

**MASARYK UNIVERSITY**  
**Faculty of Science**  
NCBR and CEITEC

# Dissertation

**Brno 2022**

**Dominik Hrebík**



M U N I  
S C I

MASARYK UNIVERSITY  
Faculty of Science  
NCBR and CEITEC

# Characterization of viral cell entry *in vivo* and *ex vivo*

Dissertation

**Dominik Hrebík**

Supervisor: Pavel Plevka, Ph.D.

Brno 2022



# Bibliografický záznam

**Autor:** Dominik Hřebík  
Přírodovědecká fakulta, Masarykova univerzita  
NCBR and CEITEC

**Název práce:** Charakterizace vstupu virů do buňky *in vivo* a *ex vivo*

**Studijní program:** Biochemie

**Studijní obor:** Biomolekulární Chemie

**Vedoucí práce:** Pavel Plevka, Ph.D.

**Akademický rok:** 2021/2022

**Počet stran:** 24 + 92

**Klíčová slova:** Kryo- elektronová mikroskopie; Röntgenová krystalografie; Bakteriofág; Pikornavirus; Enterovirus; Rhinovirus; Vypuštění genomu; Vstup do buňky



# Bibliographic Entry

**Author:** Dominik Hrebík  
Faculty of Science, Masaryk University  
NCBR and CEITEC

**Title of Thesis:** Characterization of viral cell entry *in vivo* and *ex vivo*

**Degree Programme:** Biochemistry

**Field of Study:** Biomolecular Chemistry

**Supervisor:** Pavel Plevka, Ph.D.

**Academic Year:** 2021/2022

**Number of Pages:** 24 + 92

**Keywords:** Cryo-electron microscopy; X-ray crystallography; Bacteriophage; Picornavirus; Enterovirus; Rhinovirus; Genome release; Cell entry



# Abstrakt

Všechny domény života od bakterií až po eukaryota mají své viry. Virus musí najít specifický receptor na buněčném povrchu hostitele, na který se naváže, čímž započne infekce. Následně virus doručí svůj genetický materiál do buňky hostitele. Avšak pro většinu virů je princip tohoto mechanismu na úrovni molekul pořád neznámý. V této dizertační práci se zabývám molekulárními mechanismy vstupu virů do buňky hostitele, které jsem studoval na třech virech – na bakteriofágu P68 infikující buňky *Staphylococcus aureus*, a na lidských virech z rodu *Enterovirus* – rhinoviru 14 a enteroviru 18. Struktura bakteriofága P68 byla vyřešena kombinací metod kryo-elektronové mikroskopie a röntgenové krystalografie. Srovnáním struktur bakteriofága P68 v různých fázích vypuštění genomu jsem navrhl mechanismus doručení genomu fága do buňky. Tyto výsledky pomohou k lepšímu porozumění architektury fágů z čeledě *Podoviridae* a můžou vést k zlepšení fágové terapie proti *S. aureus*.

Komplex rhinoviru 14 a jeho receptoru ICAM-1 byl vyřešen pomocí kryo-EM s rozlišením 2.4 Å. Dokázal jsem, že receptor ICAM-1 spouští molekulární spínač skrytý uvnitř virové kapsidy, který je nezbytný pro vypuštění genomu z viru. Taky jsem identifikoval strukturovanou část RNA genomu, který interaguje s kapsidovým proteinem uvnitř nativního rhinoviru 14. Na základě biochemických experimentů a struktur rhinoviru 14 v různých prostředích jsem navrhl precizně organizovaný mechanismus, který kontroluje načasování vypuštění genomu rhinoviru 14 do lidské buňky. Společně se strukturou echoviru 18, u kterého jsme popsali nový mechanismus vypuštění genomu - virus odklápí jeden nebo víc pentamer z kapsidu - ukazuje kombinace těchto dvou studií kompletní mechanismus od navázání viru až po vypuštění genomu do lidské buňky. Tyto poznatky můžou pomoci při vývoji nových antivirotik cílících na počáteční kroky infekce viru z čeledě *Enterovirus*.



# Abstract

Viruses have evolved to infect all domains of life – from bacteria to eukaryotes. The virus must encounter and bind a specific receptor on the cell's surface to start the infection. Then, the virus delivers its genetic material into the host cytoplasm. However, for most viruses, it is still unknown how the genome release mechanism is controlled at a molecular level. Here I present my structural work on three viruses – bacteriophage P68 infecting *Staphylococcus aureus*, and rhinovirus 14 and echovirus 18 from genus *Enterovirus* infecting humans. The structure of bacteriophage P68 was solved by a combination of cryo-electron microscopy and X-ray crystallography. Based on the comparison of structures of phage P68 in different stages of the genome release, I proposed a mechanism of the cell entry.

The results may lead to improved phage therapy and a better understanding of the assembly of bacteriophages from the family *Podoviridae* infecting gram-positive bacteria. The rhinovirus-14-ICAM-1 structure was solved by cryo-EM to the resolution of 2.4 Å. I proved that the receptor ICAM-1 triggers a molecular switch inside the capsid, necessary for the virus to release its genome. Additionally, a structured part of the RNA genome interacting with capsid protein was identified. I proposed orchestrated steps triggered by binding of ICAM-1, which eventually result in uncoating of the viral genome. Along with the structure of the echovirus 18 expelling pentamers to release its genome, the structural analysis of rhinovirus 14-ICAM-1 complex provides a seamless mechanism of initial stages of infection of *Enteroviruses* from the initial receptor binding to final genome release into the host cytoplasm. Our findings uncovered mechanisms of the initial steps of *Enterovirus* infection that could be targeted by antiviral drugs.

# Acknowledgment

Here I would like to express sincere thanks to my Ph.D. advisor Pavel Plevka who taught me to love science and particularly virology, X-ray crystallography, and cryo-EM. He provided me with generous support when I needed it and tremendous trust when I gained the necessary skills. He was an inspiration to me that hard work, dedication, and persistence pay off. Not always did the projects work as planned; however, he has always been supportive and has always considered my ideas and opinions. Whenever I knocked on his office door for advice or support, I was always welcome. The scientific environment he established in the group is really fascinating and unique to me. The group is highly cooperative and I consider my colleagues as my friends with whom it is so much fun to work and spend my free time. If I ever found myself in the position of a group leader, I would like to lead the group as Pavel is doing it.

The next extremely helpful and important person is my colleague and fiancée at the same time Maria Gondová. She is an extremely nice and loving person who is always ready to help anyone in the group. Whenever I found myself in trouble, she was standing next to me and helped me. We got together at Skog Urban Hub bar discussing the theory of X-ray diffraction and cryo-EM. She is always ready to say something silly or troll me, so we can laugh together. The emotional support and stability she provided me were essential to finish my Ph.D. I also need to mention here that she has three wonderful cats – Cleus, Vivi, and Levi who are really cute and funny. Especially Cleus is inspirational since he has an extremely efficient lifestyle iterating among eating, sleeping, and meowing. We also established a 'Cat Corner' in our office, together with Marta Šiborová, Liya Mukhamedova, and external members Tapu Shaikh and Vojta Dvořák.

Speaking of Marta and Liya, I would like to express my sincere thanks to them because of their emotional and science related support. Whenever I needed any advice or just a good laugh after a long day in the laboratory, they were there. I still remember our long evening/night talks about science, politics, and life in general. With Marta, we won the 11<sup>th</sup> Science slam competition where we demonstrated how bacteriophage penetrates *S.aureus* cell wall. I am really lucky to have them in the office with me.

The next two colleagues I would like to acknowledge here are David and Karel, my PhD fellows who managed to graduate before I did. With David, we started our college at Masaryk University. We finished our Masters at the same time, however in different groups. Then we joined together in the Plevka's laboratory. We mutually helped each other inside and outside of the lab all the time to cope with our experiments, drinking a tremendous amount of coffee, beer and wine, and visiting

traditional folk celebrations in Moravia - 'Hody'. I met Karel for the first time in Pavel's laboratory where he was doing his PhD. He instructed me in cryo-EM SPA reconstructions, and thanks to his support, I managed to solve the whole structure of bacteriophage P68. I believe that we became good friends, especially after the crystallography course in Erice in 2017. We had a great time together, drinking Marsala wine, Grappa, and discussing structural biology. It was kind of symbolic, we both won the prize for the 3rd best posters there. Together with Karel and David we had a trip to Deutsches PanzerMuseum in Munster in 2018. However, I booked hotel in Münster (please note that the only difference is the special ü) which is 300 km away. Thus, we drove extra 300 km to see the Panzers. It was one of the most unforgettable trips I ever experienced. After Karel and David left the lab, I really missed the time when they were around.

I would like to acknowledge another colleague of mine - Tibor Füzik. I must acknowledge his tremendous support since I have joined the lab. He has equipped me with skills both in the wet lab (e.g., cloning and protein purifications) and cryo-EM data processing. Anytime I asked for his advice or opinion, he was ready to help, although, he is an extremely busy person. His contributions in cryo-EM data collection and 3D reconstructions are priceless. He wrote many extremely useful scripts for reconstructing and manipulating cryo-EM data, including a script for localized reconstruction of subparticles that was crucial for the determination of the structure of bacteriophage P68. I am certain that without his contribution, many projects would take either much longer time or would not be finished at all.

I would like to thank the group of experimental biology from Masaryk University, particularly Roman Pantůček, Dana Štveráková, and Martin 'Ben' Benešík. They helped me a lot with the bacteriophage project, especially Dana, who prepared numerous excellent phage samples for cryo-EM.

I would like to thank Mirek Homola, Jan Bíňovský, and Pavol Bárďy, my Ph.D colleagues from our lab for making my life inside and outside of the lab so much fun and exciting. I like our pandemic beers with Mirek and our shady-Saturdays with Pavol.

I would like to thank all the people from whom I received enormous help in the Plevka lab. I would like to particularly highlight here Michaela Procházková, Jana Moravcová, Lenka Šmerdová, and Markéta Londýnová.

I would like to thank the members of cryo-EM facility at CEITEC particularly to Jirka Nováček and Michal Babiak, who helped me whenever I experienced any problems with the microscopes. Jirka patiently answered all kinds of questions regarding microscopes and microscopy, giving me the necessary skills and training to operate cryo-electron microscopes. With Michal we spent a lot of a great time either inside or outside of the microscope facility.

I would like to acknowledge the contribution of Josef 'Pepa' Houser and the crystallization facility for their support and advice with crystallization and biochemical characterization of proteins.

Finally, I would like to thank my parents, who supported me financially and emotionally during my studies. I also greatly thank them for letting me decide my own career path and their unconditional support whatever I have decided to do.



# Declaration

I hereby declare that I am the sole author of this thesis. I declare that, to the best of my knowledge, my thesis does not contain any other material from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

Brno February 7, 2022

.....  
Dominik Hřebík



## **Dominik vs. Dissertation**

I ain't gonna eat, I ain't gonna sleep  
Ain't gonna breathe 'til I see what I wanna see  
And what I wanna see is the finished dissertation  
Permanently; I am just being hurt  
Just ain't gonna work for me, it just wouldn't be.

Writing fast enough 'cause deadlines approaching  
I write as long as I breathe, I don't ever see  
Either of us comin' to terms where we can agree  
There ain't gonna be no reasonin' speakin' with me

More experiments to be finished, final figures on  
So we gonna beef, and keep on beefin' unless  
You're gonna agree to meet with me in the flesh  
And settle this face-to-face, so I can submit!  
*(Inspired by Eminem - Go To Sleep)*



*To my Parents and Maria*



# Contents

<b>1.</b>	<b>Introduction</b>	<b>1</b>
	<b>Introduction</b>	<b>1</b>
1.1	Structural Biology	1
1.2	Transmission Electron Microscopy	1
1.2.1	Working principle of TEM	3
1.2.2	Imaging with electrons - image formation and diffraction	3
1.2.3	Contrast and Contrast Transfer Function	5
1.3	Single Particle Analysis	6
1.3.1	SPA pipeline	7
1.3.2	3D reconstruction	8
1.3.3	Localized reconstruction of subparticles	9
1.4	X-ray crystallography	10
1.5	Viruses	19
1.5.1	Structural Biology of Viruses	19
1.5.2	Bacteriophages	19
1.5.3	Eucaryotic viruses – the genus of <i>Enterovirus</i>	24
1.5.4	Structure of <i>Enteroviruses</i>	24
1.5.5	The life cycle of <i>Enterovirus</i>	25
1.5.6	Rhinoviruses	26
1.5.7	Intercellular adhesion molecule 1	27
<b>2.</b>	<b>Materials and Methods</b>	<b>29</b>
2.1	Materials and Methods	29
<b>3.</b>	<b>Results and Discussion</b>	<b>31</b>
3.1	Published Articles	31
	<b>Summary</b>	<b>77</b>
	<b>Quo vadis cryo-EM?</b>	<b>79</b>



# List of Figures

1.1	Schematic representation of a Transmission Electron Microscope. . . . .	4
1.2	Electron induced damage . . . . .	5
1.3	Constructive and destructive diffraction interference . . . . .	6
1.4	Contrast Transfer Function . . . . .	13
1.5	Cryo-EM micrograph and effect of CTF on particles . . . . .	14
1.6	Fourier slice theorem . . . . .	15
1.7	Single particle analysis algorithm . . . . .	16
1.8	Localized reconstruction of subparticles . . . . .	17
1.9	Scheme of a X-ray Diffraction Experiment . . . . .	17
1.10	Bragg's law . . . . .	18
1.11	Structural overview of the order of <i>Caudovirales</i> . . . . .	20
1.12	Genome delivery mechanism by tailed phage families from the order of <i>Caudovirales</i> . . . . .	22
1.13	Bacteriophage life cycle . . . . .	23
1.14	Classification of <i>Enteroviruses</i> . . . . .	24
1.15	Structural arrangement of <i>Enteroviruses</i> . . . . .	25
1.16	Life cycle of an <i>Enterovirus</i> . . . . .	27
3.1	Bacteriophage P68 'Arstotzka' (minor capsid) protein gp21. . . . .	34



# List of Publications

- (1) **D. Hřebík**, T. Füzik, M. Gondová, L. Šmerdová, A. Adamopoulos, O. Šedo, Z. Zdráhal, and P. Plevka. ICAM-1 induced re-arrangements of capsid and genome prime rhinovirus 14 for activation and uncoating. *Proceedings of the National Academy of Sciences of the United States of America*, doi: [10.1073/pnas.2024251118](https://doi.org/10.1073/pnas.2024251118), 2021.
- (2) **D. Hřebík**, D.Štveráková, K. Škubník, T. Füzik, R. Pantůček, and P. Plevka. Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68. *Science Advances*, doi: [10.1126/sciadv.aaw7414](https://doi.org/10.1126/sciadv.aaw7414), 2019.
- (3) V. Lorin, I. Fernández, G. Masse-Ranson, M. Bouvin-Pley, L.M. Molinos-Albert, C. Planchais, T. Hieu, G. Péhau-Arnaudet, **D. Hřebík** et al., F.A. Rey, and H. Mouquet. Epitope Convergence of a Trio of Broadly HIV-1 Neutralizing IgA and IgG Antibody Lineages in a Viremic Controller. *Journal of Experimental Medicine (In press)*, 2021.
- (4) D. Buchta, Y. Levdansky, T. Füzik, L. Mukhamedova, J. Moravcová, **D. Hřebík**, T. Tollefsrud, J. Andersen, and P. Plevka. Binding of neonatal Fc receptor traps echovirus 18 in intermediate state of pocket factor release. *Journal of Virology (under review)*, 2021.
- (5) L. Mukhamedova, T. Füzik, J. Nováček, **D. Hřebík**, A. Přidal, G.A. Marti, M.A. Diego, and P. Plevka. Virion structure and *in vitro* genome release mechanism of dicistrovirus Kashmir bee virus. *Journal of Virology*, doi: [10.1128/JVI.01950-20](https://doi.org/10.1128/JVI.01950-20), 2021.
- (6) P. Bárdy, T. Füzik, **D. Hřebík**, R. Pantůček, Thomas Beatty, J., and P. Plevka. Structure and mechanism of DNA delivery of a gene transfer agent. *Nature Communications*, doi: [10.1038/s41467-020-16669-9](https://doi.org/10.1038/s41467-020-16669-9), 2020.
- (7) J. Kotouček, F. Hubatka, J. Mašek, P. Kulich, K. Velínská, J. Bezděková, M. Fojtíková, E. Bartheldyová, A. Tomečková, J. Stráská, **D. Hřebík** et al., and J. Turánek Preparation of nanoliposomes by microfluidic mixing in herring-bone channel and the role of membrane fluidity in liposomes formation. *Scientific Reports*, doi: [10.1038/s41598-020-62500-2](https://doi.org/10.1038/s41598-020-62500-2), 2020.
- (8) P. Šimečková, F. Hubatka, J. Kotouček, P. T. Knötigová, J. Mašek, J. Slavík, O. Kováč, J. Neča, P. Kulich, **D. Hřebík** et al., and J. Turánek. Gadolinium labelled nanoliposomes as the platform for MRI theranostics: *in vitro* safety study in liver cells and macrophages. *Scientific Reports*, doi: [10.1038/s41598-020-60284-z](https://doi.org/10.1038/s41598-020-60284-z), 2020.
- (9) D. Buchta, T. Füzik, **D. Hřebík**, Y. Levdansky, L. Súkeník, L. Mukhamedova, J. Moravcová, R. Vácha, and P. Plevka. Enterovirus particles expel capsid pentamers to enable genome release. *Nature Communications*, doi: [10.1038/s41467-019-09132-x](https://doi.org/10.1038/s41467-019-09132-x), 2019.
- (10) P. Bárdy\*, **D. Hřebík\***, R. Pantůček, and P. Plevka. Future prospects of structural studies to advance our understanding of phage biology. *Microbiology Australia*, doi: [10.1071/MA19009](https://doi.org/10.1071/MA19009), 2019. \* both authors contributed equally
- (11) E. Bartheldyová, R. Effenberg, J. Mašek, L. Procházka, P. T. Knötigová, P. Kulich, F. Hubatka, K. Velínská, J. Zelníčková, D. Zouharová, M. Fojtíková, **D. Hřebík** et al., and J. Turánek Hyaluronic Acid Surface Modified Liposomes Prepared via Orthogonal Aminoxy Coupling: Synthesis of Nontoxic Aminoxylipids Based on Symmetrically Branched Fatty Acids, Preparation of Liposomes by Microfluidic Mixing, and Targeting to Cancer Cells Expressing CD44. *Bioconjugate Chemistry*, doi: [10.1038/s41598-020-60284-z](https://doi.org/10.1038/s41598-020-60284-z), 2018.



# Chapter 1

## Introduction

### 1.1 Structural Biology

When you ask a structural biologist why the structures of biomolecules are important, they will probably mention their field's unofficial motto: "Structure is the function". Most regular people cannot imagine what does that mean. Here I will try to explain it with an example. Imagine a car. You can see that the car is moving forwards and backwards; it can turn right and left, it can even break. However, this is only a macroscopic view of the car. You can see the car as a big moving box. To figure out how and why the car is moving, you need to open the car's hood and look into the engine. There you will find cylinders, o-rings, pipes, etc. With that information, you can figure out the working principle of an engine and thus why the car is moving. Now we will move to the nanoscopic world of molecular machines. Probably the most famous molecular machine is a ribosome. Before the structure of the ribosome was determined, it was known that it translates information from mRNA to protein sequence [1]. However, the working principle of the ribosome remained unknown. To figure it out, decades of hard work were dedicated to the determination of ribosome structure. Ada E. Yonath, Venkatraman Ramakrishnan, and Thomas A. Steitz shared the 2009 Nobel Prize in Chemistry for the solution of the ribosome [2]. At that time, X-ray crystallography dominated the field of structural biology. Nowadays, the ribosome structure can be solved within a day using cryo-electron microscopy. This has fulfilled a 'dream' method of Richard Feynman, who once famously said, "It is very easy to answer many fundamental biological questions; you just look at the thing!" [3].

### 1.2 Transmission Electron Microscopy

The working principle of TEM is very similar to that of the light microscope. However, instead of a source of visible light, the electron microscope is equipped with a source of electrons. The rationale behind using electrons is their short wavelength which can be up to 100,000× shorter than that of the visible light [4]. The resolution of imaging instruments, including microscopes, is limited by the wavelength the instrument uses for imaging and the size of the objective lens aperture. Thus, without

using super-resolution techniques, it is not possible to resolve two points located to each other closer than half the wavelength of the imaging light [5, 6]. Conventional light microscopes use wavelengths ranging from 400-700 nm [6, 7]. Thus, the resolution limit of the visible light microscopy is somewhere around 200 nm, which is not sufficient for resolving individual atoms [7]. To beat the resolution limit, one needs to use illumination with a shorter wavelength. One of such probes is electrons [4]. The wavelength of an electron can be calculated thanks to Louis de Broglie, who showed that every particle or matter propagates like a wave [4, 8].

$$\lambda = \frac{h}{p} \quad (1.1)$$

where  $\lambda$  is the wavelength of the particle,  $h$  is Planck's constant, and  $p$  is the momentum of the particle. Since the momentum of the particle is the product of mass and the velocity of the particle, we can write the equation as

$$\lambda = \frac{h}{mv} \quad (1.2)$$

The relationship between the accelerating voltage of an electron microscope and the velocity of an electron is defined as

$$eV = \frac{1}{2}mv^2 \quad (1.3)$$

Where  $e$  is charge of an electron and  $V$  is the accelerating voltage. From which the velocity of an electron can be calculated

$$v = \sqrt{\frac{2eV}{m}} \quad (1.4)$$

When we combine it with the de Broglie equation, we get the wavelength of a propagating electron at a given accelerating voltage

$$\lambda = \frac{h}{\sqrt{2meV}} \quad (1.5)$$

Now we can calculate the wavelength of an electron because we know that the mass of an electron is  $9.1 \times 10^{-31}$ kg and  $e = 1.6 \times 10^{-19}$  C (Coulomb)

$$\lambda = \frac{6.62 \times 10^{-34}[m^2kgs^{-1}]}{\sqrt{2 \times 9.1 \times 10^{-31}[kg] \times 1.6 \times 10^{-19}[C] \times V[V]}} = \frac{12.25 \times 10^{-10}}{\sqrt{V}} \quad (1.6)$$

Therefore the wavelength of an electron accelerated by 300 kV microscope is 2.24 pm.

Due to the fact, the electrons accelerated by a 300 kV electron microscope reach velocities around 78% of the speed of light, relativistic effects on the electrons must be considered. They affect length contraction, time dilation, and an increase in mass of the electrons used for imaging. After applying the relativistic corrections, the equation has the form of

$$\lambda = \frac{12.25 \times 10^{-10}}{\sqrt{V}} \times \frac{1}{\sqrt{1 + \frac{ev}{2mc^2}}} \quad (1.7)$$

where  $c$  is the speed of light which is  $3 \times 10^8 \text{ms}^{-1}$ . Thus, the wavelength of the electrons accelerated by a 300 kV electron microscope is 1.96 pm. Such a wavelength is short enough to resolve individual atoms. However, microscope imperfections, sample quality, properties of electron detectors etc., are currently limiting the resolution of reconstructions of biological objects to 1.2 Å, which is on the edge of atomic resolution [9, 10].

### 1.2.1 Working principle of TEM

The illumination system of the TEM consists of a source of electrons, double deflection coils, apertures, two or three condenser lenses, specimen holder, objective lens, diffraction lens, projector lenses, and detector (Fig. 1.1) [4]. Every lens or group of lenses has a specific purpose. For example, the condenser system maintains parallel illumination over the specimen. The objective lens is the most important lens for high-resolution imaging because it is positioned just below the specimen holder and focuses the diffracted electrons coming from the sample onto the 'first image plane'. The focused electrons then reach the projector system which forms a magnified image of the first image plane on the detector [4, 11].

### 1.2.2 Imaging with electrons - image formation and diffraction

Unlike X-rays, electrons interact more strongly and deposit less energy per information unit to matter which makes them excellent probes for studying atomic structures of biomolecules [12]. There are several ways electrons can interact with a sample. The most important interaction event for gaining structural information is scattering [11]. Scattering can be either elastic or inelastic, whereby the interaction is energetically lossless, or part of the energy is deposited to the sample, respectively [11]. In conventional diffraction methods only elastic scattering provides meaningful information, while inelastic scattering contributes to sample damage, noise, and ultimately dampens the maximum achievable resolution [13]. Such radiation damage is observable for example on electron micrographs of large biological systems such as bacteriophages (Fig. 1.2).

A TEM image is a result of wave interference. The interference can be either constructive when the interfering waves are in phase leading to signal amplification or destructive when the waves are out of phase, leading to a signal dampening (Fig. 1.3) [11]. The conditions that need to be met to have constructive interference are described by Fraunhofer grating formula

$$d \sin \theta = n\lambda \quad (1.8)$$

where  $d$  is the spacing between diffracting planes,  $\theta$  is the diffracting angle in respect to the source, and  $\lambda$  is a wavelength of the diffracting light. As a consequence

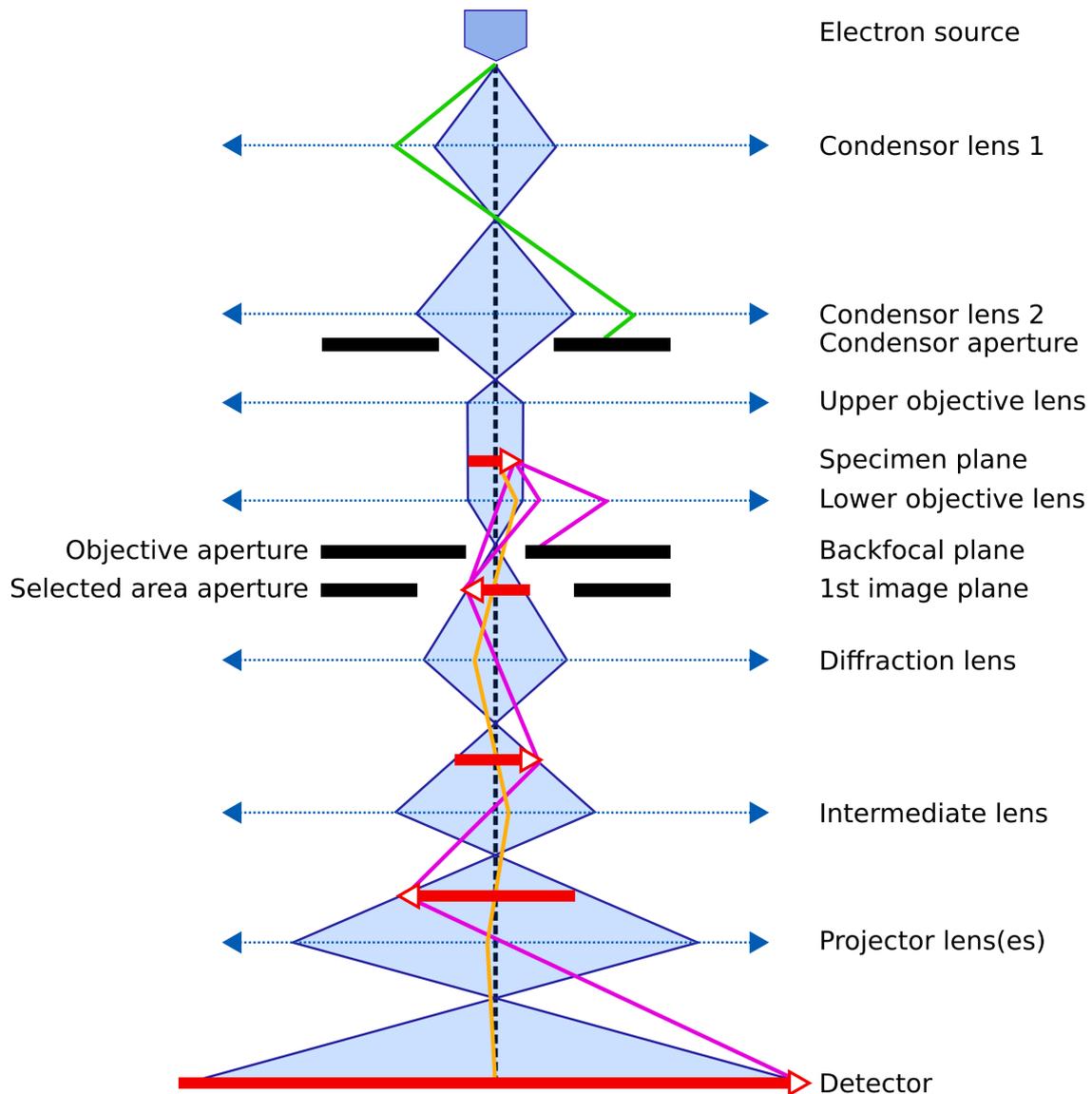


Figure 1.1: **Schematic representation of a Transmission Electron Microscope.** The electron source is producing electrons which are collected by condenser lens 1 (blue and green beam). Subsequently, the electrons are focused by condenser lens 2, while some of them are blocked by condenser aperture (green beam). The upper objective lens makes sure that the beam hitting a sample positioned at the specimen plane is parallel. The sample then diffracts the incoming electron beam, producing electrons carrying information at high and low of frequencies. The diffracted electrons are collected by the lower objective lens. Electrons diffracting at very high frequencies (one of the purple beams) are blocked by the objective aperture positioned at the back-focal plane. The beam then passes through a series of lenses from which the projector lens system is mostly contributing to the magnification. Finally, a magnified image is formed at the detector plane. (Inspired by [11])

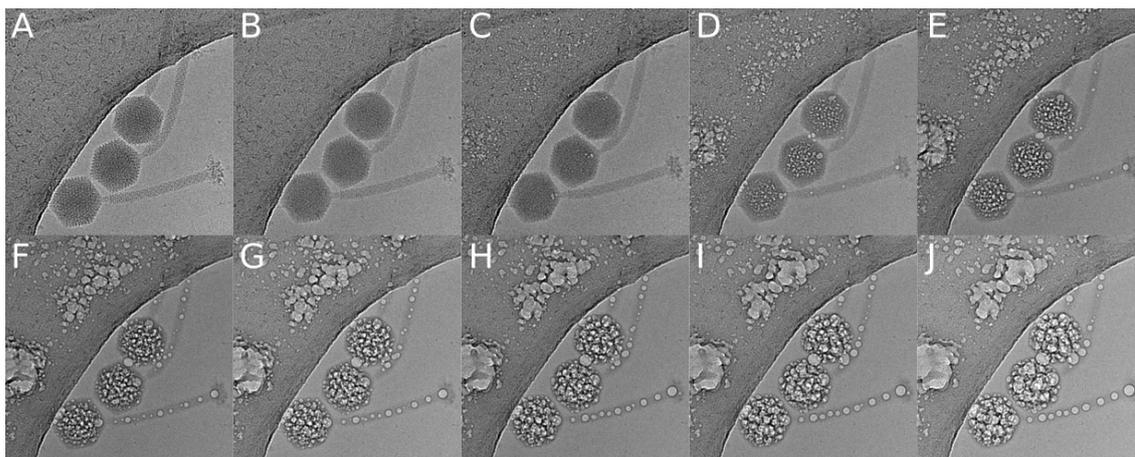


Figure 1.2: **Demonstration of damage caused by electrons to vitrified sample of a bacteriophage.** Panel (A) was exposed to  $50 \text{ e}/\text{\AA}^2$ . Each subsequent panel is exposed with the same electron dose, resulting in total accumulated dose of  $500 \text{ e}/\text{\AA}^2$  in panel (J). Hydrogen bubbles start to appear around  $150 \text{ e}/\text{\AA}^2$  (C).

of the equation, a positive interference occurs only for some scattering angles in grid spacing ( $d$ ). The grid spacing  $d$  defines the obtainable level of information (resolution) that comes from the diffraction – smaller  $d$  carries finer details about the sample. Based on the equation, to fulfill the positive interference condition, the angle of scattering  $\theta$  must be larger for smaller ( $d$ ). Therefore, in the back-focal plane, where the diffraction occurs, the information is sorted in a reciprocal way – small ( $d$ ) representing the highest level of detail will be at the edges of the diffraction pattern (high frequency), while large ( $d$ ) will be close to the center of the diffraction image. Such space is referred to as a reciprocal represented by  $(1/d)$  distribution of frequencies. In other words, the back-focal plane represents the Fourier transform of the sample image. A recorded diffraction pattern does not carry a structural phase information because the primary and scattered waves are separated. However, in the image plane (real space) the scattered and primary waves are recombined, resulting in wave interference.

### 1.2.3 Contrast and Contrast Transfer Function

Biological molecules such as proteins are mainly composed of low atomic number atoms, not much different from those of water in which they are diluted. This makes the amplitude contrast in TEM very low [14]. With the objective lens in focus, the waves that did not interact with the sample are recombined with the scattered waves. However, the amplitudes of the scattered waves are much smaller than of the unscattered ones, and also the phase difference between the scattered and unscattered beams is only  $0.25 \lambda$ , so there is very limited interference among the scattered and unscattered components and thus low contrast [11]. Therefore, a different method of contrast-enhancement in TEM is needed.

The most widely used contrast enhancing method in TEM is the utilization of phase

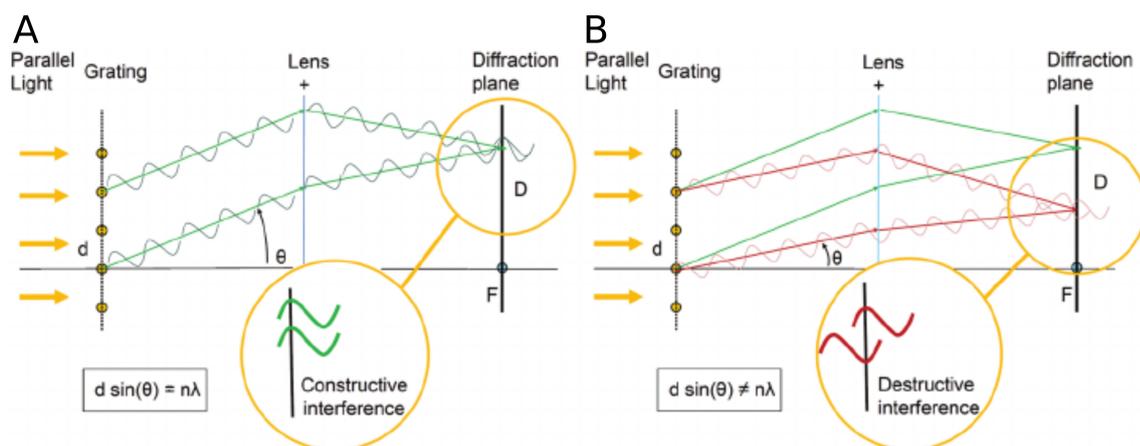


Figure 1.3: **Constructive (A) and destructive (B) interference.** The parallel illumination is diffracted at the grating (sample plane) and the scattered waves are focused by the lens at the diffraction plane (back-focal plane) of the lens. Since the condition  $n\lambda$  from equation 1.8 is fulfilled the waves in (A) are in phase, which lead to constructive interference (addition of intensities) at the diffraction plane. On the other hand, in (B) the scattering angle  $\theta$  does not suffice the condition in equation 1.8 and therefore destructive interference is observed in the diffraction plane. (Adapted from [11])

contrast. The phase contrast is introduced by defocusing the objective lens (Fig. 1.4). This moves the focal point of the objective lens from the specimen plane, usually by  $0.5\ \mu\text{m}$  to  $3.0\ \mu\text{m}$  below the sample [11]. As a consequence, a phase-shift is introduced at the back focal plane of the objective lens. In other words, phases of the scattered waves are gradually shifted with increasing frequency, creating frequency dependent interference [11]. The interference among the defocused waves forms a typical oscillatory function that modifies the amplitudes of waves forming a cryo-EM micrograph – called Contrast Transfer Function (CTF). The shape of the CTF depends on several parameters where the most important is the above-mentioned defocus. Since the function is oscillatory, it reaches zero values for some frequencies which results in a local loss of information [11]. This problem can be overcome by collecting several micrographs, each with a slightly different defocuses, which, when combined in a 3D reconstruction fills the missing spectral information [15].

### 1.3 Single Particle Analysis

'Any sufficiently advanced technology is indistinguishable from magic' – the Arthur C. Clarke's third law [16] can be easily applied to single particle analysis [15]. By using dozen to millions of noisy black-and-white images which are processed on GPU cards, one can obtain the goal of every structural biologist – a 3D density map of the molecule of interest. The possibility of going from noisy 2D images as depicted in (Fig. 1.5) to high-resolution 3D map is enabled by two phenomena – i) the conformation variability of rapidly frozen biomolecules is limited so that devi-

ation of the position between corresponding atoms, after superposition is within a few Ångstroms [15] ii) even the noisy images contain enough information for determination of the orientation of the particle [15]. The first phenomenon is necessary because biomolecules are extremely vulnerable to electron induced damage (Fig. 1.2) which limits the maximum possible electron dose that can be used for imaging [17]. Consequently, the signal-to-noise ratio (SNR) in low-dose electron micrographs is low (Fig. 1.5). To increase the SNR, averaging particles is necessary [15]. However, averaging the signal without prior orienting the particles to superimpose them would be useless. Therefore, the second phenomenon is very important as well.

Nevertheless, before the final 3D reconstruction, several preprocessing steps are necessary. We will describe them in the next section.

### 1.3.1 SPA pipeline

**Motion correction** The raw data produced by a cryo-electron microscope are in the format of so-called 'movies'. The micrograph is not acquired as a single image, but it is composed of dozens of frames, each exposed only to a few electrons per Å<sup>2</sup> [18]. Then by aligning and averaging the frames, a partial correction of the motion blur is possible. Before the introduction of direct electron detectors (DED), the beam induced motion [19] combined with low detector quantum efficiency of film and charge-coupled device (CCD) detector was the main reason which prevented achieving high-resolution reconstructions [20]. Several programs were developed to correct the beam induced motion, e.g., Unblur, MotionCorr2, FlexAlign, cryoSPARC, Warp [21, 22, 23, 24, 25].

**CTF determination and CTF correction** A CTF parameters estimation is performed after the motion correction. The first step in CTF estimation is calculating the Fourier transform of a micrograph. Such Fourier transform of a micrograph has visible rings called 'Thon rings' corresponding to the CTF function oscillations modifying the reciprocal space pixel values. The positions of Thon rings are used for fitting the CTF function (Fig. 1.4) [26]. When the CTF is fitted, a precise defocus value of the micrograph can be determined. After estimation of the CTF, the CTF correction is performed. Since the CTF has several zero values in the Fourier space, a complete inversion is impossible [15]. The most popular and the simplest correction is phase flipping. Here, all negative values of the CTF function are flipped to positive – correct phases are obtained in the Fourier space. This reverts the delocalized signal in the real space to its original position as visible in (Fig. 1.4) [15]. Software used for CTF estimation and correction are e.g., CTFfind4, CTER, and gCTF [27, 26, 28].

**Particle picking** The next step after CTF correction is particle picking and extraction of particles. Micrographs consist of useful areas containing particles and areas containing noise or contaminants, which should be avoided. The goal of particle picking is to detect positions of particles of interest in the micrograph and avoid contaminants and pure noise [15]. Several approaches were developed, from

the most primitive and tedious manual selection [29] to automatic template matching [30]. However, the most promising are algorithms based on convolutional neural networks, which are fast and accurate. Software utilizing the neural networks are for example, EMAN2, crYOLO, and TOPAZ [31, 32, 33]. When the positions of particles in the micrographs are known, they can be extracted. The extraction step involves separation of the particle from the micrograph with a sufficiently large box-size as depicted in (Fig. 1.5). Since the picking of particles is not perfect, the next steps involve classification and separation of 'good' particles from the junk.

**2D classification** 2D classification serves mostly for negative selection of contaminants and damaged particles. It also gives a good impression of heterogeneity and quality of the dataset. The principle of 2D classification relies on averaging of similarly oriented particles, which, as a result, display higher SNR. The orientations are determined by in-plane rotations (and 2D translations) of the particle images [15]. The most popular algorithms of particle classifications are principal component analysis [34] and maximum-likelihood method [30].

**3D classification** To achieve the highest possible resolution, a dataset must be as homogeneous as possible. Since every population of particles displays some degree of heterogeneity a more sensitive classification than 2D classification is necessary. Such subtle differences can be detected and separated by 3D classification. The 3D classification uses multiple 3D initial models each representing one subset (class) of particles. The number of classes is defined by the user [30]. The particles are reconstructed by iterative SPA refinement, and after a few iterations, each particle is classified into one of the specified 3D classes. Afterward, the user inspects all of the 3D classes and group particles from similarly-looking classes.

### 1.3.2 3D reconstruction

The final step is obtaining a 3D cryo-EM map from 2D projections of the same molecule. A 2D projection can be computed from a 3D map  $V$  which is composed of  $n \times n \times n$  voxels where each voxel has some value. If you draw a line through the 3D map, then sum all voxel values the line passes through, you will obtain a pixel value. If you do this for all pixels in the 2D image, you get a projection image of the given 3D map as demonstrated in (Fig. 1.6) [15]. The 3D reconstruction is based on such projection images of particles obtained from a microscope. The most instructive 3D reconstruction algorithm is based on the Fourier slice theorem. It states the relationship between a 3D map and 2D projection images of the map – a central slice (passing through the origin) of the Fourier transform of a 3D map is identical to the 2D Fourier transform of a projection image in projection direction  $p$  being orthogonal to the slice [15].

The basic algorithm works as follows:

1. A large amount of homogeneous particles (projection images) is selected from 3D classification.
2. Calculate their Fourier transforms.

3. In the 3D reciprocal space, orient the 2D Fourier transformed particles perpendicularly to their projection direction in the real space.
4. Assign a value (intensity and phase) to each voxel in the 3D reciprocal space. The value should correspond to values of the neighboring pixels of the 2D FT particles. This results in a new 3D Fourier volume  $V_2$ .
5. Calculate the inverse Fourier transform of the  $V_2$ .

Those are the basics of the 3D reconstruction in a nutshell. The most challenging task in reconstruction is the determination of particle orientations which is described in detail in the next paragraph.

**Determination of particle orientations** After sample preparation, the particles are oriented randomly in the ice, and their relative orientations (rotations) are unknown. Therefore, for each particle, three Euler angles describing its orientation need to be determined prior to the final 3D reconstruction [15]. Without precise determination of the orientations, the 3D reconstruction would be impossible because correct Fourier values could not be assigned to the corresponding voxels in the final 3D reconstruction.

The generally applied method is based on matching a set of computationally generated projections of a reference 3D map (also called "initial model") to each particle image Fig. 1.7 [35]. The reference 3D map can be a known related structure, or *de novo* generated 3D volume from a subset of particles by stochastic gradient descent algorithm. When each particle has assigned three Euler angles based on the best matching projection, then all particles are fast-Fourier transformed (FFT). Afterward, the FFT particles are combined together in Fourier space to fill the information within a 3D Fourier space. Finally, inverse Fourier transform is calculated from that 3D Fourier space, which results in a 3D real space map of the molecule. The map is then used for the generation of another set of projections to obtain even more precise orientations, and the whole process is repeated until convergence [15]. The whole process is shown in (Fig. 1.7).

### 1.3.3 Localized reconstruction of subparticles

In the previous sections, we described how SPA and 3D reconstruction work. What we omitted is considering the symmetry of the object and of the reconstruction. For example, an object with 6-fold rotational symmetry (C6) can be reconstructed with the same symmetry enforced. The presence of symmetry is of benefit since each single projection image represents multiple views of the object and thus increases the number of particles by the symmetry number. However, when parts of the complex follow different symmetries (e.g. 5-fold and 12-fold), then applying one type of symmetry leads to losing information about the part of the complex possessing the other type of symmetry. It is referred to as a "symmetry mismatch problem" [36]. One example is the case of bacteriophage, where capsid has 5-fold symmetry and the tail has 12-fold symmetry as demonstrated in (Fig. 1.8). To handle this problem, several approaches were developed, one of them which is referred to as 'Localized

reconstruction of subparticles' developed by Juha Huiskonen and his group [37]. The method relies on handling the subunits of one molecule possessing different symmetries as two different particles. Taking the bacteriophage as an example, the head is initially extracted and reconstructed using 5-fold symmetry. Then using the knowledge of orientations of the bacteriophages in the micrographs (from the 5-fold symmetry reconstruction of the capsid performed at first) we can draw a vector to extract the second subunit (sub-particle) and solve the sub-particle structure using 12-fold symmetry. Finally, the two cryo-EM maps can be combined into one, solving the whole structure of the bacteriophage P68 [38]. The whole process is described in (Fig. 1.8).

## 1.4 X-ray crystallography

Unlike electrons, X-ray radiation interacts much less with matter [12]. Therefore, with common X-ray sources, the scattering from single particles is too weak to be detected and amplification of the signal is necessary. Crystals serve as natural amplifiers of scattering. They are made of many repeating units – called unit cells – which amplify the scattering of X-rays in certain directions [39]. The basic diffraction experiment is simple – a crystal is exposed to X-ray radiation and a diffraction pattern is recorded. Afterward, the crystal is rotated and diffraction exposed until complete data are collected. The basic setup is described in (Fig. 1.9) [40]. One diffraction image obtained from a single crystal orientation represent a 2D section from the full 3D diffraction pattern. It is a concept similar to the Fourier slice theorem described in (Fig. 1.6) [41]. After hitting the crystal, the X-ray waves interfere with each other either constructively or destructively, as described in (Fig. 1.3). In crystallography, the only detectable signal can be measured from the constructive interference in the forms of diffraction exposed spots that form the diffraction pattern [39]. The conditions that need to be met for constructive interference are described by Bragg's law: [42]

$$2d \sin \theta = n\lambda \tag{1.9}$$

where  $d$  is the spacing between diffracting planes,  $\theta$  is the diffracting angle in respect to the source, and  $\lambda$  is the wavelength of the diffracting light. It is very similar to the Fraunhofer formula described in Eq. 1.8. Bragg's law shows that constructive interference of X-rays occurs only when the path difference between rays scattered from parallel crystal planes is an integer multiple of wavelengths of the radiation. In other words, when the scattering planes are separated by a distance  $d$ , then the path difference must be  $2d \sin \theta$ . Such constructive interference is necessary for amplification of the signal coming from the waves scattered by different molecules [39, 41].

