

Structure of viral membrane-penetrating machines by electron cryo-microscopy and tomography

THÈSE N° 6507 (2015)

PRÉSENTÉE LE 13 MARS 2015

À LA FACULTÉ DES SCIENCES DE BASE

LABORATOIRE DE BIOPHYSIQUE ET BIOLOGIE STRUCTURELLE

PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Sergey NAZAROV

acceptée sur proposition du jury:

Prof. C. Hébert, présidente du jury

Prof. P. Leiman, directeur de thèse

Dr T. Ishikawa, rapporteur

Dr G. Knott, rapporteur

Prof. H. Stahlberg, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Prof. Petr Leiman for the opportunity to go deep into science, and to work in his laboratory in the field of biophysics and structural biology. It is still unclear for me whether the research is more physical or biological. I have the opportunity to contribute to very interesting interdisciplinary projects. One day I could find myself purifying the new bacteriophage sample and collecting tons of images in the dark room with our F20 FEI microscope, while another days were a real “hackathon” with a huge amount of image processing softwares trying to adapt them to our needs. I would like to thank Petr for his continuous support and guidance, and innumerable amount of time he spent with us sharing his experience in science and in life.

I also would like to thank Petr for the opportunity to attend a number of international conferences and workshops, like PVA and FASEB meetings, GRC 3D Cryo-electron microscopy meeting, Regio meetings in crystallography.

I would like to thank my thesis jury members, Prof. Cécile Hebert, Prof. Henning Stahlberg, Dr. Takashi Ishikawa, and Dr. Graham Knott, for the opportunity to present my research to them.

I am very grateful to Dr. Davide Demurtas from CIME, and Dr. Mikhail Shneider. Davide introduced me into the world of practical TEM and, specifically, into the cryo-EM data collection. Mikhail has isolated astronomic amount of bacteriophages, and all of them we were able to inspect with TEM. I would like to thank him for all support related to phages propagation and purification, as well as for the scientific discussions.

I am thankful to other present and former lab members, Dr. Chris Browning, Dr. Lada Sycheva, Dr. Sergii Buth, Dr. Nicholas Taylor, Michel Plattner and Dr. Ricardo Guerrero-Ferreira.

And I want to thank my wife, my two little kids, and my parents for the constant support and atmosphere they created for these 4 years.

ABSTRACT

My PhD thesis aims to obtain the structure of several bacteriophages, including so-called “jumbo” phages, using the electron cryo-microscopy (cryo-EM) and to study the structural transformation of these viruses as they attach and infect the host cell. We aim to understand the process of host cell membrane penetration at the molecular level and to achieve the level of knowledge necessary to understand how the genome and proteins, which are originally packaged into the capsid, are delivered into the host cell. This structural information can be of great importance since bacteriophage therapy is a potentially promising alternative to antibiotics, which are urgently needed.

Bacteriophages, or phages are the most abundant and diverse form of life on Earth, and have developed various strategies through which to infect a susceptible bacterial host. A vast majority of phages have evolved to use a special organelle, called a “tail”, for host recognition, attachment and genome delivery into the cell. The host cell envelope is penetrated with the help of the tail. Unlike most eukaryotic viruses, infection of a host by tailed bacteriophages usually requires only one virion per bacterium, indicating that tailed bacteriophages have evolved an extremely efficient infection mechanism.

Bacteriophages with contractile tails are complex viruses that infect *Bacteria* and *Archaea*. They share a common evolutionary origin and some functional aspect with other large macromolecular machines, such as the *Serratia entomophila* antifeeding prophage, the *Photorhabdus* virulence cassette, and the type VI secretion system (T6SS), which function to attack other cells by translocating toxic effector proteins into the target cell's cytoplasm. The T6SS and the so-called “jumbo” phages are the most complex of these contractile injection systems. Jumbo phages have genomes exceeding 250kb and virus particle sizes of 300 nm and greater. Hundreds of different proteins are involved in the morphogenesis and assembly of a jumbo phage particle.

A number of bacteriophages infecting important pathogens were isolated and characterized by means of cryo-EM and bioinformatics tools.

Bacteriophage $\phi 92$ is a virus infecting a wide variety of acapsular *Escherichia coli* laboratory strains, *Salmonella* strains, and pathogenic *E.Coli* strains carrying a polysialic acid, which cause severe invasive infections including meningitis, pneumonia, septicaemia, osteomyelitis, septic arthritis and pyelonephritis. A cryo-EM reconstructions of the $\phi 92$ capsid, baseplate and long contractile tail were calculated, and complete three-dimensional model of the phage virion was built.

Two bacteriophages, ϕ Eco32 infecting mastitis-causing strains of *E.Coli* and 7-11 infecting *Salmonella* strains, were isolated. They are a members of the *Podoviridae* family with the rare C3

morphotype, which occurs in <1% of phage virions. Two functional states of mature ϕ Eco32 virion, before and after DNA ejection, and mature 7-11 tail were characterized by cryo-EM and cryo-ET.

Bacteriophage AR9 is a flagella-specific *Bacillus subtilis* jumbo phage with a tail of remarkable complexity. The AR9 adsorption apparatus could be used for attachment and “riding on” the motile *B. subtilis* flagellum. A three-dimensional model of the jumbo virion was built and flagella-specific adsorption apparatus was characterized.

Bacterial phage CBA120 belongs to the *Myoviridae* family distinct from T4 in morphology and behavior and specifically infects *E.Coli* strain O157, an important food-borne pathogen. CBA120 represents a newly described group of phages. Extended and contracted conformations of CBA120 tail were characterized by cryo-EM.

Mutant bacteriophage T4 lacking the tip protein of the cell-puncturing device responsible for the first interaction with the host cell membrane was isolated. Three-dimensional cryo-EM reconstruction of the baseplate and a tail fragment was calculated and compared with that of the wild type T4 phage and the distant member of T4-superfamily phage RB43.

Keywords: bacteriophage, jumbo phage, contractile injection systems, ϕ 92, ϕ Eco32, AR9, CBA120, T4, RB43, cell-puncturing device, cryo-EM.

RESUME

L'objectif de mon doctorat est de déterminer la structure de bactériophages, notamment ceux de type "jumbo", en employant la cryo-microscopie électronique. Par ailleurs ma thèse consiste aussi à étudier les transformations de la structure de ces virus lors de l'infection de leur cellule hôte. Nous aimerions comprendre à l'échelle moléculaire, comment la membrane de la cellule hôte est pénétrée et déterminer comment le génome et les protéines du bactériophage lui sont injectés. Les bactériophages offrant potentiellement une alternative aux antibiotiques, ces connaissances sur leur structure peuvent se révéler être d'une grande importance pour de futures applications thérapeutiques.

Les bactériophages ou phages, sont la forme de vie la plus abondante et diverse sur Terre, et ont développé une multitude de stratégies leur permettant d'infecter leurs bactéries hôtes. La vaste majorité des phages a acquis une organelle appelée "queue", leur servant à reconnaître et à s'attacher à la cellule hôte pour y injecter leur génome. La membrane de la cellule hôte est donc pénétrée grâce à cette queue. Contrairement à la plupart des virus eucaryotes, l'infection par les bactériophages à queue ne nécessite généralement qu'un seul phage par bactérie, indiquant donc que ces bactériophages ont évolué pour acquérir un mécanisme extrêmement efficace d'infection des bactéries.

Les bactériophages à queue contractile sont des virus complexes capables d'infecter les bactéries et les archées. Ils partagent une origine évolutionnaire commune et certains aspects fonctionnels avec d'autres machines macromoléculaires tels que la *Serratia entomophila* antifeeding prophage, la cassette *Photorhabdus* de virulence et le système de sécrétion de type VI (SST6), qui attaque les cellules par translocation des protéines toxiques dans le cytoplasme de la cellule cible. Le T6SS et les phages "jumbo" sont les plus complexes de ces systèmes d'injection contractile. Les phages "jumbo" ont un génome excédant 250kb et ces virus ont une taille de plus de 300nm. Des centaines de protéines différentes sont impliquées dans la morphogénèse et l'assemblage des phages "jumbo". Une partie de ces bactériophages infectant d'importants pathogènes a pu être isolée et caractérisée à l'aide de la cryo-microscopie électronique et d'outils de bio-informatique.

Le bactériophage $\phi 92$ est un virus infectant une grande variété de bactéries *E.Coli* acapsulaires, des souches de salmonelle, et des souches d'*E.Coli* pathogènes transportant un acide polysialique qui provoquent de graves infections notamment la méningite, la pneumonie, la septicémie, l'ostéomyélite, l'arthrite septique et la pyélonéphrite. Des reconstructions cryo-EM du capsid, de la plaque de base et de la queue contractile du $\phi 92$ ont été calculés et un modèle 3D du phage a été construit.

Deux bactériophages ont été isolés: $\phi Eco92$ qui infecte la souche *E.Coli* responsables de la

mammite et 7-11 qui infecte la *Salmonelle*. Il existe des membres de la famille des Podoviridae qui possèdent le rare morphotype C3 qui apparait chez <1% des virions phagiques. Deux états fonctionnels du virion ϕ Eco32 mature, avant et après l'éjection de l'ADN et la queue mature du 7-11 ont été caractérisés par cryo-EM et cryo-ET.

Le bactériophage AR9 est un *Bacillus subtilis* jumbo phage, flagelle-spécifique, disposant d'une queue d'une complexité remarquable. L'appareil d'adsorption de l'AR9 pourrait être utilisé pour la fixation et le "transport" sur le motile *B. subtilis* flagellum. Un modèle 3D du virion jumbo a été construit et l'appareil d'adsorption flagelle-spécifique a été caractérisé.

Le phage bactérien CBA120 appartient à la famille *Myoviridae* distincte du T4 par sa morphologie et son comportement, et infecte spécifiquement la souche O157 d'*E.Coli*, un important pathogène d'origine alimentaire. Le CBA120 représente un nouveau groupe de phages récemment étudié. Les conformations contractés et étendus du CBA120 ont été caractérisées par cryo-EM.

Un bactériophage T4 mutant, ne disposant pas de la protéine pointe lui servant à percer la membrane cellulaire, a été isolé. Des reconstructions 3D cryo-EM de la plaque de base et de la queue contractile ont été calculées et comparées avec celles d'un phage T4 de type sauvage et de type RB43, un membre distant de la superfamille-T4.

Mots-clés: système de sécrétion de type VI, SST6, bactériophage, ϕ 92, ϕ Eco32, AR9, CBA120, T4, RB43, systèmes d'injection contractiles, cryo-EM.

Table of Contents

ACKNOWLEDGEMENTS.....	2
ABSTRACT.....	3
RESUME.....	5
1.INTRODUCTION.....	9
1.1.Prokaryotic viruses.....	9
1.2.Electron cryo-microscopy.....	25
2.BACTERIOPHAGES ΦECO32 AND 7-11.....	49
2.1.Summary.....	49
2.2.Introduction.....	49
2.3.Results and discussion.....	52
2.4.Concluding remarks.....	68
2.5.Materials and methods.....	69
3.BACTERIOPHAGE Φ92.....	73
3.1.Introduction.....	73
3.2.Cryo-EM reconstruction of the ϕ 92 virion.....	74
3.3.Concluding remarks.....	81
3.4.Materials and methods.....	81
4.BACTERIOPHAGE CBA120.....	89
4.1.Summary.....	89
4.2.Introduction.....	89
4.3.Results and discussion.....	90
4.4.Concluding remarks.....	96
4.5.Materials and methods.....	97
5.BACTERIOPHAGE AR9.....	105
5.1.Summary.....	105
5.2.Introduction.....	105
5.3.Results and discussion.....	112
5.4.Concluding remarks.....	116
5.5.Materials and methods.....	116
6.BACTERIOPHAGES T4 AND RB43.....	120
6.1.Summary.....	120
6.2.Introduction.....	120
6.3.Results and discussion.....	125
6.4.Concluding remarks.....	126
6.5.Materials and methods.....	127
7.CONCLUSIONS AND PERSPECTIVES.....	130
8.DESCRPTION OF AUTHOR CONTRIBUTION TO THE PROJECTS.....	132
BIBLIOGRAPHY (REFERENCES).....	134
LIST OF ABBREVIATIONS.....	146
CURRICULUM VITAE.....	147
LIST OF PUBLICATIONS.....	148

1. Introduction

Viruses are absolute parasites of cells of living hosts, and the life of the virus is inseparably associated with the life of its host cell. Viruses arguably represents the absolute majority of the living organisms on Earth. Viruses influence global biogeochemical cycles and drive the evolution of the entire biosphere. For example, in the oceans, which cover 70% of the Earth surface, the host cells killed by the viruses become an organic matter and can be utilized by other organisms, like heterotrophs, which increases nutrient recycling in the world's oceans, particularly the ocean's carbon cycling (Rohwer, Prangishvili, and Lindell 2009).

First viruses were observed by Dimitri Ivanovsky in 1892, and by Martinus Beijerinck in 1898. But only in 1930–1940s it became clear that viruses were “living organisms” but not host products (Michael G. Rossmann and Rao 2012).

There are several stages in the life cycle of a virus: 1) recognition and attachment to a susceptible host cell; 2) penetration of the host cell envelope; 3) transfer of the viral genetic material into the cell; 4) redirection of the host cell protein and nucleic acid synthesis machinery to produce new viral subunits; 5) self-assembly of new virions; and 6) release of the newly-assembled virions from the host cell (Michael G. Rossmann and Rao 2012).

1.1. Prokaryotic viruses

Prokaryotic cell viruses include archaeal and bacterial viruses. Since prokaryotes outnumber eukaryotes, their viruses are probably the most common entities in the living world. The bacterial viruses, called bacteriophages or “phages” (Suttle 2007) occur almost everywhere, in every ecological niche on Earth, in freshwater and oceans, in soils (potentially the biggest biosphere on the planet), on plants, inside and on the outside of eukaryotes, including humans and animals (Brüssow and Hendrix 2002; Suttle 2007). Remarkably, they exist even in some of the extreme environments like hot springs and hypersaline waters (Pina et al. 2011; Rohwer, Prangishvili, and Lindell 2009). The first bacteriophage observation and description was made in 1917 by Félix d’Herelle, who proposed the term “phage” (Michael G. Rossmann and Rao 2012). However, only in 1940s the viral nature of bacteriophages was confirmed, after the invention of the electron microscope (Hans-W. Ackermann 2011). Since then, thousands of bacteriophages have been discovered (H.-W. Ackermann and Prangishvili 2012). However, structure is described for only a few selected “model system” bacteriophages.

1.1.1. Bacteriophages life cycle

All known bacteriophages have circular single-stranded(ss) DNA or double-stranded(ds) DNA, linear

dsDNA or linear ssRNA, or segmented dsRNA genomes (International Committee on Taxonomy of Viruses 2012). Most known bacteriophage particles (virions) contain dsDNA, but there are small phage groups with ssDNA, ssRNA, or dsRNA (Fig.1). Bacteriophages with linear dsDNA genomes form the largest group (H.-W. Ackermann and Prangishvili 2012). Furthermore, many of the sequenced bacterial genomes contain phage DNA (called prophage), a latent form of phage integrated into the bacterial chromosome and replicated conjointly. However, some prophages will kill the cell if activated. Many bacterial genomes contain multiple prophages, and bacteriophage genes can occupy up to 20% of the bacterial genome (Canchaya et al. 2003; S. Casjens 2003).

Depending on the dsDNA bacteriophages life cycle, specifically their replication strategy, they may roughly be divided into two main groups: lytic and temperate bacteriophages. Lytic bacteriophage infection results in synthesis of progeny virions and bacterial host cell lysis, without any bacteriophage genome incorporated into the host cell chromosome. Alternatively, temperate bacteriophages are able to integrate its dsDNA into host cell chromosome (lysogenic cycle), and replicate conjointly, or establish other relationships with its host (phage genome may exist as a circular or linear plasmid) (S. Casjens 2003).

1.1.2. Biological impact of bacteriophages

As was mentioned above, viruses, and among them bacteriophages, influence the global biogeochemical cycles, like carbon cycling in the oceans. In addition, bacteriophages perform natural selection of the infection resistant bacteria and exchange of the genetic information between hosts. Thus bacteriophages drive the evolution of the entire bacterial community.

Bacteriophages are strain-specific bacterial killers, and as a classical “Predator-Prey” system, they will kill a dominant bacterial strain and free a niche for a different strain that will be controlled by another bacteriophage (“Kill-the-winner” hypothesis). Thus, constant flux of dominant bacterial species may result in their observed diversity. However, this regulation remains ambiguous (Suttle 2007; Thingstad 2000). Also, bacteriophage DNA serves as mobile DNA elements to transfer genes between bacteria. It was shown that prophages are responsible for many virulence factors from pathogenic bacteria, and inter-strain genetic variability. For example, *botulinus* toxin or *Vibrio cholerae* toxin are prophage encoded virulence factors. The majority of *Vibrio cholerae* strains are not pathogenic until infected by the prophage with the toxin gene (Boyd and Brüssow 2002; Canchaya et al. 2003).

Several estimates of the phages impact on evolution and diversity of bacterial communities and other ecosystems were reported. The estimated population of bacterial and archaeal cells in the biosphere is $4\text{-}6 \times 10^{30}$ (Whitman, Coleman, and Wiebe 1998). Since bacteriophages are typically present in 10-15-fold excess over bacterial cells (Bergh et al. 1989; Wommack and Colwell 2000),

the global phage population is enormously big, about 10^{31} . Every milliliter of a coastal ocean water contains $\sim 10^7$ of phage particles. Similar amounts are examined in the other biological environments (Wommack and Colwell 2000). Remarkably, that $\sim 10^{24}$ successful phage infections occur in a global scale every second, consequently the turnover time of the bacteriophage global population is about a week (Suttle 2007). This means that roughly 10^{24} microorganisms ($\sim 10^8$ tons) die every minute due to phage infections (Rohwer, Prangishvili, and Lindell 2009). In the oceans microorganisms constitute more than 90% of the biomass, and bacteriophages kills on average $\sim 20\%$ of this biomass every day. However, bacteriophages themselves represent only $\sim 5\%$ of the prokaryotic biomass due to the small sizes (Suttle 2007).

The bacteriophages population is extremely dynamic. Each of the 10^{24} successful phage infections represents potential evolutionary diversity-generating event which occurs every second on Earth. Every second an astronomical amount of genes exchange opportunities occur, such as appearance of new phage mutants during replication, phage DNA recombination with a host DNA, or particularly with prophage genome integrated into host DNA, or with a DNA of co-infecting phages.

The remarkable abundance of bacteriophages results from successful strategies to infect bacterial host cells. Bacteriophages have developed intriguing mechanisms for their own survival. However, at the same time bacterial cells have evolved multiple barriers or layers to protect themselves against host immune system, environmental hostility, both chemical (extreme pHs, toxic substances, enzymes) and physical (phagocytes, bacteriophages or bacteriocines attacks).

To infect the host cell bacteriophages have to deliver their genome through a complex and thick (>400 nm) multilayer bacterial cell envelope. Gram-negative bacteria have an envelope composed of outer cell wall containing lipopolysaccharides (LPS), periplasmic space with a thin peptidoglycan monolayer, and cytoplasmic membrane made of phospholipids and proteins (Moak and Molineux 2004). In contrast, Gram-positive bacteria have an envelope lacking outer LPS cell wall, but with a thick and highly cross-linked peptidoglycan layer and similar cytoplasmic membrane. Peptidoglycan layer is essential for bacterial cell integrity and shape, and varies from ~ 30 - 80 Å thick in Gram-negative bacteria to ~ 250 Å thick in Gram-positive bacteria (Beveridge and Graham 1991). Huge variety of LPS from outer cell wall of Gram-negative bacteria is found (Whitfield 2006), and many of them identified as major virulence factors of pathogenic bacteria, causing systemic infections like septicemia, meningitis, pneumonia and others (Stummeyer et al. 2006). Bacteriophages must be equipped to penetrate through all these bacterial barriers. Indeed, phages from the order *Caudovirales* evolved a special multi-functional organelle, "tail", which have been defined as the complex of proteins in phage virions involved in bacterial host recognition, attachment and genome delivery into the cell. i.e. the host cell envelope is breached with the help of the tail (Petr G Leiman and Shneider 2012a). Remarkably, despite the 10-fold excess over

bacterial host cells, only one virion per host cell is required for infection. Tailed bacteriophages have evolved an extremely efficient infection mechanism.

Bacterial cell envelope, for one side, creates an effective barrier against animal immune system and phage infection. On the other side, these external polysaccharides provide an attachment site for specialized bacteriophages that carries enzymes with specific polysaccharides depolymerases. These enzymes are organized as multiprotein complexes and situated on the tail part of the phage virion as a tailspikes (Fig. 2) (Petr G Leiman et al. 2007). Schematically, external sugar polymers (if present) are both recognized and locally digested by these ferments, and phage creates a narrow tunnel towards the cell surface. Remarkably, that digested tunnel may be used by other bacteriophage species, designating their cooperative behavior (Bayer, Thurow, and Bayer 1979). The followed peptidoglycan layer is disrupted by a glycosidase. The glycosidase molecules are also part of the tail, but their number and possible location vary in different bacteriophages and bacteriophage families (see ϕ 92 project) (Petr G Leiman and Shneider 2012a).

1.1.3. Abundance of tailed bacteriophages

How widespread are tailed bacteriophages among all prokaryotic viruses? Morphology of 6300 prokaryote viruses has been described for the last 50 years. Among them 6196 are bacteriophages and only 88 are archaeal viruses (H.-W. Ackermann and Prangishvili 2012). However, archaeal viruses are usually found in extreme - hypersaline, hyperthermic, or anaerobic niches like hot springs or hypersaline water, which are difficult to recreate in a laboratory (Pina et al. 2011).

The overwhelming majority of observed bacteriophages are tailed (96.3 %), others 3.7 % represents different phage families (filamentous, polyhedral etc., see Fig. 1). All known phages with contractile tails contain dsDNA genome (H.-W. Ackermann and Prangishvili 2012; Petr G Leiman and Shneider 2012a). The dsDNA-containing tailed phages are not only enormously abundant, but likely the absolute majority of organisms and arguably very ancient. Archaeal viruses morphologically similar to tailed bacteriophages are found, which suggests that tailed phage-like particles appeared more than 3 billion years ago, perhaps before the separation of three domains of life (H. W. Ackermann 1998; Bamford, Grimes, and Stuart 2005).

Tailed phages constitute probably the most diversified of all virus groups. At present, tailed phages (*Caudovirales*) are subdivided into three basic families based on tail morphology: *Myoviridae* phages (myophages) with long contractile tails (24.8%, e.g. bacteriophage T4), *Siphoviridae* phages (siphophages) with a long but non-contractile tails (57.3 %, e.g. bacteriophage λ), and *Podoviridae* phages (podophages) with a short non-contractile tails (14%, e.g. bacteriophages K1-5, and ϕ 29) (H.-W. Ackermann and Prangishvili 2012). Current classification of tailed phages takes into account only the visual morphology of the tail.

By genome structure, formation of concatemers, presence or absence of unusual bases and DNA or RNA polymerase genes, and DNA sequence 15 genera corresponding to phage groups have been defined (Maniloff and Ackermann 1998):

Myoviridae (groups T4, P1, P2, Mu, SPO1, ΦH).

Siphoviridae (groups λ, T1, T5, L5, c2, ψM1).

Podoviridae (groups T7, P22, φ29)

1.1.4. Morphology of tailed bacteriophages

Numerous structures of phage virions and their protein components have been determined by structural techniques, such as cryo-electron microscopy (cryo-EM) and X-ray crystallography. Huge amount of accumulated structural, bioinformatical and biochemical data indicates that tailed double-stranded DNA phages share a common evolutionary origin.

Tailed bacteriophages usually consist of protein and DNA only. Phage virion has no envelope and is composed of an icosahedral head (capsid) with DNA genome, and a tail.

In the strict sense, phage head possess cubic symmetry, since hexagonal capsids may be icosahedra, octahedra, or dodecahedra (Caspar and Klug 1962). The overwhelming majority of studied viruses heads are isometric, but about 15% are prolate derivatives of isometric heads (H. W. Ackermann 1998). Heads are built of similar capsid protein folds, first described for bacteriophage Hong Kong 97 (HK97) (Wikoff et al. 2000).

The size of the encapsulated phage DNA varies from ~17kb (*Streptococi* phage C1) to 700kb (*Bacillus* phage G) (H-W Ackermann 2003; Donelli, Guglielmi, and Paoletti 1972). All known phages from *Myoviridae* family have double-stranded DNA genomes, which are on average larger than in other phages.

Myoviridae phages strictly utilize lytic cycle (host cell lysis) without any integration between phage and host cell genomes. The phage genome is packaged in a process that requires energy derived from ATP hydrolysis (Earnshaw and Casjens 1980; Fujisawa and Morita 1997; Sun, Rao, and Rossmann 2010). The densely packaged genome usually is wrapped axially into a spool-like or inverse spool-like structure. Since packaged genomes have similar densities, the size of the capsid varies as function of genome size (Fokine et al. 2004; Petr G Leiman et al. 2007).

Phage genomes are mosaic, with different evolutionary histories for different segments. This indicates a dominant role of horizontal genes transfer in building phage genomes (Hatfull 2008). Genetic maps are complex and may include up to ~290 genes, like in phage T4 (possibly more in larger phages) (Calendar 2006).

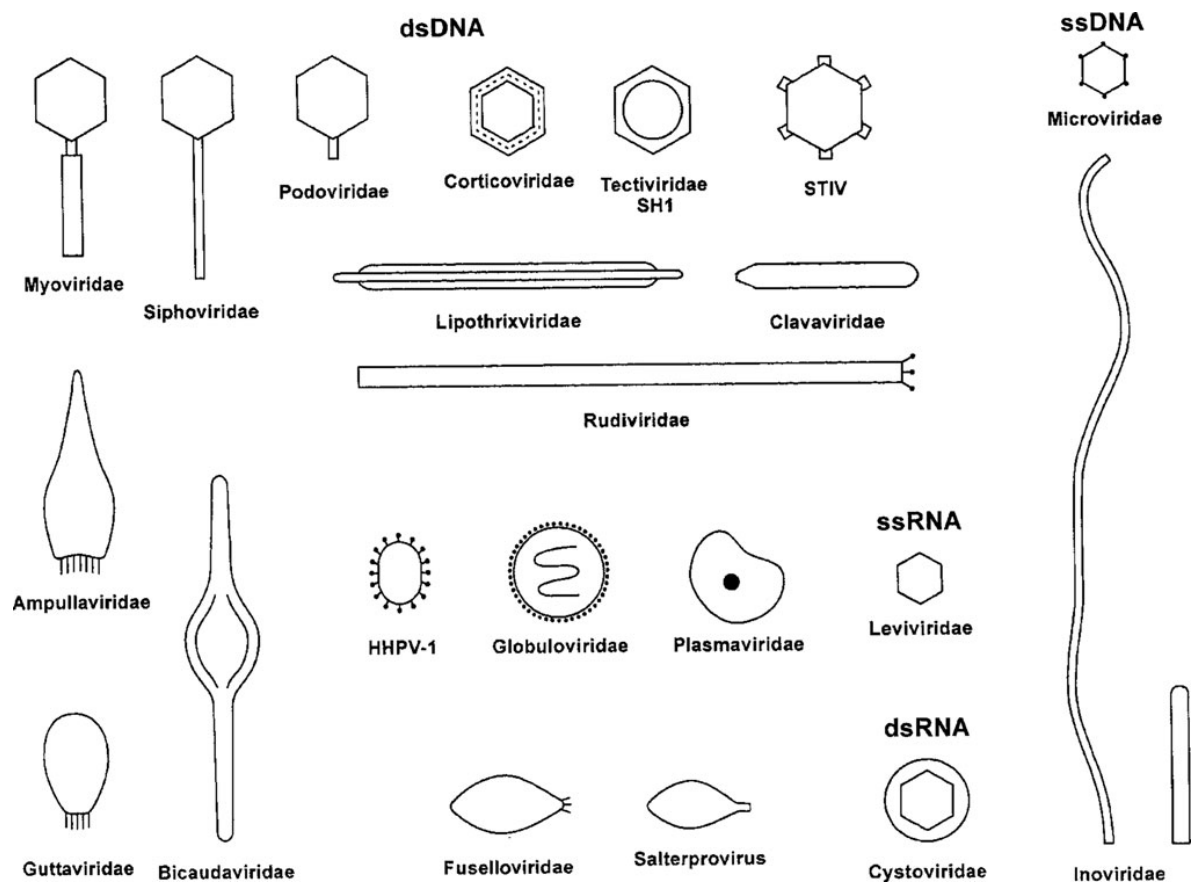


Figure 1: The present ICTV bacteriophage classification. **Source:** (Ackermann and Prangishvili, 2012).

Caudovirales bacteriophages are unique among all viruses due to the special multi-functional organelle, tail, responsible for host cell recognition, attachment and genome delivery into the cell.

Archaeal viruses morphologically similar to tailed bacteriophages have been found (Pietilä et al. 2014). The tail complexes varies from small *Podoviridae* tails (~40 proteins) to *Siphoviridae* or *Myoviridae* tails (>300 proteins), while tail length ranges from ~100 Å in some podophages to ~8000 Å in some siphophages (H-W Ackermann 2003).

Among three families of tailed bacteriophages, *Myoviridae* phages have the most complex tail structures, composed of largest number of proteins (Anastasia A. Aksyuk et al. 2009; Kostyuchenko et al. 2005; Petr G Leiman et al. 2010; Petr G. Leiman and Shneider 2012b). Contractile tails of *Myoviridae* phages generally consist of a baseplate and long non-contractile central tail tube (core) surrounded by external contractile sheath. The baseplate carries a tail fibers and/or tailspikes with a host cell-binding proteins responsible for the external capsular sugar both recognition and local digestion. In addition, there are also glycosidase molecules which break the peptidoglycan layer. They are part of the baseplate, but their number and possible location varies in different

bacteriophages and bacteriophage families.

Long non-contractile tails of *Siphoviridae* phages lack contractile sheath, and usually composed of a non-contractile tail tube and tail tip complex, to which a central fiber and side fiber proteins are attached (Fig. 2). The tail tip complex is responsible for interaction with the host cell which usually mediates by its emanating tail fibers (Davidson et al. 2012).

Podoviridae tails, despite variation in size and shape, contain a central hub structure to which 6 or 12 fibers or spikes are attached (Fig. 2). Usually, a tail hub is tubular and composed of 2 or 3 different proteins (see ϕ Eco32 and 7-11 phages description below). The upper part of the hub attached to the dodecameric portal vertex (Davidson et al. 2012).

1.1.5. Head (capsid) of bacteriophage

Phage (or any viral) capsid is a self-assembled genome container. Capsid must fulfill several properties:

- self-assemble from multiple copies of capsid protein into energetically favorable stable spherical closed shell;
- contain only one copy of the genome per capsid shell;
- have the correct size to fit the genome, without any cellular proteins inside;
- have mechanisms to strengthen the capsid shell against pressure of densely packaged genome.

1.1.5.1. Major capsid protein

Many bacterial and eukaryotic viral capsids have a similar organization. The so-called major capsid proteins (MCPs) of *Caudovirales* capsids are assembled into icosahedral shells. MCP forms pentamers and hexamers, capsomers, which are usually arranged in mature capsids in accordance with the “quasi-equivalence” principle (Caspar and Klug 1962).

The simplest icosahedral structure is composed of the 60 copies of the MCP, which forms identical bonds (interactions) with their neighbors. However, many isometric phage capsids contain more than 60 MCP copies. These bigger capsids can be built from the MCP copies with minimal distortions. Such architecture requires several MCPs to be placed into the icosahedral asymmetric unit with “quasi-equivalent” interactions. With that, the total number of MCPs is $60T$, where T is the number of quasi-equivalent positions, or the number of MCP in the icosahedral asymmetric unit. Strictly speaking, the total number equals $(60T-5)$, due to the absence of the pentameric vertex occupied by the tail.

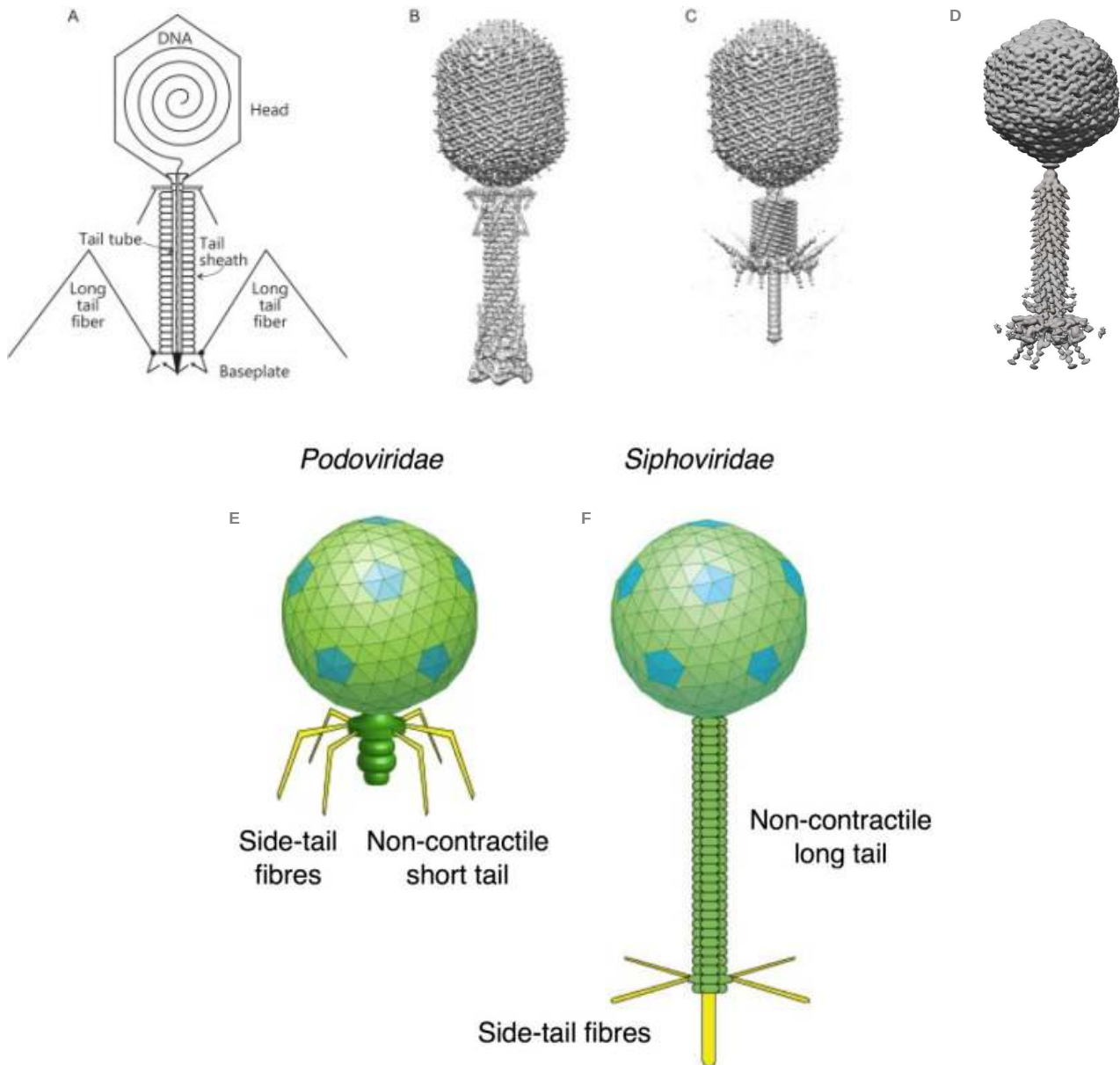


Figure 2: Typical morphology of the tailed phages. Myoviridae: T4 (**A-C**) and $\phi 92$ (**D**). Non-enveloped prolate capsid with icosahedral symmetry. The contractile tail composed of a sheath, 6 copies of long terminal fibers (LTFs), 6 copies of short tail fibers (STFs), and a small base plate. Adapted from: (Leiman et al., 2010). Podoviridae and Siphoviridae (**E**, **F**): The capsids are icosahedral with a T=7 symmetry, about 60 nm in diameter. P22-like podophage (**D**): Non-enveloped, head-tail structure. The tail is non-contractile, has 6 short sub-terminal fibers. Lambda-like siphophage (**F**): Non-enveloped, head-tail structure. The long tail is non-contractile and has 6 fibers although some Siphoviruses have only 3 or 4. **Source:** Swiss Institute of Bioinformatics, ViralZone.

Simplest icosahedral capsid has the triangulation number $T=1$, and is composed of 12 pentameric capsomers. Capsids with $T>1$ contain 12 pentameric and $10(T-1)$ hexameric capsomers. Some T numbers are not allowed, while some disallowed occur (Terje Dokland 2000)

In addition to isometric capsid organization, a small fraction (~15%) of prolate heads exists (H. W. Ackermann 1998). These elongated along 5-fold axis heads are assembled from two icosahedral end-caps and cylindrical midsection between them. Prolate capsid geometry describes by triangulation numbers T (icosahedral caps) and Q (cylindrical midsection). Similarly, the total number of the MCPs in the prolate head is $(30(T+Q)-5)$ (Moody 1999).

Remarkably, that MCPs of different tailed phages have a similar fold, first described for the bacteriophage Hong Kong 97 (HK97). HK97 fold was found in all 3D capsid structures of tailed phages, also in herpesviruses and even in bacterial compartments (encapsulins) (M. L. Baker et al. 2005; Sutter et al. 2008). Similar capsid protein fold indicates that tailed double-stranded DNA phages have a common evolutionary origin.

The majority of the *Caudovirales* capsids is composed of the copies of only one MCP, while some eukaryotic viruses (like herpesvirus, adenovirus, or some phages (like T4, ϕ Eco32, 7-11, or CBA120) encode two capsid proteins (M. G. Rossmann and Johnson 1989). In phage T4, the capsid protein gp24 forms pentameric capsomers, while another capsid protein gp23 forms hexameric capsomers (Fokine et al. 2004).

1.1.5.2. Scaffolding proteins

In complex icosahedral dsDNA bacteriophages, the correct size and geometry (correct T number) of the capsid shell is controlled by the set of virally encoded scaffolding proteins (T. Dokland 1999). Those proteins are removed after capsid shell assembly by proteolysis or probably pushed out by the genomic DNA during packaging (D.-H. Chen et al. 2011; P. G. Leiman et al. 2003).

1.1.5.3. Portal

As mentioned above, one of the 12 vertices of *Caudovirales* capsid is special and occupied by a tail complex. Tail attachment to the head accomplished with a help of the portal protein, which is also called a connector. Portal is required for genome packaging during phage virion assembly, and controls that only one copy of genome will be inserted. In addition, portal forms a channel for genome exit (ejection) during the infection process.

Accumulated structural data indicates the similar morphology of the portals among all studied *Caudovirales* and herpesviruses (Lander et al. 2006; Morais 2012; Elena V. Orlova et al. 2003). The portals are turbine-shaped molecules and form a dodecameric (C12 symmetry) ring, with a central channel which inner surface is predominantly negatively charged (Johnson and Chiu 2007). This

allows smooth passage of the negatively charged DNA molecule (Cuervo and Carrascosa 2012; Oliveira, Tavares, and Alonso 2013; Tavares, Zinn-Justin, and Orlova 2012). Despite the large variations in size and low sequence homology, all portal proteins share similar fold, suggesting that they have evolved from a common ancestor (Lebedev et al. 2007; Olia et al. 2011; Simpson et al. 2000).

1.1.5.4. Glue (and Decoration) proteins

In many *Caudovirales* dsDNA chromosome is densely packaged at ~500mg/ml as linear dehydrated dsDNA, which is self-repulsive due to the highly charged phosphate backbone. The mature phage capsid exerts internal pressure of tens of atmospheres without visible structural deformations. To stabilize and strengthen the capsid shell against internal pressure several mechanisms have been evolved. For example, HK97 capsid is stabilized by chemical cross-linking, T4 and ϕ 29 capsids are stabilized by special insertion domains in MCP, in phage epsilon 15 and adenovirus MCP subunits are held together by glue proteins (Fokine, Leiman, et al. 2005, 97; Jiang et al. 2008; H. Liu et al. 2010; Morais et al. 2005).

In addition to MCPs some tailed phages contain decorative proteins. For example, *Hoc* protein of T4 phage (Fokine et al. 2004), protruding capsid proteins of ϕ 29, or SPP1 phages (Morais et al. 2005; White et al. 2012). About ~25% of *Caudovirales* include immunoglobulin (Ig) domains on their surfaces (Fraser et al. 2006). IG-like domains probably help to bind reversibly to bacteria surface without infecting it.

1.1.5.5. Internal core proteins

Many tailed phages contain internal structures inside their capsids. The proteins which constituent them may be disordered or form ordered complexes, also called internal cores (Black and Thomas 2012). Roles of such internal proteins vary in different tailed phages. For example, some of the internal proteins are ejected into the host cell during infection, like in podophage N4 (S. R. Casjens and Molineux 2012). In myophage ϕ KZ the large internal body helps to organize the DNA during packaging (W. Wu et al. 2012). T7 phage contains a complex 3-layer internal core with different local symmetries of each layer (12-, 8-, and 4-fold). This internal core is built on the T7 portal. It was found that internal core not only forms the short tail extension during infection, but also helps with DNA packaging (Guo et al. 2013). Bacteriophage T4 has protein (gpalt), which locates close to the portal and helps with dsDNA transcription. Another protein (IPI) is dispersed within the DNA, and protects T4 genome from host cell endonucleases (Depping et al. 2005; Rifat et al. 2008).

1.1.5.6. DNA genome organization

DNA is a stiff molecule with a persistence length of ~500 Å. Inside the phage capsid it is very

tightly bent with inter-helical distance of 25-28 Å. When DNA is fully packaged into the capsid, it is organized in the hexagonal lattice. During the ejection process DNA reorganized into less ordered cholesteric state, and then into disordered isotropic state (Molineux and Panja 2013).

Several models of phage DNA arrangement were proposed (Black and Thomas 2012; Earnshaw and Casjens 1980). The most supported model is the layered inverse spool. In this model DNA is wrapped axially in concentric shells with the first layer on the outside, while the last part of the packaged DNA is in the central region of the capsid, and probably is the first to be ejected (Cerritelli et al. 1997; Earnshaw and Casjens 1980).

However, it is far from being well understood whether DNA spooling in tailed phages requires internal core or not. For example, podophages K1E, T7, SPP1 contain internal cores and dsDNA is packaged in agreement with the coaxial inverse spool model, where the axis of the DNA spool is aligned with the axis of the internal core (Guo et al. 2013; Petr G Leiman et al. 2007; X. Liu et al. 2010). For myophage T4 the spool model is also accepted, but with the spool axis orthogonal to the 5-fold elongation axis (Earnshaw et al. 1978). Remarkably, that packaged DNA of the podophage C1 is consistent with the spool model, although it does not contain an internal core (A. A. Aksyuk et al. 2012). However, proposed model is not correct for many phage capsids without internal cores (Leforestier 2013).

1.1.6. Contractile tails

Bacteriophages with contractile tails (*Myoviridae*) are of recently recognized class of “contractile injection systems” (Fig. 3). They share a common evolutionary origin and some functional aspect with other large macromolecular machines, such as the *Serratia entomophila* antifeeding prophage, the *Photobacterium* virulence cassette, R-type pyocins, and the bacterial type VI secretion system (T6SS) of Gram-negative bacteria (Browning et al. 2012; Petr G Leiman and Shneider 2012a; Pukatzki et al. 2007). Among tailed phages show up so-called “Jumbo” tailed phages, with genomes exceeding 250kb and virions of 300 nm and greater (Hendrix 2009). Hundreds of different proteins are involved in the morphogenesis and assembly of a jumbo phage particle. The DNA molecules of many of these giants are heavily modified (e.g. thymidine can be replaced by deoxyuridine) making genome sequencing either arduous or impossible. And even if sequenced, most of the predicted ORFs of these phages have no matches in the current sequence databases (Browning et al. 2012). The T6SS and the jumbo phages are the most complex of these contractile systems.

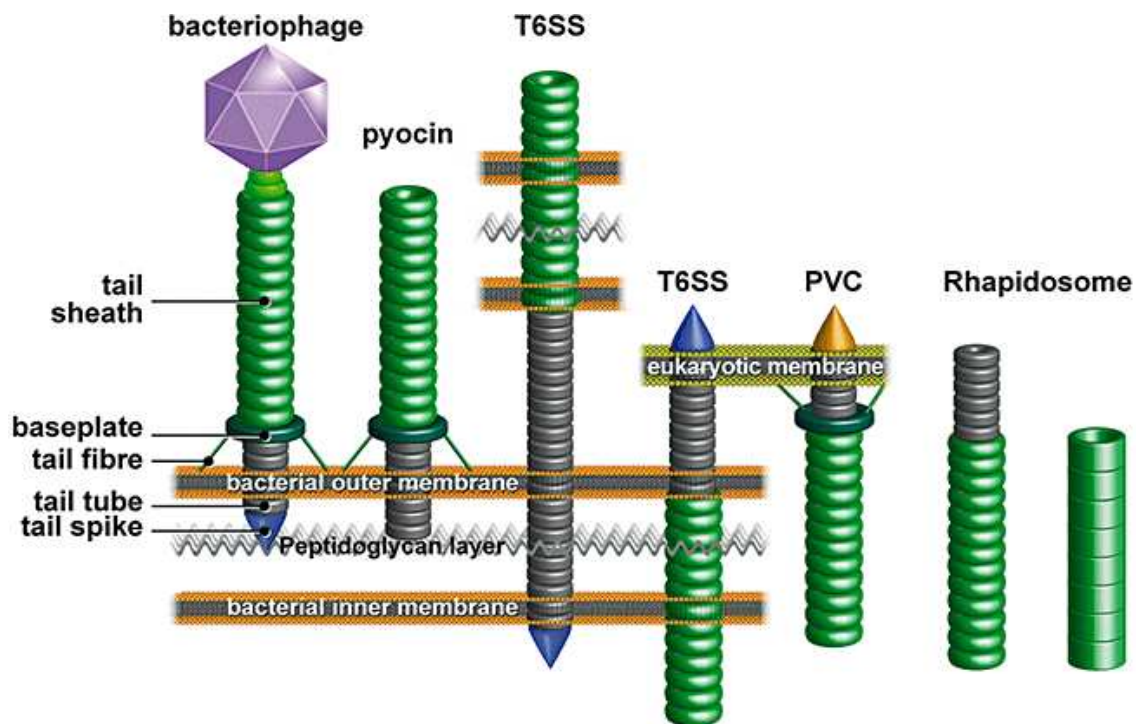


Figure 3: Examples of the contractile injection systems (tailed bacteriophage, pyocin, type VI secretion system (T6SS), *Photorhabdus virulence cassette*, rhapsosome) with main components highlighted. **Source:** (Bönemann et al., 2010).

A contractile injection system of typical *Myoviridae* is a macromolecular machine for infecting bacteria, that is to penetrate the host cell envelope and to deliver specific proteins and DNA into the host cell. These machines are composed of hundreds of copies of tens of different proteins (Petr G. Leiman and Shneider 2012b). Tail fibers, which emanate from the baseplate are the primary determinants of host cell specificity (Tétart, Desplats, and Krisch 1998; S. R. Williams et al. 2008).

In addition, T6SS and *Myoviridae* phages have been proposed to use a special protein to create an opening in the host cell membrane during infection. These special proteins, called cell-puncturing tips or devices, have a modular structure and translocate through the membrane as a rigid needle (Browning et al. 2012).

Current accumulated structural and genomic data suggests that the contractile tail systems contain a set of conserved genes and have a common ancestor. More importantly, many contractile tail genes are shared with non-contractile tails also pointing to the common ancestry (Petr G. Leiman and Shneider 2012b). The major part of the contractile tail is the internal tube and the external contractile sheath. Notable, that sheath and tail tube genes are adjacent to each

other in the genome, and serve as a genetic marker of a contractile tail. The average dimensions of contractile tails are varied in length, from 1,000 to 4,500 Å (see CBA120, AR9 phages descriptions below), but roughly has the same diameter of 220 Å.

The contractile tail of T4 bacteriophage is the best-studied contractile system and serves as the model system for all contractile sheath-like systems (Petr G Leiman et al. 2010). Its composition, assembly, structure, and conformational changes during the host cell attachment/penetration to the host cell are well-studied. Interestingly, T4 tail was the first specimen for which a three-dimensional map was derived from EM images (Petr G Leiman et al. 2010).

1.1.7. T4 contractile tail morphology

The T4 external contractile sheath is composed of 138 copies of a sheath protein gp18. The sheath of extended tail has dimensions of 925 Å x 240 Å, and is organized into a 6-start helix with the helical rise of 40.6 Å and right-handed twist of 17.2°. It is a metastable high-energy structure resembling an extended spring (Anastasia A. Aksyuk et al. 2009). The 6-fold rotational symmetry was observed in all studied *Myoviridae* phages, and might be functionally advantageous due to the interaction of the phages with the oligosaccharides which form hexagonal arrays on the outer surface of some bacterial strains (Moody 1999).

The internal non-contractile tube is believed to be composed of the same number of copies of a tube protein gp19. However, its three dimensional organization is not completely revealed by the current amount of structural data, whether it follows the helical symmetry of the sheath or not (Anastasia A. Aksyuk et al. 2009). The length of the T4 tail is controlled by the tape measure protein gp29 (Abuladze et al. 1994). Similar protein was found within the tail tube of siphophages (Davidson et al. 2012; Katsura and Hendrix 1984).

From one side, the tail sheath is attached to the baseplate, from another side to the phage capsid through the set of intermediate proteins, forming the neck region. First, the tail starts with the sheath and tube tail terminator proteins gp3 and gp15, critical components for the T4 virion assembly. *Myoviridae* tail cannot attach to the capsid without terminator proteins. Second, the tail is finished with the head completion protein gp13 and/or gp14, which form the head-tail binding interface (Petr G Leiman et al. 2010).

Infection process of *Myoviridae* phage starts with the attachment of the tail fibers to the host cell, which triggers the conformational change of the baseplate. The baseplate conformational change causes the sheath contraction, and drives the tail tube with the cell-puncturing tip through the center of the baseplate to penetrate the outer membrane. After penetration the tail tube creates a channel for DNA ejection from the capsid into the host cell (Kostyuchenko et al. 2005; Petr G Leiman et al. 2004; P. G. Leiman et al. 2003). After contraction the T4 sheath changes its

dimensions to 420 Å x 330 Å. The 6-start helical structure is preserved but with the helical rise of 16.4 Å and twist of 32.9°. Each sheath monomer gp18 moves about 50 Å radially outwards and rotates roughly 45° clockwise (Petr G Leiman et al. 2004).

While the contractile sheath and the tube are responsible for DNA delivery into the already found prey cell, the baseplate is the control center coordinating the host attachment process, as well as priming the tail assembly. The baseplate varies in size and complexity in different tailed phages (P. G. Leiman et al. 2003; Sciara et al. 2010). Very complex baseplates of bacteriophages CBA120 and AR9 have been described by us.

The T4 hexameric baseplate is composed of at least 16 different proteins, which are in the main oligomeric. The baseplate core is the trimeric central hub which contains proteins gp5 and gp27. The gp27-like proteins also present in non-contractile *Siphoviridae* tails, the fact indicating a common baseplate hub structure in phages (Sciara et al. 2010). The gp27 protein ends with a trimeric central spike protein gp5. *Myoviridae* phages, as well as T6SS, have been proposed to use these central spikes to create an opening in the host cell membrane during infection. They share a common modular structure, where the N-terminal domain contains the oligosaccharide/oligonucleotide-binding (OB) fold, the central domain has a lysozyme-like fold, and the C-terminal domain is a long intertwined β -helical structure. The T4 trimeric gp5 tip is sharpened by recently identified gp5.4 protein (Shneider et al. 2013).

Ortholog of the T4 cell-puncturing device was found in the T6SS, which is responsible for toxic effector (VgrG protein) translocation across the membrane of predatory bacteria directly into the cytoplasm of the eukaryotic prey cells (Pukatzki et al. 2007). The central cell-puncturing devices of myoviruses P2 and ϕ 92 (see ϕ 92 project), gpV and gp138, form the membrane-piercing components similar to that of gp5 of T4, but without the lysozyme domain and the tip. Instead, they have a sharp C-terminal extension coordinated by an iron ion (Browning et al. 2012).

The T4 central hub is surrounded by six copies of wedges, each made of 7 proteins: gp6, 7, 8, 10, 11, 25, 53. Also, there are copies of the long tail fiber-connecting protein (gp9), the short tail fiber protein (gp12), and gp48 and gp54 proteins, which are necessary for initiation of the tail tube and the tail sheath assembly (Kostyuchenko et al. 2005; Kostyuchenko et al. 2003).

Finally, there are long tail fibers attached to the baseplate. Each long tail fiber is a kinked stiff structure ~1450-Å-long, composed of gp34, 35, 36, 37 proteins (Petr G Leiman et al. 2010). The long tail fibers can bind reversibly to the LPS or the outer membrane protein (OmpC) of the *E. Coli* host cell (Bartual et al. 2010; Henning and Hashemolhosseini 1994). In solution the T4 particle has the long tail fibers extended without any fixed orientation, which point towards the head or move due to the thermal motion (Simon 1972).

Probably, the initial attachment to the susceptible host cell receptors is reversible and is

accomplished by one or two fibers. If the reversible adsorption is sufficiently weak, it allows the phage particle to recognize the correct host cell type and rapidly find the secondary receptors (Karam and Drake 1994, 4). Initial attachment increases the probability for the other fibers to find the host cell receptors. However, the host cell is constantly moves in solution, thus attached phage is under the influence of the various sorts of solvent movements. To stabilize the attached virion on the cell surface the other tail fibers are bind to the host cell receptors, which is possible only when all the fibers point towards the host cell surface. Thus, the phage orients perpendicular to the host cell membrane. Such configuration of fibers switches the baseplate from the high energy metastable hexagonal state to a lower energy state. Since the T4 long tail fibers are stiff structures, they transmit the conformational changes from the periphery of the baseplate through their length to the baseplate central hub. The central hub then unlocks and initiates sheath contraction. The energy of the sheath contraction moves the tail tube with the cell-puncturing device at the end to penetrate the outer host cell membrane. The contraction induced tail tube movement also includes rotation, which helps to drill rather than pierces the outer membrane (Petr G Leiman et al. 2010).

A reasonable questions arises, why the spontaneous tail contraction does not occur when the phage is free in solution? One of the explanations is the significant energies difference between extended and contracted tail states. This difference is likely higher than the average kinetic energy of molecules in solvent, and only the fibers fixation on the cell surface and the phage virion shaking by the solvent provide enough energy to overcome the barrier (Petr G Leiman et al. 2004). Another explanation, which possibly elaborates the first one, is the observed presence of divalent cations, needed for phage infections. The cations concentrations near the cell surfaces usually is very high and can cause the tail fibers extension and attachment to the host cell receptors, which further triggers the sheath contraction. Similar mechanism was proposed for the Gram-positive *Lactococcus lactis* bacteriophage p2 (Sciara et al. 2010).

Remarkably, that the energy stored in the densely packaged DNA is not used in membrane drilling. More importantly, tail sheath contraction does not triggers the DNA ejection and stable phages with DNA and contracted tails have been observed (Petr G Leiman and Shneider 2012a). This fact probably indicates the presence of unknown additional receptors on the inner membrane which triggers the DNA translocation into the host cell cytoplasm.

While for *E.Coli* phage T4 and other myophages the contraction process is well-studied (Petr G Leiman et al. 2004), for dsDNA bacteriophages without a contractile sheath and a tail tube the life cycle has not yet been studied in detail. Podophages, despite their smaller dimensions, genome sizes and simpler overall morphology, have an obscure DNA ejection mechanism. Short non-contractile tails of podophages are too short to pierce the host cell envelope. In case of Gram-positive host cells *Podoviridae* phages can digest the thick cell wall and approach to the

cytoplasmic membrane without additional transit machinery. However, Gram-negative infecting podophages must somehow extend their tails to create a trans-envelope channel used for DNA transport (S. R. Casjens and Molineux 2012).

