On the protein-nucleic acid interplay in dsRNA viruses



Serban Ilca

Division of Structural Biology

Nuffield Department of Medicine

University of Oxford

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy at the University of Oxford

Trinity term 2019

On the protein-nucleic acid interplay in dsRNA viruses

Serban Ilca, Magdalen College, University of Oxford
Doctor of Philosophy, Division of Structural Biology, Nuffield Department of Medicine
Trinity term 2019

Abstract

Viruses harbouring double-stranded RNA (dsRNA) genomes are highly diverse and their host range spans from bacteria to fungi, plants and mammals. dsRNA viruses share a number of characteristics including the possession of icosahedrally symmetric capsids which shelter their genomes. As many other icosahedral capsids, those of dsRNA viruses have been extensively characterised by cryogenic electron microscopy (cryo-EM). In contrast, *in situ* structures of genomes and genome-associated proteins such as RNA-dependent RNA polymerases (RdRPs) are scarce due to their characteristic asymmetry and high degree of heterogeneity. This thesis sets to expand the arsenal of image processing algorithms aimed at overcoming such hurdles and to resolve the architectures of capsid interiors in quiescent and mRNA-producing i.e. transcribing ϕ6 nucleocapsids (QNCs and TNCs), as well as rotavirus QNCs.

The ϕ6 genome is organised in spooled fashion, bearing strong resemblance to the canonical arrangement of viral dsDNA genomes. In ϕ6 QNCs 12,000 base pairs accounting for over 90% of the entire genome are modelled. ϕ6 TNCs depict the concerted rearrangements required to activate transcription, namely the outer capsid disassembly, inner capsid expansion and RdRP reorganisation. Rotavirus QNCs paint an entirely different picture as the RdRP-dsRNA interplay displays such extraordinary flexibility that thousands of unique capsid interior arrangements are made possible. Taken together, these results further the understanding of virus capsid interiors and highlight important disparities among dsRNA viruses.

Acknowledgements

I would first like to thank my supervisor Juha Huiskonen for not only supporting me in all endeavours, rendering all the FSC plots for publications and dealing with all EMDB map submissions, but also for sparking my interest in cryo-EM and image processing, which eventually convinced me to start a PhD. Thank you for being there to guide me on the day I first opened a Linux Terminal as a naive undergraduate, on the day I wrote my first line of code, and there to support me ever since. Particularly important were the evenings and nights we ordered take-away for dinner at work while trying to make sense of Euler angles and rotation matrices. Occasionally joining in the brainstorming (mostly to correct our nonsensical ideas) was Dave Stuart, whom I thank for many inspiring chats in the 5 years since and for additional project supervision.

None of my work on ϕ6 would have been possible without Xiaoyu Sun and Minna Poranen, our extraordinary collaborators from Helsinki. I would like to thank them for providing all the ϕ6 samples and for the plethora of insights into ϕ6, as well as viruses in general.

I am incredibly grateful to have fantastic friends who have made my life in Oxford considerably more enjoyable/bearable. First, thank you to Max Renner for his contagious passion for science, for being the kind of scientist I strive to become one day and for fantastically strange movie suggestions. Many thanks to Tim Sudmeier and Hamish Lemmey for getting me through this thesis and for the therapeutic visits to local pubs (special thanks to Tim for the invaluable chats on the way home after therapy). Thank you also to Nayab Malik, Rhys Pryce, Liz Allen, Yasa Watanabe, Sam Griffiths,

Steinar Halldorsson, Weng Ng and Vicky Avanzato for making work more worth going to and to Vicky also for the direly needed doses of Americanisms. Special thanks to Abhay Kotecha and Dimple Karia for befriending me from my early days in Oxford and for all the help ever since.

I also owe huge thanks to Annelie Muehler and Anja Đokić for their vital friendship and love, as well as occasional nights ending with just one or two beers on the sofa at 6am every time we see each other.

I would have never ended up here without my elder brother Tudor Ilca who was entirely responsible for me wanting to do well in school in the first place. By having been my fiercest competitor, Tudor pushed me all the way from picking up a chemistry book for the first time to attempting to become a scientist. Of course, this journey would have never happened without the constant love and support of my second family Gilbert Vansoi and Darren Roesch, my younger brother Dragos Ilca and my grandmother Cornelia Ilca. Most of all, I am indebted to my parents Rodica and Florin Ilca who are constantly redefining the limits of selflessness.

Lastly, I want to extend my appreciation to a truly one-of-a-kind person, Meg Hockman. Meeting you at that conference in Belgium is by far the best thing science has ever given me. Thank you for turning my life-contentment levels up a few notches and for being an absolutely terrific partner.

List of publications

Ilca SL, Sun X, El Omari K, Kotecha A, de Haas F, DiMaio F, Grimes JM, Stuart DI, Poranen MM, Huiskonen JT. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. *Nature*. 2019; 570:252-256.

De Colibus L, Roine E, Walter TS, Ilca SL, Wang X, Wang N, Roseman AM, Bamford D, Huiskonen JT, Stuart DI. Assembly of complex viruses exemplified by a halophilic euryarcheal virus. *Nat Commun*. 2019; 10:1456.

Sun X, Ilca SL, Huiskonen JT, Poranen MM. Dual role of a viral polymerase in viral genome replication and particle self-assembly. *MBio.* 2018; 9:01242-18.

Kotecha A, Perez-Martin E, Harvey Y, Zhang F, Ilca SL, Fry EE, Jackson B, Maree F, Scott K, Hecksel CW, Harmsen MM, Mioulet V, Wood B, Juleff N, Stuart DI, Charleston B, Seago J. Chimeric O1K foot-and-mouth disease virus with SAT2 outer capsid as an FMD vaccine candidate. *Sci Rep.* 2018; 8:13654.

Stass R*, Ilca SL*, Huiskonen JT. Beyond structures of highly symmetric purified viral capsids by cryo-EM. *Curr Opin Struct Biol.* 2018; 52:25-31.

Kotecha A, Wang Q, Dong X, Ilca SL, Ondiviela M, Zihe R, Seago J, Charleston B, Fry EE, Abrescia NGA, Springer TA, Huiskonen JT, Stuart DI. Rules of engagement between αvβ6 integrin and foot-and-mouth disease virus. *Nat Commun*. 2017; 8:15408.

Sun Z, El Omari K, Sun X, Ilca SL, Kotecha A, Stuart DI, Poranen MM, Huiskonen JT. Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. *Nat Commun*. 2017; 8:14814.

List of contents

List of contents.....7

List of figures 11

List of tables.....13

List of abbreviations.....14

1. Introduction 16

1.1. Replication cycle and structures of dsRNA viruses 16

1.1.1. General considerations 16

1.1.2. Cystoviridae.....16

1.1.3. Reoviridae.....22

1.2. Cryo-EM as main approach to resolve in situ structures.....26

1.2.1. Icosahedral symmetry aids capsid structure determination.....27

1.2.2. The dsRNA genome and RNA-processing proteins deviate from icosahedral symmetry 28

1.3. Study aim.....30

2. Materials & Methods 31

2.1. Sample preparation for cryo-EM.....31

2.1.1. ϕ6 quiescent nucleocapsids 31

2.1.2. ϕ6 transcribing nucleocapsids 31

2.1.3. Rotavirus quiescent nucleocapsids 32

2.2. Cryo-EM data collection 33

2.3. Single-particle image processing.....34

2.3.1. Implementation of symmetry relaxation 34
2.3.2. ϕ6 QNC.....35
2.3.2.1. Icosahedral reconstruction of the capsid.....35
2.3.2.2. Asymmetric reconstruction of the capsid interior.....36
2.3.2.3. Asymmetric reconstruction of the packaging motor P4.....38
2.3.3. ϕ6 TNC and DTNC 39
2.3.3.1. Icosahedral reconstruction of the capsid.....39
2.3.3.2. Asymmetric reconstruction of disassembly intermediates in DTNC.....41
2.3.3.3. Asymmetric reconstruction of the capsid interior.....42
2.3.4. RTV QNC 42
2.3.4.1. Icosahedral reconstruction of the capsid.....42
2.3.4.2. Combined localized reconstruction of the RdRP.....43
2.3.4.2. Asymmetric reconstruction of the RdRP-dsRNA complex 44
2.3.4.3. Individual localized reconstructions of the RdRP binding sites.....44
2.5. Model building and refinement 45
2.5.1. Fitting and refinement of capsid proteins 45
2.5.2. Fitting and refinement of other proteins 45
2.5.3. Fitting and refinement of dsRNA models.....46
2.6. Data analysis.....47
2.6.1. Identifying potential capsid-dsRNA contacts in QNC 47
2.6.2. Analysis of QNC dsRNA packing parameters.....48
2.6.3. Analysis of RTV RdRP distribution correlation 49
3. Atomic model of an encapsidated viral genome 51

3.1. Foreword 51
3.2. Summary 52
3.3. Structure of the protein shells 52
3.4. Organisation of the capsid interior 53
3.4.1. Breaking the capsid symmetry 53
3.4.2. Discerning locally distinct sub-conformations 55
3.4.3. Analysis of dsRNA packing parameters 56
3.4.4. Identification of capsid-dsRNA contacts 58
3.4.5. Resolving the RNA packaging motor P4 60
3.5. Discussion 61
4. Nucleocapsid restructuring induces ϕ6 transcription 73
4.1. Foreword 73
4.2. Summary 73
4.3. Structure of the protein shells 74
4.4. Structures of the P8 disassembly intermediates 75
4.5. Genome organisation 78
4.6. RdRP arrangement 81
4.7. Discussion 86
5. Limited yet discernible long range order inside rotavirus capsids 92
5.1. Foreword 92
5.2. Summary 93
5.3. Structure of the protein shells 93

5.4. Organisation of DLP interior 94
5.4.1. Breaking the capsid symmetry 94
5.4.2. Localized reconstruction of the 5-folds 96
5.4.3. VP1-dsRNA interactions 108
5.4.4. Analysis of VP1 orientations at neighbouring 5-folds 115
5.4.4.1. Correlations between individual classes 115
5.4.4.2. Correlations between individual classes and entire 5-folds 117
5.4.4.3. Correlations between entire 5-folds 118
5.5. Discussion 119
6. Conclusions and future directions 125
References 127
Appendix 147
1. Fourier Shell Correlation plots 147
1.1. ϕ6 QNC capsid with I1 and D3 symmetry (adapted from [130]) 147
1.2. ϕ6 QNC layer 1 with no symmetry imposed (adapted from [130]) 148
1.3. ϕ6 QNC layers 2, 3 and 4&5 after 3D refinement (no symmetry) 150
1.4. C6-symmetric and asymmetric P4 from ϕ6 QNC 152
1.5. DLP and SLP capsid from ϕ6 DTNC 153
1.6. Disassembly intermediates from ϕ6 DTNC (I1-9) 154
1.7. Peculiar intermediates from ϕ6 DTNC 159
1.8. DLP RdRP arrangements from ϕ6 DTNC 161
1.9. SLP RdRP arrangements from ϕ6 DTNC 162
2. ϕ6 P1 residue distribution in dsRNA vicinity 163

List of figures

Figure 1.1. Architecture of the ϕ6 virion 17

Figure 1.2. Overview of the ϕ6 replication cycle 18

Figure 1.3. Overview of the ϕ6 nucleocapsid assembly pathway 19

Figure 1.4. Architecture of the rotavirus virion 23

Figure 1.5. Overview of the rotavirus replication cycle 24

Figure 2.1. ϕ6 transcription reactions 32

Figure 3.1. Icosahedral reconstruction of ϕ6 QNC 53

Figure 3.2. Asymmetric reconstruction of the ϕ6 QNC genome 54

Figure 3.3. Locally distinct sub-conformations of the ϕ6 QNC genome 55

Figure 3.4. Ensemble model of the five dsRNA layers in ϕ6 QNC 57

Figure 3.5. dsRNA packing parameters in ϕ6 QNC 58

Figure 3.6. Putative capsid-dsRNA contacts in ϕ6 QNC 59

Figure 3.7. Structure of the P4 hexamer ϕ6 QNC 61

Figure 4.1. Icosahedral structures of the ϕ6 DTNC DLP and SLP 74

Figure 4.2. Comparison of the inner capsid shell in ϕ6 DLP and SLP 75

Figure 4.3. Structures of the outer shell disassembly intermediates in ϕ6 DTNC 77

Figure 4.4. Structures of peculiar P8 shell disassembly intermediates in ϕ6 DTNC 78

Figure 4.5. Structures of the DLP and SLP capsid interiors in ϕ6 DTNC 79

Figure 4.6. P4 structures in ϕ6 DTNC DLP and SLP 80

Figure 4.7. RdRP dominant arrangements in ϕ6 DTNC DLP and SLP 82

Figure 4.8. Structures of four RdRPs in ϕ6 DTNC 84

Figure 4.9. Active sites configurations of the same four RdRPs in ϕ6 DTNC 85

Figure 4.10. Proposed mechanism of transcription initiation in ϕ6 87

Figure 5.1. Pseudo-D5 organisation of capsid proximal dsRNA in RTV QNC 95

Figure 5.2. Summary of the 12 localized classifications in RTV QNC 99

Figure 5.3. VP1-dsRNA arrangements at 5-fold #1 in RTV QNC 101
Figure 5.4. VP1-dsRNA arrangements at 5-fold #2 in RTV QNC 101
Figure 5.5. VP1-dsRNA arrangements at 5-fold #3 in RTV QNC 102
Figure 5.6. VP1-dsRNA arrangements at 5-fold #4 in RTV QNC 102
Figure 5.7. VP1-dsRNA arrangements at 5-fold #5 in RTV QNC 103
Figure 5.8. VP1-dsRNA arrangements at 5-fold #6 in RTV QNC 104
Figure 5.9. VP1-dsRNA arrangements at 5-fold #7 in RTV QNC 104
Figure 5.10. VP1-dsRNA arrangements at 5-fold #8 in RTV QNC 104
Figure 5.11. VP1-dsRNA arrangements at 5-fold #9 in RTV QNC 105
Figure 5.12. VP1-dsRNA arrangements at 5-fold #10 in RTV QNC 106
Figure 5.13. VP1-dsRNA arrangements at 5-fold #11 in RTV QNC 106
Figure 5.14. VP1-dsRNA arrangements at 5-fold #12 in RTV QNC 107
Figure 5.15. VP1-dsRNA arrangement at #2.1 in RTV QNC.....110
Figure 5.16. VP1-dsRNA arrangement at #10.1 in RTV QNC.....111
Figure 5.17. VP1-dsRNA arrangement at #8.1 in RTV QNC.....112
Figure 5.18. VP1-dsRNA arrangement at #9.1 in RTV QNC.....113

List of tables

Table 1.1. Summary of ϕ6 and RTV protein functions.....20
Table 1.2. Summary of available ϕ6 structures 21
Table 1.3. Summary of available rotavirus structures 26
Table 2.1. Cryo-EM data acquisition parameters 33
Table 2.2. Summary of icosahedral refinement results.....34
Table 4.1. Summary of characteristics for the RdRP arrangements in ϕ6 DTNC 83
Table 5.1. Summary of the 12 localized classifications in RTV QNC 98
Table 5.2. Summary of the different types of map pairings in RTV QNC 116
Table 5.3. Most positively and negatively deviating map pairings in RTV QNC 116
Table 5.4. Most and least influencial classes in RTV QNC.....117
Table 5.5. Ranking of the 12 5-folds by Defining Influence in RTV QNC 118

List of abbreviations

ϕ6	bacteriophage phi6
(-)ve-sense	negative-sense
(+)ve sense	positive-sense
ARV	aquareovirus
ATP	adenosine triphosphate
BTV	bluetongue virus
CPV	cypovirus
cryo-EM	cryogenic electron-microscopy
CTP	cytosine triphosphate
DIP	disassembly-intermediate particle
DLP	double-layered particle
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DTNC	defectively transcribing nucleocapsid
EGTA	ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid
FSC	Fourier shell correlation
GTP	guanosine triphosphate
kbp	kilobase pair
kDa	kilodalton
MRV	mammalian orthoreovirs

NC	nucleocapsid
NTP	nucleoside triphosphate
NTPase	nucleoside triphosphatase
PC	polymerase complex / procapsid
QNC	quiescent nucleocapsid
RdRP	RNA-dependent RNA polymerase
RTV	rotavirus
SLP	single-layered particle
ssRNA	single-stranded RNA
TLP	triple-layered particle
TNC	transcribing nucleocapsid
UTP	uridine triphosphate
VC	virus core

1. Introduction

1.1. Replication cycle and structures of dsRNA viruses

1.1.1. General considerations

Viruses harbouring dsRNA genomes infect a wide range of hosts including bacteria, fungi, plants and animals [1] and their genomes consist of between one and twelve dsRNA segments [2]. dsRNA viruses share a remarkable number of features in terms of structure and replication cycle [1]. The genome is enclosed in the virus capsid which comprises one or more protein shells and is surrounded in some members by a lipid envelope [3]. Generally, the outer shell proteins or the proteins embedded in the lipid envelope, if present, are involved in host cell recognition and entry [4-6]. Once delivered into the cell, dsRNA viruses retain their genomes within their capsids and employ their own machinery to synthesise mRNA from within the viral cores [7]. The mRNA is translated by the host ribosomes creating new copies of the viral proteins. Immature cores are assembled from capsid proteins by packaging one positive-sense ssRNA copy of each genome segment [8, 9]. Intracapsid replication of these ssRNAs follows to form dsRNA, leading to viral cores which can either synthesise new mRNA or mature to form new infectious virions [10-12].

1.1.2. *Cystoviridae*

Cystoviridae is a dsRNA virus family consisting of species ϕ6-14 and ϕ2954 [13-15], which are the only dsRNA viruses that infect prokaryotes, namely gram-negative *Psedomonas* bacteria pathogenic to plants [13]. Cystoviruses have tri-segmented dsRNA genomes and are distinguished from other dsRNA viruses also in

that they contain a lipid envelope [16]. Nonetheless, the *Cystoviridae* show strong similarities to other dsRNA virus families and appear most closely related to *Reoviridae*, a large family of eukaryotic viruses [1]. Bacteriophage ϕ6 is the best-characterised cystovirus and serves as model organism in the study of dsRNA viruses. The lipid envelope of the ϕ6 virion (**Fig. 1.1**) surrounds a double-shelled capsid termed nucleocapsid (NC) harbouring the three dsRNA genome segments S, M and L [17-19]. Embedded in the envelope are proteins P3 and P6 which are used to bind type IV pili and to mediate membrane fusion, respectively (**Fig. 1.2**) [20-22]. P5 digests the peptidoglycan layer and allows the nucleocapsid to enter the cytoplasm [23, 24]. On entry the outer protein shell consisting of P8 dissociates leaving the transcriptionally active viral core (VC) made of the inner capsid protein P1, RdRP P2, RNA-packaging NTPase P4, assembly cofactor P7 and three dsRNA segments [10, 25, 26]. P2 uses the dsRNA template to synthesise mRNA semi-conservatively. This means that the newly synthesised (+)ve-sense ssRNA remains encapsidated by displacing the corresponding parental strand [27-32] which is extruded from the capsid with the aid of P4 [33, 34].

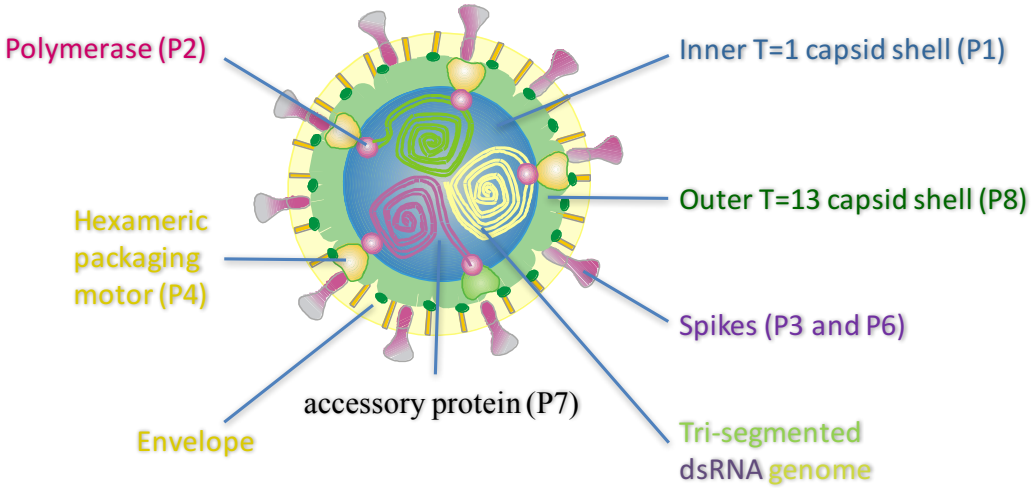


Figure 1.1. Architecture of the ϕ6 virion
Figure adapted from [3] with permission.

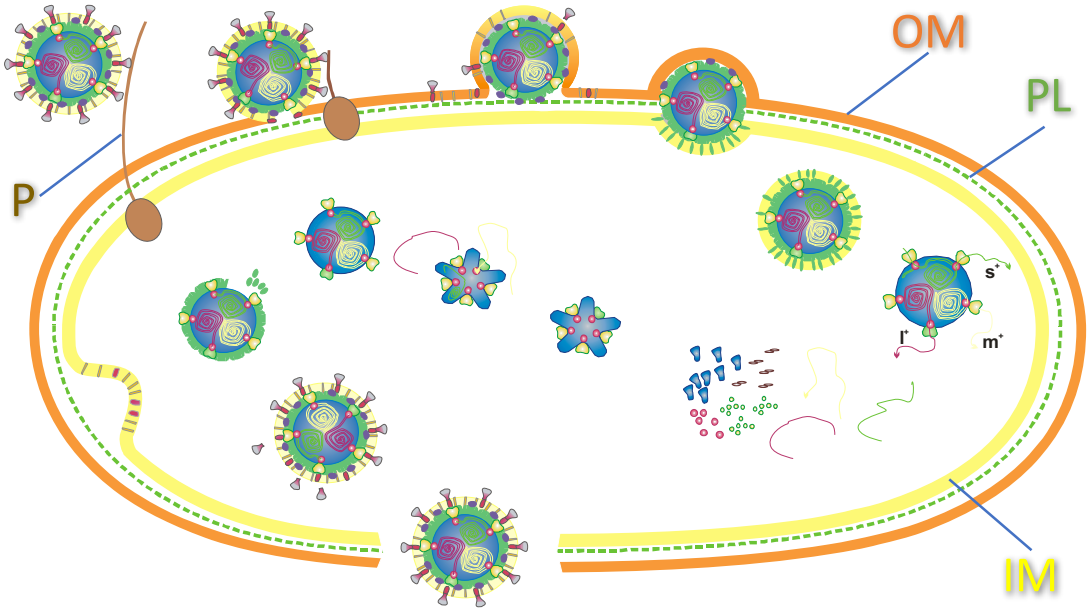


Figure 1.2. Overview of the ϕ6 replication cycle
After attachment to and retraction of a type IV pilus, the envelope fuses with the outer membrane. In the periplasm, P5 digests the peptidoglycan layer and P8 mediates entry into the cytosol. The VCs perform transcription and the newly synthesised viral proteins self-assemble into PCs. After genome packaging and replication, the P8 shell assembles and the envelope along with its associated proteins are recruited. A pilus (P), the outer membrane (OM), inner membrane (IM) and peptidoglycan layer (PL) are indicated. Figure taken with permission from [35]. Components are coloured as in **Fig. 1.1**.

Newly synthesised copies of P1, P2, P4 and P7 self-assemble into otherwise empty capsids lacking nucleic acid, termed procapsids or polymerase complexes (PC) [10, 25, 36]. The P1 shell of the originally unexpanded PC sequentially recognises the three genome segments in ssRNA form and progressively expands as the segments get packaged in the order s-m-l [8, 10, 13, 37, 38]. Packaging requires the P4 motor in a process fuelled by NTP hydrolysis [33, 34, 39]. Once the three segments are packaged, P2 performs genome replication using the (+)ve-sense ssRNAs as templates to produce dsRNAs [30, 32, 40, 41]. The core that is thus assembled is coated by P8 to form the NC [18, 19] and subsequently by the lipid envelope to form a new infectious virion [3].

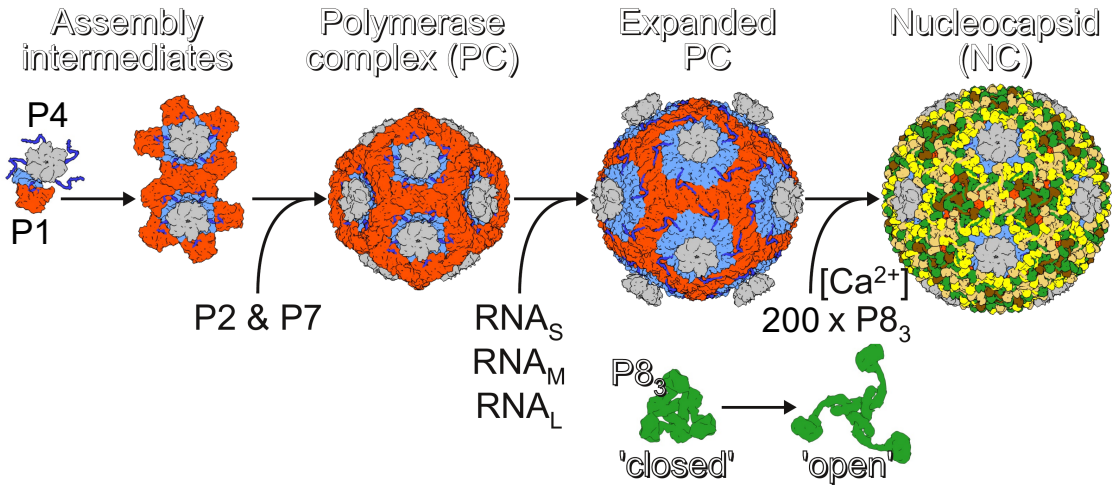


Figure 1.3. Overview of the ϕ6 nucleocapsid assembly pathway
Components of the nucleocapsid are coloured as follows: P1A orange red, P1B cornflower blue, P4 main body grey, P4 C-terminal domain dark blue, P8 trimers yellow, gold, dark green and brown. Figure adapted from [19].

An *in vitro* assembly system has been designed for ϕ6 [42, 43] which has played a key role in investigating the steps leading to the formation of the NC (**Fig. 1.3**), as well as the function and structure of the proteins involved (**Table 1.1** and **Table 1.2**). It has thus been shown that P4 hexamers nucleate the assembly of the inner capsid shell from asymmetric dimers of P1 [33]. P2 and P7 are not vital in this process though they both improve the assembly kinetics [36, 44]. Structures and biochemical analyses of complexes obtained by varying the amount of each of these proteins have provided a plethora of additional information. The T=1 dodecahedral shell is formed by 60 asymmetric dimers such that the P1A monomers surround the 5-fold symmetry axes while the P1B monomers form the regions corresponding to the 2- and 3-folds [18, 45]. Twelve P4 hexamers are located outside the capsid, mounted on the 5-folds and thus symmetry-mismatched [46, 47]. Eight to twelve copies of the monomeric P2 bind in the capsid interior at the 3-fold positions and contact two distinct P1 dimers [45, 47, 48]. An

as yet unknown number of P7 in monomeric form are thought to share the same binding site as P2 [49, 50], though ϕ12 P7 was proposed to form dimers [51]. Besides its role in PC assembly, P7 is thought to regulate the function of P2 [52].

The dodecahedral P1 shell of the PC displays heavily depressed 5-folds [25, 45]. As each of the three ssRNA segments gets encapsidated, the P1A copies progressively swing outwards thus leading to the expansion of the P1 shell [10, 13, 25]. After replication, the mature VC capsid appears roughly spherical with the P4 hexamers projecting away from the capsid [25]. The VC is then coated as 200 P8 trimers assemble by domain-swapping to form an incomplete T=13 icosahedral shell in which the P4 hexamers are embedded [18, 19]. Significant regulatory roles in NC assembly are played by divalent cations. Namely, P4 NTPase activity is inhibited by Mg^{2+} and accelerated by Ca^{2+} [39, 53]. Conversely, P2 polymerase activity is activated by Mg^{2+} and inhibited by Ca^{2+} [32, 54, 55]. In addition, the P8 shell requires Ca^{2+} for its formation and stability [19, 26].

Table 1.1. Summary of ϕ6 and RTV protein functions

Function	ϕ6 protein	RTV protein
inner T=1 capsid	P1	VP2
middle T=13 capsid	-	VP6
outer T=13 capsid	P8	VP7
spikes	P3 & P6	VP4 (VP5* & VP8*)
RdRP	P2	VP1
RNA packaging motor	P4	-
assembly cofactor	P7	-
capping enzyme	-	VP3
peptidoglycan digestion	P5	-

Table 1.2. Summary of available ϕ6 structures

✓ indicates the presence and ✖ the absence of the protein/dsRNA in the given state. The diameters of PC, VC and NC are shown in the header. The sign in the other parentheses indicates the status of the available structural information. (?) indicates unresolved with unidentified location, (✓) resolved only for ϕ12, (✓) unresolved with identified location, (✓) resolved at resolution poorer than 4 Å and (✓) resolved at resolution better than 4 Å.

	Size	State	PC (~45nm)	VC (~60 nm)	NC (~60 nm)	Purified
P1 (inner shell)	85 kDa	dimer	✓ (✓)	✓ (✓)	✓ (✓)	(?)
P2 (RdRP)	75 kDa	monomer	✓ (✓)	✓ (?)	✓ (?)	(✓)
P4 (RNA motor)	35 kDa	hexamer	✓ (✓)	✓ (✓)	✓ (✓)	(✓)
P7 (co-factor)	17 kDa	monomer	✓ (✓)	✓ (?)	✓ (?)	(✓)
P8 (outer shell)	16 kDa	trimer	✖	✖	✓ (✓)	(?)
dsRNA	2.9, 4.0 & 6.4 kbps	linear	✖	✓ (?)	✓ (✓)	-

Despite the plethora of functional and structural studies on ϕ6, key questions remain unanswered. Regarding the PC, the *in situ* structures of P4 and P7 are yet to be resolved and thus their precise roles in PC assembly are largely unknown. The segment recognition and encapsidation mechanisms, along with the structures of intermediates in the formation of VC have also eluded characterisation. No high-resolution structure of VC has been reported and while that of NC is available [19], it provides no insights into the location of P2 and P7 and very limited information on the structure of the dsRNA genome. Furthermore, the ways by which replication and transcription are activated and how they proceed upon activation are vastly unexplored. Finally, the lipid envelope and associated proteins remain the poorest characterised components, for which only low-resolution tomography data is available [56].

1.1.3. *Reoviridae*

The *Reoviridae* is the largest and most diverse family of dsRNA viruses. Members of the *Reoviridae* family (reoviruses) infect a large variety of eukaryotes including fungi, protists, plants and vertebrates [57]. Reoviruses lack a lipid envelope, have icosahedral capsids consisting of 1-3 shells and their genomes are divided into 9-12 segments [58]. They are classified into two sub-families: *Spinareovirinae* and *Sedoreovirinae* [58]. Members of *Sedoreovirinae* or "non-turreted" reoviruses are roughly spherical in appearance, whereas members of *Spinareovirinae* or "turreted" reoviruses contain additional extensions from the icosahedral 5-folds [58]. These projections contain the capping enzymes and the newly synthesised mRNA is extruded through the extensions such that capping can occur [11]. On the other hand, "non-turreted" reoviruses presumably have their capping enzymes within the capsid-RdRP-dsRNA mix and thus the capping mechanism is highly dissimilar [11].

Sedoreovirinae includes one of the best studied dsRNA virus genera: rotavirus. The rotavirus genus contains 9 species identified by letters A-I [58]. Species A-C are human pathogens with infections occurring via the faecal-oral route and being predominantly caused by species A. There are multiple strains of rotavirus A, one of which being simian rotavirus A SA11. For simplicity, this will be referred to as rotavirus in the remainder of this thesis as the concepts described should apply to the entire genus.

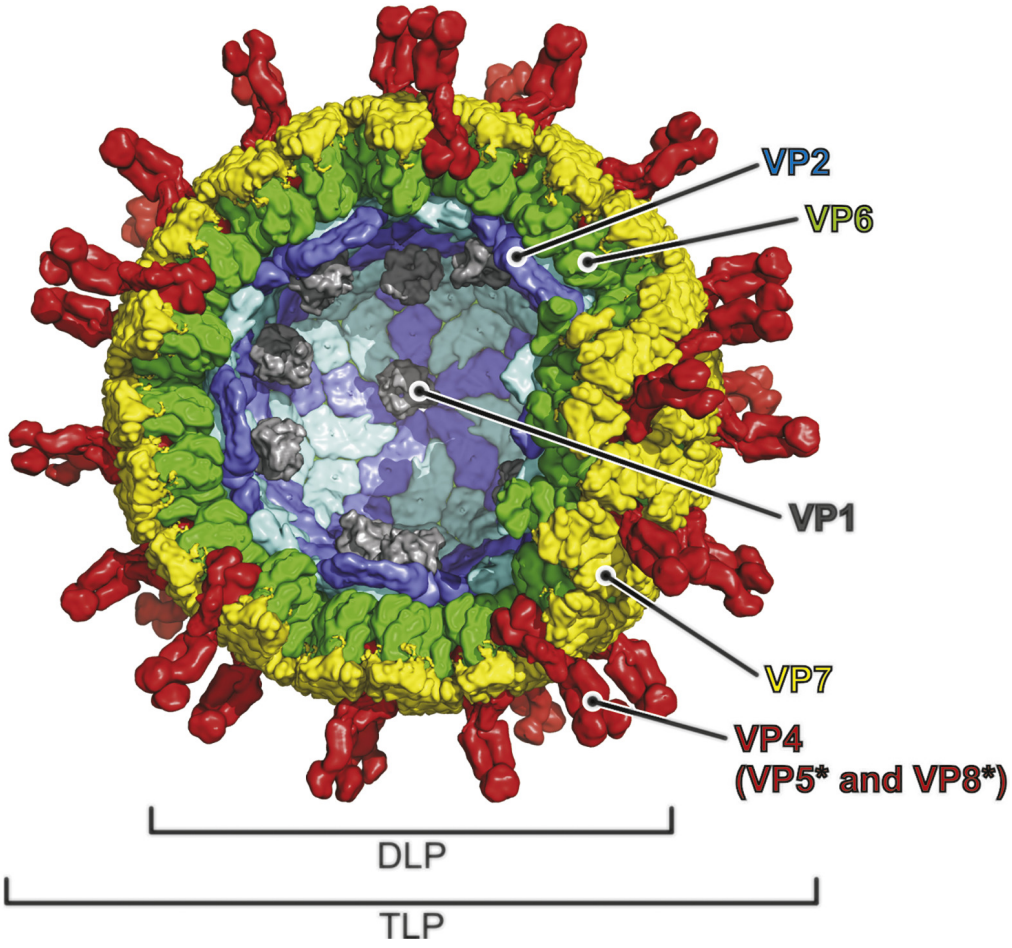


Figure 1.4. Architecture of the rotavirus virion
Missing are VP3 and the dsRNA genome both located in the DLP interior. Figure reproduced with permission from [59].

Rotavirus virions (**Fig. 1.4**) consist of a triple-shelled capsid harbouring 11 dsRNA genome segments [60]. Protein VP4 decorates the outer shell and is cleaved by trypsin-like proteases in the gut into VP5* and VP8*, rendering the particle infectious (**Fig. 1.5**) [61]. Located at the tip of the spikes, VP8* mediates attachment to the host cell via sialic acids leading to virus endocytosis [62]. In the early endosome, the outer capsid shell formed by VP7 dissociates [63, 64] and consequently the virion escapes the endosome using VP5* [65, 66]. The transcriptionally active double-layered particle (DLP) delivered into the cytoplasm consists of the VP6 intermediate shell and VP2 core

shell concealing the dsRNA genome, RdRP VP1 and capping enzyme VP3 [59, 67, 68]. Unlike ϕ6, rotavirus VP1 performs fully conservative transcription by which the newly synthesised (+)ve-sense ssRNAs are extruded from the capsid leaving both parental strands in the confines of the capsid [59, 68-70]. The transcripts acquire a 5' cap due to the activity of VP3 [71].

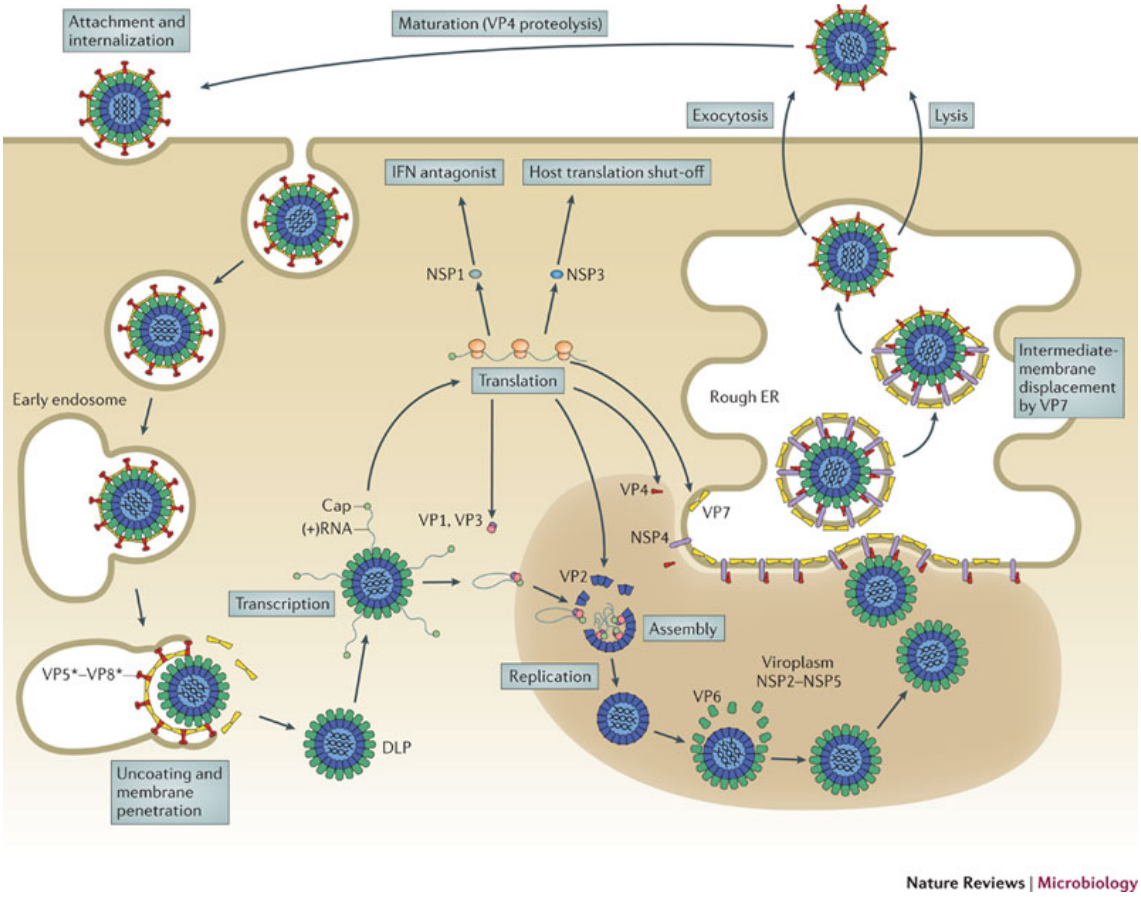


Figure 1.5. Overview of the rotavirus replication cycle
Steps involved in rotavirus replication from host cell attachment to synthesis of new viral proteins, assembly and maturation. Figure reproduced with permission from [11]. Components are coloured according to **Fig. 1.3** with the exception of VP1. Description of the replication cycle can be found in the text.

Though poorly understood compared to ϕ6, the assembly of the rotavirus VP2 shell along with VP1 and VP3 is thought to occur in tandem with packaging of the (+)ve-sense ssRNA segments [9, 11]. Initially, each segment associates with one VP1 copy and one VP3 copy, but replication only occurs once VP1 is activated by the assembling VP2 shell [72, 73]. Assembly of the VP6 shell stabilises the VP2 shell and forms the DLP [67, 74]. The DLP can either restart transcription or enter the ER and later get exocytosed to form the TLP by acquiring VP7 and VP4 with the help of the non-structural protein NSP4 [75-77].

The components of the rotavirus TLP are structurally well-characterised (**Table 1.1** and **Table 1.3**). 260 trimers of VP7 and 260 trimers of VP6 each form a complete icosahedral T=13 shell, whereas 60 asymmetric dimers of VP2 form a T=1 shell [60, 67, 78]. No substantial structural differences are observed for the inner and middle shells in the DLP and in the TLP. VP1 localizes at the icosahedral 5-folds and contacts VP2B in both the DLP and TLP [79]; however, whether each virion contains 11 or 12 VP1 copies remains elusive [59, 68]. Monomeric VP3 is believed to co-localize with VP1 though currently no structural information supporting this hypothesis is available [59, 68, 80]. The only available structural information on *Sedoreovirinae* capping enzymes is the X-ray structure of bluetongue virus VP4 [81].

A number of key structural aspects of the rotavirus replication cycle remain unresolved. The mechanism by which the VP1 and VP3 copies recognise the genome segments and the ways in which this connects to the recruitment of the VP2 shell are largely unknown. Though *in situ* structures of VP1 are available in both quiescent and transcribing DLPs [59, 68], such structures are yet to be determined for VP3 and the

dsRNA, hence leaving the transcription and capping mechanisms poorly understood. Finally, the DLP maturation process into the TLP and the release of the TLPs from the host cell are potentially the least understood processes of the entire replication cycle.

Table 1.3. Summary of available rotavirus structures
✓ indicates the presence and ✖ the absence of the protein/dsRNA in the quiescent and transcribing DLPs as well as TLP state. The sign in parenthesis indicates the status of the available structural information. (?) indicates unresolved with unidentified location, (✓) unresolved with identified location and (✓) resolved at resolution better than 4 Å.

	Size	State	Q-DLP (~70 nm)	T-DLP (~70 nm)	TLP (~80 nm)	Purified
VP1 (RdRP)	125 kDa	monomer	✓ (✓)	✓ (✓)	✓ (✓)	(✓)
VP2 (inner shell)	102 kDa	dimer	✓ (✓)	✓ (✓)	✓ (✓)	(?)
VP3 (capping)	88 kDa	monomer	✓ (?)	✓ (?)	✓ (?)	(?)
VP6 (middle shell)	45 kDa	trimer	✓ (✓)	✓ (✓)	✓ (✓)	(?)
VP7 (outer shell)	37 kDa	trimer	✖	✖	✓ (✓)	(?)
dsRNA	0.7-3.3 kbps	linear	✓ (✓)	✓ (✓)	✓ (✓)	-

1.2. Cryo-EM as main approach to resolve *in situ* structures

Single-particle cryogenic electron microscopy, simply referred to as cryo-EM here, is a technique used in structural biology to resolve structures of monodispersed biological molecules in their native states [82]. For this, the biomolecule in question diluted in buffer is flash-frozen and imaged in a high-voltage transmission electron microscope [82]. Increased particle size, homogeneity of particle population and number of unique particle views are important factors in improving the resolvability of the final structure [82, 83]. In recent years, cryo-EM has evolved into a technique that can routinely determine better than 4 Å structures of large macromolecular complexes [84] and in particular cases even of proteins as small as 50 kDa [85].

Cryo-EM was the technique used to obtain all the structural information presented thus far with the exception of the structures of purified ϕ6 P2 [32] and P4 [53] which were resolved by X-ray crystallography. In addition, the first rotavirus VP1 [73] and first rotavirus DLP structure [67] were resolved by crystallography, though since then, multiple cryo-EM structures have been reported for both [59, 60, 68, 78]. This observation is representative of the trend in virus structure determination heavily shifting from crystallography to cryo-EM. Compared to crystallography, cryo-EM does not suffer from the limitation that the protein or virus has to form ordered crystals. Most importantly however, cryo-EM is currently the only structural technique capable of tackling high levels of sample heterogeneity and symmetry mismatches between components of macromolecular complexes [45, 86-91].

