Cryo-EM in structural virology: dissecting the icosahedral capsid

PASI LAURINMÄKI

HiLIFE Helsinki Institute of Life Science - Institute of Biotechnology,
Molecular and Integrative Biosciences Research Programme,
Faculty of Biological and Environmental Sciences
and
Integrative Life Sciences Doctoral Program

University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the
Faculty of Biological and Environmental Sciences of the University of Helsinki
live streaming from Metsätalo auditorium 2 on May 28th at 14:00

HELSINKI 2020

SUPERVISOR
Professor Sarah Butcher, Ph. D.
Molecular and Integrative Biosciences Research Program
Faculty of Biological and Environmental Sciences, and
HiLIFE - Institute of Biotechnology
University of Helsinki

REVIEWERS
Linda Sandblad, Ph.D.
Umeå Core Facility for Electron Microscopy
Umeå University
Sweden

Professor Ilkka Julkunen, M.D., Ph.D.
Institute of Biomedicine
University of Turku and Turku University Hospital

OPPONENT
University Reader Erika Mancini, Ph.D.
University of Sussex
United Kingdom

FACULTY REPRESENTATIVE IN THE GRADING COMMITTEE
Professor Liisa Holm, Ph.D.
Organismal and Evolutionary Biology Research Program
Faculty of Biological and Environmental Sciences
HiLIFE - Institute of Biotechnology
University of Helsinki

CUSTOS
Professor Sarah Butcher, Ph.D.
Molecular and Integrative Biosciences Research Program
Faculty of Biological and Environmental Sciences, and
HiLIFE - Institute of Biotechnology
University of Helsinki

THESIS COMMITTEE
Research director Eija Jokitalo, Ph. D.
HiLIFE - Institute of Biotechnology
University of Helsinki

Docent Tero Ahola, Ph. D.
Department of Microbiology
Faculty of Agriculture and Forestry
University of Helsinki

The Faculty of Biological and Environmental Sciences uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

Cover illustration: Isosurface representation of Nora virus density map

ISBN 978-951-51-6125-3 (nid.)
ISBN 978-951-51-6126-0 (PDF)
Unigrafia
Helsinki 2020

“Once more unto the breach, dear friends, once more”

William Shakespeare; *Henry V (Act III)*, 1600

“Tryptophan is a good place to take a break”

Shabih Shakeel; personal communication, 2016

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

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

II Manole, V., Laurinmäki, P., Van Wyngaardt, W., Potgieter, C.A., Wright, I.M., Venter G.J., van Dijk, A.A., Sewell, B.T., and Butcher, S.J. (2012). Structural insight into African horsesickness virus infection. *Journal of Virology*, **86**, 7858-66.

III Pietilä, M.K., Laurinmäki, P., Russell, D.A., Ko, C.C., Jacobs-Sera, D., Hendrix, R.W., Bamford, D.H., and Butcher, S.J. (2013). Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10604-10609.

IV Laurinmäki, P.*, Shakeel, S.*, Ekström, J.-O.*, Mohammadi, P., Hultmark, D. and Butcher, S.J. Structure of Nora virus at 2.7 Å resolution and implications for receptor binding, capsid stability and taxonomy. *Manuscript.*

*these authors contributed equally

The articles are reprinted in the end of this dissertation.

Also unpublished data will be presented.

Description of key contributions by PL to the publications. Contributions by the first authors are described when there is overlap with PL.

I: PL (shared first authorship) vitrified the samples, calibrated the microscope for high resolution work, performed cryo-EM, calculated the reconstructions, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript. JH supported calculating the reconstructions, performed segmentation and difference imaging, developed novel tools for volume analysis, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript. This article was used in the doctoral thesis of JH.

II: VM Performed part of the cryo-EM, calculated the reconstructions, performed most of the data analysis, did most of the data visualisation and contributed to interpretation of the results and writing the manuscript. PL (second author) vitrified the samples, contributed to designing the cryo-EM experiment, calibrated the microscope for high resolution work, performed most of the cryo-EM, performed part of the data analysis, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript. This article was used in the doctoral thesis of VM.

III: MP performed cryo-EM, calculated the reconstructions, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript. PL (second author) vitrified the samples, calibrated the microscope for high resolution work, participated in the cryo-EM, performed homology modelling, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript.

IV: PL (shared first authorship) contributed to design of the experiments, vitrified the samples, calculated the reconstructions, performed atomic modelling of the amino acid chain, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript. SS supported calculating the reconstructions, supported the atomic modelling, performed refinement of the atomic modelling, did part of the data visualisation and contributed to interpretation of the results and writing the manuscript.

Methods used by PL are summarised in section 3, table 3.

LIST OF PUBLICATIONS NOT INCLUDED IN THE THESIS

Shakeel, S., J. J. T. Seitsonen, T. Kajander, P. Laurinmäki, T. Hyypia, P. Susi and S. J. Butcher (2013). "Structural and functional analysis of coxsackievirus A9 integrin alpha(v)beta(6) binding and uncoating." Journal of Virology **87**: 3943-3951.

Pietilä, M. K., P. Laurinmaki, D. A. Russell, C. C. Ko, D. Jacobs-Sera, S. J. Butcher, D. H. Bamford and R. W. Hendrix (2013). "Insights into head-tailed viruses infecting extremely halophilic archaea." Journal of Virology **87**: 3248-3260.

Hetzel, U., T. Sironen, P. Laurinmäki, L. Liljeroos, A. Patjas, H. Henttonen, A. Vaheri, A. Artelt, A. Kipar, S. J. Butcher, O. Vapalahti and J. Hepojoki (2013). "Isolation, identification, and characterization of novel arenaviruses, the etiological agents of boid inclusion body disease." Journal of Virology **87**: 10918-10935.

Psencik, J., J. B. Arellano, A. M. Collins, P. Laurinmäki, M. Torkkeli, B. Loflund, R. E. Serimaa, R. E. Blankenship, R. Tuma and S. J. Butcher (2013). "Structural and functional roles of carotenoids in chlorosomes." Journal of Bacteriology **195**: 1727-1734.

Nilsson, C., B. Barrios-Lopez, A. Kallinen, P. Laurinmäki, S. J. Butcher, M. Raki, J. Weisell, K. Bergstrom, S. W. Larsen, J. Ostergaard, C. Larsen, A. Urtti, A. J. Airaksinen and A. Yaghmur (2013). "SPECT/CT imaging of radiolabeled cubosomes and hexosomes for potential theranostic applications." Biomaterials **34**: 8491-8503.

Karjalainen, E., N. Chenna, P. Laurinmäki, S. J. Butcher and H. Tenhu (2013). "Diblock copolymers consisting of a polymerized ionic liquid and poly(N-isopropylacrylamide). Effects of PNIPAM block length and counter ion on self-assembling and thermal properties." Polymer Chemistry **4**: 1014-1024.

Hedegaard, S. F., C. Nilsson, P. Laurinmäki, S. Butcher, A. Urtti and A. Yaghmur (2013). "Nanostructured aqueous dispersions of citrem interacting with lipids and PEGylated lipids." RSC Advances **3**: 24576-24585.

Hirvonen, S. P., M. Karesoja, E. Karjalainen, S. Hietala, P. Laurinmäki, E. Vesanen, S. J. Butcher and H. Tenhu (2013). "Colloidal properties and gelation of aqueous dispersions of conductive poly(benzimidazobenzophenanthroline) derivatives." Polymer **54**: 694-701.

Seitsonen, J. J. T., S. Shakeel, P. Susi, A. P. Pandurangan, R. S. Sinkovits, H. Hyvönen, P. Laurinmäki, J. Ylä-Pelto, M. Topf, T. Hyypiä and S. J. Butcher (2012). "Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating." Journal of Virology **86**: 7207-7215.

Dearborn, A. D., P. Laurinmäki, P. Chandramouli, C. M. Rodenburg, S. F. Wang, S. J. Butcher and T. Dokland (2012). "Structure and size determination of bacteriophage P2 and P4 procapsids: function of size responsiveness mutations." Journal of Structural Biology **178**: 215-224.

Koho, T., L. Huhti, V. Blazevic, K. Nurminen, S. J. Butcher, P. Laurinmäki, N. Kalkkinen, G. Rönnholm, T. Vesikari, V. P. Hytönen and M. S. Kulomaa (2012). "Production and characterization of virus-like particles and the P domain protein of GII.4 norovirus." Journal of Virological Methods **179**: 1-7.

Koho, T., T. Mäntylä, P. Laurinmäki, L. Huhti, S. J. Butcher, T. Vesikari, M. S. Kulomaa and V. P. Hytönen (2012). "Purification of norovirus-like particles (VLPs) by ion exchange chromatography." Journal of Virological Methods **181**: 6-11.

Sarin, L. P., J. J. Hirvonen, P. Laurinmäki, S. J. Butcher, D. H. Bamford and M. M. Poranen (2012). "Bacteriophage phi 6 nucleocapsid surface protein 8 interacts with virus-specific membrane vesicles containing major envelope protein 9." Journal of Virology **86**: 5376-5379.

Seitsonen, J., P. Susi, O. Heikkilä, R. S. Sinkovits, P. Laurinmäki, T. Hyypiä and S. J. Butcher (2010). "Interaction of alpha(V)beta(3) and alpha(V)beta(6) integrins with human parechovirus 1." Journal of Virology **84**: 8509-8519.

Huiskonen, J. T., J. Hepojoki, P. Laurinmäki, A. Vaheri, H. Lankinen, S. J. Butcher and K. Grunewald (2010). "Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses." Journal of Virology **84**: 4889-4897.

Paasonen, L., T. Sipilä, A. Subrizi, P. Laurinmäki, S. J. Butcher, M. Rappolt, A. Yaghmur, A. Urtti and M. Yliperttula (2010). "Gold-embedded photosensitive liposomes for drug delivery: Triggering mechanism and intracellular release." Journal of Controlled Release **147**: 136-143.

Saarikangas, J., H. X. Zhao, A. Pykälainen, P. Laurinmäki, P. K. Mattila, P. K. J. Kinnunen, S. J. Butcher and P. Lappalainen (2009). "Molecular mechanisms of membrane deformation by I-BAR domain proteins." Current Biology **19**: 95-107.

Psencik, J., A. M. Collins, L. Liljeroos, M. Torkkeli, P. Laurinmäki, H. M. Ansink, T. P. Ikonen, R. E. Serimaa, R. E. Blankenship, R. Tuma and S. J. Butcher (2009). "Structure of chlorosomes from the green filamentous bacterium Chloroflexus aurantiacus." Journal of Bacteriology **191**: 6701-6708.

Jäälinoja, H. T., E. Roine, P. Laurinmäki, H. M. Kivelä, D. H. Bamford and S. J. Butcher (2008). "Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus." Proceedings of the National Academy of Sciences of the United States of America **105**: 8008-8013.

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

Poranen, M. M., S. J. Butcher, V. M. Simonov, P. Laurinmäki and D. H. Bamford (2008). "Roles of the minor capsid protein P7 in the assembly and replication of double-stranded RNA bacteriophage phi 6." Journal of Molecular Biology **383**: 529-538.

Arellano, J. B., M. Torkkeli, R. Tuma, P. Laurinmäki, T. B. Melo, T. P. Ikonen, S. J. Butcher, R. E. Serimaa and J. Psencik (2008). "Hexanol-induced order-disorder transitions in lamellar self-assembling aggregates of bacteriochlorophyll c in Chlorobium tepidum chlorosomes." Langmuir **24**: 2035-2041.

Strandman, S., A. Zarembo, A. A. Darinskii, P. Laurinmäki, S. J. Butcher, E. Vuorimaa, H. Lemmetyinen and H. Tenhu (2008). "Effect of the number of arms on the association of amphiphilic star block copolymers." Macromolecules **41**: 8855-8864.

Ikonen, T. P., H. Li, J. Psencik, P. A. Laurinmäki, S. J. Butcher, N. U. Frigaard, R. E. Serimaa, D. A. Bryant and R. Tuma (2007). "X-ray scattering and electron cryomicroscopy study on the effect of carotenoid biosynthesis to the structure of Chlorobium tepidum chlorosomes." Biophysical Journal **93**: 620-628.

Psencik, J., J. B. Arellano, T. P. Ikonen, C. M. Borrego, P. A. Laurinmäki, S. J. Butcher, R. E. Serimaa and R. Tuma (2006). "Internal structure of chlorosomes from brown-colored Chlorobium species and the role of carotenoids in their assembly." Biophysical Journal **91**: 1433-1440.

Psencik, J., T. P. Ikonen, P. Laurinmäki, M. C. Merckel, S. J. Butcher, R. E. Serimaa and R. Tuma (2004). "Lamellar organization of pigments in chlorosomes, the light harvesting complexes of green photosynthetic bacteria." Biophysical Journal **87**: 1165-1172.

Doan, D. N. P., K. C. Lee, P. Laurinmäki, S. Butcher, S. M. Wong and T. Dokland (2003). "Three-dimensional reconstruction of hibiscus chlorotic ringspot virus." Journal of Structural Biology **144**: 253-261.

ABBREVIATIONS

2D	two-dimensional
3D	three-dimensional
3DEM	three-dimensional electron microscopy
AHSV	African horse sickness virus
AHSV-4	African horse sickness virus, serotype 4
AHSV-7 tVP2	African horse sickness virus, serotype 7, mutant with truncated VP2
Bam35	bacteriophage Bam35
BTV	bluetongue virus
CCD	charge-coupled device
cryo-EM	cryogenic electron microscopy
CTF	contrast transfer function
DED	electron detector type: direct electron detector; synonymous to DDD
EM	electron microscopy
EMDB	Electron Microscopy Data Bank
FEG	field emission gun
FFT	fast Fourier transform
FSC	Fourier shell correlation
HK97	bacterial virus Hong Kong 97
HSTV-1	*Haloarcula sinaiiensis* tailed virus 1
kV	kilovolt, 1 kV = 1000 V
MCP	major coat protein
NORAV	Nora virus
PDB	Protein Data Bank
PRD1	bacteriophage PRD1
PSF	point spread function
ROI	region of interest
SNR	signal to noise ratio
SPA	single particle averaging
SV 40	simian virus 40
TEM	transmission electron microscope
T-number	triangulation number
Å	Ångström
VP	viral protein

SUMMARY

Technological advances especially in the last two decades have made it feasible to study virus structures to atomic or near-atomic resolution by a combination of cryogenic electron microscopy (cryo-EM) and image reconstruction or by X-ray crystallography. Drawing from these data, intertwined concepts of “viral self” and structure-based viral lineages been put forward and refined during the last decade.

In order to address these theories, I used cryo-EM, 3D image reconstruction, homology modelling and *de novo* atomic modelling of proteins to study five viruses with icosahedrally symmetric protein capsids. The capsid structures of bacteriophage Bam35, two African horse sickness virus serotypes (AHSV-4 and AHSV-7 tVP2), archaeal head-tailed virus HSTV-1, and Nora virus (NORAV) that infects *Drosophila melanogaster* were solved to subnanometer resolution.

I examine the methodological advances in the field of cryoEM and image processing that have occurred over the time span of the original articles included in the thesis, giving an empirical perspective on the important changes required to reach atomic resolution in virus structural determination. These included improved imaging, data recording, automation, and software developments. Notably, grid preparation is the area where the next strides need to be made to improve reproducibility and throughput.

The results contribute to the general field of structural virology as well as shedding light into more specific areas, such as biological membrane modulation, archaeal viruses and vaccine development. Moreover, the hosts of the viruses studied span all three domains of life (bacteria, archaea and eukaryotes). This unusually wide sampling of the viral universe, or *virosphere*, creates an excellent basis for testing the utilisation of capsid structure in structure-based virus classification as well as verification of structure-based viral lineages. The results allowed unambiguous structure-based classification of the five studied viruses into the four previously postulated virus lineages; the picornavirus-like lineage, the HK97-like tailed-phage lineage, the PRD1/adenovirus lineage and the icosahedral dsRNA virus lineage. Furthermore, the NORAV reconstruction at 2.7 Å resolution provided structural evidence suggesting that NORAV is a representative of a new virus family within the order *Picornavirales*; a result not achievable by genetic evidence alone and a benchmark example of structure-based virus classification.

TABLE OF CONTENTS

1. INTRODUCTION

The biosphere of our home, the blue planet, is in fact a sea of viruses with 1 milliliter of natural water containing typically one to ten million (Bergh et al. 1989) virus particles. These protein encoding obligatory intracellular parasites infect all cellular organisms and their estimated planetary number, 4.8×10^{31}, exceeds that of cellular life by at least one order of magnitude (Guemes et al. 2016). While the ubiquitous viruses are nanoscale in dimensions, their impact is genuinely macroscopic ranging from affecting biogeochemical cycles (Chow and Suttle 2015, Danovaro et al. 2016) to comprising a significant factor in the evolution of all cellular organisms (Koonin and Dolja 2013; Koonin and Dolja 2014; Krupovic et al. 2019).

As we sample the vast *virosphere* and the pool of characterised viruses grows, we need ways to contextualise what we detect. Along these lines, in a recent comment in the journal *Nature*, (Kuhn et al. 2019) make a strong case for investing effort on virus classification. Criteria for the classification are varied; we can use for example the host organism, the type of viral nucleic acid, symmetry of the virus capsid, virus genome architecture and virus genomic sequence (Baltimore 1971; Lefkowitz et al. 2018). While these are all valid criteria, the rapid development of structural biology methodology has opened completely new avenues for structural virology. For example, combining cryogenic electron microscopy (cryo-EM, section 1.1.) with single particle averaging (section 1.3.) allows us to look at the structure of the virus particle, or *virion*, at a near-atomic resolution. This facilitates structure-based virus classification and provides novel tools for studying deep evolutionary lineages within the *virosphere*.

I personally advocate the view, or philosophical stance, that increasing the knowledge of our universe is a valuable goal in itself and nothing less than an important contributor to the meaning of human existence. Looking at more direct, practical uses of structural virology we see that it supplies us with the much-needed tools for the constant battle with pathogenic viruses. The structure of a pathogenic virus, among other things, reveals targets for drug and vaccine design (Jiang and Tang 2017; Abdelnabi et al. 2019; Domanska et al. 2019). Furthermore, viruses are essentially efficient replicating nanomachines perfected by the evolution. Learning the principles on which they are built provides tools for emerging technologies where viruses are engineered for specific applications, such as targeting cancer cells or killing pathogenic bacteria.

1.1. Transmission electron microscopy at cryogenic temperatures

The first transmission electron microscope (TEM), or *übermikroskop* (Borries and Ruska 1939, Ruska 1939), was constructed in 1931 by electrical engineers Ernst Ruska and Max Knoll (Harris 2015). Already the early users attempted to image structural details of biological samples such as viruses (Ruska 1939). Modern TEMs have great resolving power that surpasses that of light microscopes by several orders of magnitude; however, electron microscopy (EM) has some inherent features that limit the selection of samples it can be used on:

(1) The electron microscope column is in a vacuum, limiting high resolution transmission EM to imaging non-volatile samples. If water-containing biological sample is to be imaged with traditional transmission EM, the sample has to be dehydrated. Dehydration, regardless of the sample preparation scheme used, has a tendency to alter biological samples which are typically highly hydrated, and thus may collapse (Cheng et al. 2006).

(2) In transmission EM a high-energy electron beam is used to image the sample. However, as the high-energy electrons interact with the sample they break atomic bonds within the sample molecules and create free radicals that cause secondary alterations. These beam damage effects cumulate rapidly and destroy detailed structural features. (Baker et al. 1999; Chen et al. 2008; Glaeser 2016)

(3) In transmission EM, as the name implies, the sample has to be thin enough for the electron beam to pass through.

