Electron cryo-microscopy studies of bacteriophage ϕ8 and archaeal virus SH1

Harri Jäälinoja

Institute of Biotechnology and
Department of Biological and Environmental Sciences
Division of Genetics
Faculty of Biosciences
University of Helsinki

and

National Graduate School in Informational and Structural Biology

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Biosciences of the University of Helsinki for public examination in the auditorium 2402 of Biocenter 3, Viikinkaari 1, Helsinki, on February 16th, 2007, at 12 noon

HELSINKI 2007

Supervisor

Docent Sarah J. Butcher
Institute of Biotechnology
University of Helsinki

Reviewers

Docent Roman Tuma
Faculty of Biosciences
University of Helsinki

Docent Maarit Suomalainen
Haartman Institute
University of Helsinki

Opponent

Professor Roger Burnett
The Wistar Institute
Philadelpia, U.S.A.

ISBN 978-952-10-3739-9 (paperback)
ISBN 978-952-10-3740-5 (PDF, http://ethesis.helsinki.fi/)

ISSN 1795-7079

Yliopistopaino, Helsinki University Printing House
Helsinki 2007

Tuulalle ja Heikille

Original publications

This thesis is based on the following articles[1], which are referred to in the text by their Roman numerals.

I Jäälinoja HT, Huiskonen JT, Butcher SJ. Electron cryo-microscopy comparison of the architectures of the enveloped bacteriophages ϕ6 and ϕ8. *Structure*. In press.

II Huiskonen JT, Jäälinoja HT, Briggs JA, Fuller SD, Butcher SJ. Structure of a hexameric RNA packaging motor in a viral polymerase complex. *J Struct Biol*. 2006 doi: 10.1016/j.jsb.2006.08.021.

III Jäälinoja HT, Laurinmäki P, Butcher SJ. The T=28 haloarchaeal virus SH1 has a protein shell covering a lipid membrane. Manuscript.

Unpublished results will also be presented.

[1] Articles I and II are reproduced with permission from Elsevier.

Abbreviations

2D	two-dimensional
3D	three-dimensional
BTV	Bluetongue virus
cryoEM	electron cryo-microscopy
CTF	contrast transfer function
DNA	deoxyribonucleic acid
ds	double-stranded
EM	electron microscopy
FEG	field emission gun
gp	gene product
HK97	bacteriophage Hong Kong 97
HSV-1	herpes simplex virus type 1
keV	kiloelectronvolt
NC	nucleocapsid
PBCV-1	*Paramecium bursaria chlorella* virus type 1
PC	polymerase complex
PFT	polar Fourier transform
RDV	rice dwarf virus
RNA	ribonucleic acid
ss	single-stranded
STIV	*Sulfulobus* Turreted Icosahedral Virus
T	triangulation number

Summary

Symmetry is a key principle in viral structures, especially the protein capsid shells. However, symmetry mismatches are very common, and often correlate with dynamic functionality of biological significance. The three-dimensional structures of two isometric viruses, bacteriophage ϕ8 and the archaeal virus SH1 were reconstructed using electron cryo-microscopy. Two image reconstruction methods were used: the classical icosahedral method yielded high resolution models for the symmetrical parts of the structures, and a novel asymmetric *in-situ* reconstruction method allowed us to resolve the symmetry mismatches at the vertices of the viruses. Evidence was found that the hexameric packaging enzyme at the vertices of ϕ8 does not rotate relative to the capsid. The large two-fold symmetric spikes of SH1 were found not to be responsible for infectivity. Both virus structures provided insight into the evolution of viruses. Comparison of the ϕ8 polymerase complex capsid with those of ϕ6 and other dsRNA viruses suggests that the quaternary structure in dsRNA bacteriophages differs from other dsRNA viruses. SH1 is unusual because there are two major types of capsomers building up the capsid, both of which seem to be composed mainly of single β-barrels perpendicular to the capsid surface. This indicates that the β-barrel may be ancestral to the double β-barrel fold.

Table of contents

A. INTRODUCTION

Viruses are obligatory parasites that need the cellular machinery of their host organism for creating progeny. For this reason it is debatable whether viruses can be considered as living organisms (in my opinion they can; see 1.1). Nevertheless, they are the most ubiquitous organisms (dead or alive) on this planet (Hendrix, 2002). Viruses infect organisms belonging to all three domains in the tree of life, the Archaea, the Bacteria and the Eukarya. They thrive wherever their host organism thrives, from the most extreme habitats of the hyperthermophilic archaea to the plush palaces of *Homo sapiens*. Their numbers are overwhelming, particularly those of the bacteriophages (phages for short): a millilitre of seawater can have up to 10^7 phages (Wommack and Colwell, 2000) and a gram of soil may carry as many as 4×10^9 (Williamson et al., 2005). The average turn-over time of the entire phage population in the seas is only a few days (Wilhelm et al., 2002). This means that every second, on the order of Avogadro's number of successful infections take place (Hendrix, 2003). This has obviously an enormously important role in the ecology of the oceans. Similarly, at the time of writing, we humans are alarmed at reports of new strains of avian influenza. Historically, influenza epidemics such as the 1918 Spanish flu have killed millions of people on all continents (Palese, 2004). More recently, the AIDS epidemic has had a major effect on the life expectancy of millions. For these reasons alone, the understanding of virus lifecycles and ecology is of vital importance.

We have, however, learned to use viruses also for our benefit. Bacteriophages are the natural enemies of bacteria. Phage therapy, the use of bacteriophages to fight bacterial infections, was for a long time developed simultaneously with antibiotics (Sulakvelidze and Kutter, 2004). The use of antibiotics prevailed and the research of phage therapy became an obscure sideline, best preserved in the former Soviet Union. Currently, many antibiotics are becoming obsolete: the bacteria they were supposed to fight have developed resistance to them. Methicillin-resistant *Staphylococcus aureus* (MRSA) bacteria are an increasingly common problem in hospitals (Das et al., 2006). As of yet, the scope of the problem is not great enough to really force us to look for alternatives to antibiotics, but when this happens, phage therapy may be one possibility. In addition to phage therapy, viruses or virus-derived particles have found many useful applications in biotechnology. These include the use of phage display, phage vectors in gene therapy (Clark and March, 2006) and the use of viral polymerases for production of RNA (Aalto et al., in press).

Virus research resembles the painting of a large fresco, too large for one artist to physically be able to realize his vision. Thus the task is divided into smaller pieces and delegated to experienced masters who take responsibility for a particular part of the whole, and who in turn hire less experienced, even novice painters for the various tasks at hand. During this thesis work, I have been learning to add color to ears. I have added color to the ears of a figure who together with her relatives has for a long time already been on the fresco, ϕ8 (Article I). During this time, the group of painters I am working with also started to sketch a new figure, called SH1, on the canvas, and I got to color his ears as well (Articles III). My senior colleagues have also come up with a nice new way of dabbing the color with the brush, so that the various intricate undulations of the earlobe are not smeared when the color is added (Briggs et al., 2005). This technique I have also been developing (Articles II &

III). In this Introduction I tell about the various figures, the ear-coloring, and various other things I have learnt at the studio.

1. Symmetry in viruses

At the simplest, a virus may consist of only the genome and a protein shell. The function of the protein shell is to protect the genome from environmental and enzymatic damage, and to provide means to find and infect a new host cell. The basic design criterion of the shell is that it must be large enough to accommodate the genome. It was realized early (Crick and Watson, 1956) that the most economical way to accomplish this is to have the shell consist of multiple copies of the same protein. The fact that the protein-protein interactions between the identical copies have some specific directionality necessarily usually leads to a symmetrical arrangement for the shell (not always as we now appreciate *e.g.* in case of paramyxoviruses (Bhella et al., 2004)). The two symmetries most commonly found in viral capsids are helical and icosahedral symmetry. These result in filamentous and spherical (isometric) viruses, respectively.

1.1. Helical symmetry

In helically-symmetric viruses, the coat protein is stacked in a staircase-like manner around the nucleic acid molecule that is also in a helical conformation (Watson and Crick, 1953). This results in a filamentous virus, possibly the best known example of which is the Tobacco Mosaic Virus, one of the first virus ever to have been crystallized (Watson, 1954) and assembled *in vitro* (Fraenkel-Conrat and Williams, 1955). For a helical virus, the length of the genome determines how long the virus will be. Sometimes one or both ends of the helix are capped with another protein, as in potato potyvirus (Torrance et al., 2006).

1.2. Icosahedral symmetry

An icosahedrally symmetric (also denoted as 532 symmetric) object has 15 two-fold, 10 three-fold and 6 five-fold axes of symmetry. Two platonic solids, the regular icosahedron and the regular dodecahedron, incorporate icosahedral symmetry (Figure 1). The regular icosahedron has 20 equilateral triangular faces, 12 pentagonal vertices and 30 edges. The regular dodecahedron has 12 pentagonal faces, 20 triangular vertices and 30 edges. Both polyhedra have 60 asymmetric units as defined in Figure 1.

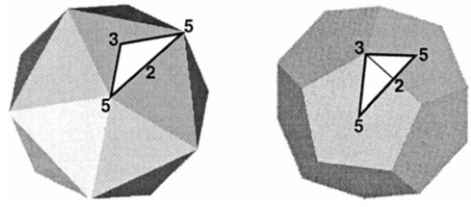


Figure 1. Icosahedral symmetry. An icosahedron (left) and a dodecahedron (right) with symmetry axes and the asymmetric unit used by microscopists. The numbers (2, 3 and 5) indicate the positions of some of the symmetry axes. The white triangle defines the asymmetric unit which is bounded by the lines joining adjacent five-fold and three-fold positions. With permission from (Baker et al., 1999) © (1999) American Society for Microbiology.

Of all the platonic solids, the regular icosahedron has the highest volume per surface area, indicating that it is the most economical geometry for building viral capsids.

1.2.1. The theory of quasi-equivalence

The simplest icosahedral virus capsids, such as that of parvovirus (Simpson et al., 1998), consist of only 60 copies of the capsid protein, one per each asymmetric unit. In this case, all subunits are in identical structural environments. The subunit positions are said to be equivalent.

In case a larger capsid is needed, it is not economical to use bigger subunits, as this would require a longer genome to produce the larger proteins and thus a yet larger capsid to contain the genome. Instead, more copies of the same subunit are used. Icosahedral symmetry can still be maintained by having pentamers (*i.e.* complex of five subunits) at the vertices and hexamers (*i.e.* complex of six subunits) distributed evenly at other locations. In this case, however, the structural environments of the subunits are no longer exactly the same. Instead of equivalent, they are said to be *quasi-equivalent* (Caspar and Klug, 1962).

The geometry of the lattice formed by the pentamers and hexamers can be described by the triangulation number *T*. $T=h^2+hk+k^2$, where the integers *h* and *k* are the lattice coordinates of a pentamer relative to another pentamer (Figure 2 left). *T* describes how each of the triangular faces is subdivided into smaller triangles. Each smaller triangle corresponds to three protein subunits, giving a total number of 60*T* subunits in the icosahedral shell. *T* is also the number of subunits per asymmetric unit. If $h \neq 0$, $k \neq 0$ and $h \neq k$, two mirror-image lattices exist for the same *T*. In this case the resulting viral capsid is either left- or right-handed, *laevo* or *dextro*, respectively.

Also prolate capsid geometries exist, where one or more layers of hexagons are added around one five-fold axis of symmetry, and thus the capsid is not strictly icosahedral. An example of such architecture an is found in bacteriophage ϕ29 (Tao et al., 1998).

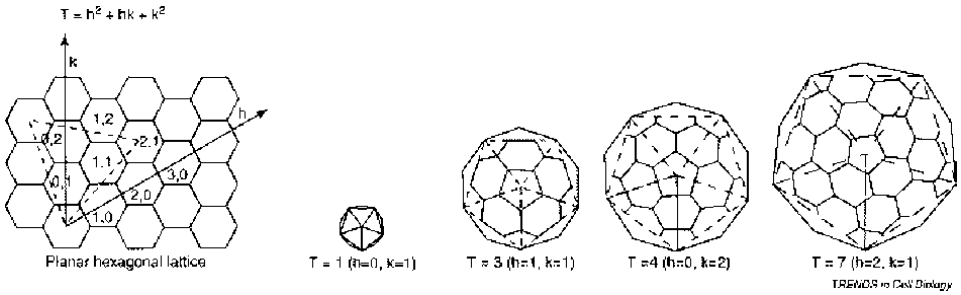


Figure 2. Quasi-equivalence of subunits in shells of icosahedral viruses: this principle (Klug and Caspar, 1960) explains how closed icosahedral capsids are constructed from multiples of 60 protein subunits (60*T*), organized as hexamer and pentamer units. Curvature can be introduced in a flat hexamer net by replacing some hexamers with pentamers. A closed shell is generated by inserting 12 pentamers in symmetrical positions. The multiplicity, *T*, depends on the vector (*h*,*k*) between the lattice points in the centres of hexagons in the sheet that become pentamers in the icosahedral shell. The subunits that make up the hexamers or pentamers interact with neighbouring subunits in *T* slightly different ways. Reprinted from (Amos and Finch, 2004), with permission from Elsevier.

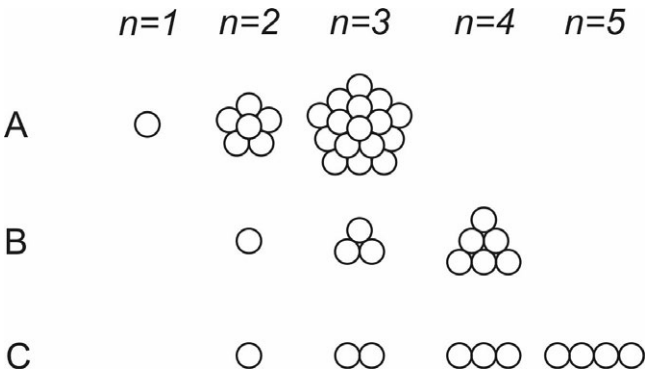


Figure 3. One possible way to divide an icosahedral capsid into symmetrons. A. Pentasymmetrons. B. Trisymmetrons. C. Disymmetrons. Modified from (Wrigley, 1969).

1.2.2. Disymmetrons, trisymmetrons and pentasymmetrons

The theoretical packing of the pentamers and hexamers on the icosahedral capsid was further analysed by Wrigley (1969) in the context of the determination of the structure of a very large virus, the *Sericesthis* Iridescent Virus. He developed a scheme to divide the capsomers in groups of local five-, three- and two-fold symmetry (termed *penta-*, *tri-* and *disymmetrons*, respectively) which could possibly correspond to assembly (or breakdown) intermediates of the virus. The immediate practical use of this scheme was that it allowed him to narrow down the set of possible triangulations of the capsid based on the size of the fragments seen on the micrographs, at a time when direct methods such as three-dimensional (3D) image reconstruction of capsids were just in their infancy (De Rosier and Klug, 1968).

In terms of the triangulation number T, the total number, N, of capsomers in the capsid is $N = 10T+2$. This can be distributed as $N=12p+20t+30d$, where p, t, and d are the numbers of capsomers in the five-, three- and two-fold symmetric groups, respectively. Assuming that the disymmetrons are continuous single lines, the trisymmetrons triangular, and the pentasymmetrons pentagonal (Figure 3), p, t and d may only assume the following set of values:

$$p_n = 1 + 5n(n\text{-}1)/2,$$
$$t_n = n(n\text{-}1) / 2 \text{ and}$$
$$d_n = n\text{-}1,$$

where n = 1,2,3, … For a given T, the values of p, t and d are unambiguously defined, as listed in the Goldberg diagram shown in Figure 3.

A possible explanation as to why the symmetrons might actually correspond to assembly intermediates can be found by studying large virus capsids for which the coat protein atomic structures are available. Paramecium bursaria Chlorella virus type 1 (PBCV-1) is a large dsDNA virus (190 nm vertex to vertex distance). Electron cryo-microscopy (cryoEM) reconstruction has shown the capsid protein to be organized on T=169d lattice, with 20 trisymmetrons of 66 capsomers and 12 pentasymmetrons of 30 capsomers (Yan et al., 2000). The crystal structure of the major coat protein Vp54 has been solved to 2.0 Å resolution (Nandhagopal et al., 2002), and shows that capsomers formed by Vp54 are trimers with double β-barrels fold (see Section 2.2.). The fact that the capsomers are trimers, only pseudo-hexameric, means that they have directionality with respect to the two-fold axis of symmetry. The trisymmetrons

reflect this directionality: all trimers within a trisymmetron face the same way, whereas those within the adjacent trisymmetron, across a two-fold axis of symmetry, face the opposite way (Nandhagopal et al., 2002). A more detailed analysis of inter-capsomer distances within the capsid indicated that the most stable unit is the trisymmetron, and that the interactions along the edges of the trisymmetrons are weaker (Simpson et al., 2003). It can, however, be argued that the inter-capsomers may not necessarily directly correlate with stability of the intersubunit contacts, because of entropic factors.

It should be kept in mind that the sets of possible values of p, t and d depend on the choice of shapes used to represent the symmetrons (*i.e.* it's purely arbitrary and perhaps tells us nothing about the real structure). For example, the trisymmetron, a group of capsomers with three-fold symmetry, need not necessarily be a triangle. A counter-example is given by the adenovirus groups-of-nine hexons (Burnett, 1985): they are certainly arranged in a three-fold symmetric arrangement, centered on the three-fold axis of symmetry, and all the trimers within the group-of-nine face the same way, like those in the trisymmetron of PBCV-1. The groups-of-nine also have a special role in the (dis)assembly of the adenovirus structure, as they have been shown to leave a dissociating virus last, after the penton and the peripentonal trimers as one unit (Prage et al., 1970).

1.2.3. The viral tiling theory

The Caspar-Klug quasi-equivalence theory (Caspar and Klug, 1962) has proven an extremely useful model for analysing virus structures. There are, however, cases that cannot readily be explained by it. For example, inner capsids of double-stranded ribonucleic acid (dsRNA) viruses (see Section 2.4) consist of 120 copies of the major capsid protein, corresponding to T=2 which is not an allowed value of T in the theory (Caspar and Klug, 1962). This is usually explained by saying that the lattice is T=1 formed of asymmetric dimers. Another classic example is simian virus 40 (Liddington et al., 1991) that has pentamers at the sites predicted for hexagons (Figure 4A). This inspired the development of the more general viral tiling theory (Twarock, 2004) of which quasi-equivalence is a special case.

Tiling refers to the covering of a surface with finite number of uniform shapes, called tiles. Essentially, the Caspar-Klug theory describes the possible ways to cover the surface of an icosahedron with equilateral triangles. In the viral tiling theory, other tile shapes also are allowed. The rhomb and kite tiles (Figure 4B) have proven especially useful.

Caspar-Klug theory predicts the locations of the protein subunits. The subunits correspond to the corners of the equilateral facet triangles, locations that are equivalent within the facet and quasi-equivalent within the capsid. The viral tiling theory introduces the *generalized principle of quasi-equivalence.* This states that on a given tile, the protein subunits only correspond to the corners that subtend the same angle, ensuring that the subunits occupy structurally or mathematically equivalent sites. Caspar-Klug theory is therefore incorporated, as the corners of an equilateral triangle naturally subtend the same angle.

Like Caspar-Klug theory, the viral tiling theory also assumes icosahedral symmetry, but there it is recognized that the symmetry can still be maintained even though the surface is described in terms of other shapes than pentamers and hexamers (consisting of equilateral triangles). With other types of tiles, it may be possible for

example to model more spherical structures in a more natural way. An example of viral tiling theory applied to the polyoma virus is shown in Figure 4C.

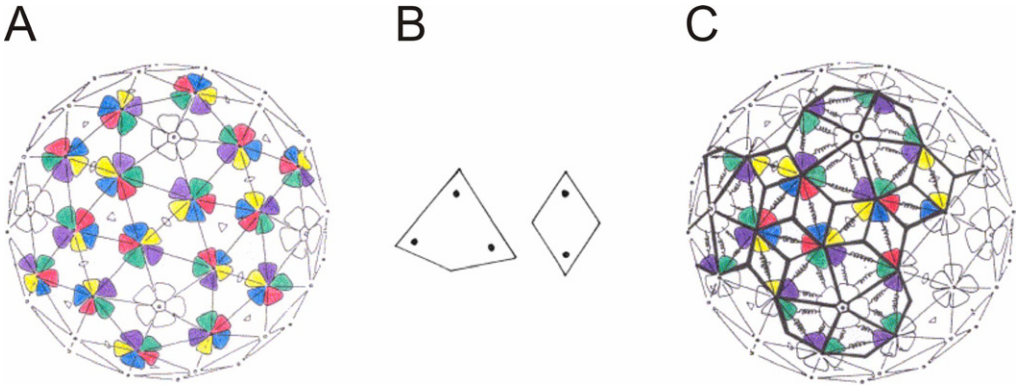


Figure 4. Viral tiling theory. A. Location of the protein subunits of polyoma virus on the 7d hexagonal lattice according to (Rayment et al., 1982). B. Tiles for the tesselation modelling the location of protein subunits for polyoma virus and Simian Virus 40. C. The tesselation for polyoma virus and Simian Virus 40 superimposed on the 7d hexagonal lattice. Spiral arms indicate the location of intersubunit bonds as observed in (Modis et al., 2002). Reprinted from (Twarock, 2004), with permission from Elsevier.

2. Typical folds of viral proteins making icosahedral capsids

As more virus structures have been solved, many different triangulation numbers have been observed. Most importantly, it has been observed that a few capsid protein folds are utilized by many of the known viruses.

2.1. Single β-barrels: +ssRNA viruses

A very common capsid protein fold is the β-barrel, an anti-parallel, eight-stranded fold, also known as the "Swiss roll" or "jelly roll" or RNA virus capsid (RVC) fold. This capsid protein is most typical of small +ssRNA viruses that all exhibit a *T*=3 capsid organization (Harrison, 1990). Examples range from plant viruses such as tomato bushy stunt virus (Hopper et al., 1984) to the black beetle virus (Wery et al., 1994) to various human pathogenic viruses in the group of picornaviruses, such as the human rhinovirus 14 (Arnold and Rossmann, 1988; Rossmann et al., 1985).

2.2. Vertical double β-barrels: adenovirus, PRD1, STIV

Distinctive to this group of capsid proteins is the so called double β-barrel fold (see Figure 5). These proteins usually form trimers where the barrels are normal to the capsid. The trimer corresponds to a hexagon in the Caspar-Klug theory, and therefore half of the monomer corresponds to a single theoretical subunit.

The double β-barrel fold was first observed in the adenovirus major capsid

protein structure (Athappilly et al., 1994; Roberts et al., 1986), but has since been found in viruses infecting organisms in all domains of life. Recent findings verified by x-ray crystallography include the archaeal virus STIV (Khayat et al., 2005) and bacteriophage PRD1 (Abrescia et al., 2004; Benson et al., 1999). The variation in capsid lattices is great, with *T*-numbers like 25 in adenovirus and PRD1 (Butcher et al., 1995; Stewart et al., 1991), 31 in STIV (Rice et al., 2004) and 169 in PBCV-1 (Yan et al., 2000). The highly conserved base of the fold easily allows the construction of capsids of various size to accommodate different sized genomes, whereas loops extending into the milieu may vary between viruses (Figure 5).

