Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Julia Peukes, M. Sc.
born in Krefeld
Oral examination: 06.11.2019

Structural studies of influenza A virus by cryo-electron tomography

Referees: Dr. Martin Beck

Prof. Dr. Hans-Georg Kräusslich

I dedicate this thesis to my parents.

“I had a little bird, it’s name was Enza,
I opened the window and in flu enza.”

Children’s Skipping Rhyme, 1918,
the year of the devastating 1918 "Spanish Flu" pandemic.

Summary

Influenza A virus (IAV) is a pleomorphic, enveloped virus known for its yearly epidemics and occasional, but fatal pandemics. The outer surface glycoprotein hemagglutinin (HA) together with the matrix protein 1 (M1) are the most abundant protein components of assembled virions. HA, located at the outside of virions, is involved in cell receptor recognition, membrane fusion and is the most relevant protein for antibody binding. Therefore the structure of isolated HA has been extensively characterised by X-ray crystallography. However, it remains unclear to which extent the structure of isolated HA corresponds to the in situ HA structure on the surface of IAV. M1 determines the morphology of the virus by forming a matrix layer underneath the viral membrane. A high resolution structure of full length M1 is missing and the lack of information about the in situ arrangement of the M1 matrix layer currently limits our understanding of how M1 functions.

Here, I set out to determine the structures of HA and M1 directly from IAV particles using high resolution cryo-electron tomography (cryoET) and subtomogram averaging. I found that virus purification can affect the integrity of the virus HA glycoprotein layer and the morphology of virus particles. I therefore adapted a workflow which allows studying the structure of viral proteins directly from viruses in the vicinity of virus-producing cells. Biosafety regulations required inactivation of IAV samples by chemical fixation prior to cryoEM imaging. To assess effects of fixation, I complemented structural studies of HA from pathogenic, fixed IAV particles with studies of HA from non-infectious, unfixed virus-like particles (VLPs). These studies revealed that fixation captures HA in an open conformation while HA structures determined from unfixed samples perfectly match the closed conformation observed in the trimeric crystal structure. In concordance with recent work by others, this observation suggests that fixation captures HA in a an open, otherwise transient conformation, which is part of a constant opening and closing motion known as breathing motion.

To characterise the in situ structure and arrangement of M1, I established a subtomogram averaging workflow to cope with the challenges presented by the small size of M1. I successfully obtained two independent structures of M1 directly from viruses and VLPs. Comparisons of my structures to existing high resolution models of the N-terminal domain (NTD) of M1 revealed that M1 monomers arrange as parallel strands, with a helical propensity and directly underneath the membrane. For the first time, my data allow to describe the M1-membrane interface as well as relevant M1-M1 interfaces within the matrix layer. Finally, I have gained first structural insights into the M1 C-terminal domain (CTD). I further combined the obtained structural information for M1 with a theoretical model of the mechanics of M1 polymerization and membrane deformation during virus assembly. The obtained results suggest that linear polymerization of M1 into multiple parallel strands efficiently provides energy to drive assembly of new virus particles.

The results presented in this thesis improve our understanding of the arrangement and structure of the two influenza proteins HA and M1 in situ which has implications for current models of HA-mediated membrane fusion, virus architecture and virus assembly.

Zusammenfassung

Influenza A Virus (IAV) ist ein pleomorphes, behuelltes Virus, das für das Ausbrechen jährlicher Epidemien und seltener Pandemien verantwortlich ist. Das Glykoprotein Hemagglutinin (HA) und das Matrix Protein 1 (M1) machen den groessten Bestandteil des Virusteilchens aus. Das Oberflächenprotein HA erkennt zelluläre Rezeptoren und ist essenziell für die Membranfusion zwischen der Endosomenmembran der Wirtszelle und der Virusmembran. Da HA das wichtigste Protein für die Antikörperbildung ist, wurde die Struktur mehrfach mit Hilfe der Kristallstrukturanalyse bestimmt. Dennoch bleibt unverstanden, ob die Struktur des isolierten HA Proteins der Struktur des HA Proteins auf der Virenoberfläche entspricht. M1 formt eine Matrix unter der viralen Membrane und beeinflusst die Morphologie der Virusteilchen. Die Struktur des M1 Proteins ist noch nicht vollständig aufgeklärt. Die fehlende Charakterisierung der Struktur sowie der M1 Anordnung innerhalb des Virusteilchens erschwert es, ein besseres Verständnis der Funktionen des Proteins zu erlangen.

In dieser Arbeit habe ich mit Hilfe der Kryoelektronentomographie in Kombination mit einer Datenauswertung durch „Subtomogram Averaging“ die in situ Strukturen der Proteine HA und M1 direkt aus den Virusteilchen bestimmt. Ich beobachtete, dass die Aufreinigung der Virusteilchen zu Artefakten auf der Virusoberfläche und zu einer Veränderung der Virusmorphologie führen kann. Daher habe ich eine Methode optimiert, die ohne die Aufreinigung der Virusteilchen auskommt. Auf Grund der geltenden Sicherheitsbestimmungen mussten IAV Proben vor dem Mikroskopieren durch eine chemische Fixierung mit Formaldehyd deaktiviert werden. Um mögliche Effekte der Fixierung auf die Probe und die Struktur des HA Proteins zu evaluieren, habe ich zusätzliche Proben mit nicht infektiösen Virus-ähnlichen Partikeln (VLP) ohne Inaktivierung präpariert. Ein direkter Vergleich der erhaltenden HA Strukturen hat gezeigt, dass die Fixierung eine offene Konformation des HA Proteins stabilisiert, während die HA Struktur nicht fixierter VLPs der geschlossenen Konformation existierender Kristallstrukturen gleicht. Die offenere HA Konformation wurde kürzlich auch von anderen beobachtet und die gezeigten Ergebnisse unterstützen die Hypothese, dass sich HA auf der Virenoberflaeche kontinuierlich öffnet und schliesst, was auch als "Atmen" des Proteins bezeichnet wird.

Für eine strukturelle in situ Charakterisierung des M1 Proteins wurden die Parameter des Subtomogram Averaging Ablaufes speziell auf die geringe Proteingröße von M1 angepasst. So habe ich sowohl eine Struktur des M1 Proteins aus Viren als auch aus VLPs bestimmen können. Ein Vergleich der in situ M1 Struktur mit höher aufgelösten Proteinkristallstrukturen der M1 N-terminalen Domäne (NTD) ergab, dass sich im Virus mehrere M1 Monomere in parallelen Strängen mit helikalem Verlauf an der Innenseite der Membran anordnen. Dank der in situ Struktur kann ich zum ersten Mal die Schnittstelle zwischen M1 und der viralen Membran sowie die bisher unbekannte Struktur der M1 C-terminalen Domäne charakterisieren. Diese strukturellen Studien wurde mit einem theoretischen Model kombiniert, das die Bildung neuer Virusteilchen und die Polymerisation des M1 Proteins in Verbindung setzt. Der Vergleich des Models mit experimentellen Daten ergab, dass die lineare M1 Polymerisation ausreichend Energie zur Bildung neuer Viren freisetzt.

Zusammenfassend tragen die vorgestellten Ergebnisse betreffend der Anordnung und der in situ Strukturen der Influenzavirenproteine HA und M1 zu einem besseren Verständnis der HA induzierten Membranfusion, der Virusteilchenarchitektur und der Virenassemblierung bei.

Acknowledgements

First of all, I would like to thank my supervisor John Briggs for giving me the opportunity to start a PhD in his group and to immerse into the world of cryoEM, cryoET and subtomogram averaging. I would like to thank John for his guidance during my PhD, many helpful advice and inspiring discussions. I have learned a lot.

I further would like to acknowledge my TAC members Hans-Georg Kräusslich, Christian Loew and Toby Gibson for their time and their helpful and encouraging comments. I also would like to thank my additional examiners Martin Beck, Alba Diz-Muños and Jochen Reinstein.

I am grateful to Hans-Georg Kräusslich for giving me the chance to perform S2 experiments in his lab and at the CIID imaging facility and for providing virus stocks and cells. In this context I also would like to thank Martin Beck for agreeing to be the S2 responsible for my S2 EM work at EMBL in 2018/19. I further would like to thank Wim Hagen for creating such a beautiful cryoET set-up at EMBL and his support, help files, trainings and answers to questions. I would like to thank Serge Dmitrieff for a fruitful and long lasting collaboration as part of the M1 project.

I would like to acknowledge all former and current members of the Briggs group at EMBL and LMB for creating a stimulating, friendly and productive scientific environment: Simo, Florian, Will, Ori, Mairi, Sveta, Cecilia, Andreas, Yury, Oleksi, Lauren Ann, Beata, Aaron, Kun, Dustin, Lucy, Kino, Simon, Alex, Zunlong, Kat. I would like to offer my special thanks to Zunlong for being such a fun and positive desk mate and for helping out with HA subtomogram averaging. I further would like to acknowledge Simon and Alex for help with protein purification and crystallization. I am in particular grateful for the assistance at the beginning of my PhD given by Simo who introduced me to subtomogram averaging. I would like to thank Simo, Sveta, Flo, Andreas and in particular Will for answering many questions about data processing and tomography. I am grateful to Ori, Mairi and Sumi for sharing cell culture and influenza work at EMBL. I further would like express my deepest appreciation for Ori who answered many of

my cell biology related questions and gave lots of helpful advice at the beginning. I am very grateful to have got the chance to start my PhD in group of such wonderful scientists who made the beginning of my PhD lots of fun and a steep learning curve.

I particularly would also like to thank everyone from the lab who was involved in the lab move to Cambridge: Kun, Lauren Ann, Aaron, Oleksi and Yury for supporting each other during and after the move. I would like to thank Mairi for her huge efforts related to the organization of the move and Beata for being always super helpful whenever I was back at EMBL. I must thank Dustin for his sleepless nights to get microscopes running smoothly on tomography and for all his efforts to improve and develop data processing scripts.

I also would like to thank the EMBL graduate office and in particular Matija as well as Jennie from the structural studies department at the LMB for being really helpful and cooperative and for lots of assistance with administrative tasks related to the move between Heidelberg and Cambridge.

I thank all my friends in Heidelberg and at EMBL for always feeling welcome back in Heidelberg as well as my friends in Cambridge who helped me feeling home there. I am deeply thankful to Klaus for critical reading of this thesis and many helpful comments and many productive discussions during the whole time of my PhD. Finally I would like to thank my family for being extraordinary supportive during every step of this journey.

Table of contents

List of figures **xix**

Abbreviations **xxiii**

1 Introduction **1**

1.1 Influenza virus 1

1.1.1 Influenza virus - a dangerous pathogen 1

1.1.2 Influenza virus basics 2

1.1.3 Life cycle of influenza A virus 5

1.1.4 Assembly of influenza A virus 9

1.1.5 Flu antivirals and vaccines 10

1.2 CryoET and subtomogram averaging 11

1.2.1 Transmission electron microscopy 11

1.2.2 The transmission electron microscope 13

1.2.3 Preparation of biological samples for TEM 16

1.2.4 Contrast and CTF 18

1.2.5 CryoEM - a tool for protein structure determination 19

1.2.6 Cryo-electron tomography 21

1.2.7 Subtomogram averaging 24

1.3 Scope of this study 26

2 Material and Methods **29**

2.1 Mammalian cell culture 29

2.2 Virus stocks and virus purification 30

2.2.1 Virus stocks 30

2.2.2 Quantification of virus titers 30

2.2.3 Virus purification 31

2.3 VLP plasmids 31

2.3.1 VLP plasmids 31

2.3.2 VLP plasmid amplification 32

2.4 Sample preparation for cryoET . . . 32
2.4.1 Preparation of grids for cell seeding . . . 32
2.4.2 Preparation of virus samples on grids . . . 32
2.4.3 Plunge freezing of grids with mammalian cells . . . 33
2.4.4 Preparation of VLP samples on grids . . . 33
2.5 CryoET . . . 34
2.5.1 CryoET data collection . . . 34
2.5.2 Data preprocessing and tomogram reconstruction . . . 35
2.6 Subtomogram Averaging . . . 36
2.6.1 Subtomogram Averaging of HA . . . 36
2.6.2 Subtomogram Averaging of M1 . . . 37
2.7 Downstream data analysis . . . 39

3 Structural studies of IAV hemagglutinin in situ 41
3.1 Introduction . . . 41
3.1.1 HA in the virus life cycle . . . 41
3.1.2 HA as a vaccine target . . . 43
3.1.3 Motivation . . . 45
3.1.4 Aims of this study . . . 46
3.2 Results . . . 47
3.2.1 Sample purification affects virus particle quality . . . 47
3.2.2 High resolution cryoET of non-purified HK68 virus . . . 50
3.2.3 HA in situ structure from non-purified, inactivated HK68 virus 52
3.2.4 Comparison of the HA reconstruction from subtomogram averaging to existing HA crystal structures . . . 55
3.2.5 In situ structure of HA from non-purified, inactivated Udorn virus . . . 55
3.2.6 In-situ structure of HA from fixed and unfixed HK68 VLPs . . 58
3.2.7 In situ HA structure from non-purified, unfixed HK68 virus . 65
3.2.8 Structural differences between HA structures from fixed samples and from unfixed samples . . . 67
3.2.9 Comparison of a HA protomer and a HA domain fit . . . 69
3.2.10 Characterisation of HA structures with an extended density in the HA2 Helix A region . . . 72
3.2.11 HA arrangement on the surface of virus and VLP filaments . . 75
3.2.12 Structure of two HA proteins with a fixed relative orientation found in fixed sample. . . . 79
3.2.13 HA density and HA tilt angle in different samples . . . 79

3.3 Discussion . . . 81
3.3.1 Aldehyde fixation captures HA in an open conformation . . . 84
3.3.2 Functional roles of open and closed HA conformations . . . 85
3.3.3 Quantification of HA arrangement . . . 86
3.3.4 Fixation might induce HA-HA crosslinking . . . 89
3.3.5 Functional roles of the HA tilt angle . . . 90
3.4 Conclusions . . . 91

4 In situ characterisation of M1 93
4.1 Introduction . . . 93
4.1.1 M1 inside the virus . . . 94
4.1.2 M1 functions during the virus life cycle . . . 96
4.1.3 Motivation and aims of this study . . . 97
4.2 Results . . . 98
4.2.1 In situ structure of M1 from non-purified HK68 VLP filaments 98
4.2.2 Existing M1 NTD models fit the obtained M1 density map . . 105
4.2.3 In situ structure of M1 from non-purified HK68 virus . . . 108
4.2.4 Structure of the IAV M1 CTD . . . 112
4.2.5 Secondary structure predictions of the IAV M1 CTD . . . 113
4.2.6 Comparison of M1 to homologue structures . . . 117
4.2.7 Attempts to crystallise the M1-CTD . . . 120
4.2.8 The membrane interface of the M1 NTD . . . 122
4.2.9 M1-M1 interfaces . . . 123
4.2.10 Comparison of identified M1-M1 interfaces to crystal contacts in M1 NTD crystals . . . 126
4.2.11 M1 arrangement in IAV filaments . . . 128
4.2.12 Polarity of M1 strands . . . 130
4.2.13 M1 in virus assembly - conclusions from a theoretical model . 134
4.3 Discussion . . . 137
4.3.1 Structure of the M1 CTD . . . 139
4.3.2 M1-membrane interactions . . . 140
4.3.3 M1-M1 interfaces and M1 arrangement inside viruses and VLPs 141
4.3.4 M1 in assembly . . . 142
4.3.5 Conclusions . . . 143

5 Discussion and future perspectives 145

References 149

List of figures

1.1 Schematic of the architecture and components of influenza A virions. 3
1.2 Variability of IAV particle morphologies. 4
1.3 Overview of the IAV life cycle. 6
1.4 A model of influenza assembly. 8
1.5 Schematic of the different interactions between electrons and matter. . 12
1.6 Principles and components of a high resolution TEM. 14
1.7 Sample preparation for transmission electron microscopy. 17
1.8 Overview of a cryoEM single particle workflow. 20
1.9 Illustration of the information transfer in tomography. 22
1.10 Diagram of the subtomogram averaging pipeline used in this study. . . 25

3.1 Crystal structure of HA. 42
3.2 Schematic of a HA-mediated membrane fusion model. 44
3.3 Comparison of virus particle quality and morphology for two different sample preparation approaches. 48
3.4 Non-purified, inactivated HK68 virus imaged by cryoEM and cryoET. 51
3.5 Subtomogram averaging of HA from non-purified, inactivated HK68 virus. 53
3.6 Comparison of the HA in situ structure from non-purified, inactivated HK68 virus to existing HA crystal structures. 56
3.7 Comparison of tomograms and HA reconstructions from non-purified, inactivated Udorn and HK68 virus. 57
3.8 Comparison of the HA structures obtained from Udorn and HK68 virus. 59
3.9 Tomogram slices and subtomogram averaging structures of HA obtained from fixed and unfixed HK68 VLPs. 61
3.10 HA structures from fixed and unfixed VLPs in comparison to HA crystal structures. 64
3.11 In situ HA structure from non-purified, unfixed HK68 viruses. 66

3.12 Direct comparison of all HA structures obtained from fixed and unfixed samples. . . . 68
3.13 Schematic of the proposed movement to transition from the closed to the open conformation of HA. . . . 70
3.14 Comparison between fitting a complete HA protomer and individual HA domains into obtained HA reconstructions. . . . 71
3.15 Correlation between the density of the HA2 helix A extension and opening of the trimer . . . 72
3.16 Comparison between the open HA in situ structure and the H5 structure with an extended HA2 helix A. . . . 75
3.17 Summary of HA arrangement from different samples. . . . 76
3.18 Reconstruction of two HA proteins in a fixed relative orientation. . . . 80
3.19 HA average density and HA tilt angles. . . . 82
3.20 Schematic of the proposed breathing motion of HA in the context of HA-mediate membrane fusion. . . . 87

4.1 Structure and arrangement of M1. . . . 94
4.2 M1 in the IAV life cycle. . . . 95
4.3 Tomograms of non-purified IAV VLP particles show a strong matrix layer density underneath the membrane. . . . 99
4.4 M1 subtomogram averaging reference over multiple iterations of alignment. . . . 100
4.5 Comparison of two different wedge weighting strategies employed during subtomogram averaging of M1. . . . 102
4.6 Geometric cleaning of subtomogram positions reduces the resolution anisotropic of M1. . . . 103
4.7 Final subtomogram averaging reconstruction of M1 obtained from HK68 VLPs. . . . 105
4.8 Comparison between the final M1 reconstruction from subtomogram averaging and the crystal structure of the M1 NTD. . . . 106
4.9 In situ structure of IAV M1 obtained by subtomogram averaging of M1 from HK68 tomograms. . . . 107
4.10 Evaluation of the M1 reconstructions from HK68 virus and HK68 VLPs. 109
4.11 Simulation of the CTF for a range of defocus values from tomograms used for M1 subtomogram averaging. . . . 111
4.12 Description of the inner lobe of the M1 reconstruction from HK68 VLPs. 112
4.13 Three models of possible connectivities for the densities found in the inner lobe of the M1 structure from HK68 VLPs. . . . 114

4.14 Comparison between secondary structure prediction of the M1 CTD and densities from the inner lobe of the subtomogram averaging reconstruction of M1. 115
4.15 Homology of M1 across different orthomyxovirus strains. 118
4.16 Comparison between the ISAV M1 crystal structure and the M1 reconstruction obtained by subtomogram averaging. 119
4.17 Expression and purification of the M1 CTD fragment. 120
4.18 Characterisation of M1-membrane and M1 NTD-CTD interfaces. . . . 122
4.19 Characterisation of M1-M1 interfaces. 124
4.20 Summary of crystal contacts in M1-NTD crystals at neutral and low pH. 127
4.21 Comparison between M1-M1 interfaces identified from the subtomogram averaging reconstruction of M1 and M1 NTD crystal contacts. . 129
4.22 M1 arrangement in virus and VLP filaments. 131
4.23 Polarity of M1 strands within virus filaments. 132
4.24 A theoretical model of virus assembly. 135

Abbreviations

AMP	Ampicillin
BSA	Bovine Serum Albumin
CCC	constraint cross correlation
CCD	charged coupled device
CPE	cytopathic effect
cryoEM	cryo-electron microscopy
CTD	C-terminal domain
CTF	Contrast Transfer Function
DDD	direct electron detecting device
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
DQE	detector quantum efficieny
EFTEM	Energy-filtered transmission electron microscopy
EM	Electron microscopy
ER	Endoplasmic Reticulum
FBS	Fetal Bovine Serum
FSC	Fourier shell correlation
HA	hemagglutinin
HEK	Human Embryonic Kidney
HK68	influenza A/Hong-Kong/1/1968 (H3N2)
IAV	Influenza A Virus
ISAV	infectious salmon anemia virus
LB	lysogeny broth
M1	matrix protein 1
M2	matrix protein 2
MDCK	Madin-Darby Canine Kidney
MOI	multiplicity of infection

mRNA	messenger RNA
NA	neuraminidase
NEP	Nuclear Export Protein
NLS	nuclear localization sequence
NMR	Nuclear Magnetic Resonance
NP	Nucleoprotein
NS1	Non-Structural Protein 1
NTD	N-terminal domain
PBS	Phosphate Buffered Saline
PFA	Paraformaldehyde
PFU	plaque forming units
PR8	influenza A/Puerto Rico/8/1934 (H1N1)
PS	Penicillin/Streptomycin
RBS	receptor binding subdomain
RMSD	root mean square deviation
RNA	Ribonucleic acid
SAXS	small angle X-ray scattering
SEC	size-exclusion chromatography
SNR	signal-to-noise ratio
SSP	secondary structure prediction
TEM	Transmission electron microscopy
TMD	transmembrane domain
Udorn	influenza A/Udorn/307/1972 (H3N2)
VLPs	virus-like particles
vRNPs	viral ribonucleoproteins
WBP	weighted back-projection
WHO	World Health Organization
WSN	influenza A/WSN/1933 (H1N1)

Chapter 1

Introduction

In this first chapter, I will give a general introduction to different aspects of influenza virus and the virus life cycle as well as to the technology and methods used for the work presented in this thesis. Additional shorter introductions, specific to the content of each chapter, will be given in the introductory part of each chapter.

1.1 Influenza virus

1.1.1 Influenza virus - a dangerous pathogen

Influenza viruses typically infect the respiratory tract of different host organisms including humans, pigs, birds and other animals. Constant evolution of influenza virus leads to a continuous emergence of new strain variants which are responsible for seasonal epidemics (*1*). Virions from substrains allow for the reoccurrence of seasonal influenza infections in the same individual due to reduced antigenic recognition of the evolved virus (*2*). In addition to continuous small changes, occasional evolutionary jumps can occur which causes the emergence of a new strain. New strains bear a high pandemic potential since they lack any antigenic recognition throughout the whole population. Consequently, viruses of new strains are highly infectious and might show different, more severe symptoms (*3*, *4*). The most devastating pandemic recorded was the 1918 Spanish flu, which killed around 50 million people worldwide. Since then, three more flu pandemic strains have emerged causing the 1957 Asian Pandemic, the 1968 Hong-Kong pandemic and the 2009 swine-flu pandemic. Around the world, seasonal epidemics are responsible for 3-5 million potentially fatal infections and 0.3 - 0.6 million deaths per year (*5*).

On the molecular level, antigenic drift, responsible for constant changes, is caused

by high error rates during genome replication while antigenic shift to generate a new strain is related to reassortment mechanisms of the segmented flu genome as well as host switching events. In addition, both mechanisms can occur together. In this case, genome reassortment occurs across strains that typically infect different hosts during the accidental infection of the same host. These cases further facilitate evolutionary jumps and the generation of new, potentially pandemic, strains. In the case of influenza, birds and pigs have played infamous roles and potentially contributed to the emergence of the 1918 influenza strain (*6*, *7*).

A century after the Spanish flu, influenza pandemics still represent a major risk. While today, flu treatments exist, a pandemic would result in a unique logistic challenge to produce and distribute enough flu medication. Besides post infection treatment, vaccinations are available today and are recommended as a protection against seasonal strains. Vaccinations target a combination of strains that are predicted to occur in the upcoming season. Due to the continuous antigenic shift, vaccinations need annual refreshment and bear a risk of strain miss matching. In addition, current vaccines will not cover the occurrence of an unpredicted pandemic strain. Therefore, vaccination strategies available do not provide complete protection. The most efficient prevention of new pandemics would be the development of a broadly effective influenza vaccine against all strains. Currently, research is on-going to identify more stable vaccine targets in order to achieve this goal (*8–10*).

1.1.2 Influenza virus basics

Influenza virus nomenclature

Influenza virus belongs to the virus family of orthomyxoviridae. By now four types of influenza viruses have been reported: Influenza A, B, C and D virus. Influenza A and B are responsible for human epidemics. Influenza A represents the most widely abundant species (*11*). Influenza A Virus (IAV) are further subdivided according to the antigenic variants of the two glycoproteins HA and NA. Currently 18 HA antigenic variants H1-H18 and 9 NA NA1-NA9 variants exist. Combination of those lead to the typical subtype names. 3 variants of these: H1N1 (Spanish Flu 1918, 2009 swine-flu pandemic), H2N2 (Asian pandemic 1957), H3N2 (Hong-Kong Pandemic 1958), are endogenous to humans while many of the other influenza A subtypes are endogenous to birds, the original host of influenza virus (*12*). Current seasonal strains comprise H1N1, H3N2 variants as well as influenza B strains (*13*). Typically after a pandemic, the pandemic subtype replaces the previously circulating seasonal strains and evolves

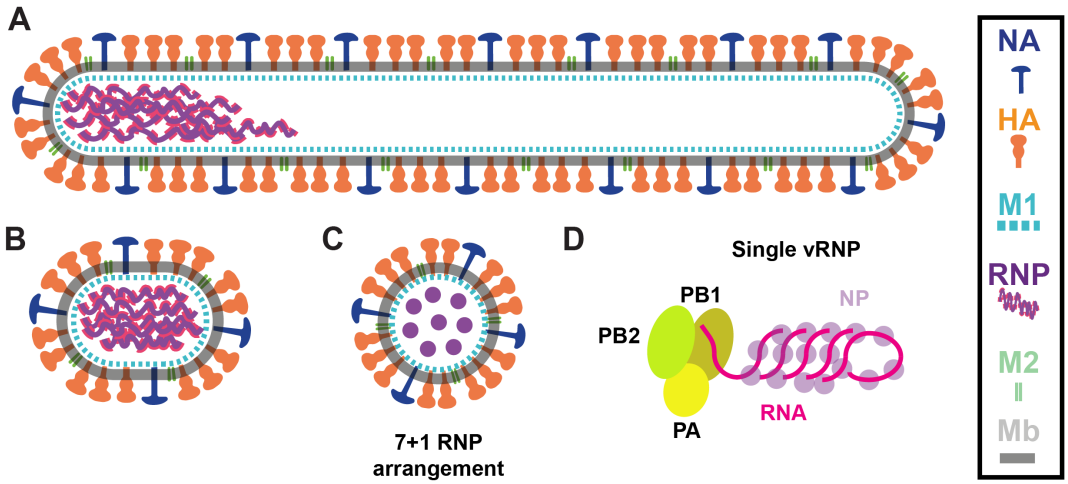

Figure 1.1 Schematic of the architecture and components of influenza A virions. A) Filamentous IAV particles. B) Shorter, bacilli-shaped IAV particles. Components are the same as in A (*15–17*). C) Vertical cross-section through a virion shown in A or B at a position where the vRNPs are present, demonstrating the circular arrangement of 7 vRNPs surrounding the 8th segment in the centre (*18*). D) Depiction of a single vRNP. The RNA (pink) is covered by NP (purple). Together they form a helix. Each vRNP carries its own heterotrimeric RNA-dependent RNA polymerase built from the three subunits PB1, PB2 and PA (*19*). In legend: Mb = membrane.

further from there (*14*). During the pandemic 1968 originating in Hong-Kong, the H3N2 subtype occurred for the first time and persisted as a seasonal subtype since then (*2*). According to the definitions from the world health organization World Health Organization (WHO), influenza virus strains are typically named by combining the following information: the type (A,B,C or D), followed by the host of origin (empty for humans), the geographic origin, the strain number and the year of isolation. For IAV the subtype is typically indicated in parenthesis. Work presented in this thesis is mostly based on the pandemic strain from Hong-Kong: influenza A/Hong-Kong/1/1968 (H3N2) (HK68)

Influenza A virus components and architecture

Influenza A virus is a negative-sense, single stranded RNA virus with a 8-fold segmented genome (*20*). The 10 Mbp-sized genome is split into 8 segments which encode at least 10 proteins since at least two segments encode for alternative splice variants. The number of described proteins is currently increasing since more splice variants are getting discovered (*21*, *22*). Inside an assembled virion, the eight RNA genome segments are each covered in multiple copies of Nucleoprotein (NP) and together they appear as elongated, helical structures, the viral ribonucleoproteins (vRNPs) (Fig. 1.1C) (*18*). The end of each genome segment is bound by the trimeric polymerase

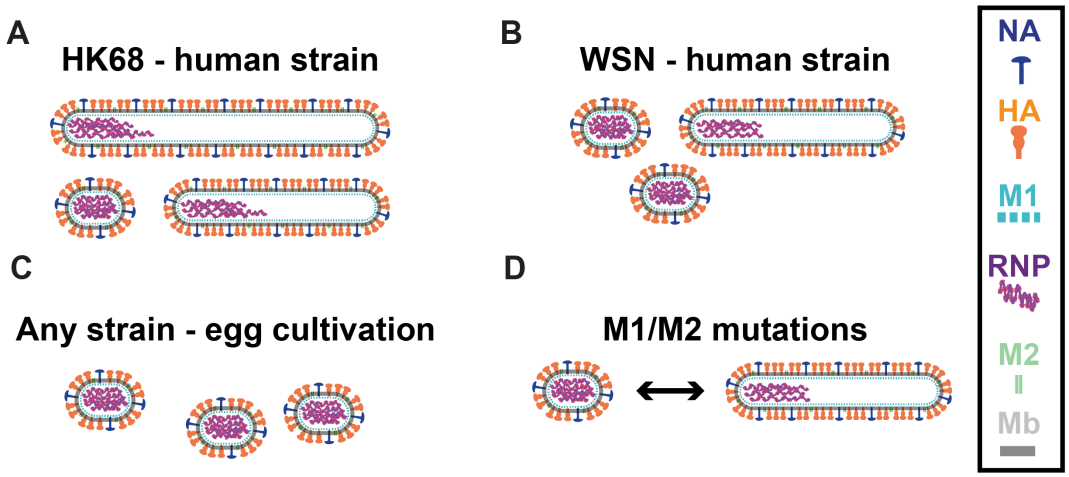

Figure 1.2 Variability of IAV particle morphologies. A) Morphology distribution of particles from the HK68 strain. A large fraction of particles is filamentous. B) Morphology distribution of particles from the WSN strain. Most particles are thought to be spherical or egg-shaped. A and B represent situations when virus is produced in mammalian cells. C) Morphology distribution of particles upon cultivation in chicken eggs, which is known to lead to the loss of filamentous morphology (*24*). D) Mutations in the M1 and M2 gene were shown previously to affect the morphology of IAV (*25*).

complex consisting of the subunits PA, PB1 and PB2 (*23*). At the inside of one tip of the virion, seven vRNPs are arranged in a circular manner surrounding the eighth vRNP (*18*) (Fig. 1.1). The viral envelope, encompassing the vRNPs is stabilized by a matrix layer made from matrix protein 1 (M1) which is located underneath the membrane bilayer. The membrane is derived from the host cell and is decorated with the glycoproteins hemagglutinin (HA), neuraminidase (NA) and the transmembrane ion channel matrix protein 2 (M2). Non-Structural Protein 1 (NS1) and Nuclear Export Protein (NEP) are not part of the assembled virions, but are expressed and crucial for infections once the virus enters the host cell.

Influenza A virus morphologies

Unlike the typical depiction of influenza viruses as spiky spheres, the appearance of influenza virions is heterogeneous and ranges from spherical particles to up to 20 um-long filaments (*26*). The filamentous form is typically found in patient's samples (*27*) while the spherical morphology is described in the context of lab-adapted strains and correlates with passaging virus in eggs (*17*, *25*) (Fig. 1.2C). Beside the influence from external factors, different genetic traits were shown to have an impact on the morphology of influenza virions (Fig. 1.2D). Different strains show tendencies for different predominant morphologies. influenza A/WSN/1933 (H1N1) (WSN) and influenza A/Puerto Rico/8/1934 (H1N1) (PR8) are examples of strains that have been

reported to be predominantly spherical while strains influenza A/Udorn/307/1972 (H3N2) (Udorn) and HK68 have been reported to be predominantly filamentous. However in all cases, particle morphologies follow a distribution of different lengths (Fig. 1.2A,B). While the length of virions vary, the diameter across virions is more conserved and ranges between 60 nm to 100 nm across different morphologies and different strains. Investigations about the causes for different morphologies have led to the identification of multiple factors that influence morphology. On the molecular level, morphological changes are mostly related to key point mutations in the M1 and M2 proteins (*28*, *29*). It was further shown that M1 mutations are acquired if virus particles are transferred from being cultured in eggs into an animal model (*25*). The specific effect of the environmental pressure onto M1 remains to be investigated.

Functions of the different morphologies are still under debate. One model for the role of filamentous virions refers to direct cell-to-cell transmission of the genome to facilitate the infection of neighbouring cells (*30*). Others have suggested that filaments are more motile in the infectious environment of an organism, the respiratory tract where newly produced virions encounter mucus and ciliated cells. This idea recently was further supported by showing that filamentous morphology favours an asymmetric distribution of the receptor binding protein HA and the receptor-degrading protein NA which allows to increase virus motility in mucus without reducing the binding stability between HA and the host cell receptors (*31*). It was further suggested that the heterogeneity in morphology allows the virus to survive in variable environments leading to an increase in viral fitness (*32*).

1.1.3 Life cycle of influenza A virus

Infection of host cells is mediated by the viral glycoprotein HA that recognizes and binds sialic acid residues of cell surface proteins to induce virus uptake (Fig. 1.3 1). For virus entry, clathrin-mediated endocytosis as well as macropinocytosis have been observed (*34*, *35*). In both cases, the virion ends up in an intracellular membrane compartment which traffics along the endosomal pathway (Fig. 1.3 2). Endosome acidification exposes the virus to low pH which causes a major conformational change of HA in preparation for membrane fusion of the viral and endosome membrane (Fig. 1.3 3). The fusion peptide of HA, which at physiological pH is buried in the core of the protein, becomes exposed at the top of the molecule to interact with the opposing endosome membrane. In a sequence of conformational changes, HA assists in pulling the viral and the endosomal membrane close enough to each other for membrane fusion to happen (for more details see section 3.1.1). At the same time,

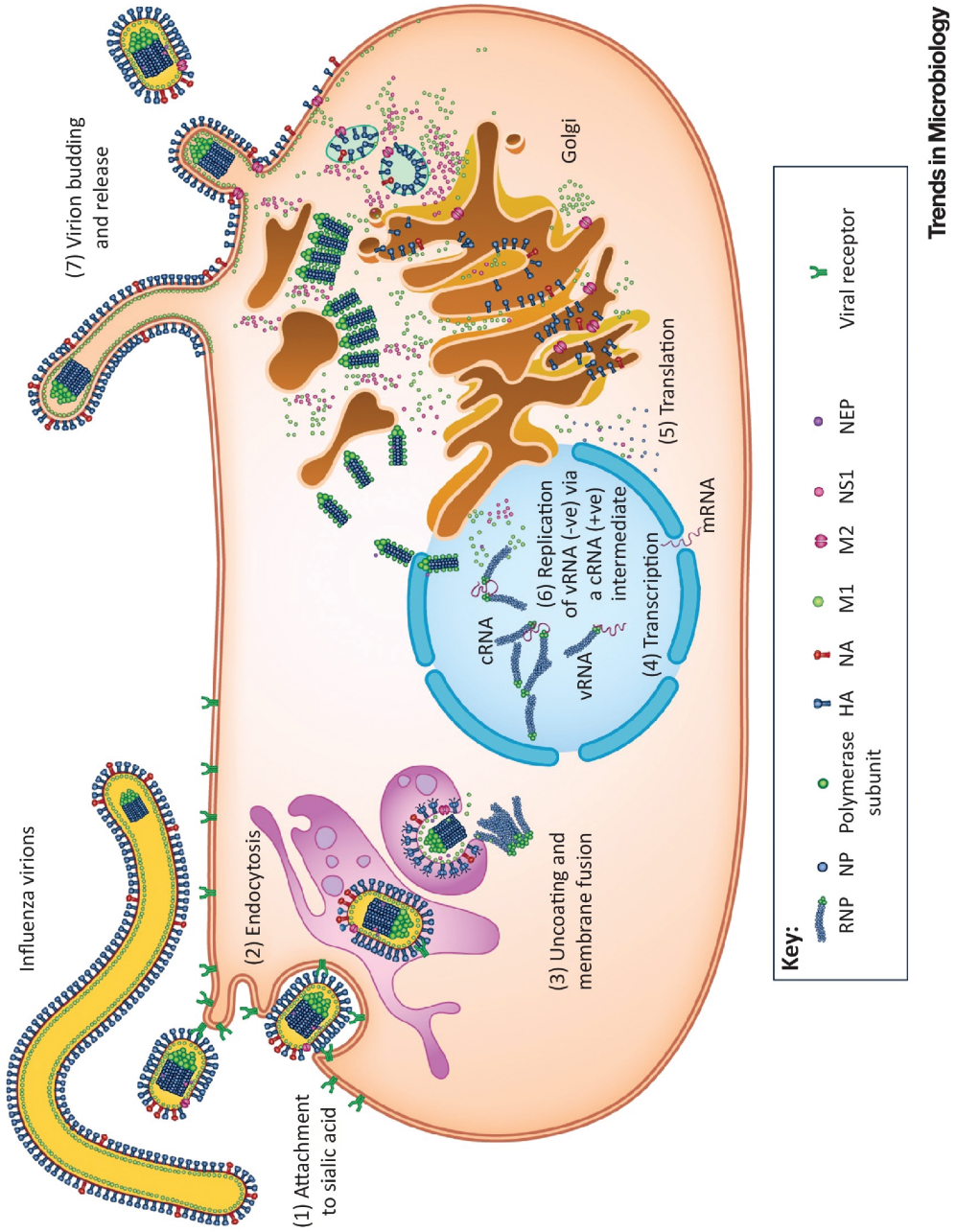

Figure 1.3 Overview of the IAV life cycle. Different steps of an IAV infection starting from the cell receptor recognition by HA (1) and the endocytosis of the bound virus particle by the host cell machinery (2). Upon HA-assisted fusion between the virus envelope and the endosomal membrane the genetic material (the vRNPs) is released into the cytosol (3). The vRNPs are further transported into the nucleus for transcription (4) and replication (6). Transcribed virus RNAs, disguised as endogenous mRNAs, are exported into the cytosol where viral proteins are translated (5). All components are transported towards the cellular plasma membrane where newly formed virions form (7). From (*33*) published under a CCBY4.0 license.

the proton-selective ion channel M2 inside the viral membrane opens upon exposure to low pH, leading to a proton influx into the virus core. Low pH inside the virus is suggested to induce conformational changes of M1 to allow the release of the vRNPs from the virus core into the cytoplasm.

Interactions of the NP protein, which covers the viral Ribonucleic acid (RNA), with host cellular proteins of the importin family initiates transmission of vRNPs through nuclear pores into the nucleus. Inside the nucleus, the RNA-dependent polymerase complex associated with each vRNP allows for independent replication and transcription of each segment (Fig. 1.3 4 and 6). Replication involves initial transcription of the single stranded RNA into the complementary, positive, cRNA strand. This strand is subsequently used as a template to produce the original negative RNA strand, that is used for vRNP formation of new virions. Before replication occurs, the vRNA is already transcribed into high amounts of messenger RNA (mRNA) copies for subsequent viral protein expression inside the cytosol. The transcription process is very efficiently initiated by a process termed 'cap-snatching' (*36*). Here, the PB2 viral polymerase subunit interacts with the cellular RNA polymerase II in order to cleave and 'steal' newly formed 5' prime caps of nascent cellular mRNAs with help of the PA subunit. These caps serve as primers which are then very efficiently elongated by the viral polymerase PB1 subunit now using the vRNA as a template. To complete mimicking of cellular mRNAs and ensure translation by the host's cellular machinery, the viral transcript also finishes with a polyA tail, added by the viral polymerase complex. Once, the polymerase encounters a stretch of 5-7 uracils towards the end of the transcript polyadenylation is achieved through a process that was described as stuttering of the polymerase. Effectively, the transcript is iteratively back-shifted relative to the template leading to annealing of more adenosins than encoded in the short poly-U stretch of the viral RNA. Being now indistinguishable from cellular mRNA, viral transcripts are exported from the nucleus for subsequent translation (*37*). Prior to nuclear export RNA transcripts are spliced by the host cell's splicing machinery to form the alternative splice variants that were discussed in section 1.1.2. (*19*).

Translation of new viral proteins relies on the host cell machinery and happens at soluble ribosomes or at Endoplasmic Reticulum (ER)-bound ribosomes in the case of the membrane proteins HA, NA and M2 (Fig. 1.3 5). The newly expressed vRNP protein components NP, PA, PB1 and PB2 are subsequently imported into the nucleus where they associate with newly replicated viral RNAs to assemble new vRNPs. Additionally, the proteins NEP (also called NS2) as well as M1 are imported into the nucleus to assist with export of assembled RNPs into the cytosol and towards assembly

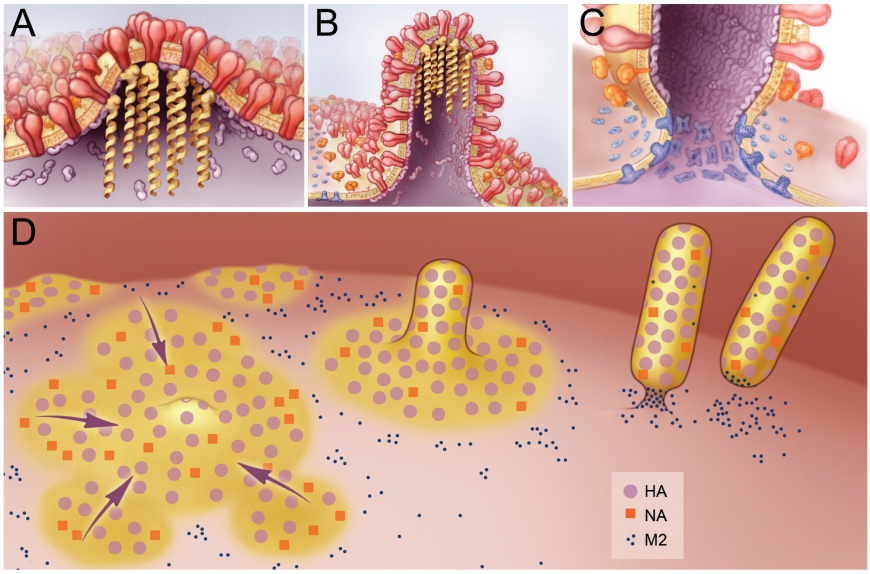

Figure 1.4 A model of influenza assembly. A) Initial formation of membrane protrusions by HA (red), the vRNPs (yellow) and M1 (purple). B) Growth of membrane protrusion. C) Scission of particles with the help of M2 (blue) and NA (orange). D) Clustering of the glycoproteins HA and NA at the plasma membrane to initiate assembly. From (*41*) with permission from the publisher.

sites. Here the role of M1 was described as preventing reimport into the nucleus once exported (*38*).

During the expression of the membrane proteins, the future extracellular domains face the ER-lumen. At this stage, where correct folding is surveyed, proteins are post-translationally modified and multimerise into trimers for HA and tetramers in the case of NA and M2. HA, NA and M2 than traffic through the Golgi network towards virus assembly sites as the plasma membrane (Fig. 1.3 7). During the transport through the Golgi network, HA0 is cleaved into HA1 and HA2. Cleavage efficiencies depend on the cleavage-site and are strain dependent. Highly pathogenic strains including HA of avian strains provide cleavage sites that are efficiently cleaved by furin an widely expressed protease in the trans-Golgi network (*39*). Less pathogenic strains are cleaved by more specifically expressed proteases such as TMPRSS2 or the human airway trypsin-like protease (HAT) which are characteristic for human respiratory epithelial cells (*40*). In ex-vivo experiments, cleavage of HA is typically achieved by trypsin addition into the growth medium.

1.1.4 Assembly of influenza A virus

Influenza virus spreads by assembling pleomorphic virions at the plasma membrane of infected cells. Assembly is driven by the interactions of the structural components HA, NA, M1, M2 and the RNPs at the host cell membrane. To date no specific host cell proteins are known to be actively contributing to the assembly process (Fig. 1.4). All components are transported to the assembly site. The membrane proteins HA, NA and M2 are transported to the plasma membrane by intracellular vesicle transfer just as cellular membrane receptors (*19*). For the 8-fold segmented genome it was shown that the vRNPs, although replicated and assembled individually, are typically found as a complete set of all 8 segments in a majority of newly assembled virions (*42*). It was further shown by deleting individual segments cellular RNA is used in order to replace the missing segment demonstrating that the presence of eight segments is specifically controlled for during transport or virus assembly in a process that is not understood in detail yet (*43*).

Different models exist about the initiation of assembly at the plasma membrane. Several studies have identified the clustering of glycoproteins and in particular HA as the initiating step for assembly. HA clustering is further thought to correlate with the presence of membrane domains. Lipid-raft dependent clustering of HA at the plasma membrane was reported previously by super resolution light microscopy studies in infected cells (*44*). In a different study, virus-like particles (VLPs) were used to specifically test the contribution of each individual protein to virus assembly and assembly initiation. VLPs are prepared by transfecting cells with the sequences encoding for the viral proteins. By transfecting only selective subsets and combinations of the viral proteins the impact of individual proteins on the assembly can be tested. The successful formation of VLPs, that closely resemble true virions based on the co-expression of HA, NA, M1 and M2 indicated that the vRNPs are not essential for IAV assembly (*9*, *45*). By further investigating VLP particle morphology in dependence on the subsets of proteins used it was shown that the expression of either of the glycoproteins is sufficient to induce the formation of irregularly shaped glycoprotein-covered, membranous particles. In contrast, all other components need to be paired with at least one glycoprotein to assure formation of new particles (*45*, *46*). These results have suggested that both glycoproteins inhere membrane bending and budding activity, while none of the other components do. Thus the presence of either of the glycoproteins during the early stages of assembly appeared to be essential to provide initial bending of the plasma membrane.

Presence of M1 at the assembly site was thought to lead to the extension of the initiated membrane protrusion (Fig. 1.4B). The recruitment of M1 was suggested to be induced by the cytoplasmic tails of the glycoproteins. In co-expression studies of HA and M1 an increased membrane binding affinity of M1 was found when HA was present as well (*47*). Previous EM studies of influenza virions have shown that M1 is located underneath the membrane in a regular arrangement. It was therefore suggested that the accumulation of matrix protein underneath the membrane contributes to the growth of budding virions (*15*). By comparing the particle morphology of VLPs formed with and without M1 it became clear, that M1 is essential for the generation of regularly shaped particles including filaments (*45*, *46*). M1 is further thought to interact and connect all components contributing to assembly. However, based on the observed heterogeneity of assembled particles even within the same population it was proposed that no specific mechanisms are in place to regulate envelope composition (*32*).

Assembly is thought to be completed by M2-mediated scission of the formed particles from the host cell membrane (*48*). Upon scission, virions are released from the host cell. At this stage, and presumably during the entire assembly process, the presence of NA prevents the reabsorption of viral progeny to the host cell. The sialiase activity of NA is essential to remove sialic acid residues from host cell receptors otherwise newly formed virus particles tend to reabsorb to cell surface receptors via HA (*19*).

Overall many aspects about the organization and succession of the different steps of assembly of influenza virus remain unclear to date. One reason for the lack of insights into the process is that studying assembly in situ and at the molecular level is challenging and limited to high resolution microscopy methods such as super resolution light microscopy, atomic force microscopy (AFM) or electron microscopy (EM). Consequently, most of the experimental insights of current assembly models result from studying the individual roles of the proteins involved which has provided an important foundation to further study assembly in situ .

1.1.5 Flu antivirals and vaccines

Four different influenza antivirals have been developed for influenza post-infection treatments: M2 channel blocker (adamantanes), NA inhibitors, HA fusion inhibitors and more recently polymerase inhibitors. NA inhibitors are the most effective and recommended treatment since resistance mechanisms remain rare in currently circulating strains (*1*, *49*). NA inhibitors, interfere with the sialidase activity of NA,

and thereby prevent the release of newly formed virus particles from the host cell surface. The most widely used NA inhibitor, and flu antiviral in general, is oseltamivir, commercially available as 'Tamiflu' (Roche). In contrast many current strains are resistant to adamantanes due to mutations acquired in the M2 gene.

The most popular vaccine target is the glycoprotein HA. The widely exposed head domain of HA represents the primary target of endogenous and exogenous antibodies. However, the continuous evolution of the HA antigen, in particular in the head region currently limits the efficacy of available vaccines and impedes the development of a broadly effective vaccine (*10*). More details about HA-targeting ABs are provided in section 3.1.2.

1.2 CryoET and subtomogram averaging

1.2.1 Transmission electron microscopy

Electron microscopy (EM) allows imaging of various materials across multiple scales and at molecular resolution. For all types of EM the image is formed upon the interaction of a beam of accelerated electrons with the sample. When electrons interact with matter a broad range of possible interactions can take place including elastic scattering, inelastic scattering and X-rays (Fig. 1.5). In Transmission electron microscopy (TEM), elastic scattering is the most relevant interaction type for image formation (*50*). The amount and type of scattering that occurs upon sample interaction depends on the material and on the length of the path the electron is travelling through the sample. A long path favours multiple scattering events and increases the chance for inelastic scattering and higher scattering angles. In order to retain electrons with small scattering angles, ideally from a single scattering event, samples for TEM have to be extremely thin, typically below 100 nm to 200 nm and have to be prepared accordingly by sectioning or alternative thinning methods (*50*).

Besides elastic scattering, interactions between electrons and the sample cause inelastic scattering (Fig. 1.5). In contrast to elastic scattering, where the energy of the exciting electron is maintained, an inelastically scattered electron has lost energy during sample interaction. Inelastically scattered electrons inherit variable amounts of energy and therefore unpredictable properties. Therefore, signals from inelastically scattered electrons contribute to the generation of image noise and are undesired for standard transmission electron microscopy imaging applications. The chances of inelastic scattering increases with sample thickness (*50*). The energy that is lost

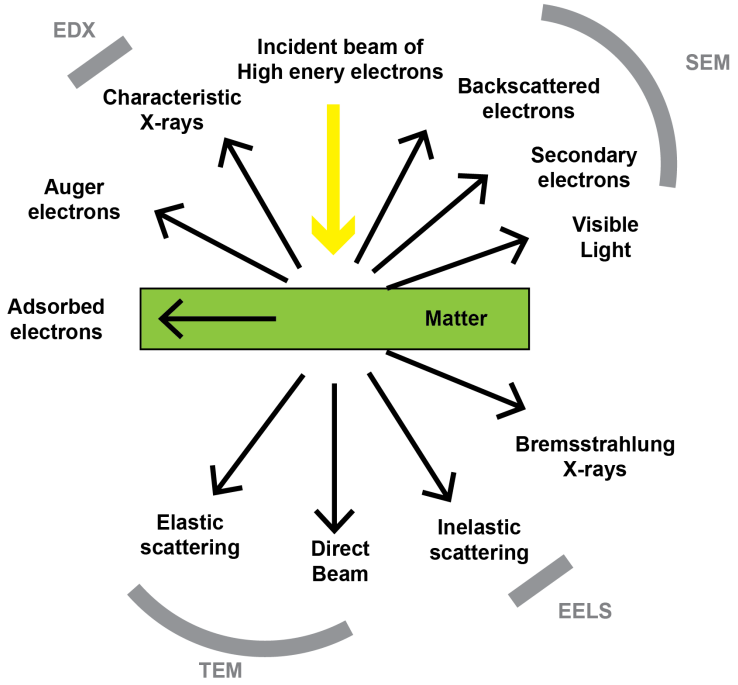

Figure 1.5 Schematic of the different interactions between electrons and matter. Black arrows illustrate the variability in the directions of the emitted signal relative to the incident beam (yellow). Different types of signals contain different information about the sample and multiple of them are used in common techniques to characterise biological and inorganic materials as indicated in grey. EELS, electron energy loss spectroscopy; SEM, scanning electron microscopy, EDX, energy dispersive X-ray spectroscopy.

during inelastic sample interaction is transferred to the sample where its effects often causes various types of radiation damage. Damage is experienced either in the form of direct molecular bond breakages or through the generation of heat and can lead to the destruction of sample features. The material-dependent tolerance to electron radiation is in particular low for biological materials. Higher acceleration voltages and thinner samples reduce the amount of inelastic interactions and therefore the amount of damage to be experienced by the sample. A combination of these precautions and the usage of very low electron doses are therefore essential for imaging of biological samples in transmission electron microscopes (*51*).

Due to the large depth of focus in TEM, a TEM image represents the sum of all interactions across the sample thickness. Information across the imaged 3D volume are projected into a single 2D image plane. Consequently, sample features which are located on top of each other will be displayed in the same pixel of the generated projection image (*50*). Confusion of the content of a projection image can be prevented by using tomographic methods, where several images from different directions are acquired which allows to relate each signal back to a specific plane of the sample.

1.2.2 The transmission electron microscope

Most TEMs are build as tall cylinders of 1 m to 4 m in height with an electron source at the top and finishing with the image detector at the bottom. The space in between is referred to as the microscope column. The electron source at the top generates the high energy electron beam (Fig. 1.6). Electrons are accelerated into the column of the microscope by an applied voltage, typically between 100 kV to 300 kV in biological research. The extraction voltage determines the velocity and therefore the energy of the electrons. To ensure that the electron beam reaches the sample it is essential to prevent collisions with any material including gas molecules along the electron's path. Therefore, high vacuum inside the microscope's column is required. To deliver the sample into the vacuum-holding column, transmission electron microscopes are equipped with gated chambers where the sample within a matching sample holder gets inserted into the microscope.

