Rossmann M.G., Rao V.B. (eds.), *Viral Molecular Machines*, Springer, US, pp. 558–600.
- Thomas J.A. & Black L.W. 2013. Mutational Analysis of the *Pseudomonas aeruginosa* myovirus ΦKZ morphogenetic protease gp175. *J Virol* 87: 8713–8725.
- Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
- Thurber R.V., Haynes M., Breitbart M., Wegley L. & Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4: 470–483.
- Torsvik T. & Dundas I.D. 1974. Bacteriophage of *Halobacterium salinarium*. *Nature* 248: 680–681.
- Tuma R., Parker M.H., Weigele P., Sampson L., Sun Y., Krishna N. R., Casjens S., Thomas G.J. Jr. & Prevelige P.E. Jr. 1998. A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. *J Mol Biol* 281: 81–94.
- Uldahl K. & Peng X. 2013. Biology, Biodiversity and Application of Thermophilic Viruses. In: Satyanarayana T., Littlechild J., Kawarabayasi Y. (eds.), *Thermophilic Microbes in Environmental and Industrial Biotechnology*, Springer, Netherlands, pp. 271–304.
- Veesler D., Ng T.S., Sendamarai A.K., Eilers B.J., Lawrence C.M., Lok S.M., Young M.G., 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.

- Wiedenheft B., Stedman K., Roberto F., Willits D., Gleske A.K., Zoeller L., Snyder J., Douglas T., Young M. 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. *J Virol* 78:1954–1961.
- Wikoff W.R., Liljas L., Duda R.L., Tsuruta H. Hendrix, R.W. & Johnson, J.E. 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 289: 2129–2133.
- Williamson M.P. 1994. The structure and function of proline-rich regions in proteins. *Biochem J* 297: 249–260.
- 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.
- Yan X., Yu Z., Zhang P., Battisti A.J., Holdaway H.A., Chipman P.R., Bajaj C., Bergoin M., Rossmann M.G. & Baker, T.S. 2009. The capsid proteins of a large, icosahedral dsDNA virus. *J Mol Biol* 385: 1287–1299.
- 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.
- Yin Y. & Fischer D. 2006. On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. *BMC Evol Biol* 6: 63.
- Youle M., Haynes M., Rohwer F. 2012. Scratching the surface of biology's dark matter. In: Witzany G. (ed.), *Viruses: essential agents of life*, Springer, Berlin, pp. 61–81
- 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-Å resolution. *Proc Natl Acad Sci U S A* 109: 18431–18436.
- Zhang Z., Liu Y., Wang S., Yang D., Cheng Y., Hu J., Chen J., Mei Y., Shen P., Bamford D.H. & Chen X. 2012. Temperate membrane-containing halophilic archaeal virus SNJ1 has a circular dsDNA genome identical to that of plasmid pHH205. *Virology* 434: 233–241.
- Zubieta C., Schoehn G., Chroboczek J. & Cusack S. 2005. The structure of the human adenovirus 2 penton. *Mol Cell* 17: 121–135.

ORIGINAL PAPERS

I

**A UNIQUE GROUP OF VIRUS-RELATED, GENOME-
INTEGRATING ELEMENTS FOUND SOLELY IN THE
BACTERIAL FAMILY *THERMACEAE* AND THE ARCHAEL
FAMILY *HALOBACTERIACEAE***

by

Matti Jalasvuori, Alice Pawlowski & Jaana K.H. Bamford
2010

Journal of Bacteriology vol 192, 3231-3234

Reprinted with kind permission of the American Society for Microbiology

©

A Unique Group of Virus-Related, Genome-Integrating Elements Found Solely in the Bacterial Family *Thermaceae* and the Archaeal Family *Halobacteriaceae*^{∇†}

Matti Jalasvuori, Alice Pawlowski, and Jaana K. H. Bamford*

Department of Biological and Environmental Science and Nanoscience Center, University of Jyväskylä,
P.O. Box 35, 40014 University of Jyväskylä, Finland

Received 4 February 2010/Accepted 9 April 2010

Viruses SH1 and P23-77, infecting archaeal *Haloarcula* species and bacterial *Thermus* species, respectively, were recently designated to form a novel viral lineage. In this study, the lineage is expanded to archaeal *Halomicrobium* and bacterial *Meiothermus* species by analysis of five genome-integrated elements that share the core genes with these viruses.

Viruses appear to form lineages that span different domains of life (1–3, 5, 13–15). Recently, a novel lineage of viruses and virus-like genetic elements of halophilic archaea and thermophilic bacteria was identified (13). The members of the lineage include the *Thermus thermophilus* virus P23-77 (11), the *Thermus aquaticus* virus IN93 (17), the *Haloarcula hispanica* virus SH1 (4), the *Haloarcula salinarium* plasmid pHH205 (29), and a genome-integrated element of *Halobacterium marismortui*. Structural analysis of P23-77 and SH1 revealed that both viruses form icosahedral capsids with triangulation number 28, decorate capsomers with similar types of tower-like structures, and contain an inner lipid membrane beneath the protein capsid (10, 11). The genomes of the viruses and the plasmid are formed of double-stranded DNA molecules that range from ~16 kbp to ~31 kbp in length. All of the above-mentioned genetic elements share three common genes, two coding for a small and a large major coat protein (sMCP and lMCP, respectively) and one coding for a putative genome-packaging ATPase, all arranged in similar orders in the genome (13). The cryoelectron microscopy structures of P23-77 and SH1 suggest that the base of the capsomer could be formed of a hexameric single beta-barrel protein (10, 11), in contrast to the pseudo-hexameric double beta-barrel capsomer structure of the PRD1 adenovirus lineage (15). However, the packaging ATPases share sequence similarity with the ATPases of this previously demonstrated virus lineage (13, 15). These notions lead to the suggestion that P23-77-like viruses might build an early divergent branch of the widespread lineage of beta-barrel capsid-containing viruses. Recently, the genomes and functions of some genes of the temperate virus IN93 were also studied (16, 17). IN93 was shown to contain four transcriptional units. Three of them are transcribed in the same direction, and they were active during the lytic cycle. One is transcribed in the

opposite direction and is active during the lysogenic cycle. Moreover, a novel thermostable lysozyme from IN93 was discovered (17).

In this study, we analyzed five new virus-related, genome-integrated elements belonging to the lineage of P23-77-like viruses. All of these elements contain genes for the major capsid proteins and the putative packaging ATPase. Two of the elements reside in the genomes of bacterial *Meiothermus* species (20), two in archaeal *Halomicrobium* species (21), and one in *Haloarcula* species (9), thus widening the distribution of members having a putative single beta-barrel capsid protein in the families of *Thermaceae* on the bacterial tree and *Halobacteriaceae* on the archaeal tree. There are currently no particle-forming viruses characterized for *Halomicrobium* or *Meiothermus*; therefore, the genetic elements described in this study provide the first evidence of genomic evolution by unique virus types in these bacterial and archaeal species.

Chromosome-integrated, virus-like sequences in the genomes of *Meiothermus* and *Halomicrobium* species. The genomes of *Meiothermus ruber* DSM 1279 (GenBank accession no. ABUF00000000) and *Meiothermus silvanus* DSM 9946 (ABUG00000000) contain genes with clear similarity to major capsid proteins of phages P23-77 and IN93. Closer analysis revealed that both *Meiothermus* genomes contain a P23-77-related provirus. They were designated MeioRubP1 and MeioSilP1, for the genome-integrated element in *M. ruber* and that in *M. silvanus*, respectively. However, the putative packaging ATPase gene, which is one of the core genes of the P23-77-like viruses, was missing in the genome-integrated provirus of *M. silvanus*, due to an incomplete genomic sequence. We sequenced the missing gap (using methods similar to those described in reference 13) in order to obtain the full provirus sequence. The gap was ~1,000 bp in length, and it contained a gene for a putative packaging ATPase. The boundaries of the integrated elements were determined by studying the genes surrounding the obvious virus-related sequence. Directly downstream of both the MeioRubP1 and the MeioSilP1 sequence are arginine tRNA genes. tRNA genes are known to be common sites for bacteriophage integration (6). Obvious host genes (OHCU [2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazole] decarboxylase- and predicted HDIG domain-containing

* Corresponding author. Mailing address: Department of Biological and Environmental Science and Nanoscience Center, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland. Phone: 358 14 260 2272. Fax: 358 14 260 2221. E-mail: jaana.bamford@jyu.fi.

