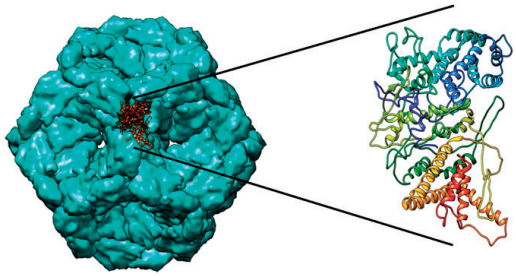
JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE
312

Sari Mäntynen

Something Old, Something New

Exploring Membrane-Containing Bacteriophages



JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 312

Sari Mäntynen

Something Old, Something New

Exploring Membrane-Containing Bacteriophages

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ambiotica-rakennuksen salissa YAA303,
tammikuun 15. päivänä 2016 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Ambiotica, hall YAA303, on January 15, 2016 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2016

Something Old, Something New

Exploring Membrane-Containing Bacteriophages

JYVÄSKYLÄ STUDIES IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE 312

Sari Mäntynen

Something Old, Something New

Exploring Membrane-Containing Bacteriophages

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2016

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

Jyväskylä Studies in Biological and Environmental Science
Editorial Board

Jari Haimi, Anssi Lensu, Timo Marjomäki, Varpu Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä

Cover picture: Photo of Ylistönrinne campus by Juho Niva;
models of φNN P1 core structure and φNN P1 protein by Janne Ravantti.

URN:ISBN:978-951-39-6461-0
ISBN 978-951-39-6461-0 (PDF)

ISBN 978-951-39-6460-3 (nid.)
ISSN 1456-9701

Jyväskylä University Printing House, Jyväskylä 2016

ABSTRACT

Mäntynen, Sari
Something old, something new - Exploring membrane-containing bacteriophages
Jyväskylä: University of Jyväskylä, 2016, 111 p.
(Jyväskylä Studies in Biological and Environmental Science
ISSN 1456-9701; 312)
ISBN 978-951-39-6460-3 (nid.)
ISBN 978-951-39-6461-0 (PDF)
Yhteenveto: Jotain uutta, jotain vanhaa – kalvorakenteen sisältävät bakteriofagit
Diss.

Bacterial viruses, also called bacteriophages or phages, form a remarkably large and diverse group of biological entities. Bacteriophage studies have traditionally centered on tailed non-enveloped species, but a multitude of other virus types is gradually coming to light. Phages with a structural membrane component vary in terms of morphologies, genome types and replication mechanisms. Possibly the most comprehensively characterized membrane-containing phage is enterobacteria infecting PRD1 of the *Tectiviridae* family. However, there are still unanswered questions concerning its assembly process. This thesis adds new pieces to the puzzle by demonstrating that the subcellular distribution of fluorescently labelled PRD1 proteins is asymmetric, possibly resulting from spatially organized phage protein oligomerization and/or virion assembly. Moreover, PRD1 non-structural proteins P17 and P33 are shown to complement a defect in host co-chaperonin, suggesting a degree of functional redundancy between the viral and bacterial proteins. This thesis also introduces two novel bacteriophages: φNN and FLiP infecting *Pseudomonas* and *Flavobacterium* species, respectively. The virion of φNN contains a tri-segmented dsRNA genome enclosed by an icosahedral protein core and an outermost membrane envelope. Structural and genetic similarities strongly suggest that φNN is a new member of the *Cystoviridae* family. Genetic and structural comparisons among the putative cystoviruses display the conservation of viral "self" components, essential for the formation of functional virions, while the sequences required for host recognition differ. On the other hand, the virion of FLiP consists of an icosahedral protein shell with an internal membrane, uniquely combined with a circular ssDNA genome. FLiP lacks any significant sequence homology to known bacteriophages, but its overall virion structure resembles strikingly that of dsDNA phage PM2, tentatively suggesting a relationship between these phages with different genome types.

Keywords: Bacteriophage PRD1; chaperonin complex; *Cystoviridae*; fluorescent fusion protein; membrane-containing virus; ssDNA phage; virus assembly.

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

Author's address Sari Mäntynen
Department of Biological and Environmental Science and Nanoscience Center
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
sari.s.mantynen@jyu.fi

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

Docent Hanna Oksanen
Institute of Biotechnology and Department of Biosciences
P.O. Box 56
FI-00014 University of Helsinki
Finland

Reviewers Docent Katri Eskelin
Department of Biosciences
P.O. Box 56
FI-00014 University of Helsinki
Finland

Reader Eugene Makeyev
MRC Centre for Developmental Neurobiology
King´s College London
London SE1 1UL
United Kingdom

Opponent Professor Harri Savilahti
Division of Genetics and Physiology
Department of Biology
FI-20014 University of Turku
Finland

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I Karttunen J., Mäntynen S., Ihalainen T.O., Lehtivuori H., Tkachenko N.V., Vihinen-Ranta M., Ihalainen J.A., Bamford J.K.H. & Oksanen H.M. 2014. Subcellular localization of bacteriophage PRD1 proteins in *Escherichia coli*. *Virus Research*. 179: 44-52.

II Karttunen J*., Mäntynen S*., Ihalainen T.O., Bamford J.K.H. & Oksanen H.M. 2015. Non-structural proteins P17 and P33 are involved in the assembly of the internal membrane-containing virus PRD1. *Virology*. 482: 225-233.

III Mäntynen S., Laanto E., Kohvakka A., Poranen M.M., Bamford J.K.H. & Ravantti J.J. 2015. New enveloped dsRNA phage from freshwater habitat. *Journal of General Virology*. 96: 1180-1189

IV Laanto E*., Mäntynen S*., Gillum A., Marjakangas J., Ravantti J.J., Sundberg L.R., Huiskonen, J.T. & Bamford J.K.H. A new virus type found from a boreal lake - a link between ssDNA and dsDNA internal-membrane containing phages. Manuscript.

* Equal contribution

RESPONSIBILITIES OF SARI MÄNTYNEN IN THE THESIS ARTICLES

I I was involved in constructing the expression vectors. I conducted the sedimentation assays together with Jenni Karttunen. I participated in the data analysis, preparation of the figures and writing the article. Jenni Karttunen performed the complementation assays and confocal microscopy studies. The fluorescence lifetime measurements were conducted and analysed by Jenni Karttunen, Heli Lehtivuori, Nikolai Tkachenko and Janne Ihalainen.

II I participated in the cloning, sedimentation assays and complementation tests. Jenni Karttunen conducted the confocal microscope imaging and FRAP experiments. Teemu Ihalainen performed the computer simulations for the FRAP assays. I analysed the data and prepared the figures with Jenni Karttunen. I wrote the article together with the co-authors.

III I co-supervised the Master´s thesis project during which the virus was purified and characterized. I participated in laboratory experiments such as electron microscopy and viral life cycle studies. I conducted the sequencing and sequence analysis. Janne Ravantti performed the protein modelling. I analysed the data, prepared the figures and wrote the article together with the co-authors.

IV I co-supervised the Master´s thesis project involving the purification and characterization of the virus. I completed the sequencing and conducted the sequence analysis. Ashley Gillum and Juha Huiskonen performed the cryo-EM imaging. I analysed the data and wrote the article together with Elina Laanto.

ABBREVIATIONS

ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
bp	base pair
CFP	cyan fluorescent protein
CFU	colony forming unit
cryo-EM	cryo-electron microscopy
DNA	deoxyribonucleic acid
ds	double-stranded
HSF	Homologous Structure Finder
FRAP	fluorescence recovery after photobleaching
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
eCFP	enhanced cyan fluorescent protein
eYPF	enhanced yellow fluorescent protein
hsp	heat shock protein
ICTV	International Committee on Taxonomy of Viruses
kb	kilo basepair
kDa	kilodalton
LPS	lipopolysaccharide
LUCA	last universal common ancestor
Mb	mega basepair
MCP	major capsid protein
MOI	multiplicity of infection
mRNA	messenger ribonucleic acid
NC	nucleocapsid
NTPase	nucleoside triphosphatase
TEM	transmission electron microscopy
ORF	open reading frame
PC	polymerase complex
p.i.	post infection
PDB id	Protein Data Bank identifier
PE	phosphatidylethanolamine
PFU	plaque forming unit
PG	phosphatidylglycerol
POG	phage orthologous group
RdRP	RNA dependent RNA polymerase
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
ss	single-stranded
T	triangulation number
TMV	tobacco mosaic virus
ts	temperature-sensitive
YFP	yellow fluorescent protein

1 INTRODUCTION

Viruses are submicroscopic nanoparticles, which are unable to carry out metabolic processes on their own. Instead, these obligate intracellular parasites depend on their host cell´s machineries for production of viral progeny (Cann 2005). An infectious virus particle, called the virion, consists of a proteinaceous capsid enclosing the genomic nucleic acid, which contains all the information required to complete viral infection cycle. The protective protein capsid is the feature, which separates viruses from other self-replicating entities, such as plasmids (Krupovič and Bamford 2010). The virion may also contain a lipid component, either covalently attached to a protein or forming a structural membrane bilayer (King *et al.* 2012). The lipid membrane, which covers or underlies the protein capsid, plays a vital role in fundamental steps of the infection cycle of membrane-containing viruses.

Viruses are thought to occupy practically every ecological niche on Earth and infect organisms ranging from animals and plants to prokaryotic (bacteria and archaea) and eukaryotic microorganisms (protozoa, fungi, algae). According to some estimates, there are over 10^{31} virus particles in the biosphere (Breitbart and Rohwer 2005), most of which infect bacteria in marine habitats (Breitbart and Rohwer 2005, Suttle 2007). If these particles would be lined end to end, the chain would stretch out over 200 million light years! And yet this is probably a gross underestimation. In fact, it has been suggested that the number of virus particles in soils equals or even exceeds that of the aquatic environments (Srinivasiah *et al.* 2008). Moreover, a substantial number of viral genomes has either integrated into host genomes or exists as independently replicating genetic elements in the cellular organisms (Casjens 2003). Undoubtedly, viruses form the most numerous type of biological entity on Earth, outnumbering their cellular hosts by at least a factor of ten (Bamford 2003). However, only a fraction of them has been discovered, let alone studied in genetic, structural or biochemical detail. Therefore we are only beginning to grasp the enormous variety displayed by the global virus population.

Viruses elicit an enormous selective pressure on their host cells. The virus-host interaction is often described as an arms-race, with cells constantly

evolving new mechanisms to prevent viral infection and viruses counter-evolving to evade these mechanisms. This, however, is only one aspect of the virus-host coevolution. The relationship between a virus and its host is not necessarily destructive, but the two can cooperate and benefit from each other.

Because viruses cannot self-replicate outside their host cell, it is debatable whether they can be considered living organisms. However, due to their profound impact on the biosphere, comprehensive characterization of viruses is essential for holistic understanding of cell-based life forms.

This doctoral thesis focuses on membrane-containing viruses, using three tailless bacterial viruses as models. Internal membrane-containing bacteriophage PRD1, with double-stranded (ds) DNA genome, is one of the best-known viruses. Still, details of its assembly mechanisms within the host cell have remained unsolved. In this thesis it is shown that plasmid-produced proteins of PRD1 are asymmetrically positioned in the host bacterium, possibly reflecting their oligomerization and/or role in assembly. It is also demonstrated that two virally encoded non-structural PRD1 proteins, P17 and P33, complement a defect in host co-chaperonin in both bacterial growth and phage multiplication, implying a connection with the host chaperonin complex. Moreover, two novel bacteriophages, φNN and FLiP, are introduced. Due to obvious similarities, we propose that φNN is a new member of the *Cystoviridae*, a family of enveloped dsRNA viruses. The unique isolation destination and host specificity of φNN demonstrate that the family is more diverse than has been previously appreciated. Comparisons between φNN and other putative cystoviruses point to conservation of essential viral elements and accelerated evolution of the parts needed in host recognition. FLiP, on the other hand, is the first described icosahedral, single-stranded (ss) DNA virus with an internal membrane under the protein shell. It shows little sequence similarity to anything in the public databases, but intriguingly the overall virion structure resembles remarkably that of marine dsDNA bacteriophage PM2.

2 REVIEW OF THE LITERATURE

2.1 The good, the bad, and the virus

The first evidence of the existence of viruses dates back to the very end of the 19th century. In 1892, a Russian scientist Dmitri Ivanovsky demonstrated that the causative agent of the tobacco mosaic disease passes through porcelain filters, which were known to retain bacteria (Lustig and Levine 1992). Six years later, a Dutch microbiologist Martinus Beijerinck repeated independently Ivanovsky´s experiments and became determined that the disease is caused by a new kind of infectious agent, which he called *virus*, the latin word for toxic or poison, distinguishing it from bacteria and other microbes. Beijerinck deduced from his observations that the disease-causing agent was in liquid state, calling it *contagium vivum fluidum* (a contagious living liquid) (Bos 1999). This theory was later discredited as the electron micrographs of tobacco mosaic virus (TMV) proved its particulate nature (Lustig and Levine 1992). Since the identification of TMV, reports of similar filterable agents, causing diseases in animals and plants, started to accumulate. Another major turning point in the virology research occurred at the beginning of the 20th century as two scientists, Frederick Twort and Felix d´Herelle, independently recognised viruses infecting bacteria, called bacteriophages, or phages for short (Summers 2011). Gradually, the predominance of viruses in the biosphere was becoming evident.

Since their discovery, viruses have had a bad reputation in the public eye. Indeed, we are all familiar with a number of mild as well as more severe diseases caused by viral infections, including the common cold, measles, rabies, influenza and acquired immunodeficiency syndrome (AIDS). Many of the viral diseases are currently in control or even completely eradicated through vaccination and antiviral drugs (Plotkin 2005, Littler and Oberg 2005). However, there are unfortunate occasions when a virus has caused a global epidemic with devastating consequences, such as in 1918 when an unusually virulent influenza virus (the Spanish flu) swept across the world and, according to estimates, caused the death of about 50 million people (Taubenberger and

Morens 2006). Recently, global effort and collaboration was needed to combat the Ebola virus outbreak in West Africa and to prevent it from spreading worldwide. It has also been speculated, whether the coronavirus causing Middle East respiratory syndrome (MERS) could have the potential to become a global health threat. Since 2012, the virus has been circulating in the Arabian Peninsula and in summer 2015 the Republic of Korea faced the largest MERS outbreak outside the Middle East [World Health Organization (WHO), 2015]. In addition to causing infectious diseases, several virus strains are also associated with human cancers. In fact, it has been estimated that 15-20 % of the cancers might be caused by viral infections (e.g. papilloma, hepatitis B and C) (Parkin 2006). Moreover, viral infections of domesticated plants and animals can have considerable social and economic consequences. For instance, the outbreak of foot-and-mouth disease, which took place in the United Kingdom in 2001, resulted in the mass slaughter of millions of sheep and cattle and an estimated economic damage of £8 billion (Law 2007).

A traditional research focus has been on viruses causing diseases on humans, animals and crop plants. However, there has been a recent shift toward more ecology- and environmental-oriented virus studies. It is now appreciated that viruses play a crucial role in maintaining a balance in different ecosystems. Viruses shape the compositions and abundances of their host populations and consequently alter the global bio- and geochemical cycles. For instance, viruses kill ~20 % of the ocean microbe biomass thus causing a fundamental impact on the marine nutrient and energy cycles (Suttle 2007). Due to their pivotal role in the ecosystems, viruses might have an effect on the climate, but it is not exactly known whether their action is exacerbating or attenuating the effects of the global climate change (Danovaro *et al.* 2011). In turn, viruses are influenced by the climate change in several ways, either indirectly via their host interactions or directly, e.g. as a result of elevated temperatures.

In addition to their ecological significance, viruses have also proven to be excellent research tools in laboratory and their study has been crucial in developing modern molecular biology. Especially bacteriophages have been particularly practical models in biological research. Their beneficial features include e.g. their inability to infect humans, relatively simple compositions, easy propagation and also the possibility to modify them genetically. Phages have therefore been instrumental in studying cellular pathways that also apply to higher organisms, and their study has elucidated many fundamental aspects of biological life, such as the nature and structure of genetic material as well as the mechanisms for replication, transcription and translation (Keen 2015). Their propagation inside the bacterial cells has many parallels with virus infection in eukaryotes and therefore phages are currently used to develop antiviral drugs. Since the increasing antibiotic resistance has become a global problem, there has also been renewed interest in using phages as therapeutic agents against bacterial infections. Other potential applications of phages include their use as vaccines and as gene therapy delivery vehicles. In addition, in the field of

biotechnology, a broad research interest has been focused on phage enzymes that may have commercial value or could be utilized in various molecular biology techniques (e.g. DNA and RNA polymerases).

2.2 Virion structure

The simplest version of a virion is composed of a protein shell encasing the viral genome. The protein shell, called the capsid, has a dual function. Firstly, it has to be sufficiently rigid to protect the viral genome against chemical and physical assaults (e.g. radiation, nucleases as well as extreme temperature and pH conditions), that the virion encounters during its passage through extracellular space. Secondly, the virus particle has also to recognise its specific receptor molecules on host cell surface and subsequently deliver the genome into the cell interior to initiate the infection cycle. Therefore, the capsid is actually a dynamic, metastable structure, which can undergo major conformational changes to initiate virus propagation in the host cell (Prasad and Schmid 2012).

Viruses are not in any way a homogenous group of particles. Unlike the genomes of cellular life forms, which are uniformly composed of DNA, the genetic information of viruses is carried by either DNA or RNA. Moreover, the nucleic acid molecule can be either single- or double-stranded and exist inside the protein capsid in different configurations (e.g. in linear, circular or segmented form). The viral genome sizes range three orders of magnitude (from <2 kb to >2 Mb) with the largest viral genomes matching those of some prokaryotes (Allan and Ellis 2000, Philippe *et al.* 2013). The viral diversity is also highly conspicuous at the level of genetic content. In fact, viruses are believed to represent the greatest reservoir of unexplored genetic diversity on Earth (Suttle 2007). The viral genetic versatility is discussed in more detail in section 2.5.1.

Even though most virus particles are about a hundred times smaller than average bacterial cells, the size range between the smallest and largest known viruses is astonishing. The ssDNA virus Porcine circovirus 1 (PCV1) is only ~20 nm in diameter (Allan and Ellis 2000), corresponding the size of a ribosome. The ~1.7-kb genome of PCV1 has the minimal number of genes required for viral reproduction: one gene encoding the replication initiation protein and another coding for the structural protein forming the capsid (Krupovič and Bamford 2010). On the other hand, there are also considerably more complex virus types that are comparable in size to small bacteria. The largest known virus, Pithovirus sibericum, has an astonishing length of 1.5 µm and its ~610-kb genome encodes approximately 467 different proteins (Legendre *et al.* 2014). Interestingly, this genome size is actually rather small compared to those of other giant dsDNA viruses. For instance, Pandoraviruses, which are about 1-1.2 µm long, harbour up to 2.8 Mb long genomes, with a coding capacity for estimated 2500 proteins [~93% of the open reading frames (ORFs) with unknown function] (Philippe *et al.* 2013). Furthermore, viruses display a variety

of shapes (King *et al.* 2012). Compared to the enormous genetic diversity, though, the number of viral morphotypes seems to be rather limited with a few prevalent categories discussed below.

2.2.1 Virion morphologies

Since the coding capacity of viral genomes is limited, capsids constitute from multiple copies of one, or alternatively a few, chemically distinct protein subunits, which are often organized symmetrically. The simplest way to form a virion is to arrange the protein subunits into a helical array around the genomic nucleic acid, which may conform into a similar helical path. These helically symmetric viruses appear as rigid rods or flexible filaments. TMV (Namba and Stubbs 1986) is the best-described member of this virus group, but also several other plant viruses, as well as a number of bacteriophages and nucleocapsids (NCs) of enveloped animal viruses have helical symmetries (King *et al*. 2012).

Another common morphological type is virions with icosahedral symmetry. These have two-, three-, and fivefold rotational symmetry axes, 20 equilateral triangular faces, 12 pentagonal vertices and 30 edges. In the simplest form of the icosahedral capsid, each triangular face consists of a trimer of a single protein subunit. Therefore 60 chemically identical subunits, clustered as 12 pentamers and making identical (equivalent) contacts with each other, constitute the shell structure (Fig. 1A). However, most viruses possess genomes that are far too large to fit into these smallest capsids and more subunits are required to form the icosahedron. Due to structural considerations, more than 60 subunits cannot occupy strictly equivalent positions on icosahedral surface. However, Caspar and Klug (1962) postulated that this problem could be solved by tolerating small changes in subunit conformation and inter-subunit bonding. In resulting capsid structures, the pentamers still occupy the vertices, but additional subunits with sixfold symmetry are arranged between the fivefold symmetric units (Fig. 1B and C). Here, protein subunits do not occupy identical environments but are in quasi-equivalent contact with each other. Each capsid has a fixed amount of pentamers (12), but the number of hexamers is 10(*T*-1), where *T* stands for the triangulation number. The triangulation number can only take distinct values (1, 3, 4, 7…), following the equation: $T = h^2 + hk + k^2$, in which h and k are integers. In some cases, triangulation rules have to be modified to describe particular capsid structures. For instance, the inner core of dsRNA viruses has an unusual $T = 1$ architecture consisting of 60 asymmetric dimers of a single polypeptide (sometimes referred to as the "$T = 2$" shell) (Grimes *et al.* 1998, Naitow *et al.* 2002, Nakagawa *et al.* 2003, Huiskonen *et al.* 2006, McClain *et al.* 2010). Also, when the capsid is composed of more than one polypeptide building block, pseudo *T*-numbers are used. For instance, adenovirus and bacteriophage PRD1 exhibit a unique pseudo $T = 25$ lattice (Stewart *et al.* 1991, Benson *et al.* 1999), in which pentameric and hexameric clusters are formed by two different protein subunits. The hexavalent positions are occupied by a trimeric capsid protein, which has a basal hexagonal shape, whereas the fivefold vertex bases are constituted by a pentameric protein.

There are a number of more elaborate virion morphologies which do not strictly fit into either of the above categories or are modifications of the prevalent morphotypes (King *et al*. 2012). For instance, some of the icosahedral capsids are prolate, i.e. elongated along the fivefold axis of the icosahedron, or consist of several concentric protein shells. Complex viruses may also contain appendages protruding from the capsid surface, such as spikes, fibres or tail-structures. In fact, the structure of most known bacterial viruses is a combination of the two basic symmetries, with a helical tail attached to the icosahedrally symmetric capsid (so called tailed phages). In addition, many viruses contain a membrane structure, which can be located either internally, underneath the protein capsid, or externally covering the capsid structure (enveloped viruses). Pleomorphism is most prominent among enveloped viruses since the membrane has the tendency to adapt different sizes and shapes, thereby preventing the averaging of virus particles. More unusual morphotypes have been found among archaeal viruses that can have the shape of e.g. a spindle, a tear drop or even a bottle (Atanasova *et al.* 2015).

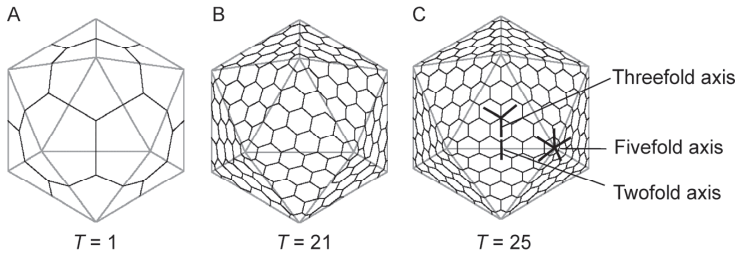


FIGURE 1 Capsids with either A) $T = 1$, B) $T = 21$ or C) $T = 25$ icosahedral symmetry. The simplest icosahedral capsids are composed of 60 identical subunits arranged into pentamers. However, larger capsids require additional hexameric building blocks inserted between the pentamers. Capsid structures were modified from the Virus Particle Explorer (Viper) database (Shepherd *et al.* 2006).

2.3 Virus assembly

Formation of virus particles exemplifies the elegance of natural engineering. Virions are assembled in the crowded intracellular environment from numerous copies (~60-10 000) of different protein subunits and other particle components (Perlmutter and Hagan 2015). Nevertheless, the process occurs with impressive precision and within biologically relevant time scales. The virion assembly includes the formation of the protein shell and the encapsidation of the genomic nucleic acid, occurring either simultaneously or subsequently. Possible presence of an internal or external membrane structure adds additional complexity to the virion morphogenesis (see section 2.6.1). The

construction of a virion may proceed in a linear fashion, with components added subsequently to the developing virion structure. Alternatively, the assembly pathway may be branched, as in bacteriophage T4, which builds up its head, tail and tail fibres separately, before combining them into infectious virus particle (Casjens and Hendrix 1988). Each assembly pathway of a virus is initiated by an accumulation of a small number of viral proteins (Prasad and Schmid 2012). Interactions within this nucleation center elicit conformational changes, consequently creating new binding sites for additional building blocks (e.g. pentamers or hexamers of capsid protein) and so forth. This enables the assembly process to proceed along the specific pathway in correct order. The final structure is at a local energy minimum compared to the free building blocks, and therefore the whole assembly pathway is driven by the laws of thermodynamics (Caspar and Klug 1962).

Many simple viruses are assembled spontaneously from their respective protein and nucleic acid components, without any additional factors influencing the process. The ability to self-assemble implies that the capsid proteins possess intrinsically all the information needed for virion formation. However, more complex viruses usually require additional proteins, encoded by the virus or the host, to ensure the correct size and shape of the virion in timely fashion. These accessory factors include scaffolding proteins, which associate with assembly intermediates transiently, but are excluded from the final virion structure (Dokland 1999). Viruses also possess a variety of other morphogenetic factors, which do not carry out a scaffolding function, but assist the virion formation in some other manner, which is often difficult to elucidate.

Viruses employ one of two strategies to encapsidate their genome: co-assembly of nucleic acid and protein components or packaging of genome into preformed empty particles. Both mechanisms include recognition of a specific region at the viral genome (packaging signal). Co-assembly is a process in which the formation of the protein capsid and the nucleic acid encapsidation take place synchronously. This assembly mechanism was demonstrated for the first time by the *in vitro* reconstruction of infective TMV in the 1950s (Fraenkel-Conrat and Williams 1955), and has since been reported for a number of filamentous viruses with ssDNA and ssRNA genomes (King *et al*. 2012). The assembly process of TMV is best known among helically symmetric viruses. The building blocks of the TMV virion are oligomeric discs, appearing as two-layered structures. The assembly is nucleated as a specific stem-loop structure of the ssRNA genome is inserted into the central hole of the two-ring disc (Butler 1999). This elicits melting of the stem structure, exposing more RNA for protein binding. The interaction between the protein and RNA components changes the conformation of the disk into a helical symmetry. As additional discs or smaller protein aggregates are assembled around the nucleic acid, the protein capsid elongates to both directions, simultaneously trapping the ssRNA genome into helical array (Butler 1999). Interestingly, simple icosahedral viruses of *Leviviridae* family also seem to apply co-assembly for virion formation. The assembly of levivirus MS2 has been shown to commence by an

interaction between a capsid protein dimer and a specific stem loop region of the ssRNA genome (Stockley *et al.* 2007). This nucleation event triggers the addition of other capsid protein dimers, ultimately forming the icosahedron.

Certain viruses employ a variation of co-assembly, in which the viral genome is condensed with specific proteins prior to encapsidation. The resulting compressed nucleoprotein complex is subsequently covered by a protein shell or alternatively a lipid envelope. For instance, the circular dsDNA genome of papovavirus Simian virus 40 (SV40) associates with cellular histones before being incorporated into capsid (Polisky and McCarthy 1975). It has been shown that a host transcription factor recruits the capsid proteins of SV40 to the packaging signal, forming a nucleation center for the assembly (Gordon-Shaag *et al.* 2002). Subsequently, building blocks assemble around the genome, resulting in the mature virion.

Packaging of genomes into preformed capsids (procapsids) has been reported for a number of different viruses, such as herpesviruses, ssDNA bacteriophage φX174, and two membrane-containing phages, φ6 (see section 2.9.3) and PRD1 (see section 2.7.3), with dsDNA and dsRNA genomes, respectively. However, it is most comprehensively described among tailed dsDNA phages. The mechanism is initiated by the formation of a procapsid, a structure consisting of an empty protein shell and a portal vertex. The portal (or the connector) is an oligomeric ring-like structure of 12 identical subunits (Feiss and Rao 2012). It resides at a specific vertex, through which the viral genome is threaded into the procapsid upon packaging. In tailed phages, the portal also doubles as the site for DNA injection into bacterial cells. The portal is usually involved in the initiation of the procapsid assembly, and has also been shown to play a role in determining the correct shape and size of phages SPP1 (Dröge *et al.* 2000) and φ29 (Guo *et al.* 1991). In the newly synthesized procapsid, the portal serves as a docking point for proteins needed in the genome packaging. In tailed phages, it also forms the site for tail attachment, after packaging has been completed.

Non-structural scaffolding proteins are universally found inside the procapsid (eg. in herpesviruses as well as in phages λ, P22 and φX174) but may also reside externally (eg. in phages P4 and φX174) [reviewed by (Dokland 1999)]. Scaffolding proteins are removed from the procapsid either concurrently with the packaging or prior to it, in order to accommodate the viral genome. They are cleaved by virally encoded proteases or alternatively recycled for further rounds of assembly, as in phage P22 (King and Casjens 1974). Phage HK97 is an exception among tailed phages, since it lacks scaffolding proteins. However, the delta domain of HK97 capsid protein has apparently adopted the scaffolding function and is cleaved from the final structure by a virally encoded protease (Duda *et al.* 1995). Another variation of scaffolding is seen in PRD1, where the internal membrane, enriched in protein P10, seems to serve as a scaffold for virion assembly (Rydman *et al.* 2001) (see section 2.7.3).

The translocation of viral genome is best characterized for tailed dsDNA phages and herpesviruses, but the same principles may also apply for tailless,

icosahedral phages. Also, dsRNA viruses and certain ssDNA viruses have been shown to use procapsid-state in their assembly, but the details of their packaging process differ notably from those of dsDNA phages (Fane *et al.* 2006, Mindich 2012). The packaging of the viral dsDNA genome at the portal vertex is powered by enzyme-catalyzed hydrolysis of adenosine triphosphate (ATP) (Feiss and Rao 2012). The viral genome is usually incorporated into procapsid as a linear molecule. This applies for the ssRNA precursors of cystoviruses (Poranen *et al.* 1999) and the unit-length genomes with terminal proteins found in PRD1 and φ29 (Grimes *et al.* 2002, Strömsten *et al.* 2003a, Strömsten *et al.* 2005). In both cases, the genomes are readily packaged into the procapsid by the packaging adenosine triphosphatase (ATPase). However, concatameric DNA genomes (e.g. in phages λ, T4, T7 and SPP1) need to be cleaved into units, which are appropriate for packaging (Feiss and Rao 2012). The cleavage is conducted by a terminase enzyme, which usually consists of two subunits with complementary functions. The small subunit attaches to a specific site at the genome, simultaneously setting the large subunit in correct position for cleavage (Rao and Feiss 2008, Feiss and Rao 2012). The large subunit contains ATPase and endonuclease activities and also mediates the interaction with the procapsid (Rao and Feiss 2008, Feiss and Rao 2012). This subunit contains two conserved motifs, Walker A and Walker B (Mitchell *et al.* 2002), required for binding ATP (Walker *et al.* 1982) and Mg^{2+}. Terminases conduct the initial cutting of the concatameric genome at or near a specific site (*pac*-site in e.g. T4 and SPP1 and *cos*-site in λ), after which the terminase translocates the generated DNA terminus into the procapsid (Feiss and Rao 2012). The packaging continues until the phage head is filled with DNA. The second cleavage may occur sequence-specifically (e.g. in λ), creating unit-length genomes. Alternatively, the concatameric DNA may be cut non-specifically as a certain amount of DNA has been packaged (in *pac* phages ~102-110 % of the genome length), resulting in terminal redundancy (Rao and Feiss 2008).

The genome packaging commonly expands the capsid, increasing its DNA capacity by 50-100 % (Rao and Feiss 2008), and elicits conformational changes both at the level of individual capsid proteins and in overall virion structure. Drastic structural changes include e.g. the cleavage of the phage T4 capsid proteins (Casjens and Hendrix 1988) or the cross-linking of the capsid proteins in phage HK97 (Duda *et al*. 1995). During virion maturation, capsid subunits may change their position or even refold certain regions, as phages P22 and HK97 (Conway *et al.* 2001, Jiang *et al.* 2003). Also, virally encoded assembly factors (e.g. proteins P10 and P17 in PRD1; Mindich *et al*. 1982a) (see section 2.7.3) and host chaperonins (Georgopoulos *et al.* 1973, Hänninen *et al*. 1997) (see section 2.13.1) are often required for achieving the correct conformation of structural proteins and consequently the proper virion structure.

2.4 Virus classification

For several decades after the first discovery of viruses (see section 2.1), there was no uniform system for their classification. Consequently, viruses were named rather randomly according to a number of distinct properties, such as the name of the person who originally discovered them (e.g. Epstein-Barr virus), the associated disease (e.g. rabies virus or poliovirus) or the location from which they were isolated (e.g. Coxsackievirus) (Flint *et al.* 2008). Originally viruses were mainly described by their clinical and pathogenic properties. Also, ecological impact and mechanism for transmission were often acknowledged when grouping viruses. Over the years, advances in virology research provided new tools for virus taxonomy. Since late 1940s, electron microscopes became more accessible to the biological research community, enabling the discrimination of viruses based on their morphological characteristics (Almeida 1963). In the 1960s, the virus classification was further augmented by new biochemical methods, which enabled the isolation of the nucleic acid from the virus, and subsequent determination of its size and nucleic acid type (Nelson 2004).

In 1962, Lwoff, Horne and Tournier proposed a comperensive system for classifying viruses using classical Linnaean system including phylum, class, order, family, genus and species (Lwoff and Tournier 1966). In this scheme, hierarchical divisions were made based on certain physical characteristics of viruses, such as the type of the nucleic acid, the symmetry of the protein capsid, the possible presence of a lipid envelope and the dimensions of the virion. Even though the original Lwoff scheme did not stand the test of time, the idea of grouping viruses based on their *own* inherent properties, instead of the properties of their hosts, laid down the ground rules for modern virus classification systems.

After the nucleic acid sequencing techniques were developed in the 1970s, the contribution of viral genomics to taxonomic systems has grown. In fact, as the genome contains all the necessary information for the virus to complete the infection cycle and to perform its biological functions, it is often considered as the most relevant feature for virus classification. In the early 1970s, David Baltimore devised a classification system, which distinguishes viruses based on their genomic nucleic acid composition as well as the strategies for replication and expression (Baltimore 1971). The scheme is founded on the fact that all viruses must synthesize positive-sense messenger RNA (mRNA) to produce proteins and reproduce themselves. The Baltimore system places viruses into classes depending on the chemical nature of the nucleic acid (DNA or RNA) and the number of nucleic acid strands (single- or double-stranded). Additionally, the ssRNA molecules are divided into two groups based on the polarity of the nucleic acid helix. Positive-sense RNA can be directly translated into proteins whereas negative-sense RNA must first be transcribed into "readable" positive-sense form. Therefore each of the virus

groups contains viruses with either dsDNA, ssDNA, dsRNA, positive-sense ssRNA or negative-sense ssRNA genome. Two additional virus classes employ reverse transcriptase in replication. These include ssRNA viruses with a DNA intermediate in the life cycle and dsDNA viruses that replicate their genetic material through an RNA intermediate.

The first global initiative to bring order in the enormous variety of viruses took place at the International Congress of Microbiology, which was organized in Moscow in 1966 (King *et al*. 2012). The meeting established a committee, which was later named the International Committee on Taxonomy of Viruses (ICTV). The ICTV seeks to develop a universal system for classifying all viruses infecting animals, plants, fungi, bacteria and archaea (King *et al*. 2012). The committee has also set internationally accepted guidelines for virus nomenclature. Viruses are predominantly discriminated on basis of the virion morphology as well as the genome type and replication strategies (thereby paralleling the Baltimore scheme). Additionally several other properties, such as the relatedness of genome sequences, host range, host cell and tissue tropism, pathogenicity, mechanism of transmission as well as several physicochemical and antigenic properties, are taken into account (King *et al*. 2012). According to these criteria viruses are assigned to the hierarchical levels of order (suffix -*virales*), family (-*viridae*), subfamily (-*virinae*), genus (-*virus*) and species, thereby resembling the classical Linnean system. Today, the ICTV recognises 2827 virus species delineated into 7 orders, 104 families, 23 subfamilies and 455 genera (ICTV 2015). However, countless virus isolates have remained to be unassigned due to inadequate structural and genetic characterization.

The ICTV classification scheme has had a profound influence on the way we envision the virus world. Nevertheless, critics have pointed out several shortcomings in this system. It has been claimed that the scheme is obsolete and largely overlooking the genomic and proteomic information accumulating in the databases (Nelson 2004). For instance, the ICTV has assigned both *Salmonella* phage P22 and *Escherichia coli* phage T7 to the *Podoviridae* family based on their short tail structures. Yet, P22 is at sequence level so closely related to the long-tailed phage λ that recombination between their respective genomes results in functional hybrids (Botstein and Herskowitz 1974). This raises a question, whether P22 should in fact be grouped together with the lambdoid-type bacteriophages of the *Siphoviridae* (Nelson 2004).

There is also pressure for establishing higher taxonomic classes. Currently, the ICTV classifies viruses into seven orders, which encompass only about one fourth of the assigned viral families. No higher taxonomic level, beyond the level of order, has been assigned. The ICTV classification scheme assumes that all members of a given taxon descended from a common ancestor. However, viruses evolve extremely fast and therefore deeper phylogenetic relationships have turned out to be difficult to detect. In addition to sequence divergence, viral evolution involves high levels of genetic recombination. This blurs the genetic relationships among viruses, further complicating the evolutionary analysis. For instance, two viruses may share practically identical polymerase

genes for replication but at the same time possess structural protein genes which are distantly related. How should these viruses be classified? The increasingly popular view in the scientific community is that the ICTV scheme should be revised and complemented by present-day genetic and structural data.

2.5 Viral phylogenetic relationships

2.5.1 Comparative genomics

Evolutionary relationships of cellular life forms are ubiquitously analysed by determining the 16S (in prokaryotes and mitochondria) and 18S (in eukaryotes) ribosomal RNA (rRNA) sequences (Woese and Fox 1977, Woese *et al.* 1990). However, viruses do not share any universally conserved genes or gene products (Rohwer and Edwards 2002), making their classification and phylogenetic studies more challenging. The analysis of virus evolution is further complicated by the apparent lack of fossil records. Nevertheless, the accumulating genetic and structural data has provided interesting new insights into the virus relationships and their evolutionary origin.

In recent years, the number of viral sequences available in the public sequence databases has increased exponentially. This is the direct consequence of more efficient techniques for virus isolation (Hurwitz *et al.* 2013) and sequencing (Hall 2007). Metagenomics has provided novel sequence data by allowing culture-independent analysis of viral genetic material directly from the environment [reviewed by (Mokili *et al.* 2012)]. The increased sequence information allows more comprehensive comparisons of viral genomic features (e.g. gene synteny, distinct genes or other genetic loci), which could be utilized in grouping viruses and establishing their phylogenetic relationships. Computational methods are commonly used to detect analogous or homologous genes from different virus species. Also, the relatively small genome size of viruses enables to extend the comparisons over the complete genome sequences. Genetic sequences, either at the nucleic acid or amino acid level, are aligned using sequence alignment algorithms and the resulting data is used to construct phylogenetic trees reflecting the evolutionary relationships. The idea is that the evolutionary history between two viruses, or at least the distinct genetic loci, can be deduced from their sequence variance. The higher the genetic variance, the more ancient was the divergence from a common ancestor. Comparisons of viral genetic elements have unambiguously denoted the enormous genetic diversity of viruses. Often, when the genome sequence of a newly isolated virus is determined, the majority of ORFs represents "dark matter" without any detectable homologs in the current sequence databases (Koonin and Dolja 2013).

The high genetic variance among viruses reflects their parasitic life style. Viruses, especially the ones infecting bacteria, replicate rapidly and are prone to

high mutation rates. New genetic compositions are frequently created by recombination and genetic material is obtained from unrelated sources via horizontal gene transfer. Genetic material is exchanged not only between two viruses co-infecting the same host cells, but also between a virus and its host cell. Bacterial viruses may also swap genetic material with a prophage residing in the host cell.

Horizontal gene transfer has had a profound impact on the evolution of viruses. In many instances it has transformed the viral genomes into mosaics of genes with differing evolutionary histories (Hendrix 2002, Filée and Chandler 2010, Krupovič and Bamford 2010). Genetic comparisons have revealed that this mosaicism is especially prominent among largely studied tailed bacteriophages, which are constantly exchanging genes, gene domains or modules, i.e. gene clusters encoding functionally related proteins (Hendrix *et al.* 1999, Brüssow and Hendrix 2002, Krupovič *et al.* 2011). Due to this genomic mosaicism, it is practically impossible to infer a single phylogeny for all the genes in the phage genome, which complicates considerably the study of virus evolution.

Despite the significant genetic variance, some level of conservation can be detected among viral genomes. For instance, a comprehensive analysis of prokaryotic virus genomes identified > 4500 sets of orthologous genes, called phage orthologous groups, POGs (Kristensen *et al.* 2013). The genes in these groups encode proteins functioning in e.g. genome packaging or virion architecture. POGs may provide means to track down deeper evolutionary relationships, even among highly mosaic phage genomes. POGs can also be used as genetic markers to detect the presence of a virus belonging to a certain taxon from metagenomic data (Kristensen *et al*. 2011, 2013).

It has been suggested that viral genes could be divided into "self" and "non-self" genes (Bamford *et al*. 2002, 2005, Bamford 2003). The "self" genes encode essential proteins for the virion architecture and assembly, such as the major coat protein (MCP) and the genome packaging nucleoside triphosphatase (NTPase). These genes are highly conserved among a group of related viruses and likely to have inherited vertically. In contrast, the functions of the "non-self" genes are usually related to genome replication or specific host interactions, either through cell recognition, cell entry or release of virus progeny. These genes are frequently swapped via horizontal gene transfer and are thought to provide means to adapt to new environments and/or hosts (Bamford *et al.* 2002, 2005, Bamford 2003, Saren *et al.* 2005, Krupovič and Bamford 2007, Abrescia *et al.* 2012).

2.5.2 Structure-based viral lineages

Comparative genomics is a feasible tool to track down phylogenetic relationships among viruses, which are relatively closely related. However, viruses are ancient, perhaps even outdating the last universal common ancestor of cellular life (LUCA) (Forterre 2010). As time passes, viral genomes evolve until no recognizable sequence similarity can be detected. Therefore other measures are needed to bring the deeper phylogenetic relationships into the

limelight. Interestingly, even when the nucleic acid and amino acid sequences have diverged beyond the point of recognition, viral structures may be conserved. High-resolution structural data on viral MCPs and entire virion structures have enabled comprehensive structural comparisons resulting in categorization of viruses based on common architectures and MCP folds (Benson *et al.* 1999, Bamford *et al.* 2002, 2005, Bamford 2003, Abrescia *et al.* 2008). The number of feasible viral architectures seems to be rather limited. This presumably reflects strict physicochemical constraints, which limit the ways to fold a native protein structure from an amino acid chain. The folding space is especially restricted for major capsid proteins, with only a small subset of folds having the potential to construct a functional virus capsid (Abrescia *et al.* 2012, Oksanen *et al.* 2012). As the high resolution structures of the viral MCPs are gradually accumulating, it would seem that most of them fit into one of four major lineages (Abrescia *et al.* 2012).

One of the structure-based viral lineages is typified by the tailed enterobacteria phage HK97 (Wikoff *et al.* 2000). A similar canonical MCP fold detected in HK97 can be found in other tailed phages, eukaryotic herpesviruses and archaeal viruses displaying icosahedral tailed morphology (Baker *et al.* 2005, Bamford *et al.* 2005, Krupovič and Bamford 2010, Pietilä *et al.* 2013). Therefore the lineage encompasses members of all three domains of life.

Another well-established viral lineage contains predominantly small icosahedral RNA viruses (Abrescia *et al.* 2012), such as rhinovirus and poliovirus (Hogle *et al.* 1985, Rossmann *et al.* 1985). In this picornavirus-like lineage, the MCP displays eight-stranded, antiparallel β-barrel fold, also called as the jellyroll fold, which is oriented parallel to the plane of the capsid. The single jellyroll fold is mostly found in positive-sense ssRNA viruses infecting animals, plants and insects, but has also been reported for ssDNA phage φX174 (Dokland *et al.* 1997) as well as for eukaryotic papilloma and polyomaviruses with dsDNA genomes (Stehle *et al.* 1996, Chen *et al.* 2000). Therefore the lineage covers different viral genome types (positive-sense ssRNA, ssDNA and dsDNA) and hosts from two different domains (Eukarya and Bacteria).

The third viral lineage encompasses a variety of dsRNA viruses. This is a highly heterogenous group with a broad spectrum of hosts, genome compositions (1-12 genome segments), capsid sizes (30-100 nm) and morphologies (Abrescia *et al.* 2012). However, all these viruses face the same fundamental challenges when infecting a cell. Firstly, host cells lack the ability to replicate the viral dsRNA genome. Secondly, the exposed dsRNA is considered to be foreign material in cellular environment and once it is detected, a strong apoptotic response is usually elicited (Bamford *et al.* 2002). To overcome these obstacles, the dsRNA genome segments are delivered into the host cell inside a conserved core structure, which carries all the enzymatic functions needed for virus replication and transcription. In addition to this core structure, the virion of a dsRNA virus may also contain additional protein layers. These outer shells facilitate specific virus-host interactions and therefore display more variability (Bamford *et al.* 2002). Bluetongue virus (BTV), an

animal virus belonging to the *Reoviridae* (Grimes *et al.* 1998), is usually considered as the type member of the dsRNA virus lineage. The segmented genome of BTV is enclosed by a core composed of 120 copies of the MCP, arranged in 60 asymmetric dimers. As mentioned in section 2.2.1, this unusual icosahedral arrangement (T = 1 structure) was not predicted by the quasi-equivalence theory of Caspar and Klug (Caspar and Klug 1962). Nevertheless, the unique inner shell architecture is shared by other members of the *Reoviridae* (Nakagawa *et al.* 2003, McClain *et al.* 2010), bacteriophage φ6 of the *Cystoviridae* family (Huiskonen *et al.* 2006) and yeast L-A virus, which is an intracellular parasite belonging to the *Totiviridae* (Naitow *et al.* 2002).

The fourth structure-based viral lineage comprises a number of icosahedral dsDNA viruses, with a distinct MCP fold. This lineage was established when the surprising topology similarity between the MCPs of human adenovirus and bacteriophage PRD1 was revealed (Athappilly *et al.* 1994, Benson *et al.* 1999). In fact, the whole concept of structure-based viral lineages was founded based on this discovery (Bamford *et al.* 2002, Bamford 2003). The MCPs of the PRD1-adenovirus lineage contain two concatenated β-barrels (jellyrolls), which are stacked against each other in a specific manner. The two β-barrels lack any significant sequence similarity but are topologically related (Benson *et al.* 1999). The MCPs exist as trimers, forming pseudohexagonal building blocks, which assemble as triangular plates to form the icosahedral lattice. While the single β-barrels of the picornavirus-like lineage orient normal to the capsid surface, the double β-barrels adopt an upright position. In addition to the common MCP fold, adenovirus and PRD1 share the same overall virion architecture [T = 25 lattice (Stewart *et al.* 1991, Butcher *et al.* 1995)], structure of the penton and spike proteins (van Raaij *et al.* 1999, Abrescia *et al.* 2004, Zubieta *et al.* 2005, Merckel *et al.* 2005), genome type (dsDNA with inverted repeats) and replication mode [protein-primed replication, reviewed by (Calendar 2006)]. This strongly implies that the two viruses descended from a common ancestor. Later similarly folded MCPs have been found in viruses infecting all three domains of life. These include crenarchaeal *Sulfolobus* turreted icosahedral virus (STIV) (Khayat *et al.* 2005), *Paramecium bursaria Chlorella* virus 1 (PBCV-1) infecting green algae (Nandhagopal *et al.* 2002) and marine bacteriophage PM2 (Abrescia *et al.* 2008). The capsid protein of phage PM2 is lacking the elaborate loop extensions, which protrude from top of other MCPs in the PRD1-adenovirus lineage. Also, the two β-barrels of PM2 MCPs resemble each other more notably than those of other members of the lineage, implying that the capsid protein of PM2 constitutes the most ancient form of the lineage (Abrescia *et al.* 2008).

Homology modelling has further expanded the lineage by detecting the canonical double β-barrel fold in a number of other virus isolates, such as in large eukaryotic dsDNA viruses Chilo iridescent virus (CIV) (Yan *et al.* 2009), Mimivirus, and African swine fewer virus (Benson *et al.* 2004) as well as in two euryarchaeal proviruses (Krupovič and Bamford 2008a). Interestingly, the distinct MCP fold is also adopted by a scaffolding protein during vaccinia virus

morphology (Bahar *et al.* 2011). This led to the assignment of vaccinia viruses into the double β-barrel lineage, despite the aberrant morphology (brick-shaped) of the mature virion. Recently, the double β-barrel MCP fold was reported for virophage Sputnik (Zhang *et al.* 2012), further emphasizing its wide distribution.