The path of the accelerated electrons along the column and towards the sample and subsequently the detector is shaped by electron lenses, apertures and deflection coils (Fig. 1.6B). In a transmission electron microscope, the optical elements can be distributed into three regions according to their location in the column and their

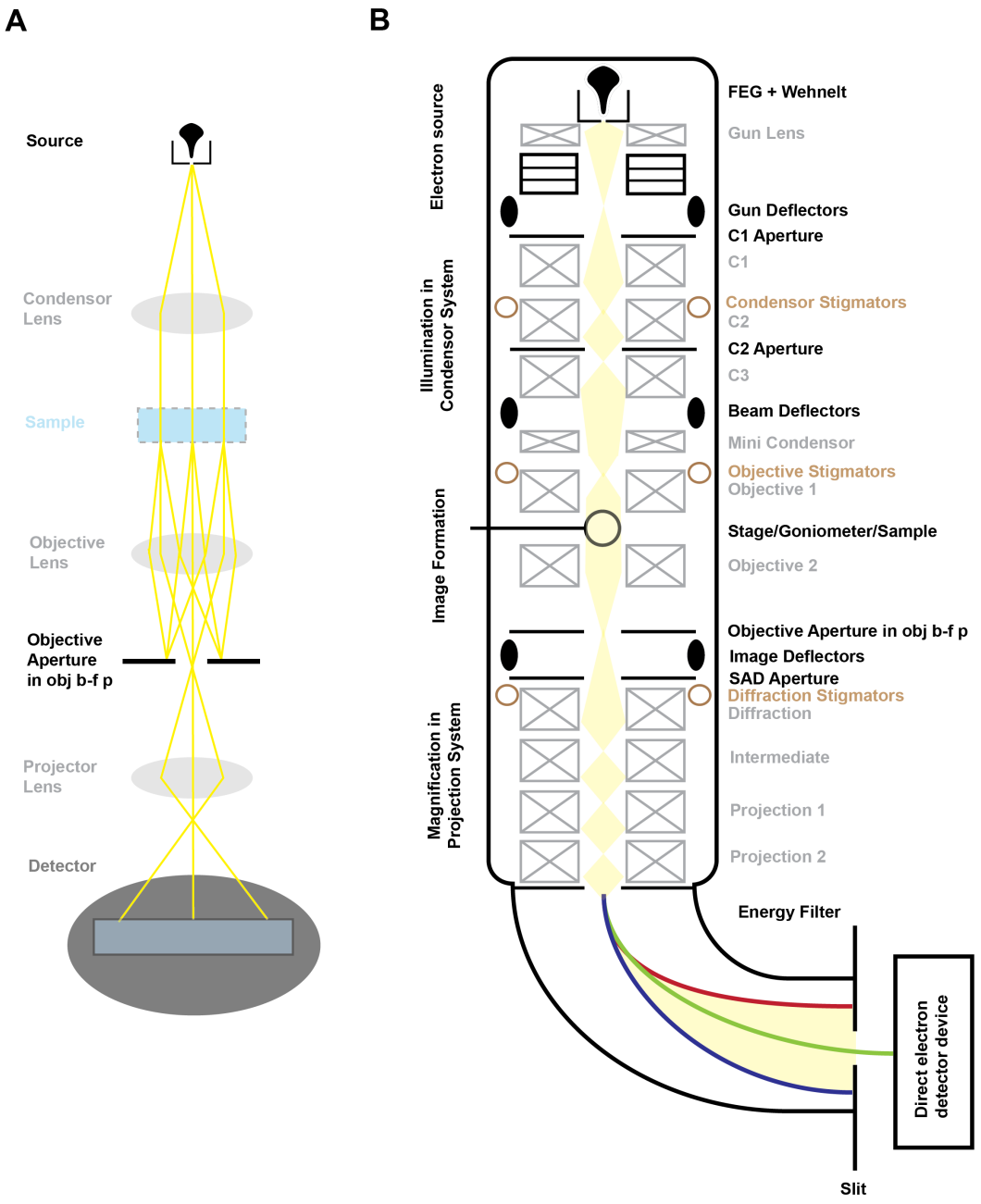

Figure 1.6 Principles and components of a high resolution TEM. A) Simplified schematic of the path of an electron beam through different optical elements in a TEM. The objective aperture is located in the back focal plane (b-f p) of the objective lens. B) Layout of a high resolution 3 condensor lens TEM. Lenses are represented as grey crossed boxes, deflection coils as black ellipsoids, stigmators as brown ellipsoids and apertures as black lines. The TEM visualized here is equipped with a post-column energy filter and a DDD. The colours of the post-column beam indicate the different energies or wavelength of the electrons. Based on (*50*, *52–54*).

functions: The condensor system at the top collects the electrons coming from the source to generate a parallel beam in preparation for interaction with the sample. The objective lenses in the middle collects emitted signals after sample interaction to generate the initial magnified image. The projection system below the objective lenses further magnifies the image and projects the image onto a detector at the bottom of the column (Fig. 1.6) (*50*).

Electron lenses are electromagnetic coils with an electric current applied. The generated electromagnetic field deflects electrons when passing through the lens. In an electron microscope, lenses are placed at fixed distances and the focal length of a lens changes depending on the amount of current applied. In comparison to glass lenses, the quality of electron lenses is very poor which manifests itself in multiple types of lens aberrations of which the most prominent ones are spherical aberration, chromatic aberration and astigmatism. (*55*). Lenses in a transmission electron microscope are complemented by apertures. Apertures are metal pieces with holes of desired width that are inserted into and aligned to the optical path to select a central, homogenous fraction of the beam (Fig. 1.6). The objective aperture is located at the back focal plane of the objective lens and specifically removes electrons with higher scattering angles to improve image contrast (Fig. 1.6A). Deflection coils in different positions along the electron's path through the microscope introduce and correct beam tilt and beam shift. Energy filters can be inserted into or after the column to remove electrons of different energies, typically from inelastic scattering, and to reduce unspecific noise in the generated image (Fig. 1.6B). In a functioning high resolution transmission electron microscope all of the described optical elements, have to be perfectly aligned with respect to each other to ensure that a straight, coherent, homogeneous, beam of the desired intensity passes along the microscope's column (Fig. 1.6B).

At the bottom, transmission electron microscopes are typically equipped with a fluorescent screen to convert electrons into light intensities for visual inspection of the sample. In addition, electronic detectors or traditionally a piece of film, can record an image of the inserted sample. For imaging of biological samples, electronic detectors are either based on charged coupled devices (CCDs) or since more recently, direct electron detecting devices (DDDs) which rely on complementary-metal-oxide-semiconductor (CMOS)-technology. In CCD-based detectors, electrons hit a scintillator which leads to the emission of photon. The photon is detected by the CCD, where a current is generated. This multi step process creates spatial inaccuracies between the electron entry point and the point where the final pixel signal is generated (*55*). In DDDs, electrons directly generate a current when they interact with a pixel.

In addition, DDDs are very sensitive and a provide better detector quantum efficieny (DQE) than CCD-based detectors (*56*). The combination of a highly sensitive detector with a low electron dose allows to accurately determine the position of each detected electron which increases the final resolution of the image. DDDs additionally provide high temporal imaging frequencies which allows splitting the exposure time onto 10-20 individual frames (*57*). The collection of multiple short exposure images allows to computationally correct for drift that builds up during sample exposure by aligning all frames to each other before combining them to generate the final image (*58*).

1.2.3 Preparation of biological samples for TEM

One of the critical requirements for successful TEM applications are thin samples. In addition, biological samples have to prepared in a way to withstand vacuum conditions. Biological samples on their own provide very low contrast since biological material scatters weakly. In order to improve contrast, contrast-enhancing materials such as heavy metal salts can be added during sample preparation (Fig. 1.7a) (*59*). During so-called negative staining, the sample, typically a purified protein or virus in solution, is deposited onto a sample support represented by a metal grid with a thin amorphous carbon film. In a next step, heavy metal salt is spread onto the sample. Upon adsorption the salt forms a cage around the deposited sample particles. Excessive liquid is removed by blotting after each step of the sample preparation process. Once inside the microscope, the heavy metal salt cage around the protein provides strong contrast and a 'negative' of the original object (Fig. 1.7a) (*55*). While negative staining allows for a quick assessment of a prepared sample it also bears the risk of generating artefacts through collapsing of the sample during staining and dehydration which limits the amount of sample detail which is maintained for subsequent imaging.

An alternative sample preparation method captures biological samples directly inside their native aqueous environment (*60*). Through a rapid drop in temperature, a thin film water or buffer containing the sample is turned into electron-transparent, vitreous ice and prevents the formation of electron dense, crystalline ice. The transition into vitreous ice maintains all structural features of the sample up to submolecular details. The introduction of vitrification of samples has started a new field of high resolution electron microscopy of biological material at cryogenic temperatures referred to as cryo-electron microscopy (cryoEM) (*61*).

A standard cryoEM sample preparation workflow is summarized in figure 1.7b.

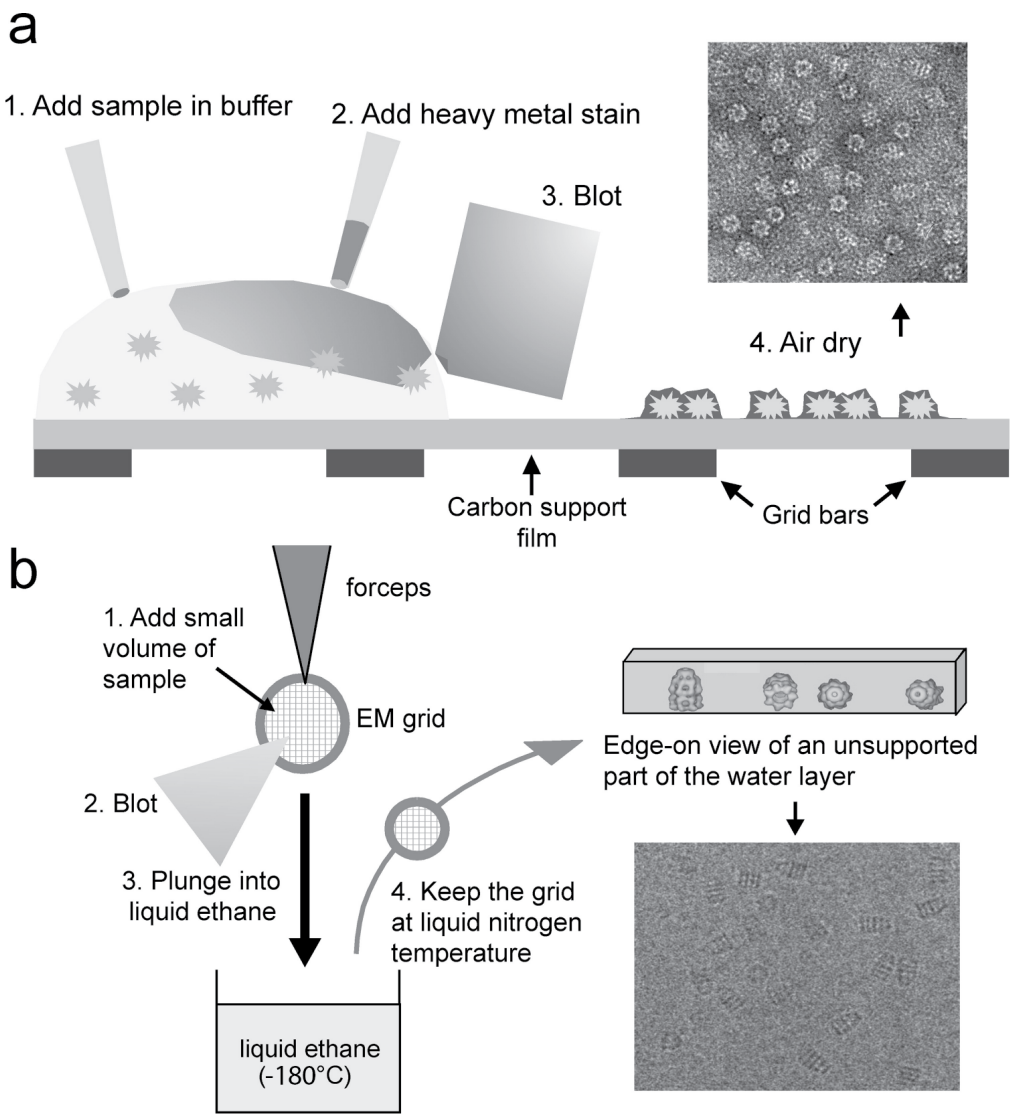

Figure 1.7 Sample preparation for transmission electron microscopy. a) Sample preparation for negative staining relies on the use of heavy metals to generate contrast. Steps include the application of the sample, blotting of excess liquid and addition of the heavy metal solution. An exemplary TEM image of a negatively stained sample of the groel protein is shown on the right. b) Sample preparation for croyEM. On the right a representative cryoEM image of groel is shown illustrating the difference in contrast relative to the image shown in a. Adapted from (*55*) with permission from the publisher.

Purified biological particles in solution are spread on a EM grid. The name refers to a small, circular metal grid with a diameter of around 2 mm. The grid is covered by a thin film of amorphous carbon with empty areas in between, e.g. holy carbon film which has a regular pattern of holes with diameters of 1-3 µm. The sample in solution is spread onto the grid and thereby covered the carbon film as well as the empty areas . Excess liquid is removed to a minimum by blotting, before the grid is immediately plunged into liquid ethane, which is typically cooled by liquid nitrogen to around -180 °C (Fig. 1.7). Ethane provides a high thermal conductivity and therefore allows for rapid enough cooling of the immersed sample to transition water into vitreous ice. To prevent dehydration of the remaining sample volume after blotting, the grid is typically kept in a humidified chamber until plunging (*62*, *63*). Upon plunge freezing samples are kept at liquid nitrogen temperatures until cryoEM imaging to prevent devitrification.

1.2.4 Contrast and CTF

Unlike samples prepared by negative staining, cryoEM samples typically exhibit low contrast due to weak scattering of the biological material (Fig. 1.7b). Scattering described using the wave formulation of electrons, relates to the introduction of an phase shift relative to the incoming wave. The phase-shifted wave interferes with the non-affected wave. For biological samples, the signal in the final images mostly relies on contrast emerging from the phase differences of the phase-shifted and non-shifted waves, which is referred to as phase-contrast. The interference signal is further modulated by imperfections of the optical system, in particular the spherical aberration of the objective lens. Therefore a TEM image is never an exact representation of the original object. One commonly used adjustable effect to manipulate the phases of the signal is the use of defocus in TEM imaging. Defocusing of the objective lens further increases the phase shift which generates an image with higher contrast (*64*). Therefore imaging using a deliberately defocused setting rather than imaging in-focus is a commonly used method to obtain extra contrast in TEM. Phase plates in cryoEM are used to increase contrast even when imaging the sample in focus by introducing an extra phase-shift (*65*).

The sum of all modulations to the signal introduced by the optical system of the microscope is described by the Contrast Transfer Function (CTF). The contrast transfer is typically described in relation to the spatial frequency range of the input signal. The CTF oscillates in a sinusoidal fashion between positive and negative signal. For frequencies where the curve passess zero, no information are transferred. Defocus and

spherical aberration directly affect the contrast transfer. Inversely, the CTF directly depends on both of these parameters. Additional imperfections in the microscope that effect the signal transfer such as reduced spatial coherence of the electron beam, sample drift, variation of lens current and sample charging among others are summarized in the envelope of the CTF, which described an extra dampening of the signal towards higher spatial frequencies (*66*). According to (*66*) and (*67*) the CTF can be expressed as follows:

$$CTF(f) = Asin(\pi\lambda f^2(z - 0.5\lambda^2 f^2 C_s) + Bcos(\pi\lambda f^2(z - 0.5\lambda^2 f^2 C_s)) \quad (1.1)$$

In the equation f represents the spatial frequency, A the envelope function λ represents the electron wavelengths, C_s represents the spherical aberration constant of the objective lens, B represents the fraction of amplitude contrast and z corresponds to the defocus value. For larger defocus values, the first peak of the CTF is shifted towards lower resolutions. This means, that for larger defocus values the first zero crossing is also shifted towards lower frequencies and that more zeros crossings will occur in the same range of frequencies. Therefore, images are preferentially acquired as close to focus as possible to minimize the number of zero transfer frequencies. Today, computational methods exist which allow to estimate the CTF directly from the power spectrum of each image and to correct for the effects of the contrast inversions resulting from alternation of the CTF between positive and negative contrast (*68*, *69*). In addition, images that were obtained at a range of different defocus values are typically combined to compensate for the information loss at the defocus dependent zero crossing of the CTF curve.

1.2.5 CryoEM - a tool for protein structure determination

The combination of low contrast obtained from biological material in a TEM and the required use of low electron doses to prevent sample damage results in images with low signal-to-noise ratio (SNR) rendering image interpretation challenging (*71*). Instead, to make use of the high resolution information present, images of the same object can be combined and averaged to increase the SNR and to recover high resolution information of the imaged object. When a purified sample in solution is vitrified, individual particles are (ideally) randomly oriented in the frozen sample and a single image contains many particles at different orientations (Fig. 1.8). Identifying the respective orientations and recombining the different projection images, allows to

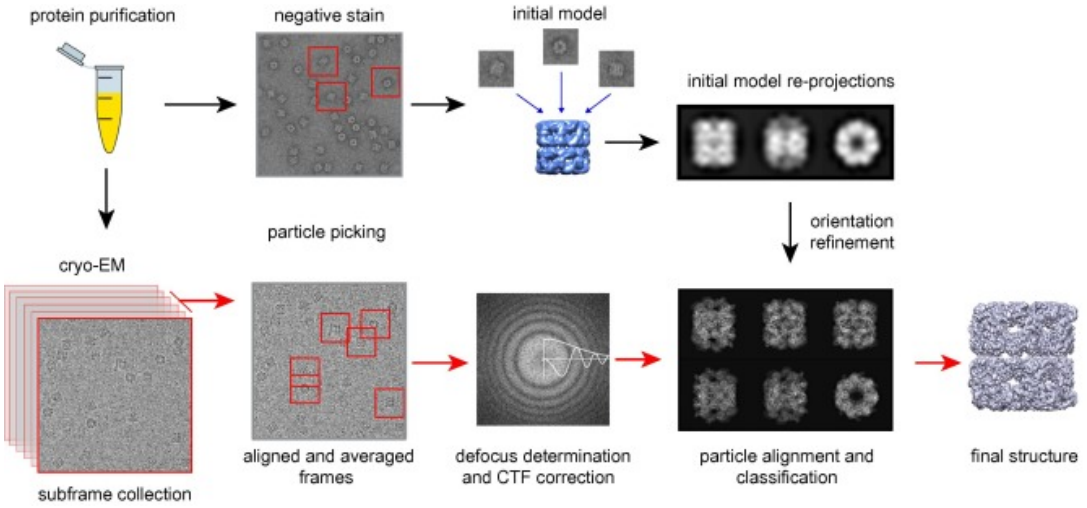

Figure 1.8 Overview of a cryoEM single particle workflow. A low resolution structure obtained from high contrast, negative strain images is used to generate an initial model. The model is used to identify different projection directions which are then compared to images obtained from cryoEM. The cryoEM workflow includes motion correction of DDD collected frame data, particle identification on the raw images, CTF determination and correction and the classification and alignment of all single particle images previously identified. 2D Class averages are compared to the projections of the initial negative stain model to refine final orientations for the reconstruction of the final structure. From (*70*) published under a CCBY4.0 license.

reconstruct a 3D volume of the original object as stated by the Fourier slice theorem (*72*). Therefore imaging a vitrified sample such as a protein in solution by cryoEM allows to reconstruct the 3D structure of protein from 2D cryoEM images. Upon acquisition of the data at high magnification conditions, obtaining the 3D structure of the sample requires the identification of single particles from each image, the alignment and averaging of particles of the same orientation to increase the SNR, the identification of each orientation and the reconstruction of the 3D volume from the different projection averages (Fig. 1.8) (*73*). In order to overcome structural heterogeneity in the sample, multiple, slightly different 3D volumes are reconstructed and further refined during 3D classification. The obtained 3D structures are evaluated by determining their resolution by calculating the Fourier Shell Correlation (FSC) from two independent structures of the same protein, typically obtained by processing two half sets of the data independently. The number of single particles that contributed to a structure is linked to the increase in SNR and therefore directly influences the achievable resolution (*74*). The completeness and isotropy of orientational sampling typically as a result of bias towards preferred orientations within the vitreous ice layer further affects the quality and resolution of the final structure (*75*, *76*).

In the last ten years several improvements of hardware in particular the introduction of DDDs (*57*) and better microscopes as well as software technology such as CTF correction (*68*, *69*), computational drift correction (*58*) (see section 1.2.2) and more accurate theoretical descriptions of the imaging process (*77*) were made and have led to a constant increase in the achievable resolution of protein structures obtained by cryoEM. Today, resolutions acquired by cryoEM and single particle processing frequently range below 3 Å which allows for de-novo protein structure determination. This has made cryoEM a popular structural biological tool applicable to proteins which are hard or impossible to crystallize such as large multi-subunit complexes, for which high resolution structural models could not be determined previously (*78*).

However, determining the structure of small proteins <100 kDA remains challenging. In addition, single particle cryoEM relies on the presence of isolated, homogeneous particles and is therefore not applicable to study proteins in a larger context such as within cells or pleomorphic particles. Targeting the structure of proteins in their native context is however achievable by combining cryoET with data processing strategies of subvolume averaging which rely on similar principles than single particle cryoEM and will be explained in more detail in the next section.

1.2.6 Cryo-electron tomography

In contrast to a 2D projection image, a tomogram allows to directly assess the 3D relationship of different features of the sample. An electron tomogram is based on acquiring multiple images of the same sample area by tilting the sample inside the microscope to a range of specified tilt angles. The acquired images are referred to as a tilt series. Based on the Fourier Slice Theorem, information from different projection directions, which correspond to different tilt images, can be combined to reconstruct a volume representation of the imaged object termed a tomogram (Fig. 1.9) (*72*).

The Fourier Slice Theorem further states that a 2D projection image represents one slice through the 3D Fourier transform of the imaged 3D object. The slice is orthogonal to the projection direction. The thickness of the slice is inversely proportional to the sample thickness. The limit of complete information transfer k_C in a tomogram according to the Crowther criteria is therefore proportional to the sample thickness and the sampling of different projections as illustrated in figure 1.9 (*79*). Frequencies beyond k_C can provide additional information but typically provide a lower SNR. Due to the design of current sample holders and the slap geometry of the sample, tilting is limited to a maximum of 70°. The missing projections at higher tilt angles represent missing slices in the 3D Fourier volume and together appear as a missing wedge in Fourier space (Fig 1.9). The missing wedge causes distortions of the reconstructed

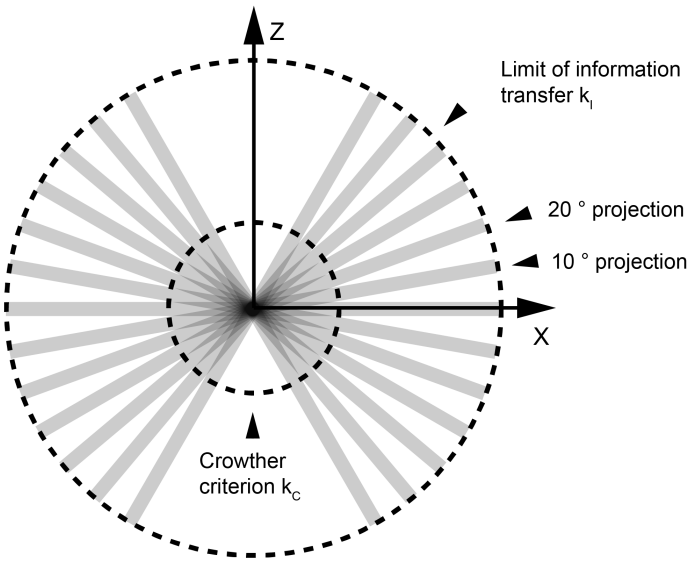

Figure 1.9 Illustration of the information transfer in tomography. Representation of a tilt series in Fourier space in 2D. According to the projection theorem, each slice or line shown here corresponds to a projection image for a different projection direction. Projection directions are sampled by tilt angles with 10° spacing. k_C indicates the limit of information transfer according to the Crowther criteria (*79*). The absolute limit of information transfer k_I is determined by the microscope and the imaging conditions. Adapted from (*80*).

volume including elongations in the direction of the missing information (*81*). The reconstruction of the tomogram from the acquired tilt series can be performed using different approaches such as back projections, direct Fourier inversion (*74*) and iterative algorithms such as the Simultaneous Iterative Reconstruction Technique (SIRT) (*82*). All tomograms reconstructed in this thesis have been reconstructed using the weighted back-projection (WBP) method. During WBP the pixel information of one region of the sample from all 2D tilt images are projected along the projection direction back into a 3D volume.

Two different types of ET can be distinguished which will be referred to as room temperature ET and cryoET. In room temperature ET, samples such as mammalian cells are high-pressure frozen and subsequently embedded into a plastic resin by freeze-substitution. The presence of the resin conserves cellular and subcellular structures. To enhance contrast for subsequent EM imaging, heavy metal salts just as used in negative staining are added. Subsequently the resin block is sectioned in thin slices for TEM imaging. While this method has been successfully used in the past to elucidate membrane structures, organelle architecture and general morphological features inside cells it fails to provide details on the molecular level (*83*, *84*).

In contrast and due to the excellent sample conservation in vitreous ice, cryoET allows studying molecular details in 3D of the chosen sample. Sample preparation for cryoET is similar to sample preparation for single particle cryoEM. Thin specimens, are typically vitrified by plunge freezing (Fig. 1.7). Small targets for thin samples include pleomorphic assemblies of proteins or membranes and proteins, such as heterogeneous, large protein complexes, enveloped viruses, cellular vesicles, small bacteria or bacterial mini cells among others (*81*, *85–87*). Medium thick sample, up so size of a single mammalian cell, can be vitrified by plunge freezing while thicker samples can get vitrified through high pressure frozen. Any type of thicker sample requires further thinning prior to imaging which can be performed either by cryo focused-ion beam (FIB) milling or cryo-sectioning (*81*).

In comparison to images acquired for subsequent single particle processing, the electron dose per tilt image has to be further reduced since multiple images of the same sample area have to be acquired. The total electron dose that is typically spent on a single image acquire for the purpose of single particle data processing (80 $e\,\text{Å}^{-2}$) has to be distributed across the complete tilt series to prevent degradation of the sample before the complete tilt series is acquired. CyoET tilt series acquired in the context of the work for this thesis, consisted of 41 2D images with a tilt angle spacing of 3° which leads to an 41-fold reduction in signal per image in comparison to a single particle image if the same total dose was used. The low contrast in the individual images of a tilt series represents significant challenges in subsequent processing steps such as image alignment. To facilitate accurate tilt series alignment despite low contrast, spherical gold particles can be added to the sample if possible to serve as high contrasty alignment markers which can be traced across the tilt series. The use of so called 'low dose methods' assures that images which are needed for autofocusing and tracking during the acquisition of a tilt series are acquired on positions next to the actual imaging position rather that exactly at the imaging positions. This procedure saves valuable electrons for increasing the contrast in each tilt series image. The acquisition of images at higher tilt angles leads to decreased image quality due to an effectively increased sample thickness in the direction of the beam. Consequently, the amount of inelastic scattering is increased for images with high tilt angles which leads to a further reduction in SNR for these images. Therefore, cryoET in particular benefits from high electron voltages and the use of an energy filter to reduce background signal from inelastic scattering (*81*, *85*).

1.2.7 Subtomogram averaging

CryoET can be combined with subvolume or subtomogram averaging to retrieve higher resolution information of a repetitive feature in the tomogram. Just like single particle processing, the benefit of subtomogram averaging relies on a significant increase in the SNR based on computationally aligning and averaging subvolumes with the same information content. The computational procedures involved in subtomogram averaging are very similar to the procedures involved and explained for single particle cryoEM data processing (see section 1.2.5). Starting from a reconstructed tomogram, the subtomograms at the position of the feature of interest are extracted and subsequently aligned to each other in 3D, before subvolumes are combined into one average. The increased SNR of the average allows to gain insights into higher resolution details relative to what is already visible in the tomogram. The subtomogram averaging workflow used for the work presented in this thesis relies on scripts which have been derived from the AV3/TOM subtomogram averaging packages (*80*, *88*). In the following I will briefly introduce the principles of subtomogram averaging pipeline using the conventions and names as described in these packages (Fig. 1.10): During alignment, all subtomograms are individually compared to the average generated from the aligned particles of the previous iteration. In the first iteration, a low pass filtered model, the non-aligned average or an artificially generated volume can be used as reference. To compare the subtomograms to the reference, the reference is tilted and rotated along the three main axes in a range which is set by the user. Rotations are described by the three Euler angles ψ, θ and ϕ. For each orientation the reference is compared to all subtomograms using a constrained cross correlation function (CCF). Constrained describes the masking step to remove the information affected by the missing wedge prior to calculating the cross correlation map. In addition low pass and high pass filters are set, to prevent noise alignment. Finally, the user defines a mask to define the region of the reference and the subtomogram to be considered during alignment. Positions and orientations with the highest cross correlation score are stored for the next iteration and used to calculate the final average. While the description of the alignment workflow follows the AV3 package, many other software packages for subtomogram averaging rely on similar principles (*89–91*). An overview of different subtomogram packages can be found in (*85*).

Concomitantly with the methodological improvements in single particle cryoEM, obtainable resolutions for structures studied by subtomogram averaging have increased in the past years. Recently, resolutions around 5 Å which allowed building a high resolution model into the obtained density maps were reported for the first time (*92–*

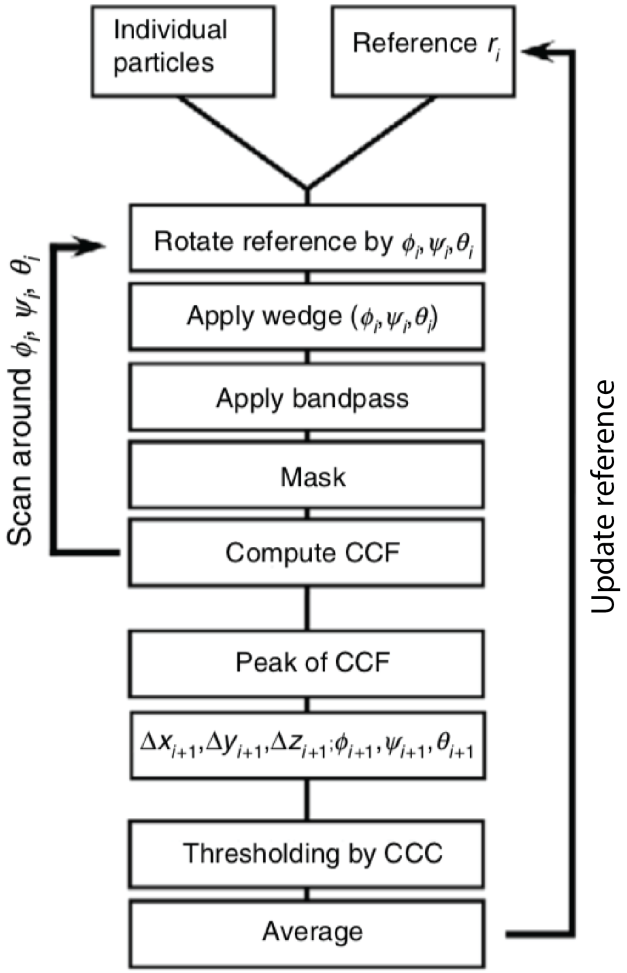

Figure 1.10 Diagram of the subtomogram averaging pipeline used in this study. Summary and order of computational steps used during subtomogram averaging starting from the extracted subvolumes. Nomenclature used in this scheme corresponds to the AV3 software package. More details for each step are provided in the text. Adapted from (*80*) with permission from the publisher.

94).

Multiple efforts have contributed to this improvement. Progress in hardware technology including the development of DDDs to improve the cryoEM image quality concomitantly has led to an improvement of tomography and subtomogram averaging methods. The development of a dose-symmetric tilt scheme has proven to increase tomogram quality in particular for subsequent subtomogram averaging (*95*). A dose-symmetric tilt scheme requires large and frequent stage movements between images which causes drift and shifts which have to be constantly checked for during tilt series acquisition in order to ensure high quality data. Therefore software to control and automate microscopes such as serialEM (*96*), Leginon (*97*) among others have been crucial to improve the quality and the automation of tilt series acquisition. On the computational side, methods which were already successfully employed in single particle data processing were transferred and adjusted for the challenges of tomography and subtomogram averaging including more accurate CTF determination and correction, including 3D-CTF correction, (*93*) as well as exposure based dose filtering (*92*, *98*).

Despite the discussed improvements, resolving amino acid side chains by subtomogram averaging for a wide variety of samples remains rare till today. Instead, most subtomogram averaging structures, give important and unique insights into the 3D distribution of proteins and the in situ arrangement of domains, subunits and secondary structural elements. Here, the structural interpretation still relies on high resolution models provided by complementary methods such as X-ray crystallography, theoretical modelling or mass spectrometry. (*87*, *99*, *100*)

1.3 Scope of this study

Protein structure provides critical information to understand protein function. Structural insights in the past were only available for samples which can be targeted by X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy such as smaller proteins and non-enveloped viruses. Recently, the development of cryoEM and single particle data processing methods has broadened the spectrum of targets for structure determination. Nevertheless, the structure of enveloped, pleomorphic particles, such as influenza virus is not accessible to single particle data processing since particle shape and size are not homogeneous. Influenza virus is a dangerous pathogen which evolves rapidly and thereby can create new strains which can be infectious enough to start a pandemic such as the 1918 Spanish flu. Current strategies

to prevent infections and ultimately the emergence of new pandemics, rely on vaccinations with limited efficiency and the need for annual refreshment to maintain partial immunity. Understanding the molecular structure of the virus is therefore essential in order to overcome our current limitations in prevention and treatment of the pathogen. Describing critical processes such as the infection of a host cell as well as the formation of new virus particles are impossible without the knowledge of the structure of the virus. Using TEM and later cryoEM to visualize enveloped viruses including influenza virus has therefore revolutionized our understanding of the virus. The obtained EM images in combination with the structural models of virus proteins allowed for the first time to generate a model of influenza virions represented as a spherical, membraneous particle with a dense surface layer of glycoproteins. In addition, only from EM images the variability of size and shape of influenza virus particles became accessible.

However, the projection image characteristics of 2D EM images have limited the correct identification of all virus components in 3D. Therefore the use of tomography for studying influenza virus helped to further increase the knowledge of the 3D arrangement of the different components within influenza virus particles. Important insights from tomography include the asymmetric distribution of the two proteins HA and NA, the arrangement of the eight genome segments at virus tips as well as the confirmation of the presence of a regular protein matrix underneath the virus envelope formed by M1. However, further details about the arrangement and details about the M1 orientation and M1-M1 interaction interfaces remain elusive. In particular the small size of M1, 28 kDA, and the respective diameter of a M1 subunit, 4 nm for the crystallized M1 NTD, is at the limit of what can be directly assessed in a tomogram. The combination of cryoET data with the data processing technique of subtomogram averaging helps to overcome the limits of tomogram interpretation by averaging repetitive features in the tomogram.

Despite the continuous improvement of the understanding of influenza virus structure, multiple questions in particular about the structure and arrangement of M1 within the virus remain open. In addition, accurate and quantitative insights of the arrangement and structural variability of the glycoproteins on the virus surface remain missing to date. Due to their high abundance within virus particles M1 and HA are both possible targets for subtomogram averaging. To approach the outlined open questions about influenza virus structure I set out to make use of the unique characteristics of cryoET in combination with subtomogram averaging to obtain structural insights of HA and M1 directly from influenza virus A particles.

Chapter 2

Material and Methods

2.1 Mammalian cell culture

MDCK cells

Madin-Darby Canine Kidney (MDCK) cells (ATCC: CCL34) were cultured in Dulbecco's Modified Eagle Medium (DMEM) Glutamax (Gibco, ThermoFisher) supplemented with 10 % Fetal Bovine Serum (FBS) (Gibco, ThermoFisher), 1 % Penicillin/Streptomycin (PS) at 37 °C, 5% CO_2 and 100 % humidity. MDCK cells were always maintained at less than 80 % confluence. To split cells, MDCK cells were washed with pre-warmed Phosphate Buffered Saline (PBS) before pre-warmed 0.05 % trypsin (Gibco, ThermoFisher) was added. Cells were incubated in the presence of trypsin for 20 min at 37 °C. When cells were detached, prewarmed media was added. Cells were subsequently washed once by centrifugation for 3 min at a maximum of 500 x g followed by replacement of supernatant with fresh media.

HEK293T cells

Human Embryonic Kidney (HEK)293T cells (ATCC: CRL-3216) were cultured as described for the MDCK cells in 2.1. In order to detach cells, incubation for 3 min at 37 °C was sufficient to achieve cell detachment. All other procedures and materials were the same.

Cell counting using an automatic cell counter

10 µL of diluted cells was mixed with 10 µL of 0.4 % tryphan blue stain (Invitrogen,ThermoFisher) and 10 µL of the mixture was added to a specific cell counting slide. The slide was inserted into the counting chamber of a CountessII Automated Cell

Counter (Invitrogen,ThermoFisher). Contrast was readjusted manually if necessary and the cell concentration was automatically determined from the counter.

2.2 Virus stocks and virus purification

2.2.1 Virus stocks

Virus stocks of A/Hong-Kong/1968/1/ (H3N2,HK68) and A/WSN/1933 (H1N1,WSN) influenza A virus (IAV) initially generated by reverse genetics were kindly provided by the lab of Hans-Georg Kraeusslich. For amplification of virus stocks, T75 flasks with 80 % confluent MDCK cells were washed twice with DMEM to remove residual serum. The virus to be amplified was diluted in 5 mL 0.3 % Bovine Serum Albumin (BSA) (fraction V, for cell culture, Gibco, ThermoFisher)/DMEM and cells were inoculated at a multiplicity of infection (MOI) < 1. After 1 h of infection, the inoculum was replaced with 5 mL 0.3 % BSA/DMEM containing 0.5 µg mL^{-1} to 5 µg mL^{-1} N-tosyl-L-phenylalanine chloromethyl ketone (TPCK)-treated trypsin (Sigma-Aldrich, Merck). 48 h post infection and when signs of cytopathic effect (CPE) were visible, the supernatant was collected and pre-cleared at 500 x g to remove cell debris. The pre-cleared supernatant was stored in small aliquots at $-80°$ until usage.

2.2.2 Quantification of virus titers

Quantification of virus titers of newly prepared virus stocks was achieved by plaque assays. The plaque assay at EMBL was performed by O. Avinoam and S. Nakatsu using the agar method and performed by L. Metskas and me at the LMB using the Avicel media as overlay (*101*). The main difference between these two protocols is the choice of the viscous substrate, while the remaining steps are the same. For the Avicel method MDCK cells were densely seeded into several 6-well plates. Cells were washed twice using 0.3 % BSA/DMEM and virus was serially diluted in 0.3 % BSA/DMEM to obtain different final dilutions and MDCK cells were subsequently inoculated with different viral dilutions in duplicates. During the first hour of infection the 6 well plates were moved every 15 min to assure a homogeneous distribution of virus particles. To prepare the Avicel overlay solution, the Avicel stock solution containing 2 mg Avicel in 100 mL water (kindly provided by R. Ulferts from the Beale lab at the Cambridge pathology department for a test experiment) was diluted 1:200 in DMEM containing 2 µg mL^{-1} TPCK-treated trypsin. Upon removal of the inoculum, 2 mL Avicel solution was added to each well and incubated for for 48 - 72 h without moving the plate. Following the incubation time, the Avicel overlay was removed

and the cell layer was fixed with 4 % Paraformaldehyde (PFA)/PBS for 30 min. After removal of the fixative, the cell layer was stained with 1 mL 0.1 % toluidine blue per well which was incubated for at least 20 min prior to rinsing the wells with water. Upon drying, the staining allows for the visualization of plaques which can be counted for the different virus dilutions to determine the original virus concentration. Titers of produced virus stocks used for the work described in this thesis were measured to correspond to $10^7 - 10^8$ plaque forming units (PFU) mL^{-1}.

2.2.3 Virus purification

Steps of infection and harvesting are the same as described for the amplification of virus stocks (section 2.2.1). Starting from the harvested supernatant, the supernatant was precleared by centrifuging for 5 min at 4° at 500 g to remove cell debris. During optimization of the purification protocol we tried adding an additional pre-clearing step of 10 min at 4° and 800 g. A sucrose cushion solution was prepared as 32.5% sucrose in TNE buffer (20 mM Tris-HCl (pH 7), 100 mM NaCl, 2 mM EDTA). 2 ml of 4° pre-cooled sucrose solution was carefully pipetted into a SW40i ultra centrifuge tube and subsequently overlaid with 10 ml pre-cleared supernatant and centrifuged for 90 min to 120 min at 25.000 - 30.000 rounds pre minute (RPM) at 4°. Tubes were carefully removed from the centrifuge and stored on ice. The supernatant as well as the sucrose solution were carefully removed and tubes were subsequently inverted and wiped dry. The pellets at the bottom of the tube were redissolved in 20 µL to 50 µL of TNE buffer, by gently shaking overnight at 4°. Dissolved virus particles from different centrifuge tubes were subsequently recombined. The dissolved virus particles were then mixed with the same volume of 2X fixative solution containing 8 % (PFA, EM solutions), 0.2 % glutaraldehye (GA, EM solutions) in PHEM buffer (600 mM PIPES, 250 mM HEPES, 100 mM EGTA and 20 mM MgCl), pH=7 and incubated on ice for 30 min. After 30 min glycine solution was added to a final glycine concentration of 2 mM to quench the fixative.

2.3 VLP plasmids

2.3.1 VLP plasmids

Plasmids used for VLP formation, encoding for the sequences of HK68 HA, HK68 NA embedded into pCAGGS vectors were previously described in (*45*). These two plasmids plus two additional plasmids encoding for HK68 M1 and HK68 M2 also cloned into pCAGGS vectors were kindly provided by the lab of Hans-Georg Kraeusslich.

2.3.2 VLP plasmid amplification

DH5α E.coli cells were thawed on ice and transformed with the respective plasmid by heat shock using 1 µL plasmid stock, which were kept at concentrations of 1 $\mu g\,\mu L^{-1}$. For heat shock transformation bacteria and plasmid were incubated on ice for 20 min and subsequently placed at 40 °C for 30 s and subsequently cooled on ice for 2 min. Subsequently 500 µL warm SOC media (super optimal broth (SOB) plus extra glucose) was added, mixed and immediately centrifuged for 3 min at 600 x g. Upon centrifugation 250 µL of the supernatant were removed from the tube and 50 µL to 100 µL of the remaining bacteria solution were plated on pre-warmed lysogeny broth (LB)/Ampicillin (AMP) (100 $\mu g\,mL^{-1}$) plates (prepared by LMB media kitchen) and incubated overnight at 37 °C. From plates with good colony density, single colonies were picked and pre-grown in 5 mL LB with 100 $\mu g\,mL^{-1}$ AMP at 37 °C for 12 h. If the culture showed visible growth, it was added to 200 mL LB-Amp to grow overnight at 37 °C. The next morning, bacteria were harvested and plasmid Deoxyribonucleic acid (DNA) was extracted using a commercially available kit for DNA preparation, (Maxi prep kit, Quiagen). The obtained plasmids were diluted to obtain a final stock solution of 1 $\mu g\,mL^{-1}$, aliquoted and new aliquots were kept at $-20°$. All newly prepared plasmids were checked by sequencing and initial expression tests by dot blot of western blot.

2.4 Sample preparation for cryoET

2.4.1 Preparation of grids for cell seeding

QF AU-200 mesh R2/2 grids (Quantifoil), stored in a closed box to minimize exposure to potential mammalian cell contaminants were glow-discharged (easy glow, Pelco) for 30 s at 20 mA to increase hydrophily and to sterilize grids. 10 nm gold fiducials in PBS were diluted 1:50 with PBS. Grids were incubated with 10 µL of diluted gold fiducials for at least 30 s before excessive liquid was removed using a piece of Whatman No. 1 (Whatman) blotting paper. A 6-well plate was prepared with 2 mL of 10 % FBS/1 % PS/DMEM Glutamax (Gibco). 4-8 grids were slid into a media-containing well while making sure, that they attach to the bottom of the plate. Grids were pre-incubated in media for at least 1 h before cells were seeded on grids.

2.4.2 Preparation of virus samples on grids

MDCK cells were trypsinized as described in section 2.1. Detached cells were counted as described in section 2.1. Prepared grids distributed in 6-well plates as described

in section 2.4.1 were used and the remaining media was removed carefully from each well to prevent dislodging of grids. 2 mL of diluted cells was added per well to seed a total number of $1.5X10^5$ cells/well and cells were incubated in the presence of EM grids at 37 °C 5% CO_2, 100% humidity for 12 h to 24 h until cells were well attached. In a biosafety level 2 cell culture room, cells on grids were washed with 0.3% BSA/10 mM HEPES in DMEM before inoculation with diluted virus. Virus was diluted to infect at an MOI<1. Virus stocks HK68 or WSN stored at −80 °C were freshly thawed and diluted in 1 mL infection media before inoculating the cells attached to EM grids. Virus and cells were incubated together for 1 h and mixed every 15 min. After 1 h, the inoculum was removed and replaced with 0.3 % BSA/DMEM containing 0.5 µg mL^{-1} to 5 µg mL^{-1} TPCK-treated trypsin, depending on trypsin activity, to ensure HA cleavage of newly formed virions. Incubation was stopped when CPE was visible, typically 24 h to 48 h post infection. If the sample was subsequently transferred into biosafety level 1 areas the virus was inactivated using 4 % PFA 0.1 % GA in 1X PHEM buffer for at least 30 min and subsequently washed 2 X with PBS prior to plunge freezing of grids.

2.4.3 Plunge freezing of grids with mammalian cells

Grids were plunge frozen using a LeicaGP cryo plunger (Leica). Upon pre-cooling with liqiud nitrogen, liquid ethane was kept at −182 °C. Humidity of the sample chamber was set to 95 % and temperature was set to 20 °C for inactivated samples and to 37 °C for living cells. 5 µL of 1:3 diluted 10 nm nm PAG-gold solution in PBS was added to the grids inside the sample chamber and prior to blotting. Blotting times varied between 1.5 s to 3 s. Upon moving to the LMB, grids were initially plunge-frozen using a Vitrobot Mark 4 (ThermoFischer) with the blotting number and blotting time set to 0 in order to disable automatic blotting. Instead, a small piece of Whatman No. 1 filter paper hold by a pair of forceps was manually inserted into the sample chamber to achieve single sided blotting. Manual blotting was performed for 10 s to 15 s while all other parameters remained the same as described above.

2.4.4 Preparation of VLP samples on grids

HEK293-T cells were trypsinized as described in 2.1. Cells were counted as described in 2.1. 400.000 cells were seeded on prepared grids (see 2.4.1) per well of a cell culture 6-well plate and incubated 15 h to 20 h before transfection. Shortly before transfection, media was replaced by 1 mL Optimem (Gibco,ThermoFisher) per well. The Transfection was performed using the lipofection agent Fugene (Promega) ac-

cording to the manufacturer's protocol. A total amount of 3.3 µg plasmid DNA was diluted in 150 µL Optimem (Gibco), before 9.9 µL FUGENE was added. DNA and FUGENE were mixed and incubated for 15 min before the mixture was added to each well drop-wise. For HA, NA, M1 and M2 VLPs, plasmids were added using the following ratio: 1:1:2:0.5 for HA, NA, M1, M2 respectively. The recommended removal of the transfection mixture after 5 h did not lead to an improvement of transfection yields or cell health and the step was omitted to reduce handling of the dishes containing EM grids to minimize the risk for grid movement inside the well. 48 hours post-transfection, grids with cells and VLPs were removed from the 6-well plates and directly plunge frozen as described in 2.4.3.

2.5 CryoET

2.5.1 CryoET data collection

All tomograms were collected on a Titan Krios cryo-electron microscope (ThermoFisher, former FEI) equipped with an in-column energy filter (Gatan) and using an K2 direct electron detector (Gatan) either at the EM facility at EMBL Heidelberg or at the facility at the MRC LMB in Cambridge. Grids of pre-screened samples were, clipped into cartridges and subsequently transferred into the loading cassette using the corresponding loading station. The cassette was loaded into the microscope and grids were transferred into by the microscope's auto loading system. All loaded grids were initially mapped automatically at a low magnification of 135x using the gridmap function in serialEM (*96*). The best grid was selected for subsequent imaging by medium magnification mapping. Montages of manually selected grid squares were automatically acquired in serialEM at a magnification of 2250x, a defocus of -100 µm and with an objective aperture of 50 µm inserted and centred. As part of the grid square mapping routine, eucentric focus was adjusted for each grid position. At the magnification used for grid square mapping, virus and VLP filaments can be identified. From the acquired montages, positions for tilt series acquisition were selected manually. Sample thickness, the absence of cells and ice contaminations in the surroundings and relative to the tilt axis were taken into account when selecting positions. Prior to data acquisition, the microscope was fine-tuned for the final imaging parameters. Coma and astigmatism was determined and corrected by taking images according to a Zemlin table (*102*) using a routine implemented in the program AutoCTF (ThermoFisher, formerly FEI). Subsequently, a 70 µm sized objective aperture was inserted. The exact magnification used was variable for different microscopes and data sets and was ranging between 81.000 and 103.000. The resulting pixel sizes for all

Table 2.1 Summary of data collection conditions of all datasets described in this thesis
'puri' indicates that the virus was purified for this sample, * indicates that the sample was inactivated by chemical fixation. Df=Defocus, Super Res=Super Resolution mode, Tilt incr=Tilt increment, TS=Tiltseries

	HK68puri	WSNpuri	HK68-I*	VLP-I	Udorn-I*	VLP-II*	HK68-II
Pixel Size (Å)	1.78	1.78	1.78	1.78	1.01	1.54	1.7
Total Dose ($eÅ^{-2}$)	145	140	150	120	100	140	140
Tilt Range (°)	-66/66	-66/66	-66/66	-60/60	-64/64	-60/60	-60/60
Tilt incr (°)	3	3	3	3	2	3	3
Super Res	yes	yes	yes	yes	no	no	no
Counting	yes	yes	yes	yes	yes	yes	yes
Df Range (µm)	-4/-6.5	-4/-6.5	-4/-6.5	-2/-6.5	-1/-2.5	-2/-5	-2/-4
Df Step (µm)	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Frames	10	10	10	10	10	20	20
TS number	60	60	40	60	30	20	100

datasets are summarized in table 2.1. All tilt series were acquired in nano probe and Energy-filtered transmission electron microscopy (EFTEM) mode with an energy filter slit set to 20 keV. The energy filter was tuned and the zeros-loss peak centred prior to data acquisition using tools provided by the Digital Micrograph program (Gatan) which controls the energy filter and the detector. Based on the desired total dose per tomogram, the dose per tilt and the total exposure time per tilt and per frame were calculated based on the dose measured in an empty area of the sample. The total dose and number of frames are summarized in table 2.1. After the manual selection of positions, tilt series were automatically acquired following a dose-symmetric tilt scheme (*95*) using serialEM (*96*) scripts provided by Wim Hagen. During tilt series acquisition autofocus and tracking was automatically executed as part of the script on manually selected areas in the surrounding of the imaging area and positions along the tilt axis.

2.5.2 Data preprocessing and tomogram reconstruction

Frame alignment and motion correction for all tilt series were performed by using the Alignframes tool in IMOD (*103*), either implemented as the serialEMCCD plug-in (*96*) for on-the-fly frame alignment during data collection or after data collection. If data were collected in super resolution mode the frame size was reduced by Fourier cropping. Aligned frame stacks were sorted by tilt angle using scripts written in Matlab (MathWorks) or usign IMOD's sort by tilt functionality (*103*). Tilts without signal were removed manually and one image stack per tomogram was generated using IMOD's newstack tool (*103*). Tilt images were exposure filtered according to the accumulated dose applied to the sample for each tilt using Matlab (MathWorks)

or more recently using the Alignframes filtering per dose function in IMOD. In both cases, exposure filtering was implemented as described in (*104*). The CTF for each tilt image was determined on non-exposure filtered tilt stacks using CTFFIND4 (*69*) and evaluated for each image. If data were 2D CTF corrected, CTF correction was performed in IMOD by phase flipping (*103*). If data were 3D CTF corrected, data were corrected using NovaCTF by CTF multiplication (*93*). Tilt series were aligned using IMOD (*103*). X-rays and high intensity pixels were removed, tilts series were aligned by cross correlation and gold fiducials were either automatically seeded and corrected manually on the zero tilt image. Beads were automatically tracked. Tilt seris were aligned based on the tracked gold fiducals, which were manually refined until the reported residual from the alignment model through all selected beads was lower or equal to the pixel size or until the residual did not improve further. All tomograms were reconstructed in IMOD by weighted back projection.

2.6 Subtomogram Averaging

Subtomogram averaging and subtomogram averaging related tasks were performed using Matlab (MathWorks) scripts based on the TOM and AV3 subtomogram averaging packages (*88*) and adapted and modified by me and other current and previous group members, in particular William Wan and Dustin Morado.

2.6.1 Subtomogram Averaging of HA

Initial positions and orientations of subtomograms

The central spline of each virus or VLP filament found in tomograms was manually traced either using the Volume Tracer function in UCSF Chimera (*105*) or in AMIRA (ThermoFisher, formerly FEI). Based on the manually determined positions a spline function was generated in Matlab (MathWorks). Using Matlab, the central spline for each filament was used to generate a tube based on the HA-HA diameter manually determined for each filament. Positions along the surface of the tube with spacings of 5 nm were used as the initial positions for subtomogram extraction. The orientation of the normal vectors at these positions along the tube surface were used to generate initial Euler angles for the generation of the first average.

Subtomogram alignment

All alignments were preformed without an external reference. An initial reference was generated from a subset of the data for each dataset individually. This generated

Table 2.2 Summary of dataset sizes processed by subtomogram averaging for HA Numbers of subtomograms and asymmetric units are final numbers per half set
,

	HK68-I*	VLP-I	Udorn-I*	VLP-II*	HK68-II
Pixelsize (Å)	1.78	1.78	1.01	1.54	1.7
Tomograms	33	60	28	12	24
Viruses	34	69	28	20	30
Subtomograms	15,000	45,000	6200	12,000	19,000
Symmetry	3	3	3	3	3
Asymmetric units	45,000	135,000	18,600	36,000	48,000
Resolution FSC(0.143) (Å)	9	8	9	9	9

reference was subsequently used for the alignment of the full dataset. Alignment masks during all stages of alignment were shaped as cylinders of different dimensions. The diameter of the cylinder was reduced over the number of iterations. Starting from wide cylinders which include a central HA protein and density of its neighbours down to cylinders which match the dimensions of a single HA protein. Masks were generated using the dynamo mask tool (*89*) and mask edges were smoothed by a Gaussian signal drop off. Initial and final number of subtomograms for each dataset processed by subtomogram averaging for HA are indicated in the corresponding results sections (see section 3.2) and summarized in table 2.2.

