Using Electron Microscopy to Understand Viral Infection and Control: Structural Studies of Murine Norovirus and a Novel Vaccine Platform Technology

Joseph Samuel Snowden

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology
School of Molecular and Cellular Biology
Faculty of Biological Sciences

March 2022

Intellectual Property and Publication Statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapters 3 and 4 of the thesis include data that has appeared in publication as follows:

***Snowden JS**, Hurdiss DL, Adeyemi OO, Ranson NA, Herod MR, Stonehouse NJ (2020). Dynamics in the murine norovirus capsid revealed by high-resolution cryo-EM. PLoS Biol. **18**(3): e3000649.*

The candidate was responsible for cultivating and purifying wtMNV and hsMNV, verification of hsMNV genome sequence and phenotype following purification, cryoEM sample preparation and data collections for wtMNV and hsMNV, and image processing and structural analysis of wtMNV, hiMNV and hsMNV. The candidate also wrote the original draft of the paper and generated figures 1, 3 and 5 as well as all supplementary figures/movies. Other authors also cultivated and purified MNV (OOA, MRH), and were responsible for selection and biochemical characterisation of hiMNV and hsMNV (OOA, MRH), sample preparation and cryoEM data collection of hiMNV (DLH), contributions to image processing of hiMNV (DLH), and supervision (NAR, MRH, NJS). All authors assisted with editing the paper.

Chapter 5 of the thesis includes data that has appeared in publication as follows:

***Snowden JS**, Alzahrani J, Sherry L, Stacey M, Rowlands DJ, Ranson NA, Stonehouse NJ (2021). Structural insight into Pichia pastoris fatty acid synthase. Sci. Rep. **11**: 9773.*

The candidate was responsible for cryoEM sample preparation, data collection, image processing, model building and structural analysis. The candidate produced all figures and wrote the paper. Other authors were responsible for maintenance of P. pastoris cultures (JA, LS), purification of FAS (JA) and supervision (LS, MS, DJR, NAR, NJS). JA, DJR, NAR and NJS assisted with data analysis and NAR and NJS assisted with writing the paper.

Acknowledgements

When I was selecting a laboratory in which to pursue my PhD, I was advised that it was important to choose a good supervisor. Having now worked with my supervisors, Profs Nicola Stonehouse and Neil Ranson, for over four years, I can confidently say I could not have been happier with my choice. Thank you both for the exciting scientific discussions, technical advice and personal encouragement. I have always felt that you were interested in my personal development, as well as my research, and for that I am extremely grateful.

I also want to thank all members of the Stonehouse and Ranson groups, past and present, for advice, support and friendship. I should give special mention to the following: Drs Daniel Hurdiss, Morgan Herod and Oluwapelumi Adeyemi for their brilliant mentorship in EM and virology; Joshua White and Dr Matthew Byrne for fun conversations (and the DSB!); Drs Lee Sherry, Natalie Kingston and Keith Grehan for both lab-related and general support (as well as copious amounts of chocolate from Nat); Drs Jake Mills and Jessica Swanson for their support with the norovirus project; Kaniz Fatema and Jehad Alzahrani, as well as Prof Dave Rowlands, for their support with the VelcroVax project; Corinna Brockhaus for enthusiastic conversations about my passion outside of science – foreign languages!

My studies were greatly supported by the excellent EM facility team at the Astbury Biostructure Laboratory. Structural biologists at Leeds are lucky to have such brilliant facility support.

I am also grateful to other mentors I have had at the University of Leeds and beyond – to Prof Adrian Whitehouse and Dr Belinda Baquero for guidance during my rotation project on m^6A reader proteins, to Drs Juan Fontana and Ryan Seipke for assessing my progress throughout my PhD, and to Drs Gillian Fraser, Owain Bryant, Yongxu Lu and Prof Geoffrey Smith, for their support in research projects during my undergraduate studies and for encouraging my interest in scientific research.

The funding for my PhD work was generously provided by Wellcome, for which I am deeply thankful. Belonging to the Wellcome scheme also placed me within a cohort of fellow Wellcome-funded students, all of whom I feel lucky to have got to know. I must give special mention to my year group – David Klebl, Isabelle Pickles (who can no longer claim she's not a biologist), Eleanor Todd, Sabine Ulamec, Daniel Van and Joshua White (again). Two weeks of intensive studying together in a windowless room is clearly the perfect fire in which to forge friendships!

Finally, I want to thank my family. My mum, Sarah, for being there for me from the beginning (and for insisting she enjoyed reading my papers, even if she didn't understand them!). My brother, Dan, and sister, Holly, for always helping me see the funnier side of life. And the most special thanks of all to my fiancée, Miho, whom I am eternally grateful to have in my life.

Abstract

Viruses have unquestionable impact on our societies. While much progress has been made in our understanding of viral disease, we remain vulnerable, and the continual risk of emerging viral pathogens highlights the importance of continued and more detailed study. Through structural investigations of an infectious virus and a vaccine platform technology, this thesis illustrates how cryogenic electron microscopy (cryoEM) can contribute to understanding and controlling viral infection.

Firstly, using murine norovirus (MNV) as a model system, cryoEM was applied to the study of noroviruses, the most common cause of acute non-bacterial gastroenteritis. Structures were determined for a number of MNV variants, including a thermally stabilised mutant with potential applications to vaccine design. This shed insight into the dynamic nature of MNV capsids, and the importance of this for viral infectivity and stability. Subsequent studies of MNV produced in an alternative cell line revealed a rigidified capsid, provoking questions around the influence of expression system on structure, and the implications for comparative structural analysis. Finally, the molecular interactions of MNV with a bile acid were investigated, and complex image processing techniques were applied to probe non-capsid components of the virion, including the viral genome.

Secondly, structural investigations were carried out into a novel vaccine platform technology. Termed ‘VelcroVax’, the technology comprises virus-like particles formed from hepatitis B virus core antigen, modified to display Affimers for the capture and presentation of target antigens. CryoEM was used to determine the structures of different prototype constructs, including one displaying a viral antigen, Junín virus glycoprotein 1. Based on analysis of these structures, recommendations are made for improving the design and expanding the functionality of VelcroVax constructs. Finally, the importance of considering contaminants within cryoEM datasets is illustrated through structural determination of a yeast fatty acid synthase identified in a VelcroVax dataset.

Table of Contents

List of Tables

List of Figures

Abbreviations

A_{260}/A_{280}	Absorbance at 260 nm/280 nm
ABSL	Astbury Biostructure Laboratory
ACP	Acyl carrier protein
AFIS	Aberration-free image shift
APS	Ammonium persulfate
ARD	Arginine-rich domain
ART	Anti-retroviral therapy
AT	Acetyl transferase
ATCC	American Type Culture Collection
BHK	Baby hamster kidney
BYDV	Barley yellow dwarf virus
CA	Cholic acid
CCD	Charge-coupled device
CD300lf/CD300ld	Cluster of differentiation 300-like family member f/d
CDCA	Chenodeoxycholic acid
cDNA	Complementary DNA
COVID-19	Coronavirus disease 2019
CPE	Cytopathic effect
CRISPR	Clustered regularly interspaced short palindromic repeats
cryoEM	Cryogenic electron microscopy
CTF	Contrast transfer function
DCA	Deoxycholic acid
DED	Direct electron detector
DH	Dehydratase
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded ribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
eIF	Eukaryotic initiation factor
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMDB	Electron Microscopy Data Bank
ER	Enoyl reductase
EtOH	Ethanol
EV	Empty vector
FAS	Fatty acid synthase
FBS	Foetal bovine serum
FEG	Field emission gun
FIB	Focussed ion beam

FMDV	Foot-and-mouth disease virus
FMN	Flavin mononucleotide
FSC	Fourier shell correlation
FUT2	Fucosyltransferase 2
GCA	Glycocholic acid
GCDCA	Glycochenodeoxycholic acid
GFP	Green fluorescent protein
GI	Gastrointestinal
GP1	Glycoprotein 1
gRNA	Genomic ribonucleic acid
HBcAg	Hepatitis B virus core antigen
HBGA	Histo-blood group antigen
HBsAg	Hepatitis B virus surface antigen
HBV	Hepatitis B virus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	High-income country
HIE	Human intestinal enteroid
hiMNV	Heat-inactivated murine norovirus
HIV	Human immunodeficiency virus
hsMNV	Heat-stable murine norovirus
HuNV	Human norovirus
ID_{50}	Median infectious dose
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
JAM-A	Junctional adhesion molecule A
JUNV	Junín virus
KS	Ketoacyl synthase
LB	Luria-Bertani
LCA	Lithocholic acid
lDDT	Local distance difference test
LMIC	Low- and middle-income country
mAb	Monoclonal antibody
Mcl-1	Induced myeloid leukaemia cell differentiation protein
MHC	Major histocompatibility complex
MIR	Major immunodominant region
ML	Maximum likelihood
MNV	Murine norovirus
MNV52	Murine norovirus selected for thermal resistance at 52°C
MOPS	3-(*N*-morpholino)propanesulfonic acid
MPT	Malonyl/palmitoyl transferase
mRNA	Messenger ribonucleic acid
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2*H*-tetrazolium inner salt
Nb	Nanobody

NHS	National Health Service
NIBSC	National Institute for Biological Standards and Control
NMR	Nuclear magnetic resonance
NS#	Non-structural protein #
NTA	N-terminal arm
NTZ	Nitazoxanide
NV	Norwalk virus
ORF	Open reading frame
P domain	Protruding domain
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PaSTRy	Particle stability thermal release
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDBeFold	Protein Data Bank in Europe Fold
PDBePISA	Protein Data Bank in Europe Protein Interfaces, Surfaces and Assemblies
PEG8000	Polyethylene glycol 8000
PFA	Paraformaldehyde
PPT	Phosphopantetheine transferase
PRR	Pathogen-recognition receptor
PSF	Point spread function
PTM	Post-translational modification
RBD	Receptor-binding domain
RdRp	Ribonucleic acid-dependent ribonucleic acid polymerase
RF	Replicative form
RHDV	Rabbit haemorrhagic disease virus
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RT-PCR	Reverse-transcription polymerase chain reaction
RT-qPCR	Reverse-transcription quantitative polymerase chain reaction
S domain	Shell domain
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SGD	Stochastic gradient descent
sgRNA	Subgenomic ribonucleic acid
SPA	Single particle analysis
ssRNA	Single-stranded ribonucleic acid
STNV	Satellite tobacco necrosis virus
SUMO	Small ubiquitin-like modifier
TBE	Tris-Borate-EDTA

TCA	Taurocholic acid
TCDCA	Taurochenodeoxycholic acid
$TCID_{50}$	Median tissue culture infectious dose
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
UCSF	University of California San Francisco
UTR	Untranslated region
VF1	Virulence factor 1
VLP	Virus-like particle
VP1	Major capsid protein
VP2	Minor capsid protein
VPg	Viral protein genome-linked
WHO	World Health Organisation
WT	Wild type
wtMNV	Wild-type murine norovirus
wtMNV:LCA	Wild-type murine norovirus in complex with lithocholic acid
wtMNV.BV2	Wild-type murine norovirus cultivated in BV-2 cells
wtMNV.RAW	Wild-type murine norovirus cultivated in RAW264.7 cells
YPD	Yeast extract-peptone-dextrose

Initials

AW	Alexander Watson
DH	Dr Daniel Hurdiss
JA	Jehad Alzahrani
KF	Kaniz Fatema
KG	Dr Keith Grehan
MH	Dr Morgan Herod
OA	Dr Oluwapelumi Adeyemi

1. Introduction

For centuries, humans have endeavoured to extend the limits of our perception. The invention of the telescope in the early 1600s catalysed a breakthrough in our understanding of the solar system, and the development of more powerful equipment like the Hubble Space Telescope has allowed us to see deeper into the universe, and even into our distant past. But there is also an imperceptible cosmos within ourselves, made accessible to us by the development of the microscope. The light microscope revealed a world of cells and microorganisms, and with electron microscopes we gained the ability to probe the inner workings of organelles and visualise viruses, orders of magnitude smaller than mammalian cells. The subsequent development of cryogenic electron microscopy (cryoEM) took us even further, allowing us to visualise proteins, diverse molecular machines that drive cellular function, at a sub-nanometre resolution – something that was previously only possible using X-ray crystallography. Today, cryoEM continues to yield valuable insight across a wide range of scientific fields, from materials science and physics, to molecular biology and pathology.

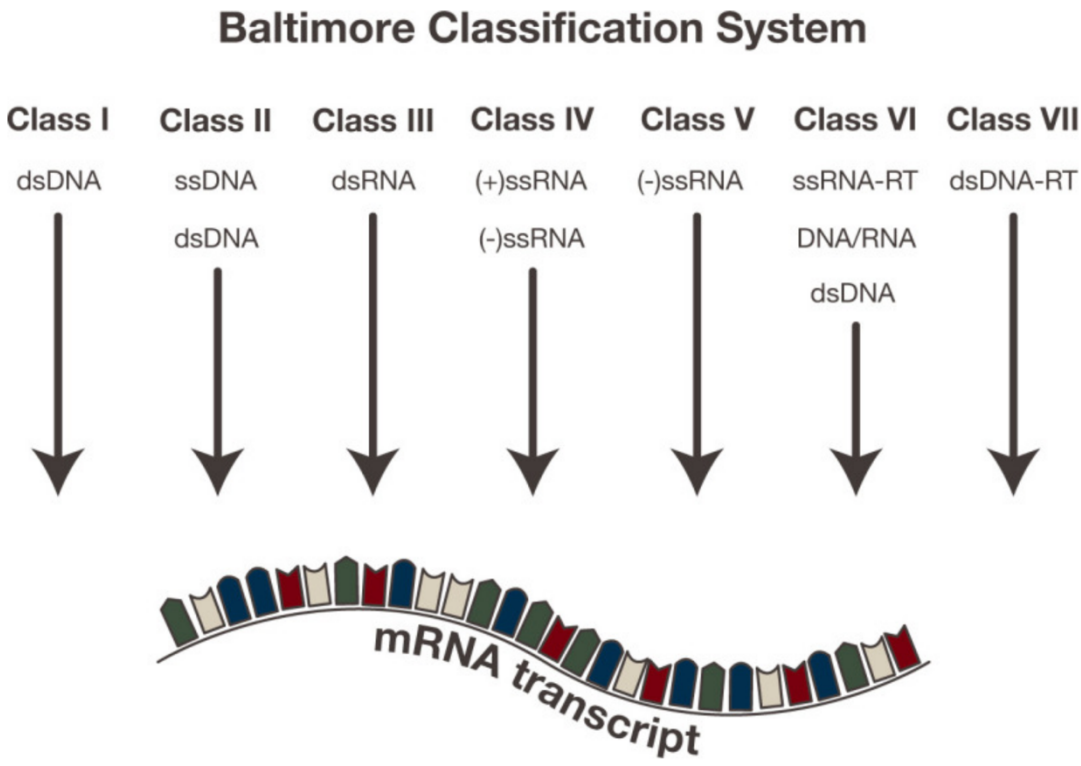
This thesis will demonstrate how cryoEM can contribute to our efforts to combat viral disease. Firstly, cryoEM is applied to the study of an infectious norovirus, revealing the dynamic nature of the norovirus capsid and its importance for viral stability and infectivity. This is followed by studies into the inner contents of the virion and the molecular interactions between the virus and host factors, critical for initiating infection. In addition to extending our understanding of the molecular basis of viral infection, which is key for developing successful antiviral agents, cryoEM can also contribute directly to the development of control measures against viral disease. This is demonstrated in the penultimate chapter, which describes structural studies of a novel vaccine platform technology based on a virus-like particle scaffold system.

1.1. Viruses and their impact

As obligate intracellular parasites, viruses are essentially molecular machines that have evolved to carry out two fundamental tasks: (i) protect the viral genome until the virus is within a permissive environment for replication (i.e., a host cell), then (ii) replicate the viral genome. To achieve this, all viruses share common elements, including genetic material in the form of RNA or DNA and a protective proteinaceous coat. However, while the fundamental tasks above are universal, there is considerable diversity in precisely how viruses fulfil them. Given that viruses must produce mRNA at some point in their infectious cycle, a particularly useful framework for categorising this diversity is the Baltimore classification system (Figure 1.1), which assigns viruses to seven different groups based on their means of mRNA synthesis (Baltimore, 1971). Alongside this conceptual system, an 'official' taxonomic system describes the classification of viruses into taxa (such as species and families) based on their evolutionary relatedness (International Committee on Taxonomy of Viruses Executive Committee, 2020).

For many viruses, a consequence of their replication within a host organism is illness in that host, either directly or indirectly (for example, as a result of the host immune response). Indeed, every Baltimore classification group contains viruses that cause disease in humans and other animals. The impact of viral disease on global health is impossible to ignore. At the time of writing (January 2022) and just over two years since the first reported case, coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is reportedly responsible for over 5.5 million deaths worldwide (WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/table). Noroviruses, the focus of Chapters 3 and 4 in this thesis, are thought to cause over 200,000 deaths every year in young children due to severe dehydration from gastroenteritis (Patel *et al.*, 2008). Viruses are also responsible for significant economic disruption. For example, noroviruses alone are estimated to impose a global economic burden of $60 billion/year, primarily as a result of productivity losses (Bartsch *et al.*, 2016). The impact of viral disease has even been a pivotal factor in historical

incidents with profound geopolitical ramifications, illustrated by the suspected link between yellow fever epidemics in North America at the end of the 18th century and the Louisiana Purchase in 1803 (Marr and Cathey, 2013).



1.1. The Baltimore classification system. *Viruses can be categorised into seven classes based on their means of mRNA synthesis. Each class is listed along with the nature of the viral genome and any intermediate forms generated prior to mRNA synthesis. Note that for members of Class IV, the viral genome itself can act as mRNA and be translated directly. ss – single-stranded; ds – double-stranded; DNA – deoxyribonucleic acid; RNA – ribonucleic acid; (+) – positive-sense; (-) – negative-sense; RT – with reverse transcriptase. Figure adapted from (Greene and Reid, 2013), published under a CC BY 4.0 licence.*

Fortunately, knowledge gained through scientific study can inform public health policies and the development of pharmaceuticals to mitigate the detrimental impact viruses can have on our societies. Anti-retroviral therapy (ART) has transformed human immunodeficiency virus (HIV) infection into a manageable condition, such that improving *access* to proper diagnosis and

treatment is now a more pressing concern. The combination of an effective vaccine with a policy of 'ring vaccination' was instrumental in the successful eradication of smallpox, declared by the WHO in 1980. More recently, eradication of the cattle disease rinderpest was achieved, in no small part due to the combined use of vaccination and prevention measures (Morens *et al.*, 2011). Polio eradication now also appears within reach, and to this end, efforts to develop improved, next-generation vaccines are ongoing in the Stonehouse laboratory and elsewhere (Marsian *et al.*, 2017; Viktorova *et al.*, 2018; Xu *et al.*, 2019; Sherry *et al.*, 2020).

This knowledge base is the result of decades of investigation across a broad range of scientific disciplines. Structural biology, the focus of this thesis, has made countless contributions to our understanding, and continues to offer new insight into viral infection and control.

1.2. Structural biology and its contributions to virology

Structural biology is concerned with the molecular structure of biological molecules and how this dictates their function. The first three-dimensional structure of a protein was solved in 1958 using X-ray crystallography, a technique that remains a cornerstone of the field (Kendrew *et al.*, 1958). In protein crystallography, X-rays incident on crystals of a protein or complex of interest are diffracted in a manner dependent on the precise arrangement of atoms within the crystal. Crystals are regular arrays of a minimal repeating unit (termed a 'unit cell'). As such, they act as natural amplifiers of the diffraction signal from the unit cell, producing measurable X-ray diffraction patterns from which structure factors can be determined and a map of electron density calculated. X-ray crystallography is useful for determining molecular structures at atomic resolution, though can present significant experimental difficulties. For example, many biological targets can be extremely challenging to crystallise, and even if a crystal is obtained and a diffraction dataset collected, additional work is needed to generate the phase information required for structure determination (e.g., through molecular replacement, isomorphous replacement or anomalous dispersion) (reviewed in (Taylor, 2010)). Aside

from X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy has proven useful for determining the molecular structures of proteins, and can also provide information on their dynamics (reviewed in (Marion, 2013)).

These techniques have provided much insight into virus structure. Indeed, the first near-atomic resolution structures of icosahedral virus capsids were determined through X-ray crystallography (Harrison *et al.*, 1978; Abad-zapatero *et al.*, 1980). NMR spectroscopy has been especially useful for investigating small viral proteins that are disordered, difficult to crystallise, or particularly dynamic. For example, NMR spectroscopy was used to map transient structural motifs and stabilising interactions for an intrinsically disordered fragment of the NS5A protein of hepatitis C virus (Feuerstein *et al.*, 2012). More recently, the technique was applied to study the structure of the main viral protease ($3CL^{pro}$) of SARS-CoV-2 (Cantrelle *et al.*, 2021).

This thesis focusses on electron microscopy (EM), a mainstay of structural virology for decades. Analyses of tobacco mosaic virus with an electron microscope gave early insight into virus morphology (Kausche *et al.*, 1939), and the technique has also long been useful for clinical diagnosis – for example, differentiating between poxvirus and herpesvirus infections in patients presenting with vesicular skin rashes (Long *et al.*, 1970). Improvements to hardware, software and sample preparation, including the development of cryoEM, have vastly increased the level of detail with which we can visualise viruses, and consequently, our understanding of virus structure has improved.

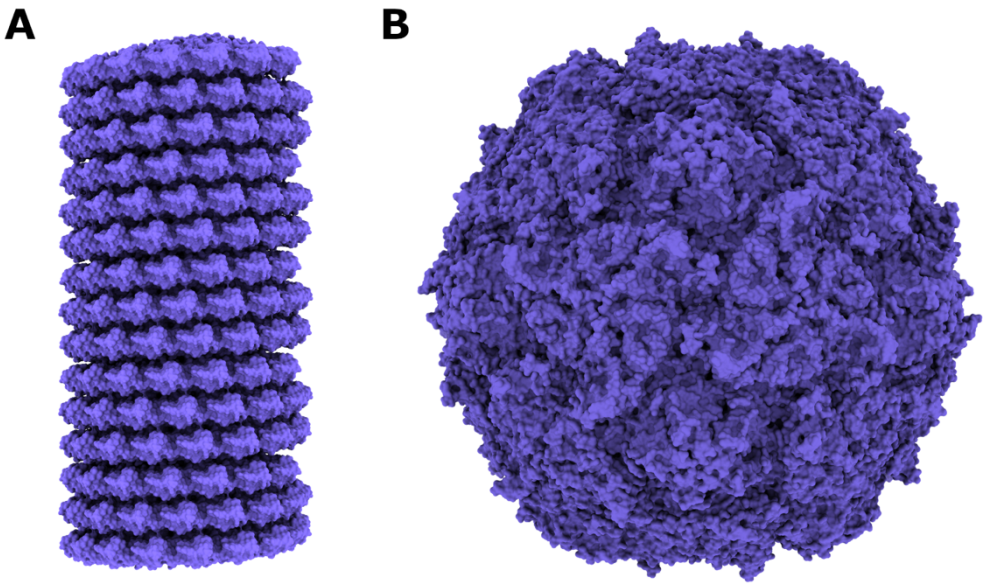
Most viruses have capsids that display either helical or icosahedral symmetry as a means of maximising genetic economy (Figure 1.2). While the length of helical capsids is often determined by the size of the RNA/DNA that they accommodate, icosahedral capsids are enclosed structures formed from sixty repeating ‘asymmetric units’, and thereby have a limited volume for accommodation of the viral genome. Some icosahedral viruses simply maintain small genomes, such as satellite tobacco necrosis virus, satellite panicum mosaic virus and mature cricket parvovirus (Jones and Liljas, 1984;

Ban and McPherson, 1995; Meng *et al.*, 2013). However, others make use of what Caspar and Klug termed 'quasi-equivalence' (Caspar and Klug, 1962). Here, rather than a single protein forming the asymmetric unit of an icosahedron, asymmetric units are formed from several proteins, each in a slightly different conformation related to its precise chemical environment (Figure 1.3A). This allows for an increase in the size of the asymmetric unit, thus increasing the distance between the vertices of the icosahedron and expanding the internal volume of the capsid. Critically, the individual proteins within the asymmetric unit can be copies of the same structural protein, such that no additional coding capacity is required in the viral genome. A useful way to think about icosahedral capsids in relation to 'quasi-equivalence' is through their triangulation (T) number, which describes the relative positioning of five-fold axes according to equation (1), where h and k give the number of unit steps (in opposing directions) separating adjacent icosahedral five-fold axes, as illustrated in Figure 1.3B.

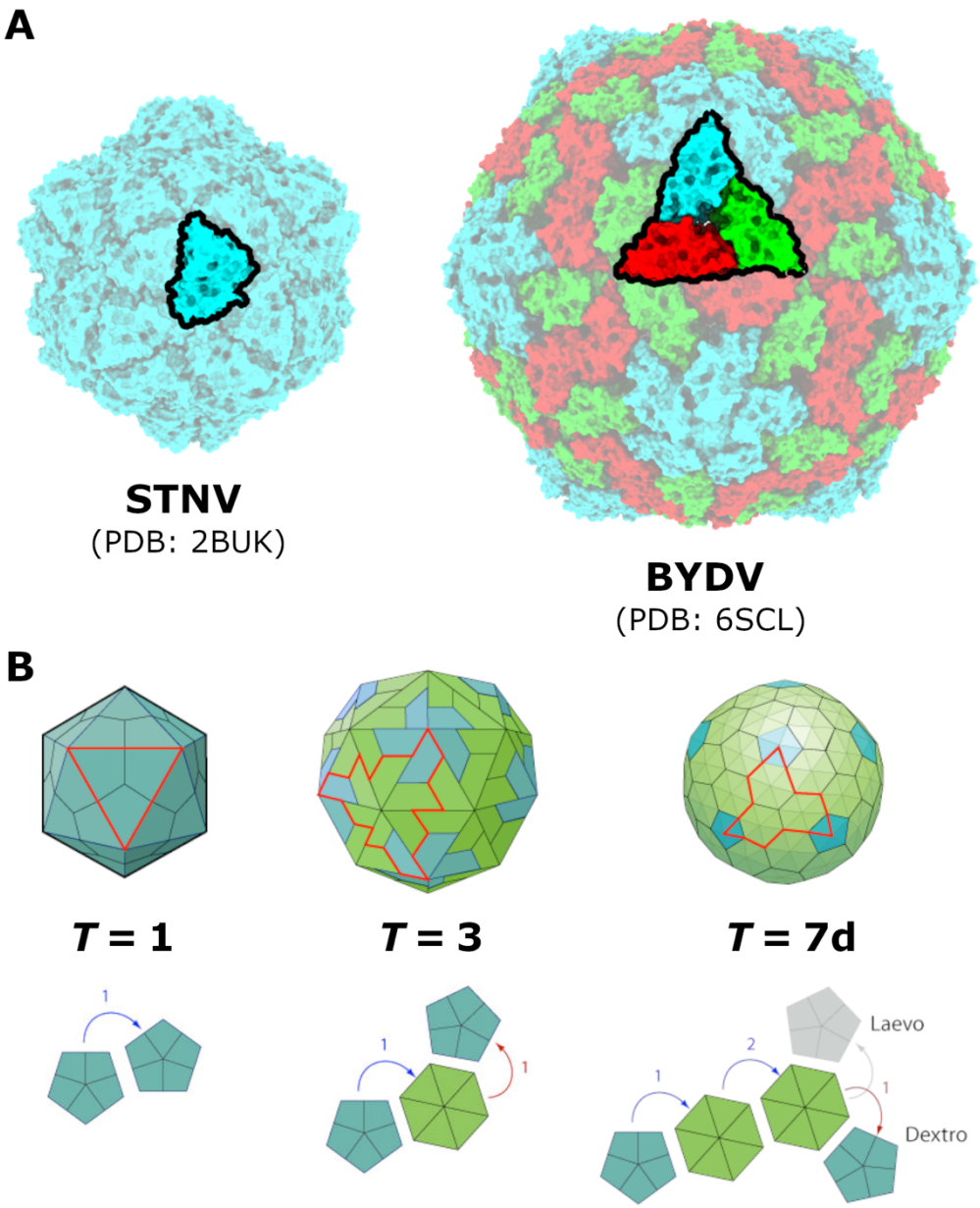
$$T = h^2 + hk + k^2 \qquad (1)$$

The contributions of structural biology to virology stretch beyond fundamental knowledge like that described above – there are myriad examples of structural data being used to support the development and improvement of control measures. Picornavirus capsid structures revealed small structural differences in a druggable hydrophobic pocket within VP1, informing the design of the antiviral agent pleconaril to have an optimised pocket fit for most picornaviruses (Pevear *et al.*, 1999). Structural data was also critical for the rational design of a stabilised virus-like particle vaccine candidate for foot-and-mouth disease virus (FMDV). Here, FMDV capsid structures were used to guide the mutagenesis of a histidine at an inter-pentamer interface to a cysteine, resulting in capsid-stabilising disulphide bonds (Porta *et al.*, 2013). Even more recently, cryoEM was used to determine the structure of the recombinant adenovirus used in the SARS-CoV-2 vaccine, ChAdOx1, and

postulate a potential molecular mechanism for extremely rare observations of clotting disorders following vaccination (Baker *et al.*, 2021).



1.2. Symmetry in virus structures. (A) *A section of the tobacco mosaic virus capsid (PDB: 6SAE, (Weis et al., 2019)), displaying helical symmetry.* ***(B)*** *The poliovirus capsid (PDB: 1HXS, (Miller et al., 2001)), displaying icosahedral symmetry. Structures are not shown to scale.*



1.3. Quasi-equivalence and triangulation numbers. (A) *Molecular surfaces generated from structures of satellite tobacco necrosis virus (STNV; PDB: 2BUK, (Jones and Liljas, 1984)) and barley yellow dwarf virus (BYDV; PDB: 6SCL, (Byrne et al., 2019)). In each case, an asymmetric unit is highlighted. BYDV displays quasi-equivalency; quasi-conformers are shown in different colours.* ***(B)*** *Calculation of triangulation (T) numbers. Steps in the h direction are shown with blue arrows, steps in the k direction are shown with red arrows. Panel adapted from ViralZone, Swiss Institute of Bioinformatics (https://viralzone.expasy.org/) licensed under a CC BY 4.0 licence. Structures not shown to scale.*

1.3. Principles of transmission electron microscopy

1.3.1. Image formation in transmission electron microscopy

The fundamental principle underlying EM is that electrons display wave-like properties with a much shorter wavelength (λ) than visible light, defined by their mass (m) and velocity (v) according to equation (2) (the de Broglie equation), where h is Planck's constant.

$$\lambda = \frac{h}{mv} \qquad (2)$$

As such, the de Broglie wavelength of an electron travelling through the column of an electron microscope is on the order of picometers, hundreds of thousands of times shorter than the wavelength of visible light, permitting a much higher resolution to be obtained. In the column, electrons can interact with the sample to be scattered elastically (i.e., where only the direction of electron, and not its kinetic energy, is altered) or inelastically (i.e., where energy from the electron is transferred to the sample). Notably, because energy is transferred to the sample during inelastic scattering, these events lead to radiation damage. By recording the incidence of electrons that have passed through the sample, images (electron micrographs) can be generated that encode information about the distribution of scattering centres within the specimen.

The contrast in these micrographs is generally created through two separate phenomena. Amplitude contrast refers to the reduction in transmitted electrons through regions of high Coulomb potential, as a result of absorption and scattering of electrons to the extent that they do not reach the detector. Conversely, phase contrast is generated as a result of interference between electron waves. While both components contribute to contrast generation in EM, their relative importance is dependent on the method of sample

preparation (described in more detail in Section 1.4), with the use of heavy metal stains increasing the relative contribution of amplitude contrast.

1.3.2. The transmission electron microscope

While there are variations in the precise design of different models of transmission electron microscope (TEM), TEMs are essentially composed of the following key components: an electron source, a series of electromagnetic lenses, a specimen holder, and an electron detector. Electrons for illumination of the sample are emitted from the electron source, usually either a tungsten filament, lanthanum hexaboride (LaB_6) crystal, or field emission gun (FEG), with the latter providing the brightest and most coherent beam (Orlova and Saibil, 2011). These emitted electrons are accelerated (typically using voltages of 100 – 300 kV) to pass through a vacuum column and series of lenses.

Electromagnetic lenses are, in essence, coils of wire through which a current is passed, generating a magnetic field that acts to alter that path of electrons within the microscope column. First in the series is a condenser lens. This parallelises the electron beam to illuminate the sample, which typically sits within the magnetic field generated by the next lens – the objective lens. This magnifies the image prior to further magnification by intermediate and projector lenses, ultimately projecting the image onto the detector. Alongside lenses, several apertures sit within the column to limit the size of the electron beam passing through. In the case of the objective aperture, this serves to increase amplitude contrast by blocking widely scattered electrons.

Several different types of detectors have been used successfully in TEMs. Photographic film was widely used for its fine spatial resolution and large detection area, however this approach is low-throughput (each individual film must be developed, then digitised) and has largely been superseded by digital detectors, such as charge-coupled devices (CCDs) and direct electron detectors (DEDs). CCD detectors record electrons indirectly; incident electrons encounter a scintillator layer that converts them to photons for

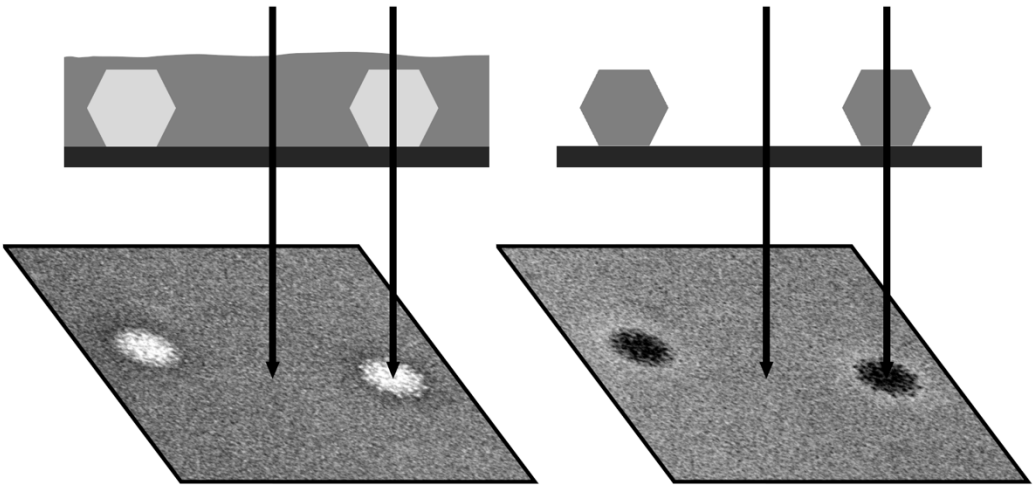
channelling through fibre optics into a CCD array. Photons that reach the CCD chip generate electrical charge, which is read as output. DEDs, which have no need for a scintillator (and thus eliminate the noise generated during this step for CCD detectors), are largely credited with bringing about the so-called 'resolution revolution' in cryoEM (Kühlbrandt, 2014). This is partly due to their fast readout, allowing for each exposure to be recorded across a number of separate frames, which allows for computational correction of any blurring from beam-induced motion (described in more detail in Section 1.5.1). In general, DEDs can be operated in one of two imaging modes – integrating mode, whereby the charge generated by incident electrons is summed over the course of a full exposure, and counting mode, which counts individual electron events. While integrating mode permits a higher dose rate and shorter exposures, there is a reduced signal-to-noise ratio because the amount of energy deposited by individual electrons can differ (Biyani *et al.*, 2017). Counting mode is slower but gives an improved signal-to-noise ratio, and with other developments that improve throughput, such as aberration-free image shift (AFIS), the speed of data collection is now rarely an issue.

1.4. Negative stain and cryogenic electron microscopy

For biological specimens, the column of an electron microscope is an extremely hostile environment. A sample would rapidly dehydrate in the vacuum, and exposure to the electron beam would cause significant radiation damage. In order to protect the sample from this environment and avoid structural degradation, there are two common sample preparation approaches that can be adopted – heavy metal staining (usually negative staining), or vitrification.

With heavy metal staining, the sample is applied to a thin carbon-backed metal grid, then coated in a stain solution – typically, 1 – 2% uranyl acetate. The stain layer is both protective and able to generate amplitude contrast for image formation. Negative staining is generally preferred to positive staining, and refers to the indirect visualisation of the sample via the displacement of stain (Figure 1.4). The maximum attainable resolution from negative stain is ~18 Å

(Scarff *et al.*, 2018), so the technique is not suitable for high-resolution structure determination, but it remains invaluable for low-resolution studies (for example, probing complex disassembly or large morphological changes) and for assessing sample quality.



1.4. Negative and positive staining. *Schematic illustrating negative and positive staining. In negative staining (left), the grid surface is coated in a layer of heavy metal stain and particles appear light against a dark background. In positive staining (right), stain is taken up by individual particles which appear dark on a lighter background. Note that the example micrograph for positive staining is simulated, generated by inverting the contrast of negative stain data, and therefore is not genuine positive stain data.*

For cryoEM, the sample is suspended in a thin layer of vitreous ice, which offers protection from the column environment. To form this vitreous ice layer, the sample is applied to a metal grid (with or without thin carbon backing), blotted to remove excess liquid, and then cooled extremely quickly (typically on the order of 10^6 K/s) by plunging the grid into liquid nitrogen-cooled liquid ethane at $-178°C$. The thickness and quality of the ice layer must be precisely controlled to ensure the sample can be visualised in the microscope, and regardless, amplitude contrast is extremely limited. As such, most contrast is derived from phase contrast, which usually requires the introduction of defocus.

1.5. Computational processing of cryoEM data

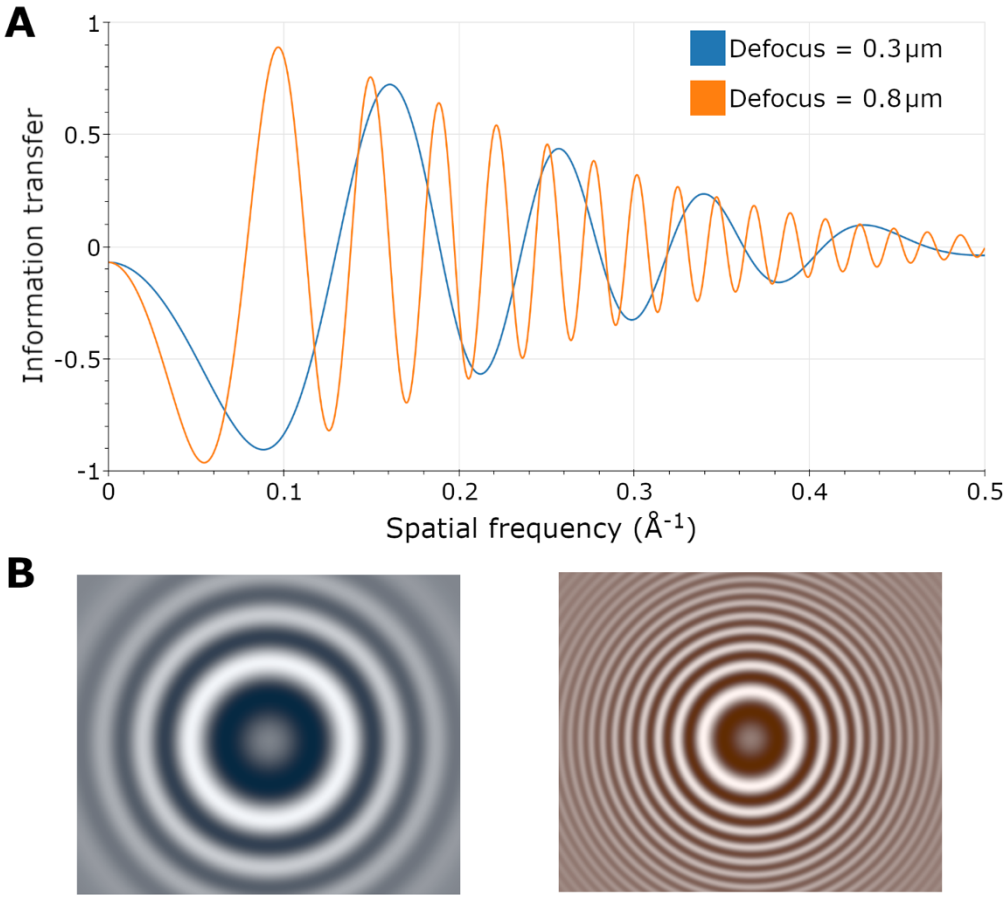
While there are multiple imaging approaches that can be applied with cryoEM (including tomography and electron crystallography), the vast majority of high-resolution structures are determined through single particle analysis (SPA). Here, micrographs are collected to generate a dataset containing many copies of the target particle in different orientations. Through computational image processing, individual images of the target can be classified, combined and reconstructed to yield a high-resolution three-dimensional structure.

1.5.1. Pre-processing

Following data collection, micrographs are first pre-processed to correct for blurring of signal as a result of stage drift and beam-induced motion (Brilot *et al.*, 2012). Using software packages such as MotionCor2 (Zheng *et al.*, 2017), individual frames captured in each micrograph 'movie' can be re-aligned to compensate for any motion, and further, frames can be weighted by accumulated dose to reduce the impact of radiation damage. Motion correction is known to enhance the final resolution of the reconstruction (Campbell *et al.*, 2012).

Subsequently, each motion-corrected micrograph is taken forward for estimation of the contrast transfer function (CTF). The CTF describes how the information conveyed in each image varies by spatial frequency, and is essentially a description of the point spread function (PSF), a consequence of imperfect microscope optics, in Fourier space (Figure 1.5). Typically, CTF plots resemble an oscillating curve with increasing frequency and decreasing amplitude as spatial frequency increases. As described in Section 1.4, most contrast in cryoEM images is generated through phase contrast, which benefits from the introduction of defocus. Increasing the defocus applied during an exposure improves the low-resolution contrast in the image (critical for particle picking and alignments), but also increases the frequency of oscillations in the CTF, and ultimately reduces the high-resolution information

contained in the image. Furthermore, for any given defocus value, there are multiple spatial frequencies where the CTF value is zero – i.e., the image contains no information at that particular resolution. To compensate for this, micrographs are usually collected at a range of different defocus values, such that reconstructions (generated using particles taken from many micrographs) contain information across a comprehensive range of spatial frequencies. In practical terms, software such as Gctf and CTFFIND4 can be used to estimate the CTF for each image, for later correction (e.g., through phase flipping) (Rohou and Grigorieff, 2015; Zhang, 2016).



1.5. The contrast transfer function. (A) *Simulated one-dimensional CTF plot for an image recorded at 0.3 µm (blue) or 0.8 µm (orange) defocus.* ***(B)*** *Simulated two-dimensional CTF plots corresponding to the curves in (A). Simulated CTFs were generated using a web app described in (Jiang and Chiu, 2001).*

Once micrographs have been motion-corrected and their CTFs estimated, individual particle images must be extracted. This can be done manually, by visually inspecting each micrograph in turn and recording the coordinates of particles, or automatically, either using cross-correlation approaches (with or without references) or deep learning object detection (Scheres, 2015; Zivanov *et al.*, 2018; Wagner *et al.*, 2019).

1.5.2. Classification

Once individual particles have been extracted from micrographs, classification can be performed. This process has multiple aims: (i) to remove poor quality particles (i.e., those that are in poor ice, show denaturation or disassembly, or are not monodisperse), (ii) to remove erroneously extracted images, such as those that correspond to contamination or background carbon, and (iii) to separate particles displaying conformational or organisational heterogeneity into distinct groups. A popular approach to this is 'maximum likelihood' (ML), as is implemented in Relion (Scheres, 2010, 2012). Here, rather than assigning each particle to a single class, every particle is assigned probability values for all classes, and these values are used to weight the relative contribution of the particle to each 'class average'. This helps to reduce the impact of erroneous assignment of particles, which arises from the low signal-to-noise ratio of individual particles in cryoEM datasets.

Classification is usually done in both two and three dimensions. For 2D classification in Relion, extracted particles are grouped based on their similarity after rotational and translational alignment. References are not provided by the user to limit the introduction of bias – instead, references are automatically generated by averaging random subsets of the total particle stack prior to any alignment (Scheres, 2010). Initially, particles are randomly assigned probability weightings to different classes, leading to subtle differences in class averages. Alignment and assignment of probability values are repeated iteratively until there are clear distinctions in class averages and classes stabilise. Good quality classes – particularly those with visible secondary structural features in the class average – are taken forward for

further processing, whereas particles without high probability values for selected classes are discarded. The precise number of classes is specified by the user, typically aiming to give enough ‘space’ to separate out heterogeneity within the dataset, but not so much that particle weightings are spread too finely between classes causing alignments to become more difficult. 3D classification in Relion is similar, although typically a much smaller number of classes is specified (e.g., 2 – 5) and one or several reference models are provided by the user, usually heavily low-pass filtered to mitigate the potential to introduce model bias.

By classifying the data several times in series, the homogeneity of the particle stack is improved, increasing the likelihood of attaining a high-resolution reconstruction.

1.5.3. Three-dimensional reconstruction

Once the particle stack has been classified sufficiently, a three-dimensional initial model (taken from a previous structure or generated from the data *de novo*) can be refined, hopefully yielding a high-resolution reconstruction. Since the reconstruction is guided by a reference, the process is susceptible to model bias, where the alignment of noise to high-resolution features present in the model can yield a reconstruction with those same features, even in the absence of experimental data to support this. As is the case for 3D classification, the reference is first heavily low-pass filtered to mitigate this risk. In Relion, the particle stack is then split into independent halves, with each used to refine separate reconstructions iteratively. At each iteration, the correlation of the two half maps across different spatial frequencies in Fourier space is compared. Since each half map is refined independently, the ‘gold standard’ Fourier shell correlation (FSC) criterion can be used to calculate a nominal resolution value, taken as the inverse of the spatial frequency where the FSC between the half maps drops below 0.143 (Rosenthal and Henderson, 2003; Scheres and Chen, 2012). When the resolution and estimated accuracy of angular assignments stops improving between iterations, the refinement is

deemed to have ‘converged’, the half-sets are combined, and a final reconstruction is generated using all of the data.

Importantly, when the symmetrical organisation of the target is known, symmetry can be imposed during the refinement to make use of the redundant information contained within each particle. While this usually has the effect of increasing the resolution of the reconstruction, it is important to consider that any asymmetric features (e.g., the genome of an icosahedral virus) would be incorrectly averaged, leading to loss of density or artefactual density.

Following refinement, the quality of the map can be further improved through a number of approaches. A solvent-excluding mask can be applied to exclude noise surrounding the map and eliminate its impact on the FSC and calculation of a nominal resolution. The map can also be ‘sharpened’. Image contrast is down-weighted at high resolutions due to experimental (such as radiation damage and optical aberrations) and computational (such as inaccurate angular assignment) factors, such that low-resolution signal dominates the reconstruction, which may appear smooth and absent of features that would be expected from the nominal resolution value. To correct this, temperature (B)-factor sharpening restores the weighting of high-resolution data in the map (Rosenthal and Henderson, 2003).

More recently developed tools permit further improvement in resolution following reconstruction and sharpening. With CTF refinement, defocus and aberrations can be estimated for individual particles based on a 3D reference structure (Zivanov *et al.*, 2018). Furthermore, having obtained a reference map along with per-particle angular assignments and defocus values, motion tracks can be fitted for each particle (Zivanov *et al.*, 2019). Since both of these processes are more effective when the 3D reference is better resolved, an iterative loop of CTF refinement, Bayesian polishing and 3D refinement with sharpening can be performed, with sizeable improvements in map resolution.

While a nominal resolution is calculated for the final reconstruction, in actuality the resolution varies across the map, for example, as a result of regional

heterogeneity. To account for this and locally filter the reconstruction, local resolution can be estimated. In Relion, this can be achieved by essentially 'sampling' each region of the map with a small spherical mask.

1.5.4. Model building

Following attainment of a high-resolution density map, an atomic model can be built to facilitate structural analysis. This process usually starts with the generation of a homology model based on the primary sequence of the target protein, which is then rigid-body fitted to the density map, though it is also possible to build the model into density *de novo*. Rigid-body fitting usually involves a translational and rotational search, with the aim of maximising the correlation of the experimental density and a hypothetical density map based on the atomic coordinates.

Following positioning of atomic coordinates within the map, comprehensive refinement can be performed to improve both the fit of the model to the map, and atomic geometry. This is often achieved through cyclical iterations of (i) visual inspection and supervised refinement of the model in software such as Coot, and (ii) automated refinement of the entire model, for example, using the real space refine tool of Phenix (Adams *et al.*, 2010; Emsley *et al.*, 2010). Model validity is assessed through a range of statistics, including (but not limited to) model-to-map cross-correlation, the prevalence of steric clashes, and the proportion of Ramachandran and rotameric outliers.

1.6. Aims of the doctoral research project

Through two distinct structural projects, this thesis aims to demonstrate how EM can contribute to our understanding of viruses and guide the development of viral control strategies.

Firstly, cryoEM was used to characterise a number of variants of murine norovirus, and in doing so attempt to explain why different particle morphologies have been observed for different norovirus genotypes.

Structural studies were then extended to visualise the interaction between murine norovirus and a key molecular co-factor, a potentially useful resource for the rational design of anti-noroviral pharmaceuticals. A number of non-standard image processing approaches were also used in an attempt to resolve non-capsid components of the infectious virion.

Secondly, a novel antigen display system under development as a vaccine platform technology in the Stonehouse group was characterised by cryoEM, with the aim that these structures would inform the development of improved construct designs.

2. Materials and methods

2.1. Materials

2.1.1. Reagents, buffers and media

The composition of the buffers and media used for the work described in this thesis are provided in Table 2.1.

Table 2.1. Compositions of buffers and media.

Buffer/Medium	Composition
Supplemented cell culture medium	High-glucose Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich), 10% (v/v) foetal bovine serum, 20 mM HEPES, 50 U/mL penicillin-streptomycin
2 × Laemmli buffer	125 mM Tris-HCl (pH 6.8), 20% (v/v) glycerol, 10% 2-mercaptoethanol, 4% sodium dodecyl sulphate (SDS), 0.004% bromophenol blue
SDS-PAGE running buffer	25 mM Tris-HCl (pH 8.0), 192 mM glycine, 0.1% SDS
EM buffer	10 mM HEPES (pH 7.6), 200 mM NaCl, 5 mM $MgCl_2$, 1 mM KCl, 1 mM $CaCl_2$
Tris-Borate-EDTA (TBE) buffer	90 mM Tris base (pH 8.3), 80 mM boric acid, 2 mM EDTA
Coomassie blue stain	45% (v/v) methanol, 45% (v/v) dH_2O, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie Blue R250
Coomassie de-stain solution	50% (v/v) dH_2O, 40% (v/v) methanol, 10% (v/v) acetic acid
Phosphate-buffered saline (PBS)	2.7 mM KCl, 1.8 mM KH_2PO_4, 137 mM NaCl, 10 mM Na_2HPO_4, pH 7.4
Paraformaldehyde (PFA) fixing solution	4% (w/v) paraformaldehyde in PBS
Luria-Bertani (LB) medium	1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract
Luria-Bertani (LB) agar	1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) agar

2.1.2. Primers

The primer sequences used for the work described in this thesis are provided in Table 2.2.

Table 2.2. Primers used for molecular cloning.

Primer name	5'-3' sequence
MNV_F6	GGTCTGGTAACACAGTCATTGCC
MNV_F8	GTGAGGCTGCCATGCATGG
MNV_F9	CGCCTGGTTTGCATGCTGTAC
MNV_F10	GGGTGCACGTTGATGGGAC
MNV_F11	GCAGAGGCGATAGACTGTGC
MNV_R1	CCTGCTACTCCCGATCTTAGG
MNV_R4	GCACAGTCTATCGCCTCTGC
MNV_VP1_L412R_fwd	CAAGGGGAACCAGTCGCCCATC
MNV_VP1_L412R_rvs	GATGGGCGACTGGTTCCCCTTG
MNV_VP1_L412E_fwd	CAAGGGGAACCAGTTCCCCATC
MNV_VP1_L412E_rvs	GATGGGGAACTGGTTCCCCTTG
MNV_VP1_L412I_fwd	CAAGGGGAACCAGAATCCCATC
MNV_VP1_L412I_rvs	GATGGGATTCTGGTTCCCCTTG
MNV_VP1_L412C_fwd	CAAGGGGAACCAGGCACCCATC
MNV_VP1_L412C_rvs	GATGGGTGCCTGGTTCCCCTTG
MNV_VP1_C143A_fwd	CATGACATGTGGGAAGGCTGTGATC
MNV_VP1_C143A_rvs	GATCACAGCCTTCCCACATGTCATG

2.2. Gel-based techniques

2.2.1. SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins for analysis of sample identity, purity and abundance. To create SDS-PAGE gels, gel plates (Bio-Rad) were first assembled according to the manufacturer's instructions. Components of the resolving layer of the gel (Table 2.3) were mixed and applied to the gel plates, leaving a 2 cm space from the top. Distilled water was carefully layered on top of the gel to ensure a flat, smooth interface and exclude any air. After the

resolving layer had set, the distilled water was removed and components of the stacking layer of the gel (Table 2.4) were mixed and poured onto the resolving layer. Wells for sample loading were created by insertion of a comb.

Samples to be loaded were mixed with 2 × Laemmli buffer (Table 2.1) and heated at 95°C for 2 min, before centrifugation (17,000 × *g*, 1 min). Gels were placed into a gel tank which was subsequently filled with SDS-PAGE running buffer (Table 2.1), then samples were loaded into wells. A protein marker (Precision Plus Protein™ Dual Color Standards [Bio-Rad]) (5 µL) was also loaded. A fixed voltage of 185 V was applied across the gel until the dye front reached the bottom of the gel.

Table 2.3. Composition of resolving layer of SDS-PAGE gels (total 10 mL).

Component	10% gel (mL)	12% gel (mL)
dH_2O	4.1	3.4
Acrylamide solution (30%)	3.3	4
1.5 M Tris.HCl (pH 8.8)	2.5	2.5
10% (w/v) sodium dodecyl sulphate (SDS)	0.1	0.1
10% (w/v) ammonium persulphate (APS)	0.1	0.1
Tetramethylethylenediamine (TEMED)	0.01	0.01

Table 2.4. Composition of stacking layer of SDS-PAGE gels (total 4 mL).

Component	6% stacking gel (mL)
dH_2O	2.7
Acrylamide solution (30%)	0.8
1 M Tris.HCl (pH 6.8)	0.5
10% (w/v) sodium dodecyl sulphate (SDS)	0.1
10% (w/v) ammonium persulphate (APS)	0.1
Tetramethylethylenediamine (TEMED)	0.01

2.2.2. Coomassie blue staining

To visualise total protein content following SDS-PAGE, gels were submersed in Coomassie blue stain (Table 2.1) for 2 hours at room temperature under gentle agitation. Stain was then removed and gels were washed in Coomassie

de-stain solution (Table 2.1) until background staining was reduced and bands became clearly visible.

2.2.3. DNA agarose gel electrophoresis

1% (w/v) agarose was dissolved in TBE buffer (Table 2.1) along with 0.0025% (v/v) SYBR Safe DNA gel stain (Invitrogen™), poured into a gel cast and allowed to set. The gel was loaded into an electrophoresis tank filled with TBE buffer (Table 2.1), and DNA samples were mixed with 6 × DNA loading dye (New England BioLabs) before being loaded into wells. A 1 kb DNA ladder (New England BioLabs) was loaded alongside samples. A fixed voltage of 100 V was applied across the gel until the dye front had progressed sufficiently. Gels were visualised using a blue light transilluminator.

2.2.4. RNA agarose-formaldehyde gel electrophoresis

0.8% (w/v) RNase-free agarose was dissolved in 1 × MOPS buffer (VWR Life Science) before adding 5% (v/v) formaldehyde and allowing the gel to set. The gel was then placed into an electrophoresis tank, which was subsequently filled with 1 × MOPS buffer. Approximately 500 ng of each RNA sample, alongside an RNA ladder (Millennium RNA Markers [Life Technologies]), was mixed with 2 × RNA loading dye and heated at 85°C for 5 min, then cooled on ice. RNA samples and ladder were loaded into the gel and a voltage of 60 V was applied until the dye front had progressed sufficiently. Bands were visualised under UV light.

2.3. DNA/RNA protocols and mutagenesis

2.3.1. Plasmid extraction from *E. coli*

Transformed stocks of *E. coli* stored at −80°C in 30% glycerol were used to inoculate 10 mL LB (containing 100 µg/mL ampicillin if a resistance marker was present in the desired plasmid) for incubation at 37°C, 180 rpm for 16-20 hours. Cultures were then taken directly for plasmid extraction, or were first streaked on LB agar plates (with 100 µg/mL ampicillin if appropriate) and

incubated overnight at 37°C, such that single colonies could be taken forward. Plasmid purification was performed using a QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions. DNA was eluted in 30 – 50 µl dH_2O and used immediately or stored at −20°C.

2.3.2. Polymerase chain reaction (PCR)

For amplification of DNA by PCR (Garibyan and Avashia, 2013), 50 µl reactions were prepared in nuclease-free dH_2O as follows: 1 × Phusion™ HF buffer, 200 µM each dNTP (dATP, dTTP, dCTP, dGTP), 0.5 µM forward primer, 0.5 µM reverse primer, 10 – 50 ng template DNA, 1 unit Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific™). Reactions were performed using a Veriti™ Thermal Cycler (Applied Biosystems) according to the schedule in Table 2.5. The anneal temperature was selected from the range specified in Table 2.5 according to the lowest predicted melting temperature of the two DNA primers used for the reaction.

Table 2.5. Thermal cycler schedule for PCR.

Step		**Temp (°C)**	**Time**
Initial denaturation		98	30 sec
Amplification (25-35 cycles)	Denature	98	5 - 10 sec
	Anneal	45 - 72	10 - 30 sec
	Extend	72	30 sec / kb
Final extension		72	5 - 10 min
Hold		4	∞

2.3.3. DNA purification from agarose gels

Following agarose gel electrophoresis (Section 2.2.3), DNA bands were visualised using a blue light transilluminator and excised from the gel with a scalpel. DNA was then extracted and purified using a QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions and eluted in a small volume (20 – 50 µl) of dH_2O. Extracted DNA was used immediately or stored at −20°C.

2.3.4. DNA sequencing

DNA was diluted in dH_2O to a final concentration of 40 – 100 ng/µl and sent to GENEWIZ for Sanger Sequencing using one or several of the following primers: MNV_F8, MNV_F9, MNV_F10, MNV_F11, MNV_R1, MNV_R4 (Table 2.2).

2.3.5. Preparation of viral genomic RNA for sequencing

To prepare viral genomic RNA for sequencing, viral genomes were extracted and converted into cDNA. A purified sample of MNV was mixed 1:3 with QIAzol lysis reagent (QIAGEN), then RNA was isolated using the Direct-zol RNA Miniprep Kit (Zymo Research) according to the manufacturer's instructions. RNA was eluted in 15 µl RNase-free water and kept on ice. cDNA was generated using a specific primer (MNV_R1 or MNV_R4) with the SuperScript™ II Reverse Transcriptase kit (Invitrogen™) according to the manufacturer's instructions, then used immediately or stored at −20°C. cDNA was subsequently diluted to 10% of its original concentration with dH_2O and amplified by PCR (Section 2.3.2). PCR products were separated by agarose gel electrophoresis and DNA bands of the expected size were extracted and purified (Section 2.3.3), then sequenced (Section 2.3.4).

2.3.6. Molecular cloning

2.3.6.1. Restriction enzyme digest

Restriction enzyme digestion reactions were performed sequentially, according to the manufacturer's instructions (New England BioLabs). Reactions were prepared in 30 or 50 µl volumes with 1 × NEB buffer, 10 – 20 units restriction enzyme (AfeI, BspEI or NotI-HF) and up to 5 µg DNA. Reactions were incubated at 37°C for 30 min. Buffer exchanges were performed between digestion reactions using a QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions. Following the final digestion reaction and only if DNA was being prepared for ligation, 2 units calf

intestinal phosphatase (New England BioLabs) were added to the reaction vessel for template DNA only, and incubated at 37°C for a further 30 – 60 min. Insert DNA was left untreated. Reaction products were analysed by agarose gel electrophoresis (Section 2.2.3) and then purified (Section 2.3.3) for ligation.

2.3.6.2. DNA ligation

DNA ligation reactions were performed using T4 DNA ligase (New England BioLabs) according to the manufacturer's instructions. 20 µl reactions were prepared with 1 × T4 ligase buffer, 1 unit T4 ligase, and a total of ~100 ng DNA at a molar ratio of 1:3 for template DNA and insert DNA. Reactions were incubated for 30 – 120 min at room temperature. As a control, an empty vector ligation reaction was performed in parallel in the absence of insert DNA to check for self-ligation.

2.3.6.3. Transformation

Chemically competent DH5α *E. coli* cells were thawed on ice, then 50 µl aliquots were added to each DNA ligation reaction vessel and incubated on ice for 30 min. Heat shock was performed at 42°C for 1 min, then reaction vessels were immediately returned to ice for 2 min before adding 500 µl LB (Table 2.1) and incubating at 37°C, 180 rpm for 30 min. Subsequently, transformed cells were pelleted (5,000 × *g*, 5 min), resuspended in 50 µl LB and streaked onto LB agar (Table 2.1) plates containing ampicillin (100 µg/mL) where appropriate. LB agar plates were incubated at 37°C for 16 – 20 hours to allow colonies to develop.

2.4. Cell culture

2.4.1. Cell propagation

RAW264.7 cells (kindly gifted from Ian Clarke, University of Southampton), BV-2 cells and BHK-21 cells (obtained from the ATCC; Manassas, VA, USA) were used for this work. RAW264.7 cells are macrophage cells transformed with Abelson leukaemia virus and derived from a BALB/c mouse (Raschke *et*

al., 1978). BV-2 cells are microglial cells transformed with a recombinant retrovirus (J2) and derived from a C57BL/6 mouse (Blasi *et al.*, 1990). BHK-21 cells are fibroblast cells derived from a golden hamster (Macpherson and Stoker, 1962). All cells were maintained in supplemented cell culture medium (Table 2.1) and incubated at 37°C, 5% CO_2. Upon reaching 80 – 90% confluency, cells were passaged using a cell scraper (RAW264.7) or by treatment with a 1% trypsin-EDTA solution (Sigma-Aldrich) at 37°C (BV-2, BHK-21).

2.4.2. Cell quantitation

To count cells for accurate seeding, cells were first detached (either by cell scraper or using 1% trypsin-EDTA solution [Sigma-Aldrich]) and a 20 µL sample was mixed 1:1 with Trypan Blue stain before being applied to a glass haemocytometer. An average live cell count (*n*) was determined from individual counts of three haemocytometer squares. Cell density in the stock solution (cells/mL) was calculated using the formula:

$$cell\ density = 2n \times 10^4$$

The stock solution was subsequently diluted with supplemented cell culture medium (Table 2.1) to achieve the desired cell density.

2.5. Virus culture and purification

2.5.1. Generation of viral RNA from an infectious clone

To generate RNA from an infectious clone, plasmid DNA was first purified from transformed *E. coli* stocks (Section 2.3.1). Purified plasmid DNA was subsequently linearised by digestion with NotI-HF (New England BioLabs) (Section 2.3.6.1) before phenol chloroform extraction. 5 µg linearised DNA was made up to a final volume of 100 µl with nuclease-free dH_2O, then vigorously mixed with an equal volume of phenol:chloroform:isoamyl alcohol mix (25:24:1) before centrifugation at 17,000 × *g* for 10 min. The aqueous

phase was transferred to a new vessel, then an equal volume of chloroform:isoamyl alcohol mix (24:1) was added and the solution mixed vigorously before further centrifugation (17,000 × *g*, 10 min). Again, the aqueous phase was recovered and incubated with 0.1 volumes of NH_4OAc and 2.5 volumes of 100% EtOH overnight at −20°C. Precipitated DNA was pelleted by centrifugation (17,000 × *g*, 4°C, 30 min) and the supernatant removed before washing DNA with an equal volume of 70% EtOH and pelleting through further centrifugation (17,000 × *g*, 4°C, 30 min). Subsequently, EtOH was removed and pelleted DNA was incubated at 42°C to dry before being resuspended in 20 µl nuclease-free dH_2O.

Extracted DNA was taken forward as a template for *in vitro* transcription using a HiScribe™ T7 ARCA mRNA Kit (New England BioLabs) according to the manufacturer's instructions, including DNase treatment and poly(A) tailing steps. RNA was purified using an RNA Clean & Concentrator Kit (Zymo Research) according to the manufacturer's instructions but without any further DNase treatment. RNA was eluted into 50 µl nuclease-free dH_2O and stored at −80°C. Prior to transfection, RNA quality was assessed by agarose-formaldehyde gel electrophoresis (Section 2.2.4).

2.5.2. Transfection of cells with viral RNA

BHK-21 cells were transfected with viral RNA to generate infectious virus. Cells were seeded into the wells of a 24-well plate or a T-75 flask and incubated in supplemented cell culture medium (Table 2.1) overnight. Upon reaching 70% – 80% confluency, spent media was removed and cells were gently washed with PBS (Table 2.1). Transfection was then performed using Lipofectin™ Transfection Reagent (Invitrogen™) according to the manufacturer's instructions, using 0.5 µg RNA/cm^2. 48 – 72 hours post-transfection, cells were lysed and nascent virus harvested through three cycles of freezing and thawing.

2.5.3. Virus propagation

MNV (MNV-1, strain CW1P3 (Ward *et al.*, 2007)) was propagated in RAW264.7 cells, as described previously (Wobus *et al.*, 2004), or in BV-2 cells. Cells were seeded in 175 cm^2 flasks and allowed to reach 80% confluency before the supplemented cell culture medium was replaced (Table 2.1) and crude stocks of MNV (originating from previous passages, or the lysate from cells transfected with MNV RNA) were added to cells. Infection was allowed to proceed for 48 – 72 hours (until confluent CPE was observed), then cells were harvested. Virus was released by three freeze-thaw cycles, then the suspension was clarified (3,300 × *g*, 10 min, 4°C) and virus pelleted by ultracentrifugation (366,000 × *g*, 60 min, 4°C). PBS was used to resuspend the virus pellet, giving a concentrated crude MNV stock for further propagation or purification. For hsMNV (Section 3.5), the concentrated crude stock was thermally stressed after each passage to maintain a selection pressure and minimise the chance of reversion.

2.5.4. Virus thermal stressing

To thermally stress virus (similar to a protocol described previously (Adeyemi *et al.*, 2017)), crude virus stocks were first clarified (17,000 × *g*, 1 min) and separated into 100-µL aliquots. The clarified viral aliquots were then heated to 52°C for 30 min (unless otherwise specified) using a Veriti™ Thermal Cycler (Applied Biosystems), before cooling to 4°C. The thermal stressing temperature of 52°C was chosen as it was found to reduce viral titres significantly but did not lead to complete loss of infectivity, permitting further passage of virus.

2.5.5. Virus purification

MNV was purified according to adapted versions of established protocols (Hwang *et al.*, 2014). Where appropriate, the basic approach detailed below was modified, as indicated in the text.

To release virus from cells, infectious media were freeze-thawed three times. The detergent NP-40 was added to the infectious lysate to a final concentration of 0.1% (v/v), then lysates were clarified (3,300 × *g*, 10 min, 4°C). The pellet was discarded, and clarification repeated two more times before the clarified supernatant was subjected to ultracentrifugation through a 30% (w/v) sucrose cushion (150,000 × *g*, 3 h, 4°C). Following this, pellets were resuspended in PBS, clarified by centrifugation (17,000 × *g*, 10 min), then loaded onto a 15% – 60% (w/v) sucrose gradient for ultracentrifugation at 300,000 × *g* (50 min, 4°C). The gradient was fractionated manually and SDS-PAGE analysis was performed to determine the peak fractions, which were combined and their contents pelleted by ultracentrifugation (366,000 × *g*, 60 min, 4°C). The virus pellet was resuspended in PBS and a second round of sucrose gradient ultracentrifugation was performed.

Sucrose reduces contrast in EM imaging (Thompson *et al.*, 2016), therefore this was removed prior to structural studies using dialysis. Peak fractions from the second round of sucrose gradient purification were combined and dialysed in 1 L EM buffer (Table 2.1) at room temperature using a 10,000 molecular weight cut-off Slide-A-Lyzer dialysis cassette (Thermo Scientific). After 1 hour, the spent EM buffer was replaced with fresh EM buffer for another hour before replacing the EM buffer again for overnight incubation at 4°C. Dialysed virus was recovered from the dialysis cassette and stored at 4°C prior to structural work.

2.6. Viral infectivity ($TCID_{50}$) assay

Median tissue culture infectious dose ($TCID_{50}$) assays were used to measure viral infectivity, according to a modified version of a previously published protocol (Hwang *et al.*, 2014). 2.0×10^4 cells (RAW264.7 or BV-2) in 100 µL supplemented cell culture medium (Table 2.1) were seeded into each well of a 96-well plate and incubated for 24 hours. 10-fold serial dilutions of MNV in supplemented cell culture medium (Table 2.1) were prepared, and 100 µL of each MNV concentration was added to 100 µL of media already present in

each well. Infection was allowed to proceed for 72 hours, then cells were fixed with PFA fixing solution (Table 2.1) and stained with crystal violet solution. Cells were analysed for CPE, then $TCID_{50}$ values were determined according to the Spearman and Kärber algorithm (Hierholzer and Killington, 1996).

2.7. Sample preparation for electron microscopy

2.7.1. EM grid glow discharge

In order to create a hydrophilic grid surface through the deposition of negatively charged ions, carbon-coated copper EM grids were glow discharged prior to sample application. Grids were placed facing upwards on a parafilm-coated glass slide, then glow discharged in air at 10 mA for 30 seconds using a PELCO easiGlow™ Glow Discharge Cleaning System or a Cressington 208carbon. When specifically indicated, glow discharge was instead performed under amylamine vapour using a Quorum GloQube® Plus Glow Discharge System to generate a positively charged grid surface (Drulyte *et al.*, 2018).

2.7.2. Negative staining procedure

For visualisation by negative staining, 3 µL aliquots of samples were applied to carbon-coated 300-mesh copper grids (Agar Scientific) for 30 seconds following glow discharge. Excess liquid was removed by blotting with filter paper. In some cases (as indicated), further applications of sample were performed to concentrate particles on the grid surface. Following sample application, grids were immediately washed twice with 10 µL dH_2O, blotting off excess liquid each time. For samples containing sucrose or glycerol, these washes were repeated to improve contrast. Grids were subsequently stained with 10 µL 1 – 2% uranyl acetate solution, which was promptly blotted away before another 10 µL 1 – 2% uranyl acetate solution was applied for 30 seconds. Excess liquid was lightly blotted away to leave a thin film of stain which was left to air dry. Grids were stored at room temperature until they could be visualised on a TEM.

2.7.3. Vitrification of grids for cryoEM

Lacey carbon 400-mesh copper grids coated with a <3-nm continuous carbon film (Agar Scientific) were first glow-discharged as described above, then loaded into the sample chamber of a Leica EM GP plunge freezing device under 80% relative humidity at 8°C. Sample was applied to the grid surface (3 µL) through a side port and left on the grid for 30 seconds, before automated blotting and vitrification in liquid nitrogen-cooled liquid ethane held at −178°C. Grids were quickly transferred into liquid nitrogen, and were stored in liquid nitrogen-filled dewars until ready to be loaded into a microscope. For each sample, several grids were prepared using different blotting time and strength parameters, leading to variation in ice thickness/quality. Grids were screened using an EM to select optimal candidates for data collection.

As indicated throughout the thesis, a multiple blotting approach was sometimes used prior to plunge freezing to increase the particle concentration on the grid surface. 30 seconds after the first sample application, instead of automatic blotting, the grid was manually blotted with filter paper and another 3 µL of sample was quickly reapplied. This was repeated once more (for a total of three applications), before automatic blotting and plunge freezing was performed.

2.8. Electron microscopy

2.8.1. EM of negatively stained samples

Negatively stained grids were imaged at ABSL, University of Leeds, using either (i) an FEI Tecnai G2-spirit (with LaB_6 as an electron source) operating at 120 kV, equipped with a Gatan Ultra Scan 4000 CCD, or (ii) an FEI Tecnai F20 (with a field emission gun [FEG]) operating at 200 kV, equipped with an FEI CETA camera. Where data was collected for further processing, micrographs were collected using a range of defocus values (in general, between −2.0 µm to −5.0 µm).

2.8.2. CryoEM

Vitrified grids were imaged at ABSL, University of Leeds, using an FEI Titan Krios microscope equipped with a Falcon III or Falcon IV camera. The microscope was operated at 300 kV and a magnification of 75,000×, giving a calibrated object sampling of 1.065 Å/pixel. Specific parameters for each data collection are indicated throughout the thesis.

2.9. Image processing and atomic model building

Unless otherwise indicated, processing was performed primarily in Relion-2.1, Relion-3.0 and Relion-3.1 pipelines (Scheres, 2012; Kimanius *et al.*, 2016; Zivanov *et al.*, 2018). General descriptions of the approaches used for this work are provided below, with specific details and modifications for each project highlighted throughout the thesis.

2.9.1. Image processing

Prior to the selection and extraction of individual particles, micrographs were processed with MotionCor2 to correct for motion-induced blurring, and the contrast transfer function (CTF) for each was estimated using Gctf or CTFFIND4 for later correction in Relion (Rohou and Grigorieff, 2015; Zhang, 2016; Zheng *et al.*, 2017). In general, a subset of 500 – 2000 particles was manually picked and extracted from motion-corrected micrographs, then classified in 2D to generate templates for automated particle picking (Scheres, 2015). To optimise this process, automated particle picking using these templates was performed multiple times on a subset of 20 – 30 micrographs, each time varying the ‘picking threshold’ and ‘minimum inter-particle distance’ parameters. Results were visually inspected, and the combination of parameters that gave the minimum amount of off-target picking, while still picking all desired target particles, was used for picking of the entire dataset. Particles were extracted with dust removal, such that individual pixels with intensity values more than three standard deviations from the image mean (in either direction) were replaced with values taken from a Gaussian distribution.

Following this, particles were classified in 2D. In general, a first round of 2D classification was performed to remove erroneous particle images selected by off-target picking, with any classes corresponding to the target particle (regardless of the apparent quality of the class average) taken forward for further classification. For larger datasets, this step was performed with CTFs ignored until the first peak (and in some cases, with the 'fast subsets' option applied) to improve processing speed. Further rounds of 2D classification were then performed to narrow down the particle stack to those particles contributing to high-quality classes (i.e., those where structural details were clear and there was no evidence of overlapping particles or contamination in class averages). Classes for each target were grouped separately for datasets with multiple targets (e.g. differently sized VLPs). Unless otherwise indicated, all 2D classifications were performed with a regularisation parameter of 2, in-plane angular sampling of 6°, offset search range of 5 pixels, and offset search step of 1 pixel.

The refined particle stack was used to generate a 3D model of the structure *de novo* through a stochastic gradient descent (SGD) algorithm implemented in Relion, similar to that developed for use in cryoSPARC (Punjani *et al.*, 2017). This was used as an initial model for 3D classification with alignments, using default parameters, including a regularisation parameter of 4. The best quality class was taken forward for high-resolution refinement (3D auto-refine). The post-2D classification particle stack (prior to any 3D classification) was also taken forward for separate 3D refinement, and in the event that this yielded a better resolved map than the 3D classified data, the post-2D classification particle stack was used for further processing, without the initial 3D classification step. For 3D refinements, the initial model was low-pass filtered to 60 Å to avoid biasing the reconstruction, and in general, default parameters were used (an initial angular sampling of 7.5°, initial offset range of 5 pixels, initial offset step of 1 pixel). Local angular searches were set to begin at angular samplings below 1.8°. Following refinement, a mask was applied to the map for sharpening, and the nominal resolution determined using the gold-

standard Fourier shell correlation (FSC) criterion (FSC = 0.143) (Scheres and Chen, 2012).

Multiple approaches were taken to improve the quality of the map, with only those yielding an improvement in resolution taken forward for each project. Firstly, 3D classification without alignments was performed using the particles and associated orientational information from a previous 3D refinement. Particles were separated into two classes, and those assigned to the least well resolved class were discarded. Secondly, CTF refinement and Bayesian polishing were performed iteratively, each time followed by 3D refinement with masking and use of solvent-flattened FSCs (Zivanov *et al.*, 2018, 2019). This was repeated, sometimes varying the precise corrections applied during CTF refinement between iterations, until map quality (judged by visual inspection and nominal resolution) stopped improving.

Following determination of a high-resolution structure, local resolution for the map was calculated with Relion's implementation of local resolution estimation, using the B-factor estimated from map sharpening.

2.9.2. Specialised image processing techniques

To solve specific issues associated with flexibility, heterogeneity, and symmetry mismatches (described in Chapters 3 – 5), focussed classification and symmetry relaxation were performed. Specific details are given at the relevant points in the thesis, but general descriptions are provided here.

For focussed classification, particles and their associated orientational information from 3D refinement were first symmetry expanded using the relion_particle_symmetry_expand function. This generated an expanded list of particles, containing identical copies of each particle assigned to every different symmetrically redundant orientation. Next, a cylindrical or spherical mask of appropriate size was generated in SPIDER (Frank *et al.*, 1996; Shaikh *et al.*, 2008), then manually aligned with the target region of the symmetrised map in UCSF Chimera and resampled onto the same coordinate system

(Pettersen *et al.*, 2004). The resampled mask was imported into Relion and a soft edge of 3 – 6 pixels was applied. 3D classification without alignments was performed using the symmetry expanded particles and mask. The original symmetrised map was used as a reference model, low-pass filtered to a resolution close to the nominal resolution of the symmetrised map. Importantly, no symmetry was applied during the classification (C1) and a relatively high value was given for the regularisation parameter (20 – 60). When the dataset was particularly large, the computational burden was reduced by using the 'fast subsets' option, and/or by downsampling the data prior to classification. Classes were inspected visually, with interesting classes separated into two random subsets, separately reconstructed using the relion_reconstruct tool, then sharpened in Relion.

For symmetry relaxation to resolve the contents of a viral capsid or VLP, UCSF Chimera was used to separate the symmetrised reconstruction into capsid density and interior density (Pettersen *et al.*, 2004). The capsid density was used as a mask to subtract capsid signal from particle images in Relion. Subtracted particles were then subjected to 3D classification with symmetry relaxed from icosahedral (I1) to asymmetric (C1), using interior density (low-pass filtered to 60 Å) as a reference model. A spherical mask was applied to exclude any remaining capsid signal during classification. In some cases, shell-shaped masks were generated in SPIDER and used instead of a spherical mask, to isolate individual layers of nucleic acid density (Frank *et al.*, 1996; Shaikh *et al.*, 2008). Resultant classes were inspected in UCSF Chimera (Pettersen *et al.*, 2004).

2.9.3. Model building and refinement

To build an atomic model from cryoEM data, a preliminary model was first generated using a homology modelling server (such as SWISS-MODEL or Phyre2) or the structure prediction tool, AlphaFold2 (Kelley *et al.*, 2015; Waterhouse *et al.*, 2018; Jumper *et al.*, 2021; Mirdita *et al.*, 2021). The preliminary model was rigid-body fitted into the density map using UCSF Chimera, then subjected to an iterative process of manual refinement in Coot

and automated refinement in Phenix, sometimes with secondary structural restraints imposed (Pettersen *et al.*, 2004; Adams *et al.*, 2010; Emsley *et al.*, 2010). To avoid incorrect positioning of side chains in unoccupied density from adjacent asymmetric units, at least one refinement in Phenix was performed with coordinates that had been symmetrised in UCSF Chimera, such that adjacent asymmetric units were occupied. After each iteration, validation statistics related to model quality were inspected, including the prevalence of steric clashes ('clashscore'), Ramachandran angle outliers, rotamer outliers, and model-to-map fit. Iterative refinement was continued until model validation statistics stopped improving.

2.10. Structural analysis and validation

Structural analysis was performed in UCSF Chimera and PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC), with atomic model validation statistics calculated in MolProbity (Pettersen *et al.*, 2004; Chen *et al.*, 2010). To calculate root-mean-square deviation (RMSD) values between atomic models, atomic coordinates were aligned and the variation between equivalent Cα atoms calculated using the 'MatchMaker' tool of UCSF Chimera with default settings. Where relevant, surface charge was calculated using PDB2PQR and APBS, implemented within PyMOL (Baker *et al.*, 2001; Dolinsky *et al.*, 2007). The PDBePISA (https://www.ebi.ac.uk/pdbe/pisa/) and PDBeFold (https://www.ebi.ac.uk/msd-srv/ssm/) web servers were used to characterise predicted molecular interfaces and identify structures with a high degree of similarity, respectively (Krissinel and Henrick, 2004, 2007). Primary sequence alignments were performed in Clustal Omega using default parameters (Sievers *et al.*, 2011).

Structural visualisation was performed in UCSF Chimera, UCSF ChimeraX and PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) (Pettersen *et al.*, 2004; Goddard *et al.*, 2018). Particle orientation distribution was visualised using an adapted version of a script from Naydenova & Russo (Naydenova and Russo, 2017; Klebl *et al.*, 2020).

2.11. Deposition of structural data in public databases

Atomic coordinates and cryoEM density maps for many of the structures described in this thesis, along with supporting files (such as half-maps and masks), were deposited in the relevant public databases, such as the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB). Accession codes for these entries are given in Table 2.6 (PDB) and Table 2.7 (EMDB).

Table 2.6. Accession codes for atomic coordinates deposited in the PDB. *For atomic coordinates used in multiple chapters, the entry is only displayed under the first chapter in which those coordinates appear.*

Model	PDB ID
Chapter 3	
VP1 from wtMNV.RAW	6S6L
Chapter 4	
VP1 from wtMNV.BV2	*To be deposited*
VP1 from wtMNV.BV2 bound to LCA	*To be deposited*
Chapter 5	
AR1 (T = 3)*	*To be deposited*
AR1 (T = 4)	*To be deposited*
n190 (T = 3)	*To be deposited*
n190 (T = 4)	*To be deposited*
n190:GP1 (T = 3)	*To be deposited*
n190:GP1 (T = 4)	*To be deposited*
P. pastoris fatty acid synthase (FAS)	7BC4
P. pastoris fatty acid synthase (FAS) acyl carrier protein (ACP) domain	7BC5

Table 2.7. Accession codes for structural data deposited in the EMDB. *For structural data used in multiple chapters, the entry is only displayed under the first chapter in which those coordinates appear.*

Map	EMDB ID
Chapter 3	
wtMNV.RAW	10103
hiMNV	10127
hsMNV	10128
Chapter 4	
wtMNV.BV2	*To be deposited*
wtMNV:LCA	*To be deposited*
Chapter 5	
AR1 (T = 3)*	*To be deposited*
AR1 (T = 4)	*To be deposited*
n190 (T = 3)	*To be deposited*
n190 (T = 4)	*To be deposited*
n190:GP1 (T = 3)	*To be deposited*
n190:GP1 (T = 4)	*To be deposited*
P. pastoris fatty acid synthase (FAS)	12138
P. pastoris FAS including ACP domain resolved by focussed classification	12139

3. The cryoEM structures of MNV variants

3.1. Introduction to noroviruses

This chapter describes my efforts to demonstrate the ability of cryoEM to further our understanding of virus structure and function, using murine norovirus (MNV) as a model system. By using cryoEM to study a number of MNV variants, I showed that the norovirus capsid is dynamic, and that this dynamic nature is important for viral stability and infectivity.

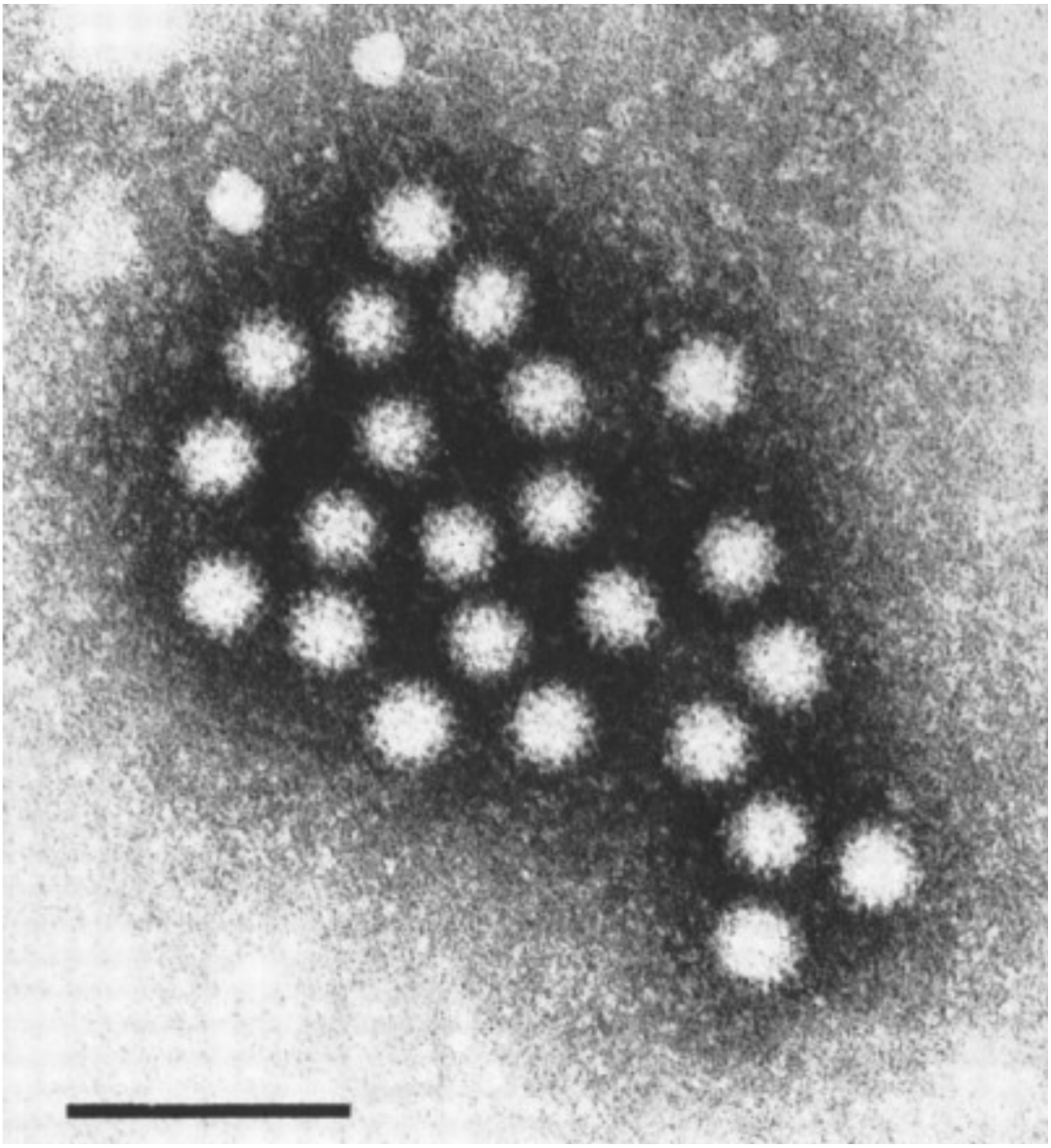
3.1.1. Norovirus biology

3.1.1.1. Norovirus overview

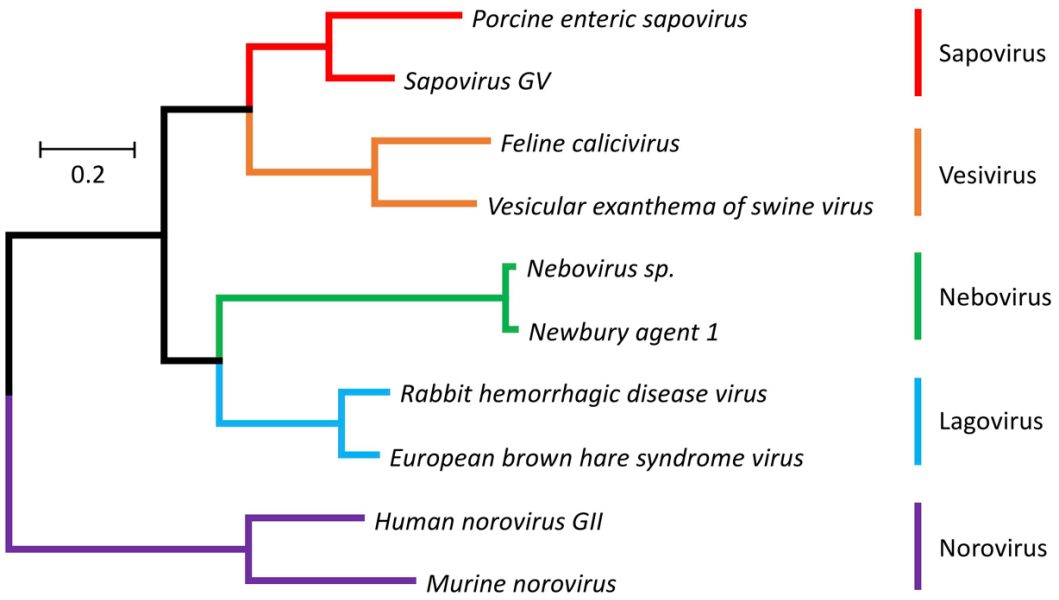
In the winter of 1968, an epidemic of 'winter vomiting disease' took place at an elementary school in Norwalk, Ohio (Adler and Zickl, 1969). At the time, the aetiological agent was unknown, but later analysis of patient stool samples using immune electron microscopy led to the discovery of the first norovirus to be described, Norwalk virus (Figure 3.1) (Kapikian *et al.*, 1972; Kapikian, 2000). Since then, it has become widely acknowledged that noroviruses are responsible for a significant burden of disease around the world, and following the development of successful vaccines against rotaviruses, they have become the leading cause of acute gastroenteritis in all age groups (Ahmed *et al.*, 2014; Bányai *et al.*, 2018).

As members of the *Caliciviridae* family, noroviruses have positive-sense single-stranded RNA genomes (class IV of the Baltimore classification) and are related to other calicivirus genera, including lagoviruses, neboviruses, sapoviruses, vesiviruses and recoviruses (Figure 3.2) (Smertina *et al.*, 2019). Even within the norovirus genus, there is considerable genetic diversity. Noroviruses can be classified based on (i) VP1 amino acid sequence into one of at least ten genogroups (GI – GX), with each subdivided into genotypes (e.g., GII.4), or (ii) based on partial nucleotide sequences for the RNA-dependent RNA polymerase (RdRp) into P-groups (GI.P – GX.P) and P-types (e.g., GII.P15) (Chhabra *et al.*, 2019). Noroviruses have been reported to infect a broad group of mammals including pigs, dogs, cats, mice, sheep and

cattle (Liu *et al.*, 1999; Karst *et al.*, 2003; Wolf *et al.*, 2009; Mesquita *et al.*, 2010; Shen *et al.*, 2012; Di Martino *et al.*, 2016), with human noroviruses (HuNVs) essentially limited to genogroups I, II, and IV (exceptions have recently been classified as genogroups VIII and IX (Chhabra *et al.*, 2019)).



3.1. Identification of Norwalk virus by immune electron microscopy. *Aggregate of Norwalk virus particles in stool filtrate incubated with human serum, visualised by electron microscopy. Scale bar represents 100 nm. Figure adapted from (Kapikian, 2000).*



3.2. Caliciviridae phylogenetic tree. *Phylogenetic tree based on RNA-dependent RNA polymerase peptide sequences from different caliciviruses. Branch length is drawn proportional to genetic change. Scale bar represents 0.2 substitutions per site. Note that not all calicivirus genera are included in the phylogenetic tree. Figure adapted from (Smertina et al., 2019), published under a CC BY 4.0 licence.*

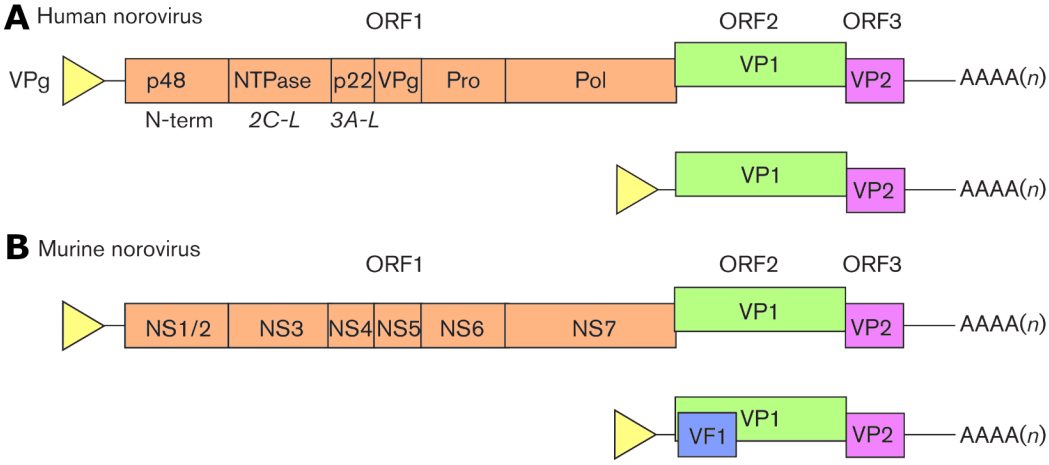
A significant obstacle to norovirus research is the lack of a simple culture system for human noroviruses, with many attempts ultimately proving unsuccessful (Duizer *et al.*, 2004; Papafragkou *et al.*, 2013; Takanashi *et al.*, 2014). Methods for successfully culturing human noroviruses have been reported, but these are difficult and time-consuming techniques based on stem-cell derived intestinal enteroids, or infection of B cells in the presence of histo-blood group antigen (HBGA-)expressing enteric bacteria (Jones *et al.*, 2014; Ettayebi *et al.*, 2016). As a result, much of our knowledge of human norovirus biology is based on work using other viruses as model systems, including feline calicivirus (a vesivirus), Tulane virus (a recovirus), and murine norovirus (Sosnovtsev and Green, 1995; Wobus *et al.*, 2006; Wei *et al.*, 2008; Vashist *et al.*, 2009; Oka *et al.*, 2014).

Murine norovirus (MNV) was first discovered using representational difference analysis in 2003, and was identified as the cause of a lethal infection in

laboratory mice deficient for STAT1 and RAG2 (Karst *et al.*, 2003). Aside from its potentially significant impact on research using laboratory mice, the importance of MNV as a model for HuNV infection has already been demonstrated, thanks to the availability of well-established cell culture and reverse genetics systems (Wobus *et al.*, 2004, 2006; Chaudhry *et al.*, 2007; Ward *et al.*, 2007). MNV is a particularly useful model system because it shares many biological/pathological features with HuNVs. For example, unlike feline calicivirus (which affects the respiratory tract), MNV is an enteric pathogen and has been observed to cause similar symptoms to HuNVs (such as acute diarrhoea) in wild-type mice (Roth *et al.*, 2020). It also mirrors HuNVs closely in terms of virion structure and genome organisation.