The issues above were significantly addressed in the early 1980s upon the introduction of transmission EM of samples fixed by cryofixation and imaged at cryogenic temperature; cryo-EM (Adrian et al. 1984; Dubochet et al. 1988; Orlova and Saibil 2011). The significance of this breakthrough was underlined in 2017 when the Nobel Prize in Chemistry was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson *"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*". (Cheng et al. 2017)

1.1.1. Sample preparation for Cryo-EM

Preparation of water-embedded sample for cryo-EM relies on controlled *vitrification* of water. When water is cooled below 0^{o}C, the water molecules in the liquid phase start to transform into solid phase, arranging into ice crystals. The key invention by the pioneers of cryo-EM was to facilitate special, extremely rapid cooling conditions under which the crystal-forming step is bypassed and an amorphous, solid state of water (known also as the *vitreous* state) is formed. This vitreous state of

water is metastable, and in consequence the temperature of the vitrified specimen has to be kept below -150°C. Warming the metastable, amorphous water causes irreversible rearrangement of the water molecules into crystalline form (ice). Ice crystal formation increases the background noise in the images and potentially damages the hydrated sample.

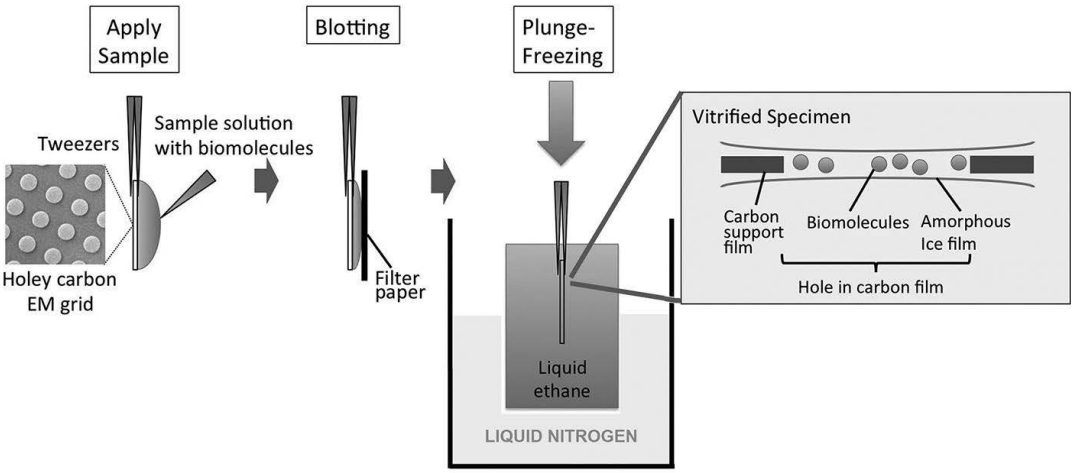


Figure 1. Preparation of vitrified sample for Cryo-EM by plunge freezing technique. Figure adapted from (Murata and Wolf 2018).

The general workflow of sample vitrification by plunge freezing (Figure 1) consists of three main steps:

- The sample is applied onto a holey carbon grid held in tweezers
- Most of the sample liquid is wicked away by blotting with filter paper
- The holey carbon grid carrying the sample in ultrathin water layer is plunged rapidly into a bath of liquid coolant kept below -170°C by a liquid nitrogen bath

These three steps form the generalised backbone of the technique. However, over the last three decades variations in the details have emerged. Starting from the left panel in (Figure 1), various grid pre-treatment protocols have been developed to manipulate the behaviour of buffer and/or sample molecules when they come into contact with the grid. This includes manipulating hydrophilicity and charge of the film as well as more refined treatments such as sample-specific molecular affinity strategies (Kelly et al. 2010; Yu et al. 2016). The cryo-EM grids themselves come in many types. Metals used in the supporting grid include (but are not limited to) copper, gold and molybdenum (Booy and Pawley 1993). The actual thin film that interacts with the sample can be holey, lacey or continuous. The most common thin film material is holey amorphous carbon, however, holey gold film combined with gold support grid has been shown to provide increased sample stability under the electron beam (Russo and Passmore 2014 and 2016). The holes in the film can be random or made to exact specifications with nanofabrication methods (Quispe et al.

2007) available by commercial manufacturers such as Quantifoil Micro Tools or Protochips Inc. This is done because the regular size and lattice of the holes is beneficial for automating the data acquisition. Examples of recent cryo-EM grid inventions can be found in (Pantelic et al. 2010), Palovcak et al. 2018) and (Naydenova et al. 2019), where a holey support film is in turn covered with ultrathin graphene oxide film. (Orlova and Saibil 2011; Murata and Wolf 2018)

The mechanical parameters of the grid blotting event (Figure 1) are varied: side of blotting, one- or two-sided blotting, blotting paper type, blotting time and blotting force can all be optimised for a particular application. Looking at the actual vitrification event, liquid ethane is the most common vitrification coolant but also ethane/propane mixture is used. The benefit of the mixture is a lowered freezing point, which some users prefer in situations where temperature control is not available for the coolant vessel. All of the variations above can be achieved with a simple, manual vitrification device closely reminiscent of the original design in 1980s (Adrian et al. 1984; Dubochet et al. 1988). Today, we also have access to commercially manufactured automated vitrification devices such as Thermo-Fischer Vitrobot, Leica EMGP and Gatan Cryoplunge 3. The benefit of using automated devices is not only better reproducibility of the blotting and plunging parameters, but they also have incorporated environmental chambers that allow control over the temperature of the sample and atmospheric humidity prior and during blotting. (Murata and Wolf 2018)

1.1.2. Operating principle, main components and recent technological breakthroughs of TEM

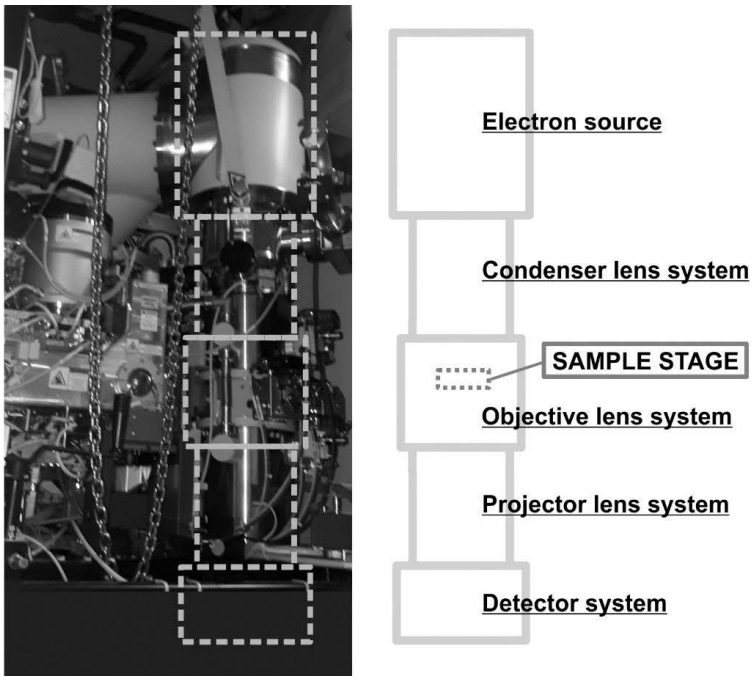


Figure 2. *(Left)* FEI Talos Arctica, a modern 200 kilovolt (kV) cryo-TEM instrument shown here with the protective covers removed. *(Right)* Schematic representation of the main components of a TEM.

In TEM imaging we use a coherent beam of high energy electrons to obtain a magnified two-dimensional (2D) projection image of a thin, electron transparent specimen. The operating principle of TEM relies on deflecting the electron paths by the use of controllable magnetic fields created by powerful magnet coils positioned around the central axis of the microscope column. An electron microscopist uses these magnetic lenses for manipulating the electron beam, achieving the desired imaging conditions in terms of magnification, focus and beam intensity. Furthermore, arrays of smaller magnet coils are used to fine-tune the shape of individual magnetic lenses as well as to tilt and shift the electron beam.

Looking at the overall construction, the primary TEM system can be divided into six main components (Figure 2):

> **Electron source**, also known as the *emitter* or *electron gun* comes in several designs; the most common for modern, high performance TEM is a field emission gun (FEG) that became commonly available in the late 1990s and was used in all original

publications (**I**, **II**, **III**, **IV**) (Figure 3) of this thesis. The core component of a FEG is a heated, monocrystalline tungsten tip coated with zirconium oxide to lower the work function of the electrons. The electrons are extracted from this tip by an electric potential gradient at the emitter surface and then accelerated via high voltage, most typically 200 kV or 300 kV. The key benefit of a FEG is the capability of producing a very bright, stable and coherent electron beam with a narrow energy distribution; these electron beam characteristics are a prerequisite for a high resolution imaging. (Orlova and Saibil 2011)

Condenser lens system comprises several magnetic lenses, condenser apertures and a selection of fine-tuning coils. Condenser lens system makes the raw, diverging electron beam parallel. An advanced three lens condenser system that makes it easier to achieve parallel illumination at variable imaging conditions was used in manuscript **IV**.

Sample holder is docked into the sample stage. For high resolution TEM, the mechanical requirements of the sample holder and sample stage are very high: they have to be exceedingly stable when the image is recorded but on the other hand they have to allow consistent and reproducible movement of the sample in X, Y and Z coordinates as well as allowing tilting of the sample. Working with cryo-EM samples adds an important additional requirement: the sample holder has to keep the sample grid stable at near liquid nitrogen temperature, typically at about -180^{o}C, in order to prevent devitrification of the sample. A side-entry type cryo-holder was used in **I**, **II** and **III**, while a cartridge type cryo-holder, loaded in a so called autoloader, was used in manuscript **IV**. A cartridge type sample holder refers to a design where the sample grid is physically detached from ambient conditions during imaging allowing better isolation. This is beneficial since any disturbances (e.g. thermal, mechanical, acoustic, electric or magnetic) cause loss of resolution in the image.

Objective lens system creates on its back focal plane the primary magnified image of the sample. Because of this pivotal role, any aberrations, limitations or inconsistencies in the objective lens system have great effect on the imaging performance of the whole microscope.

Projector lens system continues from the back focal plane providing further magnification.

Detector system includes the equipment that are used to image and/or record the final, magnified TEM projection image. Electrons are invisible to the human eye, and for live imaging several technical solutions have been developed. The simplest approach is to use a phosphorescent screen to convert the electrons to photons (visible light). This photon image can then be observed directly by eye through X-ray-shielded protective glass or indirectly via a light sensitive camera. For recording the

highest quality data, electron sensitive film was the method of choice up to about 2010, after which we have had an increasing number of options.

Electrons can also be converted to photons by a transparent scintillator that is optically coupled to a recording charge-coupled device (CCD) array operating with the same technical principle as CCD chips of the light sensitive digital cameras. The important contribution of these CCD detectors was the immediate visual feedback on the sample and the shift to a fully digitised, on-line microscopy workflow that enabled automation. However, the single most important breakthrough in the detector technology was the introduction of superior next generation detectors called direct electron detectors (DED) (also known as direct detection devices (DDD) (Orlova and Saibil 2011; Bammes et al. 2012)), that circumvent the photon conversion step altogether by a design that allows direct detection of electrons. The latest commercially available versions of DED are DE-20 and DE-64 (Direct Electron), Falcon 3 (formerly FEI, now Thermo Fischer Scientific) and K3 (Gatan Inc.). DED can also be coupled to an energy filter, a device that is used in cryo-EM to improve image contrast by filtering out inelastically scattered electrons prior to image recording (Yonekura, Braunfeld et al. 2006, Elmlund, Le et al. 2017). The combination has been commercialised by Gatan Inc. in the GIF Quantum K2 and BioQuantum K3 detectors.

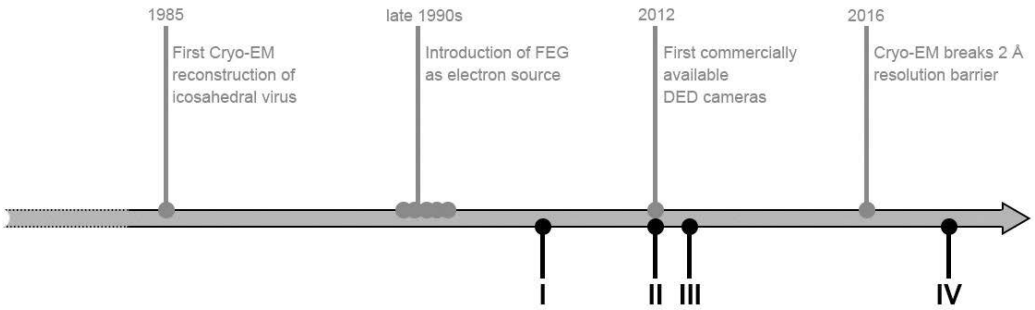


Figure 3. Timeline of selected key technological breakthroughs in cryo-EM (Frank 2017). Original publications of this thesis (**I**, **II**, **III**, **IV**) are marked in.

Integrated and auxiliary computers provide system control, user interface and downstream data management for a fully digitised microscopy workflow that enables automation for recording cryo-EM datasets that are typically one to two orders of magnitude larger than the manually collected datasets. The cumulative consequence of the two technological breakthroughs; DED and automated imaging pathway have contributed to the "resolution revolution" (Kuhlbrandt 2014) in

the three-dimensionally (3D) reconstructed cryo-EM data. For 3D-reconstruction see section 1.3., for “resolution revolution” see section 1.4.

In order to facilitate a functioning TEM system the primary components listed above have to be supported by an array of secondary, supporting equipment. A multistage vacuum pumping system provides the necessary high-quality vacuum, pneumatics system drives the mechanical valves and lead shielding protects the users and the environment from the high-energy X-rays generated in the microscope column. Importantly, constantly running cooling water circuit is required for keeping the lens magnets, detectors and heat-generating pumps at a stable temperature.

1.1.3. Beam damage and low dose technique

High energy electrons are a form of ionising radiation. In order to image the cryo-EM sample with TEM we transmit high energy electron beam through the vitrified sample. An important prerequisite for image formation, interaction of the electrons with the sample, generates a problem: the events between the high energy electrons and the sample molecules cause breakage of molecular bonds and formation of free radicals. This leads to a cascade of chemical reactions causing further entropy in the vicinity of the original event. The resulting detrimental changes in the sample structure are called beam damage (Figure 4). (Baker et al. 1999; Chen et al. 2008; Glaeser 2016)

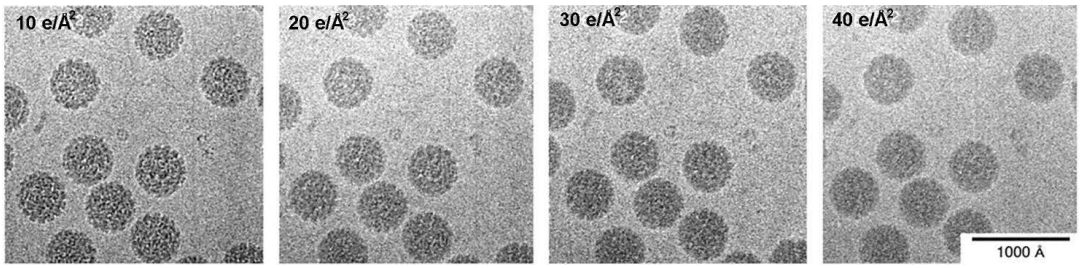


Figure 4. Beam damage; structural details of the biological material are gradually lost as the radiation dose increases. In the case of these simian virus 40 (SV 40) virions, the coat capsomers are clearly visible in the left panel, but when the electron dose reaches 40 e/Å^2, the capsid structure is lost. Adapted from (Baker, Olson et al. 1999) with permission from American Society for Microbiology.

In order to control the electron dose, we have to implement special cryo-EM data collection schemes known as the low dose technique. The details of low dose implementations vary, but the principle is always the same: the system uses different imaging modes for different microscopical tasks. Low-magnification (low electron dose) mode is used to search for areas of interest and perform other auxiliary tasks. Focusing, which requires high electron dose, is performed in the vicinity of the region of interest (ROI) but taking care that the electron beam does not irradiate the ROI. Finally, the acquisition mode is used to record the actual image. The goal of the low dose data

collection scheme is to keep the ROI intact so that close to 100% of the electrons hitting the ROI are used for recording the actual acquisition image. The data acquisition mode is set to record the image with an optimal electron dose.

What then is an optimal electron dose in cryo-EM? Here, we have to settle for a compromise: if the electron dose is increased, the signal to noise ratio (SNR) of the image improves but sample structure is progressively subjected to increasing levels of beam damage. On the other hand, if electron dose is reduced, the effect of beam damage is mitigated but SNR of the images goes down, reducing the quality of the images. The final compromise is dependent on sample buffer conditions, electron wavelength, type of electron detector, and target resolution of the experiment. One fixed value for this compromise, 15 e^- / $Å^2$, can be found from (Chen et al. 2008) where a standard biological sample, vitrified tobacco mosaic virus particles, were imaged at 200kV and subsequently reconstructed. Importantly, the use of DEDs with their rapid readout brings a change: each acquisition can be split into several fractions. This fractionation of data allows the use of high dose (better SNR) during the acquisition, and then *post-acquisition* selection or weighting of the individual fractions to optimise the dose for each image processing task. For the final reconstruction, the analyst can manage the beam damage computationally by selective downweighting or omission of the beam damaged fractions (Li et al. 2013; Cheng 2015).

1.2. Introduction to image formation and image processing concepts

Data from any imaging system is convoluted by the point spread function (PSF) of the imaging system, furthermore, a noise component is added. Image processing (arrows in Figure 5) is the mathematical toolbox that allows us to back-calculate and statistically interpolate the structure of the real object from the raw data images.

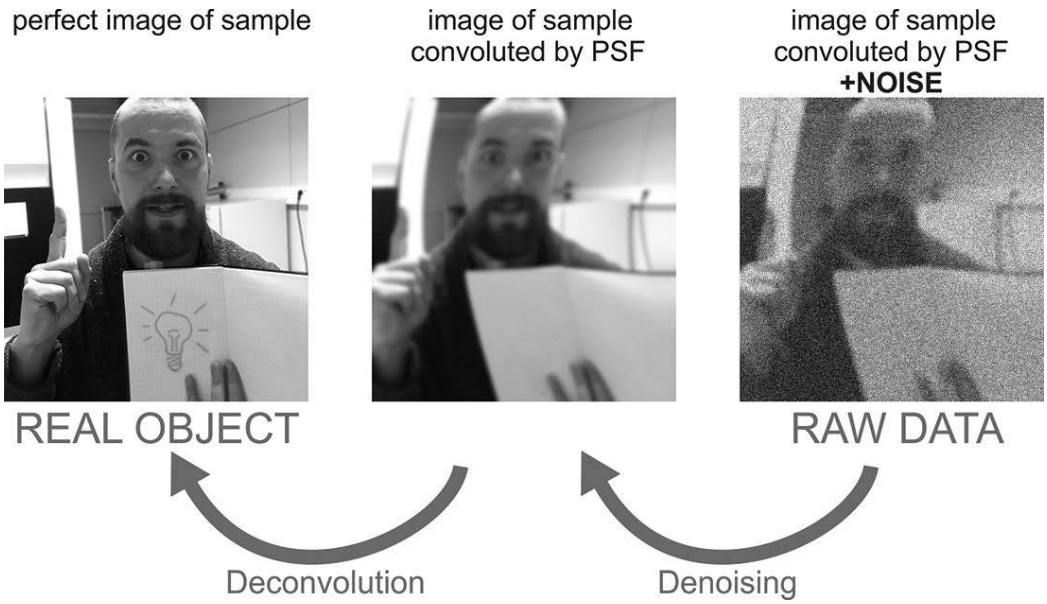


Figure 5. Conceptual relationship between real object and the raw data image produced by the imaging system.

In order to solve the high-resolution details (to see what the excited engineer has in his book in Figure 5) we need to *denoise* and *deconvolute* the raw data.

Denoising:

> Image processing methods that improve the SNR of the sample's features of interest. These include filtering methods as well as statistical approaches that rely on averaging of the raw data. When averaging, the raw data images of multiple identical sample objects are first aligned. The signal for the object features is additive but the random component (noise) is not. As a result, the SNR of the data improves.

Deconvolution:

> Image processing that mathematically corrects for the quantitatively known aberrations of the imaging system (e.g. lens aberrations) and imaging conditions (e.g. defocus). In other words, deconvolution is reverse calculation of the PSF.