1.2.1. Icosahedral symmetry aids capsid structure determination

In dsRNA viruses, the vast majority of capsid shells exhibit icosahedral symmetry [92, 93]. This means that these capsid shells are made of 60 repeating structures termed asymmetric units. In other words, each virus particle can be treated as a collection of 60 identical subunits effectively leading to a 60-fold increase in dataset size [94]. In addition, symmetry presents a number of other advantages for the structural determination process by cryo-EM. First, due to the 60-fold redundancy, resolving a symmetric object requires only view directions corresponding to a single asymmetric unit rather than the entire three-dimensional orientation space. Furthermore, molecules of high symmetry are unlikely to preferentially interact with the air-water interface while being flash-frozen, which would significantly reduce the range of attainable views [95]. These advantages simplify structural determination of most icosahedral capsids

essentially to a routine task requiring limited expertise [84]. Currently, hundreds of virus cryo-EM structures are available [84] including of members belonging to six out of twelve dsRNA virus families: *Chrysoviridae*, *Cystoviridae*, *Partitiviridae*, *Quadriviridae*, *Reoviridae* and *Totiviridae* [93]. However, most viral components not part of the capsid shells deviate from icosahedral symmetry and thus cannot be accurately resolved in these structures.

1.2.2. The dsRNA genome and RNA-processing proteins deviate from icosahedral symmetry

All dsRNA viruses contain components deviating from icosahedral symmetry. One ubiquitous symmetry mismatch in any icosahedral virus occurs between the capsid and the genome. As mentioned earlier, the capsid is made of 60 roughly identical components assembled together to form a single unit with icosahedral symmetry. Unlike this, the genome consists of 1-12 linear dsRNA molecules of varying lengths which inherently cannot form a single unit with any kind of exact symmetry. This type of symmetry mismatch between exactly two contributors shall be referred to as 'global mismatch'. Despite the plethora of capsid structures available, viruses whose genomes have been reliably reconstructed are extremely scarce: 2 for ssRNA viruses [96, 97], 2 for dsDNA viruses [98, 99] and 2 for dsRNA viruses [100-103]. The genome component was treated during image processing as a separate object lacking symmetry in all but two of these studies which employed a concept termed symmetry relaxation [100, 103].

Symmetry relaxation refers to a particular method of determining the orientation of an object of a given symmetry in relation to the orientation of another object of

different symmetry to which it is bound [90, 104]. Specifically, instead of treating the two objects as individual entities, the fact that they are bound to each other and thus not freely able to move with respect to one another allows treating them as a single entity. In turn, this means that the orientation and symmetry of one object restricts the possible orientations of the second object. This is important because determining the orientation of symmetrical objects tends to be easier than that of objects with lower-symmetry, which translates into a significant simplification of the search for the correct orientation of the lower-symmetry object.

In terms of viral capsids, the symmetrical object is the proteinaceous capsid with icosahedral symmetry and the lower-symmetry object is the asymmetric capsid interior. In general, viral nucleic acid is significantly more flexible and hence less rigid than viral capsids which further complicates its structural determination. Symmetry relaxation facilitates this computational problem by splitting it into two simpler tasks. First, the orientation of the capsid is determined by typical procedures employing icosahedral symmetry. Then, as few as 60 orientations (one per asymmetric unit), deterministically derived from the known orientation of the icosahedral component are probed to determine the correct orientation of the capsid interior. Though well-established, symmetry relaxation has not been demonstrated to perform robustly in any open-source software for image processing in cryo-EM.

A second type of symmetry mismatch is that which occurs between strictly more than two entities referred to as ‘local mismatch’. One local mismatch ubiquitous to dsRNA viruses with icosahedral capsids occurs between the capsid and the RdRP copies. This mismatch in fact has two components. First, the RdRP distribution in the

capsid interior creates a symmetry mismatch as 60 or a multiple of 60 copies of the RdRP would be required in each capsid in order to avoid deviating from icosahedral symmetry. Second, dsRNA virus RdRPs are monomeric [105] and tend to localize at or proximal to icosahedral symmetry axes, most often 5-folds [79] and in the case of ϕ6 PC, 3-folds [48], thus leading to a different symmetry mismatch. For the aforementioned reasons, virtually all dsRNA virus proteins that are not part of the icosahedral shells are locally mismatched (**Table 1.2** and **Table 1.3**). Resolving locally mismatched subunits requires their separation from the rest of the complex and individual analysis using standard image processing. A number of methods aimed at tackling local mismatches have been developed [45, 90, 106-109] and employed to resolve the *in situ* structures of RTV VP1 [59, 68], BTV VP1 [110], as well as ϕ6 P2 [45] and P4 [19].

1.3. Study aim

The aim of this thesis is to structurally characterise the capsid interior architecture in quiescent and transcribing ϕ6, as well as in quiescent rotavirus nucleocapsids. To this end, cryo-EM datasets of these three particle types were processed using a novel implementation of symmetry relaxation in conjunction with localized reconstruction [45].

2. Materials & Methods

2.1. Sample preparation for cryo-EM

2.1.1. ϕ6 quiescent nucleocapsids

Routine production of bacteriophage ϕ6 virions and purification of quiescent nucleocapsids are well-documented in literature [111]. Briefly, virions were produced by infecting the natural host of ϕ6, *Pseudomonas syringae pv.phaseolicola* HB10Y [16]. The lipid envelope and associated proteins were removed by Triton X-114 extraction [111] in 10 mM Tris-HCl pH 7.2 and 150 mM NaCl. QNCs were purified using a monolithic column in 1 mM $MgCl_2$ and 0.1 mM $CaCl_2$ [19]. After concentrating, the final preparation contained 5.3 mg/mL NC in 20 mM potassium phosphate pH 7.2, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$ and 150 mM NaCl. The solution was further diluted in the same buffer to 3.0 mg/mL. (This work was performed entirely by collaborators.)

2.1.2. ϕ6 transcribing nucleocapsids

Two transcription reactions were performed as following. Full transcription (i.e. TNC) was obtained by mixing in 1:1 ratio a NC mixture (1.0 mg/mL NC, 150 mM KCl and 20mM Tris pH 8.0) with a transcription buffer containing 100 mM Tris-HCL pH 8.0, 100 mM NH_4Ac, 50 mM KCl, 10 mM DTT, 6 mM $MnCl_2$, 2 mM $MgCl_2$, 4mM NTPs and 0.8 U/µl Ribolock RNasin. Defective transcription (i.e. DTNC) was prepared in a similar fashion with the exception that instead of 4 mM NTPs (ATP, GTP, CTP and UTP) in the transcription buffer, 4 mM of a mixture lacking CTP was used. The resulting mixtures for the two reactions were probed using agarose gel electrophoresis

confirming the presence and expected relative amounts of newly synthesised ssRNA (**Fig. 2.1**). (This work was performed entirely by collaborators.)

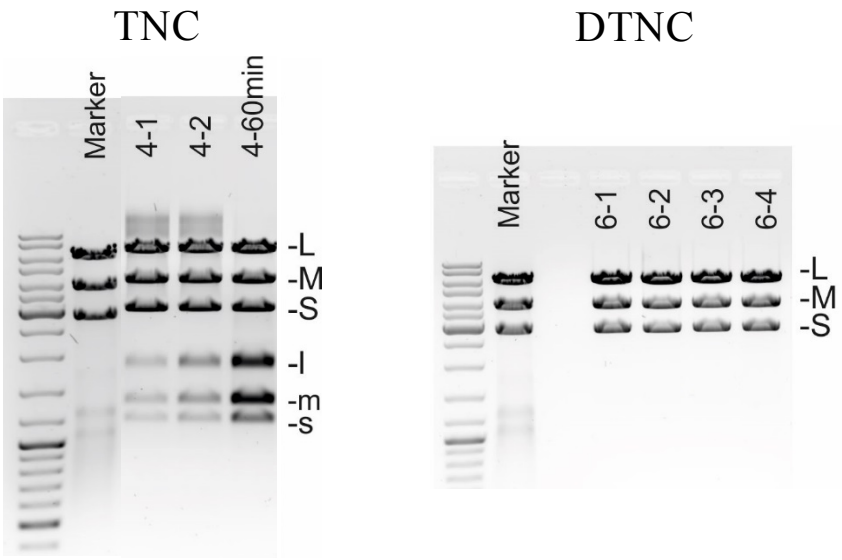


Figure 2.1. ϕ6 transcription reactions
Agarose gels indicating the presence of genome segments only in double-stranded form (S, M and L) in the DTNC sample and in both double-stranded and single-stranded form (s, m and l) in the TNC sample. 4-1, 4-2 and 4-60min were collected 5, 10 and 60 min post-incubation with the transcription buffer, respectively. 6-1, 6-2, 6-3 and 6-4 were collected 5, 10, 15 and 20 min post-incubation, respectively. Conditions 4-2 and 6-3 were used for cryo-EM data acquisition.

2.1.3. Rotavirus quiescent nucleocapsids

MA104 cells were cultured at 37°C in Glasgow's modified Eagle's Minimal Essential Medium (GMEM) supplemented with 5% (v/v) foetal bovine serum (FBS) and buffered with Hepes. Rotavirus SA11 was propagated using MA104 cells in GMEM, containing 1.5 μg/mL trypsin at 37°C [112]. At 2 days post-infection (p.i.) the culture was centrifuged at 7,000 g for 20 minutes. The supernatant was centrifuged again at 100,000 g for 45 minutes. The pellet fraction from both centrifuge spins was

pooled and resuspended in resuspension buffer (50 mM Tris HCl pH8.0, 50 mM NaCl, 1.5 mM Beta-mercaptoethanol) and freeze-thawed twice. The lipid was removed using two extractions with Vertrel XF (DuPont™). Ca^{2+} ions were removed by incubation for 24 hours in resuspension buffer containing 5 mM EGTA. DLPs were purified by CsCl density gradient (1.2 to 1.4 g/mL) at 60,000 g for 100 minutes. The DLP band was pelleted at 110,000 g for 2 hours and resuspended in 20 mM Tris HCl pH 8.0, 1 mM EGTA. (This work was performed entirely by collaborators.)

2.2. Cryo-EM data collection

For each sample, a 3-µL aliquot was applied to a glow-discharged grid (C-flat; Protochips) and plunge-frozen into liquid ethane using a Vitrobot (Thermo Fisher Scientific). Data collection parameters are summarised in **Table 2.1**. (This work was performed together with cryo-EM technicians.)

Table 2.1. Cryo-EM data acquisition parameters
The pixel sizes in parenthesis represent the uncalibrated values while the ones outside parenthesis represent the final calibrated ones.

	QNC	TNC	DTNC	DTNC	RTV
Microscope	eBIC Krios I	OPIC Polara	eBIC Krios III	OPIC Krios	OPIC Krios
Camera	Falcon II	K2	Falcon III	Falcon III	Falcon III
Mode	counting	counting	linear	linear	linear
Total dose ($e^-/Å^2$)	33	16	40	40	36
Frames	33	22	23	24	28
Exposure time (s)	2.0	0.2	0.62	0.62	0.62
Pixel size (Å)	1.42	1.35	1.40 (1.44)	1.36 (1.39)	1.36
Movies	4,400	1,400	8,900	7,300	11,700
Particles	~86,000	~6,000	~120,000	~102,000	~125,000

2.3. Single-particle image processing

2.3.1. Implementation of symmetry relaxation

The C++ source code of RELION [113, 114] was modified such that the orientational search protocol can perform symmetry relaxation. Specifically, using the XMIPP [115, 116] definitions of symmetries, the modified version of the software uses the prior orientation of each particle to calculate the orientations related by the symmetry parsed by the "--relax_symmetry" flag when running the program *relion_refine*. These orientations along with their proximal ones in three-dimensions defined by the "--sigma_ang" flag are the only orientations that are then probed by RELION instead of the orientations that would have been probed had the "--relax_symmetry" flag not been given. In the expectation step, these orientations are all given equal probability of being chosen during structure refinement and are then probed against each other to determine the one that best describes each image subjected to symmetry relaxation. All other parameters involved in running *relion_refine* were kept intact. The modified software is available at https://github.com/OPIC-Oxford/relion-relax.

Table 2.2. Summary of icosahedral refinement results

	QNC	TNC		DTNC		RTV
Particle type	DLP	DLP	SLP	DLP+DIP	SLP	DLP
Particles	~55,000	~2,500	~2,800	~74,200	~56,900	~105,400
Resolution (Å)	3.5	5.6 (2x binned)	5.5 (2x binned)	3.7	3.7	3.4
B-factor applied (Å²)	-240	-380	-310	-300	-280	-240

2.3.2. ϕ6 QNC

2.3.2.1. Icosahedral reconstruction of the capsid

Movies were corrected to account for drift and beam-induced damage in Motioncor2 [117]. Contrast transfer function parameters were estimated using CTFFIND4 [118] and only the micrographs displaying minimal drift and astigmatism were used to pick particles in ETHAN [119] for further processing. The ~88,000 picked particles were downsampled by a factor of 2 to speed up the remaining steps.

Standard 2D classification procedures in RELION were employed to curate the results of particle picking and to preliminarily separate the co-existing particle sub-populations into genome-containing "full" and genome-lacking "empty" particles. The ~55,000 "full" particles were pooled and subjected to 3D classification in RELION with icosahedral symmetry. A previously published map of the nucleocapsid (EMD-1206) [18] was scaled and low-pass filtered to 60 Å to serve as initial model. The resulting classes were highly similar in terms of reported resolution, angular accuracy and appearance and thus all ~55,000 particles were passed onto 3D refinement.

First, a standard 3D refinement employing global angular searches and icosahedral symmetry was performed in RELION for the binned particles. As the reported resolution was equivalent to the Nyquist frequency, the dataset was brought back to its original sampling of 1.42 Å/pixel and an otherwise identical round of 3D refinement was performed using local angular searches only. The final resolution and B-factor used for sharpening were determined by *relion_postprocess* using a soft-edged mask that excluded the contribution of the capsid interior (**Table 2.2** and **Appendix 1.1**).

2.3.2.2. Asymmetric reconstruction of the capsid interior

The signal corresponding to the capsid interior was hypothesised to be much weaker than that of the capsid itself due to the lack of symmetry and inherently higher degree of flexibility. Therefore, the first step in determining the structure of the capsid interior was to computationally subtract the contribution of the capsid from the acquired images. This was done using *relion_project* [88] using a mask that corresponds exclusively to the capsid (i.e. mask used in *relion_postprocess*), the output map and star file of the icosahedral refinement, and the "--ctf" flag in order to scale the map to each particle image.

The particle images now containing the contribution of the capsid interior only were subjected to 3D classification relaxing icosahedral symmetry with the incorrectly averaged map from the capsid refinement serving as initial reference. In order to prevent the prior orientation and origins from changing across iterations, the initial parameters were preserved by the addition of _rlnAngleRotPrior, _rlnAngleTiltPrior, _rlnAnglePsiPrior, _rlnOriginXPrior and _rlnOriginYPrior columns in the input star file with values equal to the _rlnAngleRot, _rlnAngleTilt, _rlnAnglePsi, _rlnOriginX and _rlnOriginY, respectively. The results were significantly improved when restricting the search to the 60 icosahedral orientations only (i.e. "--sigma_ang 0.01"). The initial rounds of classification suggested that the concentric dsRNA layers might adopt slightly different orientations in relation to one another. As a result, the approach of determining the structure of the genome was split into individual steps each corresponding to one layer only. To achieve this, for each layer two spherical masks were employed: one with

a diameter equivalent to that of the given layer ("--solvent_mask") and one with a diameter equivalent to that of the adjacent layer towards the inside of the capsid ("--solvent_mask2"). While the former mask corresponded to a region to be considered during refinement, the latter one was used to exclude the contribution of the inner genome layers thus focusing exclusively on one layer at a time.

The only layer whose orientation determination required relaxing the icosahedral symmetry of the capsid was the capsid proximal layer (i.e. outer layer or first layer). Initially, 3D classifications with symmetry relaxation were performed using 1, 2, 3 or 5 classes. The results suggested that the first layer can adopt locally distinct conformations throughout the particle population. In order to accurately align each of these locally distinct conformations in relation to the capsid, a single class was employed in symmetry relxation instead. The resulting map thus represented an overlap of the locally distinct conformations and displayed certain regions where the dsRNA strands were ambiguous. To obtain a resolution estimate for this map, 3D refinement in RELION was performed with local angular search only, leading to a 7.2-Å map (**Appendix 1.2**).

In order to separate the overlapping conformations most accurately, the further classifications were aimed specifically at the regions showing the highest degree of heterogeneity among the particle population. To achieve this, localized reconstruction [45] was performed to recenter the particles on one of these structurally heterogenous regions at a time. For this, a vector pointing to the centre of the region from the virion centre was computed and the particles were re-extracted at a 200 Å distance along this vector into boxes of 128x128x128 pixels. Furthermore, the contribution of the

remaining dsRNA in the outer layer was subtracted in similar fashion to the capsid contribution removal.

As the dsRNA layers must be connected to each other, no symmetry relaxation was required in determining the structure of the remaining layers. Rather, each layer deviated in orientation only slightly from its adjacent layers. Therefore, in order to resolve the second layer, a typical RELION 3D refinement without symmetry was performed using the orientation of the first layer as starting point and allowing the Euler angles to change by a maximum of ±9° per iteration ("--sigma_ang 3"). The same approach was pursued to resolve the third, fourth and fifth layers with the latter two being resolved in a single step. The resolutions obtained were 13.5 Å, 16.2 Å and 16.9 Å, respectively (**Appendix 1.3**).

2.3.2.3. Asymmetric reconstruction of the packaging motor P4

To resolve the hexameric NTPase P4 located on the icosahedral 5-folds [19, 46], a series of partial signal subtractions was performed. Besides the subtraction of the capsid signal, the contribution of the outer 3 genome layers was removed sequentially. First, layer 1 was subtracted using the masked map and orientations obtained from the local 3D refinement of layer 1 after symmetry relaxation. As each layer has a slightly different orientation compared to the others, the particle orientations had to match the masked map used for subtraction in order for layer 2 and 3 to be accurately removed.

Localized reconstruction was used to calculate the positions and orientations of the 12 P4 copies in each particle of the dataset. As such, ~660,000 sub-particles were obtained from the ~55,000 QNC particles. The results of the further classifications were significantly improved when only considering the locations corresponding to side views

of P4 due to the overlap of the top views with each other and with the unsubtracted genome. Two thirds of all sub-particles (~440,000) remained after discarding the top views. 3D classification relaxing C5 symmetry was performed to identify the correct orientation of each P4 hexamer in relation to the 5-fold symmetric capsid environment. Three classes were employed, of which only one corresponded to unambiguous density resembling the known P4 structure [19]. The ~145,000 sub-particles in this class were split into two halves and subjected to post-processing which lead to a ~9-Å map (**Appendix 1.4**). In order to verify the 6-fold symmetric structure previously resolved by X-ray crystallography [120], 3D refinement with C6 symmetry was performed which lead to a ~5-Å structure (**Appendix 1.4**).

2.3.3. ϕ6 TNC and DTNC

2.3.3.1. Icosahedral reconstruction of the capsid

For both TNC and DTNC, the same procedures were applied as for QNC to perform motion correction, CTF estimation and 2D classification within the scipion framework [121]. The ~6,000 TNC particles were semi-automatically picked using xmipp 3.0 [116] and classified into 10 2D classes, of which 4 were deemed to be of reliable quality for further processing. One class (46.2% of particles) corresponded to single-layered particles (SLPs), two classes (21.8% and 19.4%) to double-layered particles (DLPs) and one (12.5%) to disassembly-intermediate particles (DIPs). The SLPs and DLPs were processed similarly to QNC leading to 5.5-Å and 5.6-Å reconstructions, respectively.

The two DTNC datasets of ~120,000 and ~102,000 automatically picked particles in ETHAN were each classified into 50 2D classes. The first dataset lead to

one SLP class (27.3%), two DLP classes (16.2% and 9.0%) and one DIP class (10.0%), while the second gave two SLP classes (15.5% and 8.5%), one DLP class (23.3%) and one DIP class (8.4%). In addition, multiple distinct classes lacking the dsRNA genome were observed in both these datasets. As the DIPs displayed a similar morphology to DLPs, these two subsets were pooled and separated from SLPs. Individual 3D classifications with icosahedral symmetry identical to the one performed for QNC were carried out. As for QNC, the resulting classes were highly similar and therefore all particles were subjected to 3D refinement in identical fashion to QNC (**Table 2.2**).

The two datasets of DTNC were acquired from the same grid using almost identical microscope parameters from two different Titan Krios microscopes. However, due to the subtle particularities in optics of any one microscope, the pixel size of the two datasets was slightly different. The reported pixel size values were 1.44 Å/pixel and 1.39 Å/pixel, respectively. The two datasets were thus processed entirely using these values. The availability of atomic models for the capsid proteins of ϕ6 facilitated probing whether these values were accurate. Following a series of model fittings in UCSF Chimera [122] at pixel sizes ranging from 1.34 Å to 1.46 Å, it was concluded that the actual pixel sizes were 1.40 Å and 1.36 Å, respectively. In order to combine the two datasets, the latter one was scaled from 1.36Å to 1.40Å using xmipp functionalities from scipion while preserving the particle box size. The combined dataset was refined with icosahedral symmetry using the same procedure as for the individual ones (**Table 2.2** and **Appendix 1.5**).

2.3.3.2. Asymmetric reconstruction of disassembly intermediates in DTNC

The capsid of the DIPs was treated as an asymmetric object on an icosahedrally symmetric framework in similar fashion to the asymmetric dsRNA in relation to the capsid in QNC. As a result, the same general approach employing symmetry relaxation was followed to determine the structure of the DIPs. However, unlike the dsRNA displaying locally distinct conformations, the individual DIP conformations differed more substantially from one another due to the highly variable degree to which the outer shell was disassembled. As a result, the 3D classification with symmetry relaxation was performed using multiple classes with only one mask ("--solvent_mask") covering the entirety of the inner and outer capsid layers. As for the QNC genome, the "--sigma_ang 0.01" was used thus restricting the angular search to 60 orientations only. One parameter that proved to substantially facilitate the classification was "--strict_highres_exp 60", which defines the maximum resolution allowed to influence the orientational search in the expectation step of RELION. As the RELION classification is typically aimed at outputting the highest resolution map rather than separating the heterogeneity in the particle population, limiting this parameter to a very low resolution (i.e. 60 Å) shifted the focus to resolving heterogeneity. In terms of the DIPs, the output classes were all low-pass filtered due to this parameter regardless of corresponding number of particles and rigidity of the capsid in that conformation. In other words, this prevented particles in different stages of outer shell disassembly to cluster into the same class by being assigned orientations that would overall maximise the averaging of the outer shell. Twelve distinct maps were obtained and consequently each subjected to

local 3D refinement without symmetry leading to resolutions ranging from 6.3 to 20 Å (**Appendix 1.6** and **Appendix 1.7**).

2.3.3.3. Asymmetric reconstruction of the capsid interior

To reconstruct the capsid interior a similar procedure as for QNC was performed individually for SLP and DLP. First, the signal corresponding to the capsid was computationally subtracted from the original images. Second, 3D classification with symmetry relaxation and 1 class was performed using the QNC map filtered to 60 Å as initial model. The resulting maps for both SLP and DLP bore a strong resemblance to QNC in terms of the dsRNA organisation. However, 3 copies of the RdRP were observed at each spool end related by C3 symmetry. This suggested the potential presence of incorrect averaging and thus localized classification was performed once again to resolve it. For each of the two datasets, the two spool ends of each particle were extracted as individual sub-particles and combined into a single dataset. Subsequently, 3D classification relaxing C3 symmetry was performed with 4 classes, of which 2 corresponded to accurately aligned particles. Each of the 4 reliable classes (2 for SLP and 2 for DLP) was split into two half-sets and subjected to post-processing which lead to final resolutions between 6.6 and 7.5 Å (**Appendix 1.8** and **Appendix 1.9**).

2.3.4. RTV QNC

2.3.4.1. Icosahedral reconstruction of the capsid

The same procedures were applied as for QNC to obtain the icosahedral structure of the RTV capsid. Particles were downsampled by a factor of four. Downsampled particles were 2D classified in 100 classes using RELION. Particles belonging to well-resolved genome-containing ‘full’ classes (105,435 particles) and

'empty' classes corresponding to particles that were lacking the genome (4,784 particles) were processed separately. Particles in both sets were subjected to 3D refinement in Relion using a previously published map (EMD-5486) [123] as a starting model. 3D refinement with local searches only was continued with the original particles (i.e. before downsampling). The 3D map using 'full' particles reached 3.4 Å resolution, whereas the resolution of the map calculated from the 'empty' particles was 4.3 Å. (This work was performed entirely by collaborators.)

2.3.4.2. Combined localized reconstruction of the RdRP

The RdRp density located at the icosahedral 5-folds [59, 68, 79] was resolved using localized reconstruction approach integrated in Scipion. First the two icosahedral protein shells were subtracted from the 'empty' particles images using RELION. Twelve sub-particles, each corresponding to a five-fold vertex were then extracted from the particle images. The vertex sub-particles were classified into two classes relaxing C5 symmetry. The resolution in the expectation step was limited to 40 Å ("--strict_highres_exp 40"). One of the two classes showed RdRp density, which was used to create a RdRp mask. This mask was used in a second round of classification where symmetry was again relaxed but resolution was not restricted. The angles from this classification were assigned to sub-particles extracted from the original, unsubtracted 'empty' particle images. This allowed reconstructing an asymmetric map of the vertex with RdRp density at average resolution of 5.1 Å, which accurately defined the position and shape of the RdRp density.

To resolve the RdRp density from 'full' particles, the RdRp density determined from the 'empty' particles was used as initial model, in addition to the RdRp mask

created from the same model. Twelve sub-particles per particle were extracted from 'full' particles from which the contribution of the protein shells had been removed. The sub-particles were again classified in two classes relaxing C5 symmetry. Particles in those classes that showed clear RdRp density were further classified into eight classes without alignments (i.e. "--skip_align") to take into account possible conformational heterogeneity in the RdRp. The best ordered RdRp sub-particles were extracted from the original 'full' particle images and used to reconstruct an asymmetric map of the vertex with RdRp density at an average resolution of 3.7 Å. (This work was performed entirely by collaborators.)

2.3.4.2. Asymmetric reconstruction of the RdRP-dsRNA complex

The structure of the capsid interior was resolved by symmetry relaxation in similar fashion to that of QNC, TNC and DTNC. Briefly, the capsid signal was subtracted and 3D classification was performed with symmetry relaxation employing two masks, namely "--solvent_mask" and "--solvent_mask2". However, it was observed that using two classes instead of a single one produced significantly better results as one of the two classes consisted of poorly aligned particles.

2.3.4.3. Individual localized reconstructions of the RdRP binding sites

The well resolved class after symmetry relaxation presented numerous areas with ambiguous density and thus multiple rounds of localized reconstruction were performed, namely one per each of the 12 icosahedral 5-folds. For consistency, 6 classes were used for each of the 12 classifications alongside a soft cylindrical mask of radius 92 Å and height 98 Å. For further analysis, all 72 classes were reconstructed in

relion_reconstruct to 5.44 Å (equivalent to Nyquist frequency) for model fitting and to 12 Å for figure rendering.

2.5. Model building and refinement

2.5.1. Fitting and refinement of capsid proteins

A previously determined model of the ϕ6 QNC asymmetric unit [19] was fitted as rigid body in phenix [124] and subsequently used for refinement in REFMAC5 [125]. (This work was performed entirely by collaborators.) For ϕ6 DTNC, the resulting model was fitted in UCSF Chimera as rigid-body in both DLP and SLP maps.

2.5.2. Fitting and refinement of other proteins

In the case of ϕ6, previous models were available for the RdRP [32] and packaging motor [120]. These pdb files were used for fitting as rigid bodies in phenix [124].

For RTV, the available RdRP model at the beginning of the study was obtained from X-ray crystallography data [73]. This model was fitted in the 3.7-Å cryo-EM map obtained by combining all the sub-particles corresponding to RdRP locations. The map presented several differences compared to the model and thus further refinement in REFMAC5 was performed to account for them. (This work was performed entirely by collaborators.)

The updated model was used to identify capsid contacts and for further rigid body fittings in the maps obtained from the 12 individual localized classifications. Two types of fitting were performed, both using maps reconstructed to 5.44 Å (equivalent to Nyquist frequency). First, the 60 potential binding sites of VP1 were determined using

the ‘sym’ command in UCSF Chimera. As each localized classification corresponded to one 5-fold, the five degenerate VP1 orientations at each 5-fold were probed against each of the 6 maps from the classification at the same 5-fold. If the rigid body fit in UCSF Chimera did not displace the VP1 molecule substantially, it was denoted as successful and the fit value was recorded.

To assess the reliability of the UCSF Chimera fits, phenix.dock_in_map [124] was performed for each of the 72 maps using a randomly oriented VP1 molecule as starting model. The values reported were either below 0.10 or above 0.25 and thus the ones below 0.10 were considered unreliable. With a few exceptions, the maps with CC values higher than 0.25 were the ones giving the highest values for the fittings performed in Chimera as well.

2.5.3. Fitting and refinement of dsRNA models

The resolution of the cryo-EM maps even in the best resolved areas was too low to allow discerning the exact base pair at any position. However, as the dsRNA helices were unambiguously separated in the majority of the reconstructions, fitting and refinement of polyA-U models was performed. The map used for the ensemble model had pD3’-1 to account for layer 1 and a dsRNA model was also built for pC2.

Due to the poor resolution especially for the inner layers, the model was built starting with short polyA-U stretches, each corresponding to one discernible dsRNA helix of the map. First, 30-mers of canonical A-form dsRNA were built in COOT [126] and fitted as rigid bodies. The following steps were performed using a custom version of ROSETTA [127] provided by Frank diMaio and designed to perform extension-refinement cycles for double-stranded nucleic acid in low-resolution cryo-EM maps.

Each 30-mer was individually extended and refined until the entire corresponding helix was accounted for. As layer 1 was significantly better resolved than layers 2-5, extension of up to 100-150 base pairs could be achieved in one ROSETTA round. However, proceeding towards the centre of the virion, this value decreased progressively to 20-30 base pairs per cycle with the exception of layer 5 where the helix curvature was unusually low. The ensemble model consisted of 56 individual dsRNA strands totalling 12,245 base pairs (~92% of the 13.3 kilobase pair genome) split across layers 1-5 as following: 5274, 3250, 2340, 680 and 701, respectively. Certain areas of particularly poor density lead to slightly distorted models or could not be accounted for despite resembling dsRNA strands upon visual inspection.

2.6. Data analysis

2.6.1. Identifying potential capsid-dsRNA contacts in QNC

The contacts between the inner capsid shell and the dsRNA were analysed for pD3'-1 and pC2. For this, the two maps were reconstructed from original images i.e. from which the capsid signal was not subtracted. The capsid asymmetric unit was fitted in each of the two maps and symmetrised in UCSF Chimera using the 'sym' command and I1 symmetry. The corresponding dsRNA models for layer 1 were fitted in the maps as well, thus leading to a composite model of the capsid and layer 1 in both genome conformations. Custom python scripts were designed to calculate all pairwise distances between the atoms of the inner capsid and the non-hydrogen atoms of the dsRNA.

2.6.2. Analysis of QNC dsRNA packing parameters

Custom python scripts were written to process the pdb coordinates of the dsRNA model. The position of a base pair was considered to be the average between the positions of the adenine N1 and uracil N3 atoms. To calculate the position of each base pair *n* and account for potential distortions of the dsRNA model, the positions of the 11 consecutive base pairs from *n*-5 to *n*+5 were averaged as A-form nucleic acid has 11 base pairs per turn [128]. Curvature, polygonality and local packing density were calculated for each individual base pair position in the genome.

Curvature was denoted as the angle across 30 base pairs at the location of base pair *n* between the segments [*n*-15, *n*] and [*n*, *n*+15] [129]. Local packing density was determined by summing the molecular mass of each non-hydrogen atom within a 40 Å radius of a given base pair and dividing the sum by the Avogadro number and the sphere volume. As layer 1 was in the proximity of the inner capsid shell, the volume used in the calculations was calculated as the sphere volume minus the volume corresponding to a sphere cap of radius 40Å and height 11.5 Å (equal to the radius of A-form DNA [128]).

To calculate polygonality the following approach was performed. First, the equation for the two-dimensional plane perpendicular to the direction of the helix was determined for each base pair *n*. Next, the neighbouring helices to the one containing base pair *n* were identified. For this, the base pair belonging to a neighbouring helix had to be less than 50 Å away from base pair *n* and less than 3.4 Å away from the previously determined 2D plane. The number of neighbours was limited to 8 with a maximum of 3 in adjacent layers and 2 in the same layer as the one containing base pair *n*. Once the neighbours were identified, the following formula was used to determine

the polygonality index of base pair *n*, where *k* is the number of neighbours and a_i is the angle at base pair *n* between consecutive neighbours:

$$PI = \frac{1}{\frac{\sum_{i=1}^{k} \left| \frac{2\pi}{k} - a_i \right|}{k} + 1}$$

The neighbours that decreased polygonality were denoted as false positives likely caused by the distortions in the dsRNA model and thus excluded from the calculation. Cubicity and hexagonality were computed using the closest 4 neighbours to base pair *n* and π/3 and π/4 were used instead of 2π/*k*, respectively.

2.6.3. Analysis of RTV RdRP distribution correlation

To determine whether any correlation exists between the VP1 orientations at neighbouring 5-folds, the particle composition in the corresponding output classes from the 12 localized classifications was evaluated. For example, to define the influence of class 1 of 5-fold #1 (#1.1) on class 2 of 5-fold #2 (#2.2) the number of particles shared by #1.1 and #2.2 was identified. This was then compared to the corresponding expected number of particles shared by the two classes if there was no correlation between them (i.e. given random orientation of the 2 VP1 with respect to each other). Thus, the difference between these two values (each expressed as percentage of particles from the entire dataset) represented the absolute deviation *AD* from randomness caused by #1.1 on #2.2 or vice-versa. Dividing this difference by the random value gave the normalised deviation *ND* caused by #1.1 on #2.2 or vice-versa.

In order to calculate the individual influence *II* of #1.1 on #2 as a whole with i.e. #2 having 3 reliable classes (c1, c2 and c3) the *AD* caused by #1.1 on #2.1, #2.2 and

#2.3, respectively, were summed up. Dividing this sum by the sum of the number of particles shared by #1.1 with either of #2.1, #2.2 or #2.3 gave the normalised influence *NI* of #1.1 on #2.

Next, to determine the influence of #1 as a whole with i.e. #1 having 2 reliable classes (c1 and c2) on #2 as a whole, the *AD* caused by #1.1 on #2.1, #2.2 and #2.3 as well as those caused by #1.2 on #2.1, #2.2 and #2.3 were summed. Again, dividing this value by the sum of the number of particles shared by #1.1 with either of #2.1, #2.2 or #2.3 and those shared by #1.2 with either of #2.1, #2.2 or #2.3 lead to the pairwise influence *PI* of #1 on #2 or vice-versa.

The total influence *TI* and average influence *AI* of all classes showing reliable VP1 density from each 5-fold were also computed. For *TI* of a class i.e. #1.1, the *II* of #1.1 on all 5 of its neighbours were summed up, divided by the number of particles in #1.1 and represented as percentage. For the *AI* of #1.1, the *II* of #1.1 on each of its 5 neighbours divided by the corresponding number of reliable classes at that neighbour were summed up, divided by the number of particles in #1.1 and displayed as percentage.

Finally, the defining influence (DI) of each 5-fold was calculated as the weighed average of the *AI* values for the reliable classes at the 5-fold in question with respect to number of particles in each class.

3. Atomic model of an encapsidated viral genome

3.1. Foreword

This chapter describes the architecture of the ϕ6 nucleocapsid and discusses it from a number of different perspectives. Though the structure was published shortly before the writing of this thesis began [130], the majority of the discussion points presented below are novel. As this work involved modifying an existing image processing software, a substantial part of the project consisted of understanding the inner workings of the software and how it can be modified to perform the task at hand. Testing the modified algorithm turned out to be a challenging task in the initial stages as it was unclear whether the algorithm or the sample was causing the inability to resolve the ϕ6 genome. This confusion was cleared with the help of an artificial dataset generated by projecting from random orientations an icosahedral capsid with a cylinder locked inside it. The ability to resolve the cylinder using the modified algorithm indicated that the software was properly functioning. However, as hypothesised in other studies and described in this chapter, the organisation of the dsRNA within the ϕ6 nucleocapsid is highly heterogenous. As a result, the protocol adopted to resolve the ϕ6 genome described in the previous chapter was the outcome of hundreds of rounds of volume classification. Notably, it proved to be critical to perform symmetry relaxation using a single class and localized reconstruction to subsequently separate the now identically aligned yet different conformations. Hopefully, this protocol will turn out to be applicable in tackling other symmetry mismatches, potentially even beyond asymmetric genomes confined to icosahedral capsids.

3.2. Summary

Chapter 3 tackles the cryo-EM structure of inactive ϕ6 nucleocapsids. As the structure of the icosahedral capsids layers has been previously reported [19], the main focus here is the organisation of the capsid interior. The protocol used identifies 5 distinct dsRNA layers of similar yet not identical orientation in relation to one another. These layers follow a spooled arrangement similar to the models previously associated exclusively with dsDNA genomes [98, 99, 131-135]. Though the capsid interior maps are reliable enough to allow fitting of dsRNA models, the RdRP copies cannot be accounted for, unlike in other dsRNA viruses where they appear well-resolved [100-103]. These observations are discussed in light of their implications for the ϕ6 replication cycle and the distinct models hitherto associated with dsRNA and dsDNA genomes, respectively. Finally, the behaviour of the dsRNA is compared and contrasted to the previously reported liquid crystalline parameters of condensed nucleic acid.

3.3. Structure of the protein shells

The ϕ6 QNC contains two concentric protein shells denoted here as the inner and the outer shell. The icosahedral structure of these protein shells has been determined previously by cryo-EM on multiple occasions with the highest achieved resolution of 4.0 Å [19]. Here, the structure of the capsid shells was resolved once again using a much larger dataset leading to a 3.5 Å resolution structure (FSC plot in **Appendix 1.1**), which allowed a slight improvement of the atomic model of the asymmetric unit (data not shown). The inner T=1 shell consists of 60 asymmetric dimers of the ~85 kDa protein P1 and is surrounded by the outer incomplete T=13 shell made up of 200 trimers

of the ~16 kDa protein P8 (**Fig. 3.1)**. Embedded in the outer shell at the 12 icosahedral 5-folds are homohexamers of the ~35 kDa packaging motor P4. The icosahedral reconstruction of the QNC failed to resolve the components deviating from icosahedral symmetry: the hexameric P4 located on the 5-fold symmetry axes [19, 46], the dsRNA genome inside the capsid, the monomeric P2 present in 8-12 copies inside the capsid [36, 45, 47, 48] and the monomeric P7 also present inside the capsid in a yet unknown number of copies [49, 50].

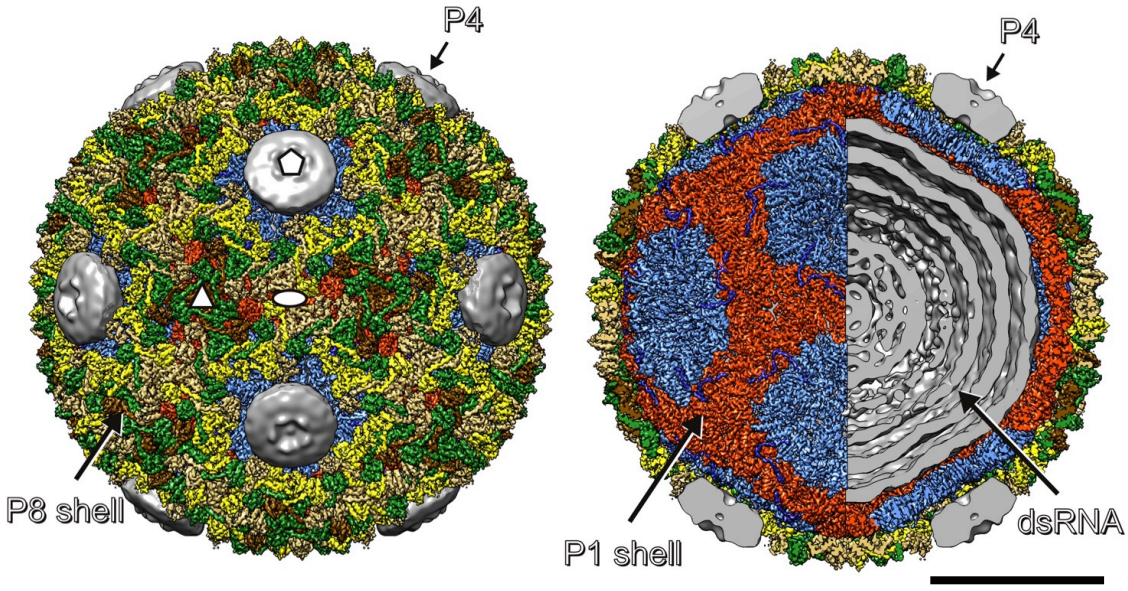


Figure 3.1. Icosahedral reconstruction of ϕ6 QNC
Coloured are the components of the two icosahedral capsid shells, namely P1 (P1A in cornflower blue and P1B in orange red) and P8 (yellow, gold, dark green and dark brown). P4 and the dsRNA deviating from icosahedral symmetry are in grey. P2 and P7 could not be located. The icosahedral 2-fold, 3-fold and 5-fold axes are indicated by the ellipse, triangle and pentagon. Figure taken with permission from [19]. Scale bar 200 Å.

3.4. Organisation of the capsid interior

3.4.1. Breaking the capsid symmetry

The structure of the capsid interior was resolved in multiple steps (**Fig. 3.2**). First, a 7.2-Å map (**Appendix 1.2**) representing all 55,000 particles (i.e. 100%) of the

capsid proximal genome layer (i.e. layer 1) was determined which exhibited pseudo-D3 symmetry (three psuedo-C2 axes rotated by 120° within one plane perpendicular to a pseudo-C3 axis i.e. pD3). The pseudo-C3 axis coincides with one of the icosahedral 3-folds and the three co-planar pseudo-C2 axes coincide with 3 icosahedral 2-folds. The structure corresponds to a right-handed spool and displays at least 12 transitions to layer 2. Most of the structure consists of unambiguously separated dsRNA helices though the spool ends and one of the pseudo-C2 location presented discontinuous density.