3.1.1.2. Norovirus genome organisation

Noroviruses have small (~7.5 kb), positive-sense single-stranded RNA genomes, capped at the 5' end with a virally encoded protein (termed VPg [viral protein genome-linked]) and polyadenylated at the 3' end. Along with very short untranslated regions (UTRs) that contain evolutionarily conserved secondary structural elements (Simmonds *et al.*, 2008), human norovirus genomes comprise three open reading frames (ORFs) (Figure 3.3A). ORF1 encodes a viral polyprotein which is co- and post-translationally processed by a virally encoded protease (Pro) to yield a range of different non-structural proteins, many of which are named after analogous proteins from picornaviruses. ORF1-encoded proteins include p48/N-term and p22/3A-like (both thought to be involved in replication complex formation), NTPase/2C-like (thought to act as an RNA helicase), VPg (which caps genomes, and is involved in translation and replication), Pro/3C-like (a viral cysteine protease) and Pol/3D^{pol} (the RdRp) (Ettayebi and Hardy, 2003; Hyde and Mackenzie, 2010; Thorne and Goodfellow, 2014). ORF2 and ORF3 encode the major (VP1) and minor (VP2) structural proteins, respectively, and are translated from subgenomic RNA (sgRNA) produced during infection (explained in more detail in Section 4.1).



3.3. Norovirus genome organisation. *Schematic showing organisation of* ***(A)*** *human norovirus and* ***(B)*** *murine norovirus genomes. Norovirus genomes are capped with VPg (yellow) and have poly(A) tails. The human norovirus genome is divided into three ORFs (ORF1 – orange, ORF2 – green, ORF3 – magenta), with ORF1 proteolytically processed to generate the viral proteins indicated (alternative names shown in italics). ORF2 and ORF3 are translated from a subgenomic RNA (which is also capped and poly(A) tailed). The murine norovirus genome has an additional fourth ORF (ORF4 – blue) which overlaps with ORF2. Figure adapted from (Thorne and Goodfellow, 2014).*

The murine norovirus genome is broadly similar to the human norovirus genome, although the nomenclature used to describe the non-structural proteins differs, with non-structural proteins termed NS1/2 – NS7 (summarised in Table 3.1) (Figure 3.3B). One unique biological feature of the MNV genome is the presence of a fourth ORF (ORF4), which overlaps with ORF2 and encodes a protein termed virulence factor 1 (VF1), previously shown to antagonise the host innate immune response (McFadden *et al.*, 2011; Borg *et al.*, 2021).

Table 3.1. HuNV and MNV protein nomenclature and function.

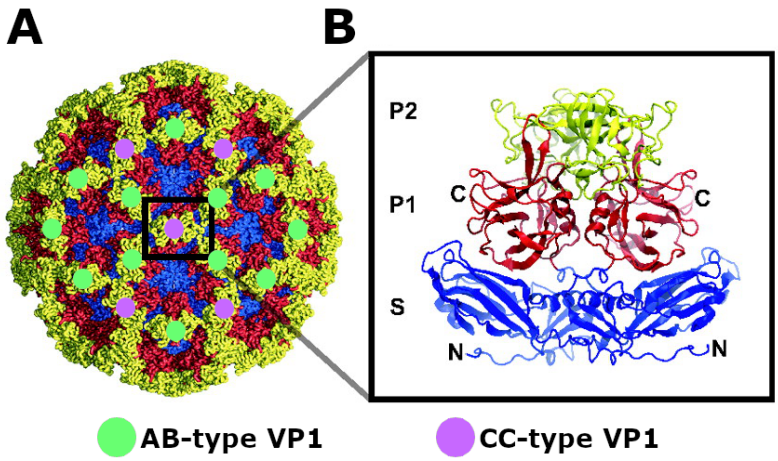
HuNV	MNV	Function
p48/N-term	NS1/2	Putative role in replication complex formation
NTPase/2C-like	NS3	Putative RNA helicase
p22/3A-like	NS4	Putative role in replication complex formation
VPg	NS5	Genome-capping protein
Pro/3C-like	NS6	Cysteine protease
Pol/3Dpol	NS7	RNA-dependent RNA polymerase
VP1	VP1	Major structural protein
VP2	VP2	Minor structural protein
-	VF1	Suppresses host innate immune response

3.1.1.3. The norovirus virion

The norovirus virion comprises the viral genome (covalently linked to VPg) and an undefined quantity of the minor structural protein, VP2, all enclosed within a non-enveloped capsid formed from the major structural protein, VP1.

Given the difficulties associated with culturing infectious HuNVs, our understanding of norovirus capsid structure was pioneered by studies of HuNV virus-like particles (VLPs) or related viruses (Prasad *et al.*, 1994b, 1994a, 1999; Katpally *et al.*, 2008, 2010; Jung *et al.*, 2019). The viral capsid is formed from 90 dimers of VP1 that assemble into a $T = 3$ icosahedral cage approximately 38 nm in diameter (Figure 3.4A) (Prasad *et al.*, 1994b). VP1 monomers can occupy one of three quasi-equivalent positions (A, B or C), each with slight differences in conformation. A-type monomers surround the icosahedral five-fold axes and dimerise with B-type monomers to form AB-type dimers, whereas C-type monomers form CC-type dimers located on icosahedral two-fold axes (Figure 3.4A). The structure of VP1 monomers can be considered as an N-terminal arm (NTA), a shell (S) domain and a protruding (P) domain, which projects away from the icosahedral shell and comprises P1

and P2 subdomains (Figure 3.4B). The S domain is known to be critical and sufficient for the formation of an icosahedral structure and is formed from an eight-stranded antiparallel β-barrel, as is common with $T = 3$ icosahedral viruses (Bertolotti-Ciarlet *et al.*, 2002). The P1 subdomain typically comprises a four-stranded antiparallel β-sheet, along with three β-strands and an α-helix. The P2 subdomain is formed by an insert between the two separate segments that form P1, and comprises a six-stranded antiparallel β-barrel with variable loops connecting each strand. The P domain (especially the P2 subdomain) is known to be important for interactions with host cellular receptors and co-factors, including HBGAs and bile salts (discussed further in Section 4.1) (Tan *et al.*, 2004; Bhella *et al.*, 2008; Nelson *et al.*, 2018).



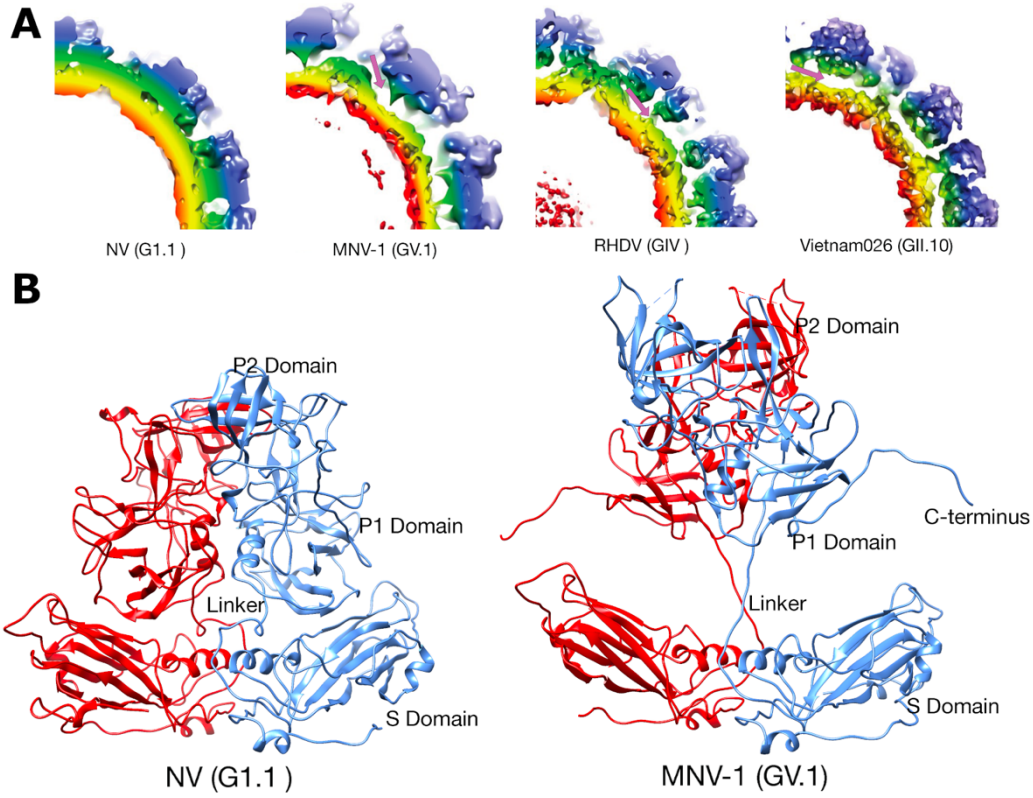
3.4. Norovirus capsid structural organisation. (A) *3.4 Å resolution X-ray crystal structure of the Norwalk virus capsid determined by (Prasad et al., 1999), coloured according to VP1 (sub)domain (blue – S domain; red – P1 subdomain; yellow – P2 subdomain). Examples of AB- and CC-type VP1 dimers are indicated by the green and purple markers, respectively.* ***(B)*** *Enlarged side view of a single VP1 dimer (N: N-terminus, C: C-terminus). Figure adapted from (Choi et al., 2008).*

The S and P domains of VP1 are connected by a flexible hinge region, which allows for variation in the positioning of the P domain relative to the shell. Specifically, most norovirus (and norovirus VLP) structures displayed a 'contracted' conformation with P domains close to the shell surface, but an

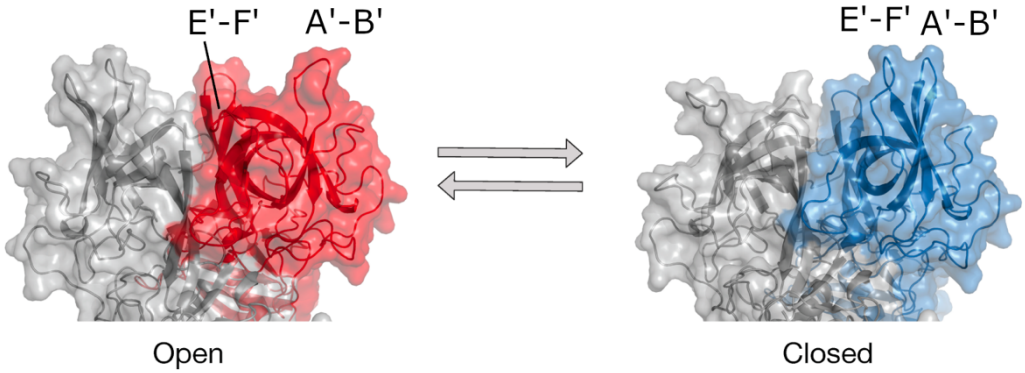
intermediate-resolution MNV structure adopted an ‘expanded’ conformation with P domains rotated and lifted ~16 Å away from the surface of the S domains (Prasad *et al.*, 1999; Katpally *et al.*, 2008, 2010). Other structures of rabbit haemorrhagic disease virus (RHDV, a lagovirus) and a GII.10 norovirus VLP appeared to show the P domains in an intermediate position between these two states (Figure 3.5) (Katpally *et al.*, 2010; Hansman *et al.*, 2012). At the time of commencing the research described in this chapter, the reasons for these differences were not clear, although the work described here, together with results reported by other research groups since, builds up a picture of P domains as highly dynamic structures that can switch between these states in response to external factors (Sherman *et al.*, 2019; Smith and Smith, 2019; Snowden *et al.*, 2020; Song *et al.*, 2020; Creutznacher *et al.*, 2021a; Williams *et al.*, 2021a). In addition to this form of flexibility (i.e., movement of the entire P domain relative to the shell), conformational changes within the P domain have also been observed. For MNV, the A’-B’ and E’-F’ loops of the P2 domain can adopt either an ‘open’ (loops splayed apart) or ‘closed’ (loops held together) conformation, with likely implications for receptor and antibody binding (Figure 3.6) (Taube *et al.*, 2010).

Aside from the capsid itself, an undefined amount of VP2 is known to be present in the virion, associated with the interior surface of VP1. Specifically, VP2 is known to interact with a highly conserved IDPWI motif in the Norwalk virus VP1 S domain (Vongpunsawad *et al.*, 2013). VP2 is known to be essential for the production of infectious virions, at least for feline calicivirus, so likely plays a critical role in virion assembly and/or initiation of infection (Sosnovtsev *et al.*, 2005). A key breakthrough was reported in 2019 with the discovery of a portal-like structure formed from VP2 in a cryoEM dataset of receptor-bound feline calicivirus (Conley *et al.*, 2019). It is thought that this VP2 portal-like assembly may provide a mechanism for calicivirus genome release, discussed in more detail in Section 4.1. Unfortunately, VP2 has not been visualised inside the capsid in any structures reported to date, likely due to symmetry mismatch with the icosahedral arrangement of VP1 and low signal intensity from the limited amount of VP2 present. Similarly, while the viral genome and its associated copy of VPg are also present in the virion,

these have not been resolved in any calicivirus structures reported to date, aside from low-resolution, radially averaged genome density. These issues are explored further in Chapter 4.



3.5. Norovirus capsid 'contracted' and 'expanded' conformations. (A) *Sectional view of density for the X-ray crystal structure of the Norwalk virus capsid (NV), and cryoEM structures of murine norovirus (MNV-1), rabbit haemorrhagic disease virus (RHDV) and a GII.10 human norovirus virus-like particle (Vietnam026) (Prasad et al., 1999; Katpally et al., 2008, 2010; Hansman et al., 2012). Purple arrows indicate the gap between P domain and S domain surface. Figure adapted from (Smith and Smith, 2019), published under a CC BY 4.0 licence.* ***(B)*** *Atomic model for a VP1 dimer from the X-ray crystal structure of the Norwalk virus capsid (NV) (Prasad et al., 1999) compared with the pseudo-atomic model based on the cryoEM structure of murine norovirus (MNV-1) (Katpally et al., 2008, 2010; Taube et al., 2010). A-type VP1 is coloured red, B-type VP1 is coloured blue. Figure adapted from (Smith and Smith, 2019), published under a CC BY 4.0 licence.*



3.6. MNV P domain 'open' and 'closed' conformations. *Atomic models and surface representations of an MNV VP1 P domain dimer in the 'open' (left, red) and 'closed' (right, blue) conformational states (Taube et al., 2010). The A'-B' and E'-F' loops of the P2 subdomain are indicated. Figure adapted from (Smith and Smith, 2019), published under a CC BY 4.0 licence.*

3.1.2. Norovirus disease and epidemiology

Upon infection by norovirus and following an incubation period of 12 – 48 hours, patients typically develop acute gastroenteritis, with only about one third of patients remaining asymptomatic (Glass *et al.*, 2009; Ahmed *et al.*, 2014). Clinical symptoms of norovirus gastroenteritis often include vomiting, abdominal cramps, watery diarrhoea and fever, sometimes accompanied with headache, chills and muscle pain, and while the majority of cases are self-resolving within 2 – 3 days, severe outcomes including hospitalisation and death can occur (Glass *et al.*, 2009). In immunocompromised patients, disease can persist for much longer periods of time, resulting in more complicated clinical outcomes (Roddie *et al.*, 2009; Bok and Green, 2012). Aside from gastroenteritis, norovirus has also been implicated in rare cases of pneumatosis intestinalis, necrotising enterocolitis, seizures and encephalopathy (together with detection of the viral genome in the cerebrospinal fluid by reverse transcription polymerase chain reaction [RT-PCR]) (Ito *et al.*, 2006; Turcios-Ruiz *et al.*, 2008; Medici *et al.*, 2010; Kim *et al.*, 2011).

Transmission of norovirus is most commonly faecal-oral or oral-oral in nature, and can be either direct or indirect (e.g., through exposure to contaminated food, water or environmental fomites). Noroviruses are highly stable in the environment, and infectious Norwalk virus has been recovered after incubation in groundwater for 61 days (Seitz *et al.*, 2011). Furthermore, a very small quantity of virus particles is required to initiate an infection. One human challenge study reported a 50% infectious dose (ID_{50}) of just 18 viruses through mathematical modelling of the dose-response relationship (Teunis *et al.*, 2008). Virus shedding from infected patients peaks around 2 – 5 days post-infection, though viral genomes have been detected in stool samples from otherwise healthy people long after the resolution of symptoms (Atmar *et al.*, 2008). It is not clear if extended shedding contributes significantly to transmission, though it was demonstrated that most infections derive from symptomatic individuals, at least in a nosocomial setting (Sukhrie *et al.*, 2012).

Norovirus disease is a global problem. Noroviruses are estimated to cause 684 million cases of diarrhoeal disease worldwide, with GII.4 strains responsible for the majority of outbreaks (Glass *et al.*, 2009; Pires *et al.*, 2015). In high-income countries (HICs), norovirus outbreaks are responsible for significant economic costs, with bed closures and staff absences costing the UK's National Health Service (NHS) around £35–49 million each year (Sandmann *et al.*, 2017). Globally, the annual societal cost of norovirus gastroenteritis is estimated to be $60 billion (Bartsch *et al.*, 2016). Norovirus mortality is non-negligible in HICs, with an estimated 56 deaths per year due to norovirus infection in the UK (Holland *et al.*, 2020), but this is a much more significant issue in low- and middle-income countries (LMICs) with around 200,000 norovirus deaths each year among children under 5 years old (Patel *et al.*, 2008).

3.1.3. Norovirus treatment and vaccine prospects

3.1.3.1. Current treatment options

In spite of the scale of morbidity and mortality caused by noroviruses, we still lack any licensed vaccines or specific antiviral agents. Current treatment is essentially limited to management of symptoms and maintaining hydration by replacing lost fluids and ensuring adequate nutritional uptake (King *et al.*, 2003). If a patient presents with indications of dehydration, oral rehydration solutions (often comprising different carbohydrates, sodium, potassium, chloride and bicarbonate ions) are administered to rehydrate the patient while maintaining an appropriate balance of electrolytes and providing nutritional support (King *et al.*, 2003). In severe cases of dehydration, or dehydration complicated with severe vomiting, intravenous fluids can be administered (King *et al.*, 2003).

3.1.3.2. Anti-noroviral agents under development

Most antiviral agents against norovirus are still in the early preclinical development stages (Netzler *et al.*, 2019). The only agent to have completed clinical trials, nitazoxanide (NTZ), was observed to reduce the duration of norovirus-related gastroenteritis symptoms in a phase II randomised double-blind trial, but a number of case reports suggest that it is ineffective against chronic norovirus infection (Rossignol and El-Gohary, 2006; Capizzi *et al.*, 2011; Kempf *et al.*, 2017).

One barrier to drug design is that the cellular receptor for HuNVs remains unknown. Without knowledge of the receptor, it is difficult to design specific compounds that work by blocking cellular entry, though a number of ‘attachment blockers’ that inhibit the interaction between VP1 and important attachment factors, HBGAs (discussed further in Section 4.1), have been identified (Zhang *et al.*, 2013; Koromyslova *et al.*, 2015, 2017). Monoclonal antibodies (mAbs) and nanobodies (Nbs) have also been investigated. While numerous mAbs and Nbs with attachment inhibition activity were observed to be relatively strain-specific (for example, through ELISAs to check for binding

against VLPs derived from noroviruses of different genotypes), one Nb with the ability to block HBGA attachment (Nano-85) showed broad binding specificity in ELISAs and isothermal titration calorimetry (ITC) experiments, and was observed to induce particle disassembly in negative stain EM analysis (Chen *et al.*, 2013; Koromyslova and Hansman, 2015, 2017; Lindesmith *et al.*, 2015). So far, studies of mAbs and Nbs have been limited to analysis of inhibition of VLP binding.

Other antiviral agents targeting key viral proteins such as the RdRp and viral cysteine protease are also under development. One subclass of RdRp inhibitors are nucleoside analogues, such as 2'-C-methylcytidine (2CMC) and its derivatives, which have been shown to protect from mortality and diarrhoea, and inhibit MNV replication in a knockout mouse model (Costantini *et al.*, 2012; Rocha-Pereira *et al.*, 2013; Kolawole *et al.*, 2016). The purine analogue T-705 (favipiravir) is already licensed for use against influenza in Japan and has been investigated for activity against SARS-CoV-2, but is thought not to be sufficiently potent against norovirus for further development, given a therapeutic index (i.e., ratio of the median toxic dose to the median effective dose) of just 4.3 was calculated from CPE-reduction and MTS assays in cell culture (Rocha-Pereira *et al.*, 2012; Furuta *et al.*, 2017; Netzler *et al.*, 2019; Du and Chen, 2020). Non-nucleoside inhibitors bind to the RdRp allosterically, and are usually limited in their spectrum of activity. A number of therapeutics already approved for use against other conditions are being investigated for potential repurposing as anti-norovirus nucleoside inhibitors (Mastrangelo *et al.*, 2012; Netzler *et al.*, 2017). Most compounds under development as specific anti-noroviral agents are viral protease inhibitors, including both transition state mimics (such as α-hydroxyphosphonates) and other inhibitors (including peptide-based and macrocyclic compounds) (Tiew *et al.*, 2011; Mandadapu *et al.*, 2013; Galasiti Kankanamalage *et al.*, 2015; Damalanka *et al.*, 2016; Netzler *et al.*, 2019).

3.1.3.3. Vaccine prospects

While antiviral agents would be useful for treating patients that have already acquired norovirus infections, an effective vaccine could prevent infection outright, and it is likely that both approaches would be needed to control the global burden of norovirus mortality and morbidity.

A number of norovirus vaccines are currently under development, as summarised in Table 3.2. They can be divided into three types: virus-like particle (VLP)-based vaccines, P particle-based vaccines, and recombinant adenovirus vector-based vaccines. VLP-based vaccines rely on the expression of norovirus VP1, which self-assembles into capsid structures in the absence of genome (Jiang *et al.*, 1992; Prasad *et al.*, 1999). P particles are small 24-mers of the VP1 P domain alone, which are stable and can be produced in *E. coli* (Tan *et al.*, 2011). Recombinant adenovirus vector-based vaccines use modified adenoviruses to express norovirus VP1 in host cells (Guo *et al.*, 2008). Notably, a number of VLP- and P particle-based vaccine candidates have been designed to include antigens from other viruses, such as rotavirus (Blazevic *et al.*, 2011). Not only was this found to induce an immune response against both norovirus and rotavirus, but rotavirus VP6 acted as an adjuvant to improve the immune response against norovirus (Blazevic *et al.*, 2011; Lappalainen *et al.*, 2015; Malm *et al.*, 2019). Thus, combination vaccines are a promising approach.

Table 3.2. Main norovirus vaccines in development. *Table adapted and updated from (Esposito and Principi, 2020).*

Vaccine type	Preclinical	Phase I	Phase II
Virus-like particle (VLP)	*Vaccine Research Center in Tampere, Finland* NoV GII.4 and GI.3 + oligomeric rotavirus VP6		*Takeda* Intramuscular, bivalent GI.1 + GII.4
	Institute Pasteur of China NoV GII.4 + enterovirus 71		*National Vaccine and Serum Institute, China* Intramuscular, bivalent GI.1 + GII.4
		Anhui Zhifei Longcom Quadrivalent NoV GI.1 + GII.3 + GII.4 + GII.17	
Recombinant adenovirus		*Vaxart* Oral, monovalent GI.1, monovalent GII.4, or bivalent GI.1 + GII.4	
P particles	*Several research centres* Various vaccines combining NoV VP1 P domain with RV, HEV, influenza virus and astrovirus		

While several norovirus vaccines are in the development pipeline, significant barriers to vaccine development remain. Firstly, natural immunity developed in response to norovirus infection is generally short-lived, with loss of

protection from disease reported 2 – 4 years after the initial infection (though mathematical models estimate the duration of immunity to be somewhat longer than this) (Parrino *et al.*, 1977; Simmons *et al.*, 2013). There is also considerable genetic diversity among (and within) norovirus genogroups, and immune responses to norovirus infection are often unable to provide cross-protection against infection from other genotypes, even within the same genogroup (Karangwa *et al.*, 2017). GII noroviruses are the predominant genogroup circulating globally, responsible for ~92% of norovirus acute gastroenteritis cases, with GI noroviruses accounting for ~8% of cases and GIV <0.1% (van Beek *et al.*, 2018). Thus, monovalent vaccines are unlikely to generate the cross-protective immunity required, such that most vaccine candidates now include more than one antigenic target (Esposito and Principi, 2020). Even within individual genotypes, antigenic drift and genetic recombination can lead to the appearance of new variants that escape existing immunity. This is particularly evident for the GII.4 genotype, with new GII.4 variants emerging and circulating approximately every 2-3 years since the 1990s (Siebenga *et al.*, 2007, 2009; Parra *et al.*, 2017). One approach designed to overcome this issue was the engineering of a 'consensus' GII.4 VLP with epitopes from three different variants (2006a [Yerseke], 2006b [Den Haag], 2002 [Houston]), which appears to generate broader cross-reactivity (Parra *et al.*, 2012).

3.1.4. Project aims

A key resource for developing our understanding of viral infection and for the rational design of antiviral agents and vaccines is high-resolution structural data on the virion. The work reported in this chapter aimed to demonstrate the utility of cryoEM in generating such structural information for norovirus, using MNV as a model system. Specifically, by combining standard cryoEM image processing with a focussed classification approach, the aim of this study was to generate a complete atomic model for MNV VP1 and address a number of unresolved questions focussing on the VP1 capsid, including:

- Why have different P domain conformations ('contracted' and 'expanded') been observed for different norovirus genotypes?
- Can individual norovirus genotypes display both of these conformations?
- What are the implications of P domain conformational changes for virus function?

In addition, through structural characterisation of MNV following heat inactivation and a thermally stabilised mutant MNV, I aimed to elucidate further alternative capsid conformations that may give mechanistic insight into the early stages of the viral lifecycle.

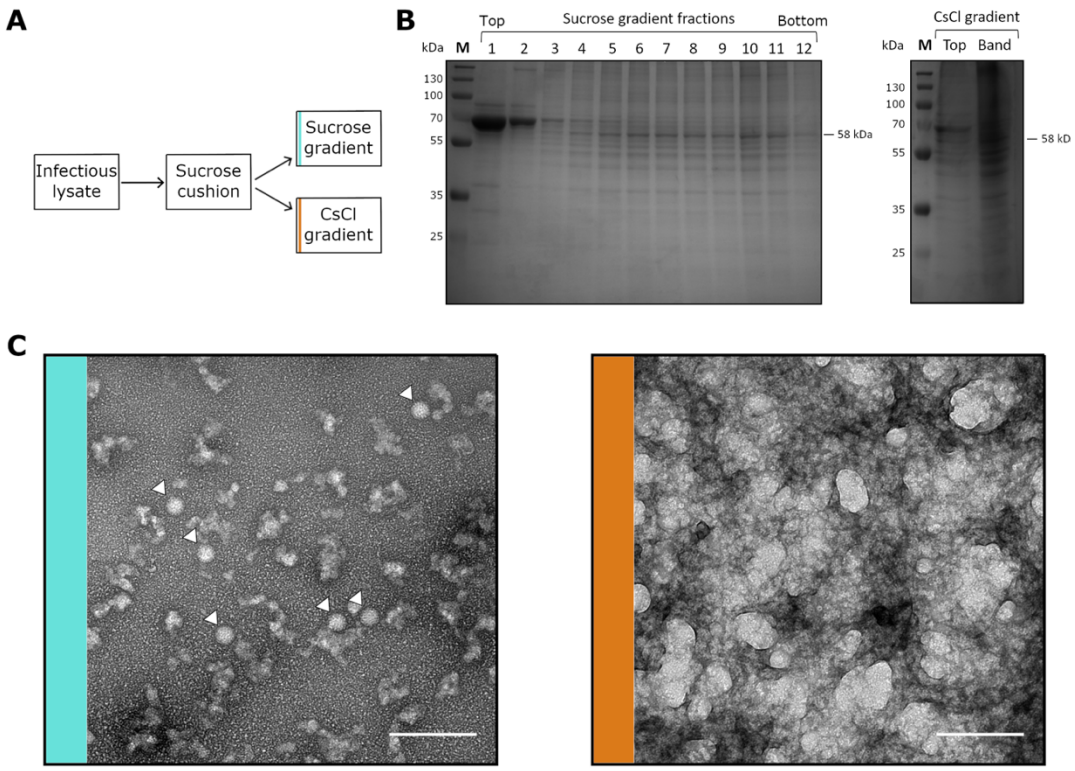
3.2. Optimisation of virus purification

MNV was purified from the lysate of infected cells for structural studies. In the first instance, four different purification strategies were compared based on both purity and yield. Each processing strategy comprised a number of initial processing steps, followed by different permutations of two gradient centrifugation stages (with each stage consisting of either a continuous 15 – 60% sucrose gradient [described in Methods 2.5.5] or a stepped 1.2 – 1.4 g cm^{-3} CsCl gradient). I.e., the purification strategies tested were:

(A) Sucrose gradient + sucrose gradient (S+S);
(B) Sucrose gradient + CsCl gradient (S+C);
(C) CsCl gradient + sucrose gradient (C+S);
(D) CsCl gradient + CsCl gradient (C+C).

The lysate of 16 × T-175 flasks of wild type (wt)MNV-infected cells was clarified, then pelleted through a sucrose cushion as described in Section 2.5.5. Virus-containing pellets were resuspended in PBS and split into two equal volumes, which were loaded onto either a continuous sucrose gradient or a stepped CsCl gradient for ultracentrifugation (Figure 3.7A). While sucrose density gradients separate components of the sample by differences in their sedimentation rate, CsCl gradients separate components on the basis of their

density (isopycnic centrifugation). Sucrose gradient ultracentrifugation was performed and gradients fractionated (Section 2.5.5). CsCl gradients were spun at 155,000 × *g* for 16 hours (15°C), then the visible band at the density step boundary was extracted. Samples were examined by SDS-PAGE to give an overview of sample concentration (judged by the strength of the stained band at the expected molecular weight of the target) and sample purity (judged by the absence of bands at other molecular weights). This showed the presence of a band corresponding to the molecular weight of VP1 (58 kDa) in the sample extracted from the CsCl gradient and spread across a large number of sucrose gradient fractions (Figure 3.7B). The CsCl gradient purified sample was less pure than the peak sucrose gradient fractions (based on the number and intensity of other protein bands). Negative stain EM analysis, which is able to provide information on sample purity, concentration, and whether or not particles are intact, supported this finding, with MNV particles easily discernible in peak fractions from the sucrose gradient-purified sample, but not for the CsCl gradient-purified sample (Figure 3.7C).

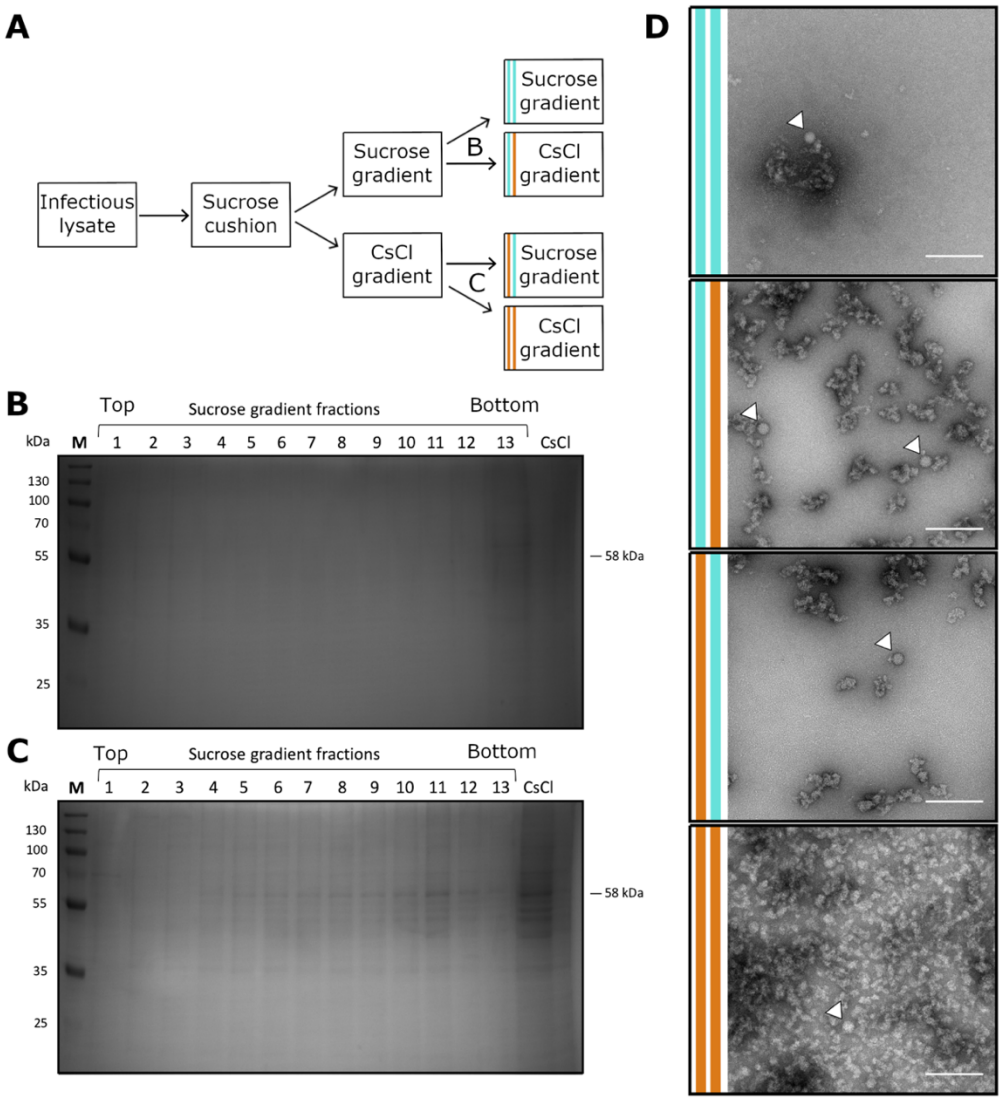


3.7. Interim analysis of MNV purification strategies. (A) *Flow chart showing how clarified infectious lysate samples were treated prior to interim analysis.* ***(B)*** *10% SDS-PAGE analysis of samples taken from sucrose gradient (left) or CsCl gradient (right), stained with Coomassie Blue. Individual fractions of equal volume were analysed from the top (1) to bottom (12) of the sucrose gradient. For the CsCl gradient, a sample from the top of the gradient (Top) was compared with the visible band of material extracted from the step boundary (Band). The expected molecular weight of MNV VP1 (58 kDa) is indicated.* ***(C)*** *Representative negative stain EM micrographs from a peak fraction of sucrose gradient-purified MNV (left, blue) and the from the band of material extracted from the step boundary of the CsCl gradient (right, orange), after dialysis into EM buffer. White arrowheads indicate MNV particles. Scale bars indicate 200 nm.*

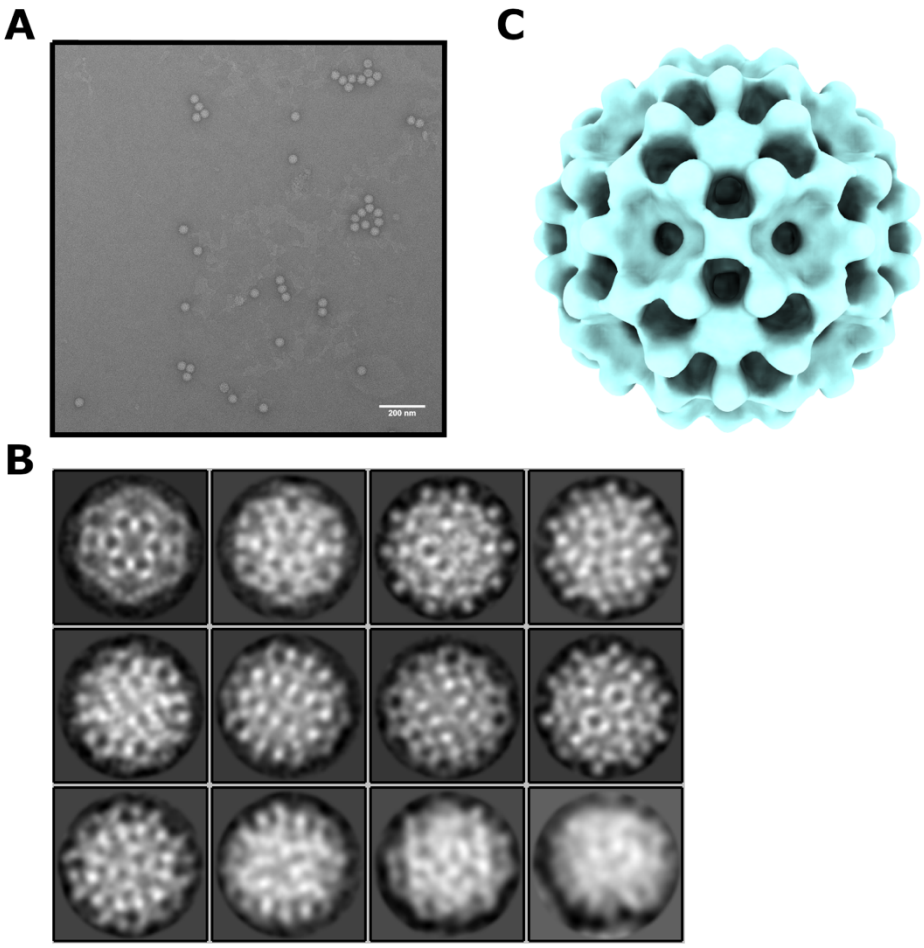
Peak fractions from sucrose and CsCl gradients were dialysed in EM buffer to remove sucrose/CsCl, then each split into two equal volumes for the second round of gradient ultracentrifugation (Figure 3.8A). It was difficult to discern any protein bands for ‘S+S’ and ‘S+C’ samples by SDS-PAGE, suggesting that a greater amount of material had been lost (Figure 3.8B), whereas protein

bands at 58 kDa were visible for ‘C+S’ and ‘C+C’ samples (Figure 3.8C). Negative stain EM analysis (Figure 3.8D) showed that the ‘C+C’ strategy was the least effective at removing residual cellular material, and in contrast to the results of SDS-PAGE analysis, it was difficult to find intact MNV particles. This could have been a result of MNV particles being hidden by residual cellular material, or because a proportion of MNV particles had dissociated during purification (which would still contribute to the signal of the VP1 band in SDS-PAGE, as was observed). Based on qualitative assessment of a large area of the EM grid for each strategy, I concluded that the ‘S+S’ strategy gave the best balance of purity and concentration.

To verify that the ‘S+S’ strategy was appropriate for preparing virus for structural work, this strategy was used to purify a fresh batch of wtMNV-infected cell lysate (from 10 × T-175 flasks). Instead of dialysis between the first and second sucrose gradients, sucrose was removed by pelleting and resuspending virus in PBS. For all other steps, the strategy was unmodified. The purified sample was concentrated by ultracentrifugation (366,000 × *g*, 1 hour, 4°C) and resuspension in a small volume of EM buffer, then analysed by negative stain EM. MNV particles were visible and sufficiently concentrated for further structural analysis of negative stain EM data (Figure 3.9A). 257 particles were manually picked from 30 micrographs, then subjected to 2D and 3D classification, and 3D refinement (with the imposition of icosahedral symmetry) in RELION-2.1. The resultant 2D classes (Figure 3.9B) and ~22-Å resolution 3D reconstruction (Figure 3.9C) had the expected appearance for MNV – in particular, protruding domains were observed in a $T = 3$ quasi-equivalent arrangement. This result confirmed that the ‘S+S’ purification strategy generated virus samples suitable for structural analysis, and therefore this approach was taken forward for subsequent structural work.



3.8. Comparison of MNV purification strategies. (A) *Flow chart showing how clarified infectious lysate samples were treated prior to analysis.* ***(B,C)*** *10% SDS-PAGE analysis of sucrose gradient fractions (top [1] to bottom [13]) and the band of material extracted from a CsCl gradient (CsCl), stained with Coomassie Blue. Sucrose gradient/CsCl gradient purification was performed after an initial sucrose gradient* ***(B)*** *or after an initial CsCl gradient* ***(C)****, as indicated in panel (A). The molecular weight of MNV VP1 (58 kDa) is indicated.* ***(D)*** *Representative negative stain EM micrographs from peak fractions of each condition (indicated by the coloured stripes, corresponding to panel (A); from top to bottom: S+S, S+C, C+S, C+C). White arrowheads indicate MNV particles. Scale bars indicate 200 nm.*



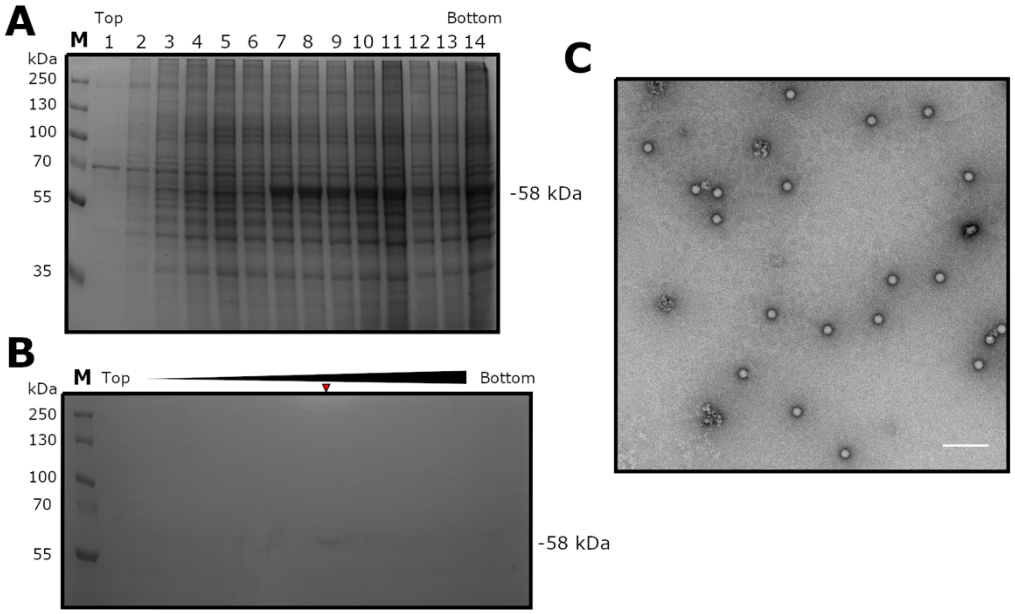
3.9. Negative stain EM of wtMNV purified by the 'S+S' strategy. (A) *Representative negative stain micrograph of wtMNV purified by 'S+S' strategy, followed by concentration by pelleting and resuspension in a small volume of EM buffer. Scale bar indicates 200 nm.* ***(B)*** *The top 12 most populated 2D classes of manually picked wtMNV particles.* ***(C)*** *The 22-Å resolution 3D reconstruction of wtMNV.*

3.3. The structure of wild-type (wt)MNV

3.3.1. Sample preparation and vitrification for cryoEM

A larger stock of wtMNV was generated for high-resolution structural work. Briefly, a crude stock of wtMNV was passaged in RAW264.7 cells to obtain an infected cell lysate from 15 × T-175 flasks, from which virus was purified by the 'S+S' strategy. Peak fractions were identified by SDS-PAGE analysis (Figure 3.10A,B) and dialysed into EM buffer. The quality of the sample was

confirmed by negative stain EM (Figure 3.10C). The stoichiometry of VP2 and gRNA within the particles was not characterised and was therefore unknown, although the sample was infectious, suggesting at least a proportion of virions contained intact VP2 and gRNA.



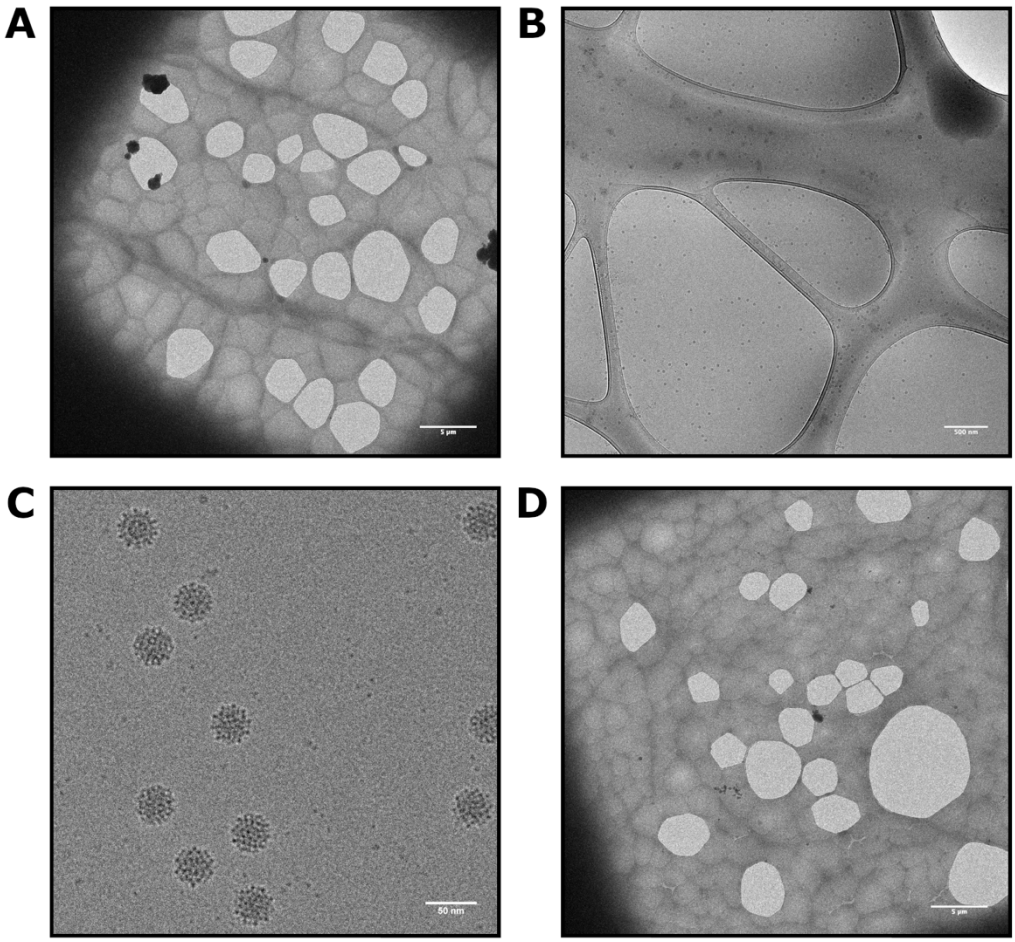
3.10. wtMNV purification for structural determination by cryoEM. (A) *10% SDS-PAGE analysis of fractions from first sucrose gradient purification of wtMNV for cryoEM. Numbers correspond to fractions from top (1; 15% sucrose) to bottom (14; 60% sucrose) of the gradient. The molecular mass of VP1 (58 kDa) is indicated.* ***(B)*** *12% SDS-PAGE analysis of fractions (from top to bottom) from second sucrose gradient purification of wtMNV for cryoEM. The molecular mass of VP1 (58 kDa) is indicated. The red arrowhead indicates the fraction selected as the 'peak fraction' and imaged in panel (C).* ***(C)*** *Representative negative stain EM micrograph of the peak fraction used for structural determination of wtMNV by cryoEM. Scale bar represents 200 nm.*

For cryoEM, grids coated with a lacey carbon film, covered in an ultra-thin (<3 nm) continuous layer of amorphous carbon were used. This grid type was chosen to maximise the concentration of particles available for imaging on the grid surface, as it permitted a 'multiple application' approach (described previously (Hurdiss *et al.*, 2018)). With each separate application of sample,

more MNV particles adhered to the continuous carbon film. A range of conditions was tested for the preparation of cryoEM grids. As well as different blotting times and blotting forces (which is thought to alter the ice thickness on the grid surface), different glow discharge conditions were examined. Grids were glow discharged under air (leading to a negatively charged grid surface), under amylamine vapour (leading to a positively charged grid surface), or used without glow discharge treatment prior to sample application and vitrification. Screening of grids at different magnifications was performed using an FEI Titan Krios microscope. Grids were evaluated based on ice quality and particle concentration/distribution, judged by visual inspection of micrographs (Figure 3.11). Overall ice quality was poor on grids with no glow discharge as might be expected, given the surface would remain hydrophobic, although some grid squares were deemed suitable for imaging. There was no clear difference between grids glow discharged under air and grids glow discharged under amylamine. A grid with optimal particle distribution and ice quality was selected for data collection.

3.3.2. wtMNV data collection

The optimised wtMNV grid was imaged using an FEI Titan Krios microscope at 300 kV, at a magnification of 75,000× and a calibrated object sampling of 1.065 Å/pixel. The total electron dose per image was 59 $e^-/Å^2$, distributed across 59 frames over 1.5 seconds. For each image, between −0.5 and −2.9 µm defocus was applied to generate phase contrast. Full details of parameters used in data collection are given in Table 3.3. Ultimately, 13,692 micrograph movies (image stacks) were collected and taken forward for image processing.



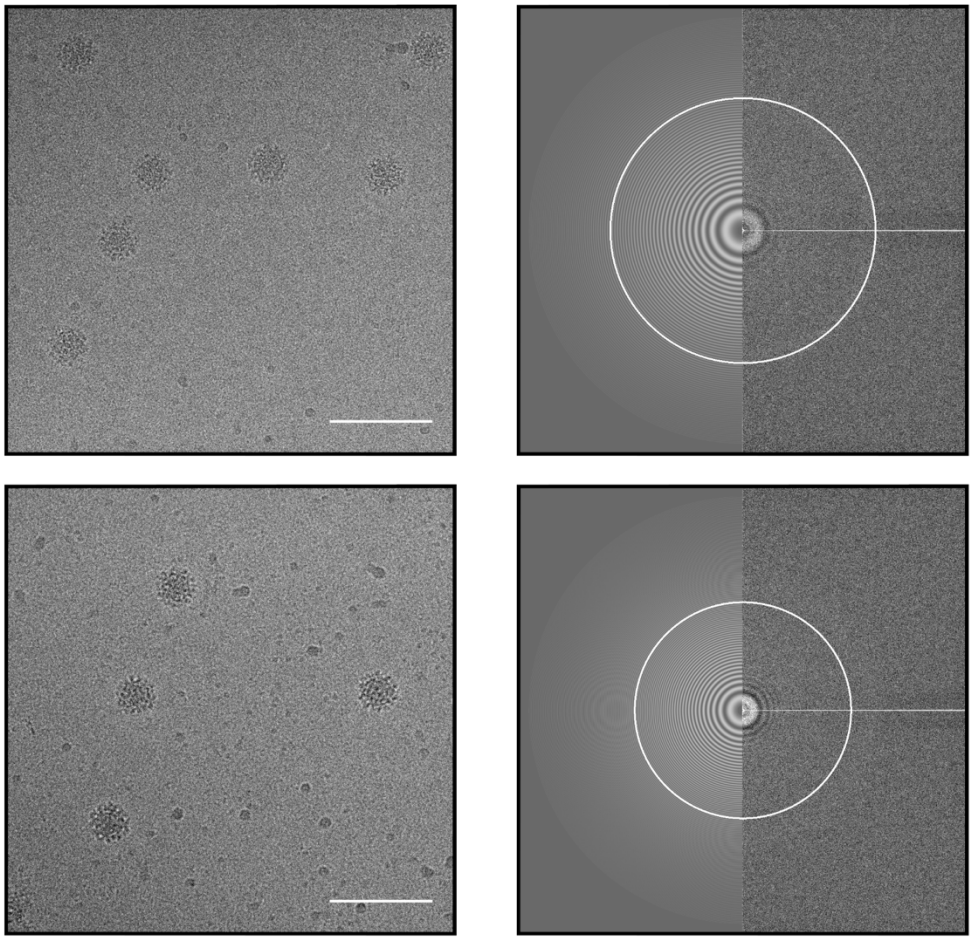
3.11. Screening of grids for wtMNV structure determination. (A-C) *Representative screening images of wtMNV cryoEM grids glow discharged under amylamine vapour. Scale bars represent* ***(A)*** *5 µm,* ***(B)*** *500 nm, and* ***(C)*** *50 nm.* ***(D)*** *Representative screening image of wtMNV cryoEM grids glow discharged under air. Scale bar represents 5 µm.*

Table 3.3. Data collection parameters for wtMNV data collection.

Microscope	FEI Titan Krios
Camera	Falcon III
Voltage (kV)	300
Pixel size (Å)	1.065
Nominal magnification	75,000×
Exposure time (s)	1.5
Total dose (e^-/Å^2)	59
Number of fractions	59
Defocus range (μm)	−0.5 to −2.9
Number of micrographs	13,692
Acquisition software	FEI EPU

3.3.3. The 3.1 Å resolution structure of wtMNV

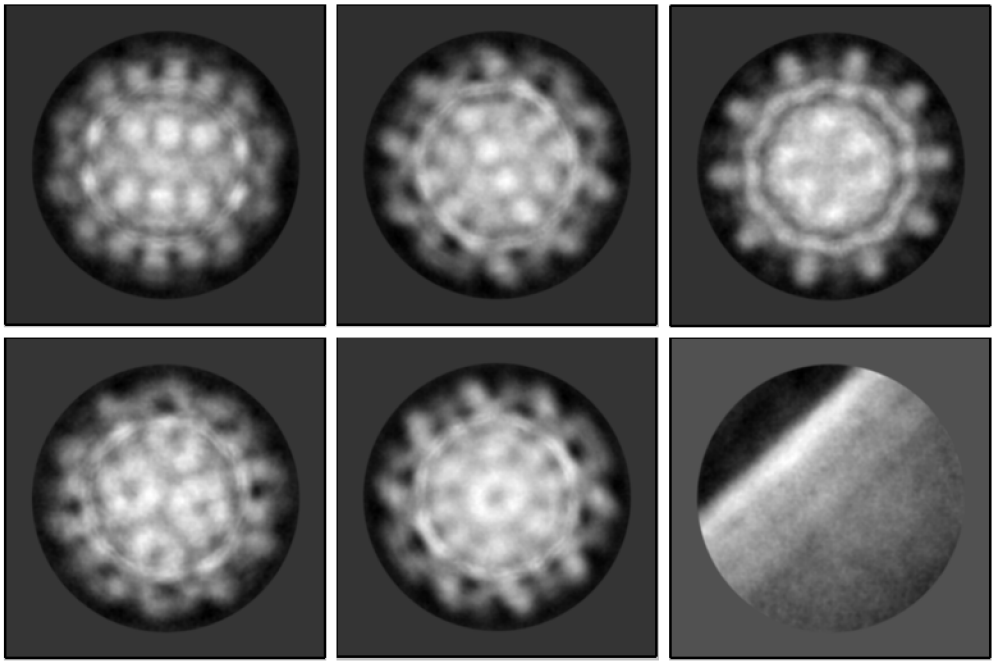
Image processing was initially performed using the Relion-2.1 pipeline, although Relion-3.0 was used for later stages following its release. Raw micrograph movie frames were first aligned and corrected for drift using MOTIONCOR2 (Zheng *et al.*, 2017), then Gctf (Zhang, 2016) was used to estimate the contrast transfer function of each micrograph (Figure 3.12). A small subset of particles was manually picked from these corrected micrographs and subjected to reference-free 2D classification, with class averages reminiscent of MNV virions taken forward as templates for autopicking in Relion. This yielded ~50,000 particles, though the majority of these corresponded to thick carbon edges and poor-quality particles (e.g. virus images over thick carbon or obscured by ice contamination). Manual removal of these particles left ~13,000 particles for further processing.



3.12. Gctf determination of CTF for wtMNV micrographs. *Representative motion- and CTF-corrected micrographs are shown on the left, scale bars represent 100 nm. For each, the output from Gctf is shown (right hand images). Gctf output includes the actual micrograph FFT (right half of image) and the CTF fitting determined by Gctf (left half of image).*

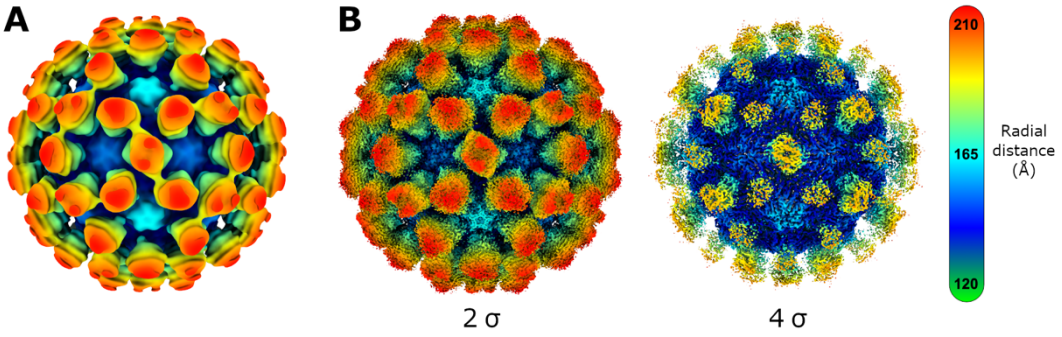
Firstly, particles were subjected to several rounds of reference-free 2D classification with poor quality classes excluded at each stage, leaving ~10,000 particles. Representative 2D classes are shown in Figure 3.13. These were combined with a small number of particles (~1,000) derived from a prior data collection on wtMNV (using the same imaging parameters, carried out by Dr Daniel Hurdiss [DH]), and used for a preliminary 3D refinement based on an initial model that was generated *de novo* (Figure 3.14A), then low-pass filtered to a resolution of 60 Å to prevent model bias. Icosahedral symmetry (I1) was imposed for both initial 3D model generation and refinement. Using

the ‘gold standard’ FSC criterion (FSC=0.143), the refined map was resolved to 3.8 Å after sharpening (Figure 3.14B).

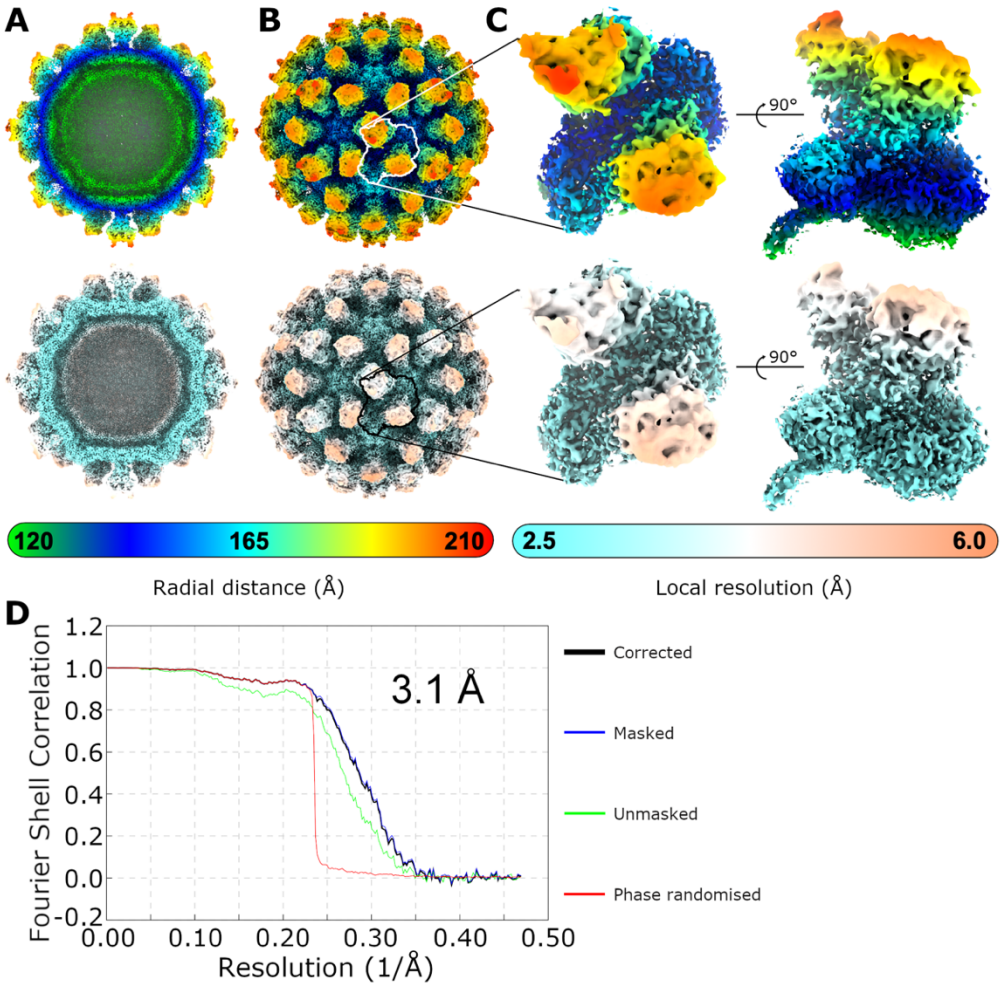


3.13. Representative 2D class averages from wtMNV data collection. *The first five classes (left to right, top to bottom) show wtMNV in different orientations, all taken forward for further processing. The bottom right class (discarded) contains thick carbon edges from the grid, picked inadvertently during autopicking.*

Two approaches were taken to improve the resolution: (i) iterative rounds of CTF refinement, Bayesian polishing and masked 3D refinement, and (ii) 3D classification without alignment. Together, these approaches yielded a 3.1 Å reconstruction from 7,811 particles (Figure 3.15). Local resolution estimation revealed S domain density to be better resolved than P domain density (~2.9 Å and ~3.7 Å, respectively), and resolution worsened in the P2 subdomain to ~5.0 – 6.0 Å (Figure 3.15), likely a result of increased flexibility in these regions.



3.14. 3D initial model and preliminary 3D reconstruction of wtMNV. (A) *Initial model used for 3D refinement of wtMNV.* ***(B)*** *Preliminary refined density map for wtMNV at 3.8 Å resolution, contoured to different thresholds. All maps were flipped in the Z plane to show the correct orientation, and are coloured according to the radial colouring scheme shown.*



3.15. Final density map for wtMNV. (A) *Cross section of wtMNV density map, filtered according to local resolution.* ***(B)*** *Full isosurface representation*

of wtMNV density map, filtered according to local resolution. The asymmetric unit is outlined. ***(C)*** *Enlarged density for asymmetric unit of wtMNV. All maps were shown at 2 σ, flipped in the Z plane to show the correct orientation, and coloured according to the radial colouring scheme (top) or local resolution colouring scheme (bottom) shown.* ***(D)*** *FSC plot for the wtMNV map. The resolution (3.1 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).*

3.3.4. Focussed classification of wtMNV P domain dimers

In an attempt to improve the relatively poor resolution of the P domain dimers, a focussed 3D classification approach was taken. A cylindrical mask was applied over an AB- or CC-type P domain dimer and used to classify a symmetry-expanded stack of particles (i.e., a stack containing 60 copies of each particle, with each copy assigned to one of 60 symmetrically redundant orientations), such that each symmetrically related P domain dimer in each particle was classified independently. Masks were positioned to include only a small section of the S domain, ensuring that classification was driven by P domain differences (Figure 3.16A,C). Unlike most standard 3D classification procedures, focussed classification does not involve any alignment of particles, such that there is no risk of 'overfitting'. As such, a greater value (20, compared to the default 4) was used for the regularisation parameter ('T' number) to increase the weighting given to the data relative to the reference model, and therefore increase the likelihood of being able to discern subtle conformational differences. Furthermore, the low-pass filter applied to the reference model was set to the global resolution of the map (3.1 Å) to ensure high-resolution differences were considered. In order to balance the need for sufficient 'space' to discern the full range of conformational variation, with the limitations imposed by the computational resources available, subparticles were split into a total of 10 classes.

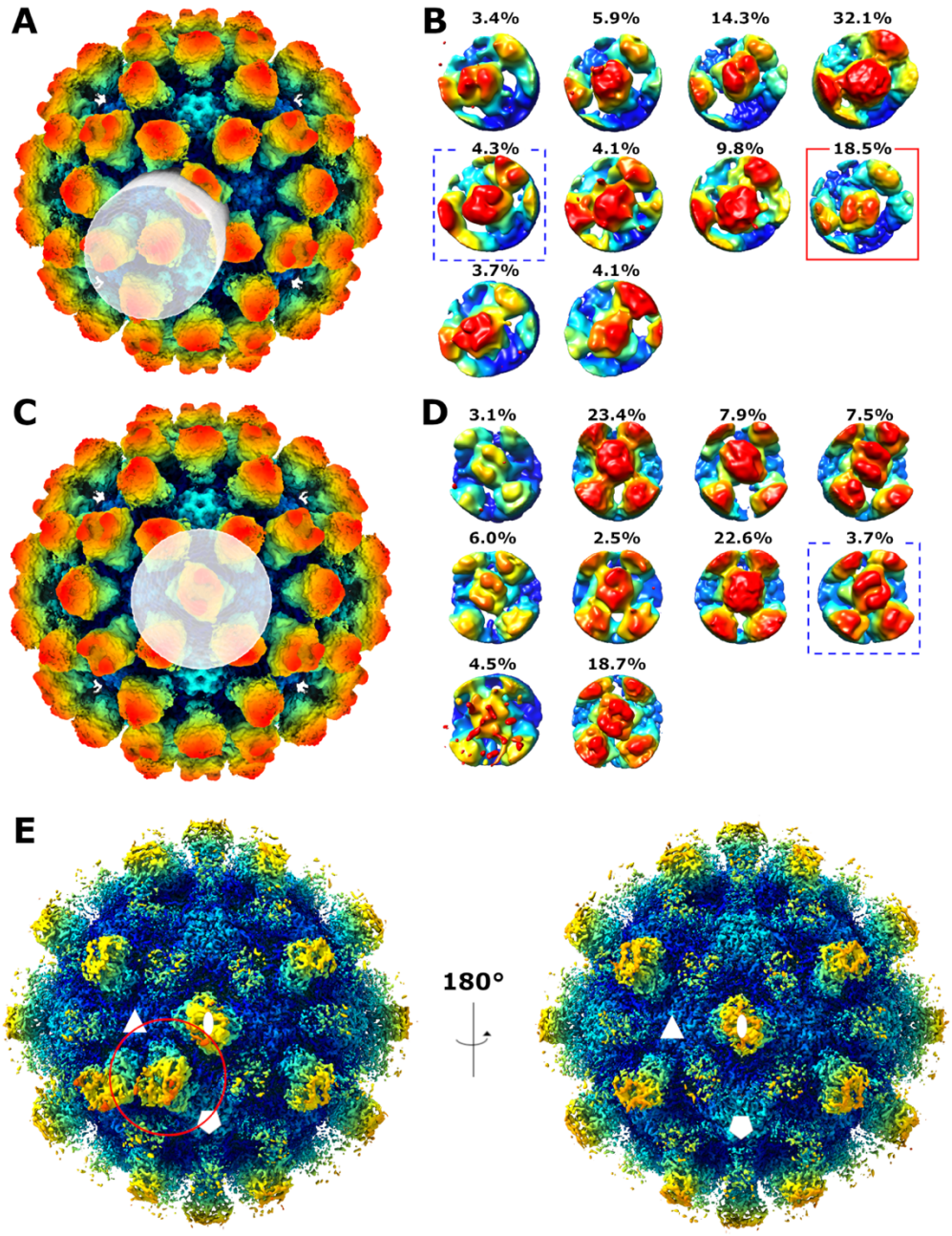
Focussed classes of AB-type P domain dimers showed clear variation in the positioning of the P domain dimer relative to adjacent P domain dimers (Figure

3.16B). In particular, there was unambiguous evidence for both rotation of the P domain dimer as well as movement in a plane parallel with the S domain surface. Conversely, CC-type P domain dimers showed less variation than AB-type P domain dimers, suggesting they were positioned more uniformly (Figure 3.16C,D). This correlated with the higher resolution of CC-type P domain dimers and the ability to resolve the flexible linker for C-type P domains in the icosahedrally averaged reconstruction of wtMNV. Interestingly, this contrasts with data reported for RHDV and Tulane virus, two other caliciviruses which were shown to have CC-type P domain dimers that were more mobile than AB-type dimers (Katpally *et al.*, 2010; Yu *et al.*, 2013a).

For both AB- and CC-type P domain dimers, all 10 classes were occupied with the least populated classes containing 3.4% (AB) or 2.5% (CC) of the total data. Most classes contained <10% of the total data, with most subparticles assigned to just one or two classes in each case. The most populated classes contained 32.1% (AB) or 23.4% (CC) of the data, though these classes were not well defined morphologically, suggesting several different conformations had been grouped together (Figure 3.16B,D). This is in contrast to other reasonably well-populated classes (~15-20% of total data) that were relatively well defined. Together, this is suggestive of a conformational landscape that individual P domain dimers are exploring continuously and independently (even within the same capsid), with certain conformations being more commonly adopted. A full capsid reconstructed from the subparticles contained within a single AB-type focussed class (without any modification of alignments or imposition of symmetry) revealed a single well-resolved AB-type P domain dimer at the site which had been contained within the mask during focussed classification, with all other AB-type P domain dimer density appearing fragmented and poorly resolved (Figure 3.16E). This supported the notion that P domain dimers are able to explore the conformational landscape independently of each other. It should be noted that for both AB- and CC-type P domain dimer focussed classifications, one class was mirrored in the *Z*-axis.

While the results of focussed classification were informative, there was no significant improvement in P domain resolution compared to the icosahedrally

averaged wtMNV reconstruction. This is likely because any potential improvements in resolution from increased homogeneity were counterbalanced by the reduction in the total number of subparticles contributing to the final reconstruction, having been divided between 10 classes.



3.16. Focussed classification of wtMNV P domain dimers. ***(A,C)*** *Positioning of mask used for focussed classification of* ***(A)*** *AB-type or* ***(C)*** *CC-type P domain dimers.* ***(B,D)*** *All focussed classes from focussed*

*classification of **(B)** AB-type or **(D)** CC-type P domain dimers. All classes are shown at ~2 σ and coloured according to height. Classes displaying an inverted Z orientation are highlighted by a dashed blue box. The proportion of P domain dimer 'subparticles' assigned is indicated above each class. **(E)** Full asymmetric reconstruction from particles contained in a single AB-type P domain dimer class (shown by the red box in (B)), displayed at ~2.8 σ with a radial colouring scheme. The two- (oval), three- (triangle) and five-fold (pentagon) icosahedral axes are indicated. (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.3.5. An atomic model for wtMNV VP1

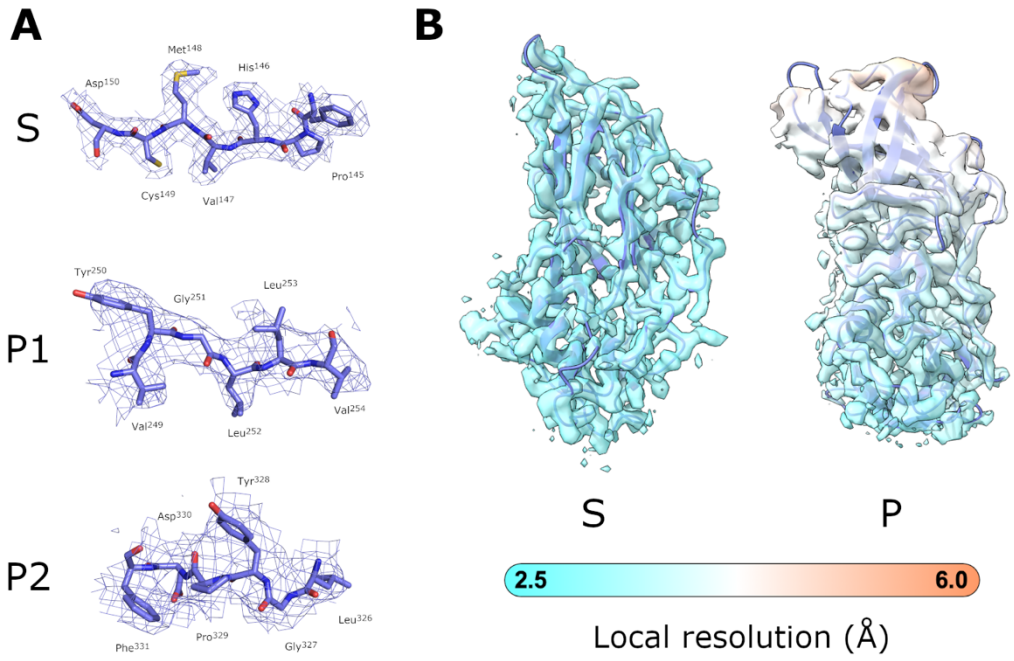
Given the difference in local resolution between the S and P domains, a hybrid approach was used to build an atomic model for wtMNV VP1.

The Phyre2 server (Kelley *et al.*, 2015) was used to generate a homology model for the S domain of MNV VP1 based on an X-ray crystal structure of the capsid from a related norovirus, Norwalk virus (PDB: 1IHM) (Prasad *et al.*, 1999). This model was rigid-body fitted into each of the three quasi-equivalent positions (A, B and C) within the asymmetric unit of the $T = 3$ capsid, then corrected manually in Coot to improve its fit to the experimental density. Automated refinement to improve fit and atomic geometry, followed by manual inspection to correct any obvious errors, were performed iteratively until there was no improvement in the quality of the model (judged by standard validation statistics).

Subsequently, an X-ray crystal structure for the MNV VP1 P domain bound to the receptor molecule CD300lf (PDB: 6C6Q) (Nelson *et al.*, 2018) was rigid-body fitted into each quasi-equivalent position. CD300lf and ligands were removed, and the peptide sequence corrected before the model was combined with the refined S domain-only model and subjected to a single round of refinement to improve fit and atomic geometry. The model was symmetrised during refinement to avoid fitting coordinates into strong density

from adjacent asymmetric units, and secondary structure restraints were imposed.

Following validation, this atomic model for VP1 was used for structural analysis of the wtMNV capsid. Validation statistics are given in Table 3.4. Representative EM densities with the VP1 atomic model fitted are shown in Figure 3.17.



3.17. Fit of VP1 atomic model to wtMNV EM density. (A) *Representative EM densities taken from the S, P1 and P2 (sub)domains of a C-type VP1 monomer are shown overlaid with the fitted VP1 atomic model.* ***(B)*** *The S domain and P domain of the atomic model of VP1 are shown fitted into EM density of a C-type VP1 monomer, coloured according to local resolution. (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

Table 3.4. Map and model validation statistics for wtMNV VP1.

Model	**wtMNV VP1**
EMDB ID	**10103**
PDB ID	**6S6L**
CryoEM map processing	
Number of particles contributing to map	7,811
Map resolution (FSC = 0.143) (Å)	3.1
Map resolution range around atom positions (Å)	2.9 – 6.5
Map sharpening B factor (Å^2)	−129
Residues modelled (from S domain homology model)	
Quasi-equivalent position A	20-191; 199-222
Quasi-equivalent position B	17-221
Quasi-equivalent position C	30-227
Residues modelled (from P domain X-ray crystal structure, PDB: 6C6Q)	
Quasi-equivalent position A	228-531
Quasi-equivalent position B	229-531
Quasi-equivalent position C	228-531
RMSD	
Bond lengths (Å)	0.0139
Bond angles (°)	1.42
Validation	
All-atom clashscore	5.43
MolProbity score	1.66
Rotamer outliers (%)	0.16
Ramachandran plot	
Favoured (%)	94.73
Allowed (%)	5.27
Outliers (%)	0.00

3.3.6. Insight into wtMNV capsid structure

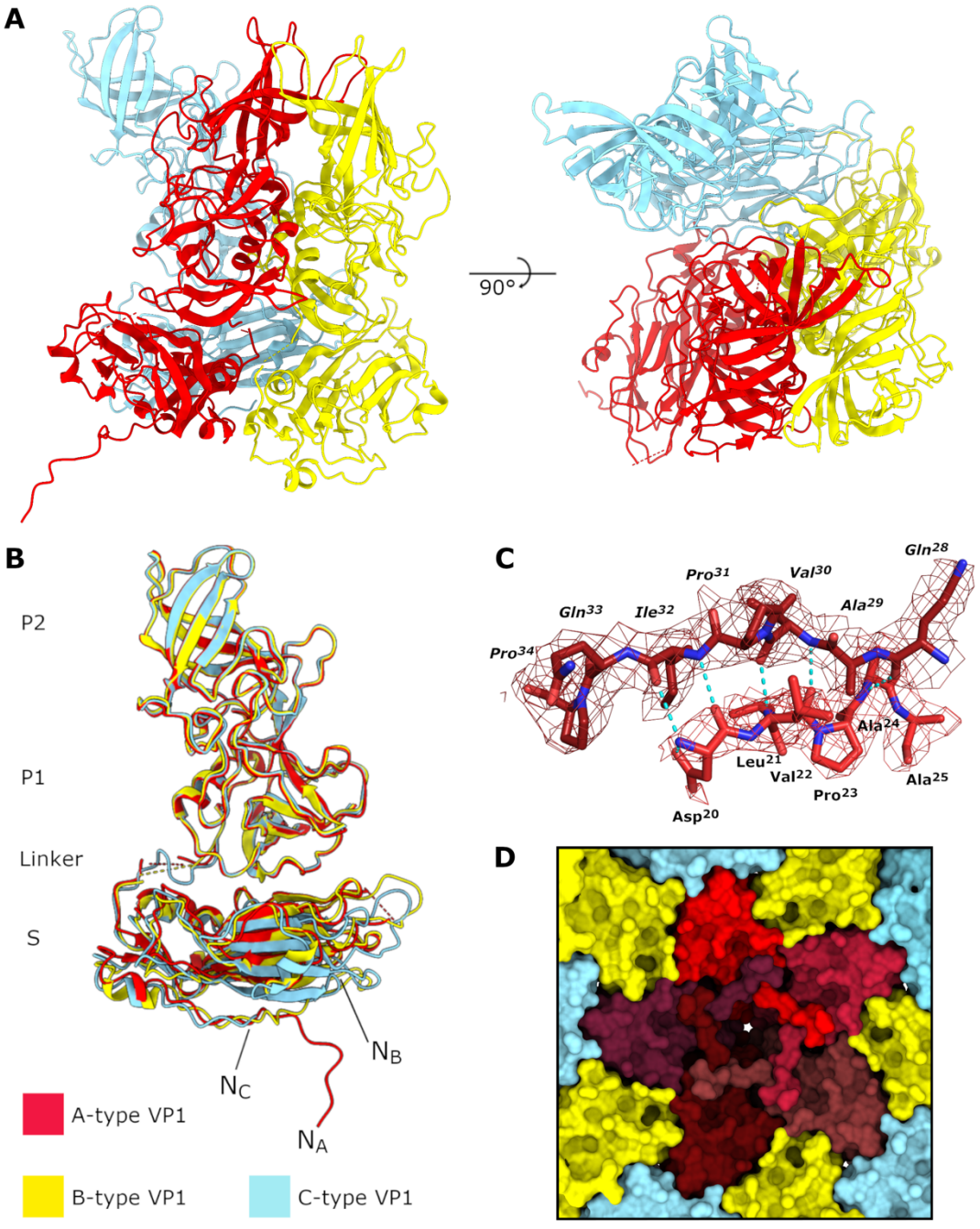
The structure of the MNV capsid shows features typical of other noroviruses. As expected, the capsid comprises 180 copies of VP1 in a *T* = 3 icosahedral arrangement. The VP1 molecules themselves display the expected domain organisation (Figure 3.18A,B): an N-terminal arm leads into the S domain,

which contains an eight-stranded jellyroll fold and is connected to the P domain by a short linker region. The P domain comprises P1 and P2 subdomains (connected by a hinge region), the latter of which comprises a six-stranded β-barrel, the strands connected by loops.

There are a number of subtle conformational differences between VP1 molecules in different quasi-equivalent positions (VP1 quasi-conformers) (Figure 3.18B). To investigate overall variation, quasi-conformers were aligned using a subset of spatially 'close-matching' atom pairs and root-mean-square deviation (RMSD) values were calculated between equivalent Cα atoms. For the subset of Cα atoms used to align the quasi-conformers, RMSD values were small (0.87 Å [A-B], 0.79 Å [A-C], 0.75 Å [B-C]). However, inclusion of all Cα atom pairs led to an increase in RMSD values (2.77 Å [A-B], 2.46 Å [A-C], 2.35 Å [B-C]), highlighting the presence of regions of substantial variability within VP1. Most striking is the variation seen between N-terminal arms. The A-type VP1 N-terminal arm (resolved from residue 20 onwards) protrudes into the capsid and interweaves with adjacent A-type N-terminal regions, creating a stabilising ring of interactions around the icosahedral five-fold axis (Figure 3.18C,D). Conversely, the N-terminal arm in B-type VP1 (resolved from residue 17 onwards) remains close to the underside of the S domain and contains a kink, appearing to extend from the icosahedral three-fold axis. The C-type VP1 N-terminal arm is mostly unresolved (missing residues 1 – 29), suggesting a greater degree of flexibility.

There is also variation in the flexible linker between S and P domains. In A- and B-type VP1, this region is unresolved, but the entire linker could be modelled for C-type VP1 (Figure 3.18B). This correlates with the overall improved quality of P domain density for C-type VP1, suggesting that C-type P domains are more rigid/fixed in position compared to A-/B-type P domains. In general, the local worsening in resolution of P domain density (compared to S domain density) for all quasi-conformers suggests that P domains are either flexible (i.e., inherent flexibility in the fold of the P domains), mobile (i.e., the overall position of the P domain may change relative to the S domains), or a

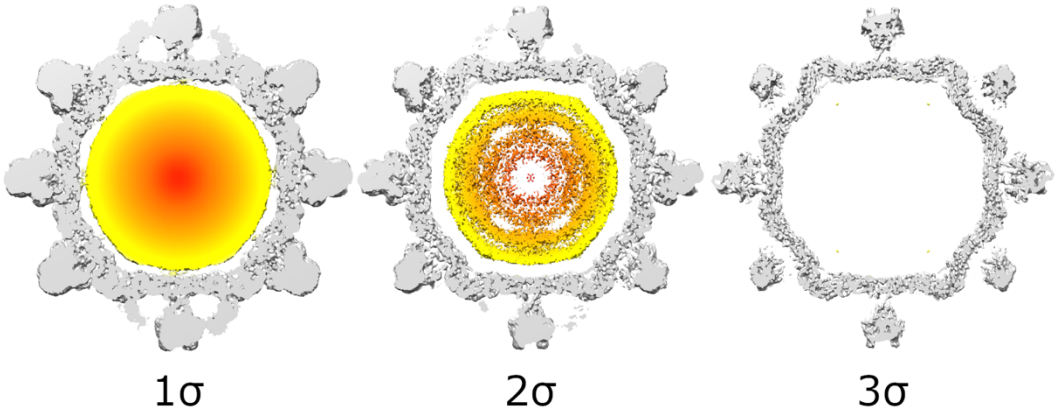
combination of the two. This is because cryoEM image processing relies on the averaging of many particle images. If P domains are not rigidly held in a fixed position relative to the S domains then P domain density becomes 'blurred' and high-resolution information is lost from the reconstruction.



3.18. Quasi-conformers of wtMNV VP1. (A) *Atomic coordinates for the asymmetric unit of the wtMNV capsid, comprising three monomers of VP1 coloured according to quasi-equivalent position (A, B and C).* ***(B)*** *VP1*

molecules from each quasi-equivalent position overlaid to highlight conformational differences. Different (sub)domains are indicated, along with the most N-terminal residue modelled for each quasi-conformer (N_A – D20, N_B – S17, N_C – V30). ***(C)*** *Atomic coordinates for N-terminal regions of adjacent A-type VP1 quasi-conformers, showing polar contacts (blue dashed lines).* ***(D)*** *Molecular surface representation of the underside of a five-fold axis (viewed from inside the capsid). Individual A-type VP1 monomers are coloured in different shades of red. Yellow and blue surfaces represent individual B- and C-type VP1 monomers.* (Parts of this figure were adapted *from (Snowden et al., 2020), published* under a CC BY 4.0 licence.)

As expected, other components of the virion (i.e., the genome, the minor structural protein VP2, and the viral protein genome-linked [VPg, also known as NS5 in MNV]) were not resolved. Only one copy of the genome and VPg are present in the virion, and only a small amount of VP2 is thought to be present, precluding these components from conforming to the icosahedral symmetry imposed during density refinement. While radial layers of amorphous genome density were visible at low contour levels in the unsharpened map (Figure 3.19), density was not present for other non-VP1 virion components, even in a map generated without the imposition of symmetry (i.e., a C1 reconstruction). This is likely because the signal for VP2, VPg and genome was weaker or less uniform than for VP1, and is discussed further in Chapter 4.



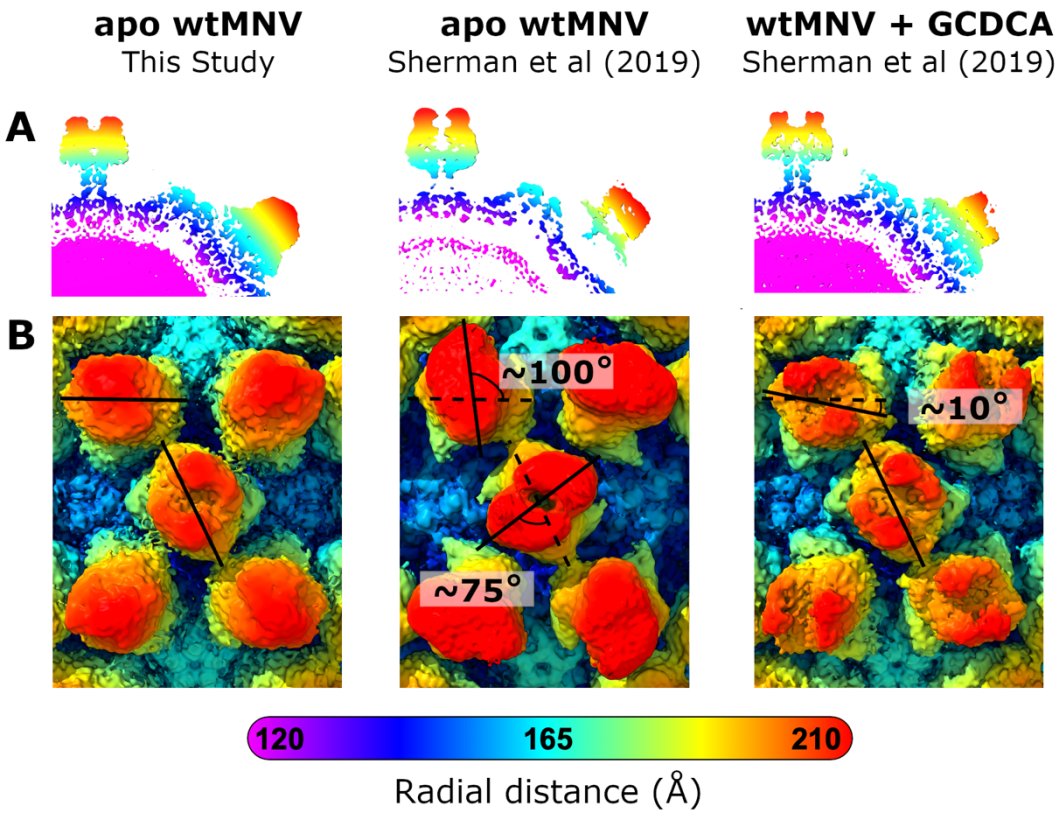
3.19. Amorphous genome density for wtMNV. *A central section through the unsharpened map of wtMNV, shown at different contour levels (approximate sigma values given below each). Amorphous density corresponding to the genome is coloured radially, the capsid (VP1) is coloured grey.*

3.3.7. Comparison with other published wtMNV structures

Interestingly, the arrangement/positioning of P domains was strikingly different to that observed for several wtMNV structures reported previously. Compared to a ~3-Å resolution cryoEM structure of wtMNV published while the analysis described here was being carried out (Sherman *et al.*, 2019), together with an 8-Å resolution cryoEM structure (Katpally *et al.*, 2010), P domain dimers in the wtMNV structure described here were in a 'contracted' state, being located much closer to the S domain surface (separated by ~6 Å rather than ~16 Å) (Figure 3.20A). They were also oriented differently – AB-type P domain dimers (surrounding icosahedral five-fold axes) were rotated ~100° clockwise relative to the previous structures and CC-type P domain dimers (at icosahedral two-fold axes) were rotated ~75° anticlockwise (Figure 3.20B).

This 'contracted' state was more in line with the vast majority of published norovirus and norovirus VLP structures from other genogroups. Similarly, the P domain positioning observed here was much closer (though not identical) to a cryoEM structure of wtMNV in complex with the bile salt

glycochenodeoxycholic acid (GCDCA) (Figure 3.20A,B) (Sherman *et al.*, 2019), despite the absence of bile salts.



3.20. Comparison of wtMNV capsid structures. (A) *Central cross-sections and* ***(B)*** *enlarged views of capsid surface centred on the icosahedral two-fold axes of the structures of apo wtMNV from this study (left), apo wtMNV from (Sherman et al., 2019) (centre) and wtMNV + GCDCA from (Sherman et al., 2019) (right). The orientation of P domain dimers in each case is indicated with a solid black line, shown with the orientation of P domain dimers from the apo wtMNV structure reported here as a dashed black line. Angles of deviation are indicated. Maps were all low-pass filtered to 4 Å resolution and coloured according to the radial colouring scheme indicated. (This figure was adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

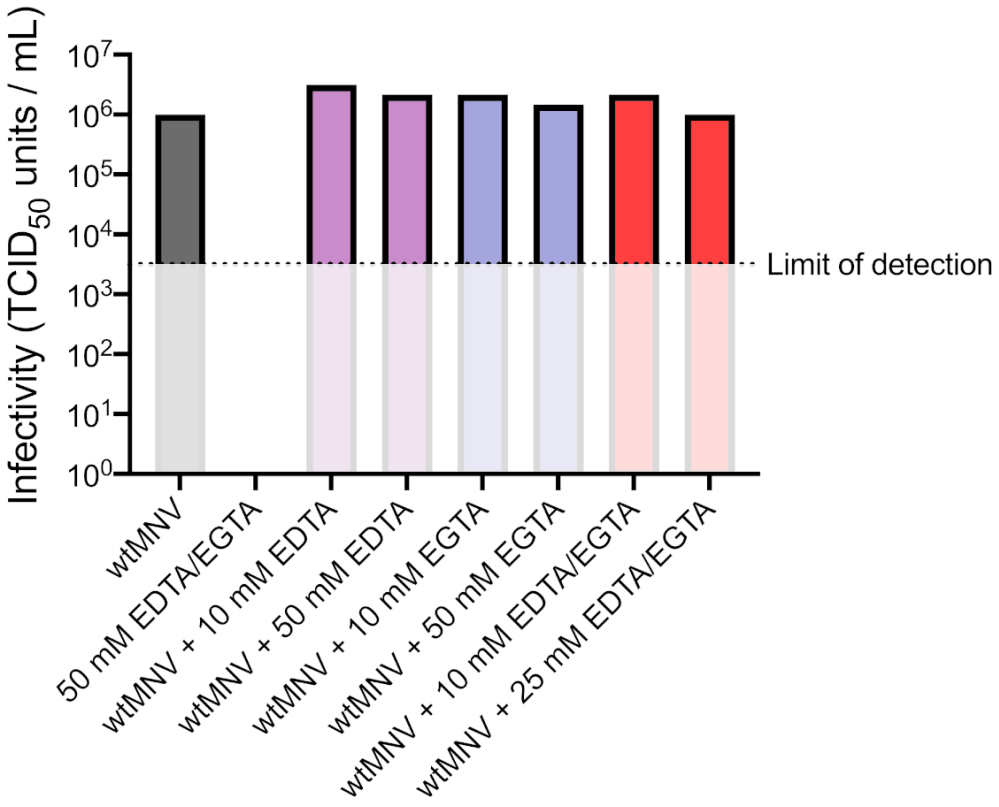
3.3.8. The effect of metal ions on P domain conformation

One hypothesis to explain the differences between different MNV structures was that the capsid adopted different conformations based on differences in buffer composition, which may be reflective of real changes in the chemical environment that the virion is exposed to during endocytosis to initiate infection (Scott and Gruenberg, 2011). To explore this, the effect of metal ions on wtMNV conformation and infectivity was examined. To sequester free metal ions in solution that may be interacting with and influencing the conformation of wtMNV, the chelating agents EDTA and EGTA were added to wtMNV in EM buffer at various concentrations up to 50 mM, either alone or in combination. EDTA has a higher affinity for Mg^{2+} ions, while EGTA has a higher affinity for Ca^{2+} ions, therefore any differences in the response of wtMNV to the two chelating agents would give insight into the relative importance of specific metal ions. As EDTA/EGTA are acidic, the pH of EDTA/EGTA-supplemented EM buffer was corrected to ensure that wtMNV was not exposed to any change in pH.

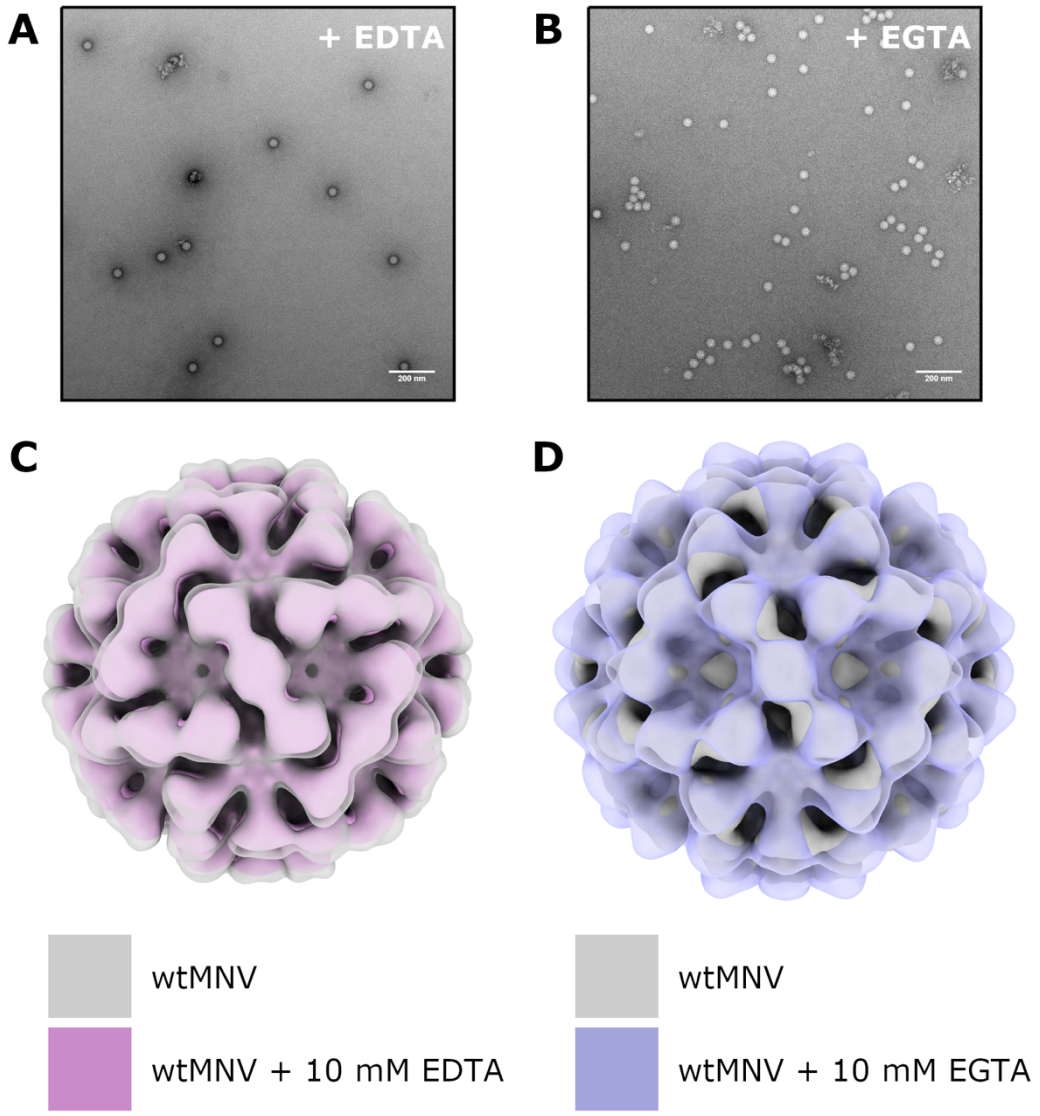
To identify any changes to infectivity, a $TCID_{50}$ assay was performed in RAW264.7 cells following incubation of wtMNV (cultivated in BV-2 cells) with EDTA/EGTA for 1 hour at 37°C. wtMNV in standard EM buffer was used as a positive control, and virus-free EM buffer supplemented with 50 mM of both EDTA and EGTA was used as a negative control to identify any potential toxic effects of EDTA/EGTA on cells. Ultimately, there were no indications that EDTA or EGTA had any effect on MNV infectivity (Figure 3.21).

