Purification and crystallization of the two major coat proteins of bacteriophage P23-77 for X-ray crystallography

Ilona Rissanen
Master's Thesis
University of Jyväskylä
Faculty of Mathematics and Science
Department of Biological and Environmental Science
Molecular biology
December 2009

PREFACE

I wish to begin with many thanks to all the people who made this thesis possible. When I began my thesis, I had just settled down to Jyväskylä and it was a great luck to get instantly involved with an inspiring and functional lab group.

Special gratitude goes to Professor Jaana Bamford for letting me in on the research first as a trainee and then as an undergraduate. It was the most interesting time of my life as a student and definitely the most educating.

And no education comes without supervisors. I'm very thankful to Jaana Bamford and Alice Pawlowski for supervising this thesis and teaching me the skills I needed to accomplish the study. It was great to learn the keys to do independent research.

Among other skillful people deserving thanks are Petri Papponen, who helped me with a ton of questions on laboratory practice, Elina Laanto for her help with sequencing, and Matti Jalasvuori for inspiring evolutionary approach and image editing advice. Salla Ruskamo's crystallization counsel was golden. I never had a question someone didn't have an answer to, and the people of the B2-corridor have made the time spent at work cozy and fun. Many thanks to you all.

Thesis wasn't pure lab work though, and I hardly would have handled the writing process and all my courses without the relaxing presence of my loved ones. I'm grateful to have you.

Jyväskylän yliopisto
Matemaattis-luonnontieteellinen tiedekunta

Pro gradu –tutkielman tiivistelmä

Tekijä: Ilona Rissanen
Tutkielman nimi: Bakteriofagin P23-77 kahden kuoriproteiinin puhdistaminen ja kiteyttäminen röntgensädekristallografiaa varten
English title: Purification and crystallization of the two major coat proteins of bacteriophage P23-77 for X-ray crystallography
Päivämäärä: 23.11.2009 **Sivumäärä:** 51 + 5

Laitos: Bio- ja ympäristötieteiden laitos
Oppiaine: Molekyylibiologia
Tutkielman ohjaaja(t): Prof. Jaana Bamford ja M.Sc. Alice Pawlowski

---

Tiivistelmä:

Virusproteiinien rakenteiden tuntemus voi tuoda vastauksen moniin kysymyksiin virusten evoluutiosta ja toiminnoista. Vallitseva menetelmä biologisten makromolekyylien rakenteiden selvittämisessä on röntgensädekristallografia. Menetelmän etukäteisvaatimuksena on, että tutkittavaa proteiinia saadaan tuotettua korkea pitoisuus erittäin puhtaana ja että puhtaasta näytteestä saadaan tarkoin valituissa olosuhteissa tuotettua diffraktoivia kiteitä. Tässä tutkimuksessa tuotettiin ja puhdistettiin bakteriofagin P23-77 kuoriproteiineja VP16 (20 kDa) ja VP17 (32 kDa). Proteiini VP16 myös kiteytettiin.

P23-77 on hännätön bakteriofagi joka tartuttaa kuumien ääriolosuhteiden *Thermus thermophilus*-bakteereja. Viruksella on ikosaedrin muotoinen, pääasiassa proteiineista VP16 ja VP17 koostuva kapsidi, sekä lipidikalvon ympäröimä kaksijuosteinen DNA-genomi. P23-77 löydettiin Uuden Seelannin kuumista lähteistä vuonna 2005 ja sen piirteissä on sittemmin havaittu yllättävää samankaltaisuutta muihin viruksiin. Huomattavimpia samankaltaisuuksia ovat P23-77:n hyödyntämä ATPaasi, jossa on PRD1-adenovirusten sukulinjalle tyypillinen motiivi, ja kuoriproteiinien rakenne, joka muistuttaa suuresti arkkivirus SH1:n kapsomeereja.

Bakteriofagi P23-77:n proteiineja ei ole aiemmin puhdistettu eikä kiteytetty. Viruksen kuoren muodostavat proteiinit ovat erityisen mielenkiinnon kohde, sillä niiden rakenteet ovat todennäköisimmin konservoituneet ja siten evolutiivisesta näkökulmasta kiinnostavat. Samankaltaisuudet bakteriofagin P23-77 ja muiden virusten välillä ovat johtaneet hypoteesiin, jonka mukaan P23-77 edustaa aiemmin tuntematonta varhaisinta haaraumaa kaksois-β-tynnyrirakenteellisten virusten sukulinjassa.

Kuoriproteiineja VP16 ja VP17 tuotettiin rekombinanttiproteiineina *Escherichia coli*-bakteereissa. Proteiineja puhdistettiin ioninvaihto- ja kokoerottelukromatografioiden avulla. Puhdistettua muokkaamatonta VP16-proteiinia (7mg/ml) käytettiin kiteyttämiseen, jossa hyödynnettiin riippuvan pisaran kaasudiffuusiotekniikkaa. VP16 muodostaa kuutiomaisia kiteitä olosuhteissa, joissa pisara- ja kaivopuskurit sisältävät polyetyleeniglykolia matalissa konsentraatioissa.

Avainsanat: bakteriofagi P23-77, virusten evoluutio, kaksois-β-tynnyri, kuoriproteiinit VP16 ja VP1, kromatografia, kiteyttäminen

University of Jyväskylä
Faculty of Mathematics and Science

Abstract of Master´s Thesis

Author: Ilona Rissanen
Title of thesis: Purification and crystallization of the two major coat proteins of bacteriophage P23-77 for X-ray crystallography
Finnish title: Bakteriofagin P23-77 kahden kuoriproteiinin puhdistaminen ja kiteyttäminen röntgensädekristallografiaa varten
Date: 23.11.2009 **Pages:** 51 + 5

Department: Department of Biological and Environmental Science
Chair: Molecular Biology
Supervisor(s): Prof. Jaana Bamford, M. Sc. Alice Pawlowski

Abstract:

Knowledge of virus protein structures can shed light to various questions about the evolutionary origin and function of viruses. Prevailing method for structural analysis of biological macromolecules is X-ray crystallography. The method has demanding requirements, since the target protein has to be highly concentrated, pure and crystallized under the exact correct conditions to produce diffracting crystals. In this study we expressed and purified VP16 (20 kDa) and VP17 (32 kDa), the two major coat proteins of bacteriophage P23-77. Protein VP16 was also crystallized.

P23-77 is a tailless bacteriophage infecting high temperature extremophile bacterium *Thermus thermophilus*. The virus has an icosahedral capsid consisting mainly of VP16 and VP17, and a circular dsDNA genome enclosed by an internal lipid membrane. P23-77 was discovered in 2005 from the hot springs of New Zealand, and has since been found to exhibit surprising reminiscence to other viruses. Most notably, P23-77 utilizes an ATPase similar to the PRD1-adenoviral lineage and has capsomer structure much alike that of an archaeal virus SH1.

None of the proteins of P23-77 have been purified or crystallized before. The proteins composing the viral coat are of special interest because their structures are most likely to be conserved and thus of evolutionary value. Given the similarities between P23-77 and other viruses, it is suggested that P23-77 could represent an unknown ancestral branch to the diverse double β-barrel lineage of viruses.

VP16 and VP17 were expressed as recombinant proteins in *Escherichia coli.* The proteins were purified using ion exchange- and size exclusion chromatography. Concentrated extract (7 mg/ml) of purified native VP16 was deployed for crystallization with hanging drop vapor diffusion technique. VP16 builds cuboid crystals when the crystallizing conditions contain polyethylene glycol in low concentrations.

Keywords: bacteriophage P23-77, evolution of viruses, double β-barrel, major coat proteins VP16 and VP17, chromatography, crystallization

TABLE OF CONTENTS

ABBREVIATIONS

AGE	Agarose gel electrophoresis
AIX	Anion-exchange chromatography
ATPase	Adenosine triphosphatase
bp	Base pairs
CIX	Cation-exchange chromatography
Da	Dalton
DNA	Deoxyribonucleic acid
ds	Double strand
HCl	Hydrochloric acid
IEX	Ion-exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	1000 base pairs
LB	Luria Bertani
MCP	Major coat protein
MES	2-(*N*-morpholino)ethanesulfonic acid
NaCl	Sodium chloride, salt
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
RNA	Ribonucleic acid
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
STIV	*Sulfolobus* turreted icosahedral virus
T-number	Triangulation number
Tris	Tris(hydroxymethyl)-aminomethane
VP	Virus protein
Å	Ångström

1. INTRODUCTION

1.1 Evolution of viruses

Viruses are parasitic agents consisting of a ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) genome enclosed in a protein coat, possibly containing an additional internal or external lipid membrane. Viruses exhibit vast diversity in their form and function and are the most abundant entities in the world, bathing all cellular life (Bergh et al., 1989; Wommack and Colwell, 2000). The largest known group among viruses are bacteriophages, the viruses infecting bacteria, with circa 5600 identified members (Ackermann, 2003). Besides infecting and living in their hosts, viruses are commonly found integrated in genomes of cellular organisms, composing plenty of the eukaryotic non-coding genetic material (Pritham et al., 2007). Thus, the viral sea is one of the greatest selection pressures affecting all organisms.

The origin and evolution of viruses have long been a puzzling question. In early 20th century, viruses were proposed to be the ancestors of cellular life (d'Herelle et al., 1924). The theory was later replaced by others taking into account the total dependence viruses have on their host cells. Since it is impossible to gather tangible evidence of the primordial times, theories have been rather vague, suggesting that the first viruses derived from pre-existing cellular structures or metabolic processes that gained independence from the cell (for review see Hendrix et al., 2000). Accumulating knowledge of viral genome sequences and protein structures have increased the understanding of the evolutionary mechanisms affecting viruses. The proposed mechanisms of bacteriophage evolution include horizontal gene transfer and moron accretion (Lucchini et al., 1999; Juhala et al., 2000). Attempts to map the evolutionary relationships of viruses have been hindered by their highly differentiated genomes. A recent approach proposes that the solution lays in conserved protein structures termed the "viral self" (Bamford, 2003). These structures govern the basic functions of viruses separate from host interaction, and reveal similarity between viruses that may infect totally different hosts.

The history of viruses can be traced back to the times before the last universal common ancestor of cells (Benson et al., 1999; Koonin et al., 2006). Based on the characteristics of viruses, it has been proposed that they originated as vehicles of genetic information transfer in the primordial world (Jalasvuori and Bamford, 2008). Among the characteristics promoting the theory are the horizontal gene transfer and genome editing capability of viruses (Kidambi et al., 1994; Comeau and Krisch, 2005; Witzany, 2006). In addition to the information transfer, proto-viruses are suggested to have had considerable influence in the appearance of multiple cellular features such as the bacterial cell wall (Jalasvuori and Bamford, 2008). As the primordial cells became more independent and isolate from the environment, the usefulness of proto-viruses declined to the point where the viruses shed their beneficial role and assumed a parasitical one.

Since viruses are in such a tight connection with cellular life, information on their evolution could provide insights to the origins of life as well as yield valuable knowledge about the molecular mechanisms governing their current functions. Such information is difficult to attain due to the vast diversity and relatively rapidly changing genetic material of viruses. Accumulating knowledge on the atomic-level structures of viral components is vital for furthering our understanding of viral evolution (Benson et al., 1999; Bamford et al., 2002; Krupovic and Bamford, 2008). Especially viral coat proteins and proteins involved in genome packaging and ejection are of evolutionary interest. This is further discussed in chapters 1.1.1 and 1.1.2.

1.1.1 Viral Self

The expanding knowledge of genomic sequences provides a valuable approach for analyzing the relationships of multiple species, but is often misleading when applied to viruses (Bamford, 2003). Genetic comparisons are feasible for systems in which sequence similarities can be detected and followed, as the organisms are reasonably close in evolutionary time or exchange genetic material. During longer evolutionary time spans, structures and functions may be conserved but sequences differentiate to suit local conditions until similarity vanishes.

Viral structures can be categorized in two separate types. The elements responsible of host interaction require plasticity and change continuously, evolving and differentiating. These elements are a source of diversity among viruses as well as a source of difficulty for virus classification. The other type besides the host interaction structures are the "viral self" elements, key structures that have evolved early to high fitness and been conserved thereafter. Identification of the "viral self" structures relies heavily on accumulation of protein structure information in atomic level. "Viral self" structures provide a way to classify viruses and determine their evolutionary relationships despite the confusing genomic differentiation. Typical "viral self" structures are the major coat proteins (MCP) and factors associated with virion assembly, such as the DNA packaging adenosine triphosphatase (ATPase). Bamford (2003) presents bacteriophages PRD1 and phi6 as model organisms for structural lineages. The characteristics linking them to supposedly unrelated viruses are summarized below.

PRD1 and adenovirus share many structural characteristics, most notably the triangulation value (T-number) of their icosahedral capsids (pseudo T=25), MCP structures and their capsid organization (Benson et al., 1999; Bamford et al., 2002). Another look into the resemblance revealed that the PRD1 virion contains other adenovirus-like features, for example a vertex spike structure resembling that of an adenovirus (Sokolova et al., 2001). The main difference between the two viruses is the internal membrane of PRD1; however, this is a structure required to deliver the genome in the bacterial host cell and thus falls in the category of host interaction elements (Grahn et al., 2002; Bamford, 2003).

Viruses of the family *Reoviridae* are double stranded (ds) RNA viruses infecting higher eukaryotes. They deploy intricate RNA-release and replication structures (Diprose et al., 2001). Phi6, a *Cystoviridae* dsRNA virus, shares a very similar replication cycle and core structure with the family *Reoviridae*. External virus layers are utilized in host interaction and thus not highly conserved, while internal core particles and their assembly are "viral self" features (Bamford, 2003). Electron microscopy-based three-dimensional structures of phi6 and reovirus cores revealed similar spherical structures related to RNA ejection, namely the five fold vertexes of the core particles (Grimes et al., 1998; Reinisch et al., 2000).

It is proposed that such strongly conserved structures appearing in viruses infecting different hosts mark lineages that have a common origin and can be traced through the evolutionary time and across all domains of life (Benson et al., 1999; Bamford et al., 2003; Krusovic and Bamford, 2008).