The structural analysis of *Thermus thermophilus* phage P23-77 gave interesting new insights into the PRD1-adenovirus lineage (Rissanen *et al.* 2013). The capsid of P23-77 contains two major protein species, VP16 and VP17, both of which contain a core fold of a single β-barrel, which resembles the ones forming a pair in the MCP of the PRD1-adenovirus lineage. The single β-barrels of P23-77 show closest structural similarity with the MCP of bacteriophage PM2, allegedly the primeval member of the double β-barrel lineage. The atomic structures of the MCPs of P23-77 fit well into the cryo-EM density map of haloarchaeal virus SH1 (Jäälinoja *et al.* 2008), implying structural resemblance (Rissanen *et al.* 2013). These findings shed light on the origin of the PRD1-adenovirus lineage. It has been hypothesized that the ancestral virus of the lineage had a capsid protein composed of a single β-barrel. Possibly the capsid protein evolved through several gene duplications. One of these duplications resulted in a virus with two distinct capsid proteins, both containing a single β-barrel fold (predecessor of P23-77 and SH1), while another duplication led to a virus with two single β-barrel proteins, which later diversified and fused together, giving rise to the present-day double β-barrel MCPs. Therefore, it would seem that viruses in the PRD1-adenovirus lineage have adopted different strategies for forming the virion lattice. The utilization of double β-barrels in constructing the virion has been proposed to be less error-prone compared to the use of single β-barrels (Jäälinoja *et al*. 2008). Moreover, double β-barrel trimers are considered to be more stable building blocks than hexamers (Krupovič and Bamford 2008b). Recent studies of *Haloarcula hispanica* icosahedral virus 2 (HHIV-2), a genetic and structural relative of SH1, imply that vertical single β-barrel MCPs use homo- and heterodimers, instead of hexamers, as building blocks (Gil-Carton *et al.* 2015). The dimers assemble into lattices, stabilized, in the case of HHIV-2, by disulfide bonds. The proper assembly requires contacts between the capsomers and the internal membrane vesicle (Gil-Carton *et al*. 2015).

The structure-based classification system seems to apply well for icosahedrally symmetric viruses. Challenges arise when considering more asymmetric pleomorphic or spindle-shaped as well as helical viruses, which are largely devoid of structural data (Abrescia *et al.* 2012). However, it was shown that archaeal dsDNA viruses of *Rudiviridae* and *Lipothrixviridae* families share the MCP fold of four helix bundles (Goulet *et al.* 2009), implying that viral lineages could also be extended to helically symmetric viruses (Abrescia *et al.* 2012). Difficulties in delineating the viral "self" in enveloped viruses hinder their classification into lineages. Nevertheless, it has been proposed that the fusogenic membrane glycoproteins and/or the nucleocapsid structure could be

utilized to trace phylogenetic relationships among enveloped viruses (Abrescia *et al.* 2012).

At least two of the established viral lineages (HK97 lineage and PRD1-adenovirus lineage) contain members of all three domains of cellular life, implying that the lineages are ancient, predating the LUCA (Abrescia *et al.* 2012). Due to the limited protein folding space, it is naturally possible that the similar MCP folds have arisen independently in different viral lineages (convergent evolution) instead of being inherited vertically (divergent evolution). The more structural details are shared between distinct viruses, the more probable is their homology (Abrescia *et al.* 2012). However, when the structural similarities between the viruses are more subtle, other "self" properties, such as packaging mechanisms, need to be compared. For instance, all membrane-containing members of the PRD1-adenovirus lineage encode a putative packaging NTPase with a specific P9/A32 motif, supporting the hypothesis of their common ancestry (Strömsten *et al*. 2005).

2.6 Bacteriophages

Bacteriophages are perhaps the most abundant and genetically divergent entities on our planet. Upon entering the host bacterium, phages are traditionally thought to undertake either lytic or lysogenic reproduction cycle. Lytic phages usually take immediate control of the host bacterium´s biosynthetic machinery; new virions are produced and ultimately released, leading to the complete lysis of the host cell (Weinbauer 2004). In contrast, temperate phages can either elicit the lytic infection or alternatively enter lysogenic cycle, in which the phage genome is integrated into the host genome or maintained extrachromosomally, either in linear or circular form, as in e.g. phages N15 (Ravin and Shulga 1970) and P1 (Lobocka *et al.* 2004). In this so called prophage stage, the virus is dormant with most of its genes shut off (Casjens 2003). The viral genome is replicated in synchrony with the cell cycle, and passed on to the next generation. Damaging conditions, such as ultraviolet light or certain chemicals, can induce the repressed prophage, triggering the lytic reproduction cycle. In addition to lytic and lysogenic cycles, there is now mounting evidence of alternative virus-host relationships. Pseudolysogeny in usually described as being a state in which viral genome resides inactive within a nutrient-depleted bacterium, without integrating into the host chromosome or triggering lytic response (Ripp and Miller 1997). Some phages may also cause chronic infection, with new virus particles constantly extruding or budding off without disrupting the physical integrity of the host bacterium (Weinbauer 2004).

To date over 6000 prokaryotic viruses have been described (Ackermann and Prangishvili 2012). In recent years, several unique morphologies (e.g. bacilli-, bottle- or droplet shapes) have been reported for archaeal viruses, but the last novel morphotype for bacteriophages was announced over 40 years

ago. In the literature, phages displaying icosahedral symmetry with a tail-structure predominate, constituting about 96 % of the isolates described (Ackermann and Prangishvili 2012). All the known tailed phages contain dsDNA genome and lack any lipid moiety. These tailed dsDNA phages constitute the order *Caudovirales*, which has been subdivided into three families (*Myoviridae*, *Siphoviridae*, *Podoviridae*) based on the tail morphologies. The remaining 3-4 % of the studied phages are either icosahedral, filamentous or pleomorphic. They lack tail-structure but may contain a structural lipid component.

2.6.1 Membrane-containing bacteriophages

The first lipid-containing phage, PM2 infecting *Pseudoalteromonas espejiana*, was isolated in the late 1960s (Espejo and Canelo 1968), several decades after the existence of phages was recognised. This delay may be partly explained by the common practise of using chloroform in virus stock solutions to prevent bacterial contaminations (Adams 1959, Atanasova *et al.* 2015).
Sensitivity to chloroform, or other organic solvents, is often the first indication of a lipid structure in the virus (Porter *et al.* 2005, Pietilä *et al.* 2009, Jaakkola *et al.* 2012). However, the chloroform treatment is a rather robust way to assay the presence of lipids, since it can also reduce the infectivity of viruses devoid of lipid components (Dyall-Smith 2011), and therefore additional tests are required. Low buoyant density of the virion, compared to that of phage particles with only protein and nucleic acid components (~1.3 g ml^{-1} and ~1.5 g ml^{-1} in cesium chloride, respectively), is another preliminary indicator of the viral lipid component (Poranen *et al.* 2015). Highly-purified virions, subjected to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), are also routinely stained with lipophilic dye Sudan Black B to detect the possible lipid moieties (Pietilä *et al.* 2012, Jaakkola *et al.* 2012). Further studies to confirm the presence of the lipid component and to determine its composition may be performed by techniques such as thin layer chromatography, mass spectrometry (e.g. electrospray ionization) and nuclear magnetic resonance, using lipid extracts (Roine and Bamford 2012, Atanasova *et al.* 2015). These techniques require highly-purified viral material, which might be challenging to obtain, if the host bacterial cells produce membrane vesicles (Roine and Bamford 2012, Atanasova *et al.* 2015).

Due to the lack of inherent lipid biosynthesis machinery, viruses acquire lipids from their hosts either selectively or non-selectively. Therefore the lipid composition of phages reflects that of their host bacteria. In some viruses, host enzymes apparently catalyse the covalent attachment of a lipid moiety to viral structural proteins (Hruby and Franke 1993). The lipid modifications are suggested to facilitate protein folding and protein-protein interactions, consequently mediating the virion assembly. However, in bacteriophages lipids seem to exist predominantly as structural bilayers (Atanasova *et al.* 2015). Enveloped phages usually acquire their external membrane layer as the newly synthesized nucleocapsids are released from their host by budding through the

cellular membrane (Garoff *et al.* 1998). Budding involves the envelopment of the viral core by host cytoplasmic membrane and the subsequent membrane fission, which releases the enveloped virions into the extracellular space. This is a delicate way of releasing virions without disrupting the host cell in the process. Another mechanism for envelopment has been suggested for phages with segmented dsRNA genomes. Studies of bacteriophage φ6 indicate, that this phage type may acquire the envelope from preformed lipid membrane vesicles which enclose the nucleocapsid by a yet unknown mechanism (Poranen *et al*. 2005) (see section 2.9.3). Similarly, host membrane-derived vesicles seem to be a prerequisite for the formation of internal-membrane containing tectiviruses (Rydman *et al*. 2001) (see section 2.7.3).

In addition to assisting virion extrusion, lipid membranes also play a vital role in the viral entry. Whereas membrane fission is required for the dissemination of enveloped viral progeny, membrane fusion mediates their entry into the host cell. Upon entry, φ6, and presumably other enveloped dsRNA phages, fuse the viral membrane with the outer membrane of their gram-negative host bacterium (Poranen and Bamford 2008). On the other hand, the membrane fusion is assumed to occur with the host cytoplasmic membrane in the case of phages infecting mycoplasmas, prokaryotes lacking the cell wall (Putzrath and Maniloff 1977). Interestingly, the internal membrane has also been shown to facilitate the viral entry. PRD1 and its gram-positive bacteria infecting relative Bam35 use their internal membranes to form a tubular conduit for the DNA injection into the host (Bamford and Mindich 1982, Gaidelytė *et al.* 2006, Peralta *et al.* 2013) (see section 2.7.3). The same entry mechanism has been proposed for other internal membrane-containing viruses (Peralta *et al.* 2013), but it does not seem to be universal since at least bacteriophage PM2 is supposed to enter the host by fusing its internal membrane with the outer membrane of the gram-negative bacterium (Kivelä *et al.* 2004, Cvirkaitė-Krupovič *et al.* 2010a).

The presence of a membrane in the virion evidently facilitates the virus-host interactions and reflects the ability of viruses to adapt to different hosts (Atanasova *et al*. 2015). Interestingly, lipid membrane is not necessarily found in all structurally similar viruses such as in case of internal membrane-containing bacteriophage PRD1 and membrane-lacking adenovirus (Stewart *et al*. 1993, Abrescia *et al*. 2004, Cockburn *et al*. 2004). It has been suggested that adenoviruses have lost the membrane in order to build larger capsids and, consequently, increase their gene content (Krupovič and Bamford 2008b). These acquired features facilitate their propagation in complex eukaryotic cells. Members of the bacteriophage families *Tectiviridae, Corticoviridae* and *Cystoviridae* are rather widely studied and therefore serve as valid models of internal membrane-containing and enveloped viruses.

2.7 Bacteriophage PRD1 – a member of the *Tectiviridae* family

2.7.1 *Tectiviridae* – a family of icosahedral, internal membrane–containing dsDNA phages

Family *Tectiviridae* consists of tailless, icosahedrally symmetric bacteriophages having a dsDNA genome of circa 15 kb (Grahn *et al.* 2006). The most distinctive structural feature of these phages is the proteinaceous internal membrane surrounding the viral genome. Tectiviruses can be divided into two subgroups based on their host bacteria, which can be either gram-positive or gram-negative. Not only have the two tectivirus subgroups retained the characteristic morphological features, but also their genome sizes and genetic organizations are remarkably similar, even though the resemblance is not detectable at the sequence level (Ravantti *et al.* 2003, Saren *et al.* 2005, Oksanen and Bamford 2012a).

Bacteriophage PRD1 (Olsen *et al.* 1974) and its close relatives PR3 (Bamford *et al.* 1981), PR4 (Stanisich 1974), PR5 (Bamford *et al.* 1981), L17 (Bamford *et al.* 1981) and PR722 (Coetzee and Bekker 1979) infect a variety of gram-negative enterobacteria, including e.g. *E. coli*, *S. typhimurium* and *Pseudomonas aeruginosa*. All the bacterial host strains harbour a conjugative multidrug-resistance plasmid of incompatibility groups P, N or W. These plasmids encode a conjugal transfer complex, which is recognized by a tectivirus as a receptor. Despite the fact that tectiviruses infecting gram-negative hosts have been isolated at globally distant locations over the course of several years, they are genetically practically identical to each other (nucleotide sequence identity ranging from 91.9 to 99.8 %), the most sequence variation being in the genes encoding proteins required for host recognition (Saren *et al.* 2005).

Unlike their more promiscuous counterparts infecting gram-negative enterobacteria, tectiviruses using gram-positive hosts tend to be highly host specific and prey only on a few *Bacillus cereus* species. These virus species include *B. anthracis* phages AP50 (Nagy *et al.* 1976) and Wip1 (Schuch *et al.* 2010), *B. thuringiensis* infecting phages Bam35 (Ackermann *et al.* 1978, Ravantti *et al.* 2003), GIL01 (Verheust *et al.* 2003) and GIL16 (Verheust *et al.* 2005), as well as φNS11, which infects *B. acidocaldarius* (Sakaki *et al.* 1977). The latter, however, is not available anymore. Also, the whole-genome sequencing of *B. cereus* reference strain ATCC 14579 revealed the presence of pBClin15 (Ivanova *et al.* 2003), a linear plasmid, which shares the genomic architecture with gram-positive bacteria-infecting tectiviruses (Verheust *et al.* 2005) and is suggested to be a defective prophage. The genetic sequences of Bam35 and GIL01 differ only by a few nucleotides (Ravantti *et al.* 2003, Verheust *et al.* 2003, Saren *et al.* 2005), and therefore it may be argued whether they can be regarded as different species. Otherwise, the tectiviruses with gram-positive hosts seem to display more genetic variance compared to those infecting gram-negative bacteria. This

may reflect their different lifestyles: whereas all tectiviruses infecting gram-negative bacteria are virulent and lyse the host cell at the end of the infection cycle, those using gram-positive hosts are temperate and can exist as autonomously replicating linear prophages within the host cell (Strömsten *et al.* 2003b). Therefore it would seem probable that the ones infecting gram-positive bacteria have been exposed to horizontal gene transfer more frequently, which could explain their more extensive genetic mosaicism. Alternatively, the greater genetic variance may also be explained by more ancient origin of tectiviruses using gram-positive bacteria as hosts.

2.7.2 Structure of Enterobacteria phage PRD1

Bacteriophage PRD1 was isolated over 40 years ago from a sewage sample, which was taken from Kalamazoo, Michigan (Olsen *et al.* 1974). The structural and functional aspects of PRD1 have been studied in great detail. For instance, its atomic-level structure has been determined at ~4 Å resolution (Abrescia *et al.* 2004, Cockburn *et al.* 2004). PRD1 is arguably the best-described viral system with an internal lipid membrane.

The genetic material of PRD1 is in the form of a linear, 14 927 bp-long dsDNA molecule (Bamford *et al.* 1991, Saren *et al.* 2005). The genome contains 110-bp inverted repeats at the ends (Savilahti and Bamford 1986), as well as a covalently joined protein P8 at both of the 5´ termini, priming the genome replication and facilitating genome packaging (Bamford *et al.* 1983, Savilahti *et al.* 1991, Caldentey *et al.* 1993, Saren *et al.* 2005, Ziedaite *et al.* 2009). The genome is organized into modules, with functionally related genes clustered together and regulated by common promoters (Grahn *et al.* 1994). The internal membrane vesicle, enclosing the viral genome, follows closely the icosahedral shape of the outmost protein capsid (Butcher *et al.* 1995, Cockburn *et al.* 2004). The host-derived lipids constitute about half of the membrane mass, whereas the other half consists of virally encoded membrane proteins (Oksanen and Bamford 2012a). In fact, approximately half of the PRD1 structural protein species are associated with the membrane. The lipid composition of the viral membrane differs somewhat from that of the host cell, implying a selective lipid acquirement upon assembly (Laurinavičius *et al.* 2004). Most notably, the ratio between phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) is higher in the viral membrane. Moreover, the phospholipids are asymmetrically distributed in the viral lipid bilayer, PG and PE being enriched in the outer and inner leaflets, respectively (Laurinavičius *et al.* 2004, Cockburn *et al.* 2004). The accumulation of zwitterionic PE in the inner leaflet probably mediates the association with the viral genome. The asymmetric distribution has been proposed to reflect the different shapes and charges of phospholipids as well as their interactions with membrane proteins (Laurinavičius *et al.* 2004, Cockburn *et al.* 2004).

The external protein capsid of the mature PRD1 virion (~66 MDa) is 740 Å in diameter, measured from vertex to vertex (Butcher *et al.* 1995, San Martin *et al.* 2001). Two hundred forty copies of the trimeric MCP P3 are arranged into

pseudo *T* = 25 icosahedral lattice (Butcher *et al.* 1995). The MCPs adopt a double β-barrel fold, giving the trimers a pseudohexagonal shape and allowing them to pack closely on the shell surface (Benson *et al.* 1999, 2002). Sixty copies of the elongated tape-measure protein P30 cement the facets together, simultaneously defining the size of the capsid (Abrescia *et al.* 2004).

The fivefold vertices of the virion are occupied by a spike structure. Five P31 proteins, organized as a penton, form the base from which trimeric spike protein P5 and host recognition protein P2 protrude (Grahn *et al.* 1999, Rydman *et al.* 1999, Bamford and Bamford 2000, Huiskonen *et al.* 2007). P31 interacts with P3 MCPs and integral membrane protein P16, which resides under the vertex (Abrescia *et al.* 2004, Jaatinen *et al.* 2004). One of the 12 vertices contains a unique composition of proteins. Instead of the common spike structure, the vertex is occupied by proteins required for DNA packaging. These proteins include integral membrane proteins P20 and P22, minor capsid protein P6 and packaging ATPase P9 (Gowen *et al.* 2003, Strömsten *et al.* 2003a).

2.7.3 Life cycle of PRD1

The PRD1 infection cycle is initiated as the virion attaches onto the host surface through an interaction between its vertex protein P2 and the cellular receptor complex (Mindich *et al.* 1982a, Grahn *et al.* 1999) (Fig. 2A). Binding to the host cell elicits structural changes, leading to the dissociation of the spike complex and the surrounding peripentonal P3 MCPs from the virion (Peralta *et al.* 2013) (Fig. 2B). This creates a large opening at the vertex, which destabilizes the association between the capsid and the membrane and triggers the transformation of the internal membrane into a tubular structure. The proteolipidic tube traverses the bacterial cell envelope, supposedly through the unique vertex, which is also used for genome packaging (Strömsten *et al.* 2003a, Peralta *et al.* 2013, Hong *et al.* 2014). Several membrane proteins (P7, P14, P18, P32) are known to be involved in the membrane transformation (Bamford and Mindich 1982, Grahn *et al.* 2002), and presumably form a scaffold for the tube formation (Peralta *et al.* 2013). Additionally, the major membrane protein P11 is supposed to be involved in the penetration of the outer membrane (Grahn *et al.* 2002), making the peptidoglycan layer more accessible for the membrane-associated enzyme P7, which is a peptidoglycan-degrading transglycosylase (Rydman and Bamford 2000, 2002). As the tip of the tube enters host cell cytoplasm, the phage DNA is released. The DNA injection is probably powered by the energy stored in the pressurized protein capsid or alternatively by changes in the osmotic pressure (Cockburn *et al.* 2004, Peralta *et al.* 2013).

After the viral genome enters the cytoplasm, transcription and subsequent translation commence (Fig. 2C and D). Both of the early operons in the PRD1 genome are transcribed first, leading to the synthesis of terminal proteins, DNA polymerase as well as two ssDNA-binding proteins, both of which function in replication (Grahn *et al.* 2006). Later, the transcription of the three late operons results in production of virion structural proteins and morphogenetic factors.

The PRD1 genome is replicated via a protein-primed sliding back mechanism, utilizing virally encoded polymerase (Savilahti *et al.* 1991, Caldentey *et al.* 1993). Similar replication mechanism has been described for adenovirus and phage φ29 [reviewed by (Calendar 2006)]. The DNA polymerase possesses 3´-5´ exonuclease activity, probably accounting for the proofreading ability. The two early produced virally encoded proteins, P12 and P19, attach to ssDNA, protecting it from nuclease attacks. These two proteins are known to stimulate PRD1 replication *in vivo* (Pakula *et al.* 1990, 1993).

About 15 minutes post infection (p.i), the first capsid precursors appear in the cytoplasm (Mindich *et al.* 1982a) (Fig. 2E). These include MCP P3 trimers, P5 trimers, P31 pentamers and P2 monomers, reflecting their oligomeric state in the final capsid structure. Simultaneously, the phage-encoded membrane proteins, such as P7, P11, P14 and P18, are recruited into the host cytoplasmic membrane (Fig. 2F). Correct folding of the soluble capsid proteins P3 and P5, as well as the assembly of several membrane proteins, are known to depend on the function of the host GroEL/GroES chaperonin complex (Hänninen *et al.* 1997). A phage-specific patch, enriched in phage-encoded protein P10, is presumably pinched off from the cytoplasmic membrane, mimicking the eukaryotic clathrin-mediated endosytosis (Rydman *et al.* 2001) (Fig. 2G). The resulting vesicle is suggested to act as a scaffold as the assembly proceeds and capsid proteins are layered onto the vesicle, ultimately displacing P10 (Fig. 2H). In addition to P10, another phage-encoded non-structural protein, P17, is required for the proper particle formation (Mindich *et al.* 1982a), even though its exact function in the assembly is not known. The resulting empty procapsid, devoid of viral genome, consists of the internal membrane and all the structural proteins of the virion except for the packaging ATPase and the genome terminal protein (Mindich *et al.* 1982a, Strömsten *et al.* 2003a). Finally the packaging ATPase P9 powers the translocation of the linear dsDNA into the procapsid through the unique vertex (Gowen *et al.* 2003, Strömsten *et al.* 2005, Karhu *et al.* 2007, Ziedaite *et al.* 2009, Hong *et al.* 2014) (Fig. 2I). Unlike other viral packaging ATPases, P9 remains associated with the mature virion after the packaging is completed. Three other proteins of the unique vertex, the packaging efficiency factor P6 and the small integral membrane proteins P20 and P22, form a conduit through which the viral genome is translocated into the internal membrane vesicle of the procapsid (Hong *et al.* 2014). As a result of DNA packaging, the internal membrane vesicle expands slightly, which increases its interactions with the surrounding capsid (Butcher *et al.* 1995). Otherwise packaging does not induce any major structural changes in the capsid, in contrast to many other dsDNA phages.

In order to release the newly synthesized virions into the environment, PRD1 utilizes phage-encoded holin – endolysin system (Rydman and Bamford 2003). Virion-associated muramidase P15 (Mindich *et al.* 1982b, Caldentey *et al.* 1994) and holin P35 (Rydman and Bamford 2003) degrade the host cytoplasmic membrane and the peptidoglycan layer in timely fashion (Mindich *et al.* 1982b, Caldentey *et al.* 1994, Rydman and Bamford 2003, Ziedaite *et al.* 2005) (Fig. 2J).

PRD1 also encodes two additional lysis proteins P36 and P37, which are required for lysis under less favorable conditions (Krupovič *et al.* 2008). Proposedly, these two proteins transform the mechanical stress of holin lesions from the cytoplasmic membrane to the outer membrane. This leads to the disruption of the outer membrane and the consequent liberation of phage progeny (Krupovič *et al.* 2008).

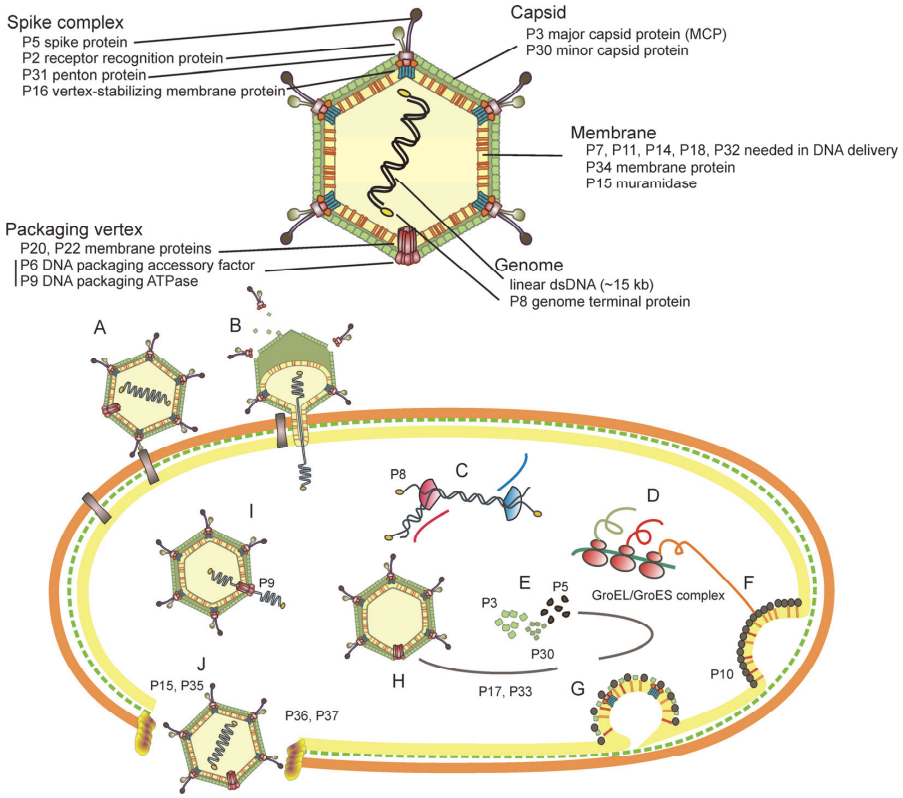


FIGURE 2 Schematic representation of the life cycle of PRD1. A) PRD1 binds to the host cell through an interaction between the spike complex and the receptor complex. B) Consequently, the spike complex and the surrounding MCPs dissociate from the virion. Internal membrane transforms into a tubular structure (involving membrane proteins P7, P14, P18 and P32), which traverses the cell outer membrane (mediated by membrane-associated protein P11). Transglycosylase P7 degrades the peptidoglycan layer, enabling the genome entry into the host cell cytoplasm. C and D) After DNA entry, transcription, translation and protein-primed replication commence. E) Capsid and spike proteins appear in the cytosol, while F) membrane proteins are recruited to the cell membrane. The correct folding/assembly of these proteins requires host GroEL/GroES chaperonin complex. G) A membrane vesicle, enriched in phage-encoded protein P10, is pinched off from the membrane. H) Subsequently, capsid proteins displace P10, forming a procapsid. Phage-encoded non-structural protein P17, and possibly P33, are needed in the formation of empty procapsids that contain an internal membrane. I) Packaging ATPase P9 powers the DNA translocation inside the procapsid, resulting in mature virion. J) Endolysin P15 and holin P35 degrade the cytoplasmic membrane and the peptidoglycan layer, leading to the cell lysis and viral progeny release. Under less favourable conditions, the lysis is ensured by accessory proteins P36 and P37. Modified from Poranen *et al*. 2015.

2.8 PM2 – the sole member of *Corticoviridae*

Virus family typified by internal membrane-containing bacteriophages with a circular dsDNA genome is the *Corticoviridae* (Oksanen and Bamford 2012b). The only representative of this family is phage PM2, isolated in Chile from a coastal seawater sample in 1968 (Espejo and Canelo 1968). PM2 has been shown to infect two gram-negative *Pseudoalteromonas* strains: the isolation host *Pseudoalteromonas espejiana* BAL-31 (Espejo and Canelo 1968) and *Pseudoalteromonas* sp. ER72M2, which was retrieved from the East River in New York (Kivelä *et al.* 1999). Despite the lack of other characterized corticoviruses, PM2-related prophages seem to reside commonly in aquatic bacterial genomes (Krupovič and Bamford 2007).

The structure of PM2 was determined at 7 Å resolution, revealing a tailless virion in which an internal, icosahedrally shaped membrane is covered by a protein shell (Abrescia *et al.* 2008). MCP P2 is organized on a pseudo $T = 21$ dextro lattice, with protein P1 forming pentameric spikes at the fivefold vertices (Huiskonen *et al.* 2004). The double β-barrel fold of the MCP links PM2 to the PRD1-adenovirus lineage (see section 2.5.2). The membrane vesicle, or lipid core, contains eight phage-encoded protein species as well as host-derived lipids, selectively acquired from the cytoplasmic membrane (Kivelä *et al.* 2002, Laurinavičius *et al.* 2007). The most abundant membrane proteins P3 and P6 are assembled into a planar lattice, lying between the membrane and the protein shell (Abrescia *et al*. 2008). The membrane vesicle covers the circular dsDNA genome of ~10 kb, which is in a highly supercoiled form (Männistö *et al.* 1999). Twenty-one ORFs are organized into three operons, the timely expression of which is under the regulation of phage-encoded transcription factors (Männistö *et al.* 2003).

PM2 life cycle is lytic. It recognises the susceptible cell via the outermost tip of the pentameric receptor recognition protein P1 (Kivelä *et al.* 2004, Abrescia *et al.* 2008). Despite efforts, the specific receptor on the bacterium cell has not been identified. However, the structural resemblance of the outermost tip of P1 to the calcium-dependent, carbohydrate-binding protein of *Saccharophagus degradans* implies a lipopolysaccharide-mediated entry (Abrescia *et al.* 2008, Kivelä *et al.* 2008). After binding, the protein capsid of PM2 dissociates, proposedly leading to the fusion between the lipid core and the host outer membrane (Kivelä *et al*. 2004).

The PM2 genome replicates via rolling-circle mechanism at the host cytoplasmic membrane (Männistö *et al.* 1999). Also, the assembly of the virion is assumed to occur in association with the cytoplasmic membrane, although the details of the process remain unclear. It has been suggested that the plate-like structure of the membrane proteins P3 and P6 acts as a template for the correct registration of the virus facets (Abrescia *et al*. 2008). P6 is assumed to interact with the supercoiled phage DNA, ultimately leading to the pinching off of genome-containing membrane vesicles (Abrescia *et al*. 2008). The putative

ATPase of PM2 is a structural component of the virion found in the viral membrane (Kivelä *et al*. 2002) and homologous to the packaging ATPases of other internal-membrane containing viruses (Strömsten *et al.* 2005). The cell lysis commences as phage-encoded protein P17 punctuates the cytoplasmic membrane, enabling another phage protein P18 as well as cellular lytic factor (CLF) to be released into the periplasm (Krupovič *et al.* 2007). CLF degrades the peptidoglycan layer, and P18 helps to disintegrate the outer membrane. Ultimately ~300 virus particles are released from an infected *Pseudoalteromonas* cell (Kivelä *et al*. 1999).

2.9 Bacteriophage φ6 and other *Cystoviridae* family members

2.9.1 *Cystoviridae* – a family of enveloped phages with segmented dsRNA genome and icosahedral core

Bacteriophage φ6 was isolated from a *Pseudomonas*-infected bean straw in Nebraska at the beginning of 1970s (Vidaver *et al.* 1973). It remained the only representative of the *Cystoviridae* family for over two decades until additional putative cystoviruses were isolated from agricultural plant samples (Mindich *et al.* 1999, Qiao *et al.* 2010). Currently, the family *Cystoviridae* is considered to encompass 10 bacteriophage species (Mindich *et al*. 1999, Qiao *et al.* 2010), although only φ6 is officially recognised by the ICTV (King *et al*. 2012). Each cystovirus is characterized by a tripartite dsRNA genome of ~13 kb and a complex capsid consisting of two concentric, icosahedrally symmetric protein shells (except for φ8, which contains only the innermost protein shell) and a protein-rich envelope. The segmented dsRNA genome and the envelope make up a unique combination among bacteriophages. All the identified cystovirus species have been isolated from leguminous plants in the U.S.A (Mindich *et al.* 1999, Qiao *et al.* 2010). The phages are specific to gram-negative bacteria, primary to plant pathogenic *Pseudomonas syringae* strains.

The widely characterized φ6 is the type species of the *Cystoviridae* (King *et al*. 2012), while other cystovirus isolates are subdivided into two groups based on the level of sequence resemblance to φ6. The close relatives of φ6 include φ7, φ9, φ10, φ11 and φ14, whereas φ8, φ12, φ13 and φ2954 are genetically more distantly related to φ6 (Mindich *et al.* 1999, Qiao *et al.* 2010). Despite the sequence variance, all cystoviral genomes share a similar gene order within each genome segment (McGraw *et al.* 1986, Gottlieb *et al.* 1988, 2002a,b, Mindich *et al.* 1988, Hoogstraten *et al.* 2000, Qiao *et al.* 2000, 2010).

The genetic distance between the two cystovirus subgroups correlates with different host specificities. Whereas φ6 attaches along type IV pilus for entry, its more distant relatives recognise a truncated lipopolysaccharide O chain of the host outer membrane [rough lipopolysaccharide (LPS)] (Mindich *et al*. 1999). An exception is φ2954, which uses type IV pilus-mediated infection mechanism (Qiao *et al*. 2010). In each case, the initial host binding is mediated

by a host attachment complex, composed of protein P3 (Bamford *et al*. 1976, Gottlieb *et al*. 1988, 2002b, Hoogstraten *et al*. 2000, Qiao *et al*. 2000, 2010). However, among identified cystoviruses, the specific association with the pilus is mediated by a homomeric complex, composed of a single P3 polypeptide (Gottlieb *et al.* 1988, Qiao *et al.* 2010), whereas the attachment to the rough LPS seems to require a heteromeric complex of two or three P3 polypeptides (Hoogstraten *et al*. 2000, Qiao *et al*. 2000, Gottlieb *et al*. 2002b). Consequently, cystoviruses φ8, φ12 and φ13 infect a broader range of hosts by recognising rough LPS of gram-negative bacteria such as *E. coli* and *S. typhimurium* (Mindich *et al.* 1999).

Cystoviruses evolve at extremely fast pace. The error-prone polymerase exposes cystoviruses to unusually high rates of spontaneous mutations (10^{-3} – 10^{-5} errors per nucleotide replication) (Chao 1988). In addition, cystoviruses experience genetic exchange, when two or more viruses co-infect the same host cell, generating hybrid progeny (Mindich 2004). The exchange may occur, when the genetic information of two template strands recombines into one daughter strand. Among cystovirus genotypes, recombination is accomplished by template switching between related (homologous recombination) or unrelated sequences (illegitimate recombination). Illegitimate recombination has been shown to require only a few identical nucleotides (usually three, but ranging from zero to 12) at the crossover site (Qiao *et al.* 1997). This recombination type seems to occur infrequently among cystovirus populations, but it can be artificially provoked by preventing minus-strand synthesis (Mindich 1996). Interestingly, template switching may also incorporate foreign genetic material into cystoviral genomes, as transcripts originating from host bacteria or other virus systems are occasionally packaged into procapsid (Onodera *et al.* 2001). Homologous recombination has been reported for φ8 (Onodera *et al.* 2001), the most distant relative of φ6. However, the packaging of cystoviruses is rigorously controlled (Mindich 1999), and therefore it is unlikely that two molecules of the same segment type would be incorporated into the same procapsid. Consequently, the homologous recombination is unusual both *in vitro* (Onodera *et al.* 2001) and in nature (Silander *et al.* 2005, O'Keefe *et al.* 2010).

Furthermore, alternatively to recombination, cystoviruses may undergo genome segment reassortment (Turner *et al.* 1999, Mindich 2004), which brings additional genetic variance to phage populations. Reassortment takes place regularly, both in laboratory and in natural settings (Onodera *et al*. 2001, Silander *et al*. 2005, O'Keefe *et al.* 2010). In fact, it has been estimated that the frequency of segment swapping equals, or even exceeds, that of spontaneous nucleotide substitutions (Silander *et al.* 2005). Surprisingly, the rate of reassortment seems to depend on the geographical location (O'Keefe *et al.* 2010). This variance could be logically explained by differing population sizes. Certain environments may also turn out to be less favourable for reassortment hybrids, leading to their rapid elimination (O'Keefe *et al.* 2010).

2.9.2 Structure of *Pseudomonas* phage φ6

The spherical virion of *Pseudomonas* phage φ6 has a rather complex architecture, with three concentric structural layers enclosing the segmented genome (Fig. 3). The three dsRNA genome segments have been named according to their size as S (small, 2948 bp), M (medium, 4063 bp) and L (large, 6374 bp) (McGraw *et al*. 1996, Gottlieb *et al*. 1988, Mindich *et al*. 1988). Lower case letters *s*, *m* and *l* are commonly used when referring to positive-sense ssRNA precursors. In each genome segment, the genes are arranged into functional groups (McGraw *et al*. 1996, Gottlieb *et al*. 1988, Mindich *et al*. 1988). The L-segment codes for proteins of the virion core (P1, P2, P4 and P7), the M-segment contains genes for membrane-associated proteins (P3, P6, P10, P13), and the S-segment encodes the protein forming the outermost capsid shell (P8), the major membrane protein (P9) and two proteins responsible for host lysis (P5 and P11) (Poranen *et al*. 2005). In addition, φ6 produces two non-structural proteins (P12 and P14), which are encoded by the S- and L-segment, respectively. The coding regions of the φ6 genome segments are flanked by noncoding sequences, which are required for packaging and replication (McGraw *et al*. 1986, Gottlieb *et al*. 1988, Mindich *et al*. 1988).

The innermost protein shell, also called as the polymerase complex (PC), is composed of four proteins: MCP P1, RNA dependent RNA polymerase (RdRP) P2, packaging NTPase P4 and minor protein P7 (Mindich and Davidoff-Abelson 1980, Bamford and Mindich 1980) (Fig. 3). Protein P1 is the main component of the PC (Ktistakis and Lang 1987, Olkkonen and Bamford 1987) and 60 asymmetric dimers of P1 organize into the dodecahedral skeleton of the virion (T = 1 architecture, contradicting the theory of Caspar and Klug; see section 2.2.1) (Butcher *et al.* 1997). Even though φ6 displays the characteristic, non-equivalent capsid geometry of dsRNA viruses, its major structural protein P1 seems to possess a unique fold, which does not bear any significant resemblance with the capsid proteins of other dsRNA viruses (Nemecek *et al.* 2013b). This raises the question of the origin of φ6, in relation to other dsRNA viruses. P1 exists in two conformation types, which form the dodecahedral lattice in ratio 1:1. Five $P1_A$ monomers occupy the fivefold axes, while $P1_B$ monomers interact at the two- and threefold axes (Huiskonen *et al.* 2006).

The dodecahedral lattice serves as a framework onto which the other components of the PC attach. NTPase P4 organizes into ring-like hexamers, which bind to the outer surface of the PC. The P4 hexamers overlie the fivefold vertices, creating a symmetry mismatch (de Haas *et al.* 1999, Huiskonen *et al.* 2006). Monomers of the polymerase P2, on the other hand, have been shown to reside at the inner surface of the PC, at sites close to the threefold axes. The 20 potential binding sites of polymerase P2 (Sen *et al.* 2008) are apparently randomly occupied, with an average of eight subunits per PC (Nemecek *et al.* 2010). It has been proposed, that during virion maturation, P2 polymerase molecules rotate to a position beneath the fivefold symmetry axis, where they catalyse replication and transcription (Sen *et al.* 2008). The fourth structural

protein of the PC is P7, which functions as a packaging cofactor (Juuti and Bamford 1995, 1997), and also contributes to virion assembly (Poranen *et al.* 2001). P7 appears as elongated dimers in solution (Juuti and Bamford 1997, Poranen *et al.* 2001), but structural studies imply that it attaches as monomers onto the interior surface of the PC, near the threefold symmetry axes (Katz *et al.* 2012, Nemecek *et al.* 2012). Due to the overlapping locations in the PC, it has been postulated that P2 and P7 compete for the same binding site upon assembly (Nemecek *et al.* 2012). Alternatively, P7 may help RdRP P2 to maintain its position prior to packaging (Katz *et al.* 2012).

The outermost shell (nucleocapsid shell, NC shell) is composed of 200 trimers of protein P8 arranged into $T = 13$ architecture (Butcher *et al.* 1997, Huiskonen *et al.* 2006) (Fig. 3). The integrity of the protein layer is disrupted at the fivefold vertices, where turret-shaped P4 hexamers protrude (Butcher *et al.* 1997, Huiskonen *et al.* 2006). The two concentric protein shells form the NC (Van Etten *et al.* 1976, Hantula and Bamford 1988). Interestingly, bacteriophage φ8 lacks the outer NC shell (Hoogstraten *et al.* 2000, Jäälinoja *et al.* 2007).

The nucleocapsid is surrounded by an envelope, consisting of host cell membrane-derived phospholipids (Laurinavičius *et al.* 2007) and four virally encoded membrane proteins P6, P9, P10, and P13 (Sinclair *et al.* 1975, Van Etten *et al.* 1976, Gottlieb *et al.* 1988) (Fig. 3). The host attachment spikes, formed by protein P3, are anchored to the viral membrane via integral membrane protein P6 (Poranen *et al.* 2005) (Fig. 3). In addition to P3, φ6 virion possesses another membrane-associated protein, lytic enzyme P5, which is loosely attached to the outer surface of the NC, beneath the viral envelope (Hantula and Bamford 1988).

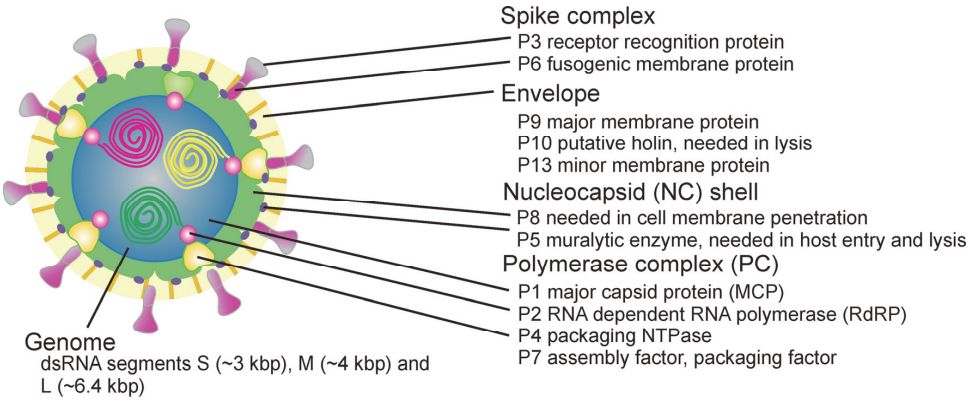


FIGURE 3 Schematic representation of the φ6 virion. Three genomic dsRNA segments (S, M and L) are surrounded by a polymerase complex, composed of four proteins (P1, P2, P4 and P7). The protein core is covered by the nucleocapsid protein P8. The outermost envelope consists of host-derived lipids, as well as phage-encoded proteins (integral membrane proteins P6, P9, P10 and P13 as well as the P3 spikes). Modified from Poranen *et al.* 2005.

2.9.3 Life cycle of φ6

The layered structure of phage φ6 is sequentially dissociated as the virion passes through the bacterial cell envelope. The initial attachment is mediated with the interaction between the spike protein P3 and type IV pilus of the host (Bamford *et al*. 1976). Upon virus attachment, the pilus retracts, bringing the virus particle into contact with the bacterial outer membrane (Bamford *et al.* 1976, Romantschuk and Bamford 1985). The spike protein P3 detaches from the virion, exposing membrane-bound protein P6. Fusogenic activity of P6 subsequently mediates the fusion between the viral envelope and the bacterial outer membrane, allowing the NC to enter the periplasmic space (Bamford *et al.* 1987). The removal of the viral envelope releases the NC-associated endopeptidase P5, which digests the peptidoglycan layer, enabling the NC to reach the host cell membrane (Mindich and Lehman 1979, Caldentey and Bamford 1992). Interaction with the NC shell triggers cell membrane invagination, delivering the NC into the host cytosol within an intracellular vesicle (Romantschuk *et al.* 1988, Poranen *et al.* 1999). It has been suggested that lowered pH or chelating calcium ions may trigger the dissociation of the NC shell, eventually exposing the PC to cytosol (Olkkonen *et al.* 1991, Cvirkaitė-Krupovič *et al.* 2010b). The release of the NC shell activates RdRP P2, initiating transcription inside the PC (Poranen *et al*. 2005). It has been demonstrated that the transcription of the L-segment is activated by the attachment of host protein YajQ to the surface of the PC (Qiao *et al.* 2008). Consequently, during this first round of transcription, full-length, polycistronic mRNAs are transcribed in equal amounts from each genome segment via a semi-conservative mechanism (Emori *et al.* 1983). Resulting positive-sense ssRNAs are either packaged as genome precursors into newly synthesized virions or used as templates for protein synthesis. However, only the *l*-transcripts are translated efficiently, leading to the accumulation of the PC protein components (Poranen *et al*. 2005). These proteins assemble into empty dodecahedral cores (Bamford and Mindich 1980), also called as procapsids (Poranen *et al*. 2005). NTPase P4 translocates the three genome segment transcripts into the procapsid in order: *s*, *m* and *l* (Qiao *et al*. 1995, Frilander and Bamford 1995, Mindich 2004). Packaging depends on the recognition of specific packaging signals (*pac* sequences, ~200 nt in length) at the 5´ end of the transcripts (Gottlieb *et al.* 1994). The packaging of each of the genome segments has been suggested to induce a conformational change in the procapsid, creating a recognition site for the next segment to be translocated (Mindich 2004). At the end of the packaging, the procapsid has transformed into a spherical, dsRNA containing particle, with a ~2.4-fold increase in the volume (Huiskonen *et al*. 2006). As the *pac* sequence of the *l*-transcript enters the procapsid, it acts as an initiation signal for the replication (Poranen and Bamford 1999). Each packaged transcript serves as a template for the minus-strand synthesis, resulting in mature, double-stranded forms of the genome segments (Poranen *et al*. 2005). Replication of the 5´ end of the L-segment triggers the switch back to plus-strand synthesis (Poranen and Bamford 1999).

During this second round of transcription, more S- and M-segments are transcribed, compared to the L-segment (Coplin *et al.* 1975), leading to the production of proteins needed in the virion assembly and maturation. Newly synthesized P8 proteins accumulate around the procapsid, forming a nucleocapsid and simultaneously switching off the transcription (Olkkonen *et al.* 1990, 1991, Poranen *et al.* 2001).

At the last step of bacteriophage φ6 maturation, viral membrane assembles around the nucleocapsid. The morphogenesis of the envelope is not known in detail, but the major membrane protein P9 and non-structural protein P12 are needed in the process (Johnson and Mindich 1994). These two virally encoded proteins have been proposed to induce the vesiculation of the host cell membrane and the subsequent association between the vesicle and the NC (Johnson and Mindich 1994). After the envelopment, the P3 spikes are attached to the virion via P6, resulting in the mature virus particle (Poranen *et al*. 2005). Ultimately, lytic enzyme P5 and membrane protein P10 facilitate the rupture of the host cell, releasing progeny viral particles into the environment (Bamford *et al.* 1976, Mindich and Lehman 1979, Caldentey and Bamford 1992, Johnson and Mindich 1994).

2.10 Icosahedral bacteriophages with ssDNA genome

ssDNA viruses have been known since the end of the 1950s, when it was demonstrated that icosahedrally symmetric bacteriophage φX174 possesses a circular ssDNA genome (Fane *et al*. 2006). Ever since, φX174 has been used as a model system and studied extensively structurally, biochemically and genetically (Fane *et al*. 2006). In fact, the first DNA genome ever to be completely sequenced was that of φX174 (Sanger *et al.* 1977). Despite these historical aspects, ssDNA viruses have been largely overshadowed by their double-stranded counterparts. The small genome size of ssDNA viruses (King *et al*. 2012) has severely complicated their examination (Tucker *et al*. 2011). Traditional methods for viral counting, such as epifluorescence microscopy and flow cytometry, do not apply for ssDNA viruses, due to their weak fluorescence signal, which is often below the detection limit of microscopes or flow cytometers (Tomaru and Nagasaki 2007, Holmfeldt *et al.* 2012). Moreover, the small circular ssDNA genomes are not detected by the pulsed-field gel electrophoresis (Steward 2001, Tucker *et al.* 2011), and also the prior metagenomic techniques excluded ssDNA viruses by only taking account dsDNA viruses (Breitbart *et al.* 2002, 2004). This created the paradigm that tailed dsDNA viruses of the *Caudovirales* order predominate in nature. However, due to advances in amplifying and subsequently sequencing ssDNA from environmental samples, it is now known that ssDNA viruses are in fact highly abundant in a variety of ecosystems (Angly *et al.* 2006, Desnues *et al.* 2008, López-Bueno *et al.* 2009, Rosario *et al.* 2009, Tucker *et al.* 2011). Especially ssDNA phages belonging to family *Microviridae* have been commonly detected

(Roux *et al*. 2012). However, most of the ssDNA phage data originates from metagenomics studies and the number of cultured species is extremely limited. Therefore more extensive knowledge of the structural and ecological aspects of ssDNA phages is still lacking.

ssDNA phages include two morphologically distinct families: the *Inoviridae* and the *Microviridae* (King *et al*. 2012). Members of the *Inoviridae* encapsidate their circular positive-sense ssDNA genome within elongated, helically symmetric capsid. Family *Inoviridae* is divided into two genera, *Inovirus* and *Plectrovirus*, which differ by their respective host ranges (Ackermann 2006). In contrast, the ssDNA genome of microviruses is enclosed by a small, icosahedrally symmetric protein capsid (Fane *et al*. 2006). The ICTV divides the *Microviridae* into two subgroups, both of which use the rolling-circle method for replication but which differ in host range, genomic characteristics, and virion composition. Similarly to tectiviruses, these two subgroups do not share a significant sequence resemblance but their overall genomic organizations are similar. Enterobacteria infecting phages belong to the genus *Microvirus*, typified by the above-mentioned φX174, whereas members of the *Gokushovirinae* subfamily infect obligate intracellular parasites, such as species of *Bdellovibrio*, *Chlamydia* and *Spiroplasma* (Brentlinger *et al.* 2002). Virions of φX174-type consist of four structural proteins: major capsid protein (F), spike protein (G), DNA-binding protein (J) and DNA pilot protein (H) (McKenna *et al.* 1992). In contrast, gokushoviruses seem to only possess two structural proteins, which are homologous to those of the MCP and the DNA pilot protein of φX174 (King *et al*. 2012). Moreover, while φX174 requires both external and internal scaffolding protein for assembly (Fane *et al*. 2006), gokushoviruses only seem to encode the latter (Clarke *et al.* 2004).