To identify any morphological changes, negative stain EM was performed on wtMNV (cultivated in RAW264.7 cells) following the addition of 10 mM EDTA or EGTA (Figure 3.22A,B). Given the scale of the difference between the ‘contracted’ and ‘expanded’ states observed for wtMNV (including an increase in particle diameter of ~2 nm), I speculated that a low-resolution negative stain EM reconstruction may be sufficient to identify any changes. For both EDTA and EGTA, particles of wtMNV were manually picked from the negative stain EM micrographs, classified in 2D to remove poor quality particles, then used

to generate low-resolution (~23 Å) 3D reconstructions (with icosahedral symmetry imposed). The reconstruction of wtMNV + EDTA overlaid very well with that of a previous negative stain EM reconstruction of wtMNV in non-supplemented EM buffer (Figure 3.22C). The reconstruction of wtMNV + EGTA overlaid less well with wtMNV alone, however this difference was not in line with a switch to the 'expanded' state – rather, the P domains looked more angular, so the difference was likely an artefact of the low resolution of the reconstruction rather than a genuine conformational change (Figure 3.22D). Ultimately, negative stain EM analysis gave no strong evidence of a morphological change.



3.21. $TCID_{50}$ assay of wtMNV in the presence of EDTA/EGTA. *wtMNV was incubated in EM buffer supplemented with EDTA, EGTA or both (EDTA/EGTA) prior to titration by $TCID_{50}$ assay (*n = *1). The limit of detection (3.15×10^3 $TCID_{50}$ units/mL) is indicated by the dotted line.*



3.22. Negative stain EM of wtMNV in the presence of EDTA/EGTA. (A, B) *Representative negative stain EM micrographs of wtMNV in the presence of 10 mM EDTA (A) or 10 mM EGTA (B). Scale bars show 200 nm.* ***(C, D)*** *Low-resolution 3D reconstructions of wtMNV in the presence of 10 mM EDTA (C) or 10 mM EGTA (D) overlaid on a similar negative stain EM 3D reconstruction of wtMNV in the absence of any chelating agent (grey). Maps were flipped in Z plane where necessary to show the correct orientation and were coloured according to the scheme indicated.*

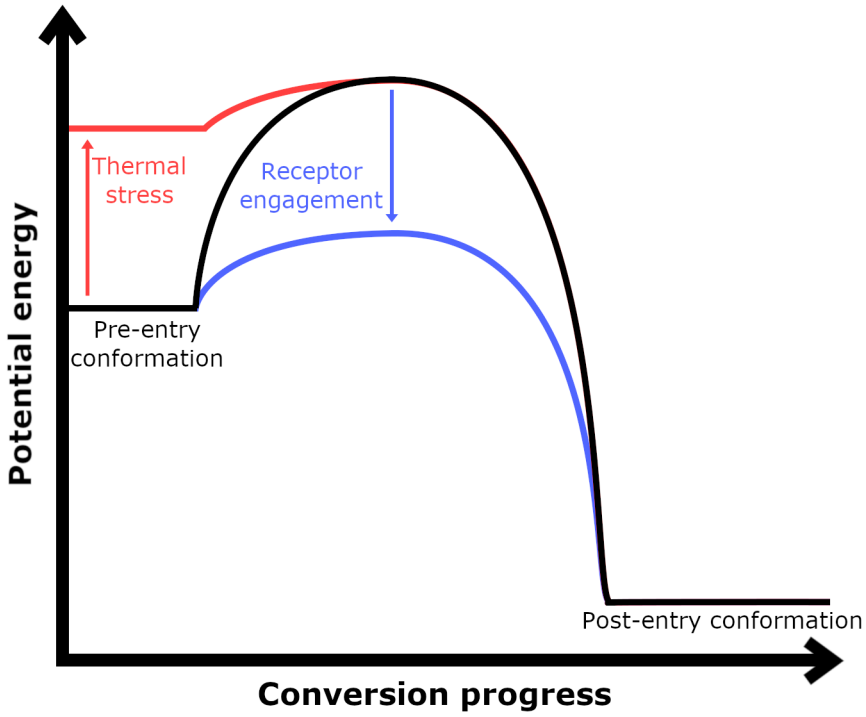
Given the lack of evidence for any change in infectivity or capsid structure induced by the chelating agents, this line of enquiry was not pursued further.

3.4. Structural changes to MNV due to heat inactivation (hiMNV)

As part of the overall structural investigation into wtMNV, efforts were made to capture the capsid in a conformational state indicative of the changes that the virion undergoes during cell entry. While there is now structural evidence from a related system (feline calicivirus) that receptor engagement triggers the assembly of a VP2 'portal' for genome release in caliciviruses (Conley *et al.*, 2019), this was not known at the time of this analysis.

In an attempt to trap a cell entry intermediate of the MNV capsid, Drs Morgan Herod (MH), Oluwapelumi Adeyemi (OA) and DH carried out a cryoEM data collection of wtMNV following exposure to thermal stress (i.e., heat-inactivated MNV, or hiMNV). Thermal stressing has already been successfully used to induce alternative capsid conformations reflective of cell entry intermediates in enteroviruses (Curry *et al.*, 1996; Bubeck *et al.*, 2005). Capsids are metastable complexes, restricted from adopting the minimum energy (most stable) conformation due to an energy barrier that must be overcome for conversion (Figure 3.23, black line). Typically, engagement with a receptor (and/or other factors, such as changes in the ionic composition of the environment) reduces the size of the energy barrier (e.g., through the induction of allosteric changes that induce strain in the 'pre-entry' form of the capsid), facilitating conversion of the capsid to a 'post-entry' conformation (Figure 3.23, blue line). Conversely, thermal stress increases the internal energy of the 'pre-entry' capsid, so that capsids are able to overcome the energetic barrier and spontaneously convert to the 'post-entry' conformation in the absence of receptor (Figure 3.23, red line). As such, thermal stress is a useful tool to induce alternative conformations that are reflective of those naturally adopted during the infection cycle.

This section describes the sample preparation and thermal stressing of MNV (performed by MH and OA), the collection of hiMNV cryoEM data (performed by DH), the processing of this data (performed by DH and the author of this thesis) and subsequent analysis (performed by the author of this thesis).



3.23. Energetic changes during conformational conversion of a capsid. *Virus capsids are kept in a metastable state, unable to overcome an energetic barrier required to convert into a more stable, low-energy post-entry conformation (black line). Receptor engagement reduces the size of the energetic barrier (blue line). Thermal stress increases the internal energy of the capsids so that they are more easily able to convert to the post-entry conformation in the absence of receptor (red line).*

3.4.1. Sample preparation, thermal stressing and vitrification of hiMNV

To generate hiMNV for structural investigation, MH and OA cultivated wtMNV by passage in RAW264.7 cells and purified the virus from crude lysate using a double sucrose gradient approach, as described above. The thermal stability of wtMNV was then assessed through $TCID_{50}$ assays (to probe infectivity) and a fluorescence-based Particle Stability Thermal Release (PaSTRy) assay (to probe capsid integrity). PaSTRy assays rely on two dyes, SYTO-9 (which detects nucleic acid) and SYPRO-Orange (which binds to hydrophobic regions of proteins), to give a simultaneous readout of genome release and capsid stability.

$TCID_{50}$ assays were performed using virus that had been heated on a 30-second constant temperature ramp, and revealed a loss of >99.9% titre at 61°C (Figure S1A). Despite this, PaSTRy assays suggested that capsids were largely intact until ~64°C, at which point SYTO-9 fluorescence peaked (Figure S1B). To confirm this result, MNV heated to 61°C then incubated with RNase showed no evidence of genome digestion, as assessed by $TCID_{50}$ assay of virus derived from transfection of extracted RNA into BHK cells (Figure S1C). As such, wtMNV heated to 61°C was termed heat-inactivated (hi)MNV, describing structurally intact virions that had become irreversibly non-infectious (Snowden *et al.*, 2020).

hiMNV was applied to ultra-thin carbon-coated lacey carbon grids by DH using a 'multiple applications' approach, and these grids were vitrified for cryoEM.

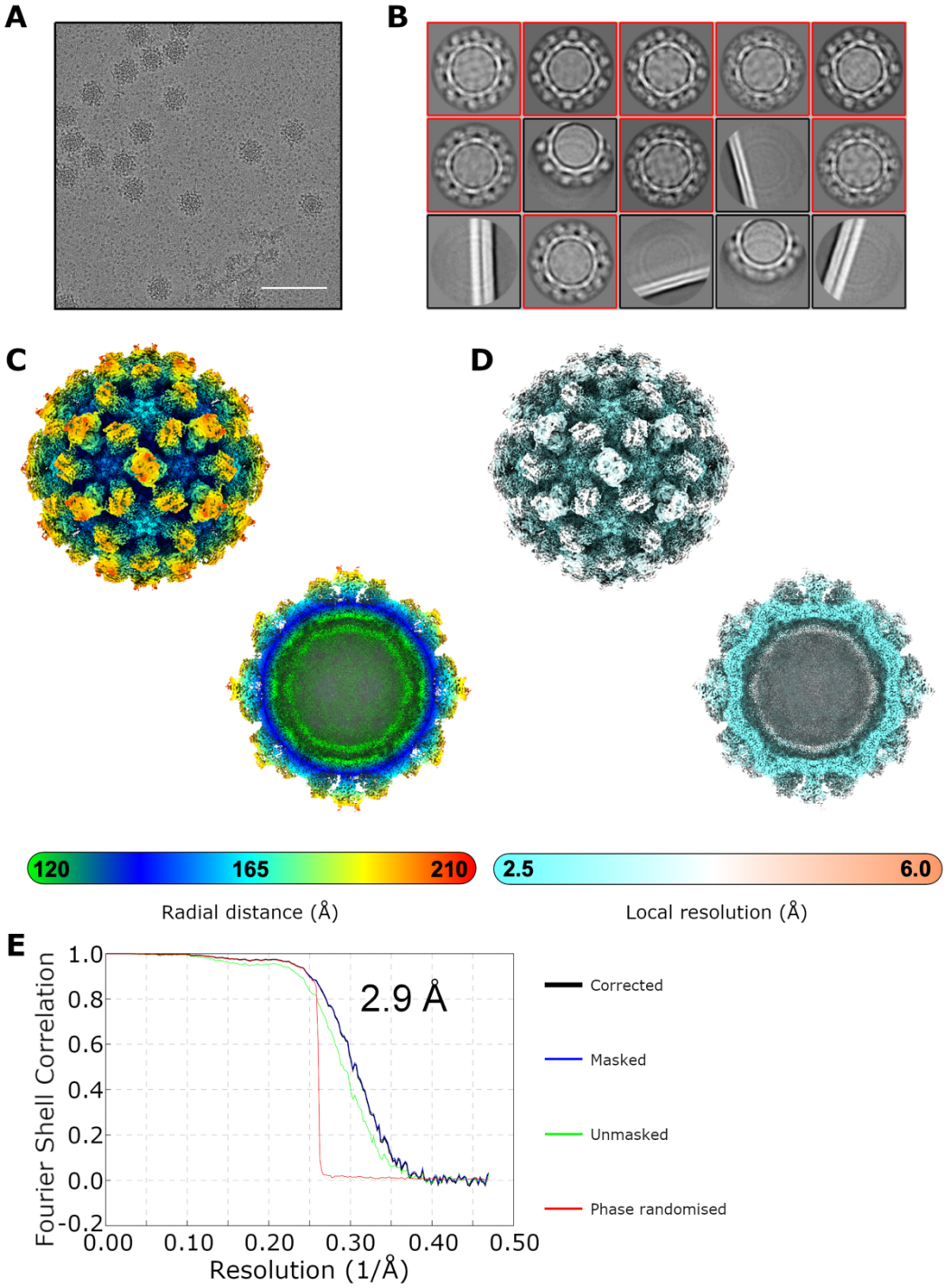
3.4.2. hiMNV cryoEM data collection and processing

hiMNV data collection was performed by DH using an FEI Titan Krios microscope at 300 kV, at a magnification of 75,000× and a calibrated object sampling of 1.065 Å/pixel. As for the wtMNV data collection, the total electron dose per image was 59 $e^-/Å^2$, distributed across 59 frames over 1.5 seconds. A defocus range of −0.5 to −2.9 µm was used to generate phase contrast. Full details of parameters used in data collection are given in Table 3.5.

From the 2,620 micrographs collected (Figure 3.24A), and after excluding poor-quality particles through 2D and 3D classification (Figure 3.24B), over 14,000 particles contributed to a 2.9-Å resolution icosahedral reconstruction following several rounds of CTF refinement and Bayesian polishing (Figure 3.24C-E). Data processing parameters are given in Table 3.5.

Table 3.5. Data collection and processing parameters for hiMNV. *The hiMNV cryoEM data collection was performed by DH.*

CryoEM data collection	
Microscope	FEI Titan Krios
Camera	Falcon III
Voltage (kV)	300
Pixel size (Å)	1.065
Nominal magnification	75,000×
Exposure time (s)	1.5
Total dose (e^-/Å^2)	59
Number of fractions	59
Defocus range (μm)	−0.5 to −2.9
Number of micrographs	2,620
Acquisition software	FEI EPU
CryoEM map processing	
EMDB ID	EMD-10127
Number of particles contributing to map	14,266
Map resolution (FSC = 0.143) (Å)	2.9
Map resolution range around atom positions (Å)	2.8 – 8.9
Map sharpening B factor (Å^2)	−136



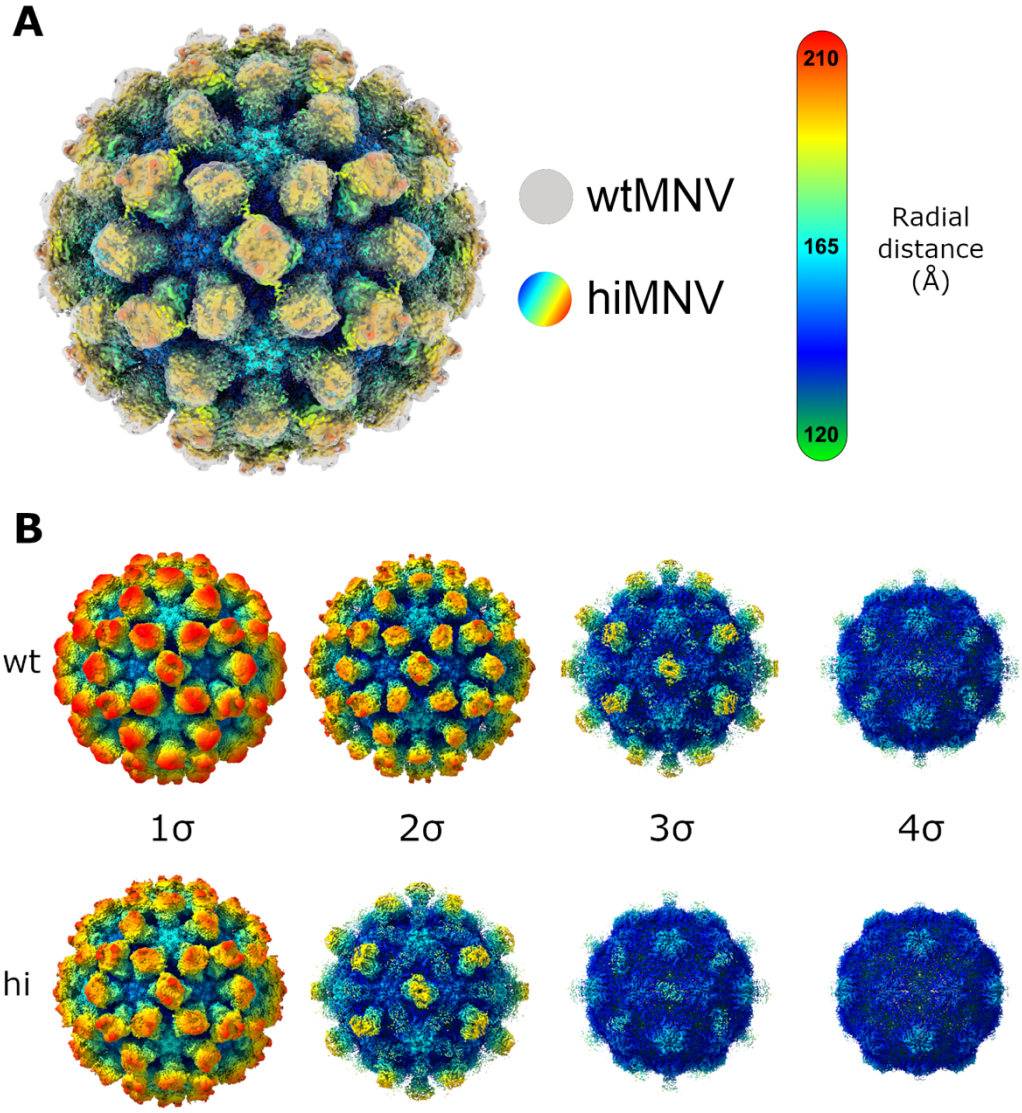
3.24. hiMNV cryoEM data collection and processing. (A) *Representative micrograph from the hiMNV dataset. Scale bar represents 100 nm.* ***(B)*** *Top 15 most populated classes from initial 2D classification of hiMNV data. Particles from the classes highlighted in red (as well as other classes not shown here) were taken forward for further processing.* ***(C,D)*** *Full and cross-sectional isosurface representations of hiMNV density map shown at ~1.5 σ, filtered according to local resolution and coloured according to* ***(C)*** *the radial*

*colouring scheme or **(D)** the local resolution colouring scheme shown. **(E)** FSC plot for the hiMNV map. The resolution (2.9 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected). (The cryoEM data presented in this figure was collected by DH. 2D classes were generated by DH. Local resolution-filtered 3D reconstructions were generated by the author of this thesis. This figure was created by the author of this thesis. Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.4.3. Heat inactivation alters the 'strength' of P domain density

Comparison of the cryoEM reconstructions of wtMNV and hiMNV suggested that heat inactivation did not lead to any changes in the gross morphology of the capsid, including the positioning of the P domain dimers relative to the S domains (Figure 3.25A). However, there was a noticeable reduction in the 'strength' of the EM density corresponding to P domain dimers (i.e., a relative reduction in the signal intensity in those voxels, such that a lower contour level was required to visualise P domain density) (Figure 3.25B). There was no equivalent reduction in density 'strength' for the S domains between wtMNV and hiMNV, suggesting the change was unique to the P domains.

One explanation for this difference is that a subpopulation of P domains became denatured or detached from the capsid as a result of heat treatment. However, no biochemical evidence was found for 'lost' P domains, despite the presence of some small P domain dimer-sized particles in raw micrographs (Figure 3.24A). Equally, there was no evidence for subsets of capsids lacking P domains, or with obviously denatured P domains, in the results of 2D and 3D classification. As such, the evidence indicated that heat treatment caused either the fold of the P domain itself to become more flexible, or the S-P linker to become more mobile (extending the range of potential positions that whole P domain dimers could occupy relative to the S domains, therefore 'blurring' the P domain density in the reconstruction).



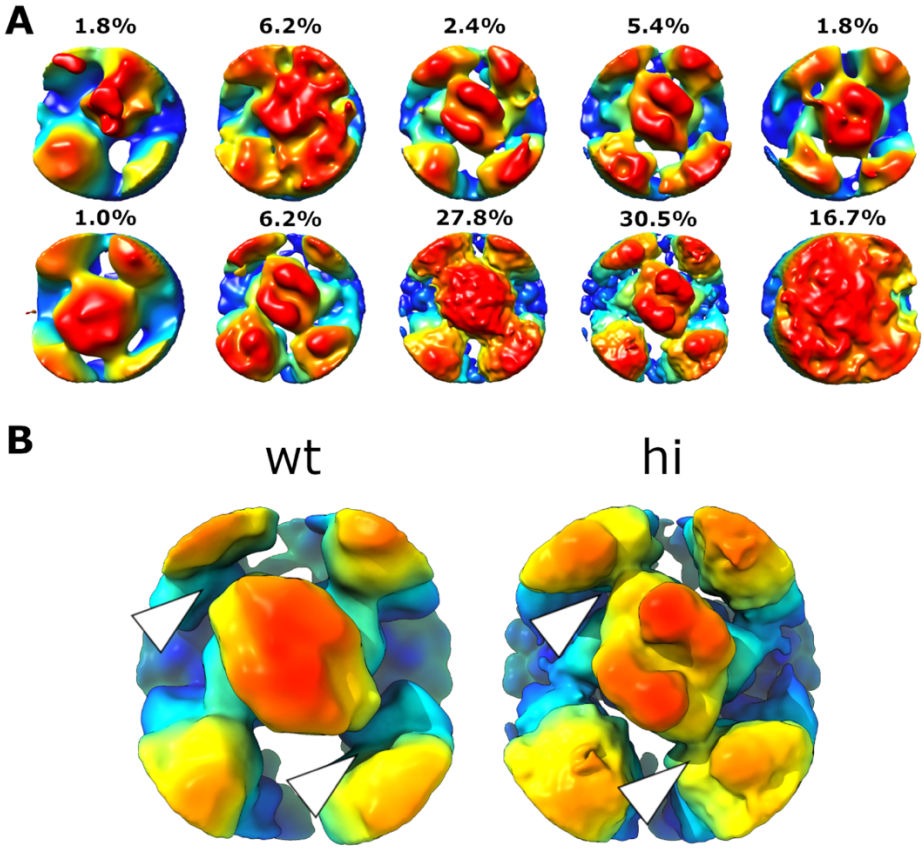
3.25. Comparison of wtMNV and hiMNV cryoEM reconstructions. (A) *Overlay of wtMNV (translucent grey) and hiMNV (radial colouring scheme) EM density maps filtered by local resolution, both shown at ~1.5 σ.* ***(B)*** *Comparison of local resolution-filtered density maps for wtMNV (wt) and hiMNV (hi) at a range of contour levels. Both maps were coloured according to the radial colouring scheme shown in (A). (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.4.4. hiMNV P dimers explore a changed conformational landscape

To further explore the variation in P domain dimer positioning/orientation in the hiMNV reconstruction, focussed classification was performed in a similar

manner as described above for wtMNV. The cylindrical mask was applied over a CC-type dimer, as CC-type dimers were better resolved than AB-type dimers.

While some subparticles were assigned to classes containing questionable CC-type dimer density (potentially indicative of P domain denaturation), most subparticles were assigned to standard CC-type dimer classes that were similar to those seen for wtMNV (Figure 3.26A). More detailed analysis revealed that the central CC-type dimer contacted adjacent AB-type dimers in all classes. This was in contrast to wtMNV, where ~8.0% of CC-type dimers populated a class without any contact between CC- and AB-type dimers (i.e., a 'non-contacting' class) (Figure 3.26B). This difference suggested that the conformational landscape explored by the P domains had been altered by the heat treatment, which may be related to the heat-induced loss of infectivity.



3.26. Focussed classification of hiMNV CC-type P domain dimers. (A) *All focussed classes from focussed classification of CC-type P domain dimers. The proportion of subparticles assigned to each class is indicated.*

(B) *Example of a 'non-contacting' class from focussed classification of wtMNV CC-type P domain dimers (wt). hiMNV only has 'contacting' classes (hi). (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.5. The structure of a heat-stable mutant (hs)MNV

Given the observations reported here and elsewhere that P domain dimers are highly mobile components of the capsid, and that a change in the flexibility or mobility of P domain dimers correlated with a loss of infectivity, it was hypothesised that virus selected for improved thermostability would acquire mutations within VP1 affecting P domain mobility or conformation and that the site of any mutation may give some indication as to which regions of VP1 are important for stability. Therefore, to support the findings reported in Section 3.4 and probe the mechanism by which MNV lost infectivity upon heating, structural analysis was performed on a mutant MNV that retained infectivity at high temperatures.

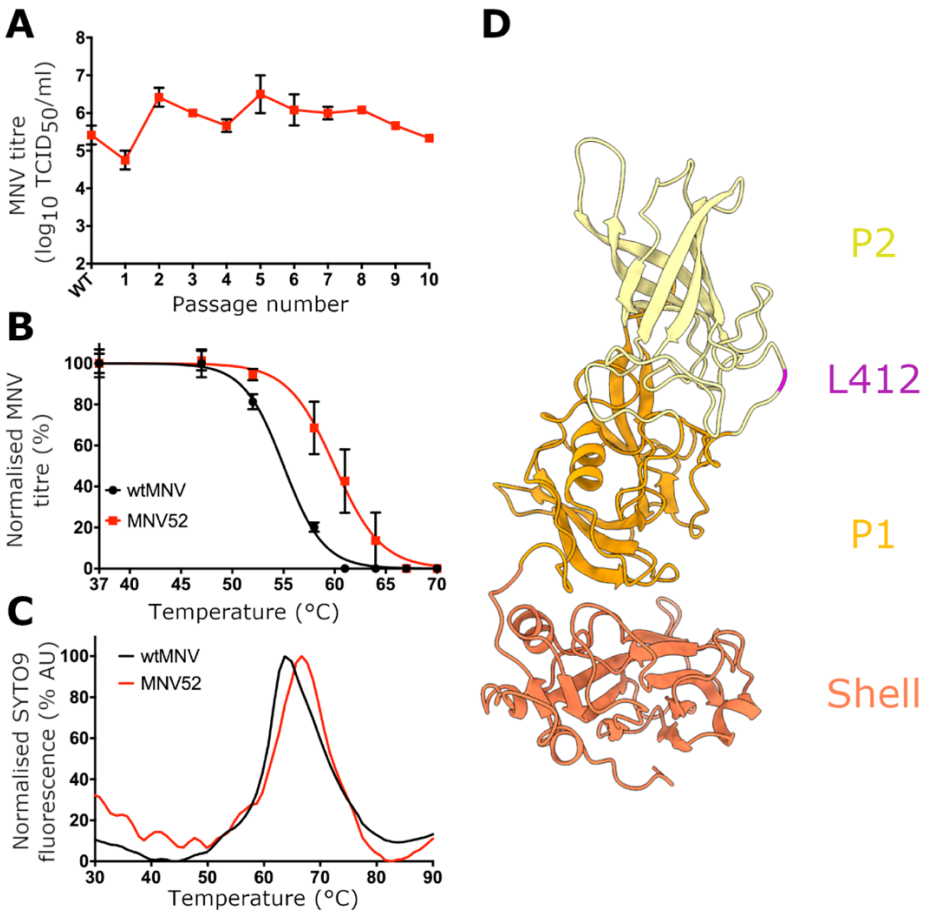
This section describes the selection and characterisation of a thermostabilised mutant MNV and subsequent generation of an infectious clone for hsMNV (performed by MH and OA), the recovery, cultivation and purification of infectious hsMNV, cryoEM data collection and image processing, and subsequent analysis of hsMNV cryoEM data, along with further mutational analysis (all performed by the author of this thesis).

3.5.1. Selection and characterisation of thermostabilised MNV mutants

In vitro evolution was performed by MH and OA to select for MNV that could retain infectivity after incubation at high temperatures. wtMNV was passaged in RAW264.7 cells with selection at 52°C for 30 min between passages (Figure 3.27A), and the resulting pool of virus (named MNV52) showed a significant improvement in thermal stability, determined by $TCID_{50}$ assay (Figure 3.27B). A fluorescence-based PaSTRy assay was also used to compare the stability of MNV52 virions with those of wtMNV. In agreement with the infectivity data,

this indicated that MNV52 viral RNA was exposed to solvent at higher temperatures than for wtMNV (Figure 3.27C).

In order to identify any mutation(s) responsible for the improvement in stability, the regions of the MNV genome that encode the structural proteins VP1 and VP2 (ORF2 and ORF3, respectively) were first amplified by RT-PCR, then sequenced at the consensus level. No changes were identified in ORF3, though ORF2 contained a single mutation leading to a substitution in VP1 – L412Q. This residue was located on a hinge loop connecting P1 and P2 subdomains (Figure 3.27D). To probe the specific effects of the L412Q mutation, it was reconstituted into an infectious clone of MNV for further characterisation.



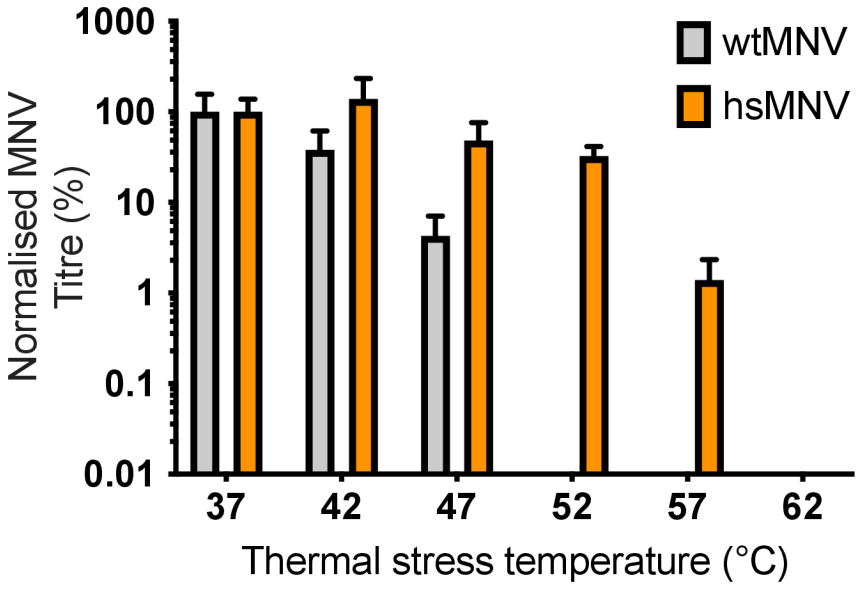
3.27. Selection and characterisation of MNV52. (A) *Thermal selection of MNV52. MNV samples were consecutively passaged for 48 hours at 37°C in RAW264.7 cells, with thermal stressing performed between passages*

(52°C, 30 min). Before selection (WT) and at each passage, the virus titre was determined by $TCID_{50}$ on RAW264.7 cells (n = 2 ± SEM). ***(B)*** *Thermal characterisation of MNV52 infectivity. The pool of virus passaged with thermal stressing (termed MNV52), as well as wtMNV, were heated to a range of different temperatures. Following thermal stress, virus titre was determined by $TCID_{50}$ on RAW264.7 cells (n = 4 ± SEM).* ***(C)*** *Thermal characterisation of MNV52 stability. Following purification by sucrose density gradient, wtMNV and MNV52 were dialysed into PBS and subjected to PaSTRy thermal stability assay using the nucleic acid dye SYTO9.* ***(D)*** *Atomic model for C-type VP1 from wtMNV, coloured by (sub)domain as indicated. The site of the mutation identified by sequencing (L412Q) is highlighted in magenta. (Experiments in (A-C) performed by MH and OA. This figure was adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.5.2. Recovery of hsMNV from an infectious clone

Infectious MNV with the L412Q mutation (termed heat-stable [hs]MNV) was recovered from the infectious clone by transfection into BHK-21 cells. BHK-21 cells do not express the cell surface receptor for MNV, CD300lf, so only permit a single infectious cycle upon transfection of viral RNA. Following transfection, cells were harvested and lysed, and the cell lysate was passaged in RAW264.7 cells to increase the viral titre. CPE was observed, suggesting that hsMNV had successfully been recovered. Between passages, hsMNV was subjected to thermal stressing (30-min incubation at 52°C) to maintain a selection pressure.

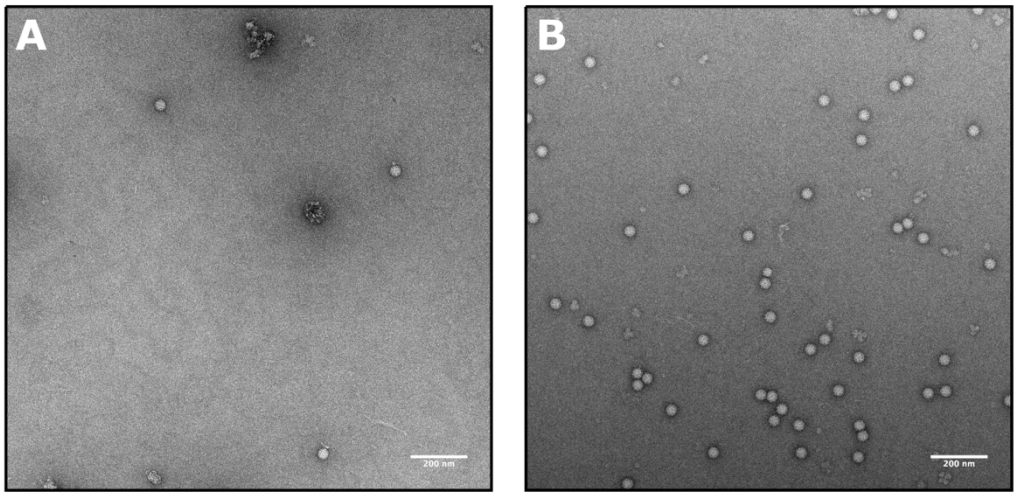
To confirm that hsMNV retained the improved thermal stability phenotype, hsMNV was titred by $TCID_{50}$ assay after incubation at a range of temperatures, and compared with wtMNV. It was clear that hsMNV retained infectivity after incubation at temperatures that rendered wtMNV non-infectious (Figure 3.28).



3.28. hsMNV retains infectivity after thermal stressing. *wtMNV (grey) or hsMNV (orange) was heated for 30 min at the indicated temperature, then its titre was determined by $TCID_{50}$ assay on RAW264.7 cells (n = 3 ± SEM). (This figure was adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.5.3. Sample preparation and vitrification of hsMNV

To generate a sufficient amount of hsMNV for structural analysis, the virus was passaged in RAW264.7 cells (with a thermal stressing step applied between each passage, as described above). Ultimately, the lysate from 16 × T-175 flasks was harvested and purified in the same way as described for wtMNV (i.e., by pelleting through a sucrose cushion, followed by two 15-60% sucrose gradients). The peak fractions of the final sucrose gradient were dialysed separately, then combined (a total volume of ~1 mL) and visualised by negative stain EM (Figure 3.29A). The concentration of particles was relatively low, so the purified stock was concentrated by pelleting (366,000 × *g*, 60 min, 4°C) and resuspending in ~50 µL EM buffer. The purified, concentrated stock was re-analysed by negative stain EM, confirming that the sample was suitable for cryoEM (Figure 3.29B).



3.29. Negative stain EM of purified hsMNV. (A) *Representative micrograph from negative stain EM of purified hsMNV prior to further concentration.* ***(B)*** *Representative micrograph from negative stain EM of purified, concentrated hsMNV. Scale bars show 200 nm.*

To confirm that hsMNV had not changed during passage, a sample of the hsMNV stock used for cryoEM was taken for genome extraction, amplification by RT-PCR and sequencing. Sequencing results for ORF2 and ORF3 (encoding the structural proteins VP1 and VP2) confirmed that the L412Q mutation was still present, and that there were no other differences from the wild-type VP1 and VP2 sequences.

To vitrify hsMNV for cryoEM, two 30-second applications of hsMNV (with manual blotting to remove excess fluid in between) were performed on glow discharged lacey carbon grids (coated with an ultrathin layer of amorphous carbon). Blotting times between 2.5 – 4.0 seconds were used to generate variable ice thickness. Grids were screened to select an optimal candidate for data collection.

3.5.4. CryoEM data collection and processing of hsMNV

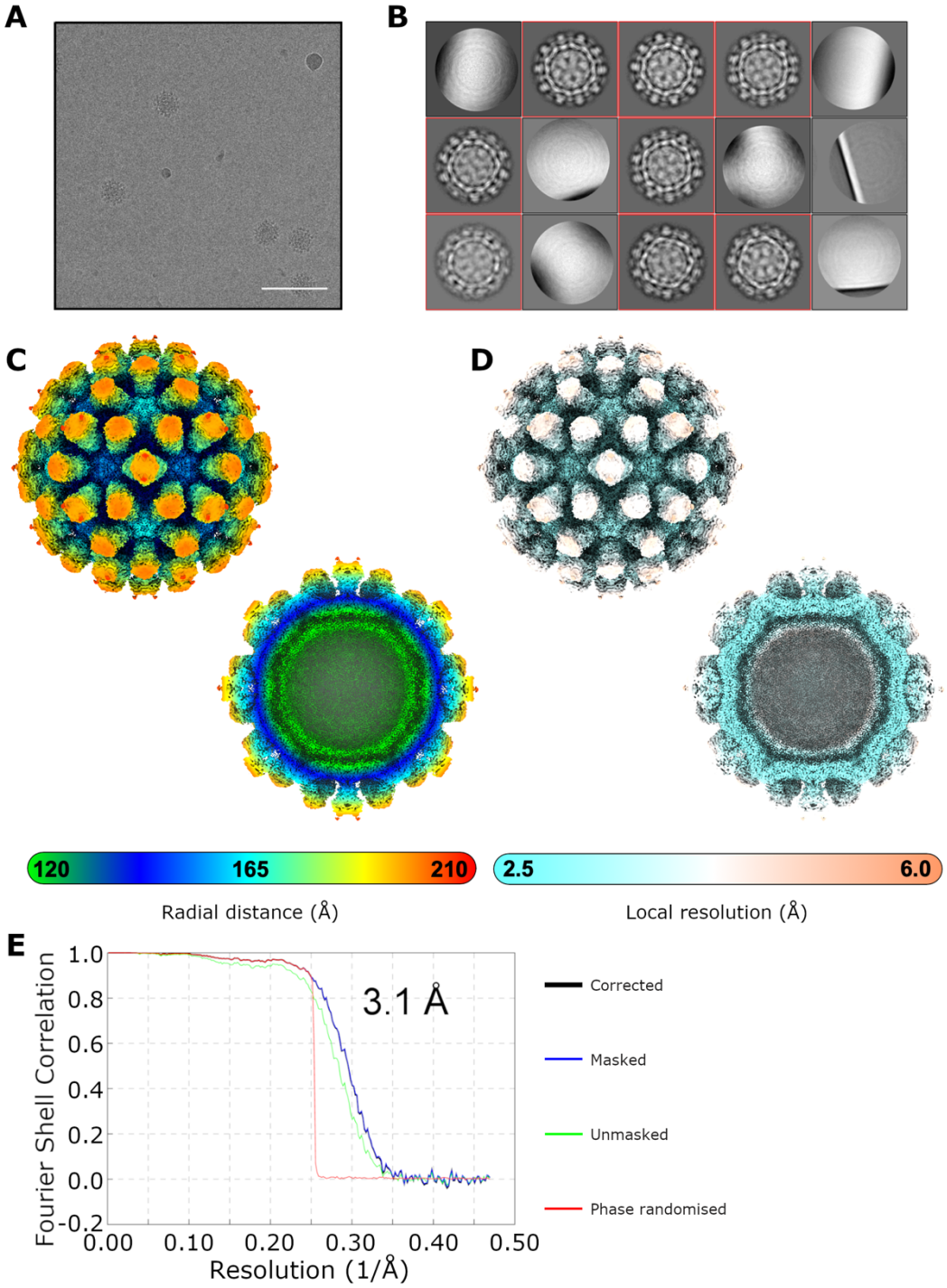
hsMNV data collection was performed using an FEI Titan Krios microscope at 300 kV, at a magnification of 75,000× and a calibrated object sampling of 1.065 Å/pixel. The total electron dose per image was 64 $e^-/Å^2$, distributed

across 59 frames over 1.9 seconds. To generate phase contrast, between −0.7 to −3.0 µm of defocus was applied to each image. Full details of parameters used in data collection are given in Table 3.6.

A total of 7,617 micrographs were collected (Figure 3.30A) which were processed to yield a total of 35,263 particles contributing to the final reconstruction, following exclusion of poor-quality particles by 2D classification (Figure 3.30B). With CTF refinement and Bayesian polishing, the final reconstruction was resolved to 3.1 Å following sharpening (Figure 3.30C-E). Full image processing details are reported in Table 3.6.

Table 3.6. Data collection and processing parameters for hsMNV.

CryoEM data collection	
Microscope	FEI Titan Krios
Camera	Falcon III
Voltage (kV)	300
Pixel size (Å)	1.065
Nominal magnification	75,000×
Exposure time (s)	1.9
Total dose (e^-/Å^2)	64
Number of fractions	59
Defocus range (µm)	−0.7 to −3.0
Number of micrographs	7,617
Acquisition software	FEI EPU
CryoEM map processing	
EMDB ID	EMD-10128
Number of particles contributing to map	35,263
Map resolution (FSC = 0.143) (Å)	3.1
Map resolution range around atom positions (Å)	2.9 – 6.0
Map sharpening B factor (Å^2)	−153

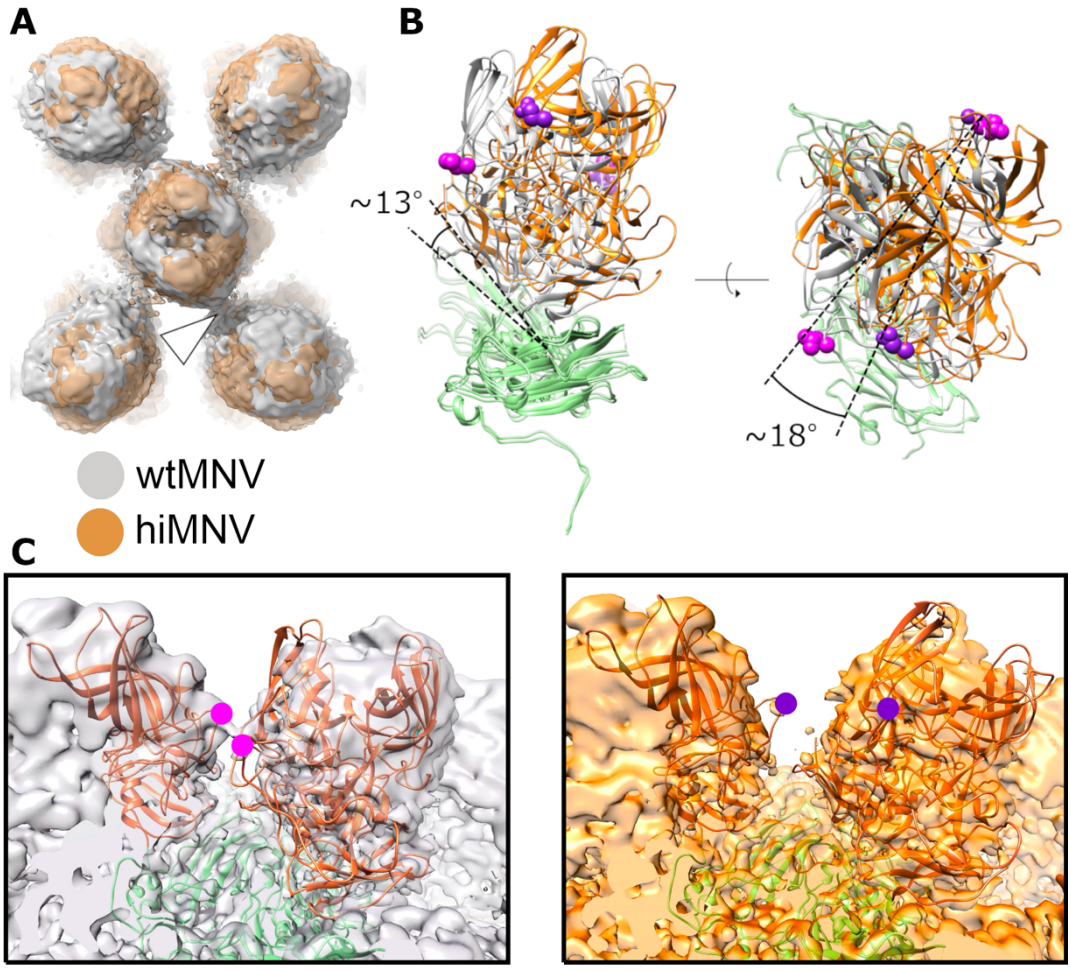


3.30. hsMNV cryoEM data collection and processing. ***(A)*** *Representative micrograph from the hsMNV dataset. Scale bar represents 100 nm.* ***(B)*** *Top 15 most populated classes from initial 2D classification of hsMNV data. Particles from the classes highlighted in red (as well as other classes not shown here) were taken forward for further processing.* ***(C,D)*** *Full and cross-sectional isosurface representations of hsMNV density map shown at ~1.5 σ, filtered according to local resolution and coloured according to* ***(C)***

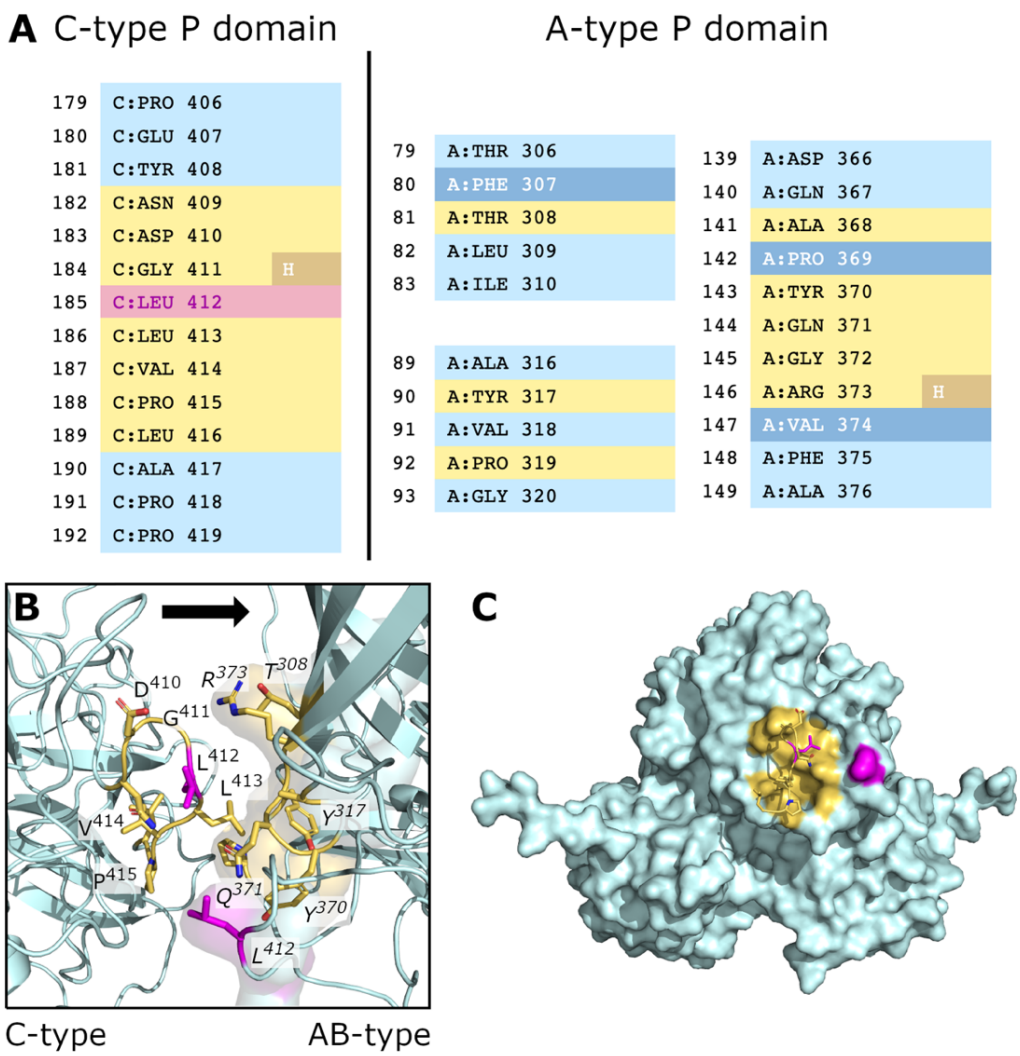
*the radial colouring scheme or **(D)** the local resolution colouring scheme shown. **(E)** FSC plot for the hiMNV map. The resolution (3.1 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected). (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.5.5. hsMNV has 'twisted' AB-type P domain dimers

Unlike with hiMNV, overlaying hsMNV with wtMNV revealed a subtle difference in the positioning of AB-type P domain dimers (Figure 3.31A). CC-type dimers overlaid very well with wild type, suggesting only AB-type dimers were affected. To examine this change in more detail, the atomic model for wtMNV VP1 was split into components (S domains, AB-type P domain dimer, C-type P domain) which were rigid-body fitted into the hsMNV map, then re-combined and subject to a single round of refinement with secondary structure restraints enabled. Analysis of the atomic coordinates fitted into hsMNV revealed that the AB-type P domain dimer had tilted upwards and away from the S domain surface by ~13°, and rotated ~18° in the anticlockwise direction (Figure 3.31B). This results in the disruption of an interface formed between A-type and C-type P domains in wtMNV (Figure 3.31C), confirmed by analysis with the interface evaluation tool PDBePISA (Protein Data Bank in Europe Proteins, Interfaces, Structures and Assemblies) (Figure 3.32). Interestingly, L412 from C-type VP1 contributes to this interface in wtMNV, suggesting that it may have been directly disrupted by the L412Q mutation in hsMNV.



3.31. hsMNV has 'twisted' AB-type P domain dimers. (A) *Isosurface representations of wtMNV (grey) and hsMNV (orange) overlaid, with the back plane clipped to show only P domain density. Both maps are displayed at ~1 σ.* ***(B)*** *AB-type VP1 atomic model after fitting P and S domains separately into density for wtMNV (grey) or hsMNV (orange). S domains are shown in green. The mutated residue is represented as magenta (L412, wtMNV) or dark purple (L412Q, hsMNV) spheres.* ***(C)*** *Atomic model for a full asymmetric unit of VP1 with P and S domains fitted separately into wtMNV (left) or hsMNV (right) density, centred on the interface highlighted by the arrowhead in (A). The mutated residue is represented as a magenta (L412, wtMNV) or dark purple (L412Q, hsMNV) circle. (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

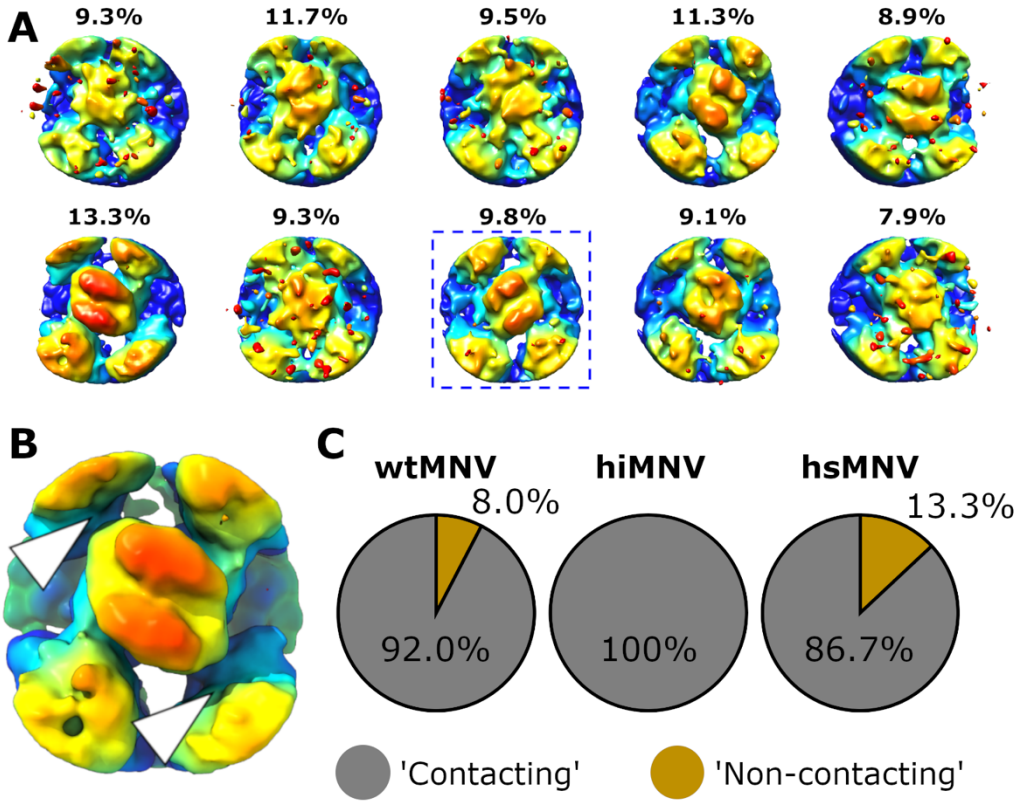


3.32. PDBePISA analysis of wtMNV VP1. (A) *Residues contributing to the wtMNV P domain interface between A-type and C-type VP1 P domains. Yellow residues were identified as contributing to the interface through PDBePISA analysis. Non-contributing residues are highlighted in light blue, solvent-inaccessible residues are highlighted in dark blue, and the residue mutated in hsMNV (L412Q) is highlighted in magenta. Residues which form hydrogen bonds are indicated with an 'H'.* ***(B)*** *Atomic model showing the A:C VP1 P domain interface, with contributing residues coloured yellow and L412 coloured magenta. Part of the molecular surface in (C) is overlaid. Standard residue labels correspond to C-type VP1, italicised labels correspond to A-type VP1.* ***(C)*** *Molecular surface of an AB-type P domain dimer viewed from the direction of the black arrow in (B). The region of the A-type P domain that interacts with the C-type P domain is coloured in*

yellow. The C-type P domain residues that interact with the A-type P domain are also shown.

3.5.6. Focussed classification of hsMNV P domain dimers

As for wtMNV and hiMNV, focussed classification was carried out on CC-type P domain dimers to probe for any evidence of changes to the conformational landscape explored by P domains in hsMNV (Figure 3.33A). Interestingly, approximately 13% of subparticles were assigned to a 'non-contacting' class, compared to ~8% for wtMNV and 0% for hiMNV (Figure 3.33B,C), again suggesting that the accessibility of different P domain dimer conformations had changed. While there is no direct evidence that the proportion of subparticles in 'non-contacting' classes affects particle thermostability (rather, it acts as a simple indicator of changes to the overall conformational landscape), it is interesting to note that the direction of change relative to wtMNV was opposite to that for hiMNV. I.e., while heat inactivation of the virus was correlated with loss of 'non-contacting' subparticles, thermostabilisation was correlated with an increase in the proportion of subparticles in 'non-contacting' classes.



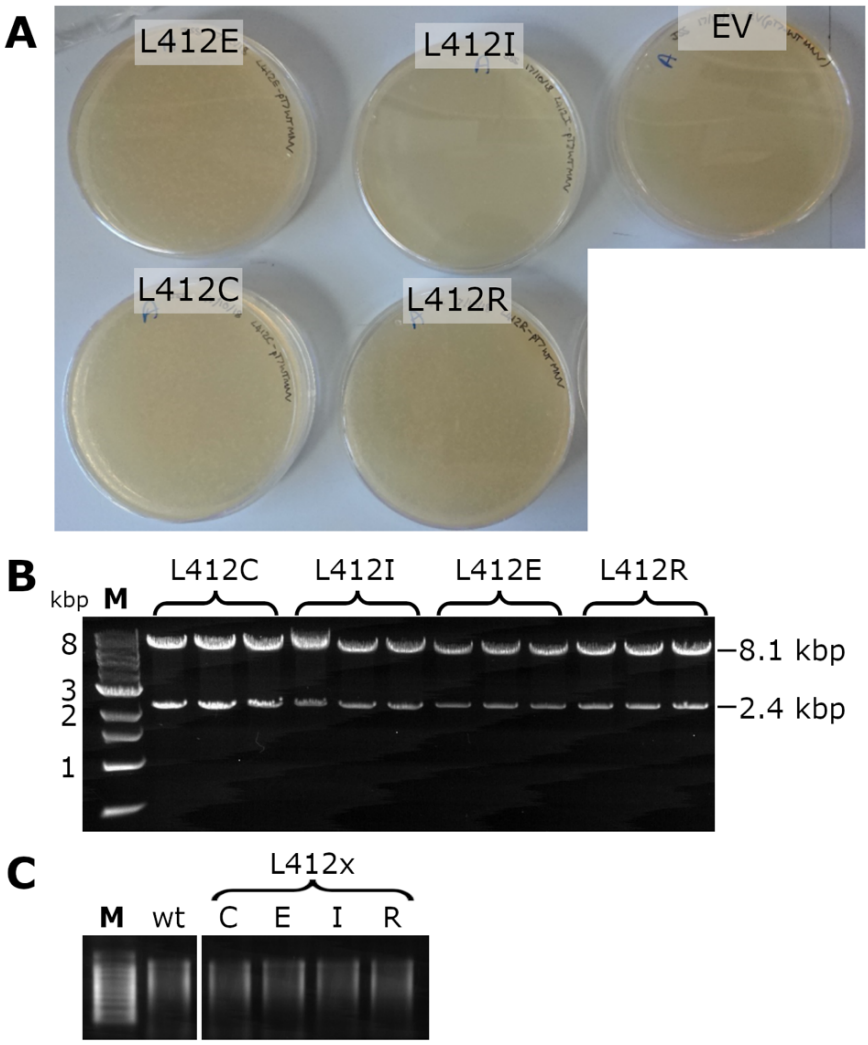
3.33. Focussed classification of hsMNV CC-type P domain dimers. (A) *All focussed classes from focussed classification of CC-type P domain dimers from hsMNV. The proportion of subparticles assigned to each class is indicated. One class displayed an inverted orientation in the Z-axis and is indicated by a blue dashed box.* ***(B)*** *Example of a 'non-contacting' class from focussed classification of hsMNV CC-type P domain dimers.* ***(C)*** *Schematic showing the proportion of subparticles (CC-type P domain dimers) assigned to either 'contacting' (grey) or 'non-contacting' (gold) classes for each of wtMNV, hiMNV and hsMNV. (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

3.5.7. Mutational analysis of VP1 residue L412

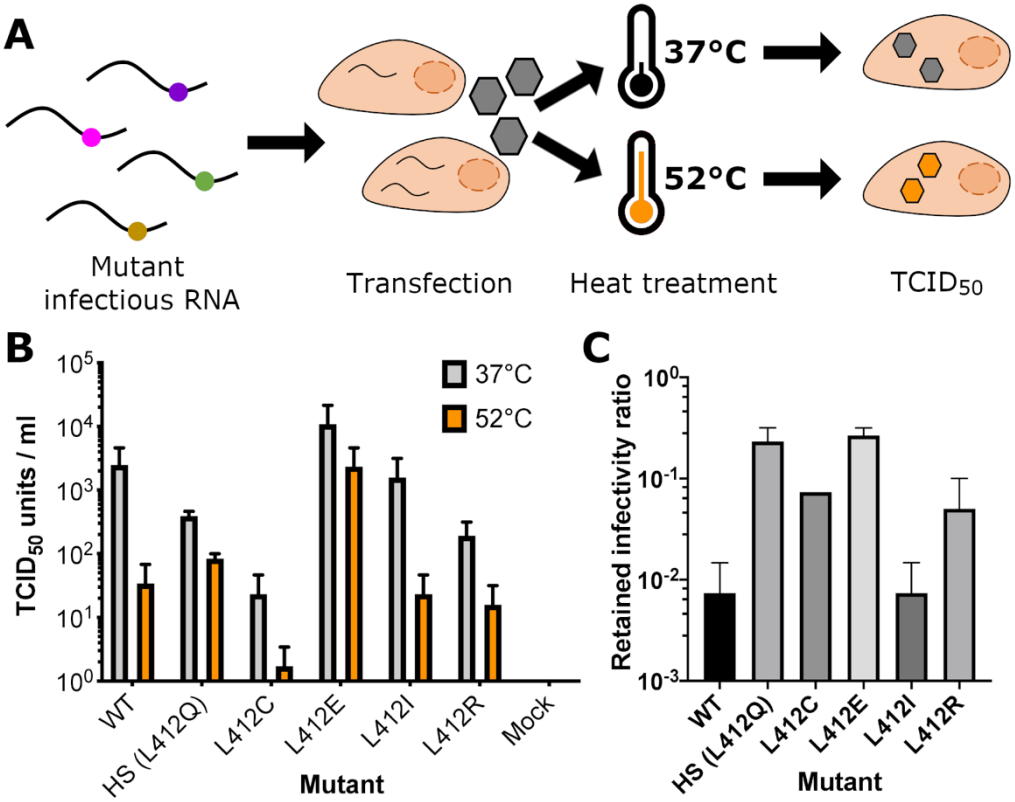
To explore the mechanism by which L412Q contributed to thermostabilisation in MNV VP1, site-directed mutagenesis was performed to generate infectious clones of MNV encoding different residues at VP1 position 412, including cysteine (L412C), glutamate (L412E, negatively charged), arginine (L412R, positively charged) and, as a control, isoleucine (L412I, hydrophobic). Overlap

extension PCR was used to introduce mutations to ORF2 (primers described in Table 2.2), and mutated inserts were ligated into a wtMNV backbone for transformation into *E. coli* (Figure 3.34A). Plasmid DNA extracted from transformed clones was endonuclease treated and analysed by agarose gel electrophoresis to rule out self-ligation (Figure 3.34B), then intact plasmid DNA was sequenced to confirm the presence of the mutations. Capped RNA was generated from linearised mutant infectious clones using a T7 polymerase reaction, and the presence of intact RNA was confirmed by agarose gel electrophoresis (Figure 3.34C). Equivalent amounts of RNA were transfected into BHK-21 cells to generate infectious virus. Importantly, BHK-21 cells do not express the cell surface receptor for MNV, so only a single round of replication was permitted, minimising the risk of reversion.

MNV-containing supernatant was harvested from the transfected cells and titrated for infectivity using a $TCID_{50}$ assay on RAW264.7 cells, following incubation at 37°C (unheated) or 52°C (heat stressed) for 30 min (Figure 3.35A). wtMNV, hsMNV (L412Q) and supernatant from mock transfected BHK-21 cells were included as controls. While overall infectious titres (regardless of heat treatment) differed between mutants, there was clear evidence that in addition to hsMNV (L412Q), L412E conferred improved thermal stability (Figure 3.35B,C). As expected, L412I behaved almost identically to wtMNV. L412C and L412R showed slight improvements in thermostability above wtMNV levels, but not to the same extent as L412Q/L412E, and for L412C this was complicated by the low overall infectious titre of recovered virus.



3.34. Site-directed mutagenesis of MNV VP1 L412. (A) *Competent DH5α cells were transformed with mutant MNV plasmid DNA (L412E, L412I, L412C, L412R) or empty vector (EV) as a negative control, then spread on agar plates and incubated at 37°C overnight.* ***(B)*** *Three colonies for each mutant were isolated and plasmid DNA extracted. Plasmid DNA was treated sequentially with the endonucleases AfeI and BspEI to cut out the mutation-containing insert from the template MNV backbone. The resultant DNA was analysed by agarose gel electrophoresis to confirm presence of the insert sequence and exclude any colonies containing self-ligated template. For all colonies, bands were detected at the expected molecular weights for insert (2.4 kb) and template (8.1 kb) DNA.* ***(C)*** *Capped and poly(A) tailed RNA was generated from MNV plasmid DNA (either wild type (*wt*) or L412C, L412E, L412I, L412R) and analysed by agarose gel electrophoresis.*



3.35. Evaluation of thermostability of L412 mutants by $TCID_{50}$ assay. ***(A)*** *Mutant infectious RNA was transfected into BHK-21 cells, leading to the production of infectious virus. Harvested virus was incubated at either 37°C or 52°C for 30 min before measuring infectious titre by $TCID_{50}$ assay.* ***(B)*** *Titres of different virus mutants determined by $TCID_{50}$ assay. Each mutant was titred after heating at 37°C (grey) or 52°C (orange). Supernatant from mock transfected cells was also titred as a negative control (Mock).* n = *2 ± SEM.* ***(C)*** *Ratio of infectivity between virus incubated at 37°C and 52°C from data shown in (B).*

3.5.8. Implications for VLP vaccine design

In addition to insight into the importance of P domains for norovirus stability and infectivity that come from these studies, hsMNV itself may be of particular benefit to VLP vaccine design. To check for any changes to antigenicity that may preclude the development of a thermostabilised VLP vaccine based on the L412Q mutation, MH and OA probed the ability of two neutralising anti-VP1 antibodies (2D3 and 4F9 (Kolawole *et al.*, 2014)) to bind to hsMNV in a

series of ELISA experiments (Snowden *et al.*, 2020). This revealed that hsMNV retained these major antigenic determinants, including at 64°C (which wtMNV was unable to tolerate) (Figure S2).

3.6. Discussion and conclusions

By determining the in-solution structures for a number of MNV variants, paired with structure-guided mutagenesis, it was possible to gain considerable insight into the dynamic nature of the capsid. As part of the work described in this chapter, I determined the first high-resolution solution structure of an infectious norovirus, also the first structure showing that a single strain of norovirus could exist in the 'contracted' conformation, as well as the 'expanded' conformation previously reported (Katpally *et al.*, 2010). While thermal inactivation did not capture an entry intermediate structure as intended, changes were observed to the conformational landscape explored by P domain dimers, and combined with similar structural analysis of hsMNV this implicated the P domain dimer conformational landscape in virion stability and infectivity.

Using wtMNV purified by an optimised 'S+S' (double sucrose gradient) approach, a 3.1-Å resolution density map of the wtMNV capsid was determined and used for atomic modelling of each VP1 quasi-conformer in the asymmetric unit. Interestingly, the most striking variation between quasi-conformers was observed in the N-terminal arms. It is possible to speculate that such differences between N-terminal arms could be important during particle assembly. For example, a structural study of a related calicivirus (feline calicivirus) proposed that VP2 may bind the capsid interior at an icosahedral 3-fold axis (Conley *et al.*, 2018). It would seem plausible that the contrasting organisation of different VP1 N-terminal arms could provide a structural cue to facilitate 'recognition' of the 3-fold axes, which are formed exclusively by B- and C-type monomers. In other viruses, such as tomato bushy stunt virus (TBSV), interactions between N-terminal arms have been found to be important for determination of differences in curvature between AB- and CC-type dimers (Harrison *et al.*, 1978). MNV clearly differs from TBSV in the arrangement of its N-terminal arms, for example, in TBSV, the C-type

N-terminal arms are more ordered than the A- or B-type N-terminal arms, and pass through the CC-dimer interface to enforce a flatter surface. In contrast, the C-type N-terminal arm in MNV was mostly disordered, and the resolved region was positioned away from the intradimer interface. Nevertheless, there are functional similarities between the organisation of N-terminal arms in MNV and TBSV, such as the formation of an ‘interlacing' network between adjacent A-type monomers at the five-fold icosahedral axis in MNV (observed for C-type N-terminal arms at the three-fold axis in TBSV). It is, therefore, plausible that N-terminal arms contribute to the determination of local curvature in a different way in MNV.

Comparison with other MNV structures revealed large differences in conformation. The wtMNV structure reported here was in a ‘contracted’ state, such that the P domains were closer to the S domains than was observed for two other ‘expanded’ state wtMNV structures (Katpally *et al.*, 2010; Sherman *et al.*, 2019). The functional relevance of these strikingly different conformations of the MNV capsid is not completely understood. At the time of this analysis, it was not clear whether the two states were both important for infection (e.g., entry intermediates) or immune evasion (e.g., providing the ability to alter the accessibility of epitopes), or whether one or both conformations were simply artefacts of purification/sample preparation. Equally, the molecular mechanism causing the capsid to adopt one particular state over another, or underlying any potential switching between the states, was not clear. The virus used to generate the structure reported here was derived from a molecular clone based on the CW1 strain of MNV-1, and genomes extracted from the same virus preparation were sequenced to confirm that no mutations had occurred during viral passage. The wtMNV used for the ‘expanded’ state wtMNV structures was also derived from the CW1 strain (Katpally *et al.*, 2010; Sherman *et al.*, 2019), but mutations may have arisen during passage so differences in the primary sequence of VP1 cannot be ruled out.

Alternatively, the different conformation may have been a result of differences in buffer composition, which may be reflective of the changing environment

that the virus is exposed to during the initial stages of infection (Scott and Gruenberg, 2011). A key difference between the buffers used here and in other studies was the metal ion content. Notably, Ca^{2+} and Mg^{2+} (which were present in the buffer used here, but not in the buffers used to generate the 'expanded' MNV structures reported elsewhere) have both been shown to enhance P domain-receptor interactions, and the authors of these studies hypothesised that this was due to stabilisation of the P domain (Katpally *et al.*, 2008, 2010; Nelson *et al.*, 2018; Sherman *et al.*, 2019). Conversely, Zn^{2+} ions (previously been implicated in norovirus VLP stability (Jung *et al.*, 2019)) were absent from the buffer used to generate the structures reported here. There was no evidence of metal ion density within the wtMNV density map, though attempts to identify any metal ions coordinated by the P domains were confounded by the poor local resolution. To probe for any potential effect of changes to ionic composition on wtMNV, negative stain EM and infectivity assays were performed following incubation of the virus with chelating agents, but these failed to reveal any clear changes to infectivity or gross morphology, perhaps because metal ions were bound sufficiently tightly to avoid being chelated by EDTA/EGTA.

Since the analysis reported here, progress in understanding the switch between 'contracted' and 'expanded' states of MNV has been made by other groups. CryoEM and NMR studies have shown that bile salts, particularly GCDCA, are able to induce wtMNV to adopt the 'contracted' conformation, and that this is important for hiding epitopes that bind neutralising antibodies (Creutznacher *et al.*, 2021a; Williams *et al.*, 2021a). However, this cannot explain the difference between the apo wtMNV structures reported here and elsewhere (Katpally *et al.*, 2010; Sherman *et al.*, 2019). Another study reported medium-resolution cryoEM structures of 'contracted' and 'expanded' MNV in PBS (without Ca^{2+} and Mg^{2+}), suggesting a reversible switch from the 'contracted' state to the 'expanded' state at pH values above 7 in the presence of 20 mM EDTA (Song *et al.*, 2020). The authors speculated that pH is the main trigger of the switch, though metal ions may bind to the capsid to stabilise the 'contracted' state. Collectively, these results indicate that MNV P domains

are highly dynamic and responsive to multiple different components of their chemical environment.

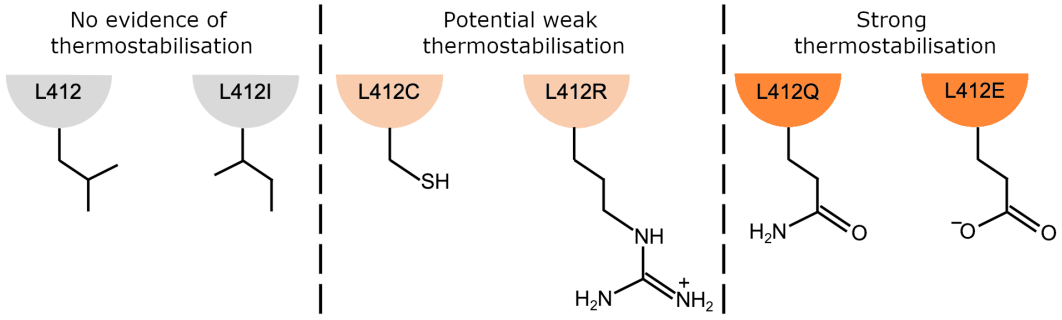
Following determination of the structure of wtMNV, a cryoEM structure of MNV inactivated through heat treatment was determined, in an attempt to capture an alternative conformation of the capsid that may be reflective of an entry intermediate. While no difference in conformation was detected, the hiMNV structure showed a relative ‘weakening’ of P domain density. In the absence of strong evidence for denaturation or detachment of P domains from the capsid, I deemed it likely that the ‘weakened’ density was a result of the fold of the P domain becoming more flexible, or the S-P linker becoming more mobile, permitting whole P domain dimers to occupy an expanded and/or altered range of positions relative to the S domains. Since cryoEM image processing is fundamentally an ‘averaging’ technique, greater variability in the fold or positioning of a domain would lead to a less uniform signal, and therefore ‘weaker’ EM density, as was observed. This was further supported by comparative focussed classification of P domain dimers from hiMNV and wtMNV. In particular, the complete absence of ‘non-contacting’ classes for hiMNV suggested a change to the range of positions/conformations that hiMNV P domains dimers were able to adopt. It is possible to speculate that, in changing this conformational landscape, certain P domain dimer conformations required for infection (e.g., for virus-receptor interactions or genome release) were made inaccessible by heat-treatment, rendering the virus non-infectious. While it is not possible to rule out changes to virion components that are unresolved in the hiMNV structure (such as VP2 or VPg), given the essential role that P domains play in initiating cellular infection and the differences observed between hiMNV and wtMNV P domain dimers following focussed classification, it is likely that the loss in infectivity was directly related to the changes observed in the P domains.

In separate experiments, a heat-stable mutant MNV was isolated and its stabilised phenotype linked to a single point mutation within the P domain, L412Q. Structural data for hsMNV clearly showed that L412Q correlated with disruption of the interface between A- and C-type P domains. This is

somewhat counter-intuitive; one would normally expect disruption of a molecular interface to lead to decreased stability, and there are a number of studies showing capsid-stabilising mutations in other non-enveloped viruses that work by stabilising molecular interfaces (Porta *et al.*, 2013; Fox *et al.*, 2017). While it is possible that L412Q may simply stabilise the fold of the P domain itself against denaturation (though there is no obvious structural explanation for this), focussed classification revealed that the conformational landscape explored by hsMNV P domain dimers had changed relative to the wtMNV. It is therefore plausible that this altered conformational landscape allows P domain dimers to access a 'protective' conformation upon heating that is not accessible (or is less accessible) to wtMNV, and which may only become apparent at high temperatures. To this end, it would be useful to determine the structure of hsMNV following heat treatment, and to compare the conformational landscape explored by its P domain dimers to that for hiMNV.

To further explore the idea that disruption of the inter-dimer interface was responsible for thermal stabilisation, additional mutagenesis was performed at VP1 residue 412, and in addition to L412Q (hsMNV), L412E was also able to provide thermostabilisation. This is perhaps unsurprising given the similarity of glutamine (Q) and glutamate (E) (Figure 3.36). Given its size and charge, it is likely that L412R would also cause disruption of the A-/C-type P domain interface, and there was some weak evidence for a slight improvement in stability. However, L412R failed to confer thermostability to the same extent as L412Q/L412E. This suggests that the mechanism behind the thermostabilisation conferred by L412Q is more complex than simply disrupting the interface. It is plausible that L412R would lead to a different conformational landscape explored by the P domain dimers than for L412Q/L412E, though structural analysis with additional focussed classification would be needed to confirm this. To extend this analysis, it would be useful to (i) mutate other residues contributing to the inter-dimer interface (including those on the opposite side of the interface) to verify that interface disruption, not limited to any single residue, results in capsid stabilisation, and

(ii) introduce salt bridges or disulphide bonds to ‘lock’ the interface together, which would be expected to reduce capsid stability.



3.36. Summary of L412 mutant phenotypes. *Schematic showing different side chains for residue 412 of VP1 across different mutants tested, categorised by thermostabilisation phenotype.*

Overall, these data build up a picture of VP1 P domains as highly dynamic components of the capsid. One facet of this dynamic nature is the synchronous (capsid-wide) conformational changes observed in response to changes in the chemical environment, including rotation by up to 100° and extension of the flexible S-P linker to lift the P domains further away from the S domain surface. At the same time, it is apparent that P domains can undergo continuous and highly asynchronous, yet limited movement of individual P domains relative to the S domains. While this constant movement might be expected to impede receptor engagement (and indeed, MNV has relatively low affinity for its cellular receptor [~200 μM] (Nelson *et al.*, 2018)), it is tempting to speculate that the dynamic nature of the P domains could give the virus an advantage in terms of immune evasion. By avoiding adoption of the ‘primed’ conformation until the virus is in the correct site for infection (recognised using cues such as bile acids and pH), MNV could minimise exposure of important conformational epitopes to the host immune system. Indeed, a recent study showed through cryoEM and ELISA experiments that three different antibodies against VP1 were unable to bind the virus in its primed, ‘contracted’ state (Williams *et al.*, 2021b). Subsequently, once the ‘contracted’ conformation has been adopted in the host gut, continuous ‘shapeshifting’ of individual P domains may be a

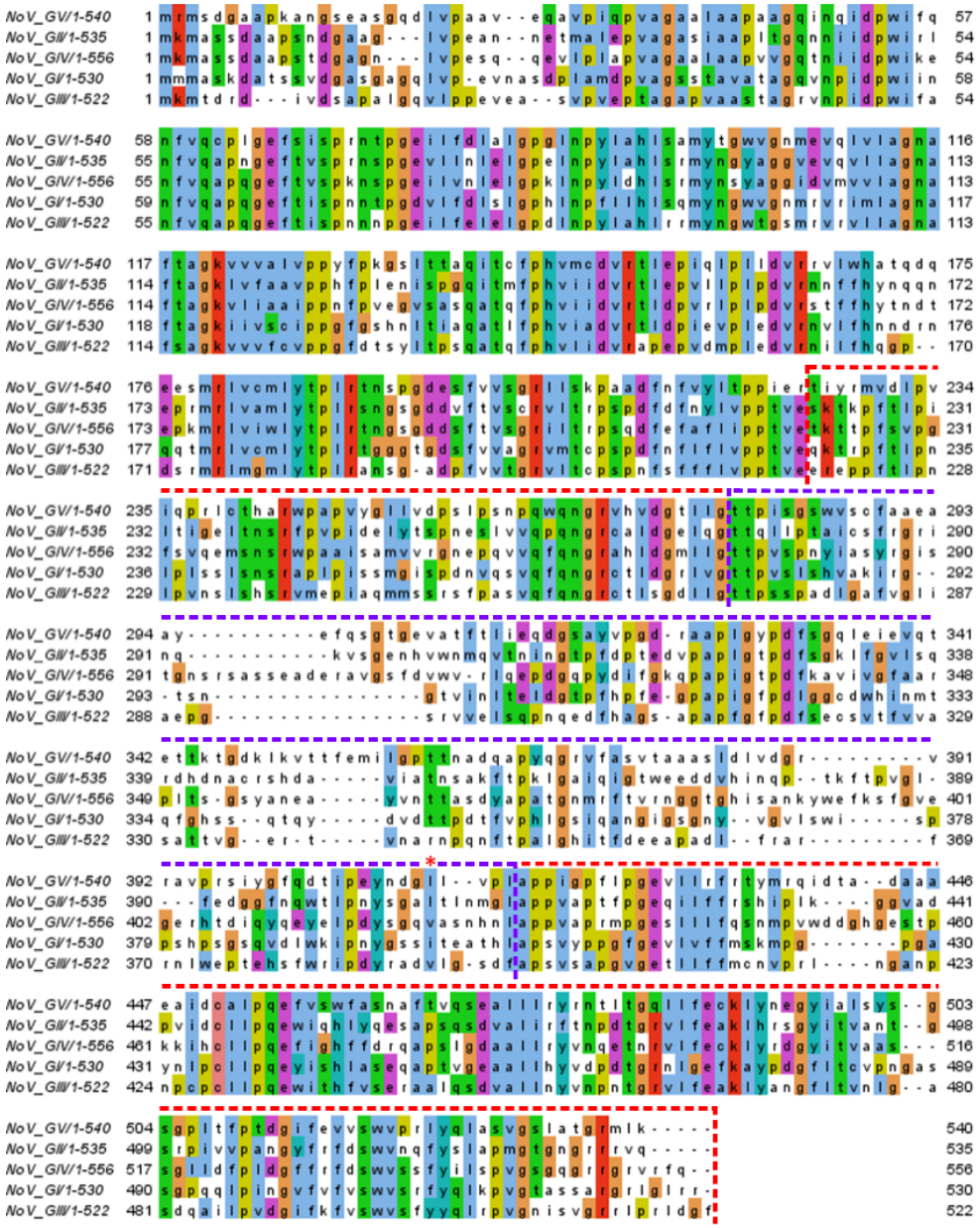
further mechanism by which the virus could counter the ability of antibodies to bind the virus.

Aside from the biological insight provided by these studies, the thermally-stabilised nature of hsMNV may prove useful for VLP vaccine design. While there are numerous examples of successful VLP vaccines (Donaldson *et al.*, 2018), and most norovirus vaccine candidates under development are VLP-based (Mattison *et al.*, 2018), VLPs do not contain the viral genome and therefore lack the potential stabilising effect that it can have on the capsid (Snijder *et al.*, 2013). This means that VLPs can be inherently unstable, which may be an issue during vaccine distribution. In fact, the most urgent need for norovirus vaccines comes from countries with warmer climates, and the cold chain is particularly difficult to maintain in rural, hard-to-reach regions. As such, stabilised VLP vaccines could be particularly useful.

It is likely that the stabilising effect of L412Q in hsMNV would also help to stabilise an MNV VLP if it were incorporated into VP1. This left two key questions as to whether L412Q could be useful for norovirus VLP vaccine development. Firstly, does hsMNV retain native antigenicity? ELISA experiments performed by MH and OA revealed that hsMNV retained the key epitopes bound by at least two neutralising antibodies, including at a temperature that wtMNV was unable to tolerate. Secondly, would a substitution equivalent to L412Q in human norovirus VP1 have a similar stabilising effect on the capsid? A sequence alignment of VP1 from norovirus genogroups GI – GV showed conservation of a hydrophobic residue at the equivalent position to MNV L412, suggesting that similar mutations may plausibly stabilise human norovirus VLPs (Figure 3.37). Thus, while a full and detailed study would be required to completely address these two questions, preliminary analyses are promising, and suggest that L412Q may be a useful tool in the development of stabilised VLP vaccines against norovirus. To explore these issues in more detail, a useful next step would be to generate HuNV VLPs with mutations equivalent to L412Q and compare their stabilities with wild-type VLPs, to confirm that this provides an equivalent stabilising effect. Ultimately, alongside *in vitro* characterisation of antigenicity, it would be

important to confirm that inoculation with mutant, stabilised VLPs can induce immune responses that protect against challenge with wild-type virus in animal immunogenicity trials.

The following chapter extends the work described above through structural investigation of MNV beyond the capsid, including analysis of non-capsid components of the virion and interactions with ligands.



3.37. Sequence alignment of norovirus VP1 from different genogroups. *VP1 sequences from different norovirus genogroups were aligned using Clustal Omega with default parameters and coloured according to the*

Clustal colouring scheme (Sievers et al., 2011; McWilliam et al., 2013; Li et al., 2015). Genogroups analysed included GV (NCBI reference sequence: YP_720002.1), GI (NP_056821.2), GII (YP_009237898.1), GIII (YP_009237901.1), and GIV (YP_009237904.1). The MNV (GV) VP1 sequence is shown in the top position, and L412 is indicated by the red asterisk. P1 and P2 subdomains are indicated by red and purple dashed lines, respectively. (This figure was adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)

4. MNV beyond the capsid – non-capsid components of the virion and interactions with ligands

4.1. Introduction

4.1.1. The limitations of symmetry averaging in cryoEM

The MNV structures described in Chapter 3 provide high-resolution insight into the molecular structure of the capsid. This high resolution in part arises from icosahedral averaging, where if the capsid is perfectly icosahedral, the information from each of the sixty copies of the asymmetric unit making up each particle can be combined to generate an 'average' asymmetric unit that is computationally symmetrised. This makes use of the redundant information in each particle, increasing the signal-to-noise ratio, leading to more accurate alignments and a better resolved reconstruction. However, this assumption is flawed. Every infectious icosahedral virion is inherently asymmetric as a result of the packaged viral genome, and many virions contain minor structural proteins organised in a way that deviates from the icosahedral symmetry of the capsid (reviewed in (Goetschius *et al.*, 2019b)). Noroviruses are no exception, with virions containing an asymmetric genome tethered to a single copy of VPg, along with a limited amount of VP2 thought not to follow the symmetry of the capsid. As such, symmetry averaging forces symmetry mismatches and therefore limits the information that can be extracted from reconstructions. In practical terms, this presents as a 'blurring' or absence of density for asymmetric features of the virion.

There are a number of approaches that can be taken to avoid symmetry mismatches or extract information about asymmetric features of a complex (reviewed in (Goetschius *et al.*, 2019a)), the most basic of which is to perform the reconstructions without any imposition of symmetry at all. However, even without imposed symmetry, the strong signal from the icosahedral capsid often dominates alignments in comparison to the usually less well-ordered asymmetric features, again leading to 'blurring' or absence of density. On top of this, there is usually a considerable reduction in resolution. Focussed classification, described extensively in Chapter 3 and elsewhere (Zhou *et al.*,

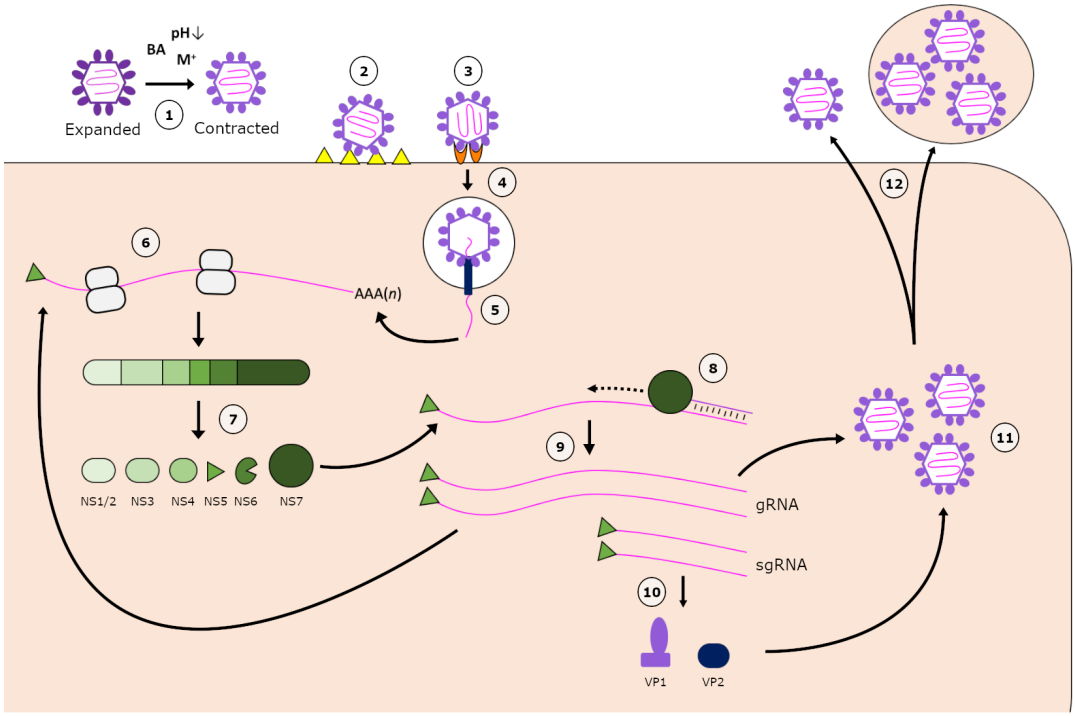
2015; Scheres, 2016; McElwee *et al.*, 2018; Conley *et al.*, 2019; Snowden *et al.*, 2020, 2021), is an alternative approach that uses orientational information generated during symmetrised reconstructions to probe individual symmetrically related positions in each particle. This is particularly useful when there is some relationship between the asymmetric feature and the symmetrical arrangement of the capsid. For example, the portal-like assembly formed by VP2 upon VP1 receptor engagement in feline calicivirus was found at a unique icosahedral three-fold axis of the capsid, such that only one of these sites was present in any individual virion (Conley *et al.*, 2019). This structure was completely absent from the symmetrised reconstruction, but focussed classification was able to resolve this.

Unfortunately, focussed classification is not useful for resolving the whole viral genome, which occupies a large volume inside the capsid spanning across all symmetrically related positions. Instead, an approach known as ‘symmetry relaxation’ can be applied. Based on the assumption that the genome, in whole or in part, is fixed in its position (or one of a number of discrete positions) relative to the capsid, symmetry relaxation uses orientational information from the capsid to guide the alignment of the genome, restricting sampling to a small range of angles around each of sixty symmetry-related positions. While this is similar to focussed classification in the sense that it makes use of orientational information from a previous symmetrised reconstruction, there is a key difference. While focussed classification treats every symmetrically related location in each particle as a separate subparticle for classification without alignments, symmetry relaxation searches a range of symmetrically related orientations to identify a single optimal alignment for each particle. This makes symmetry relaxation well-suited to resolving viral genomes, and the technique has been used successfully on a number of occasions (Liu and Cheng, 2015; Lee *et al.*, 2016; Li *et al.*, 2017; Ilca *et al.*, 2019). In one study of the *Cystoviridae* bacteriophage Φ6 and its segmented dsRNA genome, symmetry relaxation was used to resolve individual strands of dsRNA, facilitating the identification of putative interactions between RNA and the major capsid protein, P1 (Ilca *et al.*, 2019). The technique has also been used

to resolve RdRps transcribing RNA within a cypovirus (Liu and Cheng, 2015), and revealed asymmetric reorganisation of ssRNA within a picornavirus, coxsackie virus B3, upon interaction with receptor molecules embedded in lipid bilayer nanodiscs (Lee *et al.*, 2016).

4.1.2. The norovirus infectious cycle

Even when symmetry mismatches are accounted for computationally, individual cryoEM structures ultimately represent one snapshot of the virion. In reality, viruses undergo structural changes throughout the course of an infectious cycle, for example, to allow genome release, or in some cases, during maturation following assembly of nascent virions. Noroviruses are no exception, and must undergo structural changes at numerous points in the infectious cycle (Figure 4.1), many of which are triggered by interactions with host co-factors and receptor molecules.



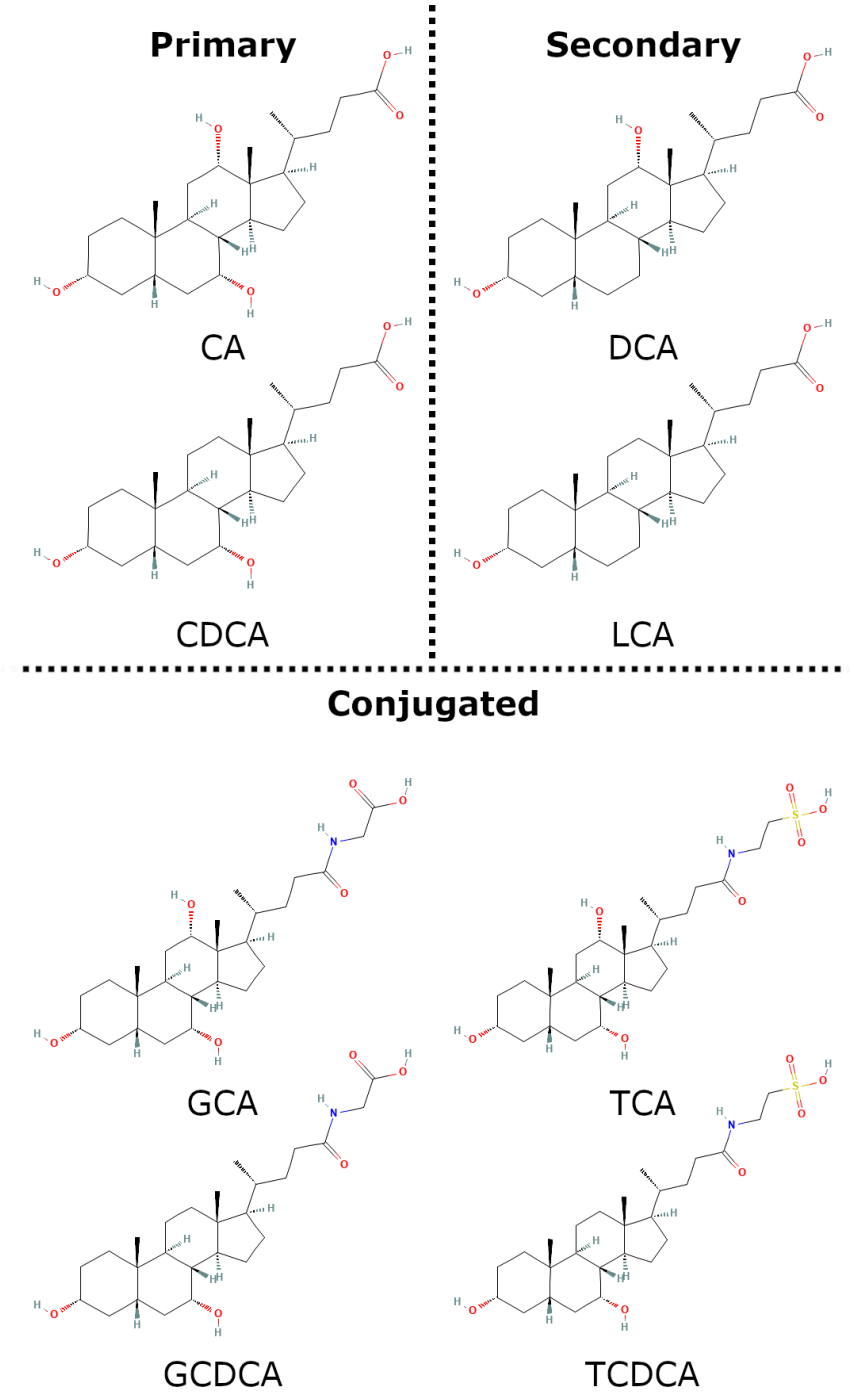
4.1. The norovirus infectious cycle. *Schematic showing an overview of the norovirus infectious cycle (not drawn to scale).* ***(1)*** *Norovirus virions interact with co-factors in the host gut such as bile acids (BA) and metal cations (M^+), which, along with a reduction in pH, favour a shift from an*

expanded conformation to a contracted conformation. ***(2)*** *Virions are concentrated at the cell surface by various attachment factors, whereby* ***(3)*** *they interact with the norovirus cellular receptor (CD300lf for MNV, unknown for HuNV), initiating* ***(4)*** *internalisation of the virion into the cell and* ***(5)*** *genome release, probably via a portal-like assembly formed from VP2.* ***(6)*** *The 'pioneer round' of translation of the viral genome results in the production of a polyprotein, which is* ***(7)*** *subsequently processed by the viral protease (NS6).* ***(8)*** *The viral polymerase (NS7) generates a negative sense copy of the viral genome which forms a dsRNA intermediate with the genomic RNA (gRNA) known as the 'replicative form'.* ***(9)*** *Many copies of the gRNA are produced, along with subgenomic RNAs (sgRNA), which are* ***(10)*** *translated to produce viral structural proteins VP1 and VP2.* ***(11)*** *Nascent virions are assembled from gRNA (linked to VPg [also termed NS5]), VP1 and VP2, and are then* ***(12)*** *released either lytically, or non-lytically in vesicles.*

4.1.2.1. Norovirus in the gut environment

Noroviruses must transit through the gastrointestinal (GI) tract of the host to reach the site of infection. Both human and murine noroviruses show dual tropisms, targeting (i) rare intestinal epithelial cells – enteroendocrine cells or tuft cells respectively (Wilen *et al.*, 2018; Green *et al.*, 2020), and (ii) immune cells within the gut lining, including macrophages, dendritic cells, and T and B cells within GI lymphoid-associated tissue, such as Peyer's patches (Karandikar *et al.*, 2016; Grau *et al.*, 2017; Wobus, 2018). During transit through the GI tract to reach the small intestine, noroviruses are exposed to changes in chemical environment, especially in the concentration of cationic metal ions and pH. They also encounter other factors present in the GI system, including bile acids (Figure 4.2), metabolites of cholesterol that are secreted into the duodenum to emulsify lipid droplets. Primary bile acids such as cholic acid (CA) and chenodeoxycholic acid (CDCA) are synthesised in the liver, and are usually conjugated with glycine or taurine to form glychocholic acid (GCA), taurocholic acid (TCA), glycochenodeoxycholic acid (GCDCA) and taurochenodeoxycholic acid (TCDCA) (reviewed in (Chiang, 2009)). Following

secretion into the intestine, these can be further metabolised by gut bacteria to form secondary bile acids, such as deoxycholic acid (DCA) and lithocholic acid (LCA).



