

**Elucidation of the *Hendra henipavirus* nucleocapsid protein
structure and its genome encapsidation architectures**

2nd Edition

Timen Casper Passchier, MSc

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

The University of Leeds

School of Molecular and Cellular Biology

The Astbury Centre for Structural Molecular Biology

September 2022

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Timen Casper Passchier to be identified as Author of this work has been asserted by Timen Casper Passchier in accordance with the Copyright, Designs and Patents Act 1988.

© 2022 The University of Leeds and Timen Casper Passchier

“People who look for the first time through a microscope say now I see this and then I see that—and even a skilled observer can be fooled. On these observations I have spent more time than many will believe, but I have done them with joy, and I have taken no notice of those who have said why take so much trouble and what good is it?—but I do not write for such people but only for the philosophical!”

- Antonie van Leeuwenhoek*

*As quoted, without citation, by Dugald Caleb Jackson and Walter Paul Jones, in *This Scientific Age: Essays in Modern Thought and Achievement* (1930), 132.

"But I am very poorly today
and very stupid and hate everybody
and everything

One lives only to make blunders"

- Charles Darwin, Oct. 1, 1861, writing to Charles Lyell

Acknowledgements

The work presented in this thesis was funded by the European Union's Horizon 2020 research and innovation programme, via the MSCA grant agreement no. 721367 (HONOURS).

First and foremost, my thanks go to Dr John Barr and Dr Thomas Edwards for their invaluable supervision, guidance, and input over the course of my doctoral studies at Leeds. Your support has been highly influential in my development as a scientist. To all the colleagues, past and present, in the Barr and Edwards groups thank you for your kind and welcoming atmosphere. It has been a pleasure to work alongside and with you all. Additional thanks go to Dr David Nicholson, Dr Samantha Hover, Dr Rebecca Thompson, and Dr Daniel Maskell for offering to proofread sections of this thesis. Your suggestions have been extremely helpful.

Furthermore, I would like to thank everyone involved in the MSCA ITN HONOURS, especially the other ESRs. Being part of this group of young scientists has been a unique and rare experience. It has been great sharing trips, science, beers, and friendships that will last a lifetime, I'm sure.

I would also like to thank the Astbury Biostructure Facility and in particular Dr Daniel Maskell for invaluable assistance in cryo-EM screening, grid-optimisation, and data collection, and for always being around for a chat.

Thanks to all the friendships I made during my time at the University of Leeds, which I will cherish forever. To Dr Daniel Maskell, thanks for all your advice and reinvigoration of my hobbies. To Danielle Charles, thank you for all the chats and life coaching. It's always good to have a co-dependent friend. To Dr David Nicholson, Dr Benjamin Chadwick, Dr Mike Davies, Katherine Murphy, and Sarah Carpenter I'm always up for some icosahedral-based shenanigans. Let's build a trebuchet! Thanks Kat and Dani, for pulling me out of dark places when I couldn't myself.

A big hug for Dr James Murphy, you're the best bud I could have hoped to find. It's been an eternal pleasure to don the Gryphons jersey and share the ice with you and although our paths diverge, I'm sure we'll see each other again wherever we are. The very same goes for Dr Ryan Seipke; James and I finally wore you down and got you to take up the skates again. To Leeds Gryphons Ice Hockey, you've been a family away from home and I love you all.

Als laatste, voor mijn ouders, broer, en grootouders. Dank jullie wel voor het oneindige aanmoedigen en geduld die jullie voor mij en het navolgen van mijn ambities hebben gehad. Ik, zo ver weg, is niet makkelijk geweest en ik ben daarom extra dankbaar voor jullie ondersteuning. En de vele pakketjes met lekkernijen van thuis.

Abstract

The paramyxovirus *Hendra henipavirus* (HeV) is an emerging, zoonotic pathogen with a high case-mortality rate and scant therapeutic interventions. Placed on the short list of priority diseases for research and development maintained by the WHO, more fundamental knowledge is needed to drive the development of novel therapeutic and preventative measures.

The viral nucleoprotein is an essential part of the paramyxoviral replication strategy as it fully encapsidates the viral genome throughout the viral life cycle. The nucleocapsid, a complex of nucleoprotein and viral RNA genome, forms the template from which the viral polymerase produces new copies. Understanding the nucleoprotein structure and nucleocapsid architecture may facilitate the design of new, specific small molecule inhibitors. A nucleoprotein structure for HeV is currently lacking in the published literature.

In this work, the HeV nucleoprotein structure was solved to 3.5 Å through the use of cryo-electron microscopy (cryo-EM) from recombinantly expressed samples. Optimisation of HeV nucleoprotein expression and purification yielded samples for cryo-EM investigation and data collection. High-resolution 3D reconstruction of the HeV nucleoprotein structure was enabled by a novel nucleoprotein assembly, here called the sauronoid, which follows D14 symmetry. This sauronoid assembly exhibits both existing and novel protein-protein and protein-RNA interfaces. The HeV nucleoprotein conforms to the paramyxoviral rule of six – binding a hexameric RNA chain in the asymmetric unit. An intercalation zone, which suffered from poor local resolution, exists between the two tetradecameric halves of the sauronoid. Furthermore, the removal of the N-terminally fused affinity purification tag enabled characterisation of a novel, but minor, nucleoprotein-nucleoprotein interaction here called the elbow interface.

These findings, combined with published data on related viruses, paint a picture of a diverse array of nucleoprotein architectures that may fulfil varied functions during the viral life cycle, including genome encapsidation, nucleocapsid end-capping, polyploid virion assembly, short RNA encapsidation, and metastable nucleoprotein storage. The data further reveals novel features of nucleoprotein-nucleoprotein interactions and adds to our fundamental knowledge on paramyxoviral nucleoproteins. Finally, the structure solved here also reveals potential targets for the development of therapeutic interventions for HeV and other related viruses that share similar targets.

Contents

Contents	VI
List of Figures	XI
Supplementary Figures	XII
List of Tables	XIII
Supplementary Tables	XIII
Abbreviation Key	XIV
Chapter 1: Introduction	1
1.1 Zoonotic emergence	1
1.2 Negative strand RNA viruses.....	2
1.2.1 Segmented negative strand RNA viruses.....	4
1.2.2 Non-segmented negative strand RNA viruses	4
1.3 Nucleoproteins and ribonucleoproteins	5
1.3.1 Nucleoproteins of the segmented negative strand RNA viruses.....	6
1.3.2 Nucleoproteins of the non-segmented negative strand RNA viruses	8
1.3.2.1 Nucleoproteins and nucleocapsids of the family <i>Paramyxoviridae</i>	9
1.4 <i>Hendra henipavirus</i> discovery and classification	12
1.5 <i>Hendra henipavirus</i> epidemiology, transmission, and symptoms.....	13
1.6 <i>Hendra henipavirus</i> organisation.....	15
1.6.1 <i>Hendra henipavirus</i> genome and genes.....	15
1.6.2 <i>Hendra henipavirus</i> virion.....	17
1.7 <i>Hendra henipavirus</i> life cycle.....	19
1.7.1 Viral tropism and entry	19
1.7.2 Viral RNA transcription and replication	20
1.7.2.1 Messenger RNA synthesis and editing.....	24
1.7.2.2 Viral genome replication and nucleocapsid assembly.....	25
1.7.3 Virus-host immune interactions	29
1.7.4 Virion maturation, assembly, and exit.....	31
1.8 <i>Hendra henipavirus</i> diagnostics and surveillance	34
1.9 <i>Hendra henipavirus</i> treatment and prevention.....	35
1.10 Thesis aims	37
Chapter 2: Materials and Methods	39
2.1 Materials	39
2.1.1 Buffers and Media Recipes	39
2.1.2 Consumables.....	40

2.1.2.1 Vectors	40
2.1.2.2 Primers	40
2.1.2.3 Bacterial cell lines.....	40
2.1.2.4 Mammalian cell lines	41
2.1.2.5 Transmission electron microscopy grids.....	42
2.2 Methods.....	42
2.2.1 Molecular biology	42
2.2.1.1 Bacterial transformation and DNA amplification.....	42
2.2.1.2 DNA purification.....	43
2.2.1.2 Cloning <i>Hendra henipavirus</i> N gene into bacterial expression vectors	43
2.2.1.2 Cloning <i>Hendra henipavirus</i> 3' NTR sequences into pUC57-support vectors.....	43
2.2.1.4 Sequence validation	43
2.2.2 Recombinant protein purification.....	44
2.2.2.1 Starter cultures	44
2.2.2.2 Protein expression and cell lysis	44
2.2.2.3 Affinity chromatography.....	44
2.2.2.4 Size exclusion chromatography	45
2.2.2.5 SDS-PAGE gels and staining	45
2.2.3 Transmission electron microscopy	46
2.2.3.1 Preparation of negative stain grids.....	46
2.2.3.2 Preparation of cryo-grids	46
2.2.3.3 Data collection on negative stain grids.....	46
2.2.3.4 Data collection on cryo-grids	46
2.2.4 Electron microscopy data processing	47
2.2.4.1 Classical single particle analysis	47
2.2.4.2 Helical single particle analysis.....	48
2.2.5 Atomic protein model refinement and interpretation	48
2.2.6 <i>Hendra henipavirus</i> replicon system.....	49
2.2.6.1 Replicon design	49
2.2.6.2 Cell maintenance	49
2.2.6.3 Cell passage and seeding	49
2.2.6.4 <i>In vitro</i> RNA transcription	50
2.2.6.3 Transfection	50
2.2.6.4 IncuCyte imaging.....	50
Chapter 3: Recombinant <i>Hendra henipavirus</i> N protein expression and purification.....	52
3.1 Introduction	52

3.1.1 Recombinant protein expression	53
3.1.1.1 pET expression vectors	54
3.1.2 Principles of chromatography	56
3.1.3 Chapter aims	58
3.2 Results	58
3.2.1 Recombinant HeV N can be successfully expressed in <i>E. coli</i> BL21 (DE3) Rosetta2 ..	58
3.2.2 Chromatography can be utilised to purify HeV N protein	59
3.2.3 Sonication improves soluble HeV N protein yield	62
3.2.4 Decreasing NaCl concentration increases the presence of high molecular weight species.....	65
3.3 Discussion.....	68
3.3.1 Recombinant expression of paramyxoviral N proteins for 3D reconstruction	68
3.3.2 Purity and stability of recombinant proteins	69
3.3.2 Summary	70
Chapter 4: Transmission electron microscopy of <i>Hendra henipavirus</i> nucleocapsid-like structures	72
4.1 Introduction	72
4.1.1 Resolving power	74
4.1.2 Anatomy of a transmission electron microscope	75
4.1.3 Negative stain EM versus cryo-EM	77
4.1.4 Grid selection and preparation	78
4.1.5 Image formation	79
4.1.6 Image Detection.....	80
4.1.7 Chapter aims	81
4.2 Results.....	81
4.2.1 Negative stain electron microscopy reveals a range of HeV N particle morphologies	81
4.2.2 HeV NLFs are stable in poorer buffer conditions.....	83
4.2.3 HeV NLFs have an affinity for carbon on cryo-grids.....	83
4.2.4 Cryo-electron microscopy datasets can be collected automatically	87
4.3 Discussion.....	87
4.3.1 Comparison with other <i>Paramyxoviridae</i>	90
4.3.2 Comparison with published <i>Henipavirus</i> microscopy data.....	90
4.3.3 Summary	92
Chapter 5: Elucidation of the <i>Hendra henipavirus</i> N structure and nucleocapsid architecture	93
5.1 Introduction	93

5.1.1 Pre-processing of micrographs	93
5.1.2 Particle selection	94
5.1.3 Three-dimensional reconstruction	96
5.1.4 Post-processing and refinement of three-dimensional maps	98
5.1.5 Model building and refinement	99
5.1.6 Chapter aims	100
5.2 Results	100
5.2.1 HeV nucleoprotein sauronoids can be extracted and 2D classified	100
5.2.2 The HeV sauronoid assembly results in a 3.5 Å electron density map	104
5.2.3 A HeV nucleoprotein model can be generated from the sauronoid map	108
5.2.4 The HeV sauronoid model elucidates N-N and N-RNA interactions	111
5.2.4.1 The HeV nucleoprotein fold consists of two globular domains and two terminal arms	112
5.2.4.2 The HeV N protein RNA binding groove contains conserved N-RNA interactions	117
5.2.4.3 The HeV sauronoid reveals conserved and novel N-N interactions	117
5.2.5 Helical reconstruction of HeV nucleocapsid-like filaments generates low-resolution maps	121
5.2.8 Recombinantly expressed HeV nucleocapsid-like filaments generate clamshell architectures	128
5.3 Discussion	129
5.3.1 The HeV nucleoprotein structure has a canonical fold exhibiting novel features ...	129
5.3.2 <i>Hendra henipavirus</i> nucleocapsid architecture	136
5.3.3 Summary	137
Chapter 6: Generation of a <i>Hendra henipavirus</i> replicon system	138
6.1 Introduction	138
6.1.1 Replicons in virus research	138
6.1.2 <i>Hendra henipavirus</i> isolation and culturing	139
6.1.3 <i>In vitro</i> experimentation with <i>Hendra henipavirus</i>	139
6.1.3 Chapter aims	140
6.2 Results	140
6.2.1 Primary design of a <i>Hendra henipavirus</i> replicon system	140
6.2.2 The designed <i>Hendra henipavirus</i> replicon system is inefficient <i>in vitro</i>	143
6.2.2.1 Using NiV replicon plasmid ratios results in poor HeV replicon performance .	143
6.2.2.2 Optimising plasmid ratios fails to improve HeV replicon performance	144
6.2.3 Alterations to the <i>Hendra henipavirus</i> replicon system support plasmids	145
6.2.4 The redesigned <i>Hendra henipavirus</i> replicon system is inefficient <i>in vitro</i>	148

6.2.5 Transfection of <i>in vitro</i> transcribed RNA for the support plasmids fails to improve replicon efficiency	151
6.3 Discussion.....	153
6.3.1 The primary HeV replicon design is inefficient in BSR T7 cells	153
6.3.1 The inclusion of viral gene NTRs improves replicon performance	154
6.3.2 Suggested optimisation steps for future work	155
6.3.4 Summary	155
Chapter 7: Discussion	156
7.1 Expression and purification of recombinant <i>Hendra henipavirus</i> nucleoprotein.....	156
7.2 The <i>Hendra henipavirus</i> N protein and nucleocapsid imaged through electron microscopy.....	157
7.3 Structural reconstruction of the <i>Hendra henipavirus</i> sauronoid assembly and N protein	158
7.3.1 Contextualising the HeV sauronoid architecture.....	161
7.3.2 Future directions for structural work involving HeV nucleoprotein.....	164
7.4 The use of a replicon system as a tool for investigating <i>Hendra henipavirus</i> molecular biology.....	165
7.4.1 Future directions for work with the HeV minigenome-based replicon system.....	165
Appendix I	167
Supplementary Figures	167
Supplementary Tables	189
Bibliography	190

List of Figures

Figure 1.1. Schematics of virus classification.....	3
Figure 1.2. Nucleoproteins of the segmented negative-sense RNA viruses.....	7
Figure 1.3. Paramyxoviral nucleoproteins.	11
Figure 1.4. Reservoir distribution and transmission.	14
Figure 1.5. <i>Henipavirus</i> clade and <i>Hendra henipavirus</i> genome organisation.	16
Figure 1.6. <i>Hendra henipavirus</i> ultrastructure.....	18
Figure 1.7. The henipavirus life cycle.....	21
Figure 1.8. A model of the paramyxoviral RNA-dependent RNA polymerase activity on a nucleocapsid template.....	23
Figure 1.9. Henipaviral P gene polycistron.	25
Figure 1.10. Interactions of the P-N ⁰ complex.....	27
Figure 1.11. Henipaviral immune antagonism.....	30
Figure 3.1. Schematic of the pET expression system.....	55
Figure 3.2. Principles of chromatography.....	57
Figure 3.3. HeV N can be successfully expressed in Rosetta 2 cells.	60
Figure 3.4. HeV N can be expressed and purified through affinity chromatography in a small-scale format.	61
Figure 3.5. HeV N multimers do not separate into discrete species on Superdex S200 size exclusion chromatography.	63
Figure 3.6. Large-scale HeV N expression and first purification.	64
Figure 3.7. Large-scale HeV N expression and first purification.	65
Figure 3.8. Large-scale HeV N expression and second purification.	66
Figure 3.9. Large-scale HeV N expression and first purification.	67
Figure 4.1. Generation of biological structures.	73
Figure 4.2. Principles of transmission electron microscopy.	76
Figure 4.3. HeV N negative stain EM reveals three distinct morphologies.	82
Figure 4.4. HeV NLFs are stable in poorer buffers.	84
Figure 4.5. HeV NLFs show affinity for carbon on cryo-grids.....	86
Figure 4.6. Data collection on a UC Lacey grid.....	88
Figure 4.7. Particle orientation and transmission artefacts.....	89
Figure 4.8. NiV nucleocapsid-like filaments resemble those from HeV.	91
Figure 5.1. Overview of the SPA reconstruction workflow.....	94
Figure 5.2. The Fourier slice theorem and concepts of 3D reconstruction.	97
Figure 5.3. HeV single particle selection through manual and automated picking.....	101
Figure 5.4. Sauronoid 2D classification of manually picked particles generates reference for automated picking.	102
Figure 5.5. HeV sauronoid 2D classification reveals a range of viewing angles and detailed 2D morphology.....	103
Figure 5.6. Impartation of symmetry generates incongruent maps.....	105
Figure 5.7. 3D classification generates favourable map for refinement.	106
Figure 5.8. Refined sauronoid electron density maps reveal global architecture and secondary structure.....	107
Figure 5.9. Automated model building in Buccaneer.	109
Figure 5.10. <i>De novo</i> RNA structure in the RNA-binding groove.....	110
Figure 5.11. Dimer of tetradecamer sauronoid model.....	113
Figure 5.12. The HeV N protein fold and protein domains.....	114

Figure 5.13. HeV N protomer surface values.	115
Figure 5.14. HeV N sauronoid surface values.	116
Figure 5.15. The HeV N protein binds RNA through electrostatic and hydrophobic interactions.	118
Figure 5.16. Two HeV N protomers form an in-base cavity.....	119
Figure 5.17. HeV N_{i+1} Nt arm-loop residue Leu31 is buried in a pocket on the N_i NTD.	120
Figure 5.18. The HeV sauronoid elbow interactions bridge non-neighbouring protomers.	122
Figure 5.19. The HeV sauronoid intercalation zone is poorly resolved.	123
Figure 5.20. HeV helical segment selection through manual picking.	125
Figure 5.21. Helical HeV NLFs generate highly flexible segments in 2D classification.	126
Figure 5.22. Helical HeV NLFs generate variable 3D classes.....	127
Figure 5.23. Refinement of 3D helical HeV NLFs results in low-quality, low-resolution 3D maps.	128
Figure 5.24. Altering extraction and 2D classification parameters fails to improve 2D classes.	129
Figure 5.25. A subset of HeV NLFs contain clam-shell transitions.....	130
Figure 5.26. Comparison of HeV and NiV N protein structures.....	131
Figure 5.27. Mapping HeV variant amino acid substitutions.....	135
Figure 6.1. Schematic of HeV replicon design.	142
Figure 6.2. HeV replicon EGFP reporter levels are low.....	144
Figure 6.3. HeV replicon EGFP reporter levels are low after 6 hours transfection.....	146
Figure 6.4. HeV replicon EGFP reporter levels are low after 24 hours transfection.....	147
Figure 6.5. Schematic of HeV replicon redesign.	149
Figure 6.6. HeV replicon redesign reporter levels are low.	150
Figure 6.7. RNA transfection of HeV replicon redesign.	152

Supplementary Figures

Figure S3.1. Plasmid map of pET28(a)-6His-SUMO-HeV-N.....	167
Figure S3.2. HeV N protein parameters.....	168
Figure S4.1. HeV N cryo-grid screening.....	169
Figure S5.1. First round 2D classification for sauronoids from reference-based automated particle picking.....	170
Figure S5.2. Sauronoid 3D reference map and 3D classification.	171
Figure S5.3. Masking of the 3D refine sauronoid map.	172
Figure S5.4. Masking of the post-processed sauronoid map.....	173
Figure S5.5. Stereographs of the final sauronoid map.	174
Figure S5.6. Symmetry refinement and post-processing for the sauronoid map.	175
Figure S5.7. Homology model for HeV nucleoprotein.	176
Figure S5.8. Stereograph of <i>de novo</i> RNA structure in the RNA-binding groove.	177
Figure S5.9. HeV nucleoprotein correlation coefficients.....	177
Figure S5.10. Stereograph and organisation of the HeV N protomer.	178
Figure S5.11. The HeV N Nt-arm contains a solvent-facing, negatively charged patch.....	179
Figure S5.12. The HeV N protein RNA binding groove and in-base cavity.....	180
Figure S5.13. The HeV elbow interface and intercalation zone.....	181
Figure S5.14. Refinement and post-processing for the helical HeV NLF map.	182
Figure S5.15. Comparison of HeV and NiV N protein structures.	183
Figure S5.16. Nucleoprotein sequence alignment of HeV and HeV-var.	184

Figure S6.1. Plasmid maps of the HeV replicon system.....	185
Figure S6.2. Agarose gel of <i>in vitro</i> transcribed replicon RNA.....	186
Figure S6.3. Fluorescence and phase-contrast images of RNA-transfected cells.....	187
Figure S7.1. Protease treatment and supra-cryogenic incubation of HeV N protein.....	188

List of Tables

Table 2.1. Buffer and media compositions	39
Table 2.2. FragmentGENE segments.....	41
Table 2.3. Bacterial expression strains.....	42
Table 5.1 HeV sauronoid nucleoprotein model validation	112

Supplementary Tables

Table S1.1 Henipaviral nucleoprotein similarities	189
Table S5.1 HeV sauronoid nucleoprotein map and model parameters	189

Abbreviation Key

Abbreviation	Definition
(-)ssRNA	negative-sense single-stranded RNA
°C	degrees Celsius
μl	microlitre; 10 ⁻⁶ litres
2D	two-dimensional
3D	three-dimensional
6His	hexahistidine
AC	affinity chromatography
agRNA	antigenomic RNA
AngV	<i>Angavokely henipavirus</i>
Au	gold
BHK21	baby hamster kidney fibroblast
BSL	biosafety level
BUNV	<i>Bunyamwera orthobunyavirus</i>
CCD	charge-coupled device
CCHFV	<i>Crimean-Congo hemorrhagic fever orthonairovirus</i>
cryo-ET	cryo-electron tomography
CoV	coronavirus
COVID-19	coronavirus infectious disease 2019
CPE	cytopathic effect
cryo-EM	cryo-electron microscopy
Ct	C-terminal
CTD	C-terminal domain
CTF	contrast transfer function
Cu	copper
DARV	<i>Daeryong henipavirus</i>
DED	direct electron detector
DNA	deoxyribonucleic acid
DQE	detective quantum efficiency
dsRNA	double-stranded RNA
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
eIF2α	eukaryotic translation initiation factor 2 alpha
ELISA	enzyme-linked immunosorbent assay
EM	electron microscopy
EMDB	Electron Microscopy Data Bank
ER	endoplasmic reticulum

Abbreviation	Definition
ESCRT	endosomal sorting complexes required for transport
F	fusion (glycoprotein)
FSC	Fourier shell correlation
FT	flow-through
G	attachment (glycoprotein)
GAKV	<i>Gamak henipavirus</i>
GCU	green calibrated unit
GE	gene end
GhV	<i>Ghanaian bat henipavirus</i>
gRNA	genomic RNA
GS	gene start
h.p.t.	hours post-transfection
HDR	hepatitis delta virus ribozyme
HeV	<i>Hendra henipavirus</i>
HIV	<i>Human immunodeficiency virus</i>
HN	haemagglutinin-neuraminidase
hRNA	host-derived ribonucleic acid
HTNV	<i>Hantaan orthohantavirus</i>
IFN	interferon
IFNAR	interferon- α/β receptor
IG	intergenic
IHC	immunohistochemistry
IMAC	immobilised metal-ion affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
IRF	interferon-regulatory factor
ISG	interferon stimulated gene
ITN	innovative training network
Jak	Janus tyrosine kinase
kDa	kilodalton; 10^3 daltons
kV	kilovolt; 10^3 volts
L	large (protein)/polymerase
LaB6	lanthanum hexaboride
LacO	<i>lac</i> operator
LaV	<i>Langya henipavirus</i>
LLC-MK2	Lilly Laboratories Culture-Monkey Kidney 2
LN	liquid nitrogen
LoG	Laplacian of Gaussian

Abbreviation	Definition
M	matrix (protein)
mAb	monoclonal antibodies
MCS	multiple-cloning site
MDA5	melanoma differentiation-associated gene 5
MDBK	Madin-Darby bovine kidney
^{me} G	methylated 5' guanine
MeV	<i>Measles morbillivirus</i>
mg	milligram; 10 ⁻³ grams
MG	minigenome
ml	millilitre; 10 ⁻³ litres
ML	maximum likelihood
mM	millimolar; 10 ⁻³ molar
MRC5	Medical Research Council cell strain 5
mRNA	messenger RNA
MuV	<i>Mumps orthorubulavirus</i>
N	nucleoprotein; among nsNSV
N ⁰	unassembled nucleoprotein
Ncore	nucleoprotein core domain
NDV	<i>Avian orthoavulavirus 1</i> / Newcastle disease virus
NES	nuclear export sequence
NF-κB	nuclear factor κB
N _i	index nucleoprotein
NiV	<i>Nipah henipavirus</i>
NLF	nucleocapsid-like filament
NLS	nuclear localisation sequence
nm	nanometre; 10 ⁻⁹ metres
NMR	nuclear magnetic resonance
NP	nucleoprotein; among sNSV
nsNSV	non-segmented (-)ssRNA virus
Nt	N-terminal
nt	nucleotide/nucleotides
Ntail	nucleoprotein tail domain
NTD	N-terminal domain
NTR	non-translated region
ORF	open reading frame
ori	origin of replication
P	phosphoprotein

Abbreviation	Definition
PCR	polymerase chain reaction
PDB	Protein Data Bank
PI(4,5)P ₂	phosphatidylinositol 4,5-bisphosphate
PIV5	<i>Mammalian orthorubulavirus 5</i> / human parainfluenza virus 5
PKR	protein kinase R
POI	protein of interest
PrE	promoter element
PS	phosphatidylserine
pSTAT	phosphoSTAT/phosphorylated STAT
PTM	post-translational modification
RBP	RNA-binding protein
RdRp	RNA dependent RNA polymerase
RIG-I	retinoic acid-inducible gene I
RK13	rabbit kidney epithelial
RNA	ribonucleic acid
RNP	ribonucleoprotein
RSV	<i>Human orthopneumovirus</i> / respiratory syncytial virus
RT	room temperature
RVFV	<i>Rift valley fever phlebovirus</i>
SARS	severe acute respiratory syndrome
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SeV	<i>Murine respirovirus</i> / Sendai virus
sG	soluble G glycoprotein
SGD	stochastic gradient descent
SMT3	suppressor of MIF2 Smt3p
SNR	signal to noise ratio
sNSV	segmented (-)ssRNA virus
SPA	single-particle analysis
STAT	signal transducers and activators of transcription
SUMO	small ubiquitin-like modifier
T7 pol	T7 RNA polymerase
TEM	transmission electron microscopy
TGN	trans-Golgi network
TLR	toll-like receptor
ULP1	ubiquitin-like-specific protease 1
UV	ultraviolet

Abbreviation	Definition
VLP	virus-like particle
VNT	virus neutralisation test
VSP4	vacuolar protein sorting-associated protein 4
VSV	<i>Vesiculovirus indiana</i> / vesicular stomatitis virus
WHO	World Health Organisation

Chapter 1: Introduction

1.1 Zoonotic emergence

The transmission of pathogens from animals to humans, called zoonosis, can have devastating effects. Perhaps one of the most destructive examples in history is the 1918 Spanish influenza (flu) pandemic, caused by influenza A virus H1N1 (Webster, 1999). At the end of the Great War, this orthomyxovirus swept across the globe killing an estimated 50-100 million people, surpassing the loss of life as a result of the war itself (Johnson and Mueller, 2002). Extensive international travel, in large part through repatriation of soldiers from the Western front, aided the world-wide spread of the Spanish flu (Oxford et al., 2005). In the 100 years since the Spanish flu pandemic, advances have been made in our collective knowledge of pathogen biology, disease prevention, and treatment. However, some key factors that can lead to the emergence of zoonotic infections with pandemic potential are still, if not more, prevalent today. The global population has increased dramatically since 1918, resulting in expansion into remote areas and higher population densities, which increases the chances of human contact with previously unknown zoonotic viruses and the dissemination of those viruses between humans. In addition, international travel has never been more commonplace. Emerging viruses can use this vast network to spread globally with relative ease, potentially reaching larger urbanised cities from some of the world's most remote places in a matter of hours (Weiss et al., 2018).

The above text was composed by the author of this thesis in the summer of 2018 for the Transfer Report associated with this project and almost predictively warns of the ongoing *Severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2)* pandemic (Dong et al., 2020). Indeed, the risk of emerging zoonoses motivated the organisation of the MSCA-ITN HONOURS (van der Hoek et al., 2018) within the framework of which the work presented in this thesis was undertaken. The SARS-CoV-2 virus and the coronavirus infectious disease 2019 (COVID-19) it causes took the globe by storm. After first being reported in December of 2019, SARS-CoV-2 has infected over half a billion people and killed over 6 million people world-wide in two and a half years (Dong et al., 2020). Through international collaboration and advances in fundamental science, the response to SARS-CoV-2 was much more rapid than previous emergent zoonotic outbreaks (Kinsella et al., 2020). A number of SARS-CoV-2 vaccines were developed in record time, owing to the years of fundamental research on the closely related SARS-CoV ahead of the current outbreak (Li et al., 2021).

The ever-increasing global population is accompanied by an ever-increasing demand for resources and living space. As a result, human exposure to pathogens infecting wildlife becomes ever more commonplace, leading to 'spillover' events (Plowright et al., 2017). With

approximately 60% of all human infectious diseases being of zoonotic origin, this risk only continues to grow (Taylor et al., 2001). Zoonoses that become capable of human-to-human transmission can have catastrophic effects. The scale of death can reach terrible proportions, such as for the Spanish flu discussed above and for *Human immunodeficiency virus* (HIV), which has taken an estimated 36 million lives (unaids.org, 2019). Others may be less deadly but instead have a massive impact on our globalised civilisation, with SARS-CoV-2 causing the largest economic crisis since the Great Depression (IMF, 2020). The need for outbreak preparedness should be crystal clear in a world where outbreaks become increasingly commonplace. A rapid response is critical in mitigating infectious disease outbreaks since the number of infections may increase exponentially in a susceptible population. Outbreak preparedness thus ensures that contingency planning is in place before an outbreak occurs, avoiding a slower *ad hoc* response and allowing intervention whilst case numbers are low. For established diseases, preparedness includes systems for surveillance, laboratory confirmation, and a tailored response (World Health Organization, 1999). In contrast, the outbreak of a novel pathogen is accompanied by gaps in fundamental knowledge and technical resources. It is therefore of paramount importance to increase these key assets in advance of a potential outbreak.

1.2 Negative strand RNA viruses

Many of the zoonotic viruses that infect humans or our livestock and agriculture, including well known examples such as *Influenza A virus* (family *Orthomyxoviridae*), *Zaire ebolavirus* (family *Filoviridae*), and *Lyssavirus rabies* (family *Rhabdoviridae*) (Figure 1.1A), have negative-sense single-stranded RNA ((-)ssRNA) genomes and thus belong to group V of the Baltimore classification (Figure 1.1B) (Baltimore, 1971). The genomes of these (-)ssRNA viruses consist of one or more segments and exhibit a wide range in overall length and organisation. Indeed, a great variety in genome segmentation, length, and organisation has been discovered among viruses isolated from arthropod hosts, reflecting the wide genomic diversity (Li et al., 2015). Many (-)ssRNA viruses are important human pathogens through zoonotic or non-zoonotic transmission, such as *Crimean-Congo hemorrhagic fever orthonavirus* (CCHFV) and *Human orthopneumovirus* (previously respiratory syncytial virus; RSV). Additionally, many plant pathogens of economic or ecological interest also belong to the group of (-)ssRNA viruses, such as *Tomato spotted wilt tospovirus* and multiple members of the family *Rhabdoviridae*. Both negative-sense and positive-sense RNA viral genomes exhibit rapid mutation and evolution compared to the genomes of DNA viruses, largely owing to the high error rate and lack of proof-reading in the viral RNA dependent RNA polymerase (RdRp) (Venkataraman et al., 2018).

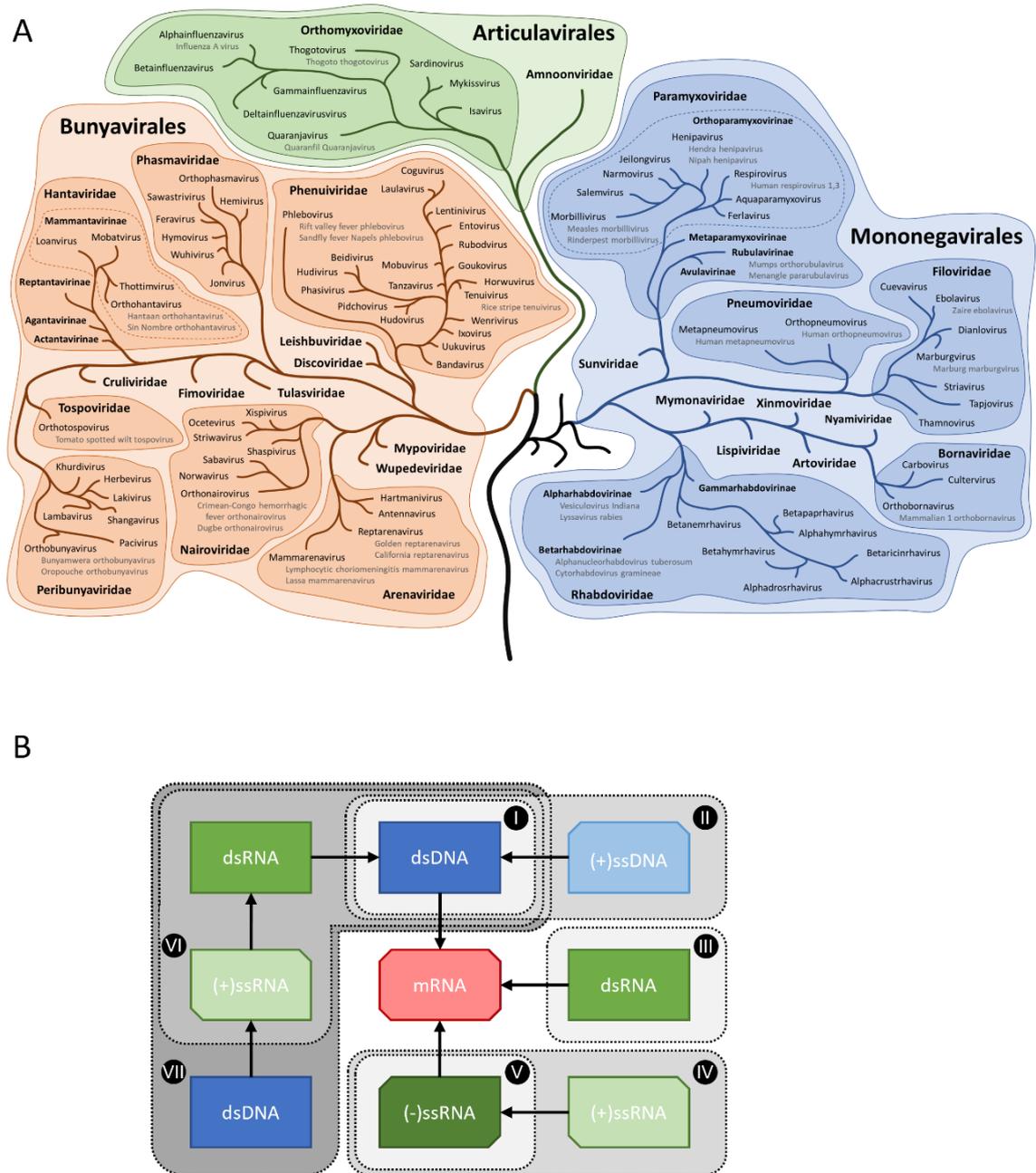


Figure 1.1. Schematics of virus classification.

A) A schematic representation of the *Negarnaviricota* phylum. The *Articulavirales* and *Bunyavirales* orders within the *Polyploviricotina* subphylum are represented in green and orange, respectively. Only the order *Mononegavirales* from the *Haploviricotina* subphylum is represented in detail (blue), with the remaining orders *Muvirales*, *Serpentovirales*, *Jingchuvirales*, and *Goujianvirales* represented by the stubbed black branches. Branch lengths do not represent phylogenetic distance. **B)** A schematic representation of the Baltimore classification system for viruses (Baltimore, 1971). Roman numerals indicate the seven viral genome expression strategies, each contained within a dashed grey box. Arrows indicate the order in which different types of nucleic acid (coloured boxes) are used, with the first in the sequence indicating the nucleic acid type of the viral genome within the virion. All strategies converge on mRNA transcription (red).

1.2.1 Segmented negative strand RNA viruses

The segmented (-)ssRNA virus (sNSV) group contains viruses that infect mammals, reptiles, birds, insects, and plants, causing a large variety of symptoms. The majority of sNSVs are found in the order *Bunyvirales*, containing nearly 500 species in 14 families (Figure 1.1A) (Fenner et al., 1974; Walker et al., 2020; International Committee on Taxonomy of Viruses (ICTV), 2020), which includes examples such as CCHFV, *Rift Valley fever phlebovirus* (RVFV), and *Bunyamwera orthobunyavirus* (BUNV). Most have bipartite or tripartite genomes, with respectively large and small segments, or large, medium, and small segments. The second major sNSV group is the order *Articulavirales*, which contains among others *Influenza A virus* and *Thogotho thogothovirus*. Members of the order *Articulavirales* have 6 to 8 genomic segments which are commonly named after the main open reading frame (ORF) they code for. In addition to the high mutation rates seen in other RNA viruses, the segmented nature of the sNSV genomes allows for reassortment during co-infection of two or more closely related viruses (Flint et al., 2009d). Through evolutionary time, antigenic drift – which is caused by error prone, RdRp-driven genome replication (in the order of one mutation per 10^5 nucleotides) – occurs at a relatively constant pace (Barr and Fearn, 2016; Sanjuán and Domingo-Calap, 2016). Contrastingly, the reassortment of whole genome segments, referred to as antigenic shift, occurs in a more stochastic fashion. The resulting reassortant virus may have drastically different host-tropism and pathogenicity compared to its founders. Furthermore, the increase in genetic diversity through RNA recombination – which results from template switching events during RNA replication – has been reported for this group of viruses, although it may be a rare occurrence (Shi et al., 2017).

1.2.2 Non-segmented negative strand RNA viruses

Most of the non-segmented (-)ssRNA viruses (nsNSVs) are part of the order *Mononegavirales*, containing nearly 500 species in 11 families (Figure 1.1A) (Fenner et al., 1974; Walker et al., 2020; International Committee on Taxonomy of Viruses (ICTV), 2020). Viruses within the nsNSV group include important zoonotic and non-zoonotic human pathogens, such as *Lyssavirus rabies*, *Zaire ebolavirus*, RSV, *Mumps orthorubulavirus* (MuV), and *Measles morbillivirus* (MeV). Viral evolution within nsNSVs is primarily driven by antigenic drift, owing to the low fidelity of the RdRp and RNA recombination to a lesser extent, since the non-segmented nature of these viral genomes precludes viral evolution through antigenic shift (Spann et al., 2003; Satharasinghe et al., 2016; Barr and Fearn, 2016; Sanjuán and Domingo-Calap, 2016; Ding et al., 2017). Furthermore, the transcriptional programme offers another

fundamental difference between the sNSVs and nsNSVs. sNSVs only generate a single mRNA transcript from each RNA template, ambi-sense segments notwithstanding, whereas nsNSVs generate multiple mRNA transcripts from the RNA template through 'start-stop' transcription (see Section 1.7.2) (Plempner and Lamb, 2021).

1.3 Nucleoproteins and ribonucleoproteins

While a variety of encapsidation strategies exists for the genomes of positive sense single strand RNA viruses, those of (-)ssRNA viruses are universally completely covered by viral nucleoproteins throughout the viral life cycle. The nucleoproteins of sNSVs and nsNSVs are often abbreviated as NP or N, respectively. When NP or N forms oligomeric complexes in association with RNA, they are generally referred to as ribonucleoprotein (RNP) complexes or nucleocapsids. The interaction between nucleoprotein and RNA in the context of a nucleocapsid or RNP must be stable yet reversible. All available structural information on the nucleoprotein suggests that the RNA bases are largely inaccessible whilst RNA is bound within the nucleocapsid or RNP. Furthermore, the polymerase structures elucidated thus far indicate that the channel leading to the polymerase active site is insufficiently wide to accept a fully encapsidated template. Access to the RNA template by the RdRp therefore likely occurs by displacing one or more nucleoprotein subunits, locally freeing up the RNA to pass through the polymerase active site. This complex approach is advantageous to the virus as complete coverage of the viral RNA by nucleoproteins likely confers protection against innate immune surveillance and degradation by cellular nucleases, and prevents the annealing of genomic RNA with *trans*-complementary sequences such as the antigenomic RNA or mRNA as well as limiting the occurrence of *cis*-complementary secondary RNA structures (Bloyet, 2021; Plempner and Lamb, 2021). The RNP or nucleocapsid is thought to bind to the viral RdRp when packaged in a virion, since the nature of the (-)ssRNA genome necessitates the presence of an RdRp for the first round of transcription following viral entry into a permissive host cell (Flint et al., 2009c; Ruigrok et al., 2011; Reguera et al., 2014; Plempner and Lamb, 2021). While describing nucleoproteins within the context of oligomeric assemblies, such as RNPs or nucleocapsids, the NP or N protomer in focus is topographically referred to as the index nucleoprotein (N_i) while the adjacent protomers upstream and downstream of the N_i are referred to as index + x nucleoprotein (N_{i+x}) or index - x nucleoprotein (N_{i-x}), respectively. The integer x represents the distance in N protomers from N_i along the RNA chain, e.g. the N protomers immediately adjacent to N_i are N_{i+1} and N_{i-1} .

1.3.1 Nucleoproteins of the segmented negative strand RNA viruses

Most of the structural information on sNSV NPs has been obtained under recombinant expression rather than using proteins purified from virus. Indeed, recombinant expression facilitates structural elucidation through protein X-ray crystallography and many of the NP monomer structures have been solved this way. Transmission electron microscopy (TEM) is an important tool to investigate NP oligomeric heterogeneity and is amenable to both recombinantly expressed and native virion samples (Arragain et al., 2019; Hopkins et al., 2022). A common feature among sNSV NPs is the ‘fist and thumb’ conformation, a conserved protein fold with a large globular domain and an extended stalk (Figure 1.2). Beyond this similar fold, the individual protein structures can be highly dissimilar. The stalk can be either C-terminal (Ct) as for BUNV, N-terminal (Nt) as for RVFV, or in the middle of the peptide sequence as for CCHFV (Figure 1.2) (Ferron et al., 2011; Carter, Surtees, et al., 2012; Li et al., 2013). Other structural features are also observed, such as the Nt tail of BUNV NP (Li et al., 2013).

Recombinantly expressed NPs can form oligomeric rings, either bound to RNA or in an RNA-free, apo-state (Ferron et al., 2011; Carter, Barr, et al., 2012; Raymond et al., 2012; Li et al., 2013). Different oligomeric forms of the same recombinant viral NP can be observed on a single electron micrograph. BUNV, for example, exhibits tetrameric, pentameric, and hexameric rings and these data agree well with crystallography-derived structures (Li et al., 2013). RVFV NP tetramers, pentamers, and hexamers have also been observed in electron micrographs and the crystal structures have been solved, illustrating the flexible nature of sNSV NP-NP interactions (Raymond et al., 2012). The rigidity of these rings makes them highly amenable to crystallisation and subsequent high-resolution structural elucidation through X-ray crystallography, however the chemical environment during crystallisation is far removed from those found within the virion or infected cell.

In electron micrographs, both recombinant and native sNSV RNPs are normally visible as flexible circles or strands, although stretches of rigidly packed RNPs are also found. One such example is the *La Crosse orthobunyavirus*, which shows putative helically coiled RNPs extracted from virions or even partially or fully supercoiled (Reguera et al., 2013). However, the RNP unwinds after incubation at 37°C to form a long string of single NPs bound to RNA; an architecture that resembles beads on a string (Reguera et al., 2013). This observation is corroborated by electron micrographs of isolated native RVFV RNPs, which show a much narrower diameter than the recombinant RVFV NP hexameric ring structure (Raymond et al., 2012). These previous studies, and others, illustrate the correlation between the crystal structures of recombinant tetrameric, pentameric, or hexameric NP rings and micrographs of putative helical native RNPs. Topologically, a split between adjacent monomers in an NP ring

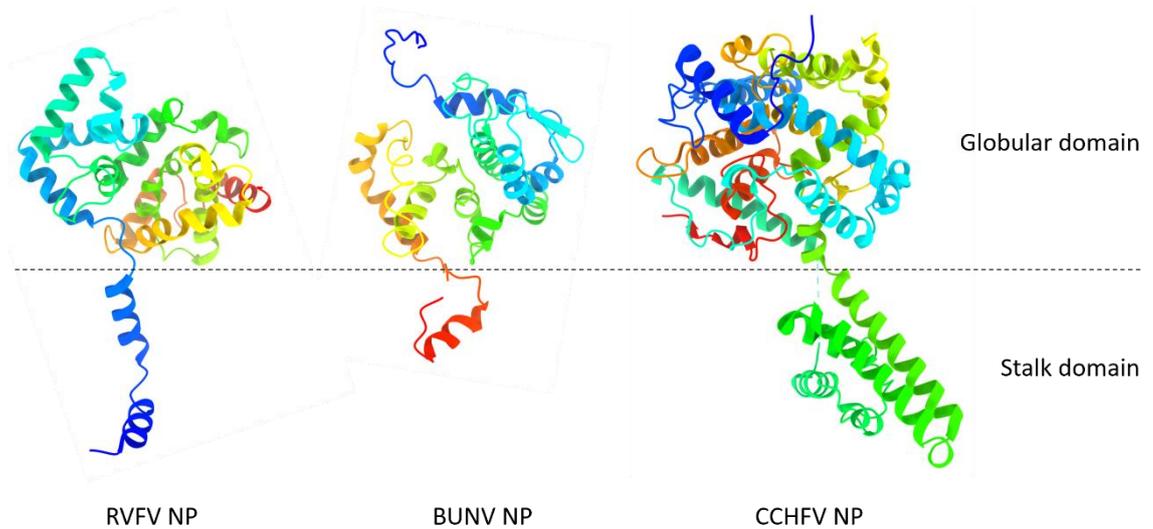


Figure 1.2. Nucleoproteins of the segmented negative-sense RNA viruses.

Single sNSV NP protomers for RVFV (left), BUNV (middle), and CCHFV (right). The NPs consist of a globular domain and an extended stalk domain. Models are presented as cartoons with rainbow colouring ranging from blue at the N-terminus to red at the C-terminus. Models were generated from RVFV NP (PDB: 3OV9) (Ferron et al., 2011), BUNV NP (PDB: 4IJS) (Li et al., 2013), and CCHFV NP (PDB: 4AKL) (Carter, Surtees, et al., 2012).

and subsequent adjustment in the angles of the flexible NP-NP interactions enables the generation of helical assemblies through stacking of split-ring subassemblies, which are analogous to the helical RNPs observed under EM. Importantly, the helical architecture of sNSV RNPs was recently solved through cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) approaches for recombinantly expressed RNPs from *Hantaan orthohantavirus* (HTNV) (Arragain et al., 2019) and native RNPs isolated from purified BUNV virions (Hopkins et al., 2022). The cryo-EM single particle analysis of recombinant HTNV RNP resulted in a 3.3 Å structure showing a clear left-handed helix (Arragain et al., 2019). Furthermore, a 13 Å cryo-EM density map, along with measurements through atomic force microscopy and cryo-ET for the native BUNV RNP unequivocally show a left-handed helical architecture. Where the HTNV structure was solved *de novo*, for BUNV the crystal structures of BUNV NP monomers were fitted into the cryo-EM density map. Both these structures highlight NP-NP and NP-RNA interactions critical for the formation of the helical assembly in recombinant and native RNPs (Arragain et al., 2019; Hopkins et al., 2022).

The RNPs found in a subset of sNSVs have been well characterised, best exemplified by those from the genus *Alphainfluenzavirus*. The 3' and 5' ends of the genomic segments in this group of viruses are *cis*-complementary and form a panhandle structure that associates with the viral RdRp (Hsu et al., 1987). The remainder of the RNA molecule is similarly encapsidated by NP, however the RNP uniquely takes on a super-coiled double helical architecture (Compans et

al., 1972). NP and RNP structures for this group have similarly been solved at medium and high-resolution through X-ray crystallography and EM approaches, respectively (Ng et al., 2008; Ruigrok et al., 2010; Moeller et al., 2012; Coloma et al., 2020).

Overall, the sNSV NPs show a high degree of flexibility in the formation of oligomers of different oligomeric states and as a result sNSV RNPs range from double helical, to helical, to beads on a string. The flexible nature of the sNSV RNP is critical due to the requirement for both ends of each RNA segment to contact a viral RdRp complex simultaneously, which is true for both alphainfluenzavirus and non-alphainfluenzavirus sNSVs. This requirement does not appear to exist for nsNSV nucleocapsids, which show overall less flexibility and more rigid helical ordering.

1.3.2 Nucleoproteins of the non-segmented negative strand RNA viruses

The N proteins of most nsNSVs are very similar in structure, generally having a core (Ncore) domain of two distinct globular sections composed predominantly of alpha-helices, with extended arms on both the N- and C-termini and beyond the Ct-arm a Ct-extended, intrinsically disordered domain called Ntail (Figure 1.3). Protein-protein interactions between neighbouring N proteins, including globular domain interactions and Nt and Ct flexible arm interactions as well as N protein-RNA interactions are involved in the formation of N protein oligomers and nucleocapsids. In most circumstances, nsNSV N proteins bind RNA to form helical nucleocapsids which can be imaged through EM. Although these helical nucleocapsids are highly ordered structures, they still retain some degree of flexibility, as is observed for the recombinant nucleocapsids of RSV (Tawar et al., 2009; Bakker et al., 2013) and MeV (Schoehn et al., 2004). The large overall size and repetitive, helical nature of the nsNSV nucleocapsids lend themselves to structural elucidation from EM data. A low-resolution structure of a recombinant RSV nucleocapsid broadly showed the organisation and left-handedness of the nucleocapsid rod (Bakker et al., 2013). The crystal structures of RSV and *Nipah henipavirus* (NiV) formed the basis of a homology model that was built into the cryo-EM density map of MeV nucleocapsids, resulting in the high-resolution pseudo-atomic model of the MeV nucleocapsid at 4.3 Å (Gutsche et al., 2015). However, high-resolution structures of native nucleocapsids are scarce, in large part owing to their flexible nature which hinders high-resolution reconstruction. Some progress has been made for *Zaire ebolavirus* and MuV where cryo-ET and cryo-EM have provided insight into both recombinant and native nucleocapsid structures and architectures (Cox et al., 2014; Wan et al., 2017). Furthermore, native nucleocapsids structures have been elucidated for *Marburg marburgvirus* and *Vesiculovirus indiana* (formerly vesicular stomatitis virus; VSV)

(Bharat et al., 2011; Si et al., 2022) as well as the molecular mechanisms of virion budding in members of the family *Filoviridae*. The architecture of the nucleocapsid may be closely linked to the overall morphology of the virion that contains them. This is particularly clear for the family *Rhabdoviridae*, for which both the nucleocapsid and the virion exhibit a characteristic ‘bullet’ shape (Riedel et al., 2020).

Beyond the helical nucleocapsid architecture, additional unique morphologies have been observed. The formation of N protein-RNA ring complexes has been described for the recombinant nucleoproteins of a variety of viruses. Although the N proteins are structurally similar, rings with different numbers of N monomers are observed. For example, RSV and VSV N-RNA rings are decameric and *Lyssavirus rabies* N-RNA rings are undecameric (Albertini et al., 2006; Green et al., 2006; Tawar et al., 2009). Not only the number of N monomers per ring can differ but also the orientation in which the viral RNA is bound, as is illustrated by the binding of RNA in the RSV and VSV rings (Green et al., 2006; Tawar et al., 2009). When a single N monomer of each is aligned, the RNA rings extend out in opposite directions. Interestingly, features that resemble N-RNA rings were observed packaged within RSV virions (Conley et al., 2022). While their biological function remains unclear, they are invaluable for structural understanding of N-N and N-RNA interactions at high resolution.

1.3.2.1 Nucleoproteins and nucleocapsids of the family *Paramyxoviridae*

The N proteins of members of the family *Paramyxoviridae* universally form helical nucleocapsids associated with the viral genomic RNA. In the helical nucleocapsids of *Murine respirovirus* (formerly Sendai virus; SeV), approximately 13 N monomers constitute a helical turn with a pitch of around 5.5 nm (Egelman et al., 1989). With slight variations, these parameters generally apply to other paramyxoviruses. The N proteins of paramyxoviruses are unusual for RNA-binding proteins as they lack the conventional RNA-binding motifs found in cellular RNA-binding proteins (Re et al., 2014). Rather, their overall acidic charge drives electrostatic and hydrophobic N-RNA interactions. The association of RNA and N proteins into a nucleocapsid is very stable and N-RNA interactions are independent of RNA sequence. Moreover, the N protein has inherent self-assembly properties as it will readily form oligomers in the absence of other viral components suggesting that N-N interactions are a driving force of nucleocapsid assembly (Plempner and Lamb, 2021). While the helical nucleocapsid assembly is the predominant form in infected cells and virions, additional morphologies have been observed, including rings and clam-shell structures (Bloyet, 2021). The clam-shells are formed of two helical turns engaged in back-to-back stacking and have been found for *Avian orthoavulavirus 1* (formerly Newcastle

disease virus; NDV), SeV, and NiV (Song et al., 2019; Zhang et al., 2021; Ker et al., 2021). Outside of their role in genome and antigenome encapsidation, an additional role for *Hendra henipavirus* (HeV) and NiV N proteins has recently been found in immune antagonism (see section 1.7.3 for more details) (Sugai et al., 2017). Other nsNSV N proteins may additionally fulfil similar or yet different roles in the viral life cycle.

Like other nsNSV N proteins, the N proteins of paramyxoviruses contains a globular core domain (Ncore) and an intrinsically disordered tail region (Ntail) (Figure 1.3). While the both the sequence and organisation of the Ncore are well conserved among members of the family *Paramyxoviridae*, the Ntail sequence and length are poorly conserved (Bloyet, 2021). Moreover, the Ntail is largely dispensable for RNA binding and formation of the nucleocapsid assembly (Curran et al., 1993; Plemper and Lamb, 2021; Bloyet, 2021). The Ntail likely imparts a degree of flexibility to the helical nucleocapsid, suggested by the formation of highly rigid nucleocapsid rods upon deletion or proteolytic removal of the Ntail (Heggeness et al., 1980; Schoehn et al., 2004; Plemper and Lamb, 2021; Bloyet, 2021). Furthermore, the Ntail appears to be critical for interactions with the RdRp during genome transcription and replication as truncation of the Ntail renders the nucleocapsid non-functional as a template in MeV replicon assays (Zhang et al., 2002; Plemper and Lamb, 2021). Indeed, the Ntail contains multiple short, conserved domains that are shown to interact with the phosphoprotein (P) – which is part of the RdRp complex – and matrix protein (M), strongly indicating a role for the Ntail in N protein recruitment to the replicase complex as well as during virion assembly (Bloyet, 2021). In particular, a disordered but conserved molecular recognition element (MoRE) in the Ntail folds into an alpha helix (α -MoRE) upon interaction with the P protein C-terminal domain (CTD) and thus participates in the recruitment of the RdRp on its template (Bloyet, 2021). Furthermore, interactions with the P protein are thought to maintain the N protein in a soluble, unassembled, and RNA-free state (N^0) (Plemper and Lamb, 2021; Bloyet, 2021).

The Ncore can be roughly split into two distinct globular domains, the N-terminal domain (NTD) and the CTD (Figure 1.3A). These two globular domains are respectively preceded and followed by the Nt and Ct flexible arms, which can differ in length between family members. The RNA sits within an exterior facing, positively charged cleft between the NTD and CTD of the Ncore (Figure 1.3B). The paramyxoviral Ncore central region contains a highly-conserved F-X4-Y-X3- ϕ -S- ϕ -A-M motif – with X being any residue and ϕ being an aromatic residue – that is essential for N-RNA self-assembly and may be involved in N-N and N-RNA interactions (Myers and Moyer, 1997; Alayyoubi et al., 2015; Gutsche et al., 2015; Matsumoto et al., 2018). Overall, N-RNA interactions are likely independent of nucleotide sequence and occur mainly through interactions with the phosphodiester backbone, aided by base stacking of the RNA nucleotides

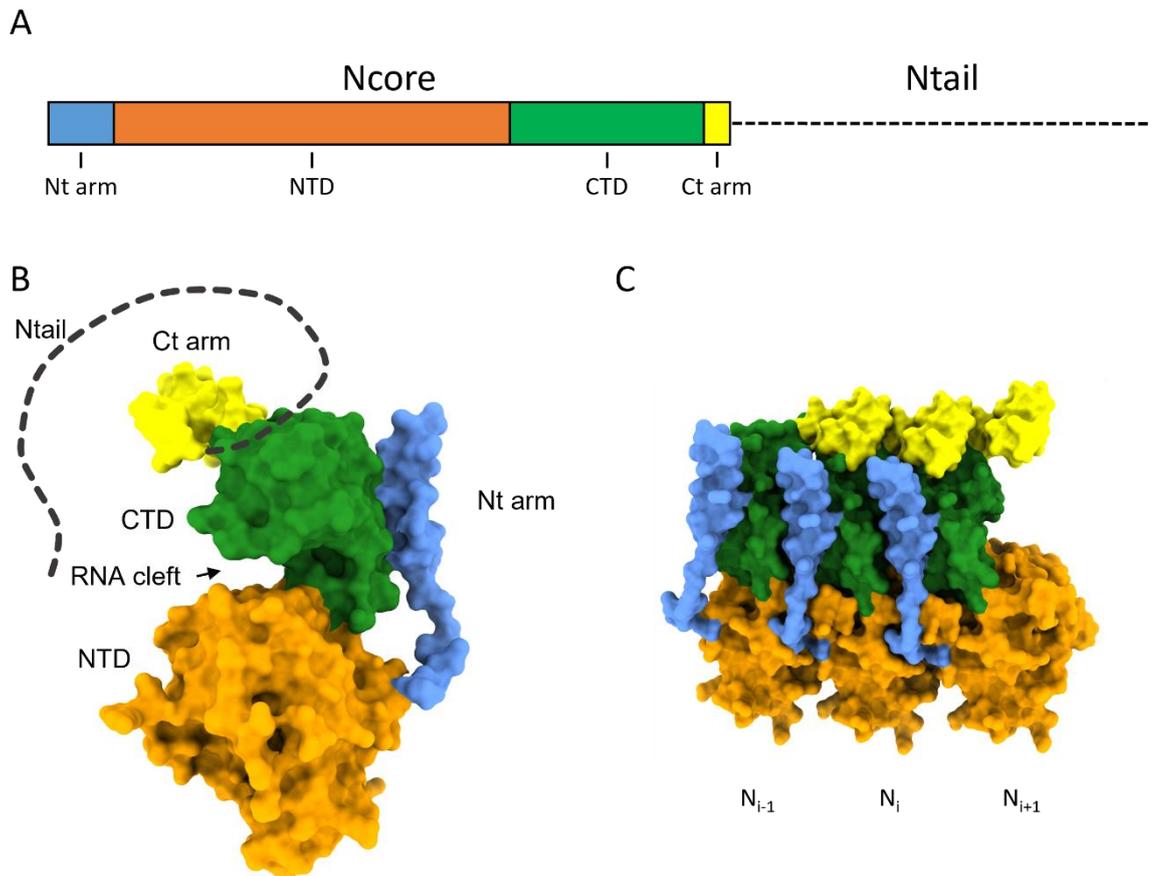


Figure 1.3. Paramyxoviral nucleoproteins.

A) Schematic, linear representation of the paramyxoviral nucleoprotein domain architecture indicating the Ncore and Ntail regions. The Ncore consists of the N-terminal domain (NTD) and C-terminal domain (CTD), each of which are flanked by a flexible arm; the N-terminal (Nt) arm and the C-terminal (Ct) arm. **B-C)** A single SeV N protomer (**B**) and SeV N trimer (**C**). (**B**) The Ncore region forms two globular domains, the NTD and CTD, separated by the RNA binding cleft. Each of the globular domains has a flexible arm, the Nt arm and Ct arm. The Ntail extends past the Ct flexible arm. The Ntail is poorly conserved and intrinsically disordered. (**C**) The SeV trimer illustrates the extensive interactions between N monomers and indicates the topology in N multimers. Models were generated from the cryo-EM SeV N protein model (PDB: 6M7D) (Zhang et al., 2021).

(Plempner and Lamb, 2021; Bloyet, 2021). However, the initiation of RNA encapsidation by the N protein appears to be, in part, sequence specific since mixing MeV N protein with RNA hexamers corresponding to the MeV 5' genomic end or poly-A hexamers results in the assembly of nucleocapsid-like assemblies and exhibiting a strong change in fluorescence anisotropy while this was not the case for poly-U hexamers (Milles et al., 2016; Bloyet, 2021). The N proteins of a number of viruses including MeV, VSV, and RSV preferentially encapsidate poly(A) RNA hexamers, but not poly(U), and RNA hexamers that correspond to the genomic 3' end (Bloyet, 2021). Further oligomerisation of N protein is not dependent on the length of RNA (Bloyet, 2021). These findings indicate that the initial binding of an N protomer to the nascent RNA is critical in ensuring full encapsidation of the RNA genome or antigenome. Paramyxoviral N

proteins bind RNA in a ‘three in three out’ conformation, where an ‘in’ three-base stack faces towards the interior of the Ncore and another ‘out’ three-base stack faces out towards the solvent. While there is some variation in the order of ‘in’ bases and ‘out’ bases, the three-base stack facing inwards is normally contained by a single N protomer while the three-base stack facing the solvent crosses the N-N boundary (Bloyet, 2021). The Ncore NTDs of each helical turn also cover the solvent-facing bases of the downstream helical turn, offering further protection for these bases.

In the context of a nucleocapsid assembly, the Nt-arm of N_i interlocks with the CTD of the N_{i-1} protomer and the N_i Ct-arm interacts with the N_{i+1} protomer, forming a stable chain-linked arrangement (Figure 1.3C) (Plempner and Lamb, 2021). The unassembled, soluble N protein is found in complex with the P protein outside of the replication machinery through domains that are distinct from those necessary for the N-P interactions during template reading (Section 1.7.2.2 and Figure 1.10) (Plempner and Lamb, 2021; Bloyet, 2021).

1.4 Hendra henipavirus discovery and classification

In 1994, an infectious agent emerged on the racing tracks of Hendra, a suburb of Brisbane, Australia (Murray et al., 1995; Selvey et al., 1995). A pregnant mare, the index case, died two days after arrival at the Hendra stables. Within 11 days, a further 17 horses fell ill and of those 14 died or were put down. Two humans – a stable hand and a trainer – developed symptoms within six days after the death of the index case. While the stable hand recovered, the trainer was admitted to hospital with exacerbated symptoms and suffered a fatal cardiac arrest a week after hospital admission (Selvey et al., 1995). A disease aetiology was established by inoculation of organ homogenates from two of the affected horses into recipient horses as well as a range of cell culture monolayers (Murray et al., 1995). Morphological, serological, and partial sequencing of the M gene suggested that the viral agent responsible belonged to the genus *Morbillivirus* and it was given the name *Equine morbillivirus* (Murray et al., 1995). Later, sequencing of the full genome suggested the virus stood sufficiently apart from the genus *Morbillivirus* and the other genera in the family *Paramyxoviridae* to warrant its own genus and was renamed *Hendra virus* (Yu et al., 1998), and currently *Hendra henipavirus* (HeV) within the family *Paramyxoviridae*, order *Mononegavirales* under the latest taxonomy release (International Committee on Taxonomy of Viruses (ICTV), 2020). After the genus *Henipavirus* was established, with HeV as the type species, discovery and sequencing of novel viruses and subsequent phylogenetic analyses have placed HeV together with its closest relative NiV (discovered in 1998) and further added *Cedar henipavirus*, *Mojiang henipavirus*, and *Ghanaian*

bat henipavirus (GhV) to the genus (Figure 1.1A) (Harcourt et al., 2000; International Committee on Taxonomy of Viruses (ICTV), 2020; B. Lee et al., 2021). The genus *Henipavirus* was further expanded over the last two years with the discovery of four new species; *Gamak henipavirus* (GAKV) and *Daeryong henipavirus* (DARV) from Korea (S.-H. Lee et al., 2021), *Langya henipavirus* (LaV) from China (X.-A. Zhang et al., 2022) and *Angavokely henipavirus* (AngV) from Madagascar (Madera et al., 2022), as well as a new variant strain of HeV (HeV-var/HeV-g2) which was discovered in Queensland, Australia in 2021 (Annand et al., 2022) (Figure 1.5A). Protein BLAST analysis of the N proteins for each of these viruses compared to HeV reveal a range of identity scores which largely follows the phylogeny, with the highest score for NiV (92%) and the lowest score for AngV (49%) (Supplementary Table S1.1).

1.5 Hendra henipavirus epidemiology, transmission, and symptoms

Symptomatic horses in the first outbreak suffered from depression, anorexia, fever, dyspnoea, ataxia, tachycardia, tachypnoea, and nasal discharge. Contact with the nasal discharge is believed to have infected the stable hand and trainer, who both suffered from a fever-like illness. The trainer developed pneumonitis, respiratory failure, renal failure, and arterial thrombosis and ultimately suffered a fatal cardiac arrest seven days after hospital admission (Selvey et al., 1995). Subsequent outbreaks along the Australian east coast have led to additional fatalities among horses and humans, with a mortality rate of around 60-75% for both. HeV exhibits a predominantly respiratory and neurological tropism, which explains the common symptoms of HeV infection including pulmonary haemorrhage and oedema, encephalitis, and meningitis (Eaton et al., 2006; Pearce et al., 2015). Although disease progression is generally rapid and most infected individuals die acutely within days, or in the case of horses are euthanised, one human case of recovery and subsequent fatal relapse 14 months after infection has been reported (O’Sullivan et al., 1997).

Fruit bats belonging to the genus *Pteropus* have been identified, through serological sampling of wild populations, as the natural reservoir of HeV (Young et al., 1996; Halpin et al., 2000). All recorded spillover events of HeV have occurred in Australia, however the natural distribution range of pteropid bats extends much further north into Asia, potentially extending the at-risk area (Figure 1.4A) (World Health Organization, 2008). Indeed, sampling among bats and humans in Africa has uncovered neutralising antibodies against HeV, NiV, and GhV in both populations (Pernet et al., 2014).

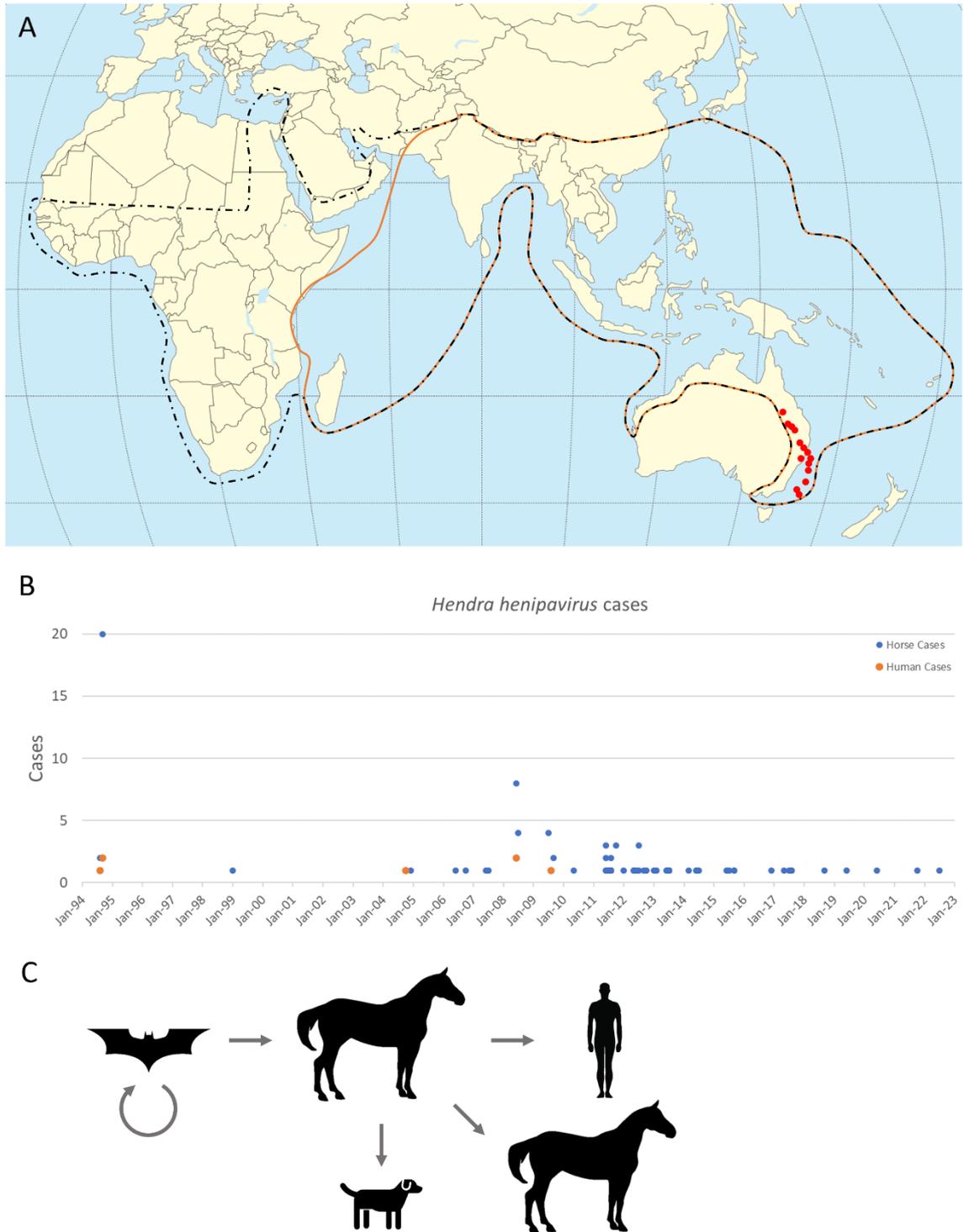


Figure 1.4. Reservoir distribution and transmission.

A) Spillover events of HeV are indicated by red dots, which may represent one or multiple spillover events. The distribution range of bats from the *Pteropus* genus is indicated in orange, while the range of the family *Pteropodidae* is indicated by the black dotted line. Data obtained from (Eaton et al., 2006; World Health Organization, 2008; Field et al., 2011). **B)** Chronology of HeV cases in both horses (blue) and humans (orange). All cases in horses ended in fatality, either from disease or euthanasia. Data obtained from (B. Lee et al., 2021; Queensland Government, 2022). **C)** Schematic of spillover HeV transmission, from the reservoir *Pteropus* spp. into horses. Subsequent HeV spread occurs through inter-equine transmission as well as into pets, such as dogs, and zoonotic transmission into humans.

Seropositivity among humans was found almost exclusively among individuals dealing in bat bush meat or living in areas of active deforestation (Pernet et al., 2014), highlighting the increasing risk of spillover of emerging zoonotic viruses due to human encroachment on natural landscapes. Between 1994 and 2022, there have been 66 recognised and suspected outbreaks of HeV, with documented annual outbreaks since 2006 (Figure 1.4B) (Playford et al., 2010; Queensland Government, 2022). Documented HeV transmission is predominantly through contact with bodily fluids and the data reported on spillover events suggests that inter-equine transmission occurs to a degree (Murray et al., 1995; Selvey et al., 1995). No direct transmission from bats to humans, or humans to humans, has yet been described with all human cases the result of contact with infected horses (Figure 1.4C). However, highly productive inter-human transmission of the closely related NiV has been observed in an outbreak with a mortality rate of around 75%, suggesting that similar outbreaks of HeV with inter-human transmission may occur in the future (Gurley et al., 2007).

1.6 Hendra henipavirus organisation

1.6.1 *Hendra henipavirus* genome and genes

As with all other members of the order *Mononegavirales*, the HeV genome consists of negative-sense single-stranded RNA. The genome is 18,234 nucleotides (nt) long and contains six genes (Figure 1.5B) which are flanked by two extracistronic sequences; a 3' leader of 56 nt and a 5' trailer of 33 nt. The leader and trailer sequences are partially complementary, with 24 of the 33 nt in the trailer aligning with the leader (Figure 1.5C) (Wang et al., 2000). Following the convention used for the family *Paramyxoviridae*, 'gene' refers to a genomic sequence encoding a single mRNA that may contain one or more open reading frames (ORFs) and encode one or more proteins (Plempner and Lamb, 2021). These genes are, from 3' to 5', the nucleoprotein (N), the phosphoprotein (P), the matrix protein (M), the fusion glycoprotein (F), the attachment glycoprotein (G), and the polymerase (L) (Figure 1.5B). The genome organisation, abbreviated to 3'-N-P-M-F-G-L-5' is identical to that of the genera *Morbillivirus* and *Respirovirus*, although in these the G gene is named HN for haemagglutinin-neuraminidase. While most henipaviral genes have a typical length for those from paramyxoviruses, the P gene is much larger than homologous genes in the family (B. Lee et al., 2021). Moreover, much of the extra length of the henipaviral genome is due in part to uniquely long untranslated intergenic sequences as well as the significantly extended 3' untranslated regions (UTRs) of the N, P, and G genes (B. Lee et al., 2021). The P gene also codes for the V, W, and C accessory proteins, which are expressed

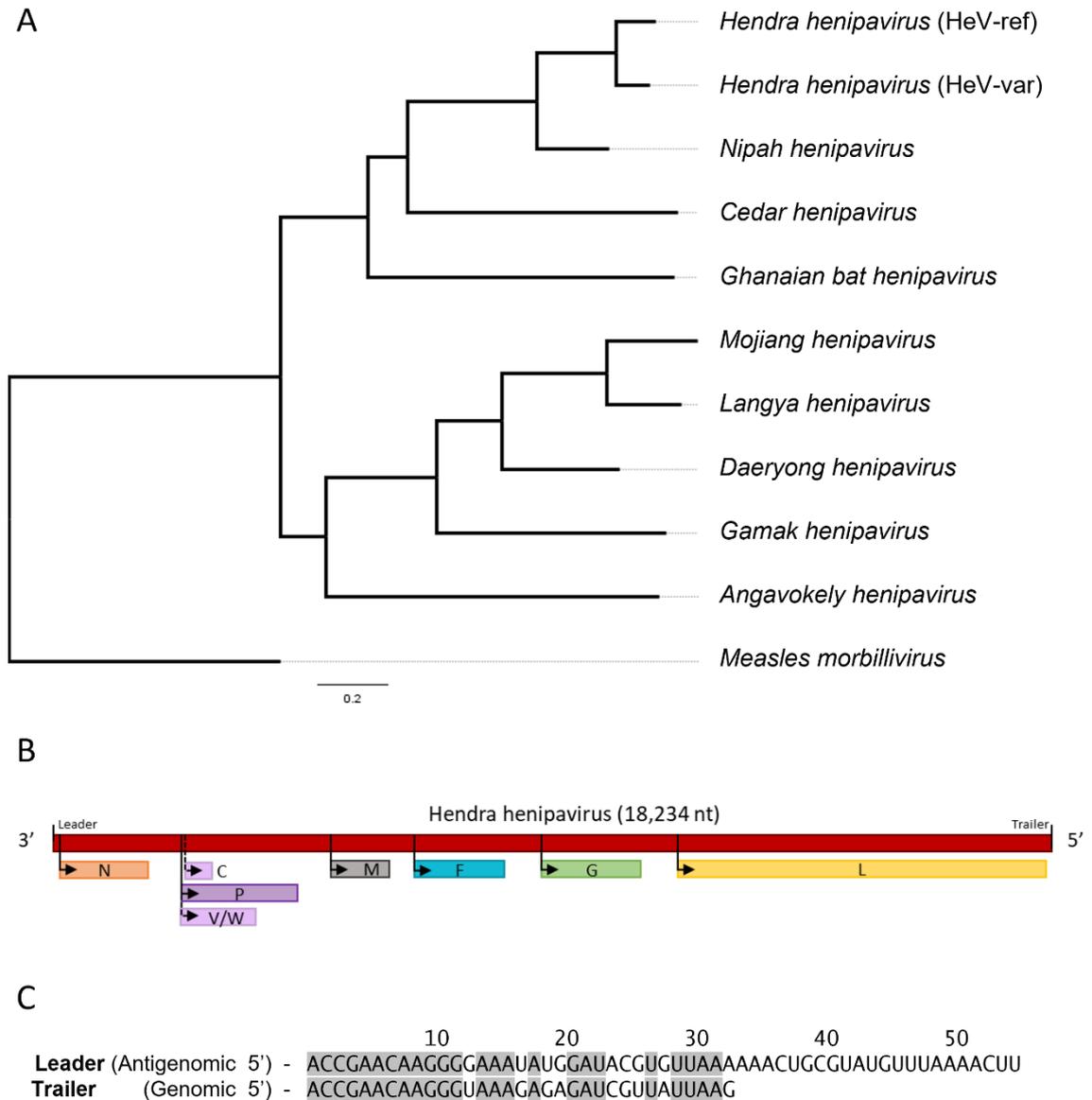


Figure 1.5. *Henipavirus* clade and *Hendra henipavirus* genome organisation.

A) Phylogenetic tree based on genomic sequences of all species in the genus *Henipavirus* as well as *Measles morbillivirus* (NCBI RefSeq: NC_001498.1) as an outgroup. The tree illustrates the genetic relationship between members of the *Henipavirus* genus and was visualised using FigTree following alignment of complete genome sequences in MAFFT and maximum likelihood tree construction in IQ-TREE. Genome accession numbers; HeV-ref (NCBI RefSeq: NC_001906.3), HeV-var (GenBank: MZ318101.1), *Nipah henipavirus* (NCBI RefSeq: NC_002728.1), *Cedar henipavirus* (GenBank: JQ001776.1), *Ghanaian bat henipavirus* (GenBank: HQ660129.1), *Mojiang henipavirus* (GenBank: KF278639.1), *Langya henipavirus* (isolate SDQD_S1801, GenBank: OM101130.1), *Daeryong henipavirus* (strain Cs17-65, GenBank: MZ574409.1), *Gamak henipavirus* (strain CI17-32, GenBank: MZ574407.1), *Angavokely henipavirus* (GenBank: ON613535.1). **B)** Schematic representation of the HeV genome (red) with the six main genes: N nucleoprotein (orange), P phosphoprotein (purple), M matrix protein (grey), F fusion glycoprotein (cyan), G attachment glycoprotein (green), and the L RNA-dependent RNA polymerase (yellow). The V, W, and C accessory proteins are alternative ORFs within the P gene and are indicated in pink. ORF location and lengths are to scale. **C)** Alignment of the HeV genome end sequences, showing the full length 3' leader (antigenomic 5') and 5' trailer (genomic 5'). Complementary nucleotides are highlighted in grey.

through alternative ORFs or mRNA editing (see Section 1.7.2). Thus, HeV conforms to the paramyxoviral 'rule of six', which stipulates that the genome length is a whole multiple of six nucleotides – 3039 repeats of six – and encodes six genes (Calain and Roux, 1993; Kolakofsky et al., 1998; Wang et al., 2000). Genomes that experimentally deviate from the 'rule of six' with regards to genome length replicate markedly less efficiently (Calain and Roux, 1993; Halpin et al., 2004).

1.6.2 Hendra henipavirus virion

HeV virions are generally presented as pleiomorphic in shape, from roughly spherical to filamentous, and range from 40 nm to 600 nm in size (Hyatt et al., 2001). This large range may be affected by experimental conditions and reflect the easily lost native filamentous shape under EM conditions as observed for the related RSV virion (Conley et al., 2022). The viral lipid envelope, which is derived from host membrane, contains the F and G glycoproteins and exhibits a characteristic double fringe when observed by EM, setting it apart from the single fringe observed in the closely related NiV (Figure 1.6) (Hyatt et al., 2001). The two fringes extend approximately 8 nm and 15 nm from the lipid bilayer, while the NiV single fringe extends approximately 17 nm (Hyatt et al., 2001; Eaton et al., 2006). The F and G glycoproteins are type I and type II membrane protein, respectively. For NiV, the G glycoprotein assembles into a tetrameric complex in which the monomers adopt one of three distinct folding patterns (Wang et al., 2022). The HeV F glycoprotein is organised in trimeric complexes in its prefusion form and retains a trimeric association throughout the maturation process (Smith et al., 2013; Wong et al., 2016) (see Section 1.7.4). The lipid bilayer is lined on the inner leaflet by the M protein. The M proteins interact with the transmembrane domains of the F and G glycoproteins in the lipid bilayer and the N proteins of the nucleocapsid in the virion lumen (Plempner and Lamb, 2021). Recent work has elucidated the virion ultrastructure of MeV and RSV through cryo-ET, both of which are related to the henipaviruses (Ke et al., 2018; Conley et al., 2022). While much of the morphology on HeV and NiV is gained through EM on isolated virions, in this case RSV was cultured directly on EM grids (see Chapter 4 Section 4.1.4) to maximise the preservation of the fragile virions. The micrographs and subsequent cryo-ET reconstruction revealed largely filamentous virions, with a clear role for M protein-driven helical ordering of the M protein lattice as well as the surface glycoproteins F and G (Conley et al., 2022). Furthermore, in native virions the MeV nucleocapsids appear to interact closely with the M protein layer and the nucleocapsid's helical pitch is in line with the periodicity of the M protein lattice (Ke et al., 2018).

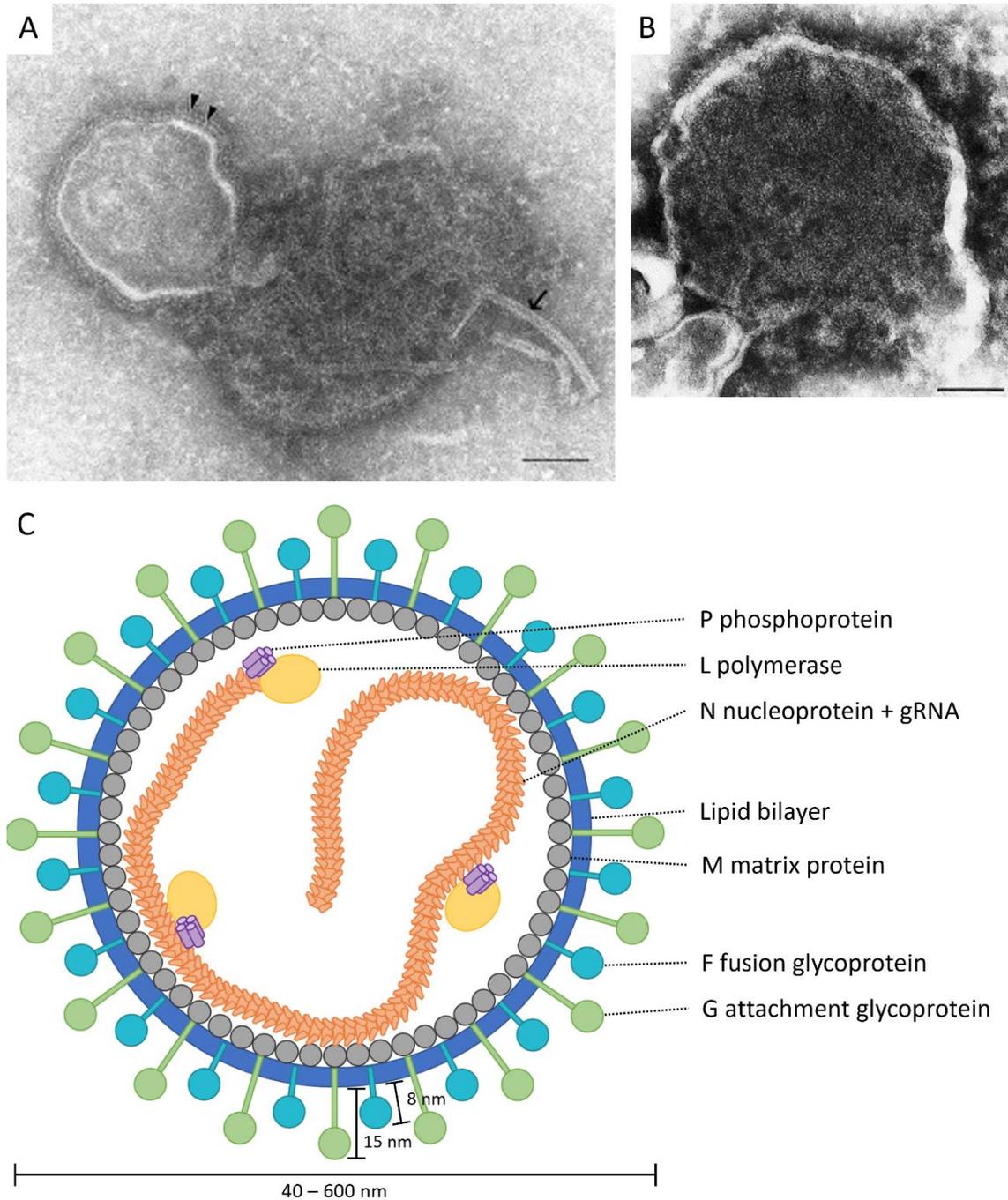


Figure 1.6. *Hendra henipavirus* ultrastructure.

A-B) Electron micrographs of negatively stained HeV (**A**) and NiV (**B**). Scale bars are 100 nm. Adapted from (Hyatt et al., 2001). **A)** The HeV virion is ruptured and nucleocapsid has spilled out (arrow). The transmembrane G attachment and F fusion glycoproteins create a characteristic double-fringed morphology (arrow heads). **C)** A schematic representation of the HeV virion, following the colour code in Figure 1.4B and based on (Eaton et al., 2006; B. Lee et al., 2021). The N nucleoprotein (orange) encapsidates the viral genomic RNA (gRNA). Interaction between the L RNA-dependent RNA polymerase (yellow) and the N nucleoproteins is facilitated by the P phosphoprotein (purple). The viral envelope is made up of a host-derived lipid bilayer (blue) which is lined on the inside by M matrix protein (grey). The transmembrane G attachment (green) and F fusion (cyan) glycoproteins form the double-fringed morphology observable through transmission electron microscopy.

These results further stress the central role for M protein in virion organisation and indicate that the often irregular morphology observed in other mononegaviruses may be artefactual, owing to the purification process while the *in situ* virions exhibiting a filamentous, helical ordering that may represent the native morphology (Conley et al., 2022). Indeed, highly ordered virion morphologies with a central role for the M protein are a hallmark of the more distantly related family *Rhabdoviridae* (Jenni et al., 2022; Si et al., 2022), supporting the notion that all members of the order *Mononegavirales* could maintain highly ordered virions as the native state. The viral RNA genome lies completely encapsidated in N proteins to form a helical nucleocapsid complex, which is the template for mRNA transcription and genomic RNA replication and the form in which the viral genomic RNA is assembled into nascent virions. The nucleocapsid combined with the P protein and L polymerase constitutes the viral replication unit, also referred to as the holonucleocapsid, which is thought to be the minimum unit of infectivity and is capable of transcribing mRNA *in vitro* (Plempner and Lamb, 2021). SeV virions isolated from cell culture using ³⁵S Methionine labelling revealed that each holonucleocapsid contains approximately 2600 N monomers, 300 P monomers, and 40 L monomers (Lamb et al., 1976). Following the rule of six, the HeV genome is encapsidated by 3039 N monomers and it is likely that the number of P and L monomers would fall within the same orders of magnitude as for SeV (Lamb et al., 1976; Plempner and Lamb, 2021). Very little structural data is available for proteins of the HeV nucleocapsid, and an N protein structure is completely absent from the literature.