FSC calculations and sharpening

All datasets were split into two half sets based on odd and even subtomogram positions. The and two half sets were processed independently after initial Bin4 alignment for which reference and subtomograms had been low pass filtered to 35 Å. The FSC was calculated in Matlab (MathWorks) between the two masked EM maps using the method described in (*106*). Masks were created to be slightly smaller than the final alignment mask and mask edges were smoothed. The reported resolution was obtained from the 0.143 or the 0.5 threshold of the FSC curve. B-factors were empirically determined as indicated for each presented structure and maps were sharpened using the indicated B-factors and filtered to the determined resolution as described in (*107*).

2.6.2 Subtomogram Averaging of M1

Initial positions and orientations of subtomograms were calculated just as described for HA in section 2.6.1. Radii which were used to calculate the cylinder for M1 subtomogram extraction were adjusted to match the M1-M1 rather than the HA-HA

Table 2.3 Summary of data processed by subtomogram averaging for M1. Numbers of subtomograms and asymmetric units is per half set. geo.cl.=geometric cleaning.

	HK68-I*	VLP-I
Pixelsize (Å)	1.78	1.78
Tomograms helix parameters	21	20
Tomograms combined for final structure	4	4
Subtomograms of combined tomograms before geo. cl.	13,200	15,900
Subtomograms of combined tomograms after geo. cl.	6000	3800
Symmetry	1	1
Resolution FSC(0.143) (Å)	8	8

diameter for each filament. Initially subtomograms were extracted at oversampled positions from Bin4 tomograms. Initial postions were generated with a spacing of 20 Å, 2 Bin4 pixels, corresponding to half the width of a M1 NTD monomer, approximately 40 Å. Depending on the orientation of the VLP or virus filament, this led to a generation of 28,000 -35,000 initial subtomogram positions were filament. Subtomograms were extracted and averaged separately for each filament. Subsequent alignments were equally performed individually for each filament in parallel due to the heterogeneity of the references obtained per filament. Alignment masks were generated as cylindrical volumes in dynamo (*89*) with varying diameters and smoothed edges. Throughout Bin4 processing, masks were almost as wide as the extracted box to include as much signal as possible. During Bin2 processing, masks were generated as volumes in Matlab to match the curvature of the filament. The width of the mask was continuously decreased to attempt to focus the alignment on individual M1 monomers. However, the signal from an individual M1 monomer was to low, so that until the end the mask included six M1 monomers. However, masks used during later stages of alignment were generated to mask only the emerging inner and outer lobe of the structure and alignment was run in parallel for a mask that includes both lobes, a mask for the outer lobe and a mask focusing on the innner lobe, with the same alignment parameters otherwise. For more details see also section 4.2.1.

FSC calculations and sharpening

For each final dataset combined from four tomograms two equal sized half sets were generated by splitting all subtomograms positions into top and bottom for each virus or VLP filament. The two half sets were processed independently after initial Bin2 alignment. FSCs between the two masked EM maps were calculated in Matlab (MathWorks) as described in (*106*). Masks were created to be slightly smaller than the final alignment mask and mask edges were smoothed. The reported resolution was

obtained from the 0.143 or the 0.5 threshold of the Fourier shell correlation (FSC) curve. B-factors were empirically determined as indicated for each presented structure and maps were sharpened using the indicated B-factors and filtered to the determined resolution as described in (*107*).

2.7 Downstream data analysis

All additional data analysis and if not indicated differently was performed in Matlab (MathWorks). All fitting of structural models into density maps was performed in UCSF Chimera (*105*) using the fit in map routine.

Chapter 3

Structural studies of IAV hemagglutinin in situ

3.1 Introduction

3.1.1 HA in the virus life cycle

The viral envelope of influenza viruses is decorated with the viral glycoproteins HA and NA of which HA is the most abundant one. The transmembrane protein HA forms a homotrimer. The 12 nm long extracellular domain points away from the virion surface (Fig. 3.1A). The ectodomain can be divided into a membrane proximal stem and a membrane-distal head region. The stem region connects to the transmembrane domain (TMD) of HA via a flexible linker. The TMD comprises a single alpha helix (*108*) and TMD extends into a short cytoplasmic tail to mediate interactions with proteins at the inner side of the membrane. Based on the high mutation rates of influenza 18 different types of HA have been identified until today. While there are significant differences on the sequence level, all types still provide a relative high level of structural homology. HA variants have been categorized into two groups based on their structural similarity. Group 1 comprises H3, H4, H7, H10, H14, H15 while H1, H2, H5, H6, H9, H8, H11-13, H16-18 are summarized as group 2 (*109*). HA is initially expressed as the HA0 precursor polypeptide which is subsequently cleaved by host-cell proteases into two chains, HA1 (58 kDa) and HA2 (26 kDa), that remain connected via disulfide bonds (*110*). HA cleavage is essential to allow the protein to undergo a drastic conformational change under the condition of low pH and in preparation for membrane fusion. Structurally, HA2 comprises the large central helix C, and a shorter helix A. Both are connected by the connective loop B (Fig. 3.1B). HA1 consists mostly of parallel beta sheets connected by loops and small

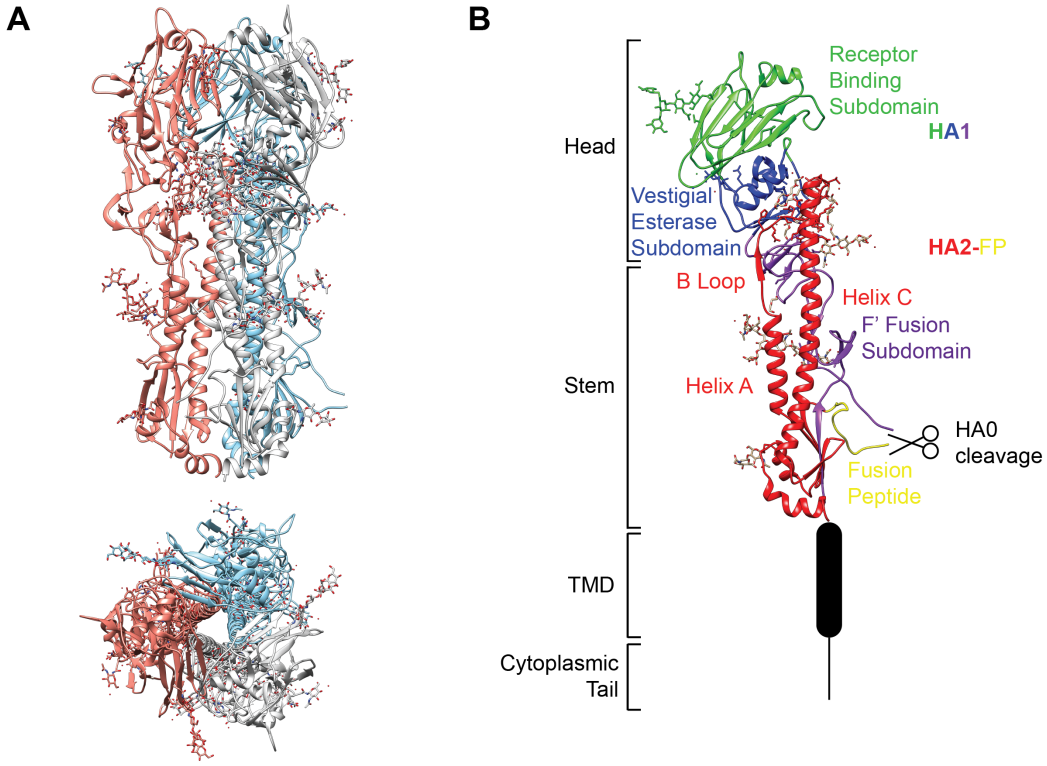

Figure 3.1 Crystal structure of HA. A) HA trimer of the H3 subtype in side view on the top and top view on the bottom. The three protomers are coloured in red, blue and white. B) Single HA monomer in side view with the different regions and domains coloured as indicated. HA2 (red) plus fusion peptide (yellow). HA1 contains the receptor binding site subdomain (green), the vestigial esterase subdomain (blue) and the F' Fusion subdomain (purple). The structure shown corresponds to HA from the H3N2 HK68 virus. PDB: 4WE4.

helices. A large part of HA1 forms the head region of HA. The top of the head region forms the receptor binding subdomain (RBS), in form of a shallow ligand binding pocket. Adjacent to the RBS and as lower part of the head region sits the so-called vestigial esterase subdomain and which is named based on the structural similarity to the esterase domain of the hemagglutinin-esterase-fusion glycoprotein of influenza C and D virus (Fig. 3.1B). The fusion subdomain of HA1 represents the only part of HA1 located in the stem region.

During the virus life cycle (see section 1.1.3) HA has multiple functions. Firstly, HA initiates interactions with host cells. The HA RBS binds to sialic acid molecules at the end of glycan chains decorating host cell surface receptors. HA variants differ in their preference for different sialic acid - galactose linkages. Predominant types of linkages depends on tissue type and species. While H5 favours $\alpha(2,3)$-sialic acid linkages, which occur widely in cells originating from birds, H3 preferentially binds to $\alpha(2,6)$ linked sialic acid, which is widely present in human lung epithelia (*111*). The difference in sialic acid linkages contributes to host specificity. The interaction between HA and sialic acid induces endocytosis of the bound virus. Upon endocytosis, the virus is exposed to the acidic interior of the endosome. Cleaved into HA1 and HA2, HA is sensitive to pH and exposure to low pH induces a major conformational change into the fusiogenic form. It was proposed that the HA1 head region dissociates from HA2 and becomes mostly unstructured. The B-loop connecting helix A and helix C transitions into a helix, and the complete HA2 chain turns into one long helix (extended intermediate, Fig.

3.1.2 HA as a vaccine target

During an influenza infection the dense glcyoprotein layer formed predominantly by HA is exposed to the host's immune system. Epitopes located on the HA surface induce the formation of neutralizing antibodies (ABs) by B-cells (*115*). Therefore, HA is also the most commonly used antigen for vaccine development and consequently all vaccines in use today rely on the recognition of HA. Currently, three different types of vaccines exist: inactivated virus, live-attenuated virus and recombinant HA vaccines. Inactivated virus is the most widely used variant which is produced be growing high titres of virus in eggs followed by harvesting, inactivation though the addition of formaldehyde or β-propiolactone and subsequent purification (*10*).

Many epitopes are found in the head region of HA and provide AB interaction interfaces with and around the RBS. While the RBS itself is rather conserved across

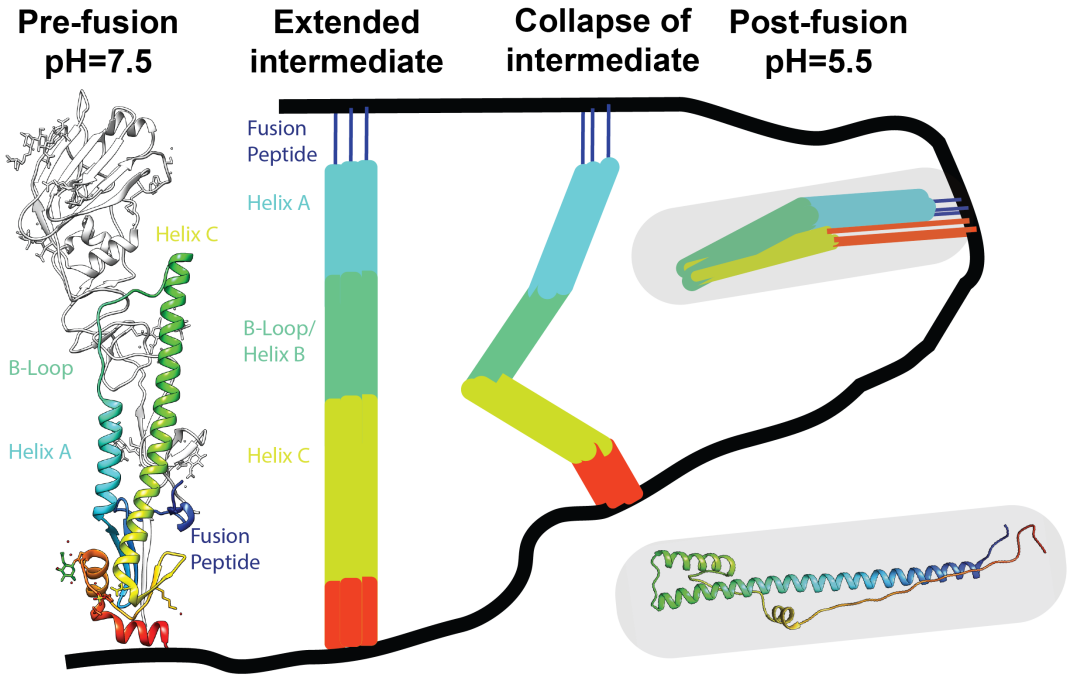

Figure 3.2 Schematic of a HA-mediated membrane fusion model. On the left: crystal structure of HA at neutral pH. HA2 is coloured in rainbow. PDB:4WE4. The black line represents the viral and the endosomal membrane. In the middle: A model of conformational changes of HA2 during membrane fusion (*112*, *113*). In the schematic, HA2 regions are coloured according to the colours in crystal structure. On the right: Schematic of the post-fusion conformation based on the low pH crystal structure of HA acquired at low pH (PDB:1htm) (*114*).

strains, regions around the RBS are hypervariable and in many cases crucial for efficient AB binding. Individual point mutations can be sufficient to prevent successful recognition by ABs. Consequently, the high antigenic variability of the head domain generated over time is one major reason for the lack of persistent immunity and the resulting re-occurrence of influenza infections in the same individual. Given the fact that current vaccines mimic an infection with one or several strains they come with the same limitations as our endogenous immune response. Essentially, vaccine-induced ABs are equally incapable of recognizing viruses that have been circulating and mutating over time. Therefore, current vaccination strategies rely on annual re-priming of the immune system. To overcome this limitation, current research is focusing on identifying broadly neutralizing antibodies that inhere the capacity to recognize HA across a broader range of, or ideally all, HA variants. In this context, ABs were identified that specifically interact with conserved residues of the RBS and thereby mimic the presence of SA in the binding pocket and which provide broader neutralizing efficiencies (*116–118*). Alternatively, possible AB interactions with the less accessible but more conserved HA stem region have been investigated (*119*).

For the characterisation of AB binding, high resolution structures as determined by crystallography or nowadays by single particle cryoEM are vital. On the other hand, AB binding in situ and in particular in vivo is more complex than interactions of two or three proteins. Stoichiometry as well as accessibility of antigens affect the interaction likelihood and thereby the neutralizing efficiency. Consequently, complementary insights from in situ methods are important in order to link structural models of AB binding to the more complex, native environment of AB binding on the virus surface.

3.1.3 Motivation

HA has several essential roles during the virus life cycle and is the most relevant protein of influenza virus for vaccine development. In the past, the structure of isolated HA proteins from different strains has been solved multiple times by X-ray crystallography. Insights into the structure of the protein were crucial to characterise antibody binding sites and to gain a better understanding of the function of the protein. While structures determined by crystallography give important insights, crystallization and protein isolation may alter the structure of a protein. For this reason, structures determined from protein crystals do not always represent the true structure on the virus surface. Further, different protein conformations of HA which might occur on the virus surface might not be captured by crystallization. While for a long time it was technically impossible to characterise structures in situ, the development of

high resolution tomography in combination with subtomogram averaging, nowadays enables structural characterisation of proteins in their native environment. This technical progress allows revisiting previously determined structures in their biological context. In addition, subtomogram averaging provides accurate information about the distribution and orientation of proteins on viruses. Understanding the glycoprotein arrangement on the virus surface is important to understand virus assembly and virus-host cell interactions. Further, insights into protein arrangement can link in vitro characterisation of AB-binding to the native context of an infection.

3.1.4 Aims of this study

To obtain a more complete understanding of the structural landscape and arrangement of HA on the virus surface we proposed to study the structure and arrangement of HA in situ on pathologically relevant influenza virus particles by high resolution cryoET and subtomogram averaging.

Specific aims of this project were to optimize and prepare samples for high resolution tomography of the human influenza strain HK68. Further, I aimed at establishing a robust data processing workflow which allows obtaining in situ structures by subtomogram averaging, ideally at sub-nm resolution. At this resolution, secondary structural elements become visible which allows detailed comparisons to existing structures determined previously. Due to biosafety regulations, virus samples require inactivation which is performed by chemical fixation, a method which potentially can effect protein structure. Virus-like particles (VLPs) represent a powerful, non-infectious system to control for potential effects of fixation. A well-characterised VLP system with sequences from the same strain relies on transfecting plasmids which encode for HK68 HA, NA, M1 and M2 (*45*). Therefore, I aimed at complementing structural studies of HA from infectious HK68 viruses, by structural studies of HA from samples of HK68 VLPs. This includes establishing a sample preparation workflow for HK68 VLPs which is suited for a subsequent characterisation by high resolution cryoET in combination with subtomogram average. By comparing in situ HA structures from infectious, inactivated viruses particles, non-infectious unfixed VLPs and previously determined crystal structure, I can assess effects of the virus and VLP sample preparation on protein structure. In addition to structural information, subtomogram averaging provides insights into protein arrangement. A direct comparison between the HA arrangement on the virus and VLP surface might further clarify if differences in the production of the two particles, such as the absence of the genome in VLPs, have any effects on the molecular arrangement and concomitantly, particle assembly.

3.2 Results

3.2.1 Sample purification affects virus particle quality

In the first part of the project I tested different approaches for preparing influenza virus particles to obtain samples suited for high resolution cryoET and subtomogram averaging.

I chose to work on the pathologically relevant strain A/Hong-Kong/1968/1/ (H3N2, HK68), which caused the 1968 Hong-Kong pandemic. As described in the section 1.1.2, influenza viruses found in patients have a predominantly filamentous morphology. The HK68 strain used here was previously shown to form long filaments in cell culture models (*45*). The filamentous morphology provides additional technical advantages for subtomogram averaging since a high number of protein copies is found per virus particle.

Typically, cryoET workflows to study virus architecture require purification of the target virus. Purification protocols involve chemical and mechanical forces to help concentrate and separate virus particles from cell debris and non-viral compartments. Any of these steps can potentially damage virus particles, particularly in the case of more fragile enveloped viruses such as influenza virus. Initially in this project, samples of HK68 were prepared for EM analysis according to standard protocols which include purification of the virus. Purification was performed by sucrose-cushioned ultra centrifugation as described (see section 2.2.3). Purified virus particles were plunge frozen and imaged by cryoET. Visual evaluation of tomogram slices however revealed that the glycoprotein layer appeared disrupted (Fig. 3.3A). I varied different parameters of the purification protocol such as centrifugation speed, length of the centrifugation step and number and intensity of pre-clearing steps. However, none of these variations resulted in significantly improved virus particle appearance.

To investigate if the observed disruption of the glycoprotein layer resulted from our purification method, I tested and optimized an alternative sample preparation method which does not require purification. Here, viruses are produced directly on EM-support grids by their host cells, which are cultured and infected with EM grids present in the culture dish (*16*). Briefly, EM-support gold grids were placed in standard tissue culture 6-well plates and MDCK cells were subsequently seeded into the dish. Initially, I optimized the cell concentration and evaluated different grid pretreatments in order to improve cell adhesion on grids. In my hands, best results were achieved when the

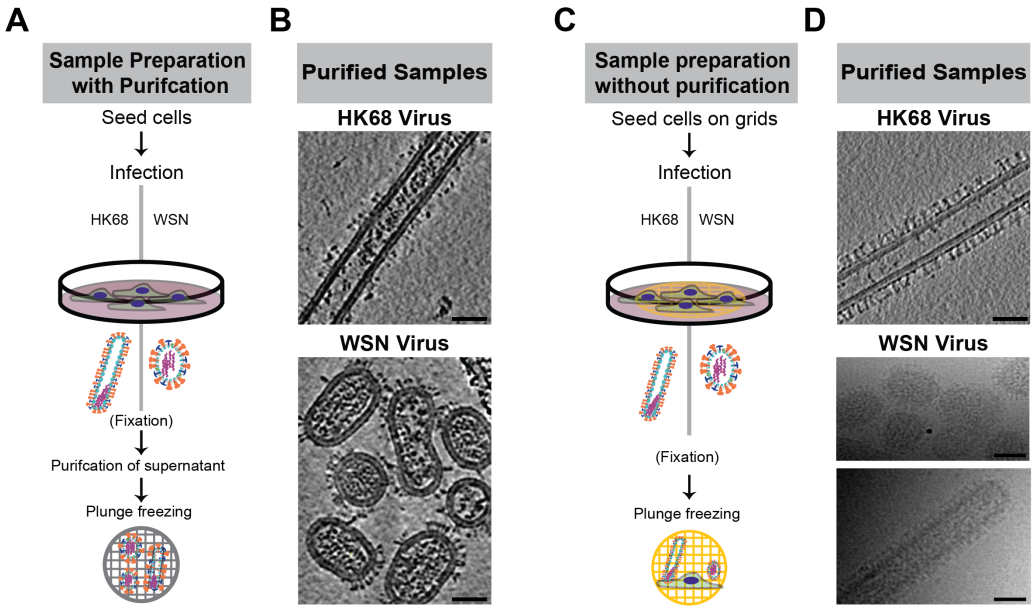

Figure 3.3 Comparison of virus particle quality and morphology for two different sample preparation approaches. A) Schematic of virus sample preparation using purification. B) Slices through representative tomograms of purified HK68 and WSN viruses. C) Schematic illustrating the workflow of sample preparation without purification, where virus particles are directly prepared on grids. D) A representative slice through a HK68 virus tomogram and 2D images of a non-purified WSN virus sample. The lower image illustrates that filamentous particles were also present in non-purified WSN samples. Scale bars are 50 nm.

grids were simply pre-incubated in serum-containing media for at least an hour prior to cell seeding. Once cells adhered well to the bottom of the plate as well as to the grids, cells were infected according to standard virus infection protocols. When clear signs of infection were visible grids containing cells and virus were removed from the plate and plunge-frozen. All virus samples prepared before the end of 2018 were required by biosafety regulations to be inactivated by virus fixation prior to plunge freezing and cryoEM imaging. For sample inactivation the grids covered with cells and viruses were incubated in 4 % paraformaldehyde (PFA) 0.1 % glutaraldehyde (GA) in PHEM buffer with pH adjusted to seven for 30 min (see section 2.4 for more details).
Samples of MDCK cells infected with HK68 virus were prepared accordingly without purification and imaged by cryoEM and cryoET. As demonstrated by tomogram slices of representative tomograms shown in figure 3.3B I observed that the glycoprotein layer appeared much more regular than in samples prepared by purification. In addition, I found that a larger fraction of virions was filamentous and that filaments appeared much longer and more straight in comparison to virus particles found in purified samples. While in purified samples HK68 filaments with a maximum length of 2 µm were observed, the majority of filaments found in non-purified samples was longer than 2 µm and filaments measuring up to 15 µm were observed.

Due to the observation that purification affected the apparent morphology of HK68 viruses, I tested the influence on the appearance of particles from another strain, A/WSN/1933 (H1N1, WSN), which is known for its small, bacilli-shaped morphology. Samples of the WSN strain were prepared with and without purification as described above for the HK68 strain and imaged by cryoET. Tomograms were acquired of purified WSN virus particles and imaged virions were exclusively spherical or bacilli-shaped as previously described in the literature (*28*, *120*). From high magnification images of purified WSN particles, the glycoprotein layer appeared irregular and disrupted, just as observed in tomograms of purified HK68 virus samples (Fig. 3.3A). Non-purified samples were only imaged by 2D cryoEM. Here we found that at least 50 % of particles found in the native WSN sample were filamentous, resembling the filamentous morphology found for native HK68 viruses (Fig. 3.3B). A comparison between the glycoprotein layers also revealed that the glycoprotein layer was much more regular in the sample prepared without purification.

In a next step I tested if subtomogram averaging was feasible from selected tomograms of data acquired for the purified and non-purified samples of the HK68 strain. Subtomograms were extracted around the surface of filaments from purified

and non-purified HK68 virus particles and averaged to obtain an initial reference. Four iterations, including one iteration of z-alignment and three iterations to refine the xy positions on twice binned data were performed for both data sets. While the obtained average from tomograms of non-purified HK68 virus particles resembled the HA ectodomain, the average obtained from tomograms of purified HK68 filaments did not show any interpretable structural features. The negative result from the subtomogram averaging test further supports the observation that the glycoprotein layer is structurally effected by the performed purification. In summary the performed tests demonstrate that the purification method used here affects the glycoprotein layer quality and the distribution of apparent virus morphologies. In particular long filaments seem to be disrupted or selected against during purification. Since the aims of this project required an optimal sample for subtomogram averaging and since a more native setting typically better reflects the in vivo situation, all further data for this project were obtained from samples of non-purified viruses for which initial subtomogram averaging tests had been successful.

3.2.2 High resolution cryoET of non-purified HK68 virus

Following the optimized sample preparation workflow described in the previous section non-purified HK68 virus particles were imaged directly in the vicinity of their host cells by cryoEM upon inactivation by fixation. In overview images, long filaments protruding from cells as well as isolated filaments were observed (Fig. 3.4A,B). For high resolution cryoET, tilt series from -66 to 66 degree with 3 degree increment and following a dose-symmetric tilt scheme ('Hagen scheme') (*95*) and with a pixel size of 1.78 Å were acquired (see section 2.5.1). Prior to tomogram reconstruction, acquired tilt images were processed by motion correction, dose-filtering and CTF-correction according to protocols employed to obtain sub-nm structural information by subtomogram averaging previously (*86*). Tomograms were then reconstructed from the preprocessed tilt-series in IMOD (*103*), after manual refinement of tilt series alignment (see section 2.5.2 for details).

In high magnification tomograms, a dense layer of glycoproteins can be observed on the surface of virus filaments. As described previously for HK68 virus particles as well as other IAV strains (*15*, *45*, *113*, *121*), we identified HA as the most abundant glycoprotein on the virus surface. In addition, individual NA molecules are identifiable along filament sides (Fig.:3.4C). According to previous description of IAV filaments, the eight genome segments are typically only found in the front tip, that assembles

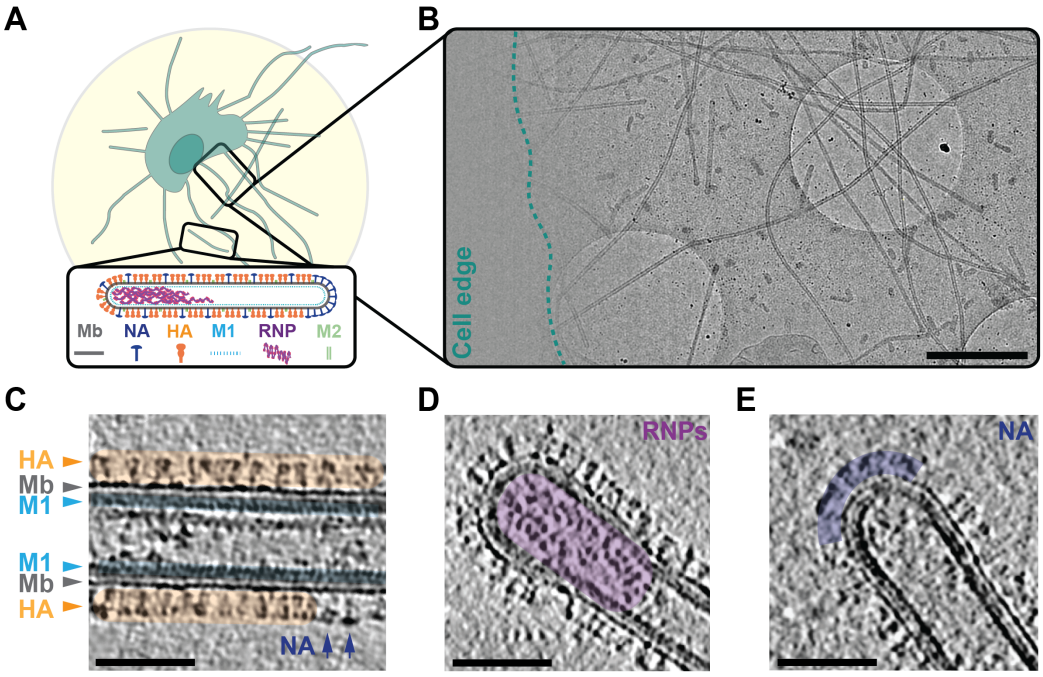

Figure 3.4 Non-purified, inactivated HK68 virus imaged by cryoEM and cryoET. A) Schematic of HK68 virions budding from an infected cell and a schematic of a HK68 virus filament with all components. B) Overview image of a virus-producing cell with virus filaments extending from the cell surface. Scale bar: 1 µm C-E) Slices through tomograms acquired on virus filaments. Scale bars: 50 nm C) Middle part of a filament, mostly HA is present on the virus surface (marked in orange). M1 is visible at the inner side of the membrane (marked in cyan). Two NA proteins are visible on the virus surface (marked by dark blue arrow heads). An actin filament is visible in the inside of the virus. D) A virus tip containing vRNPs (marked in purple). E) A virus tip without vRNPs but with a NA cluster (marked in dark blue).

first (*15*, *16*). In the data described here, four out of the eight imaged filament tips contained vRNPs. Genome-containing tips appeared to be wider than the rest of the filament. The glycoprotein layer composition on the genome tips resembles the one found along filament sides, consisting mostly of HA (Fig. 3.4C). For tips without genomes, typically the rear end of the virus, I observed NA clusters (Fig.3.4C) which matches previous descriptions of IAV filaments (*31*, *41*). Except for the space occupied by the vRNPs, most filaments appeared to be void of additional particles. However, some virus particles contained thin, long protein filaments. The general appearance and the width of these protein filaments matches the description and EM images of actin filaments (Fig.3.4C).

3.2.3 HA in situ structure from non-purified, inactivated HK68 virus

For in situ structure determination, tomographic data were processed by subtomogram averaging focusing on the glycoprotein layer. Subtomograms with a box size of 46x46x46 nm were extracted in a dense meshwork along a cylindrical surface modelled around the manually determined spline of each filament from four times binned data (4 x 4 x 4 voxel binned into one voxel, Bin4) tomograms. Initial positions were at least oversampled twice and the distance between initial subtomogram positions was set to half the spacing of neighbouring HA proteins on the virus surface, which was determined to be on average 10 nm in 3DMOD (*103*)). Alignment was performed reference-free and a single tomogram with higher defocus was chosen to generate a starting reference. 6,000 subtomograms were extracted from this single tomogram, averaged and aligned to each other (Fig. 3.5A). Using binned data for initial alignments reduces computation time for the initial search-extensive iterations. During the initial two iterations, positions were aligned along the Z-axis, the direction orthogonal to the tube surface. Subsequently, the final in-plane angle (phi angle in AV3/TOM notation) was randomized to prevent any alignment bias from the tube's curvature (Fig. 3.5A). Alignment was subsequently focused on adjusting the positions in the xy-plane for two more iterations. Within these two iterations, oversampled positions converged to positions where HA was clearly visible in the tomogram. At this stage, the HA ectodomain and its three fold symmetry became recognizable in the obtained average (Fig. 3.5A). Therefore two more iterations with 3-fold symmetry applied were run. After a total of seven iterations of alignment, duplicates were removed and positions with low cross correlation values were removed. The obtained final average was then used as the initial reference during the alignment of the complete dataset.

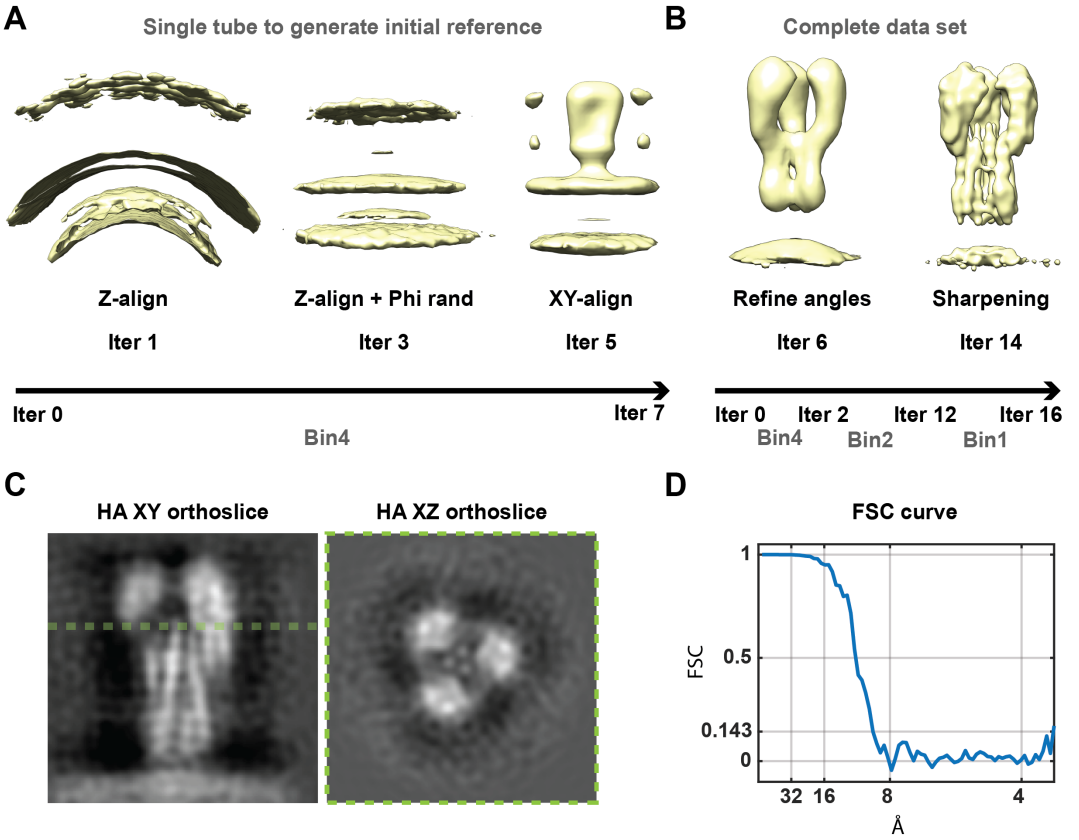

Figure 3.5 Subtomogram averaging of HA from non-purified, inactivated HK68 virus. A) Intermediate results at different stages of HA subtomogram alignment for a single tomogram. 'Phi rand' refers to the randomization of the in-plane phi angle, which generates a rotational average and thereby eliminates the contribution from the tube's curvature. B) Different stages of alignment for averages obtained from the complete dataset. Refine angles refers to the refinement of the protein's orientation by focusing the alignment on the refinement of the three euler angles. The average shown on the right corresponds to the final structure, which was sharpened using an empirically determined B-factor of -1200 C) Two orthoslices through the final, sharpened average shown in B D) FSC curve calculated based on the two half sets for the sharpened structure shown in C, indicating a final resolution of 9 Å determined at the 0.143 level of the FSC curve.

The complete dataset contained 167,002 subtomograms extracted from a total of 34 filaments from 33 CTF-corrected bin4 tomograms with an effective pixel size of 7.12 Å. The phi angle of each position was randomized from the beginning. The first two Euler angles are pre-determined based on the orientation of each subtomogram normal to the tube surface. The previously determined average from a single tomogram was used as the first reference to increase the efficiency of the initial alignment steps for the large number of subtomograms used and to reduce computational costs. One iteration using the average from the single tube as a reference and a low pass filter of 35 Å was sufficient to align positions in all three directions. The alignment was evaluated visually and showed that oversampled positions had converged to positions where protein density was clearly visible in tomogram slices. Duplicates which had converged were removed based on distance between subtomogram positions, which reduced the number of subtomograms to 36,000. After another iteration of alignment with 3-fold symmetry applied, subtomogram positions were cleaned once more based on cross correlation which reduced the total number to 32.000 subtomograms, which were re-extracted from 2x2x2 binned tomograms (Bin2) with an effective pixel size of 3.56 Å. At this stage, subtomograms were split into two half sets based on odd and even subtomogram numbers, resulting in 16,000 subtomograms each. These two half sets were subsequently processed independently with identical alignment parameters. Two iterations of alignment were performed to refine positions and orientations using a low pass filter set to 35 Å. Subsequently the low pass filter was gradually lowered to 21 Å and angular search accuracy was further refined at the same time over additional six iterations (Fig. 3.5B). The final resolution at Bin2 was calculated to be 10 Å. Additional cleaning resulted in final subtomogram numbers of 15,000 subtomograms per half set. Based on the three-fold symmetry this yields a total of 45,000 asymmetric units per half set. For the final iterations of alignments, subtomograms at updated positions were re-extracted from un-binned tomograms (Bin1) using a final box size with side lengths of 28 nm corresponding to 160 un-binned pixels. Two iterations were run using the same low pass filter of 21 Å used at Bin2. During these two iterations the resolution increased to 9 Å as determined by calculating the FSC between the two half sets (Fig. 3.5B). During subsequent alignment the resolution did not improve further.

The resolution of the final reconstruction was determined by calculating the FSC of the masked volumes from each half set. The final EM density map was then sharpened with an empirically determined B-factor of -1200 and low pass filtered based on the resolution determined at the 0.143 cut-off of the FSC curve (Fig. 3.5C) (see section 2.6.1).

3.2.4 Comparison of the HA reconstruction from subtomogram averaging to existing HA crystal structures

Consistent with the obtained resolution of the subtomogram averaging structure, α-helices were visible in the final reconstruction of HA. To understand if there are any structural differences between the HA structure obtained from subtomogram averaging and previously determined structures of HA I compared the in situ HA structure to existing crystal structures of HA from the HK68 strain, crystallized at pH=7 (PDB: 4WE4, (*122*)). The model of a single HA protomer was fit as a rigid body into the EM density map using the 'fit in map' routine in UCSF Chimera (*105*). In most regions a good overlap between the crystal structure and the HA subtomogram averaging reconstruction was observed (Fig. 3.6A). The reported cross correlation between our structure and a 9 Å map rendered from the crystal structure model was determined to be 0.92 in Chimera. However, the region around the B-loop, which connects helix A and helix C of HA2, differed between the two structures. I found that in the HA map from subtomogram averaging the density corresponding to helix A of HA2 appears to further extended than the corresponding region in the crystal structure, where helix A transitions into a loop earlier (Fig. 3.6B).

In a next step I fitted the three protomers of the HA trimer as three individual rigid bodies into the HA map obtained from subtomogram averaging (Fig. 3.6C and D). I then compared the arrangement of the three individually fitted protomers to the relative monomer arrangement in the trimeric crystal structure. I found that the three protomers when fit into the HK68 HA in situ structure were further apart from each other relative to the corresponding protomers in the crystal trimer (Fig. 3.6E, F). However, the observed difference appeared to be most prominent in the head region of HA. This observation suggests that the three subunits in the in situ HA structure obtained from subtomogram averaging are tilted outwards, away from the central three-fold axis. Our comparison further suggests, that the tilting centre is located in the stem region, so that the dislocation of the head region during tilting is stronger for the head region than for the stem region (Fig. 3.6E).

3.2.5 In situ structure of HA from non-purified, inactivated Udorn virus

Based on the observed differences between the protomer arrangement in the crystal structure and in the in situ structure of HA from HK68 virus, I tested if similar observations could be reproduced for a different virus strain in an independently prepared sample. I therefore repeated the described experiment and prepared samples containing

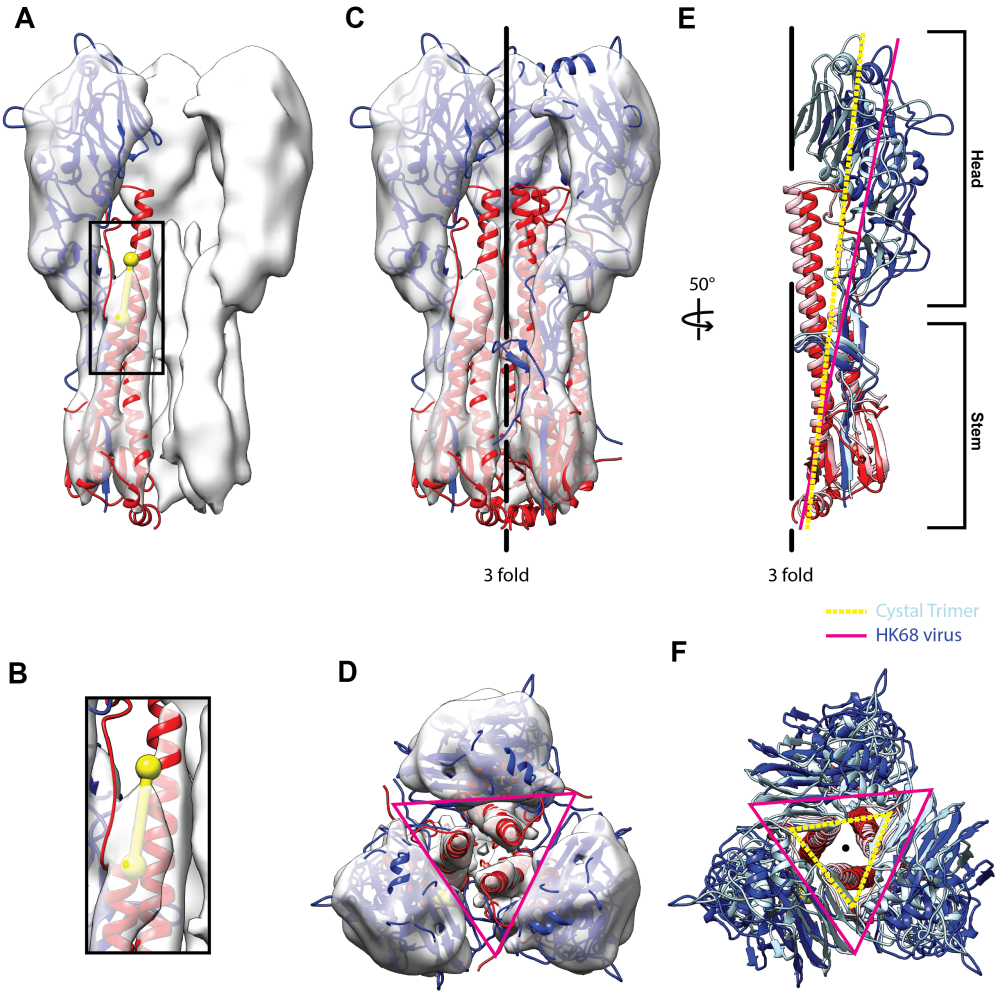

Figure 3.6 Comparison of the HA in situ structure from non-purified, inactivated HK68 virus to existing HA crystal structures. A) Single protomer of a HK68 HA crystal structure fit as a rigid body into the HK68 HA subtomogram averaging structure. HA1 is depicted in blue and HA2 is depicted in red. B) Enlarged view of the boxed area in A showing the extended density of helix A in the EM density map (yellow marker). C) Three protomers of HA fitted as 3 rigid bodies into the HA subtomogram averaging structure D) Top view of overlay shown in C. The pink triangle follows the inner edges of the fitted protomers. E and F) Overlay of the HA crystal structure protomers fitted into the HA reconstruction from subtomogram averaging (as shown in C and D) and the crystal trimer. In the crystal trimer HA1 is shown in light blue and HA2 in rose. E) Side view of two protomers only relative to the three fold axis. Axes through the protomer are shown in pink for a protomer fit in the subtomogram averaging structure and shown as a dashed yellow line for a protomer from the crystal trimer. F) Top view of all three protomers for both structures. The pink triangle is the same as in D. The yellow dashed triangle follows the inner edges of the crystal trimer and shows that the crystal trimer inner edges are closer together. PDB of crystal structure: 4we4.

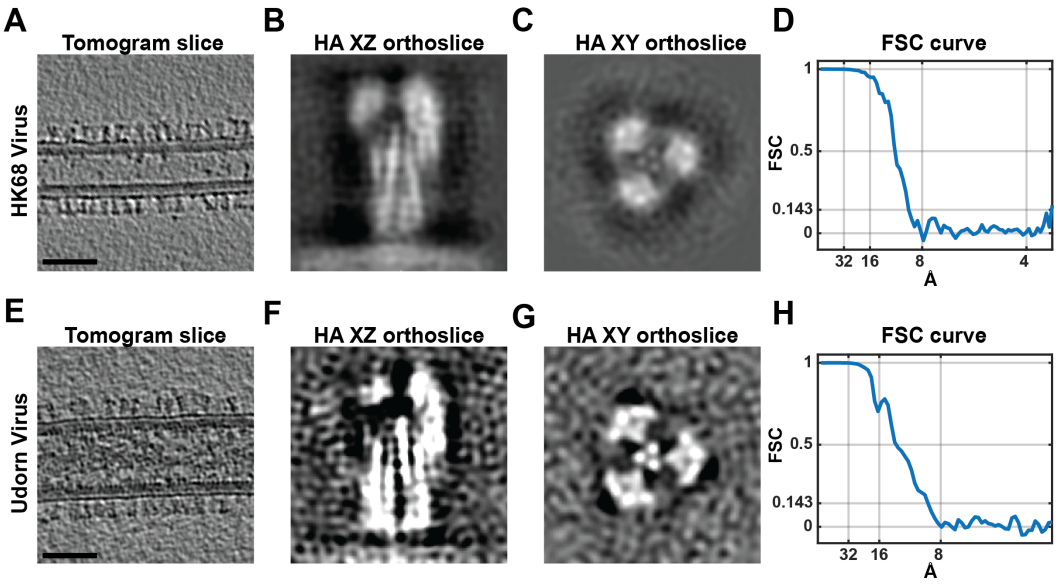

Figure 3.7 Comparison of tomograms and HA reconstructions from non-purified, inactivated Udorn and HK68 virus. A) Representative tomogram slices of HK68 virus. B) XZ orthoslice through the final reconstruction of HA from HK68 virus. C) XY orthoslice through the final HA reconstruction of HA from HK68 virus. D) FSC curve of the reconstruction shown in B and C. E) Same as A but for Udorn virus. Data were collected at lower defocus which explains the difference in contrast relative to A. F-H) Same as B-D but for Udorn virus. Scale bars:50 nm.

viruses from the A/Udorn/307/1972 (Udorn, H3N2) strain so the same subtype as HK68 (H3N2). Udorn is a commonly used lab strain with a predominantly filamentous morphology which effectively propagates in mammalian cell culture systems. Samples of Udorn virus were prepared without purification and including a step of chemical fixation for inactivation just as described for the HK68 virus (for details see section 2.4). In low magnification cryoEM images a high concentration of filaments in the surroundings of infected cells were found just as observed for the HK68 virus sample. Tilt series were acquired and preprocessed as described in section 2.5.1 and section 2.5.2. Slices through reconstructed tomograms of Udorn virus confirmed that the appearance of Udorn virus is very similar to HK68 virus (Fig. 3.7A,E). Data were further processed by subtomogram averaging just as described for the HK68 virus sample (see section 3.2.3). An reconstruction of HA from Udorn virus tomograms was obtained from a final number of 6,200 subtomograms per half set resulting in 18,600 asymmetric units. The resolution of the reference from the last alignment step was determined to be 9 Å and the FSC curve appeared similar to the FSC curve obtained for the HA reconstruction from HK68 virus data (Fig. 3.7D,H). The final reconstruction was sharpened using a B-factor of -1200 and filtered according to the determined resolution (Fig. 3.7F,G) (see section 2.6.1) and closely resembled the final

reconstruction of HA obtained from HK68 virus data (Fig. 3.7B,C).

The same crystal structure that was fitted into the HK68 HA structure was fitted into the obtained in situ HA structure from Udorn virus. Just as described for HK68 virus, a single HA protomer was fitted as a rigid body into the obtained EM density map (Fig. 3.8A). The reported cross correlation value from a comparison to a map generated at 9 Å from the crystal structure protomer was 0.9 which confirms a good overlap between the in situ structure and the model (Fig. 3.8A, B). I found extended density for the HA map from Udorn in the region of HA2 helix A just as observed for the HK68 in situ structure (Fig. 3.8B). I fitted three protomers and compared the arrangement from this fit to the arrangement when fitted into the HK68 HA structure and to the protomer arrangement found in the HA crystal trimer (Fig. 3.8). The arrangement of the three protomers fitted into the HA structure from HK68 in comparison to the three protomers fitted into the HA structure from Udorn showed that the the three protomers are arranged in a very similar manner. This comparison confirms the high similarity of the two in situ structures (Fig. 3.8D). In contrast, when the arrangement of the three fitted protomers was compared to the arrangement of the protomers in the crystal trimer, I found that the protomers fitted into the Udorn in situ structure were tilted outwards relative to the crystal trimer (Fig. 3.8C and E). In summary the observed differences in the HA protomer arrangement of the in situ HA structure from non-purified, inactivated HK68 virus relative to existing HA crystal structure trimers were reproducible in two independently prepared samples from different virus strains.

3.2.6 In-situ structure of HA from fixed and unfixed HK68 VLPs

Here I tested if the observed differences in the protomer arrangement between the two solved in situ structures on the one hand and the crystal structure trimer on the other hand are related to specific aspects of sample preparation. In standard influenza virus production protocols there are two critical steps which potentially could effect protein structure: Firstly, trypsin-induced cleavage of HA0 and secondly virus inactivation by chemical fixation. As part of a standard protocol for IAV experiments, trypsin is added during the infection to assure that cleavage of HA0 into HA1 and HA2 occurs on the surface of newly produced virions. The loop connecting HA1 and HA2 is proteolytically cleaved which allows HA to undergo a conformational change upon exposure to low pH. This process ensures that fusion proceeds between the virus particle and the endosomal membrane, a critical part of the virus life cycle (see section 3.1.1). Going through several rounds of the full virus life cycle is desired during virus production to increase the fraction of infected cells and the number of produced virus particles.

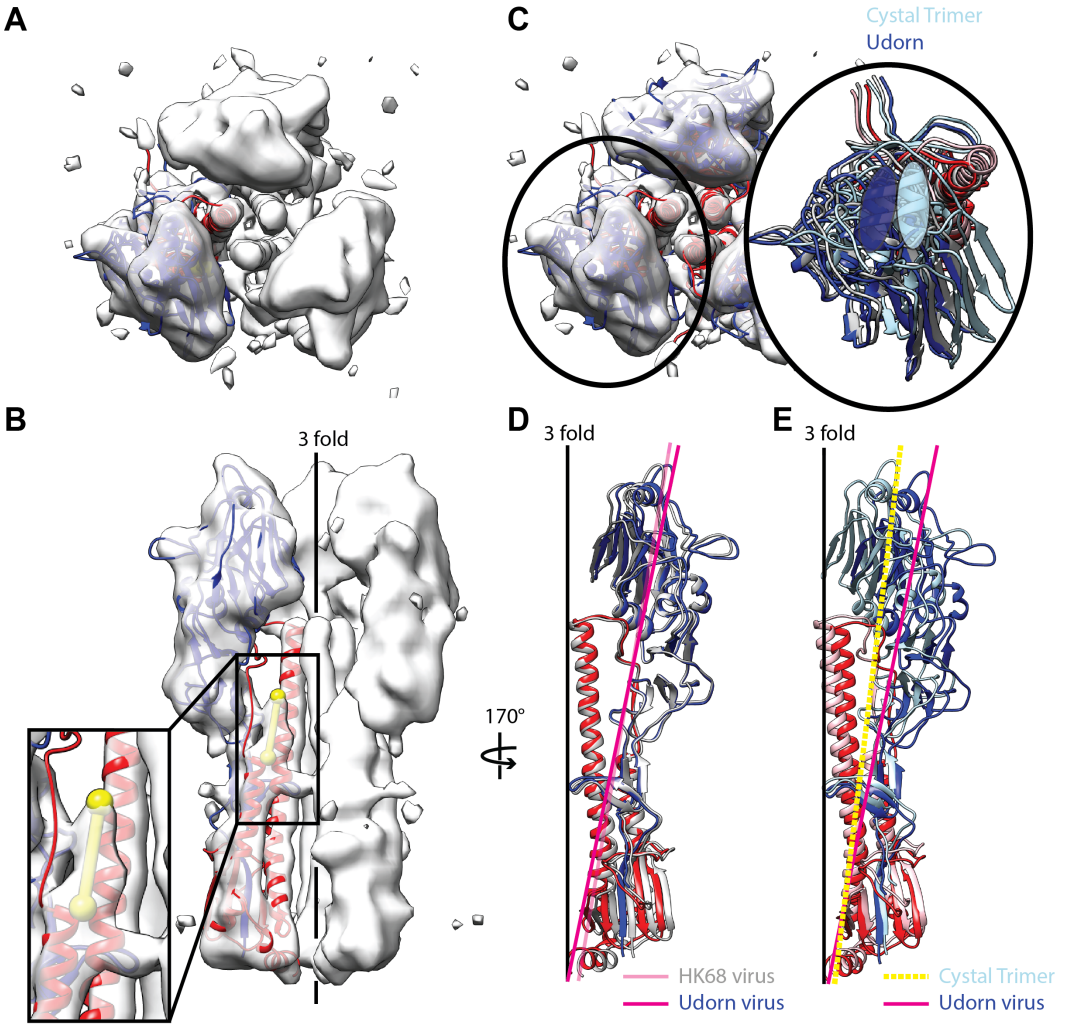

Figure 3.8 Comparison of the HA structures obtained from Udorn and from HK68 virus. A) Single protomer of a HA crystal structure fit as a rigid body into the Udorn subtomogram averaging structure. Top view. B) Side view of overlay shown in A. The box focuses on the extended density found for HA2 helix A relative to the crystal structure (yellow marker). C) Top view of three HA crystal structure protomers fit as three rigid bodies into the HA structure from Udorn. The inset shows an overlay of a single protomer fit into the Udorn structure and a single protomer from the crystal trimer. The helices on top of the head region are marked by respectively coloured ellipses to serve as a marker for the relative position of the protomer. The protomer fitted into the Udorn structure (dark blue) is moved outwards relative to the protomer in the crystal trimer (light blue). D) Overlay of two single protomers after fitting into the Udorn and the HK68 (grey model) HA in situ structures. Coloured axis indicate that the protomer arrangement in the HA structure from HK68 (rose axis) and the HA structure from Udorn (pink axis) are very similar. E) A protomer of the crystal trimer with HA1 in light blue and HA2 in rose is compared to the protomer when fit into the Udorn structure. Axes through protomers are drawn as dashed, yellow line for the crystal structure and pink for Udorn. PDB:4WE4.