4.2. Primary and secondary bile acids. *Chemical structures of several important human bile acids, taken from the NCBI PubChem database. CA – cholic acid (PubChem CID: 221493); CDCA – chenodeoxycholic acid (CID:*

10133); GCA – glychocholic acid (CID: 10140); TCA – taurocholic acid (CID: 6675); GCDCA – glycochenodeoxycholic acid (CID: 12544); TCDCA – taurochenodeoxycholic acid (CID: 387316); DCA – deoxycholic acid (CID: 222528); LCA – lithocholic acid (CID: 9903).

At the time of commencing this work, it was known that bile acids could impact norovirus infection, with bile shown to promote (or in some cases be essential for) human norovirus infection *in vitro* via an unknown mechanism (Ettayebi *et al.*, 2016). X-ray crystal structures of isolated MNV P domains had also been reported, showing that GCDCA and LCA bound to a central pocket formed between each monomer of a P domain dimer, at the junction between P1 and P2 subdomains (Nelson *et al.*, 2018). Cell binding and infectivity assays performed as part of the same study showed that the ability of MNV to attach to and infect murine microglial cells was enhanced in the presence of GCDCA (LCA was not tested). As such, bile acids were suspected to be important cofactors that would bind to noroviruses in the small intestine and promote infection. However, the effect of bile acid binding on capsid conformation (and in particular, whether this was related to the 'contracted' and 'expanded' morphologies described in Chapter 3) was unclear.

Since then, in parallel with the work described in this chapter, a number of research groups have reported breakthroughs in our understanding of how noroviruses interact with bile acids and other chemical components of the GI tract environment. A number of structural investigations of MNV indicate that bile acids, metal ions and low pH all contribute to stabilising P domains, and favour the 'contracted' conformation over the 'expanded' conformation (Sherman *et al.*, 2019; Song *et al.*, 2020; Creutznacher *et al.*, 2021a; Williams *et al.*, 2021b). This is thought to prime the virus for receptor binding and inhibit antibody binding (Williams *et al.*, 2021a). Interestingly, in human noroviruses, bile acids have been observed to bind at a different site to that observed for MNV, at the distal end of the P2 subdomain (Kilic *et al.*, 2019). However, the outcome of bile acid binding appears to be similar, with binding associated with the stabilisation of several P domain loops to prime the virion for binding to cellular attachment factors. Aside from direct effects on the capsid, bile

acids have also been observed to induce cellular responses that may promote infection, at least for some strains of HuNV in an intestinal enteroid system (Murakami *et al.*, 2020). For example, GCDCA was found to upregulate endocytosis in human small intestinal enteroid cultures (HIEs) via a fluorescent dye-based assay, whereas this effect was not observed after incubation of HIEs with a bile acid that, in contrast to GCDCA, did not promote norovirus infection (Murakami *et al.*, 2020).

4.1.2.2. Cell adhesion

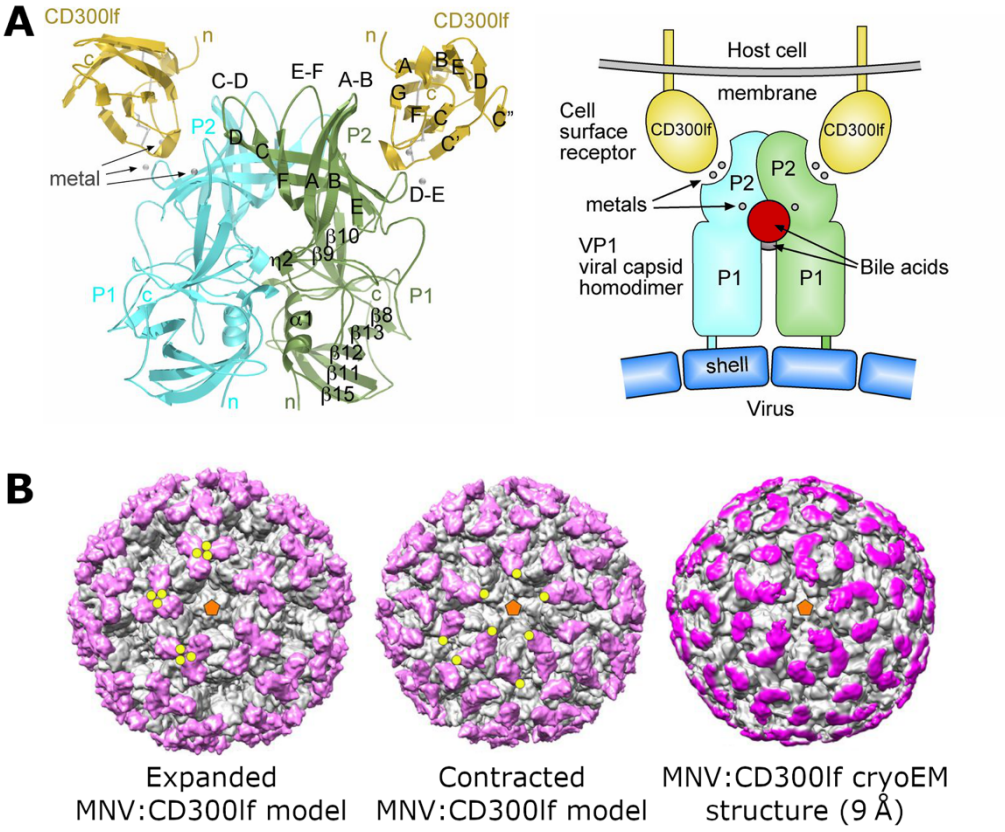
Upon reaching the small intestine and having been primed for infection through interaction with various chemical cofactors, noroviruses must attach to permissive cells and release their genome into the cytoplasm to initiate cellular infection. Histo-blood group antigens (HBGAs) are key cellular attachment factors for many HuNVs (reviewed in (Graziano *et al.*, 2019)) and appropriate HBGA expression was found to be critical for infection for some HuNV strains in an intestinal enteroid system (Ettayebi *et al.*, 2016). These complex carbohydrates are conjugated to cell surface proteins or lipids on many different types of cell, including intestinal epithelial cells (reviewed in (Hakomori, 1999)), or as free oligosaccharides in mucosal secretions. Individuals with genetic polymorphisms rendering a key HBGA biosynthetic enzyme (α-1,2-fucosyltransferase, FUT2) non-functional are highly resistant to norovirus infection (reviewed in (Nordgren and Svensson, 2019)). An abundance of structural evidence shows that HBGAs bind to the P2 subdomains of many HuNV strains (Choi *et al.*, 2008; Hansman *et al.*, 2011; Singh *et al.*, 2016). Interestingly, HBGA-expressing bacteria (but not non-expressing bacteria) were able to act as a critical co-factor for GII.4 norovirus infection in an *in vitro* infection model based on B cells, suggesting a potential role for the gut microbiota in cellular adhesion and initiation of norovirus infection (Jones *et al.*, 2014).

While HBGAs are not considered important for MNV, several studies suggest that sialic acid moieties fulfil a similar role, acting as attachment factors (Taube *et al.*, 2009, 2012). However, a recent report contradicted this, suggesting

instead that sialic acid does not interact with MNV and is not important for MNV infection. Specifically, no chemical shift perturbations were detected in the NMR spectrum for MNV P domains following the addition of 3'-sialyllactose, and there was no evidence of reduced cell surface binding or infection for MNV in a sialic acid-deficient cell line (Creutznacher *et al.*, 2021b).

4.1.2.3. Cell entry and uncoating

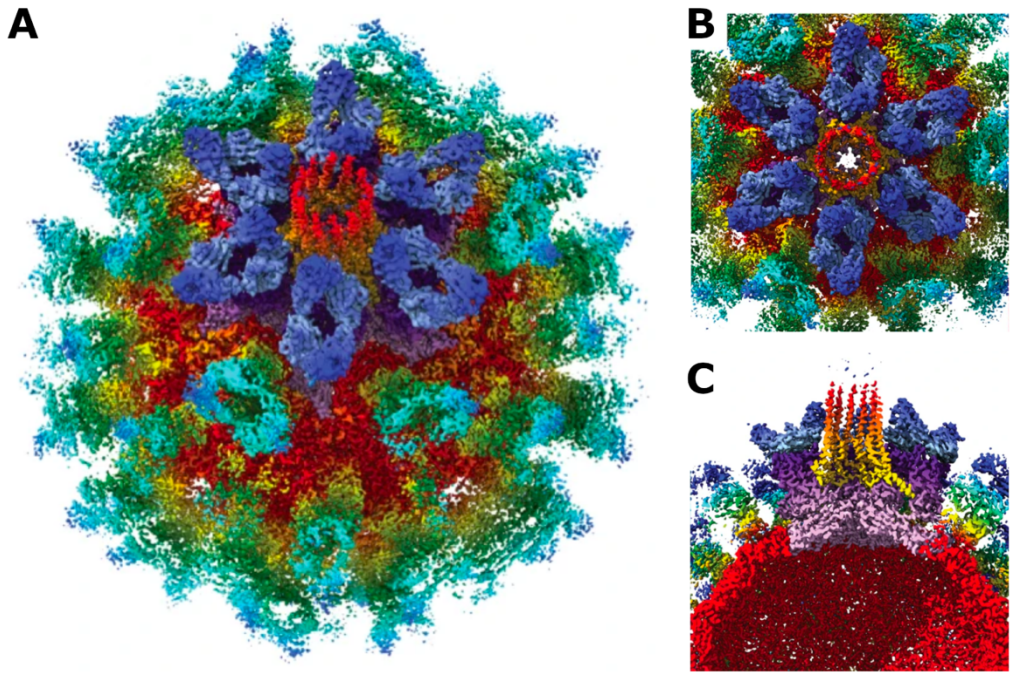
While attachment factors are important for concentrating the virus at the cell surface, receptors are the key molecules that initiate viral entry and uncoating. The cellular receptor for HuNV has not yet been identified, but CD300lf and (to a lesser extent) CD300ld, members of the immunoglobulin superfamily with roles in modulating immune cell responses, were identified as cellular receptors for MNV through genome-wide CRISPR/Cas9 screening (Borrego, 2013; Haga *et al.*, 2016). Structural data on CD300lf binding to isolated P domain dimers is available, revealing a binding footprint located between the AB and DE loops of the P2 subdomain, with each VP1 monomer engaging a monomer of CD300lf (Figure 4.3A) (Nelson *et al.*, 2018). Docking models that combined this structural information with previously reported whole-virion cryoEM structures suggested that CD300lf binding would lead to significant steric clashes in the 'expanded' conformation of the capsid, but that these would be minimised in the 'contracted' morphology (Nelson *et al.*, 2018; Sherman *et al.*, 2019). A 9-Å resolution cryoEM reconstruction of MNV decorated with CD300lf has been reported, and this was in agreement with the docking models, suggesting that CD300lf bound to the 'contracted' morphology of the capsid (Figure 4.3B) (Sherman *et al.*, 2019). No high-resolution, in-solution structure for the CD300lf-VP1 interaction has been reported to date.



4.3. The interaction between MNV and CD300lf. (A) *X-ray crystal structure (left) and schematic (right) detailing the interaction between an MNV P domain dimer and CD300lf. Figure adapted from (Nelson et al., 2018).* ***(B)*** *Simulated models of CD300lf (pink) in complex with either the 'expanded' or 'contracted' form of MNV (grey), shown next to the 9-Å resolution cryoEM reconstruction of MNV in complex with CD300lf. Figure adapted from (Sherman et al., 2019).*

Following receptor engagement, noroviruses are internalised and release their genome. For MNV, cell entry (at least in murine macrophages) is thought to be dependent on dynamin and cholesterol, but unlike feline calicivirus, is independent of clathrin, as well as caveolin and flotillin (Stuart and Brown, 2006; Gerondopoulos *et al.*, 2010; Perry and Wobus, 2010). As a result, the precise endocytic pathway that mediates uptake is unclear. While no specific uncoating mechanism has been elucidated for murine or human noroviruses thus far, the discovery of a portal-like assembly formed from VP2 upon receptor engagement in feline calicivirus (Figure 4.4) has stoked speculation

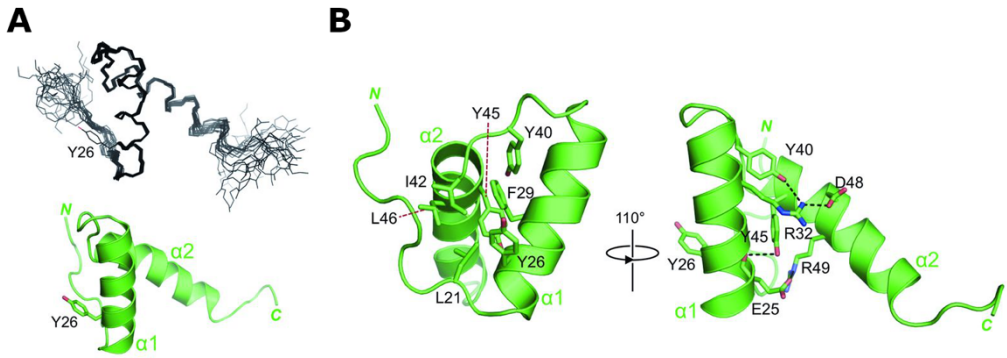
that a similar mechanism exists in noroviruses (Conley *et al.*, 2019). Receptor-bound capsids showed local rearrangements at a single icosahedral three-fold axis leading to the formation of a funnel-shaped tube from twelve copies of VP2, all extending away from the capsid with the hydrophobic N-termini presented distally. The authors postulated that these N-terminal regions may insert into the endosomal membrane to allow passage of the viral genome from the capsid interior, through the VP2 tunnel, into the host cytoplasm. In support of this, the region of ORF3 encoding the N-terminal region of VP2 in MNV was observed not to tolerate genetic insertions, suggesting they are critical to infection (Thorne *et al.*, 2012).



4.4. A portal-like assembly formed from VP2 in feline calicivirus. *Reconstructed focussed class from the cryoEM structure of feline calicivirus in complex with its cellular receptor, feline JAM-A, determined by (Conley et al., 2019). The VP2 portal-like assembly is shown* ***(A)*** *in the context of the whole capsid,* ***(B)*** *viewed down the icosahedral three-fold axis to highlight the pore in VP1, or* ***(C)*** *as a central section. Figure adapted from (Conley et al., 2019).*

4.1.2.4. Translation and genome replication

Once the norovirus genome is released into the cytoplasm, viral translation and genome replication must occur (reviewed in (Thorne and Goodfellow, 2014)). Initially, the viral RNA is translated, with cellular translation machinery recruited by the VPg covalently linked to the 5' end of the genome, which functionally substitutes for an m^7G cap. VPg has a wide range of functions and binding partners within the cell, and is known to be critical for the infectivity of calicivirus RNA (Burroughs and Brown, 1978; Guix *et al.*, 2007). In order to facilitate its multi-faceted role, the structure of VPg is highly flexible at the N- and C-termini, which flank a more compact α-helical core containing the key tyrosine residue (Y26 in MNV) that is linked to the viral RNA (Figure 4.5) (Leen *et al.*, 2013). Interactions between VPg and several components of the eukaryotic translation initiation machinery, including eIF3, eIF4E and eIF4G, have been described previously, at least for MNV, although only the interaction with eIF4G appears critical for translation (Chaudhry *et al.*, 2006; Daughenbaugh *et al.*, 2006; Chung *et al.*, 2014). It is thought that the C-terminal region of MNV VPg recruits eIF4G via its central HEAT-1 domain, and this ultimately drives recruitment of the 43S ribosomal pre-initiation complex (Leen *et al.*, 2016).



4.5. The solution structure of VPg from MNV. *The solution structure of MNV VPg determined by NMR, as reported by (Leen et al., 2013).* ***(A)*** *Backbone trace of the 20 lowest-energy conformers of VPg (upper), with representative conformer showing N- and C-termini (lower).* ***(B)*** *Enlarged view of a representative VPg conformer, showing selected side chains and electrostatic interactions. Figure adapted from (Leen et al., 2013).*

Being limited in their genomic capacity, noroviruses make use of several non-canonical expression mechanisms to achieve genetic economy. Non-structural proteins are synthesised through the translation of ORF1 to yield a single polyprotein precursor, which is co- and post-translationally processed by the viral protease (NS6) to generate mature proteins. Structural proteins are required later during infection and in greater quantities than non-structural proteins, and this is facilitated by de-coupling their production from that of the latter via expression from a polycistronic subgenomic (sg)RNA (Asanaka *et al.*, 2005; Thorne and Goodfellow, 2014). This involves a termination-reinitiation mechanism for translation of ORF3 (VP2), and in MNV uniquely, ORF4 (VF1) translation is thought to be initiated by leaky scanning and a ribosomal frameshift (Napthine *et al.*, 2009; McFadden *et al.*, 2011).

Several of the non-structural proteins produced during the 'pioneer' round of translation subsequently mediate the assembly of replication complexes within the perinuclear region (Wobus *et al.*, 2004). Genome replication occurs within these membranous vesicles, mediated by the viral RdRp (NS7). Initially, negative-sense RNA is synthesised to generate a double-stranded intermediate known as the replicative form (RF) (Qu *et al.*, 2016), and following this, positive-sense genomic RNA and sgRNA are synthesised. Initiation of positive-sense RNA synthesis is thought to be VPg-dependent, with VPg guanylylated at Y26 (MNV) or Y27 (HuNV) to serve as a primer for extension by the viral RdRp (Rohayem *et al.*, 2006; Belliot *et al.*, 2008; Olspert *et al.*, 2016). Conversely, initiation of negative-strand synthesis is thought to be independent of VPg, instead proceeding through a *de novo* initiation mechanism (Rohayem *et al.*, 2006).

4.1.2.5. Particle assembly and egress

Our understanding of norovirus particle assembly is limited, with most of our current knowledge derived from studies of VLPs. Expression of VP1 alone is sufficient for formation of VLPs with identical morphology to native virions (though it should be noted that $T = 1$ and $T = 4$ HuNV VLPs have also been

reported (Jung *et al.*, 2019)), and experiments with deletion mutants revealed that assembly is mediated by the S domains, suggesting this is also true for virion assembly in the context of infection (Bertolotti-Ciarlet *et al.*, 2002). Mass spectrometry of norovirus VP1 VLPs under conditions permissive or non-permissive for assembly was used to propose a putative assembly pathway, involving the addition of dimers to a decameric assembly nucleus that constitutes a five-fold symmetry axis (Uetrecht *et al.*, 2011). While VP2 is not necessary for particle formation, it is essential for norovirus infectivity, and circular dichroism spectroscopy of VLPs with or without VP2 suggested that it also stabilises the secondary structure of VLPs in alkaline conditions (Lin *et al.*, 2014). It has also been postulated that VP2 may be important for packaging the viral genome (Vongpunsawad *et al.*, 2013). This is partly based on the observation that the capsid interior-facing surface of VP1 lacks the abundance of basic residues found in many other positive-sense ssRNA viruses and thought to be important for RNA packaging (reviewed in (Comas-Garcia, 2019)). Given its basic nature (e.g., MNV VP2 has a predicted isoelectric point around 10.2), VP2 may be able to provide a basic surface for recruitment of genomic RNA during capsid assembly.

Following assembly, nascent virions must exit the cell. It is known that norovirus infection results in induction of apoptosis, as does expression of the MNV polyprotein alone, and this may play a role in viral egress (Bok *et al.*, 2009; Herod *et al.*, 2014). Indeed, increased levels of epithelial cell apoptosis have been observed in intestinal biopsies from norovirus-infected patients, and MNV infection of RAW264.7 cells in the presence of caspase inhibitors to block apoptosis resulted in a ten-fold reduction in viral yield (Furman *et al.*, 2009; Troeger *et al.*, 2009). However, with pharmacological inhibition of apoptosis, MNV infection led to accelerated cell death via necrosis, so the reduction in yield may have been a consequence of shortening the window of time for replication rather than through any link to viral egress (Furman *et al.*, 2009). Recent observations of norovirus-containing vesicles through negative stain EM of stool samples from infected patients (HuNV) or from cell culture (MNV) also raise the exciting possibility that non-lytic release is important for norovirus (Santiana *et al.*, 2018). In support of this, over the course of a 70-

hour infection in RAW264.7 cells, ~100% of the total yield of MNV (determined by RT-qPCR) was released prior to loss of membrane integrity, suggesting release was driven by non-lytic egress. Furthermore, HuNV- and MNV-containing exosomes were shown to be infectious in cell culture (Santiana *et al.*, 2018). Indeed, there is now evidence to indicate that vesicle-cloaked clusters of virions are more resistant to environmental stresses than naked virions (Zhang *et al.*, 2021).

4.1.3. Understanding MNV beyond the capsid

While our understanding of norovirus infection has improved considerably since the discovery of the virus fifty years ago, a number of key questions remain unresolved. CryoEM is well-suited to address some of these issues.

A number of norovirus virion structures have been reported, but none described to date contain resolved density for the viral genome, VPg, or VP2. While a VP2 portal-like assembly was observed for feline calicivirus, its structure and organisation within the virion prior to receptor engagement remain unknown (Conley *et al.*, 2019). In particular, given the finding that the vast majority of virions had only one portal-like assembly, it would be useful to know whether VP2 is distributed throughout the virion and only assembles into a portal at a single three-fold axis, or whether the location of VP2 is pre-determined and restricted to a single axis even prior to receptor engagement. It would also be useful to structurally characterise any interactions between VP1 and other capsid components, as this may yield more insight into viral assembly and genome packaging – processes that could be targeted in the development of anti-noroviral agents.

Beyond the virion itself, there remains a gap in our understanding of how the MNV virion interacts with its cellular receptor, CD300lf. A high-resolution X-ray crystal structure detailing the interaction interface between an isolated P domain dimer and CD300lf has been reported, as has a low-resolution cryoEM structure of a whole virion showing the gross morphology of CD300lf-bound MNV, but as yet there is no structure bridging the gap between them (Nelson

et al., 2018; Sherman *et al.*, 2019). A high-resolution cryoEM structure of MNV in complex with CD300lf would reveal any smaller scale conformational changes induced by receptor binding in the context of the whole capsid, and highlight any differences in the interactions between CD300lf and different quasi-conformers of VP1. Similarly, there was no high-resolution solution structure demonstrating bile acid binding at the time of commencing this work, though such structures have since been reported alongside the structure described here, as described in Sections 4.4 and 4.5.

4.1.4. Project aims

In order to address the knowledge gaps detailed above, I aimed to do the following:

- generate more MNV for structural analysis, including through use of an alternative cell line to probe for any improvements in viral yield;
- use computational approaches to resolve asymmetric features of the MNV virion, including VP2, VPg and the viral genome;
- determine the structure of MNV in complex with a bile acid and identify any effect this has on the conformation of the virus.

More generally, this would help to clarify whether VP2 is distributed throughout the capsid interior or is only present at a unique, asymmetric site, with implications for virion assembly.

4.2. The structure of wtMNV grown in BV-2 cells

The structures of MNV described in Chapter 3 were all derived from virus cultivated in RAW264.7 cells (termed MNV.RAW). While this approach was successful, it was inefficient. Most virus was lost during purification, meaning many flasks of cells had to be infected in order to generate a sufficiently concentrated virus stock for structural analysis. In parallel with standard MNV cultivation in RAW264.7 cells (murine macrophage cells), MNV was grown in an alternative cell line, BV-2 (a murine microglial cell line), to examine whether

this would offer any improvement in virus yield in preparation for structural studies of non-capsid components of MNV and of MNV in complex with ligands. Following this, the structure of the BV-2 cell-derived MNV (termed MNV.BV2) was determined, primarily to act as a control for comparison with the MNV.RAW structures determined previously.

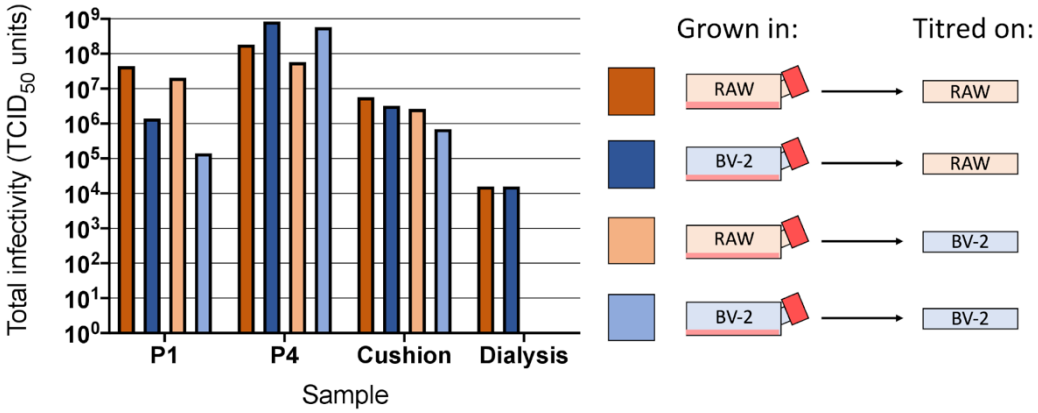
4.2.1. Cultivation of MNV in BV-2 cells

In parallel with the cultivation of wtMNV in RAW264.7 cells to generate virus used for the wtMNV structure described in Chapter 3, the same starting stock of wtMNV was used to infect BV-2 cells. In both cases, the virus was passaged four times, scaling up to 15 × T-175 flasks of infected cells. Infectious lysates were purified in parallel using the same method and dialysed into EM buffer ready for structural investigation.

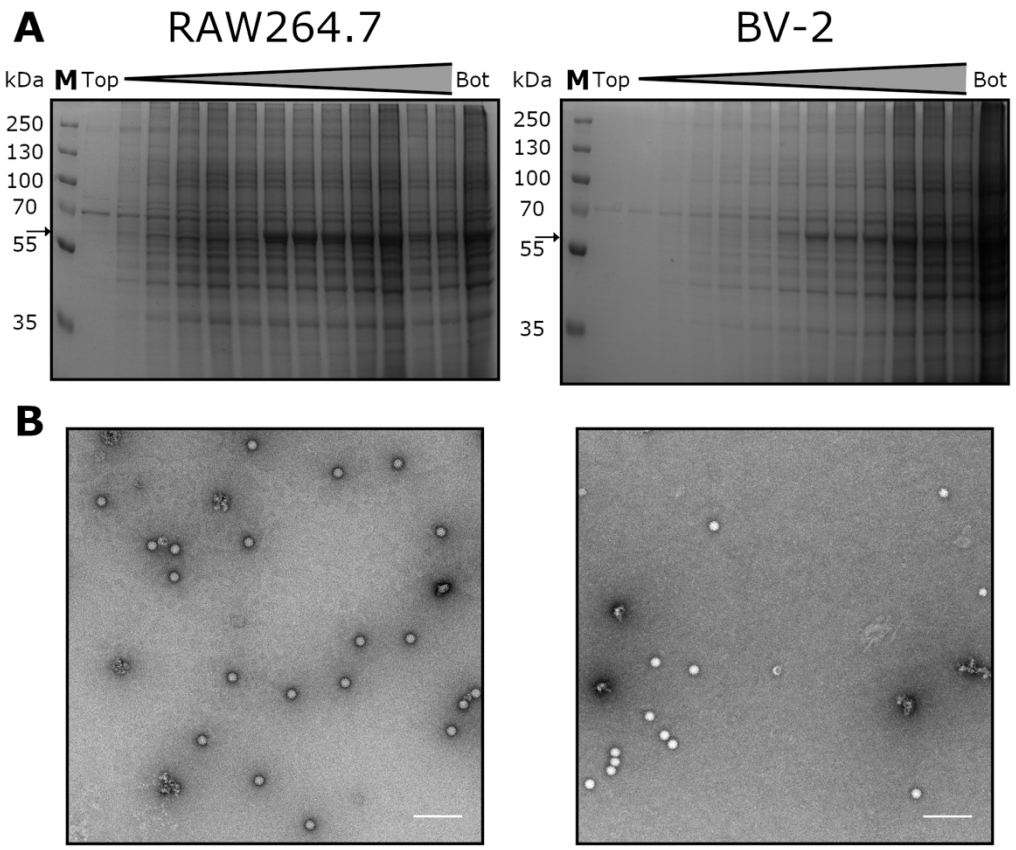
To compare yields of infectious virus from RAW264.7 and BV-2 cells, samples were taken for titration of infectivity following initial passage (P1), final passage (P4), sucrose cushion purification, and sucrose gradient purification/dialysis. These samples were titrated by $TCID_{50}$ assay on both RAW264.7 cells and BV-2 cells. While the total number of infectious units was initially lower for MNV.BV2 than for MNV.RAW following passage 1, this difference had disappeared by passage 4 at which point titres were much more comparable (Figure 4.6). Ultimately, there was no noticeable difference in total infectivity following purification, and while repeats of viral cultivation and purification would be required to make a definitive conclusion, these data suggested no clear advantage to using one cell line over another.

SDS-PAGE analysis was also performed on fractions from the initial sucrose gradient purification step for both wtMNV.RAW and wtMNV.BV2 (Figure 4.7A). Bands corresponding to VP1 (~58 kDa) were more intense for wtMNV.RAW, though this was also true of background bands, suggesting a greater degree of contamination. It also appeared that wtMNV.BV2 localised further into the gradient (peak at fraction 8) than the RAW264.7 cell-derived virus (peak at fraction 7), though this is likely a result of experimental error during sucrose

gradient set-up and fractionation rather than any difference in the sedimentation of the virus. SDS-PAGE analysis was also performed following the second sucrose gradient purification step, but the fractions were too dilute to see any bands (VP1 or background) with Coomassie staining. Negative stain EM analysis of both virus stocks following dialysis corroborated results from the initial SDS-PAGE analysis, suggesting that wtMNV.RAW was more concentrated, but also had more contamination, than wtMNV.BV2 (Figure 4.7B).



4.6. $TCID_{50}$ assay of wtMNV cultivated in RAW264.7 and BV-2 cells. *Samples of wtMNV grown in either RAW264.7 (red) or BV-2 (blue) cells were taken after the initial passage (P1), final passage (P4), pelleting through a 30% sucrose cushion (cushion), or sucrose gradient purification and dialysis (dialysis). Samples were titrated by $TCID_{50}$ assay on either RAW264.7 (strong colours) or BV-2 (pale colours) cells and total infectivity calculated (n = 1).*



4.7. Purification of wtMNV derived from RAW264.7 and BV-2 cells. (A) *10% SDS-PAGE analysis of initial 15-60% sucrose gradient purification step of wtMNV cultivated in RAW264.7 (left) and BV-2 (right) cell lines. Each lane represents a different fraction from the top (top, 15% sucrose) to the bottom (bot, 60% sucrose) of the density gradient. The black arrow indicates the expected molecular mass of VP1 (58 kDa).* ***(B)*** *Representative micrographs from negative stain of purified and dialysed wtMNV derived from RAW264.7 (left) and BV-2 (right) cell lines. Scale bars represent 200 nm.*

4.2.2. Preparation of BV-2 cell-derived wtMNV for structural analysis

Following purification, the wtMNV.BV2 was stored at 4°C for several months while wtMNV.RAW was used for other structural work. Following this period of storage, the BV-2 cell-derived virus was checked again by negative stain EM, which confirmed that the virus still appeared intact. As such, wtMNV.BV2 was taken forward for structural analysis by cryoEM, initially to serve as a 'control'

wild-type structure to compare with other planned structures of virus derived from BV-2 cells, for example, in complex with different ligands.

Multiple applications of wtMNV.BV2 (3 × 3 µl) were applied to glow discharged, ultra-thin carbon-coated lacey carbon grids for vitrification using a range of blotting times. A grid suitable for data collection was selected following screening of ice quality and particle concentration.

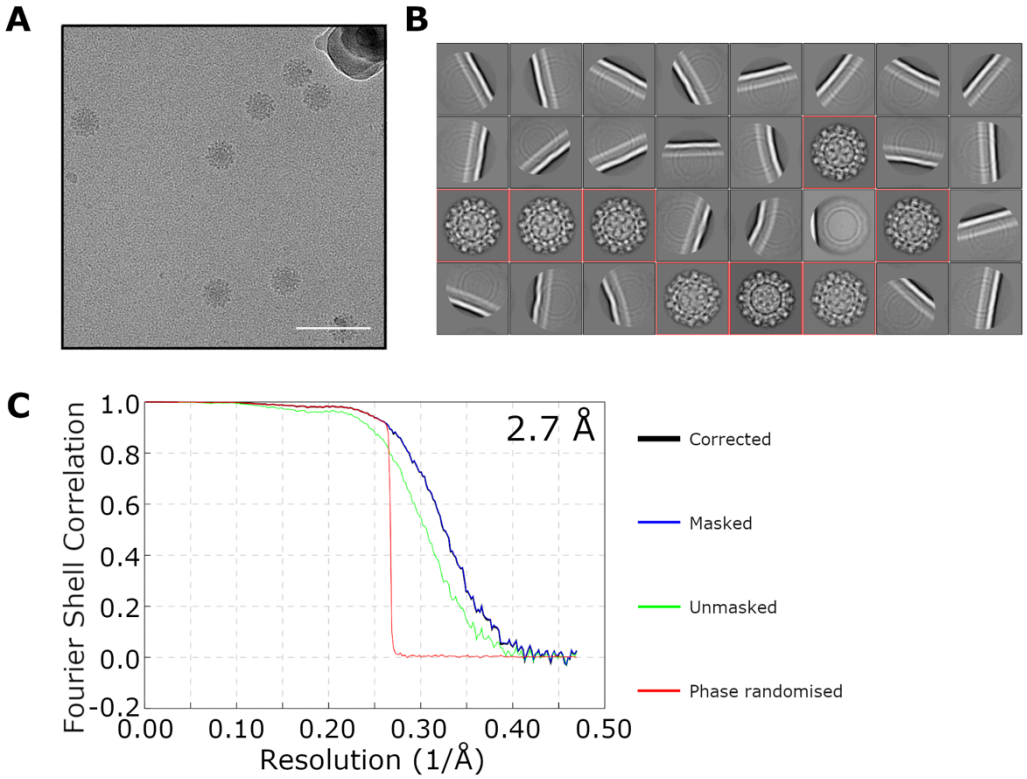
4.2.3. CryoEM data collection and image processing

As for previous MNV data collections, a dataset for wtMNV.BV2 was collected with an FEI Titan Krios microscope operating at 300 kV, at a magnification of 75,000×, giving a calibrated object sampling of 1.065 Å/pixel. Full details of parameters used for the wtMNV.BV2 data collection are given in Table 4.1.

From 11,620 micrographs collected (Figure 4.8A), low-threshold autopicking identified ~100,000 particles for extraction, although a significant proportion of this dataset corresponded to ‘junk’ particles (such as carbon backing and ice contamination, or virus particles picked over thick carbon). Two rounds of 2D classification (Figure 4.8B) refined the particle set to a stack of 24,928 contributing to the final reconstruction. Following CTF refinement and Bayesian polishing, a 2.7 Å resolution reconstruction was obtained for wtMNV.BV2 following sharpening (Figure 4.8C).

Table 4.1. Data collection and processing parameters for wtMNV.BV2. *Values marked with an asterisk (*) are approximate values. Precise values for these parameters are not available due to data loss.*

Sample	**wtMNV.BV2**
EMDB ID	*To be deposited*
PDB ID	*To be deposited*
Microscope	FEI Titan Krios
Camera	Falcon III
Voltage (kV)	300
Pixel size (Å)	1.065
Nominal magnification	75,000×
Exposure time (s)	1.0–2.0*
Total dose (e^-/Å^2)	55–80*
Number of fractions	59
Defocus range (μm)	−0.5 to −3.0*
Number of micrographs	11,620
Acquisition software	Thermo Scientific EPU
CryoEM map processing	
Number of particles contributing to map	24,928
Map resolution (FSC = 0.143) (Å)	2.7
Map resolution range around atom positions (Å)	2.4 – 4.0
Map sharpening B factor (Å^2)	−103
Residues modelled	
Quasi-equivalent position A	20 – 531
Quasi-equivalent position B	14 – 531
Quasi-equivalent position C	28 – 531
RMSD	
Bond lengths (Å)	0.0076
Bond angles (°)	1.29
Validation	
All-atom clashscore	5.30
MolProbity score	2.33
Rotamer outliers (%)	5.75
Ramachandran plot	
Favoured (%)	92.54
Allowed (%)	7.20
Outliers (%)	0.26



4.8. wtMNV.BV2 cryoEM data collection and processing. (A) *Representative micrograph from the wtMNV.BV2 dataset. Scale bar represents 100 nm.* ***(B)*** *Top 32 most populated classes from initial 2D classification of wtMNV.BV2 data. Particles from the classes highlighted in red (as well as other classes not shown here) were taken forward for further processing.* ***(C)*** *FSC plot for the wtMNV.BV2 map. The resolution (2.7 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).*

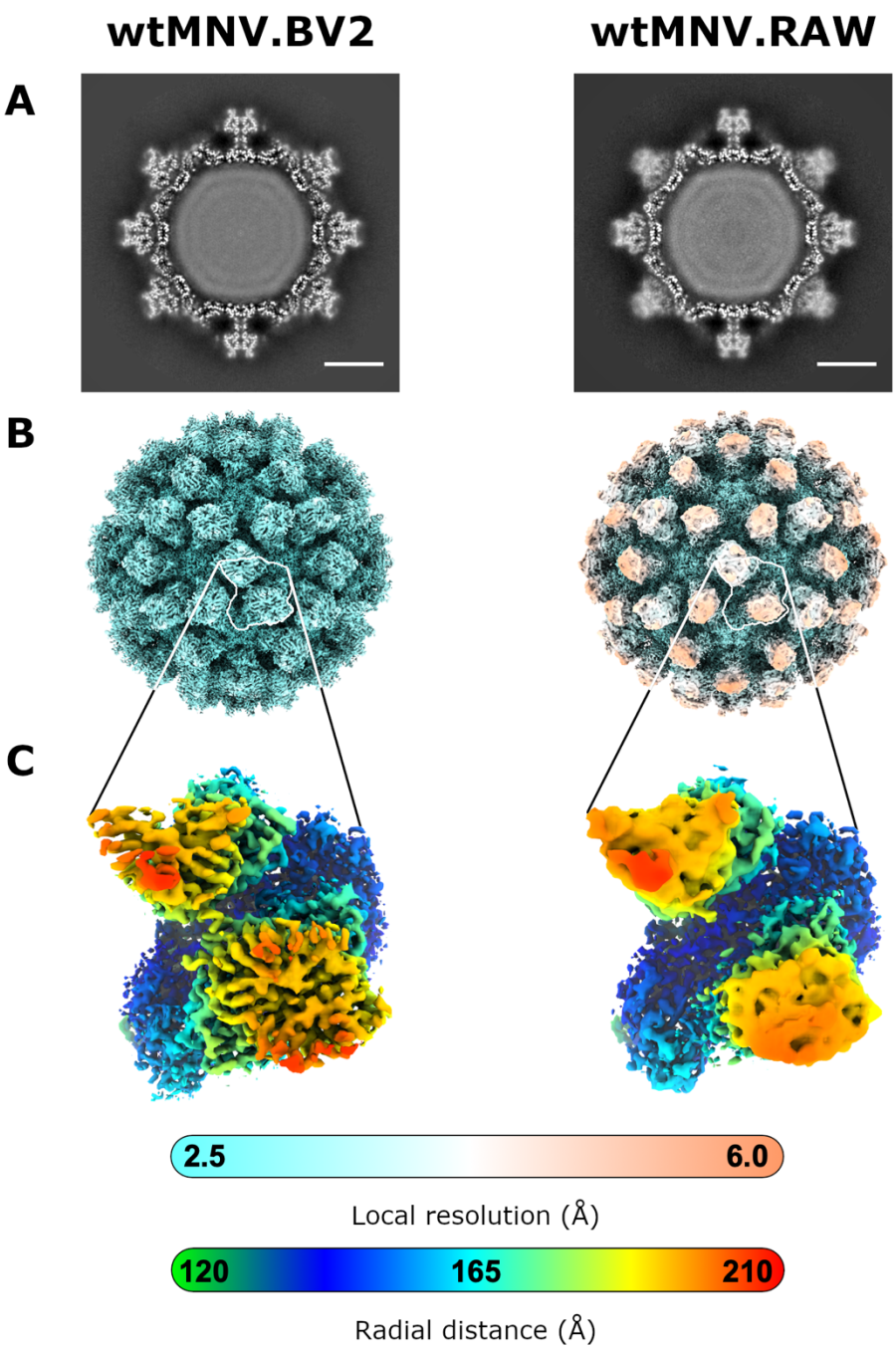
4.2.4. BV-2 cell-derived MNV shows better resolved P domains

Even before the final reconstruction was complete, it was clear that wtMNV.BV2 showed a significant improvement in local resolution for the P domains compared to wtMNV.RAW. This was evident from 2D slices of the 3D reconstruction showing improved pixel intensity and clear structural features in the P domains, as well as in the full isosurface representation (Figure 4.9A,B). While it was the case that the global resolution of the map had improved compared to wtMNV.RAW (from 3.1 Å to 2.7 Å), likely a result of more data contributing to the final reconstruction, the difference in resolution

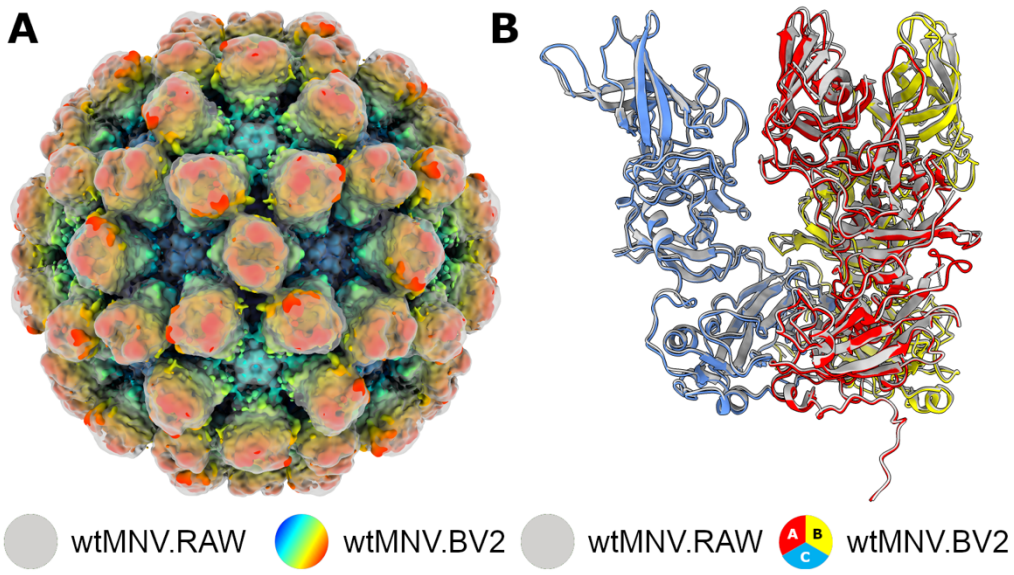
between S and P domains was much smaller for wtMNV.BV2 than it was for wtMNV.RAW (Figure 4.9B,C). Specifically, local resolution estimates for S domain and distal P domain density contoured to 1 σ were ~2.5 Å and ~4.0 Å, respectively, compared to ~2.8 Å and ~7.0 Å in wtMNV.RAW. This corresponded to a 1.8-fold increase in spatial frequency for P domain density, in contrast to a 1.1-fold increase for S domain density.

Aside from changes to local resolution, there were no significant differences observed between the two wtMNV structures (Figure 4.10). RMSD values between equivalent quasi-conformers of VP1 from atomic models of wtMNV.RAW and wtMNV.BV2 were calculated as just 0.58 Å [A-A], 0.66 Å [B-B] and 0.54 Å [C-C] using the 'MatchMaker' tool from UCSF Chimera.

Given that the dynamic nature of the P domains was thought to be responsible for their lack of resolution for wtMNV.RAW, I hypothesised that the relative improvement in P domain density may be a result of 'rigidification' of the P domains – i.e., the P domains becoming more fixed in their relative positioning. To investigate this, focussed classification was performed on the AB-type P domain dimers (identified as the more mobile type for wtMNV.RAW), using the same mask as was used for the AB-type P domain dimer focussed classification of wtMNV.RAW data described in Section 3.3.4. The wtMNV.BV2 focussed classification was performed in the same way as for wtMNV.RAW, but with 2$\times$ down-sampling and utilisation of 'fast subsets' to improve computational speed, compensating for the larger size and greater computational demand of the wtMNV.BV2 dataset.



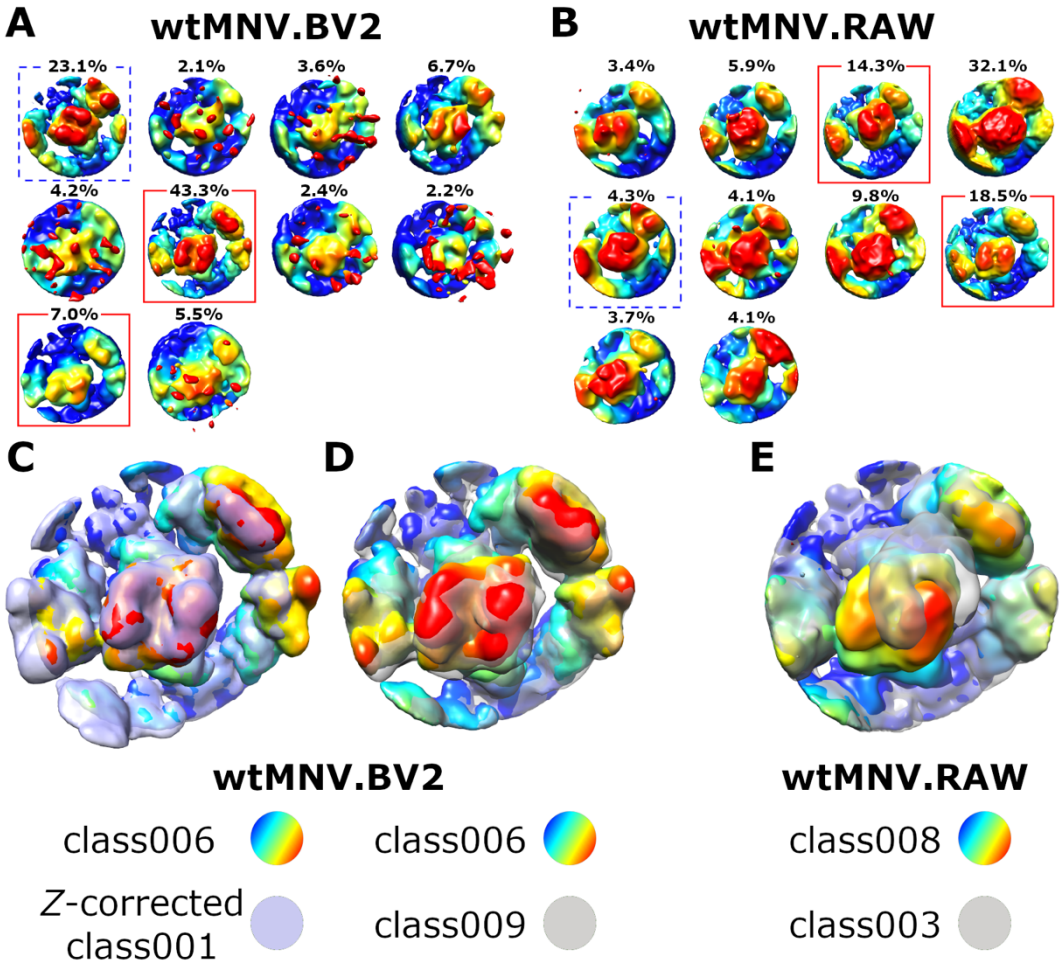
4.9. wtMNV.BV2 has better resolved P domains than wtMNV.RAW. (A) *Central sections through 3D reconstructions of wtMNV.BV2 and wtMNV.RAW (pre-sharpening). Scale bars show 10 nm.* ***(B)*** *Isosurface representations of wtMNV.BV2 and wtMNV.RAW density maps shown at ~2 σ, filtered by local resolution and coloured according to the local resolution colouring scheme indicated.* ***(C)*** *Density for the asymmetric units of wtMNV.BV2 and wtMNV.RAW, coloured according to the radial colouring scheme indicated.*



4.10. No evidence of conformational change in wtMNV.BV2. (A) *Overlaid isosurface representations of wtMNV.BV2 (radial colouring, ~1.5 σ) and wtMNV.RAW (grey, ~1 σ) density maps.* ***(B)*** *Atomic models for the three quasi-equivalent copies of VP1 overlaid after refinement into wtMNV.RAW (grey) or wtMNV.BV2 (coloured according to quasi-equivalent position: A-type [red], B-type [yellow] and C-type [blue]) density maps.*

While there was still evidence of limited variation in AB-type P domain dimer positioning, this was to a much lesser extent than for wtMNV.RAW (Figure 4.11A,B). Subparticle distribution was more concentrated into a smaller number of key classes for wtMNV.BV2, with the most populated well-defined class containing ~43% of subparticles. Interestingly, one class with an incorrect orientation in the *Z*-plane was observed and was heavily populated (~23%). As for previous focussed classifications, it is unclear why the *Z*-plane artefact occurred, but correcting the *Z* orientation revealed that the class overlaid well with another of the most heavily populated classes (Figure 4.11C). Combining these two equivalent classes covered a total of ~66% of subparticles. Thus, with the inclusion of the next most populated class, ~73% of all subparticles were assigned to one of two similar and well-defined positions (Figure 4.11D). This is in contrast to wtMNV.RAW, with the most populated class (~32%) being poorly defined, followed by two classes showing

the P domain dimer in two very different positions (~19% and ~14%) (Figure 4.11E). As such, it is clear that wtMNV.BV2 shows reduced variation in AB-type P domain dimer positioning compared to wtMNV.RAW, and this is likely the reason for the improved local resolution.



4.11. wtMNV.BV2 focussed classification. (A,B) *All focussed classes from focussed classification of AB-type P domain dimers from* ***(A)*** *wtMNV.BV2 or* ***(B)*** *wtMNV.RAW (also shown in Figure 3.16) datasets, coloured according to height and ordered based on arbitrary class name (class001 to class010 from left to right, top to bottom). Classes showing an inverted orientation in the Z plane are highlighted by a dashed blue box (wtMNV.BV2: class001; wtMNV.RAW: class005). The proportion of P domain dimer 'subparticles' assigned to each class is indicated.* ***(C)*** *Overlay of the most populated class from wtMNV.BV2 (class006, height colouring scheme) and class001 following correction to invert the Z orientation (lilac).*

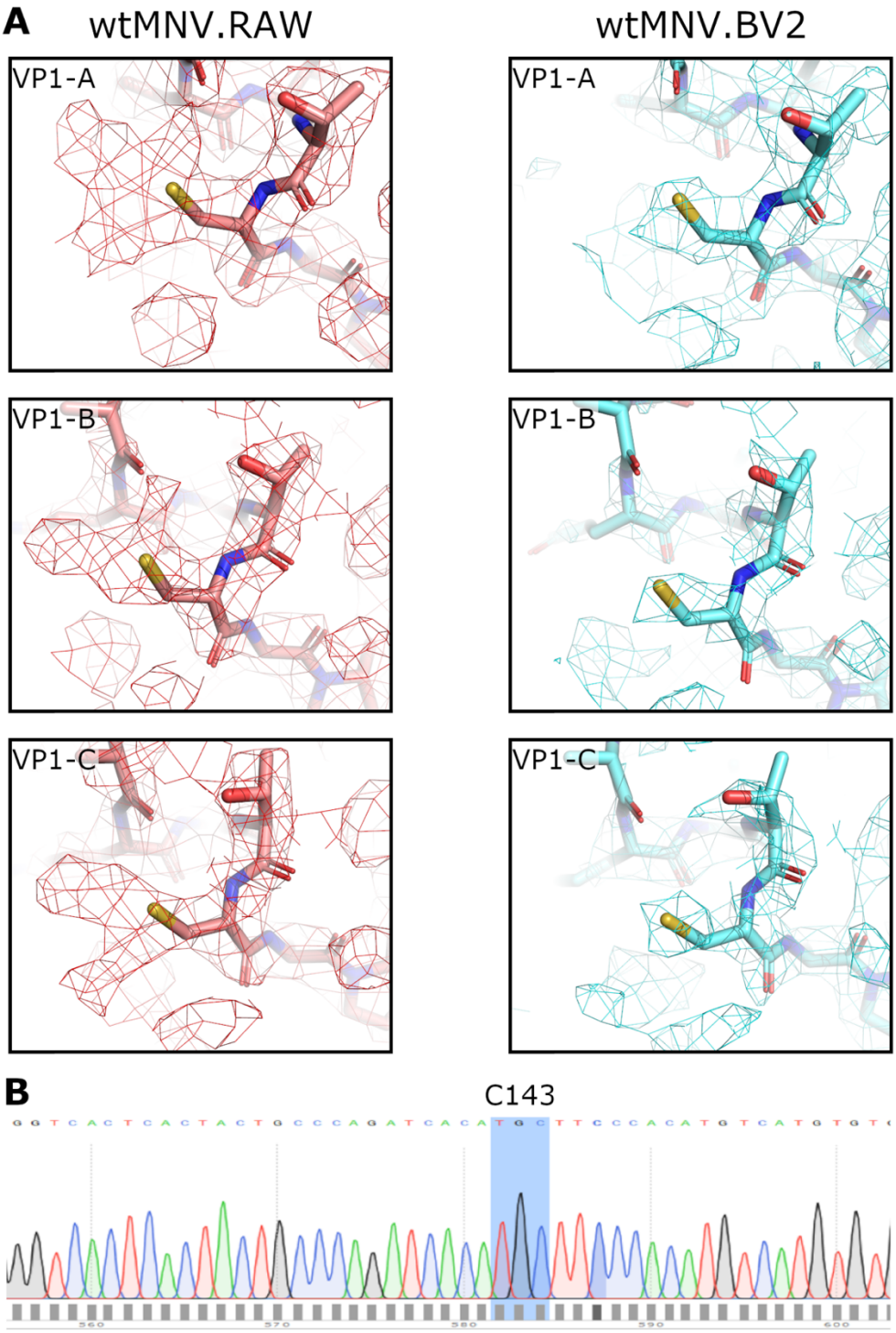
(D,E) *Overlay of the two most populated classes (excluding wtMNV.BV2 class001, which was in the incorrect Z orientation) from* ***(D)*** *wtMNV.BV2 (class006, height colouring scheme; class009, grey) and* ***(E)*** *wtMNV.RAW (class003, grey; class008, height colouring scheme). The selected classes are indicated by red boxes in (A) and (B), respectively. (Parts of this figure were adapted from (Snowden et al., 2020), published under a CC BY 4.0 licence.)*

4.2.5. Exploring a mechanistic basis for the rigidification of P domains

While these results suggested that wtMNV.BV2 P domain dimers were more rigid/fixed in their positioning than for wtMNV.RAW, it was not clear why the rigidity of the P domains had changed. To investigate the possibility that rigidification may be caused by a cell-line specific post-translational modification (PTM) to VP1 or bound chemical factor, the cryoEM density maps of wtMNV.BV2 and wtMNV.RAW were visually inspected for any evidence of additional density that could correspond to bound ligands or PTMs. This analysis was confounded by the low resolution of the wtMNV.RAW P domain density, meaning inspection was limited to the S domain for that map. Interestingly, there was strong additional density extending beyond the end of the side chain for one cysteine residue, C143, within the S domain of every quasi-conformer of VP1 in wtMNV.RAW (Figure 4.12A). To rule out mutation to a residue with a longer side chain, genomic RNA from the wtMNV.RAW sample used for cryoEM was sequenced, confirming that the sequence for ORF2 (encoding VP1) was unambiguously wild type as expected (Figure 4.12B). Interestingly, this additional density was completely absent for all copies of the same residue in wtMNV.BV2 (Figure 4.12A).

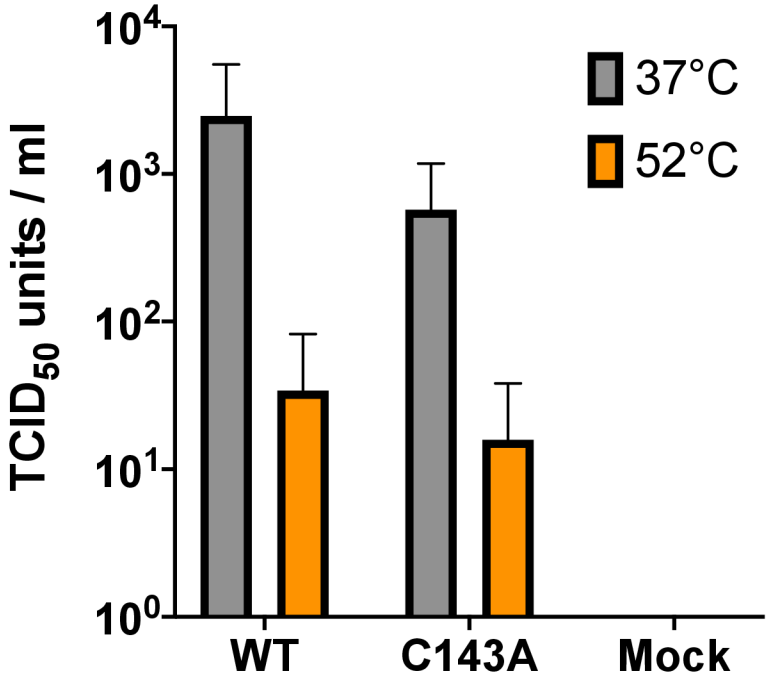
To probe the importance of this residue to virus infectivity, a point mutation was introduced to the infectious clone of MNV via molecular cloning, leading to a C143A substitution in VP1. The C143A mutant clone was transfected into BHK-21 cells before the lysate was harvested, then incubated for 30 min at either 37°C or 52°C prior to analysis by $TCID_{50}$ assay on RAW264.7 cells. At both temperatures C143A showed no statistically significant reduction in

infectivity, indicating that C143 (including any potential adduct/PTM) is not critical for MNV infectivity (Figure 4.13).



4.12. 'Adduct' density on wtMNV.RAW VP1 residue C143. (A) *Density maps for wtMNV.RAW (left, red) and wtMNV.BV2 (right, blue) are shown overlaid with the corresponding atomic models, centred around VP1 residue*

C143 in each quasi-equivalent position. In each case, maps were shown at an equivalent contour level (VP1-A: 4 σ; VP1-B, VP1-C: 6 σ). ***(B)*** *Section of the chromatogram from Sanger sequencing of cDNA generated from wtMNV.RAW genomic RNA. The nucleotides encoding VP1 residue C143 are highlighted in blue (TGC).*

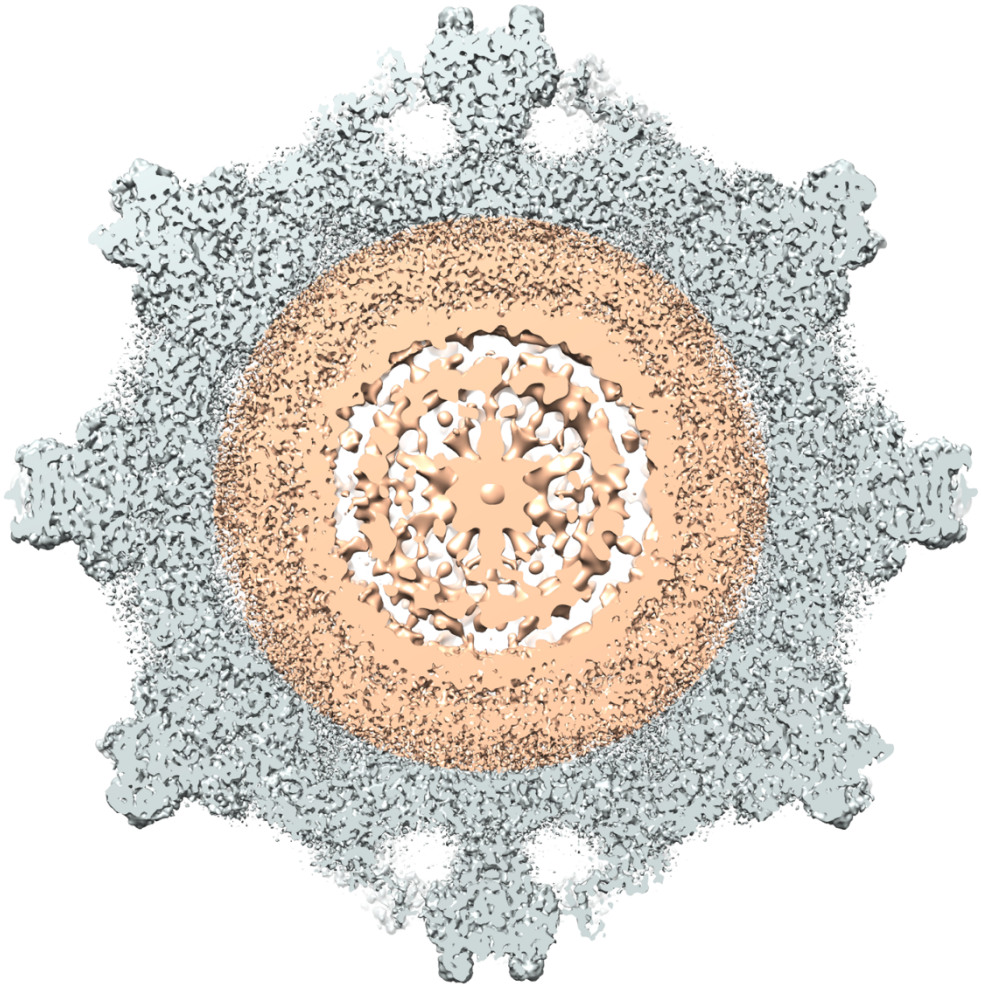


4.13. $TCID_{50}$ analysis of C143A mutant VP1 wtMNV. *Titres of wtMNV (WT) and C143A mutant MNV were determined by $TCID_{50}$ assay. Each mutant was titred after heating at 37°C (grey) or 52°C (orange). The data in this figure was collected from the same experiment shown in Figure 3.35. As such, data points for WT and Mock are the same as those shown in Figure 3.35. n = 2 ± SEM.*

4.3. Attempts to resolve the MNV genome, VP2 and VPg

While the cryoEM reconstructions described above gave useful insight into the MNV capsid and implications for norovirus biology, the information they provided on the packaged genome was limited to radial density distributions, only visible at low contour levels (Figure 4.14). Furthermore, despite the fact that both wtMNV and hsMNV were infectious, and therefore must have

contained VPg and VP2, neither of these proteins was resolved in any MNV reconstruction. In an attempt to resolve these additional components of the MNV virion, the wtMNV.BV2 dataset was taken forward for additional image processing.



4.14. Icosahedrally averaged genome density for wtMNV.BV2. *wtMNV.BV2 density map filtered by local resolution and shown as a section at ~0.8 σ. Genome density is shown in orange, capsid density is shown in grey.*

4.3.1. Asymmetric refinement to resolve the MNV genome

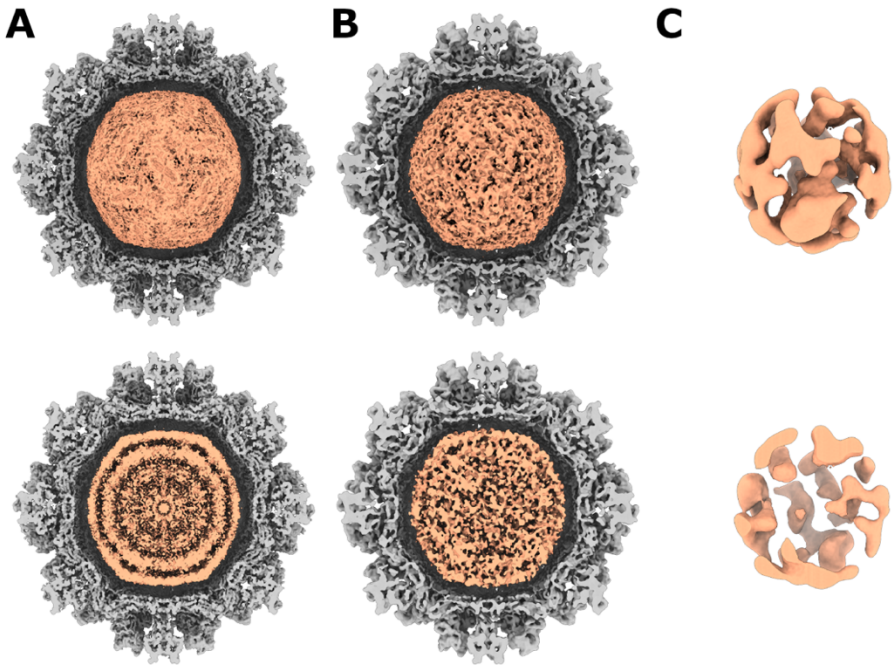
While the MNV capsid obeys icosahedral symmetry, the viral genome it contains does not, generating a symmetry mismatch. This symmetry mismatch limits the structural information available for the genome, though

reconstructions with icosahedral symmetry imposed still give an indication of the approximate distribution of genetic material along an axis projecting radially outwards from the centre of the capsid. For MNV, this revealed a multi-layered organisation with more defined layers closer to the capsid, and less defined layers closer to the capsid centre (Figure 4.15A).

To limit the effects of symmetry mismatch, asymmetric refinement of wtMNV.BV2 was performed, an approach that was previously able to resolve the ssRNA genome of bacteriophage MS2 (Dai *et al.*, 2017). In order to avoid any bias from the reference model, a new reference model was first generated without the imposition of symmetry. This reference map was low-pass filtered, then subjected to asymmetric (C1) refinement using the same particle images as before. As expected, this yielded a lower quality map than the symmetrised reconstruction, with a global resolution of 4.8 Å prior to sharpening. Capsid density overlaid well with the symmetrised reconstruction, and no obvious asymmetric features were observed. However, genome density appeared considerably different to that of the symmetrised reconstruction, likely due to reduced symmetry mismatch (Figure 4.15B). There was still a radially striated pattern visible towards the outer edge of the genome density, but this was much subtler than for the symmetrised reconstruction, probably due to the reduction in signal-to-noise ratio in an asymmetric refinement. Generally, genome density appeared more heterogeneous than before, though visual inspection did not reveal any density consistent with individual strands of RNA.

I hypothesised that even in the absence of symmetry imposition, the signal from the highly organised and symmetrical capsid may overwhelm the signal from the asymmetric genome during sampling. Thus, particles would be aligned primarily on the capsid, leading to poor alignment of the genome and weak symmetry mismatch even without symmetry imposition. To address this, a ‘genome-only’ refinement was performed. Firstly, capsid density from the symmetrised, high-resolution reconstruction of wtMNV was used as a mask to subtract capsid signal from raw particle images. These ‘subtracted’ particles were used for asymmetric refinement of a ‘genome-only’ reference model, generated using genome density extracted from the asymmetric wtMNV.BV2

reconstruction. During refinement, a spherical mask was applied to exclude any remaining signal beyond the genome (a diameter of 240 Å), ensuring that alignment proceeded based on genome signal alone. This yielded a low (>20 Å) resolution map with no obvious organisation or density consistent with individual strands of RNA (Figure 4.15C), likely a consequence of low overall signal, imperfect signal subtraction, intrinsic flexibility/disorder within the MNV genome and/or conformational heterogeneity between genomes packaged in different MNV particles.

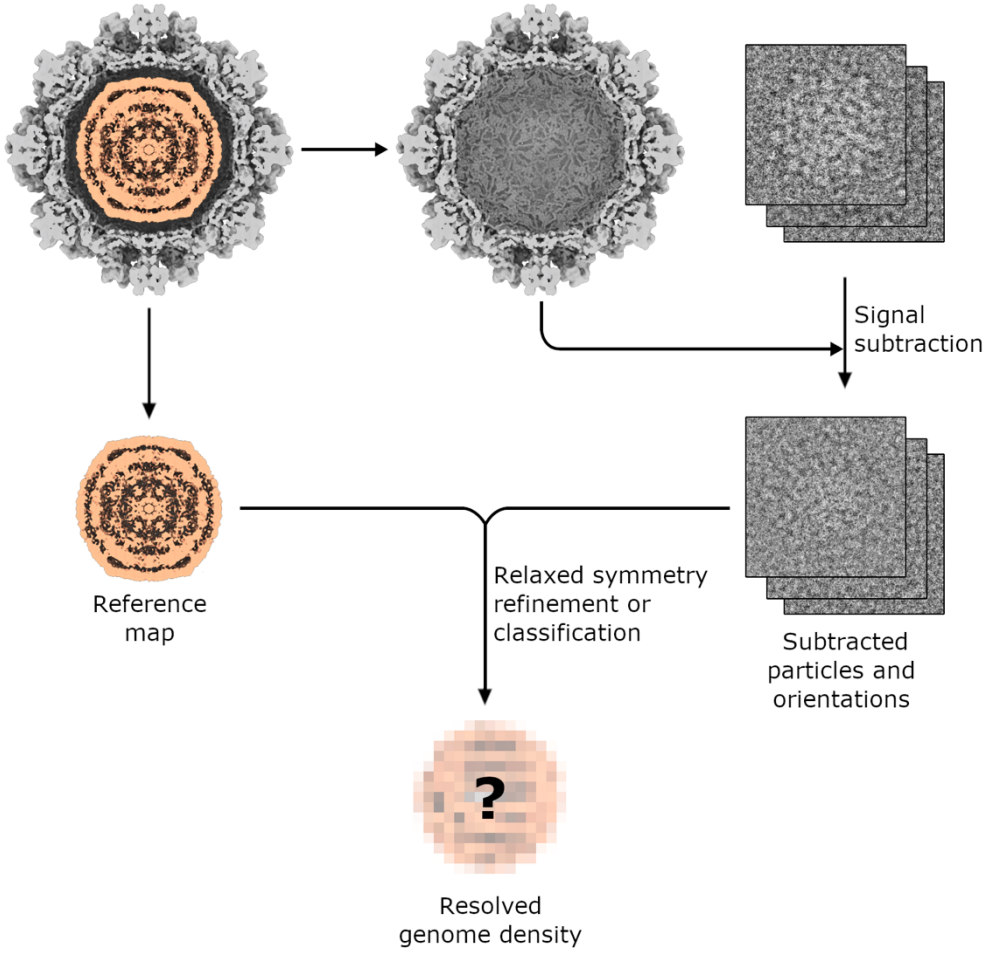


4.15. Asymmetric reconstructions of the wtMNV.BV2 genome. *Capsid (grey) and genome (orange) density from* ***(A)*** *the symmetrised (I1) reconstruction of wtMNV.BV2,* ***(B)*** *an asymmetric (C1) reconstruction of wtMNV.BV2, or* ***(C)*** *an asymmetric (C1) reconstruction of wtMNV.BV2 with subtraction of capsid signal prior to reconstruction and use of a spherical mask to exclude capsid signal during reconstruction. All density is shown prior to sharpening. For (A) and (B), capsid and genome density was separated into two individual density maps, then overlaid and shown at different contour levels for greater clarity. Capsid density is shown in sectional view at ~3 σ, with genome density overlaid in complete (upper) or sectional (lower) view, at ~5.5 σ.*

4.3.2. Symmetry relaxation to resolve the MNV genome

A significant limitation of the asymmetric refinement approach described above was the low intensity of the signal from the genome, making it difficult to achieve good alignments. To address this, a symmetry relaxation approach was pursued (Ilca *et al.*, 2019). As described in Section 4.1.1, symmetry relaxation uses the assumption that the genome, in whole or in part, is in a fixed position (or one of a number of discrete fixed positions) relative to the capsid, such that orientational information from the capsid can be used to guide alignment of the genome. Thus, particles with associated orientational information (from a previous symmetrised refinement) can be subjected to asymmetric 3D refinement or classification, with sampling restricted to a small range of angles around each of sixty symmetry-related positions rather than the global search that is performed in standard asymmetric refinement or classification.

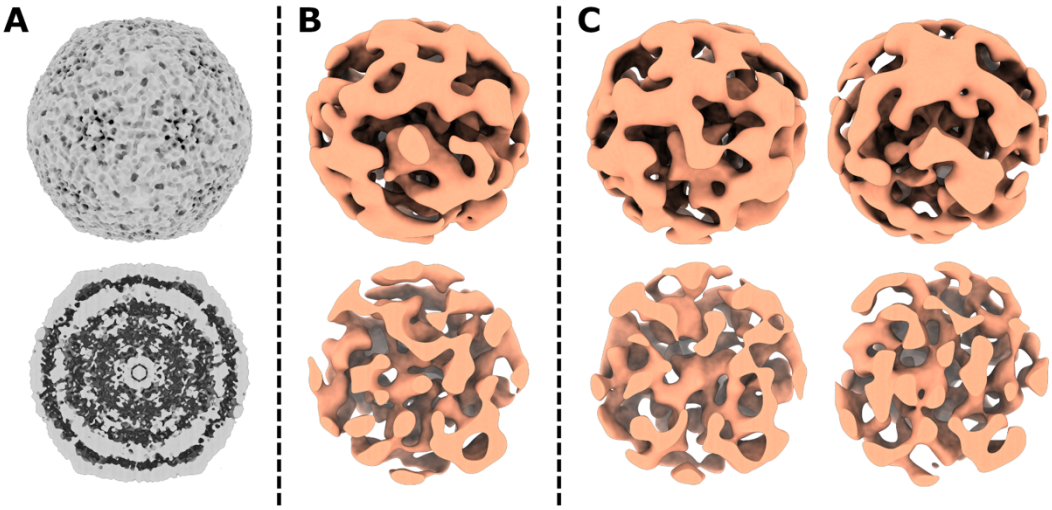
To prepare for the symmetry-relaxed classification itself, particles and a reference map were prepared following the workflow shown in Figure 4.16. The symmetrised reconstruction of wtMNV.BV2 was separated into capsid density and genome density, with the genome density taken forward for use as a reference map (following application of a low-pass filter as usual). Capsid density was used to generate a mask for subtraction of capsid signal from the particle stack used for the symmetrised reconstruction, generating a ‘subtracted’ particle stack with associated orientational information. The ‘genome-only’ reference map and subtracted particles were then subjected to 3D classification with symmetry relaxed from icosahedral (I1) to asymmetric (C1).



4.16. The symmetry relaxation workflow. *A symmetrised reconstruction of wtMNV.BV2 is separated into capsid density (grey) and genome density (orange). Capsid density is used to generate a mask for subtraction of capsid signal from particle images. Using the genome density as a reference map, signal-subtracted particles (and the orientation information generated from the original symmetrised refinement) are refined or classified in 3D using symmetry relaxation to restrict the angular search range.*

Initial results from symmetry-relaxed 3D classification of particles into a single class showed significant changes from the reference map initially supplied, but was ultimately underwhelming (Figure 4.17A,B). With a low global resolution of ~20 Å there was no possibility of observing detailed nucleic acid density, but there were also no clear continuous striations of density describing the broad organisation of the genome, as may be expected for parallel strands of RNA. Indeed, the density was fragmented and had no clear organisation. A

separate classification was attempted with data split between two classes, such that one class could act as a sink for any poor-quality particles that might be degrading the quality of the reconstruction, but there was no noticeable improvement in map quality or reported resolution, with particles split relatively evenly (57% vs 43%) between the classes (Figure 4.17C).



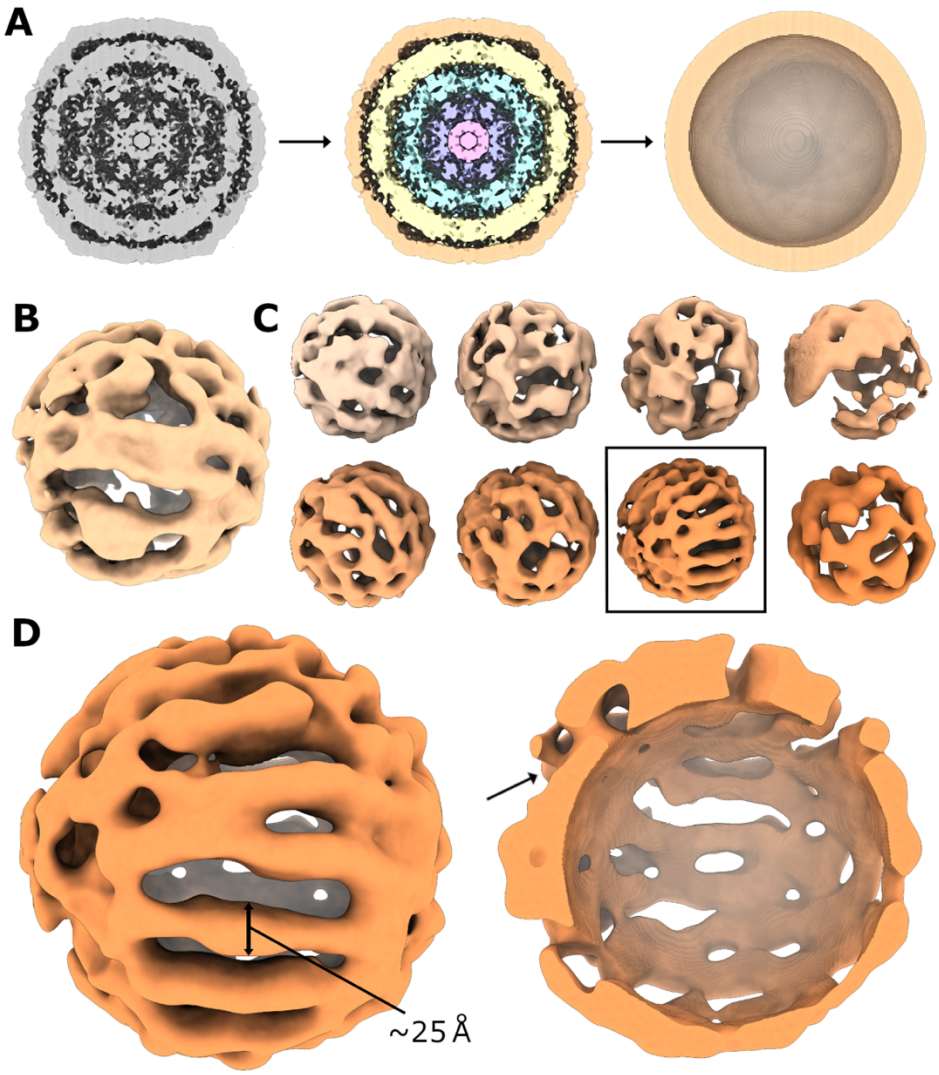
4.17. Initial symmetry-relaxed reconstructions of the MNV genome. (A) *Reference map used for symmetry-relaxed 3D classification. A 60-Å low-pass filter was applied to this map before classification.* ***(B)*** *Initial symmetry-relaxed 3D classification of wtMNV.BV2 genome using a single class.* ***(C)*** *Subsequent symmetry-relaxed 3D classification of wtMNV.BV2 genome using two classes. All maps are shown in full (upper) and as sectional (lower) views.*

The multi-layered radial distribution of density observed in the symmetrised reconstruction of wtMNV.BV2 indicated that the MNV genome may be organised into multiple layers that are not fixed relative to each other. If so, even with symmetry relaxation, it would be practically impossible to align the entire genome together in one classification, potentially explaining the results of initial symmetry-relaxed classifications. To accommodate this, the symmetrised genome-only reference map was used as a guide to design shell-shaped masks that would encompass only a single layer of genome density (Figure 4.18A). Initially, a mask for the outer layer of genome density was

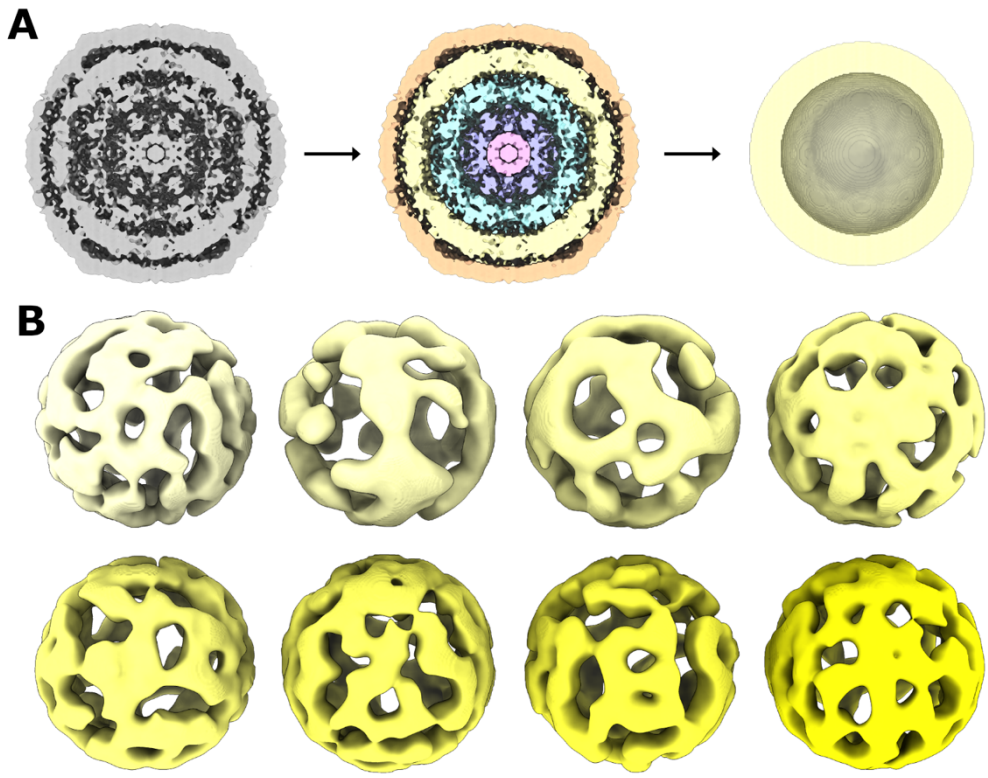
applied, given that this layer was the largest and therefore likely to yield the strongest signal. An initial classification using a single class showed an improvement in the reported resolution (from ~20 Å to ~12 Å) and estimated accuracy of angular assignments (from ~2.58° to ~2.26°). There were subtle indications of possible striated regions of density that were failing to fully separate, but overall the map was not convincingly reminiscent of packaged RNA (Figure 4.18B).

To explore the possibility that the outer layer may exist in a number of discrete positions relative to the capsid (rather than just a single fixed position), a separate classification was performed with eight classes to give sufficient space for separation of any distinct arrangements. Given the increase in class number, the classification was also allowed to proceed for 100 iterations, rather than the usual 25, to allow more time for convergence. Seven of the resultant classes showed no clear organisation, but the most populated class (with ~30% of particles) showed a much more convincing striated pattern (Figure 4.18C). However, each striation was approximately 25 Å in diameter (Figure 4.18D), larger than the expected diameter of ssRNA (~10 Å). Inspection of the density map in sectional view suggested that more than a single layer was being captured by the mask (Figure 4.18D). Further classifications were attempted with a narrower mask to exclude any potential signal from adjacent layers, but these were consistently unsuccessful. Specifically, all classes were consistently off-centre and missing large portions of density, likely because the remaining signal was too weak to keep the alignments within the boundaries of the mask.

In line with the ‘single-layer’ approach, 3D classification into eight classes was next repeated using a mask centred on the second layer of density observed in the symmetrised genome density reference map (Figure 4.19A). Unfortunately, no classes showed any convincing striated pattern like those observed for the previous classification (Figure 4.19B).



4.18. Symmetry-relaxed reconstructions of the outer layer of genome. ***(A)*** *The genome density extracted from the symmetrised reconstruction of wtMNV.BV2 contained distinct layers (different colours, middle panel). A spherical 'shell' mask was created to isolate the outermost layer of density.* ***(B)*** *Symmetry-relaxed 3D classification of wtMNV.BV2 genome using a mask to isolate only the outermost layer of density, specifying a single class.* ***(C)*** *Symmetry-relaxed 3D classification of wtMNV.BV2 genome using a mask to isolate only the outermost layer of density, specifying eight classes.* ***(D)*** *Expanded view of the boxed class from (C), shown in full and in sectional view. Individual strands of density in the striated region had a width of approximately 25 Å, as indicated. Some regions of the reconstruction appeared to include more than one layer of density (indicated by the black arrow).*



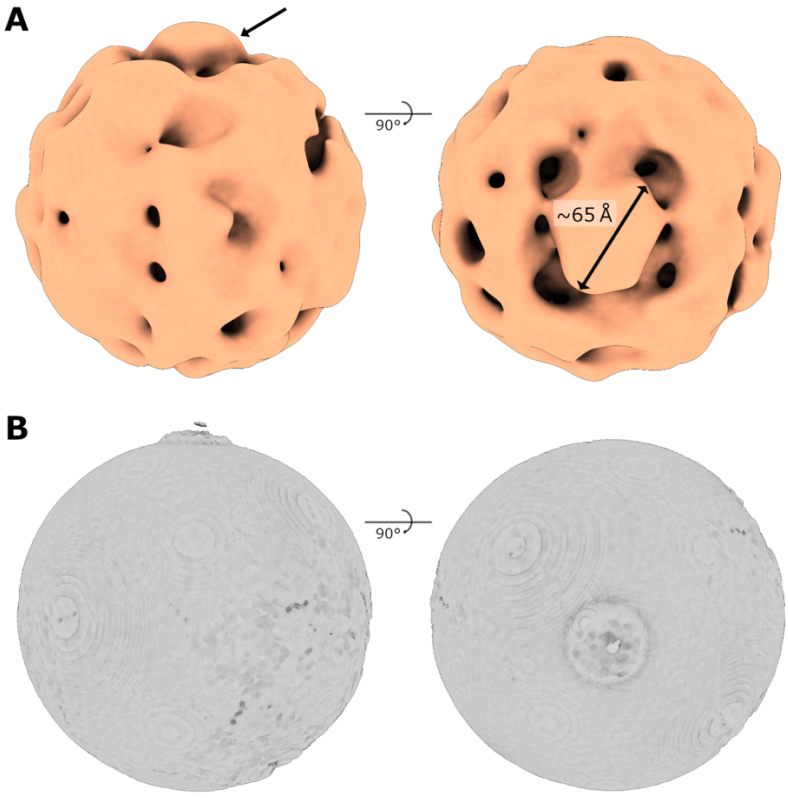
4.19. Symmetry-relaxed reconstructions of an inner layer of genome. ***(A)*** *The genome density extracted from the symmetrised reconstruction of wtMNV.BV2 contained distinct layers of density (different colours, middle panel). A spherical 'shell' mask was created to isolate an internal layer of density (yellow, right-hand panel).* ***(B)*** *Symmetry-relaxed 3D classification of wtMNV.BV2 genome using the mask shown in (A), specifying eight classes.*

4.3.3. Attempts to resolve VP2 and VPg

Aside from the genome, each infectious MNV virion is expected to contain one copy of VPg, tethered to the genome, and an undefined quantity of VP2. As for the genome, these proteins were not resolved in any MNV reconstructions described here or elsewhere to date, likely a result of low signal, symmetry mismatch, intrinsic flexibility/disorder or a combination of these factors.

Initially, reconstructions from the symmetry-relaxed reconstructions of the capsid interior described above were inspected for any signs of VPg or VP2

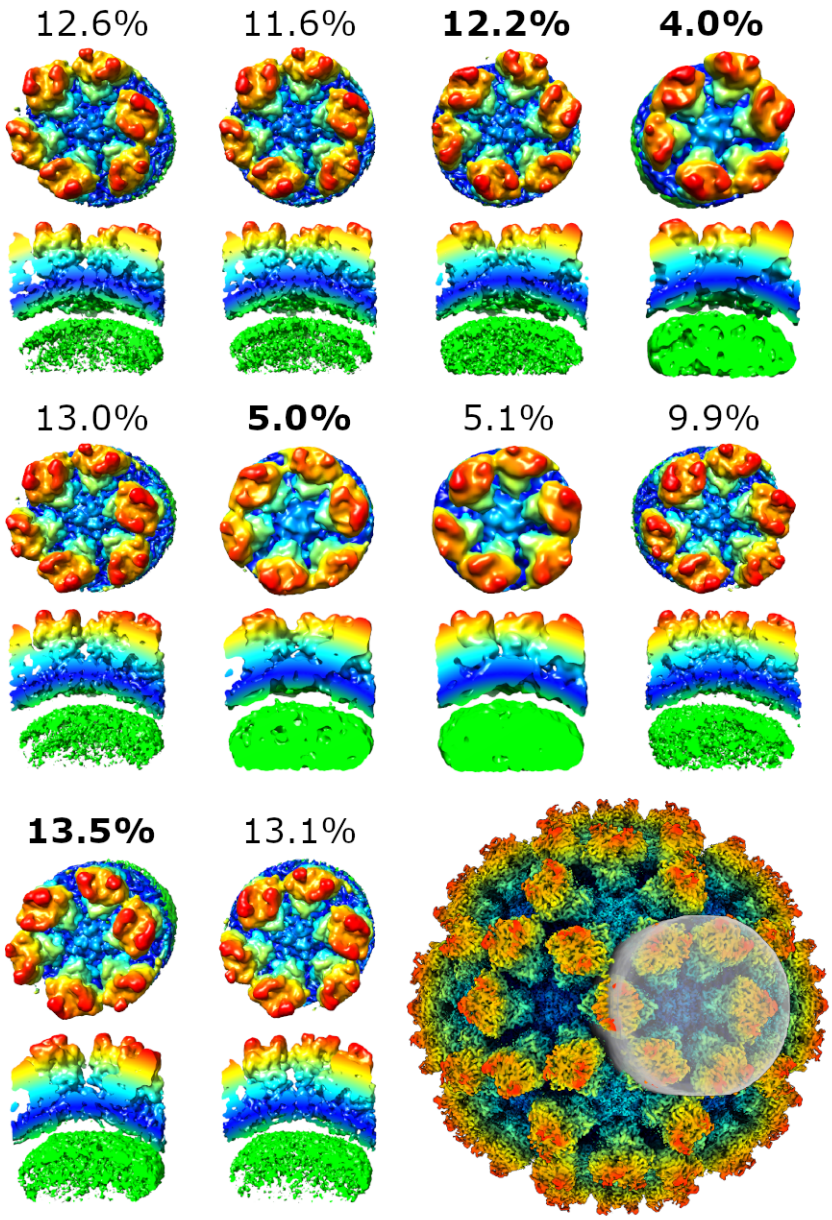
density. Some (though not all) reconstructions contained a discrete ‘bead’ of density approximately 65 Å in diameter, separated from adjacent density and protruding from the reconstruction, as might be expected for VPg or VP2 (Figure 4.20A). However, given the low resolution of the reconstruction (~20 Å), it was difficult to assess whether the size of the ‘bead’ of density was accurate. To check for potential biasing by the reference model, the reference map used for the classifications was inspected for similar density. At high contour levels, no similar density was visible, although at lower contour levels a small protrusion of density was observed at one pole of the sphere (Figure 4.20B). While the reference map was itself extracted from a symmetrised reconstruction of wtMNV.BV2, meaning that the ‘protrusion’ was also present in the symmetrised reconstruction, it was not possible to rule out the possibility that the ‘bead’ of density in the symmetry-relaxed reconstructions was artefactual.



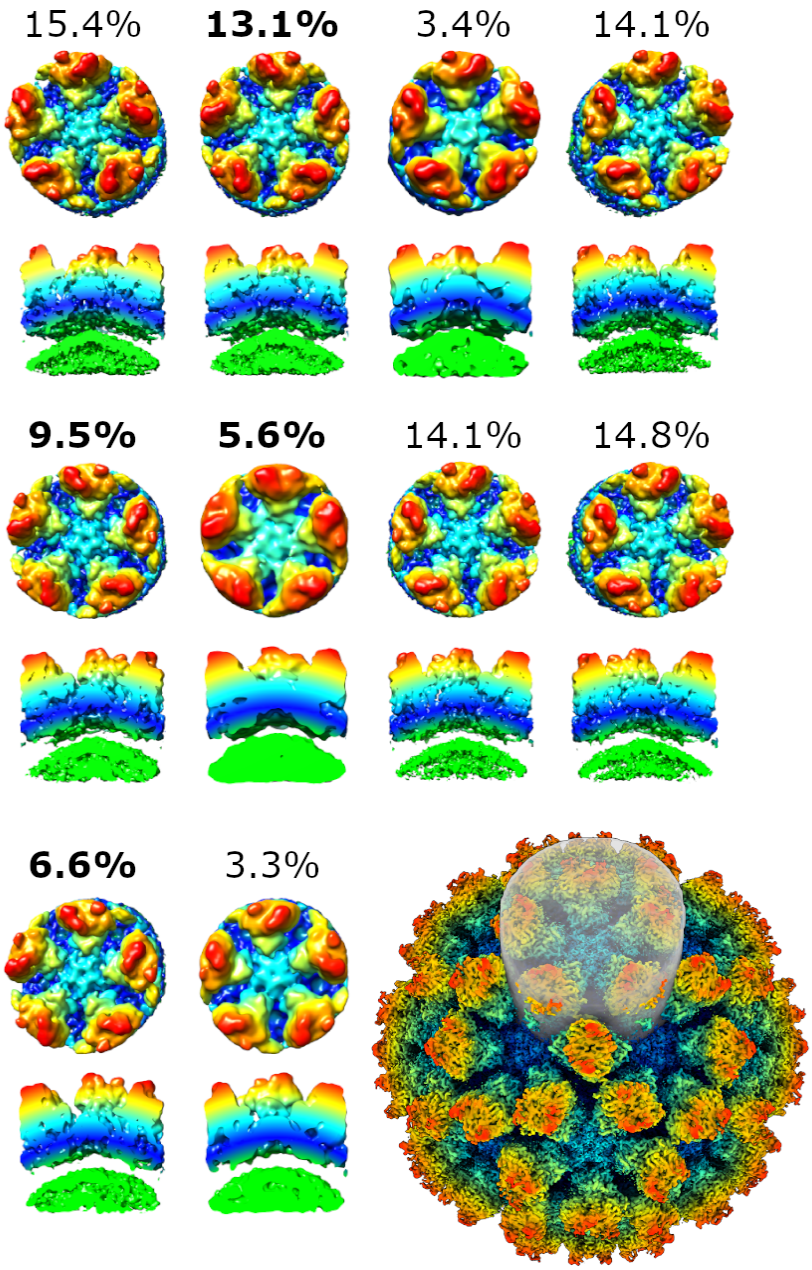
4.20. A discrete bead of density in genome reconstructions. (A) *Initial symmetry-relaxed 3D classification of wtMNV.BV2 genome using a single class (also shown in Figure 4.17B), shown at ~4 σ. A discrete bead of*

density was identified (black arrow). ***(B)*** *The genome-only reference map used for symmetry-relaxed classification (also shown in Figure 4.17A), shown at ~2 σ. A 60-Å low-pass filter was applied before classification.*

As an alternative approach, focussed classification of wtMNV.BV2 was attempted at the icosahedral three-fold axis. A cylindrical mask was created to encompass density for VP1 and the outer layers of the MNV genome, centred on the three-fold axis, and this was used for focussed 3D classification without alignments. While many resultant classes were reasonably well resolved, subparticles were split relatively evenly with the most populated class containing only 13.5% of the data, and no additional density beyond VP1 or genome was observed (Figure 4.21). Interestingly, four out of ten classes adopted the incorrect *Z* orientation. Aside from this, variation between classes was limited. Focussed classification was also performed using a mask centred on the icosahedral five-fold axis, but results were similarly underwhelming with no additional density observed (Figure 4.22). Again, four out of ten classes were assigned the incorrect *Z* orientation.