1.7 Hendra henipavirus life cycle

1.7.1 Viral tropism and entry

Host range and cell tropism is largely dictated by the F and G proteins through their roles in binding to cellular receptors and fusing the viral and host-cell membranes, respectively. While most paramyxoviruses have relatively limited host ranges, members of the genus *Henipavirus* are known to naturally infect a broad range of hosts, including cats, dogs, flying foxes, horses, pigs, and humans (Figure 1.4C) (Eaton et al., 2006). Experimental infection has extended this range to guinea pigs, hamsters, and non-human primates, as well as several cultured cell types, including Vero, MDBK, RK13, BHK21, and primary foetal equine kidney cells (Murray et al., 1995; Eaton et al., 2006; B. Lee et al., 2021). The broad species tropism is further supported by *in vitro* cell fusion studies. The F and G proteins of both HeV and NiV facilitated fusion with adjacent cells from a range of animal species, including rabbits, non-human primates, and mice, indicating that the cellular receptor is widely expressed (Bossart et al., 2002). Interestingly, the

glycoproteins of HeV and NiV also showed heterotypic functional activity between the two, but not with the relevant glycoproteins from other paramyxoviruses. Many paramyxoviral attachment glycoproteins exhibit haemagglutination and neuraminidase activities and use the ubiquitous N-acetyl neuraminic acids for cell attachment. In contrast, henipaviruses lack these activities suggesting the use of cell-surface proteins as receptors (Eaton et al., 2006; Navaratnarajah et al., 2020).

Through mass spectrometry and mRNA analysis, the cell-surface glycoprotein ephrin B2 was identified as the receptor used by HeV and NiV (Negrete et al., 2005; Bonaparte et al., 2005). Additionally, ephrin B3 was identified as an alternative henipaviral receptor (Negrete et al., 2006). Ephrin B2 is widespread amongst vertebrate species and is predominantly expressed in arterial and arteriolar endothelial cells and capillaries and a wide range of organ tissues and cell-types including neurons, smooth muscle cells, and human bronchiolar epithelia. In contrast, ephrin B3 is restricted to neuronal tissues in the brain (B. Lee et al., 2021). The distribution of these highly conserved receptors supports the observed broad species tropism and explains – to a degree – the pulmonary and neurological symptoms associated with HeV infection (Selvey et al., 1995; Eaton et al., 2006; B. Lee et al., 2021). Following G protein mediated attachment to ephrin B2 expressing permissive host-cells, the F protein drives fusion of the viral and host-cell membranes (Figure 1.7). Unlike some viruses which require the high concentrations of ions or acidic pH of the endocytic pathway (Charlton et al., 2020), HeV fusion, and indeed that of other paramyxoviruses, occurs at neutral pH on the cell surface delivering the viral genome directly into the cytosol (Plempner and Lamb, 2021). The ability to fuse at the cell-surface and at neutral pH explains the formation of syncytia, a cytopathic effect characteristic for members of the family *Paramyxoviridae* (Plempner and Lamb, 2021). The mechanism of paramyxoviral uncoating post-fusion, which entails disruption of the M-N protein interactions, remains unclear.

1.7.2 Viral RNA transcription and replication

After entry into the cell, all processes of viral replication and transcription and translation of viral proteins take place within the cytosol (Figure 1.7). While HeV-specific knowledge is scarce, the transcription and replication strategies are conserved among paramyxoviruses and generalisations can be drawn from related viruses. The viral growth cycle, which encompasses all steps necessary for the generation of progeny virions, takes place in three major phases: a primary transcription phase, a replication phase, and a secondary transcription phase. These phases ensure the transcription of two kinds of RNA from the same

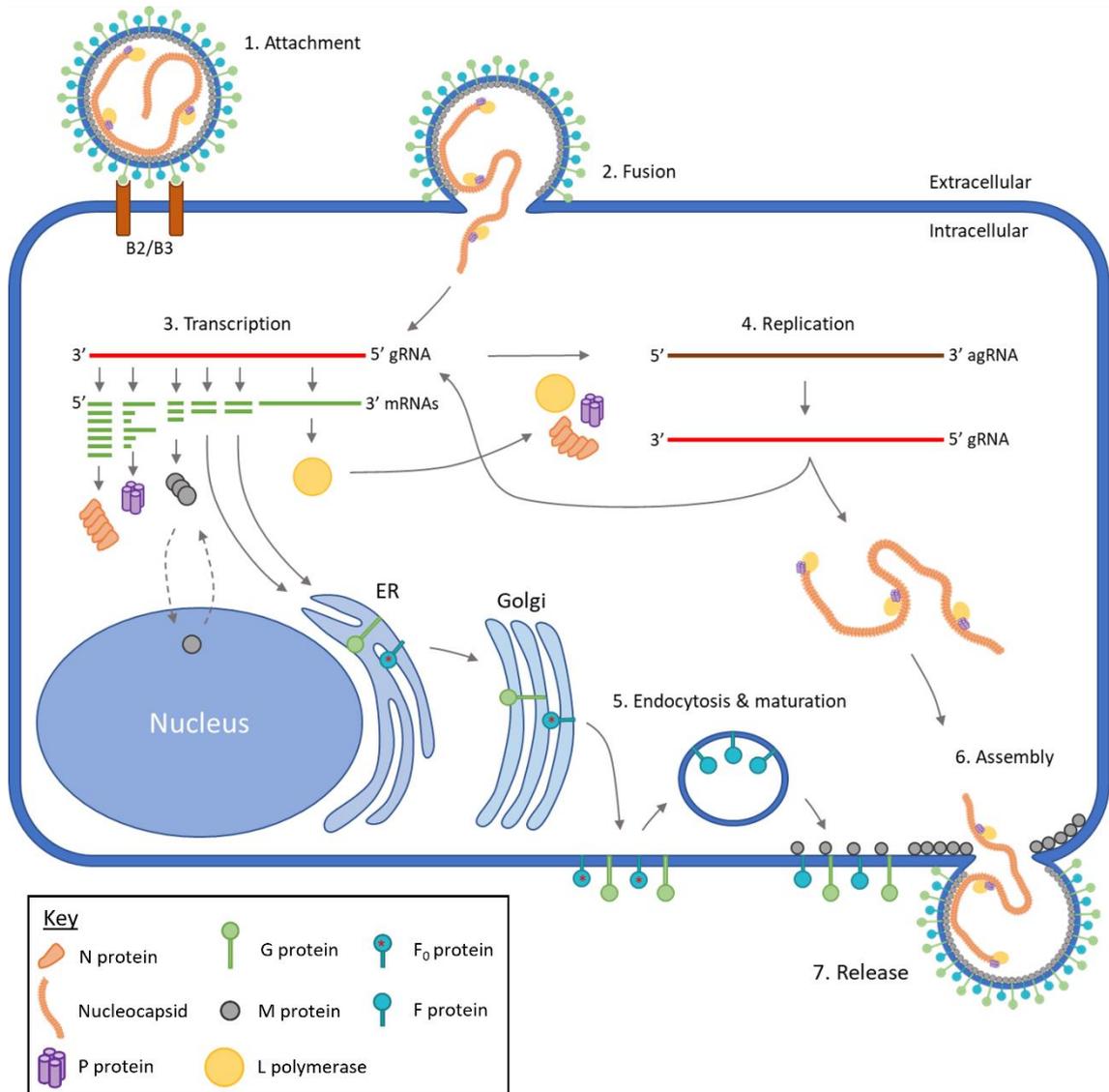


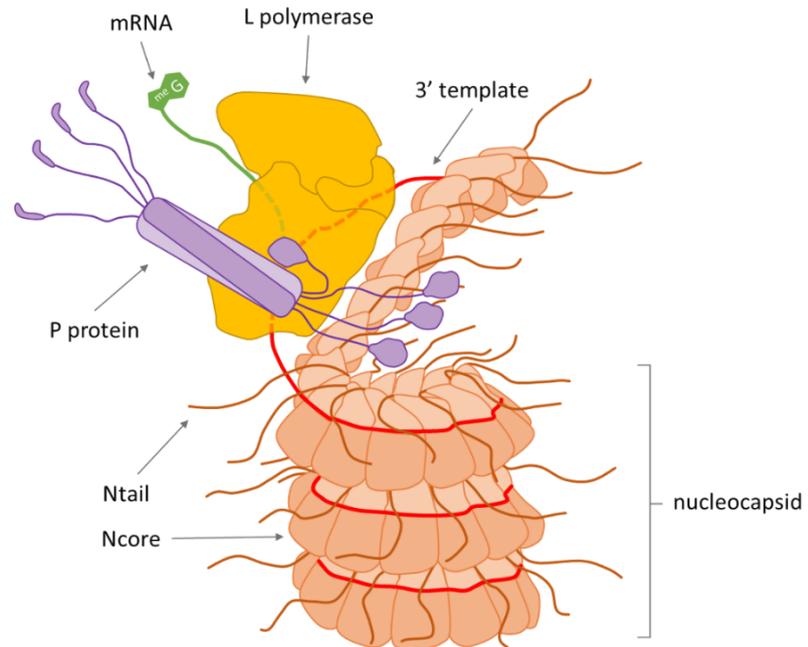
Figure 1.7. The henipavirus life cycle.

Following entry into a host, HeV virions attach to susceptible cells through interactions between the viral attachment glycoprotein (G protein; green) and the cellular Ephrin B2/B3 (1). Membrane fusion occurs on the cell surface at neutral pH and is driven by the fusion glycoprotein (F protein; cyan), releasing the holonucleocapsid into the cytosol (2). The RNA dependent RNA polymerase (RdRp), consisting of the phosphoprotein (P protein; purple) and the viral polymerase (L polymerase; yellow) engage the encapsidated genome (nucleocapsid; orange) at the 3' end. Transcription of the RNA genome template (gRNA; red) results in the generation of viral mRNAs (dark green) (3). Viral structural proteins including the nucleoprotein (N protein; orange), P, matrix protein (M protein; grey), and L as well as non-structural proteins V, W, and C are expressed in the cytosol. The transmembrane proteins F₀ and G are expressed in the endoplasmic reticulum (ER). Once cytosolic levels of N, P, and L are sufficiently high, replication of the genome can occur. The genomic nucleocapsid serves as a template for antigenome (agRNA; burgundy) synthesis. In turn, the antigenomic nucleocapsid serves as a template for progeny genome synthesis (4). The M protein performs a nuclear transit before associating with the cell membrane at the assembly site. After trafficking through the trans-Golgi network (TGN), transmembrane proteins F₀ and G localise to the cell surface. Subsequently, F₀ is endocytosed and matures to F through proteolytic cleavage by cathepsin L/B. Mature F is then recycled to the cell surface (5). Progeny virion assembly is driven by M, which interacts with the cytosolic tails of F and G in the lipid membrane and with N within the holonucleocapsid (6). M protein self-organises and drives virion budding which is mediated by host factors such as VSP4 and ESCRT to release progeny virions (7).

viral nucleocapsid template, namely non-encapsidated mRNA and encapsidated viral genomic RNA (gRNA) and antigenomic RNA (agRNA). The shift between the RdRp transcriptase (primary and secondary) and replicase modes is dependent on a number of factors. Paramyxoviruses must control the amount and type of viral RNA that is transcribed during infection, which is modulated through *cis*-acting RNA sequences and *trans*-acting viral proteins. The interplay between the *cis*-acting sequences and the RdRp – which consists of a single L protein and a P homotetramer – is particularly complex since they are only recognised within the context of the nucleocapsid (Plempner and Lamb, 2021). The paramyxoviral RdRp gains access to the viral RNA via a single entry site through the bipartite paramyxoviral promoter at the genomic and antigenomic 3' termini. The first promoter element (PrE-I) sits at the 3' RNA terminus and the second promoter element (PrE-II) follows downstream of PrE-I, falling within the N gene for the genomic promoter or the L gene for the antigenomic promoter. While the two promoter elements are distant in sequence space, they are close together in physical space as they appear on the same face of the helical nucleocapsid separated by one helical turn (Bloyet, 2021; Plempner and Lamb, 2021). Following the 'three in three out' base stacking observed in N protein structures and the helical register of the nucleocapsid, the six bases of PrE-I are fully solvent exposed and can be engaged by the RdRp, since there is no N_{i+1} neighbour to cover bases 1-3 and no preceding helical turn exists to cover the solvent exposed bases 4-6. Disruption of the nucleocapsid helical stacking at the 3' extreme then also reveals the three solvent facing bases of PrE-II which lie one helical turn directly downstream of PrE-I (Bloyet, 2021; Plempner and Lamb, 2021). The genomic promoter is responsible for the generation of both mRNA and antigenomic RNA, while the antigenomic promoter only drives the generation of genomic RNA (Plempner and Lamb, 2021).

Since the genomic and antigenomic RNAs are fully encapsidated within the nucleocapsid, promoter binding by the RdRp complex must be followed by displacement of the N proteins and threading of the exposed RNA strand through the RdRp (Figure 1.8). As the first N protomer (N_1) in the nucleocapsid has no N_{i+1} binding partner, the N_1 CTD groove that would normally be occupied by the N_{i+1} Nt-arm remains vacant. During engagement of the RdRp with the first bases of PrE-I, the P protein Nt $\alpha 1$ and $\alpha 2$ may occupy this groove to generate a pseudo P- N^0 complex and aid in the release of RNA from the first N protomer by inducing conformational changes in N_1 towards the open N^0 state (Bloyet, 2021). It is thought that RNA release from the nucleocapsid occurs only locally near the RdRp complex and that the RdRp moves through the nucleocapsid in a zipper-like fashion, reminiscent of the replication bubble seen in DNA replication (Plempner and Lamb, 2021). Upon interaction with the RdRp complex, local N

A



B

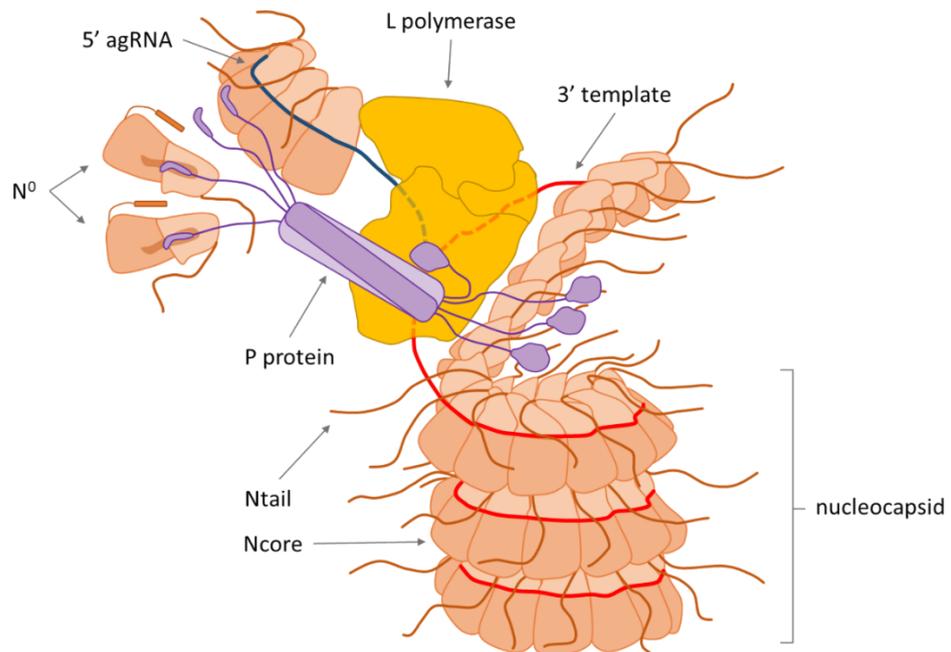


Figure 1.8. A model of the paramyxoviral RNA-dependent RNA polymerase activity on a nucleocapsid template.

The nucleocapsid, consisting of viral RNA (red) and N protein (orange), serves as the template for mRNA transcription (**A**) or antigenomic RNA (agRNA) replication (**B**) by the RdRp. The RdRp consists of the L polymerase (yellow) and a heterotetrameric P protein (purple). The P protein tetramer contains binding sites for the N protein and the L polymerase. The RdRp associates with the template 3' end and routes the N protein chain over its surface after local release of the RNA. The template RNA is re-encapsidated upon exiting the polymerase. **A**) The nascent, capped mRNA (green) exits the polymerase and remains unencapsidated. **B**) The P protein N-terminal helix interacts with the N⁰ protein to direct them for encapsidation of the nascent agRNA (navy). Adapted from (Plempner and Lamb, 2021).

protomers are thought to undergo a conformational change to release the shielded RNA towards the RdRp. The exact molecular mechanism remains obscured, however it is thought that the RdRp induces a conformational change in an N protein NTD alpha helix $\alpha 7$ and the preceding loop to release the RNA (Bloyet, 2021). In this model, the RNA-free N protomers maintain N-N interactions and are routed over the RdRp surface while the RNA is threaded through the RdRp template groove, followed by re-encapsidation as the RNA template emerges from the RdRp complex (Plempner and Lamb, 2021; Bloyet, 2021).

1.7.2.1 Messenger RNA synthesis and editing

Following initiation of RNA synthesis at the genomic 3' promoter, the RdRp complex moves towards the 5' end as it generates a transcript of the genomic leader sequence. The availability of unassembled N (N⁰)-P complexes critically dictates the mode of RdRp synthesis, which is either transcriptase mode or replicase mode. In the absence of abundant N⁰-P complexes, the leader is left unencapsidated and is released from the RdRp. In this case, the RdRp enters transcriptase mode. The RdRp then reinitiates at the N gene GS and processively interacts with *cis*-acting elements at the gene junction to produce a capped and polyadenylated mRNA. This process of termination and reinitiation of RNA synthesis at the next GS continues in a 5' direction along the template. The gene junctions consist of the upstream gene end (GE), the intergenic (IG) region, and the downstream gene start (GS). The GE contains a transcription termination signal and a poly-uridine (U tract) that serves as the polyadenylation template through RdRp stuttering. Following transcription termination, the RdRp is thought to remain associated with the template as it crosses the non-transcribed IG region. Transcription reinitiation is directed by the GS site, which also induces the addition of a methylated 5' guanine (^{me}G) cap to the nascent mRNA molecule (Figure 1.8A). RdRp transcription initiation after transcription termination is imperfect, leading to a portion of RdRp complexes that fail to reinitiate and resulting in a gradient in mRNA abundance that decreases in lockstep with the distance from the 3' end. Consequently, mRNA abundance is generally highest for the N gene and lowest for the L gene (reviewed in (Plempner and Lamb, 2021)).

Five of the six genes contain only single ORFs, however the henipaviral P gene encodes multiple proteins through alternative translation-initiation sites, overlapping reading frames, and mRNA editing. The P, V, and W ORFs share an identical N-terminal domain but have unique C-terminal domains. During mRNA synthesis, one or two pseudo-templated G nucleotides are inserted at a conserved editing site to generate transcripts for V and W, respectively (Figure 1.9).

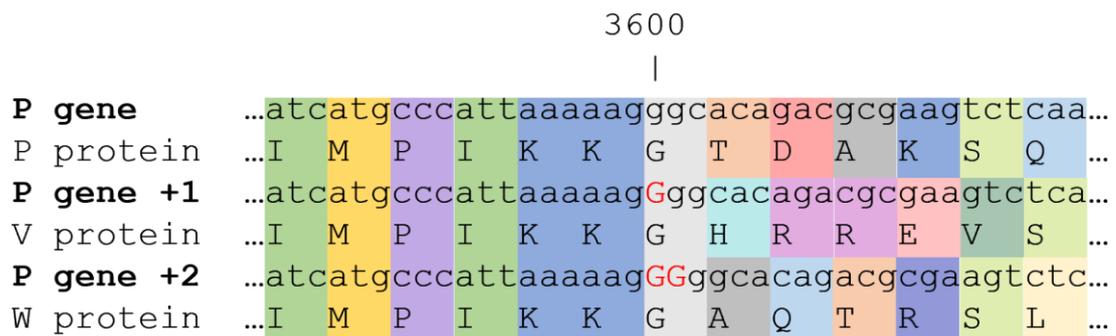


Figure 1.9. Henipaviral P gene polycistron.

The henipaviral P gene encoded P, V, and W ORFs share an identical N-terminal domain with unique C-terminal domains. One (+1) or two (+2) pseudo-templated G nucleotides are inserted (red) at a conserved editing site (nt 3600 in the HeV genome) to generate transcripts for V and W, respectively.

The insertion of the G nucleotides occurs by a co-transcriptional stuttering mechanism reminiscent of the polyadenylation of mRNAs and is dependent on relative position of the editing site within the N-bound hexamer of nucleotides. The inserted nucleotides result in a reading frame shift leading to the expression of N-terminally identical but C-terminally unique proteins (Flint et al., 2009b; Plemper and Lamb, 2021; B. Lee et al., 2021). The C protein, also encoded within the P gene, is expressed using an alternative translation initiation site from the P, V, and W mRNAs. As the full C ORF overlaps only the reading frame upstream of the mRNA editing site, no editing occurs on C ORF mRNAs (Eaton et al., 2006; Plemper and Lamb, 2021; B. Lee et al., 2021). The P protein is essential for viral RNA synthesis and contains multiple binding sites for the N protein and L polymerase, highlighting its role in bridging the N protein and L polymerase (Plemper and Lamb, 2021). In fact, a section of the P protein N-terminal domain was co-crystallised with the NiV Ncore domain (Yabukarski et al., 2014).

Early during infection, in the primary transcription phase, the absence of high-levels of viral translation products restricts the RdRp to the transcription of leader RNA and mRNAs from the incoming nucleocapsid rather than the replication of the genome. Later during infection, in a secondary transcription phase, additional mRNAs can be transcribed from progeny viral genome (see below) (Figure 1.7).

1.7.2.2 Viral genome replication and nucleocapsid assembly

Since the N protein must fully encapsidate the viral genome, its abundance is a critical limiting factor in genome replication. The transcription and translation of mRNAs leads to the

accumulation of viral proteins, which is essential for the production of the antigenome. In the presence of abundant N⁰-P complexes, the genomic leader sequence is encapsidated and the RdRp enters replicase mode to generate the encapsidated antigenomic RNA (agRNA) copy (Figure 1.7). This full length, complementary copy of the viral genome does not encode any known functional mRNAs and only serves as an intermediary in genome replication (Plempner and Lamb, 2021). A function for SeV and RSV trailer RNA, released during abortive agRNA synthesis, has been proposed in antagonising stress granule formation and cellular apoptosis, however it is uncertain if this function is widespread among other paramyxoviruses (Iseni et al., 2002; Hanley et al., 2010). Once the synthesis of antigenomic RNA is initiated, viral protein expression does not halt. Indeed, since genome synthesis and encapsidation occur concomitantly, the stores of accumulated N⁰ protein become depleted during genome replication and ongoing protein synthesis is required for further genome encapsidation. This coupling of genome assembly and synthesis leads to a self-regulatory system that modulates between mRNA transcription and genome replication. As the leader sequence contains the N encapsidation signal, termination and reinitiation at the leader-N junction (see above) must occur so that the N gene mRNA is not encapsidated and becomes untranslatable. When N⁰ protein levels are limited, such as early in the growth cycle, the RdRp preferentially transcribes mRNA. These mRNAs are translated and increase the cytosolic levels of viral protein, including N⁰ protein. When the level of P-N⁰ complex is sufficient, some RdRps switch from mRNA transcription to antigenome synthesis. In this case, encapsidation starts at the leader sequence and the leader sequence is retained in the antigenomic copy. The P protein maintains N protomers in a monomeric and open conformation, referred to as unassembled N (N⁰). Upon interacting with the Ncore of newly expressed N proteins, the P protein NTD forms a number of variable length alpha-helices (α 1-4) (Figure 1.10). In this interaction, α 1 binds a hydrophobic groove on the Ncore CTD which would normally be occupied by the Nt-arm of the adjacent N_{i+1} protomer within the nucleocapsid assembly (Bloyet, 2021). For MeV and NiV, an additional P protein Nt alpha-helix (α 2) directly follows the first one and bind the top of the Ncore CTD which is normally occupied by the Ct-arm of the N_{i-1} N protomer within the nucleocapsid assembly (Figure 1.10) (Yabukarski et al., 2014; Guryanov et al., 2016). Finally, a short, transient alpha-helix forms downstream of the phosphoprotein Nt (α 4) and together with the short, disordered peptide δ interacts with the bottom of the N protein NTD. The α 4 sequence maintains some level of conservation among members of the genera *Morbillivirus* and *Henipavirus* (Bloyet, 2021). These P protein alpha helices are linked through disordered sections and together envelop the N protomer (Figure 1.10B). While the level of N⁰-P is an important regulatory factor, it may not be the only one. The activities of the viral accessory proteins further regulate genome

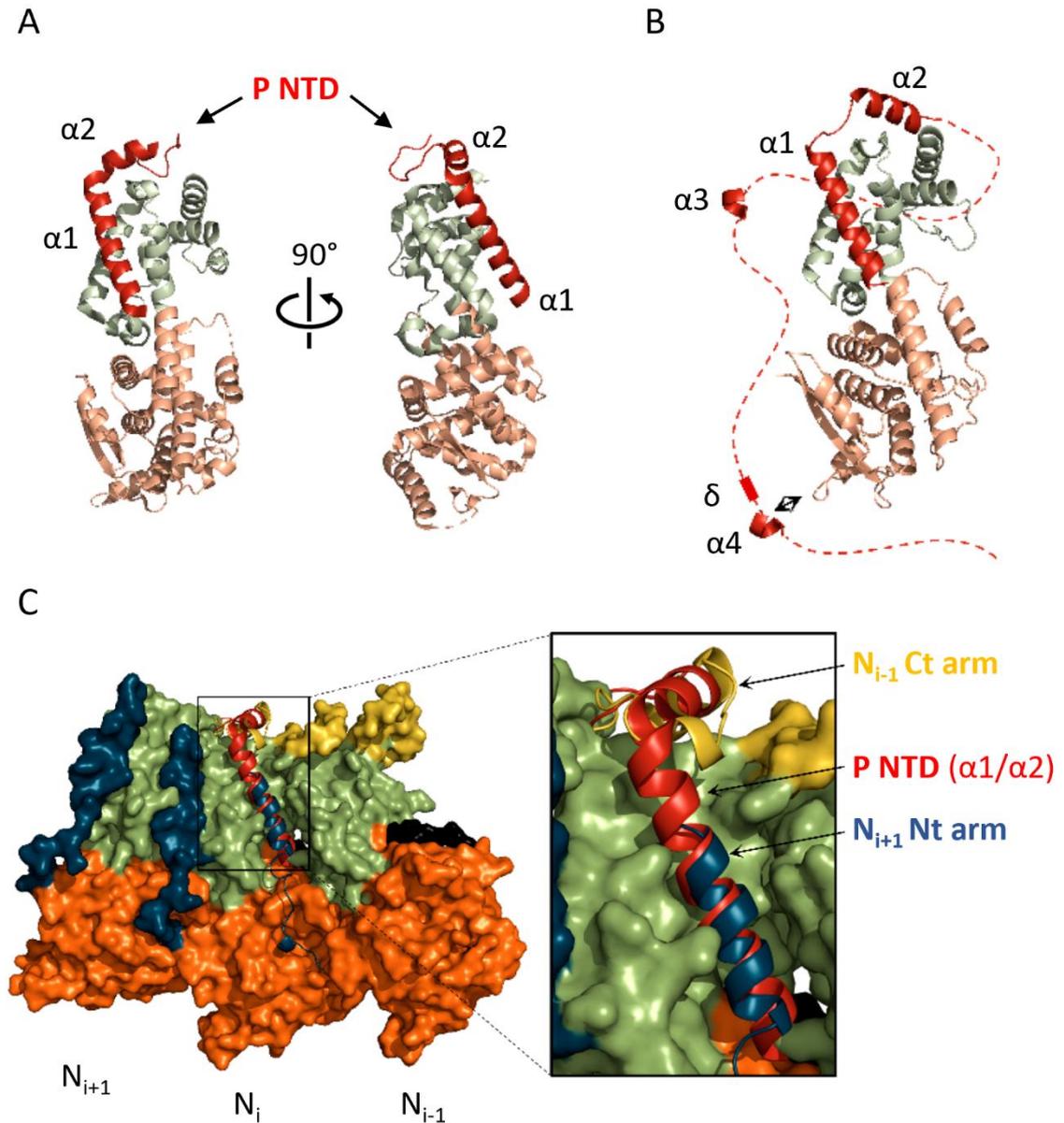


Figure 1.10. Interactions of the P- N^0 complex.

A) An Ncore domain with the NTD (orange) and CTD (green) depicting the binding of the P protein N-terminal domain alpha helices $\alpha 1$ and $\alpha 2$ (red). The Nt and Ct flexible arms are not shown. **B)** An Ncore domain with the NTD (orange) and CTD (green) and the N-terminal disordered domain of the P protein (red) illustrating the conformation of the P- N^0 complex. The transient alpha helices $\alpha 1-4$ and the disordered peptide δ are indicated. Helices $\alpha 1$ and $\alpha 2$ interact with the N protein CTD while $\alpha 4$ and δ bind the N protein NTD. In this conformation the otherwise disordered P protein NTD envelops the N protomer. The N protein Nt and Ct flexible arms are not shown. **C)** A schematic depicting three N protomers in the context of a nucleocapsid with the P protein NTD ($\alpha 1$ and $\alpha 2$; red) superimposed. The P NTD $\alpha 1$ sits in a groove on the inner face of the N_i protomer CTD that is normally occupied by the N_{i+1} Nt arm (blue). The P NTD $\alpha 2$ wraps over the top of the N CTD where the N_{i-1} Ct arm (yellow) sits in the context of a nucleocapsid. Adapted from (Bloyet, 2021).

replication. For example, SeV V protein may inhibit genome replication through binding to N⁰ protein (Horikami et al., 1996; Plemper and Lamb, 2021).

Following its synthesis, the encapsidated antigenome serves as a template for the synthesis of progeny genomes in a process similar to that of antigenome synthesis (Figure 1.7). The RdRp binds the bipartite promoter at the antigenomic 3' end which directs the generation of the trailer (or negative leader) RNA (Plemper and Lamb, 2021). If sufficient N⁰ protein and other viral proteins are available, genome synthesis begins by the encapsidation of the nascent trailer sequence. The final product is an encapsidated negative-sense viral genome. Progeny genomes can then serve three functions. Firstly, they may be a template for further mRNA synthesis in the secondary transcription phase, generating additional mRNA and viral proteins as described above. Secondly, they may be a template for the generation of additional antigenomes. Thirdly, they may associate with M proteins and be incorporated into the budding progeny virions (Figure 1.7) (Plemper and Lamb, 2021).

The encapsidation of the viral gRNA with N protein is thought to occur co-transcriptionally. Specific interactions between the N⁰-P complex and the L protein guide encapsidation. N protomers are assembled onto the 5' end of the nascent antigenomic or genomic RNA (positive leader or negative trailer, respectively) as it emerges from the RdRp. Encapsidation then proceeds in a 5' to 3' direction on the nascent chains following replication (Figure 1.8B) (Blumberg et al., 1981; Plemper and Lamb, 2021). While this regulated process ensures that only viral RNA is efficiently encapsidated, transient expression of N protein in the absence of other viral components leads to spontaneous and random encapsidation of host RNA into nucleocapsid-like structures (Plemper and Lamb, 2021). Each N protomer associates with exactly six RNA nucleotides to fully encapsidate the antigenome or genome, adhering to the rule of six and imprinting a hexameric phase onto the viral RNA. The strict adherence to the rule of six was initially proposed to reflect the requirement for the nucleocapsid template to have no free, or 'dangling', terminal bases (Plemper and Lamb, 2021). While efficient replication is indeed inhibited for many viral genomes when experimentally altered to break the rule of six, this is not universally true for all members of the family *Paramyxoviridae* (Vulliémoz and Roux, 2001; Plemper and Lamb, 2021). Instead, an alternative hypothesis posits that changes to the genome length away from the rule of six alter the phase of particular nucleotides in the context of an N protomer. In this case, the rule of six reflects the requirement of critical *cis*-acting promoter sequences to be in the correct phase and they can only be recognised while in the correct orientation within their encapsidating N protomers and within the helical nucleocapsid (Kolakofsky et al., 1998; Plemper and Lamb, 2021).

While the genomic promoter is stronger than the antigenomic promoter, genomic nucleocapsids are found at a higher relative abundance than antigenomic nucleocapsids. It is thought that the C protein interacts with the L protein and modulates polymerase activity at the promoter. This modulation generates a shift during infection from antigenome synthesis to genome synthesis (Plemper and Lamb, 2021). Reducing the levels of antigenomic copies may be important in limiting the generation of antigenome-containing progeny virions as well as stemming the depletion of N⁰-P complexes.

1.7.3 Virus-host immune interactions

Intracellular dsRNA and 5' triphosphates, hallmarks of viral infection, are sensed by protein kinase R (PKR), toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I) and RNA helicase melanoma differentiation-associated gene 5 (MDA5) (Figure 1.11). PKR activation leads, through phosphorylation of the eukaryotic translation initiation factor 2 alpha (eIF2 α), to a halt in mRNA translation known as the 'host shut-off'. PKR also mediates the downstream activation of nuclear factor κ B (NF- κ B). MDA5 and RIG-I are the first steps in a signalling pathway that activates transcription factors interferon (IFN)-regulatory factor (IRF) 3, IRF7, and NF- κ B. Once activated, IRF3 and NF- κ B migrate to the nucleus where they induce the transcription of IFN stimulated genes (ISGs), which are unaffected by the host shut-off. The cellular Jak/STAT (Janus tyrosine kinase/signal transducers and activators of transcription) pathway is activated following binding of extracellular IFN to the IFN- α/β receptors (IFNAR) 1 and 2. In turn, activation of STAT through phosphorylation into phosphoSTAT (pSTAT) precedes translocation to the nucleus where it stimulates the expression of ISGs (Figure 1.11) (Plemper and Lamb, 2021; B. Lee et al., 2021).

Proteins encoded within the henipaviral P gene modulate virulence by disrupting the cellular innate immune response. PKR is not extensively activated during infection for many paramyxoviruses. This may be in part due to the continuous encapsidation of the genome and antigenome, minimising the amount of dsRNA in the cytosol, but there is also evidence that the activation of PKR is prevented indirectly by the P gene products through their activities as modulators of RNA synthesis (Gainey et al., 2008; Boonyaratanakornkit et al., 2011; Plemper and Lamb, 2021; B. Lee et al., 2021). The use modified *Mammalian orthorubulavirus 5* (formerly *Simian virus 5*) deficient in the P and V proteins showed a marked activation of PKR and translational shutdown in cell culture and a reduction in viral mRNA levels, while reintroduction of functional P and V protein or the expression of PKR antagonists restored viral mRNA synthesis

and protein expression (Gainey et al., 2008). Furthermore, infection of cell cultures with *Human respirovirus 1* deficient in the C protein led to an increase in IRF3 and NF- κ B activation compared to wild-type virus (Boonyaratanakornkit et al., 2011). IFN-mediated responses are also disrupted by P gene products (Figure 1.11). The henipaviral V protein antagonises MDA5 and RIG-I in the cytosol through direct interactions. It is thought to compete for dsRNA binding with MDA5 and inhibit MDA5 multimerisation to its active form. In doing so, the V protein interferes with the downstream signalling and transcription of ISGs (Andrejeva et al., 2004; Shaw et al., 2005; Plemper and Lamb, 2021; B. Lee et al., 2021). The henipaviral W protein blocks TLR3 signalling and this ability is dependent on nuclear localisation of the W protein (Shaw et al., 2005). Furthermore, activated pSTAT1, pSTAT2, and IRF9 are sequestered into high molecular weight

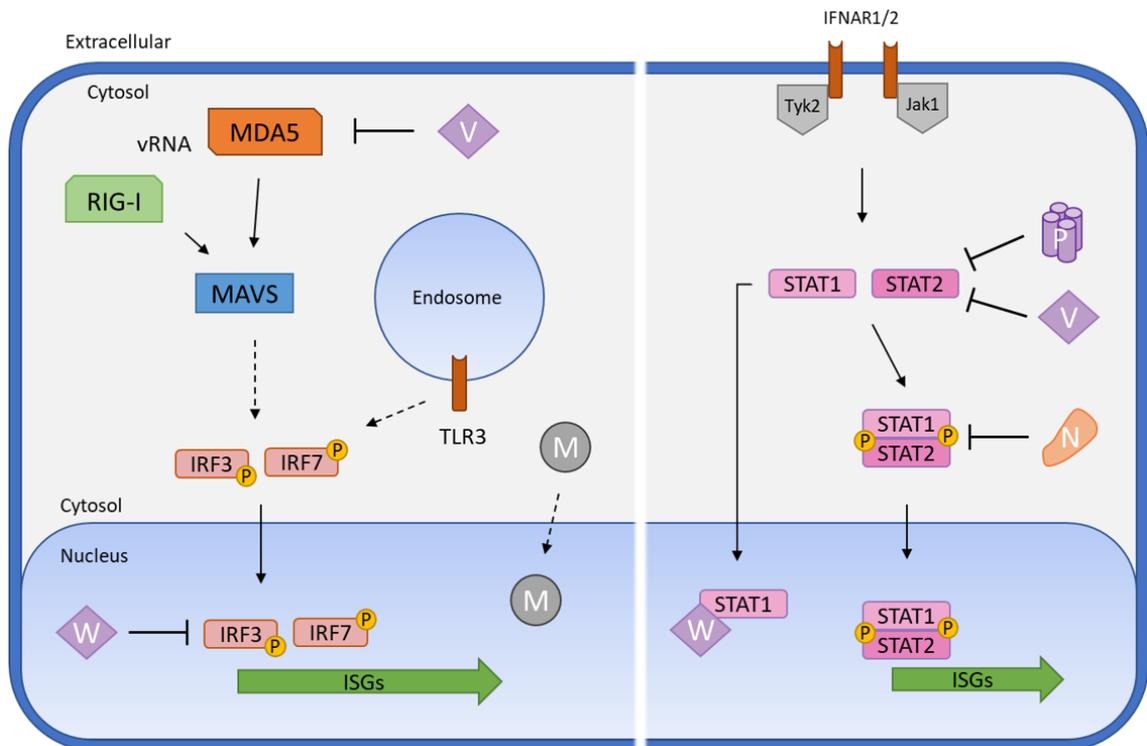


Figure 1.11. Henipaviral immune antagonism.

Left; The M, V, and W proteins antagonise interferon stimulated gene (ISG) expression through various mechanisms. Normally, the activation of RIG-I and MDA-5 through detection of viral RNA (vRNA) as well as the activation of endosomal TLR3 lead to an activating phosphorylation of IRF3 and IRF7 to generate phospho-IRF3 (pIRF3) and phospho-IRF7 (pIRF7). Once activated, pIRF3 and pIRF7 translocate to the nucleus to stimulate the expression of ISGs. Henipaviral P gene products, the V and W proteins, interfere with this signalling pathway in the cytosol (V protein) or the nucleus (W protein). Additionally, the M protein translocates to the nucleus where is associates with the nucleolar compartment. **Right;** Following binding of IFN α/β , the IFN receptors IFNAR1 and IFNAR2 activate the Jak/STAT signalling cascade. Phosphorylation and association of STAT1 and STAT2 into the phospho-STAT1/phopsho-STAT2 (pSTAT1/pSTAT2) complex enables their translocation to the nucleus to stimulate the expression of ISGs. Henipaviral P gene products, the P, V, and W proteins, interfere with this signalling pathway by sequestering STAT1/2 or pSTAT1/2 with various efficiencies. Additionally, the N protein interferes with the activation of pSTAT1. The activation and signalling of PKR/NF κ B is not shown.

complexes by henipaviral P, V, and W proteins with a range of antagonism efficiencies (Rodriguez et al., 2003; Eaton et al., 2006). Sequestration of host proteins in these high molecular weight complexes limits their phosphorylation and abrogates their function in signalling (Rodriguez et al., 2003; Shaw et al., 2005; Plemper and Lamb, 2021). The C protein also showed IFN antagonism, although to a lesser extent (Park et al., 2003). The mechanism of paramyxoviral C protein-mediated immune antagonism is thought to involve modulation of the activity of the viral polymerase to limit the production of dsRNA thus indirectly inhibiting the induction of innate immune signalling pathways (Shaw et al., 2005; Plemper and Lamb, 2021). The importance of the henipaviral V and W proteins for viral pathogenesis is highlighted by the low-pathogenic member *Cedar henipavirus* which lacks expression of V and W proteins and the ability of P protein to target STAT (Marsh et al., 2012; Lieu et al., 2015).

The HeV and NiV N protein has recently also been implicated in IFN antagonism by interfering with the activation of STAT1 (Figure 1.11). The molecular mechanism of N-mediated antagonism of pSTAT1 is not completely clear, but does not seem to rely on binding STAT1 nor preventing its phosphorylation (Sugai et al., 2017).

The henipaviral M protein interacts with host factors independent of its role in virion assembly (see below). Following nuclear localisation, the M protein associates with the nucleolar compartment (Figure 1.11). The full function of nucleolar M protein is not fully understood, but it is suggested that M modulates the DNA damage response pathway to promote a cellular environment for efficient virus replication (Wang et al., 2010; Rawlinson et al., 2018).

1.7.4 Virion maturation, assembly, and exit

Similar to most enveloped viruses, progeny virions assemble through a budding process. The structural proteins, which include the M protein, the F and G glycoproteins, and the holonucleocapsid, assemble at sites on the plasma membrane where nascent virion buds emerge and are pinched off. The congregation of viral components to the sites of assembly is coordinated through extensive protein-protein and protein-lipid interactions with a central role for the M protein, but host factors may play important roles in this process as well. Some viral components must undergo multiple steps of maturation and processing before progeny virions can emerge. For example, the nucleocapsid complex is assembled in a co-transcriptional fashion (see above).

To constitute a biologically active F protein, the precursor F₀ has to undergo proteolytic cleavage generating two disulphide-linked subunits: F₁ and F₂ (Scheid and Choppin, 1974). Most

paramyxoviral F_0 proteins are cleaved by the calcium-dependent protease furin, an abundant protein located within the trans-Golgi network (TGN) of eukaryotic cells (Hatsuzawa et al., 1990). The henipaviral F protein cleavage site is different from other paramyxoviruses, and as a consequence a role for furin in F protein maturation was excluded by the observation that furin-deficient cells support henipaviral replication and F protein maturation (Michalski et al., 2000; Pager et al., 2004). Instead, following expression and passage through the endoplasmic reticulum (ER) and TGN, F_0 is endocytosed after initial cell-surface localisation (Figure 1.7) (Meulendyke et al., 2005). The classical endocytic YXX ϕ motif and specific residues Ser490 and Tyr498 within the F protein cytoplasmic tail regulate trafficking through the early, sorting, and recycling endosomal compartments (Meulendyke et al., 2005; Popa et al., 2012). Upon reaching the lysosomal vesicle stage, lysosomal cysteine proteases cathepsin L and B cleave F_0 into F_1 and F_2 (Pager and Dutch, 2005; Diederich et al., 2012). The matured F protein is then recycled to the cell-surface for virion budding (Figure 1.7). Since viral entry occurs at neutral pH and matured F protein is fusogenically competent, cycling of the mature F protein to the cell surface and association with the G protein may initiate fusion with neighbouring (uninfected) cells. The resulting multinucleated cells – or syncytia – are a characteristic cytopathic effect (CPE) for paramyxoviruses in both cultured monolayers and isolated tissue samples (Murray et al., 1995; Selvey et al., 1995; Hyatt et al., 2001; Eaton et al., 2006). Moreover, the formation of syncytia may be a mechanism for the propagation of viral infection outside of virion budding. Interestingly, F_0 is more readily detected for HeV than for NiV, suggesting it is less efficiently cleaved by cathepsin L (Michalski et al., 2000; Wang et al., 2001). As a consequence, HeV virions may contain fewer mature F proteins and fusion with the cell surface could be less efficient. Indeed, syncytia observed in NiV-infected Vero cells are significantly larger than those in HeV-infected Vero cells (Goldsmith et al., 2003)

The progeny nucleocapsid associates with P-L protein complexes to form the holonucleocapsid. This association is essential for the generation of infective virions as the full RdRp complex is required for transcription of the nucleocapsid template to mRNA (Kingsbury et al., 1978; Plemper and Lamb, 2021). While packaging of one holonucleocapsid into a progeny virion is sufficient for infectivity, biochemical and cryo-ET studies suggest some paramyxoviruses are capable of packaging multiple holonucleocapsids at once, thus generating polyploid virions (Rager et al., 2002; Loney et al., 2009; Goff et al., 2012; Ke et al., 2018; Plemper and Lamb, 2021; Conley et al., 2022).

The M protein likely plays three major roles in the virus life cycle: organisation of the progeny virion assembly, stimulation of virion budding, and sequestration of inhibitory host

factors. The ability to form highly organised assemblies is exemplified by purified M proteins that are capable of self-association into paracrystalline arrays of sheets and tubes (Bächi, 1980; Heggeness et al., 1982). Indeed, arrays of dimeric M protein lining the viral lipid bilayer of filamentous RSV virions exhibit a helical arrangement and appear to align with putative transmembrane glycoproteins (Conley et al., 2022). Although not belonging to the family *Paramyxoviridae*, as a member of the family *Pneumoviridae* RSV is still related to paramyxoviruses (Figure 1.1A). A similar arrangement of M protein and the glycoproteins has been observed for NDV (Battisti et al., 2012). The M protein has classical nuclear localisation and nuclear export sequences (NLS, NES) that mediate nuclear trafficking (Figure 1.7). The M protein first localises to the nucleus, specifically the nucleolar compartment, before exiting and localising to the plasma membrane where it coordinates virus assembly and budding. This nuclear transit is critical for the M protein to become budding-competent (Wang et al., 2010; Monaghan et al., 2014; Pentecost et al., 2015). A central role for the M protein in virion formation and budding is further supported by henipaviral M protein expressed in isolation, which is sufficient to drive virus-like particle (VLP) formation (Patch et al., 2007). The M protein bridges the viral membrane proteins with the nucleocapsid. M associates with the viral glycoproteins F and G on the cell surface through interactions with the F and G cytoplasmic tails and it interacts with the flexible Ntail domains of the nucleocapsid complex (Figure 1.7). For many paramyxoviruses these components localise to lipid rafts on the plasma membrane of the infected cell, however assembly of viral components on these rafts does not appear to be a universal strategy (Plempner and Lamb, 2021). The dimeric henipaviral M protein anchors to the inner leaflet of the plasma membrane by synergistically associating with phosphatidylserine (PS) and phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂) through the basic M protein CTD (Norris et al., 2022). Binding these lipid head groups induces large conformational changes to the concave membrane-binding surface of the M homodimer, with the resulting flat surface inducing local plasma membrane curvature. The lipid-induced conformational change also favours M protein lattice polymerisation leading to the formation of filamentous buds (Norris et al., 2022). Furthermore, M recruits host machinery for vesicle budding, such as vacuolar protein sorting-associated protein 4 (VPS4) (Figure 1.7) (Park et al., 2016). Interestingly, the henipaviral C protein has been implicated in recruiting components of the endosomal sorting complexes required for transport (ESCRT) pathway which, in association with VPS4, catalyse membrane fission to release budding virions (Park et al., 2016; Plempner and Lamb, 2021).

In infected polarised cells, many paramyxoviruses appear to bud preferentially from the apical membrane which has implications for viral pathogenesis. Apical budding favours

restriction of viral spread to the epithelia while basolateral budding grants access to the underlying tissues and facilitates development of systemic infections (Plempner and Lamb, 2021). While the primary henipaviral replication site is unconfirmed, the respiratory route of infection suggests involvement of cells lining the oro-/nasopharyngeal and bronchiolar epithelia, pneumocytes, alveolar macrophages, or pulmonary airway dendritic cells. The distribution and timing of lesions in the vascular endothelia, the brain, and the lungs is suggestive of basolateral budding from the epithelia and secondary infection following hematogenous spread, with secondary replication occurring in the vascular endothelia (Wong et al., 2002; Parks and Alexander-Miller, 2013; de Wit and Munster, 2015; Baseler et al., 2016; B. Lee et al., 2021). Indeed, the henipaviral F and G glycoproteins are trafficked to the basolateral membrane when expressed in isolation (Weise et al., 2010).

1.8 Hendra henipavirus diagnostics and surveillance

Following symptoms consistent with viral infection, rapid diagnosis is of key importance to limit disease propagation. Isolation of the virus and establishment of an aetiology was an essential step in the initial discovery of HeV. Characterisation of the virus morphology, serology, and nucleotide sequence in any combination is the main approach to confirm infection with HeV, another member of *Henipavirus*, or another paramyxoviral agent (B. Lee et al., 2021). Specimens containing high viral loads are suitable for direct examination using EM and immunohistochemistry (IHC) for early detection and diagnosis. Either non-specifically stained or immuno-labelled EM showing morphological features of the virions aid in the identification. As discussed in Section 1.6.2, HeV virions exhibit a characteristic 'double fringe' of glycoproteins on the viral membrane (Hyatt et al., 2001). The viral helical nucleocapsid can also be used to narrow down the viral genus within the family *Paramyxoviridae*, since its overall width and helical pitch show some distinctiveness among *Paramyxoviridae* genera (Plempner and Lamb, 2021; B. Lee et al., 2021). IHC-labelled tissue samples or serology can rapidly confirm the presence of specific virus in infected cells or fluid samples. Polymerase chain reaction (PCR)-based diagnostic tools, including quantitative real-time PCR, are highly specific and can confirm infections when a particular virus is suspected, however they may fail to detect any variants of known henipaviruses or novel, undiscovered henipaviruses (B. Lee et al., 2021). If a henipaviral outbreak is suspected, virus isolation and propagation may be attempted at biosafety level (BSL) 3. The observation of CPE pathognomonic of henipavirus infection, such as syncytia, or confirmation by serology or PCR is cause for increased safety measures. Subsequent work with infected cultures and live virus should be handled under BSL4 conditions (B. Lee et al., 2021). A virus

neutralisation test (VNT) is the reference standard for outbreak investigation and disease surveillance, however conducting a VNT requires access to BSL4 facilities (B. Lee et al., 2021).

Serology represents a rapid, low-tech tool that is invaluable to not only the diagnosis of disease but also surveillance among humans and any wildlife populations that serve as viral reservoirs (Pernet et al., 2014). HeV- and NiV-specific monoclonal antibodies (mAb) can differentiate between the two viruses in pteropid populations. Since a VNT is often impractical in the field, serological tests that do not require BSL4 containment have been developed, such as the enzyme-linked immunosorbent assay (ELISA) and protein array multiplex (B. Lee et al., 2021). Surveillance may be further aided by increasingly more accessible PCR and sequencing techniques. These techniques allow for confirmation of a particular viral species, similar to serology-based approaches, yet unlike serology they may also detect variant strains of known species or facilitate discovery of unknown, novel species.

1.9 Hendra henipavirus treatment and prevention

There are no approved HeV or NiV antiviral therapies or vaccines for human use, however off-target therapies have been used in open-label studies or as compassionate treatment with varying degrees of success and henipaviral-specific treatment is under development.

Antiviral drugs have been used during the multiple HeV and NiV outbreaks. Ribavirin, which is confirmed to inhibit HeV replication *in vivo* (Wright et al., 2005), was used in an open label study to treat infections with NiV during an outbreak in Malaysia. Treatment reduced mortality by 35% in the treated group and significantly reduced the duration of hospitalisation compared to the control group (Chong et al., 2001). On the other hand, treatment of two cases of HeV infection with ribavirin failed to provide any positive health outcomes. *In vivo* testing of ribavirin in animal models of HeV or NiV infection had no significant effect on disease progression or outcome, or only served to delay mortality (B. Lee et al., 2021). The nucleotide analogue pro-drug remdesivir is currently in development to combat *Zaire ebolavirus* and its antiviral activity was also demonstrated for a wide range of viruses, including HeV and NiV (Lo et al., 2017). The antimalarial drug chloroquine has been shown to block HeV F maturation and to inhibit HeV and NiV infection in cell culture, however administration of chloroquine combined with ribavirin to a HeV-infected person had no clinical benefits (B. Lee et al., 2021).

Potential *Henipavirus*-specific treatments show promising initial results. A soluble form of the HeV G glycoprotein (sG), retaining native properties such as oligomerisation, has been shown to bind the viral receptor ephrin B2 thereby blocking viral adsorption to otherwise susceptible cells (Bossart et al., 2005). In another approach, a heptad peptide-based fusion inhibitor successfully associates with the fusogenic F protein. Binding of the peptide blocks the formation of the trimer-of-hairpins which is required for membrane fusion and subsequently prevents viral entry into susceptible cells (Bossart et al., 2001; Bossart et al., 2002; B. Lee et al., 2021). The development of *Henipavirus*-specific antivirals requires further investigation, especially in the absence of robust and widely available vaccines (see below).

Individuals naturally infected with HeV generated high levels of cross-reactive antibodies against NiV, while individuals naturally infected with NiV generated antibodies with only limited cross-reactivity (Bossart et al., 2007). In a hamster model, anti-G and anti-F antibodies afforded protection from lethal challenge with NiV and HeV following passive transfer of these antibodies to a susceptible host (Guillaume et al., 2004; Guillaume et al., 2009). An isolated human mAb against HeV G protein, m102.4, demonstrated protection from lethal challenge in a number of animal models (B. Lee et al., 2021). The m102.4 antibody has since been successfully administered on 15 occasions following an emergency use protocol due to significant HeV or NiV infection risk, and has completed phase 1 clinical trials in Australia (Broder et al., 2016; Playford et al., 2020; B. Lee et al., 2021; Wang et al., 2022). Additionally, recent investigations into the antigenicity of NiV G protein showed that combined treatment with m102.4 and another mAb – nAH1.3 – synergistically neutralised G protein-mediated entry in a VNT and offered protection from challenge in an animal model, as well as increasing the barrier for the occurrence of escape mutations (Wang et al., 2022).

The development of prophylaxes has focused on animal models and animal targets, and includes a range of immunisation strategies, including live-recombinant virus platforms, protein subunit vaccines, VLP vaccines, and DNA vaccines (Broder et al., 2016). Infection of mice and hamster models with vaccinia virus expressing NiV F and G proteins elicited neutralising antibodies, with higher titres for anti-G antibodies than for anti-F antibodies, and hamsters were protected from lethal challenge following this immunisation (Tamin et al., 2002; Guillaume et al., 2004). Further viral vector-based vaccines have been developed using platforms based on adeno-associated virus, MeV, and VSV using HeV or NiV glycoproteins with similar results (Yoneda et al., 2013; Ploquin et al., 2013; Mire et al., 2013; Lo et al., 2014; DeBuysscher et al., 2014). A single viral vector vaccine based on the NiV N protein has also been established with minimal protection from challenge but may have important effects on cell-based immunity

(DeBuysscher et al., 2014). HeV sG has been shown to elicit strong cross-reactive neutralising antibodies in rabbits against both HeV and NiV, suggesting a single sG-based vaccine may be effective against both HeV and NiV infection (Bossart et al., 2005). The HeV sG-based vaccine has been evaluated in nonhuman primates, offering complete protection from lethal challenge with no clinical signs, viral replication, or pathology observed in any of the vaccinated animals (Mire et al., 2014). These results, and the inherent safety and simplicity, support the preclinical development of HeV sG-based vaccines. Indeed, a HeV sG-based vaccine has been developed for equine use (Middleton et al., 2014). Released in 2015 under the name *Equivac HeV*, the subunit vaccine was the first commercially developed and deployed vaccine against a BSL4 pathogen and is currently the only licenced antiviral prophylaxis for henipaviral infection. Not only does vaccination directly prevent infections in horses but by extension also reduces the risk of infection in humans since all human HeV infections thus far have required an equine intermediary host. Indeed, more than 153000 horses have been fully vaccinated with *Equivac HeV* and since its release all confirmed HeV infections have occurred in unvaccinated horses (B. Lee et al., 2021). Recent vaccination and challenge studies in non-human primate animal models show that even a single dose of HeV-sG vaccine protects against lethal challenge with HeV and NiV within 7 days post-inoculation, highlighting their potential for emergency use in an outbreak scenario (Geisbert et al., 2021). Taken together, the results of both active and passive immunisation studies represent a step forward in limiting further HeV and NiV outbreaks, mortalities, and morbidities.

1.10 Thesis aims

The work presented in this thesis is part of the MSCA-ITN HONOURS network on host-switching pathogens, infectious outbreaks, and zoonosis (van der Hoek et al., 2018). The outbreak of a novel pathogen is accompanied by gaps in fundamental knowledge and technical resources. It is therefore of paramount importance to increase these key assets prior to an outbreak. While preparedness systems are falling into place for HeV, including surveillance, diagnostic tools and laboratory confirmation, and treatment and prevention, many questions regarding the molecular biology of HeV remain unanswered. In the broadest sense, the work presented in this thesis aims to reduce the gaps in fundamental knowledge and technical resources with a focus on *Hendra henipavirus*. The HeV N protein remains structurally unresolved in the literature. The elucidation of the HeV N protein structure enables the identification of differences between it and related viral N protein structures, but importantly

may also highlight commonalities that can be exploited for the design of new, broad-acting therapeutics. To that aim, four main objectives were established for this project:

1. The expression and purification of recombinant *Hendra henipavirus* nucleoprotein using a bacterial expression system and a combination of affinity and size exclusion chromatography techniques.
2. The characterisation of *Hendra henipavirus* nucleoprotein monomers and oligomers through negative stain- and cryo-electron microscopy.
3. The elucidation of the *Hendra henipavirus* nucleoprotein structure in its monomeric form and within the context of an oligomer, through cryo-electron microscopy image processing and three-dimensional (3D) model building.
4. The development of a *Hendra henipavirus* cell-based mini genome replicon system for the validation of structural data and the discovery of novel, *Henipavirus*-specific antivirals.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Buffers and Media Recipes

Table 2.1. Buffer and media compositions

Name	Contents
LB medium (in-house)	8 gr/l Tryptone Plus (Sigma Aldrich) 4 gr/l Yeast Extract (Sigma Aldrich) 8 gr/l NaCl (Fisher Scientific) Optional antibiotic: 50 µg/ml Kanamycin (Bio Basic)
2xYT medium (in-house)	16 gr/l Tryptone Plus 10 gr/l Yeast Extract 10 gr/l NaCl Optional antibiotic: 50µg/ml Kanamycin
Small-scale lysis buffer	50 mM Tris-HCl pH 8.0 (ChemCruz) 500 mM NaCl 15 mM Imidazole (Acros Organics) 0.1% Triton X-100 (SAFC) 0.1% Polysorbate 20 (Acros Organics) 0.1% Nonidet P40 Substitute (Fluka Biochemika) 10 µg/ml DNase I (Roche) 20 µg/ml RNase A (Invitrogen) 0.25 µg/ml Lysozyme (Sigma Aldrich)
Pre-lysis buffer	20 mM Tris-HCl pH 7.5 500 mM NaCl 20 mM Imidazole 0.1% Triton X-100
Large-scale lysis buffer	Pre-lysis buffer cOmplete Mini Protease Inhibitor cocktail tablets (Sigma Aldrich) 10 µg/ml DNase I 0.25 µg/ml Lysozyme Optional addition: 20 µg/ml RNase A
2x SDS loading dye	100 mM Tris-HCl pH 6.8 4% SDS (Sigma Aldrich) 20% Glycerol (Fisher Scientific) 0.2% Bromophenol blue (Fisher Scientific) 200 mM DTT (Melford)
15% SDS-PAGE resolving gel	23%v/v H ₂ O 50%v/v 30% acrylamide (Severn Biotech) 25%v/v 1.5M Tris-HCl pH 8.8 0.1% SDS 0.1% APS (AGTC Bioproducts) 0.04%v/v TEMED (Sigma Aldrich)
5% SDS-PAGE stacking gel	68%v/v H ₂ O 17%v/v 30% acrylamide 10%v/v 1.0M Tris-HCl pH 6.8 0.1% SDS 0.1% APS 0.1%v/v TEMED
Gel blue destain	45% H ₂ O 45% Methanol (Fisher Scientific) 10% Acetic Acid Glacial (Fisher Scientific)

Table 2.1. Recipes *continued*

Name	Contents
Gel blue stain	Gel blue destain 2.5 gr/l Brilliant Blue R (Sigma Aldrich)
Affinity column buffer	20 mM Tris-HCl pH 7.5 500 mM NaCl 20 mM Imidazole
Dialysis buffer	20 mM Tris-HCl pH 7.5 300 mM NaCl
SEC buffer 1	25 mM Tris-HCl pH 7.5 250 mM NaCl 0.5 mM DTT 2.5% glycerol
SEC buffer 2	25 mM Tris-HCl pH 7.5 100 mM NaCl

2.1.2 Consumables

2.1.2.1 Vectors

All HeV gene sequences were taken from the reference *Hendra henipavirus* genome (NCBI Reference Sequence: NC_001906.3). A HeV N ORF (codon optimised) was supplied in pTwist and sub-cloned into the pET28a(+)-6xHis-SUMO expression vector for recombinant protein expression and purification from bacterial systems. The HeV L, P, and N genes were synthesised by GENEWIZ in the pUC57-KanR vector for mammalian cell culture. The NTR and GS sequences for HeV L, PdC, and N were ordered through the FragmentGENE gene service by GENEWIZ, inclusive of common 5' and 3' sequences for ease of PCR amplification (Table 2.2).

2.1.2.2 Primers

Sequencing and amplification primers were designed using Benchling and ordered through Integrated DNA Technologies.

2.1.2.3 Bacterial cell lines

The *Escherichia coli* (*E. coli*) DH5- α strain (Thermo Fisher Scientific) was used for plasmid amplification. For bacterial expression of HeV N, a range of *E. coli* expression strains were used (Table 2.2). All *E. coli* strains were produced in house by Susan Matthews.

Table 2.2. FragmentGENE segments

Name	Sequence
HeV_L_UTR	GTCAGCACATGAGTGCAACGGAGCTCTAATACGACTCACTATAGGGAGGACCCAAGTCC TTAAACGTGCATGATTGAGATAGTCAGAAATTGGAGCTAATTTGAATACTAGTGAGTAA ATTAGGAGTAGTAGTATTGAACCACCTCATATCTTATTTCTTCTGAACTGCATTAATAATTT GTACCCAGGATATAACAACATGGCTCATGAATTATCTATCTCTGATATCATCTATCCTGAA TGTCACCTTGACAGTCCTATAGTATCAGGGAAATTGATTTGAGCTATAGAATATGCGCAG TTGAGACATAACCAACCGAATGGCGTAGTAGGTTGTCTGACAATCATCC
HeV_PdC_UTR	GTCAGCACATGAGTGCAACGTCTAGATAATACGACTCACTATAGGGAGGATCCAAGACC ATAAATCTAGGATCCTTTACAATCTCAGACCCTGGTGCAAGGTTATATATAGAGCCCAAT TCTCCTAATTAACAGTGTCTCAGGTTGACAAATGGACAAGTTGGATCTAGTCAACGACG GCCTAGATATTATTGACTTTATTCAGAAGAACCACAAAAGAAATACAAAAGACATACGGAC GATCAAGCATCCAACAACCAAGTACCAAAGACAGGACAAGAGCATGGGAGGACTTCTTG CAGAGCACCAAGTGGAGAACATGAACAGGCTGAGGGGGGAATGCCTAAGAATGATGGA GGTACTGAAGGAAGAAATGTGGAGGATCTATCCAGTGTACTTCTCAGATGGAACATAT TGGACAAAGAGTGTCAAACACCCGAGCTTGGGCAGAAGACCCAGATGACATACAACCTG GACCCAATGGTTACAGACGTTGTATACCATGATCATGGAGGAGAATGTACCGGACATGG ACCTTCTTCAAGCCCTGAGAGAGGGTGGAGTTATCACATGTCAGGAACACACGATGGGA ATGTACGTGCTGTACCTGATACAAAGGTGTTGCCAATGCTCCAAAACACTACAGTTCCTG AAGAAGTTAGGGAAATTGATTTAATTGGGTTGGAGGACAAATTTGCATCAGCGGGATTA AATCCAGCTGCAGCGTAGTAGGTTGTCTGACAATCATCC
HeV_N_UTR	GTCAGCACATGAGTGCAACGTCTAGATAATACGACTCACTATAGGGAGGAACCAAGACA GTGACAATTGGTCTTGGTATTGGACAATTGTTCAAGGTTCCAAAATGAGTGATATATTTG ACGAGGCGGCTAGTTTCAGAAGCTATCAATCGAAACTCGGTGAGATGGGCGGGCAAG TGCGGCAACAGCTACTTTGACTACTAAGATAAGAATTTTTGTACCAGCAACTAATAGCCC AGAACTGAGATGGGAGTTGACTTTGTTGCTCTCGATGTAATCAGATCACCAAGTGCAGC AGAATCAATGAAGATTGGTGCTGCTTTCCTCTGATATCCGATAGGTTGTCTGACAAT CATCC

2.1.2.4 Mammalian cell lines

BHK-21 cells were isolated from kidney tissue of Syrian golden hamsters (*Mesocricetus auratus*) (Stoker and Macpherson, 1964). BSR-T7 cells were derived from the BHK-21 cell line and have been stably transformed to constitutively express T7 RNA polymerase (Buchholz et al., 1999).

Table 2.3. Bacterial expression strains

Name	Manufacturer	Genotype
BL21-Gold(DE3)pLysS	Agilent	<i>E. coli</i> B F ⁻ <i>ompT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm</i> ⁺ Tet ^R <i>gal</i> λ(DE3) <i>endA Hte</i> [pLysS Cam ^R]
BL21-Star(DE3)pLysS	Thermo Fisher Scientific	<i>E. coli</i> F ⁻ <i>ompT hsdSB</i> (r _B ⁻ m _B ⁻) <i>gal dcmrne131</i> λ(DE3) [pLysS Cam ^R]
BL21-Rosetta 2(DE3)pLysS	Novagen	<i>E. coli</i> F ⁻ <i>ompT hsdSB</i> (r _B ⁻ m _B ⁻) Tet ^R <i>gal dcm</i> λ(DE3) [pLysSRARE2 Cam ^R]

2.1.2.5 Transmission electron microscopy grids

Grids used for negative stain EM were 400 mesh continuous carbon-coated copper grids that were purchased from in-house produced stock by the Astbury Biostructure Laboratory facility. Grids used for cryo-EM included lacey carbon grids with ultrathin carbon support (Agar Scientific) or holey/Quantifoil grids (Quantifoil).

2.2 Methods

2.2.1 Molecular biology

2.2.1.1 Bacterial transformation and DNA amplification

Aliquots of competent *E. coli* cells (DH5α, Rosetta2, Star, and Gold) were thawed on ice and split into 50 μl volumes. Each volume was gently mixed with 10-100 ng plasmid DNA and incubated on ice for 25 min. The mixtures were heat-shock transformed by incubating at 42°C for 45 sec, after which they were immediately placed on ice for 2 min. The cells recovered by addition of 250 μl LB medium (without antibiotic) and incubated at 37°C for 45 min in an SI500 orbital incubator (Stuart Equipment). The transformed bacteria were plated out on 10 cm LB agar (1.2% w/v) laced with the appropriate antibiotic and incubated overnight at 37°C. Colonies were picked and mixed with 10 ml LB containing the appropriate antibiotic and incubated overnight at 37°C. The cultures were used to generate glycerol stocks by mixing 750 μl of the culture with 750 μl sterile filtered 50% v/v glycerol and stored at -80°C. The remainder could be used to purify the amplified DNA as described below if desired.

2.2.1.2 DNA purification

Following amplification, plasmid DNA was purified using a QIAprep Spin Miniprep kit (Qiagen) or QIAGEN Plasmid Midi kit (Qiagen) according to the kit manuals. Where appropriate, bands of PCR or restriction products were resolved on agarose gel (see below), excised from the gel and purified using a QIAquick Gel Extraction kit (Qiagen) according to the kit manual.

2.2.1.2 Cloning *Hendra henipavirus* N gene into bacterial expression vectors

Carrier plasmids containing codon optimised nucleoprotein genes, flanked by 5' BamHI and 3' XhoI restriction sites, were obtained from Genewiz. Carrier plasmids were amplified in DH5 α competent cells. Purified DNA was digested with BamHI-HF (NEB) and XhoI (NEB) for 2 hours at 37°C. Restriction products were resolved on a 1% agarose gel (Melford) and the appropriate bands were purified. The HeV N gene fragments were then ligated to previously digested pET-28a-6His-SUMO plasmids in a 1:4 molecular ratio using T4 DNA ligase (NEB). The ligation reaction was incubated over night at 16°C and heat-shock transformed into DH5 α competent cells. pET-28a-6His-SUMO-HeVN plasmids (containing the viral gene) were then amplified and purified and stored at -20°C. HeV N protein mass was predicted at 71.97 kDa for 6His-SUMO-HeVN and 58.48 kDa for tag-free N, using ExPASy ProtParam.

2.2.1.2 Cloning *Hendra henipavirus* 3' NTR sequences into pUC57-support vectors

The FragmentGENE segments were PCR amplified using shared 5' and 3' primers: GTCAGCACATGAGTGCAACG and GGATGATTGTCAGACAACCTACTACG. Following PCR amplification, NTR segments and pUC57-HeV-N, pUC57-HeV-P, and pUC57-HeV-L were digested using the appropriate restriction enzymes and buffers. Restriction products were resolved on a 1% agarose gel (Melford) and the appropriate bands were purified. The NTR segments were then ligated to previously digested pUC57 plasmids in a 1:4 molecular ratio using T4 DNA ligase (NEB). The ligation reaction was incubated over night at 16°C and heat-shock transformed into DH5 α competent cells. pUC57-HeV-N/PdC/L plasmids (containing the viral genes, respective NTRs, and GSs) were then amplified and purified and stored at -20°C.

2.2.1.4 Sequence validation

All DNA sequences were validated via Sanger sequencing (Genewiz) and aligned to the plasmid map using MAFFT (Multiple Alignment using Fast Fourier Transform).

2.2.2 Recombinant protein purification

2.2.2.1 Starter cultures

Glycerol stocks of pET-28a-6His-SUMO-HeVN were used to inoculate 5 ml or 10 ml LB medium with kanamycin, for small- or large-scale expressions, respectively. The starter cultures were incubated overnight at 37°C in an SI500 orbital incubator.

2.2.2.2 Protein expression and cell lysis

The starter cultures were used to inoculate 2xYT medium with kanamycin in a 1/500 ratio. Inoculi were grown at 37°C in a Multitron Standard incubator (Infors HT) until reaching an OD⁶⁰⁰ of 0.5, at which point the growth temperature was reduced to 26°C and the inoculi were further grown until an OD⁶⁰⁰ of 0.8 was reached. Expression of HeV N was then induced by addition of 0.5 mM IPTG (ChemCruz) and cultures were further incubated for 4 hours at 26°C. Small- or large-scale expressions were centrifuged for 25 minutes at 4500 rpm in an AccuSpin R1 (Fisher Scientific) or Avanti J26-XP with JLA 8.1000 rotor (Beckman Coulter), respectively.

For small-scale expressions, the supernatant was discarded and pellets were resuspended in 3 ml small-scale lysis buffer and incubated for 30 minutes on ice. Lysates were subjected to three freeze-thaw cycles (-80°C to +40°C). Lysates were transferred to 2 ml Eppendorf tubes and centrifuged for 30 minutes at 13000 rpm and 4°C in a Micro Star 17R (VWR). The soluble fraction was collected in fresh Eppendorf tubes and the insoluble fractions carefully washed once with dH₂O and resuspended in dH₂O.

For large-scale expressions, the supernatant was discarded and pellets were resuspended in 30 ml prelysis buffer. Pre-lysates were subjected to one freeze-thaw cycle (-80°C to +40°C). The pre-lysis buffer was completed to make the large-scale lysis buffer and the lysates were incubated for 30 minutes on ice with a magnetic stirrer. Lysates were sonicated for 10 cycles of 15 seconds on, 45 seconds off using the SoniPrep 150 Ultrasonic Disintegrator with External Process Timer (MSE). Lysates were transferred to ultracentrifuge tubes and centrifuged for 45 minutes at 16000 rpm and 4°C in an Evolution RC with SA300 rotor (Sorvall). The soluble fraction was collected in 50 ml falcon tubes and carried forward to purification.

2.2.2.3 Affinity chromatography

Soluble fractions were passed through a primary HisTrap High Performance 5 ml column (GE Healthcare) which was previously equilibrated with column buffer, using a 101U/R peristaltic

pump (Watson Marlow). The column was washed with multiple washes of 25 ml column buffer with increasing concentrations of imidazole (20 mM, 50 mM, 100 mM, 300 mM, and 500 mM). The flow-through of the soluble fraction and each of the wash fractions was collected in full. The presence of protein was confirmed with Bradford Assay (Bio-Rad), NanoDrop One (Thermo Fischer Scientific), and SDS-PAGE analysis (see below).

Fractions containing the majority of the protein of interest were transferred into Spectra/Por 3 Standard RC Tubing (MWCO 3.5 kDa) (Spectrum) with in-house produced SUMO protease added. Fractions were dialysed overnight against dialysis buffer at 4°C. SUMO protease cleavage was confirmed via SDS-PAGE analysis (see below), and the dialysed, cleaved mix was sterile filtered using a Minisart NML Syringe filter 0.45 µm SFCA (Sartorius) and passed through a secondary HisTrap column. The flow-through was collected and kept on ice, and the column was washed with pre-chilled low- and high-imidazole washes. The flow-through was sterile filtered and concentrated to a final volume of 7 ml using Amicon Ultra-15 Centrifugal filter concentrator (Merck Millipore).

2.2.2.4 Size exclusion chromatography

Size exclusion chromatography (SEC) was then performed on the concentrate, using an ÄKTAprime Plus with HiPrep 26/60 Sephacryl 400 HR column (GE Healthcare) at 4°C, previously equilibrated with degassed and filtered SEC buffer 1 or SEC buffer 2. Fractions of 2.5 ml were collected after 80 ml had passed through. Presence of protein of interest was confirmed with SDS-PAGE analysis (see below). Concentrations and 260/280 ratios were determined for pooled or individual fractions, and fractions containing proteins of interest were either directly used for electron microscopy (see below) or flash frozen and stored at -80°C.

2.2.2.5 SDS-PAGE gels and staining

Samples were mixed with 2x SDS loading dye and incubating at 95°C for 5 minutes. Samples were resolved on 15% SDS-PAGE gels with 5% stacking gels. A Precision Plus Protein Standard (10 - 250 kDa) (Bio-Rad) was loaded along with the samples. Gels were stained with gel blue stain for up to 2 hours. The gels were de-stained with gel blue destain overnight and imaged with an inGenius Gel Doc and GeneSnap software (Syngene Bio Imaging).

2.2.3 Transmission electron microscopy

2.2.3.1 Preparation of negative stain grids

Carbon-coated 400 mesh copper electron microscopy grids (Agar Scientific) were glow discharged for 30 seconds at 10 mA on an easiGlow Glow Discharge Cleaning System (PELCO). The grids were held in place by a cross-locking forceps and 5 μl of sample (various concentrations) was applied and left for 1 min. The sample was blotted off with Whatman Filter paper and the grid was washed twice with 5 μl water and once with 5 μl 1% or 2% uranyl-acetate (AnalaR). Grids were negatively stained with 1% or 2% uranyl-acetate for 1 min before the excess was blotted off.

2.2.3.2 Preparation of cryo-grids

Copper Ultrathin Carbon (UC) Lacey grids, Quantifoil R2/2 400 mesh, or R1.2/1.5 300 mesh grids (Quantifoil) were glow discharged for 30 seconds at 10 mA on an easiGlow Glow Discharge Cleaning System. 3 μl sample at 1 mg/ml was applied to the grids and plunge frozen into liquid ethane cooled in liquid nitrogen using an FEI Vitrobot Mark IV (Thermo Fisher Scientific; blot force 6, blot time 6-8s, humidity 95-100%).

2.2.3.3 Data collection on negative stain grids

Micrographs were taken on a FEI Tecnai T12 (operated at 120 keV) which uses a lanthanum hexaboride (LaB6) filament with Gatan US4000/SP 4k x 4k CCD camera, at a nominal magnification of 68000x or 98000x with a pixel size of 0.17 or 0.11 nm, respectively. Micrographs were visualised in ImageJ.

2.2.3.4 Data collection on cryo-grids

Data was collected on an FEI Titan Krios (X-FEG) TEM (Thermo Fisher Scientific) (ABSL, University of Leeds) operated at 300 keV with a Falcon III direct electron detector in integrating mode (Thermo Fisher Scientific), at a nominal magnification of 75000x and a pixel size of 1.065 Å. Further data acquisition parameters are shown in Figure 4.6. A total of 3548 micrographs were taken at a total dose of 59.7 $\text{e}/\text{Å}^2$ split into 55 fractions and a defocus range of -0.7 to -2.8 μm . A detailed description of the data collection pipeline is has been published (Thompson et al., 2019). Micrographs were visualised in ImageJ or RELION 3.1.1.

2.2.4 Electron microscopy data processing

The dataset of 3548 micrographs was collected using an automated approach and pre-processed on the fly during collection using the RELION software suite (Scheres, 2012; Thompson et al., 2019). Motion correction and CTF estimation were carried out using RELION-integrated implementations of MotionCor2 (v1.2.1) (Zheng et al., 2017) or Gctf (v1.18) (Zhang, 2016). Following pre-processing, particles coordinates were manually selected using the manual picking job following the classical single-particle analysis (SPA) procedure. Alternatively, helical segments were manually selected using the manual picking job following the helical SPA procedure. For this, initial helical parameters were estimated from the 2D micrographs and closely related viral nucleocapsid structures. 3D maps were visualised in UCSF Chimera (Pettersen et al., 2004) and UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021)

2.2.4.1 Classical single particle analysis

For the classical SPA procedure, a total of 3500 particles were selected from a random subset of 502 micrographs to generate a reference for automated picking through 2D classification. Reference-based automated picking was then optimised for picking threshold, maximum stddev noise, and minimum avg noise values, resulting in 875,262 selected particles, extracted using a box size of 360 pixels (383.4 Å). Multiple rounds of 2D classification were carried out to enrich the particle stack with true particles and removing junk particles. The high-resolution 2D classes revealed the presence of a C2 and a C14 symmetry axis, for a combined D14 symmetry in the sauronoid particle.

An initial 3D reference map was generated *de novo* by stochastic gradient descent (SGD) and used for 3D classification and 3D refinement. Manual re-alignment and resampling of the C1 symmetry refined 3D map to the grid of the unaligned map in UCSF Chimera (Pettersen et al., 2004) enabled D14 symmetry impartation in subsequent 3D jobs. A solvent mask was generated from the 3D class in relion, using a lowpass filter of 15Å, an initial binarisation threshold of 0.012, a 3 pixel binary map extension, and a 3 pixel soft-edge. In post-processing, this solvent mask was applied to the refined sauronoid map to counter the influence of high-resolution noise to improve map quality and increase estimated resolution. The final EM map was deposited with the EMDB under accession EMD-16426.

2.2.4.2 Helical single particle analysis

For the helical SPA procedure, 88,363 start- and end-coordinates were manually selected from all 3548 micrographs. Helical segments were extracted along the line connecting the start- and end-coordinates following pre-defined, estimated helical parameters; a width of ~ 220 Å and a pitch of ~ 4.66 Å. Extraction resulted in 5,424,996 segments with a box size of 256 pixels (272.6 Å), which were down-sampled (binned) by a factor of 4 for early 2D classifications and re-extracted at nominal resolution for later 2D classifications once a satisfactory subset of classes containing 82,568 segments was generated.

Following 2D classification at nominal resolution, 66,587 segments were taken forward for 3D classification using helical reconstruction, assuming a tetradecameric helical turn. Estimates for the helical parameters were a twist of 25.7° (optimisation range 22° - 30°) and helical rise 4 Å (optimisation range 2.5 Å-5 Å). 3D refinement and post-processing of the optimal 3D class failed to generate a high-quality high-resolution reconstruction.

2.2.5 Atomic protein model refinement and interpretation

The D14 cryo-EM density map enabled atomic model refinement. SWISS-MODEL (Waterhouse et al., 2018) was used to predict an initial model based on homology search with the closest match to PDB: 4CO6 (Yabukarski et al., 2014). The homology model was input for automated model building in Buccaneer (Cowtan, 2006). The Buccaneer model was imported into Coot (Emsley and Cowtan, 2004; Emsley et al., 2010) for manual refinement, which included the removal of discontinuous peptide chains. Clear, high-resolution density enabled unambiguous, *de novo* building of a hexameric RNA chain, modelled as poly-uridine following nucleoprotein structure convention. Multiple rounds of automated real-space refinement in Phenix (Adams et al., 2010; Afonine et al., 2018) and manual adjustment in Coot resulted in the final model. Manual adjustments were carried out on a single monomer in Coot, which was symmetry expanded in UCSF Chimera according to D14 symmetry, prior to real-space refinement to prevent erroneous refinement into non-cognate density. Poor local density abrogated confident side chain placement and forced the stubbing of some amino acids, modelling them as alanines in the final model. Finally, model validation was carried out as part of the ultimate real-space refinement job through MolProbity (Chen et al., 2010; Williams et al., 2018) integrated within Phenix. Atomic models were deposited with the Protein Data Bank under accession 8C4H for a complete sauronoid assembly and 8CBW for the asymmetric protomer unit.

Interaction surfaces were calculated through PDBe PISA v1.52 (Krissinel and Henrick, 2007) and 3D atomic models were visualised in UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021).

2.2.6 Hendra henipavirus replicon system

2.2.6.1 Replicon design

The nucleotide sequences for the HeV N, P, and L genes were retrieved from NCBI (Reference Sequence: NC_001906.3) and used for the design of the HeV replicon system. The N, P, and L genes are each on separate plasmids and under transcriptional control of a T7 promoter. Mini genome design largely followed one described for *Nipah henipavirus* (Halpin et al., 2004). An EGFP sequence was used as a reporter, flanked on the 3' end by the HeV genomic 3' UTR, gene start (GS), and the 3' N gene non-translated region (NTR) and on the 5' by the 5' L gene NTR, gene end (GE) and HeV 5' UTR. The nucleotides 3'-CGUA-5' were inserted between the eGFP sequence and the 5' L gene NTR to ensure the mini genome conforms to the paramyxoviral rule-of-six. Generation of mini genomic RNA is under the control of a T7 promoter, with a hepatitis delta virus ribozyme (HDR) sequence upstream of the 3' UTR. This generates a native, negative sense eGFP mini genome upon L protein driven transcription. All sequences were ordered on pUC57 plasmids containing kanamycin resistance genes (Genewiz).

2.2.6.2 Cell maintenance

Mammalian BSR T7 cells (BHK-21 derived) were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Sigma Aldrich) supplemented with 2.5% foetal bovine serum and penicillin and streptomycin at 37°C and 5% CO₂. BSR-T7 cells were additionally supplemented with G418 (1 mg/mL) every other passage to maintain the T7P plasmid.

2.2.6.3 Cell passage and seeding

Working stocks of BSR T7 cells were maintained in T-75 flasks, which were changed every second passage. At 80% confluency, cells were washed with 1X phosphate-buffered saline (PBS), then treated with 1 mL 1% trypsin-ethylenediaminetetraacetic acid (EDTA) solution (Sigma) to detach cells and sub-cultured to a ratio between 1:10 and 1:20. Detached cells were resuspended in 5 mL complete media. Cell density was determined by diluting 10 µL of 6 mL suspension 1:1 with Trypan Blue Stain (Gibco) and counted using a haemocytometer. For

seeding, a master mix of counted suspension and complete media was prepared and 500 μ L or 1 mL was added to a 24- or 12-well plate, respectively.

2.2.6.4 *In vitro* RNA transcription

pUC57-encoded minigenome or L, PdC, or N support plasmids were linearised through restriction enzyme digest and purified using phenol-chloroform extraction. Briefly, 5 μ g of each linearised plasmid was diluted into 100 μ l nuclease-free water and an equal volume of phenol chloroform was added and mixed by vortex. Following centrifugation for 10 minutes at 13000 rpm and RT in a Micro Star 17R, the aqueous phase was removed. Sodium acetate was then added at a volume of 1/10 that of the aqueous phase and 2.5 volumes of 100% ethanol before incubation at -20 °C for 30 minutes to precipitate the DNA. Following incubation, the precipitated DNA was pelleted for 30 minutes at 13000 rpm and RT in a Micro Star 17R. The supernatant was then removed and the pellet was washed with 70% ethanol before another spin cycle for 30 minutes at 13000 rpm and RT. After removal of the ethanol, the DNA pellet was dried on a 42 °C heat block before resuspension in 20 μ l nuclease-free water. The purified, linearised DNA was then used for T7 RNA polymerase-mediated *in vitro* RNA transcription using the T7 RiboMAX Kit (Promega) and cleaned up using the RNA Clean & Concentrator-25 (Zymo Research).

2.2.6.3 Transfection

DNA and RNA transfections were carried out using Lipofectamine2000 (Thermo Fischer Scientific) following the manufacturer's protocol for either 12-well or 24-well formats, with the following changes. Transfection mixtures of the required DNA plasmids or RNA transcripts were prepared in the appropriate volume of OptiMEM (Thermo Fisher Scientific) and incubated as instructed. Transfection mixtures were incubated and added to previously PBS-washed cells and incubated at 37 °C. Transfection mixtures were optionally exchanged for cell culture media at 6 hours or 24 hours post-transfection as required by the experiment.

2.2.6.4 IncuCyte imaging

To analyse real-time growth kinetics of the HeV minigenome-based replicon system, DNA- or RNA-transfected cells were placed in an IncuCyte ZOOM (Sartorius) and incubated for up to 100 hours. Cells were then scanned hourly or once every two hours for GFP fluorescence.

Report gene (EGFP) expression was recorded as green fluorescence object counts or total green integrated intensity (green calibrated units (GCU) $\times \mu\text{m}^2$ / image) as a measure for replicon activity.

Chapter 3: Recombinant *Hendra henipavirus* N protein expression and purification

3.1 Introduction

The elucidation of protein structures by common approaches such as nuclear magnetic resonance (NMR) spectroscopy, X-ray protein crystallography, and cryo-electron microscopy (cryo-EM), as well as techniques such as mass-spectrometry require isolated protein samples to analyse. When proteins of interest cannot be isolated from their native environment for structural analysis, they may be recombinantly expressed and purified in a variety of expression systems. These systems each come with their own advantages and disadvantages, but ultimately aim to produce a pure sample. In fact, it is often preferable to use pure samples for many of the down-stream applications, ensuring confidence in the identity and composition of the sample.

When virus capsids are the object of structural inquiry, whole virus particles may be used rather than isolating their component proteins. In fact, cryo-EM single-particle analysis (SPA) approaches benefit from samples containing whole virus with symmetrical particles (see Chapter 4 Section 4.1). However, using whole virus may become experimentally or practically unfavourable, forcing the isolation of viral proteins from virions for structural elucidation. In terms of electron microscopy, SPA usually becomes the favoured method owing to its robust implementation and support in various software packages and capability to generate high-resolution 3D reconstructions.

Furthermore, recombinant expression may offer a more practical or safer approach to studying viral proteins. Research on most viruses comes with additional containment requirements, depending on their pathogen category level (Advisory Committee on Dangerous Pathogens (ACDP), 2021), designed to safeguard both the researcher as well as the environment. As a rule of thumb, higher pathogen categories contain increasingly more dangerous viruses and therefore come with more stringent containment requirements. Work with any whole virus from higher categories becomes, as a result, increasingly onerous. Indeed, this is the case for *Hendra henipavirus*, which is classed as a biosafety level (BSL) 4 pathogen (Advisory Committee on Dangerous Pathogens (ACDP), 2021) and can therefore only be directly studied in specialised facilities, such as Porton Down – the only BSL4 laboratory in the United Kingdom. In the absence of the required biosafety level 4 (BSL-4) facilities, recombinant expression and subsequent purification of the HeV N protein is the only practical route to generate samples for structural elucidation through cryo-EM data collection and SPA reconstruction.

Previous work has shown the successful expression and purification of HeV N protein using a bacterial expression system (Pearce et al., 2015). Aimed at generating reagents for immunodiagnostics, Pearce *et al.* expressed both full-length HeV N as well as a C-terminally truncated N protein which were purified through immobilised metal-ion affinity chromatography (IMAC) and size exclusion chromatography (SEC). The HeV N samples were analysed through negative-stain EM showing that recombinant HeV N is capable of self-assembly and resembled HeV nucleocapsids isolated from virions (see Chapter 4) (Pearce et al., 2015).