The contrast in a projection image can be divided into amplitude contrast and phase contrast components. For samples where the features of interest strongly absorb the illuminating beam, we have a high amplitude contrast. However, in cryo-EM we are typically imaging light element molecules embedded in water and thus the amplitude contrast component is very small, and the phase contrast is the main contributor to the final image contrast. In cryo-EM the relationship between the real structure and the recorded image is described by the contrast transfer function (CTF): (Baker et al. 1999, Orlova and Saibil 2011)

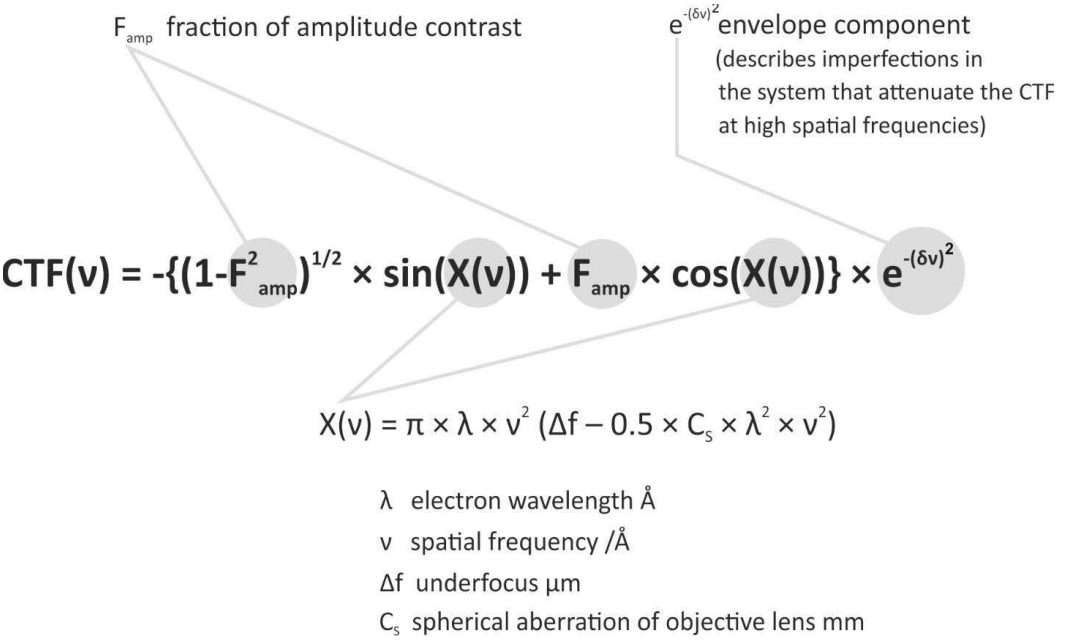


Figure 6. Contrast transfer function (CTF) broken down to components (Baker, Olson et al. 1999).

CTF (Figure 6) includes both amplitude contrast and phase contrast components. The phase contrast component $(1\text{-}F^2_{amp})^{1/2} \times \sin(X(v))$ is a sine function and importantly, while electron wavelength and spherical aberration are fixed system parameters, the phase contrast can be modulated by varying the defocus. What follows is that in cryo-EM we intentionally collect underfocused data that then has a boosted phase contrast component. This imaging strategy is necessary to yield contrast but it comes with a price: the signal for sample features is modulated by the CTF's sinusoid phase contrast component leading to frequency-related signal inversion and variation in signal strength. The resulting artefacts have to be corrected (deconvoluted) during image processing. (Baker et al. 1999)

CTF (Figure 6), is an equivalent of the PSF in frequency- or *Fourier* space. Many image processing approaches make use of Fourier space and rely on an algorithm called Fast Fourier Transform (FFT). FFT reversibly transforms the real-space image into Fourier space (also known as reciprocal space), where the spatial information of the real space image is rearranged into frequency components allowing deconvolution by multiplication with the CTF. Looking from the perspective of the Fourier space, the effect of multiplying by an accurately determined CTF is to: (1) invert image phases which are of the wrong sign due to the microscope transfer function, (2) enhance frequencies near CTF maxima, and (3) attenuate frequencies near zeroes of the CTF. Alternatively, if we look from the real space perspective, the effect is to convolute image densities with the Fourier transform of the CTF, which moves signals displaced by microscope PSF back to their true location in the specimen projection. (Rosenthal and Henderson 2003)

In addition to deconvolution by CTF multiplication, Fourier space is utilised for other useful image processing operations such as band-pass filtering of the data. Here, because frequency components are separated by FFT, we can mask away the signal from unwanted resolution range and then reverse transform the manipulated (band-pass filtered) image into real space. For details of CTF and FFT applications to cryo-EM data processing, see section 1.3.1.

1.3. 3D reconstruction from cryo-EM data by single particle averaging

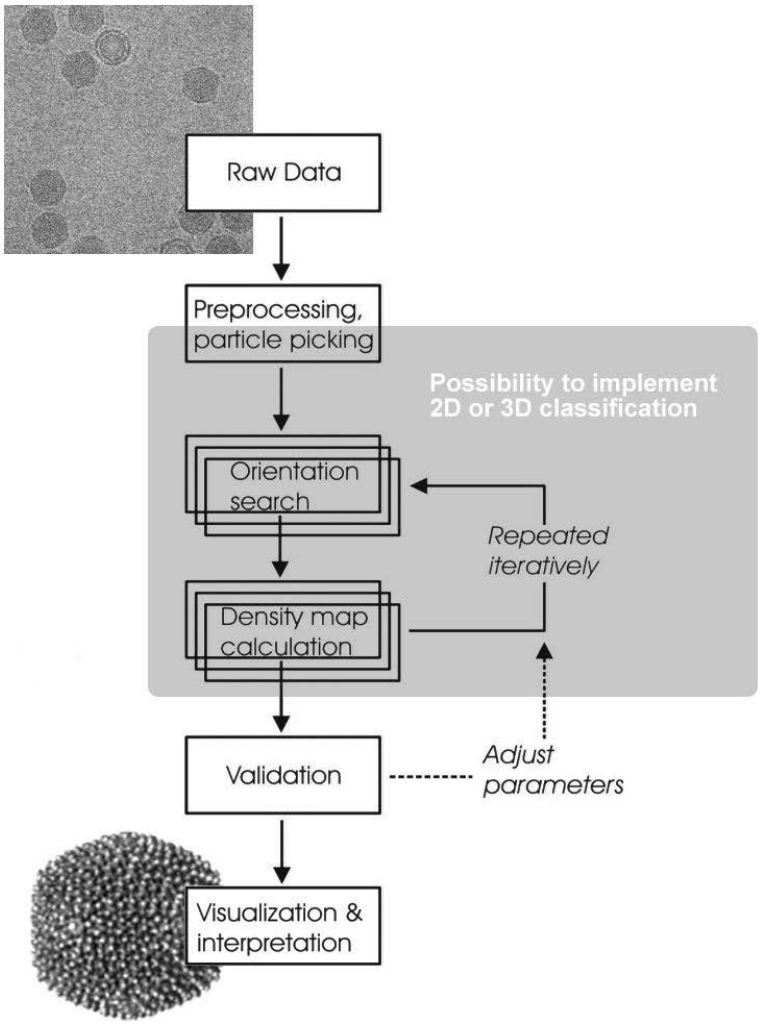


Figure 7. Generalised workflow of 3D reconstruction from cryo-EM data by single particle averaging.

Regardless of the specific algorithms and software used, the single particle averaging (SPA) process follows the workflow depicted in Figure 7. During the last two decades the SPA field has matured and the processing software, such as IMAGIC (van Heel et al. 1996), EMAN (Ludtke et al. 1999), SPIDER (Rath et al. 2003), EMAN2 (Tang et al. 2007), BSOFT (Heymann and Belnap 2007), AUTO3DEM (Yan et al. 2007) (used in **II** and **III**), SPARX (Hohn et al. 2007), XMIPP (Sorzano et al. 2004; Scheres et al. 2008) and RELION (Scheres 2012) (used in **IV**) have manifested increasing ease of operation; command-based user interfaces have been replaced by graphical user interfaces.

Instead of resorting to building *ad hoc* processing pipelines from various independent programs as was done in (**I**), we now have an option of choosing from a few unified processing packages (e.g. SCIPION (de la Rosa-Trevin et al. 2016), SPHIRE (Moriya et al. 2017), cisTEM (Grant et al. 2018), and cryoSPARC (Punjani et al. 2017)) that enable the complete process flow to be conducted and documented under the umbrella of a single user interface. Furthermore, these established reconstruction packages are actively developed (e.g. (Sorzano et al. 2018)) and supported by the user community.

1.3.1. Preprocessing

The generalised preprocessing workflow consists of the following steps:

Motion correction. When working with multiframe DED data the whole process begins with initial frame alignment where the individual frames are aligned to compensate for any image movement that has happened during the acquisition. This process is typically called motion correction. (Li et al. 2013; Cheng 2015)

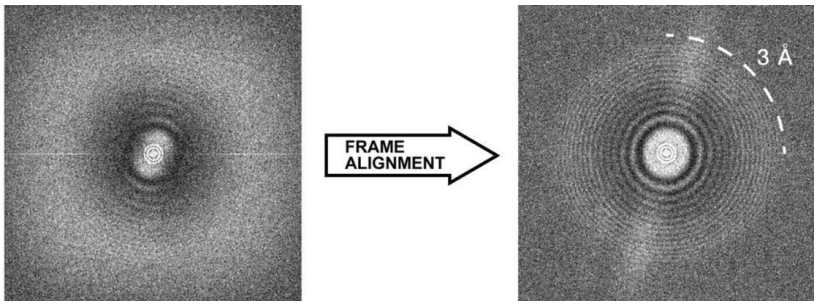


Figure 8. *Left panel:* FFT calculated from uncorrected DED frame stack with drift, showing weak Thon rings that indicate loss of information. *Right panel:* FFT calculated from motion corrected frames showing Thon rings that indicate preservation of information up to 3 Å. Adapted from (Cheng 2015) with permission from Elsevier.

FFT. Once we have a single micrograph (either as direct input or aligned, summed multiframe data from the motion correction -step), we perform FFT and analyse the Fourier transformed image as a quality control step (Figure 8, Figure 9). Micrographs with poor Thon rings, e.g. uncorrected drift (Figure 8, left panel) lacking high resolution signal or showing excessive astigmatism (Figure 9, right panel), are discarded.

Defocus measurement. The defocus of each micrograph is measured by fitting simulated CTF to FFT (Fig 9) with dedicated software, such as CTFFIND3 (Mindell and Grigorieff 2003), CTFFIND4 (Rohou and Grigorieff 2015) or GCTF (Zhang 2016). The exact defocus has to be estimated for deconvolution (CTF correction) of the data.

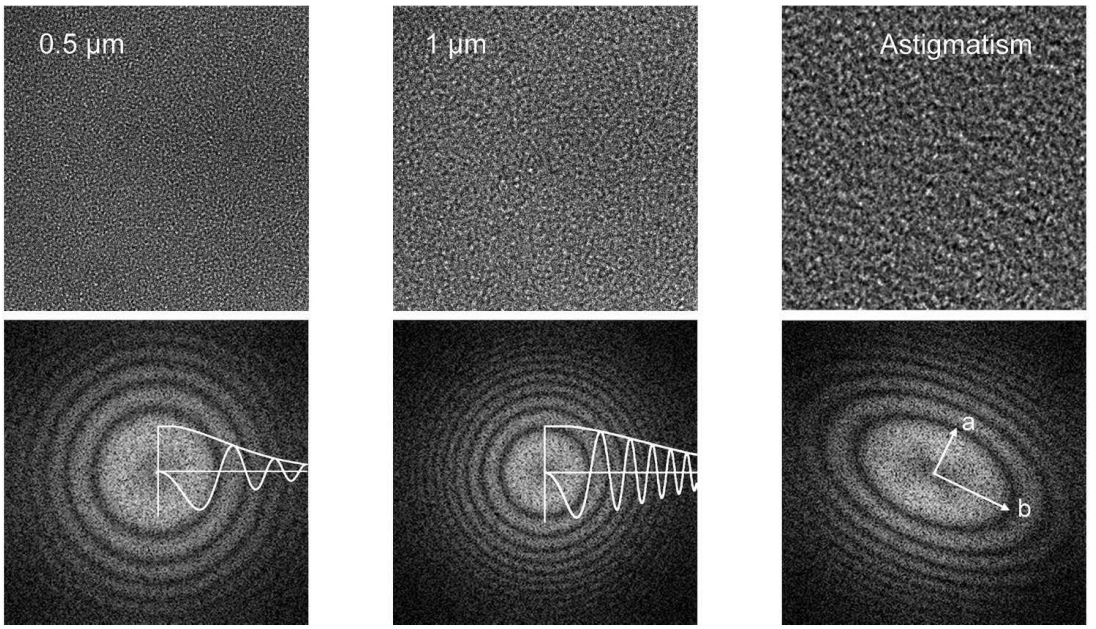


Figure 9. FFT and CTF fitting during of raw data screening and preprocessing. *Top panels*: TEM micrographs of amorphous carbon under various imaging conditions, white text shows underfocus value in µm. Astigmatic micrograph is labelled. *Bottom panels*: FFTs of the respective micrographs, the fitted CTF curve is shown in white. At greater underfocus value the Thon rings are narrower and the CTF oscillates more rapidly. *Right panel*: objective lens astigmatism originates from first-order asymmetry of the objective lens; the lens has higher focusing power in the direction of axis a, and lower in the direction of axis b. Figure adapted from https://pubs.acs.org/doi/10.1021/cr100353t (Orlova and Saibil 2011) with permission from ACS, any further permissions must be directed to ACS.

Particle picking. The simplest approach is manual picking where we browse the micrographs and manually select the particles of interest (e.g. virions). Automated particle picking software can be used to search for regular shapes such as spheres of a given size. This sphere recognition method is used in ETHAN (Kivioja et al. 2000). ETHAN works well with isometric virions and it was used in all four original publications **I**, **II**, **III**, **IV** of this thesis. Many automated particle picking algorithms not covered here utilise template matching approaches but when working with typically noisy cryo-EM data, we have to be careful to avoid the pitfall of model bias or colloquially "Einstein from noise" (Henderson 2013).

Normalisation. The particle images are extracted and normalised. This includes detecting and correcting outlier pixels (pixels with statistically abnormally high or low values), normalisation of possible background gradient(s) and normalisation of overall density distribution.

CTF correction. The normalised particle images are deconvoluted to correct for CTF artefacts. The simplest approach is just a phase correction of the CTF oscillation, but for typical high resolution project we also correct for amplitude decay that occurs at higher spatial frequencies. Furthermore, it is important to take into account that the

SNR drops near the crossovers of the sinusoid CTF. These points in the function have to be Wiener filtered to avoid amplification of noise and compensated by combining micrographs taken at different defocii. (Baker et al. 1999; Orlova and Saibil 2011)

1.3.2. Dataset classification

Classification can be done manually, relying on the subjective selection by the analyst. An easy situation would be where the dataset is of moderate size and we would go through the dataset in order to delete false positives; for example random contamination that has been erroneously selected as sample particles by an automated particle picking algorithm. However, in many cases manual approach has serious caveats: when looking at inherently noisy raw data and attempting to cluster a heterogenous pool of particles into several groups, manual classification is labour intensive and leads almost certainly to biased results (Scheres 2012). Well conducted automated classification schemes not only have the power to generate objective results, but due to the logarithmic increase in the dataset sizes, are becoming *de facto* the only option.

Details of the classification schemes vary depending on application, for example widely used reconstruction programs XMIPP (Scheres, Nunez-Ramirez et al. 2008) and RELION (Scheres 2012) perform maximum likelihood -based 2D and 3D classification steps. In the case of working with a homogenous dataset, e.g. well-purified preparation of icosahedrally symmetric virions, we can consider 2D classification simply as a tool to clean up the dataset of erroneously picked or poorly aligning particles. Essentially, we look at the averages of the classes and discard the poorly averaging classes from further processing steps (Figure 10). On the other hand, in the case of heterogenous sample in multiple oligomeric and/or conformational states, performing statistically robust classification is a critical prerequisite for SPA and 3D reconstruction (Scheres 2012 and 2016).

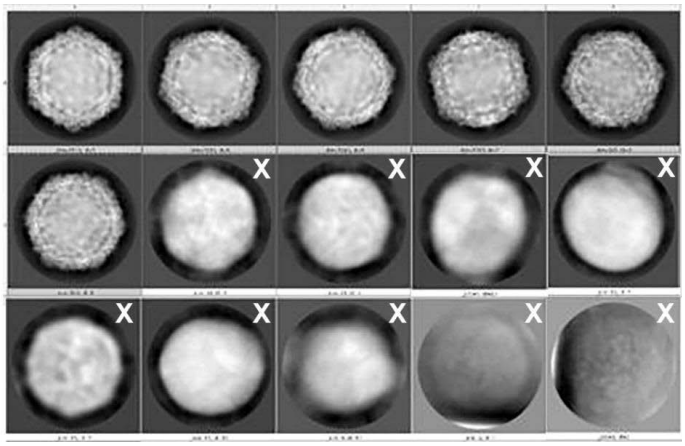


Figure 10. 2D classification as a tool to clean up the dataset of erroneously picked or poorly averaging particles. Shown here are 2D class averages from a picornavirus dataset. Particles forming classes marked with "X" would be typically discarded from further processing steps. Class average images courtesy of Benita Löflund.

1.3.3. The iterative 3D reconstruction process and icosahedral symmetry of virions

The whole SPA process relies on the principle that the 2D projection images are distinct views of structurally identical 3D objects that we then mathematically treat as views of the same object. We are averaging the data, so if previously conducted classification produced several classes we have to perform the following steps separately for each structurally unique class. (Baker et al. 1999; Orlova and Saibil 2011)

First the centres of each particle image are refined. Then, in Fourier space, the relative orientation of each 2D projection is found by correlation with the initial 3D model. This initial model can be obtained with various means, a common approach used in (**I**) is to use a low-pass filtered structure of a known, structurally similar virus. In this case heavy filtering of the initial map, e.g. to 6 nm resolution, is paramount so that any possibility of model bias can be avoided. An alternative unbiased option is to manually pick a small subset of particle images, typically 100 to 200, and use these to calculate a low resolution *ab initio* starting model (Yan et al. 2007). This random model (RM) method is implemented in AUTO3DEM (Yan et al. 2007), and used for generating starting models in (**II**, **III**, **IV**).

Once we have the initial orientations of 2D projections in the same coordinate register, we combine the 2D data and perform a reverse Fourier transform to calculate the first 3D reconstruction. This new map is then used as a new model in a new round of orientation search. The iterative cycle of orientation search and density map calculation (Figure 7) is repeated until there is no further improvement in the resolution of the 3D reconstruction. The reconstruction quality is typically assessed with algorithms that are based on Fourier shell correlation (FSC) (Harauz and van Heel 1986); for use of FSC in data validation see section 1.3.4. (Baker et al. 1999; Orlova and Saibil 2011)

The generalised process described above is descriptive of any SPA reconstruction project. In the special case of reconstructing icosahedrally (Figure 11) organised structures, such as capsids of icosahedrally symmetric virions, we gain significant boost in SNR and computing efficiency by utilising the inherent symmetry of the sample particle. An icosahedron is built from 60 identical asymmetric units (Figure 11A), what follows is that when we perform icosahedrally averaged reconstruction we only need to perform orientation search and density reconstruction within one asymmetric unit instead of the complete orientation space. Furthermore, we gain 60-fold boost in the SNR of the parts of the structure that follow icosahedral symmetry. (Baker et al. 1999; Orlova and Saibil 2011)

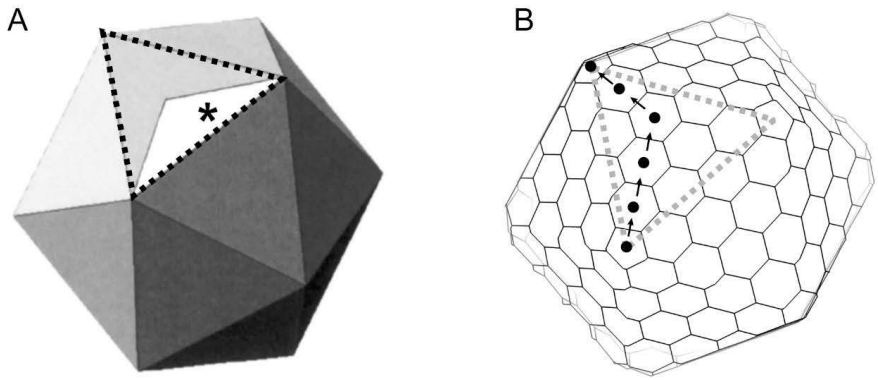


Figure 11. Icosahedrons with a single facet outlined with a dotted line. (A) A geometric icosahedron. A single asymmetric unit (out of 60, in white) is marked with an asterisk. (B) An icosahedron consisting of repeating quasi-equivalent (Caspar and Klug 1962) units organised on a T=19 lattice.