1.1.2 Double β-barrel lineage

The double β-barrel lineage of viruses is a suggested structural lineage based on the MCP structures of certain viruses. It presents a structure centric way to categorize viruses to evolutionary groups. The idea was conceived after finding the surprising similarity between the jelly roll motifs of PRD1 coat protein P3 and the adenovirus hexon (Benson et al., 1999). PRD1 is a tailless bacteriophage while adenovirus is a dsDNA virus infecting higher eukaryotes. Despite their very different hosts the viruses share the same capsid topology (Olsen et al., 1947; Stewart et al., 1991; Butcher et al., 1995). The conserved MCP structures typical to the PRD1-adenoviral lineage exhibit two jelly rolls that are essentially identical in both PRD1 and adenovirus (Benson et al., 1999, figure 1).

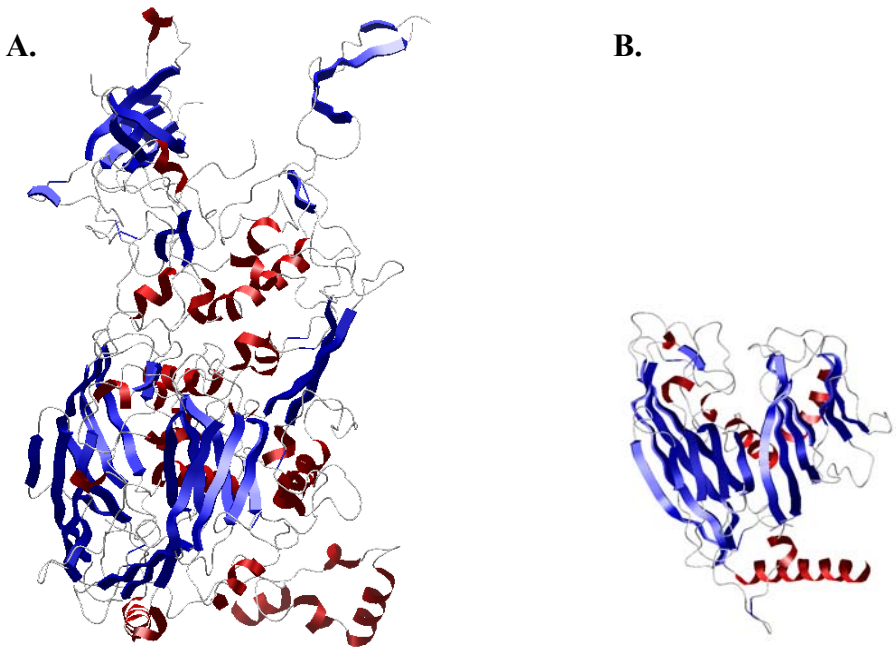

Figure 1. Double β-barrel structures exhibited by adenovirus hexon and PRD MCP P3. Part A illustrates the crystal structure of chimpanzee adenovirus major coat protein hexon (based on Protein Data Bank release by Pichla-Gollon et al., 2007) and part B illustrates the PRD1 MCP P3 (Benson et al., 2002). Structural modelling with Bodil (Lehtonen et al., 2004).

In addition to bacteriophages and viruses of eukaryotes, the double β-barrel lineage includes archaeal viruses such as the *Sulfolobus* turreted icosahedral virus (STIV) (Benson et al., 2004; Khayat et al., 2005). Among other members of the lineage are bacteriophage PM2, a group of Bam35-like phages of *Bacillus* bacteria, and viruses of eukaryotes such phycodnaviruses (Benson et al., 1999, 2004; Nandhagopal et al., 2002; Abrescia et al., 2005; Laurinmäki et al., 2005). Though the lineage is diverse, some general features can be identified beyond certain protein folds. The viruses of the double β-barrel lineage often have an icosahedral capsid with an internal lipid membrane.

1.2 P23-77

P23-77, a *Thermus* bacteriophage, was first isolated from the hot springs of New Zealand during a massive thermophilic phage characterization conducted by the Promega Corporation (Yu et al., 2005). During the research, 115 bacteriophage strains were isolated from alkaline hot springs around the world and characterized by various criteria, including capsid morphology, host range and DNA restriction endonuclease digestion patterns. P23-77-type of viruses were found to exhibit dsDNA genome packed in tailless isometric capsids with an internal lipid membrane. The primary host for P23-77 was identified as *Thermus thermophilus* (ATCC 33923). *Thermus* bacteria have optimal growth temperature ranging from +70 °C to +75 °C (Ramalay and Hixson, 1970; Dworkin et al., 2006). They contain an outer membrane and a peptidoglycan layer in their cell envelope, rendering the bacteria Gram-negative.

P23-77 was further characterized by Jaatinen et al (2008). Electron cryo-microscopy and three-dimensional image reconstruction revealed P23-77 particles to consist of an icosahedral capsid with 15 nm long stick-like spikes extending from the vertices of the virions. The dimensions of the P23-77 capsid are 78 nm measured from facet-to-facet and 87 nm vertex-to-vertex. The 6 nm thick protein capsid is well ordered with hexagonal and pentameric capsomers and has a T-number of pseudo-28. The pentameric capsomers form a base for the spike proteins. Hexagonal capsomers feature a peculiar structure consisting of a base with tower-like extensions, the position of which depends on the capsomer organization in the shell lattice (figure 2). P23-77 genome has been found to contain approxi-

mately ten genes encoding structural proteins (Jalasvuori et al., 2009). Five of these have been shown to associate with the inner membrane and three with the capsid, including MCPs VP16 (20 kDa) and VP17 (32 kDa). These two proteins are encoded by open reading frames (ORFs) 16 and 17 respectively.

The capsomer structure of P23-77 is highly similar to that of an archaeal virus SH1, as SH1 capsomeres also exhibit the "base and tower" structures (Jäälinoja et al., 2008). SH1 capsomers have been suggested to contain a base structure of six singe β-barrels (Bamford et al., 2005). Based on the structural similarity of the capsomers of P23-77 and SH1, it is assumed that P23-77 capsomers have the same organization: a base of hexameric single β-barrels of VP16 decorated by the tower proteins VP17.

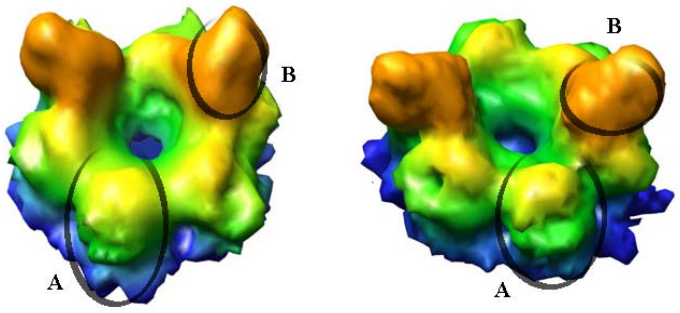

Figure 2. P23-77 capsomers illustrated in two different positions. Capsomers show the suggested hexameric single β-barrel base topped by tower proteins. "A" illustrates one of the structural units of the base (VP16), while "B" illustrates one of the overlaying structural units (VP17). The figure is based on the electron cryo-microscopy images of P23-77 capsomers by Jaatinen et al (2008).

Underneath the protein capsid lays a 3 nm thick internal lipid membrane (Jaatinen et al., 2008). Proteins connect the outer lipid bilayer leaflet with the capsid at the five-fold vertices. The membrane encloses highly ordered dsDNA genome packed with the average spacing of 31 ångström (Å). The genome is 17036 nucleotides long, circular and contains 37 putative genes (Jalasvuori et al., 2009). These ORFs encode multiple proteins besides the structural elements, including enzymes for bacterial cell wall degradation and an ATPase that has been suggested to function in DNA packaging. No viral polymerases have been found, indicating that the virus utilizes host polymerases. See table 1 for more detailed approach on the identified protein species of P23-77.

Table 1. Identified contents of the genome of P23-77. Based on the results of Jalasvuori et al. 2009.

P23-77 ORF or *gene*	**Predicted function of the product**	**Association with the virion element**
1	DNA replication initiation protein	
2		
3	Phosphoadenosine phosphosulfate reductase	
4		
5		
Gene 6	VP6	
7		
8		
9	Endolysin	
10	Amidase endolysin	
Gene 11	VP11	Coat
12		
13	ATPase	
14		
Gene 15	VP15	Membrane
Gene 16	20 kDa MCP	Coat
Gene 17	32 kDa MCP	Coat
18		
Gene 19	VP19	Membrane
Gene 20	VP20	Membrane
21		
Gene 22	VP22	Membrane
Gene 23	VP23	Membrane
24		
25		
26		
27		
28		
Gene 29	Lysozyme	
30		
31	Transglycosylase	
32		
33		
34		
35		
36		
37		

1.3 Evolutionary background of P23-77

Known features of P23-77 have been compared to a range of viruses (Jalasvuori et al., 2009). Relations revealed an interesting evolutionary background possibly predating the divergence of the three domains of life.

Structural similarity has been observed between P23-77 and viruses SH1, IN93 and STIV. Haloarchaeal virus SH1 has an icosahedral capsid with T-value of 28 (Jäälinoja et al., 2008). Electron cryo-microscopy has provided images of SH1 capsomers, revealing a structure that consists of a multimeric base decorated with two to three copies of another coat protein. P23-77 is the only known other virus exhibiting T-value of 28, and the similarity in capsid morphology between archaeal virus SH1 and bacteriophage P23-77 is considerable (Jaatinen et al., 2008; Jalasvuori et al., 2009). P23-77 and SHI1 capsomers have identical "base and tower" organization (figure 2). In addition to SH1, morphological comparison links P23-77 to the *Thermus aquaticus* virus IN93 (Jalasvuori et al., 2009). Another possible relation has been found with coat protein structure prediction: P23-77 major capsid protein VP17 structure is predicted to have 45% similarity to the major capsid protein of STIV (analyzed with PHYRE).

Relationships between P23-77 and other viruses have been further clarified by genetic comparisons of the P23-77 dsDNA genome and other known virus genomes with PSI-BLAST. P23-77 is similar to the *T. aquaticus* phage IN93, revealing overall DNA level identity of 47.1%. Most of the structural proteins of the two phages have about 75% similarity at the amino acid level. In addition to IN93, P23-77 has similarity with the plasmids and viruses of halophilic archaea. Gene products alike the two major capsid proteins and the ATPase of P23-77 were discovered in *Haloarcula hispanica* virus SH1, integrated in the genome of *Haloarcula marismortui*, and in *Halobacterium salinarium* plasmid pHH205.

Furthermore, Jalasvuori et al. (2009) discovered another important relation: bacteriophage P23-77 and bacteriophage PRD1 share a very similar ATPase motif. Since the PRD1 ATPase functions in genome packaging (Strömsten et al., 2005; Ziedaite et al., 2009) the same

can be speculated for P23-77 ATPase (Jalasvuori et al., 2009). Similar virion lipid composition provides another link between P23-77 and PRD1. The relation to PRD1, a member of the double β-barrel viral superlineage, connects P23-77 to a diverse group of viruses infecting hosts from archaea to eukarya.

Even though the genomes of viruses often show little sequence similarity, their key structures may be conserved since ancient times (Bamford et al. 2005, Krupovic and Bamford 2008). The relationships between P23-77, IN93, SH1 and pHH205 were revealed by the alignment of primary coat protein sequences, indicating that these viral protein elements were conserved before the divergence of the ancestors of P23-77 and SH1 (Jalasvuori et al., 2009). After the divergence P23-77 and SH1 have adapted to their host organisms and evolved into their current differing genomes. In case where true virion-encoding parasites do not survive, the viruses may evolve into stable plasmids (pHH205) or integrate permanently in the host chromosome, as in the genome of *Haloarcula marismortui*.

In addition to the archaeal elements described earlier, P23-77 is likely to be related to PRD1- and STIV-like viruses as indicated by their possibly similar MCP folding and similar ATPases. Jalasvuori et al. (2009) envision an ancestor that had a coating with single β-barrel proteins as the basic units, and that deployed an ATPase to deliver the genome within the coating. This ancestor later diverged into the lineage with single β-barrel coat and two MCPs (P23-77 and SH1) and the double β-barrel lineage (PRD1). According to this hypothesis, P23-77 and SH1 form the earliest separating branch of the DNA viruses of the diverse β-barrel viral lineage (Krupovic and Bamford, 2008). Acquiring information on the representatives of this ancestral branch may broaden our knowledge on the origin of viruses during the predomain era of life (Woese 2002, Jalasvuori and Bamford 2008).

1.4 Principles of crystallization

X-ray crystallography is the prevailing method for determining the three dimensional arrangement of atoms in protein molecules. The method utilizes electro-magnetic radiation with a wavelength of around 0.1 nm or 1 Å to see atomic structures. Crystals consisting of symmetrically arranged molecules are radiated with focused X-rays to produce a diffraction pattern. This is recorded in an X-ray detector and used for computational processes and modelling to extract information about the three dimensional structure. This study focuses on protein crystal acquisition, the necessary first step of X-ray crystallography.

Proteins are complex molecules and successful X-ray crystallography requires a crystal that comprises of homogenous protein in a well-ordered lattice (Bergfors 2009). Typical crystallization process is a time consuming task and consists of screening hundreds of potential crystallizing conditions and optimizing the conditions with subtle changes. Two of the most commonly used methods for protein crystallization are hanging- and sitting-drop vapour diffusion. Both methods feature a droplet of purified protein, buffer and precipitant equilibrating with a larger reservoir containing similar buffers and precipitants in higher concentrations. As the equilibration reaction progresses, water vaporizes from the drop until the precipitant concentration reaches a level suitable for crystallization. Protein crystallization begins with screening a wide catalogue of different buffer and precipitant conditions. Initial crystals are often tiny, needle- or platelike, and unsuitable for X-ray diffraction experiments. After some crystallization is observed, the size and quality of the crystals can be enhanced with a variety of methods. Typically, the protein concentration, buffer and precipitant concentrations are varied in a controlled fashion. In addition to environmental adjustments, new hanging- or sitting drops can be produced with crystal seeding. During a seeding step, nucleation points derived from previously produced crystals are introduced to new crystallization drops during their preparation. When sufficiently large crystals have been obtained, they are subjected to diffraction experiments and further optimized to provide desired resolution in X-ray diffraction crystallography.

2. AIM OF THE STUDY

The aim of this study was to purify and crystallize the two major coat proteins of the novel bacteriophage P23-77. The proteins VP16 (20 kDa) and VP17 (32 kDa) form the capsid of P23-77 and are most likely to contain "viral self" features linking the virus to an evolutionary group. It is suggested that the protein VP16 might exhibit single β-barrel structure, an unknown motif as an ancestor to the double β-barrel structure commonly found in viruses. To determine the structures of the MCPs with X-ray crystallography, well-ordered crystals of highly pure proteins are required.