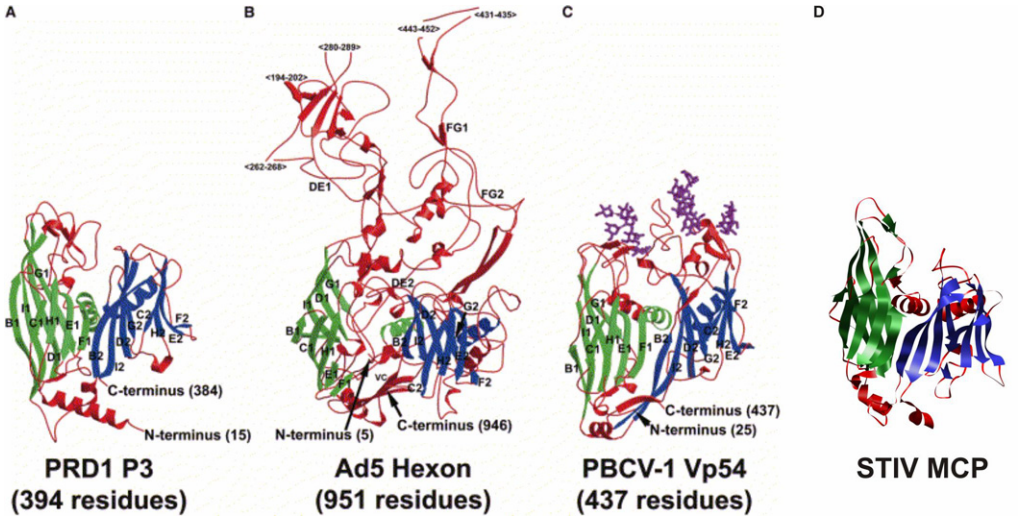


Figure 5. X-ray crystal structures of viral double-barrel trimeric major coat proteins. A. P3 of bacteriophage PRD1 (394 residues; PDB code 1hx6; Benson et al., 2002); B. Hexon of adenovirus type 5 (Ad5; 951 residues; PDB code 1p30; Rux et al., 2003); C. Vp54 of Paramecium bursaria chlorella virus 1 (PBCV-1; 436 residues; PDB code 1m3y; Nandhagopal et al., 2002). The eight β-strands and a flanking α-helix are displayed for the first (green) and second (blue) jelly rolls, and the individual strands are labeled (B1-I1 and B2-I2, respectively). The N- and C-terminal positions are marked, along with the first and last residues modeled in the structures. The four major loops (DE1, FG1, DE2, and FG2), the jelly roll separation domain (VC) and the residue numbers for the unobserved parts of the molecule (<…>) are labeled in Ad5 hexon. In Ad5 hexon, the VC domain and the DE2 loop separate the jelly rolls to produce a molecule with a broader base compared to PRD1 P3 and PBCV-1 Vp54. PBCV-1 Vp54 contains N linked sugars (magenta). Reprinted from (Benson et al., 2004), with permission from Elsevier. D. STIV MCP (PDB code 2bbd; Khayat et al., 2005).

2.3. The α/β-fold

The crystal structure of the mature empty capsid of bacteriophage HK97 (Wikoff et al., 2000) revealed a new type of viral capsid protein fold. The fold of the major capsid protein is a mix of α-helical and β-strand motifs, organized into two compact and spatially distinct domains (see Figure 6). The protein forms pentamers and hexamers on the T=7 lattice. The two main domains of the protein are the A (axial) domain that is located near the local symmetry axes of the pentamers and hexamers, and the P (peripheral) domain that is on the outer

edge of the capsomers, along with extension domains of the N-arm and the E-loop. The most striking feature of the HK97 capsid is that the E-loop is cross-linked with a hair-pin loop of neighbouring subunit within the adjacent asymmetric unit. This creates topological links between the proteins, an uncommon feature in molecular architecture. The molecular chain-mail is obviously an important stabilizing feature for the capsid (Conway et al., 1995; Duda et al., 1995).

Fairly recently it has turned out that the α/β-fold is not unique to the icosahedrally symmetric HK97, but has also been found in a virus that constructs a prolate head: the crystal structure of bacteriophage T4 major capsid protein gp24 shows a similar fold (Fokine et al., 2005). T4 capsid consists of two proteins: gp24 forms the pentamers of the capsid, and gp23 forms the hexamers. The folds of gp23 and gp24 are similar. Both proteins are proteolytically cleaved after the procapsid is assembled. The sequences of the truncated forms gp23* and gp24*, which are present in the mature virion, are similar enough to allow modelling based on the crystal structure of gp24. The models were successfully fitted into the cryoEM reconstruction (Fokine et al., 2005) of the mature virion to yield a pseudo-atomic model for the whole virion.

The observation of the conservation of the fold has been extended from the crystal structures of HK97 and T4 gp24 to existing cryoEM work (Baker et al., 2005). The same fold is probably present also in the major capsid proteins of bacteriophage P22 (Jiang et al., 2003) and the herpes simplex virus 1 that infects humans (Baker et al., 2005; Fokine et al., 2005; Zhou et al., 2000).

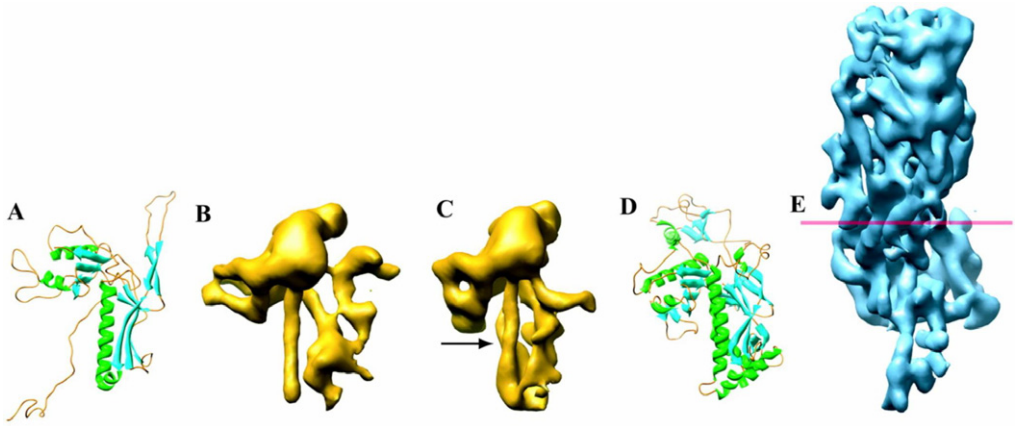


Figure 6. Helix/sheet capsid proteins. A gallery of bacteriophage capsid protein structures determined by either X-ray crystallography or cryoEM. HK97 gp5 (A), mature P22 gp5 (B), procapsid P22 gp5 (C), and T4 gp24 (D) are shown in comparison to HSV-1 VP5 (E). VP5, the 145-kDa capsid protein, was segmented from an approximately 8-Å cryoEM map of the HSV-1 capsid. The red line demarcates the boundary between the floor domain and the other two domains of VP5 (upper and middle domains). The N-terminal helix in P22 that has been proposed to undergo refolding is indicated by the arrow in panel C. With permission from (Baker et al., 2005) © (2005) American Society for Microbiology.

2.4. The highly α-helical fold: dsRNA viruses

The third common capsid protein type has a highly α-helical fold and is found in dsRNA viruses. In contrast to the β-sheet rich capsid protein types which have been found in capsids of varying sizes and triangulations, this group of proteins has only been found to create one kind of shell that always consists of 120 copies of the protein. Figure 7 shows examples of this fold. It can be seen that the structures are really very similar, with the exception of bacteriophage ϕ6, where the capsid protein monomers do not inter-digitate at all. Instead, the structure shows a clear dodecahedral cage (Huiskonen et al., 2006a). In the others there are only slight differences in the roundedness of the capsid. It should, however, be borne in mind that the ϕ6 subunit boundaries in Figure 7 are based on a manual segmentation of a 7.5-Å resolution cryoEM reconstruction (Huiskonen et al., 2006a), not on an x-ray structure like in the other examples, and therefore the possibility of misinterpretation does exist. The overall structures of these viruses are compared in Section 4.

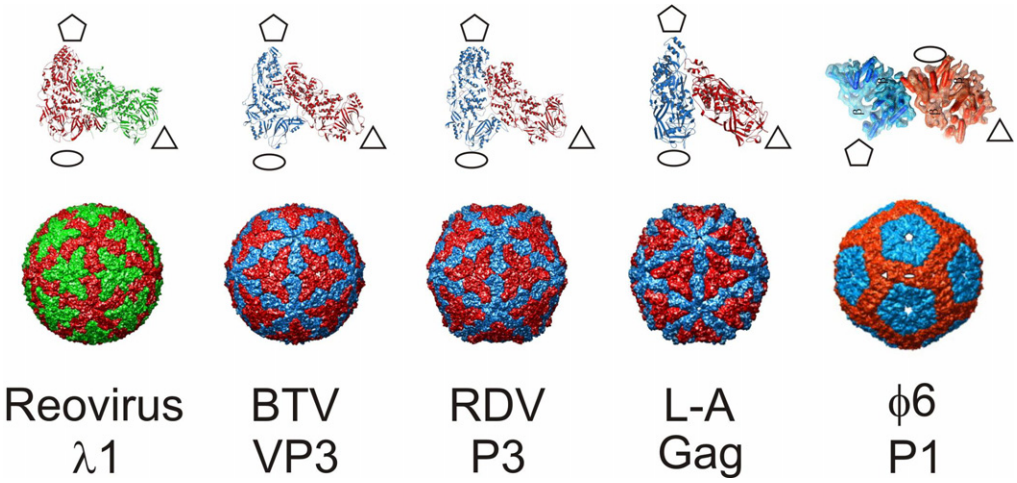


Figure 7. A gallery of dsRNA virus cores. Asymmetric unit (top row) and core *T*=1 shell (bottom row). The structures are shown for reovirus (PDB code 1ej6; 3.6Å; (Reinisch et al., 2000)), blue tongue virus (PDB code 2btv; 3.5Å; (Grimes et al., 1998)), rice dwarf virus (PDB code 1uf2; 3.5Å; (Nakagawa et al., 2003)), L-A virus (PDB code 1m1c; 3.5Å; (Naitow et al., 2002)) and bacteriophage ϕ6 (EMDB code 1206; 7.5Å; (Huiskonen et al., 2006a)). The subunits of the asymmetric dimers are shown with different colors. Illustrations A – D from ViperDB, http://viperdb.scripps.edu.

3. Symmetry mismatches in icosahedral viruses

Symmetry is a key principle in the building of viral structures. However, a highly symmetrical structure may be very static, whereas biologically interesting mechanisms require possibility for movement, a metastable configuration. For this reason, structural features related to these mechanisms often present themselves as symmetry mismatches in the overall virus structure. Figure 8 shows schematically several different ways in which the underlying icosahedral symmetry can be broken. Firstly, a feature on the virus may be unique. Secondly, the feature may only have partial occupancy (unique vertex is a special case). Thirdly, the local symmetry of the feature may differ from the global symmetry of the virus. And lastly, the feature may be flexible. Examples of each are given below. In practice, any combination of these effects may occur concurrently.

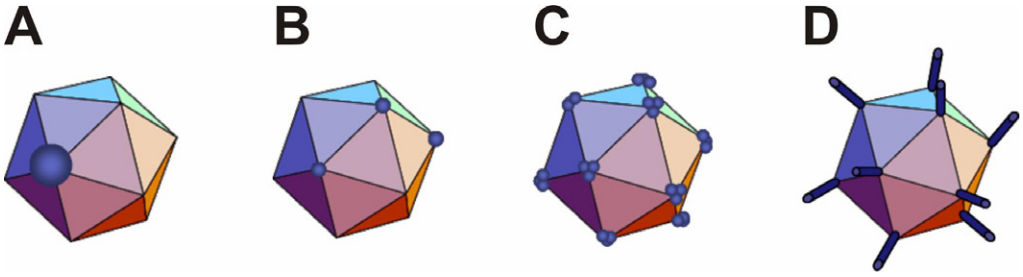


Figure 8. Types of symmetry mismatches. A. a unique structure at a unique vertex; B. partial occupancy; C. symmetry of the appendage versus symmetry of the capsid; D. a flexible structure.

3.1. Scaffolding proteins

Many viral coat proteins are able to self-assemble into complete capsids. There are, however, also many viruses whose assembly relies on the use of helper proteins, usually called scaffolding proteins. The role of the scaffolding proteins is transient: once the capsid is complete, the scaffolding exits the shell or is proteolytically degraded to make room for the genome. The capsid precursor which is void of the genome is usually called a procapsid.

A well known example of the use of scaffolding proteins is in the *Salmonella typhimurium* phage P22. The P22 capsid protein gp5 assembles into a T=7 lattice with the help of the scaffolding protein gp8. In the absence of gp8, various malformed structures are formed, such as smaller than normal T=4 capsids and spiral-like structures (Earnshaw and King, 1978; Thuman-Commike et al., 1998). gp8 is a dimer and binds two gp5 hexamers in the mature structure (unlike HK97, the coat protein of P22, gp5, does not form hexamers or pentamers alone, only within the context of capsid shell) (Morais et al., 2003; Parker et al., 1998; Parker et al., 1997). In the assembly process, both gp5 and gp8 are added to the edge of the growing capsid (Prevelige et al., 1988). After correct assembly, the procapsid contains an internal core composed of 150 to 300 copies of the scaffolding protein (Botstein et al., 1973; Casjens and King, 1974; King et al., 1973; Prevelige and King, 1993). In the procapsid, the hexamers of gp5 are in a slightly skewed

conformation, with an opening in the middle of the hexamer (Prasad et al., 1993; Thuman-Commike et al., 1999). Upon packaging, the capsid expands and the hexamers assume a more six-fold symmetric conformation, and the opening is closed (Prasad et al., 1993; Zhang et al., 2000). It is assumed that this opening is the exit path for the scaffolding (Prasad et al., 1993). The scaffolding proteins exit the procapsid intact to be reused in the assembly of further viruses (Casjens and King, 1974).

Here we are simply interested in the structure the scaffolding assumes, whether it is icosahedral or not. Thuman-Commike et al. (1996) used a temperature sensitive mutant of gp8 that was not efficiently released from the capsid. This allowed the capture of the scaffolded state of the procapsid. By a cryoEM reconstruction analysis they were able to show that the procapsid core is not likely to be icosahedrally ordered (appendage *vs* capsid symmetry mismatch, Figure 8C). In a subsequent study (Thuman-Commike et al., 1999) it was shown that gp8 seems to bind only to some of the gp5 subunits of the T=7 lattice (partial occupancy, Figure 8B).

3.2. Phage tails

Viruses encapsidate their genome either by co-assembling the capsid around the genome (*e.g.* helical viruses), or by building first an empty procapsid which is then subsequently packaged. Specialized proteins have evolved for the demanding task of packaging the genome to a high density. In tailed bacteriophages, the packaging takes place at the packaging verted located at one of the five-fold vertices of the capsid (unique structure, Figure 8A), onto which a tail is assembled after packaging. It has been known for a long time that the symmetry of the tail complex differs from that of the vertex (Hendrix, 1978). Bacteriophages P22, φ29, HK97 and λ all have this kind of symmetry mismatch (Figure 8C). The P22 tail is described below as an example.

The P22 tail machine consists of five proteins: the portal protein gp1 that is needed for DNA packaging, the tailspike protein gp9 that is needed for receptor binding and destruction, and proteins gp4, gp10 and gp26 that complete the head assembly after DNA packaging preventing the DNA from leaking out of the capsid (Tang et al., 2005). An atomic structure is available only for gp9 (Steinbacher et al., 1997), but the isolated tail complex has been reconstructed by cryoEM to 8-Å resolution (Tang et al., 2005). In combination with biochemical and sequence data, the reconstruction made it possible to localize the constituent proteins in the structure. The complex has a large central tube. The top portion of this tube is attached to the capsid, and is assigned to gp1. The symmetry in this region is 12-fold. The bottom portion of the tube consists of gp4 and gp10. gp4 binds to gp1, and gp10 binds to gp4 (Strauss and King, 1984). The masses of the proteins compared with their putative density indicate that gp4 is present as a 12mer and gp10 as a hexamer. The reconstruction also shows a clear six-fold symmetry in the region proposed as the location of gp10. The tailspike protein gp9 is known to form trimers (Steinbacher et al., 1997). The trimers are present in six copies, supposedly attached at the interface of gp4 and gp10 (Tang et al., 2005). An asymmetric reconstruction of the whole virus has also been made (Chang et al., 2006), giving a unified view of the existing structural, biochemical and sequence data, and suggesting mechanisms of assembly and DNA injection.

3.3. Viral genomes

Central sections of icosahedral reconstructions of genome-containing virus particles often show some order in the DNA or RNA. This is seen as well defined rings of density underneath the capsid or underneath the membrane bilayer, in the case of viruses like PM2 and Bam35 with an internal membrane enveloping the genome (Huiskonen et al., 2004; Laurinmaki et al., 2005). Usually, however, the detail is lost at radii closer to the center of the virus, indicating that the proximity of the capsid can induce roughly icosahedral order in the packing of the genome, especially in the outer layers, but that the symmetry is not preserved throughout the genome.

As a symmetry mismatch, the viral genome is an example of a unique substructure within the otherwise symmetrical virus (Figure 8A). The situation is even more complicated in the case of viruses with a segmented genome, as the segments may possibly constitute individual structural entities, for example separate coils, within the packaged genome. There is, however, one example where high-resolution information is available from exactly such a case. Gouet et al. (1999) were able to resolve the outer genome layers of blue tongue virus (BTV) to 6.5 Å resolution using x-ray crystallography. BTV has ten genome segments. Based on the structure, the authors propose a packaging model where the segments are packaged through different vertices into cone-like spools that meet in the center of the virus. It should be noted, however, that individual coils of RNA should be resolved at 6.5Å resolution (based on dimensions of dsRNA). Since this is not the case here, the structure is effectively at lower resolution due to imposed icosahedral averaging.

One example where already the micrographs gave a good indication of the packaging of the genome is bacteriophage T7 (Cerritelli et al., 1997). The authors noticed that a tail-deletion mutant of T7 had a preferred orientation in the micrographs: the majority of the views were along the five-fold axis of symmetry where the connector protein, *i.e.* the unique packaging vertex of the virus is located. In these views, clear concentric rings of DNA were seen, suggesting a coaxial spooling of the genome. To study this assumption, the authors made a computer simulation of DNA packed in such manner, made projections of the simulated density and compared these with the views found in the data. The results were in good agreement with various different views within the data set.

3.4. Receptor binding proteins: adenovirus

Human adenoviruses are non-enveloped dsDNA viruses that cause respiratory, ocular or enteric diseases (Dawson and Darrell, 1963; Hierholzer and Pumarola, 1976; Johansson et al., 1980). The main structural components of adenovirus capsids are the hexon (polypeptide II), the penton base (polypeptide III) and the fibre (polypeptide IV). The adenovirus capsid proteins are organized on a pseudo T=25 lattice with the pentons located at five-fold vertices and hexons at the other positions. The fibre has three distinct domains (Cusack, 2005): 1. the N-terminal tail that is attached to the penton; 2. the central shaft; and 3. the C-terminal knob that is responsible for receptor binding.

The fibre component is at odds with the global icosahedral order of the virus in two different ways. Firstly, the fibre protein is trimeric (van Raaij et al., 1999a; van Raaij et al., 1999b), leading to a mismatch relative to the five-fold

symmetric penton (Zubieta et al., 2005). Secondly, the shaft domain is kinked (Figure 8D) and for this reason lost in the icosahedral EM reconstructions save for a little stub of the capsid-proximal N-terminal domain (Stewart et al., 1991).

4. dsRNA viruses

Long double-stranded RNA is uncommon in both prokaryotic and eukaryotic cells. If long dsRNA is found in the cell, it is often a sign of a dsRNA virus infection. There are cellular mechanisms that can change the behaviour of the cell in case an infection is detected, for example by triggering programmed cell death (Goldbach et al., 2003; Jacobs and Langland, 1996) or by activating RNA silencing (Gitlin and Andino, 2003). To avoid detection, the genome of a dsRNA virus is always shielded inside a protein shell. Also replication and transcription processes take place within the shell, and therefore this innermost protein shell must also include the necessary enzymes for these activities. Another reason why the viruses must have their own dsRNA dependent polymerases, is that the cellular polymerases cannot transcribe dsRNA. Because of the multitude of functionality it incorporates, the inner core of the virus is often termed the polymerase complex (PC) or the transcription complex. The inner cores of dsRNA viruses are structurally remarkably similar. They all share a similar organization of 60 asymmetric dimers (Section 2.4).

dsRNA viruses infect a wide range of organisms, including vertebrates, invertebrates, plants, fungi and bacteria. The infection mechanisms that are needed for such a wide range of hosts differ greatly. In the virus structures this is reflected as variation in the protein layers that surround the PC, and in the case of dsRNA bacteriophages also the membrane envelope. The examples below are from the *Reoviridae* family, except for L-A virus which belongs to the family *Totiviridae,* and the only known dsRNA bacteriophages, the family *Cystoviridae.* As cystoviruses are the subject matter of Articles I and II, their lifecycle is also explained in greater detail. For the others, only the overall structures are summarized. The structures are shown schematically in Figure 9.

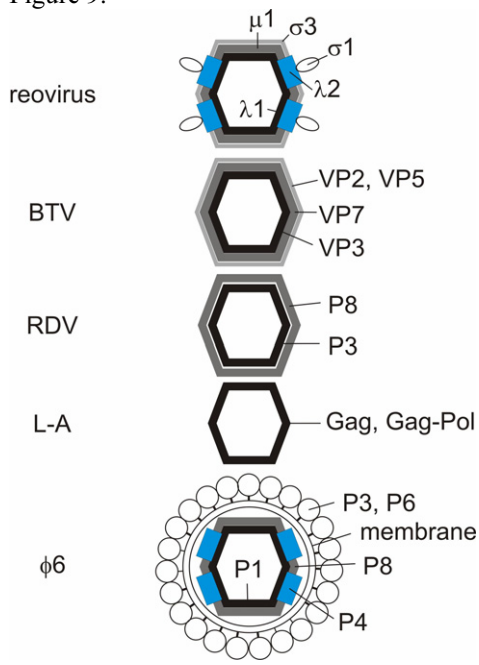


Figure 9. Comparison of structural layers in dsRNA viruses.

4.1. Reovirus

Mammalian reovirus is the type organism of the genus *Orthoreovirus* in the family *Reoviridae.* The reovirus genome consists of ten segments, each of

which contains a single gene. The segments are grouped according to their size: L1, L2, L3, M1, M2, M3, S1, S2, S3 and S4. The corresponding protein products are named λ3, λ2, λ1, μ2, μ1, μNS, σ1, σ2, σNS and σ3.

Three protein layers enclose the genome. The outermost layer is made of σ3, the middle layer of μ1, and the innermost layer of λ1. The crystal structure of σ3 in complex with μ1 has been solved, and it shows that the proteins form heterohexamers, with the 3 monomers of σ3 protruding from a trimer of μ1 (Liemann et al., 2002). Electron cryo-microscopy studies have shown that μ1 forms an incomplete T=13 lattice (Dryden et al., 1993; Metcalf et al., 1991). The lattice is broken around the five-fold axes of symmetry where pentamers of λ2, the core turret protein, are located instead. The receptor binding protein spike σ1 is also located at the five-fold positions. σ1 is shaped like a lollipop, with a 40 nm long tail and a globular head (Fraser et al., 1990). The innermost core of the virus, for which the crystal structure also is known (Reinisch et al., 2000), consists of λ1 that forms the T=1 capsid, the turret protein λ2, the RNA-dependent RNA polymerase λ3, μ2 that is needed in the synthesis of dsRNA from ssRNA (Coombs, 1996) and σ2 that binds dsRNA. The structure of the polymerase λ3 has been solved and it is known to resemble the ϕ6 polymerase P2 (Tao et al., 2002).

4.2. Blue tongue virus

BTV belongs to the orbivirus genus of the *Reoviridae* family. It infects cattle and sheep, causing high fever, excessive salivation and swelling of the face in the infected animal. In some cases, the swelling of the lips and tongue gives the tongue of the animal the name-sake blue appearance. BTV is not contagius, but is spread via an insect vector (Mertens and Diprose, 2004).