Each X-ray reflection is defined by a structure factor  $\mathbf{F}_{hkl}$  where  $h, k, l$  are coordinates of the spot on the detector. The structure factor is a mathematical description of the diffracted waves coming from the crystal, which when colliding with the detector, deposit their energy and give rise to reflections (Fig. 1.9) [40, 41]. Each scattered

wave is defined by its spatial frequency, amplitude, and phase. Since the frequency of the diffracted wave is defined by its  $h, k, l$  coordinates on the diffraction pattern, the structure factor can be expressed as [39]:

$$\mathbf{F}_{hkl} = |F_{hkl}| \exp(i\alpha_{hkl}) \quad (1.10)$$

in Eq. 1.10, the structure factor  $\mathbf{F}_{hkl}$  is a complex number vector that is composed of two scalar terms – structure factor amplitude  $|F_{hkl}|$  and a phase  $\alpha_{hkl}$ . The  $\mathbf{F}_{hkl}$  represents total wave resulting from diffraction of all atoms in the unit cell. Therefore, it can be expressed as the summation term [39]:

$$\mathbf{F}_{hkl} = \sum_{j=1}^n f_j \exp 2\pi i(hx_j + ky_j + lz_j) \quad (1.11)$$

where the sum is over all atoms in the unit cell from atom  $j = 1$  to  $n^{th}$  atom. Thus, in other words, the  $\mathbf{F}_{hkl}$  is the resultant of all waves scattered in the direction defined by  $hkl$  by all the atoms from the unit cell. The positional coordinates  $x_j, y_j, z_j$  define positions of the  $j^{th}$  atom in the real-space unit cell. The atomic scattering factor  $f_j$  defines the scattering by a given atom type. It is a measure of the X-ray scattering power of each atom, and it is independent of the position of the atom in the unit cell. It depends only on the type of the atom and the direction of scattering. The knowledge of the value of the structure factor is crucial for the determination of the final 3D electron density [41].

From the diffraction experiment as shown in (Fig. 1.9) we obtain only amplitudes  $|F_{hkl}|$  and frequencies of the waves defining the structure factor. The amplitude can be measured from the intensity of the reflection on the detector, and frequency is defined by the  $hkl$  coordinates of the reflection. There is no information about the phase which is crucial for structure determination. Therefore, the phase must be determined indirectly by other methods [39].

**Solving the phase problem** Several methods were developed to solve the phase problem of protein crystallography. The experimental methods require the acquisition of additional X-ray datasets of the same protein, however, with a known change made to the contents of the unit cell. The change involves the binding of a heavy metal ion to the protein without disturbing the conformation of the protein. Thus, the two protein structures are 'isomorphous'. The heavy atoms scatter the X-rays more strongly than N,C,O,H,S,P that form the macromolecules. Differences in structure factor amplitudes enable the calculation of heavy atom positions in the unit cell and subsequent calculation of phases. This method is called Multiple Isomorphous Replacement (MIR) and requires the preparation of several modified crystals each with a different heavy atom [43]. An alternative approach is called Multiwavelength Anomalous Dispersion (MAD) [44]. It is based on breaking Friedel's law which states that Friedel pair reflections have equal amplitudes but opposite phases.

$$|F_{hkl}| = |F_{\bar{h}\bar{k}\bar{l}}| \quad \alpha_{hkl} = -\alpha_{\bar{h}\bar{k}\bar{l}} \quad (1.12)$$

This approach takes advantage of 'anomalous dispersion', which occurs for specific atoms at specific X-ray wavelengths. Each atom type has a few 'absorption edges' in which the scattering is 'anomalous' – the Friedel mate reflections have no longer the same amplitude and phase.

The last and most popular method doesn't require any additional experimental dataset. However, it requires an already known protein structure similar to the structure of the target protein. The method is referred to as 'Molecular Replacement' [45]. The known structure must be initially rotated and translated to match the orientation and position of the protein in the crystal. Then phases of the unknown structure can be estimated using phases of the known protein.

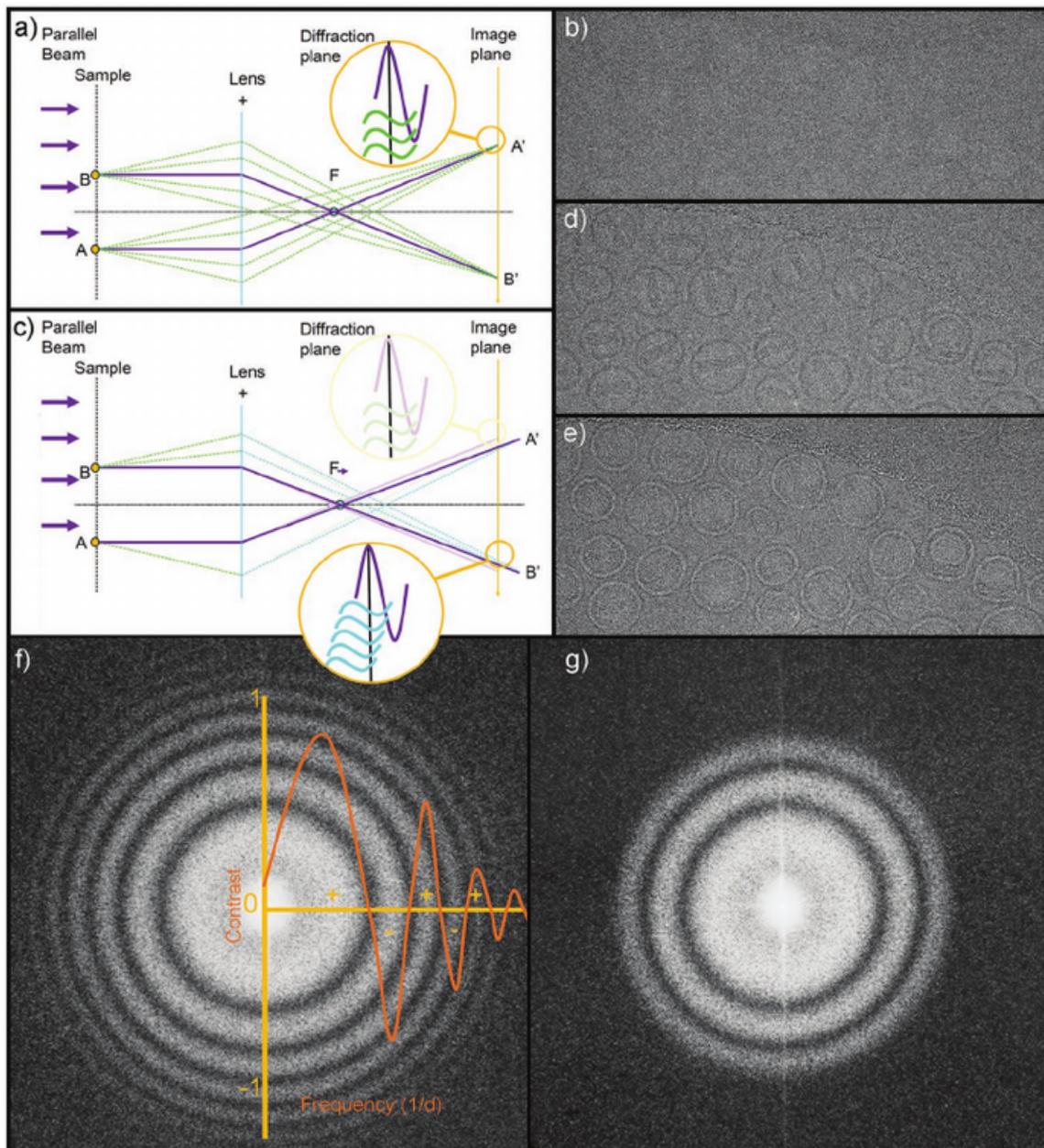


Figure 1.4: **Formation of phase contrast in TEM.** (A) Imaging in focus results in a very low contrast micrograph (B) due to lack of interference between individual waves. The unscattered wave (purple) is recombined with low amplitude scattered waves (green) in the image plane. Changing the strength of the objective lens (C) leads to either underfocused (D) or overfocused (E) micrographs. As a consequence of the defocus, a gradual phase shift is formed among the scattered waves (cyan), which leads to modification of amplitudes in reciprocal space of the micrograph by CTF and thus to phase contrast. (F) Fourier transform (FFT) of a micrograph shows an oscillating function modifying amplitudes in the reciprocal space, demonstrating the effect of the CTF (orange line). When a small objective aperture is introduced (G), it blocks the high-frequency electrons in the back focal plane of the microscope, which can be observed in the FFT of the micrograph as a loss of high-frequency amplitudes. (Adapted from [11])

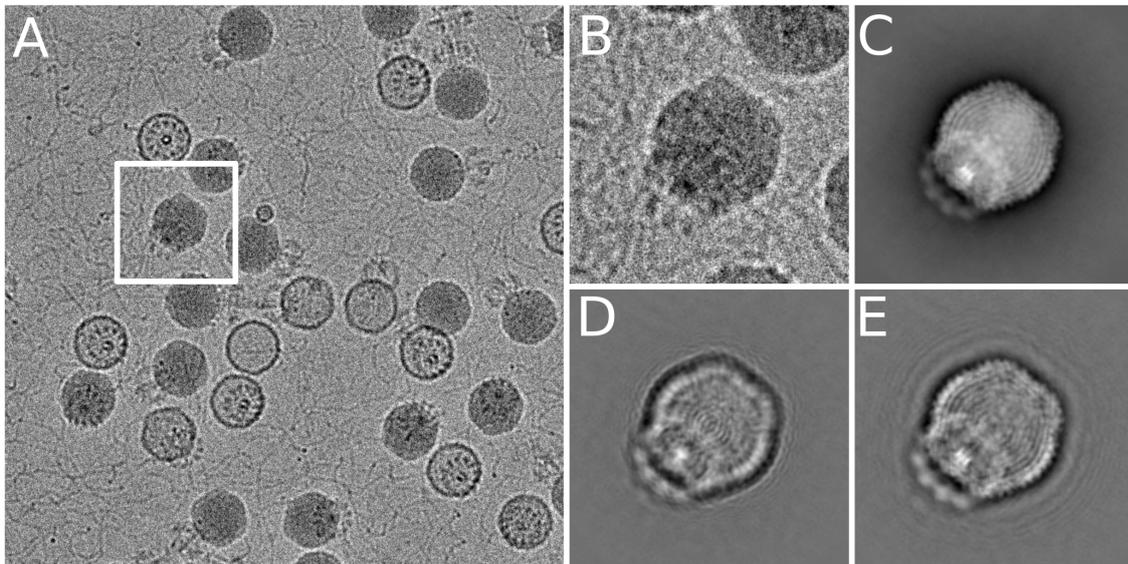


Figure 1.5: **Micrograph and effect of CTF on particles of bacteriophage P68.** (A) Cryo-EM micrograph of bacteriophage P68. (B) An Extracted particle of bacteriophage P68 from A. (C) Simulated projection of bacteriophage P68 created from already known 3D cryo-EM density map. The phage particle has the same orientation as in (A). (D) The same projection as in (C) but multiplied by CTF with defocus of  $10\ \mu\text{m}$ ,  $C_s=2.7$ , and acceleration voltage of  $300\ \text{kV}$ . Note the halo extending the particle indicating delocalization of signal. Due to the signal delocalization, a box size extending the particle by about  $1.5\times$  is necessary. (E) The same image as in (D) but after CTF correction by phase flipping. The halo effect was suppressed, and the signal was restored to its original position.

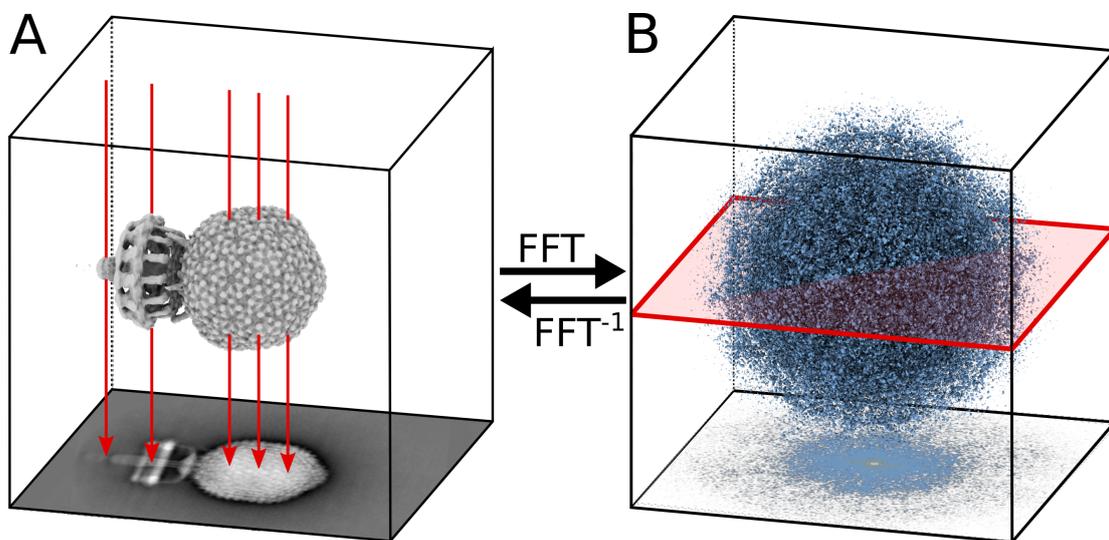


Figure 1.6: **3D reconstruction by Fourier slice theorem**, which defines the relationship between the 2D projection of a 3D object in real-space (**A**) and a 2D central slice (cutting the origin) of a 3D Fourier transform of the object in the Fourier space (**B**). The orientation of the 2D slice in Fourier space (red square in B) must be orthogonal to the projection direction (red arrows in A).

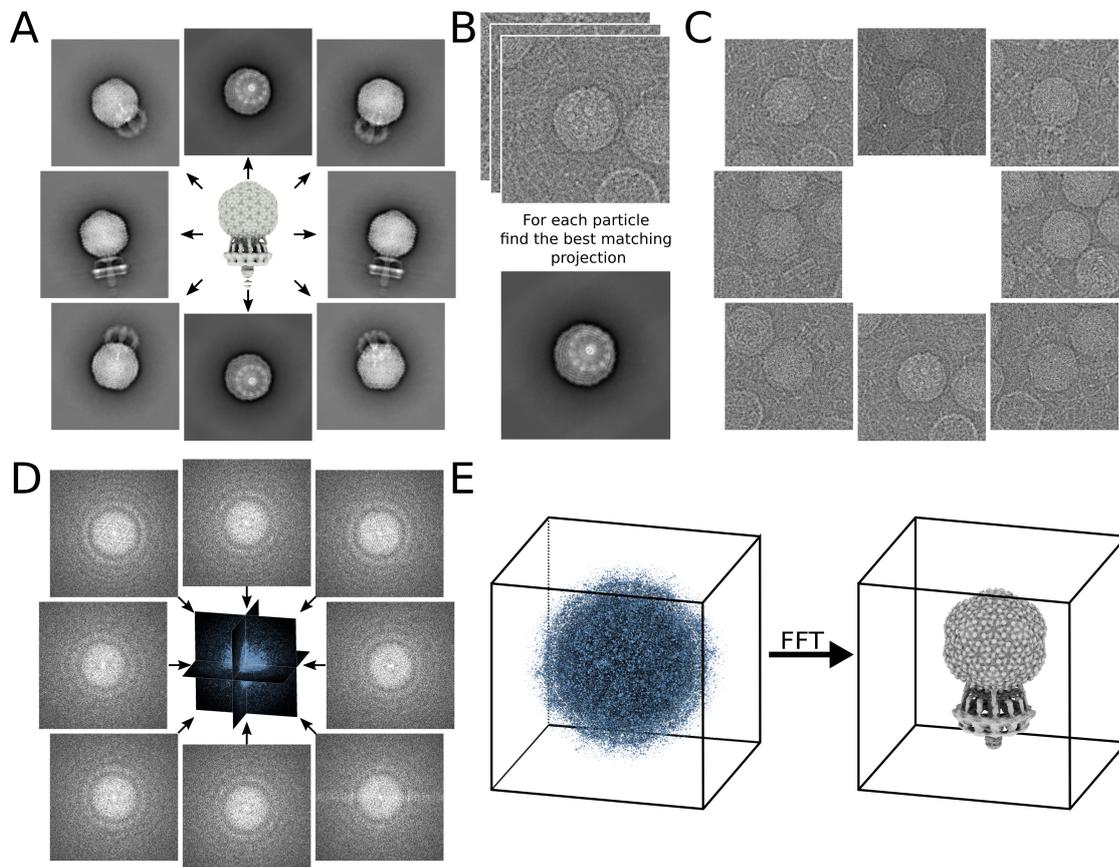


Figure 1.7: **Determination of particle orientations and 3D reconstruction in SPA.** (A) Initially, a set of reference projections of an initial model is generated in all directions. Thus, Euler angles are known for each projection. (B) Afterward, each particle is compared to all of the different projections. The best match is selected, and the particle has assigned three Euler angles of the best matching projection. (C) This rotates every particle relative to the 3D initial model. (D) Subsequently, the Fourier transform is computed for all of the particles. According to the Fourier slice theorem, a combination of many properly oriented 2D slices (the 2D Fourier transforms of all experimental projections for which we just determined their orientations) can generate the 3D Fourier transform of the 3D object. (E) To obtain the real space 3D map, the Fourier transform of the 3D reciprocal map is performed. Then the improved map is used for the generation of a new set of projections, and the steps are repeated until convergence.

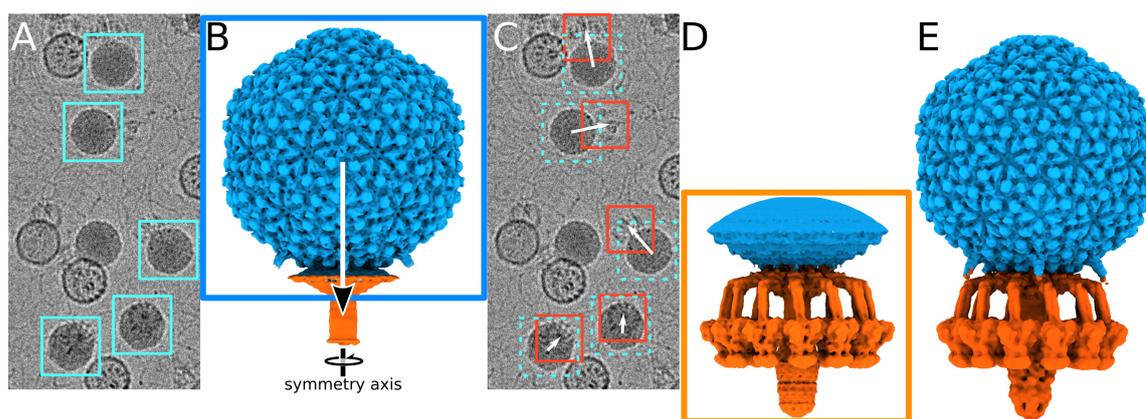


Figure 1.8: **Localized reconstruction of subparticles – the case of bacteriophage P68.** (A) Initially, only capsids possessing 5-fold symmetry are picked and extracted from the micrographs (cyan squares). (B) Afterward, 3D reconstruction using 5-fold symmetry is performed, which results in 3D cryo-EM map with capsid displaying structural details (blue), while tail (orange, 12-fold symmetry) displaying no resolvable structural features due to the symmetry-mismatch. (C) Taking advantage of the knowledge of particle orientations from the 3D reconstruction of the capsid, we can draw a vector from the capsid to tail (B, white arrow) which can be used to move the center of the original box (dashed cyan squares in C) to tail, which allows precise re-extraction of tails from the micrographs (red boxes in C). (D) Using the re-extracted tail sub-particles from (C), a 3D reconstruction with 12-fold symmetry is performed. This results in resolved structural details in parts of the structure possessing 12-fold symmetry (orange) while losing structural details of the capsid (blue). (E) Combining reconstructions from (B) and (D), we can obtain the structure of bacteriophage P68 as a whole, bypassing the symmetry-mismatch issue.

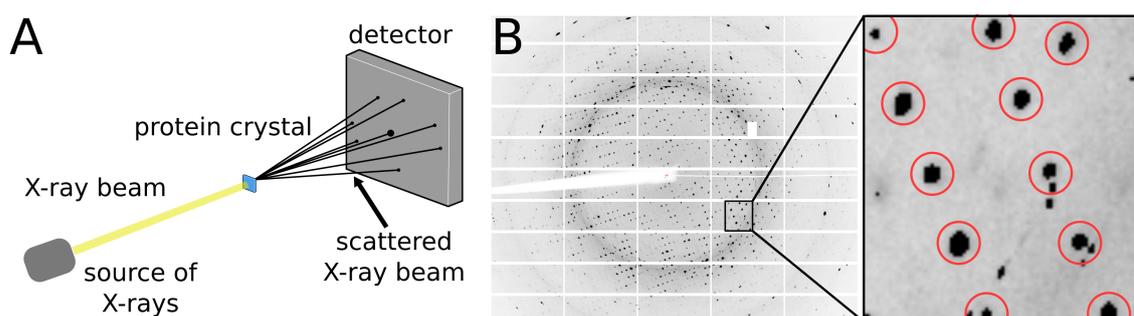


Figure 1.9: **Scheme of X-ray diffraction experiment.** (A) X-rays are produced either by home-source or by synchrotron radiation. Then the X-ray beam hits the crystal; as the beam passes through the crystallized protein, some of the rays interfere with each other constructively, forming a diffraction pattern at the detector. (B) A representative diffraction pattern from a crystal of tail-spike protein of bacteriophage P68 solved in this thesis by X-ray crystallography. The inset shows diffraction spots (reflections) in detail. The reflections are the black dots highlighted by red circles.

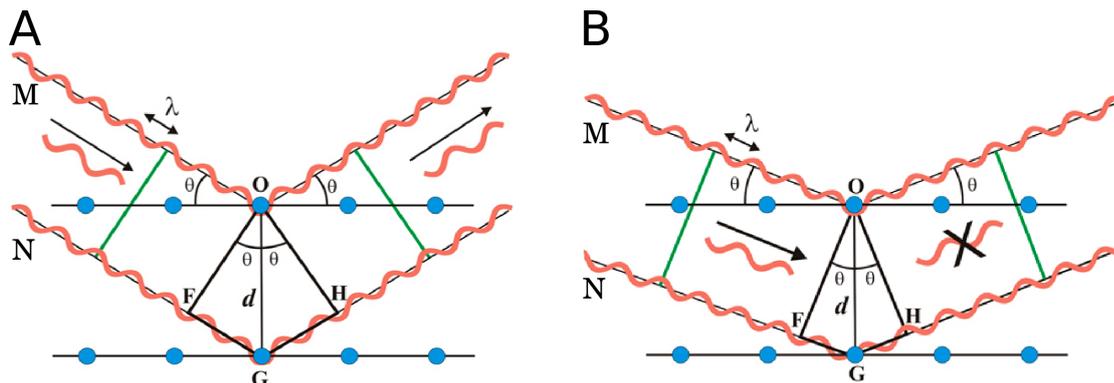


Figure 1.10: **Bragg's law.**(A,B) The two orange X-ray waves M and N of wavelengths  $\lambda$  arrive with the same phase at an incidence angle of  $\theta$  forming a wave-front (green line). The waves are interacting with the imaginary scattering 'mirrors' called as Bragg planes in the crystal lattice (black horizontal lines). Since the crystal is repetitive, the same plane is in each unit cell forming the crystal and the planes are separated by a constant distance -  $d$ . The scattering centers (O and G) - atom electron clouds - are scattering the waves in all directions. However, only certain direction results in diffraction signal on the detector. Only the direction in which the waves are in phase after scattering on the diffraction centers (green line in (A)) result in diffraction signal (A). The condition to be met to obtain a diffraction - the path extra path traveled by the wave N must correspond to an integer of the wavelength of the X-rays. Therefore, distance FG and GH must equal to  $n$  times the wavelength  $\lambda$ . In (B) the angle under which the waves arrive does not produce any diffraction signal because the extra distance traveled by the wave N does not correspond to the integer of the wavelength  $\lambda$  and thus after scattering the waves are not anymore in phase. (Adapted from [46])

## 1.5 Viruses

Viruses are the most abundant biological entities on the planet Earth. With an estimated number of more than 10 nonillion ( $10^{31}$ ) viral particles, they are outnumbering bacteria even tenfold in some ecosystems [47]. Mostly considered as pathogens, viruses are much more than that. They play a role in almost any ecosystem and affect all parts of life on the planet, ranging from our immune system [48] to the gut microbiome [49], keeping balance in the ecosystems on land and sea [50, 51], to climate regulation and the evolution of all species [52, 53, 54, 55].

Even though they are balancing on the edge of being defined as alive, they are the top predators of the microbial world [47]. Since they lack their own replication machinery, they must take over the host cell and hijack its machinery to replicate. The initial step in the infection is the injection of the viral genome into the host – a stage of infection that was studied in this dissertation thesis by the methods of structural biology.

### 1.5.1 Structural Biology of Viruses

The study of viruses from the structural biology point of view requires either solving the structure of the virus or isolated viral proteins. Structural studies revealed mechanisms of all stages of the viral life cycle ranging from the initial entry of the virus through replication of viral particles to their release from the cell [56]. Constant improvements in methodology allowed the study of the replicating virus *in vivo* without the need for fixation, utilizing state-of-the-art methods of correlative light and cryo-electron microscopy and focused ion beam milling [57, 58].

### 1.5.2 Bacteriophages

Bacteriophages are the most abundant entities in the virosphere [59]. A bacteriophage is a virus infecting a bacterium and using it for its own replication. They were discovered independently by Frederick Twort and Félix d’Hérelle in 1915 and 1917, respectively [59]. Immediately after their discovery, bacteriophages were tested for use as a medical therapy to treat bacterial infections. However, they fell out of fashion in the 1940s due to more potent and effective antibiotics [60].

Nowadays, the interest in bacteriophage therapy is back due to the emergence of antibiotic resistant strains of bacteria such as methicilin-resistant *Staphylococcus aureus* (MRSA). Therefore, it is important to know details about all stages of the bacteriophage life cycle to better regulate and even improve bacteriophages for medical treatments.

#### Classification of bacteriophages

Historically, phages have been classified based on their morphology – by observation of their appearance in an electron microscope. The most apparent morphological feature is the presence or lack of bacteriophage’s tail [61]. The tail is attached to the phage’s capsid and serves as a delivery system for the genome during infection.

One can imagine it as a tube through which the genome is pushed into the host cell. Here we will discuss in more detail only tailed bacteriophages.

The tailed phages are unified in the order of *Caudovirales* which were traditionally further divided into three families based on the tail length and contractility – *Podoviridae*, *Myoviridae*, and *Siphoviridae*. *Podoviridae* were the smallest by size and genome length, and they have a short non-contractile tail. On the other hand, *Myoviridae* had a longer tail which is contractile. The last family of *Siphoviridae* had a long non-contractile tail. The newer type of classification approved by ICTV's Bacterial and Archaeal Viruses Subcommittee introduced new genome-based families where, at the time of writing three new families of myoviruses have been ratified – *Ackermannviridae*, *Chaseviridae*, *Herellviridae*; two of siphoviruses *Demereciviridae*, and *Drexelviriidae*, and one of podoviruses, *Autographiviridae* [62].

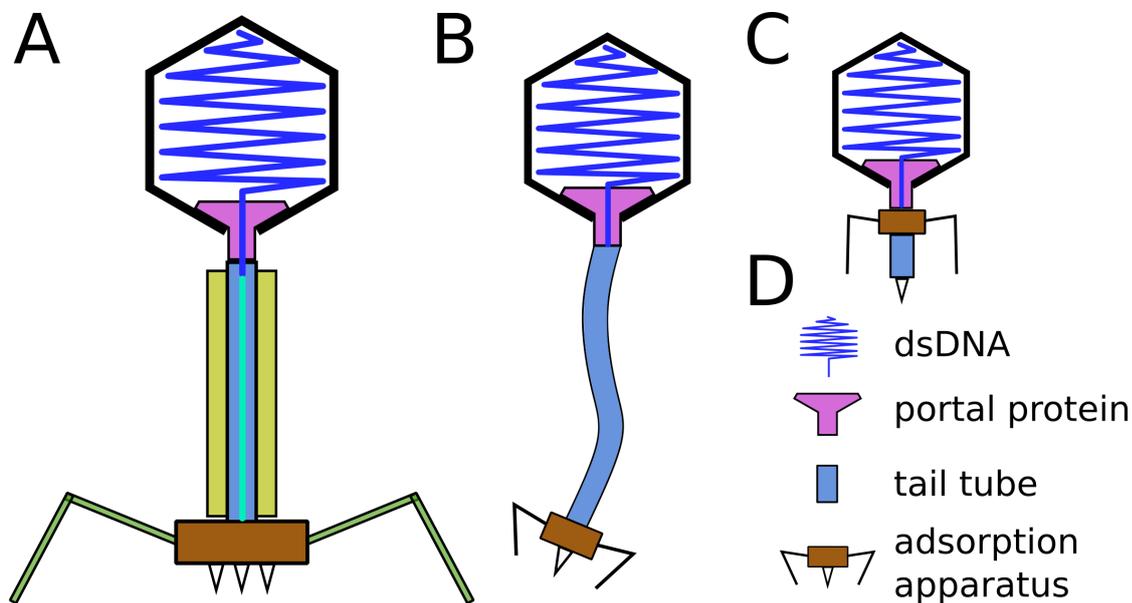


Figure 1.11: **Structural overview of the order of *Caudovirales*.** (A) Scheme of the structure of a *Myoviridae* bacteriophage. (B) Scheme of the structure of a *Siphoviridae* bacteriophage. (C) Scheme of the structure of a *Podoviridae* bacteriophage. (D) Schematic representation of individual bacteriophage components.

### Structural organization of bacteriophages

Bacteriophages from the order of *Caudovirales* are composed of three major components – capsid, tail, and baseplate at the very end of the tail [63].

**Capsid** The capsid is composed of either one or more types of proteins. The major capsid protein (MCP) is the major component of the capsid and forms the capsid shell protecting the dsDNA genome inside of the capsid. Despite low sequence identity among the phages, all known MCPs possess HK97 fold which was initially found in *Siphoviridae* phage HK97 [64]. Other proteins present on the outer or inner

surface of the capsid have either stabilizing or receptor binding functions [65]. The capsid possesses either icosahedral or 5-fold symmetry in the case of oblate or prolate capsids [66].

**Portal complex** The portal complex forms a dodecameric ring and it is composed of one or more proteins, which replace one pentameric vertex of the bacteriophage capsid. It serves as an initial gate for the dsDNA genome during its ejection into the host cell. In the phage life-cycle, it initiates the procapsid assembly and serves as an entrance for filling of the procapsid with the dsDNA genome [67]. All known portal proteins have a conserved fold with four characteristic domains – clip, stem, wing, and crown [68].

**Tail** At the distal end, the portal protein is connected to the tail complex, which has 6-fold symmetry. Some phages have additional tail-connector proteins which form an interface between the portal and tail complexes and mediate the symmetry mismatch from 12-fold to 6-fold symmetry. The tail organization depends strongly on the bacteriophage family: *Podoviridae* have short non-contractile tails, *Myoviridae* have longer contractile tails, and *Siphoviridae* have long non-contractile tails [63]. The tail structure also depends on the host type – bacteriophages infecting gram-positive bacteria have different tail organizations than those infecting gram-negative bacteria.

The *Podoviridae* tails are composed mostly of receptor binding proteins and cell penetration proteins such as peptidoglycan hydrolases and membrane penetration proteins. During the genome release *Podoviridae* tails, undergo conformational changes. Each phage type has its own cell penetration and genome release mechanism. E.g. *E. coli* phage T7 rearranges positions of whole proteins [69], while *S. aureus* phage P68 is more 'rigid', and conformational changes were observed only in protein domains [38].

The *Myoviridae* tails are composed of a tail-sheath and a baseplate located at the distal end of the tail-sheath. The tail-sheath has a double layer organization with contractile tail-sheath proteins forming the outer layer arranged in helical symmetry. The inner tail tube is rigid and is released during the sheath contraction piercing the host cell membrane like a needle. Then the dsDNA is ejected from the capsid through the inner tail tube into the host cell [70]. The baseplate of *Myoviridae* is a huge protein complex. For example baseplate of phage T4 is composed of at least 16 different proteins [71]. The baseplate proteins have a variety of functions, e.g. receptor sensing and binding, structure stabilization, and cell penetration. During the contraction, the baseplate undergoes a conformational rearrangement, changing positions of whole proteins [71].

The *Siphoviridae* tails somewhat resemble that of *Myoviridae*, where the tail is composed of a long tail-tube with a baseplate complex at the distal end. However, the tails are flexible and longer. The baseplate is usually less complex than that of *Myoviridae*, structurally closer to tails of *Podoviridae* phages [72].

That is a very general summary of bacteriophage structure. It is impossible to generalize it in more detail since even bacteriophages from the same family display

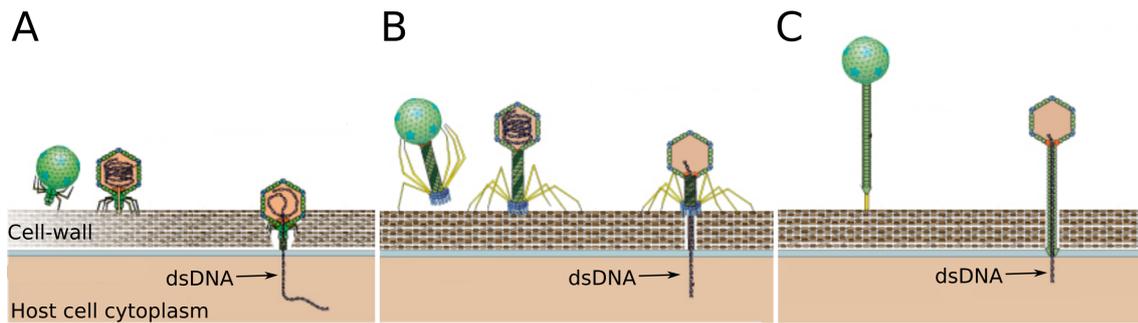


Figure 1.12: **Genome delivery mechanism by tailed phage families from the order of *Caudovirales*.** (A) Genome delivery mechanism of *Podoviridae*. Initially, the phage binds to a specific receptor through its receptor binding protein located in the tail. When its bound the, phage penetrates the cell wall and cell membrane (blue line) and releases its genome into the cell (B) Genome delivery mechanism of *Myoviridae*. Initially, the virus searches the cell surface for the receptor by receptor binding proteins located at tail-fibers (yellow sticks). When a receptor is found, the phage anchors itself on the cell surface. Afterward, the tail is contracted, and the tail tube penetrates the cell wall and cell membrane like a needle. Finally, the dsDNA is released from the phage capsid into the cell cytoplasm. (C) Genome delivery mechanism of *Siphoviridae*. Similarly, as in (A) and (B), the bacteriophage must initially find the receptor. After binding, the baseplate of the phage degrades the peptidoglycan cell wall and penetrates the cell membrane. The final step is the release of dsDNA into the host cytoplasm (Modified and adapted from [73]).

extensive conformational variability and mechanism of infection. However, some things have bacteriophages in common despite their low sequence and morphological similarity. Proteins of the same function display the same folds among all bacteriophage families. This is for example, the case of receptor binding protein of distantly related phage P68 (*Podoviridae*) and phage 80 $\alpha$  (*Siphoviridae*), where the receptor binding domains display root mean square deviation (RMSD) value of 2.4 Å. [38, 72]

### Bacteriophage life cycle

Phages can have two types of life cycles. The lytic life cycle occurs when a phage infects the host, replicates, and eventually kills the cell releasing progeny phages ready to infect other cells. The second type of phage life cycle does not kill the host cell immediately. Instead, the phage genome gets integrated into the host genome or exists as plasmids within the host cell [74]. This type of life cycle is referred to as 'lysogenic' and can last for thousand of host cell generations after the initial phage genome integration. The phage can eventually switch to the lytic cycle and lyse the cell after an induction event triggered by stressful conditions (DNA damage) [75]. A detailed description of the phage life cycle is depicted in (Fig. 1.13).

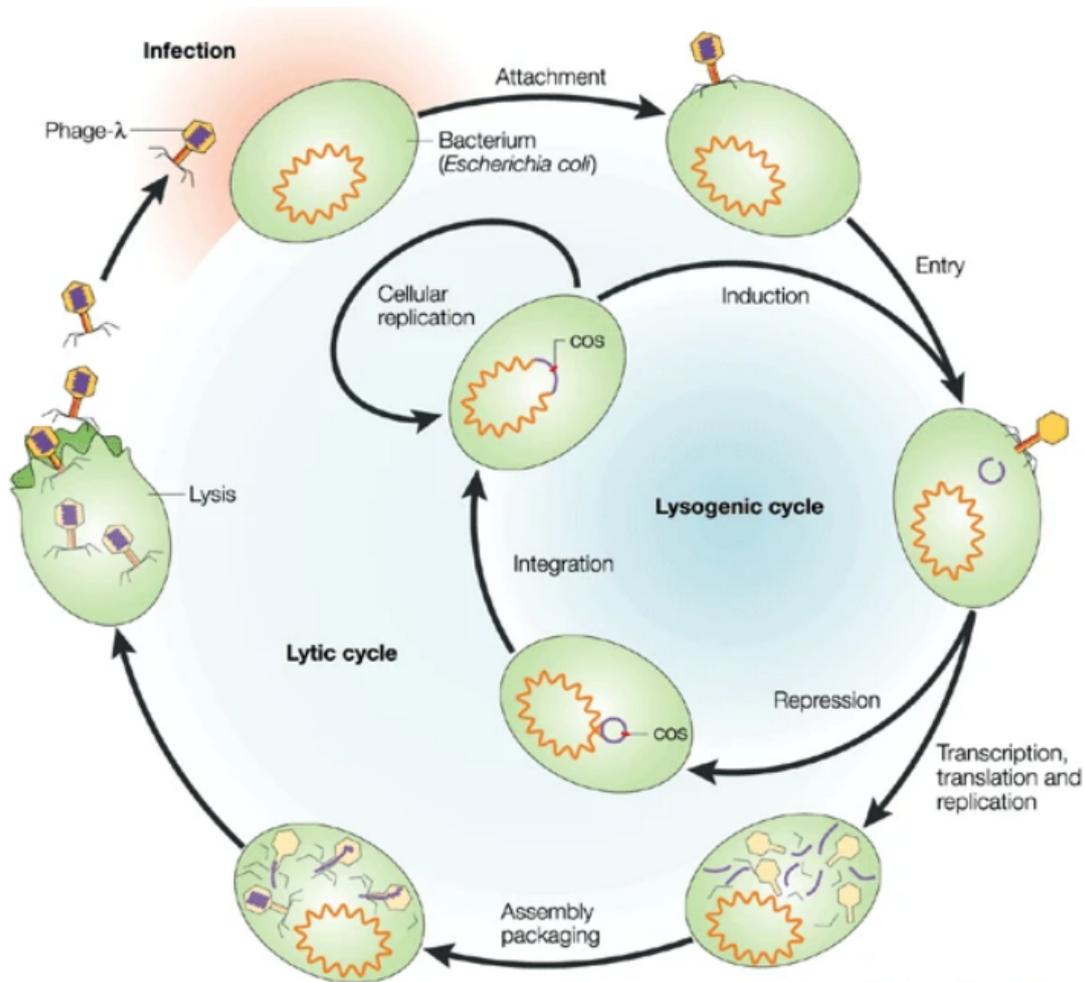


Figure 1.13: **Scheme of life cycle of bacteriophage  $\lambda$ .** The bacteriophage infection starts with attachment to the host-cell which is decorated by specific receptor. After successful binding the bacteriophage penetrates the cell envelope and delivers its genome into the cytoplasm. The genome can be integrated into the host genome which leads to a latent infection called lysogenic cycle. On the other hand, during the lytic cycle an immediate transcription, translation, and replication of the bacteriophage DNA and proteins takes place. Structural proteins are replicated as the latest leading to formation and accumulation of phage particles inside of the host cell. The final step is lysis of the host cell which releases newly made bacteriophages ready to infect neighboring cells. (Adapted from [76])

### 1.5.3 Eucaryotic viruses – the genus of *Enterovirus*

The *Picornaviridae* family contains 35 genera, one of which is *Enterovirus*. Other genera from the *Picornaviridae* family containing human pathogens are *Parechovirus*, *Hepatovirus*, *Kobuvirus*, *Cardiovirus*, and *Aphthovirus*. The genus *Enterovirus* (EV) contains echoviruses, coxsackieviruses, enteroviruses, and rhinoviruses [77]. The classification of *Enteroviruses* is depicted in (Fig. 1.14). Various diseases are linked to these viruses, including hand-foot-and-mouth disease, polio, encephalitis, paralysis, non-specific febrile illness, and respiratory diseases such as pneumonia and bronchitis. Their primary sites of replication are respiratory and gastrointestinal tracts from which they can disseminate and infect other organs, including the central nervous system causing encephalitis [78]. Apart from two EV-A71 vaccines marketed in China, there are currently no effective treatments for enterovirus infections [77].

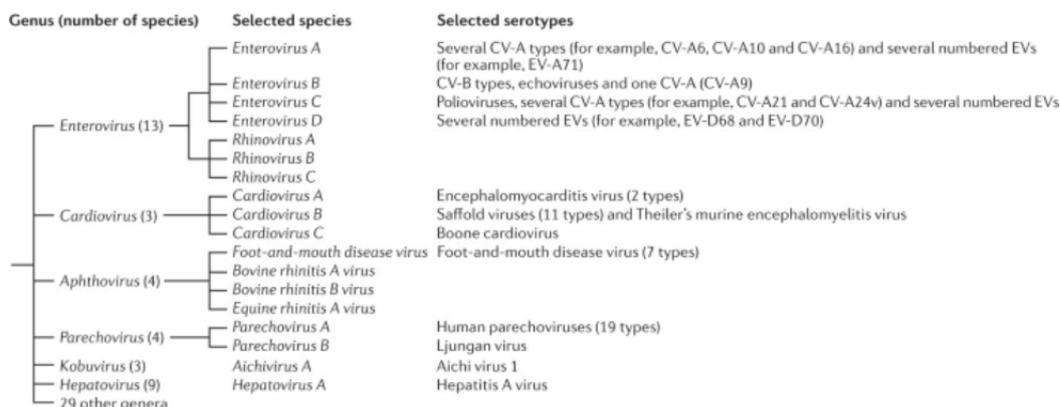


Figure 1.14: **Classification of *Enteroviruses*.** The genus *Enterovirus* is one of the 35 genera of the *Picornaviridae* family. *Enteroviruses* are classified into 13 species, with more than 100 types of *Enteroviruses* that infect humans. All of them belong to four species, *Enterovirus A-D*; besides poliovirus (*Enterovirus C* species member), these contain coxsackieviruses (CVs) A and B, echoviruses and recently discovered viruses that are named 'enterovirus' (EV) which are sequentially numbered (starting with EV-D68). Rhinoviruses were reclassified as *Enteroviruses*, and currently, more than 160 rhinovirus types are known grouped *Rhinovirus A-C* species. (Adapted from [77])

### 1.5.4 Structure of *Enteroviruses*

In 1985, X-ray crystallography structures of two *Enteroviruses* were published by James Hogle (poliovirus) and Michael Rossmann (rhinovirus 14) [79, 80]. These were the first two animal RNA viruses for which the high-resolution three-dimensional structures were determined. Structures of many other EVs were solved since then either by X-ray crystallography or cryo-EM.

It was observed that all EVs share similar conserved structural features. They are small non-enveloped viruses of 30 nm in size carrying approximately 7500 nucleotide positive sense (+) ssRNA genome. The basic structural unit – protomer – out of

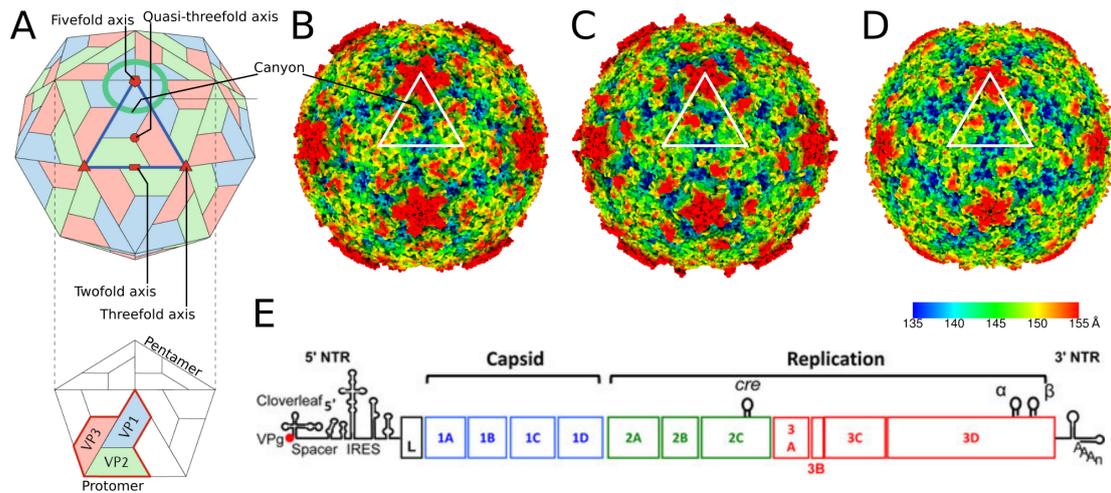


Figure 1.15: **Structural arrangement of *Enteroviruses*.** (A) Scheme of structure of an *Enterovirus* particle. The blue triangle shows the asymmetric unit of the icosahedron. Icosahedral symmetry axes are represented by the red rectangle, red triangles, and red pentagon for two, three, and five-fold symmetry axes, respectively. In the pentamer inset below the icosahedron, a protomer composed of VP1, VP2, and VP3 is highlighted. Minor capsid protein VP4 is not visible because it is on the inner surface of the capsid shell. (B,C,D) Molecular surfaces of rhinovirus 14, poliovirus, and echovirus 18 for B, C, and D, respectively (PDB IDs: 4RHV, 1HXS, 6HBG). The surfaces are radially rainbow color-coded according to the distance from the center of the particle. (E) Schematic representation of *Enterovirus* genome. (Adapted and modified from [77])

which all *Enterovirus* are constructed consists of four viral proteins VP1, VP2, VP3, and VP4 (Fig. 1.15). Sixty of such protomers form icosahedral shell with a pseudo  $T = 3$  arrangement that encapsidates the viral genome [77]. VP1, VP2, and VP3 adopt an eight-stranded antiparallel  $\beta$ -barrel fold and form the capsid shell. Minor capsid protein VP4 is myristoylated and is located on the inner surface of the capsid. The outer surface of the capsid contains a typical depression called 'canyon' encircling the 5-fold symmetry axis (Fig. 1.15). The canyon is a receptor binding site for many *Picornaviruses* [81]. In the vicinity of the canyon, a hydrophobic pocket within the VP1 can be found, usually filled with a pocket factor – a lipidic moiety. It is assumed that the pocket factor contributes to the particle stability and is involved in the uncoating of the virus during genome release [82].