All identified members of the *Microviridae* are lytic (Fane *et al*. 2006). However, proviruses related to microviruses have been found integrated into the genomes of the *Bacteroidetes* (Krupovič and Forterre 2011). Phylogenetic studies imply that these proviruses constitute a third subfamily within the *Microviridae*, tentatively named *Alpavirinae*.

2.11 Subcellular localization of bacterial proteins

Previously bacterial cells were merely viewed as simple vessels enclosing a homogenous solution of macromolecules and compounds, differing notably from the highly compartmentalized eukaryotic cells. However, our view on the internal architecture of bacterial cells has changed dramatically over the last two decades. Nowadays it is known that several fundamental biomolecules, such as the chromosome (Wang *et al.* 2013), mRNAs (Kannaiah and Amster-Choder 2014), proteins [reviewed by e.g. (Rudner and Losick 2010)] and lipids (Mileykovskaya and Dowhan 2010), as well as certain metabolites, including polyphosphate (Henry and Crosson 2013), are confined into specific subcellular compartment. Bacterial cells lack the standard set of membrane-bound

organelles (e.g. nucleus, endoplasmic reticulum, transport vesicles, Golgi apparatus, lysosomes, mitochondria and chloroplasts) which confine biomolecules, and thereby their functions, at specific sites in eukaryotic cells. However, unique organelle-like structures have been found from a number of bacterial species. These include e.g. carboxysomes, proteinaceous structures required for carbon-fixing (Yeates *et al.* 2008). It has also been shown that prokaryotic cells possess polymeric filaments, which are distant homologs of tubulin, actin and intermediate filaments [reviewed by (Shih and Rothfield 2006)]. This indicates that bacterial cells have a cytoskeleton, similarly to the more elaborate eukaryotic cells.

All in all, it would seem that the subcellular architecture of bacterial cells is especially intricate at the level of protein distribution. Due to fundamental breakthroughs in cellular imaging (e.g. high-resolution microscopy of fluorescent fusion proteins), it is now known that proteins frequently localize to distinct sites inside bacterial cells. In fact, it has been estimated that in *E. coli* almost one-fifth of the proteins are localized to a particular cellular site (Kuwada *et al.* 2015). Bacterial proteins have been found to accumulate at e.g. cellular poles, mid-cell region or along the long axis, where they display different patterns ranging from amorphous clusters to helical arrays. The asymmetric distribution of proteins is not necessarily static but can change dynamically over time in the course of cell cycle or in response to external stimuli (Shapiro *et al.* 2009, Rudner and Losick 2010). This dynamic, spatiotemporal architecture of bacterial proteins forms basis for several fundamental cellular functions, such as cell division, differentiation, chemotaxis and motility, as well as facilitates the adaptation to changing environments (Shapiro *et al.* 2009, Rudner and Losick 2010).

Even though the specific location of a significant number of bacterial proteins has been determined, mechanisms leading to this uneven distribution are largely uncovered. It would seem that so called ´diffuse and capture´ mechanism underlines the localization of most bacterial proteins (Rudner and Losick 2010). This means, that bacterial proteins reach their destination by simply diffusing randomly within the cell until encountering and subsequently being localized by a spatial cue. Perhaps the best-described mechanism for protein localization involves so called target protein, which captures the randomly diffusing protein to a specific site. An example of these target proteins is tubulin-like protein FtsZ, which forms a contractile ring-like structure at the mid-cell section, establishing the basis for divisome formation [reviewed by (Margolin 2005)]. FtsZ ring recruits other divisome proteins, which capture yet others and so forth, until cell division can be accomplished.

The ´diffuse and capture´ mechanism raises question on the ultimate positional information dictating the location of the target proteins. Dynamic self-assembly can be viewed as a unique variation of the diffusion and capture mechanism without the need for pre-existing target structures. For instance, chemoreceptors of *E. coli* preferably form large clusters as far apart from each as possible, and therefore usually locate at cell poles (Thiem *et al.* 2007, Thiem and

Sourjik 2008, Greenfield *et al.* 2009). Interestingly, some bacterial proteins are suggested to utilize bacterial cytoskeleton in targeting, similarly to their eukaryotic counterparts (Nevo-Dinur *et al.* 2012). There is compelling evidence of certain proteins possessing inherent localization signals, which determine their precise intracellular address inside the bacterial cell (Russell and Keiler 2007). On the other hand, in some cases it seems to be mRNA that is targeted, consequently determining the position of the protein products (Nevo-Dinur *et al.* 2011). Additionally, different cellular features, such as geometric cues or physical constrictions, are known to confine proteins into distinct subcellular sites (Rudner and Losick 2002).

It is possible that previously unknown mechanisms for protein localization will emerge in the future as a result of enhanced detection methods. There is already growing evidence of bacterial membranes possessing microdomains, enriched in certain lipids and/or cholesterol-like molecules. These so called lipid rafts may act as spatial cues for the positioning of a number of bacterial integral membrane proteins (Rudner and Losick 2010).

2.11.1 Proteins at the poles of bacterial cells

Numerous bacterial proteins with varying functions have been shown to localize at the chemically and physically unique cellular poles (Kirkpatrick and Viollier 2011, Laloux and Jacobs-Wagner 2014). Proteins accumulate at one or both of the poles or alternatively oscillate between them. The polar localization is controlled and can change in a temporal manner e.g. from unipolar to bipolar or from diffuse to polarly located (Laloux and Jacobs-Wagner 2014).

The cellular poles have several distinct features, which differ from the lateral sides of the cell, and can be exploited for positioning proteins. The characteristic feature of the poles is their geometry. For instance, division protein DivIVA of *B. subtilis* preferably attaches to the most concave membrane regions of the cell, i.e. the cellular poles and the newly formed cell division septum, both of which are showing high-level of negative curvature (Lenarcic *et al.* 2009, Ramamurthi and Losick 2009). Correspondingly, the positive curvature of endospore membrane is recognised by another protein of *B. subtilis*, SpoVM, during sporulation (Ramamurthi *et al.* 2009).

Another relevant feature of the cell poles is that they are mostly devoid of chromosomal DNA (Laloux and Jacobs-Wagner 2014). Due to volume-exclusion effects, the protein autoassembly is energetically more favourable in absence of bulky molecules, such as the chromosome (Ebersbach *et al.* 2008, Laloux and Jacobs-Wagner 2014). The nucleoid occlusion may also explain the polar localization of protein aggregates in *E. coli* (Winkler *et al.* 2010, Coquel *et al.* 2013). Misfolded proteins, diffusing in the cytoplasm, attach to each other by their exposed hydrophobic patches. As the aggregates get bigger, they are excluded from the vicinity of nucleoid, ending up at the poles, where the accumulations continue to expand.

The polar localization may also rely on the unique membrane composition of the cell poles. Certain proteins preferably attach to specific

phospholipids, such as cardiolipin (Mileykovskaya *et al.* 2003, Renner and Weibel 2011), which is enriched at the cellular poles and division sites of *E. coli* (Mileykovskaya and Dowhan 2000), *B. subtilis* (Kawai *et al.* 2004) and *P. putida* (Bernal *et al.* 2007). This enrichment proposedly reflects the conical shape of cardiolipin, which favours the attachment to membrane areas of high negative curvature (Huang *et al.* 2006, Renner and Weibel 2011). It has been shown, that the polar localization of two *E. coli* proteins, the mechanosensitive channel protein MscS and the osmosensory transporter ProP, correlates with the overall cardiolipid content of the cell (Romantsov *et al.* 2010). However, it remains to be revealed whether there is an actual physical interaction between these proteins and lipid moieties.

Proteins may also be directed to the poles due to the differing peptidoglycan composition (Laloux and Jacobs-Wagner 2014). In certain bacteria, such as *E. coli*, the peptidoglycan synthesis is restricted after cell septation to the lateral cell walls (de Pedro *et al.* 1997), where new peptidoglycan strains are continuously inserted between the old ones, pushing the older cell wall material progressively towards the poles (Rafelski and Theriot 2006). The difference between ´active´ and ´less active´ polar peptidoglycan has been suggested to provide a spatial cue for the polar localization (Rafelski and Theriot 2006, Laloux and Jacobs-Wagner 2014). However, more information is needed about the exact chemical compositions of polar and non-polar peptidoglycans. Also, the chemical elements triggering the polar localization are yet to be identified from peptidoglycan molecules (Laloux and Jacobs-Wagner 2014).

Interestingly, the lateral insertion of peptidoglycan may also be used as a ´shuttle` to transmit proteins towards poles (Laloux and Jacobs-Wagner 2014). Surface protein ActA of mammalian pathogen *Listeria monocytogenes* has been suggested to use this mechanism in its polar localization (Rafelski and Theriot 2006). Upon entry of the pathogenic bacterium to its host cell, ActA is secreted from the bacterial membrane, subsequently attaching to the peptidoglycan (García-del Portillo *et al.* 2011). In the progress of lateral cell wall growth, ActA presumably follows the older peptidoglycan towards the cell poles. As the cell ages, turnover of the peptidoglycan slows down, ultimately trapping ActA at the old pole (Rafelski and Theriot 2006). At the pole, it induces the host actin polymerization, resulting in an actin tail, which the bacterium uses for motility (Kocks *et al.* 1992, 1993, Lacayo *et al.* 2012). The polar localization of a bacterial protein may also be inherited from the division site (Laloux and Jacobs-Wagner 2014). If a protein is firmly positioned at the division site prior to cytokinesis, it may remain associated with the new poles of the daughter cells after cell division. This has been demonstrated with the polar landmark protein TipN of *Caulobacter crescentus* (Huitema *et al.* 2006, Lam *et al.* 2006).

2.11.2 Bacteriophages utilize the subcellular architecture

Animal viruses are known to exploit the intricate internal structure of eukaryotic cells to optimize their life cycles. In order to fulfill their functions,

viral proteins localize inside eukaryotic cells into specific compartments, such as the nucleus [e.g. the protein kinase of Epstein-Barr virus; (Gershburg *et al.* 2004)], the endoplasmic reticulum [e.g. cytomegalovirus proteins US2 and US11; (Wiertz *et al.* 1996, Lilley and Ploegh 2004)], the cell membrane [e.g. cytomegalovirus gp34 protein; (Atalay *et al.* 2002)] and the mitochondria [e.g. M11L protein of myxoma virus; (Boya *et al.* 2004)]. Similarly, there is compelling evidence of phages exploiting the bacterial cell architecture for efficient spatial coordination of the infection process. For instance, several phages have been shown to enter their bacterial host through the poles (Edgar *et al.* 2008), which are also the preferable site for DNA intake in competent cells (Chen *et al.* 2005). In case of *B. subtilis* phage SPP1 (Jakutytė *et al.* 2011) and *E. coli* infecting phage λ (Edgar *et al.* 2008), not only the DNA entry but also the initiation of replication occurs near the cellular poles. This is in contrast to the cellular replication machinery, which has been proposed to locate at mid-cell position (Li *et al.* 2002).

Interestingly, it has been shown that the cytoskeletal element MreB is required for the efficient replication of gram-positive bacteria infecting phages SPP1 and φ29 as well gram-negative bacteria infecting PRD1 (Muñoz-Espín *et al.* 2009). Whereas phages PRD1 and ϕ29 replicate their DNA genomes by the protein-primed mechanism, mediated by their terminal proteins, phage SPP1 uses initially theta mode of replication, which is later switched to rolling circle mechanism [reviewed by (Calendar 2006)]. In contrast to the replication of SPP1, which seems to occur at a single locus (Jakutytė *et al.* 2011), the replication machinery of φ29 has been shown to localize in peripheral helix-like structures, organized by MreB (Muñoz-Espín *et al.* 2009). Studies imply that after the phage DNA intake, the terminal protein recruits the φ29 genome to the nucleoid, where in encounters host RNA polymerase, leading to the transcription of proteins needed in replication (Muñoz-Espín *et al.* 2010). Later, the φ29 genome and the replication machinery redistribute to the peripheral, helix-like organization. The terminal protein of PRD1 genome similarly associates with the nucleoid (Muñoz-Espín *et al.* 2010). Rather unexpectedly, the terminal proteins of PRD1 and φ29 have been shown to contain localization signals, which direct them to the nucleus when produced in mammalian cells (Redrejo-Rodríguez *et al.* 2012). The sequence requirements for the nucleoid and nucleus localization seem to overlap to some extent, but could still be uncoupled. This implies that the two features have conserved independently during evolution, even though they probably share the same origin (Redrejo-Rodríguez *et al.* 2013).

Spatial organization may also be instrumental for the bacteriophage assembly. For instance, the virion formation of filamentous ssDNA phages, such as M13, is intimately coupled to the site of lysis (Webster and Cashman 1978). In order for the viruses to elicit chronic infection, newly synthesized virions need to be continuously released from the host. Consequently, the filamentous ssDNA phages are assembled at the cell membrane, instead of the cytosol (Webster and Cashman 1978). Interestingly, it has also been shown that

the assembly of the phage T4 head component occurs at the inner surface of the cell membrane, before being detached from the membrane, filled with DNA and ultimately connected to the tail and tail fibers in the cytosol (Casjens and Hendrix 1988).

2.12 Use of fluorescent protein fusions in bacterial protein localization studies

The isolation of autonomously fluorescing green fluorescent protein (GFP), from jellyfish *Aequerea victoria* (Shimomura *et al.* 1962), and the following gene cloning (Prasher *et al.* 1992) and protein expression (Chalfie *et al.* 1994, Inouye and Tsuji 1994) approaches have had a major influence on the biological research. The gene for GFP can be fused to the gene encoding the protein of interest, after which the resulting transgene is expressed *in vivo*. By GFP tagging, the localization of target protein can be detected in its native context by fluorescence microscopy. Fluorescent protein tagging has significantly contibuted in exploring the intracellular organization, for instance in bacterial cells, which do not possess many structures visible by electron microscopy (Margolin 2012). The significance of these studies was globally acknowledged in 2008, when Osamu Shimomura, Martin Chalfie and Roger Y. Tsien were awarded the Nobel Price in Chemisty for the discovery and development of GFP.

Over the years, the wild type GFP has been modified in many ways to enhance its fluorescence properties, accelerate folding and maturation as well as facilitate its use in different host organisms. One of these GFP mutants is enhanced green fluorescent protein, which displays increased brightness compared to the wild type one and is codon-optimized for production in mammals [reviewed by (Lippincott-Schwartz and Patterson 2003)]. Today, there are also a number of GFP variants with differing excitation and emission wavelengths, such as those emitting cyan [cyan fluorescent protein, CFP; (Heim *et al.* 1994)] and yellow [yellow fluorescent protein, YFP; (Ormö *et al.* 1996)] and their more intensively emitting variants enhanced cyan fluorescent protein (eCFP) and enhanced yellow fluorescent protein (eYFP), respectively.

Fluorescence microscopy has been traditionally utilized in studies of animal and plant viruses, while the small size of bacterial cells has been thought to limit its use in analysing phage-host interactions. However, modern high-resolution fluorescence microscopy combined with state-of-art imaging techniques allows detailed studies of the dynamic systems inside bacterial cells.

Fluorescence resonance energy transfer (FRET) is a phenomenon, which is commonly utilized to study inter- or intracellular protein-protein interactions [reviewed by (Sekar and Periasamy 2003)]. FRET occurs when energy is transferred from a fluorophore in an excited electronic state (the donor) to another fluorophore (the acceptor) through non-radiative dipole-dipole

interactions. In order for the energy to transfer, the emission spectrum of the donor fluorophore and the excitation spectrum of the acceptor fluorophore need to overlap. For instance, CFP and YFP offer such a chromophore pair (Miyawaki *et al.* 1997). By fusing two target proteins with these fluorescent proteins, the possible protein-protein interaction may be monitored in *vivo*, provided that the fluorophore-tags are in close proximity (1-10 nm) and oriented suitably in respect to each other [reviewed by (Sekar and Periasamy 2003)]. Whereas the maximum resolution in co-localization studies is limited by the optical resolution of the microscope (about 200 nm), a significantly higher resolution (less than 10 nm) is obtained by the FRET imaging, which is therefore an ideal technique for molecular interaction studies. The efficiency of FRET may be assessed by measuring the changes in the fluorescence emission of the donor and the acceptor or by monitoring the fluorescence lifetime of the donor. Since the FRET-mediated decrease in the fluorescence lifetime of the donor is independent of the fluorophore concentration or the intensity of the excitation light, it is considered as a reliable way to detect the protein-protein interaction (Lakowicz and Szmacinski 1993). In bacterial cells, FRET technique has been applied for instance in studies of chemotaxis proteins (Sourjik *et al.* 2007).

Another widely utilized imaging technique in cells expressing fluorescently tagged proteins is based on following fluorescence recovery after photobleaching (FRAP) [reviewed by (Reits and Neefjes 2001)]. In this assay, a certain area of the target cell is bleached with an intensive laser beam, the wavelength of which resembles that of the fluorophore-tag. As a result, most of the fluorophores at the region lose their fluorescence irreversibly. The fluorescence gradually recovers, as unbleached fluorophores diffuse to the bleached area. The mean fluorescence of the area is plotted against time after the photobleaching, resulting in the recovery curve. The diffusion rate of the fluorescently tagged protein may be determined from the recovery curve. The velocity of the diffusion depends on the size of the molecule, the viscosity and temperature of the cellular environment as well as possible interactions with other cellular molecules and membrane structures [reviewed by (Reits and Neefjes 2001)]. Since FRAP is an optical technique, it bears limited spatial resolution and has mostly been applied to eukaryotic systems. However, it has also been successfully utilized for studies of protein localization and kinetics in bacterial cells (Brass *et al.* 1986, Elowitz *et al.* 1999, Mullineaux *et al.* 2006).

Tagging with fluorescent proteins has always been somewhat controversial, due to the possibility of eliciting changes in the conformation and/or function of the target protein. This is especially problematic among proteins that polymerize or otherwise assemble into larger complexes. There are also concerns that the tagging could cause localization artifacts (Margolin 2012). For instance, results from a cryo-electron tomography study imply that the helical localization pattern of the MreB cytoskeletal protein, observed in various studies (Shih *et al.* 2005, Vats and Rothfield 2007, Wang *et al.* 2012), is actually an artifact caused by the YFP tag (Swulius and Jensen 2012). However, despite its

deficiencies, the fluorescent fusion protein technique has retained its fundamental role in the bacterial cell biology.

2.13 Chaperones

Protein folding is a process in which a polypeptide obtains its functional three-dimensional conformation, known as the native state. The folding results from interactions between the amino acid side chains. Hydrophobic effect is a crucial driving force behind the protein folding (Pace *et al.* 1996). In this phenomenon, hydrophobic side chains are buried in the core of the protein, where they are not exposed to the aquatic, hydrophilic environment. In addition, the formation of hydrogen bonds, salt bridges and disulfide bonds contributes in completing the thermodynamically favorable folded state. According to classical Anfinsen´s theory, the linear sequence of a polypeptide chain contains all the necessary information for forming the protein´s three-dimensional structure (Anfinsen 1973). Indeed, especially small, single-domain proteins reform efficiently into their native conformation *in vitro* under optimal conditions (reviewed by Hartl and Hayer-Hartl 2002). However, the folding of a more complex polypeptide into its respective three-dimensional structure may be more complicated due to the presence of more than one local thermodynamic minima. Furthermore, polypeptide folding *in vivo* faces additional challenges, as the extremely high concentration of proteins and other macromolecules within the cell [300-400 mg ml^{-1}, (Zimmerman and Trach 1991)] makes the folding error-prone and inefficient [reviewed by (Hartl and Hayer-Hartl 2002, 2009)]. As a result, folding-intermediates and misfolded polypeptides accumulate in the cytosol. These unfolded polypeptides expose hydrophobic patches, which stick together, leading to the formation of protein aggregates, non-specific associations of proteins, which have lost their proper biological function [reviewed by (Hartl and Hayer-Hartl 2002, 2009)]. Moreover, heat shock or other cellular stress increases substantially conformational changes and aggregation among proteins. In response, a complex cellular machinery of molecular chaperones has evolved to counteract the problems brought by the protein misfolding and aggregation.

Molecular chaperones include unrelated classes of proteins, instrumental in maintaining protein homeostasis (Ellis 1987). Many of the chaperones are called heat shock proteins (Hsps), due to their overexpression upon heat shock, even if they are constitutively expressed under normal cellular conditions. Chaperones interact transiently with the non-native protein without being present in the final folded protein structure. Recognition of non-native proteins occurs predominantly via hydrophobic residues, which are temporarilly exposed to the intracellular environment, for instance during the early stages of folding or when the misfolding occurs [reviewed by (Hartl and Hayer-Hartl 2002, 2009)]. The interaction with the chaperone molecules prevents the non-native protein from aggregating and mediates its proper folding into native

conformation. Chaperones function also in several other essential cellular processes. Instead of promoting the proper protein folding, they can for instance keep the newly synthesized protein in unfolded conformation, thereby facilitating its translocation across membranes. Chaperones may also disaggregate nonfunctional protein aggregates [reviewed by (Saibil 2013)].

Based on their molecular mass, chaperones are classified into several groups, such as Hsp60, Hsp70, Hsp90, Hsp100 and small heat shock proteins. Many of the chaperone classes are evolutionary conserved throughout all three kingdoms of life and can also be found in intracellular organelles. It would seem that the co-operation of the distinct chaperone classes is required for protein homeostasis in cellular environment. Many of the known chaperones require accessory proteins, so called co-chaperones, for proper function (Georgopoulos 1992).

Chaperonins constitute a universally conserved chaperone class, which facilitates the folding of polypeptides in an ATP-dependent manner (Hildenbrand and Bernal 2012). These proteins characteristically form an oligomeric cage around an unfolded polypeptide chain, preventing unwanted aggregation and providing favorable environment for folding. Chaperonins have been found to function in one or two-ring structures, with or without the assistance of co-chaperonins. They are divided into two classes, based on the origin and distinct structural features. Chaperonins belonging to group I exist in eubacteria, mitochondria and chloroplasts, while group II chaperonins have been found in eukaryotes and archaea (Hildenbrand and Bernal 2012).

2.13.1 GroEL

One of the best described chaperones is the group I chaperonin of *E. coli*, called GroEL. This protein is structurally and functionally connected to its co-chaperonin GroES. Genes encoding GroEL and GroES reside in the same operon and are coordinately expressed under normal physiological conditions (Fayet *et al.* 1989). In addition, their concentration is drastically increased during stressful situation, such as heat shock. Certain essential proteins of *E. coli* have been shown to utilize in their folding exclusively the GroEL/GroES complex, and none of the other existing chaperone machineries (Kerner *et al.* 2005), which makes the complex absolutely irreplaceable for the bacterial cell (Fayet *et al.* 1989).

The structures of both the chaperonin and its co-chaperonin have been solved to atomic resolution (Braig *et al.* 1994, Xu *et al.* 1997). GroEL consists of 14 identical subunits of 57 kDa, forming two toroidal-shaped heptameric rings arranged against each other back-to-back (Braig *et al.* 1994). Each subunit consists of three domains. A large equatorial domain contains the attachment site for ATP and is also responsible for most of the interactions within the ring as well as the contacts between the opposing rings (Xu and Sigler 1998). It is connected via a small hinge domain to an apical domain (Xu and Sigler 1998), which lines the opening of the GroEL central cavity and mediates the

interactions with the polypeptide substrate and the co-chaperonin (Braig *et al.* 1994).

The dome-shaped co-chaperonin GroES is a homoheptameric ring of 10 kDa subunits, which binds the chaperonin asymmetrically (Xu *et al.* 1997). From each GroES subunit protrudes a mobile stretch of 16 hydrophobic residues, presumably mediating the interaction with a hydrophobic patch on top of the GroEL apical domains (Landry *et al.* 1993, 1996, Xu *et al.* 1997). Upon binding to GroEL, the flexible mobile loop becomes structurally ordered, but the only physical interaction with the chaperonin seems to occur through a universally conserved hydrophobic tripeptide residing in the mobile loop.

Substrate proteins are usually passed on to the GroEL/GroES complex from other chaperone machines (*E. coli* Trigger factor and Hsp70) in unstable molten globule state. Hydrophobic residues, exposed by the molten globule intermediate, interact with hydrophobic surfaces of amphiphilic helices in the GroEL apical domains. This registers the substrate protein inside one of the GroEL rings (*cis*-ring) (Bukau and Horwich 1998). ATP molecules attach to each of the equatorial domains of the *cis*-ring, eliciting conformational changes in the apical domains (Hildenbrand and Bernal 2012). This triggers the binding of the co-chaperonin GroES on top of the GroEL *cis*-ring, which is holding the substrate (Ranson *et al.* 2001). The substrate molecule is released into the central cavity, and the folding commences. The folding may continue for about 10 seconds, until each of the seven ATP molecules located at the cis-ring have hydrolysed (Hildenbrand and Bernal 2012). Ascribed to this ATP hydrolysis, new ATP and substrate molecules bind to the opposing GroEL ring (*trans*-ring) (Rye *et al.* 1997), ultimately leading to the release of adenosine diphosphate (ADP), co-chaperonin and partially or fully folded polypeptide from the *cis*-ring (Rye *et al.* 1997, Hildenbrand and Bernal 2012). Simultaneously, the *trans*-ring transforms into the *cis*-ring conformation. If the released polypeptide has not attained its final conformation, it is captured by GroEL for another folding attempt. Iterative rounds of encasing, folding and release are apparently required for the polypeptide to reach its native fold.

The GroEL/GroES chaperonin complex is essential for the morphogenesis of several bacteriophages. In fact it was originally discovered as a bacterial factor required for the proper assembly of the phage λ head and phage T5 tail (Georgopoulos *et al.* 1973, Zweig and Cummings 1973), and only later shown to be equally necessary for cell viability (Fayet *et al.* 1989). The chaperonin complex functions in virion formation by facilitating the proper folding and oligomerization of viral proteins at different states of assembly. Unlike phages such as λ (Georgopoulos *et al.* 1973), T5 (Zweig and Cummings 1973) and PRD1 (Hänninen *et al.* 1997), which depend on both GroEL and GroES in their morphogenesis, bacteriophage T4 encodes its own co-chaperonin ortholog, Gp31, and thereby is only dependent on the host-encoded chaperonin (van der Vies *et al.* 1994). The presence of Gp31 is essential for the folding of the T4 MCP Gp23. Moreover, the host cell may utilize Gp31 for its own functions. Gp31 restores the growth defect in a temperature-sensitive mutant of GroES as

well as substitutes the host-encoded co-chaperonin in refolding of ribulose-1,5-bisphosphate carboxylase/oxygenase and citrate synthase (van der Vies *et al.* 1994). However, the complementation effect is not reverse, since GroES cannot replace Gp31 in the T4 head formation (van der Vies *et al.* 1994). Replacement of GroES with virally encoded Gp31 seems to expand the internal chaperonin cavity, accommodating the folding of the relatively large T4 MCP, the size of which is at the upper limit for the GroEL/GroES cage (Hunt *et al.* 1997). A homologous protein to Gp31 was discovered from coliphage RB49 (Ang *et al.* 2001). The co-chaperonin of RB49, termed *c*o-*c*haperonin *co*gnate (CocO), substitutes for GroES in the folding of citrate synthase and T4 MCP Gp23. Gp31 and CocO complement each other to some extent, but still show preference for folding their own MCPs (Keppel *et al.* 2002). Despite the lack of any significant sequence similarity between the bacterial and phage-encoded co-chaperonins, they all seem to share a similar mobile loop, which, in case of GroES and Gp31, has been shown to tether to GroEL (Landry *et al.* 1993, 1996, van der Vies *et al.* 1994, Xu *et al.* 1997, Richardson and Georgopoulos 1999, Ang *et al.* 2001, Keppel *et al.* 2002).

Bacteriophages may also possess the potential to encode equivalents for chaperonin GroEL. In fact, GroEL ortholog genes have been found in certain phage genomes (Holmfeldt *et al.* 2013, Cornelissen *et al.* 2012). The putative virally encoded chaperonin of *Pseudomonas* aeruginosa phage φEL was characterized, confirming its analogous function to GroEL (Kurochkina *et al.* 2012).

In addition to the chaperonin and co-chaperonin orthologs, viruses have also been shown to encode other protein types influencing the function of the host chaperonin complex. Protein Gp39.2, found in phages T4, RB43, RB49 and RB69, suppresses the defects in the chaperonin machinery, when the GroEL or GroES mutant in question shows weakened affinity to its binding partner (Ang and Georgopoulos 2012). Apparently this complementation effect is achieved by enhancing the open conformation of GroEL.

3 AIMS OF THE STUDY

Recent studies demonstrate that the tailless bacteriophage types have more significant roles in varying environmental settings than has been previously thought. This thesis aims to give new insights into the bacteriophage research by addressing the diversity and functionality of three tailless phages, which have a structural membrane component. The first part of the thesis focuses on the assembly of the internal-membrane containing bacteriophage PRD1. Despite the extensive research conducted with this model virus over the last decades, its assembly process has remained partly uncharacterized. The second part of the thesis introduces two novel bacteriophages: enveloped dsRNA phage φNN and internal-membrane containing ssDNA phage FLiP. The specific aims of this thesis were as follows:

1. To develop new genetic tools for detecting phage-encoded proteins in bacterial cells in an attempt to understand the dynamics of bacteriophage PRD1 proteins in its host bacterium *E. coli*.

2. To characterize contribution of PRD1 non-structural proteins P17 and P33 to *E. coli* GroEL/GroES chaperonin activity during PRD1 assembly.

3. To characterize *Pseudomonas* infecting, φ6-related phage φNN and to analyse its relationship with previously identified viruses, consequently broadening the view on the *Cystoviridae* family.

4. To conduct a genetic and structural characterization of *Flavobacterium* phage FLiP and to determine its relatedness to previously identified viruses.

4 OVERVIEW OF THE METHODS

The methods used in this thesis are summarized in the table below. Detailed descriptions of the used 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
Fluorescence lifetime microscopy	I
Time-correlated single photon counting	I
FRET assay	I
Molecular cloning and polymerase chain reaction	I, II
Production of recombinant proteins	I, II
Western Blotting	I, II
Sedimentation assay by rate zonal centrifugation	I, II
Complementation assay	I, II
Confocal microscopy	I, II
FRAP assay	II
Virus and bacterial strain isolation	III, IV
Propagation and purification of viruses	III, IV
Infection cycle analysis	III, IV
Chloroform sensitivity analysis	III, IV
Adsorption assay	III, IV
Viral genome extraction	III, IV
Agarose gel electrophoresis	III, IV
Negative staining and TEM imaging	III, IV
Plaque assay	I, II, III, IV
SDS-PAGE	I, II, III, IV
Nucleotide sequencing and annotation	I, II, III, IV
Comparative genomics	II, III, IV
Protein modelling	III
N-terminal amino acid sequencing and mass spectrometry	IV
Buoyant density determination by density gradient centrifugation	IV
Cryo-EM imaging and image processing	IV

5 RESULTS AND DISCUSSION

5.1 Fluorescent protein fusions reveal the subcellular localization of bacteriophage PRD1 proteins

5.1.1 Construction of a library of fluorescent protein expression vectors

The aim was to create genetic tools for the production of fluorescent fusion proteins in bacterial cells. Hence, a library of fluorescent protein expression vectors was constructed by cloning the genes encoding eGFP, eCFP and eYFP into bacterial vectors pSU18 (Bartolomé *et al.* 1991) and pET24 (Novagen) (I, Fig. 1, Table 1). These two backbone vectors belong to different incompatibility groups (harbouring replicons p15A and ColE1, respectively), enabling their co-expression in a cell. pET24 is a high copy-number expression vector, whereas pSU18 is a low-copy number plasmid, which has been commonly used to produce moderate levels of functional PRD1 proteins for complementation assays (Bamford and Bamford 2000, Rydman *et al.* 2001, Rydman and Bamford 2003).

The fluorescent protein genes were introduced into the multiple cloning site of the backbone vectors, under the control of either lac- (in pSU18) or T7-promotor (in pET24). A linker sequence of six glycine residues, as well as restriction enzyme recognition sites for the subsequent cloning, were designed to either the 5´ or 3´ end of the fluorescent protein gene (I, Fig. 1). This enables the fusion of a target gene into either end of the fluorescent protein gene, resulting in either N-terminal (fluorescent protein fused to the N-terminus of the target protein) or C-terminal (fluorescent protein fused to the C-terminus of the target protein) fusion proteins. Also, in case of the C-terminal fusions, a T7 ribosome-binding site is designed upstream of the target gene, enabling the translation in bacterial cells. The glycine linker was created between the fluorescent protein and the target protein to prevent steric hindrance, which could interfere with the folding of one or both of the polypeptides. Glycine has the smallest side chain and it confers great structural flexibility; consequently it

is a popular choice in polypeptide linker design (Shimozono and Miyawaki 2008).

Confocal microscopy studies demonstrated that the recombinant fluorescent proteins produced from these vectors in *E. coli* had retained their ability to emit fluorescence (data not shown), and therefore the vectors were applicable for genetic tagging. Although the library of fluorescent protein vectors was initially designed for studies of phage-specific proteins, it has the potential to be utilized in a wide range of protein studies in bacteria.

5.1.2 Production of fluorescent fusions of PRD1 proteins

To understand the dynamics of viral proteins is bacterial cells, we utilized the constructed vector library to produce cyan and yellow fusion proteins of individual bacteriophage PRD1 proteins and host chaperonin GroEL (I, II). Phage proteins included structural proteins either as an oligomer (MCP P3, penton protein P31 or spike protein P5) or a monomer (host attachment protein P2), as well as an integral membrane protein (vertex-stabilizing protein P16) and two non-structural proteins (assembly protein P17 and putative assembly factor P33).

Incorrectly positioned fluorescent fusion tag may prevent the correct folding of the target protein, leading to the accumulation of insoluble protein aggregates, which lack biological activity. However, high-resolution structural data available of the target proteins P2, P3, P5, P16 and P31 (Benson *et al.* 2002, Xu *et al.* 2003, Abrescia *et al.* 2004, Merckel *et al.* 2005) was used in designing the optimal site for the fluorescent fusion. For instance, in case of P31 monomer, both termini reside on the surface, while in P31 pentamer, the C-termini are buried in the middle of the oligomer and only the N-termini point outwards (Abrescia *et al.* 2004). This implies that the fusion of a fluorescent tag to the N-terminus of P31 would less likely disrupt the oligomer formation.

In order to monitor the solubility of the expressed fluorescent fusion proteins, the cell extracts of the bacterial cultures were subjected to high-speed centrifugation and subsequently the protein ratio between the soluble supernatant an the insoluble material was determined by SDS-PAGE and Western blotting, using monoclonal or polyclonal antibodies. According to these assays, the fluorescent fusions of the structural proteins P2, P3, P5 and P31 as well as the fusions of the non-structural proteins P17 and P33 were mainly expressed in soluble form (I, II). The soluble supernatants were subsequently subjected to rate zonal centrifugation (also known as velocity sedimentation), which separates particles on basis of their size and shape, followed by SDS-PAGE and Western blotting analysis of the obtained fractions (I, Fig. 2; II, Fig. 3). The sedimentation velocities of the fluorescent fusion proteins were compared to those of selected controls to assess their molecular mass and oligomeric state. Both of the P2 fluorescent fusions sedimented as expected for a monomer (I, Fig. 2A), corresponding to the monomeric state of P2 in the virion (Grahn *et al.* 1999). Similarly, the sedimentation velocity of the P33-eYFP fusion protein implied a monomeric state (II, Fig. 3A). It is not known, whether the

non-structural protein P33 forms oligomers. Among the fusions of oligomeric PRD1 proteins (trimeric proteins P3 and P5, tetrameric P17 and pentameric P31; I, Fig. 2B, C, D and E) and tetradecameric GroEL (II, Fig. 3D), the size distribution was wider, and both monomeric and oligomeric forms were detected in each case. The same phenomenon has previously been noticed for native P17, derived from PRD1-infected cells, implying that the protein is involved in complexes of different sizes during the infection (Caldentey *et al.* 1999). Interestingly, the sedimentation patterns differed between the corresponding C- and N-terminal fusions, indicating that certain fusions were less optimal for oligomer formation. This was evident for instance with the fusions of the penton protein P31 (I, Fig. 2C). Whereas the C-terminal fusion of P31 sedimented mostly as a monomer, the corresponding N-terminal fusion seemed to also form oligomers. As discussed above, this difference presumably results from the steric hindrance caused by the fluorescent tag at the C-terminus.

Both the N- and C-termini of GroEL are located inside the GroEL cylinder (Braig *et al*. 1994), therefore it is feasible that the attachment of a bulky fluorescent protein (~27 kDa) would have a detrimental effect on the oligomerization and function of the chaperonin. When the sedimentation of the plasmid-produced GroEL-eCFP (II, Fig. 3D) was compared to that of endogenous GroEL (II, Fig. 3E), the GroEL fusion seemed to sediment somewhat slower. However, oligomeric forms were detected, implying that the GroEL-eCFP is able to form the native cylinder-like structure. Since the endogenous GroEL is present in the expression system, we cannot exclude the possibility, that the recombinant GroEL-eCFP monomers form the oligomeric complex together with the host-encoded GroEL monomers. Also, although most of the fusion protein was expressed in soluble form, a portion of the GroEL-eCFP fusion was found in the pellet (II, Fig. 3D), implying some level of misfolding and/or aggregation.

All in all, the sedimentation assays imply that the fluorescently tagged PRD1 proteins and GroEL had retained their ability to oligomerize, at least to some extent. However, this does not necessarily mean that each fluorescently tagged protein had retained its biological activity. We analysed the functionality of the fluorescent fusion proteins of P5 and P31 by complementation assay using PRD1 mutants *sus690* and *sus525*, which have an amber mutation in gene *V* (encoding P5) or *XXXI* (encoding P31), respectively. Both C- and N-terminal fusion proteins of P5 and P31 rescued the defect in the corresponding gene, resulting in similar titers as in the suppressor strain or in strains plasmid-producing the wild type protein (I, Table 2). In each case, the titer of the mutant virus was at least five orders of magnitude higher than in the negative control strain, carrying the backbone plasmid pSU18. This implies that the fluorescently tagged structural proteins P5 and P31 are able to oligomerize and function correctly in the virion.

5.1.3 Fluorescently labelled PRD1 proteins localize in different parts of the bacterial cell

Confocal microscopy was utilized to study the subcellular localization of plasmid-produced fluorescent fusions of PRD1 proteins P2, P3, P5, P16, P17 and P31. Previously it has been shown that the terminal protein P8 of PRD1 associates with the nucleoid (Muñoz-Espín *et al.* 2010), but to my knowledge, this was the first attempt to gain a broader view on the localization of PRD1 proteins inside bacterial cells. The fusion proteins were produced in HMS174(DE3) strain of *E. coli*, the natural host bacterium of PRD1. The protein localization was studied when the bacterial cells were at stationary state and data was collected at least from three replicate cultures. In these studies, two localization patterns were seen among the fusion proteins: homologous distribution in the cytosol or accumulation into one or more loci, located most commonly at the cellular poles.

Both fluorescent fusions of the monomeric host recognition protein P2 were mostly distributed evenly inside the bacterial cell (I, Fig. 3C and D). The same diffuse distribution was seen with the fusions of the integral membrane protein P16 (I, Fig. 3A and B). This was a somewhat surprising result, since membrane proteins typically localize evenly around the circumference of the cell [e.g. (Maier *et al.* 2008, Li and Young 2012)]. In the PRD1 virion, P16 interacts with the spike protein complex at the fivefold axis, thereby connecting the vertex to the internal membrane and stabilizing the structure (Abrescia *et al.* 2004, Jaatinen *et al.* 2004). It is possible that the attachment of P16 requires some additional phage-encoded proteins, missing from the expression system, thereby explaining the distribution pattern of the fluorescent fusion protein. At least membrane protein P10 and non-structural assembly factor P17 (and possibly P33) are needed for the proper formation of the internal membrane-containing procapsid (Mindich *et al.* 1982a, Vilen *et al.* 2003). Alternatively, the fluorescent protein fusion might have prevented the proper folding of P16, affecting its ability to attach to the cell membrane.

Accumulation into distinct loci was mostly detected among the fluorescent fusions of oligomeric proteins P3 (I, Fig. 3I and J), P5 (I, Fig. 3E and F) and P17 (I, Fig. 3K), but was also seen to some extent with the penton protein P31 fusions (I, Fig. 3G and H). The formation of the fluorescent loci was especially prominent among the fusions of P5, which displayed at least one polar fluorescent locus in the majority of cells (81 and 64 % of the cells displaying loci with the C and N-terminal fusions, respectively; I, Fig. 3E and F). The corresponding C- and N-terminal fusions differed in terms of the localization patterns. For instance, the monomeric P31-eYFP (I, Fig. 2C) was distributed evenly in 95 % of the cells (I, Fig. 3G), whereas eYFP-P31, which showed some degree of oligomerization, localized somewhat more frequently to one of the cellular poles (10 % formed loci; I, Fig. 3H). As a control, the fluorescent protein eYFP was produced alone in *E. coli*. It displayed an even distribution in the cytoplasm (I, Fig. 3L), as reported previously (Deich *et al.* 2004, Scheu *et al.* 2008). This implies that the localization pattern seen with the

fusions of the PRD1 oligomeric proteins was dictated by the properties of the phage-encoded proteins, and not the fluorescent protein tag.

In addition, the fluorescent fusion protein technique was applied for co-localization studies to detect the possible association between the fluorescent fusions of P5 and P31, two structural proteins interacting at the PRD1 vertex (Caldentey *et al.* 2000). It was demonstrated that when the fusion proteins P5-eCFP and eYFP-P31 were produced simultaneously in *E. coli*, they co-localized in 66 % of the observed loci.

The protein localization was studied further by quantitating the number of loci from cells with clear localization patterns (I, Fig. 4). The fusion proteins of P31 and P5 formed mostly one polar loci (75–100 %), while the P3 and P17 fusions showed more variance in the formation of loci. eYFP-P3 displayed more than two fluorescent foci in most cases. The MCP P3 interacts in the virion with the internal membrane via its N-terminus (Benson *et al.* 1999, Abrescia *et al.* 2004). This membrane-connection may explain the formation of multiple loci, which seemed to localize around the circumference of the cell (I, Fig. 3J). Even though the samples were always taken from the parallel cultures at the stationary phase, it is naturally possible, that they included dividing cells, which could explain the presence of two or more loci. However, the formation of multiple loci did not seem to correlate with elongated cell structure, an indicator of active cell division.

To conclude, in our studies the fluorescent protein-tagged monomeric PRD1 proteins distributed homogenously (3.9 % of the cells formed fluorescent loci) in the bacterial cytosol, while the oligomeric PRD1 proteins were more prone to form distinct loci (51.0 % of the cells formed loci), mostly at the cellular poles.

The lipid composition of the PRD1 internal membrane vesicle differs from that of its host (Laurinavičius *et al.* 2004), implying that the phage acquires its lipids selectively, excluding host-encoded proteins and some of the lipids. Eukaryotic membrane-containing viruses show such as a selection as a result of viral budding in specific lipid microdomains, called lipid rafts (Suomalainen 2002, Briggs *et al.* 2003). It would be tempting to suggest, that PRD1 protein oligomerization and/or assembly of the PRD1 procapsid takes place specifically at the cellular poles, possibly mediated by a specific membrane composition. The poles have been shown to play an active role in the life cycles of a number of gram-negative and gram-positive bacteria infecting phages, being the preferable site e.g. for DNA entry (Edgar *et al.* 2008). Moreover, the replication of *B. subtilis* phage SPP1 (Jakutytė *et al.* 2011) and *E. coli* phage λ (Edgar *et al.* 2008) have been shown to occur at the poles.

It could be speculated, that the fluorescent fusion protein accumulation detected at the cellular poles was a direct consequence of aggregation. The aggregation could have resulted from misfolding, caused by the fluorescent protein fusion, and/or fusion protein overproduction. These factors have shown to cause artifactual localization patterns, particularly polar accumulations (Winkler *et al.* 2010, Landgraf *et al.* 2012, Margolin 2012, Swulius

and Jensen 2012, Coquel *et al.* 2013). However, here the fluorescent fusions of P5 and P31 were able to incorporate into infectious virions, indicating that they had retained their native fold and biological functionality, despite the fluorescent tag. Also, the used low copy-number plasmid pSU18 produces moderate levels of PRD1 proteins (Bamford and Bamford 2000, Rydman *et al.* 2001, Rydman and Bamford 2003). As eYFP was expressed alone from this backbone vector, the fluorescence signal was diffusively distributed (I, Fig. 3L), indicating that the expression level had not caused any polar aggregation. Moreover, sedimentation assay implied, that all the fusion proteins were expressed mostly in soluble state (I, Fig. 2; II, Fig. 3).

In order to reveal whether the observed localization patterns of monomeric and oligomeric viral proteins are universal, the subcellular positions of PRD1 proteins, or virally encoded proteins in general, should be explored more comprehensively. When the virally encoded recombinant proteins are produced from plasmids, their behaviour is solely dictated by their intrinsic properties, such as interactions with bacterial proteins and other cellular factors. In contrast, upon viral infection, synthesized proteins interact also with other viral proteins, which may change their behaviour. Thus, the localization patterns of virally encoded proteins may differ between viral infection and artificial expression system.

The use of less-bulky fluorescent tags and membrane-permeable organic dyes (Griffin *et al.* 1998, Charbon *et al.* 2011, Laloux and Jacobs-Wagner 2014) provides new prospects for future protein localization studies. Special attention should also be paid to the recombinant protein expression levels, which should be adjusted close to native level to obtain more physiologically relevant data (Laloux and Jacobs-Wagner 2014). Moreover, post-imaging software could be utilized to quantify the dynamics of protein localization, which is extremely challenging by simple visual measurements (Guberman *et al.* 2008, Sliusarenko *et al.* 2011, van Teeffelen *et al.* 2012, Laloux and Jacobs-Wagner 2014). Visual inspection of localization patterns may unconsciously lead to biased results, which could be avoided by the automation (Laloux and Jacobs-Wagner 2014). All in all, the constantly evolving machineries and techniques for fluorescence detection offer fascinating opportunities for the study of phage-host interactions.

5.1.4 FRET analysis of interaction between PRD1 vertex proteins P5 and P31

In addition to the co-localization studies, FRET technique was utilized for assaying the possible interaction between the fluorescent fusions of the PRD1 vertex proteins P5 and P31. The energy transfer was measured at the fluorescent locus area or the entire cell with both of the fusion protein pairs (P5-eCFP and eYFP-P31, eYFP-P5 and P31-eCFP), as well as with a negative control P31-eCFP. In these assays, a minimal energy transfer was detected between P5-eCFP and eYFP-P31 in the fluorescent locus, when measuring the changes in the lifetime of the donor fluorophore eCFP (I). Also, an additional peak at the eYFP wavelength was then detected. However, no clear interaction signal between

the fluorescent fusions of P5 and P31 was seen. The fluorescent fusions may have prevented the two virally encoded proteins from interacting, even though the fusions coexist in the same loci. It is also possible that the plasmid-produced P5 and P31 do not interact in absence of other phage-specific proteins, although they interact in the virion vertex. It would be interesting to study, whether the FRET signal could be detected between the P5 and P31 fusions by changing the linker sequence. It has been shown that the optimization of the linker in the fluorescent protein fusions may have a major impact on the success of the FRET assay (Lissandron *et al.* 2005, Shimozono and Miyawaki 2008). It is possible that the designed flexible linker of six glycine residues was not optimal for this specific interaction study, and for instance a more rigid linker would have been more suitable.

5.2 Non-structural proteins P17 and P33 participate in the assembly of bacteriophage PRD1 – possibly in association with the host chaperonin complex

5.2.1 *E. coli* exploits PRD1 non-structural proteins P17 and P33

Previously it has been shown that the assembly of infectious PRD1 virions is dependent on the host chaperonin complex GroEL/GroES (Hänninen *et al.* 1997). The chaperonin complex ensures the folding of capsid proteins P3 and P5, as well as the folding/assembly of major membrane-bound protein P11 and certain small (<10 kDa) membrane proteins (Hänninen *et al.* 1997). In addition, two virally encoded assembly factors, P10 and P17 (Mindich *et al.* 1982a, Vilen *et al.* 2003) are needed for the proper formation of PRD1 virions. Host membrane patch enriched in P10 has been suggested to act as a scaffold for the procapsid formation (Rydman *et al.* 2001). The genes encoding P17 and putative protein P33 overlap in the late operon OL2, which also comprises genes for other proteins required in the assembly or genome packaging (Grahn *et al.* 1994) (II, Fig. 2). In the absence of P33 mutants, it is not known whether this protein is essential for PRD1. However, no transposon was inserted into *XXXIII* gene (encoding P33) (Vilen *et al.* 2003), which implies an essential function for P33. Alternatively, or additionally, it may confer a selective advantage in changing environment.