Viruses as replicating entities are evolutionally pushed for optimising the efficiency of the virion construction. Viruses with protein capsid face a special optimisation problem of having to build a protein capsid that is large enough to accommodate the genome that carries the code for not only the structural proteins of the capsid, but also all other proteins needed in the full virus lifecycle. Thus, the ratio of cost (genomic information coding for the capsid) vs. capacity (the capsid size) needs to be minimised, and utilising icosahedral symmetry provides a scalable solution. In essence, an icosahedrally-symmetric capsid of any size can be constructed from an increasing number of repeating, quasi-equivalent subunits (Figure 11B). (Caspar and Klug 1962; Baker et al. 1999)

The concept of triangulation number (*T*-number, $T = h^2 + hk + k^2$) (Caspar and Klug 1962) is useful for describing the icosahedral lattice. Parameters h and k are deduced from the lattice by following the path of hexagons from one vertex pentagon to an adjacent one. In Figure 11B, arrows show three steps (h = 3) followed by two steps (k = 2) to reach the adjacent vertex. What follows is that $T = 3^2 + 3 \times 2 + 2^2 = 19$. Thus, the theoretical capsid lattice in Figure 11B can be constructed from $19 \times 60 = 1140$ copies of single subunits forming the hexamers and pentamers that populate the icosahedral lattice. In their theory of quasi-equivalence, Caspar and Klug (1962) proposed that the individual subunits can be identical at the amino acid sequence, but conformational differences would allow for minor local variation to incorporate the subunits into the icosahedral lattice. Notably, this concept was conceived prior to the determination of any viral capsid structures at atomic resolution, from interpretation of the capsomers seen in electron-micrographs of negatively-stained virus preparations. (Baker et al. 1999)

While *T*-number and quasi-equivalence are very useful concepts it should be noted that they are most suitable for describing the geometry of the virion, whereas the biological reality has shown many deviations from the simplest biological interpretation of protein homopentamers and homohexamers forming the entire capsid. These deviations include, but are not limited to, three distinct proteins with similar fold building the *pseudo* T = 3 seen in the picornaviruses (Seitsonen et

al. 2010; Seitsonen et al. 2012; Shakeel et al. 2013); hexavalent positions occupied by two-domain trimers in the *pseudo* T = 25 lattice seen in PRD1 (Abrescia et al. 2004); hexavalent positions occupied by pentamers in SV40 (Baker et al. 1989); vertices formed by different proteins altogether as seen in PRD1 (Abrescia et al. 2004), SH1 (Jäälinoja et al. 2008), P23-77 (Jaatinen et al. 2008); bluetongue virus (BTV) core following T = 1 of dimers, or "T = 2", lattice (Grimes et al. 1998); and additional minor structural proteins as seen e.g. in the capsids of HVTV-1 and HSTV-2 (Pietilä et al. 2013). As the pool of solved icosahedral lattices has grown, the data have been analysed to build on the original (Caspar and Klug 1962) theory. The results include more generalised tiling approaches (Twarock 2004) and distinct lattice classes based on hexamer complexity (Mannige and Brooks 2010).

1.3.4. Data validation

How to ensure that the reconstructed density does not include artefacts produced by the reconstruction process? The motivation for data validation is elegantly addressed in (Henderson et al. 2012), where the authors state:

> "Every 3DEM map and model has some uncertainty. Therefore, an assessment of map and model errors is essential, especially when a wide range of techniques are used by a variety of practitioners. In addition, as with all rapidly developing fields, in our enthusiasm to go further and faster, there is a risk that avoidable mistakes, both large and small, may be made in the production or interpretation of maps. Such mistakes may have the adverse effect of undermining the credibility of 3DEM methods in general. It is therefore important to develop methods for checking our conclusions and validating maps and models, with the goal of establishing a set of best practices for the field."

The recommendations in (Henderson et al. 2012) for validating single-particle cryo-EM reconstructions emphasize that each published reconstruction should be accompanied by key supporting data; it has to be possible to compare map projections with raw data and (if available) class averages of the raw data. Importantly, also coordinate metadata (the calculated orientations) of the particles included in the reconstruction should be provided so that it is possible to verify unbiased sampling of the complete orientation space. Finally, a statistically robust measurement of the map resolution, typically FSC-based, should be provided and this number should be verified by a direct observation of distinguishable features (e.g. α-helices or amino acid side chains) within the map. (Henderson et al. 2012)

A single-number resolution value for the reconstructed density map can be estimated with various means, but by far the most prevalent is the FSC (Harauz and van Heel 1986). In the FSC, the dataset is split into two, and two reconstructions are independently calculated. The 3D Fourier transforms of the two density maps are correlated with each other, and the resulting FSC curve is used for

reading a resolution value corresponding to a chosen correlation cutoff. Over the years several cutoff values have been proposed, most common ones being a conservative 0.5 used in (**I**, **II**, **III**) and 0.143 (Rosenthal and Henderson 2003) used in (**IV**) and now widely accepted. The theoretical behaviour of FSC and caveats of assessing FSC cutoff value have been addressed in great detail (van Heel and Schatz 2005).

1.4. Analysing cryo-EM reconstructions at different resolution ranges

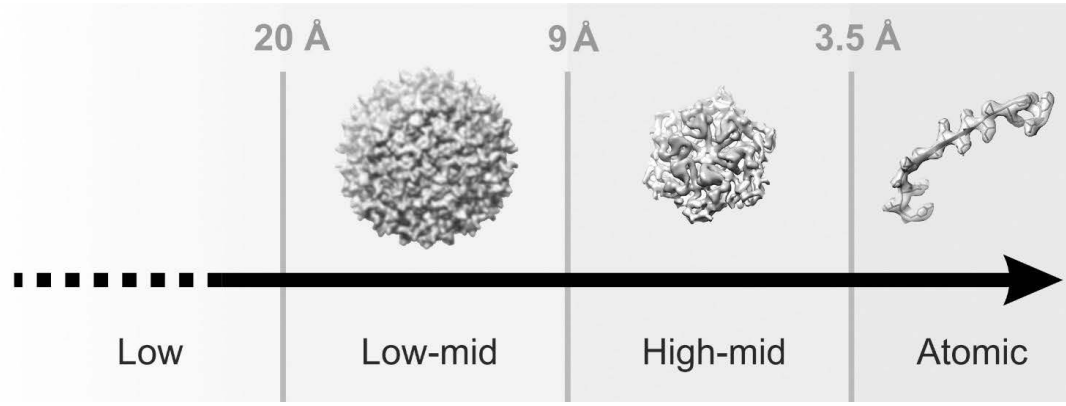


Figure 12. Resolution range reachable by cryo-EM of biological samples. The names of the resolution ranges are arbitrary and chosen for the purposes of this thesis.

The final resolution reached by the reconstruction process dictates the available analysis methods. While resolution values form a continuum, the feasibility of analysis methods have distinct threshold values; based on these I have divided the continuum into separate resolution ranges.

1.4.1. Low-mid resolution: rough morphology

For the purposes of this thesis, I consider a resolution range of 20 Å to 9 Å "low-mid". In this range, rough morphological comparison with known virus structures becomes an option and is usually a natural point to start the analysis. Towards the higher end of the range, the data analyst can detect protein boundaries with increasing reliability.

1.4.2. High-mid resolution: segmentation allows structure fitting

For the purposes of this thesis, I consider a resolution range of 9 Å to 3.5 Å "high-mid". The transition point from the low-mid range comes about at the point where the two main types of protein secondary structural, α-helices and β-sheets, can be reliably distinguished from each other. Reconstruction reaching high-mid resolution allows the analyst to segment out subvolumes corresponding to individual protein subunits. Commonly available and easy-to-use rigid body fitting tools such as "fit in map" in UCSF CHIMERA (Pettersen et al. 2004; Goddard et al. 2007) can be used for detailed comparisons with atomic models of known protein structures obtained from open databases such as Protein Data Bank (PDB) (Berman et al. 2000, Berman et al. 2003) and density maps from Electron Microscopy Data Bank (EMDB) (Lawson et al. 2011).

At high-mid resolution the analyst gains access to powerful hybrid approaches. Here, reconstruction serves as a macroscale scaffold into which atomic structures solved by other techniques can be docked, as was done in (**I**, **II**, **III**). In a classic example an atomic resolution structure of the main coat protein of a virus is solved by X-ray crystallography and combined with a cryo-EM reconstruction of the complete virus capsid (Martin et al. 2001, San Martin 2015). The result, more than a mere sum of the two, can be described as a quasi-atomic model of the virion; for detailed example of this method see (Martin et al. 2001). In an analogous approach, if amino acid sequence of the structural protein is known, the protein structure can be predicted via homology modelling and then utilised in this hybrid approach. Homology modelling with I-Tasser server (Zhang 2008; Roy et al. 2010; Yang et al. 2015) followed by docking into cryo-EM structure is exemplified in (**II**, **III**).

1.4.3. Atomic resolution: segmentation and atomic modelling of the amino acid chain

Modelling of amino acid chain becomes possible at about 3.5 Å and at higher resolutions the generated atomic coordinates become more accurate and reliable. Entering this resolution range shifts the entire paradigm of EM-structure analysis as it opens the massive analysis and verification toolbox developed over decades of protein structure science based on X-ray diffraction data collected on synchrotrons.

Densities corresponding to single amino acid chains are segmented out and coordinates of the amino acid backbone and side chains are fitted and refined with well-established and constantly developing software tools such as REFMAC (Murshudov et al. 1997) or COOT (Emsley and Cowtan 2004, Emsley et al. 2010) and PHENIX (Adams et al. 2010), which were used in (**IV**). The process produces relative atomic coordinates of the amino acid chain that describes the structure of the protein in question. These models follow standardized schema and are amenable to quantitative quality control by verification algorithms (Lawson et al. 2016). While the reconstructed density maps can be deposited into EMDB (Lawson et al. 2011), the models derived from the maps are deposited into PDB (Berman et al. 2000; Berman et al. 2003). EMDB file formats, documentation, user interface

and search tools are modelled to provide maximal synergy with the protein models in PDB. The fact that the structural models follow unified standards brings significant benefits as the analyst now has a possibility to compare the structure to other known structures in a quantitative setting.

Analysis tools for atomic resolution maps are becoming increasingly relevant as the latest generation cryo-EM TEMs with DEDs and an automated imaging pathway have facilitated the rapid growth in the number and quality of cryo-EM structures (Figure 13), shifting the field from qualitative towards quantitative science (Kuhlbrandt 2014; Kuhlbrandt 2014; Cheng 2015; Henderson 2015).

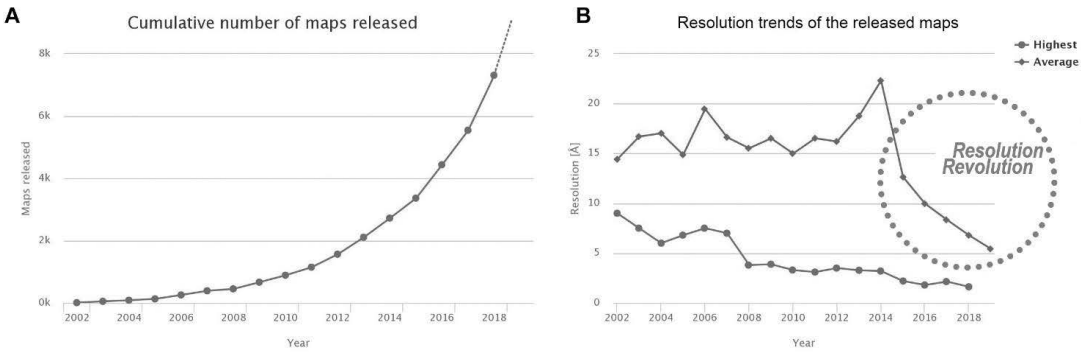


Figure 13. Logarithmic growth in the number and "resolution revolution" (Kuhlbrandt 2014) of the released cryo-EM maps. (A) Cumulative total number of released EM maps. (B) Resolution trends of the released EM maps, contribution of recent years' "resolution revolution" is marked with circle. Figure based on statistics available at EMDB (Lawson et al. 2011), website accessed on Jan 7th, 2020.

1.5. Concise introduction of the studied viruses

Bam35 virus is a member of the *Tectiviridae* infecting *Bacillus thuringiensis*, a Gram-positive bacterial host. Bam35 has an isometric protein capsid enclosing a lipid bilayer and a linear double-stranded DNA genome (Ackermann, Roy et al. 1978, Ravantti, Gaidelyte et al. 2003). The overall dimensions of Bam35 capsid and enclosed membrane, as well as genome size show clear similarity with those of PRD1, the type member of the *Tectiviridae* that infects Gram-negative bacterial host (Ravantti et al. 2003; Laurinavicius et al. 2004). The PRD1 and Bam35 genome sequences have no significant similarity limiting the possibilities of utilising genomic approaches for virus comparison (Ravantti et al. 2003). Cryo-EM image reconstruction and X-ray crystallography have been used to solve the PRD1 capsid structure (Abrescia et al. 2004, Cockburn et al. 2004), but the lack of structural information on Bam35 has prevented detailed comparison between these two tectivirus structures. An open question here is whether or not there is conservation of the protein folds between the Gram-positive and the Gram-negative viruses.

African horsesickness virus (AHSV) causes African horsesickness (AHS), a devastating arthropod-borne disease of equids with a mortality rate of up to 95% in fully susceptible horses and donkeys (Mellor and Hamblin 2004; Carpenter et al. 2017). AHSV has a complex, triple-layered, icosahedrally symmetric protein capsid enclosing a segmented, linear, double-stranded RNA genome (Roy et al. 1994). AHSV is classified into the genus *Orbivirus* within the family *Reoviridae* (Roy et al. 1994; Stuart et al. 1998; Mertens et al. 2004). The prototype species of *Orbivirus* is BTV, another important animal pathogen infecting sheep, cattle and goats (Mertens et al. 2004; Zientara and Sanchez-Vizcaino 2013). The triple-layered BTV capsid is comprised of viral protein (VP)3 inner layer, VP7 middle layer (Grimes et al. 1997; Grimes et al. 1998); and outer layer comprised of VP2 and VP5 (Hewat et al. 1992; Mertens et al. 2004). (Basak et al. 1996) solved the structure of the top domain (124 amino acids out of 349) of AHSV VP7 revealing the similarity to the structure of BTV VP7. VP7s of the two viruses are trimeric, and the top domains have similar jelly-roll folds. The top domains of the VP7s can be superimposed with an 1.2 Å rms deviation between all equivalent Cα atoms (Basak et al. 1996). Furthermore, (Stuart et al. 1998) anticipated that the similarity between AHSV and BTV would also extend to other parts of the virion.

AHSV, like BTV, is transmitted via the *Culicoides* biting midge vectors when infected *Culicoides* feed on a susceptible host (Venter et al. 2000). The European northward

spread of several *Culicoides* species from the Mediterranean area has been linked to the ongoing climate change (Purse et al. 2005). This has significant repercussions in animal health since the spread of BTV in Europe (Thiry et al. 2006; Meiswinkel et al. 2007) suggests an increasing risk of future AHS outbreaks and motivates AHS vaccine development. The efforts for novel vaccines focus on the AHSV outer shell protein VP2, a serotype determinant that can induce a protective immune response (Burrage et al. 1993; Martinez-Torrecuadrada and Casal 1995; Scanlen et al. 2002; Castillo-Olivares et al. 2011). Hence the mapping of mutations to the as yet unknown VP2 structure may help in developing better candidate vaccines.

***Haloarcula sinaiiensis* tailed virus 1 (HSTV-1)** infects *H. sinaiiensis*, an extremely halophilic archaeon. Negative-stain electron microscopy of HSTV-1 has revealed an isometric protein capsid with an attached tail structure. Based on the morphology of the virion, HSTV-1 is classified as a podovirus. (Atanasova et al. 2012)

Podoviruses from bacteria typically have a major capsid protein with the Hong Kong 97 (HK97) -fold, first described in HK97, forming both the pentamers and hexamers of the icosahedrally organised virion head (Wikoff et al. 2000; Helgstrand et al. 2003). The open question here is whether or not that fold also occurs in archaeal podoviruses, despite the large amino acid sequence diversion.

Nora virus (NORAV) is a picorna-like virus infecting several species of *Drosophila*. It has an isometric protein capsid enclosing a linear single-stranded RNA genome. The exact taxonomic classification of NORAV within the order *Picornavirales* is unclear. Furthermore, the protein composition and organisation of the capsid is unknown, as is the explanation to how NORAV capsid accommodates the almost 12 kb RNA genome, one of the largest *Picornavirales* genomes known. (Habayeb et al. 2006; Habayeb et al. 2009)

2. AIMS OF THE STUDY

The main objective of the study: to analyze the feasibility of utilising virus capsid structures in high-order virus classification. I used cryogenic electron microscopy (cryo-EM) combined with icosahedrally averaged single-particle reconstruction as the main structure determination method. The studied viruses infect hosts from all three domains of life (Figure 14, Table 1).

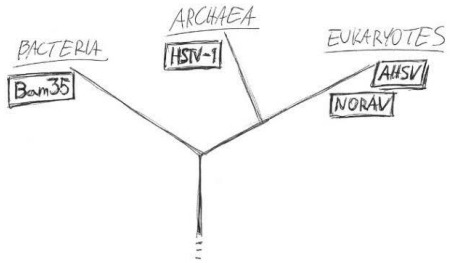


Figure 14. The hosts of the studied viruses span all three domains of life.

The specific aims were:

- To study the structure of bacteriophage Bam35 and compare it with the structure of bacteriophage PRD1, a type member of the *Tectiviridae* **(article I)**.

- To study the capsid structures of virions from two serotypes of AHSV with special emphasis on the viral coat protein VP2 in order to augment vaccine development **(article II)**.

- To study the capsid structure of archaeal head-tailed virus HSTV-1 in order to see whether or not the so-called canonical coat protein fold of bacterial head-tailed virus HK97 can be detected in an archaeal virus **(article III)**.

- To solve the capsid structure of NORAV to atomic resolution and to utilize this structure for performing structure-based classification of NORAV **(article IV)**.

3. MATERIALS AND METHODS

- Viruses studied in this dissertation are listed in Table 1.
- Cryo-EM equipment used in this dissertation are listed in Table 2.
- Experimental methods used by P. Laurinmäki in this dissertation are listed in Table 3.

The materials and methods are described in detail in the original articles and the respective references. The articles are reprinted in the end of this dissertation.

Table 1. Viruses included in the study.

Article	Virus	Virus host	Virus host type
I	Bam35	*Bacillus thuringiensis*	Bacteria (Gram positive)
II	AHSV-4	*Equidae*	Eukaryote (mammal)
II	AHSV-7 tVP2	*Equidae*	Eukaryote (mammal)
III	HSTV-1	*Haloarcula sinaiiensis*	Archaea
IV	NORAV	*Drosophila sp.*	Eukaryote (insect)

Table 2. Summary of cryo-EM equipment used in the study.