Previous studies, which employed crystallography to compare cleaved and uncleaved HA structure, suggest that the cleaved and the uncleaved structure are identical in all regions except for the cleavage site itself (*123*); however previously undescribed effects could still occur in situ . Since H3N2 influenza is categorized as a biosafety level 2 (BSL2) pathogen in the UK (and S2 in Germany) specific sample inactivation was required prior to imaging prepared virus samples by cryoEM at the electron microscopy facilities at EMBL (Germany) as well at the LMB (UK). The regulation at EMBL has changed since mid 2018 which allowed me to prepare an additional virus sample without inactivation as will be described in section 3.2.7. According to safety regulations and based on infectivity tests, inactivation of H3N2 influenza virus is achieved by incubating virus samples in 4 % paraformaldehy (PFA) and 0.1 % glutaraldehyde (GA) for at least 30 min.

To assess the impact of fixation and cleavage on the obtained in situ protein structure I made use of a IAV VLP system. VLPs are non infectious, do not go through a virus life cycle and therefore require neither inactivation nor cleavage. Here I used VLPs that correspond to the HK68 strain to maintain comparability to the data obtained from HK68 virus. HK68 VLPs have been previously introduced and characterised by Chlanda et al. (*45*). Briefly, VLPs are generated by transfecting HEK293T cells with combinations of four individual plasmids, which encode for HA, NA, M1 and M2 respectively. Expression of these four plasmids is sufficient for the production of particles that closely resemble the architecture of HK68 virions (*45*). Due to the absence of the genome in this type of VLPs, particles are not infectious. Secondly, cleavage of HA (by trypsin) is not required since transfected cells are not undergoing an infection and consequently HA0 cleavage as a prerequisite for fusion is not necessary. Instead, the application of trypsin has negative effects on sample quality due to the unspecific trypsin cleavage activity which ultimately can cause cell detachment. I therefore prepared fixed and unfixed VLP samples in parallel, while both samples remained un-cleaved and thus without trypsin-treatment.

CryoET of unfixed and fixed VLPs

As for virus samples, fixed as well as unfixed VLPs were prepared without prior purification by transfecting cells that were cultured directly on EM support grids analogous to the workflow established for preparing virus samples. Low cell concentrations were required for the on-grid sample preparation to ensure cell-free areas for imaging. Therefore, VLP sample preparation conditions were optimized to allow for efficient transfection at relatively low cell concentrations (see section 2.4.4). For the control

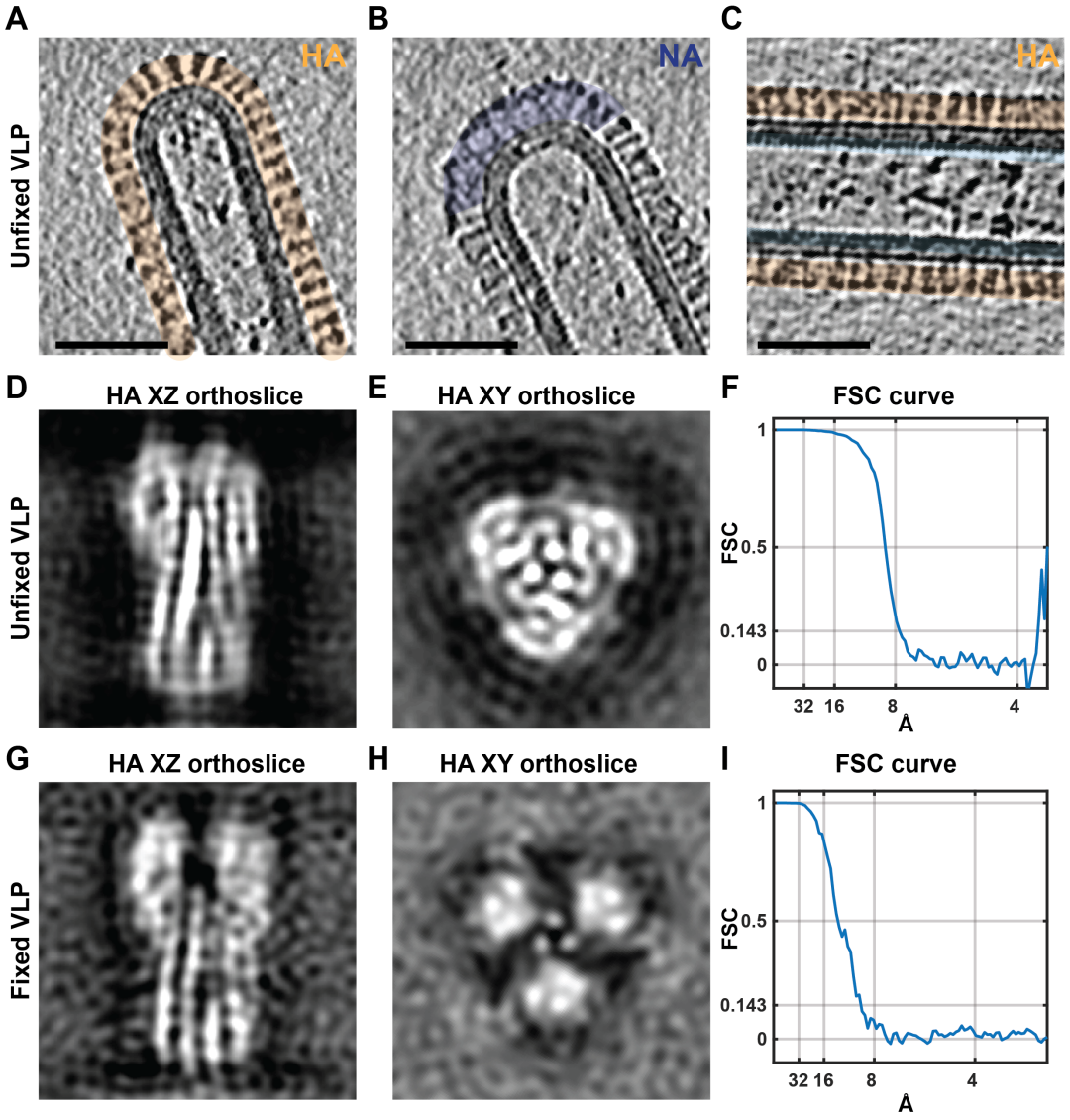

Figure 3.9 Tomogram slices and subtomogram averaging structures of HA obtained from fixed and unfixed HK68 VLPs. A-C) Representative slices through different regions of tomograms from unfixed VLPs. A) Filament tip covered in HA. B) Filament tip with a larger NA cluster. C) Middle part of a filament mostly covered in HA. M1 is visible underneath the membrane. D) XZ orthoslice through the final, sharpened HA reconstruction from non-purified, unfixed VLPs E) XY orthoslice the same HA reconstruction shown in D. F) FSC curve of the HA reconstruction presented in D and E. G-I) Same as D-F for non-purified, fixed VLPs. Scale bars: 50 nm.

sample of chemically fixed VLPs, the samples were fixed the same way samples of HK68 and Udorn virus had been treated. Samples of fixed VLPs were incubated for 30 min in 4% PFA, 0.1 % GA in PHEM buffer, directly prior to plunge freezing. After plunge-freezing, samples were imaged by cryoEM and high magnification cryoET. Low magnification images of VLPs produced directly on grids were indistinguishable from low magnification images of native HK68 or Udorn virus. No differences were found between the fixed VLP and the unfixed VLP sample from overview images. In both samples, I observed up to 20 µm long filaments surrounding the transfected cells. High resolution tomograms of fixed and unfixed VLPs were collected as described for the HK68 virus sample (see section 2.5.1). In acquired tomograms I found straight, filamentous VLPs covered in a dense glycoprotein layer. In central slices through filaments, the matrix layer was clearly visible underneath the membrane (Fig. 3.9C). Some imaged tips were covered in NA clusters resembling the rear ends of viruses (Fig. 3.9A,B).

Subtomogram averaging of HA from unfixed VLPs

Data from unfixed HK68 VLPs were processed by subtomogram averaging for HA in the same way as described for the HK68 virus sample (section 3.2.3). Alignment of subtomograms was performed reference-free and subtomograms extracted from a single tomogram were used to generate a starting reference. The starting reference was subsequently used for the alignment of the complete dataset. The initial dataset size for HA from the unfixed VLP dataset corresponded to 650,000 subtomograms extracted from 60 tomograms and 69 VLP filaments. Upon XZY alignment in Bin4 tomograms, convergence of oversampled positions was evaluated by eye and the positions were cleaned by euclidean distance to assure that only a single subtomogram remained per protein position resulting in 110,000 subtomograms which were subsequently distributed onto two half sets. Upon additional final cleaning steps, the final structure was generated from 46,000 subtomograms per half set, and a number of 138,000 asymmetric units due to the three fold symmetry applied. By calculating the FSC of the masked, two half sets, the resolution at the 0.143 FSC level was determined to be 7.5 Å (Fig. 3.9 B). The final structure was sharpened with an empirically determined B-factor of -1200 and low pass filtered to the determined resolution (Fig. 3.9 B).

Subtomogram averaging of HA from fixed VLPs

Tomograms of fixed VLPs, served as a subtomogram averaging training dataset for Zunlong Ke (Briggs group). Tomograms were processed by Zunlong and assistance of me for HA according to protocols for HA subtomogram averaging, previously established and used by me and as described for subtomogram averaging of HK68 virus, Udorn virus and the unfixed VLP dataset. The initial dataset consisted of 117,000 subtomograms extracted from 12 tomograms containing 20 VLP filaments. Upon distance cleaning and cross correlation based cleaning steps the final number of subtomograms for the final reconstruction was reduced to 12,000 subtomograms per half set corresponding to 36,000 asymmetric units due to the three fold symmetry of HA. The resolution at the 0.143 criteria from the calculated FSC of the two masked, half maps was determined to be 9 Å and the structure was sharpened using a B-factor of -1200 and low pass filtered according to the resolution (Fig. 3.9 C).

Comparison of obtained HA structures from fixed and unfixed VLPs

In a direct comparison between the density maps of HA from fixed and unfixed VLPs, differences in the amount of the opening of the trimer can be noticed in top view visualizations shown in figure 3.10A and C. The trimeric HA structure of fixed VLPs appears to be more open relative to the trimeric HA structure from unfixed VLPs. Consistent with the difference in resolution, we find that more features are resolved in the structure obtained from unfixed VLPs relative to the fixed VLP structure. The short helix comprising residues 187-196 in the receptor binding domain is well resolved in the unfixed VLP HA structure and not well resolved in the HA structure obtained from fixed VLPs (Fig. 3.10A and C, red ellipsoids). In HA from fixed VLPs, extended density relative to the crystal structure in the helix A HA2 region can be observed just as described for the in situ HA structure from fixed HK68 and fixed Udorn virus (Fig. 3.10C). In contrast, in the unfixed VLP structure the density for HA2 helix A extends into the HA1 head domain and the density of the helix does not appear to be extended.

In a next step I compared both structures to existing HA crystal structures, following the same procedures explained for the comparison of HK68 virus and Udorn virus HA to HA crystal structures (see section 3.2.4). Three crystal structure protomers were fit as three individual rigid bodies into the EM density maps obtained for HA from unfixed and fixed VLPs (Fig. 3.10A and C). I compared the arrangement of the fitted protomers to the protomer arrangement as it is found in the crystal structure trimer. I

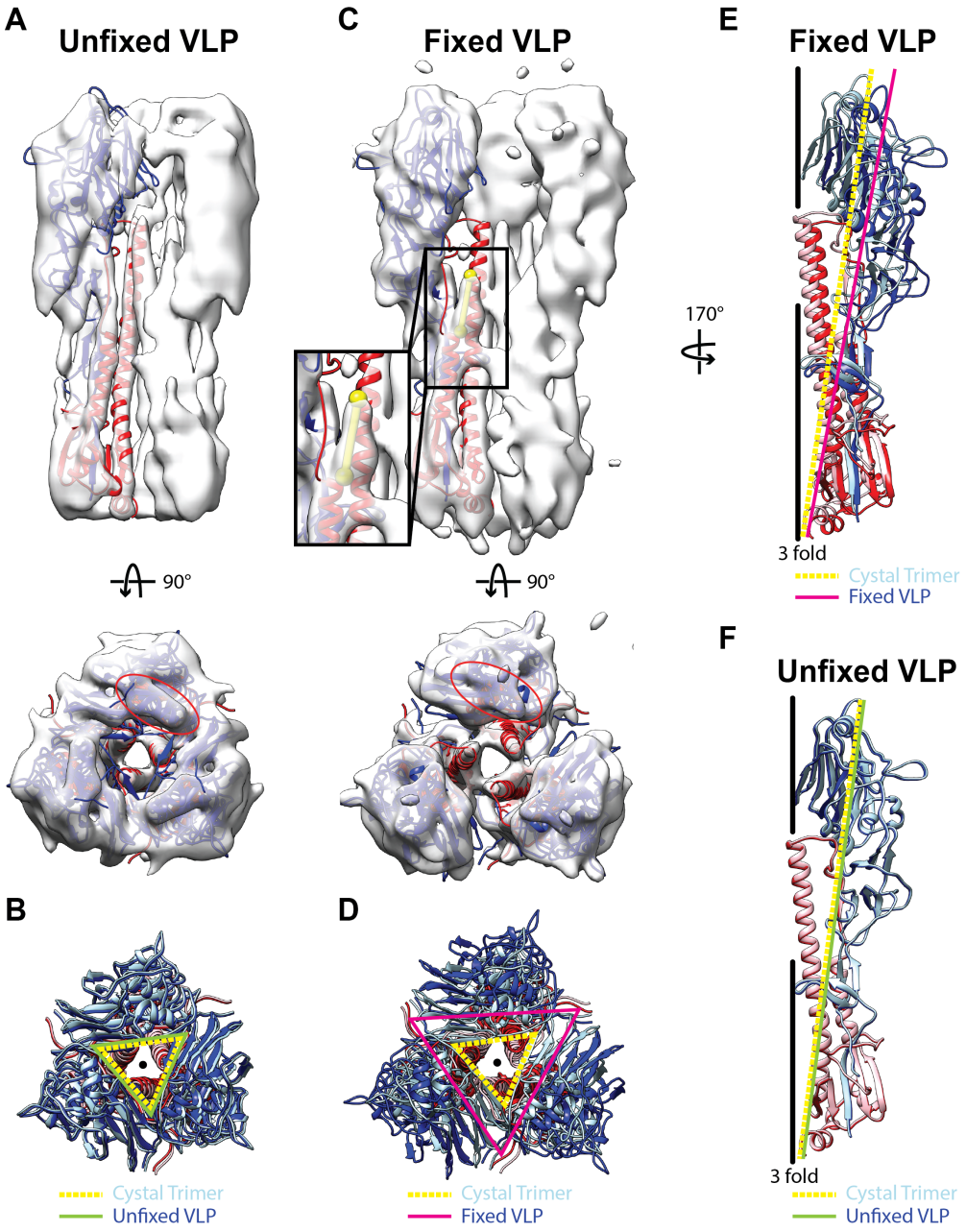

Figure 3.10 HA structures from fixed and unfixed VLPs in comparison to HA crystal structures. A) Side view of a single protomer of the HA crystal structure and top view of three protomers upon fitting into the HA in situ structure for unfixed VLPs. The red ellipse in the top view points to the small helix of the RBS. B) Overlay of three HA protomers shown in A and the crystal trimer. HA1 of the crystal structure trimer is shown in light blue and HA2 is shown in rose. As indicated at the bottom, the inner edges of the crystal trimer are followed by a triangle drawn as dashed, yellow lines while the inner edge of the protomers corresponding to the structure of the unfixed VLP are marked by a green triangle. C) Same as A but for the in situ HA structure from unfixed VLPs. The inset illustrates the extension of the density corresponding to HA2 helix A. D) Same as B but here three protomers upon fitting into the HA structure from fixed VLPs are overlaid to the crystal trimer. E) Comparison of the orientation of a single protomer upon fitting into HA from fixed VLPs relative to the central three fold axis and a single protomer from the crystal trimer. F) Similar to E but a protomer upon fitting into HA from unfixed VLPs is shown. PDB: 4WE4.

found that the protomer arrangement in the HA structure from unfixed VLPs corresponds exactly to the protomer arrangement found in the crystal trimer (Fig. 3.10B and F). In contrast, the protomer arrangement obtained upon performing the fit into the EM density map of fixed VLPs differs from the protomer arrangement in the crystal trimer (Fig. 3.10D and E); the protomers fitted into the HA structure of the fixed VLPs are further apart from each other (Fig. 3.10D). A comparison in side view shown in figure 3.10E illustrates that the protomers are tilted away from the three fold axis in comparison to the protomers of the crystal trimer. This effect is most noticeable in the head region. This effect on the protomer arrangement for the HA structure obtained from fixed VLPs is similar to what was observed and described in previous sections for the HA in situ structures from fixed HK68 and fixed Udorn virus.

3.2.7 In situ HA structure from non-purified, unfixed HK68 virus

Since mid 2018, biosafety regulations at EMBL have changed and now allow the imaging of S2 samples for a selected microscope of the cryoEM facility. In addition, a S2 cryo preparation room at the centre for infectious diseases at Heidelberg university was built which allows plunge freezing at S2 conditions and to which I was kindly granted access to by Hans-Georg Kraeusslich.

To prepare an additional control for the observed impact of the chemical fixation on the HA structure I prepared samples of unfixed HK68 viruses in the group of Hans-Georg Krauesslich (Heidelberg university). Samples were prepared without purification and without fixation. From an inspection of grids from 2D cryoEM images no difference to previously acquired images of fixed HK68 virus was noticed and I found long filaments in the surroundings of seeded MDCK cells distributed across the EM support grid. High resolution tomograms were subsequently acquired at EMBL under S2 conditions. Representative slices through acquired tomograms resemble the tomograms of fixed HK68 virus samples and I found genome containing tips (Fig. 3.11A), NA-covered tips (Fig. 3.11B), a dense glycoprotein layer of mostly HA interspaced by small NA clusters (Fig. 3.11C).

Data were subsequently processed by subtomogram averaging for HA together with Zunlong Ke (Briggs group) as part of his subtomogram averaging training and according to the previously described processing protocols. A initial reference was generated from a single, two times binned (Bin4) tomogram starting with 12,000 subtomograms extracted from two virus filaments, which was subsequently reduced to 2000 subtomograms to generate the average for the alignment of the full dataset. A subset of 24 tomograms was used for subtomogram averaging. The dataset contained 30

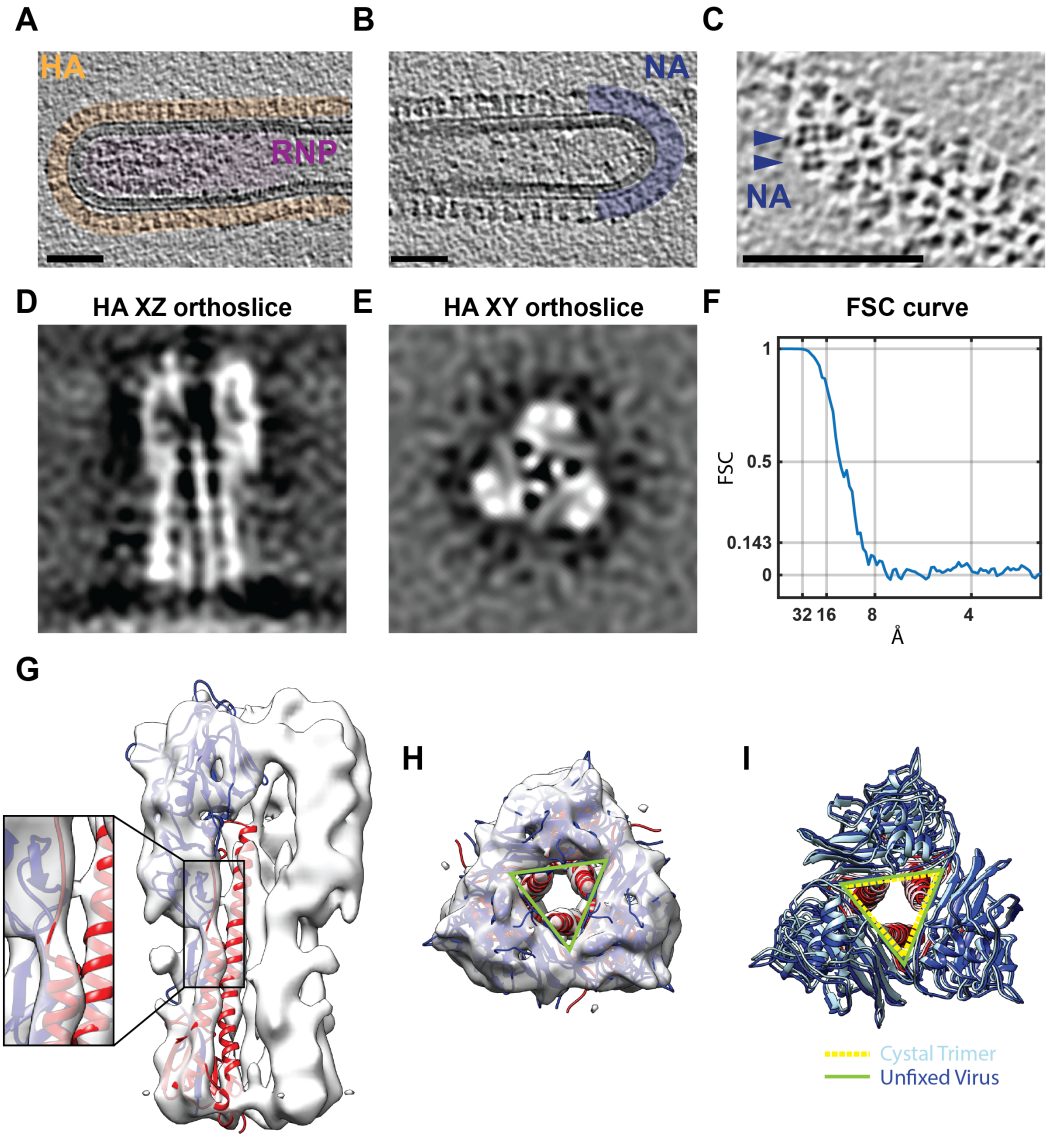

Figure 3.11 In situ HA structure from non-purified, unfixed HK68 viruses. A-C) Exemplary tomogram slices. A) Tip with RNPs B) Tip with NA cluster at the rear virus end. C) Filament surface with two NA molecules in between a dense HA layer. Scale bars: 50 nm. D) XZ orthoslice through the final, sharpened reconstruction of HA. E) XY orthoslice through the final, sharpened reconstruction of HA. F) FSC curve corresponding to the structure shown in B G) A single HA crystal structure protomer fit as a rigid body into the density map of HA from unfixed HK68 viruses. Side view. Inset: Focus on the density at the end of HA2 where no extended density is visible in this structure H) Three protomers fit as three rigid bodies into the density map. Top view. I) Overlay of three protomers fitted into the HA reconstruction and the HA crystal trimer. Inner edges of the fitted protomers are marked as green lines and the inner edges of the HA crystal trimer are marked as dashed, yellow lines. PDB: 4WE4.

virus filaments which resulted in 200,000 highly oversampled subtomogram positions. Upon convergence to the actual HA position, the number was reduced to 45,000 and based on additional cleaning by cross correlation the number was reduced to 38,000 subtomograms. Upon splitting into half sets, this resulted in 19,000 subtomograms per half set and 48,000 asymmetric units when taking the three fold symmetry into account. The resolution of the final average was determined to be 9 Å according to the 0.143 threshold of the FSC calculated between the two, masked half sets (Fig. 3.11F). The final reconstruction was sharpened with a B-factor of -1200 and filtered according to the determined resolution (Fig. 3.11B). Shown from the top, the structure resembles the closed conformation found for unfixed VLPs (Fig. 3.11E). This initial observation was confirmed by fitting three protomers in the obtained density map (Fig. 3.11H) and comparing the arrangement of the three fitted protomers to the trimeric crystal structure. The comparison shown in figure 3.11I demonstrates that the arrangement is identical. This illustrates that the in situ HA structure from unfixed viruses is captured in a close confirmation just like the HA structure obtained from unfixed VLPs. When focusing on the HA2 helix A region, no density extension of the helix density can be found (Fig. 3.11G).

3.2.8 Structural differences between HA structures from fixed samples and from unfixed samples

In a direct comparison of all obtained HA structures, it becomes clear that all structures can be divided into two groups according to structural similarities (Fig. 3.12). HA is observed in a open conformation for all structure obtained from fixed samples and its closed conformation for all structure obtained from unfixed samples. Closed HA structures from unfixed samples correspond exactly to the dimensions of the crystal trimer.

PFA and GA, which were used here for virus inactivation, are very effective protein crosslinkers and are widely applied in biological research because of this property. Typically, treatment by PFA and GA traps naturally occurring conformations and is therefore also applied to preserve cellular structures for subsequent immunostaining, resin-embedded EM or to capture transient, flexible conformations in protein complexes for subsequent structural studies (*124*, *125*). The observation of the effect of fixation on the in situ structures presented here in combination with the properties of PFA and GA indicates that virus inactivation through PFA/GA fixation affects the observed average HA protein structure and conserves a more open, potentially transient, conformation than in non-fixed samples.

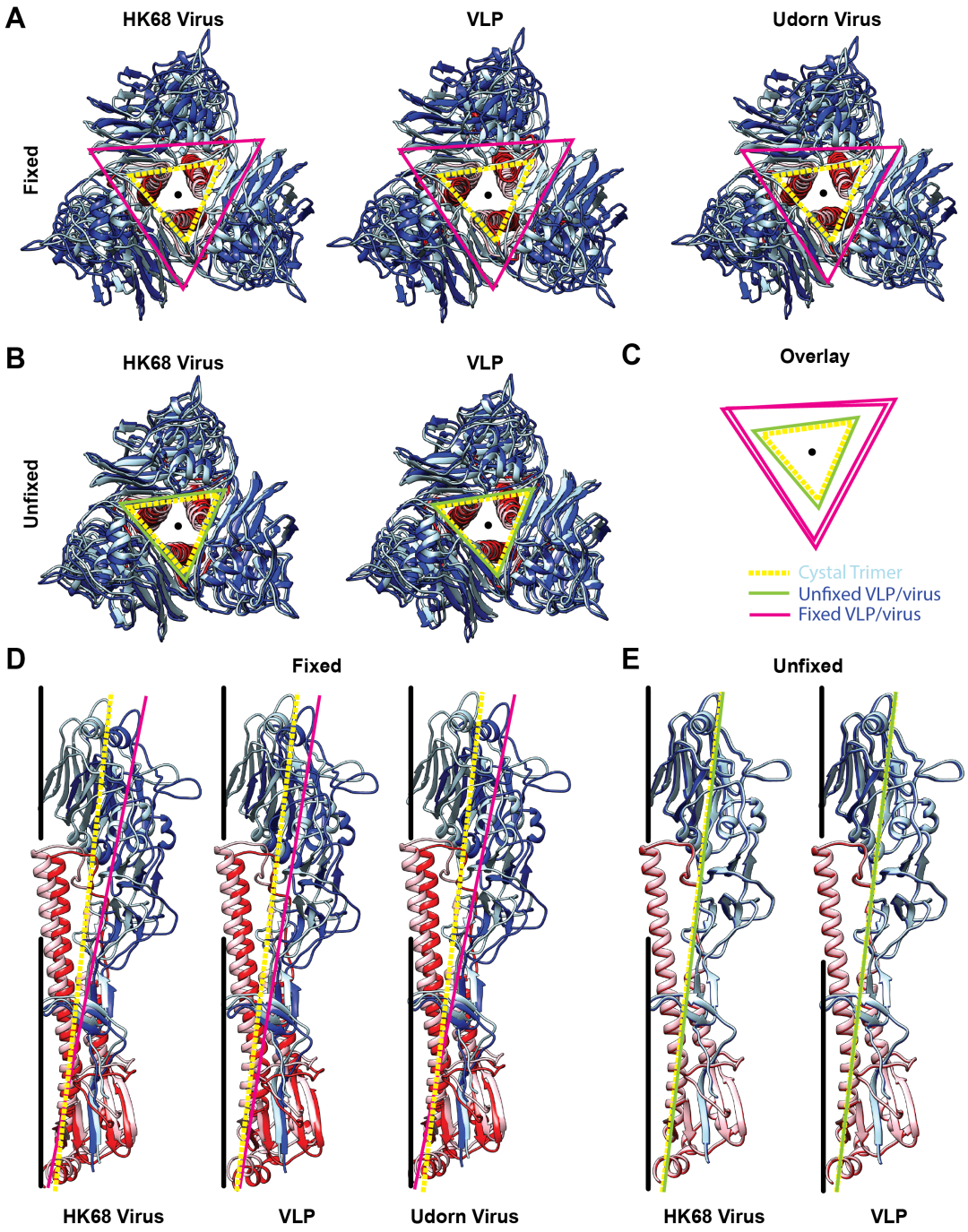

Figure 3.12 Direct comparison of all HA structures obtained from fixed and unfixed samples. In HA models used for fitting HA1 is shown in dark blue and HA2 is shown in red. For the HA crystal trimers HA1 is shown in light blue and and HA2 is shown in rose. A) Top views of three protomers upon fitting into the indicated HA structures from fixed samples overlaid to the crystal trimer. The protomer edges closest to the three fold axis (black dot) are marked in different colours as indicated in C. B) Same as A but for unfixed sample. C) Marked positions along the inner edges for all fitted protomers are overlaid to each other. D) Side views of single protomers relative to the central three fold axis upon fitting into the respective in situ HA structures from fixed samples and in comparison to protomers of the trimeric crystal trimer. The colour scheme is the same as in A. E) Same as D but for unfixed samples. PDB: 4WE4.

A more careful comparison of the open protomer arrangement as found in HA structures from fixed samples and the closed protomer arrangement as found in HA structures from unfixed samples or in the crystal trimer shows that in order to transition from the closed to the open arrangement, protomers undergo a movement outwards, sightly downwards and a small rotation of about 10° clockwise, with the rotation axis being roughly located at the outer loop of the head domain, around residue 220 (fig. 3.12 and 3.13).

3.2.9 Comparison of a HA protomer and a HA domain fit

In situ HA structures obtained from fixed samples were shown to be captured in a more open conformation. To come to this conclusion single protomers were fit as rigid bodies into the obtained EM density maps and subsequently compared to the HA protomer arrangement as it occurs in the crystal structure trimer. I consistently found that the head region of the fitted protomers seemed to be displaced further away from the three fold axis relative to the crystal trimer than the stem region. This observation can result from two types of protomer movement: Either tilting of the complete protomer around a tilt axis located in the stem region or by separate movement of the HA1 head region. To test if there is evidence for either of these models, I fitted the HA1 head and the HA2 regions independently as rigid bodies. The HA1 stem region is too small and featureless to be able to be fitted correctly on its own. Fitting of the two rigid bodies was successfully performed into the opened HA structure from fixed HK68 virus samples and the closed HA structure from unfixed HK68 virus samples as two representatives of the open and the closed HA structures (figure 3.14A). I then compared the fits from tow individual rigid bodies to the fits obtained from a single rigid body (Fig.: 3.14B). The comparison for the closed HA structure shows that both fits are identical, as expected. The comparison of the fits for the open structure shows, that the positions of the head and the stem regions after fitting two individual rigid bodies are very similar to the positions after fitting a single rigid body. To quantify the differences between the map coordinates, the root mean square deviation (RMSD) between the coordinates of the HA1 head region and the HA2 chain after fitting a single rigid body and after fitting two individual rigid bodies was calculated in Chimera resulting in 2 Å and 1 Å respectively. I then compared the positions of the individually fitted HA1 head region and the HA2 chain to the positions of the protomers in the closed crystal trimer. If the opening movement of the structure was restricted to the HA1 head region, one would expect, that the HA2 chains remain overlapping while positions of the HA1 head regions would differ. However, in this comparison I found positional differences in both regions. The relative movement of

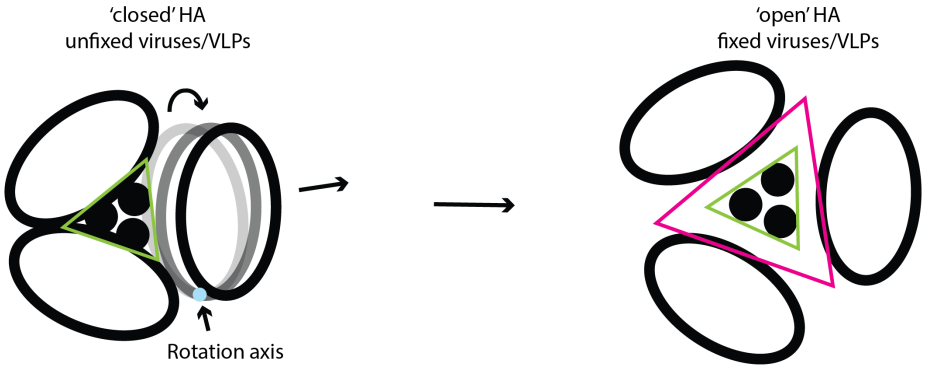

Figure 3.13 Schematic of the proposed movement to transition from the closed to the open HA conformation. Schematics of the closed HA conformation (left) and the transition to the open conformation (right). To facilitate the visualization, the rotation and the outward movement are represented as two steps. The light and dark grey representation of the protomer indicates the rotational movement, while the outward movement is illustrated by the dark grey relative to the black protomer. The green triangle marks the inner edge of the head region in the closed conformation of HA while the pink triangle marks the inner edge of the closed conformation of HA.

the HA2 chain (RMSD: 2 Å) becomes most obvious from the shift of the upper part of the central helix C in HA2 (RMSD: 3.4 Å, Fig 3.14F). Additionally, I observed that the dislocation is stronger in the head region than in the stem region (HA1 head regions RMSD: 6.1 Å) and also stronger in the upper part of HA2 (RMSD: 3.4 Å) relative to the lower part of HA2 (RMSD: 1.4 Å).The calculated RMSD values for the comparison of the multi rigid body fit and the crystal trimer are very similar (a maximum of 1.5 Å deviation) to the RMSD values calculated between the coordinates resulting from a single protomer fit and the crystal trimer, which was expected based on the high similarity between the protomer fit and the multi rigid body fit.

These results demonstrate that fitting two rigid bodies individually for different HA regions yields results very similar to fitting a complete HA protomer. This test further revealed that movement of HA2, in particular helix C, is involved in the suspected opening motion which leads to the HA structure obtained from fixed samples. Further I can conclude that given the accuracy of the fit provided by the resolution of the EM density map, fitting a single rigid body is a valid approach to describe the type of structural differences observed in this study.

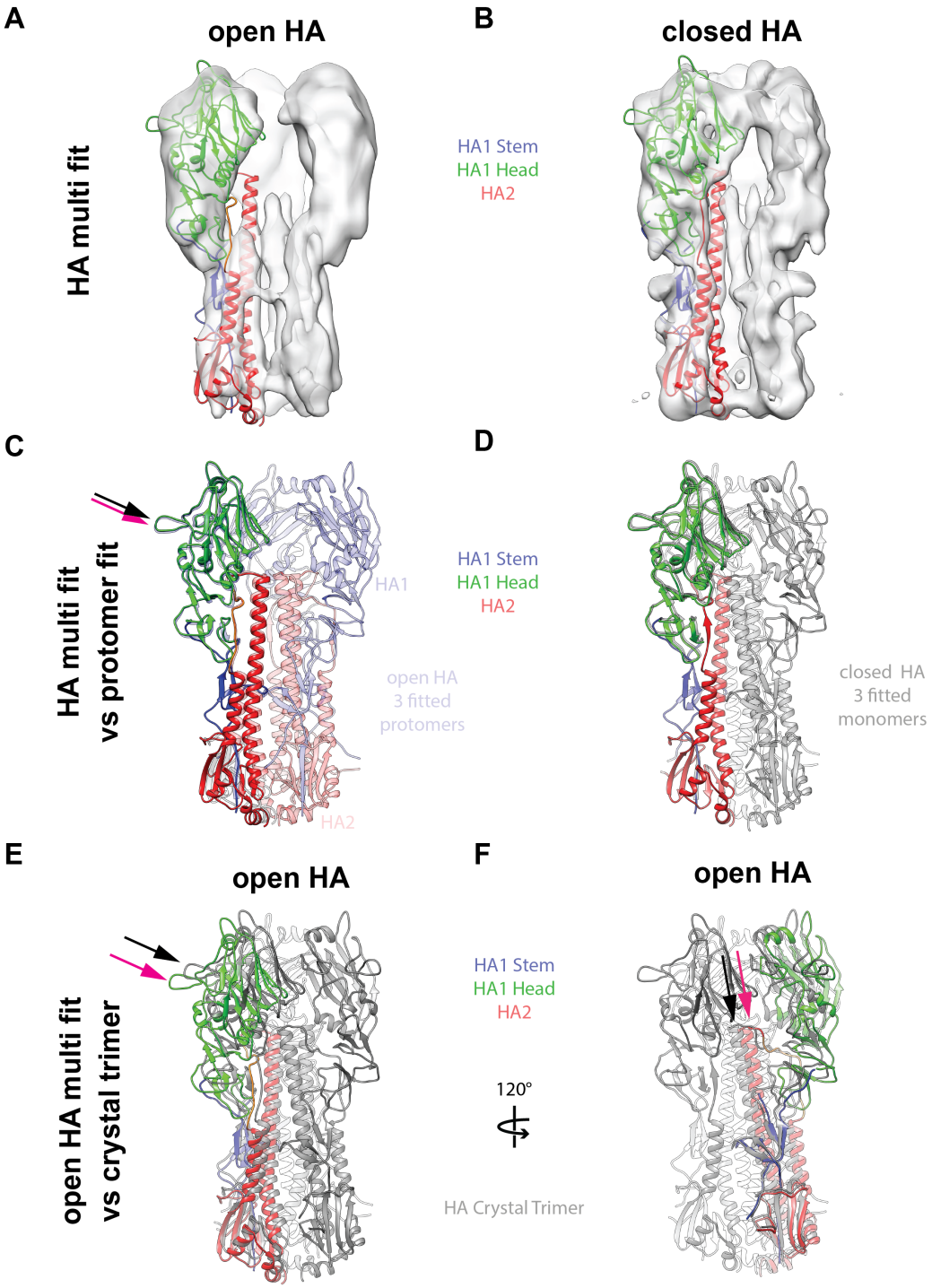

Figure 3.14 Comparison between fitting a complete HA protomer and individual HA domains into obtained HA reconstructions. A) The HA1 head domain (green) and the HA2 chain (red) were fitted as two individual rigid bodies into an open HA structure (fixed HK68 virus). B) Same as A but for a closed HA structure (unfixed HK68 HA). C) Overlay of the complete protomer fit (In transparent colours: HA1 (blue), HA2 (red), black arrow: position of head region) and the fit of two independent rigid bodies for the HA map shown in A (pink arrow) D) Overlay of the fit shown in B (closed HA) and a single rigid body fit in the closed HA structure (grey model) E) Comparison of the fit shown in A (pink arrow) and the protomer arrangement of the crystal trimer (grey, black arrow). F) Same as D but rotated by 120°. Pink arrow: position of HA2 helix C in multi rigid body fit shown in A. Black arrow: position of HA2 helix C in crystal trimer.

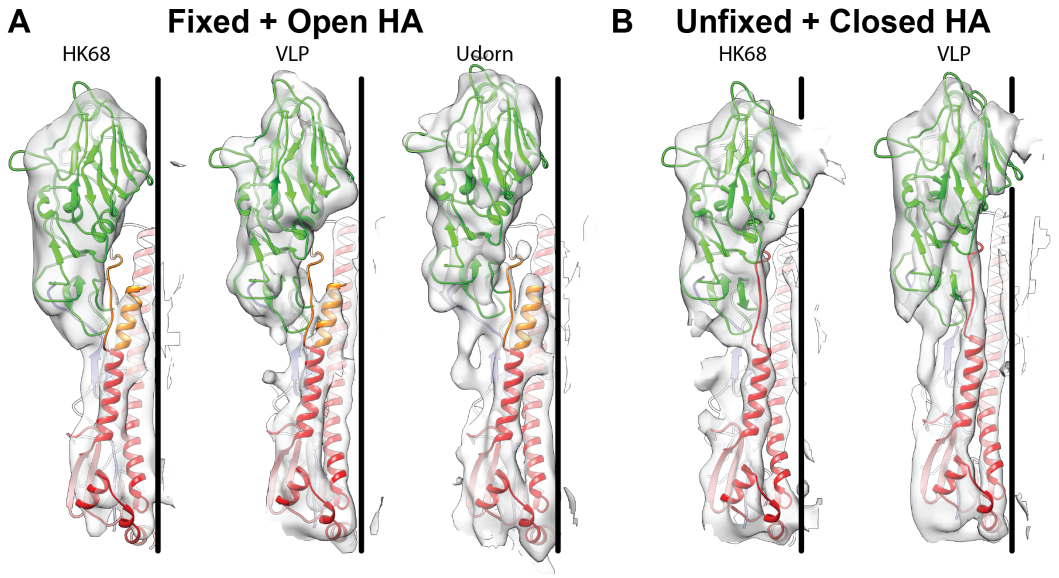

Figure 3.15 Correlation between the density of the HA2 helix A extension and opening of the trimer. A) Single, fitted protomer and the zoned EM-density map around the protomer in side view for open HA from fixed samples. The three fold symmetry axis is shown in black as a point of orientation. The head region of HA1 is shown in green and HA2 is shown in red. The B-loop is shown in orange. A short helix was modelled into the extended density and is also shown in orange. B) The same as in A for closed HA from unfixed samples.

3.2.10 Characterisation of HA structures with an extended density in the HA2 Helix A region

For all HA in situ structures solved from fixed samples with an open protomer arrangement, I found clearly extended density in the region where the HA2 helix A transitions into the B-loop in the corresponding crystal structure. The length of the extended density was measured to correspond to around 3-4 alpha helix turns. The density was found to be slightly tilting away from the head domain and from the original location of the B-loop in the crystal structure (Fig. 3.15A). In contrast, in structures from unfixed samples with a closed protomer arrangement, the density corresponding to helix A merges with the head domain where the B-loop region starts just as found in the crystal structure (Fig. 3.15B).

I subsequently tested if extensions of the HA2 helix A into the B-loop region also occur in previously published crystal structures. Together with Balaji Santhanam (Madan Babu group, MRC LMB) we compared a comprehensive set of published HA structural models with respect to the secondary structure in the B-loop region. Our initial model library consisted of all influenza HA models available in the PDB. The selection was further extended based on an additional database search based on structural homology using the protein structure comparison service at the European

bioinformatic institute (EMBL-EBI), PDBefold (*126*). The final library of models contained 400 entries including a significant amount of redundant structures. We therefore performed sequence based clustering using BLASTclust (National Center for Biotechnology Information, NCBI) to identify a set of representative structures. Representatives of all clusters were subsequently sequence aligned and secondary structure assignments in the region around the beginning of the B-loop were compared.

The analysis revealed three structures with a significantly extended helix in the B-loop region. However, across all structures identified, the end of helix A varied by up to 4 residues (Table 3.1). One class of structures which is helical throughout the complete B-loop region are low pH/post-fusion structures (see section 3.2). Here we identified a representative of influenza A (PDB:1HTM) as well as one of influenza B (PDB:4NKJ). The fact that the B-loop transitions upon low pH into a helix, reflects the general helical propensity of this region (*114*). The only other structure identified which obtains helical elements in the B-loop region found in our structure analysis corresponds to an AB-bound HA structure from the HA subtype H5 (PDB:5A3I), which crystallized as a monomer (*127*).

I compared the identified AB-bound H5 structure to three HA protomers fitted into the fixed HK68 virus structure as an representative of an open HA structure, for which we found the extended density. An overlay of the models was generated by initially minimizing the RMSD between the fitted protomer and the H5 model using Chimera's matchmaker functionality, followed by increasing the overlap with the EM density map by fitting the model into the map in Chimera. From the overlay of the H5 protomer to the arrangement of the three HA protomers, as well as from the attempted fit into the EM density maps, it becomes obvious why the structure crystallized as a monomer. Essentially, the extended HA2 helix A clashes with the central HA2 helix C of the neighbouring protomer (Fig. 3.16A). In addition, the head domain is twisted relative to the position of the head domain of the fitted protomers. The twist is illustrated based on the positions of the short helix on top of HA which is marked as respectively coloured ellipsoids in figure 3.16B. A comparison of the positions of HA2 between the fitted HA protomer and the fitted HA H5 subunit shows that the positions overlap better and that the coordinates of helix A overlap. I modelled a short piece of α-helix into the extended density of the EM-density map to allow a direct comparison between models in the region, where we found extended density in my HA structure. The orientation and position of the modelled, short helix and the position and orientation of the helix A extension found in the H5 model overlap (Fig. 3.16C). The extension in the H5 structure leads to a disruption of the interaction between the B-loop and the inner part of the head domain since the loop folds into a helix which

Table 3.1 Comparison of secondary structure elements in the B loop region
For the secondary structure assignment 'H' refers to alpha helix while '-' refers to the absence of an alpha helix. Structures where the structure information is written in bold letters were identified to have an significantly extended alpha helix relative to all other structures in the analysed region.

PDB	Sequence Secondary structure	Structure info
1HGFB	LNRVIEKTNEK HHHHHH------	Influenza A, pH=7, HA bound to cell sialic acid analogue, X-ray/NMR
1HTMB	LNRVIEKTNEK HHHHHHHHHHH	**A/Aichi H3N2, Membrane fusion pH, X-ray**
1JSDB	VNNIVDKMNKQ HHHHHHHHH--	A H9, Swine, X-ray
1RD8B	VNSVIEKMNTQ HHHHHHH----	A/1918 H1N1 HA0, uncleaved, X-ray
2RFTB	LNSLSELEVKN HHHHH------	Influenza B, with receptor analogue
3EYJB	LNRLIEKTNEK HHHHHH------	A H14, with membrane fusion inhibitor
3HTOB	VNSIIEKMNTQ HHHHHHH----	A/Avian H1N1
3KU3B	VNSVIEKMNTQ HHHHHHHH---	A/1957 H2N2
3S12B	VNSIIDKMNTQ HHHHHHH----	A H5N1
4BSAB	LNRLIEKTNQQ HHHHH------	A H7N9, isolated from humans
4JUKB	VNSIIDKMNTQ HHHHHHH----	A H1N1 (not published)
4NKJA	LNSLSELEVKN HHHHHHHHHHH	**B, membran fusion pH**
5A3I	VNSIIDKMNTQ HHHHHHHHHHH	**A H5, monomer with AB bound**

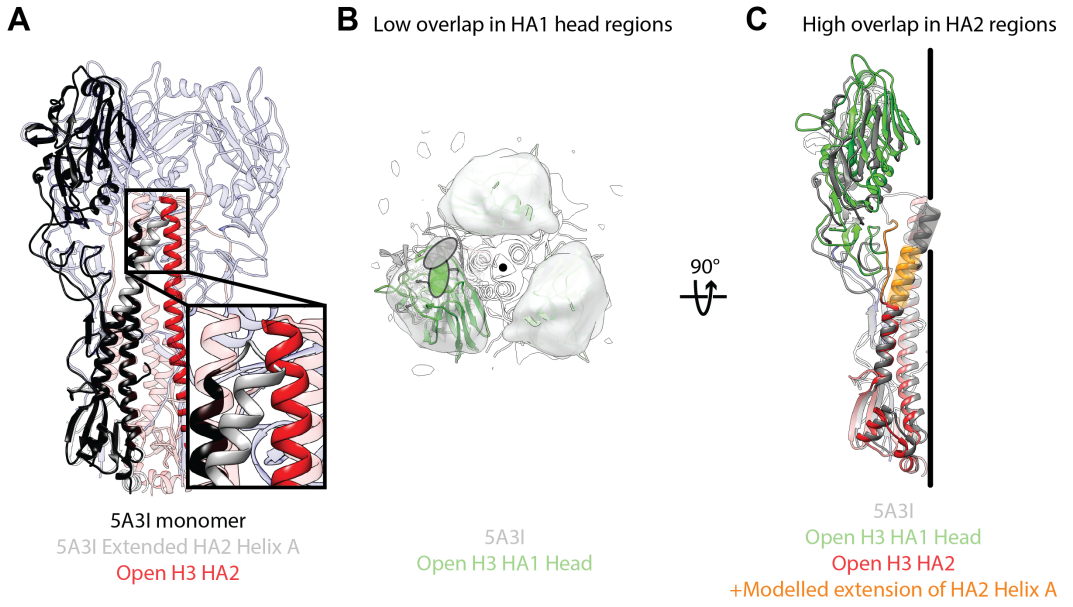

Figure 3.16 Comparison between the open HA in situ structure and the H5 structure with an extended HA2 helix A. A) Overlay of three H3 HA (PDB:4WE4) protomers fitted into the open, fixed HK68 virus HA structure (HA1: blue and HA2: red) and the H5 structure with the extended helix (black). The helix extension is coloured in grey. Inset: The helix extension (grey) clashes with helix C of HA2 (red) of the neighbouring HA subunit. B) Comparison of the positions of a fitted H3 protomer (HA1 head region shown in green) and the H5 extended helix protomer (grey) in top view. The short helices on the top of HA is marked as ellipsoids in the respective colour to demonstrate the orientation and position of the head region. C) Same as B, rotated by 90°. Fitted H3 HA2 is shown in red and a short helix modelled into the extension of the EM density map is shown in orange. The black line represents the three fold symmetry axis.

points away from the head domain (*127*).

3.2.11 HA arrangement on the surface of virus and VLP filaments

Quantification of protein arrangement

During the course of alignment for subtomogram averaging, positions and orientations of each subtomogram are iteratively refined. In addition to an averaged structure, subtomogram averaging yields accurate information about protein arrangement. In the following section I present analyses of subtomogram position data in order to quantitatively describe HA arrangement for the data sets described in the previous sections.

For initial visualization of protein positions identified by subtomogram averaging, I projected a triangular object back into the tomogram at the respective subtomogram

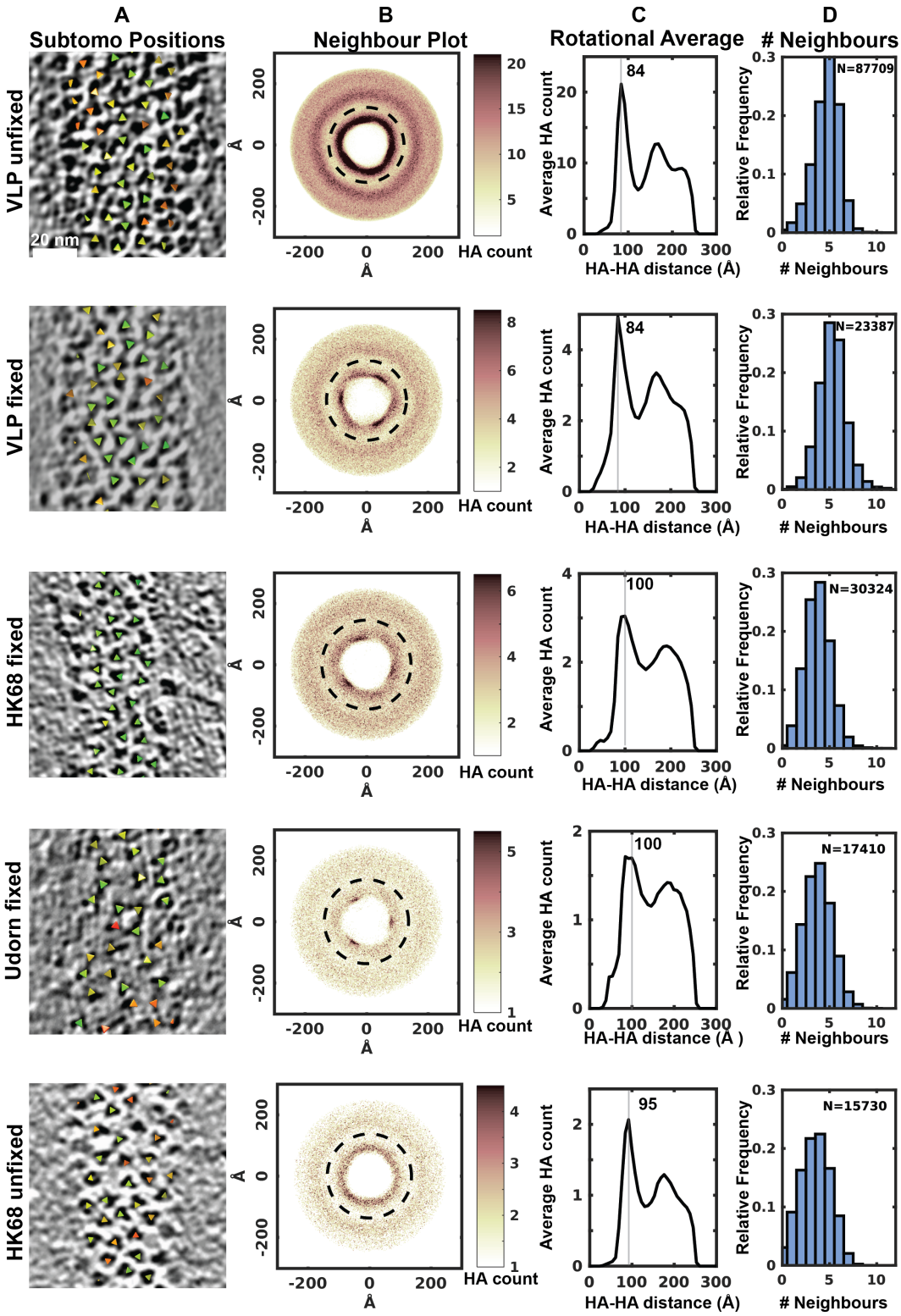

Figure 3.17 Summary of HA arrangement from different samples. A) HA subtomogram positions from representative tomograms are projected as triangles into the corresponding tomogram. For protein densities visible at the edge of the filament, subtomogram positions might not be displayed. B) Neighbour plots for all data sets. Colours correspond to the count of neighbour proteins found in one position. Dashed ring: local minimum C) Rotational average of the neighbour plot. Grey line: the most likely HA-HA distance. D) Distribution of the number of neighbours within the first ring per HA protein.

positions (fig. 3.17A). Glycoprotein densities in the tomogram are clearly matched by triangles reflecting the final positions resulting from subtomogram alignment. To estimate the false positive and false negative rates I compared positions from subtomogram averaging to positions picked manually for a small set of three filaments. I found a high degree of overlap between manual and automated HA detection with less than 1 % of false negative and less than 0.5 % of false positive positions compared to manual detection.