### 1.5.5 The life cycle of *Enterovirus*

The *Enterovirus* infection starts with the binding of the virus to one or multiple cell surface receptors. After successful binding, receptor-mediated endocytosis is initiated, which results in endocytic uptake of the virus. Different *Enteroviruses* use different cell surface receptors and endocytic routes. When in the endosome, the receptor binding and/or acidification of the endosome converts the native enterovirus

to an 'activated' state (A-particle). The changes associated with conversion to the A-particle include: externalization of VP1 N-terminal amphipathic helix, formation of pores at the 2-fold symmetry axes, and release of the minor capsid protein VP4 [83, 84, 85, 86, 82, 87]. Afterward, the virus is uncoated, resulting in genome release to the cytoplasm. It is not confirmed whether the uncoating of the virus and genome release occurs in the endosome or in the cytoplasm. Studies supporting both claims were published [88, 89, 90, 87, 91]. It is possible that different *Enteroviruses* use different genome release strategies despite their structural and sequential similarities.

After the (+)ssRNA genome is delivered to the cytosol, the whole RNA genome is translated to a single polyprotein. Afterward, the polyprotein is cleaved by viral proteinases 2A<sup>pro</sup>, 3C<sup>pro</sup>, and 3CD<sup>pro</sup> into ten proteins, including capsid proteins VP0, VP2, and VP3; replication proteins 2A–2C, and 3A–3D. The *Enteroviruses* induce the formation of membrane structures called 'replication organelles' (ROs). The replication is initiated by viral RNA-dependent RNA polymerase 3D<sup>pol</sup>. The 3D<sup>pol</sup> synthesizes a negative-strand (-) copy of the RNA genome, forming a double-stranded RNA replication intermediate [92]. Newly synthesized genome copies serve either as a template for further transcription and translation or are encapsidated to newly formed provirions. The provirions are composed of VP0, VP1, and VP3. After encapsidation of the (+)ssRNA genome, VP0 is cleaved into VP4 and VP2 by as of yet unknown mechanism [93]. Finally, the mature virions are released from the cell either by lytic or non-lytic exit pathway [94, 95, 96, 97, 98]. The *Enterovirus* life-cycle is described in (Fig. 1.16)

The *Enteroviruses* are very efficient in hijacking the cell machinery for their own purpose. The hijacking is mediated by viral proteinases 2A<sup>pro</sup> and 3C<sup>pro</sup> that cleave several host proteins to optimize virus translation, replication, and virus spread. For example they interrupt type I interferon and stress pathways (by cleaving RNA sensors and signalling proteins), induce host proteosynthesis shut-off (by cleaving cellular factors involved in transcription and cap-dependent translation of host mRNAs), disrupt nucleocytoplasmic transport (by cleaving nucleoporins), and disrupt the cellular cytoskeleton [99, 100, 101]. All these changes have one purpose, to allow the formation of as many progeny particles before the ultimate cell death as possible.

### 1.5.6 Rhinoviruses

Rhinoviruses are one of the member species of *Enteroviruses* that can cause a variety of upper and lower respiratory tract infections, including the common cold, bronchitis, pneumonia, and exacerbations of chronic respiratory illnesses such as asthma. Currently, there are more than 170 identified genotypes of rhinoviruses classified into three classes – A, B, and C according to clade-specific similarities in the genome sequence, serology, and capsid structure [102]. The majority of rhinoviruses, which constitute all rhinovirus B and the majority of rhinovirus A classes (major group) utilizes intercellular adhesion molecule 1 (ICAM-1) as a cellular receptor. A minority of rhinoviruses (minor group) require low-density lipoprotein receptor (LDLR) to initiate an infection. The recently discovered class C of rhinoviruses depends on cadherin-related protein 3 (CDHR3) to recognize and enter the host cell [103].

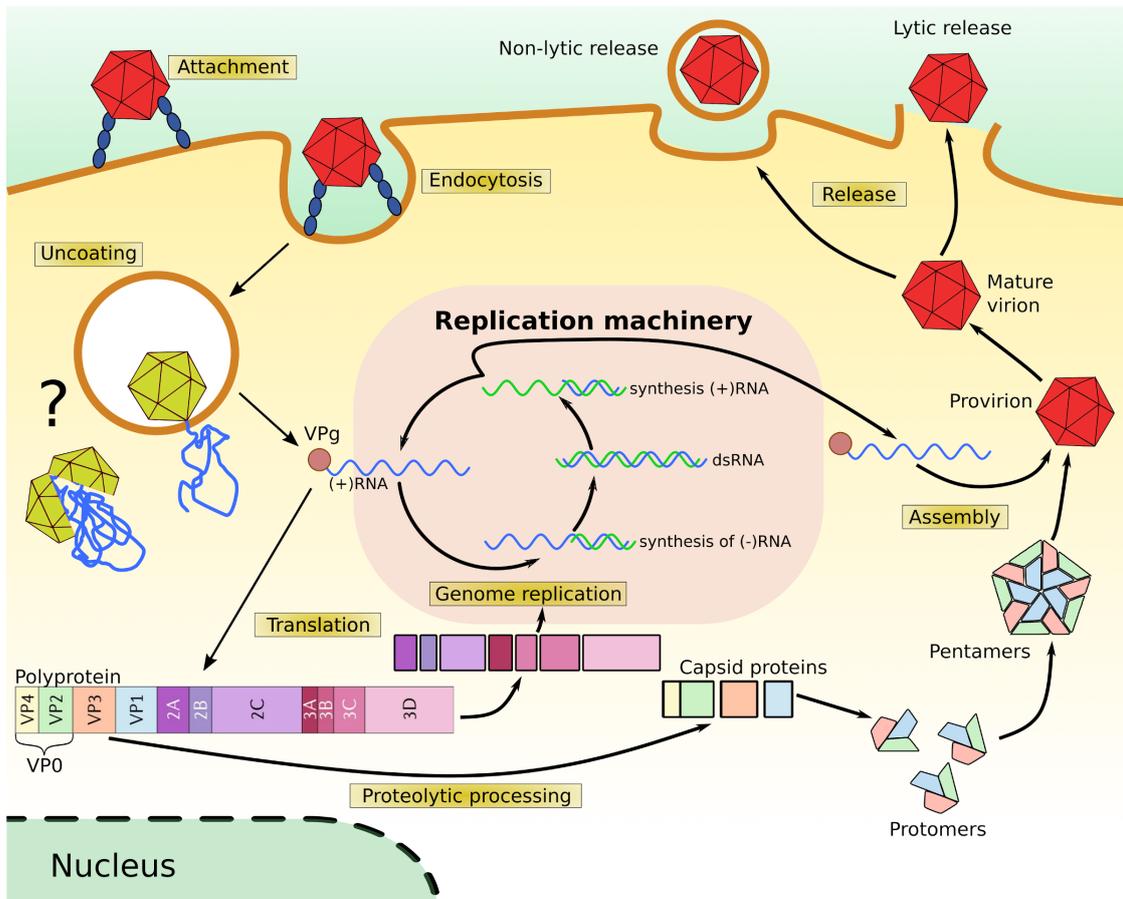


Figure 1.16: Life cycle of an *Enterovirus* (Inspired by [77])

### 1.5.7 Intercellular adhesion molecule 1

ICAM-1 is a transmembrane glycoprotein, a member of the immunoglobulin superfamily of cell adhesion proteins, and it is typically expressed by epithelial cells, endothelial cells, and leukocytes. It can be activated by chemical, or physical stress, or inflammatory stimuli. The function of ICAM-1 involves co-stimulation of cell activation, and it helps to regulate leukocyte migration from blood tissues in inflammatory sites when it binds to two integrin receptors, leukocyte function-associated antigen and macrophage-1 antigen [104].

Moreover, ICAM-1 was identified as a major rhinovirus receptor by two groups in 1989 [105, 106]. The structure of ICAM-1 is composed of five consecutive extracellular immunoglobulin domains (D1, D2, D3, D4, and D5), a carboxyl-proximal transmembrane domain (TM), and a C-terminal cytoplasmic domain [107, 108]. The N-terminal D1 domain is involved in the binding of rhinoviruses. It binds the 'canyon' region of the capsid interacting with VP1, VP2, and VP3 [109, 110].



# Chapter 2

## Materials and Methods

### 2.1 Materials and Methods

Materials and methods used in this dissertation thesis are described thoroughly in published articles presented in the next section.



# Chapter 3

## Results and Discussion

Results and discussion are presented as a compilation of published peer-reviewed articles regarding genome release and cell entry of prokaryotic and eukaryotic viruses. Each paper is shortly discussed, and main results are presented.

### 3.1 Published Articles

**Article 1: Structure and genome ejection mechanism of *Staphylococcus aureus* bacteriophage P68** The goal of the study was to determine the structure of *S. aureus* bacteriophage P68 in different stages of the genome ejection. Then, by comparison of conformational changes in the bacteriophage, we could describe the genome ejection mechanism. I started with the selection of heads of native virions from the micrographs. The structure of the head was solved to 3.5 Å resolution with C5 symmetry implemented during the reconstruction. However, all structural features of the tail were smeared because of the symmetry mismatch between the head (C5 symmetry) and the tail (C6/C12 symmetry). Therefore, I utilized localized reconstruction of subparticles to solve the symmetry mismatch problem as described in Chapter 1. I constructed a vector pointing from the central part of the head to the central part of the tail. Using the vector and known orientations of the particles from the C5 reconstruction of the head, I was able to re-extract tails of the bacteriophages from the micrographs. Then I performed 12 and 6 fold symmetry reconstructions which led to the final resolution of 3.9 Å in the tail region. Some parts of the tail were resolved very poorly, especially the C-terminal part of tail-fibers, tail-knob, and tail-spike. I performed crystallization of the tail-fibers, which was successful. Crystals diffracting to 2.0 Å resolution were obtained. The phase problem was solved by iterative molecular replacement using manually adjusted structure of receptor binding protein of phage phi11 (PDB code: 5EFV) [111]. Then all parts of the bacteriophage were fitted into the asymmetric reconstruction of the whole bacteriophage, which reached the resolution of 4.7 Å. The final native bacteriophage model consists of more than 130,000 amino acids which are composed of more than 1,000,000 non-hydrogen atoms. By using symmetry expansion and subsequent 3D classification of the tail, I discovered that one subunit of the portal is in contact with dsDNA in the central tunnel of the portal. The interaction is probably crucial for maintaining the

dsDNA inside of the bacteriophage capsid. By comparison with two additional reconstructions of the bacteriophage in different states of the genome release - namely, during and after ejection of the genome, I proposed a theory of the mechanism of the bacteriophage cell-wall, cell-membrane penetration, and subsequent release of its genome into the bacterial host.

The inner surface of the capsid of phage P68 is lined with gp21 which we call as 'Arstotzka' protein, since it resembles a flag of a fictional country Arstotzka from a videogame Papers Please (Fig: 3.1). The position of gp21 of P68 resembles that of the scaffolding protein gp7 from phage phi29 [112]. Unlike the known scaffolding proteins, gp21 of phage P68 is present in the structure of the empty particle. We speculate that the gp21 of P68 can regulate the triangulation number of the capsid as it was shown for the CpmB protein of *S. aureus* pathogenicity island that uses capsid proteins of staphylococcal phage 80 $\alpha$  [113]. The triangulation number of the phage P68 is  $T = 4$ , which is the smallest that can incorporate a portal protein into the 5-fold vertex. The portal protein gp19 of phage P68 has a fold very similar to those observed in other bacteriophages such as SPP1, P22, T4, and phi29 [114, 115, 68, 116]. Similarly to *Bacillus* phage phi29, it lacks a crown domain. Asymmetric reconstruction of the portal has shown that its incorporation into the capsid requires a conformational change of the neighboring major capsid proteins. It was shown that also local morphing of the portal structure plays a role in compensating of the symmetry-mismatch between portal and capsid [117]. The portal of P68 enables binding of receptor binding proteins (RBP) gp17 that form a skirt like structure around the tail-tube. The RBP is remarkably similar to receptor binding proteins from bacteriophages from the family of *Siphoviridae*, namely phages 80 $\alpha$  and phi11 [72, 111]. The lower collar protein gp18 of P68 contains an axial channel that is continuous with that of the portal complex. A similar protein was also found in the structure of phage phi29 [118]. Tail-knob and tail-spike proteins are attached to the lower collar protein at its distal end. Both of them were solved to lower resolution than the rest of the P68 tail machinery. Therefore, I rigid body fitted an already known structure of the tail-knob of phage C1 into the density [119].





Figure 3.1: (A) Bacteriophage P68 'Arstotzka' (minor capsid) protein gp21 and (B) flag of fictional country 'Arstotzka' from the videogame Papers Please.



## STRUCTURAL BIOLOGY

Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68Dominik Hřebík<sup>1</sup>, Dana Štveráková<sup>2</sup>, Karel Škubník<sup>1</sup>, Tibor Füzik<sup>1</sup>, Roman Pantůček<sup>2</sup>, Pavel Plevka<sup>1\*</sup>

Phages infecting *Staphylococcus aureus* can be used as therapeutics against antibiotic-resistant bacterial infections. However, there is limited information about the mechanism of genome delivery of phages that infect Gram-positive bacteria. Here, we present the structures of native *S. aureus* phage P68, genome ejection intermediate, and empty particle. The P68 head contains 72 subunits of inner core protein, 15 of which bind to and alter the structure of adjacent major capsid proteins and thus specify attachment sites for head fibers. Unlike in the previously studied phages, the head fibers of P68 enable its virion to position itself at the cell surface for genome delivery. The unique interaction of one end of P68 DNA with one of the 12 portal protein subunits is disrupted before the genome ejection. The inner core proteins are released together with the DNA and enable the translocation of phage genome across the bacterial membrane into the cytoplasm.

## INTRODUCTION

*Staphylococcus aureus* causes a range of illnesses from minor skin infections to life-threatening diseases such as pneumonia, meningitis, and sepsis (1). Medical expenses caused by *S. aureus* infections in the United States and European Union have been estimated to exceed \$2.5 billion, and *S. aureus* infections result in tens of thousands of deaths every year (2). Many *S. aureus* strains, particularly those found in hospitals, carry antibiotic resistance genes (3). In 2017, the World Health Organization listed *S. aureus* among the most threatening antibiotic-resistant pathogens for which new treatments are urgently needed (4). Mouse models showed that phages could be used to treat infections caused by antibiotic-resistant *S. aureus* strains (5). All previously described phages from the family Podoviridae that infect *S. aureus* are members of the genus *Rosenblumvirus*. P68 and several other related phages are promising phage therapy candidates because of their wide host range (6). They were analyzed for their ability to treat animals infected by methicillin-resistant *S. aureus* and to prevent healthcare-associated human infections (5). Despite the testing for the use in phage therapy, the mechanism of phage P68 infection of the Gram-positive *S. aureus* cells has not been described in detail.

Three-dimensional (3D) structures of virions of several Podoviridae phages, including T7,  $\phi$ 29, C1,  $\epsilon$ 15, P22, K1E, and K1-5, were determined using cryo-electron microscopy (cryo-EM) (7–15). Bacteriophages from the family Podoviridae have icosahedral or prolate heads and short noncontractile tails. The assembly of the heads of some bacteriophages is promoted by scaffolding proteins (14). Minor capsid proteins, attached to the outer surface of the phage head, increase the stability of capsids or, in the case of head fiber proteins, enable the initial attachment of phages to bacteria. The heads of some phages contain inner core proteins that are associated with the portal complex. The inner core proteins are released together with the phage genome and were speculated to play a role in the delivery of the DNA into the cell cytoplasm.

The phage tail is attached to one of the fivefold vertices of the head in which a pentamer of capsid proteins is replaced by a dodecahedral portal complex. The tails of podoviruses are variable in size and protein composition; however, they share common organizational motifs. The tails contain specialized protein subunits for receptor binding, cell wall degradation, and cell membrane penetration. Podoviruses that infect *S. aureus* require wall teichoic acid for their adsorption to cells (16). Podovirus tails are decorated by 6 or 12 tail fibers. Structures of tail fibers from several podoviruses, including P22,  $\phi$ 29, and K1F, were determined using x-ray crystallography (17–20). Tail fibers form trimers in which the N-terminal part is responsible for attachment to the virion and the C-terminal domain enables receptor binding. Two or more types of proteins form the central tubular part of the podovirus tail. The proximal part of the tail, attached to the portal complex, is usually a dodecamer of lower collar proteins (21). In contrast, a hexamer of tail knob proteins usually forms the distal part of the tail tube (10). The tails of some podoviruses also contain tail spike proteins with hydrolase activities, which enable phages to degrade the bacterial cell wall. Tails of podoviruses contain a trimer of tail needle proteins that is located in the tube formed by the lower collar and tail knob proteins. The tail needle is essential for retaining the genome in the phage head after packaging, and it may have a role in the penetration of the cell wall (22). The genome of phage P68 is 18,227 base pairs of double-stranded DNA (dsDNA) and contains 22 open reading frames (23).

Here, we used a combination of cryo-EM and x-ray crystallography to structurally characterize bacteriophage P68 and the mechanism of regulation of its genome ejection. An asymmetric reconstruction of the P68 virion was determined to a resolution of 4.7 Å. The structures of heads of native virions, genome release intermediates, and empty particles of P68 with imposed icosahedral symmetry were determined to resolutions of 3.3, 6.3, and 3.4 Å, respectively. The same structures with imposed fivefold symmetries were determined to resolutions of 3.8, 22.7, and 4.3 Å, respectively. Portal complexes and tails with 12-fold symmetries in native, genome release, and empty particles of P68 were characterized to resolutions of 3.9, 18.8, and 7.1 Å, respectively. X-ray crystallography enabled the determination of the structure of the C-terminal part of P68 tail fiber to a resolution of 2.0 Å. In combination, the structures enabled us to

Copyright © 2019  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic. <sup>2</sup>Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic.

\*Corresponding author. Email: pavel.plevka@ceitec.muni.cz

describe the mechanisms of attachment of the portal complex to the capsid, selectivity of binding of head fibers to hexamers of major capsid proteins adjacent to the tail vertex, and the conformational changes of the portal complex required for the genome release.

## RESULTS AND DISCUSSION

### P68 capsid structure

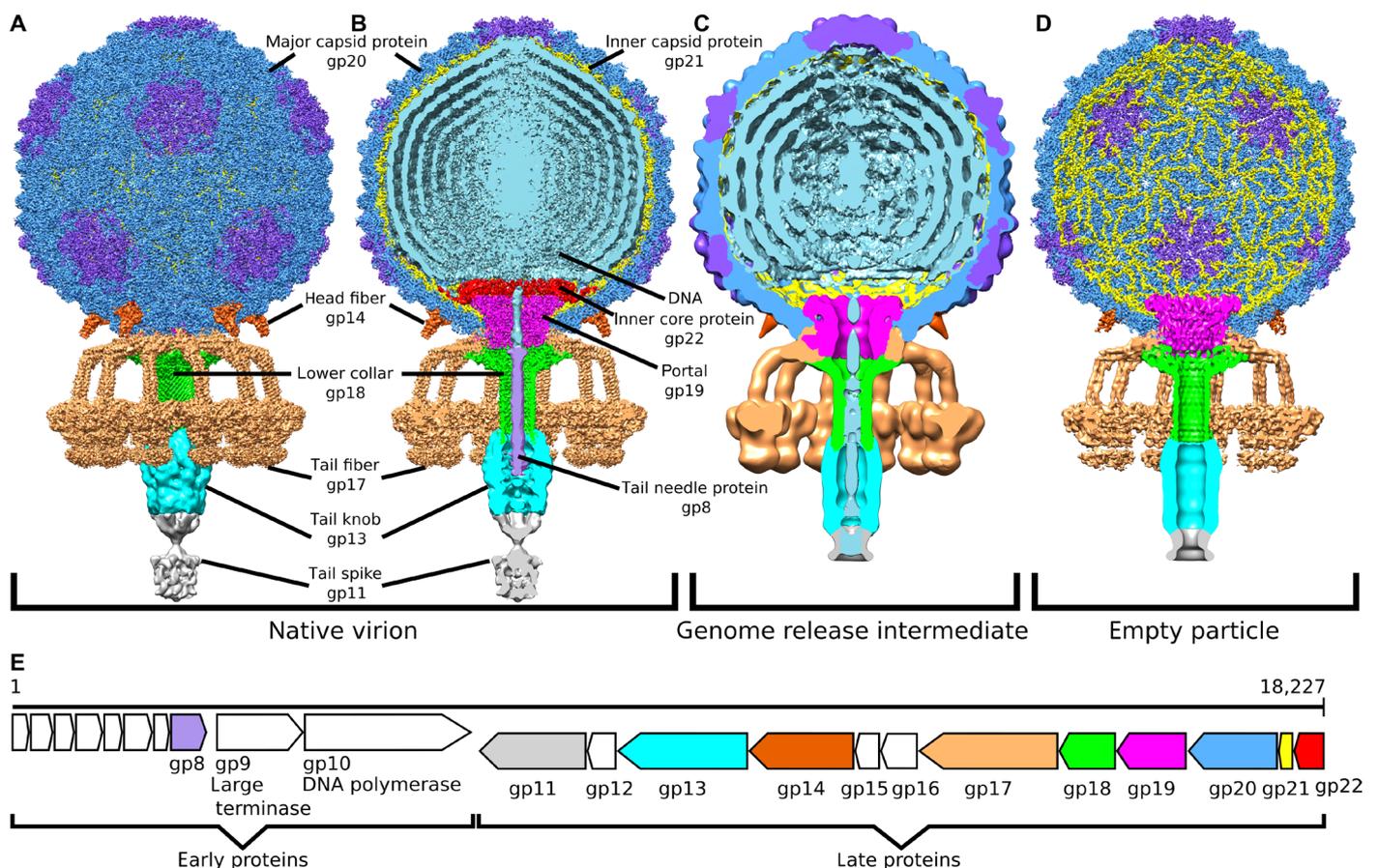
The virion of P68 has an icosahedral head with a diameter of 480 Å and a 395 Å-long tail, which is decorated with tail fibers (Fig. 1). The structures of heads of native virions, genome release intermediates, and empty particles of P68 with imposed icosahedral symmetry were determined to resolutions of 3.3, 6.3, and 3.4 Å, respectively (Fig. 1, A to D; figs. S1, A to G, and S2; and tables S1 and S2). The capsid proteins in the P68 head are organized in a  $T = 4$  icosahedral lattice (Fig. 2, A and B). The major capsid protein gp20 has the canonical HK97 fold common to numerous tailed phages and herpesviruses. According to the HK97 convention, the protein can be divided into four domains: the N-terminal arm (residues 1 to 84), extended loop (residues 85 to 124), peripheral domain (residues 125 to 182 and 346 to 388), and axial domain (residues 183 to 277, 341 to 345, and 389 to 408) (Fig. 2C). Unlike in HK97, the P68 major capsid protein also contains an insertion domain (residues 278 to 340). The insertion and peripheral domains form a cleft that binds the extended

loop of an adjacent major capsid protein and thus contribute to the capsid's stability (fig. S3, A to E).

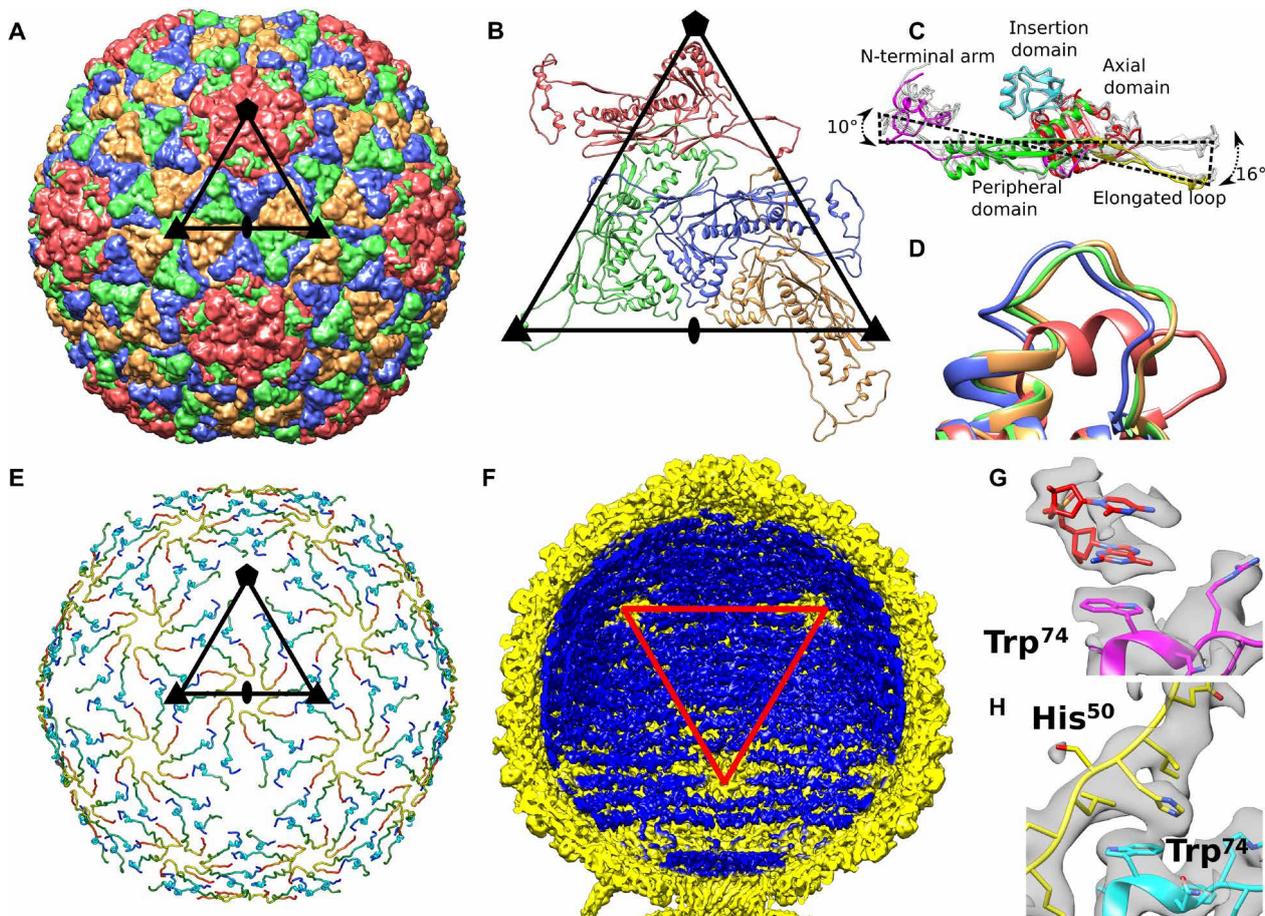
### Inner capsid proteins mediate contacts of capsid with genome

The inner face of the P68 capsid is lined by inner capsid proteins gp21, organized with icosahedral  $T = 4$  symmetry (Fig. 2E). Structures of the inner capsid proteins were determined as part of the icosahedral reconstruction of the P68 head to a resolution of 3.3 Å. Except for an  $\alpha$  helix formed by residues 14 to 21, the 55-residue-long inner capsid protein “Arstotzka” lacks secondary structure elements. The arrangement of three copies of gp21 (Fig. 2E) is reminiscent of the emblem of the fictional country Arstotzka. The inner capsid proteins related by icosahedral threefold axes and quasi-threefold axes form three-pointed stars (Fig. 2E) and are arranged so that the N terminus of one subunit is located close to the C terminus of another one within the stars (Fig. 2E).

The electron density of P68 dsDNA is resolved inside the regions of the capsid lined by the inner capsid proteins, but it is missing in the proximity of fivefold vertices, where the inner capsid proteins are not structured (Fig. 2, E and F). The electron densities of two nucleotides of single-stranded DNA are stacked against the side chains of Trp<sup>74</sup> of the major capsid proteins that form pentamers (Fig. 2G). In contrast, the side chains of Trp<sup>74</sup> of major capsid proteins that



**Fig. 1. Virion and genome organization of phage P68.** (A and B) Structures of P68 virion, (C) genome release intermediate, and (D) empty particle. The whole P68 virion is shown in (A), whereas particles without the front half are shown in (B) to (D). The structures are colored to distinguish individual types of structural proteins and DNA. (E) Schematic diagram of P68 genome organization, with structural proteins color-coded in accordance with the structure diagrams shown in (A) to (D).

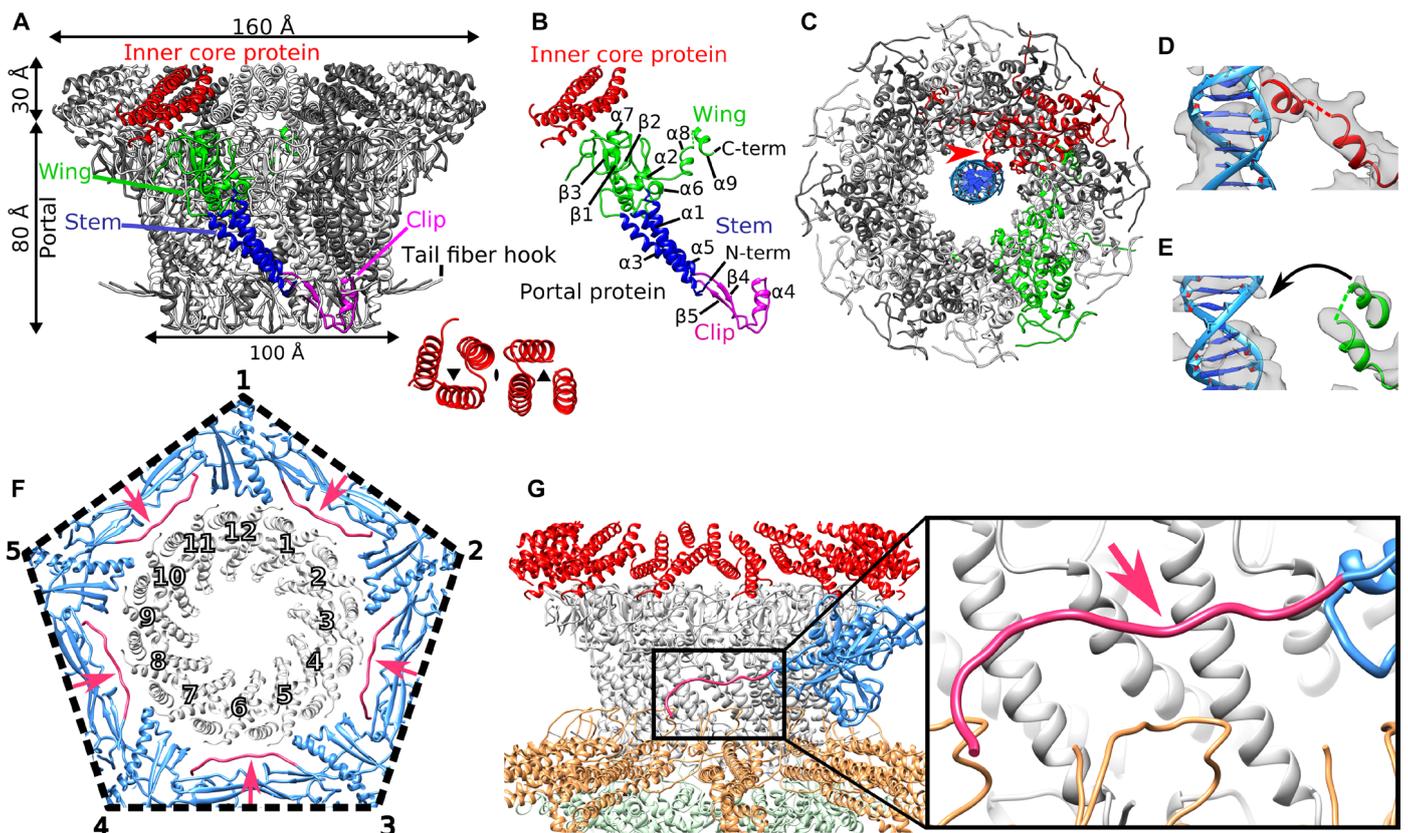


**Fig. 2. Capsid structure of P68.** (A) Major capsid proteins of P68 have HK97 fold and form  $T=4$  icosahedral lattice. Positions of selected icosahedral five-, three-, and twofold symmetry axes are indicated by pentagon, triangles, and oval, respectively. Borders of one icosahedral asymmetric unit are highlighted. (B) Cartoon representation of P68 major capsid proteins in icosahedral asymmetric unit. Positions of icosahedral symmetry axes and borders of icosahedral asymmetric unit are shown. (C) Major capsid proteins from icosahedral asymmetric unit differ in positions of elongated loops and N-terminal domains. Color coding of one of the subunits indicates division of major capsid protein to domains. (D) Residues 253 to 263 from the axial domain of major capsid proteins differ in structure. The residues form an  $\alpha$  helix in the subunit that is part of the pentamers, whereas they constitute loops in the other subunits. The color coding of subunits is the same as in (B). (E) The inner capsid protein is organized in a  $T=4$  icosahedral lattice. Proteins are rainbow-colored from N terminus in blue to C terminus in red. Subunits related by icosahedral threefold axes and quasi-threefold axes of the  $T=4$  lattice form three-pointed stars in which the C terminus of one subunit is positioned next to the N terminus of another subunit. Borders of a selected icosahedral asymmetric unit are shown. (F) The ordering of the packaged P68 dsDNA genome (shown in blue) is disrupted around the fivefold vertices of the capsid (shown in yellow). Red triangle indicates one face of icosahedron. (G) Stacking interactions of two nucleotides with side chain of Trp<sup>74</sup> of major capsid protein located next to fivefold vertex. (H) Side chains of Trp<sup>74</sup> of major capsid proteins that form hexamers bind to His<sup>50</sup> of inner capsid proteins.

form hexamers interact with the side chains of His<sup>51</sup> of the inner capsid proteins (Fig. 2H) and thus cannot bind phage DNA. Therefore, the inner capsid proteins could enable the packaging of the P68 genome in its head. The location of gp21 in the P68 genome resembles that of the scaffolding protein gp7 of phage phi29 (24). Unlike typical scaffolding proteins, gp21 subunits remain attached to the P68 capsid after maturation. Subunits of gp21 could function in determining the triangulation number of the P68 capsid during assembly, as was shown for the CpmB protein of *S. aureus* pathogenicity island 1 that is parasitical on staphylococcal phage 80 $\alpha$  (25). However, gp21 has no sequence similarity to the gp7 of phi29 or CpmB of *S. aureus* pathogenicity island 1, and the proteins also differ in their structures. The inner capsid proteins of P68 remain attached to the capsid after genome ejection, indicating that they do not participate in genome delivery (Fig. 1D).

### DNA is held inside the P68 head by interaction with one subunit of the portal complex

The structure of the portal complex from the native P68 virion with imposed 12-fold symmetry was determined to a resolution of 3.9 Å. The portal complex of P68 is 80 Å long along its 12-fold axis, with the external diameters of the upper and lower parts of 140 and 100 Å, respectively (Fig. 3, A to C). The structure of the P68 portal protein gp19 could be built except for residues 1 to 6 and 83 to 104 out of 327. According to the convention (26), it can be divided into three domains: the clip (residues 178 to 223), stem (residues 6 to 41, 156 to 177, and 227 to 248), and wing (residues 249 to 327) (Fig. 3, A and B). Unlike the portal proteins of phages SPP1, P22, and T4 but similar to that of *Bacillus* phage phi29, the portal protein of P68 lacks a crown domain (21, 27–29). The clip domain is composed of helix  $\alpha$ 4 and antiparallel strands  $\beta$ 4 and  $\beta$ 5 (Fig. 3B). It forms part of



**Fig. 3. Structure of P68 portal complex and its interaction with capsid.** (A) Twelvelfold symmetrized structure of portal and inner core complexes of native P68. One of the portal proteins is colored according to domains: clip domain in magenta, wing in green, and stem in blue. Six inner core proteins associated with one portal protein subunit are highlighted in red. The inset shows the symmetry of the arrangement of the six inner core proteins. (B) Division of portal protein into domains. Color coding is the same as in (A). (C) Asymmetric reconstruction of portal complex showing interactions of one of the portal proteins highlighted in red, with DNA shown in blue. The interaction is indicated with a red arrow. One of the portal protein subunits that does not interact with the DNA is highlighted in green. (D) Detail of interaction of helix  $\alpha_9$  of portal protein with DNA. Cryo-EM density is shown as gray transparent surface. (E) Structure of portal protein subunit that does not interact with DNA. (F) Interface between portal complex and capsid. Portal proteins are shown in gray, capsid proteins are shown in blue, and N termini of capsid proteins that mediate interactions with the portal are shown in pink and highlighted with pink arrows. (G) Side view of capsid-portal interactions. The single major capsid protein is shown in blue and its N terminus in pink, portal proteins in gray, inner core proteins in red, tail fibers in orange, and lower collar proteins in green. The inset shows detail of interactions between the N-terminal arm of the major capsid protein and stem domains of portal proteins.

the binding site for the lower collar complex and tail fibers. The stem domain of the portal protein is composed of a “tail fiber hook” (residues 6 to 11) and helices  $\alpha_1$ ,  $\alpha_3$ , and  $\alpha_5$  (Fig. 3B). The tail fiber hook enables the attachment of tail fibers to the portal complex, as discussed in detail below. Helices  $\alpha_1$  and  $\alpha_3$  of the stem domain form the outer surface of the portal complex that interacts with the capsid. The wing domain, which forms the part of the portal inside the phage head, consists of helices  $\alpha_2$  and  $\alpha_6$  to  $\alpha_9$  and strands  $\beta_1$  to  $\beta_3$ . Helix  $\alpha_6$  of the wing domain is inserted into the neighboring portal protein subunit and thus stabilizes the dodecamer structure of the portal complex (Fig. 3A).

Asymmetric reconstruction of the portal complex of native P68 virion, determined to a resolution of 6.6 Å, contains a unique interaction of one of the portal proteins, with the DNA positioned at the center of the portal channel (Fig. 3, C to E). Helix  $\alpha_9$  from the wing domain of the unique DNA binding portal subunit binds to a side of the DNA helix (Fig. 3, C and D, and fig. S2D). At a resolution of 6.6 Å, it is not possible to identify residues of the portal protein subunit interacting with the phage DNA. We speculate that the interactions of P68 genome with the portal protein and those of the tail needle

protein with the plug domain of the tail knob protein, as discussed below, hold the phage DNA inside the phage head. A similar asymmetric interaction of one of the portal protein subunits with the end of the phage genome has been previously observed for phage phi812 from the family Myoviridae (30).

### Interface between capsid and portal complex

Asymmetric reconstruction of the entire P68 virion at a resolution of 4.7 Å enabled characterization of the interface between the capsid and portal complex (Fig. 3, F and G). The binding interface is formed by the outer surface of the wing and stem domains of the portal proteins and N terminus and P domain of the major capsid protein. The interface is in agreement with the previous prediction made on the basis of docking of portal complexes into capsid structures of other phages (27). Residues 1 to 42 from the N-terminal arm of the major capsid proteins adjacent to the portal are not structured. Residues 43 to 59 of the major capsid proteins wrap around the stem domains of portal proteins (Fig. 3F and fig. S4, A and B). Capsid proteins of all tailed phages have HK97 folds that include N-terminal arms. However, there is variability in the structures of the N-terminal

arms among phage capsid proteins. Therefore, more high-resolution structures of portal complex–capsid interfaces are needed to determine whether other tailed phages use a mechanism similar to that of P68 for incorporating portal complexes into their capsids.

In the asymmetric reconstruction of the P68 virion, the portal complex exhibits 12-fold symmetry. This is in agreement with the previous studies of phage P22, for which it was speculated that its portal complex can exist in two conformations: an asymmetric one that binds the large terminase to enable genome packaging and a symmetric one present in the mature virion (31).

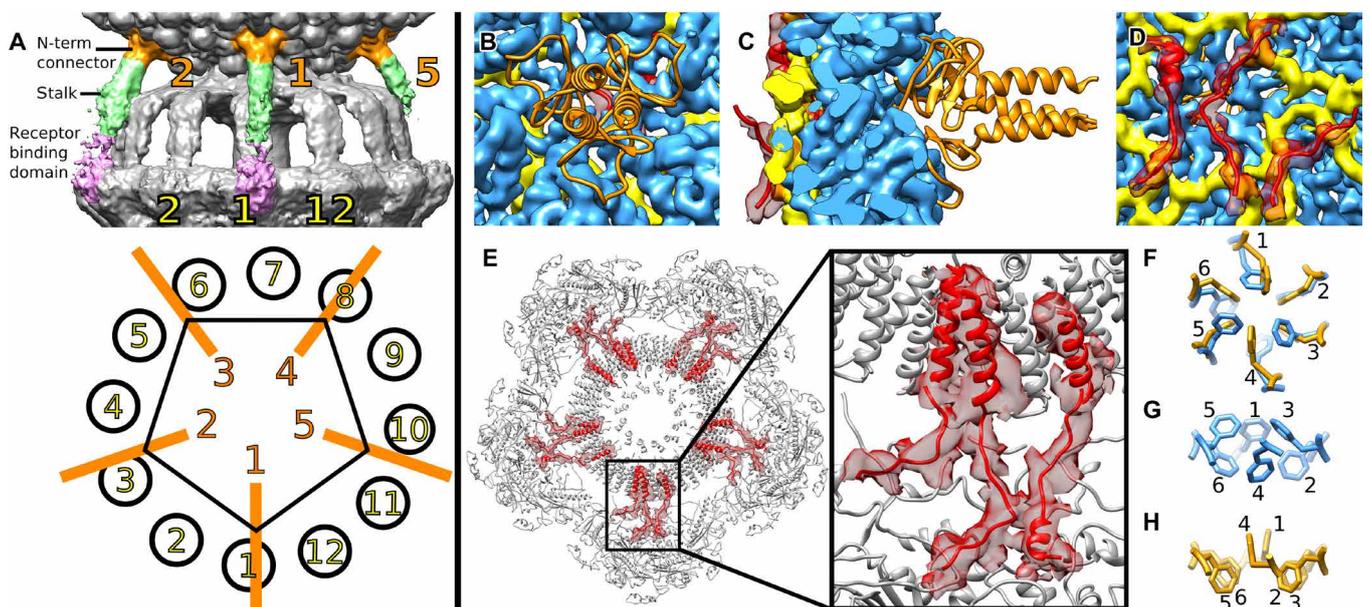
### Inner core proteins interact with the capsid and determine the attachment sites of head fibers

The surface of the portal complex facing toward the center of the P68 head is covered by 72 subunits of the inner core protein gp22 (Figs. 1B and 3, A and B). The structures of the inner core proteins were determined as part of the reconstruction of the portal complex from the native P68 virion with imposed 12-fold symmetry at a resolution of 3.9 Å. Only residues 91 to 114, which form an  $\alpha$  helix, are resolved from each 147-residue-long inner core protein. The six inner core proteins associated with each portal protein form two three-helix bundles related to a quasi-twofold rotational axis (Fig. 3A). However, the asymmetric reconstruction of the P68 virion at a resolution of 4.7 Å shows that the inner core proteins positioned closest to the capsid have additional structured residues that interact with axial domains of the adjacent major capsid proteins and, by modifying their conformations, enable the attachment of head fibers

to the capsid, as discussed below. Inner core proteins detach from the portal complex during the phage DNA release (Fig. 1C) and contain predicted pore-lining helices (fig. S3, I and J), indicating that they may enable the transport of the phage DNA across the bacterial cytoplasmic membrane.