BTV is approximately 800Å in diameter. It has three concentric protein layers encapsidating its genome of 10 dsRNA segments. Each of the ten segments codes for a single viral protein. The relatively loosely bound outermost protein layer consists of proteins VP2 and VP5 (Hewat et al., 1992; Hewat et al., 1994; Nason et al., 2004). The particle that remains after the removal of the outermost shell is called the core. The outer layer of the core consists of VP7. Its structure has been solved by x-ray crystallography and shown by electron cryo-microscopy to form a *T*=13 lattice of 720 copies (Grimes et al., 1997). The crystal structure of the inner layer of the core is also known (Grimes et al., 1998), and consists of 60 dimers of VP3 arranged on a *T*=1 lattice. The genome is highly organized, as the high packaging density renders it liquid crystalline (Gouet et al., 1999) (see Section 3.3.). The packing density is, however, much lower than that of dsDNA bacteriophages. This is because the dsRNA viruses must be able to replicate and transcribe the genome inside the capsid, and these operations require some space to move the genome.

4.3. Rice dwarf virus

Rice dwarf virus (RDV) belongs to the genus *Phytoreovirus* in the family *Reoviridae*. It is transmitted to its plant host such as rice, wheat and barley, by insect vectors, the most important of which are leafhoppers. The virus multiplies in the insect. RDV infection stunts the growth of the plant host (rice dwarf disease), leading

to great economic damage (Nakagawa et al., 2003).

The RDV genome consists of twelve segments. It is encapsidated by two protein shells. The innermost core shell is made of P3, in the familiar fashion with 60 asymmetric dimers (Nakagawa et al., 2003). The viral core also contains the proteins P1, a putative RNA polymerase (Suzuki et al., 1992); P5, a putative guanyltransferase (Suzuki et al., 1996); and P7, a non-specific nucleic acid binding protein (Ueda et al., 1997). The outer protein layer consists of P8 and P2. P8 is organized in T=13*l* lattice (Lu et al., 1998) (Nakagawa et al., 2003), and P2 is a minor protein that is needed for virus infection (Yan et al., 1996).

4.4. L-A virus

L-A virus infects the yeast *Saccharomyces cerevicae*. L-A virus is the simplest of the dsRNA viruses. It has only one genome segment that codes for two proteins, the capsid protein Gag and the RNA-dependent RNA polymerase Pol (Fauquet et al., 2005). Pol is expressed as a Gag-Pol fusion protein. The crystal structure at 3.4-Å resolution shows that the virus particle consists of 60 asymmetric dimers of Gag, and two copies of Gag-Pol (Naitow et al., 2002).

4.5. Cystoviruses

The *Cystoviridae* are the only known dsRNA bacteriophages. The type species of the family is ϕ6. In total 9 cystoviruses have been isolated (ϕ6, ϕ7, ϕ8, ϕ9, ϕ10, ϕ11, ϕ12, ϕ13 and ϕ14) (Mindich et al., 1999; Vidaver et al., 1973). ϕ6 infects *Pseudomonads syringae* pv. phaseolicola, a plant pathogen causing halo blight in beans (Pitman et al., 2005), but other members of the family have been found to infect also other gram-negative hosts such as *Eschericia coli* and *Salmonella typhimurium* (Mindich et al., 1999).

4.5.1. ϕ6 structure

The cystoviral genomes have three segments, S, M and L (Mindich et al., 1999). The L segment codes for the proteins P1, P2, P4, P7 and P14, the first four of which constitute the polymerase complex (Gottlieb et al., 1990; Poranen and Tuma, 2004). The structure and composition of ϕ6 is known from cryo-EM studies of subviral and recombinant particles (Butcher et al., 1997; de Haas et al., 1999; Huiskonen et al., 2006a). P1 is the major coat protein of the PC. It is arranged in a T=1 dodecahedral lattice, where the asymmetric unit is a dimer. The P1 monomers constituting the dimer were segmented from a high-resolution cryoEM reconstruction of the nucleocapsid (Huiskonen et al., 2006a) and fitted to a cryoEM reconstruction of the unpackaged PC (de Haas et al., 1999), leading to a model where rigid body rotations of the monomers explain the capsid conformation change from the unexpanded to the expanded form (Huiskonen et al., 2006a). P4 is the packaging enzyme (Gottlieb et al., 1992). It is located on the five-fold vertices of the P1 shell (de Haas et al., 1999). P2 is the polymerase (Makeyev and Bamford, 2000) and P7 an assembly cofactor that is also involved in packaging (Juuti and Bamford, 1997; Poranen et al., 2001). P2 monomers reside beneath the five-fold vertices, and P7 is located at a radius of 160Å from the capsid center (Ikonen et al., 2003). The PC is surrounded by a T=13 layer of P8. P8 is a

highly α-helical protein and it has two distinct domains: a flat core domain and a peripheral domain consisting of a four-helix bundle that makes i) intertrimer connections between P8 trimers and ii) connections to the P4 hexamer at the five-fold vertices (Huiskonen et al., 2006a). Cystoviruses are enveloped by a membrane bilayer. In ϕ6, spikes made of P3 protrude from the membrane where they are anchored by the fusion active protein P6 (Bamford et al., 1987; Stitt and Mindich, 1983; van Etten et al., 1976).

The atomic structure of the polymerase P2 has been solved at 2.0-Å resolution (Butcher et al., 2001). The polymerase has the canonical hand-like organization with domains corresponding to the palm, thumb and fingers. The structure has a high degree of similarity to hepatitis C virus (HCV) (Ago et al., 1999; Bressanelli et al., 1999; Lesburg et al., 1999). In fact, at the time when the structure was solved, the two polymerases were closer to each other than to any other known polymerase (Butcher et al., 2001). The ϕ6 P2 polymerase has also been co-crystallized with both ssDNA and ssRNA templates, and activated with GTP within the crystals (Butcher et al., 2001; Salgado et al., 2004). These structures suggested a model for the initiation of the replication process, where the template is inserted in a tunnel that leads to the active site, and binds so that the base of the nucleotide is placed in a binding pocket in the C-terminal domain of P2 (Butcher et al., 2001).

A 2D average of negative-stained EM images (Juuti et al., 1998), and a cryoEM reconstruction (de Haas et al., 1999), of the isolated ϕ6 packaging enzyme P4 has shown it to be a hexamer. In the case of ϕ8, preliminary cryoEM work, a 2D average of isolated ϕ8 P4, indicated that it is hexameric (Kainov et al., 2003b). Crystal structures are only available for the ϕ12 P4, alone and in complexes with adenosine diphosphate (ADP) and an adenosine triphosphate (ATP) analog (+-Mg/Mn). The complexes correspond to key points in the catalytic pathway (Mancini et al., 2004). The structures show that P4 is a Rec-A like ATPase with a central channel through which the RNA can be translocated. A comparison of the structures showed conformational changes related to ATP hydrolysis.

4.5.2. ϕ6 lifecycle

ϕ6 infection starts when P3 attaches to the type IV pili of the host. The pilus retracts and the virus is brought into contact with the outer membrane of the host (Bamford et al., 1976; Romantschuk and Bamford, 1985). P6 causes fusion of the viral envelope with the host outer membrane, leading to the release of the NC into the host cell periplasm (Bamford et al., 1987). The NC P8 layer is responsible for the penetration through the host cell plasma membrane, followed by release of the PC into the host cytoplasm (Romantschuk et al., 1988). The PC starts transcription of the genome (Coplin et al., 1975) and the positive sense mRNA are possibly released through passive channels of P4 at the PC vertices; for ϕ12 this mechanism has been verified (Kainov et al., 2004). The mRNAs are translated by the host polymerase into proteins which assemble into new empty PC particles (Emori et al., 1982). The assembly pathway is shown in Figure 10. P4 of the empty PCs packages the positive sense ssRNA, starting from the s segment and followed by the m and l segments. The PC capsid has affinity for the ssRNA segments. It has been suggested that the affinity for the different segments is related to the conformational state of the capsid, which in turn depends on the amount of ssRNA packaged at each stage (Mindich, 2004). When all segments have

been successfully packaged, the polymerase P2 replicates the complementary –ssRNA strand (Frilander et al., 1992). The P8 T=13 layer assembles on top of the packaged PCs to create NCs (Olkkonen et al., 1991). The NCs are next enveloped by a lipid membrane, with the lipids derived from the host cytoplasmic membrane (Laurinavicius et al., 2004). The viruses are released by cell lysis dependent on the lytic protein P5 and membrane protein P10 (Johnson and Mindich, 1994).

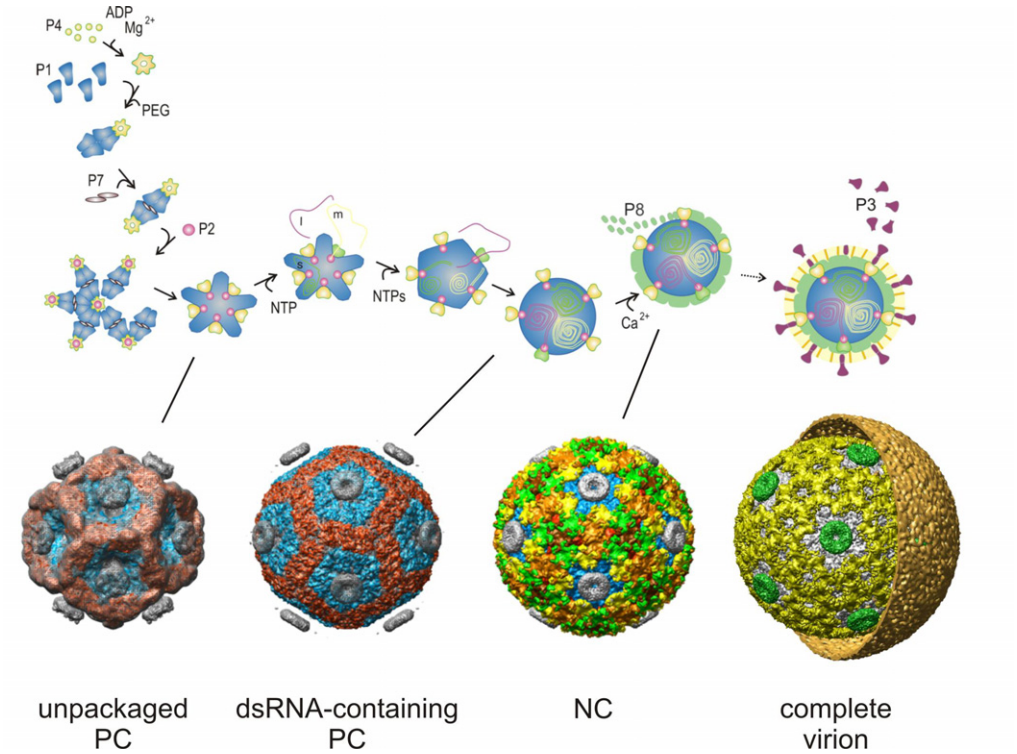


Figure 10. Cystovirus assembly pathway with cryoEM reconstructions of assembly intermediates. ϕ6 assembly is nucleated by hexamers of P4 interacting with P1 monomers. Together with P2 and P7 these form the polymerase complex which is the viral procapsid. This particle recognizes ssRNA segments and packages them.The polymerase P2 replicates the RNA inside the PC to form the dsRNA containing PC. Sometime during these processes the PC expands. Then P8 assembles onto the PC to form the NC. Finally the NC acquires the membrane and the virus induces host cell lysis releasing the virions. From left to right: unpackaged PC (de Haas et al., 1999), packaged core (Huiskonen et al., 2006a), NC (Huiskonen et al., 2006a) and complete virion (Article I).

4.5.3. ϕ8 structure

Prior to Articles I and II, no information about the 3D structure of ϕ8 has been available. Cores and virions had been imaged by cryoEM (Yang et al., 2003).

4.5.4. ϕ8 lifecycle

The lifecycle of ϕ8 largely follows that of ϕ6, but with some significant differences. The binding mechanisms of the viruses are different: ϕ8 binds directly to the lipopolysaccharide of the host cell outer membrane (Hoogstraten et al., 2000). ϕ8 P8 is lost with the membrane in a detergent treatment with Triton X-100 (Hoogstraten et al., 2000), suggesting that it is a membrane-associated protein, not a nucleocapsid outer shell protein like P8 of ϕ6 (Figure 9). The polymerase complex of ϕ8 can infect spheroplasts; in ϕ6 also the nucleocapsid protein P8 is needed to accomplish this. The assembly pathways of the viruses are also different. In particular, ϕ6 PC assembly is dependent on protein P4 (Poranen et al., 2001), whereas in ϕ8 both P2 and P4 are needed (Kainov et al., 2003a).

5. Archaeal viruses

Archaea are unicellular organisms that inhabit diverse and extreme habitats. For a long time, they were considered to belong to bacteria and were called archaebacteria. This changed with the advent of molecular sequence analysis. Comparison of ribosomal RNA from bacteria and the archaebacteria showed that archaebacteria merit their own domain alongside bacteria and eukaryotes (Woese and Fox, 1977). Since then, they have been called the archaea. The domain of the Archaea is further divided into two main kingdoms: Crenarchaeota and Euryarchaeota (Woese et al., 1990). The kingdoms correspond to the habitats of their members. Members of the Crenarchaeota are thermophilic or hyperthermophilic (optimal growth conditions at temperatures above 40° C and above 80° C, respectively), whereas the members of the Euryarchaeota are methanogenic (methane-producing) or halophilic (requiring more than 1.5 M NaCl).

In addition to the difference in the ribosomal RNA, archaea differ from bacteria also in their membrane and cell-wall structures. The lipids in archaeal membranes are ether-linked instead of ester-linked like those of bacteria and eukaryotes (Brock, 1997). In addition, archaeal cell walls do not contain a peptidoglycan layer unlike bacterial cell walls. Some methanogenic archaea have a pseudopeptidoglycan cell wall which is similar to the bacterial one. In the other archaea the cell wall consists of polysaccharide, glycoprotein and protein (Brock, 1997).

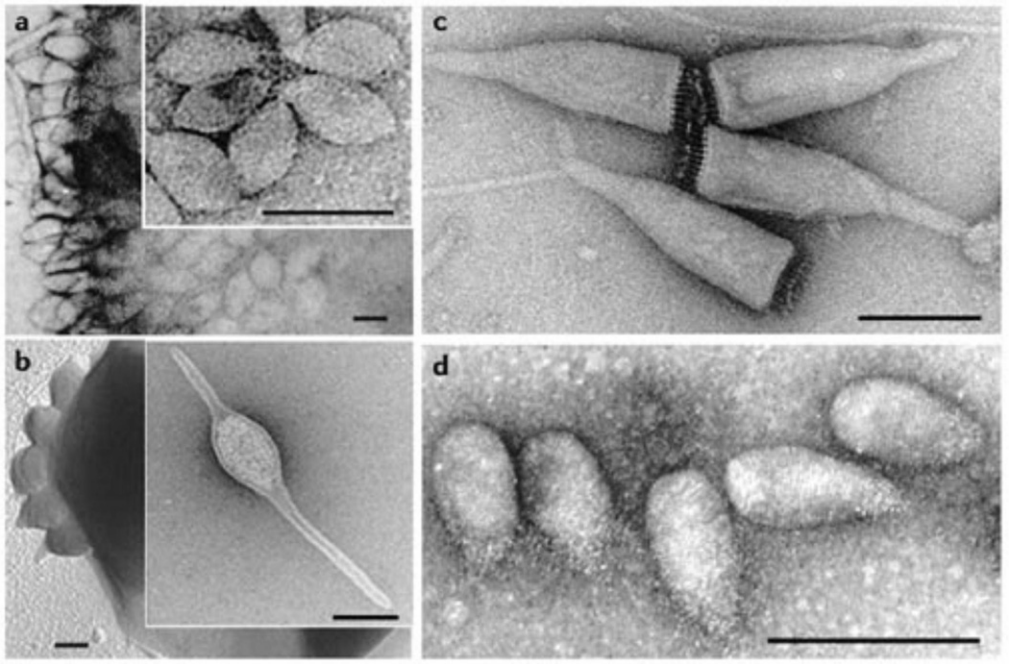


Figure 11. Viruses of hyperthermophilic archaea.

a | Sulfolobus spindle-shaped virus 1 (SSV1) (inset) and its extrusion from the host cell. b | The extracellularly developed Acidianus two tailed virus (ATV) (inset) and its extrusion from the host cell. c | Acidianus bottle-shaped virus (ABV). d | Sulfolobus neozealandicus droplet-shaped virus (SNDV). All images are negatively stained with uranyl acetate, except for part b, which was platinum-shadowed. Scale bars represent 100 nm. Parts a and d are courtesy of W. Zillig. Part b is reproduced from (Haring et al., 2005b) © (2005) Macmillan Publishers Ltd. Part c is reproduced with permission from (Haring et al., 2005a) © (2005) American Society for Microbiology. The complete figure reprinted by permission from Macmillan Publishers Ltd: Nature (Prangishvili et al., 2006a) © (2006).

5.1. Viruses of the Crenarchaeota

The previous sections on viral proteins and symmetry may have given the impression that the world of virus structures is reasonably ordered. However, a glance at the viruses of the hyperthermophilic crenarchaea will quickly convince the reader otherwise. Figure 11 shows some of the diversity of shapes. Some of the more interesting forms worth mentioning are *Sulfolobus islandicus* filamentous virus (Arnold et al., 2000b) and *Acidianus* filamentous virus 1 (Bettstetter et al., 2003), two filamentous viruses with complex structures at the ends of the filaments; the *Sulfolobus neozealandicus* droplet-shaped virus (Arnold et al., 2000a); the *Acidianus* Bottle-shaped Virus (Haring et al., 2005a); and the *Acidianus* two-tailed virus (Prangishvili et al., 2006b). The remarkable thing about the latter is that its lifecycle contains an extra-cellular phase: it grows its two tails outside the host.

5.1.1. STIV

The *Sulfolobus* Turreted Icosahedral Virus (STIV) is so far the best structurally characterized archaeal virus. As the name suggests, STIV is icosahedrally symmetric, thus rather plain in comparison to the many fancy shapes found infecting crenarchaea. STIV is a thermophilic dsDNA virus infecting *Sulfolobus solfataricus*. It was isolated in an acidic hot spring (pH 2.9 – 3.9, 72 – 92° C) in Yellow Stone National Park (Rice et al., 2001; Rice et al., 2004). Its genome has 17663 base pairs and 36 predicted open reading frames, with no known homologous proteins (Rice et al., 2004). In a more detailed characterization (Maaty et al., 2006), nine proteins have been identified by mass spectrometry. For five of these, structural prediction found possible structures or functions. Two of these, C381 and A223 were predicted to correspond to the PRD1 spike protein P5. B345 is the major capsid protein (Maaty et al., 2006; Rice et al., 2004). STIV contains a glycosyltransferase (Larson et al., 2006), and in fact, B345 is glycosylated, which has been show to increase the thermal stability of some proteins (Wang et al., 1996). STIV also contains a lipid membrane (Maaty et al., 2006). The membrane is predicted to reside under the icosahedral coat like in PRD1. There is no evidence for an icosahedrally ordered nucleocapsid (Rice et al., 2004).

The structure of STIV was first determined by cryoEM (Rice et al., 2004), and subsequently the crystal structure of B345 (Figure 5D) was solved (Khayat et al., 2005). B345 is of the double barrel type (Section 2.2), and it forms a T=31 lattice. The most striking feature of the virus is the presence of the large namesake turrets at the five-fold vertices. They are possibly required for cell entry and DNA translocation (Khayat et al., 2005; Rice et al., 2004), and may consist of proteins C381 and A223 (Maaty et al., 2006).

5.2. Viruses of the Euryarchaeota

Most information we have about viruses infecting euryarchaeota is from viruses of halophilic hosts. In a review by Reiter et al. (1988) only one virus was mentioned that infects a methanogenic host *Methanobrevibacter smithii*. In a more recent review of the haloarchaeal viruses (Dyall-Smith et al., 2003), only fourteen viruses were listed, which indicates that progress of mapping the haloarchaeal viruses has been fairly slow, given that the number of genera of the host halobacteria is about 15, with many species known within each genus (Dyall-Smith et al., 2003). ϕH (Schnabel et al., 1982) infecting *Halobacterium salinarum* is still the best known halovirus, even though it has not been actively studied recently. ϕH and most of the other known haloviruses have morphologies similar to tailed bacteriophages. Exceptions are the spindle-shaped His1, the pleiomorphic His2 and the spherical SH1, which all infect *Haloarcula hispanica* (Dyall-Smith et al., 2003).

5.2.1. SH1

SH1 was isolated from a hypersaline lake in Australia (Porter et al., 2004). It has a dsDNA genome of 31 kilobasepairs that codes for at least 11 structural proteins, and possibly three more proteins (Bamford et al., 2005b). The

analysis of the protein composition indicated proteins VP3, VP4 and VP7 as putative capsid proteins. VP4 and VP7 were also shown to make stable complexes under non-reducing conditions (Bamford et al., 2005b).

SH1 has been shown to contain lipids. The lipid composition was studied by thin-layer chromatography and electrospray ionization mass spectrometry (Bamford et al., 2005b) The virus membrane contains 81.7% phosphatidylglycerophosphate methyl ester (PGP-Me), 16.5% archaeal phosphatidylglycerol (PG) and 1.4% phosphatidylglycerosulfate (PGS). These are all present in the host as well, but with a different distribution: *H. hispanica* lipids are 13.6% PG, 56.9% PGP-Me and 24.7% PGS (Bamford et al., 2005b). The virus can be dissociated by lowering the salt concentration or by treating it with urea. When the capsid-associated proteins VP1, VP2, VP3, VP4, VP6, VP7 and VP9 are removed, they leave behind a lipid- and DNA-containing particle, a lipid core, indicating that SH1 has an internal lipid membrane around the genome {Kivelä, 2006 #33}.

The structure of SH1 is the subject matter of Article III of this thesis.

6. Evolution of viruses and viruses in evolution

When genetic information is available, phylogenetic analysis of either nucleotide or amino acid sequences is often used to chart the evolutionary relationships between organisms. Because of the redundancy of the genetic code, the amino acid sequences are preserved longer than the nucleotide sequences. This means that it is easier to detect an existing evolutionary relationship from the amino acid sequence, when the signal has already weakened in the nucleotide sequence. The purpose of the nucleotide sequence is to store the information for making the amino acid sequence, and the redundancy of the genetic code helps in doing this successfully. Similarly, the amino acid sequence codes for the 3D structure of the protein. There is redundancy also in this step: a change in the sequence does not necessarily change the fold. And when a change in the fold does occur, there is a high probability that the result is not viable. Evolution of single proteins is constrained by the task they need to accomplish: if they fail in the task, there will be no progeny. For example, a viral capsid protein must always be able to assemble into a complete capsid. Failure to do so leads to exclusion from the gene pool. This means that structural information in the fold of proteins is conserved over much longer timescales than sequence information, and that more distant relationships can be detected by comparing three dimensional structures. Thus, the observation of the four groups of viral capsid protein listed in Section 2 has led to the hypothesis that these groups actually correspond to lineages of viruses that have developed from different ancestors. The idea has been around since (Rossmann et al., 1985) reported that the single β-barrel fold is found both in human and plant viruses, but recently it has been revitalized as more evidence supporting it has surfaced. In its current form, the hypothesis was expounded for viruses with a double-barrel trimer coat protein (Bamford et al., 2002; Benson et al., 1999; Benson et al., 2004), but it has been updated to include also the HK97-type fold (Baker et al., 2005) and the α-helical T=1 fold of asymmetric dimers (Bamford et al., 2005a). The notion of the four (at least) viral lineages has quickly gained acceptance among virologists, which means a dramatic shift of perspective from not too long ago, when the observed similarities were seen as exceptions. Now the fact that the double barrel trimer capsid design is found in all domains of life

(Khayat et al., 2005) points to the possible conclusion that the ancestor of these viruses was already infecting organisms prior to the separation of the three domains.