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

∇ Published ahead of print on 16 April 2010.

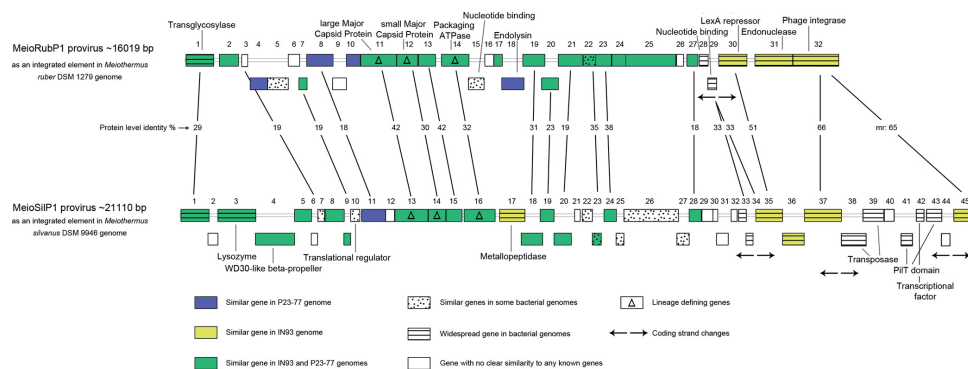


FIG. 1. Comparison of the MeiorubP1 and MeiosilP1 elements. Functions of the genes, where they are mentioned, are based on BLAST results. "nr" indicates the protein identity for the matching region in the pairwise protein alignment.

proteins) follow the tRNA genes. Upstream of the phage-related integration cassette in *M. ruber* is a gene for protoporphyrinogen oxidase, a common bacterial gene. In *M. silvanus*, the last virus-related gene (related to IN93 gene 29) encodes a protein with homology to a site-specific recombinase, which is followed by three widespread transposase genes and other host genes.

A genome-integrated, virus-like element from the genome of *Haloarcula marismortui* was discovered previously (ATCC 43049) and was designated IHP for integrated *Haloarcula* provirus (13). IHP has a set of genes in common with the *Haloarcula hispanica* virus SH1 and the *Halobacterium salinarum* plasmid pHH205. Two new elements in the genomic sequence of *Halomicrobium mukohataei* DSM 12286 (GenBank accession no. CP001688) were discovered. These new elements were designated HaloMukP1 and HaloMukP2. All three genome-integrated elements, IHP, HaloMukP1, and HaloMukP2, have at their opposite ends genes encoding a putative zinc finger protein and a phage integrase. Beyond these genes, the homologues between the genetic elements do not exist. The locations of the open reading frames (ORFs) of all of the genome-integrated elements are listed in Table S1 in the supplemental material.

Analysis of the genome-integrated elements. A comparison of the genetic elements of MeiorubP1 and MeiosilP1 is presented in Fig. 1. The elements are 16 kbp and 21 kbp for MeiorubP1 and MeiosilP1, respectively. The longer sequence of MeiosilP1 is partly due to the difference in transposases (ORFs 38 and 39) and a region that appears to have been acquired from a bacterial genome (genes 25 to 27). Indeed, it is possible that MeiosilP1 is a defective virus, due to the disturbed sequence. Matsushita and Yanase have experimentally demonstrated a transfer of transposable element lStaqTZ2 from the *Thermus thermophilus* TZ2 genome into the genome of IN93 (18), suggesting that such events occur spontaneously. Moreover, there is a short region after ORF 39 in the MeiosilP1 element which shows similarity to MeiorubP1 ORF 32 but which is not part of any ORFs of MeiosilP1. Therefore, the predicted sequences of the phage integrase

gene (ORF 37, the region after ORF 39, and ORF 42) of MeiosilP1 appear to have been divided into pieces due to integration of transposases and other non-provirus-originated genes (ORFs 40 to 44). Possibly, transposases that recognize phage integrase sequences can be favorable for the host by rendering genome-integrating viruses defective. Interestingly, the genes encoding sMCPs are only 30% identical. They were previously determined to be the most conserved genes within the lineage (13). Alignment of the genes does not indicate that a frameshift event has occurred, thus suggesting that the sMCP protein has evolved structurally to a somewhat different form or altered the interactions with other structural proteins. In line with this is, for example, the notion that some of the genes that were determined to encode minor structural components of the P23-77 virion (gene 11 and gene 29) have no homologues in the MeiorubP1 element and that the homologues with structural proteins VP19, VP20, and VP22 were only ~20% identical at the protein level. Moreover, and differently from MeiorubP1, a homolog with gene 29 of the P23-77 genome was present in the MeiosilP1 element. These aspects indicate a structural evolution in the hypothetical virions of MeiorubP1 and MeiosilP1. The cell-wall-digesting enzymes are different in MeiorubP1 and MeiosilP1. This follows the pattern in the related viruses P23-77 and IN93 (13) and, for example, some tectiviruses (26).

A comparison of the genetic elements of IHP, HaloMukP1, HaloMukP2, and pHH205 is presented in Fig. 2. In these elements, the genes encoding sMCP, IMCP, and ATPases were generally the most conserved. Indeed, the major capsid proteins of HaloMukP1 and HaloMukP2 are almost 100% identical, but many of the other common genes were only around 40% identical (interestingly, also including the putative packaging ATPase). Most of the genes in these elements were shared by at least one other member of the lineage, but many genes showed no similarities to any known genes.

The elements of the lineage have adapted to use various life strategies. SH1, IN93, and P23-77 are true viruses, pHH205 has been reported to be a plasmid, and the five other elements are integrated into the host genome. Interestingly, all of the