### Head fibers can position P68 particles for genome delivery at cell surface

The P68 head is decorated with five trimers of head fibers gp14, which are attached to the hexamers of major capsid proteins adjacent to the tail vertex (Figs. 1A and 4, A to C). Structures of head fibers are resolved in the asymmetric reconstruction of the P68 virion, which was determined to an overall resolution of 4.7 Å; however, the distant parts of the head fibers are less well resolved (Figs. 1A and 4A). Because of the mismatch of the 5-fold symmetry of the head and 12-fold symmetry of the tail, only three head fibers are stabilized by interaction with the tail fibers (Fig. 4A). P68 head fibers can be divided into the N-terminal capsid-binding domain (residues 1 to 55),  $\alpha$ -helical stalk (residues 56 to 339), and receptor binding domain (residues 340 to 481), which is positioned at the level of the receptor binding domains of tail fibers (Fig. 4A). A cryo-EM map of the fivefold symmetrized head of the native P68 virion at a resolution of 3.8 Å enabled the building of the poly-alanine structure of 55 residues of the capsid-binding domain of the head fiber, which is composed of three  $\beta$  sheets and an  $\alpha$  helix (Fig. 4, B and C). Residues of the  $\beta$  sheets mediate the attachment of the head fiber to a hexamer of major capsid proteins (Fig. 4C). The  $\alpha$  helices form a



**Fig. 4. Head of P68 is decorated with five head fibers attached to hexamers of major capsid proteins located next to tail vertex.** (A) P68 head is decorated with five head fibers that extend toward tail fibers. The head fibers can be divided into the N-terminal connector shown in orange, stalk in green, and receptor binding domain in pink. Because of the mismatch of the 5-fold symmetry of the head and 12-fold symmetry of the tail, only fibers 1, 2, and 4 are stabilized by interactions with tail fibers. (B to D) N-terminal connector domains of head fibers (shown in cartoon representation in orange) are attached to hexamers of major capsid proteins (shown as blue density). Cryo-EM density of inner capsid proteins is shown in yellow, and arms of inner core proteins, which interact with major capsid proteins, are shown in cartoon representation in red. Cryo-EM density of inner core proteins is shown as semitransparent red surface. External view of P68 head (B), section through capsid (C), and internal view of capsid (D). (E) Section through P68 head perpendicular to tail axis at level of inner core complex. Inner core proteins that interact with major capsid proteins are highlighted in red. The electron density of the inner core proteins is shown as a red semitransparent surface. (F to H) Details of organization of Phe<sup>259</sup> side chains around quasi-sixfold axis of hexamer of major capsid proteins. In hexamers that interact with inner core proteins, side chains (in blue) are organized with threefold symmetry in alternating up and down conformations (F and G). In hexamers of capsid proteins that do not interact with inner core proteins, Phe<sup>259</sup> side chains are organized with twofold symmetry, with two side chains pointing into capsid and four out (F and H).

coiled coil that enables trimerization of the head fibers (Fig. 4, B and C). The selectivity of binding of the head fibers to the hexamers of major capsid proteins adjacent to the tail vertex is determined by interactions of the inner core proteins with the inner face of the capsid (Fig. 4, D and E). Fifteen of the 72 inner core proteins present in the P68 head form structured “arms” that reach the axial domains of the closest hexamers of major capsid proteins (Fig. 4E). The interaction with the inner core proteins forces the side chains of Phe<sup>259</sup>, from the axial domains of the major capsid proteins, to adopt a threefold symmetrical alternating “in and out” conformation (Fig. 4, F and G, and fig. S3G). In contrast, in hexamers of the major capsid proteins that do not interact with the inner core proteins, two Phe<sup>259</sup> side chains point toward the center of the head and four side chains point away from the particle center (Fig. 4, F and H, fig. S3H). In summary, the binding of the inner core proteins causes a change of symmetry of the six Phe<sup>259</sup> side chains from twofold to threefold and thus defines the attachment sites for the head fiber trimers.

Residues 340 to 477 of the head fiber are homologous to the receptor binding proteins of lactococcal phages TP901-1, P2, and bIL170 (table S3) (32–34). The putative receptor binding domains of P68 head fibers are positioned next to the receptor binding domains of tail fibers (Fig. 4A). Therefore, the binding of head fibers to receptors can position P68 with its tail orthogonal to the cell surface for genome delivery. This is supported by the broader host range of phage P68 in comparison to the closely related phage 44AHJD, which lacks the gene for the head fiber (23, 35). In contrast, the head fibers of previously structurally characterized phages point in all directions and are thought to function in the reversible attachment of phages to cells in random orientations.

### Lower collar complex and tail needle of P68 tail

The structure of the 12-fold symmetrized tail of the P68 virion was determined to a resolution of 3.9 Å. The lower collar complex is attached to the portal complex and forms the central part of the P68 tail (Figs. 1, A and B, and 5A). The dodecamer of lower collar proteins (gp18) has the shape of a mushroom with a head diameter of 162 Å and a total length of 146 Å (Fig. 5, B to D). It contains an axial channel that is continuous with that of the portal complex (Figs. 1B and 5E). The structure of the lower collar protein could be built except for residues 1 and 154 to 186 out of 251. It can be divided into three parts: the curly domain (residues 3 to 116 and 222 to 251), tube domain (residues 116 to 154 and 184 to 222), and knob connector loop (residues 154 to 184) (Fig. 5D). The curly domain, formed by six  $\alpha$  helices, mediates the attachment of the lower collar complex to the portal complex and tail fibers. The tube domain is composed of two antiparallel  $\beta$  strands (Fig. 5, C and D). Twelve tube domains form a  $\beta$  barrel with 24  $\beta$  strands, which is 108 Å long (Fig. 5C). The knob connector loops enable the attachment of the tail knob complex, with sixfold symmetry, to the lower collar complex.

Portal and lower collar complexes of P68 form a channel with a total length of 270 Å (Fig. 5E). The inner diameter of the channel varies from 30 to 55 Å. The surface charge distribution inside the channel is mostly negative, but it is interrupted by neutral and positively charged layers (Fig. 5E). The mostly negative charge of the inner surfaces of the complexes enables ejection of the phage genome.

The central channel of the lower collar complex and the proximal half of that of the tail knob are filled by the tail needle trimer (Fig. 5F). Asymmetric reconstruction of the P68 particle at a resolution of 4.7 Å did not enable building of the atomic structure of the

tail needle trimer. However, the structure of the tail needle protein of *Salmonella enterica* phage P22 fits the tail needle density of P68 with a correlation coefficient of 0.8 (Fig. 5F). On the basis of the high content of  $\alpha$ -helical secondary structure elements, we speculate that the tail needle of P68 is made of gp8. It has been shown that the tail needle of bacteriophage P22 is essential for stabilizing packaged DNA inside the virion (22), and there is evidence that it enables DNA delivery by interacting with bacterial membranes.

### Tail knob and tail spike

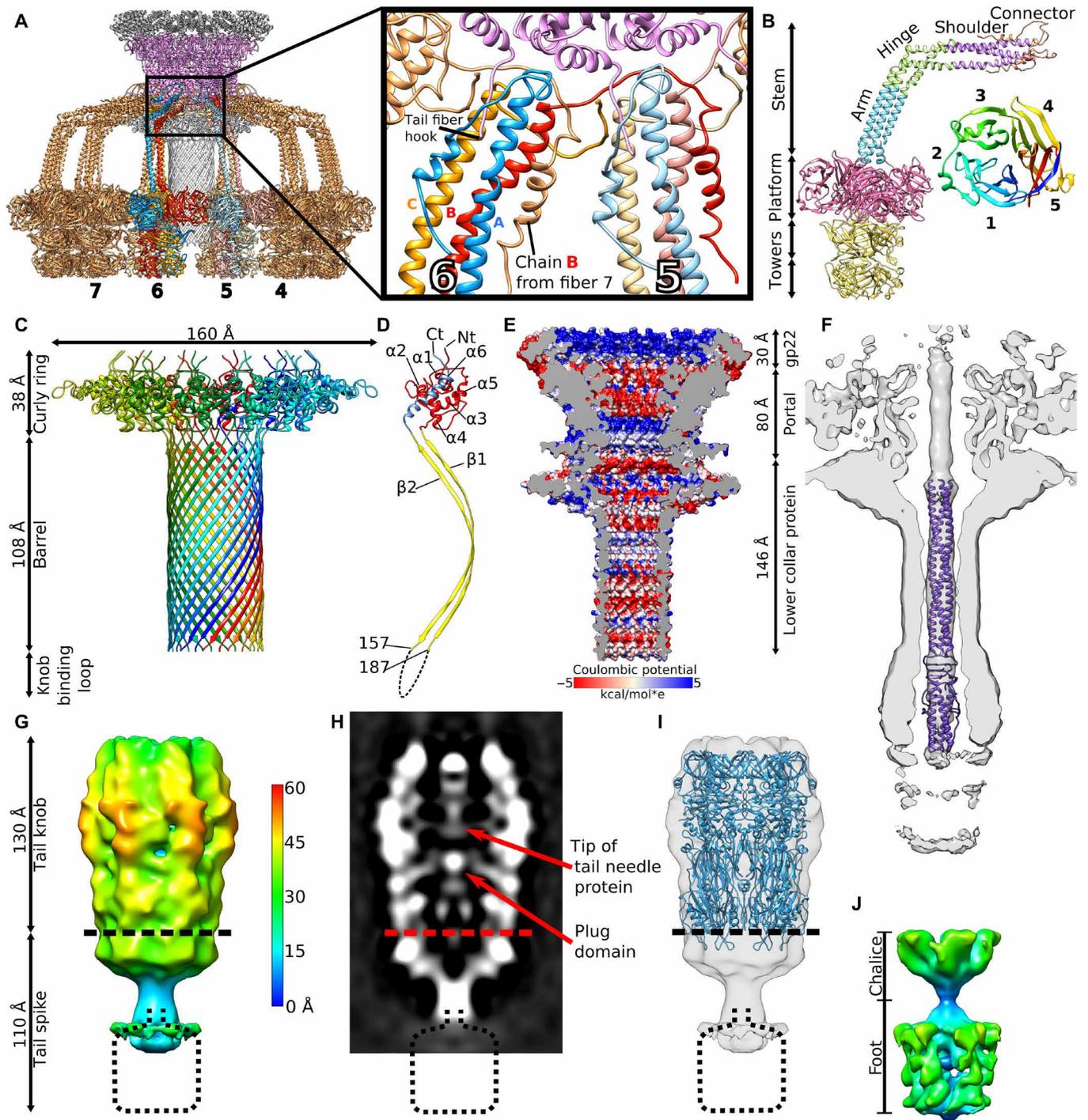
The tail of P68 continues beyond the lower collar protein by tail knob gp13 and tail spike gp11 (Fig. 1, A and B). The two complexes were reconstructed to resolutions of 11 and 7 Å, respectively, indicating that they are more flexible than the parts of the tail near the phage head (Fig. 5, G to J). This flexibility may be required to allow the putative cell wall-degrading enzymes located in the tail spike to cleave a pore in the bacterial cell wall to enable genome delivery.

The previously determined crystal structure of the tail knob of *Streptococcus* phage C1 fits into the reconstruction of the corresponding part of the P68 tail with a correlation coefficient of 0.65 (Fig. 5I) (table S3) (10). The length of the P68 tail knob is 130 Å along the tail axis. It has an outer diameter of 80 Å and an inner tube diameter of 40 Å. The channel is continuous with that of the lower collar protein (Fig. 1B). One-third of tail knob proximal to the head of native P68 contains a tubular density that probably belongs to the trimer of tail needle protein (Figs. 1B and 5, F and H), which is covalently linked to the end of P68 DNA (23).

An asymmetric reconstruction of the tail spike provides evidence that it has fivefold symmetry (fig. S5, F to H). Fivefold symmetrized, localized reconstruction of the tail spike shows that it is 110 Å long and has a maximum diameter of 70 Å (Fig. 5J). It can be divided into the chalice, which mediates attachment to the tail knob, and the distal lysis domain (Fig. 5J). Sequence comparisons indicate that the tail spike of P68 is homologous to the PlyCb lysin from phage C1 (table S3) (36). However, the structure of PlyCb does not fit into the reconstruction of the P68 tail spike (37). Other proteins with peptidoglycan degradation activities such as the amidase from *S. aureus*, peptidases from *Staphylococcus saprophyticus*, and endolysin from staphylococcal phage K are homologous to the last 130 amino acids of the P68 tail spike protein (table S3). This indicates that the tail spike proteins of P68 degrade the bacterial cell wall to enable access of the phage to the cytoplasmic membrane.

### Tail fibers

The tail fibers of P68 form a skirt-like structure around the tail (Figs. 1, A and B, and 5A). Each tail fiber is a trimer of 647-residue-long gp17 subunits (Fig. 5B). The tail fiber can be divided into the N-terminal stem domain (residues 1 to 145), platform (residues 151 to 445), and C-terminal tower (residues 446 to 647) (Fig. 5B). Cryo-EM reconstruction of the 12-fold symmetrized tail of P68 virion at a resolution of 3.9 Å enabled the building of the structure of the stem domain, which can be further subdivided into a connector (residues 1 to 45), shoulder (residues 46 to 80), hinge (residues 81 to 115), and arm (residues 116 to 145) (Fig. 5B). Because of the asymmetric shape of the tail fiber, the three constituent subunits (A, B, and C) differ in structure from each other (Fig. 5B). Functionally important differences are found in the connector regions that mediate the attachment of the tail fiber to the portal and lower collar complexes (Fig. 5A). The connector domain of subunit A and shoulder from subunit C



**Fig. 5. Structure of P68 tail.** (A) Tail fibers of P68 form skirt around tail tube. Tail fibers are shown in gold; however, individual subunits of two tail fibers are distinguished in red, blue, and orange. Portal proteins are shown in magenta, inner core proteins in dark gray, and lower collar proteins in light gray. The inset shows details of interactions of tail fibers with each other and their attachment to the portal complex. (B) Structure of P68 tail fiber trimer in cartoon representation and its division into domains. The structure of the platform and tower domains was solved by x-ray crystallography to a resolution of 2.0 Å and fitted into the cryo-EM map of the native P68 tail. The inset shows a cartoon representation of the platform domain of the P68 tail fiber rainbow-colored from N terminus in blue to C terminus in red. (C) Structure of lower collar complex with individual subunits distinguished by rainbow coloring. (D) Division of lower collar protein into domains. The curly ring domain is shown in red and blue, the barrel domain is shown in yellow, and the dashed line indicates the knob-binding loop with unknown structure. (E) Surface charge distribution in inner core, portal, and lower collar complex of native P68 particle. (F) Fit of structure of tail needle of phage P22 into 12-fold symmetrized reconstruction of native P68 tail. (G) Sixfold symmetrized reconstruction of P68 tail knob and tail spike complexes. The surface of the cryo-EM map is radially colored on the basis of the distance from the sixfold axis of the complex. (H) Distribution of electron density in central section of tail knob and tail spike complexes. (I) Fit of structure of tail knob of phage C1 into P68 reconstruction. (J) Structure of tail spike with imposed fivefold symmetry shows its chalice and foot domains.

form a noose-like structure that encircles the N-terminal tail fiber hook of the portal protein (Fig. 5A). The connector of subunit B binds to the shoulder domains of subunits A and B from the tail fiber positioned counterclockwise when looking at the tail from the direction of the head (Fig. 5A). The first structured residue of the connector of subunit C (Thr<sup>24</sup>) is located between the clamp domain of subunit B from the tail fiber positioned clockwise and the shoulder domain of subunit C and the clamp domain of subunit A positioned counterclockwise (Fig. 5A). Thus, the N terminus of C subunit mediates interactions between tail fibers that are one position removed from each other.

The shoulder domain of the tail fiber is straight until the hinge domain, which introduces a turn of 110° (Fig. 5B). The hinge of subunit C is formed by two  $\alpha$  helices connected by a short loop, which allows the chain to bend and pass under subunits A and B (Fig. 5B). After the hinge, the three subunits form a straight coiled-coil arm (Fig. 5B). The cryo-EM reconstruction of the P68 tail is complemented by the crystal structure of the tail fiber protein determined to a resolution of 2.0 Å (table S4). Although the full-length tail fiber protein was used for crystallization, only the platform and tower domains (residues 139 to 647) were resolved (table S4). The combination of cryo-EM and x-ray results allowed construction of the complete tail fiber.

The platform domain of the P68 tail fiber has a five-bladed  $\beta$ -propeller fold (Fig. 5B). Each of the blades contains four antiparallel  $\beta$  strands. The domain is cyclically enclosed, since the first N-terminal  $\beta$  strand of the domain is part of the same blade as the last three C-terminal  $\beta$  strands (Fig. 5B). It has been shown that the platform domains of various phages, including phi11, PRD1, and PhiKZ, contain receptor binding sites (38–40). The platform of the P68 tail fiber is similar in structure to that of staphylococcal phage phi11 from the family Siphoviridae, with an RMSD (root mean square deviation) of the corresponding C $\alpha$  atoms of 1.10 Å. The sequence identity of the two proteins is 24%. There are differences in the receptor binding sites within the platform domains of the two phages that may reflect their different receptor requirements (fig. S5, A to D). Whereas P68 binds to wall teichoic acid glycosylated with  $\beta$ -O-N-acetyl-glucosamine, phi11 can attach to both  $\beta$ -O-N-acetyl-glucosamine and  $\alpha$ -O-N-acetyl-glucosamine (16).

The tower domain of the P68 tail fiber is composed of two subdomains (residues 454 to 555 and 556 to 645), which are structurally similar to each other (Fig. 5B). Each subdomain is formed of a four-stranded antiparallel  $\beta$  sheet connected by loops and two short helices (Fig. 5B). The  $\beta$  sheets are positioned close to the threefold axis of the fiber, whereas the loops and helices are exposed at the surface. The subdomains are homologous to putative major teichoic acid biosynthesis protein C, muramidases, and receptor binding fibers of R-type pyocin (table S3). Thus, the tower region may be involved in binding to the cell wall or peptidoglycan digestion. Compared to the tail fiber of phage phi11, the platform and tower domains of P68 exhibit domain swapping within the trimer of the tail fiber (fig. S5E) (38).

The 12-fold symmetrized cryo-EM structure of native P68 tail at a resolution of 3.9 Å shows interactions of residues 229 to 271 of platform domain of subunit B with residues 349 to 386 of chain A and residues 180 to 268 of chain B of platform domains from the neighboring tail fiber (Fig. 5A). The interface has a buried surface area of 1100 Å<sup>2</sup>, and it is likely that it stabilizes the “skirt” structure of P68 tail fibers. The flexibility of the hinge region of tail fibers was proposed to facilitate receptor binding in other phages (38). In contrast, in P68, the structure of tail fibers appears to be rigid.

### Changes in P68 particles associated with genome release

The genome release of P68 is connected to the following major conformational rearrangements of its neck region: (i) Helix  $\alpha$ 9 from the wing domain of the DNA binding portal protein subunit (Fig. 3, C to E) detaches from the side of the DNA, and the portal complex assumes 12-fold symmetry (fig. S4, C and D). (ii) The 12-fold symmetrized reconstruction of the portal complex of the empty particle, which was determined to a resolution of 3.9 Å, contains no density for residues 257 to 267 and 307 to 327 from the wing domain of the portal protein (fig. S4, C and D). The previously determined structure of isolated portal complex of *Bacillus* phage phi29 also lacked resolved electron density for the corresponding residues from the wing domain. It is therefore possible that these residues only become structured in the presence of DNA in the portal channel. (iii) The inner core proteins are ejected from the phage head (Fig. 1, B to D). Changes in the structure of the wing domains of portal proteins associated with the decreased pressure inside the phage head after partial DNA ejection may trigger release of the inner core proteins.

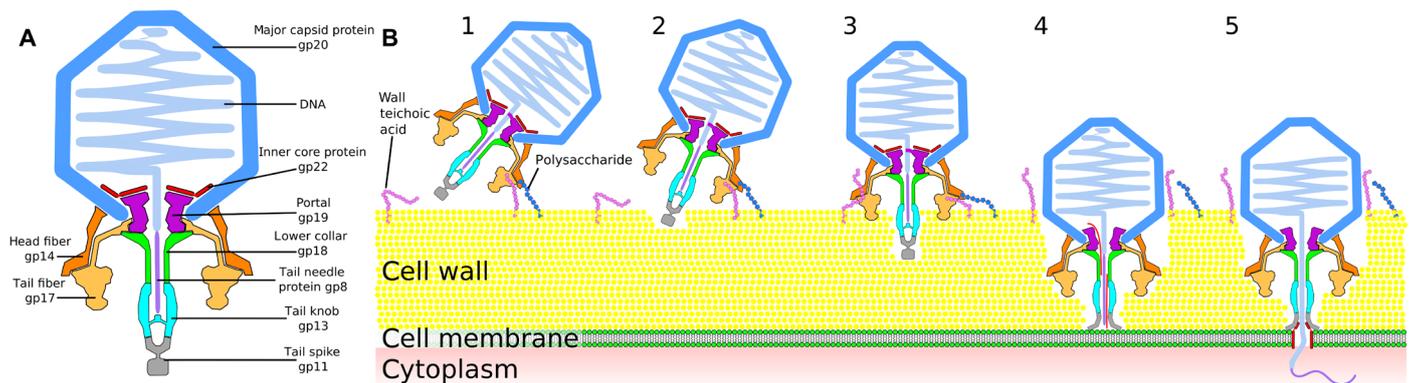
Phages have to deliver their genomes across the cell wall and cytoplasmic membrane to infect Gram-positive bacterial cells. It has been shown that, after cleavage of the peptidoglycan cell wall, phages from the family Podoviridae, including phi29, release proteins that form pores in the cytoplasmic membrane. The pore-forming proteins attach to the end of the tail tube and enable delivery of the phage DNA into the cytoplasm (41). Unlike virions of phage phi29, particles of P68 do not bind to liposomes at acidic pH (fig. S1, I to P) (41); instead, they aggregate with each other through their tails (fig. S1, I to P). However, liposomes in the mixture with P68 become distorted (fig. S1, I to P). It is possible that the ejected inner core proteins, which contain predicted pore-lining helices (fig. S3, I and J), interfered with the liposome integrity.

The heads of P68 particles in the process of genome release contain shells of packaged DNA that are spaced 26 to 30 Å apart, whereas in the full virions, the DNA spacing is 20 Å (Fig. 1, B and C, and fig. S3F). The resolved structure of the layers of dsDNA in the P68 genome release intermediate indicates that all the particles released similar amounts of DNA and provides evidence of a gradual relaxation of the DNA pressure during the genome release.

Only 2% of P68 genome release intermediates and empty particles retained their tail knobs and tail spikes in vitro (fig. S1, A to G). However, the tail knobs and tail spikes of the complete empty particles contain central channels (Fig. 1D), indicating that the complexes may remain attached to P68 virions during genome release in vivo.

### Attachment of P68 virions with *S. aureus* cell wall

Electron micrographs of P68 virions with fragments of *S. aureus* cell wall show that most of the P68 virions bind with the axes of their tails at an acute angle to the surface of the cell wall (fig. S1H). This orientation of the tail is not suitable for the delivery of the phage DNA into a cell. The initial binding of P68 virion to the cell wall may be mediated by the head fibers; however, individual head fibers cannot be distinguished in the electron micrographs. The predicted receptor binding sites of tail fibers are located at the sides of the platform domains rather than at the tips of the tower domains (Fig. 5B and fig. S5, A to D). Therefore, the receptor binding sites are located at the sides of the skirt formed by the tail fibers of P68 (Figs. 1A and 5, A and B). The tail knob and tail spike form a 20-nm-long protrusion at the center of the tail fiber skirt (Fig. 1, A and B). Therefore, the P68 virion has to locally degrade the cell wall, by the hydrolase



**Fig. 6. Mechanism of P68 genome delivery into *S. aureus* cell.** (A) Diagram of P68 virion. (B) Mechanism of P68 genome delivery. P68 virion attaches to cell surface by head or tail fibers (1). This attachment allows enzymes from tail spike to cleave bacterial cell wall (2). This degradation of *S. aureus* cell wall enables P68 to bind with its tail axis perpendicular to cell surface (3). Further cell wall digestion allows tip of P68 tail to reach cytoplasmic membrane, which triggers release of inner core proteins and DNA (4). Inner core proteins form channel in membrane for ejection of phage DNA into bacterial cytoplasm (5).

activity associated with its tail spike, to produce a cavity into which the tip of the tail knob can be inserted so that the phage can bind with its tail axis perpendicular to the cell wall.

### Mechanism of P68 genome delivery

The structures of the P68 virion, its genome release intermediate, and empty particle in combination with the observation of the attachment of P68 virions to the *S. aureus* cell wall enabled us to propose the mechanism of P68 genome delivery (Fig. 6). P68 virion binds to the *S. aureus* cell by either head or tail fibers. At the moment of the initial binding, the tail axis of the phage is not perpendicular to the cell surface (Fig. 6 and fig. S1H). After the attachment, tail spike proteins degrade the cell wall, which allows the phage to position itself with its tail axis perpendicular to the cell surface (Fig. 6). The unique attachment of P68 head fibers exclusively to the hexamers of capsid proteins located next to the tail vertex enables the phage to use its head fibers for orienting itself with the tail axis perpendicular to the cell surface. Further cell wall digestion enables the tip of the P68 tail to reach the *S. aureus* cytoplasmic membrane. The signal triggering P68 genome release is unknown; however, it may be the binding of the tail spike to a receptor in the membrane, exposure of the tail spike to the hydrophobic environment of the membrane, or a sensing of the transmembrane potential. Subsequently, conformational changes of the P68 portal and disruption of the unique interaction of one of the portal protein subunits with the end of the phage genome enable ejection of the tail needle protein, inner core proteins, and DNA through the tail channel. The tail needle and inner core proteins may form a pore in the bacterial membrane for delivering phage DNA into the bacterial cytoplasm (Fig. 6).

The quasi-equivalent structure of the  $T = 4$  icosahedral capsid includes conformational differences in the major capsid proteins from the icosahedral asymmetric unit (Fig. 2, B and C). The N-terminal arms and extended loops are in one plane in the major capsid proteins that connect two hexamers (Fig. 2C). In contrast, the same domains are bent  $16^\circ$  in the major capsid proteins that form pentamers and  $8^\circ$  in subunits that connect hexamers to pentamers (Fig. 2C). Additional differences among the capsid proteins are in the structures of residues 253 to 263 from the axial domain, which fold into  $\alpha$  helices in subunits that form pentamers and loops in subunits that belong to hexamers (Fig. 2D).

### MATERIALS AND METHODS

#### P68 growth and purification

Bacteriophage P68 was purchased from the Félix d'Hérelle Reference Center for Bacterial Viruses (Université Laval, Québec, Canada). The propagation strain of *S. aureus* RN4220  $\Delta$ tarM (16) was provided by A. Peschel from the Department of Infection Biology, University of Tübingen. Phage P68 was propagated on *S. aureus* RN4220  $\Delta$ tarM grown at  $37^\circ\text{C}$  in meat peptone broth [13 g of nutrient broth CM1 (Oxoid), 3 g of yeast extract powder L21 (Oxoid), and 5 g of peptone L37 (Oxoid) were dissolved in distilled water to a final volume of 1000 ml, and the pH was adjusted to 7.4 using 10 M NaOH].

Phage lysate from 300 ml of bacterial culture was centrifuged at  $5000g$  for 30 min at  $4^\circ\text{C}$ . The resulting supernatant was filtered through a  $0.45\text{-}\mu\text{m}$  polyether-sulfone syringe filter (Techno Plastic Products, Switzerland) to remove bacterial debris. Phages were pelleted by centrifugation at  $64,000g$  for 2.5 hours at  $4^\circ\text{C}$ . The resulting pellets were resuspended in  $350\ \mu\text{l}$  of phage buffer [50 mM tris (pH 8.0), 10 mM  $\text{CaCl}_2$ , and 10 mM NaCl] by mild shaking overnight at  $5^\circ\text{C}$ . The resulting solution was mixed with an equal amount of chloroform by gently inverting the tube  $10\times$ . The mixture was centrifuged at  $3000g$  for 10 min at room temperature. The aqueous fraction from the chloroform mixture was overlaid onto a preformed CsCl step density gradient (1 ml of each 1.45 g/ml, 1.50 g/ml, and 1.70 g/ml of CsCl in phage buffer) and centrifuged at  $194,000g$  for 4 hours at  $12^\circ\text{C}$  using an SW55Ti rotor (Beckman Coulter). Phage particles forming a visible band were collected with an 0.8-mm gauge needle and syringe. Cesium chloride was removed from the phage-containing fraction by dialysis against a  $5000\times$  excess of phage buffer at  $4^\circ\text{C}$  overnight using Visking dialysis tubing type 8/32", 0.05 mm thick (part no. 1780.1, Carl Roth, Germany).

#### Liposome preparation

Liposomes were prepared as described previously, with minor modifications (41). The phospholipids (Avanti Polar Lipids) were resuspended in chloroform and mixed in a molar ratio that mimicked the composition of the cell membrane of *S. aureus* (41): phosphatidylethanolamine, phosphatidyl-DL-glycerol, phosphatidylcholine, and cholesterol in a molar ratio of 5:6:2:9. The chloroform-liposome mixture was dried for 4 hours in a 200-ml evaporation flask in a rotary vacuum evaporator at  $37^\circ\text{C}$  and 20 rpm to create a homogenous

lipid film. The lipid film was resuspended in phage buffer [50 mM tris (pH 8.0), 10 mM CaCl<sub>2</sub>, and 10 mM NaCl], flash-frozen and thawed five times, and sonicated for 120 s using an ultrasonic cleaner (Sonica). Subsequently, the solution was passed 20 times through a membrane with 400-nm pores in the LiposoFast Basic Extruder (Avestin).

### Phage-liposome interaction

The genome ejection of P68 was induced by exposing it to pH 4.2, as described previously for phage phi29 (41). P68 at a concentration of 2 mg/ml in phage buffer [50 mM tris (pH 8.0), 10 mM CaCl<sub>2</sub>, and 10 mM NaCl] was mixed with liposomes at a concentration of 5 mM in phage buffer in a volume ratio of 1:10. The pH of the mixture was changed to acidic by three rounds of concentrating and diluting with 0.1 M sodium acetate, 300 mM ammonium sulfate (pH 4.2) buffer using 100-kDa Amicon ultra centrifugal filters. The phage was incubated in the acidic buffer for 1 hour. The mixture was applied to holey carbon-coated copper grids (R2/2, mesh 200; Quantifoil), blotted, and flash-frozen in liquid ethane.

### Identification of structural proteins of phage P68

The purified P68 virus was resuspended in Laemmli buffer and boiled for 3 min, and the proteins were separated by tricine gradient gel electrophoresis. All protein bands were cut from the gel and used for mass spectrometry (MS) analysis. In addition, P68 at a concentration of 2 mg/ml in phage buffer containing 1% SDS was boiled for 3 min and analyzed by MS. After destaining and washing, the proteins were subjected to trypsin digestion (sequencing grade, Promega). MALDI (matrix-assisted laser desorption/ionization)-MS and MS/MS analyses were performed on an Ultraextreme mass spectrometer (Bruker Daltonics, Bremen, Germany). Data processing and analysis were performed with the software FlexAnalysis 3.4 and MS BioTools (Bruker Daltonics). Mascot software (Matrix Science, London, UK) was used for sequence searches in exported MS/MS spectra against the National Center for Biotechnology Information database and a local database supplied with the expected protein sequences. The mass tolerance of peptides and MS/MS fragments for MS/MS ion searches were 50 parts per million and 0.5 Da, respectively. The oxidation of methionine and propionyl-amidation of cysteine as optional modifications and one enzyme miss-cleavage were set for all searches. Peptides with a statistically significant peptide score ( $P < 0.05$ ) were considered.

### Cryo-EM sample preparation, data acquisition, and initial data processing

Solution containing the phage (3.8  $\mu$ l) at a concentration of 2 mg/ml was pipetted onto holey carbon-coated copper grids (R2/2, mesh 200; Quantifoil), blotted, and vitrified by plunging into liquid ethane using FEI Vitrobot Mark IV. The vitrified sample was transferred to an FEI Titan Krios electron microscope operated under cryogenic conditions and at an acceleration voltage of 300 kV. The illuminating beam was aligned for parallel illumination in NanoProbe mode, and coma-free alignments were performed to remove residual beam tilt. Imaging was done under low-dose conditions with a total dose of  $21 \text{ e}^-/\text{\AA}^2 \text{ s}^{-1}$ . Data were collected with underfocus values ranging from  $-1.0$  to  $-3.0 \mu\text{m}$ . Data acquisition was performed at a nominal magnification of  $\times 75,000$ , resulting in a calibrated pixel size of 1.063  $\text{\AA}$ . Micrographs were acquired using a Falcon II direct electron detector operated in a movie mode. One-second exposure was fractionated into seven frames and saved as separate files. Automated data

acquisition was done using the acquisition software EPU (FEI). The micrographs were collected in two separate sessions (1405 and 1486 micrographs) with the same microscope with identical settings. Micrographs from both sessions were merged and processed together.

The seven-frame movies were aligned globally and locally ( $5 \times 5$  patch) using the software MotionCor2 (42), and the defocus values were estimated from aligned micrographs using the program CtfFind4 (43). Power spectra of Fourier transforms of micrographs were visually inspected, and micrographs with distorted or missing Thon rings were discarded.

### Icosahedral reconstruction of P68 capsid

P68 heads were manually boxed using e2boxer.py from the software package EMAN 2.1 (44). Images of 37,218 particles ( $600 \times 600$  pixels) were extracted from the micrographs and background-normalized using RELION 2.1 (45). Multiple rounds of 2D classification of particles were performed using RELION (45). Only particles belonging to high-resolution class averages were used for subsequent reconstruction (36,853 particles). To generate the initial model, we used the random de novo model method as implemented in EMAN2 (44). Particles were divided into two independent sets. From each half, nine subsets of  $4 \times$  binned images were generated. Each set contained 1000 particles. The particles were assigned random orientations and iteratively reconstructed into 3D volumes using the software package jspr (46). Independent initial models were chosen from both half-datasets by selecting the two models most similar to each other. Afterward, the initial models were unbinned  $4 \times$  to match the size of the original images and low-pass-filtered to a resolution of 40  $\text{\AA}$ . The refinement was performed using the program RELION 2.1 (45). After initial auto refinement, 3D classification was performed without the alignment step. Particles belonging to the best class (21,625 particles) were selected for further RELION auto refinement with imposed icosahedral symmetry and maximum allowed deviations from previous orientations of  $10^\circ$ . The resulting map was threshold-masked, divided by the modulation transfer function, and B-factor sharpened during the post-processing in RELION (45).

Capsids of genome release intermediates and empty particles were determined using the same reconstruction strategy. After 2D classification, 4040 and 10,259 images were available for the reconstructions of capsids of genome release intermediate and empty particles, respectively. After 3D classification, the numbers of particles were 2332 and 8580, respectively.

### Asymmetric reconstruction of native P68 virion

For the asymmetric reconstruction of the whole virion, the particles were extracted from the original micrographs using a box size of  $1200 \times 1200$  pixels. The images were  $2 \times$  binned in Fourier space using xmipp\_transform\_downsample (47), resulting in a pixel size of 2.126  $\text{\AA}$ . Initial 2D classification with a circular mask with a diameter of 1000  $\text{\AA}$  led to the selection of 34,717 particles for further processing. The structure of phage P22 (EMD-1222) scaled to 73% of its original size was used as the initial model for 3D reconstruction (48). Asymmetric refinement, using the auto refine procedure, was performed using the software package RELION. This was followed by 3D classification, which used orientations and center offset values from the previous round of refinement. The class with the best-resolved details contained 33,612 particles, which were used for further 3D refinement. A mask that covered the whole virion was generated using the programs UCSF Chimera, Segger, and

relion\_mask\_create (4-pixel expansion and 5-pixel soft edges) (45, 49, 50). The mask was used in the final rounds of refinement with local angular searches using the RELION auto refine procedure (45).

### Reconstruction of capsid with fivefold symmetry

Images of capsids of native virions (33,612 particles) from the 3D classification of the asymmetric reconstruction were used as the initial dataset for fivefold symmetrized reconstruction. Refinement with imposed C5 symmetry was followed by 3D classification without alignment. After 3D classification, 28,826 particles were retained for further reconstructions. These particles were re-extracted in boxes of  $600 \times 600$  pixels. A mask that included the capsid was created using the programs UCSF Chimera, Segger, and relion\_mask\_create.py (4-pixel expansion and 5-pixel soft edges) (45, 49, 50). Masked 3D refinement with local searches around the previously determined orientations was performed using the RELION auto refine procedure (45). Identical strategies were used for the reconstructions of fivefold symmetrized capsids of genome release intermediates and empty particles.

### Localized reconstruction of P68 portal and tail

Reconstruction of the P68 head with imposed fivefold symmetry provided Euler angles describing particle orientations with a phage tail along the positive direction of the coordinate  $z$  axis. The distance of the center of the tail from the center of the capsid was determined on the basis of the fivefold symmetrized map of the P68 virion using the program UCSF Chimera (49). The script localized\_reconstruction.py (51) was used to extract subparticles centered on the tail ( $512 \times 512$  pixels). 2D classification of the tails of native P68 virions resulted in the selection of 35,490 images. An initial model of the tail was calculated using relion\_reconstruct.py by applying the Euler angles determined in the C5 reconstruction of the capsid (45). The model of the tail with C5 symmetry was low-pass-filtered to  $40 \text{ \AA}$  and used as an initial reference for RELION auto refinement with imposed C12 symmetry. Segments of the reconstruction corresponding to the portal and tail were segmented using the programs UCSF Chimera and Segger and combined to generate a mask covering the portal and tail with the program relion\_mask\_create.py (4-pixel expansion and 5-pixel soft edges) (45, 49, 50). 3D classification with the applied mask, omitted alignment, and imposed C12 symmetry was used to select 21,702 particles. The particles were subjected to auto refinement and post-processing in RELION (45). The poly-alanine structure was manually built into the resolved parts of the P68 tail. The Protein Data Bank (PDB) was used to prepare an improved mask covering the portal complex, N-terminal parts of tail fibers, and lower collar protein, using the program e2pdb2mrc.py from the EMAN2 package and relion\_mask\_create.py (4-pixel expansion and 5-pixel soft edges) (44, 45). This mask was used in one more round of auto refinement, in which orientation and offset searches were limited to  $\pm 9^\circ$  and  $\pm 3$  pixels, respectively. The refinement strategy is summarized in fig. S6.

The reconstruction strategies of tails of P68 particles in the process of genome release and empty particles were identical to those of the full particle. For the genome release intermediates, the initial number of particles was 4040. After 3D classification, 1506 particles were selected for the final 3D auto refinement. The initial set of empty particles contained 10,288 images. After 2D classification, 6123 particles were selected. 3D classification resulted in the selection of 3246 particles.

2D classifications of the particles in the process of genome release and empty particles identified subsets of particles with intact tail knobs and tail spikes. 3D classifications of these subsets with imposed C12 symmetry yielded 98 and 112 particles for the genome release intermediate and empty particles, respectively.

### Localized reconstruction of P68 tail knob and tail spike

Side views of native P68 virions were selected for determining the structure of the tail knob and tail spike (21,016 particles). Boxed particles ( $1440 \times 1440$  pixels) were  $2\times$  binned using xmipp\_transform\_downsample (47). Projection images of a properly oriented reconstruction of the whole P68 virion, lacking the tail knob and tail spike, were subtracted from the particle images using the program relion\_project.py (45).

Parts of the images containing the tail knob were boxed ( $128 \times 128$  pixels) from the difference images using the program localized\_reconstruction.py (51). After 2D classification, 8589 particles were selected for reconstruction. An initial model of the tail knob was calculated using the particle orientations from the asymmetric reconstruction of the whole virion. 3D auto refinement in RELION was used to calculate the structure of the complex with imposed C6 symmetry. The resulting map was used to generate a mask (extended by 4 pixels and a soft edge of 5 pixels) using the program relion\_mask\_create.py (45). The final structure of the complexes was determined in 3D auto refinement with the use of the mask in the RELION package (45).

Parts of the images containing the tail spike were extracted from the same set of difference images as the tail knobs ( $128 \times 128$  pixels). After 2D classification, 6121 particles were selected for reconstruction. The initial model was calculated using the particle orientations from the asymmetric virion reconstruction. The asymmetric reconstruction indicated that the tail spike complex has fivefold symmetry; therefore, it was enforced during the subsequent rounds of structure refinement. A mask was generated from the tail spike map using the program relion\_mask\_create (extended by 4 pixels and a soft edge of 5 pixels) (45). The final rounds of reconstruction in RELION were performed with the use of the mask (45). The final map was threshold masked, MTF corrected, and B-factor sharpened using the program relion\_postprocess. The refinement strategy is summarized in fig. S6.

### Asymmetric reconstruction of portal

Asymmetric reconstruction of the portal complex of native P68 was determined using symmetry expansion from the C12 symmetrized localized reconstruction of the portal (21,702 particles). Initially, the particle orientations from the previous C12 reconstruction were symmetry expanded to C1 using the program relion\_particle\_symmetry\_expand.py (45). Mask covering the portal complex was generated from the C12 symmetrized structure. Focused 3D classification with 12 classes using the symmetry-expanded orientations was performed in RELION (45). The alignment step was omitted to prevent overfitting. Regularization factors in the range of 20 to 50 were tested to establish the optimal value to detect asymmetric features in the portal. The best reconstruction as judged by map interpretability was obtained with the regularization factor 33. The class from the 3D classification showing an asymmetric interaction between one of the portal protein subunits and dsDNA was chosen for further refinement. Duplicate particle images (C12 symmetry-related images of the same particle) were discarded from the reconstruction,

resulting in 7568 unique particles. Final refinement was performed using RELION auto refine, with the application of a mask, but without the orientation search (45). The refinement strategy is summarized in fig. S6.

### Cryo-EM structure determination and refinement

PDB structures were built manually using the program COOT (52). Maps of P68 capsids with imposed icosahedral symmetries were first rotated to the standard 222 orientation (one of the fivefold symmetry axes of the particle in the *xy* plane is rotated 31.717° from the *y* axis). The maps were cropped and normalized, the center of the particle was moved to the center of the box, and the map was converted to crystallographic space group *P23*. The map of the portal and tail with C12 symmetry was cropped, normalized, and interpolated to a crystallographic *P6* space group. The high-symmetry space groups were used to reduce the computational resources required for refinement of the structures.

Initially, the structures were built as poly-alanine chains. Noting the positions of residues with large side chains and aligning them to sequences of P68 structural proteins identified individual proteins in the reconstructions. Final models were iteratively refined in real space using the program PHENIX *real\_space\_refine.py* and, in reciprocal space in the program REFMAC5, corrected manually in COOT (52–54).

### Preparation of *S. aureus* cell walls

*S. aureus* strain RN4220 AtarM was grown in a meat peptone broth until the optical density at 600 nm ( $OD_{600}$ ) of 0.8. Cell culture (20 ml) was centrifuged at 8000g for 10 min. Resulting pellet was resuspended in a phage buffer to an  $OD_{600}$  of 40. Afterward, 1 ml of the cell suspension was mixed with 1 ml of Lysing Matrix B (MP Biomedicals) containing 0.1-mm silica spheres in a 2-ml FastPrep tube (MP Biomedicals). The 2-ml FastPrep tube was placed in the Microtube Homogenizer FastPrep-24 Classic Instrument (MP Biomedicals). The cells were homogenized for 5 min. The resulting lysate was centrifuged at 300g for 1 hour. The resulting supernatant was carefully pipetted to a 1.5-ml tube. The supernatant was centrifuged at 20,000g for 2 hours. The peptidoglycan appeared as a small pellet at the bottom of the tube. Last, the pellet was resuspended in 50  $\mu$ l of the phage buffer.

### Purification and crystallization of P68 tail fiber

The gene encoding the tail fiber of P68 in plasmid pMUH17 was a gift from U. Bläsi from Max F. Perutz laboratories in Vienna (35). The tail fiber protein was purified as described previously (35), with minor changes. The His-tagged tail fiber protein was expressed in *Escherichia coli* strain BL21 (DE3). After overnight cultivation in LB medium at 37°C and 200 rpm, the preculture was diluted 100 $\times$  into 1000 ml of fresh LB medium. The culture was grown at 37°C with 200 rpm shaking until an  $OD_{600}$  of 0.7 was reached. The expression was induced by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside to a final concentration of 1 mM. After 4 hours of protein expression at 37°C and 200 rpm shaking, the culture was centrifuged at 4000g for 30 min. The cell pellet was resuspended in 50 ml of lysis buffer [300 mM NaCl, 40 mM imidazole, 50 mM  $NaH_2PO_4$  (pH 8.0), 1 mM phenylmethylsulfonyl fluoride (PMSF), and lysozyme (1 mg/ml)], incubated for 1 hour at 4°C, and homogenized by sonication using a Qsonica Q700 sonicator (2-s sonication, 2-s pause, 4-min sonication, 50 W). The cell lysate was centrifuged at 20,000g for 40 min at 4°C and loaded onto a HisTrap column (GE Healthcare) equilibrated

with lysis buffer. Most of the impurities were eluted with wash buffer [70 mM imidazole, 300 mM NaCl, 50 mM  $NaH_2PO_4$  (pH 8.0), and 1 mM PMSF]. His-tagged tail fiber protein was eluted from the column using elution buffer [500 mM imidazole, 300 mM NaCl, 50 mM  $NaH_2PO_4$  (pH 8.0), and 1 mM PMSF]. The solution containing tail fiber protein was dialyzed against 25 mM tris-HCl (pH 7.5), 150 mM NaCl buffer, followed by separation in a HiLoad 16/600 Superdex 200 pg size exclusion column (GE Healthcare). The tail fiber protein eluted in the void volume of the column. The protein at a concentration of 5 mg/ml in 25 mM tris-HCl (pH 7.5), 150 mM NaCl buffer was used for crystallization screening. Crystals were obtained in a CrystalEX 96-Well conical bottom plate (Hampton Research) by the sitting-drop vapor diffusion method by mixing 0.5  $\mu$ l of protein solution with 0.5  $\mu$ l of 200 mM  $MgCl_2$ , 100 mM Hepes (pH 7.0), and 20% (w/v) PEG 6000 (polyethylene glycol, molecular weight 6000). Hexagonal crystals with dimensions up to 200  $\mu$ m formed within 30 days.

### X-ray data collection, structure determination, and refinement

Crystals of tail fiber protein in mother liquor solution were vitrified by plunging into liquid nitrogen. Diffraction data were collected at the SOLEIL synchrotron on beamline PROXIMA-1 equipped with a Pilatus 6M detector. The data were integrated and scaled using the package XDS (55). The model for molecular replacement was prepared by converting the tail fiber of phage phi11 (PDB 5EFV) to poly-alanine and rebuilding it according to the threefold averaged cryo-EM map of the P68 tail fiber (56). The quality of the cryo-EM map was sufficient for building the poly-alanine chain of residues 150 to 570, except for several loops in poorly resolved regions of the map. Molecular replacement was performed using the program Phaser (57). The trimers of the tail fiber proteins sit on threefold axes of the *R3* space group. Because of the crystal packing, the stem domain has to be bent nonphysiologically within the arm region. It is likely that the bending of the arm region differs among subunits within the crystal, which results in a local disorder. Therefore, the x-ray electron density map does not contain resolved features corresponding to the arm domain of the tail fiber protein. The structure was subjected to several rounds of automated building using the programs Phenix\_auto\_build and ARP/wARP combined with manual corrections in COOT and refinement in REFMAC5 (52, 54, 58).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/10/eaaw7414/DC1>

Fig. S1. Purified sample of P68 contains native virions, particles in process of genome release, and empty particles; attachment of P68 virions to *S. aureus* cell walls; and interactions of P68 virions with liposomes.