The concept of viral lineages based on the structure of the viruses is a novel way to group together viruses that were not previously thought to be related. Still, it does not give any possible explanation as to where the first viruses, the ancestors of the lineages, came from. As summarized recently (Forterre, 2006a), three hypotheses have been proposed for the origin of viruses: 1. viruses are relics from pre-cellular life-forms; 2. viruses have developed by reduction from cellular organisms (cell-gone-bad); 3. viruses descend from plasmids or other mobile genetic elements, that have escaped from the control of the cell (plasmid-gone-bad). None of these seems to be satisfactory on a closer inspection. Hypothesis 1 is usually readily discarded, because viruses require a host for propagation. Hypothesis 2 is refuted by examples of parasites derived from cells while retaining their cellular machineries, and by the lack of observed intermediates, although the giant mimivirus might be a possible candidate example (Xiao et al., 2005). Hypothesis 3 cannot easily explain how a plasmid or other free genetic element within a cell could acquire a protein capsid.

6.1. Hypotheses about the origin of viruses in the RNA world

The argumentation about the origin of viruses has until recently been mostly in terms of the concepts of the current biosphere. If the origin of viruses goes back to earlier stages of the development of life, such arguments are bound to be insufficient. I personally find the more hypothetical work interesting, so I try to present some of it here.

It has been proposed that life first began as the so called RNA world, where RNA molecules were both the carriers of genomic information and the active enzymes (Weiner and Maizels, 1987). RNA is known to still have an enzymatic role for example in ribosomes (Brock, 1997) and spliceosomes (Watson et al., 2004). Two recently proposed theories of virus evolution, the "Three Domains, Three Viruses" theory (Forterre, 2005; Forterre, 2006b) and the "Virus World" theory (Koonin et al., 2006) both evoke the concept of the RNA world, and most interestingly, the role of viruses in co-evolution with cellular organisms. Both theories explain the currently seen multitude of viral genome types as remnants from different stages of development from the early RNA world to the current DNA world. The theories also explain how the three domains of life emerged.

6.1.1. Three Viruses, Three Domains

In Forterre's view (Forterre, 2005; Forterre, 2006b), the early RNA world had cellular life, because cellular confinement is necessary for the development of a complicated metabolism. He divides the development from the RNA world into the DNA world into two distinct phases. The arrival of the first replicating RNA cell marks the beginning of the *first age* of the RNA world. At this stage, RNA acts as both the genome and the catalyst, there are no proteins yet. The emergence of the ribozyme ancestor of today's ribosomes, marks the beginning of the *second* age of the RNA world. From this point onwards, proteins synthesised by the ribozymes started to take over the role of the catalyst. Forterre assumes RNA viruses to be present in both ages of the RNA world, and suggests that they may have developed via parasitic reduction from out-competed cellular lineages. Furthermore, today's

RNA viruses would descend from these early RNA viruses.

According to this hypothesis, the invention of DNA took place in the second age of the RNA world. This is because the critical reaction in the RNA/DNA transition, the reduction of a ribose to a deoxyribose, could not have been performed by an RNA molecule, but would require the presence of protein enzymes. The RNA/DNA transition would go through the intermediate stage of the U-DNA world, where DNA contains uracil instead of thymidine (Forterre, 2002; Poole et al., 2000). Again, there are viruses today with U-DNA genomes, such as phage PSB-1 (Tomita and Takahashi, 1975).

The key proposal in this hypothesis is that DNA was invented by RNA viruses. The argument is that viruses were the only ones that would directly benefit evolutionarily from the modification of their genome. For RNA viruses, switching over to DNA was a way, for example, to make their genome immune to the existing RNA degrading enzymes. DNA would prove useful for cells as well, as it allows for much more stable propagation of large genomes, but this benefit is according to this argument too indirect from an evolutionary point of view. On the other hand, modification of the genome for the purpose of escaping cellular defence mechanisms is a known practise also in today's DNA viruses (Gommers-Ampt and Borst, 1995). For example, the *Xanthomonas oryzae* phage XP12 has all of its cytosines modified to 5-methylcytosines (Kuo et al., 1968).

In the next step, cellular organisms adopted the use of DNA and the related enzymes from the viruses that infected them. It is also proposed here that this is what led to the division of cellular RNA life into the three domains of cellular DNA life that exist today. The domains correspond to the progeny of RNA cells infected by three different types of viruses, with three different enzymatic toolkits for handling DNA. This difference in the basic DNA enzymes, such as the topoisomerases, is still found between the three domains today.

6.1.2. The Virus World

(Koonin et al., 2006) conducted an extensive phylogenetic analysis of viral proteins, and categorized the proteins according to their relationships, or the lack of them, to counterparts in cellular organisms and other viruses. In particular, they picked out a group of proteins that seems to only belong to viruses, to represent the "virus state". These proteins, encoded by the so called *viral hallmark genes,* are found in many diverse groups of viruses, and have only distant homologues in cellular organisms. The hallmark genes also appear to be monophylous within their respective gene families. Examples of hallmark proteins include the β-barrel capsid protein, the superfamily 3 helicase, and packaging ATPases of the FtsK family. The authors conclude that the existence of the hallmark genes is a real phenomenon (neither artifactual nor explainable by horizontal gene transfer) that refutes the "reduction-from-cell" and "escape-from-cell" theories of virus origin. The authors then propose a model of the early RNA world that would allow viruses to predate cellular life.

In the model proposed by Koonin et al. (2006), the early evolutionary processes took place in inorganic compartments. Instead of an RNA world cell proposed by Forterre (2005), the inorganic compartment provides the necessary critical concentration of biomolecules. This model readily explains how viruses can predate cellular life: in the early stages of development, all RNAs are essentially parasitic. The distinction between parasite and host arises only when co-operating ensembles of

molecules/genes are formed in some compartments. In this model, the development from RNA world to RNA-protein world to RNA-DNA retro world to DNA world is all assumed to take place within the inorganic compartments, with all of the intermediate states leaving behind typical viruses, in a similar fashion as proposed by Forterre (2005). Escape of cells from the compartment is assumed to have happened twice, once from a pre-archaeal compartment and once from a pre-bacterial compartment, producing the ancestors of archaea and bacteria, respectively. Koonin et al. (2006) adopt the view that the eukaryotic cell emerged from the fusion of archaeal and bacterial cells (Zillig et al., 1989).

6.2. Critique of the RNA world based hypotheses

The validity of the hypotheses naturally depends on the validity of the RNA world hypothesis which also has many problems, starting from a very basic question: how can RNA emerge in prebiotic conditions? For example ribose, the sugar needed for the making RNA, may not have been available in prebiotic conditions (Shapiro, 1988). If this question is solved (for one possible explanation regarding ribose, see Bielski and Tencer (2006), there remains the question of what RNA chemistry is actually capable of. For example lipids, the material used in all cell walls, are nowadays synthesized by proteins, and therefore a pre-protein cell wall as suggested by Forterre (2005) would have to have consisted of something else, or alternatively, there has to have been a prebiotic process for producing lipids, similarly as suggested for ribose. There are many interdependencies between the various macromolecular constituents of biological life today. Rather than trying to solve the chicken and egg –dilemma of which came first, RNA or protein, it may be more fruitful to consider them as arising first independently of each other, and then together, after the development of the RNA to protein -coding (Rode, 1999).

The fundamental difficulty in these hypotheses is that there is no fossil record to support any of the claims. On the other hand, sequence based studies such as that of Koonin et al. (2006) can only use the data that are available from organisms that exist now. All extinct intermediates are missing from the picture, possibly leading to a misinterpretation of the existing data. For example, it would be difficult to explain the evolution of birds, if we didn't have the fossil records suggesting a reptilian ancestor.

7. Electron cryo-microscopy

The purpose of microscopy is to get a magnified view of small things. The measure of the limit to which a given method of microscopy is applicable is resolution, the ability to distinguish two features as separate. In the study of tissues, for example, it is useful to be able to distinguish the cellular organisation that constitutes the tissue. The required resolution range for this may be something like 10 μm. In the study of cells, it is more interesting to be able to see what kind of components are at work inside the cell. For example, to study the process of viral infection, it is enough to be able to distinguish the viral particles entering the cell or inside the cell. Here the necessary resolution range is higher, around 0.1 μm. In structural biology, the interest lies in the molecular level of the structure of the biological machinery. The corresponding relevant resolution range is yet higher, from the rough determination of overall structure at the nanometer range to the ultimate goal of atomic resolution (< 4Å).

In every microscopy method, the object under study is probed with a radiation field that is able to transfer information. In light microscopy, for example, the object is illuminated with a beam of (visible) light (UV and IR extensions are now popular), and the information about the structure is conveyed in the way the light is shadowed (scattered and absorbed) by the object. The resolution of a given method, however, ultimately depends on the wavelength of the radiation used. Visible light is electromagnetic radiation at a wavelength of 400-700 nm, and the limit resolution obtainable by conventional light microscopy is around 200 nm (Ram et al., 2006). To overcome this limitation, electromagnetic radiation of a shorter wavelength can be used. X-rays range from 0.01 to 10 nm in wavelength. This is more than enough for determining atomic structures. However, electromagnetic radiation is not the only possible field for use in microscopy. The wave-particle dichotomy of quantum mechanics states that a beam of light in its entirety acts like a wave, but in its interaction with matter it acts like separate particulate entities, photons. Similarly, a beam of particles, such as electrons, also behaves in a manner of waves. The wavelengths possible for a beam of electrons are much shorter than those of visible light, therefore allowing a much greater resolution. This constitutes the advantage of electron microscopy.

7.1. The transmission electron microscope

The design of a transmission electron microscope is shown schematically in Figure 12. The components are arranged in a column with the electron source on the top. The electron source is either a heated filament or a field emission gun (FEG) source. The advantage of the FEG is that the emitted beam of electrons is highly coherent, which is important for reaching higher resolutions (see 7.4). The wavelength of the electron beam depends on the speed of the electrons. The electron that are emitted from the electron source are not fast enough to reach short wavelengths, and must therefore be accelerated in a high tension electric field, typically in the range of 60 to 400 keV. As an example, with an acceleration voltage of 200 keV, the electron wavelength λ is 0.02527 Å (Baker et al., 1999).

As electrons interact with molecules in air, a high quality vacuum must be maintained in the column for a free passage of electrons. This is accomplished with a multistage system of pumps and valves.

Analogous to optical lenses used in light microscopy, the electron beam is controlled with magnetic field lenses. The first lens in the system is the condenser lens, the purpose of which is to condense the beam onto the specimen. The beam that has passed through the sample is collected with the objective lens and the final magnified image is made with the projector lens. As with optical lenses, also magnetic lenses have imperfections that limit the resolution obtainable by the instrument. In the case of the TEM, the limiting factor is the spherical aberration (Cs) of the objective lens.

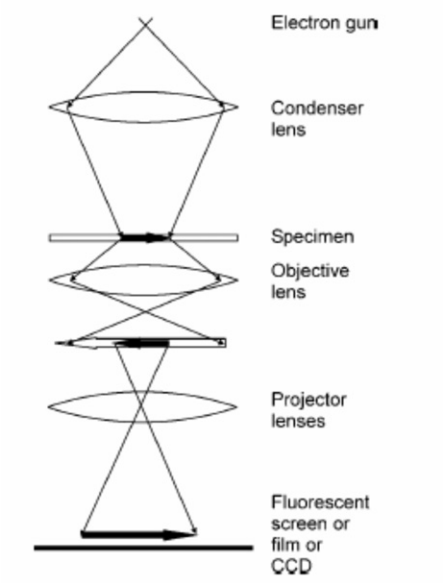


Figure 12. Schematic representation of the transmission electron microscope. Modified from (Chescoe and Goodhew, 1990) by J. Huiskonen.

For data collection the traditional approach has been to equip the microscope with a standard camera and use electron sensitive film to capture the images. The negatives are then scanned to produce digital images. Lately, however, the trend has been to replace film cameras with scintillator-CCD cameras, which produce high quality digital images without the cumbersome scanning step. This also enables highly automated data collection and processing approaches (Jiang et al., 2001b; Stagg et al., 2006).

7.2. Image formation in the TEM

The electron beam interacts with the sample in many ways, the most significant of these are elastic and inelastic scattering (Frank, 1996). Inelastic scattering occurs mostly at small angles and involves transfer of energy from the beam electrons to the sample, leading to damage of the specimen (beam damage) (Kuo and Glaeser, 1975). Inelastically scattered electrons also carry information about the sample, but as they have a continuous range of different energies, it is impossible to focus them for imaging. Instead, they contribute to the recorded image as additional noise (Frank, 1996). With an energy filter it is possible to block their passage and improve the image quality.

Elastically scattered electrons retain their kinetic energy but change their direction. The change in the direction of the electrons gives rise to two types of contrast between the sample and the background. Electrons that are deflected so much that they collide with the apertures in the TEM column never make it to the recorded image and their history is recorded as the weakened amplitude, *i.e.* amplitude contrast. The elastically scattered electrons that are deflected at a small angle reach the image, but they have travelled a longer path than the unscattered electrons. As the electron beam can also be considered a wave, this means that the scattered electrons have different phase compared to the unscattered electrons and this produces phase contrast (Frank, 1996).

The strength of elastic scattering depends on the sample. Heavy atoms give a higher amplitude contrast, whereas in unstained biological samples such as protein, which consist of light atoms, phase contrast is more significant.

7.3. The contrast transfer function

Because of the various scattering effects discussed in Section 7.2., the recorded TEM image of an object is not a straightforward projection image. Since in the further steps of processing the images we assume that the images are 2D projections of the original 3D object, we must first understand the relationship between the TEM image and the object, and correct the image to be more like a straightforward projection. The relationship between the observed image intensity $I(x,y)$ and the projected potential of the object $O(x,y)$ can be given in terms of their Fourier transforms $I(\mathbf{k}) = \mathrm{F}\{ I(x,y) \}$ and $O(\mathbf{k}) = \mathrm{F}\{ O(x,y) \}$ (where $\mathbf{k} = \{k_x, k_y\}$ is the spatial frequency) and the *contrast transfer function* (CTF) as (Frank, 1996):

$$I(\mathbf{k}) = \mathrm{CTF}(\,.\,)\, O(\mathbf{k}) \qquad (1)$$

The CTF has two terms corresponding to *phase contrast* and *amplitude contrast* (using notation of Mindell and Grigorieff (2003)):

$$\mathrm{CTF}(\chi(.)) = -w_1 \sin(\chi(.)) - w_2 \cos(\chi(.)), \qquad (2)$$

where $w_1 = \mathrm{sqrt}(1 - A^2)$ and $w_2 = A$, with A the percentage of amplitude contrast in the image. The ratio of phase contrast and amplitude contrast depends on the specimen material, however, within the spatial frequency range of practical interest it can be assumed to be constant (Toyoshima and Unwin, 1988). The effect of the lens aberrations and defocusing are modelled in the wave aberration function $\chi(.)$:

$$\begin{aligned} \chi(.) &= \chi(\lambda, \mathbf{g}, \Delta f, C_s) \\ &= \pi\lambda g^2 (\Delta f - 1/2\lambda^2 g^2 C_s) \end{aligned} \qquad (3)$$

where the scattering vector g describes the difference between the wave vectors of the scattered and unscattered electrons (Mindell and Grigorieff, 2003).

7.4. The envelope function

In the ideal case when the electron beam is completely coherent, the resolution is limited only by the aperture of the microscope. In practise, however, the source of the beam has a finite size, resulting in a finite energy spread, which in turn results in partial coherence that dampens the contrast transfer function at higher spatial frequencies. The dampening effect can be modelled with an envelope function $E(k)$, $k=|\mathbf{k}|$, with which the contrast transfer function in the partially coherent case becomes: $CTF_{pc}(.)= CTF(.)E(k)$, where $E(k) = E_i(k)\ E_c(k)$ (Frank, 1996). $E_i(k)$ describes the dampening due to partially coherent illumination, and it is a function of the electron source size. $E_c(k)$ describes the dampening due to the energy spread, and is a function of the defocus spread due to lens current fluctuations.

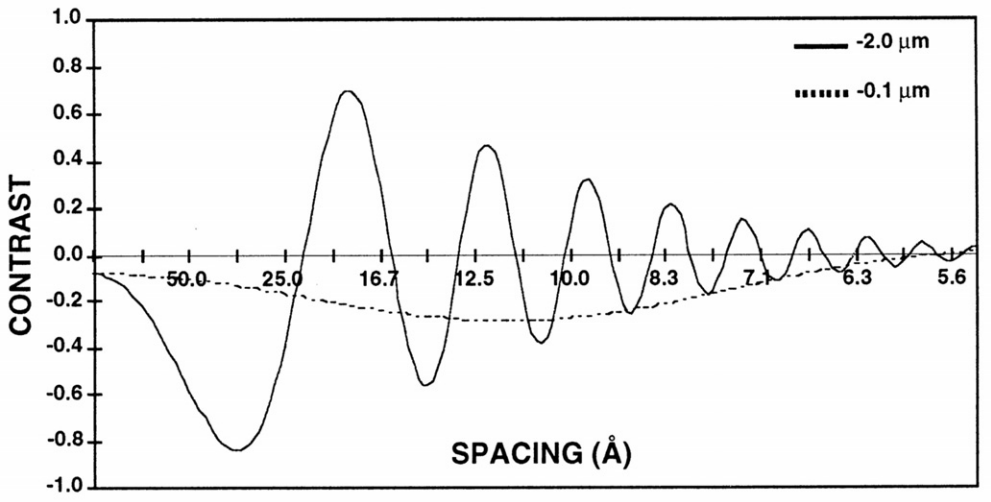


Figure 13. Contrast transfer function The CTF for a Philips CM200 FEG at 200 kV is plotted as a function of resolution in angstroms for an underfocus of 2 µm and an underfocus of 1,000 Å and a magnification of ×36,000. The decrease in the amplitude of the function with increased resolution reflects the measured attenuation due to the lack of coherence in the source, specimen movement, and other optical effects. The value of -0.1 at the origin is the amplitude contrast portion of the function. With permission from (Baker et al., 1999) © (1999) American Society for Microbiology.

7.5. Correcting for the contrast transfer function

The shape of the contrast transfer function for a 200 keV microscope with C_s=2.0 mm is shown in Figure 13. The curve oscillates around zero and is attenuated at large spatial frequencies. The fact that the CTF is negative at some spatial frequencies means that the contrast in the image is reversed, that white becomes black and black becomes white. This is a potentially serious source of artefacts and must be corrected for. As a straight division by CTF would lead to a

division by zero, there are two other commonly used approaches for the correction. The first is the so called "phase flipping", where we set

$$I'(\mathbf{k}) = -I(\mathbf{k}), \text{ when } \mathrm{CTF}(\mathbf{k}) < 0,$$
$$I'(\mathbf{k}) = I(\mathbf{k}), \text{ elsewhere} \qquad (4)$$

This does not change the amplitudes at the different spatial frequencies, and leads to an overemphasis of the low resolution information. A full CTF correction also has to account for the attenuation effect. One approach to circumvent the division by zero is to use a Wiener filter, such as (Marinescu et al., 2001):

$$I'(\mathbf{k}) = I(\mathbf{k})/[\mathrm{CTF}(\mathbf{k})+w(1-\mathrm{CTF}(\mathbf{k}))], \qquad (5)$$

where w is a small value chosen related to the noise level.

The shape of the CTF also shows that even when the phase reversals have been corrected for, no information is available at the spatial frequencies where the CTF is zero. As the CTF is also affected by the defocus level (Frank, 1996), the solution to this is to take images at different defocii, so that the regions of good information transfer are interspersed between different images. The fact that images taken at different defocii favour different resolutions also raises the technical point that there is really no objective raw data in EM, as the choice of defocus already is a way to filter the data. Overcoming this is another reason to collect images at different defocii.

7.6. Sample preparation and preservation

In their natural state, biological samples contain large amounts of water, making them inherently incompatible with the electron microscope. In the high vacuum of the electron microscope the water would immediately evaporate, leading to the collapse of the sample. One solution to this problem is the use of negative staining, where a metal replica is created of the sample and imaged (Horne and Ronchetti, 1974). This method can, however, give information only about the external detail of the object, and even these are affected by sample flattening and high salt effects on structural integrity (Harris and Scheffler, 2002). A more recent approach is to fix the sample by rapidly cooling it. If the sample is frozen fast enough, no ice crystals that could damage the sample or cause unwanted electron diffraction are formed (see Figure 14). Consequently the sample is preserved almost in its natural state, suspended in vitrified water (Adrian et al., 1984; Dubochet et al., 1988).

Vitrification is accomplished by rapidly cooling the sample, for example an aqueous solution containing viruses, in liquid ethane. The apparatus used for freezing the samples usually contains a guillotine-like plunging device that holds a pair of tweezers with the EM grid. Copper grids with a holey carbon layer are used to support the sample. The grids are first treated with a glow discharge machine to make them hydrophilic and facilitate the even distribution of the aqueous sample on the grid. A grid is then placed in a pair of tweezers and the tweezers are fixed vertically in the plunging device. A small droplet of the sample is then pipetted onto the grid. The thickness of the droplet of sample on the grid determines the thickness of the vitrified layer where the sample will be preserved. If this is very thick, virus particles may overlap and the contrast will be poor (also heat transfer may be slower and cubic ice may appear). Therefore a crucially important step is to blot excess sample away with blotting paper. On the other hand, if too much of the sample is removed, the grid may dry out completely, resulting in empty holes on the grid. Finding the optimum between these two extremes is a matter of practice. After blotting, the sample is immediately

plunged into liquid ethane that is cooled with liquid nitrogen. Ethane instead of nitrogen is used as the cryogen because it does not boil as easily as nitrogen (*i.e.* faster heat transfer). Once the sample is vitrified, it is be kept at liquid nitrogen temperature in all subsequent steps of sample management to maintain the vitrified state (Adrian et al., 1984; Baker et al., 1999).

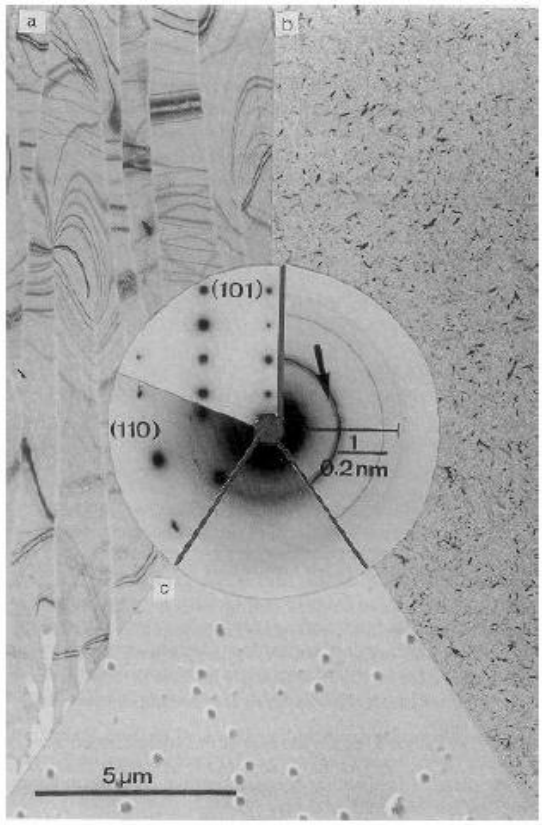


Figure 14. Typical images and electron diffractograms of three forms of solid water observed in the electron microscope. (*a*) Hexagonal ice obtained by rapid freezing of a water layer on a carbon film. The diffractograms, obtained from other specimes, show the (110) and (101) plane. (*b*) Cubic ice obtained by warming a layer of vitreous water obtained by condensation. The shoulder on the (111) reflection, possibly indicating the presence of a small amount of hexagonal ice, is marked by an arrow. (*c*) Vitreous water obtained in the microscope, by condensation of vapour on a cold carbon film supporting polystyrene spheres. The shadow effect demonstrates that the flux of water molecules was anisotropic. Reprinted from (Dubochet et al., 1982), with permission from Elsevier.

7.7. Imaging of cryo-samples

The vitrified sample grid is inserted in the microscope by using a specially designed cryo specimen holder. The stage contains a liquid nitrogen tank for keeping the sample cold. Water introduced with the sample will evaporate in the column. To keep this from condensing on the cold sample, the microscope chamber contains an anti-contaminator with a large surface area kept cold with liquid nitrogen. This traps most of the moisture from the column.