Article	Virus	Microscope type	Voltage	Detector type
I	Bam35	FEI Tecnai F20	200 kV	Film
II	AHSV-4	FEI Tecnai F20	200 kV	Film
II	AHSV-7 tVP2	FEI Tecnai F20	200 kV	Film, CCD (Gatan US4000)
III	HSTV-1	FEI Tecnai F20	200 kV	Film, CCD (Gatan US4000)
IV	NORAV	FEI Titan Krios	300 kV	DED (Gatan GIF Quantum K2)

Table 3. Summary of methods used by P. Laurinmäki in the study.

Method	Article
Virus purification by ultracentrifugation	**I**
Vitrification of virus samples	**I, II, III, IV**
Alignment of TEM for high resolution cryo-EM work	**I, II, III**
Low electron dose cryo-EM, recording high-resolution datasets	**I, II, III**
Preprocessing of cryo-EM data	**I, IV**
CTF determination and quality assessment	**I, IV**
Iterative single particle reconstruction and optimization thereof	**I, IV**
Fitting atomic model into EM density map	**II, III, IV**
Segmenting EM Density map	**III, IV**
Homology modelling	**III**
Subvolume averaging	**III**
De novo modelling of amino acid chain	**IV**
Data visualisation	**I, II, III, IV**

4. RESULTS AND DISCUSSION

The results of this thesis illustrate directly several of the key advances in cryo-EM imaging and image reconstruction that have been achieved in the past 15 years, summarized in Table 4 by considering the "resolution" of the final reconstructions included in articles **I-IV**. The resolution increased from 18 Å (**I**) to 2.7 Å (**IV**), so from the low end to the atomic level described in Figure 12. In terms of biological information, five different virus structures were described which contribute significantly to our understanding of virus capsid protein evolution, and to the development of new vaccine candidates (Sections 4.1-4.5).

Table 4. Summary of cryo-EM data and achieved resolutions of the reconstructions.

Article	Virus	Detector	Micrograph no.	Particle no.	Sampling rate (Å / pixel)	Resolution (Å) of reconstruction
I	Bam35	Film	69	4474	1.4	7.3
I	Bam35**e**	Film	32	379	1.4	18.0
II	AHSV-4	Film	307	1633	2.8[1]	14.4
II	AHSV-7 tVP2	Film	507	2798	2.8[1]	11.4
II	AHSV-7 tVP2**e**	CCD	70	361	1.8	15.8
III	HSTV-1	Film	164	7115	1.13	8.9
IV	NORAV	DED	3531*	16131	1.06	2.7[2]

e Empty particle reconstruction. In these cases the raw data contained virus particles that were devoid of the genome density inside. Such projection images were picked, classified and reconstructed as a separate dataset.

***** Total micrograph number before selection.

[1] Sampled at 1.4 Å / pixel, but only 2 x 2 binned data was processed

[2] Calculated with 0.143 FSC cutoff, if not labelled 0.5 FSC cutoff was used

In the summary of the cryo-EM datasets (Table 2, Table 4, Figure 15), it is not just the significant improvement in the resolution that stands out, but also the changes in data collection strategy that enabled a much larger data set to be collected, of higher quality, that could then be effectively classified to select only the most congruent data at high resolution using different software (Table 4, grey emphasis). Hence, this study exhibits both the enablers and the outcome of the cryo-EM "resolution revolution" (Figure 13) with the most significant enablers described in more detail below.

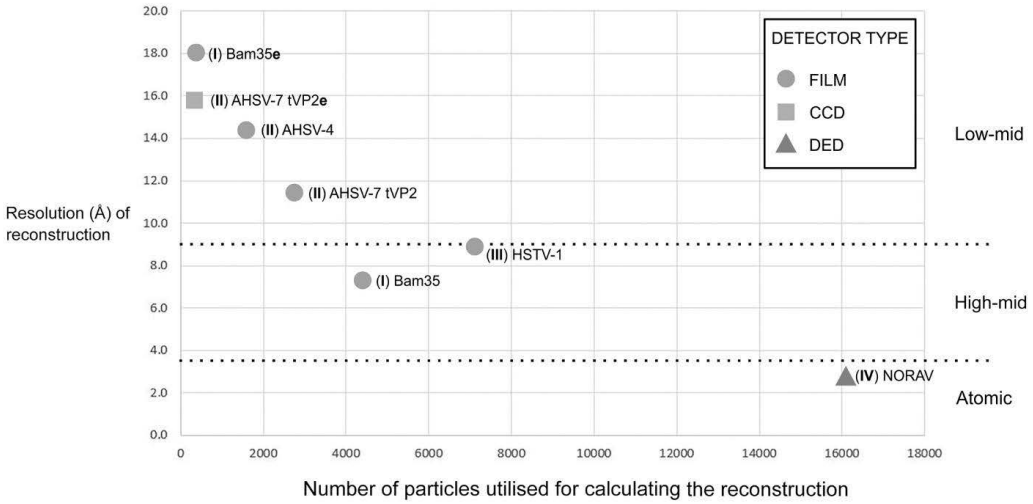


Figure 15. Plotted overview of the cryo-EM detector, dataset size and final resolutions of the reconstructions. Resolution range naming on the right side according to Figure 12, section 1.4.

The micrograph number (Table 4) describes the workload of acquiring datasets, whereas the particle number (Table 4, Figure 15) describes the actual yield of the data collection. Acquiring a single micrograph with a low dose technique (Section 1.1.3.) requires several actions: the system has to select ROI, center on the ROI, perform focus and acquire the micrograph. The micrographs then needed to be developed and scanned if recorded on film. In the case of (**I**, **II**, **III**), these steps were performed manually by the microscopist, sample optimization and data collection spread over several years. In the case of (**IV**) the latest generation microscope and DED facilitated not only better data quality, but also automated data acquisition that enabled much better data yield per hour of microscopist. The data in (**IV**) were collected over an intense period of three days, but the access to the instrument took over 6 months. The achievement of the 2.7 Å atomic resolution was only possible because the majority of the raw data had a significant signal beyond 4 Å resolution, visible in the analysis of the Thon rings from the FFTs of the micrographs, after frame alignment. In the case of data collected on CCD on the F20 (**II**), the data rarely extended to beyond 14 Å resolution, and on film the best dataset (**I**) was selected from micrographs that showed data to 5 Å resolution, but with many micrographs discarded due to this selection criterion. This reflects the sensitivity of the F20 side-entry holder to drift which when recorded on film cannot be corrected, and thus was a significant factor in discarding many micrographs. The DED data in (**IV**) were corrected for drift by frame alignment (Figure 8), and the sample grid was completely isolated from the ambient conditions, reducing environmental instability. The FEI Krios microscope used in (**IV**) also has improved electron optics compared to the FEI F20 used in (**I**, **II**, **III**), contributing to a more accurate alignment and a better beam coherence (Figure 8, Figure 9).

The sampling rate directly affects the achievable resolution, with a theoretical limit of double the sampling rate (DeRosier and Moore 1970, Bammes, Rochat et al. 2012). The film data were scanned at the highest resolution of the scanner, at 7 µm, resulting in a digitised area of

approximately 14000 x 10000 pixels. Hence, even the largest viruses were present in several copies per micrograph at 1.4 Å/pixel sampling rate. The recording arrays of CCD (4096 x 4096 pixels) and DED (3838 x 3710 pixels) equipment used in this thesis were significantly smaller, so in the case of large virions, a compromise would have to be made in the magnification (and thus sampling rate) to ensure that the micrographs produce an adequate yield of whole particles for processing. However, since NORAV had the smallest capsid of the studied viruses, we could still use the highest sampling rate (Table 4), enabling the highest theoretical resolution.

The results show a clear correlation between the dataset size and the achieved final resolution (Figure 15). This is expected as the average signal is increased with the increased number of particles. It is not a linear increase probably reflecting that the resolution criteria differed between the different datasets, as well as the reconstruction methods. The main criterion affecting the film data was the extent of the signal in the micrograph FFTs as stated above (**I**, **II**, **III**). For creating the initial icosahedrally-symmetric reconstruction, article (**I**) relied on model-based creation of the initial model (Baker and Cheng 1996), whereas the other three articles (**II**, **III**, **IV**) relied on the newer, robust, *ab initio* method (Yan et al. 2007) implemented in AUTO3DEM (Yan et al. 2007). For the DED dataset there were sufficient images to classify and use maximum likelihood algorithms implemented in RELION (Scheres 2012) to sort out some of the heterogeneity in the dataset, which significantly improved the reconstruction from about 5 Å onwards. Since the processing of our data in (**IV**), the processing methods applicable to DED data have been developed further in terms of e.g. motion correction and dose weighting (Zivanov, Nakane et al. 2019).

Closer analysis raises the question of the final resolution of HSTV-1 (**III**) which seems to be somewhat poorer than expected. One possible explanation lies in biology; the haloarchaeal host of HSTV-1 grows in high salt conditions (9% salt water containing 1.2 M NaCl and a large number of other monovalent and divalent cations, prepared from stock solution described in (Holmes and Dyall-Smith 1990)); and HSTV-1 infectivity requires high salt conditions whereas cryo-EM imaging requires low salt concentration for good contrast and SNR. To address this issue, in (**III**) it was verified that the overall morphology and infectivity were retained after HSTV-1 was subjected to low salt conditions (20 mM Tris-HCl buffer) and was thus imaged in this buffer. However, it seems feasible that such a radical reduction of the buffer ion concentration could still cause some protein capsid destabilisation, as in (Ruokolainen et al. 2019), or conformational flexibility that would then lead to the loss of high resolution information in the averaged data, although not affecting the virus titre significantly. In addition, the extent of the signal in the micrograph FFTs was not as high as in (**I**) so a less stringent micrograph selection criteria was used. Looking at the reason for this apparent lack of highest resolution signal, the instability of the side-entry holder, suboptimal alignment, thicker than normal vitrified layer due to the presence of tails in the preparation or some unknown factor(s) affecting the microscope performance cannot be excluded. Nevertheless, the 8.9 Å resolution that was reached was adequate for reaching the high-mid range (Figure 12, Figure 15) and met the biological aims of the study (Section 4.3.).

The four reconstructions that were calculated from datasets smaller than 4000 particles (Table 4, Figure 15) are likely to report an overly conservative resolution value and in reality provide useful information beyond the reported value. This originates from the FSC–based resolution determination method (Harauz and van Heel 1986); the dataset is split in two and two reconstructions are calculated. In the case of small datasets the full dataset has significantly better SNR than the halved datasets used for the resolution determination. The effect of reduced SNR on the FSC result is exacerbated when using the conservative 0.5 FSC cutoff (Table 4) as was done in (**I**, **II**, **III**). There is no simple answer to the correct FSC cut-off value that would fit every situation (van Heel and Schatz 2005). However, reproducing the full FSC curve in the publications, or on deposition to the database as is now currently the practice, allows the reader to better judge the reproducibility of the data. Validation tools that take into account the variability of resolution within the reconstruction, e.g. ResMap (Kucukelbir et al. 2014) (used in (**IV**)) and MonoRes (Vilas et al. 2018) are increasingly used. In addition, tools that use alternative validation criteria, e.g. EMRinger (Barad et al. 2015), that is based on fitting of the atomic models to the EM density provide a more detailed indication of the accuracy of the atomic model, in addition to the accuracy of the density map. Community efforts, e.g. EM validation challenges (Heymann et al. 2018; Kryshtafovych et al. 2018; Terwilliger et al. 2018), to identify the best validation methods are ongoing and reconstruction validation statistics are now available in the EMDB and PDB for deposited data, e.g. NORAV (**IV**), which is gradually helping to standardise the field.

The five virus capsid reconstructions that provided subnanometer information enabling assignment of the capsid architecture and analysis of structural details are summarized in Figure 16. The deposition codes of all seven reconstructions are given in Table 5.

Table 5. Accession numbers of open source density maps, atomic models and data.

Article	Virus	EMDB	PDB	EMPIAR
I	Bam35	EMD-1123		
I	Bam35**e**	EMD-1124		
II	AHSV-4	EMD-2076		
II	AHSV-7 tVP2	EMD-5412		
II	AHSV-7 tVP2**e**	EMD-2075		
III	HSTV-1	EMD-2279		
IV	NORAV	EMD-3528	5mm2	10088

e Empty particle reconstruction. In these cases the raw data contained virus particles that were devoid of the genome density inside. Such projection images were picked, classified and reconstructed as a separate dataset.

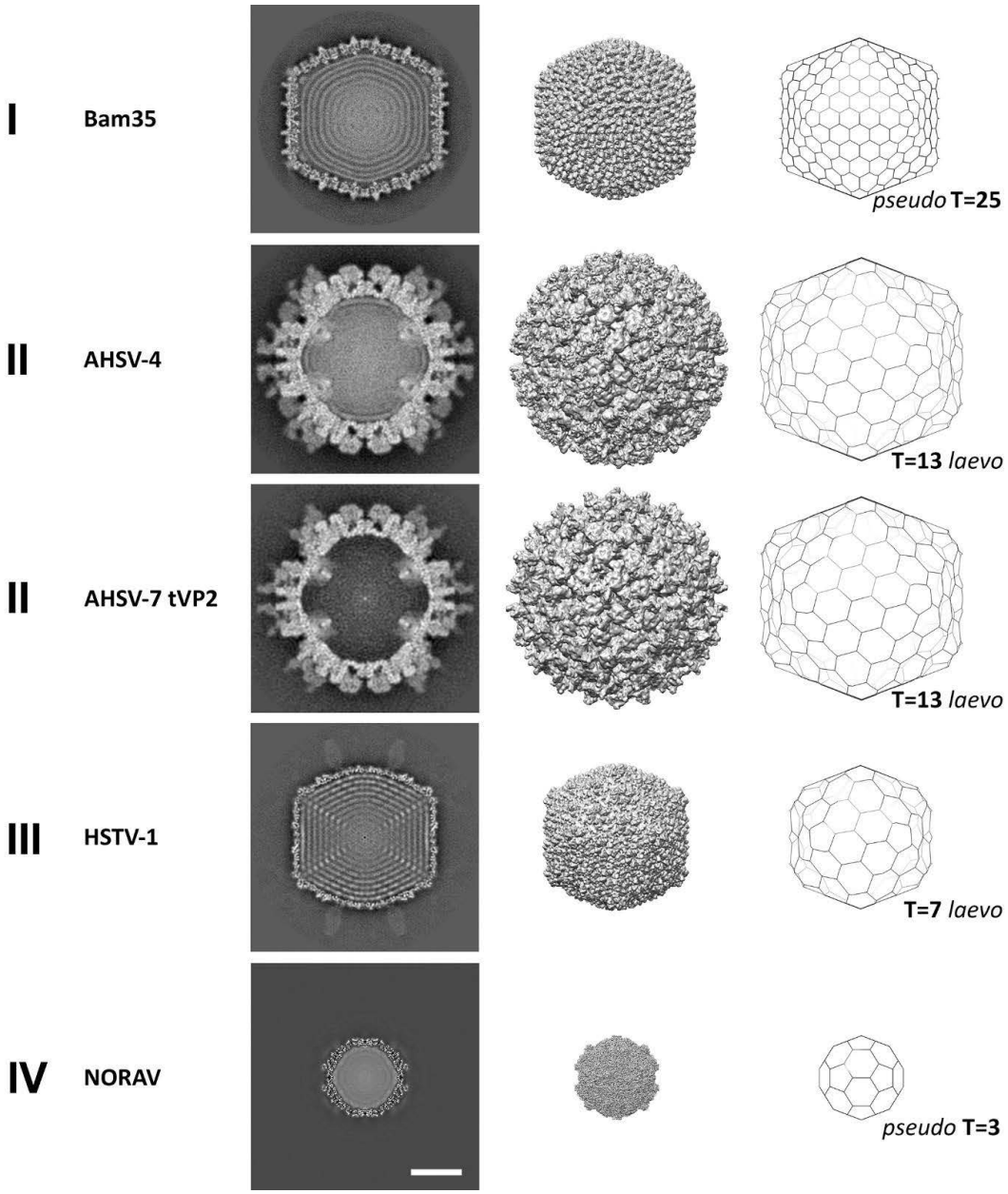


Figure 16. Capsid structures of the viruses studied in this thesis. Columns from left to right: (1) Roman numerals refer to the Table 1 in Materials and Methods; the original articles where each structure was published. (2) Name of the virus. (3) Central section through the reconstructed density map. AHSV-7 tVP2 shows the virion devoid of genome, the other central sections are from full virion reconstructions. (4) Surface view of the reconstructed density map. (5) Schematic representation of the capsid lattice architecture and T-number of the capsid lattice. Scalebar 20 nm.

4.1. Structure of Bam35 capsid reveals remarkable similarity in coat protein folds of Bam35 and PRD1

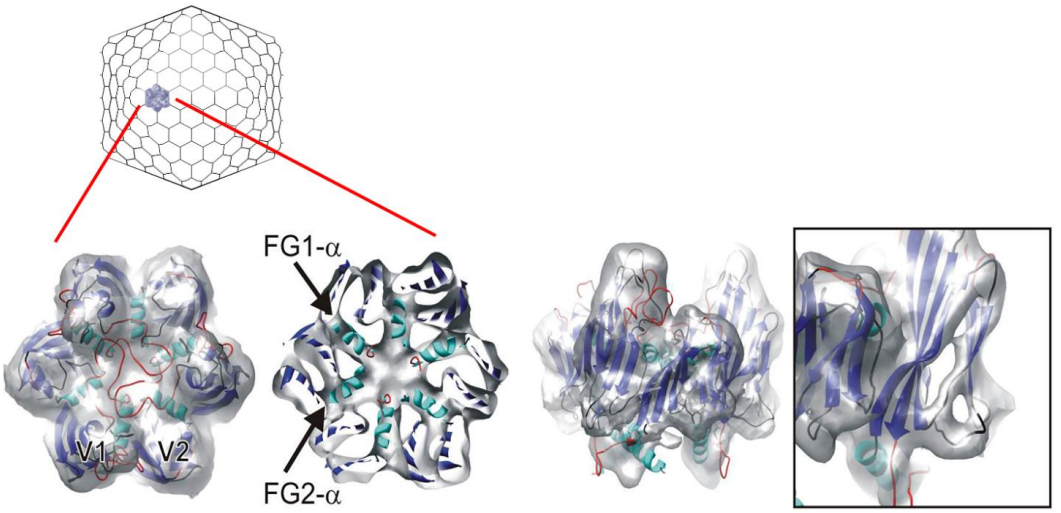


Figure 17. Comparison of the Bam35 major coat protein (MCP) with the atomic model of PRD1 MCP P3 (Abrescia et al. 2004) shows remarkable similarity. The reconstructed Bam35 MCP structure is shown as transparent gray and the PRD1 MCP structure is shown as a ribbon with α-helices coloured in cyan and β-strands coloured in blue. The loops not present in Bam35, and thus projecting out from the MCP density, are in red. The two types of β-barrels are labelled, large (V1) and small (V2). Naming of α-helices according to (Benson et al. 1999). Adapted with permission from (**I**).

In this study, Cryo-EM followed by SPA reconstruction solved the capsid structure of bacteriophage Bam35. Earlier negative stain electron micrographs of Bam35 (Ackermann et al. 1978) had indicated a low-level morphological similarity with bacteriophage PRD1, the type member of the *Tectiviridae*. The 7.3 Å resolution of Bam35 cryo-EM reconstruction proved that this was a valid hypothesis, by revealing the capsid's secondary structure and enabling a detailed comparison of structures of Bam35 MCP and PRD1 MCP, P3 (Abrescia et al. 2004). The main result of this comparison is the remarkable similarity of the overall capsid lattice architecture and the MCP fold (Figure 17).