Broken down, the aims were:

- Establishing purification protocols for VP16 and VP17
- Crystallizing protein samples produced with aforementioned purification protocols
- Experimenting on other significant characteristics of recombinant VP16 and VP17, such as their heat resistance and multimeric state

3. MATERIALS AND METHODS

3.1 DNA methods

Expression plasmids and protocols for over-expressing VP16 and VP17 were previously established (Rissanen et al., unpublished). The plasmids pIR1 and pIR2 carry the genes *ORF16* (500 base pairs, bp) and *ORF17* (900 bp) ligated to a pET22b(+)-vector (Novagen). They encode proteins VP16 or VP17, respectively (see Appendix I for plasmid maps).

3.1.1 Co-expression constructs for the capsid proteins VP16 and VP17

In addition to previously produced expression plasmids pIR1 and pIR2, a co-expression plasmid that contains both VP16 and VP17 was constructed. A segment of P23-77 genome from the beginning of the *ORF16* to the end of *ORF17* was multiplied with PCR (Finnzymes Piko Thermal Cycler). PCR yielded a fragment consisting of *ORF16-ORF17* in between of two restriction sites, *Nde*I and *Bam*HI. The 1410 bp long fragment was separated from other DNA material with 1% agarose gel electrophoresis (AGE, Appendix II) and eluted from the gel with QIAquick Gel Extraction Kit (Qiagen). The fragment was ligated in a *Sma*I-digested pUC19-vector which was transferred to *Escherichia coli* JM109 cells using standard protocols (Sambrook and Russell, 2001). After an overnight Luria Bertani-broth (LB) culture of the cells, pUC19- plasmids containing the fragment were extracted with QIAprep Spin Miniprep Kit (Qiagen).

ORF16-ORF17 fragment was separated from extracted pUC19-constructs by *Nde*I-*Bam*HI restriction digestion which was conducted overnight at +37°C. Afterwards, the fragment was separated in 1% AGE, eluted from the gel with QIAquick Gel Extraction Kit (Qiagen), and ligated in a *Nde*I-*Bam*HI digested pET22b(+)-vector. Resulting new expression plasmid, named pIR3, was transformed in *E. coli* HMS174(DE3) cells using standard protocols (Sambrook and Russell, 2001). Multiplied pIR3 plasmids were extracted from HMS174(DE3) cells with QIAprep Spin Miniprep Kit (Qiagen). Presence of the desired

frangment in pIR3 was confirmed with a *Pst*I restriction digestion. Enzymes used in the process were purchased by Fermentas.

3.2 Protein methods for VP16 and VP17

3.2.1 Expression and initial extraction

The two coat proteins of bacteriophage P23-77 were produced in T7 polymerase expressing host *E. coli* HMS174(DE3) (Studier et al., 1991). Host cells were transformed with expression plasmid pIR1 or pIR2 using standard methods (Sambrook and Russell 2001). Transformed cells were cultivated overnight in LB-agarose plates and the emergent colonies were deployed to inoculate LB-broth cultures. The cultures were grown overnight in an incubator shaker with 230 revolutions per minute (rpm) at +37°C and then utilized to inoculate 6 x 400 ml LB large scale cultures. 1mM isopropyl β-D-1-thiogalac-topyranoside (IPTG) was added during inoculation to induce protein expression. All media used in the culturing of bacteria contained ampicillin (150 µg/ml) for selection. Large scale cultures were grown overnight in an incubator shaker with 230 rpm at +28°C. Afterwards, the cells were collected with centrifugation in a Sorvall RC-5C Centrifuge using a SLA-3000 rotor, 5000 rpm for 15 minutes at +4°C. Cells were washed with 20 mM Tris-HCl – 50 mM NaCl buffer (pH 7.4) and the resulting pellets were dissolved in 1:100 of the original volume in 20 mM Tris-HCl – 50 mM NaCl buffer at pH 7.4.

Concentrated cell suspensions containing recombinant VP16 or VP17 were disrupted with French Pressure Cell treatment (Thermo electron corporation) using pressure of 2000 pounds per square inch. French Press step was conducted twice for each sample at room temperature.

Cell debris was removed from the crude lysate with centrifugation in a Sorvall RC-5C Centrifuge using a SS-34 rotor, 8000 rpm for 10 minutes at +4°C. The resulting supernatant was collected and enrolled in a high speed centrifugation in a Beckman Coulter Optima L-90K ultracentrifuge and a 70 Ti rotor, at the speed of 33000 rpm for 2 hours at

+5°C, after which the supernatant was further treated with high temperature incubation for 5 minutes at +90°C. Degraded material was removed by centrifugation in a Megafuge 1.0R (Heraeus Instruments) using a BS4402/A rotor at the speed of 5000 rpm for 15 minutes at +4°C. Different stages of the protein extraction process were analyzed in 15% Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGEs were conducted following a modified version of the procedure described by Olkkonen and Bamford (1989). For SDS-PAGE method, see Appendix III. In this study, the process of extracting soluble protein from host cells and heat incubating them as described above is referred to as initial protein extraction. This step is designed to produce protein material for other methods like chromatographic purification.

3.2.1.1 Co-expression of the capsid proteins

E. coli HMS174(DE3) cells were transformed with pIR3 following standard protocols (Sambrook and Russell, 2001). The expression and initial protein extraction were conducted using the protocol described above (Materials and methods 3.2.1), with the exception of omitting the heat incubation step. Soluble proteins present in the supernatant obtained by ultracentrifugation were analyzed in 15% SDS-PAGE.

3.2.2 Heat purification of the capsid proteins

Heat resistance of VP16 and VP17 was explored with multiple heating trials. The experiments were conducted with crude lysate and with purified proteins. Protein samples in 1ml volume were heated in a Grant BT3 heating block at temperatures ranging from +60°C to +100°C for 1-10 minutes. After heating, samples were cooled down on ice and centrifuged in a Megafuge 1.0R (Heraeus Instruments) using a BS4402/A rotor at the speed of 5000 rpm for 15 minutes to separate degraded proteins from surviving ones. High temperature treatment was integrated to the final purification protocols as an additional method for removing contaminating proteins. The heat purification steps are described as part of the purification processes (Materials and methods 3.2.1 and 3.2.3).

3.2.3 Chromatography protocols

This study focused on establishing protocols that would allow the purification of VP16 and VP17 to a level suitable for crystallization. Prevailing purification methods were ion-exchange chromatography (IEX) and size exclusion chromatography (SEC). IEX is a method for separating molecules with differing net charges and is further divided into an-ion- and cation-exchange chromatography (AIX and CIX). SEC separates the molecules according to their size. Optimal conditions for purifying VP16 and VP17 were screened for both AIX and CIX with multiple different buffer conditions (see Results, table 2). SEC was conducted after IEX. The tedious optimization process yielded the purification protocols described below.

3.2.3.1 Final purification protocol for VP16

VP16 was expressed and initially extracted as described in the section 3.2.1. After ultracentrifugation and heat purification, the protein solution in 20 mM Tris-HCl – 50 mM NaCl (pH 7.4) buffer was concentrated with Amicon ultafiltration system (Millipore) to the final volume of 5 ml. During the concentration, buffer conditions were changed to 20 mM ethanolamine – 50 mM NaCl (pH 9). Final sample of the protein suspension in pH 9 buffer was then filtrated through a Sartorius 0.45 μm filter. ÄKTAprime plus (GE Healthcare) chromatography system was equipped with a 5 ml Q HP HiTrap column (GE Healthcare) and equilibrated with 20 mM Ethanolamine – 50 mM NaCl buffer at pH 9. The filtered protein sample was injected to the ÄKTA system, and AIX was performed with flow rate of 1 ml/min. For the first half an hour the sample was washed through the column with the low-salt buffer (20 mM Ethanolamine – 50 mM NaCl buffer at pH 9) and the flow-through was collected in 1 ml fractions. After the flow-through had eluted, the buffer flowing through the column was changed to 20 mM ethanolamine – 1 M NaCl buffer at pH 9. The column was washed with this high salt buffer and a sample of the eluting proteins was collected. VP16 was present in the flow-through.

The fractions from AIX were analyzed in 15% SDS-PAGE. Flow-through fractions containing the purest VP16 in 20 mM ethanolamine – 50 mM NaCl buffer (pH 9) were pooled

and concentrated with the Amicon ultrafiltration system. Upon reaching the final volume of 5ml, the salt concentration of the protein suspension was adjusted to 150 mM NaCl. Final volume of the protein suspension was then filtrated through a Sartorius 0.45 μm filter. ÄKTAprime plus (GE Healthcare) chromatography system with a HiLoad 26/60 Superdex 200 prep grade column (GE Healtcare) was prepared for SEC by equilibrating the column with 20 mM ethanolamine – 150 mM NaCl buffer at pH 9. The filtered protein sample was injected to the ÄKTA system, and SEC was performed with a flow rate of 1 ml/min. VP16 eluted from the column between 200 and 250 minutes of running time. The protein peak was collected in 2 ml fractions, which were analyzed in 15% SDS-PAGE. Fractions containing the purest VP16 were pooled and concentrated with Amicon ultrafiltration system to the final volume of 1 ml. During concentration, the buffer conditions of the sample were changed from 20 mM ethanolamine – 150 mM NaCl (pH 9) to 20 mM Tris-HCl at pH 7.4. Obtained final sample was subjected to high temperature treatment at +95°C for 2-3 minutes. Degraded material was pelleted and the supernatant was collected and analyzed in 15% SDS-PAGE and with the Bradford protein assay (Bradford, 1976). If the purity and concentration levels were sufficient and the concentration of protein in the sample was 5-10 mg/ml, the sample was suitable for crystallization experiments.

3.2.3.2 Final purification protocol for VP17

VP17 was expressed and initially extracted as described in the section 3.2.1. After ultracentrifugation and heat purification, the protein solution in 20 mM Tris-HCl – 50 mM NaCl (pH 7.4) buffer was concentrated with Amicon ultafiltration system (Millipore) to the final volume of 5 ml. During the concentration, buffer conditions were changed to 20 mM ethanolamine (pH 9.5). Final protein suspension in pH 9.5 buffer was then filtrated through a Sartorius 0.45 μm filter. ÄKTAprime plus (GE Healthcare) chromatography system was equipped with a 5 ml Q HP HiTrap column (GE Healthcare) and equilibrated with 20 mM ethanolamine buffer at pH 9.5. The filtered protein sample was injected to the ÄKTA system, and AIX was performed with flow rate of 1 ml/min. Flow-through was washed trough the column and after it had been eluted, the salt concentration was increased manually. The salt concentration was first set to 3% and then to 4% of high salt buffer (20 mM ethanolamine – 1 M NaCl buffer at pH 9.5), washing the eluting proteins from the

column. The following 1% increase in salt gradient was a critical step since VP17 has been observed to elute in this specific salt percentage. After setting the high salt buffer concentration to 5%, eluting protein was collected in 1 ml fractions. When all of the protein had been eluted, salt percentage was set to 100% and the rest of the bound proteins were washed off the column. Fractions collected during the 5% salt peak were analyzed in 15% SDS-PAGE. The fractions contained the original 20 mM ethanolamine (pH 9.5) and 5% of 1M NaCl buffer, creating final buffer conditions of 20 mM ethanolamine – 50 mM NaCl at pH 9.5.

Fractions containing the purest VP17 were pooled and concentrated to a final volume of 5 ml with Amicon ultrafiltration system. During the concentration, buffer conditions of the protein suspension were changed to 20 mM Tris-HCl – 150 mM NaCl (pH 7.4). The suspension was then filtrated through a Sartorius 0.45 µm filter. ÄKTAprime plus (GE Healthcare) chromatography system with a HiLoad 26/60 Superdex 200 prep grade column (GE Healtcare) was prepared for SEC by equilibrating the column with 20 mM Tris-HCl – 150 mM NaCl at pH 7.4. The filtered protein sample was loaded and injected to the ÄKTA system, and SEC was performed with flow rate of 1 ml/min. VP17 eluted from the column between 200 and 250 minutes of running time. The protein peak was collected in 2 ml fractions, which were analyzed in 15% SDS-PAGE. With this protocol, VP17 was purified to a high level, but reaching crystallization purity level requires further optimization of the process.

3.2.4 Determination of the multimericity of the capsid proteins

Two methods were deployed to determine the multimericity of recombinant VP16 and VP17. Both methods, gradient ultracentrifugation and analytic SEC, rely on the molecular mass and volume difference between monomeric and multimeric forms of the proteins.

Gradient ultracentrifugation was conducted with SW41 centrifugation tubes (Beckman Coulter Instruments) containing a linear 10-14% sucrose gradient. The gradient was prepared wth Gradient Master (Biocomp) gradient former. Each tube was loaded with a single sample of protein, two tubes with 100 µl samples of purified MCPs (1 µg/µl) and three

tubes with commercial protein standards aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (44 kDa) from Gel Filtration Calibration Kit HMW 28-4038-42 (GE Healthcare). Gradient tubes were centrifuged in a Beckman Coulter Optima L-90K ultracentrifuge using a SW 41 Ti rotor for 45 hours in 32000 rpm at temperature of +10°C. After centrifugation, the gradient in each tube was fractionated as 1ml fractions and analyzed with Bradford protein assay and 15% SDS-PAGE (Bradford, 1976).

Analytical SEC was conducted with an ÄKTAprime plus chromatography system (GE Heathcare) using a special SEC column (Superdex 200 10/300 GL, GE Healthcare). A size standard mixture was prepared combining 80 µl of each standard protein, aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (44 kDa), at the concentration of 2 mg/ml (Gel Filtration Calibration Kit HMW 28-4038-42, GE Healthcare). Final volume of 160 µl was administered to the chromatography system and washed through the column with 20 mM Tris-HCl – 50 mM NaCl buffer (pH 7.4), at a flow rate of 0.5 ml/min and a pressure limit of 0.75 mPa. The absorption level of the eluting solution was monitored at the wavelength of 280 nm and the measured data were combined into a chromatogram with the PrimeView program (GE Healthcare). The standard chromatogram was saved for later reference and the process was repeated with the same parameters and a 200 µl sample of 1 mg/ml purified VP16 or VP17. Afterwards, the chromatograms were compared to determine size differences based on the elution times of the proteins. To ensure the presence of the desired protein, collected fractions were analyzed in 15% SDS-PAGE.