A closer look at the distribution of HA along filaments demonstrates that HA is densely packed but not highly ordered (Fig. 3.17A). To better describe and evaluate the arrangement of all HA proteins throughout the complete dataset, I performed further quantifications of the arrangement. I identified the positions of all neighbouring proteins for each HA subtomogram position. All positions were aligned relative to the orientation of the central HA proteins and subsequently plotted on top of each other. I term this plot 'neighbour plot'. Such a plot allows to identify regularities and preferred relative distances or orientations between proteins which are not necessarily discoverable by eye.

Neighbour plots for all data sets were calculated and are shown in figure 3.17B. For the sake of simplicity I will first introduce the data using the example of the unfixed VLPs, which has the strongest signal in the neighbour plot since it is the largest dataset (Fig. 3.17B, first row): In the unfixed VLP neighbour plot, we can observe a circular point cloud which contains two concentric rings. The absence of distinct clusters suggests that there is only a weak degree of order in the arrangement, as suggested by the impression from the subtomogram positions displayed in the tomogram (Fig. 3.17A). The presence of rings indicates preferred distances between two HA proteins. The inner ring corresponds to a preferred distance between direct neighbours. The ring is not very clearly defined, which reflects that distances between neighbours are not fixed but that there is range of possible distances. The outer ring is weaker since positions of the outer ring of neighbours are effected by the variability of the inner ring. Therefore, I will not further interpret the outer ring. In order to further facilitate the assessment of preferred distances between neighbours I calculated rotational averages of all neighbour plots, which are presented in figure 3.17C. As expected from the presence of rings in the neighbour plot, we find corresponding peaks in the rotational average. The first peak, at 84 Å for unfixed VLPs corresponds to the most likely HA-HA distance. However as indicated by the width of the peak, there is a range of up to around 122 Å within which neighbours are located.

In a final part of the analysis, I counted the number of direct neighbours, defined as being located within the first ring, for each protein position (Fig. 3.17D). The distribution of numbers of neighbours in the first ring follows a Gaussian distribution and peaks at five neighbours for the unfixed VLP dataset. If we compare the numbers from the described analysis to the raw positional data presented in figure 3.17A we can identify several positions which reflect the numbers identified: one protein surrounded by five neighbours with inter-HA distances of around 9 nm.

By comparing the neighbour plots for all data sets (Fig. 3.17B), we consistently find the occurrence of two rings and correspondingly two peaks in the rotational averages (Fig. 3.17B). In addition, I observed clustering of points within the ring in all plots calculated for data obtained from fixed samples. The presence of weakly clustered positions suggests that besides a preferred distance specific positions of neighbours relative to the orientation of each HA molecule are preferred.

By further comparing the results from the rotational averages for all data sets (fig. 3.17C) we find a group of data obtained from VLP samples, where the preferred HA-HA distance corresponds to 8.4 nm while for all data obtained from virus samples, independent on the status of fixation, the preferred distance corresponds to 10 nm. We further can observe differences in the strength of the peaks found in the rotational averages. For the two fixed virus samples (Udorn fixed, HK68 fixed), the peaks are less well separated from the baseline of the plot relative to the other data sets, indicating that the inter HA distances are more variable in comparison to the other data sets.

Finally, if we compare the number of neighbours within the first ring, we find sharp distributions around five neighbours for data obtained from VLPs, while the distribution is shifted towards slightly lower numbers, around four, for data obtained from viruses (Fig. 3.17D). This observation agrees with the slightly larger preferred HA-HA distances observed from the rotational averages. This suggests, that most HA proteins on the virus surface are further apart from each other and that HA-HA distances are more variable relative to the HA arrangement found on the surface of VLPs.

Overall the quantification and comparison of the HA arrangement for all data sets has resulted in two main observations: I identified preferred orientations of neighbouring HA proteins in fixed samples which are not detectable in unfixed samples. The preferred HA-HA distance is higher and the number of neighbours lower for virus samples relative to VLP samples.

3.2.12 Structure of two HA proteins with a fixed relative orientation found in fixed sample.

In a proof of principle experiment, that was performed together with Zunlong Ke, we tested if further evidence for preferred relative positioning of neighbouring HA proteins can be detected as observed in the neighbour plots for all fixed samples (see section 3.2.11) (Fig. 3.17). The presence of preferred relative positions of HA proteins suggests, that fixation, might lead to specific HA-HA crosslinking. To test this hypothesis we selected all HA positions for which a neighbour was found in any of the three clusters visible in the neighbour plot calculated from fixed VLPs (Fig. 3.18A). Selected subtomograms were rotationally aligned to assure that the direct neighbour is found in the same relative position in all subtomograms (Fig. 3.18A). Subsequently, subtomogram averaging and alignment to refine the average, without applying three-fold symmetry, was performed. In the resulting average structure, two HA proteins, one in the centre and one in the position indicated by the cluster, were visible. To centre the two aligned HA molecules within the subtomogram, the average was subsequently shifted and further alignments using a mask around both protein densities were performed. The density of the second protein improved further during the alignment and the relative orientations of the two HA proteins can be clearly identified (Fig. 3.18B). Individual protomers of the HK68 HA crystal structure were fitted as rigid bodies into the obtained HA densities. By comparing the relative orientation of the fitted high resolution model, we found that the closest point between the two structures is marked by a loop (140KRGPG144) at the outside of the head domain (Fig. 3.18C). The loop contains an arginine as well as an lysine, two side chains which contain amine functional groups and which represent potential targets for aldehyde crosslinking (*128*, *129*). The distance between the two opposing loops, which represents the shortest distance between two HA proteins, corresponds to 16 Å.

3.2.13 HA density and HA tilt angle in different samples

In addition to studying HA arrangement, I calculated the density based on the number of subtomogram positions per virus or VLP filaments. Since viruses are tube-shaped I can calculate the surface and consequently the density for each virus or VLP filament. The average density for each dataset is shown in figure 3.19A. Consistent with the results from the neighbour plot analyses, HA densities for viruses are lower in comparison to the HA density for VLP particles. While densities across virus and VLP

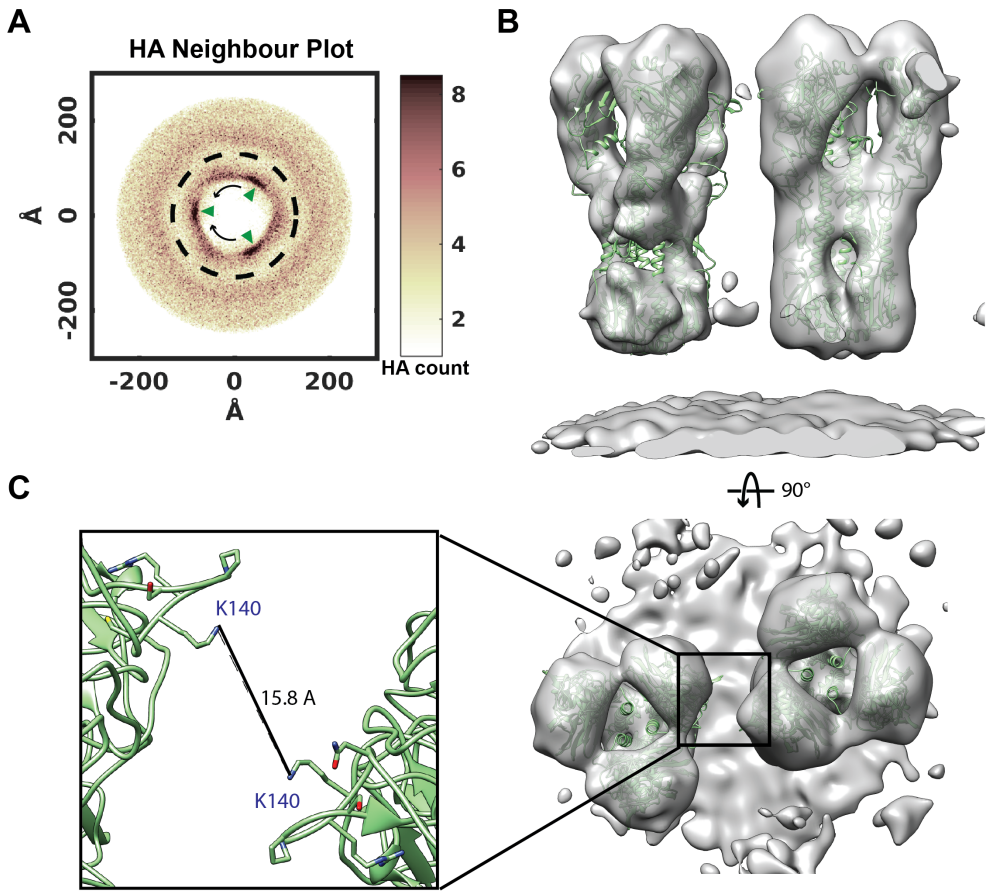

Figure 3.18 Reconstruction of two HA proteins in a fixed relative orientation. A) Neighbour plot of HA from fixed VLPs. Green arrow heads indicate clusters of neighbouring HA proteins. Black arrows indicate the rotation applied to align positions rotationally prior to averaging. B) Final, sharpened average of two neighbouring HA proteins identified from clusters in the neighbour plot shown in A. HA crystal structures were fitted into the EM density map. PDB: 4WE4. C) Top view of structure shown in B. The inset focuses on two lysine side chains in the fitted crystal structure which are pointing towards each other and which mark the closest distance between the two HA proteins.

samples show only little variation within the respective group.

From highly magnified sections from tomogram slices it occurs that HA proteins are not entirely perpendicular with respect to the filament surface but instead are mostly slightly tilted (Fig. 3.19B). To understand the range and distribution of tilt angles, I calculated the tilt angle for each HA protein on the surface. Tilt angles were calculated based on subtomogram position and orientation relative to the filament surfaces. The distribution of tilt angles for each dataset is shown in figure 3.19C. Distributions are very similar across the different data sets following a Poisson distribution with a peak at 13°. Therefore, no correlation between the calculated HA density and the tilt angle distribution can be identified.

The tilting of HA affects the shortest, possible distance to neighbouring molecules, which I will refer to as tilt distance. I calculated the tilt distance for all HA positions (Fig. 3.19B). Calculated tilt distances in correspondence to tilt angles are shown on the top x-axis in figure 3.19C. For the most abundant tilt angle of 13° the tilt distance corresponds to 1.7 nm. The calculated values for the tilt distance demonstrate that for the largest fraction of HA proteins the tilt distance ranges below 4 nm. Even if two HA proteins tilt towards each other, the total tilt distance from these two proteins (8 nm) is still shorter than the HA-HA distance identified for most HA proteins (Fig. 3.17C).

3.3 Discussion

At the beginning of the presented study I tested an alternative method for sample preparation to improve virus particle preservation. Purification negatively affected the integrity of the glycoprotein layer in comparison to non-purified samples. In addition, the predominantly occurring virus morphology was shifted towards shorter filaments or spherical particles for purified in comparison to non-purified samples. This observation is consistent with previous reports and to my knowledge, the long filaments observed here and by others in non-purified samples, have not been observed before in purified samples (*16*, *32*). The described impact from purification on morphology can have several reasons: Either longer filaments are more prone to damage and breakage through physical stress during purification or long filaments are selected against during steps of purification or potentially a combination of both effects.

The effects of purification on the structure of the glycoprotein layer observed here, match reports from previous studies where purification and the chosen host organism (e.g. cell culture vs egg) was shown to influence appearance and density of

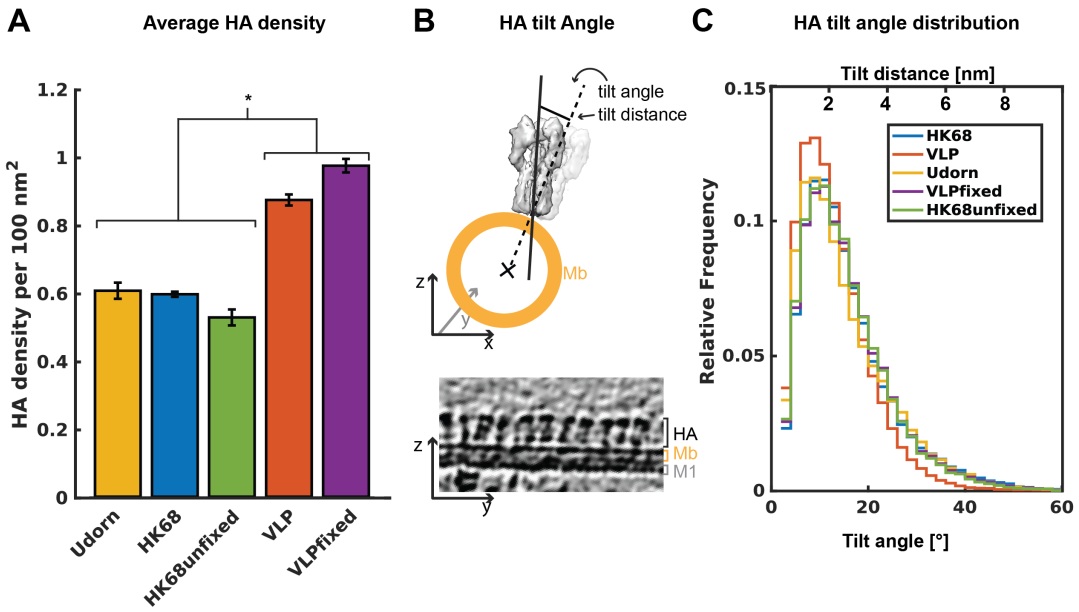

Figure 3.19 HA average density and HA tilt angles. A) The average HA density was determined for each individual virion and averaged per dataset. Error bars represent the standard error of the mean. Significance for differences in average density was tested across all data shown using an one-way ANOVA. p-values < 0.05 are indicated by *. B) Top: Illustration of the tilt angle and the tilt distance of HA. Bottom: Magnified section from a tomogram slice showing the HA layer on top of the membrane and demonstrating that HA is not oriented perfectly normal to the filament surface. C) HA tilt angle distribution. Curves are the outlines of the corresponding histograms.

the glycoprotein layer (*130*, *131*). At the same time, others have presented cryoEM images of purified influenza virus particles, where the glycoprotein layer appears more preserved indicating that further optimization of the purification protocol might improve the quality of purified samples (*15*). Overall, my data demonstrate that it remains important to be aware of possible effects resulting from virus purification, in particular when subsequent data analysis methods are used, where the effects are not directly recognizable.

Finally, the sample preparation workflow used here for influenza virus particles can be readily extended to target other enveloped viruses, which are prone to experiencing damages during purification. In a collaboration with Martin Obr and Maria Anders (Hans-Georg Kraeusslich lab), I tested the presented sample preparation protocol for HIV particles. After adapting the workflow to the non-adherent cells used for HIV particle production, we successfully imaged produced virions by cryoET directly in the vicinity of their host cells.

In summary, I have shown that an optimized sample preparation protocol without virus purification can be used to obtain well-preserved virus particles in thin and clean vitreous ice around infected cells which allows acquiring high quality tomograms by cryoET for subsequent processing by subtomogram averaging.

Using cryoET data of non-purified virus and VLP particles I have established a data processing workflow that allows to obtain HA in situ structures at sub-nanometer resolution. Subtomogram averaging of HA was successfully applied to solve HA from five independent samples and for all datasets I was able to obtain an HA reconstruction of at least 9 Å resolution. The unfixed VLP data, the by far largest dataset containing 135,000 asymmetric units, yielded a structure with a higher resolution of 7.5 Å. This suggests that a massive increase in the amount of processed data might lead to moderately improved resolutions for samples where smaller datasets were processed. However, it remains unclear which aspects are limiting for the resolution obtained for the current number of particles. The outcome of my project and the observation of an open and closed conformation in dependence on fixation, suggests, that HA occurs in multiple conformations on the virus surface. I therefore conclude that structural heterogeneity is an important limiting factor for the achievable resolution of the obtained reconstructions of HA. Despite multiple attempts from my side to reduce heterogeneity by classification none of the tests were successful to distinguish different HA conformations on the virus surface.

While subtomogram averaging structures of HA have been obtained before, resolutions previously were limited to around 20 nm, where secondary structural elements are not yet resolved (*132*). The in situ structures presented here provide sufficiently high resolution to resolve alpha helices and to compare the obtained in situ structure with high resolution crystal structures. Based on the resolution obtained, I was able to detect differences between the in situ HA structure obtained from fixed samples relative to the crystal structures which had not been reported before.

3.3.1 Aldehyde fixation captures HA in an open conformation

I performed a number of control experiments to validate that aldehyde fixation is indeed the cause for the observed structural differences between HA reconstructions directly from fixed HK68 virus samples and previously determined HA crystal structures. From a direct comparison of a set of in situ HA structures obtained from fixed and unfixed viruses and fixed and unfixed VLPs, I consistently found, that HA reconstructions from unfixed samples match the structure of existing HA crystal structure trimers, while in HA reconstructions obtained from fixed sample, the HA subunits are arranged in a open conformation unknown from crystal structures. Based on the obtained results I suggest that aldehyde fixation with PFA or GA captures HA in a transient conformation, which is present at much lower frequency and therefore essentially remains undetected in averages obtained from unfixed samples. I reason that the presence of an open and closed state relates to a continuous movement between these two states on the virus surface described previously by fluorescence techniques (*133*) and which is similar to the breathing motion known for the trimeric HIV glycoprotein (*134*).

Based on this hypothesis it might be possible that the open conformation of HA, albeit as a much lower fraction, is also present in data from unfixed samples. Classification of subtomograms could therefore be used to obtain structures for both conformations from a single dataset. I therefore attempted to show co-occurrence of different HA conformations on the same virus by structure classification. I tested different classification methods similar to what was described previously in (*135*), RELION 3D classification (*136*), multi-reference alignment as well as principle component analysis based clustering. However, so far none of the tested classification approaches was able to identify subclasses of HA subomograms exhibiting different conformations. Instead all classification attempts have led to identifying the missing wedge as the main feature to classify on. The identification of the missing wedge from classification approaches has shown that classification methods for subtomogram averaging require further optimization to be applicable in cases where small and potentially non-discrete structural differences are to be discerned as it is the case for HA.

In general, classification in subtomogram averaging so far has been mostly used to distinguish major structural differences such as the presence, absence or rearrangement of an extra domain or a larger structural element (*135*, *137*, *138*). However, further improvement of classification algorithms, signal weighting and wedge masking might enable identification of co-occurring open and closed conformations of HA on the same virus in the future.

3.3.2 Functional roles of open and closed HA conformations

In the presented study I have identified an open and closed HA structure, which I relate to the presence of breathing motion occurring on the virus surface. I compared the open HA conformation found in fixed samples to the closed conformation found in unfixed samples by performing multiple rigid body fitting experiments. Thereby I was able to accurately describe the differences between the open and the closed conformation. I found that the opening of HA does not only result from the movement of the head region outwards, but that it is accompanied by an outward movement of the central HA2 helix C and a partial refolding of the HA2 B-loop into a extension of HA2 helix A.

A very similar type of opening of the HA trimer as described here was recently suggested in a study by Turner et al., where the authors found a more open H7 structure upon interaction with a head-binding antibody (AB) (*139*). The authors came to the conclusion that the AB stabilizes a less stable, slightly more open conformation which occurs as part of HA breathing motion on the virus surface. In the context of the opened structure, the authors further identify a short extension of helix A which however is less pronounced than what I have observed in my data. The structural similarities of the HA open conformation found across two different strains described by Turner et al. for H7 and in this present study for H3 further support the interpretation, that the open HA conformation found in fixed samples relates to breathing motion.

In the HA structures obtained from fixed samples, I observed the presence of extended density corresponding the extension of the HA2 helix A. The corresponding region, which occurs as a loop in HA structures at neutral pH (pH=7), transitions into one part of an extended helix in low pH post-fusion structures (*114*) (Fig. 3.2). In addition, a comparison between HA reconstructions from fixed samples, which show extended density, and a previously published AB-bound H5 monomeric crystal structure with an almost completely extended helix A (*127*) has revealed similarities with respect to the orientation of the found helix extension. The authors of this previous study concluded that the monomeric state and the resulting lack of interaction between

the three subunits, in particular between the three central helices of HA2, is connected to a refolding of the B-loop into a α-helix. This refolding of the B-loop leads to reduced interactions between the B-loop region and HA1 (*127*). The described HA2 helix A extension is part of the final post-fusion HA structure (*127*). Their structure further shows, that a significant extension of helix A is possible while HA1 is present and folded. This suggests that the presence of HA1 seems to prevent the complete transition into the fusiogenic form, where the fusion peptide is extended at the top (*127*). The structure presented by Xiong et al. might represent a maximal extension of the helix in a situation where HA1 is still present. The amount of helix extension is however sterically hindered in a trimeric arrangement of HA where the central helices still interact. In this context, the open HA structure described here could represent an moderately extended helix within the limits of a trimeric HA arrangement. The reduced interactions between the three subunits during the opening of the structure results in partial refolding of the B-loop and consequently to a reduced interaction between HA1 and HA2 (Fig. 3.20).

To summarize, I propose that the open HA structure found here reflects the HA breathing motion on the virus surface. The detected breathing motion, reflects the metastable, fluctuating character of the neutral pH structure (*133*, *140*). My data and data from others (*139*) suggest that breathing is generated by the opening and the closing of the HA1 head region and includes movement of the HA2 central helices. At the same time, the open conformation induces a partial transition of the B-loop in an alpha helix. The extension is limited by the interaction between the central helices (HA2 helix C) as well as by the presence of the HA1 head region (*127*, *140*). The described open structure precedes the transition towards the extended intermediate (*133*) (Fig. 3.20) and further into the irreversible low pH post-fusion structure (Fig. 3.1).

3.3.3 Quantification of HA arrangement

The identification of HA positions in situ from subtomogram averaging enabled detailed analyses of HA arrangement on virus particles. These analyses revealed that HA is not randomly distributed on individual filaments. Neighbouring HA proteins are located within a preferred range of distances from each other. HA arrangement on viruses and VLPs differed with respect to the preferred HA-HA distance and shorter distances were preferred in VLP samples. This observation was confirmed by calculating the average HA density per filament which confirmed a higher HA density for VLP filaments. In comparison to previous approaches data shown here provide the most automated and extensive characterisation of HA arrangement. In a number of previous

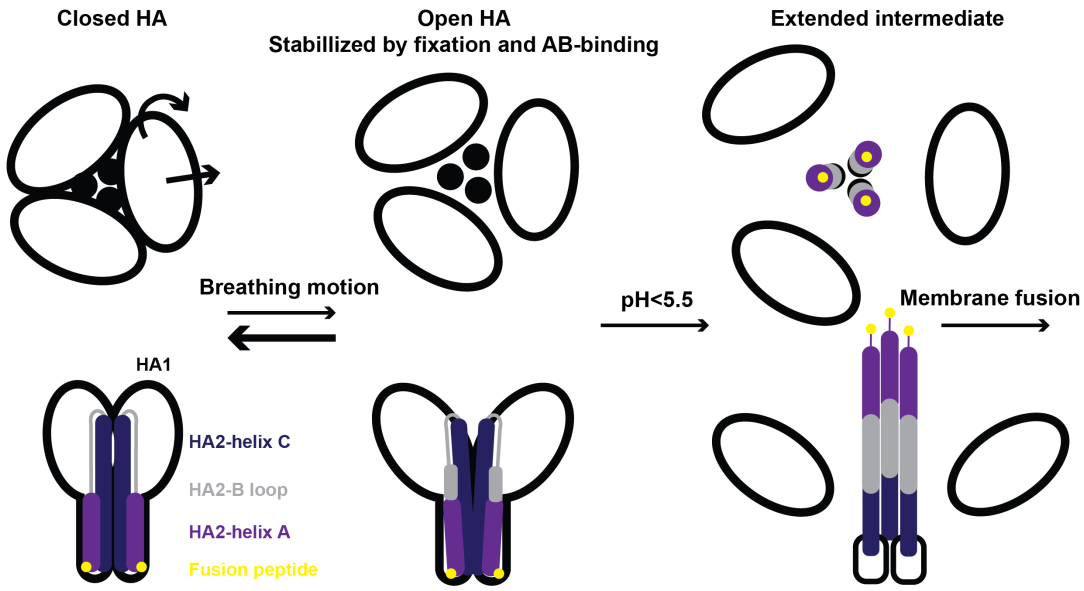

Figure 3.20 Schematic of the proposed breathing motion of HA in the context of HA-mediate membrane fusion. The scheme demonstrates how HA might continuously transition between a more stable closed and a more unstable open conformation. Upon exposure to an environment with lower pH, HA from the open conformation might irreversibly transition into the extended, intermediate to interact with the opposing host cell membrane and subsequently induce fusion of the two membranes. In this model, HA1 and HA2 dissociate to allow the transition into the extended intermediate.

studies, the density was estimated by manual counting of proteins either from 2D or 3D EM images (*130*, *141*). This type of quantification is very time consuming and limits analyses to a small subset of proteins. Additionally, measuring distances directly from projection images might underestimate the true protein-to-protein distances. Harris et al. quantified the HA distribution on the surface of egg-grown, purified X-31 (H3N2) virus and reported slightly irregular spacings of around 11 nm from manual assessment in 3D (*141*). They did not find any significant differences between spherical and slightly elongated virus morphologies. Their average spacing corresponds to what we found as preferred HA-HA distances from HK68 and Udorn virus samples. Waselewski et al. performed a more automated assessment of HA positions along the top and the bottom surface of X-31 and Udorn virus filaments. They report preferred distances of 9 nm to 10 nm between neighbouring HA molecules which is consistent with my data on HA arrangement on virus filaments (*121*).

For future analysis, the presented workflow provides a tool to compare data of HA arrangement for viruses from different strains, with introduced mutations, different morphologies or after AB-incubation. Comparing the HA arrangement for different conditions will allow to assess if and how any of the tested conditions influences the HA arrangement. This is especially relevant for studying AB binding. It remains to be shown if HA arrangement is directly affected by AB binding and if HA proteins are mobile enough to rearrange upon AB interactions. It could be possible that HA arrangement limits the binding of different ABs such as stem binding ABs.

The accurate descriptions of density and arrangement of HA found by us and also previously assessed by others (*121*, *141*) for IAV particles suggest that the dense, yet un-ordered arrangement represents the closest, energetically favourable arrangement of the trimeric HA protein. In addition, the lack of a specific lateral order in the HA arrangement makes it unlikely that HA arrangement is dictated by M1, which is expected to have a more ordered arrangement (*15*) (see chapter 4). The observation that order seems to correlate with an increase in density as observed for VLP particles further suggests that the arrangement is determined by density. The fact that neither in our data nor in the data from others particles with a significantly lower density of HA was found further suggests, that a minimal density of HA may be required during assembly to allow deformation of the host cell membrane during virus assembly. Potentially this minimum density is slightly higher for VLPs compared to viruses which would explain the differences in observed density. This difference could be further related to the absence of the vRNPs in VLPs and the resulting requirement for a higher HA concentration to allow initiation of filament formation. At the same time,

the high similarity in morphology between viruses and VLPs despite the differences in HA-density suggests that the HA arrangement does not influence virus particle shape including the particle diameter.

3.3.4 Fixation might induce HA-HA crosslinking

The analysis of HA arrangement after aldehyde fixation revealed a preferred relative orientation of HA proteins. In the data from aldehyde treated VLPs, HA proteins with neighbours in fixed preferred positions were subsequently processed by subtomogram averaging. We obtained a structure which contained two specifically oriented HA proteins. The two molecules are arranged in a two fold symmetric relation to each other and the edges of the head region are less than 20 Å apart from each other. I identified the lysine residue K140 from crystal structures fitted into the subtomogram averaging structure as a potential aldehyde crosslinking target. K140 is located in a loop at the outer surface of the head region and points into the direction of the opposing HA protein. The length of a GA molecule is 7.5 Å which corresponds to half the distance found between the two lysine side chains on the neighbouring proteins. GA, however may polymerize in an aqueous environment and form elongated species, which then might provide the length required to bridge the two residues. It remains to be shown in the future if an open HA structure, potentially resulting from HA-HA crosslinking, only occurs in the presence of GA, the longer aldehyde, or if pure PFA treatment would result in the same effect. In summary, we propose that the conservation of the open HA trimer structure in fixed samples results from inter molecular cross links between two neighbouring HA molecule, potentially via the residue K140.

Since virus inactivation by PFA/GA fixation is a widely employed method to inactivate infectious virus samples for subsequent cryoEM studies, it is likely that similar effects occurred but remained undetected in previous virus studies. It is likely that flexible and potentially mobile proteins on the outside of the virus, such as viral glycoproteins, are most prone to show structural effects based on aldehyde treatment. However, as described in this study, resulting effects are likely to be small and sufficiently high resolutions below 10 Å in case of in situ structure determination are required to be able to detect resulting structural differences. In addition, virus inactivation by aldehyde fixation is commonly used in vaccine production. Given that the differences found for the HA structure from fixed and unfixed samples are mostly related to the head domain, which is the dominant antigenic region, the structural differences could potentially affect AB binding through changes in accessibilities of the head domain in the open relative to the closed structure.

3.3.5 Functional roles of the HA tilt angle

As part of the analysis of HA arrangement I have calculated and analysed the tilt angle of HA molecules. While it is known that HA has a flexible linker between the transmembrane domain (TMD) and the ectodomain, this flexibility had been mostly ignored in the past. Only recently, the presence of flexibility in this region has been taken into account and was critical to solve the TMD of HA in micelles by single particle cryoEM (*108*). While one could imagine that there is a correlation between HA-HA distance and tilt angle, I did not find any correlations. The geometry of HA would allow tilt angles up to 60° until clashing with the neighbouring protein. However given that there preferred angle ranges around 13°, it seems likely that the tilting is limited by the structure of the flexible linker rather than by protein density. Given that HA is most likely not statically fixed in one tilt angle but is flexible, this flexibility could increase the chance of the virus to establish interactions with host cell receptors. At the same time presence of tilted HA proteins could impact AB binding considering that HAs at different tilt angles provide different steric environments. Finally, HA tilting together with the variability of HA-HA distances is likely to effect HA stem accessibility for stem-binding ABs.

3.4 Conclusions

1. IAV purification biases the distribution of observable particle morphologies towards smaller particles.

2. The in situ structure of HA is affected by aldehyde crosslinking, a common method for virus inactivation, which is used in laboratory settings as well as for vaccine production.

3. Aldehyde crosslinking, potentially between neighbouring HA proteins, captures HA in a structure where the HA trimer is more open relative to the structure from unfixed viruses.

4. The presence of an open HA structure on the virus surface upon sample fixation is another indicator of HA breathing motion. Breathing motion might facilitate further conformational changes which occur upon exposure to low pH.

5. For all in situ HA structures occurring in a more open conformation, HA2 Helix A is extended, an effect which might be induced by the loosening of the interaction between the HA1 head and the B-loop.

6. HA is arranged with lose order on the virus surface where a single protein is preferentially surrounded by four to five neighbours in distance from 8 nm to 10 nm. VLPs appear to have a more ordered arrangement and are more densely packed relative to viruses.

7. Despite the high HA density, HA is not oriented orthogonal relative to the virus surface but most proteins are tilted. The presence of tilted HA proteins reflects the suggested flexibility of the linker that connects the extracellular domain to the TMD.

8. The established analysis provides a framework for future studies of HA-antibody binding in situ.

Chapter 4

In situ characterisation of M1

4.1 Introduction

M1 is a 28 kDA-sized protein which is essential for morphology, stability and integrity of IAV particles (Fig. 4.1). During the virus life cycle, M1 is known to assist in various different processes including the mediation of the assembly process and the export and transport of newly formed RNPs from the nucleus to the assembly side (Fig. 4.2). To date, no full length structure of the protein is available. Only a partial structure comprising residues 2-165 (out of 252) was solved by X-ray crystallography (*142–146*) (Fig. 4.1C). Despite successful expression of the full-length M1 protein, proteolytic cleavage, presumably of an exposed linker, and resulting fragmentation of the polypeptide was repeatedly reported in the past (*142*, *143*). Only the N-terminal fragment crystallized which since then is termed the M1 N-terminal domain (NTD) (*142*, *143*) (Fig. 4.1C). Since the CTD fragment could not successfully by crystallized it was hypothesized that the C-terminal domain (CTD) fragment is flexible, unfolded or unstructured (*147*). Characterizations of the full-length M1 protein by small angle X-ray scattering (SAXS) have suggested that M1 appears as an elongated monomer (*147*) and characterisation by circular dichroism has shown that the CTD is at least partially helical (*143*). The M1 NTD appears as a compact, almost cube-like protein domain consisting of nine helices which are inter-connected by loops (Fig. 4.1C). The first structure of the IAV M1 NTD was obtained at pH=4 (*142*). Later, the structure of the NTD was solved again for different strains and at different pH (*143–146*). Differences in pH during crystallisation lead to differences in crystal packing and orientation of the M1 NTD monomers within the dimer which represents the asymmetric crystal unit (*144*).

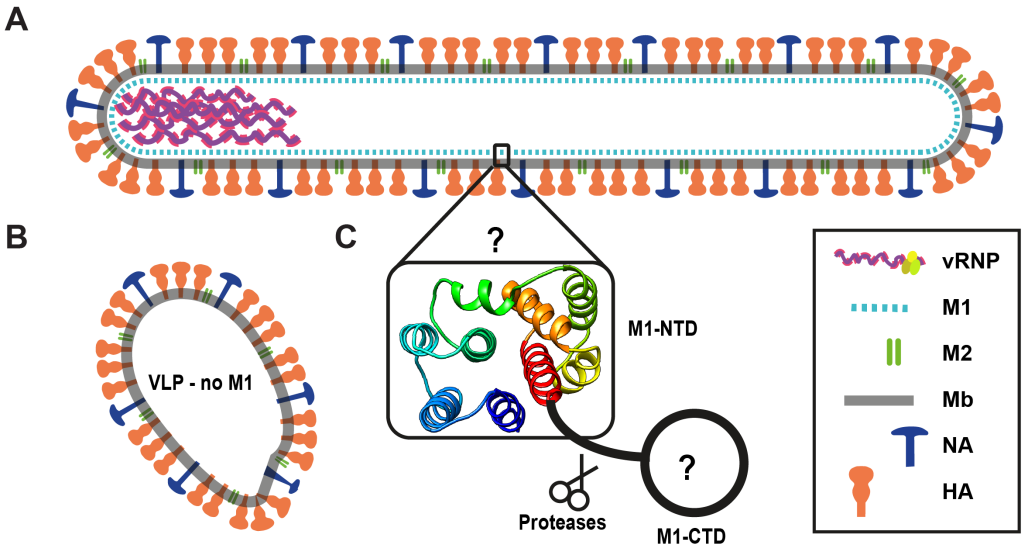

Figure 4.1 Structure and arrangement of M1. M1 forms a layer of unknown arrangement underneath the virus membrane. B) The absence of M1 leads to a irregular morphology of the formed particle (*45*). C) Crystal structure of the M1 NTD (PDB:1EA3). The structure of the M1 CTD remains unknown. Proteolytic cleavage after the M1 NTD was observed in the past. In legend: Mb=membrane.

4.1.1 M1 inside the virus

Influenza is a polymorphic, enveloped virus. Therefore, the structure of IAV particles is not accessible to crystallography but has been studied by EM (*15*, *141*, *149*). In EM images of influenza particles, a continuous protein density is visible underneath the membrane. The thickness of this density layer matches the molecular weight of M1. It was therefore reasoned that M1 forms a matrix layer underneath the viral membrane (*149*). While the distinct shape of HA is clearly recognizable from EM images, the small size and the globular shape of M1 precluded an unambiguous identification of M1 directly from EM images. In 2D projection images, the thin M1 protein layer is masked by the dense glycoprotein layer which inhibits the recognition of any additional structural features from these images (*149*). Therefore, improved visualization of the matrix protein layer was only achieved after Bromelain treatment of IAV particles, which cleaves the ectodomain of HA as well as by cryoET studies (*15*, *113*). Using a combination of these two methods and Fourier analysis of obtained images Calder et al. described the matrix layer as a regular lattice underneath the membrane (*15*).

Results from multiple in vitro studies support the hypothesis that the membrane-proximal protein layer is indeed formed by M1. In vitro flotation assays of full-length M1, M1 NTD and M1 CTD together with negative charged lipids showed that full-

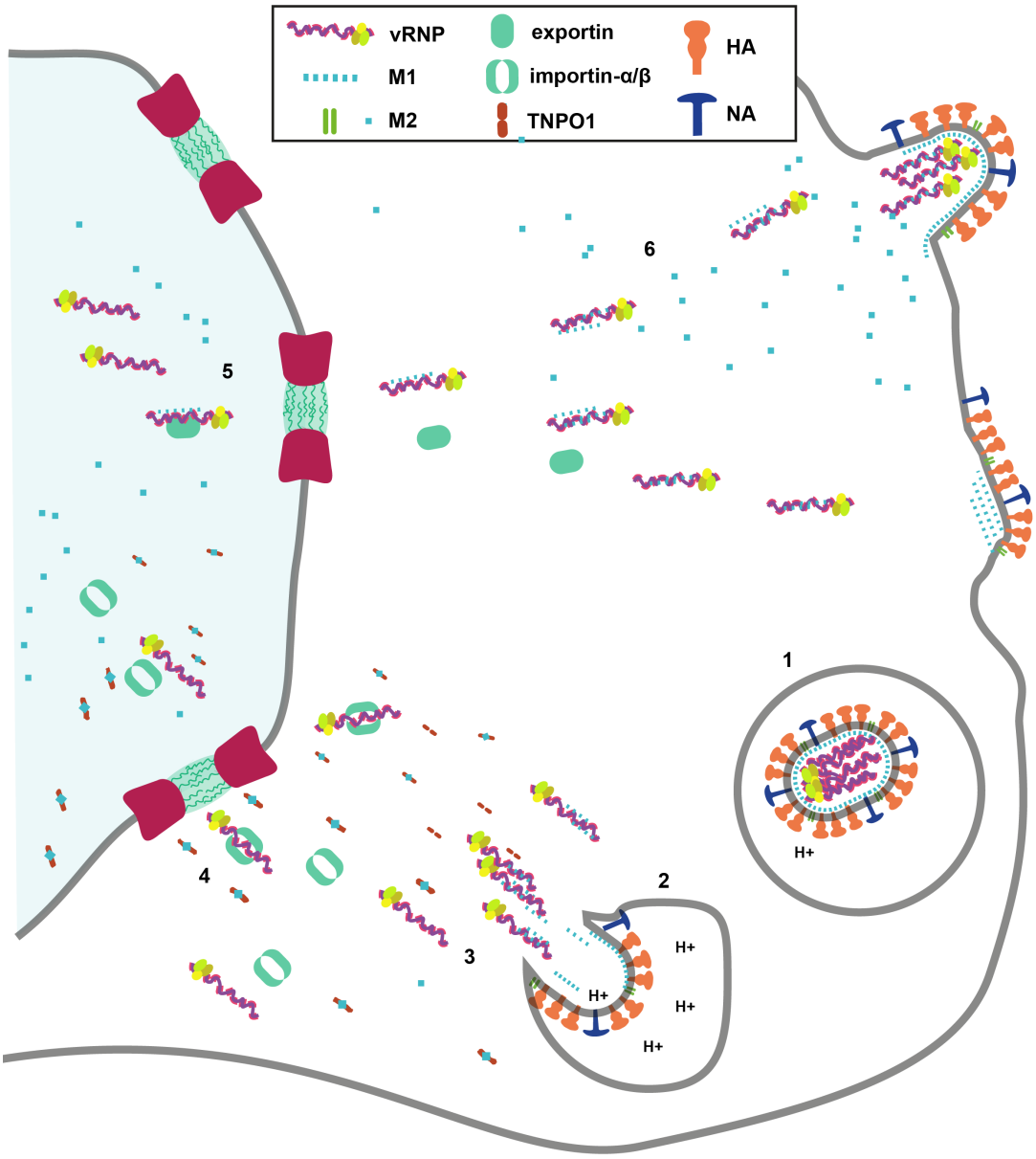

Figure 4.2 M1 in the IAV life cycle. 1) An endocytosed IAV particle within an acidified endosome. M1 form a layer underneath the viral membrane 2) The viral membrane and the endosome membrane have fused. M1 has dissociated from the membrane and RNPs still interacting with M1 have been released into the host cell cytoplasm. 3) M1 interactions with the importin protein TNPO1 result in dissociation of M1 from vRNPs (*148*). 4) Subsequent to the release from M1, M1-TNPO1 complexes and vRNP-importin complexes are transported into nucleus. 5) Replicated RNPs interact with M1 and are exported from the nucleus after binding to exportins. 6) Transport of RNPs and M1 to virus assembly sites at the host cell plasma membrane via unknown mechanisms.

length M1 as well as the M1 NTD but not the M1 CTD were found to interact with lipids (*150*). Based on distinct surface charge distribution of the M1 NTD structure, it was hypothesized that M1 interacts with membranes based on electrostatics (*143*, *150*, *151*). Instead, others suggested that M1 interacts with membranes via the insertion of hydrophobic side chains, which would require inside out folding of the protein (*142*).

In addition to the interactions with the membrane, M1 is believed to interact with the cytoplasmic tails of HA, NA and the transmembrane ion channel M2. The type and impact of interactions between M1 the glycoproteins remains disputed. While no direct interaction has been shown biochemically, it was reported that the presence of the glycoproteins increases the fraction of M1 which is being localized to the membrane during virus assembly (*47*). In contrast, others have suggested that the cytoplasmic tails are not relevant for M1 interactions since assembled virions were observed for glycoprotein mutants lacking their cytoplasmic tails (*152*). Further it was suggested that inside virions, M1 interacts with vRNPs (*150*) and is essential for the cohesion of genome segments at the tip of virus particles (*153*).

4.1.2 M1 functions during the virus life cycle

Upon entering a new cell for infection, endocytosed IAV particles undergo HA-coordinated membrane fusion of the viral and endosome membrane once exposed to low pH (Fig. 4.2 1). The membrane ion channel M2 opens to allow protons and later K^+ ions to enter the virus core. M1 dissociates from the membrane once the pH is low enough inside virus particles (*154*) (Fig. 4.2 2). In accordance with this model, in EM images of IAV particles acquired at low pH, the protein layer underneath the membrane was found to be thinner, disrupted or absent (*113*). Additional insights into the effects of pH were gained from M1 NTD protein crystals which formed at low pH conditions. The differences in the occurring dimer interfaces relative to the M1 NTD interface in neutral pH crystals have suggested that pH influences the arrangement of M1 (*144*).

Even after dissociation from the viral membrane and release of vRNPs into the host cell cytoplasm, M1 was shown to maintain interactions with the vRNPS (Fig. 4.2 2). Debundling of vRNPs is only achieved, once M1 is bound by TNP01 which releases the vRNPs. Only upon dissociation from M1, vRNPs are imported into the nucleus (*148*, *155*, *156*) (Fig. 4.2 3). At the same time, M1 itself is equipped with a nuclear localization sequence (NLS) (101RKLKR105) which allows interaction of M1 with importins for shuttling into the nucleus (*157*) (Fig. 4.2 4). Import of M1 into the nucleus

is required for subsequent M1-NEP mediated nuclear export of vRNPs (Fig. 4.2 5). Once back in the cytosol, the association of vRNPs and M1 prevents re-import of newly formed vRNPs into the nucleus. How M1 interacts with RNPs concretely however, remains unclear. Two alternative models where M1 either interacts with vRNPs via the NLS (*158*) or via its CTD (*150*) have been suggested.

M1 in virus assembly and virus morphology

In current models of influenza assembly, M1 is proposed to be the key mediator, based on its interactions with all other virus components and membranes. It is clear that M1 is an important factor for the determination of virus morphology and as such also virus stability. Using VLPs is was confirmed that regularly shaped particles in the typical size range of IAV particles only form if M1 is present (Fig. 4.2) (*45*, *46*). Others have identified specific mutations of the M1 protein which induced a switch in morphology from predominantly spherical particles to filaments as well as the other way around (*28*, *29*, *120*). Additionally it was shown that adaptions of IAV particles towards a filamenteous morphology when transferred to animal models after culturing viruses in eggs leads to the enrichment of mutations in M1 (*25*). All previous results clearly indicate, the critical role of M1 in assembly and morphology determination.

4.1.3 Motivation and aims of this study

M1 is the most abundant protein inside the IAV particles and important for many essential functions inside the virus and during the virus life cycle. The lack of a complete M1 structure, incomplete knowledge about potential structural plasticity and missing insights into the arrangement and dynamics of M1 in particular during assembly prevent the understanding of molecular details of the different functions and forms of M1. Structure determination of M1 in the past has turned out to be challenging and several attempts of classical structural biology methods have failed to describe the full-length structure of M1 leading to the assumption that the M1 CTD is partially unstructured or unfolded.

Within virions, M1 is suggested to form a protein layer underneath the membrane as observed in cryoEM and cryoET images. Insights into the in situ structural arrangement and relative orientations of M1 monomers underneath the viral envelope remain unknown. However, understanding the arrangement of M1 inside viruses is crucial to understand how M1 contributes to virus particle morphology and virus stability. Due the pleomorphy of IAV particles, studying the structure of IAV and concomitantly the matrix layer inside virions is not accessible to x-ray crystallography or cryoEM single particle processing methods. Instead, cryoET combined with subtomogram averaging

has been shown in the past to successfully provide clarity of protein arrangement inside viruses of heterogeneous morphology (*85*, *86*). In addition, subtomogram averaging has proven to be able to provide in situ structural information which is needed for M1 in order to improve our limited knowledge of the structure itself.

Here I aim at studying the structure and arrangement of M1 in filamentous particles of the pathologically relevant HK68 strain and the corresponding HK68 VLPs. Due to the small size of the protein and potential flexibility of the protein, optimization and adjustments of the subtomogram averaging workflow are needed. In addition, high resolution and good quality tomographic data are required. Based on limited structural characterizations of M1 and the absence of a full-length model, achieving medium to high resolution will be advantageous to allow interpretations of the structure even in regions which have not been structurally described before. I therefore aimed at establishing an optimized subtomogram averaging workflow to provide medium resolution in situ information of M1. Based on information about position and orientation produced during subtomogram averaging, I further aimed at describing in detail the arrangement of M1 within viruses and VLPs. These information were then to be used to further develop current models of M1 driven virus assembly.

4.2 Results

4.2.1 In situ structure of M1 from non-purified HK68 VLP filaments

To study M1 by subtomogram averaging I initially assessed cryoET data of HK68 virus and VLP particles described in chapter 3 (see table 2.1) for the presence of the matrix layer underneath the membrane. In tomograms of non-purified, unfixed HK68 VLPs I found particular strong and clearly visible protein density underneath the viral membrane for all tomograms with a defocus value closer to focus than -4.5 µm, which resulted in a subset of 21 out of 60 tomograms (Fig. 4.3B).

To assess if the protein signal underneath the viral membrane was strong enough to allow obtaining in situ structural information for the small IAV M1 protein, HK68 VLP tomograms were processed by subtomogram averaging focusing on the matrix density underneath the membrane. Starting positions and orientations of M1 subtomograms were initialized by sampling positions along the surface of a cylinder which was calculated based on a manually determined spline and manually determined M1-M1 diameter for each VLP filament.

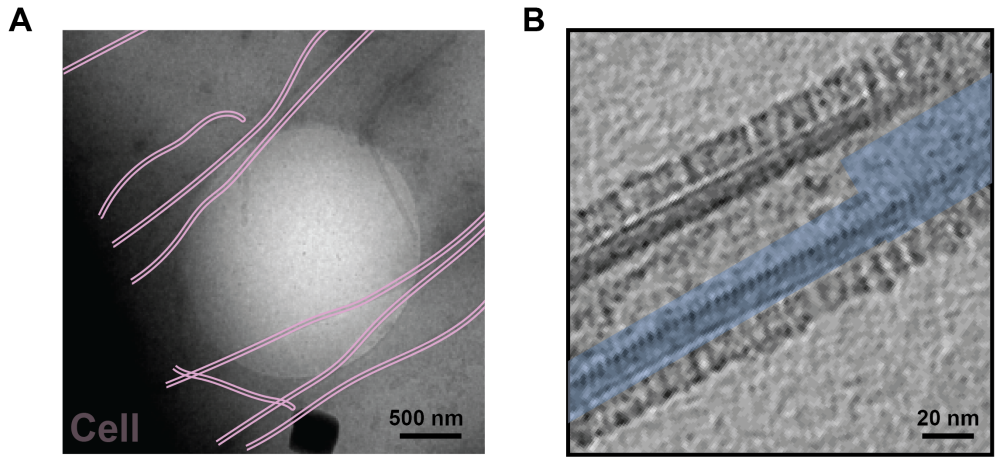

Figure 4.3 Tomograms of non-purified IAV VLP particles show a strong matrix layer density underneath the membrane. A) Representative medium magnification image. Cell edge in black in the lower and left corner. VLP filaments extending from the ell edge (outlined in rose). B) Slice through a VLP tomogram. To enhance the M1 protein signal for visualization 10 slices were summed. Regions where the matrix layer is visible are shaded in blue.

Initially, subtomograms from a subset of 20 tomograms with a box size of 46x46x46 nm were extracted at oversample positions from two times binned (Bin4) tomograms. Subtomograms were averaged and subsequently aligned to the average of each tube individually. Each tube dataset contained between 28,000 and 35,000 oversampled M1 subtomograms. The first round of alignment was focused on adjusting the Z-positions of each subtomogram utilizing the strong signal from the double layer membrane. The resulting average showed three smooth layers corresponding to the membrane bilayer plus the matrix layer (Fig. 4.4A). After another iteration of alignment focusing specifically on the matrix layer, the smooth protein density underneath the membrane evolved into several parallel strands, almost orthogonal to the tube direction with a regular spacing of 38 Å (Fig. 4.4B). Another five iterations with the same alignment parameters improved the specific signal of the strands further (Fig. 4.4C). When comparing the averages from different filaments we found that the orientation of the protein strands relative to the tube orientation was different (Fig. 4.4G). Consequently, I continued to align subtomograms from each tube separately after re-extracting subtomograms from Bin2 data. To resolve features along the strands, careful optimization of subsequent alignment parameters in particular the alignment mask and filter boundaries were required. Using a alignment mask which adopts the curvature of the tube and a band pass filter at 15 Å to 35 Å to suppress the dominating signal from the inter-strand distance was essential to improve the structure further. After three iterations using the optimized alignment parameters individual subunits along the strands with a spacing

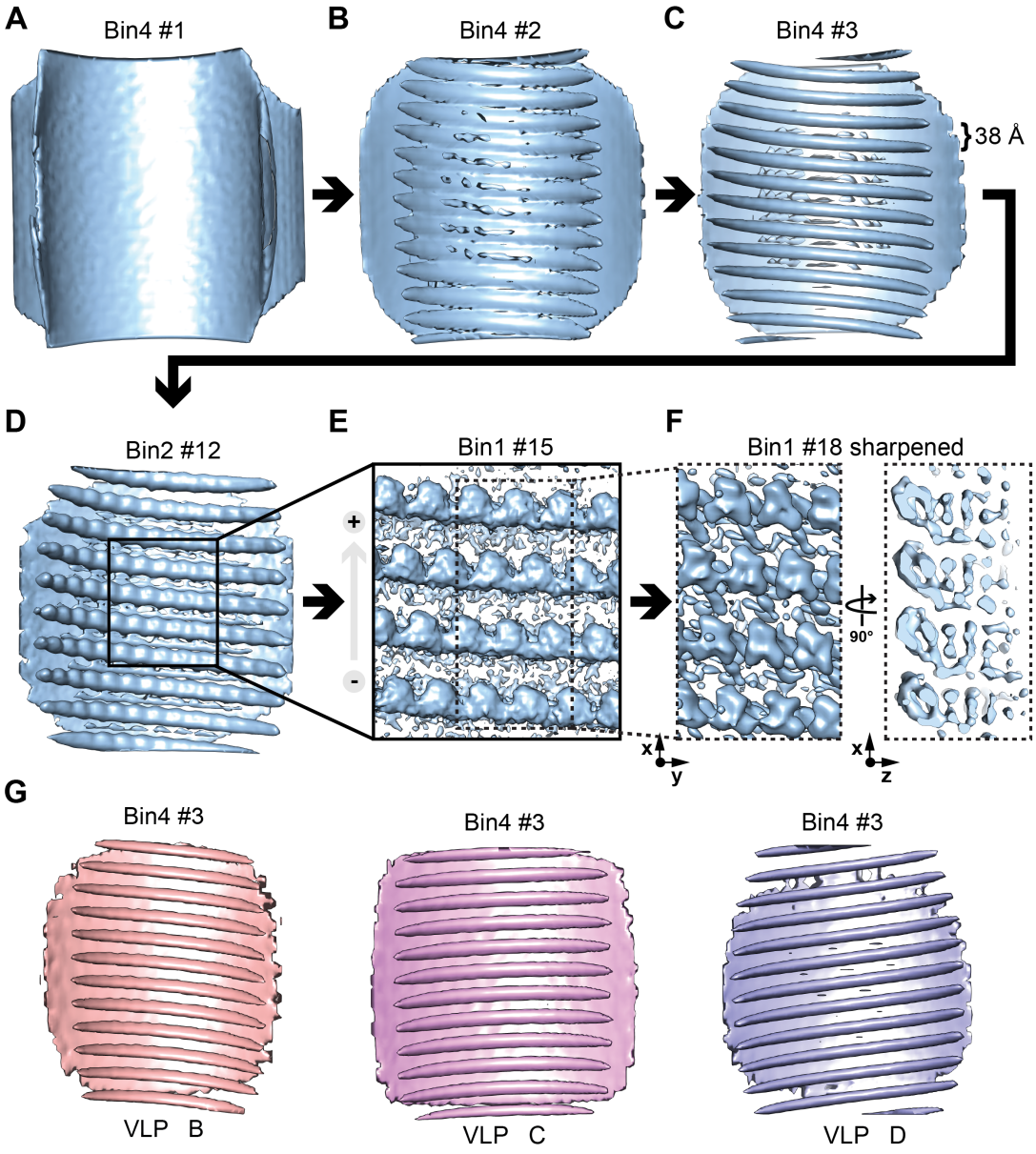

Figure 4.4 M1 subtomogram averaging reference over multiple iterations of alignment. A-D) Averaged subtomograms of the matrix layer from a single VLP filament (VLP A) at different stages of alignment. Iteration of alignment as indicated on top. Averages are shown from the center of the filament facing the inner tube layer. The orientation corresponds to the view onto the XY plane. E) Reference after iteration 15 with polarity of the M1 polymer from - to + as indicated. E,F) Averages generated after combining subtomograms from four different VLP filaments with similar M1 structures. F) The structure was sharpened using a B-factor of -800 (see section 2.6.2). G) M1 averages from different VLP filaments.

of 28 Å were resolved (Fig. 4.4D). Different settings for all employed filtering and masking steps, such as incorporation of a high pass filter were tested to identify optimal settings for alignment.