1.2. Electron cryo-microscopy

The structure of biological objects, or three-dimensional (3D) spatial arrangement of constituent molecules is necessary for complete understanding of biological processes. Since biological objects are assembled in a hierarchical manner, from macromolecules to the organisms, physical dimensions of constituent blocks range over a several orders of magnitude. The dimensions of viruses, and among them bacterial viruses, limit the methods of choice for studying them. To image objects as small as viruses we need to resolve features in the range from ~10 nanometers to several angstroms (Fig. 4).

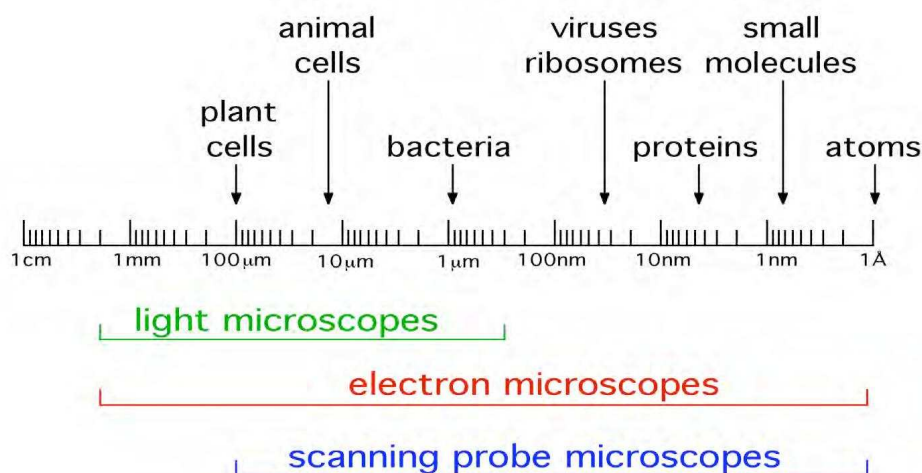


Figure 4: Resolving power of light and electron microscopy. **Source:** The Center for Cellular Imaging and NanoAnalytics, C-CINA.org.

From one side, X-ray crystallography is the most powerful tool for protein structural analysis and provides atomic-resolution structural description of many viral components (Anastasia A. Aksyuk et al. 2009; Browning et al. 2012; Helgstrand et al. 2003, 97; Olia et al. 2011). However, crystallization itself may be difficult or even impossible for “large” objects such as viruses (Michael G. Rossmann et al. 2005). From the other side, super-resolution microscopy reveals structural information at cellular level and breaks the diffraction limit for conventional optical microscopy (Huang, Babcock, and Zhuang 2010), but the resolution limit is ~50 nm and comparable to the average dimension of tailed phage virions (T. S. Baker, Olson, and Fuller 1999).

Fortunately, transmission electron microscopy (TEM) and specifically electron cryo-microscopy (cryo-EM) combined with three-dimensional (3D) image processing bridges the gap of several orders of magnitude, from near-atomic to cellular level (Hu et al. 2013; J. Liu et al. 2011). Several high-resolution “near-atomic” structures have been reported (Brown et al. 2014; B. Wu et al. 2014).

Cryo-EM combined with X-ray and optical fluorescence data provide the most complete structural description of biological specimens like viruses (Schorb and Briggs 2014). Remarkably enough, the bacteriophages were one of the first biological objects visualized by TEM (Kruger, Schneck, and Gelderblom 2000).

So, why use electrons? Electrons show both particle and wave characteristics. Particle-wave dualism was theorized by Louis de Broigle in 1925, and further proved by classical electron-diffraction experiment in 1926 by George Tompson. Since electrons, unlike X-rays, are negatively charged, they can be focused by electromagnetic/electrostatic lenses, which was discovered in 1926 by Hans Busch. After a short while, the first electron microscope was proposed in 1931 by Max Knoll and Ernst Ruska (Haguenau et al. 2003).

Electron with mass m and charge e accelerated through a potential difference V_0 has a wavelength:

$$\lambda = h / \sqrt{2 m_e V_0} = 1.22639 / \sqrt{V_0} \text{ , where } h \text{ is a Planck constant (Eq. 1).}$$

By increasing the the accelerating voltage we can decrease the electron wavelength. At high energies, 100keV or more, the electron velocity becomes higher than half of the speed of light c , and the relativistically corrected electron mass has to be considered:

$$\lambda = h / \sqrt{2 m_e e V_0 \left[1 + \frac{e V_0}{2 m_e c^2} \right]} \text{ (Eq. 2) (Spence 2013).}$$

Imaging with electrons provides the advantage of high resolution due to their short wavelength, where resolution is the ability of two closely spaced features in the object to be distinguished. For example, for a 100 kV acceleration voltage the electron wavelength is $\lambda \sim 0.004\text{nm}$ (4pm), which is much smaller than than diameter of an atom. But currently it is nowhere near to approach this wavelength limit of resolution due to the electron lenses imperfections. Without aberrations (see below) the resolution of any lens defines by the classical Rayleigh criterion: $r_0 = 0.61 \lambda f / D$, where λ is the electron wavelength, f is the lens focal length, D is the lens aperture diameter. R_0 determines the radius of the Airy disk, which results from the electrons diffraction on the aperture. A single point source will not be imaged as a point even if no aberrations or astigmatism present. The Airy Disc contains 84% of the total intensity. In reality, in the presence of spherical aberration C_s of objective lens the practical resolution is:

$$r \approx 0.91 \sqrt[4]{C_s \lambda^3}$$

Electrons are strongly interacted with the sample, the scattering cross section is $\sim 10^5$ higher than for X-rays. The nature of this interaction depends on the electron energy and sample composition. One advantage of that is the wide range of secondary signals from the specimen which are used in analytical electron microscopy (D. B. Williams and Carter 2009). But strong interaction also causes radiation damage of the sample by transferring some of the electron's energy to individuals atoms in the specimen. This level of damage comparable to that produced by an atomic explosion (Glaeser and Taylor 1978).

The incident electron beam passes through the specimen and individual electrons are either unscattered (direct beam, $\sim 80\%$) or scattered by the atoms of the specimen. Some of the electrons are scattered elastically ($\sim 5\%$) over relatively large angles and without energy loss. They deflected by the electron cloud of the sample atom. While others scattered inelastically ($\sim 15\%$) over relatively narrow angles with consequent energy loss, and transfer their energy (tens of eV's) to the specimen and damage it (E. V. Orlova and Saibil 2011).

The scattered and unscattered beams are interfered and create the interference pattern, the image. The electron microscope then becomes an interferometer (Spence 2013). Only elastically scattered electrons contain all the high-resolution information of the specimen structure and contribute to the theoretical image intensity. The out-of-focus inelastically scattered or energy-loss electrons produce an unwanted uniform background. This is one of the principal limitation to high resolution structure determination.

When the specimen becomes relatively thick, most of the electrons will scattered inelastically, and damage the sample by breaking the bonds. The average energy transferred to the secondary electrons is $\sim 20\text{eV}$ which is $\sim 5\text{-}10$ times greater than the valence bond energies. Thus, biological samples to be imaged by electrons must be very thin, a few hundreds of Å (Thach and Thach 1971).

Biological samples are extremely sensitive to the radiation damage. For example, dose of 10^{-6} electrons per Å² ($\text{e}/\text{Å}^2$) causes death of *E.Coli*, $10^{-5} \text{ e}/\text{Å}^2$ inactivates enzymes, dose of $1 \text{ e}/\text{Å}^2 - 10 \text{ e}/\text{Å}^2$ causes bond breakage in amino acids, dose of $100 \text{ e}/\text{Å}^2$ causes the destruction of nucleic acids (Reimer 1997). The solution to preserve biological samples from radiation damage is to image them at low temperature, usually at near liquid-nitrogen temperature (-170°C). In addition, the amount of electron dose must be minimize, usually to $1\text{-}20 \text{ e}/\text{Å}^2$ (T. S. Baker and Henderson 2006; B. and E. 2012). To enhance the low-resolution features the dose may be exceeded, but high resolution features are affected at the dose of $\sim 10 \text{ Å}$ (Knappek and Dubochet 1980). Unfortunately, low dose sample exposure comes at price and results in a poor signal-to-noise ratio (SNR) in the image. The compromise must be found between acceptable electron dose and the image contrast.

Low SNR strongly limits the amount of information even in a single image of biological macromolecule. Thus, it becomes necessary to average the information from $\sim 10^6$ low-contrast noisy images, or even more (Rosenthal and Henderson 2003). Things are becoming more serious since organic samples made of low atomic number atoms such as hydrogen, carbon, nitrogen and oxygen, which scatter electrons weakly. The average protein density is only 1.35 times higher than that of vitrified solvent (Fischer, Polikarpov, and Craievich 2009). Due to the weak scattering the biological samples do not modulate the intensity or amplitude of the incident electrons, but shift the phases of elastically scattered electrons instead. The unscattered and elastically scattered electrons interfere, and produce the phase contrast, which is not visible in the image. Amplitude contrast may be increased by the objective lens aperture which eliminates elastically scattered electrons outside the aperture. This is relatively low resolution contrast, since the aperture eliminates portion of elastically scattered electrons containing high-resolution information of the sample.

1.2.1. Transmission electron microscope

Typical modern transmission electron microscope is composed of the illumination system, the objective lens/stage and the imaging system. All of the systems are located in the electron column operating at high vacuum created by the vacuum system (Fig. 5). The role of the illumination system is to create the accelerated electron beam and transfer it to the specimen through the set of 2 (or 3) condenser lenses with the apertures and the stigmator. It allows to regulate the diameter of the illuminated area on the sample according to the selected magnification.

The electron beam creates by the electron source or “gun”, thermionic or field emission gun (FEG). FEG usually has a $\sim 10^3$ higher source brightness than thermionic guns, but requires a very high vacuum $\sim 10^{-10}$ Torr. FEG creates very parallel (divergence of $\sim 10^{-2}$ mrad) electron beam, with high spatial coherency and small energy spread, the conditions necessary for high-resolution imaging. Condenser lens 1 (short-focus) creates the electron source image, condenser lens 2 (long-focus) transfers this image into the sample plane, which is the front focal plane of the objective lens. Condenser apertures exclude electrons from the periphery of the beam which improves beam monochromaticity and increases the interference effects (phase contrast). Condenser stigmator corrects the beam astigmatism in the object plane.

The objective lens is the heart of the microscope. It creates the first magnified image of the specimen and focalize it. Objective lens strongly suffers from astigmatism and aberrations, the most important of which are the spherical and chromatic aberrations, C_s and C_c , which all together determines the maximal obtainable resolution of TEM image. To reduce the aberrations the specimen is usually introduced into the objective lens field (immersion lens), and the focal length of objective lens is reduced (1.5–2mm). In addition, complex C_s -correctors can be used.

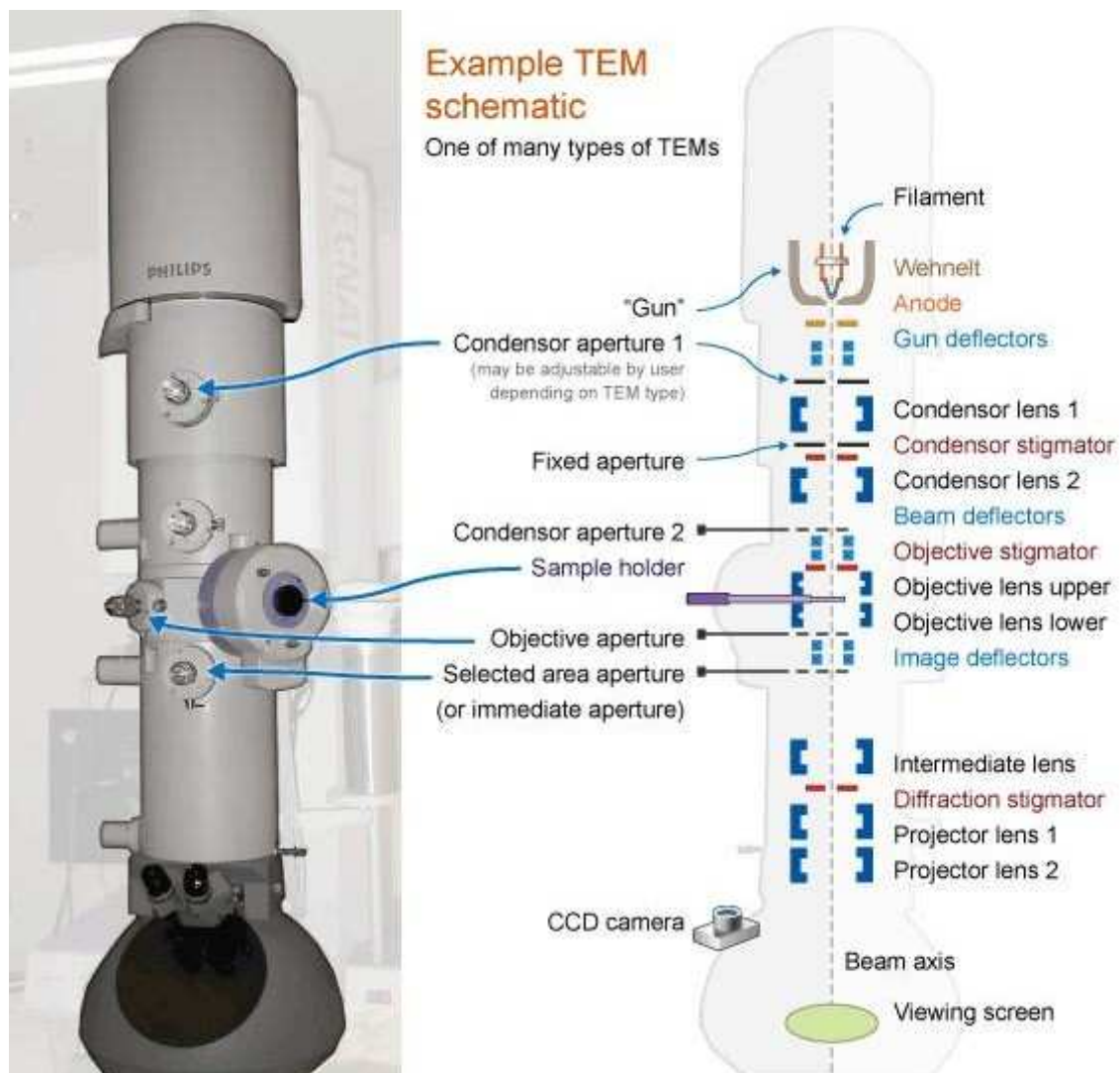


Figure 5: A schematic representation of a modern transmission electron microscope (TEM) with main components highlighted.

The objective aperture is used to improve aperture or scattering contrast in the image. Smaller aperture improves resolution limited by spherical and chromatic aberrations but reduces resolution limited by interference effects (acts as low-pass filter). Thus, an optimum aperture size must be used. Objective lens forms a diffraction pattern in the back focal plane, and the image in the image plane. Depending on which TEM mode is used, the imaging or diffraction mode, the image or the diffraction pattern will be further magnified and projected on the screen by the imaging system, which is usually the set of several lenses. The intermediate lens (usually block of 2 lenses) changes the magnification in the wide range without changing the focalization of the objective lens. If the

diffraction mode is selected the strength of intermediate (diffraction) lens below the objective is decreased, so that it is now focused on the back focal plane of the objective lens, instead of the objective lens image plane. Projective lenses 1 and 2 further magnifies the image (or diffraction pattern), and also creates big depth of focus which keeps the image in focus on the fluorescent viewing screen, on the conventional film ~10 cm below or above the viewing screen, or on the CCD or TV cameras.

TEM image of three-dimensional (3D) specimen is a two-dimensional (2D) projection in the direction of the incident beam (optical axis of the microscope). Usually small objective lens apertures are used which creates depth of field of more than 1 micron, where depth of field is the distance along the axis on both sides of the object plane within which one can move the object without detectable loss of focus in image. All of the specimen than is in focus, without depth-sensitivity which may creates confusing overlaps along z-axis. Thus, to provide relevant interpretation of structural features the 3D reconstruction from many 2D projections of the sample is required.

1.2.2. Image formation in TEM

Let's describe in more details the projection image formation in TEM. The thin (<100 nm) biological sample is illuminated with the incident electron beam. Let us set the incident wave function as Ψ_o , and Ψ_e as the exit wave, or the object transmitted wave function. Exit wave function can be written as

$$\Psi_e(\vec{r}) = \Psi_o \exp^{i\sigma\varphi_{proj}(\vec{r})} \quad (\text{Eq. 3}),$$

where $\sigma = 2\pi m_e \lambda / h^2$; m_e and λ are the relativistically corrected electron mass and wavelength, h is

the Planck's constant, and $\varphi_{proj}(\vec{r}) = \int_{-t/2}^{t/2} \varphi(\vec{r}, z) dz$ is the projected specimen potential along the z

axis (optical axis of the microscope), \vec{r} is the radius-vector in the image plane, t is the sample thickness (Spence 2013). As mentioned above, a thin biological specimen is almost the pure phase object and the weak-phase object approximation (WPOA) is valid:

$$\Psi_e(\vec{r}) \approx \Psi_o \{ 1 + i\sigma\varphi_{proj}(\vec{r}) \} = \Psi_o + i\Psi_o\sigma\varphi_{proj}(\vec{r}) \quad (\text{Eq. 4}).$$

The first term corresponds to the central unscattered electrons, and second to the scattered ones.

The image intensities can be written as:

$$I(\vec{r}) = \Psi_e(\vec{r})\Psi_e^*(\vec{r}) \approx \Psi_o^2 + \Psi_o^2(\sigma\varphi_{proj}(\vec{r}))^2 \quad (\text{Eq. 5}),$$

where $\Psi_e^*(\vec{r})$ is the complex conjugate of $\Psi_e(\vec{r})$. In the weak-phase approximation the image intensity is dominated by the first term, and the imaging mode is called the bright-field TEM. The most important consequence of this approximation is the linear dependence of the image intensities to the projected electron potential which describes density variations in the specimen (Spence 2013). However, due to the domination of the first term in Eq. 5 the resultant wave will be almost the same as the incident wave, resulting in the contrast absence in the image.

To increase the image contrast the phase contrast has to be converted into recordable amplitude contrast by shifting the phases of the scattered electrons by a quarter wave, $\pm\pi/2$ radians. This results in constructive/destructive interference of scattered and unscattered electrons, and noticeable contrast increase/decrease relative to the unscattered wave. In light microscopy the annular phase plate is placed in the diffraction plane of the objective lens. The TEM phase plates become available recently (Murata et al. 2010).

However, usually the image contrast of thin biological samples is increased by adding heavy atoms to the specimen (negative staining), which increase electron scattering and also preserves the sample from the high vacuum of the electron microscope. Another option, which is more common in cryo-EM field is to utilize phase shifts provided by the microscope imperfections, namely by the spherical aberration of the objective lens and defocus, which is focusing the electron beam under the image plane (under-focus) of the objective lens (Nagayama 2014). Image defocus and spherical aberration impose a phase shift (Scherzer formula):

$$\gamma(\vec{S}) = 2\pi \left\{ \frac{1}{4} C_s \lambda^3 \vec{S}^4 - \frac{1}{2} \Delta Z \lambda \vec{S}^2 \right\} \quad (\text{Eq. 6}),$$

where \vec{S} is the spatial frequency (\AA^{-1}), C_s is the spherical aberration coefficient (mm), λ is the wavelength of the electron beam (nm), ΔZ is the defocus value (nm), which is the distance from the true focal plane to the image plane. The bigger the ΔZ the stronger the contrast of the thin biological sample.

However, as usual, contrast enhancement comes at price. In addition to phase shift, defocalization creates artifacts described by so-called contrast transfer function (CTF) of the microscope (Scherzer 1949). CTF can be described conveniently in reciprocal (Fourier) space as:

$$CTF(\vec{S}) = \{ A \cos(\gamma(\vec{S})) - (1 - A^2) \sin(\gamma(\vec{S})) \} \propto e^{i\gamma(\vec{S})} \quad (\text{Eq. 7}),$$

where A is the fraction of the amplitude contrast, and depends on the microscope voltage and specimen fixation method (negative staining or cryo-fixation). Usually for 200 keV images the amplitude contrast is $\sim 7\text{-}10\%$, for 300 KeV images $\sim 4\%$ (Toyoshima and Unwin 1988). Usually the

$CTF(\vec{S})$ is modulated by envelope function $E(\vec{S})$:

$$CTF(\vec{S})E(\vec{S}) ,$$

where $E(\vec{S})$ is the compound envelope function, and describes dampening of the signal due to the partially coherent illumination and energy spread of the beam (Chiu and Glaeser 1977). The first term in Eq. 7 corresponds to the amplitude contrast, while second describes the phase contrast and has the major influence on phase objects. Interestingly, that since there is small fraction of amplitude contrast generated, the object is not a pure phase object (weak-phase approximation is valid) and the CTF starts off with a nonzero term A. The CTF is oscillating function, which results in the information lost in the zero transitions, reverse of the phase and damping of the high-resolution signal by the envelope function (Erickson and Klug 1971). The CTF relates to the microscope image of small point object, point spread function (PSF), through a Fourier transform:

$$F\{PSF(\vec{r})\}=CTF(\vec{S}) \text{ (Eq. 8).}$$

CTF of the microscope usually is observed as the Thon rings in Fourier transform of carbon film (Thuman-Commike and Chiu 2000).

1.2.3. Sample preparation methods

As mentioned above, the sample must be imaged in a high vacuum. In addition, electron beam sources, such as FEG, are also required a high vacuum. Since biological are samples mainly consist of water, they must be preserved to stay in the high vacuum of TEM column without significant structural changes compare to their native state. One of the simplest preparation methods is the negative staining (James R. Harris and Royal Microscopical Society (Great Britain) 1997). The small amount of specimen ($\sim 3\mu\text{l}$) is mixed with the electron-dense heavy metal salt solution, like uranium, molybdenum, tungsten etc. (J. Robin Harris and Scheffler 2002), spread on special EM grid, blotted and dried (Fig. 6). Heavy metal salt coats the sample and preserves it from high vacuum of TEM column, and also increases electron scattering, which gives a higher contrast. However, staining procedure introduces artifacts and distortions to the sample molecules, like flattening or collapse of the 3D structure of the sample. Another limitation comes from the ability of the electrons to penetrate the heavy salt coat, which usually limits the resolution to $\sim 20 \text{ \AA}$ due to the salt grain size. Thus, only solvent-excluded surface shape and size of the specimen molecule is available. However, this simple and quick procedure provides very high contrast, and mainly used for the initial sample screening. The advanced negative staining procedure is the cryo-negative staining (M. Adrian et al. 1998).

The preferable method of obtaining the high-resolution structural details is the electron cryo-microscopy (cryo-EM). The revolutionized sample fixation and preservation method was proposed by Marc Adrian and Jacques Dubochet in 1984 (Marc Adrian et al. 1984; Dubochet et al. 1988). The idea is to preserve the unstained biological sample in its native hydrated or near-native state, minimizing the distortions or artifacts. This can be achieved by rapid cooling of the sample at high rate ($\sim 10^6$ K/s), preventing the formation of cubic or hexagonal ice, which disrupts the biological

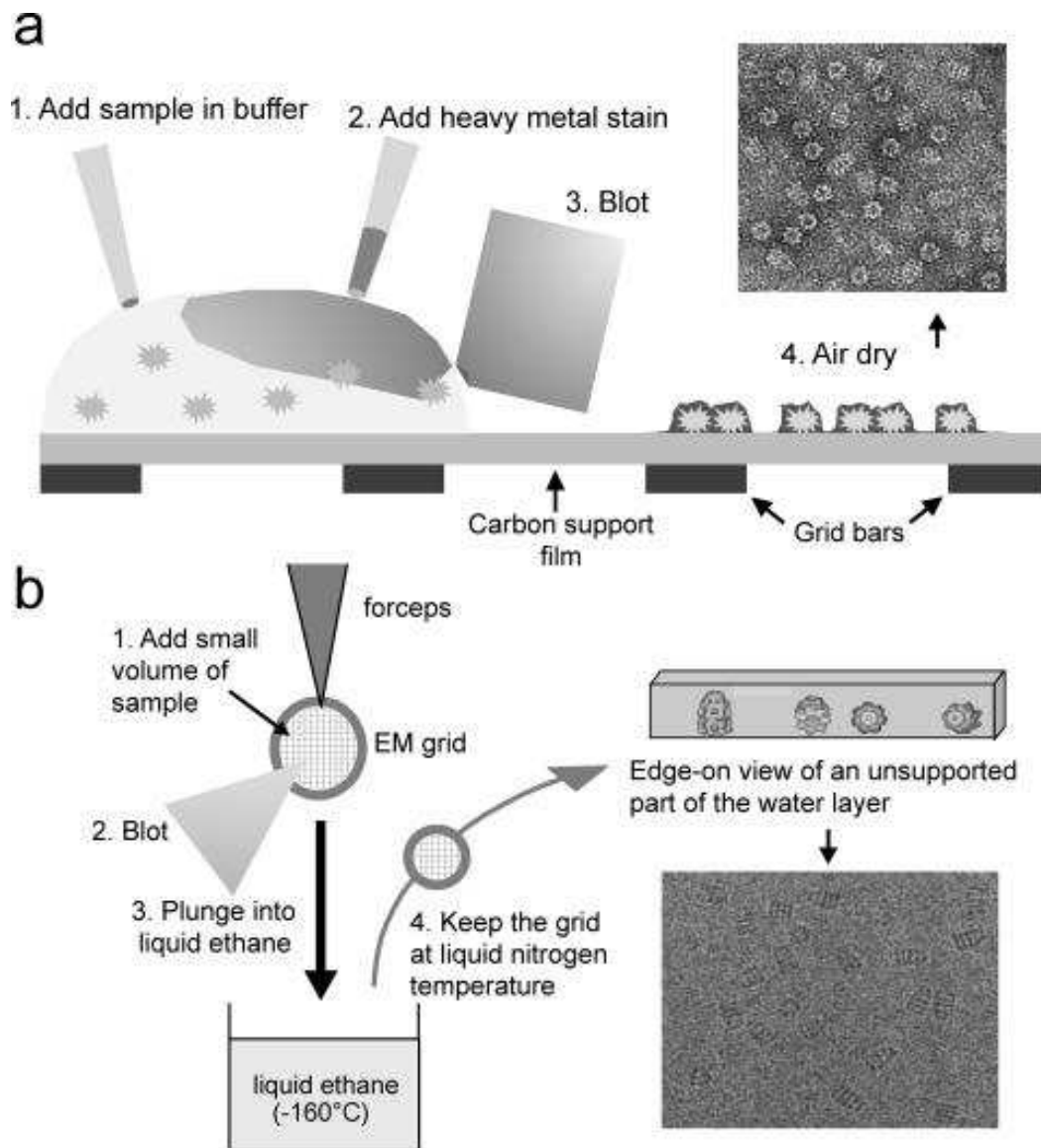


Figure 6: Biological sample preparation protocols: negative staining with heavy metal stain (a), and cryo-fixation (vitrification) by fast plunge freezing in liquid ethane (b). **Source:** (Orlova and Saibil, 2011).

sample. Instead, the process of vitrification occurs and the stable layer of amorphous, also called vitreous, glass-like ice is formed with the embedded sample molecules (Fig. 6). Vitreous ice is relatively transparent to electrons, however, vitrified sample not preserved from radiation damage, thus a special low-dose operational mode of the microscope is required (Fig. 7).

Typical cryo-EM sample preparation and image acquisition protocol is as follows. A small aliquot ($\sim 3\text{-}5\ \mu\text{l}$) of a purified sample is applied on a small holey carbon support film. The grid is discharged to reduce the hydrophobicity of the carbon and spread the sample uniformly. Next, the grid is fixed with the pairs of forceps and blotted with a filter paper. As a result, a very thin layer of sample solution is formed on the grid ($\sim 0.2\text{mm}$). The grid is rapidly plunged into pre-cooled cryo agent, usually ethane or propane (-183°C). If the plunging is fast, the ice vitrification occurs in less than 10^{-3} s (Berriman and Unwin 1994).

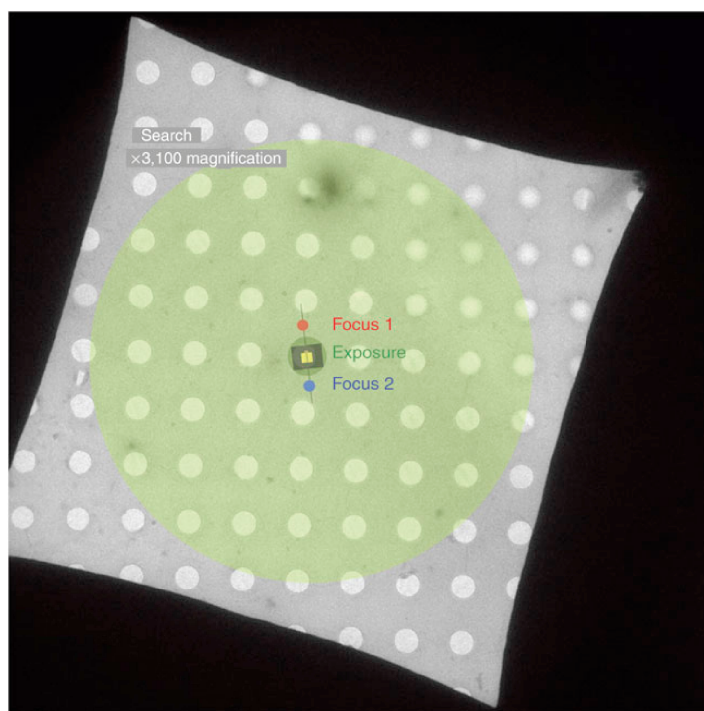


Figure 7: Low dose data collection mode. Search mode - locating the good sample at low magnification ($3,000\times$, dose $\sim 0.05\ \text{e}/\text{\AA}^2$); Focus mode - focusing, astigmatism correction and defocalization at high magnification ($100,000\times$, dose $\sim 15\text{-}20\ \text{e}/\text{\AA}^2$) $\sim 1\text{-}2\ \mu\text{m}$ apart from the sample area; Exposure mode - sample area is exposed at magnification $\sim 50,000\times$ and total dose of $<20\ \text{e}/\text{\AA}^2$. **Source:** (Grassucci et al., 2008).

The vitreous ice is a very metastable state and the frozen grid has to be handle in liquid nitrogen

during transfer to specimen cryo-holder (Cavalier, Spehner, and Humbel 2009). The cryo-holder is then inserted into the microscope. After the specimen stage stabilization (~30 min) to minimize thermal induced motion, the data collection is started. Standard low-dose dose protocol is used for the data collection (Fig. 7) (Grassucci, Taylor, and Frank 2008). Grid areas with good sample distribution and thin ice are located at low magnification (3,100x on FEI F20 TEM) under very low dose exposure (~0.05 e/Å² or less) in Search mode. Next, to preserve the sample from radiation damage the focusing and astigmatism correction is performed at high magnification (100,000x) in Focus mode ~1-2 μm apart from the sample area (Fig. 7). After focalization the defocus applied to generate contrast. The defocus level is determined by the sample size and contrast, desired resolution, and usually ranges from less than 1 μm to ~5 μm. The higher the defocus the better the low-resolution contrast, but at price of loosing high-resolution features. Several defocus groups of images required to cover a large defocus range without major gaps (Fig. 8). The sample than is exposed at magnification 25,000–50,000x in Exposure mode. The total dose is ~20 e/Å² or less, and image is recorded on film, CCD camera or direct detection device (DDD).

1.2.4. CTF estimation and correction

The collected and CTF-corrected images of the thin specimen are the true 2D projections of the 3D specimen density due to the large depth of field. Initial CTF estimation usually fulfills by inspecting calculated power spectrum (diffraction pattern, diffraction intensity) of the image, which also helps to assess the image quality (Fernández, Li, and Crowther 2006; Mindell and Grigorieff 2003). In cryo-EM the power spectrum of the image is its squared Fourier intensities:

$$F\{I^2(r)\}=M^2(s)=\Phi^2(s)CTF^2(s)E^2(s)+N^2(s) \quad (\text{Eq. 6}),$$

$M^2(s)$ is the computed diffraction pattern of the collected image, $\Phi(\vec{s})$ is the structure factor function (2D slice of 3D Fourier transform of the imaged specimen), $CTF(\vec{s})$ is a contrast transfer function, $E(\vec{s})$ is an amplitude decay (envelope) function, $N(\vec{s})$ is an additive incoherent (with the signal) noise term. In other words, $M(\vec{s})$ is the measured data, and $\Phi(s)$ represents the true structure factor of the imaged molecule, which we want to retrieve (Saad et al. 2001). CTF of the images is usually determined by cross-correlation between calculated and predicted CTF functions (Mindell and Grigorieff 2003). The most obvious and simplest CTF correction is the phase flipping, which inverts the phases of the negative regions of CTF (Fig. 8). More complex correction, both phases and amplitudes decays at high-resolution spatial frequencies of CTF – is the well-known Wiener filtration (Russ 2011). To extract signal loss near the zero crossings of CTF, many images with different defoci are averaged (Fig. 8).

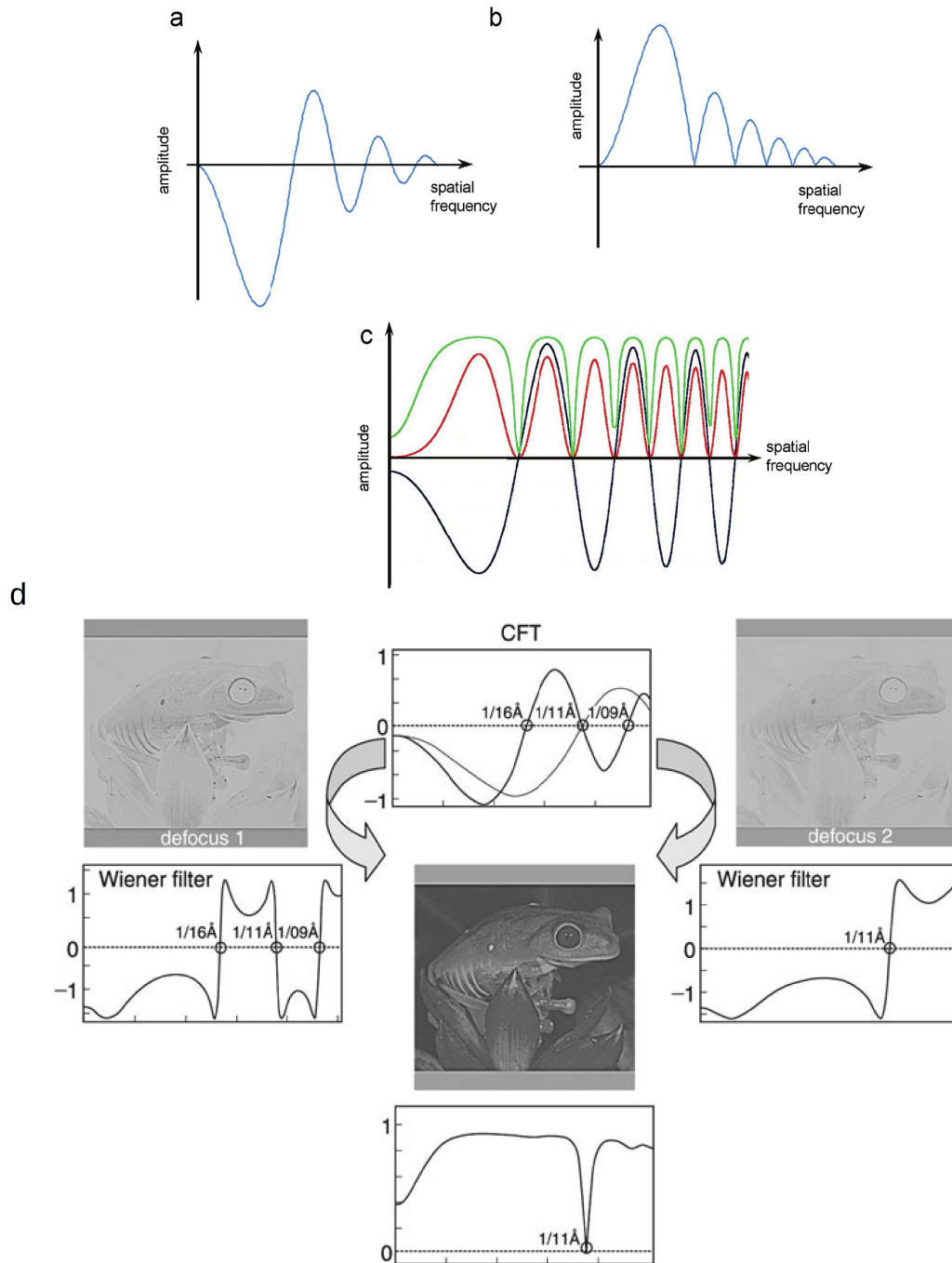


Figure 8: CTF for different types of images: raw images (**a**), phase-flipped images (**b**), phase- and amplitude-corrected images (**c**). Wiener correction of two raw images with different amounts of underfocus (**d**). Adapted from (Orlova and Saibil, 2011; Frank, 2006a)

The Wiener-corrected image is as follows:

$$M_{corr}(\vec{s}) = M(\vec{s}) \frac{CTF(\vec{s})}{CTF^2(\vec{s}) + 1/SNR(\vec{s})} \quad (\text{Eq. 7}),$$

where the first multiplier $M(\vec{s})$ is a measure data (computed diffraction pattern of the measured image), second multiplier $\frac{CTF(\vec{s})}{CTF^2(\vec{s}) + 1/SNR(\vec{s})}$ is a Wiener filter, $SNR(\vec{s})$ is a spectral signal-to-noise ratio:

$$SNR(\vec{s}) = \frac{P_{signal}(\vec{s})}{P_{noise}(\vec{s})} \quad (\text{Eq. 8}),$$

where $P_{signal}(\vec{s})$ and $P_{noise}(\vec{s})$ are the power spectra of the object and noise. The numerator of the Wiener filter corrects the phases of the image, while the denominator corrects the amplitudes. The second term in denominator prevents division by near-zero values of $CTF(\vec{s})$, viz excessive noise amplification (Frank, 2006). One can readily see that combination of multiple defocus groups or series of N images helps to avoid signal gaps in the neighborhood of the CTF zeros:

$$M_{corr}(\vec{s}) = \sum_{k=1}^N M_k(\vec{s}) \frac{CTF_k(\vec{s}) SNR_k(\vec{s})}{\sum_{k=1}^N CTF_k^2(\vec{s}) SNR_k(\vec{s}) + 1} \quad (\text{Eq. 9}),$$

where index k states for k-th image. Thus, with proper choice of defoci none of the zeros of $CTF_k(\vec{s})$ coincide, preventing signal gaps. In addition, the dampening of the signal at high spatial frequencies $E(\vec{s})$ has to be corrected (Fernández et al. 2008).

1.2.5. Averaging

After the phase and amplitude correction the images are true 2D projections of the 3D specimen. However, several pre-processing steps are required before averaging of noisy and low-contrast images of the specimen to increase the signal-to-noise ratio. Usually, only the best images (or micrographs) are selected based on power spectra inspection. Micrographs with the presence of astigmatism, drift in their power spectra, with poor or strange contrast or excessive contrast gradients should be excluded

Next, since each recorded micrograph usually contains many views, called “particles”, of the same specimen in different orientations, the best single particles should be selected. In case of the large or elongated samples like tailed bacteriophages, the “divide-and-conquer” approach could be

used, when different parts of the same particle are selected and further processed separately. The selection process is called particle picking, and manual, automated or semi-automated picking is available in the majority of the cryo-EM software packages. Finally, after creation of the stack of picked particles, they should be normalized, that is to set the mean density and the standard deviation of all particle images to the same level.

On Fig. 9 there is a Guinier plot of a protein, which shows the natural logarithm of the average structure factor amplitudes $\ln F$ as a function of resolution $1/d^2$ (Rosenthal and Henderson 2003). The plot is consisted of two main regions, $d > 10 \text{ \AA}$ low-resolution region and $d < 10 \text{ \AA}$ high-resolution region. Above the 10 \AA the amplitudes are determined by protein shape and solvent contrast and decreased strongly with the resolution. Below 10 \AA the amplitudes are determined by the random position of atoms constituent protein molecule, and to some extent by the fold-specific features. The α -helices are produced stronger amplitudes around 10 \AA resolution, while β -structures around 4.5 \AA . The average noise amplitude of one particle is shown on the right as

$$F_{\text{noise}(1)} , \text{ and after averaging of } N \text{ images is } F_{\text{noise}(N)} = \frac{F_{\text{noise}(1)}}{\sqrt{N}} .$$

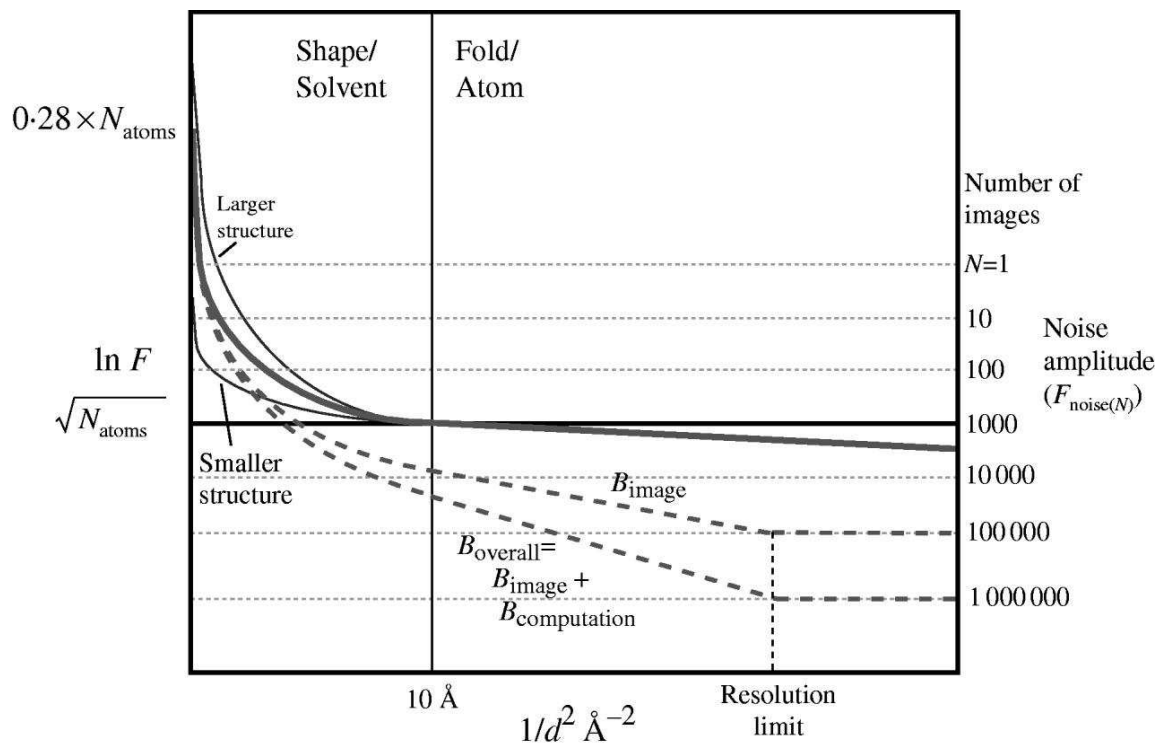


Figure 9: Guinier plot of a protein - natural logarithm of the average structure factor amplitudes $\ln F$ as a function of resolution $1/d^2$. **Source:** (Rosenthal and Henderson, 2003).

The possible achievable resolution for given number of images is the point where the structure factor curve intersects with the average noise level. The additional contrast lost due to the both imperfect images and incoherent averaging is shown as the dashed line with slope $B_{overall}$, the sum of temperature factors $B_{overall}$ and $B_{computation}$. Larger temperature factors requires a larger number of particles to achieve desired resolution. The effect of $B_{overall}$ may be shown as a $e^{-B_{overall}/4d^2}$ fall-off. The amplitudes decay at high spatial frequencies can be corrected by simply applying a negative temperature factor (sharpening), or more complex weighted sharpening according to the signal-to-noise ration of the 3D map at given resolution.

The particle mass limitation shown as a continuous thin line. Possible lower mass threshold is ~50 kDa, where the low-resolution amplitude is below the noise level even for the single image. In case of tomography, the maximum possible resolution is the intersection of the amplitude with the noise level for the single image recorded with the tomographic total dose $\sim(20-25 \text{ e}/\text{\AA}^2)$. As in single-particle cryo-EM, the averaging of the volumes extracted from reconstructed tomograms (sub-tomograms, or sub-volumes) will improve the possible achievable resolution (Schur et al. 2014).

1.2.6. Classification

For the 2D averaging to be reliable, the particles relative positions and orientations has to be known, and the particles can be aligned. Usually particles alignment (rotational and translational) is based on the maximum cross-correlation between each particle and some initial guess about the structure (projections of initial model) (Saxton and Frank 1976). After initial alignment, the particles can be averaged and resulting 2D average (class average) is used as the new initial guess (reference image) about the specimen structure, and the process is then iterated. This is so-called reference-base alignment method (Frank, 2006; Joyeux and Penczek, 2002).

However, even without initial guess about the structure the reference images can be generated by reference-free alignment. Also, reference-free alignment is served to avoid reference model bias problem, where noise correlated to the reference image (R. Henderson 2013; Stewart and Grigorieff 2004). During 2D alignment particles with similar orientations are grouped together and averaged within each group to produce the class average. To determine the unknown orientations and classify particles based on these orientations, a range of statistical approaches are available (Harauz, Boekema, and van Heel 1988; Joyeux and Penczek 2002). In case of sample heterogeneity a multi-reference approaches can be used (Scheres et al. 2005). An alternative approach to maximum cross-correlation method, especially in case of low SNR and sample heterogeneity, is a maximum likelihood (ML) alignment and classification (Scheres et al. 2005; Sigworth et al. 2010). The main idea is to find the most likely model given the observed data. This likelihood is measured

by the probability that the observations would be made, given the current model.

To determine the relative particles orientations different angle assignment methods are used. The projection matching requires an initial 3D model from which different reference projections can be generated, and compared with particles to find one with the best cross-correlation coefficient (Fig. 10) (P. A. Penczek, Grassucci, and Frank 1994).

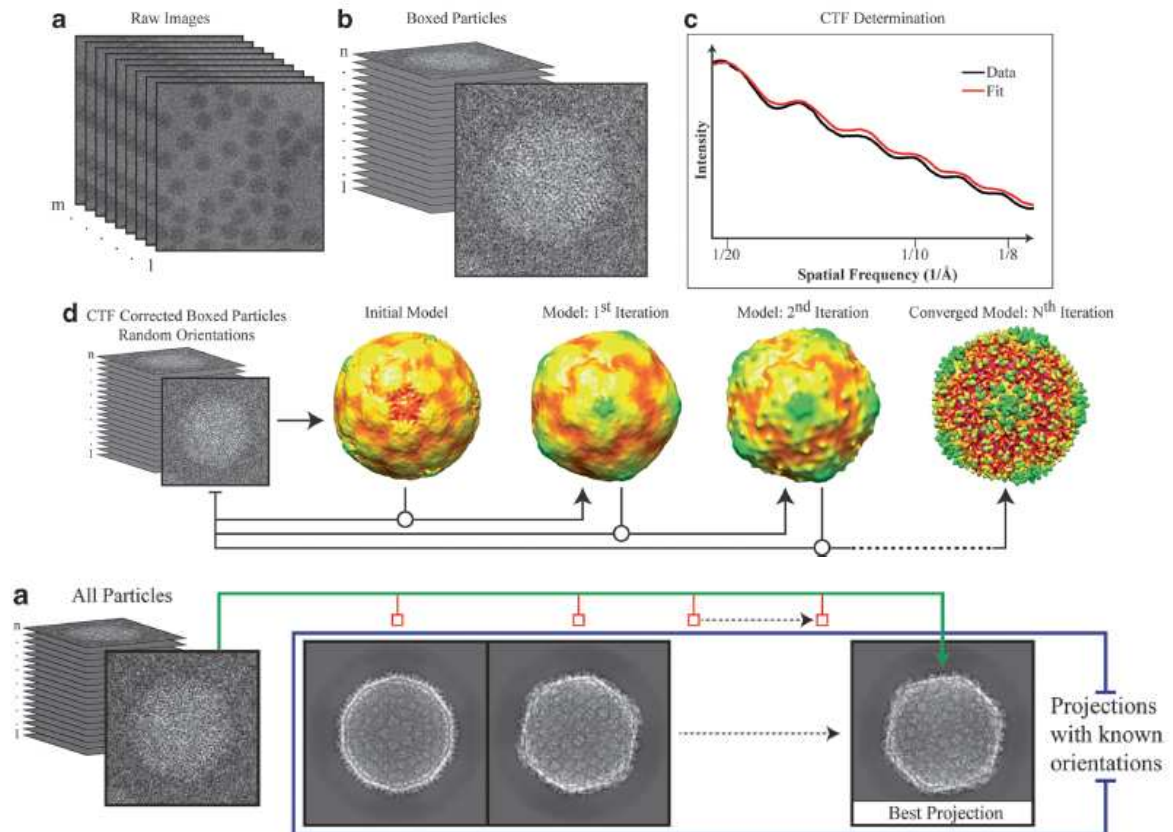


Figure 10: Generalized schematic for single-particle reconstruction (a)-(d). (Bottom) Projection matching approach for particles alignment. **Source:** (Chang et al., 2012).

Another approach is utilized the relationships between the projections in Fourier space. Since the 2D images are projections of the original 3D structure in certain orientation, the Fourier transform of this projection image is a central section of the 3D Fourier transform of the original 3D structure (Central section theorem). Two projections in Fourier space are intersected on so-called common line (Fig. 11). The presence of internal symmetries of the object results in multiple common lines, self-common lines and cross-common lines (Crowther 1971; Pawel A. Penczek, Zhu, and Frank 1996; Thuman-Commike and Chiu 2000; Van Heel 1987).

1.2.7. 3D real-space and Fourier reconstruction methods

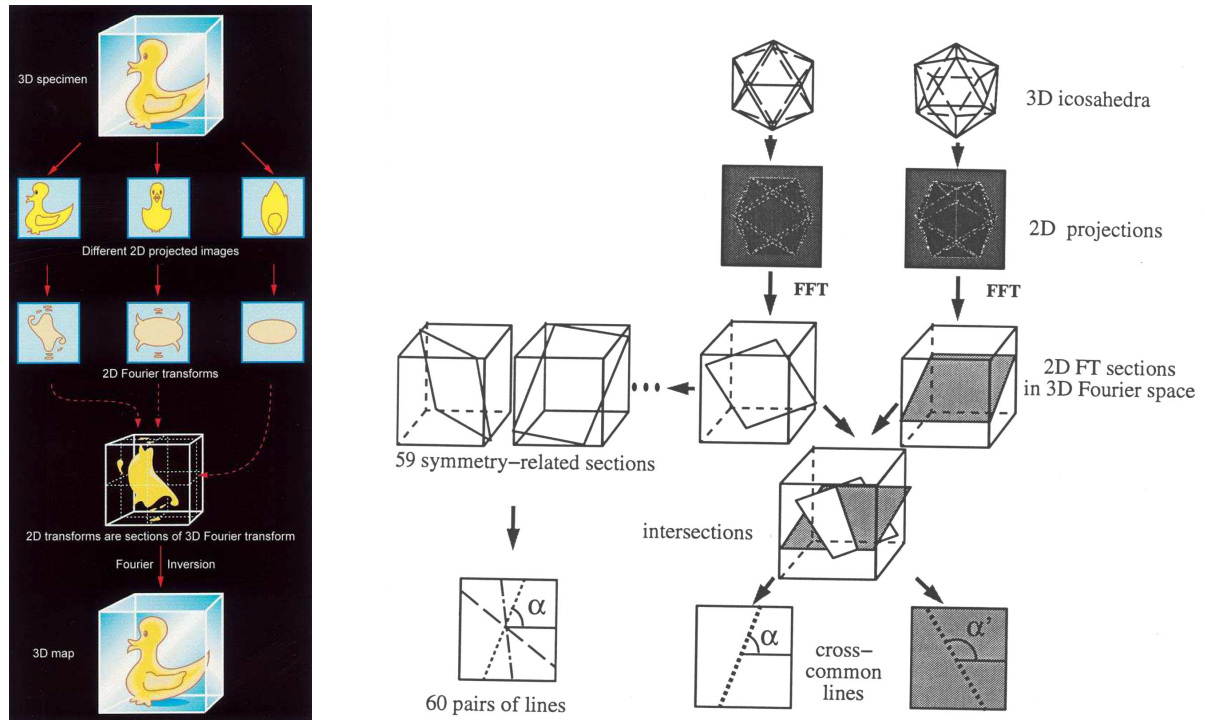


Figure 11: Implication of the central section theorem to 3D reconstruction from 2D projections (**left**). Common lines method for particles alignment (**right**). **Source:** (Baker and Henderson, 2006).

Finally, after the particles orientations has been determined the 3D reconstruction can be calculated from the 2D projections with known relative orientations.

The problem of 3D reconstruction from projections belongs to the class of incomplete, inverse, ill-posed problems. Inverse problems are more difficult to solve and usually only the approximation of the solution can be found (Pawel A. Penczek 2010a). Incompleteness comes from that part of the data was not observed experimentally, that is only a finite number of projections was observed. It was shown that the original 3D object cannot be recovered uniquely (Pawel A. Penczek 2010a). Ill-posedness means that the experimental data alone is not enough to determine a unique solution.

The reconstruction methods are divided into two main groups, namely algebraic and transform methods, both real and Fourier space. The real space methods include back-projection, or filtered back-projection methods (Frank 2006a; Herman 2009). Another part of real-space methods are algebraic iterative methods like algebraic reconstruction technique (ART), or simultaneous iterative reconstruction (SIRT), or others (Pawel A. Penczek 2010a). Algebraic methods lead a reconstruction problem into a system of linear equations, which are usually overdetermined. Algebraic methods

are applicable to different data collection geometry, and uneven distribution of projections.

Fourier methods include direct Fourier inversion, and its derivatives, like gridding-based direct Fourier reconstruction algorithm (Bracewell 2003; Pawel A. Penczek 2010a; Pawel A. Penczek, Renka, and Schomberg 2004). Objects with high symmetry, such as helical filaments, or icosahedral viral capsids usually reconstructed with modified versions of Fourier inversion. For example, Fourier-Bessel reconstruction is used for specimens with rotational or helical symmetry.

All reconstruction algorithms have to deal with several EM specific constraints. The projection distribution is random and uneven and cannot be approximated by an analytical function, which results in gaps in Fourier space. This makes the reconstruction problem not invertible, and introduces reconstruction artifacts. Also, presence of CTF artifacts and uneven distribution of spectral signal-to-noise ratio should be taken in account. Uneven data distribution should be reconstructed on evenly sampled 3D grid, thus interpolation and proper data sampling is required to prevent interpolation artifacts. Usually, the pixel size is selected $\sim 1/3$ of the desired resolution. Presence of the random and systematic errors from the orientations assignment requires an iterative refinement of 3D reconstruction.