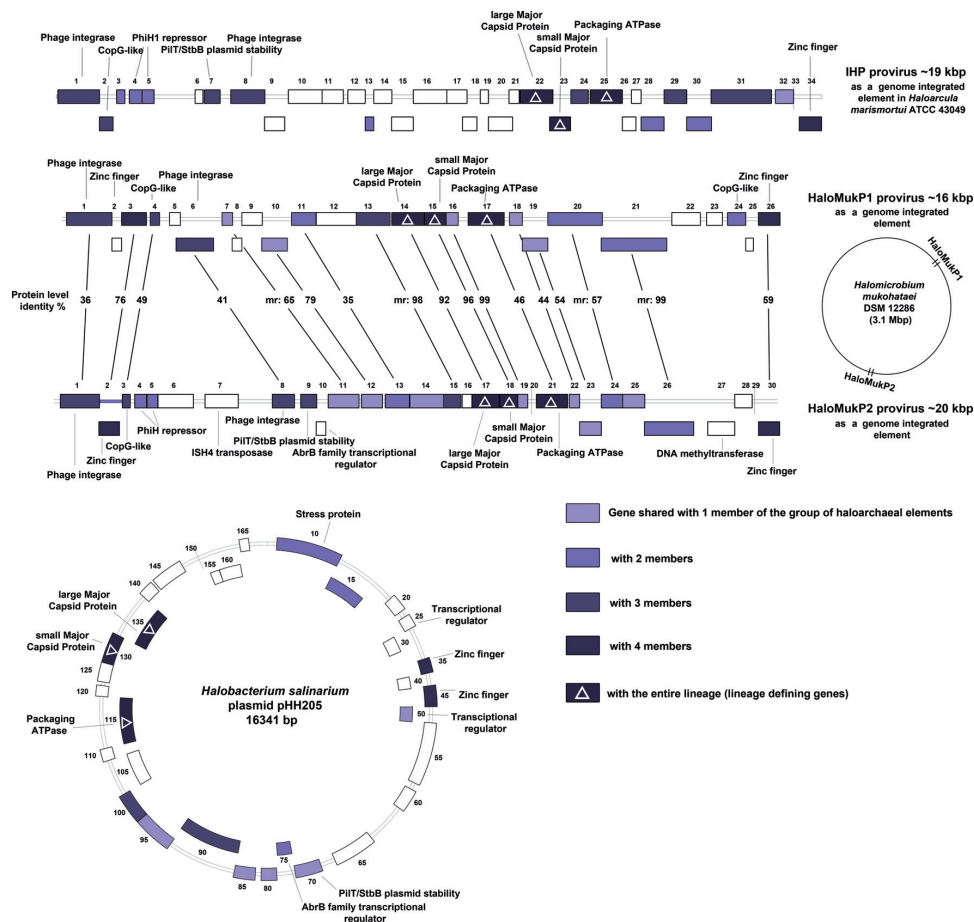


FIG. 2. Comparison of the genetic elements found in archaeal organisms. Functions of the genes, where they are mentioned, are based on BLAST results. "mr" indicates the protein identity for the matching region in the pairwise protein alignment.

genome-integrated elements have many genes with homologues to regulatory and/or DNA binding functions, suggesting that their temperate life strategies are more dependent on these genes. As opposed to the integrated elements, the lytic viruses (P23-77 and SH1), with their straightforward lytic life strategy, have no homologues to such genes.

Phylogenetic order and geographical distribution. The phylogenetic relationships and the geographical distributions of the members of the lineage are presented, respectively, in Fig. S1 and S2 in the supplemental material. The evolutionary histories were inferred as described in reference 13, using previously described methods (8, 19, 23, 24, 27, 28, 31). The

separation orders of the elements appear to be rather similar regardless of which of the lineage-defining genes are used to build the tree. However, minor differences exist, suggesting either that the evolutionary rates of these genes are not uniformly constant but may occasionally take a faster pace or that individual genes may be exchanged with related elements. The elements in *Archaea* and *Bacteria* form their own branches, suggesting that they have evolved separately. The putative single beta-barrel lineage resides in two very distantly related hosts, one group being thermophilic bacteria and the other being halophilic archaea. All of the hosts are able to thrive in moderately high temperatures, as even most of the less ther-

mophilic species of *Meiothermus* and *Halomicrobium* have optimal growth temperatures between 50 and 60°C (20, 22). It is possible that the specific pattern in which one of the hosts is a true thermophile is not a coincidence but a direct result of better chances of survival of ancient virus types in thermal environments (12). The elements are distributed equally on Earth, suggesting that they are common in any of the natural habitats of their hosts. Furthermore, they are common in the currently sequenced genomes of *Meiothermus* species and members of the order *Halobacteriales* but have not been discovered in any other cellular groups.

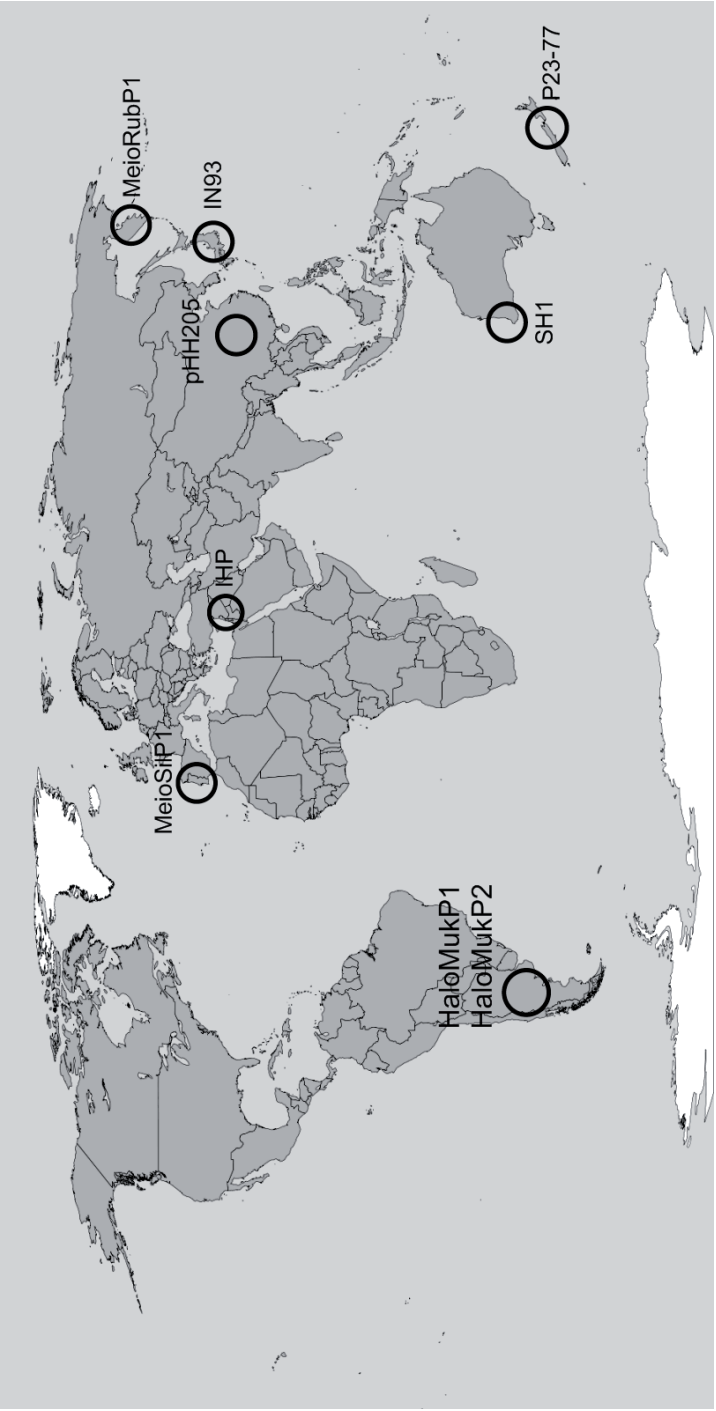
Bacterial and archaeal genomes are noted to contain many ORFans (i.e., ORFs with no matches in current databases) (25). Their share of the genomic content has remained steady despite the growing number of sequenced genomes (30). Integrative elements, such as viruses, plasmids, and transposable elements, are suggested to be responsible for a large number of the currently annotated ORFans (7). We have demonstrated here that a unique group of genome-integrating, virus-related elements can be relatively common in few groups of distantly related cellular groups and that these elements can have a number of putative genes with no matches in the previously sequenced genomes. Therefore, it is possible that a portion of the ORFans in bacterial and archaeal genomes may be related to some unique and ancient viruses.