3.1.1 Recombinant protein expression

Recombinant protein expression offers a number of advantages over protein isolation from their native environment, such as a cell or virion. These are primarily the ability to fuse the protein of interest to functional groups, such as affinity purification tags, epitope tags, or fluorescent tags, the ability to generate large quantities of protein, and the ability to produce highly pure samples. In essence, (heterologous) recombinant protein expression encompasses the synthesis of protein from externally generated nucleic acid in a biological system that may be different from its native system. These expression systems fall broadly into three categories: prokaryote cell systems, eukaryote cell systems, and cell-free systems. The selection of expression system is primarily dependent on the characteristics of the protein of interest (POI), such as size, solubility, and post-translational modifications (PTMs) but is equally dictated by the down-stream requirements. Prokaryote cell systems, such as bacterial cells, are widely used as they offer simple, easily optimisable and scalable culture conditions at a low cost. However, being prokaryotic, these systems generally lack the widespread and often complex PTMs common in eukaryotes, such as the ability to add glycosyl-, phosphoryl-, or lipid groups, or pro-peptide maturation through specific proteolytic activity, to name a few (reviewed in (Macek et al., 2019)). Eukaryotic cell systems are diverse and can cater to a wide range of requirements but are often more complicated, time consuming, and expensive.

The expression system of choice will in turn dictate in what manner the coding sequence for the POI is introduced. For example, bacterial and mammalian cell systems will accept the transformation or transfection, respectively, of plasmid DNA. This plasmid DNA carries an inducible open reading frame (ORF) for the POI for transient expression, and often also encodes for the expression of selectable (resistance) markers to ensure selection for successfully transformed or transfected cells and retention of the plasmid DNA. Insect cells, on the other hand, require the use of a complex recombinant baculovirus vector for the introduction of the POI (Hitchman et al., 2009; Jarvis, 2009). When an expression system has been selected, prior

to introduction of the genetic material, codon optimisation may be required for the nucleic acid encoding the POI. This is because the codons in the POI's native ORF may mismatch with the codon usage bias in the chosen expression system (Sharp et al., 1993; Plotkin and Kudla, 2011; Spencer and Barral, 2012). Additionally, these biases differ between systems and can even differ within systems, through the use of different strains (see below, Rosetta). Moving the expression of a POI from one system to another may therefore require additional codon optimisation steps. Another consideration is the addition of fusion tags in the coding sequence, which may facilitate down-stream applications or purification, such as through chromatographic techniques.

3.1.1.1 pET expression vectors

A broad range of recombinant protein expression vectors exists for use in bacterial expression systems, encompassing different origins of replication (*ori*), promoters, selectable markers, multiple cloning site (MCS) sequences, and fusion tags. The pET range of bacterial recombinant protein expression vectors find common use owing to their high efficiency and yield (Rosano and Ceccarelli, 2014).

In the pET system, protein expression is driven by the highly active and specific bacteriophage T7 RNA polymerase (T7 POL), where the ORF encoding the POI is under control of a T7 POL promoter and a *lac* operator. Expression of the POI under this promoter-operator dyad can be selectively induced by the non-hydrolysable allolactose analogue isopropyl β -D-1-thiogalactopyranoside (IPTG) (Figure 3.1). An essential requirement for this system is the presence of T7 POL within the host cell. A vector encoding T7 POL can be transformed in concert with the pET vector carrying the POI, however strains of *E. coli* that carry a genomic prophage (λ DE3) encoding an IPTG-inducible T7 POL gene are more commonly used (Rosano and Ceccarelli, 2014).

The *lac* repressor (LacI) is continuously expressed from the *lacI* gene in both the host genome as well as the pET vector. In the absence of allolactose or IPTG, LacI binds to the *lac* operator in the DE3 segment thereby preventing the host RNA polymerase from transcribing the T7 POL gene and halting T7 POL expression. Additionally, LacI binds to the *lac* operator on the pET vector, preventing transcription of the POI ORF (Figure 3.1A). In the presence of allolactose or IPTG, LacI binding to the genomic *lac* operator as well as the *lac* operator found of the pET vector is abrogated and T7 RNA polymerase is expressed. In turn, the T7 POL has access to the ORF coding for the POI on the pET vector allowing expression of the POI (Figure 3.1B) (Rosano and Ceccarelli, 2014).

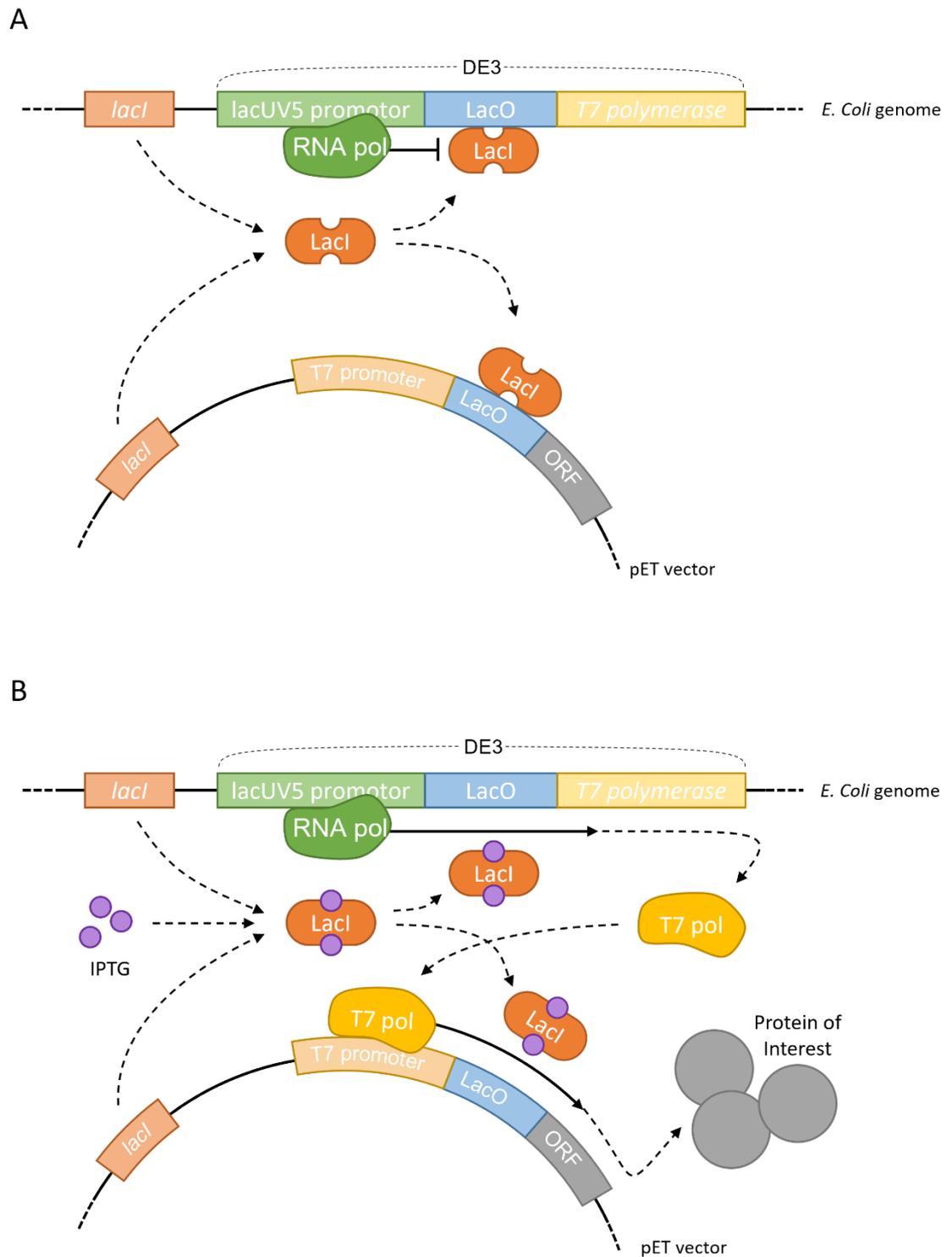


Figure 3.1. Schematic of the pET expression system.

A simplified schematic of the operation of the pET expression system within DE3 carrying *E. coli*. The *lac* repressor (LacI) is continuously expressed from ORFs in both the host genome as well as the pET vector. **A**) In the absence of IPTG, LacI interacts with the *lac* operator (LacO) on both the host genome as well as the pET vector, abrogating transcription of down-stream ORFs. **B**) In the presence of IPTG, LacI repression is lifted allowing the host RNA polymerase (RNA pol) access to T7 RNA polymerase gene (*T7 polymerase*), driving T7 RNA polymerase (T7 pol) expression. In turn, the T7 RNA polymerase drives the transcription and subsequent expression of the POI from the ORF found on the pET vector.

3.1.2 Principles of chromatography

Chromatography is the method by which components within a fluid mixture can be separated, based on the inherent properties of those components. In the broadest sense, the fluid mixture, which can be gaseous or liquid and is called the mobile phase, is separated over a system on which a solid stationary phase is fixed. Differential interactions between the components of mixture in the mobile phase with the stationary phase, based on their inherent properties, allows for the separation of the components in the mobile phase. While a plethora of chromatographic techniques exists, employing diverse ways of separating even more numerous different mixtures, they broadly fall into two categories: analytical and preparative. The aim of analytical chromatography is the separation of analytes in a mixture and determine their relative quantities while preparative chromatography is employed to isolate and purify quantities of a specific substance from a mixture. Both types are used in recombinant protein expression. Chromatographic methods allow the purification of proteins or protein complexes of interest from an extracellular media or cell lysate solution. This process selectively separates the POI, either through retention or differential migration, whilst removing any unwanted mono- or multimeric species, chemical and protein contaminants, or other impurities. The two methods employed during this work are affinity chromatography and size exclusion chromatography.

Affinity chromatography (AC) depends on the differential interaction between components, such as the POI and contaminant proteins, with the stationary phase. In many cases, this involves a so-called affinity fusion tag which interacts with the stationary phase in a non-covalent, but strong and selective manner. Some often-used examples in molecular biology include the biotin/streptavidin (Cull and Schatz, 2000; Chodosh, 2001) and glutathione S-transferase/glutathione pairs (Surtees et al., 2015). In this work, immobilised metal-ion affinity chromatography (IMAC) is used which employs the poly-histidine/ Ni^{2+} affinity pair. In short, the POI is fused to a poly-histidine tag, a hexahistidine (six histidine residues) in the case of the work presented here (Figure 3.2A). The stationary phase contains Ni^{2+} ions with which the imidazole moiety of histidine interacts, thereby retaining them within the column. The unwanted proteins and contaminants are then washed away before the POI is eluted with soluble imidazole in a concentration dependent manner. Soluble imidazole competes for binding with the histidine's imidazole moiety (Figure 3.2B) to the Ni^{2+} ions in the stationary phase, thereby releasing the histidine-tagged POI. Subsequently, the poly-histidine tag can be proteolytically removed, facilitated by the SUMO (smt3) domain that sits between the poly-histidine and the POI in these constructs (Figure 3.2A, Supplementary Figure S3.1). Proteolytic cleavage by SUMO protease

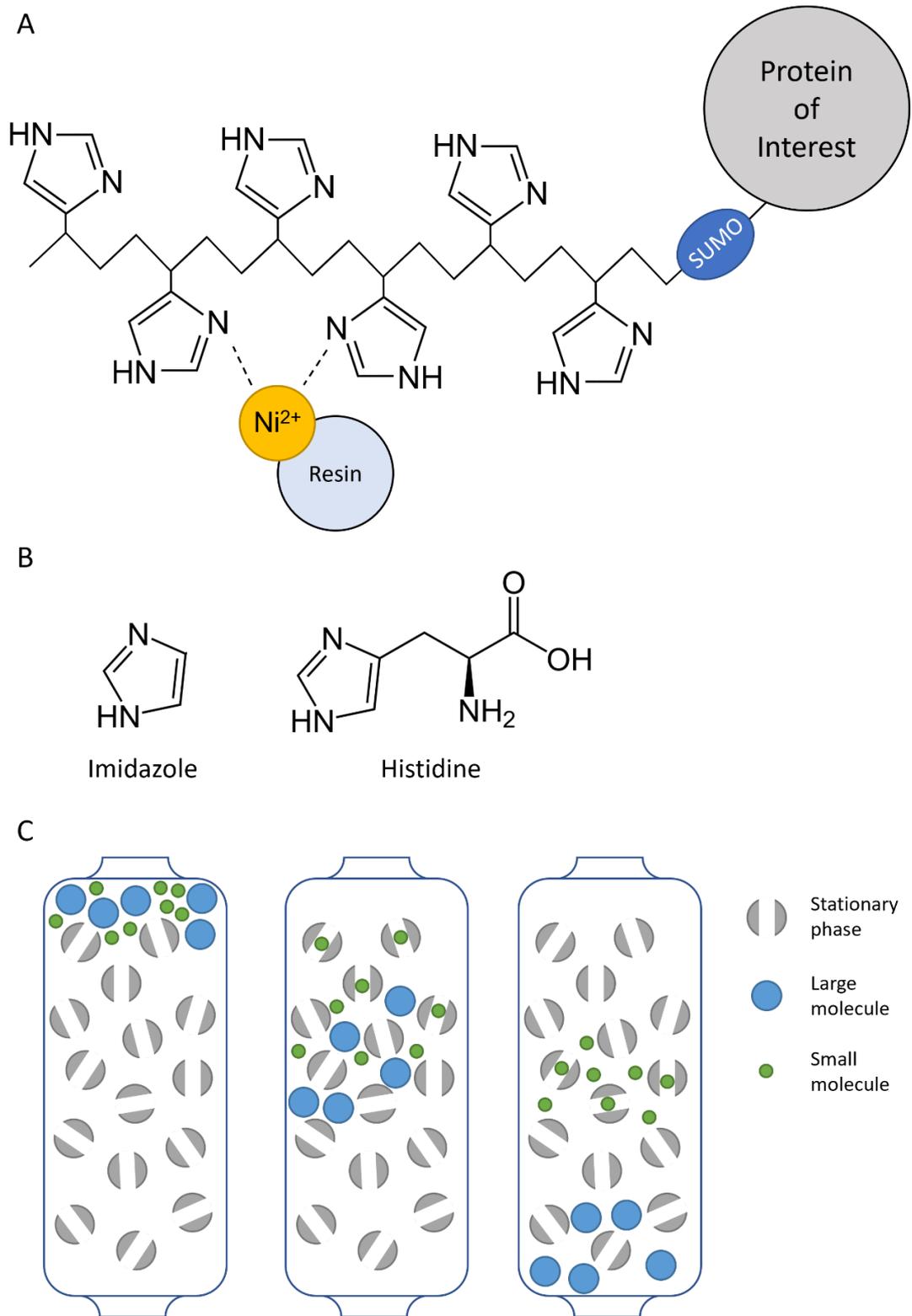


Figure 3.2. Principles of chromatography.

A) A schematic representation of a protein of interest (grey) fused to a hexahistidine SUMO tag, which interacts with resin-immobilised Ni^{2+} ions (yellow). Schematic is not to scale. **B)** Parallel display of the two-dimensional structural formulae of imidazole and histidine. Note the equivalent pentacyclic diazole form of imidazole and the histidine specific sidechain. **C)** Schematic representation of size exclusion chromatography. The direction of flow of the mobile phase is from top to bottom, with the early, middle, and late time points depicted side by side. Large molecules (blue) pass through the column faster than small molecules (green), since the latter interacts more extensively with the stationary phase (grey).

(ULP1) leaves no additional residues on the POI, reconstituting the native protein (Butt et al., 2005).

Size exclusion chromatography (SEC) separates proteins based on size. Larger molecules have a shorter retention time within the mesh of the stationary gel phase than smaller molecules (Figure 3.2C). The stationary phase consists of miniscule beads with even smaller pores. These pores trap small molecules whilst larger ones pass by, thereby retaining smaller molecules within the stationary phase for longer than larger molecules. When preceded by AC, SEC allows for the separation of multimeric species consisting of different numbers of copies of the same protein and analysis of oligomeric states by comparison to a standard curve.

3.1.3 Chapter aims

This chapter aimed to optimise and carry out the expression and purification of recombinant HeV N protein and HeV nucleocapsid-like complexes to be used for down-stream structural analysis. The protocols used to generate the results in this section are covered in Chapter 2 Materials and Methods, Sections 2.2.1 and 2.2.2.

3.2 Results

3.2.1 Recombinant HeV N can be successfully expressed in *E. coli* BL21 (DE3) Rosetta2

To recombinantly express the HeV N protein, the codon-optimised N gene was subcloned into a pET28(a)-6His-SUMO expression vector, which encodes HeV N with an N-terminally fused, sequential hexahistidine (6His) affinity tag and SUMO fusion protein (Supplementary Figure S3.1). One of the primary advantages of this vector is that following expression and purification using the hexahistidine tag, SUMO protease treatment generates a native, non-tagged HeV N protein (Butt et al., 2005). This method has been successfully used for the expression, purification, and ultimately structural elucidation of nucleoproteins of other (-)ssRNA viruses, including *Rift Valley fever phlebovirus*, *Crimean-Congo hemorrhagic fever orthonairovirus*, and *Schmallenberg orthobunyavirus* (Raymond et al., 2010; Carter, Barr, et al., 2012; Ariza et al., 2013).

Before large-scale expression could be performed, small-scale trials needed to be carried out in order to optimise expression parameters. The pET28(a)-6His-SUMO-HeV-N plasmid was first heat-shock transformed into *E. coli* BL21 (DE3) Rosetta 2, BL21 (DE3) Star, and BL21 (DE3) Gold strains in order to assess the optimal expression strain. These strains are all *E.*

coli BL21 (DE3) derived, thus including the λ prophage DE3 which allows for IPTG-inducible T7 POL-driven recombinant protein expression. BL21 strains are particularly suited for recombinant protein expression since they are deficient in the Lon and OmpT proteases. The three strains used here differ in some important aspects. The Rosetta 2 strain carried the pRARE plasmid which encodes for tRNAs specific to codons that are rare in *E. coli*, improving expression of eukaryotic proteins. The Star strain is deficient in RNase E activity, improving mRNA stability. Lastly, the Gold strain has improved transformation efficiency and allows for DNA purification.

The expression protocol briefly encompassed the following steps. Nutrient rich 2xYT media was inoculated with an overnight pre-culture. This inoculum was grown to an OD_{600} between 0.6 and 0.8, and then induced with isopropyl β -D-1-thiogalactopyranoside (IPTG) to express protein overnight at 18°C. Cells were then collected and lysed through three freeze-thaw cycles, after which the cell lysates were centrifuged to separate the soluble and insoluble (cell debris) fractions. Samples were taken pre- and post-induction, and of the supernatant and pellet of the lysed cells, which were used to assess expression using coomassie-stained SDS-PAGE (Figure 3.3). A strong band was identified in the BL21 (DE3) Rosetta 2 samples, but not the Star or Gold strains, at just under 75 kDa (Figure 3.3, black arrow). This value is consistent with the predicted molecular weight of 71.97 kDa (Supplementary Figure S3.2A) for the hexahistidine SUMO HeV N fusion protein. The lack of protein over-expression at the expected molecular weight for hexahistidine SUMO HeV N fusion protein in the Star and Gold strains suggests that they were unsuitable, however the strong band in all expected samples from the Rosetta 2 cells (induced, supernatant, and pellet) allowed this strain to be taken forward.

3.2.2 Chromatography can be utilised to purify HeV N protein

After expression and lysis of the bacterial cells, the hexahistidine fusion tag was used as a means of purifying HeV N through immobilised metal affinity chromatography (IMAC). The histidine's side chain imidazole moiety interacts with Ni^{2+} ions in a non-covalent manner, which allows for the selective retention of hexahistidine tagged POIs in a Ni^{2+} containing stationary phase.

Following selection of the *E. coli* BL21 (DE3) Rosetta 2 expression strain, a small-scale expression was performed in order to trial affinity chromatography for protein purification (Figure 3.4). After protein expression and cell lysis, the soluble fraction ('supernatant') was passed through a primary HisTrap column, which employs a Ni Sepharose resin as the stationary phase. Non-bound proteins are removed in the flow-through (FT), however the POIs are retained

in the column. After binding, the column was washed with increasing concentrations of soluble imidazole, displacing the hexahistidine SUMO HeV N fusion protein at 300 mM and 500 mM imidazole which was then eluted. The presence of a protein of expected size on coomassie stained SDS-PAGE in the primary HisTrap eluate confirmed selective retention and elution of hexahistidine SUMO HeV N fusion protein (Figure 3.4A). After freeze-thaw lysis, a substantial portion of the expressed hexahistidine SUMO HeV N fusion protein was found in the insoluble fraction (Figure 3.4A, 'pellet'), as was similarly observed in the expression trials (see Figure 3.3). This insoluble portion of hexahistidine SUMO HeV N fusion protein was therefore lost when the soluble fraction ('supernatant') was taken forward for affinity chromatography.

Despite the low yield, the 300 mM and 500 mM imidazole fractions were taken forward and firstly dialysed to reduce the Tris-HCl, NaCl, and imidazole concentrations. Concurrently, the hexahistidine-SUMO fusion tag was proteolytically removed from the N protein by addition of

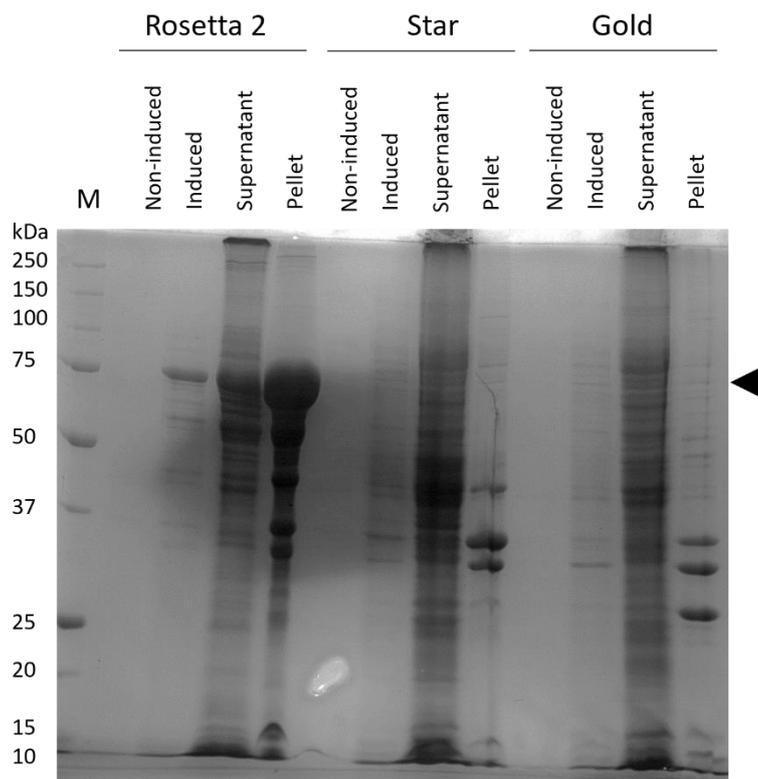


Figure 3.3. HeV N can be successfully expressed in Rosetta 2 cells.

pET28(a)-6His-SUMO-HeV-N plasmid was transformed into a range of *E. Coli* BL21 (DE3) cell lines; Rosetta 2, Star and Gold. Inoculi were incubated to mid-log phase at which point hexahistidine SUMO HeV N fusion protein expression was induced with IPTG at 18°C overnight. Inoculi were then pelleted and lysed through freeze-thaw. Lysates were clarified through centrifugation. Samples taken at each stage were resolved by SDS-PAGE alongside a protein ladder (M) and visualised using coomassie gel stain. The molecular weight of each marker is indicated (kDa) and the black arrowhead indicates the predicted location of the hexahistidine SUMO HeV N fusion protein (~72 kDa).

SUMO protease, resulting in a tag-free, native HeV N protein predicted at 58.48 kDa (Supplementary Figure S3.2B). After dialysis and protease treatment, fractions containing soluble protein were pooled, filtered, and passed over a second HisTrap column. In this second application, tag-free HeV N protein flows through the HisTrap column uninterrupted and is observed in the flow-through at the correct size (Figure 3.4B, chequered arrowhead). This shift in the protein size from around 72 kDa to around 59 kDa confirms proteolytic cleavage of the hexahistidine-SUMO fusion tag. A proportion of the cleaved hexahistidine-SUMO fusion tag however (13.5 kDa), as well as the hexahistidine-tagged SUMO protease (27 kDa), are retained within the secondary HisTrap and are eluted with soluble imidazole at 500 mM. This secondary column clarified the post-dialysis mixture resulted in untagged, native HeV N protein.

NanoDrop results from the affinity purifications showed 260/280 ratio of 0.95 and higher. A pure protein sample usually has a 260/280 ratio around 0.6, thus the obtained results indicate the presence of nucleic acid in the purified HeV N protein sample, despite the addition of DNase I and RNase A in the lysis buffer. Since HeV N is an RNA binding protein, the nucleic

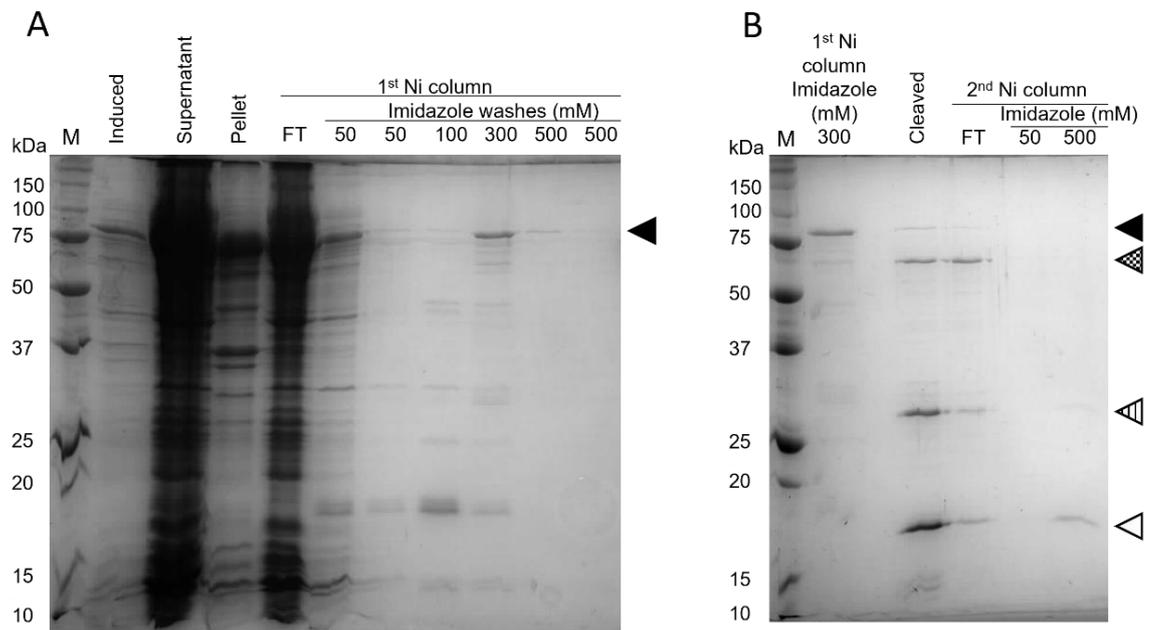


Figure 3.4. HeV N can be expressed and purified through affinity chromatography in a small-scale format.

Coomassie stained SDS-PAGE of expression and purification samples. **A)** Fractions from expression and primary HisTrap column show a large portion of insoluble HeV N protein (pellet). The soluble HeV N fraction was retained well in the HisTrap column and eluted, as expected, at 300 mM imidazole. **B)** Fractions from secondary HisTrap column, post SUMO protease treatment of HeV N. Bands indicate successful proteolytic cleavage of the hexahistidine SUMO HeV N fusion protein. Black arrowheads indicate hexahistidine SUMO HeV N fusion protein. Chequered arrowheads indicate native HeV N protein. White arrowheads indicate the post-cleavage hexahistidine-SUMO tag. Striped arrowheads indicate hexahistidine-tagged SUMO protease (27 kDa).

acid signal could represent RNA that is shielded by HeV N from degradation by RNase A. No viral RNA was added to the sample, therefore it is likely that the RNA is host-derived and may be of random length, with concomitantly random-length multimeric HeV N species. Thus, SEC was employed to separate the putative multimeric species based on size.

The flow-through fractions from the secondary HisTrap purification were concentrated and separated over a Superdex S200 column (Figure 3.5). By continuously measuring the UV absorbance of the sample flowing out of the column, a major narrow peak was observed closely following the void volume (~100 ml), indicating high molecular weight species (Figure 3.5A). The flow through was collected in 2.5 ml fractions and the presence of HeV N was confirmed via SDS-PAGE analysis (Figure 3.5B). The dotted box indicates fraction 13; the fraction which had the highest UV absorbance in SEC and the strongest band on SDS-PAGE. The protein concentration for fraction 13 was measured at 45 µg/ml. In this instance, the Superdex S200 column failed to separate species of HeV N indicating that under the current conditions HeV N does not occur as monomeric or dimeric species but rather HeV N exists as larger multimeric species. Whether these exist as discrete multimeric species or as a range of species of different multimericity is inconclusive from these results, as either type exceeds the resolution of the S200 column. Although the methodology of expression and purification by affinity and size exclusion chromatography was deemed successful, optimisation was clearly required to obtain enough native HeV N for downstream purposes.

3.2.3 Sonication improves soluble HeV N protein yield

The small-scale expression confirmed that the procedure could be utilised to purify native HeV N protein, however a significant proportion of the protein was lost in the insoluble fraction. In an attempt to increase the yield of soluble protein, the expression and purification protocols were subsequently amended with inspiration from the described recombinant expression of HeV N in Pearce *et al.*, 2015 (Pearce et al., 2015). A large-scale expression (6 litres) was performed and cultures were centrifuged as before, providing one aliquot per litre of media. Two aliquots of bacterial cells were lysed through a single freeze-thaw cycle and a sonication step was added in an attempt to improve soluble HeV N protein yield (see Chapter 2 Section 2.2.2). After affinity chromatography (as described above), SDS-PAGE showed that the addition of sonication significantly increased the yield of soluble hexahistidine SUMO HeV N fusion protein to satisfactory levels, indicated by the large protein band at ~72 kDa in the supernatant and 300 mM and 500 mM imidazole eluates (Figure 3.6A). Approximately 24 mg of tag-free,

native HeV N protein were subsequently purified from the second HisTrap column (Figure 3.6B, chequered arrowhead), and the hexahistidine-SUMO fusion tag and SUMO protease were observed at their appropriate sizes in the eluted fractions. Although some HeV N protein was not successfully cleaved and additionally some was not eluted until the addition of imidazole, the majority of HeV N was found in the flow-through and was taken forward for SEC.

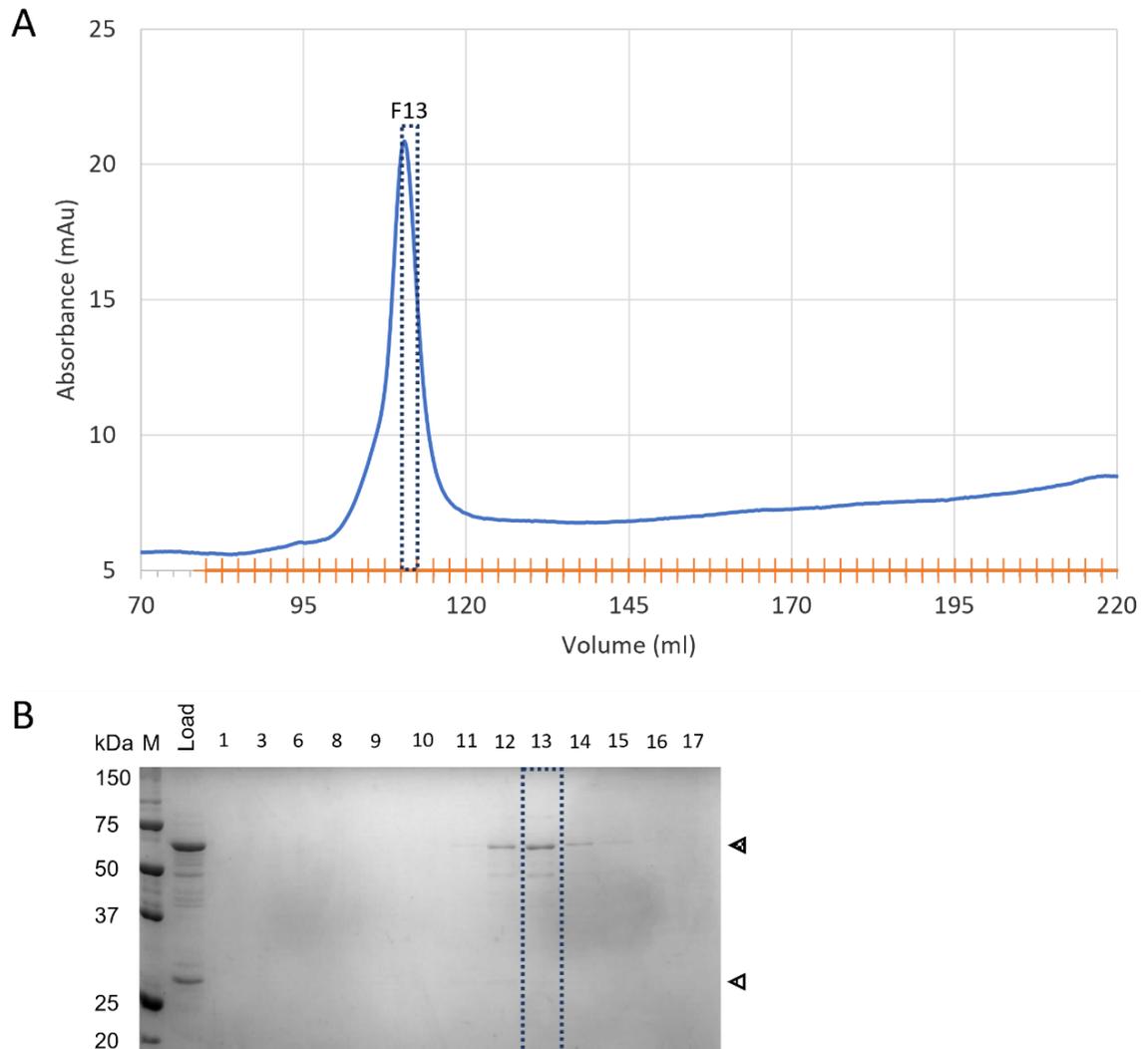


Figure 3.5. HeV N multimers do not separate into discrete species on Superdex S200 size exclusion chromatography.

A) Size exclusion chromatography (SEC) UV trace from a Sepharose S200 26/600 column. Continuous UV absorbance readings are plotted in blue. Automatically collected 2.5 ml fractions are indicated by the orange hashed line. Fraction 13 (F13) is outlined. **B)** Coomassie stained SDS-PAGE showing the presence of HeV N protein at around 59 kDa. The lane marked Load represents the sample loaded into the SEC column for reference. The numbered lanes represent SEC fractions loaded on gel and correspond to those in A. Chequered arrowheads indicate native HeV N protein (59 kDa). Striped arrowheads indicate hexahistidine-tagged SUMO protease (27 kDa).

NanoDrop results of the AC purifications again indicated the presence of nucleic acid and SEC was employed to attempt the separation of discrete multimeric species. Since the HeV N protein closely followed the void volume on the Superdex S200 column, in this instance a Sephacryl S400 column was used. As before, the flow-through was collected and analysed by measuring UV absorbance and by SDS-PAGE (Figure 3.7). A minor broad peak was observed closely following the void volume (~100 ml), indicating high molecular weight species, closely followed by a major broad peak, indicating lower molecular weight species (Figure 3.7A). The major broad peak contained the highest concentration of protein (Figure 3.7A), which was confirmed as HeV N protein by SDS-PAGE (Figure 3.7B). The broad peak of HeV N suggests that HeV N multimeric species exist in a wide range of oligomeric forms, resembling a normal distribution that would extend beyond the void volume, resulting in a minor peak or shoulder. For the purposes of cryo-EM, buffer conditions should resemble pure water as closely as possible to improve ice quality and concomitantly image quality. This required a reduction in salt concentration, which may lead to altered purification results for HeV N protein.

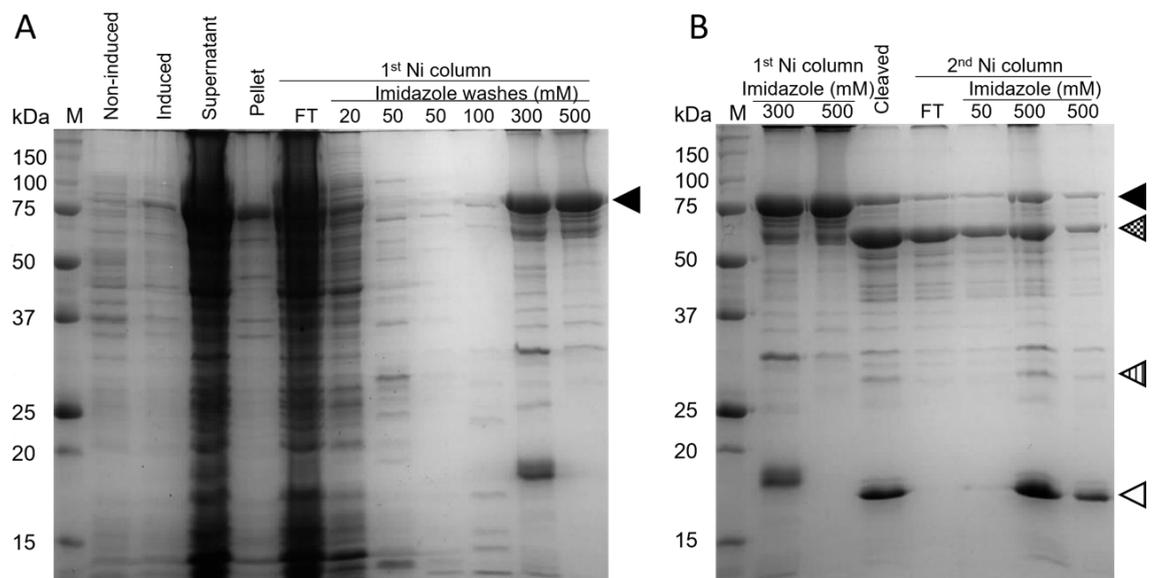


Figure 3.6. Large-scale HeV N expression and first purification.

Coomassie stained SDS-PAGE of expression and purification samples. **A)** Fractions from expression and primary HisTrap column show a large portion of soluble HeV N protein (supernatant). The soluble HeV N was retained well in the HisTrap column and eluted, as expected, at 300 mM and 500 mM imidazole. **B)** Fractions from secondary HisTrap column, post SUMO protease treatment of HeV N. Bands indicate successful proteolytic cleavage of the hexahistidine SUMO HeV N fusion protein. Black arrowheads indicate hexahistidine SUMO HeV N fusion protein. Chequered arrowheads indicate native HeV N protein. White arrowheads indicate the post-cleavage hexahistidine-SUMO tag. Striped arrowheads indicate hexahistidine-tagged SUMO protease (27 kDa).

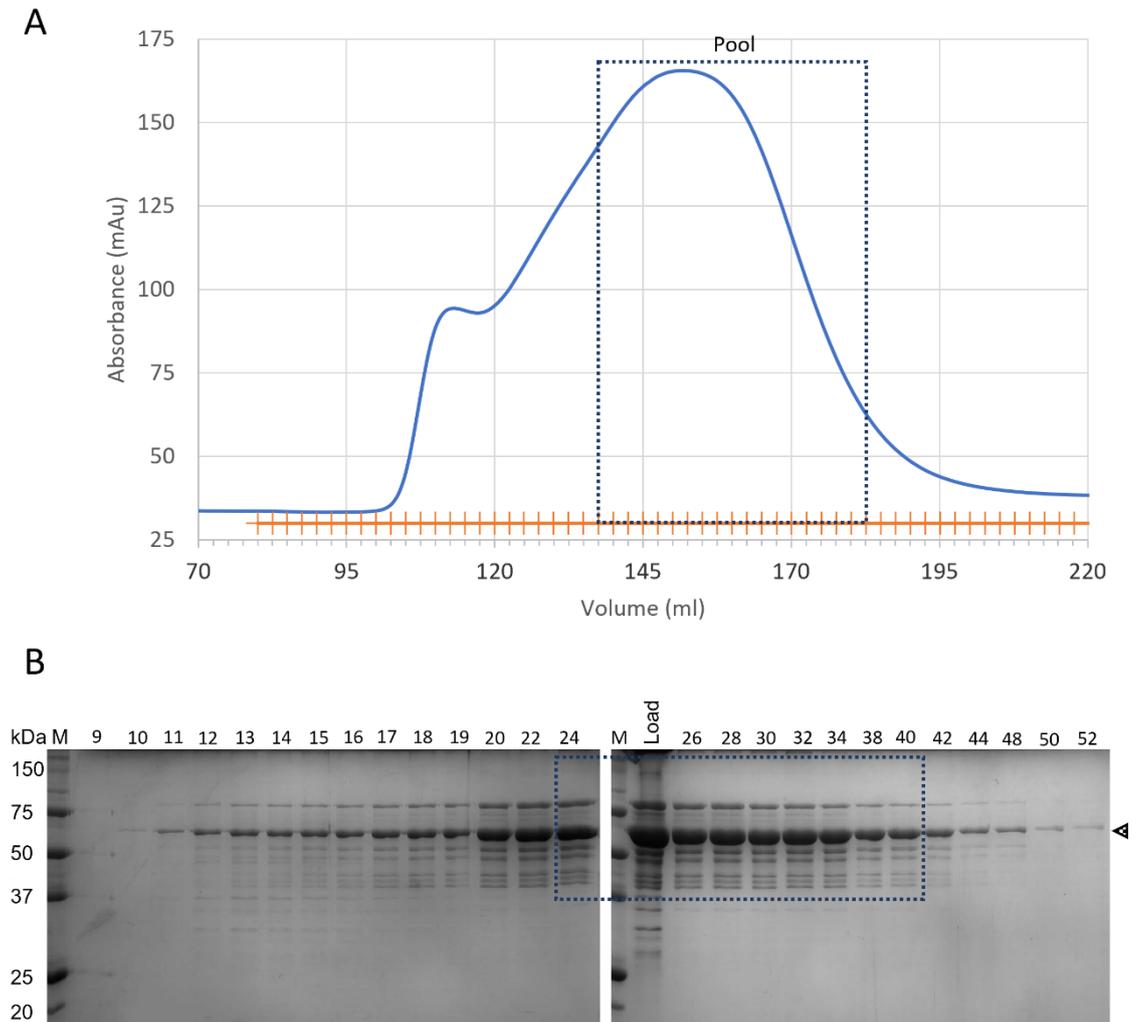


Figure 3.7. Large-scale HeV N expression and first purification.

A) Size exclusion chromatography (SEC) UV trace from a Sephacryl S400 26/600 column. Continuous UV absorbance readings are plotted in blue. Automatically collected 2.5 ml fractions are indicated by the orange hashed line. Pooled fractions (Pool) are outlined. **B)** Coomassie stained SDS-PAGE showing the presence of HeV N protein. The lane marked Load represents the sample loaded into the SEC column for reference. The numbered lanes represent SEC fractions loaded on gel and correspond to those in A. Chequered arrowheads indicate native HeV N protein.

3.2.4 Decreasing NaCl concentration increases the presence of high molecular weight species

For the two purifications detailed above, a concentration of 500 mM NaCl was used in both the dialysis buffer and the SEC buffer (see Section 3.2.4). For the second large scale purification (Figures 3.8 and 3.9), this was reduced to 200 mM NaCl in the dialysis buffer and 100 mM NaCl in the SEC buffer, so that the sample conditions are closer to pure water, improving sample suitability for cryo-EM.

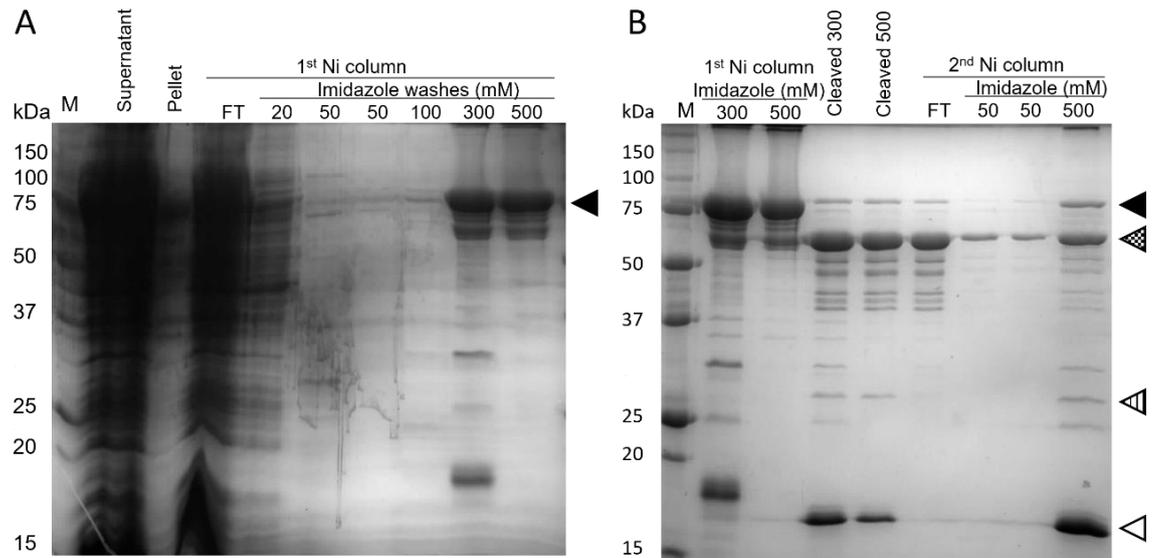


Figure 3.8. Large-scale HeV N expression and second purification.

Coomassie stained SDS-PAGE of expression and purification samples. **A)** Fractions from expression and primary HisTrap column show a large portion of soluble HeV N protein (supernatant). The soluble HeV N was retained well in the HisTrap column and eluted, as expected, at 300 mM and 500 mM imidazole. **B)** Fractions from secondary HisTrap column, post SUMO protease treatment of HeV N. Bands indicate successful proteolytic cleavage of the 6His-SUMO HeV N fusion protein. Black arrowheads indicate 6His-SUMO HeV N fusion protein. Chequered arrowheads indicate native HeV N protein. White arrowheads indicate the post-cleavage 6His-SUMO tag. Striped arrowheads indicate 6His-tagged SUMO protease (27 kDa).

After the large-scale protein expression described above, two aliquots of bacterial cells were used lysed as before. HeV N was purified by affinity chromatography as above, dialysed to 200 mM NaCl while treated with SUMO protease as before (Figure 3.8). The flow-through fractions from the secondary HisTrap purification yielded around 24 mg of HeV N protein, which was concentrated to 3 mg/ml and separated over a Sephacryl S400 column (Figure 3.9) as previously. A reduction in NaCl concentration in the SEC buffer resulted in a shift towards higher molecular weight species as shown by the large peak in the early fractions directly following the void volume and trailing out towards lower molecular weight volumes (Figure 3.9A). This shift towards a higher molecular weight indicates that longer HeV N multimers have been purified. The smaller, wide peak in later fractions (29-37) indicates a population of smaller multimers has also been collected. The presence of HeV N was confirmed via SDS-PAGE analysis as before (Figure 3.9B) and measured at 0.62 mg/ml and 47 μ g/ml for fractions 12 and 33, respectively.

The presence of RNA in these samples was confirmed via NanoDrop, with 260/280 ratios of 1.15 and 1.28 for fractions 12 and 33 respectively (Figure 3.9). The SEC profiles combined with SDS-PAGE gels and high 260/280 ratios suggest that under the present conditions, HeV N forms

large, non-covalently linked multimers bound to RNA, hereafter referred to as nucleocapsid-like filaments (NLFs).

The purification of these higher molecular weight species was favourable for downstream purposes and therefore individual or pooled SEC fractions of interest were collected as indicated by dashed boxes (Figure 3.9). Using a centrifugal concentrator, samples were concentrated to 1.425 mg/ml (F12) and 2.033 mg/ml (Pool) and subsequently stored at -80°C or used for electron microscopy approaches (see Chapter 4).

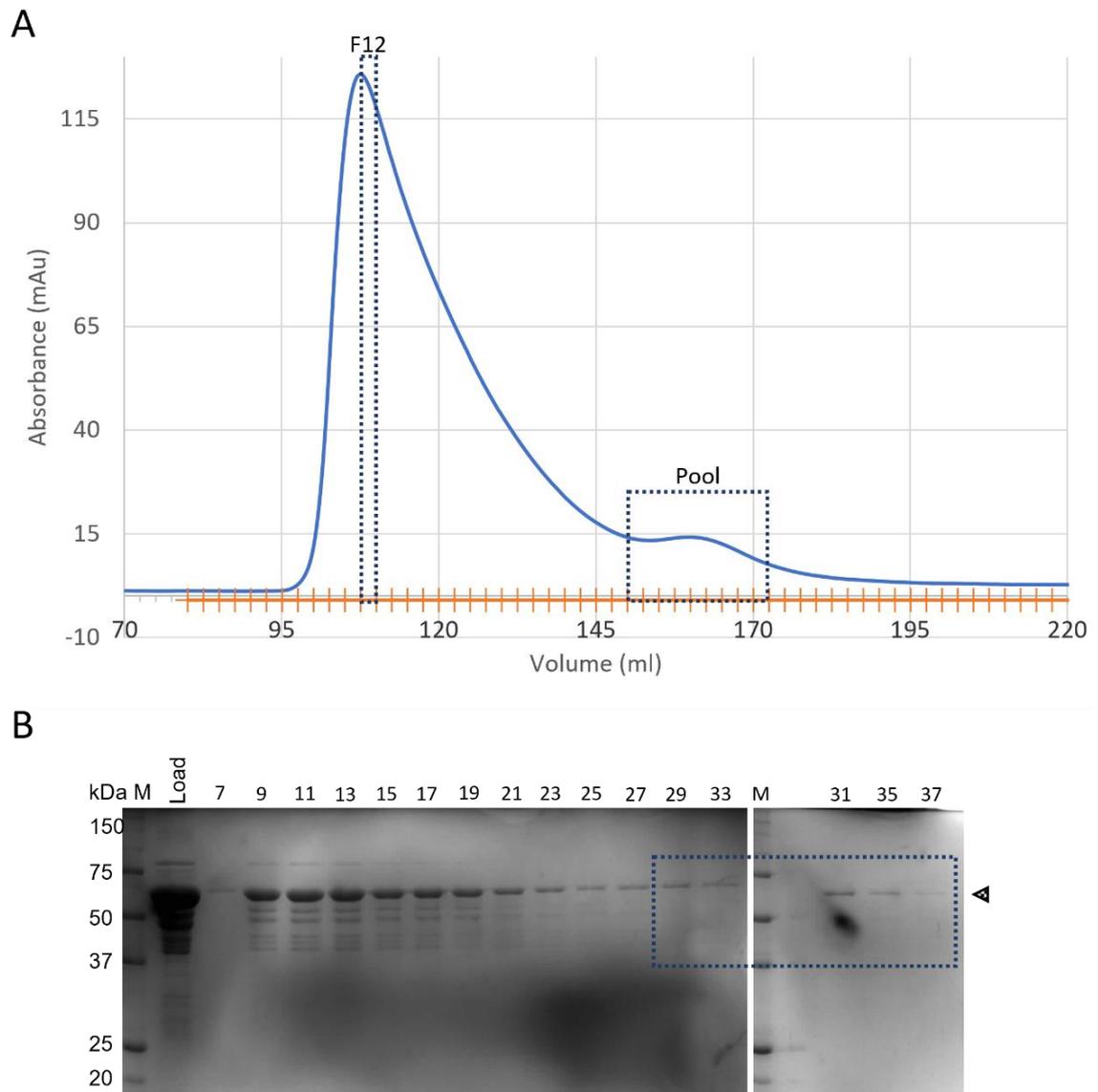


Figure 3.9. Large-scale HeV N expression and first purification.

A) Size exclusion chromatography (SEC) UV trace from a Sephacryl S400 26/600 column. Continuous UV absorbance readings are plotted in blue. Automatically collected 2.5 ml fractions are indicated by the orange hashed line. Fraction 12 (F12) and Pooled fractions (Pool) are outlined. **B)** Coomassie stained SDS-PAGE showing the presence of HeV N protein. The lane marked Load represents the sample loaded into the SEC column for reference. The numbered lanes represent SEC fractions loaded on gel and correspond to those in A. Chequered arrowheads indicate native HeV N protein.

3.3 Discussion

Most recombinant expression approaches require optimisation, and the HeV N protein is no exception. In some cases, the need to trial a wide range of conditions can be mitigated through literature research. Indeed, the use of the pET plasmid in combination with *E. coli* BL21 (DE3) strains is a tried and tested approach for recombinant expression of (viral) proteins in the Barr and Edwards research groups (Carter, Barr, et al., 2012; Carter, Surtees, et al., 2012; Surtees et al., 2015; Álvarez-Rodríguez et al., 2020; Hoste et al., 2021). The local knowledge and experience using this system, as well as the availability of necessary hardware, provided a good basis for the recombinant expression of HeV N.

Recombinant HeV N has previously been expressed in a similar system (Pearce et al., 2015), using *E. coli* expression strains and IMAC and SEC for purification. The published work provided inspiration for incubation and expression parameters such as temperature, time, and concentration of induction agent. This information not only simplified the fine-tuning of expression and purification within these systems, but it also provided a benchmark of successful expression and purification to compare the obtained results against. Important differences between the approach taken by Pearce *et al.* and the work presented here should be highlighted, however. While the published work produced both full length and C-terminally truncated HeV N protein, the work in this chapter concerned only full length HeV N. Additionally, the published study employed the *E. coli* BL21 AI strain, which is optimised for toxic protein expression. Furthermore, Pearce *et al.* used a pRSET7 C plasmid which, under T7 control, expressed the POI N-terminally fused to a hexahistidine tag and enterokinase-cleavable epitope tag analogous to the hexahistidine-SUMO-tag employed in this chapter. Finally, bacterial cell lysis buffer further contained a serine protease inhibitor in addition to the lysis buffer described here, and used benzonase rather than DNase I/RNase A to digest nucleic acids (Pearce et al., 2015).

3.3.1 Recombinant expression of paramyxoviral N proteins for 3D reconstruction

The strategy of recombinant protein expression in bacterial hosts and tandem purification through histidine fusion tags and size exclusion chromatography has similarly been employed for other, related viruses including *Mammalian orthorubulavirus 5* (formerly human parainfluenza virus 5; PIV5) (Alayyoubi et al., 2015), MeV (Desfosses et al., 2019), NiV (Ker et al., 2021), and SeV (Zhang et al., 2021) and NDV (Song et al., 2019). While mostly helical and resembling the nucleocapsids isolated from virions, in each of these studies a range of N protein

multimer lengths was observed, likely reflecting the inclusion of host-derived RNA into the nucleocapsid-like assemblies. Additional morphologies were also observed in these recombinantly expressed nucleocapsids, most strikingly the double-headed or clamshell shaped nucleocapsid boundaries (Song et al., 2019; Zhang et al., 2021; Ker et al., 2021).

Furthermore, all of these studies purified the respective N proteins through N-terminally fused poly-histidine affinity tags, with the exception of MeV which used a C-terminal poly-histidine tag instead (Desfosses et al., 2019). Excluding PIV5, none of the strategies described in these publications included an RNase treatment step at any point during the purification process. Additionally, no attempt was made to remove the poly-histidine fusion tag for any of the published structures. The approach taken here for HeV N protein ensures that no affinity purification fusion tag remains attached to the protein at the end of the purification pipeline, generating a native, tag-free protein. Chapter 5 further discusses the potential consequences of the inclusion of the N-terminal purification tag on the N protein structures and their interpretation for these viruses.

3.3.2 Purity and stability of recombinant proteins

Following expression and purification, protein samples may be used for a number of down-stream applications. In virtually all of these cases, two aspects need to be kept in mind, namely sample purity and stability. Purity is arguably the most important consideration. Sample impurities, ranging from small molecules, to nucleic acid, and peptides may interfere with antibody generation, biophysics assays, crystal formation, or electron microscopy imaging to name a few. The results show successful expression and high purity of HeV N samples.

Sample stability is another key aspect, as it will dictate what down-stream applications are not only suitable but also feasible. Additional protein bands were observed in SDS-PAGE gels for the small-scale and large-scale purifications (Figures 3.4 to 3.9). These additional bands were both larger and smaller than the predicted molecular weight for native HeV N at around 59 kDa. The lower molecular weight bands were similarly observed by Pearce *et al.* 2015 who confirmed some of them to be C-terminally truncated HeV N protomers via western blot analysis (Pearce et al., 2015). Indeed, the HeV N protein termini are predicted to consist of flexible arms which may be more susceptible to proteolytic degradation in *E. coli* hosts (Pearce et al., 2015). If these truncated protomers are incapable of multimerisation into NLFs, they should pose no detriment to cryo-EM driven structural elucidation. If they retain their capacity for multimerisation, we can expect a reduction in local resolution for the flexible arms regions in addition to any expected

reduction in resolution due to their flexible nature itself. The higher molecular weight bands are only visible following dialysis and SUMO protease treatment (Figures 3.4B, 3.6B, 3.7B, 3.8B, and 3.9B). These bands likely represent incomplete proteolysis by SUMO protease, as their molecular weight matches that of hexahistidine SUMO HeV N fusion protein. In fact, incomplete proteolysis would also explain the retention of tagged and non-tagged HeV N in the secondary HisTrap and subsequent elution in increasing free imidazole concentrations. In these cases, a small portion of hexahistidine SUMO-tagged HeV N has multimerised with non-tagged HeV N. Both are retained in the secondary HisTrap, however elute when outcompeted for binding by free imidazole. The denaturing conditions of SDS-PAGE then separates these two forms of HeV out into two bands, plus any C-terminal truncations as delineated earlier.

As a rule of thumb, keeping proteins refrigerated can extend shelf life. While freezing of samples is a viable long-term storage solution, especially when glycerol is added to the sample buffer as a cryoprotectant (Gekko and Timasheff, 1981), repeated freeze-thaw cycles may expedite protein degradation. The intended down-stream use of the recombinant HeV N protein, EM, is more tolerant of nucleic acid and peptide contaminants than most other techniques. Yet, specifically for cryo-EM, the sample buffer should be as close to pure water as possible to facilitate high-resolution reconstructions. This precludes the addition glycerol for storage, in contrast to protein crystallisation where glycerol is often a key ingredient in crystallisation conditions (Vera et al., 2011). Additionally, this requirement limits sample aliquots to a single thawing after flash-freezing in liquid nitrogen (LN) and storage at -80°C , to reduce potential protein degradation through freeze-thaw. In fact, small volume aliquots are preferred as these will more readily flash-freeze whereas larger volumes risk the formation of crystalline ice. Compared to other structural elucidation techniques, EM generally requires small sample volumes of below $5\ \mu\text{l}$ at relatively low concentrations, yet the storage and re-freezing limitations necessitate deceptively large quantities of sample aliquots. Additionally, HeV NLFs stored above cryogenic temperatures for up to eight weeks did not result in extensive protein degradation, suggesting HeV NLFs as well as HeV N protein itself are quite stable.

3.3.2 Summary

This Chapter aimed to describe the optimisation of recombinant expression and purification of HeV N protein and HeV NLFs for down-stream structural analysis. The methodology and results presented here show that approximately 12 mg HeV N protein per litre

culture media can be purified by affinity chromatography. HeV NLFs of relatively high purity were successfully obtained from SEC, at around 0.625 mg/ml for the most concentrated fraction.

Chapter 4: Transmission electron microscopy of *Hendra henipavirus* nucleocapsid-like structures

4.1 Introduction

Historically, the high-resolution structural elucidation of macromolecular protein complexes was carried out by the gold-standard techniques of X-ray protein crystallography or nuclear magnetic resonance (NMR) spectroscopy (Figure 4.1A). Indeed, protein crystallography and X-ray diffraction have a rich history at the University of Leeds. Sir William Henry Bragg pioneered X-ray diffraction at the university between 1909 and 1915, inventing the X-ray spectrometer and founding the field of X-ray crystallography, for which he shared the Nobel prize with his son. This work was further developed and applied by William Astbury at the University of Leeds between 1928 and 1961, who laid the foundation for our understanding of biomolecular structures such as the DNA helix and the protein alpha helix.

The field of electron microscopy (EM), and specifically cryo-EM, has enabled investigation of biological samples in new ways. Under this umbrella of cryo-EM, cryo-electron tomography (cryo-ET) has become a favoured technique for high-resolution *in situ* elucidation of biological samples such as whole cells or large complexes. Alternatively, isolated samples may be analysed using an SPA pipeline (Thompson et al., 2019). Indeed, in recent years advances in hardware and software have transformed the SPA approach, pushing it to the forefront of biomolecular structure determination and reaching similar structural resolution as for X-ray crystallography (Cheng, 2018). The deposition of structures resolved through EM continues to grow explosively (Figure 4.1A) and the technique is fast becoming the preferred approach for structural elucidation of virus proteins. Indeed, many viral capsids fail to form the ordered crystals necessary for X-ray crystallography and they are often too big to be suitable for NMR spectroscopy approaches, while EM is ideally set up for structural virology. In many cases, purified or isolated virus particles are sufficiently small so that they can be used whole in EM, avoiding the need for the purification of individual virus proteins. This is of particular interest for viruses containing capsids with repeating geometrical patterns since this symmetry aids in the reconstruction of the virus particle from EM data. Most capsids with repeating geometrical patterns are classically considered icosahedral in nature, either exactly or through quasi-equivalence (Caspar and Klug, 1962; Flint et al., 2009a). Additionally, members of some virus groups contain nucleocapsids which often exhibit helical geometries (Flint et al., 2009e). While the icosahedron is often observed, recent literature shows that many viruses are not described well by the Platonic geometry (convex objects made up of one type of regular polygon) but

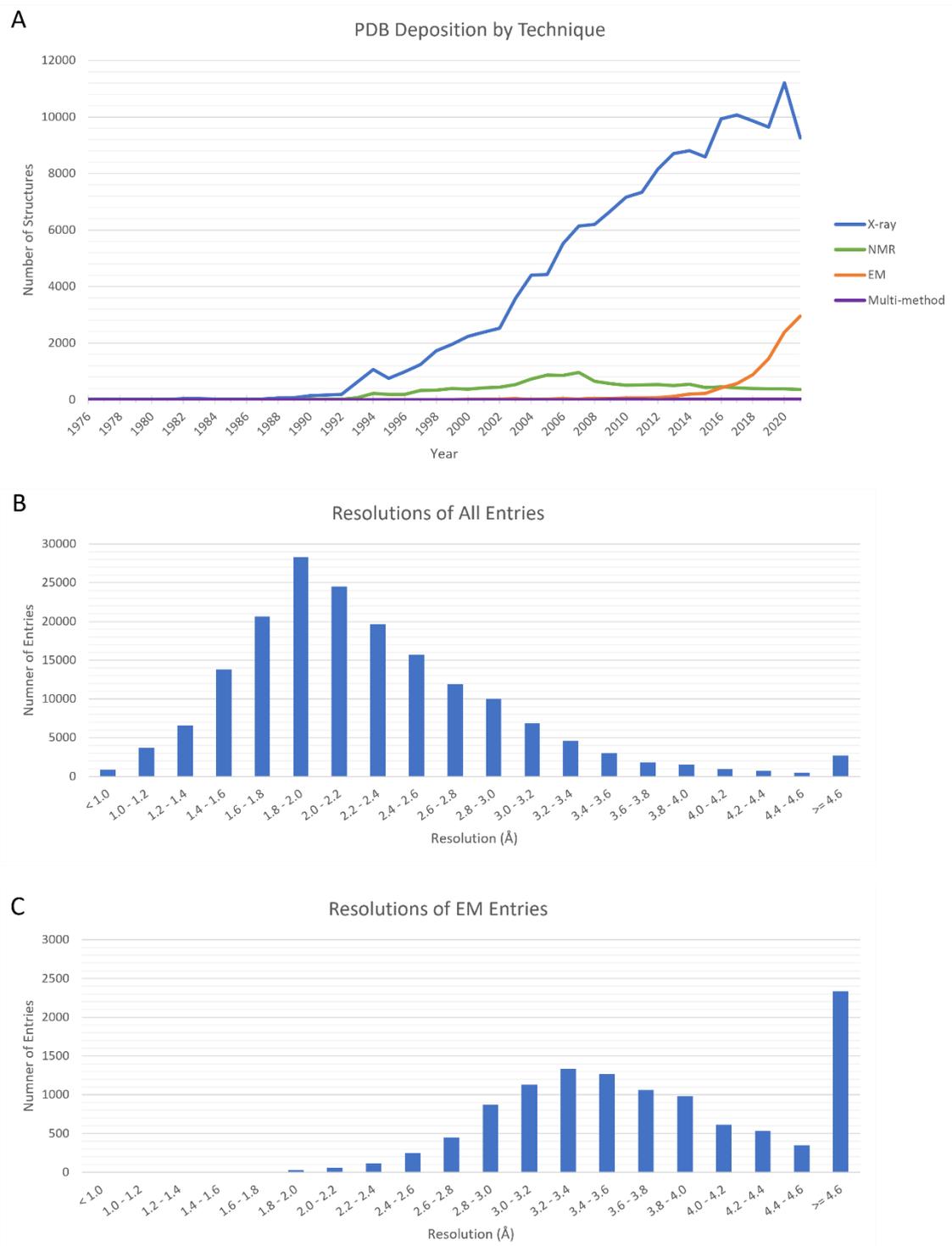


Figure 4.1. Generation of biological structures.

Deposition of structures in the Protein Data Bank (PDB) and their resolutions. **A**) Deposition of structures over time and by technique used. X-ray: X-ray crystallography, NMR: nuclear magnetic resonance, EM: electron microscopy. Data inclusive of every complete calendar year, data for 2022 is not included. **B-C**) The distribution of the reported resolution of deposited structures in the PDB for all entries taken together (**B**) or for electron microscopy entries only (**C**). Data inclusive of all structures available at the date of data retrieval. All data retrieved from the PDB website (Protein Data Bank, 2022).

rather follow Archimedean geometry (convex objects made up of two or more types of regular polygon) with an overarching icosahedral design (Twarock and Luque, 2019). Regardless, capsids constructed following Platonic as well as Archimedean geometry feature multiple axes of symmetry (Flint et al., 2009a; Twarock and Luque, 2019). These symmetry axes offer a level of symmetrical redundancy that aids a high-resolution reconstruction from a relatively small number of particles, since each particle contributes multiple asymmetric units to the reconstruction (Marsian et al., 2019; Snowden et al., 2020; Procházková et al., 2021). Additionally, cryo-electron tomography (cryo-ET) approaches (Wan and Briggs, 2016; Quemín et al., 2020) may be employed to study more pleiomorphic viruses. This approach normally leads to lower-resolution reconstructions but may further understanding of virus-host interactions (Cox and Plemper, 2017; Punch et al., 2018; Klein et al., 2020; Li, 2022). Both geometrically repeating virions as well as more pleiomorphic viruses are viable samples for EM, yet their data collection and down-stream data processing approaches may differ.

Structural information on HeV N protein and HeV nucleocapsids is scarce, collectively no more than a handful of negative-stain EM micrographs of isolated HeV nucleocapsids (Hyatt et al., 2001) or recombinantly expressed HeV nucleocapsid-like filaments (NLFs) (Pearce et al., 2015). These micrographs provide insight into the overall morphology of HeV nucleocapsids/NLFs at low resolution and in 2D. The flexible nature of the complexes exemplifies the suitability of EM to approach the HeV N structure and nucleocapsid architecture over X-ray crystallography and NMR spectroscopy.

4.1.1 Resolving power

Optical light microscopy, and by extension fluorescence microscopy, is a technique ubiquitously employed in biological research. Live cell imaging carried out in this way provides a dynamic view of biological processes, such as viral infections. Tracking the spatiotemporal localisations of specific proteins or complexes provides a wealth of information and can reach appreciable spatial resolutions in the nanometre range through single molecule fluorescence (Lelek et al., 2021) While much of virological research depends on light microscopy, cryo-EM enables the capture of snap shots of virus infected cells, virions in isolation, or even sub-viral components at resolutions unattainable through optical light microscopy.

Viruses come in a wide range of shapes and sizes, however the majority are roughly spherical and have a diameter between 20 to 300 nm, thereby falling outside the range of visible light wavelengths (700 to 400 nm) (Flint et al., 2009e; Oxford et al., 2016). As a consequence,

visible light microscopy does not have the optical resolution to image virions, let alone cellular macromolecular complexes or viral components which are often much smaller still. For example, the ribosome complex, a large and often-researched cellular macromolecular complex, is around 25 nm in diameter (Wilson and Doudna Cate, 2012) and the paramyxoviral nucleocapsids are around 20 nm wide (Plempner and Lamb, 2021).

4.1.2 Anatomy of a transmission electron microscope

Much like an optical light microscope, a transmission electron microscope employs a series of lenses to focus an illuminating beam through a sample and onto a detector. While the illuminating beam in optical light microscopes is made up of photons, transmission electron microscopes use electrons instead. While in principle they have the same aims, the technology to obtain an electron micrograph is very different to those involved in optical light microscopy.

Although there are differences between manufacturers, all electron microscopes fundamentally employ a similar arrangement of components inside a column that is being held at high vacuum (Figure 4.2A). The high vacuum is required to reduce electron scattering by air molecules. Most early electron microscopes used thermionic electron sources, such as tungsten or lanthanum hexaboride (LaB6) filaments. More recently, and of particular note in cryo-EM, field emission guns (FEGs) have supplanted the thermionic electron sources for their higher temporal and spatial coherence, and brightness and as a result FEGs produce better SNRs than thermionic devices (Thompson et al., 2016). An anode accelerates the emitted electrons, with typical acceleration voltages between 100 and 300 kV. Higher accelerating voltages generate electrons with shorter wavelengths, with 300 kV microscopes reaching electron wavelengths of around 0.02 Å (Milne et al., 2013). As a rule of thumb, shorter wavelengths correspond to higher resolution in the final image. A series of electromagnetic lenses focus the electron beam and provide magnification. Firstly, the condenser lens system focuses the electron beam to achieve parallel illumination of the sample which is critical for high-resolution cryo-EM. The sample may scatter electrons before reaching the detector at the bottom of the column. The objective lens system provides the main magnification while an objective aperture blocks high-angle scattered electrons thereby improving image contrast (Orlova and Saibil, 2011). This aperture sits at the back focal plane of the objective lens system and sample-induced high-angle scattered electrons are blocked from passing through. Consequently, sample areas of high density would result in more scattered electrons, generating darker spots since fewer electrons pass through, while sample areas of lower density generate lighter spots since most electron pass through. Finally,

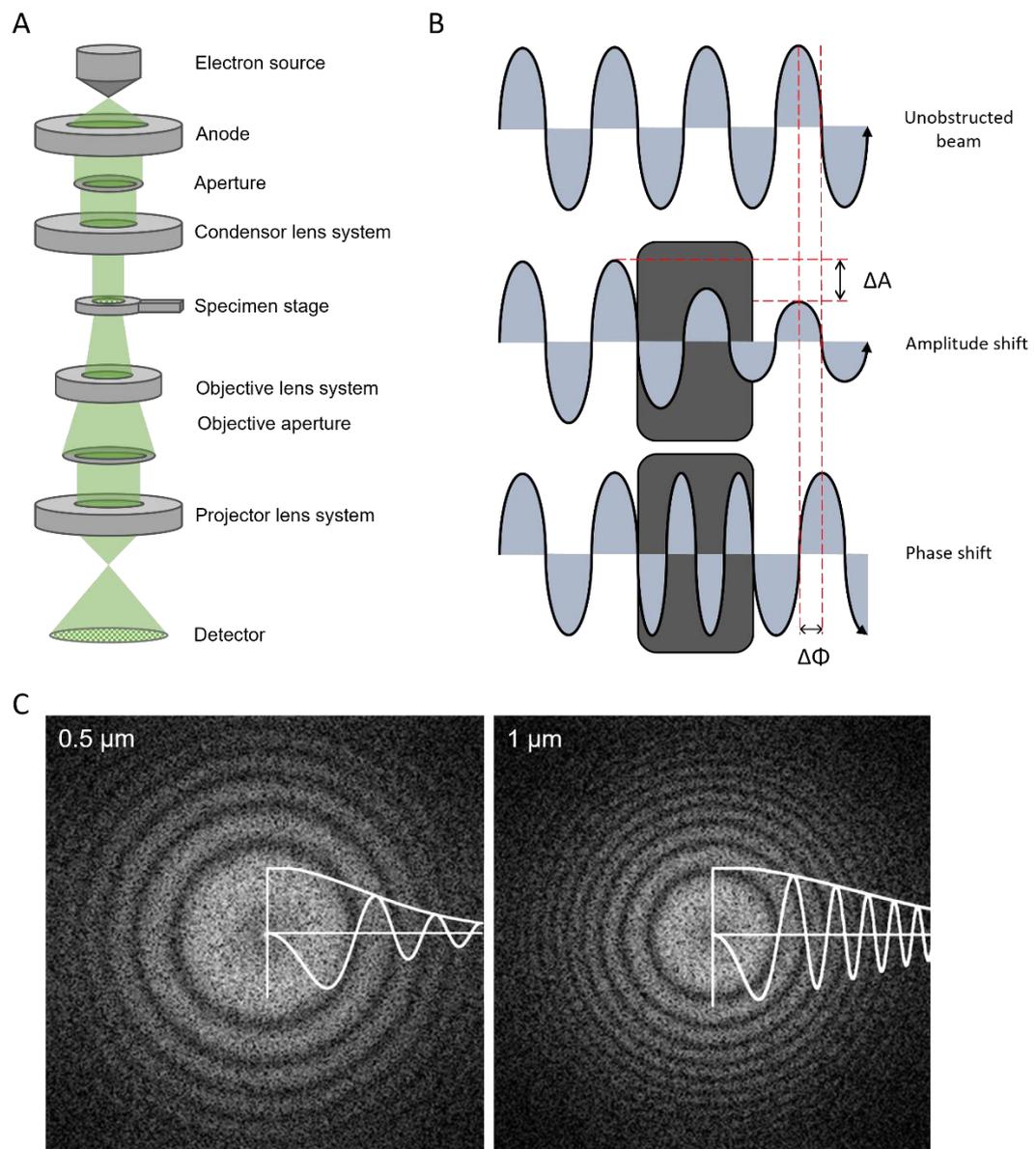


Figure 4.2. Principles of transmission electron microscopy.

A) Schematic representation of a transmission electron microscope. An electron beam (green) is generated between the electron source (cathode) and anode dyad. The cathode can be a thermionic filament or FEG. The beam passes through a set of electromagnetic lens systems and apertures that focus the electron beam and exclude high angle electrons. The parallel beam passes through the specimen after which high angle scattered electrons are excluded and further electromagnetic lens systems magnify and project the electron beam at the detector, which can be film, CCD, or DED to form an in-focus image. **B)** Schematic representation of amplitude contrast and phase contrast. After the electron beam passes through an object that generates amplitude contrast, a reduction in amplitude compared to the unobstructed beam (ΔA) creates a difference in signal intensity. In the case of negative stain-EM, the stain generates a strong shift in amplitude while the beam passing through the specimen suffers a smaller change in amplitude. After the electron beam passes through an object that generates phase contrast, the electron beam is phase shifted compared to the unobstructed beam ($\Delta\Phi$) resulting in phase-induced interference. In the case of cryo-EM, the generates a range of phase shifts while the beam passing through the vitreous ice remains unshifted. **C)** The diffraction patterns of micrographs of carbon film, showing Thon rings and corresponding CTF curves (white). The defocus values of 0.5 μm and 1 μm are indicated. The Thon rings of the 1 μm defocused image oscillate more rapidly between positive and negative contrast and are located closer to the origin. The plotted curves depict the oscillation between positive and negative contrast. Thon ring images were adapted from (Orlova and Saibil, 2011).

the projector lens system provides further magnification before electrons reach the detector.

4.1.3 Negative stain EM versus cryo-EM

Two main approaches exist for biological imaging through EM. While the machinery is in essence the same for both, with some additional components in cryo-EM systems, the usage for either system has diverged. In most biostructure projects, negative stain EM is the first port of call. The rapid grid preparation and relative ease of operation make negative stain EM a powerful visual tool for initial assessment of the morphological (composition and conformation) heterogeneity in the sample, at a range of buffer conditions (Ohi et al., 2004; Scarff et al., 2018). Negative stain EM can answer many biological questions a researcher may have about their object of interest, however negative stain EM generally fails to produce the high-resolution data available through its cryogenic counterpart; cryo-EM. In this approach, samples are imaged in the context of a thin layer of vitreous ice at temperatures around $-180\text{ }^{\circ}\text{C}$. The cryo-EM approach is generally more complex, and time and resource intensive than negative stain EM, but its data can provide more detailed answers regarding macromolecular architectures and protomer structures. Technological and methodological advancements continue to push the resolution limit of both negative stain- and cryo-EM. The resolution limit in negative stain EM is capped by the grain size of the stain used and commonly sits around $10\text{-}30\text{ }\text{\AA}$ (Peter, 2004; Scarff et al., 2018; Gallagher et al., 2019). The resolution of cryo-EM structures deposited in the PDB continues to improve each year, with the highest deposited so far a structure of apoferritin at $1.15\text{ }\text{\AA}$ but high-resolution structures in the range of $3\text{-}4\text{ }\text{\AA}$ are most common (Figure 4.1B-C) (Protein Data Bank, 2022).

Furthermore, the preparation of negative stain- or cryo-EM samples share some commonalities but also important differences. While optical light microscopes generally use glass slide to mount samples in the path of the light beam, EM systems use metal grids containing a mesh with regular rectangular spacing covered by a support film. Ahead of sample application, the grid surface must be treated to improve sample adhesion, in a process called glow-discharge. The type of mesh, support film, and glow-discharge conditions are important considerations in sample preparation and are described in more detail below.

Negative stain EM derives its name from the method by which sample contrast is enhanced. After application to a suitable grid, a number of wash steps and blotting off excess liquid, a staining solution is applied. The electron-opaque staining fluid encases the particles in the sample, effectively staining the background and leaving the specimen untouched, in a non-

hydrated state. Electron-opacity is related to the number of atomic protons. Suitable stains should strongly scatter electrons and be biologically adsorbent. As such, a number of stains exist based on uranyl, tungstate, and osmium containing compounds, which are often radioactive or toxic. At the University of Leeds, uranyl acetate stains are common practice. The small grain size of around 4-5 Å (Hayat, 1986) permits the resolution of protein subunits and domains, lending insight into overall architecture, relative orientations, and levels of flexibility within a protein or complex (Burgess et al., 2004).

Beyond the use of metal grids with carbon support films, sample preparation for cryo-EM is drastically different from negative stain EM. No staining solution is used, and samples remain in a hydrated state after being applied to the grid and the excess is blotted off, leaving a thin aqueous film of around 10 to 80 nm. Then, in a process called plunge-freezing, the grid is rapidly submerged in a cryogenic medium, commonly liquid nitrogen-cooled liquid ethane or a mix of ethane and propane (Dubochet et al., 1971; Dobro et al., 2010). This process freezes the remaining sample-containing aqueous solution on the grid so rapidly that crystalline ice does not have time to form and instead generates vitreous ice (Passmore and Russo, 2016). Critically for cryo-EM, vitreous ice is electron transparent whereas crystalline ice is not. Furthermore, the ice protects the sample from the high-vacuum conditions within the microscope and imaging at cryogenic temperatures affords the sample some protection against electron radiation damage (Bammes et al., 2010).

4.1.4 Grid selection and preparation

EM grids consist of a circular metal frame, usually made of copper (Cu) or gold (Au), with rectangular grid bars. Between these grid bars lies a regularly spaced mesh and can be covered by a support film of amorphous carbon or gold (Russo and Passmore, 2016). Typical grids are 3mm in diameter. The mesh size number for the metal grids is defined as the number of divisions per inch. Typical mesh numbers are 200, 300, and 400, with larger mesh numbers having more and smaller grid squares.

The support film may contain an array of well-defined, regularly spaced and sized, circular holes. These grids are termed 'holey' grids. Holey grid patterns have defined geometry, noted as d/s where the "d" term refers to the hole diameter in μm and the "s" term refers to the hole spacing in μm . Common hole patterns for single-particle cryo-EM are 1.2/1.3 and 2/2, but other geometries are available. Alternatively, a random pattern of holes - termed 'lacey' - is also

available for amorphous carbon support films. A further ultrathin film of continuous carbon or graphene oxide, around 2 nm thick, can be applied atop the support film.

Ahead of sample application, the grid surfaces must be glow-discharged. In this treatment, an electrically generated plasma renders the grids hydrophilic. In most cases, glow-discharge is performed in low-pressure air but defined gas mixtures including argon and oxygen, or organic compounds such as amylamine can be used to generate hydrophobic surfaces and impart an overall negative or positive charge (Passmore and Russo, 2016). Defining gas mixtures and additives may also improve reproducibility over unaltered air.

The support material, support film shape, hole pattern and size, as well as additional ultrathin film may all influence the sample behaviour during grid preparation and by extension, the suitability for high-resolution EM. As the main components of the negative stain and cryo-EM systems have a common basis but additional features, so does sample preparation for both techniques diverge in certain aspects.

4.1.5 Image formation

In order to observe biological samples in EM, contrast is a critical necessity. In EM images, contrast has one of two sources: amplitude contrast or phase contrast. Amplitude contrast refers to a difference in intensity of incident electrons. The use of amplitude contrast is a fundamental property of negative stain EM (Figure 4.2B). The electron-opaque staining fluid results in high-angle scatter of the electron beam while the more electron-transparent specimen permits increased transmission of the electron beam. Thus, the areas on the grid containing specimen particles appear brighter than the surrounding stained surface (Orlova and Saibil, 2011; Carlo and Harris, 2012). With the absence of a staining solution and similar electron-opacity between the specimen particles and surrounding vitreous ice, cryo-EM cannot rely on amplitude contrast. This is further compounded by the required low electron dose rates. As a consequence, cryo-EM depends heavily on phase contrast, which results from the interference at the detector between electrons that pass through the specimen unaltered and those that are scattered by the sample (Figure 4.2B). The thin samples permissible in cryo-EM only scatter electrons over small angles, resulting in small phase shift and thus poor contrast. The phase shift, and in turn contrast, can be increased by purposefully introducing a defocus.

The defocus value, together with the microscope-inherent spherical aberration, affect the contrast transfer function (CTF). The CTF describes the EM imaging process in Fourier space and is an oscillating function which measures contrast as a function of spatial frequency

(Erickson and Klug, 1970; Cheng, 2015; Cheng et al., 2015). The Fourier transform of a collected EM image shows this oscillating function in the form of a Thon ring (Figure 4.2C). Greater defocus values generate higher oscillation frequencies in the CTF and result in increased signal delocalisation, thereby losing high-resolution information. This interplay poses a trade-off between good contrast and true signal that must be considered ahead of data acquisition. In practice, the optimal defocus is the minimum value at which particles of interest are still visible so that high-resolution information is retained. Using a range of defocus values around this optimum compensates for specific spatial frequencies at which contrast is lost.

4.1.6 Image Detection

The detection of electrons is critical for image formation in EM, and the way in which they are detected has a profound influence of data quality. In practice, all detection methods add some noise to the captured data. This noise is described as the detective quantum efficiency (DQE), which is a function of signal to noise ratio (SNR). The DQE is defined as: $DQE = \frac{(Output\ SNR)^2}{(Input\ SNR)^2}$. The perfect detector would capture only signal and generate a DQE of 1. Early electron microscopy employed photographic film where chemical interactions between the film and incoming electrons altered the film surface. The DQE of these films was generally better than the early charge-coupled device (CCD) cameras (McMullan et al., 2009; Thompson et al., 2016).

Recent advances in direct electron detectors (DEDs) have revolutionised EM of biological samples in ice, as they allow for faster and better quality data collection (McMullan et al., 2016). Where CCD cameras indirectly convert electrons to photons for detection, DED directly translate incident electrons into electrical signals, resulting in an improved DQE. Additionally, DEDs allow for continuous, high frame-rate detection generating a movie from the frames. This in turn enables computational tracking and correction of motion in the data as a result of stage drift or beam-induced specimen motion (see Chapter 5). Most DEDs are operated in one of two modes; integrating mode or counting mode. In integrating mode, the signal is measured over a fixed time period and thus multiple incident electrons contribute to an integrated recorded signal. This mode offers high electron dose rates and reduced exposure times, but as a result has generally a reduced SNR. Contrarily, in counting mode each incident electron event is recorded individually. This mode allows for more accurate localisation of electrons, but as a consequence requires exceedingly low electron dose rates to avoid detector saturation leading to longer exposure times and slower data acquisition rates. In older DEDs,

such as the Falcon III, the difference in these modes was an important consideration. In newer DEDs such as the Falcon IV and K3, increases in frame rate have reduced required exposure times. This has sped up data acquisition considerably, mitigating the drawbacks of counting mode drastically and making it much more attractive for high-volume data acquisition.

4.1.7 Chapter aims

This Chapter aims to characterise the recombinantly expressed HeV N protein through negative stain EM, optimisation of cryo-grid preparation, and collection of cryo-EM datasets for down-stream 3D reconstruction of HeV N protein and HeV NLFs. The protocols covering the results in this section are covered in Chapter 2 Materials and Methods, Section 2.2.3.

4.2 Results

4.2.1 Negative stain electron microscopy reveals a range of HeV N particle morphologies

The SEC UV-trace data obtained for recombinantly expressed and affinity purified HeV N protein (Chapter 3 Section 3.2.4) showed the presence of high molecular weight species indicative of multimeric species. To confirm the oligomeric state of HeV N, SEC fraction 13 (Figure 3.5) was visualised through negative-stain EM. Carbon-coated copper grids were glow-discharged before HeV N sample was applied. Aliquots of SEC fraction 13 were used undiluted and at a 2-fold dilution, resulting in final concentrations of 45 $\mu\text{g/ml}$ and 22.5 $\mu\text{g/ml}$. These concentrations fall in line with previously published methods (Pearce et al., 2015). Following two wash and blot steps with water, the copper grids were then negatively stained with 2% uranyl-acetate. The stained grids were then used for imaging immediately or stored at room temperature for imaging at a later date.

The stained grids were imaged using the FEI T12 Electron Microscope. The micrographs show particles with three predominant morphologies: filamentous tubes, circular rings, and squat rings (Figure 4.3). Since no RNA was introduced to the sample, and the 260/280 ratio indicated the presence of nucleic acid (Chapter 3 Section 3.2.4), these morphologies consist of non-RNA-bound (apo) HeV N or HeV N bound to heterologous *E. coli* RNA. While the sample showed heterogeneity between HeV NLFs, the three morphologies are clearly recognisable and overall the sample shows a high degree of purity. Further optimisation in different sample buffers should provide an indication of the sample's suitability for cryo-EM.