1.2.8. Electron cryo-tomography (cryo-ET)

Single-particle analysis represents one of the two different approaches to 3D imaging of biological specimens. If the specimen is structurally homogeneous, than different 2D images represent different particle orientation in vitrified solution. Many particles can be average to enhance signal-to-noise ratio, and high resolution 3D representation of original specimen may be retrieved. However, for highly heterogeneous specimens like cells or their organelles, unique structural heterogeneity and flexibility are averaged out. For example, single-particle analysis is not suitable to study viruses without uniform structures, like non-icosahedral envelopes or decoration proteins of icosahedral capsids. Viruses infecting bacterial hosts are perfect examples of dynamic system with nonuniform properties since each viral particle is in different stage of infection process (Hu et al. 2013; J. Liu et al. 2011). Electron cryo-tomography (cryo-ET) is the leading method to study such unique biological specimens. Instead of collecting and averaging of many particles with various views in solution, we collecting entire projection set (tilt series) of one particle by tilting the stage of the microscope around a single-tilt axis at different angles (Lucić, Förster, and Baumeister 2005)). Typically angular semi-sphere is reduced to $\pm 70^\circ$ or even 60° due to the specimen holder limitation and ice thickness at higher tilt angles, with a constant 1° - 2° increment (linear tilt scheme), or with an increment proportional to the cosine of the current tilt angle (Saxton scheme) ((Saxton, Baumeister, and Hahn 1984). However, there is no escape from radiation damage limitation, now for the whole tilt series. Biological specimen can sustain limited dose, usually 20-25 $e/\text{\AA}^2$ for single-particle cryo-EM, which must be distributed over the tilt series by less than 1 $e/\text{\AA}^2$

per image. Collected images are extremely noisy with a very low contrast, and are difficult to align. Translational alignment of tilt series compensates the small specimen movement within the image during tilting. Noisy, low-contrast images are not suited for typical cross correlation-based or common lines-based alignment (Guckenberger 1982; Y. Liu et al. 1995). Another approach is to use a high-density markers (fiducials), usually a colloidal gold particles 10-20 nm wide mixed with specimen before freezing (D. N. Mastronarde 1997). Since the positions of the fiducial markers are fixed relative to the specimen, the individual images are aligned by least-squares procedures (Frank 2006b). However, marker-free alignment algorithms are currently developing (Castaño-Díez et al. 2010; Castaño-Díez et al. 2007). CTF estimation and correction in cryo-ET is difficult and not clearly defined (Fernández, Li, and Crowther 2006).

Limited tilt range results in 3D reconstruction degradation by the artifact called missing wedge

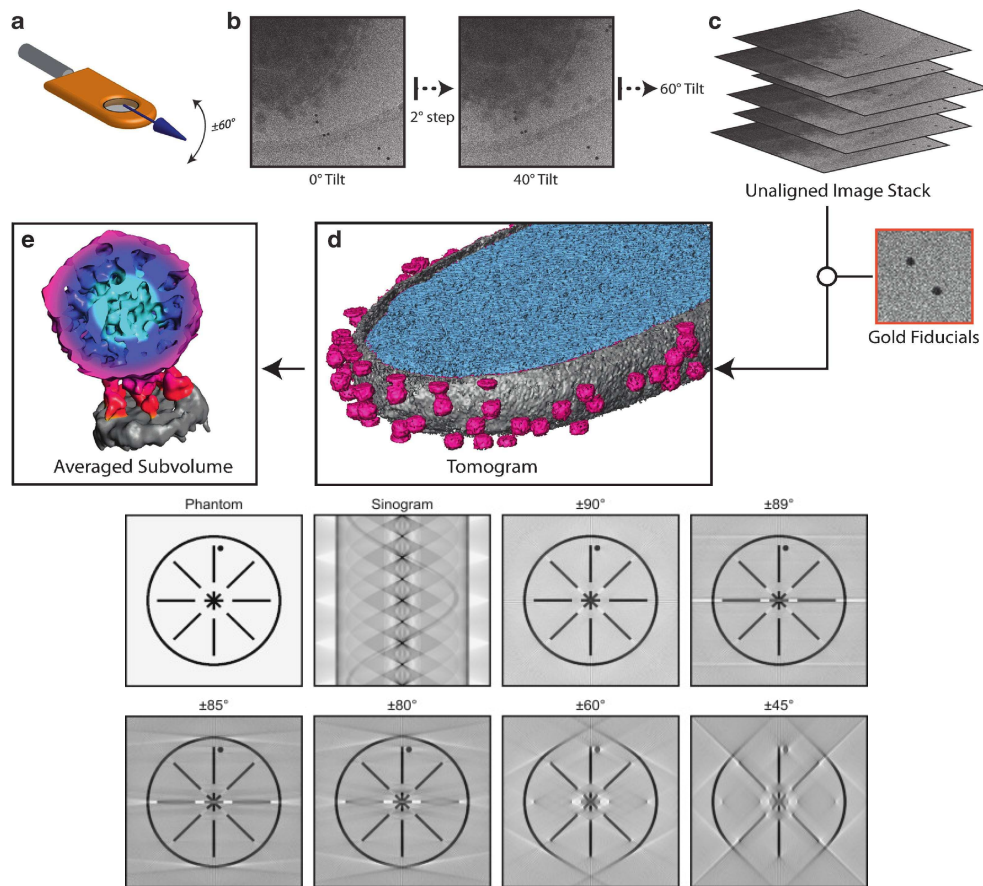


Figure 12: Generalized schematic for a tomographic reconstruction (a)-(d) and sub-tomogram averaging technique (e). Missing wedge artifacts for different tilting geometries (bottom).
Source: (Chang et al., 2012; Palmer and Löwe, 2014).

(Fig. 12), which is a wedge-shaped region in the Fourier space, empty of information. Missing wedge causes non-isotropic resolution, and reconstructed specimen appears elongated. To compensate for the missing wedge and low signal-to-noise ratio of reconstructed tomograms, the sub-tomogram averaging technique is widely used. The idea is the same as for single-particle cryo-EM, that is to detect identical compounds, align, classify and average them (Hrabe and Förster 2011; Schmid 2011). Architecture of different biological specimens has been studied by sub-tomogram averaging (Bui et al. 2008; Schur et al. 2014)

Most popular software packages for tilt series alignment, tomography reconstruction and sub-tomogram averaging include the IMOD package with PEET utility (Kremer, Mastronarde, and McIntosh 1996; Nicastro et al. 2006), Bsoft (Bernard Heymann et al. 2008), EMAN2 (Tang et al. 2007, 2), SPIDER (Renken et al. 2009). Maximum-likelihood approach now available in XMIPP package (Scheres et al. 2009). Dynamo package is dedicated completely to the sub-tomogram averaging (Castaño-Díez et al. 2012). The aligned tilt series, possibly CTF corrected, are then reconstructed into a tomogram to recover a three-dimensional information of the specimen (Fig. 12). The reconstruction methods include the most popular real-space weighted back-projection (WBP), or Fourier-space methods based on the Radon projection theorem (Frank 2006b). More robust iterative algebraic reconstruction algorithms, ART and SIRT, are getting increasing interest. However, they are computationally expensive (Gilbert 1972).

1.2.9. Resolution estimation in cryo-EM

Resolution estimation in single-particle EM is not straightforward, and currently is actively debated (Richard Henderson et al. 2012). Subjective description of the resolution is the level of reliable details which are detectable in the final reconstruction (Pawel A. Penczek 2010b). The most commonly used method is the Fourier shell correlation (FSC), a function of resolution between two independent maps, each calculated from the half of the data (Pawel A. Penczek 2010b; Rosenthal and Henderson 2003; van Heel and Schatz 2005). However, it is still debated which threshold value of the FSC should be used. The most common criteria is the spatial frequency where FSC function falls to 0.5, however other SNR-based criteria, $FSC=0.143$, has been proposed (Rosenthal and Henderson 2003). Another criteria, is the spectral signal-to-noise ratio (SSNR) (Pawel A. Penczek 2010b). This approach does not require calculation of two independent maps, instead it calculates the correlation between the signal and the noise of the map. It is useful approach in case of small number of picked particles, for example, large bacteriophages. Regardless of the threshold used, α -helices become visible at <8 Å resolution range, and their grooves at <5 Å resolution range.

2. Bacteriophages ϕ Eco32 and 7-11

2.1. Summary

Bacteriophage ϕ Eco32 is a novel, lytic virus isolated in 2004 as a phage that infects 95% of mastitis-causing strains of *Escherichia coli*. It is a member of the *Podoviridae* family with the rare C3 morphotype, characterized by unusually long heads (capsids). The mature ϕ Eco32 virion contains at least 16 different proteins and consists of a capsid, containing the genome and short non-contractile tail of remarkable complexity with two types of multicomponent fibers attached. The elongated capsid and bow tie-like arrangement of fibers around the tail give the phage particle the appearance of a torpedo. Bacteriophage 7-11 is a virus from *Podoviridae* family that infects *Salmonella* strains. It has the same as ϕ Eco32 C3 morphotype and overall morphology resembles phage ϕ Eco32. The 7-11 virion contains at least 14 different proteins, which builds the elongated capsid and short non-contractile tail with two types of radially protruding fibers. However, the capsid is more elongated, and the tail has simple topology compare to that of ϕ Eco32. Such capsid elongation of ϕ Eco32 and 7-11 phages and their tail topologies raise the question about DNA packaging, and virus-host cell interaction.

Here we report the first detailed three-dimensional structural organizations of a *E.Coli* and *Salmonella* C3 podophages.

2.2. Introduction

Bacteriophages or phages are bacterial viruses. There are the most diverse and abundant form of life on our planet (Hendrix et al. 1999; Wommack and Colwell 2000). Bacteriophages have developed various strategies to infect a bacterial host (H-W Ackermann 2003; T. S. Baker, Olson, and Fuller 1999).

Phages from *Caudovirales* family have a special multi-functional organelle, 'tail', which have been defined as the complex of proteins in phage virions involved in bacterial host recognition, attachment and genome delivery into the cell (H-W Ackermann 2003). These tailed phages are subdivided into *Myoviridae* phages with a long contractile tails (e.g. bacteriophage T4) (Browning et al. 2012; Petr G. Leiman and Shneider 2012b) *Siphoviridae* phages with a long but non-contractile tails (e.g. bacteriophage λ), and *Podoviridae* phages with a short tails (e.g. bacteriophages K1-5, and ϕ 29) (H-W Ackermann 2003; Petr G Leiman et al. 2007). For *E.Coli* phage T4 and other myoviruses with contracted tails, the contraction process is well-studied (Petr G Leiman et al. 2004). The majority of the studied tailed bacteriophages believed to have two stages of the infection mechanism: initial binding to a cell surface primary receptor by phage tail spikes or fibers, and then, interaction with a host receptor protein integral to the outer membrane.

For dsDNA bacteriophages without a contractile tail and tail tube, such as podophages and siphophages, their life cycles have not yet been studied in detail. Interestingly, the DNA transport mechanism (DNA “ejection”) for such a phages has been obscure.

Bacteriophage ϕ Eco32 was isolated in Tbilisi, Georgia, in 2004. It infects more than 95% of bovine mastitis-causing strains of *E.Coli* (Savalia et al. 2008). It is a member of the *Podoviridae* family with a rare C3 morphotype, which occurs in <1% of phage virions (H.-W. Ackermann 2006; Savalia et al. 2008, 5555). Bacteriophages of the C3 morphotype characterized by elongated capsids with an elongation (length-to-width) ratio more than 2.5. ϕ Eco32 virion consists of a prolate capsid, 1360 Å x 460 Å (elongation ratio is 3.22), lacking distinct faces (Fig. 13, 12), containing viral DNA and short non-contractile tail to which two types of multi-component fibers are attached. Prolate capsid architecture observed in about 15% of studied bacteriophages (H. W. Ackermann 1998). The genome is double-stranded DNA (dsDNA) of 77,554 base pairs. Less than 40% of encoded proteins display sequence homology to known proteins, point to the discover of the truly novel *E.Coli* phage (Savalia et al. 2008). The short non-contractile ϕ Eco32 tail is 300 Å long and 460 Å wide with a complex topology (Fig. 13, 16, 17). Two different types of kinked tail fibers are folded in an intertwined manner around the periphery of a central stub-like structure. Each fiber is likely to be composed of 2 to 3 proteins. There is a prominent needle, protruding from the center of the tail.

Bacteriophage 7-11 was isolated in Quebec, Canada, with its host *Salmonella entericac serovar* Newport (Kropinski, Lingohr, and Ackermann 2011). It is a C3 podovirus, which mature virion consists of a prolate capsid 1650 Å x 460 Å (elongation ratio is 3.57), also lacking distinct faces (Fig. 13). The capsid encapsulates a double-stranded DNA genome of 89 916 base pairs. Only ~15% of encoded 151 ORFs showed homology predominantly to ϕ Eco32 proteins (Kropinski, Lingohr, and Ackermann 2011). There are two types of straight tail fibers protruding radially from the short tail structure. The tips of the fibers are disordered pointing to the high level of flexibility compared to ϕ Eco32 fibers. Tail ends with a very short tail needle (Fig. 16, 17). In this report, we demonstrate the detailed three-dimensional organization of ϕ Eco32 virion and 7-11 tail by cryo-EM.

2.3. Results and discussion

2.3.1. Cryo-EM reconstruction of the ϕ Eco32 and 7-11 virions

Enormous capsid elongation of ϕ Eco32 and 7-11 virions forced us to use the “divide-and-conquer” approach for image processing of cryo-EM micrographs. ϕ Eco32 capsid and tail reconstructions and 7-11 tail reconstruction were calculated separately, assuming the most appropriate symmetry for each part (Fig. 13). The resolution of the ϕ Eco32 capsid and the tail (FSC 0.143 cutoff) is 20 Å and 15 Å, and of the 7-11 tail is 18 Å. Asymmetrical reconstruction of the ϕ Eco32 tail region with a

part of the capsid was calculated to establish their relative orientations. Absolute hand of the reconstructed ϕ Eco32 tail was determined by comparing with the hand of the portal X-ray model, and further proved with Tilt pair validation server (Wasilewski and Rosenthal 2014). Sub-volumes containing empty ϕ Eco32 virions were extracted from cryo-EM tomograms, aligned and averaged to highlight the internal organization of empty capsid (Fig. 14E(blue), 15B(6)).

2.3.2. Structure of the capsid and genomic DNA

2.3.2.1. ϕ Eco32

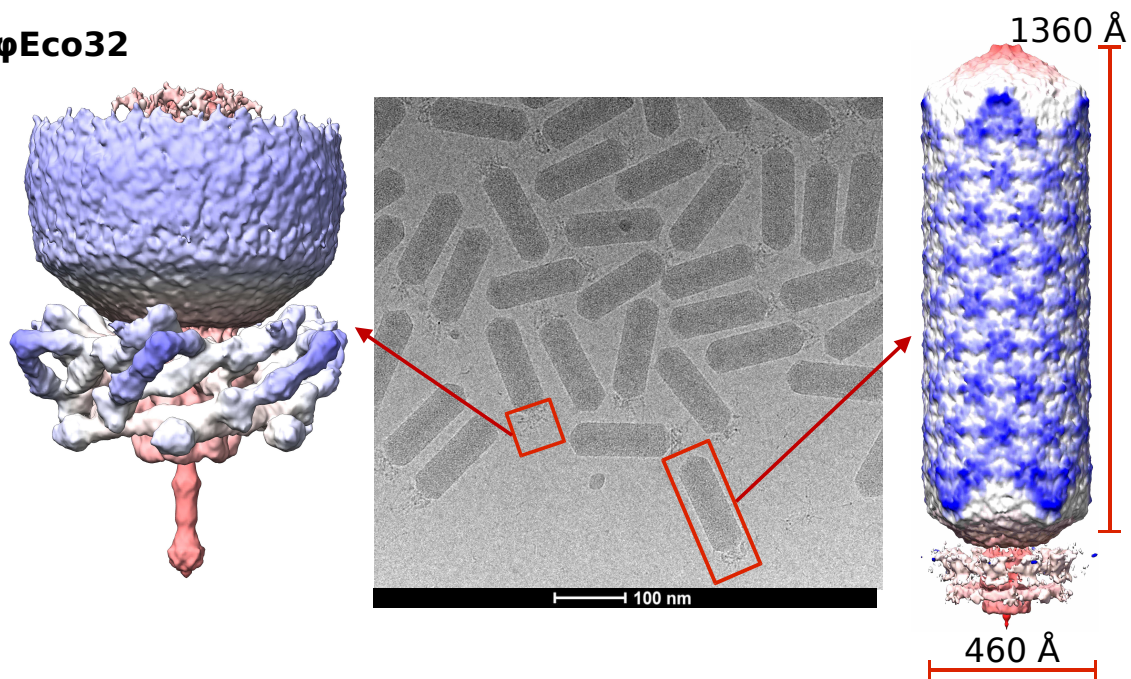
The ϕ Eco32 capsid shell is a prolate icosahedron elongated along 5-fold axis, 1360 Å long from one 5-fold to the opposite 5-fold vertex, and 460 Å wide, without distinct faces (Fig. 13, 14 A,B). Such prolate organization of a phage capsid rarely occurs among studied bacteriophages (H. W. Ackermann 1998) and raises the questions of dsDNA packaging, as well as a phage-host cell interaction.

ϕ Eco32 prolate head consists of two icosahedral caps follow $T=4$ triangulation symmetry. The equatorial midsection is ~ 1100 Å long and has $Q=20$ symmetry. The total number of major capsid protein (MCP) gp11 monomers forming the capsid, N , is given by $N = 30(T + Q) - 5$ (Aebi et al. 1974). The first term corresponds to the total number of protein subunits of phage capsid, and the last term accounts for one vertex that is occupied by the connector. Thus, the total number of the gp11 monomers in the ϕ Eco32 capsid is 715 (Fig. 14 A, B). The equatorial midsection consists of 540 copies of MCP, organizing in 9 horizontal layers of 10 hexamers each. Consequently, the top and the bottom icosahedral caps are composed of 90 and 85 copies of MCP.

HK97 capsid protein hexamers can account for a majority of the ϕ Eco32 capsid shell density, showing that the structures of the HK97 and ϕ Eco32 capsid hexamers are similar. However, ϕ Eco32 capsid pentamers are more convex than HK97 pentamers. *HHpred* search indicates the HK97 MCP (PDB ID 1OHG) as a closest ortholog of the ϕ Eco32 MCP. However, the ϕ Eco32 MCP gp11 length is 352 residues compared to HK97 MCP 285 residues, which probably indicates the domain which is not presented in HK97 MCP.

The HK97 capsid protein structure was fitted into the ϕ Eco32 capsid shell electron density using “rigid body” fitting (Fig. 14 B). The capsid shell density easily accommodates the respective oligomers of HK97 capsid protein. The distance between hexamer centers in electron density is ~ 127 Å, which, on average, was observed in HK97 (132 Å) and many others dsDNA tailed phage (Aebi et al. 1974; Petr G Leiman et al. 2007; X. Liu et al. 2010; Morais et al. 2005; Thuman-Commike and Chiu 2000; Wikoff et al. 2000), and that between equivalent positions in pentamers and adjacent hexamers is ~ 132 Å.

ϕ Eco32



7-11

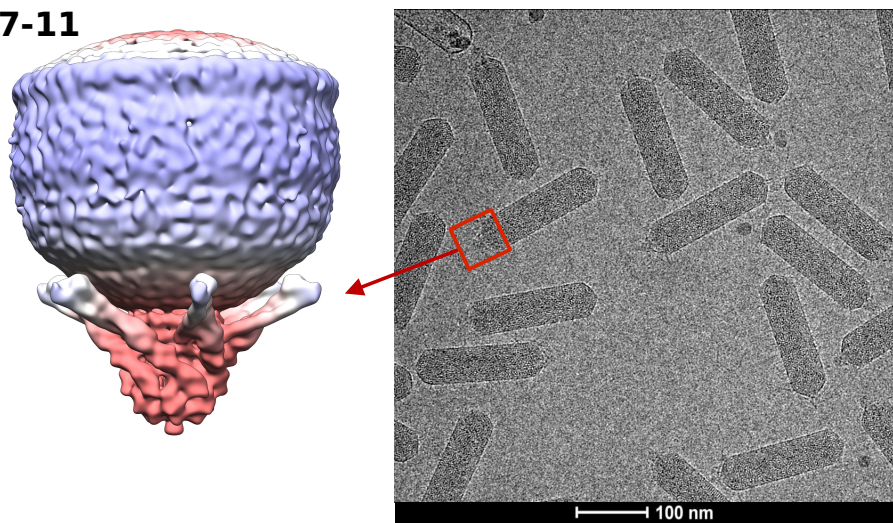


Figure 13: Cryo-EM structures of ϕ Eco32 and 7-11. **(Top)** A raw cryo-EM image of ϕ Eco32. Two separate reconstructions (“divide-and-conquer” approach) were calculated for areas highlighted with boxes colored in red. **(Bottom)** A raw cryo-EM image of 7-11. Tail reconstruction was calculated for area highlighted with box colored in red. All models colored according to the distance to the symmetry axis.

The major capsid protein of ϕ Eco32 is gp11 (39 kDa). About 20% of gp11 is likely to carry an extension IG-like domain at the C terminus, which is encoded by gene 12 and is translated because of a programmed frame shifting (Savalia et al. 2008). Similar domains are found in structural proteins of many phages (Fraser et al. 2006). However, the additional capsid decoration densities, which might corresponds to the outer capsid protein gp12, are not visible in our cryo-EM reconstruction.

Central section of cryo-EM reconstruction shows that the ϕ Eco32 capsid shell is ~ 30 Å thick on average, in a map contoured at 1.5 standard deviations (σ) from the mean, and all of the capsid is required to accommodate the genomic DNA (Fig. 14 C, D). There are clearly visible 18 layers of altering high and low density and the central straight lengthwise high density, which terminates in the portal complex corresponding density. Those densities corresponds to the 9 layers of densely packed dsDNA (Fig. 14 E(red)). They are arranged in concentric layers separated on average by 24 Å, which is common for many other *Caudovirales* phages (Cerritelli et al. 1997; Fokine, Kostyuchenko, et al. 2005).

Examination of cryo-EM reconstruction and raw micrographs reveals the layers of high density from the inner side of ϕ Eco32 empty capsid shell, ~ 26 -30 Å thick and separated by ~ 111 Å distance (Fig. 15 A, B). These layers are organized like hoops in a barrel, banding the inner shell surface, with a contrast similar to viral DNA (Fig. 15 A(4,5), 15 B(3,4,5)), and are likely to be the first packaged DNA rings. The majority of ϕ Eco32 empty virions were found having 10 internal layers space-filling the horizontal concavities forming by the MCP hexamers, connecting the imaginary hexamers centers (Fig. 15 A(4,5), 15 B(3,4,5)). However, different conformations of DNA layers were found. As an example, the partially DNA filled capsid with more frequent vertical distribution of internal layers of DNA (Fig. 15 A(1,2)), capsid with layers filling the diagonal MCP hexamers concavities (left- and right-handed) (Fig. 15 A(3,4)), capsids with less than 10 internal layers and loose hanging DNA (Fig. 15 A(5)), and completely empty phage capsids (Fig. 15 A(6)). These conformations likely correspond to distinct sequential states in DNA leakage process, from hexagonal (Molineux and Panja 2013) arrangement of the DNA domains (Fig. 13, 15 A(1), 15 B(5)), through the less ordered cholesteric state (Fig. 15 A(2)) to the isotropic DNA arrangement (Fig. 15 A(3,4,5)), and finally to the empty capsid (Fig. 15 A(6)).

Figures 15B overview all found ϕ Eco32 capsid shell internal conformations, as well the damaged capsid shells with the DNA came out. Averaged sub-volumes of ϕ Eco32 empty capsids also indicate the presence of outer DNA rings following the concavities of MCP hexamers (Fig. 15 B(6), 20 A). The diameter of the external ring of packaged DNA is ~ 375 Å, and compared to DNA persistence length ~ 500 Å, indicates the mechanically stressed DNA conformation.

Such DNA arrangement highlights a possible inverse spool-like structure with the spool axis along

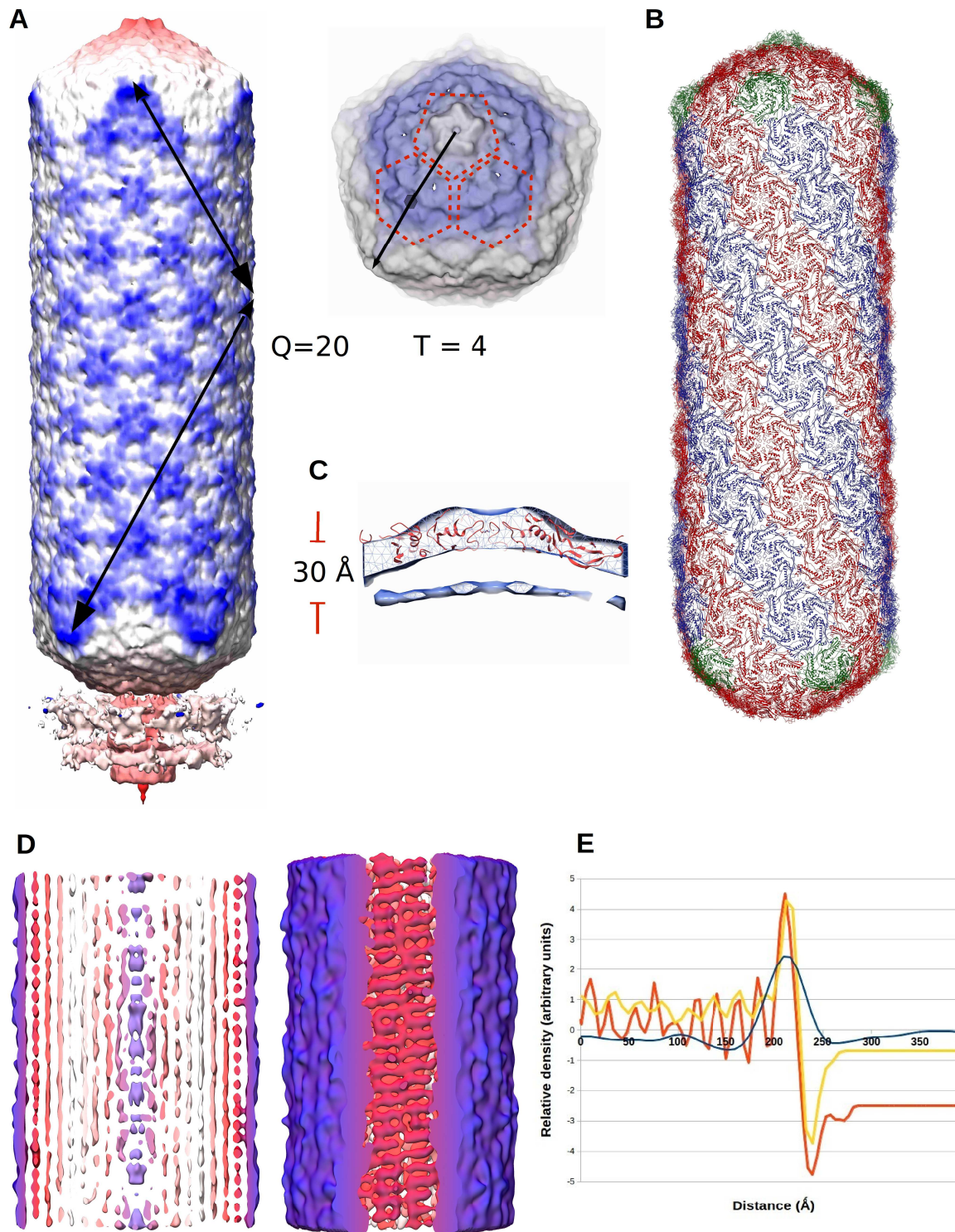


Figure 14: Cryo-EM structure of ϕ Eco32 capsid. **(A)** The structure of ϕ Eco32 capsid with the imposed C5 symmetry, and smeared density of the tail region. **(B)** Pseudo atomic capsid model with the HK97 capsid protein structure fitted. **(C)** Section of the capsid shell cryo-EM density with HK97 structure fitted. **(D)** Sections of the capsid cryo-EM density with DNA layers visible. **(E)** Radial profiles of the central sections of the cryo-EM structures of ϕ Eco32 (red), empty ϕ Eco32 (blue) and 7-11 (yellow) capsids.

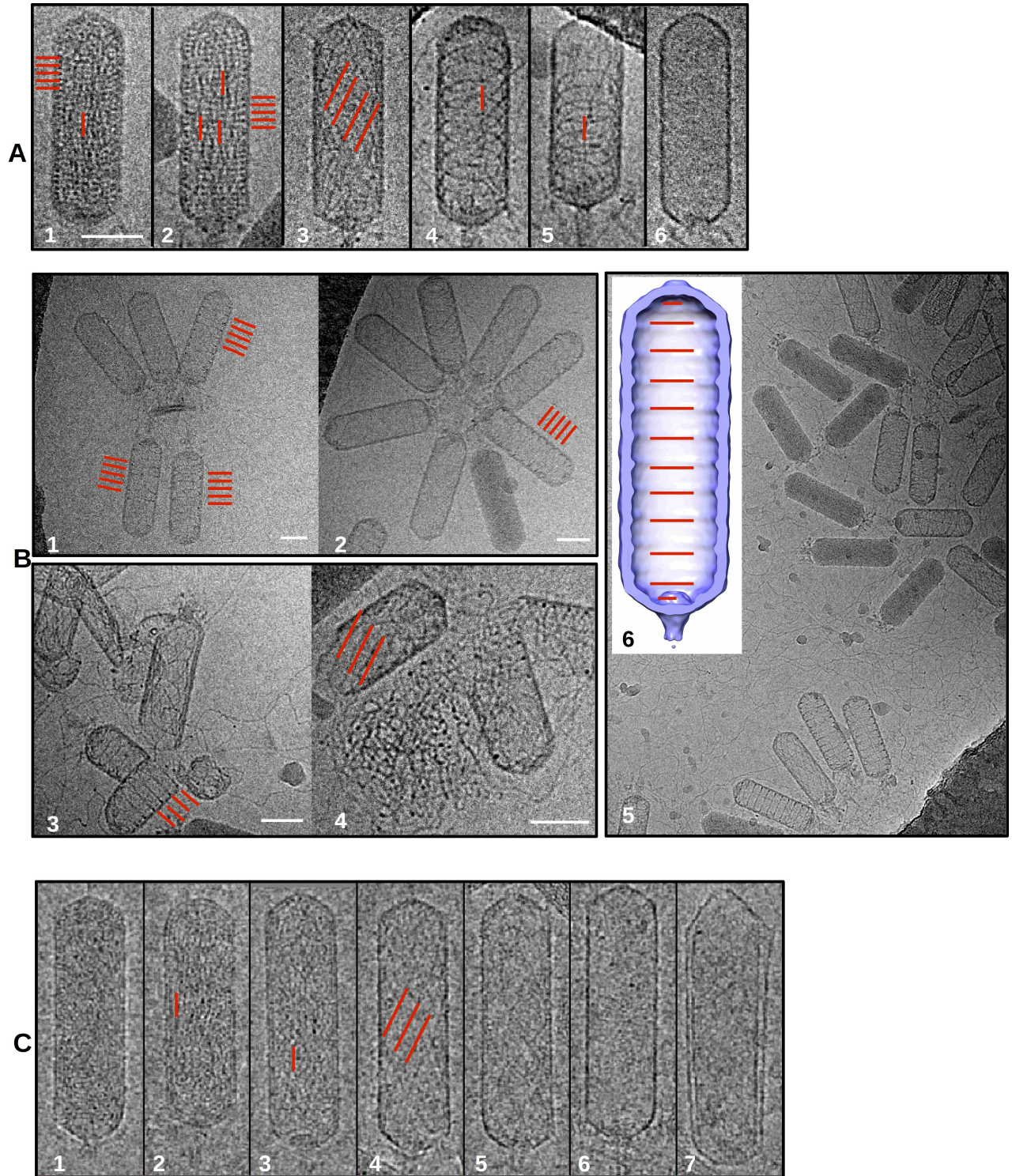


Figure 15: A raw cryo-EM images of ϕ Eco32 and 7-11, DNA layers highlighted with lines colored in red (scale bar - 460 Å). **(A)-(B)** Cryo-EM micrographs showing the structure of ϕ Eco32 DNA as it leaves the capsid. Slice of the cryo-ET structure (sub-volumes averaging) of empty ϕ Eco32 capsid with horizontal DNA layers highlighted. **(C)** Cryo-EM micrographs showing the structure of 7-11 DNA as it leaves the capsid

the capsid elongation axis, with DNA is wrapped axially in concentric shells, with a short segment of terminal DNA passing through the axis of the core and portal (Fig. 15 A(4,5)).

The density corresponding to internal core structure probably extends toward the whole capsid elongation axis, which may help orient the DNA spool (Fig. 14 D), but not visible in the cryo-EM reconstruction. The core structure built in portal complex (12 copies of gp8, see Tab. 1), with the central channel diameter of 35 Å. Portal complexes of all dsDNA viruses are dodecamers (12-fold symmetry) and have a similar cone-like shape, and are probably homologous to that of ϕ 29 phage (Fokine et al. 2004; Simpson et al. 2000) (Simpson et al., 2000). Portal complex is essential for phage assembly, genome packaging and DNA ejection. The ϕ Eco32 portal is a dodecamer which seems to be complexed with the core, and sits in one of the distal ends of prolate capsid. The 12-fold symmetry of the wing domains of the portal is clearly visible in the tail map, although only sixfold symmetry was imposed during reconstruction (Fig. 16 A). The narrow end of the portal protrudes from the ϕ Eco32 capsid and binds to the upper part of gatekeeper and possibly interacts with the long tail fibers. The portal complex is encompassed by a few rings of density, possibly corresponds to external DNA strands (Fig. 16 A). The gp8 orthologous full length P22 phage portal complex (PDB ID 3LJ5) was fitted into ϕ Eco32 cryo-EM density (Fig. 16 B). The left-handed twist of the cryo-EM density of the portal barrel domain coincides the twist P22 orthologous crystal structure twist, indicating the absolute hand of cryo-EM reconstruction (Fig. 22 C). To confirm the absolute hand the Tilt Pair validation server was used, which shows hand coincidence with the P22 portal crystal structure (Fig. 22C). Remarkably, that the averaged sub-volumes of empty virions has the same tail twist visible on cryo-ET reconstruction (Fig. 15 B(6)).

The DNA translocation channel diameters in the cryo-EM density of mature ϕ Eco32 particle and the channel in the crystal structure of isolated P22 full connector both are 35 Å in diameter. Averaged sub-volumes of ϕ Eco32 empty capsids shows the presence of portal core, but without the portal barrel structure (Fig. 15 B(6)).

The relative orientation between fivefold capsid and sixfold tail was determined by asymmetrical reconstruction. The symmetry relaxation and absence of the ϕ Eco32 top views results in preferred orientation artifacts and lower resolution structure, compare to the tail structure described below. However, the resulting density clearly indicates a fixed orientation between the fivefold capsid and the sixfold symmetric tail (Fig. 22 A, B, D), the situation observed in many tailed phages.

2.3.2.2. 7-11

The 7-11 capsid shell is a prolate icosahedron elongated along 5-fold axis, 1650 Å long from one 5-fold to the opposite 5-fold vertex, and 460 Å wide, also without distinct faces (Fig. 13, 15C, 20C). 7-11 prolate head consists of two icosahedral caps and the equatorial midsection, which is about 1300 Å long. Contrary to ϕ Eco32 bacteriophage 7-11 capsid composed of two distinct capsid

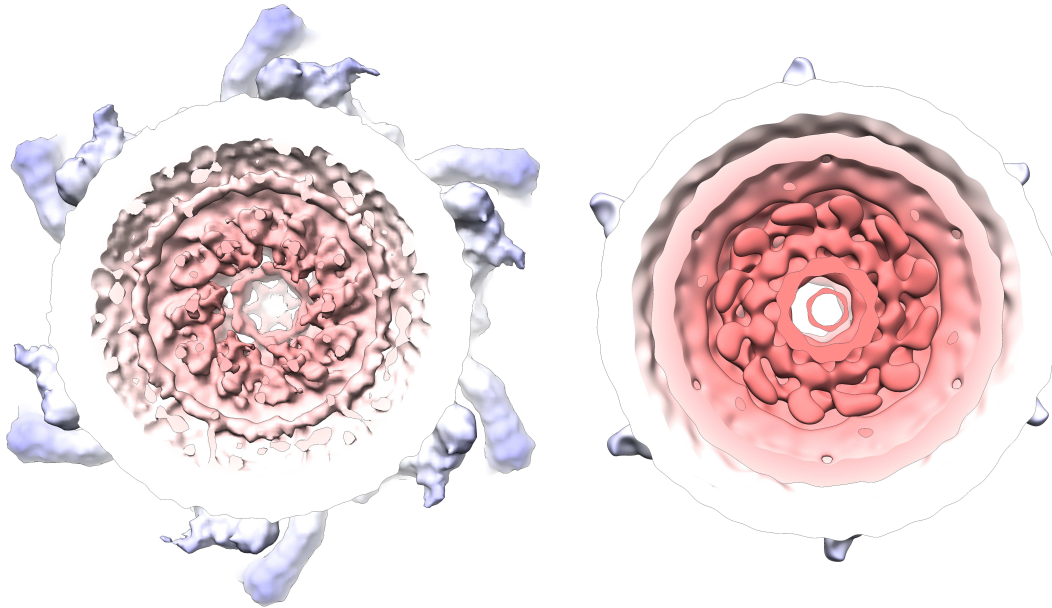
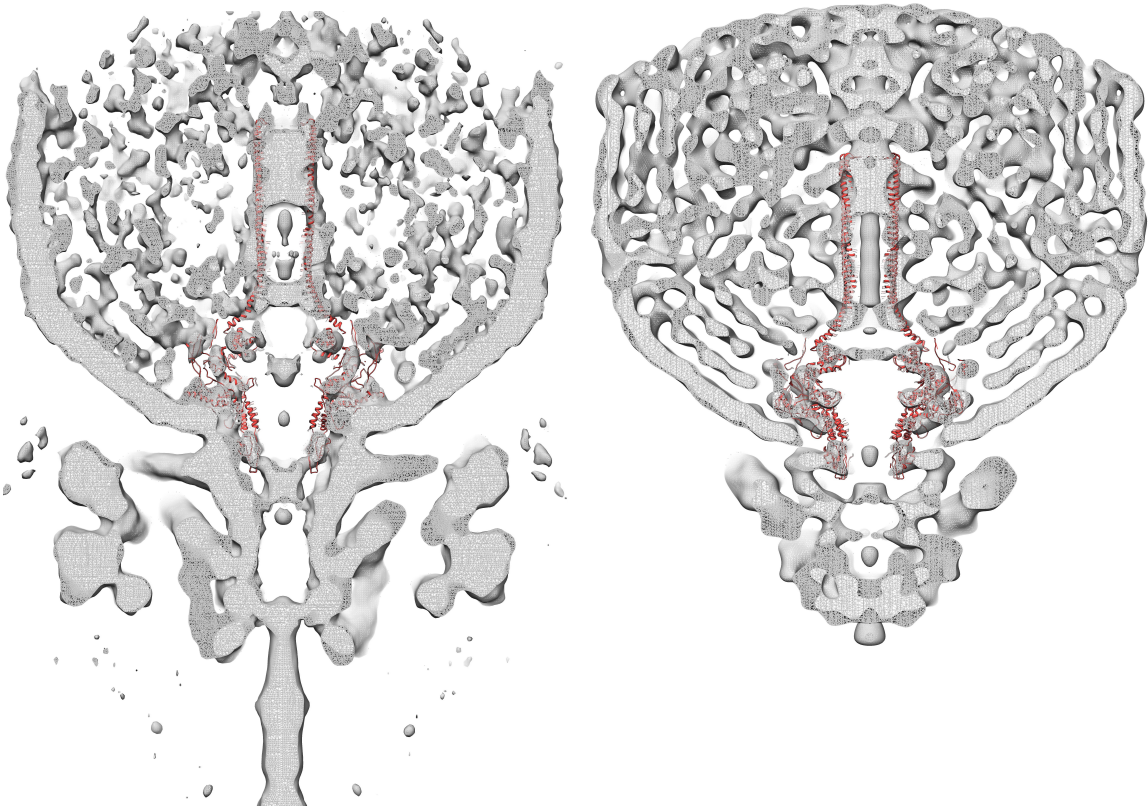
A**B**

Figure 16: Cryo-EM structures of ϕ Eco32 and 7-11 tails. **(A)** Top views of the sections of the cryo-EM structures of ϕ Eco32 and 7-11 tails with dodecameric portals visible. **(B)** Side views of the sections of the cryo-EM structures of ϕ Eco32 and 7-11 tails with P22 portal protein structure fitted.

proteins, the MCP gp8 and minor capsid protein gp9 containing IG-like domain. The MCP gp8 length is 312 residues, and its closest ortholog is HK97 capsid protein, which length is 285 residues. The second 7-11 capsid protein gp9 (585 residues) contains IG-like domain, and probably corresponds to the decoration protein at the 5-fold vertexes.

The section through the capsid part of the 6-fold tail reconstruction shows that there are clearly visible 16 layers of alternating high and low density corresponding to the 8 concentric layers of densely packed dsDNA, separated by 24 Å (Fig. 14E (yellow)), different from 9 layers of encapsulated ϕ Eco32 dsNA.

Raw micrographs of 7-11 reveal similar states of the DNA leakage process, from hexagonal DNA organization, through cholesteric and then through isotropic to the empty capsid (Fig. 15 C(1-7)). However, there are no empty virions with horizontal internal layers from the inner side of 7-11 capsid shell, organized like hoops in a barrel in ϕ Eco32. Only capsids with layers filling the diagonal MCP hexamers concavities (left- and right-handed) (Fig. 15 C(4-5), 15C(5)), which indicates different from ϕ Eco32 organization of 7-11 outer DNA layers.

2.3.3. Structure of the tail

2.3.3.1. ϕ Eco32

The reconstruction of the whole ϕ Eco32 virion including the tail region using only 5-fold averaging results in a smeared tail density of the tail due to the symmetry mismatch between the phage capsid and the tail (Fig. 13 A). Many structural studies of tailed bacteriophages show the presence of the 6-fold rotational symmetry of the tails (Savalia et al. 2008; Simpson et al. 2000; Steven et al. 2005). We use “divide-and-conquer” approach to reconstruct tail density. The tail corresponding area was cut out from the same micrographs and processing these images with 6-fold symmetry up to the resolution of 15 Å.

The tail is associated with the one of the distal ends of ϕ Eco32 prolate capsid, breaking the 5-2 symmetry of a true prolate icosahedron. There are two different types of kinked tail fibers (long and short), 6 copies of each are folded in an intertwined manner around the periphery of a tail central structure, forming a bow tie-like complex (Fig. 17, 18). The cylindrical tail structure consists of central stub-like density 175 Å long, with a wider collar region (gatekeeper), to which the long fibers are connected. This gatekeeper tapers from 140 Å, starting at the capsid, to 83 Å narrow middle tube (nozzle), which then extends to 127 Å wide hexameric terminal knob, (Fig. 17 A).

Each copy of a long tail fiber, 620 Å long and with a mass of 528 kDa, starts from the gatekeeper, and forms a 45° bend-like structure roughly at one third of its length. There is a large trimeric knob density, 45 Å and 65 Å in diameter respectively (Fig. 18 A, C, 19 D). This trimeric density has been assigned to the tail fiber component protein gp18, whose closest ortholog is the baseplate

structural protein gp10 from T4 phage. The 68 kDa trimeric C-terminal fragment of the T4 gp10 was used in the docking procedure to ϕ Eco32 gp18 assigned electron density (Fig. 19 B). The gp18 trimer estimated mass from the LC/MS/MS is 87 kDa (Savalia et al. 2008), and the corresponding ϕ Eco32 cryo-EM density volume is 111 kDa. Each long tail fiber ends with a warhead-like density, ~ 38 Å in diameter. The long fiber is likely to be composed from 2 or 3 proteins, probably gp18, gp19 and gp21.

Short tail fiber has a length of 440 Å and a mass of 534 kDa, and emanates from the hexameric tail terminal knob of the nozzle with a 30° kink at approximately half of fiber length (Fig. 18 E). The short fiber is likely to be composed from 2 proteins, gp14 and gp15, which estimated masses from the LC/MS/MS are 93 kDa and 77 kDa (Savalia et al. 2008) and from the corresponding cryo-EM densities mass estimations are 115 kDa and 60 kDa respectively. There is a possible binding site for the tail spike, 117 Å apart from the kink, which has been assigned to ϕ Eco32 gp15 tail spike protein. The observed 3-fold symmetry and the volume of the gp15 electron density suggests that gp15 is a trimeric spike (Fig. 18 E). There is an additional disordered electron density above the gp15 tail spike assigned density at map contoured at low σ value (Fig. 17 A, 18 E). The density is weaker relative to rest of the tail density, suggesting its flexible. As in other *Podoviridae* phages, the terminal domain of the tail spike probably has an enzymatic activity, and is involved in binding and/or digesting the cell wall during infection.

The ϕ Eco32 tail ends with a prominent needle, resembling P22 tail needle, protruding from the center of the tail terminal knob. The needle density ~ 105 kDa, is 220 Å long and only 32 Å wide, with a discrete wide globular C-terminal knob, and has been assigned to the ϕ Eco32 gp20 tail needle protein (Fig. 17 C, 18 A, C, E). It was shown that all characterized tail needles consists of trimeric fibers 220–320 Å in length (Bhardwaj et al. 2009). The trimeric features of the tail needle were smeared due to the 6-fold symmetry enforcement of the tail reconstruction. The closest orthologous structure of the P22 tail-needle gp26 protein was tentatively fitted into ϕ Eco32 needle density with good agreement (Fig. 19 A). However, the C-terminal knob absent in phage P22 but present in other P22-like phages. Identification of ϕ Eco32 gp20 protein (322 residues) in *PSYPRED* indicates the presence of the alpha-helix reach region, corresponding the tail needle helical core structure, followed by the C-terminal beta-strands reach region, starting from the 164 residue. The C-terminal knob easily accommodate Sf6 phage tail needle knob (Fig. 19A). This knob represents protruding tip of the ϕ Eco32 that contacts the host cell surface.

On a raw micrographs and tomograms we observed the ϕ Eco32 empty and DNA filled virions with unfolded long (Fig. 20 A(1-4),) and short tail fibers (Fig. 20 B(1-4)). ϕ Eco32 particle likely to perform initial attachment to *E.Coli* cell outer lipid membrane primary receptors with a long tail fibers, containing gp19 protein with IG-like domain. After initial binding, probably reversible, the short tail fibers, are likely to irreversibly anchor the phage to the host cell surface. The potential

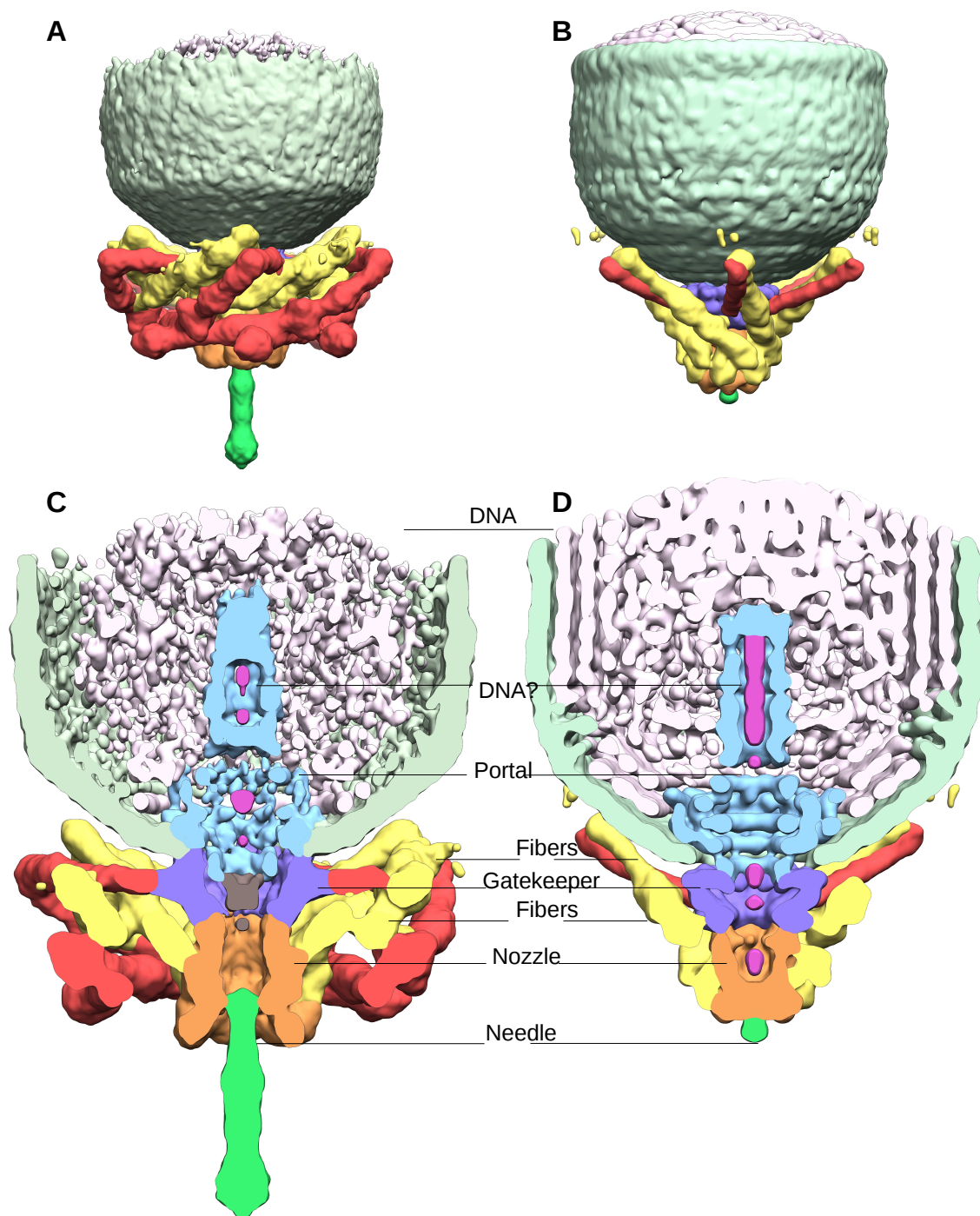


Figure 17: Cryo-EM structures of ϕ Eco32 and 7-11 tails. **(A)-(B)** Side views of the cryo-EM structures of ϕ Eco32 and 7-11 tails, the maps are interpreted and colored in terms of the component proteins that were known. **(C)-(D)** Cutaway views of the cryo-EM structures of ϕ Eco32 and 7-11 tails with main components highlighted.

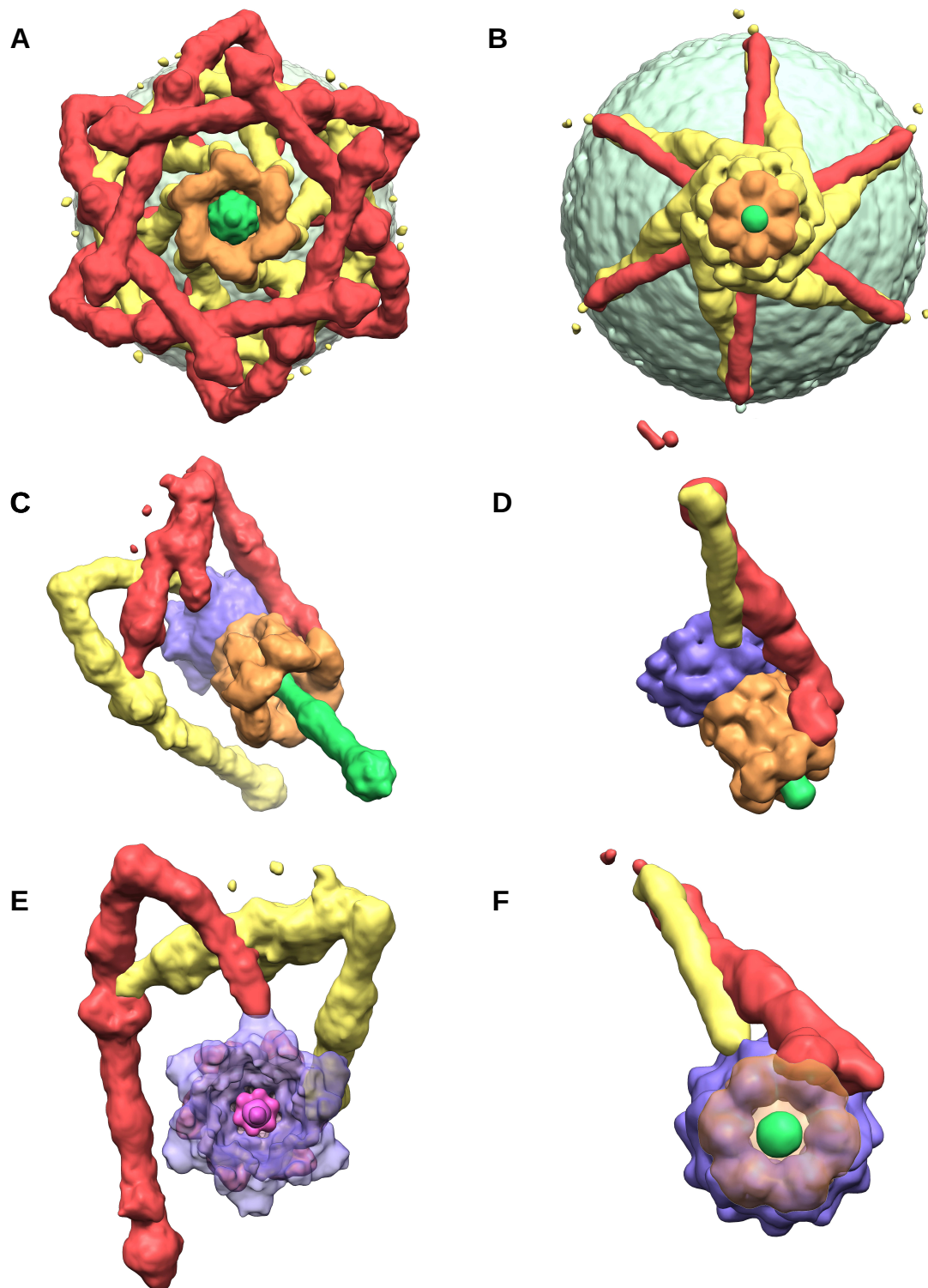


Figure 18: Cryo-EM structures of ϕ Eco32 and 7-11 tails, the maps are interpreted and colored in terms of the component proteins that were known. (A)-(B) Bottom views of the cryo-EM structures of ϕ Eco32 and 7-11 tails. (C), (E) Tilted and top views of ϕ Eco32 tail cryo-EM structure without capsid and DNA densities, and with 1 copy of each tail fiber shown. (D), (F) same for 7-11 phage.