In this study, complementation assay was used to analyse if PRD1 P17 and/or P33 could compensate the defects in GroEL or GroES of *E. coli*, similar to previously characterized phage-encoded assembly factors (van der Vies *et al.* 1994, Ang *et al.* 2001, Ang and Georgopoulos 2012). Due to the essential role of the GroEL/GroES complex in the cell (Fayet *et al.* 1989), its complete deletion would be lethal for the cell. We used conditionally lethal mutants of *E. coli* strains DW717(pLM2) (a GroEL mutant) and DW719(pLM2) (a GroES mutant), carrying a temperature-sensitive (ts) mutation *groEL59* or *groES619* respectively,

as well as a plasmid pLM2, which encodes genes for the PRD1 receptor complex. In *groES619* mutation, the glycine residue preceding the hydrophobic loop motif of GroES has been substituted for aspartic acid (Landry *et al*. 1993). This results in a more rigid loop, impairing its interaction with GroEL (Ang and Georgopoulos 2012). The mutant phenotype of GroEL or GroES is observed at the non-permissive temperature of 42 °C, causing severe growth defect, while at 37 °C the proteins adopt their normal, functional phenotypes.

To test the complementation ability, recombinant plasmids encoding PRD1 P17, P33 or both were introduced into the *groE* mutant strains. Recombinant plasmids harbouring genes for wild type GroEL or GroES were transformed into the mutant strains as positive control, while the mutant strains carrying pSU18 only, the backbone vector of the recombinant plasmids, were used as negative controls. The formation of bacterial colonies was analysed with each strain, at both the permissive and non-permissive temperatures, and compared to the colony forming ability of the wild type strain DW720(pLM2). At permissive temperature, all the *groEL* mutant strains formed similar numbers of colonies. However, at the elevated, non-permissive temperature, the growth of the negative control strain and the strains expressing P17 and/or P33 was inhibited, and only the strain carrying the recombinant *groEL* gene reached a similar level of colony formation than the wild type strain. This demonstrates that the plasmid-production of P17 and P33 did not substitute for the defective GroEL in bacterial growth.

The same analysis was conducted for the *groES* mutant strains. Again no significant difference was detected in the colony forming ability at the permissive temperature. However, at non-permissive temperature, the number of colonies on the negative control strain declined about four orders of magnitude, whereas that on the wild type strain, or the mutant strain producing recombinant GroES, remained the same. Interestingly, wild type levels of bacterial colonies were also achieved by the plasmid-production of P17 and/or P33 (II, Fig. 2A), suggesting that these two virally encoded proteins restored the growth of the bacterial host, which is producing defective GroES.

5.2.2 PRD1 P17 and P33 function in phage propagation

Additionally, the plaque formation of PRD1 was tested in each of the *groEL* or *groES* mutant strains, carrying plasmids encoding PRD1 P17, P33 or both (II, Fig. 2B). In these plaque assays, the growth constraints of PRD1 limited the non-permissive temperature to 40 °C. At permissive temperature, a similar number of plaques were detected on the *groEL* and *groES* mutant strains, compared to the wild type strain. At non-permissive temperature, the viral titer was considerably lower (at background level) in the *groEL* mutant strains expressing P17 and/or P33, compared to the titer obtained in cells with endogenous or recombinant GroEL. This implies that the production of P17 and P33 did not complement the ts *groEL* mutation in PRD1 propagation.

In the case of the *groES* mutants, at non-permissive temperature the number of plaques in the presence of recombinant GroES or both P17 and P33

resembled that on the wild type strain. When P17 and P33 were expressed in the *groES* mutant separately, the titer was somewhat lower but still about 2-4 orders of magnitude higher than in the negative control (II, Fig. 2B). Therefore the PRD1 proliferation is restored most efficiently in the *groES619* mutant when both P17 and P33 are plasmid-produced.

The possible association between the fluorescent fusions of P17 and P33 in *E. coli* was analysed by co-sedimentation assay. The co-production of P33-eYFP and P17-eCFP did not change significantly their respective sedimentation velocities (II, Fig. 3A, B and C), indicating that the fluorescent fusion proteins are not in association with each other. However, due to the varying production levels, it was difficult to detect the differences between the sedimentation velocities and therefore solid conclusions about the possible complex formation between P17 and P33 could not be drawn.

Furthermore, it was analysed whether plasmid-produced P33 could complement PRD1 mutant *sus151*, having an amber mutation in gene *XVII*, which encodes P17. The titer of *sus151* was about the same on the strain expressing P33 as on the negative control, whereas six times higher titers were seen with strains expressing P17 alone or together with P33. This suggests, that P33 cannot substitute P17 in PRD1 assembly, and that these two virally encoded proteins have different functions in the virion assembly process.

5.2.3 Possible mechanisms of action of PRD1 P17 and P33 in virus assembly

A similar complementation effect, shown here for PRD1 P17 and P33 (II, Fig. 2A and B), has previously been reported for phage-encoded co-chaperonins, such as T4 Gp31 (van der Vies *et al.* 1994) and RB49 CocO (Ang *et al.* 2001). Despite the lack of significant sequence resemblance with GroES, these virally encoded proteins have been shown to replace it in the bacterial chaperonin complex and function as a co-chaperonin. T4 Gp31 and GroES share similar structural features; importantly, both possess the mobile loop motif known to interact with GroEL (Landry *et al.* 1996, Hunt *et al.* 1997, Richardson and Georgopoulos 1999). A similar mobile stretch with a conserved hydrophobic tripeptide has also been found in RB49 CocO (Ang *et al*. 2001). However, this motif was not found in the sequences of P17 and P33 (II). Therefore it seems rather unlikely that these PRD1 proteins could function as a co-chaperonin. Since no high-resolution structural data is available for P17 and P33, their structures could not be compared to those of previously identified co-chaperonins.

Instead of physically replacing the co-chaperonin, there are alternative ways for virally encoded proteins to modulate the chaperonin complex. For instance, Gp39.2, expressed by phages T4, RB43, RB69 and RB49, suppresses the *groE* defect by enhancing the open conformation of GroEL, subsequently facilitating its interaction with the co-chaperonin (Ang and Georgopoulos 2012). Since the defective phenotype of *groES619* results from an impaired interaction with GroEL (Ang and Georgopoulos 2012), complete replacement of GroES is not necessarily needed for restoring the functionality of the complex. It is possible, that P17 and P33 function in the *groES619* mutant similarly to Gp39.2,

or facilitate the GroEL/GroES complex formation in some other, yet unknown manner.

Hänninen *et al*. (1997) showed that in certain *groE* mutants, including *groES619*, the decrease in the number of PRD1 virions (100-fold reduction) is actually not as substantial as would have been expected from the titer. A substantial amount of non-infectious PRD1 particles are produced, which are largely devoid of membrane-associated infectivity protein P11 and certain small membrane proteins (Hänninen *et al*. 1997). Moreover, previous studies imply that P17 affects neither the oligomerization of the capsid proteins nor the insertion of a number of small membrane proteins to the host membrane (Caldentey *et al*. 1999). This indicates that P17 could act later in the assembly process, possibly during the morphogenesis of the internal membrane-containing procapsids. Although P17 seems to be expressed in soluble form (II, Fig. 3B), its association with the membrane, perhaps in transient manner, cannot be excluded. In addition, purified tetrameric P17 has been shown to attach to positively charged lipids (Holopainen *et al.* 2000). In gram-negative bacteria, without positive lipids, P17 may interact via cationic lipid head group moieties or basic domains of membrane proteins. Also, it has been previously shown that GroEL binds to isolated cytoplasmic membranes of *E. coli* through an interaction with SacA membrane protein, consequently contributing to the secretion and/or composition of the membrane (Bochkareva *et al.* 1998). Therefore, the attachment of the virally encoded proteins to the membrane may be mediated by GroEL, either directly or indirectly via other proteins. So, it is possible that plasmid-production of P17, and possibly P33, could repair the compromised membrane of the mutant cell, for instance by mass action, leading to the observed complementation effect both in bacteriophage proliferation and bacterial colony formation.

Naturally it is beneficial for the phage to encode chaperone-related proteins which enable infection in mutant cells, in which the virus assembly would otherwise be impaired. For bacteriophages it is highly economical to exploit the host-encoded machinery for protein folding. Phages have evolved different ways to modulate the complex for their own benefit. This is exemplified by the T4-specific co-chaperonin Gp31, which enables the folding of the large MCP of T4, which would be difficult to achieve with the volume limits of the host-encoded GroEL-GroES cage (Hunt *et al*. 1997). Therefore it is not surprising that PRD1 has different mechanism in modulating or complementing the function of the host protein folding machinery. In order to analyse whether the complementation effect of P17 and P33 results from a direct interaction with the chaperonin complex, co-localization and FRAP assays were applied.

5.2.4 PRD1 P17 and P33 and *E. coli* GroEL co-localize in the cell

The possible subcellular co-localization between PRD1 proteins P17 and P33 and host chaperonin GroEL was analysed using the fluorescent protein vectors, described in section 5.1.1. In these co-localization assays, the fluorescent fusion

protein P33-eYFP was produced either alone, or simultaneously with the fusion of P17 (P17-eCFP) or GroEL (GroEL-eCFP). When the fluorescent fusion of P33 was produced in the bacterial cell alone, it distributed evenly along the cytosol in all parallel samples (II, Fig. 4A). Previously the same homogenous distribution had been reported for the fluorescent fusions of the monomeric PRD1 host recognition protein P2 (I, Fig. 3C and D; see section 5.1.3). In contrast, P17-eYFP formed distinct fluorescent loci at the cellular poles, as previously described (I, Fig. 3K; II, Fig. 4B; see section 5.1.3). The same polar localization had also been observed for other fusions of PRD1 oligomeric proteins (I, Fig. 3; see section 5.1.3.).

Currently, there does not seem to be any consensus concerning the subcellular localization of GroEL. It has been proposed to be distributed homogenously in the cytosol (Winkler *et al*. 2010, Charbon *et al*. 2011), co-localized with FtsZ at the septum (Ogino *et al.* 2004) or captured into large inclusion bodies (Carrio and Villaverde 2005). In our experimental setup, the plasmid-production of GroEL-eCFP resulted in two different location patterns: it was either evenly distributed throughout the cytosol or localized into the poles (II, Fig. 4C). The sedimentation assay implied that at least a portion of the recombinant GroEL-eCFP was aggregating (II, 3D), which may influence the localization pattern.

Interestingly, in co-production with either P17-eCFP or GroEL-eCFP, the localization pattern of P33-eYFP changed: instead of being homogenously distributed, it accumulated into distinct fluorescent loci at the cellular poles, co-localizing with the respective fluorescent loci of the P17 or GroEL fusion (II, Fig. 4D and E). Similar effect was not seen with eYFP; it remained evenly distributed throughout the cell, despite the presence of the P17 or GroEL fusion (data not shown). This indicates that the change in the localization pattern of P33-eYFP was triggered by the viral protein and not the fluorescent tag. It could be speculated that the localization switch of P33-eYFP resulted from interactions with the fusions of P17 and GroEL. However, this would mean that the endogenous GroEL, expressed constitutively in the cell, was not sufficient to elicit the localization change of P33-eYFP.

5.2.5 The fluorescent fusion of P33 moves slowly in *E. coli*

The diffusion dynamics of the fluorescent fusion protein P33-eYFP, and as a control eYFP, were analysed in *E. coli* by FRAP assay (II, Fig. 5). The FRAP assays were complemented by computer simulations, which were conducted from the recovery curves of the photobleached area. This enabled to determine the diffusion coefficient of the target protein and to estimate the mass of a protein complex including the fluorescent fusion protein. Therefore, the FRAP technique and computer simulations provided an indirect way to analyse the subcellular interactions of P33-eYFP.

Based on the fluorescence recovery curves and computer simulations, the diffusion coefficient of eYFP was determined to be 5 $\mu m^2 s^{-1}$. This correlates with the diffusion coefficients previously determined for GFP [7.7 ± 2.5 $\mu m^2 s^{-1}$,

(Elowitz *et al.* 1999)] and for CheY-GFP [4.6 ± 0.8 $\mu m^2 s^{-1}$, (Cluzel *et al.* 2000)]. The measured diffusion coefficient is in accordance with the theoretically simulated subcellular diffusion of an eYFP-sized molecule. The theoretically estimated diffusion coefficient of P33-eYFP is 4.6 $\mu m^2 s^{-1}$. However, the fluorescence recovery measured from the cells expressing P33-eYFP was reproduced by a simulation, in which 91% of P33-eYFP had a diffusion coefficient of 2 $\mu m^2 s^{-1}$, while the remaining 9% of the total population had a coefficient of 0.4 $\mu m^2 s^{-1}$ (II, Fig. 5C and D). The recovery of P33-eYFP in both of the populations was considerably slower than would have been estimated by a simple mass scaling. The retardation in the diffusion indicates that some cellular factors are hindering its diffusion. Computer simulations of the major population implied that the P33-eYFP complex is approximately 400 kDa. Interestingly, this correlates with the heptameric GroEL ring of 420 kDa. Previously, the diffusion coefficient of the tetradecameric GroEL was determined to be 0.16 $\mu m^2 s^{-1}$ (Charbon *et al*. 2011), which is in the same range with the diffusion coefficient estimated here for the minor population (0.4 $\mu m^2 s^{-1}$). Slow diffusion of P33-eYFP may be explained by an interaction with endogenous GroEL or some other complex.

The deceleration of the P33-eYFP diffusion was especially prominent, when the fusion protein was produced simultaneously with P17-eCFP or GroEL-eCFP (II, Fig. 5C and D). The two-component system of P33-eYFP with the diffusion coefficients of 2 $\mu m^2 s^{-1}$ and 0.4 $\mu m^2 s^{-1}$ applied for the observed data but the ratio between the populations changed. When P33-eYFP was produced together with GroEL-eCFP, the percentage of the slower diffusing population increased from 9% to 40% (II, Fig. 5D). Assuming that P33 and GroEL are in association inside the cell, the plasmid-production of GroEL would increase the number of potential binding partners for P33-eYFP, explaining the slow mobility. Such a significant change is somewhat surprising, considering the abundance of endogenous GroEL in *E. coli*. When P33-eYFP was co-expressed with P17-eYFP, the percentage of the slower population raised to 67% (II, Fig. 5D). One possible explanation is that the interaction between P33-eYFP and the chaperonin complex was enhanced in the presence of P17. This correlates with the complementation assay, in which the PRD1 propagation was restored more efficiently when both P33 and P17 were present.

The data from the FRAP assay and co-localization studies is tentatively suggesting interactions between PRD1 P17 and P33 and the host chaperonin complex. This interaction could explain their observed complementation effect on *groES* mutant. However, no interaction had been detected between P17 and GroEL in ATPase and co-sedimentation assays (Hänninen *et al*., 1997). Additional experiments, using e.g pull down technique, are needed to analyse whether the virally encoded proteins P17 and P33 are physically interacting with the host chaperonin complex, or whether their functions are connected in some indirect way.

5.3 A new freshwater bacteriophage with a tripartite dsRNA genome and an enveloped virion

5.3.1 The virion of φNN resembles those of the *Cystoviridae*

In fall 2009, bacteriophage φNN was isolated together with its host bacterium from a freshwater sample, taken from Lake Vehkalampi in Central Finland (III). Based on 16S rRNA sequence analysis, the host strain B314 belongs to *Pseudomonas* genus and, as further biophysical assays indicate, is probably *Pseudomonas syringae*, or some closely related species (III). Virus particles were purified from large-scale liquid cultures. The highest φNN yield [~3×10^{11} plaque forming units (PFU) ml^{-1}] was achieved, when the *Pseudomonas* sp. B314 culture was infected in the middle of the exponential growth with multiplicity of infection (MOI) of 2-5. Virus purification by rate zonal and equilibrium centrifugations in sucrose gradients yielded high-titer viral specimens (specific infectivity of ~6×10^{12} PFU mg^{-1}), subsequently used in virus characterization.

The chloroform sensitivity of φNN was the first indication of a lipid component in the virion (III). Transmission electron microscopy of the purified virions demonstrated that the tailless φNN particles share the diameter (~80 nm) and round shape with φ6, the type member of the *Cystoviridae* family with enveloped dsRNA viruses (King *et al*. 2012). Moreover, polyhedrally symmetric particles were detected from the samples, resembling the cystovirus cores, from which the envelope had been stripped off (III, Fig. 1B).

The structural protein profile of φNN resembles closely that of φ6 (III, Fig. 1C). Consequently, the protein species were named according to their counterparts in φ6 (P1, P2, P3 etc.), following the established nomenclature among cystoviruses (King *et al*. 2012). Detergent Triton X-100 removed φNN proteins P3, P6 and P9, which indicates that these protein species reside in the outer membrane. Moreover, φNN protein P3 was specifically removed by butylated hydroxytoluene, which is known to remove the P3 spike protein of φ6 without perturbing other membrane-associated proteins (Bamford *et al.* 1995). This further implies that φNN P3 functions as a vertex-associated spike protein, analogously to the host recognition protein P3 of φ6.

The genomic nucleic acid of φNN, extracted from the purified virions, was cleaved by RNase but not by DNase I (digests ssDNA and dsDNA) or Mung Bean nuclease (digests ssDNA or ssRNA). Moreover, when the genome was subjected to agarose gel electrophoresis, it was shown to consist of three segments, the mobilities of which resembled those of φ6 (III, Fig. 1E), implying that φNN has tri-segmented dsRNA genome, similarly to known cystoviruses. Therefore, the φNN genome segments were named as in φ6 based on their sizes as small (S), medium (M) and large (L).

Above described preliminary studies strongly indicate, that φNN is a putative new member of the *Cystoviridae*, a family of enveloped bacteriophages with a segmented dsRNA genome. Over several decades, *Pseudomonas* phage

φ6 has remained the only member of the *Cystoviridae*, which has been officially recognized by the ICTV classification system (King *et al*. 2012). However, since the end of 1990s, additional dsRNA phages have been isolated from different parts of the U.S.A (Mindich *et al.* 1999, Silander *et al.* 2005, O'Keefe *et al.* 2010, Qiao *et al.* 2010), suggesting that this virus type is more abundant in nature than has been previously known. φNN is the first putative cystovirus isolated outside North America, demonstrating a wider global distribution for enveloped dsRNA phages. To my knowledge, all the previously identified putative cystoviruses were obtained from legume plants (Mindich *et al.* 1999, Silander *et al.* 2005, O'Keefe *et al*. 2010, Qiao *et al*. 2010), while φNN was isolated from freshwater sample, pointing to the direction that the virus type may thrive in diverse environments. In earlier studies dsRNA phages were isolated by plating environmental samples on lawns of *Pseudomonas syringae* pv. phaseolicola HB10Y (Vidaver *et al*. 1973), a laboratory host of φ6, or other *Pseudomonas* strains susceptible to previously identified putative cystoviruses (Mindich *et al.* 1999, Silander *et al.* 2005, O'Keefe *et al.* 2010, Qiao *et al.* 2010). This made the sampling of phages biased toward those, which are able to grow under similar culturing conditions and on the same bacterial strains than the previously isolated dsRNA phages. By using an assortment of bacterial strains, one would expect to find more novel dsRNA phages in nature, possibly showing deviating host specificities and habitats.

5.3.2 φNN binds specifically to the isolation host cell

φNN and φ6 are not able to infect each other´s natural bacterial host strains, *Pseudomonas* sp. B314 and *P. syringae* pv. phaseolicola HB10Y, respectively (III). HB10Y has smooth polysaccharide layer in its outer membrane and type IV pili protruding from the cell surface (Mindich *et al*. 1999). It has been previously shown that φ6 uses the pilus structure for infection; it first attaches along the pilus, after which the hair-like protrusion retracts, bringing the virus particle in contact with the cell surface (Bamford *et al.* 1976, Romantschuk and Bamford 1985). A similar infection mechanism has also been reported for cystovirus φ2954 (Qiao *et al*., 2010), while the other distant relatives of φ6 (φ8, φ12 and φ13) enter the host cell directly through the rough lipopolysaccharide layer (Mindich *et al*. 1999, Hoogstraten *et al*. 2000, Qiao *et al*. 2000, Gottlieb *et al*. 2002b). TEM imaging showed φNN virions binding to the lateral sides of thin bacterial protrusions (III, Fig 1F), implying that the virus recognizes the pili, similarly to φ6 and φ2954. However, according to adsorption assays, φNN and φ6 are not able to attach to each other´s host pilus (data not shown), suggesting that their respective recognition sites on the pilus are not the same.

5.3.3 Comparative genomics demonstrating a sequence relatedness between φNN and putative cystoviruses

For φNN genome sequencing, the three segments were first converted into cDNA fragments by reverse transcriptase (III, Fig. 2). The sizes of the genome

segments were 2945, 3814 and 6503 bp, for the S-, M- and L-segments, respectively (III, Fig. 2), being in accordance with those of φ6 genome segments (2948, 4063 and 6374 bp, respectively). However, the terminal regions of the segments turned out to be difficult to copy, and consequently, we were not able to determine the 5´-end of the L-segment, which may include a few additional nucleotides. The GC-contents of the φNN and φ6 are the same (~55 % and ~56 %, respectively; III). All 13 predicted φNN ORFs are oriented in the same direction, as previously shown with other putative cystoviruses (McGraw *et al.* 1986, Gottlieb *et al.* 1988, 2002a,b, Mindich *et al.* 1988, Hoogstraten *et al.* 2000, Qiao *et al.* 2000, 2010). For each φNN ORF, an analogous gene was recognised in the φ6 genome, based on chromosomal position, size and sequence similarity (III, Fig. 2). Overall, the gene organization is conserved among cystoviruses (McGraw *et al.* 1986, Gottlieb *et al.* 1988, 2002a,b, Mindich *et al.* 1988, Hoogstraten *et al.* 2000, Qiao *et al.* 2000, 2010). The φNN ORFs initiate with the codon AUG, except for ORFs 5 and 10, which possess the alternative translation initiation codon GUG. A ribosome-binding site (Shine and Dalgarno 1974) could not be identified for ORFs 2, 5 and 12, indicating that the ribosome-loading and subsequent initiation of translation occur at the upstream ORFs. This kind of polar relationships between genes seem to be common in all the cystovirus genomes (McGraw *et al.* 1986, Gottlieb *et al.* 1988, 2002a,b, Mindich *et al.* 1988, Hoogstraten *et al.* 2000, Qiao *et al.* 2000, 2010).

Among cystoviruses, complete genome sequences are currently available for φ6, φ8, φ12, φ13 and φ2954. The nucleic acid sequence of φNN was compared with other cystovirus sequences, and phylogenetic trees were constructed from these alignments (III, Fig. 3). The φNN genome segments show a strong sequence resemblance to those of φ6, with M-segments displaying the greatest variance (84%, 55 % and 80 % nucleic acid similarity between the S-, M- and L-segments, respectively). The relatedness between φNN and φ6 was even more striking at the amino acid level (III, Table 1; Supplementary Fig. 1). The φNN ORF products of the S- and L-segments were almost identical with the corresponding φ6 ones (89-98 % amino acid similarity), whereas the translated ORFs in the M-segment differed more from their counterparts in φ6 (40-69 % similarity). The lowest amino acid similarity was detected between the spike proteins P3 (40 % similarity; III, Table 1; Supplementary Fig. 1).

Interestingly, a moderate level of sequence resemblance was detected between the L-segments of φNN and φ13 (III, Fig. 3). The relatedness was especially prominent in the ORFs/genes encoding the structural proteins of the nucleocapsid (41-60 % amino acid similarity; III, Table 1). The sequence relatedness between φNN and the previously identified cystoviruses was greatest among the ORFs/genes encoding the internal components of the virion, including essential enzymes and structural proteins. In contrast, the genes for the spike protein P3, and other membrane-associated proteins encoded by the M-segment, differed the most (III, Table 1; Supplementary Fig. 1). The closest relatives of φNN P3 protein are the corresponding proteins of φ6

and φ2954 (40 and 28 % amino acid sequence similarities, respectively). All these phages share the type IV pilus-mediated infection mechanism (Mindich *et al*. 1999, Qiao *et al*. 2010). The spike protein P3 of φNN seems to consist of a single polypeptide (III, Fig. 1C), similarly to the corresponding proteins of φ6 and φ2954 (Gottlieb *et al*. 1988, Qiao *et al*. 2010; III, Table 1). In contrast, φ8, φ12 and φ13 recognise the host by attaching to the rough lipopolysaccharide with a heteromeric complex of two or three polypeptides (Hoogstraten *et al.* 2000, Gottlieb *et al.* 2002b, Qiao *et al.* 2000).

The observed sequence variance between the M-segments of φNN and φ6 may have resulted from a segment reassortment, occurring frequently among cystovirus populations (Onodera *et al.* 2001, Silander *et al.* 2005, O'Keefe *et al.* 2010). Moreover, the M-segment is presumably under a strong selective pressure to co-evolve with the host. It has been shown that the host attachment protein P3 of φ6 is highly susceptible to spontaneous mutations, leading to a fast accumulation of host range mutations (Duffy *et al.* 2006, Ferris *et al.* 2007, Ford *et al.* 2014). It is also possible that φNN has acquired some genetic material into the M-segment via recombination with cellular transcripts or alternatively RNA strands of other viruses co-infecting the same host. However, no homologous sequences for the M-segment were found from the sequence database, outside putative cystovirus ones.

High sequence similarity between φNN and φ6 seems surprising, considering the exceptionally fast evolution of RNA viruses (Domingo and Holland 1997). In addition, the two phages were isolated from different environments (freshwater sample vs. bean plant) on distant continents at almost 40 year interval (Vidaver *et al*. 1973; III). Similar sequence conservation has previously been seen with PRD1-type phages of the *Tectiviridae* family, which were also isolated at different locations over the course of several years (Saren *et al*. 2005; see section 2.7.1). The conservation of PRD1-type viruses is presumed to reflect an optimized genetic organization.

5.3.4 Structural modelling of putative φNN proteins emphasizes the resemblance to cystoviruses

Structural modelling software I-Tasser (Roy *et al.* 2010, 2012) was utilized to predict the three-dimensional structure of each putative φNN protein from its amino acid sequence. The models of the putative P1 MCP, P2 RdRP, P4 packaging NTPase and P5 lytic enzyme (III, Fig. 1D, Table 1) were considered reliable by a ProSA-web quality quality assessment (Sippl 1993, Wiederstein and Sippl 2007). These models were primarily constructed using previously determined cystovirus protein structures as templates. For instance, the model of φNN P1 (III, Fig. 4) was built mainly based on the structure of φ6 P1 MCP [Protein Data Bank (PDB) identifier (PDB id) 4BTG; (Nemecek *et al.* 2013a); 98 % amino acid similarity]. Interestingly, the Homologous Structure Finder (HSF) program (Ravantti *et al.* 2013), detected 600 structurally equivalent residues (covering 78 % of φNN P1) between the P1 proteins of φNN and φ8 [PDBid

4BTP, (El Omari *et al.* 2013a)], even though the two proteins share only 29 % similarity at the amino acid sequence level.

The φNN P2 model (III, Fig. 4), constructed based on φ6 P2 RdRP [PDBid 1HI8, (Butcher *et al.* 2001)] shares 574 structurally equivalent amino acid residues (covering 86 % of φNN P2) with both φ6 P2 (98 % amino acid similarity) and φ12 P2 [(Ren *et al.* 2013); 34 % amino acid similarity]. This correlates with the high level of conservation detected among viral RNA polymerases (Poch *et al.* 1989, Bruenn 2003, Mönttinen *et al.* 2014).

The model of φNN P4 (III, Fig. 4) was built primarily based on φ6 P4 packaging NTPase [PDBid 4BLO; (El Omari *et al.* 2013b)], with which is shows the highest similarity. Interestingly, the φNN P4 model is structurally more similar to the corresponding protein of φ8 [PDBid 4BLQ; (El Omari *et al.* 2013b)] than that of φ12 [PDBid 1W44; (Mancini *et al.* 2004)], even though at sequence level φNN P4 is somewhat more closely related to φ12 P4 than to φ8 P4 (33 % and 29 % amino acid sequence similarities, respectively). The 172 equivalent residues detected between the φNN P4 model and the P4 packaging NTPases of φ6, φ8 and φ12 (covering 52 % of φNN P4), were mapped onto the common catalytic domain. In contrast, the N- and C-terminal domains of the cystovirus packaging NTPases confer functional specificity and varying control mechanisms (El Omari *et al.* 2013b), explaining the structural variance at these sites.

Above described results imply that the internal structural proteins and enzymes of the putative cystoviruses are even more conserved at the structural, than at the sequence level (III, Fig. 4). The conservation of the virion core, containing the RNA polymerization activity, is commonly seen among dsRNA viruses (Luque *et al.* 2010), while more variance is seen at their outermost virion layers (Bamford *et al*. 2002, Poranen and Bamford 2012). The same phenomenon is also seen in host interacting components of other virus systems. For instance influenza viruses are under constant selective pressure by the immune system, presumably eliciting the rapid evolution of the virion glycoproteins (Webster *et al.* 1992).

The genetic and structural similarities among cystoviruses, especially between φNN and φ6, highlight the conservation of viral "self" elements, required to form functional virions (Bamford *et al.* 2002, Bamford 2003, Bamford *et al.* 2005). Simultaneously, the external virion components evolve in order to improve the host interaction or to facilitate adaptation to new environments and/or hosts (Bamford *et al*. 2002, Poranen and Bamford 2012).

5.4 The first ssDNA phage with an internal membrane-containing icosahedral capsid

5.4.1 A new phage-host system from a Finnish lake

Phage FLiP (*Flavobacterium* infecting, lipid-containing phage) and its host strain B330 were isolated from a water sample taken from Lake Jyväsjärvi in Central Finland in fall 2010 (IV). The host strain produced orange- or yellow-pigmented colonies with a spreading growth, which are characteristics of *Flavobacterium* (Bernardet and Bowman 2006). Moreover, the 16S rRNA gene sequence of the strain showed strongest identity with those of a number of *Flavobacterium* sp. strains (~99% nucleotide sequence identity; IV). Consequently, the isolation host strain of FLiP was designated as *Flavobacterium* sp. B330. *Flavobacterium* is a genus of gram-negative, rod-shaped bacteria. These bacteria are widely distributed in aquatic environments and are also found in soil (Bernardet and Bowman 2006). Most species are psychrotolerant, having an optimal growth temperature of 20-30 °C but growing well at 4 °C. Therefore flavobacteria are readily isolated from temperate and polar ecosystems. Certain *Flavobacterium* species are pathogenic for freshwater fish (Bernardet and Bowman 2006) and studies have demonstrated the potential of *Flavobacterium*-specific phages to be used in aquaculture as therapeutic agents against common fish disease (Laanto *et al.* 2011, 2012, Madsen *et al.* 2013).

The binding between the phage and the isolation host occurred quickly, since the majority of the FLiP virions (~70%) had attached to B330 strains in two minutes p.i. (IV, Supplementary Fig. 2). The host range of FLiP was screened by plaque assay using 36 *Flavobacterium* strains previously isolated from Finnish freshwater environments (rivers and lakes) and inland fish farms (Laanto *et al.* 2011). In addition to the isolation host, FLiP was able to induce plaque formation on lawns of environmental *Flavobacterium* sp. strains B114 and B167, implying that its host range is reasonably narrow (IV). The 16S rRNA gene sequence of B330 shows 94% and 96% identity to those of B114 and B167, demonstrating that the strains are close relatives but not identical. No prophage elements were found in the isolation strain by using FLiP-specific primers.

In order to propagate phage FLiP in liquid culture, several different conditions were tested with varying temperature, infection time point, MOI (0.01-25), aeration conditions and nutrient composition. The turbidity of the bacterial culture did not decrease during the culturing periods (IV, Fig. 1C), suggesting that bacterial lysis was not achieved. Also, the phage titer in the bacterial culture increased only slightly (data not shown). It could be speculated that FLiP requires biofilm formation of bacterial cells for efficient infection, and therefore liquid culturing was not optimal for the propagation.

Due to the inefficient virus production in liquid cultures, FLiP was propagated on B330 lawns, after which the phage lysate was filtered and concentrated by polyethylene glycol-NaCl precipitation. The virions were

purified by rate zonal (resulting in "1 x purified" virion) and equilibrium centrifugation ("2 x purified" virions) in sucrose gradients. Sucrose was chosen for the gradient material, as its usage resulted in higher infectious phage yields compared to other tested materials (iodixanol and glycerol). With this method, highly purified, high-titer phage concentrates with specific infectivity of 4.3 x 10^{12} PFU mg^{-1} were achieved.

CsCl gradient was utilized to measure the buoyant density of FLiP. When the 1 x purified virions was subjected to equilibrium centrifugation in CsCl, a diffracting virus zone was detected at the density of 1.21 g ml^{-1}. However, the centrifugation of FLiP virions in CsCl decreased their infectivity. Therefore we can not exclude the possibility that certain virion components had partially detached from the virion, explaining the reduced infectivity. In sucrose, the virus particles equilibrated at the density of 1.18 g ml^{-1} (Fig. 1B). Low buoyant density is characteristic of lipid-containing viruses (Poranen *et al*. 2015). Furthermore, chloroform treatment decreased the infectivity of FLiP by two orders of magnitude, indicating a possible lipid moiety in the virion.

Purified FLiP virions, resuspended in potassium phosphate buffer (pH 7.2) and stored at -6°C for 22 months, had maintained their ability to form plaques on the lawns of the isolation host (IV). In the presence of calcium ions, a slight decay of infectivity was observed.

Based on negative stain TEM, the purified FLiP virions were tailless and icosahedrally symmetric, with extensions protruding from the virion vertices (IV, Fig. 1A). The highly-purified virus preparation was subjected to SDS-PAGE (IV, Fig. 2B) and Tricine-SDS-PAGE analysis (IV, data not shown), after which the gels were stained with lipophilic Sudan Black B and/or protein-binding Coomassie Brilliant Blue, to detect lipid and protein components, respectively. Positive Sudan Black B staining further indicated the presence of lipids in the highly purified phage material (IV, data not shown). Two major protein species were detected in the Coomassie-stained gel, with estimated molecular masses of ~35 kDa and ~12 kDa (IV, Fig. 2B). All detected protein bands were subjected to N-terminal amino acid sequencing and/or mass spectrometry analysis. The obtained peptide sequences were compared to the translated FLiP ORFs to identify the ones encoding structural proteins (Supplementary Table 2; see below).

5.4.2 FLiP shows a limited sequence similarity to any previously identified bacteriophages

The extracted genomic nucleic acid of the purified FLiP virions was degraded by DNase I (digests ssDNA and dsDNA) as well as Mung Bean Nuclease and S1 nuclease (both digest ssDNA and ssRNA), but remained intact after treatment with RNAse A, RNase I (both digest ssRNA) and restriction endonuclease *Eco*RI (digests dsDNA) (IV, Supplementary Fig. 3). This data indicated that the genetic material consists of an ssDNA-molecule. The hypothesis was supported by the similarity between the digestion patterns of

the FLiP genome and the ssDNA genome of microvirus φX174, treated with the same set of nucleases (IV, Supplementary Fig. 3).

We were not able to extract the replicative form of the FLiP genome from infected cells. Therefore, the genomic ssDNA of FLiP was converted into dsDNA fragments, which were subsequently sequenced and manually assembled into the complete genome sequence (IV, Fig. 2A). The determination of the genomic sequence revealed that the genome is a circular molecule of 9174 nt long. It has the GC content of 34 %, correlating with the low GC content commonly seen among flavobacteria (Gupta and Lorenzini 2007). Sixteen ORFs were identified from the genome, all of which were oriented in the same direction. Most of the ORFs did not possess any homologs in the public database, as commonly seen for environmental phage isolates (Seguritan *et al.* 2003, Holmfeldt *et al.* 2013) (IV, Supplementary Table 3). In general, viral genomes can contain a substantial number of ORFans, ORFs which show no detectable sequence similarity with any of the reported sequences in the databases (Yin and Fischer 2008). The lack of homologous sequences for FLiP ORFs may be partially explained by the limited sequence data of phages infecting *Flavobacterium* (Borriss *et al.* 2007, Castillo *et al.* 2014, Luhtanen *et al.* 2014). The closest sequence similarities for FLiP were detected among ssDNA phages of *Cellulophaga* (belonging also to the *Bacteroidetes* phylum) (IV, Supplementary Table 3). These virus isolates have been sequenced and imaged by TEM (Holmfeldt *et al*. 2013), but no higher resolution structural data is available for these phages. Most of the *Cellulophaga* ssDNA phages (e.g. φ12:2, φ12a:1, φ18:4) share genetic similarities with the members of the *Microviridae*, while one of the isolates (φ48:2) is unique among identified bacteriophages (Holmfeldt *et al*. 2013). Also, a few FLiP ORFs showed sequence resemblance to genes of *Zunongwangia profunda*, another genus of *Flavobacteriaceae* in the *Bacteroidetes* phylum.

Genes 7, 8, 9, 11 and 14 were confirmed to encode structural protein products by proteomics (IV, Supplementary Table 2) and are therefore referred to as genes. No significant matches were detected in the database for genes 7 and 11. However, both were predicted to have a transmembrane helix, suggesting presence in the membrane. Translated gene 8 showed a weak amino acid sequence similarity with a structural protein of *Cellulophaga* phage φ48:2 (E-value 0.020) and a hypothetical protein of *Zunongwangia profunda* (E-value 0.030). SDS-PAGE analysis implied a high copy number for the protein product of gene 8, which was therefore proposed to function as the MCP. The closest match for translated gene 9 was to structural proteins of *Cellulophaga* phages φ12a:1, φ12:2 and φ18:4 (E-values 5e-04, 5e-04 and 7e-04, respectively). According to sequence predictions, the protein product of gene 9 has 1-2 transmembrane helices. Gene 14 shares a significant sequence similarity with lytic transglycosylases (e.g. lytic transglycosylase of *Vibrio parahaemolyticus* with E-value of 1e-26). Also, a conserved domain of lytic transglycosylases was detected in its sequence (E-value 4.97e-04). Consequently, it would seem that lytic protein of FLiP is a structural component of the FLiP virion, and could

have a function in the bacterial cell wall penetration upon entry. Interestingly gene 14 product was also predicted to have a transmembrane region. The possible transglycosylase function of the gene product could be confirmed in the future by a zymogram analysis.

Moreover, polypeptide encoded by the longest predicted ORF (ORF15) in the genome, showed a considerable sequence similarity with a hypothetical protein of *Zunongwangia profunda* (E-value 3e-14) as well as a structural protein of *Cellulophaga* phage φ48:2 (E-value 3e-12). Also, a relative level of sequence resemblance was seen with a phage replication protein CRI of *Rivularia sp*. PCC 7116 (E-value 3e-04) and putative phage replication protein CRI of *Aneurinibacillus aneurinilyticus* ATCC 12856 (E-value 0.89).

5.4.3 Cryo-EM imaging reveals structural resemblance between ssDNA phage FLiP and dsDNA phage PM2

The three dimensional structure of FLiP virion was solved at 7.2 Å resolution using cryo-EM imaging and image processing (IV, Fig. 3). The icosahedrally symmetric FLiP virion was shown to contain pentameric extensions (13 nm) at the vertices and an internal membrane inside the protein shell. The distance between opposing facets was 52 nm, while the edge-to-edge and vertex-to-vertex dimensions were 55 nm and 62 nm, respectively. The capsid proteins are organized on T = 21 architecture, previously shown for marine dsDNA phage PM2 (Huiskonen *et al.* 2004). The similarities in the overall virion structure imply a surprising relationship between these phages with different genome types (ssDNA and dsDNA). Viruses have been traditionally classified according to their genome types to ssDNA, dsDNA, ssRNA and dsRNA viruses (King *et al*. 2012). The possible evolutionary relationship between ssDNA phage FLiP and dsDNA phage PM2 would challenge the traditional classification system. This parallels with the recent proposal of a new family Pleolipoviridae, consisting of pleomorphic archaeal viruses with both ssDNA and dsDNA genomes (Pietilä *et al*. 2015). Considering the fact that PM2 has remained the only member of the *Corticoviridae* for almost 50 years (Espejo and Canelo 1968, Oksanen and Bamford 2012), the discovery of FLiP seems interesting. However, PM2-like genetic elements are commonly detected in aquatic bacterial genomes (Krupovič and Bamford 2007), implying that similar phages may occur frequently in aquatic habitats. Discovery of FLiP demonstrates the possibility to find unique virus types even from abundant bacterial types.

6 CONCLUSIONS

The main findings of this thesis are as follows:

I Using the fluorescent fusion protein technique it was demonstrated that the subcellular distribution of fluorescently tagged phage PRD1 proteins in *E. coli* is not uniform: monomeric fusion proteins distributed evenly in the cytosol, while oligomeric fusions were more likely to accumulate in polar loci. The varying localization patterns may reflect a spatially organized viral protein oligomerization and/or PRD1 assembly. This could be confirmed in the future by exploring the localization of a more comprehensive set of virally encoded proteins.

II Two non-structural proteins of PRD1, P17 and P33, complemented a temperature-sensitive mutation of *E. coli* co-chaperonin GroES, restoring both phage multiplication and host growth. This implies that the functions of PRD1 17 and P33 are somehow connected to those of the host chaperonin complex. Furthermore, it was shown that the fluorescent fusion of P33 co-localizes with those of P17 and GroEL at the cellular poles. Also, the diffusion of the P33 fluorescent fusion was slow in *E. coli*, especially when co-expressed with the P17 and GroEL fusions. These findings further suggest associations between the PRD1 non-structural proteins and the host chaperonin complex.

III Enveloped phage φNN was isolated, purified and characterized. It was shown to share high sequence and structural similarities with previously identified cystoviruses, especially the family archetype φ6. φNN is the first putative cystovirus, which was found outside North America and isolated from a freshwater sample. The essential "self" features are preserved among putative cystoviruses, both at sequence and structure level, while

components required for specific host interactions have diversified.

IV FLiP, another novel phage introduced in this thesis, is the first described ssDNA virus with an internal membrane-containing icosahedral capsid. FLiP has only a limited level of sequence resemblance to previously identified viruses, but its overall virion structure is remarkably similar to that of dsDNA phage PM2, suggesting that these phages are related, despite their different genome types and the apparent lack of sequence similarity.

Acknowledgements

The work presented in this thesis was carried out at the University of Jyväskylä at the Department of Biological and Environmental Science between the years 2010-2015. 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), an Academy of Finland personal grant to Professor Jaana Bamford (#251106) and a Finnish Cultural Foundation personal grant to Sari Mäntynen (#00110600).

This thesis project would not have been possible without the encouragement, guidance, and support of a number of people. First, I wish to express my deepest gratitude to my supervisors Professor Jaana Bamford and Docent Hanna Oksanen. Jaana, thank you for allowing me to pursue my dream of earning a PhD degree and guiding my first steps in scientific life. The fact that you have believed in me has meant a lot. It is obvious that you truly care about the people in your group and want to secure the best possible future for all of us, which is truly appreciated. Hanna, you have been such a supportive and inspiring mentor during my Master´s and PhD studies. Your serenity and flawless logical reasoning has saved me countless times. Thank you for being so patient with me, even during my occasional "mini breakdowns". I am grateful to Professor Harri Savilahti for kindly agreeing to act as my opponent. I would also like to extend my sincere thanks to Docent Katri Eskelin and Reader Eugene Makeyev for carefully reviewing the thesis manuscript and for providing extremely insightful and helpful comments.

I want to express my great appreciation to Docents Lotta-Riina Sundberg and Minna Poranen, the members of my follow-up group. Thank you for being so supportive and for giving constructive feedback on my progress toward degree completion. Lotta-Riina is warmly thanked for acting as my custos and for providing invaluable comments and suggestions on the thesis. I have really enjoyed our collaboration. I am thankful to Minna for always finding the time to answer my never ending questions regarding the φNN-project and for letting me benefit from her vast knowledge of cystoviruses.

I wish to thank all my co-authors for fruitful collaboration. I would especially like to acknowledge Jenni Karttunen and Elina Laanto, my brilliant colleagues and dear friends. It has been such a pleasure to work with you! Jenni, you are my sound of sanity. Thank you for always finding the right words to calm me down, when I am starting to overthink and freak out. Together with our third musketeer Outi we have shared some unforgettable moments and I am sure that there are many more to come! Elina, from day one I have shamelessly exploited your kindness and skills in tackling various work-related problems. Our collaboration has always been effortless and I truly value our friendship. I also want to acknowledge Docent Janne Ravantti, my unofficial third supervisor and "scientific godfather". Thank your for backing me up during these last years and giving me that extra push when I needed it.

I am indebted to Petri Papponen for his skilled technical assistance. You have helped me a lot, outside of the department too. I turned to you for instance when I needed help in moving to a new apartment or advice on changing the fluorescent light tube in the bathroom (hey, these things can be tricky! Besides, I prefer fluorescent proteins). Thank you for your friendship!

I had the pleasure to supervise two talented undergraduate students, Annika Kohvakka and Jenni Marjakangas. I really appreciate all your hard work and commitment to our common projects. I will be following your progress with great interest in the future

I gratefully acknowledge all the former and present members of the JB lab. Thank you for creating such a joyful and warm working environment. I have enjoyed every second in your company whether pipetting side-by side in the lab, shedding sweat (occasionally also blood and tears) on the floorball field, or having one or two well-deserved after-work drinks. I feel that I have made many lifelong friends during my stay in the JB lab. Special thanks to Alice, Anni-Maria, Hanna, Nadine, Reetta and Sari, my wonderful office mates over the years. Life of a PhD student certainly has its ups and downs, but I was lucky to have you guys by my side during this journey! You have been the most amazing support group.

Alli, Anri, Merja, Sanna and all my other friends in- and outside the scientific world: Thank you for reminding me that there is more to life than the dissertation! I am also extremely proud of my two godsons, Valtteri and Noel. It is an honor and a privilege to watch you boys grow.

Finally I would like to acknowledge my parents, Anita and Sakari. Words cannot describe how grateful I am for your loving support in everything I do. You are the source of my strength no one could ask for better role models.

YHTEENVETO (RÉSUMÉ IN FINNISH)

Jotain uutta, jotain vanhaa – kalvorakenteen sisältävät bakteriofagit

Virukset ovat pieniä infektiivisiä partikkeleita, jotka tarvitsevat isäntäsolun lisääntyäkseen. Viruspartikkeli koostuu perimäaineksesta, eli DNA:sta tai RNA:sta, sekä sitä ympäröivästä proteiinikuoresta. Lisäksi osalla viruksista on rasvakalvo, joka sijaitsee proteiinikuoren ulko- tai sisäpuolella. Viruksia esiintyy käytännössä kaikkialla missä on elämää, ja niiden määrän arvellaan olevan vähintään kymmenkertainen solullisiin eliöihin verrattuna. Erityisen runsaslukuisen ja geneettisesti monimuotoisen ryhmän muodostavat bakteerivirukset (bakteriofagit tai faagit). Bakteriofagitutkimus on perinteisesti keskittynyt lähinnä hännällisiin, kaksijuosteisen DNA-genomin omaaviin faageihin, joiden on oletettu oleva dominoiva virustyyppi ympäristössä. Viimeaikaisten tutkimusten perusteella näyttäisi kuitenkin siltä, että hännättömien virustyyppien määrää, ja siten ekologista merkitystä, on selvästi aliarvioitu.

Tässä väitöskirjatutkimuksessa keskitytään hännättömiin, kalvorakenteen sisältäviin bakteriofageihin. Väitöskirjan ensimmäisessä osiossa tutkimuskohteena on enterobakteereita infektoiva PRD1. PRD1:n kaksijuosteista DNA-genomia ympäröi ikosahedraalisesti symmetrinen proteiinikuori, jonka sisäpintaa verhoaa rasvakalvo. PRD1-virusta on tutkittu varsin laajasti viimeisten vuosikymmenten aikana, mutta edelleenkään sen viruspartikkelin kokoamismekanismia ei tunneta yksityiskohtaisesti. Tässä väitöskirjaprojektissa muodostettiin fluoresoivien proteiinien vektorikirjasto, jonka avulla voitiin leimata joukko PRD1:n proteiineja ja tutkia niiden sijaintia ja dynamiikkaa *Escherichia coli* (*E. coli*) -bakteerisolussa. Leimattujen virusproteiinien sijoittumisessa bakteerisolussa havaittiin eroavaisuuksia. Monomeeriset virusproteiinit jakautuivat tasaisesti ympäri solun sytoplasmaa, kun taas oligomeeriset proteiinit kohdistuivat useammin solun napoihin. Näiden tulosten perusteella on mahdollista, että bakteerisolun rakenteellinen epäsymmetria ohjaa virusproteiinien oligomerisaatiota ja/tai viruspartikkelien kokoamista.