Fig. S2. Resolution and interpretability of cryo-EM reconstructions.

Fig. S3. Details of P68 head.

Fig. S4. Incorporation of P68 portal complex into capsid and changes in the structure of P68 portal complex upon genome release.

Fig. S5. Structures of P68 tail fiber and tail spike.

Fig. S6. Schemes of cryo-EM reconstruction strategies.

Table S1. Cryo-EM structure quality indicators.

Table S2. List of P68 proteins.

Table S3. HHpred searches for homologs of P68 tail fiber, head fiber, tail knob, and tail spike.

Table S4. X-ray structure quality indicators.

[View/request a protocol for this paper from Bio-protocol.](#)

## REFERENCES AND NOTES

- S. Chhibber, T. Kaur, K. Sandeep, Co-therapy using lytic bacteriophage and linezolid: Effective treatment in eliminating methicillin resistant *Staphylococcus aureus* (MRSA) from diabetic foot infections. *PLOS ONE* **8**, e56022 (2013).
- European Centre for Disease Prevention and Control/European Medicines Agency, "The bacterial challenge: Time to react" (ECDC/EMA Joint Technical Report, European Centre for Disease Prevention and Control/European Medicines Agency, 2009).
- J. Kurlenda, M. Grinholc, Alternative therapies in *Staphylococcus aureus* diseases. *Acta Biochim. Pol.* **59**, 171–184 (2012).
- World Health Organization, *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics* (World Health Organization, 2017).
- A. V. Aleshkin, J. Uchiyama, M. Osanai, N. Morimoto, T. Asagiri, T. Ujihara, M. Daibata, T. Sugiura, S. Matsuzaki, Experimental phage therapy against lethal lung-derived septicemia caused by *Staphylococcus aureus* in mice. *Microbes Infect.* **16**, 512–517 (2014).
- A. V. Aleshkin, O. N. Ershova, N. V. Volozhantsev, E. A. Svetoch, A. V. Popova, E. O. Rubalskii, A. I. Borzilov, V. A. Aleshkin, S. S. Afanas'ev, A. V. Karaulov, K. M. Galimzyanov, O. V. Rubalsky, S. S. Bochkareva, Phagebiotics in treatment and prophylaxis of healthcare-associated infections. *Bacteriophage* **6**, e1251379 (2017).
- P. G. Leiman, A. J. Battisti, V. D. Bowman, K. Stummeyer, M. Mühlenhoff, R. Gerardy-Schahn, D. Scholl, I. J. Molineux, The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. *J. Mol. Biol.* **371**, 836–849 (2007).
- B. Hu, W. Margolin, I. J. Molineux, J. Liu, The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* **339**, 576–579 (2013).
- Y. Xiang, M. C. Morais, A. J. Battisti, S. Grimes, P. J. Jardine, D. L. Anderson, M. G. Rossmann, Structural changes of bacteriophage phi29 upon DNA packaging and release. *EMBO J.* **25**, 5229–5239 (2006).
- A. A. Aksyuk, V. D. Bowman, B. Kaufmann, C. Fields, T. Klose, H. A. Holdaway, V. A. Fischetti, M. G. Rossmann, Structural investigations of a Podoviridae streptococcus phage C1, implications for the mechanism of viral entry. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14001–14006 (2012).
- W. Jiang, M. L. Baker, J. Jakana, P. R. Weigele, J. King, W. Chiu, Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. *Nature* **451**, 1130–1134 (2008).
- M. L. Baker, C. F. Hryc, Q. Zhang, W. Wu, J. Jakana, C. Haase-Pettingell, P. V. Afonine, P. D. Adams, J. A. King, W. Jiang, W. Chiu, Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12301–12306 (2013).
- W. Jiang, J. Chang, J. Jakana, P. Weigele, J. King, W. Chiu, Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**, 612–616 (2006).
- D.-H. Chen, M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, C. Haase-Pettingell, M. F. Schmid, W. Jiang, D. Baker, J. A. King, W. Chiu, Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1355–1360 (2011).
- G. C. Lander, R. Khayat, R. Li, P. E. Prevelige, C. S. Potter, B. Carragher, J. E. Johnson, The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit. *Structure* **17**, 789–799 (2009).
- X. Li, D. Gerlach, X. du, J. Larsen, M. Stegger, P. Kühner, A. Peschel, G. Xia, V. Winstel, An accessory wall teichoic acid glycosyltransferase protects *Staphylococcus aureus* from the lytic activity of Podoviridae. *Sci. Rep.* **5**, 17219 (2015).
- S. Steinbacher, U. Baxa, S. Miller, A. Weintraub, R. Seckler, R. Huber, Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10584–10588 (1996).
- S. Steinbacher, R. Seckler, S. Miller, B. Steipe, R. Huber, P. Reinemer, Crystal structure of P22 tailspike protein: Interdigitated subunits in a thermostable trimer. *Science* **265**, 383–386 (1994).
- Y. Xiang, P. G. Leiman, L. Li, S. Grimes, D. L. Anderson, M. G. Rossmann, Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. *Mol. Cell* **34**, 375–386 (2009).
- K. Stummeyer, A. Dickmanns, M. Muhlenhoff, R. Gerardy-Schahn, R. Ficner, Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat. Struct. Mol. Biol.* **12**, 90–96 (2005).
- A. S. Olia, P. E. Prevelige Jr., J. E. Johnson, G. Cingolani, Three-dimensional structure of a viral genome-delivery portal vertex. *Nat. Struct. Mol. Biol.* **18**, 597–603 (2011).
- A. Bhardwaj, R. S. Sankhala, A. S. Olia, D. Brooke, S. R. Casjens, D. J. Taylor, P. E. Prevelige Jr., G. Cingolani, Structural plasticity of the structural plug that traps newly packaged genomes in Podoviridae Virions. *J. Biol. Chem.* **291**, 215–226 (2016).
- D. Vybiral, M. Takáč, M. Loessner, A. Witte, U. von Ahnen, U. Bläsi, Complete nucleotide sequence and molecular characterization of two lytic *Staphylococcus aureus* phages: 44AHJD and P68. *FEMS Microbiol. Lett.* **219**, 275–283 (2003).
- M. C. Morais, S. Kanamaru, M. O. Badasso, J. S. Koti, B. A. L. Owen, C. T. McMurray, D. L. Anderson, M. G. Rossmann, Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Biol.* **10**, 572–576 (2003).
- A. D. Dearborn, E. A. Wall, J. L. Kizziah, L. Klenow, L. K. Parker, K. A. Manning, M. S. Spilman, J. M. Spear, G. E. Christie, T. Dokland, Competing scaffolding proteins determine capsid size during mobilization of *Staphylococcus aureus* pathogenicity islands. *eLife* **6**, e30822 (2017).
- W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, J. E. Johnson, Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
- L. Sun, X. Zhang, S. Gao, P. A. Rao, V. Padilla-Sanchez, Z. Chen, S. Sun, Y. Xiang, S. Subramaniam, V. B. Rao, M. G. Rossmann, Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat. Commun.* **6**, 7548 (2015).
- A. A. Lebedev, M. H. Krause, A. L. Isidro, A. A. Vagin, E. V. Orlova, J. Turner, E. J. Dodson, P. Tavares, A. A. Antson, Structural framework for DNA translocation via the viral portal protein. *EMBO J.* **26**, 1984–1994 (2007).
- A. A. Simpson, Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, M. G. Rossmann, Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408**, 745–750 (2000).
- J. Nováček, M. Šiborová, M. Benešik, R. Pantůček, J. Doškař, P. Plevka, Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9351–9356 (2016).
- R. K. Lokareddy, R. S. Sankhala, A. Roy, P. V. Afonine, T. Motwani, C. M. Teschke, K. N. Parent, G. Cingolani, Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat. Commun.* **8**, 14310 (2017).
- D. M. Tremblay, M. Tegoni, S. Spinelli, V. Campanacci, S. Blangy, C. Huyghe, A. Desmyter, S. Labrie, S. Moineau, C. Cambillau, Receptor-binding protein of *Lactococcus lactis* phages: Identification and characterization of the saccharide receptor-binding site. *J. Bacteriol.* **188**, 2400–2410 (2006).
- S. Ricagno, V. Campanacci, S. Blangy, S. Spinelli, D. Tremblay, S. Moineau, M. Tegoni, C. Cambillau, Crystal structure of the receptor-binding protein head domain from *Lactococcus lactis* phage bIL170. *J. Virol.* **80**, 9331–9335 (2006).
- A. Desmyter, C. Farenc, J. Mahony, S. Spinelli, C. Bebeacqua, S. Blangy, D. Veeler, D. van Sinderen, C. Cambillau, Viral infection modulation and neutralization by camelid nanobodies. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1371–E1379 (2013).
- M. Takáč, U. Bläsi, Phage P68 virion-associated protein 17 displays activity against clinical isolates of *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **49**, 2934–2940 (2005).
- D. Nelson, R. Schuch, P. Chahales, S. Zhu, V. A. Fischetti, PlyC: A multimeric bacteriophage lysin. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10765–10770 (2006).
- S. McGowan, A. M. Buckle, M. S. Mitchell, J. T. Hoopes, D. T. Gallagher, R. D. Heselpoth, Y. Shen, C. F. Reboul, R. H. P. Law, V. A. Fischetti, J. C. Whisstock, D. C. Nelson, X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12752–12757 (2012).
- C. Koç, G. Xia, P. Kühner, S. Spinelli, A. Roussel, C. Cambillau, T. Stehle, Structure of the host-recognition device of *Staphylococcus aureus* phage phi11. *Sci. Rep.* **6**, 27581 (2016).
- L. Xu, S. D. Benson, S. J. Butcher, D. H. Bamford, R. M. Burnett, The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. *Structure* **11**, 309–322 (2003).
- L. V. Sycheva, M. M. Shneider, N. N. Sykilinda, M. A. Ivanova, K. A. Miroshnikov, P. G. Leiman, Crystal structure and location of gp131 in the bacteriophage phiKZ virion. *Virology* **434**, 257–264 (2012).
- J. Xu, M. Gui, D. Wang, Y. Xiang, The bacteriophage phi29 tail possesses a pore-forming loop for cell membrane penetration. *Nature* **534**, 544–547 (2016).
- S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- F. Guo, W. Jiang, Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods Mol. Biol.* **1117**, 401–443 (2014).
- J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. M. Carazo, C. O. Sorzano, Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).

48. J. Chang, P. Weigele, J. King, W. Chiu, W. Jiang, Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* **14**, 1073–1082 (2006).
49. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
50. G. D. Pintilie, J. Zhang, T. D. Goddard, W. Chiu, D. C. Gossard, Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
51. S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart, J. T. Huiskonen, Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
52. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
53. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
54. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
55. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
56. G. J. Kleywegt, R. J. Read, Not your average density. *Structure* **5**, 1557–1569 (1997).
57. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
58. G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).

**Acknowledgments:** We acknowledge the CEITEC Core Facilities Cryo-Electron Microscopy and Tomography, Proteomics, and Biomolecular Interactions, supported by the CIISB project LM2015043 funded by the MEYS of the Czech Republic. X-ray data were collected at synchrotron SOLEIL beamline PROXIMA-1. **Funding:** This research was carried out under the project CEITEC 2020 (LQ1601), with financial support from the MEYS of the Czech Republic under National Sustainability Program II. This work was supported by IT4I project (CZ.1.05/1.1.00/02.0070) and funded by the European Regional Development Fund and the

national budget of the Czech Republic via the RDI-OP, as well as the MEYS via the Grant (LM2011033). The research leading to these results has received funding from the Ministry of Health of the Czech Republic grant 16-29916A to R.P., Czech Science Foundation grants 15-21631Y and 18-17810S, and EMBO installation grant 3041 to P.P. **Author contributions:** D.H. and P.P. designed research. D.H., D.S., K.S., and T.F. performed research. D.H., K.S., T.F., R.P., and P.P. analyzed data. D.H., T.F., and P.P. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Cryo-EM electron density maps of the P68 reconstructions were deposited with the Electron Microscopy Data Bank (EMDB): EMD-4438 icosahedral reconstruction of a capsid of the native phage P68; EMD-4436 5-fold symmetrized reconstruction of the native phage P68; EMD-4437 asymmetric reconstruction of the native phage P68; EMD-4440 reconstruction of head fiber of the native phage P68; EMD-4435 12-fold symmetrized reconstruction of a portal and tail of the native phage P68; EMD-4449 6-fold symmetrized reconstruction of a tail knob of the native phage P68; EMD-4450 5-fold symmetrized reconstruction of a tail spike of the native phage P68; EMD-4451 icosahedral reconstruction of a capsid of a genome release intermediate of P68; EMD-4453 5-fold symmetrized reconstruction of a capsid of the genome release intermediate of P68; EMD-4454 12-fold symmetrized reconstruction of a portal and tail of a genome release intermediate of P68; EMD-4455 12-fold symmetrized reconstruction of a portal and tail with knob of the genome release intermediate of P68; EMD-4442 icosahedral reconstruction of a capsid of the empty particle of P68; EMD-4456 5-fold symmetrized reconstruction of the empty particle of P68; EMD-4457 12-fold symmetrized reconstruction of a portal and tail of the empty particle of P68; EMD-4458 12-fold symmetrized reconstruction of a portal and tail with knob of the empty particle of P68; and EMD-4459 map of the whole native bacteriophage P68 combined from localized reconstructions. The corresponding coordinates were deposited in the PDB under the following accession numbers: 6IAT, icosahedral capsid of a native phage; 6IAW, head fiber; 6IAC, portal and tail of native P68; 6IB1, icosahedral capsid of empty particle; and 6Q3G, structure of the whole bacteriophage P68. The atomic coordinates of the tail fiber protein together with the structure factors were deposited in the PDB under the code 6IAB. Additional data related to this paper may be requested from the authors.

Submitted 21 February 2019  
Accepted 23 September 2019  
Published 16 October 2019  
10.1126/sciadv.aaw7414

**Citation:** D. Hřebík, D. Štveráková, K. Škubník, T. Füzik, R. Pantůček, P. Plevka, Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68. *Sci. Adv.* **5**, eaaw7414 (2019).



**Article 2: ICAM-1 induced rearrangements of capsid and genome prime rhinovirus 14 for activation and uncoating**

The study was initiated by observation of rhinovirus 14 forming a stable complex with its receptor intercellular adhesion molecule 1 (ICAM-1). The observation was contradictory to previous reports, which showed that ICAM-1 causes uncoating of rhinovirus 14 and rhinovirus 3 at neutral pH [120, 121, 122]. I designed a series of experiments that tested what conditions cause rhinovirus 14 to release its genome. Other enteroviruses (e.g., EV18) are sensitive to the acidic pH of the endosome, where pH 6.0 causes uncoating of the virus. Thus, I exposed rhinovirus 14 to pH 6.2. When I formed the rhinovirus 14-ICAM-1 complex and subsequently changed the pH to 6.2, most of the particles (>90%) released their genome or were in the activated state.

The next question was the determination of the mechanism - why is the receptor necessary for the uncoating? Initially, I looked into the structure of the native virion. I noticed an unidentified cryo-EM density around a 2-fold symmetry axis. I identified the cryo-EM density as part of the genome and built an RNA octanucleotide into the density where one base interacts with conserved residue Trp38 of VP2. Then, I compared the structures of the native virus and the rhinovirus 14-ICAM-1 complex. The structures are very similar except for the C-terminal (Ct) part of VP4. In the native virus, the Ct part of VP4 is located in the vicinity of the 3-fold symmetry axis of the viral capsid, while in the rhinovirus 14-ICAM-1 complex, it moves near the 2-fold symmetry axis. Moreover, after receptor binding, the VP4 replaces the RNA base in the interaction with Trp38, which is in accordance with changes of Trp fluorescence measured by nanoDSF during *in vitro* uncoating of rhinovirus 2 [123]. That also causes conformational changes to the genome. A strong genome density appears around the 3-fold symmetry axis, where the VP4 is in the native virus. All these structural changes caused by the receptor binding prime the virus for uncoating. The VP4 is ideally positioned to escape from the virus through pores that appear after acidification of the outer environment, which causes activation of the virus (release of VP4 and expansion of the viral particle), and subsequent genome release. All in all, the receptor binding causes changes to the structure of the capsid and genome, which are necessary for the subsequent steps which result in genome release into the host cytoplasm.

The structure of the rhinovirus 14-ICAM-1 complex was solved by cryo-EM in 2004 to a resolution of approximately 10 Å. At such resolution, it is not possible to identify movements of parts of proteins or RNA. Thus, at that time, it was not possible to elucidate the mechanism of uncoating of the virus. Moreover, our results are contradictory to the previous ones, which reported that binding of ICAM-1 to rhinovirus-14 causes conversion of the rhinovirus-14 to activated state and subsequent uncoating. Other experiments have shown that the ability of ICAM-1 to trigger the genome release depends on temperature, receptor concentration, and buffer composition [120, 124]. The experiments showing that ICAM-1 can induce genome release of rhinovirus-14 were performed in non-physiological concentrations of salts [120, 121, 122], which could destabilize the virus particles. It is essential for viruses to precisely control their genome release because uncoating just after binding to ICAM-1 would lead to premature genome release at the cell's surface.





# ICAM-1 induced rearrangements of capsid and genome prime rhinovirus 14 for activation and uncoating

Dominik Hřebík<sup>a</sup>, Tibor Füzik<sup>a</sup>, Mária Gondová<sup>a</sup>, Lenka Šmerdová<sup>a</sup>, Athanassios Adamopoulos<sup>a</sup>, Ondřej Šedo<sup>a</sup>, Zbyněk Zdráhal<sup>a</sup>, and Pavel Plevka<sup>a,1</sup>

<sup>a</sup>Central European Institute of Technology, Masaryk University, Brno 62500, Czech Republic

Edited by Stephen C. Harrison, Boston Children's Hospital, Boston, MA, and approved April 4, 2021 (received for review November 23, 2020)

**Most rhinoviruses, which are the leading cause of the common cold, utilize intercellular adhesion molecule-1 (ICAM-1) as a receptor to infect cells. To release their genomes, rhinoviruses convert to activated particles that contain pores in the capsid, lack minor capsid protein VP4, and have an altered genome organization. The binding of rhinoviruses to ICAM-1 promotes virus activation; however, the molecular details of the process remain unknown. Here, we present the structures of virion of rhinovirus 14 and its complex with ICAM-1 determined to resolutions of 2.6 and 2.4 Å, respectively. The cryo-electron microscopy reconstruction of rhinovirus 14 virions contains the resolved density of octanucleotide segments from the RNA genome that interact with VP2 subunits. We show that the binding of ICAM-1 to rhinovirus 14 is required to prime the virus for activation and genome release at acidic pH. Formation of the rhinovirus 14–ICAM-1 complex induces conformational changes to the rhinovirus 14 capsid, including translocation of the C termini of VP4 subunits, which become poised for release through pores that open in the capsids of activated particles. VP4 subunits with altered conformation block the RNA–VP2 interactions and expose patches of positively charged residues. The conformational changes to the capsid induce the redistribution of the virus genome by altering the capsid–RNA interactions. The restructuring of the rhinovirus 14 capsid and genome prepares the virions for conversion to activated particles. The high-resolution structure of rhinovirus 14 in complex with ICAM-1 explains how the binding of uncoating receptors enables enterovirus genome release.**

virus | structure | receptor | cryo-electron microscopy | genome release

**H**uman rhinoviruses are the cause of more than half of common colds (1). Medical visits and missed days of school and work cost tens of billions of US dollars annually (2, 3). There is currently no cure for rhinovirus infections, and the available treatments are only symptomatic. Rhinoviruses belong to the family *Picornaviridae*, genus *Enterovirus*, and are classified into species A, B, and C (4). Rhinoviruses A and B can belong to either “major” or “minor” groups, based on their utilization of intercellular adhesion molecule-1 (ICAM-1) or low-density lipoprotein receptor for cell entry (5–7). Type C rhinoviruses use CDHR3 as a receptor (8). Rhinovirus 14 belongs to the species rhinovirus B and uses ICAM-1 as a receptor. Receptors recognized by rhinoviruses and other enteroviruses can be divided into two groups based on their function in the infection process (9). Attachment receptors such as DAF, PSGL1, KREMEN1, CDHR3, and sialic acid enable the binding and endocytosis of virus particles into cells (10–13). In contrast, uncoating receptors including ICAM-1, CD155, CAR, and SCARB2 enable virus cell entry but also promote genome release from virus particles (5, 14–16).

Virions of rhinoviruses are nonenveloped and have icosahedral capsids (17). Genomes of rhinoviruses are 7,000 to 9,000 nucleotide-long single-stranded positive-sense RNA molecules (1, 17). The rhinovirus genome encodes a single polyprotein that is co- and posttranslationally cleaved into functional protein subunits. Capsid proteins VP1, VP3, and VP0, originating from one polyprotein, form a protomer, 60 of which assemble into a pseudo-T = 3 icosahedral capsid. To render the virions mature and infectious, VP0

subunits are cleaved into VP2 and VP4 (18, 19). VP1 subunits form pentamers around fivefold symmetry axes, whereas subunits VP2 and VP3 form heterohexamers centered on threefold symmetry axes. The major capsid proteins VP1 through 3 have a jelly roll  $\beta$ -sandwich fold formed by two  $\beta$ -sheets, each containing four antiparallel  $\beta$ -strands, which are conventionally named B to I (20–22). The two  $\beta$ -sheets contain the strands BIDG and CHEF, respectively. The C termini of the capsid proteins are located at the virion surface, whereas the N termini mediate interactions between the capsid proteins and the RNA genome on the inner surface of the capsid. VP4 subunits are attached to the inner face of the capsid formed by the major capsid proteins. The surfaces of rhinovirus virions are characterized by circular depressions called canyons, which are centered around fivefold symmetry axes of the capsids (21).

The VP1 subunits of most rhinoviruses, but not those of rhinovirus 14, contain hydrophobic pockets, which are filled by molecules called pocket factors (17, 21, 23, 24). It has been speculated that pocket factors are fatty acids or lipids (25). The pockets are positioned immediately below the canyons. The exposure of rhinoviruses to acidic pH induces expulsion of the pocket factors, which leads to the formation of activated particles and genome release (17, 26–32). The activated particles are characterized by capsid expansion, a reduction in interpentamer contacts, the release of VP4 subunits, externalization of N termini of VP1 subunits, and changes in the distribution of RNA genomes (17, 26–29, 33, 34). Artificial hydrophobic compounds that bind to VP1 pockets with high affinity inhibit infection by rhinoviruses (35, 36).

ICAM-1 is an endothelial- and leukocyte-associated protein that stabilizes cell–cell interactions and facilitates the movement

## Significance

**Medical visits and missed days of school and work caused by rhinoviruses cost tens of billions of US dollars annually. Currently, there are no antivirals against rhinoviruses, and the available treatments only treat the symptoms. Here, we present the molecular structure of human rhinovirus 14 in complex with its cellular receptor intercellular adhesion molecule 1. The binding of the virus to its receptor initiates the infection. Knowledge of the structure of the human rhinovirus 14–intercellular adhesion molecule 1 interface and mechanism of interaction provides the basis for the design of compounds that may block the binding of rhinoviruses to receptors and thus prevent infection.**

Author contributions: D.H. and P.P. designed research; D.H., L.Š., A.A., and O.Š. performed research; D.H., T.F., M.G., O.Š., Z.Z., and P.P. analyzed data; and D.H., T.F., and P.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: pavel.plevka@ceitec.muni.cz.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2024251118/-/DCSupplemental>.

Published May 5, 2021.

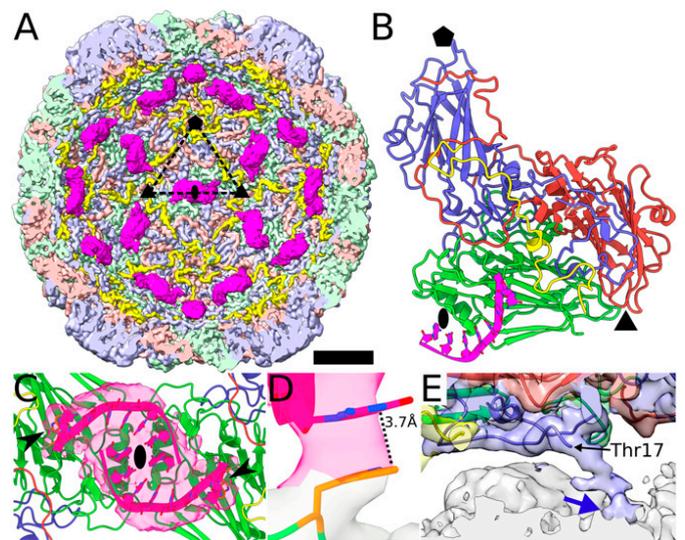
of leukocytes through endothelia (37). ICAM-1 can be divided into an extracellular amino-terminal part composed of five immunoglobulin domains, a single transmembrane helix, and a 29-residue-long carboxyl-terminal cytoplasmic domain. The immunoglobulin domains are characterized by a specific fold that consists of seven to eight  $\beta$ -strands, which form two antiparallel  $\beta$ -sheets in a sandwich arrangement (38–40). The immunoglobulin domains of ICAM-1 are stabilized by disulfide bonds and glycosylation (38–41). The connections between the immunoglobulin domains are formed by flexible linkers that enable bending of the extracellular part of ICAM-1. For example, the angle between domains 1 and 2 differs by  $8^\circ$  between molecules in distinct crystal forms (38, 42). As a virus receptor, ICAM-1 enables the virus particles to sequester at the cell surface and mediates their endocytosis (5). The structures of complexes of rhinoviruses 3, 14, and 16, and CVA21 with ICAM-1 have been determined to resolutions of 9 to 28 Å (42–46). It was shown that ICAM-1 molecules bind into the canyons at the rhinovirus surface, approximately between fivefold and twofold symmetry axes (42–46). ICAM-1 interacts with residues from all three major capsid proteins. It has been speculated that the binding of ICAM-1 triggers the transition of virions of rhinovirus 14 to activated particles and initiates genome release (45, 47). However, the limited resolution of the previous studies prevented characterization of the corresponding molecular mechanism.

Here, we present the cryo-electron microscopy (cryo-EM) reconstruction of the rhinovirus 14 virion, which contains resolved density of octanucleotide segments of the RNA genome that interact with VP2 subunits. Furthermore, we show that the binding of ICAM-1 to rhinovirus 14 induces changes in its capsid and genome, which are required for subsequent virus activation and genome release at acidic pH.

## Results and Discussion

**Interactions of Capsid of Rhinovirus 14 with Genome.** The cryo-EM reconstruction of rhinovirus 14 virion was determined to a resolution of 2.6 Å (Fig. 1A and B and *SI Appendix, Fig. S1 and Table S1*). The map enabled building of the structure of capsid proteins VP1 through 4 except for residues 1 to 17 and 290 to 293 of VP1, 1 to 6 of VP2, and 1 to 28 of VP4 (Fig. 1B). In addition to the capsid, the cryo-EM map contains resolved density corresponding to octanucleotide segments from the RNA genome (Fig. 1A–C). The quality of the map enabled building of the RNA structure; however, the nucleotide sequence could not be identified. The base of the second nucleotide from the 5' end of the RNA segment is flipped out from the RNA helix and forms a stacking interaction with the sidechain of Trp38 of VP2 (Fig. 1D and *SI Appendix, Fig. S2*). The residue Trp38 is conserved in the VP2 subunits of numerous picornaviruses, including polioviruses, rhinoviruses 2 and 16, coxsackievirus B3, coxsackievirus A9, and coxsackievirus A21 (*SI Appendix, Fig. S3*). Virion structures of these viruses contain disk-like densities that stack onto the tryptophane side chains, which were hypothesized to belong to a base of guanine nucleotide (23, 24, 48–51). The structure of the RNA segment in the virion of rhinovirus 14 provides evidence that the previous speculations about the densities stacking onto Trp38 side chains were correct. To be consistent with the previous structures, we modeled the second nucleotide of the RNA segment in rhinovirus 14 as guanine (Fig. 1D and *SI Appendix, Fig. S2*). The stacking interaction between Trp38 and the base of the second nucleotide is the only direct contact between the RNA segment and the capsid (Fig. 1A, C, and D and *SI Appendix, Fig. S2*).

Each RNA segment in the virion of rhinovirus 14 is associated with one protomer of capsid proteins VP1 through 4 (Fig. 1A and B). The RNA is positioned next to a twofold axis, and two of the oligonucleotides interact with each other (Fig. 1A–C). To enable Watson–Crick base pairing between the two segments, the nucleotide at position seven was modeled as uracil and the



**Fig. 1.** Structure of virion of rhinovirus 14 contains resolved density corresponding to octanucleotides from its RNA genome. (A) Surface representation of cryo-EM of reconstruction of virion of rhinovirus 14 with front half of the particle removed to show internal structure. Density corresponding to VP1 is shown in blue, VP2 in green, VP3 in red, VP4 in yellow, and RNA segments in pink. Borders of a selected icosahedral asymmetric unit are indicated with a dashed triangle and positions of selected twofold, threefold, and fivefold symmetry axes are represented by an oval, triangle, and pentagon, respectively. (Scale bar, 5 nm.) (B) Cartoon representation of icosahedral asymmetric unit of rhinovirus 14 viewed from the inside of the capsid. The color coding of individual virus components is the same as in A. Positions of twofold, threefold, and fivefold symmetry axes are represented by an oval, triangle, and pentagon, respectively. (C) Two RNA octanucleotides that interact with each other and with VP2 subunits. Protein and RNA coloring is the same as in A. The cryo-EM density of the RNA segments is shown as a pink semitransparent surface. RNA bases and side chains of Trp38 of VP2 are shown in stick representation, in orange, and indicated with black arrowheads. The position of a twofold symmetry axis is indicated with an oval. (D) Detail of stacking interaction between GUA2 from RNA segment and Trp38 of VP2. The cryo-EM densities of RNA and protein are shown as semitransparent surfaces in pink and gray, respectively. (E) Interaction between N terminus of VP1 and genome. Capsid proteins are shown in cartoon representation with the same coloring as in A. Cryo-EM densities of individual proteins are shown as semitransparent surfaces colored according to the chain they belong to. The density of the RNA genome is shown in gray. The blue arrow indicates the contact between the N terminus of VP1 and the genome. The position of Thr17, the first modeled residue from the N terminus of VP1, is indicated.

nucleotide at position eight as adenine (Fig. 1C). Due to the constraints imposed by the interaction of the RNA with the capsid and of the interactions of two segments of the RNA with each other, the sequence of the oligonucleotide was built as 5' UGUUUUUA 3'. Nevertheless, other sequences that fulfill the interaction conditions are equally possible.

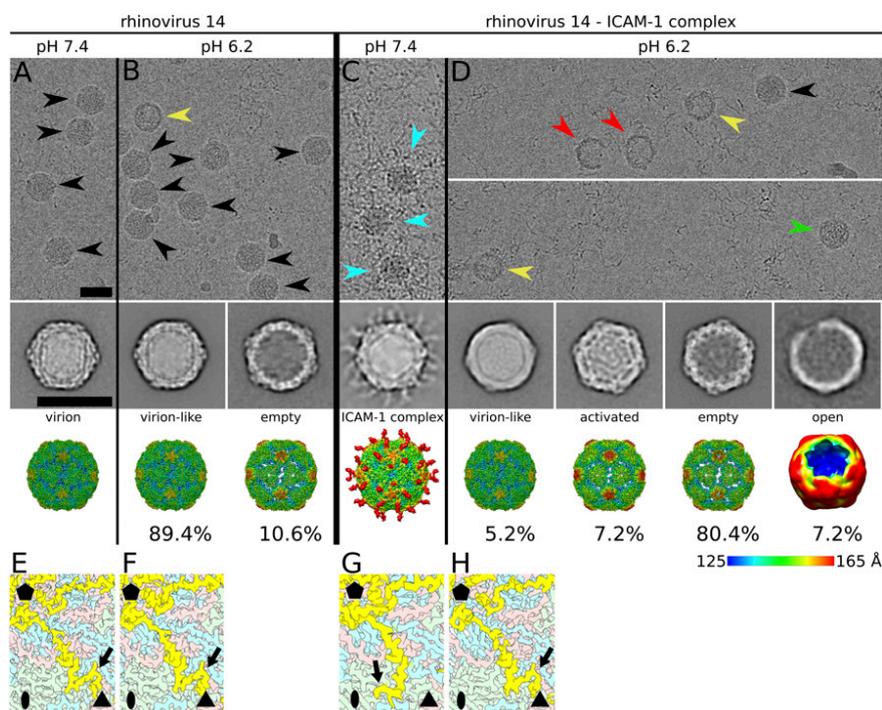
The reconstruction of virion of rhinovirus 14 provides evidence that the N termini of VP1 subunits interact with the RNA genome (Fig. 1E). A similar function of the N terminus of VP1 was described previously in rhinovirus 2 (26). None of the interactions between the genomes and capsids of enteroviruses identified to date are sequence specific. Nevertheless, even the nonspecific interactions of the N termini of VP1 and Trp38 of VP2 with the RNA may enable the packaging of the enterovirus genome into a particle. Compounds that could prevent the RNA–capsid interaction by, for instance covering the side chain of Trp38, may interfere with the production of infectious virions.

**Binding of Rhinovirus 14 to ICAM-1 Does Not Induce Genome Release.** The complex of rhinovirus 14 with ICAM-1 was prepared by mixing the components in phosphate-buffered saline (PBS) of

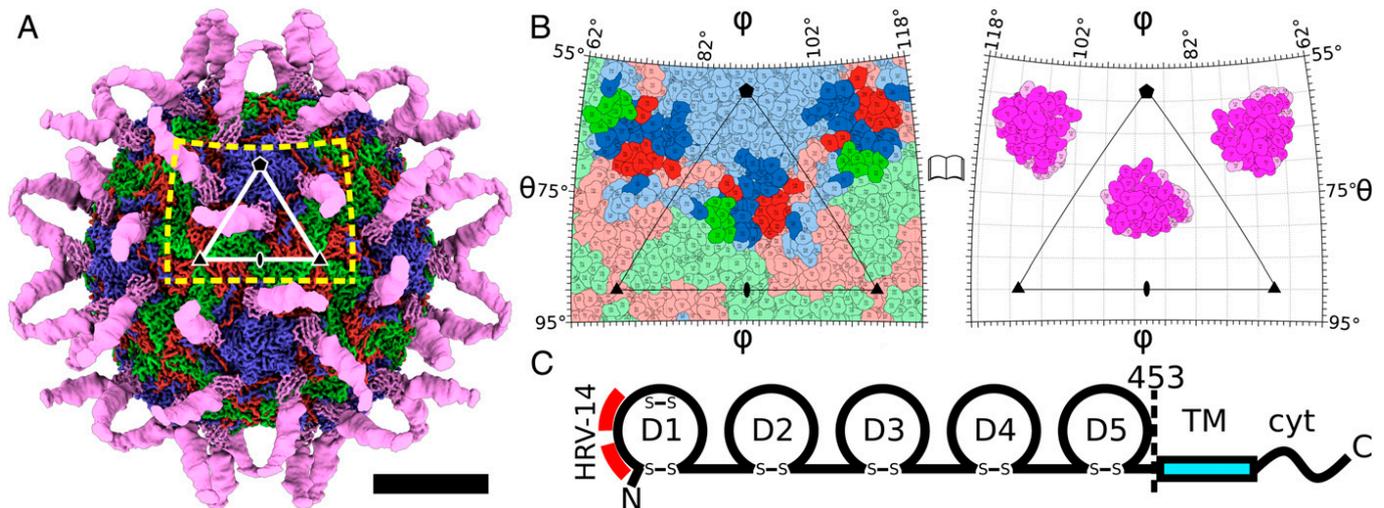
pH 7.4 and incubating them at 34 °C for 30 min (Fig. 2). The temperature was chosen to mimic that in the human upper respiratory tract (52). The binding of ICAM-1 did not induce the formation of activated particles or the genome release of rhinovirus 14 (Fig. 2 A–C). This is in agreement with previous results showing that the ability of ICAM-1 to trigger genome release depends on temperature, receptor concentration, and buffer composition (47, 53). Experiments showing that ICAM-1 could induce genome release were performed in solutions with nonphysiological concentrations of salts, which may have destabilized the virus particles (42, 46, 47). Since enteroviruses have to deliver their genomes into the cytoplasm of a host cell to initiate infection, it would be detrimental if they released their genomes immediately upon binding to receptors at the cell surface. Our results show that rhinovirus 14 is stable when bound to ICAM-1 under native-like conditions (Fig. 2C), and the induction of genome release requires exposure to acidic pH in endosomes.

**Structure of Rhinovirus 14–ICAM-1 Complex.** The structure of rhinovirus 14 in complex with the soluble ectodomain of ICAM-1 was determined to a resolution of 2.4 Å using cryo-EM and single-particle reconstruction (Fig. 3 and *SI Appendix*, Fig. S1 and Table S1). Domain 1 of ICAM-1 bound at the surface of rhinovirus 14 was resolved to a resolution of 2.6 Å (Fig. 3A and *SI Appendix*, Fig. S1). Levels of density in the map region corresponding to the domain 1 of ICAM-1 are similar to those in the capsid of the virus, indicating full occupancy of the receptors at the virus surface. Domains 2 and 3 of ICAM-1 are resolved to a resolution of 6 Å, and domains 4 and 5 are not visible in the cryo-EM reconstruction

(Fig. 3A). The low resolution of the region of the map corresponding to domains 2 and 3 and the lack of density for domains 4 and 5 are probably caused by movements of those domains relative to domain 1, which is anchored at the virus surface (Fig. 3A) (38, 42, 43). In agreement with previous studies, domain 1 of ICAM-1 binds into the canyon of rhinovirus 14, approximately in the middle between fivefold and twofold symmetry axes (Fig. 3B and C) (42, 43, 45). Previous studies of the interactions of rhinoviruses 3, 14, and 16 with ICAM-1 were limited to a resolution of 9.5 Å or lower (42–46). The interpretation of the macromolecular interactions relied on the fitting of high-resolution structures, determined by X-ray crystallography, to the cryo-EM maps of the complex (42–46). Therefore, the changes in the structures of the receptor and virus, induced upon their binding, could not be identified. Here, we show that ICAM-1 is wedged 3.4 Å deeper into the canyon and rotated 7.6° clockwise, when looking along the long axis of domain 1 toward the virus surface, relative to the structure reported previously (Fig. 4 and *SI Appendix*, Fig. S4A) (42, 43, 45). The interaction interface between ICAM-1 and rhinovirus 14 has a buried surface area of 1,500 Å<sup>2</sup>. The core of domain 1 of ICAM-1 is formed by β-sheets ABED and GFC (Fig. 4 A–C) (11, 54). Residues from the loops BC, DE, and FG and strands B, C, D, E, F, and G of ICAM-1 interact with rhinovirus 14 (*SI Appendix*, Fig. S4B). The mode of attachment of rhinovirus 14 to ICAM-1 is characteristic for uncoating receptors that bind to enterovirus canyons (5, 9, 14, 15). The uncoating receptors of enteroviruses, including ICAM-1, CD155, and CAR, have elongated molecules formed by domains with an immunoglobulin



**Fig. 2.** Binding of ICAM-1 to rhinovirus 14 is required for efficient genome release at acidic pH. (A–D) Electron micrographs (Top), reference-free two-dimensional class averages (Center), and 3D reconstructions (Bottom) of rhinovirus 14 and rhinovirus 14–ICAM-1 complex at neutral and acidic pH. The cryo-EM reconstructions are rainbow colored based on the distance of the particle surface from its center. Names above the reconstructions indicate the types of particles. Percentages below the reconstructions indicate the relative abundance of each type of particle in the sample. (A and B) Rhinovirus 14 at pH 7.4 (A) and 6.2 (B). (C and D) Rhinovirus 14–ICAM-1 complex at pH 7.4 (C) and 6.2 (D). The top row of D contains two parts of a micrograph to show all types of particles present in the sample. Black arrowheads indicate virions, yellow indicate empty particles, cyan indicate rhinovirus 14–ICAM-1 complex, green indicate activated particles, and red indicate open particles. (Scale bars, 30 nm.) (E–H) Conformational changes to VP4 are induced by ICAM-1 binding but not by acidic pH. Surface representations of cryo-EM reconstructions showing the inner faces of capsids. The surfaces are color coded according to the capsid proteins with VP1 in blue, VP2 in green, VP3 in red, and VP4 in yellow. Only particles containing VP4 are shown. Positions of icosahedral symmetry axes are indicated with a black pentagon, triangle, and oval for fivefold, threefold, and twofold, respectively. Black arrows highlight C termini of VP4.



**Fig. 3.** Structure of rhinovirus 14 in complex with ICAM-1. (A) Surface representation of cryo-EM reconstruction of rhinovirus 14-ICAM-1 complex color coded to distinguish individual proteins. Density corresponding to VP1 is shown in blue, VP2 in green, VP3 in red, and ICAM-1 in light magenta. Positions of selected icosahedral symmetry axes are indicated by a pentagon for fivefold, triangle for threefold, and an oval for twofold. The white triangle indicates the border of a selected icosahedral asymmetric unit. The yellow dashed rectangle indicates borders of the area shown in detail in B. (B) Roadmap projection showing residues forming the outer surface of rhinovirus 14 capsid (*Left*) and residues of domain 1 of ICAM-1 facing toward the virus (*Right*). Coloring is the same as in A. Residues involved in virus-receptor interactions are shown in bright colors. The polar angles  $\theta$  and  $\phi$  indicate positions at the capsid surface. (C) Schematic representation of ICAM-1. D1 to D5 indicate extracellular immunoglobulin domains; TM, transmembrane domain; cyt, cytoplasmic domain. Disulfide bridges (S-S) stabilizing the immunoglobulin domains are indicated. Red dashes highlight the binding site for rhinovirus 14. The ectodomain used in this study to determine the rhinovirus 14-ICAM-1 interactions included residues 1 to 453.

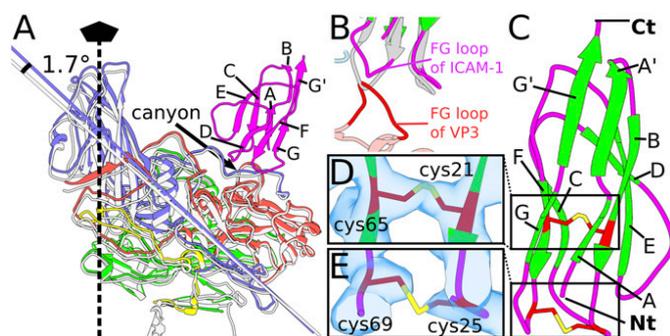
fold, which enables their insertion into the canyons of enterovirus particles (5, 14, 15).

**Conformational Changes Required for Binding of Rhinovirus 14 to ICAM-1.** The binding of rhinovirus 14 to ICAM-1 is accompanied by the local restructuring of domain 1 of ICAM-1 and surface loops of capsid proteins, as well as by overall changes in the structure of the rhinovirus-14 capsid (Figs. 4 and 5 and *SI Appendix*, Figs. S5 and S6). VP1 subunits rotate  $1.7^\circ$  toward VP2 and VP3, which results in

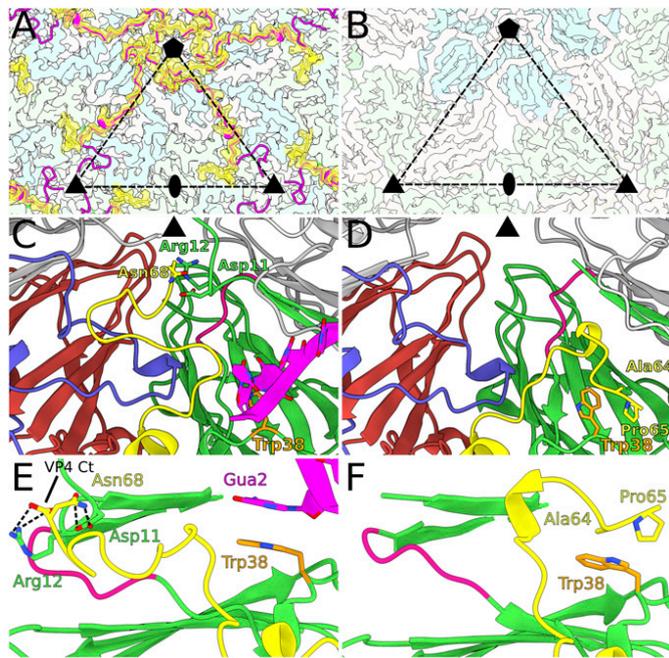
a contraction of the canyon (Fig. 4A). As a result, the capsid expands by  $5 \text{ \AA}$  in diameter (Fig. 4A). The binding of ICAM-1 to rhinovirus 14 requires the bending of the FG loop of ICAM-1  $8 \text{ \AA}$  toward the core of the immunoglobulin domain (Fig. 4B and *SI Appendix*, Fig. S5). This conformational change is necessary to prevent the clashing of the FG loop of ICAM-1 with residues 178 to 182 from the FG loop of VP3 of rhinovirus 14 (Fig. 4B). The conformational flexibility of the FG loop of ICAM-1 enables enlargement of its interaction interface with the capsid.

The structures of domain 1 of ICAM-1 determined to date contain the disulfide bonds Cys21 and Cys65 and Cys25 and Cys69 (38–40). Residues Cys25 and Cys69 are located in the vicinity of the virus surface when ICAM-1 binds to rhinovirus 14 (Fig. 4C–E). Cys69 is part of the FG loop, whereas Cys25 is part of the BC loop (Fig. 4C and E). The density connecting Cys25 and Cys69 of ICAM-1 in the complex with rhinovirus 14 is much weaker than that connecting Cys21 to Cys65 (Fig. 4C and E). However, the positions of the two cysteines in the cryo-EM density map are consistent with the linkage of their side chains by a disulfide bond (Fig. 4C and E) (55). Furthermore, mass spectrometry analysis of ICAM-1 molecules from the complex with rhinovirus 14 did not detect any peptides containing free Cys25 and Cys69 (*SI Appendix*, Fig. S7). However, peptides containing free cysteines were observed after the reduction of the disulfide bonds by dithiothreitol (DTT). This provides evidence that Cys25 and Cys69 of ICAM-1 in complex with rhinovirus 14 are linked by a disulfide bond. The low values of cryo-EM density may be caused by a higher flexibility of this part of ICAM-1, as indicated by the lower resolution than in the core of domain 1 (Fig. 3A). The binding of rhinovirus 14 to ICAM-1 also induces structural changes in the virus proteins. Residues 154 to 162 from the DE loop of VP1 shift  $2 \text{ \AA}$  toward the core of the subunit (*SI Appendix*, Fig. S6). This movement helps to accommodate ICAM-1 in the canyon of rhinovirus 14.