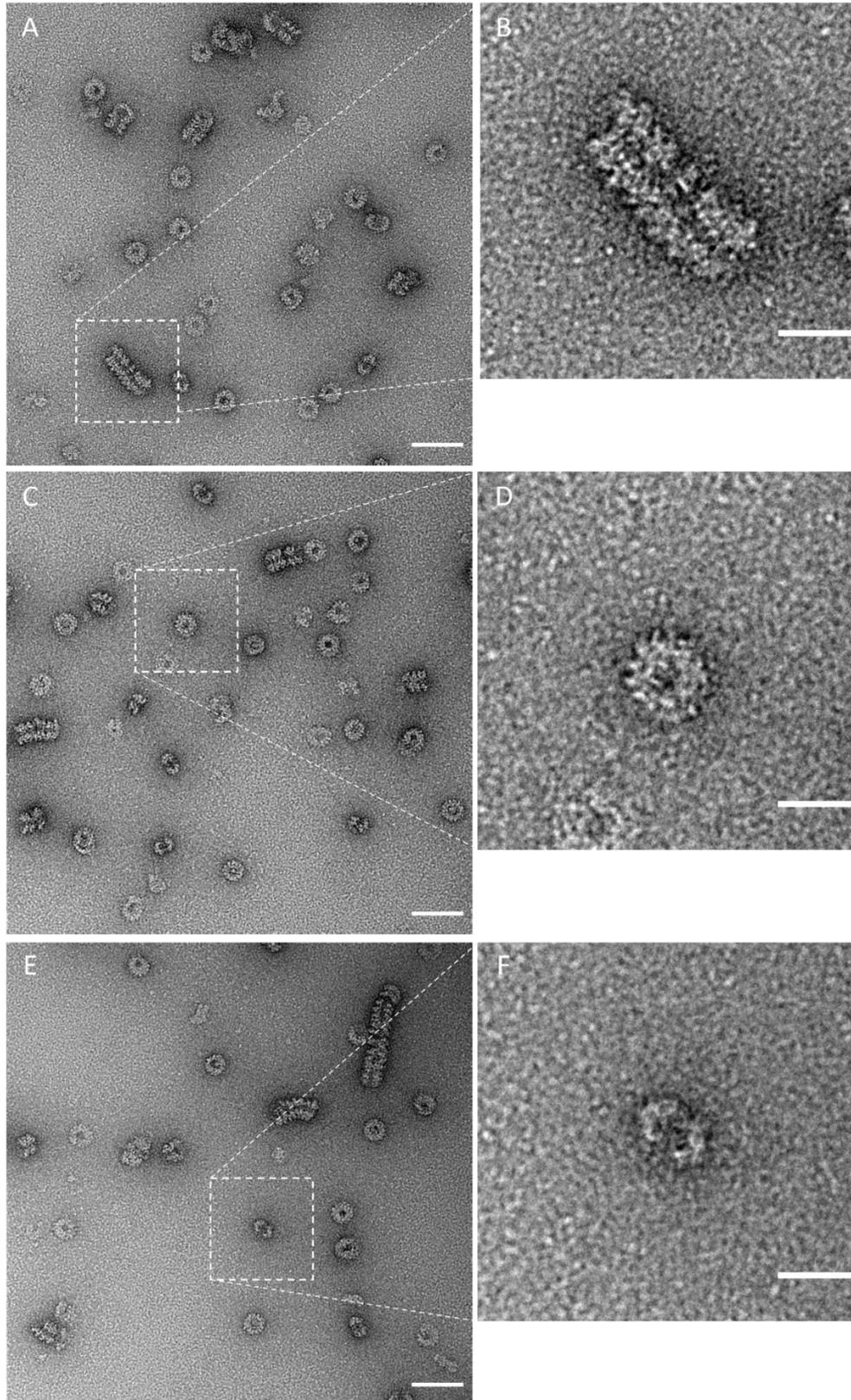


Figure 4.3. HeV N negative stain EM reveals three distinct morphologies.

In-house produced carbon-coated copper grids were prepared with HeV N protein sample from SEC. Negative staining with 2% uranyl-acetate revealed HeV NLFs of three predominant morphologies. Representative micrographs are shown, with scale bars of 50 nm (A, C, and E). B, D, and F) Insets from A, C, and E, respectively highlighting a filamentous morphology, a circular ring morphology, and a squat ring morphology. Scale bars for insets are 20 nm.

4.2.2 HeV NLFs are stable in poorer buffer conditions

High-resolution 3D reconstruction from cryo-EM micrographs requires good a SNR. Buffer components can lead to a reduction in SNR in various ways, and as such a buffer closest to pure H₂O is ideal. The purified HeV NLFs were in a buffer containing relatively high salt, DTT, and glycerol as a cryoprotectant – antithetical to the formation of vitreous ice to capture protein complexes in a soluble, hydrated state.

To facilitate the downstream high-resolution 3D reconstruction of HeV NLFs, a reduction to low salt buffer conditions – being closer to pure water – was carried out and NLF morphology was assessed through negative stain EM (Figure 4.4). Aliquots of concentrated, pooled SEC fractions (Figure 3.6; 2 mg/ml) in Tris-HCl buffer were diluted to reduce the NaCl, DTT, and glycerol concentrations. Starting at 500 mM NaCl, 0.5 mM DTT, and 5 %v/v glycerol, samples were diluted 1 in 5 (Figure 4.4A-B), 1 in 10 (Figure 4.4C-D), and 1 in 100 (Figure 4.4E-F) to result in final concentrations of 100 mM NaCl, 0.1 mM DTT, 1 %v/v glycerol; 50mM NaCl, 0.05 mM DTT, 0.5 %v/v glycerol; and 5 mM NaCl, 5 µM DTT, 0.05%v/v glycerol.

Negative stain EM micrographs indicated that at 1 in 5 and 1 in 10 dilutions (Figure 4.4A-D), HeV NLFs exhibit morphologies similar to those in the undiluted starting buffer. At a 1 in 100 dilution factor, however, the expected NLF particles were absent and instead protein aggregates were found (Figure 4.4E-F), indicating the extremely poor buffer conditions likely lead to protein denaturation and aggregation.

These results indicated that NaCl concentrations of 50 mM and higher are suitable for HeV N samples intended for cryo-EM. Ideally glycerol should be omitted entirely to promote optimal cryo-grid conditions and these results support that such an omission is likely without deleterious effects, since reducing the buffer to 0.5%v/v glycerol maintained morphologically intact HeV NLFs.

4.2.3 HeV NLFs have an affinity for carbon on cryo-grids

Since cryo-EM data collection is both time and resource intensive, it is prudent to optimise cryo-grid preparation. High quality, vitreous ice at appropriate thickness and optimal particle density, distribution, and orientation each improve the likelihood for high-resolution 3D reconstruction of the object of interest.

With good particle morphology in negative stain EM, the HeV N samples in 25 mM Tris-HCl pH 7.5 and 100 mM NaCl could be taken forward for cryo-grids preparation. Cryo-grids were

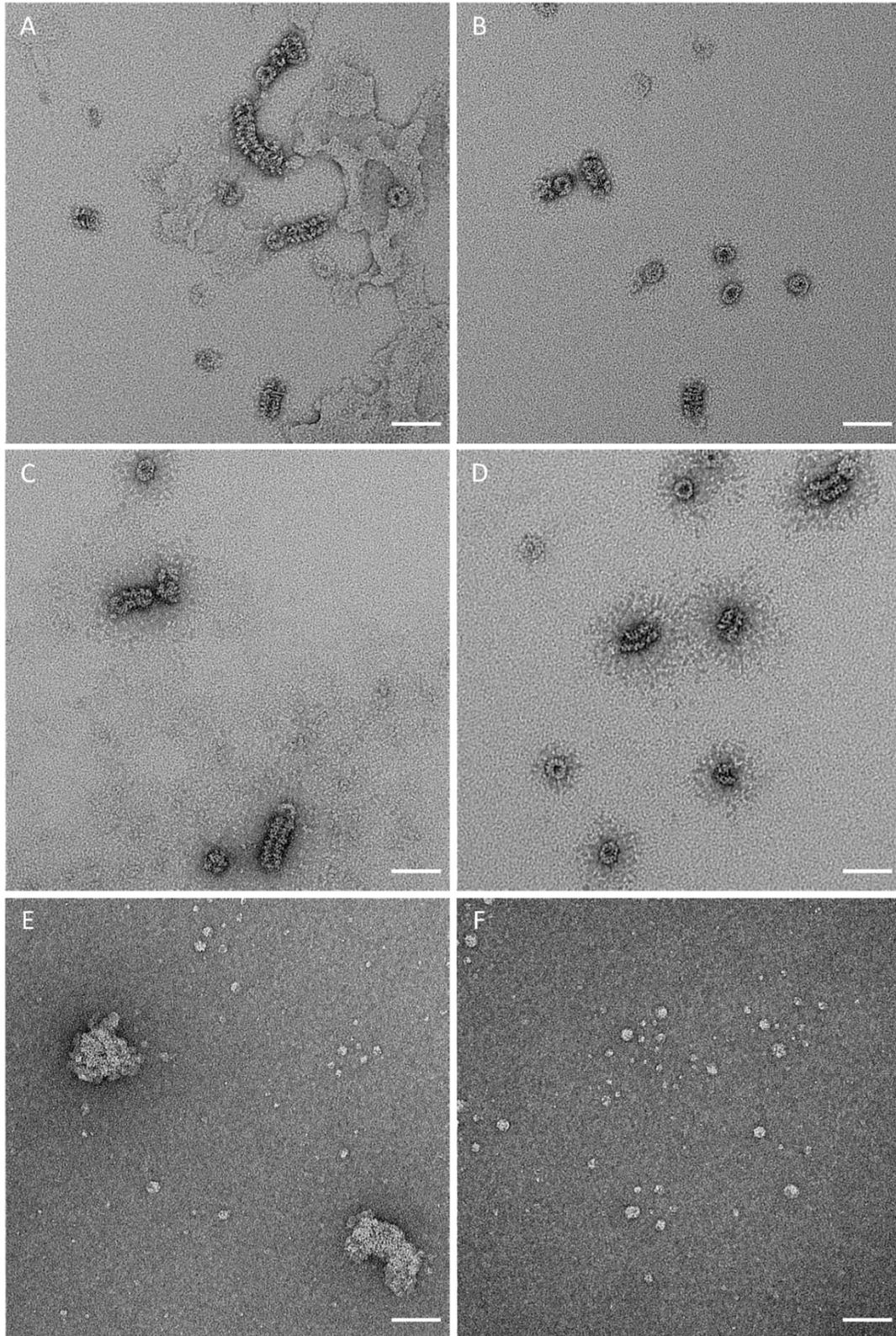


Figure 4.4. HeV NLFs are stable in poorer buffers.

Aliquots of pooled SEC fractions in Tris-HCl buffer with 500 mM NaCl, 0.5 mM DTT, and 5%v/v glycerol were diluted to reduce the NaCl, DTT, and glycerol concentrations before application onto in-house produced carbon-coated copper grids and negative staining with 2% uranyl acetate. **A-B)** a 1 in 5 dilution resulting in 100 mM NaCl, 0.1 mM DTT, and 1%v/v glycerol. **C-D)** a 1 in 10 dilution resulting in 50mM NaCl, 0.05 mM DTT, and 0.5% glycerol. **E-F)** a 1 in 100 dilution resulting in 5 mM NaCl, 5 μ M DTT, and 0.05%v/v glycerol. Scale bars are 50 nm.

initially glow-discharged in air to generate a negatively charged, hydrophilic grid surface, and optionally additional compounds (such as amylamine; see below) were used during this process to generate hydrophobic and/or positively charged surfaces (Dubochet et al., 1971; Aebi and Pollard, 1987). Using a Vitrobot IV plunge-freezing system, cryo-grids were flash frozen in liquid nitrogen-cooled liquid ethane. The system allows users to set reproducible parameters, such as chamber humidity, adhesion times, and blotting force and times. Prepared grids were then screened on one of the two FEI Titan Krios systems available at the Astbury Biostructure Laboratory, with the aim of optimising for helical HeV NLFs as the particle of interest.

For the initial grid-prep, the standard glow-discharge was applied to Quantifoil Cu 400 2/2 holey grids. The glycerol and DTT containing sample used in Figure 4.4 was used for this initial screening. Diluted (1 in 10 and 1 in 20) and undiluted samples were applied to the grids and the excess was blotted away using blot force 6 and 6 second blotting. Screening on a FEI Titan Krios revealed few grid squares with vitreous ice of appropriate thickness, and half of the prepared grids were unusable. The undiluted sample exhibited a high particle density of primarily circular ring morphology (Figure 4.5A). Screening the 1 in 10 diluted sample revealed that the majority of particles populate the carbon support film, an area in which data collection is not feasible (Figure 4.5B). Aside from suboptimal ice coverage, particle density was either too low, or too high and showing signs of preferred orientation (Figure 4.5A and B).

For the second grid preparation, Quantifoil Cu 400 2/2 and Cu 300 1.2/1.3 grids were used with the same glow-discharge as before. Concentrated samples from F12 and the Pool (Buffer 25 mM Tris-HCl pH 7.5, 100 mM NaCl; Figure 3.9) were applied to the grids using the same blotting settings as before. Both the F12 sample (Figure 4.5C) and Pool sample (Supplementary Figure S4.1) showed particles of all three morphologies, with the F12 sample tending to have longer filamentous tubes. No improvements in particle distribution were observed for either grid type, with longer filamentous tubes tending to align with the carbon edge. Additionally, the distribution and uniformity of ice thickness, as well as the particle density remained sub-optimal. As a consequence, the number of usable grid-squares was insufficient for data collection.

Next, the F12 and Pool samples were used to prepare additional Quantifoil Cu 400 2/2 grids with a range of adhesion times (5-20 sec), blotting times (4-12 sec), and blotting forces (4-8). The Pool sample provided good ice but particle density remained too high (Supplementary Figure S4.1). The F12 sample had both good distribution and uniformity of ice thickness and particle distribution (Figure 4.5D and E). The longer filamentous tubes continued to align to the

carbon edge, but shorter filamentous tubes as well as circular rings and squat rings populated the ice with good density and distribution. Additionally, particle orientations seemed better than on previous cryo-grids as well as negative stain-grids.

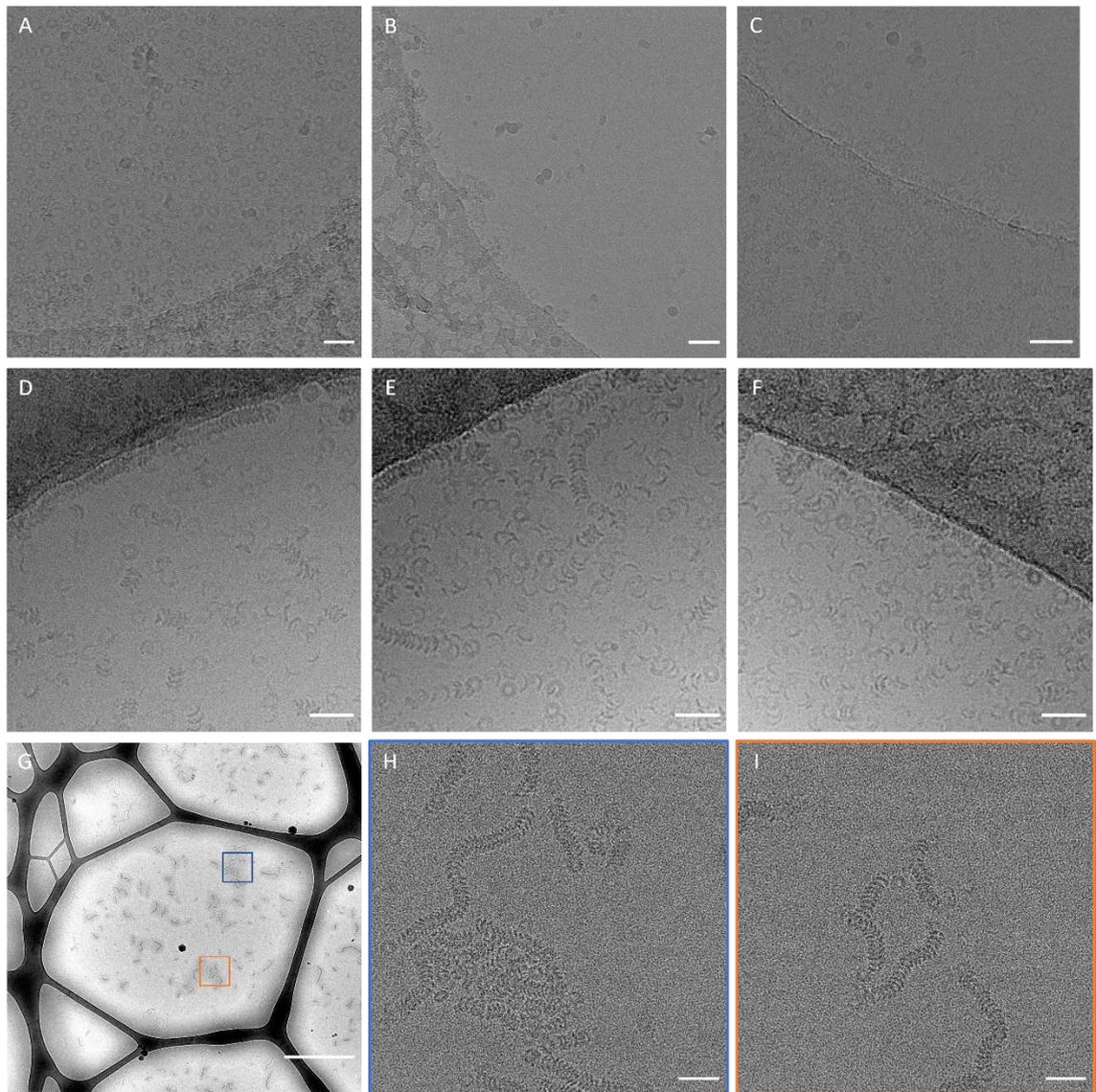


Figure 4.5. HeV NLFs show affinity for carbon on cryo-grids.

Various HeV N samples from SEC were applied to glow-discharged holey or lacey grids and plunge-frozen using an FEI Vitrobot IV. Prepared cryo-grids were screened on FEI Titan Krios. HeV N samples containing DTT and glycerol were applied undiluted (A) or diluted 1 in 10 (B) on Quantifoil Cu 400 R2/2 holey grids, showing similar morphologies as in negative stain, but with poor particle distributions. The F12 sample was applied on Quantifoil Cu 300 R1.2/1.3 (C) or Quantifoil Cu 400 R2/2 (D-E) holey grids, showing good particle distribution but with long filamentous tubes aligning to the carbon edge. F) Quantifoil Cu 400 R2/2 grids were glow-discharged in the presence of amylamine and F12 sample was applied, showing a similar particle distribution as for standard glow-discharge. G) Hole/Eucentric magnification of UC Lacey grids showing good particle distribution across the lacey holes and good ice thickness. Insets (H-I) show long filamentous tubes with good dispersion and high-resolution information. Scale bars are 50 nm for A-F, H, and I. Scale bar is 1 μm for G.

The tendency for long filamentous tubes to align along the carbon edge remained problematic, because particles along the carbon edges are unsuitable for downstream reconstruction. In an attempt to move these into the open holes, Quantifoil Cu 400 2/2 grids were glow discharged in the presence of amylamine. This treatment renders the grid surface hydrophobic and positively charged (Passmore and Russo, 2016). However, no improvement over glow discharge in air was observed for any of the grids prepared in this manner (Figure 4.5F).

Since all grids were prepared using samples from F12, differences in particle distribution and ice quality were resultant of using different types of grids as well as sample application parameters. With the observation that longer filamentous tubes had an affinity towards the carbon edge, the holey grids were replaced with lacey grids with an ultrathin carbon (UC) film. Using similar adhesion and blotting parameters, the UC lacey grids showed improved dispersion for long filamentous tubes (Figure 4.5G-I) with good ice thickness and uniformity. This grid was taken forward for data collection.

4.2.4 Cryo-electron microscopy datasets can be collected automatically

For down-stream 3D reconstruction, generally speaking the more particles the better the final resolution. Taking cryo-TEM micrographs manually becomes impractical at the numbers needed for 3D reconstruction and data collection is therefore performed using an automated system. Once an optimised grid has been identified, suitable grid squares are manually selected within the EPU software (Thompson et al., 2016). Acquisition magnification, electron dose, and defocus range are set before the EPU software automatically moves the grid and takes the micrographs.

During cryo-grid optimisation, the UC lacey grids showed the best particle distribution for filamentous tubes. This grid was used for automated data collection and the data collection parameters were recorded (Figure 4.6). A total of 3548 micrographs were collected, showing good contrast and detail of long filamentous tubes as well as a number of circular rings and squat rings (Figure 4.6A-B).

4.3 Discussion

Cryo-EM data collection is both a resource and time demanding venture. Therefore, grid-optimisation ahead of cryo-EM data collection is critical. Understanding the gross

morphologies and particle behaviours through low-resolution screening ensure that only the best grids are employed for data collection, ensuring the best chance for a high-resolution 3D reconstruction.

Negative stain EM confirmed that, as suggested by the SEC data, fractions nearer the void volume contain particles at higher oligomeric states. Additionally, the data revealed three predominant morphologies: filamentous tubes of various length, circular rings, and squat rings. It is possible that these morphologies represent different views or orientations of the same particle types. For example, the filamentous tubes may represent a longitudinal (or side) view

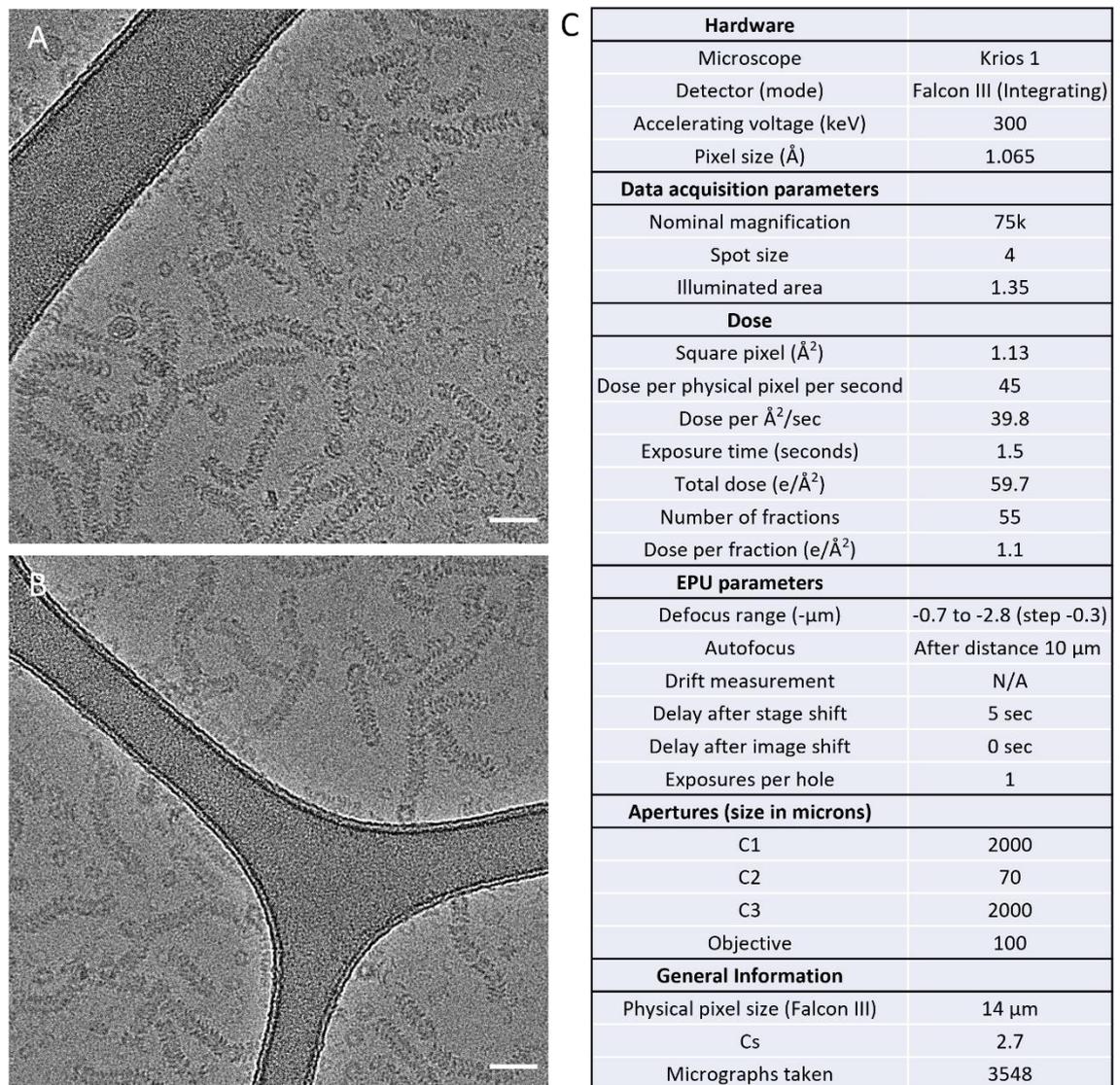


Figure 4.6. Data collection on a UC Lacey grid.

A-B) Representative micrographs from the FEI Titan Krios automated data collection through EPU, depicting HeV NLFs of all three predominant morphologies. Scale bars are 50 nm. **C)** Automated data collection parameters.

of the NLF. In turn, the circular rings could represent transverse (or top) views of shorter filamentous tubes. The squat rings look distinct from the circular rings and may represent a longitudinal view of a bottom-to-bottom stacking pair of circular rings. If this is the case, the complement of circular rings would represent a mix of transverse views of particles represented by both the filamentous tubes as well as the squat rings (Figure 4.7A).

Transitioning from negative stain grids to cryo-grids is accompanied with a major change in sample processing at the grid stage. Lessons learned from negative stain EM may not necessarily directly translate over to cryo-grid quality and particle behaviour. Therefore, cryo-grid optimisation and screening must be performed to ensure that the best grid is selected for

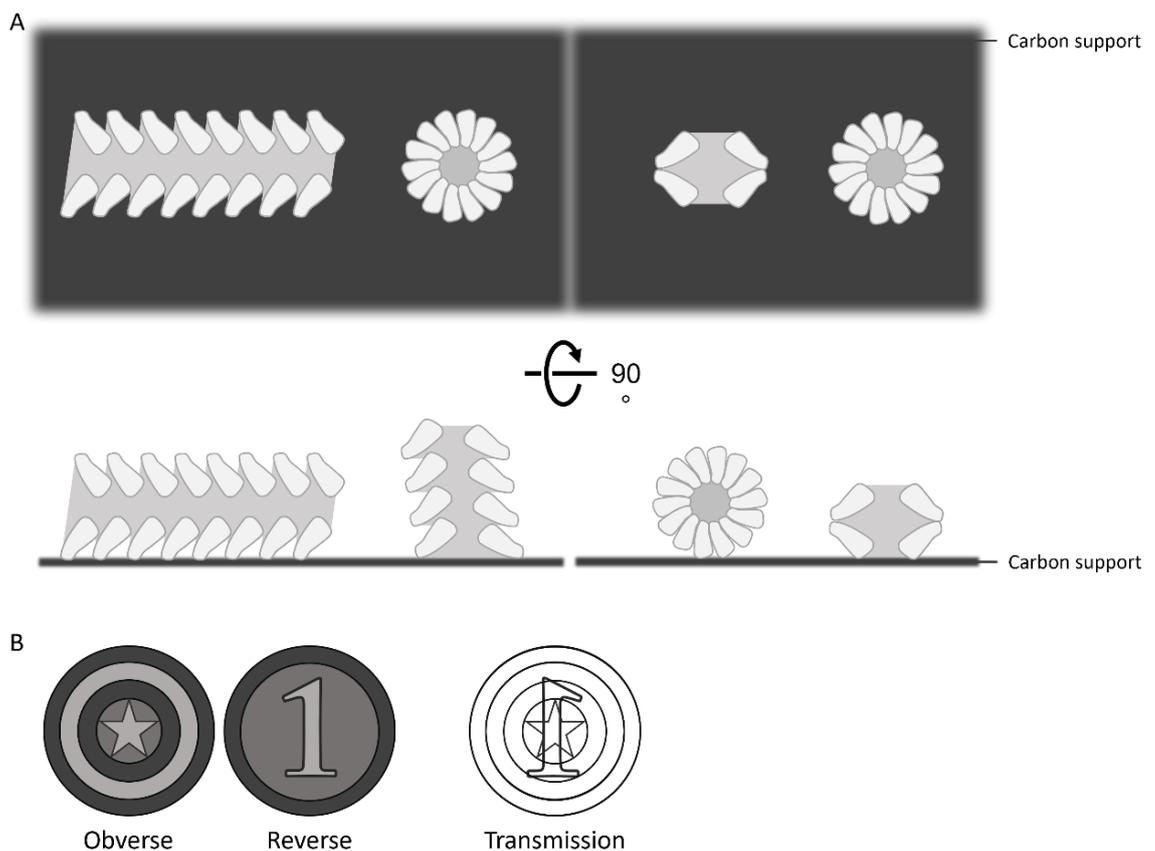


Figure 4.7. Particle orientation and transmission artefacts.

A) Schematic representation of HeV RLFs adhered to a grid. The top panels represent a top-down view through the microscope, showing the three predominant morphologies as they appear in the micrographs. From left to right; a filamentous tube, a circular ring, a squat ring, and a circular ring. The bottom panels are rotated 90° compared to the top panels to provide a side-on view of the same particles. The circular rings represent a top-down view of both filamentous tubes as well as squat rings, which are indistinguishable at low resolution. Note, projections of filaments and rings are not exactly the same and may be separated through 2D classification. **B)** A schematic of a coin showing its obverse and reverse faces. A transmission image of this coin, depicting an integration of detail from both faces and the internal structure into a single image.

data collection. Where negative stain grids have a continuous carbon layer for particles to adhere to, the aim with most holey cryo-grids is to suspend the particles in a layer of vitreous ice. For HeV NLFs, the majority of particles populated the areas of the carbon support film that are normally excluded during data collection. It seemed that HeV NLFs have an affinity to the carbon. This was supported by the observation that the longer filamentous tubes inside the holes tended to align to the carbon edge. Once lacey grids with an ultrathin carbon layer were used, HeV NLFs appeared to have a good distribution across the lacey holes that is much more suitable for data collection.

4.3.1 Comparison with other *Paramyxoviridae*

From both negative stain EM and cryo-EM data, the filamentous tubes exhibit a characteristic herringbone architecture that is ubiquitous among members of the *Paramyxoviridae* (Chui and Zhou, 2013; Rima et al., 2019; Plemper and Lamb, 2021; B. Lee et al., 2021). The herringbone architecture observed for the viral nucleocapsids in this family is a result of transmission electron microscopy of the helical nucleocapsid assembly. Since the electrons pass through the particle to reach the detector, the internal structure is superimposed along the electron beam generating an image that is the summation of the obverse and reverse faces of the particle and the internal structure (Figure 4.7B). Paramyxoviral nucleocapsid length is primarily dependent on the size of the viral genome it encapsidates, with larger genomes resulting in longer nucleocapsids. The filamentous tubes observed here are of a wide range of lengths, likely due to the inclusion of random-length *E. coli*-derived RNA in these NLFs. Particles of all three morphologies are consistent in width (the longer axis for squat rings) at around 20 nm. This falls within the range observed for other *Paramyxoviridae* family members (Chui and Zhou, 2013; Rima et al., 2019; Plemper and Lamb, 2021; B. Lee et al., 2021). Furthermore, paramyxoviral nucleoproteins exhibit a high degree of flexibility, likely owing to the Ntail domain as well as Ct and Nt flexible arm interactions. Similarly high degrees of flexibility are observed here for the HeV NLFs.

4.3.2 Comparison with published *Henipavirus* microscopy data

To date, only low-resolution negative stain micrographs of HeV nucleocapsids have been published. When imaged in 2D by negative stain EM, the recombinantly expressed HeV NLFs presented here share many similarities with other recombinantly expressed HeV NLFs (Pearce et al., 2015) as well as HeV nucleocapsids of isolated from cultured virus particles (Hyatt et al.,

2001). The overall architecture exhibits an identical herringbone morphology that is shared between recombinant NLFs and isolated nucleocapsids. Additionally, the circular rings observed here are also present in micrographs from recombinantly expressed protein elsewhere (Pearce et al., 2015). Interestingly, the published data depict filamentous tubes up to approximately 200 nm long whereas the data presented here, and in particular the cryo-EM data, show filamentous tubes vastly exceeding this length. The NLFs shown here may contain longer expression host-derived RNA, possibly due to minor difference in the expression and purification methods used. Pearce *et al.* also describe the expression of full length N protein versus a truncated N protein (Ncore) (Pearce et al., 2015), with NLFs consisting of full length N protein showing a looser coil and overall shorter filamentous tubes compared to NLFs consisting of Ncore. Here, no attempt was made to generate an Ncore-NLF. Lastly, the squat rings presented here have not been described in the published literature. This could be due to a physical absence in HeV nucleocapsid samples in the literature engendered by differences in expression and purification conditions. Alternatively, the squat rings were observed by others but not deemed of consequence to their work.

Recently published data on the closely related NiV (Ker et al., 2021) supports the observations made here (Figure 4.8). The paper describes NLFs of similar dimensions and morphology to those discussed in this work. Interestingly, the paper also investigates so-called

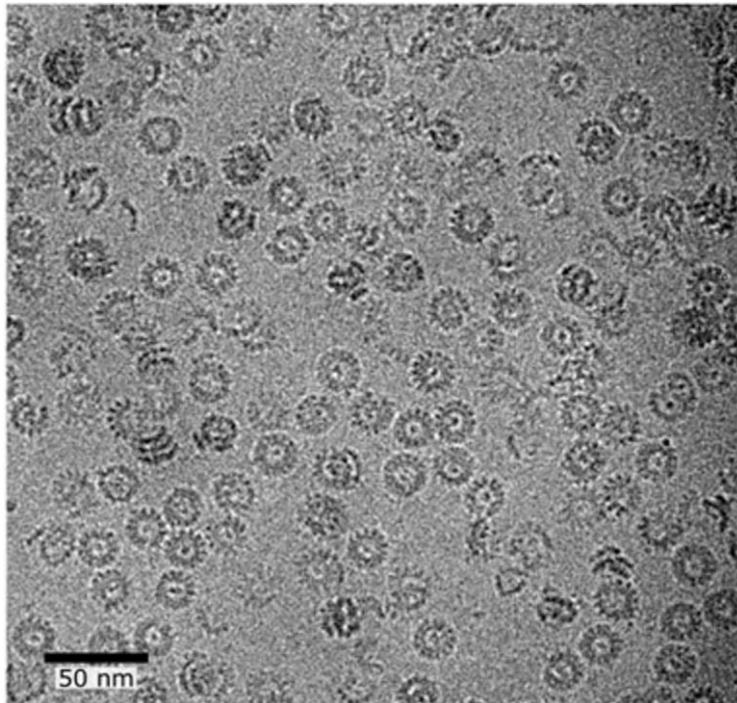


Figure 4.8. NiV nucleocapsid-like filaments resemble those from HeV.

Recombinantly expressed NiV N protein forms NLFs that resemble those formed by recombinantly expressed HeV N protein on cryo-EM grids. Scale bar is 50 nm. Adapted from (Ker et al., 2021).

clam-shape transitions within NLFs that bear resemblance to the squat rings observed here, and further investigates them through high-resolution 3D reconstruction. The similarities between the published data on NiV N protein and NiV NLFs and the high-resolution 3D reconstruction of HeV N protein and HeV NLFs will be discussed in more detail in Chapter 5 Section 5.3.

4.3.3 Summary

This Chapter aimed to characterise the recombinantly expressed HeV N protein through negative stain EM, to describe the optimisation of cryo-grid preparation, and to collect cryo-EM datasets for down-stream 3D reconstruction of HeV N protein and HeV NLFs. The methodology and results presented here reveal that HeV N NLF particles present three predominant morphologies in negative stain and cryo-EM experiments. Additionally, the cryo-grids could be used for full data collection to enable down-stream 3D reconstruction. Lastly, the gross architecture of the NLF particles presented here falls in line with published data on HeV nucleocapsids as well as those from related viruses within the family *Paramyxoviridae*.

Chapter 5: Elucidation of the *Hendra henipavirus* N structure and nucleocapsid architecture

5.1 Introduction

A wealth of information can be gleaned from two-dimensional (2D) images obtained through high-resolution light and/or electron microscopy such as subcellular localisation, molecular interactions, and protein domain organisation. Solving the 3D structure is often less straightforward yet can lead to a complete, biomechanistic understanding of a protein or macromolecular complex at a level of detail that 2D information usually cannot provide. The following sections describe the workflow for single particle analysis (SPA)-based 3D reconstruction from 2D (cryo-)EM datasets (Figure 5.1). The steps described apply to both ‘classical’ SPA as well as helical reconstruction, however it will be explicitly stated whenever these approaches drastically differ. The RELION software suite incorporates a range of popular programmes and algorithms required for 3D reconstruction from EM micrographs (Scheres, 2012). Whilst other suites exist, RELION is a robust and popular choice as it supports a wide range of steps in the 3D reconstruction pipeline; from pre-processing of the raw micrographs to the generation of a 3D electron density map. Indeed, RELION will be used extensively in this chapter.

5.1.1 Pre-processing of micrographs

Before 3D reconstruction can begin, the raw micrographs must undergo a number of pre-processing steps. As described in Chapter 4, data acquired through DEDs generates not just singular micrographs but rather a collection of multiple ‘movie’ frames from a single exposure. Stage drift or beam-induced specimen motion may have shifted the particles’ locations between subsequent frames. While stage drift generally applies a unidirectional translation to the particles, beam-induced motion is non-uniform in its direction and severity. Not only can beam-induced motion occur along different axes, it is also more severe in the early frames of the movie exposure compared to later frames (Brilot et al., 2012). Specimen motion leads to blurry micrographs, analogous to taking a picture of a moving object, and must be corrected to enable high-resolution 3D reconstruction. Further consequences of taking multiple frames should also be considered, such as radiation damage which accumulates over the course of the movie exposure and is most severe in the last frames. Dedicated, RELION-integrated software, such as MotionCor2 (Zheng et al., 2017), algorithmically aligns the micrograph frames to correct for

motion as a result of mechanical stage drift and beam-induced motion during the movie exposure (Figure 5.1). Additionally, the software also applies dose-weighting to the micrograph frames. Doing so down-weights high-resolution information to further mitigate the more severe beam-induced motion from early frames and radiation damage from late frames.

Following motion correction, the defocus and contrast transfer function (CTF) parameters of the corrected micrographs must be estimated to enable CTF-correction during the latter stages of the reconstruction pipeline (see Section 5.1.4) (Figure 5.1). Dedicated, RELION-integrated software including Gctf (Zhang, 2016) and CTFIND4 (Rohou and Grigorieff, 2015) estimate the defocus and astigmatism parameters through fitting of a theoretical CTF model to the power spectrum of the micrograph.

5.1.2 Particle selection

Following pre-processing of the micrographs, the locations of particles of interest on each micrograph must be identified (Figure 5.1). In the early days of cryo-EM reconstruction, particle picking was a time-consuming exercise performed by hand on each individual micro-

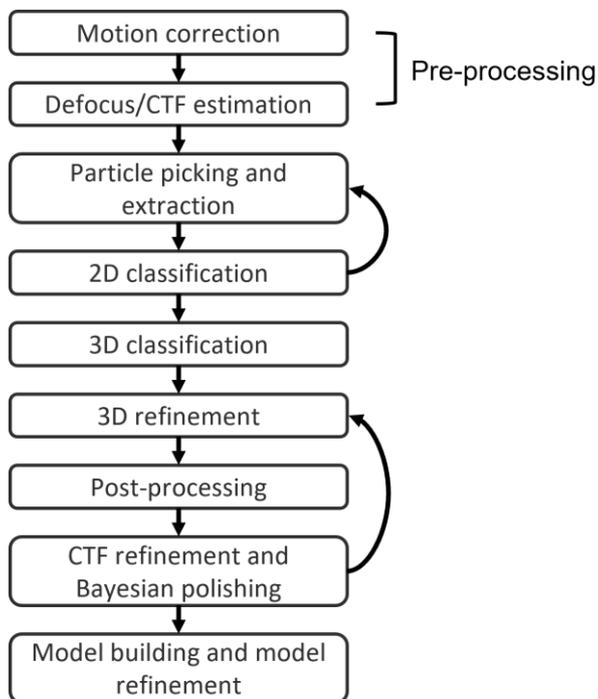


Figure 5.1. Overview of the SPA reconstruction workflow.

A schematic illustrating the typical workflow employed for an SPA-based reconstruction approach. The specifics for helical reconstruction differ from those of ‘classical’ SPA but the overall workflow is the same.

graph in a dataset. The modern procedure is, in most cases, semi-automated for classical SPA while helical reconstruction still requires a more hands-on approach. The RELION software suite incorporates some of the more popular approaches to particle picking, including reference-based and Laplacian of Gaussian (LoG) picking strategies (Scheres, 2012; Zivanov et al., 2018). In the reference-based picking strategy, a relatively small set of particles – usually around 1000 – is manually picked. The aim is to pick features on the micrograph with a high level of certainty to represent specimen particles and which are in different orientations in 3D space relative to the electron beam (Cheng et al., 2015). The selected particles are classified and averaged (see below), with each class average containing all particles with the same (or similar) orientation of the particles. These class averages are then used as a picking reference or template in an automated picking algorithm that attempts to find regions of electron density in the micrographs that match the reference density (Figure 5.1). Manual picking generally results in fewer erroneous particle selections but is exceedingly more liable to selection bias than automated picking. Alternatively, LoG picking is a reference-free approach that avoids the manual picking step entirely and instead picks all electron density features that fall within a pre-specified size range. In recent years, neural network-based particles pickers have been developed, such as crYOLO and TOPAZ (Wagner et al., 2019; Bepler et al., 2019). These algorithms require minimal user input and improve on picking speed and accuracy.

Regardless of the chosen approach, all particle picking strategies provide a set of coordinates that define the position of particles of interest on the micrographs. Once these coordinates have been defined, particles must be extracted from the micrographs (Figure 5.1). Particle coordinates are bounded by a box of user-defined size after which particles are collated to generate a particle stack or projection image for down-stream processing.

The projection image generally has a low SNR which necessitates the alignment and averaging into distinct classes of thousands of images of identical particles exhibiting a range of orientations relative to the electron beam (Cheng et al., 2015; Cheng, 2015). In doing so, the SNR is improved and it is possible to retrieve high-resolution information from the particles. Software suites such as RELION and cryoSPARC implement a maximum likelihood (ML) algorithm for 2D classification over multiple iterations (Figure 5.1). The particles are first classified in a reference-free manner. Using the ML approach ensures that all particles contribute to all classes with different probability weightings. As a result, in the early iterations no strong classifications are made while alignment is uncertain due to relatively noisy references and poor particle alignment. Subsequent iterations continually improve the SNR and alignment of particles to result in a set of ‘true’ particles and ‘junk’ particles. The true particles are classed together,

contributing strongly to a single class or set of highly similar classes with high probability while contribution to all other classes is negligible. The process of (semi-)automated particle picking will inevitably pick features that do not represent true particles. Due to poor alignment – or indeed equally good alignment – with all true particle classes, these junk particles are sorted into noisy, poorly-populated classes. Doing so allows for removal of the junk particles from the particle stack leaving a cleaner dataset for down-stream processing. In practice, during the first classification there is often a gradient of class quality, with a small number of junk classes and high-quality true particle classes, and a large number of true particle classes of intermediate quality. Performing multiple rounds of 2D classification enables the stepwise selection of the largest number of particles contributing to the highest-quality classes.

Particle selection for helical reconstruction is perhaps most similar to the template-based approach, with the exception that in RELION helical segments are picked by selecting the start and end coordinates of a helical filament. Particles, or segments, are then extracted along this straight line based on pre-defined parameters which requires *a priori* knowledge of the helical parameters describing the filaments. Modern software packages can still struggle to pick helical segments accurately and in some cases manually picking remains the only viable approach.

5.1.3 Three-dimensional reconstruction

Reconstructing a 3D model of an object from 2D projection images of the same object is non-trivial and is enabled by the Fourier slice theorem. The theorem describes that the Fourier transform of an object's 2D projection image is identical to the central slice through a 3D Fourier transform of the same object (Figure 5.2A). The central slice follows the projection vector so that, if the 2D projection images' projection vectors are known, the slices can be assembled in 3D and performing an inverse Fourier transform would reconstruct a real-space 3D object (Sigworth, 2016). With perfect data, this procedure would generate a perfect 3D reconstruction. In practice, cryo-EM data processed following the SPA workflow is imperfect. A consequence of averaging particles to improve SNR is that the projection vectors for these particles are unknown. This problem can be ameliorated by performing projection matching, where an initial low-resolution reference map of approximately the correct size and shape of the object of interest is back-projected from all angles (Figure 5.2B). These computed images can then be used to compare the images in the experimental dataset with to estimate projection vectors for the experimental dataset (Figure 5.2C). Using the projection vector estimates, the experimental

images are inversely Fourier-transformed to generate a more accurate reference map and iterating this process incrementally improves angular assignments for the particles (Figure 5.2D-E). As a result, each iteration improves upon the previous map and moves towards a reconstruction based exclusively on experimental data. In Relion, the process of angular assignment employs a ML algorithm to account for uncertain alignments in the early iterations (Nogales and Scheres, 2015).

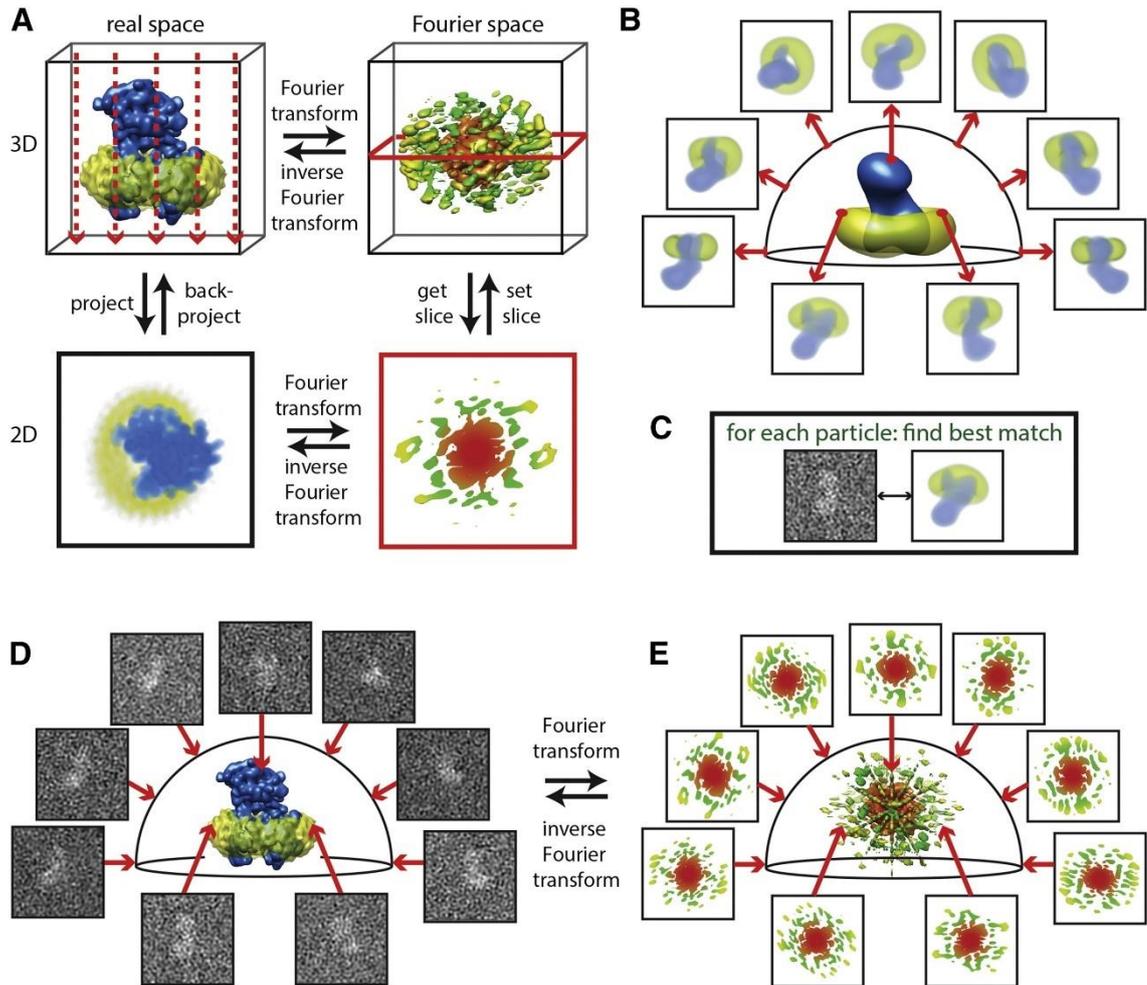


Figure 5.2. The Fourier slice theorem and concepts of 3D reconstruction.

A) The Fourier slice theorem describes how the 2D projection (bottom left) of a real-space 3D object (top left) is equivalent to the central 2D slice (bottom right) of the transformed object in 3D Fourier space (top right). The projection vector in real space (top left, dashed red arrows) is perpendicular to the Fourier slice (right, red frame). **B-E)** In the process of projection matching, experimental 2D projections can be combined into a 3D reconstruction. **B)** First, reference projections are calculated from an initial reference model of roughly the correct shape and size to establish reference projection vectors. **C)** Each experimental projection (particle image) is compared to all reference projections to find a projection match. **D)** Establishing projection matches then orients the experimental projections relative to the real-space 3D structure following the projection vectors determined between **B** and **C**. **E)** Following the Fourier slice theorem means that by positioning the 2D Fourier transforms of all experimental projections (2D slices) into the 3D Fourier transform enables the calculation of the real-space 3D reconstruction. The orientations of the projections, and thereby the resolution of the reconstruction, are gradually improved by iterating the steps visualised in panels **B-E**. Adapted from (Nogales and Scheres, 2015).

Once a particle stack has been curated through multiple rounds of 2D classification it is important to assess the presence of heterogeneity in the particles due to conformational or compositional differences. Similar to the 2D classifications, RELION employs an unsupervised 3D classification algorithm to solve this problem, using ML and requiring no prior information on heterogeneity (Figure 5.1) (Scheres, 2016). The 3D classes are first randomly populated by the particles in the selected dataset and projection matched to a low-resolution reference map, resulting in a collection of classes with minor heterogeneity (Figure 5.2C-D). Proceeding iteratively, the classification process compares all particles with all 3D class averages resulting in increasing differentiation between the 3D classes based on true structural heterogeneity in the particles rather than the initial random assignment. Multiple rounds of 3D classification may be required to separate out a truly homogeneous particle stack (Scheres, 2016).

It is then possible to attempt 3D refinement to generate a high-resolution reconstruction from the homogeneous particle stack. Refinement employs projection matching as described above and, in each iteration, the angular and translational sampling is increased until no improvements are made in estimated resolution and accuracy of angular assignment. This plateau is also referred to as convergence. During 3D refinement, the particle stack is randomly split into two subsets each of which is refined separately. Treating the two subsets independently improves the accuracy of resolution estimation through Fourier shell correlation (FSC) since the correlation of high-resolution noise between the two half maps is unlikely and the risk of overfitting is reduced (Scheres and Chen, 2012).

5.1.4 Post-processing and refinement of three-dimensional maps

A combination of physical and computational factors imparts a loss of information at high spatial frequencies on maps generated by 3D refinement. These factors are predominantly radiation damage, particle movement, and uncertainties in projection matching. Estimating the rate of loss of high-resolution information at resolution higher than 10 Å provides a blurring factor, or B-factor, which can then be used to correct for the loss of information by up-weighting the remaining high-resolution information components in the 3D reconstruction (Rosenthal and Henderson, 2003). For these reasons, and because the map resolution is underestimated in 3D refinement, maps should always be post-processed (Figure 5.1). In post-processing, the application of masks to the half maps excludes the noisy solvent regions, increasing the FSC and as a result the estimated resolution.

The reconstruction resolution may be further improved through CTF refinement and Bayesian polishing (Figure 5.1), both of which are implemented in the RELION software suite (Scheres, 2012; Zivanov et al., 2018). Firstly, CTF refinement encompasses the estimation and correction of beam tilt aberrations and symmetrical and asymmetrical aberrations on the level of individual particles. Doing so improves accuracy especially when distribution along the beam axis within the ice layer is heterogeneous. Secondly, Bayesian polishing tracks particle motion on an individual level enabling the mapping and correction of complex specimens. As both CTF refinement and Bayesian polishing benefit from higher resolution maps as the input, it is good practice to iterate a loop of 3D refinement, post-processing, CTF-refinement, and Bayesian polishing until map resolution reaches a plateau.

5.1.5 Model building and refinement

The aim of reconstructing a 3D map is to ultimately build a 3D model of the specimen (Figure 5.1) and the steps discussed in the previous sections attempt to generate a map with the best achievable resolution. Generally speaking, a higher resolution map 3D map eases model building as there are fewer ambiguities to contend with when fitting primary, secondary, and tertiary structural features. Cryo-EM density maps with resolutions of around 3.8 Å and better enable atomic model refinement as from this resolution onwards amino acid side chain densities can be visualised (Cheng, 2015). The availability of prior information on the specimen itself or close homologues will direct much of the model building approach. For example, if a model of the specimen has previously been determined it can be rigid fit into the 3D density map. Published model structures are readily available from web-based databases such as the Protein Data Bank (PDB) or the Electron Microscopy Data Bank (EMDB). In the absence of these model structures, structural prediction software such as SWISS-MODEL (Waterhouse et al., 2018) or AlphaFold2 (Jumper et al., 2021) can be employed to generate a starting model. Finally, very high-resolution maps also enable *de novo* model building where no input model is used.

Once a starting model has been generated, software packages such as Coot allow for the manual adjustment to improve the model fit within the experimental map while satisfying backbone geometry constraints (Emsley and Cowtan, 2004; Emsley et al., 2010). Following manual editing, the model is then refined in automated real-space refinement using, for example, Phenix (Adams et al., 2010; Afonine et al., 2018) or Refmac (Murshudov et al., 2011). In order to obtain the optimal model several rounds of manual adjustment and automated refinement may be required. The model quality should then be assessed by validation algorithms

such as MolProbity (Chen et al., 2010; Williams et al., 2018) which can be run separately or as integrated features within automated real-space refinement. Model building is never truly finished, but a model can be considered complete when the model quality is as good as reasonably achievable from the experimental data and within the timeframe available.

5.1.6 Chapter aims

This Chapter aims to describe reconstruction of the 3D electron density maps from the HeV NLFs found on cryo-EM micrographs and the building of a model of the HeV N protein monomer, as well as the HeV NLF architecture. The protocols covering the results in this section are covered in Chapter 2 Materials and Methods, Sections 2.2.4 and 2.2.5.

5.2 Results

5.2.1 HeV nucleoprotein sauronoids can be extracted and 2D classified

The complete dataset of 3548 micrographs was collected using an automated approach at a pixel size of 1.065 Å and pre-processed on the fly during collection using the RELION software suite (Scheres, 2012; Thompson et al., 2019). Motion correction and CTF estimation were carried out using RELION-integrated implementations of MotionCor2 (v1.2.1) (Zheng et al., 2017) and Gctf (v1.18) (Zhang, 2016).

Following pre-processing, particles were manually selected using the manual picking job which adhered to the 'classical' SPA procedure (Figure 5.3A-D). A total of 3500 particles were selected from a random subset of 502 micrographs to generate a reference for automated picking through 2D classification (Figure 5.4). Subsequently, reference-based automated picking was optimised and the final job selected 875,262 particles from all 3548 micrographs (Figure 5.3E-F) which were extracted using a box size of 360 pixels (383.4 Å). Automated picking inevitably still selects particles from the micrographs that do not represent true particles (Supplementary Figure S5.1). These 'junk' particles – as well as true particles that were not of interest for this reconstruction, such as NLFs – were filtered out over multiple rounds of 2D classification in which the input of each round of classification is a curated set of particles from the previous round. Doing so ensures the retention of the largest number of true particles while filtering out junk particles.

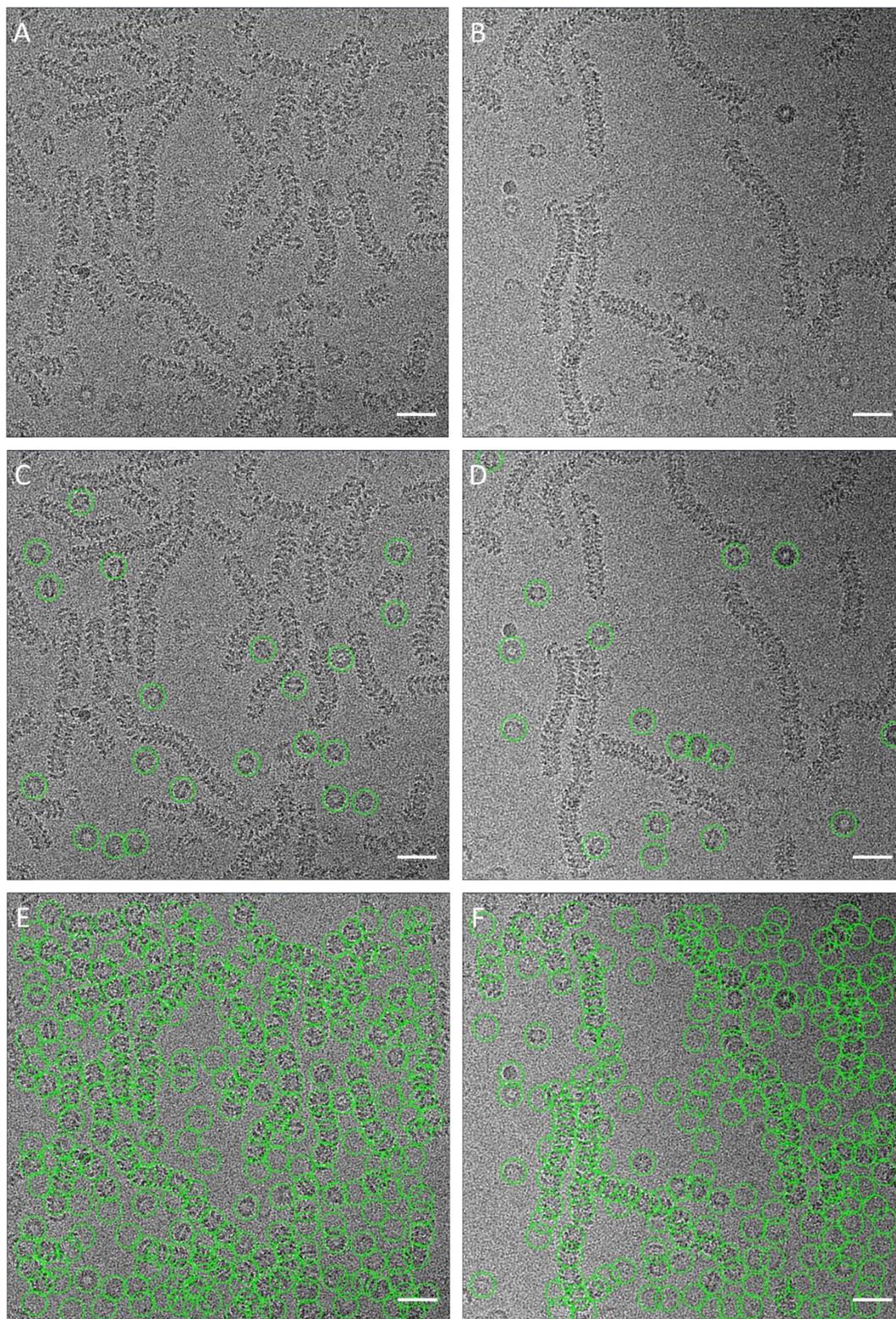


Figure 5.3. HeV single particle selection through manual and automated picking.

Representative micrographs illustrating the particle selection steps. **A-B)** Dataset micrographs depict the same morphologies observed in negative stain micrographs and cryo-EM micrographs from the screening session (see Figures 4.3 and 4.6). **C-D)** Manual picking (green circles) allows for the selection of circular rings and squat rings to generate an automated picking reference through 2D classification. **E-F)** Optimised, reference-based automated particle picking selects a high number of features per micrographs (green circles). The selected features include particles of all three morphologies and inevitably also background signal. Scale bars are 50 nm.

Over the course of multiple 2D classification rounds the morphological features of the squat rings and circular rings gained definition due to improved SNR (Figure 5.5). These 2D classes showed secondary structure elements, indicative of high-quality data. The overall morphology of both ring types indicated that they likely represent different viewing angles of the same N protein assembly, with the squat rings and circular rings providing side-on and top-down views, respectively (see Figure 4.7A). Indeed, 2D classification revealed a third angle seemingly intermediate between the side-on view and the top-down view (Figure 5.5A). The top-down view reveals a tetradecameric ring structure, however the distinctive side-on appearance of the assembly inspired the name ‘sauronoid’, after the Great Eye of J.R.R. Tolkien’s legendarium (Tolkien, 1954). The high-resolution 2D classes revealed the presence of two axes

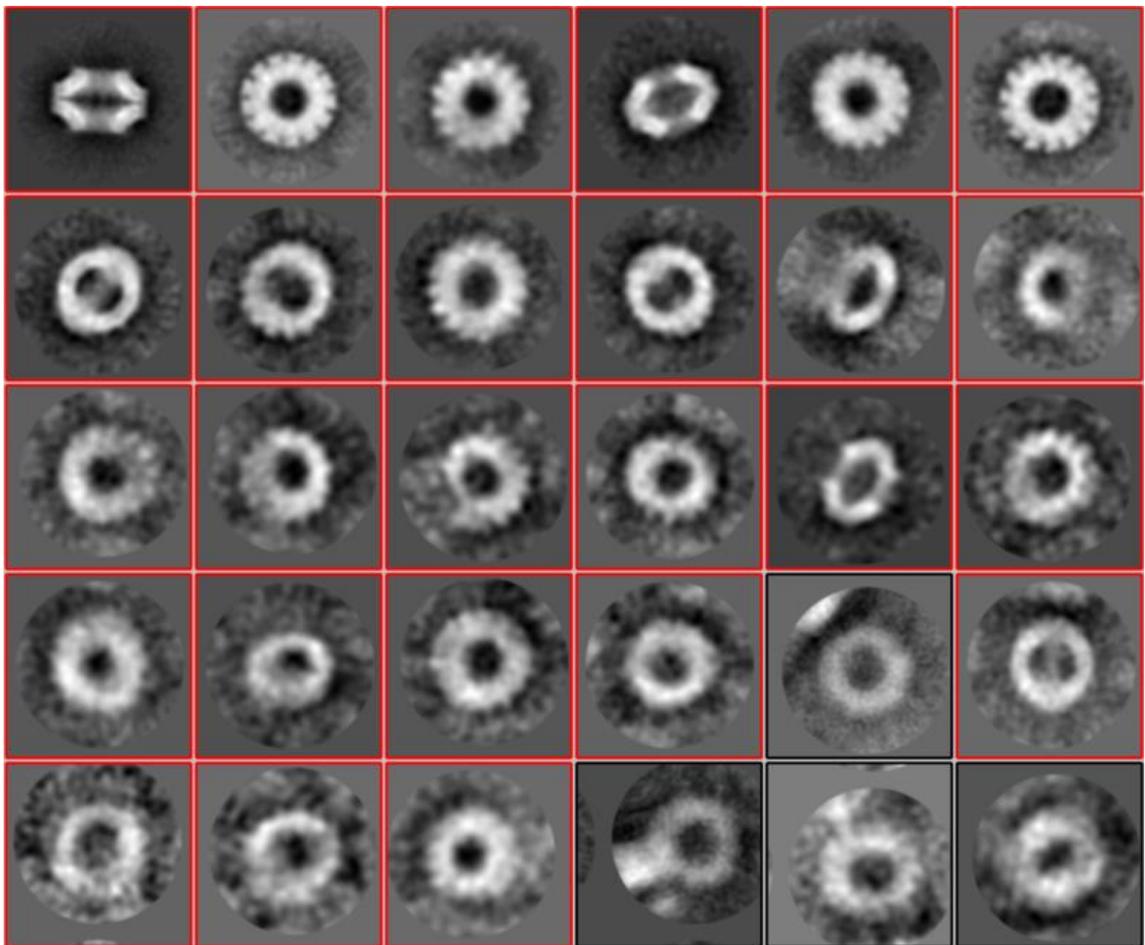


Figure 5.4. Sauronoid 2D classification of manually picked particles generates reference for automated picking.

A set of 3500 HeV sauronoid particles was manually selected following the classical SPA procedure from a random subset of 502 micrographs and 2D averaged into 30 classes. The displayed classes are sorted on ‘in class distribution’ where the most populated classes appear top left and class population decreases from left to right and top to bottom. Manual picking generally results in fewer junk particles and the 2D classes bounded by red boxes were selected to constitute the automated picking reference. Box boundaries are 360 pixels (383.4 Å).

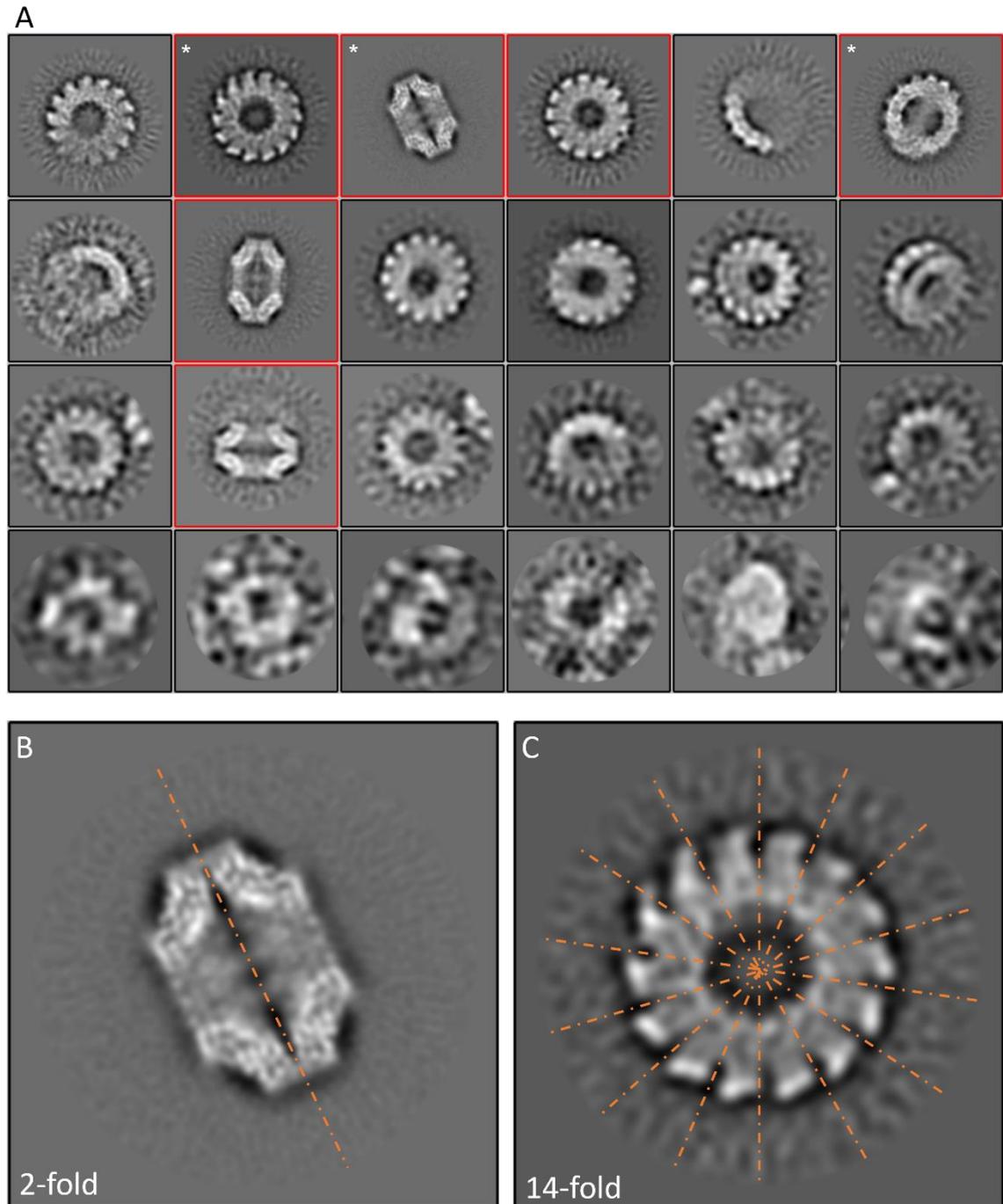


Figure 5.5. HeV sauronoid 2D classification reveals a range of viewing angles and detailed 2D morphology.

A subset of 20,809 automatically picked sauronoid particles is 2D classified into high-resolution classes revealing interior detail. Box boundaries are 383.4 Å (360 pixels). **A**) The displayed classes represent a range of viewing angles and are sorted on ‘in class distribution’ where the most populated classes appear top left and class population decreases from left to right and top to bottom. The three classes marked with an asterisk represent from left to right: the top-down or central hole view, the side-on or iris view, and an intermediate viewing angle between the central hole and iris views. The 2D classes bounded by red boxes, containing 20809 particles, were selected for 3D reference map construction and 3D classification. **B-C**) The 2D classes representing the iris view (**B**) and central hole view (**C**) reveal a 2-fold axis of symmetry and 14-fold axis of symmetry, respectively (orange dotted lines). These axes are oriented perpendicularly with respect to each other.

of symmetry (Figure 5.5B-C). The first is a two-fold axis of rotational symmetry (C2) which aligns with the side-on view, or sauronoid 'iris' (Figure 5.5B). The second is a fourteen-fold axis of rotational symmetry (C14) which is perpendicular to the C2 axis and aligns with the top-down view, or sauronoid central hole (Figure 5.5C). These two axes of symmetry combine into a dihedral symmetry of D14 for the sauronoid assembly.

5.2.2 The HeV sauronoid assembly results in a 3.5 Å electron density map

The generation of high-quality 2D classes with good SNR enabled the building of a 3D electron density map of the sauronoid assembly. An initial reference map could be generated within RELION from reference-free projection matching of the selected 2D classes (Figure 5.5A) and using the established D14 symmetry (Figure 5.5B-C).

An initial reference map was generated *de novo* by stochastic gradient descent (SGD) and used for 3D classification and 3D refinement (Supplementary Figure S5.2). Early rounds of 3D classification and 3D refinement struggled to generate accurate projection matches with the reference map. The resulting maps suffered from low particle numbers and failed to attain high resolution, not exceeding 18 Å. While improving estimated resolution, increasing degrees of imparted symmetry (C1, C2, C7, and C14) resulted in map reconstructions that are incongruent with the observed 2D classes (Figure 5.6). Inspection of the initial reference map revealed that it was grossly misaligned with respect to the two-fold and 14-fold axes of symmetry. Symmetry axes in RELION are fixed and gross misalignment of the initial reference map abrogates accurate projection matching. Manual re-alignment and resampling of the C1 symmetry refined 3D map (Figure 5.6A) to the grid of the unaligned map in UCSF Chimera (Pettersen et al., 2004) fixed axis alignment and generated a useful reference map. Employing the re-aligned reference map for symmetrised 3D classification – using D14 symmetry – resulted in an immediate and drastic improvement in assignments and map quality (Figure 5.7). The most favourable 3D class, containing only 5683 particles (Figure 5.7A), enabled initial 3D refinement to 4.0 Å resolution (Figure 5.8A). As a result of imparting D14 symmetry on the 3D classification and refinement there exists a 28-fold redundancy in particle counts. As such, an effective 157,864 asymmetric units contributed to the 3D reconstruction of the sauronoid map. In post-processing, a solvent mask was applied to the refined sauronoid map (Supplementary Figure S5.3) to counter the influence of high-resolution noise which saw an improvement in map quality and an increase in estimated resolution to 3.5 Å (Figure 5.8B, Supplementary Figure S5.4, Supplementary Figure S5.5, Supplementary Figure S5.6A).

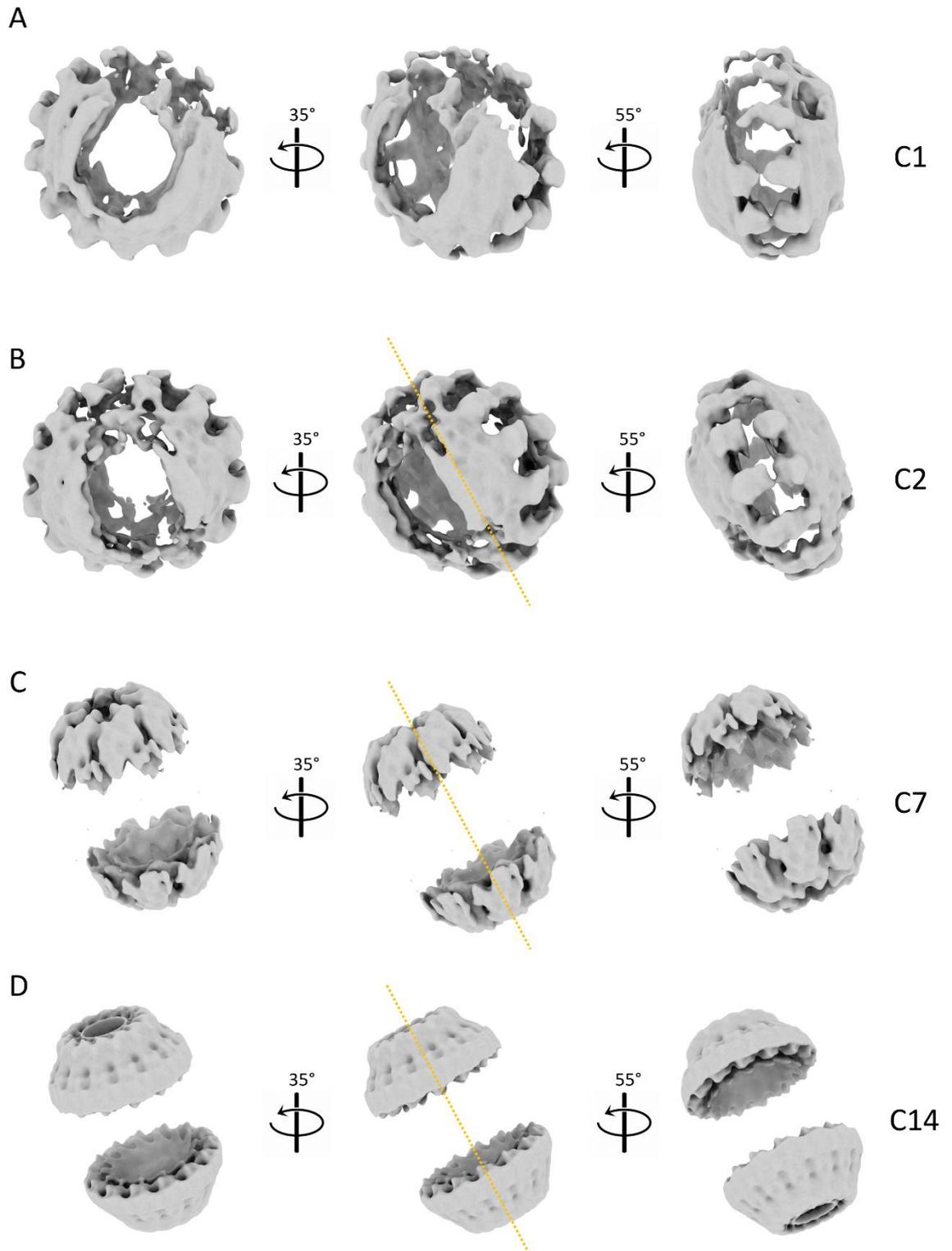


Figure 5.6. Impartation of symmetry generates incongruent maps.

Symmetrised refinement of a non-symmetrical 3D class (Supplementary Figure S5.2B) generated increasingly aberrant density maps while increasing estimated resolution. All four maps follow the native orientations, highlighting the misalignment along the two axes of symmetry. **A)** Refinement with no symmetry (C1) resulted in a map with poor density and an estimated resolution of 18.3 Å. **B)** Refinement with C2 symmetry failed to improve density and had an estimated resolution of 10.4 Å. **C-D)** Refinement with C7 or C14 symmetry generated maps incongruent with the 2D classes, with estimated resolutions of 9.1 Å and 8.7 Å, respectively. Symmetry axes are indicated in yellow.

The final reconstructed map reveals global morphological features at high-resolution. Firstly, the central-hole face shows 14 distinct N protomers (Figure 5.8, Supplementary Figure S5.5). These protomers consisting largely of structural features resembling alpha-helices. Secondly, the N protomers are seen to intercalate when viewed from the iris face, indicating potentially extensive interactions between the two tetradecameric rings (Figure 5.8). Lastly, the RNA-binding groove – identified by comparing to homologous N protein structures – shows clear electron density, suggesting it may be occupied by RNA (Figure 5.8B, Supplementary Figure

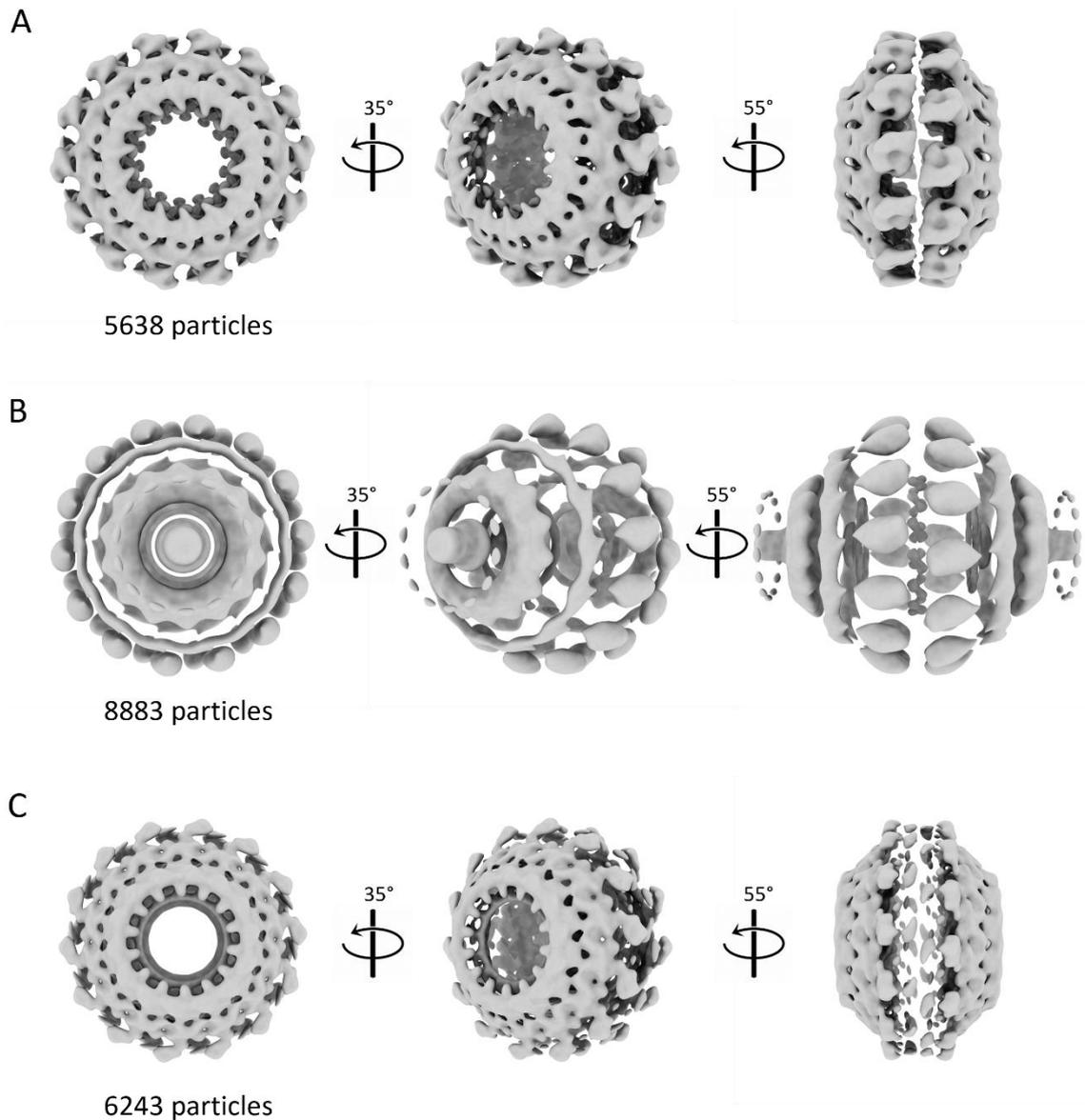


Figure 5.7. 3D classification generates favourable map for refinement.

3D classification employing D14 symmetry generated maps suitable for 3D refinement. Classification of the total particle stack (20,764 particles) into three classes resulted in comparable particle distributions (A-C). Although the first class (A) had the lowest particle number out of the three classes generated, it also had the most favourable density for 3D refinement.

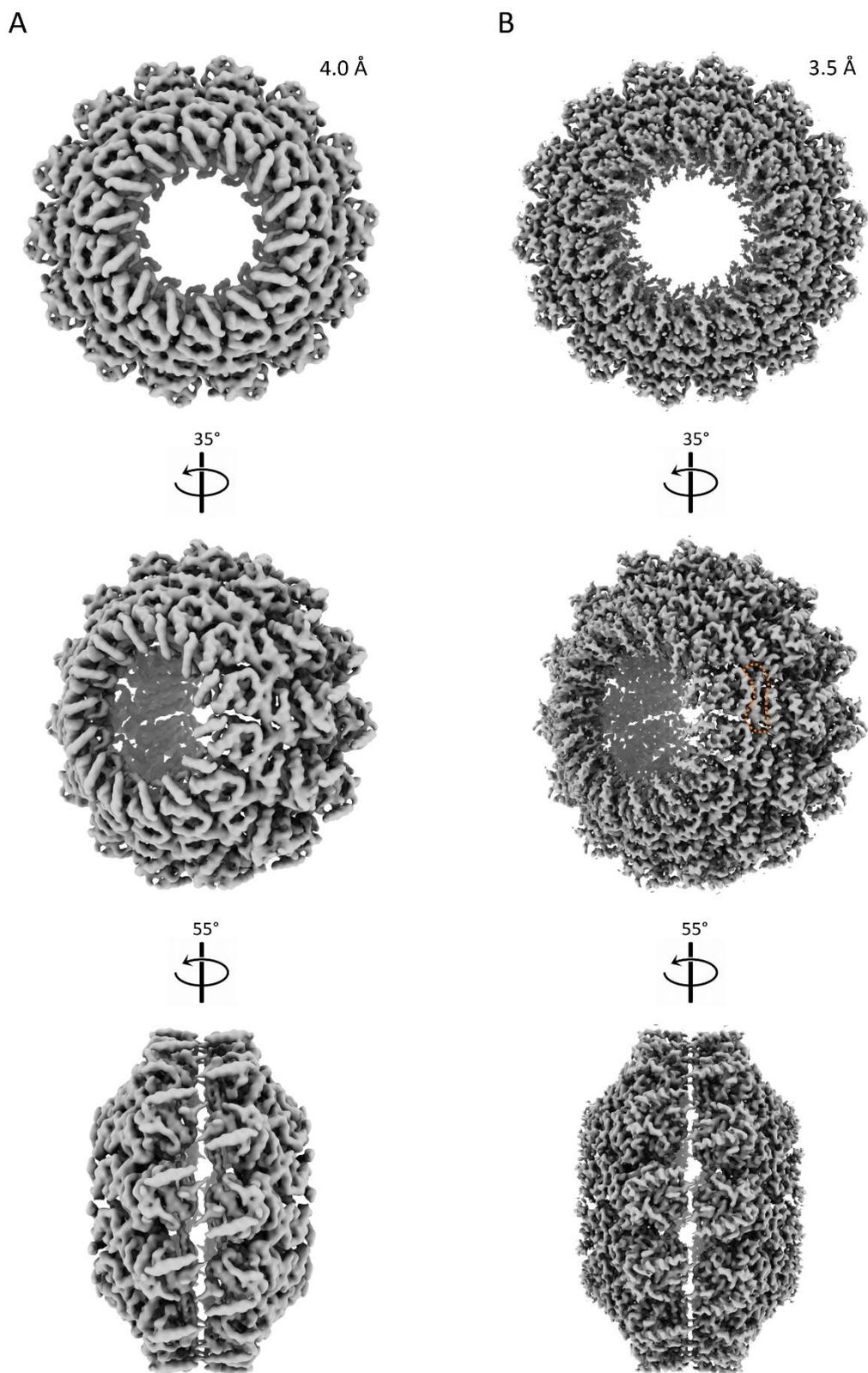


Figure 5.8. Refined sauronoid electron density maps reveal global architecture and secondary structure.

3D refinement employing D14 symmetry generated maps suitable molecular modelling. **A)** The initial 3D-refined sauronoid map, employing D14 symmetry and based on a D14 symmetrised 3D class, reached 4.0 Å resolution. The map reveals clear density for alpha helices. **B)** The final 3D-refined sauronoid map following post-processing reached 3.5 Å resolution. The RNA-binding groove is occupied by electron density (orange).

S5.5). The presence of host-derived RNA (hRNA) in the RNA-binding groove is in line with the 260/280 ratio measured for the sample used in this data collection (see Chapter 3 Section 3.2.4). Angular distribution analysis of the particles contributing to the map revealed a preferential orientation showing the iris face, however there were sufficient contributions from other angles including central-hole face and intermediate orientations (Supplementary Figure S5.6B). The final sauronoid EM map was deposited with the EMDB under accession code EMD-16426.

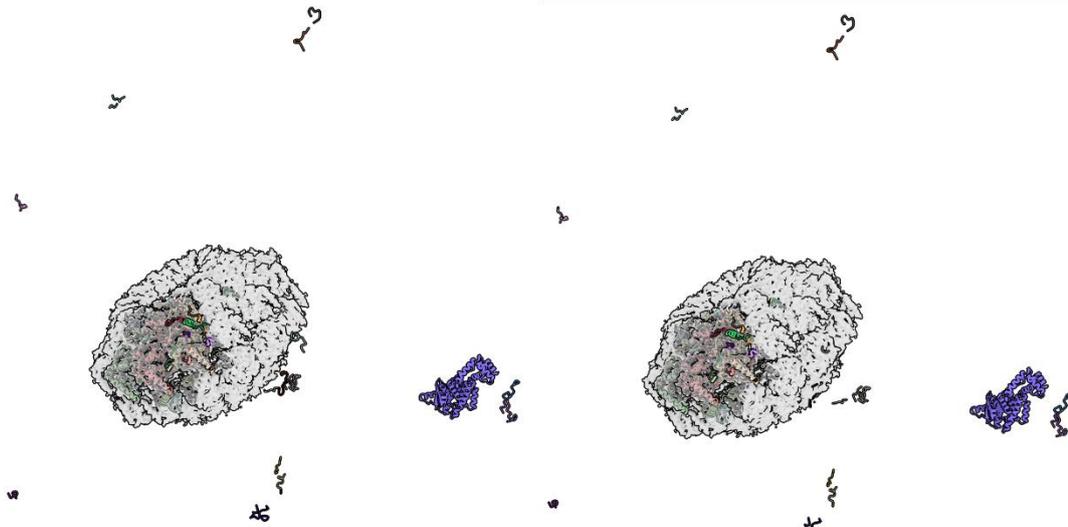
5.2.3 A HeV nucleoprotein model can be generated from the sauronoid map

At 3.5 Å, the reconstructed sauronoid electron density map was of sufficiently high resolution to enable atomic model refinement (Cheng, 2015). An initial homology model was generated using the structure prediction software SWISS-MODEL (Waterhouse et al., 2018) with the HeV N protein amino acid sequence (Supplementary Figure S3.2) as input. The SWISS-MODEL pipeline incorporates sequence alignment and homology searches, which resulted in a best match with the crystal structure of the closely related NiV N protein (PDB: 4CO6) (Yabukarski et al., 2014). This co-crystal structure features an incomplete apo-NiV Ncore missing the Nt- and Ct flexible arms and includes a liganded NiV P protein Nt domain (α 1-2, see Chapter 1 Section 1.7.2.2). Following selection of this sequence map, SWISS-MODEL constructed an initial HeV N protein homology model (Supplementary Figure S5.7). The homology model was then used as input for automated model building in Buccaneer (Cowtan, 2006) (Figure 5.9). The resulting model included 59 separate peptides and was a good starting place for manual refinement, despite discontinuous peptide chains occupying the electron density map and the large number of erroneously built short peptides (Figure 5.9).

The Buccaneer model was imported into Coot (Emsley and Cowtan, 2004; Emsley et al., 2010) for manual refinement. First, the Buccaneer model was curated by deleting peptide chains not occupying the sauronoid map. Next, peptide chains that occupied the density were either deleted or combined until a single, continuous N protomer model remained. The electron density in the RNA-binding groove was occupied by short peptide with a poor fit. The clear density was continuous between adjacent monomer densities and of high-resolution, enabling unambiguous, *de novo* building of a hexameric RNA chain into the now unoccupied density. The density around the RNA bases enabled building of the RNA chain in the 3' to 5' direction based on related N protein structures but made no suggestion towards RNA sequence and, following nucleoprotein structure convention, the six bases were modelled as a poly-uridine chain (Figure 5.10, Supplementary Figure S5.8).