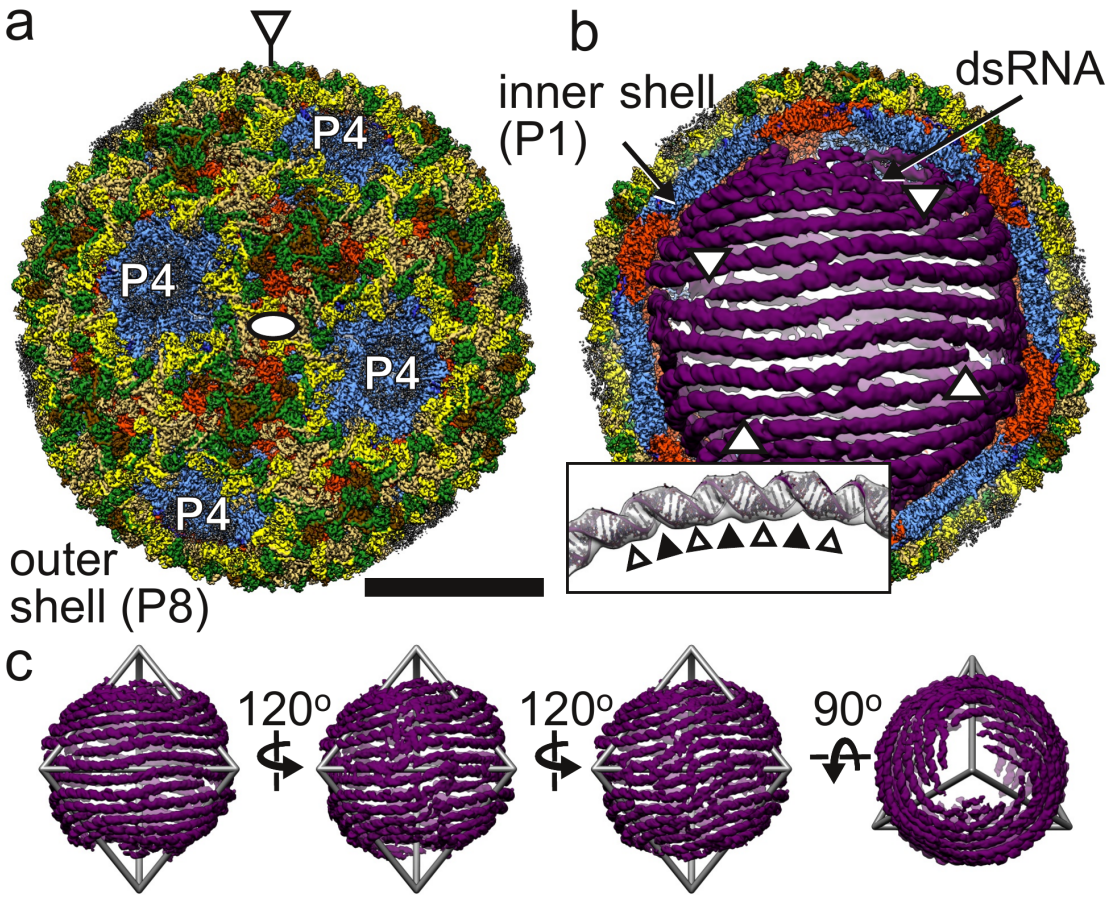


Figure 3.2. Asymmetric reconstruction of the ϕ6 QNC genome
Colouring as in **Fig. 3.1** with the dsRNA genome in dark purple. **a**, Icosahedral reconstruction of QNC seen along a 2-fold axis. **b**, Cut-open view of **a** showing the capsid proximal dsRNA. Triangles represent transitions between layers 1 and 2. The inset shows a A-form dsRNA model built into the density with major (filled triangles) and minor (empty triangles) grooves indicated. **c**, the dsRNA density in **b** viewed along the 3 pseudo-C2 and 1 pseudo-C3 axes describing the pseudo-D3 organisation. A D3-symmetric cage is shown as reference in grey. Figure taken from [130]. Scale bar 200 Å.

3.4.2. Discerning locally distinct sub-conformations

The ambiguous pseudo-C2 location was used as focus in further localized classifications to resolve structural heterogeneity (**Fig. 3.3**). The first round of classification lead to two distinct maps: one resembling the pseudo-D3 organisation of the original map (i.e. pD3' 57% of particles) and one adopting pseudo-C2 symmetry (i.e. pC2, 43% of particles, 8.0 Å, FSC plot in **Appendix 1.2**). While pC2 presented no structurally ambiguous regions, the pseudo-C2 location in pD3' was still not fully resolved. Further classifications discerned multiple pseudo-D3 sub-conformations in which this region was eventually well-resolved (i.e. pD3'-1, 20% of particles, 9.7 Å, FSC plot in **Appendix 1.2**). However, detecting these sub-conformations was limited by the number of particles subjected to classification as reliable averaging could not be achieved when splitting the dataset into more than 3 classes.

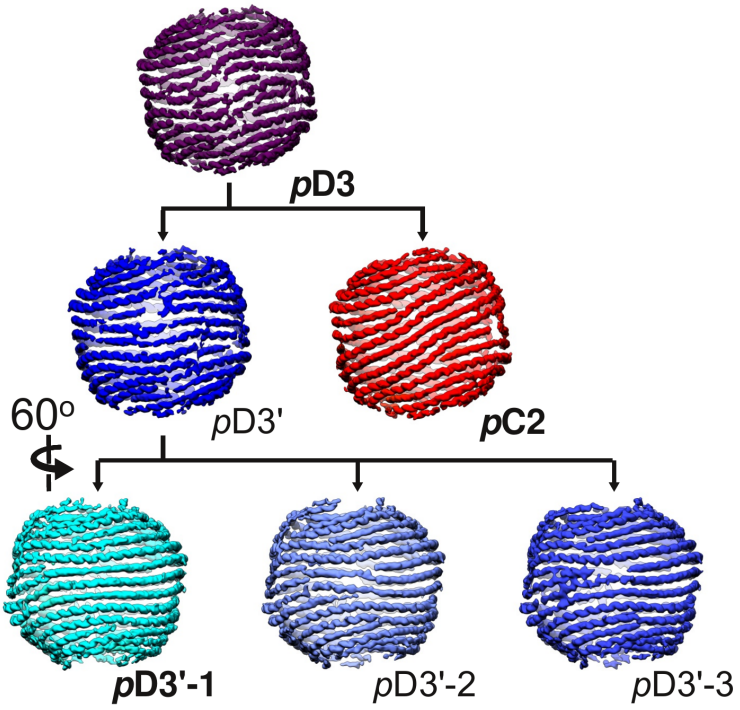


Figure 3.3. Locally distinct sub-conformations of the ϕ6 QNC genome
Results of consecutive localized classifications are shown. Adapted from [130].

3.4.3. Analysis of dsRNA packing parameters

Starting from the original pseudo-D3 map, four more genome layers were sequentially resolved, with layers 4 and 5 belonging to the same map (resolutions of 13.5 Å, 16.2 Å and 16.9 Å, FSC plots in **Appendix 1.3**). Layers 1-4 are right-handed spools of similar average inclination (81°, 77°, 77° and 82° respectively) centred at radii of 201, 170, 141 and 107 Å respectively from the capsid origin (**Fig. 3.4**). By contrast, layer 5 deviates from this nearly co-axial spooling pattern. Instead, it represents a rifling left-handed arrangement of 12 strands running much more diagonally than layers 1-4 (average inclination of 36°). The only areas devoid of dsRNA remained the spool ends and the region between the virion centre and the 5th layer. These areas are the most likely locations for the yet unresolved P2 and P7 proteins.

Layer 1 appears best resolved and is the only one where the dsRNA minor and major grooves are apparent (**Fig. 3.4**). Nonetheless, all 5 layers present dsRNA helices mostly unambiguously separated which allowed fitting of polyA-U models tracing over 90% of the entire genome (13.3 kbps total length). These models allowed quantifying a variety of dsRNA packing parameters across the entire capsid interior (**Fig. 3.5**).

The average inter-helix distance is ~26 Å within a genome layer and ~29 Å between adjacent layers. The overall dsRNA packing density is ~350 mg/mL with layers 1 and 2 being tighter packed (average local packing densities of 389±61 and 391±40 mg/mL, respectively) than layers 3-5 (average local packing densities of 355±46, 338±38 and 278±58 mg/mL, respectively). Overall, the local packing is tighter at the spool equator than at the spool ends, consistent with the cryo-EM density resolvability.

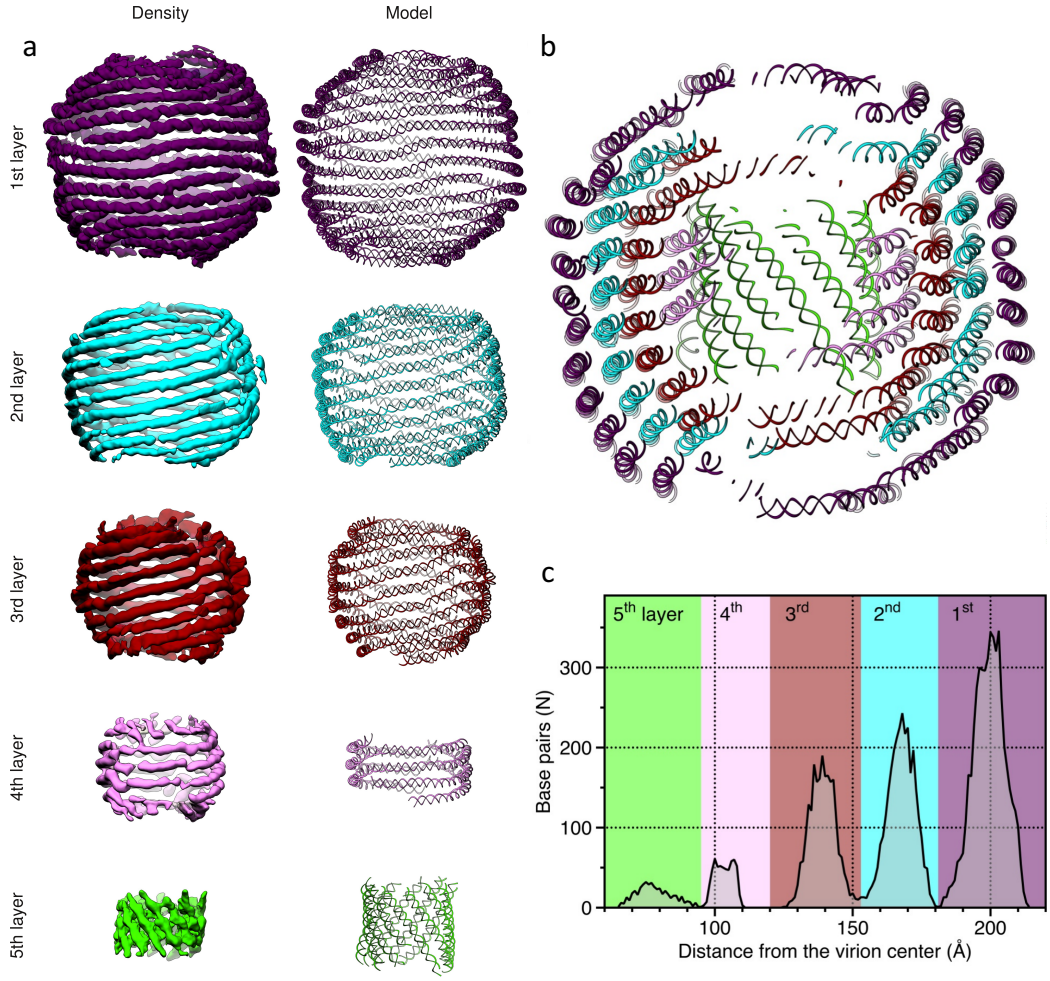


Figure 3.4. Ensemble model of the five dsRNA layers in ϕ6 QNC
a, Cryo-EM map and corresponding model for each of the 5 layers. **b**, Composite model of the ϕ6 QNC genome. **c**, Distance from the virion center for all modeled base pairs. Adapted from [130].

To quantify packing geometry, we defined dsRNA regions as being hexagonal if the angles between the closest neighbouring strands were closest to 60° or cubic if these angles were closer to either 45° or 90°. Whereas layer 1 presented almost equal amounts of these geometries, the degree of cubicity increased progressively to layers 2-5.

In terms of curvature, layers 1 and 2 appeared least curved (17±6° and 18±5°, respectively) while layers 3 and 4 are increasingly more curved (22±6° and 24±4°, respectively). Taken together, these results show that the curvature, packing density and polygonality appear to follow the same trend across the 5 genome layers.

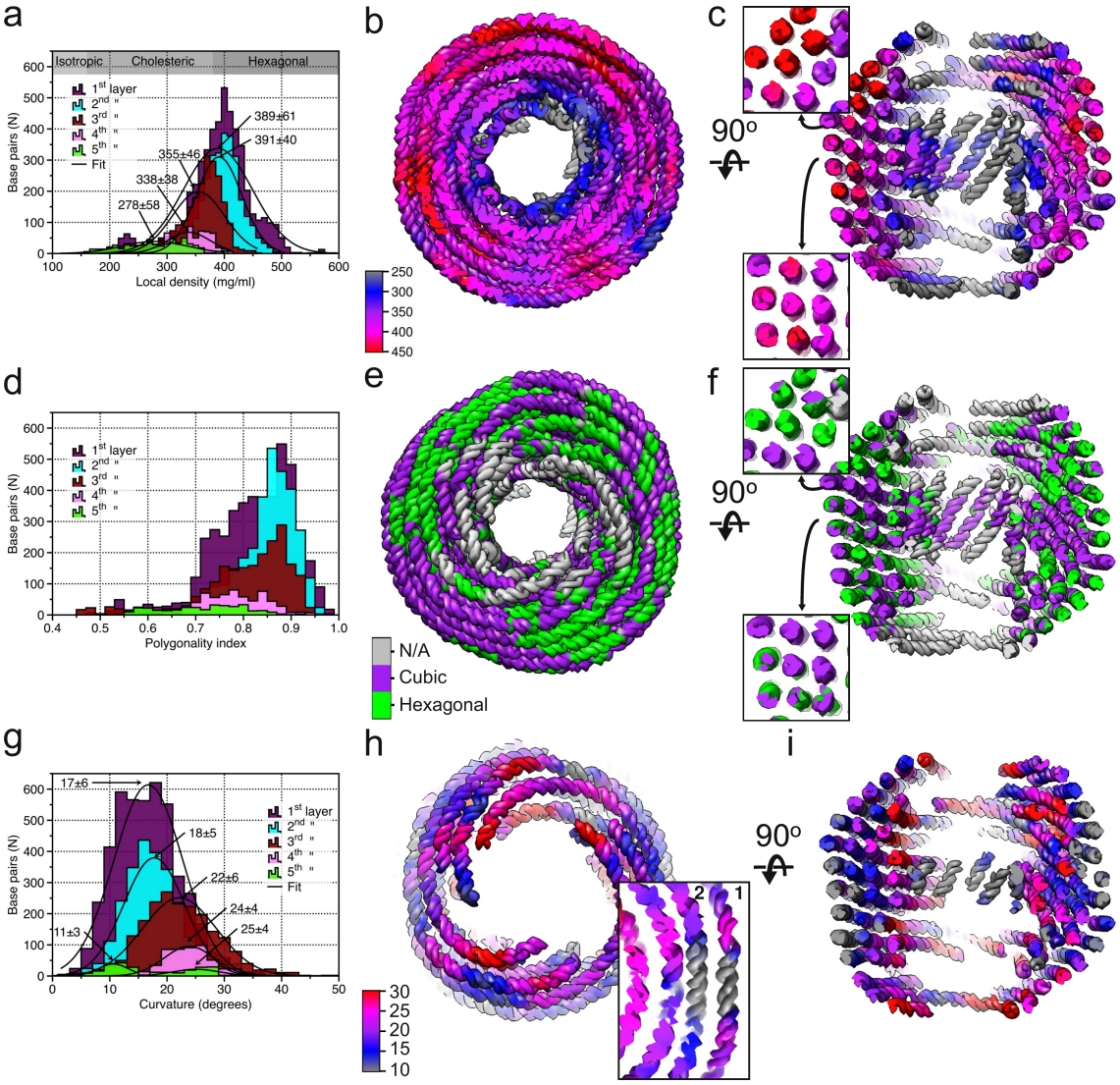


Figure 3.5. dsRNA packing parameters in ϕ6 QNC
Local packing density (**a**, **b**, **c**), polygonality (**d**, **e**, **f**) and curvature (**g**, **h**, **i**) for the ϕ6 QNC genome. Taken from [130].

3.4.4. Identification of capsid-dsRNA contacts

The atomic models of the capsid and layer 1 allowed the P1 residues in closest proximity to the dsRNA to be determined by computing all pairwise distances between all P1 residues across the 60 dimers and all base pairs in layer 1. As expected, the putative contacts to the negatively charged dsRNA are predominantly the positively charged residues lysine and arginine leading to unspecific charge-based interactions

(**Appendix 2**). In addition, three asparagine residues Asn279, Asn406 and Asn570 appeared to point towards the minor groove of the dsRNA potentially leading to specific interactions with certain base pairs [136]. As the resolution of our reconstructions only allowed fitting polyA-U dsRNA models instead of tracing the ϕ6 genome sequence, this latter observation remains to be verified.

One P1 region of particular interest in contacting the dsRNA is helix 205-217 (**Fig. 3.6**) which contains three lysine residues (Lys208, Lys215 and Lys217). Across different P1 dimers, this helix appears to be in the proximity of the dsRNA and to adopt a variety of orientations in relation to the dsRNA strands. It is thus possible that this helix might play a role in directing the nascent strands during genome replication.

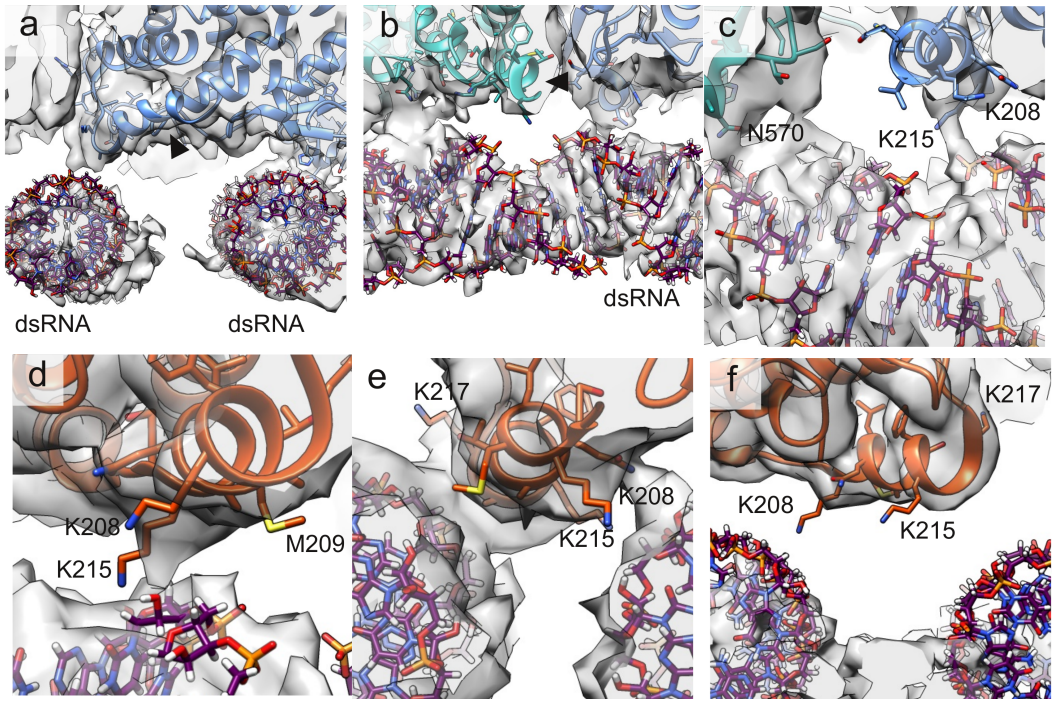


Figure 3.6. Putative capsid-dsRNA contacts in ϕ6 QNC
6 different regions corresponding to putative capsid-genome contacts. The map shown has D3 symmetry applied and is resolved to 4.0 Å (FSC plot in **Appendix 1.1**). Different P1A monomers are coloured in blue and teal, respectively, while P1B monomers are in orange. Adapted form [130].

3.4.5. Resolving the RNA packaging motor P4

The RecA helicase P4 is a homohexameric ATPase located at the icosahedral 5-fold axes of symmetry on top of the inner capsid and embedded within the outer shell [19, 120]. The C6-symmetric structure of purified P4 has been determined by X-ray crystallography [120]. Later, cryo-EM and localized reconstruction were used to resolve the structure of the hexamer from quiescent ϕ6 nucleocapsids [19]. Resolving the symmetry mismatch between the 6-fold symmetric P4 and its local 5-fold symmetric environment showed that in fact P4 deviates slightly from C6 symmetry. This observation is supported by a range of studies showing that asymmetry is crucial in the functioning of hexameric ATPases [137, 138]. Here, the large cryo-EM dataset needed to resolve the genome organisation in ϕ6 nucleocapsids is used to revisit the *in situ* structure of P4. Similarly to [19], a higher-resolution symmetric structure and a lower-resolution asymmetric structure are resolved with the purpose of validating the symmetric X-ray structure and providing more insights into the physiological structure, respectively. Notably, only ~30% of P4 sub-particles appear well-resolved after 3D classification relaxing C5 symmetry. The ~5-Å symmetric structure (**Appendix 1.4)** is in good agreement with the X-ray structure [120] as shown by a fit of the atomic model into the cryo-EM map (**Fig. 3.7**). However, the loops pointing to the centre of the hexamer were not visible in the X-ray structure and are not resolved in the cryo-EM map either. This suggests that these loops most likely deviate from the 6-fold symmetry of the hexamer and potentially display a high degree of structural heterogeneity. The ~9-Å asymmetric structure (**Appendix 1.4**) provides slightly more information into the inner pore organisation though the resolution is too low to build the missing loops.

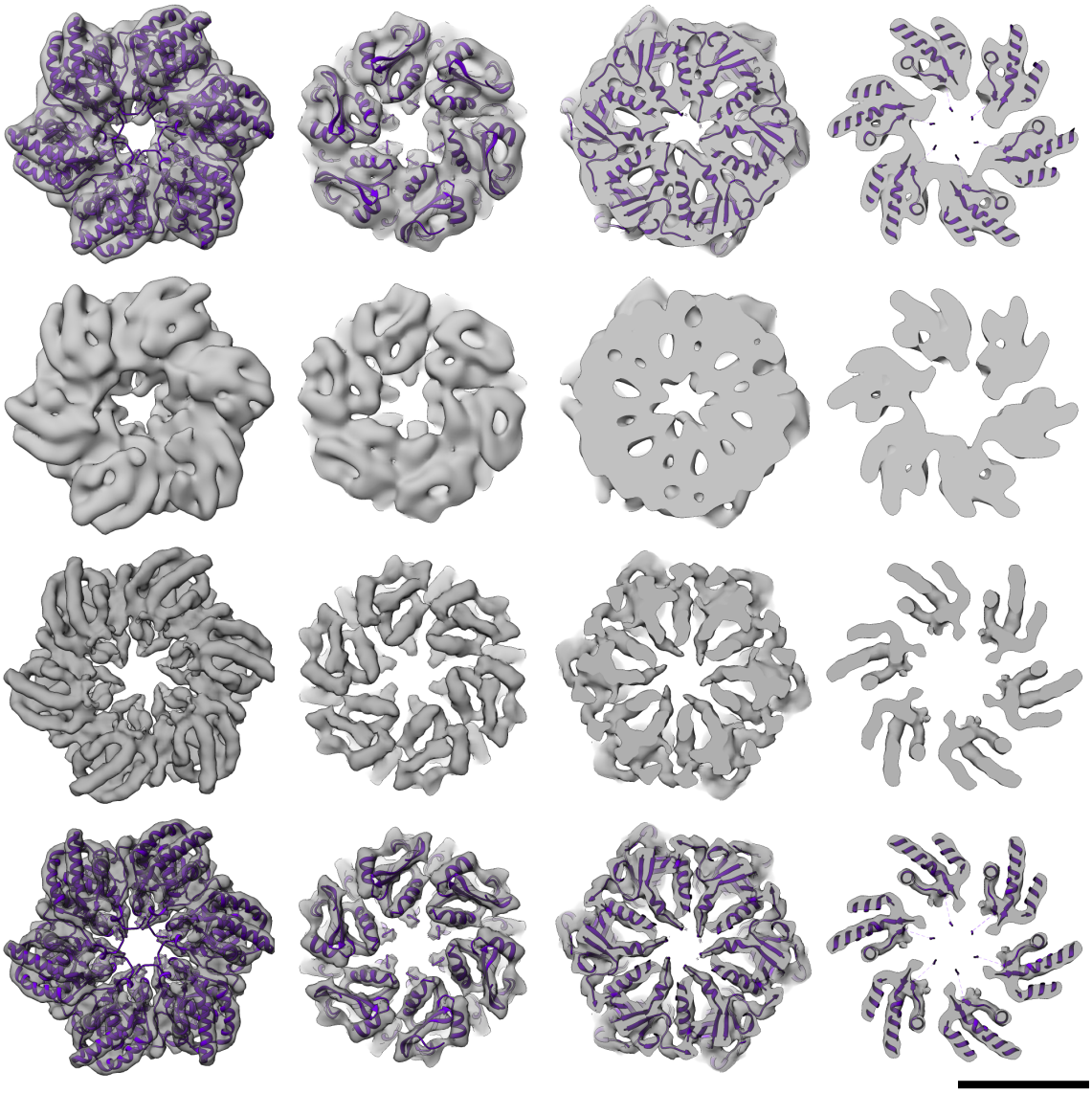


Figure 3.7. Structure of the P4 hexamer ϕ6 QNC
The asymmetric (top 2 rows) and symmetric (bottom 2 rows) structures of the P4 hexamer. The C6-symmetric X-ray structure of P4 [120] fitted as rigid body in the cryo-EM maps is shown in purple. Columns correspond to different views, namely full top view and 3 bottom view slabs, respectively. Scale bar, 50 Å.

3.5. Discussion

This chapter presents the *in situ* structure of the bacteriophage ϕ6 genome. No previous study aimed at elucidating the organisation of encapsidated double-stranded viral genome succeeded in providing a reconstruction sufficiently well resolved to

model the majority of the genome length. Several key factors contributed to the successful fitting of atomic models to describe the holistic structure of the ϕ6 genome. First, a large cryo-EM dataset of 55,000 reliable particles was needed to tackle the lack of symmetry and multitude of conformations that the nucleic acid can adopt. Second, the novel implementation of symmetry relaxation in RELION facilitated the search for the correct genome orientation. Third, localized reconstruction proved essential in separating the locally distinct genome conformations. Fourth, a custom version of ROSETTA [127] provided by Frank diMaio dramatically simplified the process of building and fitting over 12,000 base pairs of dsRNA in the cryo-EM maps.

Unexpectedly, the spooled architecture of the ϕ6 genome appears to be drastically different than the non-spooled arrangement observed in other dsRNA viruses [100-103, 139] and resembles much more closely models previously associated exclusively with dsDNA viruses [98, 99, 131-135]. At the time this thesis was being written, cryo-EM reconstructions of the capsid interior were available for only two dsRNA viruses belonging to the same family (*Reoviridae*): CPV [100, 101] and ARV [102, 103]. Members of *Reoviridae* (reoviruses) possess 9-12 genome segments in a 1:1 ratio to copies of the RdRP [9, 11]. Thus, their capsids contain 9-12 dsRNA molecules in which 9-12 RdRP copies are embedded, respectively, along with other proteins involved in assembly and/or replication. In contrast, the capsid interior consistency of dsDNA viruses studied by cryo-EM corresponds to only one dsDNA molecule and no proteins embedded in the nucleic acid. Another important difference between dsRNA and dsDNA viruses is that while the latter package nucleic acid in duplex form, the

former package single-stranded RNA, which only later gets replicated to form the duplex.

The reason behind the structural similarity between ϕ6 and dsDNA viruses and dissimilarity between ϕ6 and other dsRNA viruses may be explained by the unique composition of the capsid interior in ϕ6. Unlike the 1:1 ratio of segments to RdRP copies found in reoviruses, ϕ6 has about a 3/4:1 ratio leading to an almost equivalent number of RdRP copies but much fewer segments. In addition, while the average segment length in ϕ6 is much larger than the average segment length in reoviruses, the ϕ6 RdRP is much smaller than the reovirus RdRP (75 kDa vs 250 kDa). It may thus be possible that unlike reoviruses where the bulkier RdRPs orchestrate the arrangement of the capsid interior, in ϕ6 the larger genome segments are the ones responsible for this.

A different way of looking at these observations is by considering the nature of the polymers themselves rather than the composition of the capsid interior. Double-stranded nucleic acid can be of A-, B- or Z-form [128]. Whereas biological dsDNA is predominantly found in B-form, biological dsRNA exclusively adopts A-form helices due to the additional 2'-OH which prevents the C2'-endo sugar pucker required for B-form from occurring [140, 141]. A-form nucleic acid tends to have higher persistence length than B-form presumably due to increased rigidity conferred by the shorter and wider nature of the helical turns; indeed, A-RNA does have a higher persistence length than B-DNA (~60-65 nm vs ~40-50 nm, both dependent on buffer conditions) [142-144]. This in turn means that non-segmented dsDNA viral genomes have a much higher ratio of physical length (i.e. 44.0 kbps in P22 phage and 48.5 kbps in λ phage) to polymer persistence length compared to both ϕ6 (13.3 kbps) and reoviruses (18.0 - 25.0

kbps) though the diameter of capsid interior is reasonably similar (P22 55 nm, ϕ6 44 nm and RTV 50 nm, respectively). Furthermore, the longest ϕ6 segment (6.4 kbps) though larger than the longest segment in *Reoviridae* members (~4.0 kbps), is significantly shorter than any non-segmented dsDNA genome. Taken together, these observations alone would suggest that ϕ6 should be more similar to reoviruses than dsDNA viruses. However, this contradicts all reported structural observations. Hence, it is possible that the composition of the capsid interior plays a more important role than the nature of the nucleic acid polymer in the overall architecture of the virions.

The hypothesis that capsid content dictates genome organisation is further supported by the inability to resolve the ϕ6 RdRPs in spite of the dsRNA being resolved well-enough to allow fitting of atomic models representing over 90% of the entire genome. The only internal regions of the capsid devoid of dsRNA are the spool ends and cylindrical interior of layer 5. This suggests that it is within these areas that the RdRPs may localize, though with a high degree of flexibility in binding the capsid and/ or high degree of distributional heterogeneity. In CPV [100, 101], ARV [102, 103], RTV [59, 68] and BTV [110, 139], the RdRPs are attached to the capsid interior to allow the newly synthesised (+)ve-sense ssRNAs to be extruded from the capsid during viral transcription. Previous work on the structure of the unexpanded ϕ6 polymerase complex established and later mutagenesis work confirmed that in the absence of RNA, the RdRP contacts two different P1 molecules across the icosahedral 3-fold [44, 45]. However, the capsid expansion occurring during RNA packaging [13, 25] separates these two P1 molecules rendering the RdRP unable to maintain both P1 binding sites. Combined with the likelihood that the RdRPs can form non-specific charge-based

interactions with the dsRNA, this may provide the reason behind the inability to resolve the RdRPs embedded in the dsRNA within the expanded capsid. All things considered, the most plausible scenario is that the RdRPs localize in the proximity of the capsid within the areas at the spool ends devoid of ordered dsRNA. For transcription to occur, this would incur that the ends of at least one if not all segments have to be in the proximity of the dsRNA spool ends.

Besides the inability to locate and quantify the number of RdRP copies, another major caveat of the study is the inability to distinguish the three genome segments S (2.9 kbps), M (4.0 kbps) and L (6.4 kbps). This inability is even more pregnant because of the significant differences among the segment lengths which should in theory aid discerning them. The reason is that the structures resolved remain incomplete. Despite being able to describe the structural heterogeneity of the encapsidated nucleic acid by resolving a multitude of conformations, determining the total number of possible conformations remains a difficult task. There are two possible reasons for why this is the case. On one hand, it is possible that resolving all possible conformations is limited by the averaging process employed in determining structures by single-particle cryo-EM, which would suggest that a potentially much larger dataset would be required to achieve this. On the other hand, it is impossible to exclude the possibility that each individual virus may have a unique genome organisation, slightly different from the rest. If this is the case, resolving each individual organisation is rendered impossible by single-particle cryo-EM regardless of dataset size. Biology and statistics seem to favour the latter of the two reasons. Biologically, the dsRNA organisation is the result of genome replication or (-)ve-sense ssRNA synthesis occurring with the packaged (+)ve-

sense ssRNA molecules as templates. Before replication, the interior of the capsid contains the 8-12 RdRP copies, the three (+)ve-strand RNAs and the accessory protein P7 in an unknown copy number. This structure is presumably highly heterogeneous. As neither the RdRP nor the P7 is visible in the final structure, one can assume that their role in the formation of the final structure is negligible. If this is the case, the 3 ssRNA molecules being polymerised, the three nascent RNA duplexes and the P1 residues capable of interacting with the RNA are the only components influencing the formation of the final structure. Statistically, one can imagine that with such limited constraining factors and with such a variable starting point, the odds of reaching an identical configuration on two separate occasions is extremely slim. Nonetheless, the structures determined herein by averaging multiple images together are still predominantly well-resolved suggesting that the differences between the possible final sub-conformations are in fact relatively subtle. One plausible explanation for this observation is that the final dsRNA density is high enough to limit the structural differences between sub-conformations to local rather than global arrangement of strands. That being said, identifying the three genome segments remains an issue that the structures alone cannot achieve without input from orthogonal methods such as molecular dynamics simulations.

Proposing a mechanism for the formation of the final structure requires making several assumptions. First, the capsid interior structure pre-replication is either disordered or at least drastically more heterogenous than the final structures. Second, the three segments begin to be replicated more or less simultaneously as there is no available data that can argue the opposite (i.e. presence of virions with one segment in

ssRNA form and another in dsRNA form). Third, the rate at which the segments are being replicated is more or less constant. Fourth, the positioning and movement of the nascent duplexes is dictated mostly by the energetics of the assembly. If these assumptions hold, it is almost guaranteed that the three nascent duplexes will start filling layer 1 first as the radius of curvature is highest and layer 1 is the only one that can form favourable non-covalent interactions with the inner capsid shell. However, the high degree of curvature required to fill the spool ends may force the duplexes to fill a region around the spool equator in layer 1 and then transition to the equivalent region in layer 2 instead of completely filling layer 1 before moving to layer 2. Indeed, the observation of multiple transitions between layers 1 and 2 supports this hypothesis. This trend of inter-layer transitioning should apply to all pairs of adjacent layers, possibly even for reverse transitions from inner to outer layers. However, what happens next is slightly harder to predict as it remains unclear whether or to what extent stacking strands of similar curvature within a layer across the spool axis is more favourable than across layers at a given position along the spool axis. Nevertheless, one can imagine that the very ends of the spool which require the highest degree of dsRNA curvature are energetically the least favoured regions and will thus be filled towards the end of replication. This is supported by the observation that the density is poorest at the spool ends indicating a less ordered arrangement of strands. If this is indeed the case, it is likely that the ends of the L segment, which is significantly larger than S and M, will have to exclusively fill the regions at the very ends of the spool. Replication of the L segment is the only absolute requirement for transcription initiation [145] suggesting that it is indeed possible that L accounts for the highly curved dsRNA at the spool ends.

Assuming that the RdRPs do indeed localize around the spool ends, segment L being responsible for the dsRNA present in these regions would likely facilitate infection as L contains the genes for all the PC components. The fact that transcription of L is highly active early in the infection and significantly less active later in the infection [146] further supports this claim.

The remaining major feature of the structure not discussed thus far is the peculiar arrangement of the dsRNA helices in layer 5 and the lack of density beyond this layer. It is reasonable to expect that the dsRNA polymer is unlikely to adopt the extreme degree of curvature needed to follow the same pattern as in layers 1-4 at the radius of curvature corresponding to layer 5 and beyond. If this and the previous statements about L are accurate, it is possible that L transitions via layer 5 between spool ends in its quest to fill all remaining energetically unfavoured areas of the capsid interior. As it cannot be concluded that it is impossible for some of the RdRP copies along with (some of) the P7 copies to localize to the inside of layer 5, it remains possible that they do thus providing the “spindle” for the spool to layer itself on.

Condensed double-stranded nucleic acid forms liquid crystals [133, 147, 148]. Liquid crystalline phases refer to the phases assumed by a molecule as its concentration increases from the liquid (isotropic) phase to the solid (anisotropic) phase. Transitions between liquid crystalline phases in double-stranded nucleic acid occur at relatively discrete concentrations [147]. Currently a substantial number of studies exist into the spatial arrangement of molecules and concentration corresponding to individual liquid crystalline phases: pre-cholesteric, cholesteric and columnar hexagonal [149]. These studies have exclusively focused on purified linear dsDNA in solution at various

concentrations controlled by buffer conditions. Both short (~50 nm in length) and long (in the μm range) molecules have been analysed. This is important as the persistence length of DNA is ~40 nm [143] and polymers of length similar to the persistence length behave as rigid rods whereas polymers of much higher length than the persistence length are drastically more flexible [150]. Interestingly, the liquid crystalline properties of nucleic acid tend to be more or less similar regardless of the sequence length [149]. However, direct observations of the behaviour of any nucleic acid molecules confined to physiological containers are lacking.

The cryo-EM structures of the encapsidated ϕ6 genome described herein represent the first account of a holistic characterisation of nucleic acid in native environment. This allows the interpretation of a dsRNA model in lieu of the analysis previously performed on purified dsDNA segments and speculating about the biological implications of the liquid crystalline properties of viral genomes in double-stranded form. As previously discussed, dsRNA is predominantly found in A-form with a higher persistence length than B-form dsDNA. Whether this difference is consequential remains unclear given that the ϕ6 segments are much larger than either persistence length but the diameter of the ϕ6 capsid interior is smaller than the dsRNA persistence length. Nevertheless, there are several striking differences between the results in this chapter and previous studies. First, the nearly complete dsRNA model presented here confers the ability to characterise “individual” nucleic acid assemblies. In other words, unlike characterising the holistic properties of a unspecific mixture of an unknown number of long molecules, our model describes the specific assembly of the three linear genome segments of ϕ6. Second, the viral genome confined to the roughly spherical

contour of the capsid interior and capable of forming non-covalent interactions with the capsid is expected to behave differently than if it was simply solubilised in buffer at a similar concentration. Third, the particularities of the ϕ6 replication cycle are likely to influence the arrangement of the nucleic acid.

As mentioned previously, both short and long DNA molecules have been previously studied and the liquid crystalline phases formed are identical regardless of length. However, the concentrations at which the phase transitions occur are well-documented only for short DNAs as obtaining highly concentrated solutions of long DNAs is significantly more cumbersome [149]. That being said, it is known that the critical concentration (concentration at which the isotropic-anisotropic change occurs) significantly decreases with the increase in the polymer length and is thought to remain more or less constant at lengths much larger than the persistence length [149]. This trend tends to apply to the boundary between cholesteric and hexagonal phases with values of 380 mg/mL [149] reported for short DNAs and one value of 330 mg/mL reported for long DNA [151]. As these as well as all other measured parameters depend on buffer composition, they may only serve for the purpose of rough comparisons to ϕ6 values. Considering the packing density alone, it would appear that the ϕ6 genome approximates more accurately to columnar hexagonal rather than cholesteric phase as the majority of regions have densities higher than 330 mg/mL (**Fig. 3.5**). As the average density decreases from layer 1 to 4 (5 to be discussed later) and does not follow an exact pattern within layers, it is plausible to assume that overall the ϕ6 genome most likely forms a mixture of cholesteric and columnar hexagonal phases within the confines of the capsid.

In terms of architecture, layers 1-4 have similar average inclinations in relation to the spool axis (81°, 77°, 77° and 82° respectively), significantly different than that of layer 5 (36°) and will thus be analysed separately from layer 5. In the cholesteric phase, the DNA layers adopt a helical architecture i.e. the average inclination across layers increases progressively. In contrast, columnar hexagonal phases display no change in inclination across layers i.e. the layers are parallel and the only twist occurs along the double-helix axis. Layers 1-4 thus present similarities and differences compared to both these phases suggesting once again that the ϕ6 may be a biphasic mixture. On one hand, dsRNA helices across layers have very similar inclinations, particularly around the spool equator. However, instead of intercalating to form a typical columnar hexagonal arrangement, these strands run almost stacked onto each other therefore adopting a cubic geometry. On the other hand, moving from the spool equator towards the spool ends, the degree of cubicity decreases and that of hexagonality increases. Importantly, hexagonality here refers to the geometry of the strand arrangement rather than the liquid crystalline phase. This is reflected in the fact that it is the varying strand inclination across layers that leads to 2D planes in which the strands form a hexagonal pattern and not the intercalation of strands with similar inclination. In agreement with these observations is the distribution of local packing density and curvature across the capsid interior. The regions of highest packing density display hexagonal geometry and higher curvature. Notably, these regions are limited to short stretches highlighting the fact that the inclination of the corresponding strands across layers is indeed different.

One minor caveat to the analysis of packing geometries is the absence of modelled interlayer transitions which are responsible for additional geometric

distortions, namely from hexagonal to cubic geometry, increased packing density and likely increased local curvature. Nevertheless, the correlation of curvature, density and geometry suggests that the non-uniform curvature of the packaged dsRNA might be the main driving force for the observed genome architecture. This curvature non-uniformity is naturally caused to a large extent by the confinement of the dsRNA to a spherical volume and possibly to a smaller extent by local capsid interactions. As these factors have not been addressed in other studies, it remains to be determined whether they are are indeed the main cause behind the discrepancy in architectures when compared to the cholesteric and columnar hexagonal phases. That being said, it remains to be discussed how layer 5 fits this picture. Layer 5 is peculiar for a number of reasons, namely in that it has a much lower average density, curvature and average inclination compared to layers 1-4. This suggests that layer 5 might not be part of the assembly formed by layers 1-4 and rather influenced to a higher extent by either the incapability of the dsRNA to follow the same arrangement at such small curvature radii and/or by the potential presence of RdRP and P7 copies to its inner side.

4. Nucleocapsid restructuring induces ϕ6 transcription

4.1. Foreword

This chapter represents the natural progression from Chapter 3 by tackling ϕ6 transcription, the step in the replication cycle immediately following the delivery of the QNC into the host cytoplasm. The two main aspects tackled in this chapter are the disassembly of the outer capsid shell and the activation of transcription from the RdRP point of view. Aided by two substantial cryo-EM datasets, a plethora of intermediates in the disassembly process are characterised in addition to the first structure of the mature virion core that permits atomic modelling of P1. The previously unresolved RdRPs are also identified and shown to localize at the spool ends of the dsRNA. Pieced together, these structures describe the complex changes activating transcription in ϕ6.

4.2. Summary

Chapter 4 begins by describing the icosahedral structure of the transcribing ϕ6 nucleocapsid or double-layered particle (DLP) and virion core or single-layered particle (SLP). Symmetry relaxation is then employed to resolve asymmetric structures of nine P8 disassembly intermediates. Also by symmetry relaxation, the structure of the capsid interior is determined separately in the DLP and SLP. These two maps are then subjected to localized classifications to discern a total of four unique RdRP arrangements. Finally, a mechanism for transcription activation is proposed and discussed in relation to ϕ6 and dsRNA viruses in general.

4.3. Structure of the protein shells

Three main types of particles were observed in both TNC and DTNC datasets (**Tables 2.1** and **2.2**), namely double-layered particles (DLPs), single-layered particles (SLPs) and disassembly-intermediate particles (DIPs). As DIPs appear similar in architecture to DLPs, the two types were processed together and separately from SLPs. For simplicity, the DLP+DIP set will be referred to as DLP. The icosahedral structure of QNC is identical to the DLP structure in both the TNC (5.6 Å) and DTNC states (3.7 Å), while the SLP structure is also identical in the TNC (5.6 Å) and DTNC (3.7 Å) states (**Fig. 4.1**). However, the inner capsid shell is more expanded in SLP than in DLP. This increased expansion is the result of a ratcheting movement in each P1 asymmetric dimer by which P1B remains unmoved whereas the 5-fold proximal P1A swings away from the capsid interior by ~15 Å (**Fig. 4.2**). This new position of P1A is incompatible with the presence of a P8 shell due to a multitude of clashes between the two proteins.