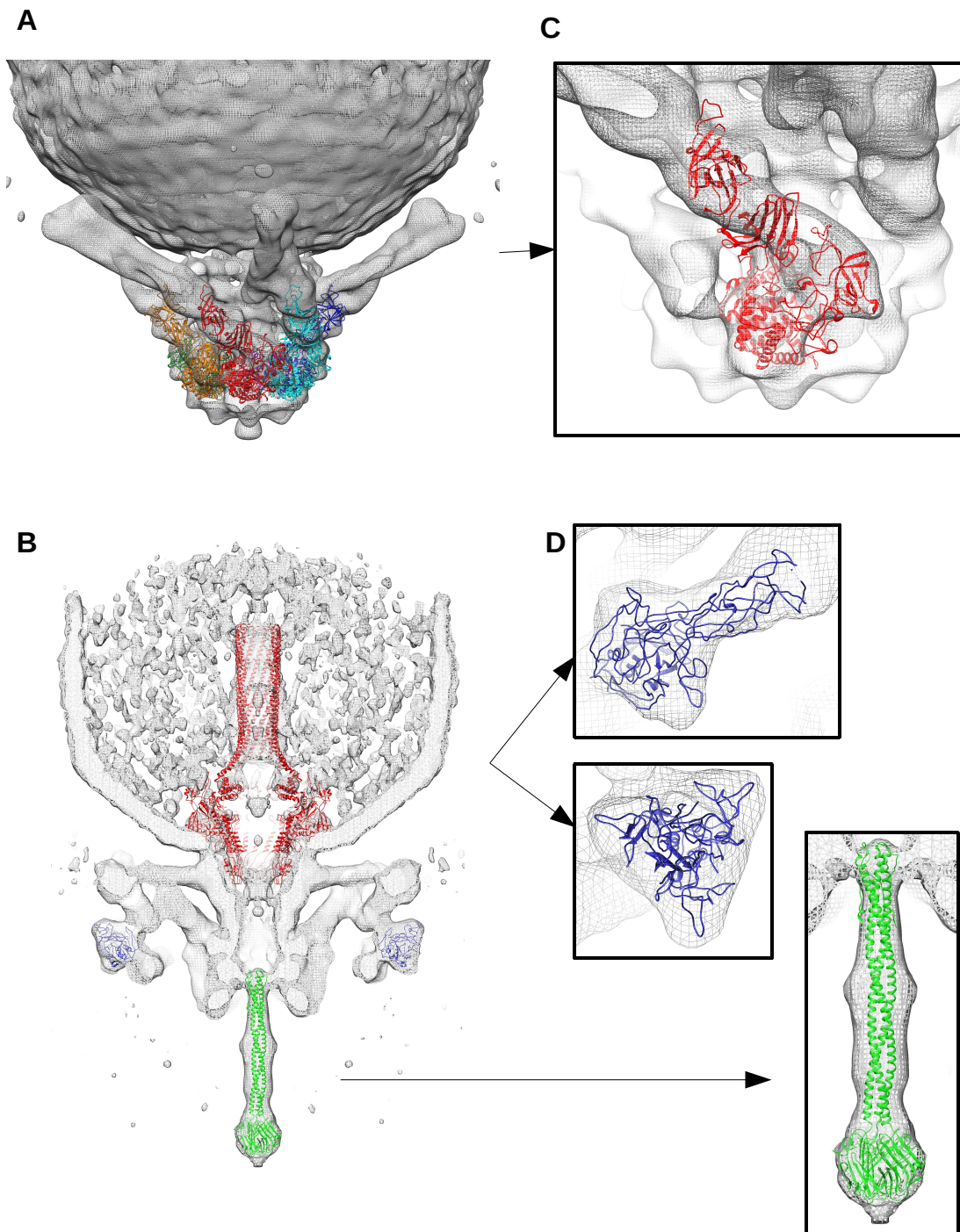


Figure 19: Cryo-EM structures of ϕ Eco32 and 7-11 tails with orthologous protein structures fitted. **(A)** Side view of the 7-11 tail cryo-EM structure with 6 copies of ϕ 92 tail spike orthologous protein gp142 fitted. **(C)** Enlarged side view of the orthologous protein gp142 structure fitted into 7-11 tail cryo-EM structure. **(B), (D)** Cutaway views of ϕ Eco32 tail cryo-EM structure with known orthologous protein structures fitted: P22 portal protein structure (red), C-terminal fragment of T4 gp10 (blue) and P22-Sf6 tail needle composite structure (green).

mechanism of tail fibers unfolding might include the stage of long tail fibers turn around the kink, and extension of the short tail fibers along the tail tube. This extension is visible only on an empty capsids. Figure 9.D represents empty ϕ Eco32 capsid during DNA leakage process. External DNA layers are organized in characteristic pattern, with a DNA terminus passed through the portal complex. The barrel region, however, is not visible except for the connector density, which is also present on averaged cryo-ET reconstruction (Fig. 3B(6)). The tail complex shows the elongation of the central hub, probably a passage formed for the DNA to pass through the cell envelope (Fig. 20 A(1, 3, 4), 20 B(1-4)).

2.3.3.2. 7-11

The same “divide-and-conquer” approach was used to reconstruct 7-11 tail density. The tail corresponding area was cutting out from the micrographs and processed with 6-fold symmetry up to the resolution of 18 Å (Fig. 13). The 7-11 short tail locates at the distal end of the 7-11 prolate capsid. It is 190 Å long and ~460 Å wide, with a simple topology, resembles that of phage T7. However, there are two different types of intertwined tail fibers (top and bottom), 6 copies of each are protrude outward from the central tail hub (Fig. 13, 17 B, 18 B, D). The disordered density of the fibers terminal domain points to the their flexibility (Fig. 17 B, 18 F, 20 C(1)). Those domains likely have enzymatic activity and involved in binding/digesting the cell wall. Phage 7-11 tail structure consists of a dodecameric portal complex 135 Å wide, a cylindrical dodecameric gatekeeper 80 Å in diameter, with a tail fibers attached (Fig. 17 D, 18 B, D, F), and a hexameric nozzle, which tapers from 77 Å to 62 Å wide middle tube, and extends to 99 Å wide terminal knob. Another type of tail fibers and small 20 A long tail needle are attached to the terminal knob. There are at least 3 possible core proteins gp23, gp24 and gp27 (Tab. 2), which are not visible in the cryo-EM reconstruction and probably are disordered. The cryo-EM tail reconstruction clearly shows the 12-fold symmetry of the portal wing domains and the gatekeeper (Fig. 16 A). The 7-11 portal barrel encompassed with a few rings of density (probably, the DNA strands), similar to ϕ Eco32. Also, as in ϕ Eco32, the closest ortholog of 7-11 gp5 portal is P22 phage portal complex, which was fitted into 7-11 tail density (Fig. 16 B). Since the 7-11 and ϕ Eco32 are from *Podoviridae* family, belong to the same genus “ ϕ Eco32-like viruses” (Kropinski, Lingohr, and Ackermann 2011) and has the similar morphology, the absolute hand of the 7-11 tail was selected to be the same as for the ϕ Eco32 tail. The characteristic twist of the 7-11 and ϕ Eco32 nozzles coincides with that of other podophages, like K1E, K1-5, T7, epsilon15 (S. R. Casjens and Molineux 2012; Jiang et al. 2006; Petr G Leiman et al. 2007). Each copy of a top tail fiber, ~230 Å long and with a mass of 180 kDa, protrudes almost radially from the gatekeeper with ~80 degrees inclination to the vertical 6-fold axis (Fig. 17 B, D, 18 B, D). Bottom tail fiber, ~ 240 Å long and with a mass of 250 kDa, starts from the hexameric terminal knob of the nozzle, and protrudes outward with ~60 degrees inclination to the 6-fold axis, and with the twist of ~45 degrees (Fig. 17 B, D, 18 B, F). Bottom tail fiber starts

with a density that has been assigned to the tail fiber component protein gp14, closest ortholog of the 103 kDa tail spike structural protein gp141 from ϕ 92 phage, which was used in the docking procedure (Fig. 19 C). The corresponding cryo-EM density mass is \sim 100 kDa.

Two copies of both fibers intertwines at \sim 2/3 of bottom fiber length and \sim 1/2 of top fiber length, at the level of the gatekeeper (Fig. 17 B, 18 B). The 7-11 tail ends with a short needle, protruding from the center of the tail terminal knob. The needle density \sim 13 kDa, is 20 Å long and has been assigned to the 7-11 gp19 tail needle protein (8kDa) (Fig. 4, 5). The 7-11 phage images shows empty and partially empty virions with unfolded fibers (Fig. 20 C), which started from the nozzle (bottom tail fibers) and extended along the tail tube. Fixed position of the nozzle fibers presents in all observed empty particles, indicating possible mechanism of host cell attachment. This mechanism might include initial reversible binding of gatekeeper tail fibers to the *Salmonella* outer lipid membrane, and anchoring the phage virion on the surface with nozzle tail fibers, which stayed fixed (Fig. 8) during the DNA ejection process.

2.3.4. Putative infection mechanism

Podoviridae bacteriophages with short, non-contractile tails are believed to have two stages infection process. First, initial reversible binding to the primary receptors (LPS) on the host cell surface by the tail spikes or fibers. This weak binding most probably is accomplished by one or two fibers, and does not usually cause the DNA release from the phage virion. However, it moves the virion close to the cell surface, and changes the orientation of the other tail fibers relative to the tail hub. Initially attached phage is performed random walk before finding a site for the irreversible binding to the secondary receptors. This type of binding orients the phage perpendicular to the cell membrane, and is thought to trigger the conformational changes and rearrangement of the tail structure and DNA ejection into the cell. However, *Podoviridae* phage tails are too short to create a channel through an outer membrane, peptidoglycan layer, and inner membrane for DNA release. Channel formation for DNA passage is a poorly understood process. However it is thought that podophages are utilized the internal core proteins, which are ejected through the rearranged portal-tail hub complex and extends the tail, forming a trans-membrane channel.

Podophage ϕ Eco32 probably performs initial reversible attachment to the *E.Coli* outer membrane with an elongated capsid. Due to the presence of the IG-like domain of gp12 protein, the capsid surface is “sticky”, and helps to bring the compact short tail with the intertwined tail fibers near the cell surface (Fig. 21 A).

We found that overwhelming majority of mature ϕ Eco32 virions have their fibers folded around the tail hub. However, we also found small fraction of virions with extended tail fibers. Interestingly, extended long tail fibers which probably moves freely around the fiber flex, are present in both free virions with packaged DNA and spontaneously emptied virions (FIG. 20 A (1-4)). While short tail

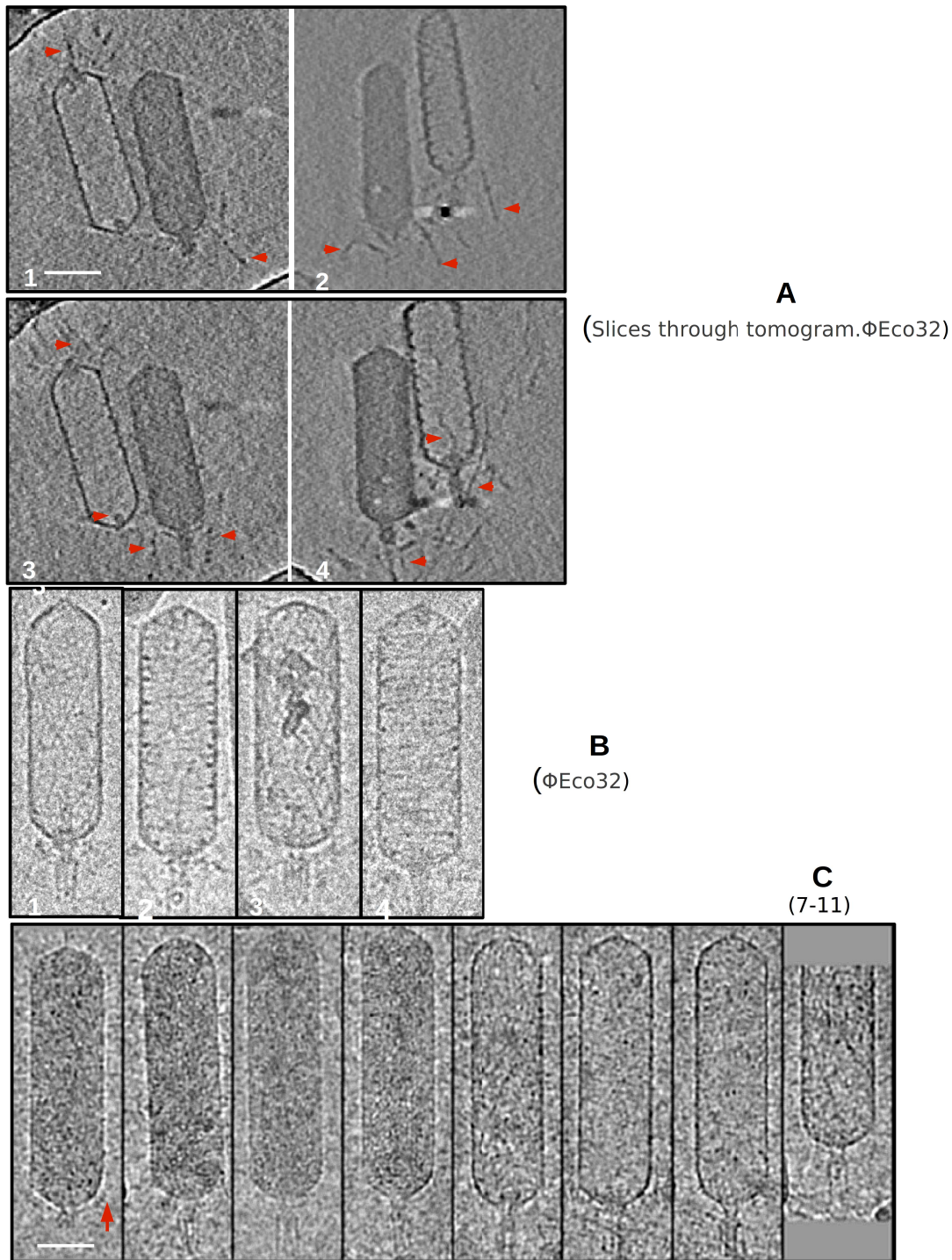


Figure 20: Raw cryo-EM images and cryo-ET slices of ϕ Eco32 and 7-11 (scale bar – 4.5 nm). **(A)** Slices through tomographic reconstruction of ϕ Eco32 virions with reoriented long and short tail fibers. DNA and tail fibers are highlighted. **(B)** Raw cryo-EM images of ϕ Eco32 virions with reoriented short tail fibers. **(C)** Raw cryo-EM images of 7-11 virions with reoriented short tail fibers. Flexible terminal domain of long tail fiber is highlighted with red arrow.

fibers appear extended only in empty virions, closely-spaced and parallel to the tail hub sixfold axis and point towards the host cell surface (FIG. 20 B (1-4)). Extended fibers are clearly visible only in cryo-ET slices, but not in cryo-EM images (FIG. 20 A (1-4)). However, it is possible that observed long tail fibers are folded around the tail hub due to the cryo-EM imaging condition, such as a low salt buffer and high concentration of phages. In native conditions they are unfolded and move slightly around the tail hub.

ϕ Eco32 has the tail needle, resembling that of P22 phage but with the Sf6-like C-terminal knob which is absent in phage P22. Distal end of the tail needle protrudes 176 Å beyond the bottom of the tail nozzle, compare to ~140 Å and ~120 Å in P22 and Sf6 phages. Most likely the gp20 trimeric needle first interacts with the outer membrane. However, no putative secondary receptor protein has been identified for ϕ Eco32.

It was shown for P22 and hypothesized for the Sf6 phages, that tail needle does not interact with a receptor and is just mechanically ejected into the cell, serving as “pressure sensor”. In this model the tail needle pierces the outer LPS, and when contacting the LPS inner core or lipids triggers conformational changes of the tail, followed by the DNA ejection.

At all events, the tail needle has to be ejected to open the channel for the DNA passage. ϕ Eco32 averaged sub-volumes does not show presence of the tail needle and any internal core proteins inside emptied virions (Fig. 15 B6), suggesting they have been ejected during the infection. However, we have not observed any tail hub extensions in free emptied virions, or virions attached to the vesicles. Most likely that interaction with the LPS or lipids triggers only tail needle ejection and DNA release, but not the tail extension (Fig. 21 G).

A ϕ Eco32 infection mechanism can be hypothesized:

- 1) Free swimming phage virion first weakly interacts with the outer LPS by a “sticky” capsid surface, and brings the tail complex near the LPS (Fig. 21 A).
- 2) Slightly protruding above the capsid diameter kinked long tail fibers reversibly bind to LPS. Phage virion is constantly shaken by the host cell movement and starts to “walk” around initial binding point (Fig. 21 B). This random walk causes the large turn of the long tail fiber, and similar binding and rotation of the other long fibers. Long tail fibers rearrangement due to the geometry of the ϕ Eco32 tail results in the turn and unfolding of the short tail fibers (Fig. 21 C).

Another possibility is the initial mechanical interaction of the tail needle with the outer cell surface, which triggers the conformational changes of the nozzle with the short tail fibers attached. Extended short tail fibers push and turn the long tail fibers, which start reversibly bind to the outer LPS and orient the phage virion. Most likely that both schemes are valid.

Attachment of the three or more long tail fibers is possible only when pointing towards the cell

surface, similar to myophages interaction with the cell surface (see T4 infection mechanism).

3) Extended short tail fibers are point toward the cell surface, and search for the putative secondary receptors (Fig. 21 D). The tail hub with attached needle protrudes 382 Å from the connector, while extended short tail fibers, closely-spaced and aligned parallel to the tail hub protrudes 468 Å below the capsid connector. As a result, the needle hides in the 6 short tail fiber copies cluster, and cannot interact with the membrane. However, we observed that unfolded short tail fibers have the same length as the tail hub with attached tail needle (Fig. 20 B1, FIG. 21 H-J). It may be possible, that the distal domain of the short tail fiber is detached or flattened along the cell surface. The resulting modified tail fiber has the length ~374 Å, similar to the tail hub-needle complex length.

To pierce the host cell membrane the short tail fibers, probably, together with the long tail fibers, move the tail hub towards the cell surface by at least the length of the protruding distal end of the needle, ~180 Å. With that, the proximal short tail half-fiber orients upward ~108° relative to the tail hub sixfold axis, situation observed for phage virions with packaged DNA (Fig. 20 A (1-4), 10 B).

4) Last step is the formation of the trans-membrane channel and DNA and the proteins ejection into the host cell cytoplasm. We have not observed formation of any tubular density for spontaneously emptied virions. However, presence of the gp26 internal protein with similarity to T7 gp15 internal core protein indicates, that ϕ Eco32 internal core is used for creation of the DNA passage. ϕ Eco32 probably does not require polymerase for DNA ejection completion, since it can eject its DNA into the vesicles (Fig. 21 G).

For bacteriophage 7-11 the infection mechanism most likely is the same.

2.4. Concluding remarks

In this present report we described the detailed cryo-EM structure of the two novel bacteriophages from the genus “ ϕ Eco32-like viruses”: *E. Coli* phage ϕ Eco32 and *Salmonella* phage 7-11. Both phages are members of the *Podoviridae* family, and share the same unusual and rare C3 morphotype, characterized by the short non-contractile tails and enormously elongated capsids. The mature virion of ϕ Eco32 contains at least 16 different proteins, compare to 14 different proteins of the 7-11 phage. Despite the unusual capsid elongation (elongation ratios are 3.22 and 3.57 for ϕ Eco32 and 7-11), both bacteriophages share a very common HK-97 fold. ϕ Eco32 and 7-11 packaged dsDNA likely form inverse spool-like structure. However, more complex DNA organization is possible. ϕ Eco32 external DNA layers forming characteristic pattern, resembling hoops inside capsid shell. This pattern is clearly visible on empty capsids images and cryo-EM reconstruction of ϕ Eco32.

Since both bacteriophages show very low level of structural homology to known proteins (~ 40 %

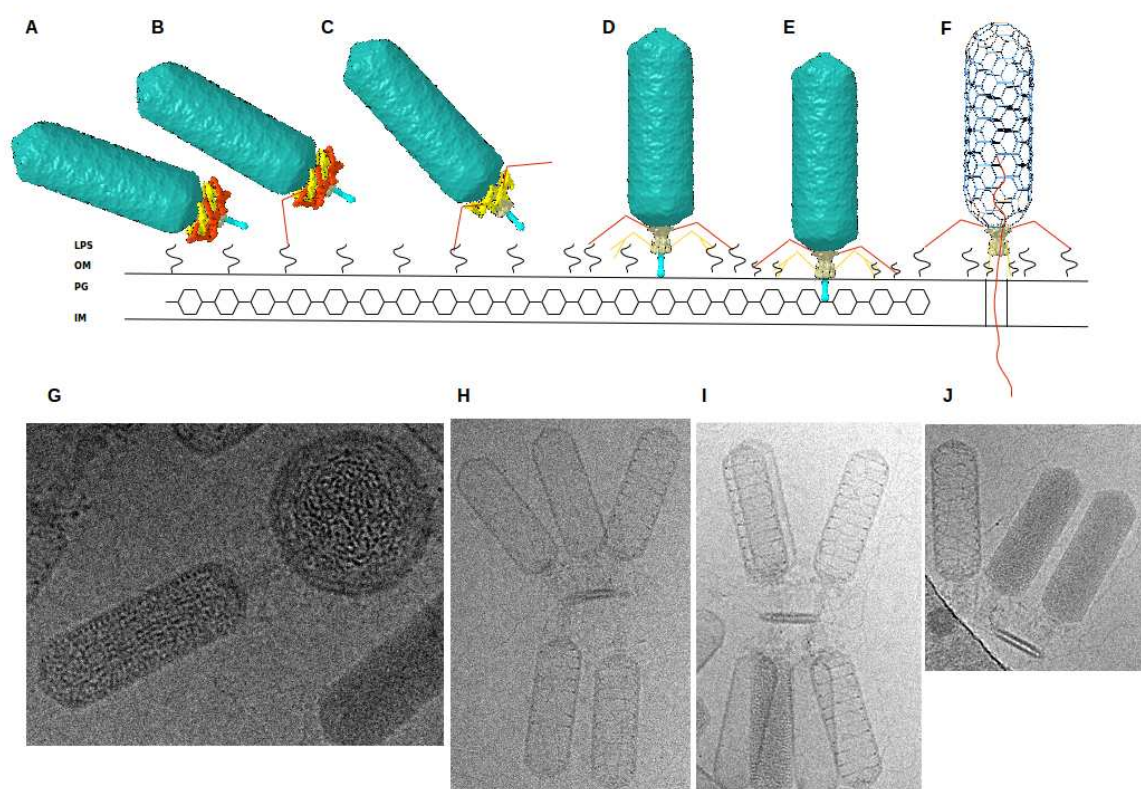


Figure 21: Hypothesized adsorption mechanism of bacteriophage ϕ Eco32. **(A)** Free ϕ Eco32 virion close to the host cell outer membrane, **(B)** initial reversible attachment with long tail fiber, **(C)** random walk on the outer membrane surface and searching for secondary receptors, **(D)** vertical orientation of the ϕ Eco32 virion and irreversible binding to the secondary receptors with short tail fibers, **(E)** piercing the host cell membrane with tail needle and formation of the trans-membrane channel from internal core proteins, and **(F)** DNA ejection process. **(G-J)** Raw cryo-EM images of ϕ Eco32 virions attached to the pieces of *E. coli* membrane.

for ϕ Eco32, and ~15 % for 7-11), it becomes difficult to annotate all structural proteins from the ϕ Eco32 and 7-11 tails. However, structural components such as the dodecameric portals and gatekeepers, hexameric nozzles are identified. Also, the mass, position and the morphology of the tail fibers, as well as the proteins constitution are clearly evident from the current cryo-EM reconstructions. Both bacteriophages have short non-contractile tails of similar dimensions, which consist of the same components: portal complex, gatekeeper and nozzle with a needle proteins attached. There are two types of tail fibers attached to the tails of ϕ Eco32 and 7-11. Each fiber is likely to be composed from 2 to 3 proteins. However, ϕ Eco32 tail has more complex topology, compare to 7-11 and other *podophages*. ϕ Eco32 possesses long protruding needle, resembles that of

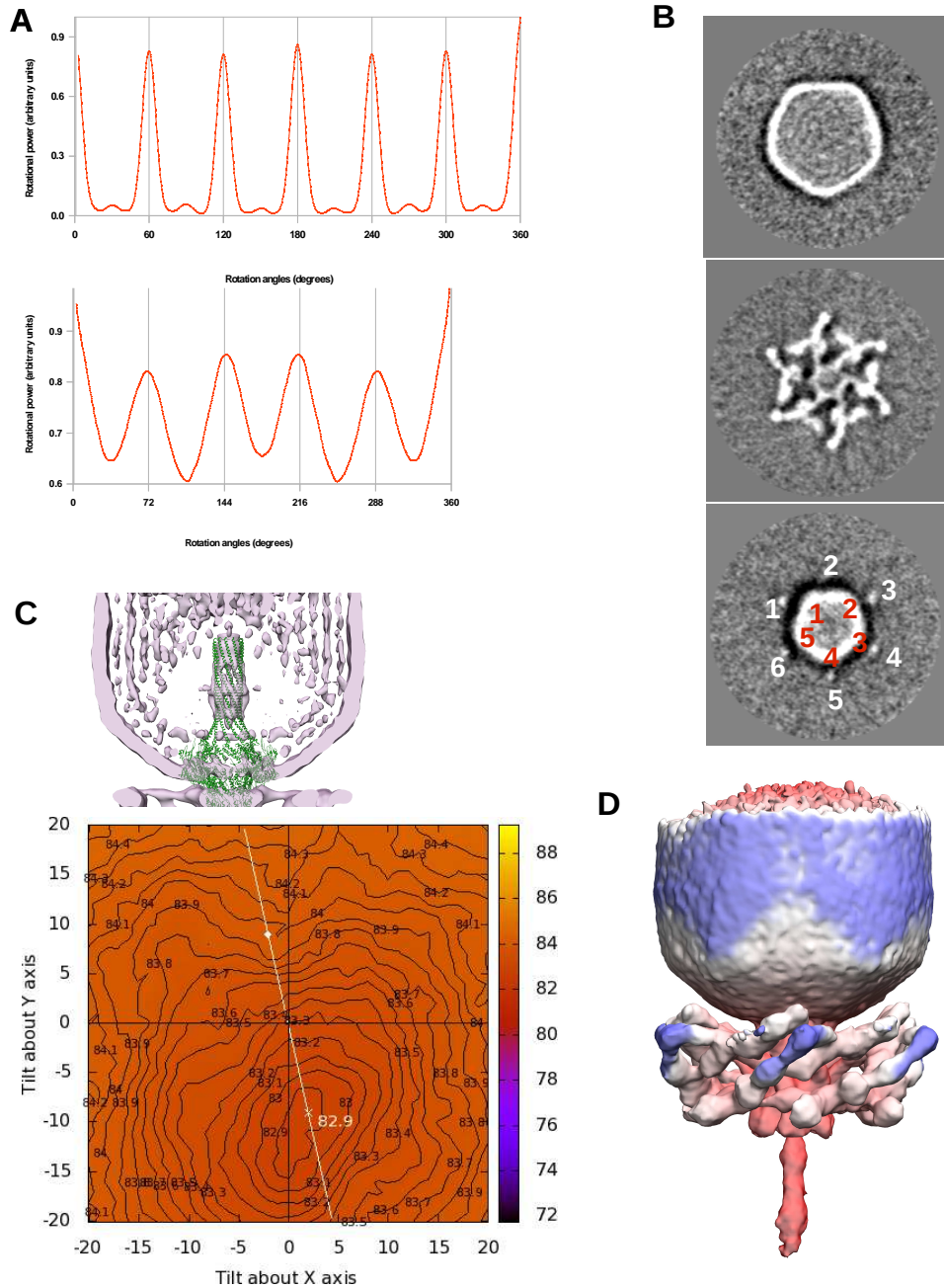


Figure 22: Symmetry-free (C1) cryo-EM reconstruction of ϕ Eco32 tail and absolute hand determination. **(A)** Rotational power spectra of the cross sections perpendicular to the central fivefold axis through the capsid and the tail regions of symmetry-free reconstruction, **(B)** cross sections through the capsid and the tail regions, **(C)** cryo-EM structure of ϕ Eco32 tail with orthologous P22 portal protein fitted, twist of the portal barrel coincides with the twist of cryo-EM structure. Average phase residual landscape from Tilt pair validation server shows the absolute hand of the cryo-EM structure, **(D)** side view of the symmetry-free ϕ Eco32 cryo-EM structure.

phage P22, which helps to pierce the *E.Coli* membrane, while 7-11 phage has very short (20 Å) needle protein attached.

2.5. Materials and methods

2.5.1. Cryo-EM and cryo-ET

Low-dose cryo-EM was performed as described elsewhere (T. S. Baker and Henderson 2006). 100 micrographs of ϕ Eco32 were collected on JEOL JEM 2200FS electron microscope using 4k CCD camera with a nominal pixel size 2.34 Å with a radiation dose of ~ 20 e-/ Å². 150 micrographs and 22 single-tilt tomograms of ϕ Eco32 were recorded on Eagle 4K CCD camera using a FEI Tecnai F20 200KV microscope with a nominal pixel size 2.26 Å with a radiation dose of ~ 20 e-/ Å. The defocus range varies from 0.6 to 3 μ m. The images were binned to give a pixel size of 4.68 Å for JEOL data and 4.42 Å for F20 data. JEOL magnification calibration was performed using the MAG*!CAL® calibration reference standard, and F20 pixel size was calibrated to the JEOL micrographs pixel size. 200 micrographs of ϕ Eco32 were collected using a K2 Summit detector and 100 micrographs of 7-11 phage with nominal pixel size 1.7 Å. The contrast transfer function (CTF) of each micrograph was estimated using CTFFIND3 (Mindell and Grigorieff 2003) and further corrected with Bsoft (Bernard Heymann et al. 2008). The quality of micrographs was checked by inspecting the averaged power spectra generated by CTFFIND3. Best images were selected based on CTFFIND3 correlation coefficient estimation.

The prolate ϕ Eco32 capsids were manually extracted from preprocessed micrographs using *helixbacthboxer* routine from EMAN (Ludtke, Baldwin, and Chiu 1999). The rectangular boxes were selected to include the whole ϕ Eco32 virion with the tail region, and further clipped to produce square boxes. The initial model was constructed with a small subset of initial data using EMAN routines. The model was then refined until convergence applying fivefold rotational symmetry. We boxed the whole particle including the tail region, which served as an anchor and helped during model refinement. The smeared density of the tail region due to the symmetry mismatches in a phage particle cut out from the capsid reconstruction, low-pass filtered and used in the ϕ Eco32 tail model refinement. The squared tail corresponding areas including small part of the capsid of ϕ Eco32 and 7-11 were manually extracted from preprocessed micrographs by *batchboxer* routine from EMAN.

ϕ Eco32 model was refined with EMAN, and 7-11 model refinement was refined with SPIDER (Shaikh et al. 2008) and with FREALIGN (Grigorieff 2007). sixfold rotational symmetry was applied to both reconstructions. Despite the large number of collected micrographs, the outcome from each micrograph was relatively small, maximum up to 10 particles at best due to the prolate shape of ϕ Eco32 and 7-11 virions. We also avoided particles with a high tilt angle, which usually caused

misalignment between capsid top and bottom caps.

Asymmetrical reconstruction of ϕ Eco32 was calculated with SPIDER, initially assuming sixfold rotational symmetry. Using already found orientations, each image was compared to six possible orientations around the sixfold tail axis. Best orientation was determined by the highest correlation coefficient. Images with refined orientations were forward to reconstruction step, assuming no symmetry, and all procedure was iteratively repeated.

The cryo-EM tomograms were collected on FEI Tecnai F20 microscope and reconstructed with IMOD package using gold fiducial markers (Kremer, Mastronarde, and McIntosh 1996). To compensate for the missing wedge due to the data collected from a limited angular tilt range (from -70° to + 70°) 30 ϕ Eco32 empty capsid sub-volumes were extracted from reconstructed tomograms, aligned and averaged using PEET routine from IMOD package (Nicastro et al. 2006). “Rigid-body” fitting in UCSF Chimera (Pettersen et al. 2004) was used for docking procedure.

Gene product number	Possible function and position in the virion	Mass of the monomer predicted by sequence, kDa			Mass estimated from cryo-EM density, kDa
11	Major head protein	39			30
12	MCP extension (frame shift), capsid surface	16			-
8	Portal protein	85			80
25	Internal DNA injection protein, P22 gp20-like protein	60			-
26	Internal structural protein, similarity with T7 gp15 protein	163			-
13	Tail fiber component, T7 gp11-like protein	29	trimers	439*	385
14	Tail fiber component	88			
15	Tail spike	77			
18	Tail fiber component, T4 gp10-like protein	29			
19	Tail fiber component	110			
21	Possible tail fiber component, T1-like fiber specific component	37			
23	Tail fiber component	36			
24	Tail fiber component, similar to gp23	33			
20	Tail needle protein	35			41

Table 1. ϕ Eco32 identified structural proteins (*including chaperons).

Gene product number	Possible function and position in the virion	Mass predicted by sequence, kDa		Mass estimated from cryo-EM density, kDa	
8	Major head protein	34		-	
9	Outer capsid protein, possible decoration protein	62		-	
5	Portal protein	89		91	
23	Internal protein, P22 gp20-like protein	61		-	
24	Tubular protein, P22 gp16-like protein	69		-	
27	Internal structural protein	46		-	
12	Tail protein, T7 gp11-like protein	26		430*	
13	Tail fiber component	28	570		
14	Tail fiber component, ϕ 92 gp141-like protein	110			
15	Tail fiber component, similar to gp13	16			
17	Tail fiber component, T4 gp10-like protein	32			trimers
18	Tail fiber component	64			
20	Tail fiber component	42			
19	Possible tail needle protein	9		13	

Table 2. 7-11 identified structural proteins (*not including disordered fiber tip).

3. Bacteriophage ϕ 92

Bacteriophage ϕ 92 is a large, lytic myovirus isolated in 1983 from pathogenic *Escherichia coli* strains that carry a polysialic acid capsule. Here we report the genome organization of ϕ 92, the cryoelectron microscopy reconstruction of its virion, and the reinvestigation of its host specificity. The genome consists of a linear, double-stranded 148,612-bp DNA sequence containing 248 potential open reading frames and 11 putative tRNA genes. Orthologs were found for 130 of the predicted proteins. Most of the virion proteins showed significant sequence similarities to proteins of myoviruses rv5 and PVP-SE1, indicating that ϕ 92 is a new member of the novel genus of rv5-like phages. Reinvestigation of ϕ 92 host specificity showed that the host range is not limited to polysialic acid-encapsulated *Escherichia coli* but includes most laboratory strains of *Escherichia coli* and many *Salmonella* strains. Structure analysis of the ϕ 92 virion demonstrated the presence of four different types of tail fibers and/or tailspikes, which enable the phage to use attachment sites on encapsulated and non-encapsulated bacteria. With this report, we provide the first detailed description of a multivalent, multispecies phage armed with a host cell adsorption apparatus resembling a nanosized Swiss army knife. The genome, structure, and, in particular, the organization of the baseplate of ϕ 92 demonstrate how a bacteriophage can evolve into a multi-pathogen-killing agent.

3.1. Introduction

Among Gram-negative bacteria, a huge variety of cell-surface polysaccharides is found (Whitfield 2006). These sugar polymers can form thick coats, which mask underlying surface structures, protect the bacterium from the host immune system, and create an efficient barrier against bacteriophage infection (Cross 1990; Taylor and Roberts 2004). However, the protective sugar coats can also serve as attachment sites for specialized phages that possess polysaccharide depolymerases as tailspike proteins, enabling these phages to penetrate the coat and to gain access to cell surface receptors (Stirm et al. 1971).

Well-studied examples of tailspike proteins with polysaccharide depolymerase activity are endosialidases or endo-N-acetyl-neuraminidases (endoN) (Buchan et al. 2010; Buchan et al. 2010; Jakobsson et al. 2012; Mühlenhoff et al. 2003; Eike Christian Schulz et al. 2010; David Schwarzer et al. 2007; Stummeyer et al. 2005). Endosialidases specifically degrade α 2,8-linked polysialic acid (polySia), the capsular polysaccharide of *Escherichia coli* K1 (*E. coli* K1) (Barry and Goebel 1957). K1-specific phages are mainly small bacteriophages of the Podoviridae family (podoviruses) with short non-contractile tails ((Petr G Leiman et al. 2007); for an overview, see reference (Jakobsson et al. 2012)). The size of podoviruses allows the incorporation of only one or two different tailspike

proteins (Petr G Leiman et al. 2007), resulting in a very limited host range. K1-specific podoviruses evolved from different progenitor types and gained their new host specificity by horizontal uptake of an endosialidase gene (Stummeyer et al. 2006). For example, the genomes and virion structures of K1F and T7 are similar except for the part responsible for host cell binding. The T7 fiber is replaced by an endoN tailspike in K1F, and this fully commits K1F to encapsulated K1 strains (Stummeyer et al. 2006). Whereas the K1 capsule is absolutely required for K1F infection (Scholl and Merrill 2005), it blocks infection by T7, which infects only nonencapsulated *E.Coli* (Scholl, Adhya, and Merrill 2005).

The larger genomes of *Myoviridae*—phages with long, contractile tails—give rise to higher complexity in structural organization and infection, mechanism as extensively studied for coliphage T4 (for an overview, see reference (Petr G. Leiman and Shneider 2012b)). The only anti-K1 myovirus described so far is $\phi 92$ (or $\phi 92$). It was isolated in 1983 (Kwiatkowski et al. 1983) from *E.Coli* K92 (Bos12 [serogroup O16:K92:H⁻]), which is encapsulated by polySia with alternating $\alpha 2,8/\alpha 2,9$ linkages (Egan et al. 1977; Furowicz and Orskov 1972). It has been further demonstrated that $\phi 92$ also degrades the $\alpha 2,8$ -linked polySia of (Kwiatkowski et al. 1983; Kwiatkowski and Stirm 1987), suggesting the presence of an $\alpha 2,8$ -specific endosialidase tailspike. To date, no further studies of $\phi 92$, including the genome, structure, or host range, have been described.

In the present report, we demonstrate that $\phi 92$ infects not only encapsulated but also non-encapsulated *E.Coli* as well as clinically important *Salmonella* strains. Sequencing of the 148-kb genome of $\phi 92$ revealed several putative tail fiber and tailspike genes. Cryo-electron microscopy (cryo-EM) reconstruction of $\phi 92$ phage particles highlighted the presence of a multivalent host adsorption apparatus providing the phage with broad host specificity.

3.2. Cryo-EM reconstruction of the $\phi 92$ virion

Due to the limited size of the cryo-EM dataset, we used the divide-and-conquer image-processing approach and the reconstructions of the $\phi 92$ capsid, tail, and baseplate were calculated separately, assuming the most appropriate symmetry for each part. Most particles were imaged with their long axis parallel to the imaging plane, thus aiding the image-processing procedure (Fig. 23 A).

3.2.1. Structure of the $\phi 92$ capsid

Neglecting the unique portal vertex of the capsid (the tail attachment vertex [gp120]), the $\phi 92$ capsid shell obeyed icosahedral symmetry, and its structure, reconstructed as such, was obtained at a resolution of 19 Å (Fig. 23 A). The capsid dimensions were found to be ~930 Å, ~795 Å, and ~810 Å along one 5-fold axis, one 3-fold axis, and one 2-fold axis, respectively. The thickness of the capsid shell ranged between 23 and 43 Å in a map contoured at a level of 1.5 standard

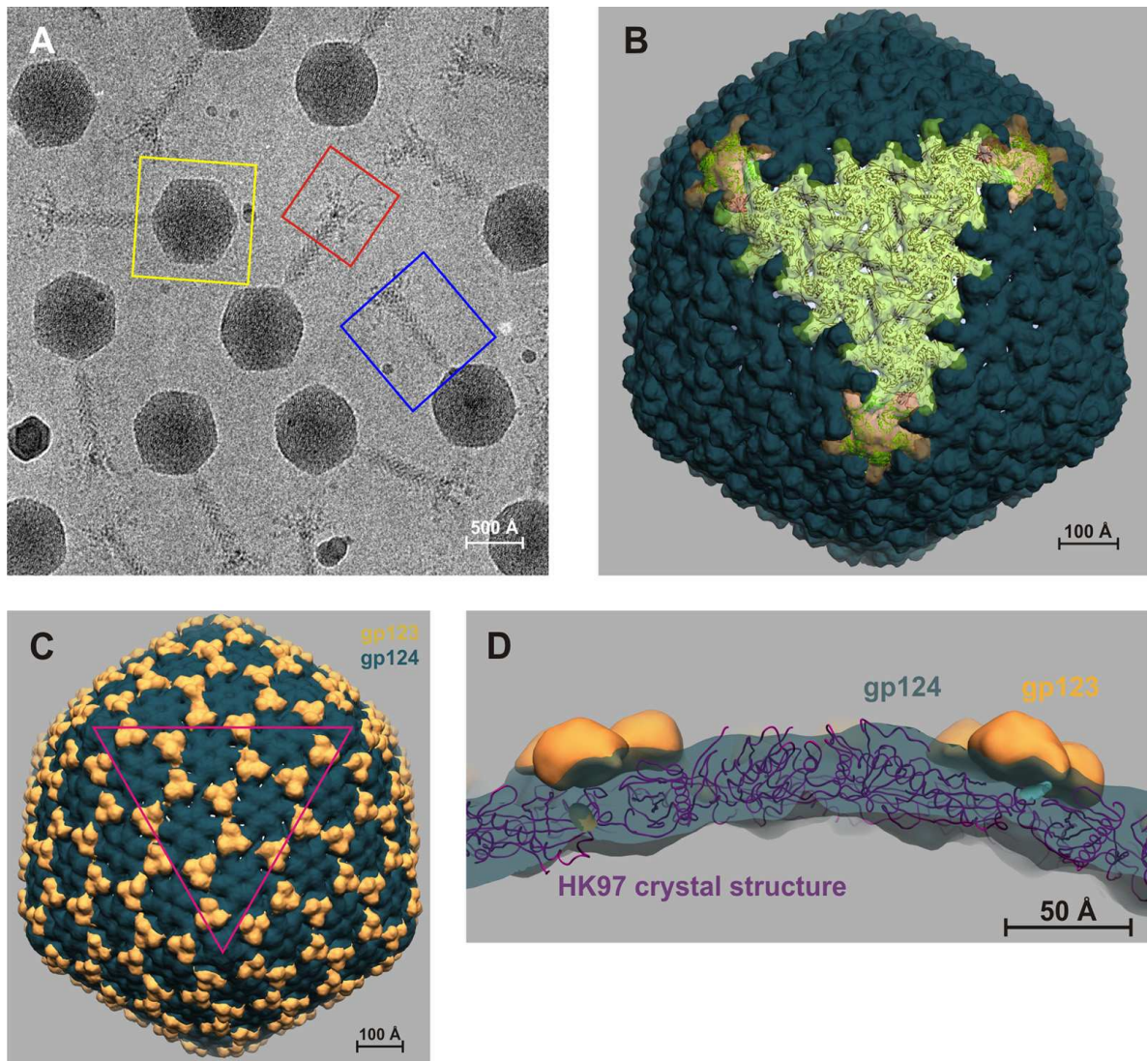


Figure 23: Cryo-EM structure of phi92. **(A)** A raw cryo-EM image of phi92. Three separate reconstructions were calculated for areas highlighted with boxes colored in yellow, blue, and red. **(B)** The structure of phi92 capsid with the imposed icosahedral symmetry. On one face of the icosahedron, the HK97 capsid protein structure is fitted into the cryo-EM density. **(C)** The location of the decoration protein gp123 (gold) on the capsid shell formed by gp124 (moss green). The magenta triangle highlights one face of the icosahedron. **(D)** Interpretation of the capsid cryo-EM density with the help of the crystal structure of the HK97 capsid protein. In panels B, C, and D, the cryo-EM map is contoured at 1.5σ from the mean.

deviations (σ) from the mean. Interestingly, even at this low contour level, the shell is permeated with holes of 13 by 21 Å (D. Schwarzer et al. 2012).

The large body of existing structural data shows that the major capsid proteins (MCPs) of all known dsDNA tailed phages have a HK97 fold and that the capsid shell is composed of flat hexamers and wide-based cone-shaped pentamers formed by the MCP (Wikoff et al. 2000). In agreement with this, the ϕ 92 capsid electron density could be segmented into hexamers and pentamers (Fig. 23B), easily accommodating the respective oligomers of the HK97 capsid protein (PDB ID [1OHG](#)). The hand of the cryo-EM map was determined by comparing the original and mirrored maps with the structure of the HK97 capsid (Kostyuchenko et al. 2003). Due to the limited resolution of the map, we used an automated fitting procedure that maximizes the number of atoms fitted into the density of the data as implemented in UCSF Chimera (Pettersen et al. 2004). The HK97 asymmetric unit comprising one full hexamer plus one subunit of the pentamer was fitted into the original and mirrored maps. Of 15,070 C α atoms, the original and mirrored maps contoured at 1.5 σ do not account for 2,891 and 3,767 atoms, respectively. The positions of the hexamer autofitted with or without the pentamer subunit into the original hand map were virtually identical. The results of the autofit procedure involving the hexamer without the pentamer subunit fitted into the mirrored map did not converge to a unique solution. As a result of these calculations, the triangulation number of the ϕ 92 icosahedron was found to be $T = 13$ *dextro* ($h = 3$, $k = 1$), resulting in 775 copies of the MCP per capsid. Interestingly, the capsids of the isometric T4 phage mutant and phage T5 were reported to have a mirrored icosahedral lattice with $T = 13$ *laevo* (Effantin et al. 2006). Apparently, the HK97 fold has enough plasticity to accommodate both arrangements of the capsid subunits, which have the same T number.

The cryo-EM map of the ϕ 92 capsid displays prominent protrusions on the periphery of the capsomers near the local 3-fold axes (Fig. 23 C). These protrusions represent either a domain which is not present in the HK97 capsid protein or an extra capsid decoration protein (Fig. 23 D). The volume of the electron density suggests that each protrusion is formed by a polypeptide chain with a length of \sim 110 residues. As the ϕ 92 MCP (gp124) is only \sim 50 residues longer than the HK97 capsid protein, the number of additional amino acids is not sufficient to account for the density of all these protrusions. Furthermore, as identified with *HHpred*, the closest ortholog of the ϕ 92 MCP with known structure is the MCP of a prophage from *E.Coli* CTF073 (PDB ID 3BQW). This capsid protein attains an HK97 fold with a short insertion formed by residues 230 to 250. Hexamers and pentamers created by superimposing the 3BQW structure onto the HK97 capsomers showed that these extra residues were at the centers of the capsomers and faced inward. The ϕ 92 cryo-EM density can accommodate this insertion. Therefore, a separate decoration protein that is not part of the MCP must form the ϕ 92 capsid protrusions (Fig. 23 D).

The density corresponding to the decoration protein in the cryo-EM map is as great as that of the

MCP, suggesting that the ratio of the two proteins in the virion is close to 1.0. As there are 775 copies of the MCP in each virus particle, the decoration protein must form a prominent band on the 1-DE gel similar to those seen with other phages carrying capsid decoration proteins. The 1-DE analysis of the ϕ 92 particle is, however, somewhat less clear in this respect (D. Schwarzer et al. 2012). The protein corresponding best to the estimated molecular mass of 13.8 kDa was identified as gp123 of the 1-DE gel in band 19 but also in eight other bands (bands 1, 6, 7, 9, 12, 13, 14, and 17; (Schwarzer et al., 2012)). The apparent molecular mass of gp123 is 13.8 kDa, which is 12% greater than the molecular mass calculated from the cryo-EM density. The observed difference can be attributed to the disorder and “averaging-out” effects, which are common to surface-exposed features in cryo-EM reconstructions (Fokine et al. 2004). The determination that the predicted secondary structure of gp123 is dominated by β -strands can be asserted with a high degree of confidence. Although the evidence is indirect, this observation supports the assignment of gp123 as the capsid decoration protein, since all known capsid decoration proteins and additional domains of MCPs are β -structural elements (Bateman, Eddy, and Mesyanzhinov 1997; Morais et al. 2005; Parent, Khayat, et al. 2010; Qin et al. 2010; Yang et al. 2000).

Orthologs of ϕ 92 decoration protein gp123 are found in phages rv5 and PVP-SE1 and are encoded by genes 61 and 72, respectively. In the absence of any additional evidence, gp72 of PVP-SE1 was annotated as a capsid decoration protein solely due to its abundance on the gel (Santos et al. 2011). The structure of the ϕ 92 capsid obtained in this study confirms the function of the ϕ 92_gp123 homolog PVP-SE1_gp72 and in addition strongly suggests that gp61 of phage rv5 is a capsid decoration protein. As the capsids of rv5-like phages are fairly large, the conservation of the decoration protein must represent a functional requirement, because decoration proteins are known to increase the stability of virus capsids (Iwasaki et al. 2000).

Capsid scaffolding proteins of most phages have two common features: they are rich in α -helical content and the genes encoding these proteins are found immediately upstream of the MCP genes in the genome (Prevelige and Fane 2012). Both criteria apply to ϕ 92 gp122, suggesting that it is the scaffolding protein. In phages rv5 and PVP-SE1, the equivalent genome positions are occupied by genes 62 and 73, respectively. These genes are annotated as scaffolding protein genes, albeit the reasons behind this annotation are unpublished. Neither displayed significant sequence similarity to ϕ 92 gp122, but they are of similar sizes and have predominantly α -helical structures.

The structure of the head-to-tail region was investigated by calculating the 5-fold symmetric reconstruction of the capsid and at the same time increasing the reconstructed volume to include a proximal part of the tail. Compared to the icosahedrally averaged reconstruction, the 5-fold symmetric map of the capsid did not show any new features and was of lower quality. Nevertheless, the 5-fold averaged reconstruction demonstrated that, unlike the extensive neck and whiskers structures found in bacteriophage T4, the ϕ 92 particle carries no collar (Petr G

Leiman et al. 2010). Genes located between the capsid and tail gene clusters are likely to participate in the attachment of the tail to the capsid by expressing either a structural protein of the neck or a chaperone. One such gene is gene 126, and gp126 was found in the phage particle by MS (Schwarzer et al., 2012). The genome location, secondary structure, and molecular mass of gp126 suggest that it is a tail-capping protein (Fig. 24) and most likely an ortholog of T4 gp15 (Petr G Leiman et al. 2010)).

3.2.2. Structure of the tail

The contractile sheath and the central tube are two major components of the phage tail. The sheath has an external diameter of 220 Å and is 840 Å long (Fig. 24). The sheath is a six-start superhelix, with each helical strand containing 24 copies of the gp130 sheath protein, resulting in a total of 144 gp130 subunits per tail. The helical rise is 38.9 Å, and the twist is 26.0 degrees. In comparison, the T4 sheath is also a six-start superhelix with 23 subunits per strand, but the parameters of the helix are slightly different: the rise is 40.6 Å and the twist is 17.2 degrees (Kostyuchenko et al. 2005). This gives the two tails almost the same length and most probably results in the two sheaths contracting and extruding the tail tube from the plane of the baseplate upon attachment to the host cell surface by the same amount. Again, similar to the observations of the T4 tail, the symmetry of the tail tube is not apparent in the reconstruction and its density partially overlaps that of the sheath (Kostyuchenko et al. 2005). Nevertheless, the ϕ 92 tail tube is in all likelihood also a six-start helix assembled of hexamers of the gp131 tube protein, as is found in all related contractile tail-like systems (Petr G. Leiman and Shneider 2012b; Veessler and Cambillau 2011).

Similar to phage capsid proteins, contractile sheath proteins have a common ancestor and appear to be built using a common conserved structure consisting of three domains with a total molecular mass of ~45 kDa (Petr G. Leiman and Shneider 2012b). Some sheath proteins such as those of T4 and its close relatives contain an additional surface-exposed domain and have a greater molecular mass. The 52-kDa ϕ 92 sheath protein gp130 is likely to have the conserved three-domain organization, which also characterizes the sheath proteins of phages Mu and P2 (Petr G. Leiman and Shneider 2012b). None of the available crystal structures of sheath proteins shows a satisfactory fit into the ϕ 92 tail cryo-EM map. The relative orientations of the three constituent domains are likely different from those found in the crystal structures. The resolution of the cryo-EM map is insufficient to fit the domains separately.

The length of the tail is determined by the size of the “tape measure” protein, which is thought to form an alpha-helical structure which extends through the entire central channel of the tail tube (Davidson et al. 2012). The tape measure proteins of ϕ 92 (gp134) and T4 (gp29) are 659 and 590 residues long, respectively, which agrees with the similar lengths of the tails. In the two related

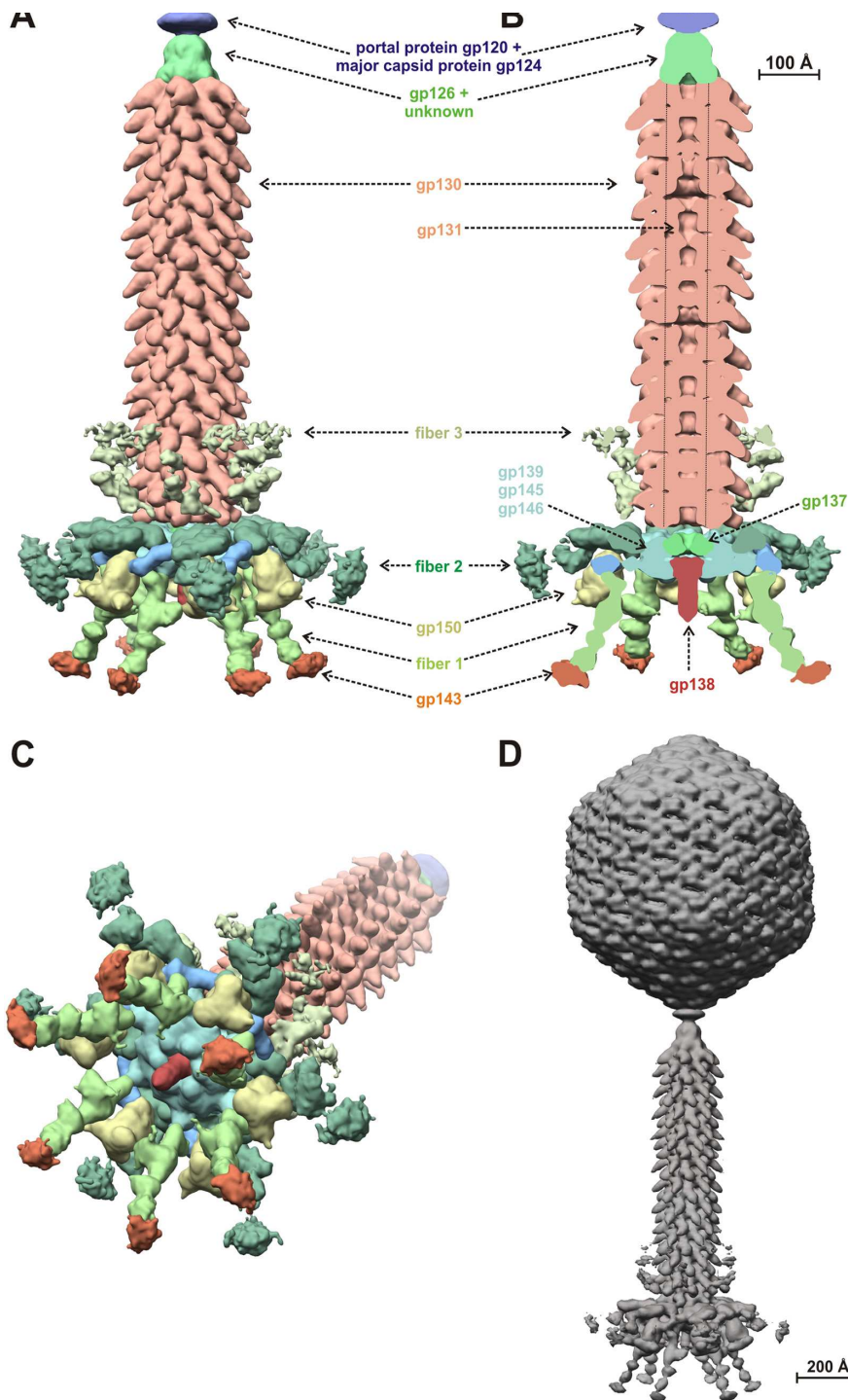


Figure 24: Structure of phi92 tail and baseplate. **(A and B)** Side and cutaway views, respectively, of the phi92 tail structure. A composite map, which includes two overlapping maps (tail sheath and the baseplate) produced separately, is shown. The map is interpreted in terms of the component proteins that were known. **(C)** A tilted view of the tail colored identically to the views in panels A and B. **(D)** The structure of phage phi92 produced by combining three reconstructions (the 5-fold-averaged capsid, the 6-fold-averaged tail, and the 6-fold-averaged baseplate). In panels A, B, and C, the cryo-EM maps are contoured at 1.5σ from the mean.

phages rv5 and PVP-SE1, the proposed tape measure proteins (gp49 and gp57) are significantly longer (778 aa and 795 aa, respectively) than gp134 of ϕ 92 (659 aa), suggesting that the tails of the rv5 and PVP-SE1 phages might be longer than that of ϕ 92.