The RNA-bound N protomer model was then used as input for automated real-space refinement in Phenix (Adams et al., 2010; Afonine et al., 2018). The monomeric N protomer was first symmetry expanded in UCSF Chimera following D14 symmetry, prior to real-space refinement. Doing so is more computationally demanding, as the algorithm has to refine 28 protomers rather than one, but prevents erroneous real-space refinement into non-cognate density in the multimeric sauronoid map as the otherwise vacant electron density neighbouring

A



B

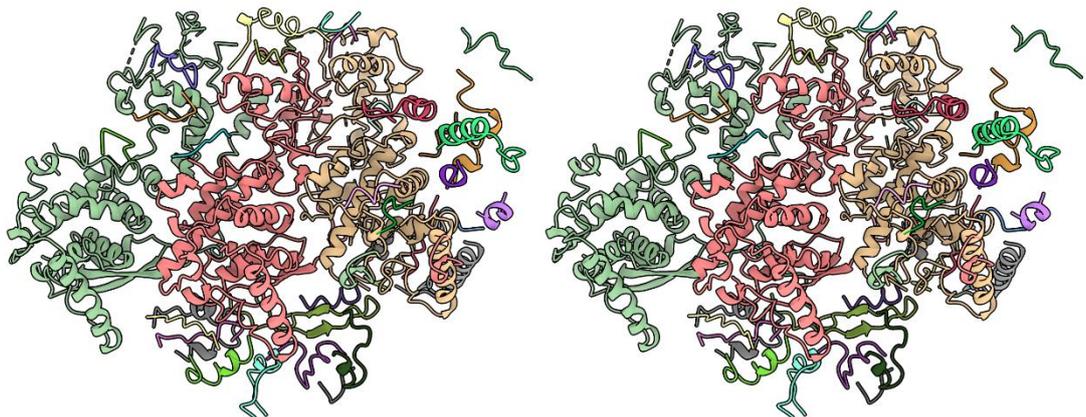


Figure 5.9. Automated model building in Buccaneer.

A HeV model was built into the sauronoid density map based on the HeV homology model. Continuous peptide chains are individually coloured. **A-B)** Stereographic depictions of the automatically built HeV N protein model. Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis. **A)** Overview of all chains built into the sauronoid density however a large number of short peptide chains as well as a nearly complete N protomer chain exist outside the map. **B)** The peptide chains occupying the sauronoid map cover most of the density, but most are discontinuous peptide chains.

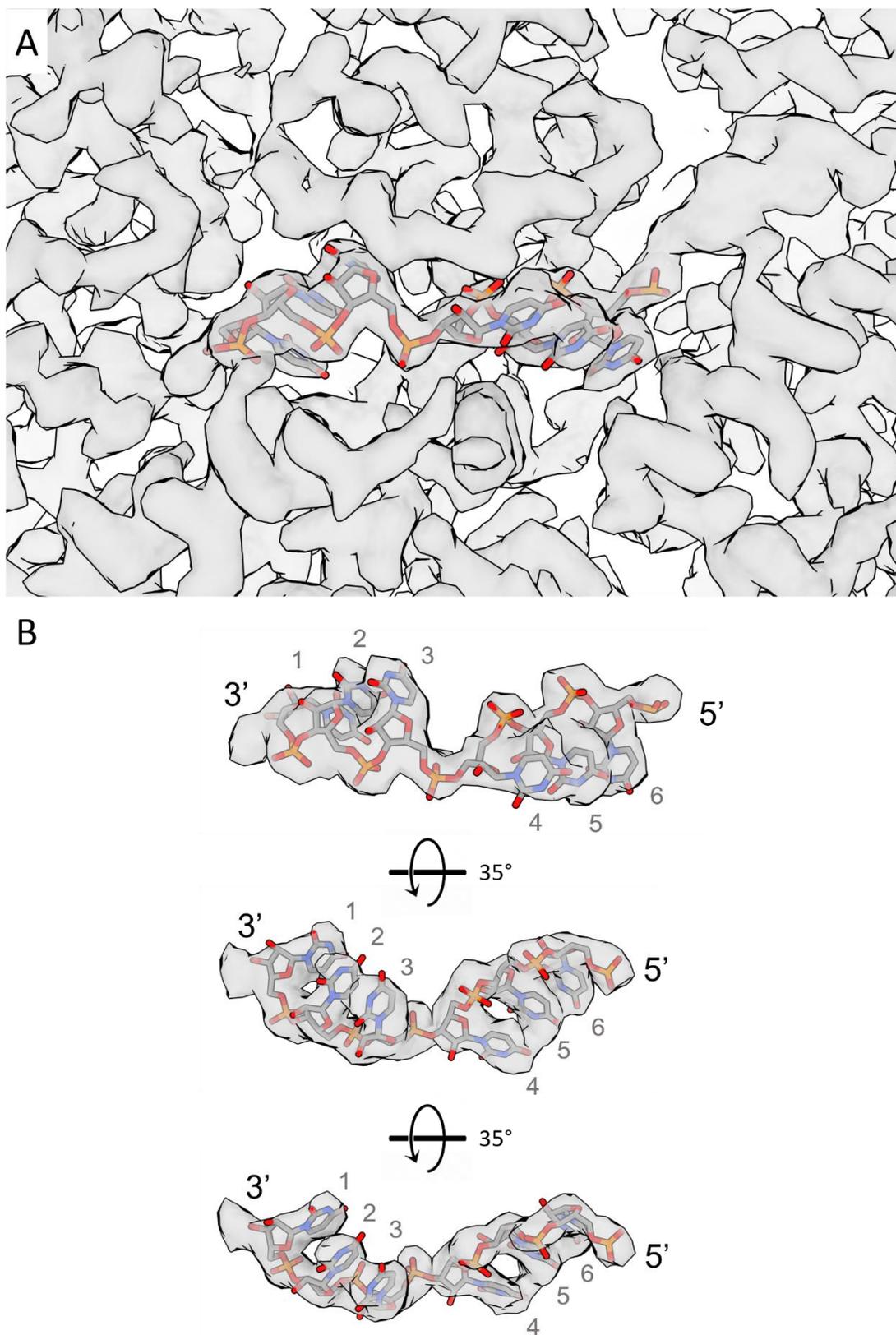


Figure 5.10. *De novo* RNA structure in the RNA-binding groove.

A) Cryo-EM density surrounding the N protein RNA-binding groove depicting the *de novo* built hexa-uridine chain (sticks). B) The density occupying the RNA-binding groove was isolated from the map and shows the built hexa-uridine RNA chain. The RNA 3' and 5' ends are indicated. Uridine nucleotides are numbered from the 3' end and exhibit two characteristic base stacks, one facing in towards the N protein density (bases 1-3) and one facing out towards the solvent (bases 4-6).

the N protomer will be occupied by other N protomers. Following real-space refinement, the dimer of tetradecamers was imported into Coot where manual adjustments were made on a single RNA-bound N protomer. Once the RNA-bound N protomer was satisfactorily adjusted by hand, the 27 unadjusted protomers were removed from the model before symmetry expansion was carried out once more. This process accelerates manual model building by focussing on only one protomer and is methodologically sound as the experimental map follows the same D14 symmetry. The manual refinement and automated real-space refinement loop was iterated 12 times before the model was considered complete. In the early loops, manual refinement focussed on improving the fit of the RNA chain and peptide backbone into the electron density, ensuring appropriate assignment of alpha-helices and beta-sheets. Later iterations focussed manual refinement to fit amino acid side chains in the density as well as improving rotamer orientations. Poor local density and resolution abrogated confident side chain placement (Supplementary Figure S5.6C) and as a result forced the stubbing of a number of amino acids, modelling them as alanines in the final model. These stubbed residues are Arg23, the loop Met114-Asp119, and the helix Leu384-Leu392. Finally, model validation was carried out as part of the ultimate real-space refinement job through MolProbity (Chen et al., 2010; Williams et al., 2018) integrated within Phenix (Table 5.1, Supplementary Table S5.1, Supplementary Figure S5.9). The final model of the asymmetric protomer unit was deposited with the PDB under accession code 8CBW. An assembly of the complete sauronoid with contiguous RNA was deposited with the PDB under accession code 8C4H.

5.2.4 The HeV sauronoid model elucidates N-N and N-RNA interactions

High-resolution map refinement and protein model building enabled investigation of the overall protein fold, as well as the protein-protein and protein-RNA interactions. The multimeric sauronoid model reveals a dimer of tetradecameric rings each of which contains 84 RNA residues, for a total of 168 RNA residues in the sauronoid assembly (Figure 5.11). No density was observed in the sauronoid map corresponding to the intrinsically disordered Ntail region (residues 395-532) and the final model includes HeV N protein residues 1 to 394 (Figure 5.12).

The multimeric assembly reveals that each tetradecameric ring is topped with the Ct-arms which cover part of the globular CTD, while the globular NTDs intercalate to form a dimer of tetradecamers (Figure 5.12C). The Nt-arm lies completely within the sauronoid central cavity where it crosses over the NTD and interfaces with the back face of the CTD (Figure 5.12D).

Table 5.1 HeV sauronoid nucleoprotein model validation

Data processing	
Symmetry imparted	D14
True particle images (no.)	5638
Symmetrised particle images (no.)	157,864
Estimated map resolution (Å)	3.48545
FSC threshold	0.143
Map sharpening B factor (Å ²)	-132.352
Model composition	
Non-hydrogen atoms	87752
Protein residues	11032
Nucleic acid residues	168
Model validation	
RMS Bond lengths (Å)	0.002 (0)
RMS Bond angles (°)	0.502 (0)
MolProbity score	1.72
Clash score	9.57
Rotamer outlier (%)	0.13
Ramachandran Favoured (%)	96.63
Ramachandran Outliers (%)	0.00
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N=10976)	0.80 (0.08)
helix (N=6524)	1.95 (0.06)
sheet (N=504)	-0.11 (0.22)
loop (N=3948)	-1.95 (0.08)
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	3.68
ADP (B-factors)	
Iso/Aniso (#)	87752/0
Protein (min/max/mean)	14.79/167.80/65.80
Nucleotide (min/max/mean)	52.67/104.15/71.05
Ligand (min/max/mean)	---
Water (min/max/mean)	---
Occupancy	
Mean	1.00
occ = 1 (%)	100.00
0 < occ < 1 (%)	0.00
occ > 1 (%)	0.00

5.2.4.1 The HeV nucleoprotein fold consists of two globular domains and two terminal arms

Globally, the HeV N protomer structure follows the common paramyxovirus nucleoprotein fold (Figure 5.12B, Supplementary Figure S5.10). Two large globular domains – the NTD and CTD – are predominantly alpha-helical and are separated by an RNA-binding groove. The N protein fold contains a total of 17 alpha-helical segments (α 1- α 17) with ten helices

in the NTD, five in the CTD, and one each in the Nt- and Ct-arms (Supplementary Figure S5.10). The NTD also contains a parallel beta sheet (Figure 5.12B, Supplementary Figure S5.10) with the first beta strand ($\beta 1$) sequentially positioned at the boundary of the NTD and the Nt-arm and the second beta sheet ($\beta 2$) sequentially separated from the first by three alpha helical stretches ($\alpha 2$ - $\alpha 4$) (Supplementary Figure S5.10B). The RNA binding groove contains hydrophobic patches and is mostly positively charged near the RNA backbone and negatively charged near the RNA bases (Figure 5.13). Two flexible regions, the Nt- and Ct-arms, extend beyond the globular domains. Both terminal arms are linked to the globular domains through flexible peptide loops and each forms an alpha helix in proximity to the CTD (Figure 5.12B). The Nt-arm alpha helix ($\alpha 1$) contains a strongly negatively charged, solvent exposed region facing the sauronoid central cavity consisting of residues Asp3, Asp6, and Glu7 (Figure 5.13A, Figure 5.14A, Supplementary Figure S5.11). The N protein surface area is overall hydrophilic, with three prominent hydrophobic patches that are covered by neighbouring N protomers in the sauronoid assembly (Figure 5.13B, Figure 5.14B).

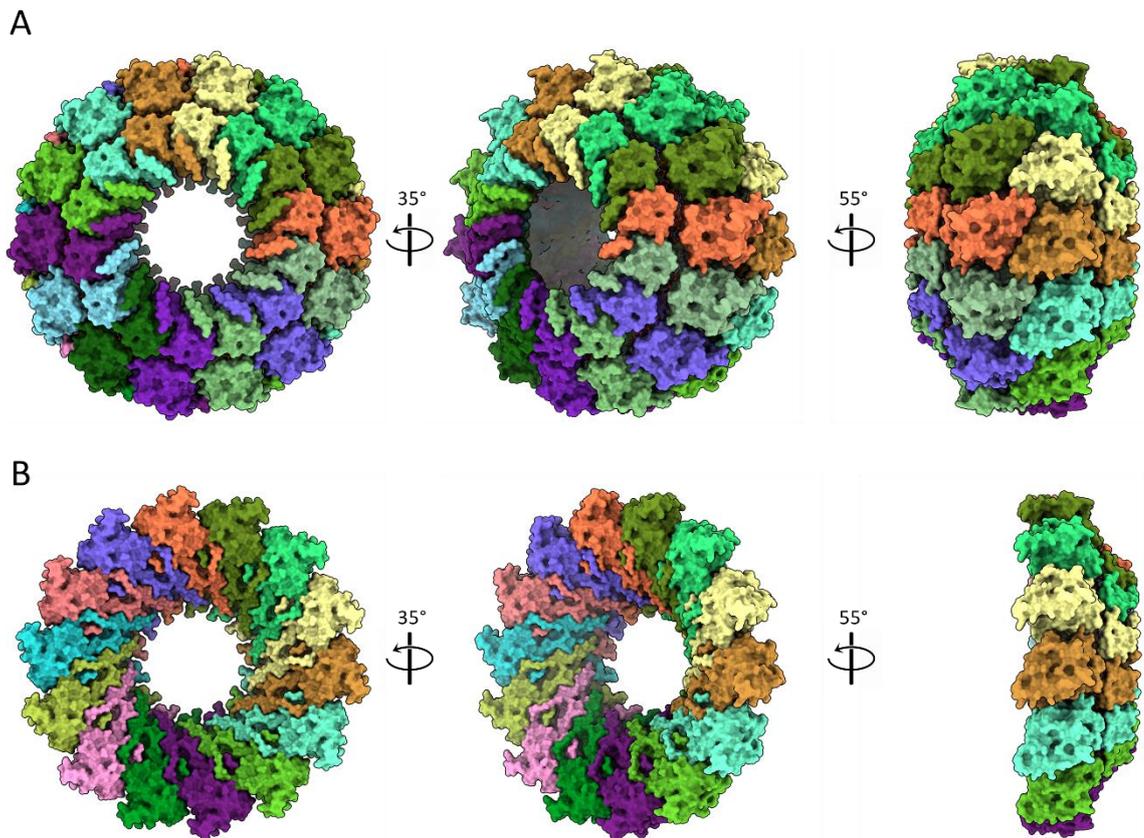


Figure 5.11. Dimer of tetradecamer sauronoid model.

A surface projection of the sauronoid assembly with each N protomer individually coloured and the RNA represented as sticks, coloured by element. **A)** The complete sauronoid reveals two bottom-stacked tetradecameric rings as well as the deep RNA-binding groove occupied by RNA. The iris face (right panel) illustrates the intercalation of the globular NTDs of each ring. **B)** The inner face of a tetradecameric ring reveals extensive overlap between neighbouring N protomers.

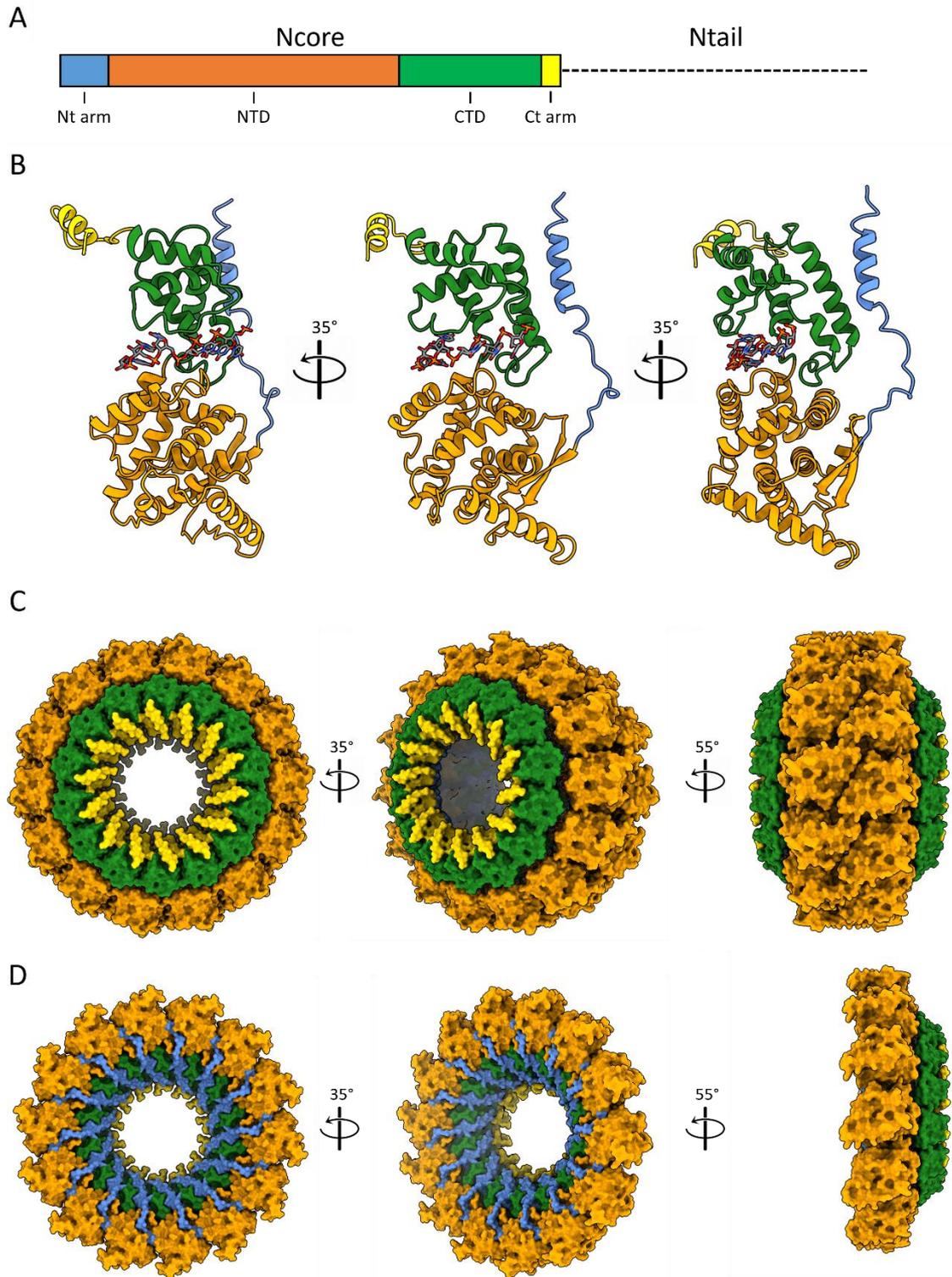


Figure 5.12. The HeV N protein fold and protein domains.

A) Schematic, linear representation of the HeV N protein domain architecture indicating the Ncore and Ntail regions. The Ncore consists of the N-terminal domain (NTD) and C-terminal domain (CTD), each of which are flanked by a flexible arm; the N-terminal (Nt) arm and the C-terminal (Ct) arm. **B)** The HeV N protein monomer model in ribbon presentation with the RNA as sticks. Protein colour coding corresponds to that in **(A)**, the RNA is coloured by element. **C-D)** The surface presentation of the multimeric assembly reveals that the top of each tetradecameric ring consists of the Ct-arm (gold) which lies on top of the globular CTD (green) **(C)**. The globular NTDs intercalate to form a dimer of tetradecamers (right panel). The Nt-arm (blue) lies completely within the sauronoid central cavity **(D)** where its loop region jumps over the NTD (orange) and its alpha-helical coil interfaces with the CTD.

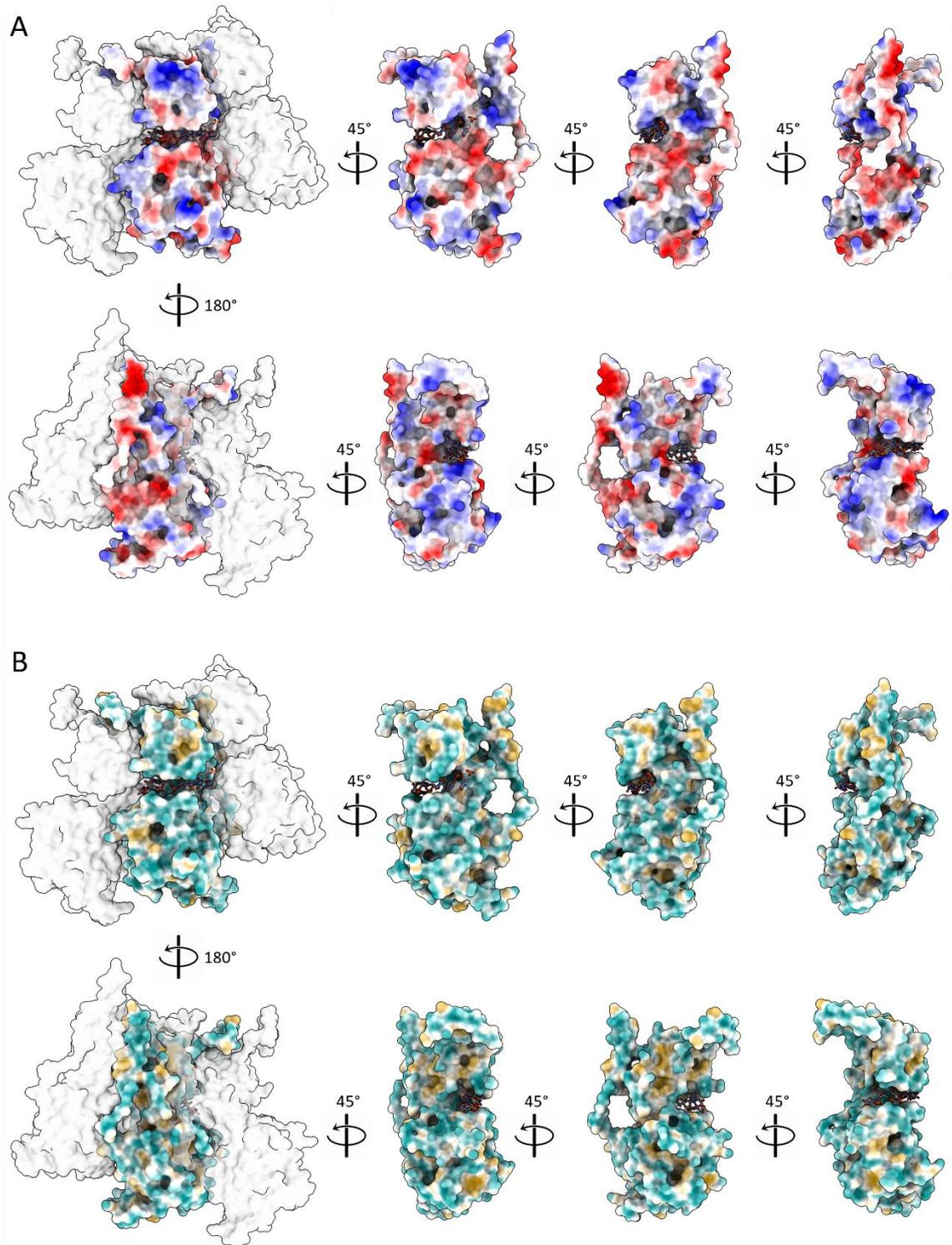


Figure 5.13. HeV N protomer surface values.

The HeV N protomer, with RNA, has been coloured according to electrostatic potential (A) and hydrophobicity (B). Neighbouring N_{i+1} and N_{i-1} protomers are represented as transparent elements in the 0° and 180° views. **A**) Electrostatic surface potentials are coloured for negative (red) and positive (blue) residues or uncoloured (white) for neutral residues. **B**) Colours are mapped to Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle, 1982) from most hydrophilic (green) to most hydrophobic (gold).

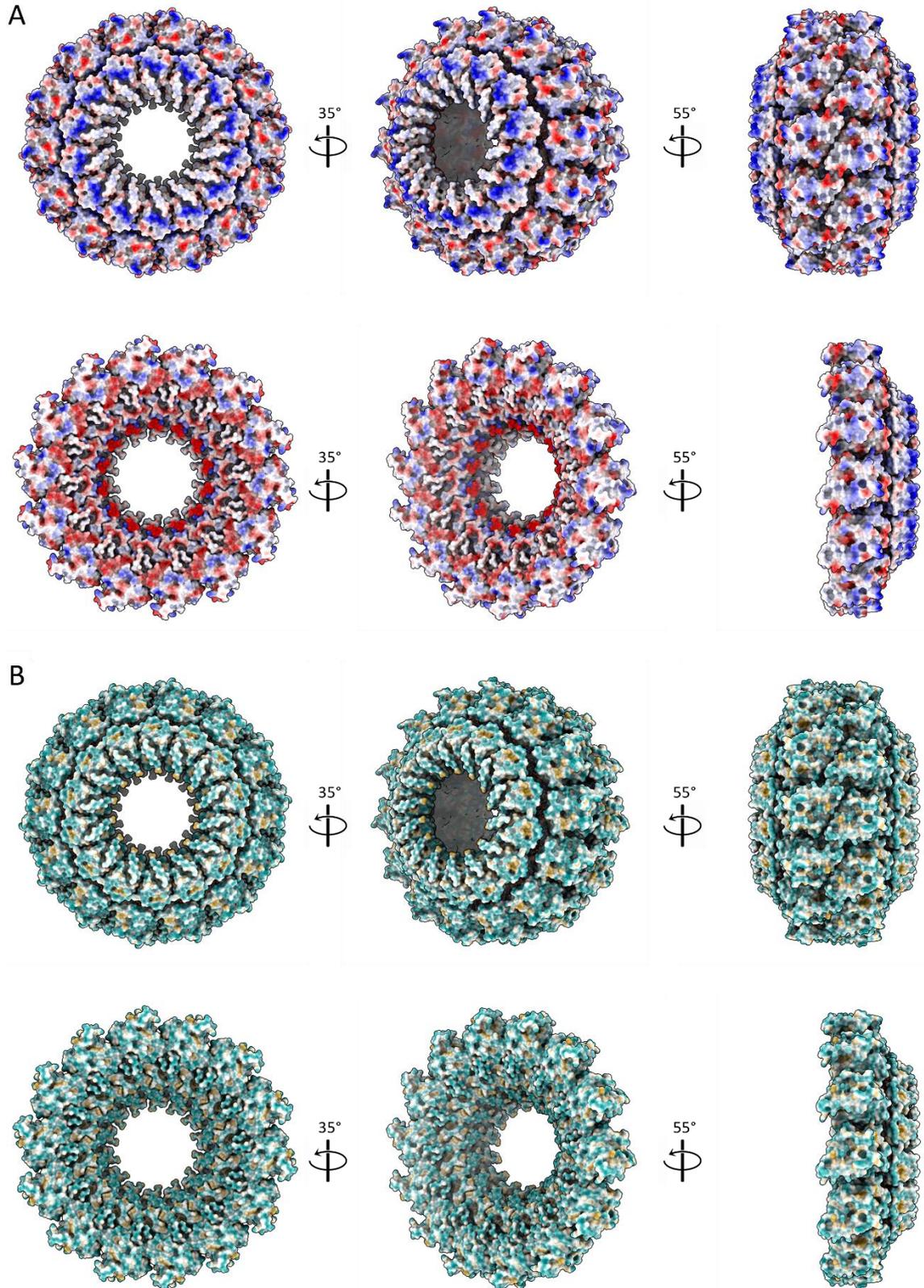


Figure 5.14. HeV N sauronoid surface values.

The HeV N sauronoid, with RNA, has been coloured according to electrostatic potential (**A**) and hydrophobicity (**B**). Top panels depict the dimer of tetradecameric sauronoid. Bottom panels depict a single tetradecameric ring. **A**) Electrostatic surface potentials are coloured for negative (red) and positive (blue) residues or uncoloured (white) for neutral residues. **B**) Colours are mapped to Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle, 1982) from most hydrophilic (green) to most hydrophobic (gold).

5.2.4.2 The HeV N protein RNA binding groove contains conserved N-RNA interactions

The sauronoid map showed clear density for single-stranded RNA in the RNA-binding groove that separates the NTD and CTD, which was modelled as a poly-uridine; one hexa-uridine per N protomer (Figure 5.10, Figure 5.15, Supplementary Figure S5.12A). The RNA molecule is oriented in the conventional ‘three-base in, three-base out’ similar to other paramyxoviruses (Gutsche et al., 2015; Ker et al., 2021). This orientation is facilitated by a 180° backbone turn every three nucleotides. In the case of HeV, the 3'-most three-base stack is oriented in towards the RNA-binding groove while the 5'-most three-base stack are oriented towards the solvent (Figure 5.15).

The charged RNA-binding groove consists of a groove floor and two ‘lips’ that follow the RNA chain’s orientation (Figure 5.15A-B). Residue segments Thr181-Gln200, containing $\alpha 8$, in the NTD and the loop Ser344-Tyr354 in the CTD form the solvent-facing groove lips. The groove floor consists of the NTD-CTD-connecting loop Glu260-Ala265 continuing into $\alpha 12$ until Arg272 as well as the $\alpha 14$ -containing stretch Ser317-Gly326. Within the RNA-binding groove, a series of conserved basic (Lys178, Arg192, Arg193, Arg352) and polar (Thr185, Gln199, Gln200, Tyr258, Gln319, Ser344, Tyr354) residues likely interact with the RNA chain through electrostatic and hydrophobic interactions (Figure 5.15B-C) (Gutsche et al., 2015; Ker et al., 2021).

Adjacent to and contiguous with the RNA binding groove floor is a large cavity surrounding the ‘three-base in’ stack flanked by the N_i NTD and CTD as well as $\alpha 13$ of the N_{i+1} CTD and $\alpha 1$ of the N_{i+1} Nt-arm (Figure 5.16, Supplementary Figure S5.12B). While the modelled RNA bases are all uridines, this cavity – here called the ‘in-base cavity’ – is sufficiently spacious to accommodate the larger purine bases and, correspondingly, N_i residues Tyr258, Glu260, Glu261, Phe268, Arg272, Ser344, and Met345, and N_{i+1} residue Lys321 may be involved in interactions with these bases (Figure 5.16C-D).

5.2.4.3 The HeV sauronoid reveals conserved and novel N-N interactions

The sauronoid assembly is formed through extensive lateral interactions in each tetradecameric ring as well as intercalating interactions between two tetradecameric rings. The lateral interfaces consist of electrostatic and hydrophobic interactions and are mediated primarily through the Nt-arms, NTDs, and CTDs of two neighbouring N protomers, with contributions from the Ct-arm, over a total buried surface area of $\sim 2930 \text{ \AA}^2$, calculated via PDBe PISA (Krissinel and Henrick, 2007) (Figure 5.13, Figure 5.14). At a calculated buried surface area of $\sim 1195 \text{ \AA}^2$, the N_i - N_{i+1} interface mediated by globular domains represents just over a third of

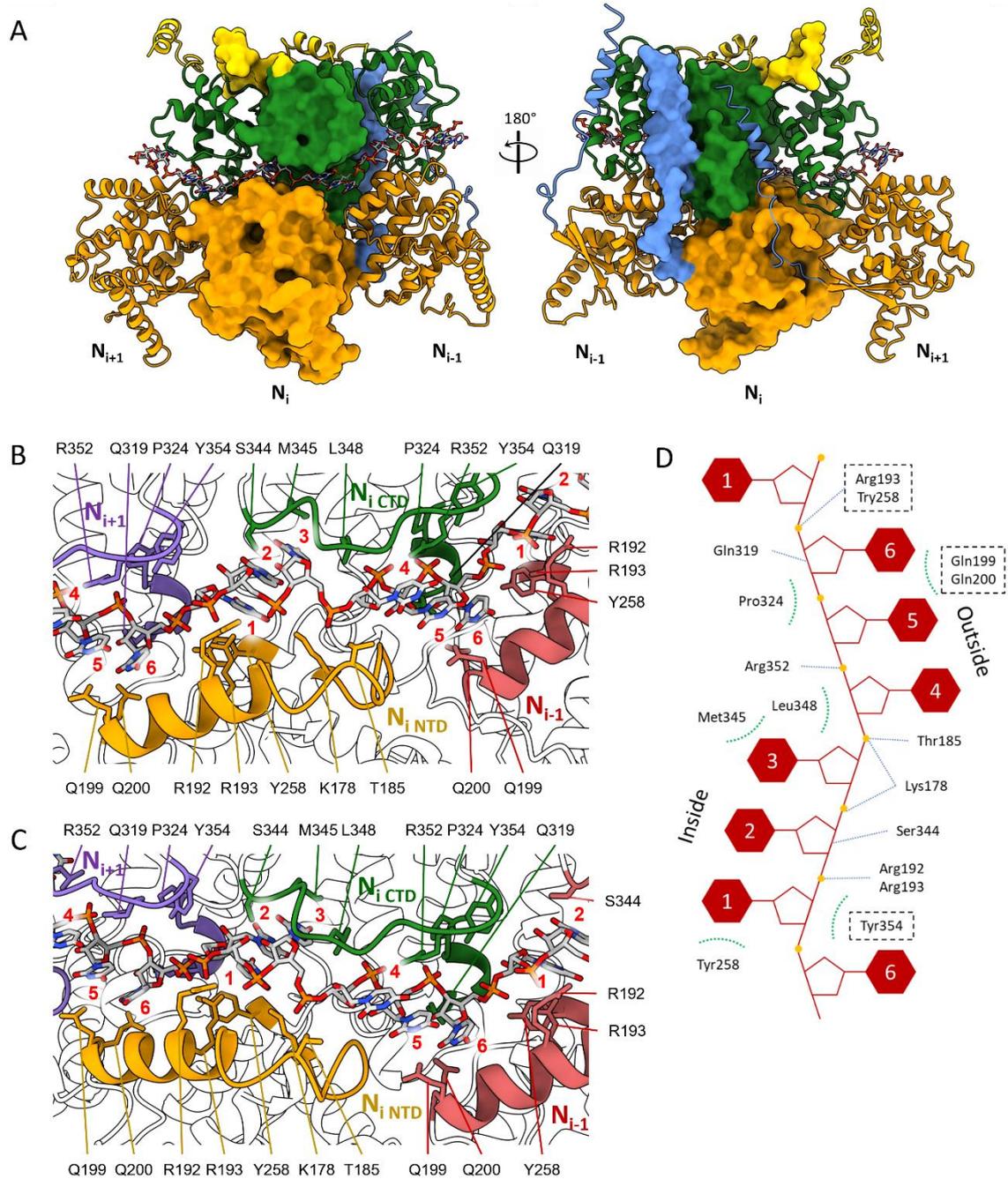


Figure 5.15. The HeV N protein binds RNA through electrostatic and hydrophobic interactions.

A) A trimer of N protomers with the N_i protomer as a surface render and the neighbouring N_{i+1} and N_{i-1} protomers as ribbons. The N protein domains are individually coloured; the Nt-arm in blue, the NTD in orange, the CTD in green, and the Ct-arm in gold. The 18-meric RNA chain is visualised as sticks and follows heteroatom colouring. **B-C**) Two different close ups of the RNA binding groove centred on the N_i protomer. N proteins are visualised as ribbons and the RNA chain as sticks. Helices and loops containing putative RNA interacting residues are coloured for the N_i (NTD; gold, CTD; green), the N_{i+1} (purple), and the N_{i-1} (coral). RNA bases in each hexamer are numbered 1 to 6 on a per-N protomer basis from the 3'- to the 5'-end. **D**) A topological schematic of the RNA-interacting residues centred around the N_i protomer and the corresponding RNA hexamer. RNA base stack orientations are shown and bases in each hexamer are numbered 1 to 6 from the 3'- to the 5'-end. Putative protein-RNA interactions are indicated for electrostatic interactions (blue dashed line) and hydrophobic interactions (green dashed arc). Residues belonging to neighbouring N protomers are bounded by a dashed box.

the total buried surface area between the neighbouring protomers. The globular domain interface consists of extensive electrostatic interactions, with the N_i surface containing mostly positively charged patches and the N_{i+1} surface mostly negatively charged patches (Figure 5.13A).

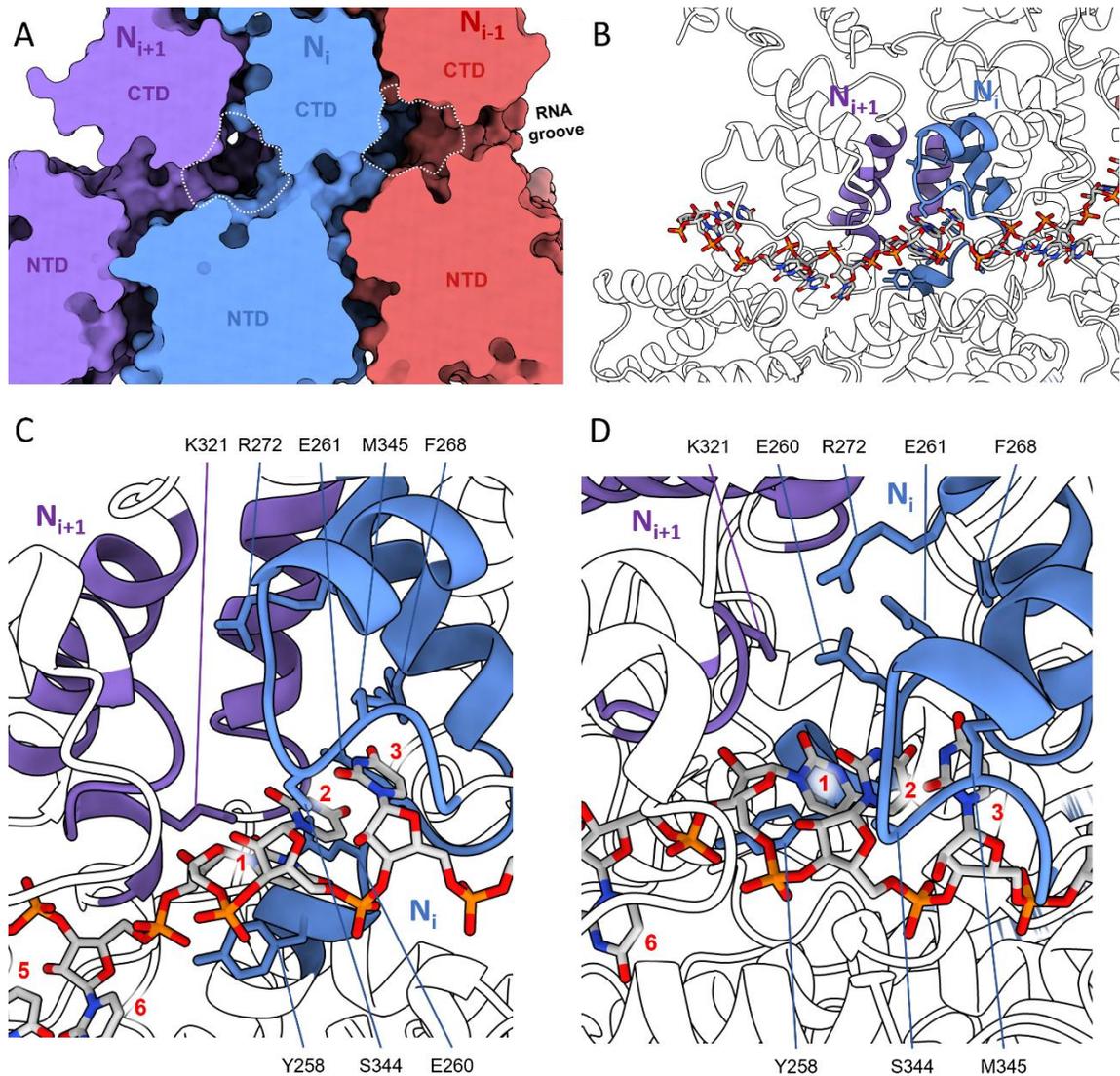


Figure 5.16. Two HeV N protomers form an in-base cavity.

A) A trimer of N protomers as surface renders and individually coloured; N_i (blue), N_{i+1} (purple), and N_{i-1} (coral). Each N protomer's NTD and CTD is indicated. The trimer is sliced down the frontal face to reveal the RNA binding groove and in-base cavities (white dashed area). Each in-base cavity is formed between two neighbouring N protomers. **B-D**) An N protein dimer centred on the in-base cavity with an overview (**B**) and close ups (**C-D**). Helices and loops that form the in-base cavity are coloured for the N_i (blue) and the N_{i+1} (purple). Putative RNA-base interacting residue side chains are visualised as sticks. **C-D**) N protein residues putatively interacting with bases in the in-base cavity are visualised as sticks and individually indicated. RNA bases in each hexamer are numbered 1 to 6 on a per-N protomer basis from the 3'- to the 5'-end.

The Nt- and Ct-arms contribute differentially to the lateral interactions in each decameric ring. While the contribution of the Ct-arm is minor, with a calculated buried surface area of $\sim 340 \text{ \AA}^2$, the Nt-arm contributes extensively to the lateral interactions. At a calculated buried surface area of $\sim 1740 \text{ \AA}^2$, this interface represents just over half of the total lateral interface between N_i and N_{i+1} . Aside from electrostatic patches, a conserved hydrophobic region consisting of a triad of aromatic residues (Phe267, Phe268, and Tyr301) in the CTD of the N_i protomer and Phe11 in $\alpha 1$ of Nt-arm N_{i+1} protomer forms key interface (Gutsche et al., 2015; Ker et al., 2021). A further hydrophobic groove in the N_i NTD (Tyr80, Ser81, Leu87, Val162, Val165, Ile166, Ile234, and Val249) is occupied by the N_{i+1} Nt-arm loop Ala27-Thr33. Within this N_{i+1} Nt-arm loop, residue Leu31 is buried inside a pocket in the N_i NTD consisting of Asn43, Arg48, Tyr80, Glu82, Met158, Leu161, Val162, and Val165 (Figure 5.17).

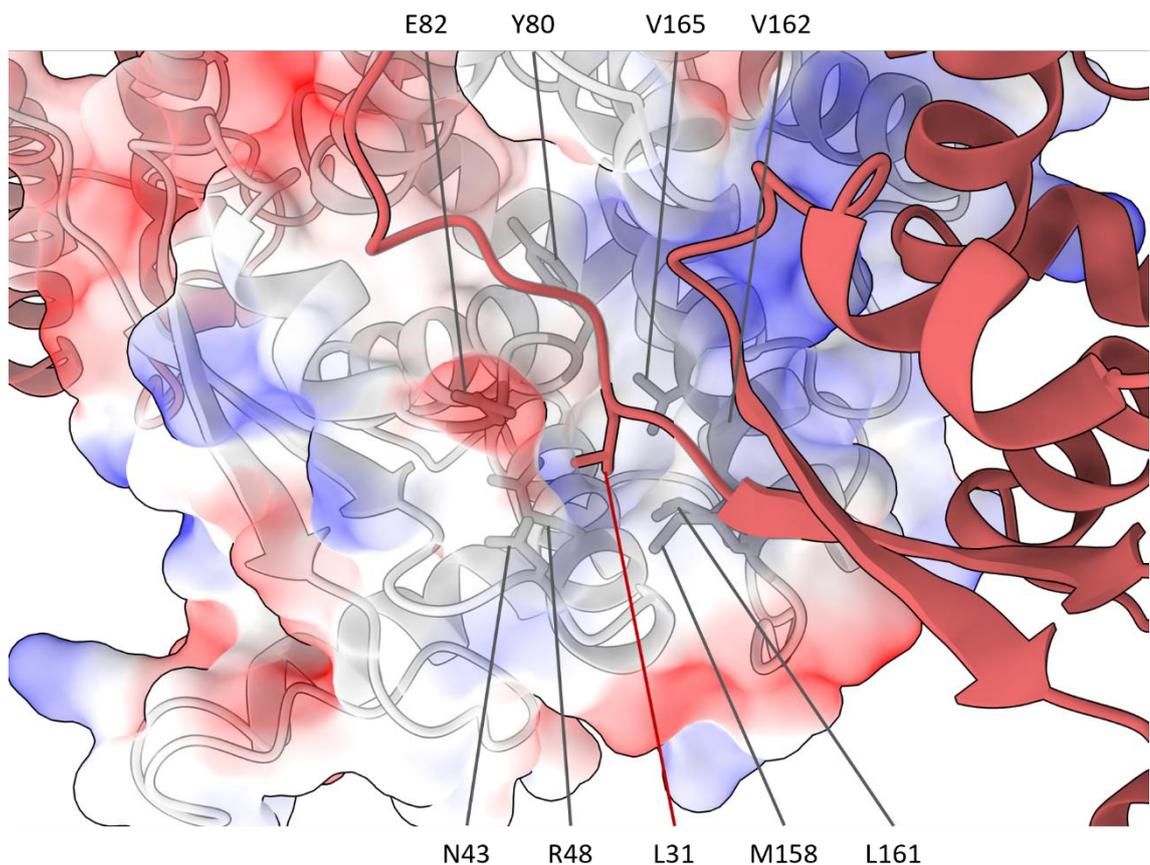


Figure 5.17. HeV N_{i+1} Nt arm-loop residue Leu31 is buried in a pocket on the N_i NTD.

The HeV N_{i+1} residue Leu31 (coral) sits within a pocket consisting of the neighbouring N_i protomer's NTD residues Asn43, Arg48, Tyr80, Glu82, Met158, Leu161, Val162, and Val165 (dark grey sticks). The NTD sporting the binding pocket shows transparent electrostatic surface potentials for negative (red) and positive (blue) residues or uncoloured (white) for neutral residues.

Beyond the extensive N_{i+1} Nt-arm interface with the N_i NTD and CTD, a minor contact exists between the utmost amino-terminal residues of the N_{i+1} Nt-arm (Met1, Ser2, Ile4, and Phe5) and the N_{i-1} Ct-arm loop region Gln376-Lys381 (Figure 5.18, Supplementary Figure S5.13A). This N_{i+1} - N_{i-1} interface represents a novel interaction that, to the author's knowledge, has not been described in the literature for paramyxoviral nucleocapsids and will be referred to as the 'elbow interface'. Since the elbow interface only has a calculated buried surface area of $\sim 80 \text{ \AA}^2$, it is unlikely to contribute strongly to N-N interactions.

A unique feature of the sauronoid assembly is the gear-wheel interface between the two tetradecameric rings. This interface, here called the 'intercalation zone', consists solely of residues in the NTDs of N protomers in the top ring and N protomers in the bottom ring (indicated by the lower-case n) (Figure 5.19, Supplementary Figure S5.13B). The N_i protomer contacts two protomers in the bottom ring (n_i and n_{i+1}), with a total calculated buried surface area of $\sim 545 \text{ \AA}^2$. A clear, reciprocal electrostatic interface exists between N_i and n_{i+1} with the residues Lys120, Glu123, and Glu124 where the two Glu residues in N_i Glu form a tab that binds the Lys residue in n_{i+1} and *vice versa* (Figure 5.19B). The poor electron density surrounding loop Met114-Asp119 forced the stubbing of side chains and, accordingly, these were modelled as alanines (Figure 5.19B-C). An electrostatic interaction may exist between the N_i Asp157 and the stubbed n_i Arg116, as they are in proximity and there is ample space for the Arg side chain. Similarly, N_i residues Ser44 and Glu46 form a patch that may interact with the stubbed n_i Met114 and Glu115 residues which are in close proximity. The absence of clear electron density neither supports nor refutes these putative interactions and improved local resolution may further elucidate the intercalation zone and increase the calculated buried surface area.

5.2.5 Helical reconstruction of HeV nucleocapsid-like filaments generates low-resolution maps

The dataset of 3548 micrographs which enabled elucidation of the N protein monomer through sauronoid assemblies also contain a large quantity of NLFs. The pre-processed micrographs (see Section 5.2.1) were similarly used for manual particle picking, however now a helical picking strategy was employed. Start and end coordinates of continuous NLFs were selected within RELION which automatically joins the two coordinates with a straight line (Figure 5.20). The generation of a template for automated picking was attempted but failed to provide a useful picking reference. Instead, 88,363 coordinates were manually selected from all 3548 micrographs in dataset.

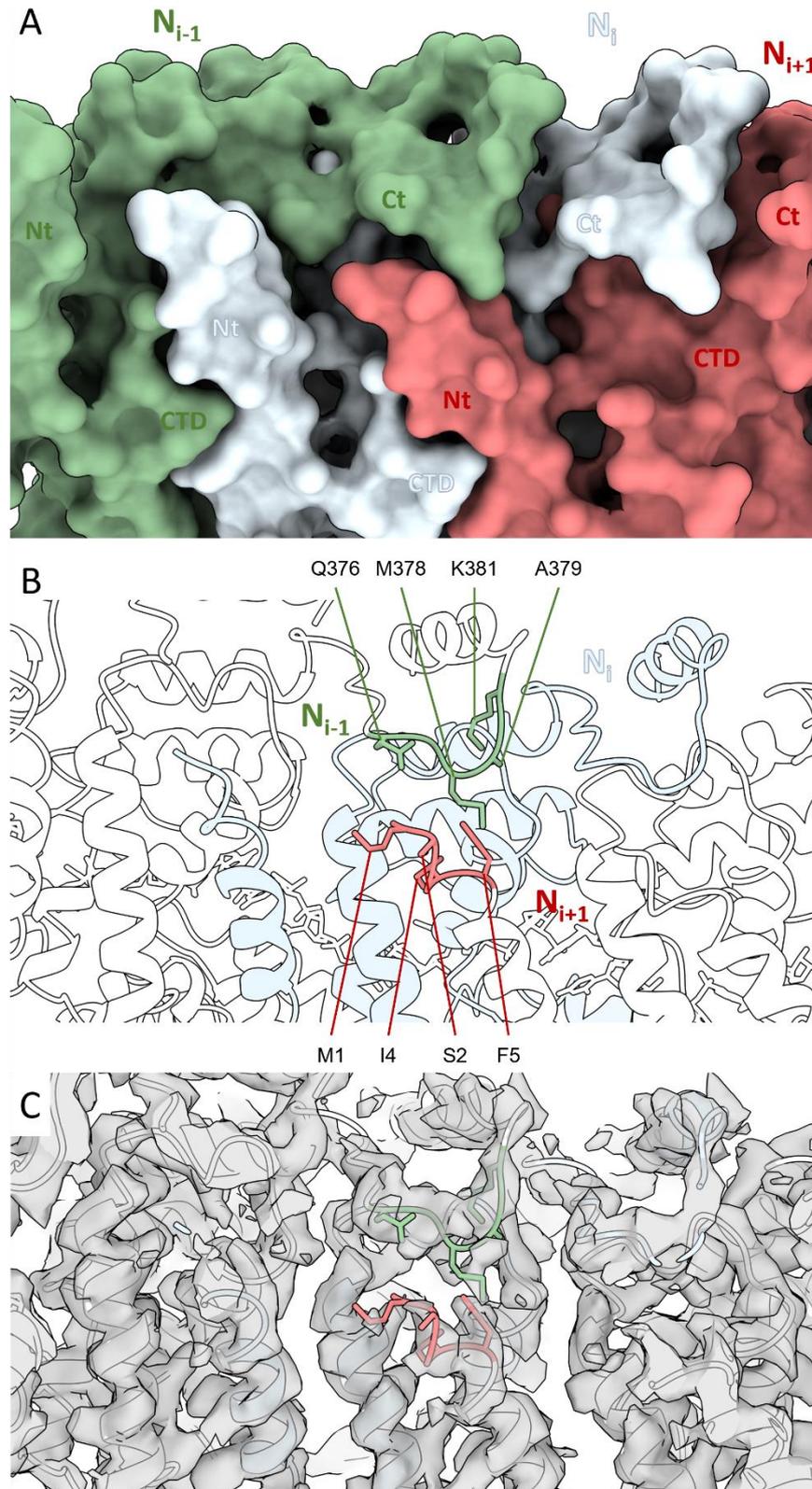


Figure 5.18. The HeV sauronoid elbow interactions bridge non-neighbouring protomers.

A trimer of N protomers showing the N_i (pale blue), N_{i+1} (coral), and N_{i-1} (green) protomers, centred on the elbow interaction between the N_{i+1} Nt-arm and the N_{i-1} Ct-arm. **A)** A surface render of the N protein trimer with the Nt-arm (Nt), CTD, and Ct-arm (Ct) indicated. **B-C)** A ribbon visualisation of the N protein trimer. Protomers N_{i+1} and N_{i-1} are white. Putative interacting residues are coloured coral and green and presented as sticks. **B)** Residues putatively involved in the elbow interactions are shown as sticks and indicated. **C)** The cryo-EM sauronoid density map is shown as a transparent overlay.

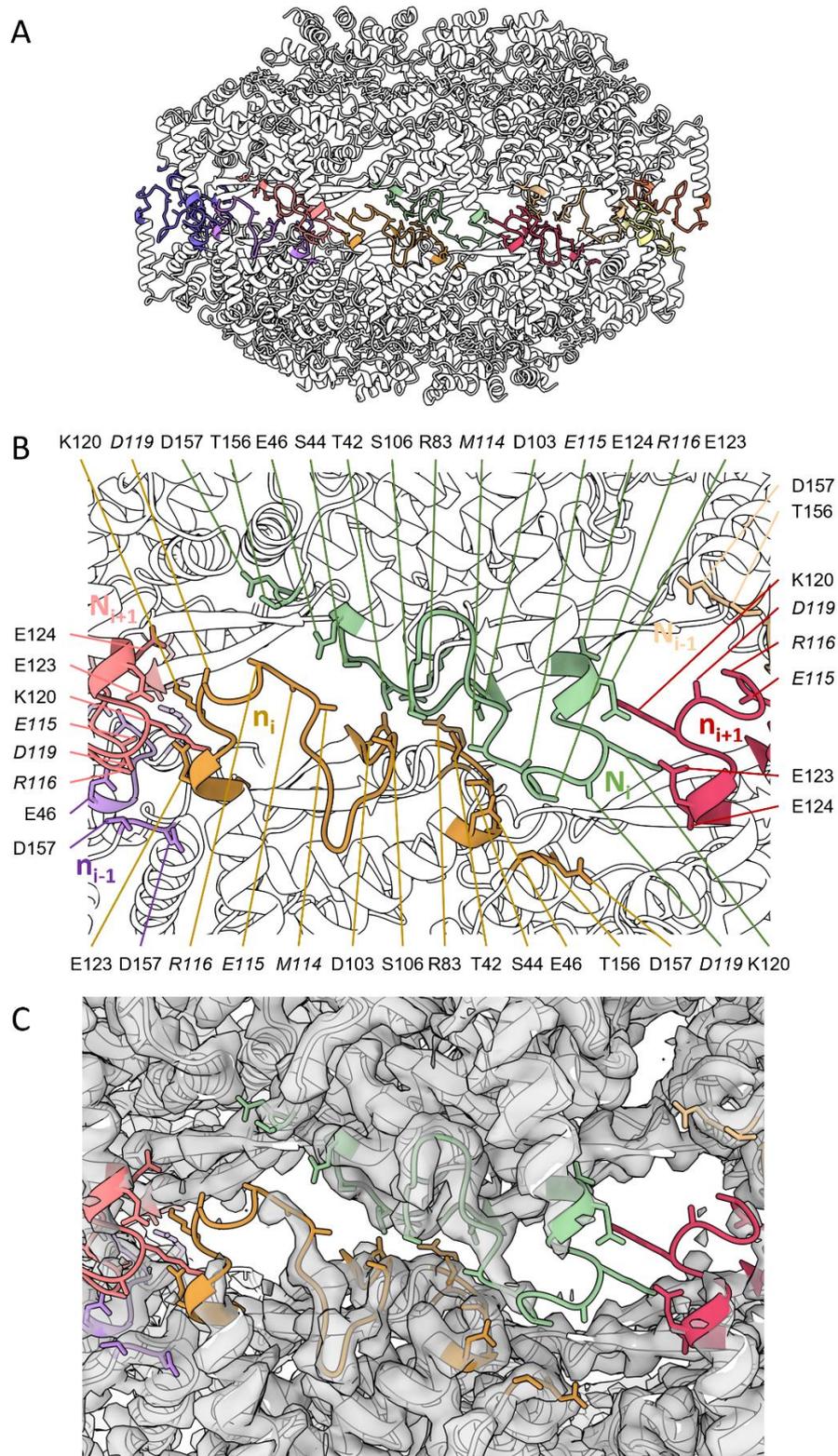


Figure 5.19. The HeV sauronoid intercalation zone is poorly resolved.

A ribbon visualisation of a sauronoid. N protomers are coloured white with loops belonging to the intercalation zone coloured per protomer. Putative interacting residues are presented as sticks. **A**) An overview of the dimer of pentamers assembly showing the intercalation zone. **B-C**) The intercalation zone centred around the N_i protomer (lime) of the top ring, showing the N_{i+1} (pink) protomer and the bottom ring protomers N_i (n_i ; gold), N_{i-1} (n_{i-1} ; purple), and N_{i+1} (n_{i+1} ; coral). **B**) Residues putatively involved in the intercalation zone are shown as sticks and indicated. Italicised residues are stubbed and modelled as Ala. **C**) The cryo-EM sauronoid density map is shown as a transparent overlay revealing areas of poor density.

Particle extraction for helical coordinates takes a more nuanced approach compared to classical SPA. The particles, here referred to as segments, are extracted along the line connecting the start and end coordinates following pre-defined parameters. This requires *a priori* knowledge of the helical parameters describing the HeV NLFs including width and pitch, which were estimated from the raw micrographs in ImageJ to be a width of ~ 220 Å and a pitch of ~ 4.66 Å. Doing so resulted in the extraction of 5,424,996 segments with a box size of 256 pixels (272.6 Å). Increasing the box size results in the extraction of fewer but longer segments. The alignment of particles in large box sizes is computationally demanding, often taking up vast computing resources and time. Down-sampling, or binning, of the data drastically reduces the computational demands and speeds up 2D classification at the cost of high-resolution information. Accordingly, the extracted helical segments were initially binned by a factor of 4, for an effective box size of 64 pixels, to reduce computational demands and speed up initial 2D classification at the cost of high-resolution information. This also allowed junk particles to be filtered out more rapidly since binned classes are aligned on lower-resolution features. Still, there were markedly fewer junk particles due to the manual picking approach taken during coordinate selection (Figure 5.21A). Multiple rounds of 2D classification were carried out, each time visually inspecting the resultant classes on linearity and resolved detail and selecting the optimal classes to move forward. NLF segments were then re-extracted at nominal resolution (unbinning) once a satisfactory subset of classes containing 82,568 segments was generated to enable increased accuracy in the assignment of 2D classes (Figure 5.21B).

Following 2D classification at nominal resolution, 66,587 segments were taken forward for 3D classification using helical reconstruction. Helical reconstruction requires further *a priori* estimates of the helical parameters and allows for optimisation of these parameters during classification. Assuming a tetradecameric helical turn, based on the tetradecameric semi-sauronoid ring, the helical twist was estimated to be around 25.7° with an optimisation range of 22° to 30° . The helical rise was re-estimated based on the 2D classes and set to 4 Å with an optimisation range of 2.5 Å to 5 Å. The ensuing 3D classification produced three classes with varying segment assignments and optimised helical parameters, revealing a putative left-handed spiral (based on related paramyxoviral nucleocapsids) with around 12 to 13 N protomers per helical turn (Figure 5.22). The most populated class was selected for 3D refinement following visual inspection and accordance of helical parameters to those estimated from the 2D classes, with an optimised rise of 4.29 Å and twist of 27.35° corresponding to approximately 13.2 N protomers per helical turn.

The refinement and post-processing, however, failed to generate a high-quality high-resolution reconstruction (Figure 5.23, Supplementary Figure S5.14). The 3D refinement estimated a resolution of 7.4 Å and the ensuing post-processing improved the estimated resolution to 5.8 Å (Figure 5.22). However, visual inspection of the corresponding maps casts doubt on the resolution estimates, in particular that of the post-processed map. This is further supported by an aberrant FSC curve (Supplementary Figure S5.14).

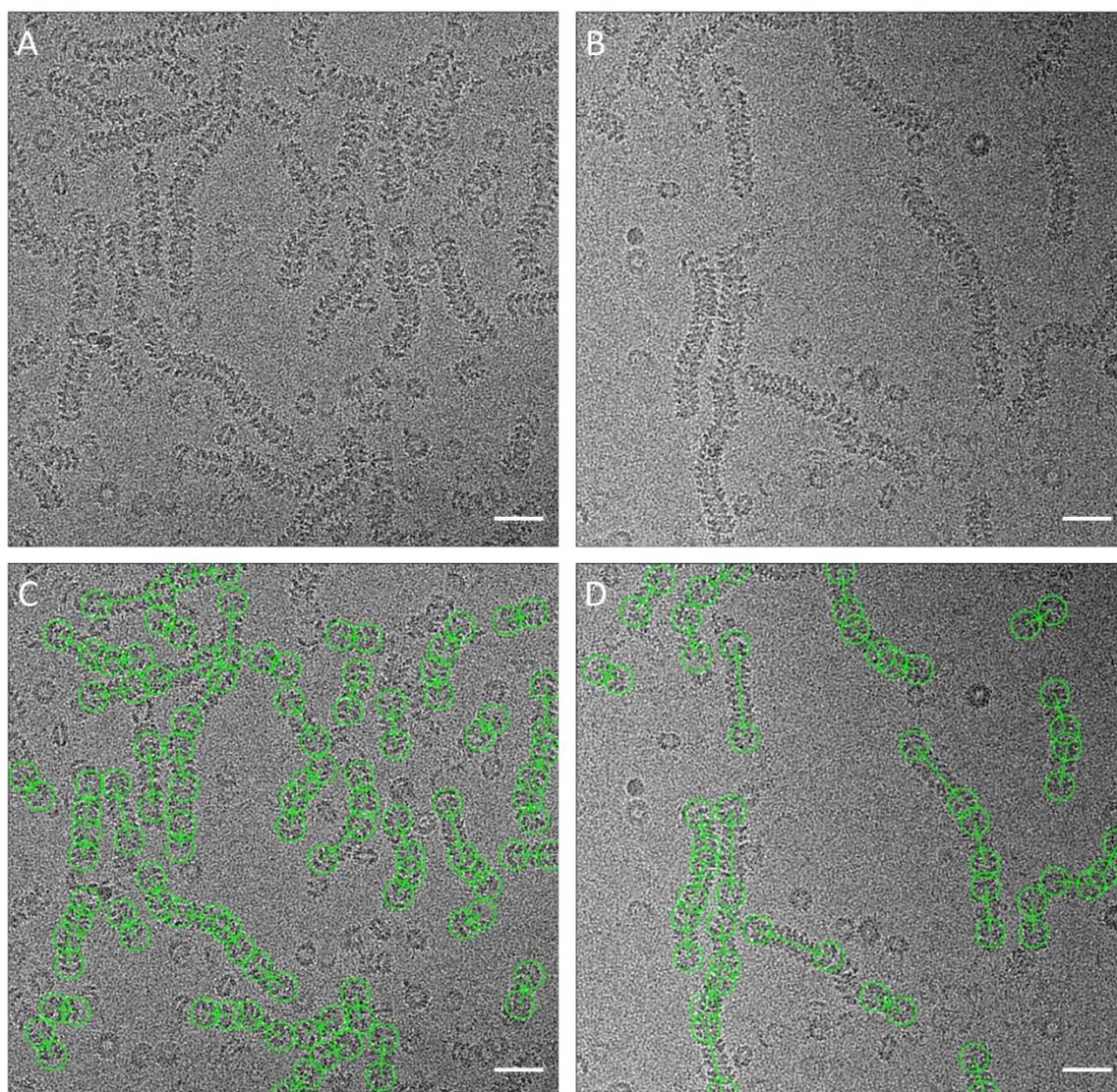


Figure 5.20. HeV helical segment selection through manual picking.

Representative micrographs, identical to those in Figure 5.3, illustrating the helical segment selection steps. Scale bars are 50 nm. **A-B)** Dataset micrographs depict the same morphologies observed in negative stain micrographs and cryo-EM micrographs from the screening session (see Figures 4.3 and 4.6). **C-D)** Manual picking allows for the selection of filamentous tubes. The start and end coordinated of a continuous filament are selected (green circles paired through green line). Helical segments were extracted along the green line following pre-defined helical parameters.

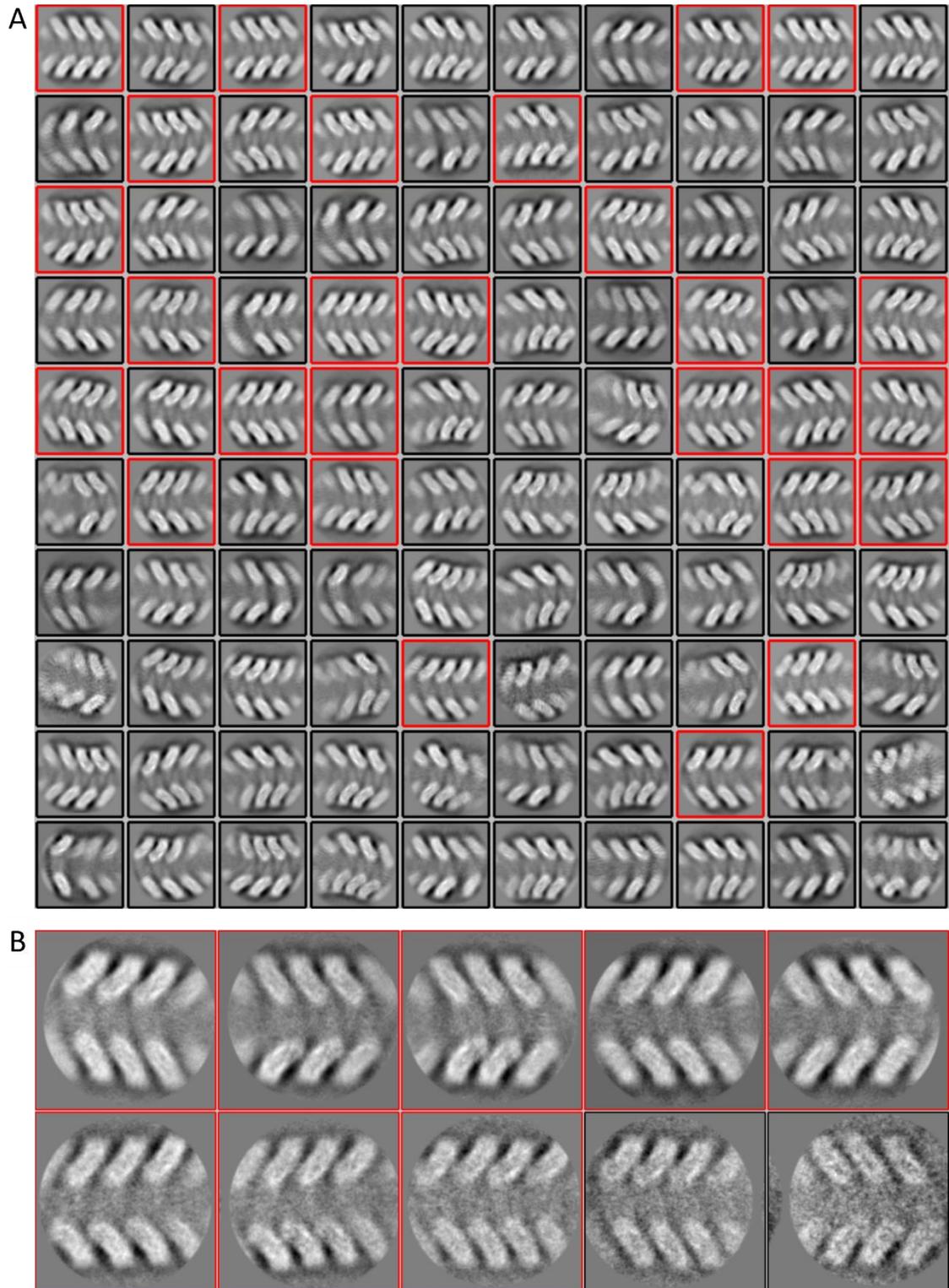


Figure 5.21. Helical HeV NLFs generate highly flexible segments in 2D classification.

Extracted and 2D classified HeV NLF segments. The displayed classes are sorted on ‘in class distribution’ where the most populated classes appear top left and class population decreases from left to right and top to bottom. Box boundaries are 272.6 Å (256 pixels). **A)** A set of 5,424,996 HeV NLF segments was extracted with a binning factor of 4 from manually selected coordinates following the helical SPA procedure from 3548 micrographs and 2D averaged into 100 classes. Classes bounded by red boxes were selected for further processing. **B)** A set of 82,568 unboxed HeV NLF segments was extracted from previous 2D classifications and 2D averaged into 10 classes. Classes bounded by red boxes were selected for further processing.

Further 2D classifications were attempted to enrich the segment stack with high-quality segments, as well as narrowing the initial helical parameters and optimisation ranges in 3D classification. Additionally, segments were re-extracted in a larger box size (360 pixels; 383.4 Å) to aid in 2D alignment and 2D classification was performed with a higher regularisation parameter of T30 (compared to the standard of T2). None of these attempts improved the visual quality of 3D classes and 3D refinements nor did they improve the resolution estimates.

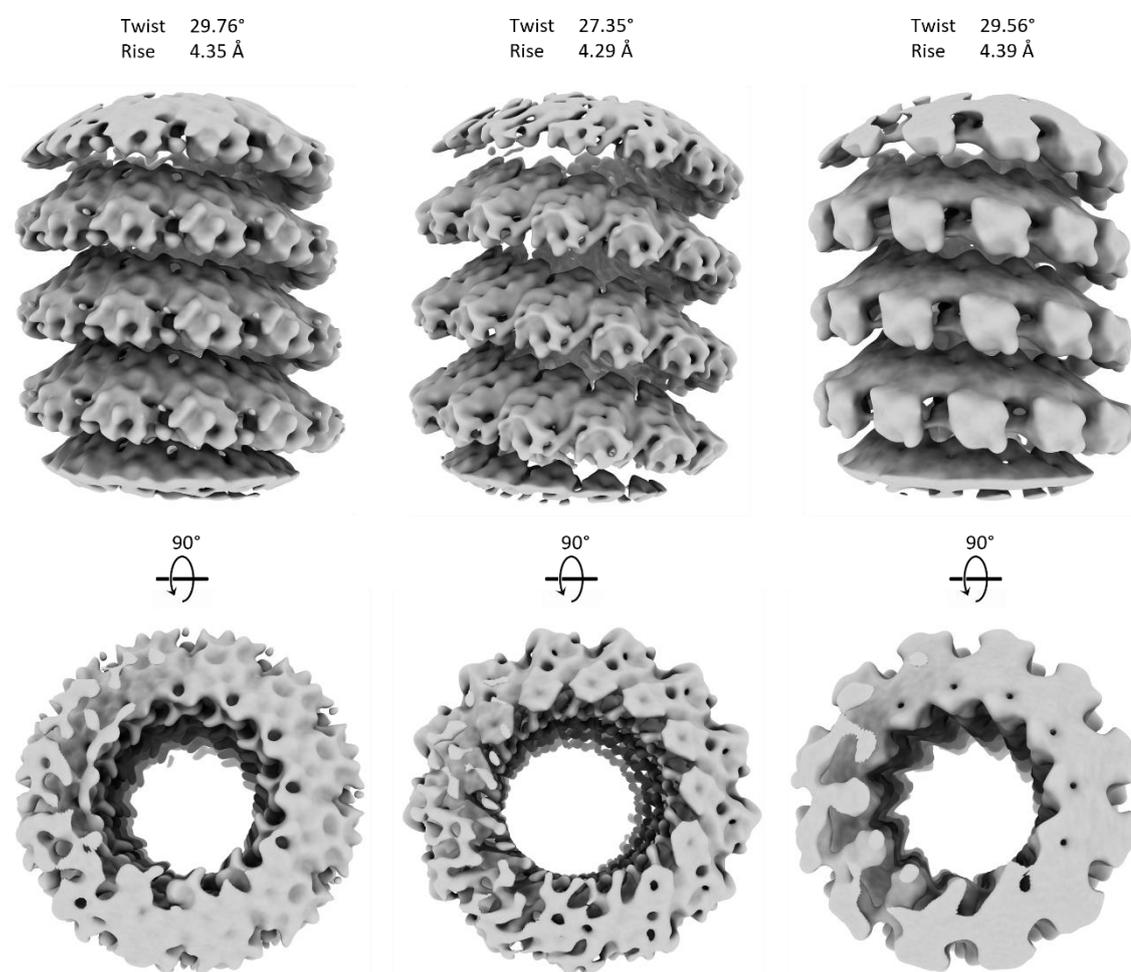


Figure 5.22. Helical HeV NLFs generate variable 3D classes.

The 3D classification of helical segments resulted in three variable classes with different helical parameters and segment assignments. From left to right, classes 1, 2, and 3 contain respectively 18,778 particles, 45,022 particles, and 2807 particles. The central class is the most populated and its helical parameters best match those estimated from 2D classes, resulting in an estimated 13.2 monomers per helical turn.

5.2.8 Recombinantly expressed HeV nucleocapsid-like filaments generate clamshell architectures

The segmentation of HeV NLFs during the helical SPA approach revealed an additional morphology in the sample. Selection and classification generated 2D classes distinct from the regular helical NLFs, which revealed the presence of a clam-shell transition between two NLFs with opposite polarities (Figure 5.25). While 2D classification was promising, no useful 3D classes could be generated for structural elucidation.

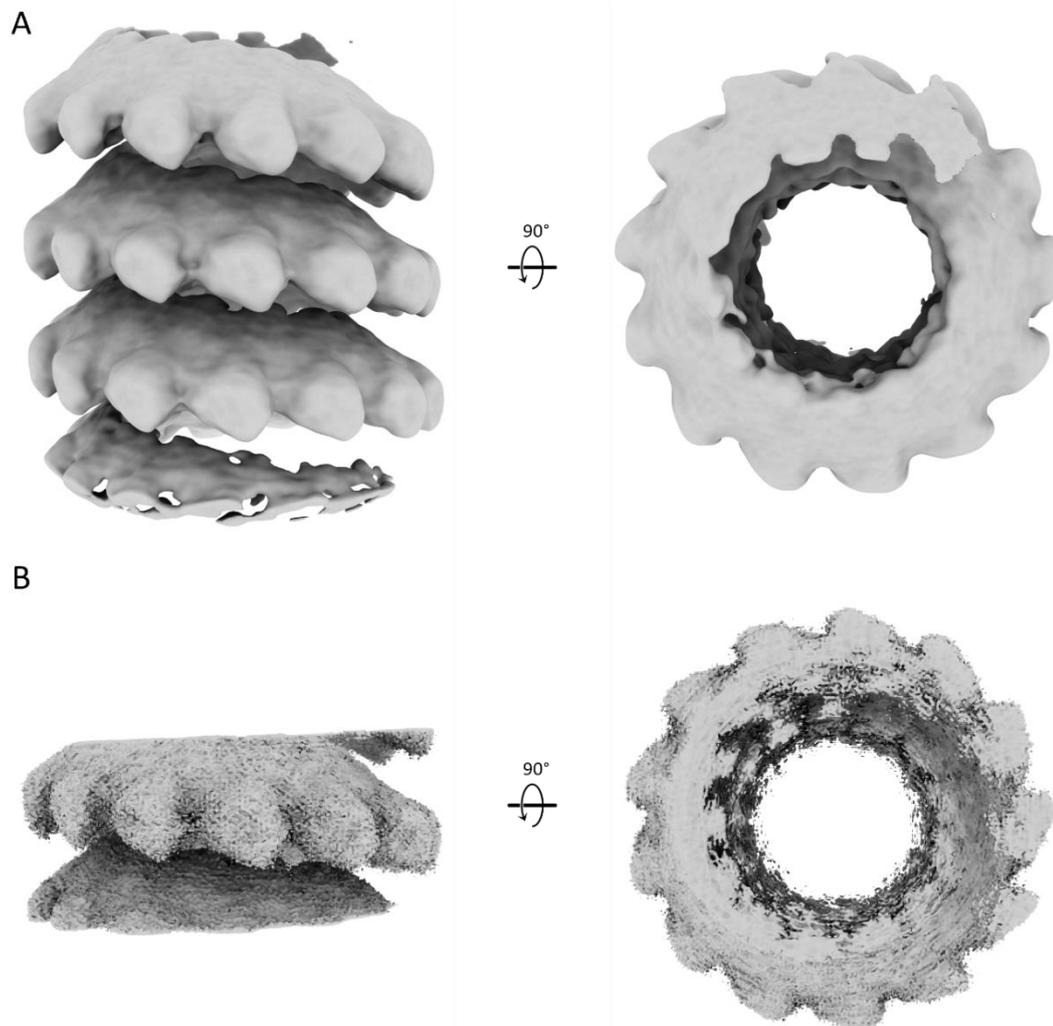


Figure 5.23. Refinement of 3D helical HeV NLFs results in low-quality, low-resolution 3D maps.

Refinement (A) and post-processing (B) of the optimal 3D class resulted in low-quality, low-resolution maps. A) The 3D refinement estimated a resolution of 7.4 Å. B) A mask was applied to the central 30% along the z-axis for post-processing, with an estimated a resolution to 5.8 Å.

5.3 Discussion

5.3.1 The HeV nucleoprotein structure has a canonical fold exhibiting novel features

Following cryo-EM data collection, a HeV sauronoid map was resolved to 3.5 Å resolution which enabled the building and refinement of an atomic N protein model. The sauronoid N protein model revealed both conserved and novel features.

The overall HeV N protein fold resembles other paramyxoviral N proteins. The N protein exists of an Ncore region and an intrinsically disordered Ntail, and – as is common for paramyxoviral N proteins – only the Ncore region was well resolved with no clear cryo-EM density representing the Ntail. The HeV N protein fold resembles in particular the closely related NiV N protein with which it shares around 91% DNA sequence identity (Yabukarski et al., 2014; Ker et al., 2021). The crystal structure for NiV N was solved as a truncated protein – missing both the Nt- and Ct-arms – and in complex with the Nt region of the NiV P protein (Yabukarski et al., 2014). More recently, the full NiV Ncore structure was solved through cryo-EM from helical

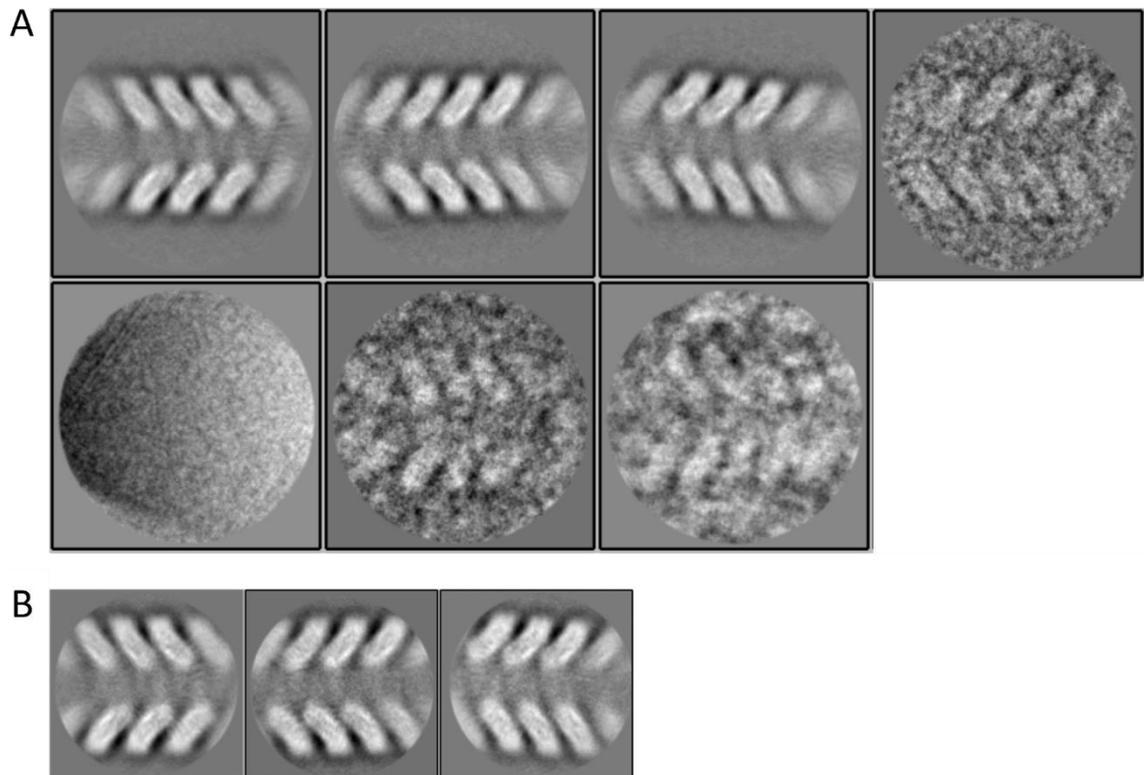


Figure 5.24. Altering extraction and 2D classification parameters fails to improve 2D classes. The 2D classes of NLF segments are sorted on ‘in class distribution’ where the most populated classes appear top left and class population decreases from left to right and top to bottom. Classes containing no particles are not shown. **A)** The curated segment stack was re-extracted using a larger box size of 360 pixels (383.4 Å). **B)** The segment stack in a box size of 256 pixels (272.6 Å) was 2D classified using a regularisation parameter of T30.

particles (Ker et al., 2021). This structure was published whilst the work presented here was in progress, more specifically at the model refinement stage. Since the helical NiV N protein structure was unavailable during the automated model building stage, a homology model was constructed through SWISS-Model based on the NiV N protein crystal structure.

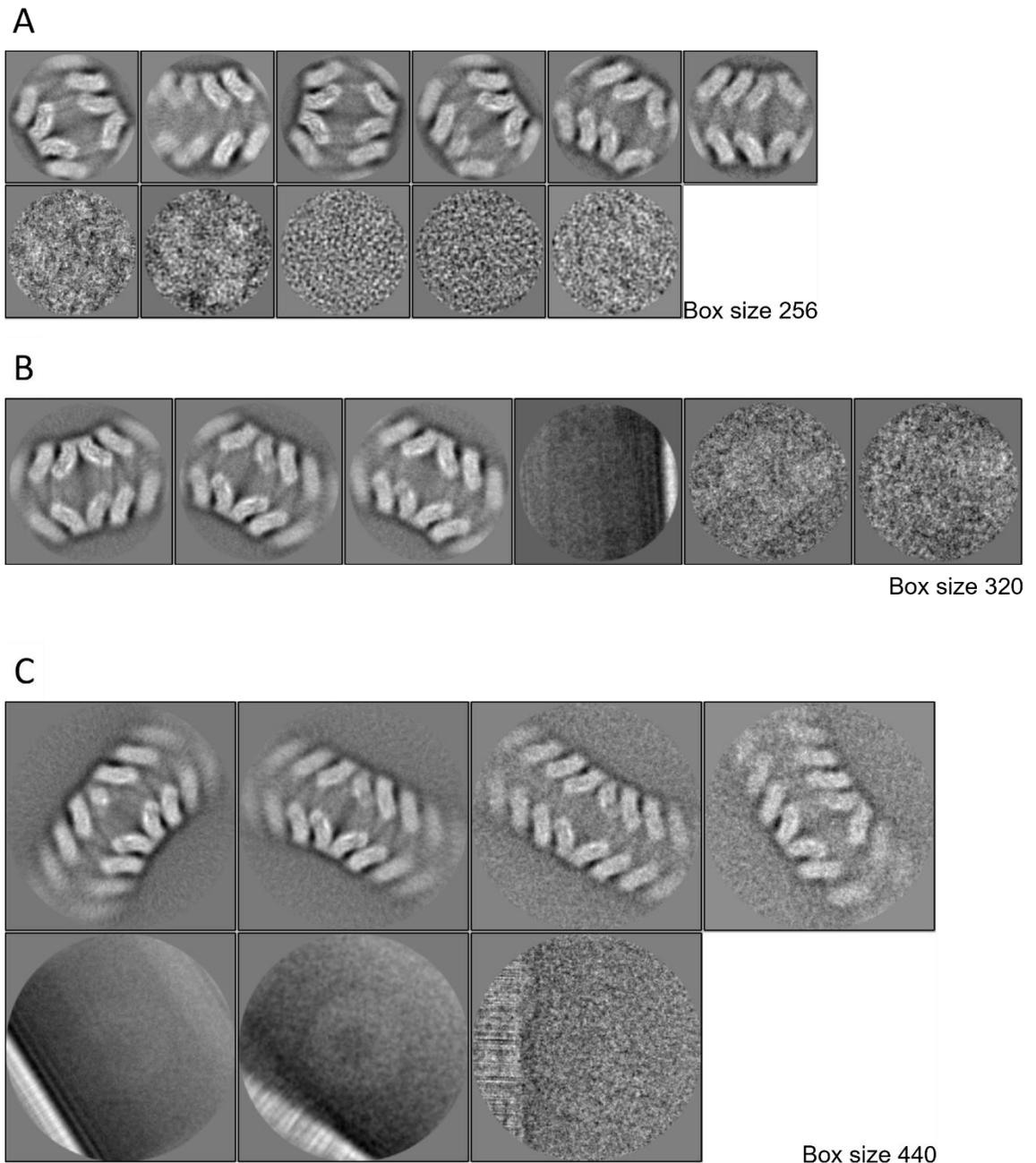


Figure 5.25. A subset of HeV NLFs contain clam-shell transitions.

Extracted and 2D classified HeV NLF segments containing clam-shell transitions. The displayed classes were sorted on 'in class distribution' where the most populated classes appear top left and class population decreases from left to right and top to bottom. Classes containing no particles are not shown. The same segment stack was extracted into increasing box sizes of (A) 256 pixels (272.6 Å), (B) 320 pixels (340.8 Å), and (C) 440 pixels (468.6 Å).

Structural alignment of the final HeV N protein model with the NiV N protein crystal structure reveals some key differences (Figure 5.26A, Supplementary Figure S5.15A). With an overall RMSD of 0.783 Å between the HeV and NiV structures, the NiV NTDs and CTDs are rotated approximately 25° with respect to each other compared to the HeV N structure. This movement is likely induced by binding to the P protein Nt region which keeps the N monomers in an ‘open’ RNA-free conformation. Indeed, similar rotations have been observed for other N proteins such as NiV (Ker et al., 2021) and MeV (Gutsche et al., 2015). The RNA-bound HeV N monomer structure more closely resembled the RNA-bound NiV structure, with an overall RMSD of 0.756 Å (Figure 5.26B, Supplementary Figure S5.15B). From this alignment, three regions of

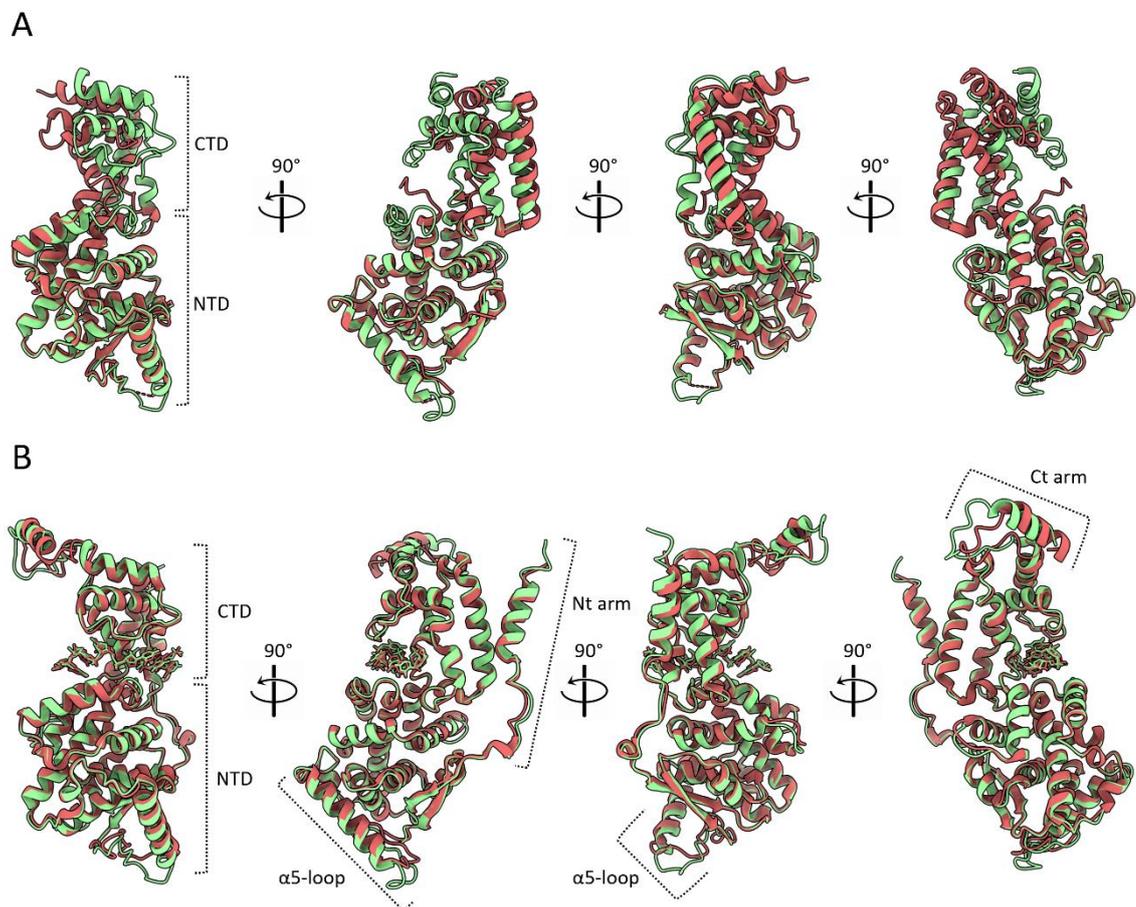


Figure 5.26. Comparison of HeV and NiV N protein structures.