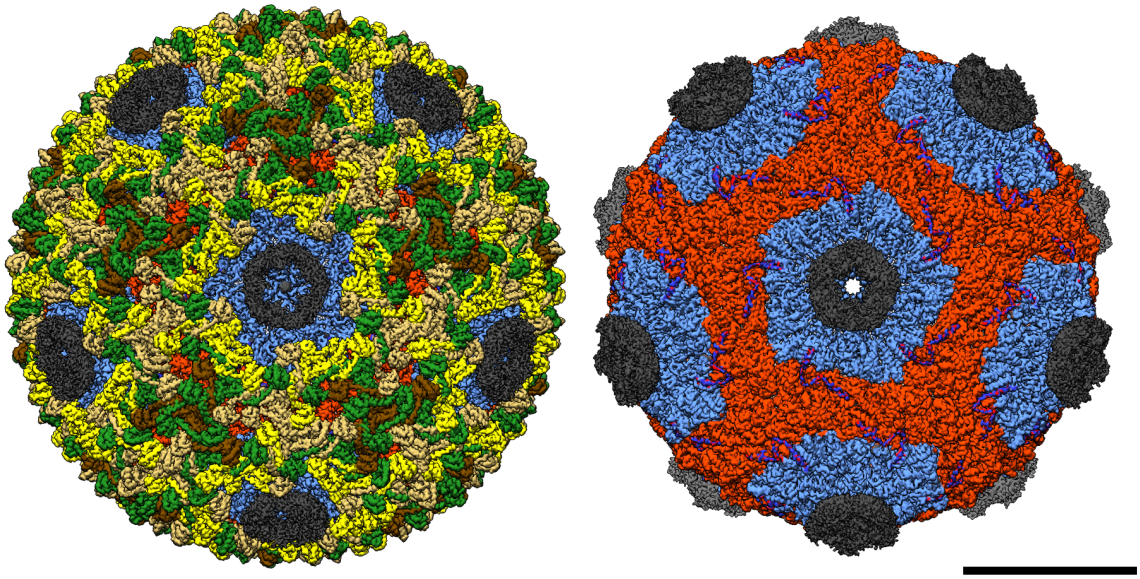


Figure 4.1. Icosahedral structures of the ϕ6 DTNC DLP and SLP
DLP (left) and SLP (right) displayed according to the colour scheme in Chapter 3. FSC plots can be found in **Appendix 1.5**. Scale bar, 200 Å.

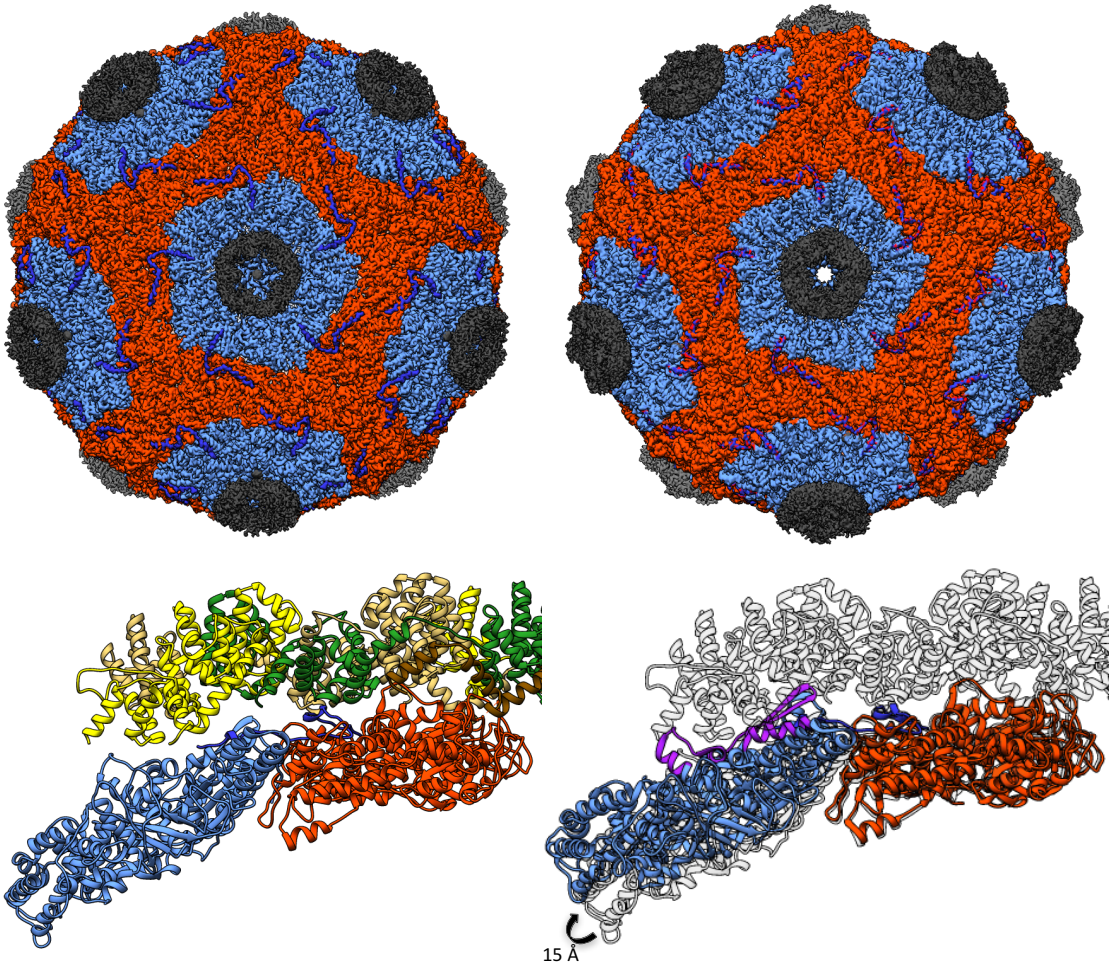


Figure 4.2. Comparison of the inner capsid shell in ϕ6 DLP and SLP
DLP full capsid with hidden outer capsid shell (top left), DLP asymmetric unit (bottom left), SLP full capsid (top right) and SLP asymmetric unit displayed according to the colour scheme in Chapter 3. In transparent grey is the superposed DLP asymmetric unit and in magenta are the regions of SLP P1A that would clash with the outer capsid shell. Scale bar for bottom panels, 50 Å.

4.4. Structures of the P8 disassembly intermediates

The DLP population was subjected to symmetry relaxation in order to pursue the asymmetric structure of P8 disassembly intermediates. Nine intermediates are observed, displaying a complete P1 shell and a variable degree of remaining P8 shell (*I1-9* as in

Fig. 4.3, FSC plots in **Appendix 1.6**). Despite not necessarily depicting a disassembly timeline, the presence of these nine intermediates indicates a potential stepwise mechanism for P8 disassembly. Unexpectedly, only one of the 20 3D classes used in 3D classification exhibits a roughly intact P8 shell (*I1*). *I4* exhibits a disassembly roughly centred on a 5-fold axis and contains 31% of the particle population, the highest proportion by a large margin. *I3*, *I6* and *I9* lack one of the P4 hexamers at or slightly offset from the origin of the disassembled region suggesting its possible shedding from the capsid. In *I3* the P1 shell at the 5-fold location lacking the P4 resembles the P1 shell in the DLP (coloured blue in **Fig. 4.3**). In contrast, in *I6* and *I9* the 5-fold in question is as expanded as in the SLP (coloured dark grey in **Fig. 4.3**). Analysing all nine intermediates, it becomes apparent that the ones lacking a larger portion of the outer shell have an increased number of inner shell 5-folds in the expanded state. Specifically, in *I4-9* the regions proximal to the disassembly origins are predominantly expanded, whereas those proximal to the remaining outer shell are not. This suggests that the inner shell expansion and outer shell disassembly may be correlated.

In addition to *I1-9*, three additional classes were resolved that exhibit incomplete P8 and P1 shells at the map threshold of 0.01 used for *I1-9* (**Fig. 4.4** and **Appendix 1.7**). Upon lowering the threshold even further to 0.001, the second and third maps still show incomplete P1 shells. Notably, in these 3 maps even at 0.01 threshold the P8 shell is predominantly present and appears to be missing only in the region also missing P1. It may thus be possible that P8 disassembly can erroneously force the disassembly of P1 in some particles. Alternatively, the Triton extraction performed to remove the lipid envelope may have caused these artifactual particles.

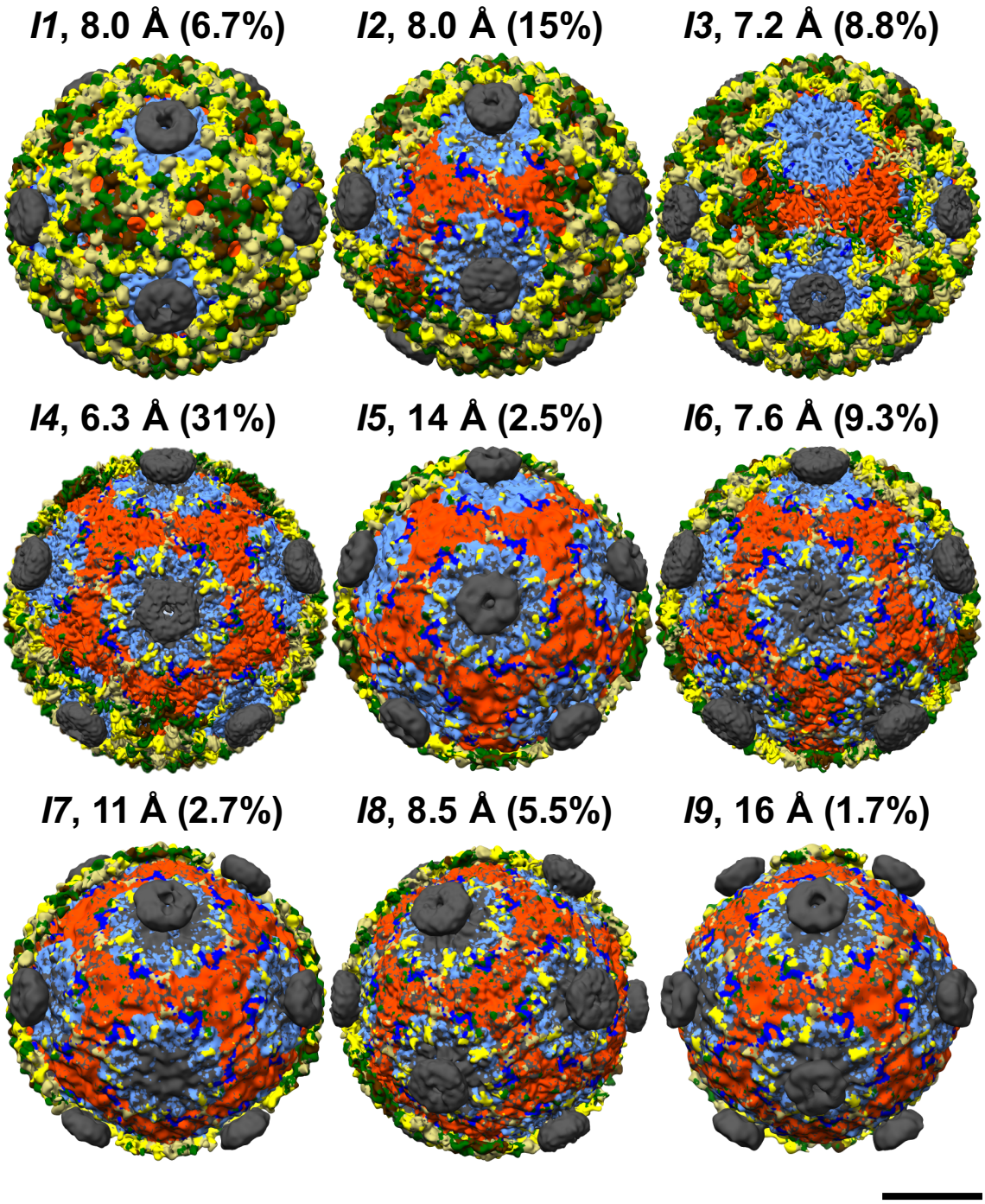


Figure 4.3. Structures of the outer shell disassembly intermediates in ϕ6 DTNC
The nine main asymmetric reconstructions as obtained by 3D classification (see Chapter 2) are ranked from least to most disassembled. They are presented at the same threshold level (0.01) and are shown along the view axis at the origin of the disassembled region. The resolution obtained after 3D refinement and post-processing along with the corresponding particle percentage of each class are also indicated. Regions in grey correspond to density farther than 5 Å from the capsid shells of QNC. FSC plots can be found in **Appendix 1.6**. Scale bar, 200 Å.

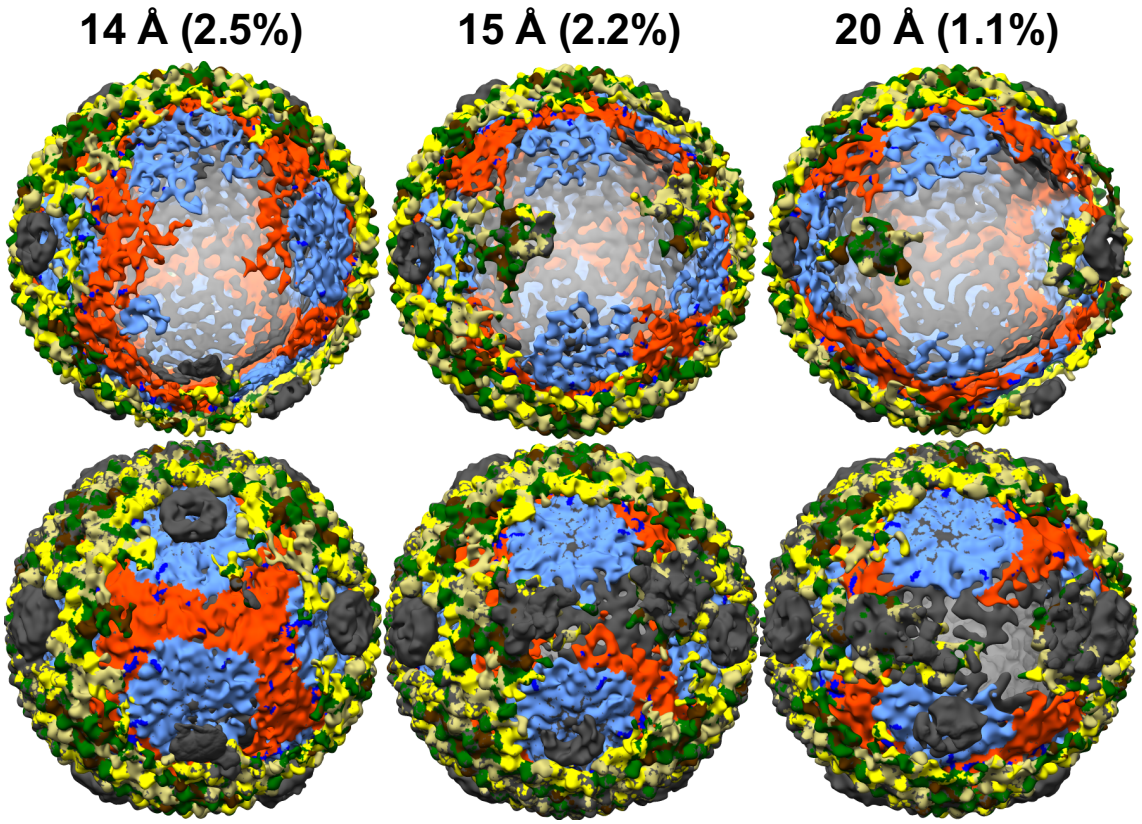


Figure 4.4. Structures of peculiar P8 shell disassembly intermediates in ϕ6 DTNC
Three reconstructions corresponding to intermediates with incomplete P1 shells. The maps are rendered at 0.01 threshold (top row) and 0.001 threshold (bottom row). Regions in grey correspond to density farther than 5 Å from the capsid shells of QNC.

4.5. Genome organisation

The organisation of the capsid proximal dsRNA in both DLP and SLP is highly similar to the pseudo-D3 structure observed in QNC (**Fig. 4.5**). Nonetheless, the previously unresolved spooled ends each contain 3 copies of the RdRP P2. These 3 copies together with a triskelion-like formation that appears to be anchoring the RdRP molecules form a unique arrangement in both the DLP and SLP structures. This novel framework is hypothesised to consist of the accessory protein P7 whose localisation within the NC has not been previously addressed. In addition, SLP exhibits additional density compared to DLP in the region proximal to the pseudo-C3 axis.

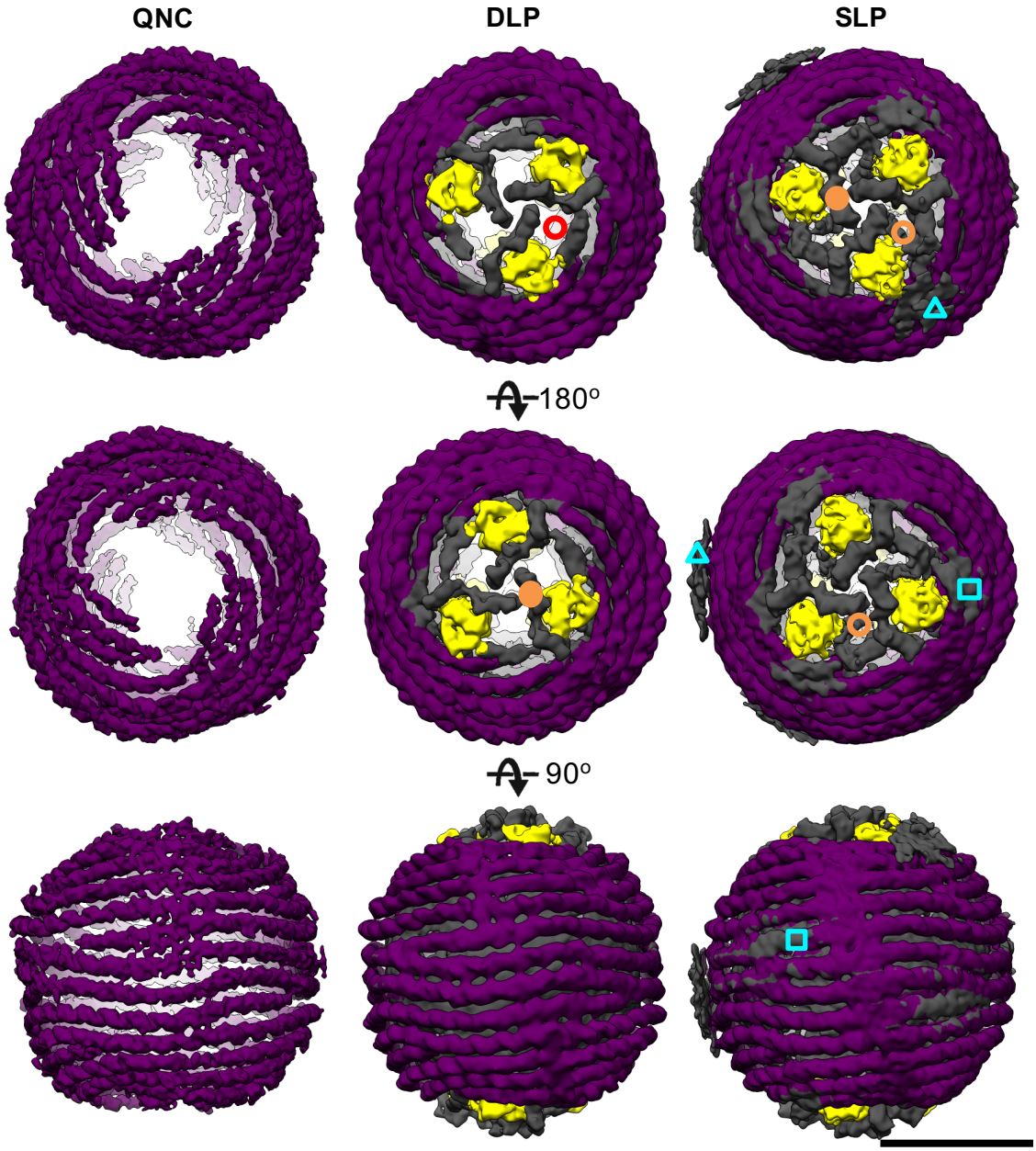


Figure 4.5. Structures of the DLP and SLP capsid interiors in ϕ6 DTNC
DLP and SLP genome structures compared to the QNC pD3' structure. Views are along the pseudo-C3 axis (top and middle) and along one pseudo-C2 axis (bottom). The dsRNA model of layer 1 and models of fitted RdRPs were used to colour the DLP and SLP maps in dark purple and yellow, respectively. In dark grey is density farther than 8 and 10 Å away from dsRNA and RdRPs, respectively. The triskelion formation putatively consisting of protein P7 is denoted by full orange circle and the density around the pseudo-C3 axis present only in SLP and not in DLP is marked by empty orange circle (empty red circle is the corresponding region in DLP). In SLP, several regions of the dsRNA corresponding to the expanded 5-folds are dark grey (square) along with several regions outside the dsRNA spool which represent capsid remnants (triangle). The P4 hexamers are not shown. Scale bar, 200 Å.

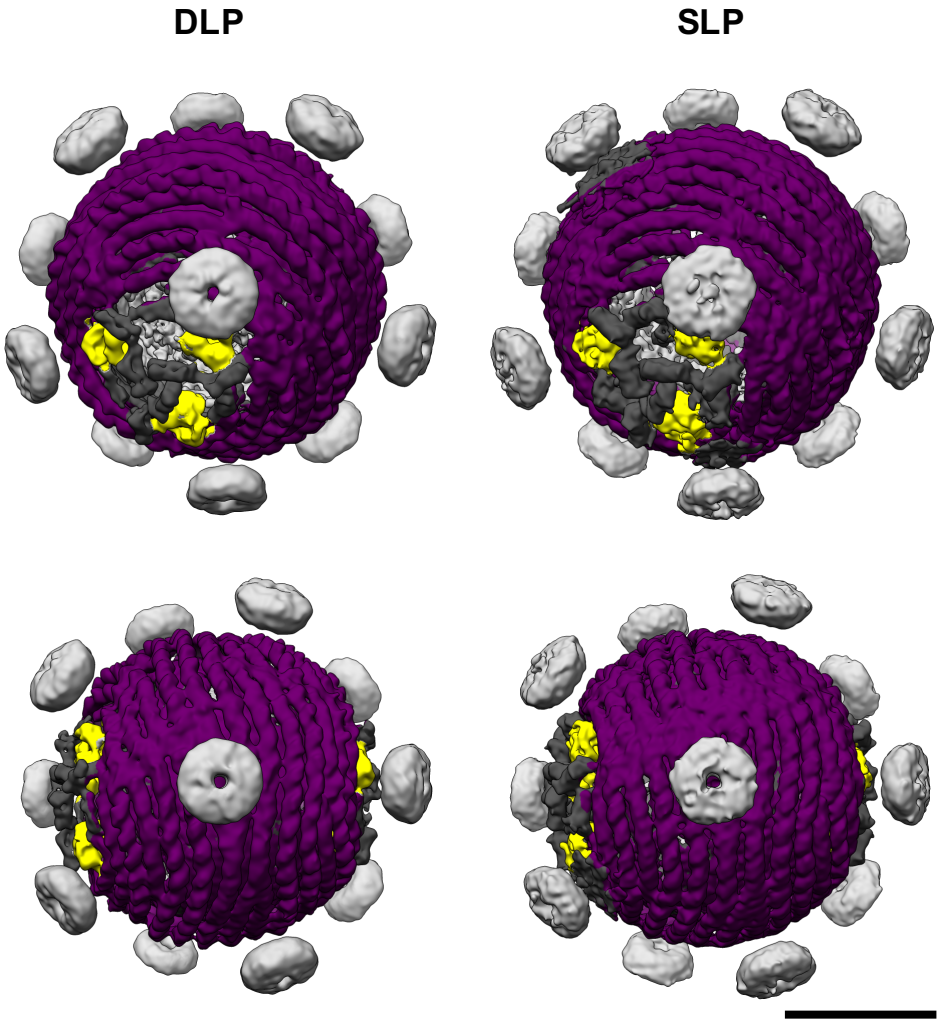


Figure 4.6. P4 structures in ϕ6 DTNC DLP and SLP
Representative P4 morphologies (light grey) in DLP and SLP at the 5-folds proximal to the spool ends (top row) and at those proximal to the spool equator (bottom row). Remaining components coloured according to the scheme in **Fig. 4.5**. Scale bar, 200 Å.

In regards to P4 structure, it is worth mentioning that despite not applying any symmetry on the genome reconstructions, the P4 hexamers are still symmetry mismatched; this indicates the lack of symmetry coherence between the capsid interior and RNA packaging motors. In spite of the symmetry mismatch, it is apparent that the P4s at 5-folds proximal to spool ends have filled pores in SLP and hollow pores in DLP, whereas the ones distal to the spool ends are hollow in both SLP and DLP (**Fig. 4.6**).

4.6. RdRP arrangement

The pseudo-D3 symmetry of the SLP and DLP genome organisations and the lessons from the QNC processing suggested the presence of potential incoherent averaging in the reconstructions. Indeed, localized classification of the spool ends relaxing C3 symmetry distinguishes two predominant RdRP arrangements in both DLP and SLP (**Fig. 4.7** and **Table 4.1**, FSC plots can be found in **Appendix 1.8** and **Appendix 1.9,** respectively). DLP gives rise to one arrangement with only one RdRP copy and another with three RdRP copies, whereas SLP presents two arrangements each with three RdRP copies. The only RdRP present in DLP #1 and the ones occupying the equivalent site in DLP #2, SLP #1 and SLP #2 will be denoted by the suffix '.1' (i.e. DLP#1.1) with the remaining two RdRPs denoted with the suffixes '.2' and '.3' in clockwise direction. Notably, DLP #1.1, DLP #2.1-3, SLP #1.1 and SLP #2.1-2 are resolved equally well and oriented roughly identically. SLP #1.2 and SLP #2.3 are roughly equally well resolved though oriented differently compared to each other and to the aforementioned ones. SLP #1.3 is significantly worse resolved and its orientation is strikingly different than all others. The model used as reference is the X-ray structure of the ϕ6 RdRP trapped in the initiation stage [32]. The model is coloured by domain, namely red for the fingers domain (1-30, 104-276, 333-397), green for the palm domain (277-332, 398-517), blue for the thumb domain (37-91, 518-600) and yellow for the C-terminal domain (601-664).

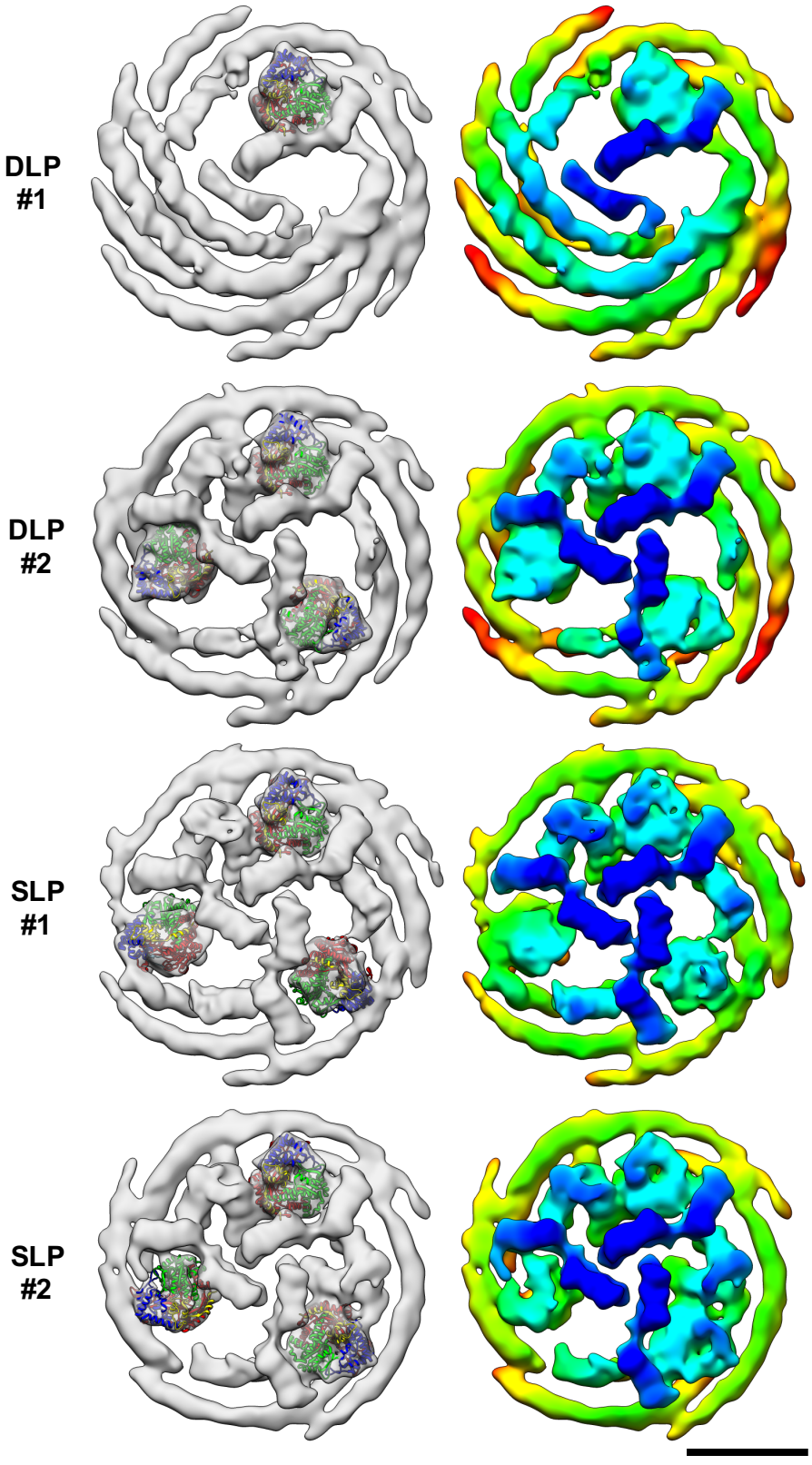


Figure 4.7. RdRP dominant arrangements in ϕ6 DTNC DLP and SLP
The two main arrangements of spool ends in DLP and SLP shown with fitted RdRP models (left) and coloured by height from red in the background to dark blue in the foreground (right). Rendered are the maps output by 3D classification. Scale bar, 100 Å.

Table 4.1. Summary of characteristics for the RdRP arrangements in ϕ6 DTNC
Shown are the number of particles, percentage of particles subjected to the classification, final resolution achieved after refinement and post-processing along with cross-correlations to RdRP models for the four RdRP arrangements.

	Particles	Particle %	Resolution (Å)	RdRP 1	RdRP 2	RdRP 3
DLP #1	40,000	28.1	6.9	0.76	-	-
DLP #2	32,000	22.8	7.5	0.76	0.71	0.77
SLP #1	42,000	30.5	6.6	0.76	0.53	0.25
SLP #2	32,000	24.3	7.3	0.73	0.73	0.48

The P7 triskelion formation consists of three roughly dimeric subunits and appears fully assembled in DLP #2, SLP #1 as well as SLP #2 and only partially assembled in DLP #1. Each dimeric subunit of the formation contains one monomer in the immediate proximity of both the spool axis and the capsid likely forming extensive contacts to the P1 shell. The second monomer interacts with one RdRP via the palm domain (**Fig. 4.8**). This observation is in agreement with a previous report suggesting that P2 and P7 share the same binding sites to P1 in PC [50]. P2 also forms a contact via the fingers domain to a neighbouring dsRNA strand and in some cases to the U-shaped density that was present in the SLP and absent in the DLP global reconstructions (**Fig. 4.5**). This U-shaped density is not present in DLP #1 and present once in DLP #2, once in SLP #1 and twice in SLP #2. In PC, it was shown that the RdRP has separate binding sites to two distinct P1 dimers, namely W570 of the thumb domain and E444 and R470 of the palm domain [44, 45]. W570 is located in the region of the RdRP closest to the capsid, putatively forming a similar contact to P1 as in PC, while E444 and R470 appear to contact P7.

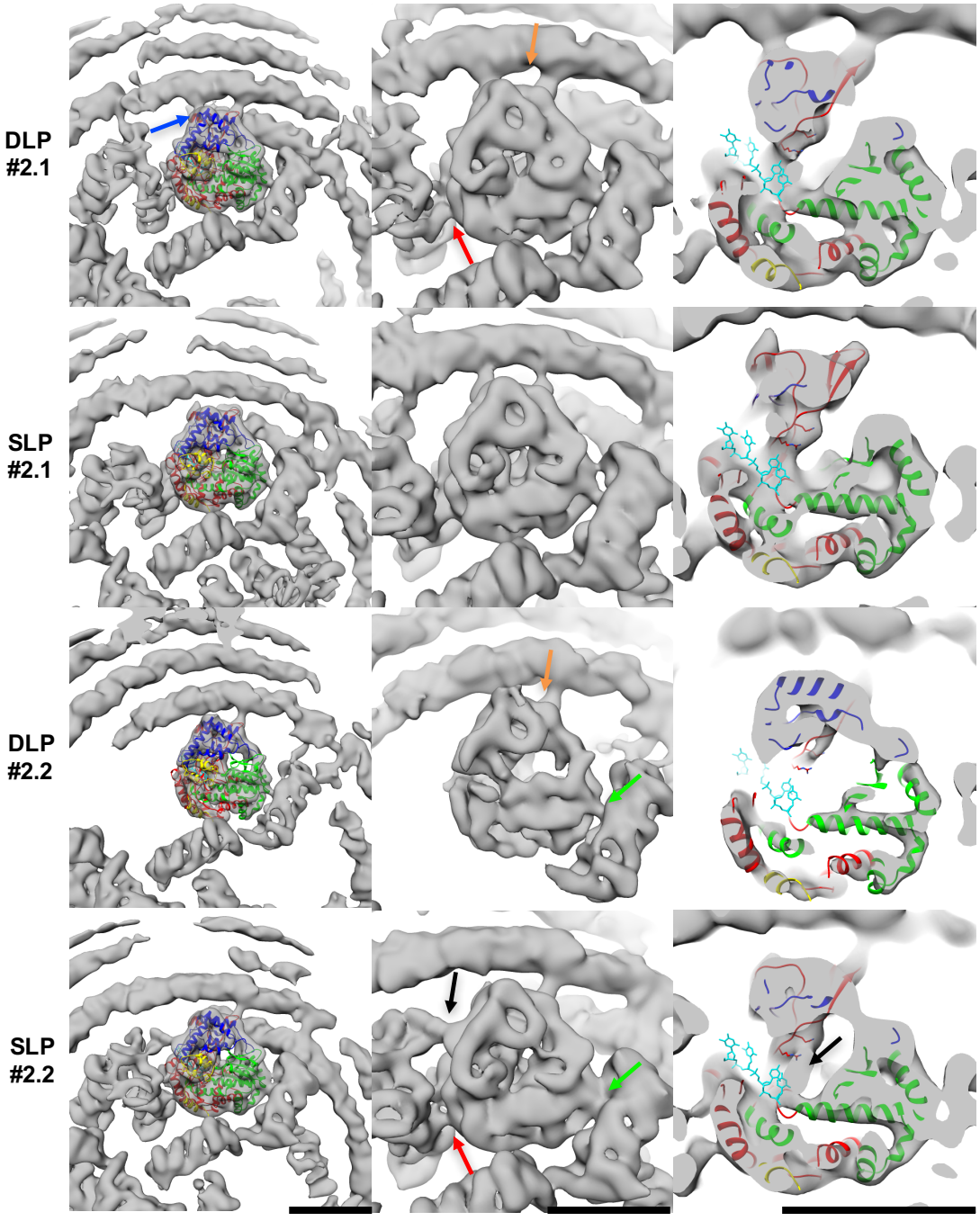


Figure 4.8. Structures of four RdRPs in ϕ6 DTNC
The DLP #2.1, SLP #2.1, DLP #2.2 and SLP #2.2 RdRPs are shown from a zoomed-out view (left column), an intermediate view (middle column) and a cut-open zoomed-in view (right column). The left and right columns are rendered at a higher threshold (0.030), while the middle column is rendered at a lower threshold (0.025). The template is coloured in cyan. The RdRP sites contacting neighbouring molecules are indicated by coloured arrows: P1 via thumb (blue), U-shaped density via fingers (red), dsRNA via fingers (orange) and P7 via palm (green). Densities found only in SLP #2.2 (black arrows) are present at the template entry site and next to the active site, respectively. Rendered are the maps output from refinement and post-processing. Scale bars, 50 Å.

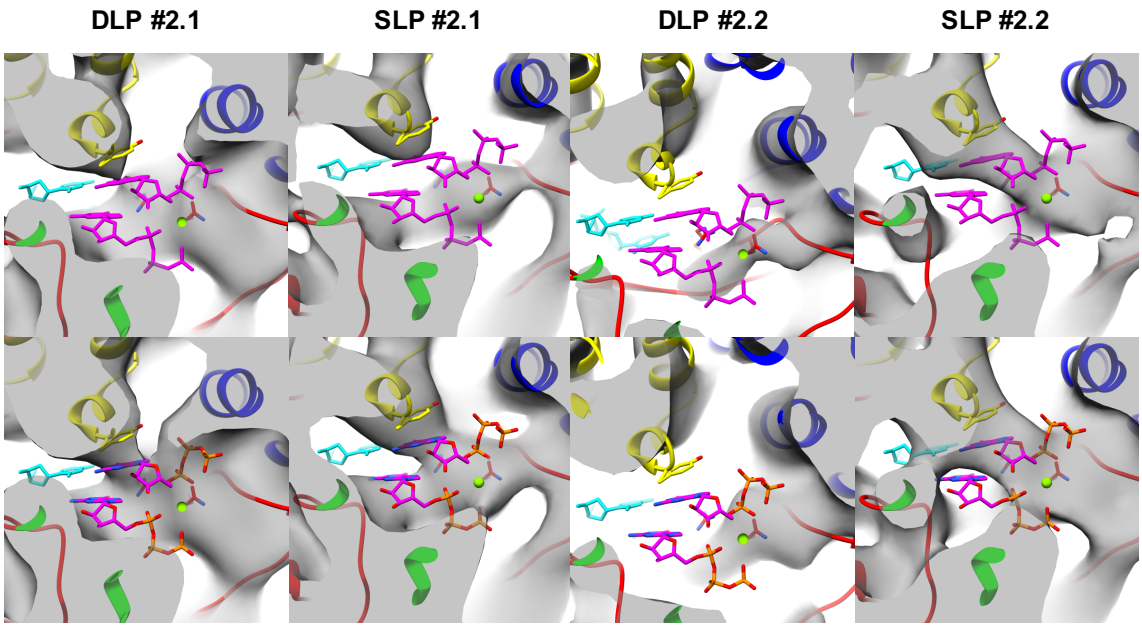


Figure 4.9. Active sites configurations of the same four RdRPs in ϕ6 DTNC
Zoomed-in views on the active sites of the four RdRPs shown in **Fig. 4.8**. The maps in the top row are rendered at the higher threshold (0.030) and the ones in the bottom row are rendered at a lower threshold (0.025). Two GTPs thought to be part of the initiation complex in the X-ray structures used as reference are coloured in magenta. Mg^{2+} ions are shown in light green. The RdRP residues indicated are Y630 which base-stacks to one GTP as well as R204 and 206 which are also proximal to the same GTP.
.

Further analyses of the RdRPs reveal a few noteworthy aspects. Unlike all other RdRPs, SLP #2.2 displays strong additional density in the proximity of the template entry channel at both the higher and lower thresholds (**Fig. 4.8**). In addition, even at the higher threshold SLP #2.2 presents density potentially corresponding to a GTP molecule (**Fig. 4.9**) that was shown to be part of the initiation complex [32]. Other RdRPs containing a proximal U-shaped density such as DLP #2.1 and SLP #2.1 manifest density only at lower thresholds that is still different than in SLP #2.2. The RdRPs lacking the U-shaped density such as DLP #2.2 do not present any density in the active site and are significantly displaced away from the capsid compared to the other RdRPs. Taken together, these results suggest that RdRPs may require both the ordered P7

formation and U-shaped density to be fully prepared for transcription; nonetheless, most RdRPs fulfilling these requirements are yet to be in a template-engaged state.

4.7. Discussion

Taken together, the structures determined in this chapter suggest a mechanism for transcription initiation in ϕ6 (**Fig. 4.10**). It has been shown previously that P8 disassembly is required for transcription to occur [152, 153] which can be triggered by lowered pH or decreased Ca^{2+} [19, 153, 154]. On the other hand, the RdRPs performing transcription are inhibited by Ca^{2+} and activated by Mg^{2+} [39, 53]. It is thus feasible that the Mg^{2+} in the transcription buffer used for the samples studied in this chapter outcompetes the Ca^{2+} needed for P8 stability. Destabilising the outer capsid shell allows the P1 shell to expand and kickstart the disassembly of P8. Meanwhile, the encapsidated copies of P2 and P7 start to rearrange at the dsRNA spool ends as a result of the conformational changes in P1. Though disordered in QNC, P2 and P7 which are thought to share the same P1 binding sites in PC [49, 50] have likely stayed attached to the P1 shell ever since the PC assembly; their arrangement is thus heavily impacted by the P1 expansion and by the dsRNA which is exclusively tightly packed with the exception of the spool ends. The rearrangement of the spool ends is complete once the entire P1 shell is expanded and the entire P8 shell is disassembled. The final configuration activates transcription and the parental dsRNA plus-strands are extruded from the P4 hexamers at the 5-folds proximal to the spool ends.

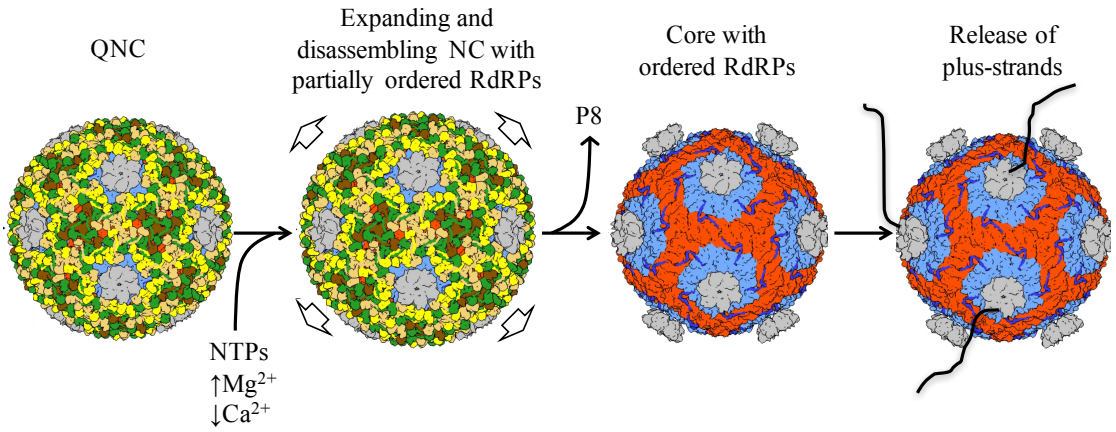


Figure 4.10. Proposed mechanism of transcription initiation in ϕ6
Components are coloured using the standard colour scheme and the extruded plus-strands are coloured in black

The main potential caveat of the study presented herein is that it is mostly based on a sample obtained by incubating QNCs only with A, G and U rather than with the full complement of nucleotides (i.e. DTNC). The first G is the 10th base from the 3' ends of all three ϕ6 segments meaning that transcription should at least in theory be triggered, continue for the first nine bases and then stop. This naturally restricts the amount of insights into transcription progression that can be gained. More importantly, though the structures obtained for QNCs incubated with all four nucleotides (i.e. TNC) appear similar to the ones shown here, the possibility that no artifactual events occur cannot be discarded with absolute certainty.

Additionally, a number of aspects regarding the mechanism proposed here cannot be accounted for with absolute certainty by the data presented herein. The dsRNA recognition by the RdRP, the separation of parental strands prior to template processing and path taken by the resulting dsRNA remain largely unknown. A study of purified ϕ6 RdRP by X-ray crystallography [32] has suggested that a 'plough' formation

separates the incoming duplex and that the emerging dsRNA exits the RdRP as the C-terminal domain rearranges, hypotheses that cannot be tackled using the data in this chapter. If the emerging duplex does indeed exit in this fashion, the structures here suggest that it would be pointed towards the capsid (**Fig. 4.7** and **4.8**), which seems problematic given that virtually no room is available between the RdRP and the capsid.