3.2.3. Structure of the ϕ 92 baseplate

The ϕ 92 baseplate with emanating fibers and tailspikes has a complex structure and resembles an open Swiss army knife (Fig. 24). The bioinformatic analysis outlined above shows that the genes comprising the core part of the ϕ 92 baseplate—139, 145, and 146—are orthologous to those of phages Mu and P2, although slightly larger (e.g., ϕ 92 gp145 is 54 kDa compared to P2 gpI's 34 kDa).

ϕ 92 gp138 and gp137 are orthologs of the gp5-gp27 cell-puncturing complex of phage T4 (Kanamaru et al. 2002; Petr G. Leiman and Shneider 2012b) and thus must form the “cell-puncturing device” or the central hub of the baseplate (Kanamaru et al. 2002). Although not identified by MS, gp138 is part of the phage particle and constitutes the baseplate central spike protein that is essential for breaching the host's outer membrane during infection (Browning et al. 2012). There are only 3 copies of gp138 polypeptides per particle, and they form a densely folded structure resistant to SDS and proteases, which might have prevented its identification in the MS analysis. gp137 forms a stable trimer in solution which is likely to have the same fold and donut-like shape as Mu gp44, T4 gp27, and many other central hub proteins, including those from non-contractile tail phages (Davidson et al. 2012; Petr G. Leiman and Shneider 2012b). Therefore, the ring-shaped density immediately on top of gp138 in the cryo-EM map can be attributed to gp137. Unlike T5 gp5 and gp27 (Kanamaru et al. 2002), the trimers of ϕ 92 gp137 and gp138 form a weak complex in solution.

A combination of bioinformatics, MS, and cryo-EM analyses showed that each ϕ 92 particle carries at least five different tailspike and tail fiber proteins (or receptor binding proteins [RBPs]), gp141, gp142, gp143, gp150, and gp151, which are located at the 3' end of the tail operon. Three sets of RBPs emanating from the baseplate can be distinguished in the cryo-EM map. They are directed downward, sideward, and upward (fibers 1, 2, and 3 in Fig. 24). The downward-pointing fibers have the strongest cryo-EM density, whereas those pointing upward have the weakest. The cryo-EM map is a representation of all particle images that were included in the 3D reconstruction procedure. As a consequence, the RBPs pointing downward are likely to have the most similar orientations in all particles whereas those pointing upward are likely to have the most dissimilar and thus are “averaged out” in the final 3D maps. It is also possible that each particle carries a full complement of the downward-pointing RBPs but only a few of those pointing sideward and even fewer of those pointing upward. The latter hypothesis is less likely, as it has never been reported for any other virus or phage. The quality of the cryo-EM map does not allow reliable identification

of the gps comprising these RBPs. Nevertheless, the genome of $\phi 92$ contains enough gps to account for all the RBPs, with each having a full occupancy per particle .

The RBP regions of the $\phi 92$, PVP-SE1, and rv5 genomes show the highest degree of mosaicism compared to the rest of the structural operon (Fig. 2A), a feature common to all groups or families of phages (S. R. Casjens and Molineux 2012). All known RBPs have a modular architecture and consist of up to three different parts: (i) a small N-terminal domain or peptide for attachment of the RBP to the baseplate or tail region; (ii) a large central part that forms an extended trimeric structure with a coiled coil or beta-helical fold and is the key factor for host specificity; and (iii) a small C-terminal domain that usually functions as an intramolecular chaperone. The three parts of an RBP might be encoded by different ORFs or belong to a single polypeptide chain. In the three rv5-like phages, the N-terminal and C-terminal parts of several RBPs are significantly better conserved than their central parts. For example, the N-terminal regions of $\phi 92$ gp150 and gp151 show similarity to those of rv5 gp28 and gp41, respectively, and the C-terminal parts of rv5 gp28 and $\phi 92$ gp151 are also similar at the amino acid level (see Fig. 2A; see also Table S1 in the supplemental material (D. Schwarzer et al. 2012)). Phages are able to exchange the host-specific modules, while maintaining the N-terminal tail attachment domain unchanged, to increase their host range or to adapt to a new environment (Stummeyer et al. 2006). The C-terminal intramolecular chaperones of some RBPs are known to facilitate folding of foreign RBPs (Mühlenhoff et al. 2003; Eike C. Schulz et al. 2010; David Schwarzer et al. 2007; Xiang et al. 2009).

$\phi 92$ infects K1 and K92 strains of *E.Coli* (Kwiatkowski et al. 1983), which are encapsulated by polySia with $\alpha 2,8$ - or alternating $\alpha 2,8/\alpha 2,9$ -linkages, respectively (Barry and Goebel 1957; Furowicz and Orskov 1972; McGuire and Binkley 1964). The polySia capsule-digesting tailspike of the $\phi 92$ particle is likely to be encoded by gene 143. It displays 54% sequence identity to the well-studied endosialidase of coliphage K1F (endoNF) through the whole length of the protein, suggesting that gp143 has the same structure as the mushroom-shaped endoNF (Mühlenhoff et al. 2003; Stummeyer et al. 2005). By analogy with other endosialidases (endoNF, endoNE, etc.), we refer to it as endoN92. It is located at the distal end of the downward-pointing fibers in the cryo-EM map (Fig. 24). This fiber ends in a bulge that corresponds to the sialidase domain of endoN92. The narrow stalk domain of endoNF forms the tip of the 650 Å-long fiber and thus is disordered (Fig. 24).

At present, the other putative fibrous proteins (e.g., gp141, gp142, and gp151) cannot be located in the cryo-EM map without additional information regarding their structure or interactions with other proteins (fibers 1 to 3 in Fig. 24 A). However, gene 150 of $\phi 92$ is likely to encode a tailspike protein and the cryo-EM map contains a density that might correspond to gp150. HHpred analysis shows that gp150 is likely to have the trimeric β -helical fold that characterizes many other phage tailspikes (e.g., $\phi 29$, P22, Sf6, HK620, or K5A), and therefore is likely to have a similar compact

and trimeric shape (Petr G. Leiman and Molineux 2008). The cryo-EM map contains a tailspike-like density with a clear 3-fold symmetry and a volume sufficient to account for three polypeptide chains of gp150 (Fig. 24).

The sequence of gp150 shows similarity to that of a colanic acid-degrading enzyme and, in analogy to other glycosidases, can be referred to as “colanidase.” The ϕ 92 host strain *E.Coli* K92 produces colanic acid at lower temperatures than the K92 polySia (Navasa et al. 2009), and one of the possible functions of gp150 is likely that of breaching the colanic acid layer. Interestingly, the catalytic domain of gp49 from phage PVP-SE1 has a significant sequence similarity to that of gp150, whereas its N-terminal region is related to the N-terminal part of ϕ 92 gp151, suggesting that the colanidase might be attached to the PVP-SE1 baseplate at a site different from that of ϕ 92 gp150. It is possible that the ϕ 92 colanidase attachment site in PVP-SE1 is occupied by gp41, since their N-terminal domains share high sequence similarity (64.2%).

3.3. Concluding remarks

This report describes the genome, structure, and host range of the *Enterobacteria* phage ϕ 92. The ϕ 92 genome displays mosaicism similar to that of other phages and was found to be closely related to the genomes of phages rv5 and PVP-SE1, favoring suggestions to group them into the new genus of rv5-like phages. The structural operons are well conserved, strongly suggesting that the structures of the capsid (including the location of the decoration protein) and the tail and the organization of the major part of the baseplate are conserved. The baseplates of rv5 and PVP-SE1 are likely to have a similar Swiss army knife-like structure, although the numbers of tail fibers and tailspikes emanating from the baseplate and their exact attachment sites on the baseplate might differ. This report illustrates that phages display an enormous potential to change host recognition proteins by intra- and interspecies recombination and by mutations, resulting in their ability to recognize new host receptors. This phenomenon plays an important role in the natural ecology of phage-host interactions and can be exploited in the selection of therapeutic phages having desired “target ranges.”

3.4. Materials and methods

3.4.1. Phage propagation

E.Coli strain Bos12 (O16:K92:H⁻) was used for propagation of coliphage ϕ 92 (Kwiatkowski et al. 1983). *Enterobacteria* phage K1F was grown on *E.Coli* U9/41 (O2:K1:H4) (Gerardy-Schahn et al. 1995; Gross, Cheasty, and Rowe 1977). Purification of phage particles and isolation of phage DNA were performed as described previously (Gerardy-Schahn et al. 1995). The protein content of phage particles was estimated using the Bio-Rad protein assay following the manufacturer's

guidelines.

3.4.2. Cryo-electron microscopy and reconstruction of ϕ 92

Low-dose cryo-EM was performed as described previously (T. S. Baker and Henderson 2006). The images were recorded on Kodak SO-163 film using a Philips CM300 FEG microscope at a magnification of $\times 33,000$ with a radiation dose of $\sim 20 \text{ e}^-/\text{\AA}^2$ and a defocus of -3.5 to $-2.0 \text{ }\mu\text{m}$. The images were scanned using a Zeiss SCAI scanner with a $7\text{-}\mu\text{m}$ step size and were binned to obtain a pixel size of $2.108 \text{ }\text{\AA}$. The tail and baseplate images were binned further to give a pixel size of $4.216 \text{ }\text{\AA}$, whereas the final capsid reconstruction was calculated with a pixel size of $2.811 \text{ }\text{\AA}$. The capsid and baseplate/tail datasets contained 1,137 and 985 boxed images, respectively, which were picked from 38 micrographs. The contrast transfer function (CTF) was calculated and then corrected with either EMAN (Ludtke, Baldwin, and Chiu 1999) or CTFFIND3 (Mindell and Grigorieff 2003) software.

The icosahedral reconstruction of the ϕ 92 capsid was calculated with EMAN from scratch. The initial model was constructed with a small subset of initial data using EMAN routines. This model was then subjected to several rounds of refinement until convergence. The reconstruction of the ϕ 92 tail was obtained in two steps. Initially, the structure of the distal part of the tail containing the baseplate and about one-third of the sheath was calculated. Then, the reconstructed volume was increased to include the entire sheath, excluding the fibers emanating from the baseplate.

The T4 baseplate-tail tube complex (EMDB-1048) was used as the initial model for the baseplate reconstruction. The SPIDER (Shaikh et al. 2008) and EMAN software suites were employed in image processing. The reconstruction was initiated with EMAN, but due to the large discrepancy between the initial model and the ϕ 92 images, the refinement procedure was quickly trapped in a local minimum, with tail fibers being aligned to the central spike and the tail sheath layers incorrectly cross-correlated with each other. This problem was overcome by modifying this “ensemble” structure with a threshold-based Gaussian-filtered mask and limiting the search angles to a region of in-plane orientations. The search angles for the tilt were restricted to between 40 and 90 degrees, as most phage particles had their long axes roughly parallel to the imaging plane. These operations and subsequent refinement were performed with SPIDER. The resolution of the reconstruction was found to be $26 \text{ }\text{\AA}$ using the Fourier shell correlation coefficient method, assuming the limit of resolution to be at the point where the correlation coefficient drops to below 0.5 (van Heel and Schatz 2005).

The tail reconstruction was calculated by increasing the box size of the baseplate map to include the sheath in its entirety. The center of this box was then shifted to roughly coincide with the sheath's center. The reconstruction was refined with FREALIGN (Grigorieff 2007). sixfold symmetry was used for all orientation searches and reconstructions of the baseplate and the tail. As the final

maps had shown the features expected to be seen at their medium resolutions, the 6-fold symmetry assumption was correct. The left-right structural orientation ("hand") of the reconstruction was confirmed with atomic force microscopy (AFM) analysis as described in reference (Browning et al. 2012). Fitting of the HK97 capsid structure into the ϕ 92 cryo-EM map was performed with UCSF, Chimera (Pettersen et al. 2004).

4. Bacteriophage CBA120

4.1. Summary

Escherichia phage CBA120, which belongs to a newly described genus “*Viunalikevirus*” within the family *Myoviridae*, is morphologically very similar to the classic phage Vil of *Salmonella enterica* serovar Typhi. Recently, such phages, related to *Salmonella* Typhi phage Vil, have been discovered infecting a variety of different *Enterobacteriaceae*. CBA120 specifically infects *E.Coli* strain O157, an important food-borne pathogen. The mature CBA120 virion contains at least 37 different proteins, 11 of which show structural homology to coliphage T4. The baseplate and method of adhesion to the host are, however, very different from those of either T4 or the cyanophages. At recent, CBA120 is the most host specific phage that has observed. It is a dsDNA myovirus with a genome size 157.3kb. The baseplate is very complex with numerous protruding tail spikes and fibers. At present and to our knowledge, it is the most complex tail structure investigated by anyone in the world. One of the CBA120 tailspikes is a candidate for biofilm-forming bacterial infections therapy.

Here we report the first detailed three-dimensional structural organizations of *E.Coli Myoviridae* bacteriophage CBA120, the first structural insight into the “*Viunalikeviruses*” genus.

4.2. Introduction

Bacteriophages are the most abundant form of life in the biosphere. Tailed phages, or *Caudovirales*, represent the overwhelming majority of bacteriophages population, and are traditionally divided into three families: *Myo*-, *Sipho*- and *Podoviridae*. Among them, the myophages with long contractile tails have the most complex tail structures, composed of hundreds copies of tens of different proteins. *Myoviridae* phage tails are complex molecular machines, generally consisting of a baseplate with tail fibers and a long, non-contractible tube surrounded by a contractile sheath. However, current morphology-based classification of bacteriophages requires clarification, since unusual *Myoviridae* bacteriophages have been observed.

Bacteriophage vb_EcoM_CBA120 (CBA120) was isolated from a cattle feedlot in a systematic search for the group of phages targeting important food-borne pathogenic *E.Coli* O157:H7. It is the first observed phage, which is highly specific for *E.Coli* O157:H7 serotype, and infects 13 of 17 tested O157:H7 strains and only one among 72 non-O157:H7 *E.Coli* strains (Kutter et al. 2011).

CBA120 has a long contractile tail, thus it is a *Myoviridae* phage, with a genome size 157.3 kb. Genomic and cryo-EM analysis revealed that CBA120 is an unusual myophage. It lacks outer baseplate proteins and the long tail fibers, which appears to be characteristic of *Myoviridae* family.

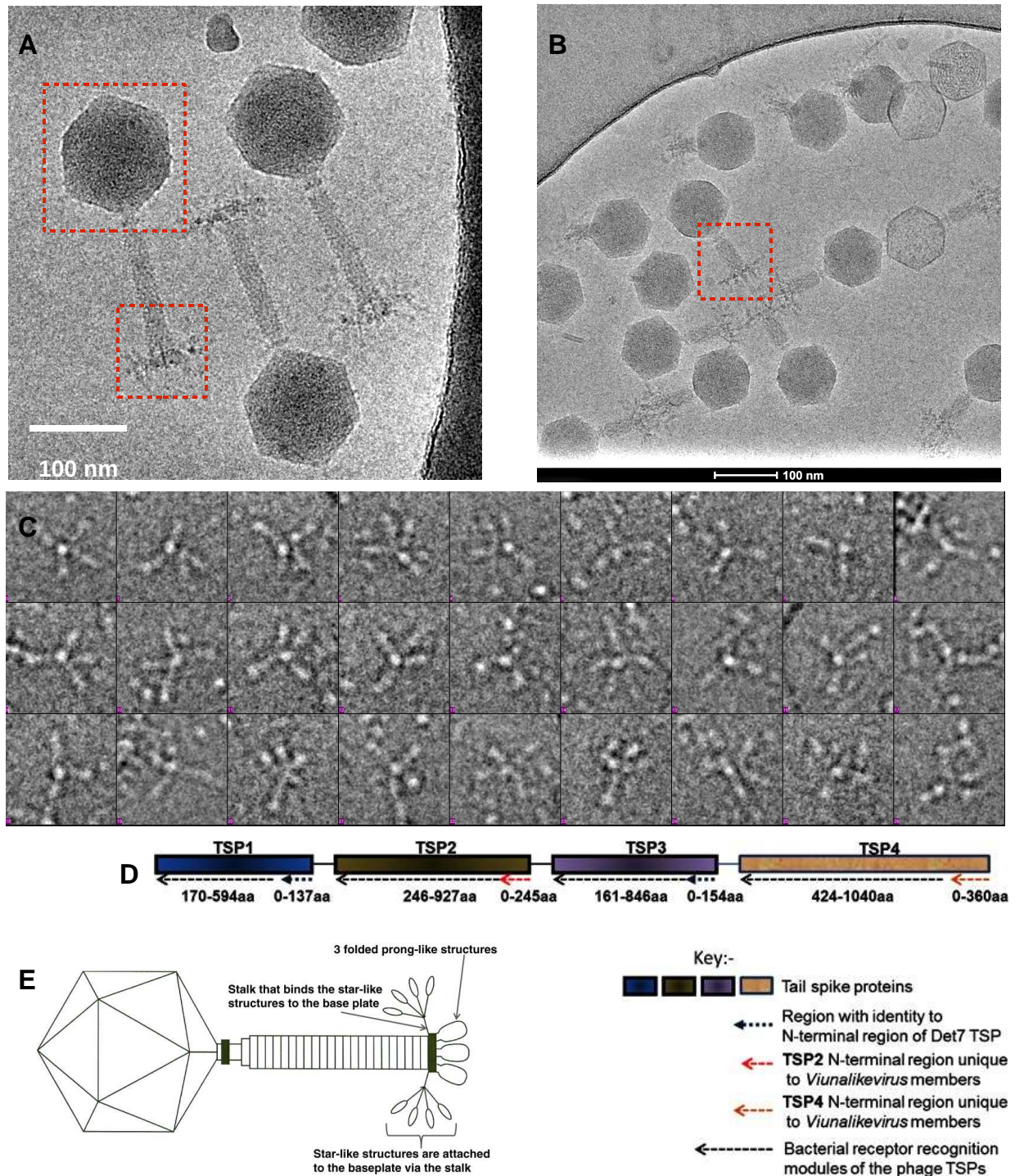


Figure 25: Cryo-EM images of the coliphage CBA120 in extended (**A**), and contracted conformation (**B**). Images of the CBA120 tailspikes boxed from over-focused cryo-EM micrographs (**C**). Generalized description of the tailspike proteins of the genus “*Viunalikeviruses*” (**D**). Proposed morphology of the “*Viunalikeviruses*” genus with star-like tailspikes (**E**). (D and E are adapted from (Adriaenssens et al., 2012)).

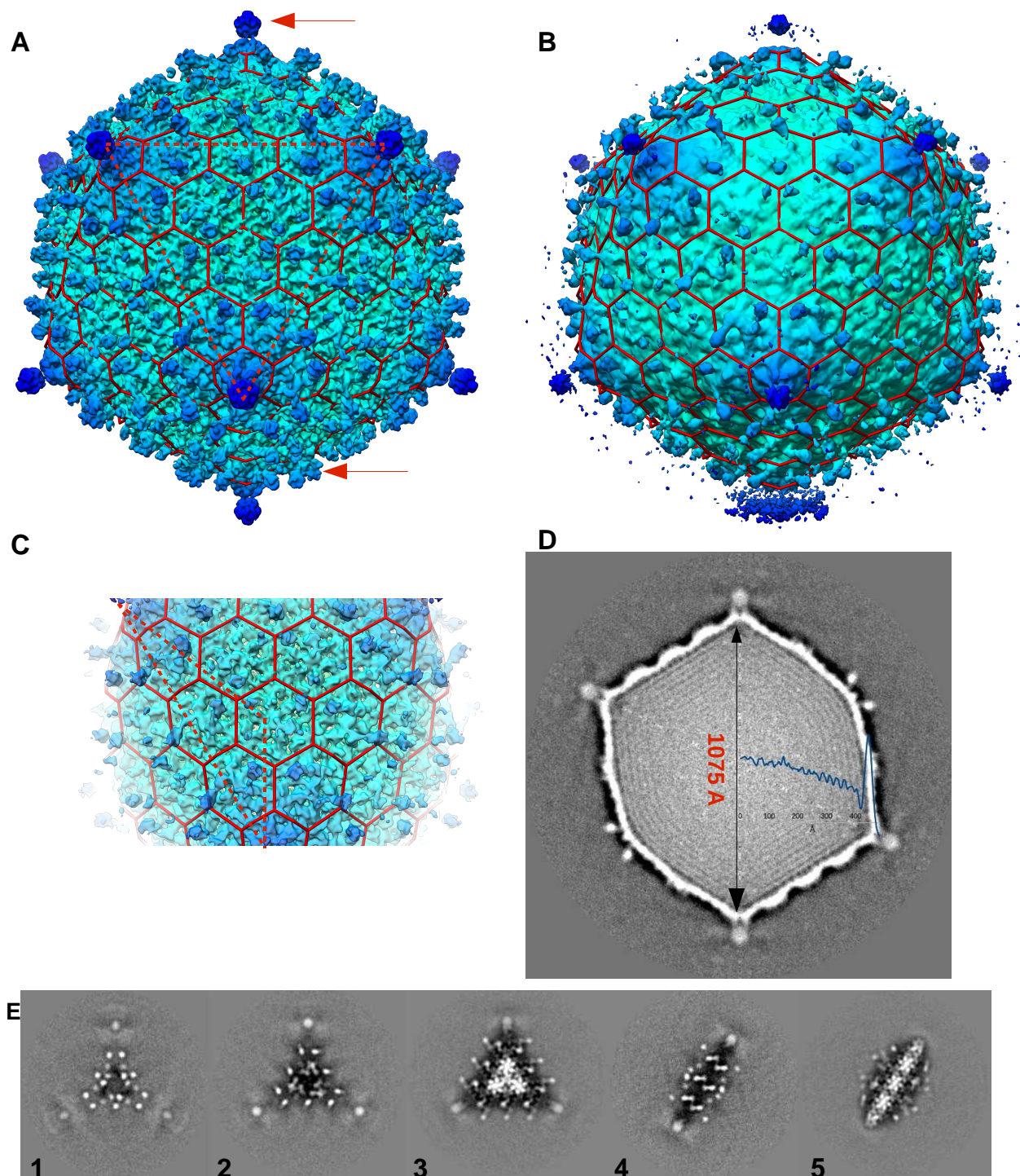


Figure 26: Cryo-EM structure of CBA120. The structure of CBA120 capsid with the imposed icosahedral (A) and 5-fold (B) symmetries. Decoration proteins highlighted with red arrows. T=16 icosahedral shell highlighted in red. (C) Face of icosahedron with asymmetric triangle highlighted in red. (D) Central section through 5-fold axis of capsid cryo-EM structure with radial profile colored in blue. (E) Sections through the icosahedron face orthogonal to 3-fold axis (1-3), and through the icosahedron edge orthogonal to 2-fold axis (4-5).

In contrast, it contains several genes for putative tailspikes, which are more common for *Podoviridae* phages, phages with short non-contractile tails. Comparative genomic analysis revealed five other closely related myophages with *Podoviridae*-like multi-spikes, like a classical *Salmonella* Typhi phage Vil, which was described in 1930s, and used to treat typhoid fever. All seven sequenced myophages constitute a new genus “*Viunalikeviruses*” after the phage Vil. Despite the similarity to the T4 superfamily group, “*Viunalikeviruses*” genus morphologically and genomically differs from the other phage world, and is characterized by several distinguishing features. First, they are incorporated a non-canonical nucleotide, probably hydroxymethyluracil (HMdU) in place of thymine (Adriaenssens et al. 2012). But the most important and unique signature of the “*Viunalikeviruses*” is the genome organization, namely arrangement of the tailspike genes, which now believed to encode four tailspike proteins (TSPs) (Fig. 25). Lack of the long tail fibers, and presence of the TSPs result in a different from T4-like phages strategy of host recognition.

CBA120 is the first encountered “*Viunalikeviruses*” phage infecting *E.Coli*. The virion consists of isometric 1075-Å-diameter icosahedral capsid, enclosed 157.3kb dsDNA genome, and 1010-Å-long contractile tail, with tiny baseplate and umbrella-like tailspikes network. The infection process likely includes two stages, as for other studied tailed phages. However, the combination of *Myoviridae* tail core and *Podoviridae*-like tailspikes makes the infection mechanism more intriguing and obscure. In this report, we demonstrate the detailed three-dimensional organization of CBA120 virion by cryo-EM, which reveals complex molecular machinery evolved to specifically kills pathogenic *E.Coli* strain O157:H7.

4.3. Results and discussion

4.3.1. Cryo-EM reconstruction of the CBA120 virion

Due to the large size and symmetry mismatches between different parts of CBA120 virion, we used the divide-and-conquer (Fig. 25) image-processing approach and the reconstructions of the CBA120 capsid, extended and contracted tail-baseplate complexes were calculated separately with the most appropriate symmetry for each part. The capsid is reconstructed to resolution 14 Å (FSC=0.143 criterion) with icosahedral symmetry applied, and to 22 Å with only 5-fold rotational symmetry applied, the tail-baseplate complex to 25 Å in the contracted state, and to 16 Å in the extended state, and extended tail sheath to 12 Å with helical symmetry applied.

4.3.1.1. Structure of the CBA120 capsid and genomic DNA

CBA120 capsid shell is an isometric icosahedron, and was reconstructed to 14 Å resolution. The icosahedral shell dimensions are 1074 Å, 913 Å, and 934 Å along one 5-fold axis, one 3-fold axis, and one 2-fold axis (Fig. 26). Central section of cryo-EM reconstruction shows that the CBA120

capsid shell is 33-42 Å thick in a map contoured at 1.5 standard deviations (σ) from the mean, and all of the capsid is required to accommodate the genomic DNA. The CBA120 capsid electron density could be segmented into flat hexamers and wide-based cone-shaped pentamers formed by the major capsid protein (MCP), according to quasi-equivalence principle, and the triangulation number of the CBA120 icosahedron was found to be $T = 16$ ($h = 4$, $k = 0$), resulting in 960 copies of the MCP per capsid (Fig. 26).

The major capsid protein gp23 has a length of 440 residues and a molecular weight (MW) of 48 kDa from the gp23 sequence. From the corresponding cryo-EM map the mass estimations are 50kDa and 53 kDa for pentameric and hexameric capsomer subunits. The closest ortholog of CBA120 MCP gp23 is the head vertex protein gp24 of phage T4 (P. G. Leiman et al. 2003), except for the 50-residues insertion domain similar to the ferric uptake regulator (Fur) protein. *Fur* protein consist of two domains, N-terminal DNA-binding domain linked by a hinge region to a C-terminal dimerization domain (Butcher et al. 2012).

The large body of existing structural data shows that the heads of all known dsDNA tailed phages are built of similar capsid protein HK97 fold. The gp24 of T4 phage has similar HK97 fold, except from 60-residues insertion domain (Fokine, Leiman, et al. 2005). Interestingly, all 7 sequenced phages currently assigned to the "*Viunlikeviruses*" genus (*Salmonella* phages Vil, SFP10, PhiSH19 coliphages PhaXI, CBA120, ECML-4, *Shigella* phage Ag3 and *Dickeya* phage Limiestone) has a 440-residues-long major capsid protein similar to T4 gp24 in their genomes.

The distance between CBA120 capsid shell adjacent hexamers is 140 Å, and that between equivalent positions in pentamers and adjacent hexamers is 135 Å. The same arrangement was observed in T4 phage (Fokine, Leiman, et al. 2005) In HK97 phage the distance between adjacent hexamers is 132 Å, which, on average, was observed in many others dsDNA tailed phages (Aebi et al. 1974; Petr G Leiman et al. 2007; X. Liu et al. 2010; Morais et al. 2005; Thuman-Commike and Chiu 2000; Wikoff et al. 2000). 11 pentameric vertices of the CBA120 icosahedral capsid are occupied with the spherical 60-Å-diameter mace-head-like decoration densities, which protrude radially along the 5-fold axes. The estimated mass from the cryo-EM map is ~70 kDa (Fig. 26, Tab. 3).

In addition, there are protruding doubled knob-like densities, ~50-Å-long with a mass of ~25 kDa (cryo-EM map), 180 copies of which follows the edges of the icosahedral capsid. Each of doubled protrusions emanates from two opposite MCP subunits, forming hexameric capsomers (Fig 26). Probably, similar protrusion-like disordered densities are found on hexameric capsomers adjacent to the local 3-fold axes (Fig. 26 A, C). Densities disorder designate the violation of the truly icosahedral symmetry, which was observed only in marine virus Syn5 (Gipson et al. 2014).

CBA120 capsid cryo-EM map was reconstructed assuming most appropriate icosahedral symmetry,

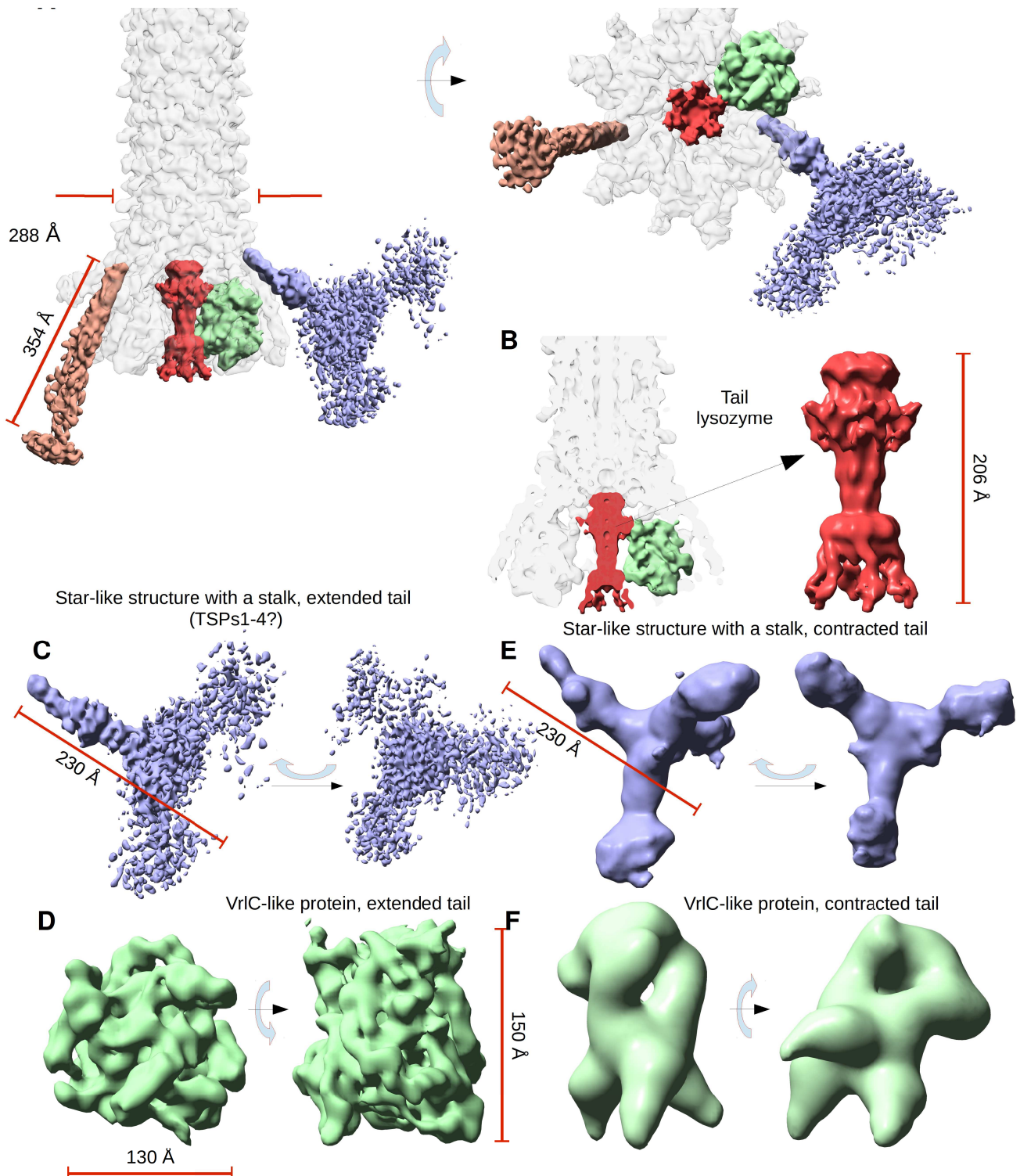


Figure 27: Cryo-EM structure of CBA120. **(A)** The structure of CBA120 tail with 6-fold symmetry imposed, and putative tailspikes and fibers highlighted, side and bottom views. **(B)** Cutaway view of the cryo-EM tail structure, and enlarged central spike. Side and bottom views of the star-like structure with a stalk composed of TSP1-TSP4 tailspikes in extended **(C)** and contracted **(E)** tail conformations. Bottom and side views of the VrlC-like tailspike in extended **(D)** and contracted **(F)** tail conformations.

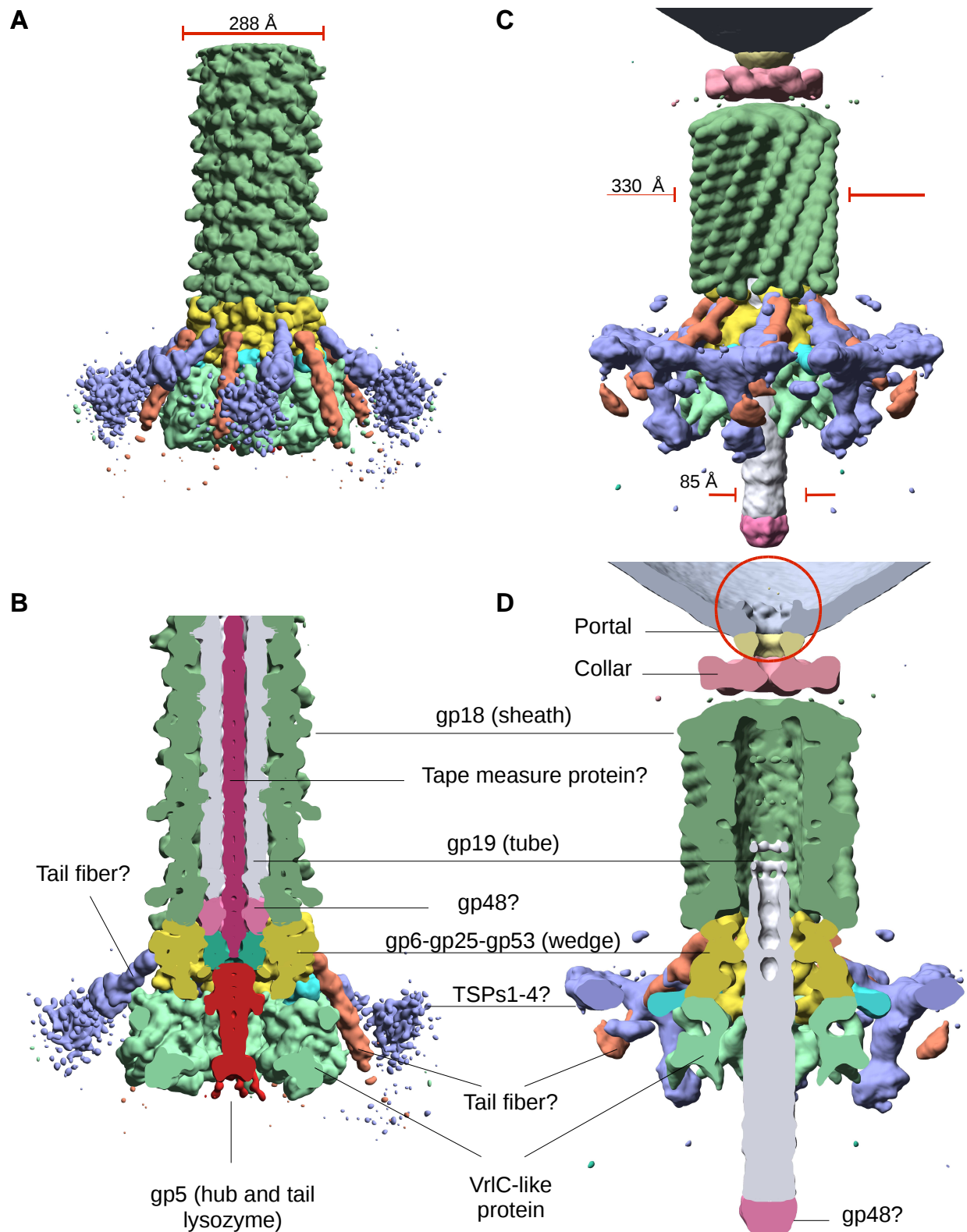


Figure 28: Cryo-EM structure of CBA120. Main components are segmented, colored and highlighted. **(A)**, **(B)** Side and cutaway views of CBA120 tail structure in extended conformation with the imposed 6-fold symmetry. **(C)**, **(D)** Side and cutaway views of CBA120 tail structure in contracted conformation with the imposed 6-fold symmetry.

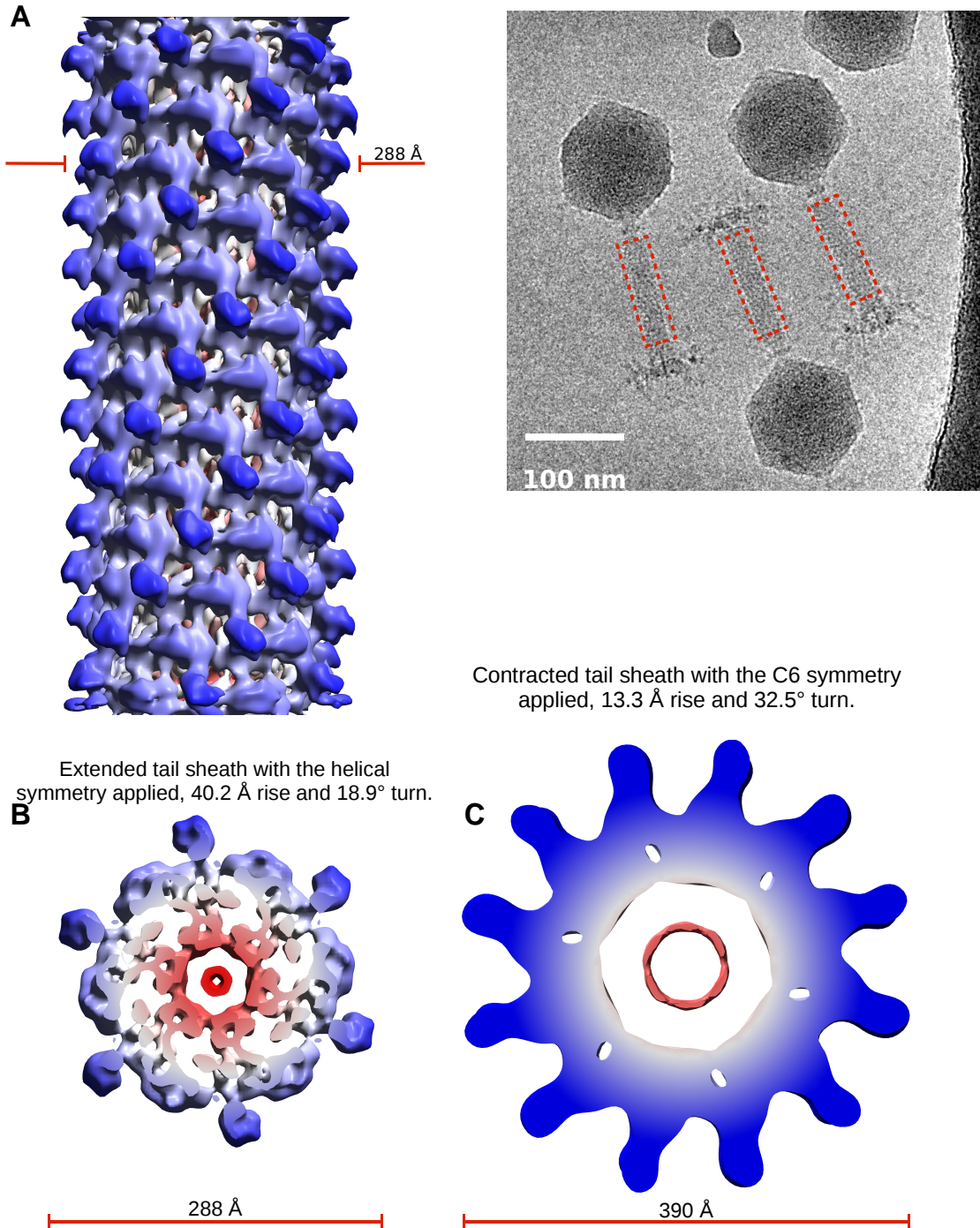


Figure 29: Cryo-EM structure of CBA120. **(A)** A raw cryo-EM image of CBA120 with reconstructed area highlighted in red. Cryo-EM structure of CBA120 tail sheath with imposed helical symmetry. Top views of CBA120 tail sheath structure in extended conformation with helical symmetry imposed **(B)** and contracted conformation with 6-fold symmetry imposed **(C)**.

which observed in all studied dsDNA bacteriophages. However, reconstruction with the symmetry relaxation from icosahedral to 5-fold rotational symmetry (C5) also shows disordered pattern of the 3-fold axis adjacent decoration proteins (Fig. 26 B). Probably, decoration densities do not follow the local 3-fold or 2-fold symmetries, or have a fixed orientations for all hexameric capsomers in the asymmetric triangle (Fig. 26 B). Possible candidates for the decoration proteins should be β -strands-reach proteins, since all known capsid decoration proteins and additional domains of MCPs are β -structural elements (Bateman, Eddy, and Mesyanzhinov 1997; Fokine, Leiman, et al. 2005; Morais et al. 2005; Parent, Khayat, et al. 2010; Qin et al. 2010; Yang et al. 2000), with masses of ~ 70 kDa and ~ 24 kDa, estimated from the pentameric and hexameric decoration densities in cryo-EM structure (Table 3).

The central section of the capsid structure makes visible at least 7 concentric layers of densely packaged 158 kb dsDNA genome. The layers are separated by 25 Å, similar to all other studied tailed phages. The central part of the packaged DNA is very dense, without distinguishable structural features, which is common for dsDNA phages.

4.3.1.2. Structure of the CBA120 contractile tail

CBA120 extended and contracted tails were reconstructed to 16 Å and 25 Å resolutions assuming 6-fold rotational symmetry. Helical reconstruction of the tail sheath to 12 Å resolution was performed to enforce the average signal from the individual sheath subunit, and refine sheath helical parameters.

The CBA120 contractile tail is attached to one of the 12 pentameric vertices through a T4-like neck with a collar, breaking the truly icosahedral symmetry of the capsid. The contractile tail length is 1010 Å and diameter is 228 Å, similar to dimensions reported to other “*Viunaliikeviruses*” phages (Adriaenssens et al. 2012). Being *Myoviridae* bacteriophage, the CBA120 major components of the contractile tail are the contractile sheath and the central non-contractile tube.

Contractile sheath show 23 transverse striations surrounding a tail tube, resembles that of T4 phage (Petr G Leiman et al. 2010). The tail sheath is formed with 138 copies of 631-residues-long protein orf197 (MW = 68 kDa), ortholog of 659-residues-long T4 tail sheath protein gp18 (MW=71kDa). Corresponding cryo-EM map has a mass of 69 kDa. The orthologous T4 sheath protein has 3 domains which can move relative to each other during contraction process (Kostyuchenko et al. 2005). The tail sheath is a right-hand 6-fold 6-start helix, with 40.2 Å helical rise and 18.9° helical turn (Fig. 29). T4 sheath has similar helical parameters (40.6 Å rise, 17.2° rotation), however CBA120 sheath subunit is tilted downward $\sim 28^\circ$ with respect to the 6-fold tail axis, compare to the horizontal T4 sheath subunits orientation (Kostyuchenko et al. 2005).

The non-contractile tube is composed of copies of 177-residues-long (MW=20kDa) protein orf194,

whose closest ortholog is 163-residues-long T4 tail tube protein gp19 (MW=18kDa). The tail tube is smooth, without distinct features as seen from the cryo-EM map, however, it likely consists of the same number subunits as there are sheath subunits, and follows the same helical parameters. There is a density along the center of the tail tube, probably corresponding to the tape-measure protein, which determines the length of the tail. The corresponding protein is still not annotated in the genome.

The sheath is connected to the capsid by a 132-Å-long and 90-100-Å-diameter neck, with a 250-Å-diameter and 50-Å-thick collar, which has a hexameric features visible in cryo-EM map of the CBA120 contracted tail (Fig. 28 C). The extended tail reconstruction does not include the neck region, thus its dimensions have been estimated from the capsid reconstruction with the part of the tail. Because of the 5-fold capsid rotational symmetry the 6-fold features of the tail and the neck were averaged out, and only shape and linear dimensions are visible. Nonetheless, the CBA120 genome contains T4 neck structural proteins homologs: gp3 and gp15 tail completion and sheath stabilizer proteins, gp13 and gp14 neck proteins. No fibrin related densities has been observed. The contracted sheath is a 390 Å-long and 330 Å-diameter, with a 960 Å-long, 85 Å diameter tube. The tail tube pass through the 100 Å-wide hole in the center of the baseplate and protrudes about 530 Å. The uniformly contracted gp18 sheath subunits is a right-hand 6-fold 6-start helix, with axial rise 13.3 Å and a rotation angle of 32.5° (16.4 Å and 32.9° in T4 contracted sheath). The sheath does not interact with the 85 Å diameter tail tube, since the inner sheath diameter is 110 Å.

4.3.1.3. Host cell adhesion apparatus of CBA120

CBA120 is a bacteriophage from *Myoviridae* family, i.e. phages with long contractile tails. *Myoviridae* phages have the most complex tail structures, with a complex adhesion apparatus, which generally includes baseplate with tail fibers. Tail fibers are trimeric proteins, which help bacteriophages reversibly bind primary receptors and then irreversibly bind a secondary receptor. Irreversible binding triggers the sheath contraction and release of phage DNA into the host cell. Tail fibers usually lack catalytic activity.

As mentioned above coliphage CBA120 appears to be an unusual myophage, as it lacks the outer baseplate proteins and the long tail fibers, representative of *Myoviridae* phages, and contains putative tailspikes genes, which are more common for *Podoviridae* phages with short non-contractile tails. Tailspike proteins (TSPs) are attached to the baseplate and have enzymatic activity (endoglycosidase), which is essential to assist the phage in attachment to primary receptors and degradation of a thick layer of long lipopolysaccharides (LPSs). After penetration through the polymer tailspikes irreversibly bind to the secondary receptors of outer membrane, which triggers DNA ejection. Tailspike proteins has a short, 100-150-residues-long N-terminal head-

binding domain which attaches to the baseplate, and longer (400-600 residues) C-terminal receptor binding domain with a LPS binding site and catalytic site (Fig. 25). The bacteriophage host range and specificity is determined, obviously, by the divergence of the receptor binding domains (S. R. Casjens and Molineux 2012).

CBA120 shows conservation of the proteins found in the central part of the T4 baseplate (Kutter et al. 2011), namely gp25, gp6, gp53 and gp48, which can be identified in the cryo-EM densities of extended and contracted tails (Fig. 28). The outer baseplate proteins gp7, gp8, gp9, gp10, and gp11, gp12 tail fibers which are parts of T4's long and short tail fibers, responsible for reversible and irreversible absorption of the T4 virion, are all missing in the Vil group. However, most likely there are non-annotated proteins in CBA120 genome which replace some of the missing proteins, like orf45, orf47, orf215, orf216, detected in proteomic studies (Kutter et al. 2011).

The unique "*Viunalikeviruses*" adhesion apparatus locates at the distal end of the CBA120 contractile tail, and is composed of a thin baseplate with a terminal tailspikes attached. In contracted tail the baseplate is a 85 Å thick and 215 Å in diameter. The tailspikes arrangement is complex, and resembles an umbrella-like structure (Fig. 27, 28), similar to *Salmonella* Typhi phage Vil (Pickard et al. 2010).

Compare to proposed "*Viunalikeviruses*" general morphology, CBA120 has 6 copies of prong-like structures, which has been associated with the biggest protein in the CBA120 genome, 1612-residues-long orf209 (MW=177kDa). The protein is a homolog of the coliphage ECML-4 virulence-related (VrIC) protein. Cryo-EM map reveals 6 copies of massive folded prong-like structures 150 Å long and 130 Å thick, with distinct 3-fold rotational symmetry (Fig. 27, 28). Trimeric features of these VrIC-like proteins become more and more apparent in cryo-EM map of the contracted tail, while suffered structural changes after the contraction (Fig. 27, 28). The VrIC-like proteins mass estimations from the extended and contracted tails reconstructions are 180 kDa and 176 kDa (Tab. 3).

There are 6 copies of the star-like structure, which are attached to the baseplate via the stalk (Fig. 27, 28). These structures have been associated with four distinct putative tailspike proteins, TSP1-4 (ORFs 210-213), encoded in CBA120 genome (Adriaenssens et al. 2012), which umbrella-like arrangement is visible on electron micrographs (Fig. 25). At recent, Vil-like phages are believed to have four tails spike proteins, which supports by cryo-EM observations (Fig. 25). Each sequenced Vil-like phage contains unique TSPs for which the targets have not yet been identified. Four tailspikes are located on six copies of the stalk, resulting in 24 copies of TSPs1-4 per CBA120 particle.

The three CBA120 spike proteins are distantly related to each other, and are very similar in their N-terminal regions. TSPs1-3 encoded by orfs210-212 share similar modules, while TSP4 is much

longer (1036 residues) and differs from TSPs1-3. TSPs1-3 N-terminal regions are highly conserved, and are possible base-plate-binding domains similar to the N-terminal of phage Det7 spike, which contains a K1F-like endosialidase segment (Stummeyer et al. 2006). Interestingly, the *Salmonella* phage Det7 was the first observed *Myoviridae* virus with a podovirus-like tailspikes (Walter et al. 2008, 7). The central and C-terminal segments of TSPs1-3 determine the adhesion properties of CBA120 tail. All seven sequenced “Viunlikephages” have N-terminal segments conserved while the C-terminal parts divergent enormously (Fig. 25 D).

TSP1 is a 770-residues-long protein (MW=82kDa) and has a synteny with TSP1 of SFP10 phage, but the target is unknown. The TSP1 crystal structure reveals a polysaccharides-binding β -helix domain. It was shown that TSP1 is able to degrade biofilm polysaccharides, and is possible candidate for biofilm-forming bacteria therapy (C. Chen et al. 2014).

TSP2 (921 residues, MW=98kDa) has a synteny with TSP2 of PhaxI and SFP10 phages, which targets *E.Coli* O157, while TSP3 (627 residues, MW=68kDa) is unique for CBA120. TSP4 (1036 residues, MW=110kDa) has a synteny with TSP4 of PhaxI, which target is putatively ETEC or *Citrobacter freundii* (*E.Coli* H10407) (Adriaenssens et al. 2012). TSP4 N-terminal segment possibly could transmit binding signal to trigger the contraction.

The star-like density connected to the baseplate with a stalk probably consists of all four TSPs (MW=(357)₃kDa), and is ~230 Å long with a mass estimation of ~382 kDa from the extended tail cryo-EM map, and ~230 Å long with a mass estimation of ~373 kDa from contracted tail cryo-EM map. The C-terminal region of a star-like composite density is disordered in the cryo-EM map of extended tail, suggesting it is flexible. In opposite, the contracted tail reconstruction is a fixed structure without presence of any flexibility.

The stalk has a length of 220 Å, with a mass estimation of 101 kDa, and probably corresponds to the largest tailspike, TSP4 (MW=110kDa). In contracted tail reconstruction the stalk is more flatten, ~240 Å long and with estimated mass of ~120 kDa. The N-terminal segment of stalk is attached to the CBA120 conserved baseplate region, probably to the gp6-gp25-gp53 complex (MW=101kDa), resembles that of the T4 baseplate central part (Fig. 27, 28).

In addition, there are 354-Å-long and 50-Å-thick downward-pointing mushroom-shaped fibers visible in cryo-EM map, with disordered bulge at the distal end of the fiber. The tip resembles that of the endosialidase (endoN92, gp143) of the myophage ϕ 92. Estimated mass from cryo-EM map is 118 kDa, however possible structural protein is not annotated in the genome. After the sheath contraction the thin tail fiber and the stalk with the TSPs are aligned side by side (Fig. 28).

4.3.1.4. CBA120 central spike

In addition to the TSPs1-4 CBA120 has a key baseplate protein, 537-residues-long tail-associated

lysozyme protein gp5 (59kDa), which is part of a baseplate hub, and serves as a cell-puncturing device. From cryo-EM map the mass estimation is 67kDa. Tail lysozyme protein is essential to assist phage in penetration of the peptidoglycan layer. T4 cell-puncturing device is formed by trimeric gp5 protein with monomeric gp5.4 protein, which caps the gp5 β -helix tip (Petr G Leiman and Shneider 2012a). This needle-like structure drills the outer cell membrane, and entire hub moves into the periplasmic space. The gp5 β -helix with gp5.4 protein dissociates, and the peptidoglycan layer is then digested by three lysozyme domains of gp5 trimer. The resulting hole opens the path for tail tube to the cytoplasmic membrane, and initiates dsDNA transfer.

CBA120 gp5 protein does not contain identifiable lysozyme, however contains a putative CHAP domain, which acts as a peptidoglycan hydrolase (Kutter et al. 2011). CBA120 lysozyme probably has a bacterial rather than a phage origin. Morphologically it is very different from the T4 gp5 spike. It was shown that the central spike proteins of bacteriophages infecting Gram-negative bacteria split into two groups (Browning et al. 2012). The first group of spikes has a tip part as a separate protein (T4 gp5.4 monomeric cap on gp5 trimer), forming a sharp extension. Most of those spikes are PAAR-repeat proteins (Shneider et al. 2013). Second group of spikes have a tip which is a part of the spike gene itself, and have HxH motif close to the C-terminus. Since CBA120 does not contain a PAAR-repeat protein in the genome, the gp5 spike belongs to the second group and has a HxH motif.

The cryo-EM map of CBA120 extended tail shows the presence of the central spike trimer which does not have a sharp tip, however, there is additional density which follows the 6-fold rotational symmetry of the tail (Fig. 27). The density likely corresponds to the extra domain linked to the CBA120 central spike C-terminus, which is supposed to be sharp by itself. Probably the extra domain is trimeric and flexible, and moves out during the infection.

4.4. Concluding remarks

We described the first detailed three-dimensional organization of the bacteriophage CBA120 from the unique "*Viunalikeyiruses*" genus. It is the first encountered Vil-like phage infecting *E.Coli*, and the most specific to the important food-borne pathogenic strain O157:H7. CBA120 shows substantial conservation of the T4 structural proteins, and carries 38 genes from 39 genes of T4 superfamily core genome, which include 21 structural proteins responsible for capsid and tail formation. Even so, the coliphage CBA120 is the unusual member of the *Myoviridae* family since it combines structural features of the *Myoviridae* phages with long contractile tails and *Podoviridae* phages with short non-contractile tail.

As a consequence, the baseplate and method of adhesion to the host of this T4-like coliphage are very different from those of either T4 or cyanophages. Instead of binding to the host cell primary receptors through various tail fibers, CBA120 utilize *Podoviridae*-like tailspikes, which incorporates

polysaccharide binding sites and endoglycosidase catalytic site into their structures.

CBA120 phage has a tail structure have not been reported in other phages. Current cryo-EM structures of the CBA120 virion allow segmentation and annotation of major structural components, including tail sheath and tube, baseplate components, including wedge and hub with central tail lysozyme attached, and emanating four tailspike proteins TSPs1-4, assembled into a star-like structure and attached to the baseplate wedge via the stalk. In addition, there is possible tail fiber/spike visible in cryo-EM structure with endosialidase-like tip

Such a diverse tailspikes setup resembles an open Swiss army knife (see myophage ϕ 92 for comparison), and results in a more extensive host range than observed before. Probably, CBA120 evolved from O157 *Salmonella* to infect *E.Coli* O157 host, but preserved virion morphology similar to other *Salmonella*-infecting phages from “*Viunalikevirus*” genus.