3.3 Nucleic and amino acid sequencing

Expression vectors pIR1 (*ORF17*/pET22b(+)), pIR2 (*ORF16*/pET22b(+)) and pIR3 (*ORF16-ORF17*/pET22b(+)) were sequenced with ABI prism 3130XL Genetic Analyzer to confirm the DNA sequence of the recombinant protein coding regions. Sequencing and analyzing of the results were conducted by Elina Laanto and Alice Pawlowski, respectively.

To ascertain that the purification processes were focused on the correct proteins, amino acid sequencing was conducted for samples of VP16 and VP17. Two samples containing

presumed target proteins were purified with ultracentrifugation, heat purification and AIX, and sent to the Protein Chemistry Research Group and Core Facility at the Institute of Biotechnology at Helsinki University, where the samples were run in 12% SDS-PAGE and electroblotted into polyvinylidene fluoride membrane by Nisse Kalkkinen. The membrane was treated with the Coomassie Brilliant Blue reagent, after which a section of the target protein band was cut off and sequenced with Procise 494HT Sequencer (Applied Biosystems).

3.4 Protein crystallization

3.4.1 Initial screening

Crystallization experiments were initiated with commercial crystal screens and utilizing the hanging-drop vapour diffusion technique. Purified VP16 (7 mg/ml) in 20 mM Tris-HCl buffer at pH 7.4 was used as a starting material and administered to the screening well plates with a Mosquito Nanodrop Crystallization Robot (TTP Labtech). The commercial screens included Crystal Screen HT (HR2-130, Hampton Research) and Salt Screen (Hampton Research), both containing 96 different buffer conditions. The Mosquito robot was utilized to create the hanging drops containing the protein solution and crystallizing buffer in 1:1 ratio. After protein and buffer administration, the plates were sealed and stored at room temperature. Crystal formation was first observed daily and later monthly with a light microscope (GWB Olympus BH-2).

3.4.2 Optimization of crystallization

Conditions where initial crystallization was observed were manually set up and adjusted. Adjustments of the crystallization conditions were conducted using 1% changes of the buffer component concentrations, in hanging drops of higher volume than the ones produced with mosquito (Appendices IV). In addition, a seeding procedure was performed (see Introduction 1.4).

After administering 1 ml of customized buffer to each well of the optimization plate, siliconized cover slides (Hampton Research) were equipped with the droplet of protein and buffer, placed on top of the wells facing the buffer, and sealed with oil (Silicone DC 200 fluid, Serva). 1 µl of 7 mg/ml protein sample were mixed together with 1 µl of custom-made buffer to produce each hanging drop. Crystal formation was observed with a light microscope (Olympus SZ61 with an 110AL2X-2 WD38 objective) and progress was recorded with an integrated camera system (Altra 20 Soft Imaging System). To reduce the amount of small crystals in a sample, a seeding step was conducted. Final crystallization protocol is described below.

3.4.3 Final crystallization protocol for VP16

VP16 crystallization was conducted with highly pure protein sample in 20 mM Tris-HCl buffer (pH 7.4) and utilizing the hanging drop vapour diffusion technique. Well plates were prepared by pipetting 1 ml of buffer into each well. One well was prepared with 10 % PEG 1500 (Sigma) – 10% PEG 8000 (Sigma) water based buffer. Respective crystallization drop was prepared by mixing 1 µl of protein sample (~ 7 mg/ml) and 1 µl of well buffer and sealing the lid containing the drop on top of the well. After 24 hours, the drop contained a shower of tiny crystals. The well was then opened and the drop was diluted in 50 µl of the well buffer. Resulting solution contains crystal shards, and is further diluted 1:1000 in a buffer containing 6 % PEG 1000 and 5% PEG 8000. This dilution was then utilized to produce a new crystallization drop by mixing 1 µl of the dilution and 1 µl of protein sample (~ 7 mg/ml). The drop was sealed facing a well with water-based buffer containing 6 % PEG 1000 and 5% PEG 8000. Crystal formation was observed daily and expected to take place in two weeks. Produced crystals were single cuboids approximately 0.1 – 0.2 mm long and 0.05 – 0.1 mm in other dimensions.

3.4.4 Diffraction experiments

Optimization of the crystallization yielded single, cuboid crystals large enough to pick them from the hanging drop and subject to analyzing. The protein content of the crystals was ensured in 15% SDS-PAGE. After this, the diffraction potential of the crystals was in-

vestigated. The test of the diffraction power of the protein crystals were done using Bruker Kappa Apex II diffractometer with graphite-monochromatized molybdenum-K_α (λ = 0.71073 Å) and copper-K_α (λ = 1.5418 Å) radiation; sealed tubes with 2.0 kW and 1.5 kW power, respectively. The crystals, sized circa 0.1 x 0.15 x 0.20 mm, were mounted on a nylon loop using a cryo protectant (5% PEG1500, 5% PEG8000, 20% glycerol) and placed on the goniometer equipped with Oxford cryostream liquid N_2 cooling device, 123 K for Mo- and 173 K for Cu-radiation. The mounted and centered crystals were then exposed to X-rays with various exposure times ranging from 1 minute to 4 minutes.

4. RESULTS

Three-dimensional structures of virus proteins contain valuable information about the evolutionary origin and function of viruses. An often used method of structural analysis is the X-ray crystallography, which requires extensive preparations ranging from protein expression to purification and finally crystallization. Protein purification process has to be optimized to yield a sample of highly homogenous and concentrated native protein for crystallization. The two major coat proteins VP16 (20 kDa) and VP17 (32 kDa) of *Thermus thermophilus* phage P23-77 were expressed as recombinant proteins in *E.coli*, initially extracted and purified with ultracentrifugation, heat incubation, IEX and SEC (figure 3).

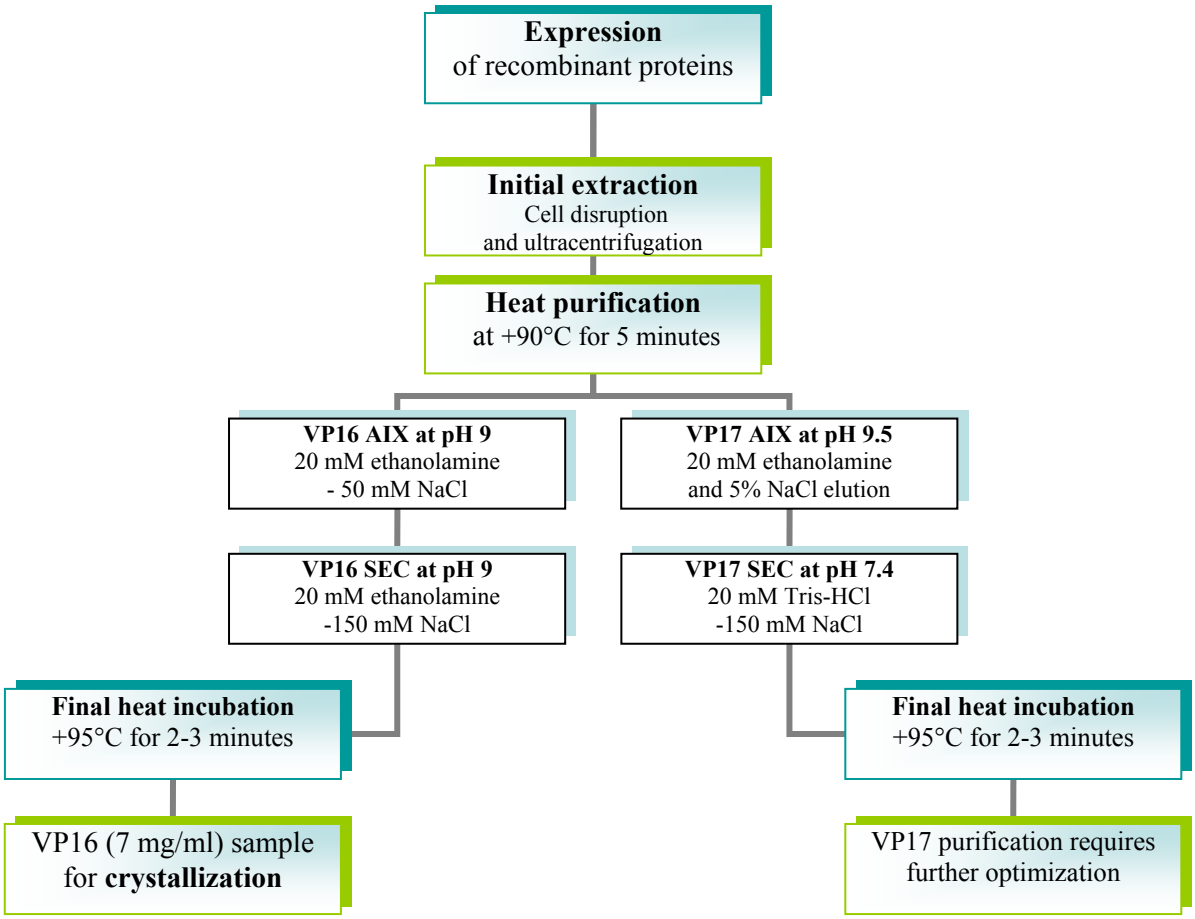

Figure 3. Purification of P23-77 coat proteins. Flowchart presents the processes utilized for purifying P23-77 coat proteins VP16 and VP17. Most favorable conditions found during the research are presented. AIX stands for anion-exchange chromatography and SEC for size exclusion chromatography.

Conditions and methods for successfully separating recombinant coat proteins were refined to suit the distinct chemical and physical properties of each protein. In addition to establishing purification protocols for VP16 and VP17, the multimeric state of the proteins was studied. Also, a plasmid was constructed that allows co-expression of both proteins in the same cell. Purified VP16 was introduced into crystallization experiments ranging from initial screening to manual optimization and diffraction tests.

4.1 Protein purification

pET22b(+)-based plasmids pIR1 and pIR2 were used to produce protein material for purification experiments. Proteins were expressed in separate cultures and the cells were disrupted utilizing French pressure cell, after which water soluble proteins were collected centrifugations as described in Materials and Methods 3.2.1. Throughout the experiments, samples of multiple purification steps were analyzed to follow the purity of the protein samples. Prevailing method for analysis was 15% SDS-PAGE. Resulting gels were photographed with Bio-Rad imaging system.

4.1.1 Heat purification

Despite not being a complex purification process, high temperature treatment can be efficient if the target protein has an unusual heat resistance. P23-77 is a phage for thermophilic bacteria, which might indicate high thermostability, especially for the MCPs. A series of high temperature treatments was conducted to observe the heat resistance of VP16 and VP17. The protein samples were incubated on a heating block at temperatures ranging from +70°C to +100°C for several minutes, cooled on ice and centrifuged to remove degraded material. Aliquots of the original sample and the supernatant obtained after heat purification were analyzed in 15% SDS-PAGE.

VP16 and VP17 were found to withstand several minutes in temperatures up to +90°C without significant degradation while various host proteins degraded. The limit for the heat resistance of the two coat proteins was +95°C, which they withstood for 2-3 minutes be-

fore severe degradation begun. Heat purification was found to be an efficient additional purification method and utilized in the final purification protocols twice: once before IEX, in 5 ml volume at +90°C for 5 minutes, and after all purification steps, including SEC, in 1 ml volume at +95°C for 2-3 minutes. The first heat purification considerably decreased the occurrence of non-target proteins in the sample containing all soluble proteins from the expression strain. In the final stage of the purification process, additional lingering impurities could be removed with temperature treatment near the degradation limit of the coat proteins (figure 9).

4.1.2 Ion-exchange chromatography (IEX)

VP16 and VP17 were obtained from the host cells as highly concentrated soluble protein and partially purified with heat incubation. IEX was chosen as a method for further purification. To determine suitable conditions for IEX, theoretical isoelectric point (pI) was calculated for both proteins, yielding 8.14 for VP16 and 5.76 for VP17. pI presents the pH value at which a molecule has no net electric charge. When the environment of the molecule has pH conditions higher than pI, the molecule will assume a negative net charge; in lower pH the charge will be positive. AIX and CIX retain negatively and positively charged molecules, respectively. Various buffers with different pH values were utilized in IEX to analyze the binding characteristics of VP16 and VP17 (table 2). IEX methods were performed with ÄKTAprime plus chromatography system as described in chapter 3.2.3.

Table 2. Binding characteristics of VP16 and VP17 in IEX
Binding of VP16 and VP17 to Anion-exchange (AIX) and Cation-exchange (CIX) columns in different buffers and pH values. Conditions chosen for final purification protocols are emphasized in bold. pI stands for the theoretical isoelectric point, MES is short for 2-(*N*-morpholino)ethanesulfonic acid and Tris for tris(hydroxymethyl)-aminomethane and n.d stands for undetermined condition.

Protein	pI	Buffer		Binding to AIX	Binding to CIX
VP16	8,14	50 mM MES – 50 mM NaCl	pH 6,5	-	n.d
		20 mM Tris-HCl – 50 mM NaCl	pH 7,4	-	-
		20 mM ethanolamine – 50 mM NaCl	**pH 9.0**	**-**	-
VP17	5,76	50 mM MES – 50 mM NaCl	pH 6,5	-	n.d.
		20 mM Tris-HCl – 50 mM NaCl	pH 7,4	-	-
		20 mM ethanolamine – 50 mM NaCl	pH 9	-	-
		20 mM ethanolamine	**pH 9,5**	+	n.d.

4.1.2.1 IEX of VP16

Smaller coat protein VP16 (20 kDa), with a pI value of 8.14, should bind to AIX column when the pH of the environment exceeds 8.14, and bind to CIX column in a lower pH. Experiments with buffers ranging from pH 6.5 to 9 revealed no binding to either AIX or CIX columns (table 2).

As illustrated in figure 4, AIX with a buffer at pH 9.0 (20 mM ethanolamine – 50 mM NaCl) proved to be a functional purification step compared to other experimented conditions, since most protein impurities bind to the column. Typically, IEX is utilized to bind and specifically elute the target protein; in this case, most of the other proteins bound, leaving the purified VP16 to be collected from the flow-through. After multiple IEX experiments, the optimal purification procedure for VP16 was defined as AIX at with buffer 20 mM ethanolamine – 50 mM NaCl in pH 9.0, in which the target protein was collected from the flow-through.