As stated above, inelastic scattering events damage the sample. To minimize the occurrence of these events, the sample is subjected to high intensity electron beam only when necessary. The proper area for imaging is first located by scanning the sample grid in low dose mode at low magnification. Focusing is done with a high intensity beam, but adjacent to the area where the actual image will be next taken. In the recording of the image, the choice of the electron beam intensity imposes a trade-off between improved contrast and possible deformation of the sample (Baker et al., 1999).

7.8. Quality control and particle image extraction

Both the sample handling and the imaging steps described above can go wrong so that the resulting micrographs are not acceptable for further processing. If the sample is allowed to warm up at any stage prior to imaging, it will lead to the formation of ice crystals which may damage the sample and decrease the contrast. Another problem is the movement of the sample during imaging (drift), which causes loss of information in the direction of the movement. There are different sources of drift. The changes caused by the electron beam in the large-scale structure of the ice layer can cause sample movement during imaging (beam-induced drift). The boiling of the liquid nitrogen in the cryo stage dewar can cause vibration of stage (stage-induced drift). Nearby construction work may cause the whole microscope to vibrate (microscope-induced drift). To the expert eye, all these effects are discernible in the micrographs, and more quantitatively in their Fourier transforms, and thus poor quality images can be removed. In particular, the contrast reversals caused by the CTF are seen as Thon rings (Thon, 1971) in the radially averaged Fourier power spectrum. The extent of the rings from the center reflects the resolution range of the image: the further the rings reach out, the higher resolution information is available. On the other hand, if the rings are elliptical, the image is astigmatic, or if the rings are weak in one direction, the image is drifted.

The next processing step (see Figure 15) is to identify and extract the particle sub-images from the microscope images. This can be done either manually or using a computational approach. One approach that is especially useful in the case of icosahedral viruses is the method to locate spherical features in the images (Kivioja et al., 2000). The output of the computational method must still be verified manually, but that is a much smaller task than manually selecting the particles. In highly automated systems even this verification step is skipped, though at the possible cost of additional noise in the results and need for extensive iterative refinements.

8. 3D reconstruction methods

EM images are two-dimensional (2D) projections of a 3D object (see Sections 7.2 and 7.3). A 3D reconstruction is an approximative model of the original object, based on the information in the projections. Intuitively, it is necessary to have multiple different views of the object for the reconstruction to be possible: a book and a row of books, for example, look the same from one view, but two perpendicular views reveal the difference. The method used to get the different views depends on the specimen. If the sample is unique, such as a cell, we must rotate it in the microscope to see it from different sides, or equivalently, if the specimen is a human undergoing a magnetic resonance imaging scan, we rotate the camera around patient. This is tomography. In a case where the specimen consists of multiple copies of identical particles, such as viruses, the multiple views are readily available in one single micrograph, as the particles are randomly oriented in the vitrified water. If the orientations are not random, if the particles are ordered in a 2D crystal (Glaeser, 1999) or if they have some feature that causes them to always align in a particular way in the vitrified water, tilting of the sample in the microscope is necessary. For example the isolated tail machine of P22 would always orient with the axis of the tail normal to the sample plane (Tang et al., 2005).

The symmetry of the specimen is helpful in obtaining a sufficient number of views necessary for reconstructing the original object at a reasonable resolution, as multiple copies of the asymmetric unit can be seen from different views within one particle. It is for example possible to reconstruct the 3D structure of a filamentous phage from just one virus image, given that the filament is long enough (De Rosier and Klug, 1968).

Low signal-to-noise ratio is a major problem in the analysis of electron cryo-microscopy images. Symmetry helps in overcoming this problem as it increases the amount of data, *e.g.* one image of an icosahedral virus contains 60 views of the asymmetric unit. Furthermore, since the angular relationships between the asymmetric units are known, determining them will not introduce any errors, as it will in the case of unconnected images of an asymmetric object (Crowther, 1970). In the case of the asymmetric object classification methods can be used to find views that are similar and can thus be averaged (van Heel et al., 1996).

In the remainder of this section I will explain the principals of how the 3D reconstruction is calculated (see Figure 15). I will also introduce the classification approach that is helpful in case the object is asymmetrical or the symmetry is unknown. Together they form the basis of the novel in-situ reconstruction method developed and applied in this thesis. To put this work in context, I will also touch upon previously existing methods to tackle symmetry mismatches.

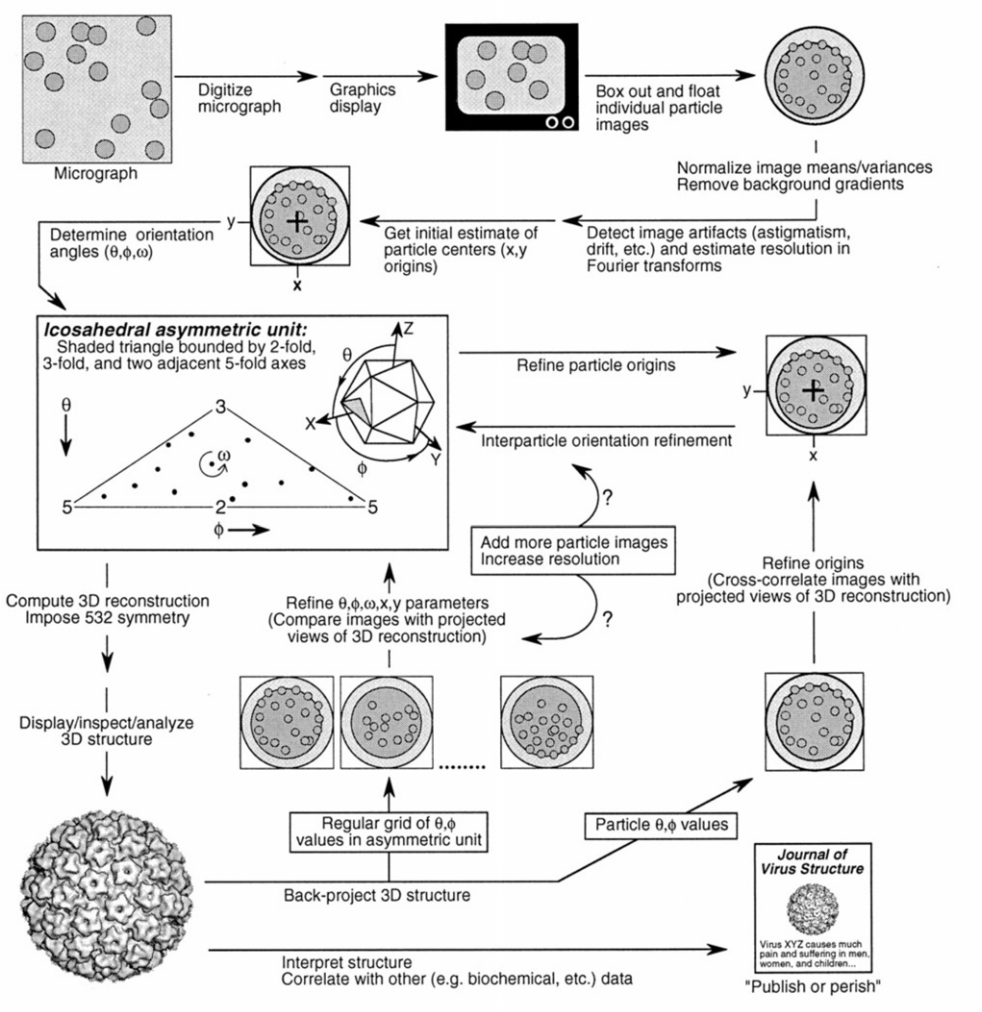


Figure 15. Model-based reconstruction process. Schematic diagram of the 3D image reconstruction process from digitization of the micrograph to the dissemination of the structural results. Though some steps such as the boxing out of individual particle images lend themselves to automation (*e.g.*, see (Boier Martin et al., 1997) and (Thuman-Commike and Chiu, 1996)), many steps including the determination of particle origins and orientations must be repeated and involve some trial-and-error decision making (*e.g.*, to determine which data should or should not be included). The scheme depicted here, which uses both cross-common lines (Fuller et al., 1996) and model-based (Baker and Cheng, 1996) approaches to determine and refine particle origin and orientation parameters, is but one of many suitable schemes. With permission from (Baker et al., 1999) © American Society for Microbiology.

8.1. Orientation search and density map calculation

To be able to make a 3D reconstruction of 2D projections we must know the angular relationship between the projected views. In tomography the

projection angles are known, but in the case of many identical particles they need to be determined from the data. There are basically two approaches to this: we can either find out the angles relative to an external model, or in pair-wise fashion within the group of projections. The first approach is termed "model-based" and the latter the "common lines" approach.

Both real (Van Heel, 1987) and reciprocal space (Fourier) algorithms (Fuller et al., 1996) exist for the common lines method. The basis of the reciprocal space methods is the projection theorem. The projection theorem states that 2D Fourier transforms of projections of a 3D object are central sections through the 3D Fourier transform of the original 3D object. This implies that the transforms of two projections taken at different angles intersect. The line of intersection is called the "common line", whence the name of the method. Equivalently, in real space, the sinograms of two projections taken from the same object contain an identical line. By searching for the identical line in the transforms or the sinograms it is possible to find out the angle between the projections. Three views are required to span the 3D space to give a unique solution. However, it should be noted that when symmetrical objects are projected, the asymmetric units correspond to "particles within the particle" and thus give rise to "self-common lines". These can be used to find the axes of symmetry in individual projections (Fuller et al., 1996).

In the model-based approach (Baker and Cheng, 1996), projections of the initial model are made at an appropriate angular sampling, and the images are compared with these. The projection angle of the model projection that best matches a particle image is taken as the orientation of the particle. In the Polar Fourier Transform (PFT) method (Baker and Cheng, 1996), the images and model projections are first transformed into polar coordinates (r,γ), and one dimensional Fourier transforms are calculated for each r; the PFT is thus an array of one-dimensional transforms. The measure of similarity is the correlation between the PFTs, which as a reciprocal space measure is not sensitive to the absolute contrast in the images or the density. Other distance measures can also be defined (Ji et al., 2006).

Real and reciprocal space methods exist also for the task of computing the 3D reconstruction. An example of the real space methods is the filtered back-projection algorithm used in IMAGIC software (van Heel et al., 1996). Reciprocal space algorithms are used in PFT (Baker and Cheng, 1996) and P3DR (Marinescu et al., 2001).

In the reciprocal space methods, two methods are used for "summing" the projection transforms to create an approximation of the 3D transform of the object. In the Fourier-Bessel methods (Crowther, 1971), the density to be reconstructed is expressed in terms of (yet undetermined) cylindrical functions. This has two benefits: i) it gives rise to a natural way to sample the transform from the data; and ii) it presents a computationally effective way to apply 52-symmetry to the reconstruction (Crowther, 1971). The actual task of reconstruction consists of determining the coefficients of the Fourier-Bessel expansion of the transform, given the image data. From these, the real space cylindrical functions can be determined, and they in turn give the real space reconstructed density. In contrast, P3DR (Marinescu et al., 2001) uses Cartesian, instead of cylindrical, coordinates for sampling the data. This has the benefit that the algorithm is easily parallelizable.

Once the 3D reconstruction has been calculated, it is used as the model in the next round of iteration to refine the orientation and origin parameters. The process is repeated until the resolution of the model no longer improves.

8.2. Multivariate statistical analysis

In multivariate statistical analysis (MSA) as it is applied here, a digital 2D image of *N* pixels, is considered a point (or a vector) in *N*-dimensional space. The main principle of classification is to find which points are close to one another in this space, so that they can be grouped together. This is illustrated in Figure 16A for images consisting of two pixels.

Figure 16A also shows that the set of images may incorporate different kinds of variation. It can be readily seen, that most, but not all, of the variation is in the direction of the x-axis. In the analysis of EM images the situation is much more complex as the dimensionality of the problem is much greater. There are many sources of variation within the images: noise, absence of features, orientation, etc., and all of these are represented in the relations of the image-vectors in *N*-dimensional space. In order to classify the data, we first need to determine which kind of variation exists in the set of images, and which kind of variation is meaningful. In terms of the simple example in Figure 16, we are looking for the vectors that best describe the variation. Mathematically, the variation within the set can be described with the eigenvectors of its variance covariance matrix. In the current application, as vectors in the *N*-dimensional space correspond to images, the eigenvectors are usually termed *eigenimages*. The computation of eigenimages and the further steps described here are implemented in IMAGIC (van Heel et al., 1996).

Figure 16B shows a set of eigenimages for the SH1 vertex reconstruction project. The eigenimages clearly represent different kinds of variation: the one on the left shows something about the capsid of the virus, whereas the others something about the spike. The next step in the process is to select the eigenimages which are relevant to the study, and to run the classification algorithm, in which the principal sources of variation are now set by the selected eigenvectors. The result of the classification step is a set of class-averages (Figure 16C). These class-averages can now be subjected to either model-based or common-lines based orientation determination.

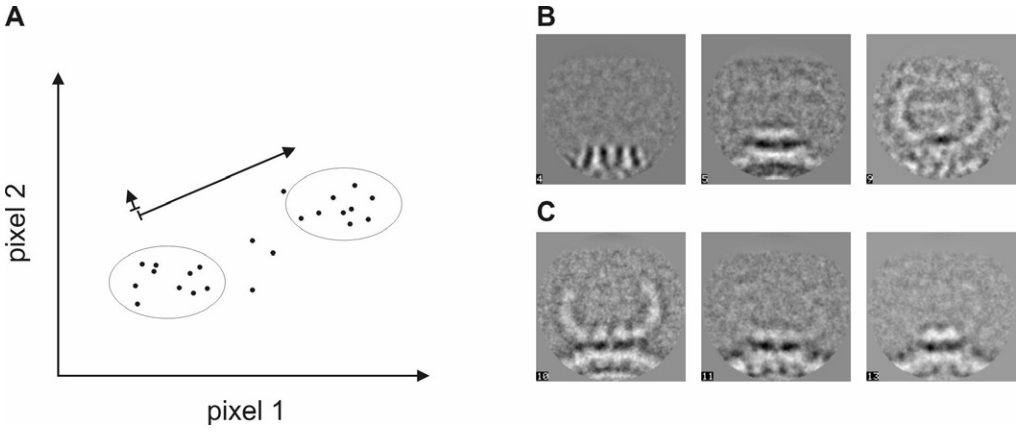


Figure 16. Multi-variate statistical analysis. A. Classification of images consisting of two pixels. B. Eigenimages of SH1 spike data (Article III). C. Class-averages of SH1 spike data (Article III).

8.3. Previous methods used to tackle symmetry mismatches

Symmetry mismatches are of great biological interest and several attempts have been made at solving symmetry mismatched features of icosahedral viruses (Section 2). Some of these are summarized here.

8.3.1. Reducing symmetry

In 3D reconstruction algorithms such as PFT or P3DR the symmetry that is applied on the reconstruction can be set by the user. Therefore one possibility is to relax the symmetry from full icosahedral symmetry, or drop it altogether. de Haas et al. (1999) studied various recombinant versions of the unpackaged cystovirus ϕ6 polymerase complex in this manner, in order to localize the PC proteins P1, P2, P4 and P7, and to analyze the substructure at the vertices of the dodecahedral PC, which they showed did not comply with icosahedral symmetry. The C1 and C5 reconstructions did not reveal any detail about the vertex structure, but the authors were nevertheless able to make statements about its occupancy. The obvious drawback is that data from the twelve vertices was not combined.

Another way to approach icosahedral viruses without imposing the symmetry is to look at only one particle at a time by tomography. This proved fruitful in the case of herpes simplex virus type 1 (HSV-1) (Grünewald et al., 2003). The capsid of HSV-1 is icosahedrally symmetric (Zhou et al., 2000), and is enveloped by a membrane bilayer. Between the membrane and the capsid is a proteinaceous layer, the tegument. The tomography study (Grünewald et al., 2003) showed that the tegument does not evenly surround the capsid, but forms a “cap” on one side of it. Due to the presence of the cap, the capsid is also not in the center of the virion. The tomographic reconstructions were of high enough resolution to distinguish individual spikes protruding from the membrane. The authors could estimate the resolution of the tomograms by comparison with the known cryoEM reconstruction (Zhou et al., 2000) as 6 to 9 nm, depending on the direction. The difference is caused by the lack of data from a range of angles in the tilt series (due to mechanical limitations of the stage), leading to a “missing wedge” effect. The Fourier shell correlation method (Harauz and van Heel, 1986) that is generally used for estimating the resolution of EM reconstructions is not applicable in tomography.

Tomography has also been used to study the structure of human immunodeficiency virus 1 (HIV-1) (Briggs et al., 2006). The tomograms showed the cone-shaped cores of the virus inside the mostly spherical membrane layer. The structure allowed the authors to propose a model for the formation of the core, where the core starts to grow at its narrow end and proceeds until it meets the membrane.

The resolution attainable by tomography is limited by the maximum electron dose that the sample can be subjected to before damaging it. Otherwise the method works equally well on symmetric and asymmetric samples.

8.3.2. Looking through the symmetry

The greatest advantage the high 532 symmetry offers is, in icosahedral viruses, the great boost in the signal to noise ratio. This also means that features that adhere to the symmetry dominate in the particle orientation search, whereas other features have much less impact. One way to overcome this is a model-based

orientation search using only projections that are equivalent from the point of view of the symmetry. This way the symmetrical parts in the images correlate equally with all model projections and become, in a way, transparent to the processing. Now the symmetry-mismatched features determine the orientation. This method assumes that the features of different symmetry are in a locked position with each other.

This kind of approach was first applied on the tailed phage ϕ29 (Morais et al., 2001; Tao et al., 1998). Morais et al. (2001) first used C5 symmetry to align all the tails of the phages in the same direction and to compute an initial model. The subsequent processing steps were aimed at resolving the symmetry mismatch between the head and the connector of the tail. The model projections were taken perpendicular to the five-fold axis, at intervals of $2\pi/5$. No symmetry was applied in the reconstruction, and the process was iterated in the model-based manner. In the end, the method produced an improved, 33Å resolution model, where the symmetry-mismatch between the head and the 6-fold symmetric tail assembly was resolved.

A similar strategy was used to reconstruct the tailed phages epsilon 15 (Jiang et al., 2006) and P22 (Chang et al., 2006) to 20Å resolution without applying any symmetry. In these cases the head of the virus is not prolate, so the artificial initial model could be projected in sixty icosahedrally equivalent views.

8.3.3. Erasing the symmetry

In the previous method, the symmetry related features in the images are made transparent to the projection matching by sampling the orientation space only at symmetry-equivalent angles. The drawback is that the angular resolution is limited in the process. An alternative approach is to erase the symmetry-related feature altogether, by subtracting projections of the reconstruction with an imposed symmetry from the images.

(Morais et al., 2003) used this approach to study the organization of the scaffolding protein within the bacteriophage ϕ29 procapsid. Without applying any symmetry, they found concentric layers of density with different axial symmetries inside the head. Viewed from the side, the structure forms a lattice where the crystal structure of the scaffolding protein (Morais et al., 2003) could be fitted.

B. AIMS OF THE STUDY

1. To chart the virosphere

Viruses are an essential player in the ecosystem of the Earth. Even though viruses have been studied for decades, our understanding of them is still far from complete. We lack understanding both of the depth, of the working principles of individual viruses, and of width, of the total landscape of viruses. Therefore, also this thesis is a tiny contribution into the growing picture. Specifically, the aim was to bring in new information about three viruses, the bacteriophages ϕ6 and ϕ8, and the archaeal virus SH1.

1.1. Deepen our knowledge of the cystoviruses

The assembly pathway of ϕ6 and the corresponding structural intermediates have been studied extensively, but there was a clear gap in the structural information: we had not yet reconstructed the complete virion (Figure 10). Filling this was one goal. Additionally, in 2002 when this study was started, nothing was known about the structure of ϕ8, or any of the other cystoviruses apart from ϕ6, for that matter. We clearly needed to have a system to compare ϕ6 to, in order to be able to make conclusions about common principles on the level of the *Cystoviridae* family. Thus another goal was to gather as much structural information as possible from ϕ8. Specifically, we wanted to reconstruct the complete virion and the polymerase complex, and study the packaging enzyme P4 structure within the intact polymerase complex.

1.2. Go to new territories, take a look at the Euryarchaeal viruses

When this thesis project started, no 3D structures of archaeal viruses were available. As the initial purification of SH1 had been started, the goal was to complement this work by solving the structure of SH1. As the biochemistry work progressed, it became clear that it would be possible to further complement the dissociation studies with structural work, and so the identification and detailed structural analysis of the SH1 spike was taken as the final goal of this thesis.

1.3. Learn about virus evolution

Both of the goals above, even though they have a well defined content, are also brushstrokes in the bigger picture, and in both cases we had a reason to believe that we would be able to learn something new about virus evolution. In the case of SH1 the goal was to find out if it belongs to the proposed lineage of viruses with the double β-barrel trimer capsids (Benson et al., 2004). This would be significant in the long time-scale evolution of viruses: especially so in the beginning of the project, when the structure of the crenarchaeal STIV (Khayat et al., 2005; Rice et al., 2004) was not known, but even now, as SH1 is the first structure known from the kingdom of Euryarchaea. For the cystoviruses, the goal was to find out about similarities and differences within the family, and from there to learn about evolution in a shorter timeframe.

2. Study symmetry mismatches in virus structure

Icosahedral reconstruction methods have been the bread and butter in this field since 1971 (Crowther, 1971), but as structural virology has progressed in depth, it has become evident that they miss out on some interesting biological aspects. This is simply because the symmetric structure is not conducive to movement and action; its core competence is to be stable and protect the genome. Many phases in the viral life cycle, such as genome packaging and receptor binding, on the other hand may require dynamic action, and it has been indicated that the structure corresponding to these may not comply with the overall icosahedral symmetry. For this reason it is important to find new methods for dealing with symmetry mismatches.

2.1. Method development

The goal in this thesis project was method development in vertex reconstruction (Briggs et al., 2005). In particular, the published method only dealt with partial occupancy of five-fold symmetric objects in icosahedral viruses. Here the goal was to develop functionality to address the issue of rotational symmetry mismatch. Another goal was to improve the usability of the method, in order to make it possibly more appealing to a broader user base.

2.2. Test the method and apply it to the target viruses

Experience with other cystoviruses led us to believe that the ϕ8 packaging enzyme would be symmetry mismatched relative to the polymerase complex. Therefore we had an obvious test case with a large dataset in our hands, and motivation to develop the method in the manner stated above. Furthermore, as the initial three dimensional structures of SH1 were calculated, it soon turned out that we actually had another excellent test case for the method, and we thus set as a further goal the asymmetric reconstruction of the SH1 spike.

C. MATERIALS AND METHODS

The exact methods for the propagation and purification of ϕ6 and ϕ8, as well as the method for producing ϕ8 cores, can be found in Article I. For SH1, the protocols have been documented previously {Kivelä, 2006 #33}. SH1 and the host *Haloarcula hispanica* were grown in modified growth medium containing 18% (w/v) artificial salt water. Cell debris was removed from the lysate and the virus particles were concentrated and purified by rate zonal centrifugation. SH1 subviral particles VP36$^-$ (where proteins VP3 and VP6 are missing) and VP236$^-$ (where VP2, VP3 and VP6 are missing) were created by incubating purified 1× purified virions (1 mg ml^{-1}). To produce VP236$^-$, incubation was done in 40 mM Tris-HCl, pH 7.2, 1 M NaCl and 40 mM $MgCl_2$ at 37°C for 20 hours. For the production of VP36$^-$, the buffer was modified to contain 40 mM Tris-HCl pH 7.2, 0.75 M Na_2SO_4, and 40 mM $MgSO_4$ and incubation was continued for 42 to 48 hours at room temperature. The dissociation products were purified by rate zonal centrifugation and analyzed by SDS-PAGE (Schagger and von Jagow, 1987). For cryoEM imaging, the sample was resuspended in 40 mM Tris-HCl, pH 7.2, 1 M NaCl , 40 mM $MgCl_2$ and 5 mM $CaCl_2$, to improve the stability of the virus particles.