Gp or orf number	Possible function and position in the virion	Mass of the monomer predicted by sequence, kDa		Mass estimated from cryo-EM map, kDa
gp23	Major capsid protein	48		53(pentamer) 50(hexamer)
-	Pentameric vertex decoration protein	-		70
-	Hexameric capsomer decoration protein	-		25
20	Portal protein	63		
gp18	Sheath	68		69
gp19	Tube	35		-
gp48	Tail tube associated protein, tube initiator	36		37
orf210	Tail spike (TPS) 1	82	357	382 extended tail; 373 contracted tail.
orf211	TPS2	98.2		
orf212	TPS3	68		
orf213	TPS4	110		
orf209	Conserved protein, Vrlc-like	177		176 ext, 176.3 contr.
gp5	Baseplate hub subunit and tail lysozyme	59		55
-	Long, thin fiber with endosialidase-like domain	-		(40) ₃ , trimer?

Table 3. Assignment of CBA120 proteins with cryo-EM and bionformatical data.

4.5. Materials and methods

4.5.1. Phage propagation

CBA120 sample was isolated and purified by Dr. Mikhail Shneider. Host *E.Coli* cells were grown to

an $OD_{600\text{ nm}} \sim 0.6$ before adding phage to an $MOI \sim 0.01$. After lysis completion the chloroform ($CHCl_3$) was added in the ratio 1:1000. Phage sample was centrifuged for 20 min at $8000 \times g$ to remove bacterial debris, followed by 2 h centrifugation at $20\,000 \times g$ to precipitate the phage. Phage debris was resuspended in 1 ml of SM buffer for 24 h. The chloroform was added in the ratio 5:1, resuspended and centrifuged at $10\,000 \times g$. Phage sample was overlaid onto Optiprep density gradient medium and centrifuged at $100\,000 \times g$ overnight. The visible viral band was extracted using a needle syringe and then dialysed in SM buffer.

4.5.2. Data collection

3.5 μl drop of purified CBA120 phage sample with high concentration of 5 mg/ml was applied to a Quantifoil R2/2 200-mesh holey carbon grid (Quantifoil Micro Tools GmbH, Germany). The grid was glow-discharged for 1 min in a home-made glow-discharger. The grid with the sample drop was blotted for 3 sec with filter paper, and plunged into liquid ethane at -180° with a FEI Vitrobot Mark III. Frozen CBA120 phage virions were imaged using a FEI Tecnai F20 electron microscope, operated at an acceleration voltage of 200KV at low-dose regime with an average dose of $\sim 20\text{ e}/\text{\AA}^2$. Totally 1079 cryo-EM micrographs of frozen hydrated CBA120 virions were recorded on Eagle 4K CCD camera. The images were collected both manually and with automated acquisition software SerialEM (D. Mastronarde 2003). 772 micrographs were collected with a nominal pixel size 2.26 \AA , 120 with a nominal pixel size 1.89 \AA , and 187 with a nominal pixel size 1.37 \AA , all with a radiation dose of $\sim 20\text{ e}/\text{\AA}^2$, and a defocus range of 1 - 3 μm (Tab. 4).

4.5.3. Image processing and models reconstruction

For computationally intensive jobs we used Aries cluster (scitas.epfl.ch). It is a 44-nodes (48 cores per node) Linux cluster managed by PBSPro, and currently by SLURM.

4.5.3.1. Pre-processing

Very bright pixels in micrographs have been removed with *speckle_filter.py* utility from EMAN1 package. The contrast transfer function (CTF) of each micrograph was estimated with CTFFIND3, and verified with CTFTILT. The micrographs were binned 4 times to increase the contrast, and reduce the computational cost of CTF estimation. The quality of micrographs was assessed by inspecting the rotationally averaged Fourier power spectra generated by CTFFIND3 for each micrograph. Best micrographs were selected based on the correlation coefficient (CC) between the simulated and experimental Fourier power spectrum in CTFFIND3. It was shown a direct correlation between the visibility of the Thon rings in the rotationally averaged power spectra and the CC from CTFFIND3 (Fotin et al. 2006). Micrographs with the CC higher than 0.2, were selected for further refinement. Full amplitude correction of the best micrographs was performed with *bctf* utility from

Bsoft by application of the Wiener filter based on already determined CTF parameters.

CBA120 part	Symmetry	N_o micrographs	N_o particles	Pixel size, Å	Resolution, Å	Software
Capsid	Icosahedral	872	5122	2.26	14	EMAN, cross-common lines
Capsid	C5	872	5122	2.26	22	SPIDER, FREALIGN, projection matching
Extended tail-baseplate	C6	885	5575	2.26	16	SPIDER, FREALIGN, projection matching
Contracted tail-baseplate	C6	319	1195	2.26	25	SPIDER, FREALIGN, projection matching
Extended tail	Helical	159	5403	1.37	12	SPIDER, IHRSR+

Table 4. Cryo-EM reconstruction of CBA120 virion.

CBA120 virions were extracted from the best micrographs corrected with the Wiener filter. We used divide-and-conquer approach to reconstruct CBA120 virion parts due to the symmetry mismatches in the phage particle. CBA120 baseplates including part of the extended and contracted tails and protruding tail fibers were selected manually from the micrographs with the *batchboxer*/*e2batchboxer.py* utilities from EMAN1/EMAN2. Icosahedral capsids were selected with automated particle selection software ETHAN (Kivioja et al. 2000), and manually with a small fraction of the tail using a *helixboxer* utility from EMAN1. Manual particles selection was necessary to roughly orient all particles before the refinement. Extended tail sheaths were selected with a *helixboxer* and include only central part of the sheath, ~80% of the tail length. All selected particles were picked from the best micrographs using *batchboxer*, *helixbatchboxer*, or *e2helixbatchboxer* from EMAN1/EMAN2 packages. Custom-made c-shell scripts were used, which implemented particles boxing, stacks creations, stacks post-processing including contrast inverting, padding, masking etc. Totally, 1195 best particles of contracted tails, 5575 best particles of extended tails, 5403 best particles of tail sheaths from 80K images, 5122 best particles of capsids were picked and merged into stacks.

To get the initial impression of the data, all particles were subjected to iterative multi-reference classification and alignment using the ML2D (maximum-likelihood multi-reference refinement) or CL2D (standard clustering algorithm aiming at subdividing the original dataset into a given number of subclasses) utilities from the XMIPP package (Scheres et al. 2005). Particles from bad class-averages without well-defined features were excluded. Good class-averages were used to reconstruct an initial 3D model.

4.5.3.2. Capsid reconstruction

CBA120 initial capsid model was build using *randomeuler.py* utility from EMAN1, which assign random orientation (Euler angles) to subset of particles and build a 3D model with icosahedral symmetry enforced.

Future refinement of the initial 3D model was performed using cross-common line searching, which implements in *crossCommonLineSearch.py* routine from EMAN1. Prepared stack of best particles was binned two times, which results in a 4.53 Å pixel size. Initially, search and model refinement was made with relatively small number of the projections (<5), large orientation and center search ranges, 35° and 10 pixels, 1° radial step size for collecting data on common line, and random initial orientations assignment for every new iteration. After each iteration the reconstructed 3D map was low-pass filtered to 20 Å to enforce the low-resolution shape-dependent contrast. After the 3D reconstruction convergence, which was monitored by calculation of the Fourier shell correlation coefficient (FSC) between two consecutive iterations using *proc3d* routine from EMAN, the images were subjected for the final refinement.

Initial particles stack with 2.26 Å pixel size was used. Orientations file from previous refinement was modified to double the particles shifts. Refinement was made with a large number of projections (>15) and using previously assigned Euler angles. Orientation and center search ranges, as well as the radial step size were reduced to 5°, 5 pixels, 0.1 respectively. No low-pass filtering of reconstructed 3D map was made. The 12 Å nominal resolution of the final reconstruction was determined by FSC falloff to 0.143. A negative B-factor of 1000 Å² was applied (Fernández et al. 2008).

CBA120 capsid reconstruction with symmetry relaxation from icosahedral to C5 rotational was calculated with *asymrefine.py* utility from EMAN1. Orientations already found in previous refinement were used, and each image was compared to 12 possible orientations of the 5-fold axis. Best orientation was determined by the highest correlation coefficient. Images with refined orientations were forward to reconstruction step, assuming C5 symmetry, and all procedure was iteratively repeated.

4.5.3.3. Tail reconstruction

CBA120 initial extended tail-baseplate model was build using *startcsym* utility from EMAN1, which generates a preliminary 3D model for objects with Cn or Dn symmetry, C6 in case of CBA120 contractile tail. The *startcsym* will search for particles which are top views (along the Cn axis) and side views of the phage. It then generates a class average for both of these views, then builds a rough 3d model from the 2 class averages. Since we had already aligned and classified particles with XMIPP, we extracted good-looking class-averages, representing side views. Top view was generated as a Gaussian circle with diameter corresponding to CBA120 baseplate. Initial model was built and low-pass filtered to 100 Å.

Similar approach was used to generate initial CBA120 contracted tail-baseplate model. The difference between two cases is the position of the box/cube centers. In case of extended tail the box center was placed in the center of the baseplate. CBA120 contracted virions tends to stick to each other with protruding tail tube tips, resulting in virions overlapping. To remove the unnecessary signal the box center was placed to the contracted tail sheath. The box also includes small fraction of the capsid with neck region, and serves as an anchor during alignment.

We developed algorithm to reconstruct contractile phage tails. For each case (extended and contracted tails) two particles stacks were created. First stack consist of particles extracted from Wiener-filtered micrographs, inverted (white virions on dark background), binned 2 or 4 times, low-pass filtered to 20 Å or lower. Second stack contains same particles extracted from non-corrected raw micrographs, and has black virions on white background.

The first stack was subjected for projection-matching refinement with SPIDER package. We created a SPIDER procedures which include all of the refinement steps. First, initial model was projected according to the proposed symmetry with PJ 3Q operation and stack of reference images was created. AP SH operation determines optimal shifts and rotations to align a stack of particles with a stack of reference images. For each particle, it finds the in-plane Euler rotation angle, and X, Y translational shifts which align the particles with the most-similar reference projection. Next, only particles with high correlation coefficients calculated during alignment were selected and rotated and shifted with RT SF operation. This operation is based on a high-accuracy Fourier spline interpolation, which use a single interpolation of the input particle, without any accumulation of interpolation errors. Best aligned particles were subjected to reconstruction step with BR RP operation, which performs a volume reconstruction from aligned particles using SIRT-WBP (Simultaneous Iterative Reconstruction Technique - Weighted Back Projection) method. SIRT produces the least disturbing artifacts in the presence of angular gaps in projections distribution. Reconstructed model was post-processed by setting pixels below 3σ to the 3D model density average value, which removes surrounding noise while preserving high-resolution features. This set of operations was iteratively repeated, and convergence was judged stable average CC value,

calculated with AP SH operation. The angular search step size was decreased, which resulted in a larger number of reference projections and more precise orientation determination.

After refinement convergence in SPIDER, all of the data was subjected to further refinement in FREALIGN, which determines angular and shift parameters, correct for the CTF and computes a 3D reconstruction. FREALIGN requires raw particles extracted from non-corrected micrographs, which we prepared in second stack, and input text file with orientation assignment and CTF parameters. We used custom-made c-shell script to prepare the input file by merging data from CTFFIND3 and SPIDER text outputs. Output 3D reconstruction from SPIDER was used as an input to FREALIGN. At first iterations we used all particles from the stack, and limited resolution range 800 Å - 30 Å for search and refinement. After convergence, judged by the FREALIGN stable phase residual value (PRES), the resolution range was extended to 20, 15, 9.07 Å (pixel size 4.54 Å, Nyquist limit 9.07 Å). The nominal resolution of the final reconstruction was determined by FSC falloff to 0.143 (Rosenthal and Henderson 2003), 16 Å for extended tail and 26 Å for contracted tail.

4.5.3.4. Tail sheath reconstruction

Best tail sheath particles were extracted from micrographs with nominal pixel size 1.37 Å, using *helixbatchboxer/e2helixbatchboxer.py* routines from EMAN1/EMAN2. Particles were subjected to *helix_automation* SPIDER procedure from IHRSR++ (Iterative Helical Real Space Reconstruction) package (Parent, Sinkovits, et al. 2010). Prepared particles were cut into overlapping segments with *helix_cutstk* and combined into a single stack file, centered with *center_tubes*, sorted by diameter with *get_tube_diameter*. Only diameters from the narrow range were determined, and resulting stack was inverted, padded and low-passed to 20 Å. Stack was subjected for subsequent refinement with IHRSR++.

Initial model was built from extended tail reconstruction by segmenting and extracting the sheath from the map in UCSF Chimera. Initial helical parameters were determined in Chimera by maximizing CC value (max CC=0.95) of the sheath model with itself.

Initial refinement was performed without search for best matching out-of-plane projection. Refinement was continued until the stable values of helical rise and turn were observed. Additional refinement was performed, but the helical rise and turn were fixed, and out-of-plane search was extended to +/-10°. Refinement was iterated with helical symmetry applied until convergence. The convergence was controlled by average CC value generated from IHRSR++ search (AP SH operation in SPIDER). The 12 Å nominal resolution was estimated from the spatial frequency at which the FSC fell to 0.143.

4.5.3.5. Models visualization and interpretation

Models visualization was performed in UCSF Chimera package. Models were segmented using

Segger routine from Chimera, mass were estimated by measuring volumes of segments in Å³, and converting into kDa (Tab. 3).

5. Bacteriophage AR9

5.1. Summary

Jumbo phages have genomes exceeding 200 kb and virus particle sizes of 300 nm and greater. Hundreds of different proteins are involved in the morphogenesis and assembly of a jumbo phage particle. The DNA molecules of many of these giants are heavily modified (e.g. thymidine can be replaced by deoxyuridine) making genome sequencing either arduous or impossible. The jumbo *Myoviridae* phages are the most complex contractile injection systems. Surprisingly, small Gram-positive bacterium *Bacillus subtilis* is a host to several jumbo phages.

Bacteriophage AR9 is a *B. subtilis* jumbo phage with a tail of remarkable complexity. The AR9 tail carries long and thick curly fibers that could be used for attachment and “riding on” the *B. subtilis* flagellum, suggesting that AR9 is a flagella-specific (flagellotropic) phage. Throughout its length, the tail is densely decorated with long and thin fibers that give it an appearance of a bottle brush in the contracted state. The AR9 genome is being sequenced.

We report the first cryo-EM reconstruction of the AR9 virion, which reveals a number of unique morphological components and astonishing flagella-specific mechanism of infection.

5.2. Introduction

Bacteriophages have been classified according to their morphology and genetic composition. *Caudovirales*, or tailed bacteriophages, represent the overwhelming majority of all currently studied bacterial viruses. Among them, *Myoviridae* tailed phages with long contractile tails are the most complex bacterial viruses.

One of the most striking examples of *Myoviridae* phages are the jumbo phages. Jumbo phages have genomes exceeding 200 kb and virus particle sizes of 300 nm and greater (Hendrix 2009). Hundreds of different proteins are involved in the morphogenesis and assembly of a jumbo phage particle. The DNA molecules of many of these giants are heavily modified (e.g. thymidine can be replaced by deoxyuridine) making genome sequencing either arduous or impossible, because all high-throughput sequencing techniques, including Next-generation sequencing, fail. And even if sequenced, most of the predicted ORFs of these phages have no matches in the current sequence databases (Gillis and Mahillon 2014). Among ~400 sequenced dsDNA tailed phages only six have 200+kb genomes (longer than 200kb). No genomes have sequenced for Gram-positive bacteria. The biggest known jumbo phage is *Bacillus megaterium* myophage G (497kb). There is a possibility that jumbo phages are under-sampled, despite their major contribution to microbial

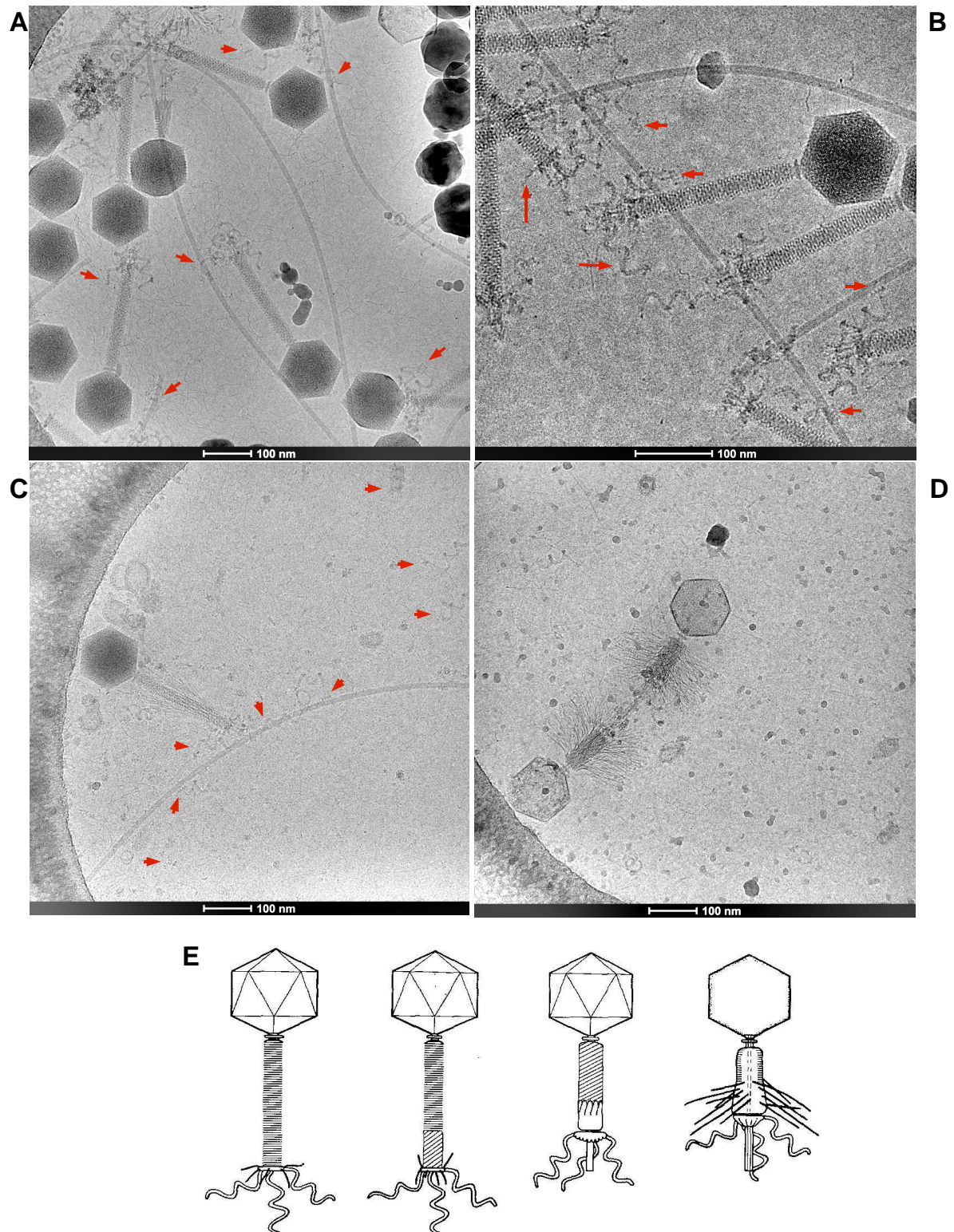


Figure 30: Raw cryo-EM images of *B. subtilis* jumbo flagellotropic phage AR9 in extended (**A-C**), and contracted (**D**) conformations. Curly tail fibers and terminal tailspikes are highlighted by arrows. (**E**) A schematic representation of the different structural forms of phage PBS1, close relative of AR9. **Source:** (Eiserling, 1967, p. 1).

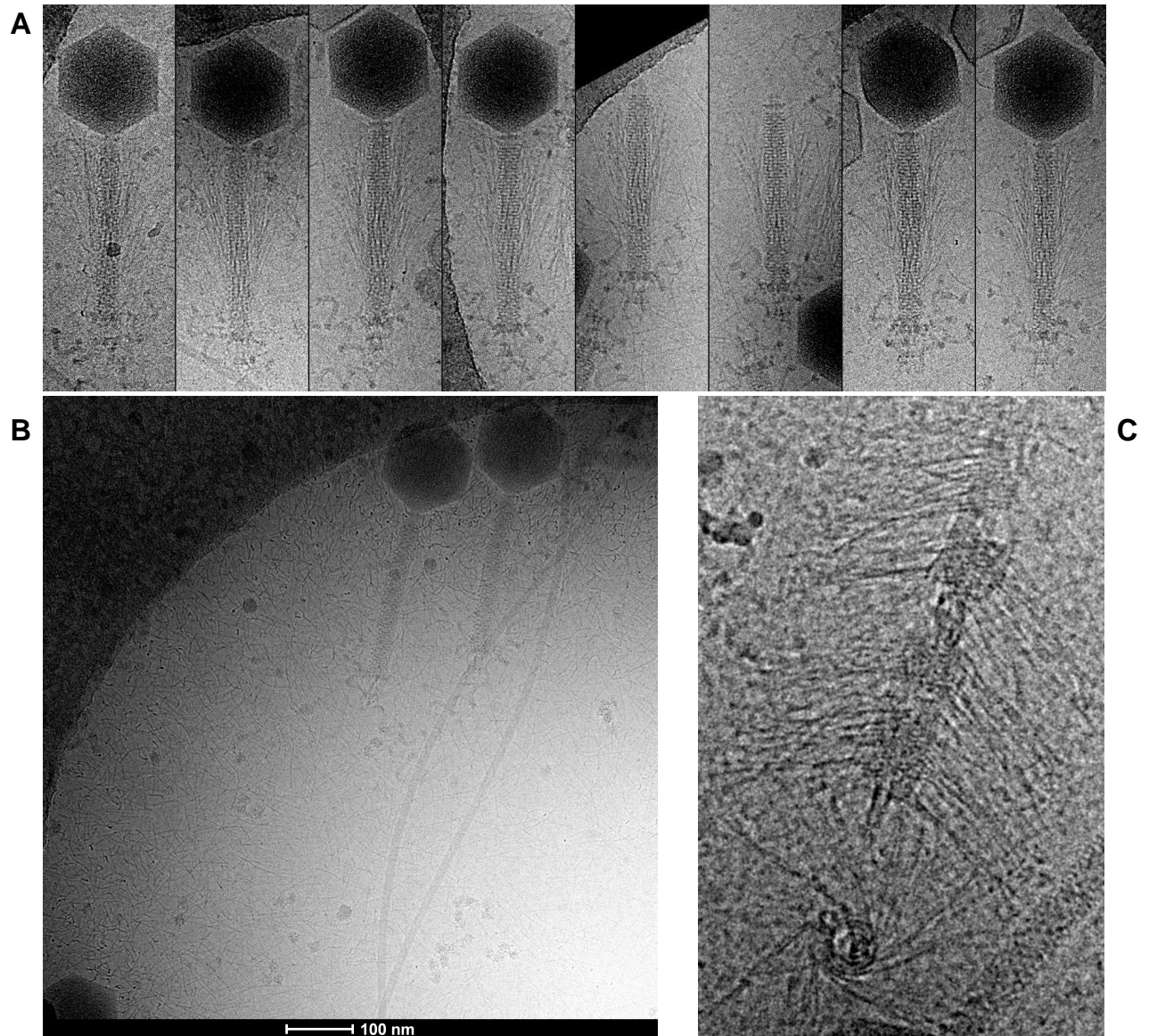


Figure 31: Raw cryo-EM images of the *B. Subtilis* phage AR9 in extended (**A**, **B**), and contracted conformations (**C**). Protruding helical fibers branch off from the sheath subunits (**A**), helical fibers detached from the sheath subunits (**B**). Contracted sheath disassembled into segments with helical fibers protruding downward (**C**).

evolution due to the enormous genome size (Serwer et al. 2007). The jumbo *Myoviridae* phages and the bacterial secretion system type VI (T6SS) are the most complex of contractile injection systems (Petr G. Leiman and Shneider 2012b). A number of jumbo phages infecting *Escherichia coli* and *Pseudomonas aeruginosa* have also been isolated. Surprisingly, small Gram-positive bacterium *Bacillus subtilis* is a host to several jumbo phages (AR9, PBS1).

The initial event of *Myoviridae* phage infection is a tail binding to a specific host cell receptor(s), usually called adsorption. Many of the primary phage receptors have been identified, including membrane proteins, or lipopolysaccharides (LPSs) (Quirk, Sletten, and Hignett 1976; Rakhuba et al. 2010). The majority of specific receptor sites are found on cell wall of both Gram-negative and Gram-positive bacteria. However, it was shown that the cell wall is not the only site for phage adsorption. Another bacterium part known to adsorb phages is the flagellum. Several flagella-specific, or flagellatropic, phages have been identified, and most of them are *Siphoviridae* phages with long non-contractile tails, including phages χ (Chi), ϕ CP34, ϕ 6, ϕ X7, PBP1, SP15, iEPS5 (Appelbaum, Hugo, and Coetzee 1971; Choi et al. 2013; Lovett 1972). Only few *Myoviridae* bacteriophages have been identified, all morphologically similar to phage PBS1, which has been studied most extensively (Belyaeva and Azizbekyan 1968; Eiserling 1967, 1; Joys 1965, 1). However, the structure of the PBS1 phage has remained unknown, and none of the flagellatropic phages have been investigated structurally.

Jumbo *B. subtilis* bacteriophage AR9 was isolated from soil samples in 1967, however, haven't been studied for decades. AR9 phage virion has unique structural features compare to other *B. subtilis* phages. Large genome possibly allows the specification of a large number of structural elements. Despite its similarity to another *Bacillus* phage PBS1, AR9 has different sizes of the virion components and exhibits unique structural elements. Together with PBS1 phage they probably represent a new genus in *Myoviridae* family. AR9 only adsorb to the *B. subtilis* flagella and cannot infect flagella lacking mutants of *B. subtilis*.

Bacillus phage AR9 is temperate phage, which means it has lysogenic activity. In addition, AR9 was found to possess transduction activity, since it mediate the transfer of DNA pieces having a size of about 5 to 8 % of that of the bacterial chromosome (Ferrari et al. 1978).

AR9 consists of 1380-Å-diameter capsid, 2500-Å-long tail and baseplate, which on average are larger than those for bacteriophage phikZ, which is the largest phage studied in detail previously.

The AR9 genome is being sequenced, and is estimated to be 250+kb. Studied phikZ myophage has 1200-Å-diameter capsid encapsulating 280kb genome (Fokine, Kostyuchenko, et al. 2005), thus AR9 genome may be even larger than 300kb.

AR9 tail has a sheath, which was observed in both extended and contracted conformations on cryo-EM micrographs (Fig. 30). According to the current bacteriophages classification, the AR9

phage with long contractile tail belongs to the *Myoviridae* family of tailed phages. However, large dimensions and presence of the unique structural elements not observed in any other studied phages, even in PBS1, indicate that AR9 phage is unique myophage.

Unique structural features include special helical fibers, which probably branch off from each sheath subunit. Similar helical fibers have been observed on phage PBS1 only. The baseplate is small, 300-Å-wide, with at least three types of tail fibers/tailspikes attached. Most exceptive are the twisted flagellum-shaped thick tail fibers with helical pitch of 200 Å, three copies of each are protrudes from the baseplate. These ~1200-Å-long and 84-Å-thick curly fibers were found wrapped around the bacterial flagellum (Fig. 30). In addition, at the distal end of each curly tail fiber there is a 185-Å-wide comb-like density with 60-Å-long 4 teeth, which is usually attached to the flagellum.

The available data suggest that *Bacillus* phage AR9 is PBS1-like jumbo phage, with unique morphology and tail of unusual complexity, which hasn't been observed in any other studied phages. Most likely, it utilizes curly fibers for initial reversible attachment to motile *Bacillus* flagella. Possibly, additional structural elements such as helical fibers or comb-like tailspikes participate in the process of AR9 adsorption to the host cell.

5.3. Results and discussion

5.3.1. Cryo-EM reconstruction of the AR9 virion

Generally, symmetry mismatches between different parts of *Myoviridae* phages complicate the single-particle analysis, and force to use the divide-and-conquer image-processing approach. Different components of the phage virion are extracted and processed separately, assuming the most appropriate symmetry for each component. *Bacillus* phage AR9 components, such as the capsid, the baseplate-tail complex, the tail sheath were reconstructed using divide-and-conquer approach. In addition, reconstruction of *Bacillus* flagellum was performed, and special comb-like tailspike at the distal end of the curly tail fibers was analyzed with 2D classification and averaging. The resolutions of reconstructions were measured with Fourier shell correlation FSC=0.143 criterion, and are 12 Å for the capsid, 20 Å for the baseplate and 15 Å for the tail sheath (Tab. 6).

5.3.1.1. Structure of the AR9 capsid and genomic DNA

AR9 has a gigantic regular icosahedral capsid, 1380 Å in diameter excluding decoration spikes, and 1140 Å along the 2-fold axis. Capsid is so large that individual capsomers are easily seen on the empty capsids (Fig. 30). Capsid geometry characterized by triangulation number $T=27$ ($h=3$, $k=3$). Since $h=k=3$, the capsid is not skewed, or non-chiral. In comparison, T4 phage capsid is skewed with triangulation numbers $T_{\text{end}} = 13/aevo$ and $Q_{5F} = 20$. The AR9 capsid triangulation number is the same as for the 280kb jumbo phage ϕKZ (Fokine, Kostyuchenko, et al. 2005),

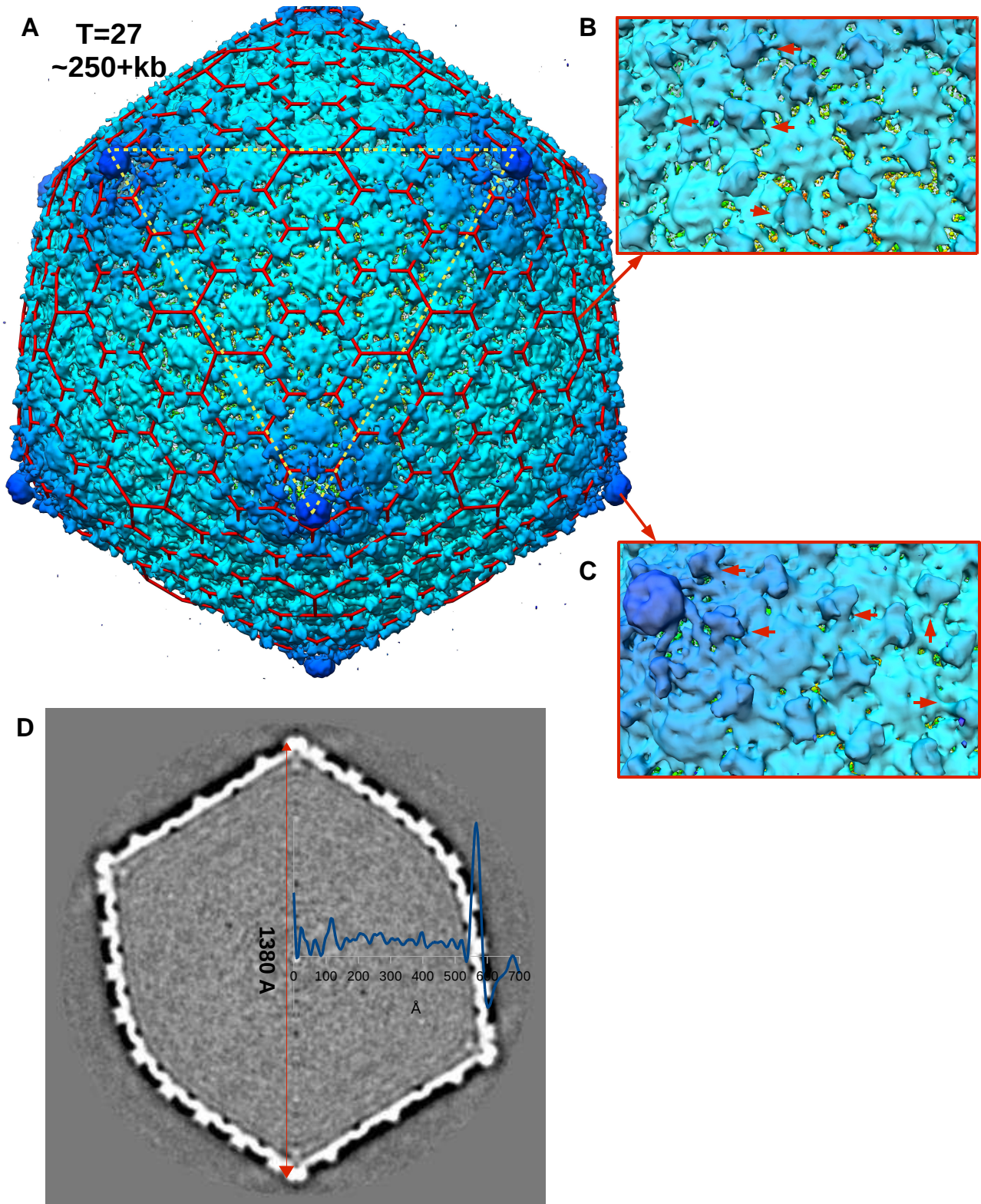


Figure 32: Cryo-EM structure of AR9. **(A)** The structure of AR9 capsid with icosahedral symmetry imposed. Decoration proteins highlighted with red arrows. T=27 icosahedral shell highlighted in red. Face of icosahedron with asymmetric triangle highlighted in yellow. **(B)** Enlarged view of capsid hexameric capsomers **(B)** and pentameric capsomers **(C)** with decoration proteins highlighted. **(D)** Central section through 5-fold axis of capsid cryo-EM structure with radial profile colored in blue.

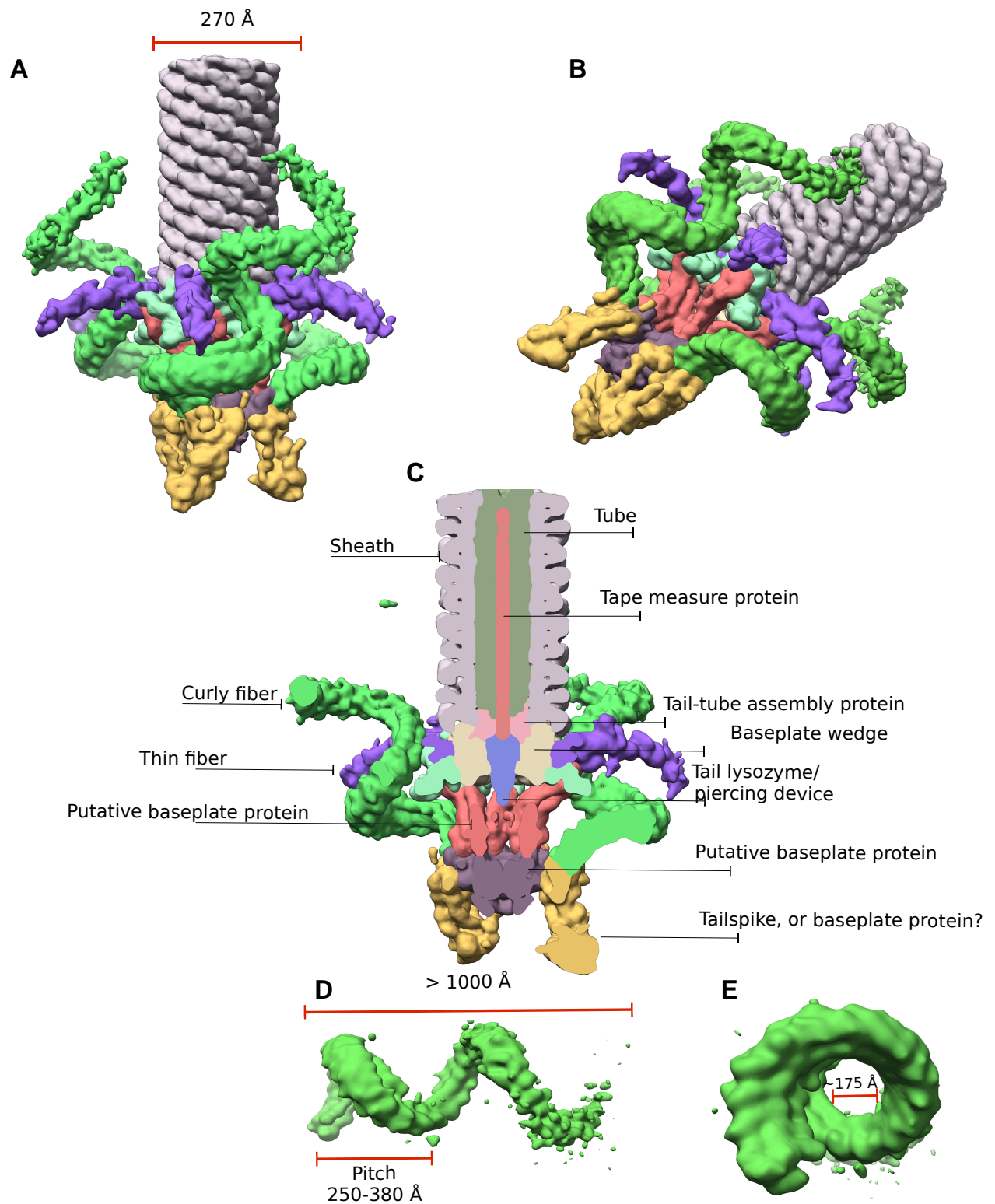


Figure 33: Cryo-EM structure of AR9. Main components are segmented and colored. **(A)**, **(B)** Side and tilted views of AR9 tail structure in extended conformation with 3-fold symmetry imposed. **(C)** Cutaway view of AR9 tail structure in extended conformation with main components highlighted. N-terminal part of thick curly tail fiber, side **(D)** and top views **(E)**.

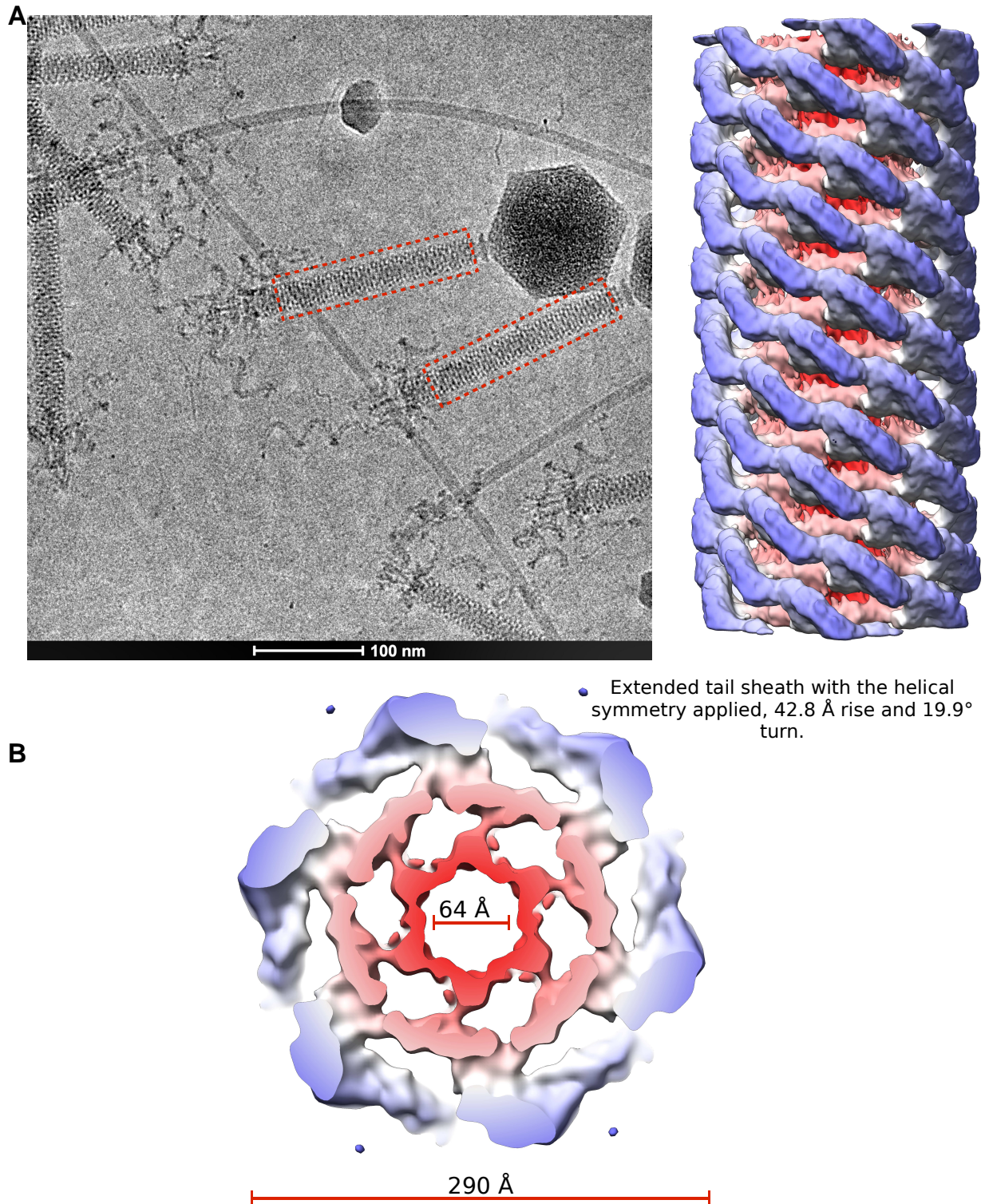


Figure 34: Cryo-EM structure of AR9. **(A)** A raw cryo-EM image of AR9 with reconstructed area highlighted in red. Side view of cryo-EM structure of AR9 tail sheath with helical symmetry imposed. **(B)** Top view of AR9 tail sheath structure in extended conformation with helical symmetry imposed.

however the capsid is ~ 180 Å bigger in diameter, which probably indicates larger major capsid protein (MCP) with extra domain to stabilize the capsid structure. Triangulation number $T=27$ results in 60T or 1620 MCP monomers organized into a lattice of hexamers and pentamers. The distance between adjacent hexamer centers is 146 Å (145 Å in ϕ KZ capsid), and between adjacent pentamers and hexamers is 136 Å (130 Å in ϕ KZ capsid) (Fig. 32).

There are two types of capsid protrusions, or decoration densities, visible on cryo-EM map of the capsid shell, which may contribute to capsid stabilization. The first type is facet-decorating densities, which are located at local 3-fold axes and follows the same icosahedral geometry. These densities probably correspond to the dimeric proteins, which outlines the hexameric capsomer. Each dimeric protein connected to the inner layer of the capsid via two thin cylinders, and probably contributes to capsid stabilization (Fig. 32 B, C). The corresponding volume estimation from cryo-EM map suggests that the mass of this dimeric protein is at least 44kDa. It is a common feature of large phages, and big jumbo phages most likely all utilize other proteins in addition to the MCP for capsid assembly and stabilization.

The second type of decoration densities is an additional protein covering the pentameric capsomer, which mass estimated from cryo-EM map is 96kDa. The density interacts only with the tip of the pentameric capsomer, thus, does not aid in capsid stabilization. However, those protruding densities can have an immunoglobulin (Ig) domains, and serve to attach to host cell wall.

The capsid central section shows the ~ 35 -Å-thick capsid shell, and concentric layers of DNA, unfortunately, with lower contrast than the shell. The rotationally averaged 1D profile reveals at least 13 peaks corresponding to DNA layers, separated by ~ 26 Å. The AR9 genome packaging is quite common in tailed phages, and larger number of layers corresponds to larger genome size.

The AR9 genome is being sequenced, and is an arduous process since it contains hypermodified bases. Current estimation of the genome is 250+kb, however the AR9 genome may be even larger than 300kb.

5.3.1.2. Structure of the AR9 contractile tail

AR9 baseplate-tail complex and tail sheath were reconstructed assuming most appropriate symmetry of each part to 20 Å and 15 Å resolutions.

The AR9 2500-Å-long extended tail is attached to the icosahedral capsid through a 155-Å-long neck with 274-Å-wide collar resembles that of T4 phage (Fig. 33). The tail sheath is 2025-Å-long and has a diameter which varies from 290 Å at the distal part to 270 Å at proximal part. The sheath is formed by 54 hexameric rings, resulting in 324 tail sheath monomers. Same sheath organization was observed for bacteriophage PBS1 (Belyaeva and Azizbekyan 1968) To our

knowledge, the AR9 has the largest phage tail that has been studied structurally.

The 324 copies of the tail sheath protein form a right-hand 6-fold 6-start helix. Helical rise is 42.8 Å and helical turn is 19.9°. Tail sheath protein has a mass of 107 kDa estimated from cryo-EM map (Fig. 34, Tab. 5). Compare with 71kDa T4 or 77kDa phiKZ sheath proteins, the AR9 sheath protein probably has additional tip domain, which interacts with the appendages wrapped around the outside of the sheath. These ~40-Å-thick and ~1000-Å-long appendages, or helical fibers, appear to be unique feature of the AR9 and PBS1 phages. These fibers are spaced along the tail, and protrude upward, or fixed at the neck region, and most likely branch off from each sheath subunit (Belyaeva and Azizbekyan 1968). In the contracted sheath they protrude downward (Fig. 30). These fibers are not visible on every particle, because have been lost during purification, or have aligned with the sheath (Fig. 31).

These additional filaments definitely involved in the process of phage tail adsorption, and may assist the gigantic AR9 virion to stay attached on the host cell membrane until the complete 250+kb genome ejection. Large genome requires longer ejection time, and thus additional filaments may require keeping AR9 virion attached. However, they do not interact with the host cell until the sheath contraction, even after initial binding to flagellum with curly fibers (Fig. 30). AR9 phage virions are able to bind to flagellum even in the absence of the helical fibers on the sheath (Fig. 30). The contracted sheath is 1110-Å-long and 460-Å-wide cylinder, with tail tube protruding ~700 Å below the distal end of the baseplate. Tail sheath contracts uniformly and makes the right-hand 6-fold 6-start helix, however does not interact with the tail tube. As visible in the cryo-EM micrographs, after contraction the sheath disassembles into 330-500-Å-long segments, which run off from the tube (Fig. 31). The tail tube beneath the helical sheath is smooth, and likely consists of the same number subunits as there are sheath subunits, and follows the same helical parameters. The density along the center of the tail tube probably corresponds to the tape-measure protein, which determines the length of the tail (Fig. 33).

5.3.1.3. Flagellum-shaped fibrils and AR9 interaction with host cell

Most exceptive structural features of the AR9 phage are the strongly twisted flagellum-shaped tail fibers made up of subunits visible in cryo-EM micrographs, three copies of which branch off from the baseplate. These ~1200-Å-long and 84-Å-thick curly fibers with helical pitch varying from 250 Å to 380 Å, were found wrapped around the ~170-Å-diameter bacterial flagellum. The diameter of the curly fiber helix is ~175 Å which corresponds well to the *B. subtilis* flagellum diameter (Fig. 35). The AR9 curly fiber has three helical turns, however only two are visible in cryo-EM map while the distal end is flexible and, thus, appears disordered. The flexible tip of the fibers may consist of additional proteins, since it usually observed unfolded in mature virions (Fig. 30, 31). The fiber mass estimated from corresponding cryo-EM map is at least 6MDa. One helical pitch is made up of

~20 subunits, ~150kDa each. Each fiber terminates with 185-Å-wide comb-like tailspike composed of four teeth separated by ~60 Å.

Bacterial flagella are long (~15µm) and ~200-Å-thin helical filaments that protrudes from the cell body of bacteria at various sites. They serve as a rotary motor with hundreds of revolutions per second for bacteria swimming (Yonekura, Maki-Yonekura, and Namba 2003). It was shown that the adsorption of flagellotropic phages, like PBS1, Chi, 7-7-1, initiate by attachment to the bacterial flagellum. Interestingly, that *Caulobacter crescentus* phage phiCbK use the head filament instead of tail fibers for primary adsorption to flagella (Guerrero-Ferreira et al. 2011).

The primary receptors are located on flagella, and phage binds reversibly to flagella (Fig. 35). Successful adsorption requires motile strains of bacteria and functional flagella (Lotz, Acker, and Schmitt 1977; Raimondo, Lundh, and Martinez 1968, 7-7-1). We have observed AR9 virions with curly fibers wrapped around the bacterial flagellum. This initial event of phage adsorption probably initiates with comb-like tailspikes, which have been found attached to the flagella (Fig. 30). The distance between adjacent teeth of the tailspike is ~60 Å, which corresponds to the average flagellum helical rise of ~64 Å (Fig. 35) (Yonekura, Maki-Yonekura, and Namba 2003). After initial attachment the curly fiber wrap around the active flagellum, and allows the attachment of the other curly fibers. Such interaction facilitates concentration of phage particles around the receptor on the bacterial cell surface. After initial reversible attachment phage virions has to move to the base of the flagellum to eject the DNA into the host. The nut-and-bolt model was proposed, in which an adsorbed phage moves toward the cell body due to the rotation of the flagellum, using flagellum helical grooves (Lee et al. 2013). Remarkable observation for PBS1 phage was made, that flagellar movement blocks mechanically by a single attached virion, even without genetic material (Raimondo, Lundh, and Martinez 1968).

The irreversible adsorption occurs near the flagellar base, which was reported for PBS1, Chi and 7-7-1 phages. However, iEPS5 phage seems to directly inject its DNA into the flagellar filament (Choi et al. 2013). It is unlikely that AR9 myophage utilize direct DNA injection into the flagellum, since the contractile sheath and tube are too long to pierce the ~170-Å-diameter flagellum with only ~37-Å-diameter channel precisely.

In addition to the characteristic curly fibers, AR9 baseplate has 6 copies of ~300-Å-long thin tail fibers, protruding radially from the baseplate. Fibers are trimeric proteins, and has a mass of (348)₃ kDa estimated from the cryo-EM map.

The baseplate is narrow, with the diameter of 300 Å, slightly bigger than that of the sheath. It is a complex two-level structure. The first level consists of the hexameric baseplate wedge and, likely, trimeric hub with central sharp-pointed spike, 93-Å-long and mass of 105 kDa estimated from the cryo-EM map. Six copies of thin tail fibers attached to the baseplate wedge. The second level is

built from three subunits, serving as a bask for three copies of curly fibers. Since no central hole is present, there is a possibility that second level changes its conformation after secondary irreversible attachment by long thin fibers, and opens the path for the tube during sheath contraction. Another possible mechanism is that the second baseplate level with curly fibers detaches from the tail complex after the phage virion reached the flagellar base. To complete adsorption and start DNA ejection phage has to detach from the flagellum and bind to secondary receptors on the cell outer membrane. Admissible evidences for proposed mechanism are the cryo-EM images of contracted AR9 virions. All observed contracted tails lack distal baseplate structure with curly fibers, while isolated curly fibers around have been found attached to flagella or in vitreous ice around phage virions (Fig. 30, 31). In addition, possible candidates for irreversible binding, the long thin fibers, are not long enough to fix the virion perpendicular to the outer membrane in case of presence of the distal part of the baseplate (Fig. 33).

After attachment to the secondary receptors, most likely by long thin fibers, the proximal baseplate part changes its conformation and triggers the sheath contraction, similar to other *Myoviridae* phages. Unfolded helical fibers protrude downward to the peptidoglycan layer of the *B. subtilis*, and possibly stabilize gigantic contracted virion during DNA ejection process, which may require cellular energy. The genome ejection process was found to be at least four times longer for PBS1 phage than for smaller flagellotropic phage Chi (Meynell 1961).

5.4. Concluding remarks

The first structural studies of the flagellotropic *Bacillus subtilis* bacteriophage AR9 has been made. AR9 is a jumbo phage with 250+kb genome, and complex morphology resembles that of another *B. subtilis* bacteriophage PBS1. However, AR9 components have different sizes, and, in addition to structural components not detected in PBS1 phage make AR9 unique among the whole *Myoviridae* family.

The method of adhesion to the host appears to be complex and requires motile bacterial strain with active flagella. Initial reversible tail adsorption to the flagella primary receptors occur with twisted tail fibers made up of subunits. The distal end of each curly fiber is decorated with the comb-like tailspike responsible for the initial attachment to the primary receptors. The nut-and-bolt model propose explanation of the further phage virion movement to the flagellar base, where AR9 tail irreversibly attaches to the secondary receptors on the thick peptidoglycan (PG) layer of *B. subtilis*. Complex conformational changes of the baseplate occur for successful virion fixation on the PG layer, sheath contraction, membrane piercing, and dsDNA genome ejection into *B. subtilis* cytoplasm. The major structural components has been annotated in the baseplate-tail cryo-EM map.

Possible function and position in the virion	Mass estimated from cryo-EM map, kDa
Curly fiber	>6000
Curly fiber subunit	150
Thin tail fiber	348
Central spike	35
Sheath	107

Table 5. AR9 tail proteins assignment.

5.5. Materials and methods

5.5.1. Phage AR9 propagation

AR9 sample was isolated and purified by Dr. Mikhail Shneider. Host *B. subtilis* cells infected with phage were grown to an OD_{600 nm} ~0.6 on a shaker at 220rpm. Cell were lysed in 2 hours. Phage lysate was resuspended for 30 min in MgSO₄ to 20mM concentration and DNase was added. Cell debris was centrifuged for 20 min at 5 000 × g. PEG 8K was added to the lysate to final concentration of 10% and NaCl to 0.5M, and lysate was kept overnight at 4°C. Phage lysate was centrifuged for 20 min at 8 000 × g, and precipitate was resuspended in SM buffer with DNase addition. The chloroform was added in the ratio 5:1, resuspended and centrifuged at 10 000 × g. Supernatant was overlaid The chloroform was added in the ratio 5:1, resuspended and centrifuged at 10 000 × g overnight. The visible viral band was extracted using a needle syringe and then dialysed in SM buffer.

5.5.2. Data collection

3 µl of purified high-concentrated AR9 sample was applied to glow discharged Quantifoil R2/2 200-mesh holey carbon grid. The grid was plunge freeze in liquid ethane using a FEI Vitrobot Mark III. Phage particles were imaged using a 200 kV FEI Tecnai F20 electron microscope at low-dose regime. 1224 cryo-EM micrographs were recorded both manually on Eagle 4K CCD camera. 1011 micrographs were collected with a nominal pixel size 2.26 Å, 213 micrographs with a nominal pixel size 1.37 Å. The defocus range was set to 1 - 3 µm (Tab. 6).

AR9 part	Symmetry	N _e micrographs	N _e particles	Pixel size, Å	Resolution, Å	Software
Capsid	Icosahedral	573	1922	2.26	14	EMAN, cross-common lines
Extended tail-baseplate	C6	607	2530	2.26	20	SPIDER, FREALIGN, projection matching
Extended tail	Helical	496	10146 segments	1.37	15	SPIDER, IHRSR++
Comb-like spike	No symmetry	17	171	2.26	-	ML2D, XMIPP
<i>B. subtilis</i> flagellum	Helical	22	389	2.26	-	ML2D, XMIPP

Table 6. Single-particle reconstruction of AR9 virion.

5.5.3. Image processing and models reconstruction

For computationally intensive jobs we used 44-nodes (48 cores per node) Linux cluster Aries (scitas.epfl.ch).

5.5.3.1. Pre-processing

Images were preprocessed, as described for CBA120 phage. CTF parameters were estimated with CTFFIND3 for each micrograph, which were further corrected by application of Wiener filter with Bsoft. Best micrographs were selected for further processing based on the CTF parameter estimation correlation coefficient between the simulated and experimental Fourier power spectra.

Divide-and-conquer approach was used to reconstruct AR9 virion. This approach appears to be general for three-dimensional reconstruction of large macromolecular complexes, such as Myoviridae or even larger *jumbo* bacteriophages.

AR9 virions were extracted from the best CTF corrected micrographs with the *batchboxer/e2batchboxer.py* or *helixbatchboxer/e2.helixbatchboxer.py* utilities from EMAN1/EMAN2, automated particle selection software ETHAN. Totally, 2530 best particles of extended tails with 2.26 Å nominal pixel size, 1922 best particles of capsids with nominal pixel size 2.26 Å, 10146 segments of helical sheath with nominal pixel size 1.37 Å, 171 particles of the distal comb-like tailspikes, and 389 particles of the distal *B. subtilis* flagellum with nominal pixel size 2.26 Å were picked and merged into stacks.