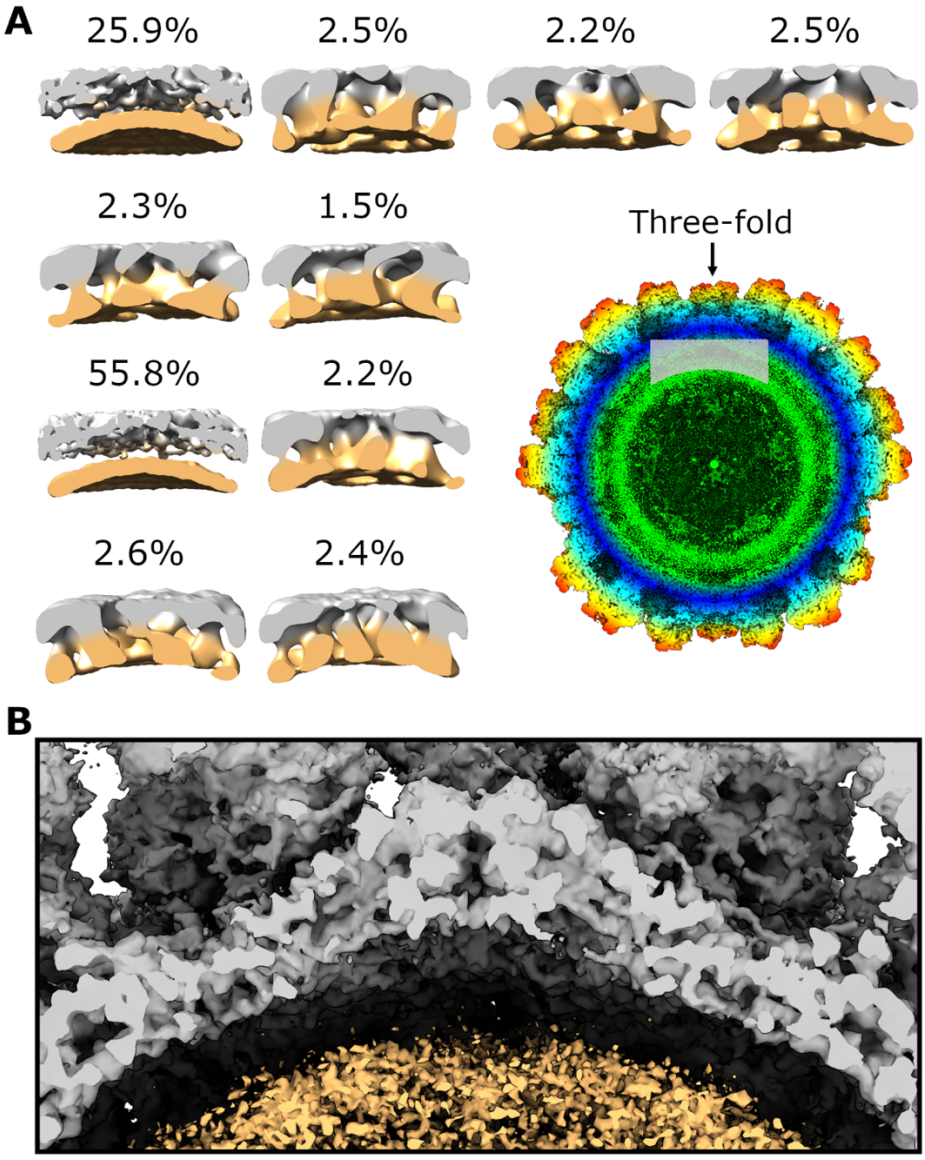


4.21. Focussed classification at the icosahedral 3-fold axis. *All focussed classes from focussed classification of wtMNV.BV2 centred on the icosahedral 3-fold axis, with the proportion of subparticles assigned to each class indicated. Each class is shown viewed down the axis (upper) and as a central section (lower) at ~6 σ. The position of the mask used for focussed classification (grey) is indicated on the reference map. Classes are coloured according to radial distance (relative to the complete wtMNV.BV2 map). Classes with subparticle proportions in bold were assigned the incorrect Z orientation.*



4.22. Focussed classification at the icosahedral 5-fold axis. *All focussed classes from focussed classification of wtMNV.BV2 centred on the icosahedral 5-fold axis, with the proportion of subparticles assigned to each class indicated. Each class is shown viewed down the axis (upper) and as a central section (lower) at ~6 σ. The position of the mask used for focussed classification (grey) is indicated on the reference map. Classes are coloured according to radial distance (relative to the complete wtMNV.BV2 map). Classes with subparticle proportions in bold were assigned the incorrect Z orientation.*

I hypothesised that one reason for the lack of any additional density may be domination of the classification by the strong signal from VP1. Given that so much of the masked volume was occupied by VP1 signal, any subtle differences in weak signal that could correspond to VP2 or VPg may be concealed and therefore not contribute significantly to classification of subparticles. Based on this assumption, focussed classification was repeated at the icosahedral three-fold axis but using a much shorter mask (a cylinder with radius ~65 Å and height ~50 Å, compared to ~65 Å and ~160 Å previously). The mask was centred on the apparent gap between genome density and VP1 shell domain density, such that it contained only a small amount of each. Of the resultant classes, two were heavily populated (>75% of the data) and showed no evidence of additional density (Figure 4.23A). The other eight classes each contained <3% of the data and appeared poorly resolved based on the lack of detail in the density for the VP1 shell domain. While these classes contained poorly defined density that appeared continuous with the shell domain of VP1, this was likely a consequence of the low resolution rather than a meaningful finding. To confirm this and rule out the possibility that the density corresponded to VP2 or VPg, and was only poorly defined because subparticles were spread too thinly between the eight classes, subparticles from each of the eight poorly defined classes were combined and subjected to asymmetric reconstruction, followed by sharpening. Again, no additional density was observed in the focussed region before or after sharpening (Figure 4.23B).



4.23. Focussed classification at the 3-fold axis using a narrow mask. ***(A)*** *All focussed classes from focussed classification of wtMNV.BV2 using a narrower mask centred on the icosahedral 3-fold axis, with the proportion of subparticles assigned to each class indicated. Each class is shown as a central section at ~10 σ, with density corresponding to VP1 coloured grey and non-VP1 density coloured orange. The position of the mask used for focussed classification (grey) is indicated on a section through a reference map. Due to the lack of structural detail, it was not possible to determine whether classes were assigned the correct Z orientation.* ***(B)*** *Subparticles from the eight least populated classes were combined and used for asymmetric reconstruction. The unsharpened map is shown as a central section at ~3 σ.*

4.4. The structure of wtMNV.BV2 in complex with the bile acid LCA

Following the determination of a high-resolution structure of wtMNV.BV2, the same BV-2 cell-derived virus stock was used to explore the structural basis for bile acid binding to the MNV capsid. At the time of the investigation, X-ray crystal structures had been solved for isolated P domain dimers in complex with bile acids (Nelson *et al.*, 2018), but no solution structures for complete capsids in complex with bile acids had been reported. Given the observations of dynamic P domains described here and elsewhere, I hypothesised that bile acid binding may alter the conformational landscape explored by P domain dimers – for example, by stabilising one particular conformation. To investigate this, I determined the structure of wtMNV.BV2 in complex with the bile acid, lithocholic acid (LCA).

4.4.1. Preparation of wtMNV:LCA for structural determination

LCA has previously been shown to bind to P domain dimers (Nelson *et al.*, 2018), so this interaction was attempted in solution with wtMNV.BV2 to give bile acid-bound wtMNV (termed wtMNV:LCA) ready for vitrification. LCA has poor solubility in aqueous solution, reported at just 0.38 µg/mL in water at room temperature (National Center for Biotechnology Information, 2021). Therefore, in order to maximise the amount of LCA present in solution and avoid diluting the wtMNV.BV2 stock, an excess of LCA (40 mg) was first dissolved in 2 mL EtOH and allowed to evaporate at 37°C, leaving a thin coating of LCA on the base of the tube. wtMNV.BV2 (in EM buffer [80 µl]) was then added to the tube and pipetted up and down to dissolve LCA at room temperature.

There were two key limitations with this approach. Firstly, assuming that the solubility of LCA in EM buffer was the same as its reported solubility in water, the maximum concentration that could be achieved in solution would give an approximate molar ratio of 1:2 for LCA to VP1, such that many binding sites would not be occupied. Secondly, because of the small volumes involved, it

was not possible to precisely control for any subtle changes to pH induced by the bile acid. Regardless, the approach was chosen as a proof-of-principle to optimise sample and grid preparation for cryoEM of wtMNV in complex with a bile acid and show that bile acid density could be visualised, with a view to using more soluble, conjugated bile acids (such as GCDCA) after optimisation.

Approximately 30 min after addition of wtMNV.BV2 to the LCA-coated tube, the wtMNV:LCA sample was centrifuged (17,000 × *g*, 3 min, 4°C) to pellet any insoluble material, then applied to glow discharged, ultra-thin carbon-coated lacey carbon grids for vitrification. wtMNV:LCA was applied as either a single 3 µl application or using a multiple applications approach with three separate 3 µl applications. Upon screening, grids prepared with the multiple application approach were all found to have relatively poor ice quality, so a single application grid was taken forward for data collection.

4.4.2. CryoEM data collection, image processing and analysis

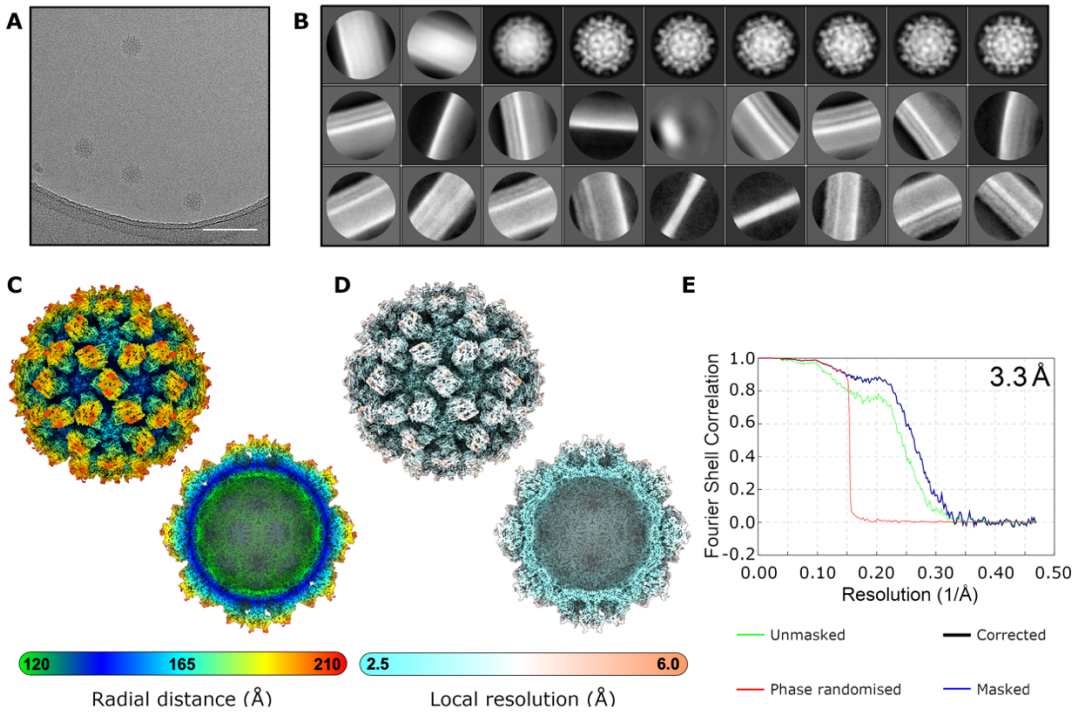
In line with previous data collections for MNV, data for wtMNV:LCA was collected using an FEI Titan Krios microscope operating at 300 kV at a magnification of 75,000×, leading to a calibrated object sampling of 1.065 Å/pixel. Full data collection parameters for wtMNV:LCA are given in Table 4.2.

In total, 10,294 micrographs were collected, from which autopicking identified ~68,000 particles for extraction (Figure 4.24A). Initial 2D classification revealed that many of these particles corresponded to background carbon or contamination. Through further rounds of classification, the particle stack was ultimately reduced to ~4,800 particles used for the final reconstruction (Figure 4.24B). Following CTF refinement and Bayesian polishing, a final reconstruction of 3.3 Å resolution (post-sharpening) was obtained for wtMNV:LCA (Figure 4.24C-E). Notably, despite the small number of particles contributing to the final reconstruction, the density map for wtMNV:LCA also contained detailed P domain density (Figure 4.24C,D), in line with the previous

observation of improved relative resolution for P domains in BV-2 cell-derived MNV (Section 4.2).

Table 4.2. Data collection and processing parameters for wtMNV:LCA.

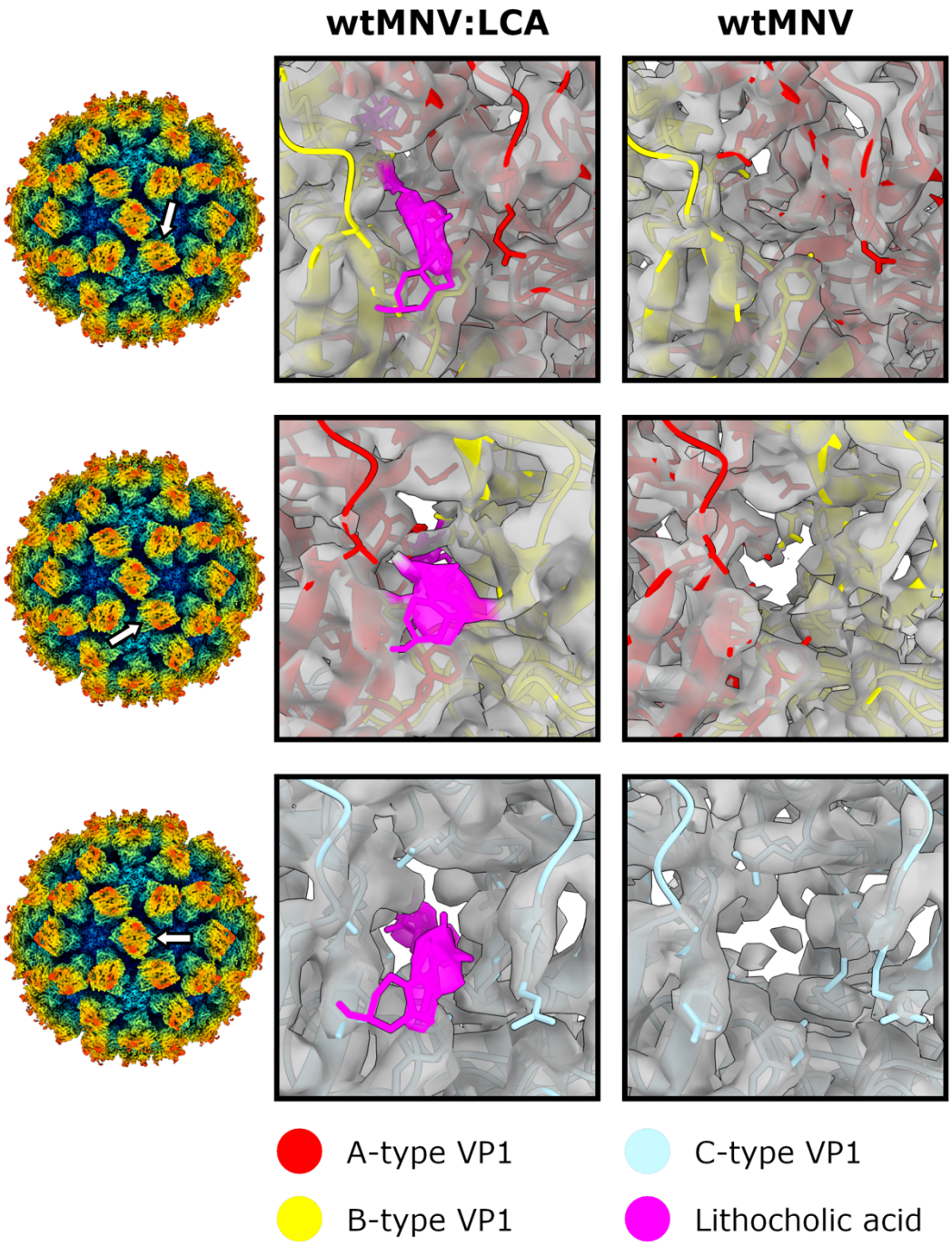
Sample	**wtMNV:LCA**
EMDB ID	*To be deposited*
PDB ID	*To be deposited*
Microscope	FEI Titan Krios
Camera	Falcon III
Voltage (kV)	300
Pixel size (Å)	1.065
Nominal magnification	75,000×
Exposure time (s)	2.0
Total dose (e^-/Å^2)	59.3
Number of fractions	59
Defocus range (μm)	−0.7 to −2.9
Number of micrographs	10,294
Acquisition software	FEI EPU
CryoEM map processing	
Number of particles contributing to map	4,830
Map resolution (FSC = 0.143) (Å)	3.3
Map resolution range around atom positions (Å)	3.1 – 5.4
Map sharpening B factor (Å^2)	−132
Residues modelled	
Quasi-equivalent position A	20 – 531
Quasi-equivalent position B	14 – 531
Quasi-equivalent position C	28 – 531
RMSD	
Bond lengths (Å)	0.0246
Bond angles (°)	1.58
Validation	
All-atom clashscore	33.86
MolProbity score	2.93
Rotamer outliers (%)	2.52
Ramachandran plot	
Favoured (%)	87.43
Allowed (%)	12.11
Outliers (%)	0.46



4.24. wtMNV:LCA cryoEM data processing. ***(A)*** *Representative micrograph from the wtMNV:LCA data collection. Scale bar shows 100 nm.* ***(B)*** *Representative 2D class averages of auto-picked particles from the wtMNV:LCA dataset.* ***(C-D)*** *Density map for wtMNV:LCA filtered according to local resolution and shown at ~2.5 σ, coloured by* ***(C)*** *radial distance or* ***(D)*** *local resolution according to the scales indicated.* ***(E)*** *FSC plot for wtMNV:LCA. Resolution was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).*

In order to check for the presence of bile acid density, the atomic model for wtMNV.BV2 was rigid-body fitted into wtMNV:LCA density and the expected bile acid binding site was inspected at each quasi-equivalent position. Promisingly, extra density was observed in each binding site when comparing wtMNV:LCA with unliganded wtMNV.BV2 at equivalent contour levels, and this density was the correct size for LCA (Figure 4.25). Atomic coordinates for LCA were added to the atomic model for wtMNV.BV2 and this was rigid-body fitted into wtMNV:LCA density, then subjected to one round of real space refinement in Phenix. Subsequent comparison of equivalent Cα atoms

showed that all quasi-conformers had RMSD values <0.5 Å, indicating that there was no evidence for bile acid-induced conformational change.



4.25. Density for LCA is present in the wtMNV:LCA reconstruction. *Expanded views of the bile acid binding pocket from local-resolution filtered maps of wtMNV.BV2 in complex with LCA (left) compared to unliganded wtMNV.BV2 (right). Density is shown at ~2.5 σ, with density corresponding*

to LCA highlighted in magenta. The position of each binding pocket shown is indicated by the arrow on the wtMNV:LCA reconstruction. Atomic coordinates for VP1 are shown, coloured according to the scheme indicated. To generate atomic coordinates for wtMNV:LCA, the coordinates for wtMNV.BV2 and LCA were rigid-body fitted into wtMNV:LCA density and subjected to a single round of real space refinement in Phenix.

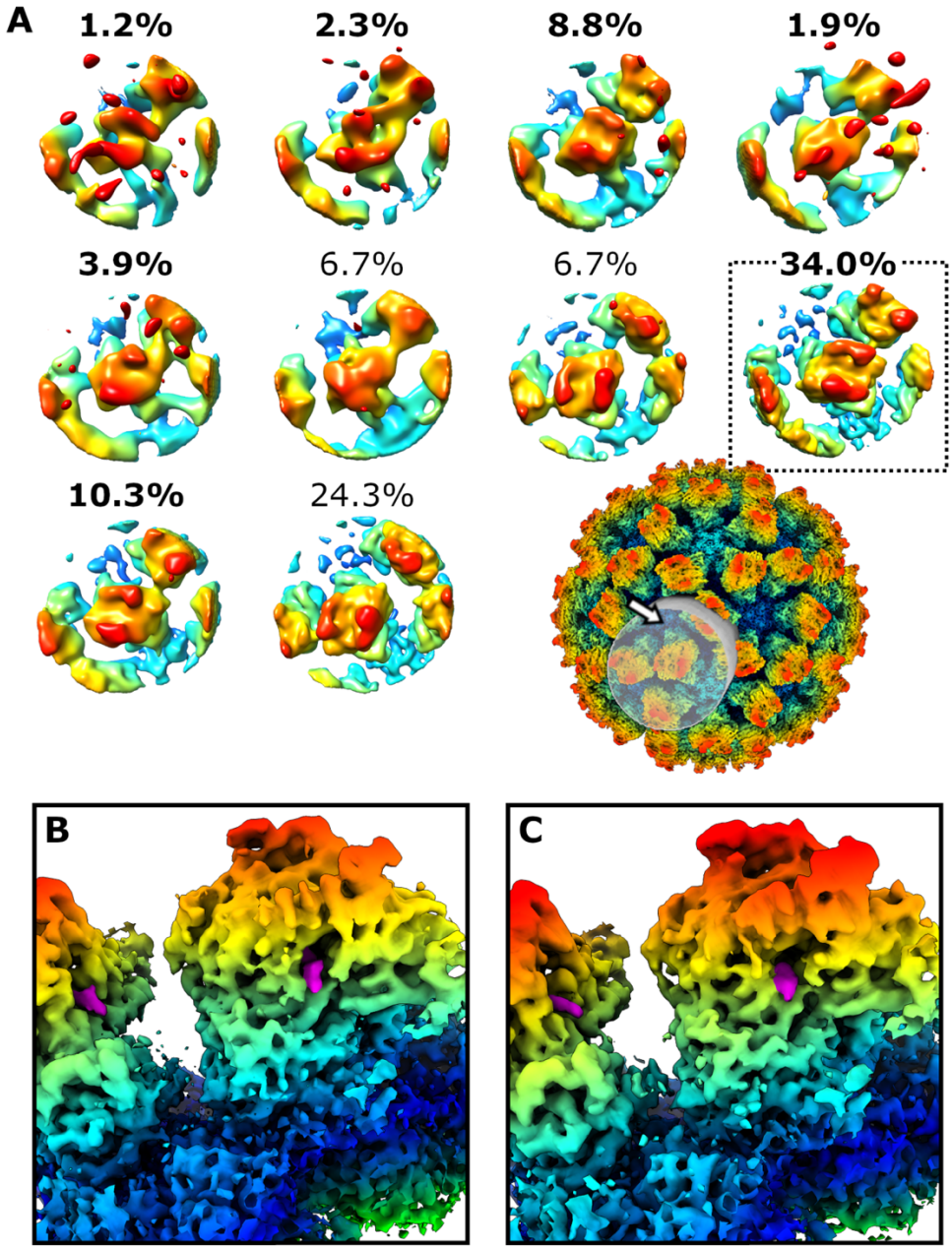
Given the resolution of the map, it was difficult to determine with confidence which residues of VP1 were forming interactions with LCA, but the approximate positioning of LCA density was consistent with an X-ray crystal structure of an isolated P domain dimer in complex with LCA, reported elsewhere (Nelson *et al.*, 2018). Based on the proximity of bile acid density to adjacent residues, LCA could plausibly form interactions with Y250 and Y435 from one monomer of VP1 and with R392 from the other, as was reported previously, in addition to R437, which was previously only reported as an interacting residue for GCDCA (Nelson *et al.*, 2018). This was observed to be the case for all quasi-equivalent binding sites.

4.4.3. Focussed classification at the bile acid-binding pocket

Given the expectation that only a fraction of total binding sites would have been occupied by LCA, a single AB-type P domain dimer was targeted for focussed classification in an attempt to separate LCA-bound sites from unbound sites, and thereby improve LCA density. Resultant classes showed the majority of subparticles were assigned to two classes with the most populated class containing 34.0% of the data (Figure 4.26A). Interestingly, the majority of classes (including the most populated) were assigned the incorrect *Z* orientation. Visual inspection of classes following correction of *Z* orientations indicated good alignment, suggesting minimal variation between classes that had reliable P domain density in line with previous focussed classification results for unliganded wtMNV.BV2.

It was not possible to distinguish whether bile acid density was present or absent from class averages, therefore subparticles from the most populated

class were taken forward for asymmetric reconstruction and sharpening. This showed that, while bile acid density was clearly present, it was no better resolved than was observed for the symmetrised reconstruction (Figure 4.26B,C). The next most populated class was also taken forward for asymmetric reconstruction to see if a class with unoccupied binding sites could be isolated, but bile acid density was similarly present.



4.26. Focussed classification at the bile acid-binding pocket. (A) *All focussed classes from focussed classification of wtMNV:LCA with the proportion of subparticles assigned to each class indicated. Each class is*

coloured according to radial distance (relative to the complete wtMNV:LCA map) and shown at ~15 σ. The position of the mask used for focussed classification (grey) is indicated on reference map. Classes with subparticle proportions in bold were assigned the incorrect Z-orientation. ***(B,C)*** *Enlarged view of VP1 viewed in the direction indicated by the white arrow on the wtMNV:LCA reference map for* ***(B)*** *the symmetrised wtMNV:LCA reconstruction or* ***(C)*** *an asymmetric reconstruction using particles from the most populated focussed class, indicated by the black dashed box in (A). Maps in both (B) and (C) were filtered by local resolution and shown at ~2 σ. Density corresponding to LCA is highlighted in magenta.*

4.5. Discussion and conclusions

In this chapter, I extended structural investigations into MNV beyond analysis of the capsid alone to consider (i) other components of the MNV virion and (ii) interactions of the capsid with a bile acid. While asymmetric refinement, focussed classification and symmetry relaxation failed to resolve VP2 or VPg, promising signs were observed for genomic RNA, indicating that the work described here could serve as a useful 'blueprint' for future studies using an expanded dataset. The structure of MNV in complex with a bile acid, LCA, revealed no conformational differences to unliganded wtMNV.BV2. This is in agreement with studies from other groups reported after the work described in this chapter was completed, which indicate that bile acids stabilise the 'contracted' conformation of MNV, a conformation already adopted by the wtMNV.BV2 described here (Sherman *et al.*, 2019).

With a nominal resolution of 2.7 Å, the wtMNV.BV2 structure was the best resolved to date, but most striking was the level of detail observed within the P domains. This was in contrast to wtMNV.RAW, which had poorly resolved P domains. Compared to wtMNV.RAW, the relative improvement in local resolution was greater for the P domains than the rest of the wtMNV.BV2 map, likely a result of P domain 'rigidification' or greater homogeneity in the positioning of the P domains relative to the S domains, as indicated by focussed classification. Various hypotheses were considered as explanations

for the changes observed in the P domains, including that the difference was related to the extended duration of wtMNV.BV2 storage at 4°C prior to structural investigation, compared to wtMNV.RAW. However, in the absence of any plausible explanation as to why P domains would stabilise (rather than degrade) over time, it was deemed more likely that the change was due to a cell line-related difference, such as a cell line-specific chemical factor binding to VP1, or a cell line-specific post-translational modification to VP1. This could be tested in future by cultivating and purifying fresh MNV from RAW264.7 and BV-2 cells, then determining structures of wtMNV.BV2 immediately after purification and wtMNV.RAW after an extended period of storage at 4°C (i.e., opposite to the wtMNV.RAW and wtMNV.BV2 structures described here). If the difference is cell line specific, then one would expect rigidified P domains only for the structure from wtMNV.BV2, regardless of the duration of storage.

Close inspection of wtMNV.RAW and wtMNV.BV2 density maps revealed additional unoccupied density at residue C143 for wtMNV.RAW, but not wtMNV.BV2, as would be expected for a cell line-specific post-translational modification. However, mutational analysis indicated that this residue had limited importance for viral infectivity. Based on the size of the additional density extending from the thiol group of C143, it is possible to exclude larger modifications from consideration, such as palmitoylation, prenylation and glutathionylation. Rather, nitrosylation or oxidation are more likely candidates, although further investigation would be required to precisely characterise any putative modification. Specifically, mass spectrometry of VP1 from wtMNV.RAW could be compared to that from wtMNV.BV2 to look for a difference in mass corresponding to different post-translational modifications. This could be followed by further tandem mass spectrometry to determine the precise position and identity of any modification(s).

Regardless of the mechanism underlying the change in P domain resolution between wtMNV.RAW and wtMNV.BV2, it was clear that there was a difference, and this was also observed for the BV2 cell-derived wtMNV:LCA structure, which was reconstructed from a much smaller number of particles.

The specific system that is used to cultivate a virus or express a protein for structural studies is sometimes overlooked as a minor experimental detail, with the assumption that this would have no or minimal impact on the final structure. The finding of an apparent difference in P domain rigidity between structures of the same virus from different cell lines, despite being purified in parallel, using identical purification approaches and buffers, highlights the importance of considering differences in expression/culture system during any comparative structural analysis.

Given the improved resolution of the wtMNV.BV2 structure, this was taken forward for analysis of non-VP1 components of the capsid, starting with the viral genome. Only limited information about the radial distribution of genome signal could be derived from the symmetrised reconstruction, with defined layers of density near the interior surface of the capsid and less defined layers closer to the centre of the virion. This may be indicative of less ordering in the centre of the capsid, or may simply be an artefact of the imposition of icosahedral symmetry, given that symmetry axes become much closer together nearer to the centre. Initial asymmetric reconstructions failed to shed any more insight, possibly due to low signal from the genome. When the strong and uniform signal from the ordered and symmetrical capsid was present, this dominated alignments. Even when capsid signal was subtracted from the particles, the reconstruction failed to achieve a resolution better than 20 Å, suggesting genome signal was simply too weak for good alignments.

To address this, symmetry relaxation was used to restrict the search range explored for alignments during 3D classification, guided by the orientational information yielded from a symmetrised reconstruction. Unlike focussed classification, which treats every symmetrically-related copy of a particle as an individual unit for classification and does not perform alignments, symmetry relaxation searches for one optimal alignment for each particle from a limited range of symmetrically related orientations, so is well-suited for resolving an asymmetrical genome within a symmetrical capsid. Initial results were not convincing, but a mask-based ‘layer-by-layer’ approach (which was previously successful for resolving the segmented genome of a dsRNA virus (Ilca *et al.*,

2019)) yielded a class with a striated pattern, reminiscent of parallel strands of nucleic acid as may be expected for a packaged genome. While the diameter of each strand was too big for ssRNA, this may be a consequence of insufficient resolution to separate adjacent strands of ssRNA, such that each striation in the density map represented combined density for two adjacent strands.

While improved relative to non-symmetry relaxed reconstructions, this approach did not succeed in clearly resolving the packaged viral genome. There are both biological and computational explanations that could account in whole or in part for this result, including: (i) insufficient signal as a result of low particle numbers (particularly problematic for asymmetric reconstructions); (ii) imperfect subtraction of capsid signal from particles, introducing more noise into the genome signal and reducing the signal-to-noise ratio; (iii) heterogeneity/poor ordering of the genome within the capsid; (iv) the positioning of the genome not being fixed relative to the capsid (a central assumption of symmetry relaxation). As such, it would be sensible to collect a much larger dataset for future efforts to resolve the MNV genome. If an improvement is seen, this would suggest that signal was a limiting factor, whereas no improvement would give more credence to the idea that biological factors are limiting. Notably, while one low-to-intermediate resolution structure of the bacteriophage MS2 with genomic RNA density used a similar size dataset to that used here (~25,000 particles), suggesting that some unique property of MS2 or its genome made it particularly amenable to genome resolution, successful attempts to resolve the genomic RNA of MS2 and Qβ to high resolution have required much larger datasets (~330,000, ~250,000 and ~86,000 particles) (Koning *et al.*, 2016; Cui *et al.*, 2017; Dai *et al.*, 2017; Chang *et al.*, 2022).

Symmetry relaxed reconstructions of the virion interior were also examined for any signs of VP2 or VPg density, and although many reconstructions contained a discrete ‘bead’ of low-resolution density at one pole of the genome, it was not possible to rule out that this was artefactual. Instead, focussed

classification was performed, initially at the icosahedral three-fold axis. This site was chosen as a previous study reported the formation of a VP2 portal-like assembly at the three-fold axis following receptor engagement in a related calicivirus (Conley *et al.*, 2019). The icosahedral five-fold axis, thought to be a point of nucleation during norovirus capsid assembly (Uetrecht *et al.*, 2011), was also investigated by focussed classification, but neither location showed signs of additional density in any class, and use of a narrower mask to limit signal from VP1 (thereby increasing the relative importance of any VP2/VPg signal) yielded no improvement.

While these computational approaches were unsuccessful in resolving VP2/VPg, it may be possible to visualise VP2/VPg using alternative approaches in future studies. For example, it may be possible to genetically insert a tag sequence into the VPg- or VP2-encoding regions of ORF1 or ORF3, respectively, to increase the size of VPg/VP2 and act as a fiducial marker, generating a stronger signal for alignments. Increasing the size of a target protein, usually through complex formation (e.g., with a Fab fragment), has previously been demonstrated to make structure determination more tractable for small, low-signal targets (Wu *et al.*, 2012; Wu and Rapoport, 2021). However, it is unclear whether a tagging approach would be successful for proteins within the context of a virion. A transposon mutagenesis study showed that MNV VPg only tolerated insertions stably at the C-terminal end, and the insertions investigated were much smaller (~15 nt) than would be required to encode a protein tag large enough to act as a fiducial marker in cryoEM (Thorne *et al.*, 2012). For VP2, ~30% of detected input insertions were maintained after three passages. However, all attempts to generate GFP- or luciferase-tagged MNV at a number of validated insertion sites were reported to be unsuccessful, suggesting that larger insertions may not be tolerated, and even a small FLAG tag insertion into VP2 was unstable, with the insertion lost after two passages (Thorne *et al.*, 2012).

Alternatively, despite being limited to relatively low-resolution information, tomography of MNV virions may provide insight into VP2 and VPg, given that individual virions could be characterised without the need for signal averaging.

Specifically, tomography may help to address the issue of VP2 distribution – whether VP2 is spread across the capsid interior, or is restricted to a single portal-assembly site, determined at the time of particle assembly. If the latter is true, one may speculate that it would be possible to observe a small 'bead' of density representing a cluster of VP2 molecules at a single site within the capsid interior for each virion, which could then be explored further using sub-tomogram averaging.

In addition to non-capsid components of the virion, the interaction of wtMNV with the bile acid LCA was also examined. Despite the low aqueous solubility of LCA and therefore predicted limitations on maximum occupancy of bile acid binding sites, clear density was observed for LCA in the 3.3 Å reconstruction of wtMNV:LCA. Focussed classification did not yield an improvement in the resolution of LCA density, and also failed to produce a class with any unoccupied binding sites (at least for the most populated classes). This is likely because of two factors. Firstly, LCA itself is a small molecule so is only responsible for a small fraction of the total signal. Thus, the presence or absence of LCA signal was unlikely to be a driving factor in assigning subparticles to different classes. Secondly, the masked region contained three binding sites, such that each subparticle would display one of eight unique combinations of LCA binding, making it difficult to generate classes with homogeneous occupancies at each site.

As expected, LCA density at every quasi-equivalent binding site was consistent with a previous X-ray crystal structure of an isolated P domain dimer in complex with LCA (Nelson *et al.*, 2018). No bile acid-induced conformational changes were observed relative to the unliganded wtMNV.BV2 structure, and this is consistent with a more recent cryoEM study that suggests bile acid-binding favours (but is not essential for) the 'contracted' conformation of MNV (Sherman *et al.*, 2019). Through a combination of size exclusion chromatography and NMR experiments, another report showed that GCDCA stabilises MNV P domain dimers, although it should be noted that these experiments were performed with isolated P domain dimers under acidic conditions (Creutznacher *et al.*, 2021a). The relevance of this finding for P

domains in the context of a capsid is supported by cryoEM structures of MNV in complex with GCDCA and TCA (Sherman *et al.*, 2019), and it is plausible that in the work described here, an LCA-induced stabilising effect may have contributed to the relatively high resolution achieved for the reconstruction of wtMNV:LCA, given it was derived from the lowest number of particles of any MNV reconstruction reported in this thesis.

Originally, the wtMNV:LCA structure described here was intended as a proof-of-principle that bile acid binding (and its effects on the capsid) could be demonstrated using cryoEM, with plans to optimise the process for further structures using more soluble, conjugated bile acids, such as GCDCA. However, structures of wtMNV in complex with GCDCA and TCA were subsequently reported (Sherman *et al.*, 2019). Overall, there is convincing evidence that bile acids are one of a number of physiological signals, including pH and metal ions (as discussed in Chapter 3), that interact with the capsid in the host gut to favour a ‘contracted’ and ‘closed’ conformation primed for infection, with increased P domain affinity for CD300lf and inhibition of antibody binding (Sherman *et al.*, 2019; Song *et al.*, 2020; Williams *et al.*, 2021a, 2021b).

5. Structural studies of a novel hepatitis B core antigen-based vaccine platform

5.1. Introduction

In the previous two chapters, I aimed to demonstrate the utility of cryoEM for gaining insight into the structure and function of an infectious virus. Here, cryoEM was used to determine several structures of a novel antigen display technology, in an effort to illustrate the potential of the technique when applied to vaccine design.

5.1.1. Vaccines against viral disease

Vaccines are a key tool in the control of infectious disease and are widely regarded as one of the most impactful public health measures to date. Their widespread use has saved countless lives and led to the successful eradication of smallpox, declared by the World Health Organisation in 1980. Since then, innovations in vaccine technology have made it possible to produce vaccines that are cheaper, safer and more effective, all in a shorter time scale, recently illustrated by the assortment of mRNA and viral vector-based vaccines developed against COVID-19.

Early vaccines against viral pathogens primarily involved attenuating (e.g., through continuous passage in cell culture) or inactivating (chemically or thermally) the virus, although a variety of vaccine technologies are now licensed for use, as summarised in Table 5.1. Each technology has specific advantages and disadvantages associated with its use. While live attenuated vaccines closely mimic natural infection, therefore often induce robust, multi-faceted immunity, they tend to cause more severe side effects and require a cold chain during distribution. There is also a risk of reversion to a virulent strain and of vaccinated people shedding infectious virus, currently a key problem for polio eradication efforts (https://polioeradication.org/polio-today/polio-now/). Inactivated vaccines eliminate the risk of reversion, but tend to be less immunogenic and often require booster doses. Furthermore, large

quantities of live virus must be produced prior to inactivation, leading to the possibility of an accidental release of infectious virus into the environment.

Table 5.1. Vaccine technologies in vaccines licensed for viral disease.

Technology	Examples
Live attenuated virus	Polio (Sabin), MMR, yellow fever
Inactivated virus	Polio (Salk), rabies, influenza
Recombinant protein/virus-like particle (VLP)	Human papilloma virus, hepatitis B virus
RNA-based	COVID-19
Viral vector	Ebola, COVID-19

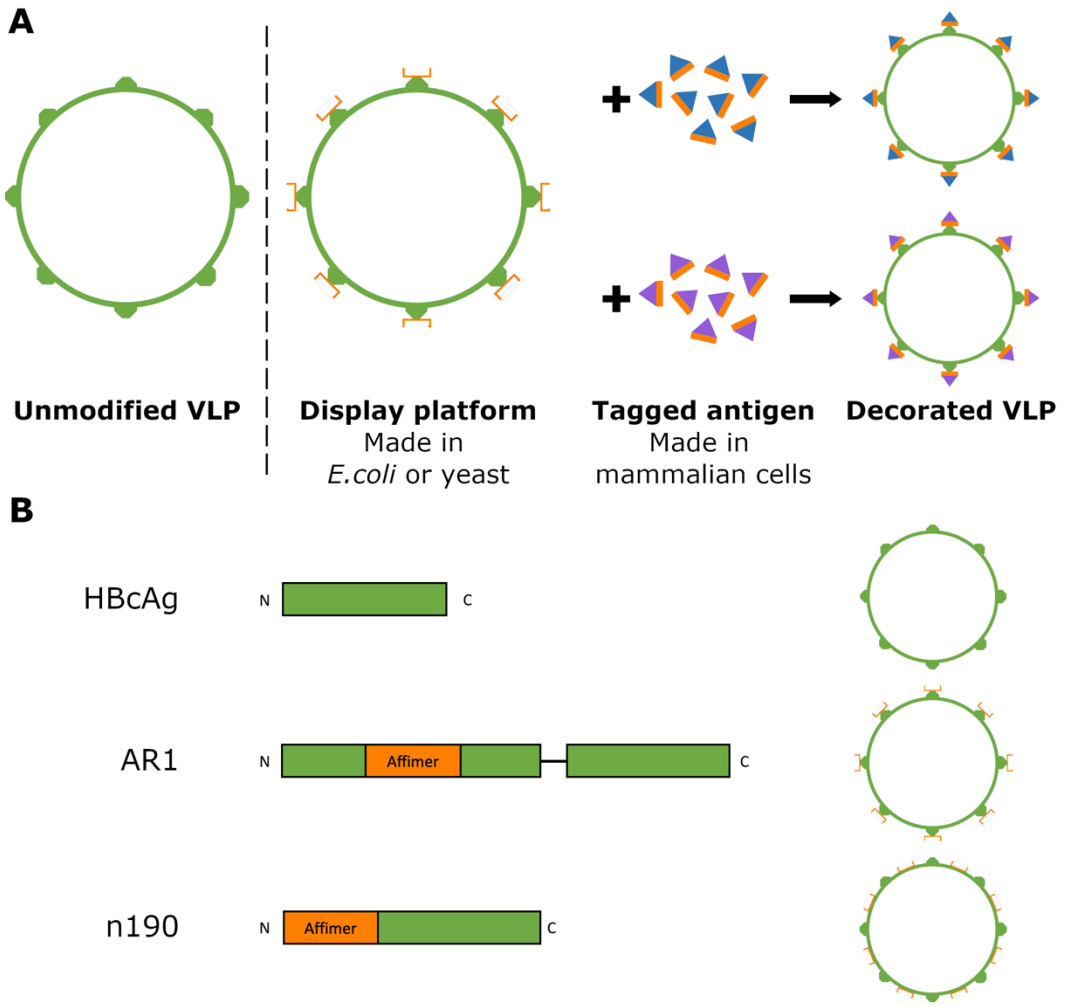
Virus-like particle (VLP)-based vaccines constitute macromolecular assemblies that resemble viruses, but contain no viral genetic material so are non-infectious. This technology resolves many of the issues associated with more traditional approaches: there is no risk of reversion or environmental contamination with infectious virus, and they tend to be more immunogenic than inactivated vaccines. The simplest VLP vaccines use VLPs assembled from the viral proteins against which an immune response is desired (reviewed in (Noad and Roy, 2003)), although it is also possible to decorate VLPs with antigens from other viruses (or even non-viral pathogens) that would not intrinsically organise into VLPs. Presenting antigens on a VLP surface elicits an improved immune response compared to immunisation with free soluble antigen, often involving both cellular and humoral immunity (Francis *et al.*, 1990; Chackerian *et al.*, 2001; Guo *et al.*, 2019; Marini *et al.*, 2019). This is likely because VLPs are approximately the same size as infectious virus and are essentially repetitive and highly-organised arrays of antigen. As such, they have an enhanced ability to cross-link B cell receptors and are more prone to internalisation by antigen presenting cells (Bachmann and Jennings, 2010). Indeed, a previous immunisation study using antigen-conjugated nanoparticles of different sizes showed that particulate immunogens around 20 – 60 nm in diameter appeared optimal for uptake by dendritic cells in mice (Fifis *et al.*, 2004).

VLPs are typically 'decorated' by genetically inserting the target antigen sequence into a scaffold protein construct (e.g., a viral structural protein), however, this raises a number of practical issues. Firstly, it is difficult to predict the effect of such genetic modification on the solubility of the recombinant protein for any given expression system, and based on unpublished empirical data acquired by the Stonehouse/Rowlands groups, chimeric VLP-antigen constructs seem to be particularly sensitive to this issue. This represents a significant hurdle to standardisation of the approach (e.g., such that decorated VLP vaccines could be updated regularly to account for changes in the predominant circulating virus strains). Each new genetic construct would need to be thoroughly tested for expression/solubility and modified if found to be insoluble. Secondly, many of the target antigens that would benefit from this approach are viral glycoproteins, which require production in expensive mammalian expression systems to achieve an appropriate glycosylation pattern. Linking the glycoprotein to a VLP scaffold protein increases the resources consumed to generate a given amount of antigen, and therefore raises production costs considerably.

5.1.2. ‘VelcroVax’ - a novel antigen presentation platform

To address these problems, the Stonehouse and Rowlands groups have developed a 'plug-and-play' VLP-based antigen presentation platform, termed 'VelcroVax'. Rather than genetically inserting the target antigen into the VLP protein, the VLP protein has been genetically modified to display a capture system that can form a non-covalent interaction with a small, defined recognition 'tag' (Figure 5.1A). Any desired target antigen can be modified to contain this tag, such that the VLP scaffold can capture the target antigen for display just by mixing the two species in solution. Since the scaffold remains unchanged regardless of the target antigen, this system is much less prone to the solubility issues associated with continuous redesign of the VLP construct. Furthermore, the scaffold (which does not require a mammalian pattern of glycosylation) can be produced separately from the target antigen in a cheaper bacterial or yeast expression system, reducing production costs.

The 'VelcroVax' scaffold is based on the core antigen (HBcAg) from hepatitis B virus (HBV), which naturally assembles into VLPs when expressed recombinantly (Cohen and Richmond, 1982). The protein is primarily α-helical in nature, with dimerisation leading to the formation of four-helix bundle ‘spikes’ that straddle the dimer interface (Crowther *et al.*, 1994; Böttcher *et al.*, 1997; Conway *et al.*, 1998). HBcAg dimers then assemble into both $T = 3$ and $T = 4$ VLPs. HBcAg is particularly well suited for use as a scaffold because (i) it is highly immunogenic (Milich and Mclachlan, 1986; Guo *et al.*, 2019), and (ii) it contains multiple sites amenable to genetic insertion of a capture system (Schödel *et al.*, 1992). One of these sites is the major immunodominant region (MIR), located at the tip of each four-helix bundle. Previous studies have shown that foreign antigenic sequences inserted at the MIR induce strong immune responses (Brown *et al.*, 1991). Aside from the MIR, antigen connected via a linker to the N-terminus of HBcAg is also surface exposed in the context of the VLP, so this is another site suitable for insertion of a capture system (Francis *et al.*, 1990; Schödel *et al.*, 1992). The C-terminus contains an arginine-rich domain (ARD), and is thought to be internalised in the HBV nucleocapsid, facilitating the recruitment of the viral genome (Birnbaum and Nassal, 1990; Nassal, 1992; Patel *et al.*, 2017). However, if the C-terminus was truncated, foreign antigenic sequences inserted at this site were found to be surface exposed and not to disrupt VLP formation, although immunogenicity was sub-optimal (Schödel *et al.*, 1992). Truncations to the C-terminus have also been observed to shift the proportion of $T = 3$ to $T = 4$ particles, suggesting an important role for the C-terminal region in particle assembly (Zlotnick *et al.*, 1996).



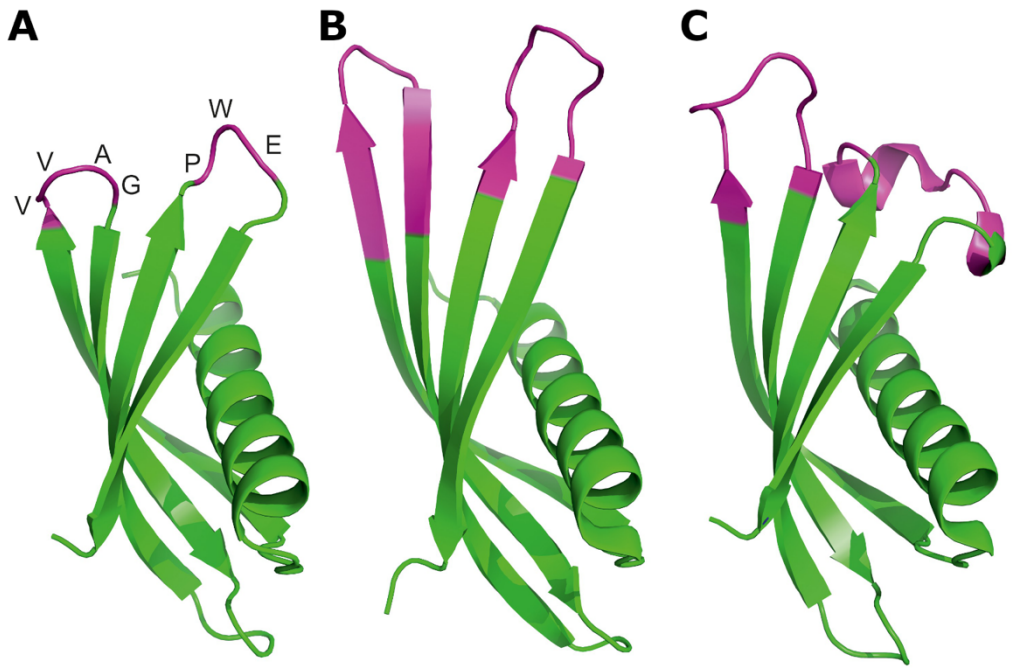
5.1. The VelcroVax concept and constructs under investigation. (A) *Schematic showing the VelcroVax concept. VLPs are modified to include a capture mechanism (such as an Affimer [orange]), generating a display platform. Target antigens, tagged with an epitope recognised by the capture system, are mixed with the display platform, generating decorated VLPs.* ***(B)*** *Schematic showing constructs under investigation. Compared with unmodified HBcAg (top), AR1 and n190 have anti-SUMO Affimers genetically inserted into the HBcAg backbone. AR1 comprises two HBcAg monomers, one with an Affimer inserted at the MIR and one with an unmodified MIR, linked by a flexible linker sequence. For n190, the Affimer is linked to the N-terminus of HBcAg.*

5.1.3. 'VelcroVax' constructs: AR1 and n190

In line with these alternative insertion sites, a number of different 'VelcroVax' constructs were developed by the Stonehouse and Rowlands groups, of which two were selected for structural examination and are described here (Figure 5.1B). The first construct was designed to include the capture system at the MIR, and was termed 'AR1'. Since the MIRs of each HBcAg are located in close proximity at the tip of the four-helix bundle in each dimer, AR1 was designed as a 'tandem core' construct, i.e. two HBcAg monomers were genetically fused and only the first monomer was modified to include a capture system (Figure 5.1B). This ensured that each four-helix bundle only contained a single capture system, avoiding steric clashes that can disrupt VLP formation (Peyret *et al.*, 2015). The second construct, termed 'n190', had a capture system inserted at the N-terminus of HBcAg. Given the absence of steric clash issues associated with insertion at the MIR, the construct was not converted into a tandem core, meaning that one capture system would be displayed on each monomer. The HBcAg component of n190 itself was largely unmodified, although a single cysteine residue (equivalent to C61 in unmodified HBcAg) was mutated to an alanine in order to abrogate a disulphide bridge at the intradimer interface. This interaction was previously shown to have a destabilising effect on the HBcAg capsid; specifically, VLPs formed from a greater proportion of oxidised HBcAg (i.e., HBcAg containing the disulphide bridge) assembled more slowly and were more sensitive to urea treatment (Selzer *et al.*, 2014).

For both AR1 and n190, the capture system chosen for antigen display was an Affimer against a small ubiquitin-like modifier (SUMO) tag. Affimers are small (~12-14 kDa) binding proteins that can be used as molecular recognition tools in a manner analogous to antibodies, and have been used across a number of different applications (Tiede *et al.*, 2014, 2017). Structurally, they comprise a highly stable core formed from an α-helix and a four-stranded anti-parallel β-sheet, with two variable loops between β-strands forming the binding interface against a target protein (Figure 5.2) (Tiede *et al.*, 2014). By inserting the sequence of an Affimer selected for binding to SUMO (here termed a 'SUMO Affimer') into the MIR or N-terminus of HBcAg, AR1 and n190 VLPs

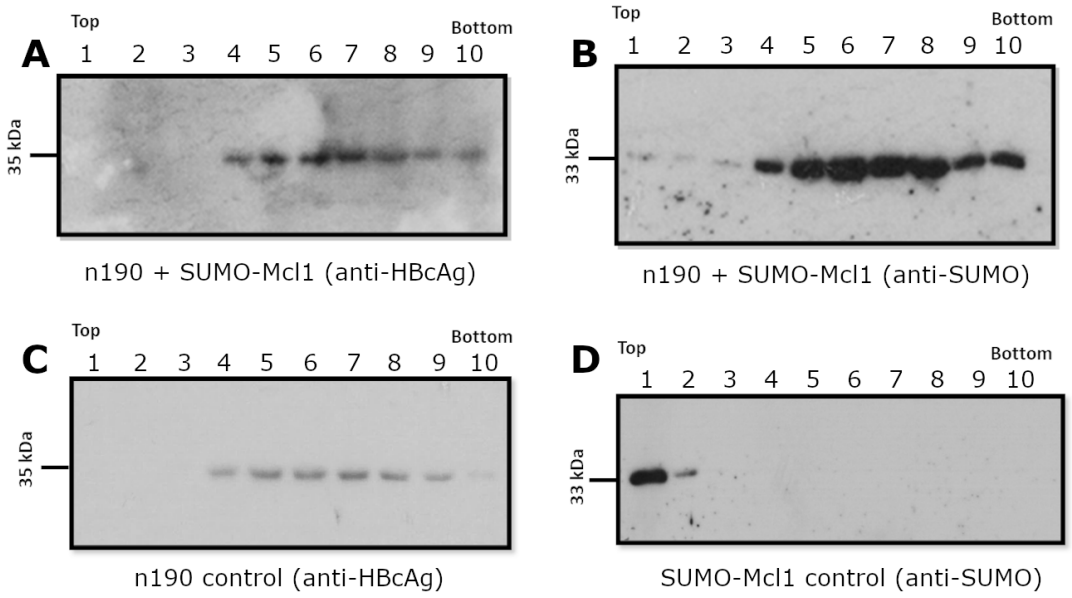
are, in principle, able to capture any protein that has been modified to contain a SUMO tag.



5.2. X-ray crystal structures of Affimer molecules. *Three Affimer molecules, showing the core α-helix and β-sheet (green) and variable binding loops (pink).* ***(A)*** *Affimer scaffold (PDB: 4N6T).* ***(B)*** *Affimer against p300 (PDB: 5A0O).* ***(C)*** *Affimer against human SUMO proteins (PDB: 5ELJ). Figure adapted from (Tiede et al., 2017), published under a CC BY 4.0 licence.*

Using AR1 expressed in the yeast strain *Pichia pastoris,* and n190 expressed in *E. coli*, Jehad Alzahrani (JA) and Kaniz Fatema (KF) generated a range of biochemical data showing empirically that these constructs are able to capture a range of SUMO-tagged antigens. For example, as proof of principle, SUMO-tagged induced myeloid leukaemia cell differentiation protein (Mcl-1) co-sedimented with n190 in a sucrose density gradient when the two were mixed prior to ultracentrifugation (Figure 5.3A,B). In contrast, SUMO-Mcl1 remained at the top of the gradient in the absence of n190 (Figure 5.3C,D). Other experiments performed by JA and KF with both n190 and AR1, using a variety of target antigens and across a range of techniques (including ELISA and

isothermal titration calorimetry (ITC)), have generated similar results (data not shown).



5.3. n190 is able to interact with SUMO-tagged antigens. *n190 and SUMO-tagged Mcl-1 (SUMO-Mcl1) were separated by sucrose gradient ultracentrifugation either alone or following mixing. Sucrose gradients were fractionated, then the fractions analysed by western blot with either a mouse monoclonal antibody against HBcAg (10E11, anti-HBc), or polyclonal rabbit antibodies against the SUMO tag (anti-SUMO).* ***(A-B)*** *Western blots performed following sucrose gradient ultracentrifugation of mixed n190 and SUMO-Mcl1, blotted with (A) anti-HBc or (B) anti-SUMO to show the locations of n190 and SUMO-Mcl1, respectively.* ***(C)*** *Western blot performed following sucrose gradient ultracentrifugation of n190 alone, blotted with anti-HBc.* ***(D)*** *Western blot performed following sucrose gradient ultracentrifugation of SUMO-Mcl1 alone, blotted with anti-SUMO. Experiments performed and figure generated by KF.*

5.1.4. The value of structural data in the development of 'VelcroVax'

While biochemical and immunological data are key to vaccine development, structural analysis of vaccine candidates can provide valuable additional insight. In the case of the 'VelcroVax' system, high-resolution structural data

would permit detailed visualisation of n190 and AR1 VLPs and their capture systems, to examine whether or not the SUMO Affimers are optimally positioned to maximise accessibility to target antigens. Structural analysis would also provide a means by which to confirm whether AR1 and n190 both produce $T = 3$ and $T = 4$ VLPs, as is the case for unmodified HBcAg, and whether there is any change in the proportion of these different arrangements. The effect of VLP size on immunogenicity is not clear, so having a robust method for validating the $T = 3/T = 4$ composition of n190 and AR1 preparations would be particularly useful for future immunogenicity trials aimed at addressing this question. Aside from analysis of unliganded 'VelcroVax' VLPs, structural data for a target antigen-bound VLP would provide useful validation of capture system functionality, and may yield insight into Affimer occupancy and flexibility. Furthermore, antigen-bound VLP structures would be useful in beginning to define and explain the limitations of the system, for example, in estimating what antigens may not be suitable for 'VelcroVax' presentation (based on antigen size or other factors).

Ultimately, a library of 'VelcroVax' structures would be a useful resource for explaining biochemical and immunological results, and for further hypothesis generation. These structures could be fed back into the design of new and improved 'VelcroVax' constructs, maximising the likelihood of success for the system as a vaccine platform technology.

5.1.5. Project aims

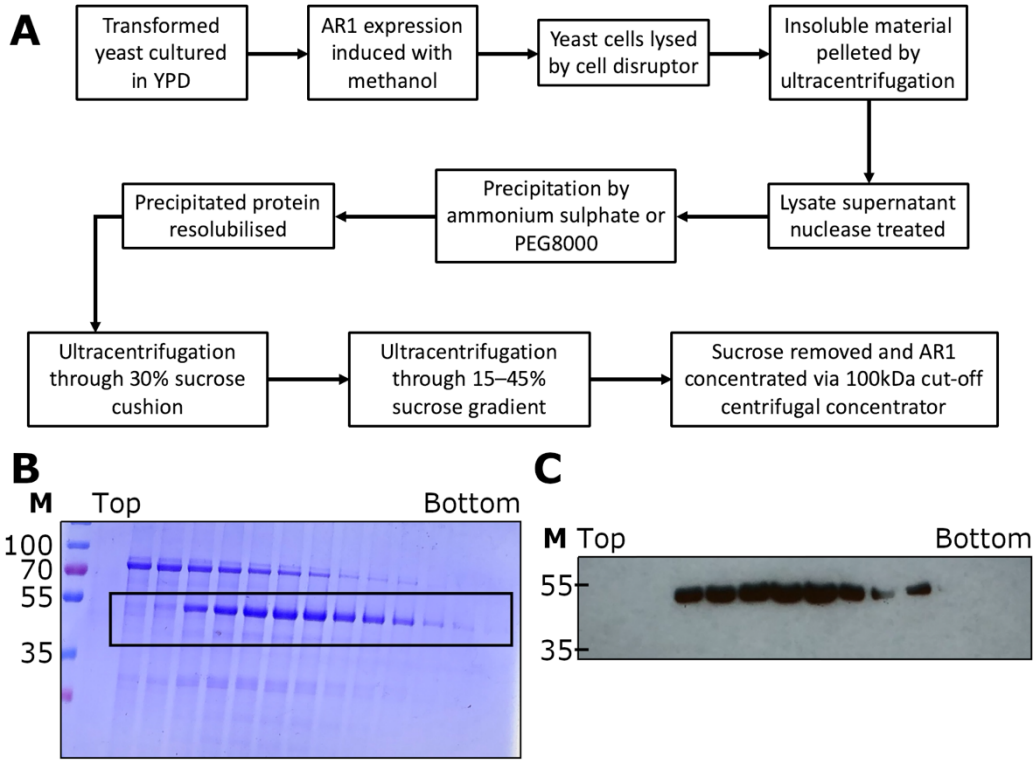
In this chapter, I aimed to do the following:

- determine the structure of unliganded AR1;
- determine the structure of unliganded n190 and compare this with AR1;
- determine the structure of a 'VelcroVax' construct in complex with a target antigen and compare this with the unliganded structures;
- use this structural information to inform the design of alternative 'VelcroVax' constructs.

5.2. The structure of unliganded AR1

5.2.1. AR1 expression and purification

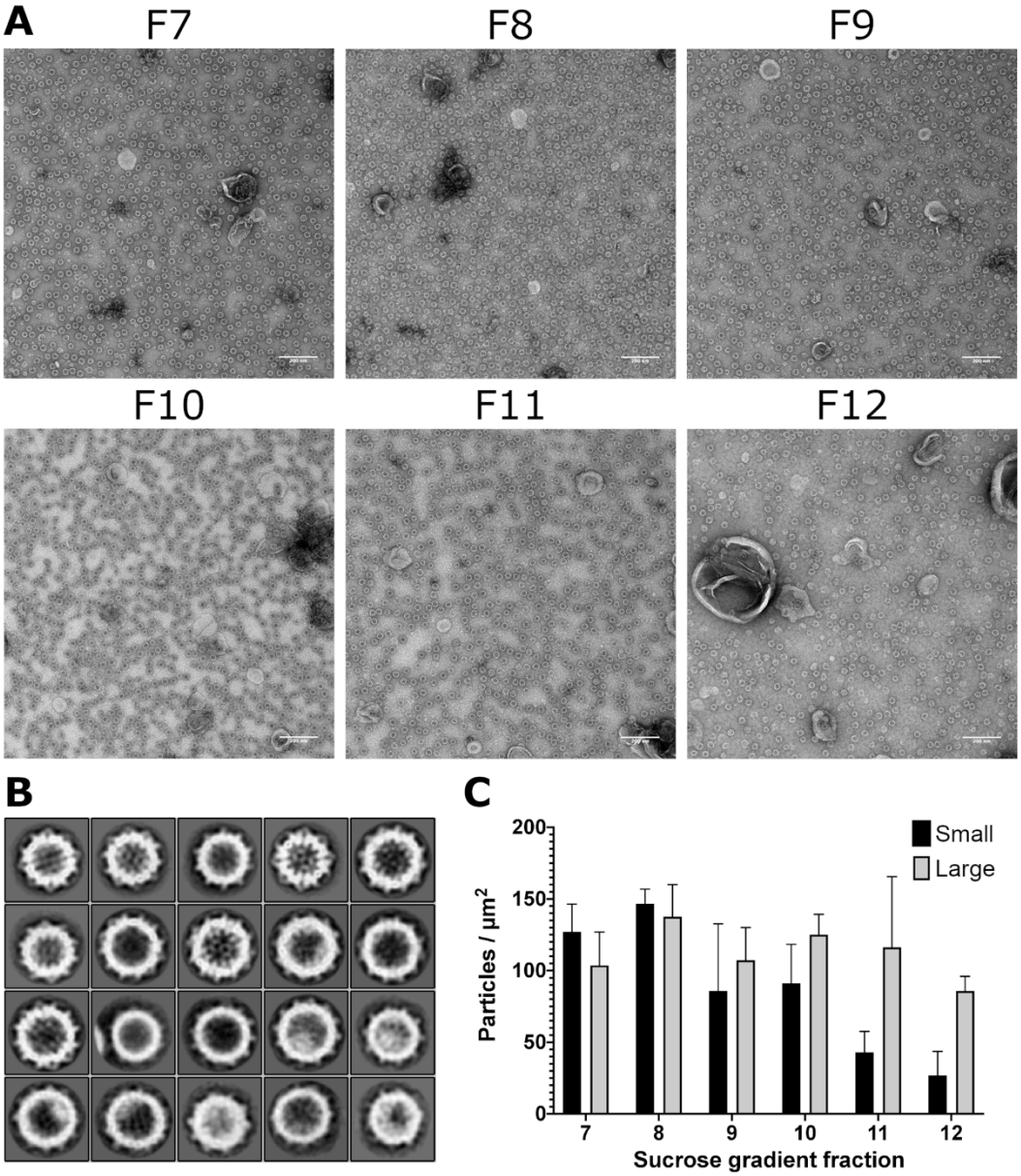
AR1 was expressed in *Pichia pastoris* and purified by JA (Snowden *et al.*, 2021). Yeast were cultured in yeast extract-peptone-dextrose (YPD) and expression of AR1 induced with 0.5% (v/v) methanol. Following that, yeast cells were lysed using a cell disruptor and AR1 was purified via ammonium sulphate/PEG8000 precipitation, followed by resolubilisation and sucrose density gradient ultracentrifugation (Figure 5.4).



5.4. Expression and purification of AR1. (A) *Simplified flow diagram for expression and purification of AR1 in P. pastoris.* ***(B)*** *SDS-PAGE analysis of fractions from sucrose gradient purification of AR1. The contents of fractions 5 (top) – 18 (bottom) were separated by gel electrophoresis and the gel stained with Coomassie blue. Bands corresponding to the expected molecular weight of AR1 are indicated with the black box.* ***(C)*** *Western blot of the same sucrose gradient fractions shown in (B), using a mouse monoclonal antibody against HBcAg. M – molecular weight markers (kDa).*

Expression and purification of AR1 and subsequent analysis was performed by JA. Figure panels (B) and (C) were provided by JA.

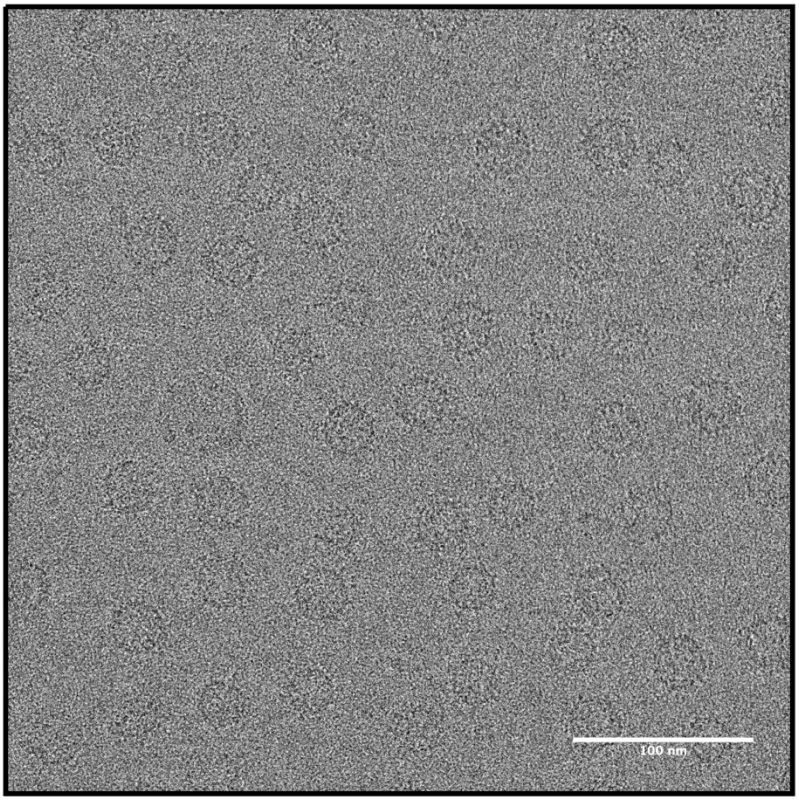
To gain further insight into the quality of AR1 and select an appropriate gradient fraction for dialysis and structural work, negative stain EM was performed for several of the fractions identified as containing AR1 (fractions 7 – 12). All fractions contained AR1 particles, with greater concentrations in fractions 7 and 8 (Figure 5.5A). 2D classification of ~7,800 particles manually picked from fraction 7 micrographs yielded relatively high-quality classes with the expected morphology and HBcAg 'spikes' clearly visible (Figure 5.5B). It was also clear that two different sizes of particle were present, likely corresponding to the $T = 3$ and $T = 4$ configurations naturally formed by HBcAg. While these different forms are indistinguishable by western blot/denaturing SDS-PAGE analysis, counts of the relative amounts of small (putative $T = 3$) and large (putative $T = 4$) particles in micrographs from each fraction suggested that each particle type migrates somewhat differently during sucrose gradient centrifugation (Figure 5.5C). Small particles showed a clear peak in shallower fractions (7-8), while the distribution of large particles was flatter. Thus, there was an almost 50:50 split between small and large particles in fractions 7 – 8, but large particles clearly outnumbered small particles in fraction 12, further down the gradient. Ultimately, fraction 7 was chosen for further structural investigation of AR1.



5.5. Negative stain EM analysis of fractionated AR1. (A) *Representative negative stain EM micrographs of AR1 sucrose gradient fractions (7 – 12). Scale bars show 200 nm.* ***(B)*** *The 20 most populated classes from 2D classification of ~7,800 particles picked from negative stain EM micrographs of AR1 (fraction 7).* ***(C)*** *Individual small (putative T = 3) and large (putative T = 4) particles in five 800x800 pixel spaces at the corners and centre of each micrograph shown in (A) were counted and used to calculate particle 'density' for each fraction. Error bars show standard deviation for particle counts from the five separate regions inspected (n = 5 ± SD).*

5.2.2. AR1 sample preparation, cryoEM data collection and processing

AR1 was vitrified for cryoEM using continuous ultra-thin carbon-coated lacey grids. Given the high concentration of AR1 particles in the sample, only a single 3-µl application was applied to each grid before blotting (for between 1 – 4 seconds) and plunge freezing. Screening confirmed that the grids had acceptable ice thickness, allowing identification of AR1 particles even at low applied defocus values (Figure 5.6), so the grid was taken forward for a 16-hour data collection.



5.6. Screening of AR1 grids for cryoEM. *Representative micrograph taken during screening at a magnification used for data collection (75,000×) with a low defocus applied (–0.5 µm). Scale bar shows 100 nm.*

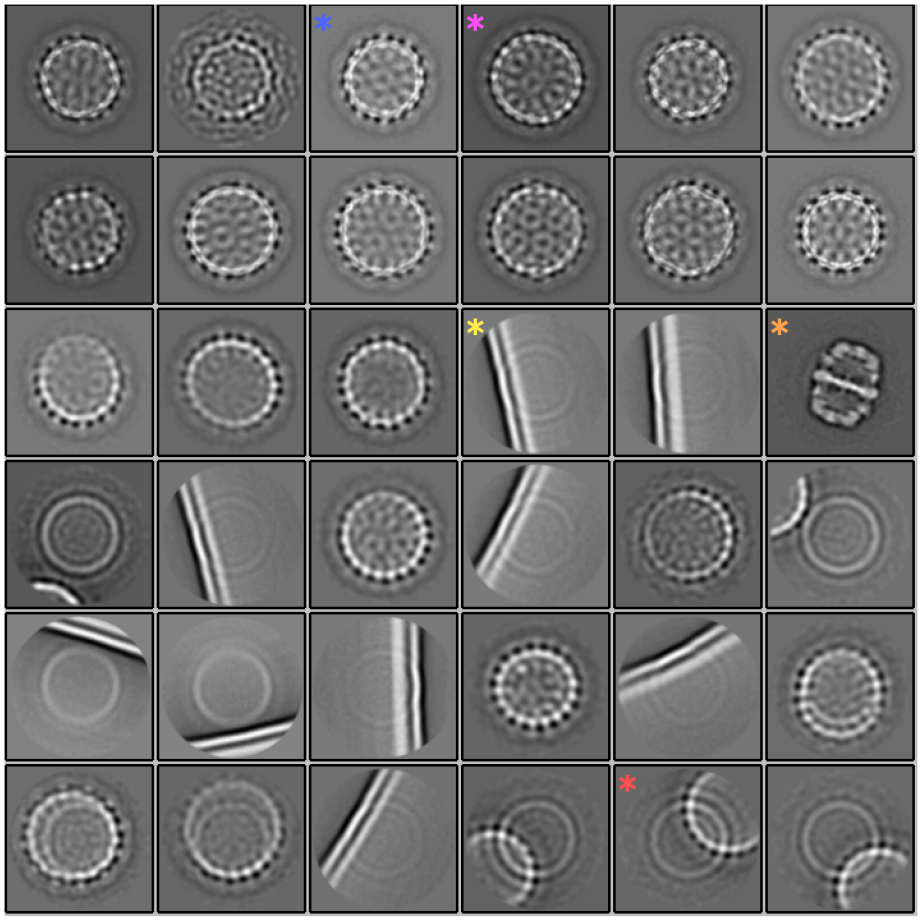
Unfortunately, a subsequent western blot of the AR1 sample used for cryoEM (performed by JA) showed evidence of partial degradation of AR1 in the time between purification and sample vitrification (data not shown). To solve this, fresh AR1 was expressed and purified (by JA) with vitrification performed quickly after purification for a new cryoEM data collection. A western blot of

the remaining sample after vitrification showed no evidence of degradation (data not shown).

Ultimately, an AR1 dataset containing 3,643 micrographs was collected on the FEI Titan Krios (a full set of data collection parameters is given in Table 5.2). Initially, just under 250,000 particles were extracted from these micrographs using the autopicking function of RELION, but a large proportion of these were 'junk' particles (e.g. areas of thick carbon backing or ice contamination) or non-AR1 contaminants (discussed in Section 5.5). After an initial round of 2D classification to remove non-AR1 particles (Figure 5.7), 49,489 $T = 4$ particles and 51,376 $T = 3^*$ particles were extracted into separate particle stacks and used to generate 3D reconstructions.

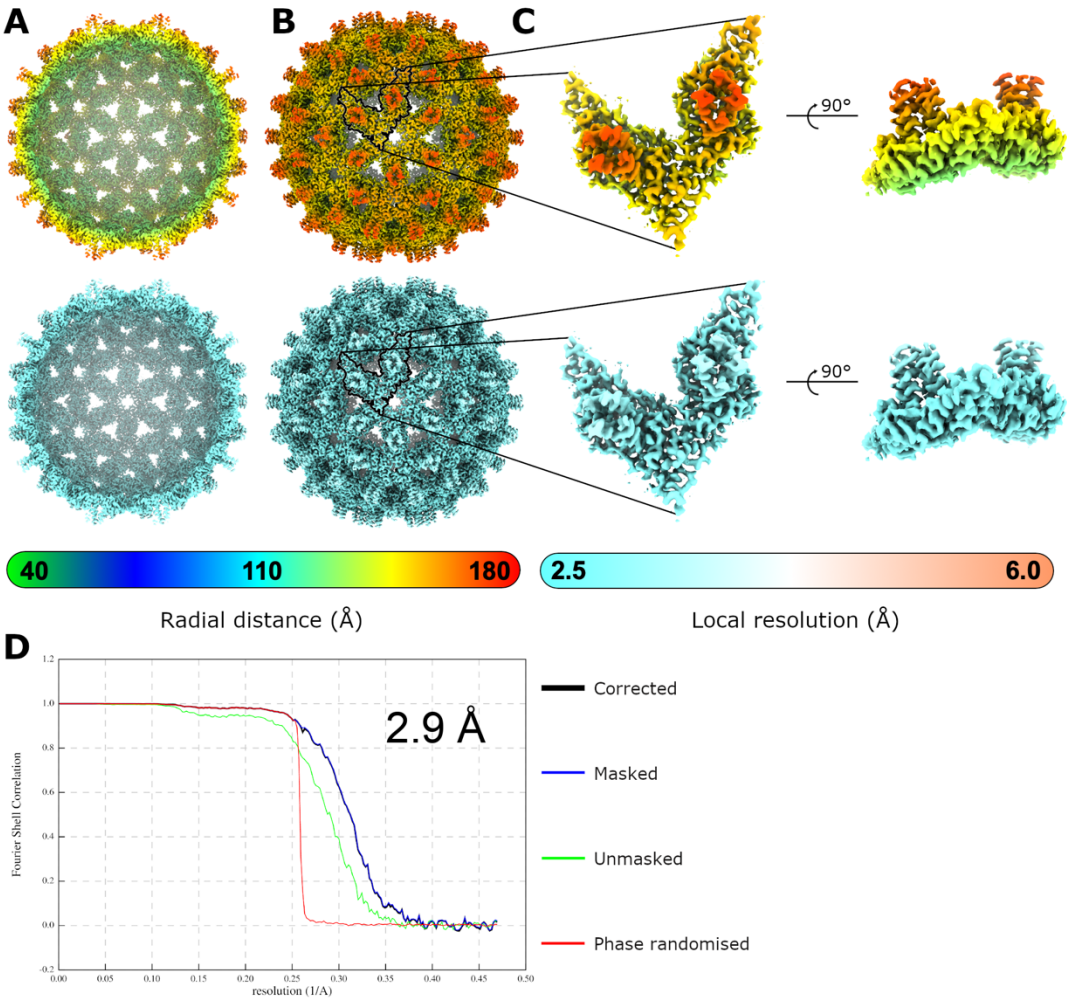
Table 5.2. Data collection parameters for VelcroVax samples.

Sample	**AR1**	**n190**	**n190-GP1**
Microscope	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Detector mode	Linear	Linear	Linear
Camera	Falcon III	Falcon III	Falcon IV
Voltage (kV)	300	300	300
Pixel size (Å)	1.065	1.065	1.065
Nominal magnification	75,000×	75,000×	75,000×
Exposure time (s)	1.3	1.0	1.3
Total dose (e^-/Å^2)	60	43	54.7
Number of fractions	40	30	40
Defocus range (μm)	−0.8 to −3.0	−0.5 to −2.9	−0.5 to −2.9
Number of micrographs	3,643	12,797	23,966
Acquisition software	Thermo Scientific EPU	Thermo Scientific EPU	Thermo Scientific EPU



5.7. Initial 2D classification of AR1. *Particles were two-fold down-sampled and subjected to 2D classification. Class averages for the 36 most populated classes are shown. Coloured asterisks show example classes for T = 3* (i.e., pseudo-T = 3, illustrated in Figure 5.9A and explained further in Section 5.2.2) VLPs (blue), T = 4 VLPs (pink), and 'junk' classes including carbon support (yellow), yeast fatty acid synthase (orange) and overlapping AR1 VLPs (red).*

Initially, 3D refinement of *T* = 4 AR1 (with icosahedral [I1] symmetry imposed) yielded a 3.5-Å resolution structure following sharpening, and this was improved through two rounds of CTF refinement and Bayesian polishing to 2.9 Å resolution (Figure 5.8).



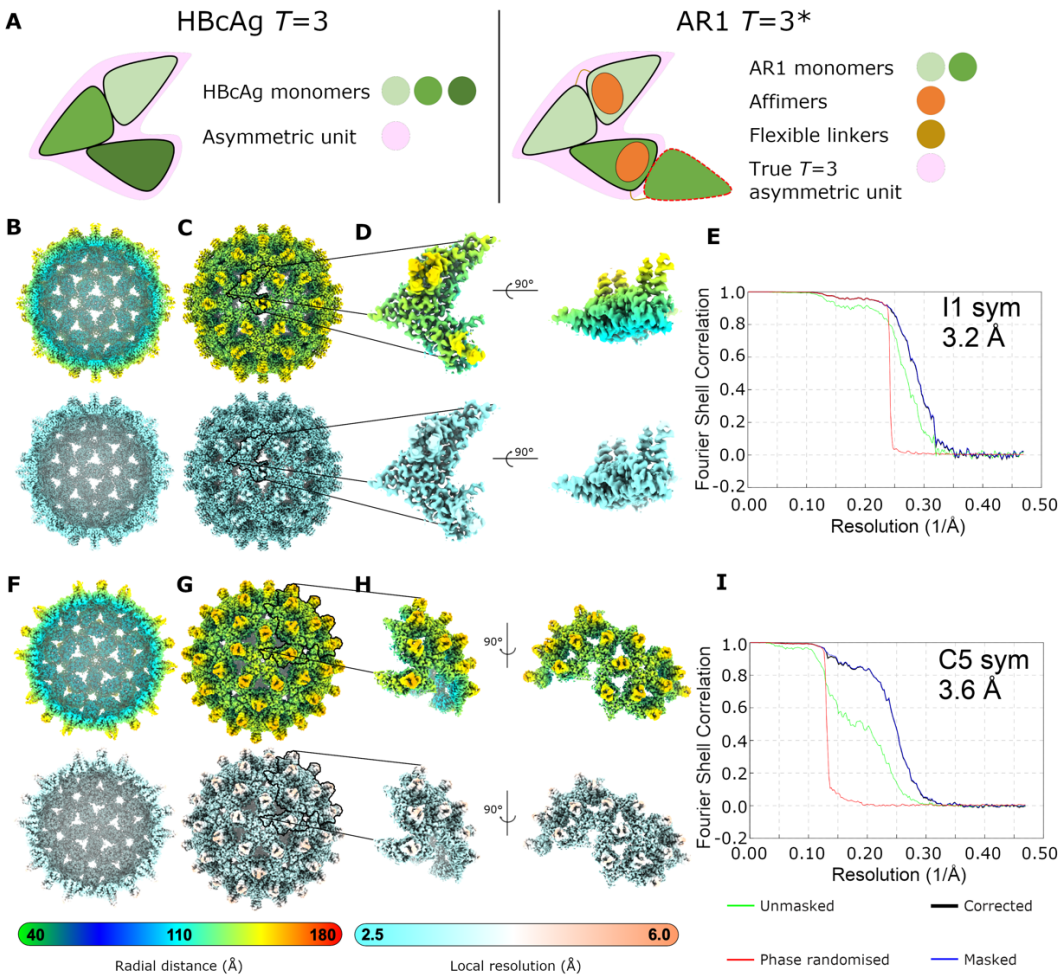
5.8. CryoEM density map for the T = 4 configuration of AR1. (A) *Cross section of T = 4 AR1 density map, filtered according to local resolution.* ***(B)*** *Full isosurface representation of T = 4 AR1 density map, filtered according to local resolution. The asymmetric unit is outlined.* ***(C)*** *Enlarged density for the asymmetric unit of T = 4 AR1. All maps in (A-C) were shown at 4 σ and coloured according to the radial colouring scheme (top) or local resolution colouring scheme (bottom) shown.* ***(D)*** *FSC plot for the T = 4 AR1 map. The resolution (2.9 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).*

Structural determination for the *T* = 3* configuration of AR1 was more complex. Since each subunit of AR1 comprises two different HBcAg monomers connected by a flexible linker, strictly speaking, the *T* = 3* configuration does

not conform to icosahedral symmetry (and is therefore denoted here as $T = 3^*$). True $T = 3$ icosahedral symmetry would require three HBcAg monomers per asymmetric unit, but since the two HBcAg monomers that form each subunit of AR1 are connected by a flexible linker and are intrinsically different (one is modified to include an Affimer, while the other maintains the standard MIR), this leads to an imbalance between what would be icosahedral asymmetric units (Figure 5.9A). Some asymmetric units would contain two copies of the Affimer-coupled HBcAg monomer and one Affimer-free HBcAg monomer, and others would contain the inverse arrangement – meaning they would not truly be asymmetric units. Regardless, imposition of icosahedral (I1) symmetry yielded a high-resolution reconstruction (3.2 Å post-sharpening) (Figure 5.9B-E), likely because the components of AR1 that render the $T = 3^*$ configuration non-icosahedral (i.e., the linker sequence and the Affimer) are flexible, and therefore mostly unresolved in the map anyway. To generate a technically correct structure, the $T = 3^*$ configuration was reprocessed with five-fold (C5) symmetry imposed, leading to a 3.6-Å resolution reconstruction after sharpening (Figure 5.9F-I).

5.2.3. The cryoEM structures of AR1

To build atomic models for the $T = 4$ and $T = 3^*$ arrangements of AR1, SWISS-MODEL (Waterhouse *et al.*, 2018) was used to generate a homology model based on the peptide sequence of AR1. The homology model was rigid-body fitted into each density map, and unresolved segments (notably, the Affimer and flexible linker) were removed. Copies of the model were then fitted appropriately to generate a complete asymmetric unit (2 copies of AR1 in total for $T = 4$, 18 copies for $T = 3^*$). The atomic models were each inspected and refined iteratively, including refinement following symmetrisation to avoid erroneous fitting of the model into strong unoccupied density from adjacent asymmetric units. Model validation statistics are given in Table 5.3.



5.9. CryoEM density maps for the T = 3* configuration of AR1. (A) *The T = 3* configuration of AR1 does not conform to icosahedral symmetry. The schematic shows the asymmetric unit for the T = 3 configuration of unmodified HBcAg VLPs (left) with individual HBcAg monomers in different shades of green. This contrasts with AR1 (right), which does not conform to true icosahedral symmetry in this configuration.* ***(B-D)*** *Density map for T = 3* AR1 determined with the imposition of icosahedral (I1) symmetry, filtered according to local resolution. This includes* ***(B)*** *a central cross section,* ***(C)*** *a complete isosurface representation (with pseudo-asymmetric unit outlined) and* ***(D)*** *enlarged views of density for the pseudo-asymmetric unit alone.* ***(E)*** *FSC plot for T = 3* AR1 determined with the imposition of icosahedral (I1) symmetry.* ***(F-H)*** *Density map for T = 3* AR1 determined with the imposition of cyclical five-fold (C5) symmetry, filtered according to local resolution. This includes* ***(F)*** *a central cross section,* ***(G)*** *a complete isosurface*

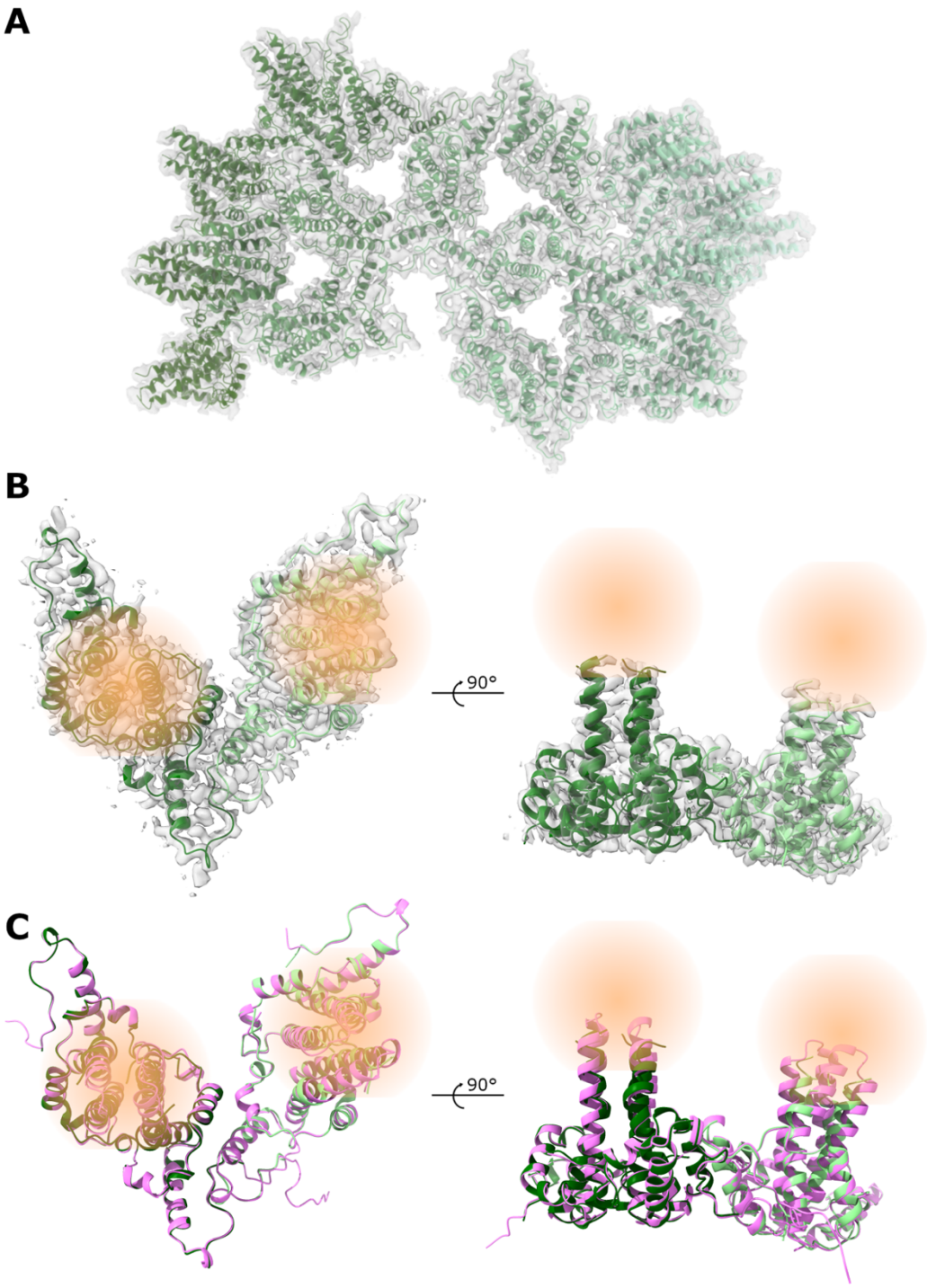
*representation (with asymmetric unit outlined) and **(H)** enlarged views of density for the asymmetric unit alone. **(I)** FSC plot for T = 3* AR1 determined with the imposition of cyclical five-fold (C5) symmetry. All maps are shown at 4 σ, coloured according to the radial colouring scheme or local resolution colouring scheme shown. Resolutions were determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).*

Table 5.3. Map and model validation statistics for AR1.

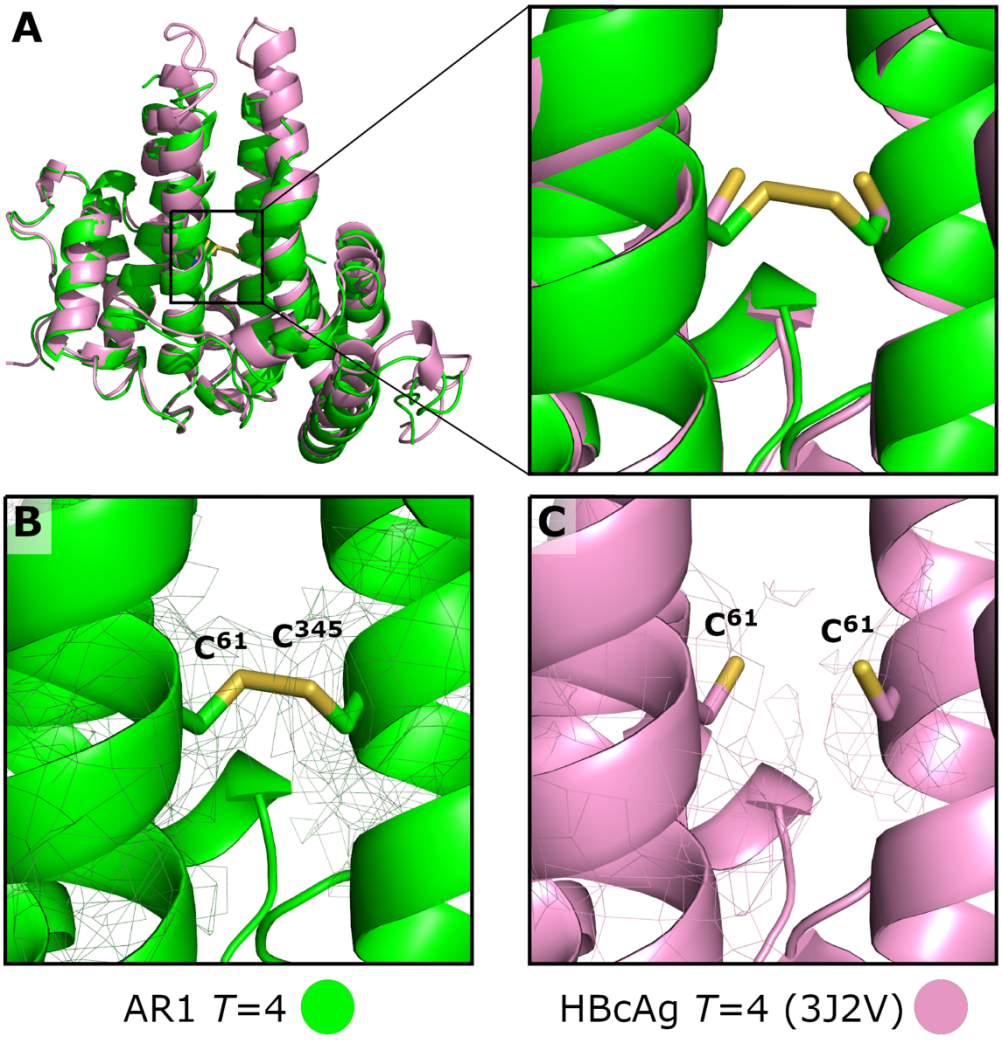
Model	**AR1 (*T* = 4)**	**AR1 (*T* = 3*)**
EMDB ID	*To be deposited*	*To be deposited*
PDB ID	*To be deposited*	*To be deposited*
CryoEM map processing		
Symmetry imposed	I1	C5
Number of particles contributing to map	49,489	51,376
Map resolution (FSC = 0.143) (Å)	2.9	3.6
Map resolution range around atom positions (Å)	2.9 – 3.4	3.4 – 5.1
Map sharpening B factor (Å^2)	−143	−183
Residues modelled		
Chain A	1–75; 195–255; 285–358; 375–431.	1–76; 195–255; 285–357; 375–431.
Chain B	1–76; 195–255; 285–360; 372–431.	1–76; 195–255; 285–359; 372–431.
Chain C	N/A	1–76; 192–255; 285–358; 375–431.
Chain D	N/A	1–75; 195–255; 285–358; 375–431.
Chain E	N/A	1–76; 195–255; 285–358; 375–431.
Chain F	N/A	1–76; 195–255; 285–356; 377–431.
Chain G	N/A	1–76; 195–255; 285–358; 375–431.
Chain H	N/A	1–76; 195–254; 285–360; 372–431.
Chain I	N/A	1–75; 195–255; 285–360; 373–431.
Chain J	N/A	1–76; 195–255; 285–356; 375–431.
Chain K	N/A	1–76; 195–255; 285–358; 375–431.
Chain L	N/A	1–76; 199–255; 285–360; 372–431.
Chain M	N/A	1–76; 195–255; 285–358; 375–431.
Chain N	N/A	1–76; 195–255; 285–360; 373–431.
Chain O	N/A	1–76; 195–255; 285–358; 375–431.
Chain P	N/A	1–76; 195–255; 285–358; 375–431.
Chain Q	N/A	1–76; 195–255; 285–360; 372–431.
Chain R	N/A	1–76; 195–255; 285–358; 375–431.