The precise role and functioning of P4 in transcription remains poorly understood. First and foremost, the implications of P8 disassembly and P1 expansion on the structure and activity of P4 are yet to be determined. Whether the enzymatic activity of P4 is required in mRNA extrusion in ϕ6 and the mechanism of mRNA extrusion itself remain poorly understood as well. ϕ6 nucleocapsids and cores are known to display NTPase activity [39], which is activated by Ca^{2+} and inhibited by Mg^{2+} [39, 53]; however, in ϕ12 it has been shown that P4 acts as a passive rather than active motor during transcription [155]. In addition, the number of P4 hexamers per capsid required for minimal and optimum transcription, respectively, remain elusive. The results here suggest that the six P4 hexamers distal to the dsRNA spool ends may not be needed at all for transcription and that the remaining six may be capable of acting concomitantly. The presence of P8 disassembly intermediates lacking the full complement of P4 supports the hypothesis that not all twelve P4s are required as some appear to incidentally detach from the inner shell. Reducing P4 incorporation to one copy per capsid has been shown to increase the packaging rate, preserve the replication rate and virtually abolish transcription [34]. This observation somewhat supports the previous ideas; however, alternative scenarios such as requiring at least one hexamer at each spool end are also plausible. These caveats as well as the ones discussed in the previous

paragraphs are expected to be at least partially elucidated by obtaining higher-resolution structures of both fully transcribing and defectively transcribing ϕ6 particles.

The ϕ6 RdRP displays striking dissimilarities to prototypical reovirus RdRPs in terms of structure, function and behaviour throughout the virus replication cycle. With the exception of some members of *Totiviridae*, *Birnaviridae* and *Partitiviridae* [156-160], cystoviruses are the only dsRNA viruses known to perform semi-conservative transcription [27-29], whereas reoviruses exclusively perform conservative transcription [69, 70]. The ϕ6 and reovirus RdRPs must thus possess distinct double helix unwinding mechanisms in order to use as mRNA the parental positive-strand and the newly synthesised positive-strand, respectively. Reovirus RdRPs are believed to recognise dsRNA termini using a cap binding site [68, 73, 161] and to hold them in a duplex-open state [68]; however, ϕ6 does not require mRNA capping and thus this option is not feasible. The results described in this chapter suggest the existence of cooperativity in the substantial RdRP rearrangements required for transcription initiation. In contrast, reovirus RdRPs appear to function more or less independently from one another [59, 68, 110]. This observation may be connected to the fact that while reoviruses have one RdRP copy per genome segment [100, 103, 139, 162], ϕ6 has multiple RdRPs capable of concomitantly transcribing a given segment [28]. In summary, it may be possible that transcriptional cooperativity in ϕ6 RdRPs may provide an alternative to the more complex recognition of each genome segment by one RdRP copy in reoviruses [32]. Nonetheless, better structures are required to elucidate how up to six RdRPs as observed here can function together. Whether or how this contrasting

aspect between ϕ6 and reoviruses is connected to the type of transcription performed remains to be determined as well.

Unlike reovirus RdRPs, the ϕ6 RdRPs undergo dramatic rearrangements during the virus replication cycle. The RdRP initially assembles together with P1 and P4 into a RNA-devoid PC that only subsequently acquires the genome by sequentially packaging the three segments [13, 37]. In contrast, reovirus packaging is widely believed to be the result of co-assembly between the RdRPs, capping enzymes, inner capsid proteins and genome segments [9, 11, 73]. Notably, the ϕ6 RdRPs get encapsidated due to direct contacts with the inner capsid [44, 45] and the distribution of copies within the capsid appears remarkably variable throughout the replication cycle. After being tightly bound to two distinct P1 dimers in PC, they lose one of these contacts during the maturation of PC into NC which confers them an extraordinary degree of flexibility. Upon P8 disassembly, they are then reanchored with the help of P7 as P1 expands further. In sharp contrast, reovirus RdRPs are believed to be indirectly encapsidated by associating with the genome segments and likely remain anchored to the capsid in more or less the same position throughout the replication cycle [59, 68, 72, 163]. The significantly reduced complexity and size of the ϕ6 RdRP (664 amino acids in length [32] compared to 1088 in RTV [73]) may facilitate these ample dynamics.

Finally, purified ϕ6 RdRP can perform both negative- and positive-strand synthesis [31, 164], whereas purified RTV RdRP cannot achieve either [73, 165] meaning that the two also have distinct regulatory mechanisms. A common theme in a number of dsRNA viruses including both ϕ6 and RTV is the presence of a T=13 capsid shell whose disassembly is required to activate transcription [19, 25, 60, 67, 166, 167].

In ϕ6 and RTV, this disassembly is facilitated by reduced Ca^{2+} levels [19, 26, 64]. However, while in RTV the shape of the DLP remains constant in the presence or absence of the outer shell [60, 67], the ϕ6 inner shell expands upon outer shell disassembly. The most likely explanation for this is that the RTV DLP is significantly more rigid than the ϕ6 inner capsid due to the simple fact that it consists of two separate protein shells that stabilise each other [67, 74]. Nonetheless, this difference may explain the discrepancy in RdRP regulation in the two viruses; whereas RTV prevents premature transcription by occluding the exit of the transcripts from the DLP, ϕ6 does so by holding the inner shell to which the RdRP is attached in a state that renders the RdRP inactive. Given the plethora of intermediates resolved in this chapter, it is likely that ϕ6 disassembly proceeds in stepwise fashion to facilitate the structural rearrangements of the capsid interior required to activate transcription. As no major rearrangements of the capsid interior are needed to kickstart RTV transcription, the outer shell disassembly may occur instantly after the prerequisites are met. If this is indeed the case, it may also explain the lack of available information on disassembly intermediates in RTV.

5. Limited yet discernible long range order inside rotavirus capsids

5.1. Foreword

This chapter marks the change of focus from ϕ6 NC to RTV DLP. Originally the project consisted of determining a high resolution structure of the encapsidated RdRP as well as elucidating the overall architecture of the dsRNA. However, soon before the writing of this thesis began, two papers reported the *in situ* structure of VP1 [59, 68] using the same procedures as those performed here. This part of the project will thus be only briefly touched on as the structure determined here is highly similar to the reported ones. The main focus of the chapter is the interplay between the VP1 copies and the dsRNA. This interplay turns out to be highly dynamic and problematic to tackle. Unlike ϕ6, where breaking the symmetry of the capsid suffices to resolve the majority of the capsid proximal interior, in RTV the resulting map represents the average of a myriad of different conformations and thus provides limited information. As this could not be predicted originally, a good amount of work was put into attempting to resolve a reliable global structure. After numerous failures, a separate approach was followed in which the global structure was split into subparts that were individually analysed by localized classifications [45]. Though this proved to be significantly more successful, the results provide only a rudimentary take on the intricacies of the VP1-dsRNA complex.

5.2. Summary

This chapter presents the structural characterisation of the RTV DLP capsid interior by cryo-EM and asymmetric reconstruction. The main section starts by briefly touching on the protein shells and continues with the partially resolved pseudo-D5 organisation of the capsid proximal dsRNA obtained by symmetry relaxation. This is followed by individual analysis using localized reconstruction of the 12 icosahedral 5-folds harbouring the RdRP binding sites. The plethora of different VP1-dsRNA conformations is then described along with the probabilistic analysis of correlations in VP1 orientations between neighbouring 5-folds. These results are discussed in the context of known RTV and genome structures, the biology of related viruses and implications for RTV replication and transcription.

5.3. Structure of the protein shells

In similar fashion to ϕ6, the icosahedral structure of the rotavirus DLPs has been reported on several occasions [67, 78] with the highest achieved resolution being 2.6 Å. The 3.4 Å resolution obtained here is poorer, as the main aim of the data collection scheme (i.e. linear mode, large pixel size) was not to obtain the highest possible resolution, but rather to obtain enough particles for the asymmetric reconstruction of the DLP interior. Aside from the resolution aspect, the structure determined here is in good agreement with previously reported structures and does not provide any significant novel insights into the capsid structure (data not shown). The outer T=13 layer of the DLP is formed by 260 trimers of protein VP6 (13 monomers per asymmetric unit) and the inner T=1 layer by 120 copies of protein VP2 (2 monomers VP2A and VP2B per

asymmetric unit). However, the DLP interior which deviates from icosahedral symmetry consisting of the 11 dsRNA segments, supposedly 11 RdRP copies (VP1) and supposedly 11 capping enzyme copies (VP3) remains unresolved [11].

5.4. Organisation of DLP interior

5.4.1. Breaking the capsid symmetry

The structure of the DLP interior was resolved in multiple steps (see Chapter 2). First, a map representing 87,217 particles (i.e. 82.72% of the 105,435 used for the final icosahedral refinement) was determined (**Fig. 5.1**) which exhibited pseudo-D5 symmetry (5 pseudo-C2 axes rotated by 72° within one plane perpendicular to a pseudo-C5 axis). The pseudo-C5 axis coincides with one of the icosahedral 5-folds and the five co-planar pseudo-C2 axes coincide with 5 icosahedral 2-folds. The genome reconstruction shows a partially ordered dsRNA and disordered density for all icosahedral 5-folds. The ordered dsRNA is present only in the capsid-proximal genome layer and constitutes ~50% of the whole layer with all other layers being fully disordered. This indicated a high degree of heterogeneity for the majority of the genome and even more so for the RdRP binding sites.

The 12 icosahedral 5-fold sites are indistinguishable within the framework of the icosahedral capsid reconstruction and can only be distinguished within that of the asymmetric capsid interior reconstruction. To ease the discussion, these 12 locations are numbered 1-12 in regards to the D5 frame as following: #1-#2, #3-#4, #5-#6, #7-#8 and #9-#10 are the ten equatorial locations corresponding to the 5 pseudo-C2 axes, while #11 and #12 are the south and north poles of the pseudo-C5 axis, respectively (**Fig 5.1**).

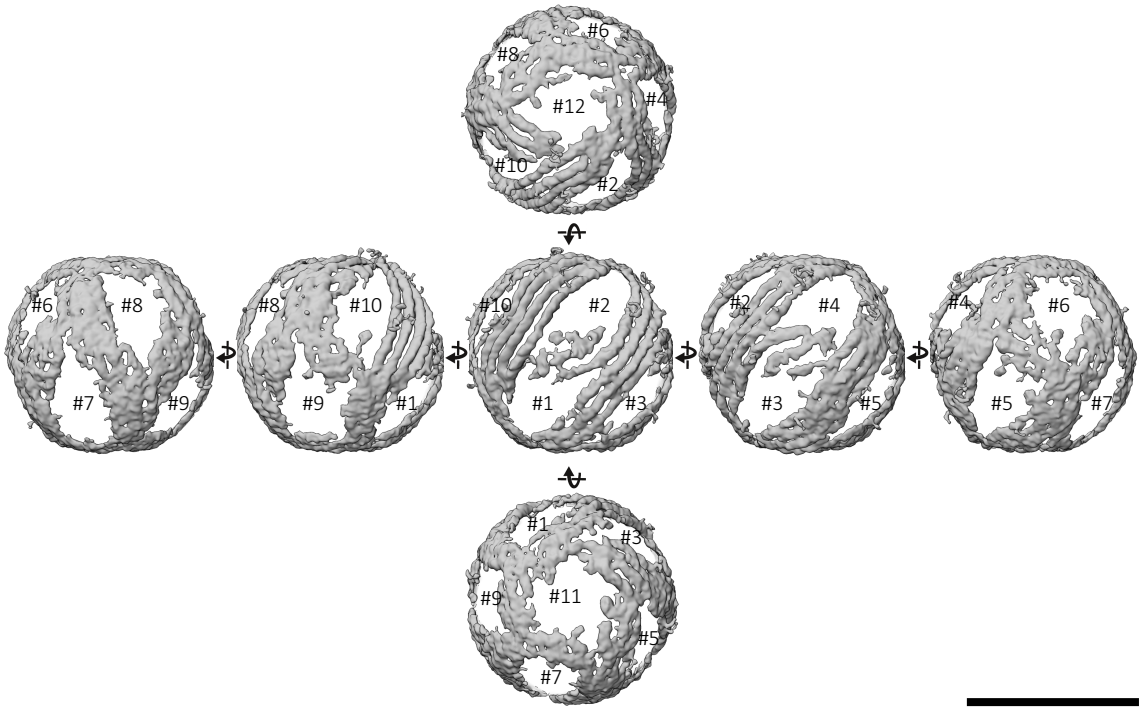


Figure 5.1. Pseudo-D5 organisation of capsid proximal dsRNA in RTV QNC
The two 180° rotated pseudo-C5 views are shown at the bottom and at the top. 90° rotated to them is the plane containing the 5 pseudo-C2 views, which are 72° rotated with respect to each other. The numbering of the icosahedral 5-folds in accord with the text is also indicated. Scale bar, 400 Å.

Unambiguous separation of dsRNA strands is observed only in the region centred on #2 delimited by #1-#3-#4-#12-#10, whereas the opposite i.e. 180° rotated area centred on #7 delimited by #8-#6-#5-#11-#9 presented significantly poorer dsRNA density. Consequently, the dsRNA bridging the corresponding 5-fold pairs across the pseudo-C2 axes is ordered only for the #1-#2 and #3-#4 pairs. Notably, the inclination of these bridging strands is very similar for the #1-#2 and #3-#4 pairs and roughly 90° rotated for the #5-#6, #7-#8 and #9-#10 pairs. Furthermore, the dsRNA strands lining the #1-#2 and #3-#4 regions connect the two poles with little interference from the dsRNA around the corresponding pseudo-C2 axes. This appears to not be the case for

the strands lining the #5-#6, #7-#8 and #9-10 regions, observation best depicted by the dsRNA delimited by #6-#7-#9-#8.

5.4.2. Localized reconstruction of the 5-folds

As described in Chapter 2, two distinct procedures were performed to resolve the symmetry mismatched VP1. First, to obtain the highest resolution of the RdRP itself, the 12 icosahedral 5-fold locations were extracted and combined into a single dataset. This was then subjected to 3D classification relaxing C5 symmetry and 3D refinement leading to a 3.7-Å *in situ* structure of VP1. The resulting map allowed fitting the atomic model of purified VP1 previously obtained by X-ray crystallography. Notably, several differences were detected between the *in situ* and purified VP1 structures. These structural differences have been described by two studies published soon before the writing of this thesis began [59, 68]. The results obtained here are in good agreement with these reports and do not provide any significant novel insights into the VP1 structure or the VP1-VP2 interactions (data not shown).

The second procedure was aimed at reconstructing the VP1 together within its local environment of the capsid interior. Therefore, instead of combining the 12 icosahedral 5-fold locations into a single dataset, they were kept separate while being subjected to individual 3D classifications. 6 classes were employed in each of the 12 volume classifications and the classification parameters were kept identical. In the following section, 'matching a VP1 orientation' refers to an achievable rigid body fit in the density using UCSF Chimera [122] of one of the 60 icosahedrally-related VP1 orientations without significant displacement of the molecule (see Chapter 2). Out of the 72 output classes, 57 presented density that matched one VP1 orientation, 8 matched 2

VP1 orientations and 7 matched no VP1 orientation. Out of the 60 possible VP1 positions, 56 could be accounted for by the 72 classes with 41 by exactly one class, 13 accounted for by exactly 2 classes and 2 by exactly 3 classes (73 fits total). The cross-correlation values of the fits ranged in value from 0.363 to 0.665 with an average of 0.540, median of 0.548 and standard deviation of 0.076. The results varied widely across the 12 classifications and the map reliability did not correlate perfectly with the VP1 fit or the corresponding number of particles.

Visual inspection of the 72 maps aided by VP1 cross-correlation values determined in phenix [124] concluded that only 42 were of reliable quality (**Table 5.1** and **Figure 5.2**). These accounted for 38 unique and 2 degenerate VP1 orientations. These 42 maps will be referred to as 'good' and the remaining 30 as 'bad'. All 21 maps with UCSF Chimera fits equal to or better than 0.600 were preserved, along with all 5 with fits between 0.575 and 0.599, 9 out of 10 with fits between 0.550 and 0.574, 4 out of 6 with fits between 0.525 and 0.549, 1 out of 8 with fits between 0.500 and 0.524 and 2 out of 23 with fits below 0.500 (0.499 and 0.484). As a number of maps exhibited significant noise levels, it cannot be excluded and it is actually rather likely that some of the fits, especially the ones with particularly low values, may correspond to noise and not VP1 signal. In addition, the maps that lead to poor fits or no fits at all cannot be assigned with certainty to the corresponding 5-fold lacking a VP1 copy altogether. This is due to the fact that they do not present reliable density resembling either dsRNA or another protein at the VP1 binding sites.

21 out of the 42 'good' maps (50.0%) originated from the six 5-folds lined by ordered dsRNA (#1, #2, #3, #4, #10 and #12 which will be denoted as 'secured') and 21

from the other 6 5-folds (denoted as 'unsecured'). The highest 4 fits originate from the secured 5-folds, as well as only 1 out of 5 with fits between 0.30 and 0.34 and 1 out of 3 below 0.30. In terms of average CC value, 4 of the top 5 were secured with the remaining 2 ranking 8th and 9th. These results indicate a slight yet robust advantage provided by ordered neighbouring dsRNA on VP1 structure. Based on the arrangement of VP1 proximal dsRNA, the 42 maps can be classified as representing a 'caged in' VP1 with dsRNA spooled around it (i.e. #1.1 and #1.2 in **Fig. 5.3**) or a 'free' VP1 lacking this dsRNA pattern (i.e. #2.1 in **Fig. 5.4**). There appears to be a tendency for the 5-folds without highly ordered dsRNA in the overall structure to preferentially accommodate 'caged in' VP1s as these 5-folds are responsible for 15 out of 25 such VP1s.

Table 5.1. Summary of the 12 localized classifications in RTV QNC
For each 5-fold, several parameters are indicated. The degree of dsRNA order ranges from 5 (best) to 1 (worst). The particle distribution refers to all 6 classes employed, while reliable particles refers to the total % of particles in the reliable classes only.

	Shape	dsRNA order	Reliable fits	'Caged in' VP1s	Fit range	Mean fit	Particle distribution (%)	Reliable particles (%)
#1	triangle	4	6	3	0.29 - 0.45	0.39	12.4 - 28.4	100
#2	triangle	5	3	1	0.32 - 0.57	0.45	11.8 - 33.5	65.0
#3	triangle	4	3	1	0.42 - 0.52	0.46	17.4 - 25.4	62.9
#4	triangle	4	3	2	0.39 - 0.51	0.44	13.9 - 24.6	60.2
#5	elongated triangle	2	5	4	0.31 - 0.49	0.40	10.6 - 26.7	89.8
#6	pentagon	1	4	1	0.27 - 0.40	0.35	15.7 - 24.9	72.4
#7	elongated triangle	2	3	1	0.32 - 0.42	0.37	15.7 - 20.2	53.9
#8	elongated triangle	2	3	3	0.42 - 0.50	0.46	14.5 - 24.4	57.7
#9	elongated pentagon	1	4	4	0.32 - 0.47	0.40	16.5 - 20.2	73.1
#10	elongated triangle	3	2	0	0.43 - 0.54	0.49	15.2 - 26.1	41.6
#11	pentagon	1	2	2	0.27 - 0.32	0.30	13.6 - 14.6	28.1
#12	pentagon	3	4	3	0.35 - 0.40	0.38	14.0 - 20.3	68.9

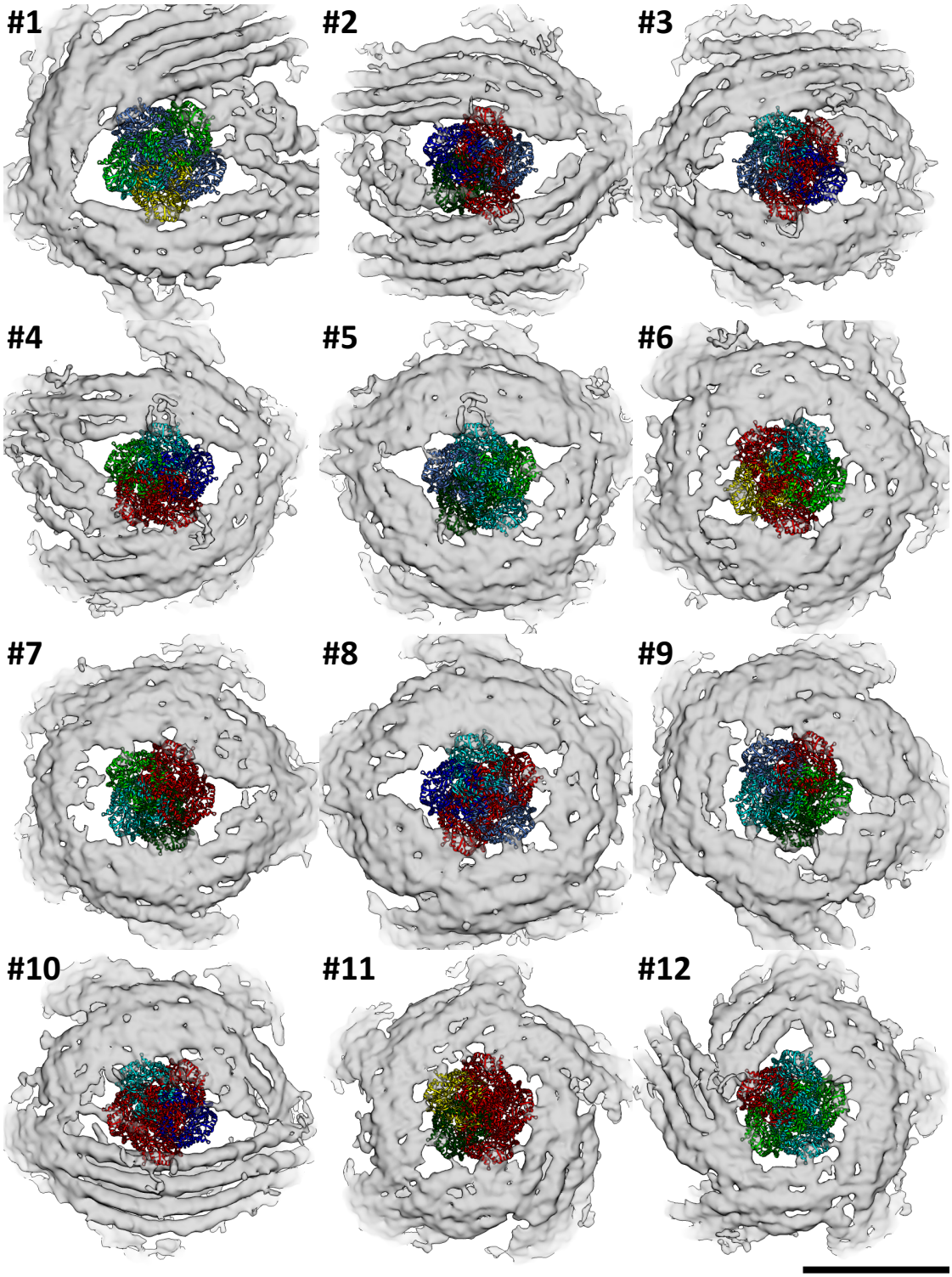


Figure 5.2. Summary of the 12 localized classifications in RTV QNC
The 60 VP1 binding sites are superimposed onto 5-folds #1-#12 of the pseudo-D5 structure of the capsid interior. The VP1s are coloured by CC value as follows: CC>=0.5 (dark blue), 0.50>CC>=0.45 (light blue), 0.45>CC>=0.40 (cyan), 0.40>CC>=0.35 (green), 0.35>CC>=0.30 (dark green), 0.30>CC>=0.25 (yellow), 0.10>CC (red). Scale bar, 200 Å.

Generally, in the maps with ordered dsRNA lining the 5-fold the CC value tends to be lower for the VP1 orientations roughly orthogonal to the dsRNA than for those roughly parallel to the dsRNA. Each map exhibits unique particularities in the dsRNA arrangement in the immediate proximity of the VP1, while the dsRNA lining the 5-fold and the dsRNA bridging across the pseudo-C2 axes appears less variable. In the following sections, blue corresponds to maps with phenix CC values over 0.50, cornflower blue to ones over 0.45, cyan to ones over 0.40, green to those over 0.35, dark green to those over 0.30 and yellow to those bellow 0.30.

#1 is the only 5-fold in which all six classes present reliable density for VP1 with four representing unique VP1 orientations and two corresponding to one degenerate orientation (**Fig. 5.3**). Three of the six maps appear to be 'caged in' by dsRNA spooling around them (#1.1, #1.2 and #1.5), whereas the other three appear less tightly contained (#1.3, #1.4 and #1.6). Of the six maps, the four that are aligned more or less in parallel to the ordered dsRNA strands in the overall map appear best ordered (#1.1-4), while the two that are roughly orthogonal to the strands are least ordered (#1.5-6). While #1.1-5 appear to preserve the overall architecture of the dsRNA strands lining the 5-fold, #1.6 lacks one of these strands in order to accommodate the VP1 (**Fig 5.3 #1.6***). Even maps #1.3 and #1.4 which correspond to the same VP1 orientation display subtle differences in RNA arrangement from the other. Notably, there is strong unknown density at the template entry site in map #1.4 (**Fig 5.3 #1.4***), which map #1.3 lacks. It is however possible that this density may be the result of inaccurate classification.

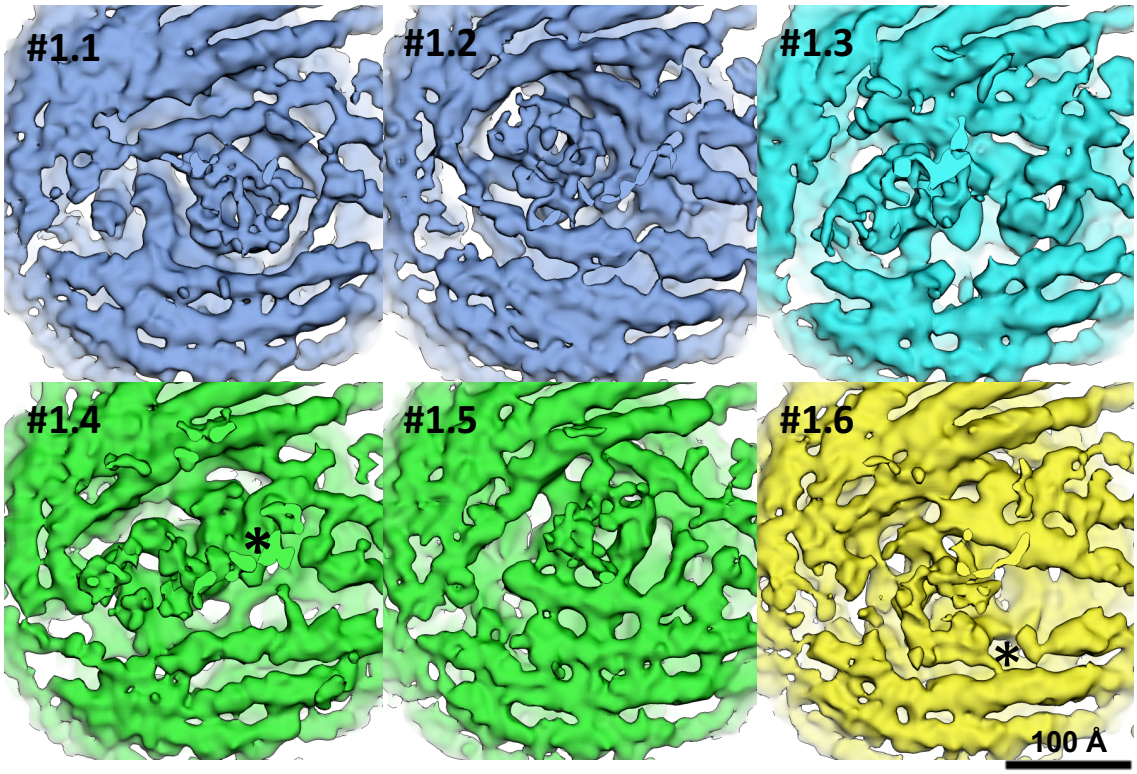


Figure 5.3. VP1-dsRNA arrangements at 5-fold #1 in RTV QNC

#2 presents reliable density for three VP1 orientations, each represented by one map (**Fig. 5.4**). The two VP1 orientations that could not be reliably detected are the only two that appear to clash with the ordered dsRNA strands (**Fig. 5.1**). One of the three reliable maps is 'caged in' by dsRNA (#2.3) and similarly to #1, the two that are aligned more or less in parallel to the ordered dsRNA strands appear best ordered (#2.1-2), while the one that is roughly orthogonal to the strands is least ordered (#2.3).

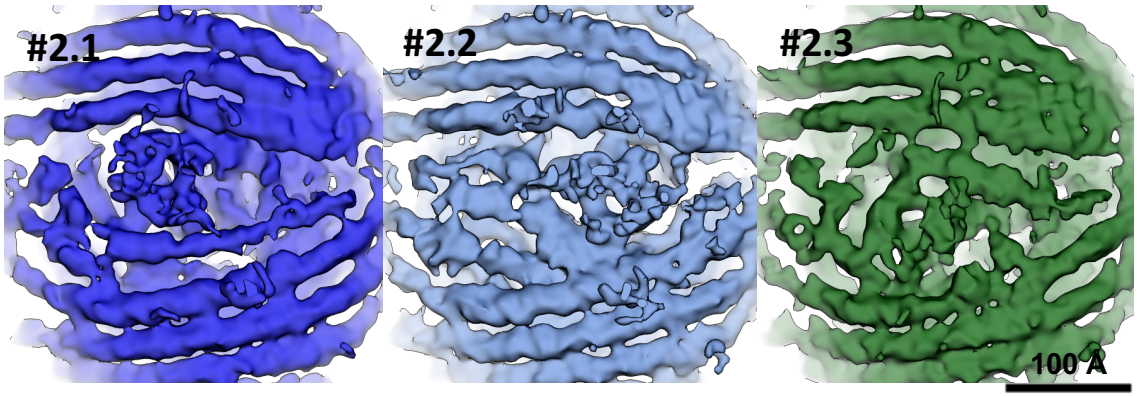


Figure 5.4. VP1-dsRNA arrangements at 5-fold #2 in RTV QNC

#3 is very similar to #2 in that it also accounts for three VP1 orientations, of which one (#3.3) is 'caged in', orthogonal to the dsRNA strands lining the 5-fold and has the poorest CC value (**Fig. 5.5**).

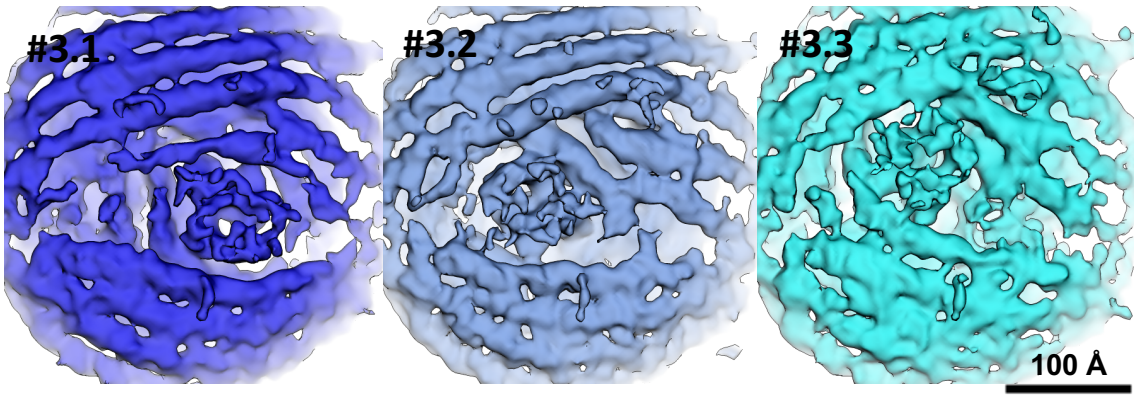


Figure 5.5. VP1-dsRNA arrangements at 5-fold #3 in RTV QNC

#4 also accounts for three VP1 orientations, though unlike #2 and #3, two of these are 'caged in' and the one orthogonal to the dsRNA lining the 5-fold (**2**) does not give the poorest CC value (**Fig. 5.6**). As seen in #1.6, this map lacks one of the side dsRNA strands in order to accommodate the VP1.

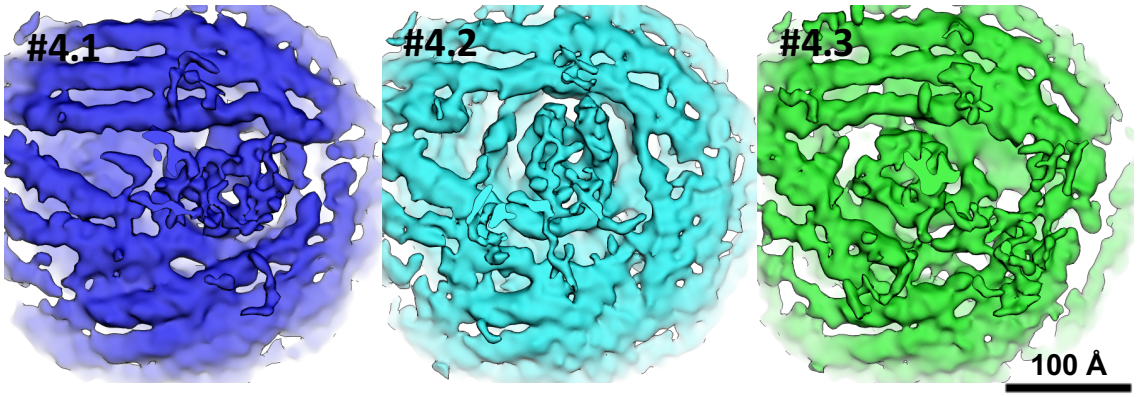


Figure 5.6. VP1-dsRNA arrangements at 5-fold #4 in RTV QNC

#5 accounts for five VP1 orientations and all 5 maps exhibit significantly poorer resolved proximal dsRNA compared to #1-#4. Four of these five maps (#5.2-5) represent 'caged in' VP1s (**Fig. 5.7**).

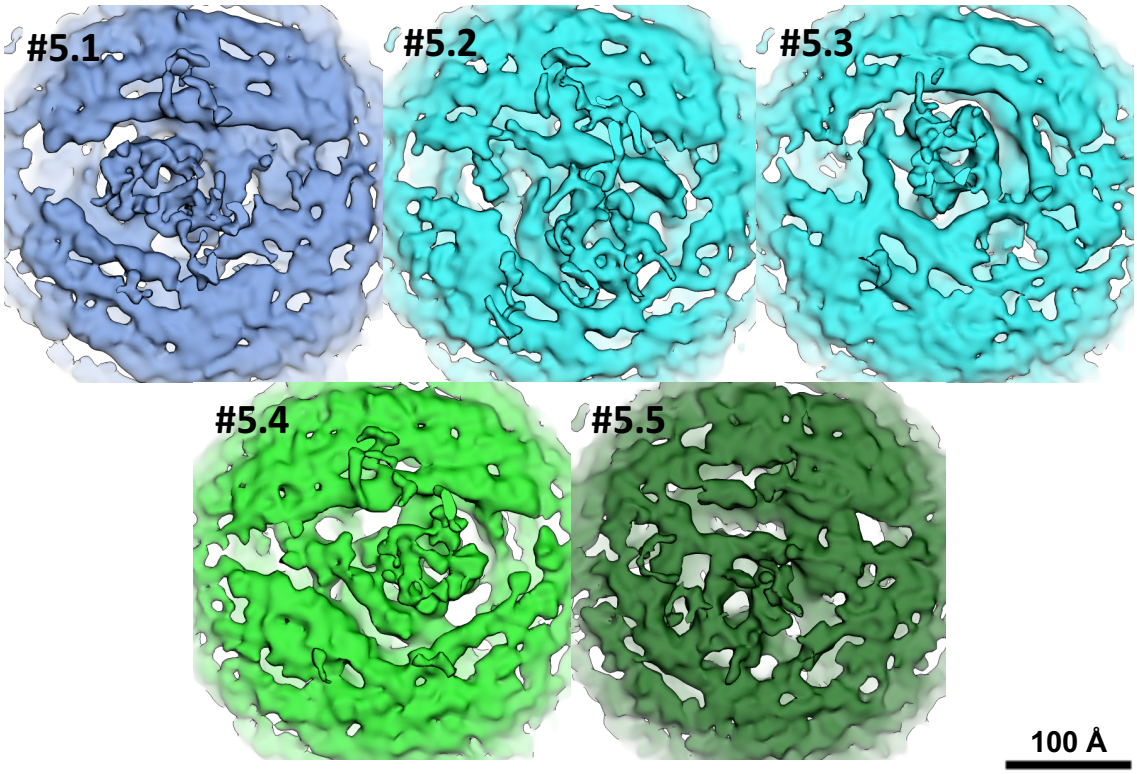


Figure 5.7. VP1-dsRNA arrangements at 5-fold #5 in RTV QNC

The four maps obtained for 5-fold #6 account for three VP1 orientations, one of which is degenerate and overall #6 ranks as one of the poorest 5-folds with respect to the CC value of its maps (**Fig. 5.8**). Likewise, the dsRNA is poorly resolved and appears to be highly inclined with respect to the 5-fold. The two maps corresponding to the degenerate orientation #6.2-3 show significant differences in dsRNA arrangement both in the proximity of VP1 as well as in the more remote regions.

Similarly to #5 and #6, the dsRNA in the reliable maps from #7 is predominantly disordered with short stretches slightly more ordered in the immediate vicinity of VP1 (**Fig. 5.9**). These latter strands appear to wrap around the corresponding VP1s thus mimicking the VP1 orientation. While in #7.1 and #7.2 the ordered dsRNA strands point towards #8, in #7.3 they point towards #6. However, the general orientation of the disordered dsRNA farther from VP1 is roughly similar in the 3 maps.

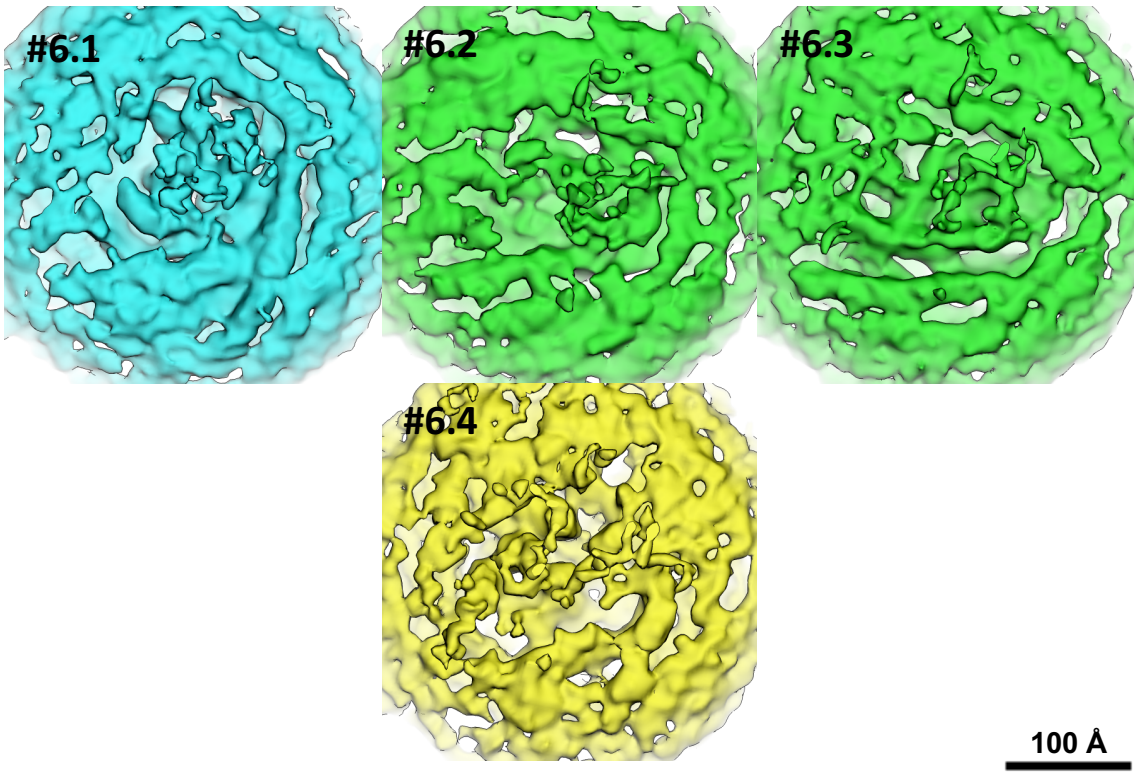


Figure 5.8. VP1-dsRNA arrangements at 5-fold #6 in RTV QNC

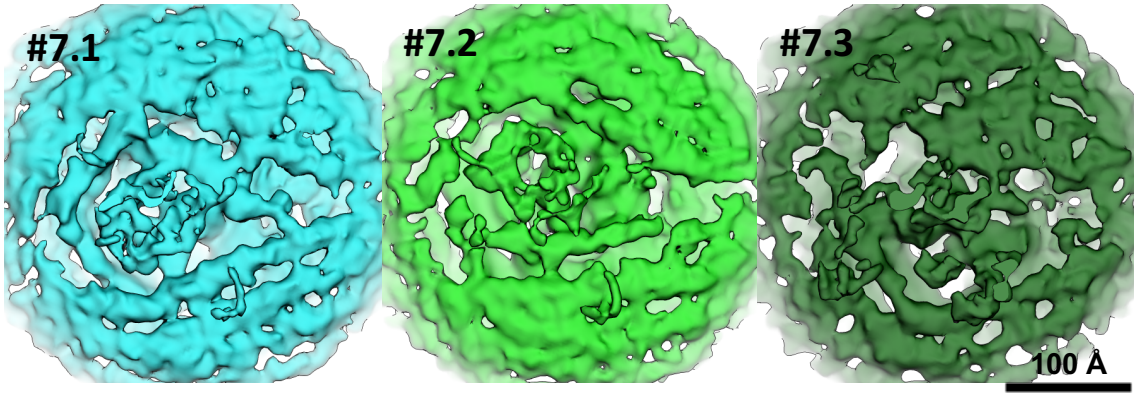


Figure 5.9. VP1-dsRNA arrangements at 5-fold #7 in RTV QNC

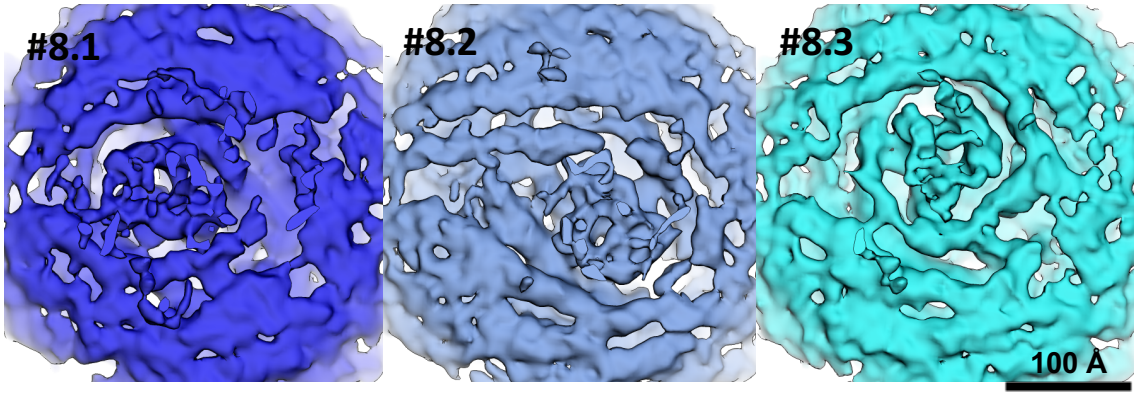


Figure 5.10. VP1-dsRNA arrangements at 5-fold #8 in RTV QNC

Of the 5-folds with poorly resolved dsRNA in their vicinity, #8 gives the best results in terms of the CC values for its maps by a significant margin (**Fig. 5.10**). All three corresponding VP1s are 'caged in' and the dsRNA in their proximity is better ordered than in #5, #6 and #7. The more remote dsRNA is best resolved in #8.2 though overall it still displays limited strand separation compared to the maps originating from the secured 5-folds.