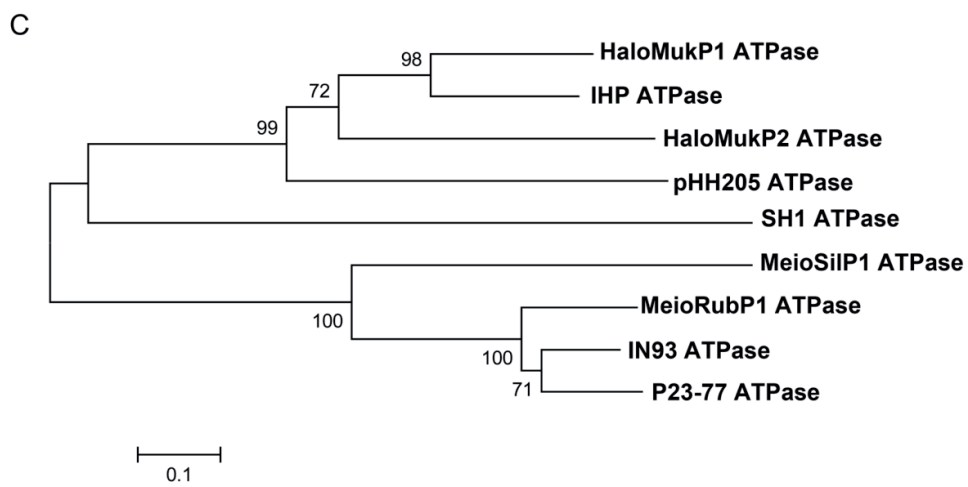
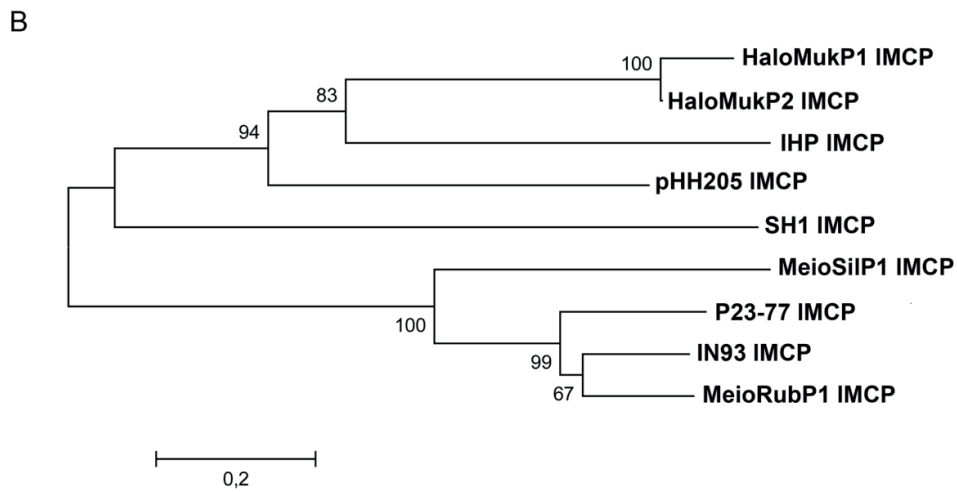
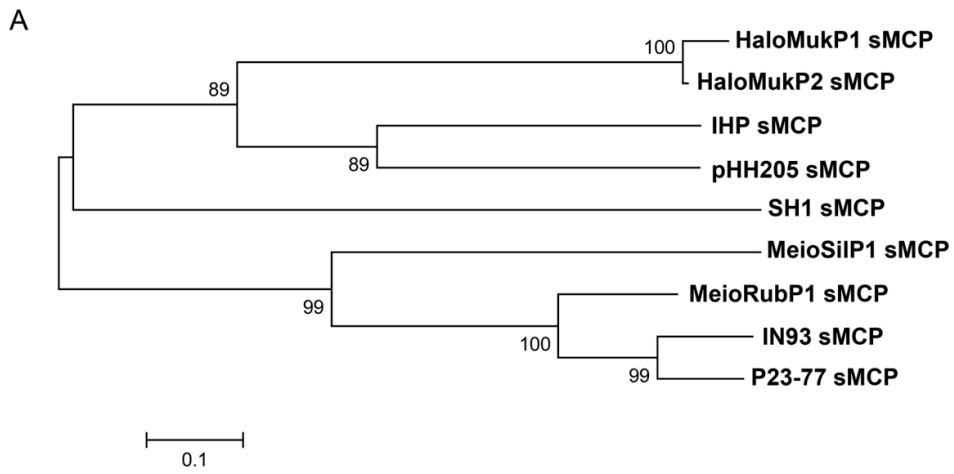
Nucleotide sequence accession numbers. The *M. silvanus* provirus sequence has been deposited in GenBank under accession number HM140848.

This work was supported by the Finnish Centre of Excellence Program of the Academy of Finland (2006-2011), grant 1129648 (J.K.H.B.).

REFERENCES

- Abrescia, N. G., J. J. Cockburn, J. M. Grimes, G. C. Sutton, J. M. Diprose, S. J. Butcher, S. D. Fuller, C. San Martin, R. M. Burnett, D. I. Stuart, D. H. Bamford, and J. K. Bamford. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**:68–74.
- Akita, F., K. T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara, and A. Nakagawa. 2007. The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *J. Mol. Biol.* **368**:1469–1483.
- Bamford, D. H. 2003. Do viruses form lineages across different domains of life? *Res. Microbiol.* **154**:231–236.
- Bamford, D. H., J. J. Ravanetti, G. Rönholm, S. Laurinavicius, P. Kukkaro, M. Dyal-Smith, P. Somerharju, N. Kalkinen, and J. K. Bamford. 2005. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J. Virol.* **79**:9097–9107.
- Benson, S. D., J. K. Bamford, D. H. Bamford, and R. M. Burnett. 2004. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* **16**:673–685.
- Canchaya, C., G. Fournous, and H. Brüssow. 2004. The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* **53**:9–18.
- Cortez, D., P. Forterre, and S. Gribaldo. 2009. A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol.* **10**:R65.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Ihara, K., S. Watanabe, and T. Tamura. 1997. *Haloarcula argentinensis* sp. nov. and *Haloarcula mukohataei* sp. nov., two new extremely halophilic archaea collected in Argentina. *Int. J. Syst. Bacteriol.* **47**:73–77.
- Jaalinoja, H. T., E. Roine, P. Laurinmaki, H. M. Kivela, D. H. Bamford, and S. J. Butcher. 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., L. J. Happonen, P. Laurinmaki, S. J. Butcher, and D. H. Bamford. 2008. Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic *Thermus thermophilus*. *Virology* **379**:10–19.
- Jalasvuori, M., and J. K. Bamford. 2009. Did the ancient crenarchaeal viruses from the dawn of life survive exceptionally well the cons of meteorite bombardment? *Astrobiology* **9**:131–137.
- Jalasvuori, M., S. T. Jaatinen, S. Laurinavicius, E. Ahola-Iivarinen, N. Kalkinen, D. H. Bamford, and J. K. Bamford. 2009. The closest relatives of icosahedral viruses of thermophilic bacteria are among viruses and plasmids of the halophilic archaea. *J. Virol.* **83**:9388–9397.
- Khayat, R., L. Tang, E. T. Larson, C. M. Lawrence, M. Young, and J. E. Johnson. 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.
- Krupovic, M., and D. H. Bamford. 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat. Rev. Microbiol.* **6**:941–948.
- Matsushita, I., and H. Yanase. 2008. A novel thermophilic lysozyme from bacteriophage phiN93. *Biochem. Biophys. Res. Commun.* **377**:89–92.
- Matsushita, I., and H. Yanase. 2009. The genomic structure of thermus bacteriophage phiN93. *J. Biochem.* **146**:775–785.
- Matsushita, I., and H. Yanase. 2009. A novel insertion sequence transposed to thermophilic bacteriophage phiN93. *J. Biochem.* **145**:797–803.
- Nei, M., and S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY.
- Nobre, M. F., H. G. Triiper, and M. S. da Costa. 1996. Transfer of *Thermus ruber* (Logoinova et al. 1984), *Thermus silvanus* (Tenreiro et al. 1995), and *Thermus chliarophilus* (Tenreiro et al. 1995) to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., respectively, and emendation of the genus *Thermus*. *Int. J. Syst. Bacteriol.* **46**:604–606.
- Oren, A., R. Eleveli, S. Watanabe, K. Ihara, and A. Corcelli. 2002. *Halomicrobium mukohataei* gen. nov., comb. nov., and emended description of *Halomicrobium mukohataei*. *Int. J. Syst. Evol. Microbiol.* **52**:1831–1835.
- Robinson, J. L., B. Pyzyra, R. G. Atrasz, C. A. Henderson, K. L. Morrill, A. M. Burd, E. Desoucy, R. E. Foglieman III, J. B. Naylor, S. M. Steele, D. R. Elliott, K. J. Leyva, and R. F. Shand. 2005. Growth kinetics of extremely halophilic archaea (family *Halobacteriaceae*) as revealed by Arrhenius plots. *J. Bacteriol.* **187**:923–929.
- Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum evolution trees. *Mol. Biol. Evol.* **9**:945–967.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Siew, N., Y. Azaria, and D. Fischer. 2004. The ORFanage: an ORFan database. *Nucleic Acids Res.* **32**:D281–D283.
- Sozhamannan, S., M. McKinsty, S. M. Lentz, M. Jalasvuori, F. McAfee, A. Smith, J. Dabbs, H. W. Ackermann, J. K. Bamford, A. Maticzun, and T. D. Read. 2008. Molecular characterization of a variant of *Bacillus anthracis*-specific phage AP50 with improved bacteriolytic activity. *Appl. Environ. Microbiol.* **74**:6792–6796.
- Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
- Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
- Ye, X., J. Ou, L. Ni, W. Shi, and P. Shen. 2003. Characterization of a novel plasmid from extremely halophilic Archaea: nucleotide sequence and function analysis. *FEMS Microbiol. Lett.* **221**:53–57.
- Yin, Y., and D. Fischer. 2006. On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. *BMC Evol. Biol.* **6**:63.
- Zackerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins, p. 97–166. *In* V. Bryson and H. J. Vogel (ed.), *Evolving genes and proteins*. Academic Press, New York, NY.