RMSD		
Bond lengths (Å)	0.0076	0.0068
Bond angles (°)	1.13	1.11
Validation		
All-atom clashscore	2.44	3.19
MolProbity score	1.12	1.38
Rotamer outliers (%)	0.21	0.21
Ramachandran plot		
Favoured (%)	97.52	96.04
Allowed (%)	2.48	3.96
Outliers (%)	0.00	0.00

AR1 showed a high level of structural similarity to unmodified structures of HBcAg (Figure 5.10). Following alignment of the *T* = 4 asymmetric unit of AR1 with the best-matched subunit of a 2.8 Å resolution cryoEM structure of a *T* = 4 HBcAg VLP (PDB: 7OD4) (Makbul *et al.*, 2021), the RMSD value between equivalent Cα atoms was only ~1.5 Å. Visual inspection suggested that most variation occurred within the four-helix bundles, the site of Affimer insertion in AR1 (Figure 5.10C). In line with previous studies of HBcAg described in Section 5.1 (Nassal *et al.*, 1992; Zheng *et al.*, 1992; Wynne *et al.*, 1999), the AR1 structures reported here contained a disulphide bond linking C61 and C345, straddling the interface between each copy of HBcAg within the four-helix bundle, as expected (Figure 5.11).



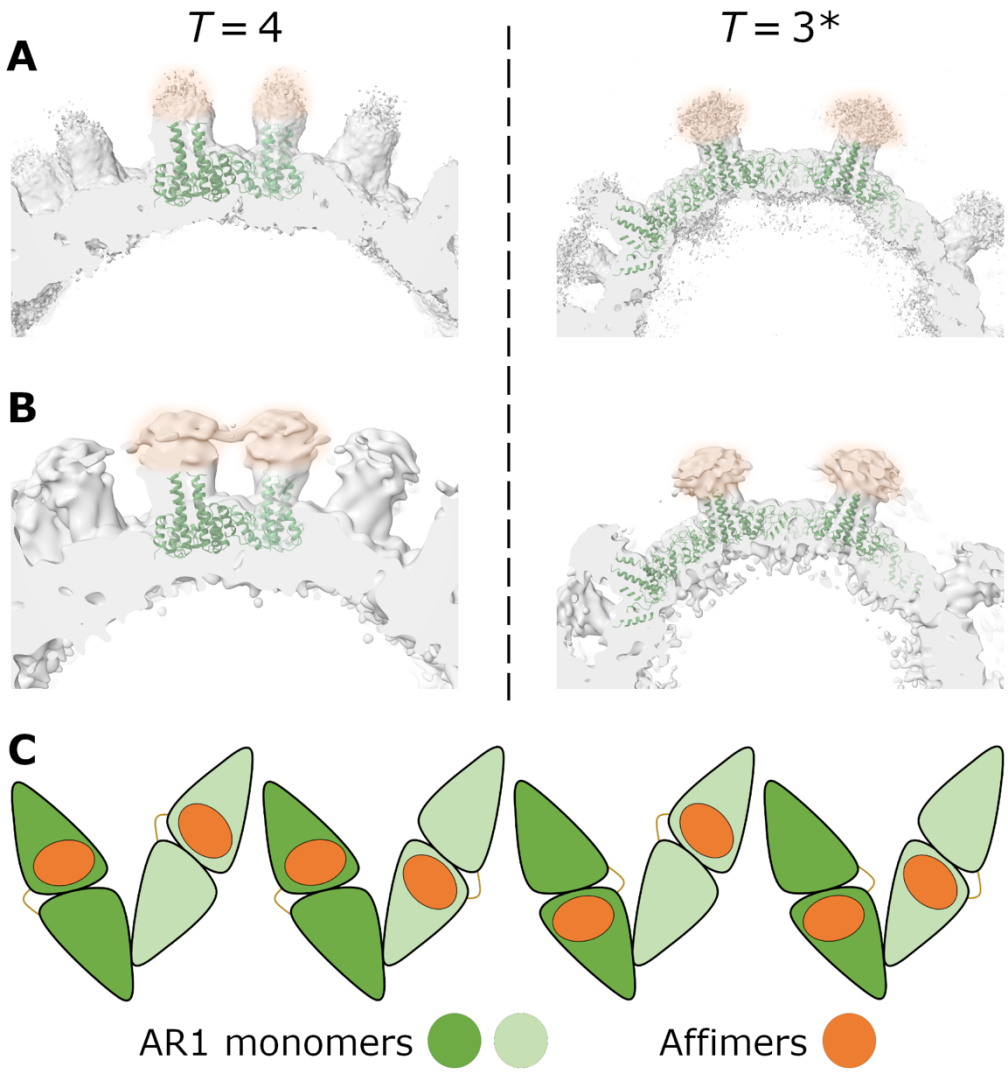
5.10. AR1 atomic models. ***(A)*** *Atomic model for AR1 T = 3* asymmetric unit, with EM density overlaid at ~4 σ.* ***(B)*** *Atomic model for AR1 T = 4 asymmetric unit, with EM density overlaid at ~4 σ. Approximate positions of the unresolved Affimers are shown by the orange shaded regions.* ***(C)*** *The atomic model for AR1 T = 4 shown in (B) (green) with an atomic model for unmodified HBcAg overlaid (PDB: 7OD4, (Makbul et al., 2021), pink).*



5.11. AR1 contains a disulphide bond linking C61-C345. (A) *Overlay of a single AR1 monomer with a single dimer of HBcAg (PDB: 3J2V, (Yu et al., 2013b)). Inset shows enlarged image of the four-helix bundle, centred around the C61-C345 (AR1) or C61-C61 (HBcAg) interaction.* ***(B,C)*** *The same atomic models shown in (A) with density maps overlaid (2 σ) for* ***(B)*** *AR1 and* ***(C)*** *HBcAg (PDB: 3J2V).*

Even at low contour levels, density corresponding to the Affimers was not visible in either $T = 4$ or $T = 3^*$ density maps, so Affimers could not be modelled. Fragmented density was visible above the MIR in the unsharpened density maps (Figure 5.12A), and amorphous density in this position was observed in a map low-pass filtered to 10 Å resolution (Figure 5.12B), suggesting the Affimers were present but were too flexible to be resolved. This was likely exacerbated by irregular positioning of the Affimers due to the

inherent asymmetry within each AR1 subunit. Only one of the two copies of HBcAg within the AR1 subunit contains an inserted Affimer. As such, when AR1 subunits are incorporated into the VLP (in either orientation) it is likely there would be an even split in the positioning of the Affimer relative to each four-helix bundle (Figure 5.12C). This would result in a reduction of the 'effective occupancy' of each potential Affimer site to 50%, further accounting for the poor Affimer density. As expected, the flexible linkers connecting each copy of HBcAg within the AR1 subunits were also unresolved and therefore not included in the atomic models.



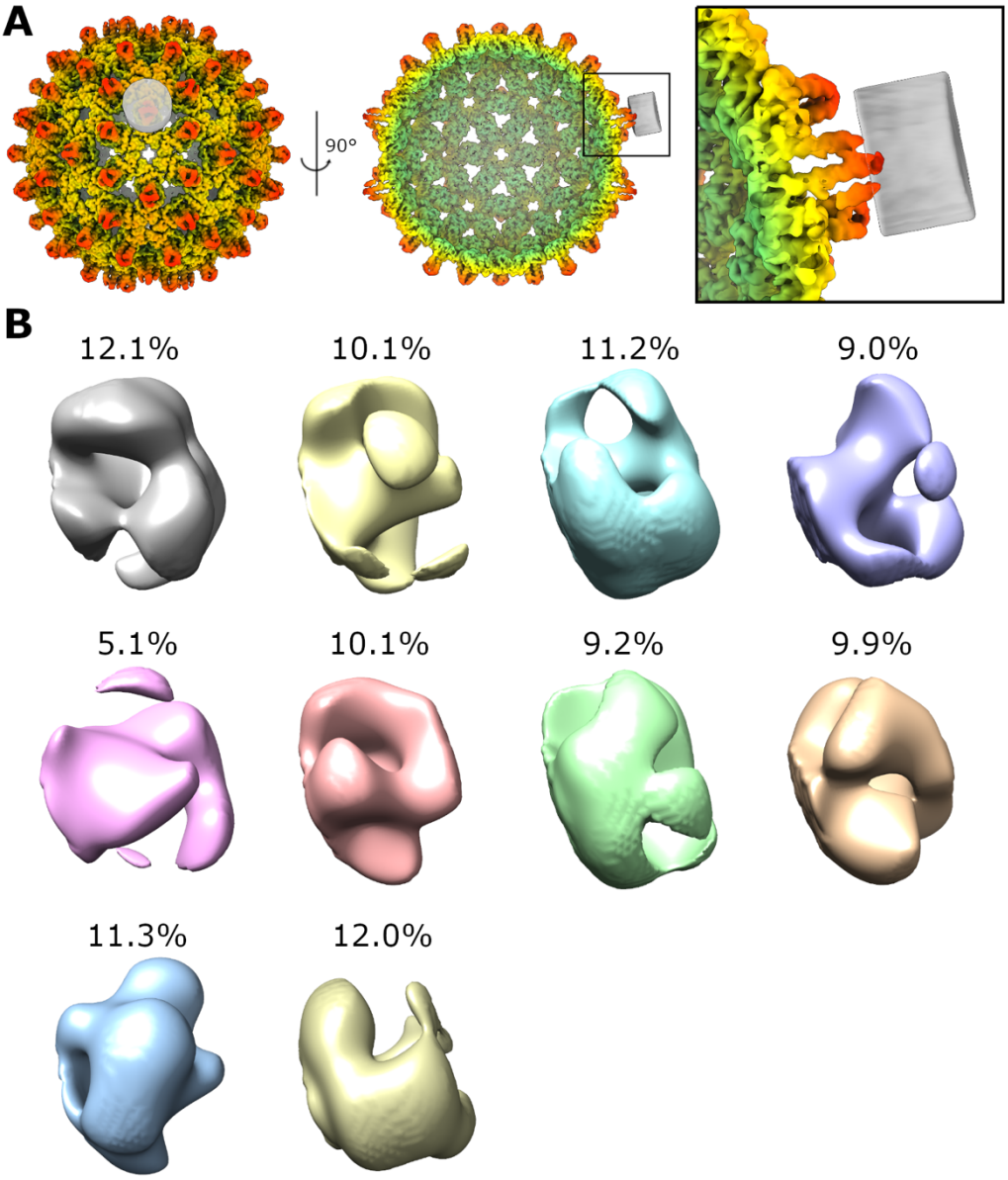
5.12. Affimers are unresolved in AR1 density maps. (A) *Sections of unsharpened density maps from T = 4 and T = 3* arrangements of AR1 shown at a low contour level (~1 σ) and overlaid with atomic models of the*

AR1 asymmetric units (lacking the Affimer residues). Density highlighted in orange corresponds to putative fragmented Affimer density. ***(B)*** *Sections of local resolution-filtered density maps for AR1 VLPs with 10 Å low-pass filter applied, showing amorphous Affimer density (highlighted orange) above the AR1 four-helix bundles. Density maps shown at ~0.5 σ.* ***(C)*** *Schematic showing the different orientations that each monomer of AR1 could adopt within the T = 4 asymmetric unit. For each monomer, the Affimer could be positioned in one of two positions relative to the central four-helix bundle.*

5.2.4. Focussed classification of AR1 Affimers

In an attempt to resolve Affimer density, the *T* = 4 particle stack was taken forward for focussed 3D classification, with a cylindrical mask positioned over density corresponding to a four-helix bundle (Figure 5.13A). There was a considerable level of variation between focussed classes, but no single class contained density reminiscent of an Affimer and all were relatively poorly resolved (Figure 5.13B). Subparticles were split reasonably evenly between all ten classes, suggesting that there was no dominant conformation. Thus, focussed classification results indicated that the Affimers contained within AR1 were indeed highly flexible and variable in their positioning.

Because of the unique symmetrical arrangement of the *T* = 3* capsid (i.e., with five-fold [C5] symmetry as the highest order symmetry), focussed classification would only give five symmetrically redundant subparticles per Affimer position instead of the sixty generated for the *T* = 4 capsid (icosahedral [I1] symmetry). As such, the chance of success would be considerably lower. Given the failure of focussed classification to resolve any Affimer density for the *T* = 4 configuration of AR1 and the high computational demand of focussed classification, this process was not performed for the *T* = 3* configuration.



5.13. T = 4 AR1 focussed classification. ***(A)*** *Mask placement for focussed classification of T = 4 AR1 Affimers. The mask used for focussed classification is shown in grey. The inset shows an enlarged image of the region highlighted by the black box.* ***(B)*** *All focussed classes from focussed classification of T = 4 AR1, with the proportion of subparticles assigned to each class indicated. Classes are oriented in the same way as the mask shown in the inset in (A) and coloured according to arbitrary class number.*

5.2.5. Implications for further structural work

While biochemical evidence supported the fact that Affimers genetically inserted at the distal end of the four-helix bundles were intact at the time of vitrification, these were not resolved in the cryoEM density maps, even following focussed classification. These findings indicate that AR1 is unlikely to be a useful platform for further structural studies of 'VelcroVax' technology. While it is possible that ligand binding may stabilise the Affimers to some extent, it is unlikely that such structures would be sufficiently resolved to provide useful structural information on Affimer-ligand interactions, given the issues described above. Furthermore, the nature of tandem core technology introduces additional complexity to the processing of *T* = 3* VLPs, as they do not obey icosahedral symmetry and therefore need to be treated as five-fold (C5) symmetrical complexes. This means that not all of the redundant information contained within each copy of the strict icosahedral asymmetric unit is combined, reducing the final resolution of the map for any given number of input particles.

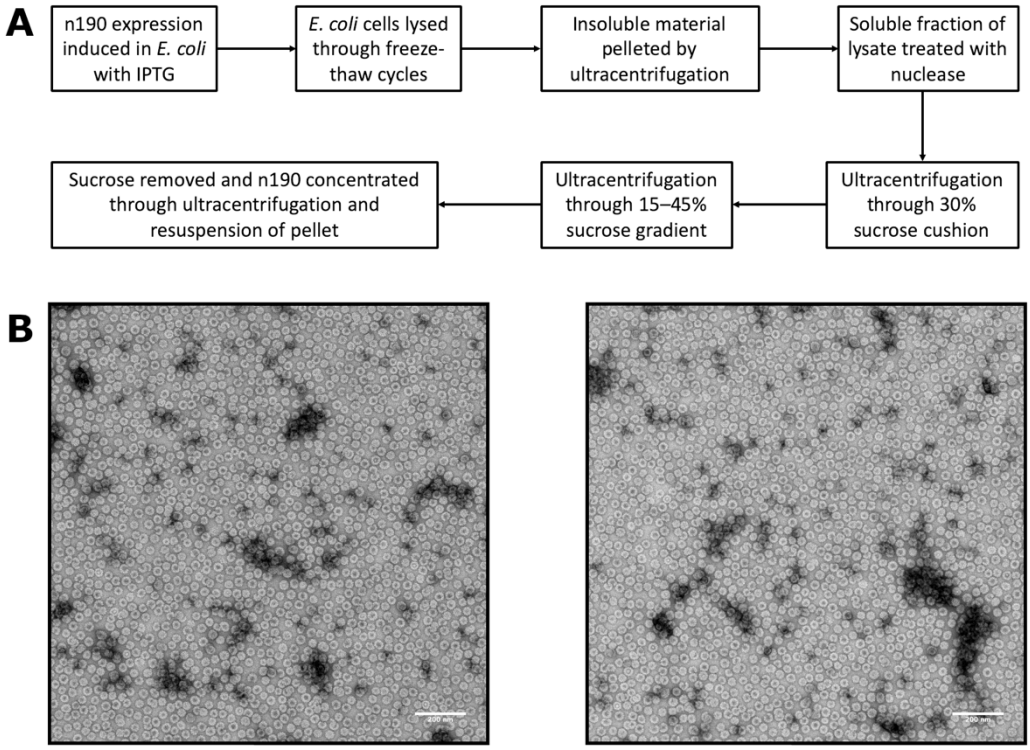
5.3. The structure of unliganded n190

As described in Section 5.1.3, the n190 construct differed from AR1 in that it comprised just a single monomer of HBcAg, with an anti-SUMO Affimer inserted at the N-terminus rather than the MIR. Structural characterisation of n190 was performed in order to compare n190 VLPs with AR1 VLPs, and to examine the suitability of n190 as a target for visualisation of the inserted Affimer using cryoEM.

5.3.1. n190 expression and purification

n190 was expressed and purified by KF. Importantly, in contrast to AR1 (which was produced in the yeast strain *P. pastoris*), n190 expression was induced in *E. coli* cells using IPTG. Following expression of n190, *E. coli* cells were lysed via repeated freeze-thaw cycles and the lysate was nuclease treated, before purification via a 30% sucrose cushion and a 15–45% sucrose gradient (Figure

5.14A). Negative stain EM analysis was performed to confirm that n190 was sufficiently pure and concentrated for cryoEM (Figure 5.14B).

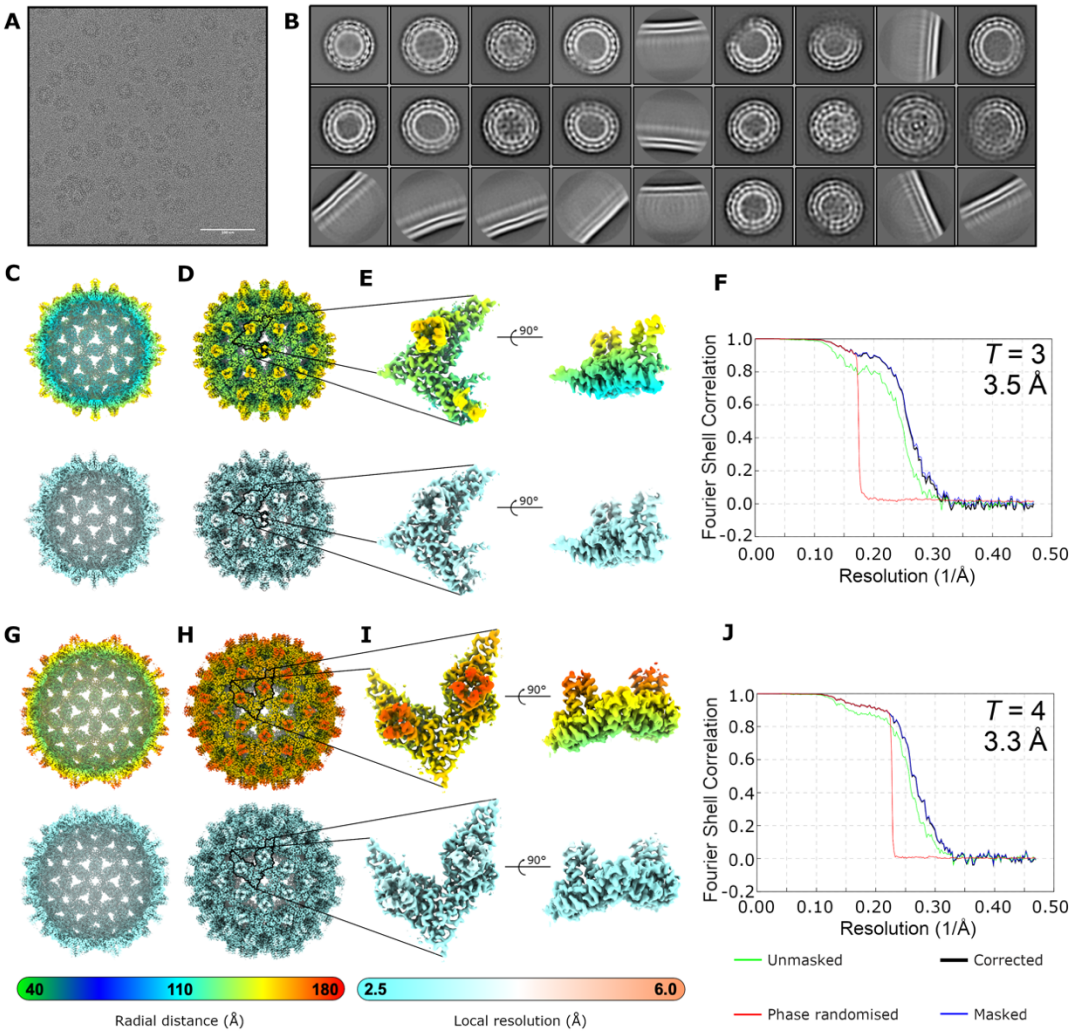


5.14. Expression and purification of n190. (A) *Simplified flow diagram for expression and purification of n190 in E. coli.* ***(B)*** *Representative negative stain EM micrographs of n190 following purification. Scale bar shows 200 nm. n190 expression and purification performed by KF. Negative stain EM performed by the author of this thesis.*

5.3.2. n190 sample preparation, cryoEM data collection and processing

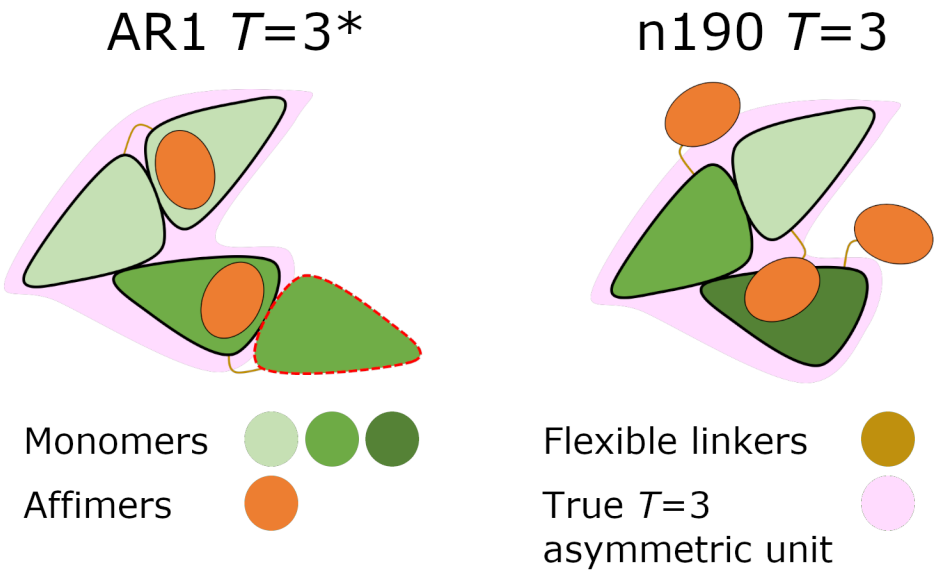
n190 was vitrified for cryoEM in the same way as described for AR1 above. Several grids were prepared using a range of blotting times (between 1 – 4 seconds) and were screened to select for optimal ice thickness and particle distribution. An optimised grid was taken forward for a 72-hour data collection on an FEI Titan Krios microscope. Full details of cryoEM data collection parameters are given in Table 5.2.

12,797 micrographs were collected and used to extract 407,492 particles for image processing (Figure 5.15A). This particle stack included a large number of ‘junk’ particles, so the number of useful particles corresponding to true n190 VLPs was thought to be significantly lower. Due to the computational demands of such a large particle stack, particles were 2× down-sampled for initial classification, including several rounds of 2D classification to progressively remove ‘junk’ particles and separate $T = 3$ and $T = 4$ VLPs. Notably, 2D classes showed evidence of additional density beyond the main icosahedral shell formed by HBcAg, at the same radial distance as the four-helix bundles, indicative of Affimer density (Figure 5.15B). Some class averages also looked distorted in one dimension, suggesting that n190 VLPs may be less stable than AR1 VLPs, or that n190 VLPs were exposed to a greater level of physical stress during expression/purification. Following classification, final particle stacks were re-extracted without down-sampling, yielding 40,254 $T = 3$ particles and 56,416 $T = 4$ particles for generation of 3D reconstructions. Iterative rounds of CTF refinement and Bayesian polishing were also performed, ultimately leading to 3.5 Å ($T = 3$) (Figure 5.15C-F) and 3.3 Å ($T = 4$) (Figure 5.15G-J) resolution reconstructions. In contrast to the tandem organisation of the AR1 construct, the non-tandem nature of n190 meant that the $T = 3$ VLP (as well as $T = 4$) obeyed strict icosahedral symmetry, simplifying the processing workflow and maximising the final resolution achievable (Figure 5.16).



5.15. CryoEM reconstructions of n190 VLPs. ***(A)*** *Representative micrograph from n190 data collection. Scale bar shows 100 nm.* ***(B)*** *Most populated classes from 2D classification of all auto-picked particles from the n190 dataset following two-fold down-sampling.* ***(C-E)*** *Density map for T = 3 n190 filtered according to local resolution. This includes* ***(C)*** *a central cross section,* ***(D)*** *a complete isosurface representation (with asymmetric unit outlined) and* ***(E)*** *enlarged views of density for the asymmetric unit alone.* ***(F)*** *FSC plot for T = 3 n190.* ***(G-I)*** *Density map for T = 4 n190, filtered according to local resolution. This includes* ***(G)*** *a central cross section,* ***(H)*** *a complete isosurface representation (with asymmetric unit outlined) and* ***(I)*** *enlarged views of density for the asymmetric unit alone.* ***(J)*** *FSC plot for T = 4 n190. All maps were shown at 4* σ *and coloured according to the radial colouring scheme or local resolution colouring scheme shown. Resolutions*

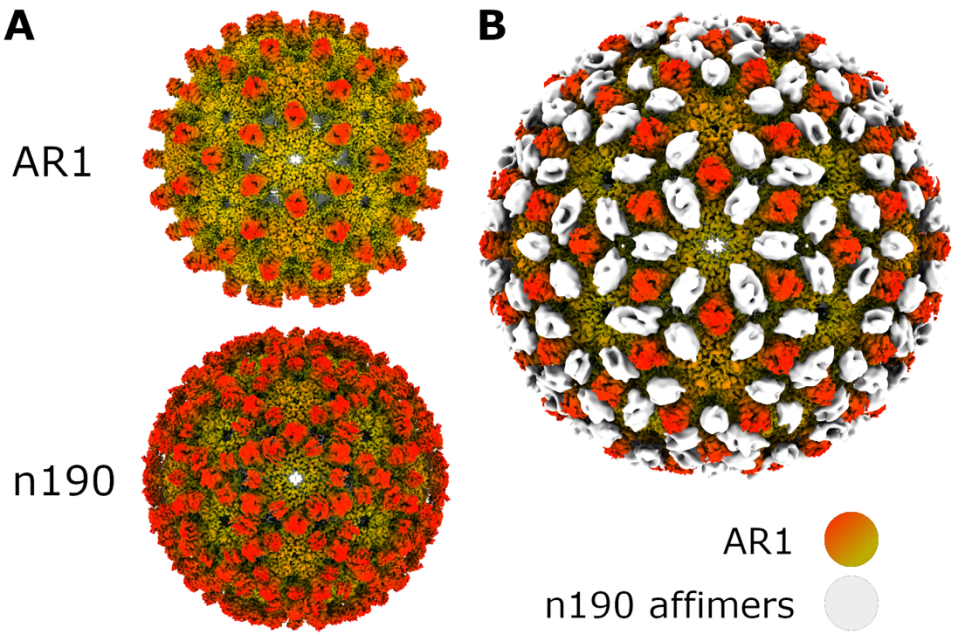
were determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected).



5.16. The T = 3 configuration of n190 obeys icosahedral symmetry. *Schematic of the T = 3 asymmetric unit of n190 (right) in comparison to the T = 3* arrangement of AR1 (left).*

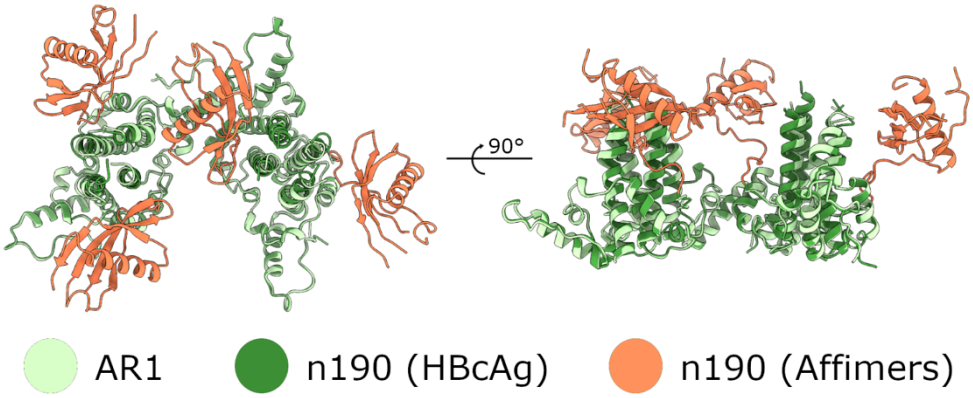
5.3.3. The cryoEM structures of n190

The most striking feature of the n190 VLP cryoEM reconstructions was the presence of additional densities between four-helix bundles at low contour levels, for both *T* = 3 and *T* = 4 density maps (Figure 5.17). These densities were of the correct size to accommodate an Affimer, and were positioned above the N-terminus of the HBcAg monomer. While the densities were not of sufficient resolution for detailed molecular modelling, in the *T* = 4 density map secondary structural features (specifically, the core α-helix and anti-parallel β sheet) could be identified, facilitating rigid-body fitting of homology models for the Affimer.



5.17. Additional Affimer density is observed for n190. (A) *T = 4 VLP density maps for AR1 (upper) and n190 (lower), filtered by local resolution and shown at a low contour level (~2 σ).* ***(B)*** *The maps shown in (A) were subjected to a 6 Å low-pass filter, then used to calculate a difference map. The difference map is shown in white, overlaid on the AR1 density map shown in (A), highlighting the additional densities in n190 that correspond to N-terminal linked Affimers.*

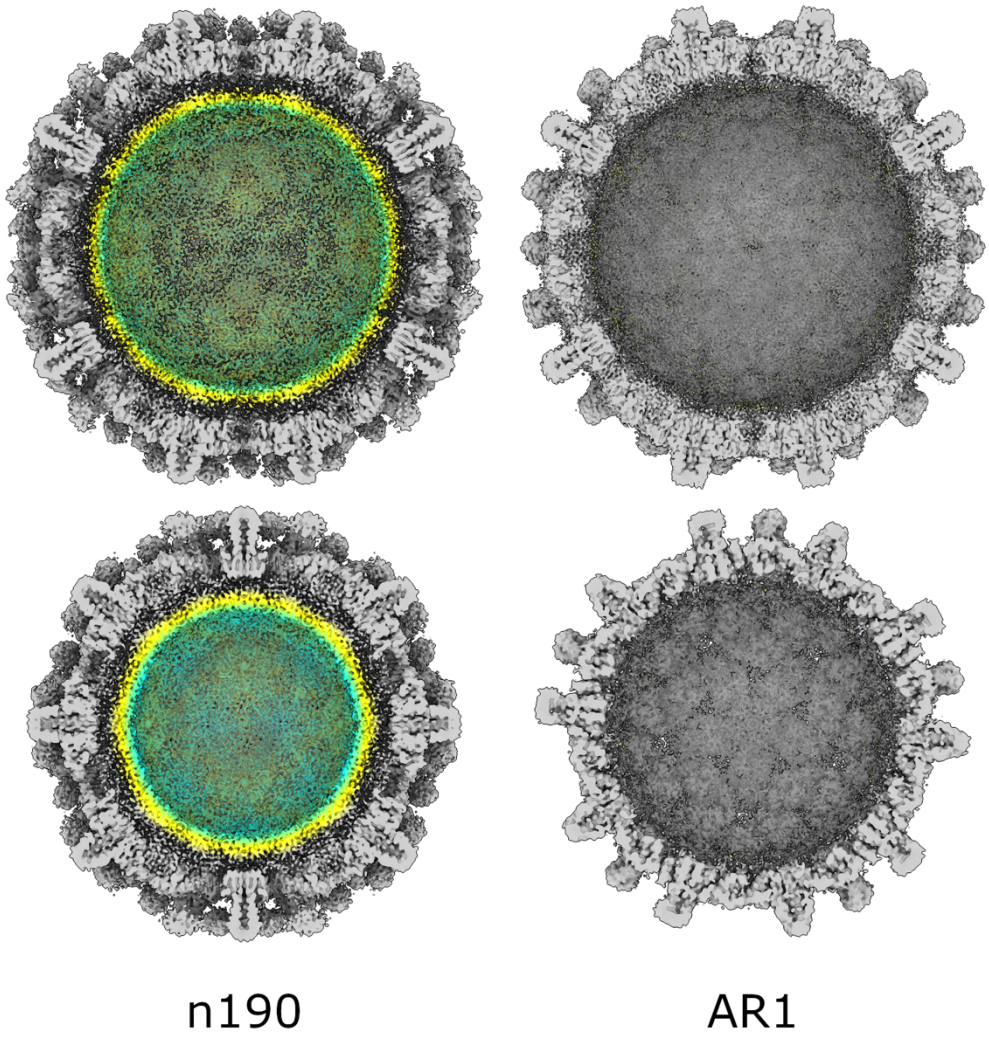
The HBcAg components of n190 overlaid well with the AR1 structure. For the *T* = 4 VLPs, the asymmetric units showed an RMSD value of only 0.92 Å following alignment based on equivalent Cα atoms. It was clear from both RMSD analysis and visual inspection that the greatest deviation between the two structures was in the four-helix bundles (Figure 5.18). This may be related to the alternative Affimer positioning, either attached to the four-helix bundles (AR1) or positioned between them (n190).



5.18. Comparison of T = 4 VLP structures for n190 and AR1. *T = 4 atomic models for n190 and AR1 were aligned using the 'MatchMaker' tool from UCSF Chimera. Models were coloured according to the colouring scheme indicated.*

5.3.4. n190 VLPs contain nucleic acid

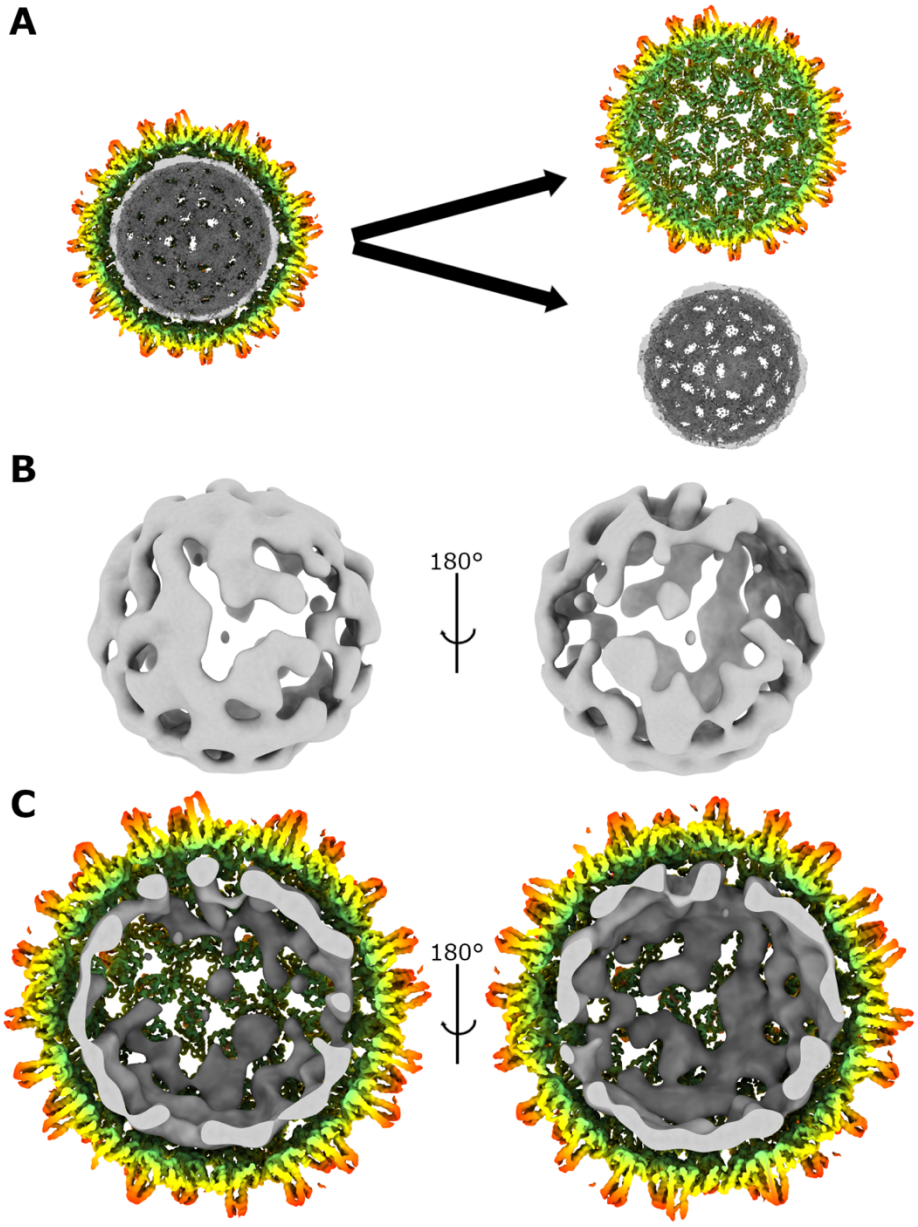
Another difference between n190 and AR1 VLPs was observed within the capsid interior. While AR1 had no observable density within the core of the VLP, low-resolution density could be identified close to the internal surface of n190 in both *T* = 3 and *T* = 4 VLPs (Figure 5.19). This density likely corresponds to nucleic acid packaged during VLP assembly. To confirm this structural finding, an A_{260}:A_{280} ratio was calculated for the n190 sample. Nucleic acid shows peak absorption of light at 260 nm, whereas protein (primarily through tyrosine and tryptophan residues) has an absorption peak at 280 nm. As such, a the A_{260}:A_{280} ratio gives an indication of how much nucleic acid is present within the sample. For n190, this ratio was ~1.8 (compared to ~1.0 for AR1), suggesting the presence of a considerable amount of nucleic acid within the sample (either packaged inside the VLPs or as unpackaged nucleic acid that simply co-purified).



5.19. Putative nucleic acid density within the interior of n190 VLPs. *Local-resolution filtered density maps for n190 and AR1 are shown for T = 4 (upper, ~1.25 σ) and T = 3 (lower, ~2 σ) VLPs. Density corresponding to AR1 and n190 proteins is coloured grey, with putative nucleic acid density coloured radially (separate scale for T = 4 and T = 3).*

In an attempt to gain more information on the internal density and with the assumption that it represented packaged nucleic acid, the n190 dataset was processed using symmetry relaxation, as was described in Chapter 4. Briefly, capsid density was used to subtract capsid signal from *T* = 4 particles, then subtracted particles were subjected to symmetry-relaxed 3D classification (with a single class) using a reference map comprising only nucleic acid density (Figure 5.20A). This yielded a reconstruction with some apparent

organisation, but no obvious regions of striation (as was observed for MNV in Section 4.3.2) (Figure 5.20B,C). Density was weighted towards one half of the reconstruction, with a large gap in density at the opposite pole. These results suggested that the nucleic acid packaged within different n190 VLPs was unlikely to have a strict and consistent organisation, and that 'coverage' of the capsid interior by nucleic acid was likely incomplete.



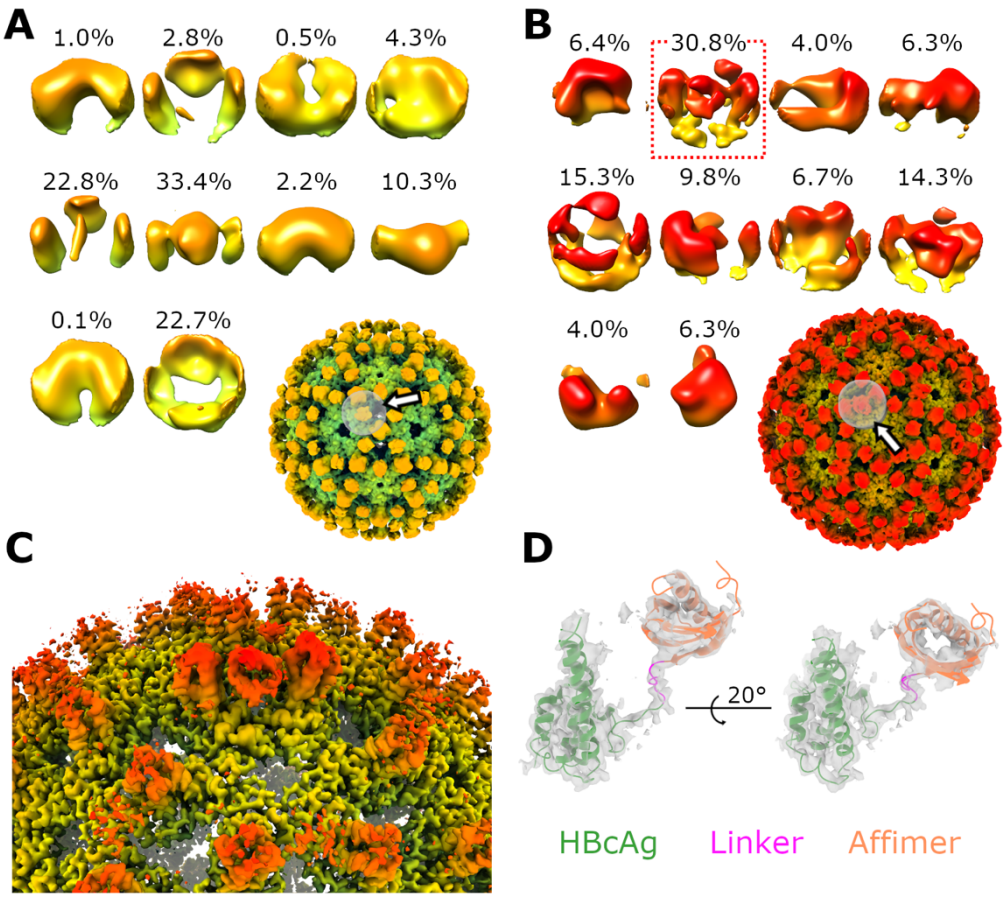
5.20. Attempt to resolve n190 nucleic acid using symmetry relaxation. ***(A)*** *The unsharpened 3D reconstruction of T = 4 n190 was split into capsid density (radial colouring) and nucleic acid density (grey). The capsid density was used to make a mask for subtraction of capsid signal from individual*

particles. The nucleic acid density was used as a reference for symmetry-relaxed 3D classification. ***(B,C)*** *Symmetry-relaxed density map for T = 4 n190, shown either* ***(B)*** *alone or* ***(C)*** *as a section in the context of the n190 capsid.*

5.3.5. Focussed classification of n190 Affimers

In an attempt to improve the resolution of Affimer density, focussed 3D classification was performed using masks centred on Affimers for both *T* = 3 and *T* = 4 density maps (Figure 5.21A,B). For the *T* = 3 VLP, there was no apparent improvement in Affimer density, although 33.4% of subparticles were grouped into a single class with clear Affimer-sized density present. Interestingly, 22.7% of subparticles were assigned to a class with density bordering the masked region (likely corresponding to adjacent four-helix bundles) but without density in the expected position of the Affimer (Figure 5.21A). This indicates that a proportion of Affimers may have denatured or been lost from the VLPs, for example, through protease-mediated cleavage.

For the *T* = 4 VLP, 30.8% of subparticles were assigned to a class with density showing clear structural features expected for an Affimer (Figure 5.21B). These subparticles were taken forward for full asymmetric reconstruction, leading to a map with clear secondary structural features for the Affimer including the central α-helix and separation between adjacent β-strands (Figure 5.21C). This permitted more accurate fitting of a homology model of the SUMO Affimer into the density (Figure 5.21D), although the variable loops involved in antigen capture remained unresolved. Given that the reconstruction was derived from focussed classification of a single Affimer, and therefore was not suitable for modelling a full asymmetric unit, the atomic coordinates built for this n190 monomer (including Affimer) were replicated and placed within different quasi-equivalent positions in the original symmetrised map to generate a complete asymmetric unit. This atomic model was then refined into the symmetrised density map to generate a final Affimer-containing atomic model for *T* = 4 n190 (model validation parameters provided in Table 5.4).

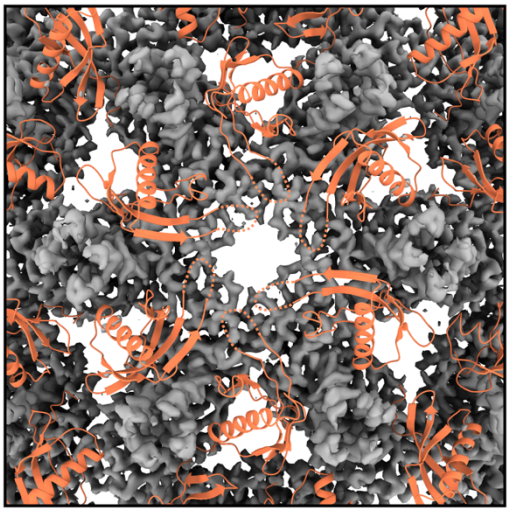


5.21. Focussed classification of n190 Affimers. (A,B) *All focussed classes from focussed classification of* ***(A)*** *T = 3 n190 and* ***(B)*** *T = 4 n190, with the proportion of subparticles assigned to each class indicated. The position of the mask (grey) is indicated on reference maps for each case. Classes are coloured according to radial distance (relative to the complete n190 VLP) and are oriented from the viewpoint indicated by the white arrow on each reference model.* ***(C)*** *Asymmetric reconstruction from particles contained in the class highlighted by the red dashed box in (B), following sharpening and masking.* ***(D)*** *Density for a single n190 monomer from the map shown in (C), overlaid with atomic coordinates for a single n190 monomer (including HBcAg (green), linker sequence (magenta) and Affimer (orange)).*

Table 5.4. Map and model validation statistics for n190.

Model		**n190 (*T* = 4)**	**n190 (*T* = 3)**
EMDB ID		*To be deposited*	*To be deposited*
PDB ID		*To be deposited*	*To be deposited*
CryoEM map processing			
Number of particles contributing to map		56,416	40,254
Map resolution (FSC = 0.143) (Å)		3.3	3.5
Map resolution range around atom positions (Å)		3.1 – 5.4	3.2 – 4.2
Map sharpening B factor (Å^2)		−179	−187
Residues modelled			
Quasi-equivalent position A	14-52, 59-83, 89-186, 198-258		111-186, 198-258
Quasi-equivalent position B	14-52, 59-83, 89-184, 194-258		111-186, 198-257
Quasi-equivalent position C	14-52, 59-83, 89-186, 198-257		111-185, 198-258
Quasi-equivalent position D	14-52, 59-83, 89-185, 194-258		N/A
RMSD			
Bond lengths (Å)		0.0102	0.0071
Bond angles (°)		1.27	1.07
Validation			
All-atom clashscore		6.39	2.61
MolProbity score		1.67	1.05
Rotamer outliers (%)		0.39	0.55
Ramachandran plot			
Favoured (%)		95.49	98.74
Allowed (%)		4.51	1.26
Outliers (%)		0.00	0.00

While the variable binding loops of the Affimer were unresolved and not included in the atomic model, it was possible to determine their approximate position based on the orientation of the Affimer (Figure 5.22). The loops projected towards the icosahedral three- and five-fold axes, and based on the size of the two unresolved loops (6 and 5 residues long, respectively), it was deemed likely that antigen binding would require reorientation of the Affimers to avoid steric clashes.



5.22. Affimer positioning in the n190 T = 4 VLP. *Atomic coordinates for Affimers (orange) within the n190 T = 4 VLP (grey density map), centred on an icosahedral three-fold axis. The expected approximate position of one of the two unresolved binding loops within each Affimer is indicated with a dashed orange line (not to scale). For clarity, the second unresolved loop is not indicated.*

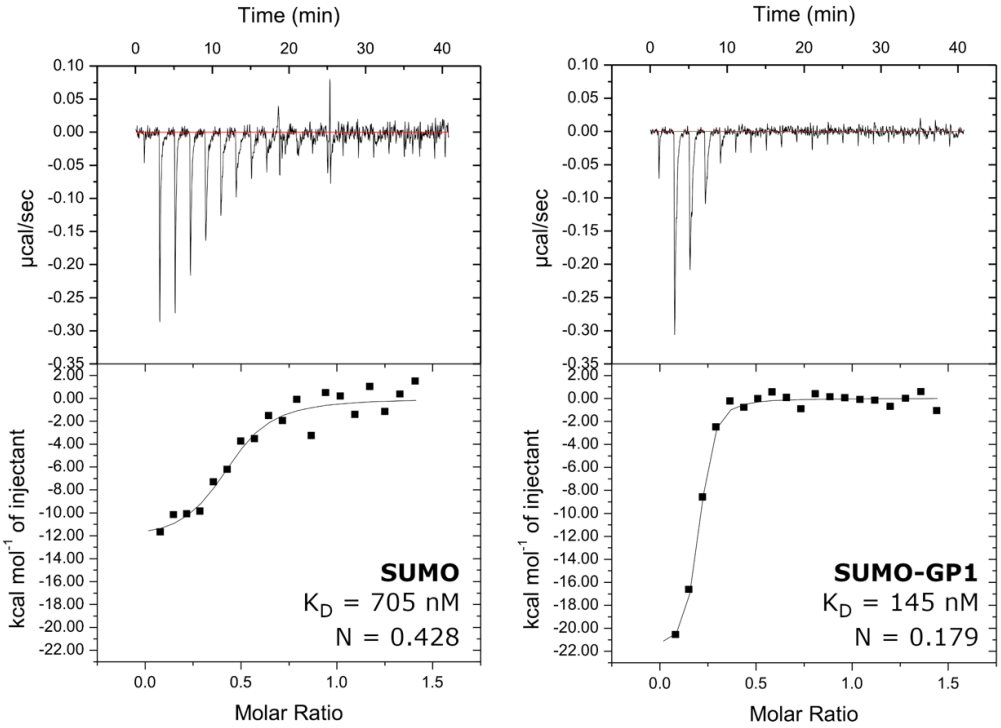
5.4. The structure of n190 in complex with Junín virus GP1

For structural characterisation of n190 in complex with a ligand, the interaction between n190 and glycoprotein 1 (GP1) from Junín virus (JUNV) was investigated. JUNV is an enveloped virus with a segmented, ambisense RNA genome. A member of the *Arenaviridae* family (reviewed in (Burri *et al.*, 2012; Grant *et al.*, 2012)), it is endemic to central Argentina, with infections occurring mainly among agricultural workers after exposure to bodily fluids or excretions from chronically infected rodents (Grant *et al.*, 2012). Most infections resolve after mild illness, but without treatment the case fatality rate can range between 15-30%, following progression of the disease to a neurologic-haemorrhagic phase (Enria *et al.*, 2008). A live attenuated vaccine (Candid #1) has been licensed for use in Argentina, though this suffers from the general issues associated with live attenuated vaccines described in Section 5.1.1 (McKee *et al.*, 1993).

A key feature of the VelcroVax platform is the ability to produce the target antigen separately from the scaffold, for example, in mammalian cells, to ensure the antigen is modified with a more appropriate glycosylation pattern. Therefore, as a viral glycoprotein, JUNV GP1 was deemed a suitable representative antigen for use with the VelcroVax system.

5.4.1. Preliminary analysis of the n190:SUMO-GP1 interaction

A recombinant version of GP1 linked at its N-terminus to a SUMO tag (SUMO-GP1) was designed and produced by Dr Thomas Bowden (Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford). The ability of n190 to capture SUMO-GP1 was then confirmed biochemically by KF. Further analysis by ITC indicated a binding affinity (K_D) of ~145 nm, although interestingly, stoichiometry (N) was determined as 0.179, indicating that only a fraction of Affimers were occupied (Figure 5.23).



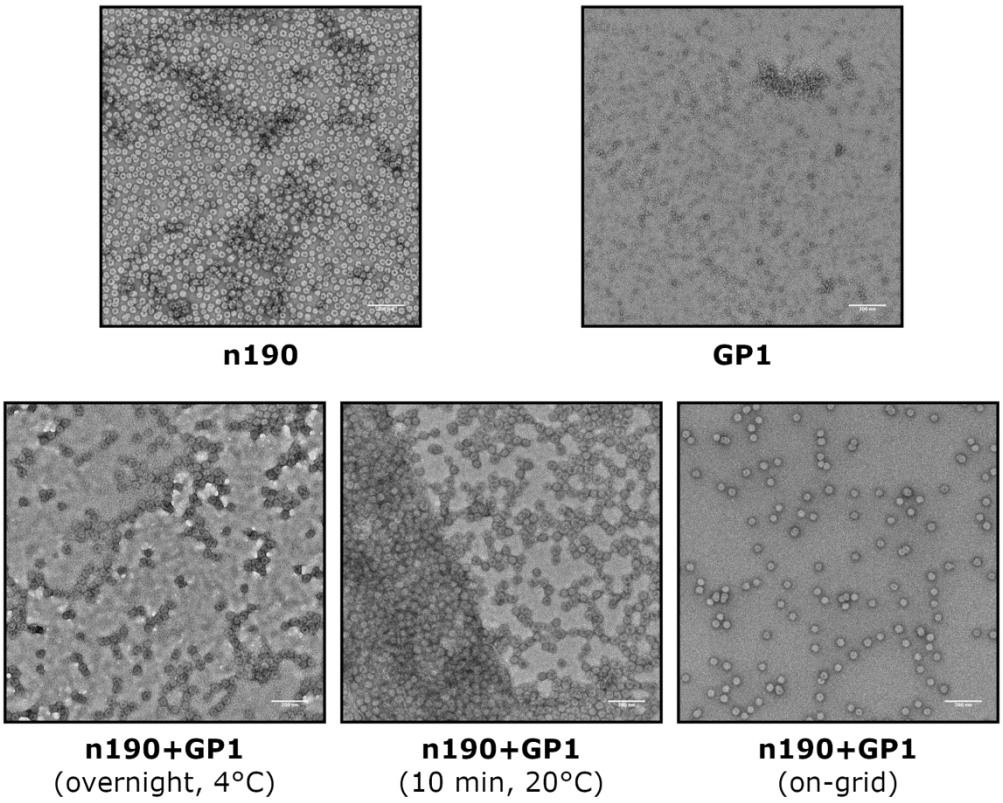
5.23. n190 binds to SUMO-tagged GP1 from Junín virus. *Isothermal titration calorimetry (ITC) analysis of the interactions of SUMO (left) and SUMO-GP1 (right) with n190. For each plot, the upper panel shows the heat*

changes resulting from successive injections of SUMO or SUMO-GP1 in the presence of n190. The lower panel shows the integrated heat plot with best fit line. K_D values were calculated as 705 nM (n190:SUMO) and 145 nM (n190:SUMO-GP1), stoichiometry (N) values were calculated as 0.428 (n190:SUMO) and 0.179 (n190:SUMO-GP1). Experiment performed and plots generated by KF.

Following this, the n190:SUMO-GP1 (termed n190-GP1) interaction was characterised by negative stain EM. Initial attempts at performing the interaction in solution led to aggregation of n190 (Figure 5.24). Aggregates were also seen in samples of SUMO-GP1 alone, although it was not clear if these were aggregates of SUMO-GP1 or of some contaminant. To circumvent the aggregation problem and progress to cryoEM, the interaction was performed 'on grid' rather than in solution. *I.e.,* n190 was applied directly to the grid surface, excess liquid wicked away, then SUMO-GP1 was applied on top. As n190 would already be immobilised by the carbon film backing, GP1 would be unable to drive any aggregation. This was confirmed in negative stain EM (Figure 5.24).

5.4.2. CryoEM data collection and processing

To prepare samples for cryoEM, n190 and SUMO-GP1 were diluted 1:2 in dH_2O before being applied to continuous ultra-thin carbon-coated lacey grids in separate applications (*i.e.,* n190 was applied, excess liquid wicked away after 30 seconds, then SUMO-GP1 was applied). A number of grids were vitrified using a range of blotting time (1 – 4 seconds) to maximise the chance of achieving optimal ice thickness. Following screening, one grid was taken forward for a 72-hour data collection on an FEI Titan Krios microscope (a complete set of cryoEM data collection parameters is given in Table 5.2).



5.24. Negative stain EM analysis shows GP1 aggregates n190. *Representative negative stain EM micrographs of n190 and GP1, alone or in combination. The n190-GP1 interactions were performed either in solution (overnight at 4°C or for 10 min at 20°C) or on-grid, as indicated. For the on-grid interaction, both n190 and GP1 were diluted 1:3 prior to their application to the grid. Scale bars indicate 200 nm.*

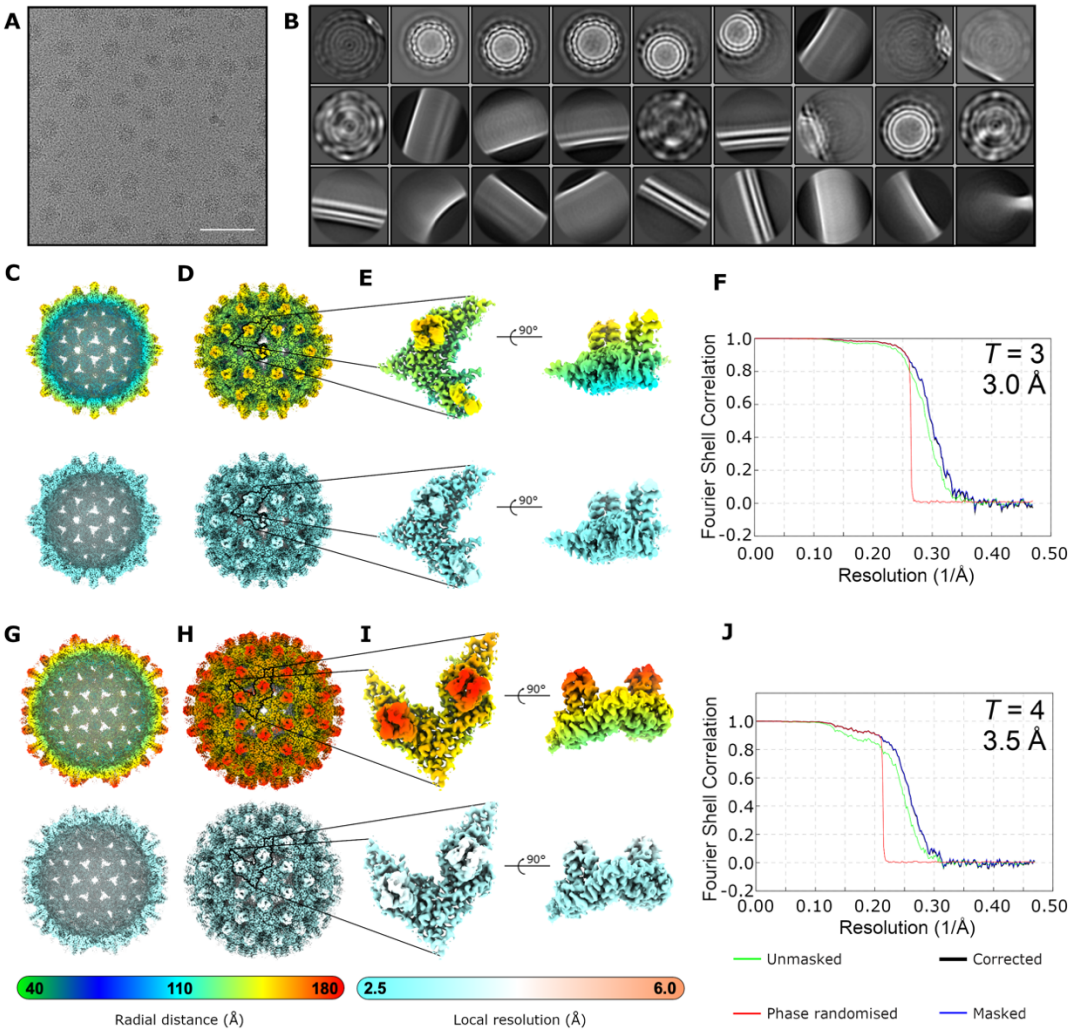
Image processing was performed in collaboration with an undergraduate student, Alexander Watson (AW). Overall, 23,966 n190-GP1 micrographs were collected (Figure 5.25A) yielding 1.5 million particles, of which ~299,000 corresponded to high-quality $T = 3$ VLPs and ~58,000 corresponded to high-quality $T = 4$ VLPs, judged by 2D classification. Other particles were either low-quality VLP particles or corresponded to contamination/background carbon. Notably, there was a striking difference in the ratio of T = 3 to $T = 4$ particles for the n190-GP1 dataset (~5:1) compared to the unliganded n190 dataset (~0.7:1). An excess of $T = 3$ particles was also observed in raw micrographs for n190-GP1, suggesting that the difference was not a result of bias introduced at the autopicking or classification stages of processing, but

was a true reflection of the sample, perhaps influenced by the specific sucrose gradient fraction that was selected during purification of n190.

Given the size of the initial particle stack, classification was performed after down-sampling the data by a factor of either 4 or 5 (Figure 5.25B), with re-extraction at the original pixel size only performed prior to 3D reconstruction with the computationally lighter, clarified particle stacks. Iterative rounds of CTF refinement were performed for both $T = 3$ and $T = 4$ VLPs, ultimately yielding 3.0 Å (Figure 5.25C-F) and 3.5 Å (Figure 5.25G-J) reconstructions. Quantitative parameters relating to n190-GP1 density maps are given in Table 5.5.

Table 5.5. Quantitative parameters related to n190-GP1 density maps.

Model	**n190-GP1 (*T* = 4)**	**n190-GP1 (*T* = 3)**
EMDB ID	*To be deposited*	*To be deposited*
CryoEM map processing		
Number of particles contributing to map	57,537	298,802
Map resolution (FSC = 0.143) (Å)	3.5	3.0
Map resolution range around atom positions (Å)	3.3 – 6.0	2.8 – 3.6
Map sharpening B factor (Å^2)	−223	−203



5.25. CryoEM reconstructions of n190 in complex with GP1. (A) *Representative micrograph from n190-GP1 data collection. Scale bar shows 100 nm.* ***(B)*** *Most populated classes from initial 2D classification of all auto-picked particles from the n190-GP1 dataset following five-fold down sampling.* ***(C-E)*** *Density map for T = 3 n190-GP1 filtered according to local resolution. This includes* ***(C)*** *a central cross section,* ***(D)*** *a complete isosurface representation (with asymmetric unit outlined) and* ***(E)*** *enlarged views of density for the asymmetric unit alone.* ***(F)*** *FSC plot for T = 3 n190-GP1.* ***(G-I)*** *Density map for T = 4 n190-GP1, filtered according to local resolution. This includes* ***(G)*** *a central cross section,* ***(H)*** *a complete isosurface representation (with asymmetric unit outlined) and* ***(I)*** *enlarged views of density for the asymmetric unit alone.* ***(J)*** *FSC plot for T = 4 n190-GP1. All maps were shown at 3 σ and coloured according to the radial*

colouring scheme or local resolution colouring scheme shown. Resolution values were determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected). n190-GP1 image processing was performed in collaboration with AW.

5.4.3. Structural analysis of the n190-GP1 cryoEM reconstruction

Intriguingly, neither the *T* = 3 nor the *T* = 4 reconstructions showed any density for SUMO-GP1, although Affimers were visible at low contour levels as for the unliganded n190 reconstructions. Visual inspection of n190 and n190-GP1 density maps gave no indication of any conformational changes to the HBcAg component of n190, although it appeared as though Affimer densities were closer together for n190-GP1, particularly for the *T* = 4 VLP (Figure 5.26). I hypothesised that SUMO-GP1 was bound to n190, but that (i) occupancy was sub-optimal, perhaps due to the combination of the on-grid binding approach used during grid preparation and potential steric exclusion blocking SUMO-GP1 binding to adjacent binding sites, and (ii) SUMO-GP1 was flexibly positioned. SUMO-GP1 itself was suspected to be inherently flexible as a result of the flexible linker connecting the SUMO tag to GP1. Atomic models predicted by AlphaFold 2 showed considerable variation in the relative positioning of GP1 and SUMO, with RMSD values between models ranging from ~20 Å to ~40 Å (Figure 5.27A). However, it is not clear whether this represents true flexibility between the domains or is just a result of a limited ability to predict the structure of the linker based on available structural data (Figure 5.27B). In the context of n190, SUMO-GP1 would be bound to an Affimer via a non-covalent interaction, with the Affimer itself connected to the rest of n190 by a flexible linker. Thus, while the average position of the Affimers had likely changed as a result of SUMO-GP1 binding and could be visualised, the signal from SUMO-GP1 itself was likely too weak to be visualised in the symmetrised reconstruction.

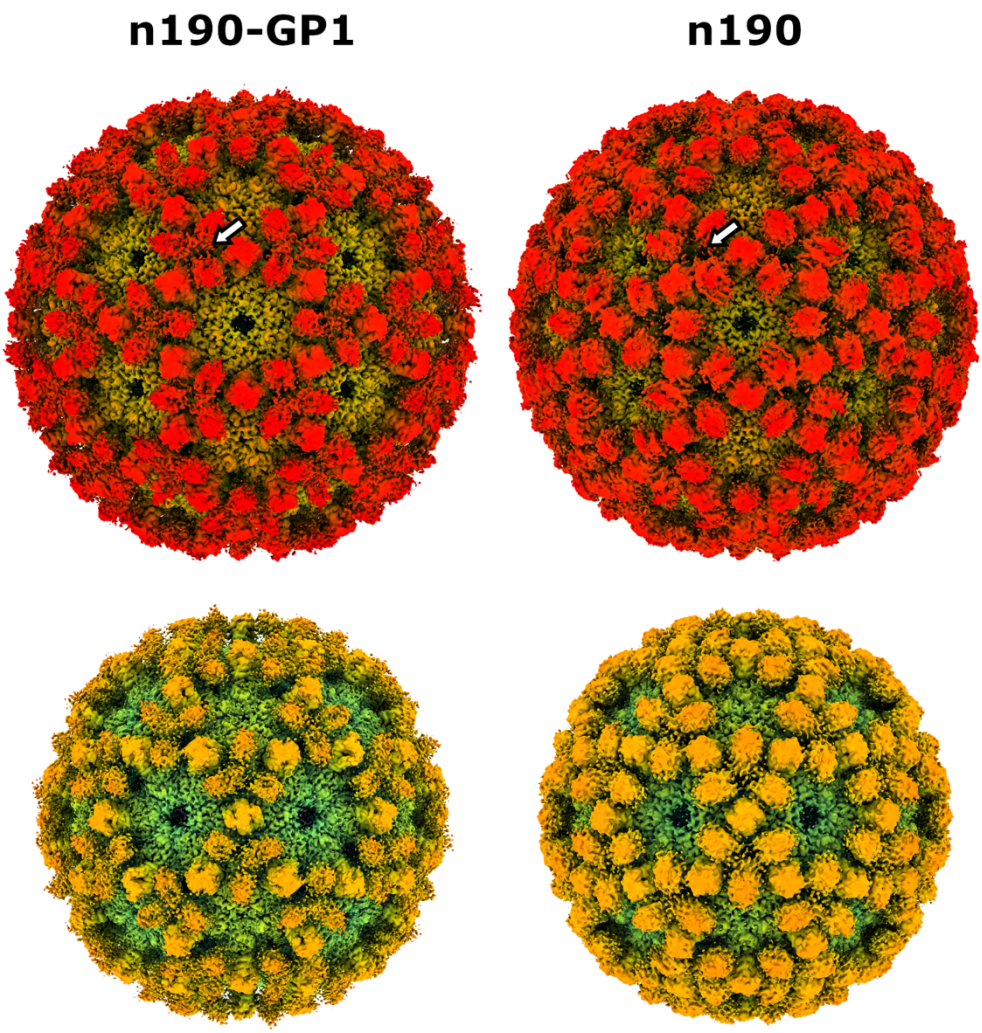
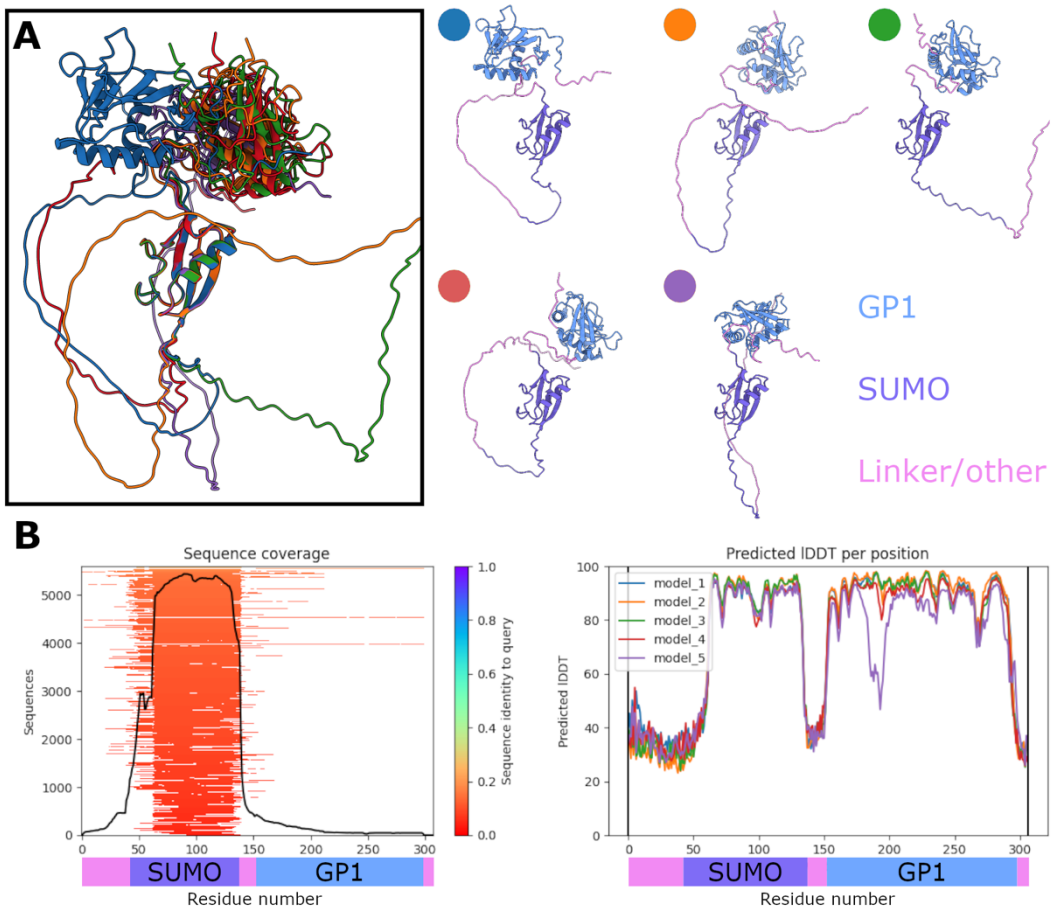


Figure 5.26. Affimers are closer together in n190-GP1. *Local resolution-filtered density maps for n190-GP1 (left) and n190 (right) in both T = 4 (above) and T = 3 (below) forms, all shown at ~2 σ and coloured radially. There is a subtle but clear change in Affimer positioning in the T = 4 VLPs, as indicated by the white arrow.*

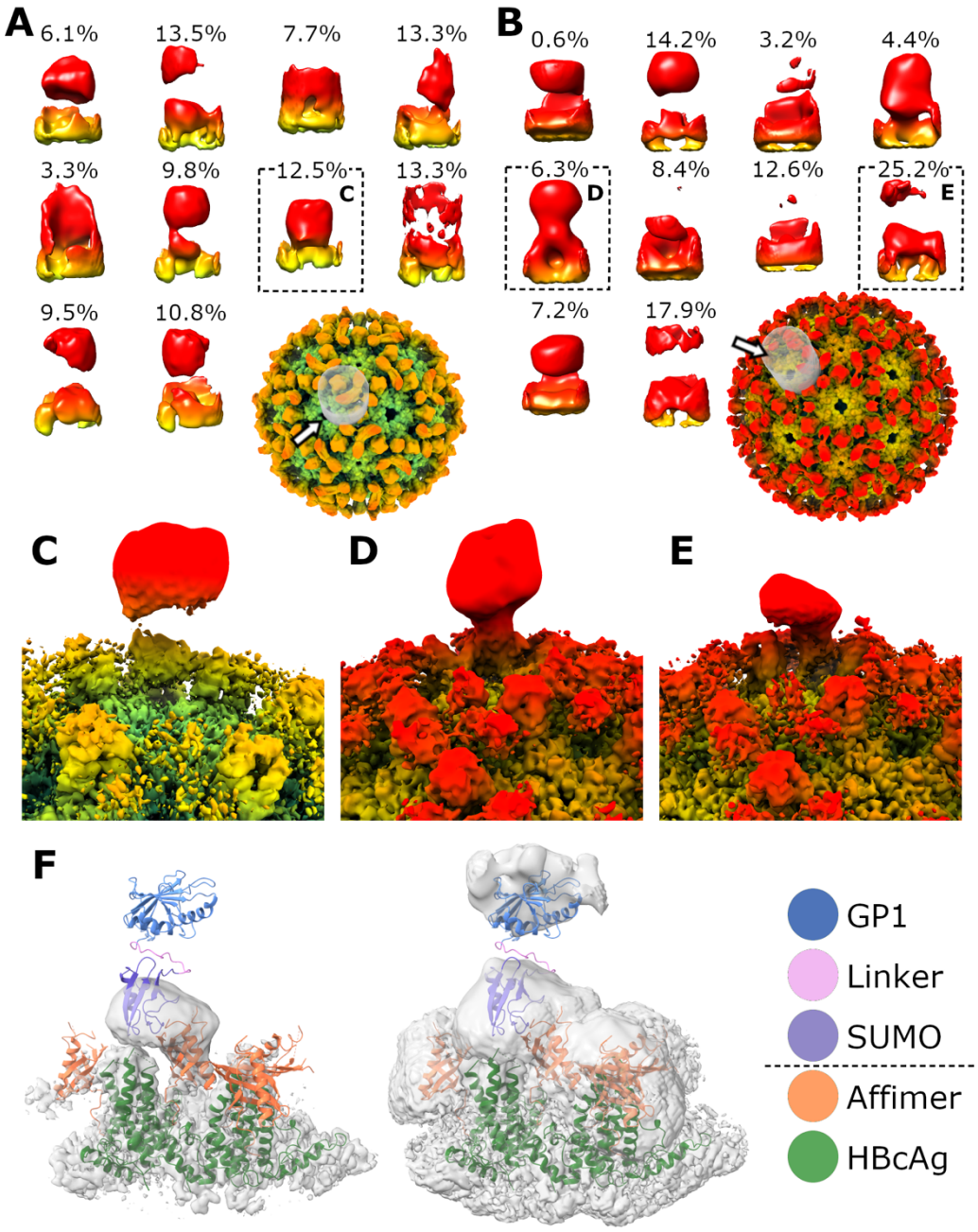


5.27. AlphaFold 2 predictions of SUMO-tagged GP1 structure. (A) *Five predicted models of SUMO-GP1 generated by AlphaFold 2, aligned based on the SUMO tag and coloured according to model number (left). Each model is also shown individually, coloured according to domain organisation (GP1 – blue, SUMO tag – violet, linker/other sequence – pink) (right).* ***(B)*** *Model analysis reported by AlphaFold 2 for the predicted models shown in (A), including 'depth' of the multi-sequence alignment stage by residue number (sequence coverage, left) and predicted local distance difference test (lDDT) value per residue, a metric of prediction confidence, for each model (right). Regions with a predicted lDDT greater than 90 are thought to be modelled to a high accuracy. Regions with a predicted lDDT below 70 are low confidence. Schematics below each graph contextualise the residue number within the domain organisation of SUMO-GP1.*

In an attempt to resolve additional density for SUMO-GP1, focussed classification was performed on 2× down-sampled data, using a mask centred

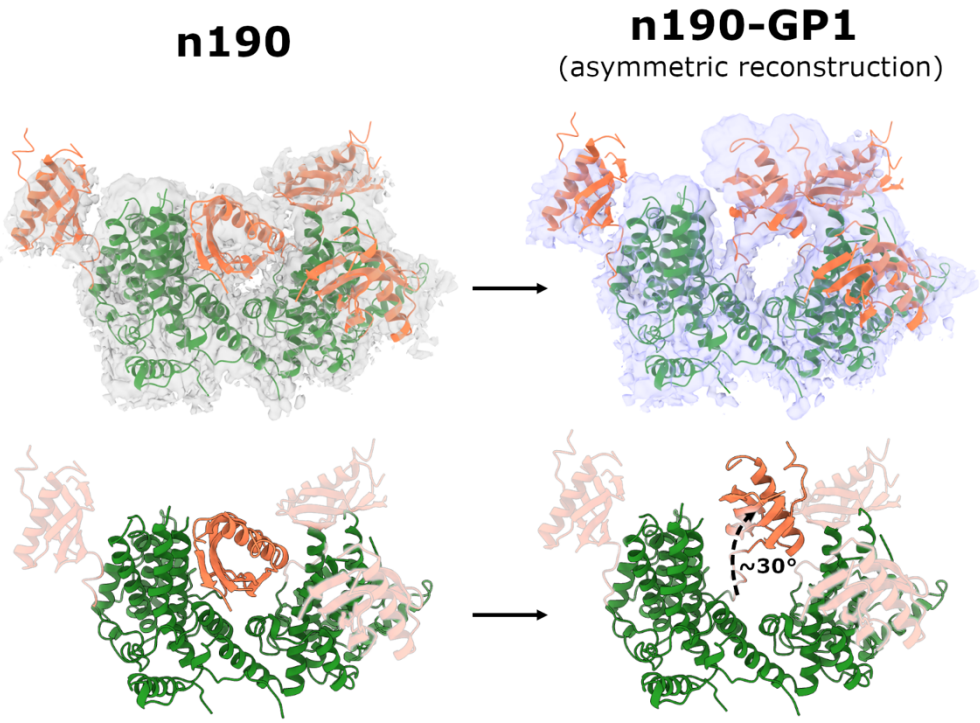
on an n190 Affimer and the region above this, for both *T* = 3 and *T* = 4 VLPs. Interestingly, both classifications revealed multiple classes with low-resolution density above and, in several cases, continuous with, Affimer density (Figure 5.28A,B). Full asymmetric reconstructions using data from some of these classes gave clear evidence of ligand binding (Figure 5.28C-E), indicating that a subpopulation of Affimers had captured SUMO-GP1, despite the absence of SUMO-GP1 density in the symmetrised reconstruction. Manual fitting of one of the predicted models generated for SUMO-GP1 into the additional density at a higher contour level (showing only stronger density) revealed that the additional density was approximately the correct size for a single SUMO tag (Figure 5.28F). At a lower contour level (showing weaker density), disjointed and low-resolution density also become apparent, separated from the SUMO tag density by a distance consistent with the length of the flexible linker sequence between SUMO and GP1 (Figure 5.28F). Based on these results, I concluded that the flexibility of the Affimer:SUMO-GP1 axis contributed to the lack of SUMO-GP1 density in the symmetrised reconstruction. As expected, the flexibility of each component in sequence compounded the effect moving distally away from the surface of the VLP, such that GP1 density was the weakest and least well resolved.

Interestingly, asymmetric reconstructions also indicated that Affimers involved in the interaction with SUMO-GP1 had reoriented, with the binding loops tilted approximately 30° upwards (Figure 5.29). It is likely that this reorientation was necessary to accommodate the ligand binding. Despite this, asymmetric reconstruction suggested that even after reorientation, there would be insufficient space for adjacent Affimers to capture SUMO-GP1.



5.28. Focussed classification of n190-GP1 reveals binding. (A,B) *All focussed classes from focussed classification of* ***(A)*** *T = 3 n190-GP1 and* ***(B)*** *T = 4 n190-GP1, with the proportion of subparticles assigned to each class indicated. The position of the mask (grey) is indicated on reference maps for each case. Classes are coloured according to radial distance (relative to the complete VLP) and are oriented from the viewpoint indicated by the white arrow on each reference model.* ***(C-E)*** *Asymmetric reconstructions from particles contained in the classes highlighted by the dashed boxes in (A-B), filtered by local resolution.* ***(F)*** *Atomic model for the*

asymmetric unit of T = 4 n190 (green and orange) overlaid with density from the asymmetric reconstruction shown in (E). A model of SUMO-GP1 (model_2) generated by AlphaFold 2 was trimmed to include only the SUMO tag (purple), GP1 (blue) and the linker (pink) between them, and manually positioned within unoccupied density. Density is shown at high (left) and low (right) contour levels.

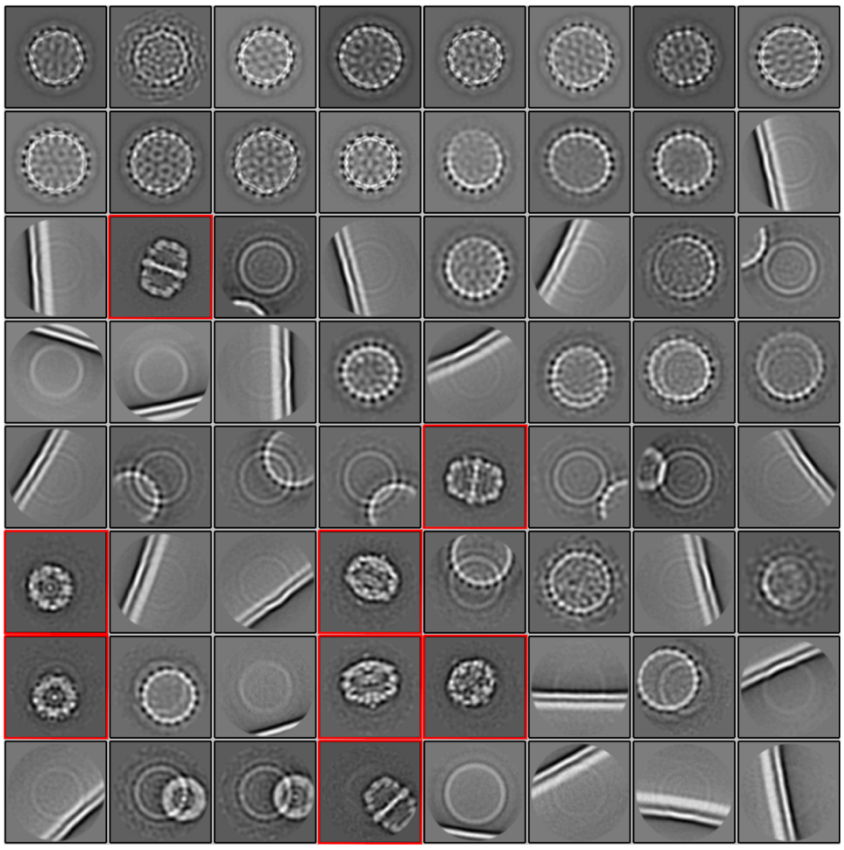


5.29. GP1 binding is associated with reorientation of n190 Affimers. *Atomic model for the T = 4 asymmetric unit of n190 before and after SUMO-GP1 binding. To generate the atomic model for n190-GP1, Affimers were manually repositioned into density for the full asymmetric reconstruction derived from a single focussed class of T = 4 n190-GP1, and refined to correct any obvious geometrical inconsistencies. Models are shown overlaid with local resolution-filtered cryoEM density (~2 σ) within 4 Å of atomic coordinates (upper row), or with Affimers not involved in the interaction faded out (lower row). The angle of tilt is indicated. Green – HBcAg component of n190; orange – Affimer/linker component of n190.*

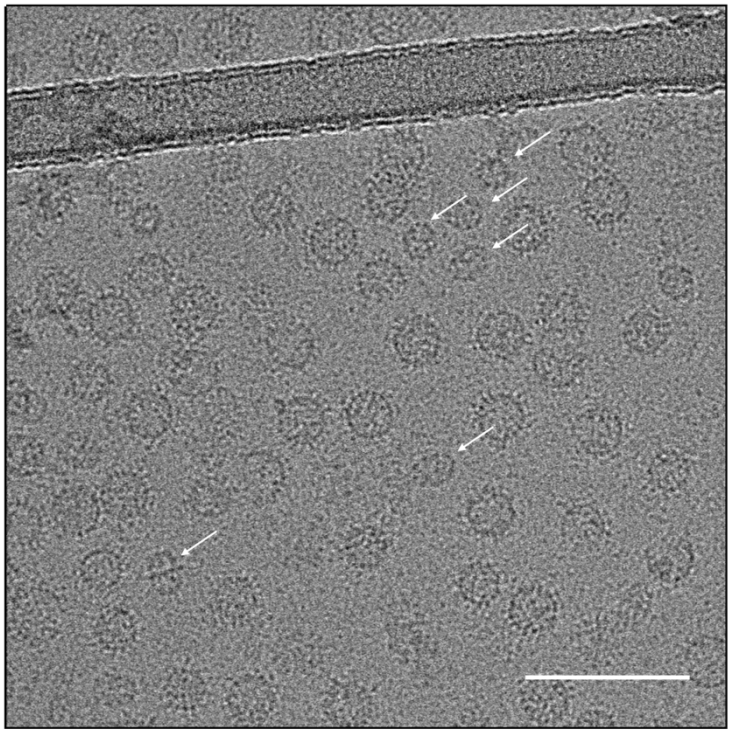
5.5. The importance of considering 'contaminant' data – the structure of *P. pastoris* fatty acid synthase

5.5.1. Identification of fatty acid synthase in an AR1 cryoEM dataset

Interestingly, the initial 2D classification performed for the unliganded AR1 dataset (described in Section 5.2.2) revealed a number of distinctly non-VLP-like classes (Figure 5.30). These were recognised as corresponding to yeast fatty acid synthase (FAS) due to their characteristic morphology, and individual FAS particles were subsequently identified in raw micrographs (Figure 5.31). No direct interactions between FAS and AR1 could be seen in 2D class averages and FAS particles were well separated from AR1 in raw micrographs, suggesting that the FAS contamination was likely a result of co-sedimentation rather than a specific interaction.



5.30. Identification of FAS in 2D classes from AR1 dataset. *Following early processing steps and autopicking for AR1 particles, all extracted particles were subjected to 2D classification. Representative classes, as shown above, included classes of yeast FAS (outlined in red).*



5.31. Presence of FAS in raw micrographs from AR1 dataset. *Representative micrograph from the AR1 dataset. FAS particles can be identified alongside (but separated from) AR1 VLPs. Example FAS particles are indicated by white arrows. Scale bar shows 100 nm. (Figure adapted from* (Snowden *et al.*, 2021)*, published under a CC BY 4.0 licence.)*

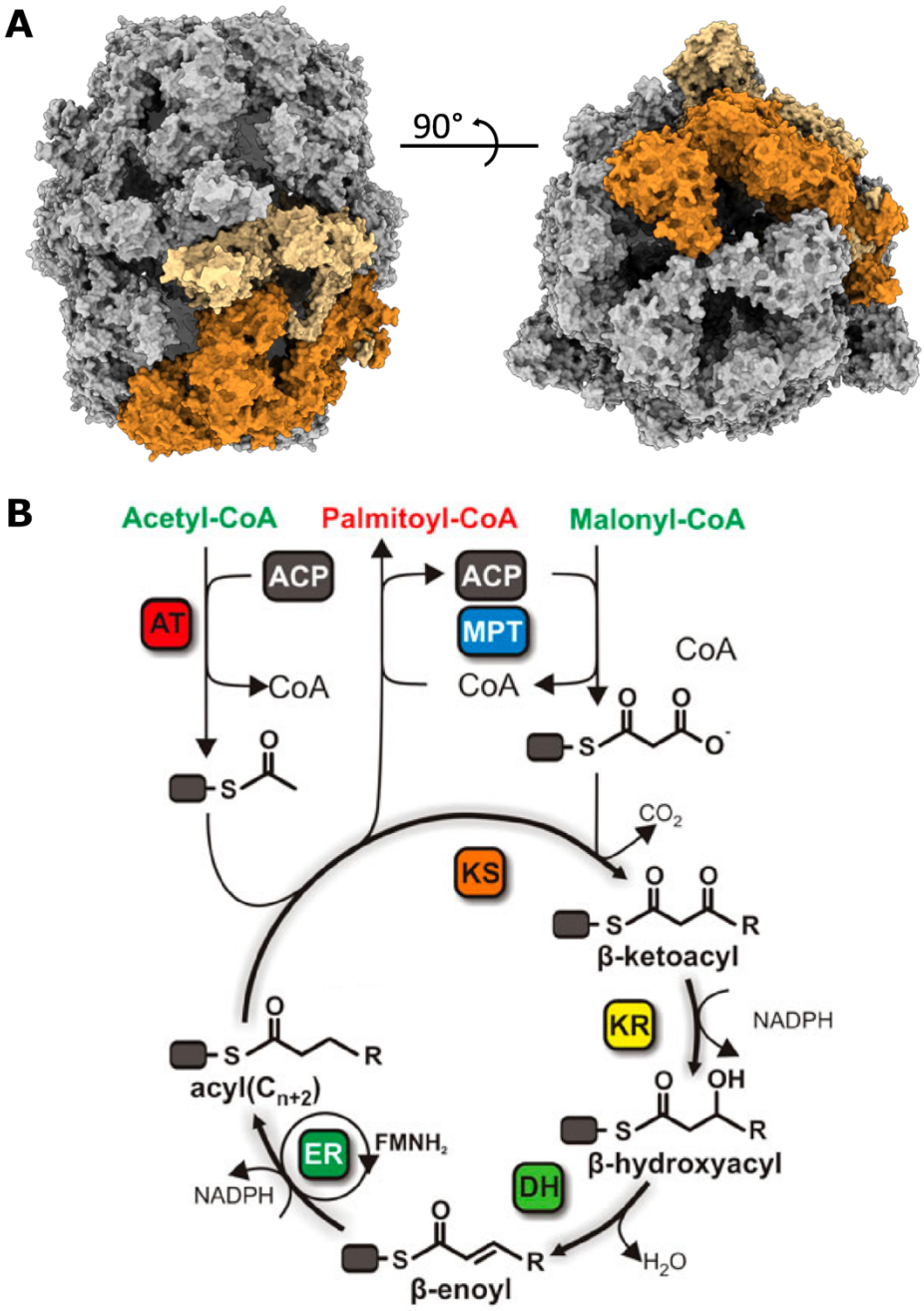
5.5.2. The physiological importance of FAS and its industrial potential

While yeast FAS was identified here as a contaminant in the AR1 dataset, it is a critical physiological enzyme complex in its own right, catalysing the synthesis of long-chain fatty acids important for the synthesis of membrane bilayers, protein lipid anchors and various signalling compounds (Tehlivets *et al.*, 2007). FAS is also a common target of bioengineering approaches to generate short chain fatty acids for use as chemical feedstock in industry (Beld *et al.*, 2015; Gajewski *et al.*, 2017a, 2017b; Zhu *et al.*, 2017; Heil *et al.*, 2019).

Yeast FASs are hetero-dodecameric complexes with an $\alpha_6\beta_6$ configuration (Figure 5.32A). The α and β subunits each contain several enzymatic domains, most of which face the hollow interior of the two domes on either side of a

central platform (Stoops *et al.*, 1992; Jenni *et al.*, 2006, 2007; Lomakin *et al.*, 2007; Gipson *et al.*, 2010). The acetyl transferase (AT) domain activates acetyl CoA to initiate fatty acid synthesis, before the substrate is acted on in turn by the ketoacyl synthase (KS), ketoacyl reductase (KR), dehydratase (DH) and enoyl reductase (ER) domains in an iterative cycle (Figure 5.32B) (Maier *et al.*, 2010). During this cycle, the growing fatty acid chain is carried between catalytic centres by the mobile acyl carrier protein (ACP) domain, which is itself connected to the rest of the α subunit by two long, flexible linkers (Gipson *et al.*, 2010; Crosby and Crump, 2012; Lou and Mazhab-Jafari, 2020). Upon reaching a certain chain length (usually C_{16}), the nascent fatty acid chain is released by the bifunctional malonyl/palmitoyl transferase (MPT) domain. The phosphopantetheine transferase (PPT) domain is unusual in that it is located on the exterior of the complex and does not contribute directly to the fatty acid elongation cycle. Instead, it catalyses the addition of a phosphopantetheinyl group to the active serine residue of the ACP domain, essential for FAS activity (Fichtlscherer *et al.*, 2000; Lomakin *et al.*, 2007; Johansson *et al.*, 2009).

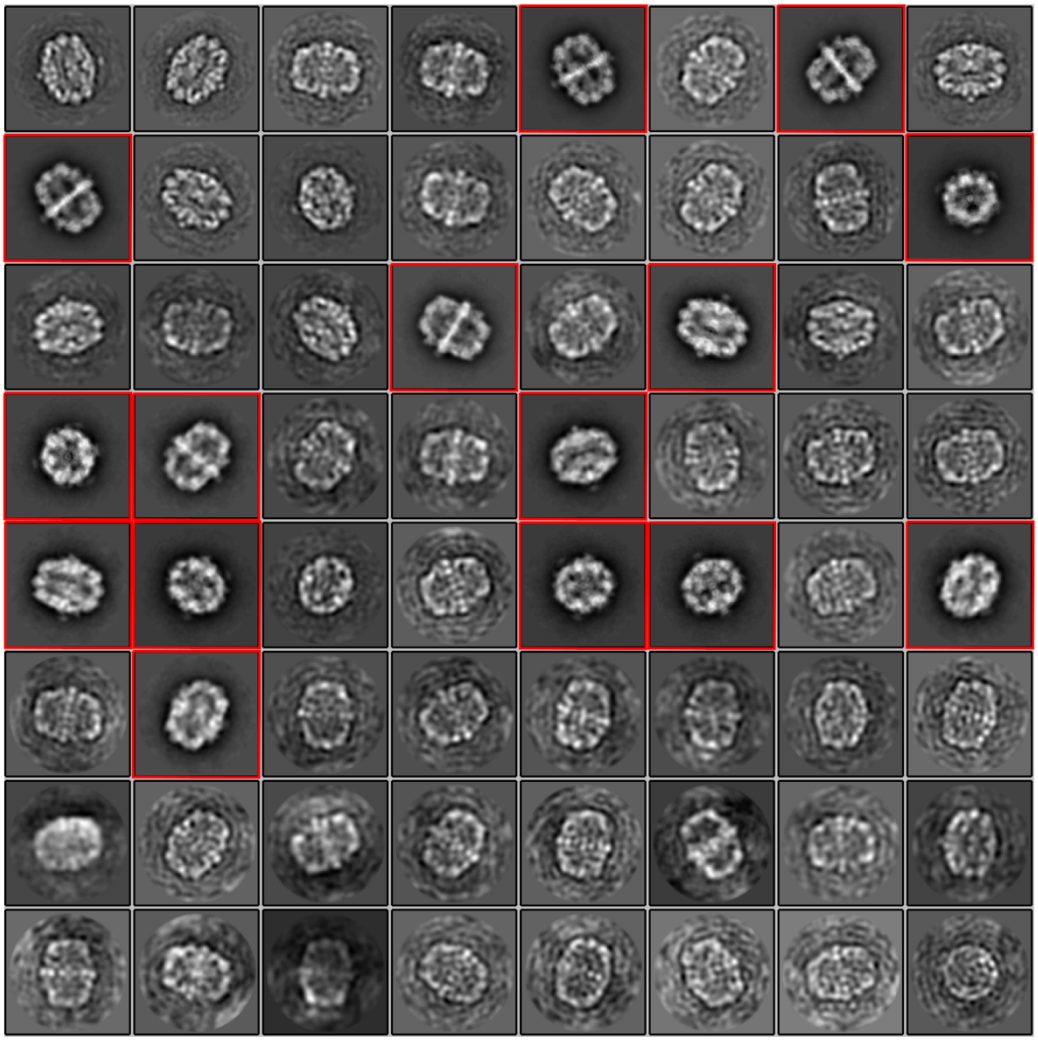
While there are many molecular structures reported for yeast FASs (Stoops *et al.*, 1992; Jenni *et al.*, 2006, 2007; Lomakin *et al.*, 2007; Gipson *et al.*, 2010; D'Imprima *et al.*, 2019; Lou *et al.*, 2019; Joppe *et al.*, 2020; Singh *et al.*, 2020), no structures of *P. pastoris* FAS were available at the time of this study. *P. pastoris* in particular is used extensively for recombinant protein production, and has also been applied to metabolic engineering in recent years (Ahmad *et al.*, 2014; Yang and Zhang, 2018). While *S. cerevisiae* is the prototypical yeast species used in most FAS engineering work, a major advantage of *P. pastoris* over *S. cerevisiae* is its relatively low production of ethanol, making the maintenance of productive, high-cell density fermenter cultures more feasible (Cereghino and Cregg, 2000). As such, specific structural information on *P. pastoris* FAS is likely to be particularly useful for future efforts to engineer yeast FAS.