Väitöskirjatutkimuksessa analysoitiin myös PRD1:n kahden ei-rakenteellisen proteiinin, P17:n ja P33:n, vaikutusta *E. coli* -bakteerin GroEL/GroES-chaperoniinikompleksin toimintaan. Chaperoniinit, sekä niiden toimintaa avustavat co-chaperoniinit, ehkäisevät solussa polypeptidien aggregoitumista ja edistävät niiden laskostumista toiminnallisiksi proteiineiksi. Tässä väitöskirjatyössä osoitettiin P17:n ja P33:n komplementoivan lämpötilasensitiivistä mutaatiota GroES-co-chaperoniinissa, minkä seurauksena faagin monistuminen ja bakteeri-isännän kasvu palautuivat. Fluoresoivien fuusioproteiinien konfokaalimikroskopiatutkimusten perusteella PRD1 P33-proteiini sijaitsee samassa kohdassa solua PRD1 P17- tai *E. coli* GroEL-proteiinien kanssa. Lisäksi fluoresoivan P33-fuusioproteiinin osoitettiin liikkuvan hitaasti bakteerisolussa, erityisesti silloin kun se tuotetaan yhdessä fluoresoivien P17- tai GroEL-fuusioproteiinien kanssa. Edellä olevat havainnnot viittaavat vuorovaikutuk-

seen PRD1:n ei-rakenteellisten proteiinien sekä bakteerin chaperoniinikompleksin välillä.

Väitöskirjatyön toisessa osiossa esitellään kaksi uutta kalvorakenteen sisältävää bakteriofagia. Kumpikin virus on eristetty järvivesinäytteestä Jyväskylässä. *Pseudomonas*-faagi φNN:n viruspartikkelissa on kolme kaksijuosteista RNA-segmenttiä, joita ympäröi kaksi ikosahedraalista proteiinikuorta sekä ulkoinen kalvorakenne. φNN muistuttaa sekä rakenteellisesti että geneettisesti *Cystoviridae*-virusheimon jäseniä, etenkin heimon φ6-mallivirusta. Erityisen konservoituneita cystovirusten keskuudessa ovat ne rakenneproteiinit ja entsyymit, joita tarvitaan toiminnallisten viruspartikkelien muodostamiseen. Vastaavasti isäntäsolun tunnistavissa osissa on suurempia eroavaisuuksia eri cystovirusten välillä. Kaikki tunnetut cystovirukset on eristetty kasvinäytteistä Pohjois-Amerikassa. φNN:n eristäminen jyväskyläläisestä vesistönäytteestä osoittaa, että *Cystoviridae*-virusheimo on monipuolisempi ja laajemmalle levinnyt, kuin on aiemmin tiedetty.

Toinen väitöskirjassa esiteltävä uusi virus on *Flavobacterium*-bakteeria infektoiva FLiP. FLiP:n yksijuosteista DNA-genomia ympäroi ikosahedraalinen proteiinikuori. Proteiinikuoren sisäinen kalvorakenne tekee FLiP:n rakenteesta ainutlaatuisen tunnettujen yksijuosteisten DNA-faagien joukossa. FLiP:llä ei ole merkittävää sekvenssisamankaltaisuutta tunnettuihin viruksiin, mutta sen virusrakenne muistuttaa huomattavasti kaksijuosteisen DNA-genomin omaavan *Pseudoalteromonas*-faagi PM2:n rakennetta. Perinteisesti virukset on luokiteltu niiden nukleiinihappotyypin mukaan. Mahdollinen sukulaisuus FLiP- ja PM2-faagien välillä kuitenkin kyseenalaistaa tämän perinteisen luokittelujärjestelmän.

REFERENCES

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., 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. 2006. Classification of bacteriophages. In: Calendar R. (ed.), The Bacteriophages. 2nd ed., Oxford University Press, New York, USA, pp. 8–16.

Ackermann H.W. & Prangishvili D. 2012. Prokaryote viruses studied by electron microscopy. *Arch Virol* 157: 1843-1849.

Adams M.H. 1959. *Bacteriophages.* Interscience Publishers Inc, New York.

Allan G.M. & Ellis J.A. 2000. Porcine circoviruses: a review. *J Vet Diagn Invest* 12: 3-14.

Almeida J.D. 1963. A classification of virus particles based on morphology. *Can Med Assoc J* 89: 787-798.

Anfinsen C.B. 1973. Principles that govern the folding of protein chains. *Science* 181: 223-230.

Ang D. & Georgopoulos C. 2012. An ORFan no more: the bacteriophage T4 39.2 gene product, NwgI, modulates GroEL chaperone function. *Genetics* 190: 989-1000.

Ang D., Richardson A., Mayer M.P., Keppel F., Krisch H. & Georgopoulos C. 2001. Pseudo-T-even bacteriophage RB49 encodes CocO, a cochaperonin for GroEL, which can substitute for *Escherichia coli*'s GroES and bacteriophage T4's Gp31. *J Biol Chem* 276: 8720-8726.

Angly F.E., Felts B., Breitbart M., Salamon P., Edwards R.A., Carlson C., Chan A.M., Haynes M., Kelley S., Liu H., Mahaffy J.M., Mueller J.E., Nulton J., Olson R., Parsons R., Rayhawk S., Suttle C.A. & Rohwer F. 2006. The marine viromes of four oceanic regions. *PLoS Biol* 4: e368.

Atalay R., Zimmermann A., Wagner M., Borst E., Benz C., Messerle M. & Hengel H. 2002. Identification and expression of human cytomegalovirus transcription units coding for two distinct Fcγ receptor homologs. *J Virol* 76: 8596-8608.

Atanasova N.S., Senčilo A., Pietilä M.K., Roine E., Oksanen H.M. & Bamford D.H. 2015. Comparison of lipid-containing bacterial and archaeal viruses. *Adv Virus Res* 92: 1-61.

Athappilly F.K., Murali R., Rux J.J., Cai Z. & Burnett R.M. 1994. The refined crystal structure of hexon, the major coat protein of adenovirus type 2, at 2.9 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* 35: 235-241.

Bamford D.H., Ojala P.M., Frilander M., Wallin L. & Bamford J.K.H. 1995. Isolation, purification, and function of assembly intermediates and subviral particles of bacteriophage PRD1 and φ6. *Methods Mol Genet* 6: 455-474.

Bamford D. & Mindich L. 1982. Structure of the lipid-containing bacteriophage PRD1: disruption of wild-type and nonsense mutant phage particles with guanidine hydrochloride. *J Virol* 44: 1031-1038.

Bamford D., McGraw T., MacKenzie G. & Mindich L. 1983. Identification of a protein bound to the termini of bacteriophage PRD1 DNA. *J Virol* 47: 311-316.

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

Bamford D.H. & Mindich L. 1980. Electron microscopy of cells infected with nonsense mutants of bacteriophage φ6. *Virology* 107: 222-228.

Bamford D.H., Grimes J.M. & Stuart D.I. 2005. What does structure tell us about virus evolution? *Curr Opin Struct Biol* 15: 655-663.

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

Bamford D.H., Romantschuk M. & Somerharju P.J. 1987. Membrane fusion in prokaryotes: bacteriophage φ6 membrane fuses with the *Pseudomonas syringae* outer membrane. *EMBO J* 6: 1467-1473.

Bamford D.H., Palva E.T. & Lounatmaa K. 1976. Ultrastructure and life cycle of the lipid-containing bacteriophage φ6. *J Gen Virol* 32: 249-259.

Bamford D.H., Rouhiainen L., Takkinen K. & Söderlund H. 1981. Comparison of the lipid-containing bacteriophages PRD1, PR3, PR4, PR5 and L17. *J Gen Virol* 57: 365-373.

Bamford J.K., Hänninen A.L., Pakula T.M., Ojala P.M., Kalkkinen N., Frilander M. & Bamford D.H. 1991. Genome organization of membrane-containing bacteriophage PRD1. *Virology* 183: 658-676.

Bamford J.K.H. & Bamford D.H. 2000. A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to the vertex. *J Virol* 74: 7781-7786.

Bartolomé B., Jubete Y., Martínez E. & de la Cruz F. 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene* 102: 75-78.

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.

Benson S.D., Bamford J.K., Bamford D.H. & Burnett R.M. 2002. The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65 Å resolution. *Acta Crystallogr D Biol Crystallogr* 58: 39-59.

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.

Bernal P., Muñoz-Rojas J., Hurtado A., Ramos J.L. & Segura A. 2007. A *Pseudomonas putida* cardiolipin synthesis mutant exhibits increased sensitivity to drugs related to transport functionality. *Environ Microbiol* 9: 1135-1145.

Bernardet J. & Bowman J.P. 2006. The genus *Flavobacterium.* In: Dworkin M., Falkow S., Rosenberg E., Schleifer K.H. & Stackebrandt E. (eds.), The Prokaryotes. 3rd ed., Springer, New York, USA, pp. 481–531.

Bochkareva E.S., Solovieva M.E. & Girshovich A.S. 1998. Targeting of GroEL to SecA on the cytoplasmic membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A* 95: 478-483.

Borriss M., Lombardot T., Glockner F.O., Becher D., Albrecht D. & Schweder T. 2007. Genome and proteome characterization of the psychrophilic *Flavobacterium* bacteriophage 11b. *Extremophiles* 11: 95-104.

Bos L. 1999. Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philos Trans R Soc Lond B Biol Sci* 354: 675-685.

Botstein D. & Herskowitz I. 1974. Properties of hybrids between *Salmonella* phage P22 and coliphage lambda. *Nature* 251: 584-589.

Boya P., Pauleau A.L., Poncet D., Gonzalez-Polo R.A., Zamzami N. & Kroemer G. 2004. Viral proteins targeting mitochondria: controlling cell death. *Biochim Biophys Acta* 1659: 178-189.

Braig K., Otwinowski Z., Hegde R., Boisvert D.C., Joachimiak A., Horwich A.L. & Sigler P.B. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371: 578-586.

Brass J.M., Higgins C.F., Foley M., Rugman P.A., Birmingham J. & Garland P.B. 1986. Lateral diffusion of proteins in the periplasm of *Escherichia coli*. *J Bacteriol* 165: 787-795.

Breitbart M. & Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* 13: 278-284.

Breitbart M., Miyake J.H. & Rohwer F. 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* 236: 249-256.

Breitbart M., Salamon P., Andresen B., Mahaffy J.M., Segall A.M., Mead D., Azam F. & Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 99: 14250-14255.

Brentlinger K.L., Hafenstein S., Novak C.R., Fane B.A., Borgon R., McKenna R. & Agbandje-McKenna M. 2002. *Microviridae*, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J Bacteriol* 184: 1089-1094.

Briggs J.A., Wilk T. & Fuller S.D. 2003. Do lipid rafts mediate virus assembly and pseudotyping? *J Gen Virol* 84: 757-768.

Bruenn J.A. 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res* 31: 1821-1829.

Brüssow H. & Hendrix R.W. 2002. Phage genomics: small is beautiful. *Cell* 108: 13-16.

Bukau B. & Horwich A.L. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* 92: 351-366.

Butcher S.J., Bamford D.H. & Fuller S.D. 1995. DNA packaging orders the membrane of bacteriophage PRD1. *EMBO J* 14: 6078-6086.

Butcher S.J., Grimes J.M., Makeyev E.V., Bamford D.H. & Stuart D.I. 2001. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 410: 235-240.

Butcher S.J., Dokland T., Ojala P.M., Bamford D.H. & Fuller S.D. 1997. Intermediates in the assembly pathway of the double-stranded RNA virus φ6. *EMBO J* 16: 4477-4487.

Butler P.J. 1999. Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philos Trans R Soc Lond B Biol Sci* 354: 537-550.

Caldentey J. & Bamford D.H. 1992. The lytic enzyme of the *Pseudomonas* phage φ6. Purification and biochemical characterization. *Biochim Biophys Acta* 1159: 44-50.

Caldentey J., Tuma R. & Bamford D.H. 2000. Assembly of bacteriophage PRD1 spike complex: role of the multidomain protein P5. *Biochemistry* 39: 10566-10573.

Caldentey J., Hänninen A.L. & Bamford D.H. 1994. Gene XV of bacteriophage PRD1 encodes a lytic enzyme with muramidase activity. *Eur J Biochem* 225: 341-346.

Caldentey J., Blanco L., Bamford D.H. & Salas M. 1993. *In vitro* replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. *Nucleic Acids Res* 21: 3725-3730.

Caldentey J., Hänninen A.L., Holopainen J.M., Bamford J.K., Kinnunen P.K. & Bamford D.H. 1999. Purification and characterization of the assembly factor P17 of the lipid-containing bacteriophage PRD1. *Eur J Biochem* 260: 549-558.

Calendar R. 2006. *The Bacteriophages, 2nd ed.* Oxford University Press, New York, USA.

Cann A.J. 2005. *Principles of molecular biology, 4th ed.* Elsevier Academic Press, London, UK.

Carrio M.M. & Villaverde A. 2005. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J Bacteriol* 187: 3599-3601.

Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49: 277-300.

Casjens S. & Hendrix R.W. 1988. Control mechanisms in dsDNA bacteriophage assembly. In: Calendar R. (ed.), The Bacteriophages. 1st ed., Plenum Press, New York, USA, pp. 15–91.

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

Castillo D., Espejo R. & Middelboe M. 2014. Genomic structure of bacteriophage 6H and its distribution as prophage in *Flavobacterium psychrophilum* strains. *FEMS Microbiol Lett* 351: 51-58.

Chalfie M., Tu Y., Euskirchen G., Ward W.W. & Prasher D.C. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263: 802-805.

Chao L. 1988. Evolution of sex in RNA viruses. *J Theor Biol* 133: 99-112.

Charbon G., Wang J., Brustad E., Schultz P.G., Horwich A.L., Jacobs-Wagner C. & Chapman E. 2011. Localization of GroEL determined by *in vivo* incorporation of a fluorescent amino acid. *Bioorg Med Chem Lett* 21: 6067-6070.

Chen I., Christie P.J. & Dubnau D. 2005. The ins and outs of DNA transfer in bacteria. *Science* 310: 1456-1460.

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.

Clarke I.N., Cutcliffe L.T., Everson J.S., Garner S.A., Lambden P.R., Pead P.J., Pickett M.A., Brentlinger K.L. & Fane B.A. 2004. Chlamydiaphage Chp2, a skeleton in the φX174 closet: scaffolding protein and procapsid identification. *J Bacteriol* 186: 7571-7574.

Cluzel P., Surette M. & Leibler S. 2000. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* 287: 1652-1655.

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.

Coetzee W.F. & Bekker P.J. 1979. Pilus-specific, lipid-containing bacteriophages PR4 and PR772: comparison of physical characteristics of genomes. *J Gen Virol* 45: 195-200.

Conway J.F., Wikoff W.R., Cheng N., Duda R.L., Hendrix R.W., Johnson J.E. & Steven A.C. 2001. Virus maturation involving large subunit rotations and local refolding. *Science* 292: 744-748.

Coplin D.L., Van Etten J.L., Koski R.K. & Vidaver A.K. 1975. Intermediates in the biosynthesis of double-stranded ribonucleic acids of bacteriophage φ6. *Proc Natl Acad Sci U S A* 72: 849-853.

Coquel A.S., Jacob J.P., Primet M., Demarez A., Dimiccoli M., Julou T., Moisan L., Lindner A.B. & Berry H. 2013. Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLoS Comput Biol* 9: e1003038.

Cornelissen A., Hardies S.C., Shaburova O.V., Krylov V.N., Mattheus W., Kropinski A.M. & Lavigne R. 2012. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse φKZ-related phages. *J Virol* 86: 1844-1852.

Cvirkaitė-Krupovič V., Krupovič M., Daugelavicius R. & Bamford D.H. 2010a. Calcium ion-dependent entry of the membrane-containing bacteriophage PM2 into its *Pseudoalteromonas* host. *Virology* 405: 120-128.

Cvirkaitė-Krupovič V., Poranen M.M. & Bamford D.H. 2010b. Phospholipids act as secondary receptor during the entry of the enveloped, double-stranded RNA bacteriophage φ6. *J Gen Virol* 91: 2116-2120.

Danovaro R., Corinaldesi C., Dell'anno A., Fuhrman J.A., Middelburg J.J., Noble R.T. & Suttle C.A. 2011. Marine viruses and global climate change. *FEMS Microbiol Rev* 35: 993-1034.

de Haas F., Paatero A.O., Mindich L., Bamford D.H. & Fuller S.D. 1999. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage φ6. *J Mol Biol* 294: 357-372.

de Pedro M.A., Quintela J.C., Holtje J.V. & Schwarz H. 1997. Murein segregation in *Escherichia coli*. *J Bacteriol* 179: 2823-2834.

Deich J., Judd E.M., McAdams H.H. & Moerner W.E. 2004. Visualization of the movement of single histidine kinase molecules in live *Caulobacter* cells. *Proc Natl Acad Sci U S A* 101: 15921-15926.

Desnues C., Rodriguez-Brito B., Rayhawk S., Kelley S., Tran T., Haynes M., Liu H., Furlan M., Wegley L., Chau B., Ruan Y., Hall D., Angly F.E., Edwards R.A., Li L., Thurber R.V., Reid R.P., Siefert J., Souza V., Valentine D.L., Swan B.K., Breitbart M. & Rohwer F. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452: 340-343.

Dokland T. 1999. Scaffolding proteins and their role in viral assembly. *Cell Mol Life Sci* 56: 580-603.

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.

Domingo E. & Holland J.J. 1997. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 51: 151-178.

Dröge A., Santos M.A., Stiege A.C., Alonso J.C., Lurz R., Trautner T.A. & Tavares P. 2000. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. *J Mol Biol* 296: 117-132.

Duda R.L., Martincic K., Xie Z. & Hendrix R.W. 1995. Bacteriophage HK97 head assembly. *FEMS Microbiol Rev* 17: 41-46.

Duffy S., Turner P.E. & Burch C.L. 2006. Pleiotropic costs of niche expansion in the RNA bacteriophage φ6. *Genetics* 172: 751-757.

Dyall-Smith M.L. 2011. Genus *Salterprovirus*. In: King A.M.Q., Adams M.J., Carstens E.B. & Lefkowitz E.J. (eds.), *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier Academic Press, San Diego, USA, pp. 183–186.

Ebersbach G., Briegel A., Jensen G.J. & Jacobs-Wagner C. 2008. A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell* 134: 956-968.

Edgar R., Rokney A., Feeney M., Semsey S., Kessel M., Goldberg M.B., Adhya S. & Oppenheim A.B. 2008. Bacteriophage infection is targeted to cellular poles. *Mol Microbiol* 68: 1107-1116.

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. 2013a. Plate tectonics of virus shell assembly and reorganization in phage φ8, a distant relative of mammalian reoviruses. *Structure* 21: 1384-1395.

El Omari K., Meier C., Kainov D., Sutton G., Grimes J.M., Poranen M.M., Bamford D.H., Tuma R., Stuart D.I. & Mancini E.J. 2013b. Tracking in atomic detail the functional specializations in viral RecA helicases that occur during evolution. *Nucleic Acids Res* 41: 9396-9410.

Ellis J. 1987. Proteins as molecular chaperones. *Nature* 328: 378-379.

Ellis R.J. & Minton A.P. 2006. Protein aggregation in crowded environments. *Biol Chem* 387: 485-497.

Elowitz M.B., Surette M.G., Wolf P., Stock J.B. & Leibler S. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *Journal of Bacteriology* 181: 197-203.

Emori Y., Iba H. & Okada Y. 1983. Transcriptional regulation of three double-stranded RNA segments of bacteriophage φ6 *in vitro*. *J Virol* 46: 196-203.

Espejo R.T. & Canelo E.S. 1968. Properties and characterization of the host bacterium of bacteriophage PM2. *J Bacteriol* 95: 1887-1891.

Fane B., Brentlinger K.L., Burch A.D., Chen M., Hafenstein S., Moore E., Novak C.R. & Uchiyama A. 2006. φX174 et al., The *Microviridae*. In: Calendar R. (ed.), The Bacteriophages, Oxford University Press, Oxford, UK, pp. 129–145.

Fayet O., Ziegelhoffer T. & Georgopoulos C. 1989. The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J Bacteriol* 171: 1379-1385.

Feiss M. & Rao V.B. 2012. The bacteriophage DNA packaging machine. *Adv Exp Med Biol* 726: 489-509.

Fenton W.A., Kashi Y., Furtak K. & Horwich A.L. 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* 371: 614-619.

Ferris M.T., Joyce P. & Burch C.L. 2007. High frequency of mutations that expand the host range of an RNA virus. *Genetics* 176: 1013-1022.

Filée J. & Chandler M. 2010. Gene exchange and the origin of giant viruses. *Intervirology* 53: 354-361.

Flint J.S., Enquist L.W., Racaniello V.R. & Skalka A.M. 2008. *Principles of virology*. ASM Press.

Ford B.E., Sun B., Carpino J., Chapler E.S., Ching J., Choi Y., Jhun K., Kim J.D., Lallos G.G., Morgenstern R., Singh S., Theja S. & Dennehy J.J. 2014. Frequency and fitness consequences of bacteriophage φ6 host range mutations. *PLoS One* 9: e113078.

Forterre P. 2010. Defining life: the virus viewpoint. *Orig Life Evol Biosph* 40: 151-160.

Fraenkel-Conrat H. & Williams R.C. 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc Natl Acad Sci U S A* 41: 690-698.

Frilander M. & Bamford D.H. 1995. *In vitro* packaging of the single-stranded RNA genomic precursors of the segmented double-stranded RNA bacteriophage φ6: the three segments modulate each other's packaging efficiency. *J Mol Biol* 246: 418-428.

Gaidelytė A., Cvirkaitė-Krupovič V., Daugelavicius R., Bamford J.K. & Bamford D.H. 2006. The entry mechanism of membrane-containing phage Bam35 infecting *Bacillus thuringiensis*. *J Bacteriol* 188: 5925-5934.

García-del Portillo F., Calvo E., D'Orazio V. & Pucciarelli M.G. 2011. Association of ActA to peptidoglycan revealed by cell wall proteomics of intracellular *Listeria monocytogenes*. *J Biol Chem* 286: 34675-34689.

Garoff H., Hewson R. & Opstelten D.J. 1998. Virus maturation by budding. *Microbiol Mol Biol Rev* 62: 1171-1190.

Georgopoulos C. 1992. The emergence of the chaperone machines. *Trends Biochem Sci* 17: 295-299.

Georgopoulos C.P., Hendrix R.W., Casjens S.R. & Kaiser A.D. 1973. Host participation in bacteriophage lambda head assembly. *J Mol Biol* 76: 45-60.

Gershburg E., Marschall M., Hong K. & Pagano J.S. 2004. Expression and localization of the Epstein-Barr virus-encoded protein kinase. *J Virol* 78: 12140-12146.

Gil-Carton D., Jaakkola S.T., Charo D., Peralta B., Castaño-Díez D., Oksanen H.M., Bamford D.H. & Abrescia N.G. 2015. Insight into the assembly of viruses with vertical single β-barrel major capsid proteins. *Structure* 23: 1866-1877.

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

Gottlieb P., Potgieter C., Wei H. & Toporovsky I. 2002a. Characterization of φ12, a bacteriophage related to φ6: nucleotide sequence of the large double-stranded RNA. *Virology* 295: 266-271.

Gottlieb P., Wei H., Potgieter C. & Toporovsky I. 2002b. Characterization of φ12, a bacteriophage related to φ6: nucleotide sequence of the small and middle double-stranded RNA. *Virology* 293: 118-124.

Gottlieb P., Qiao X., Strassman J., Frilander M. & Mindich L. 1994. Identification of the packaging regions within the genomic RNA segments of bacteriophage φ6. *Virology* 200: 42-47.

Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J., Bamford D.H., Kalkkinen N. & Mindich L. 1988. Nucleotide sequence of the middle dsRNA segment of bacteriophage φ6: placement of the genes of membrane-associated proteins. *Virology* 163: 183-190.

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.

Grahn A.M., Daugelavicius R. & Bamford D.H. 2002. Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. *Mol Microbiol* 46: 1199-1209.

Grahn A.M., Butcher S.L., Bamford J.K.H. & Bamford D.H. 2006. PRD1: Dissecting the genome, structure, and entry. In: Calendar R. (ed.), The Bacteriophages. 2nd ed., Oxford University Press, Oxford, pp. 161-170.

Grahn A.M., Caldentey J., Bamford J.K. & Bamford D.H. 1999. Stable packaging of phage PRD1 DNA requires adsorption protein P2, which binds to the IncP plasmid-encoded conjugative transfer complex. *J Bacteriol* 181: 6689-6696.

Grahn A.M., Bamford J.K., O'Neill M.C. & Bamford D.H. 1994. Functional organization of the bacteriophage PRD1 genome. *J Bacteriol* 176: 3062-3068.

Greenfield D., McEvoy A.L., Shroff H., Crooks G.E., Wingreen N.S., Betzig E. & Liphardt J. 2009. Self-organization of the *Escherichia coli* chemotaxis network imaged with super-resolution light microscopy. *PLoS Biol* 7: e1000137.

Griffin B.A., Adams S.R. & Tsien R.Y. 1998. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281: 269-272.

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

Grimes S., Jardine P.J. & Anderson D. 2002. Bacteriophage φ29 DNA packaging. *Adv Virus Res* 58: 255-294.

Guberman J.M., Fay A., Dworkin J., Wingreen N.S. & Gitai Z. 2008. PSICIC: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution. *PLoS Comput Biol* 4: e1000233.

Guo P.X., Erickson S., Xu W., Olson N., Baker T.S. & Anderson D. 1991. Regulation of the phage φ29 prohead shape and size by the portal vertex. *Virology* 183: 366-373.

Gupta R.S. & Lorenzini E. 2007. Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. *BMC Evol Biol* 7: 71.

Hall N. 2007. Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* 210: 1518-1525.

Hänninen A.L., Bamford D.H. & Bamford J.K. 1997. Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. *Virology* 227: 207-210.

Hantula J. & Bamford D.H. 1988. Chemical crosslinking of bacteriophage φ6 nucleocapsid proteins. *Virology* 165: 482-488.

Hartl F.U. & Hayer-Hartl M. 2009. Converging concepts of protein folding *in vitro* and *in vivo*. *Nat Struct Mol Biol* 16: 574-581.

Hartl F.U. & Hayer-Hartl M. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295: 1852-1858.

Heim R., Prasher D.C. & Tsien R.Y. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* 91: 12501-12504.

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

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.

Henry J.T. & Crosson S. 2013. Chromosome replication and segregation govern the biogenesis and inheritance of inorganic polyphosphate granules. *Mol Biol Cell* 24: 3177-3186.

Hildenbrand Z.L. & Bernal R.A. 2012. Chaperonin-mediated folding of viral proteins. *Adv Exp Med Biol* 726: 307-324.

Hogle J.M., Chow M. & Filman D.J. 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 229: 1358-1365.

Holmfeldt K., Odic D., Sullivan M.B., Middelboe M. & Riemann L. 2012. Cultivated single-stranded DNA phages that infect marine *Bacteroidetes* prove difficult to detect with DNA-binding stains. *Appl Environ Microbiol* 78: 892-894.

Holmfeldt K., Solonenko N., Shah M., Corrier K., Riemann L., Verberkmoes N.C. & Sullivan M.B. 2013. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci U S A* 110: 12798-12803.

Holopainen J.M., Saily M., Caldentey J. & Kinnunen P.K. 2000. The assembly factor P17 from bacteriophage PRD1 interacts with positively charged lipid membranes. *Eur J Biochem* 267: 6231-6238.

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.

Hoogstraten D., Qiao X., Sun Y., Hu A., Onodera S. & Mindich L. 2000. Characterization of φ8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to φ6. *Virology* 272: 218-224.

Hruby D.E. & Franke C.A. 1993. Viral acylproteins: greasing the wheels of assembly. *Trends Microbiol* 1: 20-25.

Huang K.C., Mukhopadhyay R. & Wingreen N.S. 2006. A curvature-mediated mechanism for localization of lipids to bacterial poles. *PLoS Comput Biol* 2: e151.

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 φ6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* 14: 1039-1048.

Huiskonen J.T., Manole V. & Butcher S.J. 2007. Tale of two spikes in bacteriophage PRD1. *Proc Natl Academy Sci* 104: 6666-6671.

Huitema E., Pritchard S., Matteson D., Radhakrishnan S.K. & Viollier P.H. 2006. Bacterial birth scar proteins mark future flagellum assembly site. *Cell* 124: 1025-1037.

Hunt J.F., van der Vies S.M., Henry L. & Deisenhofer J. 1997. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell* 90: 361-371.

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 2015. *Virus Taxonomy: 2014 Release*. http://www.ictvonline.org/virusTaxonomy.asp. viewed 18.9.2015

Inouye S. & Tsuji F.I. 1994. *Aequorea* green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett* 341: 277-280.

Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B., Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A., Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T., Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D., Overbeek R. & Kyrpides N. 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* 423: 87-91.

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.

Jäälinoja H.T., Huiskonen J.T. & Butcher S.J. 2007. Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages φ6 and φ8. *Structure* 15: 157-167.

Jäälinoja 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., 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.

Jakutytė L., Baptista C., São-José C., Daugelavičius R., Carballido-López R. & Tavares P. 2011. Bacteriophage infection in rod-shaped gram-positive

bacteria: evidence for a preferential polar route for phage SPP1 entry in *Bacillus subtilis*. *J Bacteriol* 193: 4893-4903.

Jiang W., Li Z., Zhang Z., Baker M.L., Prevelige P.E.,Jr & Chiu W. 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol* 10: 131-135.

Johnson M.D. & Mindich L. 1994. Plasmid-directed assembly of the lipid-containing membrane of bacteriophage φ6. *J Bacteriol* 176: 4124-4132.

Juuti J.T. & Bamford D.H. 1997. Protein P7 of phage φ6 RNA polymerase complex, acquiring of RNA packaging activity by *in vitro* assembly of the purified protein onto deficient particles. *J Mol Biol* 266: 891-900.

Juuti J.T. & Bamford D.H. 1995. RNA binding, packaging and polymerase activities of the different incomplete polymerase complex particles of dsRNA bacteriophage φ6. *J Mol Biol* 249: 545-554.

Kannaiah S. & Amster-Choder O. 2014. Protein targeting via mRNA in bacteria. *Biochim Biophys Acta* 1843: 1457-1465.

Karhu N.J., Ziedaite G., Bamford D.H. & Bamford J.K. 2007. Efficient DNA packaging of bacteriophage PRD1 requires the unique vertex protein P6. *J Virol* 81: 2970-2979.

Katz G., Wei H., Alimova A., Katz A., Morgan D.G. & Gottlieb P. 2012. Protein P7 of the cystovirus φ6 is located at the three-fold axis of the unexpanded procapsid. *PLoS One* 7: e47489.

Kawai F., Shoda M., Harashima R., Sadaie Y., Hara H. & Matsumoto K. 2004. Cardiolipin domains in *Bacillus subtilis* marburg membranes. *J Bacteriol* 186: 1475-1483.

Keen E.C. 2015. A century of phage research: bacteriophages and the shaping of modern biology. *Bioessays* 37: 6-9.

Keppel F., Rychner M. & Georgopoulos C. 2002. Bacteriophage-encoded cochaperonins can substitute for *Escherichia coli*'s essential GroES protein. *EMBO Rep* 3: 893-898.

Kerner M.J., Naylor D.J., Ishihama Y., Maier T., Chang H.C., Stines A.P., Georgopoulos C., Frishman D., Hayer-Hartl M., Mann M. & Hartl F.U. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122: 209-220.

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.Q., Adams M.J., Carstens E.B. & Lefkowitz E.J. 2012. *Virus Taxonomy, ninth report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego.

King J. & Casjens S. 1974. Catalytic head assembling protein in virus morphogenesis. *Nature* 251: 112-119.

Kirkpatrick C.L. & Viollier P.H. 2011. Poles apart: prokaryotic polar organelles and their spatial regulation. *Cold Spring Harb Perspect Biol* 3: 10.1101/cshperspect.a006809.

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., Männistö R.H., Kalkkinen N. & Bamford D.H. 1999. Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated. *Virology* 262: 364-374.

Kivelä H.M., Madonna S., Krupovič M., Tutino M.L. & Bamford J.K. 2008. Genetics for *Pseudoalteromonas* provides tools to manipulate marine bacterial virus PM2. *J Bacteriol* 190: 1298-1307.

Kivelä H.M., Daugelavicius R., Hankkio R.H., Bamford J.K. & Bamford D.H. 2004. Penetration of membrane-containing double-stranded-DNA bacteriophage PM2 into *Pseudoalteromonas* hosts. *J Bacteriol* 186: 5342-5354.

Kocks C., Hellio R., Gounon P., Ohayon H. & Cossart P. 1993. Polarized distribution of *Listeria monocytogenes* surface protein ActA at the site of directional actin assembly. *J Cell Sci* 105 (Pt 3): 699-710.

Kocks C., Gouin E., Tabouret M., Berche P., Ohayon H. & Cossart P. 1992. *L. monocytogenes*-induced actin assembly requires the *actA* gene product, a surface protein. *Cell* 68: 521-531.

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.

Krupovič M. & Forterre P. 2011. *Microviridae* goes temperate: microvirus-related proviruses reside in the genomes of *Bacteroidetes*. *PLoS One* 6: e19893.

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

Krupovič M. & Bamford D.H. 2008a. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum *Euryarchaeota*. *Virology* 375: 292-300.

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

Krupovič 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.

Krupovič M., Cvirkaitė-Krupovič V. & Bamford D.H. 2008. Identification and functional analysis of the Rz/Rz1-like accessory lysis genes in the membrane-containing bacteriophage PRD1. *Mol Microbiol* 68: 492-503.

Krupovič M., Daugelavicius R. & Bamford D.H. 2007. A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. *Mol Microbiol* 64: 1635-1648.

Krupovič 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.

Ktistakis N.T. & Lang D. 1987. The dodecahedral framework of the bacteriophage φ6 nucleocapsid is composed of protein P1. *J Virol* 61: 2621-2623.

Kurochkina L.P., Semenyuk P.I., Orlov V.N., Robben J., Sykilinda N.N. & Mesyanzhinov V.V. 2012. Expression and functional characterization of the first bacteriophage-encoded chaperonin. *J Virol* 86: 10103-10111.

Kuwada N.J., Traxler B. & Wiggins P.A. 2015. Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Mol Microbiol* 95: 64-79.

Laanto E., Sundberg L.R. & Bamford J.K. 2011. Phage specificity of the freshwater fish pathogen *Flavobacterium columnare*. *Appl Environ Microbiol* 77: 7868-7872.

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

Lacayo C.I., Soneral P.A., Zhu J., Tsuchida M.A., Footer M.J., Soo F.S., Lu Y., Xia Y., Mogilner A. & Theriot J.A. 2012. Choosing orientation: influence of cargo geometry and ActA polarization on actin comet tails. *Mol Biol Cell* 23: 614-629.

Lakowicz J.R. & Szmacinski H. 1993. Fluorescence lifetime-based sensing of pH, Ca2+, K+ and glucose. *Sensors Actuators B: Chem* 11: 133-143.

Laloux G. & Jacobs-Wagner C. 2014. How do bacteria localize proteins to the cell pole? *J Cell Sci* 127: 11-19.

Lam H., Schofield W.B. & Jacobs-Wagner C. 2006. A landmark protein essential for establishing and perpetuating the polarity of a bacterial cell. *Cell* 124: 1011-1023.

Landgraf D., Okumus B., Chien P., Baker T.A. & Paulsson J. 2012. Segregation of molecules at cell division reveals native protein localization. *Nat Methods* 9: 480-482.

Landry S.J., Taher A., Georgopoulos C. & van der Vies S.M. 1996. Interplay of structure and disorder in cochaperonin mobile loops. *Proc Natl Acad Sci U S A* 93: 11622-11627.

Landry S.J., Zeilstra-Ryalls J., Fayet O., Georgopoulos C. & Gierasch L.M. 1993. Characterization of a functionally important mobile domain of GroES. *Nature* 364: 255-258.

Laurinavičius S., Bamford D.H. & Somerharju P. 2007. Transbilayer distribution of phospholipids in bacteriophage membranes. *Biochim Biophys Acta* 1768: 2568-2577.

Laurinavičius S., Käkelä R., Somerharju P. & Bamford D.H. 2004. Phospholipid molecular species profiles of tectiviruses infecting Gram-negative and Gram-positive hosts. *Virology* 322: 328-336.

Law J. 2007. Pinboards and Books: Juxtaposing, Learning and Materiality. In: Kritt D.W. & Winegar L.T. (eds.), Education and Technology: Critical perspectives, possible futures, Lexington Books, UK, pp. 125-150.

Legendre M., Bartoli J., Shmakova L., Jeudy S., Labadie K., Adrait A., Lescot M., Poirot O., Bertaux L., Bruley C., Couté Y., Rivkina E., Abergel C. & Claverie J.M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A* 111: 4274-4279.

Lenarcic R., Halbedel S., Visser L., Shaw M., Wu L.J., Errington J., Marenduzzo D. & Hamoen L.W. 2009. Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J* 28: 2272-2282.

Li G. & Young K.D. 2012. Isolation and identification of new inner membrane-associated proteins that localize to cell poles in *Escherichia coli*. *Mol Microbiol* 84: 276-295.

Li Y., Sergueev K. & Austin S. 2002. The segregation of the *Escherichia coli* origin and terminus of replication. *Mol Microbiol* 46: 985-996.

Lilley B.N. & Ploegh H.L. 2004. A membrane protein required for dislocation of misfolded proteins from the ER. *Nature* 429: 834-840.

Lippincott-Schwartz J. & Patterson G.H. 2003. Development and use of fluorescent protein markers in living cells. *Science* 300: 87-91.

Lissandron V., Terrin A., Collini M., D'alfonso L., Chirico G., Pantano S. & Zaccolo M. 2005. Improvement of a FRET-based indicator for cAMP by linker design and stabilization of donor-acceptor interaction. *J Mol Biol* 354: 546-555.

Littler E. & Oberg B. 2005. Achievements and challenges in antiviral drug discovery. *Antivir Chem Chemother* 16: 155-168.

Lobocka M.B., Rose D.J., Plunkett G.,3rd, Rusin M., Samojedny A., Lehnherr H., Yarmolinsky M.B. & Blattner F.R. 2004. Genome of bacteriophage P1. *J Bacteriol* 186: 7032-7068.

López-Bueno A., Tamames J., Velazquez D., Moya A., Quesada A. & Alcami A. 2009. High diversity of the viral community from an Antarctic lake. *Science* 326: 858-861.

Luhtanen A.M., Eronen-Rasimus E., Kaartokallio H., Rintala J.M., Autio R. & Roine E. 2014. Isolation and characterization of phage-host systems from the Baltic Sea ice. *Extremophiles* 18: 121-130.

Luque D., González J.M., Garriga D., Ghabrial S.A., Havens W.M., Trus B., Verdaguer N., Carrascosa J.L. & Castón J.R. 2010. The T=1 capsid protein of *Penicillium chrysogenum* virus is formed by a repeated helix-rich core indicative of gene duplication. *J Virol* 84: 7256-7266.

Lustig A. & Levine A.J. 1992. One hundred years of virology. *J Virol* 66: 4629-4631.

Lwoff A. & Tournier P. 1966. The classification of viruses. *Annu Rev Microbiol* 20: 45-74.

Madsen L., Bertelsen S.K., Dalsgaard I. & Middelboe M. 2013. Dispersal and survival of *Flavobacterium psychrophilum* phages *in vivo* in rainbow trout

and *in vitro* under laboratory conditions: implications for their use in phage therapy. *Appl Environ Microbiol* 79: 4853-4861.

Maier K.S., Hubich S., Liebhart H., Krauss S., Kuhn A. & Facey S.J. 2008. An amphiphilic region in the cytoplasmic domain of KdpD is recognized by the signal recognition particle and targeted to the *Escherichia coli* membrane. *Mol Microbiol* 68: 1471-1484.

Mancini E.J., Kainov D.E., Grimes J.M., Tuma R., Bamford D.H. & Stuart D.I. 2004. Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* 118: 743-755.

Männistö R.H., Grahn A.M., Bamford D.H. & Bamford J.K. 2003. Transcription of bacteriophage PM2 involves phage-encoded regulators of heterologous origin. *J Bacteriol* 185: 3278-3287.

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.

Margolin W. 2012. The price of tags in protein localization studies. *J Bacteriol* 194: 6369-6371.

Margolin W. 2005. FtsZ and the division of prokaryotic cells and organelles. *Nat Rev Mol Cell Biol* 6: 862-871.

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.

McGraw T., Mindich L. & Frangione B. 1986. Nucleotide sequence of the small double-stranded RNA segment of bacteriophage φ6: novel mechanism of natural translational control. *J Virol* 58: 142-151.

McKenna R., Xia D., Willingmann P., Ilag L.L., Krishnaswamy S., Rossmann M.G., Olson N.H., Baker T.S. & Incardona N.L. 1992. Atomic structure of single-stranded DNA bacteriophage φX174 and its functional implications. *Nature* 355: 137-143.

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

Mileykovskaya E. & Dowhan W. 2000. Visualization of phospholipid domains in *Escherichia coli* by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange. *J Bacteriol* 182: 1172-1175.

Mileykovskaya E., Fishov I., Fu X., Corbin B.D., Margolin W. & Dowhan W. 2003. Effects of phospholipid composition on MinD-membrane interactions *in vitro* and *in vivo*. *J Biol Chem* 278: 22193-22198.

Mindich L. 2012. Packaging in dsRNA viruses. In: Rossmann M.G. & Rao V.B. (eds.), Viral Molecular Machines, Springer USA, pp. 601-608.

Mindich L. 2004. Packaging, replication and recombination of the segmented genome of bacteriophage φ6 and its relatives. *Virus Res* 101: 83-92.

Mindich L. 1999. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage φ6. *Microbiol Mol Biol Rev* 63: 149-160.

Mindich L. 1996. Heterologous recombination in the segmented dsRNA genome of bacteriophage φ6. *Sem Virol* 7: 389–397.

Mindich L. & Davidoff-Abelson R. 1980. The characterization of a 120 S particle formed during φ6 infection. *Virology* 103: 386-391.

Mindich L. & Lehman J. 1979. Cell wall lysin as a component of the bacteriophage φ6 virion. *J Virol* 30: 489-496.

Mindich L., Lehman J. & Huang R. 1979. Temperature-dependent compositional changes in the envelope of φ6. *Virology* 97: 171-176.

Mindich L., Bamford D., McGraw T. & Mackenzie G. 1982a. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. *J Virol* 44: 1021-1030.

Mindich L., Bamford D., Goldthwaite C., Laverty M. & Mackenzie G. 1982b. Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. *J Virol* 44: 1013-1020.

Mindich L., Qiao X., Qiao J., Onodera S., Romantschuk M. & Hoogstraten D. 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. *J Bacteriol* 181: 4505-4508.

Mindich L., Nemhauser I., Gottlieb P., Romantschuk M., Carton J., Frucht S., Strassman J., Bamford D.H. & Kalkkinen N. 1988. Nucleotide sequence of the large double-stranded RNA segment of bacteriophage φ6: genes specifying the viral replicase and transcriptase. *J Virol* 62: 1180-1185.

Mitchell M.S., Matsuzaki S., Imai S. & Rao V.B. 2002. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Res* 30: 4009-4021.

Miyawaki A., Llopis J., Heim R., McCaffery J.M., Adams J.A., Ikura M. & Tsien R.Y. 1997. Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* 388: 882-887.

Mokili J.L., Rohwer F. & Dutilh B.E. 2012. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2: 63-77.

Mönttinen H.A., Ravantti J.J., Stuart D.I. & Poranen M.M. 2014. Automated structural comparisons clarify the phylogeny of the right-hand-shaped polymerases. *Mol Biol Evol* 31: 2741-2752.

Mullineaux C.W., Nenninger A., Ray N. & Robinson C. 2006. Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. *J Bacteriol* 188: 3442-3448.

Muñoz-Espín D., Holguera I., Ballesteros-Plaza D., Carballido-López R. & Salas M. 2010. Viral terminal protein directs early organization of phage DNA replication at the bacterial nucleoid. *Proc Natl Acad Sci U S A* 107: 16548-16553.

Muñoz-Espín D., Daniel R., Kawai Y., Carballido-López R., Castilla-Llorente V., Errington J., Meijer W.J. & Salas M. 2009. The actin-like MreB cytoskeleton organizes viral DNA replication in bacteria. *Proc Natl Acad Sci U S A* .

Nagy E., Pragai B. & Ivanovics G. 1976. Characteristics of phage AP50, an RNA phage containing phospholipids. *J Gen Virol* 32: 129-132.

Naitow H., Tang J., Canady M., Wickner R.B. & Johnson J.E. 2002. L-A 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 R.H. & Tsukihara T. 2003. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* 11: 1227-1238.

Namba K. & Stubbs G. 1986. Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science* 231: 1401-1406.

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.

Nelson D. 2004. Phage taxonomy: we agree to disagree. *J Bacteriol* 186: 7029-7031.

Nemecek D., Plevka P. & Boura E. 2013a. Using cryoEM reconstruction and phase extension to determine crystal structure of bacteriophage φi6 major capsid protein. *Protein J* 32: 635-640.

Nemecek D., Qiao J., Mindich L., Steven A.C. & Heymann J.B. 2012. Packaging accessory protein P7 and polymerase P2 have mutually occluding binding sites inside the bacteriophage φ6 procapsid. *J Virol* 86: 11616-11624.

Nemecek D., Heymann J.B., Qiao J., Mindich L. & Steven A.C. 2010. Cryo-electron tomography of bacteriophage φ6 procapsids shows random occupancy of the binding sites for RNA polymerase and packaging NTPase. *J Struct Biol* 171: 389-396.

Nemecek D., Boura E., Wu W., Cheng N., Plevka P., Qiao J., Mindich L., Heymann J.B., Hurley J.H. & Steven A.C. 2013b. Subunit folds and maturation pathway of a dsRNA virus capsid. *Structure* 21: 1374-1383.

Nevo-Dinur K., Govindarajan S. & Amster-Choder O. 2012. Subcellular localization of RNA and proteins in prokaryotes. *Trends Genet* 28: 314-322.

Nevo-Dinur K., Nussbaum-Shochat A., Ben-Yehuda S. & Amster-Choder O. 2011. Translation-independent localization of mRNA in *E. coli*. *Science* 331: 1081-1084.

Ogino H., Wachi M., Ishii A., Iwai N., Nishida T., Yamada S., Nagai K. & Sugai M. 2004. FtsZ-dependent localization of GroEL protein at possible division sites. *Genes Cells* 9: 765-771.

O'Keefe K.J., Silander O.K., McCreery H., Weinreich D.M., Wright K.M., Chao L., Edwards S.V., Remold S.K. & Turner P.E. 2010. Geographic differences in sexual reassortment in RNA phage. *Evolution* 64: 3010-3023.

Oksanen H., Pietilä M.K., Senčilo 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, Dordrecht: Springer Science+Business Media, pp. 83-105.

Oksanen H.M. & Bamford D.H. 2012a. Family *Tectiviridae*. In: King A.M.Q., Adams M.J., Carstens E.B. & Lefkowitz E.J. (eds.), Virus taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses, Elsevier Academic Press, San Diego, USA, pp. 317-322.

Oksanen H.M. & Bamford D.H. 2012b. Family *Corticoviridae*. In: King A.M.Q., Adams M.J., Carstens E.B. & Lefkowitz E.J. (eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, Elsevier Academic Press, San Diego, USA, pp. 179-182.

Olkkonen V.M. & Bamford D.H. 1987. The nucleocapsid of the lipid-containing double-stranded RNA bacteriophage φ6 contains a protein skeleton consisting of a single polypeptide species. *J Virol* 61: 2362-2367.

Olkkonen V.M., Ojala P.M. & Bamford D.H. 1991. Generation of infectious nucleocapsids by *in vitro* assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage φ6. *J Mol Biol* 218: 569-581.

Olkkonen V.M., Gottlieb P., Strassman J., Qiao X.Y., Bamford D.H. & Mindich L. 1990. *In vitro* assembly of infectious nucleocapsids of bacteriophage φ6: formation of a recombinant double-stranded RNA virus. *Proc Natl Acad Sci U S A* 87: 9173-9177.

Olsen R.H., Siak J.S. & Gray R.H. 1974. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. *J Virol* 14: 689-699.

Onodera S., Sun Y. & Mindich L. 2001. Reverse genetics and recombination in φ8, a dsRNA bacteriophage. *Virology* 286: 113-118.

Ormö M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y. & Remington S.J. 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 273: 1392-1395.

Pace C.N., Shirley B.A., McNutt M. & Gajiwala K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J* 10: 75-83.

Pakula T.M., Caldentey J., Gutierrez C., Olkkonen V.M., Salas M. & Bamford D.H. 1993. Overproduction, purification, and characterization of DNA-binding protein P19 of bacteriophage PRD1. *Gene* 126: 99-104.

Pakula T.M., Caldentey J., Serrano M., Gutierrez C., Hermoso J.M., Salas M. & Bamford D.H. 1990. Characterization of a DNA binding protein of bacteriophage PRD1 involved in DNA replication. *Nucleic Acids Res* 18: 6553-6557.

Parkin D.M. 2006. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 118: 3030-3044.

Peralta B., Gil-Carton D., Castaño-Díez D., Bertin A., Boulogne C., Oksanen H.M., Bamford D.H. & Abrescia N.G. 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biol* 11: e1001667.

Perlmutter J.D. & Hagan M.F. 2015. Mechanisms of virus assembly. *Annu Rev Phys Chem* 66: 217-239.

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

Pietilä M.K., Roine E., Senčilo A., Bamford D.H. & Oksanen H.M. 2015 Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with signle-stranded or double-stranded DNA genomes.Arch Virol.

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.

Plotkin S.A. 2005. Vaccines: past, present and future. *Nat Med* 11: S5-11.