MeioRubP1		MeioSIIP1		HaloMukP1		HaloMukP2		IHP	
Host GenBank id	ABUF000000000	Host GenBank id	ABUG000000000	Host GenBank id	CP001688	Host GenBank id	CP001688	Host GenBank id	A Y596297
Integrated element position in GenBank file	1703253 - 1719272	Integrated element position in GenBank file	Contig 7 (NZ_ABUG01000004); 219435 - 227946; Contig 8 (NZ_ABUG01000017); 1 - 12918	Integrated element position in GenBank file	447446 - 463160	Integrated element position in GenBank file	1628217 - 1647275	Integrated element position in GenBank file	514712 - 533695
Element GC% Host GC%	60 63	Element GC% Host GC%	59 62	Element GC% Host GC%	64 65	Element GC% Host GC%	62 65	Element GC% Host GC%	59 62
Prophage ORF	ORF Location*	ORF	ORF Location*	ORF	ORF Location*	ORF	ORF Location*	ORF	ORF Location*
1	1-683	1	Contig 7: 1-747	1	1-1008	1	1-1035	1	1-1032
2	818-1288	2	725-985	2	1008-1223	2	1031-1564	2	1028-1386
3	1445-1576	3	985-1995	3	1223-1774	3	1649-1852	3	1456-1659
4	1576-1895	4	1988-3034	4	1860-2066	4	1970-2290	4	1777-2091
5	2008-2511	5	3037-3489	5	2286-2525	5	2297-2594	5	2098-2385
6	2508-2786	6	3480-3650	6	2431-3264	6	2592-3533	6	3411-3599
7	2776-2967	7	3656-3841	7	3447-3677	7	3830-4720	7	3632-4030
8	2887-3608	8	3855-4358	8	3677-3886	8	5619-6203	8	4301-5143
9	3895-3836	9	4352-4525	9	3889-4338	9	6366-6788	9	5134-5664
10	3838-4276	10	4525-4776	10	4329-4898	10	6788-7030	10	5724-6560
11	4291-5166	11	4821-5463	11	4932-5528	11	7100-7909	11	6565-7092
12	5179-5688	12	5468-5707	12	5538-6419	12	7976-8536	12	7211-7639
13	5704-6120	13	5710-6591	13	6425-7171	13	8613-9257	13	7639-7851
14	6269-6943	14	6603-7081	14	7207-7929	14	9265-10152	14	7851-8300
15	6933-7316	15	7074-7502	15	7937-8416	15	10157-10609	15	8291-8830
16	7339-7548	16	8501	16	8419-8679	16	10644-10904	16	8830-9689
17	7593-7757	17	8505-9179	17	8906-9703	17	10912-11637	17	9674-10174
18	7794-8290	18	9088-9697	18	9825-10100	18	11645-12124	18	10063-10416
19	8271-8804	19	9593-9955	19	10100-10672	19	12127-12387	19	10501-10698
20	8725-9135	20	9955-10422	20	10872-11677	20	12463-12620	20	10898-11300
21	9142-9732	21	10507-10862	21	11582-13309	21	12620-13447	21	11210-11470
22	9737-10050	22	10713-10979	22	13414-14046	22	13468-13759	22	11488-12300
23	10659-10432	23	10979-11216	23	14189-14596	23	13739-14534	23	12219-12746
24	10463-10788	24	11284-11623	24	14581-15055	24	14534-15003	24	12732-13186
25	10798-11024	25	11624-11920	25	15052-15222	25	15006-15094	25	13235-14092
26	11268-12024	26	11927-12653	26	15536-15612	26	15094-15779	26	14254-14394
27	12268-12569	27	13237-13669	27	17159-17574	27	17159-17574	27	14256-14494
28	12882-12893	28	13669-13869	28	17970-18304	28	18304-18508	28	14408-15067
29	12809-13018	29	13910-14208	29	18306-18508	29	18306-18508	29	15075-15632
30	13074-13766	30	14208-14612	30	18508-19056	30	18508-19056	30	15629-16240
31	13966-14880	31	14681-14846	31	14681-14846	31	18508-19056	31	16240-17293
32	14895-16019	32	14886-15089	32	15085-16270	32	18508-19056	32	17635-18287
		33	15085-16270	33	16055-16633	33		33	18287-18430
		34	16055-16633	34	16650-17627	34		34	18430-18981
		35	16650-17627	35	17627-18283	35			
		36	17627-18283	36	18196-18750	36			
		37	18196-18750	37	18768-19031	37			
		38	18768-19031	38	19213-19506	38			
		39	19213-19506	39	19619-19825	39			
		40	19619-19825	40	19897-20289	40			
		41	19897-20289	41	20289-20156	41			
		42	20289-20156	42	20619-21104	42			
		43	20619-21104	43		43			
		44		44		44			
		45		45		45			

II

CRYSTALLIZATION AND PRELIMINARY CRYSTALLOGRAPHIC ANALYSIS OF THE MAJOR CAPSID PROTEINS VP16 AND VP17 OF BACTERIOPHAGE P23-77

by

Ilona Rissanen, Alice Pawlowski, Karl Harlos, Jonathan M. Grimes, Jaana K.H.
Bamford & David I. Stuart
2012

Acta Crystallographica Section F Structural Biology and Crystallization
communications vol 68, 580-583

Reprinted with kind permission of the International Union of Crystallography

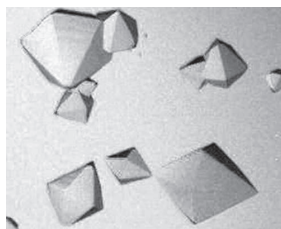
©

Ilona Rissanen,^a Alice
Pawlowski,^a Karl Harlos,^b
Jonathan M. Grimes,^{b,c}
David I. Stuart^{b,c} and
Jaana K. H. Bamford^{a*}