After two additional iterations, subtomograms from tubes with averages where the filament orientation relative to the tube direction were sufficiently similar and where subunits had become clearly visible (a subset of 4 out of 20 tomograms) were combined for subsequent steps of alignment (Fig. 4.4E). By now sufficient features to recognize that resolved protein strands were polarized became apparent which was not visible from previous low er resolution averages (Fig. 4.4F). Whenever necessary, the polarity was adjusted, prior to combining subtomograms. To removed oversamples positions, subtomograms which were closer together than a threshold distance of 14 Å, half the subunit-subunit distance, were removed. This cleaning step reduced the number of subtomograms to 8,000 per filament.

The dataset was split into two half sets, top and bottom of each tube containing 16,000 asymmetric units from four filaments per half set. Alignment was continued independently for both half sets. During the final stages of alignment utilizing unbinned data, the low pass filter was reduced to 8 Å. An additional cleaning step after one iteration reduced the number of positions to 13,000 subtomograms per half set. The resolution of the resulting average was determined by calculating the FSC between the two half sets as 8 Å at this alignment stage (Fig. 4.4F).

In most subtomogram averaging packages including the workflow used here, the similarity between subtomograms and the reference is evaluated by calculating the constraint cross correlation (CCC) function between the two volumes (see section 1.2.7). Prior to calculating the cross correlation function as well as during the calculation of the final average, the range of spatial frequencies which is affected by the missing wedge is masked out, typically by multiplying the Fourier transform of each subtomogram and the reference volume with a binary wedge mask (Fig. 4.5B). Here we replaced the binary wedge mask with the amplitude spectrum of each tomogram (implemented by Dustin Morado, Briggs group). Besides information about the orientation and width of the missing wedge, the amplitude spectra provides information about the 'weight' of each frequency in the tomogram (Fig. 4.5D). Amplitude spectra were calculated from 100 randomly selected noise positions in each tomogram. When using a binary wedge mask, the M1 reconstruction appeared to be distorted into the Z direction. This effect was reduced in the M1 reconstruction calculated using amplitude spectra. This improvement of the structure facilitated subsequent interpretations (Fig. 4.5A,C).

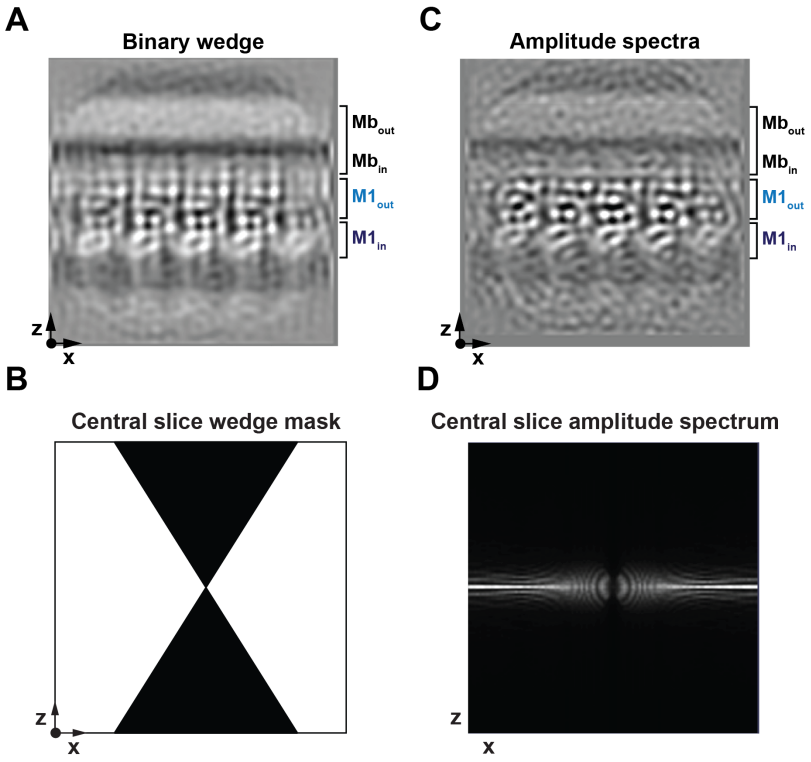

Figure 4.5 Comparison of two different wedge weighting strategies employed during subtomogram averaging of M1. A) Orthoslice through the M1 reference obtained when using a binary wedge mask during subtomogram averaging and alignment. B) Orthoslice through a binary wedge volume displayed in Fourier space used for wedge masking. C) Same as A but after using amplitude spectra during alignment and averaging. D) Orthoslice through amplitude spectrum (Fourier space) calculated form a representative tomogram of the dataset.

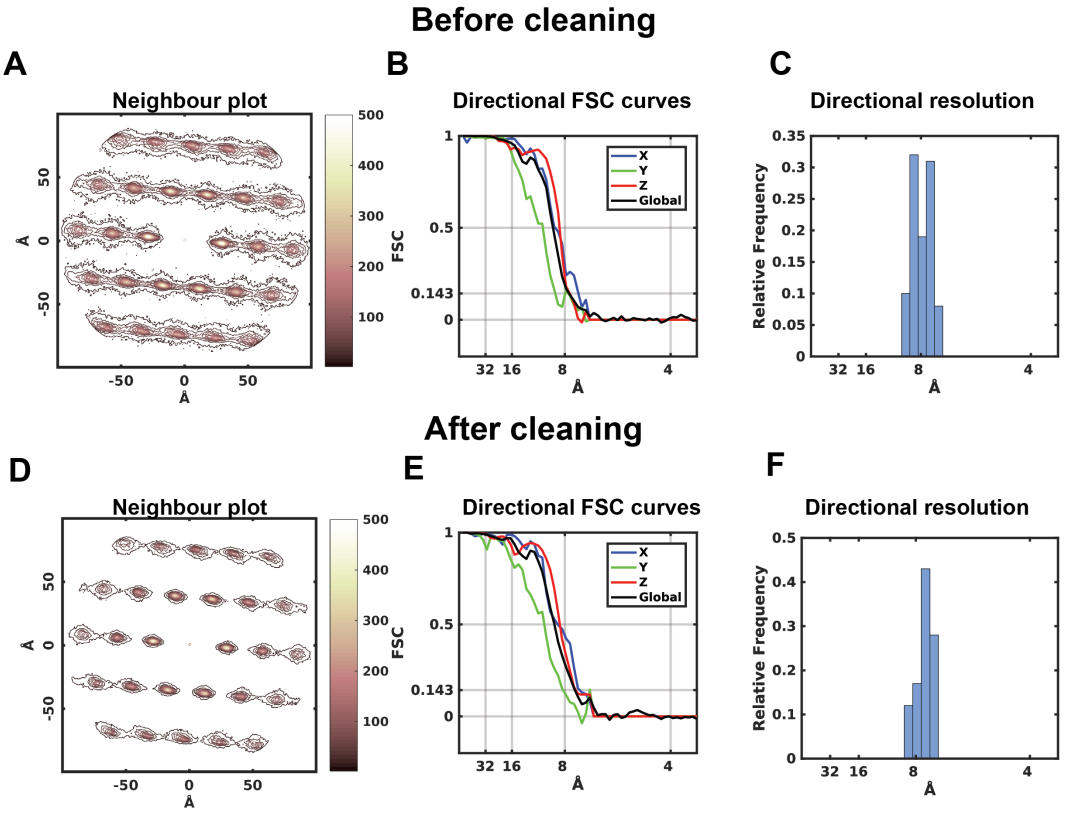

Figure 4.6 Geometric cleaning of subtomogram positions reduces the resolution anisotropic of M1 A-C) Neighbour plot and results from directional FSC calculations (*75*) before geometric cleaning. B) Calculated FSC in the three main directions and the global FSC for the structure before geometric cleaning. C) Histogram of the resolution values determined at the 0.143 level for 100 different directions. D-F) Same as in A-C but for data after geometric cleaning.

Despite optimized computation of the final average by the means described above, we found that features were still not well separated along the direction of the polymer (roughly corresponding to the Y direction indicated in figure 4.4). I suspected that this effect results from difficulties to align individual subunits correctly, due the small size, or from flexibility in the arrangement of the protein layer. I assessed the anisotropic resolution of the final structure using the 3DFSC program developed by Tan et al. (*75*). The 3DFSC tool performs directional resolution measurements by determining the FSC for individual cones rather than for the complete sphere. By comparing the resolutions for the three different main directions I found that the resolution was indeed worse in the Y-direction relative to all other directions as expected from the visualization of the reconstruction (Fig. 4.6B). The distribution of 0.143 FSC measurements for 100 different directions further indicated that the resolution varied between 7 Å to 10 Å for the different directions (Fig. 4.6C).

To improve the resolution in the direction along the strands (Y-direction) I removed positions which diverged from the average arrangement by geometric cleaning. As part of the geometric cleaning procedure I first assessed the positions of all neighbouring subtomograms for each M1 position and plotted the neighbour positions on top of each other in a neighbour plot (see section 3.2.11). By evaluating the neighbour plot for M1 I found a regular pattern of positions. However, rather than being defined, circular spots, neighbour position appeared as elliptically clustered spots (Fig. 4.6A). Clusters were more extended in the Y-direction than in the X-direction which correlates with the anisotropy determined directly from the structure. This observed range of relative distances between neighbouring M1 subunits reflects either flexibility in the arrangement or inaccuracy in the performed alignment with respect to the Y-direction. Independent of the source of this variability, specifically removing these positions in a geometric cleaning step could potentially increase the anisotropy of resolution.

To perform geometric cleaning, subsets of positions were selected for which at least three neighbour positions were located within the range of a defined sphere around the cluster centre. Initially, different variants of this cleaning step were tested with regards to the number and position of fixed neighbour positions. Best results were obtained when requiring three neighbours with no fixed position resulting in reduced resolution anisotropy while retaining 50 % of subtomograms. When calculating the neighbour plot for the cleaned set of positions a reduction in the width of the point cloud became clearly visible (Fig. 4.6D). In correspondence to the observations from the neighbour plot, the average calculated from the cleaned subset of subtomograms showed an improved resolution in the Y-direction and better separation of distinct density elements along the direction of the M1 polymer strands was observed (Fig. 4.6E,F).

Further improvement of the structure was obtained from specifically masking selected areas of the reference during alignment. If displayed as a XZ orthoslice, the obtained structure appears as a bi-lobe density underneath the membrane, where one lobe of density is located directly underneath the membrane and one lobe of density is located further towards the centre of the tube (Fig. 4.7A). The two lobes were masked individually during the last rounds of alignment and each region improved. The resolution of the final average combined from the inner and outer focused alignment and after geometric cleaning ranged between 9 Å to 7 Å for different directions (Fig. 4.7B). The final map was sharpened using an empirically determined B-factor of -800 and filtered according to the global resolution (see section 2.6.2).

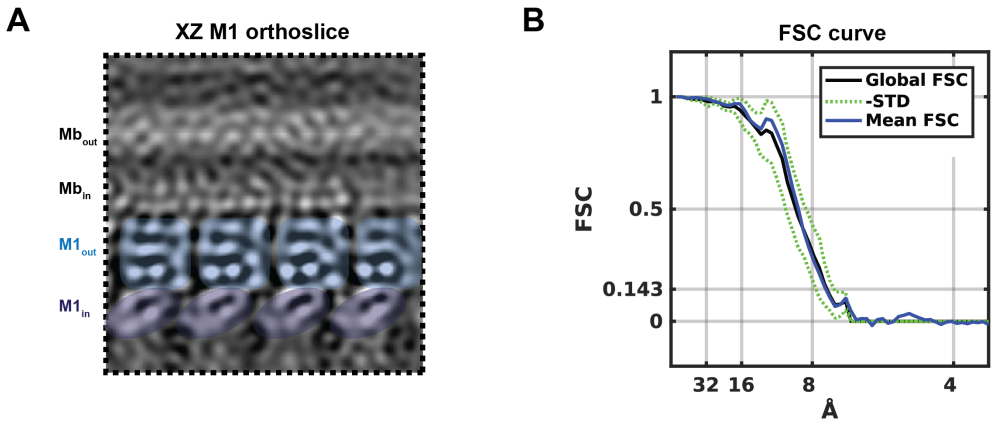

Figure 4.7 Final subtomogram averaging reconstruction of M1 obtained from HK68 VLPs. A) XZ orthoslice through the M1 subtomogram averaging reconstruction obtained from HK68 VLPs. The inner and the outer lobe visible in the orthoslice are marked in light blue and dark blue respectively B) FSC curve for the structure shown in A. The mean FSC and the standard deviation around the mean based on resolution measurements in different directions are shown.

4.2.2 Existing M1 NTD models fit the obtained M1 density map

As expected for the obtained final resolution, alpha helices were resolved in our structure which allows to fit high resolution structural models into the density. While no models exist that describe the complete influenza M1 structure, the structure of the M1 NTD [residue 1-165] has been determined before by X-ray crystallography from different strains and at different pH values (pH=4,PR8,PDB:1AA7; pH=7, PR8,PDB:1EA4; pH=4,WSN,PDB:2Z16) (*142–144*) (see section 4.1). While the relative arrangement of monomers in the asymmetric unit is different for structures determined at different pH all models share a highly similar model of the NTD monomer (see section 4.1). Here I initially selected the only M1 NTD crystal structure that was obtained at neutral pH (pH=7, PR8, PDB:1EA4 (*143*)) for fitting into the obtained M1 in situ structure.

The M1 NTD monomer was fitted as a rigid-body into the membrane proximal part of the M1 density map using the fit in map tool in UCSF Chimera (*105*). All nine alpha helices of the M1 NTD monomer matched nine densities in the membrane proximal lobe of the M1 in situ structure obtained from HK68 VLPs (Fig. 4.8A). When I further evaluated the performed fit of the model into the EM density map, I found that the two cylindrical densities corresponding to alpha helix 6 and helix 5, which are located at the surface facing the membrane, appeared to be moved closer towards the membrane relative to the positions of the helices in the crystal structure (Fig. 4.8C).

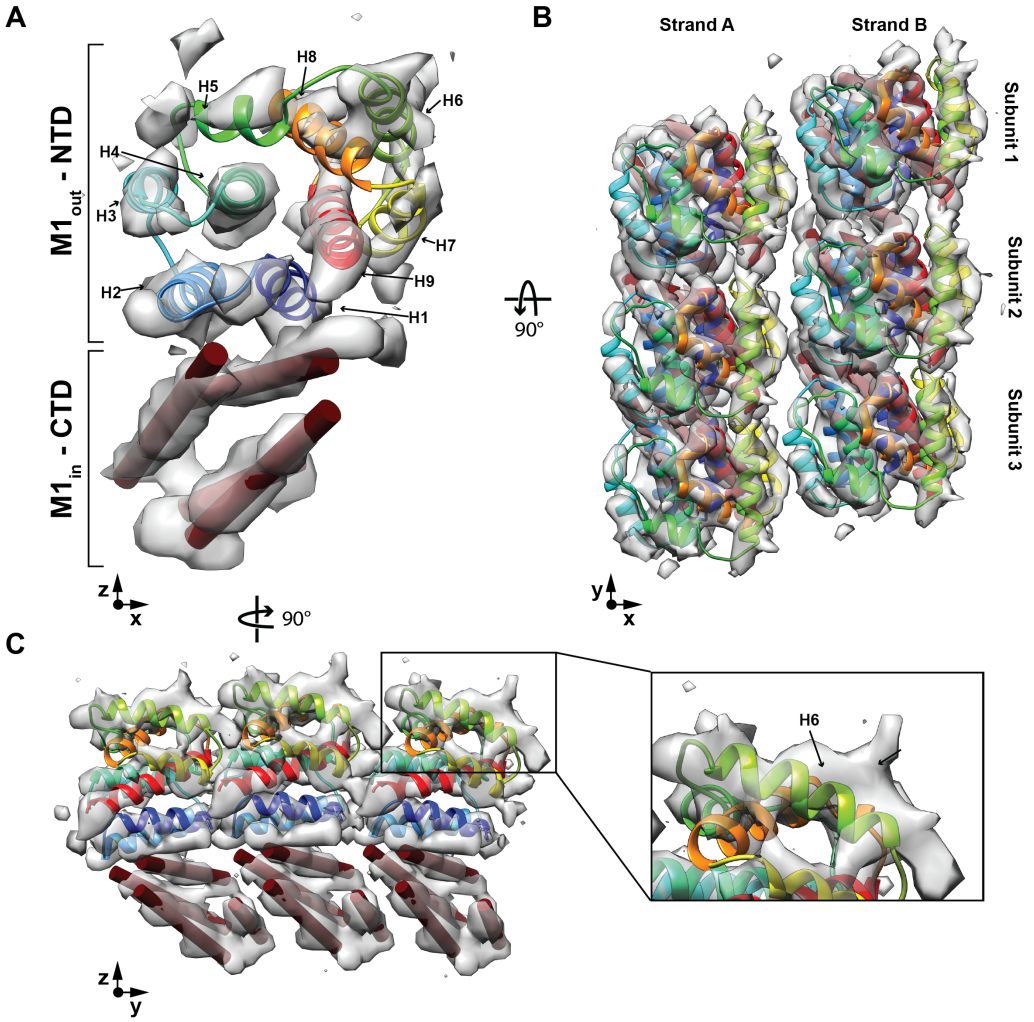

Figure 4.8 Comparison between the final M1 reconstruction from subtomogram averaging and the crystal structure of the M1 NTD. A) Subsection of the final, sharpened M1 reconstruction from HK68 VLPs overlaid to a single M1 NTD monomer fitted as a rigid body into the outer lobe of the structure, $M1_{out}$. Cylindrical densities in the inner lobe of the structure ($M1_{in}$) were marked by displaying red cylinders inside the density. B) A subsection of the final M1 structure overlaid to six NTD monomers fitted as rigid bodies. The membrane interface points to the front C) A subset of the average with three NTD monomers fitted and the CTD densities filled with red cylinders. The inset focuses on the density which located directly underneath the membrane and which corresponds to helix six of the crystal structure.

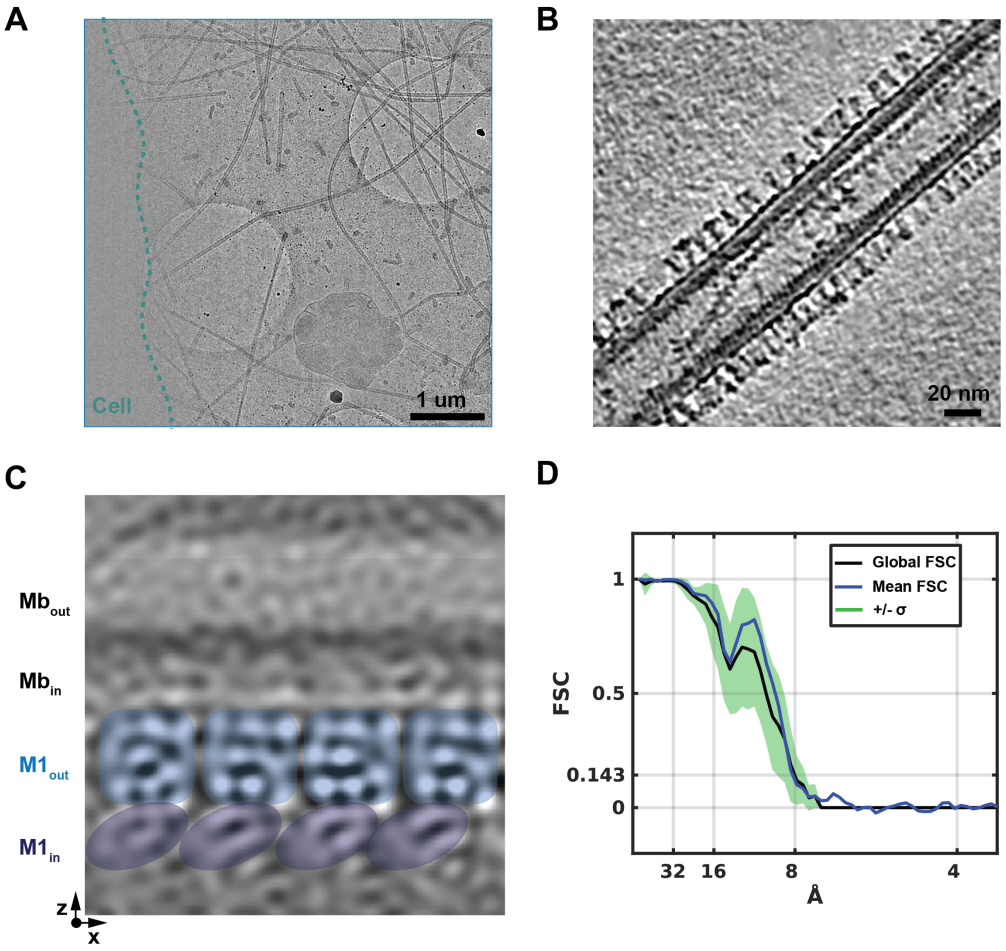

Figure 4.9 In situ structure of IAV M1 obtained by subtomogram averaging of M1 from HK68 tomograms. A) Representative cryoEM images at medium magnification showing the edge of a cell surrounded by HK68 virus filaments. B) Sum of ten slices through a tomogram to further enhance the M1 protein density underneath the membrane for visualization purposes. C) Final sharpened in situ structure of M1. D) Mean FSC curve with standard deviation from directional FSC measurements.

4.2.3 In situ structure of M1 from non-purified HK68 virus

To confirm that the M1 structure obtained from VLP filaments corresponds to the M1 structure found inside HK68 virus particles I repeated the processing workflow described for VLPs using tomographic data of HK68 virus presented in chapter 3 (see also table 2.1). In 16 out of 40 acquired HK68 virus tomograms a protein density underneath the membrane was clearly visible (Fig. 4.9B). Subtomogram averaging of M1 from this subset of tomograms was performed just as described for the data from HK68 VLPs. Subtomograms were initially aligned and averaged independently for each virus filament. Reconstructions of M1 from early stages of alignment appeared very similar to initial averages obtained from VLP tomograms. Parallel strands of protein density lining the inside of the membrane bilayer were visible. The direction of strands relative to the virus tube orientation was variable across the dataset and alignment was continued to be performed for each filament independently. However, during further steps of alignment the distinction into subunits along the filaments (Y-direction) was much weaker in M1 from HK68 virus data than M1 from HK68 VLP data and did remain very weak throughout all following steps. Subtomograms from four filaments were combined at a later stage of alignment, based on the similarity of the references from these four virus filaments. During subsequent steps of alignment the structure evolved further. In the XY orthoslice through the final, sharpened reconstruction (B-factor of -800, see section 2.6.2) an inner and on outer lobe were clearly visible just as observed in the VLP M1 structure (Fig. 4.9C compare to Fig. 4.7). However, the inner lobe region of the M1 average structure obtained from HK68 virus particles remained less defined than the corresponding region in the M1 average from HK68 VLP particles (Fig. 4.10). I tested if the inferior appearance of the structure was related to an anisotropic resolution by determining the directional FSC (*75*). From directional FSC measurements I found a larger standard deviation around the mean FSC curve across different directions for the M1 reconstruction obtained from HK68 virus relative to the M1 structure obtained from HK68 VLPs (Fig. 4.9D). From directional FSC measurements I identified that the obtained resolution in the Y-direction was much lower than in the X and the Z-direction (Fig. 4.10D). In the coordinate system of the reconstruction, the Y-direction corresponds to the direction along protein strands. As for the VLP data I attempted to reduce resolution anisotropy through geometric cleaning of subtomogram positions based on the results from calculated neighbour plots (see section 4.2.1). However, geometric cleaning of M1 subtomograms from HK68 virus resulted only in minor improvements not sufficient to recover structural details at the level previously achieved for VLP particles.

Except the lower resolution in the Y-direction all other directions were resolved

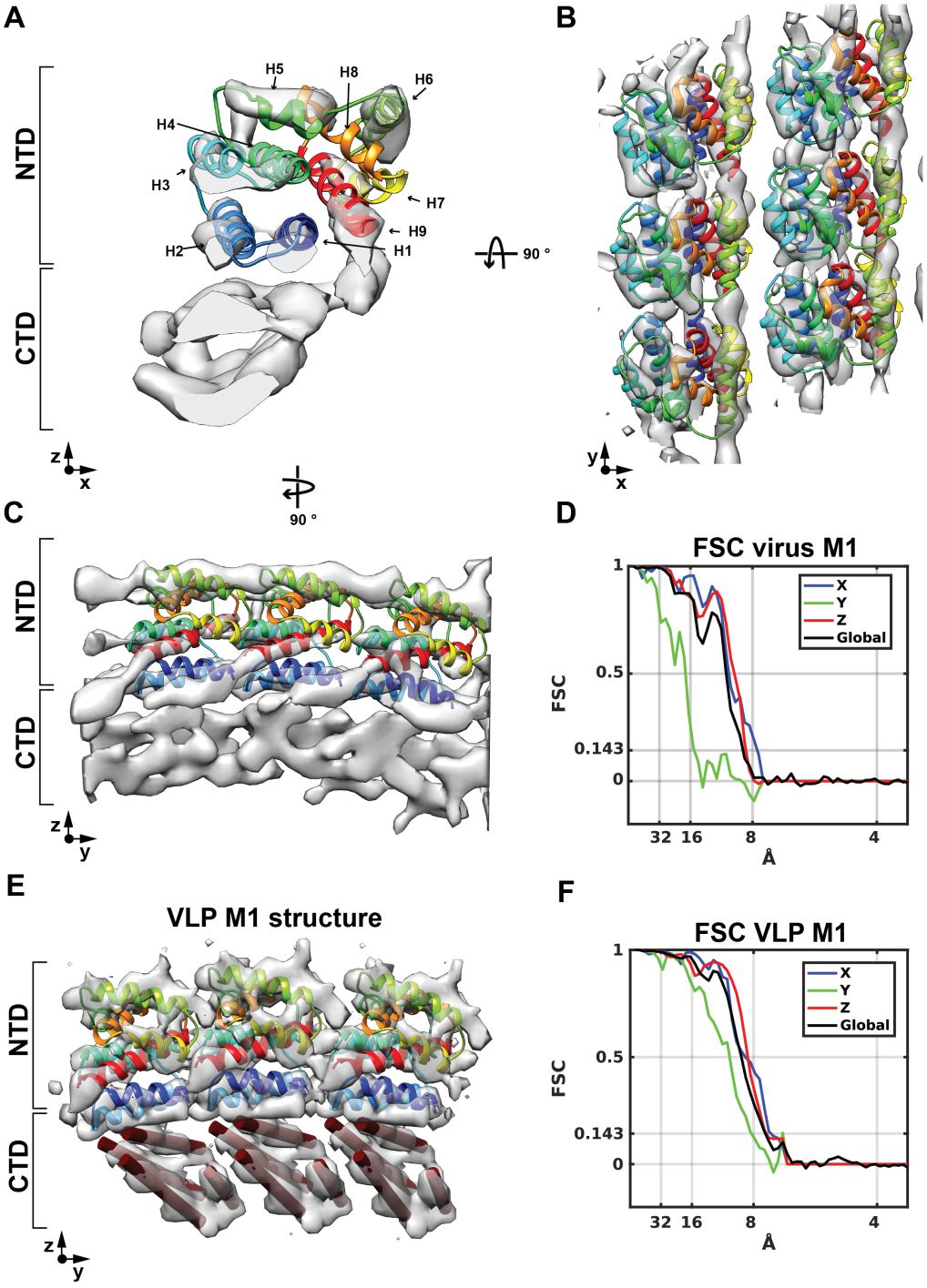

Figure 4.10 Evaluation of the M1 reconstructions from HK68 virus and HK68 VLPs. A) Single subunit of the M1 reconstruction from HK68 virus displayed together with a M1 NTD crystal structure monomer fitted into the volume. B) Six M1 NTD monomers fitted into a subsection of the final M1 reconstruction from HK68 virus. The side facing the membrane is turned towards the front. C) Subsection of the final M1 reconstruction from HK68 virus overlaid to three M1 NTD monomers fitted into the volume. D)'Directional FSC curve of the structure shown in A-C. E) Final M1 reconstruction from HK68 VLPs and crystal structure of the M1 NTD fitted into the volume. F) Directional FSC curves for the structure shown in E. PDB:1EA3.

to 8 Å. Despite the measured anisotropy, the obtained resolution was sufficient to be able to fit NTD crystal structure monomers as rigid bodies into the outer lobe of the structure just as described for the M1 structure obtained from HK68 VLPs (Fig. 4.10A-D). Helices H1-H6 and H9 fit well into the obtained density. The respective densities matching H7 and H8 are however weaker and appear to be less well resolved in the virus structure in comparison to the VLP structure (Fig. 4.10A). Despite the presence of weaker densities in comparison to the VLP M1 structure, the fit of the NTD crystal structure into the outer lobe of the M1 reconstruction from HK68 virus clearly resembles the fit of the NTD crystal structure into the outer lobe of the M1 reconstruction from HK68 VLPs. The comparison the inner lobe regions from both reconstruction is impeded by the lack of a high resolution crystal structure of the CTD to which each reconstruction could be compared to individually. Overall, the comparison of the two M1 structures obtained from HK68 VLPs and HK68 viruses to existing crystal structures suggests that the NTD regions of both in situ structures are identical. The lower overall resolution as well as the higher resolution anisotropy of the virus structure is likely to explain the observable differences in the inner lobe region of the structure.

Influence of imaging conditions on reconstruction quality

To identify reasons for the difference in quality between the reconstruction of M1 from HK68 viruses and HK68 VLPs, I visualized the contrast transfer function (CTF) to characterise the signal transfer for the defocus range of the subsets of tomograms used for the respective M1 reconstructions. The subset of tomograms of HK68 VLPs that were processed for M1 had been collected using a defocus range of -2 µm to -4 µm, while the subset of tomograms of HK68 virus which were processed for M1 were collected at a defocus range from -4 µm to -6 µm. The CTF for four different defocus values ranging from -2 µm to -6 µm was therefore calculated in Matlab (MathWorks) as described in equation 1.1 without using an envelope function (see section 1.2.4).

The simulated CTFs demonstrate that for defocus values <-4 µm, as used for recording the HK68 virus dataset, the first zero crossing of the CTF occurs at spatial frequencies corresponding to distances of 28 Å or larger (Fig. 4.11). The main features for initial alignment of M1 are the distance between M1 strands, 38 Å, and the distance between M1 subunits along strands, 28 Å, which however requires additional high pass filtering to suppress the strong signal from the M1 inter-strand distance. Consequently,

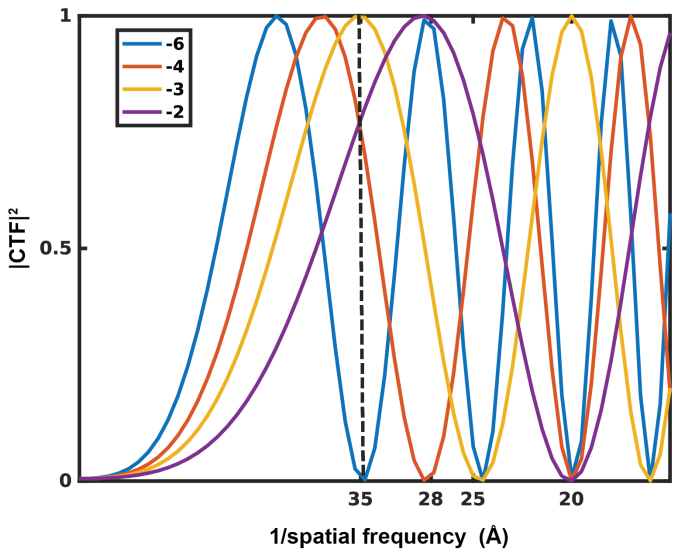

Figure 4.11 Simulation of the CTF for a range of defocus values from tomograms used for M1 subtomogram averaging. The CTF was calculated according to the properties of the employed microscope (Titan Krios) and specific imaging conditions (see section 1.2.4). The fraction of amplitude contrast was set to 0.07, the spherical abberation constant of the objective lens of the microscope is 2.7 mm and the electron wavelength at 300 keV was determined as 1.97 pm. The legend indicates the defocus value for each curve in µm. The dashed black line indicates the limit of the high pass filter used during subtomogram averaging alignment of M1.

alignment features with features smaller than the 35 Å threshold of the high pass filter, are crucial to drive any further alignments and ultimately to resolve individual M1 subunits. Therefore, in data collected with a defocus <-4 µm the signal corresponding to the first peak of the CTF is mostly cut out by the employed high pass filter. Instead, at a defocus of -2 µm a larger fraction of the first CTF peak signal extends beyond the limit of the employed high pass filter (Fig. 4.11. Additionally, for data collected with a defocus at -2 µm the first peak of the CTF occurs at 28 Å which corresponds to the distances between two M1 subunits (Fig. 4.11. Consequently, it appears that the range of defocus values <-4 µm in the data of HK68 viruses (which were initially collected for processing of the data for HA described in chapter 3) is not ideal for subtomogram averaging alignment of M1 and might therefore explain the differences in quality of the M1 reconstruction. In the future, these information can be taken into account when choosing parameters for acquiring new datasets for M1 subtomogram averaging.

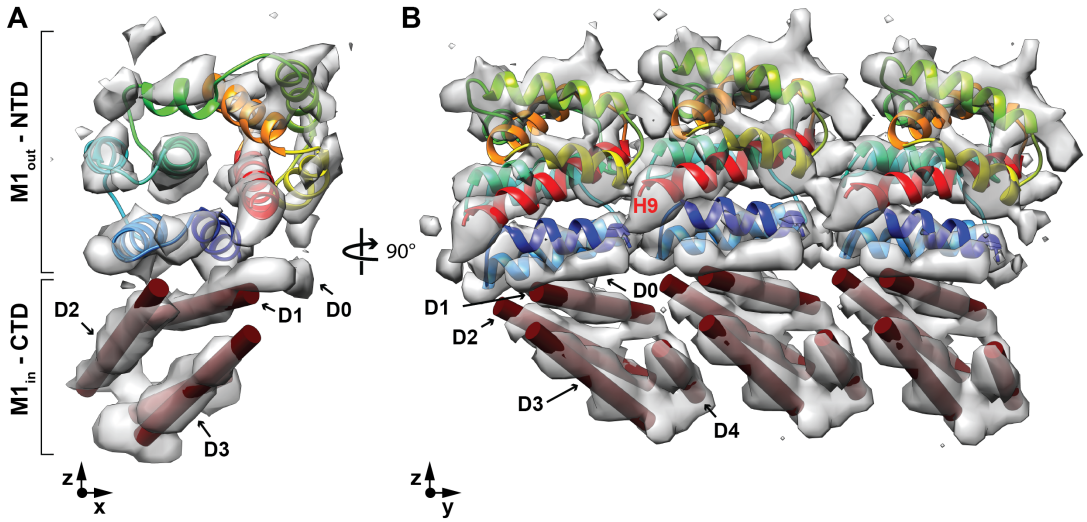

Figure 4.12 Description of the inner lobe of the M1 reconstruction from HK68 VLPs. A) Subsection of final HK68 VLP M1 structure which fits a single M1 NTD crystal structure. B) Subsection of the final HK68 VLP M1 structure which fits three M1 NTD monomers. Red cylinders are displayed into cylindrical densities of the inner lobe of the M1 structure which remains unoccupied from the fit. The densities were named D0-D4.

4.2.4 Structure of the IAV M1 CTD

Due to the better quality of the M1 reconstruction from HK68 VLP samples in comparison to the M1 reconstruction from HK68 viruses, further interpretation regarding the inner lobe of the structure were performed using the data from HK68 VLPs (see section 4.2.1).

In the the inner lobe of the M1 structure obtained for HK68 VLPs defined, cylindrical densities were visible, which remained unoccupied after fitting the M1 NTD crystal structure into the M1 in situ structure. I therefore attempted to further interpret the obtained in situ structure in this region. We identified four tightly packed cylindrical densities which I expect to correspond to the M1 CTD. For better visualization I placed red cylinders into each density and numbered the four densities D0-D4 (Fig. 4.12). D2 and D3 were measured to be about 15 and 8 residues in lengths, respectively. Densities D1 and D3 appeared less regular in diameter. The length of these densities were therefore estimated to be between 18-24 (D1) and 12-18 (D3) residues. In addition to the four central cylindrical densities, I observed un-occupied density extending from the last helix (H9) of the M1 NTD and connecting to D1 in the CTD region which I termed D0 (Fig. 4.12). Two out of the five densities (D3 to D4) are connected by visible density. The connectivities between densities D0-D2 and the connectivity to the M1 NTD regions are less defined and different options are in agreement with the

constraints of the obtained structure. I identified three different scenarios describing the connectivity of the density elements to each other and to the NTD. In one scenario one NTD is connected to the CTD located under the neighbouring M1 NTD monomer and I will therefore term this scenario the 'trans' connectivity (Fig. 4.13A). In the other two scenarios, the NTD is connected to the CTD located directly underneath and consequently we term these two connectivities 'CisA' and 'CisB'.

In the trans connectivity model, one NTD from the outer layer of the structure is connected to the CTD (inner layer of the structure) that is stacked under the adjacent NTD. In this scenario D1 would be connected by D0 to H9 of the NTD (Fig. 4.13A). In the Cis A connectivity model one NTD is connected to the CTD located directly underneath itself. The connectivity of the other elements is the same as described for the trans connectivity model. Here D1 would connect via a short loop to H9 of the NTD crossing D0. D0 therefore would be mostly empty (Fig. 4.13B). In the Cis B connectivity model just like in the Cis A model, the NTD is connected to a CTD which is located directly underneath. In contrast to the Cis A model, the NTD connects to D2. Consequently the connectivity for the other elements is different than in the first two cases. Essentially, orientation and connectivity of D1 and D2 would be inverted, while connections between D3 and D4 are the same. In this scenario, D0 would be only partially filled by turning loop connecting D1 and D3 (Fig. 4.13C).

4.2.5 Secondary structure predictions of the IAV M1 CTD

To test if any of the above described models could be supported by properties of the known sequence of the HK68 M1 CTD, I performed secondary structure prediction (SSP) of the HK68 M1 CTD sequence. I first compared the results from three different SSP tools PsiPred (*159*), Porter5 (*160*) and Lomets (*161*) and generated a consensus prediction based on the predictions and the confidence values for each residue (Fig. 4.14). The consensus prediction indicates evidence for four helical stretches, H1-H4 along the sequence. Overall, the majority of the sequence was predicted to be helical. The first helical stretch (H1) in the prediction is very short and only comprises less than ten residues. The subsequent three helical stretches (H2-H4) are longer measuring about 20 residues each. To compare results from secondary structure predictions more carefully to the identified densities, I listed the measured lengths of all density elements according to the three different connectivity scenarios (Fig. 4.14). If no connecting density was present between densities I measured the distance between subsequent density elements and estimated the number of residues. I then overlaid the identified structural elements along the sequence for a subsequent comparison to the structure prediction.

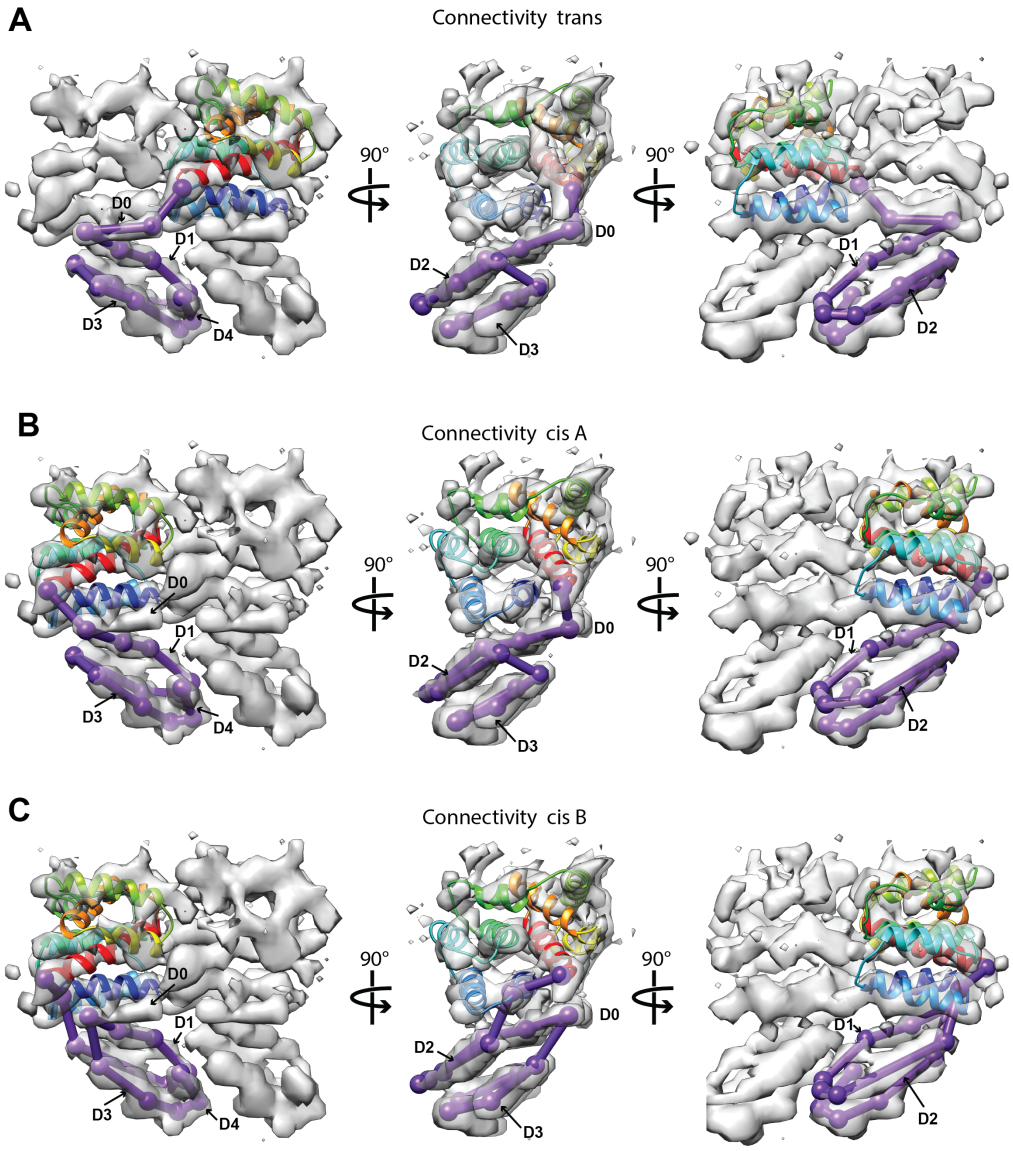

Figure 4.13 Three models of possible connectivities for the densities found in the inner lobe of the M1 structure from HK68 VLPs. A) The NTD and CTD connect in a 'trans' configuration via the unoccupied densities D0, D1, D2, D3, D4 in this order. B) The NTD and CTD connect in a cis configuration. The connectivity within the densities otherwise are the same as shown in A. The NTD connects to density D1 followed by D2, D3 and D4 C) The NTD and CTD connect in a cis configuration as shown in B. In contrast to B the NTD connects to the density D2 which connects to D1,D3 and D4.

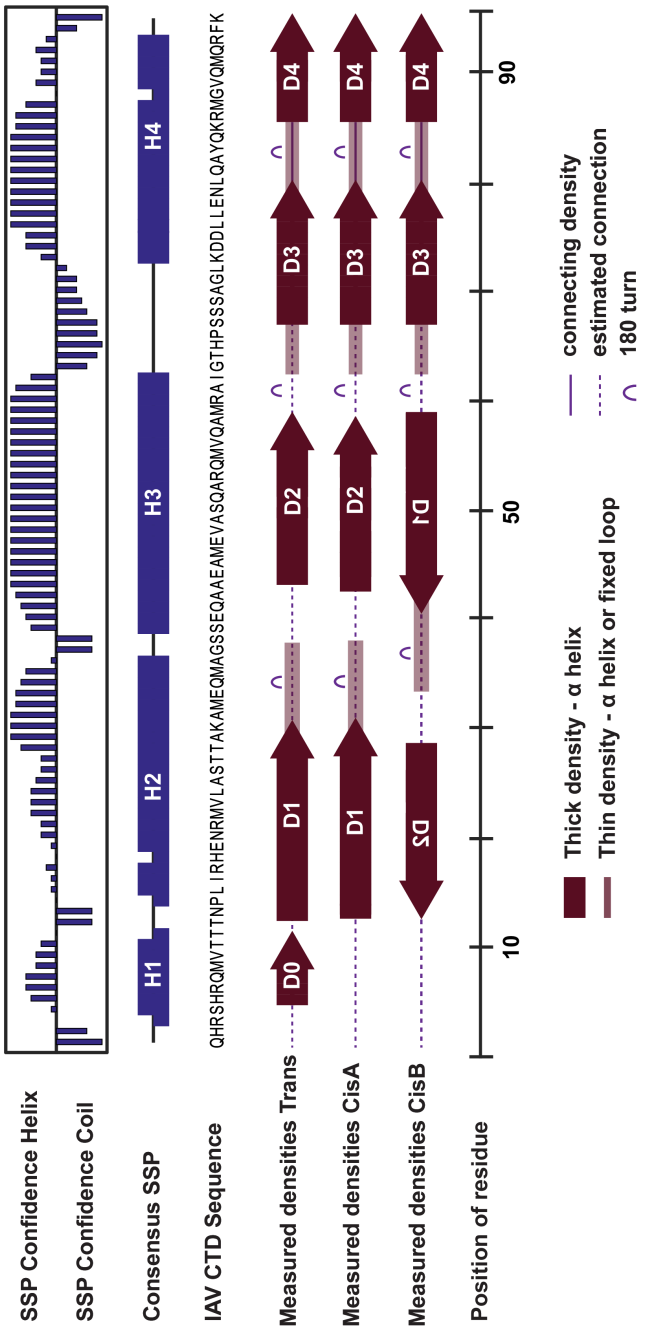

Figure 4.14 Comparison between secondary structure prediction of the M1 CTD and densities from the inner lobe of the subtomogram averaging reconstruction of M1. Predicted secondary structure and corresponding confidence for helix and coil assignments based on predictions from three different tools (see text) are outlined in comparison to density elements in the in situ M1 structure for different possible connectivity models at estimated positions along the sequence.

The prediction as well as the characterisation of density elements directly from the subtomogram reconstruction suggests that the M1 CTD is predominantly helical and void of beta sheets. Further, the predictions suggest an arrangement of four alpha helices connected by short loops. This matches the observed four to five tightly packed cylindrical densities in the inner lobe of the M1 reconstruction from cryoET data. However, upon comparing the exact positions and lengths of the predicted secondary structure elements to the interpretation of structure from the subtomogram reconstruction of M1, several differences can be observed. Starting from the beginning of the sequence, I found low confidence prediction for a short helical element within the first ten residues. This prediction would match the identification of a short helix matching D0 involved in the connection between the NTD and CTD as observed for the trans connectivity model (Fig. 4.14). In the region between residue 10 and 60, all employed algorithms predicted two long helical elements, H2 and H3. However, in particular predicted helix 2 provides mostly low confidence, while predicted helix 3 is assigned with high confidence. In between the two predicted helices 2 and 3, a short coil region is predicted. In all connectivity models based on the density elements in the M1 reconstruction from subtomogram averaging, I found two density elements located between residue 10 and 60, D1 and D2. In addition, in all connectivity scenarios these two densities have to be connected by a 180° degree turn. Although the order of D1 and D2 is reversed in the scenario CisB relative to the other two scenarios, the prediction in this region cannot distinguish between any of the three connectivity models. Based on the M1 reconstruction from subtomogram averaging, I suggested two connected cylindrical density elements termed D3 and D4 to be located at the end of the sequence for all connectivity scenarios. In particular the last density D4 is shorter than the others. D3 and D4 are connected by density representing a connecting turn. Comparing densities D3 and D4 to the sequence based prediction at the C-terminus of the sequence, there is weak theoretical evidence for one longer predicted helical element H4. H3 and H4 in the prediction are connected by a longer stretch of coil with lower confidence. This coil region could correspond to the beginning of the density element D3, for which only thinner density was found and which therefore could represent either a fixed loop or a helix. In all connectivity models, D3 has to be connected by a turning loop to the previous density element (either D2 and D1) which is roughly located in the region where a coil region is predicted between H3 and H4.

In summary, a detailed comparison between the prediction and the identified density elements appears to be challenging due to the inaccuracies of the density measurements and low confidence regions from SSP. Nevertheless, the SSP results

support the interpretation of finding mostly alpha helical elements which roughly match the overall length of the M1 CTD sequence. Thereby the SSP results provide a first link between the HK68 M1 CTD sequence and the inner lobe of the M1 subtomogram reconstruction.

4.2.6 Comparison of M1 to homologue structures

Due to the absence of IAV M1 CTD high resolution models I compared the obtained M1 in situ structure to an available homologue matrix structure from infectious salmon anemia virus (ISAV). The M1 structure from this distant relative of influenza virus is the only full-length structural model of a related M1 protein which includes the CTD (Fig. 4.15A) (*162*). However, the protein sequence identity between ISAV M1 and IAV M1 is only 10 % and no degree of homology can be detected from pure sequence alignment (Fig. 4.15B). In addition, the ISAV protein sequence is overall 25% shorter than the IAV M1 sequence. From structure and sequence comparisons it occurs that the missing 25% of the protein sequence are not located in one particular region of the structure but are distributed along the complete length. Despite this high divergence on the sequence level, the ISAV M1 NTD and the IAV M1 NTD share structural similarities (Fig. 4.15C). Relative positions of alpha helices are conserved in most parts between the two structures, however some of the corresponding helices and loops are slightly twisted relative to each other. Consistent with the ISAV sequence being 25% shorter than the IAV sequence, several of the ISAV NTD helices are shorter, as well (Fig. 4.15C) (*162*). Even for more closely related virus strains such as influenza C virus, no clear sequence homology can be detected from performed sequence alignments while the structures show a remarkably high degree of similarity (Fig. 4.15).

To gain further support for any of the connectivity models suggested for the IAV M1 CTD and based on the structural similarities in the NTD regions between ISAV M1 and IAV M1 I attempted to compare the CTD of ISAV M1 to the four densities identified in the inner lobe of the M1 in situ structure from HK68 VLPs. The ISAV M1 CTD consists of four alpha helices measuring H1=11, H2=19, H3=16 and H4=11 residues. The CTD is connected by a loop to its NTD, which however is located on top of the neighbouring CTD similar to the trans connectivity scenario proposed for M1 reconstruction from subtomogram averaging (Fig. 4.16A). Side-by-side visualization of the ISAV and my density map shows that the relative packing of the four cylindrical densities D1-D4 appears to be similar to the arrangement of the four helices of the ISAV CTD (Fig. 4.16A,B). I consequently tried to fit the ISAV M1 CTD model as a rigid body directly into the CTD region of my structure. The three best fits for which

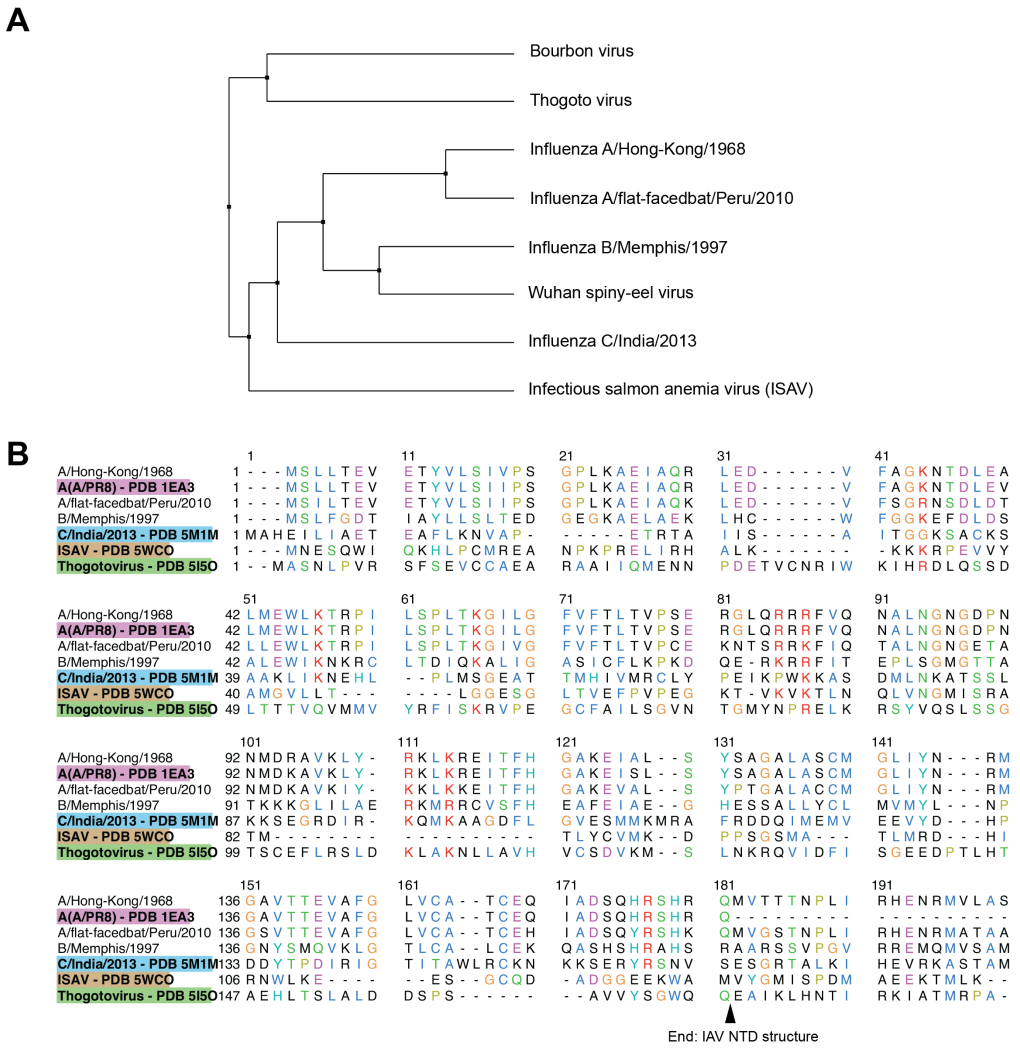

Figure 4.15 Homology of M1 across different orthomyxovirus strains. A) Phylogenetic tree of a small subset of orthomyxovirus strains based on multiple sequence alignment using ClustalW. B) Sequence alignment for the NTD regions of a subset of the strains shown in A. Coloured strain names indicate that a structure of the M1 NTD is available. Confidence of sequence alignment decreases significantly for any strain more distantly related than influenza B virus. C) Overlay of structural models of the M1 NTD from different strains performed in Chimera. The influenza A virus M1 NTD monomer is shown in red and compared to influenza C virus (PBD:5M1M) in blue, ISAV (PDB:5WCO) in beige and Thogoto virus (PBD:5I5D) in green. Despite the failed sequence alignments, structures for these three strains show similarities between each other.