Poch O., Sauvaget I., Delarue M. & Tordo N. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* 8: 3867-3874.

Polisky B. & McCarthy B. 1975. Location of histones on simian virus 40 DNA. *Proc Natl Acad Sci USA* 72: 2895-2899.

Poranen M.M., Bamford D.H. & Oksanen H.M. 2015. Membrane-containing bacteriophages. In: Anonymous Encyclopedia of Life Sciences, John Wiley & Sons, pp. 1-11.

Poranen M.M. & Bamford D.H. 2012. Assembly of large icosahedral double-stranded RNA viruses. *Adv Exp Med Biol* 726: 379-402.

Poranen M.M. & Bamford D.H. 2008. Entry of a segmented dsRNA virus into the bacterial cell. In: Patton J.T. (ed.), Segmented double-stranded RNA viruses, Caister Academic Press, Norfolk, UK, pp. 215-226.

Poranen M.M. & Bamford D.H. 1999. Packaging and replication regulation revealed by chimeric genome segments of double-stranded RNA bacteriophage φ6. *RNA* 5: 446-454.

Poranen M.M., Tuma R. & Bamford D.H. 2005. Assembly of double-stranded RNA bacteriophages. *Adv Virus Res* 64: 15-43.

Poranen M.M., Paatero A.O., Tuma R. & Bamford D.H. 2001. Self-assembly of a viral molecular machine from purified protein and RNA constituents. *Mol Cell* 7: 845-854.

Poranen M.M., Daugelavicius R., Ojala P.M., Hess M.W. & Bamford D.H. 1999. A novel virus-host cell membrane interaction. Membrane voltage-

dependent endocytic-like entry of bacteriophage straight φ6 nucleocapsid. *J Cell Biol* 147: 671-682.

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.

Prasad B.V. & Schmid M.F. 2012. Principles of virus structural organization. *Adv Exp Med Biol* 726: 17-47.

Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G. & Cormier M.J. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111: 229-233.

Putzrath R.M. & Maniloff J. 1977. Growth of an enveloped mycoplasmavirus and establishment of a carrier state. *J Virol* 22: 308-314.

Qiao X., Qiao J. & Mindich L. 1997. An *in vitro* system for the investigation of heterologous RNA recombination. *Virology* 227: 103-110.

Qiao X., Sun Y., Qiao J. & Mindich L. 2008. The role of host protein YajQ in the temporal control of transcription in bacteriophage φ6. *Proc Natl Acad Sci U S A* 105: 15956-15960.

Qiao X., Qiao J., Onodera S. & Mindich L. 2000. Characterization of φ13, a bacteriophage related to φ6 and containing three dsRNA genomic segments. *Virology* 275: 218-224.

Qiao X., Casini G., Qiao J. & Mindich L. 1995. *In vitro* packaging of individual genomic segments of bacteriophage φ6 RNA: serial dependence relationships. *J Virol* 69: 2926-2931.

Qiao X., Sun Y., Qiao J., Di Sanzo F. & Mindich L. 2010. Characterization of φ2954, a newly isolated bacteriophage containing three dsRNA genomic segments. *BMC Microbiol* 10: 55-2180-10-55.

Rafelski S.M. & Theriot J.A. 2006. Mechanism of polarization of *Listeria monocytogenes* surface protein ActA. *Mol Microbiol* 59: 1262-1279.

Ramamurthi K.S. & Losick R. 2009. Negative membrane curvature as a cue for subcellular localization of a bacterial protein. *Proc Natl Acad Sci U S A* 106: 13541-13545.

Ramamurthi K.S., Lecuyer S., Stone H.A. & Losick R. 2009. Geometric cue for protein localization in a bacterium. *Science* 323: 1354-1357.

Ranson N.A., Farr G.W., Roseman A.M., Gowen B., Fenton W.A., Horwich A.L. & Saibil H.R. 2001. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107: 869-879.

Rao V.B. & Feiss M. 2008. The bacteriophage DNA packaging motor. *Annu Rev Genet* 42: 647-681.

Ravantti J., Bamford D. & Stuart D.I. 2013. Automatic comparison and classification of protein structures. *J Struct Biol* 183: 47-56.

Ravantti J.J., Gaidelytė A., Bamford D.H. & Bamford J.K. 2003. Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. *Virology* 313: 401-414.

Ravin V.K. & Shulga M.G. 1970. Evidence for extrachromosomal location of prophage N15. *Virology* 40: 800-807.

Redrejo-Rodríguez M., Muñoz-Espin D., Holguera I., Mencía M. & Salas M. 2013. Nuclear and nucleoid localization are independently conserved functions in bacteriophage terminal proteins. *Mol Microbiol* 90: 858-868.

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

Reits E.A. & Neefjes J.J. 2001. From fixed to FRAP: measuring protein mobility and activity in living cells. *Nat Cell Biol* 3: E145-7.

Ren Z., C Franklin M. & Ghose R. 2013. Structure of the RNA-directed RNA polymerase from the cystovirus φ12. *Proteins* 81: 1479-1484.

Renner L.D. & Weibel D.B. 2011. Cardiolipin microdomains localize to negatively curved regions of *Escherichia coli* membranes. *Proc Natl Acad Sci U S A* 108: 6264-6269.

Richardson A. & Georgopoulos C. 1999. Genetic analysis of the bacteriophage T4-encoded cochaperonin Gp31. *Genetics* 152: 1449-1457.

Ripp S. & Miller R.V. 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* 143: 2065-2070.

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.

Rohwer F. & Edwards R. 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* 184: 4529-4535.

Roine E. & Bamford D.H. 2012. Lipids of archaeal viruses. *Archaea* 2012: 384919.

Romantschuk M. & Bamford D.H. 1985. Function of pili in bacteriophage φ6 penetration. *J Gen Virol* 66 (Pt 11): 2461-2469.

Romantschuk M., Olkkonen V.M. & Bamford D.H. 1988. The nucleocapsid of bacteriophage φ6 penetrates the host cytoplasmic membrane. *EMBO J* 7: 1821-1829.

Romantsov T., Battle A.R., Hendel J.L., Martinac B. & Wood J.M. 2010. Protein localization in *Escherichia coli* cells: comparison of the cytoplasmic membrane proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *J Bacteriol* 192: 912-924.

Rosario K., Nilsson C., Lim Y.W., Ruan Y. & Breitbart M. 2009. Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol* 11: 2806-2820.

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. 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317: 145-153.

Roux S., Krupovič M., Poulet A., Debroas D. & Enault F. 2012. Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. *PLoS One* 7: e40418.

Roy A., Yang J. & Zhang Y. 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 40: W471-7.

Roy A., Kucukural A. & Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725-738.

Rudner D.Z. & Losick R. 2010. Protein subcellular localization in bacteria. *Cold Spring Harb Perspect Biol* 2: a000307.

Rudner D.Z. & Losick R. 2002. A sporulation membrane protein tethers the pro-sigmaK processing enzyme to its inhibitor and dictates its subcellular localization. *Genes Dev* 16: 1007-1018.

Russell J.H. & Keiler K.C. 2007. Peptide signals encode protein localization. *J Bacteriol* 189: 7581-7585.

Rydman P.S. & Bamford D.H. 2003. Identification and mutational analysis of bacteriophage PRD1 holin protein P35. *J Bacteriol* 185: 3795-3803.

Rydman P.S. & Bamford D.H. 2002. The lytic enzyme of bacteriophage PRD1 is associated with the viral membrane. *J Bacteriol* 184: 104-110.

Rydman P.S. & Bamford D.H. 2000. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol Microbiol* 37: 356-363.

Rydman P.S., Bamford J.K. & Bamford D.H. 2001. A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. *J Mol Biol* 313: 785-795.

Rydman P.S., Caldentey J., Butcher S.J., Fuller S.D., Rutten T. & Bamford D.H. 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. *J Mol Biol* 291: 575-587.

Rye H.S., Burston S.G., Fenton W.A., Beechem J.M., Xu Z., Sigler P.B. & Horwich A.L. 1997. Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. *Nature* 388: 792-798.

Saibil H. 2013. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* 14: 630-642.

Sakaki Y., Yamada K., Oshima M. & Oshima T. 1977. Bacteriophage φNS11: a lipid-containing phage of acidophilic thermophilic bacteria. II. Purification and some properties of the phage. *J Biochem* 82: 1451-1456.

San Martin C.S., Burnett R.M., de Haas F., Heinkel R., Rutten T., Fuller S.D., Butcher S.J. & Bamford D.H. 2001. Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. *Structure* 9: 917-930.

Sanger F., Nicklen S. & Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Saren A.M., Ravantti J.J., Benson S.D., Burnett R.M., Paulin L., Bamford D.H. & Bamford J.K. 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. *J Mol Biol* 350: 427-440.

Savilahti H. & Bamford D.H. 1986. Linear DNA replication: inverted terminal repeats of five closely related *Escherichia coli* bacteriophages. *Gene* 49: 199-205.

Savilahti H., Caldentey J., Lundström K., Syväoja J.E. & Bamford D.H. 1991. Overexpression, purification, and characterization of *Escherichia coli* bacteriophage PRD1 DNA polymerase. *In vitro* synthesis of full-length PRD1 DNA with purified proteins. *J Biol Chem* 266: 18737-18744.

Scheu P., Sdorra S., Liao Y.F., Wegner M., Basche T., Unden G. & Erker W. 2008. Polar accumulation of the metabolic sensory histidine kinases DcuS and CitA in *Escherichia coli*. *Microbiology* 154: 2463-2472.

Schuch R., Pelzek A.J., Kan S. & Fischetti V.A. 2010. Prevalence of *Bacillus anthracis*-like organisms and bacteriophages in the intestinal tract of the earthworm *Eisenia fetida*. *Appl Environ Microbiol* 76: 2286-2294.

Seguritan V., Feng I.W., Rohwer F., Swift M. & Segall A.M. 2003. Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. *J Bacteriol* 185: 6434-6447.

Sekar R.B. & Periasamy A. 2003. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol* 160: 629-633.

Sen A., Heymann J.B., Cheng N., Qiao J., Mindich L. & Steven A.C. 2008. Initial location of the RNA-dependent RNA polymerase in the bacteriophage φ6 procapsid determined by cryo-electron microscopy. *J Biol Chem* 283: 12227-12231.

Shapiro L., McAdams H.H. & Losick R. 2009. Why and how bacteria localize proteins. *Science* 326: 1225-1228.

Shepherd C.M., Borelli I.A., Lander G., Natarajan P., Siddavanahalli V., Bajaj C., Johnson J.E., Brooks C.L.,3rd & Reddy V.S. 2006. VIPERdb: a relational database for structural virology. *Nucleic Acids Res* 34: D386-9.

Shih Y.L. & Rothfield L. 2006. The bacterial cytoskeleton. *Microbiol Mol Biol Rev* 70: 729-754.

Shih Y.L., Kawagishi I. & Rothfield L. 2005. The MreB and Min cytoskeletal-like systems play independent roles in prokaryotic polar differentiation. *Mol Microbiol* 58: 917-928.

Shimomura O., Johnson F.H. & Saiga Y. 1962. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* 59: 223-239.

Shimozono S. & Miyawaki A. 2008. Engineering FRET constructs using CFP and YFP. *Methods Cell Biol* 85: 381-393.

Shine J. & Dalgarno L. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* 71: 1342-1346.

Silander O.K., Weinreich D.M., Wright K.M., O'Keefe K.J., Rang C.U., Turner P.E. & Chao L. 2005. Widespread genetic exchange among terrestrial bacteriophages. *Proc Natl Acad Sci U S A* 102: 19009-19014.

Sinclair J.F., Tzagoloff A., Levine D. & Mindich L. 1975. Proteins of bacteriophage φ6. *J Virol* 16: 685-695.

Sippl M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins* 17: 355-362.

Sliusarenko O., Heinritz J., Emonet T. & Jacobs-Wagner C. 2011. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol Microbiol* 80: 612-627.

Sourjik V., Vaknin A., Shimizu T.S. & Berg H.C. 2007. *In vivo* measurement by FRET of pathway activity in bacterial chemotaxis. *Methods Enzymol* 423: 365-391.

Srinivasiah S., Bhavsar J., Thapar K., Liles M., Schoenfeld T. & Wommack K.E. 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol* 159: 349-357.

Stanisich V.A. 1974. The properties and host range of male-specific bacteriophages of *Pseudomonas aeruginosa*. *J Gen Microbiol* 84: 332-342.

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.

Steven A.C., Heymann J.B., Cheng N., Trus B.L. & Conway J.F. 2005. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr Opin Struct Biol* 15: 227-236.

Steward G.F. 2001. Fingerprinting viral assemblages by pulsed field gel electrophoresis (PFGE). In: Anonymous *Marine Microbiology. Methods in Microbiology Series*, Academic Press, London, UK, pp. 85-103.

Stewart P.L., Burnett R.M., Cyrklaff M. & Fuller S.D. 1991. Image reconstruction reveals the complex molecular organization of adenovirus. *Cell* 67: 145-154.

Stockley P.G., Rolfsson O., Thompson G.S., Basnak G., Francese S., Stonehouse N.J., Homans S.W. & Ashcroft A.E. 2007. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol* 369: 541-552.

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.

Strömsten N.J., Bamford D.H. & Bamford J.K. 2003a. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *J Virol* 77: 6314-6321.

Strömsten N.J., Benson S.D., Burnett R.M., Bamford D.H. & Bamford J.K. 2003b. The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state. *J Bacteriol* 185: 6985-6989.

Summers W.C. 2011. In the beginning.. *Bacteriophage* 1: 50-51.

Suomalainen M. 2002. Lipid rafts and assembly of enveloped viruses. *Traffic* 3: 705-709.

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

Swulius M.T. & Jensen G.J. 2012. The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-terminal yellow fluorescent protein tag. *J Bacteriol* 194: 6382-6386.

Taubenberger J.K. & Morens D.M. 2006. 1918 Influenza: the mother of all pandemics. *Emerg Infect Dis* 12: 15-22.

Thiem S. & Sourjik V. 2008. Stochastic assembly of chemoreceptor clusters in *Escherichia coli*. *Mol Microbiol* 68: 1228-1236.

Thiem S., Kentner D. & Sourjik V. 2007. Positioning of chemosensory clusters in *E. coli* and its relation to cell division. *EMBO J* 26: 1615-1623.

Tomaru Y. & Nagasaki K. 2007. Flow cytometric detection and enumeration of DNA and RNA viruses infecting marine eukaryotic microalgae. *J Oceanogr* 63: 215-221.

Tucker K.P., Parsons R., Symonds E.M. & Breitbart M. 2011. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J* 5: 822-830.

Turner P.E., Burch C.L., Hanley K.A. & Chao L. 1999. Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage φ6. *J Virol* 73: 2420-2424.

van der Vies S.M., Gatenby A.A. & Georgopoulos C. 1994. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature* 368: 654-656.

Van Etten J.V., Lane L., Gonzalez C., Partridge J. & Vidaver A. 1976. Comparative properties of bacteriophage φ6 and φ6 nucleocapsid. *J Virol* 18: 652-658.

van Raaij M.J., Mitraki A., Lavigne G. & Cusack S. 1999. A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* 401: 935-938.

van Teeffelen S., Shaevitz J.W. & Gitai Z. 2012. Image analysis in fluorescence microscopy: bacterial dynamics as a case study. *Bioessays* 34: 427-436.

Vats P. & Rothfield L. 2007. Duplication and segregation of the actin (MreB) cytoskeleton during the prokaryotic cell cycle. *Proc Natl Acad Sci U S A* 104: 17795-17800.

Verheust C., Fornelos N. & Mahillon J. 2005. GIL16, a new gram-positive tectiviral phage related to the *Bacillus thuringiensis* GIL01 and the *Bacillus cereus* pBClin15 elements. *J Bacteriol* 187: 1966-1973.

Verheust C., Jensen G. & Mahillon J. 2003. pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar israelensis. *Microbiology* 149: 2083-2092.

Vidaver A.K., Koski R.K. & Van Etten J.L. 1973. Bacteriophage φ6: a lipid-containing virus of *Pseudomonas phaseolicola*. *J Virol* 11: 799-805.

Vilen H., Aalto J.M., Kassinen A., Paulin L. & Savilahti H. 2003. A direct transposon insertion tool for modification and functional analysis of viral genomes. *J Virol* 77: 123-134.

Walker J.E., Saraste M., Runswick M.J. & Gay N.J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* 1: 945-951.

Wang S., Furchtgott L., Huang K.C. & Shaevitz J.W. 2012. Helical insertion of peptidoglycan produces chiral ordering of the bacterial cell wall. *Proc Natl Acad Sci U S A* 109: E595-604.

Wang X., Montero Llopis P. & Rudner D.Z. 2013. Organization and segregation of bacterial chromosomes. *Nat Rev Genet* 14: 191-203.

Webster R.E. & Cashman J.S. 1978. Morphogenesis of the filamentous single-stranded DNA phages. In: Denhardt D.T., Dressler D. & Ray D.S. (eds.), The Single-stranded DNA phages, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 557–569.

Webster R.G., Bean W.J., Gorman O.T., Chambers T.M. & Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol Rev* 56: 152-179.

Weinbauer M.G. 2004. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28: 127-181.

Wiederstein M. & Sippl M.J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35: W407-10.

Wiertz E.J., Jones T.R., Sun L., Bogyo M., Geuze H.J. & Ploegh H.L. 1996. The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* 84: 769-779.

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.

Winkler J., Seybert A., Konig L., Pruggnaller S., Haselmann U., Sourjik V., Weiss M., Frangakis A.S., Mogk A. & Bukau B. 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J* 29: 910-923.

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.

World Health Organization (WHO). 2015. *Middle East respiratory syndrome coronavirus*. http://www.who.int/emergencies/mers-cov/en/ viewed on 18.9.15.

Xu L., Benson S.D., Butcher S.J., Bamford D.H. & Burnett R.M. 2003. The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. *Structure* 11: 309-322.

Xu Z. & Sigler P.B. 1998. GroEL/GroES: structure and function of a two-stroke folding machine. *J Struct Biol* 124: 129-141.

Xu Z., Horwich A.L. & Sigler P.B. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 388: 741-750.

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.

Yeates T.O., Kerfeld C.A., Heinhorst S., Cannon G.C. & Shively J.M. 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol* 6: 681-691.

Yin Y. & Fischer D. 2008. Identification and investigation of ORFans in the viral world. *BMC Genomics* 9: 24-2164-9-24.

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.

Ziedaite G., Kivelä H.M., Bamford J.K. & Bamford D.H. 2009. Purified membrane-containing procapsids of bacteriophage PRD1 package the viral genome. *J Mol Biol* 386: 637-647.

Ziedaite G., Daugelavicius R., Bamford J.K. & Bamford D.H. 2005. The holin protein of bacteriophage PRD1 forms a pore for small-molecule and endolysin translocation. *J Bacteriol* 187: 5397-5405.

Zimmerman S.B. & Trach S.O. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* 222: 599-620.

Zubieta C., Schoehn G., Chroboczek J. & Cusack S. 2005. The structure of the human adenovirus 2 penton. *Mol Cell* 17: 121-135.

Zweig M. & Cummings D.J. 1973. Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein. *J Mol Biol* 80: 505-518.

ORIGINAL PAPERS

I

SUBCELLULAR LOCALIZATION OF BACTERIOPHAGE PRD1 PROTEINS IN *ESCHERICHIA COLI*

by

Jenni Karttunen, Sari Mäntynen, Teemu O. Ihalainen, Heli Lehtivuori, Nikolai V. Tkachenko, Maija Vihinen-Ranta, Janne A. Ihalainen, Jaana K.H. Bamford & Hanna M. Oksanen 2014.

Virus Research 179: 44-52.

Virus Research 179 (2014) 44–52

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Subcellular localization of bacteriophage PRD1 proteins in *Escherichia coli*

Jenni Karttunen[a], Sari Mäntynen[a], Teemu O. Ihalainen[b], Heli Lehtivuori[b], Nikolai V. Tkachenko[c], Maija Vihinen-Ranta[b], Janne A. Ihalainen[b], Jaana K.H. Bamford[a], Hanna M. Oksanen[d,*]

[a] *Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science and Nanoscience Center, P.O. Box 35, 40014 University of Jyväskylä, Finland*
[b] *Nanoscience Center, Department of Biological and Environmental Science, P.O. Box 35, 40014 University of Jyväskylä, Finland*
[c] *Department of Chemistry and Bioengineering, Tampere University of Technology, P.O. Box 541, 33101 Tampere, Finland*
[d] *Institute of Biotechnology and Department of Biosciences, P.O. Box 56, 00014 University of Helsinki, Finland*

ARTICLE INFO

Article history:
Received 17 September 2013
Received in revised form
19 November 2013
Accepted 19 November 2013
Available online 28 November 2013

Keywords:
Membrane virus
Confocal microscopy
Protein interactions
Virus assembly
Bacteria

ABSTRACT

Bacteria possess an intricate internal organization resembling that of the eukaryotes. The complexity is especially prominent at the bacterial cell poles, which are also known to be the preferable sites for some bacteriophages to infect. Bacteriophage PRD1 is a well-known model serving as an ideal system to study structures and functions of icosahedral internal membrane-containing viruses. Our aim was to analyze the localization and interactions of individual PRD1 proteins in its native host *Escherichia coli*. This was accomplished by constructing a vector library for production of fluorescent fusion proteins. Analysis of solubility and multimericity of the fusion proteins, as well as their localization in living cells by confocal microscopy, indicated that multimeric PRD1 proteins were prone to localize in the cell poles. Furthermore, PRD1 spike complex proteins P5 and P31, as fusion proteins, were shown to be functional in the virion assembly. In addition, they were shown to co-localize in the specific polar area of the cells, which might have a role in the multimerization and formation of viral protein complexes.

1. Introduction

Bacterial cells have been mainly regarded as amorphous reaction vessels concealing a homogenous solution of proteins. Due to advances in bacterial cell biology, this traditional view has changed dramatically. Similarly to eukaryotes, bacteria deploy macromolecules such as proteins, lipids and nucleic acids into specific subcellular locations. This asymmetric architecture is spatially and temporally dynamic, enabling cells to respond to changing demands during their life cycle (Rudner and Losick, 2010). Accumulated data on bacterial proteins have revealed a variety of localization patterns (Amster-Choder, 2011). Whereas certain proteins oscillate from pole to pole (Gerdes et al., 2010; Leonardy et al., 2010; Loose et al., 2011), others form clusters on the bacterial cell surface or at specific sub-cellular locations (Amster-Choder, 2011). In addition, it is known that certain bacterial proteins, especially cytoskeletal, assemble into helical structures extending along the cell or construct ring-like structures at the mid-cell position (Vats et al., 2009). However, interpretation of the localization pattern data with fluorescent tagged proteins has been challenging and some artifacts have emerged (Swulius and Jensen, 2012).

Recent studies have elucidated factors governing the asymmetric protein distribution in bacteria, which is presumably most commonly mediated by 'diffusion and capture', when proteins diffuse freely until interacting with other, so-called target proteins (Deich et al., 2004; Rudner and Losick, 2002). This raises the question about the primary factors directing the target proteins to their specific cellular sites, and emphasizes the need to reveal other mechanisms for protein targeting. For instance, self-assembly is a unique variation of the 'diffusion and capture' positioning, which does not require any pre-existing anchor structures. Cellular factors such as geometric cues and physical constrictions have a role in positioning a number of bacterial proteins into their specific intracellular sites (Rudner and Losick, 2010). Also localization signals can be found in certain bacterial proteins, similarly to their eukaryotic analogs (Russell and Keiler, 2007). Correspondingly, there is evidence of subcellular localization of certain mRNA transcripts correlating with the localization of their protein products (Nevo-Dinur et al., 2011). Yet another mechanism for protein positioning was brought up by the discovery of cytoskeletal proteins

* Corresponding author at: Institute of Biotechnology and Department of Biosciences, Biocenter 2, P.O. Box 56 (Viikinkaari 5), FIN-00014, University of Helsinki, Helsinki, Finland. Tel.: +358 9 191 59104; fax: +358 9 19159098.
E-mail address: hanna.oksanen@helsinki.fi (H.M. Oksanen).

0168-1702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2013.11.015

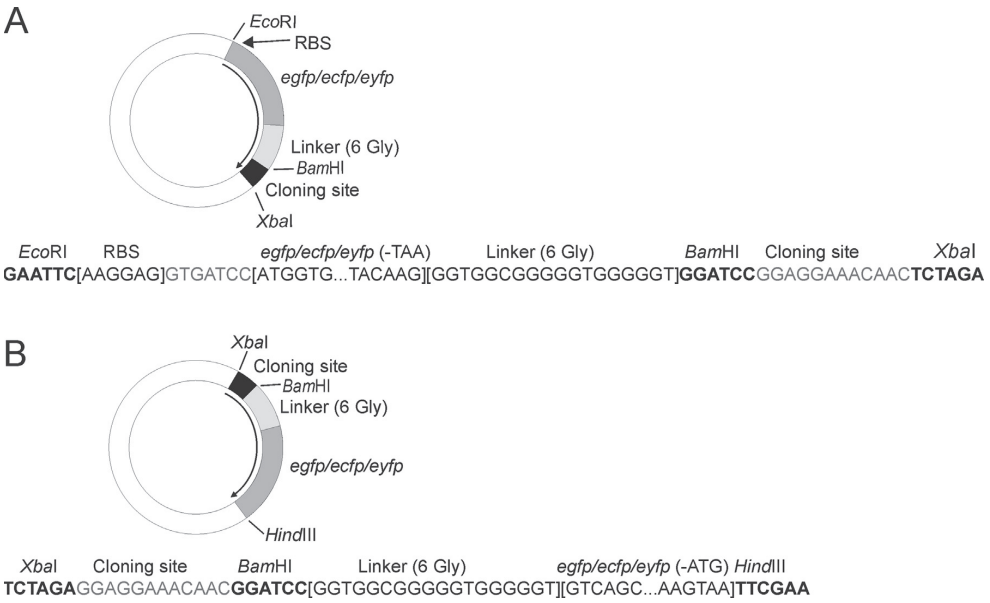


Fig. 1. Vector library. Vectors and sequence of mutated area shown by arrow for production of fluorescent fusion proteins. (A) Vectors for production of N-terminal fusion proteins containing *egfp* (pKM57 and pJK22), *ecfp* (pKM51 and pJK24) or *eyfp* (pKM54 and pJK28). Genes for fluorescent proteins are indicated by dark gray. The cloning site (*Bam*HI-*Xba*I) and the glysine linker are shown by black and light gray, respectively. (B) Vectors for C-terminal fusion proteins containing *egfp* (pKM47 and pKM67), *ecfp* (pKM41 and pKM61) or *eyfp* (pKM44 and pKM64). Colors are like in (A). See also Table 1.

in bacteria (Vats et al., 2009). These structures are also suggested to provide a track for other proteins to locate, resembling again the situation in eukaryotes (Nevo-Dinur et al., 2012). However, the underlying principles of targeting for most bacterial proteins remain elusive.

The studies indicate that the asymmetric protein distribution is particularly conspicuous at the chemically and physically unique cell poles (Lai et al., 2004). For instance, unequally distributed lipid composition and negative curvature of the membrane (Ramamurthi, 2010) are believed to contribute to encompassing proteins into these cellular areas (Nevo-Dinur et al., 2012). It has also been shown that a number of bacteriophages infect preferably at these extreme regions. These viruses bind to distinct cellular receptors on the surface of their Gram negative hosts such as *Escherichia coli*, *Yersinia pseudotuberculosis*, or *Vibrio cholera* (Edgar et al., 2008; Rothenberg et al., 2011) or Gram positive ones such as *Bacillus subtilis* (Jakutyte et al., 2011). This implies that the cell poles contain cellular components essential for DNA intake (Edgar et al., 2008). The hypothesis is supported by the fact that the poles are the preferred site of DNA intake in natural competent cells (Chen et al., 2005; Edgar et al., 2008). Moreover, in studies of *B. subtilis* infecting phage SPP1 (Jakutyte et al., 2011) and *E. coli* phage lambda (Rothenberg et al., 2011) it was observed that in addition to being injected, viral DNA is replicated at the poles. However, several proteins of replication machinery of bacteriophage ϕ29 infecting *B. subtilis* have been found localized in helix-like pattern near the membrane. It was also shown that the ϕ29 replication is dependent on cytoskeleton protein MreB, as also with phage PRD1 (Muñoz-Espín et al., 2009). Terminal protein of these phages has been shown to associate with bacterial nucleoid independently of other phage-coded proteins as well as localize in the nucleus of mammalian cells (Muñoz-Espín et al., 2010; Redrejo-Rodriguez et al., 2012). It seems that bacteriophages have evolved to exploit the internal asymmetry of their host cells in order to make the infection process more efficient.

One of the most extensively studied bacteriophages is Enterobacteria phage PRD1 (family: *Tectiviridae*, genus: *Tectivirus*), which infects various Gram negative bacteria, such as *E. coli* and *Salmonella typhimurium*, carrying P-, W- or N-type conjugative plasmid, whereas other tectiviruses can also infect Gram positive bacteria such as *Bacillus* (Grahn et al., 2006; Oksanen and Bamford, 2012). The PRD1 virion is formed by an icosahedral protein capsid surrounding a protein-rich membrane which, in turn, encloses the linear dsDNA genome (Abrescia et al., 2004; Cockburn et al., 2004; Olsen et al., 1974). Based on X-ray crystallographic analyses, PRD1 belongs to a certain structure based viral lineage with several other viruses such as: adenovirus, *Paramecium bursaria* chlorella virus 1 (PBCV-1) and *Sulfolobus* turreted icosahedral virus (STIV) (Abrescia et al., 2012; Benson et al., 1999; Khayat et al., 2005; Nandhagopal et al., 2002). All these viruses have a major capsid protein with a double β-barrel fold and similar virion architecture. In PRD1, the capsid is mainly composed of the major capsid protein P3, which forms pseudo-hexameric trimers (Abrescia et al., 2004; Benson et al., 1999). The receptor binding spike complex at the virion vertices contains the pentameric penton protein P31 forming the base structure from which the trimeric spike protein P5 and the monomeric receptor binding protein P2 protrude (Caldentey et al., 2000; Merckel et al., 2005; Rydman et al., 1999; Xu et al., 2003). The spike structure complex is stabilized by the integral membrane protein P16 linking the vertex to the underlying viral membrane (Jaatinen et al., 2004). In addition to several other PRD1 structural proteins, also a number of non-structural proteins have been identified, such as the tetrameric assembly protein P17 required for virion formation (Caldentey et al., 1999; Holopainen et al., 2000; Mindich et al., 1982). Despite the intensive structural and functional characterization, the interactions of a number of predicted PRD1 proteins are yet to be revealed.

Our aim was to analyze the localization and interactions of PRD1 proteins in its native host bacterium *E. coli*. The study included viral monomeric and multimeric structural proteins, an integral

membrane protein and a soluble assembly protein. Special attention was paid to the receptor binding spike complex proteins P5 and P31 for which the structures at atomic resolution are known (Abrescia et al., 2004; Caldentey et al., 2000; Rydman et al., 1999). We localized the proteins in living cells by exploiting fluorescent fusion protein technology and confocal microscopy.

2. Results and discussion

2.1. Construction of bacterial expression vector library to produce fluorescent fusion proteins

We created a bacterial vector library for convenient production of fluorescent fusion proteins (Fig. 1; Table 1). The vectors were constructed by cloning genes encoding eGFP and its cyan and yellow variants eCFP and eYFP. We used two bacterial vectors pSU18 and pET24 bearing replicons p15A and ColE1, respectively, enabling simultaneous expression of two proteins. According to confocal microscopy the expression of the fluorescent protein genes in bacterial cells produced functional proteins (for eYFP see Section 2.3, for eGFP and eCFP data not shown). Using these vectors, it is possible to insert any gene of interest into either end of the fluorescent protein gene, thereby creating N-terminal fluorescent fusion protein (the fluorescent protein is linked to the N-terminus of the target protein) or C-terminal fluorescent fusion protein (the fluorescent protein is linked to the C-terminus of the protein). A linker sequence of six glycines was designed to these vectors to separate the fluorescent protein from the protein of interest reducing steric hindrance. In this study, we exploited the vector library to create cyan and/or yellow fluorescent fusion proteins of bacteriophage PRD1 proteins P2 (receptor binding protein), P3 (major capsid protein), P5 (spike protein), P16 (vertex stabilizing integral membrane protein), P17 (non-structural assembly protein), and P31 (penton protein) (Table 1). Genes were cloned into both vector types (Fig. 1) using either pSU18 or pET24 to produce both N-terminal and C-terminal fusion proteins, except gene *XVII*, for which only fusion P17-eYFP was available. Sequencing of the vectors revealed only minor changes in PRD1 genes (Supplementary Table S1).

2.2. Solubility and multimericity of viral fluorescent fusion proteins

In the fusion protein studies, the first concern is whether the fusion affects on the folding and functionality of the native protein. One way to evaluate this is to monitor changes in the protein solubility and find out whether the known multimeric proteins form multimers with fluorescent protein tags. The majority of the fusion proteins (P2, P3, P5, P17 and P31) were expressed as soluble (data not shown). These proteins were directed to sedimentation assay by a rate zonal centrifugation for the size determination.

Monomeric receptor binding protein P2 (Grahn et al., 1999; Xu et al., 2003) was expressed as a fusion protein in a monomeric form (Fig. 2A). Small fraction of smaller side-product was detected with both P2 fusions. More variation in the molecular mass distribution was detected with proteins, which can be released as multimers from the virion. The individual PRD1 spikes composed of the trimeric protein P5 form an elongated structure (Bamford and Bamford, 2000; Caldentey et al., 2000; Huiskonen et al., 2007; Merckel et al., 2005), and there is no obvious reason that the fluorescent tag at the C-terminus of the protein would interfere the folding. The N-terminal fusion protein eYFP-P5 (~61 kDa) sedimented as a monomer (Fig. 2B), but also a smaller multimeric side product (~45 kDa) was detected by an antibody against P5 (data not shown). The C-terminal fusion protein P5-eYFP showed two separate peaks in the sedimentation assay indicating that the protein was in two different forms (monomer and multimer) (Fig. 2B). It also had a smaller P5-specific side-product (~55 kDa) in fractions representing monomeric and multimeric proteins (data not shown). The rate zonal centrifugation indicated that the C-terminal fusion of the penton protein P31 was mainly monomeric whereas the fluorescent protein attached to the N-terminus of P31 formed larger multimers (Fig. 2C). This correlates well with the known P31 X-ray structure as its C-termini are located in the middle of the pentamer (Abrescia et al., 2004) and therefore the added fluorescent tag can hinder the formation of the multimeric complex. However, the N-termini of P31 are pointing outwards from the pentamer (Abrescia et al., 2004). Thus, the N-terminal fluorescence tag should not interfere with the formation of the multimer.

Both fluorescent protein fusions with the major capsid protein P3 were broadly distributed in the multimericity assay starting from the monomeric forms, but also trimeric molecules were detected (P3 is a trimer; Benson et al., 1999) (Fig. 2D). Sedimentation analysis of the assembly protein P17 fusion (P17-eYFP) revealed both monomeric and multimeric forms (Fig. 2E).

The functionality of the produced fluorescent fusion proteins P5 and P31 was tested by complementation assay using PRD1 virus mutants, *sus690* and *sus525*, having amber mutation either in gene *V* (protein P5) or gene *XXXI* (protein P31), respectively. Both N- and C-terminal versions of P5 and P31 fusion proteins

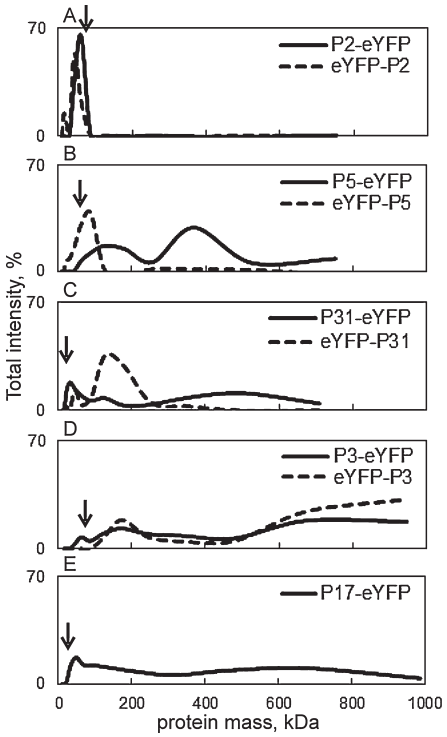


Fig. 2. Sedimentation assay to determine the multimericity of the fluorescent fusion proteins. The masses of the expressed proteins from the soluble fraction of the cell extracts were analyzed by rate zonal centrifugation using standard proteins as a control (see Section 4), SDS-PAGE and Western blotting using specific antibodies against P2, P5, P31 and P3. The P17 fusion protein was identified with antibody against GFP. Calculated monomeric masses are shown by arrows. Only fusion proteins with correct monomeric molecular mass were taken account when creating the image, smaller side products were seen with proteins P5 and P2. (A) Receptor binding protein P2 (monomeric in the virion). (B) Spike protein P5 (trimeric in the virion). (C) Penton protein P31 (pentameric in the virion). (D) Major capsid protein P3 (trimer in the virion). (E) Assembly factor P17 (tetrameric in its native form).

Table 1
Phages, bacterial strains and plasmids used in this study.

Phages, bacterial strains and plasmids	Relevant genotype or description (nt coordinates in PRD1 genome)[a]	Relevant phenotype	Source or reference
Phages			
PRD1 wt			Olsen et al. (1974)
PRD1 *sus690*	Amber mutation in gene *V*		Bamford and Bamford (2000)
PRD1 *sus525*	Amber mutation in gene *XXXI*		Rydman et al. (1999)
Bacterial strains			
Escherichia coli K-12			
HB101	*supE44 hsdS20* ($r_B^-m_B$) *recA13 ara14 proA2 lacY1 galK2 rpsL20 xyl5 mtl1*	Cloning host	Bolivar and Backman (1979)
HMS174(DE3)	*recA1 hsd* Rif^r	Expression host	Campbell et al. (1978)
Salmonella enterica serovar Typhimurium LT2			
DS88	*SL5676 ΔH2 H1-i::Tn 10 (TcS)* (pLM2)	Non-suppressor host for PRD1	Bamford and Bamford (1990)
PSA(pLM2)	*supE*	Suppressor host for *sus690*	Mindich et al. (1976)
DB7156(pLM2)	*leuA414(Am) hisC527(Am) supF30*	Suppressor host for *sus525*	Winston et al. (1979)
Plasmids			
pLM2		Encodes PRD1 receptor	Mindich et al. (1976)
pSU18	Low-copy-number cloning vector; p15A replicon, Cm^R		Bartolome et al. (1991)
pET24	High-level-expression vector; ColE1 replicon, Km^R		Novagen
pJB500	pSU18 + PRD1 *XXXI* + *V*	P5 and P31	Bamford and Bamford (2000)
pEGFP-N3			Clontech
pECFP-N3			Clontech
pEYFP-N3			Clontech
pJK5	pSU18Δ(*Eco*RI–*Hind*III)Ω(T7 RBS + *egfp* from pEGFP-N3)	eGFP	This study
pJK6	pSU18Δ(*Eco*RI–*Hind*III)Ω(T7 RBS + *ecfp* from pECFP-N3)	eCFP	This study
pJK7	pSU18Δ(EcoRI–HindIII)Ω(T7 RBS + *eyfp* from pEYFP-N3)	eYFP	This study
pSSM1	pET24Δ(*Eco*RI–*Hind*III)Ω(T7 RBS + *egfp* from pEGFP-N3)	eGFP	This study
pSSM2	pET24Δ(*Eco*RI–*Hind*III)Ω(T7 RBS + *ecfp* from pECFP-N3)	eCFP	This study
pSSM3	pET24Δ(*Eco*RI–*Hind*III)Ω(T7 RBS + *eyfp* from pEYFP-N3)	eYFP	This study
pKM57[b]	pJK5Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pKM51[b]	pJK6Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pKM54[b]	pJK7Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pKM67[b]	pJK5Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pKM61[b]	pJK6Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pKM64[b]	pJK7Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pJK22[b]	pSSM1Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pJK24[b]	pSSM2Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pJK28[b]	pSSM3Ω(*Bam*HI–*Xba*I)Ω(6 × Gly)	Cloning vector for fusion protein production	This study
pKM47[b]	pSSM1Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pKM41[b]	pSSM2Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pKM44[b]	pSSM3Ω(*Xba*I–*Bam*HI)Ω(6 × Gly)Δ(T7 RBS)	Cloning vector for fusion protein production	This study
pSSM22	pKM41Δ(*XbaI–BamHI*)Ω(PRD1 gene *V* (5287–6309))	P5-eCFP	This study

Table 1 (*Continued*).

Phages, bacterial strains and plasmids	Relevant genotype or description (nt coordinates in PRD1 genome)[a]	Relevant phenotype	Source or reference
pJK8	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *V* (5287–6309))	P5-eYFP	This study
pJK10	pKM54Δ(*Bam*HI–*Xba*I)Ω(PRD1 gene *V* (5287–6309))	eYFP-P5	This study
pJK12	pKM41Δ(*Xba*I–*Bam*HI)Ω(PRD1 gene *XXXI* (4907–5287))	P31-eCFP	This study
pSSM20	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *XXXI* (4907–5287))	P31-eYFP	This study
pSSM21	pKM54Δ(*Bam*HI–*Xba*I)Ω(PRD1 gene *XXXI* (4907–5287))	eYFP-P31	This study
pSSM30	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *II* (3128–4903))	P2-eYFP	This study
pSSM32	pKM54Δ(*Bam*HI–*Xba*I)Ω(PRD1 gene *II* (3128–4903))	eYFP-P2	This study
pSSM39	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *III* (8595–9782))	P3-eYFP	This study
pSSM41	pKM54Δ(*Bam*HI–*Xba*I)Ω(PRD1 gene *III* (8595–9782))	eYFP-P3	This study
pSSM43	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *XVII* (6328–6588))	P17-eYFP	This study
pSSM49	pKM64Δ(*Xba*I–*Bam*HI)Ω(T7 RBS + PRD1 gene *XVI* (11,836–12,189))	P16-eYFP	This study
pSSM34	pKM54Δ(BamHI–XbaI)Ω(PRD1 gene *XVI* (11,836–12,189))	eYFP-P16	This study

[a] Gene Bank Acc. No. AY848689 (Bamford et al., 1991; Saren et al., 2005).
[b] See details in Fig. 1.

complemented the defect in the corresponding gene at the same level as with the plasmid-produced wt protein or when the mutant was grown on suppressor host (Table 2) showing that the folding of the proteins was not compromised. According to the results fluorescent fusions did not seem to interfere the viral proteins significantly and multimerization was altered mostly in cases where the protein structures suggested steric hindrance between protein subunits.

2.3. Multimeric PRD1 proteins localize in the cell poles of E. coli

Localization of the fluorescent fusion proteins in *E. coli* cells was studied by confocal microscopy using living cells at the stationary phase of the bacterial growth. As a fusion protein, P16 was expressed all over the cytoplasm (Fig. 3A and B). Localization around the circumference of the cells, which is typical for membrane proteins, was not observed (Li and Young, 2012; Maier et al., 2008). In the virion, protein P16 locks the vertex complex to the inner membrane stabilizing the vertex structure and is found in the virus membrane (Abrescia et al., 2004; Jaatinen et al., 2004). During virus assembly, the interaction of P16 with the virus membrane might require other viral proteins, which may explain the localization of the P16 fusion protein. It is known that the formation of the procapsid including also P16 is dependent on the non-structural scaffolding protein P10 and assembly factor(s) P17 (and most probably P33) (Bamford et al., 2002; Mindich et al., 1982; Rydman et al., 2001).

Although both fusions of the monomeric receptor binding protein P2 were also evenly distributed inside the bacterium, P2 was occasionally found specifically localized in the cell poles (Fig. 3C and D). Clear loci (a locus is used here to describe the specific localization of fluorescence in a cell) were detected mainly with multimeric fusion proteins, especially with the spike protein P5 (Fig. 3E and F). Both fusions were clearly localized in one specific polar locus in the majority of the cells (~64% and ~81%; Fig. 3E and F).

More notable variations between C- and N-terminal fusion proteins were detected with the penton protein P31 and the major capsid protein P3. Like P2 fusion proteins, the fluorescence of P31-eYFP was evenly spread in the cytoplasm in the most of the cells, but about 5% of the cells had specific fluorescence locus (Fig. 3G). Most of the cells producing eYFP-P31 (~90%) had very low intensity level and the fluorescence was spread throughout the cytoplasm (Fig. 3H). However, rest of the cells (~10%) were having high fluorescence intensity and eYFP-P31 was specifically localized in the polar end of the bacteria (Fig. 3H). It can be concluded that the multimeric P31 fusion proteins were found mostly localized in specific polar regions more frequently than the monomeric ones (Figs. 2C and 3G, H).

The C-terminal fusion of the major capsid protein P3 (P3-eYFP) formed clear polar loci in half of the analyzed cells in all cultivations (Fig. 3I). With the eYFP-P3 fusion protein, the specific localization was detected only in around half of the parallel cultures. The other half of the cultivations had cells with fluorescence equally distributed. The parallel samples used for data collection on eYFP-P3 (Fig. 3J) were taken from the cultures with clear localization. Specific loci were detected not only in the polar region, but eYFP-P3 was also found specifically localized in several other regions inside the cell (~69% of the cells had loci) (Fig. 3J). Fluorescence was also found in specific loci in the fusion of assembly protein P17 (P17-eYFP) on most of the cells (Fig. 3K). P17 is tetrameric (Caldentey et al., 1999), but its function in the virus assembly is rather unknown. As a control, eYFP was produced alone and it was distributed evenly across the bacterial cell, as also previously reported (Fig. 3L) (Deich et al., 2004; Edgar et al., 2008).

The amount of loci in one cell was calculated from the samples, in which ~10% or more of the cells were with loci (Fig. 4). Protein P31 and most of the protein P5 were localized in one specific locus.

Table 2
Complementation titers of PRD1 mutants on strains producing either P5 or P31 fusion proteins.

Strain	Description	Titers (pfu/ml)	
		sus690 (gene *V* mutant)	*sus525* (gene *XXXI* mutant)
DS88	Non-suppressor host	2×10^5	1×10^5
PSA	Suppressor host	3×10^{11}	
DB7156	Suppressor host		2×10^{11}
HMS174(pLM2)(pSU18)	Negative control	2.6×10^4	3.3×10^4
HMS174(pLM2)(pJB500)	Positive control	1×10^{10}	1.6×10^{10}
HMS174(pLM2)(pJK10)	eYFP-P5	1×10^{10} [a]	
HMS174(pLM2)(pJK8)	P5-eYFP	2.3×10^{10}	
HMS174(pLM2)(pSSM20)	P31-eYFP		2.4×10^{10}
HMS174(pLM2)(pSSM21)	eYFP-P31		4.9×10^{10}

[a] A bit weaker plaques than in other plates.

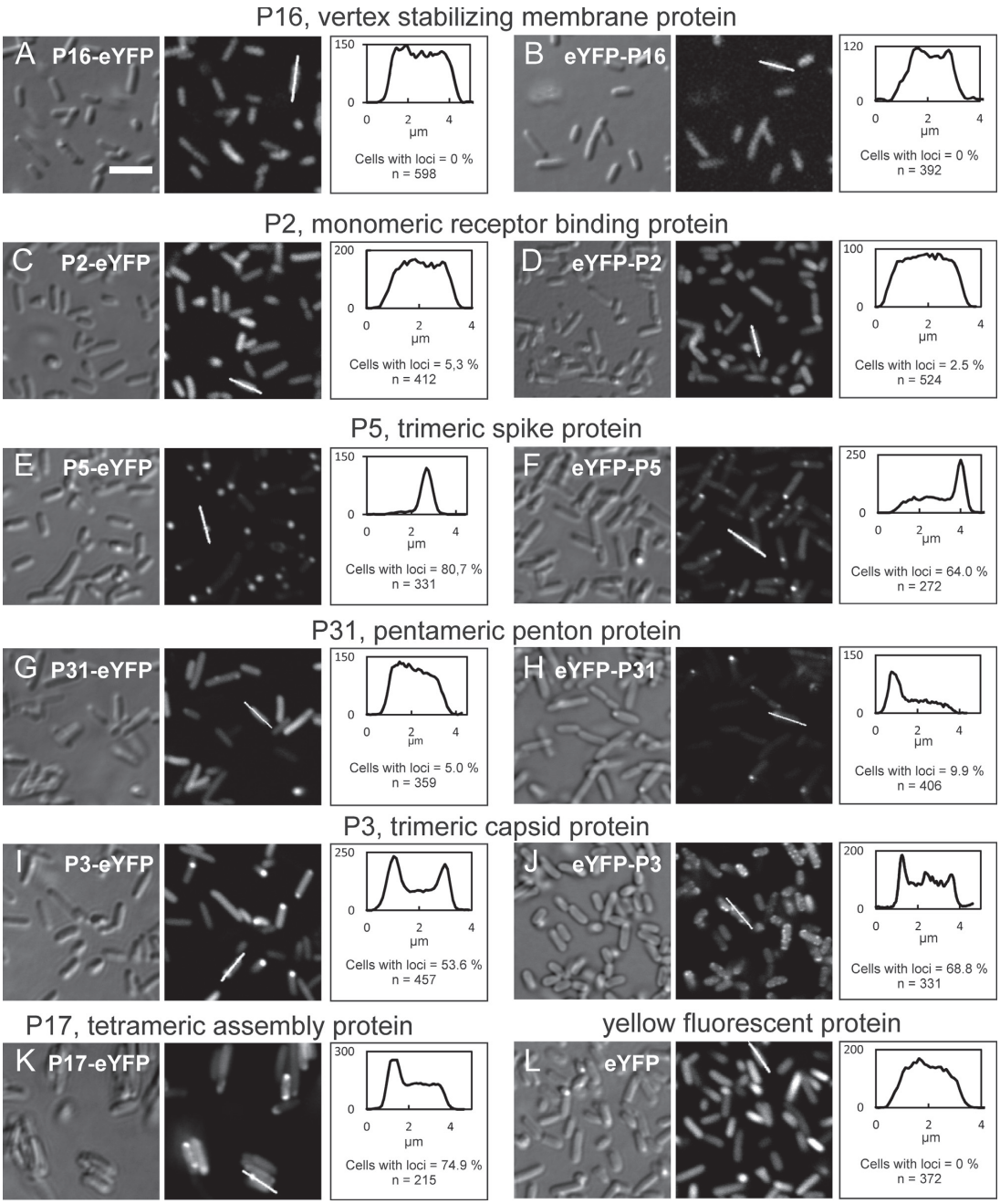


Fig. 3. Localization of PRD1 proteins. Confocal microscope images of different fluorescent fusion proteins of PRD1 overproduced in *E. coli* HMS174 (DE3) cells. In the left column differential interference contrast (DIC) image, in the middle fluorescence image and on the right intensity (*y* axes) profile of a single cell marked with a white line (see the fluorescence image) and percentage of cells with loci are shown. The scale bar applicable to all images in (A) is 5 µm.