^aDepartment of Biological and Environmental Science and Nanoscience Center, University of Jyväskylä, PO Box 35, 40014 University of Jyväskylä, Finland, ^bDivision of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, England, and ^cScience Division, Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot OX11 0DE, England

Correspondence e-mail:
jaana.k.h.bamford@jyu.fi

Received 26 January 2012
Accepted 8 March 2012



Crystallization and preliminary crystallographic analysis of the major capsid proteins VP16 and VP17 of bacteriophage P23-77

Members of the diverse double- β -barrel lineage of viruses are identified by the conserved structure of their major coat protein. New members of this lineage have been discovered based on structural analysis and we are interested in identifying relatives that utilize unusual versions of the double- β -barrel fold. One candidate for such studies is P23-77, an icosahedral dsDNA bacteriophage that infects the extremophile *Thermus thermophilus*. P23-77 has two major coat proteins, namely VP16 and VP17, of a size consistent with a single- β -barrel core fold. These previously unstudied proteins have now been successfully expressed as recombinant proteins, purified and crystallized using hanging-drop and sitting-drop vapour-diffusion methods. Crystals of coat proteins VP16 and VP17 have been obtained as well as of a putative complex. In addition, virus-derived material has been crystallized. Diffraction data have been collected to beyond 3 Å resolution for five crystal types and structure determinations are in progress.

1. Introduction

Analysis of virus structures determined over the past 30 years has led to the concept of 'viral self' elements that can be used to trace ancient evolutionary relationships (Benson *et al.*, 1999; Bamford *et al.*, 2005). 'Viral self' elements, vertically inherited elements that are required to assemble a virion, such as capsid proteins, are key to the survival of the virus; thus, while viral DNA may change, the structure and architecture of the virus and its capsid proteins are preserved. One such lineage, identified by conserved protein structures, containing adenovirus and the bacterial virus PRD1 traces its roots back to the divergence of the domains of life (Krupovic & Bamford, 2008; Bamford *et al.*, 2005). Modern members of this lineage are identified primarily by a capsid protein fold containing a double β -barrel. This structural signature has been observed in viruses such as adenovirus, PRD1, STIV, PBCV-1, PM2 and recently vaccinia virus (Abrescia *et al.*, 2004, 2008; Bahar *et al.*, 2011; Khayat *et al.*, 2005; Nandhagopal *et al.*, 2002; Rux *et al.*, 2003; Liu *et al.*, 2010; Reddy *et al.*, 2010), that infect hosts from different domains of life.

P23-77 is a bacteriophage which appears to possess structural elements that bear a novel relationship to the adeno/PRD1 lineage. It has a $T = 28$ icosahedral capsid, ~ 800 Å across, with an internal lipid membrane enclosing a circular dsDNA genome (Yu *et al.*, 2006; Jaatinen *et al.*, 2008; Jalasvuori *et al.*, 2009). The genome is approximately 17 000 nucleotides in length and contains 37 putative ORFs, ten of which code for structural proteins (Jaatinen *et al.*, 2008; Jalasvuori *et al.*, 2009). P23-77 has two major coat proteins, VP16 (20 kDa) and VP17 (32 kDa); they form, in an approximate 1:1 ratio, a viral capsid with an unusual base-and-tower architecture reminiscent of the capsomeric structure observed in some members of the adeno/PRD1 lineage but sited at positions in the icosahedral virus that are not allowed for the adeno/PRD1 trimeric capsomers. Thus, while in the adeno/PRD1 lineage the pseudo-hexameric capsomers are formed by trimers composed of subunits harbouring a double- β -barrel fold, in P23-77 some capsomers lie on icosahedral twofold axes. This organization has only recently been identified in extremophile viruses and has not been characterized at the level of protein structure (Jaatinen *et al.*, 2008; Jääliñoja *et al.*, 2008). In addition, sequence analysis of the packaging ATPase of P23-77

suggests a relationship to the adeno/PRD1 double- β -barrel lineage. P23-77 may belong to an unknown ancestral branch of the double- β -barrel lineage that utilizes single- β -barrel core structures instead of the duplicate version and has modern members among extremophile viruses, including the archaeal virus SH1 (Jalasvuori *et al.*, 2009, 2010). Here, we report the crystallization and preliminary diffraction analysis of VP16 and VP17, with the aim of ultimately solving their structures by X-ray crystallography, in order to shed light on the origins of P23-77 and its relatives.

2. Materials and methods

2.1. Plasmid construction

Genes *ORF16* and *ORF17* were PCR-amplified from the P23-77 genome using primers with restriction sites for *NdeI* and *HindIII* and corresponding to full-length VP16 (1–173) and VP17 (1–291), respectively. Purified restricted PCR fragments were ligated with *NdeI*–*HindIII*-restricted expression vector pET22b(+) (Novagen) and the resulting recombinant plasmids pIR1 (*ORF17*/pET22b) and pIR2 (*ORF16*/pET22b) were used to transform competent *Escherichia coli* HMS174 (DE3) for high-level recombinant protein expression using standard methods.

2.2. Protein expression and purification

Large-scale cultures of *E. coli* HMS174 (DE3) transformed with each plasmid were grown (12×400 ml LB medium with $150 \mu\text{g ml}^{-1}$ ampicillin, 310 K, 230 rev min^{-1}) until the absorbance at 550 nm reached 0.5, at which point recombinant protein expression was induced by the addition of IPTG to a final concentration of 1 mM. Cultures were grown for 22 h. Cells were collected by centrifugation, resuspended to one hundredth of the original volume in 20 mM Tris pH 7.4, 50 mM NaCl buffer and disrupted with a French press (Thermo Fisher Scientific). Soluble fractions containing target proteins were separated from cell debris by centrifugation (Beckman Ti-70 rotor, 33 000 rev min^{-1} , 2 h, 278 K).

Supernatants containing VP16 or VP17 were incubated at 363 K for 10 min, which caused degradation of the less heat-stable host-derived proteins. Degraded material was removed by centrifugation, after which samples were concentrated and buffer-exchanged (20 mM ethanolamine pH 9, 50 mM NaCl for VP16 and 20 mM ethanolamine pH 9.5 for VP17) using an Amicon ultrafiltration system (Millipore).

VP16 was loaded onto an anion-exchange chromatography column [5 ml Q HP HiTrap column (GE Healthcare)] equilibrated with 20 mM ethanolamine pH 9 at 295 K]. The flowthrough containing VP16 was further purified by size-exclusion chromatography [HiLoad 26/60 Superdex 200 prep-grade column (GE Healthcare)] equilibrated with 20 mM Tris pH 7.4, 150 mM NaCl at 295 K]. VP17 was purified with a similar anion-exchange chromatography protocol in which the column was equilibrated with 20 mM ethanolamine pH 9.5 and the protein was eluted specifically with 50 mM NaCl. Fractions containing VP17 were purified by size-exclusion chromatography as for VP16. After size-exclusion chromatography, fractions containing pure VP16 or VP17 were pooled, concentrated, exchanged into 20 mM Tris pH 7.4 buffer and stored at 280 K. Purified VP16 and VP17 were concentrated using 10 kDa molecular-weight microconcentrators (Amicon).

2.3. P23-77 virus purification

Virus particles were produced in *T. thermophilus* strain ATCC33923 and purified as described previously (Jaatinen *et al.*, 2008). In brief, cells were infected at a cell density of 7×10^8 cfu ml^{-1}

with a multiplicity of infection (MOI) of around 10. Viral particles were precipitated from the lysate with 12% polyethylene glycol (PEG) 6000 and 0.5 M NaCl and concentrated to one twentieth of the original lysate volume in TV buffer (20 mM Tris–HCl pH 7.5, 5 mM MgCl_2 , 150 mM NaCl). Viruses were purified by rate zonal centrifugation [linear 5–20% (*w/v*) sucrose gradient in TV buffer, 23 000 rev min^{-1} , 45 min, 298 K], followed by equilibrium centrifugation in 1.30 mg ml^{-1} CsCl_2 in TV buffer (21 000 rev min^{-1} , 16 h, 298 K). $2 \times$ purified viral particles were collected by differential centrifugation (32 000 rev min^{-1} , 4 h, 298 K) and the virus pellet was suspended in TV buffer. Purified virus samples were stored at 295 K.