**Molecular Details of the Interface between Rhinovirus 14 and ICAM-1.** The interaction between ICAM-1 and rhinovirus 14 is formed by 36 residues from domain 1 of ICAM-1 and 31, 8, and 13 residues



**Fig. 4.** Conformational changes associated with binding of rhinovirus 14 to ICAM-1. (A) Cartoon representation of icosahedral asymmetric unit of rhinovirus 14 in complex with ICAM-1. The VP1 subunit is shown in blue, VP2 in green, VP3 in red, VP4 in yellow, and domain 1 of ICAM-1 in magenta. The capsid proteins from virion of rhinovirus 14 are superimposed onto those of the rhinovirus 14-ICAM-1 complex and are shown in white. The binding of ICAM-1 to rhinovirus 14 induces a  $1.7^\circ$  tilt of VP1 toward VP2 and VP3, which results in a narrowing of the canyon relative to the virion structure. Least-squares planes fitted to VP1 are shown to highlight the rotation of VP1. (B) Conformational change of FG loop of ICAM-1, shown in magenta, is required to prevent clashes with FG loop of VP3, shown in red. The native structure of ICAM-1 clashing with VP2 is shown in gray. (C) Cartoon representation of structure of ICAM-1 bound to rhinovirus 14 with side chains of cysteine residues shown in stick representation with red carbon atoms and yellow sulfur atoms. (D and E) Detail of disulfide bridge between Cys21 and Cys65 (D) and Cys25 and Cys69 (E). Cryo-EM density is shown as a semitransparent blue surface.



**Fig. 5.** Changes of structure of C terminus of VP4 induced by ICAM-1 binding to rhinovirus 14. (A) Surface representation of cryo-EM reconstruction of capsid of rhinovirus 14 in complex with ICAM-1 viewed from inside the virion. Density corresponding to VP1 is shown in pale blue, VP2 in pale green, VP3 in pale red, and VP4 in semitransparent yellow. The structure of VP4 in the rhinovirus 14–ICAM-1 complex is shown in cartoon representation in yellow, whereas the structure of VP4 in the virion of rhinovirus 14 is shown in magenta. The positions of selected icosahedral symmetry axes are indicated with a pentagon for fivefold, triangle for threefold, and oval for twofold. Borders of a selected icosahedral asymmetric unit are indicated with a dashed triangle. (B) Capsid structure of an empty particle of rhinovirus 14 containing pores around twofold symmetry axes and between twofold and fivefold symmetry axes through which VP4 may be released from the particle. (C–F) Differences in structure of VP4 subunits in virion (C and E) and rhinovirus 14–ICAM-1 complex (D and F). Capsid proteins are shown in cartoon representation. VP1 is shown in blue, VP2 in green, VP3 in red, VP4 in yellow, and RNA segments in pink. (C and E) Asn68 from C terminus of VP4 interacts with Asp11 and Arg12 of VP2 in virion of rhinovirus 14. The residues Asp11 and Arg12 are stabilized in position by the underlying loop of VP2 formed by residues 27 to 32 (highlighted in magenta). The side chain of Trp38 (highlighted in orange) forms a stacking interaction with Gua2 that is part of the resolved RNA segment positioned next to a twofold axis. (D and F) Binding of rhinovirus 14 to ICAM-1 induces conformational changes of virus capsid that include movement of residues 27 to 32 of VP2 toward particle center, which prevents interaction of C terminus of VP4 with residues Asp11 and Arg12 of VP2. The C terminus of VP4 acquires a new conformation, which covers the side chain of Trp38 of VP2 and blocks its interaction with RNA.

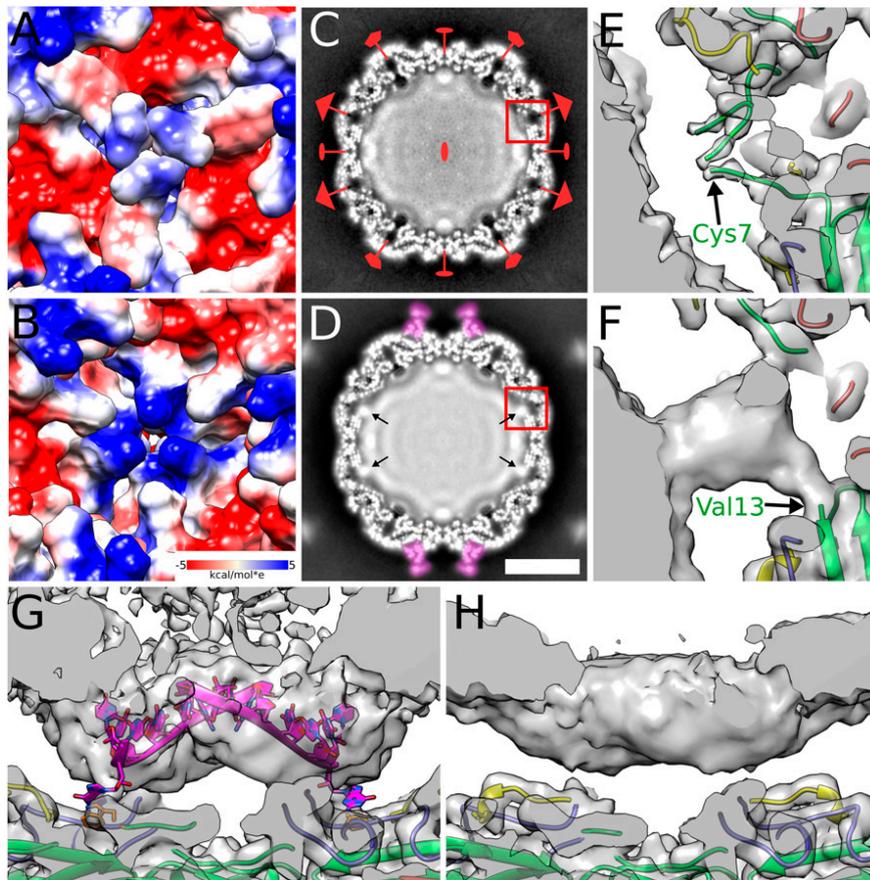
of VP1, VP2, and VP3 of rhinovirus 14, respectively (Fig. 3B). The specificity of the interaction is controlled by a combination of the complementarity of the electrostatic interactions, a network of hydrogen bonds, and the positions of patches of hydrophobic interfaces. There are salt bridges between Lys77 of ICAM-1 and Glu210 from VP1 and Lys39 of ICAM-1 and C-terminal carboxyl group of Glu236 from VP3 (*SI Appendix, Fig. S8 A and B*). The interface includes a network of 18 hydrogen bonds. Furthermore, the resolution of the cryo-EM reconstruction of the complex was sufficient to enable the identification of water molecules, some of which form hydrogen bonds with both ICAM-1 and rhinovirus 14 and thus mediate interactions between the receptor and virus (*SI Appendix, Fig. S8 C and D*). For example, the amino group from the side chain of Lys29 of ICAM-1 interacts with two water molecules that form hydrogen bonds with side chains of Thr105 of

VP3 and Asn68 of ICAM-1 (*SI Appendix, Fig. S8D*). It has been shown previously that the mutation of Thr75 of ICAM-1 to Ala reduces the efficiency of binding of rhinovirus 14 by more than 50% (54, 56). No ions were identified at the rhinovirus 14–ICAM-1 interface (7, 57).

The interface between rhinovirus 14 and ICAM-1 contains complementary patches of hydrophobic interactions (*SI Appendix, Fig. S8 E and F*). Previous studies have shown that most mutations of Pro70 from the FG loop of ICAM-1 prevent the binding of rhinovirus 14 (54, 56). We show that Pro70 fits into a small hydrophobic pocket formed by Pro178, Phe86, and Thr180 of VP3 (*SI Appendix, Fig. S8F*). Fitting Pro70 of ICAM-1 into the hydrophobic cavity in VP3 requires movement and restructuring of the FG loop of ICAM-1 (Fig. 4 C and E and *SI Appendix, Fig. S8F*). Another residue of ICAM-1 that is critical for the binding of rhinovirus 14 is Leu30 (54, 56). In the complex, the side chain of Leu30 is situated between the side chains of Ile215 and Val217 from VP1, which form a hydrophobic pocket for the leucine side chain (*SI Appendix, Fig. S8E*). This explains why the mutation of Leu30 to Ser eliminates the binding of ICAM-1 to rhinovirus 14 (54, 56).

**ICAM-1 Binding Prepares VP4 Subunits for Release from Activated Particles.** The structure of the C terminus of VP4 subunit in the rhinovirus 14–ICAM-1 complex differs from that in the native virion (Figs. 2 E–H and 5). The two structures of VP4 subunits are similar for residues 29 to 57, with rmsd of C $\alpha$  atoms of the corresponding residues of 0.44 Å. However, residues 58 to 65 of VP4 extend toward a threefold symmetry axis of the capsid in the native virion, whereas the same residues point toward a twofold symmetry axis in the rhinovirus 14–ICAM-1 complex (Fig. 5A). The movement of the C terminus of VP4 is induced by conformational changes of the major capsid proteins, which are triggered by ICAM-1 binding to the capsid. In the rhinovirus-14 virion, the C-terminal carboxyl group of Asn68 from VP4 forms a salt bridge with the side chain of Arg12 of VP2 (Fig. 5 C and E). Additionally, the side chain of Asn68 forms two hydrogen bonds with the side chain of Asp11 of VP2 (Fig. 5 C and E). As discussed above, the binding of ICAM-1 to rhinovirus 14 induces a rotation of VP1 toward VP2 and VP3 (Fig. 4A). These movements of capsid proteins bring residues 27 to 33 from the N terminus of VP2 into the space that is occupied by Arg12 of VP2 in the native virion (Fig. 5 D and F). This frees the C terminus of VP4 from the interaction with Arg12 of VP2 and probably enables its translocation toward a twofold axis (Fig. 5 A, D, and F). The restructuring of the C-terminal part of VP4 to point toward a twofold symmetry axis prepares the protein for release through either of the holes that form at and next to the twofold symmetry axes upon particle activation (Fig. 5 A and B) (29, 34).

**ICAM-1 Binding Induces Changes in Genome Organization of Rhinovirus 14.** The binding of ICAM-1 into the canyon of rhinovirus 14 induces the relocation of the C terminus of VP4 toward a twofold symmetry axis of the capsid (Fig. 5A). The movement of the C terminus of VP4 uncovers a patch of positively charged residues at the inner face of the capsid, adjacent to a threefold symmetry axis (Fig. 6 A and B). The positively charged surface attracts genomic RNA, which is represented in the cryo-EM reconstruction as a cylindrical appendage emanating from the spherical genome density filling the center of the virus particle (Fig. 6 C–F). This indicates that parts of the RNA genome in various conformations interact nonspecifically with the positively charged regions of the capsid. Furthermore, the N termini of VP2 subunits probably interact with the RNA density positioned on a threefold axis (Fig. 6F). The C terminus of VP4 positioned next to a twofold axis of the capsid covers the side chain of Trp38 of VP2, which in the native virion forms a stacking interaction with a nucleotide from the RNA genome (Figs. 1D and 5 E and F and *SI Appendix, Fig.*

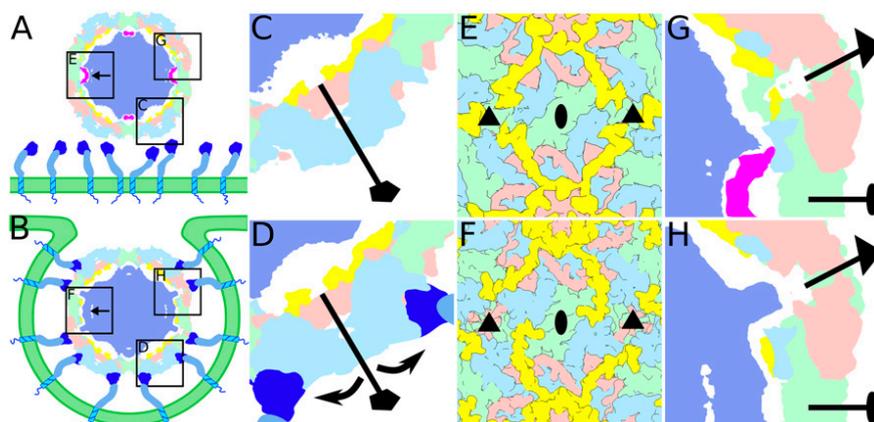


**Fig. 6.** Conformational changes of capsid of rhinovirus 14 that are induced by binding to ICAM-1 trigger redistribution of RNA genome in the particle. (A and B) Detail of inner capsid surface around threefold symmetry axis of virion (A) and rhinovirus 14–ICAM-1 complex (B). The surfaces are colored according to charge. (C and D) Electron densities of central slices of cryo-EM reconstructions of virion (C) and rhinovirus 14–ICAM-1 complex (D) with a thickness of 1 Å. White represents high density values. Density representing ICAM-1 is highlighted in magenta. Positions of icosahedral symmetry axes are indicated with an oval, triangle, and pentagon for twofold, threefold, and fivefold axes, respectively. Black arrows in D point toward densities on threefold symmetry axes, which are not present in the virion. Red squares indicate positions of details shown at higher magnification in E and F. (Scale bar, 10 nm.) (E and F) Details of cryo-EM density distribution at the inner face of the capsid on a threefold symmetry axis. Capsid proteins are shown in cartoon representation with VP1 in blue, VP2 in green, VP3 in red, and VP4 in yellow. Cryo-EM density is shown as a semitransparent gray surface. Positions of the first resolved residues from the N termini of VP2 subunits are indicated. E and F show sections of particles with a thickness of 20 Å. (G and H) Comparison of structures of RNA genome interacting with VP2 subunits in virion (G) and rhinovirus 14–ICAM-1 complex (H). The virion contains resolved cryo-EM density corresponding to octanucleotides (G). In contrast, there is a featureless density in the rhinovirus 14–ICAM-1 complex (H). Capsid proteins are shown in cartoon representation, colored as in E and F.

S2). The loss of Trp38–RNA contact relaxes the ordering of segments of the RNA genome that interact with the capsid around twofold symmetry axes in native virions (Fig. 6 G and H). A density corresponding to RNA at the twofold axis in the rhinovirus 14–ICAM-1 complex does not have resolved features (Fig. 6 G and H). The interactions of the N termini of VP1 subunits with the RNA genome remain preserved even after the binding of rhinovirus 14 to ICAM-1 (Fig. 1E). The binding of rhinovirus 14 to ICAM-1 induces a rearrangement of the RNA genome, which may play a role in particle activation, as discussed below.

**Binding of ICAM-1 Primes Particles of Rhinovirus 14 for Activation and Genome Release.** The binding of ICAM-1 to rhinovirus 14 triggers a cascade of structural changes that prepare the particle for activation and subsequent genome release (Fig. 7 A and B). The rotation of VP1 subunits results in a narrowing of the canyon and transmits the conformational changes to the inside of the capsid (Fig. 7 C and D). C termini of VP4 subunits reposition toward twofold symmetry axes, where they are optimally poised for externalization upon particle activation (Fig. 7 E and F). The conformational change to C termini of VP4 subunits uncovers patches of positively

charged residues that attract genomic RNA toward threefold symmetry axes of the capsid (Fig. 7 G and H). The same conformational change prevents the interaction of Trp38 from VP2 with bases from the ordered RNA segments of the genome positioned next to twofold symmetry axes of the capsid (Fig. 7 G and H). Both of these effects result in reorganization of the RNA genome (Fig. 7 G and H). These changes in the capsid and genome structure of rhinovirus 14 induced by ICAM-1 binding are required for efficient genome release at acidic pH (Fig. 2). All in all, 90% of virions of rhinovirus 14 exposed to pH 6.2 remained in their native conformation, whereas the remaining particles were empty (Fig. 2 A, B, E, and F and SI Appendix, Table S1). The structures of rhinovirus 14 in their native conformation and empty capsids at acidic pH were determined to resolutions of 2.8 and 3.9 Å, respectively (Fig. 2 A and B and SI Appendix, Table S1). The exposure of rhinovirus 14 to acidic pH did not induce structural changes in VP4 subunits (Fig. 2 E and F). In contrast, the exposure of rhinovirus 14–ICAM-1 complex to pH 6.2 resulted in activation and genome release from 95% of particles (Fig. 2 C, D, E, and H). The structure of the activated particle was determined to a resolution of 4.0 Å, empty particle to 3.9 Å, and open particle to 22 Å



**Fig. 7.** Overview of structural changes to rhinovirus 14 induced by binding of ICAM-1 that prepare the virus for activation and genome release. (A) Native virion diffuses toward cell membrane (green ribbon) decorated with ICAM-1 molecules (blue sticks with dark blue heads representing domain 1). The virus particle is represented by a central slice with the electron density of VP1 shown in light blue, VP2 in light green, VP3 in light red, VP4 in yellow, the genome in purple, and resolved RNA segments in pink. (B) Rhinovirus 14 is endocytosed by the cell after binding to ICAM-1. (C–H) Sequence of structural changes in virion induced by binding to ICAM-1. C, E, and G represent native virion, whereas D, F, and H show rhinovirus 14–ICAM-1 complex. (C and D) Binding of ICAM-1 induces rotation of VP1 subunit toward VP2 and VP3. Virus components are colored as in A. (E and F) ICAM-1 binding induces disruption of interactions between C terminus of VP4 and N terminus of VP2. C terminus of VP4 repositions from a threefold symmetry axis (indicated with a triangle) toward a twofold symmetry axis (oval). (G and H) Movements of C termini of VP4 subunits uncover positively charged residues around twofold symmetry axes, which attract negatively charged RNA genome. Furthermore, the C terminus of VP4 in the altered conformation covers Trp38 of VP2 and prevents its specific interaction with structured segments of the RNA genome, which relaxes the structure of RNA adjacent to the twofold symmetry axis.

(Fig. 2 C and D and *SI Appendix*, Table S1). The changes in the capsid and genome of rhinovirus 14, which were induced by ICAM-1 binding, may lower the energy barrier of particle activation so that it can be overcome by random fluctuations in particle structure due to thermal motions termed “capsid breathing” (58, 59). This provides a putative explanation of how the reduction of capsid dynamics by antiviral compounds, which target VP1 pockets (59), blocks the activation and genome release of enteroviruses.

The sample of complex of rhinovirus 14 with ICAM-1 exposed to acidic pH contained 7% empty particles missing a pentamer of capsid protein protomers (Fig. 2D). Open particles were previously speculated to enable enterovirus genome release (33). The expulsion of pentamers of capsid proteins results in the formation of a large hole in the capsid, which enables the diffusion of the RNA genome from the capsid within a microsecond (33, 60). The rapid release of a genome may be connected to its subsequent transport across the endosome membrane into the cytoplasm (61, 62).

Structural characterization of the rhinovirus 14–ICAM-1 complex at atomic resolution provides detailed information about the conformational changes of both the receptor and virus that are required for its binding. Additionally, it provides insight into the structural changes of the virus that enable subsequent particle activation and genome release. In combination, these results provide the basis for the design of compounds that block enterovirus infection.

## Materials and Methods

**Expression and Purification of ICAM-1 D1 to D5.** The extracellular part of ICAM-1 containing domains D1 to D5 was produced using the MutiBac system (Geneva Biotech). The full-length gene of ICAM-1 was a gift from Timothy Springer (Harvard Medical School, Boston, MA) (Addgene plasmid No. 8632; <http://addgene.org/8632>; RRID: Addgene 8632). The sequence encoding domains D1 to D5, the secretion signal peptide, and the C-terminal 10 His-tag were inserted into the pACEBac1 vector at the restriction site BamHI. The recombinant bacmid with the target sequence was prepared by recombination in DH10EMBaY *Escherichia coli* cells (Geneva Biotech). The recombinant baculovirus was prepared by transfecting SF9 cells with the recombinant bacmid. A total of 250 mL of the culture of SF9 cells were infected with the recombinant baculovirus and incubated for 96 h at 27 °C with 120 rpm shaking. The produced protein was secreted into the medium. Cells and cell debris were pelleted by centrifugation at 20,000 × *g* at 4 °C for 15 min. The supernatant was filtered through a 0.2 μm

filter (Corning) and loaded into a HisTrap column (GE Healthcare) equilibrated in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4). Most of the impurities were removed by washing with PBS containing 70 mM imidazole. His-tagged ICAM-1 D1 to D5 was eluted using PBS with 500 mM imidazole. The eluted protein was buffer exchanged into PBS using 30 kDa cutoff centrifugal concentrators (Millipore, Merck). The target protein was further purified by size-exclusion chromatography, using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). Fractions containing ICAM-1 D1 to D5 were pooled and concentrated using centrifugal concentrators (Millipore, Merck) to a final concentration of 3.5 mg/mL.

**Rhinovirus 14 Purification.** Rhinovirus-14 strain 1059, obtained from ATCC, was propagated in HeLa-H1 (ATCC CRL195) cells cultivated in Dulbecco’s modified Eagle’s medium enriched with 10% fetal bovine serum. HeLa cells grown to 100% confluence (100 tissue culture dishes, 150 mm diameter) were infected with a multiplicity of infection of 0.1. The infection was allowed to progress for 36 h until a complete cytopathic effect was observed. The media and cells were harvested and centrifuged at 15,000 × *g* at 10 °C for 30 min. The resulting pellet was subjected to three freeze–thaw cycles and resuspended in 5 mL PBS followed by homogenization in a Dounce tissue grinder. Cell debris was separated from the supernatant by centrifugation at 4,000 × *g* for 30 min at 10 °C. The virus-containing supernatant was combined with the media from infected cells. The virus particles were precipitated by the addition of PEG-8000 and NaCl to final concentrations of 12.5% (wt/vol) and 500 mM, respectively, and incubation overnight at 10 °C. The precipitated virus was pelleted by centrifugation at 15,000 × *g* at 10 °C for 30 min. The pellet was resuspended in 20 mL PBS with 5 mM MgCl<sub>2</sub>. The sample was subjected to DNase (10 μg/mL final concentration) and RNase (10 μg/mL final concentration) treatment at room temperature for 30 min. Subsequently, trypsin was added to a final concentration of 0.5 μg/mL, and the sample was incubated at 37 °C for 15 min. EDTA (pH = 9.5) and Nonidet P-40 (Sigma-Aldrich) were added to final concentrations of 15 mM and 1% (vol/vol), respectively. The virus was pelleted through a 30% (wt/vol) sucrose cushion by centrifugation at 210,000 × *g* in an Optima ×80 ultracentrifuge (Beckman Coulter) using a Ti50.2 rotor. The pelleted virus particles were resuspended in 2 mL PBS, loaded on the top of a 60% (wt/vol) CsCl solution in PBS, and centrifuged at 160,000 × *g* in an Optima ×80 ultracentrifuge using a Beckman Coulter SW41Ti rotor at 10 °C for 24 h. Opaque bands containing virus particles were harvested and subjected to buffer exchange in PBS using a Centricon Plus-70 centrifugal filter (Millipore) with a 100 kDa cutoff. The final concentration of rhinovirus 14 was 0.5 mg/mL. Purified virus was kept at 4 °C.

**Preparation of Rhinovirus 14–ICAM-1 Complex and Exposure to Acidic pH.** The complex of rhinovirus 14 with ICAM-1 was prepared by mixing rhinovirus 14 with ICAM-1 at a molar ratio of 1:100 at pH 7.4 and incubating the mixture

for 30 min at 34 °C. Rhinovirus 14 and the rhinovirus 14–ICAM-1 complex were transferred to acidic pH using DyeEx 2.0 (QIAGEN) spin columns containing PBS with pH 6.2. The samples were applied onto the columns and eluted by 1 min of centrifugation at 1,200 × g. The samples were incubated at pH 6.2 at 34 °C for 2.5 min, including the centrifugation step.

**Cryo-EM Sample Preparation and Data Collection.** For vitrification, 3 μL virus samples were applied onto a holey carbon-coated copper grid (R2/1, mesh 300, Quantifoil), blotted for 2 s, and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids for virion reconstruction were prepared by vitrifying a virus sample with a concentration of 0.5 mg/mL. The grids (except for those with rhinovirus 14 at pH 6.2) were then transferred to a Titan Krios electron microscope, operating at 300 kV at cryogenic conditions, equipped with a Falcon III direct electron detector (Thermo Fisher Scientific). The illuminating beam was aligned for parallel illumination in NanoProbe mode. Low-dose imaging was used with a total dose of 84.7 e<sup>-</sup>/Å<sup>2</sup>. Nominal magnification was set to 75,000×, resulting in a calibrated pixel size of 1.063 Å. The dataset was recorded automatically using EPU software (Thermo Fisher Scientific) in fast acquisition mode, using large image shifts. The samples of rhinovirus 14 and rhinovirus 14–ICAM-1 complex were recorded using five acquisitions per hole, nine holes per stage shift. The sample of rhinovirus 14–ICAM-1 complex at pH 6.2 was recorded using seven acquisitions per hole, nine holes per stage shift. The exposure time was set to 1 s, and each micrograph was recorded as a movie containing 40 frames. The target defocus range was –0.5 to –2.4 μm.

Electron micrographs from the sample of rhinovirus 14 at pH 6.2 were collected using a Talos Arctica electron microscope (Thermo Fisher Scientific), operated at 200 kV under cryogenic conditions, equipped with a Falcon III direct electron detector (Thermo Fisher Scientific). The illuminating beam was aligned for parallel illumination in NanoProbe mode. Low-dose imaging was used with a total dose of 34.1 e<sup>-</sup>/Å<sup>2</sup>. Nominal magnification was set to 120,000×, resulting in a calibrated pixel size of 1.22 Å. The dataset was recorded automatically using EPU software (Thermo Fisher Scientific) in fast acquisition mode, using five acquisitions per hole, nine holes per stage shift. The exposure time was set to 1 s, and each micrograph was recorded as a movie containing 40 frames. The target defocus range was –0.5 to –2.4 μm.

**Image Processing.** The beam-induced movements within one micrograph were corrected with the software MotionCorr2 using 5 × 5 patches (63). The motion-corrected micrographs were dose weighted, and defocus values were estimated using the program gCTF (64). Using crYOLO box manager (65), 200 particles were boxed manually and used as a template for ab initio model training. The resulting crYOLO model was used to pick particles. The particles were extracted using Relion3.1 (66) with a box size of 546 px. The particles were binned to a box size of 150 × 150 px and subjected to reference-free two-dimensional classifications in Relion3.1 (66). Particles from classes exhibiting high-resolution features were used for de novo model calculation with imposed icosahedral symmetry, using stochastic gradient descent as implemented in Relion3.1 (66). The resulting three-dimensional (3D) volume was used as a starting model for autorefinement in Relion3.1. After initial autorefinement, 3D classification in Relion3.1 was performed, omitting the alignment step. Particles belonging to the best class were reextracted and recentered box-size 512 × 512 px for rhinovirus 14 particles without ICAM-1 and 546 × 546 px for the rhinovirus 14–ICAM-1 complex. Reextracted particles were subjected to another round of autorefinement in Relion3.1. Particles were then sorted into nine optic groups. The optic groups were determined by the position of the image shift used for the acquisition, whereas all the acquisition areas from the same foil hole were considered as one optic group. Therefore, only large (interhole) image shifts were considered as separate optic groups. Magnification correction was performed using Relion3.1, followed by beam-tilt correction, and subsequently by the estimation of third- and fourth-order Zernike polynomials. The aberration-corrected particles were further subjected to per-particle defocus and astigmatism correction and estimation of the CTF envelope function (CTF B-factor fitting). The particles were subjected to autorefinement with imposed icosahedral symmetry. Ewald sphere correction was performed as implemented in `relion_reconstruct.py` in Relion3.1 (67). The resulting map was used for Bayesian polishing of particles with default parameters. The polished particles were used for 3D autorefinement and CTF refinement followed by another 3D autorefinement. Finally, Ewald sphere correction was performed. The final map was threshold masked, divided by a modulation transfer function, and B-factor sharpened using Relion3.1. Local resolutions were estimated using the program MonoRes implemented in the Scipion software package (68, 69). Map B-factor sharpening based on local resolution estimation from MonoRes was performed using the program LocalDeblur (70). The dataset of rhinovirus 14 with ICAM-1 exposed to pH 6.2 contained empty capsids missing

pentamers. These were identified by 3D classification with C5 symmetry, using the complete empty capsid as an initial model. Subsequent 3D autorefinement was performed with C5 symmetry. Neither CTF refinement nor Bayesian polishing were applied to this subset of particles.

**Building and Refinement of Atomic Structure.** The electrostatic potential map from cryo-EM reconstruction was oriented to the standard 222 icosahedral crystallographic orientation. The origin of the map was moved from the 0,0,0, coordinate to the center of the particle using the program mapman (71). The map was normalized and converted to crystallographic space group P23 using the CCP4i software suite (72). The higher-symmetry space group was used to reduce the computational demands of the model refinement. Crystal structures of rhinovirus 14 (Protein Data Bank [PDB]: 4RHV) and domain 1 of ICAM-1 (PDB: 1IC1) were manually fitted into the cryo-EM maps using the program Chimera and refined with the tool “Fit in map” (73). The cryo-EM structure of an empty particle of rhinovirus 14 in complex with a Fab fragment of antibody (PDB: 5W3O) was used as a starting model for the building of activated and empty particles (74). The fitted models were subjected to multiple rounds of real-space refinement in Phenix (version dev-3765), reciprocal-space refinement in REFMAC5, combined with manual corrections in Coot 0.9 and ISOLDE (75–78). Hydrogen atoms were taken into account during the real-space refinement, whereas they were ignored in the reciprocal-space refinement. Waters were added automatically by the program “find waters” in Coot 0.9 and validated manually. Model validation parameters were calculated using MolProbity server and EMringer as implemented in phenix (79, 80). The RNA octanucleotide sequence in the native virion of rhinovirus 14 was initially built using the program Coot and refined with restraints using the program ISOLDE (78). Structural comparisons were performed in Chimera (73). Hydrogen bonds, salt bridges, and residues involved in the binding interface and buried surface areas were calculated using the program PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa>). Roadmaps were produced using the program Rivem (81).

**Mass Spectrometry Analyses.** Purified samples of ICAM-1 and the complex of rhinovirus 14 with ICAM-1 were digested with alpha-lytic protease (EC 3.4.21.12, Sigma-Aldrich catalog No. A6362) for 2 h at 37 °C with shaking at 700 rpm. Half of the volume of each sample was then reduced using 10 mM DTT (for 45 min at 57 °C with shaking at 700 rpm). After adding polyethylene glycol to a final concentration of 0.001%, the peptides were extracted from the vials using 25% formic acid/acetonitrile (1:1 vol/vol mixture) and vacuum concentrated. The peptide mixture was subjected to liquid-chromatography-mass spectrometry (LC-MS)/MS analysis using a RSLCnano system (ThermoFisher Scientific) coupled to an Impact II Qq-Time-Of-Flight mass spectrometer (Bruker). Prior to LC separation, peptides were online concentrated in a trap column (100 μm × 30 mm) filled with 3.5 μm X-Bridge BEH 130 C18 sorbent (Waters). The peptides were separated using an Acclaim Pepmap100 C18 column (3 μm particles, 75 μm × 500 mm; ThermoFisher Scientific) by the following LC gradient program (mobile phase A: 0.1% formic acid in water; mobile phase B: 0.1% formic acid in 80% acetonitrile; 300 nL/min): the gradient elution started at 1% of mobile phase B and increased to 56% over the first 50 min, then increased linearly to 80% of mobile phase B over the next 5 min and remained at this state for the final 10 min. Equilibration of the trapping column and the column was done prior to sample injection into the sample loop. The analytical column outlet was directly connected to a CaptiveSpray nanoBooster ion source (Bruker). The nanoBooster was filled with acetonitrile. MS and MS/MS spectra were acquired in a data-dependent strategy with a 3 s cycle time. The mass range was set to 150 to 2,200 m/z and precursors were selected from 300 to 2,000 m/z. The acquisition speed of MS and MS/MS scans was 2 Hz and 4 to 16 Hz, respectively. The speed of MS/MS spectra acquisition was based on precursor intensity. The preprocessing of the mass spectrometric data including recalibration, compound detection, and charge deconvolution was carried out using DataAnalysis software (version 4.2 SR1; Bruker).

The obtained data were searched with an in-house Mascot search engine (version 2.4.1; Matrixscience) against a custom database involving the ICAM-1 sequence and cRAP entries (downloaded from <https://www.thegpm.org/crap/>). The database searches were done without enzyme specificity and with oxidation (M) as a variable modification. The mass tolerances for peptides and MS/MS fragments were 10 ppm and 0.1 Da, respectively. Only peptides with a statistically significant peptide score ( $P < 0.05$ ) were considered, and the obtained MS/MS data were validated manually.

**Multiple Sequence Alignment.** Multiple sequence alignment of capsid proteins of selected viruses from the family *Picornaviridae* was performed in the Clustal Omega server (82). The multiple sequence alignment was visualized in the software Jalview 2.11.1.3 (83).

**Data Availability.** The cryo-EM maps and coordinates were deposited under the following accession codes: virion of rhinovirus 14 at neutral pH Electron Microscopy Data Bank [EMD-12171](#) and PDB [7BG6](#); rhinovirus 14-ICAM-1 complex at neutral pH [EMD-12172](#) and PDB [7BG7](#); rhinovirus 14 in native conformation at acidic pH [EMD-12599](#) and PDB [7NUQ](#); empty particle of rhinovirus 14 at acidic pH [EMD-12597](#) and PDB [7NUQ](#); rhinovirus 14 in native conformation at acidic pH originating from complex with ICAM-1 [EMD-12596](#) and PDB [7NUN](#); activated particle originating from complex with ICAM-1 at acidic pH [EMD-12594](#) and PDB [7NUL](#); empty particle originating from complex with ICAM-1 at acidic pH [EMD-12595](#) and PDB [7NUM](#); and open particle originating from complex with ICAM-1 at acidic pH [EMD-12598](#).

1. S. E. Jacobs, D. M. Lamson, K. St George, T. J. Walsh, Human rhinoviruses. *Clin. Microbiol. Rev.* **26**, 135–162 (2013).
2. A. M. Fendrick, A. S. Monto, B. Nightengale, M. Sarnes, The economic burden of non-influenza-related viral respiratory tract infection in the United States. *Arch. Intern. Med.* **163**, 487–494 (2003).
3. J. S. Bertino, Cost burden of viral respiratory infections: Issues for formulary decision makers. *Am. J. Med.* **112** (suppl. 6A), 42S–49S (2002).
4. A. C. Palmenberg *et al.*, Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science* **324**, 55–59 (2009).
5. J. M. Greve *et al.*, The major human rhinovirus receptor is ICAM-1. *Cell* **56**, 839–847 (1989).
6. C. R. Uncapher, C. M. DeWitt, R. J. Colonna, The major and minor group receptor families contain all but one human rhinovirus serotype. *Virology* **180**, 814–817 (1991).
7. F. Hofer *et al.*, Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1839–1842 (1994).
8. Y. A. Bochkov *et al.*, Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5485–5490 (2015).
9. A. I. Wells, C. B. Coyne, Enteroviruses: A gut-wrenching game of entry, detection, and evasion. *Viruses* **11**, 460 (2019).
10. J. M. Bergelson *et al.*, Decay-accelerating factor (CD55), a glycosylphosphatidylinositol-anchored complement regulatory protein, is a receptor for several echoviruses. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6245–6248 (1994).
11. Y. Nishimura *et al.*, Human P-selectin glycoprotein ligand-1 is a functional receptor for enterovirus 71. *Nat. Med.* **15**, 794–797 (2009).
12. J. Staring *et al.*, KREMEN1 is a host entry receptor for a major group of enteroviruses. *Cell Host Microbe* **23**, 636–643.e5 (2018).
13. P. Y. Su *et al.*, Cell surface sialylation affects binding of enterovirus 71 to rhabdomyosarcoma and neuroblastoma cells. *BMC Microbiol.* **12**, 162 (2012).
14. C. L. Mendelsohn, E. Wimmer, V. R. Racaniello, Cellular receptor for poliovirus: Molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily. *Cell* **56**, 855–865 (1989).
15. J. M. Bergelson *et al.*, Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**, 1320–1323 (1997).
16. S. Yamayoshi *et al.*, Scavenger receptor B2 is a cellular receptor for enterovirus 71. *Nat. Med.* **15**, 798–801 (2009).
17. T. J. Tuthill, E. Groppelli, J. M. Hogle, D. J. Rowlands, Picornaviruses. *Curr. Top. Microbiol. Immunol.* **343**, 43–89 (2010).
18. R. Basavappa *et al.*, Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: Structure of the empty capsid assembly intermediate at 2.9 Å resolution. *Protein Sci.* **3**, 1651–1669 (1994).
19. J. J. Holland, E. D. Kiehn, Specific cleavage of viral proteins as steps in the synthesis and maturation of enteroviruses. *Proc. Natl. Acad. Sci. U.S.A.* **60**, 1015–1022 (1968).
20. M. Procházková *et al.*, Virion structures and genome delivery of honeybee viruses. *Curr. Opin. Virol.* **45**, 17–24 (2020).
21. M. G. Rossmann *et al.*, Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145–153 (1985).
22. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–373 (1978).
23. N. Verdaguer, D. Blaas, I. Fita, Structure of human rhinovirus serotype 2 (HRV2). *J. Mol. Biol.* **300**, 1179–1194 (2000).
24. M. A. Oliveira *et al.*, The structure of human rhinovirus 16. *Structure* **1**, 51–68 (1993).
25. M. Smyth, T. Pettitt, A. Symonds, J. Martin, Identification of the pocket factors in a picornavirus. *Arch. Virol.* **148**, 1225–1233 (2003).
26. A. Pickl-Herk *et al.*, Uncoating of common cold virus is preceded by RNA switching as determined by X-ray and cryo-EM analyses of the subviral A-particle. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20063–20068 (2013).
27. H. C. Levy, M. Bostina, D. J. Filman, J. M. Hogle, Catching a virus in the act of RNA release: A novel poliovirus uncoating intermediate characterized by cryo-electron microscopy. *J. Virol.* **84**, 4426–4441 (2010).
28. D. Garriga *et al.*, Insights into minor group rhinovirus uncoating: The X-ray structure of the HRV2 empty capsid. *PLoS Pathog.* **8**, e1002473 (2012).
29. X. Wang *et al.*, A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nat. Struct. Mol. Biol.* **19**, 424–429 (2012).
30. K. L. Shingler *et al.*, The enterovirus 71 A-particle forms a gateway to allow genome release: A cryoEM study of picornavirus uncoating. *PLoS Pathog.* **9**, e1003240 (2013).
31. J. J. Seitsonen *et al.*, Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *J. Virol.* **86**, 7207–7215 (2012).
32. D. M. Belnap *et al.*, Molecular tectonic model of virus structural transitions: The putative cell entry states of poliovirus. *J. Virol.* **74**, 1342–1354 (2000).
33. D. Buchta *et al.*, Enterovirus particles expel capsid pentamers to enable genome release. *Nat. Commun.* **10**, 1138 (2019).
34. J. Ren *et al.*, Picornavirus uncoating intermediate captured in atomic detail. *Nat. Commun.* **4**, 1929 (2013).
35. L. Bauer, H. Lyoo, H. M. van der Schaar, J. R. Strating, F. J. van Kuppeveld, Direct-acting antivirals and host-targeting strategies to combat enterovirus infections. *Curr. Opin. Virol.* **24**, 1–8 (2017).
36. T. J. Smith *et al.*, The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* **233**, 1286–1293 (1986).
37. A. van de Stolpe, P. T. van der Saag, Intercellular adhesion molecule-1. *J. Mol. Med. (Berl.)* **74**, 13–33 (1996).
38. J. M. Casasnovas, T. Stehle, J. H. Liu, J. H. Wang, T. A. Springer, A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4134–4139 (1998).
39. C. Chothia, E. Y. Jones, The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **66**, 823–862 (1997).
40. J. Wang, T. A. Springer, Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunol. Rev.* **163**, 197–215 (1998).
41. Y. Harpaz, C. Chothia, Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528–539 (1994).
42. P. R. Kolatkar *et al.*, Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. *EMBO J.* **18**, 6249–6259 (1999).
43. C. Xiao *et al.*, Discrimination among rhinovirus serotypes for a variant ICAM-1 receptor molecule. *J. Virol.* **78**, 10034–10044 (2004).
44. C. Xiao *et al.*, Interaction of coxsackievirus A21 with its cellular receptor, ICAM-1. *J. Virol.* **75**, 2444–2451 (2001).
45. N. H. Olson *et al.*, Structure of a human rhinovirus complexed with its receptor molecule. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 507–511 (1993).
46. L. Xing, J. M. Casasnovas, R. H. Cheng, Structural analysis of human rhinovirus complexed with ICAM-1 reveals the dynamics of receptor-mediated virus uncoating. *J. Virol.* **77**, 6101–6107 (2003).
47. H. Hoover-Litty, J. M. Greve, Formation of rhinovirus-soluble ICAM-1 complexes and conformational changes in the virion. *J. Virol.* **67**, 390–397 (1993).
48. J. M. Hogle, M. Chow, D. J. Filman, Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229**, 1358–1365 (1985).
49. J. K. Muckelbauer *et al.*, Structure determination of coxsackievirus B3 to 3.5 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **51**, 871–887 (1995).
50. E. Hendry *et al.*, The crystal structure of coxsackievirus A9: New insights into the uncoating mechanisms of enteroviruses. *Structure* **7**, 1527–1538 (1999).
51. C. Xiao *et al.*, The crystal structure of coxsackievirus A21 and its interaction with ICAM-1. *Structure* **13**, 1019–1033 (2005).
52. E. R. McFadden Jr. *et al.*, Thermal mapping of the airways in humans. *J. Appl. Physiol.* **58**, 564–570 (1985).
53. L. Xing *et al.*, Distinct cellular receptor interactions in poliovirus and rhinoviruses. *EMBO J.* **19**, 1207–1216 (2000).
54. A. McClelland *et al.*, Identification of monoclonal antibody epitopes and critical residues for rhinovirus binding in domain 1 of intercellular adhesion molecule 1. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7993–7997 (1991).
55. M. A. Sun *et al.*, Prediction of reversible disulfide based on features from local structural signatures. *BMC Genomics* **18**, 279 (2017).
56. R. B. Register, C. R. Uncapher, A. M. Naylor, D. W. Lineberger, R. J. Colonna, Human-murine chimeras of ICAM-1 identify amino acid residues critical for rhinovirus and antibody binding. *J. Virol.* **65**, 6589–6596 (1991).
57. N. Verdaguer, I. Fita, M. Reithmayer, R. Moser, D. Blaas, X-ray structure of a minor group human rhinovirus bound to a fragment of its cellular receptor protein. *Nat. Struct. Mol. Biol.* **11**, 429–434 (2004).
58. Q. Li, A. G. Yafal, Y. M. Lee, J. Hogle, M. Chow, Poliovirus neutralization by antibodies to internal epitopes of VP4 and VP1 results from reversible exposure of these sequences at physiological temperature. *J. Virol.* **68**, 3965–3970 (1994).
59. J. K. Lewis, B. Bothner, T. J. Smith, G. Siuzdak, Antiviral agent blocks breathing of the common cold virus. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6774–6778 (1998).
60. K. Škubník *et al.*, Capsid opening enables genome release of flaviviruses. *Sci. Adv.* **7**, eabd7130 (2021).
61. M. Bostina, H. Levy, D. J. Filman, J. M. Hogle, Poliovirus RNA is released from the capsid near a twofold symmetry axis. *J. Virol.* **85**, 776–783 (2011).

Hřebík *et al.*

ICAM-1 induced rearrangements of capsid and genome prime rhinovirus 14 for activation and uncoating

PNAS | 9 of 10

<https://doi.org/10.1073/pnas.2024251118>

62. M. Strauss, H. C. Levy, M. Bostina, D. J. Filman, J. M. Hogle, RNA transfer from poliovirus 135S particles across membranes is mediated by long umbilical connectors. *J. Virol.* **87**, 3903–3914 (2013).
63. S. Q. Zheng *et al.*, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
64. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
65. T. Wagner *et al.*, SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
66. S. H. Scheres, A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
67. C. J. Russo, R. Henderson, Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy* **187**, 26–33 (2018).
68. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
69. J. L. Vilas *et al.*, MonoRes: Automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* **26**, 337–344.e4 (2018).
70. E. Ramirez-Aportela *et al.*, Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* **36**, 765–772 (2020).
71. G. J. Kleywegt, T. A. Jones, xdlMAPMAN and xdlDATAMAN—Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 826–828 (1996).
72. E. Potterton, P. Briggs, M. Turkenburg, E. Dodson, A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1131–1137 (2003).
73. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
74. Y. Dong *et al.*, Antibody-induced uncoating of human rhinovirus B14. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8017–8022 (2017).
75. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
76. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
77. P. D. Adams *et al.*, PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
78. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018).
79. V. B. Chen *et al.*, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
80. B. A. Barad *et al.*, EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
81. C. Xiao, M. G. Rossmann, Interpretation of electron density with stereographic roadmap projections. *J. Struct. Biol.* **158**, 182–187 (2007).
82. F. Madeira *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
83. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).