The remaining cells producing P5 had two loci. With P17-eYFP the number of loci in one cell was slightly more diverse (from one to five loci), but majority of the cells had one locus (60% of cells with localization). Both P3 fusion proteins were often found in several specific positions (Fig. 4). Especially, eYFP-P3 was localized more often in three or more loci per cell than in one specific locus. In the virion, the major capsid protein P3 has a connection to the viral inner membrane by its N-terminus (Abrescia et al., 2004; Benson et al., 1999). Both N- and C- termini are located outwards from the trimeric protein capsomer, nevertheless they have roles in locking trimers together to form larger capsid facets (Abrescia et al., 2004). The membrane connection might partly explain the observed

differences in the localization of P3 fusion proteins. In addition, the loci of eYFP-P3 were occasionally in shuffling motion (data not shown).

These observations indicated that the viral proteins had a specific intracellular distribution and the multimeric ones seemed to accumulate into polar areas of the host cells similarly with bacterial proteins, for example the chemotaxis protein CheA (Sourjik and Berg, 2000) and chaperon protein GroES (Li and Young, 2012) (Fig. 3, Supplementary Fig. S1). However, there were differences in the number of loci between viral proteins thus the process leading to localization might not be the same for all proteins. The specific polar regions for protein localization might be the assembly sites of protein multimers. It has also been reported that the protein aggregation has led to a similar polar localization (Lindner et al., 2008; Lloyd-Price et al., 2012). However, the low copy-number plasmid pSU18 used here has been widely utilized for the production of functional PRD1 structural proteins (Bamford and Bamford, 2000; Bartolome et al., 1991; Rydman et al., 2001). In addition, the fluorescent fusion proteins of P5 and P31 complemented the defect of virus mutants (Table 2). Also the detected localization varied between proteins and for example the clear difference between monomeric and multimeric proteins indicates that the proteins were produced as soluble (Fig. 2).

2.4. Proteins P5 and P31 co-localize within a specific locus area

The observed protein localization was studied further with proteins P5 and P31. In the virion, P5 and P31 are known to interact as a part of the spike vertex complex (Caldentey et al., 2000). To find out whether the proteins co-localize in the same locus, P5 and P31 were fused with eCFP and eYFP, respectively, and their co-expression was imaged by confocal microscopy. The background was manually removed, yielding images only from the higher intensity loci. The loci were considered to be co-localized if the locus emission had contribution from both eCFP and eYFP labels. When P5-eCFP and eYFP-P31 were co-produced, 66% of the loci were identical ($n = 140/211$). Rest of the loci contained only either eYFP-P31 (16%, $n = 34/211$) or P5-eCFP (18%, $n = 37/211$), which is mostly explained by a production of only one type of fusion protein in a cell.

Based on PRD1 structural data the theoretical maximum distance observed in the virion between the C-terminus of P5 and the N-terminus of P31 is around 30 nm (Abrescia et al., 2004; Huiskonen et al., 2007). Protein P5 is an elongated and flexible trimer with a collagen like region in the middle of the protein (Bamford et al., 1991; Bamford and Bamford, 2000; Caldentey et al., 2000; Sokolova et al., 2001). In addition, the glycine linker region in the fusion protein also allows the fluorescent protein tag to move and interact suggesting that protein–protein interaction could be followed by förster resonance energy transfer (FRET).

We tested the energy transfer between proteins P5 and P31 by fluorescence lifetime microscope (FLIM) measurements using three samples: (i) co-expression of P5-eCFP and eYFP-P31, which represents the ideal combination for the interaction studies, (ii) co-expression of eYFP-P5 and P31-eCFP, where the fluorescent proteins hinder the interaction, and (iii) expression of P31-eCFP, used as a control to observe the lifetime of eCFP in the absence of FRET. With samples (i) and (iii) the lifetime of eCFP was measured from locus area, outside the locus area and from entire cell (Supplementary Fig. S1A). For sample (ii) only entire cells were measured. The results showed a minimal energy transfer in the locus area of sample (i) comparing to other measured samples, and so no significant FRET could be observed (Supplementary Fig. S1B and C). Similar results were obtained when fluorescence spectra and decays were measured by confocal microscopy and time-correlated single photon counting (TCSPC) from liquid cell samples, respectively (Supplementary Fig. S1D and E). In studies by others (Onuki et al., 2002) the efficiency of the energy transfer between CFP and YFP has been observed to be significantly more efficient than we observed here, although the orientation of P5 and P31 as fusion proteins was theoretically ideal for detection FRET based on their X-ray structures and orientation in the virion.

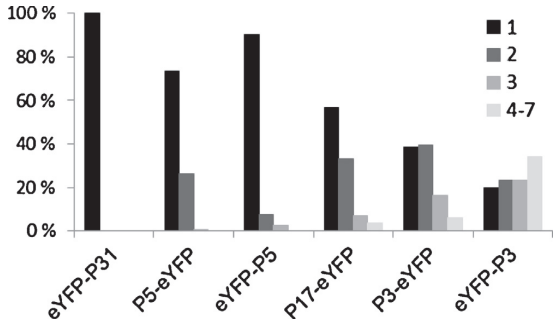


Fig. 4. Amount of loci in single cells. The distribution of amount of loci in one cell calculated from images used in Fig. 3. Only cells with loci were taken account.

3. Conclusions

Numerous bacterial proteins and functions have been localized to the poles of bacteria (Gestwicki et al., 2000; Li and Young, 2012; Maddock et al., 1993) and in other specific regions (Maier et al., 2008; Nevo-Dinur et al., 2012; Russell and Keiler, 2008). The current knowledge of the complexity of bacterial cells provides also a new aspect to the study of functions and life cycle of bacterial viruses. Animal viruses exploit the organization of host cells in very efficient way and it is likely that bacteriophages do the same. We designed and created a vector library (Fig. 1) and utilized it in expression of virus specific proteins, but the approach can be used easily for other research frames to produce proteins with fluorescent tags. We observed mainly polar localization of several PRD1 viral proteins (Fig. 3). The clear localization was observed only with multimeric proteins as monomeric proteins seemed to be evenly distributed. We also showed that the host receptor recognition vertex associated proteins, the spike protein P5 and the penton protein P31 co-localize within specific cell areas in *E. coli*. These polar areas might play a role in the multimerization and formation of viral protein complexes.

During the virus life cycle, viral proteins are expressed and function according to precise scheme in a close interaction with each other. When a single viral protein is produced from a plasmid, the protein loses these interactions occurring during the viral replication cycle and this might have an influence on the protein functions. However, in the absence of the natural virus infection context viral proteins tested here showed a clear polarized localization in the host cells. The specific localization pattern of these virus proteins suggests that localization is determined either by the proteins themselves and/or by their interactions with other proteins. This indicates that viral proteins are not randomly distributed in the host cell and their (polar) localization might be explained by interaction with specific bacterial proteins. Based on our results, we suggest that viral proteins are interacting with specific bacterial proteins essential for the viral infection. However, many fundamental aspects regarding the molecular mechanisms of interactions as well as the specific bacterial interactions partners remain to be elucidated.

4. Materials and methods

4.1. Bacteria, plasmids, molecular cloning and protein expression

The strains and plasmids used in this study are listed in Table 1. Cells were grown on Luria-Bertani (LB) medium with chloramphenicol (25 μg/ml) and kanamycin (25 μg/ml) when appropriate. *E. coli* HB101 was used as a host for plasmid propagation and molecular cloning. Genes *egfp*, *eyfp* and *ecfp* were amplified by PCR using pEGFP-N3, pECFP-N3 and pEYFP-N3 as templates with primers including a ribosomal binding site (RBS) and restriction enzyme cleavage sites. The fragments were cloned between *Eco*RI and *Hind*III restriction sites in pSU18 and pET24 vectors. Site-directed mutagenesis was used to insert a linker encoding six glycines and new restriction enzyme cutting sites in one end of fluorescent protein genes resulting in 12 new vectors (Fig. 1; Table 1). These vectors were used to construct plasmids for production of C- and N-terminal fusions with yellow or cyan fluorescent proteins. For fusion construction PRD1 genes *II*, *III*, *V*, *XVI*, *XVII* and *XXXI* were amplified by PCR using the phage genome as a template and specific primers, containing restriction enzyme cutting sites and when needed RBS. The fragments were cloned between *Xba*I and *Bam*HI restriction sites in pKM41, pJK24, pKM54 or pKM64 (Fig. 1; Table 1). The plasmids were sequenced using the Sanger sequencing method with an automatic sequencer (Applied Biosystems 3130xl Genetic Analyzer) and a BigDye Terminator, version 3.1, Cycle Sequencing Kit (Applied Biosystems). Base calling and sequence refining were performed with Sequencing Analysis, version 5.2.0 (Applied Biosystems). Plasmids were transformed to *E. coli* HMS174(DE3) cells, which were used for protein expression.

The cells were grown at 28 °C. The protein production was induced at $A_{550} = 0.75$ by adding isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration of 1 mM) and the growth was continued for 16–20 h at 18 °C. The strains used for the co-expression of P5-eCFP and eYFP-P31 and for co-expression of eYFP-P5 and P31-eCFP were HMS174(DE3)(pSSM21)(pSSM22) and HMS174(DE3)(pJK10)(pJK12), respectively.

4.2. Solubility and multimericity of the proteins

Cells were grown as described above and concentrated 1:100 by centrifugation (Sorvall SLA3000, 4200 g, 10 min, 5 °C) in 50 mM Tris–HCl, pH 7.2. Cells were disrupted by a French pressure cell. To analyze the solubility of the proteins, the protein ratio between the supernatant and pellet was determined after centrifugation (Sorvall SLA3000 rotor, 10,800 g, 15 min, 5 °C) by SDS-PAGE and Western blotting. SDS-PAGE was performed according to previously reported method (Olkkonen and Bamford, 1989) and for Western blotting the proteins were transferred onto a PVDF membrane (Millipore). Monoclonal 16A201 (anti-P16) serum (Hänninen et al., 1997) polyclonal antisera agaist PRD1 proteins P2, P3, P5 and P31 antisera (Grahn et al., 1999; Hänninen et al., 1997; Rydman et al., 1999, 2001) or anti-GFP (Invitrogen) were used as primary antibodies. Proteins were visualized with the Thermo Scientific Supersignal West Pico Chemiluminescent Substrate kit using HRP-conjugated swine anti-rabbit Igs (Dako) as a secondary antibody.

For the protein multimericity assay the supernatant was applied on a top of a 10–40% (w/v) linear sucrose gradient in 50 mM Tris–HCl, pH 7.2 and centrifuged (Beckmann SW41 rotor, 210,000 × g, 42 h, 15 °C). Lysozyme (14.3 kDa), albumin (66 kDa), bovine serum albumin (68 kDa), lactate dehydrogenase (140 kDa), aldolase (158 kDa), catalase (232 kDa), ferritin (440 kDa) and thyroglobulin (669 kDa) were used as molecular mass markers. After centrifugation twelve 1 ml-fractions and the pellet were collected and analyzed by SDS-PAGE and Western blotting (see above).

PRD1 mutants *sus690* (amber mutation in gene *V*) and *sus525* (amber mutation in gene *XXXI*) were propagated on *Salmonella enterica* suppressor strain PSA (*supE*) or DB7156 (*supF30*) harboring pLM2, respectively (Table 1). The titers of the viruses were determined on their suppressor strain and on the non-suppressing strain *S. enterica* serovar Typhimurium LT2 DS88 (wt host; Table 1). The functions of the fluorescent fusion proteins (eYFP-P5, P5-eYFP, P31-eYFP and eYFP-P31) were tested with the mutant viruses by *in vivo* complementation assay using plasmids carrying either the genes for the fusion proteins or the corresponding PRD1 wt genes *V* and *XXXI* (Table 1). The PRD1 sensitive strain carrying only the cloning vector was used as a negative control.

4.3. Confocal microscopy and localization

The cells were grown as described and diluted 1:1 in phosphate-buffered saline buffer (PBS). Plates were coated with poly-L-lysine (0.01%, MW 70,000–150,000) and cell suspension was applied to the plates and incubated for 10–20 min. Excess of cell suspension was removed. The samples were covered with LB-soft-agar and imaged immediately.

The imaging was performed with an Olympus FV1000 laser scanning confocal microscope attached to an IX81 inverted microscope frame (Olympus, Japan) with an UPLSAPO 60x water immersion objective having a numerical aperture (NA) 1.20 for live samples or UPLSAPO 60× oil immersion objective (NA = 1.3) for fixed samples. eGFP was excited with 488 nm laserline, eCFP with the 405 nm laserline, and eYFP with 515 nm laserline, and the fluorescence signals were collected with 500–600 nm, 425–525 nm, and 530–630 nm band-pass filters, respectively. In co-localization imaging 458 nm laserline was used for eCFP excitation and the fluorescence was collected with 465–505 nm band-pass filter. Images were captured with an image size of 512 pixels × 512 pixels. For presentation purpose (Fig. 3) the images were cropped further to the size of 18 μm × 18 μm. The number of cells with localization loci and the average amount of loci in one cell were calculated from the images from three separate cultivations.

Acknowledgements

We thank Kati Mökkönen for the help in DNA cloning and Dr. Vladimir Chukharev for technical support. This study was supported by the Academy of Finland Centre of Excellence Program in Virus Research (11296841, 2006–2011 J.K.H.B.), the Centre of Excellence Program in Biological Interactions (252411, 2012–2014 J.K.H.B.), Academy of Finland grants 127665 (H.M.O.), 138063 (J.A.I.) and 251106 (J.K.H.B.) and a grant from the Finnish Cultural Foundation (S.M.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2013.11.015.

References

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., 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.

Amster-Choder, O., 2011. The compartmentalized vessel: the bacterial cell as a model for subcellular organization (a tale of two studies). Cell. Logist. 1, 77–81.

Bamford, J.K., Bamford, D.H., 1990. Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane. Virology 177, 445–451.

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.

Bamford, J.K., Hänninen, A.L., Pakula, T.M., Ojala, P.M., Kalkkinen, N., Frilander, M., Bamford, D.H., 1991. Genome organization of membrane-containing bacteriophage PRD1. Virology 183, 658–676.

Bamford, J.K.H., Bamford, D.H., 2000. A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to the vertex. J. Virol. 74, 7781–7786.

Bartolome, B., Jubete, Y., Martinez, E., de la Cruz, F., 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene 102, 75–78.

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.

Bolivar, F., Backman, K., 1979. Plasmids of *Escherichia coli* as cloning vectors. Methods Enzymol. 68, 245–267.

Caldentey, J., Hänninen, A.L., Holopainen, J.M., Bamford, J.K., Kinnunen, P.K., Bamford, D.H., 1999. Purification and characterization of the assembly factor P17 of the lipid-containing bacteriophage PRD1. Eur. J. Biochem. 260, 549–558.

Caldentey, J., Tuma, R., Bamford, D.H., 2000. Assembly of bacteriophage PRD1 spike complex: role of the multidomain protein P5. Biochemistry 39, 10566–10573.

Campbell, J.L., Richardson, C.C., Studier, F.W., 1978. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc. Natl. Acad. Sci. U.S.A. 75, 2276–2280.

Chen, I., Christie, P.J., Dubnau, D., 2005. The ins and outs of DNA transfer in bacteria. Science 310, 1456–1460.

Cockburn, J.J., Abrescia, N.G., Grimes, J.M., Sutton, G.C., Diprose, J.M., Benevides, J.M., Thomas Jr., G.J., 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.

Deich, J., Judd, E.M., McAdams, H.H., Moerner, W.E., 2004. Visualization of the movement of single histidine kinase molecules in live Caulobacter cells. Proc. Natl. Acad. Sci. U.S.A. 101, 15921–15926.

Edgar, R., Rokney, A., Feeney, M., Semsey, S., Kessel, M., Goldberg, M.B., Adhya, S., Oppenheim, A.B., 2008. Bacteriophage infection is targeted to cellular poles. Mol. Microbiol. 68, 1107–1116.

Gerdes, K., Howard, M., Szardenings, F., 2010. Pushing and pulling in prokaryotic DNA segregation. Cell 141, 927–942.

Gestwicki, J.E., Lamanna, A.C., Harshey, R.M., McCarter, L.L., Kiessling, L.L., Adler, J., 2000. Evolutionary conservation of methyl-accepting chemotaxis protein location in Bacteria and Archaea. J. Bacteriol. 182, 6499–6502.

Grahn, A.M., Butcher, S.J., Bamford, J.K.H., Bamford, D.H., 2006. PRD1 – dissecting the genome, structure and entry. In: Calendar, R. (Ed.), The Bacteriophages. Oxford University Press, New York, pp. 161–170.

Grahn, A.M., Caldentey, J., Bamford, J.K., Bamford, D.H., 1999. Stable packaging of phage PRD1 DNA requires adsorption protein P2, which binds to the IncP plasmid-encoded conjugative transfer complex. J. Bacteriol. 181, 6689–6696.

Hänninen, A.L., Bamford, D.H., Bamford, J.K., 1997. Probing phage PRD1-specific proteins with monoclonal and polyclonal antibodies. Virology 227, 198–206.

Holopainen, J.M., Saily, M., Caldentey, J., Kinnunen, P.K., 2000. The assembly factor P17 from bacteriophage PRD1 interacts with positively charged lipid membranes. Eur. J. Biochem. 267, 6231–6238.

Huiskonen, J.T., Manole, V., Butcher, S.J., 2007. Tale of two spikes in bacteriophage PRD1. Proc. Natl. Acad. Sci. U.S.A. 104, 6666–6671.

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.

Jakutyte, L., Baptista, C., Sao-Jose, C., Daugelavicius, R., Carballido-Lopez, R., Tavares, P., 2011. Bacteriophage infection in rod-shaped gram-positive bacteria: evidence for a preferential polar route for phage SPP1 entry in *Bacillus subtilis*. J. Bacteriol. 193, 4893–4903.

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.

Lai, E.M., Nair, U., Phadke, N.D., Maddock, J.R., 2004. Proteomic screening and identification of differentially distributed membrane proteins in *Escherichia coli*. Mol. Microbiol. 52, 1029–1044.

Leonardy, S., Miertzschke, M., Bulyha, I., Sperling, E., Wittinghofer, A., Sogaard-Andersen, L., 2010. Regulation of dynamic polarity switching in bacteria by a Ras-like G-protein and its cognate GAP. EMBO J. 29, 2276–2289.

Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., Taddei, F., 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc. Natl. Acad. Sci. U.S.A. 105, 3076–3081.

Lloyd-Price, J., Häkkinen, A., Kandhavelu, M., Marques, I.J., Chowdhury, S., Lihavainen, E., Yli-Harja, O., Ribeiro, A.S., 2012. Asymmetric disposal of individual protein aggregates in *Escherichia coli*, one aggregate at a time. J. Bacteriol. 194, 1747–1752.

Li, G., Young, K.D., 2012. Isolation and identification of new inner membrane-associated proteins that localize to cell poles in *Escherichia coli*. Mol. Microbiol. 84, 276–295.

Loose, M., Kruse, K., Schwille, P., 2011. Protein self-organization: lessons from the min system. Annu. Rev. Biophys. 40, 315–336.

Maddock, J.R., Alley, M.R., Shapiro, L., 1993. Polarized cells, polar actions. J. Bacteriol. 175, 7125–7129.

Maier, K.S., Hubich, S., Liebhart, H., Krauss, S., Kuhn, A., Facey, S.J., 2008. An amphiphilic region in the cytoplasmic domain of KdpD is recognized by the signal recognition particle and targeted to the *Escherichia coli* membrane. Mol. Microbiol. 68, 1471–1484.

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

Mindich, L., Cohen, J., Weisburd, M., 1976. Isolation of nonsense suppressor mutants in *Pseudomonas*. J. Bacteriol. 126, 177–182.

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.

Muñoz-Espín, D., Daniel, R., Kawai, Y., Carballido-Lopez, R., Castilla-Llorente, V., Errington, J., Meijer, W.J., Salas, M., 2009. The actin-like MreB cytoskeleton organizes viral DNA replication in bacteria. Proc. Natl. Acad. Sci. U.S.A. 106, 13347–13352.

Muñoz-Espín, D., Holguera, I., Ballesteros-Plaza, D., Carballido-Lopez, R., Salas, M., 2010. Viral terminal protein directs early organization of phage DNA replication at the bacterial nucleoid. Proc. Natl. Acad. Sci. U.S.A. 107, 16548–16553.

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.

Nevo-Dinur, K., Govindarajan, S., Amster-Choder, O., 2012. Subcellular localization of RNA and proteins in prokaryotes. Trends Genet. 28, 314–322.

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., Amster-Choder, O., 2011. Translation-independent localization of mRNA in *E. coli*. Science 331, 1081–1084.

Oksanen, H.M., Bamford, D.H., 2012. Family *Tectiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Oxford, pp. 317–322.

Olkkonen, V.M., Bamford, D.H., 1989. Quantitation of the adsorption and penetration stages of bacteriophage phi 6 infection. Virology 171, 229–238.

Olsen, R.H., Siak, J.S., Gray, R.H., 1974. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. J. Virol. 14, 689–699.

Onuki, R., Nagasaki, A., Kawasaki, H., Baba, T., Uyeda, T.Q., Taira, K., 2002. Confirmation by FRET in individual living cells of the absence of significant amyloid beta-mediated caspase 8 activation. Proc. Natl. Acad. Sci. U.S.A. 99, 14716–14721.

Ramamurthi, K.S., 2010. Protein localization by recognition of membrane curvature. Curr. Opin. Microbiol. 13, 753–757.

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

Rothenberg, E., Sepúlveda, L.A., Skinner, S.O., Zeng, L., Selvin, P.R., Golding, I., 2011. Single-virus tracking reveals a spatial receptor-dependent search mechanism. Biophys. J. 100, 2875–2882.

Rudner, D.Z., Losick, R., 2002. A sporulation membrane protein tethers the pro-sigmaK processing enzyme to its inhibitor and dictates its subcellular localization. Genes Dev. 16, 1007–1018.

Rudner, D.Z., Losick, R., 2010. Protein subcellular localization in bacteria. Cold Spring Harb. Perspect. Biol. 2, a000307.

Russell, J.H., Keiler, K.C., 2007. Peptide signals encode protein localization. J. Bacteriol. 189, 7581–7585.

Russell, J.H., Keiler, K.C., 2008. Screen for localized proteins in *Caulobacter crescentus*. PLoS ONE 3, e1756.

Rydman, P.S., Bamford, J.K., Bamford, D.H., 2001. A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. J. Mol. Biol. 313, 785–795.

Rydman, P.S., Caldentey, J., Butcher, S.J., Fuller, S.D., Rutten, T., Bamford, D.H., 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. J. Mol. Biol. 291, 575–587.

Saren, A.M., Ravantti, J.J., Benson, S.D., Burnett, R.M., Paulin, L., Bamford, D.H., Bamford, J.K., 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. J. Mol. Biol. 350, 427–440.

Sokolova, A., Malfois, M., Caldentey, J., Svergun, D.I., Koch, M.H., Bamford, D.H., Tuma, R., 2001. Solution structure of bacteriophage PRD1 vertex complex. J. Biol. Chem. 276, 46187–46195.

Sourjik, V., Berg, H.C., 2000. Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. Mol. Microbiol. 37, 740–751.

Swulius, M.T., Jensen, G.J., 2012. The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-Terminal yellow fluorescent protein tag. J. Bacteriol. 194, 6382–6386.

Vats, P., Yu, J., Rothfield, L., 2009. The dynamic nature of the bacterial cytoskeleton. Cell Mol. Life Sci. 66, 3353–3362.

Winston, F., Botstein, D., Miller, J.H., 1979. Characterization of amber and ochre suppressors in *Salmonella typhimurium*. J. Bacteriol. 137, 433–439.

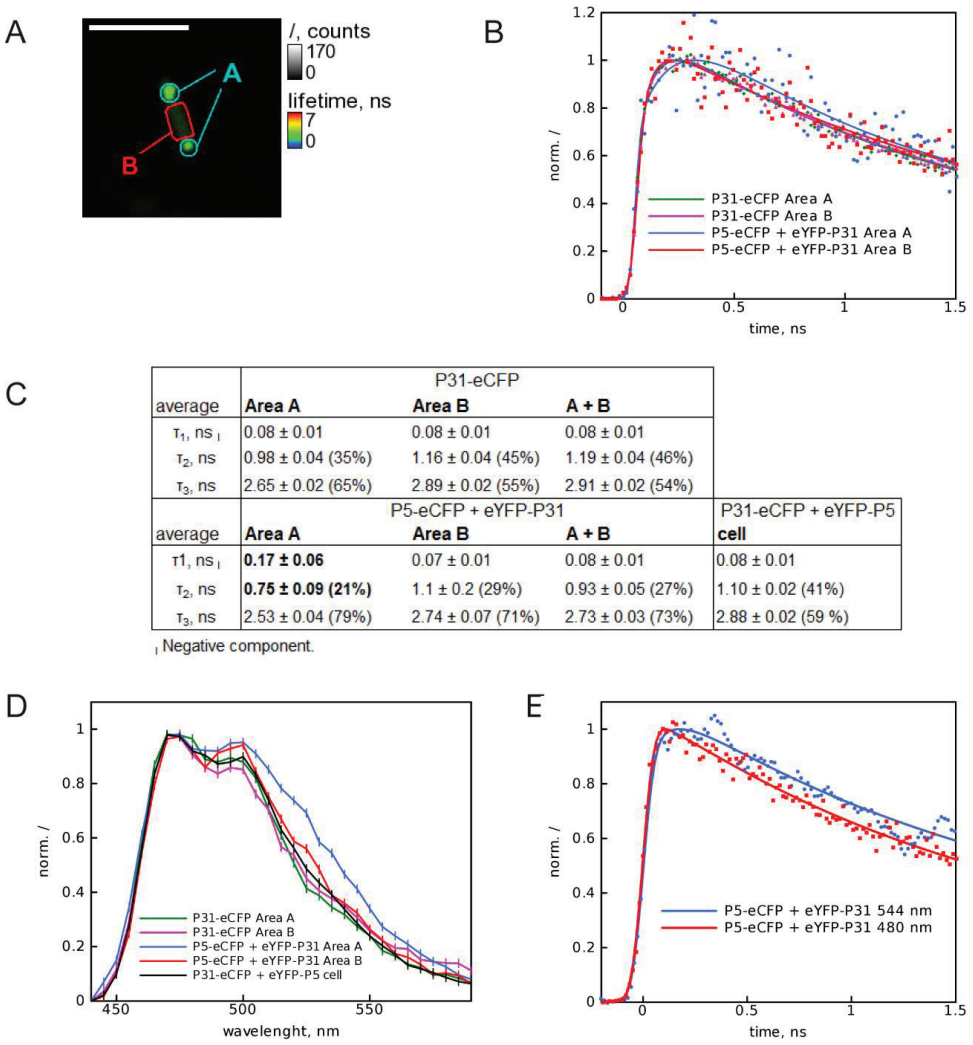
Xu, L., Benson, S.D., Butcher, S.J., Bamford, D.H., Burnett, R.M., 2003. The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. Structure 11, 309–322.

Appendix A. Supplementary data

Table S1. Mutations in PRD1 genes

Plasmid	Relevant phenotype	Mutation in PRD1 gene[a]
pSSM30	P2-eYFP	G555A
pSSM32	eYFP-P2	
pSSM39	P3-eYFP	No mutations
pSSM41	eYFP-P3	No mutations
pJK8	P5-eYFP	T211I
pJK10	eYFP-P5	T211I
pSSM34	eYFP-P16	P63L
pSSM49	16-eYFP	P63L
pSSM43	P17-eYFP	No mutations
pSSM20	P31-eYFP	No mutations
pSSM21	eYFP-P31	No mutations

[a)] according to the genome sequence



C

average	P31-eCFP Area A	P31-eCFP Area B	P31-eCFP A + B	
τ_1, ns [I]	0.08 ± 0.01	0.08 ± 0.01	0.08 ± 0.01	
τ_2, ns	0.98 ± 0.04 (35%)	1.16 ± 0.04 (45%)	1.19 ± 0.04 (46%)	
τ_3, ns	2.65 ± 0.02 (65%)	2.89 ± 0.02 (55%)	2.91 ± 0.02 (54%)	
average	P5-eCFP + eYFP-P31 Area A	P5-eCFP + eYFP-P31 Area B	P5-eCFP + eYFP-P31 A + B	P31-eCFP + eYFP-P5 cell
τ1, ns [I]	**0.17 ± 0.06**	0.07 ± 0.01	0.08 ± 0.01	0.08 ± 0.01
τ_2, ns	**0.75 ± 0.09 (21%)**	1.1 ± 0.2 (29%)	0.93 ± 0.05 (27%)	1.10 ± 0.02 (41%)
τ_3, ns	2.53 ± 0.04 (79%)	2.74 ± 0.07 (71%)	2.73 ± 0.03 (73%)	2.88 ± 0.02 (59 %)

[I] Negative component.

Fig. S1. FRET measurements of interaction between co-produced P5 and P31 fusion proteins. (A) A typical FLIM image of an *E. coli* cell co-producing P5-eCFP and eYFP-P31 (P5-eCFP + eYFP-P31) on a microscope cover glass substrate. At each pixel, brightness and color denote total intensity (*I*, counts) and average lifetime (lifetime, ns), respectively. The scale bar is 10 μm. The fluorescence decay curves were measured from area A (blue A) and area B (red B). (B) Typical normalized fluorescence decay curves corresponding to area A and area B of samples P31-eCFP and P5-eCFP + eYFP-P31. The solid lines show tri-exponential fits of the data. The emission decay curve in the polar locus of P5-eCFP + eYFP-P31 (area A)

differed a bit from the other samples measured. (C) Emission decay fit results (n=5) of samples P31-eCFP, P5-eCFP + eYFP-P31 and P31-eCFP + eYFP-P5. τ_i are the lifetimes, τ_1 is the rise component and the percentages of the amplitudes of the lifetimes, τ_2 and τ_3 are presented. Areas A, B and A+B were fitted separately. The first lifetime τ_1 was slightly longer in the area A of the cell co-producing P5-eCFP and eYFP-P31 than of the control samples. In addition, shortening of the lifetimes of τ_2 and τ_3 was observed in the same area. These observations suggest that a slightly excitation of eYFP is followed by energy transfer from eCFP (Tramier *et al.* 2002, Shimozono *et al.* 2006, Hellwig *et al.* 2008).

(D) The normalized emission spectra (n = 5) obtained by confocal microscopy. An additional emission of eYFP at 520–540 nm could be detected in the area A containing the co-produced P5-eCFP and eYFP-P31, whereas no additional emission was observed in other regions or samples showing that the additional emission observed was not due to the direct excitation of eYFP.

(E) The normalized fluorescence decays of co-produced P5-eCFP and eYFP-P31 excited at 440 nm and measured at 480 and 544 nm. The solid lines show tri-exponential fits (fitted numbers are not shown) of the data. The fluorescence decay of the sample at 544 nm (eCFP + eYFP detection) is slightly longer than at 480 nm (eCFP detection).

Supplementary methods:

For spectroscopic single cell measurements samples were prepared similar as live samples except that the cells were fixed with 4% paraformaldehyde (20 min, 22 °C) prior to the mounting of the samples with Mowiol 4–88 - DABCO (Sigma-Aldrich). For *time-correlated single photon counting (TCSPC)* measurements cells were grown as described, transferred to Tris-HCl buffer (50 mM, pH 7.2.). The samples were diluted so that the absorption was sufficiently low to prevent inner filter effect.

Fluorescence lifetime images were acquired by inverse time-resolved fluorescence lifetime microscope MicroTime-200 (PicoQuant GmBH, Germany) coupled with the inverted microscope Olympus IX71 (Olympus, Japan) with a 100x objective (used with the immersion oil) having a NA 0.8, which enables a minimum spatial resolution of 0.3 μm and a maximum scan area of 80x80 μm^2. The excitation wavelength and the time resolution were 405 nm and 60–70 ps (FWHM of the instrument response function (IRF)), respectively (Tkachenko 2006). Emission was acquired with 415 nm long-pass filter. The SymPhoTime v. 4.7 software was used to calculate the lifetime map images.

Emission spectra were measured from fixed samples with Olympus confocal microscope described in section 4.3. eCFP was excited with 405 nm laserline and the emitted fluorescence was collected between 440–590 nm.

Fluorescence decays of the cell sample P5-eCFP + eYFP-P31 in the sub-nanosecond and nanosecond time scales were measured using a TCSPC system consisting of a HydraHarp 400 controller and a PDL 800-B driver (PicoQuant GmBH). The samples were excited at 440 nm (spectral FWHM 20 nm) with the pulsed diode laser head LDH-P-C-440 at a repetition frequency of 40 MHz driven by the PDL 800-B. The output power of the laser was 0.55 mW/cm^2 for 440 nm excitation. The 460 nm long-pass filter and the band-pass filters were used to detect the emission above 480 ± 10 nm and 544 ± 10 nm with a micro channel plate (MCP, R1564-07). The electrical signal obtained from the MCP detector was amplified by a pre-amplifier (PAM 102-M). The time resolution of the experiment was determined to be approximately 80 ps (FWHM of the IRF). Measurements were carried out at room temperature and ambient conditions.

References:

Hellwig, D., Münch, S., Orthaus, S., Hoischen, C., Hemmerich, P., Diekmann, S., 2008. Live cell imaging reveals sustained centromere binding of CENP-T via CENP-A and CENP-B. *J. Biophotonics* 1, 245-254.

Shimozono, S., Hosoi, H., Mizuno, H., Fukano, T., Tahara, T., Miyawaki, A., 2006. Concatenation of cyan and yellow fluorescent proteins for efficient resonance energy transfer. *Biochemistry* 45, 6267-6271.

Tkachenko, N.V., 2006. Optical Spectroscopy: Methods and Instrumentations. Elsevier, 523 Amsterdam, pp. 115.

Tramier, M., Gautier, I., Piolot, T., Ravalet, S., Kemnitz, K., Coppey, J., Durieux, C., Mignotte, V., Coppey-Moisan, M., 2002. Picosecond-hetero-FRET microscopy to probe protein-protein interactions in live cells. *Biophys. J*. 83, 3570-3577.

II

NON-STRUCTURAL PROTEINS P17 AND P33 ARE INVOLVED IN THE ASSEMBLY OF THE INTERNAL MEMBRANE-CONTAINING VIRUS PRD1

by

Jenni Karttunen, Sari Mäntynen, Teemu O. Ihalainen, Jaana K.H. Bamford & Hanna M. Oksanen 2015.

Virology 482: 225-233.

Virology 482 (2015) 225–233

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Non-structural proteins P17 and P33 are involved in the assembly of the internal membrane-containing virus PRD1

Jenni Karttunen [a,1,2], Sari Mäntynen [a,2], Teemu O. Ihalainen [b], Jaana K.H. Bamford [a], Hanna M. Oksanen [c,*]

[a] Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science and Nanoscience Center, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland
[b] Stem Cells in Neurological Applications Group, BioMediTech, University of Tampere, Tampere, Finland
[c] Institute of Biotechnology and Department of Biosciences, University of Helsinki, Biocenter 2, P.O. Box 56 (Viikinkaari 5), FIN-00014 Helsinki, Finland

ARTICLE INFO

Article history:
Received 3 October 2014
Returned to author for revisions
30 November 2014
Accepted 22 March 2015
Available online 14 April 2015

Keywords:
Bacteriophage
Membrane virus
Assembly
Chaperonin
Fluorescent protein
Protein localisation
Fluorescence recovery after photobleaching

ABSTRACT

Bacteriophage PRD1, which has been studied intensively at the structural and functional levels, still has some gene products with unknown functions and certain aspects of the PRD1 assembly process have remained unsolved. In this study, we demonstrate that the phage-encoded non-structural proteins P17 and P33, either individually or together, complement the defect in a temperature-sensitive GroES mutant of *Escherichia coli* for host growth and PRD1 propagation. Confocal microscopy of fluorescent fusion proteins revealed co-localisation between P33 and P17 as well as between P33 and the host chaperonin GroEL. A fluorescence recovery after photobleaching assay demonstrated that the diffusion of the P33 fluorescent fusion protein was substantially slower in *E. coli* than theoretically calculated, presumably resulting from intermolecular interactions. Our results indicate that P33 and P17 function in procapsid assembly, possibly in association with the host chaperonin complex GroEL/GroES.

Introduction

The assembly of viral capsids is a highly efficient and elaborately controlled process, which provides a useful model system to study the factors governing macromolecular assembly. The mature virion is formed after several steps, which can include proteolytic cleavages, conformational reorganisations, covalent bonding and post-translational modifications. Several accessory factors are usually needed to assure correct formation of the capsid. These factors include scaffolding proteins, which are necessary for assembly, but are expelled from the structure during the late stages of the process. However, the exact functions of other types of accessory proteins, so-called assembly factors, are largely unknown.

Interestingly, host-encoded chaperones are essential in the morphogenesis of several bacterial viruses, such as λ (Georgopoulos et al., 1972, 1973 and Sternberg, 1973), T4 (van der Vies et al., 1994), PRD1 (Hänninen et al., 1997) and RB49 (Ang et al., 2001 and Keppel et al., 2002). In fact, the GroEL/GroES complex of *Escherichia coli* (*E. coli*) was originally discovered by studying host mutants that blocked bacteriophage capsid assembly (Georgopoulos et al., 1972, 1973 and Sternberg, 1973). Phages λ and PRD1 require host-encoded GroEL and GroES in their life cycles, whereas T4 and its distant relative RB49 encode co-chaperonin orthologs Gp31 and CocO, respectively and only require GroEL. Gp31 and CocO do not share significant sequence similarity with GroES, although they are functionally analogous to GroES and can replace it in the chaperonin complex (Ang et al., 2001, Keppel et al., 2002 and van der Vies et al., 1994). In addition to GroES analogues, a few GroEL ortholog-encoding genes have been predicted from bacteriophage genomes (Cornelissen et al., 2012, Hertveldt et al., 2005). One example is the GroEL ortholog of the *Pseudomonas aeruginosa* phage EL, protein gp146, which functions as a chaperonin (Kurochkina et al., 2012). The chaperonin and co-chaperonin analogues are not the only bacteriophage-encoded gene products known to modulate the GroEL/GroES complex. Recently, protein Gp39.2 of bacteriophage RB69 was shown to suppress defects in GroEL and GroES function (Ang and Georgopoulos, 2012). However, this occurs only when either the GroEL or GroES mutants have a reduced affinity to bind to the chaperonin complex. Homologous gene products of the bacteriophages T4, RB43 and RB49 were shown to function similarly to Gp39.2 of RB69 (Ang and Georgopoulos, 2012). Evidently, many

* Corresponding author. Fax: +358 2941 59098.
E-mail address: hanna.oksanen@helsinki.fi (H.M. Oksanen).
[1] Present address: A.I. Virtanen Institute for Molecular Sciences, Department of Neurobiology, University of Eastern Finland, P.O. Box 1627, Kuopio, Finland.
[2] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.virol.2015.03.049
0042-6822/

phages utilise the host-encoded chaperonins, but some viruses have their own encoded proteins to ensure proper folding of virus-specific proteins during virus assembly under different conditions.

PRD1, which infects Gram-negative bacteria, is a well-studied model virus of the *Tectiviridae,* which includes icosahedral dsDNA viruses with an internal membrane underneath the protein capsid. The internal membrane of PRD1 is acquired from the host cytoplasmic membrane. However, its lipid composition is not identical to that of the host bacterium (Laurinavičius et al., 2004), implying that the lipid acquisition of PRD1 is selective upon particle assembly. The lipid molecules are distributed asymmetrically between the membrane leaflets, the outer leaflet being enriched in phosphatidyl glycerol and cardiolipin, whereas the inner leaflet is predominantly composed of zwitterionic phosphatidylethanolamine molecules (Cockburn et al., 2004 and Laurinavičius et al., 2007). The membrane encloses the linear dsDNA genome with covalently linked 5′ terminal proteins (Bamford et al., 1983 and Bamford and Mindich, 1984). The structure of the virion has been solved at ~4 Å resolution using X-ray crystallography (Abrescia et al., 2004 and Cockburn et al., 2004). The trimeric major coat protein (MCP) P3 is organised in a pseudo $T=25$ lattice. Vertices contain a spike complex, formed of the spike protein P5 and receptor binding protein P2, which is anchored to the capsid by the penton protein P31 (Bamford and Bamford, 2000; Grahn et al., 1999; Huiskonen et al., 2007 and Rydman et al., 1999). One of the 12 vertices possesses a unique composition of proteins required for DNA packaging (Hong et al., 2014 and Strömsten et al., 2003). PRD1 has no tail, but during infection, the internal membrane transforms to a tail tube through which the DNA is translocated into the host cell (Grahn et al., 2002 and Peralta et al., 2013).

Assembly of PRD1 proceeds through the formation of an internal membrane-containing procapsid, into which the genome is packaged. The host chaperonin complex GroEL/GroES promotes folding of the capsid proteins P3 and P5 but it also has a role in the folding/assembly pathway of virus-encoded membrane proteins (Hänninen et al., 1997). The MCP P3 uses virus-specific membrane vesicles coated with the assembly factor P10 as a platform for particle formation (Rydman et al., 2001). The tape measure protein P30 cements the P3 facets together and controls the size of the virus particle (Abrescia et al., 2004 and Rydman et al., 2001). DNA is packaged through the special vertex consisting of the packaging ATPase P9, packaging efficiency factor P6 and small membrane proteins P20 and P22 (Hong et al., 2014 and Strömsten et al., 2003). The procapsid is devoid of P9 and, in contrast to several other packaging ATPases, P9 remains in the viral particle after DNA is packaged (Mindich et al., 1982b and Strömsten et al., 2003). In addition to protein P10, PRD1 encodes the non-structural assembly factor P17, which plays an essential role in PRD1 assembly (Mindich et al., 1982b and Vilen et al., 2003). P17 is a soluble tetramer (Caldentey et al., 1999), and it binds to positively charged lipid membranes in its purified form (Holopainen et al., 2000). One hypothesis is that P17 could substitute for co-chaperonin GroES and its function could be analogous to that of Gp31 in phage T4 (van der Vies et al., 1994). However, results from ATPase assays and co-sedimentation experiments do not support this. In addition, it was shown that plasmid-produced P17 does not complement PRD1 assembly defect associated with the *groEL* mutant (Hänninen et al., 1997). Genes essential for virus assembly and particle formation reside in one late operon (OL2) in the PRD1 genome (Grahn et al., 1994). These genes encode the assembly factors P10 and P17 and packaging vertex proteins P6 and P9 together with the predicted protein P33 (Fig. 1). The coding regions of P17 and P33 overlap by a few nucleotides. The function of P33 is unknown, but its gene location indicates a role in the virus assembly.

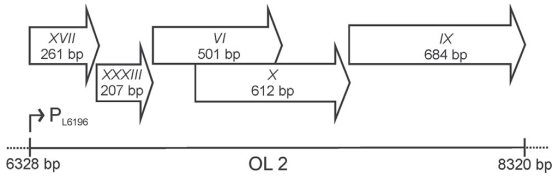


Fig. 1. Late operon OL2 (6328 to 8320 bp) of PRD1. Operon contains genes *X*, *XVII* and *XXXIII* encoding proteins for assembly (P10, P17, P33) and genes *VI* and *IX* for DNA packaging (P6, P9). P refers to the promotor.

Although the assembly process of PRD1 has been studied extensively, the functions of certain non-structural assembly factors, as well as the mechanism of the viral membrane morphogenesis are largely unknown. In the present study, we demonstrate that PRD1 proteins P17 and P33, individually or together, complement the defect in GroES of *E. coli* for bacterial growth and propagation of PRD1. Confocal microscopy of fluorescent fusion proteins revealed that the fluorescent fusion of P33 co-localises with the fluorescent fusions of GroEL and P17, which indicate possible associations between these proteins. In addition, we demonstrate that the fluorescent P33 fusion protein experiences slow mobility in *E. coli,* further suggesting that some associations are hindering its motion.

Results

PRD1 proteins P33 and P17 complement the GroES defect in E. coli

We tested whether proteins P33 and P17 of PRD1 could complement the defect in *E. coli* GroEL or GroES for bacterial growth. For this, we used *E. coli* strains DW717(pLM2) and DW719 (pLM2), which have a temperature-sensitive *groEL*59 or *groES*619 mutation, respectively, and carry the plasmid encoding the receptor of PRD1. Under non-permissive conditions, these strains cannot produce functional GroEL or GroES, causing severe growth inhibition. The mutant strains were transformed with a recombinant plasmid encoding P17, P33 or both (pSH33, pPR3 or pSH35, respectively). The positive control strains carried either plasmid pJBEL6 or pJBES4, encoding wild-type *E. coli* GroEL and GroES, respectively. The mutant strains carrying the backbone vector pSU18 of the recombinant plasmids, were used as the negative control strains [DW717(pLM2)(pSU18), DW719(pLM2)(pSU18)]. The plasmid-produced protein P17 complemented the defect of the corresponding protein in the PRD1 mutant *sus*151 (data not shown), showing that the recombinant P17 was functional. Since no mutant for PRD1 gene *XXXIII* is available, the complementation ability of plasmid-produced P33 could not be tested.

The colony-forming ability of the above-mentioned *groEL* and *groES* mutant strains carrying either a complementing plasmid or non-complementing control vector was compared to that of the wild-type strain DW720(pLM2)(pSU18) at permissive (37 °C) and non-permissive (42 °C) temperatures (Fig. 2A). At the permissive temperature, all strains formed similar numbers of colonies. When the temperature was elevated to the non-permissive temperature, neither the negative control strain DW717(pLM2)(pSU18) nor any of the *groEL* mutant strains carrying a complementing plasmid [DW717(pLM2)(pPR3), DW717(pLM2)(pSH33), DW717(pSH35)] were growing. Only the positive control strain, encoding the wild-type GroEL [DW717(pLM2)(pJBEL6)] had a similar colony number to wild-type. This demonstrates that plasmid-produced P17 and P33 can not substitute GroEL for bacterial growth.

In case of the *groES* mutant, the negative control strain DW719 (pLM2)(pSU18) formed around four orders of magnitude fewer

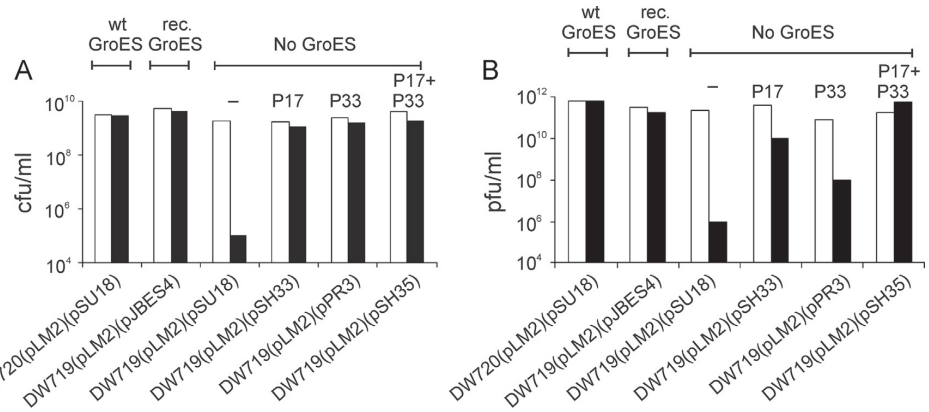


Fig. 2. Complementation of *E. coli groES* mutant strains by PRD1 *XVII* and/or *XXXIII* genes. Complementation tests were performed by spot tests. *E. coli* strain DW719(pLM2) with a temperature-sensitive mutation in *groES* gene and carrying the plasmid encoding receptor complex was complemented by a recombinant plasmid carrying a gene encoding PRD1 P17 (pSH33), PRD1 P33 (pPR3) or both (pSH35). As a comparison, the strain was transformed by a recombinant plasmid encoding a wild-type (wt) *E. coli* GroES [DW719(pLM2)(pJBES4)]. DW720(pLM2)(pSU18) is the wild-type strain and DW719(pLM2)(pSU18) was used as a negative control. The recombinant gene expression was induced with IPTG, and bacterial cultures were grown at permissive or non-permissive temperatures. (A) Colony-forming ability of *E. coli groES* mutant strains. Each *groES* mutant strain carrying one of the recombinant plasmids was incubated overnight at permissive (37 °C, white bars) or non-permissive (42 °C, black bars), after which colony formation was monitored. (B) Plaque-forming ability of bacteriophage PRD1 on *E. coli groES* mutant strains. The growth of PRD1 was similarly studied by plating dilutions on *groES* mutant strains carrying one of the recombinant plasmids. After overnight incubation at permissive (37 °C, white bars) or non-permissive (40 °C, black bars), plaques were counted.

colonies compared to wild-type (Fig. 2A). However, when the *groES* mutant carried the recombinant plasmid encoding either wild-type GroES [DW719(pLM2)(pJBES4)] or one or both PRD1 proteins P17 and P33 [DW719(pLM2)(pSH33), DW719(pLM2)(pPR3), or DW719(pLM2)(pSH35)], the number of colony-forming units was very similar to that of the wild-type (Fig. 2A). This shows that the two PRD1 proteins can restore the growth of *E. coli groES*619 mutant strain, either separately or together.