Structural alignment of the sauronoid HeV N protein model and two NiV N protein models. Models are presented in ribbon visualisation, with the HeV structure in green and the NiV structures in red. The globular NTD and CTD are indicated. **A)** Alignment of the sauronoid HeV N protein with the crystal structure of NiV N protein (PDB: 4CO6) with an RMSD of 0.783 Å. The RNA hexamer was removed and the HeV N structure was truncated by removing the Nt- and Ct-arms to resemble the truncated apo-NiV N protein. The NiV P protein domain present in the crystal structure was also removed. **B)** Alignment of the sauronoid HeV N protein with the helical cryo-EM structure of NiV N protein (PDB: 7NT5) with an RMSD of 0.756 Å. The Nt- and Ct-arms as well as the α 5-loop region are indicated.

interest appear. The first two regions constitute the Nt- and Ct-arms. The HeV sauronoid Nt- and Ct-arms are shifted compared to the helical NiV Nt- and Ct-arms. Helical spiralling of a single sauronoid tetradecameric ring would generate a NLF helix and inevitably incurs some minor conformational changes. The flexible nature of the Nt- and Ct-arm put them forward as suitable candidates to provide those changes. Additionally, more Ct-arm residues are resolved in the helical NiV N protein structure than in the sauronoid HeV N protein structure, while the HeV N protein structure shows the very N-terminal Met residue where the NiV N protein structure cuts off at Ile4 (Ker et al., 2021). Indeed, the extreme N-terminal residues are often absent in N protein structures from related viruses such as *Mammalian orthorubulavirus 5* (formerly human parainfluenza virus 5; PIV5) (Alayyoubi et al., 2015), MeV (Desfosses et al., 2019), NiV (Ker et al., 2021), and SeV (Zhang et al., 2021) or face towards the central cavity as in NDV (Song et al., 2019). This absence or shift in orientation may be due to the presence of remaining N-terminal affinity tags that limit cryo-EM resolution. Indeed, an N-terminal affinity purification tag remained attached to resolved N protein structures for PIV5, NDV, NiV, and SeV (Alayyoubi et al., 2015; Song et al., 2019; Ker et al., 2021; Zhang et al., 2021) while the MeV purification used a C-terminal tag (Desfosses et al., 2019). The resolution of the HeV N protein N-terminus enabled the characterisation of a novel henipaviral N-N interface, here called the elbow interface, between the N_{i+1} and N_{i-1} protomers. This interface likely exists in NiV N protein assemblies as well, and the data suggest a similar interface may occur in more distantly related viruses such as PIV5 and MeV (Alayyoubi et al., 2015; Desfosses et al., 2019).

Finally, the solvent facing NTD alpha-helix 5 (HeV $\alpha 5$) and its succeeding loop are also shifted compared to the homologous helix-loop in the NiV structure (Figure 5.26). In fact, this helix-loop region shows significant structural and sequence variability among paramyxoviruses. Both its position and length are highly variable among the N proteins of closely related paramyxoviruses, including MeV, NDV, NiV, PIV5, and SeV (Yabukarski et al., 2014; Alayyoubi et al., 2015; Desfosses et al., 2019; Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). For example, the HeV N protein $\alpha 5$ -loop region contains two amino acid substitutions compared with the NiV N protein; L108V and E137D, both of which are solvent exposed (Ker et al., 2021). Whether there is a functional relevance to these substitutions is unknown, however the $\alpha 5$ -loop region is, in the context of a helical NLF, the main solvent exposed section of the N protomer and is likely important for interactions with both the RdRp complex during transcription and replication as well as the M protein during virion assembly (Plempner and Lamb, 2021). Furthermore, the $\alpha 5$ -loop region has also been identified as the main interface in intercalating interactions in the sauronoid (Figure 5.19), as well as clam-shell formation in NDV, NiV, and SeV

NLFs (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). Similar clam-shell assemblies were observed in the data presented here, however failed to materialise into a useful 3D structure. The intercalating interface observed in the sauronoid assembly suffered from local poorly resolved electron density (Supplementary Figure S5.6C), which abrogated confident model building and forced the stubbing of residue side chains for Met114, Glu115, Arg116, Arg117, and Asp119. This may reflect the weak interactions that exist within the interface as stronger interactions would rigidify the loop side chains and may produce clearer density. The modelled alanines likely do not accurately portray the potential for extensive interactions these residues may engage in with neighbouring N protomers. Indeed, residues homologous to these have been identified as putative drivers of clam-shell interactions in NDV, NiV, and SeV (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). Focussed refinement on this area failed to improve map resolution, and likely reflects – in part – the inherent flexibility of both this loop and the intercalating interface. Similarly, local resolution around the HeV N protein Ct-arm α 17 was poor and while an alpha helix could confidently be built into the density, a number of side chains were also stubbed and modelled as alanines.

The electron density in the RNA binding groove enabled unambiguous, *de novo* modelling of an RNA chain. With one RNA hexamer per N protomer, HeV N resembles closely related viral N proteins and adheres to the paramyxoviral ‘rule of six’ (Kolakofsky et al., 1998). Furthermore, the typical ‘three in, three out’ base stacking orientation observed in other paramyxovirus N proteins, where the first three bases from the 3’-end face in towards the N protomer and the last three bases face the solvent, is also found in HeV N protein. The density representing the RNA chain, which almost certainly is bacterial host-derived, was not clear enough to distinguish specific bases likely due to averaging and symmetrisation of a heterogeneous sequence and the RNA was modelled – as is the convention – as a poly-uridine. It is unlikely that each sauronoid particle in the data set was fully occupied by two RNA chains of 84 nucleotides each. It is more likely that sauronoids have differential RNA occupation and that the RNA present was of varying length. The clear RNA density within each N protomer as well as the clear density surrounding the phosphate group bridging the RNA hexamers suggests that overall RNA occupancy in the sauronoids was high. A series of basic (Lys178, Arg192, Arg193, Arg352) and polar (Thr185, Gln199, Gln200, Tyr258, Gln319, Ser344, Tyr354) residues putatively interact with the RNA chain through electrostatic and hydrophobic interactions with both the backbone and the RNA bases. The majority of these interactions are with the RNA backbone, but a key interaction occurs between the amino acid dyad Gln199 and Gln200 and

RNA bases 5 and 6. This conserved interaction is also found among other paramyxoviruses, including MeV, NiV, and PIV5 (Alayyoubi et al., 2015; Desfosses et al., 2019; Ker et al., 2021).

Contiguous with the RNA binding groove exists a large cavity that is made up of two neighbouring N protomers. RNA bases 1, 2, and 3 which form the 'in' base stack point in towards this cavity, here referred to as the 'in-base cavity'. The in-base cavity can accommodate the larger purine bases compared to the modelled uridines and additional amino acid residues (Tyr258, Glu260, Glu261, Phe268, Arg272, Lys 321, Ser344, and Met345) may facilitate purine binding where pyrimidines are too distant. This same cavity exists in closely related N proteins, such as NiV, NDV, and MeV (Song et al., 2019; Desfosses et al., 2019; Ker et al., 2021).

Recently, a novel strain of HeV was identified in Australia, called HeV-var, which shares around 83% sequence identity with the reference strain used in this work (Annand et al., 2022). The N gene sequence is a little more conserved at 88% sequence identity, resulting in 18 amino acid substitutions of which 13 lie in the intrinsically disordered Ntail region (Supplementary Figure S5.16). Five substitutions are located in the Ncore region; S2G, D6E, and S13N in the Nt-arm, L153M in the NTD, and S386P in the Ct-arm (Figure 5.27, Supplementary Figure S5.16). Both D6E and S13N face the solvent of the sauronoid central hole, however Ser2 forms a putative elbow interface with its +2 neighbour. The substitution to Gly2 in the HeV variant could abrogate this elbow interface. The NTD-located L153M substitution sits at the upper surface of the NTD and faces the exterior solvent. It is unlikely to be involved in intercalation zone interactions. Finally, the S386P substitution in the Ct-arm faces the solvent in the sauronoid assembly, but likely forms part of the interface between successive helical turns in the nucleocapsid. These interactions are thought to contribute minimally to nucleocapsid stability due to the nucleocapsid's inherent flexibility and the observation that each helical turn consists of a non-integral number of N protomers (Cox et al., 2014). Taken together, the amino acid substitutions in the Ncore observed in the HeV variant strain are likely to have a minimal impact on HeV nucleocapsid formation and architecture. The large number of substitutions in the intrinsically disordered Ntail region, especially towards the C-terminus, could indicate a changed interaction with the RdRp complex and M proteins however these assumptions still await experimental validation.

The impartation of D14 symmetry to the sauronoid particle stack was essential to enable high-resolution reconstruction and ultimately protein model building and refinement. Doing so has clear advantages, since it generates a redundancy in the data and providing, in the case of D14 symmetry, a 28-fold increase in particle contributions to the final map. The experimental

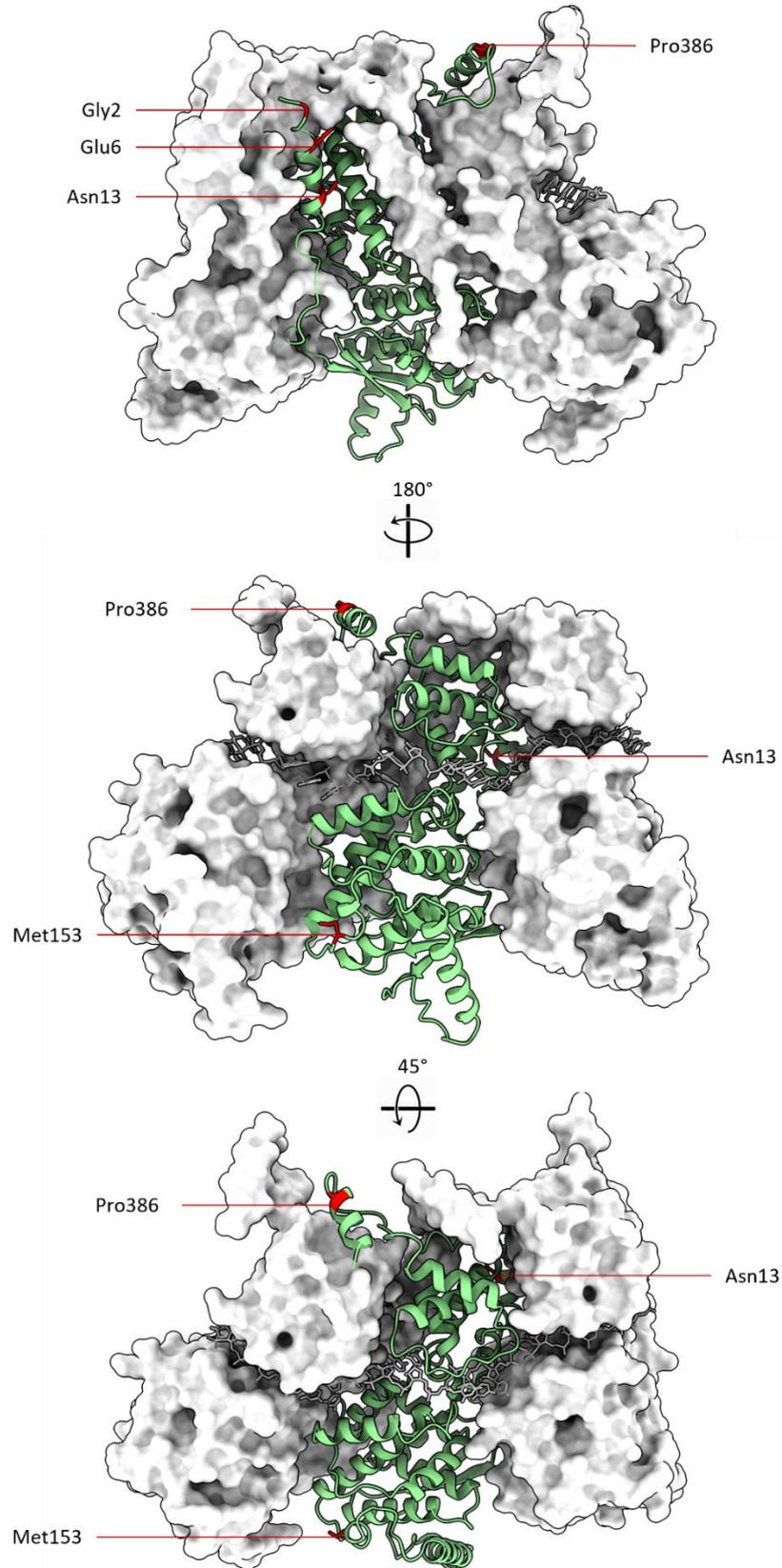


Figure 5.27. Mapping HeV variant amino acid substitutions.

Amino acid substitutions between the HeV reference strain and the HeV variant strain are mapped (red) onto the sauronoid HeV N monomer (lime) within a trimer assembly (white). The N_i protomer (lime) is visualised as ribbons and the N_{i+1} and N_{i-1} protomers are surface rendered (white). Residues present in the HeV variant strain are indicated.

stack only contained 5683 particles, but through symmetry impartation this resulted in an effective 157,864 asymmetric units. The drawbacks of symmetry impartation must also be considered. Symmetrised reconstructions struggle to detect any structural or compositional heterogeneity. Firstly, and as alluded to earlier in this section, the sauronoid assemblies are unlikely to be completely occupied by bacterial host-derived RNA. Any apo-state N protomers in the sauronoid assembly would be outnumbered by RNA-bound protomers and as a result are averaged out against them, evidenced by the clear RNA density. Secondly, there exist a large number of loop regions that are not predicted to be alpha-helical or beta-sheets. These loops may exhibit some conformational heterogeneity in otherwise uniform sauronoid assemblies, and as a result, their signal will be diluted through symmetry impartation. Indeed, some of these regions, foremost Met114-Asp119, are poorly resolved and as a result hindered confident model building. Non-symmetrised particles would likely shed more light on both the uniformity of RNA occupancy as well as regions of structural flexibility with the requirement for larger particle stacks.

5.3.2 *Hendra henipavirus* nucleocapsid architecture

The pursuit of a helical HeV NLF structure has resulted in only low-resolution electron density maps. These maps still offer information on HeV nucleocapsid architecture in the absence of high-resolution helical HeV NLF models. From the low-resolution maps generated through helical SPA overall NLF architecture can be deduced and globally compared to the helical structures of related viruses. The HeV NLF likely has a left-handed spiral turn, in line with closely and distantly related viruses such as PIV5, RSV, SeV, MeV, MuV, NDV, and NiV to name a few (Egelman et al., 1989; Schoehn et al., 2004; Bakker et al., 2013; Cox et al., 2014; Gutsche et al., 2015; Alayyoubi et al., 2015; Song et al., 2019; Ker et al., 2021). Like many of these viruses, the HeV NLF helical turn constituted approximately 13.2 N protomers. This non-integer number is indicative of the minimal contribution of interactions between successive helical turns and emphasises the stabilisation offered by lateral N-N contacts which maintain the linear integrity of NLFs. Indeed, the cryo-EM data set presented in Chapter 4 and the 2D classification of helical segments presented in this chapter illustrate the inherent and extensive flexibility of the NLFs even over short linear distances along the helix. These helical lateral contacts are likely to closely resemble the lateral interfaces described in this work for the sauronoid assembly. Further attempts should be made to obtain the helical HeV NLF structure. Continuing the helical SPA approach would involve generating higher-quality 2D classes through improved linear alignments. Alternatively, classical SPA may be attempted to generate helical non-symmetrised

segments similar to the approach taken for NiV (Ker et al., 2021) or cryo-ET may be employed to investigate the helical architecture of HeV NLFs.

Classification of helical segments also unearthed a small population of clam-shell transitions between NLFs with opposite polarities. Similar clam-shell structures have been described in greater detail and better resolution for NiV, NDV, and SeV (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). These clam-shell structures appear to form between two bottom-stacking NLFs and rely on many of the same interactions observed in the intercalation zone of the sauronoids described in this work. While the clam-shell interactions for NDV are only minor, those for NiV are extensive and at ~ 670 Å the buried surface area falls within the range of that calculated for the HeV sauronoid (Ker et al., 2021).

5.3.3 Summary

The work in this Chapter aimed to reconstruct a 3D electron density map from the HeV NLFs found on cryo-EM micrographs and build a model of the HeV N protein monomer as well as the HeV NLF architecture. During this process, a novel paramyxovirus multimeric N protein assembly was discovered and named the sauronoid assembly. The sauronoid assembly enabled high-resolution reconstruction of the HeV N protein structure to 3.5 Å. The data set also yielded NLF reconstructions but only to low-resolution.

Chapter 6: Generation of a *Hendra henipavirus* replicon system

6.1 Introduction

Gaining an understanding of the molecular mechanisms that govern a viral infectious cycle, from receptor binding to the budding of progeny virions, includes structural studies such as the work presented in Chapters 3, 4, and 5. The knowledge gained through structural elucidation should be functionally validated which often includes working with the virus itself. When laboratory-based experimentation with the virus of interest is impossible or impractical, due to for example restrictive biosafety measures, alternative solutions must be found. In many cases, experiments may be carried out with related but less infectious or pathogenic viral species. A good example is the work done on *Hazara orthonairovirus*, a close relative of *Crimean-Congo hemorrhagic fever orthonairovirus* (CCHFV) (Fuller et al., 2019; Mega et al., 2020). When no suitable relatives are available for study, research can instead be carried out using virus systems that seek to recapitulate selected stages of the viral replication cycle in isolation. For example, the study of viral attachment and fusion with the host membrane can be carried out with pseudotyped viruses or in cell fusion assays whilst being replication-incompetent (Cronin et al., 2005; Li et al., 2018). Similarly, viral replicon systems can be used to study the stages of virus-specific RNA synthesis in cell culture systems without generating replication-competent infectious particles.

6.1.1 Replicons in virus research

Viral replicon systems, alternatively called minigenome systems, offer a safe and effective approach to studying virus-specific RNA synthesis and gene expression. A viral replicon system generally consists of a sub-genomic nucleic acid that is a truncated version of the original viral genome with at least one gene essential for the viral multiplication cycle removed, but for which essential *cis*-acting signal sequences remain. Doing so generates a non-infectious replicon, for which replication and transcription depends on the expression, in *trans*, of supporting protein components that their truncated genomes no longer encode. Often, the deleted genes are substituted with non-viral reporter genes, which are chosen to permit easy detection and quantification of gene expression from the replicon template. Additionally, replicons can persist in cell lines and may be passed on during cell division. Replicon systems have been generated for a large variety of RNA viruses, including paramyxoviruses, some of which are reviewed in (Hannemann, 2020). Beyond the clear use in generating a fundamental, mechanistic understanding of viral genome expression and replication, these replicon systems also represent

a key tool for the development of novel therapies and vaccines. Their relative ease of use and self-replicating nature enables a rapid, efficient, and high-throughput system for the screening of candidate therapeutics that target viral transcription, translation, and replication (Hannemann, 2020).

6.1.2 *Hendra henipavirus* isolation and culturing

Shortly after the initial HeV outbreak in 1994, propagation of isolated virus was instrumental in the identification of the new pathogen. Indeed, the ability to isolate, culture, and passage a pathogen to establish an aetiology following Koch's postulates is still as applicable today as it was when he originally published them in 1890 (Lustig and Levine, 1992). The inoculation of spleen and lung homogenates from infected horses, collected during the initial outbreak, into a recipient horse resulted in symptomatic disease (Murray et al., 1995). The lung homogenate was also successful in infecting cell culture monolayers, specifically Vero cells, and causing syncytia formation. Post-mortem homogenates of infected human organs – including kidney, liver, lung, and spleen – successfully infected LLC-MK2 and MRC5 cultures and formed similar syncytia (Murray et al., 1995). Further experimentation in animal models has added guinea pigs, hamsters, and non-human primates to the list of species susceptible to HeV infection (Eaton et al., 2006; B. Lee et al., 2021). Furthermore, additional cultured cell types were used for HeV propagation, including human (HeLa, A549, PCI 13, SAEC) and non-human cells (MDBK, RK13, BHK21, and primary foetal equine kidney)(Eaton et al., 2006; B. Lee et al., 2021).

6.1.3 *In vitro* experimentation with *Hendra henipavirus*

Owing to its status as a hazard level 4 pathogen, experimentation with infectious HeV is restricted to high-containment biosafety level (BSL) 4 facilities. Yet, many key discoveries have been made in live virus systems using either wild-type or recombinant virus (reviewed in (B. Lee et al., 2021)). Nevertheless, the use of transient transfection of viral genes, pseudotyped virus, and replicon systems have increased accessibility to henipaviral research and elucidated many aspects of viral entry, replication, and egress.

For example, investigations into the cell tropism and viral entry mechanisms were carried out through cell fusion assays (Bossart et al., 2001; Bossart et al., 2002). A wide range of target cells, including those from rabbits, non-human primates, and mice were infected using a vaccinia virus-based platform with HeV F and G proteins to simulate HeV infection. In many cell

types, infection led to the formation of syncytia with neighbouring cells, indicating a wide cell tropism and confirming that the F and G proteins are both necessary and sufficient to drive adhesion and fusion (Bossart et al., 2001; Bossart et al., 2002). Furthermore, these pseudotyped viruses enabled the assessment of novel henipaviral therapeutic strategies in the form of a fusion inhibitor peptide (Bossart et al., 2002). A different pseudotyped virus, based on VSV, also employed the henipaviral glycoproteins to assess the anti-henipaviral activity of chloroquine during entry (Porotto et al., 2009).

Additionally, minigenome-based replicon assays designed for NiV confirmed the necessity of non-structural proteins for the replication of viral RNA (Halpin et al., 2004; Freiberg et al., 2008). Furthermore, the NiV non-structural proteins also rescued HeV minigenome replication in a heterotypic fashion, highlighting the close genetic and functional relationship between HeV and NiV. No complete, homotypic HeV minigenome replicon system exists in the published literature.

6.1.3 Chapter aims

This chapter aims to design and characterise a minigenome-based HeV replicon system for the functional validation of HeV N protein function. Following the establishment of the replicon system, site directed changes to the cDNA-expressed N gene within the replicon system will be guided by the HeV N protein structure. The protocols used to generate the results in this section are covered in Chapter 2 Materials and Methods, Sections 2.2.1 and 2.2.6.

6.2 Results

6.2.1 Primary design of a *Hendra henipavirus* replicon system

The initial design of a HeV replicon system was based on a published system for NiV (Halpin et al., 2004). This NiV replicon system uses a minigenome encoding a chloramphenicol acetyltransferase (CAT) reporter gene flanked by the NiV genomic termini as well as the NiV N gene and L gene 3' and 5' non-translated regions (NTRs), respectively. The minigenome is encoded on a DNA plasmid under transcriptional control of a T7 RNA polymerase (T7 pol) promoter, such that the initial T7 pol transcript is of negative, noncoding sense. Additionally, three support plasmids encoding the NiV N, P, and L genes are also under T7 pol transcriptional control. The minigenome and support plasmids were transfected into cells previously infected with the recombinant vaccinia virus MVA-T7 that encodes for T7 pol (Wyatt et al., 1995; Halpin

et al., 2004). When all components are present in the cells, T7 pol transcribes mRNA from the N, P, and L genes as well as generating the negative sense RNA minigenome. Translation of the N, P, and L mRNAs leads to the expression of the N protein, P protein, and L polymerase, all three of which are required and sufficient for paramyxoviral gene expression (Plempner and Lamb, 2021). The primary T7 pol transcript is assumed to be encapsidated with N. The viability of an unencapsidated RNA template is best characterised for *Alphainfluenzavirus influenzae* (formerly *Influenza A virus*) where only short unencapsidated RNA segments are replicated by the viral RdRp (Turell et al., 2013). Following T7 pol transcription, the viral RdRp complex (P protein and L polymerase) can then use the RNA minigenome as a template for transcription of CAT mRNA as well as the generation anti-minigenomes following the paramyxoviral replication strategies (Halpin et al., 2004; Plempner and Lamb, 2021). The expression of the CAT reporter is monitored through enzyme linked immunosorbent assay (ELISA) and northern blot as a measure of replicon efficiency (Halpin et al., 2004).

Following the strategy for the NiV replicon, largely the same approach was taken for the primary design of the HeV replicon system with a number of notable exceptions (Figure 6.1). Firstly, the need for coinfection with MVA-T7 was abrogated by the use of the BSR T7 cell line, a BHK-21 derivative that stably expresses T7 pol (Buchholz et al., 1999). BHK-21 cells were previously shown to be permissible for NiV infection and support NiV replication *in vitro* (Yoneda et al., 2006). Both the minigenome as well as the three support genes were placed under T7 transcriptional control, similar to the NiV system (Halpin et al., 2004). Cell lines that do not stably express T7 pol can be additionally transfected with a plasmid encoding T7 pol to broaden the range of suitable cell lines. Alternatively, RNA can be directly transcribed *in vitro* from the pUC57 plasmids for transfection into suitable cell lines.

The protein coding RNA sequences for the N, P, and L supports were taken from the published reference genome for HeV (NCBI Reference Sequence: NC_001906.3)(Wang et al., 2000) and no codon optimisation or other mutations were carried out. Similarly, the RNA sequences for the HeV genomic 3' leader and 5' trailer sequences, the HeV N gene start (GS) and 3' non-translated region (NTR), and the HeV L gene's 5' NTR and gene end (GE) were taken from the same reference genome (Wang et al., 2000). The CAT reporter gene in the NiV minigenome was replaced with an enhanced green fluorescent protein (EGFP) reporter gene (Cormack et al., 1996) (Figure 6.1). Using EGFP as a reporter allows for the use of fluorescence light microscopy techniques and live cell imaging to track replicon activity *in vitro*.

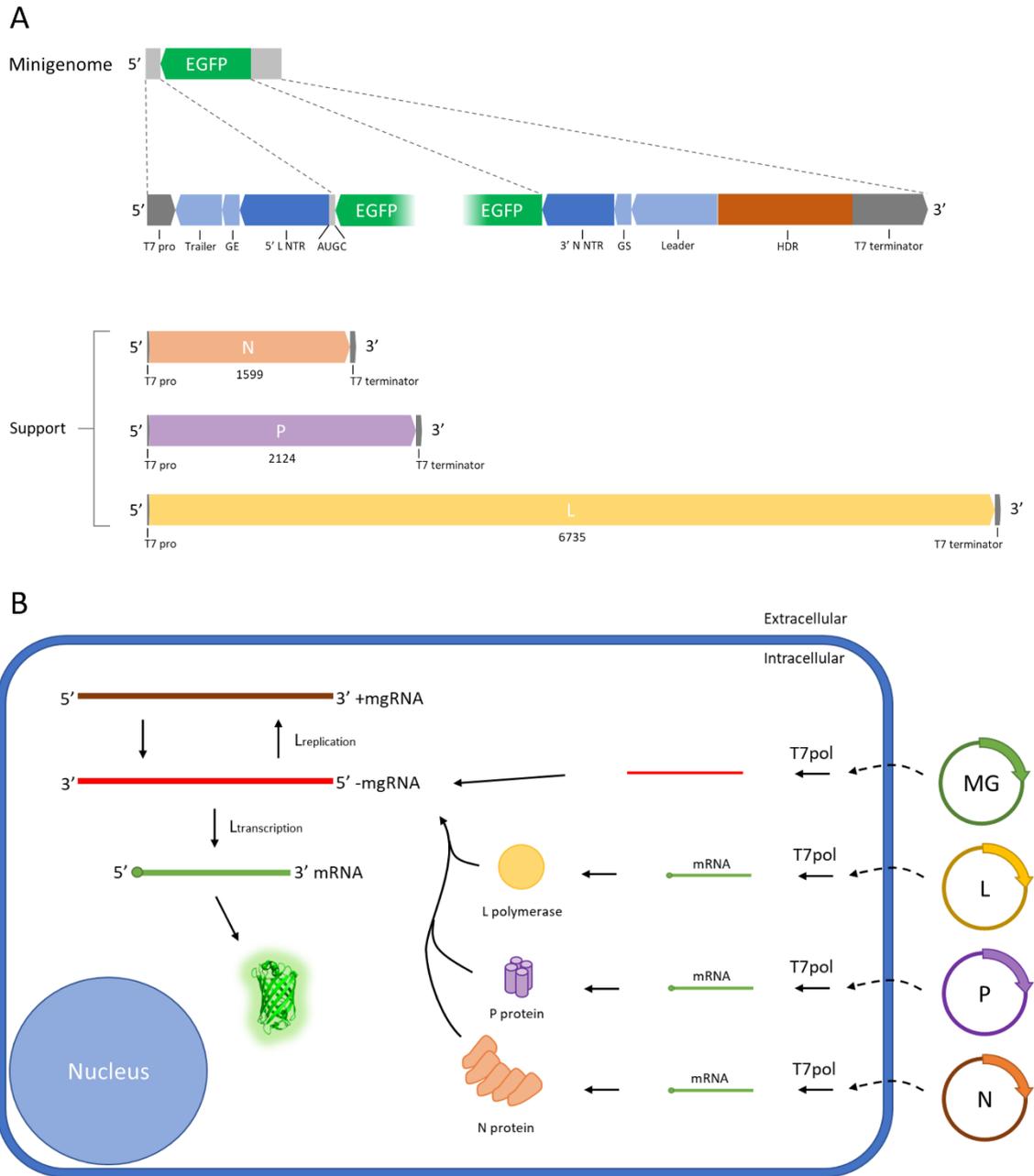


Figure 6.1. Schematic of HeV replicon design.

A) Schematic representation of the HeV replicon design. The minigenome and the three supports are all under T7 transcriptional control in a 5' to 3' direction. The supports encode for the nucleoprotein (N), phosphoprotein (P), and polymerase (L) in sense and are flanked by a T7 promoter (T7 pro) and terminator pair. The minigenome is similarly flanked by a T7 promoter and terminator pair but encodes for an antisense reporter genome followed by a hepatitis delta virus ribozyme cleavage site (HDR). The reporter genome contains, from 3' to 5', the HeV genomic 3' Leader sequence, the N gene start (GS) sequence and 3' non-translated region (NTR), an ORF encoding enhanced green fluorescent protein (EGFP), an AUGC spacer, the L gene 5' NTR and gene end (GE) sequence, and the genomic 5' Trailer sequence. Sections are to scale. **B)** Schematic representation of the HeV replicon approach. Following transfection, the T7 RNA polymerase (T7pol) present in the cell transcribes the minigenome (MG) and the three support plasmids, generating an unencapsidated negative sense RNA transcript of the MG and mRNA transcripts of the L, P, and N support genes. The support gene proteins are expressed and support replication of encapsidated negative sense minigenomic RNA (-mgRNA) and the 'antigenome' positive sense minigenomic RNA (+mgRNA) in replication mode ($L_{\text{replication}}$), or EGFP mRNA in transcription mode ($L_{\text{transcription}}$). EGFP is then expressed from this mRNA.

Finally, similar to the NiV minigenome, a hepatitis delta virus ribozyme (HDR) cleavage site was added following the 3' leader (Figure 6.1). The hepatitis delta virus ribozyme (HDR) is capable of self-cleavage leaving no additional nucleotides (Kuo et al., 1988). T7 pol termination can sometimes result in the addition of extra terminal nucleotides, resulting in a homogeneous 3' sequence. Since the HeV 3' leader is a defined sequence and likely essential for recognition by the viral RdRp, inclusion of the HDR directly adjacent should ensure the generation of a linear T7 transcribed RNA that terminates at the precise start of the 3' leader sequence.

The HeV minigenome and N, P, and L supports were cloned into pUC57 plasmids for use in mammalian cell systems (Figure 6.1, Supplementary Figure S6.1).

6.2.2 The designed *Hendra henipavirus* replicon system is inefficient *in vitro*

6.2.2.1 Using NiV replicon plasmid ratios results in poor HeV replicon performance

The DNA plasmid carrying the HeV minigenome (MG) as well as the three support plasmids coding for the HeV N, P, and L genes were transfected into BSR T7 cells before being placed in an IncuCyte live cell analysis system. This system automatically records multiple visual light and fluorescence images per well. Since the EGFP ORF is encoded in antisense in the minigenome, any EGFP fluorescence signal served as a read out of successful N protein encapsidation of the mini-genome template, followed by subsequent transcription of the EGFP reporter gene by the HeV RdRp. Further amplification of reporter gene expression indicates RdRp-specific replication that amplifies the initial T7 pol transcript.

BSR T7 cells were seeded in a 12-well format. The initial transfection followed the relative plasmid quantities used in the optimised NiV replicon system; NiV N 1.25 mg, NiV P 0.8 mg, NiV L 0.4 mg, NiV minigenome 3.5 mg (Halpin et al., 2004) (Figure 6.2). Normalised for the L plasmid, this equates to a ratio of 3.125:2:1:8.75 for N:P:L:MG. A pCMV-EGFP plasmid served as a transfection control while the exclusion of either the L support plasmid (no-L) or the minigenome (no-MG) served as negative controls, for which no HeV-specific reporter activity was expected. The pET-28a-6His-SUMO-N plasmid used for recombinant bacterial expression in Chapter 3 was used as a mock. The transfection mixture was replaced at 6 hours post-transfection (h.p.t.) with cell culture media. Following transfection, phase contrast images and green fluorescence were recorded hourly for 74 hours and analysed for green fluorescence object counts (Figure 6.2). While green fluorescence signal was high for pCMV-EGFP, the signal was low for the other transfections. In fact, the full replicon complement – consisting of the minigenome and the L, P, and N support plasmids – showed only marginally higher green object

counts than the no-L negative control, indicating either poor transfection or poor minigenome activity (Figure 6.2).

6.2.2.2 Optimising plasmid ratios fails to improve HeV replicon performance

Next, the ratios between the minigenome and support plasmids were optimised. Starting from the normalised ratios used before – which were each set to 1x – altered transfection mixtures were prepared with either 1.5x or 0.5x the starting concentration of one of the three support plasmids or the minigenome. BSR T7 cells were seeded in a 24-well format and transfected with the full replicon at 1x concentrations, or with 1.5x or 0.5x minigenome and support plasmids. The pCMV-EGFP served as a transfection control, the no-L as a negative control, and the pET-28a-6His-SUMO-N plasmid as a mock. The transfection mixtures were replaced at 6 or 24 h.p.t with cell culture media. Following transfection, phase contrast images and green fluorescence were recorded hourly for 100 hours and analysed for green fluorescence

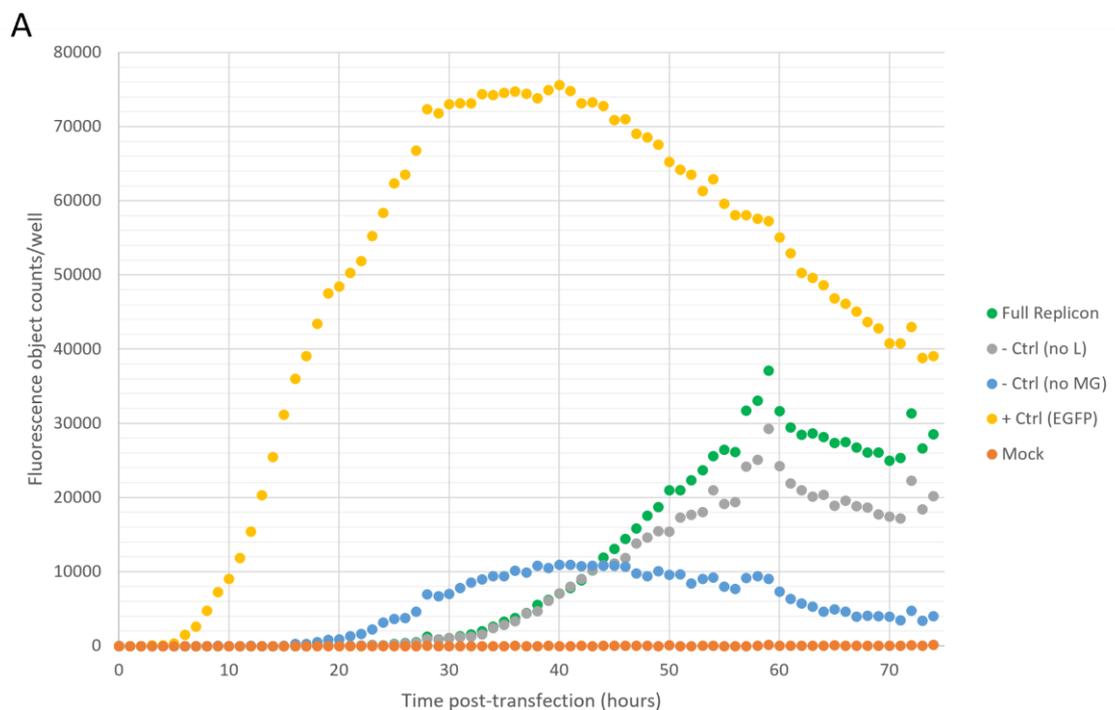


Figure 6.2. HeV replicon EGFP reporter levels are low.

Analysis of the HeV replicon using continuous live cell imaging. BSR T7 cells were seeded in a 12-well format and transfected with a complete set of minigenome and support plasmids (Full Replicon), transfected without the L support plasmid (no L) or the minigenome (no MG) as negative controls, or mock transfected with pET28a-6His-SUMO-N (Mock). Transfection with pCMV-eGFP served as a positive control for transfection. The transfection mixture was replaced at 6 h.p.t with cell culture media. Cells were scanned hourly for phase contrast and green fluorescence, representing EGFP expression, using an IncuCyte ZOOM and analysed for total green object counts per well. Biological repeats n=1.

object counts (Figure 6.3, Figure 6.4). Again, transfection efficiency for pCMV-EGFP was good in both the 6-hour (Figure 6.3A) and 24-hour (Figure 6.4A) transfections. The curve representing green fluorescence object counts for the full replicon at 1x ratios closely matched the no L negative control after a 6-hour transfection (Figure 6.3A) or fell below the no L negative control after a 24-hour transfection (Figure 6.4A). For both the 6- and 24-hour transfections, increasing the N support to 1.5x had a negative effect on fluorescence count, however decreasing the N support to 0.5x either had no effect (Figure 6.3B) or a positive effect (Figure 6.4B) on fluorescence counts compared to the 1x baseline. Altering the P support concentration to 1.5x or 0.5x had either a negative effect or no effect on fluorescence counts at 24-hour transfection (Figure 6.4C), however both increase and decreasing the concentration of the P support had a negative effect on fluorescence counts for the 6-hour transfection (Figure 6.3C). Similar to P support at 24-hour transfection, altering the L support concentration to 1.5x or 0.5x had no strong effect on fluorescence counts at 6-hour or 24-hour transfections (Figure 6.3D, Figure 6.4D). Finally, reducing the concentration of the MG plasmid to 0.5x had a strong negative effect on total fluorescence counts for both 6- and 24-hour transfections, however increasing the concentration to 1.5x had no appreciable effect for the 6-hour transfection but a strong positive effect for the 24-hour transfection (Figure 6.3E, Figure 6.4E). Taken together, these results suggest that the full replicon complement at 1x ratios performs poorly at both 6-hour and 24-hour transfection windows, generating a similar fluorescence object count as the no-L negative control at 6-hour transfection or fewer counts at 24-hour transfection. Furthermore, increasing the relative concentration of the MG or decreasing the relative concentration of the N support in a 24-hour transfection window had a positive effect on overall green fluorescence object counts.

6.2.3 Alterations to the *Hendra henipavirus* replicon system support plasmids

The poor performance of the replicon system could be due to inefficient expression of the N, P, and/or L supports. The support plasmids in the initial design contained only the protein coding ORFs of the genes in question and each of the genes' gene start (GS) and non-translated region (NTR) were absent. Additionally, the AUG start codon followed on directly from the T7 promoter sequence and missed a crucial G-spacer. Finally, the P support also contains the ORFs for the C, V, and W non-structural proteins which are expressed through alternative translation initiation or mRNA editing involving the insertion of pseudo-templated G nucleotides (Plempner and Lamb, 2021; B. Lee et al., 2021) (see Chapter 1 Section 1.7.2.1). The P gene products C, V, and W are known modulators, through different actions, of RNA synthesis during the

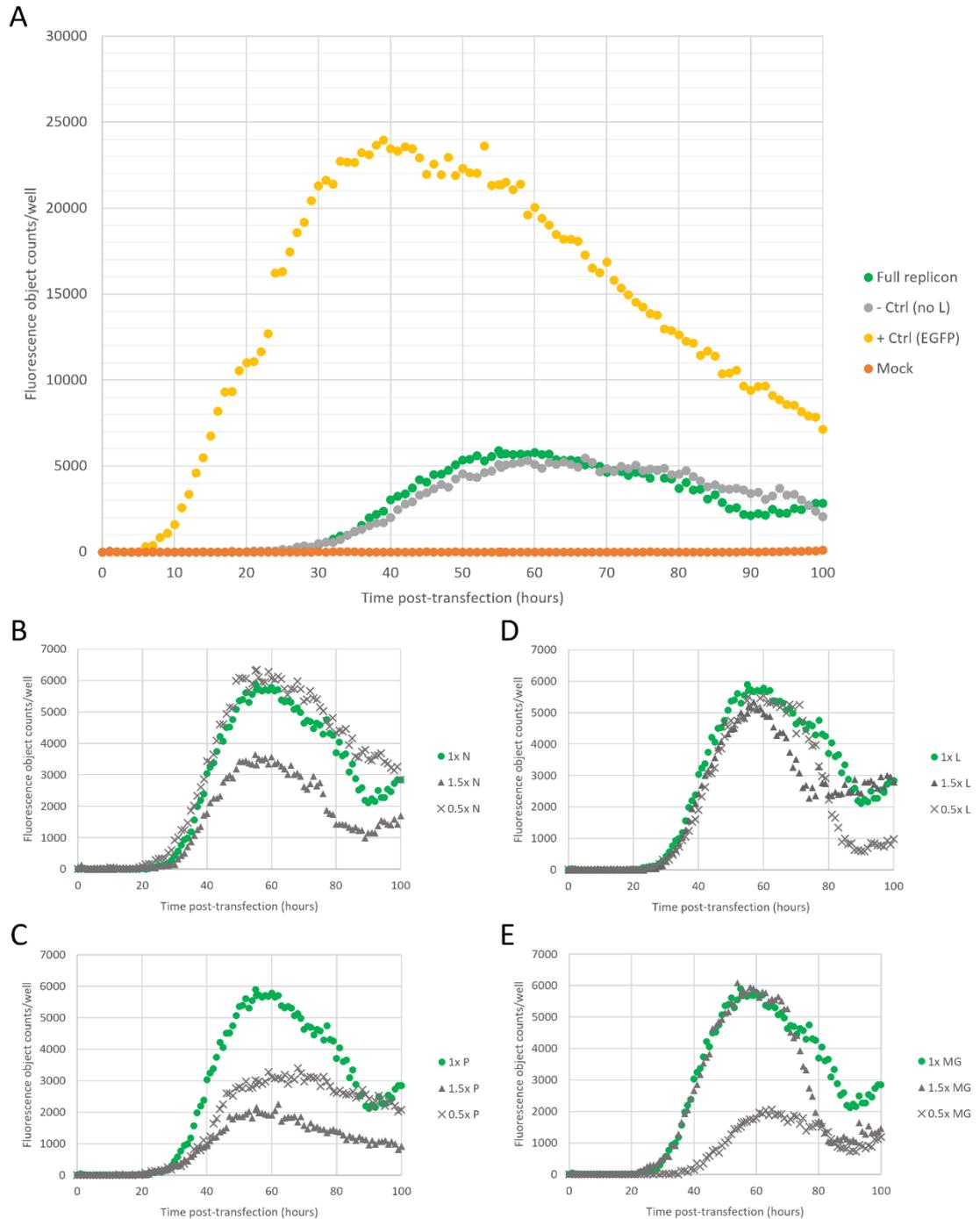


Figure 6.3. HeV replicon EGFP reporter levels are low after 6 hours transfection.

Analysis of the HeV replicon using continuous live cell imaging. BSR T7 cells were seeded in a 24-well format and transfected with a complete set of minigenome and support plasmids at the 1x concentrations (Full replicon) or a complete set where the amount of one of the four plasmids was increased to 1.5x or decreased to 0.5x. Cells were alternatively transfected without the L support plasmid (no L) as a negative control, or mock transfected with pET28a-6His-SUMO-N (Mock). Transfection with pCMV-eGFP served as a positive control for transfection. The transfection mixture was replaced at 6 h.p.t with cell culture media. Cells were scanned hourly for phase contrast and green fluorescence, representing EGFP expression, using an IncuCyte ZOOM and analysed for total green object counts per well. **A**) The green fluorescence object counts were plotted for the Full replicon as well as the positive and negative controls and the Mock. **B-E**) The green fluorescence object counts were plotted for the Full replicon, as in **(A)**, and cells transfected with the variable concentrations of the N support **(B)**, P support **(C)**, L support **(D)**, or minigenome (MG) **(E)**. Biological repeats n=1.

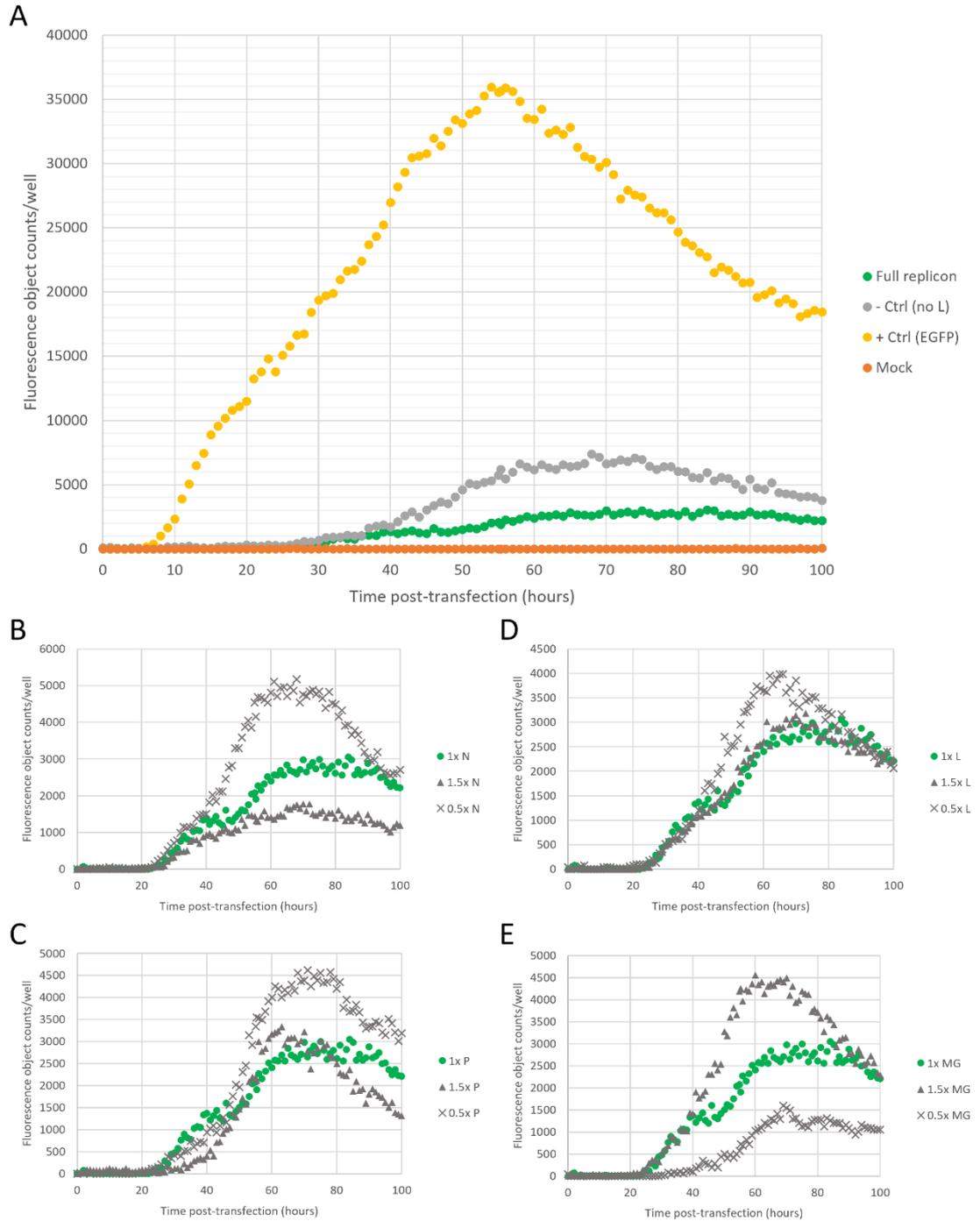


Figure 6.4. HeV replicon EGFP reporter levels are low after 24 hours transfection.

Analysis of the HeV replicon using continuous live cell imaging. BSR T7 cells were seeded in a 24-well format and transfected with a complete set of minigenome and support plasmids at the 1x concentrations (Full replicon) or a complete set where the amount of one of the four plasmids was increased to 1.5x or decreased to 0.5x. Cells were alternatively transfected without the L support plasmid (no L) as a negative control, or mock transfected with pET28a-6His-SUMO-N (Mock). Transfection with pCMV-eGFP served as a positive control for transfection. The transfection mixture was replaced at 24 h.p.t with cell culture media. Cells were scanned hourly for phase contrast and green fluorescence, representing EGFP expression, using an InCuCyte ZOOM and analysed for total green object counts per well. **A)** The green fluorescence object counts were plotted for the Full replicon as well as the positive and negative controls and the Mock. **B-E)** The green fluorescence object counts were plotted for the Full replicon, as in **(A)**, and cells transfected with the variable concentrations of the N support **(B)**, P support **(C)**, L support **(D)**, or minigenome (MG) **(E)**. Biological repeats n=1.

paramyxoviral replication cycle (Gainey et al., 2008; Boonyaratanakornkit et al., 2011; Plemper and Lamb, 2021; B. Lee et al., 2021). Indeed, the C ORF was also silenced in the NiV replicon system in order to study the function of the P protein in the absence of the C protein (Halpin et al., 2004).

Consequently, a number of alterations were made to the initial replicon design (Figure 6.5, Supplementary Figure S6.1). Firstly, the minigenome remained unchanged and is identical to the one presented above (Figure 6.1). The N, P, and L support plasmids were modified to include the respective GS and NTR sequences upstream of the ORFs (Figure 6.5A). Finally, two consecutive putative start codons for the C ORF were identified in the P gene (Wang et al., 2000). These start codons were mutated from ATG ATG to ACG ACG and the TCG codon in position 4 was mutated to a TAG stop codon to generate a C protein deficient P gene (PdC) that maintains P ORF protein coding (Figure 6.5B).

DNA segments containing the designed alterations were ordered with additional 5' and 3' sequences matching the respective support plasmids to enable their use as an insert for restriction digestion and ligation, as well as being flanked by shared PCR primer target sequences to allow for easy amplification. Following PCR amplification, the insert sequences as well as the original pUC57 support plasmids were digested with the appropriate restriction enzymes, ligated, and heat-shock transformed into DH5 α cells for amplification. These cells were then cultured overnight, and plasmids were prepared for further transfection experiments.

6.2.4 The redesigned *Hendra henipavirus* replicon system is inefficient *in vitro*

Following their redesign, the L, PdC, and N support plasmids as well as the HeV MG were transfected into BSR T7 cells, previously seeded in a 24-wells format, using the 1x ratios before being placed in the live cell analysis system. The pCMV-EGFP served as a transfection control, the exclusion of the L support (no-L) as a negative control, and the pET-28a-6His-SUMO-N plasmid as a mock. The transfection mixtures were replaced at 6 h.p.t with cell culture media. Following transfection, phase contrast images and green fluorescence were recorded every two hours for 70 hours and analysed for confluency-adjusted green fluorescence integrated intensity (GCU x $\mu\text{m}^2/\text{Image}$) rather than green object counts (Figure 6.6). A strong signal for the pCMV-EGFP control indicated good transfection (Figure 6.6A). In addition to the no-L control, the PdC and N support plasmids were also individually excluded (no-PdC and no-N, respectively), with the strongest attenuation of green fluorescence integrated intensity in the no-L control, a reduced attenuation for no-PdC, and no discernible effect for no-N (Figure 6.6B).

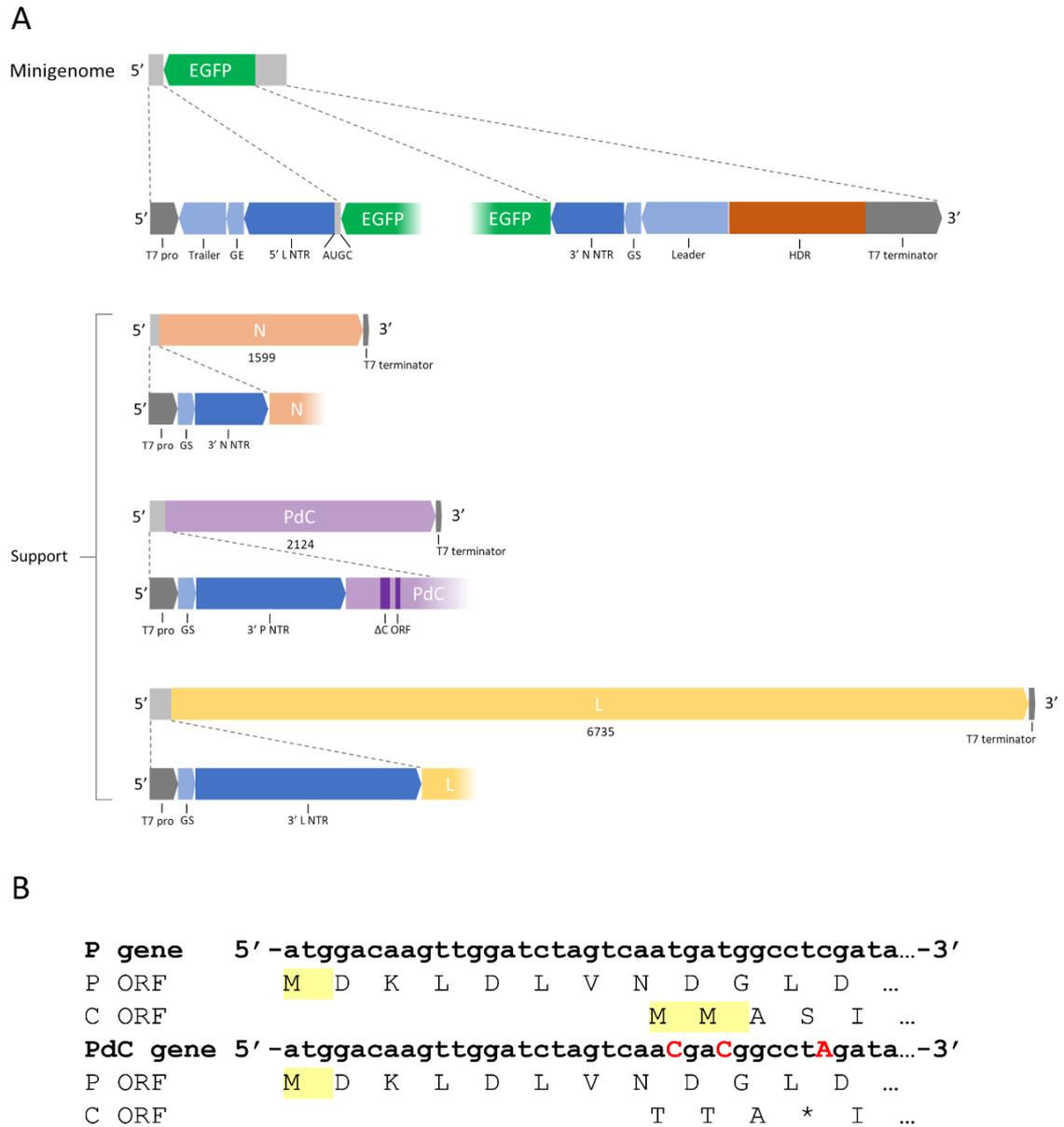


Figure 6.5. Schematic of HeV replicon redesign.

A) Schematic representation of the HeV replicon redesign. The minigenome has remained unchanged and it as well as the modified supports follow the same nomenclature as in Figure 6.1. The minigenome as well as the three modified supports are under T7 transcriptional control in a 5' to 3' direction. The modified supports encode for the nucleoprotein (N), phosphoprotein (P), and polymerase (L) in sense and are flanked by a T7 promoter (T7 pro) and terminator pair. They also include each of the genes' gene start (GS) sequence and 3' non-translated region (NTR) and an additional G spacer nucleotide between the T7 promoter and the GS sequence. Finally, two putative start codons and a the Ser4 codon in the C ORF have been mutated (dark purple bars) to generate a C deficient P gene (PdC). Sections are to scale. **B)** Sequence alignment and translation of the start of the P ORF and C ORF in the native P gene and the C ORF-deficient P gene (PdC). Mutated nucleotides are capitalised and indicated in red. Putative start codons are indicated in yellow in the respective translated sequences. Stop codons are marked with *.

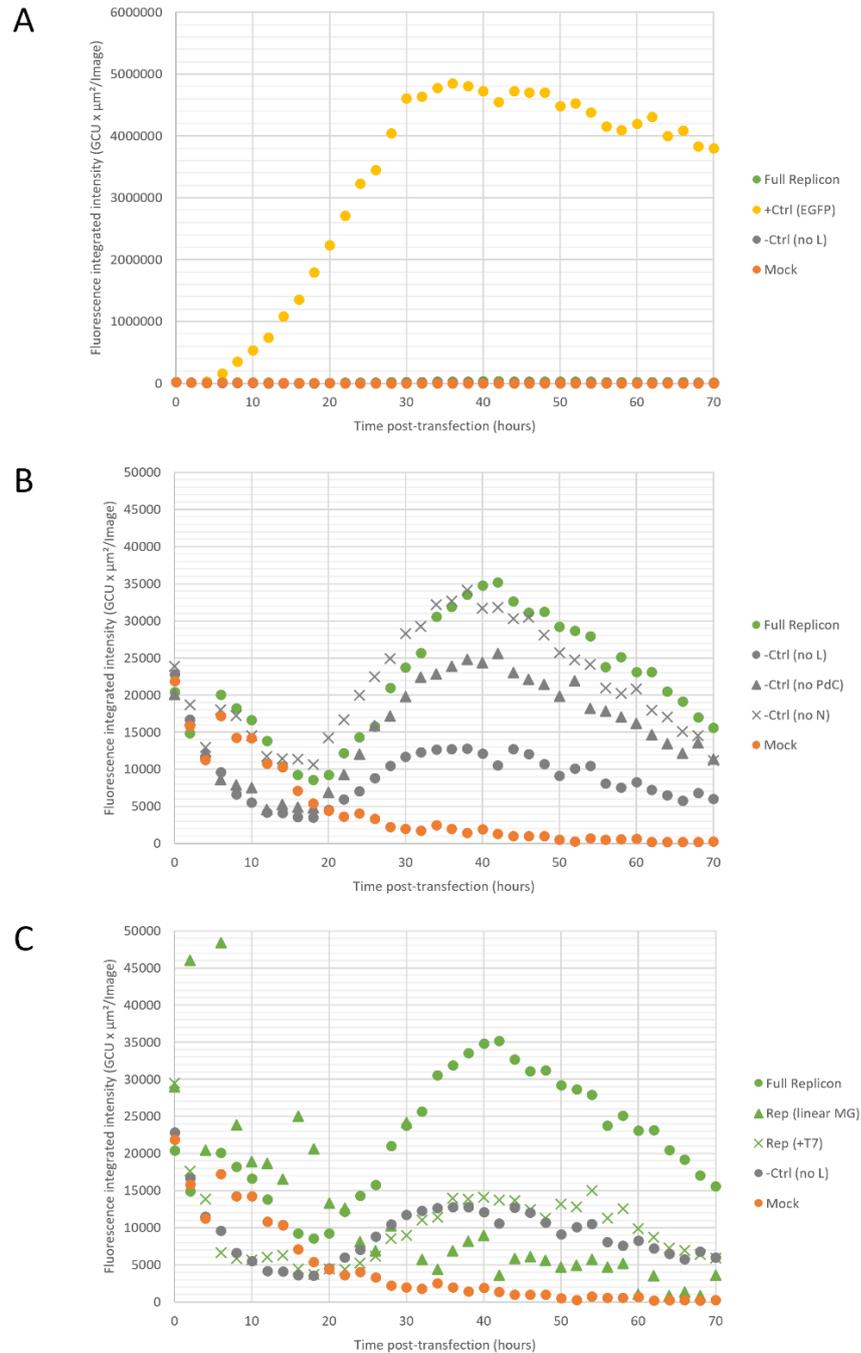


Figure 6.6. HeV replicon redesign reporter levels are low.

Analysis of the redesigned HeV replicon using continuous live cell imaging. BSR T7 cells were seeded in a 24-well format and transfected with a complete set of minigenome and support plasmids. Cells were alternatively transfected without the L support plasmid (no L) as a negative control, or mock transfected with pET28a-6His-SUMO-N (Mock). Transfection with pCMV-eGFP served as a positive control for transfection. The transfection mixture was replaced at 6 h.p.t with cell culture media. Cells were scanned hourly for phase contrast and green fluorescence, representing EGFP expression, using an IncuCyte ZOOM and analysed fluorescence integrated intensity (GCU x $\mu\text{m}^2/\text{Image}$) which was adjusted for cell confluency. **A)** The fluorescence integrated intensity was plotted for the full replicon as well as the positive and negative controls and the mock. **B)** The fluorescence integrated intensity was plotted for the full replicon, the no-L negative control and the Mock. Additionally, transfections without the PdC support (no PdC) or N support (no N) were also carried out and plotted. **C)** The fluorescence integrated intensity was plotted for the full replicon, the no-L negative control and the Mock. Additionally, transfections with a previously linearised minigenome (Rep (linear MG)) and the full replicon complement with the addition of a T7 polymerase-expressing plasmid (Rep (+T7)) were also carried out and plotted. Biological repeats n=1.

Concomitantly, the efficiency EGFP expression from the pUC57-HeV-minigenome was assessed. Firstly, an additional T7 polymerase-expressing plasmids was co-transfected into BSR T7 cells with the full replicon complement (consisting of the minigenome and the L, PdC, and N support plasmids) (Figure 6.6C). Secondly, a previously linearised pUC57-HeV-minigenome plasmid was transfected into BSR T7 cells along with the L, PdC, and N supports as a replacement for the circular minigenome (Figure 6.6C). In the event of a weak T7 terminator, the circular minigenome plasmid could result in concatenated minigenome RNA transcripts that fail to express EGFP. Linearisation was performed through restriction digestion down-stream of the T7 terminator. Both the linearisation of the minigenome plasmid as well as the addition of the T7 polymerase-expressing plasmid had a negative effect on green fluorescence integrated intensity (Figure 6.6C).

Overall, the redesigned support plasmids seem to have a positive effect on replicon performance, with a maximum EGFP signal approximately three-fold higher than the no-L negative control (Figure 6.6) suggesting RdRp-specific transcription of the EGFP reporter gene. The HeV replicon system should enable the functional investigation of the effect of mutations in the HeV N protein on reporter output. However, the EGFP signal for the no-N control was similar to that of the full replicon complement (Figure 6.6B). The no-N control is a critical control as its exclusion should reflect the need for the N protein within the replication cycle, and this necessity is not observed presently.

6.2.5 Transfection of *in vitro* transcribed RNA for the support plasmids fails to improve replicon efficiency

In the pursuit of a better performing HeV replicon system, with higher overall reporter gene transcription, the pUC57-HeV-minigenome and L, PdC, and N supports were *in vitro* transcribed to RNA ahead of transfection into BSR T7 cells. *In vitro* transcription was performed using the T7 polymerase and following the reaction the RNA was purified and used for transfections. Cells were previously seeded in a 24-well format and the exclusion of the L support transcript (no-L) as a negative control while yeast tRNA transfection served as a mock. The transfection mixture was replaced at 6 h.p.t with cell culture media or not replaced for the duration of the experiment (NoX). Following transfection, phase contrast images and green fluorescence were recorded hourly for 50 hours and analysed for confluency-adjusted green fluorescence integrated intensity (Figure 6.7). The transcribed replicon performed poorly in both the media-replaced and NoX transfections, with fluorescence integrated intensities similar to or

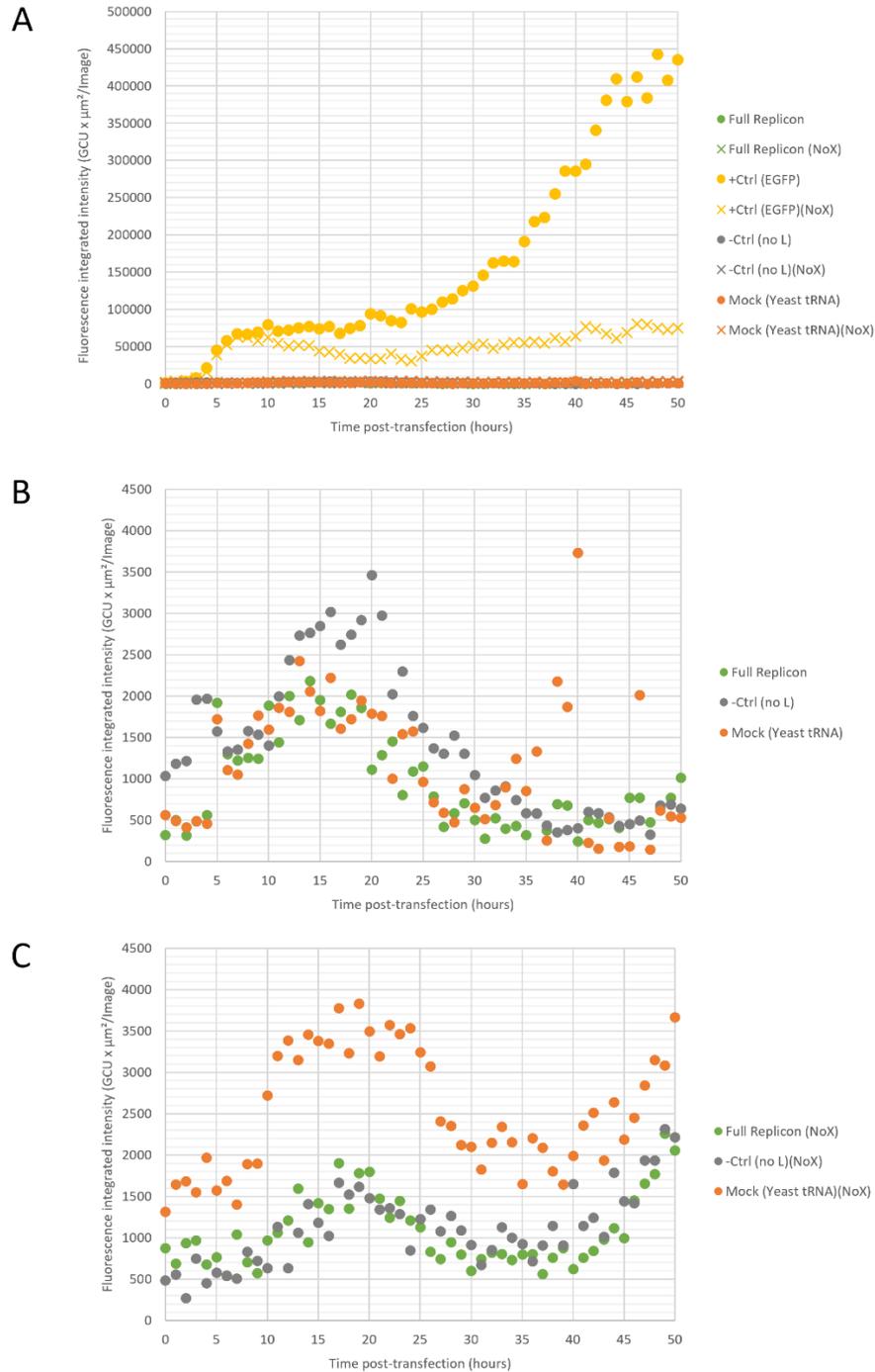


Figure 6.7. RNA transfection of HeV replicon redesign.

Analysis of the redesigned HeV replicon using continuous live cell imaging. BSR T7 cells were seeded in a 24-well format and transfected with a complete set of T7-transcribed RNA minigenome and support RNA transcripts. Cells were alternatively transfected without the L support transcript (no L) as a negative control, or mock transfected with Yeast tRNA (Mock). Transfection with pCMV-eGFP served as a positive control for transfection. The transfection mixture was replaced at 6 h.p.t with cell culture media or not replaced for the duration of the experiment (NoX). Cells were scanned hourly for phase contrast and green fluorescence, representing EGFP expression, using an IncuCyte ZOOM and analysed fluorescence integrated intensity (GCU x $\mu\text{m}^2/\text{Image}$) which was adjusted for cell confluency. **A**) The fluorescence integrated intensity was plotted for the full replicon as well as the positive and negative controls and the mock for both the 6-hour media exchanged and non-exchanged (NoX) wells. **B**) The fluorescence integrated intensity was plotted for the 6-hour transfection of the full replicon, the no-L negative control and the mock. **C**) The fluorescence integrated intensity was plotted for the media non-exchanged (NoX) transfection of the full replicon, the no-L negative control and the mock. Biological repeats n=1.

lower than the Yeast tRNA mock (Figure 6.7). It is likely these results inaccurately reflect the performance of transcribed replicon plasmids as the experiment suffered from both poor RNA quality (Supplementary Figure S6.2) and low cell viability (Supplementary Figure S6.3).

6.3 Discussion

A homotypic minigenome-based replicon system for HeV does not exist in the published literature. Previous work has been done on minigenome-based replicon systems for the closely related NiV (Halpin et al., 2004; Freiberg et al., 2008). In these systems, the NiV N, P, and L genes were able to support the NiV minigenome as well as showing heterotypic support for a HeV minigenome, highlighting the close genetic relationship between HeV and NiV (Halpin et al., 2004). While paramyxoviral genomes are normally fully encapsidated within the nucleocapsid, a NiV L-P fusion protein was shown to produce RNA transcripts from unencapsidated templates upon recognition of the first promoter element (PrE-I) (Jordan et al., 2018), validating the DNA plasmid-based replicon system approach. Although the NiV-based replicon system enables henipaviral research in the absence of BSL-4 facilities, the genetic nature of the support plasmids – in particular the N support – abrogates their use for the functional validation of the structural data obtained for the HeV N protein described in Chapter 5 and as a consequence, the design and characterisation of a homotypic HeV minigenome-based replicon system is treated in this chapter.

6.3.1 The primary HeV replicon design is inefficient in BSR T7 cells

The primary design followed the strategy employed for the NiV system (Halpin et al., 2004), with some differences in the minigenome to enable fluorescence-based reporter gene analysis. Overall, this system performed poorly, with fluorescence levels close to that of the negative control which missed the L polymerase. Leaky expression of EGFP could explain the low levels of fluorescence irrespective of the presence of the L support. However, the EGFP ORF is present in reverse on the minigenome plasmid and would first require the generation of a negative sense RNA transcript before a positive RNA transcript could serve as mRNA for EGFP expression. In the absence of an RdRp, the expression of EGFP in this manner is exceedingly unlikely. Additionally, the systems described for NiV uses a vaccinia virus vector to introduce T7 pol (Halpin et al., 2004), which may exhibit different polymerase stoichiometry compared to the stably transfected BSR T7 used in this work.

Instead, the low efficiency of the HeV replicon system could be due to the sensitivity of paramyxoviral systems to the relative ratios of the minigenome and support plasmids. Indeed, a T7-free NiV minigenome-based replicon system showed high sensitivity for the ratio between the N, P, and L supports (Freiberg et al., 2008). Due to limitations in time and resources, a limited plasmid ratio optimisation was carried out for the HeV replicon system. The minigenome, L, P, or N plasmids were individually increased or decreased 1.5-fold or 0.5-fold, respectively, compared to the starting concentrations. While in most cases there was no strong difference, a few results stood out. As expected, increasing the minigenome concentration resulted in a stronger EGFP signal while decreasing the concentration reduced the EGFP signal, since either more or fewer cells would be transfected with an EGFP carrying plasmid. Interestingly, a reduction in the N support concentration also had a noticeable positive effect on EGFP signal. The viral RdRp alternates between transcription and replication modes, and this alternation is – in large part – dependent on the relative N protein concentration (see Chapter 1 Section 1.7.2) (Plempner and Lamb, 2021). Reducing the availability of N protein in the replicon system could favour transcription of the EGFP ORF over generation of anti-minigenome.

6.3.1 The inclusion of viral gene NTRs improves replicon performance

The importance of relative ratios of the support plasmids echoes a more basal consequence of the effective transformation of a non-segmented viral genome into a segmented replicon. Normally, the ratios of N, P, and L protein are governed by imperfect transcription reinitiation after a gene boundary, which is the result of both the distance from the genomic start as well as the efficiency of the specific gene's gene boundary and NTR (see Chapter 1 Section 1.7.2) (Plempner and Lamb, 2021). The generation of the replicon system removes this inherent control of the relative protein concentrations. In an attempt to improve the efficiency of support protein expression, the support plasmids were altered to include each gene's respective NTR sequence. Furthermore, the C protein, which is encoded within the P gene, is known to have a negative effect on L polymerase activity (Plempner and Lamb, 2021). Indeed, the C ORF was also silenced in the NiV replicon system, albeit to study the function of the P protein in the absence of C (Halpin et al., 2004). Therefore, the P support plasmid was further altered to generate a C-silenced P gene; PdC.

Following these alterations, the performance of the replicon system improved slightly. EGFP signal was now around 3-fold higher than the negative control, however this difference is insufficient to be considered a robust platform for experimentation with N protein mutations or

therapeutic candidates. *In vitro* RNA transcription of the minigenome and support plasmids failed to produce a better result, likely due to both low quality of the RNA and poor cell viability.

6.3.2 Suggested optimisation steps for future work

Further assessment of the support plasmid ratios should be undertaken to ensure the optimal ratio between each plasmid is used. A large-scale titration experiment could be informative, as has been done for the T7-independent NiV system (Freiberg et al., 2008), although carrying this out will be both time and resource intensive.

Furthermore, cell lines other than BSR T7 may be tested. Indeed, a large number of primary and immortalised cell lines are permissive for HeV replication (B. Lee et al., 2021). Since the HeV replicon system presented here is T7 polymerase dependent, any alternative cell lines used that do not stably express T7 polymerase either need to be co-transfected with a T7-expressing plasmid, infected with a T7-expressing recombinant virus (Halpin et al., 2004), or transfected with previously *in vitro* generated RNA transcripts of the minigenome and support genes. Both the inclusion of an additional T7 polymerase-expressing plasmid or the *in vitro* transcription of the minigenome and support genes has seen a limited attempt here, with few results. As such, improving the quality of RNA transcripts and ensuring the optimal viability of the cell monolayer should enable the assessment of this approach more robustly.

6.3.4 Summary

The aims for this chapter were to design a minigenome-based HeV replicon system for HeV for the functional validation of structural data. While the initial design performed poorly, a redesign of the support plasmids showed a small but promising improvement. Further optimisation is needed to generate a robust system for down-stream experimentation.

Chapter 7: Discussion

The work presented in this thesis was undertaken within the framework of the MSCA-ITN HONOURS consortium, its organisation motivated by the continuous risk of emerging zoonotic viruses (van der Hoek et al., 2018). In particular, the pages of this work deliver a disquisition on the structural and molecular investigations into the *Hendra henipavirus* nucleoprotein and the nucleocapsid multimers it forms. Importantly, the genus *Henipavirus* features on the very short list of priority diseases for research and development, maintained by the WHO (World Health Organization, 2022). The HeV N protein structure was solved by recombinant protein expression and transmission-electron microscopy, revealing a novel nucleocapsid-like architecture here referred to as the ‘sauronoid’. Additionally, a HeV minigenome-based replicon system was designed and trialled to enable functional investigation of the structural data. This chapter will summarise and contextualise the results obtained within this project and provide a guide for future inquiries based on this data.

7.1 Expression and purification of recombinant *Hendra henipavirus* nucleoprotein

Recombinant protein expression in bacterial hosts enables the study of viral proteins in the absence of high containment facilities. Indeed, the recombinant expression of HeV N protein was performed in an *E. coli* expression system for this reason, with the aim of enabling high-resolution reconstruction of the HeV N protein through cryo-EM. The use of (pET-)plasmid-based approaches has proved successful for the generation of recombinant viral protein for downstream applications (Carter, Barr, et al., 2012; Carter, Surtees, et al., 2012; Surtees et al., 2015; Alayyoubi et al., 2015; Pearce et al., 2015; Song et al., 2019; Desfosses et al., 2019; Álvarez-Rodríguez et al., 2020; Hoste et al., 2021; Ker et al., 2021; Zhang et al., 2021). In many of these, a tandem purification approach was taken where a poly-histidine fusion tag enabled primary purification over Ni-NTA columns and secondary purification was carried out through size exclusion chromatography.

Building on a strong basis of recombinant expression of viral nucleocapsid proteins, the HeV N protein was expressed from a pET28a plasmid using a hexahistidine-SUMO fusion tag for purification. The initial expression parameters were based on those described for HeV N in the literature (Pearce et al., 2015), but required further optimisation using locally available systems. Following tandem purification, this robust expression system yielded high quantities of pure HeV

N protein for structural elucidation.

In recent years, the recombinant expression and purification of paramyxoviral N proteins have yielded high-resolution cryo-EM reconstructions (Alayyoubi et al., 2015; Desfosses et al., 2019; Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). While all of these studies purified the respective N proteins through N-terminally fused poly-histidine affinity tags, with the exception of MeV which used a C-terminal poly-histidine tag instead (Desfosses et al., 2019), none of the strategies described in these publications attempted to remove the poly-histidine fusion tag. The effect of the inclusion of this tag on the N protein structures and their interpretation is discussed in detail in Chapter 5 and in Section 7.3 below.

7.2 The *Hendra henipavirus* N protein and nucleocapsid imaged through electron microscopy

Electron microscopy, as the prominent tool used in this thesis, enabled firstly the visualisation of protein samples following recombinant expression and purification. Understanding the gross morphologies present in the sample as well as sample behaviour in different physiological conditions through low-resolution negative stain and cryo-EM screening ensures that only high-quality sample is taken forward.

Negative stain EM revealed three predominant morphologies for recombinantly expressed HeV N protein: filamentous tubes of various lengths, circular rings, and squat rings. From the relatively low-resolution 2D images it was reasonable to assume that these morphologies represent different views or orientations of two particle types (filaments and sauronoids), providing longitudinal and transverse views of both.

The generation of cryo-grids for high-resolution data collection and structural elucidation required a modicum of optimisation, revealing a tendency for the HeV N particles to associate with the carbon layers. Ultimately, ultra-thin carbon lacey grids performed best during cryo-grid screening and enabled the collection of 3548 micrographs for 3D reconstruction.

Ahead of 3D reconstruction, the unprocessed micrographs of HeV N revealed many similarities with other recombinantly expressed HeV NLFs (Pearce et al., 2015), HeV nucleocapsids of isolated from cultured virus particles (Hyatt et al., 2001), and recombinantly expressed NLFs from the closely related NiV (Ker et al., 2021). The overall architecture exhibits an identical herringbone morphology that is shared between recombinant NLFs and isolated nucleocapsids of HeV and is ubiquitous among members of the *Paramyxoviridae* family (Chui

and Zhou, 2013; Rima et al., 2019; Plemper and Lamb, 2021; B. Lee et al., 2021). The range in NLF lengths likely reflects the inclusion of random-length *E. coli*-derived RNA. Particles of all three morphologies observed in negative stain- and cryo-EM share a consistent width at around 20 nm, which is in line with the range observed for other *Paramyxoviridae* family members (Chui and Zhou, 2013; Rima et al., 2019; Plemper and Lamb, 2021; B. Lee et al., 2021).

7.3 Structural reconstruction of the *Hendra henipavirus sauronoid* assembly and N protein

The cryo-EM data led to the resolution of a multimeric HeV N protein cryo-EM density map, here called the sauronoid owing to its distinct morphology. At an estimated resolution of 3.5 Å, the sauronoid enabled the building and refinement of an atomic N protein model, revealing both conserved and novel features.

Overall, the HeV N protein fold resembles closely and more distantly related paramyxoviral N protein structures, including the recently solved N protein structures of PIV5 (Alayyoubi et al., 2015), MeV (Desfosses et al., 2019), NDV (Song et al., 2019), and SeV (Zhang et al., 2021), with its closest structural and genetic match being NiV N (Yabukarski et al., 2014; Ker et al., 2021). Indeed, in the process of generating a homology model as input for automated model building for HeV N, the crystal structure of NiV N (PDB: 4CO6) (Yabukarski et al., 2014) was found to be the closest match at the time. Since then, a NiV N protein cryo-EM structure was published (PDB: 7NT5) (Ker et al., 2021) while the HeV N structure was undergoing refinement. A repeat of the homology model generation in SWISS-model now finds the best result based on the cryo-EM NiV N structure instead.

The electron density in the RNA binding groove, identified through comparison with homologous structures, enabled unambiguous, *de novo* modelling of a hexameric RNA chain in the asymmetric unit, yet was not sufficiently clear to distinguish specific bases. This is likely due to both the presence of random RNA and symmetrised averaging of that RNA, leading to good overall density for the RNA chain including the location of the bases, but not their identities. Therefore, as is the convention for paramyxoviral N protein, the RNA was modelled as one hexanucleotide per N protomer, adhering to the paramyxoviral 'rule of six' (Kolakofsky et al., 1998). The typical 3' to 5' 'three in, three out' base stacking orientation observed in other paramyxoviral N proteins is also found in the HeV N protein, where bases 1, 2, and 3 face in towards the protein and bases 4, 5, and 6 are solvent exposed. Furthermore, a series of basic (Lys178, Arg192,

Arg193, Arg352) and polar (Thr185, Gln199, Gln200, Tyr258, Gln319, Ser344, Tyr354) residues were identified as putatively interacting, through electrostatic and hydrophobic interactions, with the RNA chain. While the majority of these interactions are with the RNA backbone, a key, conserved interaction occurs between RNA bases 5 and 6 and the amino acid dyad Gln199 and Gln200 which HeV N shares with the N proteins from MeV, NiV, and PIV5 (Alayyoubi et al., 2015; Desfosses et al., 2019; Ker et al., 2021). Additionally, a conserved in-base cavity is contiguous with the RNA binding groove and can accommodate larger purines in the in-base stack comprising nucleotides 1, 2, and 3. Putative interactions with residues Tyr258, Glu260, Glu261, Phe268, Arg272, Lys 321, Ser344, and Met345 may facilitate purine binding where pyrimidines are too distant.

While N-RNA interactions are thought to be sequence non-specific and HeV N model building has progressed in a sequence-agnostic manner, some evidence for MeV suggests RNA encapsidation exhibits a preference for purine-rich sequences and importantly the antigenomic leader sequence (Milles et al., 2016). Similar preferences for purine rich and leader-specific RNA were also observed for the more distantly related RSV and VSV, respectively (Blumberg et al., 1983; Gao et al., 2020). Indeed, the terminal six nucleotides in the HeV genomic and antigenomic leader sequences are identical and purine rich (5'-ACCGAA), and both sequences overall are purine rich as well, with ~70% and ~63% purines for the genomic and antigenomic leaders, respectively (Wang et al., 2000). Additional interactions with these purine bases both in the in-base cavity as well as in the region of bases 5 and 6 may drive this difference in specificity and be important factors in the initial stages of 5' encapsidation of the nascent genome as it emerges from the RdRp complex. As a result, the in-base cavity may present a target for the development of therapeutics that abrogate the critical step of RNA-encapsidation in the paramyxoviral life cycle.

Structural alignment of the final HeV N protein model with both the NiV N protein crystal structure and cryo-EM structure revealed some key differences and similarities. With an overall RMSD of 0.783 Å and 0.756 Å between HeV and NiV crystal structure and cryo-EM structure, respectively. While both structures showed interesting differences to the HeV N structure solved here, differences between HeV N and the NiV N crystal structure were most prominent. These differences largely stem from the open, apo-configuration of the P protein-bound NiV N (P-N⁰) complex in the case of the crystal structure (Yabukarski et al., 2014). Since the cryo-EM NiV N structure is RNA-bound (Ker et al., 2021), it offers a better comparison with the similarly RNA-bound HeV N structure. A slight shift in the orientations of both the Nt- and Ct-arms likely stems from the differences in overall multimeric morphology, since the NiV structure originates from

a helical segment with ~ 13.2 N protomers per helical turn, while the HeV structure is drawn from a 14 N protomer-containing flat ring within the sauronoid particle. Helical spiralling of a single sauronoid tetradecameric ring would generate a NLF helix resembling that of NiV and inevitably incur some minor conformational changes likely engendered by the flexible Nt- and Ct-arms whilst maintain a relatively static globular domain fold.

Furthermore, the NiV N Ct-arm has resolved four additional residues compared to the sauronoid HeV N protein structure, while the HeV N protein structure resolves the extreme N-terminal Met residue where the NiV N protein structure cuts off at Ile4 (Ker et al., 2021). Similarly, the extreme N-terminal residues are often absent in N protein structures from other, related viruses, including PIV5, MeV, and SeV (Alayyoubi et al., 2015; Desfosses et al., 2019; Zhang et al., 2021). The resolution of the extreme N-terminal residues in the HeV N structure enabled the identification of a novel, albeit likely weak, interaction surface between the N_{i+1} and N_{i-1} protomers here called the elbow interface. The presence of an elbow interface in the N protein assemblies of PIV5, MeV, NiV, and SeV is likely obscured due to the absence of the extreme N-terminal residues in these structures. Where the expression and purification pipeline for HeV N included an affinity-tag removal step to generate a native protein sequence, those for the viral N proteins listed above did not. As a result, the presence of remaining, unstructured, N-terminal affinity tags may locally limit cryo-EM resolution and fail to depict the elbow interface. This hypothesis is further supported by the N protein structure of NDV which does resolve the extreme N-terminal residues but these face away from the putative elbow interface and towards the central cavity (Song et al., 2019), likely due to the presence of the N-terminal affinity tag.

Finally, the solvent facing HeV NTD alpha-helix 5 (HeV $\alpha 5$) and its succeeding loop show significant structural and sequence variability among paramyxoviruses and its position and length are highly variable among the N proteins of closely related paramyxoviruses, including MeV, NDV, NiV, PIV5, and SeV (Yabukarski et al., 2014; Alayyoubi et al., 2015; Desfosses et al., 2019; Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). Compared to the cryo-EM NiV N structure, the HeV N protein $\alpha 5$ -loop region contains two amino acid substitutions compared with the NiV N protein; L108V and E137D, both of which are solvent exposed (Ker et al., 2021). In the context of a helical nucleocapsid, this solvent-exposed helix-loop region is likely important for interactions with both the RdRp complex and the M protein at various stages of the viral life cycle (Plempner and Lamb, 2021). Furthermore, the $\alpha 5$ -loop region is also part of the intercalating interface in the sauronoid, as well as the clam-shell assemblies in NDV, NiV, and SeV NLFs (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021).