**Article 3: Enterovirus particles expel capsid pentamers to enable genome release**

We discovered and described a new mechanism of genome release in Enterovirus 18 (EV18), which may be common to most enteroviruses. It was known before that receptor binding and exposure to acidic pH causes uncoating of enteroviruses. However, the exact mechanism was not well understood. With my colleagues David Buchta (first-author) and Yevgen Levdansky, we observed that when EV18 is exposed to acidic pH, it releases its genome. That was something we expected already. However, after a more detailed inspection of the data, we noticed that some of the particles have "curly" density in their vicinity. We concluded that it is probably genomic RNA being released from the particles. Thus, we selected such particles from the micrographs and started the reconstruction. After 3D classifications, we gained three classes of particles – particles lacking one capsid pentamer, two capsid pentamers, and three capsid pentamers. Now it was clear that the EV18 must expel one or more pentamers to release its genome. We collaborated with the computational group of Robert Vácha, who performed coarse grained molecular dynamics simulation of the genome release. It confirmed the observed data.

It was speculated before that the genome release of *Enteroviruses* occurs through one of the pores that appear on the virus capsid after activation of the viral particle caused by the acidic environment of the endosome [85, 87, 125, 126, 127, 128]. However, this would require the RNA to uncoil its double stranded segments and to lose its secondary structures. On the other hand, the large opening formed by losing up to three pentamers enables the genome to escape at once, which is an entropically favored process. Such release is fast (approximately 100 ns), which limits the time the RNAses have to reach the genome for degradation. The open particles were also identified for other *Enteroviruses* such as the above-mentioned rhinovirus 14 or Enterovirus 30. Thus, it is possible that the genome release mechanism by expelling one or more pentamers is widespread in the genus of *Enterovirus*.



ARTICLE

<https://doi.org/10.1038/s41467-019-09132-x>

OPEN

# Enterovirus particles expel capsid pentamers to enable genome release

David Buchta <sup>1</sup>, Tibor Füzik <sup>1</sup>, Dominik Hřebík <sup>1</sup>, Yevgen Levdansky<sup>1,3</sup>, Lukáš Sukeník<sup>1,2</sup>, Liya Mukhamedova<sup>1</sup>, Jana Moravcová<sup>1</sup>, Robert Vácha<sup>1,2</sup> & Pavel Plevka <sup>1</sup>

Viruses from the genus *Enterovirus* are important human pathogens. Receptor binding or exposure to acidic pH in endosomes converts enterovirus particles to an activated state that is required for genome release. However, the mechanism of enterovirus uncoating is not well understood. Here, we use cryo-electron microscopy to visualize virions of human echovirus 18 in the process of genome release. We discover that the exit of the RNA from the particle of echovirus 18 results in a loss of one, two, or three adjacent capsid-protein pentamers. The opening in the capsid, which is more than 120 Å in diameter, enables the release of the genome without the need to unwind its putative double-stranded RNA segments. We also detect capsids lacking pentamers during genome release from echovirus 30. Thus, our findings uncover a mechanism of enterovirus genome release that could become target for antiviral drugs.

<sup>1</sup>Central European Institute of Technology, Masaryk University, Kamenice 5, Brno 625 00, Czech Republic. <sup>2</sup>Faculty of Science, Masaryk University, Kamenice 5, Brno 625 00, Czech Republic. <sup>3</sup>Present address: Max Planck Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany. Correspondence and requests for materials should be addressed to P.P. (email: [pavel.plevka@ceitec.muni.cz](mailto:pavel.plevka@ceitec.muni.cz))

Viruses from the family *Picornaviridae* form non-enveloped icosahedral virions that are about 30 nm in diameter. Picornavirus capsids protect 8000-nucleotide-long single-stranded RNA genomes, which are translated into polyproteins that are co-translationally and post-translationally cleaved by viral proteases to produce structural (capsid-forming) and non-structural proteins<sup>1</sup>. The capsid proteins VP1, VP2, VP3, and VP4 originating from a single polyprotein form a protomer, the basic building block of the icosahedral capsid. The entire capsid consists of 60 such protomers, arranged in 12 pentamer units.

The interactions of enteroviruses with receptors or exposure to acidic pH in endosomes induce conformational changes in virions into an activated state characterized by increased particle diameter, reduced contact areas between pentamers of capsid protein protomers, release of VP4 subunits from particles, and externalization of the N termini of VP1 subunits<sup>1–5</sup>. The activated particles of numerous enteroviruses were shown to contain openings along two-fold ( $5 \times 10 \text{ \AA}$ ) or five-fold (diameters of up to  $8 \text{ \AA}$ ) axes of icosahedral symmetry of their capsids<sup>2–6</sup>. It has been speculated that these pores serve for the release of enterovirus genomes<sup>3,6–10</sup>. However, capsids of viruses from the families *Reoviridae* and *Totiviridae* that release single-stranded RNAs as part of their replication cycles contain circular pores larger than  $15 \text{ \AA}$  in diameter<sup>11,12</sup>. The size of the pores in enterovirus capsids is not sufficient for the passage of single-stranded RNA<sup>2–6,13–15</sup>. Furthermore, enterovirus genomes contain sequences that form double-stranded RNA segments, which fold into three-dimensional (3D) structures, such as the internal ribosomal entry site required to initiate translation of viral RNA<sup>16</sup>. If these double-stranded RNA segments were present inside enterovirus particles, then the genome release would require either the opening of pores larger than  $40 \text{ \AA}$  in diameter, or a mechanism to unwind the double-stranded RNA. However, there is no evidence of an association between enzymes with RNA helicase activity and enterovirus virions<sup>1,17</sup>. The structures of enterovirus particles before and after genome release have been characterized at high resolution<sup>3,6–10</sup>. Most of these cryo-electron microscopy (cryo-EM) and X-ray crystallography studies imposed icosahedral symmetry during the structure determination process and were not aimed at identifying the unique site of genome exit<sup>3–7,9</sup>. Asymmetric single-particle reconstruction and sub-tomogram averaging studies, at a resolution of  $50 \text{ \AA}$ , were used to indicate that RNA exits poliovirus particles close to a two-fold axis<sup>8</sup>. The end stage of the enterovirus genome release are the empty capsids, the structures of which were determined for several enteroviruses<sup>4,5</sup>.

Here we show that the exit of the RNA from the particles of echovirus 18 and 30 requires capsid opening and results in a loss of up to three adjacent capsid protein pentamers. The large openings in the capsid enable the release of the genomes without uncoiling their double-stranded RNA segments.

## Results and Discussion

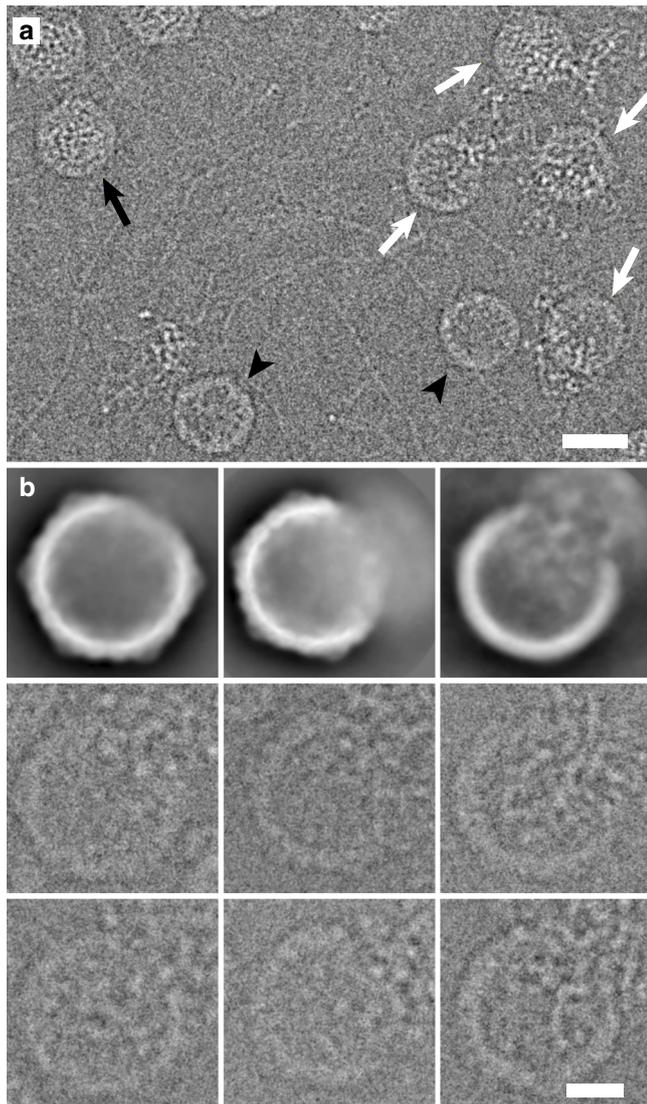
### Opening of particles enables genome release of echovirus 18.

We imaged enterovirus particles in the process of genome release by cryo-EM. Specifically, we performed cryo-EM of echovirus 18 virions exposed to pH 6.0 for 10 min, mimicking the acidic environment that the virus encounters in endosomes (Fig. 1a, Supplementary Fig. 1). Reference-free two-dimensional (2D) class averages show that the particles releasing genomes lack parts of their capsids (Fig. 1b). Asymmetric reconstruction combined with 3D classification identified subpopulations of echovirus 18 particles that lacked up to three pentamers of capsid protein protomers (Fig. 2a–c, Supplementary Fig. 2a). The missing pentamers always formed a single compact opening through the

capsid (Fig. 2a–c). We call the particles lacking one or several pentamers open particles. The remaining particles with complete capsids were either activated particles or empty capsids (Supplementary Fig. 2a). We did not detect native echovirus 18 virions that lacked pentamers at neutral pH. The asymmetric reconstructions of the open particles were determined to resolutions better than  $9 \text{ \AA}$  (Supplementary Table 1, Supplementary Figs 3–5). The absence of one pentamer of capsid protein protomers creates a  $120 \text{ \AA}$ -diameter pore in the capsid (Fig. 2a). The openings formed by the removal of one or more pentamers are sufficiently large to allow release of the viral RNA, even if the genome contains double-stranded RNA segments. Individual micrographs of the open particles frequently show multiple strands of RNA passing through the pore (Fig. 1b). In the 3D reconstructions, the capsid openings contain featureless electron densities with average values two times lower than those of the protein capsid (Fig. 2d–i). This diffuse electron density corresponds to an average of the RNA genomes escaping from the virions, which have unique conformations in each of the particles included in the reconstructions (Fig. 1b). In contrast to the formation of the open particles of echovirus 18, complete capsids were observed in the cryo-EM study of the genome release of poliovirus<sup>8</sup>. However, the poliovirus uncoating was induced by the exposure of particles to  $56 \text{ }^\circ\text{C}$ , which may have affected the secondary structure of the genome and the mechanism of its release. A comparison of the distribution of charges and hydrophobic areas at the inter-pentamer contacts (Supplementary Fig. 6) and a comparison of the inter-pentamer buried surface areas (Supplementary Table 2) of enteroviruses do not provide evidence why distinct enteroviruses should employ different genome release mechanisms.

### High-resolution structures of open particles of echovirus 18.

Reconstructions of the open particles missing one, two, or three pentamers with imposed five-fold, two-fold, and three-fold symmetries, respectively, were determined to resolutions of  $3.8$ ,  $4.1$ , and  $3.7 \text{ \AA}$ , which allowed their molecular structures to be built (Fig. 3, Supplementary Fig. 4, Supplementary Table 1). Except for the missing pentamers, the open particles are similar in structure to that of the activated echovirus 18 particle: increased diameter relative to the native virus, reduced inter-pentamer contacts, absence of VP4 subunits, and holes along icosahedral two-fold symmetry axes (Fig. 3, Supplementary Fig. 7)<sup>2–5,14,15</sup>. The capsid proteins next to the missing pentamers are more mobile than the rest of the capsid, as indicated by the three times higher temperature factors than the remainder of the capsid. Residues 1–42 from the N termini of VP1 subunits, which were previously shown to interact with membranes and facilitate enterovirus genome delivery into the cytoplasm<sup>1,18,19</sup>, are not resolved in the electron density map, indicating that their structures differ among particles. The N-termini of some of the VP1 subunits could reach out of the capsid through the openings formed by the missing pentamers. It has been shown that the RNA genome of poliovirus is protected against RNase A degradation during uncoating and transfer across the membrane<sup>20</sup>. It may seem that opening the particles of echovirus 18 exposes their genomes for degradation. Nevertheless, in silico simulations (Fig. 4, Supplementary Movie 1) and considerations of genome diffusion from the capsid (for details, see Methods) show that the large capsid opening results in a microsecond release time of the genome. The short time required for genome release limits the potential for its degradation. The externalized VP4 subunits and N termini of VP1 were shown to be required for the subsequent transport of enterovirus genomes across the endosome membranes<sup>18,21–24</sup>.



**Fig. 1** Particles of echovirus 18 in the process of genome release lack parts of their capsids. **a** Cryo-electron micrograph of echovirus 18 particles captured in the process of genome release after incubation at acidic pH. Black arrow indicates activated particle, white arrows particles in the process of genome release, and black arrowheads empty particles. Scale bar represents 25 nm. **b** Reference-free two-dimensional class averages showing echovirus 18 particles in the process of RNA release lacking parts of their capsids. Two representative electron micrographs of particles are shown for each class average. Scale bar represents 10 nm

### Re-structuring of echovirus 18 genome enables capsid opening.

Empty particles of echovirus 18 formed after genome release, induced by exposure to acidic pH for 2 h, were used to calculate their icosahedral reconstruction to a resolution of 3.2 Å (Supplementary Table 1, Supplementary Figs 4, 7). This is consistent with previous experiments, which show that empty capsids of enteroviruses are stable *in vivo* and *in vitro*<sup>1–5,20</sup>. Because the empty capsids of echovirus 18 are stable under the experimental conditions, the discharged pentamers can bind back to the capsid openings after the genome release<sup>25</sup>. It is likely that the re-assembly of the empty particles was favored by the high (0.5 mg/mL) concentration of echovirus 18 particles in samples that were prepared for cryo-EM observations. Nevertheless, the fate of the empty capsids after genome delivery is unimportant for the infection process *in vivo*.

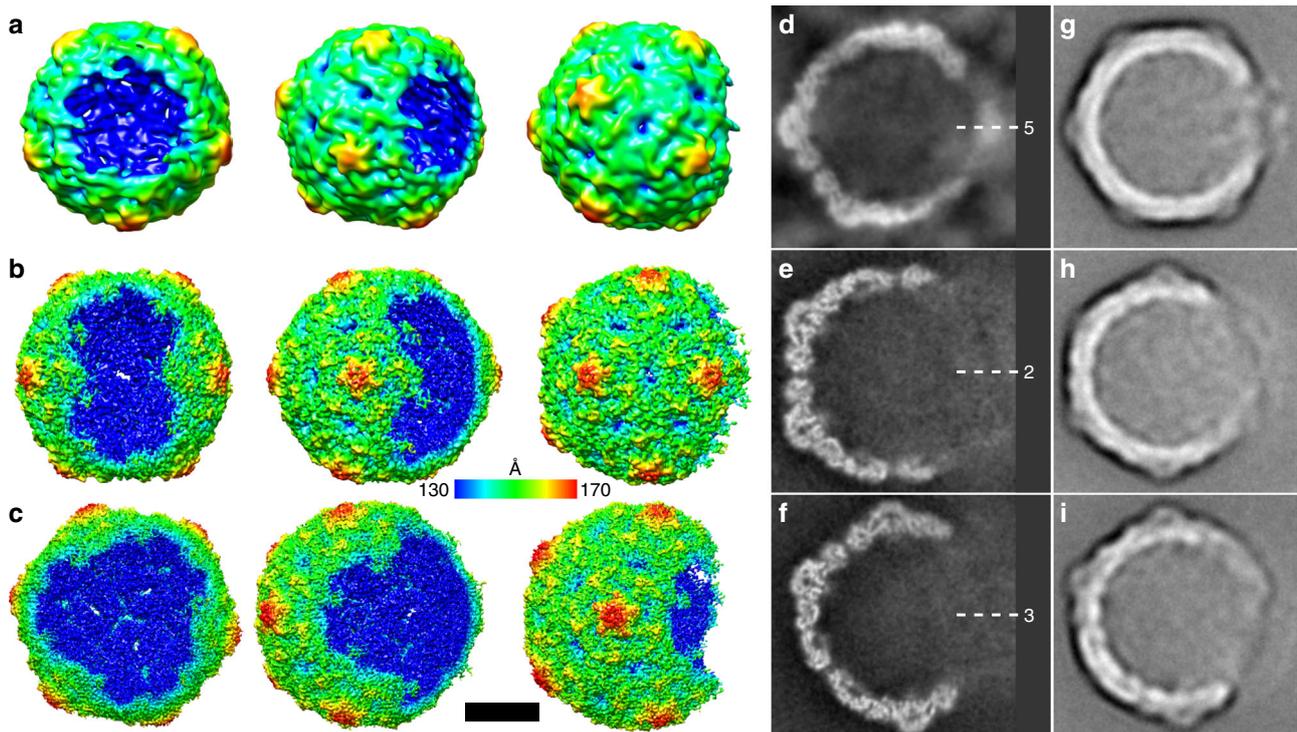
The stability of echovirus 18 empty capsids under the conditions promoting genome release provides evidence that the force for the expulsion of pentamers from the activated particles is provided by the RNA genome. Comparing the cryo-EM images of native virions with activated particles of echovirus 18 (Fig. 5), as well as other enteroviruses<sup>3,26,27</sup>, reveals that their genomic RNAs undergo conformational changes upon exposure to acidic pH. The conversion of echovirus 18 virions to activated particles occurred in <3 min after exposure to acidic pH at 37 °C, however some of the particles also released their RNA and aggregated (Supplementary Fig. 8). This rapid conversion to activated particles and genome release are consistent with previous experiments showing that human rhinovirus 2 delivers its genome into the cell cytoplasm within 2 min<sup>28</sup>. The electron density of the genomes is distributed uniformly in echovirus 18 virions at neutral pH, but transforms to a grainy appearance in activated particles at acidic pH (Fig. 5). During the shift from neutral to acidic pH, the side chains of histidines of capsid proteins (Supplementary Fig. 9), and probably also parts of the genomic RNA, become protonated, and thus acquire more positive charge. The reduction in the negative charge may disrupt interactions between the RNA and positively charged polyamines present within the virions<sup>17,29,30</sup>. The changes in the charge distribution and putative release of the polyamines from virions may result in the observed changes in the genome structure and increased pressure on the capsid from the inside. The inter-pentamer contacts of 11,000 Å<sup>2</sup> in native echovirus 18 virion are reduced to 5400 Å<sup>2</sup> in the activated particles. Correspondingly, the interaction between two pentamers in an activated particle of echovirus 18 is 25% weaker than that in a native virion (Supplementary Fig. 10). The weakening of the inter-pentamer contacts together with the changes in the genome organization, probably enable the opening of the activated particles for genome release.

### Molecular dynamics simulation of genome release.

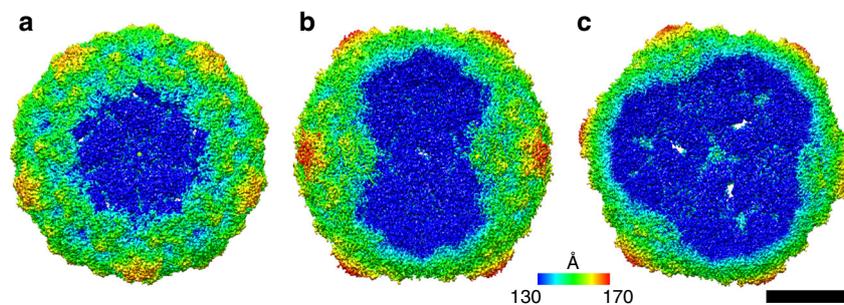
The dynamics of the capsid opening and genome release were investigated using *in silico* simulations of a simplified model of the picornavirus capsid with its genome (Fig. 4). Using certain parameters in the model (for details, see Methods), we observed genome release after which a pentamer was separated from a capsid (Fig. 4, Supplementary Movie 1). In the simulations, the capsid first cracked open to allow the release of part of the genome. During this process, one or a few pentamers separated from the rest of the capsid. Subsequently, the two fragments of the capsid closed, resulting in an open capsid missing one or a few pentamers (Fig. 4). The pressure from the inside of the particle required to crack open the capsid is two-thirds of that required to expel a pentamer (for details, see Methods). Consequently, enterovirus capsids are more likely to rupture into two halves than to expel a single pentamer. After capsid rupture the genome pressure decreases, so that the two halves of the capsid may not separate completely and can quickly reassemble. The escaping genome can break off some pentamers, as observed in our cryo-EM experiments and simulations.

### Open particles were also observed for other enteroviruses.

Open particles are also present in the samples of echovirus 30 (Fig. 6). Furthermore, Harutyunyan et al. detected particles lacking pentamers in a sample of human rhinovirus 2 with crosslinked RNA genomes exposed to acidic pH<sup>26</sup>. These observations indicate that enteroviruses may release their genomes through openings formed by the removal of pentamers of capsid protein protomers from their capsids. Numerous capsid-binding antiviral compounds inhibit the genome release of



**Fig. 2** Open particles of echovirus 18 lacking one, two, or three adjacent pentamers. **a–c** Asymmetric three-dimensional reconstructions of open particles lacking one (**a**), two (**b**), or three (**c**) pentamers. The electron density maps are rainbow colored based on the distance of the electron density surface from the particle center. **d–f** Electron density distributions in central sections of asymmetric reconstruction of open particles missing one (**d**), two (**e**), or three (**f**) pentamers. The directions of five-fold, two-fold, and three-fold symmetry axes are indicated. Diffuse density in the areas of the missing pentamers probably belong to the average of the RNA molecules escaping from the particles. Example reference-free two-dimensional class averages of final three-dimensional refinement with C1 symmetry of open particles missing one (**g**), two (**h**), or three (**i**) pentamers. Scale bar represents 10 nm



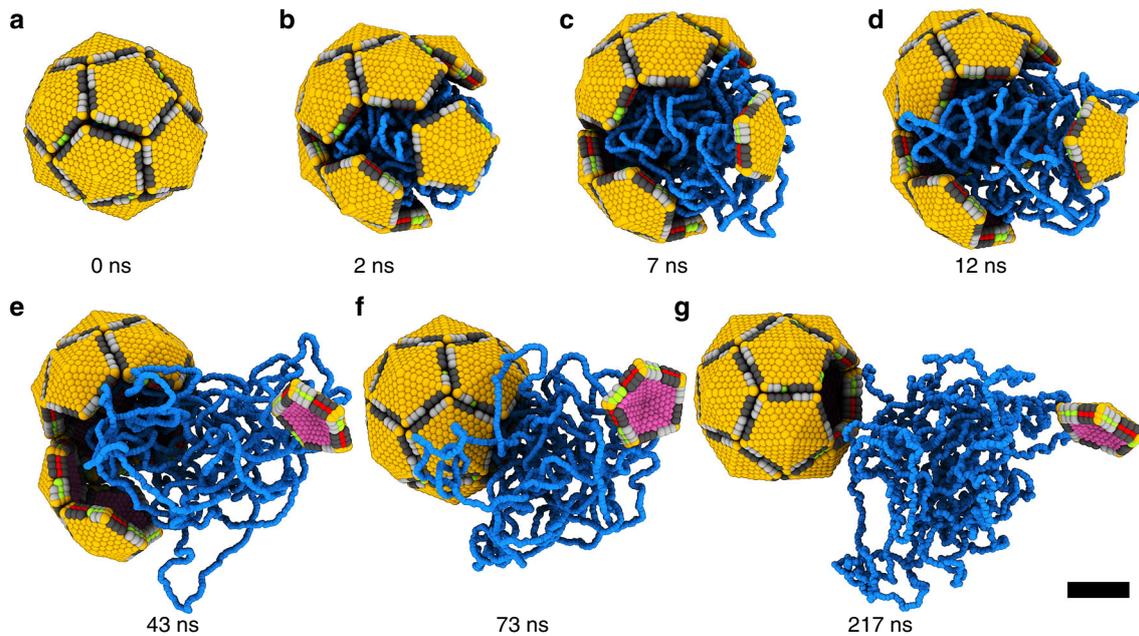
**Fig. 3** Symmetrized three-dimensional reconstructions of open particles. Echovirus 18 particles lacking one (**a**), two (**b**), or three (**c**) pentamers. Five-fold (**a**), two-fold (**b**), and three-fold (**c**) symmetries were employed during the reconstructions. The electron density maps are rainbow colored based on the distance of the electron density surface from the particle center. Scale bar represents 10 nm

enteroviruses<sup>31–34</sup> and an improved understanding of the process may facilitate drug development. In summary, our data show that enterovirus genome release requires several consecutive structural changes in a capsid (Fig. 7). Receptor binding or exposure to acidic pH induces the transformation of native virions into activated particles with reduced inter-pentamer interfaces<sup>1,3,7–9,14,15</sup>. The weakening of inter-pentamer contacts and changes in the genome structure enable opening of the capsids and release of the RNA genome.

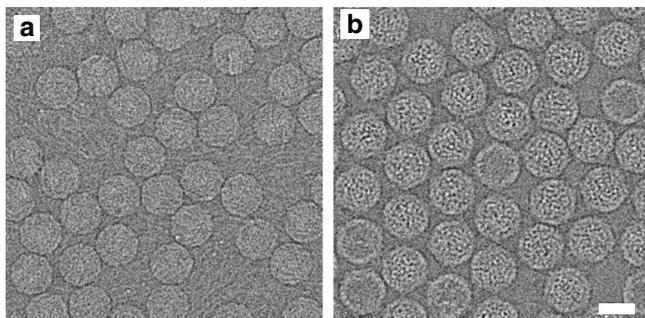
## Methods

**Production and purification of echovirus 18.** Echovirus 18 (strain METCALF, ATCC-VR-852<sup>TM</sup>) was propagated in immortalized African green monkey kidney (GMK, 84113001 Sigma) cells cultivated in Dulbecco's modified Eagle's medium

enriched with 10% fetal bovine serum. For virus preparation, 50 tissue culture dishes with a diameter of 150 mm of GMK cells grown to 100% confluence were infected with echovirus 18 with a multiplicity of infection of 0.01. The infection was allowed to proceed for 5–7 days, at which point more than 90% of the cells exhibited the cytopathic effect. The cell media were harvested and any remaining attached cells were removed from the dishes using cell scrapers. The cell suspension was centrifuged at 15,000 × *g* in a Beckman Coulter Allegra 25R centrifuge, rotor A-10 at 10 °C for 30 min. The resulting pellet was re-suspended in 10 mL of phosphate-buffered saline (PBS) (10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 137 mM NaCl, and 2.7 mM KCl, pH 7.4). The solution was subjected to three rounds of freeze-thawing by transfer between –80 and 37 °C, and homogenized using a Dounce tissue grinder. Cell debris was separated from the supernatant by centrifugation at 3100 × *g* in a Beckman Coulter Allegra 25R centrifuge, rotor A-10 at 10 °C for 30 min. The resulting supernatant was combined with media from the infected cells. Virus particles were precipitated by the addition of PEG-8000 and NaCl to final concentrations of 12.5% (w/v) and 0.6 M, respectively. The precipitation was allowed to proceed overnight at 10 °C and with mild shaking. The



**Fig. 4** Molecular dynamics simulation of echovirus 18 genome release. Seven snapshots from the process were selected. **a** Compact capsid just before opening. **b** Initial cracking of the particle with one pentamer separated from the rest of the capsid. **c–e** Continued release of the genome. **f** Re-assembly of the capsid missing one pentamer. **g** Remainder of the genome diffuses from the open capsid. Genome is shown in blue, outer capsid surface in orange, inner capsid surface in purple, beads at pentamer edges shown in dark and light gray, green, and red represent attractive inter-pentamer interfaces. Scale bar represents 10 nm



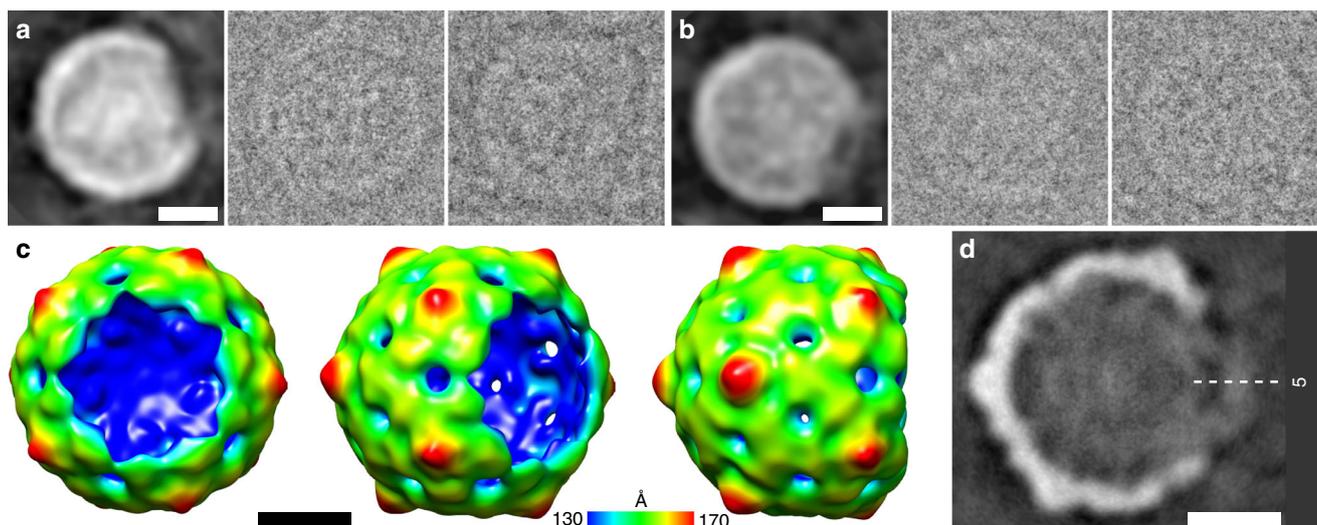
**Fig. 5** Exposure to acidic pH for 3 min at 4 °C induces conformational changes in echovirus 18 genome. **a** Electron micrograph of native echovirus 18 virions with uniformly distributed density of RNA genome. **b** Electron micrograph of echovirus 18 activated particles with grainy distribution of electron density belonging to RNA genome. Scale bar represents 25 nm

following day, the precipitate was centrifuged at 15,000 × g in a Beckman Coulter Allegra 25R centrifuge, rotor A-10 at 10 °C for 30 min. The pelleted white precipitate was re-suspended in 12 mL of PBS. MgCl<sub>2</sub> was added to a final concentration of 5 mM, and the sample was subjected to DNase (10 μg/mL final concentration) and RNase (10 μg/mL final concentration) treatment for 30 min at ambient temperature. Subsequently, trypsin was added to a final concentration of 0.5 μg/mL and the mixture was incubated at 37 °C for 10 min. EDTA at pH 9.5 was added to a final concentration of 15 mM and non-ionic detergent, NP-40™ (Sigma Aldrich Inc.), was added to a final concentration of 1%. The virus particles were pelleted through a 30% (w/v) sucrose cushion in re-suspension buffer (0.25 M HEPES, pH = 7.5, and 0.25 M NaCl) by centrifugation at 210,000 × g in an Optima X80 ultracentrifuge using a Beckman Coulter™ Ti50.2 rotor at 10 °C for 2 h. The pellet was re-suspended in 1.5 mL of PBS and loaded onto 60% (w/w) CsCl solution in PBS. The CsCl gradient was established by ultracentrifugation at 160,000 × g in an Optima X80 ultracentrifuge using a Beckman Coulter™ SW41Ti rotor at 10 °C for 18 h. The opaque bands containing the virus was extracted with a 20-gauge needle mounted on a 5 mL disposable syringe. The virus was transferred into 10 mM Tris, pH = 7.4, and 0.1 M NaCl by multiple rounds of buffer exchange using a centrifugal filter device with a 100-kDa molecular weight cutoff.

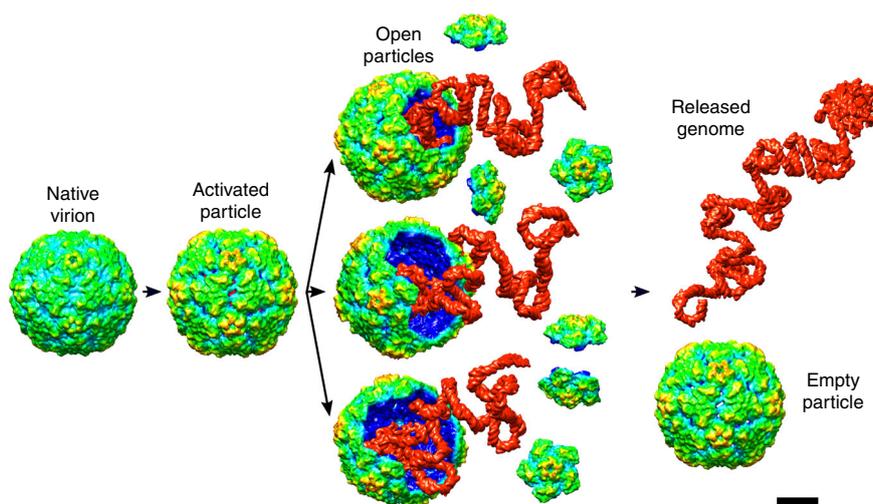
**Cryo-EM sample preparation and data acquisition.** For cryo-EM, 3.5 μL of echovirus 18 solution (2 mg/mL) were blotted and vitrified using a Vitrobot Mark IV on Quantifoil R2/1300 mesh holey carbon grids (vitrobot settings blot-force 2, blotting time 2 s). To observe echovirus 18 particles in the process of genome release, 5 μL of virus solution at a concentration of 2 mg/mL in 10 mM Tris, pH = 7.4, and 0.1 M NaCl was diluted in 15 μL of 50 mM Mes, pH 6.0. To obtain samples for single-particle analysis, the virus was incubated in acidic pH for 10 min at 4 °C. To measure the speed of formation of activated particles, the sample was incubated at 37 °C. Electron micrographs of virus particles were collected using an FEI Titan Krios transmission electron microscope operated at 300 kV. The sample in the column of the microscope was kept at −196 °C. Images were recorded with an FEI Falcon III direct electron detection camera under low-dose conditions (22.6 e<sup>−</sup>/Å<sup>2</sup>) with defocus values ranging from −1.0 to −3.0 μm at a nominal magnification of ×75,000, resulting in a pixel size of 1.061 Å/px. Each 1 s of exposure was recorded in movie mode and saved as 39 separate movie frames. The frames from each exposure were aligned to compensate for drift and beam-induced motion during image acquisition using the program motioncor2<sup>35</sup>. The resulting dose-weighted sum of aligned frames was used in the subsequent image processing steps, except for estimating contrast transfer function (CTF) parameters, which were determined from non-dose-weighted micrographs using the program gCTF<sup>36</sup>.

**Single-particle reconstructions of echovirus 18.** Particles of echovirus 18 (512 × 512 pixels) were automatically selected from micrographs by Gautomatch. The images were processed using the package RELION 2.1<sup>37</sup>. The dataset of autopicked particles of echovirus 18 was subjected to 2D classification. Classes containing full and empty particles were selected for processing in parallel. Classes containing full particles were 3D-classified to obtain sets of native particles and activated particles. Echovirus 7 [PDB ID: 2X51] served as the initial model for these classifications<sup>38</sup>. Similarly, classes containing empty particles were subjected to 3D classification to obtain a homogeneous set of empty particles. Finally, refinement of the selected particles was performed for native, activated particles and empty particles, using the RELION 3dautorefine procedure<sup>37</sup>. Icosahedral symmetry was imposed on the volumes during the refinement process.

**3D reconstructions of open particles.** Schemes of the image processing, classification and reconstruction of echovirus 18 and echovirus 30 are shown in Supplementary Fig. 2. Particles of echovirus 18 (512 × 512 pixels) were automatically selected by Gautomatch. Particles of echovirus 30 in the process of genome release (512 × 512 pixels) were manually boxed using the program e2boxer.py from EMAN2<sup>39</sup>. The images were processed using RELION 2.1<sup>37</sup>. The dataset of autopicked echovirus 18 particles was subjected to 2D classification. Classes containing particles releasing their genomes were selected for further 3D classification.



**Fig. 6** Open particle of echovirus 30. **a, b** Reference-free two-dimensional class averages showing echovirus 30 particles lacking parts of their capsids. For each class, average images of two representative particles are shown. **c** Three-dimensional reconstruction of open particle of echovirus 30 lacking one pentamer with imposed five-fold symmetry. The electron density map is rainbow colored based on the distance from the particle center. **d** Electron density distribution in central section of reconstruction of open particle. The diffuse density in the pore formed by the missing pentamer probably belongs to the average of RNA molecules escaping from the particles. Scale bars represent 10 nm



**Fig. 7** Scheme of enterovirus genome release. Binding to receptors or exposure to acidic pH in endosomes induces conformational transition of virions to activated particles. The structural changes within the capsid and virus RNA enable the expulsion of pentamers from the capsid, resulting in the formation of open particles. The RNA genomes are released from the open particles. After the genome release, the pentamers may re-associate with the open capsids. Scale bar represents 10 nm

The structure of the echovirus 18 or echovirus 30 “A” particle, low-pass filtered to a resolution of 50 Å, was used to initiate asymmetrical 3D classification of echovirus 18 and echovirus 30, respectively. Classes with particles exhibiting defects corresponding to missing one, two, or three pentamers were selected for further 3D classification. In these classification steps and in the final refinement step, the previous 3D class reconstructions served as the initial models. These reconstructions were rotated so that the symmetry axes of the open particles were aligned with the z-axis. The resulting 3D classes with homogenous particles missing one, two, or three pentamers were independently refined using the RELION 3dautorrefine procedure as asymmetric reconstructions or by imposing the appropriate symmetries.

**Post-processing of the refined maps.** The volumes resulting from the 3D refinement were threshold masked, detector modulation transfer function corrected and B-factor sharpened using the RELION postprocess procedure. To avoid over-masking, the masked maps were visually inspected to exclude the possibility

of the clipping of electron densities belonging to the virus capsid. Additionally, the occurrence of over-masking was monitored by inspecting the shapes of Fourier shell correlation (FSC) curves. Furthermore, the shapes of the FSC curves of phase-randomized half-datasets with the applied mask were checked. The resulting resolutions of the reconstructions were estimated as the values at which the FSC curves fell to 0.143.

**Structure determination of open particles of echovirus 18.** The initial model, the activated particle of echovirus 18, was rigid-body fitted into the B-factor-sharpened cryo-EM map of echovirus 18 open particles and subjected to manual rebuilding using Coot<sup>40</sup>, and coordinate and B-factor refinement using Phenix<sup>41</sup>.

**Charge calculation - Monte Carlo simulations.** We performed Metropolis Monte Carlo (MC) simulations using the Faunus framework<sup>42</sup>. The spherical cell with a radius of 45 nm contained one copy of the capsid described with an implicit-solvent coarse-grained model, where every residue was treated as a spherical bead

(located at the center of mass of the residue) with a radius derived from the amino-acid molecular weight and the common density of 0.9 g/mL. The N- and C termini of both proteins were represented as separate residues. The solvent was treated as a dielectric continuum using the Debye–Hückel approximation with a relative permittivity of 78.7 for the interaction of charged residues<sup>43,44</sup>. The capsid was placed in the middle of the simulation sphere with all degrees of motion frozen. Each amino acid was allowed to change its protonation state by titration move, where protons are allowed to move between the bead and solution. The energy associated with the exchange is determined by the change in local electrostatic energy  $\pm (pH - pK_0)\ln 10$ , where  $pK_0$  is the dissociation constant of the isolated amino acid, and  $pH$  is that of the system<sup>45</sup>. The plus and minus in the equation is associated with protonation and de-protonation, respectively. Titratable residues with their  $pK_a$  values are: C terminus (2.6), Asp (4.0), Glu (4.4), His (6.3), N-terminus (7.5), Tyr (9.6), Lys (10.4), Cys (10.8), and Arg (12.0). The total number of moves, in which there are attempts to protonate/deprotonate residues, was at least 1000 per each residue in all simulations. The temperature of our NVT ensemble was set to 298 K. We performed calculations of capsids of both native virions and activated particles with structures determined from cryo-EM. The average charges of amino acids were determined for  $pH = 7.4$  and 6.0, and monovalent salt solutions of concentrations 150 and 40 mM.

**Comparison of forces required to open a capsid.** The reason why the capsids are more likely to initially crack open rather than directly expel pentamers can be shown by a comparison of the forces holding the capsids together to the force exerted by the genome, which acts to rupture the capsid from the inside. Assuming that the inside pressure generated by the genome is homogeneous, the force acting on each pentamer is proportional to its area of  $1/12$  of the sphere surface:

$f_{\text{pentamer}} = \pi r^2 p/3$ , where  $r$  is the capsid radius and  $p$  is the excess pressure from inside to outside. Each pentamer in the capsid interacts with five others, generating a force of 5 F that holds the pentamer in the capsid. Therefore, the pressure of the genome required to break a pentamer away is  $p = 15 F/(\pi r^2)$ . However, the pressure to separate two halves of a capsid is only  $p = 10 F/(\pi r^2)$ , because two half-capsids interact with each other through 10 inter-pentamer interfaces, resulting in a holding force of 10 F and the pressure force exerted by the genome is  $f_{\text{half}} = \pi r^2 p$  due to the half-capsid projection in the direction of the force.

**Molecular dynamics simulations.** All-atom molecular dynamics simulations were performed using GROMACS version 2016.4<sup>46,47</sup>. The initial structure from cryo-EM was minimized and equilibrated for 10 ns using the all-atom Amber99SB-ILDN force field<sup>48</sup>. During the all-atom equilibration, the system was kept in an isothermal-isobaric ensemble using position restraints on backbone atoms. The spring constant of the position restraints was 1000 kJ/mol/nm<sup>2</sup>. Temperature was held at 300 K with a velocity-rescaling thermostat<sup>49</sup>. The time constant for temperature coupling was set to 0.1 ps. We employed an isotropic Berendsen barostat<sup>50</sup> set to 1 bar using a 1 ps coupling constant. Lennard-Jones forces were calculated with a cutoff radius of 1.1 nm. The same cutoff was applied for real-space electrostatic interactions, while the long-range contribution was evaluated with particle mesh Ewald<sup>51</sup>. All bonds were constrained using the LINCS algorithm<sup>52</sup>, apart from TIP3P water, for which analytical SETTLE was used<sup>53</sup>. A 2-fs time step was used for the production run. The system consisted of four echovirus 18 protomers in aqueous solution (of roughly 200,000 water molecules) with added NaCl ions at a concentration of 150 mM. We simulated capsids of both native virions and activated particles in a rectangular box of  $16 \times 16 \times 21$  and  $17 \times 17 \times 25$  nm, respectively. Periodic boundary conditions were applied.

For the binding free energy calculations between two pentamers, we used a computationally efficient, coarse-grained MARTINI 2.2 force field<sup>54–56</sup>. The resulting structure from all-atom equilibration was converted into a MARTINI model using the *martinize.py* script. As a consequence of coarse-graining, the MARTINI model does not explicitly describe backbone hydrogen bonds. Thus, the secondary structure has to be imposed on the peptides and maintained throughout the simulation. The assignment of secondary structure for both native virion and activated particle echovirus 18 was done with the program DSSP<sup>57</sup>. To help preserve the higher-order structure, an elastic network was added to the standard MARTINI topology. Harmonic bonds were generated between backbone beads by the *martinize.py* script using the option *-elastic*. The elastic bond force constant was set to 500 kJ/mol/nm<sup>2</sup> (*-ef 500*), and the lower and upper elastic bond cutoff radius to 0.5 and 0.9 nm (*-el 0.5* and *-eu 0.9*), respectively, and the elastic bond decay factor and decay power both to 0 (*-ea 0* and *-ep 0*). Furthermore, elastic bonds were deleted for residues exhibiting a high degree of flexibility in the electron density map (Supplementary Table 3). Additionally, the mapping of histidines with an average charge over 0.4 e, determined from MC simulations, were changed from the uncharged to charged form. The simulation time step was set to 30 fs. A velocity-rescaling thermostat with a coupling constant of 1.0 ps was employed to maintain the temperature at 310 K<sup>49</sup>. Protein and solvent beads were coupled to separate baths to ensure the correct temperature distribution. The pressure was kept at 1 bar with a Parrinello-Rahman barostat with an isotropic coupling scheme with a coupling constant of 12 ps<sup>58</sup>. All non-bonded interactions were cutoff at 1.1 nm and the van der Waals potential was shifted to zero. The relative dielectric constant was set to 15. Periodic boundary conditions were employed, yielding a rectangular box of dimensions  $17.7 \times 17.7 \times 28.5$  nm for the activated particle and

$17.7 \times 17.7 \times 31.4$  nm for the native virion. The System consisted of 4 protomers of capsid proteins of echovirus 18 in water with added NaCl ions at a concentration of 150 mM.

The umbrella sampling method was employed to determine the free energy of binding between two pentamers. We defined the reaction coordinate as the  $z$ -distance between the centers of mass of protomers 1–2 and 3–4 (Supplementary Fig. 11). We restrained the position of protomers 1–2 through the use of harmonic potentials on backbone beads, excluding the flexible residues from Supplementary Table 3. Cylindrical flat-bottomed position restraints were applied to the backbone beads of protomers 3–4, excluding the residues from Supplementary Table 3, to keep the protomers from tilting and moving in the  $XY$  plane. The cylinders were parallel to the  $z$ -axis. The force constant was set to 1000 kJ/mol/nm<sup>2</sup> and the radius of all cylinders was 0.3 nm. The reference configuration for the cylindrical flat-bottomed position restraint was selected from a 1000-ns equilibration run. For the last 500 ns of the equilibration run, the structure of protomers 3–4 was averaged. The reference configuration was selected from the trajectory based on the lowest root mean squared deviation toward the averaged structure. For the native virion, 74 umbrella windows were simulated for 2000 ns each, which was necessary to get a convergence (Supplementary Fig. 12). The spring constant of the umbrella harmonic potential was set to 50,000 kJ/mol/nm<sup>2</sup> for the first 40 windows, with a spacing of 0.02 nm. The next 34 windows were spaced by 0.025 nm and the spring constant was set to 10,000 kJ/mol/nm<sup>2</sup>. For the activated particle of enterovirus 18, 149 umbrella windows were simulated for 1000 ns each. The first 34 windows were spaced by 0.02 nm and a harmonic spring of 50,000 kJ/mol/nm<sup>2</sup> was applied. The next 115 windows had the spring constant set to 10,000 kJ/mol/nm<sup>2</sup> with a spacing of 0.025 nm. To analyze the probability distributions of states from each window, iterative WHAM was used, implemented in the GROMACS tool *gmx wham*<sup>59,60</sup>.