PRD1 requires a large chaperone pool in the temperature-sensitive groES mutant of E. coli at elevated temperature

The various *groEL*59 and *groES*619 mutant strains (see above) were also analysed for their ability to support the growth of PRD1. Complementation tests were performed by pipetting viral dilutions onto bacterial lawn and analysing the appearance of bacterial lysis. At the permissive temperature (37 °C), PRD1 formed a similar number of plaques on the *groEL* (data not shown) and *groES* strains (Fig. 2B) as on the wild-type strain. When temperature was elevated to non-permissive, the plaque-formation on the *groEL* mutant expressing the wild-type GroEL [DW717(pLM2)(pJBEL6] corresponded to that in wild-type (data not shown). However, the titer in the strains carrying a complementing plasmid [DW717 (pLM2)(pPR3), DW717(pLM2)(pSH33), DW717(pSH35)] was considerably lower, resembling the titer of the negative control [DW717(pLM2)(pSU18)], implying that the expression of P17 and P33 does not complement the defect in GroEL for PRD1 growth.

Interestingly, the situation was different for *groES* mutants: At the non-permissive temperature (40 °C), the number of plaque-forming units was almost the same in the mutant strains expressing wild-type GroES [DW719(pLM2)(pJBES4)] or co-expressing P17 and P33 [DW719(pLM2)(pSH35)] as in the wild-type strain. Somewhat lower titres were seen in strains DW719(pLM2)(pSH33) and DW719(pLM2)(pPR3), which express either P17 or P33, respectively. However, the number of plaques with these strains was several (2 – 4) orders of magnitude higher than with the negative control [DW719(pLM2)(pSU18)]. These data indicate that the phage-encoded proteins P17 and P33, produced from both viral and plasmid transcripts, enable PRD1 to propagate in the GroES deficient *E. coli* strain.

Similarly, we tested whether the expression of PRD1 P33 [HMS174(pLM2)(pPR3)] could complement the defective phenotype of PRD1 mutant *sus*151. In this strain, the titer of *sus*151 was the same as in the negative control strain HMS174(pLM2)(pSU18), whereas in strains expressing P17 [HMS174(pLM2)(pSH33)] or co-expressing P17 and P33 [HMS174(pLM2)(pSH35)], the titer was about six orders of magnitude higher. This implies that the over-production of P33 can not substitute for P17 in PRD1 assembly.

Protein P33 co-localises with P17 and GroEL in E. coli

Fluorescent fusions of PRD1 proteins P33 and P17, as well as *E. coli* GroEL, were produced to study their cellular localisations in *E. coli*. In these fusions, either eYFP or eCFP was attached to the C-terminus of the viral protein using a linker of six glycine residues. The expressed P33, P17 and GroEL fusion proteins were mainly soluble and were assayed by rate zonal centrifugation. P33-eYFP (34.4 kDa) was expressed either as a monomer or a small multimer (Fig. 3A), similarly to P17-eCFP (36.4 kDa; Fig. 3B). When P33-eYFP and P17-eCFP were co-expressed, no major change was detected in the mobility of P33-eYFP (Fig. 3C). Also, the major signals of P17-eCFP were detected in the same positions as when the fusion protein was produced alone (Fig. 3B and C). However, the difference in the production levels hinder interpretation leaving the question of the P17–P33 complex formation open. According to the sedimentation data, GroEL-eCFP (83.9 kDa) formed multimers similarly to endogenous GroEL (Fig. 3D and E). However, compared to endogenous GroEL, a portion of GroEL-eCFP sedimented slower. Although most of the recombinant protein was found in the soluble fractions, there was also a portion in the pellet, suggesting partial misfolding and/or aggregation.

Localisation of the P33, P17 and GroEL fusion proteins, as well as co-localisation of different fusion protein pairs, was monitored in *E. coli* cells using confocal microscopy. The fusion protein P33-eYFP was evenly distributed throughout the cell in all parallel samples (Fig. 4A). P17-eYFP formed clear polar loci as previously reported [Fig. 4B; (Karttunen et al., 2014)]. The production of the

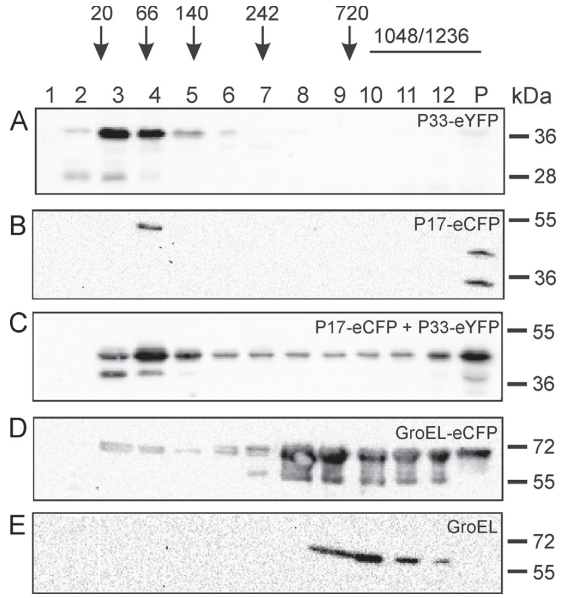


Fig. 3. Oligomerization/complex formation of (A) P33-eYFP, (B) P17-CFP, (C) P33-eYFP and P17-eCFP, (D) GroEL-eCFP and (E) endogenous GroEL analysed using rate zonal centrifugation in sucrose, SDS-PAGE and Western blotting. The fluorescent fusion proteins P33-eYFP, P17-eCFP and GroEL-eCFP were produced and P33-eYFP and P17-eCFP co-produced in *E. coli* strain HMS174(DE3). The fractions (1–12) and the pellet (P) of the gradients are shown. The molecular masses (kDa) and the positions of the standard proteins for centrifugation are shown with arrows on top and for SDS-PAGE on right. GFP(FL) HRP (Santa Cruz Biotechnology) was used as a primary antibody for P33-eYFP and P17-eCFP, and polyclonal anti-Hsp60 (anti-GroEL) serum (Poranen et al., 2006) for GroEL-eCFP and endogenous GroEL.

GroEL-eCFP fusion protein in *E. coli* led to two different localisation patterns in parallel cultures. The fusion was either evenly distributed around the bacterial cell or localised into the cell poles (Fig. 4C). We also observed that the cells were occasionally elongated that is possibly resulting from a shock reaction (data not shown). When P33-eYFP was co-produced either with P17-eCFP or GroEL-eCFP, localisation of P33-eYFP was changed to a co-localisation signal at the cell poles (Fig. 4D and E). No co-localisation was observed when eYFP was co-produced with either P17-eCFP or GroEL-eCFP (data not shown). This indicates that the co-localisation may result from the association of P33 with either P17 or GroEL.

Diffusion dynamics of P33-eYFP in E. coli

To study the mobility of P33-eYFP in a bacterial cell, we followed the recovery of fluorescence after photobleaching (FRAP assay). In addition, we conducted computer simulations from the measured recovery curves. The mobility of P33-eYFP was compared to that of eYFP or P33-eYFP co-produced together with either GroEL-eCFP or P17-eCFP. eYFP was used as a control to rule out changes not caused by P33. In all samples, only the fluorescence emission of eYFP was monitored regardless of the presence of cyan fluorescent protein. Although P17-eCFP or GroEL-eCFP localised into specific loci in some cells, only cells with evenly distributed fluorescence were selected for this assay (Fig. 5).

The selected areas at the end of an individual cell were photobleached and the recovery was monitored (Fig. 5A and B). According to the averaged ($n=13$) recoveries and computer simulations, the diffusion coefficient (D) of eYFP was determined to be 5 $\mu m^2/s$ (Fig. 5C), which is in accordance with the reported diffusion coefficient values of 7.7 ± 2.5 $\mu m^2/s$ for GFP (Elowitz et al., 1999) or 4.6 ± 0.8 $\mu m^2/s$ for CheY-GFP (Cluzel et al., 2000). According to mass scaling, the theoretical diffusion coefficient for free P33-eYFP (34.4 kDa) is 4.6 $\mu m^2/s$. The production of P33-eYFP in *E. coli* resulted in two distinct populations with different diffusion constants. The measured recovery of P33-eYFP ($D=2$ $\mu m^2/s$) in the faster population (91% of the total population) differed significantly from the theoretical value (Fig. 5C). This coefficient equals a complex with a mass of approximately 400 kDa. In addition, the simulations revealed that the other population representing 9% of the total population was slower with $D=0.4$ $\mu m^2/s$ (Fig. 5C). The co-production of P33-eYFP with either GroEL-eCFP or P17-eCFP (Fig. 5D) changed the mobility of P33-eYFP even more. In both cases, the ratio between the faster and slower populations was changed. The co-production of P33-eYFP with GroEL-eCFP raised the percentage of the slower-moving population of P33-eYFP from 9% to 40%, whereas co-production with P17-eCFP increased the percentage of the slow population to 67%. When eYFP was co-produced with GroEL-eCFP or P17-eCFP, the recoveries were same as observed for eYFP alone (Fig. 5D).

Discussion

The assembly of the tailless icosahedral viruses with an internal membrane possesses additional complexity when compared to widely-studied head-tailed viruses with protein and nucleic acid components only. Evidently, the incorporation of the virus-specific membrane patch to the forming particles is a crucial step. For bacteriophage PRD1, a comprehensive model of the virion formation has been proposed (reviewed in Atanasova et al. (2015) and Grahn et al. (2006)), and it serves as an assembly model for other icosahedral, internal membrane-containing viruses infecting bacteria, archaea or eukaryotes.

During PRD1 assembly, the first events are visible about 15 min after infection (Mindich et al., 1982b), when the MCP P3 and the spike-complex proteins P2, P5 and P31 appear in a soluble form in the host cytosol, whereas the phage-encoded membrane proteins (P7, P11, P14 and P18 etc.) are found in the host cytoplasmic membrane (Mindich et al., 1982b). It is known that the GroEL/GroES chaperonin complex is essential for correct folding of the capsid proteins P3 and P5 as well as the assembly/folding pathway of the membrane proteins (Hänninen et al., 1997). Most probably, the virus-specific membrane patches in the cytoplasmic membrane act as nucleation sites for procapsid formation. It has been proposed, that the non-structural scaffolding protein P10 attaches to the host cell membrane, and initiates membrane budding together with the MCP P3 trimers. Consequently, the virus-specific batch of the cell membrane is pinched off, similarly to clathrin-coated pits in eukaryotic cells. In this process, lipid vesicles formed are first covered with P10 and a small portion of P3, and then P10 is replaced by the tape measure protein P30, which stabilises interactions between the capsomers (Rydman et al., 2001). Finally, the resulting procapsid is packaged with the viral DNA powered by the packaging ATPase P9, which stays in the mature virion (Hong et al., 2014; Mindich et al., 1982b and Strömsten et al., 2003). In addition to P10, two other non-structural proteins P17 [Fig. 2B; (Mindich et al., 1982b)] and P33 function in virion formation (Fig. 2B). P33 and P17 enabled PRD1 to propagate in *E. coli* GroES temperature-sensitive mutant *groES*619, which would otherwise block the bacteriophage multiplication (Fig. 2B). The proteins together restored the phage growth more effectively than each alone. In addition, P33 and P17 complemented the growth defect of *E. coli* (*groES*619) when expressed either individually or simultaneously (Fig. 2A). This effect of P33 and P17 seems to be specific to GroES, since their expression did not

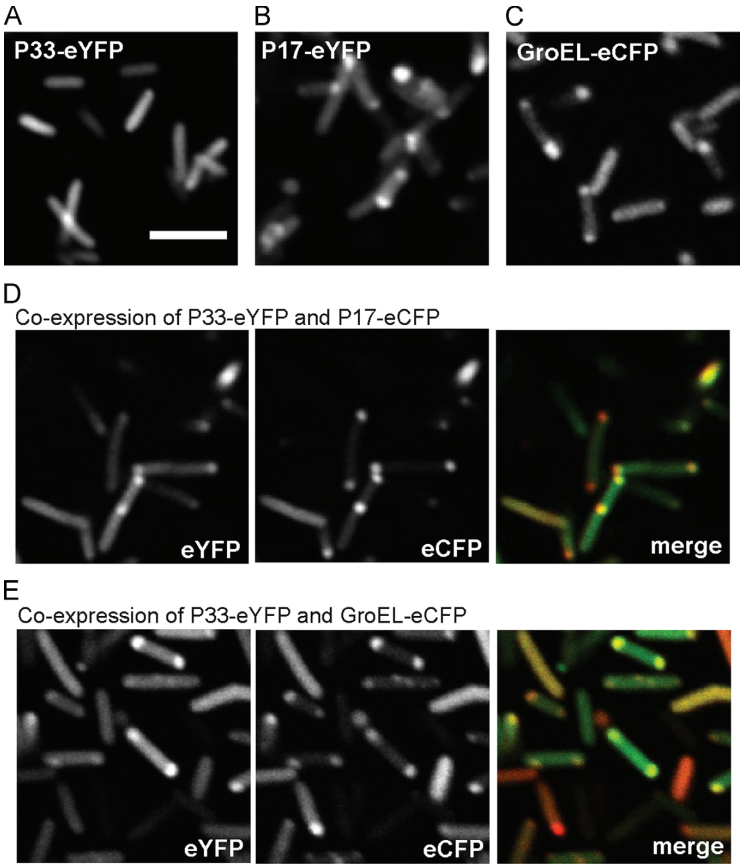


Fig. 4. Confocal microscope images of proteins (A) P33-eYFP, (B) P17-eYFP and (C) GroEL-eCFP produced in HMS174(DE3) cells and measured by the corresponding wavelengths. (D) Co-production of P33-eYFP and P17-eCFP measured by the eYFP (left) and eCFP (centre) detection channels; the merged channels are on the right, where P33-eYFP and P17-eCFP are in green and red and co-localised pixels are in yellow. (E) Co-production of P33-eYFP and GroEL-eCFP. The eYFP and eCFP detection channels and the merged channels are on the left, centre and right, where P33-eYFP and GroEL-eCFP are in green and red and co-localised pixels are in yellow. The scale bar (5 μm) in A is applicable to all images.

complement the defect of an *E. coli* GroEL temperature-sensitive mutant *groEL*59. Our data indicate that both P33 and P17 guide PRD1 particle formation, and their functions may be related to that of the GroEL/GroES chaperonin complex. However, the inability of plasmid-produced PRD1 P33 to complement the defect of PRD1 P17 mutant *sus*151 suggests that these two proteins do not have a functional redundancy in PRD1 assembly. To generate amber mutations in the PRD1 genome by targeted in vitro mutagenesis is challenging (Bamford and Bamford, 2000) and to date, no PRD1 P33 deficient mutant is available. Therefore, it is not known, whether P33 is essential for PRD1 propagation in standard laboratory conditions. However, the gene *XXXIII* does not tolerate transposon insertion (Vilen et al., 2003) implying that the function of the gene product is essential.

The localisation of GroEL inside bacterial cells has puzzled researchers for decades. It has been suggested to be diffusely distributed in the cytoplasm (Charbon et al., 2011 and Winkler et al., 2010), co-localised with the cell-division protein FtsZ at the cleavage furrow (Ogino et al., 2004), or trapped within inclusion bodies (Carrio and Villaverde, 2005). In this study, a GroEL-eCFP fusion protein was either seen evenly distributed around the cell or localised into the cell poles of *E. coli* (Fig. 4C). However, due to the structural constraints in GroEL, it is possible that its folding and, consequently, its function might be altered with the fluorescent protein attached. Confocal microscopy of fluorescent fusion proteins also demonstrated that P33 alone was evenly distributed along the cell, whereas P33 in the presence of P17 or GroEL resulted in polar co-localisation (Fig. 4), indicating possible associations between these proteins. The polar co-localisation could result from aggregation rather than the tendency of the proteins to interact (Lindner et al., 2008; Lloyd-Price et al., 2012). However, the low-copy number plasmid pSU18 used in this study produces proteins at a moderate level, and it has been widely utilised to produce functional PRD1 proteins (Bamford and Bamford, 2000; Bartolomé et al., 1991 and Rydman et al., 2001). Therefore, it seems unlikely that all the observed fusion proteins were aggregating.

In FRAP assay, the mobility of the P33-eYFP in *E. coli* was considerably slower than theoretically calculated (Fig. 5), indicating that some factors hinder its motion. Interestingly, the diffusion coefficient of the slower moving major population of P33-eYFP notably resembles the reported reference value of the GroEL complex ($0.4\ \mu m^2/s$ versus $0.16 \pm 0.15\ \mu m^2/s$, respectively) (Charbon et al., 2011), which might be explained by the association between the fusion protein and endogenous GroEL. The relatively low diffusion constant of the major population (compared to the theoretically calculated value of P33-eYFP) corresponds to a complex where the P33-eYFP is bound to a partner with a mass of approximately

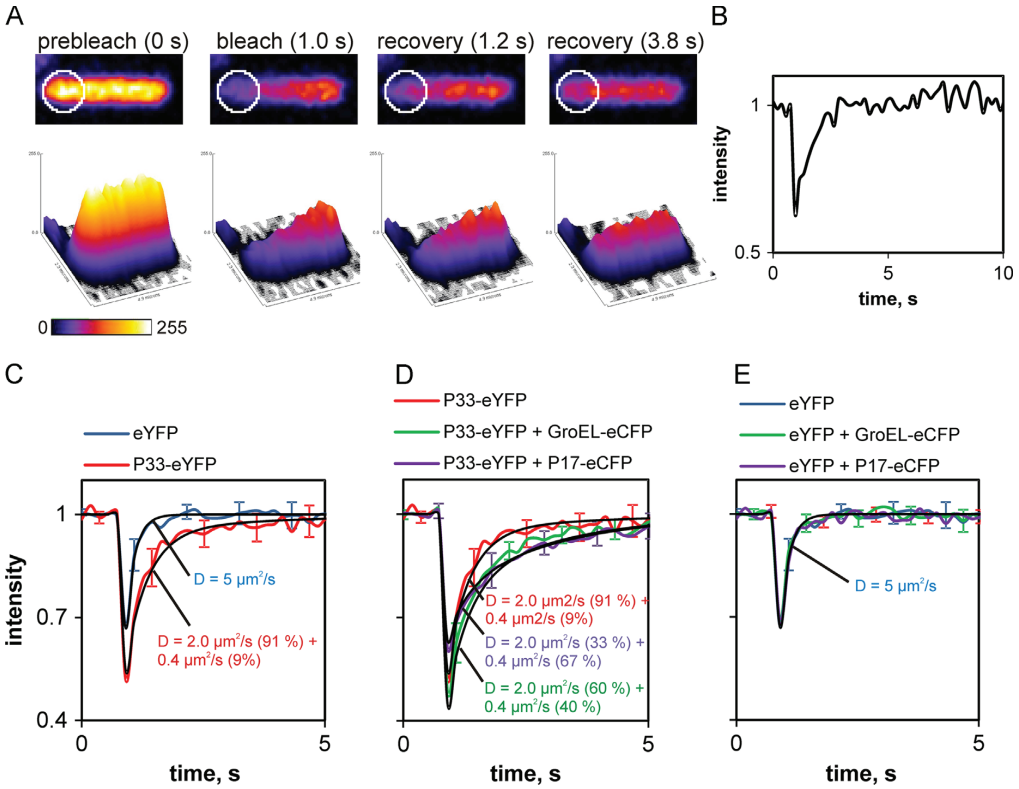


Fig. 5. Diffusion dynamics of P33-eYFP. (A) An example of one measured cell co-producing GroEL-eCFP and P33-eYFP. The area marked with the white line was photobleached with 100% laser intensity, and the recovery of the eYFP fluorescence was monitored. The scale of the used lookup table (LUT) is presented. (B) The normalised fluorescence recovery curve from the single cell shown in section A. (C) eYFP and P33-eYFP fluorescence recoveries (blue and red, respectively) and the corresponding simulation curves (black). The diffusion coefficients (D) of the simulations are presented in C, D and E. (D) P33-eYFP fluorescence recoveries expressed individually (red), in co-expression with GroEL-eCFP (green) or P17-eCFP (violet) and corresponding simulation curves (black). (E) eYFP fluorescence recoveries expressed individually (blue) and in co-expression with GroEL-eCFP (green) or P17-eCFP (violet).

400 kDa, which resembles the mass of a GroEL heptamer of 420 kDa. Co-production of P33-eYFP with GroEL-eCFP increased the percentage of the slower population from 9% to 40% (Fig. 5). This might be explained by the increase in the number of association partners for P33-eYFP. Interestingly, when P33-eYFP was produced together with P17-eCFP, the percentage of its slower population increased to 67%, which indicates an interaction between the fusions.

Bacteriophage T4 and RB49 encoded proteins Gp31 and CocO, respectively, can act as co-chaperonins and replace GroES (Ang et al., 2001; Keppel et al., 2002 and van der Vies et al., 1994). However, P17 or P33 do not share any significant level of sequence identity with *E. coli* GroES co-chaperonin and those bacteriophage-encoded homologues. In GroES, three consecutive hydrophobic residues located in a mobile loop are known to interact with GroEL (Landry et al., 1993, 1996 and Xu et al., 1997). This region seems to be conserved within the previously identified phage-encoded co-chaperonins (Ang et al., 2001; Hunt et al., 1997; Landry et al., 1996 and Richardson and Georgopoulos, 1999), but one could not be found in P17 or P33.

E. coli mutant strain (*groES*619) produces a mutant GroES protein, in which a glycine preceding the hydrophobic loop is substituted for an aspartic acid resulting in a reduced affinity for GroEL interaction (Landry et al., 1993). When screening for different *groES* and *groEL* mutants blocking the bacteriophage propagation, most mutations interfere with normal GroEL-GroES interaction by either reducing the affinity to the corresponding binding partner (chaperonin or co-chaperonin) or prolonging the interaction (Ang and Georgopoulos, 2012). Obviously, mutations causing complete abolishment of the GroES function would be detrimental for the cell. Bacteriophage T4 and its close relatives RB69, RB43 and RB49 encode a gene product (Gp39.2) that can suppress defects in either GroES or GroEL, when the mutation is weakening interaction between them (Ang and Georgopoulos, 2012). It has been suggested that Gp39.2 accomplishes this by promoting an open conformation of GroEL (Ang and Georgopoulos, 2012). It is possible that PRD1 P33 and P17 function similarly. However, no physical interaction between P17 and GroEL has been detected (Hänninen et al., 1997). This does not eliminate the possibility that there could be a direct contact between P33 and GroEL or that GroEL interacts with P17 and P33 indirectly, possibly mediated by some other cellular or phage-encoded molecules.

Hänninen et al., 1997 demonstrated that in certain *E. coli* mutants of GroEL or GroES, such as in mutant *groES*619, non-infectious PRD1 virus particles were formed, probably due to the reduced amounts of small membrane-associated proteins and major membrane protein P11 in the virion. This suggests that it might be possible that P17 and P33 function later in assembly, perhaps during the viral membrane morphogenesis. The viral membrane may be compromised in *groES*619 mutant, thus contributing to the observed defective virus assembly. It is known that GroEL interacts with the isolated cytoplasmic membranes of *E. coli*

by specifically targeting to SecA, and thus participates directly or indirectly in proper protein transport and membrane composition (Bochkareva et al., 1998). In addition, P17 binds to positively charged membranes (Holopainen et al., 2000). Although gram-negative bacteria do not have positively charged lipids, such as sphingosine, P17 might interact with the membrane through the cationic moieties of the lipid head groups and/or the basic domains of certain membrane proteins (Holopainen et al., 2000) or via an interaction with the membrane-associated GroEL. Overproduction of recombinant P17 as well as P33 may complement the *groES*619 defect indirectly, perhaps by mass action, enabling an otherwise compromised assembly step to proceed.

The functions of PRD1 proteins P17 and P33 exemplifies how viruses can refine host machineries for their own purposes. Interestingly, these phage-encoded proteins P17 and P33 restored the bacterial growth even more efficiently than the virus propagation. This demonstrates that the host cell can correspondingly evolve to exploit viral proteins.

Materials and methods

Bacteria, phages and plasmids

The bacterial strains, phages and plasmids used in this study are listed in Table 1. The cells were grown in Luria-Bertani (LB) medium with antibiotic supplements when appropriate (25 μg/ml chloramphenicol or kanamycin; 10 μg/ml tetracycline). Wild-type PRD1 was grown in *Salmonella enterica* serovar Typhimurium DS88 as described earlier (Bamford and Bamford, 1991). PRD1 mutant *sus*151 (amber mutation in gene *XVII*; Mindich et al., 1982a) was propagated on *Salmonella* Typhimurium suppressor strains PSA (*supE*) or DB7154 (*supD*) harbouring the pLM2 plasmid. The PRD1 genes *XVII* and *XXXIII* (encoding proteins P17 and P33, respectively) were amplified by PCR using the phage genome as a template and specific primers hybridising to the ends of the genes. The fragments were cloned into the pSU18 vector between the *Eco*RI and *Hind*III restriction sites. The functionality of plasmid-produced P17 was tested with the mutant virus *sus*151 by an in vivo complementation assay in *E. coli* HMS174(pLM2) using plasmid pSH33, encoding the PRD1 wild-type gene *XVII*. Similarly, the plasmids pPR3 (including PRD1 gene *XXXIII*) and pSH35 (including both PRD1 genes *XVII* and *XXXIII*) were introduced into HMS174(pLM2) strain. The constructs were also used in complementation assays in *E. coli* temperature-sensitive *groEL*59 and *groES*619 mutant strains. Protein production was induced by adding 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). When monitoring the colony-forming ability, the *groEL* and *groES* mutant strains were incubated at either permissive (37 °C) or non-permissive (42 °C) temperatures. When analysing the plaque formation on the mutant strains, the non-permissive temperature was decreased to 40 °C due to the growth requirement of the virus. To construct plasmids for the production of fluorescent fusion proteins, the *E. coli* gene *groEL* and the PRD1 genes *XVII* and *XXXIII* were amplified using plasmid pOF39 (Fayet et al., 1986) and the PRD1 genome as template, respectively, with specially designed primers binding to the ends of the genes. The amplicons were ligated into vectors pKM64 and pKM41 between the *Xba*I and

Table 1
Bacterial strains, phages and plasmids used in this study.

Bacterial strains, phages and plasmids	Relevant genotype or description (nt coordinates in PRD1 genome)[a]	Relevant phenotype	Source or reference
Bacterial strains			
Salmonella enterica Typhimurium LT2			
DS88	SL5676 Δ*H2 H1-i:* Tn *10* (Tc^S)(pLM2)	Nonsuppressor host	Bamford and Bamford, 1990
PSA(pLM2)	*sup*E	Suppressor host	Mindich et al., 1982a
DB7154(pLM2)	*supD*	Suppressor host	Winston et al., 1979
Escherichia coli K-12			
DW720	Wild-type		Sherwood Casjens, University of Utah, US
DW717	*groEL*59 (ts)		Sherwood Casjens
DW719	*groES*619 (ts)		Sherwood Casjens
HB101	*supE*44 *hsdS*20 ($r_B^-m_B^-$) *recA*13 *ara*14 *proA*2 *lacY*1 *galK*2 *rpsL*20 *xyl*5 *mtl*1	Cloning host	Bolivar and Backman, 1979
HMS174	*recA*1 *hsdR* Rf^r	Expression host	Campbell et al., 1978
HMS174(DE3)	*recA*1 *hsdR* Rf^r	Expression host	Campbell et al., 1978
Phages			
PRD1	Wild-type		Olsen et al., 1974
PRD1 *sus*151	Amber mutation in gene *XVII*		Mindich et al., 1982a
Plasmids			
pLM2	Encodes PRD1 receptor; IncPα replicon, Km^r		Mindich et al., 1976
pSU18	Low-copy-number cloning vector; p15A replicon, Cm^r		Bartolomé et al., 1991
pET24	High-level-expression vector; ColE1 replicon, Km^r		Novagen
pSH35	pSU18+PRD1 genes *XVII, XXXIII*		This study
pSH33	pSU18+PRD1 gene *XVII*		This study
pPR3	pSU18+PRD1 gene *XXXIII*		This study
pJBEL6	pSU18+gene *groEL*		Hänninen et al., 1997
pJBES4	pSU18+gene *groES*		Hänninen et al., 1997
pOF39	genes *groEL*+*groES*		Fayet et al., 1986
pKM41	pET24+gene *ecfp*+6 Gly linker	Cloning vector	Karttunen et al., 2014
pKM64	pSU18+gene *eyfp*+6 Gly linker	Cloning vector	Karttunen et al., 2014
pJK7	pSU18+gene *eyfp*	eYFP	Karttunen et al., 2014
pSSM43	pKM64+PRD1 gene *XVII*	P17-eYFP	Karttunen et al., 2014
pSN1	pKM41+gene *groEL* from pOF39	GroEL-eCFP	This study
pSSM51	pKM64+PRD1 gene *XXXIII*	P33-eYFP	This study
pAK17	pKM41+PRD1 gene *XVII*	P17-eCFP	This study

[a] GenBank Acc No AY848689 (Bamford et al., 1991; Saren et al., 2005).

*Bam*HI restriction sites by fusing them at the 5′end of the gene encoding yellow or cyan fluorescent protein. Fusion proteins were produced in *E. coli* HMS174(DE3). Cells were grown at 28 °C, induced at A_{550}=0.75 by adding 1 mM IPTG and grown overnight at 18 °C.

Complementation assays

When analysing complementation by PRD1 P17 and P33, the temperature-sensitive *E. coli groEL*59 mutant DW717(pLM2) and *groES*619 mutant DW719(pLM2), both of which encode the PRD1 receptor complex, were used. The mutant strains were transformed with a low-copy number recombinant plasmid encoding PRD1 P17, P33 or both (pSH33, pPR3 or pSH35, respectively) or, alternatively, wild-type *E. coli* GroEL or GroES (pJBEL6 or pJBES4, respectively) (Table 1). Growth of *E. coli* mutant strains was monitored by serially diluting overnight cultures (10-fold dilution steps) in LB broth and then pipetting a spot of 10 μl of each dilution on selective LB plates. After overnight incubation at permissive (37 °C) or non-permissive (42 °C) temperatures, bacterial colonies were counted.

Growth of bacteriophage PRD1 on *E. coli groEL* or *groES* mutant strains was similarly studied by plating 10 microliters of bacteriophage serial dilutions on mutant strains growing as lawns on LB plates. The plaque formation was analysed after overnight incubation, at permissive (37 °C) and non-permissive (40 °C) temperatures.

Solubility and oligomerization of the fusion proteins

Strains producing fusion proteins [HMS174(DE3) containing the corresponding plasmids; Table 1] were cultured as described above and concentrated 1:50 by centrifugation (Sorvall SLA3000 rotor, 4200*g*, 10 min, 5 °C) in 20 mM Tris–HCl, pH 7.2. Cells were disrupted by sonication, and cell debris was removed by centrifugation (Thermo IEC MicroCL 17 centrifuge, 17000*g*, 5 min, 22 °C). The pellet was resuspended in 50 mM Tris–HCl, pH 7.2, and the protein ratio between the supernatant and pellet was determined using sodium dodecyl sulphate polyacrylamide gel electrophoresis [SDS-PAGE, 15% (w/v) acrylamide] (Olkkonen and Bamford, 1989) and Western blotting. SDS-PAGE and Western blotting were carried out as previously described (Karttunen et al., 2014). GFP (FL) HRP (Santa Cruz Biotechnology) and a polyclonal anti-Hsp60 (anti-GroEL) serum (Poranen et al., 2006) were used as primary antibodies.

For the oligomerization/complex formation assay by rate zonal centrifugation, the supernatant was applied on a linear 10 – 40% (w/v) sucrose gradient in 20 mM Tris–HCl, pH 7.2 and then centrifuged (Beckmann SW41 rotor, 210,000*g*, 21 h, 15 °C). A NativeMark™ Unstained Protein Standard (Life Technologies) was used as a molecular mass marker. After centrifugation, twelve 1 ml fractions and the pellet were collected and analysed by SDS-PAGE and Western blotting.

Microscopy

Live samples for confocal microscopy were prepared and imaged as described in Karttunen et al., 2014. Images were captured with an image size of 512 × 512 pixels. For fluorescence recovery after photobleaching (FRAP) experiments, images were captured with an image size of 128 × 128 pixels. A time-lapse sequence of eYFP images consisting of 100 frames was taken. After capturing five frames, a circular area of the cells with a diameter of 1.5 μm was photobleached with a 515 nm laser (one iteration of 100% laser power for 8 μs/pixel) and the time-lapse continued for 95 frames. Image data was analysed using ImageJ software and an average of 13 cells was calculated. FRAP data was normalised according to the following formula (Phair and Misteli, 2000): $F(t)=(\mathrm{ROI}(t)/\mathrm{ROI}(0))/(\mathrm{Bacteria}(t)/\mathrm{Bacteria}(0))$, where fluorescence at time t [$F(t)$] was calculated from fluorescence of the bleached region of interest (ROI) at time t [ROI(t)], average fluorescence of the bleached ROI before bleaching [ROI(0)], total fluorescence of the bacteria at time t [Bacteria(t)] and average fluorescence of the bacteria before bleaching [Bacteria(0)].

Virtual cell simulations

Virtual Cell software was used to simulate the FRAP recoveries (Schaff et al., 1997). Bacterial cell geometry was simulated with a rectangle (1 μm wide and 3 μm high). The time step of the simulation was 20 ms, and images were collected with 180 ms time steps to have the same frame rate as in the FRAP experiments. The bleaching pulse was modelled using a laser light-induced reaction, and the bleaching ROI was set to correspond to a 1.3-μm area at the end of the cell. The bleach pulse was adjusted to 54 ms, and the first recovery image was collected 90 ms after the bleach phase to simulate image acquisition of the FRAP experiments. Free diffusions of eYFP and P33-eYFP were simulated with a single diffusive population, but when they were co-produced with GroEL or P17 fusions, two diffusive components were used in order to reproduce the FRAP recovery. The determined P33-eYFP free diffusion coefficient was fixed as the diffusion coefficient of the faster diffusing population for the simulations of the co-expressed protein samples. The relative concentrations of the diffusive components and the diffusion coefficient of the slower population were adjusted until the recovery curve fitted the data.

Acknowledgements

We thank Dr. Pia Niinimäki, Sari Korhonen, Sanna Niinivehmas, Annika Kohvakka and Petri Papponen for their valuable help in constructing the strains. This work was supported by the Academy of Finland Centre of Excellence Program in Virus Research (2006–2011 J.K.H.B.), the Centre of Excellence Program in Biological Interactions (#252411, 2012–2017 J.K.H.B.), Academy of Finland Grant 251106 (J.K.H.B.) and Grant from the Finnish Cultural Foundation (00110600, S.M.). We thank Academy of Finland (Grants 271413 and 272853) and University of Helsinki for the support to EU ESFRI Instruct Centre for Virus Production (ICVIR).

References

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.H., 2004. Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432, 68–74.

Ang, D., Georgopoulos, C., 2012. An ORFan no more: the bacteriophage T4 39.2 gene product, NwgI, modulates GroEL chaperone function. Genetics 190, 989–1000.

Ang, D., Richardson, A., Mayer, M.P., Keppel, F., Krisch, H., Georgopoulos, C., 2001. Pseudo-T-even bacteriophage RB49 encodes CocO, a cochaperonin for GroEL, which can substitute for *Escherichia coli*'s GroES and bacteriophage T4's Gp31. J. Biol. Chem. 276, 8720–8726.

Atanasova, N.S., Senčilo, A., Pietilä, M.K., Roine, E., Oksanen, H.M., Bamford, D.H., 2015. Comparison of lipid-containing bacterial and archaeal viruses. Adv. Virus Res. 92, 1–61.

Bamford, J.K.H., Bamford, D.H., 1990. Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane. Virology 177, 445–451.

Bamford, J.K.H., Bamford, D.H., 1991. Large-scale purification of membrane-containing bacteriophage PRD1 and its subviral particles. Virology 181, 348–352.

Bamford, J.K.H., Bamford, D.H., 2000. A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to the vertex. J. Virol. 74, 7781–7786.

Bamford, J.K.H., Hänninen, A.-L., Pakula, T.M., Ojala, P.M., Kalkkinen, N., Frilander, M., Bamford, D.H., 1991. Genome organization of membrane-containing bacteriophage PRD1. Virology 183, 658–676.

Bamford, D., McGraw, T., MacKenzie, G., Mindich, L., 1983. Identification of a protein bound to the termini of bacteriophage PRD1 DNA. J. Virol. 47, 311–316.

Bamford, D.H., Mindich, L., 1984. Characterization of the DNA-protein complex at the termini of the bacteriophage PRD1 genome. J. Virol. 50, 309–315.

Bartolomé, B., Jubete, Y., Martinez, E., de la Cruz, F., 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene 102, 75–78.

Bochkareva, E.S., Solovieva, M.E., Girshovich, A.S., 1998. Targeting of GroEL to SecA on the cytoplasmic membrane of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 95, 478–483.

Bolivar, F., Backman, K., 1979. Plasmids of *Escherichia coli* as cloning vectors. Methods Enzymol. 68, 245–267.

Caldentey, J., Hänninen, A.L., Holopainen, J.M., Bamford, J.K.H., Kinnunen, P.K., Bamford, D.H., 1999. Purification and characterization of the assembly factor P17 of the lipid-containing bacteriophage PRD1. Eur. J. Biochem. 260, 549–558.

Campbell, J.L., Richardson, C.C., Studier, F.W., 1978. Genetic recombination between bacteriophage T7 and cloned fragments of T7 DNA. Proc. Natl. Acad. Sci. USA 75, 2276–2280.

Carrio, M.M., Villaverde, A., 2005. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J. Bacteriol. 187, 3599–3601.

Charbon, G., Wang, J., Brustad, E., Schultz, P.G., Horwich, A.L., Jacobs-Wagner, C., Chapman, E., 2011. Localization of GroEL determined by in vivo incorporation of a fluorescent amino acid. Bioorg. Med. Chem. Lett. 21, 6067–6070.

Cluzel, P., Surette, M., Leibler, S., 2000. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. Science 287, 1652–1655.

Cockburn, J.J., Abrescia, N.G., Grimes, J.M., Sutton, G.C., Diprose, J.M., Benevides, J.M., Thomas Jr, G.J., Bamford, J.K.H., Bamford, D.H., Stuart, D.I., 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature 432, 122–125.

Cornelissen, A., Hardies, S.C., Shaburova, O.V., Krylov, V.N., Mattheus, W., Kropinski, A.M., Lavigne, R., 2012. Complete genome sequence of the giant virus OBP and comparative genome analysis of the diverse φKZ-related phages. J. Virol. 86, 1844–1852.

Elowitz, M.B., Surette, M.G., Wolf, P., Stock, J.B., Leibler, S., 1999. Protein mobility in the cytoplasm of *Escherichia coli*. J. Bacteriol. 181, 197–203.

Fayet, O., Louarn, J.M., Georgopoulos, C., 1986. Suppression of the *Escherichia coli dnaA*46 mutation by amplification of the *groES* and *groEL* genes. Mol. Gen. Genet. 202, 435–445.

Georgopoulos, C.P., Hendrix, R.W., Casjens, S.R., Kaiser, A.D., 1973. Host participation in bacteriophage lambda head assembly. J. Mol. Biol. 76, 45–60.

Georgopoulos, C.P., Hendrix, R.W., Kaiser, A.D., Wood, W.B., 1972. Role of the host cell in bacteriophage morphogenesis: effects of a bacterial mutation on T4 head assembly. Nat. New Biol. 239, 38–41.

Grahn, A.M., Bamford, J.K.H., O'Neill, M.C., Bamford, D.H., 1994. Functional organization of the bacteriophage PRD1 genome. J. Bacteriol. 176, 3062–3068.

Grahn, A.M., Butcher, S., Bamford, J.K.H., Bamford, D.H., 2006. PRD1-dissecting the genome structure and entry. In: Calendar, R. (Ed.), The Bacteriophage. Oxford University Press, New York, pp. 161–170.

Grahn, A.M., Caldentey, J., Bamford, J.K.H., Bamford, D.H., 1999. Stable packaging of phage PRD1 DNA requires adsorption protein P2, which binds to the IncP plasmid-encoded conjugative transfer complex. J. Bacteriol. 181, 6689–6696.

Grahn, A.M., Daugelavicius, R., Bamford, D.H., 2002. Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. Mol. Microbiol. 46, 1199–1209.

Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N., Kurochkina, L., Korchevskii, R., Robben, J., Mesyanzhinov, V., Krylov, V.N., Volckaert, G., 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. J. Mol. Biol. 354, 536–545.

Holopainen, J.M., Saily, M., Caldentey, J., Kinnunen, P.K., 2000. The assembly factor P17 from bacteriophage PRD1 interacts with positively charged lipid membranes. Eur. J. Biochem. 267, 6231–6238.

Hong, C., Oksanen, H.M., Liu, X., Jakana, J., Bamford, D.F., Chiu, W., 2014. A structural model of the genome packaging process in a membrane-containing double-stranded DNA virus. PLoS Biol. 12, e1002024.

Huiskonen, J.T., Manole, V., Butcher, S.J., 2007. Tale of two spikes in bacteriophage PRD1. Proc. Natl. Acad. Sci. USA 104, 6666–6671.

Hunt, J.F., van der Vies, S.M., Henry, L., Deisenhofer, J., 1997. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. Cell 90, 361–371.

Hänninen, A.L., Bamford, D.H., Bamford, J.K.H., 1997. Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. Virology 227, 207–210.

Karttunen, J., Mäntynen, S., Ihalainen, T.O., Lehtivuori, H., Tkachenko, N.V., Vihinen-Ranta, M., Ihalainen, J.A., Bamford, J.K.H., Oksanen, H.M., 2014. Subcellular localization of bacteriophage PRD1 proteins in *Escherichia coli*. Virus Res. 22, 44–52.

Keppel, F., Rychner, M., Georgopoulos, C., 2002. Bacteriophage-encoded cochaperonins can substitute for *Escherichia coli*'s essential GroES protein. EMBO Rep. 3, 893–898.

Kurochkina, L.P., Semenyuk, P.I., Orlov, V.N., Robben, J., Sykilinda, N.N., Mesyanzhinov, V.V., 2012. Expression and functional characterization of the first bacteriophage-encoded chaperonin. J. Virol. 86, 10103–10111.

Landry, S.J., Taher, A., Georgopoulos, C., van der Vies, S.M., 1996. Interplay of structure and disorder in cochaperonin mobile loops. Proc. Natl. Acad. Sci. USA 93, 11622–11627.

Landry, S.J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., Gierasch, L.M., 1993. Characterization of a functionally important mobile domain of GroES. Nature 364, 255–258.

Laurinavičius, S., Bamford, D.H., Somerharju, P., 2007. Transbilayer distribution of phospholipids in bacteriophage membranes. Biochim. Biophys. Acta 1768, 2568–2577.

Laurinavičius, S., Käkelä, R., Somerharju, P., Bamford, D.H., 2004. Phospholipid molecular species profiles of tectiviruses infecting Gram-negative and Gram-positive hosts. Virology 322, 328–336.

Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., Taddei, F., 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proc. Natl. Acad. Sci. USA 105, 3076–3081.

Lloyd-Price, J., Häkkinen, A., Kandhavelu, M., Marques, I.J., Chowdhury, S., Lihavainen, E., Yli-Harja, O., Ribeiro, A.S., 2012. Asymmetric disposal of individual protein aggregates in *Escherichia coli*, one aggregate at a time. J. Bacteriol. 194, 1747–1752.

Mindich, L., Bamford, D.H., Goldthwaite, C., Laverty, M., Mackenzie, G., 1982a. Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. J. Virol. 44, 1013–1020.

Mindich, L., Bamford, D., McGraw, T., Mackenzie, G., 1982b. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. J. Virol. 44, 1021–1030.

Mindich, L., Cohen, J., Weisburd, M., 1976. Isolation of nonsense suppressor mutants in Pseudomonas. J. Bacteriol. 126, 177–182.

Ogino, H., Wachi, M., Ishii, A., Iwai, N., Nishida, T., Yamada, S., Nagai, K., Sugai, M., 2004. FtsZ-dependent localization of GroEL protein at possible division sites. Genes Cells 9, 765–771.

Olkkonen, V.M., Bamford, D.H., 1989. Quantitation of the adsorption and penetration stages of bacteriophage φ6 infection. Virology 171, 229–238.

Olsen, R.H., Siak, J.S., Gray, R.H., 1974. Characteristics of PRD1, a plasmid-dependent broad host range DNA bacteriophage. J. Virol. 14, 689–699.

Peralta, B., Gil-Carton, D., Castaño-Diez, D., Bertin, A., Boulogne, C., Oksanen, H.M., Bamford, D.H., Abrescia, N.G., 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. PLoS Biol. 11, e1001667.

Phair, R.D., Misteli, T., 2000. High mobility of proteins in the mammalian cell nucleus. Nature 404, 604–609.

Poranen, M.M., Ravantti, J.J., Grahn, A.M., Gupta, R., Auvinen, P., Bamford, D.H., 2006. Global changes in cellular gene expression during bacteriophage PRD1 infection. J. Virol. 80, 8081–8088.

Richardson, A., Georgopoulos, C., 1999. Genetic analysis of the bacteriophage T4-encoded cochaperonin Gp31. Genetics 152, 1449–1457.

Rydman, P.S., Bamford, J.K.H., Bamford, D.H., 2001. A minor capsid protein P30 is essential for bacteriophage PRD1 capsid assembly. J. Mol. Biol. 313, 785–795.

Rydman, P.S., Caldentey, J., Butcher, S.J., Fuller, S.D., Rutten, T., Bamford, D.H., 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. J. Mol. Biol. 291, 575–587.

Saren, A.M., Ravantti, J.J., Benson, S.D., Burnett, R.M., Paulin, L., Bamford, D.H., Bamford, J.K.H., 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. J. Mol. Biol. 350, 427–440.

Schaff, J., Fink, C.C., Slepchenko, B., Carson, J.H., Loew, L.M., 1997. A general computational framework for modeling cellular structure and function. Biophys. J. 73, 1135–1146.

Sternberg, N., 1973. Properties of a mutant of *Escherichia coli* defective in bacteriophage lambda head formation (groE). II. The propagation of phage lambda. J. Mol. Biol. 76, 25–44.

Strömsten, N.J., Bamford, D.H., Bamford, J.K.H., 2003. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. J. Virol. 77, 6314–6321.

van der Vies, S.M., Gatenby, A.A., Georgopoulos, C., 1994. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. Nature 368, 654–656.

Vilen, H., Aalto, J.M., Kassinen, A., Paulin, L., Savilahti, H., 2003. A direct transposon insertion tool for modification and functional analysis of viral genomes. J. Virol. 77, 123–134.

Winkler, J., Seybert, A., Konig, L., Pruggnaller, S., Haselmann, U., Sourjik, V., Weiss, M., Frangakis, A.S., Mogk, A., Bukau, B., 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. EMBO J. 29, 910–923.

Winston, F., Botstein, D., Miller, J.H., 1979. Characterization of amber and ochre suppressors in *Salmonella typhimurium*. J. Bacteriol. 137, 433–439.

Xu, Z., Horwich, A.L., Sigler, P.B., 1997. The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. Nature 388, 741–750.

III

NEW ENVELOPED dsRNA PHAGE FROM FRESHWATER HABITAT

by

Sari Mäntynen, Elina Laanto, Annika Kohvakka, Minna M. Poranen, Jaana K.H. Bamford & Janne J. Ravantti 2015.

Journal of General Virology 96: 1180-1189.

IV

A NEW VIRUS TYPE FOUND FROM A BOREAL LAKE - A LINK BETWEEN ssDNA AND dsDNA INTERNAL MEMBRANE-CONTAINING PHAGES

by

Elina Laanto, Sari Mäntynen, Ashley Gillum, Jenni Marjakangas, Janne J. Ravantti, Lotta-Riina Sundberg, Juha T. Huiskonen & Jaana K.H. Bamford

Manuscript