In similar fashion to #8, all four classes corresponding to #9 exhibit 'caged in' VP1s and display similar degree of dsRNA resolvability as the other maps originating from unsecured 5-folds (**Fig. 5.11**).

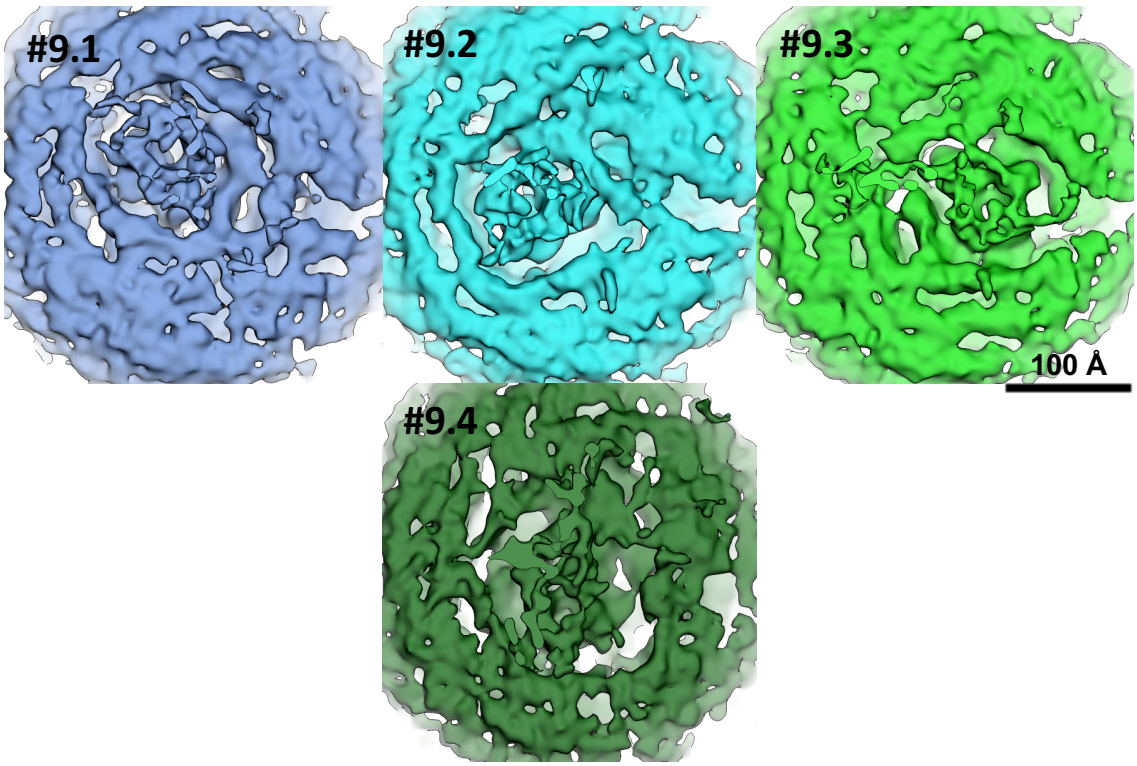


Figure 5.11. VP1-dsRNA arrangements at 5-fold #9 in RTV QNC

Despite providing only two maps with reliable VP1 density, the two maps corresponding to #10 rank 2nd and 12th of all 42 maps in terms of CC value. Neither of the two VP1s is 'caged in' and both maps display well resolved dsRNA towards #1, #2 and #12, and poorly resolved dsRNA towards #8 and #9 (Figure **5.12**).

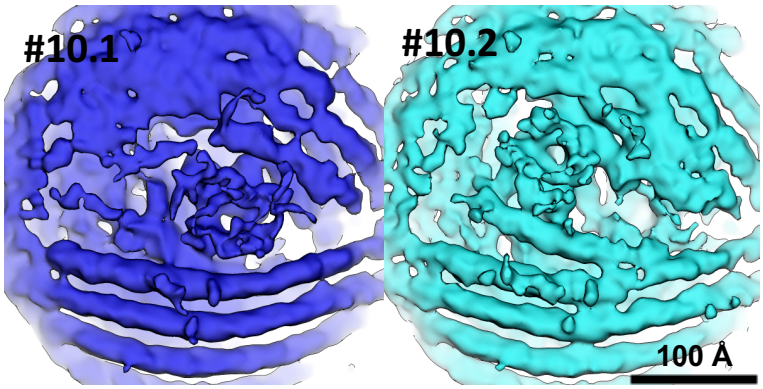


Figure 5.12. VP1-dsRNA arrangements at 5-fold #10 in RTV QNC

#11 is overall the lowest ranking of all 5-folds in terms of CC values and only 28.2% of the particle population fall within the reliable classes. However, #11.1 and #11.2 both exhibit dsRNA on average better resolved than in the maps originating from unsecured 5-folds (**Fig. 5.13**). Notably, the dsRNA in both maps is roughly equally well resolved towards the secured and unsecured 5-folds. Unlike 5-folds #1-#10, both proximal and remote dsRNA to the VP1 in both maps appear to spool around the VP1.

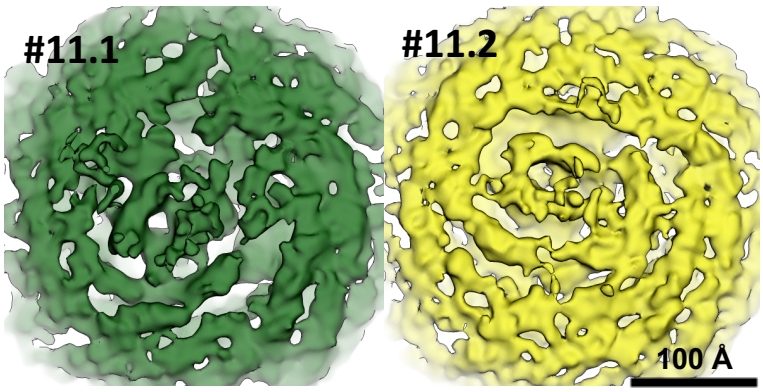


Figure 5.13. VP1-dsRNA arrangements at 5-fold #11 in RTV QNC

Similarly to #11, the dsRNA in the maps from #12 spools around the VP1 though in #12.4 the proximal dsRNA does not lead to a 'caged in' VP1 as one side of the protein remains exposed (**Fig. 5.14**). In all four maps, the dsRNA is best resolved towards the secured 5-folds and significantly worse resolved towards the unsecured ones.

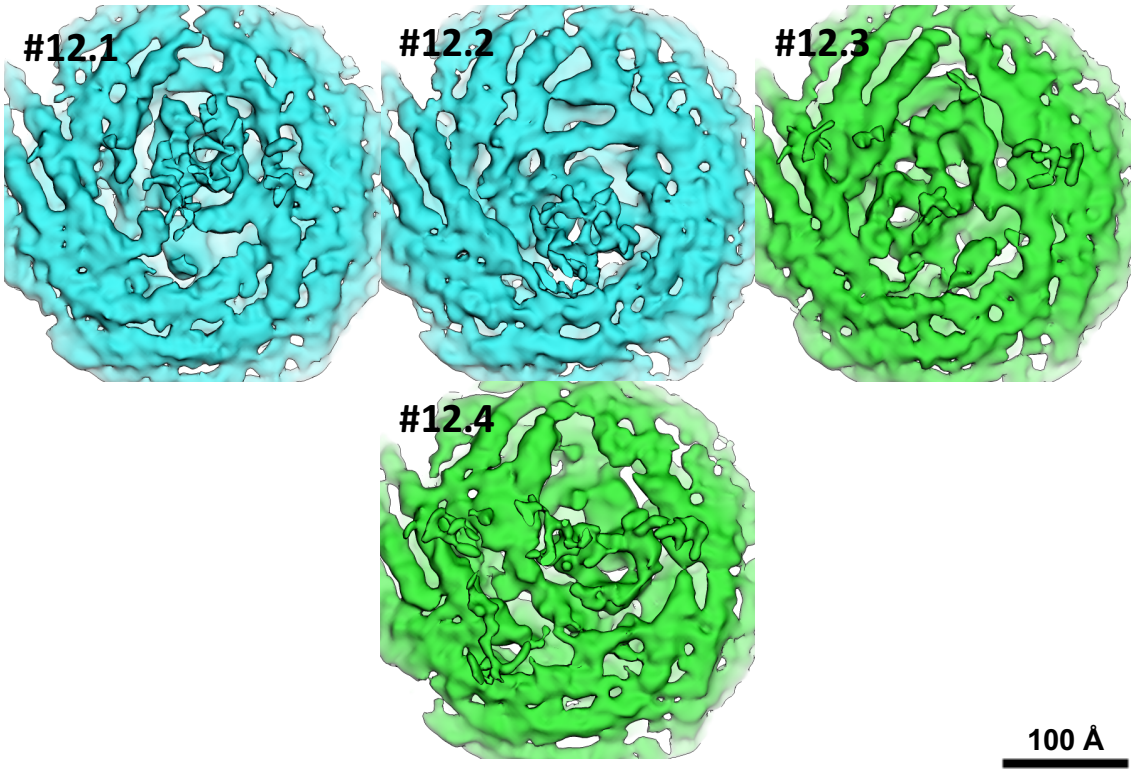


Figure 5.14. VP1-dsRNA arrangements at 5-fold #12 in RTV QNC

Taken together, these data paint a complicated picture of the RTV capsid interior. On one hand, numerous reconstructions show a significant degree of short-range order for the dsRNA strands in the proximity of well-resolved VP1s. On the other hand, there are numerous examples of well-resolved VP1 surrounded by poorly resolved dsRNA and vice-versa. Furthermore, the VP1 reconstructions for #11 are significantly poorer than for the other eleven 5-folds suggesting that #11 is the site showing the highest degree of disorder in the VP1 structure.

5.4.3. VP1-dsRNA interactions

The interaction modes between VP1 and dsRNA were further investigated in an attempt to identify possible similarities and differences among the 42 reliable maps. For this, the maps were oriented such that the VP1s are in the same orientation and several close-ups from different views were taken. VP1 was coloured by domain as following: dark khaki for the N-terminal domain (residues 1-332), blue for the fingers domain (333-488, 524-595), red for the palm domain (489-523, 596-685), green for the thumb domain (686-778) and purple for the bracelet domain (779-1088).

Across the 42 maps, six putative types of VP1-dsRNA interactions are identified and will be classified based on location relative to the view direction from the capsid down the transcript exit channel (**Fig. 5.15**). There are two types of interactions on the sides of VP1; one (type I) is made by the N-terminal domain (154-159 stretch N, D, V, A, E, K) and a second (type II) by a region between the fingers and bracelet domains (429-433 stretch A, N, E, R, Y and 805-811 stretch S, S, T, F, N, Y). The remaining four interaction types occur on the side of VP1 pointed towards the capsid interior and are categorised based on their proximity to the template entry channel. Types III and IV happen distal to the entry channel with type III around the 50-53 stretch (K, N, G, L) of a helix-loop-helix subdomain at the N-terminus and type IV between the N-terminal and thumb domains (79-83 stretch Y, D, K, Y, N and 744-748 stretch I, T, G, S, L). Types V and VI are located in the vicinity of the template entry channel and are formed by separate loops of the fingers domain (stretch 406-415 R, Q, L, K, F, G, R, K, T and stretch 577-583 D, G, N, V, I, K, K, respectively).

#2.1 has the best CC value (0.57), originates from a secured 5-fold and the VP1 is not 'caged in' by the proximal dsRNA (**Fig. 5.15**). In terms of orientation, VP1 is roughly parallel to the dsRNA lining the 5-fold and its template exit channel points towards the pseudo-C2 axis and #1. Both the substrate entry and transcript exit channels appear to not be obstructed by nearby dsRNA. The VP1 forms all 6 types of interactions with four distinct dsRNA strands. The type I and II are made with distinct strands while type III and IV are made with the same dsRNA. The type V contact involves what appears to be a dsRNA terminus. This strand appears to run parallel to the capsid surface and lowering the map threshold reveals a potential type VI contact. This dual-binding of a dsRNA molecule is in accordance with a previously proposed model in which the dsRNA is held in a duplex-open state in preparation for transcription [68].

The second best CC value (0.54) arises from #10.1, which displays a very similar organisation to #2.1 though the VP1 proximal dsRNA strands exhibit a slightly different inclination relative to the protein (**Fig. 5.16**). VP1 forms dsRNA contacts of type II, IV and V. The type V contact mimics well that from #2.1 though it is unclear what the origin of the dsRNA strand in question is. If this dsRNA is indeed a segment terminus, it can either come from the dsRNA responsible for the type IV contact or from a strand similar to the one in #2.1 running slightly more diagonally. The data also cannot exclude the option that the dsRNA making the type V contact is not a terminus, in which case transcription activation by the corresponding VP1 may not be as straightforward.

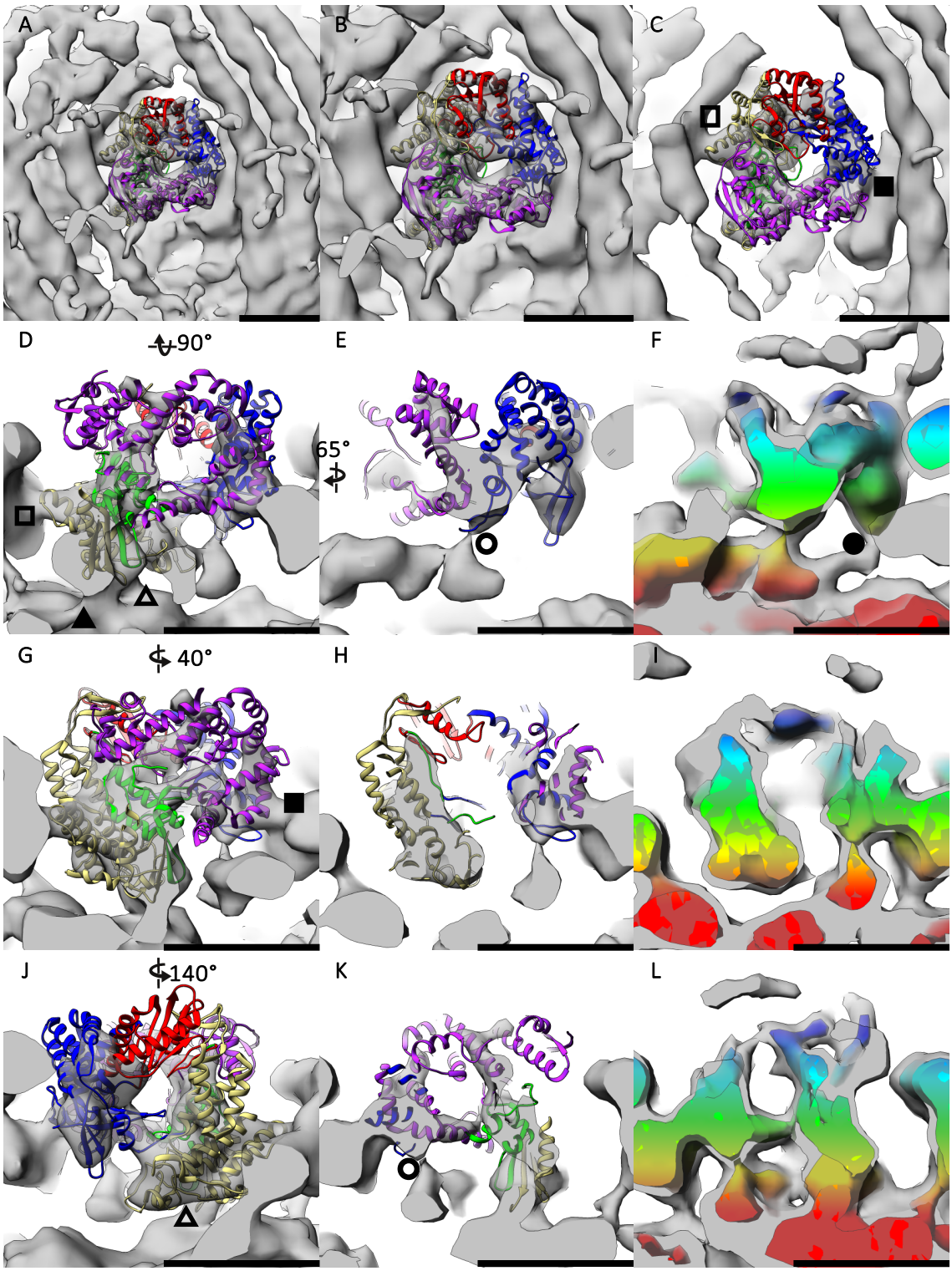


Figure 5.15. VP1-dsRNA arrangement at #2.1 in RTV QNC
VP1 is coloured by domain as described in the text. Panels A-C are top views aligned to the transcript exit channel, D is aligned to the template exit channel (90° rotated from A-C), E-F are 65° rotated from D, while G-I are 40° rotated from D in the opposite direction and J-L are aligned to the substrate entry channel (180° rotated from D). The grey map in A, B, F, I and L is low threshold, while the coloured map in F, I and L, as

well as the grey map in the other panels are high threshold. Type I interactions are marked as empty square, type II as full square, type III as full triangle, type IV as empty triangle, type V as empty circle and type VI as full circle. Scale bars, 50 Å.

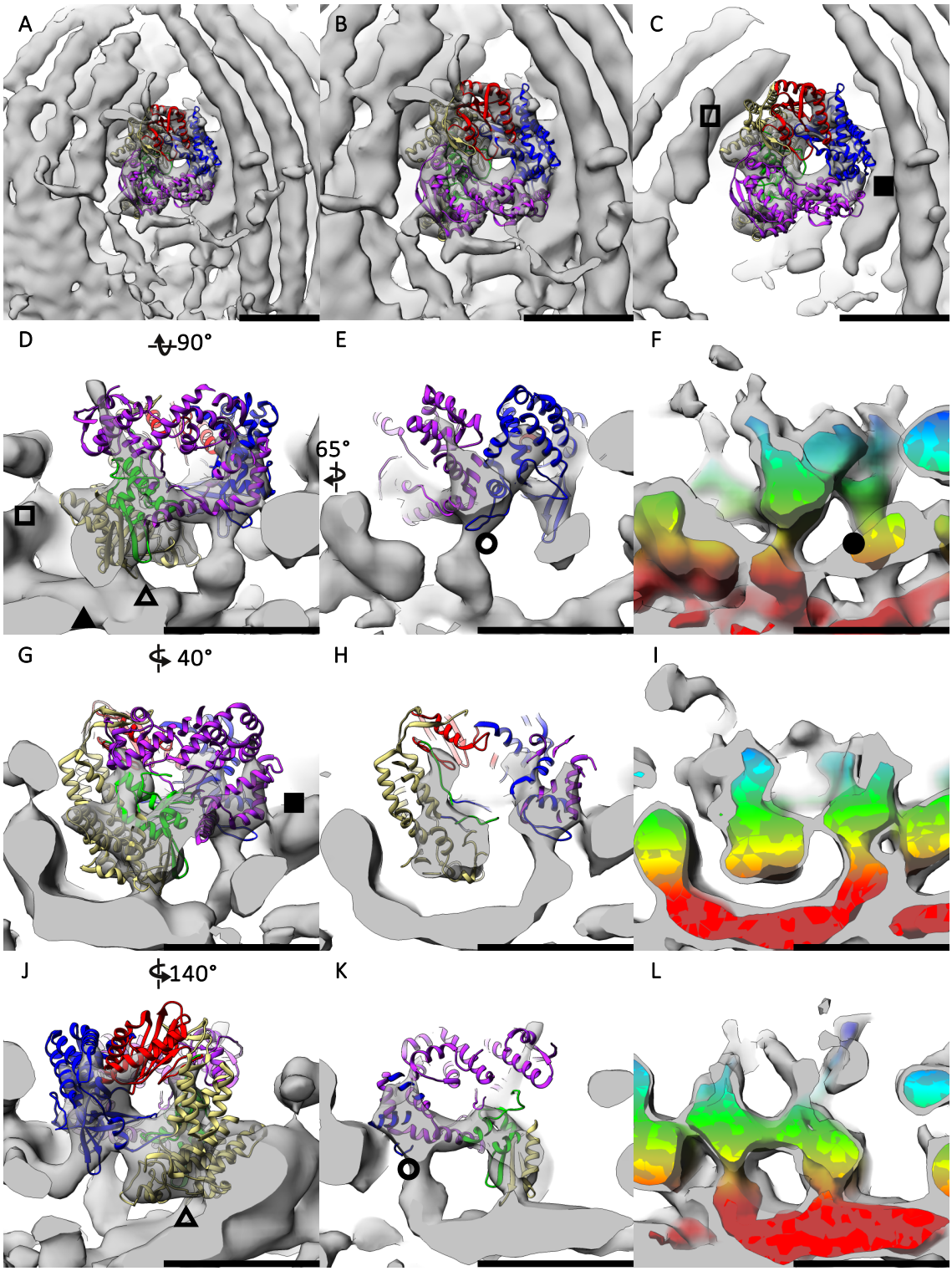


Figure 5.16. VP1-dsRNA arrangement at #10.1 in RTV QNC

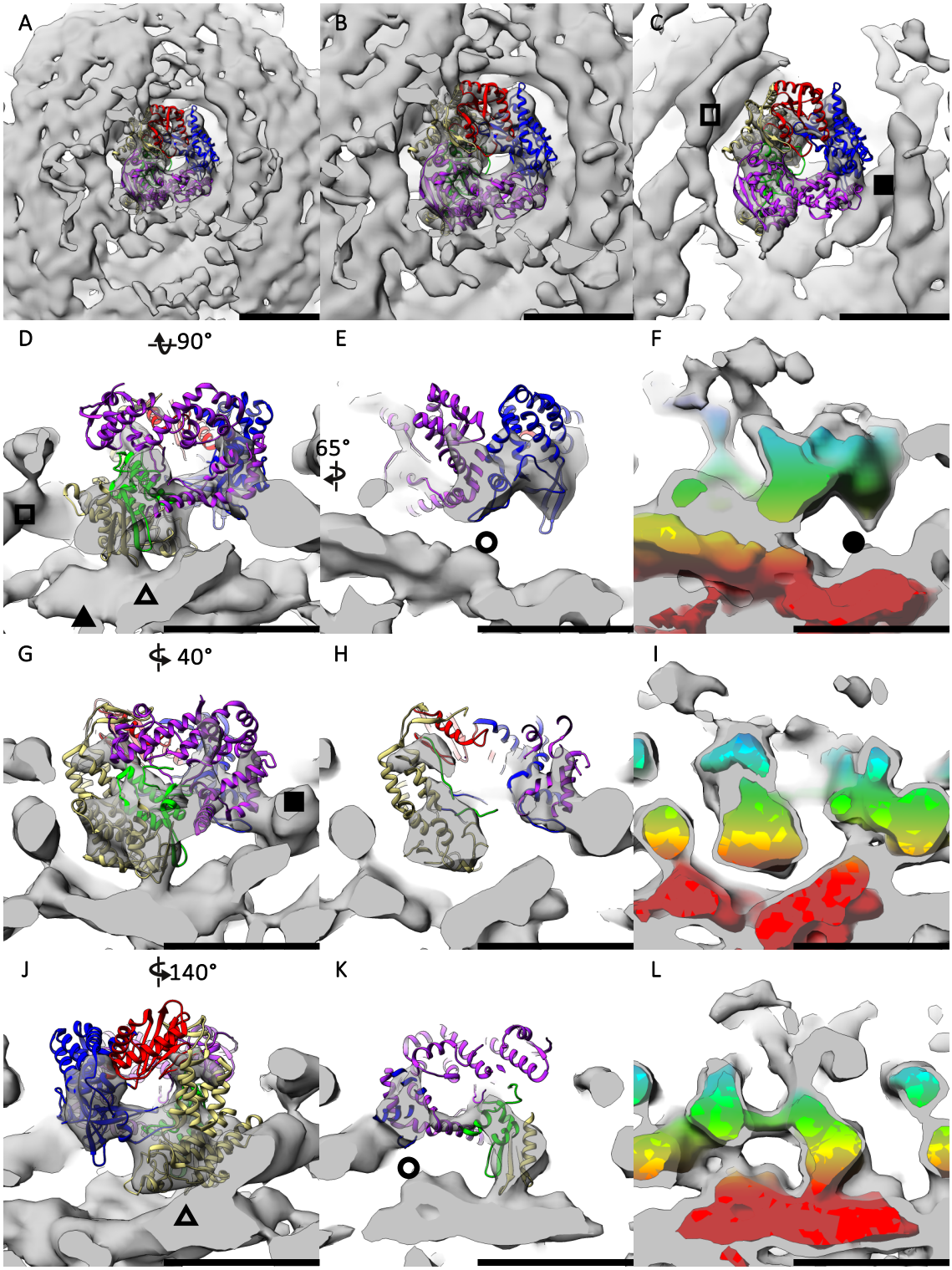


Figure 5.17. VP1-dsRNA arrangement at #8.1 in RTV QNC

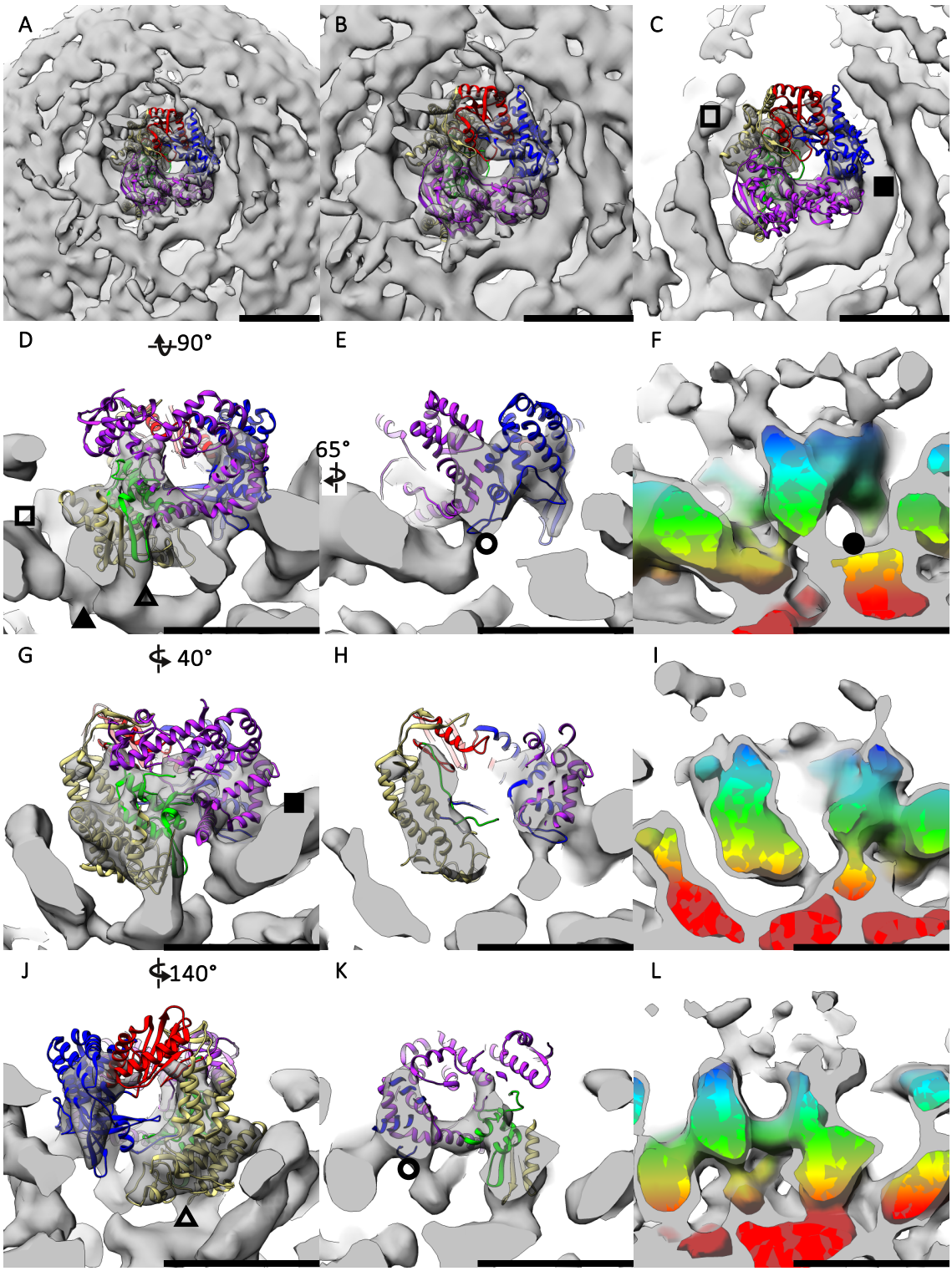


Figure 5.18. VP1-dsRNA arrangement at #9.1 in RTV QNC

#3.1 provided the third best CC value (0.52) and yet another arrangement at the 5-fold highly similar to that in #2.1 and #10.1 with a slight change in inclination for the VP1 proximal dsRNA strands. In terms of VP1-dsRNA contacts and origin of the dsRNA contacting the fingers domain, #3.1 is virtually identical to #10.1 though a type VI contact appears to also occur with a dsRNA strand running to the side of VP1.

#8.1 accounts for the fifth highest CC value (0.50) despite having a strikingly different dsRNA arrangement with respect to VP1 (**Fig. 5.17**). The resolvability of the dsRNA is significantly poorer than in #2.1, #10.1 and #3.1 and VP1 appears to be 'caged in' resulting in an obstructed template exit channel. Though the type I-IV VP1-dsRNA contacts are very similar to those in #2.1, no dsRNA appears to lead to any type V or VI contacts. In fact, there seems to be no dsRNA anywhere in the immediate proximity of the template entry channel; potential contacts slightly farther away from the channel can only be observed at much lower map density thresholds.

#4.2 is another map with a high CC value (0.42) displaying a 'caged in' VP1 though the dsRNA is better resolved and highly dissimilar in arrangement to that in #8.1. Nonetheless, the template entry channel in #4.2 is even more clearly devoid of dsRNA than in #8.1 as even at a very low threshold there is no detectable dsRNA in the vicinity of the channel. Another example of a map virually identical in this aspect is #3.2, in which the VP1 is slightly less 'caged in' by neighbouring dsRNA. Nonetheless, not all 'caged in' maps lack VP1-dsRNA contacts proximal to the template entry channel. One example of such a map is #9.1 which displays a similar architecture to #8.1 with an even more 'caged in' VP1 (**Fig. 5.18**). Unlike #8.1, #9.1 presents strong density corresponding to a dsRNA terminus attached to the fingers domain in

reminiscent fashion of #2.1. The only difference from #2.1 seems to be the orientation of the dsRNA in question which runs significantly more diagonally with respect to the capsid plane. Furthermore, at low map density threshold this strand in #9.1 appears to also make a previously unobserved contact to the N-terminal domain (stretch 188-192 K, Y, R, Y, E). This indicates another possible configuration of the duplex-open state of the dsRNA to be transcribed. However, this dsRNA thus participates in 'caging in' VP1 by obstructing the template exit channel which means it is plausible that the reorganisation occurring during transcription might lead to a partial uncaging of VP1.

5.4.4. Analysis of VP1 orientations at neighbouring 5-folds

5.4.4.1. Correlations between individual classes

The six classes at each 5-fold regardless of map quality or VP1 fit value were subjected to statistical analysis with the six classes at each of the 5 neighbouring 5-folds probing for potential correlations. The highest correlations or Normalised Deviations *ND* (see Chapter 2) were observed for pairings of maps unable to reliably account for any VP1 orientations i.e. 'bad' maps. Often these highly correlating pairs involved at least one map failing to provide a reliable fit even in UCSF Chimera with the highest such correlation of +68.0% occurring between #12.6 and #8.6 (both with no fit). For ease of discussion, pairings with correlations above 40% will be defined as tier-1, ones between 40% and 30% as tier-2, ones between 30% and 20% as tier-3, ones between 20% and 10% as tier-4 and ones below 10% as tier-5. 'Bad' maps appear to correlate best with each other followed by pairs of one 'good' and one 'bad' map and eventually by pairs of two 'good' maps, which appear significantly less correlated (**Table 5.2**).

Table 5.2. Summary of the different types of map pairings in RTV QNC
Maps with VP1 CC >= 0.25 are defined as 'good' and with CC < 0.1 as 'bad'. Tiers 1-5 described in the text. Values expressed as percentage of all 1080 pairings are shown outside parenthesis and values expressed as percentage of number of pairings in the corresponding tier are shown in parenthesis.

N = 1080	%	tier-1 (%)	tier-2 (%)	tier-3 (%)	tier-4 (%)	tier-5 (%)
total	100	1.9	2.0	2.2	15.5	78.3
good-good	32.9	0 (0)	0.19 (9.1)	0.46 (20.8)	2.78 (18.0)	29.4 (37.6)
good-bad	50.9	0.56 (28.6)	1.39 (68.2)	1.30 (58.3)	8.80 (56.9)	38.9 (49.6)
bad-bad	16.2	1.39 (71.4)	0.46 (22.7)	0.46 (20.8)	3.89 (25.1)	10.0 (12.8)

Table 5.3. Most positively and negatively deviating map pairings in RTV QNC
Classes involved in the pairings with highest *ND* values along with their CC values in parenthesis.

Pairing	*ND*	Pairing	*ND*
(0.37) #1.4 - #9.4 (0.32)	+31.2	**(0.42) #7.1 - #9.4 (0.32)**	-14.6
(0.37) #1.4 - #2.2 (0.46)	+30.8	**(0.29) #1.6 - #9.4 (0.32)**	-14.3
(0.32) #7.3 - #9.4 (0.32)	+26.3	**(0.39) #4.3 - #12.3 (0.36)**	-14.0
(0.41) #5.3 - #7.3 (0.32)	+26.0	**(0.36) #1.5 - #9.4 (0.32)**	-13.9
(0.32) #7.3 - #8.1 (0.41)	+24.9	**(0.46) #2.2 - #4.3 (0.39)**	-13.9
(0.39) #4.3 - #6.4 (0.27)	+23.8	**(0.32) #9.4 - #10.2 (0.43)**	-13.5
(0.46) #2.2 - #12.3 (0.36)	+22.6	**(0.39) #4.3 - #5.3 (0.41)**	-13.1

The highest negative correlations are much lower in magnitude than the positive ones though the majority of tier-4 pairings between reliable classes are negative rather than positive (**Table 5.3**). As such, there are 0/0 tier-1, 0/2 tier-2, 0/5 tier-3 and 21/30 tier-4 negatively correlating pairs of reliable maps. Among all 16 tier-4 and higher pairings between reliable maps with positive correlations, #9.4, #8.1 and #4.3 appear three times. Among all 21 tier-4 and higher pairings between reliable maps with negative correlations, #9.4 appears 11 times and #4.3 appears 8 times, while #8.1 does

not appear at all. As the 5-folds neighbouring #9 account for 16 reliable maps and those neighbouring #4 for 19, these results indicate that #4.3 and especially #9.4 may indeed significantly influence the VP1 orientations at their neighbouring 5-folds.

5.4.4.2. Correlations between individual classes and entire 5-folds

As expected given the *ND* between individual classes, #9.4 (CC=0.32) and #4.3 (CC=0.39) exert the strongest influence on neighbouring 5-folds as a whole as well (**Table 5.4** see Chapter 2 for definitions). The average CC for all 42 reliable classes is 0.40, whereas for the 8 most influencing classes is 0.36 and for the 8 least influencing classes is 0.44. Indeed, when comparing the 6 classes at each 5-fold the trend becomes clear that the most influencing classes are those with lowest CC values and vice-versa.

Table 5.4. Most and least influencial classes in RTV QNC
Classes involved in the pairings with highest and lowest *AI/TI* values along with their CC values in parenthesis.

Pairing	*AI*	*TI*	Pairing	*AI*	*TI*
#9.4 (0.32)	12.0	38.0	**#2.1 (0.57)**	2.6	9.7
#4.3 (0.39)	10.5	39.7	**#4.2 (0.46)**	3.1	11.9
#1.4 (0.37)	8.3	25.4	**#3.1 (0.52)**	3.1	12.7
#7.3 (0.32)	7.9	30.4	**#9.3 (0.44)**	3.2	10.5
#6.4 (0.27)	7.9	28.0	**#10.2 (0.43)**	3.3	14.3
#12.3 (0.36)	7.5	23.1	**#6.2 (0.37)**	3.4	11.6
#3.3 (0.42)	6.8	26.4	**#5.5 (0.31)**	3.5	9.9
#2.2 (0.46)	6.0	24.0	**#1.2 (0.45)**	3.5	10.0

When comparing the influence of a given class on each of the 5 neighbouring 5-folds, no clear trend is observed. However, several examples do exist of classes that appear to influence their neighbours with some degree specificity. #1.4 deviates

significantly more in relation to #2 (14.4%) and #9 (11.9%) compared to #3 (7.1%), #11 (4.7%) and #10 (4.5%), whereas #1.6 does so more in relation to #11 (7.8%) and #9 (6.8%) than to #3 (3.3%), #10 (2.4%) and #2 (0.8%). All four classes at #6 (#6.1-#6.4) deviate more in relation to #4 than to all other neighbours (5, 7, 8 and 12), with #6.1, #6.2 and #6.4 doing so by a significant margin.

5.4.4.3. Correlations between entire 5-folds

In terms of the influence of each 5-fold on its neighbouring 5-folds, the twelve 5-folds can be classified into 4 groups with respect to their defining influence *DI* as shown in **Table 5.5** (see Chapter 2 for definition).

Table 5.5. Ranking of the 12 5-folds by Defining Influence in RTV QNC

	DI	dsRNA order	Reliable fits	Fit range	Mean fit	Particle distribution (%)	Reliable particles (%)
#4	5.96	4	3	0.39 - 0.51	0.44	13.9 - 24.6	60.2
#12	5.93	3	4	0.35 - 0.40	0.38	14.0 - 20.3	68.9
#9	5.66	1	4	0.32 - 0.47	0.40	16.5 - 20.2	73.1
#1	5.31	4	6	0.29 - 0.45	0.39	12.4 - 28.4	100
#7	5.25	2	3	0.32 - 0.42	0.37	15.7 - 20.2	53.9
#6	4.58	1	4	0.27 - 0.40	0.35	15.7 - 24.9	72.4
#3	4.56	4	3	0.42 - 0.52	0.46	17.4 - 25.4	62.9
#8	4.42	2	3	0.42 - 0.50	0.46	14.5 - 24.4	57.7
#5	4.34	2	5	0.31 - 0.49	0.40	10.6 - 26.7	89.8
#11	4.22	1	2	0.27 - 0.32	0.30	13.6 - 14.6	28.1
#2	3.99	5	3	0.32 - 0.57	0.45	11.8 - 33.5	65.0
#10	3.55	3	2	0.43 - 0.54	0.49	15.2 - 26.1	41.6

5.5. Discussion

Unlike previously reported [59], the results presented in this chapter suggest that long range order inside the RTV capsid does exist albeit short range order is considerably more abundant. As is the case for ϕ6, relaxing the icosahedral symmetry of the capsid provides a global reconstruction of the RTV RdRP-dsRNA complex. In this map, the dsRNA is poorly resolved and the icosahedral 5-fold locations are even more disordered. The 12 5-folds provide a total of 60 possible VP1 binding sites, 5 at each of the 12 5-folds. As the 5 sites at each 5-fold clash with one another, a maximum of 1 VP1 copy can exist at each 5-fold and thus up to 12 VP1 copies could theoretically coexist within a capsid. The structures resolved here indicate the presence of a highly dynamic interplay between the VP1 copies and neighbouring dsRNA and suggests that there is an extremely high number of possible VP1-dsRNA arrangements within the RTV DLPs. As single-particle cryo-EM relies on averaging homogenous particle populations, the capability of obtaining a global structure of the capsid interior free of incorrect averaging artefacts is significantly hampered even given a dataset of over 100,000 DLPs. Individually analysing capsid subparts drastically reduces the sample heterogeneity in each analysis to manageable levels for the available image processing tools. As a result, the possible orientations of VP1 and respective dsRNA organisations at each 5-fold can be elucidated. However, details about the general capsid interior architecture can only be indirectly inferred by interpreting together the results of the individual analyses.

The global organisation of the capsid proximal dsRNA presented here albeit poorly resolved is sufficient to provide a number of biological insights. Genome structures have been determined for CPV [100, 101], ARV [102, 103] and ϕ6 [130]; these all adopt pseudo-D3 symmetry unlike the pseudo-D5 symmetry presented here for RTV. Though RTV, CPV and ARV are all members of the *Reoviridae* family, the RTV structure presented here shows striking differences even beyond the adopted symmetry. Specifically, the RdRPs in RTV are not rigidly anchored to the dsRNA framework but rather form a highly dynamic framework with the dsRNA. Moreover, the dsRNA strands appear to provide only a limited degree of long range order across the capsid interior. As shown by the results of the 12 individual classifications, though some output classes display a certain degree of similarity, the 5-folds are indeed unique. This emphatically refutes the hypothesis that the arrangement of the VP1 copies is stochastic [59]. Indeed, the 42 resolved and possibly other elusive 5-fold arrangements compared to the 12 or so in CPV and ARV highlight a drastically enhanced degree of flexibility in the capsid interior architecture of RTV. In addition, some level of correlation occurs among VP1-dsRNA arrangements at neighbouring 5-folds. Though limited, these correlations show that the 5-folds themselves are not fully unaffected nor equally affected by their neighbours, and that certain VP1 orientations may favour specific VP1 orientations at neighbouring sites.

Members of *Reoviridae* have 9-12 genome segments and it is thought that the capsid contains exactly one RdRP copy for each segment [11, 100, 101, 139, 162]. The attractive claim that this 1:1 ratio ensures the stoichiometric processing of the genome segments is supported by the global structures of the capsid interior in the 10-segmented

CPV and 11-segmented ARV. The length of the segments in any member of *Reoviridae* varies substantially and naturally each segment encodes different proteins. In terms of packaging, these aspects are inconsequential as exactly one copy of each segment is needed in each virion. However, in terms of replication and transcription, the variable length and the fact that not all viral proteins are needed in the same amount brings into question whether indeed all these viruses contain exactly one RdRP copy per segment. In addition, the vastly elusive mechanism of packaging in *Reoviridae* [9] does not unequivocally explain how viruses with fewer than 12 segments manage to prevent all 12 RdRP binding sites from being occupied. Therefore, there is a possibility that this 1:1 ratio may not hold at least in some of the *Reoviridae* members. One of the two recently published papers on the *in situ* structure of rotavirus VP1 [59] claims the existence of 11 copies stochastically arranged across the 12 5-folds. However, the analysis performed there investigated the 5-folds as part of the icosahedral framework in which they are indistinguishable rather than as unique parts of an asymmetric object. The other such paper in which an almost identical procedure was followed [68] as well as another similar study of the BTV RdRP structure [110] refrained from commenting on the issue.

The analysis presented in this chapter shows that 5-fold #11 is one of two 5-folds for which only two VP1 orientations can be accounted for. In addition, the corresponding maps fit the VP1 model significantly worse than the maps at the other 11 5-folds and combined the two maps represent only 28.2% of the particle population. Together with the fact that one map at #11 fails to account for any VP1 orientation, these results suggests that #11 is the most likely 5-fold to be devoid of VP1. However,

this remains a mere possibility rather than a fact due to the high degree of heterogeneity in the data as well as potential limitations of the image processing and fitting software.