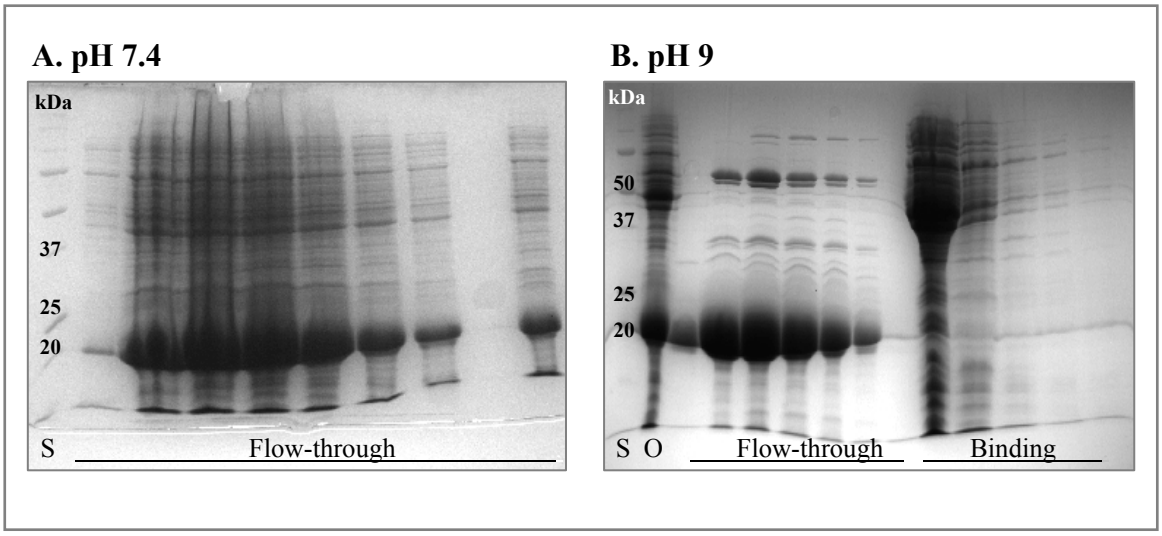

Figure 4. AIX experiments of VP16 at pH 7.4 and pH 9. Part A presents the AIX of a VP16 extract in 20 mM Tris-HCl – 50 mM NaCl buffer at pH 7.4. No binding to the column was observed and all the proteins in the sample eluted instantly ("Flow-through" in the figure A). Part B presents the AIX of a VP16 extract 20 mM ethanolamine – 50 mM NaCl at pH 9. During this AIX, notable binding to the column is observed: the VP16 still eluted instantly ("Flow-through" in part B) but most non-target proteins bind to the column and can be eluted by increasing the salt gradient, here at 100% ("Binding" in part B). Original sample enrolled in AIX is labeled O. Samples were analyzed in 15% SDS-PAGE (Appendix II) with Bio-Rad Precision Plus Protein Standard as a molecular mass reference (presented with S in figure).

4.1.2.2 IEX of VP17

Larger coat protein VP17 (35 kDa), with a pI of 5.76, should bind to an AIX column when the pH of the environment exceeds 5.76, and bind to a CIX column in a lower pH. Multiple different conditions with buffers ranging from pH 6.5 to 9.5 were tested for AIX and CIX (table 2). VP17 was found in the flow-through in most conditions, but observed to bind in AIX with saltless 20 mM ethanolamine buffer at pH 9.5 (figure 5).

Eluting protein was observed as peaks in the chromatogram produced with PrimeView program (figure 6). The program produces chromatograms based on the absorption level of the solution emanating from the column. Proteins increase the absorption and thus are presented as peaks. Peaks eluting after the beginning of the salt gradient are presented in the figures as "binding peaks". The gradient is created with a low salt buffer (20 mM ethanolamine buffer at pH 9.5) and increasing concentration of s high salt buffer (20 mM ethanolamine buffer – 1 M NaCl at pH 9.5). High salt buffer is indicated as the percentage it composes of the total buffer flowing through the column.

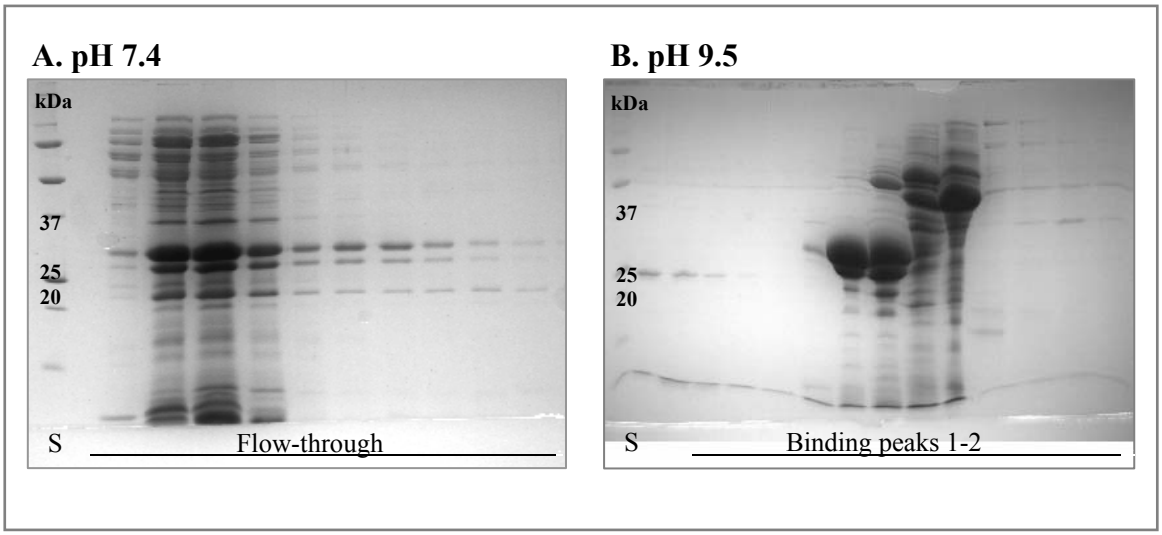

Figure 5. AIX experiments of VP17 at pH 7.4 and pH 9.5. Part A presents the AIX of a VP17 extract in 20 mM Tris-HCl – 50 mM NaCl buffer at pH 7.4. No binding to the column was observed and all the proteins in the sample eluted instantly ("Flow-through" in the figure A). Part B presents the AIX of a VP17 extract 20 mM ethanolamine pH 9.5. In these conditions, VP17 binds to the column and is eluted with low salt percentage ("Bindind peaks 1-2" in part B). Most impurities eluted after VP17, at higher salt concentration. Analyzed as in figure 4.

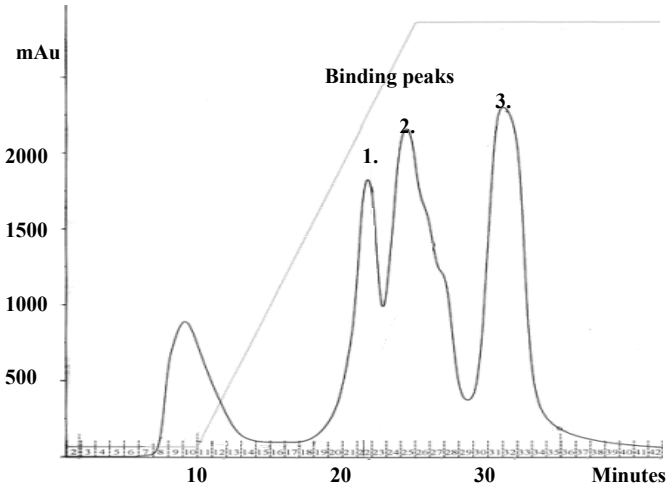

Figure 6. VP17 AIX chromatogram in pH 9.5 and a steadily increasing salt gradient. VP17 AIX chromatogram illustrates steadily increasing salt gradient based on 20 mM ethanolamine sample buffer in pH 9,5. The peaks in the chromatogram are caused by eluting proteins, and "binding peaks" illustrate proteins that initially bound to the column and eluted only in an increased salt concentration. Binding peaks are numbered 1-3, and the peaks 1 and 2 (presented in the figure 5 B) reveal the elution of VP17. mAu stands for milliabsorbance units. The chromatogram was produced with PrimeView program by GE Healthcare.

The initial NaCl gradient deployed for analyzing the binding of VP17 in 20 mM ethanolamine buffer at pH 9.5 was rapidly and continuously increasing, and thus unsuitable for determining the specific eluting conditions for VP17. The gradient revealed that VP17 binds to the column and elutes at a low salt percentage, while more detailed experiments on the eluting conditions were conducted afterwards with manually adjusted stepwise salt concentration trials. After experiments with 10%, 5%, 2% and 1% salt concentration increments, the elution of VP17 from the AIX column was pinpointed at 5% of high salt buffer, which is the same as 50 mM NaCl concentration in the column (figure 7).

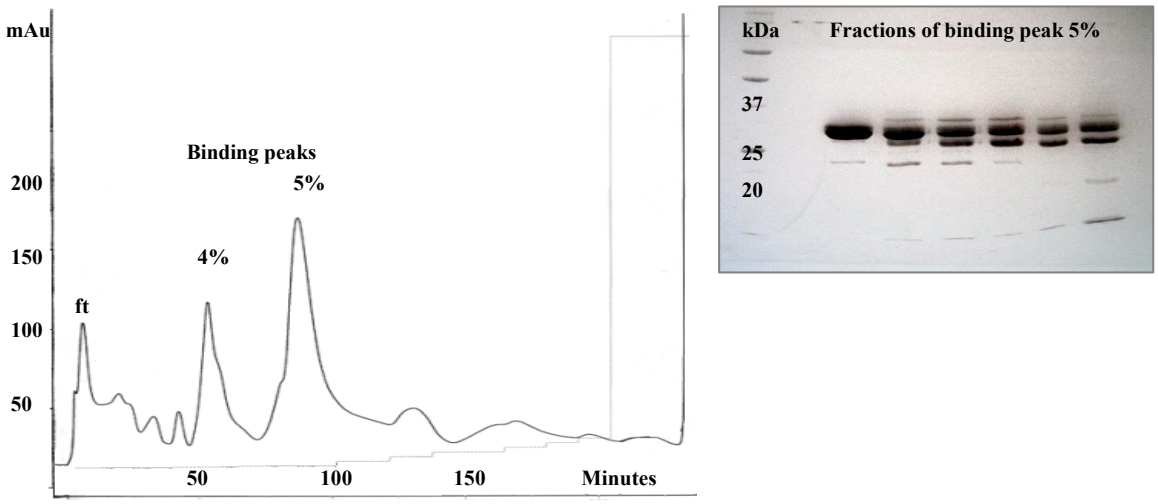

Figure 7. VP17 AIX with manually adjusted salt gradient ranging with 1% increments in the range of 4%-11%. The gradient was based on 20 mM ethanolamine buffer with pH 9.5. The chromatogram, produced with PrimeView (GE Healtcare) presents the flow-through (ft) and two notable low-salt binding peaks (4% and 5%). An SDS-PAGE analysis (conducted as in figure 4) revealed that VP17 elutes at 5% of 1 M salt.

For VP17, the purification effect was highest when AIX was conducted with 20 mM Ethanolamine buffer at pH 9.5 and with a stepwise low NaCl gradient, collecting the protein eluting at 5% of 1 M salt.

4.1.3 Size exclusion chromatography (SEC)

IEX step was not sufficient to remove all impurities, which led to the usage of size exclusion chromatography. SEC was utilized to remove protein impurities with similar chemical nature but a different size than VP16 (20 kDa) or VP17 (32 kDa). The protein material used in SEC consisted of fractions collected during IEX. To prepare the protein samples for SEC, the IEX fractions containing high amounts of VP16 or VP17 were pooled and concentrated utilizing Amicon ultrafiltration system (Millipore). The salt concentration of the final samples was set to 150 mM to suit the buffers utilized in SEC.

SEC was conducted in 20 mM Ethanolamine – 150 mM NaCl buffer (pH 9) for VP16 and in 20 mM Tris-HCl – 150 mM NaCl buffer (pH 7.4) for VP17. Flow rate during SEC was 1 ml/min. After running the protein samples in SEC, chromatograms were used to determine the elution time for the coat proteins. For both VP16 and VP17, a notable peak in the

absorption curve was observed between 200 to 250 minutes. Fractions collected during this time were analyzed in 15% SDS-PAGE. The fractions contained target proteins with considerably decreased levels of impurities (figure 8). SEC provides effective purification for IEX treated samples of VP16 and VP17. The method caused some target protein loss but yielded highly purified protein.

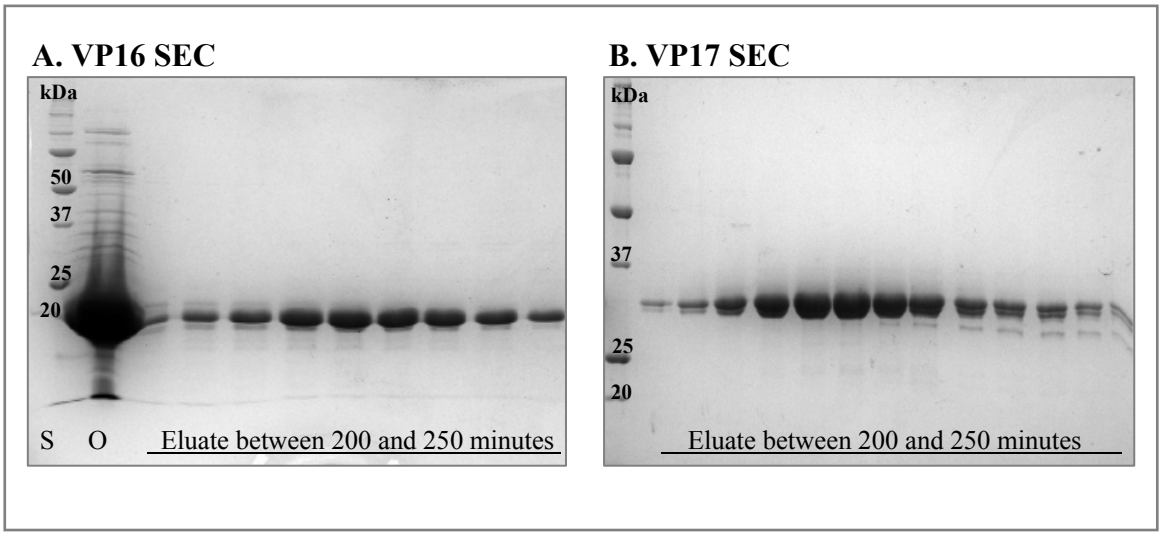

Figure 8. SEC purification of IEX purified VP16 and VP17. Part A presents VP16 SEC eluate and a sample of the original material deployed for SEC ("O"). Part B presents VP17 SEC eluate. SDS-PAGE analysis revealed high purification effect. Analyzed as in figure 4.