The recent discovery of a novel HeV strain in Australia, called HeV-var (Annand et al., 2022), as well as the addition of four new henipaviral species, highlights the needs for continued surveillance and research into this genus and HeV specifically. With around 88% sequence identity for the HeV-var N gene sequence compared to the HeV reference strain (NCBI RefSeq: NC_001906), a total of five amino acid substitutions were mapped to the Ncore region. Three of the five substitutions occur in the first half of the Nt-arm, with two facing the solvent of the central hole while a third (S2G) could abrogate the putative elbow interface. Furthermore, 13 amino acid substitutions occur in the intrinsically disordered Ntail region. While these are unlikely to affect the sauronoid assembly, they may be important for N-RdRp and N-M interactions during the replication/transcription and virion assembly stages of the viral life cycle, respectively, and await experimental confirmation.

7.3.1 Contextualising the HeV sauronoid architecture

The HeV sauronoid assembly, a double tetradecameric ring, is a novel architecture undescribed in the current literature on paramyxoviral N proteins. However, related N protein architectures have been observed. Clamshell interfaces have recently been described, in detail, for NDV, NiV, and SeV NLFs (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021). In these cases, recombinant expression of N proteins revealed clamshell interfaces as both discrete particles reminiscent of but different to the sauronoids (Song et al., 2019; Ker et al., 2021) as well as part of a so-called double-headed helix (Song et al., 2019; Zhang et al., 2021). While a similar loop region is involved in these interactions in all three, the specifics of these clamshell architectures differ between the three reports. Since the clamshell joins two anti-parallel NLFs, the overall shape in essence two anti-parallel NLFs stacked bottom-to-bottom and follows the NLF's helical register. In the case of NDV, the N protomers that constitute the clamshell are positioned vertically in line and do not intercalate (Song et al., 2019). Those described for SeV appear to have two forms, called 'closed' and 'hyper-closed', where the level of intercalation is higher for the hyper-closed state (Zhang et al., 2021). Finally, the NiV clamshell exhibits extensive intercalation similar to the sauronoid (Ker et al., 2021). Furthermore, while in each of these cases the two faces of the clamshell follow the helical register of the NLF, an additional small population of NiV clamshells have been observed where the ~ 13.2 N protomer helix is capped with a flat tetradecameric ring (Ker et al., 2021). This ring closely matches one half of the sauronoid assembly. Furthermore, similar double-headed NLFs were observed for HeV, but these failed to progress past 2D classification. Importantly, the presence of double-headed helices was confirmed from virus-extracted nucleocapsids of SeV (Zhang et al., 2021) and the

inability to form double-headed nucleocapsids was detrimental to EGFP reporter gene expression in an NDV minigenome-based replicon assay (Song et al., 2019).

Interestingly, a recent paper describes the presence of single-faced closed-ring N protein assemblies inside RSV virions (Conley et al., 2022). Although of lower resolution due to the use of cryo-ET, these rings bear a striking resemblance to the decameric rings observed for recombinantly expressed RSV N protein (Tawar et al., 2009; Bakker et al., 2013) and indeed one of the two HeV sauronoid tetradecameric rings. Their abundance in these RSV virions hints at a potential physiological relevance. While the identity, and indeed occupancy, of RNA in these virion-enclosed rings was not confirmed, it is known that in RSV-infected cells short, 21 to 25 nucleotide-long RNAs generated from the genomic leader and trailer sequences are among the most abundant viral RNAs present (Noton et al., 2012; Tremaglio et al., 2013; Braun et al., 2017). Furthermore, some of these RNAs are resistant to nuclease treatment, indicating they may be encapsidated (Tremaglio et al., 2013).

The occurrence of sauronoids and indeed the related clamshells, in recombinantly expressed N protein samples could be an artefactual result of the method of sample generation. However, the presence of clamshells in virion-extracted nucleocapsids (Zhang et al., 2021) and of N protein ring assemblies inside virions (Conley et al., 2022) suggests otherwise. A picture emerges where the sauronoid assembly may be indicative of one or multiple of a number of viral processes.

Firstly, it is hypothesised that the occurrence of clamshell interfaces both protects the viral genome from nuclease and protease attack and drives the inclusion of multiple copies of the viral genome into nascent virions (Song et al., 2019; Zhang et al., 2021). Indeed, these clamshells appear to join only the 5' ends, which are the first to emerge from the RdRp during genome replication and may be the most vulnerable to degradation. Furthermore, polyploidy is increasingly recognised in a wide range of members of the order *Mononegavirales* (Rager et al., 2002; Loney et al., 2009; Beniac et al., 2012; Goff et al., 2012) and clamshell interfaces may explain this occurrence at least partially. The sauronoid resembles these clamshells, in particular the flat ring of the semi-spiral clamshell interface found in NiV N particles, and the sauronoid may therefore be a by-product of the viral need for genomic 5' nucleocapsid capping.

Secondly, the high abundance of compatible leader and trailer RNAs in virus-infected cells could result in the activation of the innate immune response. The encapsidation of these sequences is hinted at for RSV (Tremaglio et al., 2013), which could shield them from detection by the innate immune system. Interestingly, the 56 nt-long HeV leader sequence, which is

transcribed and separated from the mRNA transcripts during the RdRp transcription mode, falls within the occupancy range for a single tetradecameric ring (84 nt). This purine-rich sequence is also likely preferentially encapsidated over RNA chains with an even purine-pyrimidine ratio (Blumberg et al., 1983; Milles et al., 2016; Gao et al., 2020). The HeV trailer sequence is shorter than the leader and similarly rich in purine bases. An antagonistic function in stress granule formation and cellular apoptosis and has been reported for the (released) trailer sequence in SeV and RSV (Iseni et al., 2002; Hanley et al., 2010) and a similar function could exist for the HeV trailer. Whether putative encapsidation of the trailer aids or hinders this function is unclear, but encapsidation of both the leader and the trailer could permit inclusion into progeny virions and subsequent release in newly infected cells to perform their putative functions.

Thirdly, the N-RNA rings such as the HeV sauronoid, could play a role in nucleating liquid-liquid phase separation for biomolecular condensates and aid in the formation of viral transcription/replication centres (Rincheval et al., 2017; Roden and Gladfelter, 2021). Evidence has been found for this process in RSV and has recently been investigated in detail for MuV where a combination of cryo-ET, fluorescence microscopy, and mass spectrometry reveal a stress-induced modulation of phase-separated N-P complexes (X. Zhang et al., 2022). A similar role may exist for the sauronoid during phase separation in HeV infected cells.

Finally, the sauronoid, and indeed N protein rings observed elsewhere, may reflect the need for short-term, reversible storage of N protomers in the absence of abundant P protein chaperones. Indeed, the P protein engages in extensive interactions with N protomers to maintain them in an open, apo-state (N^0). Since, due to inefficient re-initiation of the RdRp, viral mRNAs decrease in abundance the further they are from the genomic 3'-end, the abundance of N protein is likely much higher than that of P protein and there is the possibility for excessive N protein concentrations in the cell. The formation of sauronoids, either RNA-bound or in the apo-state, could therefore provide a short-term, metastable multimeric N protein storage where P protein shuttles between them and the RdRp complex during active encapsidation of nascent genomic or antigenomic RNA.

Taken together, sauronoids may represent a convergent architecture that reflects the need for 5' capping of genomic nucleocapsids, the need to encapsidate leader sequences, and the need for short-term metastable N protein storage.

7.3.2 Future directions for structural work involving HeV nucleoprotein

The high-resolution reconstruction of the HeV N protein was facilitated by the impartation of D14 symmetry on the sauronoid particle stack. However, symmetrised reconstructions do not allow for the detection of structural or compositional heterogeneity. The intercalating interface in the sauronoid assembly was poorly resolved, which abrogated confident model building and forced the stubbing of residue side chains for Met114, Glu115, Arg116, Arg117, and Asp119. This loop was identified as a putative driver of clam-shell interactions in NDV, NiV, and SeV (Song et al., 2019; Ker et al., 2021; Zhang et al., 2021) and mutation to alanines prevented the formation of double-headed NDV nucleocapsids (Song et al., 2019). Similarly, RNA occupancy in the sauronoids may be non-uniform and is obscured by symmetrisation. The resolution of non-symmetrised sauronoid particles would likely elucidate both the uniformity of RNA occupancy as well as regions of structural flexibility but necessitates much larger particle stacks and thus further cryo-EM data collection to reach a similar overall resolution.

The HeV NLF reconstruction presented in Figure 5.23 resembles those of other paramyxoviruses (for a review see (Bloyet, 2021)). However, the generation of a high-resolution reconstruction of the helical HeV NLF has eluded conclusion in this work, although not for a lack of perseverance. Instead, obtaining a high-resolution HeV N helix may require more rigid NLFs, facilitated through altering the protein sample buffer conditions or removal of the Ntail region which effects much of the nucleocapsid's flexibility. Indeed, both salt and pH conditions affect paramyxoviral NLF flexibility (Heggeness et al., 1980). Furthermore, the intrinsically disordered Ntail imparts a high degree of flexibility, since after its proteolytic removal from MeV N those nucleocapsids become more rigid (Schoehn et al., 2004; Bhella et al., 2004). Interestingly, extended incubation above cryogenic temperatures (4-37°C) also had a rigidifying effect on SeV and MuV nucleocapsids (Cox et al., 2014; Severin et al., 2016; Zhang et al., 2021). Similar proteolytic 'pruning' of HeV N with trypsin resulted in extensive protein degradation but HeV N was similarly stable for up to 8 weeks at 4 °C (Supplementary Figure S7.1). Alternatively, a new expression construct with a deletion of the Ntail (residues 399 to 532) may result in rigid HeV nucleocapsids that are more amenable to helical reconstruction and require minimal optimisation of the expression and purification protocol. Indeed, the expression of a similarly truncated HeV N protein (Ncore) was described and NLFs of Ncore showed tighter coiling compared to those consisting of full length N (Pearce et al., 2015).

7.4 The use of a replicon system as a tool for investigating *Hendra henipavirus* molecular biology

A homotypic HeV minigenome-based replicon system does not exist in the published literature, but previous work with a NiV minigenome-based replicon systems showed heterotypic support for a HeV minigenome by NiV support genes (Halpin et al., 2004). The need for a homotypic HeV system arose since it would enable the molecular investigation of HeV N, P, and L proteins in the absence of BSL-4 facilities.

The primary HeV replicon design followed the strategy employed for the NiV system (Halpin et al., 2004), with some differences in the minigenome to enable fluorescence-based reporter gene analysis. Overall, this system performed poorly, with fluorescence levels close to that of the negative controls. This low efficiency could be due to the sensitivity of paramyxoviral systems to the relative ratios of the minigenome and support plasmids. Indeed, a T7-free NiV minigenome-based replicon system showed high sensitivity for N, P, and L support stoichiometry (Freiberg et al., 2008). Limited plasmid ratio optimisation failed to provide a strong direction, with a positive effect on overall performance by decreasing the relative concentration of the N support. This could be explained by the effect N protein levels have on modulating the viral RdRp between transcription and replication modes (Plempner and Lamb, 2021). The overabundance of N protein in the transfected cells could favour the generation of anti-minigenome over transcription of the EGFP ORF, resulting in reduced fluorescence signal.

In a similar pursuit, the support genes were amended to include each gene's non-translated region to resemble the native transcription and translation control of these genes more closely. In addition, the P gene-encoded C protein is known to have a negative effect on L polymerase activity (Plempner and Lamb, 2021) and was silenced to generate a PdC gene. These alterations resulted in a minor improvement in replicon performance, with an EGFP signal around 3-fold higher than the negative control. Unfortunately, this difference is insufficient to consider the HeV replicon system as a robust platform for experimentation with N protein mutations or therapeutic candidates.

7.4.1 Future directions for work with the HeV minigenome-based replicon system

Since the stoichiometry of viral genes is essential for efficient minigenome replication and transcription, further optimisation of the support plasmid ratios should be undertaken. A large-scale optimisation has been done for the T7-independent NiV system (Freiberg et al., 2008)

and a similar approach may shed more light on the optimal ratios for the HeV system, which should be undertaken with the NTR-inclusive support plasmids. Particular attention should be paid to both the ratio and time of addition of the N support, since N protein levels are known to modulate RdRp activity.

A large number of primary and immortalised cell lines are known to permit HeV replication (B. Lee et al., 2021) and therefore a wider range of cell lines should be tested for the HeV replicon system. The reports on NiV minigenome-based replicon systems used CV-1 cells (Halpin et al., 2004) or HeLa, 293T, and BHK-T7 cells (Freiberg et al., 2008). Thus, the suitability of these or similar cell lines should be investigated for the HeV replicon system. The T7 polymerase-dependent nature of the HeV replicon system presented here necessitates – in cells that do not stably express T7 polymerase – the co-transfection of a T7 polymerase-expressing plasmid, co-infected with a T7 polymerase-expressing recombinant virus (Halpin et al., 2004), or transfected with previously *in vitro* generated RNA transcripts of the minigenome and support genes. The transfection of T7 polymerase-transcribed RNA was attempted here with limited results, but likely suffered from poor RNA quality and cell viability and re-examination of this approach may provide rapid, positive results.

Appendix I

Supplementary Figures

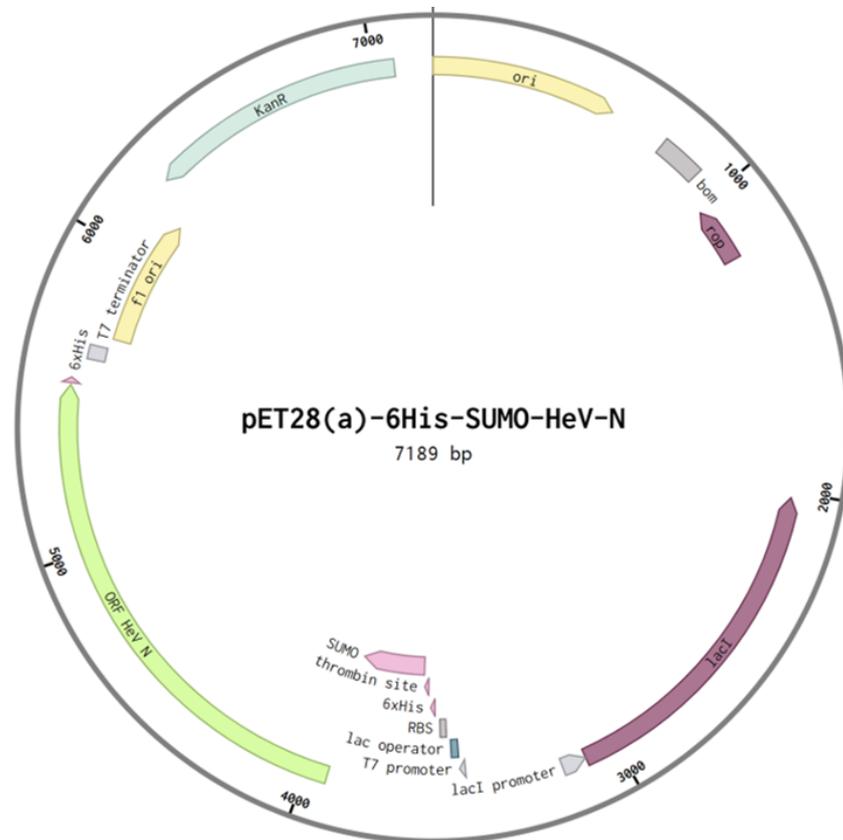


Figure S3.1. Plasmid map of pET28(a)-6His-SUMO-HeV-N.

Plasmid features are annotated, and annotations provide the ORF directions. The vertical black line indicates the index nucleotide. Starting at the index nucleotide going clockwise, in order of appearance, the annotations indicate the following. **ori**, the origin of replication. **bom**, the basis of mobility region. **rop**, the repression of primer region. **lacI**, gene of the lac repressor LacI. **lacI promoter**, promoter sequence for the lacI gene. **T7 promoter**, T7 RNA polymerase promoter sequence. **lac operator**, region where the lac repressor binds in absence of allolactose. **RBS**, ribosomal binding site. **6xHis**, hexahistidine repeat. **thrombin site**, thrombin cleavage site. **SUMO**, small ubiquitin-like modifier (SMT3). **ORF HeV N**, open reading frame encoding the codon-optimised *Hendra henipavirus* nucleoprotein. **T7 terminator**, T7 RNA polymerase terminator sequence. **f1 ori**, phage-derived origin of replication. **KanR**, kanamycin resistance gene. Map was generated in Benchling (Benchling, 2022).

A

ProtParam

User-provided sequence:

```

10      20      30      40      50      60
MGSSHHHHHHH SSSLVPRGSH MSDESEVQEA KPEVKEVVKP ETRINLKVSD GSSEIFFKIK
70      80      90     100     110     120
KTTPLRLRIIE AFARQRQKEM DSLRFLYDSI RIQADQTPED LDMEDNDIIE AHRDQIGGSH
130     140     150     160     170     180
SDIFDEAASF RSYQSKLGRD GRASAAATL TTKIRIFVPA TNSPELWEL TLFALDVIK
190     200     210     220     230     240
PSAAESMKIG AAFLLISMV ERPGALIRSL LNDPDIKAVI IDVGSMLNGI FVMERRGDKA
250     260     270     280     290     300
QEEMGLMRI LKTARESSKG KTFPFVDSRAY GLRITDMSTL VSAVITIEAQ IWILLIAKAVI
310     320     330     340     350     360
APDTAESET RRWAKYVQK RVNPFALTO QWLTEMRNLL SQSLVSRKFM VEILMEVKKG
370     380     390     400     410     420
GSAKRAVEI ISDIGNYVEE TGMAGFFATI RFGLETRIFPA LALNEFQSDL NTKIGMLLLI
430     440     450     460     470     480
REIGPRAPYM VLLLESIQTK FAPGGYPLLW SFAMGVATTI DRSMGALNIN RGVLEPMYFR
490     500     510     520     530     540
LGQKSARHHA GGIDONWANK LGLNSDQVAE LAAAVQETSU GRQDNMQRAR EAKFAAGGVL
550     560     570     580     590     600
VVGGEQDIDE EEEPIEHSGR QSVTFKREMS MSSLADSVPS SSVSTSGGTR LTNLSLLNLR
610     620     630     640     650
RLAAKAIKES TAQSSSERNP PNNRPQADSG RKDDQEPKPA QNDLDFVRAD V

```

Number of amino acids: 651

Molecular weight: 71969.54

Theoretical pI: 5.96

Amino acid composition: [CSV format](#)

Ala (A)	60	9.2%
Arg (R)	46	7.1%
Asn (N)	24	3.7%
Asp (D)	35	5.4%
Cys (C)	0	0.0%
Gln (Q)	27	4.1%
Glu (E)	52	8.0%
Gly (G)	46	7.1%
His (H)	12	1.8%
Ile (I)	40	6.1%
Leu (L)	54	8.3%
Lys (K)	33	5.1%
Met (M)	26	4.0%
Phe (F)	23	3.5%
Pro (P)	28	4.3%
Ser (S)	61	9.4%
Thr (T)	34	5.2%
Trp (W)	5	0.8%
Tyr (Y)	12	1.8%
Val (V)	33	5.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 87
 Total number of positively charged residues (Arg + Lys): 79

Atomic composition:

Carbon	C	3137
Hydrogen	H	5040
Nitrogen	N	902
Oxygen	O	984
Sulfur	S	26

Formula: C₃₁₃₇H₅₀₄₀N₉₀₂O₉₈₄S₂₆
 Total number of atoms: 10089

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient	45380
Abs 0.1% (=1 g/l)	0.631

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
 >20 hours (yeast, in vivo).
 >10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 54.32
 This classifies the protein as unstable.

Aliphatic index: 80.23

Grand average of hydropathicity (GRAVY): -0.411

B

ProtParam

User-provided sequence:

```

10      20      30      40      50      60
MSDIFDEAAS FRSYQSKLGR DGRASAAAT LTKIRIFVFP ATNSPELRWE LTLFALDVIK
70      80      90     100     110     120
SPSAAESMKI GAAPTLLISMV SERPGALIRSL LNDPDIKAVI IDVGSMLNGI IPVMERRGDK
130     140     150     160     170     180
AQEEMGLMRI ILKTARESSK GRTPFVDSRAY YGLRITDMST LVSAVITIEA QIWILLIAKAV
190     200     210     220     230     240
TAPDTAEESR TRRWAKYVQK KRWNPFALTO QWLTEMRNLL LSQSLVSRKF MVEILMEVKK
250     260     270     280     290     300
GGSAKRAVEI IISDIGNYVEE ETGMAGFFATI IRFGLLETRYP ALALNEFQSD LNTIKIGMLLL
310     320     330     340     350     360
YREIGPRAPY NVLLESIQTK KFAPGGYPLLW SFAMGVATTI IDRSMGALNI NRGVLEPMYF
370     380     390     400     410     420
RLGQKSARHH AGGIDQNMAN KGLNSDQVA ELAAAVQETS VGRQDNMQR REAKFAAGGV
430     440     450     460     470     480
LVGGEQDID EEEPIEHSGR RQSVTFKREM SMSSLADSVP SSVSTSGGTR RLNSLLNLR
490     500     510     520     530
SRLAAKAIKE STAQSSSERNP PNNRPQADS GRKDDQEPK QNDLDFVRA DV

```

Number of amino acids: 532

Molecular weight: 58475.47

Theoretical pI: 5.95

Amino acid composition: [CSV format](#)

Ala (A)	55	10.3%
Arg (R)	39	7.3%
Asn (N)	21	3.9%
Asp (D)	26	4.9%
Cys (C)	0	0.0%
Gln (Q)	22	4.1%
Glu (E)	40	7.5%
Gly (G)	38	7.1%
His (H)	3	0.6%
Ile (I)	32	6.0%
Leu (L)	47	8.8%
Lys (K)	24	4.5%
Met (M)	21	3.9%
Phe (F)	19	3.6%
Pro (P)	22	4.1%
Ser (S)	49	9.2%
Thr (T)	30	5.6%
Trp (W)	5	0.9%
Tyr (Y)	11	2.1%
Val (V)	28	5.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 66
 Total number of positively charged residues (Arg + Lys): 63

Atomic composition:

Carbon	C	2556
Hydrogen	H	4119
Nitrogen	N	727
Oxygen	O	798
Sulfur	S	21

Formula: C₂₅₅₆H₄₁₁₉N₇₂₇O₇₉₈S₂₁
 Total number of atoms: 8221

Extinction coefficients:

Extinction coefficients are in units of M⁻¹ cm⁻¹, at 280 nm measured in water.

Ext. coefficient	43890
Abs 0.1% (=1 g/l)	0.751

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
 >20 hours (yeast, in vivo).
 >10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 54.83
 This classifies the protein as unstable.

Aliphatic index: 83.52

Grand average of hydropathicity (GRAVY): -0.296

Figure S3.2. HeV N protein parameters.

HeV N protein parameters were calculated in ExPASy ProtParam for the hexahistidine SUMO HeV N fusion protein (A) and native HeV N protein (B). Hexahistidine (green) and SUMO (yellow) are indicated.

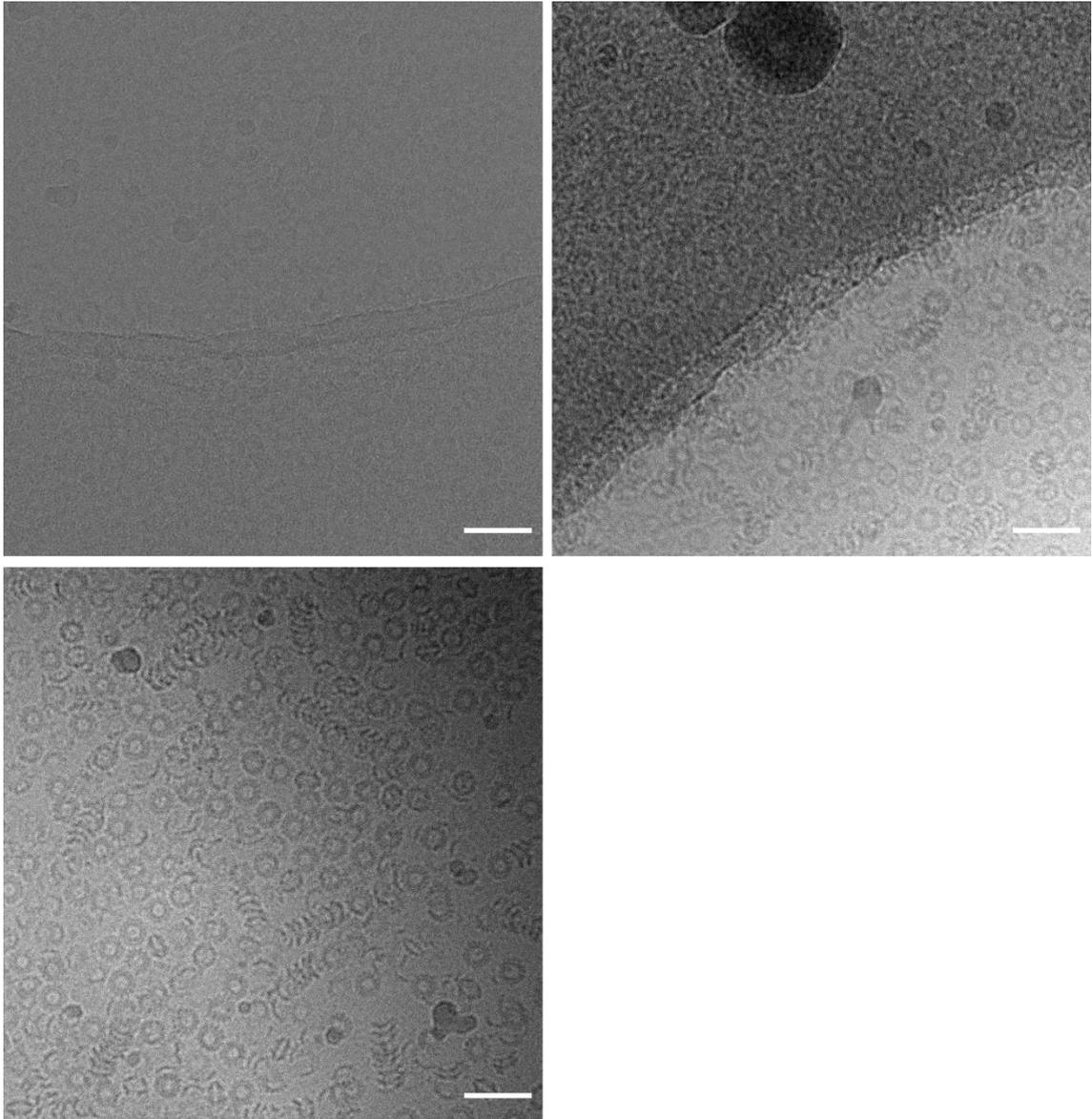


Figure S4.1. HeV N cryo-grid screening.

Pool HeV N samples from SEC were applied to glow-discharged Quantifoil Cu 400 R2/2 holey grids and plunge-frozen using an FEI Vitrobot IV. Prepared cryo-grids were screened on FEI Titan Krios showing highly dense particle distributions. Scale bars are 50 nm.

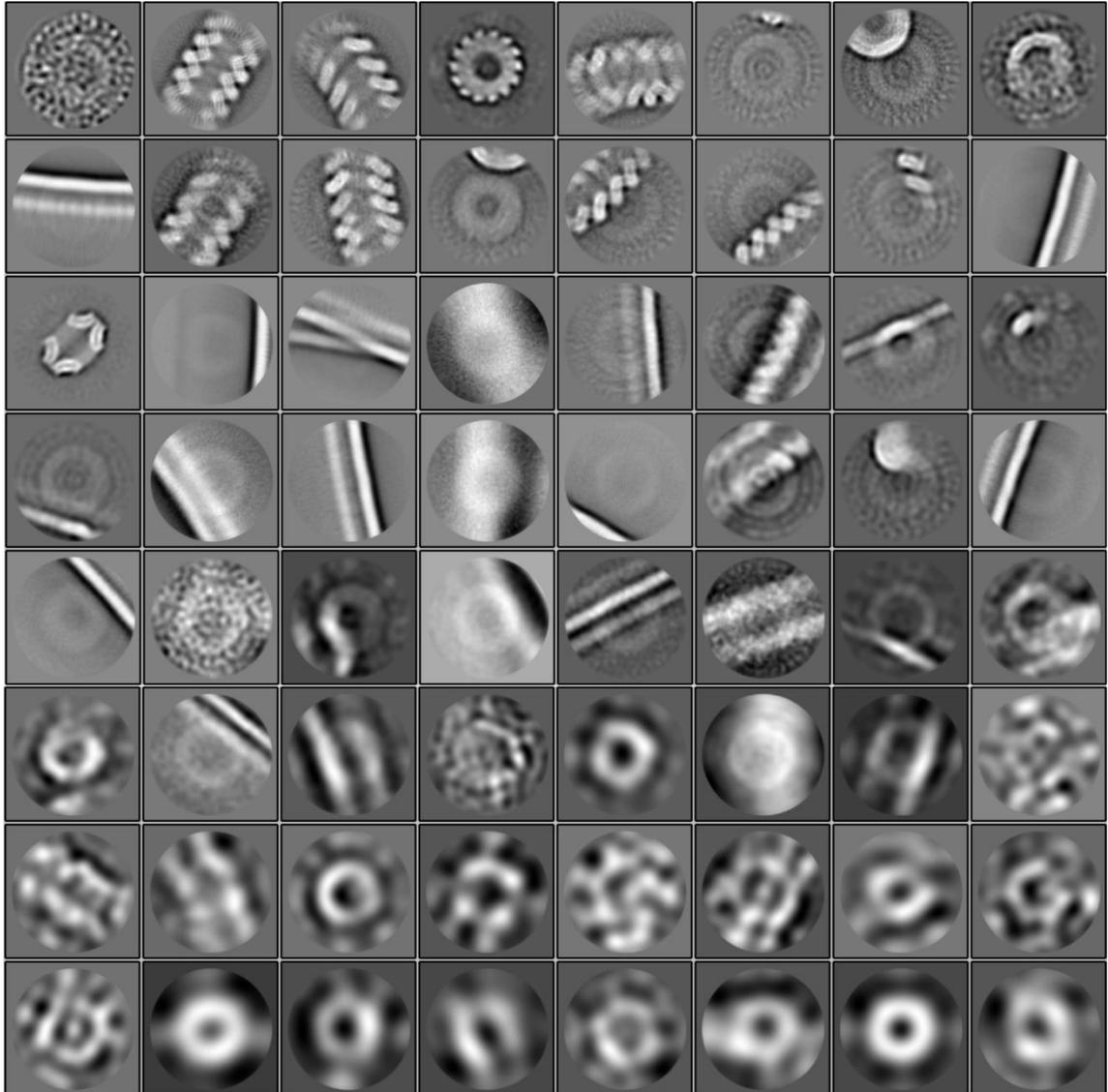


Figure S5.1. First round 2D classification for sauronoids from reference-based automated particle picking.

A set of 875,262 electron density features were automatically selected from 3548 micrographs following reference-based automated particle picking and 2D averaged into 64 classes. The displayed classes were sorted on 'in class distribution' where the most populated classes appear top left and class population decreases from left to right and top to bottom. Automated picking selected features that do not represent true sauronoid particles, including carbon edges, carbon film, nucleocapsid-like filaments, and unidentified features. Box boundaries are 383.4 Å (360 pixels).

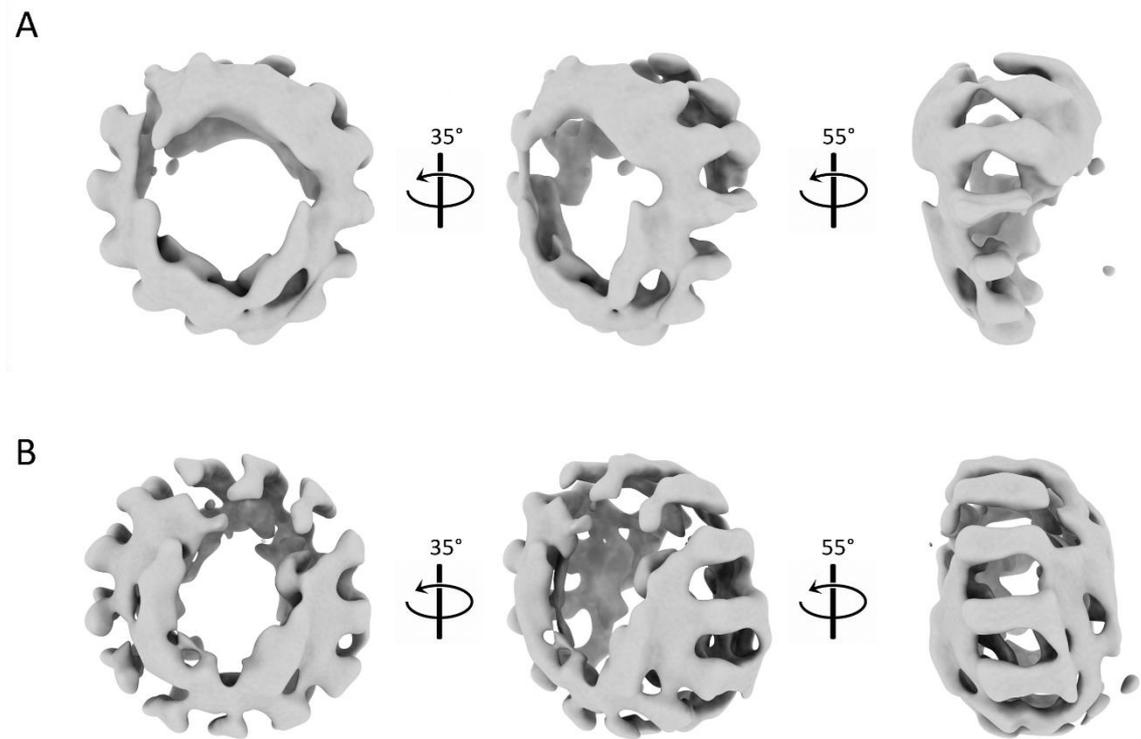


Figure S5.2. Sauronoid 3D reference map and 3D classification.

A) An initial reference map was generated *de novo* from the curated 2D classified particles by stochastic gradient descent (SGD). **B)** The reference map was used for 3D classification without symmetry impartation. The most favourable 3D class, depicted here, could subsequently be used for 3D refinement.

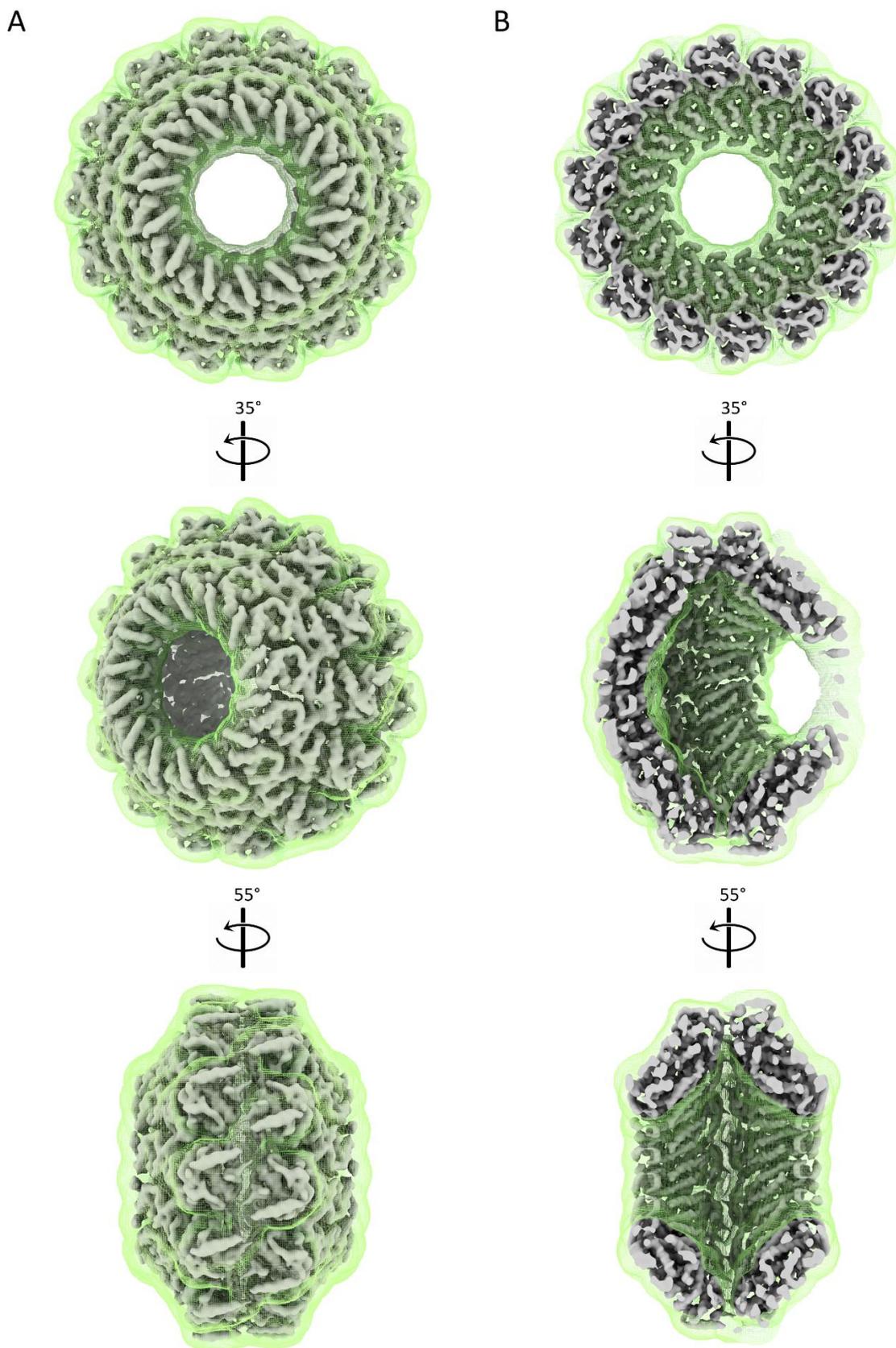


Figure S5.3. Masking of the 3D refine sauronoid map.

The initial refinement of the sauronoid map was post-processed using a solvent mask. The solvent mask covers the whole particle but importantly does not fill the central cavity. The initial refinement map, before post-processing, is depicted surrounded by the solvent mask (green mesh) as a complete map (A) and sliced in half down the axis of rotation (y-axis) (B) showing the mask extension beyond the map.

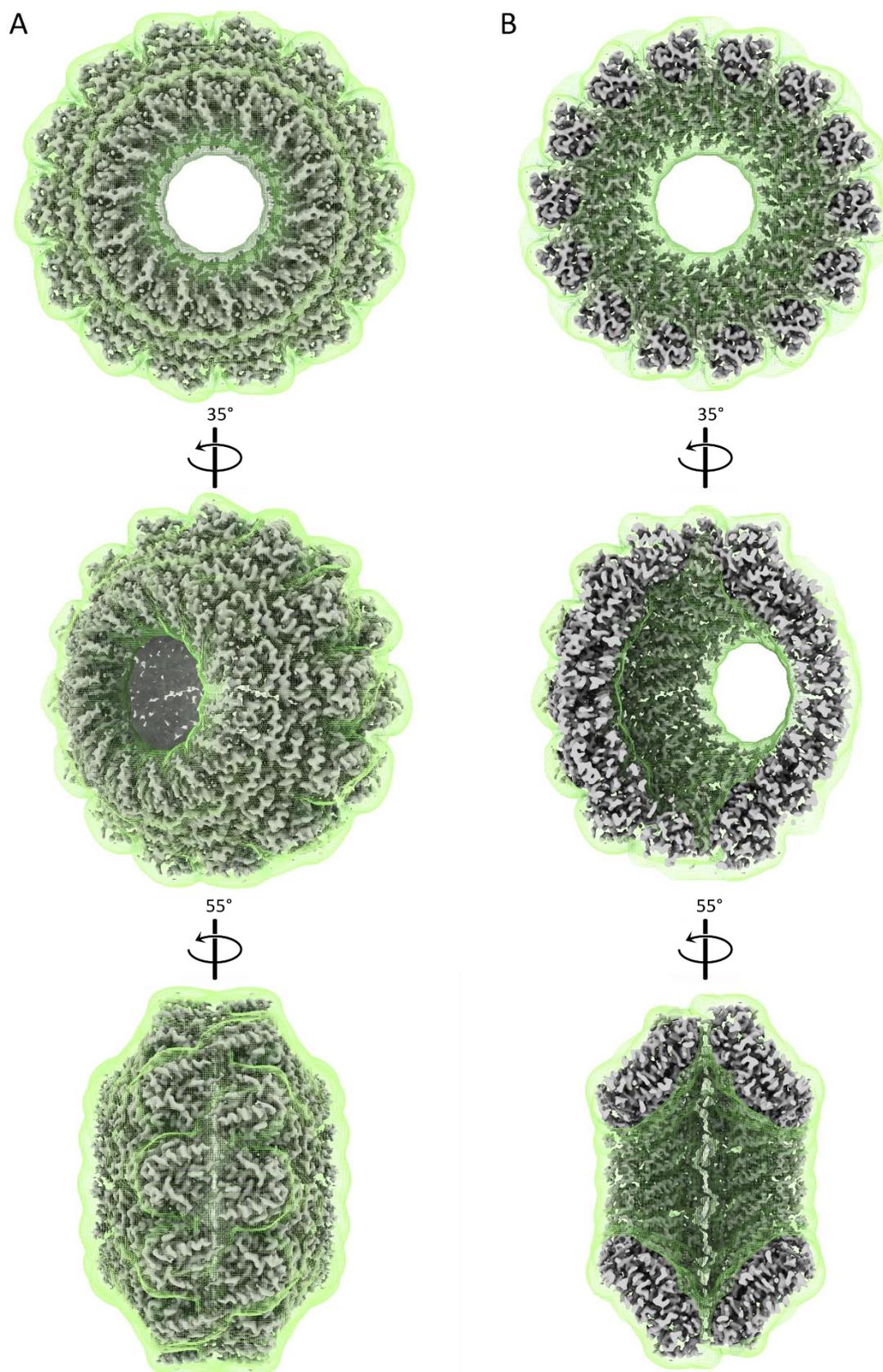


Figure S5.4. Masking of the post-processed sauronoid map.

The initial refinement of the sauronoid map was post-processed using a solvent mask. The solvent mask covers the whole particle but importantly does not fill the central cavity. The refined map, after post-processing, is depicted surrounded by the solvent mask (green mesh) as a complete map (A) and sliced in half down the axis of rotation (y-axis) (B) showing the mask extension beyond the map.

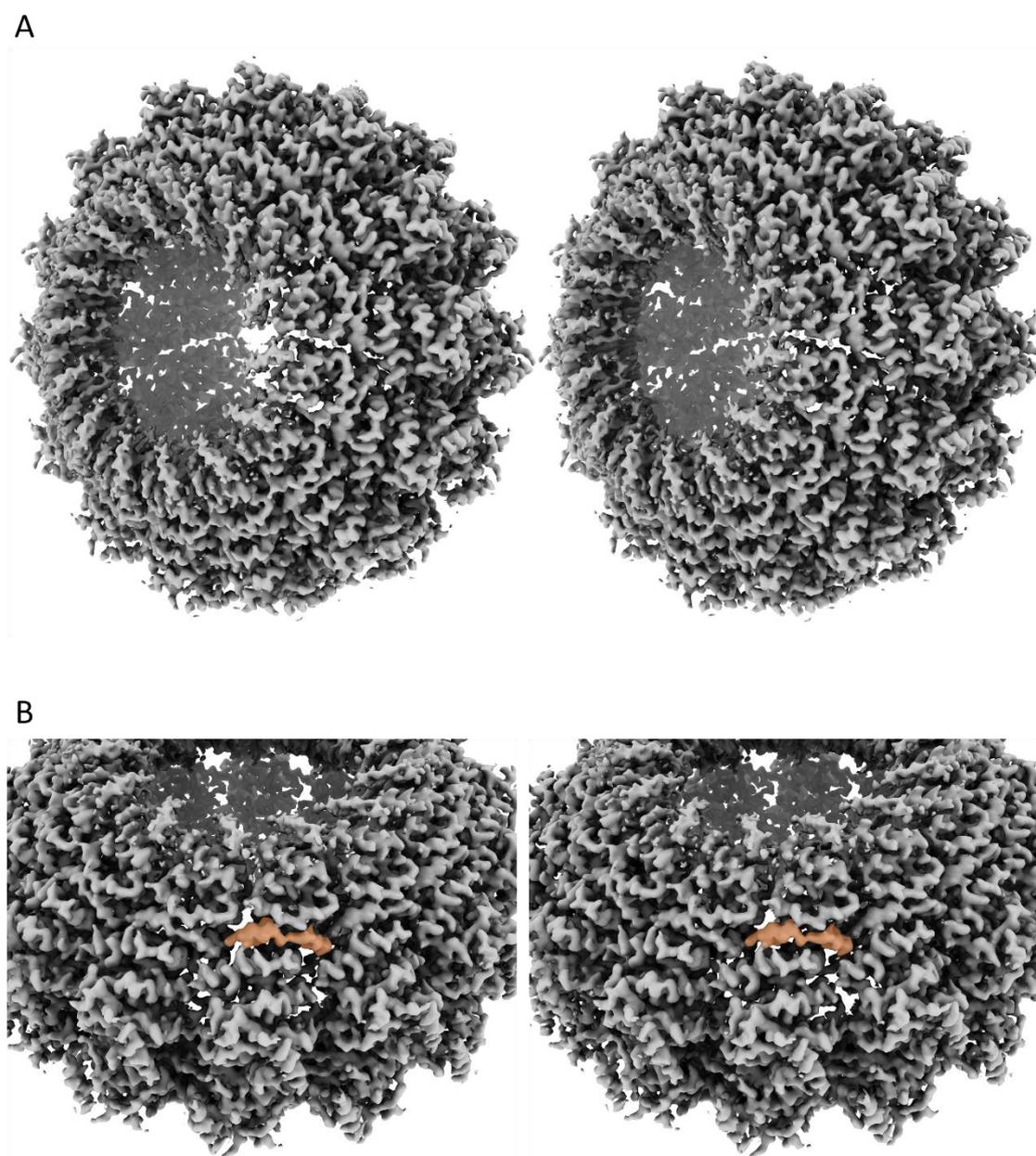


Figure S5.5. Stereographs of the final sauronoid map.

Stereographic depictions of the final, post-processed sauronoid map. Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis. **A)** The final 3D-refined sauronoid map following post-processing reached 3.5 Å resolution. **B)** The RNA-binding groove is occupied by electron density, with the electron density sitting within a single N protomer indicated in orange.

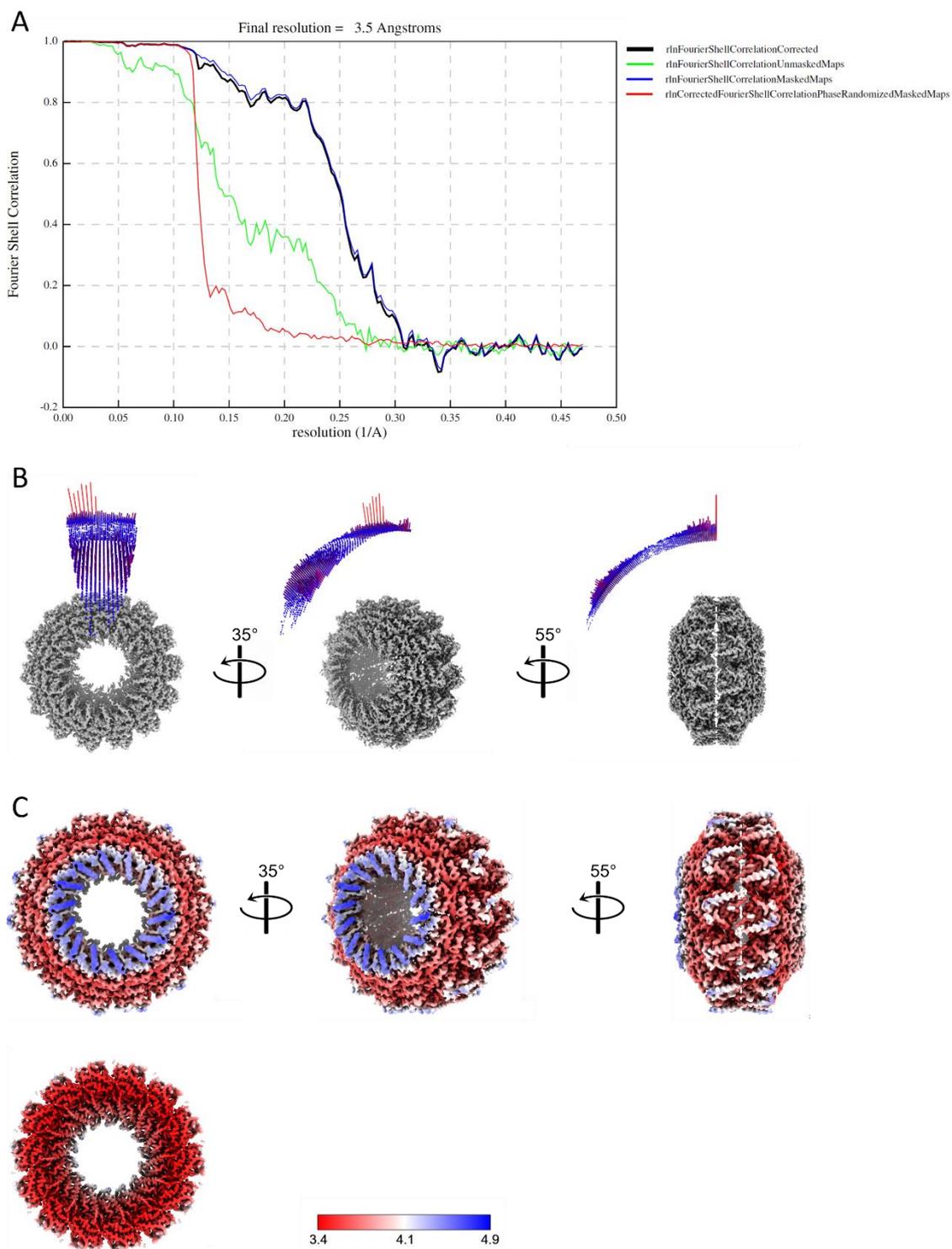


Figure S5.6. Symmetry refinement and post-processing for the sauronoid map.

A) Fourier shell correlation plot from post-processing of the initial sauronoid map refinement. **B)** Angular distribution plot for the post-processed sauronoid map. Angular distributions appear in a wedge shape due to the imparted D14 symmetry. **C)** Local resolution representation of the sauronoid map. Colours range from 3.4 Å to 4.9 Å. The top three images are of the complete sauronoid map. The bottom image represents the central hole view slices across the z-plane to reveal the internal detail.

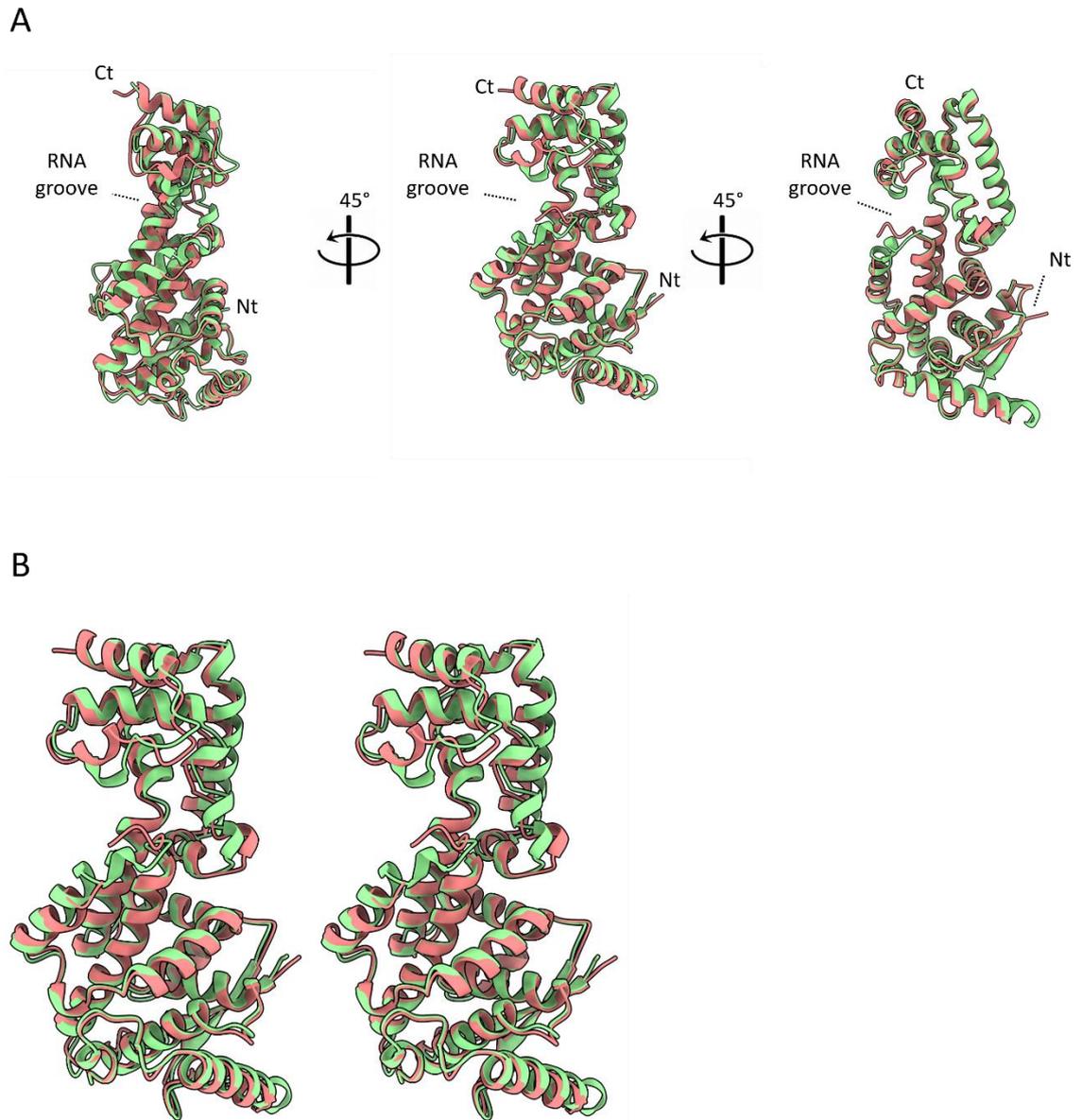


Figure S5.7. Homology model for HeV nucleoprotein.

A HeV N protein homology model (green) was generated with SWISS-MODEL based on the crystal structure of a truncated NiV N protein (red) (PDB 4CO6), which lacks the Nt and Ct flexible arms. **A)** N- and C- termini are indicated, as well as the RNA binding groove. The RNA binding groove does not have RNA modelled within. **B)** Stereographic depictions of the HeV N protein homology model (green) and the NiV N protein crystal structure (red) it was based on. Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis.

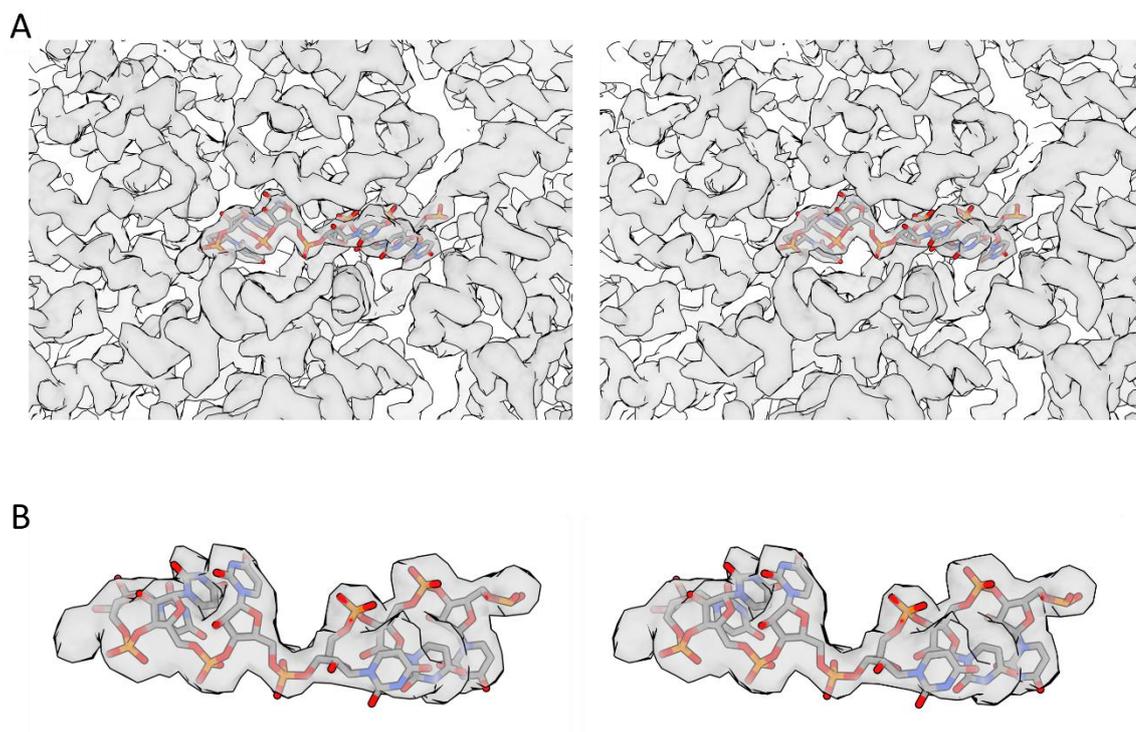


Figure S5.8. Stereograph of *de novo* RNA structure in the RNA-binding groove.

Stereographic representation of the cryo-EM density surrounding the N protein RNA-binding groove (**A**) and the isolated density occupying the RNA-binding groove (**B**) depicting the *de novo* built hexa-uridine chain (sticks). Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis.

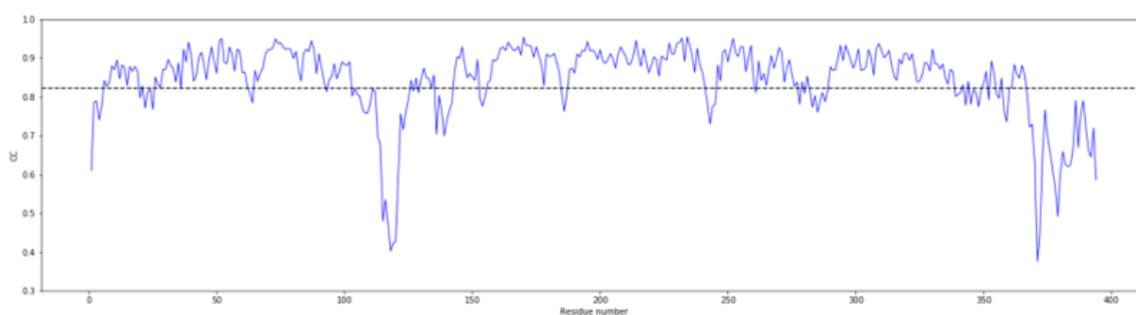
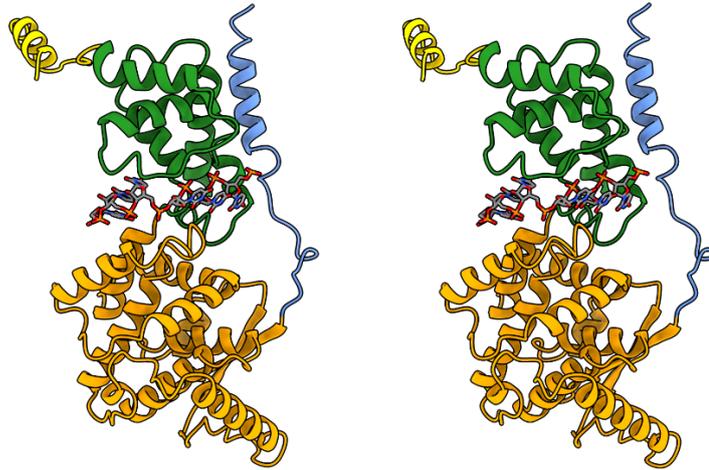


Figure S5.9. HeV nucleoprotein correlation coefficients.

Correlation graph for a single N protomer chain, with correlation values plotted over the residue numbers. The average correlation coefficient (0.84) is indicated by the dashed line. The poorly resolved regions Met114-Asp119 and Leu384-Leu392 also exhibit lower correlation coefficients.

A



B

```

1 MSDIFDEAASFRSYQSKLGRDGRASAATATLTTKIRIFV PATNSPELRWELTLFALDVIR
61 SPSAAESMKIGAAFTLISMYSERPGALIRSL LNDPDI EAVIIDVGSMLNGIPVMERRGDK
121 AQEEMEGLMRILKTARESSK GKTPFVDSRAYGLRITDMSTLVS AVITIEAQIWI LIAKAV
181 TAPDTAE ESETRRWAKYVQQRVNPFFALTQQWLTEM RNLLSQSL SVRKFMVEILMEVKK
241 GGSAKGRAVEIISDIGNYVEETGMAGFFATIRFGL ETRYPALALNEFQSDLNTIKGLMLL
301 YREIGPRAPYMLLEE SIQTKFAPGGY PLLWSFAMGVATTI DRSMGALNINRGYLEPMYF
361 RLGQKSARHHAGGIDQNMNKLGLNSDQVAELAA

```

C

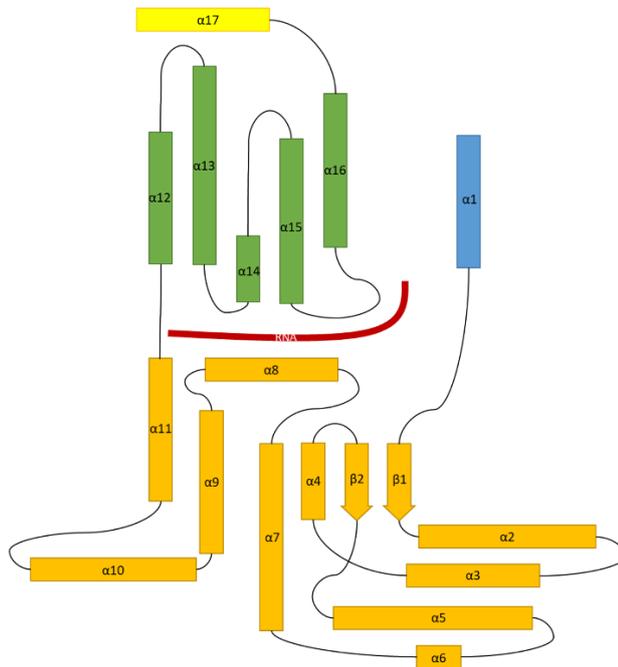


Figure S5.10. Stereograph and organisation of the HeV N protomer.

A) Stereographic ribbon representation of the HeV N protomer structure including the bound RNA hexanucleotide (sticks). The N protein domains are coloured; Nt-arm (blue), globular NTD (orange), globular CTD (green), Ct-arm (gold). Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis. B) The built HeV N protein sequence (B) and topographic representation (C) showing alpha-helical (yellow) and beta-sheet (blue) secondary structure elements. C) Secondary structure elements are scaled appropriately, the peptide loops (black lines) connecting these elements are not to scale. The RNA binding groove is depicted in red.

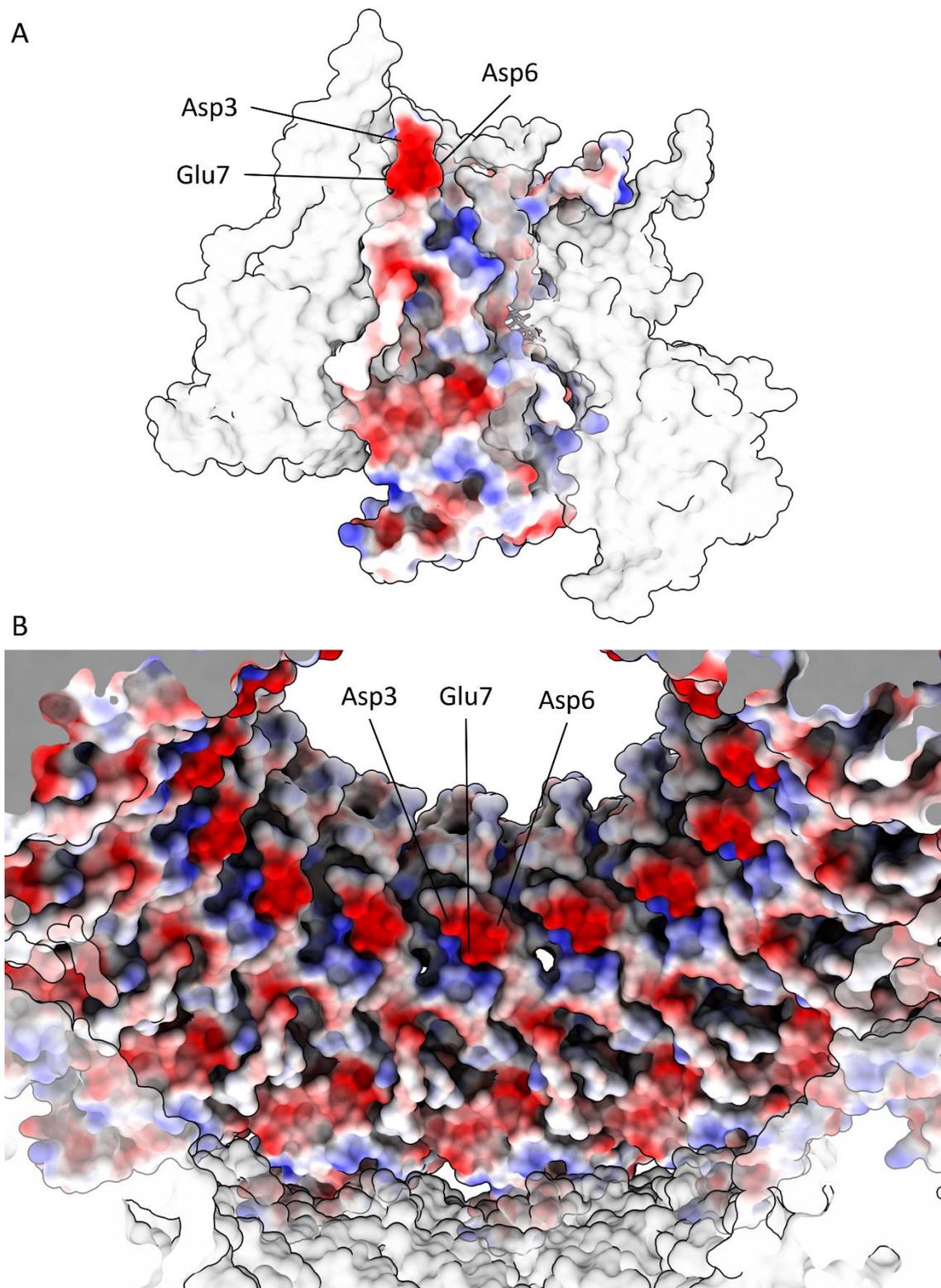


Figure S5.11. The HeV N Nt-arm contains a solvent-facing, negatively charged patch.

The presence of a strongly negatively charged patch (Asp3, Asp6, Glu7) is indicated within the context of a HeV N protomer (**A**) or a tetradecameric ring (**B**). The protomer or ring of interest are coloured according to electrostatic surface potentials for negative (red) and positive (blue) residues or uncoloured (white) for neutral residues. HeV N protomers or tetradecameric ring adjacent to the protomer or ring of interest, respectively, are transparent.

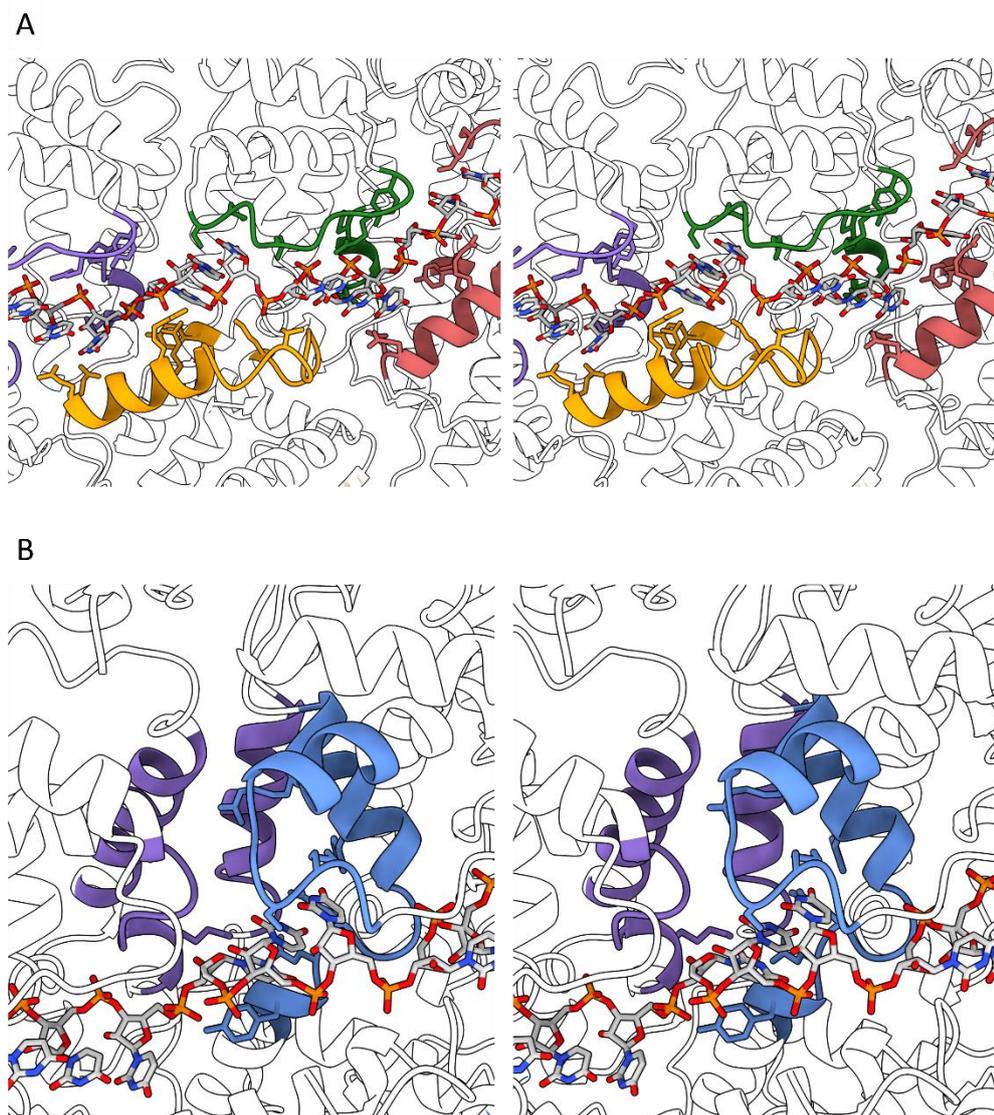


Figure S5.12. The HeV N protein RNA binding groove and in-base cavity.

Stereographic ribbon representations of the RNA binding groove (**A**) and in-base cavity (**B**). The RNA chain is visualised as sticks and follows heteroatom colouring. Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis. **A**) An N protein trimer centred on the N_i protomer RNA binding groove. Helices and loops containing RNA interacting residues are coloured for the N_i (NTD; gold, CTD; green), the N_{i+1} (purple), and the N_{i-1} (coral). RNA interacting residue side chains are visualised as sticks. **B**) An N protein dimer centred on the in-base cavity. Helices and loops that form the in-base cavity are coloured for the N_i (blue) and the N_{i+1} (purple). Putative RNA-base interacting residue side chains are visualised as sticks.

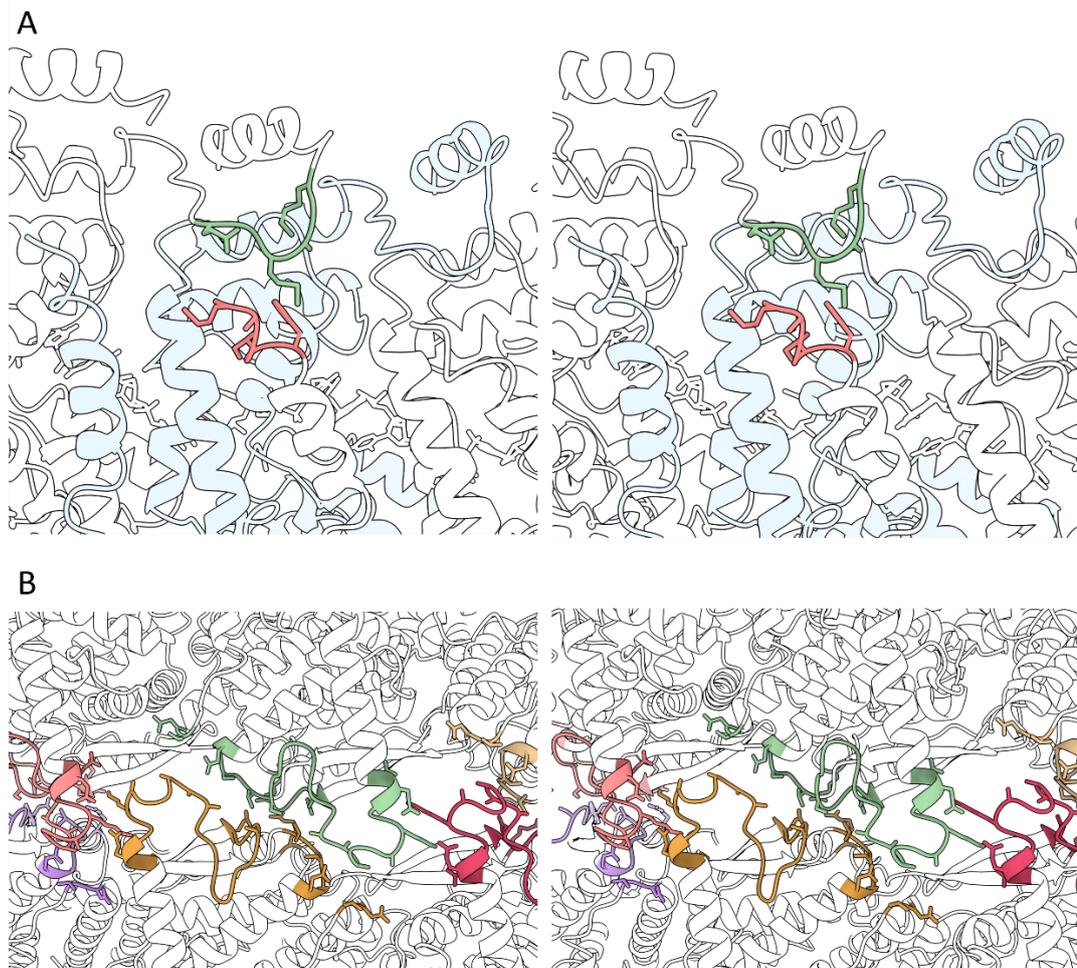


Figure S5.13. The HeV elbow interface and intercalation zone.

Stereographic ribbon representations of the elbow interface and interaction zone. Stereographs were generated by taking two images with a relative rotation of 5° about the y -axis. **A)** A ribbon visualisation of the N protein trimer showing the N_i (pale blue), N_{i+1} (coral), and N_{i-1} (green) protomers, centred on the elbow interface between the N_{i+1} Nt-arm and the N_{i-1} Ct-arm. Protomers N_{i+1} and N_{i-1} are coloured white and the N_i is coloured pale blue. Putative interacting residues are coloured coral and green and presented as sticks. **B)** The intercalation zone between a dimer of pentamers. N protomers are coloured white with loops belonging to the intercalation zone coloured per protomer. Putative interacting residues are presented as sticks. Poorly resolved amino acid side chains are modelled as alanines. The representation is centred around the N_i protomer (lime) of the top ring, showing the N_{i+1} (pink) protomer as well as the bottom ring protomers N_i (n_i ; gold), N_{i-1} (n_{i-1} ; purple), and N_{i+1} (n_{i+1} ; coral).

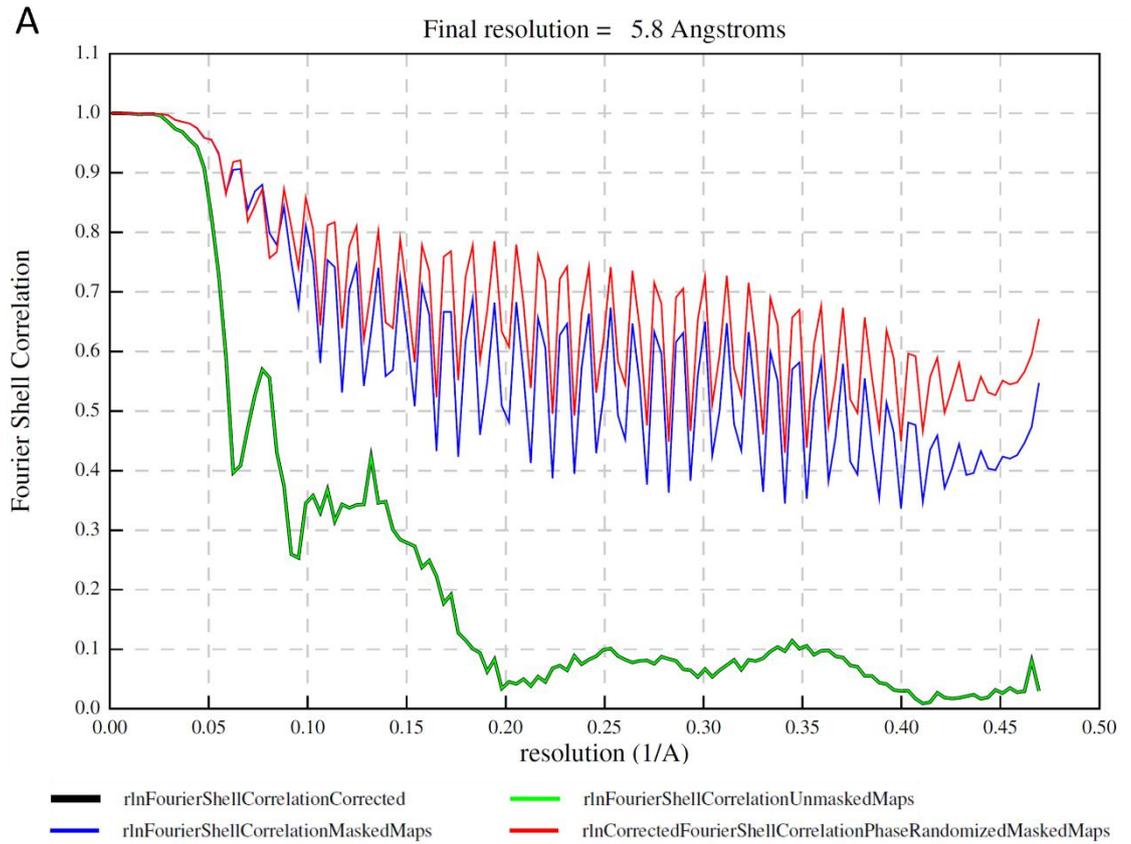


Figure S5.14. Refinement and post-processing for the helical HeV NLF map.
Fourier shell correlation plot from post-processing of the helical HeV NLF map refinement.

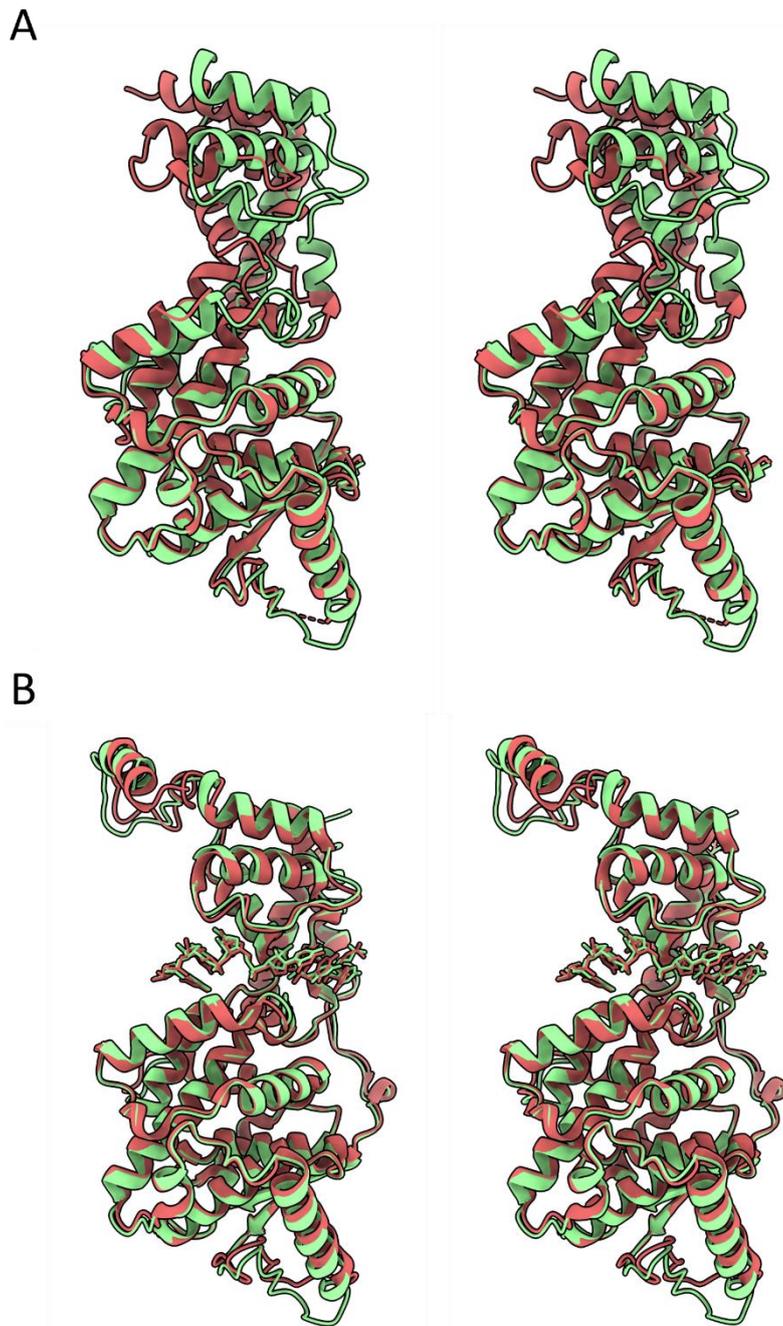


Figure S5.15. Comparison of HeV and NiV N protein structures.

Stereographic ribbon representations of the structural alignment of the sauronoid HeV N protein model and two NiV N protein models. Models are presented in ribbon visualisation, with the HeV structure in green and the NiV structures in red. Stereographs were generated by taking two images with a relative rotation of 5° about the y-axis. **A**) Alignment of the sauronoid HeV N protein with the crystal structure of NiV N protein (PDB: 4CO6) with an RMSD of 0.783 Å. The RNA hexamer was removed and the HeV N structure was truncated by removing the Nt- and Ct-arms to resemble the truncated apo-NiV N protein. The NiV P protein domain present in the crystal structure was also removed. **B**) Alignment of the sauronoid HeV N protein with the helical cryo-EM structure of NiV N protein (PDB: 7NT5) with an RMSD of 0.756 Å.

HeV_reference	M	S	D	I	F	E	E	A	A	S	FR	S	Y	Q	S	K	L	G	R	D	G	R	A	S	A	A	T	A	T	L	T	T	K	I	R	I	F	V	P	A	T	N	S	P	E	L	R	W	E	L	T	L	F	A	L	D	V	I	R	60			
HeV_variant	M	G	D	I	F	E	E	A	A	S	FR	N	Y	Q	S	K	L	G	R	D	G	R	A	S	A	A	T	A	T	L	T	T	K	I	R	I	F	V	P	A	T	N	S	P	E	L	R	W	E	L	T	L	F	A	L	D	V	I	R	60			
	*	.	*	*	*	*	*	*	*	*	*	.	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*						
HeV_reference	S	P	S	A	A	E	S	M	K	I	G	A	A	F	T	L	I	S	M	Y	S	E	R	P	G	A	L	I	R	S	L	L	N	D	P	I	E	A	V	I	I	D	V	G	S	M	L	N	G	I	P	V	M	E	R	R	G	D	K	120			
HeV_variant	S	P	S	A	A	E	S	M	K	I	G	A	A	F	T	L	I	S	M	Y	S	E	R	P	G	A	L	I	R	S	L	L	N	D	P	I	E	A	V	I	I	D	V	G	S	M	L	N	G	I	P	V	M	E	R	R	G	D	K	120			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	A	Q	E	E	M	E	G	L	M	R	I	L	K	T	A	R	E	S	S	K	G	K	T	P	F	V	D	S	R	A	Y	G	L	R	I	T	D	M	S	T	L	V	S	A	V	I	T	I	E	A	Q	I	W	I	L	I	A	K	A	V	180		
HeV_variant	A	Q	E	E	M	E	G	L	M	R	I	L	K	T	A	R	E	S	S	K	G	K	T	P	F	V	D	S	R	A	Y	G	M	R	I	T	D	M	S	T	L	V	S	A	V	I	T	I	E	A	Q	I	W	I	L	I	A	K	A	V	180		
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	T	A	P	D	T	A	E	E	S	E	T	R	R	W	A	K	Y	V	Q	Q	K	R	V	N	P	F	F	A	L	T	Q	Q	W	L	T	E	M	R	N	L	L	S	Q	S	L	S	V	R	K	F	M	V	E	I	L	M	E	V	K	K	240		
HeV_variant	T	A	P	D	T	A	E	E	S	E	T	R	R	W	A	K	Y	V	Q	Q	K	R	V	N	P	F	F	A	L	T	Q	Q	W	L	T	E	M	R	N	L	L	S	Q	S	L	S	V	R	K	F	M	V	E	I	L	M	E	V	K	K	240		
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	G	G	S	A	K	G	R	A	V	E	I	I	S	D	I	G	N	Y	V	E	E	T	G	M	A	G	F	F	A	T	I	R	F	G	L	E	T	R	Y	P	A	L	A	L	N	E	F	Q	S	D	L	N	T	I	K	G	L	M	L	L	300		
HeV_variant	G	G	S	A	K	G	R	A	V	E	I	I	S	D	I	G	N	Y	V	E	E	E	T	G	M	A	G	F	F	A	T	I	R	F	G	L	E	T	R	Y	P	A	L	A	L	N	E	F	Q	S	D	L	N	T	I	K	G	L	M	L	L	300	
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	Y	R	E	I	G	P	R	A	P	Y	M	V	L	L	E	E	S	I	Q	T	K	F	A	P	G	G	Y	P	L	L	W	S	F	A	M	G	V	A	T	T	I	D	R	S	M	G	A	L	N	I	N	R	G	Y	L	E	P	M	Y	F	360		
HeV_variant	Y	R	E	I	G	P	R	A	P	Y	M	V	L	L	E	E	S	I	Q	T	K	F	A	P	G	G	Y	P	L	L	W	S	F	A	M	G	V	A	T	T	I	D	R	S	M	G	A	L	N	I	N	R	G	Y	L	E	P	M	Y	F	360		
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	R	L	G	Q	K	S	A	R	H	H	A	G	G	I	D	Q	N	M	A	N	K	L	G	L	N	S	D	Q	V	A	E	L	A	A	A	V	Q	E	T	S	V	G	R	Q	D	N	N	M	Q	A	R	E	A	K	F	A	A	G	G	V	420		
HeV_variant	R	L	G	Q	K	S	A	R	H	H	A	G	G	I	D	Q	N	M	A	N	K	L	G	L	N	P	D	Q	V	A	E	E	L	A	A	A	V	Q	E	T	S	V	V	G	R	Q	D	N	N	M	Q	A	R	E	A	K	F	A	A	G	G	V	420
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	L	V	G	G	G	E	Q	D	I	D	E	E	E	P	I	E	H	S	G	R	Q	S	V	T	F	K	R	E	M	S	M	S	L	A	D	S	V	P	S	S	S	V	S	T	S	G	G	T	R	L	T	N	S	L	L	N	L	R	480				
HeV_variant	L	V	G	G	G	E	Q	D	I	D	E	D	E	E	P	I	E	H	S	G	R	Q	S	V	T	F	K	R	E	M	S	M	S	L	A	D	S	L	P	S	S	S	V	S	T	S	G	G	T	R	L	T	N	S	L	L	N	L	R	480			
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					
HeV_reference	S	R	L	A	A	K	A	I	K	E	S	T	A	Q	S	S	E	R	N	P	P	N	N	R	P	Q	A	D	S	G	R	K	D	D	Q	E	P	K	P	A	Q	N	D	L	D	F	V	R	A	D	V	532											
HeV_variant	S	R	L	A	A	K	A	M	K	D	N	A	A	Q	S	S	V	E	K	N	T	P	N	N	R	P	Q	A	D	S	R	G	K	E	D	Q	E	S	K	P	A	Q	N	D	L	D	F	V	R	A	D	D	532										
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*					

Figure S5.16. Nucleoprotein sequence alignment of HeV and HeV-var.