3. Results

3.1. Crystallization

Crystallization conditions for purified full-length VP16 (1–173) and full-length VP17 (1–291) and the P23-77 virion were initially screened by hanging-drop vapour diffusion using a Mosquito Nanodrop Crystallization Robot (TTP LabTech) at the University of Jyväskylä, Finland. Subsequently, final crystallization experiments were performed at the Division of Structural Biology, Oxford University, England either by setting up crystallization experiments manually or by using the nanolitre high-throughput facility with sitting-drop sizes of 1 $\mu\text{l} + 1 \mu\text{l}$ and 100 nl + 100 nl (protein solution and crystallization screen), respectively (Walter *et al.*, 2005). Commercially available crystallization screening kits (Hampton Research, California, USA, Molecular Dimensions, UK and Emerald BioStructures, Washington, USA) were used for all initial experiments. 576 crystallization conditions were tested for VP16 and 288 conditions were tested for VP17, whilst the VP16–VP17 complex was screened against 984 conditions. Virion crystallization was tested against 480 conditions. All crystallizations were set up at room temperature (293–295 K). A number of different crystal forms were obtained.

VP16 type 1 crystals grew within 1–2 weeks from microlitre drops of protein ($2\text{--}3 \text{ mg ml}^{-1}$) mixed in a 1:1 (*v:v*) ratio with a solution consisting of 5% (*w/v*) PEG 1000 and 5% (*w/v*) PEG 8000 dissolved in autoclaved water (the crystallization drop was pH 7.4). Additional screening experiments carried out in 96-well plates in the high-throughput crystallization facility (Walter *et al.*, 2005) yielded VP16 type 2 crystals. These crystals grew within days from 20% (*w/v*) polyethylene glycol 6000, 0.1 M citrate pH 4.

VP17 crystals were obtained from drops consisting of 1 μl protein solution ($2\text{--}3 \text{ mg ml}^{-1}$) mixed with 1 μl 1.9 M sodium formate, 0.1 M bis-Tris buffer pH 7.0. Crystals grew to full size in two weeks.

In an attempt to obtain the structure of a VP16–VP17 complex, VP16 and VP17 were mixed at concentrations of 1.7 and 2.0 mg ml^{-1} , respectively (*i.e.* a 1:1 molar ratio); one well diffracting crystal, which was used for data collection, took three months to grow from 1.1 M diammonium tartrate pH 7.

Crystals were also obtained from experiments using the whole P23-77 virion. Crystallization conditions were screened in 96-well plates with virion concentrations ranging from 1.0 to 2.5 mg ml^{-1} . Thin needles appeared in various conditions, all of which contained 0.1 M citric acid pH 3.5 and PEG. Crystals appeared within 12 h to 8 d depending on the buffer conditions, virion concentration and virus preparation. Crystals used for diffraction analysis were grown in microlitre drops which were set up manually at a virus concentration of 2.4 mg ml^{-1} and a virus:reservoir ratio of 1:1 (*v:v*). The reservoir solution consisted of 0.1 M citric acid pH 3.5, 20 mM Tris–HCl pH 7.5, 5 mM MgCl_2 , 150 mM NaCl and 25% PEG 3350. All five crystal types

crystallization communications

used for diffraction analysis of VP16, VP17, their complex and virion-derived material are illustrated in Fig. 1.

3.2. X-ray characterization and data collection

All crystals were cooled in liquid nitrogen using glycerol mixed with reservoir solution at 25% (v/v) as a cryoprotectant and exposed to X-rays at 100 K. Initially, crystals of VP16 type 1 and VP17 were

characterized at Jyväskylä and Oxford using in-house X-ray equipment. Subsequently, definitive data sets for all crystals were collected at the Diamond Light Source synchrotron, Didcot, England as follows. VP16 type 1 data were collected on beamline I03 ($\lambda = 0.979 \text{ \AA}$) in high- and low-resolution sweeps, with the high-resolution sweep consisting of 360 images with 0.5° oscillation per image and an exposure time of 1 s per image. VP16 type 2 data were collected on beamline I04 ($\lambda = 1.000 \text{ \AA}$) as 360 images with 1° and 1 s

Table 1

Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell. Each data set was collected from one crystal, except for the virus-derived crystal data set, which was collected from two.

	VP16 type 1	VP16 type 2	VP17	Putative complex	Virus-derived crystals
Space group	<i>P</i> 6 ₂ 22	<i>C</i> 2	<i>P</i> 6 ₂ 22	<i>C</i> 2	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit-cell parameters					
<i>a</i> (Å)	61.9	76.6	107.2	76.8	41.8
<i>b</i> (Å)	61.9	68.6	107.2	69.6	78.2
<i>c</i> (Å)	251.2	31.6	233.8	81.6	405.8
α (°)	90	90	90	90	90
β (°)	90	96.4	90	105.0	90
γ (°)	120	90	120	90	90
Resolution (Å)	62.8–1.80 (1.85–1.80)	34.3–1.26 (1.30–1.26)	59.7–2.26 (2.32–2.26)	39.6–1.53 (1.57–1.53)	51.2–2.92 (2.99–2.92)
$R_{\text{merge}}^{\dagger}$	0.075 (1.045)	0.053 (0.626)	0.082 (0.917)	0.064 (1.015)	0.301 (1.492)
$\langle I/\sigma(I) \rangle$	30.7 (3.3)	23.1 (2.6)	38.6 (5.3)	17.7 (2.9)	6.4 (1.5)
Completeness (%)	100 (100)	85.8 (41.4)	100 (100)	98.4 (83.3)	99.9 (99.8)
Multiplicity	28.5 (21.0)	7.5 (6.5)	35.5 (36.6)	6.3 (4.4)	6.3 (5.6)

$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th measurement and $\langle I(hkl) \rangle$ is the weighted mean of all measurements $I_i(hkl)$.

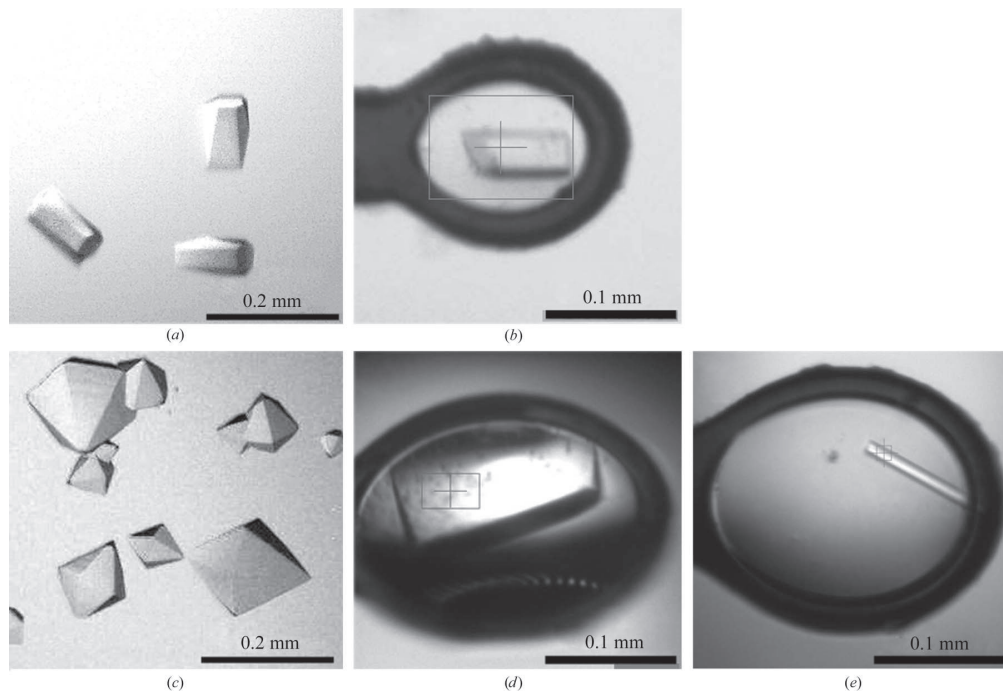


Figure 1

Crystals of the major coat proteins of P23-77: (a) VP16 type 1, (b) VP16 type 2, (c) VP17, (d) putative complex, (e) virus-derived crystals. (b), (d) and (e) show crystals frozen in loops that were used for data collection at the beamline.