4.1.4 Final polishing

Final protocols for purifying VP16 and VP17 included initial extraction with French pressure cell, ultracentrifugation, initial heat purification (5 minutes in +90°C for 5 ml samples), refined AIX with distinguished pH values and ion concentration, SEC, and final heat purification (2-3 minutes in +95°C for 1 ml samples). For final protocols, see Materials and methods 3.2.3 and 3.4.3. In between the purification steps, samples were concentrated with Amicon Ultrafiltration system. After all purification procedures, final samples of VP16 and VP17 in 20 mM Tris-HCl buffer (pH 7.4) were analyzed in 15% SDS-PAGE (figure 9). The protein concentration of the samples was determined with Bradford protein assay, yielding 7 mg/ml for VP16 and 3 mg/ml for VP17.

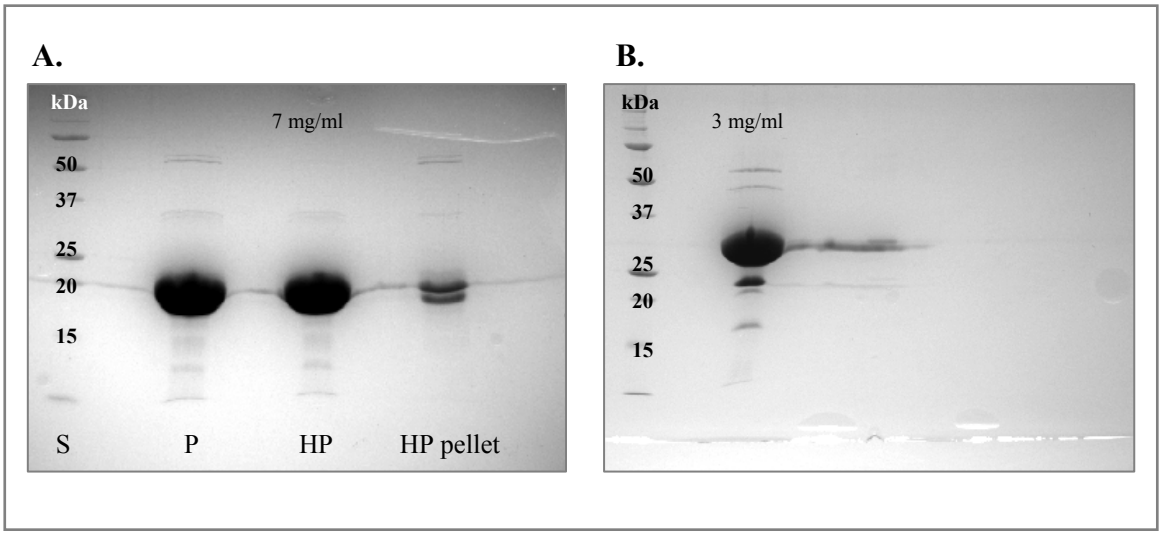

Figure 9. Final polishing step of the highly concentrated and purified VP16 and VP17. Proteins have been purified with IEX and SEC, and here the final polishing is given to VP16 with heat purification. Part A presents the purified VP16 protein ("P") that was subjected to final heat purification for 2-3 minutes at +95°C. The heat purified supernatant, containing native protein and the heat degraded pellet are presented as "HP" and "HP pellet" respectively. Part B presents VP17 in purest state that was acquired during the study. Both proteins are in 20 mM Tris-HCl buffer at pH 7.4. Analyzed as in figure 4.

The purified final sample of VP17 still contained impurities rendering the sample insufficient for crystallization, and VP17 purification protocol has to be further optimized. Purified recombinant VP16 was concentrated and deemed pure enough to introduce the sample in crystallization trials.

4.2 Determination of multimericity

Separately expressed VP16 and VP17 were found to be very stable, heat resistant, soluble after ultracentrifugation and susceptible to purification with IEX and SEC. All these characteristics are beneficial when aiming for protein crystallization, but as the crystallization requires a protein sample as homogenous as possible, other factors should also be taken into consideration. Knowledge about the multimeric state of the target proteins can help to pinpoint possible complications in the crystallization process. The multimericity of recombinant VP16 and VP17 was studied with two methods: gradient centrifugation and analytical SEC. Monomers of VP16 and VP17 have molecular masses of 20 and 32 kDa respectively.

Rate zonal gradient centrifugation separates the molecules in a sample according to their mass. VP16 and VP17 were centrifuged in a 10-40% linear sucrose gradient along with three commercial molecular mass standards, aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (44 kDa). Each sample occupied its own gradient tube. After centrifugation, the gradients in each tube were fractionated (1 ml fractions) and the protein content of the fractions was studied with Bradford protein assay and 15% SDS-PAGE. Analysis revealed that the proteins had stayed in the uppermost parts of the gradient (table 3).

Table 3. Gradient centrifugation of VP16, VP17 and mass standards. The table shows the gradient fraction in which the proteins were found. Large X presents majority of the protein, small x a lesser amount.

Fraction of the gradient from top to bottom	**Protein (kDa) found**				
	VP16	**VP17**	**aldolase (158)**	**conalbumin (75)**	**ovalbumin (44)**
1 (top 1 ml)	X	X		x	X
2			x	X	
3			X		
4					
5					
- 12 (bottom 1 ml)					

The standards were found in the following fractions: 1 (ovalbumin), 1-2 (conalbumin) and 2-3 (aldolase). The fractions are named 1-12 from topmost of the gradient in SW41 tubes to the bottommost. Both VP16 and VP17 were present in the fraction 1. Results indicate that the recombinant proteins are of the same or lighter mass than ovalbumin (44 kDa).

In addition to gradient centrifugation, an analytical SEC was conducted. The principle of analytical SEC is the same as in conventional SEC: the separation of molecules according to their size. The main difference to SEC utilized in the purification process is the usage of a smaller, more refined column (Superdex 200 10/300 GL, GE Healthcare) and a slower eluting buffer flow rate (0.5 ml/min) for the analytic SEC. A mixture of the protein mass standards aldolase (158 kDa), conalbumin (75 kDa) and ovalbumin (44 kDa) was injected into the analytic SEC column. After the SEC, produced chromatogram was stored and subsequent analytical SECs with the same parametres were conducted for VP16 and VP17 samples. Their eluting times were compared to the chromatogram to determine their mass (figure 10). Buffer deployed in the SEC and for all the protein samples was 20 mM Tris-HCl at pH 7.4.

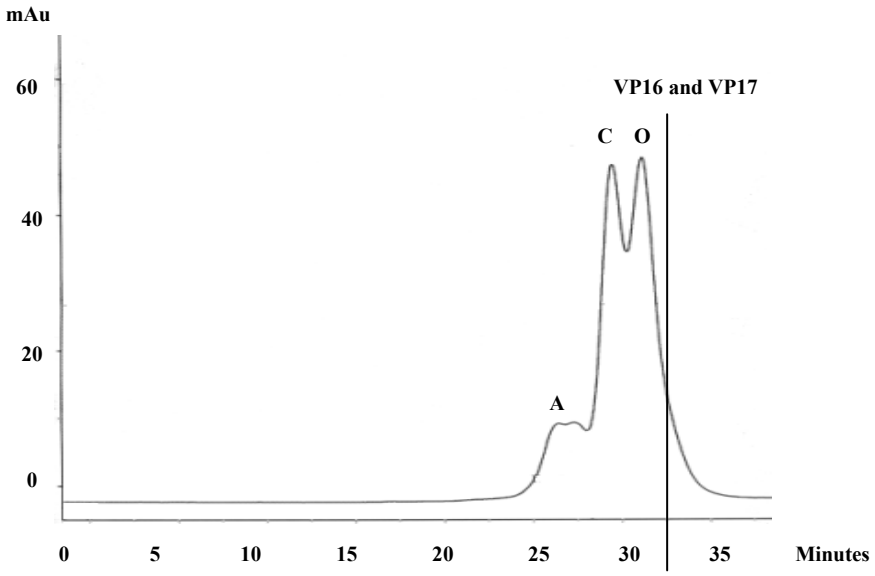

Figure 10. Analytic SEC chromatogram of the protein mass standards with an illustration of the eluting time of proteins VP16 and VP17. Protein mass standards are marked with letters: A is for aldolase (158 kDa), C for conalbumin (75 kDa) and O for ovalbumin (44 kDa). VP16 and VP17 elute around 32,5-33 minutes very close to each other. The chromatogram show both MCPs to have smaller mass than 44 kDa. The chromatogram was produced by PrimeView for milli-absorbtion units (mAu) per minutes.

Results from analytical SEC revealed the mass standards eluting after 20 minutes (aldolase), 29 minutes (conalbumin) and 32 minutes (ovalbumin). VP16 elutes at 32.5 minutes and VP17 at circa 33 minutes, and MCPs were thus revealed to be smaller in mass than the smallest mass standard ovalbumin (44 kDa). The results from analytical SEC were consistent with those from gradient centrifugation.

4.3 Protein crystallization

Crystallization of a target protein is required for structural analysis with X-ray crystallography (Introduction 1.4). In this study, hanging-drop vapor diffusion technique was utilized as a method for driving the equilibrium state of target-protein containing solution drops to a state allowing crystal formation. Purified VP16 sample (7 mg/ml, storage at +7C) in 20 mM Tris-HCl buffer (pH 7.4) was introduced in crystallization trials, first in initial screening conducted with commercial crystallization screens and later in manual optimization with varying buffer component concentrations

4.3.1 Initial screening

During initial screening, commercial crystallization screens providing multiple different buffer conditions were utilized as described in Materials and methods section 3.4.1. Initial screening revealed crystal formation in 24 hours. The conditions that induced crystal formation, found with Crystal Screen HT (HR2-130, Hampton Research), had buffers with the composition of 0.05 M Potassium phosphate monobasic - 20% w/v Polyethylene glycol 8000 and 10% w/v Polyethylene glycol 1000 - 10% w/v Polyethylene glycol 8000. When utilized for hanging-drop vapour diffusion crystallization with VP16 (7 mg/ml, in 20 mM Tris-HCl at pH 7.4), these buffer conditions produced needle- and plate-like crystals (figure 11). To improve crystal quality, found crystallization conditions were further optimized manually.

4.3.2 Optimization of crystallization

Optimization of crystallization aims to enlarge the crystal size and enhance the occurrence of single crystals. As the optimization of VP16 crystallization progressed, enlargement and singularization were accomplished with specific buffer conditions. Larger crystals took longer to form, and the emergence of crystals varied from 24 hours for small crystals to the maximum of three weeks for the larger ones.

From the original two crystallization conditions found with initial screening, the conditions with potassium phosphate failed to produce larger crystals. Efforts were focused to crystallization with PEG mixtures in water. Decreasing the PEG concentrations, utilizing PEG 1000 instead of the initially deployed PEG 1500, and applying a seeding step lead to steady increase in the crystal quality. Protein concentration was kept constant while PEG composition trials with concentrations ranging between 4%-10% for each PEG were conducted (Appendixes IV). The optimization progressed from hanging drops filled with showers of tiny crystals to drops containing 10-20 single, symmetrical crystals sized circa 0.1 x 0.15 x 0.20 mm (figure 11). For the final VP16 crystallization protocol, see Materials and methods 3.4.3.

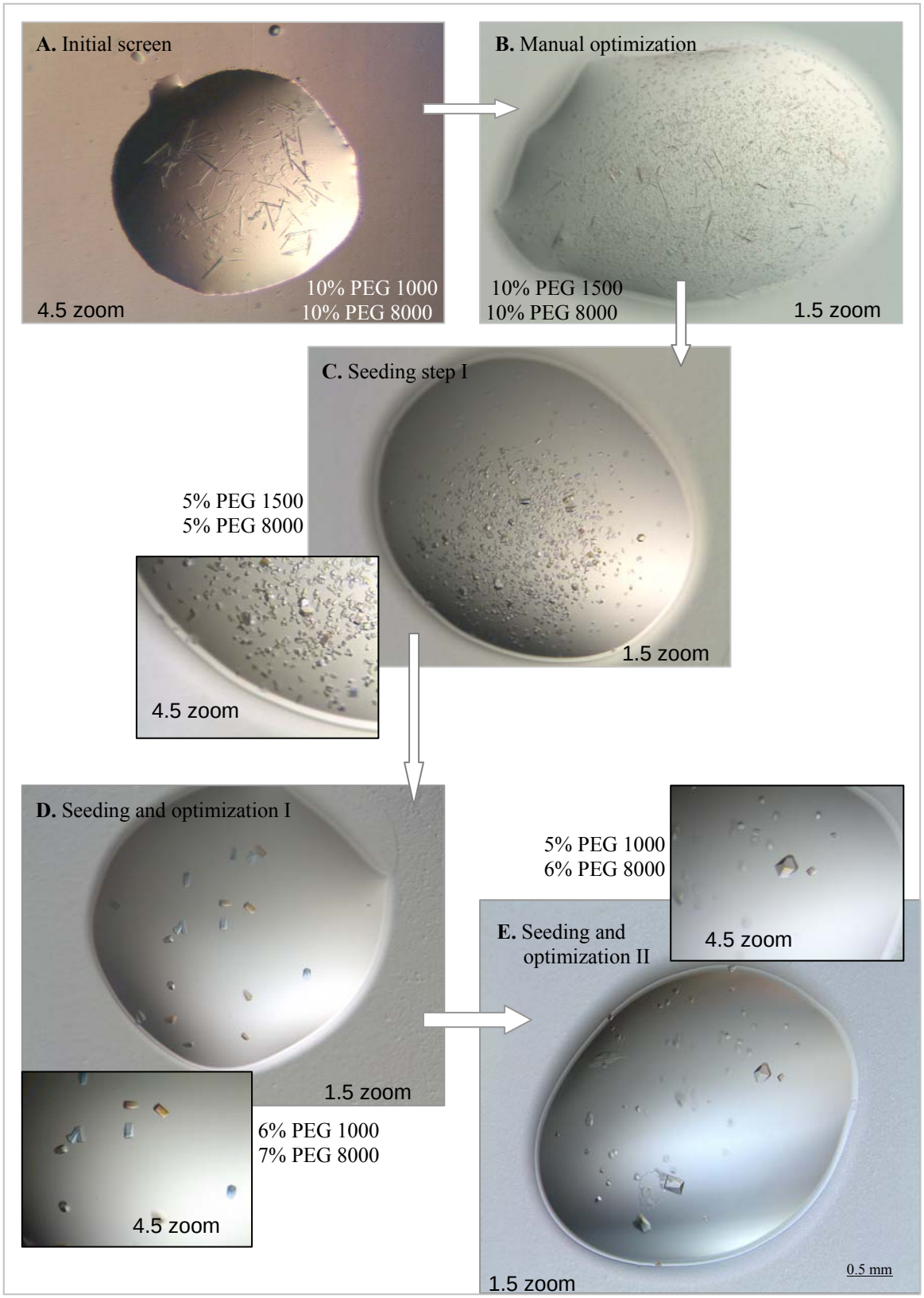

Figure 11. The optimization process of VP16 crystallization. Optimization progressed from the initial screening with Crystal Screen (Hampton research) (A) to the manual optimization (B) which was advanced with the addition of a seeding step (C).Conducting a seeding step with PEG 1000 instead of PEG 1500 enhanced crystal formation (D). Finally, with subtle buffer concentration changes and weeks of crystallization time larger crystals were obtained (E). Pictures were taken with Altra 20 Soft Imaging System.