The full-length nucleoprotein sequences of HeV (HeV_reference; NCBI NC_001906.3) and HeV-var (HeV_variant; GenBank MZ318101) were aligned in Clustal Omega. Amino acid substitutions are indicated in red. N protein domains are indicated for Ncore; Nt-arm (blue), NTD (orange), CTD (green), Ct-arm (yellow), and the Ntail (grey).

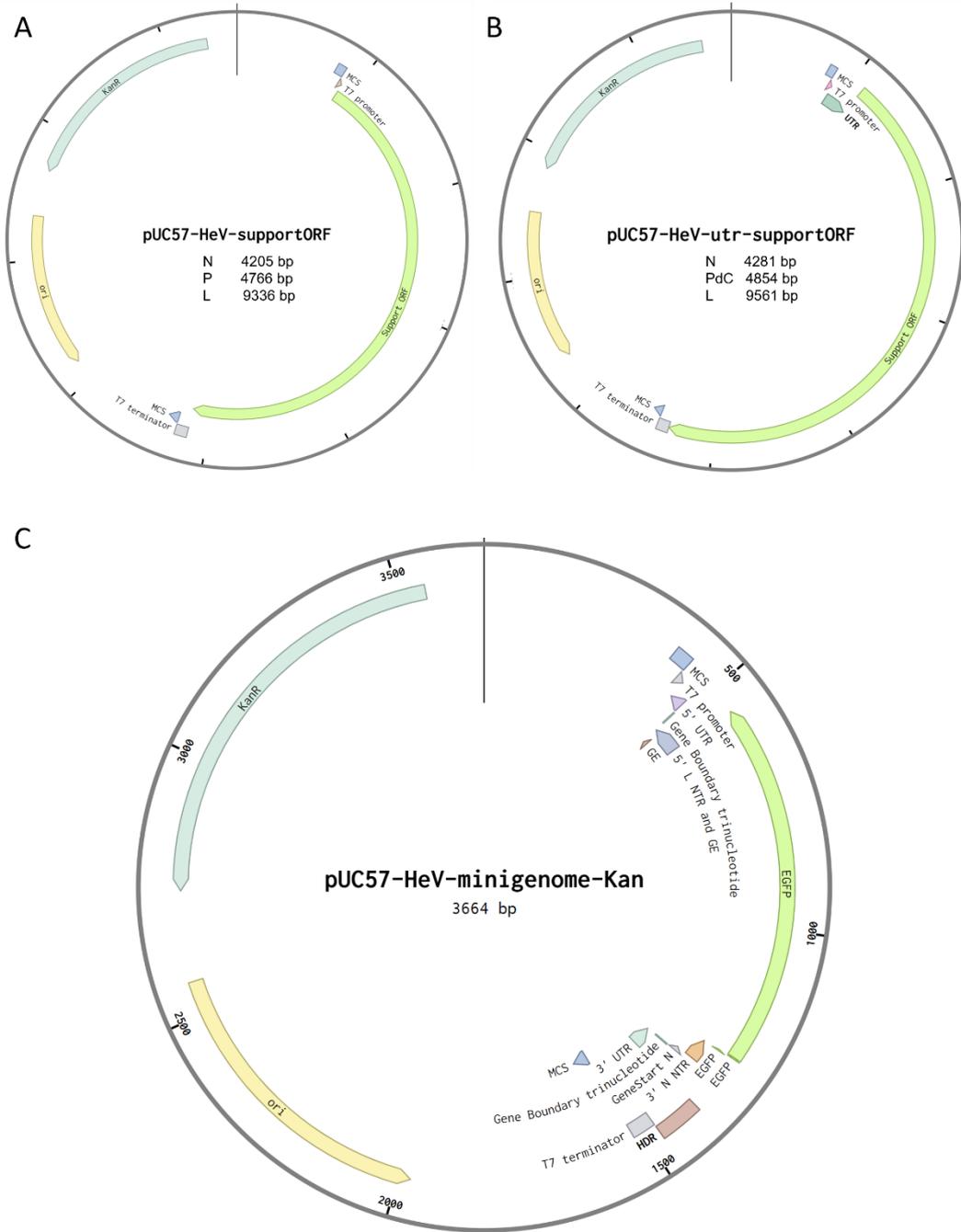


Figure S6.1. Plasmid maps of the HeV replicon system.

Plasmid maps of the pUC57-HeV support ORFs before (A) and after (B) UTR addition and C ORF silencing (PdC support only), and the pUC57-HeV-minigenome (C). Features are annotated and annotations provide the ORF directions. The vertical black line indicates the index nucleotide. Starting at the index nucleotide going clockwise, in order of appearance, the annotations indicate the following. **A-B** **MCS**, multiple cloning site. **T7 promoter**, T7 RNA polymerase promoter sequence. **UTR**, untranslated region/non-translated region (only in B). **Support ORF**, open reading frame for the support genes of the nucleoprotein, phosphoprotein, or polymerase. **T7 terminator**, T7 RNA polymerase terminator sequence. **ori**, the origin of replication. **KanR**, kanamycin resistance gene. **C** **MCS**, multiple cloning site. **T7 promoter**, T7 RNA polymerase promoter sequence. **5' UTR**, the HeV 5' genomic untranslated region. **5' L NTR and GE**, the L gene 5' non-translated region and gene end sequences. **EGFP**, ORF for EGFP. **3' N NTR**, the N gene 3' non-translated region. **GeneStart N**, the N gene start sequence. **3' UTR**, the genomic 3' untranslated region. **HDR**, hepatitis delta virus ribozyme. **T7 terminator**, T7 RNA polymerase terminator sequence. **ori**, the origin of replication. **KanR**, kanamycin resistance gene. Maps were generated in Benchling (Benchling, 2022).

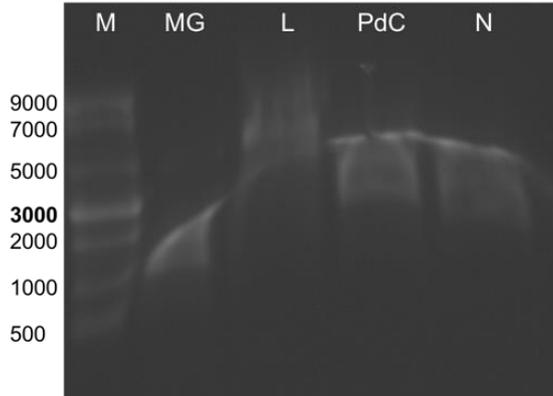


Figure S6.2. Agarose gel of *in vitro* transcribed replicon RNA.

The pUC57-HeV plasmids carrying the minigenome (MG) and the L, PdC, and N support genes were *in vitro* transcribed to RNA using the T7 promoter within the pUC57-HeV plasmids. The RNA was purified post-reaction and a sample was resolved on agarose gel. M is an RNA gel marker. Expected transcript lengths are; MG: 1073, L: 6939, PdC: 2310, N: 1707.

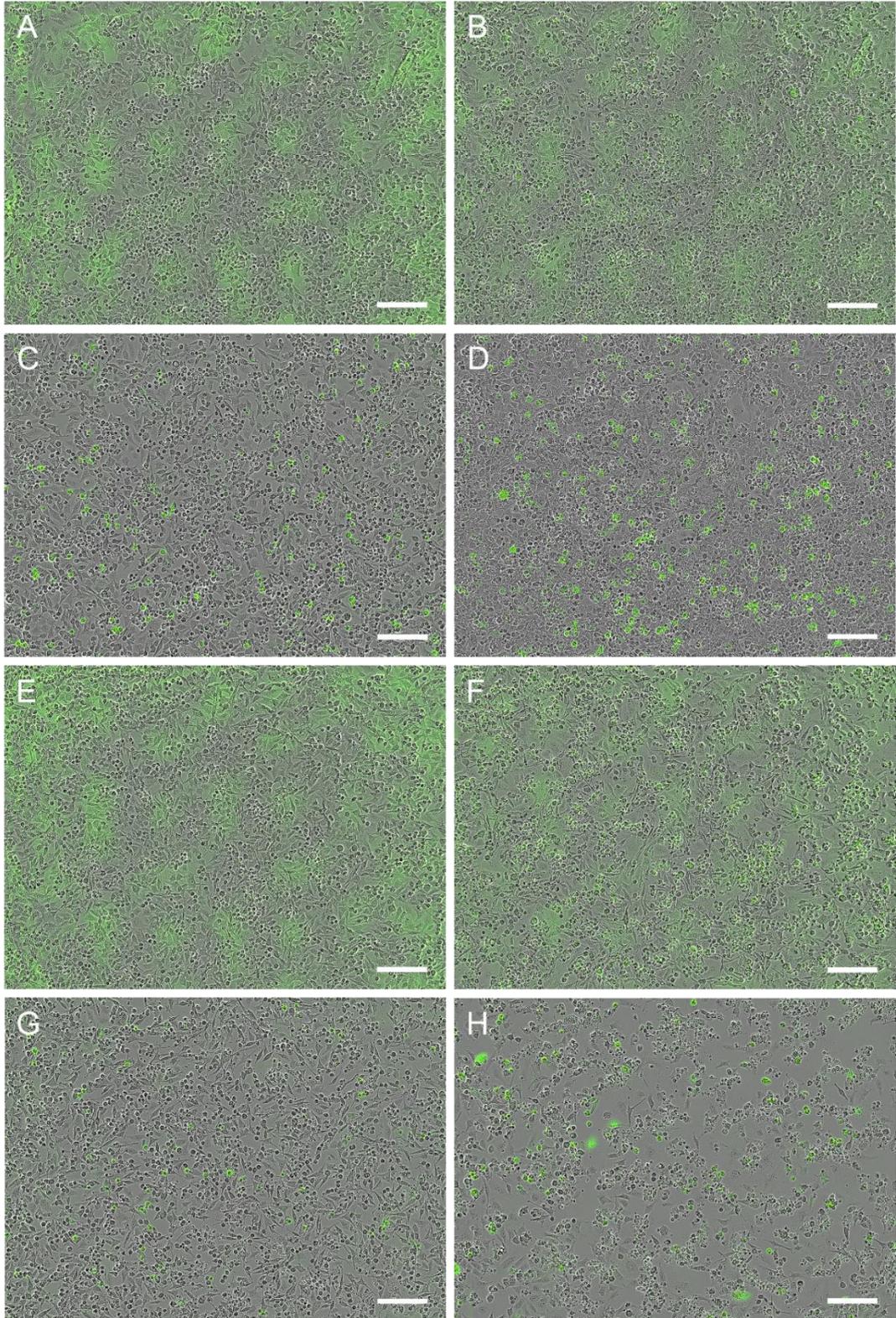


Figure S6.3. Fluorescence and phase-contrast images of RNA-transfected cells.

BSR T7 cells were transfected with *in vitro* transcribed RNA of the HeV minigenome and the L, PdC, and N supports. Representative images depict phase contrast with green fluorescence overlay adjusted for background signal on a per-image basis. Transfection mixture was replaced with media at 6 h.p.t. (A-D) or not replaced (E-H). A-B Full replicon transfected cells at 20 (A) or 50 (B) h.p.t. with a media change at 6 h.p.t. C-D EGFP positive control transfected cells at 20 (C) or 50 (D) h.p.t. with a media change at 6 h.p.t. E-F Full replicon transfected cells at 20 (E) or 50 (F) h.p.t. with no media exchange. G-H EGFP positive control transfected cells at 20 (G) or 50 (H) h.p.t. with no media exchange.

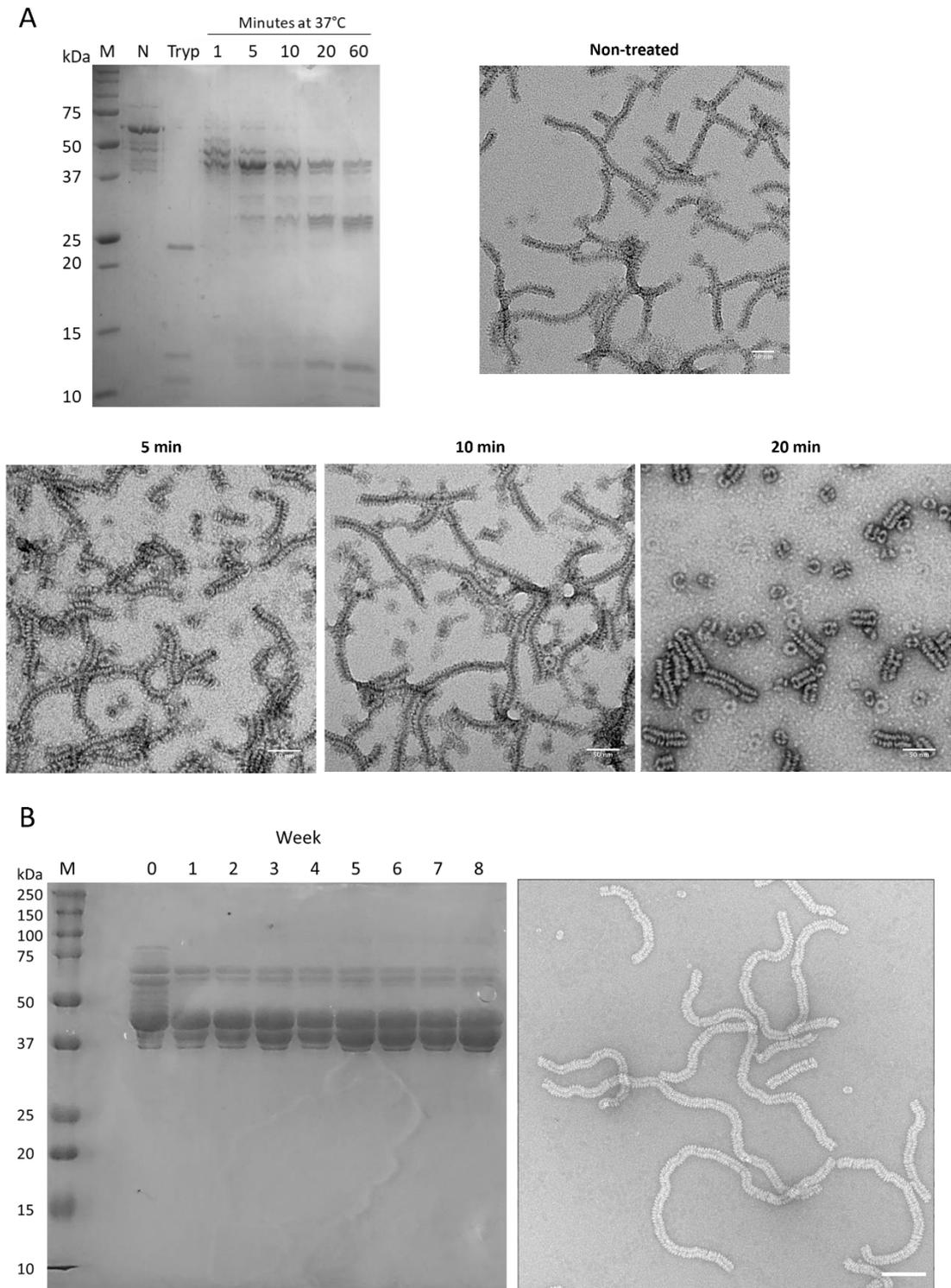


Figure S7.1. Protease treatment and supra-cryogenic incubation of HeV N protein.

A) Purified HeV N protein (SEC Fraction 12) was treated with 100 mg/ml final concentration trypsin for 1, 5, 10, 20, or 60 minutes at 37°C and the reaction was stopped by flash freezing in LN. Samples were then thawed on ice and immediately analysed by SDS-PAGE and negative stain EM. HeV N exhibits extensive degradation after short incubation with trypsin, but plateaus with three main species at ~40 kDa, ~28 kDa, and ~12 kDa. Trypsin digest had no rigidifying effect on NLFs but generated exceedingly shorter helices.

B) Purified HeV N protein (SEC Fraction 12) was incubated at 4 °C for up to 8 weeks. Aliquots were taken every week and flash frozen in LN. Samples were then thawed on ice and immediately analysed by SDS-PAGE and negative stain EM. HeV NLFs exhibit no appreciable rigidification after incubation at 4 °C. Scale bars are 50 nm in A and 100 nm in B.

Supplementary Tables

Table S1.1 Henipaviral nucleoprotein similarities

Protein sequences were retrieved from henipaviral genome reference sequences and compared to HeV N protein in NCBI BLAST (blastp suite), providing identity percentages

N protein	Percent Identity v HeV N protein
<i>Hendra henipavirus</i>	100
<i>Nipah henipavirus</i>	92
<i>Cedar henipavirus</i>	62
<i>Ghanaian Bat henipavirus</i>	56
<i>Mojiang henipavirus</i>	52
<i>Langya henipavirus</i>	59
<i>Daeryong henipavirus</i>	57
<i>Gamak henipavirus</i>	56
<i>Angavokely henipavirus</i>	49

Table S5.1 HeV sauronoid nucleoprotein map and model parameters

Data		
Box		
Lengths (Å)	224.72, 224.72, 139.52	
Angles (°)	90.00, 90.00, 90.00	
Supplied resolution (Å)	3.5	
Map Resolution Estimates	Masked	Unmasked
Using map alone (d99)	3.41	3.40
Overall Biso	60.00	60.00
d_model	3.30	3.30
d_model (B factors = 0)	3.90	3.90
FSC (model) = 0	3.14	3.14
FSC (model) = 0.143	3.23	3.23
FSC (model) = 0.5	3.65	3.52
Mask smoothing radius (Å)	6.98	
Map min/max/mean	-0.05/0.12/0.01	
Correlation Coefficients		
CC (mask)	0.84	
CC (box)	0.87	
CC (peaks)	0.80	
CC (volume)	0.84	
Mean CC for ligands	---	

Bibliography

- 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., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*. **66**(Pt 2), pp.213–221.
- Advisory Committee on Dangerous Pathogens (ACDP) 2021. *The Approved List of Biological Agents* [Online]. Available from: <http://www.hse.gov.uk/pubns/misc208.pdf>.
- Aebi, U. and Pollard, T.D. 1987. A glow discharge unit to render electron microscope grids and other surfaces hydrophilic. *Journal of electron microscopy technique*. **7**(1), pp.29–33.
- Afonine, P. V., Poon, B.K., Read, R.J., Sobolev, O. V., Terwilliger, T.C., Urzhumtsev, A. and Adams, P.D. 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*. **74**(6), pp.531–544.
- Alayyoubi, M., Leser, G.P., Kors, C.A. and Lamb, R.A. 2015. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein-RNA complex. *Proceedings of the National Academy of Sciences of the United States of America*. **112**(14), pp.E1792–E1799.
- Albertini, A.A.V., Wernimont, A.K., Muziol, T., Ravelli, R.B.G., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W.H. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*. **313**(5785), pp.360–363.
- Álvarez-Rodríguez, B., Tiede, C., Hoste, A.C.R., Surtees, R.A., Trinh, C.H., Slack, G.S., Chamberlain, J., Hewson, R., Fresco, A., Sastre, P., Tomlinson, D.C., Millner, P.A., Edwards, T.A. and Barr, J.N. 2020. Characterization and applications of a crimean-congo hemorrhagic fever virus nucleoprotein-specific affimer: Inhibitory effects in viral replication and development of colorimetric diagnostic tests. *PLoS Neglected Tropical Diseases*. **14**(6), pp.1–20.
- Andrejeva, J., Childs, K.S., Young, D.F., Carlos, T.S., Stock, N., Goodbourn, S. and Randall, R.E. 2004. The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(49), pp.17264–17269.
- Annand, E.J., Horsburgh, B.A., Xu, K., Reid, P.A., Poole, B., de Kantzow, M.C., Brown, N., Tweedie, A., Michie, M., Grewar, J.D., Jackson, A.E., Singanallur, N.B., Plain, K.M., Kim, K., Tachedjian, M., van der Heide, B., Crameri, S., Williams, D.T., Secombe, C., Laing, E.D., Sterling, S., Yan, L., Jackson, L., Jones, C., Plowright, R.K., Peel, A.J., Breed, A.C., Diallo, I., Dhand, N.K., Britton, P.N., Broder, C.C., Smith, I. and Eden, J.-S. 2022. Novel Hendra Virus Variant Detected by Sentinel Surveillance of Horses in Australia. *Emerging infectious diseases*. **28**(3), pp.693–704.
- Ariza, A., Tanner, S.J., Walter, C.T., Dent, K.C., Shepherd, D.A., Wu, W., Matthews, S. V., Hiscox, J.A., Green, T.J., Luo, M., Elliott, R.M., Fooks, A.R., Ashcroft, A.E., Stonehouse, N.J., Ranson, N.A., Barr, J.N. and Edwards, T.A. 2013. Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic Acids Research*. **41**(11), pp.5912–5926.
- Arragain, B., Reguera, J., Desfosses, A., Gutsche, I., Schoehn, G. and Malet, H. 2019. High resolution cryo-em structure of the helical rna-bound hantaan virus nucleocapsid reveals its assembly mechanisms. *eLife*. **8**, pp.1–14.

- Bächi, T. 1980. Intramembrane structural differentiation in Sendai virus maturation. *Virology*. **106**(1), pp.41–49.
- Bakker, S.E., Duquerroy, S., Galloux, M., Loney, C., Conner, E., Eléouët, J.-F., Rey, F.A. and Bhella, D. 2013. The respiratory syncytial virus nucleoprotein-RNA complex forms a left-handed helical nucleocapsid. *The Journal of general virology*. **94**(Pt 8), pp.1734–1738.
- Baltimore, D. 1971. Expression of animal virus genomes. *Bacteriological reviews*. **35**(3), pp.235–241.
- Bammes, B.E., Jakana, J., Schmid, M.F. and Chiu, W. 2010. Radiation damage effects at four specimen temperatures from 4 to 100 K. *Journal of Structural Biology*. **169**(3), pp.331–341.
- Barr, J.N. and Fearn, R. 2016. Genetic Instability of RNA Viruses. *Genome Stability*., pp.21–35.
- Baseler, L., Scott, D.P., Saturday, G., Horne, E., Rosenke, R., Thomas, T., Meade-White, K., Haddock, E., Feldmann, H. and de Wit, E. 2016. Identifying Early Target Cells of Nipah Virus Infection in Syrian Hamsters. *PLoS neglected tropical diseases*. **10**(11), p.e0005120.
- Battisti, A.J., Meng, G., Winkler, D.C., McGinnes, L.W., Plevka, P., Steven, A.C., Morrison, T.G. and Rossmann, M.G. 2012. Structure and assembly of a paramyxovirus matrix protein. *Proceedings of the National Academy of Sciences*. **109**(35), pp.13996–14000.
- Benchling 2022. Benchling [Biology Software].
- Beniac, D.R., Melito, P.L., Devarenes, S.L., Hiebert, S.L., Rabb, M.J., Lamboo, L.L., Jones, S.M. and Booth, T.F. 2012. The organisation of Ebola virus reveals a capacity for extensive, modular polyploidy. *PLoS one*. **7**(1), p.e29608.
- Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J. and Berger, B. 2019. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods*. **16**(11), pp.1153–1160.
- Bharat, T.A.M., Riches, J.D., Kolesnikova, L., Welsch, S., Krähling, V., Davey, N., Parsy, M.-L., Becker, S. and Briggs, J.A.G. 2011. Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS biology*. **9**(11), p.e1001196.
- Bhella, D., Ralph, A. and Yeo, R.P. 2004. Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. *Journal of molecular biology*. **340**(2), pp.319–331.
- Bloyet, L.-M. 2021. The Nucleocapsid of Paramyxoviruses: Structure and Function of an Encapsidated Template. *Viruses*. **13**(12).
- Blumberg, B.M., Giorgi, C. and Kolakofsky, D. 1983. N protein of vesicular stomatitis virus selectively encapsidates leader RNA in vitro. *Cell*. **32**(2), pp.559–567.
- Blumberg, B.M., Leppert, M. and Kolakofsky, D. 1981. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell*. **23**(3), pp.837–845.
- Bonaparte, M.I., Dimitrov, A.S., Bossart, K.N., Crameri, G., Mungall, B.A., Bishop, K.A., Choudhry, V., Dimitrov, D.S., Wang, L.F., Eaton, B.T. and Broder, C.C. 2005. Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proceedings of the National Academy of Sciences of the United States of America*. **102**(30), pp.10652–10657.
- Boonyaratanakornkit, J., Bartlett, E., Schomacker, H., Surman, S., Akira, S., Bae, Y.-S., Collins, P., Murphy, B. and Schmidt, A. 2011. The C proteins of human parainfluenza virus type 1

- limit double-stranded RNA accumulation that would otherwise trigger activation of MDA5 and protein kinase R. *Journal of virology*. **85**(4), pp.1495–1506.
- Bossart, K.N., Cramer, G., Dimitrov, A.S., Mungall, B.A., Feng, Y.-R., Patch, J.R., Choudhary, A., Wang, L.-F., Eaton, B.T. and Broder, C.C. 2005. Receptor binding, fusion inhibition, and induction of cross-reactive neutralizing antibodies by a soluble G glycoprotein of Hendra virus. *Journal of virology*. **79**(11), pp.6690–6702.
- Bossart, K.N., McEachern, J.A., Hickey, A.C., Choudhry, V., Dimitrov, D.S., Eaton, B.T. and Wang, L.-F. 2007. Neutralization assays for differential henipavirus serology using Bio-Plex protein array systems. *Journal of virological methods*. **142**(1–2), pp.29–40.
- Bossart, K.N., Wang, L.-F., Flora, M.N., Chua, K.B., Lam, S.K., Eaton, B.T. and Broder, C.C. 2002. Membrane Fusion Tropism and Heterotypic Functional Activities of the Nipah Virus and Hendra Virus Envelope Glycoproteins. *Journal of Virology*. **76**(22), pp.11186–11198.
- Bossart, K.N., Wang, L.F., Eaton, B.T. and Broder, C.C. 2001. Functional expression and membrane fusion tropism of the envelope glycoproteins of Hendra virus. *Virology*. **290**(1), pp.121–135.
- Braun, M.R., Deflubé, L.R., Noton, S.L., Mawhorter, M.E., Tremaglio, C.Z. and Fearn, R. 2017. RNA elongation by respiratory syncytial virus polymerase is calibrated by conserved region V. *PLoS pathogens*. **13**(12), p.e1006803.
- 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. *Journal of structural biology*. **177**(3), pp.630–637.
- Broder, C.C., Weir, D.L. and Reid, P.A. 2016. Hendra virus and Nipah virus animal vaccines. *Vaccine*. **34**(30), pp.3525–3534.
- Buchholz, U.J., Finke, S. and Conzelmann, K.K. 1999. Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. *Journal of virology*. **73**(1), pp.251–259.
- Burgess, S.A., Walker, M.L., Thirumurugan, K., Trinick, J. and Knight, P.J. 2004. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *Journal of Structural Biology*. **147**(3), pp.247–258.
- Butt, T.R., Edavettal, S.C., Hall, J.P. and Mattern, M.R. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*. **43**(1), pp.1–9.
- Calain, P. and Roux, L. 1993. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *Journal of Virology*. **67**(8), pp.4822–4830.
- Carlo, S. De and Harris, J.R. 2012. TEM et cryTEM. *Micron*. **42**(2), pp.117–131.
- Carter, S.D., Barr, J.N. and Edwards, T.A. 2012. Expression, purification and crystallization of the Crimean-Congo haemorrhagic fever virus nucleocapsid protein. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. **68**(5), pp.569–573.
- Carter, S.D., Surtees, R., Walter, C.T., Ariza, A., Bergeron, É., Nichol, S.T., Hiscox, J.A., Edwards, T.A. and Barr, J.N. 2012. Structure, Function, and Evolution of the Crimean-Congo Hemorrhagic Fever Virus Nucleocapsid Protein. *Journal of Virology*. **86**(20), pp.10914–10923.

- Caspar, D.L. and Klug, A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor symposia on quantitative biology*. **27**, pp.1–24.
- Charlton, F.W., Pearson, H.M., Hover, S., Lippiat, J.D., Fontana, J., Barr, J.N. and Mankouri, J. 2020. Ion Channels as Therapeutic Targets for Viral Infections: Further Discoveries and Future Perspectives. *Viruses*. **12**(8).
- Chen, V.B., Arendall, W.B. 3rd, 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 crystallographica. Section D, Biological crystallography*. **66**(Pt 1), pp.12–21.
- Cheng, Y. 2018. Single-particle cryo-EM-How did it get here and where will it go. *Science*. **361**(6405), pp.876–880.
- Cheng, Y. 2015. Single-particle Cryo-EM at crystallographic resolution. *Cell*. **161**(3), pp.450–457.
- Cheng, Y., Grigorieff, N., Penczek, P.A. and Walz, T. 2015. A primer to single-particle cryo-electron microscopy. *Cell*. **161**(3), pp.438–449.
- Chodosh, L.A. 2001. Purification of DNA-binding proteins using biotin/streptavidin affinity systems. *Current protocols in molecular biology*. **Chapter 12**, Unit 12.6.
- Chong, H.T., Kamarulzaman, A., Tan, C.T., Goh, K.J., Thayaparan, T., Kunjapan, S.R., Chew, N.K., Chua, K.B. and Lam, S.K. 2001. Treatment of acute Nipah encephalitis with ribavirin. *Annals of neurology*. **49**(6), pp.810–813.
- Chui, S.T. and Zhou, L. 2013. Helical Structures *In: Electromagnetic Behaviour of Metallic Wire Structures*. Washington, DC: ASM Press, pp.45–61.
- Coloma, R., Arranz, R., de la Rosa-Trevín, J.M., Sorzano, C.O.S., Munier, S., Carlero, D., Naffakh, N., Ortín, J. and Martín-Benito, J. 2020. Structural insights into influenza A virus ribonucleoproteins reveal a processive helical track as transcription mechanism. *Nature microbiology*. **5**(5), pp.727–734.
- Compans, R.W., Content, J. and Duesberg, P.H. 1972. Structure of the ribonucleoprotein of influenza virus. *Journal of virology*. **10**(4), pp.795–800.
- Conley, M.J., Short, J.M., Burns, A.M., Streetley, J., Hutchings, J., Bakker, S.E., Power, B.J., Jaffery, H., Haney, J., Zanetti, G., Murcia, P.R., Stewart, M., Fearn, R., Vijaykrishnan, S. and Bhella, D. 2022. Helical ordering of envelope-associated proteins and glycoproteins in respiratory syncytial virus. *The EMBO journal*. **41**(3), p.e109728.
- Cormack, B.P., Valdivia, R.H. and Falkow, S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. **173**(1 Spec No), pp.33–38.
- Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*. **62**(9), pp.1002–1011.
- Cox, R., Pickar, A., Qiu, S., Tsao, J., Rodenburg, C., Dokland, T., Elson, A., He, B. and Luo, M. 2014. Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(42), pp.15208–15213.
- Cox, R.M. and Plemper, R.K. 2017. Structure and organization of paramyxovirus particles. *Current Opinion in Virology*. **24**, pp.105–114.

- Cronin, J., Zhang, X.-Y. and Reiser, J. 2005. Altering the tropism of lentiviral vectors through pseudotyping. *Current gene therapy*. **5**(4), pp.387–398.
- Cull, M.G. and Schatz, P.J. 2000. Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods in enzymology*. **326**, pp.430–440.
- Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W. and Kolakofsky, D. 1993. The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *Journal of virology*. **67**(7), pp.4358–4364.
- DeBuysscher, B.L., Scott, D., Marzi, A., Prescott, J. and Feldmann, H. 2014. Single-dose live-attenuated Nipah virus vaccines confer complete protection by eliciting antibodies directed against surface glycoproteins. *Vaccine*. **32**(22), pp.2637–2644.
- Desfosses, A., Milles, S., Jensen, M.R., Guseva, S., Colletier, J.-P., Maurin, D., Schoehn, G., Gutsche, I., Ruigrok, R.W.H. and Blackledge, M. 2019. Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proceedings of the National Academy of Sciences*. **116**(10), pp.4256–4264.
- Diederich, S., Sauerhering, L., Weis, M., Altmepfen, H., Schaschke, N., Reinheckel, T., Erbar, S. and Maisner, A. 2012. Activation of the Nipah virus fusion protein in MDCK cells is mediated by cathepsin B within the endosome-recycling compartment. *Journal of virology*. **86**(7), pp.3736–3745.
- Ding, N.-Z., Xu, D.-S., Sun, Y.-Y., He, H.-B. and He, C.-Q. 2017. A permanent host shift of rabies virus from Chiroptera to Carnivora associated with recombination. *Scientific reports*. **7**(1), p.289.
- Dobro, M.J., Melanson, L.A., Jensen, G.J. and McDowell, A.W. 2010. Plunge freezing for electron cryomicroscopy. In: G. J. B. T.-M. in E. Jensen, ed. *Methods in Enzymology* [Online]. Academic Press, pp.63–82. Available from: <https://www.sciencedirect.com/science/article/pii/S0076687910810031>.
- Dong, E., Du, H. and Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious diseases*. **20**(5), pp.533–534.
- Dubochet, J., Ducommun, M., Zollinger, M. and Kellenberger, E. 1971. A new preparation method for dark-field electron microscopy of biomacromolecules. *Journal of Ultrastructure Research*. **35**(1–2), pp.147–167.
- Eaton, B.T., Broder, C.C., Middleton, D. and Wang, L.F. 2006. Hendra and Nipah viruses: Different and dangerous. *Nature Reviews Microbiology*. **4**(1), pp.23–35.
- Egelman, E.H., Wu, S.S., Amrein, M., Portner, A. and Murti, G. 1989. The Sendai virus nucleocapsid exists in at least four different helical states. *Journal of Virology*. **63**(5), pp.2233–2243.
- Emsley, P. and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*. **60**(Pt 12 Pt 1), pp.2126–2132.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*. **66**(4), pp.486–501.
- Erickson, H.P. and Klug, A. 1970. The Fourier Transform of an Electron Micrograph: Effects of Defocussing and Aberrations, and Implications for the Use of Underfocus Contrast Enhancement. *Philosophical Transactions of the Royal Society of London B*. **261**(11),

pp.105–118.

- Fenner, F., Ginsberg, H.S., Maurin, J. and Valenta, V. 1974. International committee on taxonomy of viruses. *Intervirology*. [Online]. **3**(3), p.199. Available from: <https://ictv.global/taxonomy/>.
- Ferron, F., Li, Z., Danek, E.I., Luo, D., Wong, Y., Coutard, B., Lantez, V., Charrel, R., Canard, B., Walz, T. and Lescar, J. 2011. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS pathogens*. **7**(5), p.e1002030.
- Field, H., de Jong, C., Melville, D., Smith, C., Smith, I., Broos, A., Kung, Y.H. (Nina), McLaughlin, A. and Zeddeman, A. 2011. Hendra Virus Infection Dynamics in Australian Fruit Bats. *PLOS ONE*. **6**(12), pp.1–6.
- Flint, S.J., Enquist, L.W., Racaniello, V.R. and Skalka, A.M. 2009a. Capsids or Nuceocapsids with Icosahedral Symmetry *In: Principles of Virology*. Washington, DC: ASM Press, pp.92–106.
- Flint, S.J., Enquist, L.W., Racaniello, V.R. and Skalka, A.M. 2009b. Editing of Viral mRNAs *In: Principles of Virology*. Washington, DC: ASM Press, pp.354–356.
- Flint, S.J., Enquist, L.W., Racaniello, V.R. and Skalka, A.M. 2009c. Genomes and Genetics *In: Principles of Virology*. Washington, DC: ASM Press, pp.53–80.
- Flint, S.J., Enquist, L.W., Racaniello, V.R. and Skalka, A.M. 2009d. Origins of Diversity in RNA Virus Genomes *In: Principles of Virology*. Washington, DC: ASM Press, pp.195–199.
- Flint, S.J., Enquist, L.W., Racaniello, V.R. and Skalka, A.M. 2009e. Structure *In: Principles of Virology*. Washington, DC: ASM Press, pp.83–91.
- Freiberg, A., Dolores, L.K., Enterlein, S. and Flick, R. 2008. Establishment and characterization of plasmid-driven minigenome rescue systems for Nipah virus: RNA polymerase I- and T7-catalyzed generation of functional paramyxoviral RNA. *Virology*. **370**(1), pp.33–44.
- Fuller, J., Surtees, R.A., Shaw, A.B., Álvarez-Rodríguez, B., Slack, G.S., Bell-Sakyi, L., Mankouri, J., Edwards, T.A., Hewson, R. and Barr, J.N. 2019. Hazara nairovirus elicits differential induction of apoptosis and nucleocapsid protein cleavage in mammalian and tick cells. *The Journal of general virology*. **100**(3), pp.392–402.
- Gainey, M.D., Dillon, P.J., Clark, K.M., Manuse, M.J. and Parks, G.D. 2008. Paramyxovirus-induced shutoff of host and viral protein synthesis: role of the P and V proteins in limiting PKR activation. *Journal of virology*. **82**(2), pp.828–839.
- Gallagher, J.R., Kim, A.J., Gulati, N.M. and Harris, A.K. 2019. Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. *Current protocols in microbiology*. **54**(1), p.e90.
- Gao, Y., Cao, D., Ahn, H.M., Swain, A., Hill, S., Ogilvie, C., Kurien, M., Rahmatullah, T. and Liang, B. 2020. In vitro trackable assembly of RNA-specific nucleocapsids of the respiratory syncytial virus. *The Journal of biological chemistry*. **295**(3), pp.883–895.
- Geisbert, T.W., Bobb, K., Borisevich, V., Geisbert, J.B., Agans, K.N., Cross, R.W., Prasad, A.N., Fenton, K.A., Yu, H., Fouts, T.R., Broder, C.C. and Dimitrov, A.S. 2021. A single dose investigational subunit vaccine for human use against Nipah virus and Hendra virus. *NPJ vaccines*. **6**(1), p.23.
- Gekko, K. and Timasheff, S.N. 1981. Mechanism of Protein Stabilization by Glycerol: Preferential Hydration in Glycerol-Water Mixtures. *Biochemistry*. **20**(16), pp.4667–4676.

- 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 Science*. **27**(1), pp.14–25.
- Goff, P.H., Gao, Q. and Palese, P. 2012. A majority of infectious Newcastle disease virus particles contain a single genome, while a minority contain multiple genomes. *Journal of virology*. **86**(19), pp.10852–10856.
- Goldsmith, C.S., Whistler, T., Rollin, P.E., Ksiazek, T.G., Rota, P.A., Bellini, W.J., Daszak, P., Wong, K.T., Shieh, W.-J. and Zaki, S.R. 2003. Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches. *Virus research*. **92**(1), pp.89–98.
- Green, T.J., Zhang, X., Wartz, G.W. and Luo, M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science*. **313**(5785), pp.357–360.
- Guillaume, V., Contamin, H., Loth, P., Georges-Courbot, M.-C., Lefevre, A., Marianneau, P., Chua, K.B., Lam, S.K., Buckland, R., Deubel, V. and Wild, T.F. 2004. Nipah virus: vaccination and passive protection studies in a hamster model. *Journal of virology*. **78**(2), pp.834–840.
- Guillaume, V., Wong, K.T., Looi, R.Y., Georges-Courbot, M.-C., Barrot, L., Buckland, R., Wild, T.F. and Horvat, B. 2009. Acute Hendra virus infection: Analysis of the pathogenesis and passive antibody protection in the hamster model. *Virology*. **387**(2), pp.459–465.
- Gurley, E.S., Montgomery, J.M., Hossain, M.J., Bell, M., Azad, A.K., Islam, M.R., Molla, M.A.R., Carroll, D.S., Ksiazek, T.G., Rota, P.A., Lowe, L., Comer, J.A., Rollin, P., Czub, M., Grolla, A., Feldmann, H., Luby, S.P., Woodward, J.L. and Breiman, R.F. 2007. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerging Infectious Diseases*. **13**(7), pp.1031–1037.
- Guryanov, S.G., Liljeroos, L., Kasaragod, P., Kajander, T., Butcher, S.J. and García-Sastre, A. 2016. Crystal Structure of the Measles Virus Nucleoprotein Core in Complex with an N-Terminal Region of Phosphoprotein. *Journal of Virology*. **90**(6), pp.2849–2857.
- Gutsche, I., Desfosses, A., Effantin, G., Ling, W.L., Haupt, M., Ruigrok, R.W.H.H., Sachse, C. and Schoehn, G. 2015. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science*. **348**(6235), pp.704–707.
- Halpin, K., Bankamp, B., Harcourt, B.H., Bellini, W.J. and Rota, P.A. 2004. Nipah virus conforms to the rule of six in a minigenome replication assay. *Journal of General Virology*. **85**(3), pp.701–707.
- Halpin, K., Young, P.L., Field, H.E. and Mackenzie, J.S. 2000. Isolation of Hendra virus from pteropid bats: A natural reservoir of Hendra virus. *Journal of General Virology*. **81**(8), pp.1927–1932.
- Hanley, L.L., McGivern, D.R., Teng, M.N., Djang, R., Collins, P.L. and Fearn, R. 2010. Roles of the respiratory syncytial virus trailer region: effects of mutations on genome production and stress granule formation. *Virology*. **406**(2), pp.241–252.
- Hannemann, H. 2020. Viral replicons as valuable tools for drug discovery. *Drug discovery today*. **25**(6), pp.1026–1033.
- Harcourt, B.H., Tamin, A., Ksiazek, T.G., Rollin, P.E., Anderson, L.J., Bellini, W.J. and Rota, P.A. 2000. Molecular characterization of Nipah virus, a newly emergent paramyxovirus. *Virology*. **271**(2), pp.334–349.

- Hatsuzawa, K., Hosaka, M., Nakagawa, T., Nagase, M., Shoda, A., Murakami, K. and Nakayama, K. 1990. Structure and expression of mouse furin, a yeast Kex2-related protease. Lack of processing of coexpressed prorenin in GH4C1 cells. *Journal of Biological Chemistry*. **265**(36), pp.22075–22078.
- Hayat, M.A. 1986. Negative Staining *In: Basic Techniques for Transmission Electron Microscopy*. Elsevier, pp.232–264.
- Heggeness, M.H., Scheid, A. and Choppin, P.W. 1980. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proceedings of the National Academy of Sciences of the United States of America*. **77**(5), pp.2631–2635.
- Heggeness, M.H., Smith, P.R. and Choppin, P.W. 1982. In vitro assembly of the nonglycosylated membrane protein (M) of Sendai virus. *Proceedings of the National Academy of Sciences of the United States of America*. **79**(20), pp.6232–6236.
- Hitchman, R.B., Possee, R.D. and King, L.A. 2009. Hitchman (2009) Baculovirus Expression Systems for Recombinant Protein Production in Insect cells.pdf. *Recent Patents on Biotechnology*. **3**(1), pp.46–54.
- van der Hoek, L., Verschoor, E., Beer, M., Höper, D., Wernike, K., Van Ranst, M., Matthijnsens, J., Maes, P., Sastre, P., Rueda, P., Drexler, J.F., Barr, J., Edwards, T., Millner, P., Vermeij, P., de Groof, A., Thiel, V., Dijkman, R., Suter-Riniker, F., Leib, S., Koller, R., Ramette, A., Engler, O. and Beuret, C. 2018. Host switching pathogens, infectious outbreaks and zoonosis: A Marie Skłodowska-Curie innovative training network (HONOURs). *Virus Research*. **257**, pp.120–124.
- Hopkins, F.R., Álvarez-Rodríguez, B., Heath, G.R., Panayi, K., Hover, S., Edwards, T.A., Barr, J.N. and Fontana, J. 2022. The Native Orthobunyavirus Ribonucleoprotein Possesses a Helical Architecture. *mBio*, p.e0140522.
- Horikami, S.M., Smallwood, S. and Moyer, S.A. 1996. The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology*. **222**(2), pp.383–390.
- Hoste, A.C.R., Ruiz, T., Fernández-Pacheco, P., Jiménez-Clavero, M.Á., Djadjovski, I., Moreno, S., Brun, A., Edwards, T.A., Barr, J.N., Rueda, P. and Sastre, P. 2021. Development of a multiplex assay for antibody detection in serum against pathogens affecting ruminants. *Transboundary and Emerging Diseases*. **68**(3), pp.1229–1239.
- Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M. and Palese, P. 1987. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proceedings of the National Academy of Sciences of the United States of America*. **84**(22), pp.8140–8144.
- Hyatt, A.D., Zaki, S.R., Goldsmith, C.S., Wise, T.G. and Hengstberger, S.G. 2001. Ultrastructure of Hendra virus and Nipah virus within cultured cells and host animals. *Microbes and Infection*. **3**(4), pp.297–306.
- IMF 2020. World Economic Outlook, April 2020: The Great Lockdown. *World Economic Outlook*. [Online]. (May), p.177. Available from: <https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/weo-april-2020>.
- International Committee on Taxonomy of Viruses (ICTV) 2020. Taxonomy. *Taxonomy*. [Online]. Available from: <https://ictv.global/taxonomy/>.
- Iseni, F., Garcin, D., Nishio, M., Kedersha, N., Anderson, P. and Kolakofsky, D. 2002. Sendai

- virus trailer RNA binds TIAR, a cellular protein involved in virus-induced apoptosis. *The EMBO journal*. **21**(19), pp.5141–5150.
- Jarvis, D.L. 2009. *Chapter 14 Baculovirus-Insect Cell Expression Systems* [Online] 1st ed. Elsevier Inc. Available from: [http://dx.doi.org/10.1016/S0076-6879\(09\)63014-7](http://dx.doi.org/10.1016/S0076-6879(09)63014-7).
- Jenni, S., Horwitz, J.A., Bloyet, L.M., Whelan, S.P.J. and Harrison, S.C. 2022. Visualizing molecular interactions that determine assembly of a bullet-shaped vesicular stomatitis virus particle. *Nature Communications*. **13**(1), 2022.04.07.487545.
- Johnson, N.P.A.S. and Mueller, J. 2002. Updating the accounts: global mortality of the 1918-1920 'Spanish' influenza pandemic. *Bulletin of the history of medicine*. **76**(1), pp.105–115.
- Jordan, P.C., Liu, C., Raynaud, P., Lo, M.K., Spiropoulou, C.F., Symons, J.A., Beigelman, L. and Deval, J. 2018. Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLoS Pathogens*. **14**(2), pp.1–23.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**(7873), pp.583–589.
- Ke, Z., Strauss, J.D., Hampton, C.M., Brindley, M.A., Dillard, R.S., Leon, F., Lamb, K.M., Plemper, R.K. and Wright, E.R. 2018. Promotion of virus assembly and organization by the measles virus matrix protein. *Nature Communications*. **9**(1), p.1736.
- Ker, D.S., Jenkins, H.T., Greive, S.J. and Antson, A.A. 2021. CryoEM structure of the Nipah virus nucleocapsid assembly. *PLoS Pathogens*. **17**(7), pp.1–8.
- Kingsbury, D.W., Hsu, C.H. and Murti, K.G. 1978. Intracellular metabolism of Sendai virus nucleocapsids. *Virology*. **91**(1), pp.86–94.
- Kinsella, C.M., Santos, P.D., Postigo-Hidalgo, I., Folgueiras-González, A., Passchier, T.C., Szillat, K.P., Akello, J.O., Álvarez-Rodríguez, B. and Martí-Carreras, J. 2020. Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19. *PLoS Pathogens*. **16**(10).
- Klein, S., Cortese, M., Winter, S.L., Wachsmuth-Melm, M., Neufeldt, C.J., Cerikan, B., Stanifer, M.L., Boulant, S., Bartenschlager, R. and Chlanda, P. 2020. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature Communications*. **11**(1), pp.1–10.
- Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J. and Roux, L. 1998. Paramyxovirus RNA Synthesis and the Requirement for Hexamer Genome Length: the Rule of Six Revisited. *Journal of Virology*. **72**(2), pp.891–899.
- Krissinel, E. and Henrick, K. 2007. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*. **372**(3), pp.774–797.
- Kuo, M.Y., Sharmeen, L., Dinter-Gottlieb, G. and Taylor, J. 1988. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *Journal of virology*. **62**(12), pp.4439–4444.
- Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a

- protein. *Journal of molecular biology*. **157**(1), pp.105–132.
- Lamb, R.A., Mahy, B.W. and Choppin, P.W. 1976. The synthesis of sendai virus polypeptides in infected cells. *Virology*. **69**(1), pp.116–131.
- Lee, B., Broder, C.C. and Wang, L.-F. 2021. Henipaviruses: Hendra and Nipah Viruses *In: Fields Virology Volume 1: Emerging Viruses*. Philadelphia: Wolters Kluwer, pp.559–590.
- Lee, S.-H., Kim, K., Kim, J., No, J.S., Park, K., Budhathoki, S., Lee, S.H., Lee, J., Cho, S.H., Cho, S., Lee, G.-Y., Hwang, J., Kim, H.-C., Klein, T.A., Uhm, C.-S., Kim, W.-K. and Song, J.-W. 2021. Discovery and Genetic Characterization of Novel Paramyxoviruses Related to the Genus Henipavirus in Crocidura Species in the Republic of Korea. *Viruses*. **13**(10).
- Lelek, M., Gyparaki, M.T., Beliu, G., Schueder, F., Griffié, J., Manley, S., Jungmann, R., Sauer, M., Lakadamyali, M. and Zimmer, C. 2021. Single-molecule localization microscopy. *Nature Reviews Methods Primers*. **1**(1), p.39.
- Li, B., Wang, Q., Pan, X., De Castro, I.F., Sun, Y., Guo, Y., Tao, X., Risco, C., Sui, S.F. and Lou, Z. 2013. Bunyamwera virus possesses a distinct nucleocapsid protein to facilitate genome encapsidation. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(22), pp.9048–9053.
- Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C. and Zhang, Y.Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife*. **2015**(4), pp.1–26.
- Li, Q., Liu, Q., Huang, W., Li, X. and Wang, Y. 2018. Current status on the development of pseudoviruses for enveloped viruses. *Reviews in medical virology*. **28**(1).
- Li, S. 2022. Cryo-electron tomography of enveloped viruses. *Trends in Biochemical Sciences*. **47**(2), pp.173–186.
- Li, Y., Tenchov, R., Smoot, J., Liu, C., Watkins, S. and Zhou, Q. 2021. A Comprehensive Review of the Global Efforts on COVID-19 Vaccine Development. *ACS Central Science*. **7**(4), pp.512–533.
- Lieu, K.G., Marsh, G.A., Wang, L.-F. and Netter, H.J. 2015. The non-pathogenic Henipavirus Cedar paramyxovirus phosphoprotein has a compromised ability to target STAT1 and STAT2. *Antiviral research*. **124**, pp.69–76.
- Lo, M.K., Bird, B.H., Chattopadhyay, A., Drew, C.P., Martin, B.E., Coleman, J.D., Rose, J.K., Nichol, S.T. and Spiropoulou, C.F. 2014. Single-dose replication-defective VSV-based Nipah virus vaccines provide protection from lethal challenge in Syrian hamsters. *Antiviral research*. **101**, pp.26–29.
- Lo, M.K., Jordan, R., Arvey, A., Sudhamsu, J., Shrivastava-Ranjan, P., Hotard, A.L., Flint, M., McMullan, L.K., Siegel, D., Clarke, M.O., Mackman, R.L., Hui, H.C., Perron, M., Ray, A.S., Cihlar, T., Nichol, S.T. and Spiropoulou, C.F. 2017. GS-5734 and its parent nucleoside analog inhibit Filo-, Pneumo-, and Paramyxoviruses. *Scientific reports*. **7**, p.43395.
- Loney, C., Mottet-Osman, G., Roux, L. and Bhella, D. 2009. Paramyxovirus ultrastructure and genome packaging: cryo-electron tomography of sendai virus. *Journal of virology*. **83**(16), pp.8191–8197.
- Lustig, A. and Levine, A.J. 1992. One hundred years of virology. *Journal of virology*. **66**(8), pp.4629–4631.
- Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C. and Mijakovic, I.

2019. Protein post-translational modifications in bacteria. *Nature Reviews Microbiology*. **17**(11), pp.651–664.
- Madera, S., Kistler, A., Ranaivoson, H.C., Ahyong, V., Andrianaina, A., Andry, S., Raharinosy, V., Randriambolamanantsoa, T.H., Ravelomanantsoa, N.A.F., Tato, C.M., DeRisi, J.L., Aguilar, H.C., Lacoste, V., Dussart, P., Heraud, J.-M. and Brook, C.E. 2022. Discovery and Genomic Characterization of a Novel Henipavirus, Angavokely Virus, from Fruit Bats in Madagascar. *Journal of virology*, p.e0092122.
- Marsh, G.A., de Jong, C., Barr, J.A., Tachedjian, M., Smith, C., Middleton, D., Yu, M., Todd, S., Foord, A.J., Haring, V., Payne, J., Robinson, R., Broz, I., Cramer, G., Field, H.E. and Wang, L.-F. 2012. Cedar virus: a novel Henipavirus isolated from Australian bats. *PLoS pathogens*. **8**(8), p.e1002836.
- Marsian, J., Hurdiss, D.L., Ranson, N.A., Ritala, A., Paley, R., Cano, I. and Lomonossoff, G.P. 2019. Plant-made nervous necrosis virus-like particles protect fish against disease. *Frontiers in Plant Science*. **10**(July), pp.1–11.
- Matsumoto, Y., Ohta, K., Kolakofsky, D. and Nishio, M. 2018. The control of paramyxovirus genome hexamer length and mRNA editing. *RNA (New York, N.Y.)*. **24**(4), pp.461–467.
- McMullan, G., Chen, S., Henderson, R. and Faruqi, A.R. 2009. Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy*. **109**(9), pp.1126–1143.
- McMullan, G., Faruqi, A.R. and Henderson, R. 2016. Direct Electron Detectors *In*: R. A. B. T.-M. in E. Crowther, ed. *Methods in Enzymology* [Online]. Academic Press, pp.1–17. Available from: <https://www.sciencedirect.com/science/article/pii/S0076687916300921>.
- Mega, D.F., Fuller, J., Álvarez-Rodríguez, B., Mankouri, J., Hewson, R. and Barr, J.N. 2020. Mutagenic Analysis of Hazara Nairovirus Nontranslated Regions during Single- and Multistep Growth Identifies both Attenuating and Functionally Critical Sequences for Virus Replication. *Journal of virology*. **94**(17).
- Meulendyke, K.A., Wurth, M.A., McCann, R.O. and Dutch, R.E. 2005. Endocytosis Plays a Critical Role in Proteolytic Processing of the Hendra Virus Fusion Protein. *Journal of Virology*. **79**(20), pp.12643–12649.
- Michalski, W.P., Cramer, G., Wang, L.F., Shiell, B.J. and Eaton, B. 2000. The cleavage activation and sites of glycosylation in the fusion protein of Hendra virus. *Virus Research*. **69**(2), pp.83–93.
- Middleton, D., Pallister, J., Klein, R., Feng, Y.R., Haining, J., Arkinstall, R., Frazer, L., Huang, J.A., Edwards, N., Wareing, M., Elhay, M., Hashmi, Z., Bingham, J., Yamada, M., Johnson, D., White, J., Foord, A., Heine, H.G., Marsh, G.A., Broder, C.C. and Wang, L.F. 2014. Hendra virus vaccine, a one health approach to protecting horse, human, and environmental health. *Emerging Infectious Diseases*. **20**(3), pp.372–379.
- Milles, S., Jensen, M.R., Communie, G., Maurin, D., Schoehn, G., Ruigrok, R.W.H. and Blackledge, M. 2016. Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence Dependence. *Angewandte Chemie (International ed. in English)*. **55**(32), pp.9356–9360.
- Milne, J.L.S., Borgnia, M.J., Bartesaghi, A., Tran, E.E.H., Earl, L.A., Schauder, D.M., Lengyel, J., Pierson, J., Patwardhan, A. and Subramaniam, S. 2013. Cryo-electron microscopy - A primer for the non-microscopist. *FEBS Journal*. **280**(1), pp.28–45.
- Mire, C.E., Geisbert, J.B., Agans, K.N., Feng, Y.-R., Fenton, K.A., Bossart, K.N., Yan, L., Chan, Y.-

- P., Broder, C.C. and Geisbert, T.W. 2014. A recombinant Hendra virus G glycoprotein subunit vaccine protects nonhuman primates against Hendra virus challenge. *Journal of virology*. **88**(9), pp.4624–4631.
- Mire, C.E., Versteeg, K.M., Cross, R.W., Agans, K.N., Fenton, K.A., Whitt, M.A. and Geisbert, T.W. 2013. Single injection recombinant vesicular stomatitis virus vaccines protect ferrets against lethal Nipah virus disease. *Virology journal*. **10**, p.353.
- Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B. and Wilson, I.A. 2012. Organization of the influenza virus replication machinery. *Science (New York, N.Y.)*. **338**(6114), pp.1631–1634.
- Monaghan, P., Green, D., Pallister, J., Klein, R., White, J., Williams, C., McMillan, P., Tilley, L., Lampe, M., Hawes, P. and Wang, L.-F. 2014. Detailed morphological characterisation of Hendra virus infection of different cell types using super-resolution and conventional imaging. *Virology journal*. **11**, p.200.
- Murray, K., Selleck, P., Hooper, P., Hyatt, A., Gould, A., Gleeson, L., Westbury, H., Hiley, L., Selvey, L., Rodwell, B. and Ketterer, P. 1995. A morbillivirus that caused fatal disease in horses and humans. *Science*. **268**(5207), pp.94–97.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography*. **67**(Pt 4), pp.355–367.
- Myers, T.M. and Moyer, S.A. 1997. An amino-terminal domain of the Sendai virus nucleocapsid protein is required for template function in viral RNA synthesis. *Journal of virology*. **71**(2), pp.918–924.
- Navaratnarajah, C.K., Generous, A.R., Yousaf, I. and Cattaneo, R. 2020. Receptor-mediated cell entry of paramyxoviruses: Mechanisms, and consequences for tropism and pathogenesis. *Journal of Biological Chemistry*. **295**(9), pp.2771–2786.
- Negrete, O.A., Levroney, E.L., Aguilar, H.C., Bertolotti-Ciarlet, A., Nazarian, R., Tajyar, S. and Lee, B. 2005. EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature*. **436**(7049), pp.401–405.
- Negrete, O.A., Wolf, M.C., Aguilar, H.C., Enterlein, S., Wang, W., Mühlberger, E., Su, S. V, Bertolotti-Ciarlet, A., Flick, R. and Lee, B. 2006. Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. *PLoS pathogens*. **2**(2), p.e7.
- Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., Liu, J., Chan, P.K.-S., Li, S.-M., Chan, W.-Y., Au, S.W.-N., Joachimiak, A., Walz, T., Wang, J.-H. and Shaw, P.-C. 2008. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **22**(10), pp.3638–3647.
- Nogales, E. and Scheres, S.H.W. 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular cell*. **58**(4), pp.677–689.
- Norris, M.J., Husby, M.L., Kiosses, W.B., Yin, J., Saxena, R., Rennick, L.J., Heiner, A., Harkins, S.S., Pokhrel, R., Schendel, S.L., Hastie, K.M., Landeras-Bueno, S., Salie, Z.L., Lee, B., Chapagain, P.P., Maisner, A., Duprex, W.P., Stahelin, R. V and Saphire, E.O. 2022. Measles and Nipah virus assembly: Specific lipid binding drives matrix polymerization. *Science advances*. **8**(29), p.eabn1440.

- Noton, S.L., Deflubé, L.R., Tremaglio, C.Z. and Fearn, R. 2012. The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. *PLoS pathogens*. **8**(10), p.e1002980.
- O'Sullivan, J.D., Allworth, A.M., Paterson, D.L., Snow, T.M., Boots, R., Gleeson, L.J., Gould, A.R., Hyatt, A.D. and Bradfield, J. 1997. Fatal encephalitis due to novel paramyxovirus transmitted from horses. *Lancet*. **349**(9045), pp.93–95.
- Ohi, M., Li, Y., Cheng, Y. and Walz, T. 2004. Negative staining and image classification - Powerful tools in modern electron microscopy. *Biological Procedures Online*. **6**(1), pp.23–34.
- Orlova, E. V. and Saibil, H.R. 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical Reviews*. **111**(12), pp.7710–7748.
- Oxford, J., Kellam, P. and Collier, L. 2016. Virology: how it all began and where it will go next *In: Human Virology*. Oxford: Oxford University Press, pp.4–7.
- Oxford, J.S., Lambkin, R., Sefton, A., Daniels, R., Elliot, A., Brown, R. and Gill, D. 2005. A hypothesis: The conjunction of soldiers, gas, pigs, ducks, geese and horses in Northern France during the Great War provided the conditions for the emergence of the 'Spanish' influenza pandemic of 1918-1919. *Vaccine*. **23**(7), pp.940–945.
- Pager, C.T. and Dutch, R.E. 2005. Cathepsin L Is Involved in Proteolytic Processing of the Hendra Virus Fusion Protein. *Journal of Virology*. **79**(20), pp.12714–12720.
- Pager, C.T., Wurth, M.A. and Dutch, R.E. 2004. Subcellular Localization and Calcium and pH Requirements for Proteolytic Processing of the Hendra Virus Fusion Protein. *Journal of Virology*. **78**(17), pp.9154–9163.
- Park, A., Yun, T., Vigant, F., Pernet, O., Won, S.T., Dawes, B.E., Bartkowski, W., Freiberg, A.N. and Lee, B. 2016. Nipah Virus C Protein Recruits Tsg101 to Promote the Efficient Release of Virus in an ESCRT-Dependent Pathway. *PLoS pathogens*. **12**(5), p.e1005659.
- Park, M.-S., Shaw, M.L., Muñoz-Jordan, J., Cros, J.F., Nakaya, T., Bouvier, N., Palese, P., García-Sastre, A. and Basler, C.F. 2003. Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. *Journal of virology*. **77**(2), pp.1501–1511.
- Parks, G.D. and Alexander-Miller, M.A. 2013. Paramyxovirus activation and inhibition of innate immune responses. *Journal of molecular biology*. **425**(24), pp.4872–4892.
- Passmore, L.A. and Russo, C.J. 2016. Specimen preparation for high-resolution cryo-EM. *Methods in Enzymology*. (579), pp.51–86.
- Patch, J.R., Cramer, G., Wang, L.-F., Eaton, B.T. and Broder, C.C. 2007. Quantitative analysis of Nipah virus proteins released as virus-like particles reveals central role for the matrix protein. *Virology journal*. **4**, p.1.
- Pearce, L.A., Yu, M., Waddington, L.J., Barr, J.A., Scoble, J.A., Cramer, G.S. and McKinstry, W.J. 2015. Structural characterization by transmission electron microscopy and immunoreactivity of recombinant Hendra virus nucleocapsid protein expressed and purified from *Escherichia coli*. *Protein Expression and Purification*. **116**, pp.19–29.
- Pentecost, M., Vashisht, A.A., Lester, T., Voros, T., Beaty, S.M., Park, A., Wang, Y.E., Yun, T.E., Freiberg, A.N., Wohlschlegel, J.A. and Lee, B. 2015. Evidence for ubiquitin-regulated nuclear and subnuclear trafficking among Paramyxovirinae matrix proteins. *PLoS*

pathogens. **11**(3), p.e1004739.

- Pernet, O., Schneider, B.S., Beaty, S.M., LeBreton, M., Yun, T.E., Park, A., Zachariah, T.T., Bowden, T.A., Hitchens, P., Ramirez, C.M., Daszak, P., Mazet, J., Freiberg, A.N., Wolfe, N.D. and Lee, B. 2014. Evidence for henipavirus spillover into human populations in Africa. *Nature Communications*. **5**(1), p.5342.
- Peter, B. 2004. Negative-stain spreading of protein-protein and protein-lipid complexes for EM. [Accessed 4 July 2022]. Available from: [https://www2.mrc-lmb.cam.ac.uk/groups/hmm/techniqs/Negative stain protocol.html](https://www2.mrc-lmb.cam.ac.uk/groups/hmm/techniqs/Negative%20stain%20protocol.html).
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. **25**(13), pp.1605–1612.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. and Ferrin, T.E. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*. **30**(1), pp.70–82.
- Playford, E.G., McCall, B., Smith, G., Slinko, V., Allen, G., Smith, I., Moore, F., Taylor, C., Kung, Y.H. and Field, H. 2010. Human Hendra virus encephalitis associated with equine outbreak, Australia, 2008. *Emerging infectious diseases*. **16**(2), pp.219–223.
- Playford, E.G., Munro, T., Mahler, S.M., Elliott, S., Gerometta, M., Hoger, K.L., Jones, M.L., Griffin, P., Lynch, K.D., Carroll, H., El Saadi, D., Gilmour, M.E., Hughes, B., Hughes, K., Huang, E., de Bakker, C., Klein, R., Scher, M.G., Smith, I.L., Wang, L.-F., Lambert, S.B., Dimitrov, D.S., Gray, P.P. and Broder, C.C. 2020. Safety, tolerability, pharmacokinetics, and immunogenicity of a human monoclonal antibody targeting the G glycoprotein of henipaviruses in healthy adults: a first-in-human, randomised, controlled, phase 1 study. *The Lancet. Infectious diseases*. **20**(4), pp.445–454.
- Plempner, R.K. and Lamb, R.A. 2021. Paramyxoviridae: The Viruses and Their Replication *In: Fields Virology Volume 1: Emerging Viruses*. Philadelphia: Wolters Kluwer, pp.504–549.
- Ploquin, A., Szécsi, J., Mathieu, C., Guillaume, V., Barateau, V., Ong, K.C., Wong, K.T., Cosset, F.-L., Horvat, B. and Salvetti, A. 2013. Protection against henipavirus infection by use of recombinant adeno-associated virus-vector vaccines. *The Journal of infectious diseases*. **207**(3), pp.469–478.
- Plotkin, J.B. and Kudla, G. 2011. Synonymous but not the same. *National Review of Genetics*. **12**(1), pp.32–42.
- Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L. and Lloyd-Smith, J.O. 2017. Pathways to zoonotic spillover. *Nature Reviews Microbiology*. **15**(8), pp.502–510.
- Popa, A., Carter, J.R., Smith, S.E., Hellman, L., Fried, M.G. and Dutch, R.E. 2012. Residues in the hendra virus fusion protein transmembrane domain are critical for endocytic recycling. *Journal of virology*. **86**(6), pp.3014–3026.
- Porotto, M., Orefice, G., Yokoyama, C.C., Mungall, B.A., Realubit, R., Sganga, M.L., Aljofan, M., Whitt, M., Glickman, F. and Moscona, A. 2009. Simulating henipavirus multicycle replication in a screening assay leads to identification of a promising candidate for therapy. *Journal of virology*. **83**(10), pp.5148–5155.
- Procházková, M., Füzik, T., Grybchuk, D., Falginella, F.L., Podešvová, L., Yurchenko, V., Vácha, R. and Plevka, P. 2021. Capsid Structure of Leishmania RNA Virus 1. *Journal of Virology*.

95(3).

- Protein Data Bank 2022. PDB Statistics. [Accessed 4 July 2022]. Available from: <https://www.rcsb.org/stats/>.
- Punch, E.K., Hover, S., Blest, H.T.W., Fuller, J., Hewson, R., Fontana, J., Mankouri, J. and Barr, J.N. 2018. Potassium is a trigger for conformational change in the fusion spike of an enveloped RNA virus. *Journal of Biological Chemistry*. **293**(26), pp.9937–9944.
- Queensland Government 2022. Summary of Hendra virus incidents in horses. *Business Queensland Website*. [Online]. [Accessed 14 July 2022]. Available from: <https://www.business.qld.gov.au/industries/service-industries-professionals/service-industries/veterinary-surgeons/guidelines-hendra/incident-summary>.
- Quemin, E.R.J., MacHala, E.A., Vollmer, B., Pražák, V., Vasishtan, D., Rosch, R., Grange, M., Franken, L.E., Baker, L.A. and Grünewald, K. 2020. Cellular Electron Cryo-Tomography to Study Virus-Host Interactions. *Annual Review of Virology*. **7**, pp.239–262.
- Rager, M., Vongpunsawad, S., Duprex, W.P. and Cattaneo, R. 2002. Polyploid measles virus with hexameric genome length. *The EMBO journal*. **21**(10), pp.2364–2372.
- Rawlinson, S.M., Zhao, T., Rozario, A.M., Rootes, C.L., McMillan, P.J., Purcell, A.W., Woon, A., Marsh, G.A., Lieu, K.G., Wang, L.-F., Netter, H.J., Bell, T.D.M., Stewart, C.R. and Moseley, G.W. 2018. Viral regulation of host cell biology by hijacking of the nucleolar DNA-damage response. *Nature communications*. **9**(1), p.3057.
- Raymond, D.D., Piper, M.E., Gerrard, S.R., Skiniotis, G. and Smith, J.L. 2012. Phleboviruses encapsidate their genomes by sequestering RNA bases. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(47), pp.19208–19213.
- Raymond, D.D., Piper, M.E., Gerrard, S.R. and Smith, J.L. 2010. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(26), pp.11769–11774.
- Re, A., Joshi, T., Kulberkyte, E., Morris, Q. and Workman, C.T. 2014. RNA-protein interactions: an overview. *Methods in molecular biology (Clifton, N.J.)*. **1097**, pp.491–521.
- Reguera, J., Cusack, S. and Kolakofsky, D. 2014. Segmented negative strand RNA virus nucleoprotein structure. *Current Opinion in Virology*. **5**(1), pp.7–15.
- Reguera, J., Malet, H., Weber, F. and Cusack, S. 2013. Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(18), pp.7246–7251.
- Riedel, C., Hennrich, A.A. and Conzelmann, K.-K. 2020. Components and Architecture of the Rhabdovirus Ribonucleoprotein Complex. *Viruses*. **12**(9).
- Rima, B., Balkema-Buschmann, A., Dundon, W.G., Duprex, P., Easton, A., Fouchier, R., Kurath, G., Lamb, R., Lee, B., Rota, P., Wang, L. and Ictv Report Consortium 2019. ICTV Virus Taxonomy Profile: Paramyxoviridae. *The Journal of general virology*. **100**(12), pp.1593–1594.
- Rincheval, V., Lelek, M., Gault, E., Bouillier, C., Sitterlin, D., Blouquit-Laye, S., Galloux, M., Zimmer, C., Eleouet, J.-F. and Rameix-Welti, M.-A. 2017. Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nature Communications*. **8**(1), p.563.

- Roden, C. and Gladfelter, A.S. 2021. RNA contributions to the form and function of biomolecular condensates. *Nature Reviews Molecular Cell Biology*. **22**(3), pp.183–195.
- Rodriguez, J.J., Wang, L.-F. and Horvath, C.M. 2003. Hendra Virus V Protein Inhibits Interferon Signaling by Preventing STAT1 and STAT2 Nuclear Accumulation. *Journal of Virology*. **77**(21), pp.11842–11845.
- Rohou, A. and Grigorieff, N. 2015. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. **192**(2), pp.216–221.
- Rosano, G.L. and Ceccarelli, E.A. 2014. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*. **5**(APR), pp.1–17.
- Rosenthal, P.B. and Henderson, R. 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology*. **333**(4), pp.721–745.
- Ruigrok, R.W.H., Crépin, T., Hart, D.J. and Cusack, S. 2010. Towards an atomic resolution understanding of the influenza virus replication machinery. *Current opinion in structural biology*. **20**(1), pp.104–113.
- Ruigrok, R.W.H., Crépin, T. and Kolakofsky, D. 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Current Opinion in Microbiology*. **14**(4), pp.504–510.
- Russo, C.J. and Passmore, L.A. 2016. Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *Journal of Structural Biology*. **193**(1), pp.33–44.
- Sanjuán, R. and Domingo-Calap, P. 2016. Mechanisms of viral mutation. *Cellular and molecular life sciences : CMLS*. **73**(23), pp.4433–4448.
- Satharasinghe, D.A., Murulitharan, K., Tan, S.W., Yeap, S.K., Munir, M., Ideris, A. and Omar, A.R. 2016. Detection of Inter-Lineage Natural Recombination in Avian Paramyxovirus Serotype 1 Using Simplified Deep Sequencing Platform. *Frontiers in microbiology*. **7**, p.1907.
- Scarff, C.A., Fuller, M.J.G., Thompson, R.F. and Iadaza, M.G. 2018. Variations on negative stain electron microscopy methods: Tools for tackling challenging systems. *Journal of Visualized Experiments*. **2018**(132), pp.1–8.
- Scheid, A. and Choppin, P.W. 1974. Identification of biological activities of paramyxovirus glycoproteins. Activation of cell fusion, hemolysis, and infectivity by proteolytic cleavage of an inactive precursor protein of Sendai virus. *Virology*. **57**(2), pp.475–490.
- Scheres, S.H.W. 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods in enzymology*. **579**, pp.125–157.
- Scheres, S.H.W. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. **180**(3), pp.519–530.
- Scheres, S.H.W. and Chen, S. 2012. Prevention of overfitting in cryo-EM structure determination. *Nature methods*. **9**(9), pp.853–854.
- Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A. and Ruigrok, R.W.H. 2004. The 12 Å structure of trypsin-treated measles virus N-RNA. *Journal of Molecular Biology*. **339**(2), pp.301–312.
- Selvey, L.A., Wells, R.M., McCormack, J.G., Ansford, A.J., Murray, K., Rogers, R.J., Lavercombe,

- P.S., Selleck, P. and Sheridan, J.W. 1995. Infection of humans and horses by a newly described morbillivirus. *Medical Journal of Australia*. **162**(12), pp.642–645.
- Severin, C., Terrell, J.R., Zengel, J.R., Cox, R., Plemper, R.K., He, B. and Luo, M. 2016. Releasing the Genomic RNA Sequestered in the Mumps Virus Nucleocapsid. *Journal of Virology*. **90**(22), pp.10113–10119.
- Sharp, P.M., Stenico, M., Peden, J.F. and Lloyd, A.T. 1993. Codon usage: Mutational bias, translational selection, or both? *Biochemical Society Transactions*. **21**(4), pp.835–841.
- Shaw, M.L., Cardenas, W.B., Zamarin, D., Palese, P. and Basler, C.F. 2005. Nuclear localization of the Nipah virus W protein allows for inhibition of both virus- and toll-like receptor 3-triggered signaling pathways. *Journal of virology*. **79**(10), pp.6078–6088.
- Shi, J., Hu, S., Liu, X., Yang, J., Liu, D., Wu, L., Wang, H., Hu, Z., Deng, F. and Shen, S. 2017. Migration, recombination, and reassortment are involved in the evolution of severe fever with thrombocytopenia syndrome bunyavirus. *Infection, Genetics and Evolution*. **47**, pp.109–117.
- Si, Z., Zhou, K., Tsao, J., Luo, M. and Zhou, Z.H. 2022. Locations and in situ structure of the polymerase complex inside the virion of vesicular stomatitis virus. *Proceedings of the National Academy of Sciences of the United States of America*. **119**(18), p.e2111948119.
- Sigworth, F.J. 2016. Principles of cryo-EM single-particle image processing. *Microscopy (Oxford, England)*. **65**(1), pp.57–67.
- Smith, E.C., Smith, S.E., Carter, J.R., Webb, S.R., Gibson, K.M., Hellman, L.M., Fried, M.G. and Dutch, R.E. 2013. Trimeric transmembrane domain interactions in paramyxovirus fusion proteins: roles in protein folding, stability, and function. *The Journal of biological chemistry*. **288**(50), pp.35726–35735.
- 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 Biology*. **18**(3), pp.1–23.
- Song, X., Shan, H., Zhu, Y., Hu, S., Xue, L., Chen, Y., Ding, W., Niu, T., Gu, J., Ouyang, S., Shen, Q.T. and Liu, Z.J. 2019. Self-capping of nucleoprotein filaments protects the newcastle disease virus genome. *eLife*. **8**, pp.1–19.
- Spann, K.M., Collins, P.L. and Teng, M.N. 2003. Genetic Recombination during Coinfection of Two Mutants of Human Respiratory Syncytial Virus. *Journal of Virology*. **77**(20), pp.11201–11211.
- Spencer, P.S. and Barral, J.M. 2012. Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal*. **1**(1), p.e201204006.
- Stoker, M. and Macpherson, I. 1964. Syrian Hamster Fibroblast Cell Line BHK21 and its Derivatives. *Nature*. **203**(4952), pp.1355–1357.
- Sugai, A., Sato, H., Takayama, I., Yoneda, M. and Kai, C. 2017. Nipah and Hendra Virus Nucleoproteins Inhibit Nuclear Accumulation of Signal Transducer and Activator of Transcription 1 (STAT1) and STAT2 by Interfering with Their Complex Formation. *Journal of virology*. **91**(21).
- Surtees, R., Ariza, A., Punch, E.K., Trinh, C.H., Dowall, S.D., Hewson, R., Hiscox, J.A., Barr, J.N. and Edwards, T.A. 2015. The crystal structure of the Hazara virus nucleocapsid protein. *BMC Structural Biology*. **15**(1), pp.1–13.

- Tamin, A., Harcourt, B.H., Ksiazek, T.G., Rollin, P.E., Bellini, W.J. and Rota, P.A. 2002. Functional properties of the fusion and attachment glycoproteins of Nipah virus. *Virology*. **296**(1), pp.190–200.
- Tawar, R.G., Duquerroy, S., Vonrhein, C., Varela, P.F., Damier-Piolle, L., Castagné, N., MacLellan, K., Bedouelle, H., Bricogne, G., Bhella, D., Eléouët, J.-F. and Rey, F.A. 2009. Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus. *Science*. **326**(5957), pp.1279–1283.
- Taylor, L.H., Latham, S.M. and Woolhouse, M.E.J. 2001. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **356**(1411), pp.983–989.
- Thompson, R.F., Iadanza, M.G., Hesketh, E.L., Rawson, S. and Ranson, N.A. 2019. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nature Protocols*. **14**(1), pp.100–118.
- 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**(2016), pp.3–15.
- Tolkien, J.R.R. 1954. *The Fellowship of the Ring* 1st ed. London: George Allen & Unwin.
- Tremaglio, C.Z., Noton, S.L., Deflubé, L.R. and Fearn, R. 2013. Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. *Journal of virology*. **87**(6), pp.3196–3207.
- Turrell, L., Lyall, J.W., Tiley, L.S., Fodor, E. and Vreede, F.T. 2013. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nature communications*. **4**, p.1591.
- Twarock, R. and Luque, A. 2019. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nature Communications*. **10**(1), pp.1–9.
- unaids.org 2019. Global HIV & AIDS statistics — 2019 fact sheet | UNAIDS. [Http://www.unaids.org/en/resources/fact-sheet](http://www.unaids.org/en/resources/fact-sheet). [Online]. **1**(June), pp.1–6. Available from: <https://www.unaids.org/en/resources/fact-sheet>.
- Venkataraman, S., Prasad, B.V.L.S. and Selvarajan, R. 2018. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*. **10**(2).
- Vera, L., Czarny, B., Georgiadis, D., Dive, V. and Stura, E.A. 2011. Practical use of glycerol in protein crystallization published as part of the crystal growth & design virtual special issue on the 13th international conference on the crystallization of biological macromolecules (ICCBM13). *Crystal Growth and Design*. **11**(7), pp.2755–2762.
- Vulliémoz, D. and Roux, L. 2001. ‘Rule of six’: how does the Sendai virus RNA polymerase keep count? *Journal of virology*. **75**(10), pp.4506–4518.
- Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T.R., Lill, P., Gatsogiannis, C. and Raunser, S. 2019. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications biology*. **2**, p.218.
- Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Dempsey, D.M., Dutilh, B.E., Harrach, B., Harrison, R.L., Hendrickson, R.C., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M., Orton, R.J., Rubino, L., Sabanadzovic,

- S., Simmonds, P., Smith, D.B., Varsani, A., Zerbini, F.M. and Davison, A.J. 2020. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). *Archives of Virology*. **165**(11), pp.2737–2748.
- Wan, W. and Briggs, J.A.G. 2016. *Cryo-Electron Tomography and Subtomogram Averaging* [Online] 1st ed. Elsevier Inc. Available from: <http://dx.doi.org/10.1016/bs.mie.2016.04.014>.
- Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S. and Briggs, J.A.G. 2017. Structure and assembly of the Ebola virus nucleocapsid. *Nature*. **551**(7680), pp.394–397.
- Wang, L.-F., Yu, M., Hansson, E., Pritchard, L.I., Shiell, B., Michalski, W.P. and Eaton, B.T. 2000. The Exceptionally Large Genome of Hendra Virus: Support for Creation of a New Genus within the Family Paramyxoviridae. *Journal of Virology*. **74**(21), pp.9972–9979.
- Wang, L., Harcourt, B.H., Yu, M., Tamin, A., Rota, P.A., Bellini, W.J. and Eaton, B.T. 2001. Molecular biology of Hendra and Nipah viruses. *Microbes and infection*. **3**(4), pp.279–287.
- Wang, Y.E., Park, A., Lake, M., Pentecost, M., Torres, B., Yun, T.E., Wolf, M.C., Holbrook, M.R., Freiberg, A.N. and Lee, B. 2010. Ubiquitin-regulated nuclear-cytoplasmic trafficking of the Nipah virus matrix protein is important for viral budding. *PLoS pathogens*. **6**(11), p.e1001186.
- Wang, Z., Amaya, M., Addetia, A., Dang, H. V, Reggiano, G., Yan, L., Hickey, A.C., DiMaio, F., Broder, C.C. and Veessler, D. 2022. Architecture and antigenicity of the Nipah virus attachment glycoprotein. *Science (New York, N.Y.)*. **375**(6587), pp.1373–1378.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. **46**(W1), pp.W296–W303.
- Webster, R.G. 1999. 1918 Spanish influenza: The secrets remain elusive. *Proceedings of the National Academy of Sciences of the United States of America*. **96**(4), pp.1164–1166.
- Weise, C., Erbar, S., Lamp, B., Vogt, C., Diederich, S. and Maisner, A. 2010. Tyrosine residues in the cytoplasmic domains affect sorting and fusion activity of the Nipah virus glycoproteins in polarized epithelial cells. *Journal of virology*. **84**(15), pp.7634–7641.
- Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T.C.D., Howes, R.E., Tusting, L.S., Kang, S.Y., Cameron, E., Bisanzio, D., Battle, K.E., Bhatt, S. and Gething, P.W. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*. **553**(7688), pp.333–336.
- Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B. 3rd, Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S. and Richardson, D.C. 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein science : a publication of the Protein Society*. **27**(1), pp.293–315.
- Wilson, D.N. and Doudna Cate, J.H. 2012. The structure and function of the eukaryotic ribosome. *Cold Spring Harbor perspectives in biology*. **4**(5).
- de Wit, E. and Munster, V.J. 2015. Animal models of disease shed light on Nipah virus pathogenesis and transmission. *The Journal of pathology*. **235**(2), pp.196–205.

- Wong, J.J.W., Paterson, R.G., Lamb, R.A. and Jardetzky, T.S. 2016. Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form. *Proceedings of the National Academy of Sciences*. **113**(4), pp.1056–1061.
- Wong, K.T., Shieh, W.-J., Kumar, S., Norain, K., Abdullah, W., Guarner, J., Goldsmith, C.S., Chua, K.B., Lam, S.K., Tan, C.T., Goh, K.J., Chong, H.T., Jusoh, R., Rollin, P.E., Ksiazek, T.G. and Zaki, S.R. 2002. Nipah virus infection: pathology and pathogenesis of an emerging paramyxoviral zoonosis. *The American journal of pathology*. **161**(6), pp.2153–2167.
- World Health Organization 2008. Geographic distribution of Marburgh haemorrhagic fever outbreaks and fruit bats of Pteropodidae Family. Available from: https://www.who.int/images/default-source/health-topics/marburg-virus-disease/marburg-virus-disease-map-2009.png?sfvrsn=2718bf75_0.
- World Health Organization 1999. *Guidelines for Epidemic Preparedness and Response to Measles Outbreaks* [Online]. World Health Organization. Available from: <http://www.who.int/emc>.
- World Health Organization 2022. Prioritizing diseases for research and development in emergency contexts. [Accessed 15 December 2022]. Available from: <https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>.
- Wright, P.J., Cramer, G. and Eaton, B.T. 2005. RNA synthesis during infection by Hendra virus: an examination by quantitative real-time PCR of RNA accumulation, the effect of ribavirin and the attenuation of transcription. *Archives of virology*. **150**(3), pp.521–532.
- Wyatt, L.S., Moss, B. and Rozenblatt, S. 1995. Replication-deficient vaccinia virus encoding bacteriophage T7 RNA polymerase for transient gene expression in mammalian cells. *Virology*. **210**(1), pp.202–205.
- Yabukarski, F., Lawrence, P., Tarbouriech, N., Bourhis, J.M., Delaforge, E., Jensen, M.R., Ruigrok, R.W.H., Blackledge, M., Volchkov, V. and Jamin, M. 2014. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nature Structural and Molecular Biology*. **21**(9), pp.754–759.
- Yoneda, M., Georges-Courbot, M.-C., Ikeda, F., Ishii, M., Nagata, N., Jacquot, F., Raoul, H., Sato, H. and Kai, C. 2013. Recombinant measles virus vaccine expressing the Nipah virus glycoprotein protects against lethal Nipah virus challenge. *PloS one*. **8**(3), p.e58414.
- Yoneda, M., Guillaume, V., Ikeda, F., Sakuma, Y., Sato, H., Wild, T.F. and Kai, C. 2006. Establishment of a Nipah virus rescue system. *Proceedings of the National Academy of Sciences of the United States of America*. **103**(44), pp.16508–16513.
- Young, P.L., Halpin, K., Selleck, P.W., Field, H., Gravel, J.L., Kelly, M.A. and Mackenzie, J.S. 1996. Serologic Evidence for the presence in Pteropus Bats of a Paramyxovirus Related to Equine Morbillivirus. *Emerging Infectious Diseases*. **2**(3), pp.239–240.
- Yu, M., Hansson, E., Shiell, B., Michalski, W., Eaton, B.T. and Wang, L.F. 1998. Sequence analysis of the Hendra virus nucleoprotein gene: Comparison with other members of the subfamily Paramyxovirinae. *Journal of General Virology*. **79**(7), pp.1775–1780.
- Zhang, K. 2016. Gctf: Real-time CTF determination and correction. *Journal of structural biology*. **193**(1), pp.1–12.
- Zhang, N., Shan, H., Liu, M., Li, T., Luo, R., Yang, L., Qi, L., Chu, X., Su, X., Wang, R., Liu, Y., Sun, W. and Shen, Q.T. 2021. Structure and assembly of double-headed Sendai virus

- nucleocapsids. *Communications Biology*. **4**(1), pp.1–10.
- Zhang, X.-A., Li, H., Jiang, F.-C., Zhu, F., Zhang, Y.-F., Chen, J.-J., Tan, C.-W., Anderson, D.E., Fan, H., Dong, L.-Y., Li, C., Zhang, P.-H., Li, Y., Ding, H., Fang, L.-Q., Wang, L.-F. and Liu, W. 2022. A Zoonotic Henipavirus in Febrile Patients in China. *The New England journal of medicine*. **387**(5), pp.470–472.
- Zhang, X., Glendening, C., Linke, H., Parks, C.L., Brooks, C., Udem, S.A. and Oglesbee, M. 2002. Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. *Journal of virology*. **76**(17), pp.8737–8746.
- Zhang, X., Sridharan, S., Zagoriy, I., Oegema, C.E., Ching, C., Pflaesterer, T., Fung, H.K.H., Poser, I., Müller, C.W., Hyman, A.A., Savitski, M.M. and Mahamid, J. 2022. Molecular mechanisms of stress-induced reactivation in mumps virus condensates. *bioRxiv*, 2021.07.10.451879.
- 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. *Nature methods*. **14**(4), pp.331–332.
- Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E. and Scheres, S.H.W. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*. **7**, pp.1–38.