5.32. Yeast FAS structure and function. (A) *Atomic model of S. cerevisiae FAS (PDB: 6TA1) showing the subunit organisation of fungal FAS (Joppe et al., 2020). One α subunit is shown in light orange, one β subunit is shown in dark orange.* ***(B)*** *Schematic detailing the catalytic cycle of yeast FAS. ACP – acyl carrier protein; AT – acetyl transferase; MPT – malonyl/palmitoyl transferase; KS – ketoacyl synthase; KR – ketoacyl reductase; DH – dehydratase; ER – enoyl reductase. Schematic adapted from (Maier et al., 2006).*

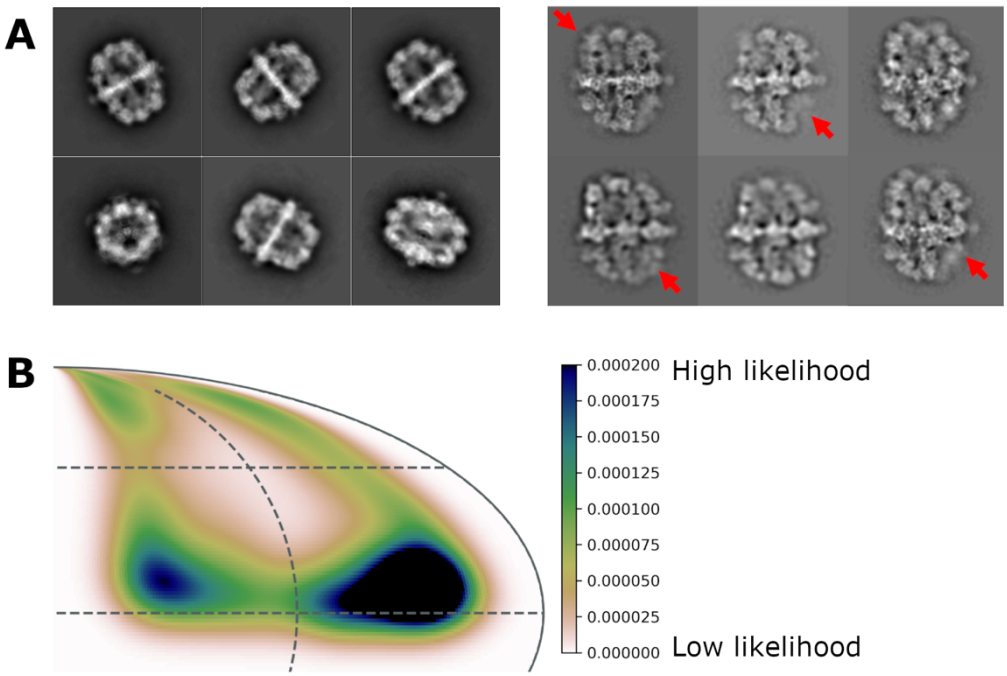
5.5.3. CryoEM structure determination of *P. pastoris* FAS

Given the potential utility of structural information on yeast FAS, the decision was made to process the 'contaminant' FAS particles to a high-resolution structure. Autopicking was repeated, this time using FAS 2D class averages as templates, resulting in a stack of ~550,000 particles, of which ~160,000 corresponded to actual FAS particles (rather than inadvertently picked VLPs or thick carbon edges). Two rounds of 2D classification yielded ~37,000 particles from high-quality classes (Figure 5.33). Most of the discarded particles were excluded because they were picked over thick carbon.



5.33. Second round of FAS 2D classification. *Classes taken forward for further processing are outlined in red.*

Interestingly, unlike a previously reported structure of *S. cerevisiae* FAS (D'Imprima *et al.*, 2019), we did not observe any indications of denaturation in 2D class averages (Figure 5.34A), though it is difficult to give a confident assessment due to the large number of FAS particles that were discarded for the reason above. The use of a continuous thin carbon film here may have sequestered particles away from the air-water interface by adsorption, protecting them from denaturation. Using support films can sometimes introduce problems related to adoption of a preferred orientation by the target particle, although while there were some indications of preferential orientation effects here, a reasonable coverage of potential orientations within the D3 asymmetric unit was achieved (Figure 5.34B).

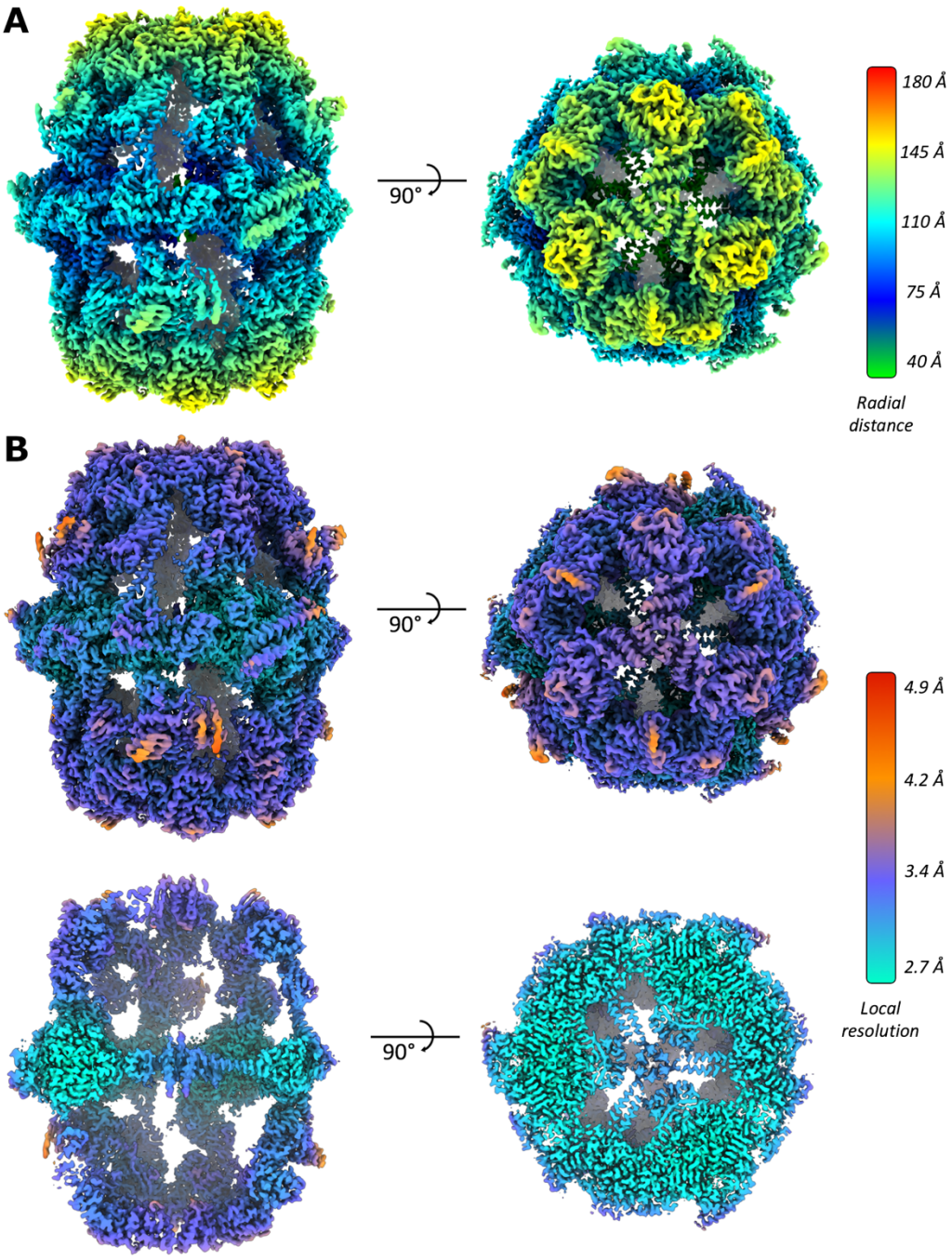


5.34. FAS particle quality. (A) *Representative 2D class averages of P. pastoris FAS reported here (left) and S. cerevisiae FAS (right), showing signs of partial denaturation indicated by red arrows (D'Imprima et al., 2019) (licensed under CC BY 4.0).* ***(B)*** *Distribution map of orientations assigned to particles contributing to the final reconstruction of the FAS asymmetric unit. Scale indicates the likelihood that any particle in was assigned the orientation indicated on the map. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

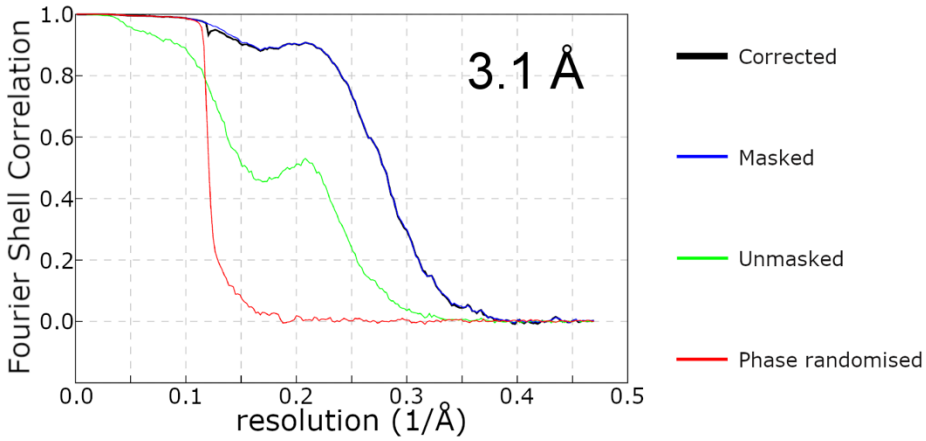
The final stack of ~37,000 particles was used for 3D refinement with D3 symmetry imposed, based on a *de novo* initial model. This led to a 3.6 Å reconstruction following map sharpening. Three iterative cycles of CTF refinement, Bayesian polishing and 3D refinement with a mask and solvent-flattened FSCs were performed, each time improving the resolution of the density to a final global resolution of 3.1 Å (Figure 5.35A, Figure 5.36). Resolution was highest in the central platform, and worsened in the more distal parts of the complex (range 2.7 – 5.4 Å) (Figure 5.35B).

5.5.4. An atomic model for *P. pastoris* FAS

To build an atomic model for *P. pastoris* FAS, a homology model was generated using SWISS-MODEL (Waterhouse *et al.*, 2018) and fitted to the available density. Unresolved regions were removed from the atomic model, before the fit-to-density was visually inspected and corrected in Coot. The model was then symmetrised and refined in Phenix to improve its fit-to-density and geometry. Iterative cycles of visual inspection/correction and refinement were performed to improve the quality of the atomic model. Validation statistics for the FAS atomic model are given in Table 5.6.



5.35. CryoEM reconstruction of P. pastoris FAS at 3.1 Å resolution. *Isosurface representations of the cryoEM reconstruction of P. pastoris FAS, shown from two different orientations and coloured by* ***(A)*** *radial distance, or* ***(B)*** *local resolution, according to the scales indicated. A central cross-section, highlighting the central platform, is also shown for (B). (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

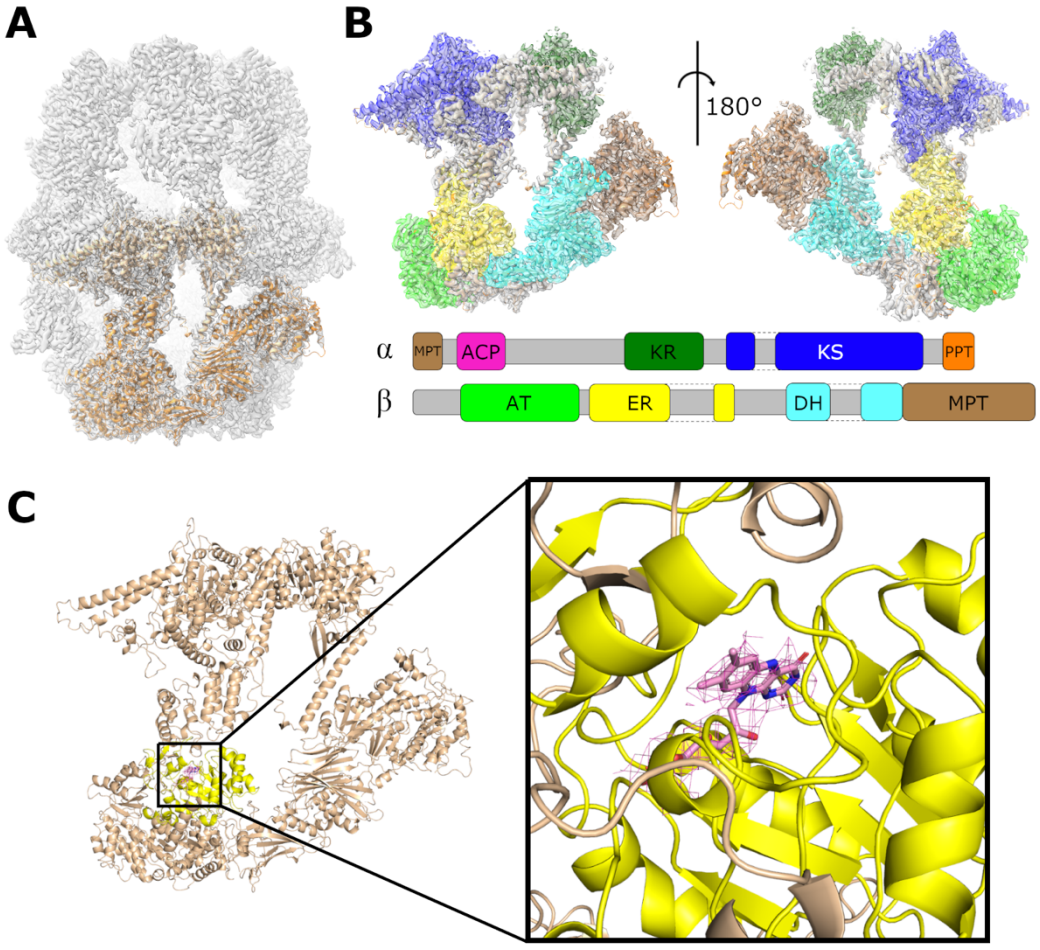


5.36. FSC plot for P. pastoris FAS density map. *Map resolution (3.1 Å) was determined using the FSC = 0.143 criterion with high-resolution noise substitution to correct for any overfitting (black line, corrected). (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

Table 5.6. Map and model validation statistics for P. pastoris FAS.

Model	**FAS**	**FAS ACP domain**
EMDB ID	**12138**	**12139**
PDB ID	**7BC4**	**7BC5**
CryoEM map processing		
Number of particles contributing to map	37,054	N/A
Map resolution (FSC = 0.143) (Å)	3.1	N/A
Map resolution range around atom positions (Å)	2.7 – 5.4	N/A
Map sharpening B factor (Å^2)	−112	N/A
Residues modelled		
Subunit α	1-95, 323-534, 599-1751	139-299
Subunit β	10-2063	N/A
RMSD		
Bond lengths (Å)	0.0082	0.0055
Bond angles (°)	1.20	1.32
Validation		
All-atom clashscore	5.38	7.83
MolProbity score	1.77	1.92
Rotamer outliers (%)	1.19	0.76
Ramachandran plot		
Favoured (%)	93.61	91.82
Allowed (%)	6.39	8.18
Outliers (%)	0.00	0.00

In addition to many of the enzymatic and structural domains of FAS (including AT, KR, KS, DH, ER and MPT domains), density was also observed for a ligand bound within the ER domain active site, likely a molecule of FMN (Figure 5.37). While the β subunit was essentially complete (only missing residues 1-9 and 2064-2069), the α subunit was missing several segments (residues 96-322, 535-598 and 1752-1879), including the acyl carrier protein (ACP) and phosphopantetheinyl transferase (PPT) domains. In an attempt to resolve some of these 'missing regions', 3D focussed classifications were performed with masks applied to the areas of the map expected to contain ACP and PPT density.



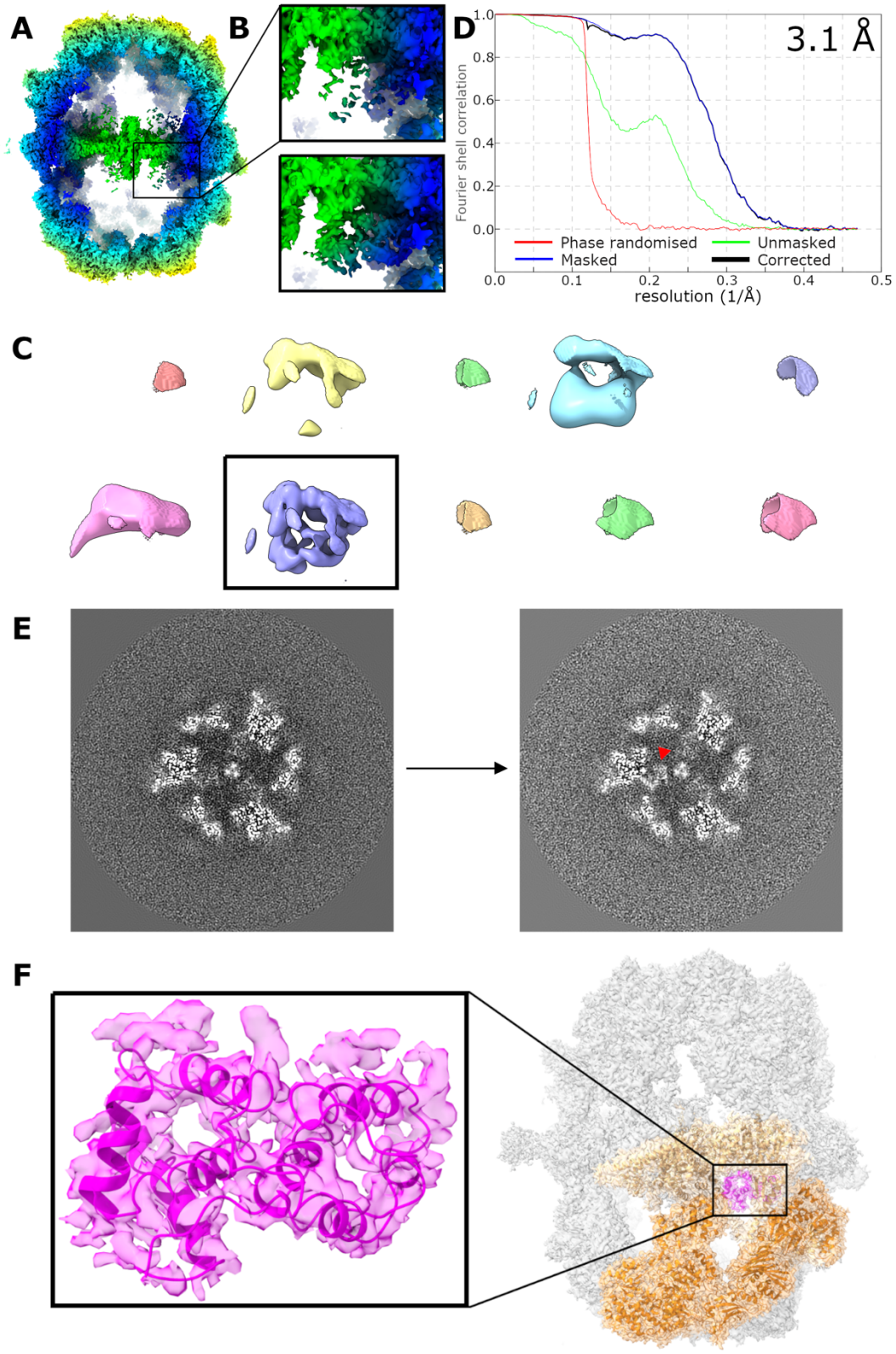
5.37. An atomic model for P. pastoris FAS. (A) *The atomic model for an asymmetric unit of FAS fitted into the full density map. The α subunit is coloured light orange, the β subunit is coloured dark orange.* ***(B)*** *Enlarged image of the asymmetric unit of FAS. The atomic model is fitted into density*

coloured based on enzymatic domain, according to the domain schematic for FAS α and β subunits shown below. KS – ketoacyl synthase, KR – ketoacyl reductase, AT – acetyl transferase, ER – enoylreductase, DH – dehydratase, MPT – malonyl/palmitoyl transferase. ***(C)*** *Density for a ligand bound in the ER domain active site. The ER domain is coloured in yellow. FMN is shown in pink, with corresponding density overlaid. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

5.5.5. Resolving the mobile ACP domain by focussed classification

In contrast to much of the FAS structure, density for the mobile ACP domain was poor quality, fragmented and only visible at low contour levels, meaning it was insufficient for atomic modelling (Figure 5.38A,B). This was somewhat unsurprising given the mobile nature of the ACP domain. Particles contributing to the D3 symmetric reconstruction of FAS were subjected to symmetry expansion and focussed classification with a cylindrical mask over the ACP domain-containing region of the complex interior. Initially, a regularisation parameter ('T' number) of 20 was used, but no well-resolved ACP domain-containing classes were identified. Upon increasing the regularisation parameter to 40 (i.e., placing a higher relative weighting on the experimental data than on the reference map), one class containing ACP domain density with a reasonable level of structural detail was identified (Figure 5.38C).

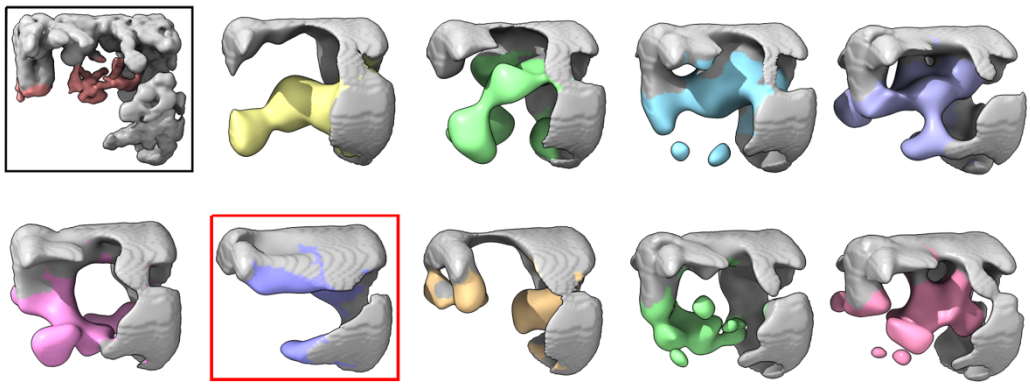
Particles from this class were randomly divided into two subgroups which were reconstructed independently and sharpened, yielding a 3.1-Å resolution map (Figure 5.38D). This map showed a clear improvement in ACP domain density, with a continuous backbone trace and quality sufficient for atomic modelling (Figure 5.38B,E). The ACP domain of the *P. pastoris* FAS homology model was rigid-body fitted, then iteratively refined to improve fit-to-density and atomic geometry (Figure 5.38F). Interestingly, the ACP domain was positioned adjacent to the KS domain with its active serine (S178) located ~18 Å away from the active site, in agreement with the length of the phosphopantetheine arm (~18 Å) that tethers the growing fatty acid chain (Leibundgut *et al.*, 2007).



5.38. Focussed classification to resolve the ACP domain of FAS. (A) *Clipped view of the sharpened FAS density map, coloured according to radius. The box highlights the region of the complex interior selected for*

focussed 3D classification. ***(B)*** *Enlarged image of the region depicted by the box in (A) (upper) compared with the sharpened, asymmetric reconstruction of FAS derived from the ACP density-containing class identified by the black solid box in (C) (lower).* ***(C)*** *All focussed classes from the initial ACP domain focussed classification, shown at the same contour level.* ***(D)*** *Fourier shell correlation (FSC) plot for the asymmetric reconstruction of FAS-ACP derived from the focussed class containing improved ACP density.* ***(E)*** *Slice through the sharpened D3-symmetric density map of FAS at the level of KS and ACP domains (left) compared with the equivalent slice through the sharpened, asymmetric reconstruction of FAS following focussed classification of ACP domain density (right). The red arrowhead indicates the position of the improved ACP domain density.* ***(F)*** *Atomic coordinates for the ACP domain of FAS fitted into ACP density from the map shown in (B, lower). ACP is also shown in the context of the full FAS complex. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

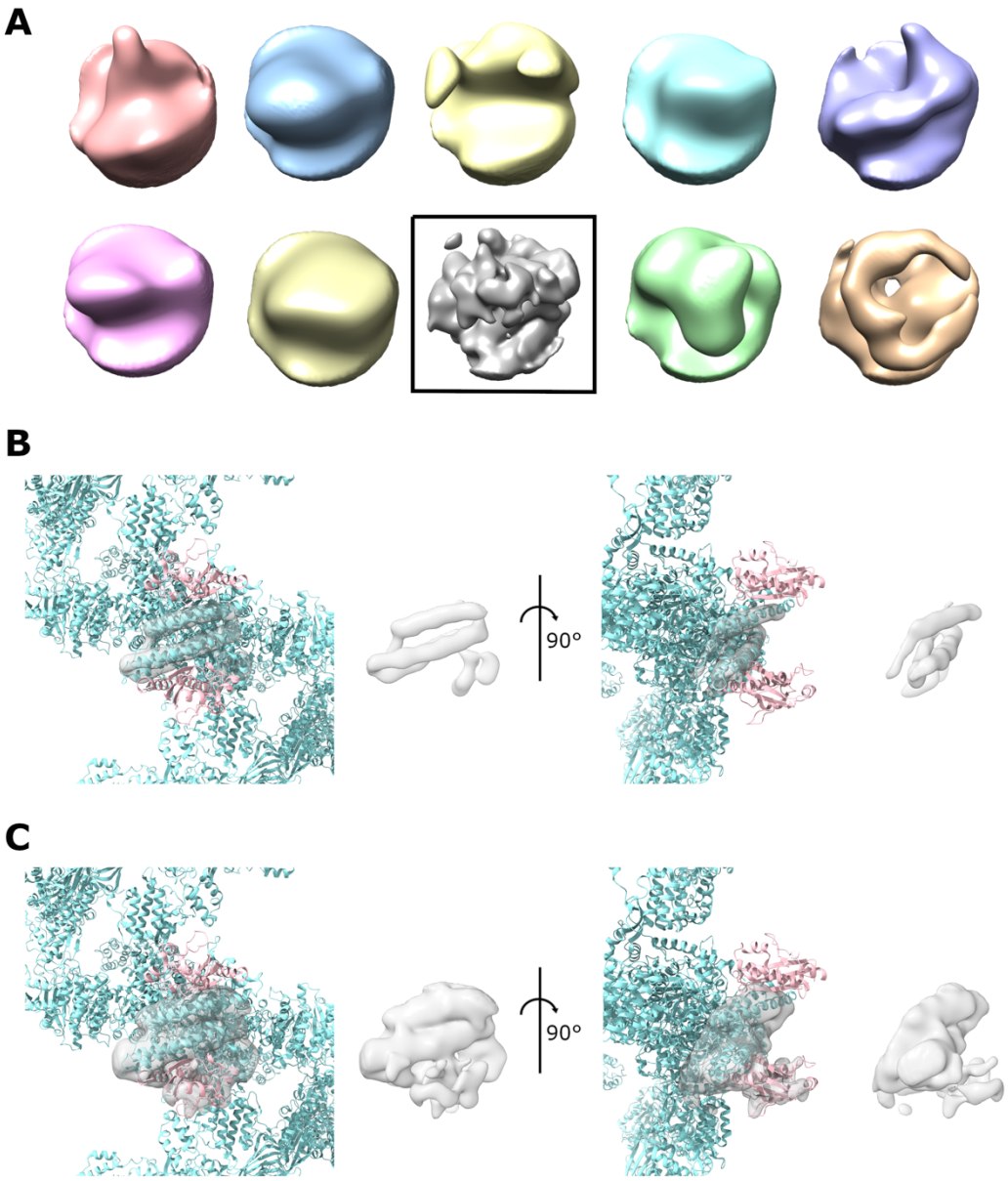
Further examination of all ACP domain focussed classes at a low contour level revealed that some classes contained additional density at the edge of the masked region. Speculating that these densities may correspond to alternative positions of the ACP domain (which naturally migrates between different catalytic centres), focussed classification was repeated using a larger mask. Several classes were identified containing additional density of the approximate size expected for an ACP domain in alternative locations (Figure 5.39), however these classes contained a relatively small number of subparticles suggesting that most ACP domains were localised to the KS domain. Classes with no apparent density for an ACP domain were also identified, suggesting denaturation or loss of the ACP domain from a subpopulation of reaction chambers (Figure 5.39).



5.39. Potential alternative ACP domain locations. *All classes from the second ('expanded') focussed 3D classification are shown. Grey regions correspond to density accounted for by non-ACP components of the FAS atomic model (i.e., the outer wall and central platform that surround the ACP domain-containing interior chamber). All classes are shown at the same contour level. The class indicated by the black box shows the ACP domain localised to the KS domain. The class indicated by the red box has no additional density for an ACP domain. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

5.5.6. Focussed classification to discern the PPT domain

Focussed classification was also performed using a mask over the region expected to contain the PPT domain. This domain is often missing from FAS structures, probably a result of partial denaturation during purification (D'Imprima *et al.*, 2019; Joppe *et al.*, 2020). Most classes were very poorly resolved and showed no convincing additional density (Figure 5.40A). However, one class showed additional density with features reminiscent of the PPT domain. While full asymmetric reconstruction of the particles from this class did not provide any improvement in the density, such that atomic modelling was not possible, the additional density from focussed classification overlaid very well with the PPT domain in an atomic model of *S. cerevisiae* FAS (PDB: 6TA1), suggesting that a subpopulation of PPT domains were likely to be intact (Figure 5.40B,C).



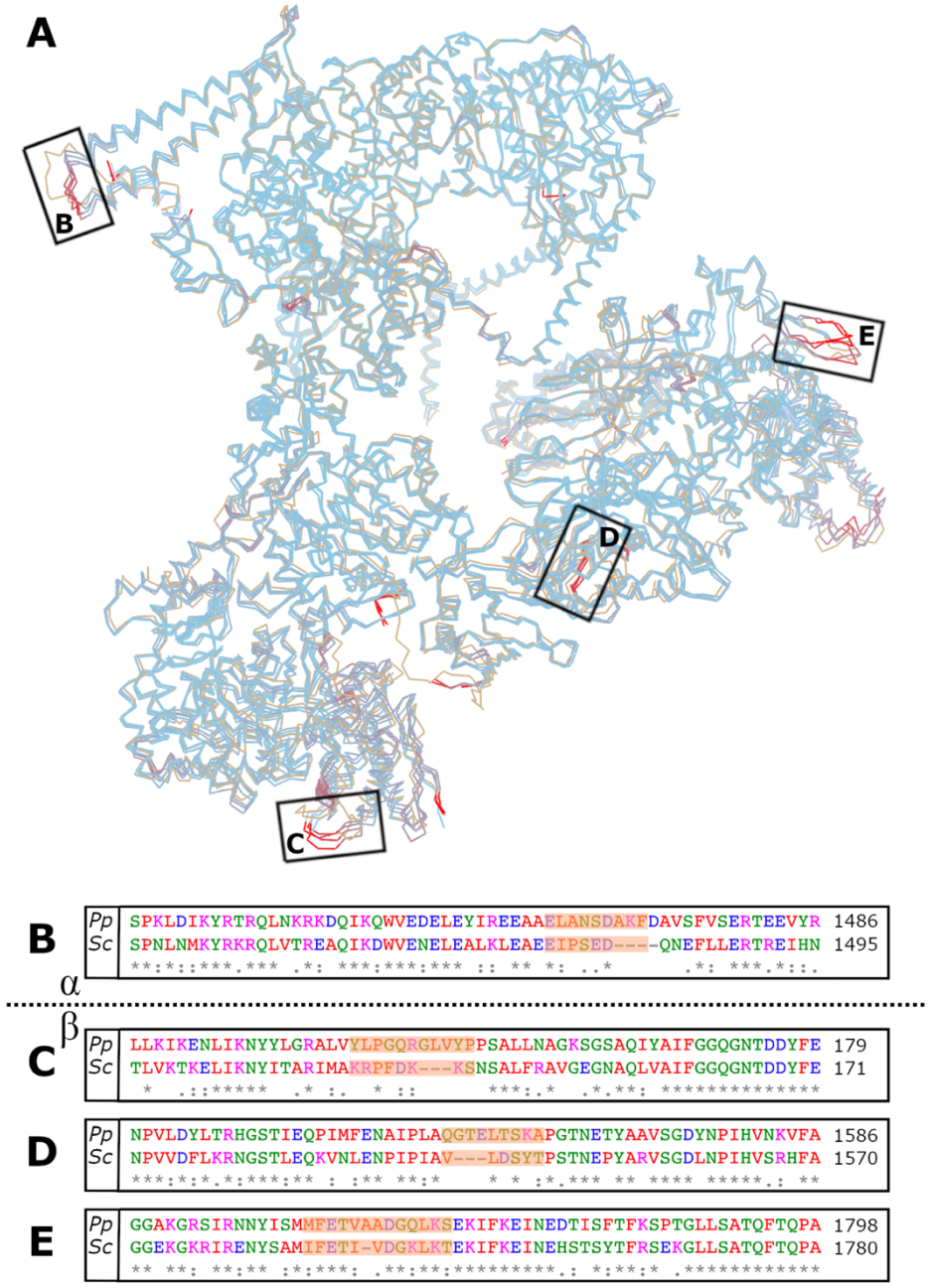
5.40. A subpopulation of FAS PPT domains are intact. (A) *All classes from focussed 3D classification of the PPT domain.* ***(B,C)*** *The PPT density-containing class (indicated by the black box in (A)) shown overlaid with an atomic model of S. cerevisiae FAS (including the PPT domain, highlighted in pink). Density is shown at both* ***(B)*** *high and* ***(C)*** *low contour thresholds. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

5.5.7. Structural comparison of *P. pastoris* FAS and *S. cerevisiae* FAS

To identify unique structural features in *P. pastoris* FAS, the atomic model was compared with a recent *S. cerevisiae* FAS structure determined to the same resolution (PDB: 6TA1) (Joppe *et al.*, 2020). While *C. albicans* FAS (73% [α] and 68% [β] sequence identity with *P. pastoris* FAS) was a closer structural match than *S. cerevisiae* FAS (69% [α] and 63% [β] sequence identity with *P. pastoris* FAS) when compared using PDBeFold (Krissinel and Henrick, 2004; Lou *et al.*, 2019; Joppe *et al.*, 2020), *S. cerevisiae* FAS is better characterised and has greater relevance to metabolic engineering attempts. Following alignment using the 'MatchMaker' tool in UCSF Chimera (Pettersen *et al.*, 2004), the RMSD value between Cα atom pairs was relatively high at 7.1 Å, and a similar level of variation was also observed between *P. pastoris* FAS and several other *S. cerevisiae* FAS atomic models derived from both X-ray crystallography and cryoEM (PDBs: 6QL6, 6QL9, 2UV8) (Figure 5.41) (Leibundgut *et al.*, 2007; Singh *et al.*, 2020). Alignment of individual subunits gave substantially lower RMSD values (1.7 Å [α-α] and 2.2 Å [β-β]), indicating that a large proportion of the variation may be a result of differences in the spatial relationship between the α and β subunits.

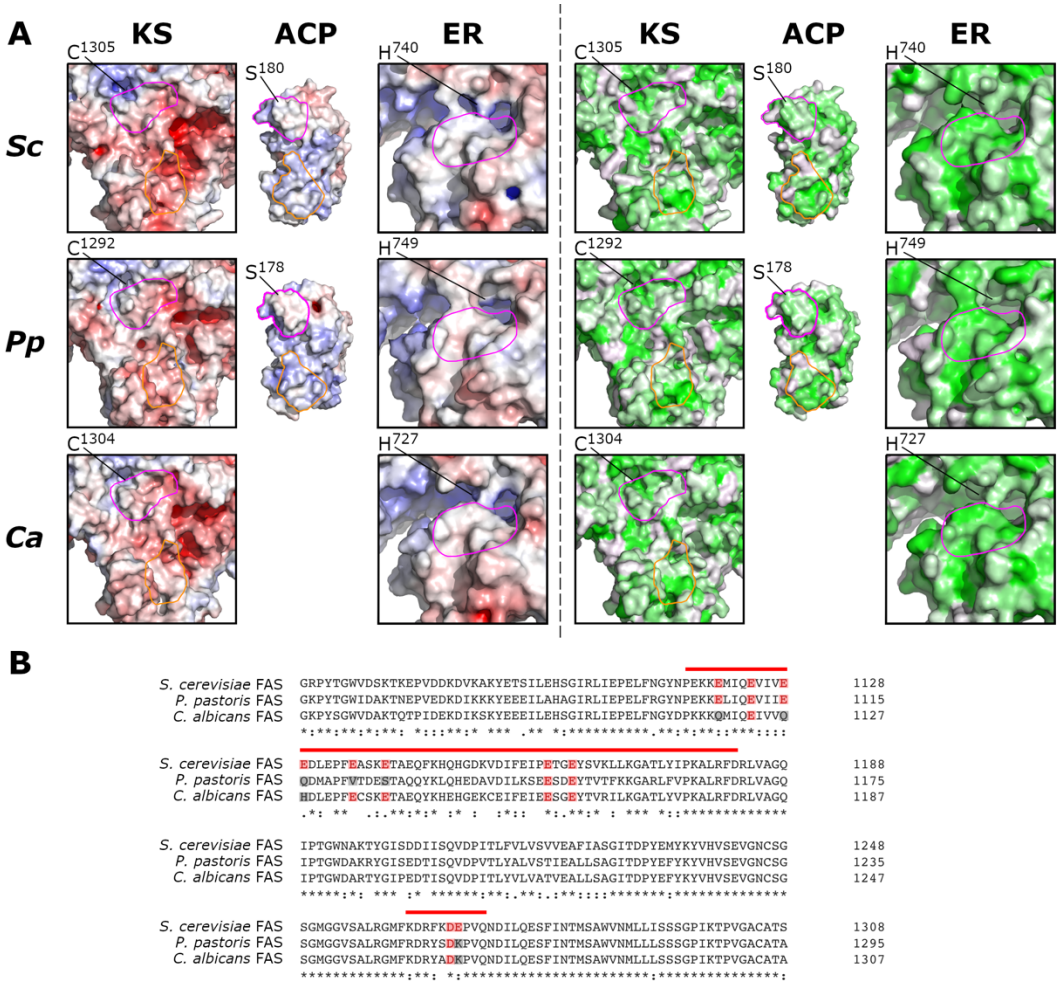
In general, the domain organisation of *P. pastoris* FAS matched well with that of *S. cerevisiae* FAS. This included the ACP domain which was positioned adjacent to the KS domain in both structures (RMSD of 2.1 Å between ACP domain Cα atom pairs after alignment based on KS domains). This contrasted with the ACP domains in *C. albicans* FAS (PDB: 6U5V) and *C. thermophilum* FAS (EMDB: 3757), which were observed to localise to the ER domain (Kastritis *et al.*, 2017; Lou *et al.*, 2019). It has previously been suggested that this difference is driven by the strength of electrostatic interactions between the surfaces of the ACP and KS domains, given that the *C. albicans* FAS KS domain has a weakened negative surface charge relative to that of *S. cerevisiae* (Anselmi *et al.*, 2010; Lou *et al.*, 2019). Interestingly though, while some of the key acidic residues involved in the ACP-KS interaction that are absent in *C. albicans* FAS are retained in *P. pastoris* FAS, other residues contributing to the negative surface charge of KS that are conserved in both

S. cerevisiae FAS and *C. albicans* FAS are replaced with non-charged residues in *P. pastoris* FAS (Figure 5.42).



5.41. Structural comparison of P. pastoris FAS and S. cerevisiae FAS. ***(A)*** *The atomic model of P. pastoris FAS (beige) overlaid with a range of S. cerevisiae FAS structures (PDB: 6TA1, 6QL6, 6QL9, 2UV8; blue). Regions of high RMSD between Cα pairs (≥15 Å) are highlighted in red. Segments of the backbone that were obviously misaligned (e.g. due to missing residues) were removed for clarity. Labels on the overlay correspond to the*

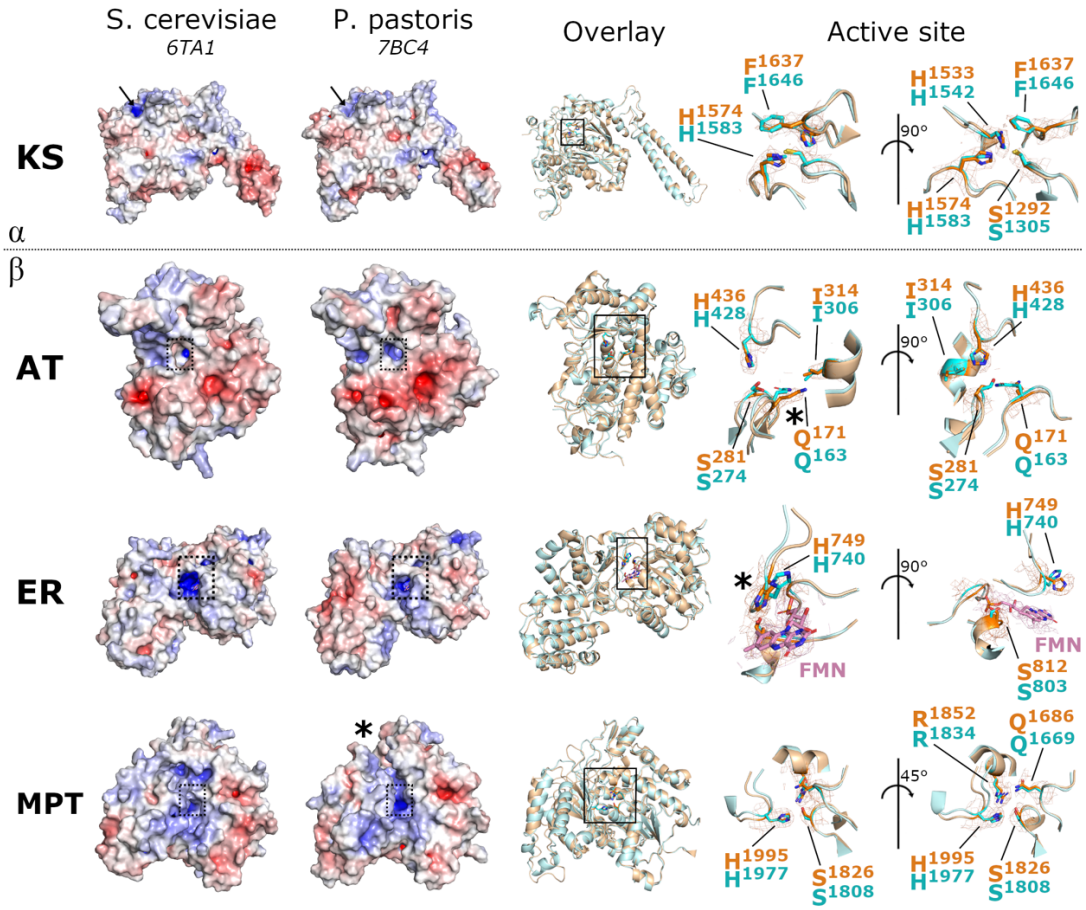
*primary sequence alignments of **(B)** α and **(C-E)** β subunits. Sequence alignments between P. pastoris FAS (UniProt references C4QY10-1 [α] and C4QVT8-1 [β]) and S. cerevisiae FAS (UniProt references P19097-1 [α] and P07149-1 [β]) were generated using Clustal Omega (Sievers et al., 2011) with default parameters, coloured according to the default Clustal colouring scheme. Sections of the primary sequence corresponding to the areas of high RMSD in (A) are highlighted in red. Pp – P. pastoris FAS; Sc – S. cerevisiae FAS. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*



***5.42. Comparative analysis of ACP-KS and ACP-ER interactions. (A)** Molecular surfaces of ketoacyl synthase (KS), acyl carrier protein (ACP) and enoyl reductase (ER) domains of FAS complexes from S. cerevisiae (Sc; PDB: 6TA1), P. pastoris (Pp) and C. albicans (Ca; PDB: 6U5V). Surfaces*

are coloured by surface charge using PDB2PQR (Dolinsky et al., 2007) and APBS (Baker et al., 2001) (left; blue – positive charge, red – negative charge) or hydrophobicity according to the Eisenberg hydrophobicity scale (Eisenberg et al., 1984) (right; green – hydrophobic). Approximate molecular footprints are outlined in magenta and orange. ACP domain side chains were not included in the C. albicans FAS model, so this was excluded. ***(B)*** *A multiple sequence alignment between FAS α subunits from S. cerevisiae (UniProt reference P19097-1), P. pastoris (UniProt reference C4QY10-1), and C. albicans (UniProt reference P43098-1) was performed using Clustal Omega (Sievers et al., 2011) with default parameters. Parts of the primary sequence of the KS domain that contribute to the ACP-KS interface are indicated with the red horizontal line. Key acidic residues for S. cerevisiae FAS (previously identified by (Lou et al., 2019)) are highlighted in red, with altered residues highlighted in grey for P. pastoris FAS and C. albicans FAS. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

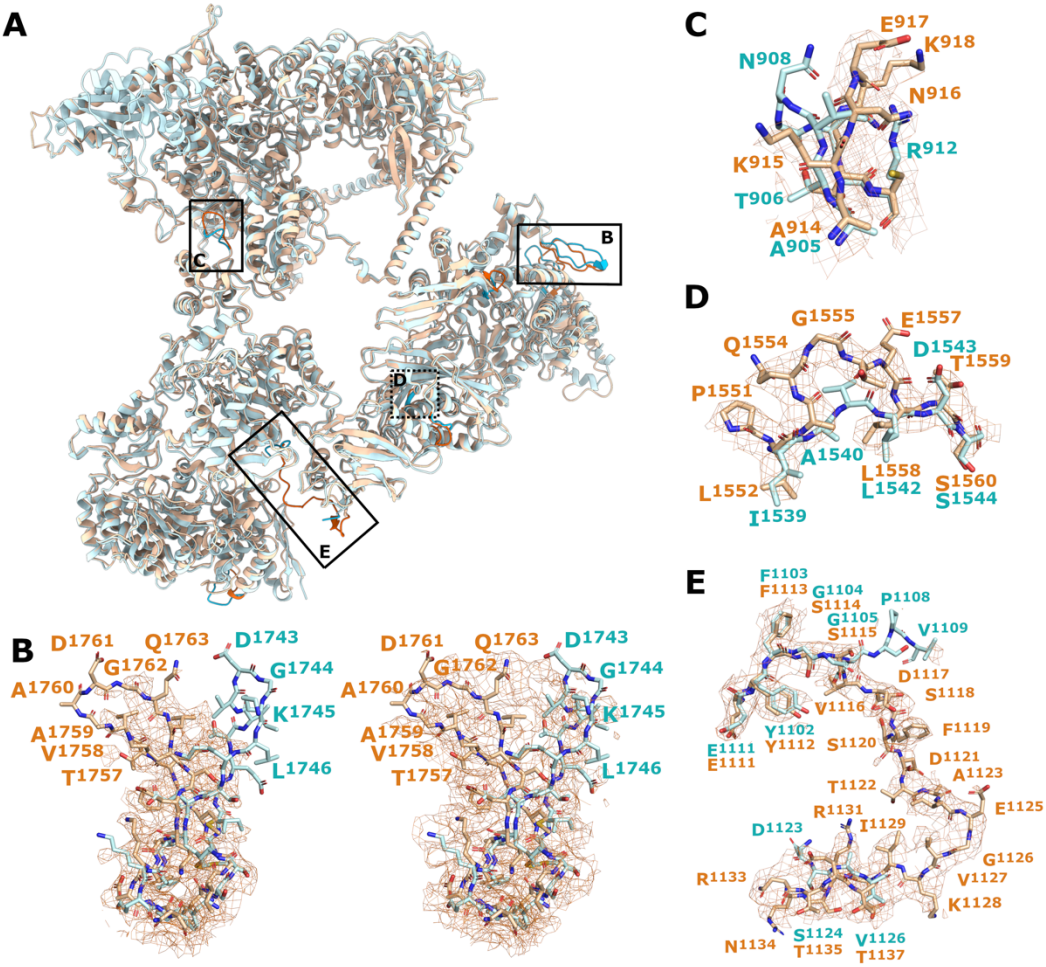
In contrast to the ACP domain, the *P. pastoris* FAS MPT domain showed a clear deviation from its *S. cerevisiae* equivalent. Specifically, the ferredoxin-like subdomain was tilted away from the α/β-hydrolase fold, creating extra space between the subdomains around the active site (Figure 5.43).



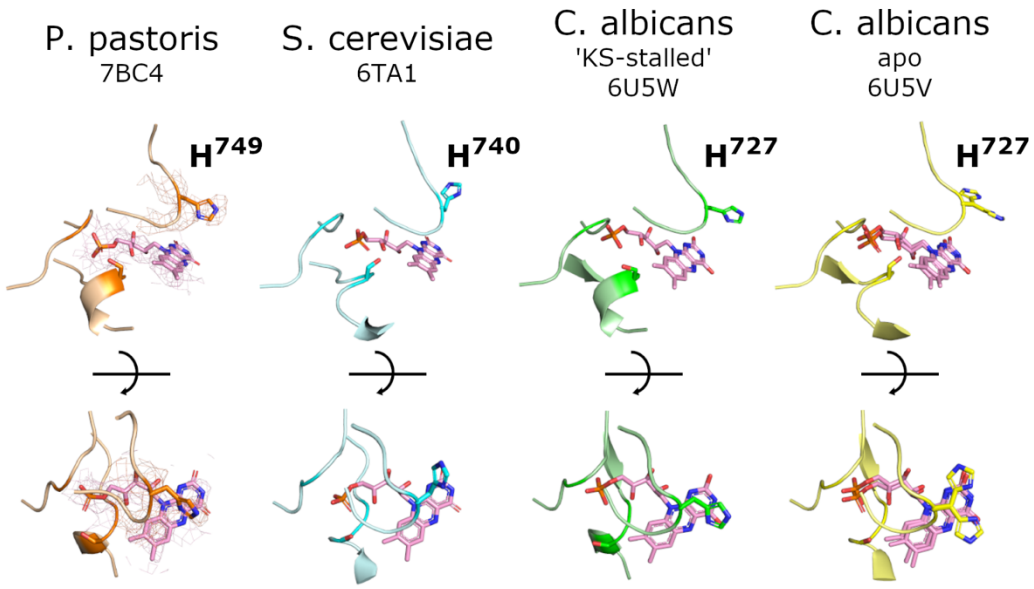
5.43. Structural details of key FAS enzymatic domains. *Key enzymatic domains (KS – ketoacyl synthase, AT – acetyl transferase, ER – enoyl reductase, MPT – malonyl/palmitoyl transferase) from both P. pastoris FAS and S. cerevisiae FAS (PDB: 6TA1) are shown as molecular surfaces coloured according to electrostatic charge (blue – positive, red – negative) with active sites highlighted by dashed boxes (AT/ER/MPT) or black arrows (KS). Atomic coordinates for the enzymatic domains are also overlaid (beige – P. pastoris FAS, light blue – S. cerevisiae FAS) and key active site residues are shown as enlarged views (corresponding regions indicated by solid black boxes). Density from the P. pastoris FAS density map is shown around highlighted active site residues. FMN and corresponding EM density is shown in pink. Key differences referred to in the main text are highlighted by asterisks (Q171 in the AT domain, H749 in the ER domain, MPT subdomain arrangement). (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

At the level of the peptide backbone, there were numerous differences between *P. pastoris* FAS and *S. cerevisiae* FAS scattered throughout the structure, mostly within the β subunit (Figure 5.44A). This included a loop within the MPT domain which was shifted by 16 Å at its turning point (D1761) (Figure 5.44B). Interestingly, density for the path traced by the loop in *S. cerevisiae* FAS was also observed at low contour levels, suggesting some level of exchange between the two conformations. Differences observed for other regions of the peptide backbone included changes in the DH domain (F1283-V1289, E1505-I1511, A1553-L1558), MPT domain (A1859-Y1870), and non-enzymatic domains/linker regions (G75-N78, A914-G919) (Figure 5.44C,D). A sequence of residues linking the ER domain to the non-enzymatic domain adjacent to the DH domain (F1113-T1135) was mostly unresolved in *S. cerevisiae* FAS (PDB: 6TA1), but was well-resolved for *P. pastoris* FAS and the path traced by the peptide backbone differed significantly from the terminal parts of the linker resolved for *S. cerevisiae* FAS (Figure 5.44E).

While most enzymatic domains matched closely between FAS structures, subtle rotameric differences were observed for key residues in the AT and ER domains. Specifically, Q171 was shifted in the active site of the AT domain, potentially impacting the arrangement of a tight hydrogen bonding network at the catalytic centre of the AT domain (Figure 5.43). Similarly, the catalytic histidine of the ER domain (H749) was oriented in a different direction to the equivalent residue in *S. cerevisiae* FAS (Figure 5.43). Interestingly, the orientation of H749 observed in *P. pastoris* FAS overlaid well with the equivalent histidine from *C. albicans* FAS in the KS-stalled state, whereas the positioning observed for *S. cerevisiae* FAS matched more closely to one of the two rotamers identified in the apo (non-KS-stalled) state of *C. albicans* FAS (Figure 5.45) (Lou *et al.*, 2019).



5.44. Differences in the peptide backbone of P. pastoris FAS. (A) *Overlay of P. pastoris FAS (beige) and S. cerevisiae FAS (blue, PDB: 6TA1), with differences highlighted by stronger colours. Enlarged images are indicated on the overlay by solid black boxes.* ***(B)*** *Loop region comprising residues 1753 – 1768 of subunit β, highlighting the different positioning compared to the equivalent region from S. cerevisiae FAS. The P. pastoris density map is shown overlaid at both high (left) and low (right) contour levels.* ***(C-E)*** *Other regions from different parts of the complex are shown overlaid with the equivalent regions from S. cerevisiae FAS, including the loop comprising residues 914 – 919* ***(C)****, the β-strand comprising residues 1553 – 1558* ***(D)*** *and the additional resolved linker region comprising residues 1113 – 1135* ***(E)****, all from the β subunit. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*



5.45. Comparison of yeast FAS ER domain catalytic histidine rotamers. *The catalytic histidine of various yeast FASs are shown in relation to FMN (pink). Density is shown for P. pastoris FAS. Structures for both KS (ketoacyl synthase)-stalled and apo states of C. albicans FAS are shown. The apo state of C. albicans FAS has two alternative rotamers, both are shown. (Figure adapted from (Snowden et al., 2021), published under a CC BY 4.0 licence.)*

5.6. Discussion and conclusions

5.6.1. VelcroVax

Using cryoEM, I was able to determine the structures of two different VelcroVax constructs, AR1 and n190, and derive structural information on the interaction between n190 and an example target antigen, SUMO-GP1. This information permitted comparison of VelcroVax constructs with unmodified HBcAg, and offered insight into the functioning of the VelcroVax platform and how the constructs might be improved.

The high-resolution structures reported here (in line with other tandem core constructs described in previous low-resolution structural studies (Peyret *et al.*, 2015)) confirmed that both AR1 and n190 formed intact capsids in $T = 3$ and

T = 4 organisations that were highly similar to VLPs assembled from unmodified HBcAg (Yu *et al.*, 2013b; Makbul *et al.*, 2021). Previous studies of HBcAg revealed a disulphide bond linking cysteine residues from adjacent monomers at the dimerisation interface within the four-helix bundle (C61-C61) (Nassal *et al.*, 1992; Zheng *et al.*, 1992; Wynne *et al.*, 1999). In contrast, a cryoEM structure of the HBV capsid vitrified quickly after purification (<24 h) showed no disulphide bond at this position, providing structural evidence that it was not required for capsid assembly, with later biochemical studies indicating that oxidation occurs with a half-time ($t_{1/2}$) of ~53 h in the context of an assembled capsid (Yu *et al.*, 2013b; Selzer *et al.*, 2014). In line with this finding, the AR1 structures reported here (vitrified 72 – 96 h after purification) showed evidence of a disulphide bond linking the equivalent cysteine residues (C61-C345) within each AR1 subunit. As expected, density for this disulphide bond was not present in the n190 structure because the equivalent cysteine was mutated to an alanine in the n190 construct, as described in Section 5.1.3.

For AR1, Affimers were not resolved in the cryoEM density maps even following focussed classification, despite a range of biochemical evidence supporting that they were intact at the time of vitrification. This was likely due to a combination of (i) flexibility of the Affimer relative to the four-helix bundle, and (ii) inherent asymmetry within each AR1 subunit, meaning each four-helix bundle would have an Affimer attached in one of two possible orientations. While n190 was highly similar to AR1 overall, additional density for Affimers was observed. Unlike AR1, n190 is not a tandem core construct, but comprises a single modified HBcAg monomer. Thus, n190 subunits do not have the same problems with inherent asymmetry. Furthermore, the positioning of the Affimers at the N-terminus of HBcAg means that they are spatially surrounded by four-helix bundles, rather than extending distally away from them as in AR1. It is plausible that the flanking four-helix bundles restricted the movement of the Affimers in n190, reducing the structural heterogeneity in the dataset.

While the inherent flexibility of the Affimers in AR1 was problematic for structural studies, the implications for immunogenicity are unclear. For example, it is plausible that displaying antigens via Affimers with more flexible

positioning may facilitate more efficient cross-linking of immune cell receptors. Alternatively, if the immune response against the target antigen is primarily driven by professional antigen presenting cells via MHC II molecules following phagocytosis and degradation of ligand-bound AR1/n190, Affimer flexibility may not be expected to have a major effect. Similarly, the effect of particle size (T = 3 vs T = 4) on immunogenicity is unclear and remains to be determined empirically. Another key question that must be addressed is the effect of prior immunity to HBV on the immunogenicity of VelcroVax. While the HBV vaccine relies on HBsAg, so does not induce an immune response against HBcAg, infection with HBV induces a broad response to multiple viral proteins, including HBcAg. Given that the HBcAg component of AR1/n190 was found to be extremely similar in its structure/conformation to unmodified HBcAg, it is likely that pre-existing infection-induced immunity to HBV would recognise HBcAg in VelcroVax and affect the generation of an immune response to the target antigen.

An interesting observation for the n190 dataset was evidence for the presence of nucleic acid within n190 VLPs, which was not observed for AR1. The precise identity of any packaged nucleic acid was unclear, although it is plausible that this may correspond specifically to the n190 mRNA present within *E. coli* when n190 is expressed. Alternatively, n190 may simply package non-specific nucleic acid during assembly. To explore this further, it would be useful to analyse nucleic acid extracted from n190 VLPs by agarose gel electrophoresis, before and after treatment with DNase or RNase. If the nucleic acid within the particles was found to be RNA, this could be followed up by northern blot analysis to determine whether or not the VLP is packaging n190 mRNA specifically. It is also unclear why only n190, and not AR1, packaged nucleic acid. The two most plausible explanations are that this was a result of (i) differences in the expression system (*E. coli* compared with *P. pastoris* for AR1), or (ii) differences in the design of the constructs affecting packaging efficiency – notably, AR1 only has half as many ARDs as n190. Inverting the expression systems used to generate AR1 and n190 would provide useful insight, though AR1 has previously shown solubility issues when expressed in *E. coli* which would make any sort of structural characterisation difficult.

Regardless of the reason for this difference, the ability to package nucleic acid within the VLPs would present interesting opportunities to be explored. It is possible that the presence of nucleic acid may confer self-adjuvating activity, given that viral nucleic acids are commonly detected as pathogen-associated molecular patterns (PAMPs) by different pathogen-recognition receptors (PRRs) to initiate the innate immune response (Luecke and Paludan, 2017). Previous studies of truncated HBcAg (Singh and Zlotnick, 2003), along with unpublished data from the Stonehouse/Rowlands laboratory (data not shown), indicate that it is possible to disassemble, then reassemble n190 VLPs *in vitro*. By introducing a nuclease treatment step between disassembly and reassembly, it may be possible to produce nucleic acid-free n190 VLPs for comparison with nucleic acid-containing n190 VLPs in immunological trials, to check for evidence of a self-adjuvating effect. If it were possible to package specific nucleic acid sequences (for example, by disassembly followed by reassembly in the presence of specific target DNA/RNA), this system may also show potential as a gene delivery vehicle. Notably, a previous study combined RNA SELEX with bioinformatic analysis of the HBV pre-genomic RNA to identify packaging signals, short RNA sequences that could form stem-loop structures to mediate high-affinity interactions with HBcAg, primarily via the ARD (Patel *et al.*, 2017). Incorporation of these packaging signals into target sequences of RNA/DNA may facilitate specific and efficient packaging of nucleic acid for gene delivery.

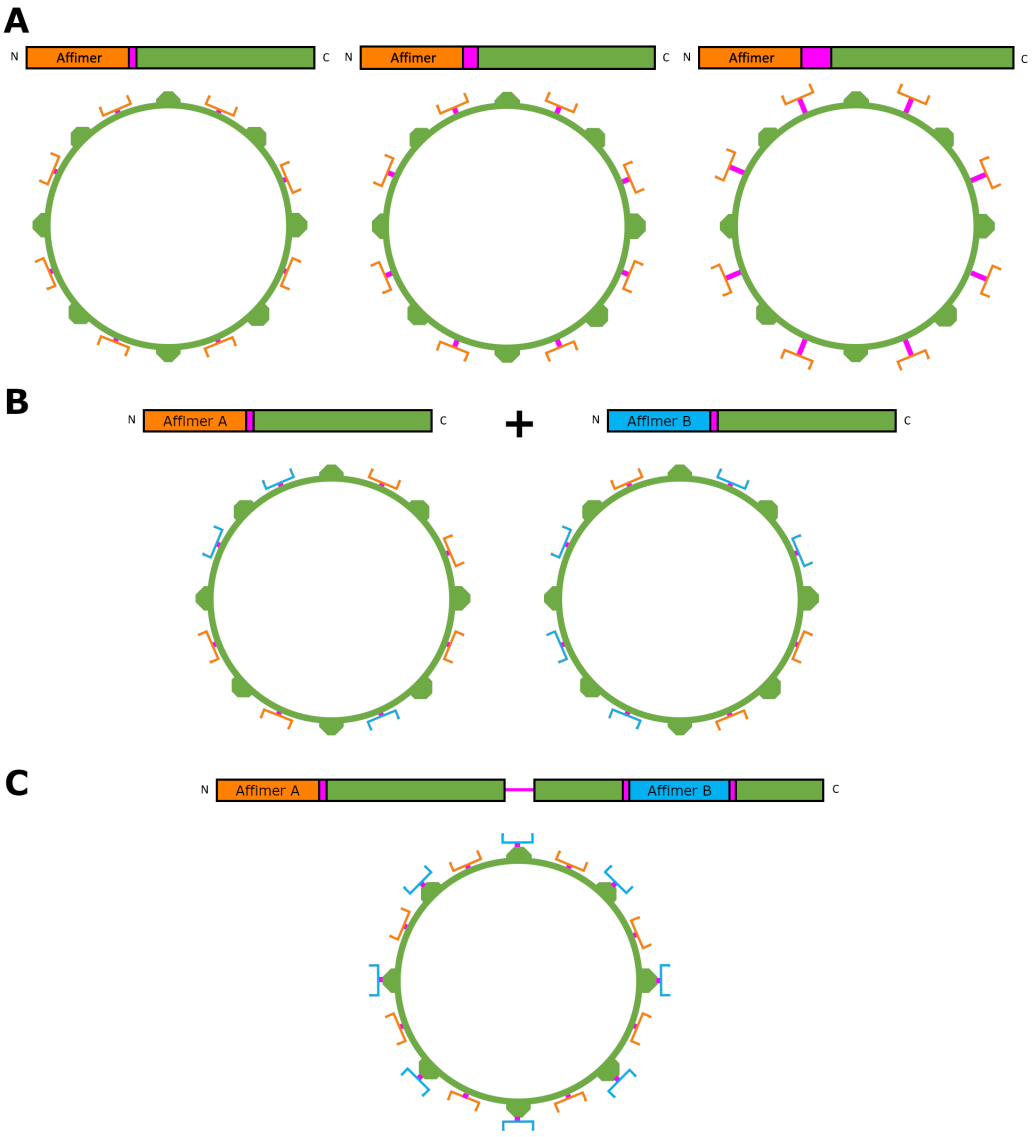
Based on the ability to visualise Affimers in n190 VLPs and the fact that n190 can assemble into true $T = 3$ VLPs (rather than the pseudo-symmetrical $T = 3^*$ VLPs generated by AR1), n190 was selected for further structural analysis of VelcroVax constructs in complex with the target antigen SUMO-GP1. SUMO-GP1 was found to induce aggregation when mixed with n190 in solution, so an ‘on-grid’ binding approach was used to prepare samples for cryoEM. While this approach offered a quick solution to prevent aggregation, it also presented downsides. Firstly, the maximum binding occupancy was limited because the initial binding partner (n190 in this case) was in contact with the grid surface and therefore partially inaccessible to the second binding partner (SUMO-

GP1). On-grid binding also necessitates the use of grids with a continuous support film, which reduces the signal-to-noise ratio of the data to some extent. In spite of these issues, ‘on-grid’ binding was pursued in order to avoid a lengthy optimisation process with no guarantee of success.

The symmetrised reconstruction of n190 in complex with SUMO-GP1 showed no density for SUMO-GP1 – indeed, the only evidence of binding was a small shift in the average position of the Affimers. Using focussed classification, density for SUMO-GP1 was visualised, although not at a resolution sufficient for atomic modelling, likely a result of both flexibility and low occupancy. Moving forward, determination of the structure of n190 in complex with a target antigen that does not induce aggregation (such that the interaction can be performed in solution) may yield improved occupancy and better resolution at the interaction interface. One of the key findings from this dataset was that steric hindrance of adjacent ligands would likely prevent full occupancy from being achieved. This correlated with ITC data for n190 and SUMO-GP1, which indicated a low stoichiometry with a predicted maximum of 17.9% of total binding sites occupied. Interestingly, ITC data for the interaction between n190 and unconjugated SUMO (collected by KF) indicated a stoichiometry of 0.428 (Figure 5.23). Thus, a greater occupancy of n190 Affimers was achieved when using a smaller ligand, highlighting the importance of ligand size.

While the specific effect of occupancy on immunogenicity has not been investigated, it is plausible that reduced occupancy would lead to less efficient cross-linking of immune cell receptors and a weaker immune response. As such, selecting appropriately sized target antigens is likely to be particularly important for VelcroVax, and it may be possible to determine approximate size limitations/guidance empirically. It may also be possible to improve occupancy through other approaches. In principle, the spacing between Affimers would be greater for $T = 4$ VLPs than for $T = 3$ VLPs, so isolating or enriching for $T = 4$ VLPs may offer improved occupancy. Likewise, Affimers at the tips of four-helix bundles are at a greater radial distance and therefore have a greater degree of separation than those at the N-terminus, so AR1 may give improved occupancy compared to n190. However, as a tandem core construct with only

one Affimer for every two HBcAg monomers, it would also have fewer total binding sites. Alternatively, constructs could be designed with longer flexible linker sequences between HBcAg and Affimer to increase the spacing between them and to afford greater mobility to accommodate target antigens (Figure 5.46A).



5.46. Potential alternative 'VelcroVax' construct designs. *Representative schematics of potential constructs showing* ***(A)*** *constructs with different lengths of linker (magenta) between Affimer (orange) and HBcAg (green);* ***(B)*** *constructs with different Affimers or capture systems (orange/blue) that can be co-expressed to produce heterotypic VLPs;* ***(C)*** *a tandem core construct with one Affimer/capture system (orange) inserted at*

the N-terminus of the first HBcAg monomer, and one Affimer/capture system (blue) inserted at the MIR of the second HBcAg monomer. This would generate a VLP decorated with an equal quantity of Affimer A and Affimer B, as illustrated, although note that the total quantity of N-terminal Affimers displayed would be half that for the VLPs expressed from the non-tandem constructs shown in (A,B).

Another consideration is whether it would be possible to present a number of different antigens at once. This could offer broader protection against different serotypes or different pathogens. The simplest approach would be to incubate VelcroVax VLPs with a mix of different antigens, all modified to contain the epitope tag recognised by the VelcroVax capture system. However, it would be difficult to control the relative proportions of each antigen displayed by the VLP. As was shown through ITC for unconjugated SUMO and SUMO-GP1, the affinity of the interaction and maximum occupancy can change even for the same epitope tag, based on what the tag is conjugated to. Alternatively, it would be possible to express different constructs, each with a unique capture system, to create heterotypic VLPs through co-expression, but it would be very difficult to control the precise composition of different VLPs (Figure 5.46B). To maintain a fixed stoichiometry, constructs could be engineered to contain independent capture systems within each subunit. For example, a tandem core construct could be designed to include an Affimer against SUMO at the N-terminus, and an Affimer against an alternative epitope tag at the MIR of one copy of HBcAg (Figure 5.46C). Affimers could also be replaced with alternative capture systems, such as nanobodies or SpyTag/SpyCatcher (Zakeri *et al.*, 2012; Peyret *et al.*, 2015, 2020; Brune *et al.*, 2016). Incorporating two co-existing capture systems into a single construct may lead to steric clashes that would be problematic for occupancy, although the structures reported here indicate that capture systems at the N-terminus and MIR would be at different radial distances, perhaps providing more separation between them.

While there is still room for VelcroVax constructs to be improved, initial immunisation trials performed by collaborators at the National Institute for

Biological Standards and Control (NIBSC) have shown promising signs for the platform in general. Preliminary results indicated that mice inoculated with AR1 decorated using a SUMO-tagged version of SARS-CoV-2 spike protein receptor binding domain (RBD) generated a more potent antibody response against SUMO-RBD than those inoculated with soluble SUMO-RBD, determined by ELISAs performed by Dr Keith Grehan (KG) (Figure 5.47). With further improvement to specific constructs guided in part by the structural data reported here, VelcroVax technology has strong potential to contribute to our arsenal of vaccines against infectious disease.

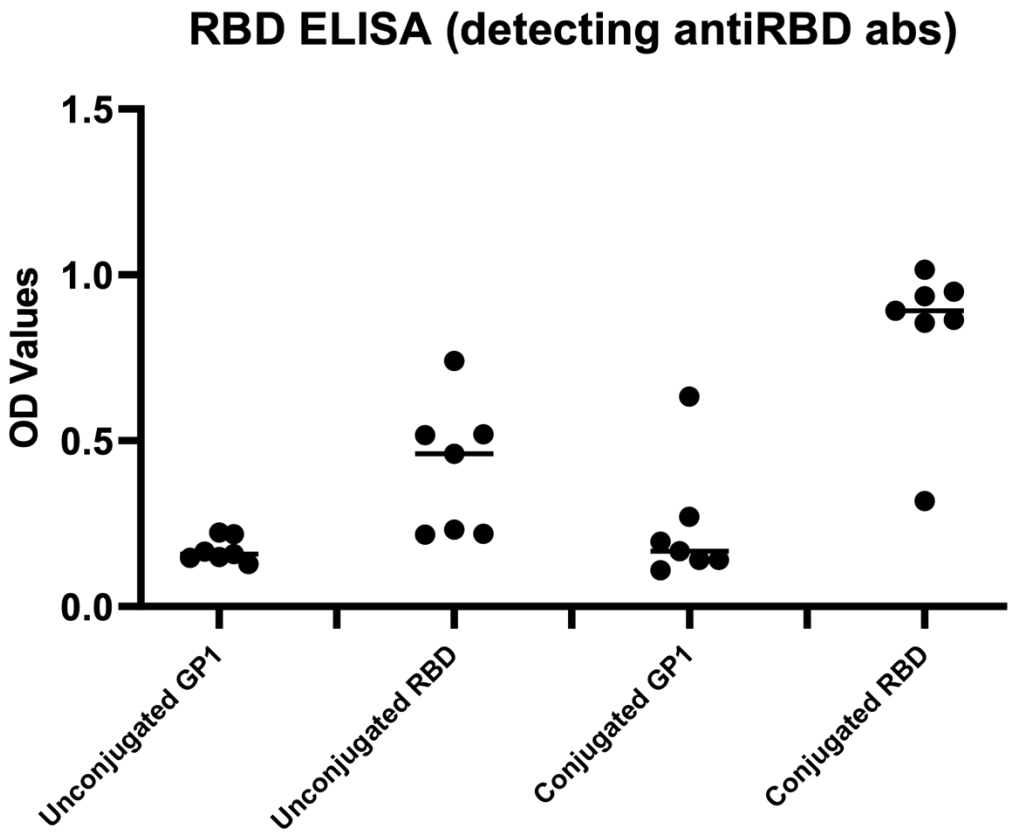


Figure 5.47. Antigen bound to AR1 has improved immunogenicity. *Mice were immunised with different constructs (SUMO-tagged GP1 or RBD alone [unconjugated] or bound to AR1 [conjugated]) and their sera analysed for antibodies against RBD by ELISA. Immunisation experiment performed by collaborators at NIBSC. ELISA performed and figure generated by KG.*

5.6.2. Yeast fatty acid synthase

FAS was identified in 2D classification results from a cryoEM dataset of AR1, expressed in *Pichia pastoris*. There was no evidence of a specific interaction between FAS and AR1, and FAS particles were generally well separated from AR1 particles in raw micrographs. As such, it is likely that FAS co-sedimented with AR1 during sucrose gradient purification as a result of its similar molecular mass (2.6 MDa, compared with 4.7 MDa and 6.3 MDa for $T = 3$ and $T = 4$ AR1 VLPs, respectively).

A previous study of FAS from the yeast *S. cerevisiae* reported evidence of FAS denaturation, and hypothesised that this was due to interactions between the particles and the air-water interface. In that study, this issue was solved through coating grids with hydrophilised graphene, which is thought to reduce such denaturation-inducing contacts with the air-water interface (Drulyte *et al.*, 2018; D'Imprima *et al.*, 2019). Here, grids coated with a continuous thin carbon film were used, and this is likely to have had a similar effect, trapping particles away from the air-water interface by adsorption to the support film. It is also possible that the presence of VLPs may have favourably altered the properties of the ice, but this is thought to be less likely. A problem that sometimes arises when using continuous support films is the adoption of a preferred orientation by the target particle due to interactions with the support film. However, while there was evidence of some preferential orientation effects, this was not particularly problematic and did not impede high resolution structure determination. Similarly, continuous support films lead to a reduction in the signal-to-noise ratio of the data as a result of additional background contrast (Drulyte *et al.*, 2018), although again, this was not a significant issue for a large complex like yeast FAS.

As a well characterised yeast FAS and the target of most metabolic engineering attempts, FAS from *S. cerevisiae* was chosen for comparative analysis with the *P. pastoris* FAS structure reported here. This revealed that the mobile ACP domain localised to the KS domain, in line with *S. cerevisiae* FAS, but in contrast to FAS from *C. albicans*, where the ACP was observed to localise to the ER domain (Lou *et al.*, 2019; Joppe *et al.*, 2020). The ACP-KS

interaction is thought to be dependent on electrostatic interactions between the two domains, and *C. thermophilum* FAS was previously shown to lack some of the key acidic residues present in the *S. cerevisiae* FAS KS domain, potentially explaining the difference in ACP localisation. Interestingly, *P. pastoris* FAS maintained some of the acidic residues that were absent in *C. thermophilum* FAS, but others present in FAS from both *S. cerevisiae* and *C. thermophilum* were absent. This suggested that some acidic residues at the ACP-KS interface (including E1107 and E1115) may be more important in driving the ACP-KS interaction.

Other than the ACP domain, three key structural differences were observed in enzymatic domains compared to *S. cerevisiae* FAS. Alternative conformers were observed for key active site residues in both AT and ER domains, although perhaps more interestingly, there was a shift in the relative positioning of subdomains within the MPT domain. This difference is particularly noteworthy because the MPT domain has previously been targeted as part of a successful attempt to re-engineer FAS to favour the production of short-chain fatty acids (Gajewski *et al.*, 2017b).

Ultimately, what was initially a serendipitous observation of contamination in a VLP cryoEM dataset led to the determination of a 3.1-Å resolution structure and novel insight into the structural biology of yeast FAS complexes. In addition to providing a useful resource for rational engineering of FAS in *P. pastoris*, this highlights the importance of considering the value of contaminant data in cryoEM datasets.

6. Summary and concluding remarks

6.1. Overview

Through structural studies of murine norovirus and a novel vaccine platform technology, this thesis illustrates the utility of cryoEM in our efforts to understand and overcome viruses and viral disease.

In Chapters 3 and 4, MNV was used as a model system to explore norovirus biology with cryoEM. Following the selection of an optimised purification protocol, structures for wtMNV, hiMNV and a mutant hsMNV were determined by cryoEM and compared with other MNV structures. MNV P domains were found to be highly dynamic, with both independent (i.e., variation in positioning of individual P domain dimers even within the same capsid) and synchronous (i.e., switching between contracted and expanded conformations) components to their dynamic nature. Changes in the conformational landscape explored by P domain dimers correlated with differences in infectivity and thermal stability, emphasising the importance of P domains for norovirus infection. Structural and mutagenesis data was also combined to suggest a potential mechanism by which the L412Q substitution may stabilise hsMNV, and the possible application of this finding to norovirus vaccine design was highlighted.

In contrast to wtMNV from RAW264.7 cells, wtMNV from BV-2 cells was observed to have rigidified P domain dimers, leading to a significant improvement in local P domain resolution. Although a cause was not conclusively identified, the observation of ‘adduct’ density at a particular cysteine residue in VP1 for wtMNV.RAW but not for wtMNV.BV2 raised the possibility of a cell-line specific PTM, which could plausibly affect P domain rigidity. Rigidified P domains were also observed for BV-2 cell-derived wtMNV in complex with the bile acid, LCA. Density for LCA was present and supported findings from X-ray crystal structures of isolated P domain dimers in complex with bile acids (Nelson *et al.*, 2018). While attempts to resolve VP2, VPg and the viral genome were ultimately unsuccessful, the processing pipelines described in this thesis serve as a useful blueprint for resolving asymmetric

features of icosahedral viruses, and may yield exciting results with larger data sets.

In Chapter 5, cryoEM was applied to the 'VelcroVax' vaccine platform technology. Structures were determined for two different constructs based on HBcAg, termed AR1 and n190. While AR1 is based on a 'tandem core' design, with an Affimer presented at the MIR of only one of the two HBcAg monomers in each AR1 subunit, n190 is a monomeric construct with an Affimer attached to the HBcAg N-terminus. Affimers were not resolved for AR1, likely a result of complications introduced by the tandem core design and increased flexibility from the positioning of the Affimers at the tips of four-helix bundles, however n190 showed additional Affimer density close to the N-terminus of HBcAg monomers. Further structural investigation of n190 in complex with a target antigen (SUMO-GP1) revealed issues with occupancy, in line with biochemical and biophysical experiments performed by others. This was, at least in part, likely a result of insufficient spacing between adjacent Affimers to accommodate SUMO-GP1. Based on this and other findings, potential approaches to improve the design of VelcroVax constructs were suggested. Finally, contaminant data from the AR1 dataset was processed separately to yield a high-resolution structure of a yeast FAS – the first described for *P. pastoris*. This revealed a number of conformational differences in key catalytic domains from the prototypical *S. cerevisiae* FAS structure, with potential implications for metabolic engineering approaches, and ultimately highlighted the importance of carefully considering the value of contaminant data in cryoEM datasets.

Perhaps one of the most striking features of the cryoEM field is the rate at which the technique has developed. In 2017, when the work described in this thesis commenced, the average resolution of single particle analysis (SPA) cryoEM maps submitted to the EMDB was 8.32 Å, and the highest was 2.14 Å. Just four years later, in 2021, the average resolution of submitted SPA maps had improved to 5.26 Å, with the best-resolved at 1.27 Å (the overall best-resolved SPA map in the EMDB was submitted in 2020, at 1.15 Å resolution [EMD-11668]). It is not just the quality of maps that has improved; there has

been rapid growth in the number of published cryoEM structures too. While it took almost 15 years (2002 – 2017) for the total number of maps deposited in the EMDB to exceed 5,000, this number has now more than trebled to over 17,000 at the time of writing, in just four years. A number of factors have contributed to this change, including improvements in both hardware (such as high-throughput cameras) and software (such as algorithms that can computationally correct aberrations in the images). In particular, the development of more user-friendly and intuitive software packages, such as Relion (Scheres, 2012; Zivanov *et al.*, 2018), has made it easier for a wider range of users to process their data independently.

Aside from the standard data processing tools, a suite of non-canonical processing techniques has expanded our ability to extract information from cryoEM data through SPA. In particular, focussed classification and symmetry relaxation (along with other techniques not described in this thesis) provide a means to use symmetry information to resolve components that (i) display a high degree of flexibility or conformational/positional heterogeneity, or (ii) do not conform to the symmetry of the rest of the complex. In the work reported here, focussed classification was used to show variability in MNV P domains (Chapter 3, Chapter 4), and to resolve Affimers attached to HBcAg via flexible linkers in n190 (Chapter 5), as well as the mobile acyl carrier protein domain of *P. pastoris* fatty acid synthase (Chapter 5). Symmetry relaxation yielded some promising indications for the ssRNA genome of MNV (Chapter 4), although was ultimately unsuccessful, as was the case for the putative genetic material packaged by n190 (Chapter 5). Whether the inability to resolve genetic material was caused primarily by experimental factors (such as insufficient data) or biological factors (including poor ordering of the genetic material or variation in the positioning of the genetic material relative to the capsid) was not clear. Repeating symmetry relaxation with a larger dataset may help to provide more clarity. Regardless, the technique has been used successfully elsewhere and is a useful tool for resolving viral genome density (Lee *et al.*, 2016; Ilca *et al.*, 2019).

6.2. Future directions and concluding remarks

Following on from the work described here, a number of unanswered questions on norovirus biology could be addressed through both structural and complementary techniques. One putative explanation proposed for the rigidification of P domains in wtMNV.BV2 compared to wtMNV.RAW was cell-line specific post-translational modifications to VP1, and while additional density was observed extending from a cysteine residue in only one of the maps, it was not clear what the density described specifically. By using mass spectrometry to compare molecular masses for VP1 from RAW264.7 and BV-2 cells, it may be possible to determine the identity of any potential PTM. Conversely, no difference in mass would suggest that the additional density was potentially a result of modifications occurring during sample preparation for cryoEM, rather than during virus cultivation. To extend the attempts to resolve non-VP1 components of the virion described here, it would be useful to perform tomography on MNV. This could also be performed in the context of cell surface binding and infection. By growing cells on gold EM grids, then incubating these with MNV on ice (to inhibit progression of infection beyond cell surface binding), it may be possible to visualise MNV bound to CD300lf at the thin edge of infected cells through tomography, and in principle, to generate high-resolution reconstructions through sub-tomogram averaging. This could be extended to consider the initial stages of virus internalisation by releasing the temperature lock shortly before vitrification, and performing focussed ion beam (FIB) milling on target cells to generate sufficiently thin sections to visualise virions undergoing endocytosis. Aside from virus biology, the work described here raised the possibility of developing a thermally stabilised VLP-based norovirus vaccine. A logical next step to explore this avenue would be to introduce equivalent mutations to L412Q into VLP constructs for a range of human norovirus genotypes and probe their stability and antigenicity in comparison to wild-type VLPs.

The work presented in Chapter 5 describing a novel vaccine platform technology could also be extended. Putative density for genetic material was observed in n190, but not AR1 – determining cryoEM structures of n190

produced in *P. pastoris* or of AR1 produced in *E. coli* could help to identify whether the packaging of nucleic acid is determined by the construct or the expression system. While a cryoEM structure for n190 in complex with SUMO-GP1 was determined and SUMO-GP1 density identified through focussed classification, using an antigen that does not induce aggregation (such as SUMO-RBD) would permit the interaction to be performed in solution rather than on-grid, improving occupancy and likely yielding better resolved density at the interaction interface. CryoEM revealed differences between n190 and AR1, and showed that both existed as a mix of $T = 3$ ($T = 3^*$ for AR1) and $T = 4$ VLPs. It would be useful to compare constructs in immunisation trials – n190 with AR1, and by enriching for a particular size during purification, $T = 3$ VLPs with $T = 4$ VLPs. Initial immunisation trials with AR1 showed promising results, with increased antibody titres generated in response to the spike protein RBD from SARS-CoV-2 when conjugated to AR1 than when presented alone. It would also be useful to test sera for neutralising titre, and ultimately, to perform challenge studies in mice to examine whether VelcroVax-conjugated antigens generate improved protection against disease. To this end, a murine model for SARS-CoV-2 infection has been reported (Dinnon *et al.*, 2020). Aside from AR1 and n190, the structural results described here suggested potential ways to improve the VelcroVax constructs. For example, constructs could be developed with increased linker length between HBcAg and Affimer to provide more space for ligand binding (and thereby improve occupancy), or using alternative capture systems such as nanobodies or SpyTag/SpyCatcher and its variants (Zakeri *et al.*, 2012; Peyret *et al.*, 2015, 2020; Brune *et al.*, 2016; Keeble *et al.*, 2017). Capture systems could also be combined to generate multivalent constructs that display different antigens at fixed stoichiometries. Any new constructs would need to be probed for ability to capture antigen, and in some cases may benefit from structural analysis.

Ultimately, this thesis demonstrates that cryoEM is an extremely useful tool in the fields of virology and vaccinology. As well as providing insight into virus biology and vaccine structure, analysis of cryoEM data often leads to the formulation of new and exciting questions to guide further investigation and improve vaccine design. At a time when the global impact of viral disease

seems to extend into every aspect of our lives, it is important to make full use of this exciting technique.

7. List of references

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G., Rayment, I., Rossmann, M. G., Suck, D. and Tsukihara, T. (1980) 'Structure of southern bean mosaic virus at 2.8 Å resolution', *Nature*, 286(5768), pp. 33–39. doi: 10.1038/286033a0.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., *et al.* (2010) 'PHENIX: A comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D: Biol Crystallogr*, 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

Adeyemi, O. O., Nicol, C., Stonehouse, N. J. and Rowlands, D. J. (2017) 'Increasing Type 1 Poliovirus Capsid Stability by Thermal Selection', *J Virol*, 91(4), pp. e01586-16. doi: 10.1128/JVI.01586-16.

Adler, J. L. and Zickl, R. (1969) 'Winter vomiting disease', *J Infect Dis*, 119(6), pp. 668–673. doi: 10.1093/infdis/119.6.668.

Ahmad, M., Hirz, M., Pichler, H. and Schwab, H. (2014) 'Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production', *Appl Microbiol Biotechnol,* pp. 5301–5317. doi: 10.1007/s00253-014-5732-5.

Ahmed, S. M., Hall, A. J., Robinson, A. E., Verhoef, L., Premkumar, P., Parashar, U. D., Koopmans, M. and Lopman, B. A. (2014) 'Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis', *Lancet Infect Dis*, 14(8), pp. 725–730. doi: 10.1016/S1473-3099(14)70767-4.

Anselmi, C., Grininger, M., Gipson, P. and Faraldo-Gómez, J. D. (2010) 'Mechanism of substrate shuttling by the acyl-carrier protein within the fatty acid mega-synthase', *J Am Chem Soc*, 132(35), pp. 12357–12364. doi: 10.1021/ja103354w.

Asanaka, M., Atmar, R. L., Ruvolo, V., Crawford, S. E., Neill, F. H. and Estes, M. K. (2005) 'Replication and packaging of Norwalk virus RNA in cultured mammalian cells', *Proc Natl Acad Sci USA*, 102(29), pp. 10327–10332. doi: 10.1073/pnas.0408529102.

Atmar, R. L., Opekun, A. R., Gilger, M. A., Estes, M. K., Crawford, S. E., Neill, F. H. and Graham, D. Y. (2008) 'Norwalk virus shedding after experimental human infection', *Emerg Infect Dis*, 14(10), pp. 1553–1557. doi: 10.3201/eid1410.080117.

Bachmann, M. F. and Jennings, G. T. (2010) 'Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns', *Nat Rev Immunol*, pp. 787–796. doi: 10.1038/nri2868.

Baker, A. T., Boyd, R. J., Sarkar, D., Teijeira-Crespo, A., Chan, C. K., Bates, E., Waraich, K., Vant, J., Wilson, E., Truong, C. D., *et al.* (2021) 'ChAdOx1 interacts with CAR and PF4 with implications for thrombosis with thrombocytopenia syndrome', *Sci Adv*, 7(49), p. 8213. doi: 10.1126/sciadv.abl8213.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) 'Electrostatics of nanosystems: application to microtubules and the ribosome', *Proc of the Natl Acad USA*, 98(18), pp. 10037–10041. doi: 10.1073/pnas.181342398.

Baltimore, D. (1971) 'Expression of animal virus genomes', *Bacteriol*

Rev, 35(3), pp. 235–241. doi: 10.1128/mmbr.35.3.235-241.1971.

Ban, N. and McPherson, A. (1995) 'The structure of satellite panicum mosaic virus at 1.9 Å resolution', *Nat Struct Biol*, 2(10), pp. 882–890. doi: 10.1038/nsb1095-882.

Bányai, K., Estes, M. K., Martella, V. and Parashar, U. D. (2018) 'Viral gastroenteritis', *Lancet*, 392(10142), pp. 175–186. doi: 10.1016/S0140-6736(18)31128-0.

Bartsch, S. M., Lopman, B. A., Ozawa, S., Hall, A. J. and Lee, B. Y. (2016) 'Global Economic Burden of Norovirus Gastroenteritis', *PLoS One*, 11(4), p. e0151219. doi: 10.1371/journal.pone.0151219.

van Beek, J., de Graaf, M., Al-Hello, H., Allen, D. J., Ambert-Balay, K., Botteldoorn, N., Brytting, M., Buesa, J., Cabrerizo, M., Chan, M., *et al.* (2018) 'Molecular surveillance of norovirus, 2005–16: an epidemiological analysis of data collected from the NoroNet network', *Lancet Infect Dis*, 18(5), pp. 545–553. doi: 10.1016/S1473-3099(18)30059-8.

Beld, J., Lee, D. J. and Burkart, M. D. (2015) 'Fatty acid biosynthesis revisited: structure elucidation and metabolic engineering', *Mol BioSyst*, pp. 38–59. doi: 10.1039/c4mb00443d.

Belliot, G., Sosnovtsev, S. V., Chang, K. O., McPhie, P. and Green, K. Y. (2008) 'Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein', *Virology*, 374(1), pp. 33–49. doi: 10.1016/j.virol.2007.12.028.

Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V. and Estes, M. K. (2002) 'Structural requirements for the assembly of Norwalk virus-like particles', *J Virol*, 76(8), pp. 4044–4055. doi: 10.1128/jvi.76.8.4044-4055.2002.

Bhella, D., Gatherer, D., Chaudhry, Y., Pink, R. and Goodfellow, I. G. (2008) 'Structural insights into calicivirus attachment and uncoating', *J Virol*, 82(16), pp. 8051–8058. doi: 10.1128/jvi.00550-08.

Birnbaum, F. and Nassal, M. (1990) 'Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein', *J Virol*, 64(7), pp. 3319–3330. doi: 10.1128/jvi.64.7.3319-3330.1990.

Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N. and Stahlberg, H. (2017) 'Focus: The interface between data collection and data processing in cryo-EM', *J Struct Biol*, 198(2), pp. 124–133. doi: 10.1016/j.jsb.2017.03.007.

Blasi, E., Barluzzi, R., Bocchini, V., Mazzolla, R. and Bistoni, F. (1990) 'Immortalization of murine microglial cells by a v-raf/v-myc carrying retrovirus', *J Neuroimmunol*, 27(2–3), pp. 229–237. doi: 10.1016/0165-5728(90)90073-V.

Blazevic, V., Lappalainen, S., Nurminen, K., Huhti, L. and Vesikari, T. (2011) 'Norovirus VLPs and rotavirus VP6 protein as combined vaccine for childhood gastroenteritis', *Vaccine*, 29(45), pp. 8126–8133. doi: 10.1016/j.vaccine.2011.08.026.

Bok, K., Prikhodko, V. G., Green, K. Y. and Sosnovtsev, S. V. (2009) 'Apoptosis in murine norovirus-infected RAW264.7 cells is associated with downregulation of survivin', *J Virol*, 83(8), pp. 3647–3656. doi: 10.1128/jvi.02028-08.

Bok, K. and Green, K. Y. (2012) 'Norovirus gastroenteritis in immunocompromised patients', *N Engl J Med*, 367(22), pp. 2126–2132. doi: 10.1056/nejmra1207742.

Borg, C., Jahun, A. S., Thorne, L., Sorgeloos, F., Bailey, D. and

Goodfellow, I. G. (2021) ‘Murine norovirus virulence factor 1 (VF1) protein contributes to viral fitness during persistent infection’, *J Gen Virol*, 102(9), p. 001651. doi: 10.1099/jgv.0.001651.

Borrego, F. (2013) ‘The CD300 molecules: an emerging family of regulators of the immune system’, *Blood*, 121(11), pp. 1951–1960. doi: 10.1182/blood-2012-09-435057.

Böttcher, B., Wynne, S. A. and Crowther, R. A. (1997) ‘Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy’, *Nature*, 386(6620), pp. 88–91. doi: 10.1038/386088a0.

Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R. and Grigorieff, N. (2012) ‘Beam-induced motion of vitrified specimen on holey carbon film’, *J Struct Biol*, 177(3), pp. 630–637. doi: 10.1016/j.jsb.2012.02.003.

Brown, A. L., Francis, M. J., Hastings, G. Z., Parry, N. R., Barnett, P. V., Rowlands, D. J. and Clarke, B. E. (1991) ‘Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted’, *Vaccine*, 9(8), pp. 595–601. doi: 10.1016/0264-410X(91)90248-5.

Brune, K. D., Leneghan, D. B., Brian, I. J., Ishizuka, A. S., Bachmann, M. F., Draper, S. J., Biswas, S. and Howarth, M. (2016) ‘Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization’, *Sci Rep*, 6, p. 19234. doi: 10.1038/srep19234.

Bubeck, D., Filman, D. J., Cheng, N., Steven, A. C., Hogle, J. M. and Belnap, D. M. (2005) ‘The structure of the poliovirus 135S cell entry intermediate at 10-angstrom resolution reveals the location of an externalized polypeptide that binds to membranes.’, *J Virol*, 79(12), pp. 7745–55. doi: 10.1128/JVI.79.12.7745-7755.2005.

Burri, D. J., da Palma, J. R., Kunz, S. and Pasquato, A. (2012) ‘Envelope glycoprotein of arenaviruses’, *Viruses*, 4(10), pp. 2162–2181. doi: 10.3390/v4102162.

Burroughs, J. N. and Brown, F. (1978) ‘Presence of a covalently linked protein on calicivirus RNA’, *Journal of General Virology*, 41(2), pp. 443–446. doi: 10.1099/0022-1317-41-2-443.