per image. VP17 data were collected on beamline I04 ($\lambda = 1.000 \text{ \AA}$) as 614 images with 0.5° and 1 s per image. The data for the putative complex were collected on beamline I24 ($\lambda = 1.071 \text{ \AA}$) as 1800 images with 0.2° and 0.2 s per image. Data for the virion-derived crystal were collected on beamline I24 ($\lambda = 0.969 \text{ \AA}$) as 2×450 images with 0.2° and 0.2 s per image. All crystals diffracted to better than 3 \AA resolution. Data were automatically processed with *xia2/XDS* (Winter, 2010; Kabsch, 1993) and the processing statistics are summarized in Table 1.

4. Discussion

P23-77 major capsid proteins VP16 and VP17 have been purified, crystallized and native data sets collected. The host of bacteriophage P23-77 is the extremophile *T. thermophilus* and consequently the thermal stability of the P23-77 proteins made purification straightforward; raising the temperature of the clarified supernatant to 363 K for 10 min degraded and precipitated most cellular proteins.

Preliminary data analyses indicate that the asymmetric units of VP16 type 1 and VP16 type 2 are likely to contain one subunit of VP16 each (corresponding to solvent contents of 65 and 41%, respectively); the asymmetric unit of VP17 probably contains two subunits of VP17 (corresponding to a solvent content of 59%) and the asymmetric unit of the putative complex could accommodate one subunit each of VP16 and VP17 (corresponding to a solvent content of 40%). The unit cell of the crystal derived from virus crystallization is far too small to contain the whole virus (which is some 800 \AA across, exceeding a complete unit cell in every direction). It is very likely to contain VP16 and/or VP17 (for example, six subunits of VP16 in the crystallographic asymmetric unit would correspond to a solvent content of 55%, whereas four subunits of VP17 would correspond to 52% solvent). A search for heavy-atom derivatives is in progress. Structures of major capsid proteins VP16 and VP17, which are the building blocks of P23-77, and their complexes will provide details of both their oligomeric states and how they assemble to form part of the capsid of P23-77. This will contribute to the overall picture of the evolutionary relationships in this diverse group of dsDNA viruses.

Petri Papponen and Salla Ruskamo are thanked for valuable technical assistance, and Professors Jari Yläne and Kari Rissanen for providing facilities for crystallization screening and X-ray crys-

tallography. We thank the staff of beamlines I03, I04 and I24 at the Diamond Light Source for technical support. DIS and KH are supported by the UK MRC. This work was enabled by the UK MRC, European Commission contract No. 031220FP6 (SPINE2-COMPLEXES) and Finnish Centre of Excellence Program of the Academy of Finland (2006–2011) grant 1129648 (JKHB) and was supported by the EU P-CUBE program (grant 227764). The Wellcome Trust is acknowledged for providing administrative support (grant 075491/Z/04).

References

- Abrescia, N. G., Cockburn, J. J., Grimes, J. M., Sutton, G. C., Diprose, J. M., Butcher, S. J., Fuller, S. D., San Martín, C., Burnett, R. M., Stuart, D. I., Bamford, D. H. & Bamford, J. K. (2004). *Nature (London)*, **432**, 68–74.
- Abrescia, N. G., Grimes, J. M., Kivelä, H. M., Assenberg, R., Sutton, G. C., Butcher, S. J., Bamford, J. K., Bamford, D. H. & Stuart, D. I. (2008). *Mol. Cell*, **31**, 749–761.
- Bahar, M. W., Graham, S. C., Stuart, D. I. & Grimes, J. M. (2011). *Structure*, **19**, 1011–1020.
- Bamford, D. H., Grimes, J. M. & Stuart, D. I. (2005). *Curr. Opin. Struct. Biol.* **15**, 655–663.
- Benson, S. D., Bamford, J. K., Bamford, D. H. & Burnett, R. M. (1999). *Cell*, **98**, 825–833.
- Jääliñoja, H. T., Roine, E., Laurinmäki, P., Kivelä, H. M., Bamford, D. H. & Butcher, S. J. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 8008–8013.
- Jaatinen, S. T., Happonen, L. J., Laurinmäki, P., Butcher, S. J. & Bamford, D. H. (2008). *Virology*, **379**, 10–19.
- Jalasvuori, M., Jaatinen, S. T., Laurinavicius, S., Ahola-Iivarinen, E., Kalkkinen, N., Bamford, D. H. & Bamford, J. K. (2009). *J. Virol.* **83**, 9388–9397.
- Jalasvuori, M., Pawlowski, A. & Bamford, J. K. (2010). *J. Bacteriol.* **192**, 3231–3234.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Khayat, R., Tang, L., Larson, E. T., Lawrence, C. M., Young, M. & Johnson, J. E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 18944–18949.
- Krupovic, M. & Bamford, D. H. (2008). *Nature Rev. Microbiol.* **6**, 941–948.
- Liu, H., Jin, L., Koh, S. B., Atanasov, I., Schein, S., Wu, L. & Zhou, Z. H. (2010). *Science*, **329**, 1038–1043.
- Nandhagopal, N., Simpson, A. A., Gurnon, J. R., Yan, X., Baker, T. S., Graves, M. V., Van Etten, J. L. & Rossmann, M. G. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 14758–14763.
- Reddy, V. S., Natchiar, S. K., Stewart, P. L. & Nemerow, G. R. (2010). *Science*, **329**, 1071–1075.
- Rux, J. J., Kuser, P. R. & Burnett, R. M. (2003). *J. Virol.* **77**, 9553–9566.
- Walter, T. S. *et al.* (2005). *Acta Cryst.* **D61**, 651–657.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Yu, M. X., Slater, M. R. & Ackermann, H.-W. (2006). *Arch. Virol.* **151**, 663–679.

III

**GAMMASPHAEROLIPOVIRUS, A NEWLY PROPOSED
BACTERIOPHAGE GENUS, UNIFIES VIRUSES OF
HALOPHILIC ARCHAEA AND THERMOPHILIC BACTERIA
WITHIN THE NOVEL FAMILY SPHAEROLIPOVIRIDAE**

by

Alice Pawlowski, Ilona Rissanen, Jaana K.H. Bamford, Mart Krupovic &
Matti Jalasvuori
2014

Archives of Virology vol 159, 1541-1554

Reprinted with kind permission of Springer

©

IV

THE MINOR CAPSID PROTEIN VP11 OF THERMOPHILIC BACTERIOPHAGE P23-77 FACILITATES CAPSID ASSEMBLY USING LIPID-PROTEIN INTERACTIONS

by

Alice Pawlowski, Anni M. Moilanen, Ilona Rissanen, Juha A.E. Määttä, Vesa P.
Hytönen, Janne A. Ihalainen & Jaana K.H. Bamford
2015

submitted manuscript