The high similarity between the MCP folds is significant since Bam35 infects Gram-positive hosts whereas PRD1 infects Gram-negative hosts. Both bacteriophages are obligate parasites of their host bacteria, so the common ancestor of these viruses is likely to reside in very distant history. In line with the long evolutionary distance, similarity in the genomic sequence between the two viruses is low; when both local and global alignments were used, the overall nucleotide identity across the entire genome of Bam35 and PRD1 was 42%, with further analysis showing that the

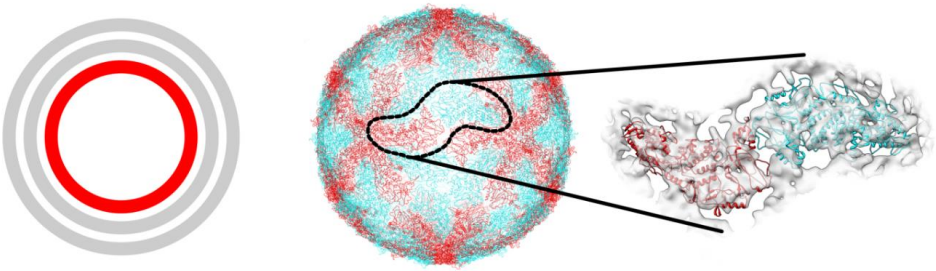
regions of identity were very patchy (Ravantti et al. 2003). Taken together, the genetic and structural comparisons between Bam35 and PRD1 build a strong case for the MCP fold being conserved between the two viruses, as a result of a divergent rather than convergent evolution.

The similarity between the MCPs of Bam35 and PRD1 is striking (Figure 17). Furthermore, PRD1 minor structural proteins, penton protein P31 and tape measure protein P30, have structural counterparts in Bam35 (**I**). These similarities in the icosahedrally organised capsid lattice are contradicted by the dissimilarity in the membrane-associated proteins, membrane organisation and the receptor binding proteins, all of which clearly differ between Bam35 and PRD1 (**I**). Adaptation to very different hosts (Gram positive vs. Gram negative bacteria) is the most likely explanation for these observed differences and supports the view of the highly conserved MCP fold as the "true self" of the virus. In agreement with later studies (Abrescia et al. 2012), MCP structural analysis is thus a promising tool for detecting distant evolutionary connections within the *virosphere*.

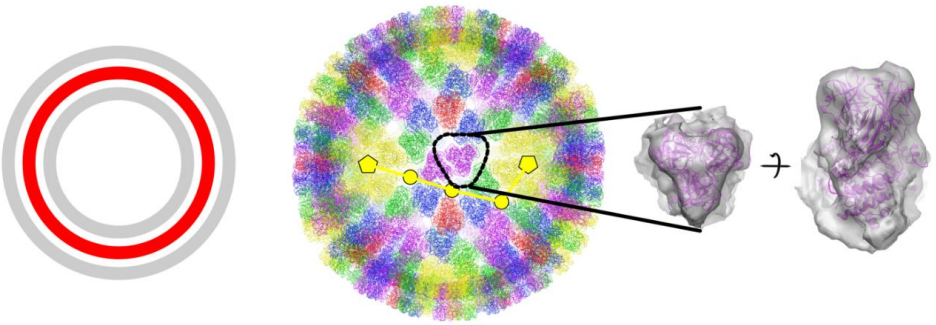
In terms of future prospects, while the achieved 7.3 Å resolution was a very good result at the time of the study, it would be very interesting to analyse the structural details of Bam35 at the atomic resolution reachable with the present-day cryo-EM technology. A good example of the state-of-the-art can be found in the recent 2.3 Å cryo-EM structure of a close relative of PRD1, bacteriophage PR772 (Reddy et al. 2019). Significantly, in that study, local reconstruction of the vertices without the icosahedral symmetry indicated that the pentamer base is a mixed pentamer of two copies of P31, and three copies of P5, with P2 attached to the pentamer base, thus resolving a long-standing structural question (Abrescia et al. 2004; Merckel et al. 2005; Huiskonen et al. 2007) in the field. This is an excellent example illustrating the advantage of SPA over crystallography for virus capsids where one set of data can now be analysed in several different ways, revealing the position and organisation of minor components, and nucleic acid that deviate from icosahedral symmetry (Dai and Zhou 2018; De Colibus et al. 2019; Ilca et al. 2019; Liu et al. 2019). In X-ray crystallography, the crystal packing for capsids does not necessarily orient all of these minor components identically, and thus when the structure is solved using icosahedral symmetry, the components are averaged out, just as with the data in articles (**I**, **II**, **III**, **IV**).

4.2. Structures of AHSV-4 and AHSV-7 tVP2 solve the triple-layered AHSV capsid architecture

(A) Inner layer; VP3 dimers organised on a T=1 lattice



(B) Middle layer; VP7 trimers organized on a T=13 *laevo* lattice



(C) Outer layer; VP5 trimers organized on a T=13 *laevo* lattice decorated by VP2 trimers

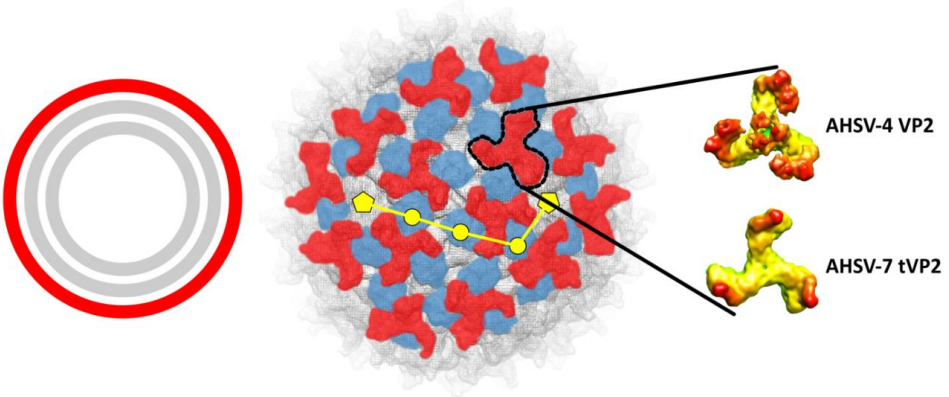


Figure 18. Structure of AHSV reveals organisation of all three protein layers. (A) Homology model of the VP3 shell, zoom-in shows fitting of homology model with the cryo-EM density. (B) Homology model of the VP7 shell, zoom-in shows fitting of homology model with the cryo-EM density. (C) Surface layer of AHSV structure showing VP5 trimers in blue and VP2 trimers in red. Zoom-in shows comparison between full-size VP2 of AHSV4 next to the truncated VP2 of the AHSV7. Adapted from (**II**) with permission from American Society for Microbiology.

In this study, Cryo-EM followed by SPA reconstruction, volume segmenting and homology modelling solved for the first time the organisation of the complete AHSV virion to the conservatively estimated resolutions of 14.4 Å (AHSV-4) and 11.4 Å (AHSV-7 tVP2). The complex, triple-layered AHSV capsid (Figure 18) is composed of four major structural proteins; VP3 dimers form the inner layer, VP7 trimers form the middle layer and VP5 and VP2 trimers form the outer layer. The capsid structure has a close similarity to that of BTV, an important animal pathogen and a type member of a postulated BTV-like structural lineage (Abrescia et al. 2012). Based on the achieved resolution and description of the resolution ranges (Figure 12, Figure 15) it would seem that it in the case of AHSV reconstructions, segmenting of volumes corresponding to individual proteins is not possible. However, the closely similar BTV structure provided critical *a priori* information that facilitated reliable segmenting of the AHSV protein volumes.

The outermost structural protein of BTV and AHSV, VP2 (Figure 18 C), is believed to mediate the initial cell attachment during the infection cycle and is a pathogenicity determinant. Importantly, AHSV-7 tVP2 has a truncated VP2 and outgrows AHSV-4 in cell culture. Comparison of VP2 structures from AHSV-7 tVP2, AHSV-4 and BTV allowed mapping of two domains in AHSV4 VP2 and one domain in BTV VP2 that are important in mediating the initial cell attachment during infection and are thus promising targets for further studies leading towards vaccine development (Aksular et al. 2018).

In conclusion, while the aims of the present study were met, it would be very interesting to utilise the latest generation cryo-EM technology to solve the AHSV virion structure to an atomic resolution. The ability to analyse the structural proteins at amino acid level would reveal not only the complex network of interactions within the triple-layered capsid, but would also allow a direct study of individual mutations in the structure of the pathogenically important VP2. Along these lines, a recent atomic resolution cryo-EM structure of the BTV RNA-dependent RNA polymerase in the context of the virion illustrates the present state-of-the art of *Orbivirus* cryo-EM (He et al. 2019).

4.3. Structure of HSTV-1 capsid reveals HK97-type capsid protein fold

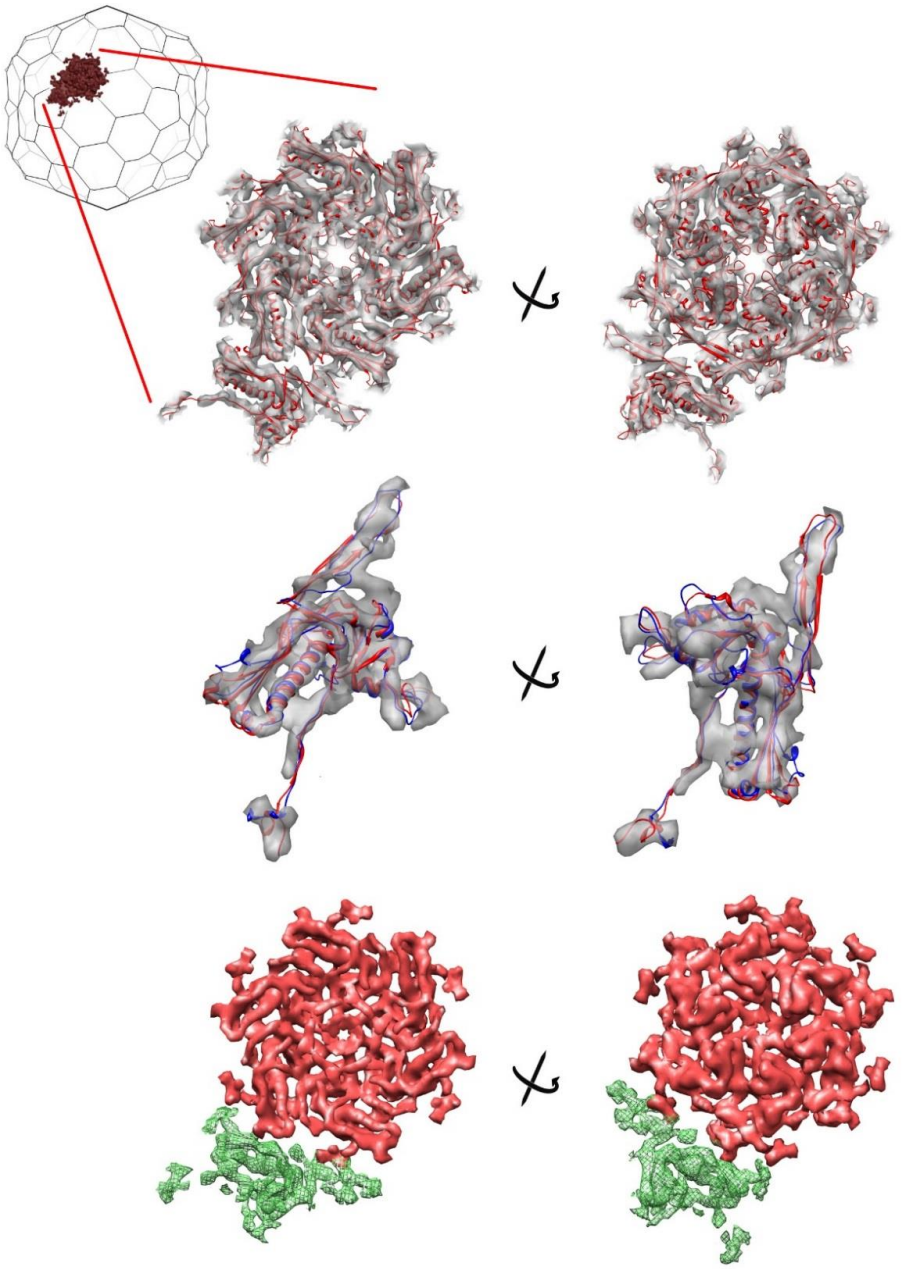


Figure 19. MCP of HSTV1 has a canonical coat protein fold of the bacterial virus Hong Kong 97 (HK97). *(Left)* view from capsid interior; *(Right)* view from capsid exterior. *(Top)* gray transparent: cryo-EM density of HSTV-1 capsid asymmetric unit, red ribbon: atomic model of HK97 asymmetric unit (Helgstrand et al. 2003). *(Middle)* gray transparent: averaged density of the six subunits making up the hexamer, red ribbon: HK97 chain E atomic model (Helgstrand et al. 2003), blue ribbon: HSTV-1 capsid protein homology model. *(Bottom)* red surface: capsid hexamer built from six copies of the averaged subunit. Green surface: the pentameric subunit that differs in conformation from the hexameric units and thus was not included in the averaging. Adapted with permission from (**III**).

In this study, Cryo-EM followed by SPA reconstruction and homology modelling produced the capsid structure of HSTV-1 at 8.9 Å resolution, solving for the first time the MCP fold of an archaeal head-tailed virus. First inspection of the reconstructed capsid density revealed a capsid lattice and MCP topology closely reminiscent of the bacterial virus HK97. It was possible to directly fit the atomic model of the HK97 capsid asymmetric unit (Helgstrand et al. 2003) into the corresponding reconstructed density of HSTV-1 (Figure 19, *top*). The densities corresponding to the six MCP monomers within a capsid hexamer were then computationally extracted and averaged together. This improved the definition of the MCP fold allowing us to verify the independently generated homology model of MCP (Figure 19, *middle*).

It has been proposed (Abrescia et al. 2012) that when looking at the MCP fold, the vast multitude of viruses can be reduced to a handful of MCP fold classes, one of them exemplified by the MCP fold of the tailed bacteriophage HK97 and including human herpesvirus (Baker et al. 2005). While genomic analysis of archaeal virus HSTV-1 showed some similarity with HK97 (**III**), the analysis of the MCP structure (Figure 19) revealed undisputable homology between the MCP folds of these evolutionary distant viruses, expanding the postulated HK97-like virus lineage (Abrescia et al. 2012) to all three domains of life. The experimental verification provided by the cryo-EM structure was pivotal as the sequence similarity alone was too low for reliable confirmation of the fold of HSTV-1 MCP. This result is in line with similar conclusions from study (**I**) but connects viruses over an even longer evolutionary distance providing structural background for e.g. theories about origin of viruses (Krupovic et al. 2019).

4.4. Structure of Nora virus capsid suggests a new virus family within the order *Picornavirales*

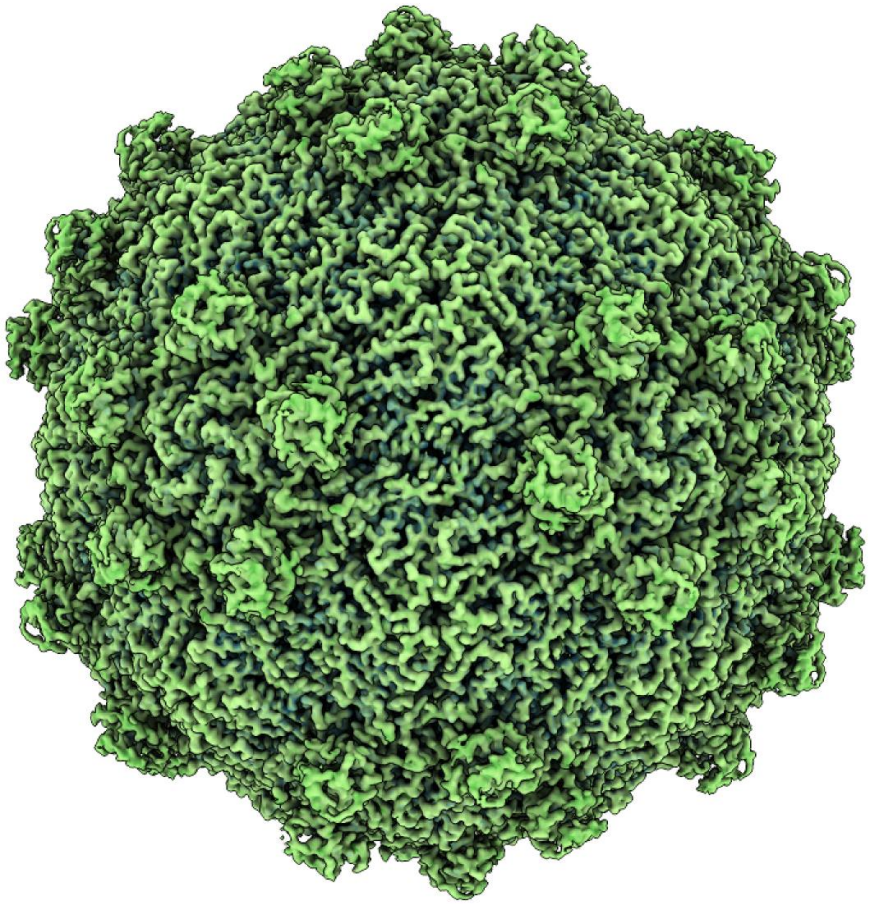


Figure 20. Nora virus cryo-EM reconstruction; isosurface representation of NORAV density map. Surface threshold three standard deviations above mean. Image courtesy of Shabih Shakeel.

In this study, Cryo-EM followed by SPA reconstruction produced the capsid structure of NORAV at a 2.7 Å resolution (Figure 20), allowing *de novo* atomic modelling of the MCPs (Figure 21). We were able to model residues 1-249 (out of 264) for VP4A, 2-242 (out of 251) for VP4B and 1-364 (out of 416) for VP4C. Analysis of the capsid structure revealed a *pseudo T*=3 capsid organisation, a hallmark of the *Picornavirale*s order. In terms of the virus structural lineage–hypothesis, the characteristic wedge-shaped β-jelly roll core domain of the MCPs (Figure 21) identifies NORAV as a member of the postulated picornavirus-like virus lineage (Abrescia et al. 2012).

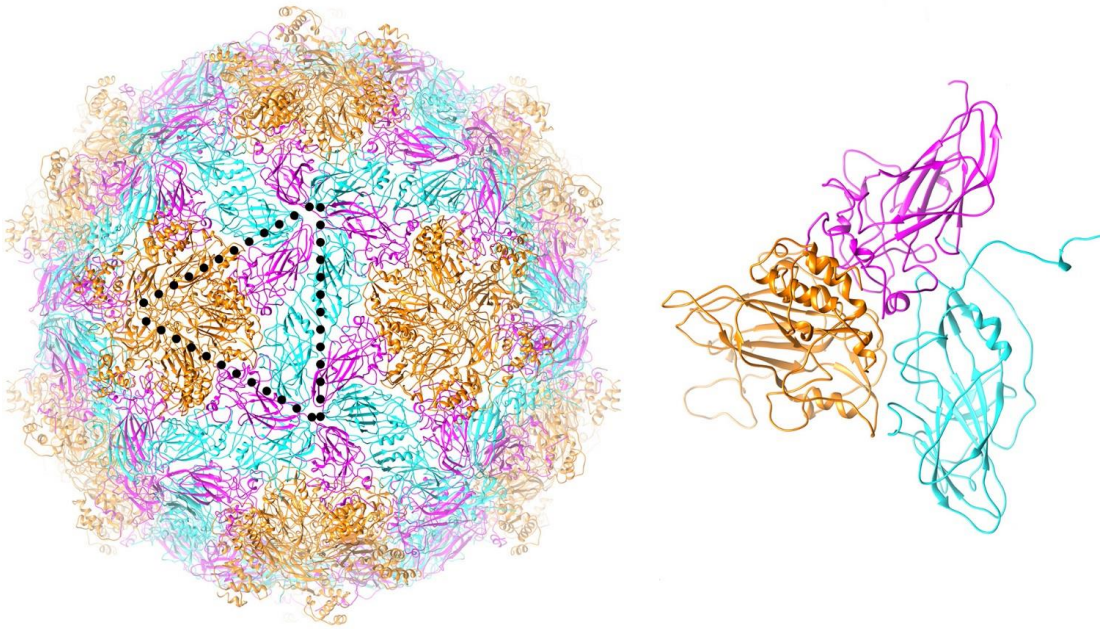


Figure 21. Atomic model of NORAV capsid (left) as a ribbon representation. One asymmetric unit is marked with dotted line and magnified (right). Cyan: VP4A; magenta: VP4B; and orange: VP4C.