All particles were subjected to iterative multi-reference classification and alignment using the ML2D utility from XMIPP. Particles from bad class-averages without well-defined features were excluded. Good class-averages were used to reconstruct an initial 3D model.

5.5.3.2. Capsid reconstruction

AR9 initial capsid model was built as described for CBA120 capsid. Future refinement of the initial 3D model was performed using cross-common line searching using *crossCommonLineSearch.py* routine from EMAN1. The 14 Å nominal resolution of the final reconstruction was determined by FSC falloff to 0.143.

5.5.3.3. Tail reconstruction

AR9 initial extended tail-baseplate model was build as described for CBA120 tail. Two stacks of particles were created. First stack particles extracted from Wiener-filtered micrographs, inverted (white virions on dark background), binned 2 or 4 times, low-pass filtered to 20 Å or lower. Second stack contains same particles extracted from non-corrected raw micrographs, and has black virions on white background.

The first stack was subjected for projection-matching refinement with SPIDER package. After refinement convergence in SPIDER, all of the data was subjected to further refinement in FREALIGN, which determines angular and shift parameters, correct for the CTF and computes a 3D reconstruction. FREALIGN requires raw particles extracted from non-corrected micrographs, which we prepared in second stack, and input text file with orientation assignment and CTF parameters. We used custom-made c-shell script to prepare the input file by merging data from CTFFIND3 and SPIDER text outputs. Output 3D reconstruction from SPIDER was used as an input to FREALIGN. At first iterations we used all particles from the stack, and limited resolution range 800 Å - 30 Å for search and refinement. After convergence, judged by the FREALIGN stable phase residual value (PRES), the resolution range was extended to 20, 15, 9.07 Å (pixel size 4.54 Å, Nyquist limit 9.07 Å). The 20 Å resolution of the final reconstruction was determined by FSC falloff to 0.143.

5.5.3.4. Tail sheath reconstruction

Best tail sheath particles were extracted from micrographs with nominal pixel size 1.37 Å, using *helixbatchboxer/e2helixbatchboxer.py* routines from EMAN1/EMAN2. Particles were subjected to *helix_automation* SPIDER procedure from IHRSR++ package. Prepared particles were cut into overlapping segments with *helix_cutstk* and combined into a single stack file, centered with *center_tubes*, sorted by diameter with *get_tube_diameter*. Only diameters from the narrow range were determined, and resulting stack was inverted, padded and low-passed to 20 Å. Stack was subjected for subsequent refinement with IHRSR++. Initial model was built from extended tail reconstruction by segmenting and extracting the sheath from the map in UCSF Chimera [Chimera].

Initial helical parameters were determined in Chimera by maximizing CC value (max CC=0.95) of the sheath model with itself. Refinement was iterated with helical symmetry applied until convergence. The convergence was controlled by average CC value generated from IHRSR++ search (AP SH operation in SPIDER). The 15 Å resolution was estimated from the spatial frequency at which the FSC fell to 0.143.

5.5.3.5. Models visualization and interpretation

Models visualization was performed in UCSF Chimera package. Models were segmented using *Segger* routine from Chimera, mass were estimated by measuring volumes of segments in Å³, and converting into kDa (Tab. 5).

6. Bacteriophages T4 and RB43

6.1. Summary

Long-tailed bacteriophages from *Myoviridae* family are of recently recognized class of “contractile injection systems”, and share a common evolutionary origin and some functional aspect with other large macromolecular machines, such as the *Serratia entomophila* antifeeding prophage, the *Photorhabdus* virulence cassette, R-type pyocins, and the bacterial type VI secretion system (T6SS) of Gram-negative bacteria.

A contractile injection system of typical *Myoviridae* is a macromolecular machine for infecting bacteria. *Myoviridae* phages, as well as T6SS have been proposed to use these central spikes, which share a common modular structure, to pierce the host cell membrane during infection.

Bacteriophage T4 with a long contractile tail is a readily recognizable member of the *Myoviridae* family. Major portion of information of proteins involved in host attachment for all contractile systems comes from studies of the T4 tail.

The T4 central spike, or cell-puncturing device is formed by gene product gp27, gp5 and recently identified gp5.4 “tip” protein, which is responsible for the first interaction with the host cell membrane.

Using electron cryo-microscopy we highlighted the differences in the structural organization of wild type T4 phage and mutant T4 phage lacking gp5.4 tip protein, In addition, T4 cell-puncturing device was compared with that of T4-like pseudo-T-even phage RB43.

6.2. Introduction

Contractile tails of *Myoviridae* phages generally consists of a baseplate, long non-contractile central tail tube (core), surrounded by external contractile sheath. The baseplate carrier tail fibers and/or tailspikes with host cell-binding proteins. Tail fibers, which emanate from the baseplate, are the primary determinants of host cell specificity, and responsible for the external capsular sugar (if present) both recognition and local digestion. In addition, T6SS and *Myoviridae* phages have been proposed to use a special protein, called cell-puncturing tip or device, to create an opening in the host cell membrane during infection. The T4 hexameric baseplate is composed of at least 16 different proteins. The baseplate core is the trimeric central hub, which contains proteins gp5, and gp27. The gp27-like proteins also present in non-contractile *Siphoviridae* tails (Sciara et al. 2010). The gp27 protein ends with a trimeric central spike protein, gp5, which has a modular structure where the N-terminal domain contains an oligosaccharide/oligonucleotide-binding (OB) fold, the

central domain has a lysozyme-like fold, and the C-terminal domain is a long intertwined β -helical structure. The T4 gp5 spike's tip β -helix sharpened by recently identified gp5.4 protein, which pierces the *E.Coli* host cell membrane (Shneider et al. 2013). Gp5.4 protein is a monomeric protein, which belongs to the PAAR (Proline-Alanine-Alanine- α Rginine) repeat superfamily. Its polypeptide chain is folded into three loops of different length that slightly wrap around each other (Fig. 36).

This needle-like structure drills the outer cell membrane, and entire hub moves into the periplasmic space. The gp5 β -helix with gp5.4 protein dissociates, and the peptidoglycan layer is then digested by three lysozyme domains of gp5 trimer (Petr G Leiman et al. 2010). The resulting hole opens the path for tail tube to the cytoplasmic membrane, and initiate dsDNA transfer.

It was shown, that the cell-puncturing devices are a conserved feature of contractile tail-like systems (Petr G. Leiman and Shneider 2012b). Orthologs of the T4 cell-puncturing device is found in the T6SS, which is responsible for toxic effector (VgrG protein) translocation across the membrane of predatory bacteria directly into the cytoplasm of the eukaryotic prey cells (Pukatzki et al. 2007). The central cell-puncturing devices of myophages P2 and ϕ 92, gpV and gp138, form the membrane-piercing components similar to that of gp5 of T4, but without the lysozyme domain and the monomeric tip. Instead, they have a sharp C-terminal extension coordinated by an iron ion (Browning et al. 2012). At recent, \sim 170 T4-like phages have been isolated on a wide range of bacterial hosts that grow in diverse environments (H. W. Ackermann and Krisch 1997; Tétart et al. 2001). T4 phage is probably the best-understood virulent phage, while little is known about all the other T4-type phages (Desplats et al. 2002).

RB43 is a virulent bacteriophage that infects *E.Coli* host strain B^E. It has an indistinguishable from T4 morphology, but the genome have blocks of sequence that differentiate it from other known T4-like phages. Recently, such T4-type phages have been divided into the Pseudo T-even subgroup in T4 supefamily (Monod et al. 1997). RB43 coliphage has 180 kb genome, which brings him into the top 100 of largest known viral genomes. No genomes larger than that of RB43 have been reported among the *E.Coli*-specific T4-type phages. RB43 genome contains only 112 T4-like ORFs from 292 ORFs genome (Petrov et al. 2006). At recent, little is known about these most distant members of the T4 phage family. Study of the RB43 will facilitate the structural analysis of these subgroup of the T4-type phages.

Using electron cryo-microscopy we show that coliphage RB43, distant member of T4-type phages, has structural difference in cell-piercing device responsible for the first interaction with the host cell membrane. Cryo-EM reconstruction of the mutant T4 phage reveals the absence of the cell-piercing device tip protein gp5.4, which results in the mutant's reduced fitness.

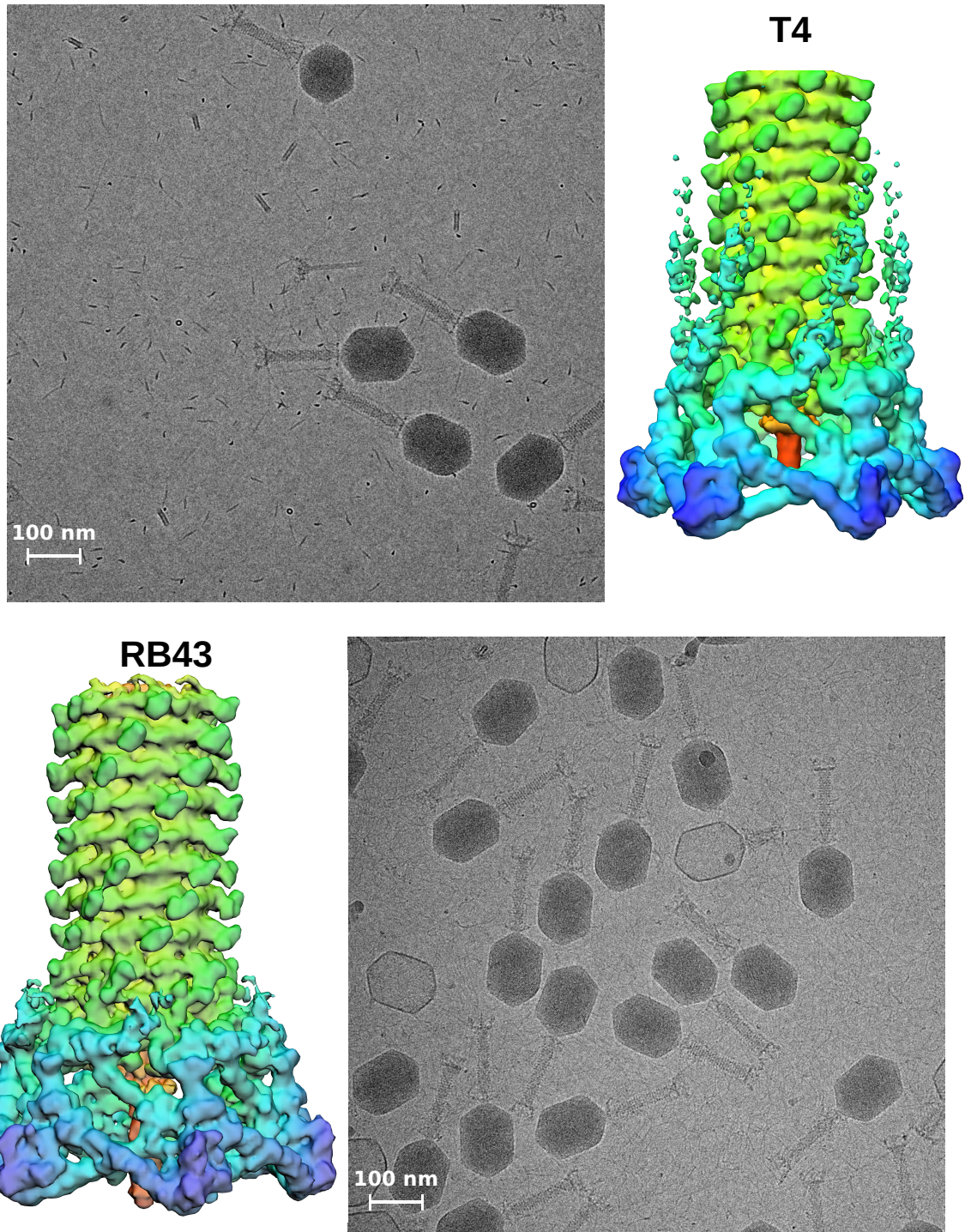


Figure 36: Cryo-EM structures of T4 and RB43. **(Top)** A raw cryo-EM image of gp5.4am mutant T4, and cryo-EM structure of gp5.4am T4 baseplate with tail segment with 6-fold symmetry imposed. **(Bottom)** A raw cryo-EM image of wt RB43 from Pseudo T-even subgroup in T4-superfamily, and cryo-EM structure of RB43 baseplate with tail segment with 6-fold symmetry imposed.

6.3. Results and discussion

6.3.1. Cryo-EM reconstruction T4 and RB43 baseplate/tail complexes

The structures of the distal part of the T4 and RB43 tail, containing the baseplate and part of the sheath were reconstructed using divide-and-conquer approach (Fig. 36). The baseplate corresponding densities have been extracted from the cryo-EM micrographs and processed assuming sixfold rotational symmetry. The resolutions of reconstructions are 15 Å and 17 Å for T4 and the RB43 baseplates, FSC=0.143 criterion (Tab. 7).

6.3.1.1. Structure of the mutant T4 and RB43 contractile tails

T4 amber mutant with mutation in the gp5.4 gene (gp5.4am) has been used for imaging and cryo-EM reconstruction. The resulting structure of the gp5.4am T4 baseplate with the fraction of extended sheath is essentially the same as previously reported (Kostyuchenko et al. 2005). Cryo-EM map shows expected T4 baseplate structure, and include 240-Å-diameter sheath surrounding the 90-Å-diameter tube, the baseplate with an external diameter of 517 Å and six long tail fibers (LTFs) and six short tail fibers (STFs) attached.

T4 infection is started with the reversible attachment of the long tail fibers to the lipopolysaccharide receptors. The STFs are unfolded and anchor the baseplate irreversibly on the cell surface. Tail fibers rearrangement initiates the baseplate transition from the hexagonal to a star-like shape, which triggers the sheath contraction. The rigid non-contractile tube extends and rotates with contracting sheath, and breaches the outer cell membrane with the monomeric gp5.4 tip protein. The entire hub (gp27)₃-(gp5)₃-gp5.4 moves into the periplasmic space and the inter-membrane peptidoglycan (PG) layer is digested by the lysozyme domains of gp5 trimer. The dsDNA genome is ejected through the tail tube into the *E.Coli* cytoplasm. However, little is known about the interaction of the tube with the cytoplasmic (inner) membrane.

The cell-piercing device-corresponding density in the cryo-EM map is conclusive about gp5.4 being the tip protein of T4, since there is no density corresponding to gp5.4 protein (Fig. 37, 38). Gene 5.4 amber mutant of T4 phage is unable to form complete hubs (gp27)₃-(gp5)₃-gp5.4, which results in reduced phage fitness (unpublished data). The difference map between wild type (wt) and gp5.4am T4 baseplate cryo-EM reconstructions shows the density 49-Å-long, with mass of 10.3 kDa estimated from the cryo-EM map, that perfectly corresponds to the gp5.4 protein parameters (MW=10kDa, 51-Å-long). Cryo-EM micrographs of gp5.4am T4 phage show the presence of the virions with an unusual morphology (Fig. 38). Roughly 3% of all observed virions have capsids with two tails, which originate from non-apex vertices of the capsid. Remarkably, that in addition to

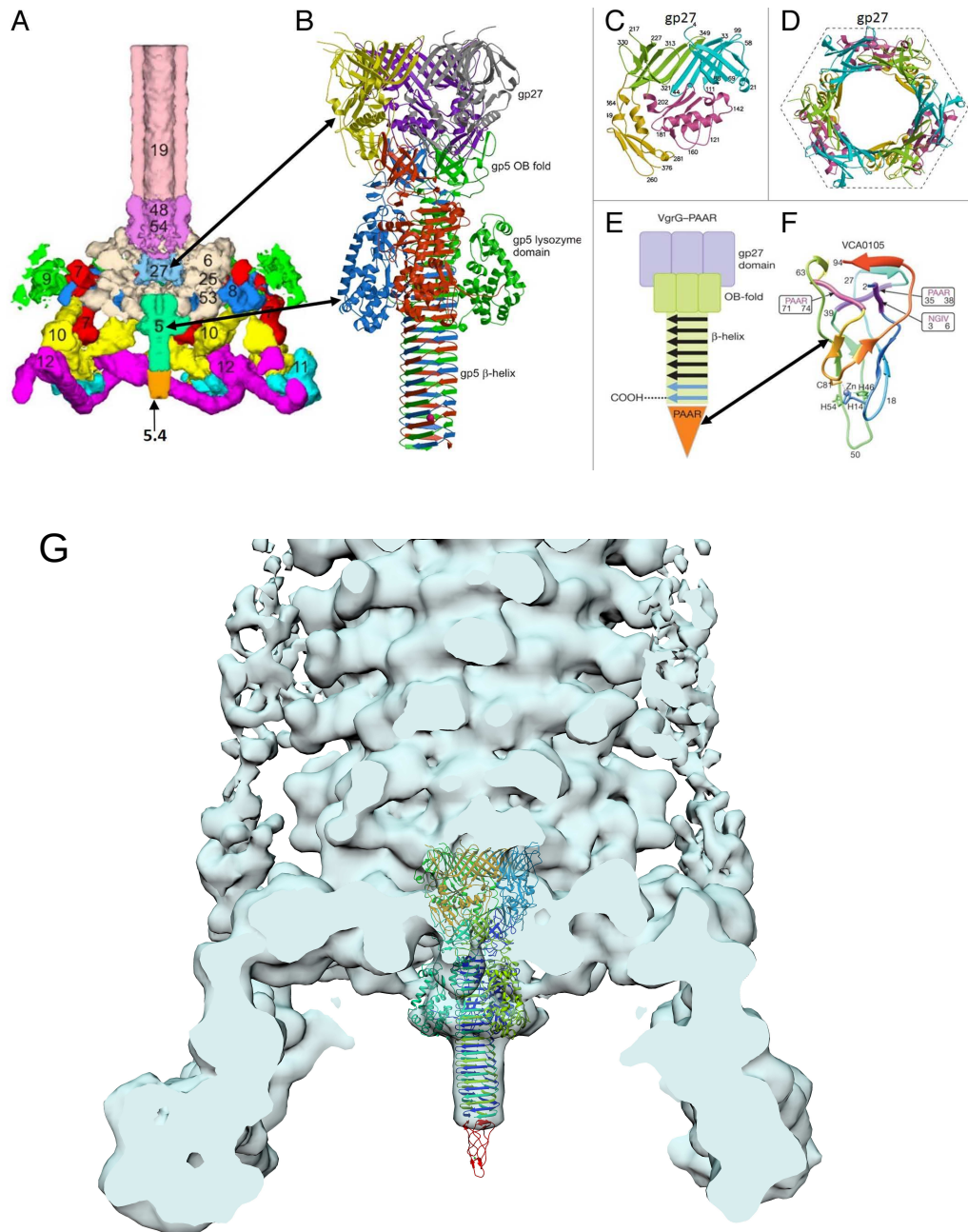


Figure 37: Cryo-EM structure of T4. **(A)** Cross-section view of the T4 baseplate. Constituent proteins are shown in different colors and are identified with their corresponding gene names. **(B)** Structure of the gp5-gp27 trimer is shown as a ribbon diagram in which each chain is shown in a different color. **(C)** Domains of gp27. The 2 homologous domains are colored in light green and cyan. These domains are similar to the tail-tube proteins. **(D)** The pseudohexameric ring formed by the tail-tube like domains of the gp27 trimer. **(E)** Schematic representation of a spike from the bacterial secretion system VI. The spike consists of the VgrG and PAAR-repeat proteins (F). **(A-F)** **Source:** (Leiman et al., 2010; Shneider et al., 2013). **(G)** Cryo-EM structure of the gp5.4am mutant T4 baseplate with fitted (gp27)₃-(gp5)₃-gp5.4 proteins structure.

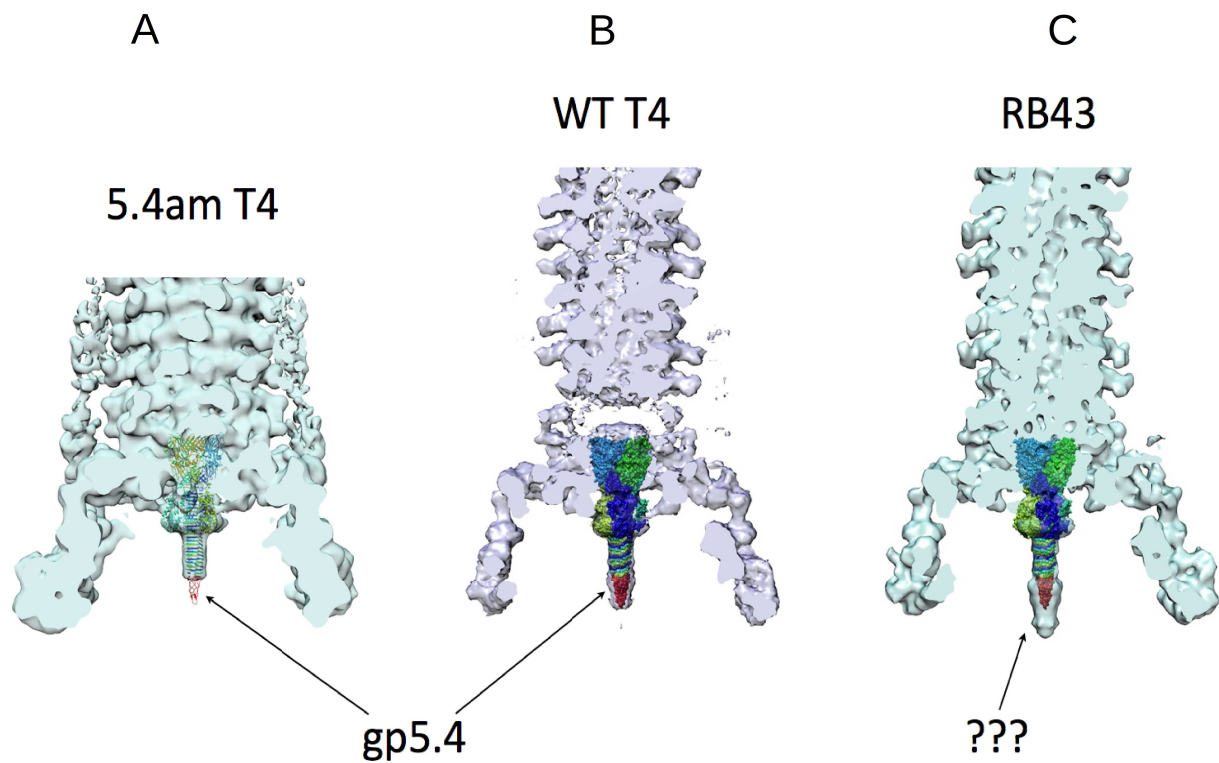
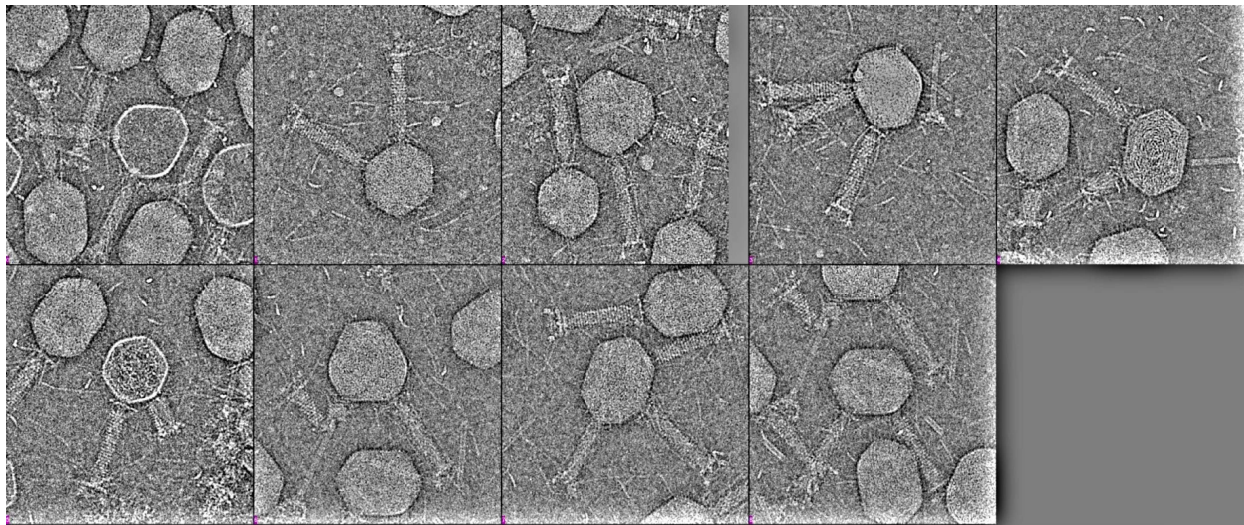


Figure 38: Cryo-EM structure of T4 and RB43. (**Top**) Aberrant gp5.4am mutant T4 virions with multiple tails and odd-shaped capsids. (**A-C**) Cryo-EM structures of tail - base plate complexes with fitted crystal structure of complete T4 spike gp5-gp27-gp5.4, (**A**) cryo-EM structure of mutant T4 phage without gp5.4 tip protein (central spike tip), (**B**) cryo-EM structure of wild type T4 phage (V. Kostyuchenko, *Nat Struct Mol Biol* 2005), (**C**) cryo-EM structure of wild type RB43 phage (Pseudo T-even phage, T4 superfamily) with additional protein domain on the tip of central tail spike.

protein not at the apex vertex. Such prolate capsid morphology has never been observed for T4 phage. It is unlikely that gene 5.4 amber mutation can cause such divergent morphology. Probably, there is a mutation in regulatory genes or an assembly chaperons.

Wild type RB43 phage infecting *E.Coli* host strain B^E has been used for imaging and cryo-EM reconstruction. The resulting cryo-EM map is essentially the same as for the T4 phage except for the cell-piercing device (Fig. 36, 38). Using the same reasoning as for T4 phage we hypothesized that RB43 central spike is formed by a protein homologous to T4 gp5, with N-terminal domain as OB-fold and C-terminal domain is β -helix. The cryo-EM corresponding density shows similarity in (gp27)₃-(gp5) segment, however the tip density is different. RB43 gp5.4 homologous protein is probably is a spike-shaped trimer ~75-Å-long and 44-Å-diameter in the widest part. The mass of the corresponding cryo-EM density is estimated as 38kDa. No other significant structural differences from T4 baseplate have been observed.

Despite the fact that RB43 genome contains only 112 T4-like ORFs from 292 ORFs and has unique organization of the clusters of replication-related genes in the genome, the main difference of this distal T4-type coliphage lays in genes for phage DNA metabolism, but not in structural genes.

6.4. Concluding remarks

In this report we described the structures of *E.Coli* bacteriophages, mutant T4 phage and T4-type phage RB43. Using electron cryo-microscopy we highlighted the differences in the structural organization of wild type T4 phage and mutant T4 phage lacking gp5.4 tip protein. In addition, T4 cell-puncturing device was compared with that of T4-like pseudo-T-even phage RB43.

Cryo-EM reconstruction of the mutant T4 phage is conclusive about gp5.4 being the tip protein of T4. In comparison, the distant member of the T4-superfamily phage RB43 has different density corresponding to the tip protein, which is ~4 times bigger than that from wild type T4 phage, and probably trimeric. RB43 thus interact with the host cell surface in a specific way, different from T4.

6.5. Materials and methods

6.5.1. Phages T4/RB43 propagation

T4 and RB43 samples were isolated and purified by Dr. Mikhail Shneider. Host *E.Coli* cells infected with phage were grown to an OD_{600 nm} ~0.6 before adding phage to an MOI ~0.01. After lysis completion the chloroform (CHCl₃) was added in the ratio 1:1000. Phage sample was centrifuged for 20 min at 5000 × g to remove bacterial debris. PEG 8K was added to the lysate to final concentration of 10% and NaCl to 0.5M, and lysate was kept overnight at 4°C. Phage lysate was centrifuged for 20 min at 8 000 × g, and precipitate was resuspended in SM buffer with DNase

addition. The chloroform was added in the ratio 5:1, resuspended and centrifuged at $10\,000 \times g$. Phage sample was precipitated by centrifugation for 2 h at $20\,000 \times g$ and resuspended in 1 ml of SM buffer overnight. Phage sample was overlaid onto Optiprep density gradient medium and centrifuged at $100\,000 \times g$ overnight. The visible viral band was extracted using a needle syringe and then dialysed in SM buffer.

6.5.2. Data collection

3.3 μ l drop of purified high-concentrated T4/RB43 sample was applied to 30s glow discharged Quantifoil R1.2/1.3 300-mesh holey carbon grid. The grid was plunge frozen in liquid ethane using FEI Vitrobot Mark III. Phage particles were imaged using a 200KV FEI Tecnai F20 electron microscope with an average dose of $\sim 16 \text{ e}/\text{\AA}^2$. 215 cryo-EM micrographs of T4 phage, and 154 cryo-EM micrographs of RB43 phage were recorded manually on Eagle 4K CCD camera with a nominal pixel size 2.26 \AA (Tab. 7).

T4/RB43	Symmetry	N _o micrographs	N _o particles	Pixel size, \AA	Resolution, \AA	Software
T4 extended tail	C6	215	1819	2.26	15	SPIDER, FREALIGN, projection matching
RB43 extended tail	C6	154	1678	2.26	17	SPIDER, FREALIGN, projection matching

Table 7. Single-particle reconstruction of T4 and RB43 baseplates.

6.5.3. Image processing and models reconstruction

For computationally intensive jobs we used 44-nodes (48 cores per node) Linux cluster Aries (scitas.epfl.ch).

6.5.4. Pre-processing

Images were preprocessed, as described for CBA120 phage. CTF parameters were estimated with CTFFIND3 for each micrograph, which were further corrected by application of Wiener filter with Bsoft. Best micrographs were selected for further processing based on the CTF parameter estimation correlation coefficient between the simulated and experimental Fourier power spectra.

Divide-and-conquer approach was used to reconstruct T4/RB43 baseplates. This approach appears to be general for three-dimensional reconstruction of large macromolecular complexes, such as

Myoviridae or even larger *jumbo* bacteriophages. T4/RB43 baseplates were extracted from the best CTF corrected micrographs with the *batchboxer* utility from EMAN1 package. Totally, 1819 best particles of T4 extended tails, and 1678 best particles of RB43 extended tails were picked and merged into stacks. All particles were subjected to iterative multi-reference classification and alignment using the ML2D utility from XMIPP. Particles from bad class-averages without well-defined features were excluded. Good class-averages were used to reconstruct an initial 3D model.

6.5.5. Tail reconstruction

T4/RB43 initial extended tail-baseplate models were build from previously reported T4 baseplate reconstruction (EMDB-1126). Model was re-centered, clipped and low-pass filtered to 100 Å.

Two stacks of particles were created. First stack consists of particles extracted from Wiener-filtered micrographs, inverted (white virions on dark background), binned 4 times, low-pass filtered to 20 Å or lower. Second stack contains same particles extracted from non-corrected raw micrographs, and has black virions on white background. The first stack was subjected for projection-matching refinement with SPIDER package. After refinement convergence in SPIDER, all of the data was subjected to further refinement in FREALIGN, which determines angular and shift parameters, correct for the CTF and computes a 3D reconstruction. FREALIGN requires raw particles extracted from non-corrected micrographs, which we prepared in second stack, and input text file with orientation assignment and CTF parameters. We used custom-made c-shell script to prepare the input file by merging data from CTFFIND3 and SPIDER text outputs.

Output 3D reconstruction from SPIDER was used as an input to FREALIGN. At first iterations we used all particles from the stack, and limited resolution range 800 Å - 30 Å for search and refinement. After convergence, judged by the FREALIGN stable phase residual value (PRES), the resolution range was extended to 20, 15, 9.07 Å (pixel size 4.54 Å, Nyquist limit 9.07 Å). The 17 Å and 15 Å resolution of the final reconstructions were estimated from the spatial frequency at which the FSC fell to 0.143. Models visualization and rigid-body fitting of crystal structures were performed using UCSF Chimera package.

7. Conclusions and perspectives

We studied a number of bacterial viruses, including so-called “jumbo” phages, infecting important pathogens. They were isolated and characterized by means of single-particle electron cryo-microscopy, electron tomography, biochemical and bioinformatical tools. Viral components responsible for the cell attachment and genome delivery were studied.

The detailed cryo-EM structures of the two novel bacteriophages from the genus ϕ Eco32-like viruses”: *E.Coli* phage ϕ Eco32 and *Salmonella* phage 7-11 have been determined. Structural components of the tail, such as the portal, gatekeeper and nozzle were identified. Morphology of the tail fibers and their proteins constitution is clearly evident from the current cryo-EM reconstructions. The mechanism of the phage adsorption to the host cell has been hypothesized.

Future structural studies of ϕ Eco32 and 7-11 phages include:

1. Determination of the 7-11 capsid structure and improvement of ϕ Eco32 capsid reconstruction using helical reconstruction technique. The ϕ Eco32 and 7-11 virions are enormously elongated and have preferred orientations in vitreous ice. The helical reconstruction of the mature capsids is complicated by the presence of several helical symmetries. It might be possible to prepare and reconstruct the ϕ Eco32 and 7-11 samples without DNA.
2. Complete assignment of the tail hub and tail fiber proteins in ϕ Eco32 and 7-11 virions. Such assignment will explain in more details the reversible and further irreversible attachment to the host cell, and formation of the trans-membrane channel for DNA passage.
3. Cryo-ET of the ϕ Eco32 penetration into *E.Coli* mini-cells to visualize the phage-cell interaction process in detail. We are confident that we will be able to prepare mini-cells, which can be infected by ϕ Eco32, but the mini-cell preparation might represent a certain difficulty.

The other project was the structural characterization of bacteriophage ϕ 92 infecting a wide variety of acapsular *Escherichia coli* laboratory strains and large number of *Salmonella* strains. In addition, it infects pathogenic capsular *E.Coli* strains causing severe invasive infections. Complete three-dimensional model of the ϕ 92 virion was built, and adsorption apparatus was characterized. The baseplate of ϕ 92 has a Swiss army knife-like structure with several emanating tail fibers and tailspikes. This project illustrates that phages display an enormous potential to change host recognition proteins. Future work requires structural investigation of the disordered tail fibers, and process of the phage penetration into *E.Coli* mini-cells by cryo-tomography and sub-volume averaging. In addition, the symmetry-free (asymmetric) reconstruction of the baseplate might be

possible to calculate.

One of important aspects of this thesis was the structural characterization of the jumbo phages. Flagella-specific jumbo bacteriophage AR9 infecting *Bacillus subtilis* strain has the tail of remarkable complexity. The complete three-dimensional model of the AR9 virion was built. To our knowledge, it is the largest calculated three-dimensional reconstruction of tailed bacteriophage. Baseplate composition indicates that AR9 phage could utilize curly tail fibers for attachment and “riding on” the motile *B. subtilis* flagellum. Future aims in AR9 structural studies include high-resolution reconstruction of the curly fibers made up from subunits, and determination of the functions of unique helical fibers surrounding tail sheath.

The project of structural characterization of bacterial phage CBA120 specifically infecting an important food-borne pathogenic *E.Coli* strain O157 resulted in the cryo-EM models of virion with extended and contracted tail. The baseplate is very complex with numerous tail spikes and fibers protruding from it. At present and to our knowledge, it is the most complex tail structure investigated by anyone in the world. The receptor-binding part of the tailspikes was disordered in the cryo-EM map of extended tail, but fixed in low-resolution contracted tail map.

Future plans are:

1. Calculate high-resolution reconstruction of the star-like complex, probably made up from four tailspikes. Fixed conformation of the tailspikes was observed in CBA120 contracted tail. However, it is a complicated study because phage virions are very fragile and tend to disassemble after tail contraction. One of the solutions is the isolation of tailspikes, and calculation of cryo-EM reconstruction of joint tailspikes separately.
2. Improvement of CBA120 capsid reconstruction and determination of the spatial organization of decoration proteins. These decoration proteins do not follow the icosahedral symmetry of the capsid shell, and asymmetric reconstruction may require. For that the large amount of data has to be collected.

Interesting project was the cryo-EM reconstruction of the mutant T4 bacteriophage. The mutant baseplate lacks the tip protein of the cell-puncturing device responsible for the first interaction with the host cell membrane. Three-dimensional cryo-EM reconstruction of the baseplate and a tail fragment is conclusive about gp5.4 being the tip protein of T4. Cryo-EM map was compared with that of the wild type T4 phage and the distant member of T4-superfamily phage RB43. Interesting direction of future research is to determine the high-resolution structure of the RB43 tip protein, which probably interacts with the host cell surface in a specific way.

8. Description of author contribution to the projects

Bacteriophages ϕ Eco32 and 7-11

Purification of bacteriophages ϕ Eco32 and 7-11 (supervised by Dr. Mikhail Shneider). Samples preparation for EM data collection using negative staining, cryo-negative staining and cryo-fixation (cryo-EM), conditions optimization. Manual and automated EM images collection both on CCD detector and DDD. Cryo-ET data collection. Processing of collected micrographs: CTF correction, filtering, particles boxing both manual and automated, drift correction. Initial models creation, iterative refinement. 3D reconstructions interpretation (segmentation of reconstructions, preparation of figures). Processing of collected tomograms: alignment, reconstruction, denoising, sub-volumes picking, extraction and alignment, sub-volumes averaging, interpretation. Absolute hand determination using tilt pairs technique. Creation of custom scripts and modification of available programs. Bioinformatical analysis of phages genomes. Assignment of proteins in cryo-EM maps. Paper preparation for publication.

Bacteriophage ϕ 92

Processing of scanned micrographs: CTF correction, filtering, particles boxing, both manual and automated. Improvement of previously reconstructed models (supervised by Dr. Sergii Buth). Interpretation of 3D reconstructions (segmentation of reconstructions, preparation of figures). Processing of collected tomograms: alignment, reconstruction, denoising, sub-volumes picking, extraction and alignment, sub-volumes averaging, interpretation. Creating of custom scripts and modification of available programs. Bioinformatical analysis of phages genome. Assignment of proteins in cryo-EM maps. Paper publication.

Bacteriophage CBA120

Purification of bacteriophage CBA120. Conditions optimization for phage contraction. Samples preparation for EM data collection using negative staining and cryo-fixation (cryo-EM), conditions optimization. Manual and automated EM images collection both on CCD detector and DDD. Processing of collected micrographs: CTF correction, filtering, particles boxing, both manual and automated, drift correction. Initial models creation, iterative refinement. 3D reconstructions interpretation (segmentation of reconstructions, preparation of figures). 2D analysis of disassembled tailspikes. Creation of custom scripts and modification of available programs. Bioinformatical analysis of phage genome. Assignment of proteins in cryo-EM maps.

Bacteriophage AR9

Purification of bacteriophage AR9. Samples preparation for EM data collection using negative staining and cryo-fixation (cryo-EM), conditions optimization. Manual EM images collection on CCD detector. Processing of collected micrographs: CTF correction, filtering, particles boxing, both manual and automated. Initial models creation, iterative refinement. 3D reconstructions interpretation (segmentation of reconstructions, preparation of figures). Creation of custom scripts and modification of available programs. Preliminary proteins assignment in cryo-EM maps.

Bacteriophages T4 and RB43

Purification of bacteriophages T4 and RB43. Samples preparation for EM data collection using negative staining and cryo-fixation (cryo-EM), conditions optimization. Manual EM images collection on CCD detector. Processing of collected micrographs: CTF correction, filtering, particles boxing, both manual and automated. Initial model creation, improvement of previously reconstructed model, iterative refinement. 3D reconstructions interpretation (segmentation of reconstructions, preparation of figures). Creation of custom scripts.

BIBLIOGRAPHY (REFERENCES)

- Abuladze, N. K., M. Gingery, J. Tsai, and F. A. Eiserling. 1994. "Tail Length Determination in Bacteriophage T4." *Virology* 199 (2): 301-10. doi:10.1006/viro.1994.1128.
- Ackermann, Hans-W. 2011. "The First Phage Electron Micrographs." *Bacteriophage* 1 (4): 225-27. doi:10.4161/bact.1.4.17280.
- Ackermann, H W. 1998. "Tailed Bacteriophages: The Order Caudovirales." *Advances in Virus Research* 51: 135-201.
- Ackermann, H-W. 2003. "Bacteriophage Observations and Evolution." *Research in Microbiology* 154 (4): 245-51. doi:10.1016/S0923-2508(03)00067-6.
- Ackermann, H.-W. 2006. "5500 Phages Examined in the Electron Microscope." *Archives of Virology* 152 (2): 227-43. doi:10.1007/s00705-006-0849-1.
- Ackermann, H. W., and H. M. Krisch. 1997. "A Catalogue of T4-Type Bacteriophages." *Archives of Virology* 142 (12): 2329-45.
- Ackermann, H.-W., and D. Prangishvili. 2012. "Prokaryote Viruses Studied by Electron Microscopy." *Archives of Virology* 157 (10): 1843-49. doi:10.1007/s00705-012-1383-y.
- Adriaenssens, Evelien M., Hans-Wolfgang Ackermann, Hany Anany, Bob Blasdel, Ian F. Connerton, David Goulding, Mansel W. Griffiths, et al. 2012. "A Suggested New Bacteriophage Genus: 'Viunaliavirus.'" *Archives of Virology* 157 (10): 2035-46. doi:10.1007/s00705-012-1360-5.
- Adrian, Marc, Jacques Dubochet, Jean Lepault, and Alasdair W. McDowell. 1984. "Cryo-Electron Microscopy of Viruses." *Nature* 308 (5954): 32-36. doi:10.1038/308032a0.
- Adrian, M., J. Dubochet, S. D. Fuller, and J. R. Harris. 1998. "Cryo-Negative Staining." *Micron (Oxford, England: 1993)* 29 (2-3): 145-60.
- Aebi, U, R Bijlenga, J v d Broek, H v d Broek, F Eiserling, C Kellenberger, E Kellenberger, et al. 1974. "The Transformation of Tau Particles into T4 Heads. II. Transformations of the Surface Lattice and Related Observations on Form Determination." *Journal of Supramolecular Structure* 2 (2-4): 253-75. doi:10.1002/jss.400020218.
- Aksyuk, A. A., V. D. Bowman, B. Kaufmann, C. Fields, T. Klose, H. A. Holdaway, V. A. Fischetti, and M. G. Rossmann. 2012. "Structural Investigations of a Podoviridae Streptococcus Phage C1, Implications for the Mechanism of Viral Entry." *Proceedings of the National Academy of Sciences* 109 (35): 14001-6. doi:10.1073/pnas.1207730109.
- Aksyuk, Anastasia A., Petr G. Leiman, Lidia P. Kurochkina, Mikhail M. Shneider, Victor A. Kostyuchenko, Vadim V. Mesyanzhinov, and Michael G. Rossmann. 2009. "The Tail Sheath Structure of Bacteriophage T4: A Molecular Machine for Infecting Bacteria." *The EMBO Journal* 28 (7): 821-29. doi:10.1038/emboj.2009.36.
- Appelbaum, P. C., N. Hugo, and J. N. Coetzee. 1971. "A Flagellar Phage for the Proteus-Providence Group." *Journal of General Virology* 13 (1): 153-62. doi:10.1099/0022-1317-13-1-153.
- Baker, M. L., W. Jiang, F. J. Rixon, and W. Chiu. 2005. "Common Ancestry of Herpesviruses and Tailed DNA Bacteriophages." *Journal of Virology* 79 (23): 14967-70. doi:10.1128/JVI.79.23.14967-14970.2005.
- Baker, T. S., and R. Henderson. 2006. "Electron Cryomicroscopy." In *International Tables for Crystallography*, edited by M. G. Rossmann and E. Arnold, 1st ed., F:451-63. Chester, England: International Union of Crystallography. http://xrpp.iucr.org/cgi-bin/itr?url_ver=Z39.88-2003&rft_dat=what%3Dchapter%26volid%3DFa%26chnum%3D19o6%26chvers%3Dv0001.
- Baker, T S, N H Olson, and S D Fuller. 1999. "Adding the Third Dimension to Virus Life Cycles: Three-Dimensional Reconstruction of Icosahedral Viruses from Cryo-Electron Micrographs." *Microbiology and Molecular Biology Reviews: MMBR* 63 (4): 862-922, table of contents.
- Bamford, Dennis H., Jonathan M. Grimes, and David I. Stuart. 2005. "What Does Structure Tell Us about Virus Evolution?" *Current Opinion in Structural Biology* 15 (6): 655-63. doi:10.1016/j.sbi.2005.10.012.
- Barry, G. T., and W. F. Goebel. 1957. "Colominic Acid, a Substance of Bacterial Origin Related to Sialic Acid." *Nature* 179 (4552): 206.
- Bartual, Sergio G., José M. Otero, Carmela Garcia-Doval, Antonio L. Llamas-Saiz, Richard Kahn, Gavin C. Fox, and Mark J. van Raaij. 2010. "Structure of the Bacteriophage T4 Long Tail Fiber Receptor-Binding Tip." *Proceedings of the National Academy of Sciences of the United States of America* 107 (47): 20287-92. doi:10.1073/pnas.1011218107.
- Bateman, A., S. R. Eddy, and V. V. Mesyanzhinov. 1997. "A Member of the Immunoglobulin Superfamily in Bacteriophage T4." *Virus Genes* 14 (2): 163-65.
- Bayer, Manfred E., Horst Thurow, and Margret H. Bayer. 1979. "Penetration of the Polysaccharide Capsule of Escherichia Coli (B1161/42) by Bacteriophage K29." *Virology* 94 (1): 95-118. doi:10.1016/0042-6822(79)90441-0.

- B., David, and James E. 2012. "Low-Dose Imaging Techniques for Transmission Electron Microscopy." In *The Transmission Electron Microscope*, edited by Khan Maaz. InTech. <http://www.intechopen.com/books/the-transmission-electron-microscope/low-dose-imaging-techniques-for-transmission-electron-microscopy>.
- Belyaeva, Nina N., and Rudolf R. Azizbekyan. 1968. "Fine Structure of New Bacillus Subtilis Phage AR9 with Complex Morphology." *Virology* 34 (1): 176–79. doi:10.1016/0042-6822(68)90023-8.
- Bergh, O., K. Y. Børshheim, G. Bratbak, and M. Haldal. 1989. "High Abundance of Viruses Found in Aquatic Environments." *Nature* 340 (6233): 467–68. doi:10.1038/340467a0.
- Bernard Heymann, J., Giovanni Cardone, Dennis C. Winkler, and Alasdair C. Steven. 2008. "Computational Resources for Cryo-Electron Tomography in Bsoft." *Journal of Structural Biology* 161 (3): 232–42. doi:10.1016/j.jsb.2007.08.002.
- Berriman, John, and Nigel Unwin. 1994. "Analysis of Transient Structures by Cryo-Microscopy Combined with Rapid Mixing of Spray Droplets." *Ultramicroscopy* 56 (4): 241–52. doi:10.1016/0304-3991(94)90012-4.
- Beveridge, T J, and L L Graham. 1991. "Surface Layers of Bacteria." *Microbiological Reviews* 55 (4): 684–705.
- Bhardwaj, Anshul, Nancy Walker-Kopp, Sherwood R Casjens, and Gino Cingolani. 2009. "An Evolutionarily Conserved Family of Virion Tail Needles Related to Bacteriophage P22 gp26: Correlation between Structural Stability and Length of the Alpha-Helical Trimeric Coiled Coil." *Journal of Molecular Biology* 391 (1): 227–45. doi:10.1016/j.jmb.2009.05.069.
- Black, Lindsay W., and Julie A. Thomas. 2012. "Condensed Genome Structure." *Advances in Experimental Medicine and Biology* 726: 469–87. doi:10.1007/978-1-4614-0980-9_21.
- Boyd, E. Fidelma, and Harald Brüssow. 2002. "Common Themes among Bacteriophage-Encoded Virulence Factors and Diversity among the Bacteriophages Involved." *Trends in Microbiology* 10 (11): 521–29.
- Bracewell, Ronald N. 2003. *Fourier Analysis and Imaging*. New York: Kluwer Academic/Plenum Publishers.
- Brown, A., A. Amunts, X.-c. Bai, Y. Sugimoto, P. C. Edwards, G. Murshudov, S. H. W. Scheres, and V. Ramakrishnan. 2014. "Structure of the Large Ribosomal Subunit from Human Mitochondria." *Science* 346 (6210): 718–22. doi:10.1126/science.1258026.
- Browning, Christopher, Mikhail M Shneider, Valorie D Bowman, David Schwarzer, and Petr G Leiman. 2012. "Phage Pierces the Host Cell Membrane with the Iron-Loaded Spike." *Structure (London, England: 1993)* 20 (2): 326–39. doi:10.1016/j.str.2011.12.009.
- Brüssow, Harald, and Roger W. Hendrix. 2002. "Phage Genomics: Small Is Beautiful." *Cell* 108 (1): 13–16.
- Buchan, D. W. A., S. M. Ward, A. E. Lobley, T. C. O. Nugent, K. Bryson, and D. T. Jones. 2010. "Protein Annotation and Modelling Servers at University College London." *Nucleic Acids Research* 38 (Web Server issue): W563–68. doi:10.1093/nar/gkq427.
- Bui, Khanh Huy, Hitoshi Sakakibara, Tandis Movassagh, Kazuhiro Oiwa, and Takashi Ishikawa. 2008. "Molecular Architecture of Inner Dynein Arms in Situ in Chlamydomonas Reinhardtii Flagella." *The Journal of Cell Biology* 183 (5): 923–32. doi:10.1083/jcb.200808050.
- Butcher, James, Sabina Sarvan, Joseph S. Brunzelle, Jean-François Couture, and Alain Stintzi. 2012. "Structure and Regulon of Campylobacter Jejuni Ferric Uptake Regulator Fur Define Apo-Fur Regulation." *Proceedings of the National Academy of Sciences of the United States of America* 109 (25): 10047–52. doi:10.1073/pnas.1118321109.
- Calendar, Richard, ed. 2006. *The Bacteriophages*. 2nd ed. Oxford ; New York: Oxford University Press.
- Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. 2003. "Prophage Genomics." *Microbiology and Molecular Biology Reviews* 67 (2): 238–76. doi:10.1128/MMBR.67.2.238-276.2003.
- Casjens, Sherwood. 2003. "Prophages and Bacterial Genomics: What Have We Learned so Far?" *Molecular Microbiology* 49 (2): 277–300.
- Casjens, Sherwood R, and Ian J Molineux. 2012. "Short Noncontractile Tail Machines: Adsorption and DNA Delivery by Podoviruses." *Advances in Experimental Medicine and Biology* 726: 143–79. doi:10.1007/978-1-4614-0980-9_7.
- Caspar, D. L., and A. Klug. 1962. "Physical Principles in the Construction of Regular Viruses." *Cold Spring Harbor Symposia on Quantitative Biology* 27: 1–24.
- Castañó-Díez, Daniel, Ashraf Al-Amoudi, Anne-Marie Glynn, Anja Seybert, and Achilleas S. Frangakis. 2007. "Fiducial-Less Alignment of Cryo-Sections." *Journal of Structural Biology* 159 (3): 413–23. doi:10.1016/j.jsb.2007.04.014.
- Castañó-Díez, Daniel, Mikhail Kudryashev, Marcel Arheit, and Henning Stahlberg. 2012. "Dynamo: A Flexible, User-Friendly Development Tool for Subtomogram Averaging of Cryo-EM Data in High-Performance Computing Environments." *Journal of Structural Biology* 178 (2): 139–51. doi:10.1016/j.jsb.2011.12.017.
- Castañó-Díez, Daniel, Margot Scheffer, Ashraf Al-Amoudi, and Achilleas S. Frangakis. 2010. "Alignator: A GPU Powered Software Package for Robust Fiducial-Less Alignment of Cryo Tilt-Series." *Journal of Structural Biology* 170 (1): 117–26. doi:10.1016/j.jsb.2010.01.014.

- Cavalier, Annie, Daniele Spehner, and Bruno M. Humbel, eds. 2009. *Handbook of Cryo-Preparation Methods for Electron Microscopy*. Methods in Visualization. Boca Raton: CRC Press.
- Cerritelli, M E, N Cheng, A H Rosenberg, C E McPherson, F P Booy, and A C Steven. 1997. "Encapsidated Conformation of Bacteriophage T7 DNA." *Cell* 91 (2): 271-80.
- Chen, Chen, Patrick Bales, Julia Greenfield, Ryan D. Heselpoth, Daniel C. Nelson, and Osnat Herzberg. 2014. "Crystal Structure of ORF210 from E.Coli O157:H1 Phage CBA120 (TSP1), a Putative Tailspike Protein." *PLoS ONE* 9 (3): e93156. doi:10.1371/journal.pone.0093156.
- Chen, D.-H., M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, et al. 2011. "Structural Basis for Scaffolding-Mediated Assembly and Maturation of a dsDNA Virus." *Proceedings of the National Academy of Sciences* 108 (4): 1355-60. doi:10.1073/pnas.1015739108.
- Chiu, W., and R. M. Glaeser. 1977. "Factors Affecting High Resolution Fixed-Beam Transmission Electron Microscopy." *Ultramicroscopy* 2 (2-3): 207-17.
- Choi, Younho, Hakdong Shin, Ju-Hoon Lee, and Sangryeol Ryu. 2013. "Identification and Characterization of a Novel Flagellum-Dependent Salmonella-Infecting Bacteriophage, iEPS5." *Applied and Environmental Microbiology* 79 (16): 4829-37. doi:10.1128/AEM.00706-13.
- Cross, A. S. 1990. "The Biologic Significance of Bacterial Encapsulation." In *Bacterial Capsules*, edited by Prof Dr Klaus Jann and Dr Barbara Jann, 87-95. Current Topics in Microbiology and Immunology 150. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-74694-9_5.
- Crowther, R. A. 1971. "Procedures for Three-Dimensional Reconstruction of Spherical Viruses by Fourier Synthesis from Electron Micrographs." *Philosophical Transactions of the Royal Society B: Biological Sciences* 261 (837): 221-30. doi:10.1098/rstb.1971.0054.
- Cuervo, Ana, and José L Carrascosa. 2012. "Viral Connectors for DNA Encapsulation." *Current Opinion in Biotechnology* 23 (4): 529-36. doi:10.1016/j.copbio.2011.11.029.
- Davidson, Alan R., Lia Cardarelli, Lisa G. Pell, Devon R. Radford, and Karen L. Maxwell. 2012. "Long Noncontractile Tail Machines of Bacteriophages." *Advances in Experimental Medicine and Biology* 726: 115-42. doi:10.1007/978-1-4614-0980-9_6.
- Depping, Reinhard, Christiane Lohaus, Helmut E. Meyer, and Wolfgang Rüger. 2005. "The Mono-ADP-Ribosyltransferases Alt and ModB of Bacteriophage T4: Target Proteins Identified." *Biochemical and Biophysical Research Communications* 335 (4): 1217-23. doi:10.1016/j.bbrc.2005.08.023.
- Desplats, Carine, Christophe Dez, Françoise Tétart, Heïdy Eleaume, and H. M. Krisch. 2002. "Snapshot of the Genome of the Pseudo-T-Even Bacteriophage RB49." *Journal of Bacteriology* 184 (10): 2789-2804.
- Dokland, T. 1999. "Scaffolding Proteins and Their Role in Viral Assembly." *Cellular and Molecular Life Sciences: CMLS* 56 (7-8): 580-603.
- Dokland, Terje. 2000. "Freedom and Restraint: Themes in Virus Capsid Assembly." *Structure* 8 (8): R157-62. doi:10.1016/S0969-2126(00)00181-7.
- Donelli, G., F. Guglielmi, and L. Paoletti. 1972. "Structure and Physico-Chemical Properties of Bacteriophage G." *Journal of Molecular Biology* 71 (2): 113-25. doi:10.1016/0022-2836(72)90341-5.
- Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowell, and P. Schultz. 1988. "Cryo-Electron Microscopy of Vitrified Specimens." *Quarterly Reviews of Biophysics* 21 (2): 129-228.
- Earnshaw, W. C., and S. R. Casjens. 1980. "DNA Packaging by the Double-Stranded DNA Bacteriophages." *Cell* 21 (2): 319-31.
- Earnshaw, W.C., J. King, S.C. Harrison, and F.A. Eiserling. 1978. "The Structural Organization of DNA Packaged within the Heads of T4 Wild-Type, Isometric and Giant Bacteriophages." *Cell* 14 (3): 559-68. doi:10.1016/0092-8674(78)90242-8.
- Effantin, G., P. Boulanger, E. Neumann, L. Letellier, and J. F. Conway. 2006. "Bacteriophage T5 Structure Reveals Similarities with HK97 and T4 Suggesting Evolutionary Relationships." *Journal of Molecular Biology* 361 (5): 993-1002. doi:10.1016/j.jmb.2006.06.081.
- Egan, W., T. Y. Liu, D. Dorow, J. S. Cohen, J. D. Robbins, E. C. Gotschlich, and J. B. Robbins. 1977. "Structural Studies on the Sialic Acid Polysaccharide Antigen of Escherichia Coli Strain Bos-12." *Biochemistry* 16 (16): 3687-92.
- Eiserling, F. A. 1967. "The Structure of Bacillus Subtilis Bacteriophage PBS 1." *Journal of Ultrastructure Research* 17 (3-4): 342-47. doi:10.1016/S0022-5320(67)80053-4.
- Erickson, H. P., and A. Klug. 1971. "Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs." *Philosophical Transactions of the Royal Society B: Biological Sciences* 261 (837): 105-18. doi:10.1098/rstb.1971.0040.