For all the projects, EM data was collected on a FEI Tecnai F20 microscope operating at 200 keV, with spherical aberration of 2.0 mm. The nominal magnification used was 50000x. Micrographs were scanned on a Zeiss Photoscan TD scanner.

Contrast transfer function estimation was done with CTFFIND3 (Mindell and Grigorieff, 2003). Icosahedral reconstructions were calculated with PFT2/EM3DR (Baker and Cheng, 1996) and POR/P3DR (Marinescu et al., 2001). The in-situ reconstruction method is implemented using IMAGIC (van Heel et al., 1996). BSoft (Heymann, 2001) and EMAN (Ludtke et al., 1999) were used for various image processing tasks. The details of the processing are described in the original articles. All calculations were run on a Linux cluster.

Segmentation of molecular boundaries in the capsid was done with qsegment of the EMAN package (Ludtke et al., 1999). In the case of ϕ8, the repeating subunit was found by visual inspection of spherical shells of the reconstruction, and by comparison with the analysis of ϕ6 (Huiskonen et al., 2006a). In SH1 the capsomers were easily discernible, but there it was desirable to improve the signal by averaging capsomers of the same type. This was done by fitting pseudo-atomic models of type 2 and type 3 capsomers into the asymmetric unit using CoLoRes (Chacon and Wriggers, 2002), and by applying the inverse transformations on the asymmetric unit to position the superimpose the similar capsomer.

Helix assignment for ϕ8 was done with helixhunter (Jiang et al., 2001a), and for visualization, Chimera (Huang et al., 1996) was used.

D. RESULTS

The key points of the structural studies are listed here. For details the reader is referred to the original articles.

1. The structure of ϕ6 virion

The structure of the intact ϕ6 virion was determined to 14-Å resolution (Article I). The structure shows the same structural layers as the nucleocapsid reconstruction (Huiskonen et al., 2006a), plus the membrane and spike layers. However, no distinct proteins were revealed in the membrane or spike, or for the lysozyme P5 that is located between the P8 and the membrane (Hantula and Bamford, 1988). There is a notable difference between the P8 T=13 layers of the virion and the nucleocapsid: the four-helix bundle of the peripheral domain of the P8 is more ordered in the former. The change in conformation may play a part in disassembly or membrane recognition during assembly.

2. The structures of ϕ8 virion and ϕ8 core

The structures of ϕ8 virion and ϕ8 core were determined to 18 and 8.5 Å resolutions, respectively (Article I). The core capsid protein P1 is organized in a T=1 lattice of asymmetric dimers. The detailed structures of the core P1 shells are different in ϕ6 and ϕ8, as revealed by modelling of α-helices into the densities. The hexameric packaging enzyme P4 is symmetry mismatched at the five-fold vertices. The symmetry mismatch was resolved for ϕ8 in Article II. Neither the polymerase P2 nor the assembly cofactor P7 could be localized in the reconstructions. The genome has at least two ordered layers with an average spacing of 2.9 nm.

The virion shows the same structural layers as the core, with the addition of the membrane and spike layers. There is no T=13 layer corresponding to the ϕ6 P8 in ϕ8. Instead, the virion has 60 copies of a putative membrane associated protein attached to the P1 shell, approximately at the intersection of the monomers in the P1 dimer. These are not present in the ϕ8 core.

ϕ8 is similar to ϕ6 in size and shape. The spike layer of ϕ8 is thicker than that of ϕ6, giving the virus a slightly larger diameter, otherwise the equivalent layers are located at the same radii. No individual spikes or transmembrane proteins were resolved.

3. Method development for symmetry mismatch studies

The vertex reconstruction method introduced by Briggs et al. (2005) was developed further to allow analysis of rotational symmetry matches at the vertices of the virus (Article II). The method was successfully applied in the study of ϕ8 packaging enzyme (Article II) and the SH1 spike (Article III).

4. The structure of the ϕ8 packaging enzyme

The structure of the ϕ8 P4 was determined to 15 Å resolution (Article II). The asymmetric reconstruction shows that the P4 is indeed hexameric within the core. There is an opening at the base of the hexamer, and one putative connection to the P1 shell. The hexamer appears to be fixed relative to the capsid, indicating that it does not rotate during packaging, contrary to what has been suggested (Hendrix, 1978).

5. The structure of the SH1 virion

The structure of the SH1 virion was determined to 9.5 Å resolution (Article III). SH1 has a novel capsid architecture with triangulation *T*=28. Noticeably, the capsid has two types of capsomers: type 2 capsomers at two-fold axes and at one of the adjacent positions have two distal towers, whereas capsomers at all other locations (type 3 capsomers) have three towers. The towers appear to consist of small β-barrel domains that are located on top of the hexameric base of the capsomer which in turn consists of six β-barrels. The hexameric base is slightly skewed in type 2 capsomers. All capsomers make connections to the membrane. A smaller, five-fold symmetric structure closes the capsid at the vertices. Each of the monomers in this structure could accommodate a single β-barrel, akin to the adenovirus penton (Zubieta et al., 2005) or PRD1 P31 (Abrescia et al., 2004). Below the vertices is a large transmembrane complex that penetrates both layers of the membrane, strongly modulating the membrane curvature.

Outside the capsid at the five-fold vertices are large horn-like spikes. These were incorrectly averaged in the icosahedral reconstruction, but were reconstructed with the in-situ method to 32 Å resolution using two-fold symmetry. This was clearly the appropriate symmetry, as suggested by the initial reconstruction using no symmetry. The spikes consist of ten tubular densities stacked side by side.

6. The structure and composition of the SH1 vertex

Two subviral particles of SH1 were created: VP236$^-$, lacking proteins VP2, VP3 and VP6; and VP36$^-$, lacking proteins VP3 and VP6. The structures of VP236$^-$ and VP36$^-$ were determined to 10.5 Å resolution (Figure 17). The spikes are missing in both particles. When compared with the intact SH1, the capsid and membrane organization does not change in VP36$^-$, but in VP236$^-$ the membrane order is lost. A protein analysis shows that the spikes are formed of proteins VP3 and VP6. An infectivity assay shows that VP36$^-$ is infectious, indicating that the spikes are not needed for receptor binding or other critical step in the infection process. In contrast, VP236$^-$ is not infectious, indicating that VP2 has an active role in the infection process, or that it is so crucial to the integrity of the particle that infection is no longer possible without it. The VP3 and VP6 from both types of dissociation experiments ran together in gel filtration, but VP2 did not co-elute with them. When cryoEM images were recorded of the dissociated spikes, isolated before and after the gel filtration, they appeared identical. This confirmed the results from difference imaging that the spikes are formed of proteins VP3 and VP6. As VP3 is one of the major proteins of the virion, it probably makes up most of the ten domains, and VP6 may well connect the spike to the capsid.

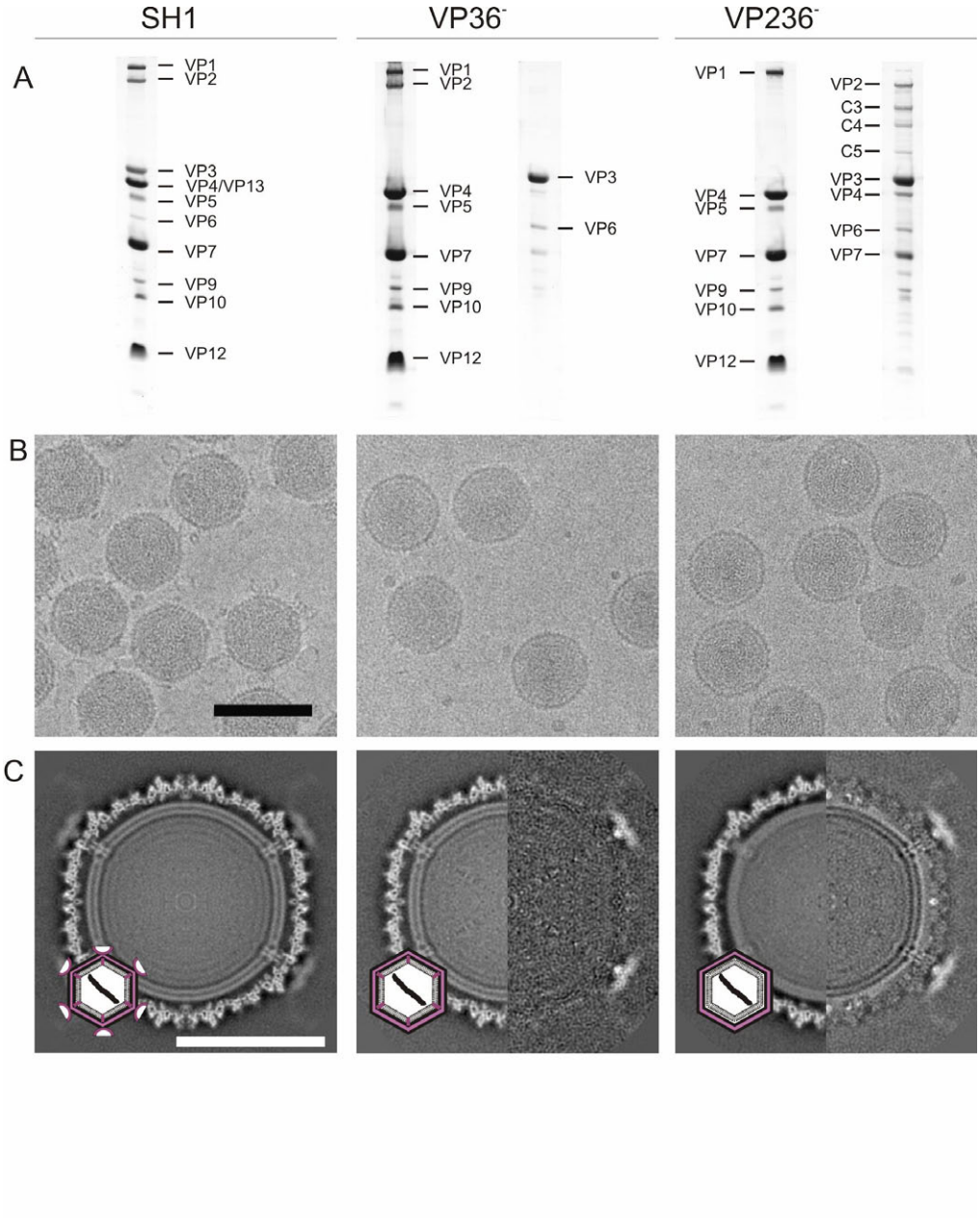


Figure 17. Characterisation of the SH1 dissociation products. A. SH1, the subviral particles VP36$^-$, VP236$^-$, and the corresponding released proteins analysed by SDS-PAGE. B. Electron micrographs of the particles SH1, VP36$^-$ and VP236$^-$ from left to right. Scalebar represents 100 nm. C. Central sections of the 3D reconstructions and difference images. From left to right: intact SH1 virion; VP36$^-$ and difference between SH1 and VP36$^-$; VP236$^-$ and difference image between SH1 and VP236$^-$. D. Schematic diagrams highlighting the changes in the particles: the loss of spikes in both VP36$^-$ and VP236$^-$, and the disordering of the vertex transmembrane complex and the slumping down of the membrane in VP236$^-$.

E. DISCUSSION

1. The novel capsid architecture of SH1

We lost the race to reconstruct the first archaeal virus to the researchers of STIV (Rice et al., 2004), but came in a good second with the first structure of a euryarchaeal virus (Article III). The structure of SH1 has some interesting features which make it novel also in other ways. Firstly, the *T*=28 lattice has not been observed before. It is the smallest handed *T*-number for which the Wrigley scheme predicts a symmetron that would place a capsomer on the two-fold axis of symmetry (Simpson et al., 2003; Wrigley, 1969). Most interestingly, in SH1 the capsomers that correspond to the predicted disymmetron are actually different from the other capsomers. This is the first time a disymmetron has been observed in a virus structure. Another point with regard to the Wrigley scheme is that the way the disymmetrons are oriented in SH1 makes the trisymmetrons not triangles but triskelions (Figure 18).

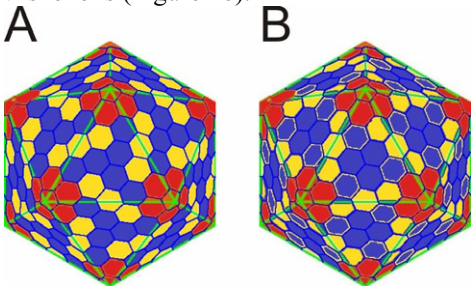


Figure 18. Symmetrons in *T*=28 capsid. Pentasymmetrons are colored red, trisymmetrons blue and disymmetrons yellow. A. The pattern of symmetrons that is observed in SH1 structure, with triskelion-like trisymmetrons. B. The "standard" pattern, with triangular trisymmetrons.

The SH1 capsid consists of two clearly different types of capsomers. That the bases of the capsomers appear similar apart from the skew in the type 2 capsomers is suggestive of an assembly consisting of heterohexamers (the bases) and decorating proteins (the towers) bound in a conformation-dependent manner. This is also supported by the observation that the towers seem to consist of small individual β-barrels, not of the tops of taller barrel as in the known double β-barrel coat proteins (Section A.2.2). However, given the limited resolution of the reconstruction of the model, these observations must be considered with caution. VP3, VP4 and VP7 have previously been considered the putative capsid proteins (Bamford et al., 2005b). As we showed VP3 to be a spike protein, the capsomers probably consist of VP4 and VP7.

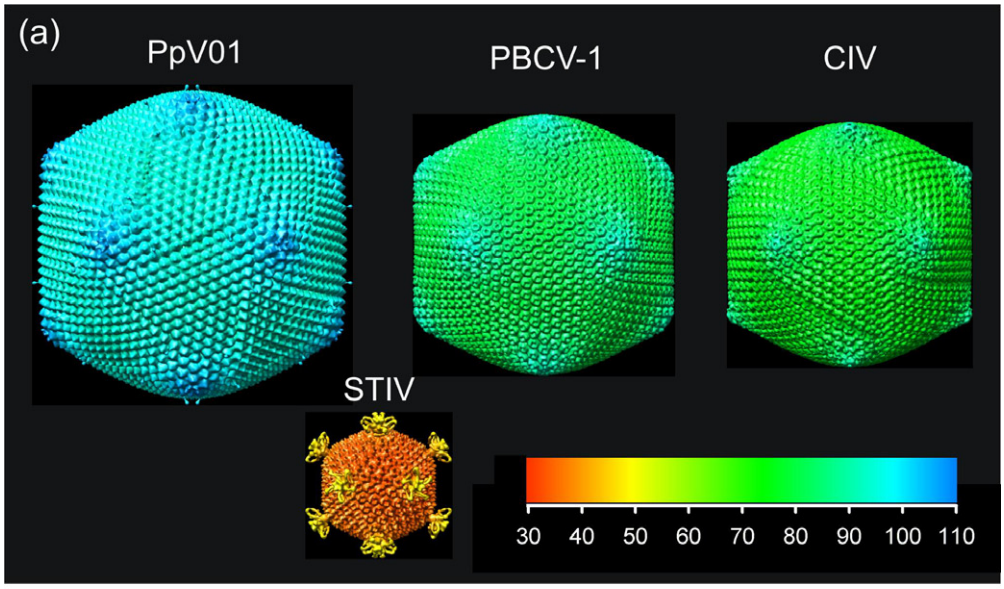
Skewed capsomers, such as the SH1 type 2 capsomers, have been observed in the procapsids of bacteriophages HK97 (Wikoff et al., 2006), P22 (Zhang et al., 2000) and λ (Dokland and Murialdo, 1993), and herpes simplex virus type 1 (Heymann et al., 2003). In all the cases, the procapsid are more round than the mature capsid where the capsomers have assumed a non-skewed conformation. The skew of the type 2 capsomers probably explains the rather round form of SH1 (Figure 19). SH1 is much less angular than Bam35, for example (Laurinmäki et al., 2005).

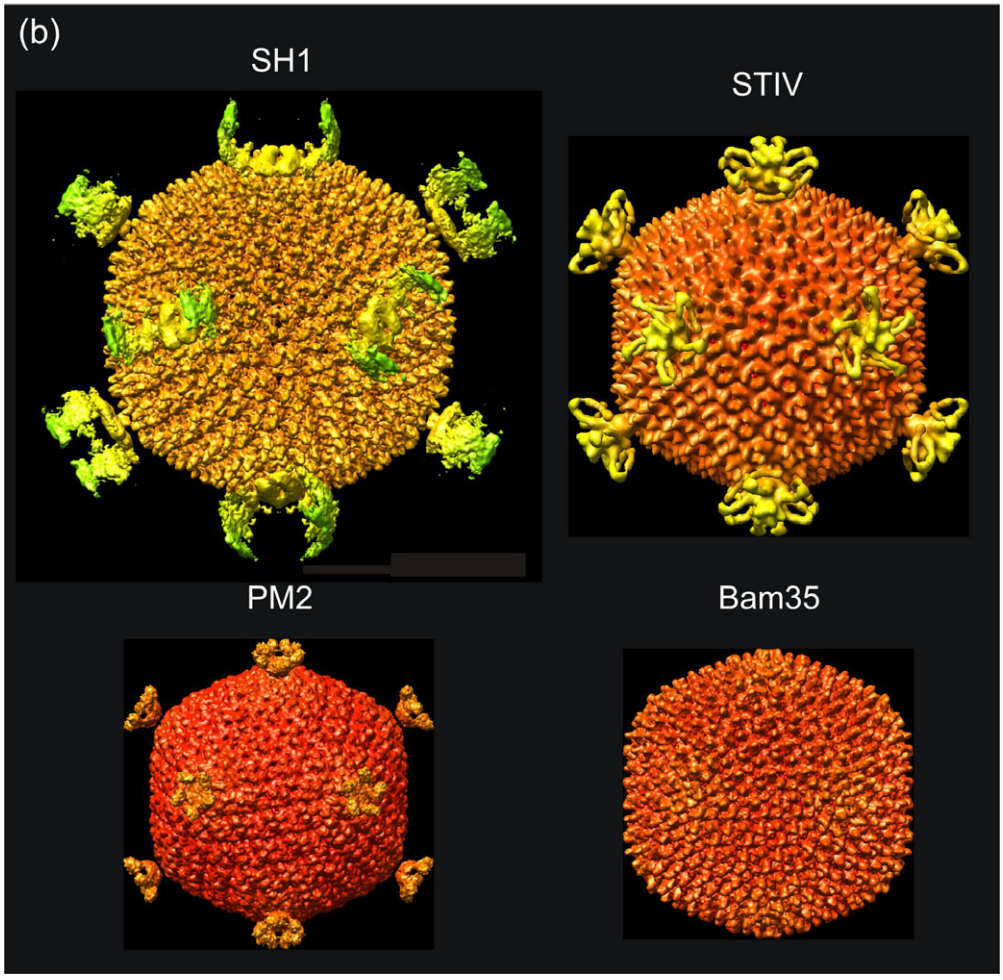
The SH1 membrane is icosahedrally organized. The average thickness of the SH1 membrane is 24 Å. Despite the differences in lipid composition, this is quite similar to the membranes of PRD1, Bam35 and PM2 which are also modulated by membrane proteins (Laurinmaki et al., 2005).

When the first micrographs of SH1 showed large spikes, the immediate assumption was that they would be

involved in receptor binding, as many known virus spike structures are. However, we have showed that this is not the case, which immediately raises the question of their actual function. The SH1 spike bears some similarity to the spikes of the *Acidianus* filamentous virus 1 (AFV1) (Bettstetter et al., 2003). The 20 nm wide "claws" of AFV1 were observed to bind to the pili of the host. Similarly, *Acidianus* two-tailed virus (ATV) has anchor like structures at the end of its tails (Prangishvili et al., 2006b). The remarkable thing about ATV is that it grows its tails outside the host. The tails also did not appear to affect infectivity. The authors proposed that the purpose of the tails might be to increase the probability of host contact. The same might be true for the SH1 spikes.

Figure 19. Comparison of dsDNA membrane-containing viruses. (a) Isosurface representations of large dsDNA virus reconstructions are shown for PpV01 (T=219, diameter=220 nm) (Yan et al., 2005), PBCV-1 (T=169, diameter=190 nm) (Yan et al., 2000) and CIV (T=147, diameter=185 nm) (Yan et al., 2000). Mimivirus (diameter=460 nm) had to be omitted from the figure due to lack of space. The surfaces are colored radially as indicated by the color bar (units in nanometers). The much smaller STIV is shown for comparison (see *b*). (b) Isosurface representations of SH1 (Article III), STIV (T=31, diameter=100 nm) (Rice et al., 2004), PM2 (T=21, diameter=74 nm) (Huiskonen et al., 2004), and Bam35 (T=25, diameter=70 nm) (Laurinmaki et al., 2005) reconstructions are shown. STIV and PM2 both have prominent five-fold symmetric spikes attached to the vertices. The coloring is as in a. The scale is 300% larger than that in a. The view in all images is along an icosahedral two-fold axis of symmetry. The figure was made in Chimera (Huang et al., 1996) by J. Huiskonen from the original electron density maps kindly provided by the original authors or from those deposited in the EM database (http://www.ebi.ac.uk/msd-srv/emsearch/).


(a)
PpV01
PBCV-1
CIV
STIV
30 40 50 60 70 80 90 100 110


(b)
SH1
STIV
PM2
Bam35

2. Methods of resolving symmetry mismatches

Simultaneously with this work, method development similar to the in-situ asymmetric method has been going on. Zhang et al. (2006) used a similar approach to study the structure of flavivirus core. The authors initially made an icosahedral reconstruction of the whole particle. In this, the core did not show any detail. They next created projections of the symmetric protein and membrane layers of the reconstruction, and subtracted these from the particle images, using the same scale factor k for all images on a given micrograph. The scale factor was chosen so that the 3D reconstruction of the difference images with the selected k gave the lowest correlation with the original icosahedral reconstruction in the symmetric protein and membrane regions. This ensured that the difference images are only noise in these regions. The difference images were then subjected to multivariate statistical analysis, and the resulting class averages were used to compute a 3D reconstruction. The process was iterated in a model-based manner until no further refinement was possible. In summary, what the approach of Zhang et al. and our method (as it as used in Article II) have in common is the subtraction of icosahedral information followed by a classification analysis of the symmetry mismatched components. However, in our case we could take better advantage of the icosahedral orientations by minimizing the uncertainties of the orientation angles, by aligning the spikes with respect to one angle and by grouping them with respect to another angle, wheras Zhang et al. had to start the single particle analysis from scratch. Our approach gave a clear advantage over the asymmetric reconstruction method used earlier by de Haas et al. (1999). We could combine information from all vertices to one reconstruction, whereas they ended up with twelve reconstructions, one at each vertex.

The successful work on the structures of the tailed bacteriophages epsilon 15 (Jiang et al., 2006) and P22 (Chang et al., 2006) is another recent example of asymmetric cryoEM work. In the case of tailed phages, the symmetry mismatch is always at one position in the capsid, and the unique vertex can be found by projection-matching using symmetry-equivalent projections. However, this would not work for a structure like the ϕ8 core, where the symmetry mismatch is present at all vertices, and thus the applicability of the method is limited. On the other hand, the method developed in (Briggs et al., 2005) and Article II would be directly applicable to reconstruct a tailed phage, given that the head of the phage is icosahedral.

Tomography is maybe the most promising approach to study the structure of individual spikes in viruses such as the cystoviruses, where a layer of spikes protrude from the membrane. An icosahedral reconstruction can at the best reveal the average length of the spike as measured from the membrane layer. With a tomographic reconstruction on the other hand, it is possible to locate and extract individual spikes from the tomogram and average these to get 3D information about the structure. Such studies have been done on simian immunodeficiency virus 1 (Zanetti et al., 2006; Zhu et al., 2006). Each of the two groups proposes somewhat different structure for the spike. This may be partly explained by methodological differences, as the field of electron cryo tomography is still fairly new.