In more detailed analysis, several notable features emerge (Figure 21). Firstly, C-terminal domains of VP4B and VP4C form prominent, mainly α-helical protrusions around the five-fold vertices of the NORAV capsid. Secondly, NORAV lacks the capsid – stabilising annulus formed by VP3 in dicistrovirus, picornavirus and iflavirus capsids. Thirdly, NORAV VP4A N-termini form a crossover across the two-fold symmetry axis while the symmetry-related VP4A α-helices are further apart than in other *Picornavirales*. The conformation of the MCPs allows a capsid inner volume that is largest reported in *Picornavirales*, a feature we link to the need to accommodate the unusually large 12kb genome of NORAV. Taken together, both global and detailed features of the capsid structure, in agreement with the genome organisation and genetic distance to other viruses, suggest that NORAV is a representative of a new virus family within the order *Picornavirales*.

4.5. The MCP folds facilitate reliable structure-based classification of the five studied viruses.

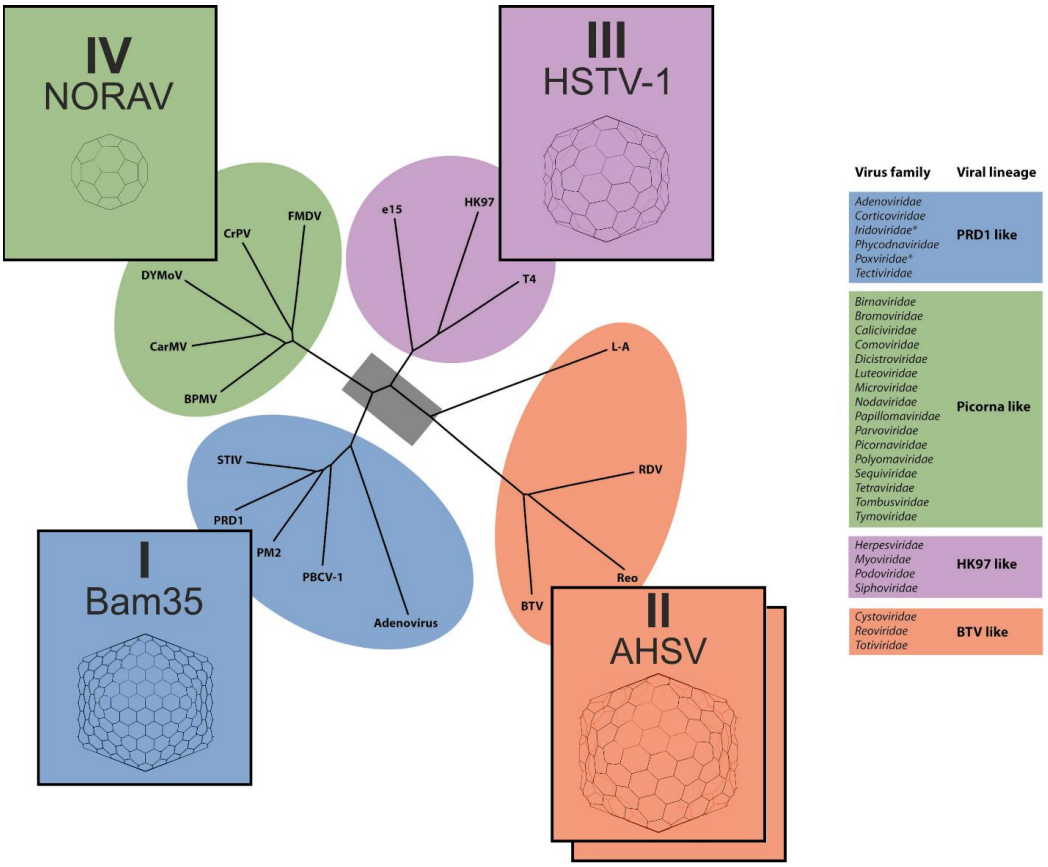


Figure 22. Capsid protein structures of the studied viruses in the context of capsid protein lineages. Figure adapted from (Abrescia et al. 2012) with permission of Annual Reviews.

Our results allow unambiguous structure-based classification of the five studied viruses into four virus lineages that have been postulated and are illustrated in figure 22 (Abrescia et al. 2012). The higher hierarchical levels of virus classification such as type of nucleic acid (Baltimore 1971) and capsid symmetry are well characterized in a relatively unambiguous manner (Lefkowitz et al. 2018), although deviations do exist, e.g. in the *Pleolipoviridae* family that includes both ssDNA and dsDNA viruses (Pietilä et al. 2016). At the lowest hierarchical levels such as species recognition and strain variation representing recent evolutionary events we can utilize very powerful and quantitative phylogenetic analyses. However, there is a large middle ground where we lack good quantitative classification tools. Looking from the bottom up we are essentially limited with the high mutation rates of viral genomes and the lack of a viral fossil record. The results reported in this thesis support the notion that the structures of viral capsid proteins are well conserved. In an attempt to detect connections between distantly-related viral species, a comparison of the coat protein fold is a robust

approach compared to just following the shorter-spanned evolutionary trace stored in the genomic nucleic acid sequence. Furthermore, we can create a powerful tool for studying virus evolution by combining structural data such as presented here with the analysis of genomic organisation (Krupovic et al. 2019).

The five capsid structures can be utilised as experimental results for testing theoretical approaches to, for example, mathematical principles that underpin the available icosahedral capsid lattice architecture. Looking at the reconstructed capsid lattices in the light of the seminal (Caspar and Klug 1962) theory, we see that HSTV-1 provides a biological representation of the original concept of quasi-equivalence, with the HSTV-1 MCP homopentamers and homohexamers populating the entire icosahedrally averaged capsid lattice with the aid of the predicted conformational flexibility. The *pseudo* lattices of Bam35 and NORAV represent biological alternatives to fulfilling the geometric lattice whereas the "T=2", or T=1 of dimers, in the inner layer of AHSV was not predicted by (Caspar and Klug 1962) and shows that generalisation of the original theory is required as was done in (Twarock 2004).

Mannige and Brooks introduced a division of theoretically available geometric *T*-numbers into three distinct capsid classes (Figure 23) (Mannige and Brooks 2010). Furthermore, they introduced a term "hexamer complexity" that predicts certain *T*-numbers to be evolutionally underrepresented (Figure 23, grey shading). It is interesting to note that our reconstructed capsid lattices sample all three capsid classes, but none are in the predicted underrepresented domain; a result in perfect agreement with the theory of Mannige and Brooks (2010).

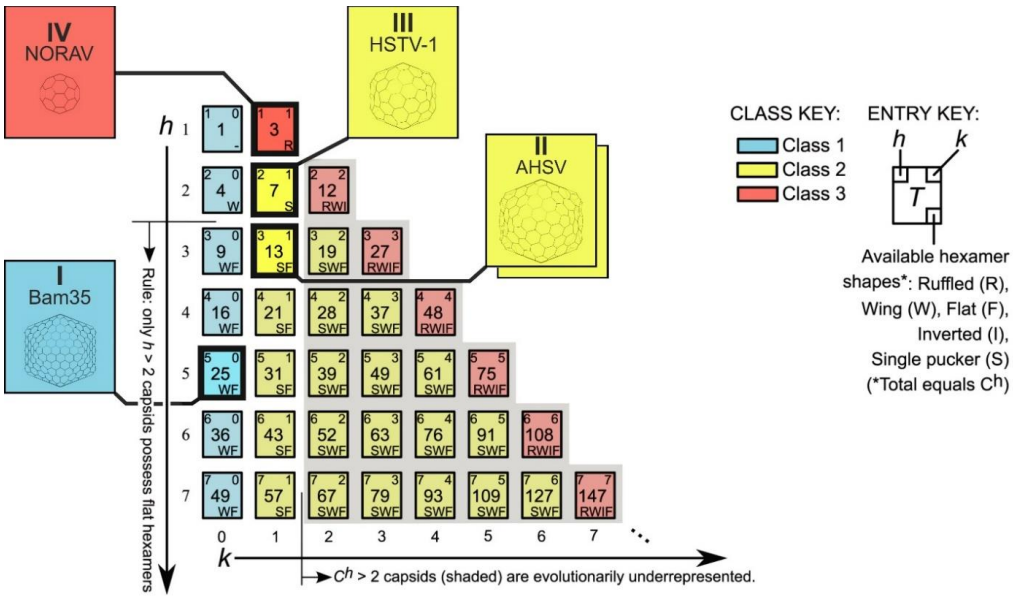


Figure 23. Capsid protein organisation of the studied viruses in the context of the hypothetical icosahedral capsid lattice classes. Figure adapted from (Mannige and Brooks 2010).

5. CONCLUDING REMARKS

Looking from the perspective of the cryo-EM technology we see that the timescale covered by this thesis is unusually long, the first paper was published in 2005 and technology has leaped forward in the 15 years that followed. It would be very interesting to see what could be uncovered in the viruses studied in (**I**, **II**, **III**) if they were studied with equipment and methods utilised in (**IV**) or e.g. (Dai and Zhou 2018; De Colibus et al. 2019; He et al. 2019; Ilca et al. 2019; Liu et al. 2019). Another fascinating idea is to extrapolate the same 15 years into the future; what would be my vision for cryo-EM science of viruses in 2035? Predicting the future is always difficult, even futile, but I feel safe in predicting further increase in automation, fully automated processing pipelines, automated analysis of structures, and constantly growing databases accessible to datamining algorithms utilising at least simple forms of artificial intelligence.

In my view, the current technology area in the highest need of development seems to be the sample preparation (Carragher et al. 2019). We need technology for more automated and more repeatable sample vitrification with better control over the conditions of the actual vitrification event. Along these lines, recent developments focusing on sample and grid preparation have been described (Earl et al. 2017). Furthermore, a promising new generation of vitrification devices utilising microfluidics are emerging (Ravelli et al. 2019; Schmidli et al. 2019).

The results presented in this thesis contribute to the field of structural virology as well as shedding light to more specific areas, such as biological membrane modulation, archaeal viruses and vaccine development. Moreover, the hosts of the viruses studied span all three domains of life (bacteria, archaea and eukaryotes). This unusually wide sampling of the viral universe, or *virosphere*, provides an excellent basis for testing the utilisation of capsid structure in structure-based virus classification as well as verification of structure-based viral lineages. The results allowed unambiguous structure-based classification of the five studied viruses into the four previously postulated virus lineages; the picorna-like, the HK97-like, the PRD1-like and the BTV-like. Furthermore, the NORAV reconstruction at 2.7Å resolution provided structural evidence suggesting that NORAV is a representative of a new virus family within the order *Picornavirales*; a result not achievable by genetic evidence alone and a benchmark example of structure-based virus classification.

6. ACKNOWLEDGEMENTS

The work for described in this thesis was carried out at the Institute of Biotechnology, University of Helsinki over a time period spanning from 2001 to 2020. This research as well as my doctoral studies (2013-2019) in the integrative life sciences doctoral program were conducted alongside my work as a technical support specialist for the Institute of Biotechnology cryo-EM core service. The research was supported by funding from the Academy of Finland including the Centre of Excellence Programme, project funding and infrastructure funding, the Sigrid Juselius Foundation and HiLIFE (to SJB). HiLIFE, Biocenter Finland, Instruct, eBic - Diamond UK and iNEXT are acknowledged for access to facilities.

First and foremost I want to express my deepest gratitude to professor Sarah Butcher. We met in the spring 2000 when I (with a fresh B.Eng. degree) was interviewed by her as I applied for an open research assistant's position at the institute. In the 20 years that have followed Sarah has been my supervisor, my mentor, a key support for this thesis project and a wonderful person. I have hard time imaging that any single person would have similar impact on my overall scientific career no matter where the future takes me.

I thank my thesis committee members Eija jokitalo and Tero Ahola as well as faculty professor Kari Keinänen for their helpful and supportive attitude.

I thank Linda Sandblad and professor Ilkka Julkunen for reviewing my thesis.

I thank all my coauthors over the years but especially the contributors to the thesis articles (in no particular order): Juha Huiskonen, Dennis Bamford, Violeta Manole, Wouter van Wyngaardt, Christiaan Potgieter, Isabella Wright, Gert Venter, Alberdina van Dijk, Trevor Sewell, Maija Pietilä, Daniel Russell, Ching-Chung Ko, Deborah Jacobs-Sera, Roger Hendrix, Shabih Shakeel, Jens-Ola Ekström, Pezhman Mohammadi and Dan Hultmark.

I thank the Butcher group alumni for all those wonderful discussions with and without beer back when I was a still a young man and the world was shiny and fresh out of the box. In this context I want to especially mention Juha Huiskonen, Harri Jäälinoja, Lotta Happonen, Heini Hyvönen, Violeta Manole, Jani Seitsonen, Lassi Liljeroos, Veli-Pekka Kestilä and Shabih Shakeel. Professionally, Juha and Harri taught me Linux and single particle data processing, Lassi taught me subtomogram averaging and Shabih taught me protein modelling from cryo-EM data. Thank you guys!

I thank the present members of the Butcher group for keeping up the warm and welcoming atmosphere, it is always nice to visit upstairs. Benita Löflund and Eevakaisa Vesanen are warmly thanked for flexible and efficient co-operation in our technical support tasks. I couldn't have hoped for better tech support team members!

The Institute of Biotechnology EM-Unit is thanked for help and co-operation; I want to especially bring up Eija Jokitalo, Mervi Lindman, and Helena Vihinen.

I thank my wonderful parents Heikki and Marja, as well as my brother Vesa for all the support and kind words over the years. Last but not least, I'd like to thank my wonderful wife Milla for support, good discussions, being my closest friend in this journey and making our home a place where it is always a pleasure to return. Our daughters Anni and Emma are thanked for bringing their happiness and energy as well as profound sense of meaning to my life.

Pasi Laurinmäki
Helsinki, April 2020

7. REFERENCES

Abdelnabi, R., J. A. Geraets, Y. P. Ma, C. Mirabelli, J. W. Flatt, A. Domanska, L. Delang, D. Jochmans, T. A. Kumar, V. Jayaprakash, B. N. Sinha, P. Leyssen, S. J. Butcher and J. Neyts (2019). "A novel druggable interprotomer pocket in the capsid of rhino- and enteroviruses." PLOS Biology **17**: e3000281.

Abrescia, N. G., D. H. Bamford, J. M. Grimes and D. I. Stuart (2012). "Structure unifies the viral universe." Annual Review of Biochemistry **81**: 795-822.

Abrescia, N. G. A., J. J. B. Cockburn, J. M. Grimes, G. C. Sutton, J. M. Diprose, S. J. Butcher, S. D. Fuller, C. S. Martin, R. M. Burnett, D. I. Stuart, D. H. Bamford and J. K. H. Bamford (2004). "Insights into assembly from structural analysis of bacteriophage PRD1." Nature **432**: 68-74.

Ackermann, H. W., R. Roy, M. Martin, M. R. V. Murthy and W. A. Smirnoff (1978). "Partial characterization of a cubic bacillus phage." Canadian Journal of Microbiology **24**: 986-993.

Adams, P. D., P. V. Afonine, G. Bunkoczi, 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 and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**: 213-221.

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

Aksular, M., E. Calvo-Pinilla, A. Marin-Lopez, J. Ortego, A. C. Chambers, L. A. King and J. Castillo-Olivares (2018). "A single dose of African horse sickness virus (AHSV) VP2 based vaccines provides complete clinical protection in a mouse model." Vaccine **36**: 7003-7010.

Atanasova, N. S., E. Roine, A. Oren, D. H. Bamford and H. M. Oksanen (2012). "Global network of specific virus-host interactions in hypersaline environments." Environmental Microbiology **14**: 426-440.

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

Baker, T. S. and R. H. Cheng (1996). "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy." Journal of Structural Biology **116**: 120-130.

Baker, T. S., J. Drak and M. Bina (1989). "The capsid of small papova viruses contains 72 pentameric capsomeres - direct evidence from cryo-electron-microscopy of Simian Virus-40." Biophysical Journal **55**: 243-253.

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

Baltimore, D. (1971). "Expression of animal virus genomes." Bacteriological Reviews **35**: 235-.

Bammes, B. E., R. H. Rochat, J. Jakana, D. H. Chen and W. Chiu (2012). "Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency." Journal of Structural Biology **177**: 589-601.

Barad, B. A., N. Echols, R. Y. Wang, Y. Cheng, F. DiMaio, P. D. Adams and J. S. Fraser (2015). "EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy." Nature Methods **12**: 943-946.

Basak, A. K., P. Gouet, J. Grimes, P. Roy and D. Stuart (1996). "Crystal structure of the top domain of African horse sickness virus VP7: Comparisons with bluetongue virus VP7." Journal of Virology **70**: 3797-3806.

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

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

Berman, H., K. Henrick and H. Nakamura (2003). "Announcing the worldwide Protein Data Bank." Nature Structural & Molecular Biology **10**: 980.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Research **28**: 235-242.

Booy, F. P. and J. B. Pawley (1993). "Cryo-crinkling - what happens to carbon-films on copper grids at low-temperature." Ultramicroscopy **48**: 273-280.

Borries, B. v. and E. Ruska (1939). "Ein übermikroskop für forschungsintitute." Naturwissenschaften **27**: 577-582.

Burrage, T. G., R. Trevejo, M. Stonemarschat and W. W. Laegreid (1993). "Neutralizing epitopes of African horsesickness virus serotype-4 are located on VP2." Virology **196**: 799-803.

Carpenter, S., P. S. Mellor, A. G. Fall, C. Garros and G. J. Venter (2017). "African horse sickness virus: History, transmission, and current status." Annual Review of Entomology **62**: 343-358.

Carragher, B., Y. Cheng, A. Frost, R. M. Glaeser, G. C. Lander, E. Nogales and H. W. Wang (2019). "Current outcomes when optimizing 'standard' sample preparation for single-particle cryo-EM." Journal of Microscopy **276**: 39-45.

Caspar, D. L. and A. Klug (1962). "Physical principles in the construction of regular viruses." Cold Spring Harbour Symposia on Quantitative Biology **27**: 1-24.

Castillo-Olivares, J., E. Calvo-Pinilla, I. Casanova, K. Bachanek-Bankowska, R. Chiam, S. Maan, J. M. Nieto, J. Ortego and P. P. C. Mertens (2011). "A modified vaccinia Ankara virus (MVA) vaccine expressing African horse sickness virus (AHSV) VP2 protects against AHSV challenge in an IFNAR -/- mouse model." PLOS One **6**: e16503.

Chen, J. Z., C. Sachse, C. Xu, T. Mielke, C. M. T. Spahn and N. Grigorieff (2008). "A dose-rate effect in single-particle electron microscopy." Journal of Structural Biology **161**: 92-100.

Cheng, Y., R. M. Glaeser and E. Nogales (2017). "How cryo-EM became so hot." Cell **171**: 1229-1231.

Cheng, Y. F. (2015). "Single-particle cryo-EM at crystallographic resolution." Cell **161**: 450-457.

Cheng, Y. F., E. Wolf, M. Larvie, O. Zak, P. Aisen, N. Grigorieff, S. C. Harrison and T. Walz (2006). "Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques." Journal of Molecular Biology **355**: 1048-1065.

Chow, C. E. T. and C. A. Suttle (2015). "Biogeography of viruses in the sea." Annual Review of Virology **2**: 41-66.

Cockburn, J. J. B., N. G. A. Abrescia, J. M. Grimes, G. C. Sutton, J. M. Diprose, J. M. Benevides, G. J. Thomas, J. K. H. Bamford, D. H. Bamford and D. I. Stuart (2004). "Membrane structure and interactions with protein and DNA in bacteriophage PRD1." Nature **432**: 122-125.

Dai, X. H. and Z. H. Zhou (2018). "Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes." Science **360**: eaao7298.

Danovaro, R., A. Dell'Anno, C. Corinaldesi, E. Rastelli, R. Cavicchioli, M. Krupovic, R. T. Noble, T. Nunoura and D. Prangishvili (2016). "Virus-mediated archaeal hecatomb in the deep seafloor." Science Advances **2**: e1600492.