Byrne, M. J., Steele, J. F. C., Hesketh, E. L., Walden, M., Thompson, R. F., Lomonossoff, G. P. and Ranson, N. A. (2019) ‘Combining Transient Expression and Cryo-EM to Obtain High-Resolution Structures of Luteovirid Particles’, *Structure*, 27(12), pp. 1761-1770.e3. doi: 10.1016/j.str.2019.09.010.

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., *et al.* (2012) ‘Movies of ice-embedded particles enhance resolution in electron cryo-microscopy’, *Structure*, 20(11), pp. 1823–1828. doi: 10.1016/j.str.2012.08.026.

Cantrelle, F. X., Boll, E., Brier, L., Moschidi, D., Belouzard, S., Landry, V., Leroux, F., Dewitte, F., Landrieu, I., Dubuisson, J., *et al.* (2021) ‘NMR Spectroscopy of the Main Protease of SARS-CoV-2 and Fragment-Based Screening Identify Three Protein Hotspots and an Antiviral Fragment’, *Angew Chem Int Ed Engl*, 60(48), pp. 25428–25435. doi: 10.1002/anie.202109965.

Capizzi, T., Makari-Judson, G., Steingart, R. and Mertens, W. C. (2011) ‘Chronic diarrhea associated with persistent norovirus excretion in patients with chronic lymphocytic leukemia: report of two cases’, *BMC Infect Dis*, 11, p. 131. doi: 10.1186/1471-2334-11-131.

Caspar, D. L. and Klug, A. (1962) 'Physical principles in the construction of regular viruses', *Cold Spring Harb Symp Quant Biol*, 27, pp. 1–24. doi: 10.1101/SQB.1962.027.001.005.

Cereghino, J. L. and Cregg, J. M. (2000) 'Heterologous protein expression in the methylotrophic yeast Pichia pastoris', *FEMS Microbiol Rev*, 24(1), pp. 45–66. doi: 10.1111/j.1574-6976.2000.tb00532.x.

Chackerian, B., Lowy, D. R. and Schiller, J. T. (2001) 'Conjugation of a self-antigen to papillomavirus-like particles allows for efficient induction of protective autoantibodies', *J Clin Invest*, 108(3), pp. 415–423. doi: 10.1172/JCI11849.

Chang, J. Y., Gorzelnik, K. V., Thongchol, J. and Zhang, J. (2022) 'Structural Assembly of Qβ Virion and Its Diverse Forms of Virus-like Particles', *Viruses*, 14(2), p. 225. doi: 10.3390/v14020225.

Chaudhry, Y., Nayak, A., Bordeleau, M. E., Tanaka, J., Pelletier, J., Belsham, G. J., Roberts, L. O. and Goodfellow, I. G. (2006) 'Caliciviruses differ in their functional requirements for eIF4F components', *J Biol Chem*, 281(35), pp. 25315–25325. doi: 10.1074/jbc.M602230200.

Chaudhry, Y., Skinner, M. A. and Goodfellow, I. G. (2007) 'Recovery of genetically defined murine norovirus in tissue culture by using a fowlpox virus expressing T7 RNA polymerase', *J Gen Virol*, 88(8), pp. 2091–2100. doi: 10.1099/vir.0.82940-0.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography.', *Acta Crystallogr D Biol Crystallogr*, 66(Pt 1), pp. 12–21. doi: 10.1107/S0907444909042073.

Chen, Z., Sosnovtsev, S. V., Bok, K., Parra, G. I., Makiya, M., Agulto, L., Green, K. Y. and Purcell, R. H. (2013) 'Development of Norwalk virus-specific monoclonal antibodies with therapeutic potential for the treatment of Norwalk virus gastroenteritis', *J Virol*, 87(17), pp. 9547–9557. doi: 10.1128/jvi.01376-13.

Chhabra, P., de Graaf, M., Parra, G. I., Chan, M. C., Green, K., Martella, V., Wang, Q., White, P. A., Katayama, K., Vennema, H., *et al.* (2019) 'Updated classification of norovirus genogroups and genotypes', *J Gen Virol*, 100(10), pp. 1393–1406. doi: 10.1099/JGV.0.001318.

Chiang, J. Y. (2009) 'Bile acids: regulation of synthesis', *J Lipid Res*, 50(10), pp. 1955–1966. doi: 10.1194/jlr.R900010-JLR200.

Choi, J.M., Hutson, A. M., Estes, M. K. and Prasad, B. V. (2008) 'Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus', *Proc Natl Acad Sci USA*, 105(27), pp. 9175–80. doi: 10.1073/pnas.0803275105.

Chung, L., Bailey, D., Leen, E. N., Emmott, E. P., Chaudhry, Y., Roberts, L. O., Curry, S., Locker, N. and Goodfellow, I. G. (2014) 'Norovirus translation requires an interaction between the C Terminus of the genome-linked viral protein VPg and eukaryotic translation initiation factor 4G', *J Biol Chem*, 289(31), pp. 21738–21750. doi: 10.1074/jbc.M114.550657.

Cohen, B. J. and Richmond, J. E. (1982) 'Electron microscopy of hepatitis B core antigen synthesized in E. coli', *Nature*, 296(5858), pp. 677–678. doi: 10.1038/296677a0.

Comas-Garcia, M. (2019) 'Packaging of Genomic RNA in Positive-

Sense Single-Stranded RNA Viruses: A Complex Story', *Viruses*, 11(3), p. 253. doi: 10.3390/v11030253.

Conley, M., Mcelwee, M., Azmi, L., Gabrielsen, M., Byron, O., Goodfellow, I. G. and Bhella, D. (2018) 'Calicivirus VP2 forms a portal to mediate endosome escape', *BioRxiv [Preprint]*, p. 397901. doi: 10.1101/397901.

Conley, M. J., McElwee, M., Azmi, L., Gabrielsen, M., Byron, O., Goodfellow, I. G. and Bhella, D. (2019) 'Calicivirus VP2 forms a portal-like assembly following receptor engagement', *Nature*, 565(7739), pp. 377–381. doi: 10.1038/s41586-018-0852-1.

Conway, J. F., Cheng, N., Zlotnick, A., Stahl, S. J., Wingfield, P. T., Belnap, D. M., Kanngiesser, U., Noah, M. and Steven, A. C. (1998) 'Hepatitis B virus capsid: localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens', *J Mol Biol*, 279(5), pp. 1111–1121. doi: 10.1006/jmbi.1998.1845.

Costantini, V. P., Whitaker, T., Barclay, L., Lee, D., McBrayer, T. R., Schinazi, R. F. and Vinjé, J. (2012) 'Antiviral activity of nucleoside analogues against norovirus', *Antivir Ther*, 17(6), pp. 981–991. doi: 10.3851/IMP2229.

Creutznacher, R., Maaß, T., Dülfer, J., Feldmann, C., Hartmann, V., Knickmann, J., Torben Westermann, L., Smith, T. J., Uetrecht, C., Mallagaray, A., *et al.* (2021a) 'Murine norovirus capsid plasticity – Glycochenodeoxycholic acid stabilizes P-domain dimers and triggers escape from antibody recognition', *BioRxiv [Preprint]*, p. 2021.02.27.433148. doi: 10.1101/2021.02.27.433148.

Creutznacher, R., Maass, T., Ogrissek, P., Wallmann, G., Feldmann, C., Peters, H., Lingemann, M., Taube, S., Peters, T. and Mallagaray, A. (2021b) 'NMR Experiments Shed New Light on Glycan Recognition by Human and Murine Norovirus Capsid Proteins', *Viruses*, 13(3), p. 416. doi: 10.3390/v13030416.

Crosby, J. and Crump, M. P. (2012) 'The structural role of the carrier protein - active controller or passive carrier', *Nat Prod Rep*, pp. 1111–1137. doi: 10.1039/c2np20062g.

Crowther, R. A., Kiselev, N. A., Böttcher, B., Berriman, J. A., Borisova, G. P., Ose, V. and Pumpens, P. (1994) 'Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy', *Cell*, 77(6), pp. 943–950. doi: 10.1016/0092-8674(94)90142-2.

Cui, Z., Gorzelnik, K.V., Chang, J.Y., Langlais, C., Jakana, J., Young, R. and Zhang, J. (2017) 'Structures of Qβ virions, virus-like particles, and the Qβ-MurA complex reveal internal coat proteins and the mechanism of host lysis.', *Proc Natl Acad Sci USA*, 114(44), pp. 11697–11702. doi: 10.1073/pnas.1707102114.

Curry, S., Chow, M. and Hogle, J. M. (1996) 'The poliovirus 135S particle is infectious.', *J Virol*, 70(10), pp. 7125–31. doi: 10.1128/JVI.70.10.7125-7131.1996.

D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M. and Kühlbrandt, W. (2019) 'Protein denaturation at the air-water interface and how to prevent it', *eLife*, 8, p. e42747. doi: 10.7554/eLife.42747.

Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z. H. and Sun, R. (2017) 'In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus', *Nature*, 541(7635), pp. 112–116. doi:

10.1038/nature20589.

Damalanka, V. C., Kim, Y., Alliston, K. R., Weerawarna, P. M., Galasiti Kankanamalage, A. C., Lushington, G. H., Mehzabeen, N., Battaile, K. P., Lovell, S., Chang, K. O., *et al.* (2016) 'Oxadiazole-Based Cell Permeable Macrocyclic Transition State Inhibitors of Norovirus 3CL Protease', *J Med Chem*, 59(5), pp. 1899–1913. doi: 10.1021/acs.jmedchem.5b01464.

Daughenbaugh, K. F., Wobus, C. E. and Hardy, M. E. (2006) 'VPg of murine norovirus binds translation initiation factors in infected cells', *Virol J*, 3, p. 33. doi: 10.1186/1743-422X-3-33.

Dinnon, K. H., Leist, S. R., Schäfer, A., Edwards, C. E., Martinez, D. R., Montgomery, S. A., West, A., Yount, B. L., Hou, Y. J., Adams, L. E., *et al.* (2020) 'A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures', *Nature*, 586(7830), pp. 560–566. doi: 10.1038/s41586-020-2708-8.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. and Baker, N. A. (2007) 'PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations', *Nucleic Acids Res*, 35(Web Server issue), pp. W522–W525. doi: 10.1093/nar/gkm276.

Donaldson, B., Lateef, Z., Walker, G. F., Young, S. L. and Ward, V. K. (2018) 'Virus-like particle vaccines: immunology and formulation for clinical translation', *Expert Rev Vaccines*, 17(9), pp. 833–849. doi: 10.1080/14760584.2018.1516552.

Drulyte, I., Johnson, R. M., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. and Thompson, R. F. (2018) 'Approaches to altering particle distributions in cryo-electron microscopy sample preparation', *Acta Crystallogr D Struct Biol*, 74(6), pp. 560–571. doi: 10.1107/S2059798318006496.

Du, Y. X. and Chen, X. P. (2020) 'Favipiravir: Pharmacokinetics and Concerns About Clinical Trials for 2019-nCoV Infection', *Clin Pharmacol Ther*, 108(2), pp. 242–247. doi: 10.1002/cpt.1844.

Duizer, E., Schwab, K. J., Neill, F. H., Atmar, R. L., Koopmans, M. P. G. and Estes, M. K. (2004) 'Laboratory efforts to cultivate noroviruses', *J Gen Virol*, 85(1), pp. 79–87. doi: 10.1099/vir.0.19478-0.

Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) 'Analysis of membrane and surface protein sequences with the hydrophobic moment plot', *J Mol Biol*, 179(1), pp. 125–142. doi: 10.1016/0022-2836(84)90309-7.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', *Acta Crystallogr D Biol Crystallogr*, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Enria, D. A., Briggiler, A. M. and Sánchez, Z. (2008) 'Treatment of Argentine hemorrhagic fever', *Antiviral Res*, 78(1), pp. 132–139. doi: 10.1016/j.antiviral.2007.10.010.

Esposito, S. and Principi, N. (2020) 'Norovirus Vaccine: Priorities for Future Research and Development', *Front Immunol*, 11, p. 1383. doi: 10.3389/fimmu.2020.01383.

Ettayebi, K., Crawford, S. E., Murakami, K., Broughman, J. R., Karandikar, U., Tenge, V. R., Neill, F. H., Blutt, S. E., Zeng, X.L., Qu, L., *et al.* (2016) 'Replication of human noroviruses in stem cell-derived human

enteroids', *Science*, 353(6306), pp. 1387–1393. doi: 10.1126/science.aaf5211.

Ettayebi, K. and Hardy, M. E. (2003) 'Norwalk virus nonstructural protein p48 forms a complex with the SNARE regulator VAP-A and prevents cell surface expression of vesicular stomatitis virus G protein', *J Virol*, 77(21), pp. 11790–11797. doi: 10.1128/jvi.77.21.11790-11797.2003.

Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D. and Brutscher, B. (2012) 'Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A', *J Mol Biol*, 420(4–5), pp. 310–323. doi: 10.1016/j.jmb.2012.04.023.

Fichtlscherer, F., Wellein, C., Mittag, M. and Schweizer, E. (2000) 'A novel function of yeast fatty acid synthase. Subunit α is capable of self-pantetheinylation', *Eur J Biochem*, 267(9), pp. 2666–2671. doi: 10.1046/j.1432-1327.2000.01282.x.

Fifis, T., Gamvrellis, A., Crimeen-Irwin, B., Pietersz, G. A., Li, J., Mottram, P. L., McKenzie, I. F. and Plebanski, M. (2004) 'Size-dependent immunogenicity: therapeutic and protective properties of nano-vaccines against tumors', *J Immunol*, 173(5), pp. 3148–3154. doi: 10.4049/jimmunol.173.5.3148.

Fox, H., Knowlson, S., Minor, P. D. and Macadam, A. J. (2017) 'Genetically Thermo-Stabilised, Immunogenic Poliovirus Empty Capsids; a Strategy for Non-replicating Vaccines.', *PLoS Pathog*, 13(1), p. e1006117. doi: 10.1371/journal.ppat.1006117.

Francis, M. J., Hastings, G. Z., Brown, A. L., Grace, K. G., Rowlands, D. J., Brown, F. and Clarke, B. E. (1990) 'Immunological properties of hepatitis B core antigen fusion proteins', *Proc Natl Acad Sci USA*, 87(7), pp. 2545–2549. doi: 10.1073/pnas.87.7.2545.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) 'SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields', *J Struct Biol*, 116(1), pp. 190–199. doi: 10.1006/jsbi.1996.0030.

Furman, L. M., Maaty, W. S., Petersen, L. K., Ettayebi, K., Hardy, M. E. and Bothner, B. (2009) 'Cysteine protease activation and apoptosis in Murine norovirus infection', *Virol J*, 6, p. 139. doi: 10.1186/1743-422X-6-139.

Furuta, Y., Komeno, T. and Nakamura, T. (2017) 'Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase', *Proc Jpn Acad Ser B Phys Biol Sci*, 93(7), pp. 449–463. doi: 10.2183/pjab.93.027.

Gajewski, J., Buelens, F., Serdjukow, S., Janßen, M., Cortina, N., Grubmüller, H. and Grininger, M. (2017a) 'Engineering fatty acid synthases for directed polyketide production', *Nat Chem Biol*, 13(4), pp. 363–365. doi: 10.1038/nchembio.2314.

Gajewski, J., Pavlovic, R., Fischer, M., Boles, E. and Grininger, M. (2017b) 'Engineering fungal de novo fatty acid synthesis for short chain fatty acid production', *Nat Commun*, 8, p. 14650. doi: 10.1038/ncomms14650.

Galasiti Kankanamalage, A. C., Kim, Y., Weerawarna, P. M., Uy, R. A., Damalanka, V. C., Mandadapu, S. R., Alliston, K. R., Mehzabeen, N., Battaile, K. P., Lovell, S., *et al.* (2015) 'Structure-guided design and optimization of dipeptidyl inhibitors of norovirus 3CL protease. Structure-activity relationships and biochemical, X-ray crystallographic, cell-based, and in vivo studies', *J Med Chem*, 58(7), pp. 3144–3155. doi: 10.1021/jm5019934.

Garibyan, L. and Avashia, N. (2013) ‘Polymerase chain reaction’, *J Invest Dermatol*, 133(3), pp. 1–4. doi: 10.1038/jid.2013.1.

Gerondopoulos, A., Jackson, T., Monaghan, P., Doyle, N. and Roberts, L. O. (2010) ‘Murine norovirus-1 cell entry is mediated through a non-clathrin-, non-caveolae-, dynamin- and cholesterol-dependent pathway’, *J Gen Virol*, 91(6), pp. 1428–1438. doi: 10.1099/vir.0.016717-0.

Gipson, P., Mills, D. J., Wouts, R., Grininger, M., Vonck, J. and Kühlbrandt, W. (2010) ‘Direct structural insight into the substrate-shuttling mechanism of yeast fatty acid synthase by electron cryomicroscopy’, *Proc Natl Acad Sci USA*, 107(20), pp. 9164–9169. doi: 10.1073/pnas.0913547107.

Glass, R. I., Parashar, U. D. and Estes, M. K. (2009) ‘Norovirus gastroenteritis’, *N Engl J Med*, 361(18), pp. 1776–1785. doi: 10.1056/nejmra0804575.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. and Ferrin, T. E. (2018) ‘UCSF ChimeraX: Meeting modern challenges in visualization and analysis’, *Protein Sci*, 27(1), pp. 14–25. doi: 10.1002/PRO.3235.

Goetschius, D. J., Lee, H. and Hafenstein, S. (2019a) ‘CryoEM reconstruction approaches to resolve asymmetric features’, *Adv Virus Res*, 105, pp. 73–91. doi: 10.1016/bs.aivir.2019.07.007.

Goetschius, D. J., Parrish, C. R. and Hafenstein, S. (2019b) ‘Asymmetry in icosahedral viruses’, *Curr Opin Virol*, pp. 67–73. doi: 10.1016/j.coviro.2019.05.006.

Grant, A., Seregin, A., Huang, C., Kolokoltsova, O., Brasier, A., Peters, C. and Paessler, S. (2012) ‘Junín virus pathogenesis and virus replication’, *Viruses*, pp. 2317–2339. doi: 10.3390/v4102317.

Grau, K. R., Roth, A. N., Zhu, S., Hernandez, A., Colliou, N., DiVita, B. B., Philip, D. T., Riffe, C., Giasson, B., Wallet, S. M., *et al.* (2017) ‘The major targets of acute norovirus infection are immune cells in the gut-associated lymphoid tissue’, *Nat Microbiol*, 2(12), pp. 1586–1591. doi: 10.1038/s41564-017-0057-7.

Graziano, V. R., Wei, J. and Wilen, C. B. (2019) ‘Norovirus Attachment and Entry’, *Viruses*, 11(6), p. 495. doi: 10.3390/v11060495.

Green, K. Y., Kaufman, S. S., Nagata, B. M., Chaimongkol, N., Kim, D. Y., Levenson, E. A., Tin, C. M., Yardley, A. B., Johnson, J. A., Barletta, A. B. F., *et al.* (2020) ‘Human norovirus targets enteroendocrine epithelial cells in the small intestine’, *Nat Commun*, 11(1), p. 2759. doi: 10.1038/s41467-020-16491-3.

Greene, S. E. and Reid, A. (2013) ‘Viruses Throughout Life & Time: Friends, Foes, Change Agents: A Report on an American Academy of Microbiology Colloquium San Francisco // July 2013’, *American Academy of Microbiology Colloquia Reports*. doi: 10.1128/AAMCol.Jul.2013

Guix, S., Asanaka, M., Katayama, K., Crawford, S. E., Neill, F. H., Atmar, R. L. and Estes, M. K. (2007) ‘Norwalk virus RNA is infectious in mammalian cells’, *J Virol*, 81(22), pp. 12238–12248. doi: 10.1128/jvi.01489-07.

Guo, J., Zhou, A., Sun, X., Sha, W., Ai, K., Pan, G., Zhou, C., Zhou, H., Cong, H. and He, S. (2019) ‘Immunogenicity of a Virus-Like-Particle Vaccine Containing Multiple Antigenic Epitopes of *Toxoplasma gondii* Against Acute and Chronic Toxoplasmosis in Mice’, *Front Immunol*, 10, p. 592. doi:

10.3389/fimmu.2019.00592.

Guo, L., Wang, J., Zhou, H., Si, H., Wang, M., Song, J., Han, B., Shu, Y., Ren, L., Qu, J., *et al.* (2008) 'Intranasal administration of a recombinant adenovirus expressing the norovirus capsid protein stimulates specific humoral, mucosal, and cellular immune responses in mice', *Vaccine*, 26(4), pp. 460–468. doi: 10.1016/j.vaccine.2007.11.039.

Haga, K., Fujimoto, A., Takai-Todaka, R., Miki, M., Doan, Y. H., Murakami, K., Yokoyama, M., Murata, K., Nakanishi, A. and Katayama, K. (2016) 'Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells', *Proc Natl Acad Sci USA*, 113(41), pp. E6248–E6255. doi: 10.1073/pnas.1605575113.

Hakomori, S. (1999) 'Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer', *Biochim Biophys Acta*, pp. 247–266. doi: 10.1016/S0304-4165(99)00183-X.

Hansman, G. S., Biertümpfel, C., Georgiev, I., McLellan, J. S., Chen, L., Zhou, T., Katayama, K. and Kwong, P. D. (2011) 'Crystal structures of GII.10 and GII.12 norovirus protruding domains in complex with histo-blood group antigens reveal details for a potential site of vulnerability', *J Virol*, 85(13), pp. 6687–6701. doi: 10.1128/JVI.00246-11.

Hansman, G. S., Taylor, D. W., McLellan, J. S., Smith, T. J., Georgiev, I., Tame, J. R., Park, S.Y., Yamazaki, M., Gondaira, F., Miki, M., *et al.* (2012) 'Structural basis for broad detection of genogroup II noroviruses by a monoclonal antibody that binds to a site occluded in the viral particle', *J Virol*, 86(7), pp. 3635–3646. doi: 10.1128/JVI.06868-11.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. and Bricogne, G. (1978) 'Tomato bushy stunt virus at 2.9 Å resolution', *Nature*, 276(5686), pp. 368–373. doi: 10.1038/276368a0.

Heil, C. S., Wehrheim, S. S., Paithankar, K. S. and Grininger, M. (2019) 'Fatty Acid Biosynthesis: Chain-Length Regulation and Control', *ChemBioChem*, 20(18), pp. 2298–2321. doi: 10.1002/cbic.201800809.

Herod, M. R., Salim, O., Skilton, R. J., Prince, C. A., Ward, V. K., Lambden, P. R. and Clarke, I. N. (2014) 'Expression of the murine norovirus (MNV) ORF1 polyprotein is sufficient to induce apoptosis in a virus-free cell model', *PLoS One*, 9(3), p. e90679. doi: 10.1371/journal.pone.0090679.

Hierholzer, J. C. and Killington, R. A. (1996) 'Virus isolation and quantitation', *Virology Methods Manual*, pp. 25–46. doi: 10.1016/B978-012465330-6/50003-8.

Holland, D., Thomson, L., Mahmoudzadeh, N. and Khaled, A. (2020) 'Estimating deaths from foodborne disease in the UK for 11 key pathogens', *BMJ Open Gastroenterol*, 7(1), p. e000377. doi: 10.1136/bmjgast-2020-000377.

Hurdiss, D. L., Frank, M., Snowden, J. S., Macdonald, A. and Ranson, N. A. (2018) 'The Structure of an Infectious Human Polyomavirus and Its Interactions with Cellular Receptors', *Structure*, 26(6), pp. 839-847.e3. doi: 10.1016/j.str.2018.03.019.

Hwang, S., Alhatlani, B., Arias, A., Caddy, S. L., Christodoulou, C., Cunha, J. B., Emmott, E., Gonzalez-Hernandez, M., Kolawole, A., Lu, J., *et al.* (2014) 'Murine norovirus: propagation, quantification, and genetic manipulation', *Curr Protoc Microbiol*, 33, pp. 15K.2.1–61. doi: 10.1002/9780471729259.mc15k02s33.

Hyde, J. L. and Mackenzie, J. M. (2010) 'Subcellular localization of the MNV-1 ORF1 proteins and their potential roles in the formation of the MNV-1 replication complex', *Virology*, 406(1), pp. 138–148. doi: 10.1016/j.virol.2010.06.047.

Ilca, S. L., Sun, X., El Omari, K., Kotecha, A., de Haas, F., DiMaio, F., Grimes, J. M., Stuart, D. I., Poranen, M. M. and Huiskonen, J. T. (2019) 'Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus', *Nature*, 570(7760), pp. 252–256. doi: 10.1038/s41586-019-1229-9.

International Committee on Taxonomy of Viruses Executive Committee (2020) 'The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks', *Nat Microbiol*, pp. 668–674. doi: 10.1038/s41564-020-0709-x.

Ito, S., Takeshita, S., Nezu, A., Aihara, Y., Usuku, S., Noguchi, Y. and Yokota, S. (2006) 'Norovirus-associated encephalopathy', *Pediatr Infect Dis J*, 25(7), pp. 651–652. doi: 10.1097/01.inf.0000225789.92512.6d.

Jenni, S., Leibundgut, M., Maier, T. and Ban, N. (2006) 'Architecture of a fungal fatty acid synthase at 5 Å resolution', *Science*, 311(5765), pp. 1263–1267. doi: 10.1126/science.1123251.

Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B. and Ban, N. (2007) 'Structure of fungal fatty acid synthase and implications for iterative substrate shuttling', *Science*, 316(5822), pp. 254–261. doi: 10.1126/science.1138248.

Jiang, W. and Chiu, W. (2001) 'Web-based Simulation for Contrast Transfer Function and Envelope Functions', *Microsc Microanal*, 7(4), pp. 329–334. doi: 10.1007/s10005-001-0004-4.

Jiang, X., Wang, M., Graham, D. Y. and Estes, M. K. (1992) 'Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein', *J Virol*, 66(11), pp. 6527–6532. doi: 10.1128/jvi.66.11.6527-6532.1992.

Johansson, P., Mulinacci, B., Koestler, C., Vollrath, R., Oesterhelt, D. and Grininger, M. (2009) 'Multimeric options for the auto-activation of the Saccharomyces cerevisiae FAS type I megasynthase', *Structure*, 17(8), pp. 1063–1074. doi: 10.1016/j.str.2009.06.014.

Jones, M. K., Watanabe, M., Zhu, S., Graves, C. L., Keyes, L. R., Grau, K. R., Gonzalez-Hernandez, M. B., Iovine, N. M., Wobus, C. E., Vinjé, J., *et al.* (2014) 'Enteric bacteria promote human and mouse norovirus infection of B cells.', *Science*, 346(6210), pp. 755–759. doi: 10.1126/science.1257147.

Jones, T. A. and Liljas, L. (1984) 'Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution', *J Mol Biol*, 177(4), pp. 735–767. doi: 10.1016/0022-2836(84)90047-0.

Joppe, M., D'Imprima, E., Salustros, N., Paithankar, K. S., Vonck, J., Grininger, M. and Kühlbrandt, W. (2020) 'The resolution revolution in cryoEM requires high-quality sample preparation: a rapid pipeline to a high-resolution map of yeast fatty acid synthase', *IUCrJ*, 7(2), pp. 220–227. doi: 10.1107/S2052252519017366.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Jung, J., Grant, T., Thomas, D. R., Diehnelt, C. W., Grigorieff, N. and

Joshua-Tor, L. (2019) ‘High-resolution cryo-EM structures of outbreak strain human norovirus shells reveal size variations’, *Proc Natl Acad Sci USA*, 116(26), pp. 12828–12832. doi: 10.1073/pnas.1903562116.

Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R. and Chanock, R. M. (1972) ‘Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis’, *J Virol*, 10(5), pp. 1075–1081. doi: 10.1128/jvi.10.5.1075-1081.1972.

Kapikian, A. Z. (2000) ‘The discovery of the 27-nm Norwalk virus: an historic perspective’, *J Infect Dis*, 181(Suppl 2), pp. S295–S302. doi: 10.1086/315584.

Karandikar, U. C., Crawford, S. E., Ajami, N. J., Murakami, K., Kou, B., Ettayebi, K., Papanicolaou, G. A., Jongwutiwes, U., Perales, M. A., Shia, J., *et al.* (2016) ‘Detection of human norovirus in intestinal biopsies from immunocompromised transplant patients’, *J Gen Virol*, 97(9), pp. 2291–2300. doi: 10.1099/jgv.0.000545.

Karangwa, C. K., Parra, G. I., Bok, K., Johnson, J. A., Levenson, E. A. and Green, K. Y. (2017) ‘Sequential Gastroenteritis Outbreaks in a Single Year Caused by Norovirus Genotypes GII.2 and GII.6 in an Institutional Setting’, *Open Forum Infect Dis*, 4(4), p. ofx236. doi: 10.1093/ofid/ofx236.

Karst, S. M., Wobus, C. E., Lay, M., Davidson, J. and Virgin IV, H. W. (2003) ‘STAT1-dependent innate immunity to a Norwalk-like virus’, *Science*, 299(5612), pp. 1575–1578. doi: 10.1126/science.1077905.

Kastritis, P. L., O’Reilly, F. J., Bock, T., Li, Y., Rogon, M. Z., Buczak, K., Romanov, N., Betts, M. J., Bui, K. H., Hagen, W. J., *et al.* (2017) ‘Capturing protein communities by structural proteomics in a thermophilic eukaryote’, *Mol Syst Biol*, 13(7), p. 936. doi: 10.15252/msb.20167412.

Katpally, U., Wobus, C. E., Dryden, K., Virgin IV, H. W. and Smith, T. J. (2008) ‘Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles’, *J Virol*, 82(5), pp. 2079–2088. doi: 10.1128/JVI.02200-07.

Katpally, U., Voss, N. R., Cavazza, T., Taube, S., Rubin, J. R., Young, V. L., Stuckey, J., Ward, V. K., Virgin IV, H. W., Wobus, C. E., *et al.* (2010) ‘High-resolution cryo-electron microscopy structures of murine norovirus 1 and rabbit hemorrhagic disease virus reveal marked flexibility in the receptor binding domains’, *J Virol*, 84(11), pp. 5836–5841. doi: 10.1128/JVI.00314-10.

Kausche, G. A., Pfankuch, E. and Ruska, H. (1939) ‘Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop’, *Naturwissenschaften*, 27, pp. 292–299. doi: 10.1007/BF01493353.

Keeble, A. H., Banerjee, A., Ferla, M. P., Reddington, S. C., Anuar, I. N. A. K. and Howarth, M. (2017) ‘Evolving Accelerated Amidation by SpyTag/SpyCatcher to Analyze Membrane Dynamics’, *Angew Chem Int Ed Engl*, 56(52), pp. 16521–16525. doi: 10.1002/anie.201707623.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. (2015) ‘The Phyre2 web portal for protein modeling, prediction and analysis’, *Nat Protoc*, 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.

Kempf, B., Edgar, J. D., Mc Caughey, C. and Devlin, L. A. (2017) ‘Nitazoxanide Is an Ineffective Treatment of Chronic Norovirus in Patients With X-Linked Agammaglobulinemia and May Yield False-Negative Polymerase Chain Reaction Findings in Stool Specimens’, *J Infect Dis*, 215(3), pp. 486–487. doi: 10.1093/infdis/jiw497.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958) 'A three-dimensional model of the myoglobin molecule obtained by x-ray analysis', *Nature*, 181(4610), pp. 662–666. doi: 10.1038/181662a0.

Kilic, T., Koromyslova, A. and Hansman, G. S. (2019) 'Structural Basis for Human Norovirus Capsid Binding to Bile Acids', *J Virol*, 93(2), pp. e01581-18. doi: 10.1128/JVI.01581-18.

Kim, M. J., Kim, Y. J., Lee, J. H., Lee, J. S., Kim, J. H., Cheon, D. S., Jeong, H. S., Koo, H. H., Sung, K. W., Yoo, K. H., *et al.* (2011) 'Norovirus: a possible cause of pneumatosis intestinalis', *J Pediatr Gastroenterol Nutr*, 52(3), pp. 314–318. doi: 10.1097/MPG.0b013e3181ebfa01.

Kimanius, D., Forsberg, B. O., Scheres, S. H. and Lindahl, E. (2016) 'Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2', *eLife*, 5, p. e18722. doi: 10.7554/eLife.18722.

King, C. K., Glass, R., Bresee, J. S. and Duggan, C. (2003) 'Managing acute gastroenteritis among children: oral rehydration, maintenance, and nutritional therapy.', *MMWR Recomm Rep*, 52(RR-16), pp. 1–16. doi: 10.1542/peds.114.2.507.

Klebl, D. P., Gravett, M. S. C., Kontziampasis, D., Wright, D. J., Bon, R. S., Monteiro, D. C. F., Trebbin, M., Sobott, F., White, H. D., Darrow, M. C., *et al.* (2020) 'Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales', *Structure*, 28(11), pp. 1238–1248.e4. doi: 10.1016/j.str.2020.07.018.

Kolawole, A. O., Xia, C., Li, M., Gamez, M., Yu, C., Rippinger, C. M., Yucha, R. E., Smith, T. J. and Wobus, C. E. (2014) 'Newly isolated mAbs broaden the neutralizing epitope in murine norovirus', *J Gen Virol*, 95(Pt 9), pp. 1958–1968. doi: 10.1099/vir.0.066753-0.

Kolawole, A. O., Rocha-Pereira, J., Elftman, M. D., Neyts, J. and Wobus, C. E. (2016) 'Inhibition of human norovirus by a viral polymerase inhibitor in the B cell culture system and in the mouse model', *Antiviral Res*, 132, pp. 46–49. doi: 10.1016/j.antiviral.2016.05.011.

Koning, R. I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., Carazo, J. M. and Koster, A. J. (2016) 'Asymmetric cryo-EM reconstructions of phage MS2 reveals genome structure in situ', *Nat Commun*, 7, p. 12524. doi: 10.1038/ncomms12524.

Koromyslova, A., Tripathi, S., Morozov, V., Schroten, H. and Hansman, G. S. (2017) 'Human norovirus inhibition by a human milk oligosaccharide', *Virology*, 508, pp. 81–89. doi: 10.1016/j.virol.2017.04.032.

Koromyslova, A. D. and Hansman, G. S. (2015) 'Nanobody binding to a conserved epitope promotes norovirus particle disassembly', *J Virol*, 89(5), pp. 2718–2730. doi: 10.1128/jvi.03176-14.

Koromyslova, A. D. and Hansman, G. S. (2017) 'Nanobodies targeting norovirus capsid reveal functional epitopes and potential mechanisms of neutralization', *PLoS Pathog*, 13(11), p. e1006636. doi: 10.1371/journal.ppat.1006636.

Koromyslova, A. D., White, P. A. and Hansman, G. S. (2015) 'Treatment of norovirus particles with citrate', *Virology*, 485, pp. 199–204. doi: 10.1016/j.virol.2015.07.009.

Krissinel, E. and Henrick, K. (2004) 'Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions',

Acta Crystallogr D Biol Crystallogr, 60(12 Pt 1), pp. 2256–2268. doi: 10.1107/S0907444904026460.

Krissinel, E. and Henrick, K. (2007) 'Inference of macromolecular assemblies from crystalline state', *J Mol Biol*, 372(3), pp. 774–797. doi: 10.1016/J.JMB.2007.05.022.

Kühlbrandt, W. (2014) 'The resolution revolution', *Science*, 343, pp. 1443–1444. doi: 10.1126/science.1251652.

Lappalainen, S., Pastor, A. R., Malm, M., López-Guerrero, V., Esquivel-Guadarrama, F., Palomares, L. A., Vesikari, T. and Blazevic, V. (2015) 'Protection against live rotavirus challenge in mice induced by parenteral and mucosal delivery of VP6 subunit rotavirus vaccine', *Arch Virol*, 160(8), pp. 2075–2078. doi: 10.1007/s00705-015-2461-8.

Lee, H., Shingler, K. L., Organtini, L. J., Ashley, R. E., Makhov, A. M., Conway, J. F. and Hafenstein, S. (2016) 'The novel asymmetric entry intermediate of a picornavirus captured with nanodiscs', *Sci Adv*, 2(8), p. e1501929. doi: 10.1126/sciadv.1501929.

Leen, E. N., Kwok, K. Y., Birtley, J. R., Simpson, P. J., Subba-Reddy, C. V., Chaudhry, Y., Sosnovtsev, S. V., Green, K. Y., Prater, S. N., Tong, M., *et al.* (2013) 'Structures of the compact helical core domains of feline calicivirus and murine norovirus VPg proteins', *J Virol*, 87(10), pp. 5318–5330. doi: 10.1128/jvi.03151-12.

Leen, E. N., Sorgeloos, F., Correia, S., Chaudhry, Y., Cannac, F., Pastore, C., Xu, Y., Graham, S. C., Matthews, S. J., Goodfellow, I. G., *et al.* (2016) 'A Conserved Interaction between a C-Terminal Motif in Norovirus VPg and the HEAT-1 Domain of eIF4G Is Essential for Translation Initiation', *PLoS Pathog*, 12(1), p. e1005379. doi: 10.1371/journal.ppat.1005379.

Leibundgut, M., Jenni, S., Frick, C. and Ban, N. (2007) 'Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase', *Science*, 316(5822), pp. 288–290. doi: 10.1126/science.1138249.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N. and Lopez, R. (2015) 'The EMBL-EBI bioinformatics web and programmatic tools framework.', *Nucleic Acids Res*, 43(W1), pp. W580–W584. doi: 10.1093/nar/gkv279.

Li, X., Zhou, N., Chen, W., Zhu, B., Wang, X., Xu, B., Wang, J., Liu, H. and Cheng, L. (2017) 'Near-Atomic Resolution Structure Determination of a Cypovirus Capsid and Polymerase Complex Using Cryo-EM at 200 kV', *J Mol Biol*, 429(1), pp. 79–87. doi: 10.1016/j.jmb.2016.11.025.

Lin, Y., Fengling, L., Lianzhu, W., Yuxiu, Z. and Yanhua, J. (2014) 'Function of VP2 protein in the stability of the secondary structure of virus-like particles of genogroup II norovirus at different pH levels: function of VP2 protein in the stability of NoV VLPs', *J Microbiol*, 52(11), pp. 970–975. doi: 10.1007/s12275-014-4323-6.

Lindesmith, L. C., Beltramello, M., Swanstrom, J., Jones, T. A., Corti, D., Lanzavecchia, A. and Baric, R. S. (2015) 'Serum Immunoglobulin A Cross-Strain Blockade of Human Noroviruses', *Open Forum Infect Dis*, 2(3). doi: 10.1093/ofid/ofv084.

Liu, B. L., Lambden, P. R., Günther, H., Otto, P., Elschner, M. and Clarke, I. N. (1999) 'Molecular characterization of a bovine enteric calicivirus: relationship to the Norwalk-like viruses', *J Virol*, 73(1), pp. 819–825. doi: 10.1128/jvi.73.1.819-825.1999.

Liu, H. and Cheng, L. (2015) ‘Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus’, *Science*, 349(6254), pp. 1347–1350. doi: 10.1126/science.aaa4938.

Lomakin, I. B., Xiong, Y. and Steitz, T. A. (2007) ‘The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together’, *Cell*, 129(2), pp. 319–332. doi: 10.1016/j.cell.2007.03.013.

Long, G. W., Nobel, J., Murphy, F. A., Herrmann, K. L. and Lourie, B. (1970) ‘Experience with electron microscopy in the differential diagnosis of smallpox’, *Appl Microbiol*, 20(3), pp. 497–504. doi: 10.1128/aem.20.3.497-504.1970.

Lou, J. W., Iyer, K. R., Hasan, S. M. N., Cowen, L. E. and Mazhab-Jafari, M. T. (2019) ‘Electron cryomicroscopy observation of acyl carrier protein translocation in type I fungal fatty acid synthase’, *Sci Rep*, 9(1), p. 12987. doi: 10.1038/s41598-019-49261-3.

Lou, J. W. and Mazhab-Jafari, M. T. (2020) ‘Steric occlusion regulates proximal interactions of acyl carrier protein domain in fungal fatty acid synthase’, *Commun Biol*, 3(1), p. 274. doi: 10.1038/s42003-020-0997-y.

Luecke, S. and Paludan, S. R. (2017) ‘Molecular requirements for sensing of intracellular microbial nucleic acids by the innate immune system’, *Cytokine*, 98, pp. 4–14. doi: 10.1016/j.cyto.2016.10.003.

Macpherson, I. and Stoker, M. (1962) ‘Polyoma transformation of hamster cell clones–an investigation of genetic factors affecting cell competence’, *Virology*, 16, pp. 147–151. doi: 10.1016/0042-6822(62)90290-8.

Maier, T., Leibundgut, M., Boehringer, D. and Ban, N. (2010) ‘Structure and function of eukaryotic fatty acid synthases’, *Q Rev Biophys*, 43(3), pp. 373–422. doi: 10.1017/S0033583510000156.

Maier, T., Jenni, S. and Ban, N. (2006) ‘Architecture of mammalian fatty acid synthase at 4.5 Å resolution’, *Science*, 311(5765), pp. 1258–1262. doi: 10.1126/science.1123248.

Makbul, C., Khayenko, V., Maric, H. M. and Böttcher, B. (2021) ‘Conformational Plasticity of Hepatitis B Core Protein Spikes Promotes Peptide Binding Independent of the Secretion Phenotype’, *Microorganisms*, 9(5), p. 956. doi: 10.3390/microorganisms9050956.

Malm, M., Diessner, A., Tamminen, K., Liebscher, M., Vesikari, T. and Blazevic, V. (2019) ‘Rotavirus VP6 as an Adjuvant for Bivalent Norovirus Vaccine Produced in *Nicotiana benthamiana*’, *Pharmaceutics*, 11(5), p. 229. doi: 10.3390/pharmaceutics11050229.

Mandadapu, S. R., Gunnam, M. R., Galasiti Kankanamalage, A. C., Uy, R. A., Alliston, K. R., Lushington, G. H., Kim, Y., Chang, K. O. and Groutas, W. C. (2013) ‘Potent inhibition of norovirus by dipeptidyl α-hydroxyphosphonate transition state mimics’, *Bioorg Med Chem Lett*, 23(21), pp. 5941–5944. doi: 10.1016/j.bmcl.2013.08.073.

Marini, A., Zhou, Y., Li, Y., Taylor, I. J., Leneghan, D. B., Jin, J., Zaric, M., Mekhaiel, D., Long, C. A., Miura, K., *et al.* (2019) ‘A Universal Plug-and-Display Vaccine Carrier Based on HBsAg VLP to Maximize Effective Antibody Response’, *Front Immunol*, 10, p. 2931. doi: 10.3389/fimmu.2019.02931.

Marion, D. (2013) ‘An introduction to biological NMR spectroscopy’, *Mol Cell Proteomics*, 12(11), pp. 3006–3025. doi: 10.1074/mcp.O113.030239.

Marr, J. S. and Cathey, J. T. (2013) ‘The 1802 Saint-Domingue yellow fever epidemic and the Louisiana Purchase’, *J Public Health Manag Pract*,

19(1), pp. 77–82. doi: 10.1097/PHH.0b013e318252eea8.

Marsian, J., Fox, H., Bahar, M. W., Kotecha, A., Fry, E. E., Stuart, D. I., Macadam, A. J., Rowlands, D. J. and Lomonossoff, G. P. (2017) 'Plant-made polio type 3 stabilized VLPs—a candidate synthetic polio vaccine', *Nat Commun*, 8(1), p. 245. doi: 10.1038/s41467-017-00090-w.

Di Martino, B., Di Profio, F., Melegari, I., Sarchese, V., Cafiero, M. A., Robetto, S., Aste, G., Lanave, G., Marsilio, F. and Martella, V. (2016) 'A novel feline norovirus in diarrheic cats', *Infect Genet Evol*, 38, pp. 132–137. doi: 10.1016/j.meegid.2015.12.019.

Mastrangelo, E., Pezzullo, M., Tarantino, D., Petazzi, R., Germani, F., Kramer, D., Robel, I., Rohayem, J., Bolognesi, M. and Milani, M. (2012) 'Structure-based inhibition of Norovirus RNA-dependent RNA polymerases', *J Mol Biol*, 419(3–4), pp. 198–210. doi: 10.1016/j.jmb.2012.03.008.

Mattison, C. P., Cardemil, C. V. and Hall, A. J. (2018) 'Progress on norovirus vaccine research: public health considerations and future directions', *Expert Rev Vaccines*, 17(9), pp. 773–784. doi: 10.1080/14760584.2018.1510327.

McElwee, M., Vijayakrishnan, S., Rixon, F. and Bhella, D. (2018) 'Structure of the herpes simplex virus portal-vertex', *PLoS Biol*, 16(6), p. e2006191. doi: 10.1371/journal.pbio.2006191.

McFadden, N., Bailey, D., Carrara, G., Benson, A., Chaudhry, Y., Shortland, A., Heeney, J., Yarovinsky, F., Simmonds, P., Macdonald, A., *et al.* (2011) 'Norovirus regulation of the innate immune response and apoptosis occurs via the product of the alternative open reading frame 4', *PLoS Pathog*, 7(12), p. e1002413. doi: 10.1371/journal.ppat.1002413.

McKee, K. T., Oro, J. G., Kuehne, A. I., Spisso, J. A. and Mahlandt, B. G. (1993) 'Safety and immunogenicity of a live-attenuated Junin (Argentine hemorrhagic fever) vaccine in rhesus macaques', *Am J Trop Med Hyg*, 48(3), pp. 403–411. doi: 10.4269/ajtmh.1993.48.403.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R. (2013) 'Analysis Tool Web Services from the EMBL-EBI', *Nucleic Acids Res*, 41(W1), pp. W597–W600. doi: 10.1093/nar/gkt376.

Medici, M. C., Abelli, L. A., Dodi, I., Dettori, G. and Chezzi, C. (2010) 'Norovirus RNA in the blood of a child with gastroenteritis and convulsions–A case report', *J Clin Virol*, 48(2), pp. 147–149. doi: 10.1016/j.jcv.2010.03.001.

Meng, G., Zhang, X., Plevka, P., Yu, Q., Tijssen, P. and Rossmann, M. G. (2013) 'The structure and host entry of an invertebrate parvovirus', *J Virol*, 87(23), pp. 12523–12530. doi: 10.1128/jvi.01822-13.

Mesquita, J. R., Barclay, L., Nascimento, M. S. and Vinjé, J. (2010) 'Novel norovirus in dogs with diarrhea', *Emerg Infect Dis*, 16(6), pp. 980–982. doi: 10.3201/eid1606.091861.

Milich, D. R. and McLachlan, A. (1986) 'The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen', *Science*, 234(4782), pp. 1398–1401. doi: 10.1126/science.3491425.

Miller, S. T., Hogle, J. M. and Filman, D. J. (2001) 'Ab initio phasing of high-symmetry macromolecular complexes: successful phasing of authentic poliovirus data to 3.0 Å resolution. *J Mol Bio*, 307(2), pp. 499–512. doi: 10.1006/jmbi.2001.4485.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and

Steinegger, M. (2021) ‘ColabFold - Making protein folding accessible to all’, *bioRxiv [Preprint]*, p. 2021.08.15.456425. doi: 10.1101/2021.08.15.456425.

Morens, D. M., Holmes, E. C., Davis, A. S. and Taubenberger, J. K. (2011) ‘Global rinderpest eradication: lessons learned and why humans should celebrate too’, *J Infect Dis*, 204(4), pp. 502–505. doi: 10.1093/infdis/jir327.

Murakami, K., Tenge, V. R., Karandikar, U. C., Lin, S.C., Ramani, S., Ettayebi, K., Crawford, S. E., Zeng, X. L., Neill, F. H., Ayyar, B. V., *et al.* (2020) ‘Bile acids and ceramide overcome the entry restriction for GII.3 human norovirus replication in human intestinal enteroids’, *Proc Natl Acad Sci USA*, 117(3), pp. 1700–1710. doi: 10.1073/pnas.1910138117.

Napthine, S., Lever, R. A., Powell, M. L., Jackson, R. J., Brown, T. D. and Brierley, I. (2009) ‘Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy’, *PLoS One*, 4(12), p. e8390. doi: 10.1371/journal.pone.0008390.

Nassal, M. (1992) ‘The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly’, *J Virol*, 66(7), pp. 4107–4116. doi: 10.1128/jvi.66.7.4107-4116.1992.

Nassal, M., Rieger, A. and Steinau, O. (1992) ‘Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking’, *J Mol Biol*, 225(4), pp. 1013–1025. doi: 10.1016/0022-2836(92)90101-O.

National Center for Biotechnology Information (2021) ‘PubChem Compound Summary for CID 9903, Lithocholic acid’, from https://pubchem.ncbi.nlm.nih.gov/compound/Lithocholic-acid (Accessed: 9 December 2021).

Naydenova, K. and Russo, C. J. (2017) ‘Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy’, *Nat Commun*, 8(1), p. 629. doi: 10.1038/s41467-017-00782-3.

Nelson, C. A., Wilen, C. B., Dai, Y. N., Orchard, R. C., Kim, A. S., Stegeman, R. A., Hsieh, L. L., Smith, T. J., Virgin, H. W. and Fremont, D. H. (2018) ‘Structural basis for murine norovirus engagement of bile acids and the CD300lf receptor’, *Proc Natl Acad Sci USA*, 115(39), pp. E9201–E9210. doi: 10.1073/pnas.1805797115.

Netzler, N. E., Enosi Tuipulotu, D., Eltahla, A. A., Lun, J. H., Ferla, S., Brancale, A., Urakova, N., Frese, M., Strive, T., Mackenzie, J. M., *et al.* (2017) ‘Broad-spectrum non-nucleoside inhibitors for caliciviruses’, *Antiviral Res*, 146, pp. 65–75. doi: 10.1016/j.antiviral.2017.07.014.

Netzler, N. E., Enosi Tuipulotu, D. and White, P. A. (2019) ‘Norovirus antivirals: Where are we now?’, *Med Res Rev*, 39(3), pp. 860–886. doi: 10.1002/med.21545.

Noad, R. and Roy, P. (2003) ‘Virus-like particles as immunogens’, *Trends Microbiol*, 11(9), pp. 438–444. doi: 10.1016/S0966-842X(03)00208-7.

Nordgren, J. and Svensson, L. (2019) ‘Genetic Susceptibility to Human Norovirus Infection: An Update’, *Viruses*, 11(3), p. 226. doi: 10.3390/v11030226.

Oka, T., Takagi, H. and Tohya, Y. (2014) ‘Development of a novel single step reverse genetics system for feline calicivirus’, *J Virol Methods*, 207, pp. 178–181. doi: 10.1016/j.jviromet.2014.07.004.

Olspert, A., Hosmillo, M., Chaudhry, Y., Peil, L., Truve, E. and

Goodfellow, I. (2016) ‘Protein-RNA linkage and posttranslational modifications of feline calicivirus and murine norovirus VPg proteins’, *PeerJ*, 4, p. e2134. doi: 10.7717/peerj.2134.

Orlova, E. V. and Saibil, H. R. (2011) ‘Structural analysis of macromolecular assemblies by electron microscopy’, *Chem Rev*, 111(12), pp. 7710–7748. doi: 10.1021/cr100353t.

Papafragkou, E., Hewitt, J., Park, G. W., Greening, G. and Vinjé, J. (2013) ‘Challenges of culturing human norovirus in three-dimensional organoid intestinal cell culture models’, *PLoS One*, 8(6), p. e63485. doi: 10.1371/journal.pone.0063485.

Parra, G. I., Bok, K., Taylor, R., Haynes, J. R., Sosnovtsev, S. V., Richardson, C. and Green, K. Y. (2012) ‘Immunogenicity and specificity of norovirus Consensus GII.4 virus-like particles in monovalent and bivalent vaccine formulations’, *Vaccine*, 30(24), pp. 3580–3586. doi: 10.1016/j.vaccine.2012.03.050.

Parra, G. I., Squires, R. B., Karangwa, C. K., Johnson, J. A., Lepore, C. J., Sosnovtsev, S. V. and Green, K. Y. (2017) ‘Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity’, *PLOS Pathog*, 13(1), p. e1006136. doi: 10.1371/journal.ppat.1006136.

Parrino, T. A., Schreiber, D. S., Trier, J. S., Kapikian, A. Z. and Blacklow, N. R. (1977) ‘Clinical immunity in acute gastroenteritis caused by Norwalk agent’, *N Engl J Med*, 297(2), pp. 86–89. doi: 10.1056/nejm197707142970204.

Patel, M. M., Widdowson, M. A., Glass, R. I., Akazawa, K., Vinjé, J. and Parashar, U. D. (2008) ‘Systematic literature review of role of noroviruses in sporadic gastroenteritis’, *Emerg Infect Dis*, 14(8), pp. 1224–31. doi: 10.3201/eid1408.071114.

Patel, N., White, S. J., Thompson, R. F., Bingham, R., Weiß, E. U., Maskell, D. P., Zlotnick, A., Dykeman, E. C., Tuma, R., Twarock, R., *et al.* (2017) ‘HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly’, *Nat Microbiol*, 2, p. 17098. doi: 10.1038/nmicrobiol.2017.98.

Perry, J. W. and Wobus, C. E. (2010) ‘Endocytosis of murine norovirus 1 into murine macrophages is dependent on dynamin II and cholesterol’, *J Virol*, 84(12), pp. 6163–6176. doi: 10.1128/JVI.00331-10.

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’, *J Comput Chem*, 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Pevear, D. C., Tull, T. M., Seipel, M. E. and Groarke, J. M. (1999) ‘Activity of pleconaril against enteroviruses’, *Antimicrob Agents Chemother*, 43(9), pp. 2109–2115. doi: 10.1128/aac.43.9.2109.

Peyret, H., Gehin, A., Thuenemann, E. C., Blond, D., El Turabi, A., Beales, L., Clarke, D., Gilbert, R. J., Fry, E. E., Stuart, D. I., *et al.* (2015) ‘Tandem fusion of hepatitis B core antigen allows assembly of virus-like particles in bacteria and plants with enhanced capacity to accommodate foreign proteins’, *PLoS One*, 10(4), p. e0120751. doi: 10.1371/journal.pone.0120751.

Peyret, H., Ponndorf, D., Meshcheriakova, Y., Richardson, J. and Lomonossoff, G. P. (2020) ‘Covalent protein display on Hepatitis B core-like particles in plants through the in vivo use of the SpyTag/SpyCatcher system’,

Sci Rep, 10(1), p. 17095. doi: 10.1038/s41598-020-74105-w.

Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., Duarte, A. S., Black, R. E. and Angulo, F. J. (2015) 'Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food', *PLoS One*, 10(12), p. e0142927. doi: 10.1371/journal.pone.0142927.

Porta, C., Kotecha, A., Burman, A., Jackson, T., Ren, J., Loureiro, S., Jones, I. M., Fry, E. E., Stuart, D. I. and Charleston, B. (2013) 'Rational engineering of recombinant picornavirus capsids to produce safe, protective vaccine antigen', *PLoS Pathog*, 9(3), p. e1003255. doi: 10.1371/journal.ppat.1003255.

Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G. and Estes, M. K. (1999) 'X-ray crystallographic structure of the Norwalk virus capsid', *Science*, 286(5438), pp. 287–290. doi: 10.1126/SCIENCE.286.5438.287.

Prasad, B. V., Matson, D. O. and Smith, A. W. (1994a) 'Three-dimensional structure of calicivirus', *J Mol Biol*, 240(3), pp. 256–264. doi: 10.1006/jmbi.1994.1439.

Prasad, B. V., Rothnagel, R., Jiang, X. and Estes, M. K. (1994b) 'Three-dimensional structure of baculovirus-expressed Norwalk virus capsids', *J Virol*, 68(8), pp. 5117–25. doi: 10.1128/JVI.68.8.5117-5125.1994.

Punjani, A., Rubinstein, J. L., Fleet, D. J. and Brubaker, M. A. (2017) 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination', *Nat Methods*, 14(3), pp. 290–296. doi: 10.1038/nmeth.4169.

Qu, L., Murakami, K., Broughman, J. R., Lay, M. K., Guix, S., Tenge, V. R., Atmar, R. L. and Estes, M. K. (2016) 'Replication of Human Norovirus RNA in Mammalian Cells Reveals Lack of Interferon Response', *J Virol*, 90(19), pp. 8906–8923. doi: 10.1128/jvi.01425-16.

Raschke, W. C., Baird, S., Ralph, P. and Nakoinz, I. (1978) 'Functional macrophage cell lines transformed by Abelson leukemia virus', *Cell*, 15(1), pp. 261–267. doi: 10.1016/0092-8674(78)90101-0.

Rocha-Pereira, J., Jochmans, D., Dallmeier, K., Leyssen, P., Nascimento, M. S. and Neyts, J. (2012) 'Favipiravir (T-705) inhibits in vitro norovirus replication', *Biochem Biophys Res Commun*, 424(4), pp. 777–780. doi: 10.1016/j.bbrc.2012.07.034.

Rocha-Pereira, J., Jochmans, D., Debing, Y., Verbeken, E., Nascimento, M. S. and Neyts, J. (2013) 'The viral polymerase inhibitor 2'-C-methylcytidine inhibits Norwalk virus replication and protects against norovirus-induced diarrhea and mortality in a mouse model', *J Virol*, 87(21), pp. 11798–11805. doi: 10.1128/jvi.02064-13.

Roddie, C., Paul, J. P., Benjamin, R., Gallimore, C. I., Xerry, J., Gray, J. J., Peggs, K. S., Morris, E. C., Thomson, K. J. and Ward, K. N. (2009) 'Allogeneic hematopoietic stem cell transplantation and norovirus gastroenteritis: a previously unrecognized cause of morbidity', *Clin Infect Dis*, 49(7), pp. 1061–1068. doi: 10.1086/605557.

Rohayem, J., Robel, I., Jäger, K., Scheffler, U. and Rudolph, W. (2006) 'Protein-primed and de novo initiation of RNA synthesis by norovirus 3Dpol', *J Virol*, 80(14), pp. 7060–7069. doi: 10.1128/jvi.02195-05.

Rohou, A. and Grigorieff, N. (2015) 'CTFFIND4: Fast and accurate defocus estimation from electron micrographs', *J Struct Biol*, 192(2), pp. 216–

221. doi: 10.1016/j.jsb.2015.08.008.

Rosenthal, P. B. and Henderson, R. (2003) 'Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy', *J Mol Biol*, 333(4), pp. 721–745. doi: 10.1016/j.jmb.2003.07.013.

Rossignol, J. F. and El-Gohary, Y. M. (2006) 'Nitazoxanide in the treatment of viral gastroenteritis: a randomized double-blind placebo-controlled clinical trial', *Aliment Pharmacol Ther*, 24(10), pp. 1423–1430. doi: 10.1111/j.1365-2036.2006.03128.x.

Roth, A. N., Helm, E. W., Mirabelli, C., Kirsche, E., Smith, J. C., Eurell, L. B., Ghosh, S., Altan-Bonnet, N., Wobus, C. E. and Karst, S. M. (2020) 'Norovirus infection causes acute self-resolving diarrhea in wild-type neonatal mice', *Nat Commun*, 11(1), p. 2968. doi: 10.1038/s41467-020-16798-1.

Sandmann, F. G., Jit, M., Robotham, J. V. and Deeny, S. R. (2017) 'Burden, duration and costs of hospital bed closures due to acute gastroenteritis in England per winter, 2010/11-2015/16', *J Hosp Infect*, 97(1), pp. 79–85. doi: 10.1016/j.jhin.2017.05.015.

Santiana, M., Ghosh, S., Ho, B. A., Rajasekaran, V., Du, W. L., Mutsafi, Y., De Jésus-Diaz, D. A., Sosnovtsev, S. V., Levenson, E. A., Parra, G. I., *et al.* (2018) 'Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral Transmission', *Cell Host Microbe*, 24(2), pp. 208-220.e8. doi: 10.1016/j.chom.2018.07.006.

Scarff, C. A., Fuller, M. J. G., Thompson, R. F. and Iadanza, M. G. (2018) 'Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems', *J Vis Exp*, (132), p. 57199. doi: 10.3791/57199.

Scheres, S. H. (2010) 'Classification of structural heterogeneity by maximum-likelihood methods', *Methods Enzymol*, 482, pp. 295–320. doi: 10.1016/S0076-6879(10)82012-9.

Scheres, S. H. (2012) 'RELION: implementation of a Bayesian approach to cryo-EM structure determination', *J Struct Biol*, 180(3), pp. 519–530. doi: 10.1016/j.jsb.2012.09.006.

Scheres, S. H. (2015) 'Semi-automated selection of cryo-EM particles in RELION-1.3', *J Struct Biol*, 189(2), pp. 114–122. doi: 10.1016/j.jsb.2014.11.010.

Scheres, S. H. (2016) 'Processing of Structurally Heterogeneous Cryo-EM Data in RELION', *Methods Enzymol*, 579, pp. 125–157. doi: 10.1016/bs.mie.2016.04.012.

Scheres, S. H. and Chen, S. (2012) 'Prevention of overfitting in cryo-EM structure determination', *Nat Methods*, 9(9), pp. 853–854. doi: 10.1038/nmeth.2115.

Schödel, F., Moriarty, A. M., Peterson, D. L., Zheng, J. A., Hughes, J. L., Will, H., Leturcq, D. J., McGee, J. S. and Milich, D. R. (1992) 'The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity', *J Virol*, 66(1), pp. 106–114. doi: 10.1128/jvi.66.1.106-114.1992.

Scott, C. C. and Gruenberg, J. (2011) 'Ion flux and the function of endosomes and lysosomes: pH is just the start: the flux of ions across endosomal membranes influences endosome function not only through regulation of the luminal pH', *BioEssays*, 33(2), pp. 103–110. doi: 10.1002/bies.201000108.

Seitz, S. R., Leon, J. S., Schwab, K. J., Lyon, G. M., Dowd, M., McDaniels, M., Abdulhafid, G., Fernandez, M. L., Lindesmith, L. C., Baric, R. S., *et al.* (2011) 'Norovirus infectivity in humans and persistence in water', *Appl Environ Microbiol*, 77(19), pp. 6884–6888. doi: 10.1128/AEM.05806-11.

Selzer, L., Katen, S. P. and Zlotnick, A. (2014) 'The hepatitis B virus core protein intradimer interface modulates capsid assembly and stability', *Biochemistry*, 53(34), pp. 5496–5504. doi: 10.1021/bi500732b.

Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Boisset, N., Leith, A. and Frank, J. (2008) 'SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs', *Nat Protoc*, 3(12), pp. 1941–1974. doi: 10.1038/nprot.2008.156.

Shen, Q., Zhang, W., Yang, S., Cui, L. and Hua, X. (2012) 'Complete genome sequence of a new-genotype porcine norovirus isolated from piglets with diarrhea', *J Virol*, 86(12), pp. 7015–7016. doi: 10.1128/jvi.00757-12.

Sherman, M. B., Williams, A. N., Smith, H. Q., Nelson, C., Wilen, C. B., Fremont, D. H., Virgin, H. W. and Smith, T. J. (2019) 'Bile Salts Alter the Mouse Norovirus Capsid Conformation; Possible Implications for Cell Attachment and Immune Evasion', *J Virol*, 93(19), p. e00970-19. doi: 10.1128/JVI.00970-19.

Sherry, L., Grehan, K., Snowden, J. S., Knight, M. L., Adeyemi, O. O., Rowlands, D. J. and Stonehouse, N. J. (2020) 'Comparative Molecular Biology Approaches for the Production of Poliovirus Virus-Like Particles Using *Pichia pastoris*', *mSphere*, 5(2), p. e00838-19. doi: 10.1128/msphere.00838-19.

Siebenga, J. J., Vennema, H., Renckens, B., de Bruin, E., van der Veer, B., Siezen, R. J. and Koopmans, M. (2007) 'Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006', *J Virol*, 81(18), pp. 9932–9941. doi: 10.1128/jvi.00674-07.

Siebenga, J. J., Vennema, H., Zheng, D. P., Vinjé, J., Lee, B. E., Pang, X. L., Ho, E. C., Lim, W., Choudekar, A., Broor, S., *et al.* (2009) 'Norovirus illness is a global problem: emergence and spread of norovirus GII.4 Variants, 2001–2007', *J Infect Dis*, 200(5), pp. 802–812. doi: 10.1086/605127.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Mol Syst Biol*, 7, p. 539. doi: 10.1038/msb.2011.75.

Simmonds, P., Karakasiliotis, I., Bailey, D., Chaudhry, Y., Evans, D. J. and Goodfellow, I. G. (2008) 'Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses', *Nucleic Acids Res*, 36(8), pp. 2530–2546. doi: 10.1093/nar/gkn096.

Simmons, K., Gambhir, M., Leon, J. and Lopman, B. (2013) 'Duration of immunity to norovirus gastroenteritis', *Emerg Infect Dis*, 19(8), pp. 1260–1267. doi: 10.3201/eid1908.130472.

Singh, B. K., Leuthold, M. M. and Hansman, G. S. (2016) 'Structural Constraints on Human Norovirus Binding to Histo-Blood Group Antigens', *mSphere*, 1(2), p. e00049-16. doi: 10.1128/mSphere.00049-16.

Singh, K., Graf, B., Linden, A., Sautner, V., Urlaub, H., Tittmann, K., Stark, H. and Chari, A. (2020) 'Discovery of a Regulatory Subunit of the Yeast Fatty Acid Synthase', *Cell*, 180(6), pp. 1130-1143.e20. doi: 10.1016/j.cell.2020.02.034.

Singh, S. and Zlotnick, A. (2003) 'Observed hysteresis of virus capsid

disassembly is implicit in kinetic models of assembly', *J Biol Chem*, 278(20), pp. 18249–18255. doi: 10.1074/jbc.M211408200.

Smertina, E., Urakova, N., Strive, T. and Frese, M. (2019) 'Calicivirus RNA-Dependent RNA Polymerases: Evolution, Structure, Protein Dynamics, and Function', *Front Microbiol*, 10, p. 1280. doi: 10.3389/fmicb.2019.01280.

Smith, H. Q. and Smith, T. J. (2019) 'The Dynamic Capsid Structures of the Noroviruses', *Viruses*, 11(3), p. 235. doi: 10.3390/v11030235.

Snijder, J., Uetrecht, C., Rose, R. J., Sanchez-Eugenia, R., Marti, G. A., Agirre, J., Guérin, D. M., Wuite, G. J., Heck, A. J. and Roos, W. H. (2013) 'Probing the biophysical interplay between a viral genome and its capsid', *Nat Chem*, 5(6), pp. 502–509. doi: 10.1038/nchem.1627.

Snowden, J. S., Hurdiss, D. L., Adeyemi, O. O., Ranson, N. A., Herod, M. R. and Stonehouse, N. J. (2020) 'Dynamics in the murine norovirus capsid revealed by high-resolution cryo-EM', *PLoS Biol*, 18(3), p. e3000649. doi: 10.1371/journal.pbio.3000649.

Snowden, J. S., Alzahrani, J., Sherry, L., Stacey, M., Rowlands, D. J., Ranson, N. A. and Stonehouse, N. J. (2021) 'Structural insight into *Pichia pastoris* fatty acid synthase', *Sci Rep*, 11(1), p. 9773. doi: 10.1038/s41598-021-89196-2.

Song, C., Takai-Todaka, R., Miki, M., Haga, K., Fujimoto, A., Ishiyama, R., Oikawa, K., Yokoyama, M., Miyazaki, N., Iwasaki, K., *et al.* (2020) 'Dynamic rotation of the protruding domain enhances the infectivity of norovirus', *PLoS Pathog*, 16(7), p. e1008619. doi: 10.1371/journal.ppat.1008619.

Sosnovtsev, S. and Green, K. Y. (1995) 'RNA transcripts derived from a cloned full-length copy of the feline calicivirus genome do not require VpG for infectivity', *Virology*, 210(2), pp. 383–390. doi: 10.1006/viro.1995.1354.

Sosnovtsev, S. V., Belliot, G., Chang, K. O., Onwudiwe, O. and Green, K. Y. (2005) 'Feline calicivirus VP2 is essential for the production of infectious virions', *J Virol*, 79(7), pp. 4012–4024. doi: 10.1128/jvi.79.7.4012-4024.2005.

Stoops, J. K., Kolodziej, S. J., Schroeter, J. P., Bretaudiere, J. P. and Wakil, S. J. (1992) 'Structure-function relationships of the yeast fatty acid synthase: negative-stain, cryo-electron microscopy, and image analysis studies of the end views of the structure', *Proc Natl Acad Sci USA*, 89(14), pp. 6585–6589. doi: 10.1073/pnas.89.14.6585.

Stuart, A. D. and Brown, T. D. (2006) 'Entry of feline calicivirus is dependent on clathrin-mediated endocytosis and acidification in endosomes', *J Virol*, 80(15), pp. 7500–7509. doi: 10.1128/jvi.02452-05.

Sukhrie, F. H., Teunis, P., Vennema, H., Copra, C., Thijs Beersma, M. F., Bogerman, J. and Koopmans, M. (2012) 'Nosocomial transmission of norovirus is mainly caused by symptomatic cases', *Clin Infect Dis*, 54(7), pp. 931–937. doi: 10.1093/cid/cir971.

Takanashi, S., Saif, L. J., Hughes, J. H., Meulia, T., Jung, K., Scheuer, K. A. and Wang, Q. (2014) 'Failure of propagation of human norovirus in intestinal epithelial cells with microvilli grown in three-dimensional cultures', *Arch Virol*, 159(2), pp. 257–266. doi: 10.1007/s00705-013-1806-4.

Tan, M., Huang, P., Xia, M., Fang, P. A., Zhong, W., McNeal, M., Wei, C., Jiang, W. and Jiang, X. (2011) 'Norovirus P particle, a novel platform for vaccine development and antibody production', *J Virol*, 85(2), pp. 753–764. doi: 10.1128/jvi.01835-10.

Tan, M., Hegde, R. S. and Jiang, X. (2004) 'The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors', *J Virol*, 78(12), pp. 6233–6242. doi: 10.1128/jvi.78.12.6233-6242.2004.

Taube, S., Perry, J. W., Yetming, K., Patel, S. P., Auble, H., Shu, L., Nawar, H. F., Lee, C. H., Connell, T. D., Shayman, J. A., *et al.* (2009) 'Ganglioside-linked terminal sialic acid moieties on murine macrophages function as attachment receptors for murine noroviruses', *J Virol*, 83(9), pp. 4092–4101. doi: 10.1128/jvi.02245-08.

Taube, S., Rubin, J. R., Katpally, U., Smith, T. J., Kendall, A., Stuckey, J. A. and Wobus, C. E. (2010) 'High-resolution X-ray structure and functional analysis of the murine norovirus 1 capsid protein protruding domain', *J Virol*, 84(11), pp. 5695–5705. doi: 10.1128/JVI.00316-10.

Taube, S., Perry, J. W., McGreevy, E., Yetming, K., Perkins, C., Henderson, K. and Wobus, C. E. (2012) 'Murine noroviruses bind glycolipid and glycoprotein attachment receptors in a strain-dependent manner', *J Virol*, 86(10), pp. 5584–5593. doi: 10.1128/jvi.06854-11.

Taylor, G. L. (2010) 'Introduction to phasing', *Acta Crystallogr D Biol Crystallogr*, 66(4), pp. 325–338. doi: 10.1107/S0907444910006694.

Tehlivets, O., Scheuringer, K. and Kohlwein, S. D. (2007) 'Fatty acid synthesis and elongation in yeast', *Biochim Biophys Acta*, 1771(3), pp. 255–270. doi: 10.1016/j.bbalip.2006.07.004.

Teunis, P. F., Moe, C. L., Liu, P., Miller, S. E., Lindesmith, L., Baric, R. S., Le Pendu, J. and Calderon, R. L. (2008) 'Norwalk virus: how infectious is it?', *J Med Virol*, 80(8), pp. 1468–1476. doi: 10.1002/jmv.21237.

Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. and Ranson, N. A. (2016) 'An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology', *Methods*, 100, pp. 3–15. doi: 10.1016/j.ymeth.2016.02.017.

Thorne, L., Bailey, D. and Goodfellow, I. (2012) 'High-resolution functional profiling of the norovirus genome', *J Virol*, 86(21), pp. 11441–56. doi: 10.1128/JVI.00439-12.

Thorne, L. G. and Goodfellow, I. G. (2014) 'Norovirus gene expression and replication', *J Gen Virol*, 95(Pt 2), pp. 278–291. doi: 10.1099/vir.0.059634-0.

Tiede, C., Tang, A. A., Deacon, S. E., Mandal, U., Nettleship, J. E., Owen, R. L., George, S. E., Harrison, D. J., Owens, R. J., Tomlinson, D. C., *et al.* (2014) 'Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications', *Protein Eng Des Sel*, 27(5), pp. 145–155. doi: 10.1093/protein/gzu007.

Tiede, C., Bedford, R., Heseltine, S. J., Smith, G., Wijetunga, I., Ross, R., AlQallaf, D., Roberts, A. P., Balls, A., Curd, A., *et al.* (2017) 'Affimer proteins are versatile and renewable affinity reagents', *eLife*, 6, p. e24903. doi: 10.7554/eLife.24903.

Tiew, K. C., He, G., Aravapalli, S., Mandadapu, S. R., Gunnam, M. R., Alliston, K. R., Lushington, G. H., Kim, Y., Chang, K. O. and Groutas, W. C. (2011) 'Design, synthesis, and evaluation of inhibitors of Norwalk virus 3C protease', *Bioorg Med Chem Lett*, 21(18), pp. 5315–5319. doi: 10.1016/j.bmcl.2011.07.016.

Troeger, H., Loddenkemper, C., Schneider, T., Schreier, E., Epple, H. J., Zeitz, M., Fromm, M. and Schulzke, J. D. (2009) 'Structural and functional

changes of the duodenum in human norovirus infection', *Gut*, 58(8), pp. 1070–1077. doi: 10.1136/gut.2008.160150.

Turcios-Ruiz, R. M., Axelrod, P., St John, K., Bullitt, E., Donahue, J., Robinson, N. and Friss, H. E. (2008) 'Outbreak of necrotizing enterocolitis caused by norovirus in a neonatal intensive care unit', *J Pediatr*, 153(3), pp. 339–344. doi: 10.1016/j.jpeds.2008.04.015.

Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E. and Heck, A. J. (2011) 'Interrogating viral capsid assembly with ion mobility-mass spectrometry', *Nat Chem*, 3(2), pp. 126–132. doi: 10.1038/nchem.947.

Vashist, S., Bailey, D., Putics, A. and Goodfellow, I. (2009) 'Model systems for the study of human norovirus Biology', *Future Virol*, pp. 353–367. doi: 10.2217/fvl.09.18.

Viktorova, E. G., Khattar, S. K., Kouiavskaia, D., Laassri, M., Zagorodnyaya, T., Dragunsky, E., Samal, S., Chumakov, K. and Belov, G. A. (2018) 'Newcastle Disease Virus-Based Vectored Vaccine against Poliomyelitis', *J Virol*, 92(17), p. e00976-18. doi: 10.1128/jvi.00976-18.

Vongpunsawad, S., Venkataram Prasad, B. V. and Estes, M. K. (2013) 'Norwalk Virus Minor Capsid Protein VP2 Associates within the VP1 Shell Domain', *J Virol*, 87(9), pp. 4818–25. doi: 10.1128/JVI.03508-12.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., *et al.* (2019) 'SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM', *Commun Biol*, 2, p. 218. doi: 10.1038/s42003-019-0437-z.

Ward, V. K., McCormick, C. J., Clarke, I. N., Salim, O., Wobus, C. E., Thackray, L. B., Virgin IV, H. W. and Lambden, P. R. (2007) 'Recovery of infectious murine norovirus using pol II-driven expression of full-length cDNA', *Proc Natl Acad Sci USA*, 104(26), pp. 11050–11055. doi: 10.1073/pnas.0700336104.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., *et al.* (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Res*, 46(W1), pp. W296–W303. doi: 10.1093/nar/gky427.

Wei, C., Farkas, T., Sestak, K. and Jiang, X. (2008) 'Recovery of infectious virus by transfection of in vitro-generated RNA from tulane calicivirus cDNA', *J Virol*, 82(22), pp. 11429–11436. doi: 10.1128/jvi.00696-08.

Weis, F., Beckers, M., von der Hocht, I. and Sachse, C. (2019) 'Elucidation of the viral disassembly switch of tobacco mosaic virus', *EMBO Rep*, 20(11), p. e48451. doi: 10.15252/embr.201948451.

Wilen, C. B., Lee, S., Hsieh, L. L., Orchard, R. C., Desai, C., Hykes, B. L., McAllaster, M. R., Balce, D. R., Feehley, T., Brestoff, J. R., *et al.* (2018) 'Tropism for tuft cells determines immune promotion of norovirus pathogenesis', *Science*, 360(6385), pp. 204–208. doi: 10.1126/science.aar3799.

Williams, A. N., Sherman, M. B., Smith, H. Q., Taube, S., Pettitt, B. M., Wobus, C. E. and Smith, T. J. (2021a) 'A Norovirus Uses Bile Salts To Escape Antibody Recognition While Enhancing Receptor Binding', *J Virol*, 95(13), p. e0017621. doi: 10.1128/JVI.00176-21.

Williams, A. N., Sherman, M. B., Smith, H. Q., Taube, S., Pettitt, B. M., Wobus, C. E. and Smith, T. J. (2021b) 'Multiple Signals in the Gut Contract

the Mouse Norovirus Capsid To Block Antibody Binding While Enhancing Receptor Affinity', *J Virol*, 95(22), p. e0147121. doi: 10.1128/JVI.01471-21.

Wobus, C. E., Karst, S. M., Thackray, L. B., Chang, K. O., Sosnovtsev, S. V, Belliot, G., Krug, A., Mackenzie, J. M., Green, K. Y. and Virgin, H. W. (2004) 'Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages', *PLoS Biol*, 2(12), p. e432. doi: 10.1371/journal.pbio.0020432.

Wobus, C. E. (2018) 'The Dual Tropism of Noroviruses', *J Virol*, 92(16). doi: 10.1128/jvi.01010-17.

Wobus, C. E., Thackray, L. B. and Virgin IV, H. W. (2006) 'Murine norovirus: a model system to study norovirus biology and pathogenesis', *J Virol*, 80(11), pp. 5104–5112. doi: 10.1128/jvi.02346-05.

Wolf, S., Williamson, W., Hewitt, J., Lin, S., Rivera-Aban, M., Ball, A., Scholes, P., Savill, M. and Greening, G. E. (2009) 'Molecular detection of norovirus in sheep and pigs in New Zealand farms', *Vet Microbiol*, 133(1–2), pp. 184–189. doi: 10.1016/j.vetmic.2008.06.019.

Wu, S., Avila-Sakar, A., Kim, J., Booth, D. S., Greenberg, C. H., Rossi, A., Liao, M., Li, X., Alian, A., Griner, S. L., *et al.* (2012) 'Fabs enable single particle cryoEM studies of small proteins', *Structure*, 20(4), pp. 582–592. doi: 10.1016/j.str.2012.02.017.

Wu, X. and Rapoport, T. A. (2021) 'Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies)', *Proc Natl Acad Sci USA*, 118(41), p. e2115001118. doi: 10.1073/pnas.2115001118.

Wynne, S. A., Crowther, R. A. and Leslie, A. G. (1999) 'The crystal structure of the human hepatitis B virus capsid', *Mol Cell*, 3(6), pp. 771–780. doi: 10.1016/S1097-2765(01)80009-5.

Xu, Y., Ma, S., Huang, Y., Chen, F., Chen, L., Ding, D., Zheng, Y., Li, H., Xiao, J., Feng, J., *et al.* (2019) 'Virus-like particle vaccines for poliovirus types 1, 2, and 3 with enhanced thermostability expressed in insect cells', *Vaccine*, 37(17), pp. 2340–2347. doi: 10.1016/j.vaccine.2019.03.031.

Yang, Z. and Zhang, Z. (2018) 'Engineering strategies for enhanced production of protein and bio-products in Pichia pastoris: A review', *Biotechnol Adv*, 36(1), pp. 182–195. doi: 10.1016/j.biotechadv.2017.11.002.

Yu, G., Zhang, D., Guo, F., Tan, M., Jiang, X. and Jiang, W. (2013a) 'Cryo-EM structure of a novel calicivirus, Tulane virus', *PLoS One*, 8(3), p. e59817. doi: 10.1371/journal.pone.0059817.

Yu, X., Jin, L., Jih, J., Shih, C. and Zhou, Z. H. (2013b) '3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein', *PLoS One*, 8(9), p. e69729. doi: 10.1371/journal.pone.0069729.

Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T. and Howarth, M. (2012) 'Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin', *Proc Natl Acad Sci USA*, 109(12), pp. E690–E697. doi: 10.1073/pnas.1115485109.

Zhang, K. (2016) 'Gctf: Real-time CTF determination and correction', *J Struct Biol*, 193(1), pp. 1–12. doi: 10.1016/j.jsb.2015.11.003.

Zhang, M., Ghosh, S., Kumar, M., Santiana, M., Bleck, C. K. E., Chaimongkol, N., Altan-Bonnet, N. and Shuai, D. (2021) 'Emerging Pathogenic Unit of Vesicle-Cloaked Murine Norovirus Clusters is Resistant to Environmental Stresses and UV_{254} Disinfection', *Environ Sci Technol*, 55(9), pp. 6197–6205. doi: 10.1021/acs.est.1c01763.

Zhang, X. F., Tan, M., Chhabra, M., Dai, Y. C., Meller, J. and Jiang, X. (2013) 'Inhibition of histo-blood group antigen binding as a novel strategy to block norovirus infections', *PLoS One*, 8(7), p. e69379. doi: 10.1371/journal.pone.0069379.

Zheng, J., Schödel, F. and Peterson, D. L. (1992) 'The structure of hepadnaviral core antigens. Identification of free thiols and determination of the disulfide bonding pattern', *J Biol Chem*, 267(13), pp. 9422–9429. doi: 10.1016/s0021-9258(19)50440-9.

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y. and Agard, D. A. (2017) 'MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy', *Nat Methods*, 14(4), pp. 331–332. doi: 10.1038/nmeth.4193.

Zhou, M., Li, Y., Hu, Q., Bai, X. C., Huang, W., Yan, C., Scheres, S. H. and Shi, Y. (2015) 'Atomic structure of the apoptosome: mechanism of cytochrome c- and dATP-mediated activation of Apaf-1', *Genes Dev*, 29(22), pp. 2349–2361. doi: 10.1101/gad.272278.115.

Zhu, Z., Zhou, Y. J., Krivoruchko, A., Grininger, M., Zhao, Z. K. and Nielsen, J. (2017) 'Expanding the product portfolio of fungal type I fatty acid synthases', *Nat Chem Biol*, 13(4), pp. 360–362. doi: 10.1038/nchembio.2301.

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, p. e42166. doi: 10.7554/eLife.42166.

Zivanov, J., Nakane, T. and Scheres, S. H. W. (2019) 'A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis', *IUCrJ*, 6(Pt 1), pp. 5–17. doi: 10.1107/S205225251801463X.

Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. and Wingfield, P. T. (1996) 'Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein', *Biochemistry*, 35(23), pp. 7412–7421. doi: 10.1021/bi9604800.

Appendix A. Supplementary figures.

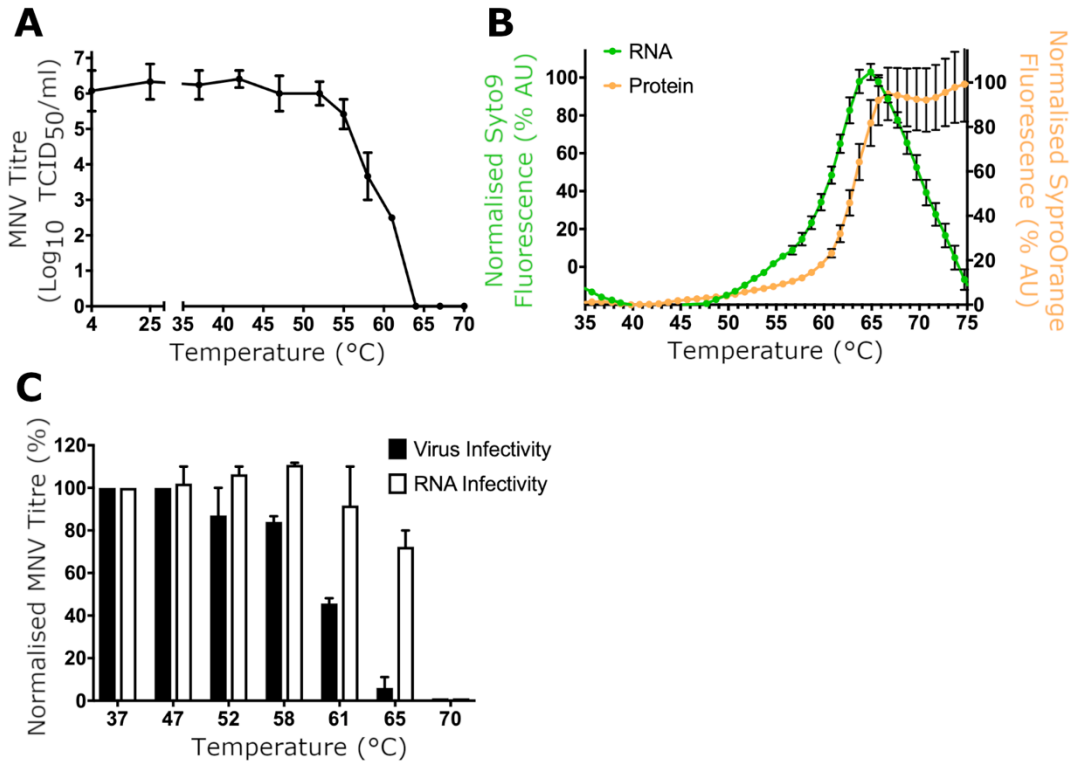


Figure S1. Characterisation of wtMNV thermal stability. ***(A)*** *$TCID_{50}$ assay of wtMNV on RAW264.7 cells following heating on a 30-second constant temperature ramp at the indicated temperature (n = 2 ± SEM).* ***(B)*** *PaSTRy assay of wtMNV during heating on a 30-second constant temperature ramp (n = 3 ± SEM). Fluorescence from the nucleic acid dye SYTO-9 is indicated in green. Fluorescence from the protein dye SYPRO-Orange is indicated in orange.* ***(C)*** *Comparison of virus and RNA genome 'infectivity' via $TCID_{50}$ assay (n = 2 ± SEM). wtMNV was heated at the indicated temperature then treated with RNase A before titration by $TCID_{50}$ assay on RAW264.7 cells ('Virus Infectivity') or extraction of total RNA. Extracted RNA was transfected into BHK cells (which only permit a single round of replication) and resultant virus was harvested and titrated by $TCID_{50}$ assay on RAW264.7 cells ('RNA Infectivity'). (This figure was adapted from* (Snowden *et al.*, 2020)*, published under a CC BY 4.0 licence. Experiments were performed by MH and OA.)*

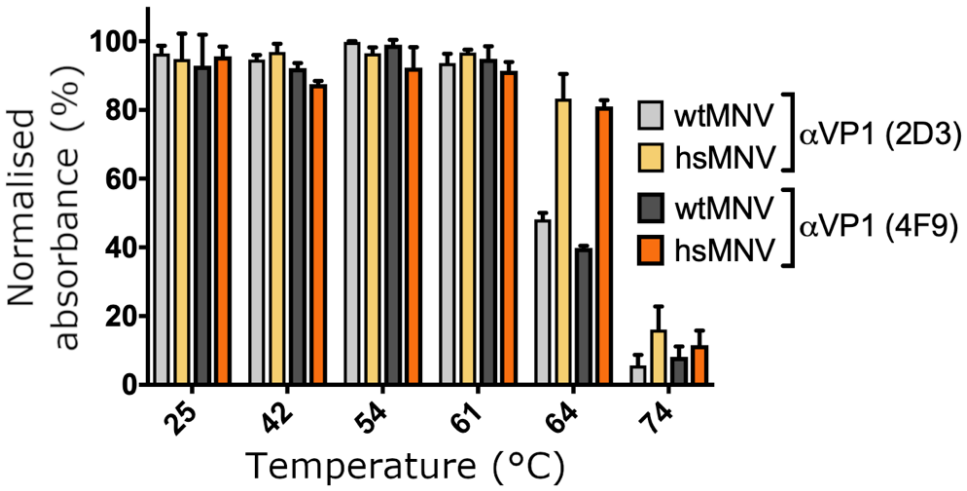


Figure S2. hsMNV retains two key antigenic determinants. *wtMNV and hsMNV antigenicity was analysed by ELISA using the neutralising anti-VP1 antibodies, 2D3 and 4F9, following incubation at the temperature indicated (n = 2 ± SD). (Experiment performed by MH and OA. This figure was adapted from* (Snowden *et al.*, 2020)*, published under a CC BY 4.0 licence.)*

Appendix B. Additional publications.

During the course of the PhD, the candidate also contributed to the following co-authored publications:

(1) Hurdiss D.L., Frank M., **Snowden J.S.**, Macdonald A. and Ranson N.A. (2018). The Structure of an Infectious Human Polyomavirus and Its Interactions with Cellular Receptors. *Structure*. **26**(6):839-847.e3.

Abstract:
BK polyomavirus (BKV) causes polyomavirus-associated nephropathy and hemorrhagic cystitis in immunosuppressed patients. These are diseases for which we currently have limited treatment options, but potential therapies could include pre-transplant vaccination with a multivalent BKV vaccine or therapeutics which inhibit capsid assembly or block attachment and entry into target cells. A useful tool in such efforts would be a high-resolution structure of the infectious BKV virion and how this interacts with its full repertoire of cellular receptors. We present the 3.4-Å cryoelectron microscopy structure of native, infectious BKV in complex with the receptor fragment of GT1b ganglioside. We also present structural evidence that BKV can utilize glycosaminoglycans as attachment receptors. This work highlights features that underpin capsid stability and provides a platform for rational design and development of urgently needed pharmacological interventions for BKV-associated diseases.

Contributions of the candidate:
The candidate repeated aspects of the processing for the BK-heparin structure and assisted with editing the manuscript.

(2) Sherry L., Grehan K., **Snowden J.S.**, Knight M., Adeyemi O.O., Rowlands D.J. and Stonehouse N.J. (2020). Comparative Molecular Biology Approaches for the Production of Poliovirus Virus-Like Particles Using *Pichia pastoris*. *mSphere*. **5**:e00838-19.

Abstract:

For enteroviruses such as poliovirus (PV), empty capsids, which are antigenically indistinguishable from mature virions, are produced naturally during viral infection. The production of such capsids recombinantly, in heterologous systems such as yeast, have great potential as virus-like particle (VLP) vaccine candidates. Here, using PV as an exemplar, we show the production of VLPs in *Pichia pastoris* by co-expression of the structural precursor protein P1 and the viral protease 3CD. The level of expression of the potentially cytotoxic protease relative to that of the P1 precursor was modulated by three different approaches: expression of the P1 precursor and protease from different transcription units, separation of the P1 and protease proteins using the Thosea asigna virus (TaV) 2A translation interruption sequence, or separation of the P1 and protease-coding sequences by an internal ribosome entry site sequence from Rhopalosiphum padi virus (RhPV). We also investigate the antigenicity of VLPs containing previously characterized mutations when produced in *P. pastoris*. Finally, using transmission electron microscopy and two-dimensional classification, we show that yeast-derived VLPs exhibited the classical icosahedral capsid structure displayed by enteroviruses.

Contributions of the candidate:

The candidate prepared samples for and performed all electron microscopy and image processing, and analysed structural data. The candidate produced figure 3 and assisted with editing the manuscript.

(3) Adeyemi O.O., Ward J.C., **Snowden J.S.**, Kingston N.J., Sherry L., Herod M.R., Rowlands D.J. and Stonehouse N.J. (2021). Functional advantages of triplication of the 3B coding region of the FMDV genome. *FASEB J*. **35**(2):e21215.

Abstract:

For gene duplication to be maintained, particularly in the small genomes of RNA viruses, this should offer some advantages. We have investigated the functions of a small protein termed VPg or 3B, which acts as a primer in the replication of foot-and-mouth disease virus (FMDV). Many related picornaviruses encode a single copy but uniquely the FMDV genome includes three (nonidentical) copies of the 3B coding region. Using sub-genomic replicons incorporating nonfunctional 3Bs and 3B fusion products in competition and complementation assays, we investigated the contributions of individual 3Bs to replication and the structural requirements for functionality. We showed that a free N-terminus is required for 3B to function as a primer and although a single 3B can support genome replication, additional copies provide a competitive advantage. However, a fourth copy confers no further advantage. Furthermore, we find that a minimum of two 3Bs is necessary for trans replication of FMDV replicons, which is unlike other picornaviruses where a single 3B can be used for both cis and trans replication. Our data are consistent with a model in which 3B copy number expansion within the FMDV genome has allowed evolution of separate cis and trans acting functions, providing selective pressure to maintain multiple copies of 3B.

Contributions of the candidate:

The candidate analysed previously published structures of FMDV 3B and 3D to propose a mechanistic model that could explain results observed in N-terminal fusion experiments. The candidate produced figure S8, wrote the corresponding section of the Discussion, and assisted with editing the manuscript.

(4) Kingston N.J., Grehan K., **Snowden J.S.**, Shegdar M., Macadam A.J., Rowlands D.J. and Stonehouse N.J. Development of an ELISA for the detection of the native conformation of enterovirus A71. *In preparation*.

Abstract:

EVA71 is a medically important virus which is commonly associated with hand, foot, and mouth disease (HFMD) resulting in periodic outbreaks with significant economic impact and loss of life. Vaccination offers the potential to curtail future outbreaks and vaccine development has been increasingly the focus of global research efforts. However, antigenic characterisation of vaccine candidates is challenging as there are very limited effective tools to characterise the different antigenic forms of EV71. As with other picornaviruses, EVA71 virions exist in two antigenic states, native (NAg) and expanded (HAg). It is likely that the composition of vaccines, in terms of the proportions of NAg and HAg, will be a critical determinant for vaccine efficacy and batch-to-batch consistency. This paper describes the development of a single chain fused variable domain fragment and the optimisation of a sandwich ELISA for the specific detection of the NAg conformation of EVA71. Samples of purified EVA71 were used to show NAg specificity. The conversion of NAg to HAg (with the application of heat) resulted in a loss of binding. We have thus developed an effective tool for characterisation of the specific antigenic state of EVA71.

Contributions of the candidate:

The candidate performed negative stain EM on unheated and heated EVA71 samples (shown in Figure 4e), and assisted with editing the manuscript.

(5) Kingston N.J., Shegdar M., **Snowden J.S.**, Fox H., Groppelli E., Macadam A., Rowlands D.J., Stonehouse N.J. Thermal stabilisation of enterovirus A 71 and production of antigenically stabilised empty capsids. *In preparation*.

Abstract:

Enterovirus A71 (EVA71) infection can result in paralysis and may be fatal. In common with other picornaviruses, empty capsids are produced alongside infectious virions during the viral lifecycle. These empty capsids are antigenically indistinguishable from infectious virus, however, at moderate temperatures they are converted to an expanded conformation. In the closely related poliovirus, native and expanded antigenic forms of particle have different long-term protective efficacies when used as vaccines. The native form provides long-lived protective immunity, while expanded capsids fail to generate immunological protection. Whether this is true for EVA71 remains to be determined. Here, we selected an antigenically stable EVA71 virus population using successive rounds of heating and passage and characterised the antigenic conversion of both virions and empty capsids. The mutations identified within the heated passaged virus were dispersed across the capsid, including at key sites associated with particle expansion. The data presented here indicates the mutant sequence may be a useful resource to address the importance of antigenic conformation in EVA71 vaccines.

Contributions of the candidate:

The candidate carried out structural visualisation of the mutations identified during passaging of EVA71 under thermal selection pressure, generated figure 2, and assisted with editing the manuscript.