3. Evolutionary considerations

3.1. Observations from the ϕ6, ϕ8 and SH1 structures

As can be seen in Figure 7, and as was confirmed by the similarity of the structures of ϕ6 and ϕ8 (Article I), the bacteriophage version of the *T*=1 fold of asymmetric dimers differs considerably from the examples that infect eukaryotic hosts. This might reflect the long evolutionary time span separating the groups, or it might equally well be simply a result of under-sampling of the virosphere (*e.g.* the viral structures we know are not representative of the two classes of viruses). However, even though the eukaryotic dsRNA viruses seem to be very similar, the similarity is only at the tertiary (subunit) and quaternary (capsid) levels. They all have in common the division of the subunits into apical, carapace and dimerization domains, but the structures still cannot be readily superimposed (Reinisch et al., 2000). In Article I we compare the P1 shells of ϕ6 and ϕ8, and come to the conclusion that only the quaternary structure is preserved. We cannot convincingly overlay the modelled groups of helices. This may be an artefact of working at a lower resolution and crystal structures may prove us wrong, but the fact that the helices in the two independent subunits of the dimer are modelled similarily in both ϕ6 and ϕ8 does lend credibility to the modelling. With the present data and resolution, we cannot say if cystoviral capsid protein subunits also share a common pattern of organization as do those of the other dsRNA viruses. However, one thing all dsRNA phages do seem to have in common is the hexameric packaging enzyme at the five-fold vertices of the PC (Article II) (de Haas et al., 1999; Huiskonen et al., 2006b). This implies an assembly pathway where the genome is packaged into a preformed empty procapsid. In ϕ6, we have evidence that the procapsid expands in the process, and that the expansion is related to a hinging movement at the interface of the subunits of the P1 dimer (Huiskonen et al., 2006a). The fact that ϕ8 PC has a similar quaternary structure suggests that it can accommodate a similar expansion. The assembly pathways of the eukaryotic dsRNA viruses are not known. It is possible, for example, that they do not proceed via an empty procapsid state, and thus there would be no need for an expansion mechanism. This might be reflected in their distinctly different interdigitating *T*=1 shells.

If we assume that the two different kinds of dsRNA virus quaternary structures (cystovirus *vs* reovirus) have evolved from a common ancestor, it becomes clear that the highly α-helical fold is much more amenable to change than the double β-barrel fold, all examples of which appear to be very much similar in their base. This may be because the barrel fold is more restricted by its three-dimensionality, whereas the dsRNA virus fold is more like a 2D object and free to fill the surface of the virus whichever way it pleases. However, the fact that the α-helical fold has only been found to make capsids with a *T*=1 lattice of dimers suggests that the fold is strictly constrained by the quaternary structure, and cannot make the jump to new triangulations.

The crystal structure of the STIV major coat protein (Khayat et al., 2005) showed that the double β-barrel fold is found also in an archaeal virus. It may be safe to say that the jury is still out when it comes to SH1 having the same fold, until we have crystal structure of the SH1 coat proteins. However, already the fact that there are two major coat proteins that form stable complexes (Bamford et al., 2005b)

suggests that the picture is more complicated than in the case of PRD1 for example. The type 2 capsomers of SH1, with two distal towers, cannot be homotrimers. The type 3 capsomers could, but the molecular weights of the coat proteins do not seem to support this idea, when compared to the capsomer mass estimate based on the reconstructed density. Therefore unlike in the case of STIV, where a relatively low-resolution reconstruction allowed a prediction of the major capsid protein fold that was verified by x-ray crystallography, we would have a difficult time trying to prove the presence of the double β-barrel fold with SH1. One can fit the PRD1 P3 into the structure of the type 3 capsomers (Figure 20), but on a closer examination it is not convincing as the position of the towers does not match. In addition, the modelled capsomers do not then pack together, but are spaced by the distance of the bridges seen between them at the surface of the capsid. In reality, the capsomers do actually interdigitate significantly.

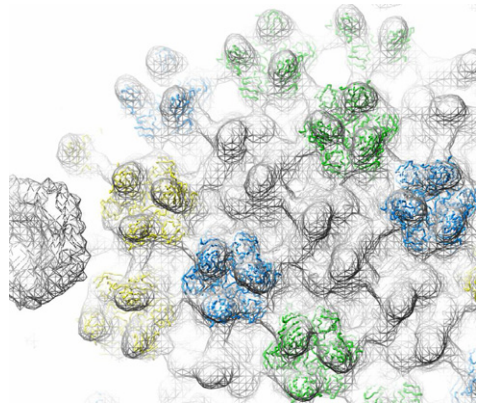


Figure 20. PRD1 P3 (PDB ID code 1hx6; (Abrescia et al., 2004)) fitted in SH1 capsid. Courtesy of J. Huiskonen.

3.2. Viruses, cells and the Way

Daoism, an ancient Chinese system of philosophy, considers that all things in the universe manifest according to the same universal principle. This way of manifesting is in Chinese called Dao, the Way, thus Daoism is the philosophy of the Way. From the point of view of human life, the Daoist ideal is to reach an understanding of the Way things unfold, so as to be able to always act in accordance, and never in opposition, to what naturally happens. However, as the same principal applies everywhere, the philosophy has also deeper philosophical and cosmological reaches. Here I would like to attempt to show how the philosophy corresponds with the "Virus World" theory of the emergence of life. Regrettably, my understanding of neither side of the comparison is very solid, so this attempt will necessarily be shallow and erroneous, but I still find the idea fascinating enough to give it a try.

According to the Daoist view, all phenomena can be viewed as the interplay of two opposing extremes: light and heavy, hard and soft, male and female, movement and stillness, existence and non-existence and so on. The two sides are termed the yin and the yang. Their relation is symbolized in the ubiquitous yin-yang symbol (Figure 21). The symbol shows that the extremes vary in a cyclical fashion: when one increases, the other decreases. It also shows that each of the extremes contains something of the other extreme. The symbol captures, however, only a part of the explanation of how things manifest. The whole story is summarized in a poem in the Daoist classic *Daodejing* by the sage Laozi (translations from (Davis, 2004)):

> *The Dao gave birth to one,*
> *one gave birth to two,*
> *two gave birth to three,*
> *and three gave birth to the myriad of things.*

The first two lines can also be stated as (attributed to Wang Zongyue):

Taiji is born from wuji; it is the mother of yin and yang.

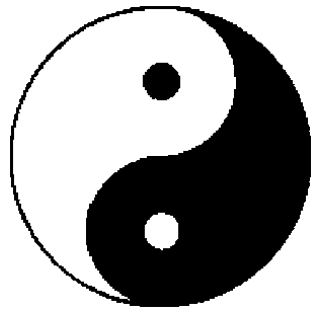


Figure 21. The taiji symbol.

Wuji (the Dao) is the original state, void of all phenomena, formless, also sometimes termed the primordial chaos. Taiji (the one) is an intermediate state: something manifested from wuji, but does not yet have any polarity. Taiji contains a potential of polarity, which when realized, gives rise to yin and yang (the two). From the interplay of the two, the third thing is formed, and from the three "a myriad of things" arise.

From this, the correspondence to the Viral World is straightforward. The primordial soup, with no biomolecules yet formed is wuji, the original state, devoid of form, the primordial chaos. The early, "virus-like" stage is taiji. There is no polarity between the parasite and the host, as all RNAs are parasitic, completely selfish. The polarity of yin and yang, of the host and the virus, arises with the emergence of relatively stable, co-operating ensembles of molecules, localized in some compartments. An RNA molecule now either belongs to the co-operative or does not belong, and those that do not, can assume a parasitic relation to it. These two can now give birth to three, the third being for example

something akin to an RNA-world equivalent of a plasmid, something not an integral part of the co-operative, but not hostile to it either. And from the interplay of these three, myriads of things then arise.

In summary, we come to the conclusion that viruses are the yang aspect of biological life. Light, small, dense, hard, masculine, warrior-like, these are attributes that are usually considered yang, all descriptive of viruses. Cells, in contrast, are clearly yin: heavy, large, loosely filled, soft, feminine, mother-like. Their cyclical interplay is nowadays masked by the myriad of interactions between the myriad of things that have manifested from the original interaction, but it can maybe still be seen almost in the original form in the oceans, in the interaction between the marine phages and their algal hosts. Continuously, massive numbers of marine microbes are infected and destroyed by phages, their constituent vital resources and energy reserves being converted into progeny phages and debris. Sooner or later, the debris and the phages will sink to the bottom of the sea, turn into sedimented nutrients that sooner or later will be released by *e.g.* volcanic activity. The released nutrients can then be taken advantage of by other cells, the yin side of life, with the capacity to accumulate and grow new life, only to be destroyed once more by viruses that lack this capacity. In addition, that yin always contains some yang and vice versa can here be exemplified by a lysogenic virus integrated into the host genome, and by a virus carrying genes acquired from the host. Both of these processes promote horizontal gene transfer, an obviously important factor in evolution. This also underlines the fact that the interaction between the virus and the host is not merely one of animosity, but at its root one of mutual interdependence. This is not an entirely new idea. That viruses depend on the hosts is clear, but it has also been previously stated that viruses are the single most powerful evolutionary driving force (Bubanovic et al., 2005), which points to very much the same thing.

Does this approach perhaps also offer something to answer the persistent question of whether viruses are alive or not? I think so. Viewed like this, the picture changes: we no longer focus only on the cellular, yin aspect of life. Viruses are clearly not that. But the yin is only one of the two. In fact, this way the question changes: we must now ask if our definition of biological life has been too narrow and if we should maybe accept viruses as the complementary aspect of cells in a broader definition. I think the reasoning is valid regardless of the validity of the Virus World hypothesis that inspired it: the relationship between cellular organisms and viruses as observed today fits the yin/yang description regardless of the exact way they evolved.

Having ventured this far out of my depth, I can find consolation in that “the Dao that can be spoken of is not the Dao”, *i.e.* even the best explanations by the best experts would anyway only be pointers to true understanding. A better way to approach that understanding may be to shut up and spend a minute in silence after this full stop.

F. FUTURE WORK

Crystallization trials of various cystoviral P1 proteins are underway. If successful, these will certainly shed light on the variation of structure within the virus family.

In the case of SH1, an interesting question is the distribution of the major coat proteins VP4 and VP7 in the capsid. Crystallization would probably solve this issue, but while waiting for the crystals to grow, a dissociation trick that hits only one of the two proteins might do the job. The complexity of the structure will require many complementary methods in order to understand it completely and to link it to the virus life cycle.

G. ACKNOWLEDGEMENTS

This thesis project was carried out at the Finnish Academy Centre of Excellence in Structural Virology / Virus Research (2000-2005 and 2006-2011), at the Institute of Biotechnology and the Department of Biological and Environmental Sciences, at the University of Helsinki, under the supervision of Dr Sarah Butcher.

First and foremost, I thank Sarah. Not only for her superb supervision and expertise, but also for her motivation, dedication and sheer hard work to see through the projects, exemplified recently by her unfathomable willingness to wake up before 6AM to read my thesis drafts. Her students are a truly lucky bunch!

I am much indebted to Dr Juha Huiskonen, who has been my mentor in all aspects of the work. Juha was quick to pioneer all the methods used in the projects, so for a slow person, there was much less learning to do. He also gave the invaluable demonstration that writing a thesis need not take too long a time.

My task of making the 3D reconstructions is really only the tip of the iceberg of the whole operation. I thank Professor Dennis Bamford, the head of the Centre of Excellence, for keeping the big wheels turning. More specifically, I thank Terhi Kemppinen, Dr Hanna Kivelä, Sari Korhonen, Benita Löflund, Dr Elina Roine and others for the wet lab part of the SH1 project, and Pasi Laurinmäki and Benita for taking care of microscopy. I thank Benita also for her wet lab work in the ϕ8 project. Joonatan Kaartinen, Teppo Kankaanpää and Risto Tetri took good care of our computers, many thanks for that.

My thanks also go to our international collaborators in the ϕ8 vertex reconstruction project, Dr John Briggs and Professor Stephen Fuller.

I thank Juha, Dr Denis Kainov, Dr Nelli Karhu, Dr Minni Koivunen, Pasi, Dr Jiří Lísal, Dr Janne Ravantti and Jani Seitsonen for their inspiring books, for showing me by example what kind of stuff goes into a thesis.

Many thanks go to Dr Maarit Suomalainen and Dr Roman Tuma for quickly but thoroughly reviewing the manuscript of this thesis.

I thank Professor Tapio Palva for well chosen licenciate exam questions and for managing all the paperwork related to my studies.

Many of my international conference travels during the project were funded by the University of Helsinki chancellor's travel grant, sincere thanks for that. Similarly I am thankful for having had the chance to participate in the domestic conferences organized and funded by the National Graduate School in Informational and Structural Biology.

I thank the kitchen staff at Kipsari, Siamintie 18, Tähkä, Unicafe restaurants in Viikki, and Viikin Kartano, for feeding me.

I thank our little group (Sarah, Juha, Pasi, Benita, Violeta Manole, Dr Ari Ora and Jani) for the group thing. And the big group, everybody at the CoE, for the big group thing.

I thank all my friends for their friendship, for making life more like life. Especially I thank Miisuliina Bombaluriina Valkotassu Sulo-Kusti Josefiina, a.k.a. Miisu, for her sane presence during the most intense period of writing.

I thank all my teachers and masters, all helpful people, for guiding me to this point in my life.

I thank my parents, Tuula and Heikki Jäälinoja, for the occasional bonus funding, the berries, the fish and everything.

H. REFERENCES

Aalto, A. P., Sarin, P., van Dijk, A. A., Saarma, M., Poranen, M. M., Arumäe, U., and Bamford, D. (in press). Large scale production of dsRNA and siRNA pools for RNA interference utilizing bacteriophage ϕ6 RNA-dependent RNA polymerase. RNA.

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

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A. W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32-36.

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K., and Miyano, M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure *7*, 1417-1426.

Amos, L. A., and Finch, J. T. (2004). Aaron Klug and the revolution in biomolecular structure determination. Trends Cell Biol *14*, 148-152.

Arnold, E., and Rossmann, M. G. (1988). The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. Acta Crystallogr A *44 (Pt 3)*, 270-282.

Arnold, H. P., Ziese, U., and Zillig, W. (2000a). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. Virology *272*, 409-416.

Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P., Nelson, K. E., and Fraser, C. M. (2000b). A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. Virology *267*, 252-266.

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

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

Baker, T. S., and Cheng, R. H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. J Struct Biol *116*, 120-130.

Baker, T. S., Olson, N. H., and Fuller, S. D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol Rev *63*, 862-922.

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

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

Bamford, D. H., Palva, E. T., and Lounatmaa, K. (1976). Ultrastructure and life cycle of the lipid-containing bacteriophage phi 6. J Gen Virol *32*, 249-259.

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

Bamford, D. H., Romantschuk, M., and Somerharju, P. J. (1987). Membrane fusion in prokaryotes: bacteriophage phi6 membrane fuses with the *Pseudomonas syringae* outer membrane. EMBO J *6*, 1467-1473.

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

Benson, S. D., Bamford, J. K., Bamford, D. H., and Burnett, R. M. (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 D Biol Crystallogr *58*, 39-59.

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

Bettstetter, M., Peng, X., Garrett, R. A., and Prangishvili, D. (2003). AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. Virology *315*, 68-79.

Bhella, D., Ralph, A., and Yeo, R. P. (2004). Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. J Mol Biol *340*, 319-331.

Bielski, R., and Tencer, M. (2006). A Possible Path to the RNA World: Enantioselective and Diastereoselective Purification of Ribose. Orig Life Evol Biosph.

Boier Martin, I. M., Marinescu, D. C., Lynch, R. E., and Baker, T. S. (1997). Identification of spherical virus particles in digitized images of entire electron micrographs. J Struct Biol *120*, 146-157.

Botstein, D., Waddell, C. H., and King, J. (1973). Mechanism of head assembly and DNA encapsulation in Salmonella phage p22. I. Genes, proteins, structures and DNA maturation. J Mol Biol *80*, 669-695.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R. L., Mathieu, M., De Francesco, R., and Rey, F. A. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc Natl Acad Sci U S A *96*, 13034-13039.

Briggs, J. A. G., Grunewald, K., Glass, B., Forster, F., Krausslich, H. G., and Fuller, S. D. (2006). The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. Structure *14*, 15-20.

Briggs, J. A. G., Huiskonen, J. T., Fernando, K. V., Gilbert, R. J., Scotti, P., Butcher, S. J., and Fuller, S. D. (2005). Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. J Struct Biol *150*, 332-339.

Brock, T. (1997). Biology of microorganisms, Prentice-Hall Inc.).

Bubanovic, I., Najman, S., and Andjelkovic, Z. (2005). Origin and evolution of viruses: escaped DNA/RNA sequences as evolutionary accelerators and natural biological weapons. Med Hypotheses *65*, 868-872.

Burnett, R. M. (1985). The structure of the adenovirus capsid. II. The packing symmetry of hexon and its implications for viral architecture. J Mol Biol *185*, 125-143.

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

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

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

Casjens, S., and King, J. (1974). P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct *2*, 202-224.

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

Cerritelli, M. E., Cheng, N., Rosenberg, A. H., McPherson, C. E., Booy, F. P., and Steven, A. C. (1997). Encapsidated conformation of bacteriophage T7 DNA. Cell *91*, 271-280.

Chacon, P., and Wriggers, W. (2002). Multi-resolution contour-based fitting of macromolecular structures. J Mol Biol *317*, 375-384.

Chang, J., Weigele, P., King, J., Chiu, W., and Jiang, W. (2006). Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure *14*, 1073-1082.

Chescoe, D., and Goodhew, P. J. (1990). The operation of transmission and scanning electron microscopes. (New York, Oxford University Press).

Clark, J. R., and March, J. B. (2006). Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials. Trends Biotechnol *24*, 212-218.

Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W., and Steven, A. C. (1995). Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J Mol Biol *253*, 86-99.

Coombs, K. M. (1996). Identification and characterization of a double-stranded RNA- reovirus temperature-sensitive mutant defective in minor core protein mu2. J Virol *70*, 4237-4245.

Coplin, D. L., Van Etten, J. L., Koski, R. K., and Vidaver, A. K. (1975). Intermediates in the biosynthesis of double-stranded ribonucleic acids of bacteriophage phi 6. Proc Natl Acad Sci U S A *72*, 849-853.

Crick, F. H., and Watson, J. D. (1956). Structure of small viruses. Nature *177*, 473-475.

Crowther, R. A. (1970). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc Roy Soc Lond *A*, 319-340.

Crowther, R. A. (1971). Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. Phil Trans R Soc Lond *B. 261*, 221-230.

Cusack, S. (2005). Adenovirus complex structures. Curr Opin Struct Biol *15*, 237-243.

Das, I., O'Connell, N., and Lambert, P. (2006). Epidemiology, clinical and laboratory characteristics of *Staphylococcus aureus* bacteraemia in a university hospital in UK. J Hosp Infect.

Davis, B. (2004). The taijiquan classics (Berkeley, North Atlantic Books).

Dawson, C., and Darrell, R. (1963). Infections due to adenovirus type 8 in the United States. I. An outbreak of epidemic keratoconjunctivitis originating in a physician's office. N Engl J Med *268*, 1031-1034.

de Haas, F., Paatero, A. O., Mindich, L., Bamford, D. H., and Fuller, S. D. (1999). A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage phi6. J Mol Biol *294*, 357-372.

De Rosier, D. J., and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. Nature *217*, 130-134.

Dokland, T., and Murialdo, H. (1993). Structural transitions during maturation of bacteriophage lambda capsids. J Mol Biol *233*, 682-694.

Dryden, K. A., Wang, G., Yeager, M., Nibert, M. L., Coombs, K. M., Furlong, D. B., Fields, B. N., and Baker, T. S. (1993). Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. J Cell Biol *122*, 1023-1041.

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Q Rev Biophys *21*, 129-228.

Dubochet, J., Groom, M., and Mueller-Neuteboom, S. (1982). The mounting of macromolecules for electron microscopy with particular reference to surface phenomena and the treatment of support films by glow discharge. In Advances in optical and electron microscopy, R. Barer, and V. E. Cosslett, eds. (London, Academic Press), pp. 107-135.

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

Dyall-Smith, M., Tang, S. L., and Bath, C. (2003). Haloarchaeal viruses: how diverse are they? Res Microbiol *154*, 309-313.

Earnshaw, W., and King, J. (1978). Structure of phage P22 coat protein aggregates formed in the absence of the scaffolding protein. J Mol Biol *126*, 721-747.

Emori, Y., Iba, H., and Okada, Y. (1982). Morphogenetic pathway of bacteriophage phi 6. A flow analysis of subviral and viral particles in infected cells. J Mol Biol *154*, 287-310.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberg, U., and Ball, L. A., eds. (2005). Eight Report of the International Committee on Taxonomy of Viruses (San Diego, Elsevier Academic Press).

Fokine, A., Leiman, P. G., Shneider, M. M., Ahvazi, B., Boeshans, K. M., Steven, A. C., Black, L. W., Mesyanzhinov, V. V., and Rossmann, M. G. (2005). Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. Proc Natl Acad Sci U S A *102*, 7163-7168.

Forterre, P. (2002). The origin of DNA genomes and DNA replication proteins. Curr Opin Microbiol *5*, 525-532.

Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie *87*, 793-803.

Forterre, P. (2006a). The origin of viruses and their possible roles in major evolutionary transitions. Virus Res *117*, 5-16.

Forterre, P. (2006b). Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc Natl Acad Sci U S A *103*, 3669-3674.

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

Frank, J. (1996). Three-dimensional electromicroscopy of macromolecular assemblies, Academic Press).

Fraser, R. D., Furlong, D. B., Trus, B. L., Nibert, M. L., Fields, B. N., and Steven, A. C. (1990). Molecular structure of the cell-attachment protein of reovirus: correlation of computer-processed electron micrographs with sequence-based predictions. J Virol *64*, 2990-3000.

Frilander, M., Gottlieb, P., Strassman, J., Bamford, D. H., and Mindich, L. (1992). Dependence of minus-strand synthesis on complete genomic packaging in the double-stranded RNA bacteriophage phi 6. J Virol *66*, 5013-5017.

Fuller, S. D., Butcher, S. J., Cheng, R. H., and Baker, T. S. (1996). Three-dimensional reconstruction of icosahedral particles--the uncommon line. J Struct Biol *116*, 48-55.

Gitlin, L., and Andino, R. (2003). Nucleic acid-based immune system: the antiviral potential of mammalian RNA silencing. J Virol *77*, 7159-7165.

Glaeser, R. M. (1999). Review: electron crystallography: present excitement, a nod to the past, anticipating the future. J Struct Biol *128*, 3-14.

Goldbach, R., Bucher, E., and Prins, M. (2003). Resistance mechanisms to plant viruses: an overview. Virus Res *92*, 207-212.

Gommers-Ampt, J. H., and Borst, P. (1995). Hypermodified bases in DNA. Faseb J *9*, 1034-1042.

Gottlieb, P., Strassman, J., and Mindich, L. (1992). Protein P4 of the bacteriophage phi 6 procapsid has a nucleoside triphosphate-binding site with associated nucleoside triphosphate phosphohydrolase activity. J Virol *66*, 6220-6222.

Gottlieb, P., Strassman, J., Qiao, X. Y., Frucht, A., and Mindich, L. (1990). In vitro replication, packaging, and transcription of the segmented double-stranded RNA genome of bacteriophage phi 6: studies with procapsids assembled from plasmid-encoded proteins. J Bacteriol *172*, 5774-5782.

Gouet, P., Diprose, J. M., Grimes, J. M., Malby, R., Burroughs, J. N., Zientara, S., Stuart, D. I., and Mertens, P. P. (1999). The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. Cell *97*, 481-490.

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

Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I., and Prasad, B. V. (1997). An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. Structure *5*, 885-893.

Grünewald, K., Desai, P., Winkler, D. C., Heymann, J. B., Belnap, D. M., Baumeister, W., and Steven, A. C. (2003). Three-dimensional structure of herpes simplex virus from cryo-electron tomography. Science *302*, 1396-1398.