De Colibus, L., E. Roine, T. S. Walter, S. L. Ilca, X. X. Wang, N. Wang, A. M. Roseman, D. Bamford, J. T. Huiskonen and D. I. Stuart (2019). "Assembly of complex viruses exemplified by a halophilic euryarchaeal virus." Nature Communications **10:** 1456.

de la Rosa-Trevin, J. M., A. Quintana, L. Del Cano, A. Zaldivar, I. Foche, J. Gutierrez, J. Gomez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Oton, G. Sharov, J. L. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. O. Sorzano and J. M. Carazo (2016). "Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy." Journal of Structural Biology **195**: 93-99.

DeRosier, D. J. and P. B. Moore (1970). "Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry." Journal of Molecular Biology **52**: 355-369.

Domanska, A., J. W. Flatt, J. J. J. Jukonen, J. A. Geraets and S. J. Butcher (2019). "A 2.8-Ångstrom-resolution cryo-electron microscopy structure of human parechovirus 3 in complex with Fab from a neutralizing antibody." Journal of Virology **93**: e01597-18.

Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall and P. Schultz (1988). "Cryo-electron microscopy of vitrified specimens." Quarterly Reviews of Biophysics **21**: 129-228.

Earl, L. A., V. Falconieri, J. L. Milne and S. Subramaniam (2017). "Cryo-EM: beyond the microscope." Current Opinion in Structural Biology **46**: 71-78.

Elmlund, D., S. N. Le and H. Elmlund (2017). "High-resolution cryo-EM: the nuts and bolts." Current Opinion in Structural Biology **46**: 1-6.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**: 2126-2132.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**: 486-501.

Frank, J. (2017). "Advances in the field of single-particle cryo-electron microscopy over the last decade." Nature Protocols **12**: 209-212.

Glaeser, R. M. (2016). "Specimen behavior in the electron beam." Methods in Enzymology **579**: 19-50.

Goddard, T. D., C. C. Huang and T. E. Ferrin (2007). "Visualizing density maps with UCSF Chimera." Journal of Structural Biology **157**: 281-287.

Grant, T., A. Rohou and N. Grigorieff (2018). "cisTEM, user-friendly software for single-particle image processing." Elife **7:** e35383.

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

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

Guemes, A. G. C., M. Youle, V. A. Cantu, B. Felts, J. Nulton and F. Rohwer (2016). "Viruses as winners in the game of life." Annual Review of Virology **3**: 197-214.

Habayeb, M. S., R. Cantera, G. Casanova, J. O. Ekstrom, S. Albright and D. Hultmark (2009). "The Drosophila Nora virus is an enteric virus, transmitted via feces." Journal of Invertebrate Pathology **101**: 29-33.

Habayeb, M. S., S. K. Ekengren and D. Hultmark (2006). "Nora virus, a persistent virus in Drosophila, defines a new picorna-like virus family." Journal of General Virology **87**: 3045-3051.

Harauz, G. and M. van Heel (1986). "Exact filters for general geometry 3-dimensional reconstruction." Optik **73**: 146-156.

Harris, J. R. (2015). "Transmission electron microscopy in molecular structural biology: A historical survey." Archives of Biochemistry and Biophysics **581**: 3-18.

He, Y., S. Shivakoti, K. Ding, Y. X. Cui, P. Roy and Z. H. Zhou (2019). "In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating." Proceedings of the National Academy of Sciences of the United States of America **116**: 16535-16540.

Helgstrand, C., W. R. Wikoff, R. L. Duda, R. W. Hendrix, J. E. Johnson and L. Liljas (2003). "The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution." Journal of Molecular Biology **334**: 885-899.

Henderson, R. (2013). "Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise." Proceedings of the National Academy of Sciences of the United States of America **110**: 18037-18041.

Henderson, R. (2015). "Overview and future of single particle electron cryomicroscopy." Archives of Biochemistry and Biophysics **581**: 19-24.

Henderson, R., A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schroder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt and C. L. Lawson (2012). "Outcome of the first electron microscopy validation task force meeting." Structure **20**: 205-214.

Hewat, E. A., T. F. Booth and P. Roy (1992). "Structure of bluetongue virus particles by cryoelectron microscopy." Journal of Structural Biology **109**: 61-69.

Heymann, J. B. and D. M. Belnap (2007). "Bsoft: image processing and molecular modeling for electron microscopy." Journal of Structural Biology **157**: 3-18.

Heymann, J. B., R. Marabini, M. Kazemi, C. O. S. Sorzano, M. Holmdahl, J. H. Mendez, S. M. Stagg, S. Jonic, E. Palovcak, J. P. Armache, J. H. Zhao, Y. F. Cheng, G. Pintilie, W. Chiu, A. Patwardhan and J. M. Carazo (2018). "The first single particle analysis map challenge: a summary of the assessments." Journal of Structural Biology **204**: 291-300.

Hohn, M., G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams and S. J. Ludtke (2007). "SPARX, a new environment for cryo-EM image processing." Journal of Structural Biology **157**: 47-55.

Holmes, M. L. and M. L. Dyall-Smith (1990). "A plasmid vector with a selectable marker for halophilic archaebacteria." Journal of Bacteriology **172**: 756-761.

Huiskonen, J. T., V. Manole and S. J. Butcher (2007). "Tale of two spikes in bacteriophage PRD1." Proceedings of the National Academy of Sciences of the United States of America **104**: 6666-6671.

Ilca, S. L., X. Y. Sun, K. El Omari, A. Kotecha, F. de Haas, F. DiMaio, J. M. Grimes, D. I. Stuart, M. M. Poranen and J. T. Huiskonen (2019). "Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus." Nature **570**: 252-256.

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

Jiang, W. and L. Tang (2017). "Atomic cryo-EM structures of viruses." Current Opinion in Structural Biology **46**: 122-129.

Jäälinoja, H. T., E. Roine, P. Laurinmäki, H. M. Kivela, D. H. Bamford and S. J. Butcher (2008). "Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus." Proceedings of the National Academy of Sciences of the United States of America **105**: 8008-8013.

Kelly, D. F., D. Dukovski and T. Walz (2010). "Strategy for the use of affinity grids to prepare non-His-tagged macromolecular complexes for single-particle electron microscopy." Journal of Molecular Biology **400**: 675-681.

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

Koonin, E. V. and V. V. Dolja (2013). "A virocentric perspective on the evolution of life." Current Opinion in Virology **3**: 546-557.

Koonin, E. V. and V. V. Dolja (2014). "Virus world as an evolutionary network of viruses and capsidless selfish elements." Microbiology and Molecular Biology Reviews **78**: 278-303.

Krupovic, M., V. V. Dolja and E. V. Koonin (2019). "Origin of viruses: primordial replicators recruiting capsids from hosts." Nature Reviews Microbiology **17**: 449-458.

Kryshtafovych, A., P. D. Adams, C. L. Lawson and W. Chiu (2018). "Evaluation system and web infrastructure for the second cryo-EM model challenge." Journal of Structural Biology **204**: 96-108.

Kucukelbir, A., F. J. Sigworth and H. D. Tagare (2014). "Quantifying the local resolution of cryo-EM density maps." Nature Methods **11**: 63-65.

Kuhlbrandt, W. (2014). "Biochemistry. The resolution revolution." Science **343**: 1443-1444.

Kuhlbrandt, W. (2014). "Cryo-EM enters a new era." Elife **3**: e03678.

Kuhn, J. H., Y. I. Wolf, M. Krupovic, Y. Z. Zhang, P. Maes, V. V. Dolja and E. V. Koonin (2019). "Classify viruses - the gain is worth the pain." Nature **566**: 318-320.

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

Lawson, C. L., M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman and W. Chiu (2011). "EMDataBank.org: unified data resource for CryoEM." Nucleic Acids Research **39**: D456-464.

Lawson, C. L., A. Patwardhan, M. L. Baker, C. Hryc, E. S. Garcia, B. P. Hudson, I. Lagerstedt, S. J. Ludtke, G. Pintilie, R. Sala, J. D. Westbrook, H. M. Berman, G. J. Kleywegt and W. Chiu (2016). "EMDataBank unified data resource for 3DEM." Nucleic Acids Research **44**: D396-403.

Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell and D. B. Smith (2018). "Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)." Nucleic Acids Research **46**: D708-D717.

Li, X., P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard and Y. Cheng (2013). "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM." Nature Methods **10**: 584-590.

Liu, Y. T., J. Jih, X. H. Dai, G. Q. Bi and Z. H. Zhou (2019). "Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome." Nature **570**: 257-261.

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

Mannige, R. V. and C. L. Brooks, 3rd (2010). "Periodic table of virus capsids: implications for natural selection and design." PLOS One **5**: e9423.

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

Martinez-Torrecuadrada, J. L. and J. I. Casal (1995). "Identification of a linear neutralization domain in the protein VP2 of African horse sickness virus." Virology **210**: 391-399.

Meiswinkel, R., P. van Rijn, P. Leijs and M. Goffredo (2007). "Potential new Culicoides vector of bluetongue virus in northern Europe." Veterinary Record **161**: 564-565.

Mellor, P. S. and C. Hamblin (2004). "African horse sickness." Veterinary Research **35**: 445-466.

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

Mertens, P. P., J. Diprose, S. Maan, K. P. Singh, H. Attoui and A. R. Samuel (2004). "Bluetongue virus replication, molecular and structural biology." Veterinaria Italiana **40**: 426-437.

Mindell, J. A. and N. Grigorieff (2003). "Accurate determination of local defocus and specimen tilt in electron microscopy." Journal of Structural Biology **142**: 334-347.

Moriya, T., M. Saur, M. Stabrin, F. Merino, H. Voicu, Z. Huang, P. A. Penczek, S. Raunser and C. Gatsogiannis (2017). "High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE." Journal of Visualized Experiments (123): 55448.

Murata, K. and M. Wolf (2018). "Cryo-electron microscopy for structural analysis of dynamic biological macromolecules." Biochimica et Biophysica Acta General Subjects **1862**: 324-334.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**: 240-255.

Naydenova, K., M. J. Peet and C. J. Russo (2019). "Multifunctional graphene supports for electron cryomicroscopy." Proceedings of the National Academy of Sciences of the United States of America **116**: 11718-11724.

Orlova, E. V. and H. R. Saibil (2011). "Structural analysis of macromolecular assemblies by electron microscopy." Chemical Reviews **111**: 7710-7748.

Palovcak, E., F. Wang, S. Q. Zheng, Z. Yu, S. Li, M. Betegon, D. Bulkley, D. A. Agard and Y. Cheng (2018). "A simple and robust procedure for preparing graphene-oxide cryo-EM grids." Journal of Structural Biology **204**: 80-84.

Pantelic, R. S., J. C. Meyer, U. Kaiser, W. Baumeister and J. M. Plitzko (2010). "Graphene oxide: a substrate for optimizing preparations of frozen-hydrated samples." Journal of Structural Biology **170**: 152-156.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." Journal of Computational Chemistry **25**: 1605-1612.

Pietilä, M. K., P. Laurinmäki, D. A. Russell, C. C. Ko, D. Jacobs-Sera, S. J. Butcher, D. H. Bamford and R. W. Hendrix (2013). "Insights into head-tailed viruses infecting extremely halophilic archaea." Journal of Virology **87**: 3248-3260.

Pietilä, M. K., E. Roine, A. Sencilo, D. H. Bamford and H. M. Oksanen (2016). "Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes." Archives of Virology **161**: 249-256.

Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nature Methods **14**: 290-296.

Purse, B. V., P. S. Mellor, D. J. Rogers, A. R. Samuel, P. P. C. Mertens and M. Baylis (2005). "Climate change and the recent emergence of bluetongue in Europe." Nature Reviews Microbiology **3**: 171-181.

Quispe, J., J. Damiano, S. E. Mick, D. P. Nackashi, D. Fellmann, T. G. Ajero, B. Carragher and C. S. Potter (2007). "An improved holey carbon film for cryo-electron microscopy." Microscopy and Microanalysis **13**: 365-371.

Rath, B. K., R. Hegerl, A. Leith, T. R. Shaikh, T. Wagenknecht and J. Frank (2003). "Fast 3D motif search of EM density maps using a locally normalized cross-correlation function." Journal of Structural Biology **144**: 95-103.

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

Ravelli, R. B. G., F. J. T. Nijpels, R. J. M. Henderikx, G. Weissenberger, S. Thewessem, A. Gijsbers, B. W. A. M. M. Beulen, C. López-Iglesias and P. J. Peters (2019). "Automated cryo-EM sample preparation by pin-printing and jet vitrification." bioRxiv :651208.

Reddy, H. K. N., M. Carroni, J. Hajdu and M. Svenda (2019). "Electron cryo-microscopy of bacteriophage PR772 reveals the elusive vertex complex and the capsid architecture." Elife **8**.

Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." Journal of Structural Biology **192**: 216-221.

Rosenthal, P. B. and R. Henderson (2003). "Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy." Journal of Molecular Biology **333**: 721-745.

Roy, A., A. Kucukural and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nature Protocols **5**: 725-738.

Roy, P., P. P. C. Imertens and I. Casal (1994). "African horse sickness virus structure." Comparative Immunology Microbiology and Infectious Diseases **17**: 243-273.

Ruokolainen, V., A. Domanska, M. Laajala, M. Pelliccia, S. J. Butcher and V. Marjomäki (2019). "Extracellular albumin and endosomal ions prime enterovirus particles for uncoating that can be prevented by fatty acid saturation." Journal of Virology **93:** e00599-19.

Ruska, H. B., Bodo v.; Ruska, E. (1939). "Die bedeutung der übermikroskopie für die virusforschung." Archiv für die gesamte Virusforschung **1**: 155-169.

Russo, C. J. and L. A. Passmore (2014). "Electron microscopy: ultrastable gold substrates for electron cryomicroscopy." Science **346**: 1377-1380.

Russo, C. J. and L. A. Passmore (2016). "Progress towards an optimal specimen support for electron cryomicroscopy." Current Opinion in Structural Biology **37**: 81-89.

San Martin, C. (2015). "Transmission electron microscopy and the molecular structure of icosahedral viruses." Archives of Biochemistry and Biophysics **581**: 59-67.

Scanlen, M., J. T. Paweska, J. A. Verschoor and A. A. van Dijk (2002). "The protective efficacy of a recombinant VP2-based African horsesickness subunit vaccine candidate is determined by adjuvant." Vaccine **20**: 1079-1088.

Scheres, S. H. (2012). "A Bayesian view on cryo-EM structure determination." Journal of Molecular Biology **415**: 406-418.

Scheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." Journal of Structural Biology **180**: 519-530.

Scheres, S. H. (2016). "Processing of structurally heterogeneous cryo-EM data in RELION." Methods in Enzymology **579**: 125-157.

Scheres, S. H., R. Nunez-Ramirez, C. O. Sorzano, J. M. Carazo and R. Marabini (2008). "Image processing for electron microscopy single-particle analysis using XMIPP." Nature Protocols **3**: 977-990.

Schmidli, C., S. Albiez, L. Rima, R. Righetto, I. Mohammed, P. Oliva, L. Kovacik, H. Stahlberg and T. Braun (2019). "Microfluidic protein isolation and sample preparation for high-resolution cryo-EM." Proceedings of the National Academy of Sciences of the United States of America **116**: 15007-15012.

Seitsonen, J., P. Susi, O. Heikkilä, R. S. Sinkovits, P. Laurinmäki, T. Hyypiä and S. J. Butcher (2010). "Interaction of alpha(V)beta(3) and alpha(V)beta(6) integrins with human parechovirus 1." Journal of Virology **84**: 8509-8519.

Seitsonen, J. J. T., S. Shakeel, P. Susi, A. P. Pandurangan, R. S. Sinkovits, H. Hyvönen, P. Laurinmäki, J. Ylä-Pelto, M. Topf, T. Hyypiä and S. J. Butcher (2012). "Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating." Journal of Virology **86**: 7207-7215.

Shakeel, S., J. J. T. Seitsonen, T. Kajander, P. Laurinmäki, T. Hyypiä, P. Susi and S. J. Butcher (2013). "Structural and functional analysis of coxsackievirus A9 integrin alpha(v)beta(6) binding and uncoating." Journal of Virology **87**: 3943-3951.

Sorzano, C. O. S., R. Marabini, J. Velazquez-Muriel, J. R. Bilbao-Castro, S. H. W. Scheres, J. M. Carazo and A. Pascual-Montano (2004). "XMIPP: a new generation of an open-source image processing package for electron microscopy." Journal of Structural Biology **148**: 194-204.

Sorzano, C. O. S., J. Vargas, J. M. de la Rosa-Trevin, A. Jimenez, D. Maluenda, R. Melero, M. Martinez, E. Ramirez-Aportela, P. Conesa, J. L. Vilas, R. Marabini and J. M. Carazo (2018). "A new algorithm for high-resolution reconstruction of single particles by electron microscopy." Journal of Structural Biology **204**: 329-337.

Stuart, D. I., P. Gouet, J. Grimes, R. Malby, J. Diprose, S. Zientara, J. N. Burroughs and P. P. C. Mertens (1998). "Structural studies of orbivirus particles." Archives of Virology: 235-250.

Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. J. Ludtke (2007). "EMAN2: an extensible image processing suite for electron microscopy." Journal of Structural Biology **157**: 38-46.

Terwilliger, T. C., P. D. Adams, P. V. Afonine and O. V. Sobolev (2018). "Map segmentation, automated model-building and their application to the Cryo-EM Model Challenge." Journal of Structural Biology **204**: 338-343.

Thiry, E., C. Saegerman, H. Guyot, P. Kirten, B. Losson, F. Rollin, M. Bodmer, G. Czaplicki, J. F. Toussaint, K. De Clercq, J. M. Dochy, J. Dufey, J. L. Gilleman and K. Messeman (2006). "Bluetongue in northern Europe." Veterinary Record **159**: 327-327.

Twarock, R. (2004). "A tiling approach to virus capsid assembly explaining a structural puzzle in virology." Journal of Theoretical Biology **226**: 477-482.

van Heel, M., G. Harauz, E. V. Orlova, R. Schmidt and M. Schatz (1996). "A new generation of the IMAGIC image processing system." Journal of Structural Biology **116**: 17-24.

van Heel, M. and M. Schatz (2005). "Fourier shell correlation threshold criteria." Journal of Structural Biology **151**: 250-262.

Venter, G. J., S. D. Graham and C. Hamblin (2000). "African horse sickness epidemiology: vector competence of South African Culicoides species for virus serotypes 3, 5 and 8." Medical and Veterinary Entomology **14**: 245-250.

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

Vilas, J. L., J. Gomez-Blanco, P. Conesa, R. Melero, J. M. de la Rosa-Trevin, J. Oton, J. Cuenca, R. Marabini, J. M. Carazo, J. Vargas and C. O. S. Sorzano (2018). "MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps." Structure **26**: 337-344.

Yan, X., R. S. Sinkovits and T. S. Baker (2007). "AUTO3DEM--an automated and high throughput program for image reconstruction of icosahedral particles." Journal of Structural Biology **157**: 73-82.

Yan, X. D., K. A. Dryden, J. H. Tang and T. S. Baker (2007). "Ab initio random model method facilitates 3D reconstruction of icosahedral particles." Journal of Structural Biology **157**: 211-225.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang (2015). "The I-TASSER Suite: protein structure and function prediction." Nature Methods **12**: 7-8.

Yonekura, K., M. B. Braunfeld, S. Maki-Yonekura and D. A. Agard (2006). "Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300 kV." Journal of Structural Biology **156**: 524-536.

Yu, G., K. Li, P. Huang, X. Jiang and W. Jiang (2016). "Antibody-based affinity cryoelectron microscopy at 2.6-Å resolution." Structure **24**: 1984-1990.

Zhang, K. (2016). "Gctf: Real-time CTF determination and correction." Journal of Structural Biology **193**: 1-12.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics **9**: 40.

Zientara, S. and J. M. Sanchez-Vizcaino (2013). "Control of bluetongue in Europe." Veterinary Microbiology **165**: 33-37.

Zivanov, J., T. Nakane and S. H. W. Scheres (2019). "A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis." IUCrJ **6**: 5-17.