The global structure of the RTV capsid interior here is the only one reported for a member of the *Sedoreovirinae* sub-family of the *Reoviridae* virus family. This sub-family consists of viruses with so-called "non-turreted" capsids and differs from the *Spinareovirinae* sub-family of "turreted" viruses in a number of aspects [58]. Most importantly for this discussion, the capping enzyme VP3 localizes in the capsid projections of the "turreted" viruses, away from the RdRP, whereas in "non-turreted" viruses it is contained within within the capsid-RdRP-dsRNA mix. The different localization of the capping enzyme leads to distinct capping mechanisms in the two sub-families [70]. In "turreted" members, the capping enzyme processes the newly transcribed ssRNA strand as it passes from the RdRP to the turret and towards the capsid exterior. As shown here and by others [59, 68, 79], "non-turreted" members have their RdRP bound to the capsid with the transcript exit channel pointed to the capsid. This likely means that capping the nascent ssRNA strand is slightly more problematic as VP3 cannot directly access the RdRP active site or transcript exit channel. A previously well-established assumption was that VP3 is found in the immediate proximity of both VP1 and the capsid in order to facilitate accessing the transcript [11]. However, the structures shown here as well as the two RTV and one BTV RdRP structures reported recently provide a good argument against this hypothesis by failing to identify the location of the capping enzyme. Even more strikingly, the reported structures of actively transcribing particles [59, 68] identify the nascent transcript and capsid pore required for transcript extrusion yet fail to locate the capping enzyme all the same. This could

suggest that a more complicated capping mechanism is in place involving complex dynamics between the transcript, VP3, VP1 and VP1 proximal dsRNA. If so, the flexibility required to facilitate the dynamics may be the reason behind the drastically increased heterogeneity of the VP1-dsRNA complex described herein compared to those described for *Spinareovirinae*. One important caveat is that no insights can be gained from this study into the positioning of VP3 even at 5-folds with well-resolved VP1-dsRNA arrangements.

Another attractive hypothesis described in one of the recent RTV VP1 publications [68] is that in the quiescent state, the termini of the 11 dsRNA molecules are in a so-called 'duplex-open state'. This entails that the very last section of each genome segment has unpaired bases such that the (+)ve strand is attached via the cap to the VP1 N-terminal domain and the (-)ve strand template attached to the VP1 fingers domain in the vicinity of the entry channel. A cap binding site has indeed been identified in the X-ray structures of RTV and MRV RdRP [73, 161]. This would represent a slick way for each genome segment to be recognised and prepared for transcription by a VP1 copy. However, the 42 structures of 5-folds determined in this chapter tell a slightly more intricate story. Some VP1s appear to indeed bind via the fingers domain to dsRNA termini; nevertheless, others bind to dsRNA that does not appear to be a segment terminus and others do not present this interaction mode at all. In addition, the interactions made by the N-terminal domain are not always with the dsRNA piece that contacts the fingers domain. These variations among VP1s are complemented by the observations of different types of VP1 proximal dsRNA arrangement i.e. 'caged in' vs. 'uncaged' and different types of 5-folds i.e. 'secured' vs

'unsecured'. Taken together, these results suggest that not all VP1s are in the same state with regards to the dsRNA to be used as transcription template. Whether this is characteristic for the DLPs in the quiescent state or caused by the virus purification procedure remains to be determined.

6. Conclusions and future directions

This thesis furthers the understanding of virus capsid interiors and highlights major differences between families of dsRNA viruses as well as between sub-families of the same dsRNA virus family. These differences are thought to stem from particularities in individual virus replication cycles. The capping mechanism in *Reoviridae* appears to play an important role in dictating the interplay between the RdRP copies and dsRNA genome. Nonetheless, more striking are the differences between reoviruses and ϕ6 and the similarities between ϕ6 and dsDNA viruses. dsRNA and dsDNA viruses in general have dramatically dissimilar replication cycles. However, ϕ6 exhibits a number of characteristics that resemble dsDNA more strongly than prototypical dsRNA viruses. ϕ6 belongs to the only dsRNA virus family that infects bacteria with most other bacteriophages having dsDNA genomes. ϕ6 has fewer and larger genome segments than most dsRNA viruses thus being more similar to dsDNA viruses which are predominantly non-segmented. In contrast to the genome segments, ϕ6 RdRPs are more numerous and much smaller than typical reovirus RdRPs. Cystoviruses are also the only dsRNA viruses that contain a genome packaging motor, a common characteristic of dsDNA viruses. Genome packaging in ϕ6 follows a variation of the canonical 'headful' packaging in dsDNA viruses, whereas reoviruses package their segments as a result of protein-genome co-assembly. Notably, the genome and RdRPs in ϕ6 are severely less well-connected than in reoviruses as their packaging processes occur independently, their arrangements in mature nucleocapsids lack virtually any coherence and segment transcription is not RdRP-specific. ϕ6 transcription also occurs in a more rudimentary,

semi-conservative fashion which is also in contrast to the conservative mechanism in most other dsRNA viruses. Taken together, these aspects hint towards the possibility that cystoviruses may share a different line of evolutionary ancestry with dsDNA viruses than all other dsRNA viruses. Nevertheless, making such claims about evolutionary relatedness requires much more thorough investigation and thus it remains to be determined whether this is indeed the case.

In terms of methodologies to visualise virus capsid interiors, cryo-EM is now firmly established as the only technique capable of achieving this feat. The relative ease of acquiring vast datasets combined with recent developments in image processing have empowered cryo-EM to tackle the extraordinary degree of heterogeneity and mixture of symmetries characteristic to capsid interiors. Particularly, localized reconstruction and symmetry relaxation have been successfully employed in determining a multitude of *in situ* structures and are expected to continue doing so in the future. Significantly larger datasets, possibly consisting of millions of particles, will most likely be required to resolve at resolutions better than 4 Å components such as nucleic acid or structurally heterogeneous proteins such as the ϕ6 RdRP and the RTV capping enzyme. Nonetheless, there are yet no clues suggesting that some or any viruses with icosahedral capsids may not be amenable to studies similar to those in this thesis. As structural insights into capsid interior architecture are lacking for the vast majority of viruses, there is no doubt that a myriad of spectacular discoveries are imminent. Hopefully, the procedures described in this thesis will at least to some extent contribute to a few such discoveries.

References

1. Poranen MM, Bamford DH. Assembly of large icosahedral double-stranded RNA viruses. Adv Exp Med Biol. 2012;726:379-402.

2. Prasad BVV, Prevelige PE. Viral Genome Organization. Virus Structure. 2003:219-58.

3. Poranen MM, Tuma R. Self-assembly of double-stranded RNA bacteriophages. Virus Res. 2004;101(1):93-100.

4. Jayaram H, Estes MK, Prasad BVV. Emerging themes in rotavirus cell entry, genome organization, transcription and replication. Virus Res. 2004;101(1):67-81.

5. Daugelavicius R, Cvirkaite V, Gaidelyte A, Bakiene E, Gabrenaite-Verkhovskaya R, Bamford DH. Penetration of Enveloped Double-Stranded RNA Bacteriophages 13 and 6 into Pseudomonas syringae Cells. Journal of Virology. 2005;79(8):5017-26.

6. Danthi P, Guglielmi KM, Kirchner E, Mainou B, Stehle T, Dermody TS. From touchdown to transcription: the reovirus cell entry pathway. Curr Top Microbiol Immunol. 2010;343:91-119.

7. Patton JT, Spencer E. Genome replication and packaging of segmented double-stranded RNA viruses. Virology. 2000;277(2):217-25.

8. Mindich L. Packaging in dsRNA viruses. Adv Exp Med Biol. 2012;726:601-8.

9. Borodavka A, Desselberger U, Patton JT. Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. Curr Opin Virol. 2018;33:106-12.

10. Kainov DE, Butcher SJ, Bamford DH, Tuma R. Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. J Mol Biol. 2003;328(4):791-804.

11. Trask SD, McDonald SM, Patton JT. Structural insights into the coupling of virion assembly and rotavirus replication. Nat Rev Microbiol. 2012;10(3):165-77.

12. Roy P. Bluetongue virus structure and assembly. Curr Opin Virol. 2017;24:115-23.

13. Mindich L. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage phi6. Microbiol Mol Biol Rev. 1999;63(1):149-60.

14. Qiao X, Sun Y, Qiao J, Di Sanzo F, Mindich L. Characterization of Phi2954, a newly isolated bacteriophage containing three dsRNA genomic segments. BMC Microbiol. 2010;10:55.

15. Mäntynen S, Sundberg L-R, Poranen MM. Recognition of six additional cystoviruses: Pseudomonas virus phi6 is no longer the sole species of the family Cystoviridae. Arch Virol. 2018;163(4):1117-24.

16. Vidaver AK, Koski RK, Van Etten JL. Bacteriophage phi6: a Lipid-Containing Virus of Pseudomonas phaseolicola. J Virol. 1973;11(5):799-805.

17. Etten JV, Lane L, Gonzalez C, Partridge J, Vidaver A. Comparative properties of bacteriophage phi6 and phi6 nucleocapsid. J Virol. 1976;18(2):652-8.

18. Huiskonen JT, de Haas F, Bubeck D, Bamford DH, Fuller SD, Butcher SJ. Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. Structure. 2006;14(6):1039-48.

19. Sun Z, El Omari K, Sun X, Ilca SL, Kotecha A, Stuart DI, et al. Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. Nat Commun. 2017;8:14814.

20. Stitt BL, Mindich L. The structure of bacteriophage φ6: Protease digestion of φ6 virions. Virology. 1983;127(2):459-62.

21. Romantschuk M, Bamford DH. Function of pili in bacteriophage Ø6 penetration. J Gen Virol. 1985;66(11):2461-9.

22. Kenney JM, Hantula J, Fuller SD, Mindich L, Ojala PM, Bamford DH. Bacteriophage phi 6 envelope elucidated by chemical cross-linking, immunodetection, and cryoelectron microscopy. Virology. 1992;190(2):635-44.

23. Romantschuk M, Olkkonen VM, Bamford DH. The nucleocapsid of bacteriophage phi 6 penetrates the host cytoplasmic membrane. EMBO J. 1988;7(6):1821-9.

24. Poranen MM, Daugelavičius R, Ojala PM, Hess MW, Bamford DH. A Novel Virus--Host Cell Membrane Interaction: Membrane Voltage--Dependent Endocytic-like Entry of Bacteriophage φ6 Nucleocapsid. J Cell Biol. 1999;147(3):671-82.

25. Butcher SJ, Dokland T, Ojala PM, Bamford DH, Fuller SD. Intermediates in the assembly pathway of the double-stranded RNA virus φ6. EMBO J. 1997;16(14):4477-87.

26. Tuma R, Bamford JK, Bamford DH, Thomas GJ, Jr. Assembly dynamics of the nucleocapsid shell subunit (P8) of bacteriophage phi6. Biochemistry. 1999;38(45):15025-33.

27. Emori Y, Iba H, Okada Y. Semi-Conservative Transcription of Double-Stranded RNA Catalyzed by Bacteriophage ø6 RNA Polymerase. J Biochem. 1980;88(6):1569-75.

28. Usala SJ, Brownstein BH, Haselkorn R. Displacement of parental RNA strands during in vitro transcription by bacteriophage phi 6 nucleocapsids. Cell. 1980;19(4):855-62.

29. Van Etten JL, Burbank DE, Cuppels DA, Lane LC, Vidaver AK. Semiconservative synthesis of single-stranded RNA by bacteriophage phi 6 RNA polymerase. J Virol. 1980;33(2):769-73.

30. Van Dijk AA, Frilander M, Bamford DH. Differentiation between Minus- and Plus-Strand Synthesis: Polymerase Activity of dsRNA Bacteriophage φ6 in an in Vitro Packaging and Replication System. Virology. 1995;211(1):320-3.

31. Makeyev EV, Bamford DH. The polymerase subunit of a dsRNA virus plays a central role in the regulation of viral RNA metabolism. EMBO J. 2000;19(22):6275-84.

32. Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI. A mechanism for initiating RNA-dependent RNA polymerization. Nature. 2001;410(6825):235-40.

33. Paatero AO, Mindich L, Bamford DH. Mutational analysis of the role of nucleoside triphosphatase P4 in the assembly of the RNA polymerase complex of bacteriophage phi6. J Virol. 1998;72(12):10058-65.

34. Pirttimaa MJ, Paatero AO, Frilander MJ, Bamford DH. Nonspecific Nucleoside Triphosphatase P4 of Double-Stranded RNA Bacteriophage φ6 Is Required for Single-Stranded RNA Packaging and Transcription. J Virol. 2002;76(20):10122-7.

35. Poranen MM, Bamford DH. Cystovirus. In: Tidona C, Darai G, editors. The Springer Index of Viruses. New York, NY: Springer New York; 2011. p. 431-7.

36. Sun X, Bamford DH, Poranen MM. Probing, by self-assembly, the number of potential binding sites for minor protein subunits in the procapsid of double-stranded RNA bacteriophage Φ6. J Virol. 2012;86(22):12208-16.

37. Qiao X, Qiao J, Mindich L. Stoichiometric packaging of the three genomic segments of double-stranded RNA bacteriophage Φ6. Proc Natl Acad Sci U S A. 1997;94(8):4074-9.

38. Qiao X, Qiao J, Mindich L. Analysis of specific binding involved in genomic packaging of the double-stranded-RNA bacteriophage phi6. J Bacteriol. 2003;185(21):6409-14.

39. Paatero AO, Syväoja JE, Bamford DH. Double-stranded RNA bacteriophage phi 6 protein P4 is an unspecific nucleoside triphosphatase activated by calcium ions. J Virol. 1995;69(11):6729-34.

40. Makeyev EV, Grimes JM. RNA-dependent RNA polymerases of dsRNA bacteriophages. Virus Res. 2004;101(1):45-55.

41. Mindich L. Packaging, replication and recombination of the segmented genome of bacteriophage Phi6 and its relatives. Virus Res. 2004;101(1):83-92.

42. Olkkonen VM, Gottlieb P, Strassman J, Qiao XY, Bamford DH, Mindich L. In vitro assembly of infectious nucleocapsids of bacteriophage phi 6: formation of a recombinant double-stranded RNA virus. Proceedings of the National Academy of Sciences. 1990;87(23):9173-7.

43. Poranen MM, Paatero AO, Tuma R, Bamford DH. Self-Assembly of a Viral Molecular Machine from Purified Protein and RNA Constituents. Molecular Cell. 2001;7(4):845-54.

44. Sun X, Ilca SL, Huiskonen JT, Poranen MM. Dual Role of a Viral Polymerase in Viral Genome Replication and Particle Self-Assembly. MBio. 2018;9(5).

45. Ilca SL, Kotecha A, Sun X, Poranen MM, Stuart DI, Huiskonen JT. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat Commun. 2015;6:8843.

46. Haas Fd, de Haas F, Paatero AO, Mindich L, Bamford DH, Fuller SD. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage φ6. Journal of Molecular Biology. 1999;294(2):357-72.

47. Nemecek D, Bernard Heymann J, Qiao J, Mindich L, Steven AC. Cryo-electron tomography of bacteriophage φ6 procapsids shows random occupancy of the binding sites for RNA polymerase and packaging NTPase. Journal of Structural Biology. 2010;171(3):389-96.

48. Sen A, Heymann JB, Cheng N, Qiao J, Mindich L, Steven AC. Initial location of the RNA-dependent RNA polymerase in the bacteriophage Phi6 procapsid determined by cryo-electron microscopy. J Biol Chem. 2008;283(18):12227-31.

49. Katz G, Wei H, Alimova A, Katz A, Morgan DG, Gottlieb P. Protein P7 of the Cystovirus φ6 Is Located at the Three-Fold Axis of the Unexpanded Procapsid. PLoS ONE. 2012;7(10):e47489.

50. Nemecek D, Qiao J, Mindich L, Steven AC, Heymann JB. Packaging Accessory Protein P7 and Polymerase P2 Have Mutually Occluding Binding Sites inside the Bacteriophage 6 Procapsid. Journal of Virology. 2012;86(21):11616-24.

51. Eryilmaz E, Benach J, Su M, Seetharaman J, Dutta K, Wei H, et al. Structure and Dynamics of the P7 Protein from the Bacteriophage ϕ12. Journal of Molecular Biology. 2008;382(2):402-22.

52. Alphonse S, Arnold JJ, Bhattacharya S, Wang H, Kloss B, Cameron CE, et al. Cystoviral Polymerase Complex Protein P7 Uses Its Acidic C-Terminal Tail to Regulate the RNA-Directed RNA Polymerase P2. Journal of Molecular Biology. 2014;426(14):2580-93.

53. Juuti JT, Bamford DH, Tuma R, Thomas GJ. Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage φ6. Journal of Molecular Biology. 1998;279(2):347-59.

54. Ojala PM, Bamford DH. In Vitro Transcription of the Double-Stranded RNA Bacteriophage φ6 Is Influenced by Purine NTPs and Calcium. Virology. 1995;207(2):400-8.

55. Salgado PS, Makeyev EV, Butcher SJ, Bamford DH, Stuart DI, Grimes JM. The structural basis for RNA specificity and Ca2+ inhibition of an RNA-dependent RNA polymerase. Structure. 2004;12(2):307-16.

56. Jäälinoja HT, Huiskonen JT, Butcher SJ. Electron Cryomicroscopy Comparison of the Architectures of the Enveloped Bacteriophages ϕ6 and ϕ8. Structure. 2007;15(2):157-67.

57. Knipe DM, Howley P. Fields Virology: Lippincott Williams & Wilkins; 2013 2013/5/21. 2664 p.

58. International Committee on Taxonomy of V, King AMQ. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses: Elsevier; 2011 2011/11/8. 1338 p.

59. Jenni S, Salgado EN, Herrmann T, Li Z, Grant T, Grigorieff N, et al. In situ structure of rotavirus VP1 RNA-dependent RNA polymerase. J Mol Biol. 2019;431(17):3124-38.

60. Settembre EC, Chen JZ, Dormitzer PR, Grigorieff N, Harrison SC. Atomic model of an infectious rotavirus particle. The EMBO Journal. 2011;30(2):408-16.

61. Estes MK, Graham DY, Mason BB. Proteolytic enhancement of rotavirus infectivity: molecular mechanisms. J Virol. 1981;39(3):879-88.

62. Fiore L, Greenberg HB, Mackow ER. The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin. Virology. 1991;181(2):553-63.

63. Ludert JE, Michelangeli F, Gil F, Liprandi F, Esparza J. Penetration and uncoating of rotaviruses in cultured cells. Intervirology. 1987;27(2):95-101.

64. Chemello ME, Aristimuno OC, Michelangeli F, C. Ruiz M. Requirement for Vacuolar H -ATPase Activity and Ca2 Gradient during Entry of Rotavirus into MA104 Cells. Journal of Virology. 2002;76(24):13083-7.

65. Yoder JD, Trask SD, Vo PT, Binka M, Feng N, Harrison SC, et al. VP5* Rearranges when Rotavirus Uncoats. Journal of Virology. 2009;83(21):11372-7.

66. Trask SD, Kim IS, Harrison SC, Dormitzer PR. A Rotavirus Spike Protein Conformational Intermediate Binds Lipid Bilayers. Journal of Virology. 2010;84(4):1764-70.

67. McClain B, Settembre E, Temple BRS, Richard Bellamy A, Harrison SC. X-ray Crystal Structure of the Rotavirus Inner Capsid Particle at 3.8 Å Resolution. Journal of Molecular Biology. 2010;397(2):587-99.

68. Ding K, Celma CC, Zhang X, Chang T, Shen W, Atanasov I, et al. In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. Nature Communications. 2019;10(1).

69. Levin DH, Mendelsohn N, Schonberg M, Klett H, Silverstein S, Kapuler AM, et al. Properties of RNA Transcriptase in Reovirus Subviral Particles. Proceedings of the National Academy of Sciences. 1970;66(3):890-7.

70. Lawton JA, Estes MK, Prasad BV. Mechanism of genome transcription in segmented dsRNA viruses. Adv Virus Res. 2000;55:185-229.

71. Chen D, Luongo CL, Nibert ML, Patton JT. Rotavirus Open Cores Catalyze 5′-Capping and Methylation of Exogenous RNA: Evidence That VP3 Is a Methyltransferase. Virology. 1999;265(1):120-30.

72. Patton JT. Rotavirus VP1 alone specifically binds to the 3' end of viral mRNA, but the interaction is not sufficient to initiate minus-strand synthesis. J Virol. 1996;70(11):7940-7.

73. Lu X, McDonald SM, Alejandra Tortorici M, Tao YJ, Carpio RV-D, Nibert ML, et al. Mechanism for Coordinated RNA Packaging and Genome Replication by Rotavirus Polymerase VP1. Structure. 2008;16(11):1678-88.

74. Mathieu M, Petitpas I, Navaza J, Lepault J, Kohli E, Pothier P, et al. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. EMBO J. 2001;20(7):1485-97.

75. Poruchynsky MS. Calcium depletion blocks the maturation of rotavirus by altering the oligomerization of virus-encoded proteins in the ER. The Journal of Cell Biology. 1991;114(4):651-6.

76. Lopez T, Camacho M, Zayas M, Najera R, Sanchez R, Arias CF, et al. Silencing the Morphogenesis of Rotavirus. Journal of Virology. 2005;79(1):184-92.

77. Silvestri LS, Tortorici MA, Vasquez-Del Carpio R, Patton JT. Rotavirus Glycoprotein NSP4 Is a Modulator of Viral Transcription in the Infected Cell. Journal of Virology. 2005;79(24):15165-74.

78. Grant T, Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. eLife. 2015;4.

79. Estrozi LF, Settembre EC, Goret G, McClain B, Zhang X, Chen JZ, et al. Location of the dsRNA-dependent polymerase, VP1, in rotavirus particles. 2012.

80. Lawton JA, Estes MK, Prasad BV. Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles. Nat Struct Biol. 1997;4(2):118-21.

81. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue virus VP4 is an RNA-capping assembly line. Nature Structural & Molecular Biology. 2007;14(5):449-51.

82. Frank J. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State: Oxford University Press; 2006. 432 p.

83. Elmlund D, Elmlund H. Cryogenic electron microscopy and single-particle analysis. Annu Rev Biochem. 2015;84:499-517.

84. Jiang W, Tang L. Atomic cryo-EM structures of viruses. Curr Opin Struct Biol. 2017;46:122-9.

85. Fan X, Wang J, Zhang X, Yang Z, Zhang J-C, Zhao L, et al. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. Nat Commun. 2019;10(1):2386.

86. Scheres SHW. Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzymol. 2010;482:295-320.

87. Bai X-C, McMullan G, Scheres SHW. How cryo-EM is revolutionizing structural biology. Trends Biochem Sci. 2015;40(1):49-57.

88. Bai X-C, Rajendra E, Yang G, Shi Y, Scheres SHW. Sampling the conformational space of the catalytic subunit of human γ-secretase. eLife Sciences. 2015;4:e11182.

89. Li X, Liu H, Cheng L. Symmetry-mismatch reconstruction of genomes and associated proteins within icosahedral viruses using cryo-EM. Biophys Rep. 2016;2(1):25-32.

90. Huiskonen JT. Image processing for cryogenic transmission electron microscopy of symmetry-mismatched complexes. Biosci Rep. 2018.

91. Stass R, Ilca SL, Huiskonen JT. Beyond structures of highly symmetric purified viral capsids by cryo-EM. Curr Opin Struct Biol. 2018;52:25-31.

92. Jia H, Dong K, Zhou L, Wang G, Hong N, Jiang D, et al. A dsRNA virus with filamentous viral particles. Nat Commun. 2017;8(1):168.

93. Luque D, Mata CP, Suzuki N, Ghabrial SA, Caston JR. Capsid Structure of dsRNA Fungal Viruses. Viruses. 2018;10(9).

94. Grigorieff N, Harrison SC. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr Opin Struct Biol. 2011;21(2):265-73.

95. Noble AJ, Wei H, Dandey VP, Zhang Z, Tan YZ, Potter CS, et al. Reducing effects of particle adsorption to the air-water interface in cryo-EM. Nat Methods. 2018;15(10):793-5.

96. Gorzelnik KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J. Asymmetric cryo-EM structure of the canonical Allolevivirus Qbeta reveals a single maturation protein and the genomic ssRNA in situ. Proc Natl Acad Sci U S A. 2016;113(41):11519-24.

97. Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, et al. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature. 2017;541(7635):112-6.

98. Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC. Encapsidated Conformation of Bacteriophage T7 DNA. Cell. 1997;91(2):271-80.

99. Jiang W, Chang J, Jakana J, Weigele P, King J, Chiu W. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature. 2006;439(7076):612-6.

100. Liu H, Cheng L. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. Science. 2015;349(6254):1347-50.

101. Zhang X, Ding K, Yu XK, Chang W, Sun JC, Zhou ZH. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. 2015.

102. Ding K, Nguyen L, Zhou ZH. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. J Virol. 2018;92(21).

103. Wang X, Zhang F, Su R, Li X, Chen W, Chen Q, et al. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. Proc Natl Acad Sci U S A. 2018;115(28):7344-9.

104. Morais MC, Tao Y, Olson NH, Grimes S, Jardine PJ, Anderson DL, et al. Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29. J Struct Biol. 2001;135(1):38-46.

105. Ferrero D, Ferrer-Orta C, Verdaguer N. Viral RNA-Dependent RNA Polymerases: A Structural Overview. Subcell Biochem. 2018;88:39-71.

106. Briggs JA, Huiskonen JT, Fernando KV, Gilbert RJ, Scotti P, Butcher SJ, et al. Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles. J Struct Biol. 2005;150(3):332-9.

107. Huiskonen JT, Jaalinoja HT, Briggs JA, Fuller SD, Butcher SJ. Structure of a hexameric RNA packaging motor in a viral polymerase complex. J Struct Biol. 2007;158(2):156-64.

108. Scheres SH. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods Enzymol. 2016;579:125-57.

109. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife. 2018;7.

110. He Y, Shivakoti S, Ding K, Cui Y, Roy P, Zhou ZH. In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating. Proc Natl Acad Sci U S A. 2019;116(33):16535-40.

111. Bamford DH, Ojala PM, Frilander M, Walin L, Bamford JKH. [25] Isolation, purification, and function of assembly intermediates and subviral particles of bacteriophages PRD1 and σ6. In: Adolph KW, editor. Methods in Molecular Genetics. 6: Academic Press; 1995. p. 455-74.

112. Patton JT, Chizhikov V, Taraporewala Z, Chen D. Virus replication. Methods Mol Med. 2000;34:33-66.

113. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180(3):519-30.

114. Scheres SHW. A Bayesian View on Cryo-EM Structure Determination. J Mol Biol. 2012;415(2):406-18.

115. Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, et al. Xmipp: An Image Processing Package for Electron Microscopy. J Struct Biol. 1996;116(1):237-40.

116. de la Rosa-Trevin JM, Oton J, Marabini R, Zaldivar A, Vargas J, Carazo JM, et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J Struct Biol. 2013;184(2):321-8.

117. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods. 2013;10(6):584-90.

118. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol. 2003;142(3):334-47.

119. Kivioja T, Ravantti J, Verkhovsky A, Ukkonen E, Bamford D. Local average intensity-based method for identifying spherical particles in electron micrographs. J Struct Biol. 2000;131(2):126-34.

120. El Omari K, Meier C, Kainov D, Sutton G, Grimes JM, Poranen MM, et al. Tracking in atomic detail the functional specializations in viral RecA helicases that occur during evolution. Nucleic Acids Res. 2013;41(20):9396-410.

121. de la Rosa-Trevin JM, Quintana A, Del Cano L, Zaldivar A, Foche I, Gutierrez J, et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. J Struct Biol. 2016;195(1):93-9.

122. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J Struct Biol. 2007;157(1):281-7.

123. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veesler D, et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure. 2012;20(11):1823-8.

124. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):213-21.

125. Nicholls RA, Long F, Murshudov GN. Low-resolution refinement tools in REFMAC5. Acta Crystallogr D Biol Crystallogr. 2012;68(Pt 4):404-17.

126. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2126-32.

127. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. Nat Methods. 2015;12(4):361-5.

128. Dickerson RE. DNA structure from A to Z. Methods Enzymol. 1992;211:67-111.

129. Goodsell DS, Dickerson RE. Bending and curvature calculations in B-DNA. Nucleic Acids Res. 1994;22(24):5497-503.

130. Ilca SL, Sun X, El Omari K, Kotecha A, de Haas F, DiMaio F, et al. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. Nature. 2019.

131. Lepault J, Dubochet J, Baschong W, Kellenberger E. Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. EMBO J. 1987;6(5):1507-12.

132. Black LW. DNA packaging in dsDNA bacteriophages. Annu Rev Microbiol. 1989;43:267-92.

133. Booy FP, Newcomb WW, Trus BL, Brown JC, Baker TS, Steven AC. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. Cell. 1991;64(5):1007-15.

134. LaMarque JC, Le TV, Harvey SC. Packaging double-helical DNA into viral capsids. Biopolymers. 2004;73(3):348-55.

135. Speir JA, Johnson JE. Nucleic acid packaging in viruses. Curr Opin Struct Biol. 2012;22(1):65-71.

136. Frugier M, Schimmel P. Subtle atomic group discrimination in the RNA minor groove. Proc Natl Acad Sci U S A. 1997;94(21):11291-4.

137. Yang S, Yu X, VanLoock MS, Jezewska MJ, Bujalowski W, Egelman EH. Flexibility of the rings: structural asymmetry in the DnaB hexameric helicase. J Mol Biol. 2002;321(5):839-49.

138. Ye J, Osborne AR, Groll M, Rapoport TA. RecA-like motor ATPases--lessons from structures. Biochim Biophys Acta. 2004;1659(1):1-18.

139. Gouet P, Diprose JM, Grimes JM, Malby R, Burroughs JN, Zientara S, et al. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. Cell. 1999;97(4):481-90.

140. Bamford JK, Bamford DH, Li T, Thomas GJ, Jr. Structural studies of the enveloped dsRNA bacteriophage phi 6 of Pseudomonas syringae by Raman spectroscopy. II. Nucleocapsid structure and thermostability of the virion, nucleocapsid and polymerase complex. J Mol Biol. 1993;230(2):473-82.

141. Caston JR, Trus BL, Booy FP, Wickner RB, Wall JS, Steven AC. Structure of L-A virus: a specialized compartment for the transcription and replication of double-stranded RNA. J Cell Biol. 1997;138(5):975-85.

142. Abels JA, Moreno-Herrero F, van der Heijden T, Dekker C, Dekker NH. Single-molecule measurements of the persistence length of double-stranded RNA. Biophys J. 2005;88(4):2737-44.

143. Gross P, Laurens N, Oddershede LB, Bockelmann U, Peterman EJG, Wuite GJL. Quantifying how DNA stretches, melts and changes twist under tension. Nature Physics. 2011;7(9):731-6.

144. Jin L, Shi YZ, Feng CJ, Tan YL, Tan ZJ. Modeling Structure, Stability, and Flexibility of Double-Stranded RNAs in Salt Solutions. Biophys J. 2018;115(8):1403-16.

145. Frilander M, Poranen M, Bamford DH. The large genome segment of dsRNA bacteriophage phi6 is the key regulator in the in vitro minus and plus strand synthesis. RNA. 1995;1(5):510-8.

146. Coplin DL, Van Etten JL, Koski RK, Vidaver AK. Intermediates in the biosynthesis of double-stranded ribonucleic acids of bacteriophage phi 6. Proc Natl Acad Sci U S A. 1975;72(3):849-53.

147. Strzelecka TE, Davidson MW, Rill RL. Multiple liquid crystal phases of DNA at high concentrations. Nature. 1988;331(6155):457-60.

148. Rey AD. Liquid crystal models of biological materials and processes. Soft Matter. 2010;6(15):3402-29.

149. Livolant F, Leforestier A. Condensed phases of DNA: Structures and phase transitions. Prog Polym Sci. 1996;21(6):1115-64.

150. Knobler CM, Gelbart WM. Physical chemistry of DNA viruses. Annu Rev Phys Chem. 2009;60:367-83.

151. Podgornik R, Strey HH, Gawrisch K, Rau DC, Rupprecht A, Parsegian VA. Bond orientational order, molecular motion, and free energy of high-density DNA mesophases. Proc Natl Acad Sci U S A. 1996;93(9):4261-6.

152. Olkkonen VM, Bamford DH. The nucleocapsid of the lipid-containing double-stranded RNA bacteriophage phi 6 contains a protein skeleton consisting of a single polypeptide species. J Virol. 1987;61(8):2362-7.

153. Olkkonen VM, Ojala PM, Bamford DH. Generation of infectious nucleocapsids by in vitro assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage phi 6. J Mol Biol. 1991;218(3):569-81.

154. Cvirkaite-Krupovic V, Poranen MM, Bamford DH. Phospholipids act as secondary receptor during the entry of the enveloped, double-stranded RNA bacteriophage phi6. J Gen Virol. 2010;91(Pt 8):2116-20.

155. Kainov DE, Lísal J, Bamford DH, Tuma R. Packaging motor from double-stranded RNA bacteriophage ϕ12 acts as an obligatory passive conduit during transcription. Nucleic Acids Res. 2004;32(12):3515-21.

156. Buck KW. Replication of double-stranded RNA in particles of Penicillium stoloniferum virus S. Nucleic Acids Res. 1975;2(10):1889-902.

157. Ratti G, Buck KW. Semi-conservative transcription in particles of a double-stranded RNA mycovirus. Nucleic Acids Res. 1978;5(10):3843-54.

158. Spies U, Muller H, Becht H. Properties of RNA polymerase activity associated with infectious bursal disease virus and characterization of its reaction products. Virus Res. 1987;8(2):127-40.

159. Kim JW, Kim SY, Kim KM. Genome organization and expression of the Penicillium stoloniferum virus S. Virus Genes. 2003;27(3):249-56.

160. Kozlakidis Z, Herrero N, Ozkan S, Bhatti MF, Coutts RH. A novel dsRNA element isolated from the Aspergillus foetidus mycovirus complex. Arch Virol. 2013;158(12):2625-8.

161. Tao Y, Farsetta DL, Nibert ML, Harrison SC. RNA synthesis in a cage--structural studies of reovirus polymerase lambda3. Cell. 2002;111(5):733-45.

162. Yazaki K, Miura K. Relation of the structure of cytoplasmic polyhedrosis virus and the synthesis of its messenger RNA. Virology. 1980;105(2):467-79.

163. Patton JT, Jones MT, Kalbach AN, He YW, Xiaobo J. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. J Virol. 1997;71(12):9618-26.

164. Qiao J, Mindich L. The template specificity of bacteriophage Phi6 RNA polymerase. J Virol. 2013;87(18):10190-4.

165. Guglielmi KM, McDonald SM, Patton JT. Mechanism of intraparticle synthesis of the rotavirus double-stranded RNA genome. J Biol Chem. 2010;285(24):18123-8.

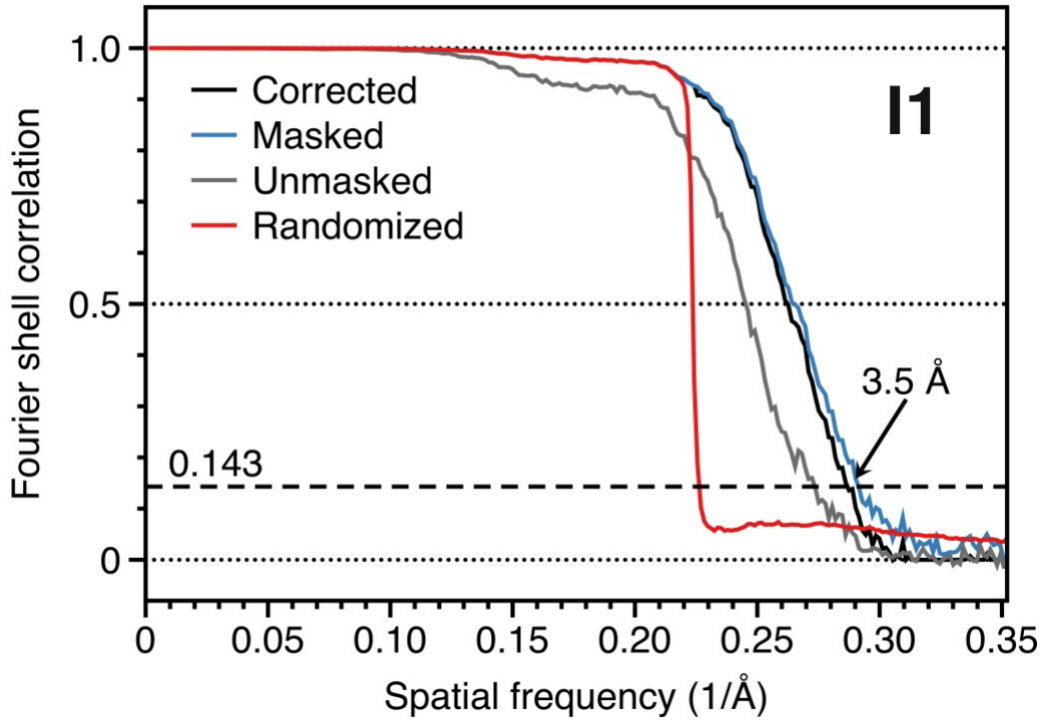
166. Reinisch KM, Nibert ML, Harrison SC. Structure of the reovirus core at 3.6 A resolution. Nature. 2000;404(6781):960-7.

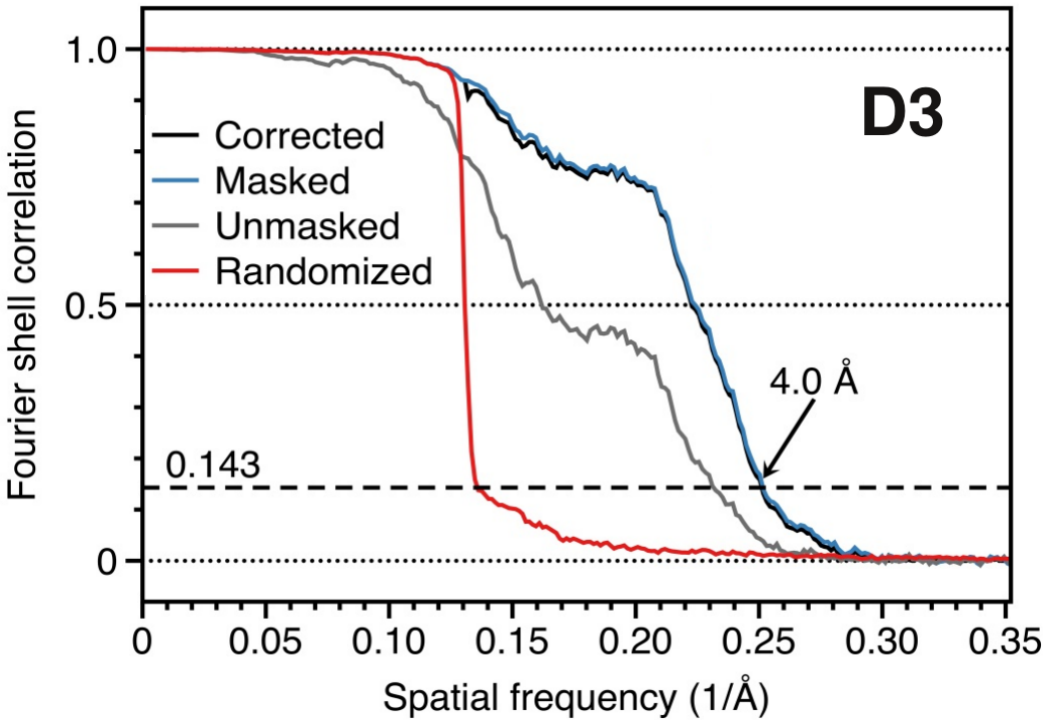
167. Zhang X, Ji Y, Zhang L, Harrison SC, Marinescu DC, Nibert ML, et al. Features of reovirus outer capsid protein mu1 revealed by electron cryomicroscopy and image reconstruction of the virion at 7.0 Angstrom resolution. Structure. 2005;13(10):1545-57.