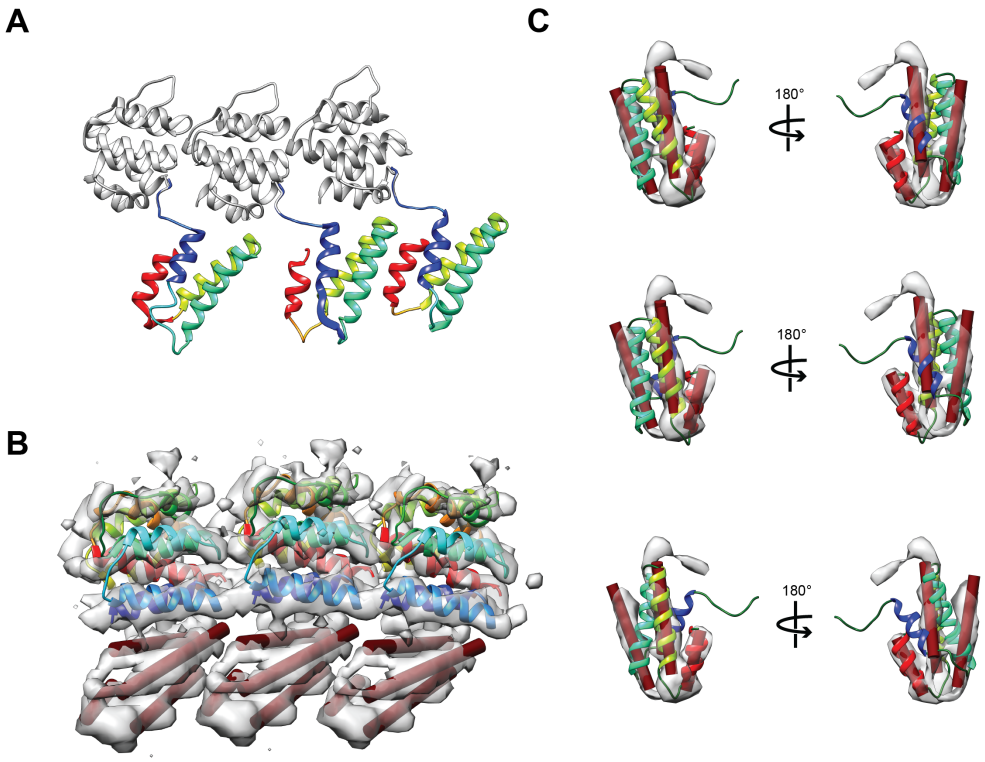

Figure 4.16 Comparison between the ISAV M1 crystal structure and the M1 reconstruction obtained by subtomogram averaging. A) Crystal structure of three subunits of the ISAV M1 (PDB:5WCO). The M1 CTD is coloured in rainbow colours. From NTD to CTD H1 blue, H2 cyan, H3 light green, H4 red. B) Three subunits of M1 reconstruction from IAV with three NTD subunits (PDB:1EA3) fitted into the density. The densities in the inner lobe are marked by red cylinders. C) The three best reasonable fits of the ISAV M1 CTD model into the inner lobe of the IAV M1 reconstruction. The four densities are marked by four red cylinders.

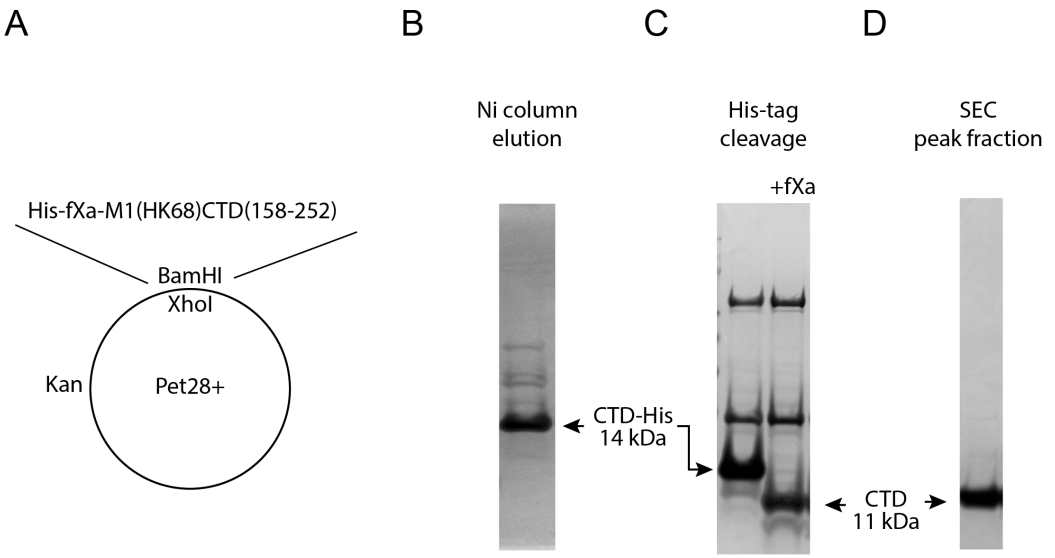

Figure 4.17 Expression and purification of the M1 CTD fragment. A) Schematic of the used expression construct. B) SDS PAGE of the elute from a Ni-column, the first step of purification after expression of the construct shown in A C) SDS PAGE showing the the pre-purified protein before and after 1 h incubation with the His-tag cleavage enzyme factorXa (fXa). D) SDS PAGE of the peak fraction obtained by size exclusion chromatography (SEC) after tag cleavage.

the beginning of the CTD was found to be in a realistic position relative to the NTD are displayed in figure 4.16C. A comparison of the three fitting results demonstrates that all three fits are very similar and that the overall orientation of the ISAV model relative to the densities of my M1 structure is the same in all three fitting results (Fig. 4.16C). In all cases, ISAV H4 matches density D4, ISAV H3 is placed into density D3, ISAV H2 is placed into density D2 and ISAV H1 is placed into density D1. Mainly the orientation of each ISAV helix within the density elements varies slightly for each fit. However, in none of the three fits, the four helices from the ISAV model match the maps perfectly since relative orientations of the four helices are not identical. These dissimilarities in the helix orientation corresponds to the degree of structural divergence expected from the comparison of the NTD (Fig. 4.15).

4.2.7 Attempts to crystallise the M1-CTD

In previous attempts to crystallise full-length M1, the CTD was reported to be cleaved during the purification and crystallisation process (see section 4.1). The sequence between amino acid 158 and 164 was reported to be especially vulnerable to cleavage by proteases (*142*, *143*). Based on the obtained subtomogram averaging structure and comparison to results from SSP as well as to the homologue M1 CTD structure

from ISAV, I proposed that the IAV M1 CTD is formed from four helices that form a bundle and the core of the M1 CTD. This proposed fold suggested that the protein fragment could be stable if expressed as an isolated fragment. A high resolution model of the CTD would dramatically improve the interpretation of the M1 reconstruction from subtomogram averaging. I therefore attempted to prepare the CTD fragment for crystallisation. I generated two different plasmid variants of the M1 CTD region. One construct was designed to include the full CTD while the second construct was designed to exclude the first part of the sequence, M1(172-252)CTDdH0. According to the trans connectivity model, the missing region from this construct would correspond to a short helix connecting the M1 NTD and the CTD and therefore would not be part of the core fold of the structure. Constructs were further modified by the addition of a N-terminal His-tag via a factorXa protease cleavage side synthesized and cloned into pET-28a(+) vectors. The designed constructs were purchased from GenScript Biotech.

With the help from Simon Erlendsson and Xiaoli Xiong, I established a protocol to express and purify the CTD from the obtained plasmids in *E.coli*. Expression worked efficiently and the expressed protein fragment was purified from lysed cells via its His-tag via a Nickel (Ni) column which resulted in effective isolation of the protein fragment from the cell lysate (Fig. 4.17B). Subsequently, we optimized removal of the His-tag via factorXa cleavage. A maximum incubation time of 1 h at room temperature was critical to prevent complete degradation of the protein (Fig. 4.17C). The cleaved tag as well as remaining impurities of the protein preparation were subsequently removed by size-exclusion chromatography (SEC) (Fig. 4.17D). Pooled fractions were analysed by circular dichroism to assess the distribution of secondary structure elements. CD spectra suggested a large fraction of the protein to be helical, which confirms our previous interpretation from the M1 reconstruction from subtomogram averaging as well as previous results from others (*143*). Upon concentrating the protein to 20 $\mathrm{mg\,mL^{-1}}$ the protein solution up was prepared for crystallization under multiple different conditions. Crystals were observed for two different conditions, 24 h after setting up crystal plates (pH 4, 0.8 M Ammonium sulfate, 0.1 M citric acid and condition E7 in the Morpheus HT-96 MD1-47 screen) and were send to the national structural biology facility Diamond (UK). Using standard settings for acquiring diffraction pattern with help from the X-ray facility at the MRC-LMB, crystals were exposed to X-rays but no diffraction pattern was detected.

In parallel to crystallisation tests, the construct was expressed for subsequent analysis by NMR by Simon Erlendsson. Obtained spectra confirmed the presence of secondary structure but no tertiary structure was visible, suggesting that the monomeric

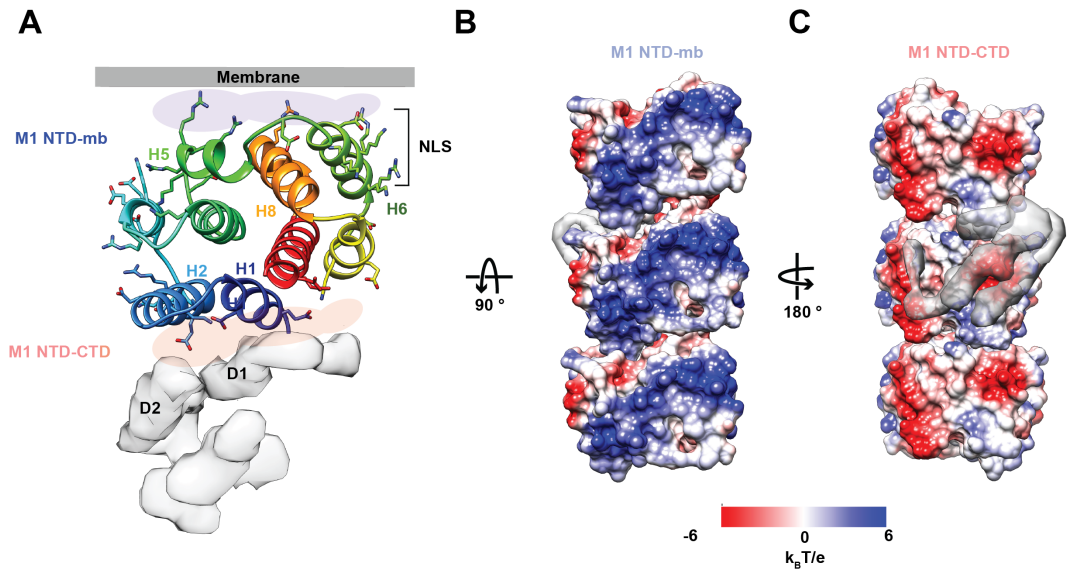

Figure 4.18 Characterisation of M1-membrane and M1 NTD-CTD interfaces. A) Single M1 NTD crystal structure in side view after fitting into the M1 subtomogram reconstruction of which only the inner lobe corresponding to the CTD is shown. All charged residues of the NTD model are displayed. The interfaces between the M1 NTD and the membrane and the NTD and the CTD are indicated in blue and red respectively. B) Charge distribution along the membrane-facing surface of three neighbouring M1 NTDs displayed after fitting the NTD models into the M1 subtomogram averaging reconstruction. C) Same as B but displayed from the opposite side of the NTD, which according to the fit into the subtomogram averaging reconstruction faces the M1 CTD. EM density which I expect to correspond to the CTD of a single M1 subunit is displayed in the middle to demonstrate the relative positioning of the NTD and CTD but to prevent visual distraction from the displayed surface charge. PBD:1EA3, Mb=membrane.

M1 CTD fragment appears as a ribbon of alpha helical elements. In summary, the attempts for high resolution structure determination of the M1 CTD by X-ray crystallography and NMR were not successful to date and no structure was obtained by these means.

4.2.8 The membrane interface of the M1 NTD

An important aspect regarding the structure-function relationship of M1 is how M1 interacts with membranes and to date mechanistic understanding of this process is missing. By fitting high resolution models of the M1 NTD into the M1 reconstruction from subtomogram averaging, I can describe the M1-membrane interface as it occurs inside IAV viruses. From the fit of the NTD crystal structure, I found that the M1 surface facing the membrane comprises helices 5,6 and 8 and the loop between helix 4 and 5 of the M1 NTD (Fig. 4.18A,B). The CTD on the other hand is not located in

membrane proximity and Membrane-CTD interactions inside the virus can be excluded. A high proportion of residues in the membrane interface contain positively charged side chains. Presented data confirm that residues that are part of the NLS (101-105,H6) and other positively charged side chains that have been previously suggested to be relevant for interactions with the membrane such as R76, R78 (H5), R101, K104 (H6) and R134 (H8) are found on the surface facing the membrane (*163*). These residues create a wide, positively charged surface suited to bind negatively charged heads of phospholipids through electrostatic interactions (Fig. 4.18). In previous mutation studies, several of these positively charged side chains at the membrane interface had been identified to reduce or inhibit virus budding when replaced with non-charged residues such as K95A, K98A, K101A, K102A (reduced budding) and K104A (no budding) (all H6) (*120*).

4.2.9 M1-M1 interfaces

To describe the interfaces relevant for the M1 arrangement in viruses, I fitted multiple models of M1 NTD monomers as rigid bodies into the outer lobe of the obtained subtomogram reconstruction of M1. M1 subunits were found to arrange in parallel strands underneath the membrane and I identified three different M1-M1 interfaces along and across strands: Interfaces between the M1 NTD and the M1 CTD (fig. 4.18A,C), intra-strand M1-M1 interfaces (fig. 4.19A,B) and inter-strand M1-M1 interfaces (fig. 4.19A,D).

NTD-CTD interface

From the fit of the M1 NTD model into the M1 subtomogram reconstruction, I observed that the surface of the NTD opposing the membrane faces the inner lobe of the M1 in situ reconstruction which I expect to correspond to the M1 CTD. The interface comprises H1 and H2 of the M1 NTD and the cylindrical densities, D1 and D2 of the expected CTD region in the M1 subtomogram averaging reconstruction (Fig. 4.18A). I found that in contrast to the membrane interface, the NTD surface directed towards the CTD is widely negatively charged (Fig. 4.18C). Due to the absence of a high resolution model of the CTD I cannot analyse the charge distribution along the corresponding site of the CTD. From sequence characterisation of the M1 CTD I found that the IAV M1 CTD also contains a large fraction (18%) of charged residues. It is therefpre likely that the NTD and the CTD interact based on electrostatic interactions.

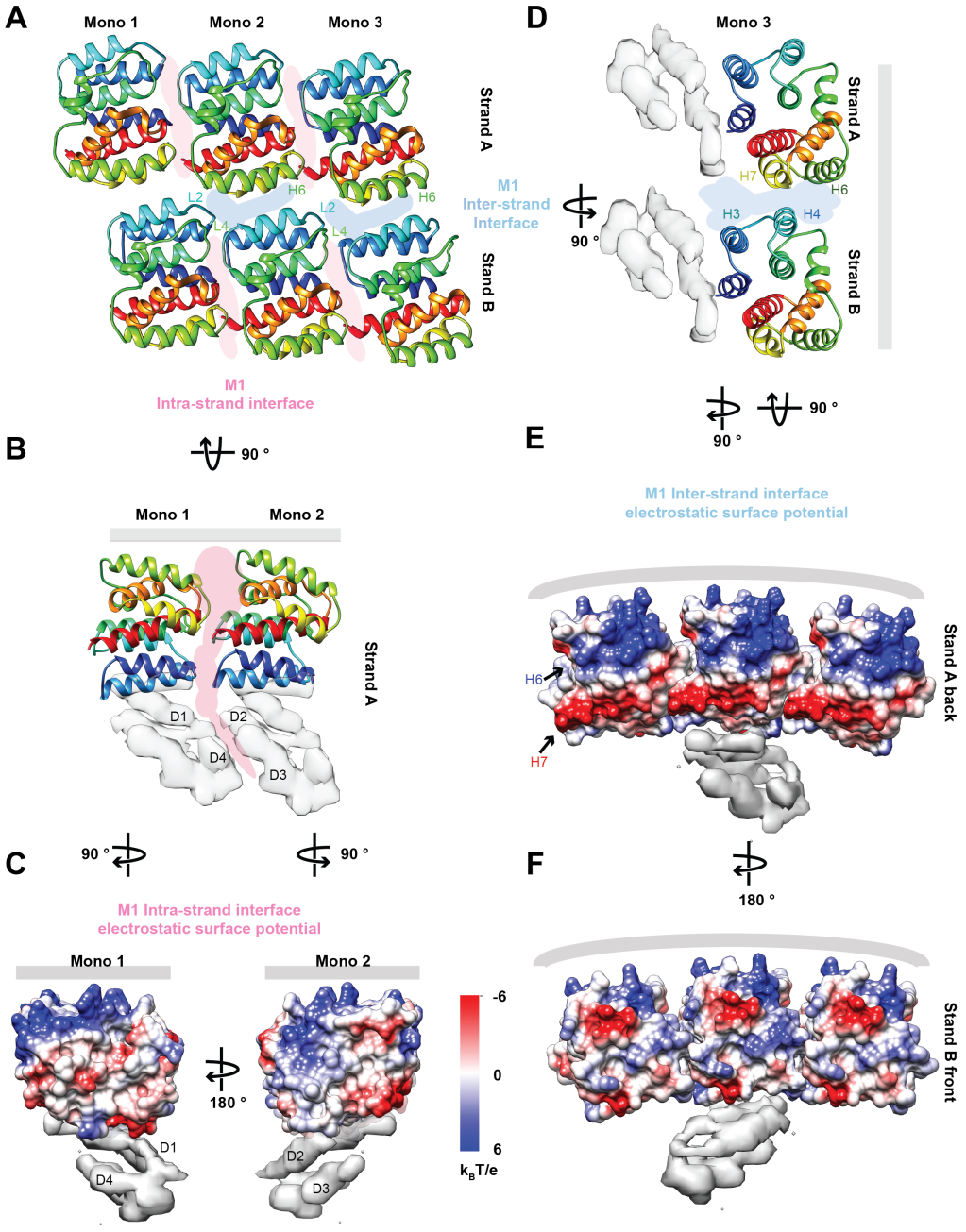

Figure 4.19 Characterisation of M1-M1 interfaces. A) Six M1 NTD monomers arranged according to the fit into the final M1 reconstruction. Intra-strand and inter-strand interfaces are marked in rose and light blue respectively. The membrane interface is displayed towards the front. B) Two neighbouring monomers from the same strand. The rose coloured field describes the M1-M1 intra-strand interface. The grey line in B-E represents the membrane. C) Electrostatic surface potential for the same two monomers shown in B with the interface for each monomer turned towards the front. D) Two opposing monomers from two neighbouring strands. The blue field highlights the inter-strand interface. Only two monomers are shown for clarity while a second monomer on strand B is thought to provide an additional interface as indicated in A. E) Electrostatic surface potential for three monomers from two neighbouring strands. The interface between the strands is turned towards the front. F) Same as E rotated by 180°.

M1 intra-strand interfaces

The M1-M1 intra-strand interface can be further divided into contributions from the NTD-NTD interaction and contributions from the CTD-CTD interaction. The intra-strand NTD-NTD interface comprises almost all of the turning loops that connect the nine alpha helices of the NTD. In detail, the beginning of H1 as well as loops connecting H2 to H3 (L2), H4 to H5 (L4), H6 to H7 (L6) and H8 to H9 (L8) on the interface of monomer 1 and H1 to H2 (L1), H3 to H4 (L3), the end of H9 and H3 to H4 (L3) on the interface of monomer 2 (Fig. 4.19B). The visualization of the charge distribution along the intra-strand interface indicates that no particularly strong charge differences are located on the two opposing surfaces (Fig. 4.19C). Instead an analysis of the potentially involved residues at the positions listed above suggests that the interface mostly involves hydrophobic interactions and hydrogen bonds. The interface between neighbouring CTDs include the cylindrical densities D2 and D3 on one side and the cylindrical densities D1 and D4 on the next monomer which form tilted, parallel planes in close proximity to each other (Fig. 4.19B). In the case of a trans-connectivity between the NTD and the CTD, the above-described NTD-CTD interactions, would additionally contribute to intra-strand M1-M1 interactions, while in the case of cis connectivities, the NTD-CTD interactions would be contributing to intra-molecular interactions.

M1 inter-strand interfaces

Based on the fit of multiple M1 NTD subunits into the described subtomogram averaging reconstruction of M1, the interfaces between M1 strands can be described in more detail. The inter-strand interface includes the surface of one NTD monomer on one strand (monomer 3, strand A) which is in proximity to two monomers on the adjacent strand (monomer 2 and monomer 3, strand B, Fig. 4.19A). We find that H6 and H7 of monomer 3 on strand A form a potential interface with the loop connecting H4 and H5 (L4) and the loop H2-H3 (L2) of monomer 2 on strand B (Fig. 4.19A) and H3 and H4 of monomer 3, strand B (Fig. 4.19D). I visualized the electrostatic potential of the inter-strand interface of the NTDs and found charged surfaces on the opposing sides between strands. The contribution of Helix 6 to this interface provides a patch of positive charge at the backside of one strand (strand A) while helix 7 provides a region of negative charge on the same side (Fig. 4.19E). On the front side of the subsequent strand (strand B) helix 3 contains three negatively charged residues which are responsible for the spot of concentrated negative charge on the front side (Fig. 4.19F). This distribution of opposing charges along the inter-strand interfaces suggests that the inter-strand interaction is driven by electrostatics.

4.2.10 Comparison of identified M1-M1 interfaces to crystal contacts in M1 NTD crystals

I compared the identified and described M1-M1 interfaces to the crystal contacts from previously published crystal structures. To facilitate the comparison, I first have summarized crystal contacts from low and neutral pH crystal structures, that have been previously suggested to be relevant for M1-M1 interactions (Fig. 4.20). M1 at neutral pH was shown to crystallize in the P1 space group. The corresponding asymmetric unit is represented by a dimer without symmetry as shown in figure 4.20C (*143*, *146*). Within the crystal, M1 monomers arrange in sheets and a cross section is shown in figure 4.20A. In contrast, in structures solved at low pH, the asymmetric unit is represented by a dimer, where monomers are related to each other by two fold symmetry (Fig. 4.20D). It was reported, that M1 at low pH appears as a dimer during purification. Based on this observation, M1 is assumed to also physiologically occur as a dimer at low pH (*142*). The low pH dimer arrangement is also referred to as the face-to-face arrangement (*144*). The low pH fact-to-face dimer interface comprises the surface, which I identified to be facing the membrane in the in situ M1 structure and is therefore very different to any M1-M1 interfaces I identified to be relevant in the virus (Fig. 4.20D,E and Fig. 4.21D). The face-to-face arrangement of M1 at low pH is typically compared to what is referred to as the face-to-back interface (Fig. 4.20B), a subset of the neutral pH M1 crystal contacts, which was previously discussed to be relevant for M1-M1 interactions within the virus as demonstrated in figure 4.20A (*145*). From my data, I did not find the face-to-back interface to be a relevant M1-M1 interface in situ. In other studies the interface between two low pH dimers was suggested as relevant for M1 polymerization (Fig. 4.20E,F), while the respective interface does not resemble any of the interfaces identified by us (*145*).

Instead I found, that the dimer interface of the asymmetric unit in the neutral pH M1 NTD crystal structure matches the M1-M1 intra-strand interface found in the virus. However, inside the virus, M1 strands are curved and therefore M1 monomers along strands interact at a slight angle whereas the arrangement in the crystal is planar (Fig. 4.21A). The observation that a very similar arrangement between two monomers occurs inside the virus with an angle but without an angle in the crystal suggests that the corresponding interface is flexible to adjust to different curvatures of the virus filament. I further identified a set of lateral interfaces between multiple asymmetric units within the neutral pH M1 crystal (Fig. 4.20A) which are similar but not identical to the inter-strand interfaces identified in our structure (Fig. 4.21A). In the crystal lattice, the lateral interaction between monomers appears to be restricted to the directly opposing monomer. In contrast, in the inter-strand interface described based on the

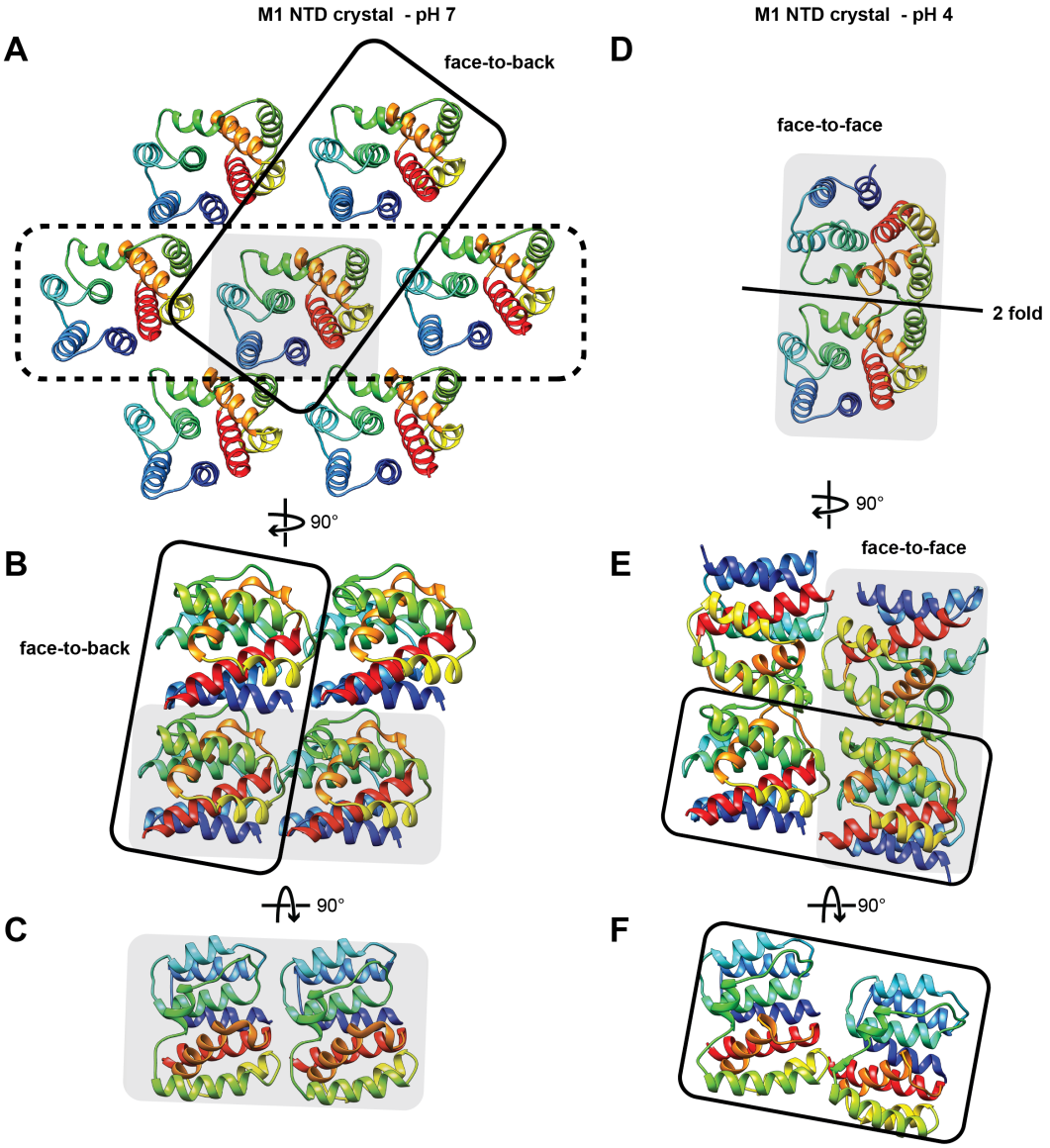

Figure 4.20 Summary of crystal contacts in M1-NTD crystals at neutral and low pH. A-C) Subsets of M1 monomers from a pH 7 M1 NTD crystal, PDB:1EA3, (*143*). Asymmetric units of crystals are shaded in grey. A) Cross section through crystal sheets of the M1 NTD crystal at neutral pH. The interface which is referred to as face-to-back interface is framed in black. The crystal sheet which includes the interfaces with the highest similarity to interfaces that I have identified in the virus is marked by a dashed frame. B) Four M1 NTD monomers from the crystal in A. Two asymmetric units are shown which share the face-to-back interface. C) Two neighbouring monomers, the asymmetric unit in the crystal represented in A. D-F) Subsets of the M1 NTD as found in a crystal of the M1 NTD at pH 4 (PDB:1AA7). D) Dimer with a two fold symmetry axis representing the asymmetric unit of the low pH M1 NTD crystal termed the face-to-face arrangement (*142*). E) Interface between two dimers in the low pH M1 NTD crystal. F) Rotated view of the two monomers framed in black shown in C.

M1 reconstruction presented in this study, one monomer interacts with two opposing monomers (Fig. 4.21B). Essentially, the two arrangements differ by a small shift (about a third of a monomer) of the three shown strands relative to each other. This observation suggests that the M1 inter-strand interface might be flexible and allows for relative shifts between M1 strands in the virus.

To summarize, two types of crystal contacts within the neutral pH M1 NTD crystal have been identified to be similar to M1-M1 interface in the virus and can therefore be interpreted in the context of M1 arrangement inside the virus.

4.2.11 M1 arrangement in IAV filaments

During subtomogram averaging the orientation and position of each M1 subtomogram are iteratively refined. I used this information from each subtomogram to display the obtained M1 reconstruction back into the tomogram to analyse the larger scale arrangement of M1 in VLPs and viruses.

M1 arrangement in HK68 VLPs

I found that in the context of IAV VLP filaments, M1 monomers are arranged in parallel strands with a helical propensity, densely lining the inner surface of the membrane (Fig. 4.22A). To understand how conserved this arrangement is between VLP filaments, I compared the arrangement of M1 across individual VLPs in the dataset. Most strikingly, I observed that the larger scale arrangement of M1 was conserved. The spacing of M1 subunits along strands was identified to be 28 Å and the spacing between strands was found to be 38 Å across all tubes analysed. However, in the majority of VLP filaments, more than one single strand was present. For different VLP tubes I counted between one and seven strands that wind in parallel along the inside of the membrane (Fig. 4.22). To confirm this finding I processed a second dataset collected on the same VLP sample by subtomogram averaging in the same way the first VLP dataset had been processed. For 22 filaments from this dataset, the number of helix strands could be assessed from the evaluation of the subtomogram positions after subtomogram alignment. By combining the numbers from these two data sets of the HK68 VLP sample, the total number of VLP filaments analysed with respective to their larger scale arrangement of M1 increased to 44. In 41 out of 44 filaments, the number of strands was found to not change along the 800 nm of the filament, captured in a single tomogram. Interestingly, in three filaments I observed changes in the number of helix strands within the imaged filament section. If I extrapolate this observation to the complete length of VLP filaments imaged in this dataset, this would correspond to a change rate of M1 strand numbers per 5.6 µm

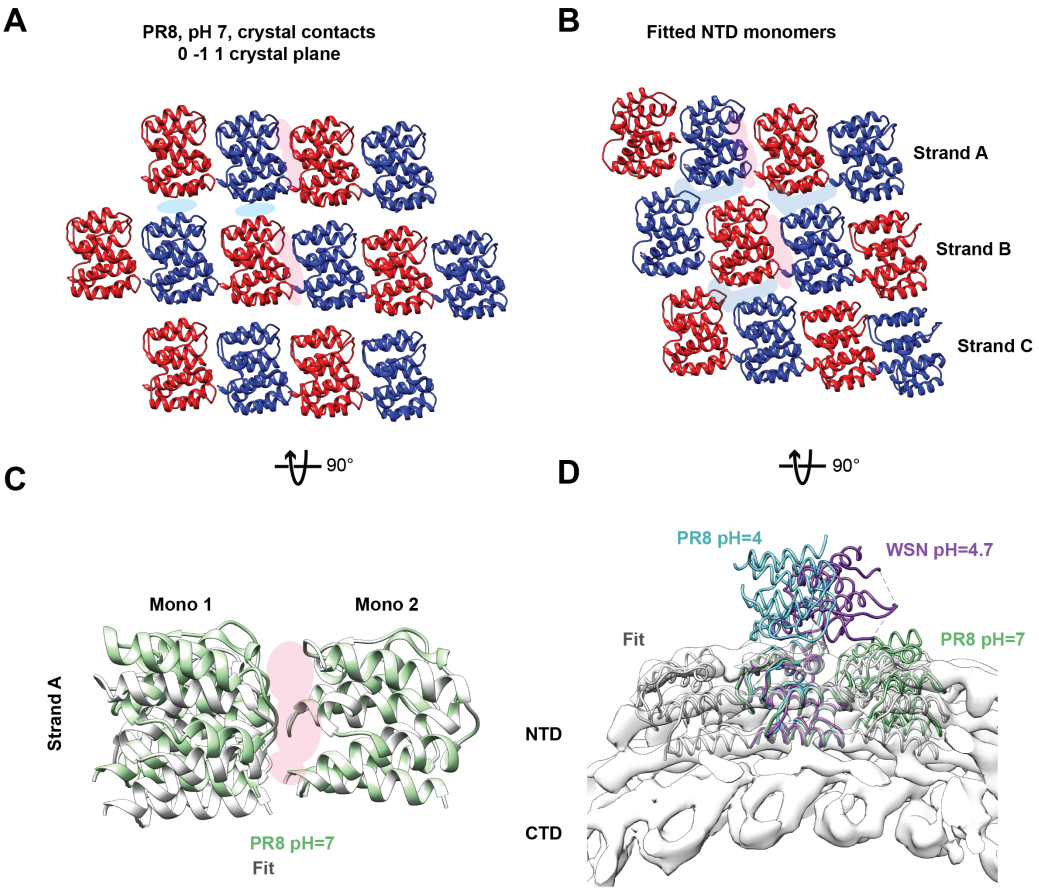

Figure 4.21 Comparison between M1-M1 interfaces identified from the subtomogram averaging reconstruction of M1 and M1 NTD crystal contacts. A) Crystal contacts within one crystal sheet of the neutral pH M1 NTD crystal. The two monomers representing the asymmetric unit are coloured in red and blue, PDB:1EA3. B) Twelve NTD monomers arranged according to the fit of each monomer as a rigid body into three parallel strands of the M1 in situ structure. To facilitate visual comparisons, neighbouring monomers are also coloured in red and blue. In exemplary positions the lateral interface between strands is marked in light blue and the interfaces between neighbouring monomers in rose. Corresponding interfaces in the crystal in A are coloured accordingly. C) Two M1 NTD monomers fitted as two rigid bodies into the M1 subtomogram averaging reconstruction (white) are overlaid to the dimer representing the asymmetric unit in the neutral pH M1 NTD crystal (green). The dimer interface corresponding to the M1-M1 intra-strand interface in the subtomogram reconstruction is coloured in rose. D) Three M1 NTD monomers fit into the outer lobe of the M1 reconstruction from subtomogram averaging (light grey). Three different crystal dimers obtained under different conditions are overlaid to the fitted NTD monomers. PDBs: 1EA3 (green) 1AA7 (light blue) and 5WCO (purple). Crystallization conditions and virus strains as in the corresponding colour.

filament length. Since filaments can be up to 10 µm to 20 µm long it appears that these events occur at a considerable frequency. In two other filaments, I found regions where the VLP filaments were not straight but instead bended or kinked. I detected that in these cases a set of parallel strands would finish and new strands would start and initiate a slight change in direction. In 41 out of the 44 VLP filaments, I found M1 to be arranged in strands with a right-handed helical propensity while for three filaments, I found M1 to be arranged with a left-handed helical propensity (Fig. 4.22B,C).

The arrangement of M1 in viruses was consistent with how M1 arranges in VLPs. Essentially, M1 occurs as multiple, parallel strands with a helical propensity following the inside of the virus membrane. The center-to-center distance between neighbouring M1 strands is identical to the distance of 38 Å determined for M1 polymers in VLPs. Comparing individual virus filaments I found that the number of strands varies between one and seven and shows a similar distribution as observed in VLPs. In all but one filaments, strands were found to have a right-handed helical propensity, while a single virus filaments M1 strands exhibited a left-handed helical propensity. Based on the described similarities between virus and VLP M1 data I conclude, that the M1 arrangement is indistinguishable between viruses and VLPs.

4.2.12 Polarity of M1 strands

Virus and VLP tubes have an orientation (Fig. 4.23A). All filaments have a tip that assembles first and is typically dome shaped. In virus filaments the tip is where the genome is located. It appears that the radius in the region around the genome can be slightly wider than in the rest of the filament (Fig. 3.4, Fig. 3.11). The rear end of the tube is also dome shaped and forms during scission of virus and VLP tubes from the host cell. Tube rear ends were previously suggested to display a high concentration of NA molecules which was also found in several cases in my data (Fig. 3.4, Fig. 3.11, Fig. 3.9) (*32*).

M1 strands are polar and have a directionality. Here, I demonstrate the polarity of M1 by using the above-described scenario of a trans connectivity between the NTD and the CTD. If several M1 monomers polymerize into one strand a free NTD will be located at one of the ends of the polymer and a free CTD will be located and the other end of the polymer (Fig. 4.23B). By displaying the M1 reconstruction back into the tomogram at the subtomogram positions, the M1 polarity relative to the orientation of the tomogram can be assessed (Fig. 4.23B,C).

In a next step I tried to assess if the M1 strand polarity is fixed relative to the orientation of the corresponding virus or VLP by comparing the virus orientation

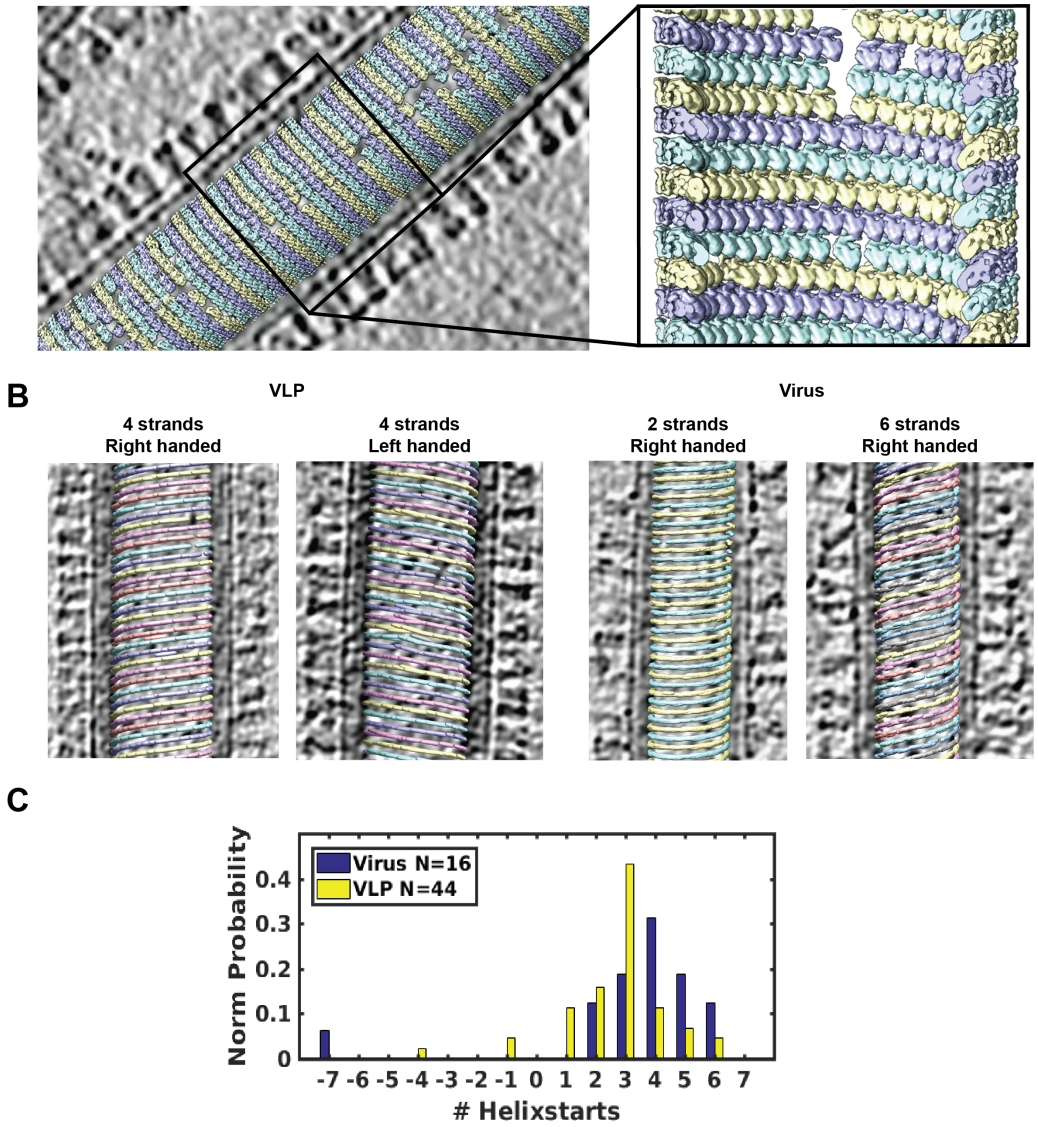

Figure 4.22 M1 arrangement in virus and VLP filaments. A) Visualization of the arrangement of M1 in a VLP filament with three parallel M1 strands (yellow, cyan, purple) generated by projecting the final M1 reconstruction back into the tomogram at the M1 subtomogram positions from the final alignment step. The inset shows a zoomed in tube section where the inside of the tube is oriented towards the viewer. B) The arrangement of different numbers of parallel M1 strands for two exemplary VLP and two exemplary virus filaments. In the second VLP example, strands exhibit a left-handed propensity. C) Distribution of the number of parallel M1 strands (number of helixstarts) found in viruses and VLPs. Negative numbers of helixstarts correspond to a left handed propensity.

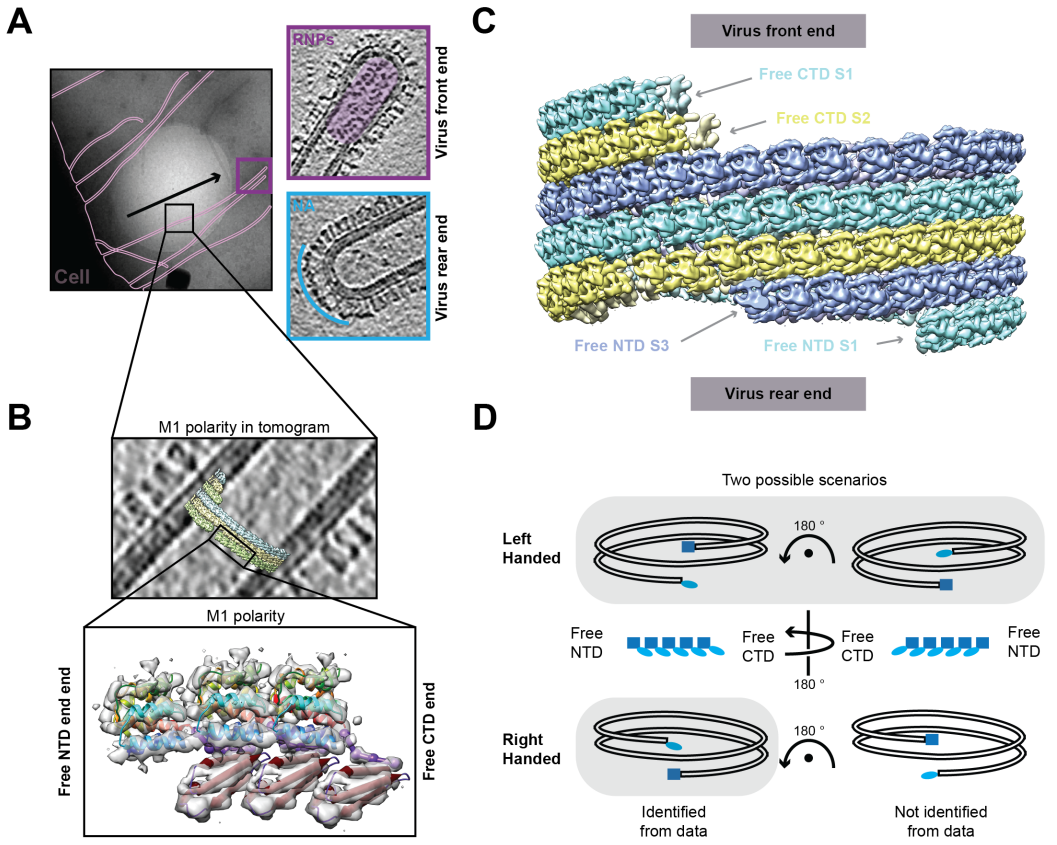

Figure 4.23 Polarity of M1 strands within virus filaments. A) Orientation of virus filaments. Left panel: medium magnification image of virus filaments before scission. The tip of the budding filament is highlighted in purple. Right panel: Image of a virus front end with vRNPs (purple frame) and image of a virus rear end with a NA cluster (cyan frame). B) Upper panel: Projection of the M1 structure at subtomogram positions into the respective tomogram allows to determine the M1 polarity in the corresponding tomogram. Lower panel: Three M1 monomers visualized according to the trans connectivity model. The polarity of M1 is clearly visible, since one end of the strand is represented by the free CTD and the other end of the strand is indicated by the free NTD. C) Three M1 strands as arranged underneath the membrane displayed for a subsection of a virus tube. The relative orientation of the virus and the polarity of the M1 strands is indicated. D) A model to visualize the connection between polarity and handedness demonstrating the potential M1 arrangement in cases where M1 strands exhibit a left handed helical propensity.

and the M1 polarity in my data. Across the virus and VLP data sets, I found three virus and two VLP tubes, for which I was able to clearly identify the orientation of the corresponding virus or VLP filament. The virus orientation can be determined either through the presence of the genome (in the virus case only) or through a visible host-cell connection identifiable from lower magnification maps of the corresponding tomogram positions (Fig. 4.23A,B). I compared the polarity of the VLP M1 strand in these tomograms to the orientation of the corresponding virus and VLP filament. I consistently found that the unbound CTD points in the direction of the filament's tip, while the free NTD faces the base of the filament (Fig. 4.23C).

Assuming that M1 polymerizes from one end of the strand only, my observation that the free NTD faces the rear end of the filament which is closer to the cytosol, suggests that M1 strands polymerize from the NTD end. It also appears unlikely that the polymerization occurs at the front end of the virus, since this would require large amounts of M1 monomers to passively diffuse or be actively transported through a narrow tube to reach the polymerization site. In such a scenario one would further expect the virus tube to be filled with M1 monomers, a situation which is not reflected in my data.

The identified relation between virus orientation and M1 polarity refers to cases where M1 strands exhibit a right handed propensity. Handedness and polarity are connected. Due to the small number of filaments (4 out of 60) where M1 strands exhibit a left handed propensity all cases for which the polarity of M1 was examined relative to the orientation of the virus or VLP filaments refers to M1 strands with a right handed helical propensity. Based on the polarity analysis for M1 strands with a right handed helical propensity I deduced two different scenarios for M1 strands with a left handed helical propensity (Fig. 4.23D).

To convert the handedness of a strand with a right handed helical propensity into a strand with a left handed helical propensity, the free NTD that was facing the base of the filament in the right handed helical case would now face the tip of the filament (Fig. 4.23D). If we assume that polymerization happens at the free NTD only, this would mean, that strands with a left handed propensity polymerize from at the tip of the assembling virus. Alternatively, one can imagine that in the left handed case monomers are rotated by 180° around the z-axis, orthogonal to the membrane tube surface (Fig. 4.23D). Polymerization from the free NTD side could then lead to the formation of strands with a left handed helical propensity.

4.2.13 M1 in virus assembly - conclusions from a theoretical model

The presented results about the arrangement of M1 provide important, new insights to extend our understanding of IAV virus assembly. The organization of M1 into a single or several parallel strands suggests a model in which parallel polymerization of M1 at multiple strands provides the energy to drive protrusions of new virus particles from the host cell. I therefore collaborated with Serge Dmitrieff (Nedelec group EMBL, now institut Jacques Monod) who developed a theoretical model to describe M1 polymerization and the virus assembly process. Based on this theoretical model, we can make several predictions which we can then compared experimentally obtained data about M1 arrangement and the appearance of assembled viruses.

Description of the model

The model describes a simplified two-component system consisting of membrane plus glycoproteins as one component and the M1 polymer as another component.

In the model, M1 generates force through polymerization, similar to force generated when actin polymerizes. Mathematically this can be expressed as the polymerization force $f1 = df/dL$ which depends on L the length of the added monomer and an energy gain $df < 0$ for addition of a monomer to the polymer (Fig. 4.24A). The M1 polymer is described as a 3D curve that has freedom to bend around all three axis described by the two curvatures C_v, C_u and the torsion T (Fig. 4.24B). There are three different rigidities one associated to each bending direction.
Deformation of the glycoprotein-covered membrane into tube-like virus filaments requires energy. The energy required to deform a patch of membrane depends on the mechanical properties of the membrane determined by the membrane tension σ, the bending rigidity κ and the spontaneous membrane curvature C_0 (Fig. 4.24A) (*164*).

The theoretical description for deforming a piece of membrane and the energy gain from polymerizing M1 are combined to calculate the total energy of the system. Based on the evaluation of the total energy, the model makes several predictions that I tested using my data. The model predicts, that within a cylindrical IAV membrane tube, the M1 polymer adopts the shape of a regular helix with flexibility in the direction of torsion (Fig 4.24). Indeed, the obtained data on M1 arrangement show that inside virus filaments, M1 has the shape of a regular helix. The experimental data further show that for M1 the curvature $C_u = 0$ which suggests a high rigidity in this direction (Fig. 4.24A). Since $C_u = 0$, the relevant curvature of M1 in the context of the virus relates to C_v to which I will refer to as 'the curvature of M1' in the following.The experimental data further show that M1 arranges in multiple parallel strands with a helical propensity where the number of strands effectively affects the incline and

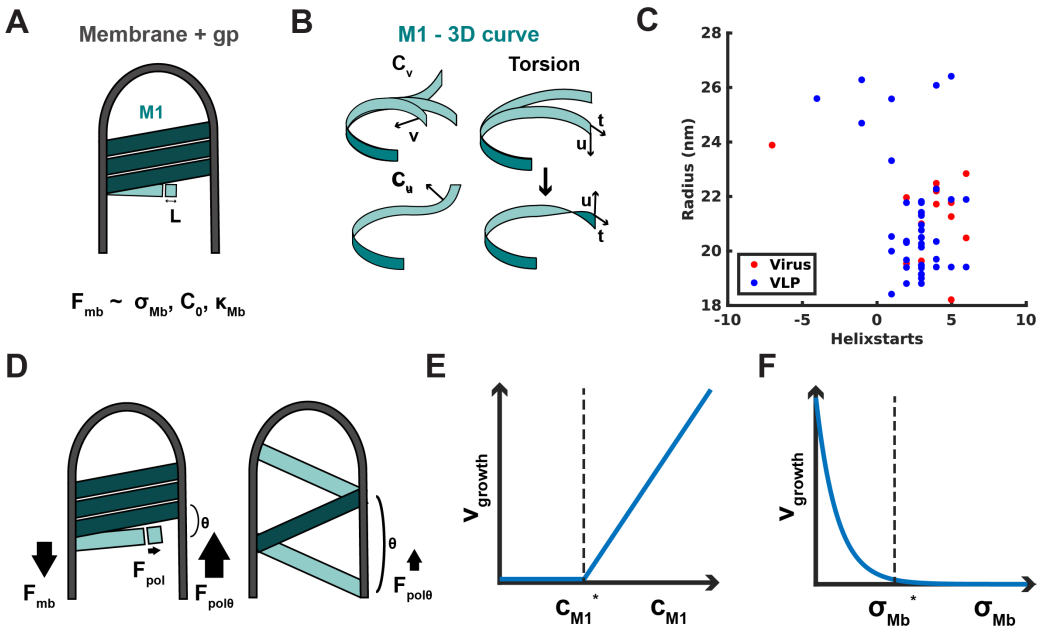

Figure 4.24 A theoretical model of virus assembly. A) Components of the theoretical model: Tube-shaped glycoprotein covered membrane and the M1 polymer. The force required to deform membrane into a tube F_{mb} depends on the membrane tension σ_{mb}, the spontaneous curvature of the membrane C_0 and the bending rigidity of the membrane κ_{mb}. The M1 polymer extends through the addition of monomers with length L. B) The polymer M1 is described as a 3D curve defined by two curvatures C_v and C_u and the torsion T with associated rigidities. Torsion is affected by the helical incline of the strand. C) Scatterplot of filament radius and number of parallel M1 strands for each virus and VLP filament analysed. Negative helix start numbers correspond to left handed helices. D) M1 polymerization generates force in the direction of polymerization. The amount of usable force for membrane extension is anti-correlated to the polymerization angle θ. E) Theoretically expected relationship between the growth velocity V_{growth} and the concentration of M1 C_{M1}. A critical concentration $C_{M1}*$ has to be reached to initiate polymerization. F) Theoretically expected relationship between V_{growth} and the membrane tension σ_{mb}. If the tension exceeds σ_{mb} no filaments are formed.

therefore the torsion of each strand (Fig. 4.23B). This observation means that M1 is flexible in the direction of torsion to accommodate different number of parallel M1 strands just as predicted by the model.