- Fernández, J. J., S. Li, and R. A. Crowther. 2006. "CTF Determination and Correction in Electron Cryotomography." *Ultramicroscopy* 106 (7): 587–96. doi:10.1016/j.ultramic.2006.02.004.
- Fernández, J. J., D. Luque, J. R. Castón, and J. L. Carrascosa. 2008. "Sharpening High Resolution Information in Single Particle Electron Cryomicroscopy." *Journal of Structural Biology* 164 (1): 170–75. doi:10.1016/j.jsb.2008.05.010.
- Ferrari, E., U. Canosi, A. Galizzi, and G. Mazza. 1978. "Studies on Transduction Process by SPP1 Phage." *The Journal of General Virology* 41 (3): 563–72.
- Fischer, Hannes, Igor Polikarpov, and Aldo F. Craievich. 2009. "Average Protein Density Is a Molecular-Weight-Dependent Function." *Protein Science* 13 (10): 2825–28. doi:10.1110/ps.04688204.
- Fokine, Andrei, Paul R Chipman, Petr G Leiman, Vadim V Mesyanzhinov, Venigalla B Rao, and Michael G Rossmann. 2004. "Molecular Architecture of the Prolate Head of Bacteriophage T4." *Proceedings of the National Academy of Sciences of the United States of America* 101 (16): 6003–8. doi:10.1073/pnas.0400444101.
- Fokine, Andrei, Victor A Kostyuchenko, Andrey V Efimov, Lidia P Kurochkina, Nina N Sykilinda, Johan Robben, Guido Volckaert, et al. 2005. "A Three-Dimensional Cryo-Electron Microscopy Structure of the Bacteriophage phiK2 Head." *Journal of Molecular Biology* 352 (1): 117–24. doi:10.1016/j.jmb.2005.07.018.
- Fokine, Andrei, Petr G. Leiman, Mikhail M. Shneider, Bijan Ahvazi, Karen M. Boeshans, Alasdair C. Steven, Lindsay W. Black, Vadim V. Mesyanzhinov, and Michael G. Rossmann. 2005. "Structural and Functional Similarities between the Capsid Proteins of Bacteriophages T4 and HK97 Point to a Common Ancestry." *Proceedings of the National Academy of Sciences of the United States of America* 102 (20): 7163–68. doi:10.1073/pnas.0502164102.
- Fotin, Alexander, Tomas Kirchhausen, Nikolaus Grigorieff, Stephen C. Harrison, Thomas Walz, and Yifan Cheng. 2006. "Structure Determination of Clathrin Coats to Subnanometer Resolution by Single Particle Cryo-Electron Microscopy." *Journal of Structural Biology* 156 (3): 453–60. doi:10.1016/j.jsb.2006.07.001.
- Frank, J. 2006a. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. 2nd ed. Oxford ; New York: Oxford University Press.
- . 2006b. *Electron Tomography Methods for Three-Dimensional Visualization of Structures in the Cell*. New York: Springer. <http://public.eblib.com/choice/publicfullrecord.aspx?p=371740>.
- Fraser, James S, Zhou Yu, Karen L Maxwell, and Alan R Davidson. 2006. "Ig-like Domains on Bacteriophages: A Tale of Promiscuity and Deceit." *Journal of Molecular Biology* 359 (2): 496–507. doi:10.1016/j.jmb.2006.03.043.
- Fujisawa, H., and M. Morita. 1997. "Phage DNA Packaging." *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 2 (9): 537–45.
- Furowicz, A. J., and F. Orskov. 1972. "Two New Escherichia Coli O Antigens, O150 and O157, and One New K Antigen, K92, in Strains Isolated from Veterinary Diseases." *Acta Pathologica Et Microbiologica Scandinavica. Section B: Microbiology and Immunology* 80 (3): 441–44.
- Gerardy-Schahn, R., A. Bethe, T. Brennecke, M. Mühlenhoff, M. Eckhardt, S. Ziesing, F. Lottspeich, and M. Frosch. 1995. "Molecular Cloning and Functional Expression of Bacteriophage PK1E-Encoded Endoneuraminidase Endo NE." *Molecular Microbiology* 16 (3): 441–50.
- Gilbert, Peter. 1972. "Iterative Methods for the Three-Dimensional Reconstruction of an Object from Projections." *Journal of Theoretical Biology* 36 (1): 105–17. doi:10.1016/0022-5193(72)90180-4.
- Gillis, Annika, and Jacques Mahillon. 2014. "Phages Preying on Bacillus Anthracis, Bacillus Cereus, and Bacillus Thuringiensis: Past, Present and Future." *Viruses* 6 (7): 2623–72. doi:10.3390/v6072623.
- Gipson, Preeti, Matthew L. Baker, Desislava Raytcheva, Cameron Haase-Pettingell, Jacqueline Piret, Jonathan A. King, and Wah Chiu. 2014. "Protruding Knob-like Proteins Violate Local Symmetries in an Icosahedral Marine Virus." *Nature Communications* 5 (July). doi:10.1038/ncomms5278.
- Glaeser, R. M., and K. A. Taylor. 1978. "Radiation Damage Relative to Transmission Electron Microscopy of Biological Specimens at Low Temperature: A Review." *Journal of Microscopy* 112 (1): 127–38.
- Grassucci, Robert A., Derek Taylor, and Joachim Frank. 2008. "Visualization of Macromolecular Complexes Using Cryo-Electron Microscopy with FEI Tecnai Transmission Electron Microscopes." *Nature Protocols* 3 (2): 330–39. doi:10.1038/nprot.2007.474.
- Grigorieff, Nikolaus. 2007. "FREALIGN: High-Resolution Refinement of Single Particle Structures." *Journal of Structural Biology* 157 (1): 117–25. doi:10.1016/j.jsb.2006.05.004.
- Gross, R. J., T. Cheasty, and B. Rowe. 1977. "Isolation of Bacteriophages Specific for the K1 Polysaccharide Antigen of Escherichia Coli." *Journal of Clinical Microbiology* 6 (6): 548–50.
- Guckenberger, R. 1982. "Determination of a Common Origin in the Micrographs of Tilt Series in Three-Dimensional Electron Microscopy." *Ultramicroscopy* 9 (1-2): 167–73. doi:10.1016/0304-3991(82)90239-X.
- Guerrero-Ferreira, Ricardo C., Patrick H. Viollier, Bert Ely, Jeanne S. Poindexter, Maria Georgieva, Grant J. Jensen, and

- Elizabeth R. Wright. 2011. "Alternative Mechanism for Bacteriophage Adsorption to the Motile Bacterium *Caulobacter Crescentus*." *Proceedings of the National Academy of Sciences of the United States of America* 108 (24): 9963–68. doi:10.1073/pnas.1012388108.
- Guo, Fei, Zheng Liu, Frank Vago, Yue Ren, Weimin Wu, Elena T. Wright, Philip Serwer, and Wen Jiang. 2013. "Visualization of Uncorrelated, Tandem Symmetry Mismatches in the Internal Genome Packaging Apparatus of Bacteriophage T7." *Proceedings of the National Academy of Sciences of the United States of America* 110 (17): 6811–16. doi:10.1073/pnas.1215563110.
- Haguenau, F., P.W. Hawkes, J.L. Hutchison, B. Satiat-Jeunemaître, G.T. Simon, and D.B. Williams. 2003. "Key Events in the History of Electron Microscopy." *Microscopy and Microanalysis* 9 (02): 96–138. doi:10.1017/S1431927603030113.
- Harauz, G., E. Boekema, and M. van Heel. 1988. "Statistical Image Analysis of Electron Micrographs of Ribosomal Subunits." *Methods in Enzymology* 164: 35–49.
- Harris, James R., and Royal Microscopical Society (Great Britain). 1997. *Negative Staining and Cryoelectron Microscopy: The Thin Film Techniques*. Oxford, UK; Herndon, VA: BIOS Scientific Publishers in association with the Royal Microscopical Society.
- Harris, J. Robin, and Dirk Scheffler. 2002. "Routine Preparation of Air-Dried Negatively Stained and Unstained Specimens on Holey Carbon Support Films: A Review of Applications." *Micron (Oxford, England: 1993)* 33 (5): 461–80.
- Hatfull, Graham F. 2008. "Bacteriophage Genomics." *Current Opinion in Microbiology* 11 (5): 447–53. doi:10.1016/j.mib.2008.09.004.
- Helgstrand, Charlotte, William R. Wikoff, Robert L. Duda, Roger W. Hendrix, John E. Johnson, and Lars Liljas. 2003. "The Refined Structure of a Protein Catenane: The HK97 Bacteriophage Capsid at 3.44 Å Resolution." *Journal of Molecular Biology* 334 (5): 885–99.
- Henderson, R. 2013. "Avoiding the Pitfalls of Single Particle Cryo-Electron Microscopy: Einstein from Noise." *Proceedings of the National Academy of Sciences* 110 (45): 18037–41. doi:10.1073/pnas.1314449110.
- Henderson, Richard, Andrej Sali, Matthew L. Baker, Bridget Carragher, Batsal Devkota, Kenneth H. Downing, Edward H. Egelman, et al. 2012. "Outcome of the First Electron Microscopy Validation Task Force Meeting." *Structure* 20 (2): 205–14. doi:10.1016/j.str.2011.12.014.
- Hendrix, R. W. 2009. "Jumbo Bacteriophages." *Current Topics in Microbiology and Immunology* 328: 229–40.
- Hendrix, R W, M C Smith, R N Burns, M E Ford, and G F Hatfull. 1999. "Evolutionary Relationships among Diverse Bacteriophages and Prophages: All the World's a Phage." *Proceedings of the National Academy of Sciences of the United States of America* 96 (5): 2192–97.
- Henning, U., and S. Hashemolhosseini. 1994. "Receptor Recognition by T-Even-Type Coliphages." *Molecular Biology of Bacteriophage* 4: 291–98.
- Herman, Gabor T. 2009. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. 2nd ed. Advances in Pattern Recognition. Dordrecht; New York: Springer.
- Hrabe, Thomas, and Friedrich Förster. 2011. "Structure Determination by Single Particle Cryo-Electron Tomography." In *Encyclopedia of Life Sciences*, edited by John Wiley & Sons, Ltd. Chichester, UK: John Wiley & Sons, Ltd. <http://www.els.net/WileyCDA/ElsArticle/refid-a0023175.html>.
- Huang, Bo, Hazen Babcock, and Xiaowei Zhuang. 2010. "Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells." *Cell* 143 (7): 1047–58. doi:10.1016/j.cell.2010.12.002.
- Hu, B., W. Margolin, I. J. Molineux, and J. Liu. 2013. "The Bacteriophage T7 Virion Undergoes Extensive Structural Remodeling During Infection." *Science* 339 (6119): 576–79. doi:10.1126/science.1231887.
- International Committee on Taxonomy of Viruses. 2012. *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Edited by Andrew M. Q. King. London; Waltham, MA: Academic Press.
- Iwasaki, K., B. L. Trus, P. T. Wingfield, N. Cheng, G. Campusano, V. B. Rao, and A. C. Steven. 2000. "Molecular Architecture of Bacteriophage T4 Capsid: Vertex Structure and Bimodal Binding of the Stabilizing Accessory Protein, Soc." *Virology* 271 (2): 321–33. doi:10.1006/viro.2000.0321.
- Jakobsson, Elina, David Schwarzer, Anne Jokilampi, and Jukka Finne. 2012. "Endosialidases: Versatile Tools for the Study of Polysialic Acid." *Topics in Current Chemistry*, August. doi:10.1007/128_2012_349.
- Jiang, Wen, Matthew L. Baker, Joanita Jakana, Peter R. Weigele, Jonathan King, and Wah Chiu. 2008. "Backbone Structure of the Infectious epsilon15 Virus Capsid Revealed by Electron Cryomicroscopy." *Nature* 451 (7182): 1130–34. doi:10.1038/nature06665.
- Jiang, Wen, Juan Chang, Joanita Jakana, Peter Weigele, Jonathan King, and Wah Chiu. 2006. "Structure of epsilon15 Bacteriophage Reveals Genome Organization and DNA Packaging/injection Apparatus." *Nature* 439 (7076): 612–

16. doi:10.1038/nature04487.
- Johnson, John E., and Wah Chiu. 2007. "DNA Packaging and Delivery Machines in Tailed Bacteriophages." *Current Opinion in Structural Biology* 17 (2): 237–43. doi:10.1016/j.sbi.2007.03.011.
- Joyeux, Laurent, and Pawel A. Penczek. 2002. "Efficiency of 2D Alignment Methods." *Ultramicroscopy* 92 (2): 33–46.
- Joys, Terence M. 1965. "Correlation Between Susceptibility to Bacteriophage PBS1 and Motility in *Bacillus Subtilis*." *Journal of Bacteriology* 90 (6): 1575–77.
- Kanamaru, Shuji, Petr G. Leiman, Victor A. Kostyuchenko, Paul R. Chipman, Vadim V. Mesyanzhinov, Fumio Arisaka, and Michael G. Rossmann. 2002. "Structure of the Cell-Puncturing Device of Bacteriophage T4." *Nature* 415 (6871): 553–57. doi:10.1038/415553a.
- Karam, Jim D., and John W. Drake, eds. 1994. *Molecular Biology of Bacteriophage T4*. Washington, DC: American Society for Microbiology.
- Katsura, Isao, and Roger W. Hendrix. 1984. "Length Determination in Bacteriophage Lambda Tails." *Cell* 39 (3): 691–98. doi:10.1016/0092-8674(84)90476-8.
- Kivioja, T., J. Ravantti, A. Verkhovsky, E. Ukkonen, and D. Bamford. 2000. "Local Average Intensity-Based Method for Identifying Spherical Particles in Electron Micrographs." *Journal of Structural Biology* 131 (2): 126–34. doi:10.1006/jsbi.2000.4279.
- Knappek, E., and J. Dubochet. 1980. "Beam Damage to Organic Material Is Considerably Reduced in Cryo-Electron Microscopy." *Journal of Molecular Biology* 141 (2): 147–61. doi:10.1016/0022-2836(80)90382-4.
- Kostyuchenko, Victor A., Paul R. Chipman, Petr G. Leiman, Fumio Arisaka, Vadim V. Mesyanzhinov, and Michael G. Rossmann. 2005. "The Tail Structure of Bacteriophage T4 and Its Mechanism of Contraction." *Nature Structural & Molecular Biology* 12 (9): 810–13. doi:10.1038/nsmb975.
- Kostyuchenko, Victor A., Petr G. Leiman, Paul R. Chipman, Shuji Kanamaru, Mark J. van Raaij, Fumio Arisaka, Vadim V. Mesyanzhinov, and Michael G. Rossmann. 2003. "Three-Dimensional Structure of Bacteriophage T4 Baseplate." *Nature Structural Biology* 10 (9): 688–93. doi:10.1038/nsb970.
- Kremer, J R, D N Mastronarde, and J R McIntosh. 1996. "Computer Visualization of Three-Dimensional Image Data Using IMOD." *Journal of Structural Biology* 116 (1): 71–76. doi:10.1006/jsbi.1996.0013.
- Kropinski, Andrew M., Erika J. Lingohr, and Hans-Wolfgang Ackermann. 2011. "The Genome Sequence of Enterobacterial Phage 7-11, Which Possesses an Unusually Elongated Head." *Archives of Virology* 156 (1): 149–51. doi:10.1007/s00705-010-0835-5.
- Kruger, Dh, P Schneek, and Hr Gelderblom. 2000. "Helmut Ruska and the Visualisation of Viruses." *The Lancet* 355 (9216): 1713–17. doi:10.1016/S0140-6736(00)02250-9.
- Kutter, Elizabeth M, Kyobi Skutt-Kakaria, Bob Blasdel, Ayman El-Shibiny, Anna Castano, Daniel Bryan, Andrew M Kropinski, et al. 2011. "Characterization of a Vil-like Phage Specific to *Escherichia Coli* O157:H7." *Virology Journal* 8 (September): 430. doi:10.1186/1743-422X-8-430.
- Kwiatkowski, B, B Boschek, H Thiele, and S Stirm. 1983. "Substrate Specificity of Two Bacteriophage-Associated Endo-N-Acetylneuraminidases." *Journal of Virology* 45 (1): 367–74.
- Kwiatkowski, B., and S. Stirm. 1987. "Polysialic Acid Depolymerase." *Methods in Enzymology* 138: 786–92.
- Lander, Gabriel C., Liang Tang, Sherwood R. Casjens, Eddie B. Gilcrease, Peter Prevelige, Anton Poliakov, Clinton S. Potter, Bridget Carragher, and John E. Johnson. 2006. "The Structure of an Infectious P22 Virion Shows the Signal for Headful DNA Packaging." *Science (New York, N.Y.)* 312 (5781): 1791–95. doi:10.1126/science.1127981.
- Lebedev, Andrey A., Margret H. Krause, Anabela L. Isidro, Alexei A. Vagin, Elena V. Orlova, Joanne Turner, Eleanor J. Dodson, Paulo Tavares, and Alfred A. Antson. 2007. "Structural Framework for DNA Translocation via the Viral Portal Protein." *The EMBO Journal* 26 (7): 1984–94. doi:10.1038/sj.emboj.7601643.
- Lee, Ju-Hoon, Hakdong Shin, Younho Choi, and Sangryeol Ryu. 2013. "Complete Genome Sequence Analysis of Bacterial-Flagellum-Targeting Bacteriophage Chi." *Archives of Virology* 158 (10): 2179–83. doi:10.1007/s00705-013-1700-0.
- Leforestier, Amélie. 2013. "Polymorphism of DNA Conformation inside the Bacteriophage Capsid." *Journal of Biological Physics* 39 (2): 201–13. doi:10.1007/s10867-013-9315-y.
- Leiman, Petr G, Fumio Arisaka, Mark J van Raaij, Victor A Kostyuchenko, Anastasia A Aksyuk, Shuji Kanamaru, and Michael G Rossmann. 2010. "Morphogenesis of the T4 Tail and Tail Fibers." *Virology Journal* 7 (1): 355. doi:10.1186/1743-422X-7-355.
- Leiman, Petr G, Anthony J Battisti, Valorie D Bowman, Katharina Stummeyer, Martina Mühlenhoff, Rita Gerardy-Schahn, Dean Scholl, and Ian J Molineux. 2007. "The Structures of Bacteriophages K1E and K1-5 Explain Processive Degradation of Polysaccharide Capsules and Evolution of New Host Specificities." *Journal of Molecular Biology* 371 (3): 836–49. doi:10.1016/j.jmb.2007.05.083.

- Leiman, Petr G, Paul R Chipman, Victor A Kostyuchenko, Vadim V Mesyanzhinov, and Michael G Rossmann. 2004. "Three-Dimensional Rearrangement of Proteins in the Tail of Bacteriophage T4 on Infection of Its Host." *Cell* 118 (4): 419–29. doi:10.1016/j.cell.2004.07.022.
- Leiman, Petr G., and Ian J. Molineux. 2008. "Evolution of a New Enzyme Activity from the Same Motif Fold." *Molecular Microbiology* 69 (2): 287–90. doi:10.1111/j.1365-2958.2008.06241.x.
- Leiman, Petr G, and Mikhail M Shneider. 2012a. "Contractile Tail Machines of Bacteriophages." *Advances in Experimental Medicine and Biology* 726: 93–114. doi:10.1007/978-1-4614-0980-9_5.
- Leiman, Petr G., and Mikhail M. Shneider. 2012b. "Contractile Tail Machines of Bacteriophages." *Advances in Experimental Medicine and Biology* 726: 93–114. doi:10.1007/978-1-4614-0980-9_5.
- Leiman, P. G., S. Kanamaru, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. 2003. "Structure and Morphogenesis of Bacteriophage T4." *Cellular and Molecular Life Sciences: CMLS* 60 (11): 2356–70. doi:10.1007/s00018-003-3072-1.
- Liu, Hongrong, Lei Jin, Sok Boon S. Koh, Ivo Atanasov, Stan Schein, Lily Wu, and Z. Hong Zhou. 2010. "Atomic Structure of Human Adenovirus by Cryo-EM Reveals Interactions among Protein Networks." *Science (New York, N.Y.)* 329 (5995): 1038–43. doi:10.1126/science.1187433.
- Liu, Jun, Cheng-Yen Chen, Daisuke Shiomi, Hironori Niki, and William Margolin. 2011. "Visualization of Bacteriophage P1 Infection by Cryo-Electron Tomography of Tiny Escherichia Coli." *Virology* 417 (2): 304–11. doi:10.1016/j.virol.2011.06.005.
- Liu, Xiangnan, Qinfen Zhang, Kazuyoshi Murata, Matthew L Baker, Matthew B Sullivan, Caroline Fu, Matthew T Dougherty, et al. 2010. "Structural Changes in a Marine Podovirus Associated with Release of Its Genome into Prochlorococcus." *Nature Structural & Molecular Biology* 17 (7): 830–36. doi:10.1038/nsmb.1823.
- Liu, Yu, Pawel A Penczek, Bruce F McEwen, and Joachim Frank. 1995. "A Marker-Free Alignment Method for Electron Tomography." *Ultramicroscopy* 58 (3-4): 393–402. doi:10.1016/0304-3991(95)00006-M.
- Lotz, W., G. Acker, and R. Schmitt. 1977. "Bacteriophage 7-7-1 Adsorbs to the Complex Flagella of Rhizobium Lupini H13-3." *The Journal of General Virology* 34 (1): 9–17.
- Lovett, P. S. 1972. "PBPI: A Flagella Specific Bacteriophage Mediating Transduction in Bacillus Pumilus." *Virology* 47 (3): 743–52.
- Lucić, Vladan, Friedrich Förster, and Wolfgang Baumeister. 2005. "Structural Studies by Electron Tomography: From Cells to Molecules." *Annual Review of Biochemistry* 74: 833–65. doi:10.1146/annurev.biochem.73.011303.074112.
- Ludtke, S J, P R Baldwin, and W Chiu. 1999. "EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions." *Journal of Structural Biology* 128 (1): 82–97. doi:10.1006/jsbi.1999.4174.
- Maniloff, J., and H. W. Ackermann. 1998. "Taxonomy of Bacterial Viruses: Establishment of Tailed Virus Genera and the Order Caudovirales." *Archives of Virology* 143 (10): 2051–63.
- Mastronarde, David. 2003. "SerialEM A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position." *Microscopy and Microanalysis* 9 (Supplement S02): 1182–83. doi:null.
- Mastronarde, D. N. 1997. "Dual-Axis Tomography: An Approach with Alignment Methods That Preserve Resolution." *Journal of Structural Biology* 120 (3): 343–52. doi:10.1006/jsbi.1997.3919.
- Mcguire, E. J., and S. B. Binkley. 1964. "THE STRUCTURE AND CHEMISTRY OF COLOMINIC ACID." *Biochemistry* 3 (February): 247–51.
- Meynell, E. W. 1961. "A Phage, Phi Chi, Which Attacks Motile Bacteria." *Journal of General Microbiology* 25 (June): 253–90.
- Mindell, Joseph A, and Nikolaus Grigorieff. 2003. "Accurate Determination of Local Defocus and Specimen Tilt in Electron Microscopy." *Journal of Structural Biology* 142 (3): 334–47.
- Moak, Michael, and Ian J. Molineux. 2004. "Peptidoglycan Hydrolytic Activities Associated with Bacteriophage Virions." *Molecular Microbiology* 51 (4): 1169–83.
- Molineux, Ian J., and Debabrata Panja. 2013. "Popping the Cork: Mechanisms of Phage Genome Ejection." *Nature Reviews. Microbiology* 11 (3): 194–204. doi:10.1038/nrmicro2988.
- Monod, C., F. Repoila, M. Kutateladze, F. Tétart, and H. M. Krisch. 1997. "The Genome of the Pseudo T-Even Bacteriophages, a Diverse Group That Resembles T4." *Journal of Molecular Biology* 267 (2): 237–49. doi:10.1006/jmbi.1996.0867.
- Moody, Michael F. 1999. "Geometry of Phage Head Construction." *Journal of Molecular Biology* 293 (2): 401–33. doi:10.1006/jmbi.1999.3011.
- Morais, Marc C. 2012. "The dsDNA Packaging Motor in Bacteriophage ø29." *Advances in Experimental Medicine and Biology* 726: 511–47. doi:10.1007/978-1-4614-0980-9_23.
- Morais, Marc C., Kyung H. Choi, Jaya S. Koti, Paul R. Chipman, Dwight L. Anderson, and Michael G. Rossmann. 2005. "Conservation of the Capsid Structure in Tailed dsDNA Bacteriophages: The Pseudoatomic Structure of ø29." *Molecular Cell* 18 (2): 149–59. doi:10.1016/j.molcel.2005.03.013.

- Mühlenhoff, Martina, Katharina Stummeyer, Melanie Grove, Markus Sauerborn, and Rita Gerardy-Schahn. 2003. "Proteolytic Processing and Oligomerization of Bacteriophage-Derived Endosialidases." *The Journal of Biological Chemistry* 278 (15): 12634-44. doi:10.1074/jbc.M212048200.
- Murata, Kazuyoshi, Xiangang Liu, Radostin Danev, Joanita Jakana, Michael F. Schmid, Jonathan King, Kuniaki Nagayama, and Wah Chiu. 2010. "Zernike Phase Contrast Cryo-Electron Microscopy and Tomography for Structure Determination at Nanometer and Subnanometer Resolutions." *Structure (London, England: 1993)* 18 (8): 903-12. doi:10.1016/j.str.2010.06.006.
- Nagayama, Kuniaki. 2014. "Biological Applications of Phase-Contrast Electron Microscopy." *Methods in Molecular Biology (Clifton, N.J.)* 1117: 385-99. doi:10.1007/978-1-62703-776-1_18.
- Navasa, Nicolás, Leandro Rodríguez-Aparicio, Honorina Martínez-Blanco, Mario Arcos, and Miguel Angel Ferrero. 2009. "Temperature Has Reciprocal Effects on Colanic Acid and Polysialic Acid Biosynthesis in E.Coli K92." *Applied Microbiology and Biotechnology* 82 (4): 721-29. doi:10.1007/s00253-008-1840-4.
- Nicastro, Daniela, Cindi Schwartz, Jason Pierson, Richard Gaudette, Mary E Porter, and J Richard McIntosh. 2006. "The Molecular Architecture of Axonemes Revealed by Cryoelectron Tomography." *Science (New York, N.Y.)* 313 (5789): 944-48. doi:10.1126/science.1128618.
- Olia, Adam S, Peter E Prevelige, John E Johnson, and Gino Cingolani. 2011. "Three-Dimensional Structure of a Viral Genome-Delivery Portal Vertex." *Nature Structural & Molecular Biology* 18 (5): 597-603. doi:10.1038/nsmb.2023.
- Oliveira, Leonor, Paulo Tavares, and Juan C. Alonso. 2013. "Headful DNA Packaging: Bacteriophage SPPI as a Model System." *Virus Research* 173 (2): 247-59. doi:10.1016/j.virusres.2013.01.021.
- Orlova, Elena V., Brent Gowen, Anja Dröge, Asita Stiege, Frank Weise, Rudi Lurz, Marin van Heel, and Paulo Tavares. 2003. "Structure of a Viral DNA Gatekeeper at 10 Å Resolution by Cryo-Electron Microscopy." *The EMBO Journal* 22 (6): 1255-62. doi:10.1093/emboj/cdg123.
- Orlova, E. V., and H. R. Saibil. 2011. "Structural Analysis of Macromolecular Assemblies by Electron Microscopy." *Chemical Reviews* 111 (12): 7710-48. doi:10.1021/cr100353t.
- Parent, Kristin N., Reza Khayat, Long H. Tu, Margaret M. Suhanovsky, Juliana R. Cortines, Carolyn M. Teschke, John E. Johnson, and Timothy S. Baker. 2010. "P22 Coat Protein Structures Reveal a Novel Mechanism for Capsid Maturation: Stability without Auxiliary Proteins or Chemical Cross-Links." *Structure (London, England: 1993)* 18 (3): 390-401. doi:10.1016/j.str.2009.12.014.
- Parent, Kristin N., Robert S. Sinkovits, Margaret M. Suhanovsky, Carolyn M. Teschke, Edward H. Egelman, and Timothy S. Baker. 2010. "Cryo-Reconstructions of P22 Polyheads Suggest That Phage Assembly Is Nucleated by Trimeric Interactions among Coat Proteins." *Physical Biology* 7 (4): 045004. doi:10.1088/1478-3975/7/4/045004.
- Penczek, P. A., R. A. Grassucci, and J. Frank. 1994. "The Ribosome at Improved Resolution: New Techniques for Merging and Orientation Refinement in 3D Cryo-Electron Microscopy of Biological Particles." *Ultramicroscopy* 53 (3): 251-70.
- Penczek, Pawel A. 2010a. "Fundamentals of Three-Dimensional Reconstruction from Projections." In *Methods in Enzymology*, 482:1-33. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/S0076687910820014>.
- . 2010b. "Resolution Measures in Molecular Electron Microscopy." *Methods in Enzymology* 482: 73-100. doi:10.1016/S0076-6879(10)82003-8.
- Penczek, Pawel A., Robert Renka, and Hermann Schomberg. 2004. "Gridding-Based Direct Fourier Inversion of the Three-Dimensional Ray Transform." *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 21 (4): 499-509.
- Penczek, Pawel A., Jun Zhu, and Joachim Frank. 1996. "A Common-Lines Based Method for Determining Orientations for N > 3 Particle Projections Simultaneously." *Ultramicroscopy* 63 (3-4): 205-18. doi:10.1016/0304-3991(96)00037-X.
- Petrov, Vasiliy M., James M. Nolan, Claire Bertrand, Dawn Levy, Carine Desplats, H. M. Krisch, and Jim D. Karam. 2006. "Plasticity of the Gene Functions for DNA Replication in the T4-like Phages." *Journal of Molecular Biology* 361 (1): 46-68. doi:10.1016/j.jmb.2006.05.071.
- Pettersen, Eric F, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. 2004. "UCSF Chimera--a Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25 (13): 1605-12. doi:10.1002/jcc.20084.
- Pickard, Derek, Ana Luisa Toribio, Nicola K. Petty, Andries van Tonder, Lu Yu, David Goulding, Bart Barrell, et al. 2010. "A Conserved Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi Capsular Receptor of Salmonella Enterica Serovar Typhi." *Journal of Bacteriology* 192 (21): 5746-54. doi:10.1128/JB.00659-10.
- Pietilä, Maija K., Tatiana A. Demina, Nina S. Atanasova, Hanna M. Oksanen, and Dennis H. Bamford. 2014. "Archaeal Viruses and Bacteriophages: Comparisons and Contrasts." *Trends in Microbiology* 22 (6): 334-44. doi:10.1016/j.tim.2014.02.007.

- Pina, Mery, Ariane Bize, Patrick Forterre, and David Prangishvili. 2011. "The Archeoviruses." *FEMS Microbiology Reviews* 35 (6): 1035–54. doi:10.1111/j.1574-6976.2011.00280.x.
- Prevelige, Peter E., and Bentley A. Fane. 2012. "Building the Machines: Scaffolding Protein Functions during Bacteriophage Morphogenesis." *Advances in Experimental Medicine and Biology* 726: 325–50. doi:10.1007/978-1-4614-0980-9_14.
- Pukatzki, Stefan, Amy T. Ma, Andrew T. Revel, Derek Sturtevant, and John J. Mekalanos. 2007. "Type VI Secretion System Translocates a Phage Tail Spike-like Protein into Target Cells Where It Cross-Links Actin." *Proceedings of the National Academy of Sciences of the United States of America* 104 (39): 15508–13. doi:10.1073/pnas.0706532104.
- Qin, Li, Andrei Fokine, Erin O'Donnell, Venigalla B. Rao, and Michael G. Rossmann. 2010. "Structure of the Small Outer Capsid Protein, Soc: A Clamp for Stabilizing Capsids of T4-like Phages." *Journal of Molecular Biology* 395 (4): 728. doi:10.1016/j.jmb.2009.10.007.
- Quirk, A. V., A. Sletten, and R. C. Hignett. 1976. "Properties of Phage-Receptor Lipopolysaccharide from *Pseudomonas Morsprunorum*." *Journal of General Microbiology* 96 (2): 375–81.
- Raimondo, Linda M., Nancy P. Lundh, and Rafael J. Martinez. 1968. "Primary Adsorption Site of Phage PBS1: The Flagellum of *Bacillus Subtilis*." *Journal of Virology* 2 (3): 256–64.
- Rakhuba, D. V., E. I. Kolomiets, E. Szwajcer Dey, and G. I. Novik. 2010. "Bacteriophage Receptors, Mechanisms of Phage Adsorption and Penetration into Host Cell." *Polish Journal of Microbiology / Polskie Towarzystwo Mikrobiologów = The Polish Society of Microbiologists* 59 (3): 145–55.
- Reimer, Professor Dr Ludwig. 1997. "Specimen Damage by Electron Irradiation." In *Transmission Electron Microscopy*, 463–94. Springer Series in Optical Sciences 36. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-662-14824-2_11.
- Renken, Christian, Chyong-ere Hsieh, Michael Marko, Bimal Rath, Ardean Leith, Terry Wagenknecht, Joachim Frank, and Carmen A. Mannella. 2009. "Structure of Frozen-Hydrated Triad Junctions: A Case Study in Motif Searching inside Tomograms." *Journal of Structural Biology* 165 (2): 53–63. doi:10.1016/j.jsb.2008.09.011.
- Rifat, Dalin, Nathan T. Wright, Kristen M. Varney, David J. Weber, and Lindsay W. Black. 2008. "Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target." *Journal of Molecular Biology* 375 (3): 720–34. doi:10.1016/j.jmb.2007.10.064.
- Rohwer, Forest, David Prangishvili, and Debbie Lindell. 2009. "Roles of Viruses in the Environment." *Environmental Microbiology* 11 (11): 2771–74. doi:10.1111/j.1462-2920.2009.02101.x.
- Rosenthal, Peter B., and Richard Henderson. 2003. "Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy." *Journal of Molecular Biology* 333 (4): 721–45.
- Rossmann, M. G., and J. E. Johnson. 1989. "Icosahedral RNA Virus Structure." *Annual Review of Biochemistry* 58: 533–73. doi:10.1146/annurev.bi.58.070189.002533.
- Rossmann, Michael G., Marc C. Morais, Petr G. Leiman, and Wei Zhang. 2005. "Combining X-Ray Crystallography and Electron Microscopy." *Structure (London, England: 1993)* 13 (3): 355–62. doi:10.1016/j.str.2005.01.005.
- Rossmann, Michael G., and Venigalla B. Rao. 2012. "Viruses: Sophisticated Biological Machines." *Advances in Experimental Medicine and Biology* 726: 1–3. doi:10.1007/978-1-4614-0980-9_1.
- Russ, John C. 2011. *The Image Processing Handbook*. 6th ed. Boca Raton: CRC Press.
- Saad, A., S. J. Ludtke, J. Jakana, F. J. Rixon, H. Tsuruta, and W. Chiu. 2001. "Fourier Amplitude Decay of Electron Cryomicroscopic Images of Single Particles and Effects on Structure Determination." *Journal of Structural Biology* 133 (1): 32–42. doi:10.1006/jsbi.2001.4330.
- Santos, Sílvia B., Andrew M. Kropinski, Pieter-Jan Ceyssens, H.-W. Ackermann, Andre Villegas, Rob Lavigne, Victor N. Krylov, Carla M. Carvalho, Eugénio C. Ferreira, and Joana Azeredo. 2011. "Genomic and Proteomic Characterization of the Broad-Host-Range *Salmonella* Phage PVP-SE1: Creation of a New Phage Genus φ ." *Journal of Virology* 85 (21): 11265–73. doi:10.1128/JVI.01769-10.
- Savalia, Dhruvi, Lars F Westblade, Manisha Goel, Laurence Florens, Priscilla Kemp, Natalja Akulenko, Olga Pavlova, et al. 2008. "Genomic and Proteomic Analysis of ϕ Eco32, a Novel *Escherichia Coli* Bacteriophage." *Journal of Molecular Biology* 377 (3): 774–89. doi:10.1016/j.jmb.2007.12.077.
- Saxton, W.O., W. Baumeister, and M. Hahn. 1984. "Three-Dimensional Reconstruction of Imperfect Two-Dimensional Crystals." *Ultramicroscopy* 13 (1-2): 57–70. doi:10.1016/0304-3991(84)90057-3.
- Saxton, W.O., and J. Frank. 1976. "Motif Detection in Quantum Noise-Limited Electron Micrographs by Cross-Correlation." *Ultramicroscopy* 2 (January): 219–27. doi:10.1016/S0304-3991(76)91385-1.
- Scheres, Sjors H. W., Roberto Melero, Mikel Valle, and Jose-Maria Carazo. 2009. "Averaging of Electron Subtomograms and Random Conical Tilt Reconstructions through Likelihood Optimization." *Structure (London, England: 1993)* 17 (12): 1712–22. doi:10.1016/j.str.2009.09.005.

- 1563-72. doi:10.1016/j.str.2009.10.009.
- Scheres, Sjors H. W., Mikel Valle, Rafael Nuñez, Carlos O. S. Sorzano, Roberto Marabini, Gabor T. Herman, and Jose-Maria Carazo. 2005. "Maximum-Likelihood Multi-Reference Refinement for Electron Microscopy Images." *Journal of Molecular Biology* 348 (1): 139-49. doi:10.1016/j.jmb.2005.02.031.
- Scherzer, O. 1949. "The Theoretical Resolution Limit of the Electron Microscope." *Journal of Applied Physics* 20 (1): 20. doi:10.1063/1.1698233.
- Schmid, Michael F. 2011. "Single-Particle Electron Cryotomography (cryoET)." *Advances in Protein Chemistry and Structural Biology* 82: 37-65. doi:10.1016/B978-0-12-386507-6.00002-6.
- Scholl, Dean, Sankar Adhya, and Carl Merrill. 2005. "Escherichia Coli K1's Capsule Is a Barrier to Bacteriophage T7." *Applied and Environmental Microbiology* 71 (8): 4872-74. doi:10.1128/AEM.71.8.4872-4874.2005.
- Scholl, Dean, and Carl Merrill. 2005. "The Genome of Bacteriophage K1F, a T7-like Phage That Has Acquired the Ability to Replicate on K1 Strains of Escherichia Coli." *Journal of Bacteriology* 187 (24): 8499-8503. doi:10.1128/JB.187.24.8499-8503.2005.
- Schorb, Martin, and John A.G. Briggs. 2014. "Correlated Cryo-Fluorescence and Cryo-Electron Microscopy with High Spatial Precision and Improved Sensitivity." *Ultramicroscopy* 143 (August): 24-32. doi:10.1016/j.ultramic.2013.10.015.
- Schulz, Eike C., Achim Dickmanns, Henning Urlaub, Andreas Schmitt, Martina Mühlenhoff, Katharina Stummeyer, David Schwarzer, Rita Gerardy-Schahn, and Ralf Ficner. 2010. "Crystal Structure of an Intramolecular Chaperone Mediating Triple-Beta-Helix Folding." *Nature Structural & Molecular Biology* 17 (2): 210-15. doi:10.1038/nsmb.1746.
- Schulz, Eike Christian, David Schwarzer, Martin Frank, Katharina Stummeyer, Martina Mühlenhoff, Achim Dickmanns, Rita Gerardy-Schahn, and Ralf Ficner. 2010. "Structural Basis for the Recognition and Cleavage of Polysialic Acid by the Bacteriophage K1F Tailspike Protein EndoNF." *Journal of Molecular Biology* 397 (1): 341-51. doi:10.1016/j.jmb.2010.01.028.
- Schur, Florian K. M., Wim J. H. Hagen, Michaela Rumlová, Tomáš Ruml, Barbara Müller, Hans-Georg Kräusslich, and John A. G. Briggs. 2014. "Structure of the Immature HIV-1 Capsid in Intact Virus Particles at 8.8 Å Resolution." *Nature*, November. doi:10.1038/nature13838.
- Schwarzer, David, Katharina Stummeyer, Rita Gerardy-Schahn, and Martina Mühlenhoff. 2007. "Characterization of a Novel Intramolecular Chaperone Domain Conserved in Endosialidases and Other Bacteriophage Tail Spike and Fiber Proteins." *The Journal of Biological Chemistry* 282 (5): 2821-31. doi:10.1074/jbc.M609543200.
- Schwarzer, D., F. F. R. Buettner, C. Browning, S. Nazarov, W. Rabsch, A. Bethe, A. Oberbeck, et al. 2012. "A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: A Comprehensive Genomic and Structural Analysis." *Journal of Virology* 86 (19): 10384-98. doi:10.1128/JVI.00801-12.
- Sciara, Giuliano, Cecilia Bebeacua, Patrick Bron, Denise Tremblay, Miguel Ortiz-Lombardia, Julie Lichère, Marin van Heel, Valérie Campanacci, Sylvain Moineau, and Christian Cambillau. 2010. "Structure of Lactococcal Phage p2 Baseplate and Its Mechanism of Activation." *Proceedings of the National Academy of Sciences of the United States of America* 107 (15): 6852-57. doi:10.1073/pnas.1000232107.
- Serwer, Philip, Shirley J. Hayes, Julie A. Thomas, and Stephen C. Hardies. 2007. "Propagating the Missing Bacteriophages: A Large Bacteriophage in a New Class." *Virology Journal* 4 (1): 21. doi:10.1186/1743-422X-4-21.
- Shaikh, Tanvir R, Haixiao Gao, William T Baxter, Francisco J Asturias, Nicolas Boisset, Ardean Leith, and Joachim Frank. 2008. "SPIDER Image Processing for Single-Particle Reconstruction of Biological Macromolecules from Electron Micrographs." *Nature Protocols* 3 (12): 1941-74. doi:10.1038/nprot.2008.156.
- Shneider, Mikhail M., Sergey A. Buth, Brian T. Ho, Marek Basler, John J. Mekalanos, and Petr G. Leiman. 2013. "PAAR-Repeat Proteins Sharpen and Diversify the Type VI Secretion System Spike." *Nature* 500 (7462): 350-53. doi:10.1038/nature12453.
- Sigworth, Fred J., Peter C. Doerschuk, Jose-Maria Carazo, and Sjors H. W. Scheres. 2010. "An Introduction to Maximum-Likelihood Methods in Cryo-EM." *Methods in Enzymology* 482: 263-94. doi:10.1016/S0076-6879(10)82011-7.
- Simon, Lee D. 1972. "Infection of Escherichia Coli by T2 and T4 Bacteriophages as Seen in the Electron Microscope: T4 Head Morphogenesis*." *Proceedings of the National Academy of Sciences of the United States of America* 69 (4): 907-11.
- Simpson, A A, Y Tao, P G Leiman, M O Badasso, Y He, P J Jardine, N H Olson, et al. 2000. "Structure of the Bacteriophage phi29 DNA Packaging Motor." *Nature* 408 (6813): 745-50. doi:10.1038/35047129.
- Spence, John C. H. 2013. *High-Resolution Electron Microscopy*. Fourth edition. Oxford: Oxford University Press.
- Steven, Alasdair C., J Bernard Heymann, Naiqian Cheng, Benes L Trus, and James F Conway. 2005. "Virus Maturation: Dynamics and Mechanism of a Stabilizing Structural Transition That Leads to Infectivity." *Current Opinion in*

- Structural Biology* 15 (2): 227–36. doi:10.1016/j.sbi.2005.03.008.
- Stewart, Alex, and Nikolaus Grigorieff. 2004. "Noise Bias in the Refinement of Structures Derived from Single Particles." *Ultramicroscopy* 102 (1): 67–84. doi:10.1016/j.ultramic.2004.08.008.
- Stirm, Stephan, Wolfgang Bessler, Frank Fehmel, and Elisabeth Freund-Mölbart. 1971. "Bacteriophage Particles with Endo-Glycosidase Activity." *Journal of Virology* 8 (3): 343–46.
- Stummeyer, Katharina, Achim Dickmanns, Martina Mühlenhoff, Rita Gerardy-Schahn, and Ralf Ficner. 2005. "Crystal Structure of the Polysialic Acid-Degrading Endosialidase of Bacteriophage K1F." *Nature Structural & Molecular Biology* 12 (1): 90–96. doi:10.1038/nsmb874.
- Stummeyer, Katharina, David Schwarzer, Heike Claus, Ulrich Vogel, Rita Gerardy-Schahn, and Martina Mühlenhoff. 2006. "Evolution of Bacteriophages Infecting Encapsulated Bacteria: Lessons from Escherichia Coli K1-Specific Phages." *Molecular Microbiology* 60 (5): 1123–35. doi:10.1111/j.1365-2958.2006.05173.x.
- Sun, Siyang, Venigalla B. Rao, and Michael G. Rossmann. 2010. "Genome Packaging in Viruses." *Current Opinion in Structural Biology* 20 (1): 114–20. doi:10.1016/j.sbi.2009.12.006.
- Sutter, Markus, Daniel Boehringer, Sascha Gutmann, Susanne Günther, David Prangishvili, Martin J. Loessner, Karl O. Stetter, Eilika Weber-Ban, and Nenad Ban. 2008. "Structural Basis of Enzyme Encapsulation into a Bacterial Nanocompartment." *Nature Structural & Molecular Biology* 15 (9): 939–47. doi:10.1038/nsmb.1473.
- Suttle, Curtis A. 2007. "Marine Viruses — Major Players in the Global Ecosystem." *Nature Reviews Microbiology* 5 (10): 801–12. doi:10.1038/nrmicro1750.
- Tang, Guang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. 2007. "EMAN2: An Extensible Image Processing Suite for Electron Microscopy." *Journal of Structural Biology* 157 (1): 38–46. doi:10.1016/j.jsb.2006.05.009.
- Tavares, Paulo, Sophie Zinn-Justin, and Elena V. Orlova. 2012. "Genome Gating in Tailed Bacteriophage Capsids." *Advances in Experimental Medicine and Biology* 726: 585–600. doi:10.1007/978-1-4614-0980-9_25.
- Taylor, Clare M., and Ian S. Roberts. 2004. "Capsular Polysaccharides and Their Role in Virulence." In *Contributions to Microbiology*, edited by W. Russell and H. Herwald, 12:55–66. Basel: KARGER. <http://www.karger.com/Article/Abstract/81689>.
- Tétart, F., C. Desplats, and H. M. Krisch. 1998. "Genome Plasticity in the Distal Tail Fiber Locus of the T-Even Bacteriophage: Recombination between Conserved Motifs Swaps Adhesin Specificity." *Journal of Molecular Biology* 282 (3): 543–56. doi:10.1006/jmbi.1998.2047.
- Tétart, F., C. Desplats, M. Kutateladze, C. Monod, H. W. Ackermann, and H. M. Krisch. 2001. "Phylogeny of the Major Head and Tail Genes of the Wide-Ranging T4-Type Bacteriophages." *Journal of Bacteriology* 183 (1): 358–66. doi:10.1128/JB.183.1.358-366.2001.
- Thach, Robert E., and Sigrid S. Thach. 1971. "Damage to Biological Samples Caused by the Electron Beam during Electron Microscopy." *Biophysical Journal* 11 (2): 204–10. doi:10.1016/S0006-3495(71)86208-2.
- Thingstad, T. Frede. 2000. "Elements of a Theory for the Mechanisms Controlling Abundance, Diversity, and Biogeochemical Role of Lytic Bacterial Viruses in Aquatic Systems." *Limnology and Oceanography* 45 (6): 1320–28. doi:10.4319/lo.2000.45.6.1320.
- Thuman-Commike, P A, and W Chiu. 2000. "Reconstruction Principles of Icosahedral Virus Structure Determination Using Electron Cryomicroscopy." *Micron (Oxford, England: 1993)* 31 (6): 687–711.
- Toyoshima, Chikashi, and Nigel Unwin. 1988. "Contrast Transfer for Frozen-Hydrated Specimens: Determination from Pairs of Defocused Images." *Ultramicroscopy* 25 (4): 279–91. doi:10.1016/0304-3991(88)90003-4.
- Van Heel, Marin. 1987. "Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3D Reconstruction." *Ultramicroscopy* 21 (2): 111–23. doi:10.1016/0304-3991(87)90078-7.
- Van Heel, Marin, and Michael Schatz. 2005. "Fourier Shell Correlation Threshold Criteria." *Journal of Structural Biology* 151 (3): 250–62. doi:10.1016/j.jsb.2005.05.009.
- Veesler, David, and Christian Cambillau. 2011. "A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries." *Microbiology and Molecular Biology Reviews: MMBR* 75 (3): 423–33. doi:10.1128/MMBR.00014-11.
- Walter, Monika, Christian Fiedler, Renate Grassl, Manfred Biebl, Reinhard Rachel, X. Lois Hermo-Parrado, Antonio L. Llamas-Saiz, Robert Seckler, Stefan Miller, and Mark J. van Raaij. 2008. "Structure of the Receptor-Binding Protein of Bacteriophage Det7: A Podoviral Tail Spike in a Myovirus." *Journal of Virology* 82 (5): 2265–73. doi:10.1128/JVI.01641-07.
- Wasilewski, Sebastian, and Peter B. Rosenthal. 2014. "Web Server for Tilt-Pair Validation of Single Particle Maps from Electron Cryomicroscopy." *Journal of Structural Biology* 186 (1): 122–31. doi:10.1016/j.jsb.2014.02.012.

- White, Helen E., Michael B. Sherman, Sandrine Brasilès, Eric Jacquet, Philippa Seavers, Paulo Tavares, and Elena V. Orlova. 2012. "Capsid Structure and Its Stability at the Late Stages of Bacteriophage SPP1 Assembly." *Journal of Virology* 86 (12): 6768–77. doi:10.1128/JVI.00412-12.
- Whitfield, Chris. 2006. "Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia Coli." *Annual Review of Biochemistry* 75: 39–68. doi:10.1146/annurev.biochem.75.103004.142545.
- Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. "Prokaryotes: The Unseen Majority." *Proceedings of the National Academy of Sciences of the United States of America* 95 (12): 6578–83.
- Wikoff, W. R., L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, and J. E. Johnson. 2000. "Topologically Linked Protein Rings in the Bacteriophage HK97 Capsid." *Science (New York, N.Y.)* 289 (5487): 2129–33.
- Williams, David B., and C. Barry. Carter. 2009. *Transmission Electron Microscopy a Textbook for Materials Science*. New York: Springer. <http://dx.doi.org/10.1007/978-0-387-76501-3>.
- Williams, S. R., D. Gebhart, D. W. Martin, and D. Scholl. 2008. "Retargeting R-Type Pyocins To Generate Novel Bactericidal Protein Complexes." *Applied and Environmental Microbiology* 74 (12): 3868–76. doi:10.1128/AEM.00141-08.
- Wommack, K E, and R R Colwell. 2000. "Virioplankton: Viruses in Aquatic Ecosystems." *Microbiology and Molecular Biology Reviews: MMBR* 64 (1): 69–114.
- Wu, Bin, Alys Peisley, David Tetrault, Zongli Li, Edward H. Egelman, Katharine E. Magor, Thomas Walz, Pawel A. Penczek, and Sun Hur. 2014. "Molecular Imprinting as a Signal-Activation Mechanism of the Viral RNA Sensor RIG-I." *Molecular Cell* 55 (4): 511–23. doi:10.1016/j.molcel.2014.06.010.
- Wu, W., J. A. Thomas, N. Cheng, L. W. Black, and A. C. Steven. 2012. "Bubblegrams Reveal the Inner Body of Bacteriophage KZ." *Science* 335 (6065): 182–182. doi:10.1126/science.1214120.
- Xiang, Ye, Petr G. Leiman, Long Li, Shelley Grimes, Dwight L. Anderson, and Michael G. Rossmann. 2009. "Crystallographic Insights Into the Autocatalytic Assembly Mechanism of a Bacteriophage Tail Spike." *Molecular Cell* 34 (3): 375–86. doi:10.1016/j.molcel.2009.04.009.
- Yang, F., P. Forrer, Z. Dauter, J. F. Conway, N. Cheng, M. E. Cerritelli, A. C. Steven, A. Plückthun, and A. Wlodawer. 2000. "Novel Fold and Capsid-Binding Properties of the Lambda-Phage Display Platform Protein gpD." *Nature Structural Biology* 7 (3): 230–37. doi:10.1038/73347.
- Yonekura, Koji, Saori Maki-Yonekura, and Keiichi Namba. 2003. "Complete Atomic Model of the Bacterial Flagellar Filament by Electron Cryomicroscopy." *Nature* 424 (6949): 643–50. doi:10.1038/nature01830.

LIST OF ABBREVIATIONS

3D three-dimensional

Å Angstrom

bp base pair

cryo-EM electron cryo-microscopy

CTF contrast transfer function

Da Dalton

DNA deoxyribonucleic acid

E.coli Escherichia coli

EM electron microscope/microscopy

eV electron volt

gp gene product

kb thousand of nucleic acid bases

kbp thousand of nucleic acid base pairs

kDa thousand of Daltons

LPS lipopolysaccharide

mM milimole

MW molecular weight

nm nanometer

OB-fold oligonucleotide/oligosaccharide binding fold

PDB Protein Data Bank

phage bacteriophage

RB/RBP receptor binding/receptor binding protein

TEM transmission electron microscope

CURRICULUM VITAE

Sergey U. Nazarov

Avenue Recordon, 16

1004 Lausanne, Switzerland

Phone: +41 78 70 92049

E-mail: **sergey.nazarov@epfl.ch**

Education

Ph.D in Physics

December 2010 - January 2015 - École Polytechnique Fédérale de Lausanne

Laboratory of structural biology and biophysics (Lausanne, Switzerland)

M. Sc. In Applied Mathematics and Physics

2006 - 2008 - Moscow Institute of Physics and Technology

Department of High technologies and Innovations (Moscow, Russia)

B. Sc. In Applied Mathematics and Physics

2002 - 2006 - Moscow Institute of Physics and Technology

Department of Molecular and Biological Physics (Moscow, Russia)

Languages

English - upper-intermediate;

French - upper-beginner;

Spanish - beginner;

Russian - native;

LIST OF PUBLICATIONS

- **A Multivalent Adsorption Apparatus Explains the Broad Host Range of Phage phi92: a Comprehensive Genomic and Structural Analysis.** D. Schwarzer, F. F. R. Buettner, C. Browning, **S. Nazarov**, W. Rabsch, c A. Bethe, A. Oberbeck, V. D. Bowman, K. Stummeyer, M. Mulenhoff, P. G. Leiman, and R. Gerardy-Schahna. Journal of Virology p. 10384-10398, October 2012, Volume 86, Number 19.
- **Structure of bacteriophages ϕ Eco32 and 7-11** – to be submitted.