4.3.3 Diffraction experiments

VP16 crystals presented in figure 11 part D were enrolled in diffraction experiments described in Materials and methods section 3.4.4. Prior to the diffraction experiments, the protein content of crystals produced with identical parameters was confirmed with SDS-PAGE. Though the crystals did contain protein of the expected molecular mass (20 kDa), no diffraction pattern was obtained with in house facilities.

4.4 Co-expression of VP16 and VP17

Recombinant VP16 and VP17, expressed separately, were utilized for protein purification. In addition, an expression plasmid containing the genes for both MCPs was constructed. The aim was to produce a functional expression plasmid that provides a way to express VP16 and VP17 in the same host. Such a culture could be used to observe whether recombinant VP16 and VP17 form complexes when expressed in the same cell.

4.4.1 Plasmid construction

A 1410 bp DNA fragment containing *ORF16* and *ORF17* was amplified by PCR using genomic DNA of phage P23-77 as a template. Gene fragment contained restriction sites for *Nde*I at the 5' terminus and for *Bam*HI at the 3' terminus, allowing cloning into the final expression vector pET22b(+) (Novagen). PCR product was eluted from 1% agarose gel. The fragment was ligated into *Sma*I digested pUC19-vector and transformed in *E. coli* JM109 cells with blue-white selection for screening. DNA was isolated from white colonies and inserts were verified with *Nde*I and *Bam*HI restrictions. A map of the resulting plasmid is shown in figure 12 part A.

The *ORF16-ORF17* containing fragment was cut from the pUC19-construct with *Nde*I- and *Bam*HI-restriction enzymes, ligated to similarly digested pET22b(+) (Novagen)-vector and transformed in *E. coli* HMS174 cells. The expression plasmid, named pIR3 (figure 12 part B), was verified with restriction analysis utilizing *Pst*I restriction enzyme. In addition,

the *ORF16-ORF17* insert in pIR 3 was sequenced and the sequence was confirmed to be correct (Elina Laanto and Alice Pawlowski).

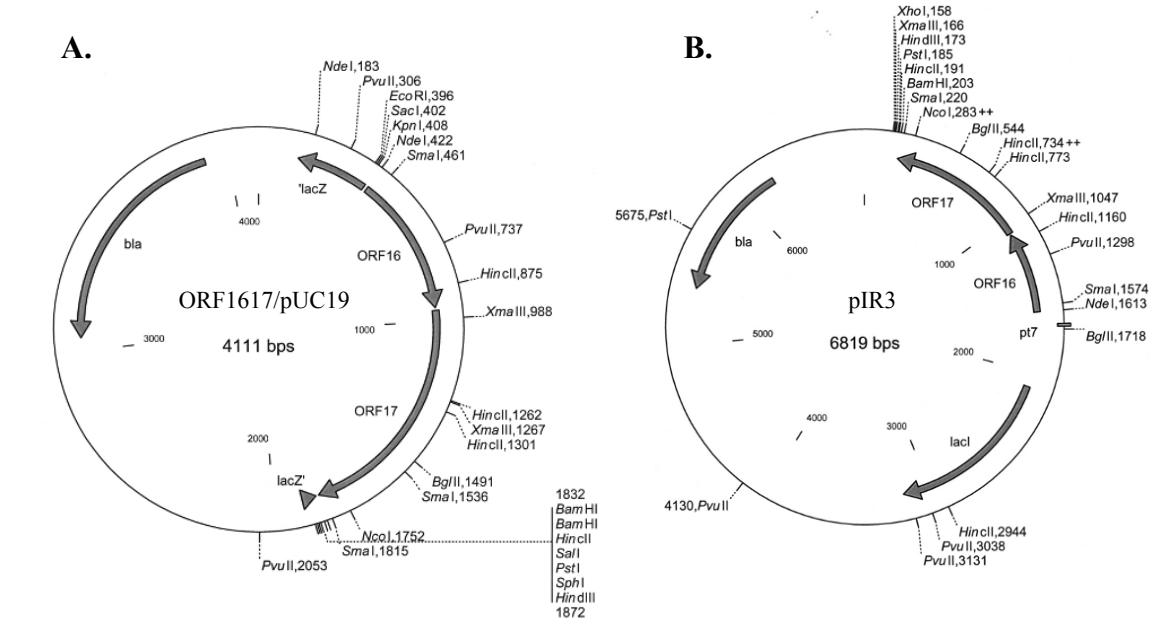

Figure 12. Plasmid maps of the mediating co-expression construct (A) and the final co-expression construct pIR3 (B). Constructs consist of a respective vector, pUC19 in part A and pET22b(+) (Novagen) in part B, and an insert presenting MCPs coding segment *ORF16* to *ORF17* from bacteriophage P23-77 genome.

4.4.2 Co-expression experiment

Plasmid pIR3 was transferred in competent *E. coli* HMS174(DE3) cells with standard protocols, expressed and proteins were initially extracted (Materials and methods 3.2.1 and 3.2.1.1). After ultracentrifugation, samples collected during different stages of the process were analyzed in 15% SDS-PAGE. VP16 was observed to be expressed in higher amount than VP17, but both were present in the sample and soluble. The co-expression construct is suitable for producing VP16 and VP17 in the same cell with aforementioned protocols.

4.5 Nucleic and amino acid sequencing

Nucleic acid sequencing proved expression plasmids pIR1 (*ORF17*), pIR2 (*ORF16*) and pIR3 (*ORF16* + *ORF17*) to contain unmodified and complete target genes. Amino acid se-

quencing was conducted for AIX purified VP16 and VP17 samples (in 20 mM Tris-HCl – 50 mM NaCl buffer at pH 7.4, 2-3 mg protein/ml) at the Institute of Biotechnology at Helsinki University. Amino acid sequencing proved the authenticity of the recombinant proteins. The amino-terminal amino-acids of VP16 were MQEAFNR. For VP17, the N-terminal amino-acid sequence revealed the cleavage of methionine as the sequence was GVFDRIRGAL.

5. DISSCUSSION

The origin of viruses is in close connection with the origin of life (Hendrix et al., 2000; Jalasvuori et al. 2009). Despite its importance, the evolutionary history of viruses is unclear, since their vastly diverse morphologies and differentiated genomes resist feasible classification. Recent approach emphasizes conserved key structures as a method for identifying groups of related viruses (Bamford et al., 2002).

This study focused on a novel bacteriophage P23-77, that exhibits a curious combination of structural features and is speculated to represent an unknown ancestral branch of the double β-barrel superlineage of viruses (Jaatinen et al., 2008; Bamford and Krupociv, 2008; Jalasvuori et al., 2009). The capsid of P23-77 is suggested to consist of two protein species, VP16 (20 kDa) and VP17 (32 kDa), that are arranged in capsomeres closely resembling those of an archaeal virus SH1. The atomic-level structures of the coat proteins of both viruses are unknown. However, the proteins are envisioned to exhibit single β-barrel folding, since these units would correspond to the capsomer morphology (Jäälinoja et al., 2008). In addition to SH1, P23-77 has multiple similarities with other viruses. Notably, it has an ATPase motif found in bacteriophage PRD1, a member of the double β-barrel lineage (Jalasvuori et al., 2009). The double β-barrel lineage is a group of viruses suggested to be related due to their highly similar and conserved vertical jelly roll MCP structures (Benson et al., 1999; Bamford., 2003). Viruses of the lineage are diverse and infect hosts from all three domains of life.

These two findings, the similarity to arhaeal virus SH1 and to the PRD1-adenoviral lineage, present an interesting issue: do P23-77 MCPs exhibit the unprecedented single β-barrel fold as well as similarity to the double β-barrel lineage? What does it imply if there is such a similarity? Since there is relation between P23-77 and archaeal elements, and possible relation to eukaryote viruses through the double β-barrel lineage, it can be speculated that the viruses had a common ancestor predating the divergence of the domains of life. Information about the atomic structures of the two MCPs of P23-77 can reveal whether a single β-barrel fold is present. This might define P23-77 a representative of a

branch that gave rise to the double β-barrel lineage. Thus, research on P23-77 MCPs aspires to shed light to events in the distant evolutionary history of viruses. If the coat proteins of P23-77 exhibit the predicted single β-barrel structure similar to the double β-barrel lineage, evolutionary conclusions are extended to a wide group of viruses. The double β-barrel structure itself can be a duplication derivative of the motifs present in early single β-barrel viruses.

X-ray crystallography presents an effective method for VP16 and VP17 structural analysis, but requires the proteins to be highly purified and crystallized in a well-ordered lattice. In this study, the exact conditions for purifying the two MCPs of P23-77 were examined with multiple AIX, CIX and SEC experiments. During the research, extremely high heat resistance was observed for VP16 and VP17. High thermostability of the coat proteins is an adaptation to the extreme environment of the *Thermus* host, and exceeds the optimal growth temperature of *Thermus thermophilus* (+70 °C to +75 °C, Ramalay and Hixson, 1970; Dworkin et al., 2006) by 20 degrees. The heat resistance of VP16 and VP17 was exploited in the protein purification process.

VP16 and VP17 did not bind in IEX columns in conditions predicted by their isoelectric points. For VP16, this is a beneficial effect: in a high pH AIX most of the host-derived protein impurities bind to the column while the purified VP16 can be collected from the flow-through (buffer conditions: 20 mM Ethanolamine – 50 mM NaCl at pH 9). For VP17, conditions that induced binding in AIX were discovered: with 20 mM Ethanolamine buffer at pH 9.5, VP17 binds to the column and eluates specifically at 5% of 1M NaCl buffer gradient (concentration of 50 mM NaCl in the column). Since most buffers that were initially utilized in IEX experiments had 50 mM NaCl, it would be informative to repeat the experiments with salt free buffers. Despite difficulties in finding binding conditions for the capsid proteins, efficient AIX purification protocols for VP16 and VP17 were established.

SEC experiments advanced VP16 and VP17 purification. Final protocols that allowed purification of high concentration of proteins were established (see Materials and methods 3.2.3). SDS-PAGE analysis of the purified VP16 revealed the protein pure enough to be enrolled in crystallization trials. Even purified VP16 showed a double band in SDS-PAGE,

but since this band seemed to share the exact same charge, size and heat resistance properties than VP16, it was concluded to be an isomorph of the native protein. The double band did not hinder crystallization. Crystallization was observed in salt free conditions containing low concentrations of PEG1000 and PEG8000. VP16 crystals were proven to contain protein but experiments with in house X-ray sources yielded no diffraction pattern. It could be that the crystals were too small, the equipment inefficient, or the protein lattice in the crystals unorganized. Nevertheless, the diffraction experiments conducted at Jyväskylä University should not be taken too conclusively, since the equipment is usually deployed for the analysis of small organic compounds instead of biological macromolecules. In addition, X-ray exposure only lasted for a few minutes.

Besides establishing the purification procedures for VP16 and VP17, research was conducted on their other significant characteristics. Gradient centrifugation and analytical SEC revealed that the recombinant VP16 and VP17 are expressed in a form that sediments and elutes slower than 44 kDa mass reference. This implies that VP17 is a monomer (32 kDa) and VP16 is either a monomer (20 kDa) or a dimer. For more specific results, standard proteins with smaller molecular mass are required; in addition, gradient centrifugation time should be elongated to provide enhanced separation.

Since VP16 and VP17 are the MCPs of P23-77, they are likely to have interaction if expressed in the same host. To observe such interactions, a co-expression construct pIR3 was produced. pIR3 features both MCP-encoding ORFs inserted in a pET22b(+) (Novagen) vector. *E. coli* HMS174(DE3) cells transformed with pIR3 were observed to produce both VP16 and VP17, though VP16 was expressed in higher amounts. This could be caused by an effect of the T7 promoter on VP16, or a higher natural copy number of VP16 in the P23-77 virion. VP16 is suggested to form the hexameric single β-barrel base of the P23-77 capsomers, while VP17 molecules would be positioned on top of this base in two bodies (Jaatinen et al., 2008). This arrangement is highly similar to the capsomer structure of the arhaeal virus SH1 (Jäälinoja et al., 2008).

To summarize, the P23-77 coat proteins VP16 and VP17 were studied to establish purification protocols yielding crystallization-grade native protein material. This was achieved for

VP16, though the purification could still be optimized. Furthermore, crystallization conditions were established for VP16. The conditions contained no salt and thus eliminated the possibility of salt pseudocrystals. VP17 was purified to a high level, but the purification process requires further optimizing before crystallization trials can begin.

Future aspects of the project include optimizing the purification of VP17 and establishing a crystallization protocol for the protein. The crystals featuring VP16 should be enrolled in more informative diffraction experiments after further optimization to enhance crystal size. The double band of VP16 should also be identified. The protocols and experiments of this study have considerably facilitated the production of diffracting VP16 and VP17 crystals. Finally, their three-dimensional atomic arrangement should be determined and analyzed to extract the information they undoubtedly hold.