Hantula, J., and Bamford, D. H. (1988). Chemical crosslinking of bacteriophage phi 6 nucleocapsid proteins. Virology *165*, 482-488.

Harauz, G., and van Heel, M. (1986). Similarity measures between images. Exact filters for general geometry of 3D reconstructions. Optik *73*, 146-156.

Haring, M., Rachel, R., Peng, X., Garrett, R. A., and Prangishvili, D. (2005a). Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. J Virol *79*, 9904-9911.

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

Harris, J. R., and Scheffler, D. (2002). Routine preparation of air-dried negatively stained and unstained specimens on holey carbon support films: a review of applications. Micron *33*, 461-480.

Harrison, S. (1990). Principles of virus structure. In Virology, B. N. Fields, ed. (New York, Raven Press, Ltd), pp. 37-61.

Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large bacteriophages. Proc Natl Acad Sci U S A *75*, 4779-4783.

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

Hendrix, R. W. (2003). Bacteriophage genomics. Curr Opin Microbiol *6*, 506-511.

Hewat, E. A., Booth, T. F., and Roy, P. (1992). Structure of bluetongue virus particles by cryoelectron microscopy. J Struct Biol *109*, 61-69.

Hewat, E. A., Booth, T. F., and Roy, P. (1994). Structure of correctly self-assembled bluetongue virus-like particles. J Struct Biol *112*, 183-191.

Heymann, J. B. (2001). Bsoft: image and molecular processing in electron microscopy. J Struct Biol *133*, 156-169.

Heymann, J. B., Cheng, N., Newcomb, W. W., Trus, B. L., Brown, J. C., and Steven, A. C. (2003). Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. Nat Struct Biol *10*, 334-341.

Hierholzer, J. C., and Pumarola, A. (1976). Antigenic characterization of intermediate adenovirus 14-11 strains associated with upper respiratory illness in a military camp. Infect Immun *13*, 354-359.

Hoogstraten, D., Qiao, X., Sun, Y., Hu, A., Onodera, S., and Mindich, L. (2000). Characterization of phi8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to phi6. Virology *272*, 218-224.

Hopper, P., Harrison, S. C., and Sauer, R. T. (1984). Structure of tomato bushy stunt virus. V. Coat protein sequence determination and its structural implications. J Mol Biol *177*, 701-713.

Horne, R. W., and Ronchetti, I. P. (1974). A negative staining-carbon film technique for studying viruses in the electron microscope. I. Preparative procedures for examining icosahedral and filamentous viruses. J Ultrastruct Res *47*, 361-383.

Huang, C. C., Couch, G. S., Pettersen, E. F., and Ferrin, T. E. (1996). Chimera: an extensible molecular modeling application constructed using standard components. Pacific Symposium on Biocomputing *1*, 724.

Huiskonen, J. T., de Haas, F., Bubeck, D., Bamford, D. H., Fuller, S. D., and Butcher, S. J. (2006a). Structure of the bacteriophage ϕ6 nucleocapsid suggests a mechanism for sequential RNA packaging. Structure *14*, 1039-1048.

Huiskonen, J. T., Jäälinoja, H. T., Briggs, J. A., Fuller, S. D., and Butcher, S. J. (2006b). Structure of a hexameric RNA packaging motor in a viral polymerase complex. Journal of Structural Biology.

Huiskonen, J. T., Kivelä, H. M., Bamford, D. H., and Butcher, S. J. (2004). The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. Nat Struct Mol Biol *11*, 850-856.

Ikonen, T., Kainov, D. E., Timmins, P., Serimaa, R., and Tuma, R. (2003). Locating the minor components of double-stranded RNA bacteriophage ϕ6 by neutron scattering. J Appl Cryst, 525-529.

Jacobs, B. L., and Langland, J. O. (1996). When two strands are better than one: the mediators and modulators of the cellular responses to double-stranded RNA. Virology *219*, 339-349.

Ji, Y., Marinescu, D. C., Zhang, W., Zhang, X., Yan, X., and Baker, T. S. (2006). A model-based parallel origin and orientation refinement algorithm for cryoTEM and its application to the study of virus structures. J Struct Biol *154*, 1-19.

Jiang, W., Baker, M. L., Ludtke, S. J., and Chiu, W. (2001a). Bridging the information gap: computational tools for intermediate resolution structure interpretation. J Mol Biol *308*, 1033-1044.

Jiang, W., Chang, J., Jakana, J., Weigele, P., King, J., and Chiu, W. (2006). Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature *439*, 612-616.

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

Jiang, W., Li, Z., Zhang, Z., Booth, C. R., Baker, M. L., and Chiu, W. (2001b). Semi-automated icosahedral particle reconstruction at sub-nanometer resolution. J Struct Biol *136*, 214-225.

Johansson, M. E., Uhnoo, I., Kidd, A. H., Madeley, C. R., and Wadell, G. (1980). Direct identification of enteric adenovirus, a candidate new serotype, associated with infantile gastroenteritis. J Clin Microbiol *12*, 95-100.

Johnson, M. D., 3rd, and Mindich, L. (1994). Isolation and characterization of nonsense mutations in gene 10 of bacteriophage phi 6. J Virol *68*, 2331-2338.

Juuti, J. T., and Bamford, D. H. (1997). Protein P7 of phage phi6 RNA polymerase complex, acquiring of RNA packaging activity by in vitro assembly of the purified protein onto deficient particles. J Mol Biol *266*, 891-900.

Juuti, J. T., Bamford, D. H., Tuma, R., and Thomas, G. J., Jr. (1998). Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage phi 6. J Mol Biol *279*, 347-359.

Kainov, D. E., Butcher, S. J., Bamford, D. H., and Tuma, R. (2003a). Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. J Mol Biol *328*, 791-804.

Kainov, D. E., Lisal, J., Bamford, D. H., and Tuma, R. (2004). Packaging motor from double-stranded RNA bacteriophage phi12 acts as an obligatory passive conduit during transcription. Nucleic Acids Res *32*, 3515-3521.

Kainov, D. E., Pirttimaa, M., Tuma, R., Butcher, S. J., Thomas, G. J., Jr., Bamford, D. H., and Makeyev, E. V. (2003b). RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. J Biol Chem *278*, 48084-48091.

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

King, J., Lenk, E. V., and Botstein, D. (1973). Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J Mol Biol *80*, 697-731.

Kivelä, H. M., Roine, E., Kukkaro, P., Laurinavicius, S., Somerharju, P., and Bamford, D. H. (2006). Quantitative dissociation of archaeal virus SH1 reveals distinct capsid proteins and a lipid core. Virology.

Kivioja, T., Ravantti, J., Verkhovsky, A., Ukkonen, E., and Bamford, D. (2000). Local average intensity-based method for identifying spherical particles in electron micrographs. J Struct Biol *131*, 126-134.

Klug, A., and Caspar, D. L. (1960). The structure of small viruses. Adv Virus Res *7*, 225-325.

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2006). The ancient Virus World and evolution of cells. Biology Direct *1*.

Kuo, I. A., and Glaeser, R. M. (1975). Development of methodology for low exposure, high resolution electron microscopy of biological specimens. Ultramicroscopy *1*, 53-66.

Kuo, T. T., Huang, T. C., and Teng, M. H. (1968). 5-Methylcytosine replacing cytosine in the deoxyribonucleic acid of a bacteriophage for Xanthomonas oryzae. J Mol Biol *34*, 373-375.

Larson, E. T., Reiter, D., Young, M., and Lawrence, C. M. (2006). Structure of A197 from Sulfolobus turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. J Virol *80*, 7636-7644.

Laurinavicius, S., Kakela, R., Bamford, D. H., and Somerharju, P. (2004). The origin of phospholipids of the enveloped bacteriophage phi6. Virology *326*, 182-190.

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

Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F., and Weber, P. C. (1999). Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat Struct Biol *6*, 937-943.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L., and Harrison, S. C. (1991). Structure of simian virus 40 at 3.8-Å resolution. Nature *354*, 278-284.

Liemann, S., Chandran, K., Baker, T. S., Nibert, M. L., and Harrison, S. C. (2002). Structure of the reovirus membrane-penetration protein, Mu1, in a complex with is protector protein, Sigma3. Cell *108*, 283-295.

Lu, G., Zhou, Z. H., Baker, M. L., Jakana, J., Cai, D., Wei, X., Chen, S., Gu, X., and Chiu, W. (1998). Structure of double-shelled rice dwarf virus. J Virol *72*, 8541-8549.

Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82-97.

Maaty, W. S., Ortmann, A. C., Dlakic, M., Schulstad, K., Hilmer, J. K., Liepold, L., Weidenheft, B., Khayat, R., Douglas, T., Young, M. J., and Bothner, B. (2006). Characterization of the archaeal thermophile Sulfolobus turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. J Virol *80*, 7625-7635.

Makeyev, E. V., and Bamford, D. H. (2000). Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage phi6. Embo J *19*, 124-133.

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

Marinescu, D. C., Ji, Y., and Lynch, R. E. (2001). Space-time tradeoffs for parallel 3D reconstruction algorithms for atomic virus structure determination. Concurrency Computat: Pract Exper *13*, 1083-1106.

Mertens, P. P., and Diprose, J. (2004). The bluetongue virus core: a nano-scale transcription machine. Virus Res *101*, 29-43.

Metcalf, P., Cyrklaff, M., and Adrian, M. (1991). The three-dimensional structure of reovirus obtained by cryo-electron microscopy. EMBO J *10*, 3129-3136.

Mindell, J. A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol *142*, 334-347.

Mindich, L. (2004). Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. Virus Res *101*, 83-92.

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

Modis, Y., Trus, B. L., and Harrison, S. C. (2002). Atomic model of the papillomavirus capsid. Embo J *21*, 4754-4762.

Morais, M. C., Kanamaru, S., Badasso, M. O., Koti, J. S., Owen, B. A., McMurray, C. T., Anderson, D. L., and Rossmann, M. G. (2003). Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. Nat Struct Biol *10*, 572-576.

Morais, M. C., Tao, Y., Olson, N. H., Grimes, S., Jardine, P. J., Anderson, D. L., Baker, T. S., and Rossmann, M. G. (2001). Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29. J Struct Biol *135*, 38-46.

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

Nakagawa, A., Miyazaki, N., Taka, J., Naitow, H., Ogawa, A., Fujimoto, Z., Mizuno, H., Higashi, T., Watanabe, Y., Omura, T.*, et al.* (2003). The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. Structure *11*, 1227-1238.

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

Nason, E. L., Rothagel, R., Mukherjee, S. K., Kar, A. K., Forzan, M., Prasad, B. V., and Roy, P. (2004). Interactions between the inner and outer capsids of bluetongue virus. J Virol *78*, 8059-8067.

Olkkonen, V. M., Ojala, P. M., and Bamford, D. H. (1991). Generation of infectious nucleocapsids by in vitro assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage phi 6. J Mol Biol *218*, 569-581.

Palese, P. (2004). Influenza: old and new threats. Nat Med *10*, S82-87.

Parker, M. H., Casjens, S., and Prevelige, P. E., Jr. (1998). Functional domains of bacteriophage P22 scaffolding protein. J Mol Biol *281*, 69-79.

Parker, M. H., Stafford, W. F., 3rd, and Prevelige, P. E., Jr. (1997). Bacteriophage P22 scaffolding protein forms oligomers in solution. J Mol Biol *268*, 655-665.

Pitman, A. R., Jackson, R. W., Mansfield, J. W., Kaitell, V., Thwaites, R., and Arnold, D. L. (2005). Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. Curr Biol *15*, 2230-2235.

Poole, A., Penny, D., and Sjoberg, B. (2000). Methyl-RNA: an evolutionary bridge between RNA and DNA? Chem Biol *7*, R207-216.

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

Poranen, M. M., and Tuma, R. (2004). Self-assembly of double-stranded RNA bacteriophages. Virus Res *101*, 93-100.

Porter, K., Kukkaro, P., Bamford, J. K. H., Bath, C., and Dyall-Smith, M. L. (2004). SH1: A novel , spherical halovirus isolated from an australian hypersaline lake.

Prage, L., Pettersson, U., Hoglund, S., Lonberg-Holm, K., and Philipson, L. (1970). Structural proteins of adenoviruses. IV. Sequential degradation of the adenovirus type 2 virion. Virology *42*, 341-358.

Prangishvili, D., Forterre, P., and Garrett, R. A. (2006a). Viruses of the Archaea: a unifying view. Nat Rev Microbiol *4*, 837-848.

Prangishvili, D., Vestergaard, G., Haring, M., Aramayo, R., Basta, T., Rachel, R., and Garrett, R. A. (2006b). Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J Mol Biol *359*, 1203-1216.

Prasad, B. V., Prevelige, P. E., Marietta, E., Chen, R. O., Thomas, D., King, J., and Chiu, W. (1993). Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J Mol Biol *231*, 65-74.

Prevelige, P. E., Jr., and King, J. (1993). Assembly of bacteriophage P22: a model for ds-DNA virus assembly. Prog Med Virol *40*, 206-221.

Prevelige, P. E., Jr., Thomas, D., and King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. J Mol Biol *202*, 743-757.

Ram, S., Ward, E. S., and Ober, R. J. (2006). Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. Proc Natl Acad Sci U S A *103*, 4457-4462.

Rayment, I., Baker, T. S., Caspar, D. L., and Murakami, W. T. (1982). Polyoma virus capsid structure at 22.5 A resolution. Nature *295*, 110-115.

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

Reiter, W. D., Zillig, W., and Palm, P. (1988). Archaebacterial viruses. Adv Virus Res *34*, 143-188.

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

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T., and Young, M. (2004). The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc Natl Acad Sci U S A *101*, 7716-7720.

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

Rode, B. M. (1999). Peptides and the origin of life. Peptides *20*, 773-786.

Romantschuk, M., and Bamford, D. H. (1985). Function of pili in bacteriophage phi 6 penetration. J Gen Virol *66 (Pt 11)*, 2461-2469.

Romantschuk, M., Olkkonen, V. M., and Bamford, D. H. (1988). The nucleocapsid of bacteriophage phi6 penetrates the host cytoplasmic membrane. EMBO J *7*, 1821-1829.

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

Rux, J. J., Kuser, P. R., and Burnett, R. M. (2003). Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution x-ray crystallographic, molecular modeling, and sequence-based methods. J Virol *77*, 9553-9566.

Salgado, P. S., Makeyev, E. V., Butcher, S. J., Bamford, D. H., Stuart, D. I., and Grimes, J. M. (2004). The structural basis for RNA specificity and Ca^{2+} inhibition of an RNA-dependent RNA polymerase. Structure *12*, 307-316.

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem *166*, 368-379.

Schnabel, H., Zillig, W., Pfaffle, M., Schnabel, R., Michel, H., and Delius, H. (1982). *Halobacterium halobium* phage ϕH. EMBO J *1*, 87-92.

Shapiro, R. (1988). Prebiotic ribose synthesis: a critical analysis. Orig Life Evol Biosph *18*, 71-85.

Simpson, A. A., Chipman, P. R., Baker, T. S., Tijssen, P., and Rossmann, M. G. (1998). The structure of an insect parvovirus (Galleria mellonella densovirus) at 3.7 Å resolution. Structure *6*, 1355-1367.

Simpson, A. A., Nandhagopal, N., Van Etten, J. L., and Rossmann, M. G. (2003). Structural analyses of Phycodnaviridae and Iridoviridae. Acta Crystallogr D Biol Crystallogr *59*, 2053-2059.

Stagg, S. M., Lander, G. C., Pulokas, J., Fellmann, D., Cheng, A., Quispe, J. D., Mallick, S. P., Avila, R. M., Carragher, B., and Potter, C. S. (2006). Automated cryoEM data acquisition and analysis of 284742 particles of GroEL. J Struct Biol *155*, 470-481.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., and Huber, R. (1997). Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J Mol Biol *267*, 865-880.

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

Stitt, B. L., and Mindich, L. (1983). The structure of bacteriophage phi 6: protease digestion of phi 6 virions. Virology *127*, 459-462.

Strauss, H., and King, J. (1984). Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J Mol Biol *172*, 523-543.

Sulakvelidze, A., and Kutter, E. (2004). Bacteriophage therapy in humans. In Bacteriophages (CRC).

Suzuki, N., Kusano, T., Matsuura, Y., and Omura, T. (1996). Novel NTP binding property of rice dwarf phytoreovirus minor core protein P5. Virology *219*, 471-474.

Suzuki, N., Tanimura, M., Watanabe, Y., Kusano, T., Kitagawa, Y., Suda, N., Kudo, H., Uyeda, I., and Shikata, E. (1992). Molecular analysis of rice dwarf phytoreovirus segment S1: interviral homology of the putative RNA-dependent RNA polymerase between plant- and animal-infecting reoviruses. Virology *190*, 240-247.

Tang, L., Marion, W. R., Cingolani, G., Prevelige, P. E., and Johnson, J. E. (2005). Three-dimensional structure of the bacteriophage P22 tail machine. EMBO J *24*, 2087-2095.

Tao, Y., Farsetta, D. L., Nibert, M. L., and Harrison, S. C. (2002). RNA synthesis in a cage--structural studies of reovirus polymerase lambda3. Cell *111*, 733-745.

Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G., and Baker, T. S. (1998). Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell *95*, 431-437.

Thon, F. (1971). Phase contrast electron microscopy. In Electron microscopy in material science, U. Valdre, ed. (New York, Academic Press), pp. 570-625.

Thuman-Commike, P. A., and Chiu, W. (1996). PTOOL: a software package for the selection of particles from electron cryomicroscopy spot-scan images. J Struct Biol *116*, 41-47.

Thuman-Commike, P. A., Greene, B., Jakana, J., Prasad, B. V., King, J., Prevelige, P. E., Jr., and Chiu, W. (1996). Three-dimensional structure of scaffolding-containing phage p22 procapsids by electron cryo-microscopy. J Mol Biol *260*, 85-98.

Thuman-Commike, P. A., Greene, B., Malinski, J. A., Burbea, M., McGough, A., Chiu, W., and Prevelige, P. E., Jr. (1999). Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. Biophys J *76*, 3267-3277.

Thuman-Commike, P. A., Greene, B., Malinski, J. A., King, J., and Chiu, W. (1998). Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. Biophys J *74*, 559-568.

Tomita, F., and Takahashi, I. (1975). DNase specific for uracil-containing bacteriophage DNA. J Virol *15*, 1081-1087.

Torrance, L., Andreev, I. A., Gabrenaite-Verhovskaya, R., Cowan, G., Makinen, K., and Taliansky, M. E. (2006). An unusual structure at one end of potato potyvirus particles. J Mol Biol *357*, 1-8.

Toyoshima, C., and Unwin, N. (1988). Contrast transfer for frozen-hydrated specimens: determination from pairs of defocused images. Ultramicroscopy *25*, 279-291.

Twarock, R. (2004). A tiling approach to virus capsid assembly explaining a structural puzzle in virology. J Theor Biol *226*, 477-482.

Ueda, S., Masuta, C., and Uyeda, I. (1997). Hypothesis on particle structure and assembly of rice dwarf phytoreovirus: interactions among multiple structural proteins. J Gen Virol *78 (Pt 12)*, 3135-3140.

van Etten, J., Lane, L., Gonzalez, C., Partridge, J., and Vidaver, A. (1976). Comparative properties of bacteriophage phi6 and phi6 nucleocapsid. J Virol *18*, 652-658.

Van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111-123.

van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J Struct Biol *116*, 17-24.

van Raaij, M. J., Louis, N., Chroboczek, J., and Cusack, S. (1999a). Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. Virology *262*, 333-343.

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

Wang, C., Eufemi, M., Turano, C., and Giartosio, A. (1996). Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry *35*, 7299-7307.

Watson, J. D. (1954). The structure of tobacco mosaic virus. I. X-ray evidence of a helical arrangement of sub-units around the longitudinal axis. Biochim Biophys Acta *13*, 10-19.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. (2004). Molecular biology of the gene, 5 edn (San Francisco, Jim Smith).

Watson, J. D., and Crick, F. H. (1953). The structure of DNA. Cold Spring Harb Symp Quant Biol *18*, 123-131.

Weiner, A. M., and Maizels, N. (1987). tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. Proc Natl Acad Sci U S A *84*, 7383-7387.

Wery, J. P., Reddy, V. S., Hosur, M. V., and Johnson, J. E. (1994). The refined three-dimensional structure of an insect virus at 2.8 Å resolution. J Mol Biol *235*, 565-586.

Vidaver, A. K., Koski, R. K., and Van Etten, J. L. (1973). Bacteriophage phi6: a lipid-containing virus of *Pseudomonas phaseolicola*. J Virol *11*, 799-805.

Wikoff, W. R., Conway, J. F., Tang, J., Lee, K. K., Gan, L., Cheng, N., Duda, R. L., Hendrix, R. W., Steven, A. C., and Johnson, J. E. (2006). Time-resolved molecular dynamics of bacteriophage HK97

capsid maturation interpreted by electron cryo-microscopy and X-ray crystallography. J Struct Biol *153*, 300-306.

Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. Science *289*, 2129-2133.

Wilhelm, S. W., Brigden, S. M., and Suttle, C. A. (2002). A dilution technique for the direct measurement of viral production: a comparison in stratified and tidally mixed coastal waters. Microb Ecol *43*, 168-173.

Williamson, K. E., Radosevich, M., and Wommack, K. E. (2005). Abundance and diversity of viruses in six Delaware soils. Appl Environ Microbiol *71*, 3119-3125.

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

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

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

Wrigley, N. G. (1969). An electron microscope study of the structure of *Sericesthis* iridescent virus. Journal of General Virology, 123-134.

Xiao, C., Chipman, P. R., Battisti, A. J., Bowman, V. D., Renesto, P., Raoult, D., and Rossmann, M. G. (2005). Cryo-electron microscopy of the giant Mimivirus. J Mol Biol *353*, 493-496.

Yan, J., Tomaru, M., Takahashi, A., Kimura, I., Hibino, H., and Omura, T. (1996). P2 protein encoded by genome segment S2 of rice dwarf phytoreovirus is essential for virus infection. Virology *224*, 539-541.

Yan, X., Chipman, P. R., Castberg, T., Bratbak, G., and Baker, T. S. (2005). The marine algal virus PpV01 has an icosahedral capsid with T=219 quasisymmetry. J Virol *79*, 9236-9243.

Yan, X., Olson, N. H., Van Etten, J. L., Bergoin, M., Rossmann, M. G., and Baker, T. S. (2000). Structure and assembly of large lipid-containing dsDNA viruses. Nat Struct Biol *7*, 101-103.

Yang, H., Makeyev, E. V., Butcher, S. J., Gaidelyte, A., and Bamford, D. H. (2003). Two distinct mechanisms ensure transcriptional polarity in double-stranded RNA bacteriophages. J Virol *77*, 1195-1203.

Zanetti, G., Briggs, J. A., Grunewald, K., Sattentau, Q. J., and Fuller, S. D. (2006). Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. PLoS Pathog *2*, e83.

Zhang, Y., Kostyuchenko, V. A., and Rossmann, M. G. (2006). Structural analysis of viral nucleocapsids by subtraction of partial projections. J Struct Biol.

Zhang, Z., Greene, B., Thuman-Commike, P. A., Jakana, J., Prevelige, P. E., Jr., King, J., and Chiu, W. (2000). Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. J Mol Biol *297*, 615-626.

Zhou, Z. H., Dougherty, M., Jakana, J., He, J., Rixon, F. J., and Chiu, W. (2000). Seeing the herpesvirus capsid at 8.5 Å. Science *288*, 877-880.

Zhu, P., Liu, J., Bess, J., Jr., Chertova, E., Lifson, J. D., Grise, H., Ofek, G. A., Taylor, K. A., and Roux, K. H. (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature *441*, 847-852.

Zillig, W., Klenk, H. P., Palm, P., Puhler, G., Gropp, F., Garrett, R. A., and Leffers, H. (1989). The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. Can J Microbiol *35*, 73-80.

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