**Phenomenological model.** We developed a phenomenological coarse-grained model of a *Picornavirus* family based on human echovirus 18 (Supplementary Fig. 13). The capsid was a regular dodecahedron comprised of 12 pentagonal subunits. Each subunit assumes the role of a stable pentameric intermediate. The pentamers were made of beads (pseudoatoms) organized in three layers.

The outer circumscribed sphere of the capsid had a radius of 16 nm and the inner was 14 nm. Each pentamer was composed of 317 beads connected with 1311 harmonic bonds keeping the structure. The spring constant of the harmonic potential was 250 kJ/mol. Beads within the capsids were only interacting via harmonic bonds. There were six types of beads (Supplementary Fig. 14), all of which were interacting with Weeks-Chandler-Anderson repulsive potential with an epsilon set to 1.0 [<https://doi.org/10.1063/1.1674820>]. In addition, beads at the pentamer edges had an attractive interaction range of 0.3 to 2.0 nm based on the free energy calculation with the Martini model. The interaction decreased to zero with a  $\cos^2$  dependence. The attraction strength was weak, 0.5 kJ/mol, for the inner and outer layer, while the middle layer had stronger attraction strength that varied from 2 to 20 kJ/mol. The attraction only acts between the types, which are in contact in the assembled capsid structure. These interactions represent specific contacts between the protomers in the capsid (Supplementary Fig. 11).

To investigate RNA genome release from the capsid, we modeled the RNA as a chain of repulsive beads. A single bead represented two nucleotides with a radius of 0.6 nm connected by a 1.1-nm-long harmonic bond, which is about twice the distance between phosphates of adjacent nucleotides. All beads were interacting with a shifted truncated Lennard-Jones potential, i.e. Weeks-Chandler-Anderson potential with an epsilon set to 1.0<sup>61</sup>.

Simulations were performed in LAMMPS<sup>62</sup>, with the use of a Langevin thermostat<sup>63–65</sup>. Center of mass motion of the entire system caused by the thermostat was eliminated using the option “zero yes”. Additionally, the “gif yes” option was turned on, applying Gronbech-Jensen/Farago time-discretization for the Langevin model to enable longer time steps, while still producing the correct Boltzmann distribution of atom positions<sup>65</sup>. The viscous damping term was set to 10,000 time steps. The reduced temperature in our simulations was  $T^* = 1 k_B T$ . The box size  $150 \times 150 \times 150$  nm was constant and the same for all simulations.

The simulation protocol was as follows. First the capsid was generated from pentamers, and then the chain representing the genome was generated within the capsid using a random walk. The equilibration started with the chain equilibration alone, which was simulated with Langevin dynamics for  $10^8$  steps, while the capsid shell was motionless. The second step was the equilibration of both the capsid and genome. The attraction between pair B-C begun at 35 kJ/mol and was gradually decreased by a rate of 0.25 kJ/mol every 200,000 time steps. The production genome release simulations were repeated 10 times, with different conditions, for each set of parameters (attraction strength, attraction range, flexibility of the capsid, and the size of the genome beads) each of which were concluded at  $10^5$  time steps or sooner if the release had occurred.

To estimate the timescale of the simulations, we performed 50 independent simulations of the full virus capsid starting from different conditions. Each simulation was performed for  $10^8$  steps. The parameters for the capsid were chosen to prevent the genome release within the duration of the simulation. We analyzed the averaged square displacement of the full capsid center of mass. Using Fick’s second law, we calculated the diffusion coefficient of the capsid to be  $1.37 \times 10^{-4} \pm 2.0 \times 10^{-8} \text{Å}^2/\text{step}$  (Supplementary Fig. 15). We also estimated the diffusion coefficient using the Stokes-Einstein relation, with a capsid radius of 16.5 nm, the

diffusion coefficient is  $15 \mu\text{m}^2/\text{s}$ . Comparing the estimated coefficient with the simulation one, the simulation time step corresponded to 92 fs (there were 10,901  $\pm 1$  steps in a ns).

**Timescale of genome release.** To estimate the timescale of genome release from the virus capsid, we analyzed the average times of this process in our simulations. The genome was released from the capsid in the order of  $10^6$  simulation steps corresponding to 100 ns. Another approach is to estimate the time required for the genome to diffuse freely from an open capsid. Using Einstein-Stokes relation, we determined that the genome diffuses 50 nm in microsecond timescale, which makes the observation of the genome release experimentally challenging.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Cryo-EM electron density maps have been deposited in the Electron Microscopy Data Bank, <https://www.ebi.ac.uk/pdbe/emdb/> (accession numbers EMD-0181, EMD-0182, EMD-0183, EMD-0184, EMD-0185, EMD-0186, EMD-0187, EMD-0188, EMD-0189, and EMD-0217), and the fitted coordinates have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID codes 6HBG, 6HBH, 6HBJ, 6HBK, 6HBL, and 6HHT). The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information files, or are available from the authors upon request.

Received: 30 August 2018 Accepted: 17 February 2019

Published online: 08 March 2019

### References

- Tuthill, T. J., Groppelli, E., Hogle, J. M. & Rowlands, D. J. Picornaviruses. *Curr. Top. Microbiol. Immunol.* **343**, 43–89 (2010).
- Ren, J. et al. Picornavirus uncoating intermediate captured in atomic detail. *Nat. Commun.* **4**, 1929 (2013).
- Levy, H. C., Bostina, M., Filman, D. J. & Hogle, J. M. Catching a virus in the act of RNA release: a novel poliovirus uncoating intermediate characterized by cryo-electron microscopy. *J. Virol.* **84**, 4426–4441 (2010).
- Garriga, D. et al. Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. *PLoS Pathog.* **8**, e1002473 (2012).
- Wang, X. et al. A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nat. Struct. Mol. Biol.* **19**, 424–429 (2012).
- Shingler, K. L. et al. The enterovirus 71 A-particle forms a gateway to allow genome release: a cryoEM study of picornavirus uncoating. *PLoS Pathog.* **9**, e1003240 (2013).
- Seitonen, J. J. et al. Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *J. Virol.* **86**, 7207–7215 (2012).
- Bostina, M., Levy, H., Filman, D. J. & Hogle, J. M. Poliovirus RNA is released from the capsid near a twofold symmetry axis. *J. Virol.* **85**, 776–783 (2011).
- Belnap, D. M. et al. Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. *J. Virol.* **74**, 1342–1354 (2000).
- Hadfield, A. T. et al. The refined structure of human rhinovirus 16 at 2.15 Å resolution: implications for the viral life cycle. *Structure* **5**, 427–441 (1997).
- Reinisch, K. M., Nibert, M. L. & Harrison, S. C. Structure of the reovirus core at 3.6 Å resolution. *Nature* **404**, 960–967 (2000).
- Naitow, H., Tang, J., Canady, M., Wickner, R. B. & Johnson, J. E. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nat. Struct. Biol.* **9**, 725–728 (2002).
- Rich, A. & Davies, D. R. A new two stranded helical structure: polyadenylic acid and polyridylic acid. *J. Am. Chem. Soc.* **78**, 3548–3549 (1956).
- Hewat, E. A., Neumann, E. & Blaas, D. The concerted conformational changes during human rhinovirus 2 uncoating. *Mol. Cell* **10**, 317–326 (2002).
- Hewat, E. A. & Blaas, D. Cryoelectron microscopy analysis of the structural changes associated with human rhinovirus type 14 uncoating. *J. Virol.* **78**, 2935–2942 (2004).
- Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).
- Jiang, P., Liu, Y., Ma, H. C., Paul, A. V. & Wimmer, E. Picornavirus morphogenesis. *Microbiol. Mol. Biol. Rev.* **78**, 418–437 (2014).
- Fricks, C. E. & Hogle, J. M. Cell-induced conformational change in poliovirus: externalization of the amino terminus of VP1 is responsible for liposome binding. *J. Virol.* **64**, 1934–1945 (1990).
- Tuthill, T. J., Bubeck, D., Rowlands, D. J. & Hogle, J. M. Characterization of early steps in the poliovirus infection process: receptor-decorated liposomes induce conversion of the virus to membrane-anchored entry-intermediate particles. *J. Virol.* **80**, 172–180 (2006).
- Groppelli, E. et al. Picornavirus RNA is protected from cleavage by ribonuclease during virion uncoating and transfer across cellular and model membranes. *PLoS Pathog.* **13**, e1006197 (2017).
- Chow, M. et al. Myristylation of picornavirus capsid protein VP4 and its structural significance. *Nature* **327**, 482–486 (1987).
- Kirkegaard, K. Mutations in VP1 of poliovirus specifically affect both encapsidation and release of viral RNA. *J. Virol.* **64**, 195–206 (1990).
- Krausslich, H. G., Holscher, C., Reuer, Q., Harber, J. & Wimmer, E. Myristoylation of the poliovirus polyprotein is required for proteolytic processing of the capsid and for viral infectivity. *J. Virol.* **64**, 2433–2436 (1990).
- Moscufo, N., Yafal, A. G., Rogove, A., Hogle, J. & Chow, M. A mutation in VP4 defines a new step in the late stages of cell entry by poliovirus. *J. Virol.* **67**, 5075–5078 (1993).
- Rapaport, D. C. Role of reversibility in viral capsid growth: a paradigm for self-assembly. *Phys. Rev. Lett.* **101**, 186101 (2008).
- Harutyunyan, S. et al. Viral uncoating is directional: exit of the genomic RNA in a common cold virus starts with the poly-(A) tail at the 3'-end. *PLoS Pathog.* **9**, e1003270 (2013).
- Pickl-Herk, A. et al. Uncoating of common cold virus is preceded by RNA switching as determined by X-ray and cryo-EM analyses of the subviral A-particle. *Proc. Natl Acad. Sci. USA* **110**, 20063–20068 (2013).
- Harutyunyan, S., Kowalski, H. & Blaas, D. The Rhinovirus subviral a-particle exposes 3'-terminal sequences of its genomic RNA. *J. Virol.* **88**, 6307–6317 (2014).
- Fout, G. S., Medappa, K. C., Mapoles, J. E. & Rueckert, R. R. Radiochemical determination of polyamines in poliovirus and human rhinovirus 14. *J. Biol. Chem.* **259**, 3639–3643 (1984).
- Mounce, B. C., Olsen, M. E., Vignuzzi, M. & Connor, J. H. Polyamines and their role in virus infection. *Microbiol Mol Biol Rev* **81**, e00029-17 (2017).
- Rossmann, M. G. Viral cell recognition and entry. *Protein Sci.* **3**, 1712–1725 (1994).
- Grant, R. A. et al. Structures of poliovirus complexes with anti-viral drugs: implications for viral stability and drug design. *Curr. Biol.* **4**, 784–797 (1994).
- Smith, T. J. et al. The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* **233**, 1286–1293 (1986).
- Shih, S. R. et al. Selective human enterovirus and rhinovirus inhibitors: An overview of capsid-binding and protease-inhibiting molecules. *Med. Res. Rev.* **24**, 449–474 (2004).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Plevka, P. et al. Interaction of decay-accelerating factor with echovirus 7. *J. Virol.* **84**, 12665–12674 (2010).
- Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Stenqvist, B., Thuresson, A., Kurut, A., Vácha, R. & Lund, M. Faunus—a flexible framework for Monte Carlo simulation. *Mol. Simul.* **39**, 1233–1239 (2013).
- Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic Press, 2011) <https://www.elsevier.com/books/intermolecular-and-surface-forces/israelachvili/978-0-12-391927-4>.
- Bagotsky, V. S. *Fundamentals of Electrochemistry* (John Wiley & Sons, Hoboken, 2005).
- Lund, M., Åkesson, T. & Jönsson, B. Enhanced protein adsorption due to charge regulation. *Langmuir* **21**, 8385–8388 (2005).
- Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In *Solving Software Challenges for Exascale*, Vol. 8759 (eds Markidis, S. & Laure, E.) 3–27 (Springer, Switzerland London, 2015).
- Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
- Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
- Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

50. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
51. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
52. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
53. Miyamoto, S. & Kollman, P. A. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
54. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
55. Monticelli, L. et al. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).
56. de Jong, D. H. et al. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).
57. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
58. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
59. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
60. Zhang, C., Lai, C. L. & Pettitt, B. M. Accelerating the weighted histogram analysis method by direct inversion in the iterative subspace. *Mol. Simul.* **42**, 1079–1089 (2016).
61. Weeks, J. D., Chandler, D. & Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **54**, 5237–5247 (1971).
62. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
63. DÜNweg, B. & Paul, W. Brownian dynamics simulations without Gaussian random numbers. *Int. J. Mod. Phys. C* **02**, 817–827 (1991).
64. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17**, 1302–1322 (1978).
65. Grønbech-Jensen, N., Hayre, N. R. & Farago, O. Application of the G-JF discrete-time thermostat for fast and accurate molecular simulations. *Comput. Phys. Commun.* **185**, 524–527 (2014).
66. Stenqvist, B., Thuresson, A., Kurut, A., Vácha, R. & Lund, M. Faunus—a flexible framework for Monte Carlo simulation. *Mol. Simul.* **39**, 1233–1239 (2013).

## Acknowledgements

We wish to thank the Central European Institute of Technology Core Facilities of Cryo-Electron Microscopy and Tomography and of Proteomics, supported by the Czech Infrastructure for Integrative Structural Biology (LM2015043 funded by the Ministry of Education, Youth and Sports of the Czech Republic) for their assistance in obtaining the scientific data presented in this paper. This research was carried out under the project CEITEC 2020

(LQ1601), with financial support from the Ministry of Education, Youth and Sports of the Czech Republic under National Sustainability Program II. This work was supported by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Super-computing Center – LM2015070". Computational resources were provided by CESNET LM2015042, and the CERIT Scientific Cloud LM2015085 provided under the program Projects of Large Research, Development, and Innovations Infrastructures of The Ministry of Education, Youth and Sports. The Titan Xp used for this research was donated by the NVIDIA Corporation. The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Program Grant (FP7/2007-2013)/ERC Grant Agreement 335855 (to PP), Czech Science Foundation Grant Agreement GX19-25982X (to P.P.), and EMBO Grant Agreement IG 3041 (to P.P.).

## Author contributions

P.P. and R.V. designed research; D.B., T.F., D.H., Y.L., L.S., L.M. and J.M. performed research; D.B., T.F., D.H., L.S., R.V. and P.P. analyzed data; and D.B., T.F., L.S., R.V. and P.P. wrote the paper.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-09132-x>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks David Rowlands and the other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019





# Summary

Taking advantage of the 'resolution-revolution' and rapid development of cryo-EM methods, which happened just before and during my Ph.D. studies, I was able to describe molecular mechanisms of viral genome release *in vitro* in prokaryotic and eukaryotic viruses. Bacteriophage P68 was among the first bacteriophages whose whole native structure achieved resolution sufficient enough to describe it in molecular detail. Using the localized reconstruction of subparticles, we established a protocol that solves the symmetry mismatch problem between bacteriophage capsid and tail. I was able to reconstruct the intermediate and final stages of the phage genome release, which were present in the sample. Then, by comparison of the different 3D structures, I proposed a mechanism that controls the genome release and mechanism of phage penetration into the host cell.

The structure of the rhinovirus 14-ICAM-1 complex has revealed a molecular switch inside of the virus that is triggered by the binding of ICAM-1. We were able to describe the conformational changes of the capsid and genome necessary for the virus to uncoat. Together with the discovery of enterovirus 30 particles expelling pentamers, it gives a picture about molecular mechanisms behind genome release in the genus of *Enterovirus*. However, it is possible that there is more than one common mechanism of genome release in *Enteroviruses*.

All in all, the above-mentioned articles provide new insights into the early stages of infection in both prokaryotic and eukaryotic viruses.



## Quo vadis cryo-EM?

Since the early days of structural biology, the studied molecules must have been prepared in high amounts and in extreme purity. That was possible to achieve for some protein structures of which were solved by X-ray crystallography. However, many important structural targets were impossible to study due to their instability, complexity, or they did not produce crystals of sufficient quality.

The 'resolution revolution' in cryo-EM opened a gate for many previously intractable molecules, now being solved almost routinely. That is mostly because cryo-EM does not require such high amounts of sample, and no crystallization is required. Instead, single molecules are imaged, averaged, and eventually, 3D reconstructed. Besides the rapid progress, one still needs to isolate and amplify the molecule of interest to solve its 3D structure.

Only recently, people started to solve structures of protein complexes and viruses enriched directly from the endogenous cellular milieu and even from native vitrified cells. The latter is a natural extension of cryo-electron microscopy. The development of a new generation of detectors, energy filters, and more precise sample stages will allow to perform very fast and precise tomography on whole cells, which allow structure determination in the cellular context. However, solving structures of proteins with low copy numbers, such as membrane proteins and signal molecules, is still a challenge. Development of hybrid methods, such as super-resolution correlated light and electron microscopy, mass spectrometry, and genetics, would be necessary to precisely target such proteins. Simultaneous development of software is important as well. For example, finding molecules of interest in noisy cryo-electron tomograms of whole cells is still a challenging task. New deep-learning algorithms working in combination with alpha-fold structure prediction will be necessary to tackle this challenge. Additionally, software predicting the flexibility of protein complexes combined with sensitive 3D classification will provide additional information to the structures. The rate of developments in cryo-EM is astonishing. Things that were only several years ago considered as 'visionary' are now a reality. Therefore, I am very optimistic about the future of cryo-EM and the community as a whole. I am looking forward to what surprises the future holds and if any of the above-mentioned visions become true.



# Bibliography

- [1] Thomas A. Steitz. A structural understanding of the dynamic ribosome machine. *Nature Reviews Molecular Cell Biology*, 9(3):242–253, Mar 2008.
- [2] Nobel prize 2009. <https://www.nobelprize.org/prizes/chemistry/2009/summary/>. Accessed: 2021-12-12.
- [3] John J. Hopfield. *Feynman and Computation*, page 3–6. Perseus Books, USA, 1999.
- [4] Ray Egerton. *Physical principles of electron microscopy: An introduction to TEM, SEM, and AEM*, volume 1. Springer, Boston, MA, 2005.
- [5] Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor P. C. Drummen. Super-resolution microscopy demystified. *Nature Cell Biology*, 21(1):72–84, Jan 2019.
- [6] James Pawley. *Handbook of biological confocal microscopy*, volume 236. Springer Science & Business Media, 2006.
- [7] Bartosz Turkowyd, David Virant, and Ulrike Endesfelder. From single molecules to life: microscopy at the nanoscale. *Analytical and Bioanalytical Chemistry*, 408(25):6885–6911, Oct 2016.
- [8] Pavan M. V. Raja and Andrew R. Barron. Transmission Electron Microscopy, dec 21 2021. [Online; accessed 2021-12-03].
- [9] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, Nov 2020.
- [10] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M. G. E. Brown, Ioana T. Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, Tomasz Uchański, Lingbo Yu, Dimple Karia, Evgeniya V. Pechnikova, Erwin de Jong, Jeroen Keizer, Maarten Bischoff, Jamie McCormack, Peter Tiemeijer, Steven W. Hardwick, Dimitri Y. Chirgadze, Garib Murshudov, A. Radu Aricescu, and Sjors H. W. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, Nov 2020.

- [11] Linda E. Franken, Kay Grünewald, Egbert J. Boekema, and Marc C. A. Stuart. A technical introduction to transmission electron microscopy for soft-matter: Imaging, possibilities, choices, and technical developments. *Small*, 16(14):1906198, 2020.
- [12] Richard Henderson. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*, 28(2):171–193, 1995.
- [13] Mikhail Kudryashev, Daniel Castaño-Díez, and Henning Stahlberg. Limiting factors in single particle cryo electron tomography. *Computational and Structural Biotechnology Journal*, 1(2):e201207002, 2012.
- [14] Martin Obermair, Simon Hettler, Chyongere Hsieh, Manuel Dries, Michael Marko, and Dagmar Gerthsen. Analyzing contrast in cryo-transmission electron microscopy: Comparison of electrostatic and hole-free phase plates. *Ultramicroscopy*, 218:113086, 2020.
- [15] Fred J. Sigworth. Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1):57–67, 12 2015.
- [16] Arthur C Clarke. *Profiles of the Future*. Hachette UK, 2013.
- [17] Lindsay A Baker and John L Rubinstein. Chapter fifteen - radiation damage in electron cryomicroscopy. In Grant J. Jensen, editor, *Cryo-EM Part A Sample Preparation and Data Collection*, volume 481 of *Methods in Enzymology*, pages 371–388. Academic Press, 2010.
- [18] David Bhella. Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophysical Reviews*, 11(4):515–519, Aug 2019.
- [19] Axel F. Brilot, James Z. Chen, Anchi Cheng, Junhua Pan, Stephen C. Harrison, Clinton S. Potter, Bridget Carragher, Richard Henderson, and Nikolaus Grigorieff. Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*, 177(3):630–637, 2012.
- [20] Benjamin E. Bammes, Ryan H. Rochat, Joanita Jakana, Dong-Hua Chen, and Wah Chiu. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 nyquist frequency. *Journal of Structural Biology*, 177(3):589–601, 2012.
- [21] Timothy Grant and Nikolaus Grigorieff. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife*, 4:e06980, may 2015.
- [22] Shawn Q. Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332, Apr 2017.

- [23] David Střelák, Jiří Filipovič, Amaya Jiménez-Moreno, Jose María Carazo, and Carlos Óscar Sánchez Sorzano. FlexAlign: An accurate and fast algorithm for movie alignment in cryo-electron microscopy. *Electronics*, 9(6), 2020.
- [24] Ali Punjani. Algorithmic advances in single particle cryo-EM data processing using cryoSPARC. *Microscopy and Microanalysis*, 26(S2):2322–2323, Aug Aug 2020. Copyright - Copyright © Microscopy Society of America 2020.
- [25] Dimitry Tegunov and Patrick Cramer. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods*, 16, 11 2019.
- [26] Pawel A. Penczek, Jia Fang, Xueming Li, Yifan Cheng, Justus Loerke, and Christian M.T. Spahn. CTER—rapid estimation of CTF parameters with error assessment. *Ultramicroscopy*, 140:9–19, 2014.
- [27] Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015. Recent Advances in Detector Technologies and Applications for Molecular TEM.
- [28] Kai Zhang. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1):1–12, 2016.
- [29] Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007. Software tools for macromolecular microscopy.
- [30] Jasenko Zivanov, Takanori Nakane, Björn O Forsberg, Dari Kimanius, Wim JH Hagen, Erik Lindahl, and Sjors HW Scheres. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, 7:e42166, nov 2018.
- [31] James M. Bell, Muyuan Chen, Tunay Durmaz, Adam C. Fluty, and Steven J. Ludtke. New software tools in EMAN2 inspired by EMDatabank map challenge. *Journal of Structural Biology*, 204(2):283–290, 2018.
- [32] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, Dennis Quentin, Daniel Roderer, Sebastian Tacke, Birte Siebolds, Evelyn Schubert, Tanvir R. Shaikh, Pascal Lill, Christos Gatsogiannis, and Stefan Raunser. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology*, 2(1):218, Jun 2019.
- [33] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1153–1160, Nov 2019.

- [34] Pawel A. Penczek, Marek Kimmel, and Christian M.T. Spahn. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure*, 19(11):1582–1590, 2011.
- [35] Pawel A. Penczek, Robert A. Grassucci, and Joachim Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3d cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251–270, 1994.
- [36] Vahid Abrishami, Serban L. Ilca, Josue Gomez-Blanco, Ilona Rissanen, José Miguel de la Rosa-Trevín, Vijay S. Reddy, José-Maria Carazo, and Juha T. Huiskonen. Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. *Progress in Biophysics and Molecular Biology*, 160:43–52, 2021. Cryo-EM microscopy developments and their biological applications.
- [37] Serban L. Ilca, Abhay Kotecha, Xiaoyu Sun, Minna M. Poranen, David I. Stuart, and Juha T. Huiskonen. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature Communications*, 6(1):8843, Nov 2015.
- [38] Dominik Hřebík, Dana Štveráková, Karel Škubník, Tibor Füzik, Roman Pantůček, and Pavel Plevka. Structure and genome ejection mechanism of Staphylococcus aureus phage P68. *Science Advances*, 5(10):eaaw7414, 2019.
- [39] Jan Drenth. *Theory of X-ray Diffraction by a Crystal*, pages 64–108. Principles of Protein X-Ray Crystallography. Springer New York, New York, NY, 2007.
- [40] Monaco Hugo L. and Gilberto Artioli. *Experimental methods in X-ray and neutron crystallography*. Fundamentals of Crystallography. Oxford University Press, Oxford, 3 edition, 2011.
- [41] Carmelo Giacovazzo. *The diffraction of X-rays by crystals*. Fundamentals of Crystallography. Oxford University Press, Oxford, 3 edition, 2011.
- [42] William H. Bragg and William L. Bragg. The reflection of x-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 88(605):428–438, 1913.
- [43] D. W. Green, V. M. Ingram, and M. F. Perutz. The structure of haemoglobin. iv. sign determination by the isomorphous replacement method. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 225(1162):287–307, 1954.
- [44] Wayne A. Hendrickson. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, 254(5028):51–58, 1991.

- [45] M. G. Rossmann. The molecular replacement method. *Acta Crystallographica Section A*, 46(2):73–82, Feb 1990.
- [46] Bragg’s law. [https://www.xtal.iqfr.csic.es/Cristalografia/parte\\_05\\_5-en.html](https://www.xtal.iqfr.csic.es/Cristalografia/parte_05_5-en.html). Accessed: 2021-12-15.
- [47] Microbiology by numbers. *Nature Reviews Microbiology*, 9(9):628–628, Sep 2011.
- [48] Amanda Carroll-Portillo and Henry C. Lin. Bacteriophage and the innate immune system: Access and signaling. *Microorganisms*, 7(12):625, Nov 2019.
- [49] Pilar Manrique, Benjamin Bolduc, Seth T. Walk, John van der Oost, Willem M. de Vos, and Mark J. Young. Healthy human gut phageome. *Proceedings of the National Academy of Sciences*, 113(37):10400–10405, 2016.
- [50] Curtis A. Suttle. Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10):801–812, Oct 2007.
- [51] Maureen A. O’Malley. The ecological virus. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 59:71–79, 2016.
- [52] Britt Koskella and Michael A. Brockhurst. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, 38(5):916–931, 2014.
- [53] Rebecca K. French and Edward C. Holmes. An ecosystems perspective on virus evolution and emergence. *Trends in Microbiology*, 28(3):165–175, Mar 2020.
- [54] Karin Moelling and Felix Broecker. Viruses and evolution – viruses first? a personal perspective. *Frontiers in Microbiology*, 10:523, 2019.
- [55] David Enard, Le Cai, Carina Wennap, and Dmitri A Petrov. Viruses are a dominant driver of protein adaptation in mammals. *eLife*, 5:e12469, may 2016.
- [56] Madhumati Sevana, Thomas Klose, and Michael G. Rossmann. Principles of virus structure. In Dennis H. Bamford and Mark Zuckerman, editors, *Encyclopedia of Virology (Fourth Edition)*, pages 257–277. Academic Press, Oxford, fourth edition edition, 2021.
- [57] Pascal de Boer, Jacob P. Hoogenboom, and Ben N. G. Giepmans. Correlated light and electron microscopy: ultrastructure lights up! *Nature Methods*, 12(6):503–513, Jun 2015.
- [58] Alexander Rigort, Felix J. B. Bäuerlein, Elizabeth Villa, Matthias Eibauer, Tim Laugks, Wolfgang Baumeister, and Jürgen M. Plitzko. Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proceedings of the National Academy of Sciences*, 109(12):4449–4454, 2012.

- [59] Eric C. Keen. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(1):6–9, Jan 2015.
- [60] Kaitlyn E. Kortright, Benjamin K. Chan, Jonathan L. Koff, and Paul E. Turner. Phage therapy: A renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe*, 25(2):219–232, 2019.
- [61] Hans-W. Ackermann. Tailed bacteriophages: The order caudovirales. volume 51 of *Advances in Virus Research*, pages 135–201. Academic Press, 1998.
- [62] Dann Turner, Andrew M. Kropinski, and Evelien M. Adriaenssens. A roadmap for genome-based phage taxonomy. *Viruses*, 13(3), 2021.
- [63] Helen E. White and Elena V. Orlova. Bacteriophages: Their structural organisation and function. In Renos Savva, editor, *Bacteriophages*, chapter 2. IntechOpen, Rijeka, 2020.
- [64] William R. Wikoff, Lars Liljas, Robert L. Duda, Hiro Tsuruta, Roger W. Hendrix, and John E. Johnson. Topologically linked protein rings in the bacteriophage hk97 capsid. *Science*, 289(5487):2129–2133, 2000.
- [65] Corynne L. Dedeo, Carolyn M. Teschke, and Andrei T. Alexandrescu. Keeping it together: Structures, functions, and applications of viral decoration proteins. *Viruses*, 12(10), 2020.
- [66] Kristin N. Parent, Jason R. Schrad, and Gino Cingolani. Breaking symmetry in viral icosahedral capsids as seen through the lenses of x-ray crystallography and cryo-electron microscopy. *Viruses*, 10(2), 2018.
- [67] Peter E Prevelige and Juliana R Cortines. Phage assembly and the special role of the portal protein. *Current Opinion in Virology*, 31:66–73, 2018.
- [68] Lei Sun, Xinzheng Zhang, Song Gao, Prashant A. Rao, Victor Padilla-Sanchez, Zhenguo Chen, Siyang Sun, Ye Xiang, Sriram Subramaniam, Venigalla B. Rao, and Michael G. Rossmann. Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nature Communications*, 6(1):7548, Jul 2015.
- [69] Bo Hu, William Margolin, Ian J. Molineux, and Jun Liu. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science*, 339(6119):576–579, 2013.
- [70] Jiří Nováček, Marta Šiborová, Martin Benešík, Roman Pantůček, Jiří Doškař, and Pavel Plevka. Structure and genome release of Twort-like myoviridae phage with a double-layered baseplate. *Proceedings of the National Academy of Sciences*, 113(33):9351, Aug 2016.

- [71] Nicholas M. I. Taylor, Nikolai S. Prokhorov, Ricardo C. Guerrero-Ferreira, Mikhail M. Shneider, Christopher Browning, Kenneth N. Goldie, Henning Stahlberg, and Petr G. Leiman. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature*, 533(7603):346–352, May 2016.
- [72] James Kizziah, Keith Manning, Altaira Dearborn, and Terje Dokland. Structure of the host cell recognition and penetration machinery of a staphylococcus aureus bacteriophage. *PLOS Pathogens*, 16:e1008314, 02 2020.
- [73] Chantal Hulo, Edouard de Castro, Patrick Masson, Lydie Bougueleret, Amos Bairoch, Ioannis Xenarios, and Philippe Le Mercier. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Research*, 39(1):576–582, 10 2010.
- [74] Janis Doss, Kayla Culbertson, Delilah Hahn, Joanna Camacho, and Nazir Barekzi. A review of phage therapy against bacterial pathogens of aquatic and terrestrial organisms. *Viruses*, 9(3), 2017.
- [75] Andrea Du Toit. Phage induction in different contexts. *Nature Reviews Microbiology*, 17(3):126–127, Mar 2019.
- [76] Beata Orzechowska and Manal Mohammed. The war between bacteria and bacteriophages. In Madhusmita Mishra, editor, *Growing and Handling of Bacterial Cultures*, chapter 6. IntechOpen, Rijeka, 2019.
- [77] Jim Baggen, Hendrik Jan Thibaut, Jeroen R. P. M. Strating, and Frank J. M. van Kuppeveld. The life cycle of non-polio enteroviruses and how to target it. *Nature Reviews Microbiology*, 16(6):368–381, Jun 2018.
- [78] Julian W. Tang and Christopher W. Holmes. Acute and chronic disease caused by enteroviruses. *Virulence*, 8(7):1062–1065, 2017.
- [79] J. M. Hogle, M. Chow, and D. J. Filman. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science*, 229(4720):1358–1365, 1985.
- [80] Michael G. Rossmann, Edward Arnold, John W. Erickson, Elizabeth A. Frankengerger, James P. Griffith, Hans-Jürgen Hecht, John E. Johnson, Greg Kamer, Ming Luo, Anne G. Mosser, Roland R. Rueckert, Barbara Sherry, and Gerrit Vriend. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, 317(6033):145–153, Sep 1985.
- [81] Michael G. Rossmann, Yongning He, and Richard J. Kuhn. Picornavirus–receptor interactions. *Trends in Microbiology*, 10(7):324–331, 2002.
- [82] Xiangxi Wang, Wei Peng, Jingshan Ren, Zhongyu Hu, Jiwei Xu, Zhiyong Lou, Xumei Li, Weidong Yin, Xinliang Shen, Claudine Porta, Thomas S. Walter, Gwyndaf Evans, Danny Axford, Robin Owen, David J. Rowlands, Junzhi Wang, David I. Stuart, Elizabeth E. Fry, and Zihe Rao. A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nature Structural & Molecular Biology*, 19(4):424–429, Apr 2012.

- [83] Tobias J. Tuthill, Elisabetta Gropelli, James M. Hogle, and David J. Rowlands. Picornaviruses. *Current topics in microbiology and immunology*, 343:43–89, 2010.
- [84] Jingshan Ren, Xiangxi Wang, Zhongyu Hu, Qiang Gao, Yao Sun, Xuemei Li, Claudine Porta, Thomas S. Walter, Robert J. Gilbert, Yuguang Zhao, Danny Axford, Mark Williams, Katherine McAuley, David J. Rowlands, Weidong Yin, Junzhi Wang, David I. Stuart, Zihe Rao, and Elizabeth E. Fry. Picornavirus uncoating intermediate captured in atomic detail. *Nature Communications*, 4(1):1929, Jun 2013.
- [85] Hazel C. Levy, Mihnea Bostina, David J. Filman, and James M. Hogle. Catching a virus in the act of RNA release: a novel poliovirus uncoating intermediate characterized by cryo-electron microscopy. *Journal of Virology*, 84(9):4426–4441, 2010.
- [86] Damià Garriga, Angela Pickl-Herk, Daniel Luque, Jürgen Wruss, José Castón, Dieter Blaas, and Núria Verdaguer. Insights into minor group rhinovirus uncoating: The X-ray structure of the HRV2 empty capsid. *PLoS pathogens*, 8:e1002473, 01 2012.
- [87] Kristin Shingler, Jennifer Yoder, Michael Carnegie, Robert Ashley, Alexander Makhov, James Conway, and Susan Hafenstein. The enterovirus 71 A-particle forms a gateway to allow genome release: A cryoEM study of picornavirus uncoating. *PLoS pathogens*, 9:e1003240, 03 2013.
- [88] Rick Conzemius, Haleh Ganjian, Dieter Blaas, Renate Fuchs, and R. M. Sandri-Goldin. ICAM-1 binding rhinoviruses A89 and B14 uncoat in different endosomal compartments. *Journal of Virology*, 90(17):7934–7942, 2016.
- [89] Dieter Blaas and Renate Fuchs. Mechanism of human rhinovirus infections. *Molecular and Cellular Pediatrics*, 3(1):21, Jun 2016.
- [90] Seiya Yamayoshi, Seii Ohka, Ken Fujii, and Satoshi Koike. Functional comparison of SCARB2 and PSGL1 as receptors for enterovirus 71. *Journal of virology*, 87(6):3335–3347, 2013.
- [91] Khairunnisa Mohamed Hussain, Kim Leong, Mary Ng, and Jang Hann Chu. The essential role of clathrin-mediated endocytosis in the infectious entry of human enterovirus 71. *The Journal of biological chemistry*, 286:309–21, 10 2010.
- [92] Aniko Paul and Eckard Wimmer. Initiation of protein-primed picornavirus RNA synthesis. *Virus research*, 206, 01 2015.
- [93] Ravi Basavappa, D.J. Filman, R. Syed, O. Flore, J.P. Icenogle, and James Hogle. Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: Structure of the empty capsid assembly intermediate at 2.9 Å resolution. *Protein Science*, 3:1651 – 1669, 10 1994.

- [94] Jeffrey K. F. Lai, I-Ching Sam, and Yoke Fun Chan. The autophagic machinery in enterovirus infection. *Viruses*, 8(2), 2016.
- [95] Scott M Robinson, Ginger Tsueng, Jon Sin, Vrushali Mangale, Shahad Rahawi, Laura L McIntyre, Wesley Williams, Nelson Kha, Casey Cruz, Bryan M Hancock, et al. Coxsackievirus b exits the host cell in shed microvesicles displaying autophagosomal markers. *PLoS pathogens*, 10(4):e1004045, 2014.
- [96] Sara Whitney Bird, Nathaniel D Maynard, Markus W Covert, and Karla Kirkegaard. Nonlytic viral spread enhanced by autophagy components. *Proceedings of the National Academy of Sciences*, 111(36):13081–13086, 2014.
- [97] Ying-Han Chen, WenLi Du, Marne C Hagemeyer, Peter M Takvorian, Cyrilla Pau, Ann Cali, Christine A Brantner, Erin S Stempinski, Patricia S Connelly, Hsin-Chieh Ma, et al. Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. *Cell*, 160(4):619–630, 2015.
- [98] Issac Horng Khit Too, Huimin Yeo, October Michael Sessions, Benedict Yan, Eshale Anak Libau, Josephine LC Howe, Ze Qin Lim, Shalini Suku-Maran, Wei-Yi Ong, Kaw Bing Chua, et al. Enterovirus 71 infection of motor neuron-like NSC-34 cells undergoes a non-lytic exit pathway. *Scientific reports*, 6(1):1–16, 2016.
- [99] Richard E. Lloyd. Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses. *Virology*, 479-480:457–474, 2015. 60th Anniversary Issue.
- [100] Richard E Lloyd. Enterovirus control of translation and rna granule stress responses. *Viruses*, 8(4):93, 2016.
- [101] Qian Feng, Martijn A Langereis, and Frank JM van Kuppeveld. Induction and suppression of innate antiviral responses by picornaviruses. *Cytokine & growth factor reviews*, 25(5):577–585, 2014.
- [102] Christopher C. Stobart, Jenna M. Nosek, and Martin L. Moore. Rhinovirus biology, antigenic diversity, and advancements in the design of a human rhinovirus vaccine. *Frontiers in Microbiology*, 8:2412, 2017.
- [103] Sarmila Basnet, Ann C. Palmenberg, and James E. Gern. Rhinoviruses and their receptors. *Chest*, 155(5):1018–1025, 2019.
- [104] Triet M. Bui, Hannah L. Wiesolek, and Ronen Sumagin. ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis. *Journal of Leukocyte Biology*, 108(3):787–799, 2020.
- [105] Jeffrey M. Greve, Gary Davis, Ann M. Meyer, Carla P. Forte, Susan Connolly Yost, Christopher W. Marlor, Michael E. Kamarck, and Alan McClelland. The major human rhinovirus receptor is ICAM-1. *Cell*, 56(5):839–847, 1989.

- [106] Donald E. Staunton, Vincent J. Merluzzi, Robert Rothlein, Randall Barton, Steven D. Marlin, and Timothy A. Springer. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell*, 56(5):849–853, 1989.
- [107] Cyrus Chothia and E. Yvonne Jones. The molecular structure of cell adhesion molecules. *Annual Review of Biochemistry*, 66(1):823–862, 1997.
- [108] Jia-huai Wang and Timothy A. Springer. Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunological Reviews*, 163(1):197–215, 1998.
- [109] Jim Baggen, Daniel L. Hurdiss, Georg Zocher, Nitesh Mistry, Richard W. Roberts, Jasper J. Slager, Hongbo Guo, Arno L. W. van Vliet, Maryam Wahedi, Kimberley Benschop, Erwin Duizer, Cornelis A. M. de Haan, Erik de Vries, José M. Casasnovas, Raoul J. de Groot, Niklas Arnberg, Thilo Stehle, Neil A. Ranson, Hendrik Jan Thibaut, and Frank J. M. van Kuppeveld. Role of enhanced receptor engagement in the evolution of a pandemic acute hemorrhagic conjunctivitis virus. *Proceedings of the National Academy of Sciences*, 115(2):397–402, 2018.
- [110] Dominik Hrebík, Tibor Füzik, Mária Gondová, Lenka Šmerdová, Athanassios Adamopoulos, Ondrej Šedo, Zbyněk Zdráhal, and Pavel Plevka. ICAM-1 induced rearrangements of capsid and genome prime rhinovirus 14 for activation and uncoating. *Proceedings of the National Academy of Sciences*, 118(19), 2021.
- [111] Cengiz Koç, Guoqing Xia, Petra Kühner, Spinelli Silvia, Alain Roussel, Christian Cambillau, and Thilo Stehle. Structure of the host-recognition device of Staphylococcus aureus phage 11. *Scientific Reports*, 6:27581, 06 2016.
- [112] Marc C Morais, Shuji Kanamaru, Mohammed O Badasso, Jaya S Koti, Barbara AL Owen, Cynthia T McMurray, Dwight L Anderson, and Michael G Rossmann. Bacteriophage  $\varphi$ 29 scaffolding protein gp7 before and after prohead assembly. *Nature Structural & Molecular Biology*, 10(7):572–576, 2003.
- [113] Altaira D Dearborn, Erin A Wall, James L Kizziah, Laura Klenow, Laura K Parker, Keith A Manning, Michael S Spilman, John M Spear, Gail E Christie, and Terje Dokland. Competing scaffolding proteins determine capsid size during mobilization of staphylococcus aureus pathogenicity islands. *Elife*, 6:e30822, 2017.
- [114] Andrey A Lebedev, Margret H Krause, Anabela L Isidro, Alexei A Vagin, Elena V Orlova, Joanne Turner, Eleanor J Dodson, Paulo Tavares, and Alfred A Antson. Structural framework for DNA translocation via the viral portal protein. *The EMBO journal*, 26(7):1984–1994, 2007.
- [115] Adam S Olia, Peter E Prevelige, John E Johnson, and Gino Cingolani. Three-dimensional structure of a viral genome-delivery portal vertex. *Nature structural & molecular biology*, 18(5):597–603, 2011.

- [116] Alan A Simpson, Yizhi Tao, Petr G Leiman, Mohammed O Badasso, Yongning He, Paul J Jardine, Norman H Olson, Marc C Morais, Shelley Grimes, Dwight L Anderson, et al. Structure of the bacteriophage  $\varphi$ 29 DNA packaging motor. *Nature*, 408(6813):745–750, 2000.
- [117] Qianglin Fang, Wei-Chun Tang, Pan Tao, Marthandan Mahalingam, Andrei Fokine, Michael Rossmann, and Venigalla Rao. Structural morphing in a symmetry-mismatched viral vertex. *Nature Communications*, 11, 04 2020.
- [118] Jingwei Xu, Dianhong Wang, Miao Gui, and Ye Xiang. Structural assembly of the tailed bacteriophage 29. *Nature Communications*, 10(1):2366, May 2019.
- [119] Anastasia A. Aksyuk, Valorie D. Bowman, Bärbel Kaufmann, Christopher Fields, Thomas Klose, Heather A. Holdaway, Vincent A. Fischetti, and Michael G. Rossmann. Structural investigations of a Podoviridae streptococcus phage C1, implications for the mechanism of viral entry. *Proceedings of the National Academy of Sciences*, 109(35):14001–14006, 2012.
- [120] H Hoover-Litty and J M Greve. Formation of rhinovirus-soluble ICAM-1 complexes and conformational changes in the virion. *Journal of Virology*, 67(1):390–397, 1993.
- [121] Chuan Xiao, Tobias J. Tuthill, Carol M. Bator Kelly, Lisa J. Challinor, Paul R. Chipman, Richard A. Killington, David J. Rowlands, Alister Craig, and Michael G. Rossmann. Discrimination among rhinovirus serotypes for a variant ICAM-1 receptor molecule. *Journal of Virology*, 78(18):10034–10044, 2004.
- [122] Li Xing, José M. Casasnovas, and R. Holland Cheng. Structural analysis of human rhinovirus complexed with ICAM-1 reveals the dynamics of receptor-mediated virus uncoating. *Journal of Virology*, 77(11):6101–6107, 2003.
- [123] Antonio Real-Hohn, Martin Groznica, Nadine Löffler, Dieter Blaas, and Heinrich Kowalski. nanoDSF: In vitro label-free method to monitor picornavirus uncoating and test compounds affecting particle stability. *Frontiers in Microbiology*, 11:1442, 2020.
- [124] Li Xing, Karin Tjarnlund, Birgitta Lindqvist, Gerardo G Kaplan, Dino Feigelstock, R Holland Cheng, and José M Casasnovas. Distinct cellular receptor interactions in poliovirus and rhinoviruses. *The EMBO Journal*, 19(6):1207–1216, 2000.
- [125] Jani JT Seitsonen, Shabih Shakeel, Petri Susi, Arun P Pandurangan, Robert S Sinkovits, Heini Hyvönen, Pasi Laurinmäki, Jani Ylä-Pelto, Maya Topf, Timo Hyypiä, et al. Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *Journal of virology*, 86(13):7207–7215, 2012.

- [126] Mihnea Bostina, Hazel Levy, David J Filman, and James M Hogle. Poliovirus RNA is released from the capsid near a twofold symmetry axis. *Journal of virology*, 85(2):776–783, 2011.
- [127] Andrea T Hadfield, Wai-ming Lee, Rui Zhao, Marcos A Oliveira, Iwona Minor, Roland R Rueckert, and Michael G Rossmann. The refined structure of human rhinovirus 16 at 2.15 Å resolution: implications for the viral life cycle. *Structure*, 5(3):427–441, 1997.
- [128] David M Belnap, David J Filman, Benes L Trus, Naiqian Cheng, Frank P Booy, James F Conway, Stephen Curry, Chaitanya N Hiremath, Simon K Tsang, Alasdair C Steven, et al. Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus. *Journal of virology*, 74(3):1342–1354, 2000.