6. REFERENCES

Abrescia, N. G. A., H. M. Kivelä, J. M. Grimes, J. K. H. Bamford, D. H. Bamford, and D. I. Stuart. 2005. Preliminary crystallographic analysis of the major capsid protein P2 of the lipid-containing bacteriophage PM2. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 61:762–765.

Ackermann, H. W. 2003. Bacteriophage observations and evolution. *Res. Microbiol.* 154:245–251.

Bamford, D. H., R. M. Burnett, and D. I.Stuart. 2002. Evoluton of viral structure. *Theor. Pop. Biol.* 61:461–470.

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

Bamford, D. H., J. J. Ravantti, G. Rönnholm, S. Laurinavicius, P. Kukkaro, M. Dyall-Smith, P. Somerharju, N. Kalkkinen, and J. K. H. Bamford. 2005. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J. Virol.* 79:9097–9107.

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

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

Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 2004. Does common architecture reveal a viral lineage spanning all three domains of life? *Mol. Cell* 16:673–685.

Bergfors, T. M. 2009. Protein Crystallization, 2nd edition. International University Line, La Jolla. 474 pp.

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

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.

Brock, T.D., 2005. Genus Thermus. In: Holt, J.G. (Ed.), Bergey's manual of systematicbacteriology, vol. 1. Williams & Wilkins, pp. 333–337.

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

Butcher, S. J., T. Dokland, P. Ojala, D. H. Bamford, and S. D. Fuller. 1997. Intermediates in the assembly pathway of the double stranded RNA virus phi6. *EMBO J.* 16:4477–4487.

Comeau, A. M., and H. M. Krisch. 2005. War is peace-dispatches from the bacterial and phage killing fields. *Curr. Opin. Microbiol.* 8:488–494.

Diprose, J. M., J. N. Burroughs, G. C. Sutton, A. Goldsmith, P. Gouet, R. Malby, I. Overton, S. Zientara, P. P. C. Mertens, D. I. Stuart, and J. M. Grimes. 2001. Translocation portals for the substrates and products of a viral transcription complex: The bluetongue virus core. *EMBO J.* 20:7229–7239.

Dworkin, M., S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt. 2006. The Prokaryotes volume 7, 3rd edition. Springer, New York. 1074 p.

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

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

d'Herelle, F. 1924. Immunity in Natural Infectious Disease, Williams & Wilkins Co., Baltimore. 399 pp.

Jalasvuori, M., and J. K. H. Bamford. 2008. Structural co-evolution of viruses and cells in the primordial world. *Orig. Life Evol. Biosph.* 38:165–181.

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

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

Juhala, R.J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299: 27–52.

Jäälinoja, H. T., E. Roine, P. Laurinmäki, H. M. Kivelä, D. H. Bamford, and S. J. Butcher. 2008. Structure and host-cell interaction of SH1, a membranecontaining, halophilic euryarchaeal virus. *Proc. Natl. Acad. Sci. USA* 105: 8008–8013.

Kidambi, S.P., S. Ripp, and R. V. Miller. 1994. Evidence for phage-mediated gene transfer among pseudomonas aeruginosa strains on the phylloplane. *Appl. Environ. Microbiol.* 60:496–500.

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

Brock, T. D, 2005. Genus Thermus. In: Holt, J.G. (Ed.), Bergey's manual of systematicbacteriology, vol. 1. Williams & Wilkins, pp. 333–337

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

Laurinmäki, P.A., J. T. Huiskonen, D. H. Bamford, and S. J. Butcher. 2005. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure* 13:1819–1828.

Lehtonen, J. V., D. J. Still, V. V. Rantanen, J. Ekholm, D. Björklund, Z. Iftikhar, M. Huhtala, S. Repo, A. Jussila, J. Jaakkola, O. Pentikäinen, T. Nyrönen, T. Salminen, M. Gyllenberg, and M. Johnson. 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J. Comput. Aided Mol. Des.* 18:401-419.

Lucchini, S., F. Desiere, and H. Brüssow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* 73:8647–8656.

Matsushita, I., and H. Yanase. 2008. A novel thermophilic lysozyme from bacteriophage phiIN93. *Biochem. Biophys. Res. Commun.* 377:89–92.

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

Olsen, R.H., J.-S. Siak, and R. H. Gray. 1974. Characteristics of PRD1, a plasmid dependent broad host range DNA bacteriophage. *J. Virol*. 14689–699.

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

Pichla-Gollon, S.L., M. Drinker, X. Zhou, F. Xue, J. J. Rux, G.-P. Gao, J. M. Wilson, H. C. J. Ertl, R. M. Burnett, and J. M. Bergelson. 2007. Structure-based identification of a major neutralizing site in an adenovirus hexon. *J.Virol*. 81:1680-1689.

Pritham, E. J, T. Putliwala, and C. Feschotte. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390:3–17.

Ramaley, R. F., and J. Hixson. 1970. Isolation of a nonpigmented, thermophilic bacterium similar to Thermophilic bacterium similar to *Thermus aquaticus*. *J. Bacteriol.* 103:527–528.

Reinisch, K. M., M. L. Nibert, and S. C. Harrison. 2000. Structure of the reovirus core at 3.6 Å resolution. *Nature* 404:960–967.

Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual, 3rd edition. Cold Spring Harbor Laboratory Press, New York. 2344 p.

San Martin, C., R. Burnett, F. de Haas, R. Heinkel, T. Rutten, S. D. Fuller, S. J. Butcher, and D. H. Bamford. 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.

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

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

Strömsten, N. J., D. H. Bamford, and J. K. Bamford. 2005. In vitro DNA packaging of PRD1: a common mechanism for internal-membrane viruses. *J. Mol. Biol.* 348:617–629.

Studier, F.W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219:37–44.

Urbonavicius, J., S. Auxilien, H. Walbott, K. Trachana, B. Golinelli-Pimpaneau, C. Brochier-Armanet, and H. Grosjean. 2008. Acquisition of a bacterial RumA-type tRNA(uracil-54, C5)-methyltransferase by *Archaea* through an ancient horizontal gene transfer. *Mol. Microbiol.* 67:323–335.

Witzany, G. 2006. Natural genome-editing competences of viruses. *Acta Biotheor.* 54:235–253.

Woese, C. R. 1998. The universal ancestor. Proc. *Natl. Acad. Sci. USA* 95:6854–6859.

Woese, C. R. 2002. On the evolution of cells. *Proc. Natl. Acad. Sci. USA* 99:8742–8747.

Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64:69–114.

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

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

7. APPENDICES

Appendix I:

Plasmid maps of the expression plasmids for VP16 and VP17.

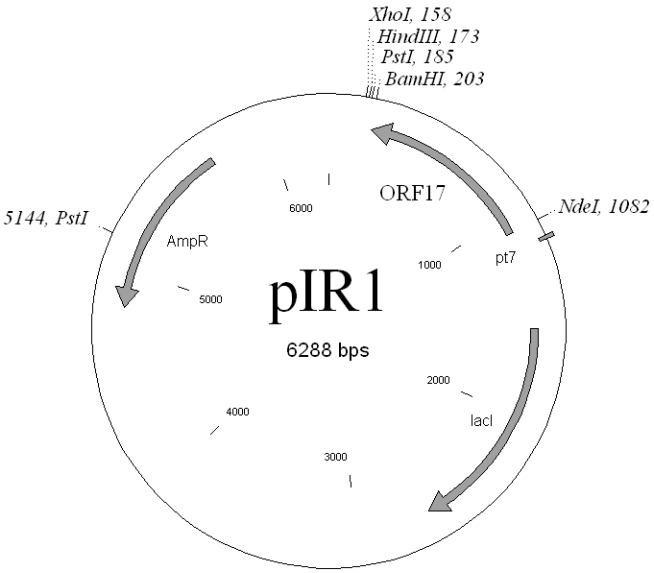

Figure A1. Expression plasmid pIR1. The plasmid consists of an pET22b(+) (Novagen) vector with *ORF17* as an insert. Plasmid was utilized to express VP17, a MCP of bacteriophage P23-77.

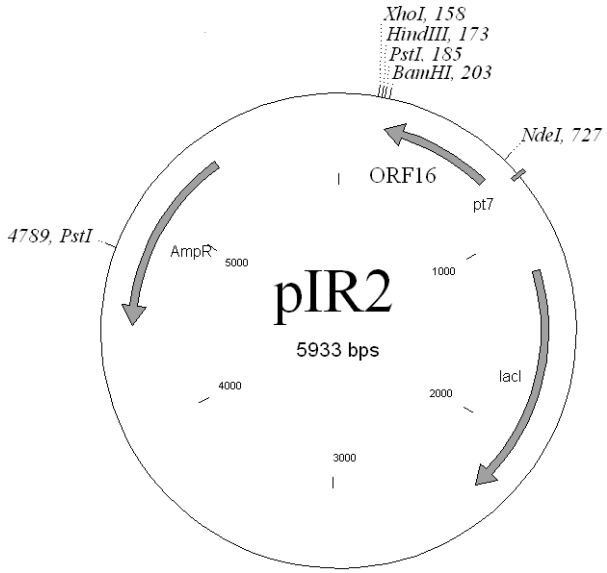

Figure A2. Expression plasmid pIR2. The plasmid consists of an pET22b(+) (Novagen) vector with *ORF16* as an insert. Plasmid was utilized to express VP16, a MCP of bacteriophage P23-77.

Appendix II:

AGE protocol

Buffers :

50 x TAE buffer :

- Trizma base (Sigma T1503) 242 g
- Glacial acetic acid 57.1 µl
- 0.5 M EDTA pH 8 100 µl

1 x TAE buffer was deployed in AGE. To prepare 1 x TAE, 50 x TAE was diluted in water 1:50.

5 x Loading buffer :

- 20 % Ficol 400
- 20 mM EDTA pH 8
- 0.1 % Bromphenol blue
- 0.1 % Xylene cyanol

1% agarose gel was prepared dissolving 1 weight percentage of agarose (EuroClone) in 1 x TAE buffer, adding ethidiumbromide (0.4 µg/ml) and casting the gel with respective well combs. DNA-containing samples were mixed with 5 x Loading buffer and applied to the gel with a size standard (Gene Ruler 1 kb Plus DNA Ladder, Fermentas). AGE was conducted with a current of 100 volts until the samples had separated (typically 1-2 hours). Afterwards, the gel was observed and photographed under ultraviolet lighting.

Appendix III:

SDS-PAGE protocol

Buffers :

Buffer I : Stacking gel buffer
- 363 g Trizma base (Sigma T1503)
- pH adjusted to 8.6 with 37% HCl
- Volume adjusted to 1000 ml with water

Buffer II : Running gel buffer
- 30 g Trizma base (Sigma T1503)
- 30.34 ml NaH_2PO_4 x H_20
- pH adjusted to 7.8 with 37% HCl
- Volume adjusted to 500 ml with water

3 x Sample buffer
- 10 ml buffer I (Stacking gel buffer)
- 10 ml 10% SDS (BDH, 442444H, diluted 1:10 in water)
- 0.4 ml 0.5 M EDTA
- 1 ml 2-mercaptoethanol
- 10 ml 87% glycerol
- 10 mg Bromphenol blue
- Mix and store in -20°C

RB: Running buffer
- 7.71 g Trizma base (Sigma T1503)
- 36.99 g glycine
- 30 ml 10% SDS (BDH, 442444H, diluted 1:10 in water)
- Volume adjusted to 3 l with water
- pH adjusted to 8.5

1:4 buffer
- 100 ml buffer II (Running gel buffer)
- 4 ml 10% SDS (BDH, 442444H, diluted 1:10 in water)
- 100 ml water
- pH adjusted to 8.55 - 8.6.

MR-Spacer buffer
- 31.23 ml 40% acrylamid (Bio-Rad, 161-0148)
- 31.3 ml Buffer I (Stacking gel buffer)
- 2.3 ml 10% SDS (BDH, 442444H, diluted 1:10 in water)
- Volume adjusted to 250 ml with water

SDS-PAGE gels were cast with the aforementioned buffers according to the following description.

Concentration gel:

- MR Spacer 32,5 ml
- TEMED 32,5 µl
- 10 % APS 325 µl

Separation gel:

- 40% acrylamid (Bio-Rad, 161-0148) 18,8 ml
- 1:4 buffer 15 ml
- H2O 16,2 ml
- TEMED 30 µl
- 10% APS 300 µl

Protein samples for SDS-PAGE were prepared mixing 20 µl of sample and 7 µl of 3 x SB. The mixture was boiled for 5-10 minutes and mixed thoroughly. SDS-PAGE system was set up and equipped with a gel in RB-buffer. Samples and a molecular weigh standard (Precision Plus Protein Standard, Bio-Rad) were administered to the wells on the upper-most part of the concentration gel and SDS-PAGE was conducted for 60 min with current of 200 volts and 200 milliamperes.

After the electrophoresis, the gel was dyed with Coomassie stain, washed with 10% acetic acid and photographed with Bio-Rad imaging system.

Appendix IV :

Table of crystallization conditions in manual optimization experiments

Table A1. Conditions tested in VP16 crystallization optimization
Table 2 features all of the conditions used in the crystallization optimization process of VP16. The buffer components present the percentages of polyethylene glycol (PEG) used in each buffer. "1000" stands for PEG 1000, "1500" for PEG 1500 and "8000" for PEG 8000 (given values represent the PEG % in buffer). "Phos" stands for potassium phosphate (given values represent the mM concentration of phosphate in buffer).

Plate name	Buffer components				Seeding dilution ratio					
	1000	1500	8000	Phos.	none	1:1	1:10	1:100	5:100	1:1000
First optimization, PEG		8	8, 9, 10, 11		X					
		10			X					
		12			X					
First optimization, p. phosphate			15,18, 20,22	0,04	X					
				0,05	X					
				0,06	X					
First seeding		4,	4, 5, 6, 7			X				
		5					X			
		6						X		
		7							X	
		8, 9								X
Second seeding		5	5			X	X	X		X
		7	6			X	X	X		X
		9	4			X	X	X		X
Third seeding		5	4				X	X		
		5	5				X	X		
		6	7				X	X		
Seeding 1.2	4		4, 5, 6, 7				X			
	5							X		
	6									X
Log 1	4		4, 5, 6, 7					X		X
	5							X		X
	6							X		X
Multiply seed	6		4							X
			5							X
			6							X
			7							X