Any pushing or pulling force exerted by M1 onto a piece of membrane will lead to the deformation of membrane into a tube. The tube dimensions depend on the mechanical properties of the membrane and the applied force. Based on this relationship, we estimated the energy required to form a membrane tube based on the tube dimensions observed in the experimental data and assuming typical values for the mechanical properties of membranes such as the membrane tension ($\sigma = 10^{-5} - 10^{-3}$ $\mathrm{N\,m^{-1}}$), the bending rigidity ($\kappa > 20 k_B T$) and the spontaneous membrane curvature ($C_0 < 1/100\ \mathrm{nm^{-1}}$). If we consider a IAV filament radius of 30 nm we find that 20 pN to 200 pN force are required for tube formation. According to polymerization energies of other small proteins such as actin, we assume a low gain of energy, around 1 pN for the addition of one monomer to the polymer. If M1 polymerized into the direction of the growing tube, this energy would not be sufficient to generate enough force to induce formation of tubes with dimensions observed in the experimental data. However, we found that M1 arranges as strands with a helical propensity and therefore with an angle θ relative to the central tube axis. The force generated into the direction of the growing tube scales by the polymerization angle θ and is largest if M1 polymerizes orthogonal to the tube axis: $F_\theta = f 1/sin(\theta)$ (Fig. 4.23D).

This prediction can be compared to the experimental data and we find that helical strands in all cases are very tightly packed and thus that the helix incline angle corresponding to θ is always small. The incline angle θ however varies slightly for different virus particles, since we find different numbers of parallel protein strands. According to the model, a single protein strand would be the most efficient and as such the situation expected to be found. However, overall the incline angle θ might remain small enough for scenarios of one and up to seven M1 strands found in my data to still exert enough force to induce tube formation. This suggests, that mechanics in the end is not determining the number of M1 strands but instead that the number of M1 strands as well as the handedness is determined by the kinetics of the assembly process.

According to the theoretical model, the radius of the virus filament is controlled by the contributions of the curvature of M1 and the curvature of the membrane. The curvature of M1 is determined by the bending rigidity and the spontaneous curvature of the M1 polymer while the membrane curvature is determined by the bending rigidity κ_{mb} of the membrane, the spontaneous curvature C_0 of the membrane but not membrane tension σ_{mb} as long as the membrane is covered with glycoproteins. To

evaluate the predictions from the model, I have measured the radius of all virus and VLP filaments, for which the arrangement of M1 was determined. The results show that the radius varies for different filaments independent on other parameters such as the number of M1 strands or particle type (Fig. 4.23C). Based on the results from experimental data in combination with the prediction from the theoretical model, we expect that the observed distribution in the IAV particle radii results from variability in the membrane curvature due to local changes in membrane composition. We further suggest that flexibility of the M1 polymer allows M1 to adapt to these variations in curvature within certain thresholds.

Finally, the model makes additional predictions about the assembly process which we cannot yet test with the type of experimental data at hand. The theoretical model predicts, that membrane tension and M1 concentration do not affect the geometry of virus filaments but rather that critical thresholds have to be reached for filament growth to occur. Thus, too low M1 monomer concentration as well as too high membrane tension can slow down filament growth and eventually stop filament extension (Fig. 4.23D).

In summary, predictions from a theoretical model which describe the mechanics of M1 in the context of a growing membrane protrusion are consistent with the presented experimental data on in situ arrangement of M1. Based on our results, we propose that efficient polymerization of M1 into a densely packed arrangement like the one found here, generates sufficient energy to drive the growth of a budding virus filament and represents the driving force for IAV assembly.

4.3 Discussion

I have described and evaluated two M1 reconstructions obtained directly from IAV virus and VLP filaments which provide first insights into a full-length structure of M1. Despite the small size of 28 kDa and the resulting low signal in EM images I demonstrated that a reconstruction of M1 by subtomogram averaging from cryoET data at a resolution below 1 nm can be obtained. CryoET and subtomogram averaging typically result in lower resolution structures in comparison to single particle methods. However, subtomogram averaging provides the advantage to not exclusively rely on the signal from a single protein, randomly distributed within an ice layer, but instead can make use of a priori information from the context of the protein. This can be a strong advantage for studying small proteins. In the case of M1, I made use of this advantage by deliberately placing initial subtomogram positions beneath the virus

membrane which significantly reduced the parameters search space during alignment. Further, the signal for initial alignment of M1 subtomograms is not exclusively defined by the features of an individual protein but rather includes larger scale features such the arrangement of M1. The regular spacing of M1 strands in my case was a strong feature for initial subtomogram alignment. Nevertheless, the small size still represented substantial challenges and needed extensive fine-tuning and specific improvements of the established subtomogram averaging workflow. In particular the separation of individual subunits along the M1 strands was challenging and success varied between tomograms which resulted in M1 reconstructions with reduced quality and which were hard to interpret. However, employing amplitude spectra weighting instead of binary wedge masking for subtomogram averaging and geometric cleaning helped to improve quality of the final M1 reconstruction and I expect both tools to be suitable for other cases where subtomogram averaging with conventional approaches will turn out to be challenging.

Despite considerable efforts to optimize M1 subtomogram averaging processing in the case of the HK68 virus sample, the final M1 reconstruction obtained did not reach the same quality as the structure obtained for M1 from HK68 VLPs. While the global reported resolution of both reconstructions was comparable, I found that the resolution of M1 from HK68 virus was more anisotropic, which resulted in lower quality and reduced interpretability of the final M1 reconstruction. To understand these differences I determined the optimal defocus value range of data for M1 subtomogram averaging and found that the chosen settings in the dataset of HK68 virus were not ideal for subtomogram averaging of M1. However, when the data were initially collected the arrangement and structure of M1 and therefore the criteria which need to be considered to design the data collection parameters were not known yet. Therefore, the obtained insights into arrangement and consequently the spatial frequencies which are important for subtomogram alignment can be used in the future to improve data collection settings specifically for subtomogram averaging of M1.

Based on these requirements, I have recently collected a large dataset of unfixed HK68 virus filaments using a defocus range from -1.5 µm to -4 µm which I expect to give an improved M1 structure from HK68 virus data (see section 3.2.7). I therefore plan to use the established subtomogram averaging workflow for M1 to process the recently acquired tomograms. Due to the heterogeneous arrangement of M1 for different filaments, only a small subset of filaments is typically homogeneous enough to be combined and to calculate a final M1 reconstruction. Consequently, the relatively large size of the new unfixed HK68 virus dataset (110 tomograms) might increase the

number of filaments with a homogeneous M1 arrangement and therefore the quality and resolution of the obtainable M1 reconstruction. Finally, processing a dataset of unfixed HK68 viruses for M1 might help to understand if the fixation, for which an effect was discovered for the in situ structure of HA (see chapter 3) has any effects on the in situ structure of M1

.

4.3.1 Structure of the M1 CTD

The obtained M1 reconstruction from HK68 VLPs has provided a direct structural representation of the full-length structure of the IAV M1 including the IAV M1 CTD inside HK68 VLPs. While previous reports have suggested the M1 CTD to be unstructured or highly flexible (*147*) my data clearly show that the CTD of membrane-bound M1 inside the virus has a defined structure which I expect to correspond to a core of four alpha helices, similar to the M1 CTD structure of ISAV (*162*). This interpretation of the M1 CTD is further supported by performed SSP of the M1 CTD which also suggests the presence of 4 alpha helices connected by short loops.

From the M1 structure obtained by subtomogram averaging, I was not able to unambiguously determine the connectivity of the identified density elements in the M1 CTD region of the M1 reconstruction. I therefore presented three possible connectivity models for the connection of M1 CTD helices and the connection between the M1 CTD and the M1 NTD. In the M1 reconstruction from HK68 VLPs I identified the density, D0, which could be connecting density between the M1 NTD and density elements of the M1 CTD. Only the trans connectivity model would include a structure which fits this density D0. In the other two models, this density would remain mostly empty. In addition, a connection of the M1 NTD and the CTD in trans resembles how the two domains are connected in the ISAV M1 structure. However, due to the low sequence homology of IAV and ISAV no clear assignment between densities from the IAV M1 structure presented here and the ISAV M1 structure was possible.

Based on the obtained results the trans connectivity model is best supported by my data, but none of the models could definitely be excluded. To confirm the suggested connectivity model and ultimately the structure of the M1 CTD, clarification through additional experiments and ideally a higher resolution structure of the M1 CTD are required. While I tried to approach the lack of a high resolution structural model by attempting to crystallize the M1 CTD, these attempts have not been successful. Nevertheless, I established an efficient expression and purification workflow for the

IAV M1 CTD fragment. Consequently, more variants of the construct as well as more crystallization conditions could be tested in the future. Alternatively, further improvements of the subtomogram averaging workflow and optimization of tomography parameters specifically for M1 might allow to increase the obtainable resolution and quality of an in situ M1 structure (see also section 4.3). Finally, it could be tested if and how M1 polymerizes in vitro and if in vitro assemblies are homogeneous enough to be targeted by single particle cryoEM or helical reconstruction to obtain a higher resolution structure of M1 including the M1 CTD.

4.3.2 M1-membrane interactions

The obtained subtomogram averaging reconstruction of M1 in combination with existing high resolution models of the M1 NTD allowed me to characterise the interface between the M1-NTD and the membrane inside viruses. I observed that one side of the NTD faces the membrane which is lined by a number of positive side chains, including the nuclear localization signal (NLS) located on H6 of the NTD. In previous studies, interactions between M1 and membrane were observed for lipid mixtures containing high concentrations of posphatidylserine in vitro as well as inside cells (*150*, *151*, *165*). Additional mutation studies have identified that removing charged residues which I found to be located at the membrane interface reduces virus fitness and for most of them, completely inhibits virus formation (*120*). Therefore, in concordance with previous M1-membrane interaction studies, my data confirm the hypothesis that the membrane M1 interaction is of electrostatic nature and is essential for virus budding.

Until now it remained unclear if the NTD or the CTD or both domains are involved in membrane interactions. In recent interaction studies M1-membrane interactions were only observed in full-length M1 while truncated versions, of either the NTD or the CTD lower the membrane affinity (*166*). From the presented data I can exclude a direct contribution of the M1 CTD to M1-membrane interactions inside assembled viruses. However, it seems possible, that M1 polymerization relies on interactions which include the CTD. It was further shown that M1 preferentially oligomerises when membranes are present (*165*) which further supports the idea of a link between M1 oligomerisation and membrane binding. Electrostatic M1 membrane interactions could be easily susceptible to pH changes and could explain the observed pH-induced membrane detachment and reorganization of M1 once the pH inside the virus is lowered upon entering the endosomal system of host cells (*167*, *168*).

4.3.3 M1-M1 interfaces and M1 arrangement inside viruses and VLPs

Here I have shown, that M1 is arranged in parallel strands of M1 polymers underneath the membrane. I identified three M1-M1 interfaces from fitted M1 NTD models in the M1 reconstruction from subtomogram averaging. Since the in situ M1 structure includes the CTD, I can estimate the contribution from the CTD to any of the interfaces. I found indications that in particular the intra-strand M1-M1 interface seems to include interactions of neighbouring CTDs based on their close proximity. In addition, the NTD-NTD intra-strand interface on its own, does not appear to possess any features to mediate a strong interaction but consists mostly of residues which interact via hydrophobic interactions. A comparison to the corresponding interface in the neutral pH crystal has revealed that it is likely that the M1 intra-strand interface is flexible enough to adapt to different virus filament radii observed in the data without compromising the M1-M1 interaction along M1 strands. Unlike the M1 intra-stand interface, the M1 inter-strand interface includes residues of opposing charge. I therefore concluded that the inter-strand interactions of M1 are based on electrostatics. The comparison to crystal contacts from the neutral pH M1 NTD crystal has revealed a similar M1 NTD-NTD interface, where the corresponding 'strands' of subunits are slightly shifted against each other. This shows that the subunits can interact at multiple positions along the strands which supports the hypothesis that the inter-strand interface is flexible enough to allow sliding of the strands relative to each other. Flexibility between the strands would allow to adapt to irregularities in the arrangement, such as the identified changes of the the M1 strand number.

Interestingly, I found that helix 6, which contains a set of positively charged side chains including the NLS, is located at the M1 inter-strand interface. Helix 6 is also located at the M1-membrane interface. Therefore it seems likely that some residues along helix 6 are involved in M1-membrane interactions while others are involved in M1-M1 interactions along the inter-strand interface. Due to previously suggested contributions of helix 6 to membrane binding, RNA binding and virus assembly helix 6 targeted in several previous mutagenesis studies (*120*, *169*, *170*). Several positions in the sequence of helix 6 have been found to potentially affect virus morphology such as 95K/R and 41V/A which are two morphology determining sites in the M1 sequence (*28*, *29*). The occurrence of morphology determining mutations in this interface suggests that modulations of the inter-strand interface can affect M1 arrangement in a way that affects virus morphology. In filaments, as described here, parallel M1 strands are arranged straight next to each other. In spherical particles and dome-shaped

filament tips, one can imagine that the parallel strands have to be arranged with a tilt or a shift relative to each other in order to generate additional curvature. Therefore, it seems imaginable that such an effect could be induced by mutations located within the inter-strand interface.

4.3.4 M1 in assembly

The combination of the presented data on M1 arrangement with a theoretical model suggests that assembly can be driven by M1 polymerization. We further suggest that the arrangement of polymerized M1 in combination with the mechanical properties of the protein-covered membrane determines the dimension of the budding virion. My data on HK68 VLPs, in accordance with data by others (*45*) show that HK68 VLPs, which do not contain vRNPs, are highly similar in architecture and in their M1 arrangement to HK68 virus particles. This high level of similarity supports a model where the presence of the vRNPs is not essential for assembly initiation. We instead propose that M1 polymerization underneath the glycoprotein-covered plasma membrane might initiate formation and elongation of a membrane protrusion. It was previously shown that dense coverage of the membrane with glycoproteins affects membrane properties such as membrane curvature (*171*). In this context, we suggest, that the presence of glycoproteins at the membrane generates an environment with a lowered energy barrier at the assembly site and M1 polymerization can initiate the formation of membrane protrusions.

The presented data have provided insights into the flexibility of the M1 arrangement to adjust to variations in curvature and to allow rearrangement of parallel M1 strands. The obtained insights into the M1 arrangement along filament sides from experimental data as well as from the theoretical model allows us to make further suggestions about the M1 arrangement in filament tips. In all virus and VLP tips found in tomograms, M1 is clearly present and looks very similar to the M1 layer at filament sides (Fig. 3.9, Fig. 3.11). To arrange tightly and in a similar fashion within dome-shaped tips, M1 has to be able to firstly further increase the curvature along the direction of the radius and secondly M1 has to be able to accommodate to a second direction of curvature along the inter strand interface, as described above. It is possible that the arrangement of M1 within dome-shaped filament tips results from additional geometrical restrictions of the membrane layer at the initiation of tube formation. At the rear end of filaments, it is possible that clusters of NA induce extra curvature which contributes to generating the dome-shaped base.

4.3.5 Conclusions

1. Using an optimized subtomogram averaging workflow has resulted in in situ structures of IAV M1 at 8 Å resolution from HK68 virus tomogram data and HK68 VLP tomogram data.

2. By fitting existing crystal structures of the M1 NTD in the outer lobe of the obtained M1 subtomogram averaging reconstruction I confirmed that the M1 NTD interacts with the virus membrane via a positively charged interface while the M1 CTD does not interact with the membrane.

3. Density in the obtained M1 reconstruction in combination with results from SSP and a comparison to a distantly homologue M1 structure gave first insights into the structure of the M1 CTD.

4. M1 arranges in one or multiple parallel strands mostly exhibiting a right handed helical propensity underneath the membrane.

5. M1 Intra-strand interfaces via the M1 NTD are dominated by hydrophobic interactions and allow M1 to adapt to different radii of virus filaments, which might result from variation in membrane properties.

6. M1 Inter-strand interfaces involve electrostatic interactions and I suspect that flexibility in this interface allows M1 strands to slide relative to each other to increases the flexibility of the virus filament and react to spontaneous changes in M1 strand number.

7. In conclusion with a theoretical model we propose that linear polymerization of M1 into tightly packed strands at the glycoprotein covered plasma membrane drives IAV assembly.

Chapter 5

Discussion and future perspectives

Understanding the structure of influenza virus is crucial to understand many aspects of the virus life cycle and in particular host cell infection and virus assembly. CryoET in combination with subtomogram averaging is the only approach to gain direct structural information on pleomorphic particles such as influenza virus. Here I used cryoET and subtomogram averaging to investigate the structure and arrangement of the two proteins HA and M1 directly from tomograms of the human HK68 influenza virus strain and corresponding HK68 VLPs.

In the first part of this study the sample preparation workflow was adapted. I have demonstrated that influenza virus purification can be detrimental to virus morphology and particle integrity. Therefore an alternative sample preparation approach where the virus is directly produced on the EM grid, in the vicinity of the host cell, was shown to be suitable to prepare influenza virus samples for cryoET and subsequent subtomogram averaging. I ,as well as others before, observed up to 20 µm long filaments in non-purified samples which are absent in purified samples. The observed shift in the detectable virus morphology suggests that many experimental set-ups, which include purification and potentially the use of chicken eggs for virus amplification, rely on virus particles with a strong bias towards spherical morphology (*130*, *131*, *172*). This systematic experimental bias might be one explanation why the filamentous morphology despite its association with human infection remains less widely known and studied than spherical influenza virus particles (*17*).

Virus samples prepared without purification exhibit additional advantages since the native context of produced viruses including the connection to the host cell is maintained. This allowed me to determine the virus orientation relative to the M1 polarity. In the future, the presence of the native environment of the virus may allow to address additional questions regarding aspects of host cell infection and virus assembly

such as the organization of proteins at the inner membrane where virions assemble and how scission of virus particles occurs. In addition, influenza A virus samples prepared on EM grids can be combined with cryo focused-ion-beam milling to target various aspects of the virus life cycle within the host cell such as vRNP transport or the effects on membrane alteration reported to be induced by viral proteins inside the host cell. Overall, the approach used here appears to be a better alternative to study the structure of other enveloped, pleomorphic viruses by cryoET and provides various opportunities to study additional aspects of the virus life cycle.

CryoET in combination with subtomogram averaging tailored to the two IAV proteins HA and M1 has allowed me to obtain in situ structures of both proteins at resolutions which resolve secondary structure elements. Thanks to the achieved resolution I was able to compare in situ reconstructions of both proteins to existing high resolution crystal structures. Thereby I deduced information about structural variability of HA on the virus surface and the structure of the influenza matrix layer formed by M1. The in situ structure of M1 allowed me to describe the M1-membrane and M1-M1 interfaces for the first time. In addition, protein positions could be determined for both proteins for the same virus particles in the same data. The quantification of the distribution of HA on the virus surface and the arrangement of M1 underneath the membrane has elucidated several aspects of virus architecture which have not been shown before.

A comparison of non-infectious VLPs with infectious virus allowed me to assess the effects of virus inactivation on protein structure. Virus inactivation through chemical fixation is a common technique to treat viruses for subsequent work outside of high biosafety levels. Additionally, virus inactivation through chemical fixation is a commonly used method to prepare influenza virus vaccines. While chemical fixation through aldehyde crosslinking has long been implicated in altering protein structure, the exact effects in many cases and in particular directly on the virus surface remained unknown. I provided data which describe the molecular details of surface exposed HA on viruses upon addition of an aldehyde-based fixative. I have proposed that aldehyde fixation induces specific inter-molecular cross links in the head region of the protein. Thereby crosslinking seems to stabilize an open HA conformation resembling conformations which have been recently described to occur upon antibody binding and in the context of HA breathing motion. Understanding the structural details of breathing motion might help to uncover overlooked, and potentially more conserved antibody binding sites on the HA surface. Together with data from others, the insights into the conformational changes on HA shown here provide a first structural foundation

for subsequent and higher resolution structures describing breathing motion. While in this study, fixation was initially employed for biosafety reasons, it remains possible that fixation under specifically tuned conditions might potentially be useful to capture additional structural intermediates along the conformational path towards and during membrane fusion on the virus surface.

From the accurate and precise descriptions of protein localization within the virus, I found that HA arranges with a low degree of order and an average protein-protein distance of around 8 nm to 12 nm. In the same virus particles, I found that M1 arranges in parallel strands with helical propensities of variable helix number and variable strand number. Distances in the M1 arrangement where highly conserved to 3.8 nm distance between strands and 2.8 nm between subunits along strands. From a comparison of the arrangement of HA and M1 I did not find any evidence that M1 and HA arrange in an orderly fashion with respect to each other. Nevertheless, these insights do not exclude transient interactions of the two proteins. These observations are in concordance with studies describing that neither the cytoplasmic tails nor specifically HA, but either of the glycoproteins, are required at the assembly site for particles to form (*45*, *152*). This leads to the hypothesis that unspecific contributions from the presence of the glycoproteins rather than specific interactions between HA and M1 are relevant for virus and VLP assembly.

In conclusion with the previous section and work from others I propose a model of virus assembly, where the presence of glycoprotein in high concentrations at the assembly site is mainly required to modulate the mechanical properties of the membrane to lower the energy barrier for membrane deformation. I propose that M1 polymerization within such an environment is sufficient to deform the membrane and induce the formation of a membrane protrusion. Based on the data presented in this study it appears very likely that M1 in competition with membrane determines particle shape rather than HA or the vRNPs as proposed previously. It remains possible that the presence of the vRNPs increases the efficiency of the initiation process for example by further decreasing the energy barrier. Thereby, the presence of the vRNPs could compensate lower glycoprotein concentrations. Interestingly, the observation that filaments are wider in regions where the vRNPs are located, could suggest that the width of the vRNPs is larger than the preferred radius of M1 and the membrane. Therefore, the radius of the filament decreases at the end of the vRNPs where M1 relaxes back to its favourite curvature.

Based on presented results on M1-M1 interfaces together with previous mutation

studies of M1 I speculated that effects from single point mutations within the M1 inter-strand interface induce alterations in the relative orientations of two neighbouring strands which support a spherical virus particle structure. Suspected similarities in the M1 arrangement between filament tips and spherical virus particles suggest that extra mechanical forces induced by the membrane during assembly initiation introduce an effect on M1 arrangement which resembles the effect from mutations within the M1 inter-strand interface. In the future, it would be exciting to study the arrangement in spherical particles as well as in filament tips to validate this hypothesis. While this study has contributed to increase our understanding of influenza virus architecture focusing on HA and M1, the distribution and in situ structure of NA is to be investigated further in the future. Due to the lower concentration of NA on the virus surface, NA appears to be more challenging but not impossible to target by an adapted subtomogram averaging analysis similar to the subtomogram averaging workflow described for HA.

I believe that more quantitative studies describing the in situ structure of influenza virus in combination with further progress in obtaining high resolution structures for all components will allow us to obtain an accurate 3D model of influenza virus in the future. Besides EM data, quantitative mass spectrometry has already contributed to quantifying virus and host cell proteins in virus particles, while methods such as subtomogram averaging will continue to exploit the arrangement and in situ structures of influenza virus proteins. Additional dynamic data such as recently obtained on influenza virus assembly via super resolution fluorescence microscopy (*31*, *32*) can further contribute to increase our understanding of the kinetics of the assembly process. Finally all quantitative data can be integrated and combined with additional models and dynamic simulation of influenza virus assembly to obtain not only a accurate structural but also a time-resolved model of IAV assembly.

References

(*1*) Paules, C., and Subbarao, K. (2017). Influenza. *The Lancet 390*, 697–708.

(*2*) Petrova, V. N., and Russell, C. A. (2018). The evolution of seasonal influenza viruses. *Nature reviews. Microbiology 16*, 47–60.

(*3*) Kilbourne, E. D. (2006). Influenza pandemics of the 20th century. *Emerging infectious diseases 12*, 9–14.

(*4*) Trombetta, C., Piccirella, S., Perini, D., Kistner, O., and Montomoli, E. (2015). Emerging Influenza Strains in the Last Two Decades: A Threat of a New Pandemic? *Vaccines 3*, 172–185.

(*5*) Iuliano, A. D. et al. (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet (London, England) 391*, 1285–1300.

(*6*) Neumann, G., Noda, T., and Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature 459*, 931–939.

(*7*) Worobey, M., Han, G.-Z., and Rambaut, A. (2014). A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature 508*, 254–257.

(*8*) Horimoto, T., and Kawaoka, Y. (2005). Influenza: lessons from past pandemics, warnings from current incidents. *Nature reviews. Microbiology 3*, 591–600.

(*9*) Cohen, J. (2018). Universal flu vaccine is 'an alchemist's dream'. *Science 362*, 1094.

(*10*) Yamayoshi, S., and Kawaoka, Y. (2019). Current and future influenza vaccines. *Nature Medicine 25*, 212–220.

(*11*) Asha, K., and Kumar, B. (2019). Emerging Influenza D Virus Threat: What We Know so Far! *Journal of Clinical Medicine 8*.

(*12*) Taubenberger, J. K., and Morens, D. M. (2010). Influenza: the once and future pandemic. *Public health reports (Washington, D.C. : 1974) 125 Suppl 3*, 16–26.

(*13*) WHO Recommended composition of influenza virus vaccines for use in the 2019-2020 northern hemisphere influenza season., https://www.who.int/influenza/vaccines/virus/recommendations/2019_20_north/en,last visited 25.05.2019.

(*14*) Furuse, Y., and Oshitani, H. (2016). Mechanisms of replacement of circulating viruses by seasonal and pandemic influenza A viruses. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases 51*, 6–14.

(*15*) Calder, L. J., Wasilewski, S., Berriman, J. A., and Rosenthal, P. B. (2010). Structural organization of a filamentous influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America 107*, 10685–10690.

(*16*) Vijayakrishnan, S., Loney, C., Jackson, D., Suphamungmee, W., Rixon, F. J., and Bhella, D. (2013). Cryotomography of budding influenza A virus reveals filaments with diverse morphologies that mostly do not bear a genome at their distal end. *PLOS Pathogens 9*, e1003413.

(*17*) Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D., and Hutchinson, E. C. (2016). Filamentous influenza viruses. *J Gen Virol 97*, 1755–1764.

(*18*) Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., and Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature 439*, 490–492.

(*19*) Dou, D., Revol, R., Östbye, H., Wang, H., and Daniels, R. (2018). Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology 9*, 1581.

(*20*) McGeoch, D., Fellner, P., and Newton, C. (1976). Influenza virus genome consists of eight distinct RNA species. *Proceedings of the National Academy of Sciences 73*, 3045–3049.

(*21*) Jagger, B. W. et al. (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science 337*, 199–204.

(*22*) Wise, H. M., Hutchinson, E. C., Jagger, B. W., Stuart, A. D., Kang, Z. H., Robb, N., Schwartzman, L. M., Kash, J. C., Fodor, E., Firth, A. E., Gog, J. R., Taubenberger, J. K., and Digard, P. (2012). Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain. *PLOS Pathogens 8*, e1002998–14.

(*23*) Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B., and Wilson, I. A. (2012). Organization of the influenza virus replication machinery. *Science 338*, 1631–1634.

(*24*) Choppin, P. W., MURPHY, J. S., and TAMM, I. (1960). Studies of two kinds of virus particles which comprise influenza A2 virus strains. III. Morphological characteristics: independence to morphological and functional traits. *The Journal of experimental medicine 112*, 945–952.

(*25*) Seladi-Schulman, J., Steel, J., and Lowen, A. C. (2013). Spherical influenza viruses have a fitness advantage in embryonated eggs, while filament-producing strains are selected in vivo. *Journal of Virology 87*, 13343–13353.

(*26*) Kilbourne, E. D., and Murphy, J. S. (1960). Genetic studies of influenza viruses. I. Viral morphology and growth capacity as exchangeable genetic traits. Rapid in ovo adaptation of early passage Asian strain isolates by combination with PR8. *The Journal of experimental medicine 111*, 387–406.

(*27*) Nakajima, N., Hata, S., Sato, Y., Tobiume, M., and Katano, H. (2010). The first autopsy case of pandemic influenza (A/H1N1pdm) virus infection in Japan: detection of a high copy number of the virus in type II alveolar epithelial cells ... *Jpn J Infect Dis*.

(*28*) Bourmakina, S. V., and Garcıa-Sastre, A. (2003). Reverse genetics studies on the filamentous morphology of influenza A virus. *Journal of General Virology 84*, 517–527.

(*29*) Elleman, C. J., and Barclay, W. S. (2004). The M1 matrix protein controls the filamentous phenotype of influenza A virus. *Virology 321*, 144–153.

(*30*) Cifuentes-Muñoz, N., Dutch, R. E., and Cattaneo, R. (2018). Direct cell-to-cell transmission of respiratory viruses: The fast lanes. *PLOS Pathogens 14*, e1007015.

(*31*) Vahey, M. D., and Fletcher, D. A. (2019). Influenza A virus surface proteins are organized to help penetrate host mucus. *eLife 8*.

(*32*) Vahey, M. D., and Fletcher, D. A. (2019). Low-Fidelity Assembly of Influenza A Virus Promotes Escape from Host Cells. *Cell 176*, 281–294.e19.

(*33*) Hutchinson, E. C. (2018). Influenza Virus. *Trends in Microbiology 26*, 809–810.

(*34*) Roy, A. M., Parker, J. S., Parrish, C. R., and Whittaker, G. R. (2000). Early stages of influenza virus entry into Mv-1 lung cells: involvement of dynamin. *Virology 267*, 17–28.

(*35*) de Vries, E., Tscherne, D. M., Wienholts, M. J., Cobos-Jiménez, V., Scholte, F., Garcıa-Sastre, A., Rottier, P. J. M., and de Haan, C. A. M. (2011). Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway. *PLOS Pathogens 7*, e1001329.

(*36*) Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell 23*, 847–858.

(*37*) Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., and Digard, P. (2001). Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway. *J Virol 75*, 408–419.

(*38*) Bui, M., Whittaker, G., and Helenius, A. (1996). Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J Virol 70*, 8391–8401.

(*39*) Sun, X., Tse, L. V., Ferguson, A. D., and Whittaker, G. R. (2010). Modifications to the hemagglutinin cleavage site control the virulence of a neurotropic H1N1 influenza virus. *Journal of Virology 84*, 8683–8690.

(*40*) Böttcher, E., Matrosovich, T., Beyerle, M., Klenk, H.-D., Garten, W., and Matrosovich, M. (2006). Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium. *J Virol 80*, 9896–9898.

(*41*) Rossman, J. S., and Lamb, R. A. (2011). Influenza virus assembly and budding. *Virology 411*, 229–236.

(*42*) Nakatsu, S., Sagara, H., Sakai-Tagawa, Y., Sugaya, N., Noda, T., and Kawaoka, Y. (2016). Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *mBio 7*.

(*43*) Noda, T., Murakami, S., Nakatsu, S., Imai, H., Muramoto, Y., Shindo, K., Sagara, H., and Kawaoka, Y. (2018). Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nature Communications 9*, 54.

(*44*) Hess, S. T., Gould, T. J., Gudheti, M. V., Maas, S. A., Mills, K. D., and Zimmerberg, J. (2007). Dynamic clustered distribution of hemagglutinin resolved at 40 nm in living cell membranes discriminates between raft theories. *Proceedings of the National Academy of Sciences 104*, 17370–17375.

(*45*) Chlanda, P., Schraidt, O., Kummer, S., Riches, J., Oberwinkler, H., Prinz, S., Kräusslich, H.-G., and Briggs, J. A. G. (2015). Structural Analysis of the Roles of Influenza A Virus Membrane-Associated Proteins in Assembly and Morphology. *Journal of Virology 89*, 8957–8966.

(*46*) Lai, J. C. C., Chan, W. W. L., Kien, F., Nicholls, J. M., Peiris, J. S. M., and Garcia, J.-M. (2010). Formation of virus-like particles from human cell lines exclusively expressing influenza neuraminidase. *J Gen Virol 91*, 2322–2330.

(*47*) Gómez-Puertas, P., Albo, C., Pérez-Pastrana, E., Vivo, A., and Portela, A. (2000). Influenza virus matrix protein is the major driving force in virus budding. *J Virol 74*, 11538–11547.

(*48*) Rossman, J. S., Jing, X., Leser, G. P., and Lamb, R. A. (2010). Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell 142*, 902–913.

(*49*) Davidson, S. (2018). Treating Influenza Infection, From Now and Into the Future. *Frontiers in Immunology 9*, 517–14.

(*50*) Williams, D. B., and Carter, C. B., *Transmission Electron Microscopy*; A Textbook for Materials Science; Springer Science & Business Media: 2009.

(*51*) Fujiyoshi, Y., *Low Dose Techniques and Cryo-Electron Microscopy*; Schmidt-Krey, I., and Cheng, Y., Eds.; Humana Press: Totowa, NJ, 2013, pp 103–118.

(*52*) FEI, Titan Condensor Manual FEI Online help.

(*53*) FEI, Alignment Manual FEI Online help.

(*54*) Russo, C., Lecture on EM Physics and Optics delivered at the LMB in 2017.

(*55*) Orlova, E. V., and Saibil, H. R. (2011). Structural analysis of macromolecular assemblies by electron microscopy. *Chemical reviews 111*, 7710–7748.

(*56*) McMullan, G., Chen, S., Henderson, R., and Faruqi, A. R. (2009). Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy 109*, 1126–1143.

(*57*) Faruqi, A. R., and McMullan, G. (2011). Abstract. *Quarterly Reviews of Biophysics 44*, 357–390.

(*58*) Rubinstein, J. L., and Brubaker, M. A. (2015). Alignment of cryo-EM movies of individual particles by optimization of image translations. *J Struct Biol 192*, 188–195.

(*59*) BRENNER, S., and HORNE, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica et biophysica acta 34*, 103–110.

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

(*61*) Dubochet, J., Adrian, M., Lepault, J., and McDowall, A. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends in Biochemical Sciences 10*, 143–146.

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

(*63*) Passmore, L. A., and Russo, C. J. (2016). Specimen Preparation for High-Resolution Cryo-EM. *Methods in enzymology 579*, 51–86.

(*64*) Scherzer, O. (1949). The Theoretical Resolution Limit of the Electron Microscope. *Journal of Applied Physics 20*, 20–29.

(*65*) Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America 111*, 15635–15640.

(*66*) Wade, R. (1992). A brief look at imaging and contrast transfer. *Ultramicroscopy 46*, 145–156.

(*67*) Erickson, H. P., and Klug, A. (1971). Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 261*, 105–118.

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

(*69*) Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol 192*, 216–221.

(*70*) Carroni, M., and Saibil, H. R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods 95*, 78–85.

(*71*) Henderson, R. (1992). Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice. *Ultramicroscopy 46*, 1–18.

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

(*73*) Sigworth, F. J. (2016). Principles of cryo-EM single-particle image processing. *Microscopy (Oxford, England) 65*, 57–67.

(*74*) Penczek, P. A. (2010). Resolution measures in molecular electron microscopy. *Methods in enzymology 482*, 73–100.

(*75*) Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods 14*, 793–796.

(*76*) Glaeser, R. M. (2019). How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? *Annual Review of Biophysics 48*, 45–61.

(*77*) Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife 7*, 163.

(*78*) Kühlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science 343*, 1443–1444.

(*79*) Crowther, R. A., DeRosier, D. J., and Klug, A. (1970). The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science 317*, 319–340.

(*80*) Förster, F., and Hegerl, R. In *Cellular Electron Microscopy*; Elsevier: 2007, pp 741–767.

(*81*) Lučić, V., Rigort, A., and Baumeister, W. (2013). Cryo-electron tomography: the challenge of doing structural biology in situ. *The Journal of Cell Biology 202*, 407–419.

(*82*) Fernández, J. J. (2012). Computational methods for electron tomography. *Micron (Oxford, England : 1993) 43*, 1010–1030.

(*83*) Romero-Brey, I., and Bartenschlager, R. (2015). Viral Infection at High Magnification: 3D Electron Microscopy Methods to Analyze the Architecture of Infected Cells. *Viruses 7*, 6316–6345.

(*84*) Bykov, Y. S., Cortese, M., Briggs, J. A. G., and Bartenschlager, R. (2016). Correlative light and electron microscopy methods for the study of virus-cell interactions. *FEBS LETTERS 590*, 1877–1895.

(*85*) Wan, W., and Briggs, J. A. G. (2016). Cryo-Electron Tomography and Subtomogram Averaging. *Methods in enzymology 579*, 329–367.

(*86*) Schur, F. K. M., Hagen, W. J. H., Rumlová, M., Ruml, T., Müller, B., Kräusslich, H.-G., and Briggs, J. A. G. (2015). Structure of the imma ture HIV-1 capsid in intact virus particles at 8.8 A resolution. *Nature 517*, 505–508.

(*87*) Dodonova, S. O., Diestelkoetter-Bachert, P., von Appen, A., Hagen, W. J. H., Beck, R., Beck, M., Wieland, F., and Briggs, J. A. G. (2015). VESICULAR TRANSPORT. A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science (New York, N.Y.) 349*, 195–198.

(*88*) Nickell, S., Förster, F., Linaroudis, A., Net, W. D., Beck, F., Hegerl, R., Baumeister, W., and Plitzko, J. M. (2005). TOM software toolbox: acquisition and analysis for electron tomography. *Journal of structural biology 149*, 227–234.

(*89*) Castaño-Dıez, D., Kudryashev, M., Arheit, M., and Stahlberg, H. (2012). Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *Journal of Structural Biology 178*, 139–151.

(*90*) Hrabe, T., Chen, Y., Pfeffer, S., Cuellar, L. K., Mangold, A.-V., and Förster, F. (2012). PyTom: A python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *Journal of Structural Biology 178*, 177–188.

(*91*) Winkler, H. (2007). 3D reconstruction and processing of volumetric data in cryo-electron tomography. *Journal of Structural Biology 157*, 126–137.

(*92*) Schur, F. K. M., Obr, M., Hagen, W. J. H., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., Sachse, C., Kräusslich, H.-G., and Briggs, J. A. G. (2016). An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science 353*, 506–508.

(*93*) Turoňová, B., Schur, F. K. M., Wan, W., and Briggs, J. A. G. (2017). Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J Struct Biol 199*, 187–195.

(*94*) Hutchings, J., Stancheva, V., Miller, E. A., and Zanetti, G. (2018). Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nature Communications*, 1–8.

(*95*) Hagen, W. J. H., Wan, W., and Briggs, J. A. G. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J Struct Biol 197*, 191–198.

(*96*) Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology 152*, 36–51.

(*97*) Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. A., Potter, C. S., Pulokas, J., and Reilein, A. (2000). Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens. *Journal of Structural Biology 132*, 33–45.

(*98*) Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A., and Scheres, S. H. W. (2015). Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure 23*, 1743–1753.

(*99*) von Appen, A. et al. (2015). In situ structural analysis of the human nuclear pore complex. *Nature 526*, 140–143.

(*100*) Qu, K., Glass, B., Doležal, M., Schur, F. K. M., Murciano, B., Rein, A., Rumlová, M., Ruml, T., Kräusslich, H.-G., and Briggs, J. A. G. (2018). Structure and architecture of immature and mature murine leukemia virus capsids. *Proceedings of the National Academy of Sciences of the United States of America 115*, E11751–E11760.

(*101*) Matrosovich, M., Matrosovich, T., Garten, W., and Klenk, H.-D. (2006). New low-viscosity overlay medium for viral plaque assays. *Virology Journal 3*, 63.

(*102*) Zemlin, F. (1979). A practical procedure for alignment of a high resolution electron microscope. *Ultramicroscopy 4*, 241–245.

(*103*) Kremer, J., Mastronarde, D. N., and McIntosh, J. R. (1996). Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology 116 1*, 71–6.

(*104*) Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife 4*, e06980.

(*105*) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of computational chemistry 25 13*, 1605–12.

(*106*) Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., and Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy 135*, 24–35.

(*107*) Rosenthal, P. B., and Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology 333*, 721–745.

(*108*) Benton, D. J., Nans, A., Calder, L. J., Turner, J., Neu, U., Lin, Y. P., Ketelaars, E., Kallewaard, N. L., Corti, D., Lanzavecchia, A., Gamblin, S. J., Rosenthal, P. B., and Skehel, J. J. (2018). Influenza hemagglutinin membrane anchor. *Proceedings of the National Academy of Sciences of the United States of America 115*, 10112–10117.

(*109*) Kirkpatrick, E., Qiu, X., Wilson, P. C., Bahl, J., and Krammer, F. (2018). The influenza virus hemagglutinin head evolves faster than the stalk domain. *Scientific Reports*, 1–14.

(*110*) Skehel, J. J., and Waterfield, M. D. (1975). Studies on the primary structure of the influenza virus hemagglutinin. *Proceedings of the National Academy of Sciences 72*, 93–97.

(*111*) Rogers, G. N., and Paulson, J. C. (1983). Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology 127*, 361–373.

(*112*) Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology 15*, 690–698.

(*113*) Fontana, J., Cardone, G., Heymann, J. B., Winkler, D. C., and Steven, A. C. (2012). Structural changes in Influenza virus at low pH characterized by cryo-electron tomography. *Journal of Virology 86*, 2919–2929.

(*114*) Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature 371*, 37–43.

(*115*) Chen, X., Liu, S., Goraya, M. U., Maarouf, M., Huang, S., and Chen, J.-L. (2018). Host Immune Response to Influenza A Virus Infection. *Frontiers in Immunology 9*, e1179–13.

(*116*) Lee, P. S., Yoshida, R., Ekiert, D. C., Sakai, N., Suzuki, Y., Takada, A., and Wilson, I. A. (2012). Heterosubtypic antibody recognition of the influenza virus hemagglutinin receptor binding site enhanced by avidity. *Proceedings of the National Academy of Sciences of the United States of America 109*, 17040–17045.

(*117*) Ekiert, D. C. et al. (2012). Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature 489*, 526–532.

(*118*) Lee, P. S., Ohshima, N., Stanfield, R. L., Yu, W., Iba, Y., Okuno, Y., Kurosawa, Y., and Wilson, I. A. (2014). Receptor mimicry by antibody F045-092 facilitates universal binding to the H3 subtype of influenza virus. *Nature Communications 5*, 3614.

(*119*) Impagliazzo, A. et al. (2015). A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science 349*, 1301–1306.

(*120*) Burleigh, L. M., Calder, L. J., Skehel, J. J., and Steinhauer, D. A. (2005). Influenza a viruses with mutations in the m1 helix six domain display a wide variety of morphological phenotypes. *J Virol 79*, 1262–1270.

(*121*) Wasilewski, S., Calder, L. J., Grant, T., and Rosenthal, P. B. (2012). Distribution of surface glycoproteins on influenza A virus determined by electron cryotomography. *Vaccine 30*, 7368–7373.

(*122*) Yang, H., Carney, P. J., Chang, J. C., Guo, Z., Villanueva, J. M., and Stevens, J. (2015). Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology 477*, 18–31.

(*123*) Chen, J., Lee, K. H., Steinhauer, D. A., Stevens, D. J., Skehel, J. J., and Wiley, D. C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell 95*, 409–417.

(*124*) Hoffman, E. A., Frey, B. L., Smith, L. M., and Auble, D. T. (2015). Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes. *Journal of Biological Chemistry 290*, 26404–26411.

(*125*) Kastner, B. et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods 5*, 53–55.

(*126*) Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D 60*, 2256–2268.

(*127*) Xiong, X. et al. (2015). Structures of complexes formed by H5 influenza hemagglutinin with a potent broadly neutralizing human monoclonal antibody. *Proceedings of the National Academy of Sciences of the United States of America 112*, 9430–9435.

(*128*) Migneault, I., Dartiguenave, C., Bertrand, M. J., and Waldron, K. C. (2004). Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques 37*, 790–6– 798–802.

(*129*) Salem, M., Mauguen, Y., and Prangé, T. (2010). Revisiting glutaraldehyde cross-linking: the case of the Arg-Lys intermolecular doublet. *Acta crystallographica. Section F, Structural biology and crystallization communications 66*, 225–228.

(*130*) Moulès, V. et al. (2011). Importance of viral genomic composition in modulating glycoprotein content on the surface of influenza virus particles. *Virology 414*, 51–62.

(*131*) Sugita, Y., Noda, T., Sagara, H., and Kawaoka, Y. (2011). Ultracentrifugation deforms unfixed influenza A virions. *J Gen Virol 92*, 2485–2493.

(*132*) Tran, E. E. H., Podolsky, K. A., Bartesaghi, A., Kuybeda, O., Grandinetti, G., Wohlbold, T. J., Tan, G. S., Nachbagauer, R., Palese, P., Krammer, F., and Subramaniam, S. (2016). Cryo-electron Microscopy Structures of Chimeric Hemagglutinin Displayed on a Universal Influenza Vaccine Candidate. *mBio 7*, e00257.

(*133*) Das, D. K., Govindan, R., Nikić-Spiegel, I., Krammer, F., Lemke, E. A., and Munro, J. B. (2018). Direct Visualization of the Conformational Dynamics of Single Influenza Hemagglutinin Trimers. *Cell 174*, 926–937.e12.

(*134*) Munro, J. B., Gorman, J., Ma, X., Zhou, Z., Arthos, J., Burton, D. R., Koff, W. C., Courter, J. R., Smith, A. B., Kwong, P. D., Blanchard, S. C., and Mothes, W. (2014). Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science 346*, 759–763.

(*135*) Förster, F., Pruggnaller, S., Seybert, A., and Frangakis, A. S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. *J Struct Biol 161*, 276–286.

(*136*) Scheres, S. H. W. (2012). A Bayesian View on Cryo-EM Structure Determination. *J Mol Biol 415*, 406–418.

(*137*) Unverdorben, P., Beck, F., Śledź, P., Schweitzer, A., Pfeifer, G., Plitzko, J. M., Baumeister, W., and Förster, F. (2014). Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America 111*, 5544–5549.

(*138*) Obbineni, J. M., Yamamoto, R., and Ishikawa, T. (2017). A simple and fast approach for missing-wedge invariant classification of subtomograms extracted from filamentous structures. *Journal of Structural Biology 197*, 145–154.

(*139*) Turner, H. L., Pallesen, J., Lang, S., Bangaru, S., Urata, S., Li, S., Cottrell, C. A., Bowman, C. A., Crowe, J. E., Wilson, I. A., and Ward, A. B. (2019). Potent anti-influenza H7 human monoclonal antibody induces separation of hemagglutinin receptor-binding head domains. *PLoS biology 17*, e3000139.

(*140*) Carr, C. M., Chaudhry, C., and Kim, P. S. (1997). Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proceedings of the National Academy of Sciences 94*, 14306–14313.

(*141*) Harris, A., Cardone, G., Winkler, D. C., Heymann, J. B., Brecher, M., White, J. M., and Steven, A. C. (2006). Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences 103*, 19123–19127.

(*142*) Sha, B., and Luo, M. (1997). Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nature structural biology 4*, 239–244.

(*143*) Arzt, S., Baudin, F., Barge, A., Timmins, P., Burmeister, W. P., and Ruigrok, R. W. (2001). Combined results from solution studies on intact influenza virus M1 protein and from a new crystal form of its N-terminal domain show that M1 is an elongated monomer. *Virology 279*, 439–446.

(*144*) Safo, M. K., Musayev, F. N., Mosier, P. D., Zhou, Q., Xie, H., and Desai, U. R. (2014). Crystal structures of influenza A virus matrix protein M1: variations on a theme. *PLOS ONE 9*, e109510.

(*145*) Harris, A., Forouhar, F., Qiu, S., Sha, B., and Luo, M. (2001). The crystal structure of the influenza matrix protein M1 at neutral pH: M1-M1 protein interfaces can rotate in the oligomeric structures of M1. *Virology 289*, 34–44.

(*146*) Chiang, M.-J., Musayev, F. N., Kosikova, M., Lin, Z., Gao, Y., Mosier, P. D., Althufairi, B., Ye, Z., Zhou, Q., Desai, U. R., Xie, H., and Safo, M. K. (2017). Maintaining pH-dependent conformational flexibility of M1 is critical for efficient influenza A virus replication. *Emerging microbes & infections 6*, e108.

(*147*) Shtykova, E. V., Baratova, L. A., Fedorova, N. V., Radyukhin, V. A., Ksenofontov, A. L., Volkov, V. V., Shishkov, A. V., Dolgov, A. A., Shilova, L. A., Batishchev, O. V., Jeffries, C. M., and Svergun, D. I. (2013). Structural Analysis of Influenza A Virus Matrix Protein M1 and Its Self-Assemblies at Low pH. *PLOS ONE 8*, e82431–10.

(*148*) Miyake, Y., Keusch, J. J., Decamps, L., Ho-Xuan, H., Iketani, S., Gut, H., Kutay, U., Helenius, A., and Yamauchi, Y. (2019). Influenza virus uses transportin 1 for vRNP debundling during cell entry. *Nature Microbiology 4*, 578–586.

(*149*) Fujiyoshi, Y., Kume, N. P., Sakata, K., and Sato, S. B. (1994). Fine structure of influenza A virus observed by electron cryo-microscopy. *The EMBO journal 13*, 318–326.

(*150*) Baudin, F., Petit, I., Weissenhorn, W., and Ruigrok, R. W. (2001). In vitro dissection of the membrane and RNP binding activities of influenza virus M1 protein. *Virology 281*, 102–108.

(*151*) Ruigrok, R. W., Barge, A., Durrer, P., Brunner, J., Ma, K., and Whittaker, G. R. (2000). Membrane interaction of influenza virus M1 protein. *Virology 267*, 289–298.

(*152*) Jin, H., Leser, G. P., and Lamb, R. A. (1994). The influenza virus hemagglutinin cytoplasmic tail is not essential for virus assembly or infectivity. *The EMBO journal 13*, 5504–5515.

(*153*) Zhirnov, O. P., Manykin, A. A., Rossman, J. S., and Klenk, H. D. (2016). Intravirion cohesion of matrix protein M1 with ribonucleocapsid is a prerequisite of influenza virus infectivity. *Virology 492*, 187–196.

(*154*) Stauffer, S., Feng, Y., Nebioglu, F., Heilig, R., Picotti, P., and Helenius, A. (2014). Stepwise priming by acidic pH and a high K+ concentration is required for efficient uncoating of influenza A virus cores after penetration. *Journal of Virology 88*, 13029–13046.

(*155*) Boulo, S., Akarsu, H., Ruigrok, R. W. H., and Baudin, F. (2007). Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. *Virus Research 124*, 12–21.

(*156*) Wang, S., Zhao, Z., Bi, Y., Sun, L., Liu, X., and Liu, W. (2013). Tyrosine 132 phosphorylation of influenza A virus M1 protein is crucial for virus replication by controlling the nuclear import of M1. *Journal of Virology 87*, 6182–6191.

(*157*) Ye, Z., Robinson, D., and Wagner, R. R. (1995). Nucleus-targeting domain of the matrix protein (M1) of influenza virus. *J Virol 69*, 1964–1970.

(*158*) Ye, Z., Liu, T., Offringa, D. P., McInnis, J., and Levandowski, R. A. (1999). Association of influenza virus matrix protein with ribonucleoproteins. *J Virol 73*, 7467–7473.

(*159*) Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology 292*, 195–202.

(*160*) Torrisi, M., Kaleel, M., and Pollastri, G. (2018). Porter 5: fast, state-of-the-art ab initio prediction of protein secondary structure in 3 and 8 classess. *bioRxiv*, DOI: 10.1101/289033.

(*161*) Wu, S., and Zhang, Y. (2007). LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research 35*, 3375–3382.

(*162*) Zhang, W., Zheng, W., Toh, Y., Betancourt-Solis, M. A., Tu, J., Fan, Y., Vakharia, V. N., Liu, J., McNew, J. A., Jin, M., and Tao, Y. J. (2017). Crystal structure of an orthomyxovirus matrix protein reveals mechanisms for self-polymerization and membrane association. *Proceedings of the National Academy of Sciences of the United States of America 114*, 8550–8555.

(*163*) Kerviel, A., Dash, S., Moncorgé, O., Panthu, B., Prchal, J., Décimo, D., Ohlmann, T., Lina, B., Favard, C., Decroly, E., Ottmann, M., Roingeard, P., and Muriaux, D. (2016). Involvement of an Arginine Triplet in M1 Matrix Protein Interaction with Membranes and in M1 Recruitment into Virus-Like Particles of the Influenza A(H1N1)pdm09 Virus. *PLOS ONE 11*, e0165421.

(*164*) Helfrich, W. (1973). Elastic properties of lipid bilayers: theory and possible experiments. *Zeitschrift fur Naturforschung. Teil C: Biochemie, Biophysik, Biologie, Virologie 28*, 693–703.

(*165*) Hilsch, M., Goldenbogen, B., Sieben, C., Höfer, C. T., Rabe, J. P., Klipp, E., Herrmann, A., and Chiantia, S. (2014). Influenza A matrix protein M1 multimerizes upon binding to lipid membranes. *Biophys J 107*, 912–923.

(*166*) Höfer, C., Lella, S. D., Dahmani, I., Jungnick, N., Bordag, N., Bobone, S., Huang, Q., Keller, S., Herrmann, A., and Chiantia, S. (2019). Structural determinants of the interaction between influenza A virus matrix protein M1 and lipid membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes 1861*, 1123–1134.

(*167*) Calder, L. J., and Rosenthal, P. B. (2016). Cryomicroscopy provides structural snapshots of influenza virus membrane fusion. *Nature Structural & Molecular Biology 23*, 853–858.

(*168*) Shtykova, E. V., Dadinova, L. A., Fedorova, N. V., Golanikov, A. E., Bogacheva, E. N., Ksenofontov, A. L., Baratova, L. A., Shilova, L. A., Tashkin, V. Y., Galimzyanov, T. R., Jeffries, C. M., Svergun, D. I., and Batishchev, O. V. (2017). Influenza virus Matrix Protein M1 preserves its conformation with pH, changing multimerization state at the priming stage due to electrostatics. *Scientific Reports*, 1–16.

(*169*) Noton, S. L., Medcalf, E., Fisher, D., Mullin, A. E., Elton, D., and Digard, P. (2007). Identification of the domains of the influenza A virus M1 matrix protein required for NP binding, oligomerization and incorporation into virions. *Journal of General Virology 88*, 2280–2290.

(*170*) Elster, C., Larsen, K., Gagnon, J., Ruigrok, R. W., and Baudin, F. (1997). Influenza virus M1 protein binds to RNA through its nuclear localization signal. *Journal of General Virology 78 (Pt 7)*, 1589–1596.

(*171*) Chlanda, P., Mekhedov, E., Waters, H., Sodt, A., Schwartz, C., Nair, V., Blank, P. S., and Zimmerberg, J. (2017). Palmitoylation contributes to membrane curvature in Influenza A virus assembly and hemagglutinin-mediated membrane fusion. *Journal of Virology*, JVI.00947–17–62.

(*172*) Harris, A. K., Meyerson, J. R., Matsuoka, Y., Kuybeda, O., Moran, A., Bliss, D., Das, S. R., Yewdell, J. W., Sapiro, G., Subbarao, K., and Subramaniam, S. (2013). Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies. *Proceedings of the National Academy of Sciences of the United States of America 110*, 4592–4597.