Appendix

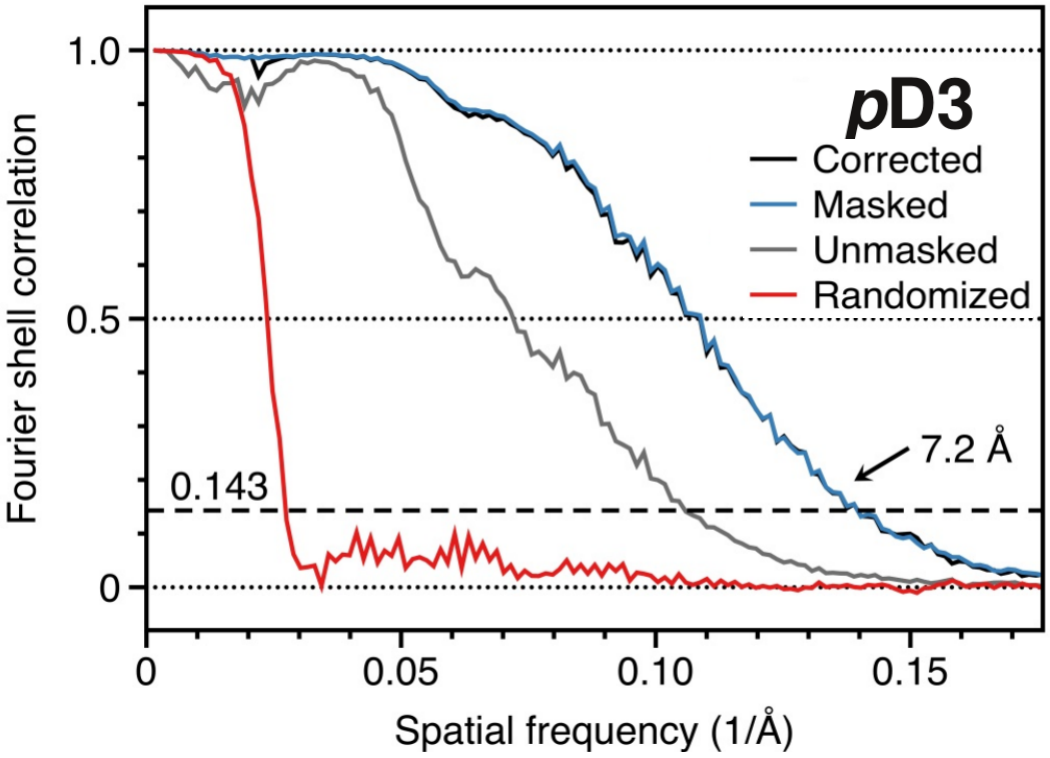
1. Fourier Shell Correlation plots

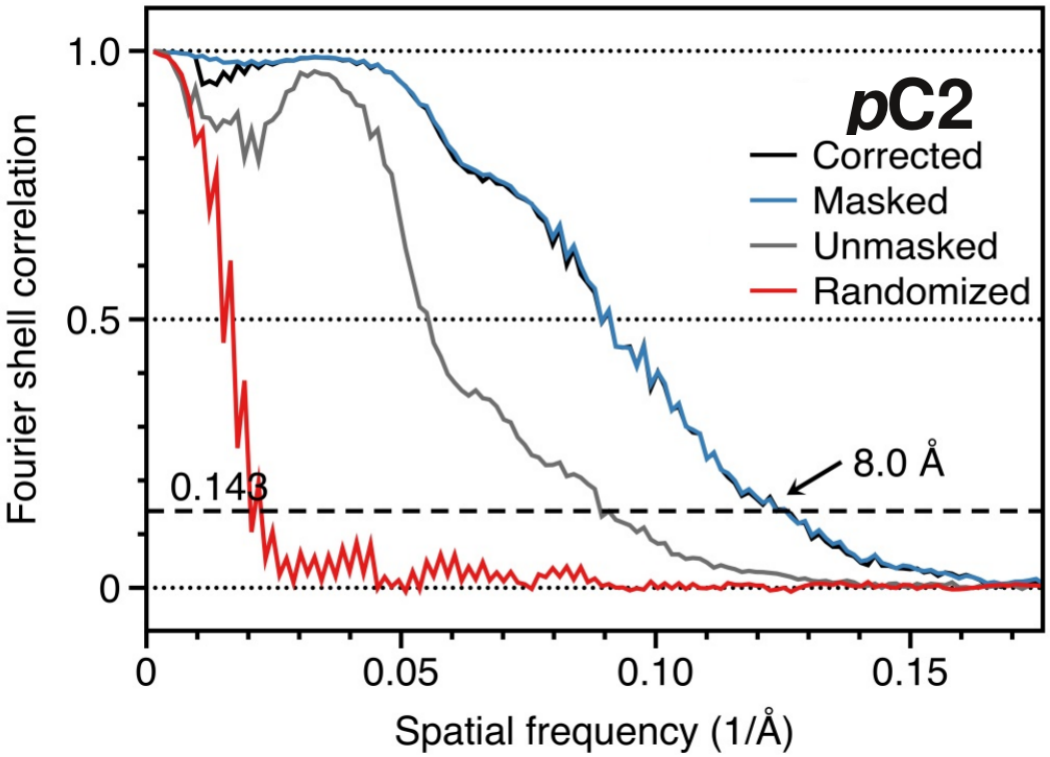
1.1. ϕ6 QNC capsid with I1 and D3 symmetry (adapted from [130])

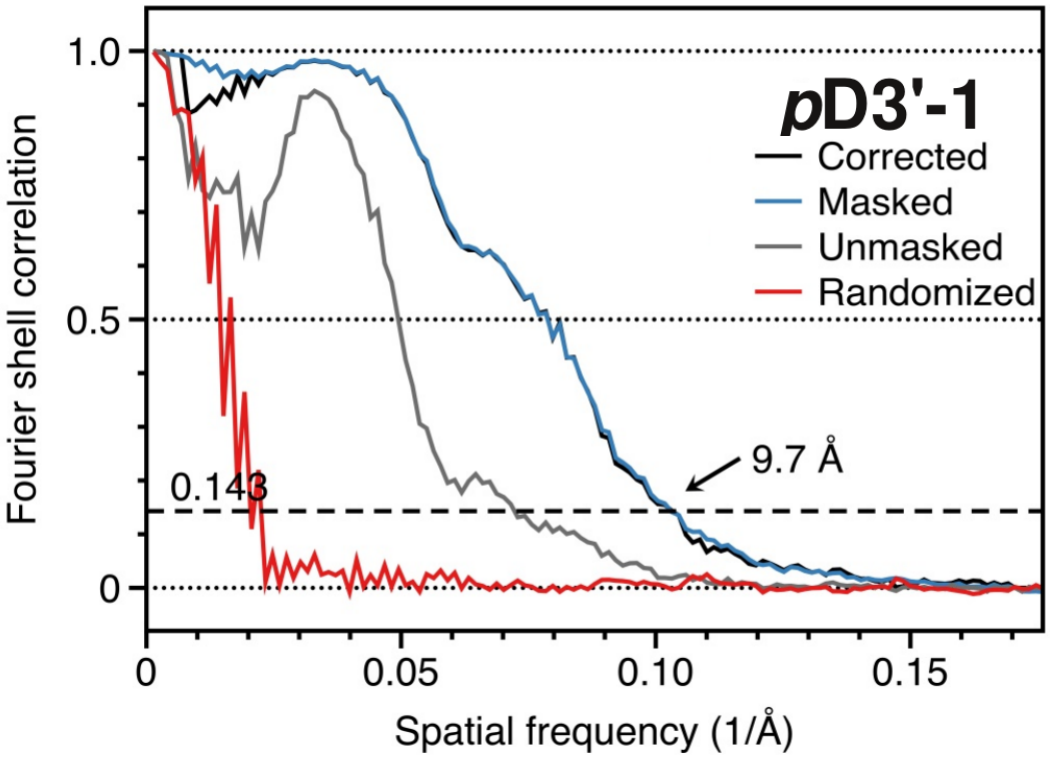




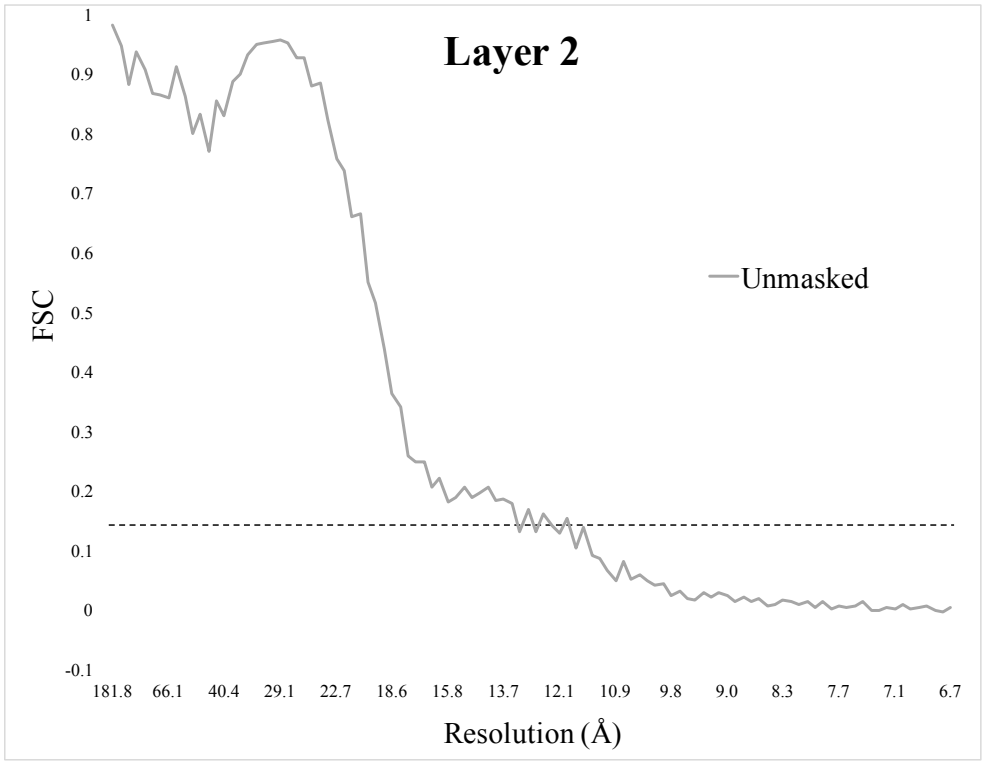
1.2. ϕ6 QNC layer 1 with no symmetry imposed (adapted from [130])

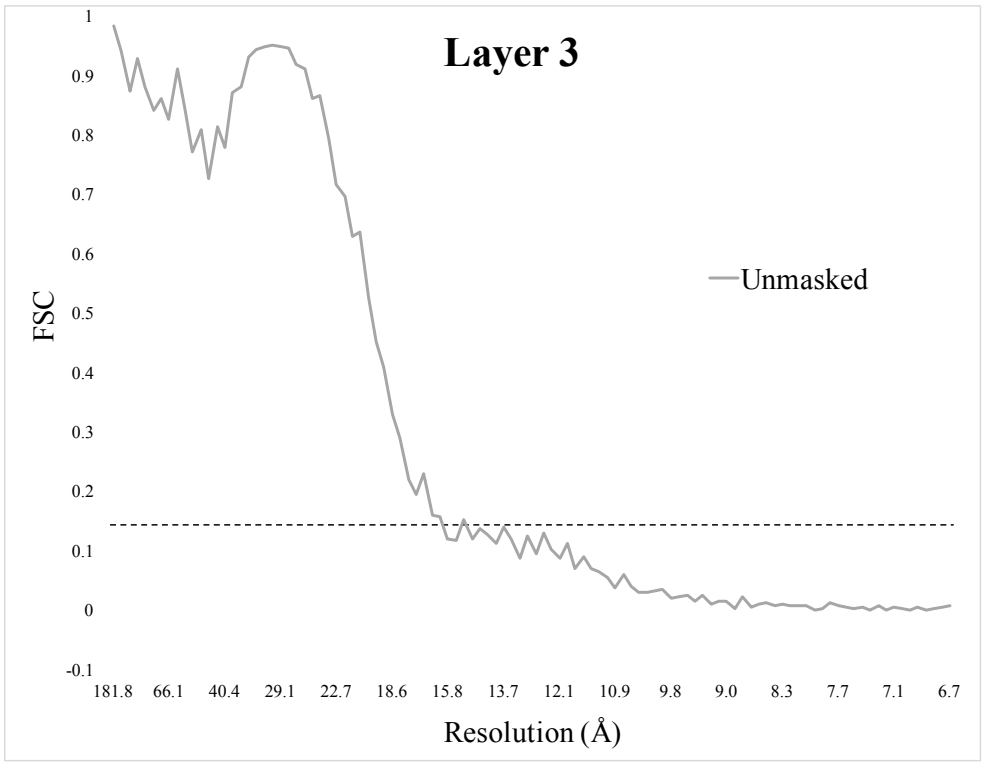


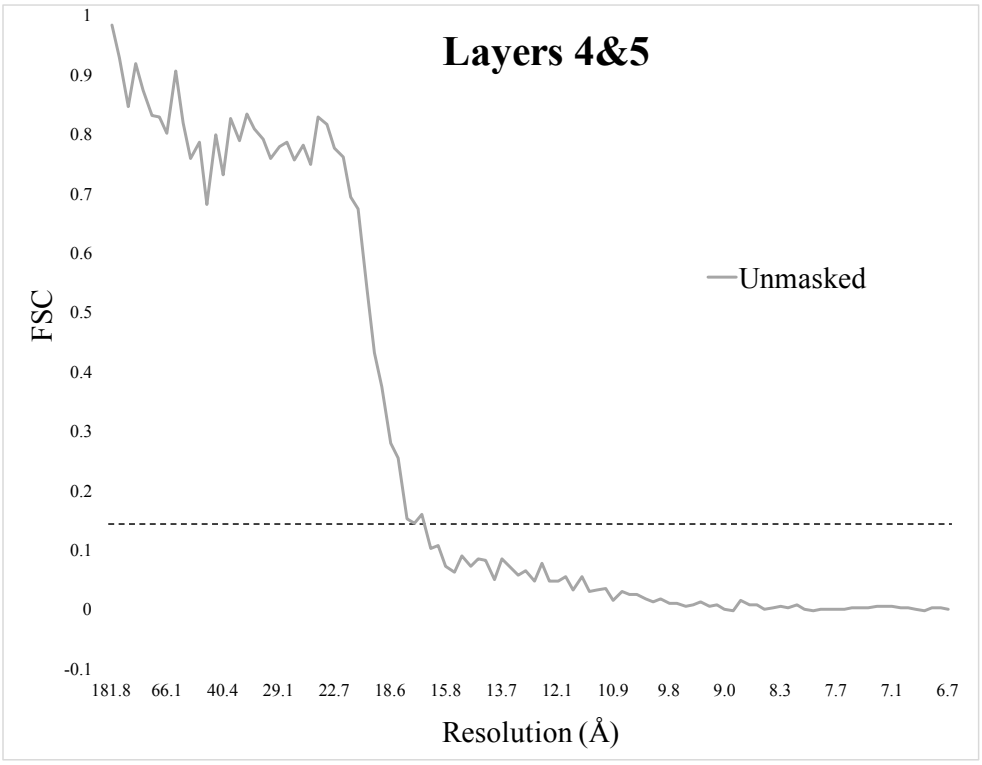



pD3'-1
Corrected
Masked
Unmasked
Randomized
1.0
0.5
0
0.143
9.7 Å
0
0.05
0.10
0.15
Fourier shell correlation
Spatial frequency (1/Å)

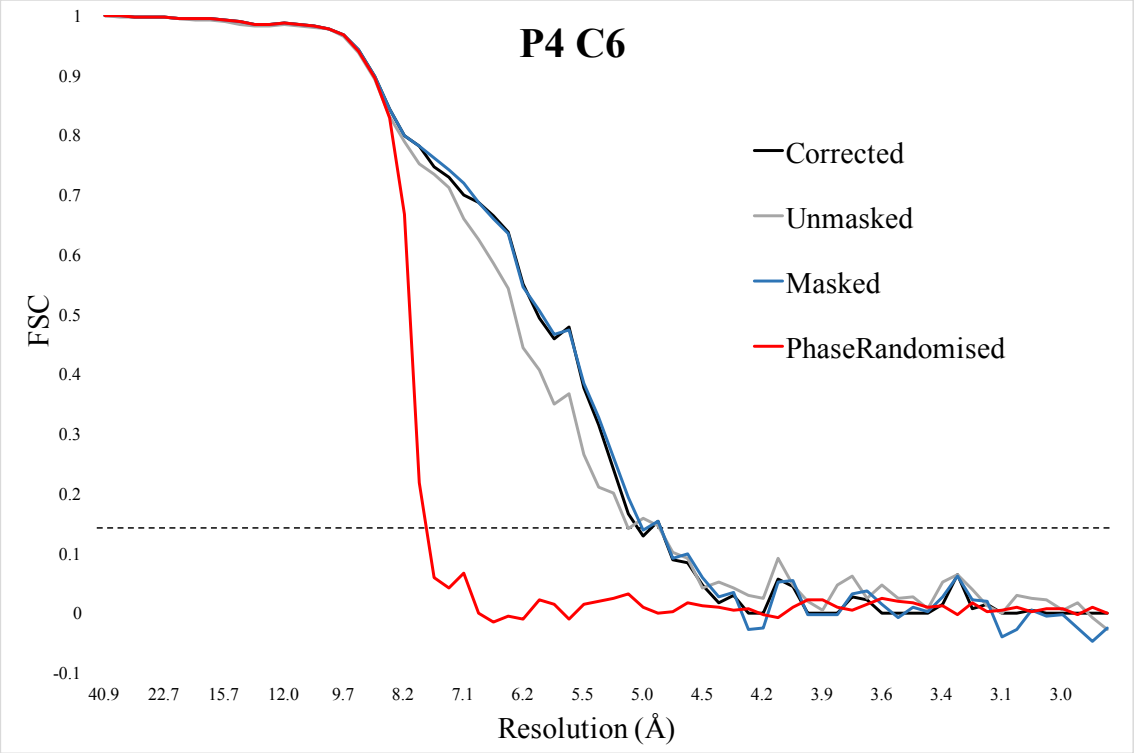
1.3. ϕ6 QNC layers 2, 3 and 4&5 after 3D refinement (no symmetry)

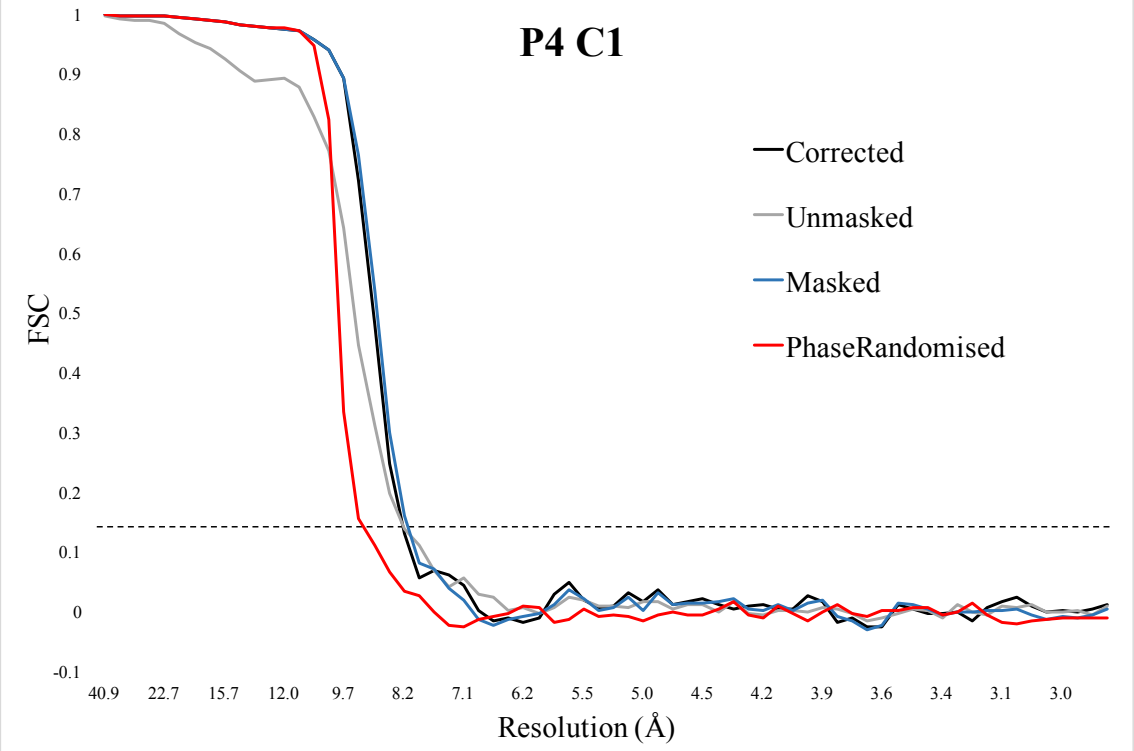




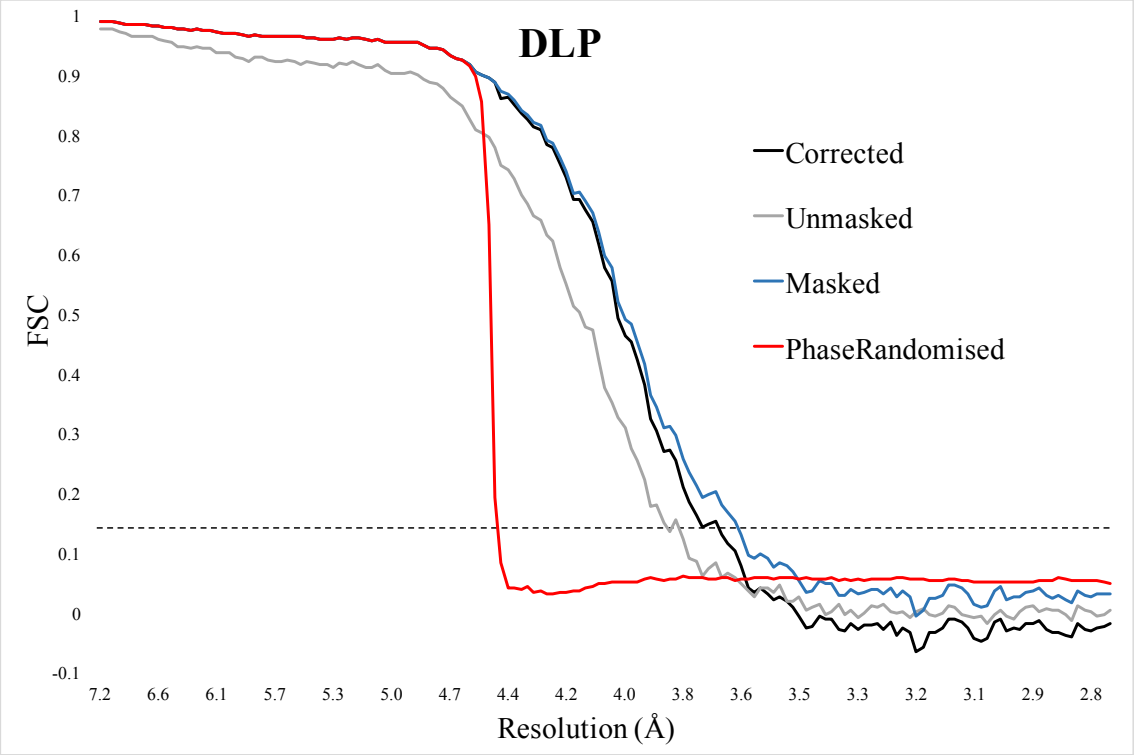

Layers 4&5
FSC
Resolution (Å)
Unmasked
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
181.8
66.1
40.4
29.1
22.7
18.6
15.8
13.7
12.1
10.9
9.8
9.0
8.3
7.7
7.1
6.7

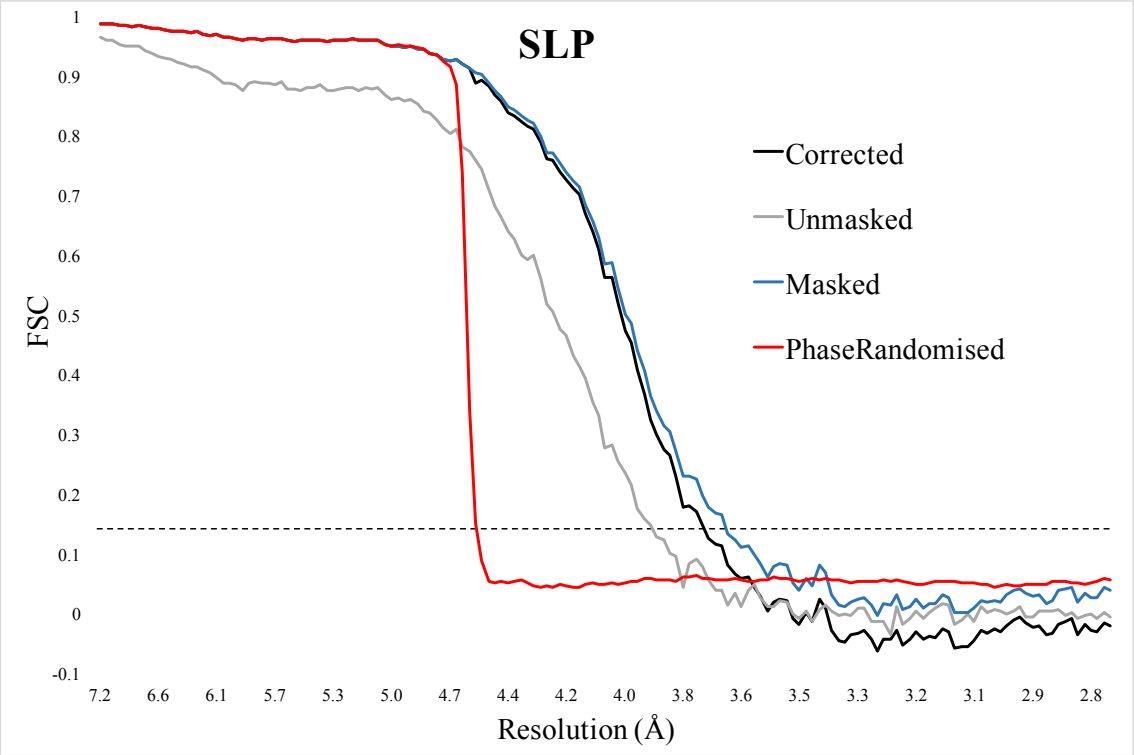
1.4. C6-symmetric and asymmetric P4 from ϕ6 QNC



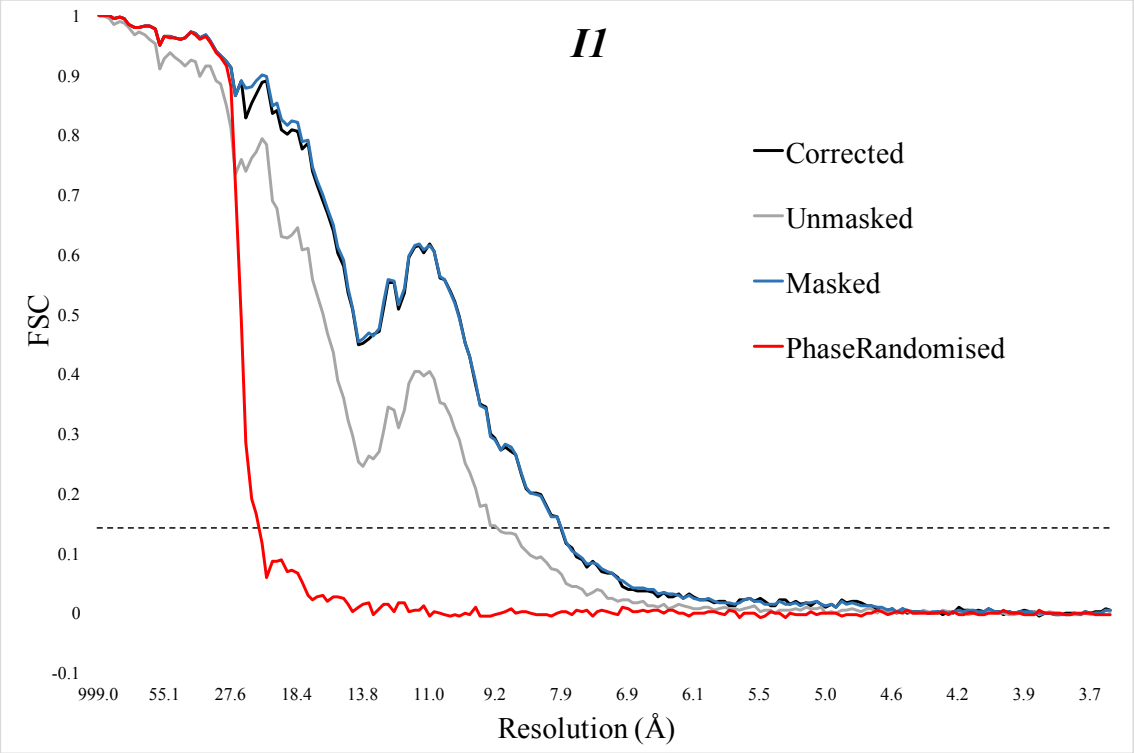


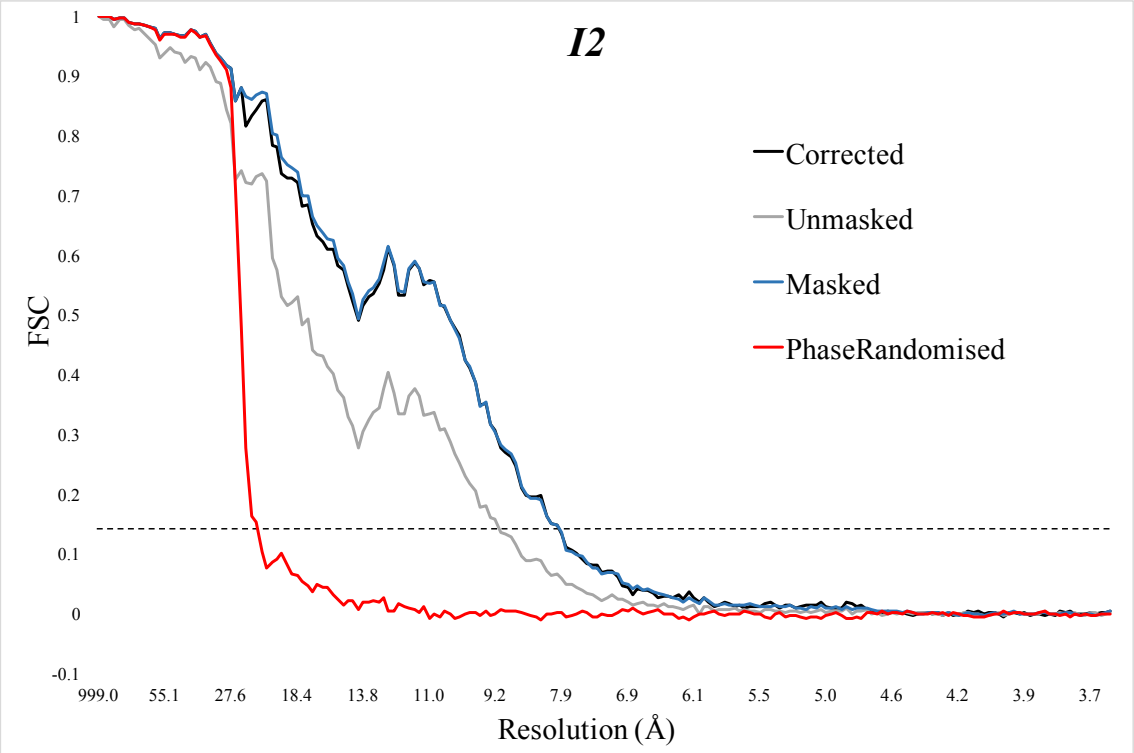
1.5. DLP and SLP capsid from ϕ6 DTNC

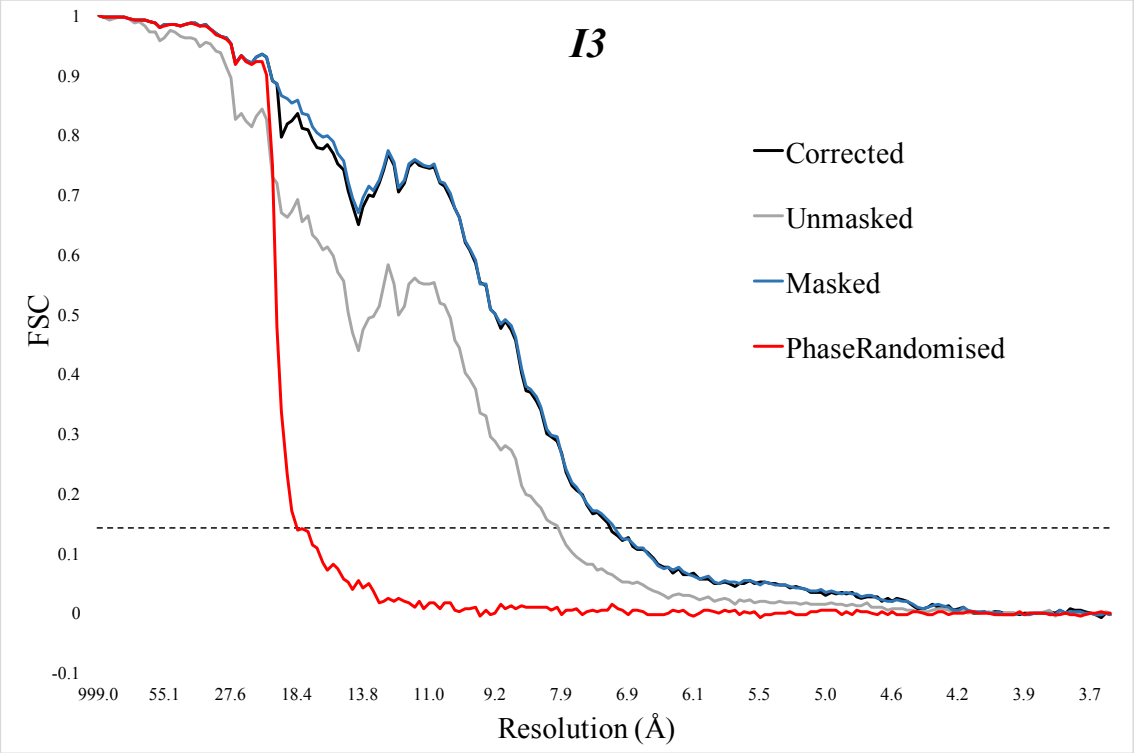


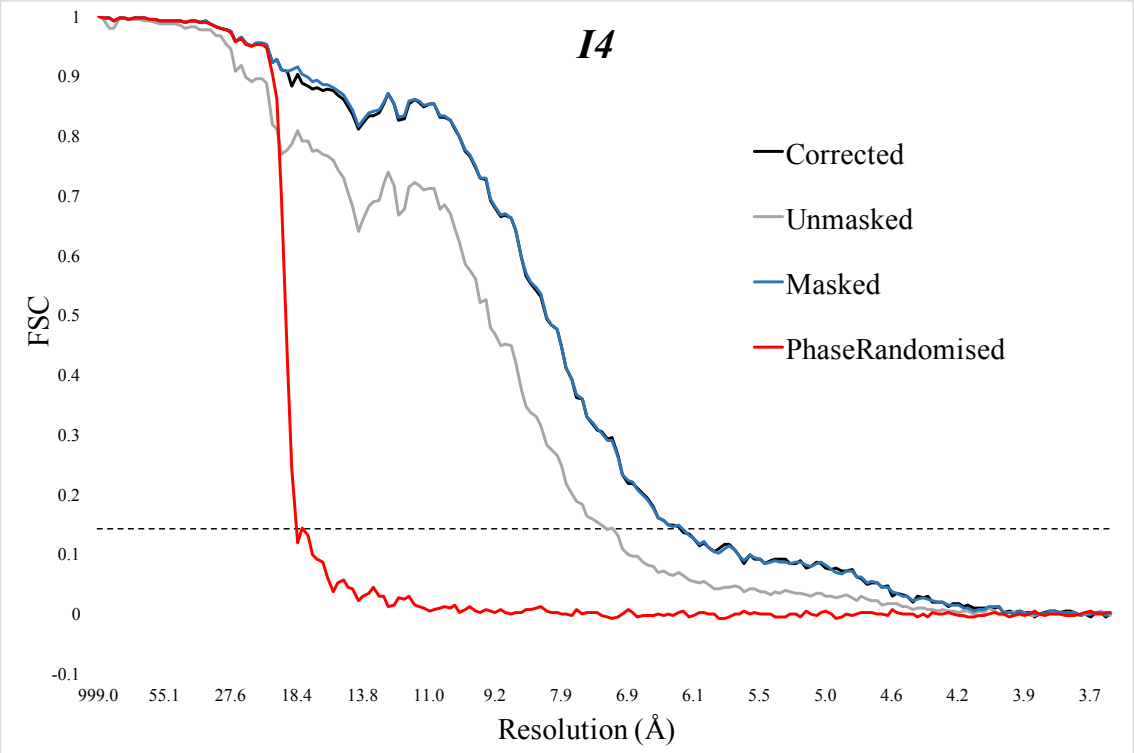


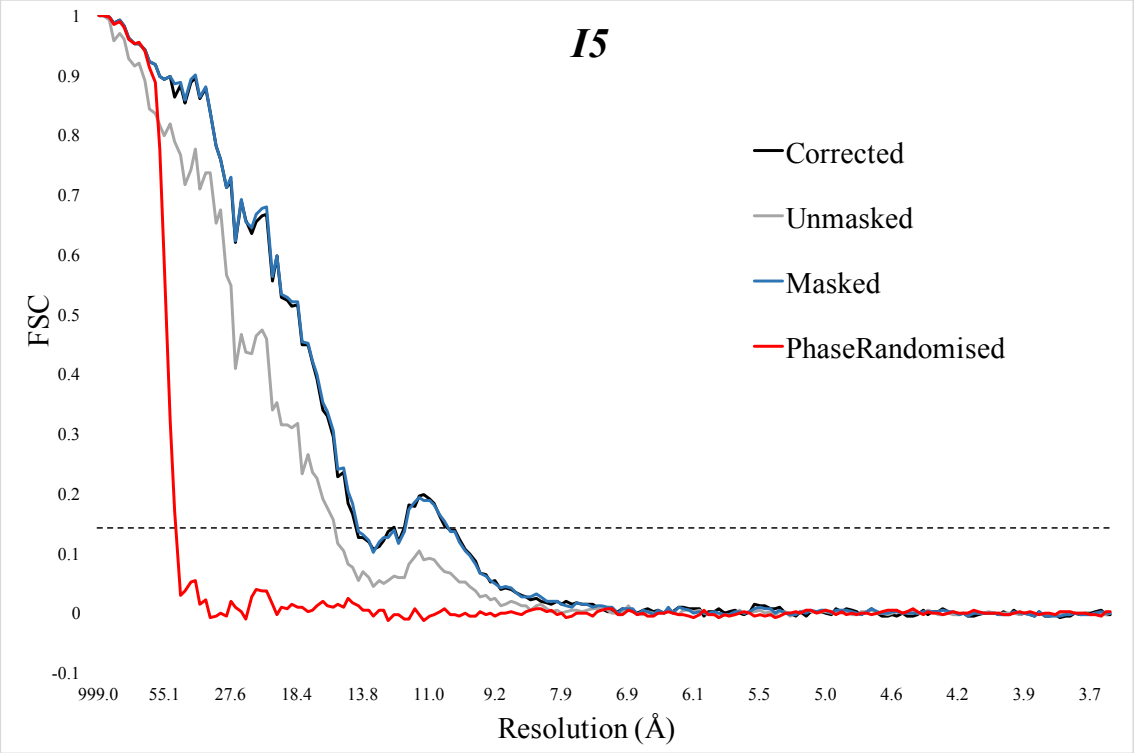
1.6. Disassembly intermediates from ϕ6 DTNC (I1-9)

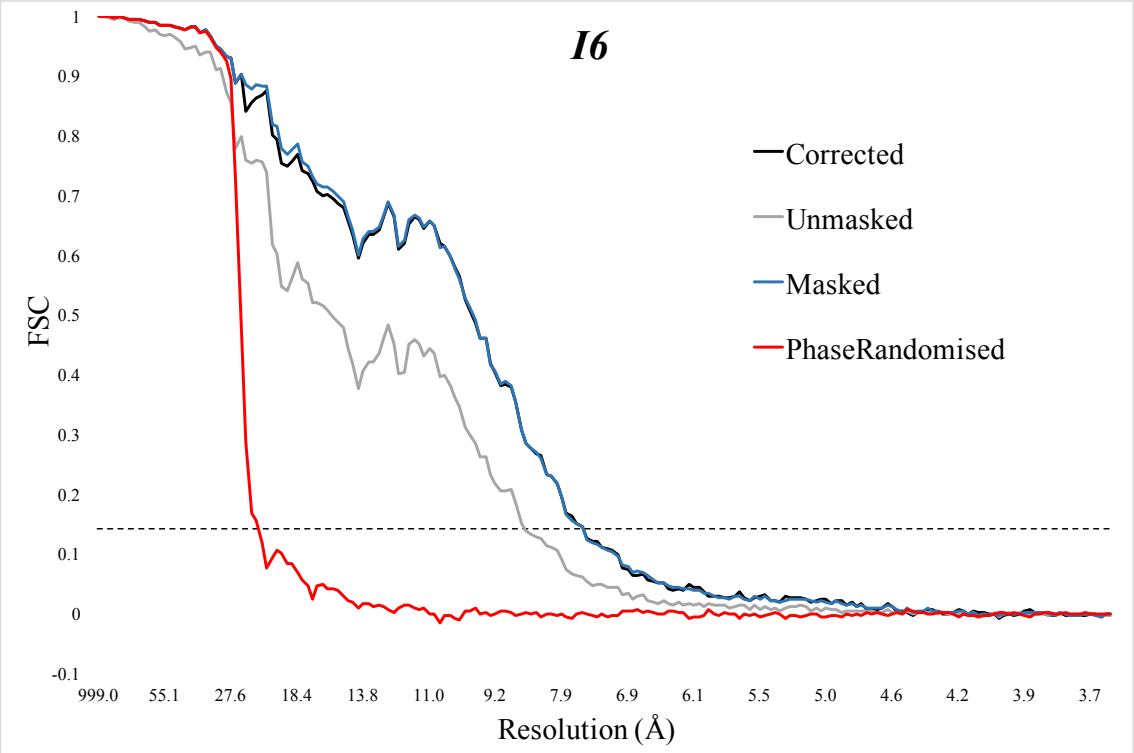


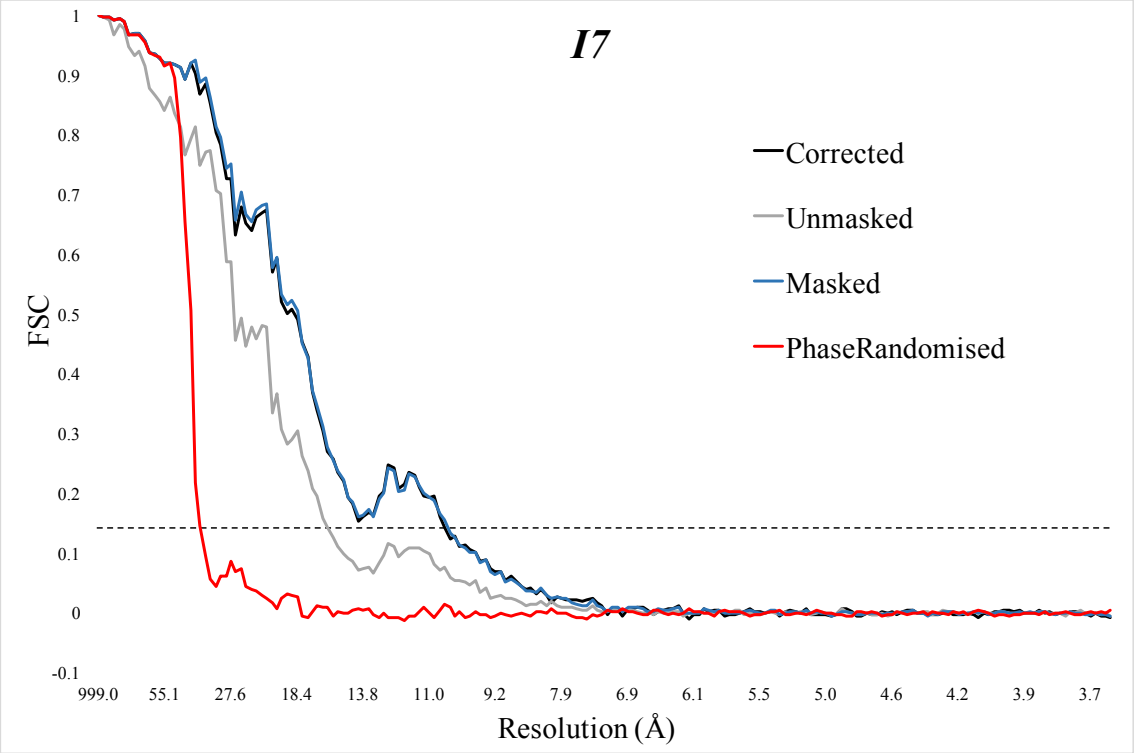


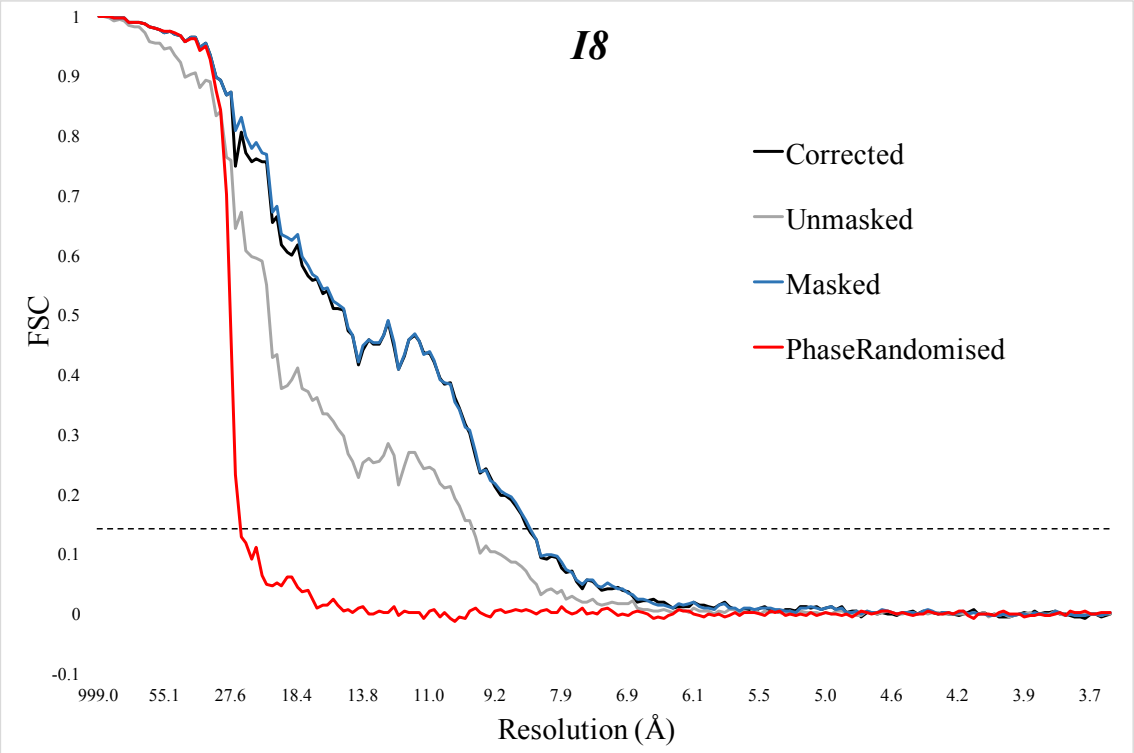

I3
Corrected
Unmasked
Masked
PhaseRandomised
FSC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0 55.1 27.6 18.4 13.8 11.0 9.2 7.9 6.9 6.1 5.5 5.0 4.6 4.2 3.9 3.7
Resolution (Å)

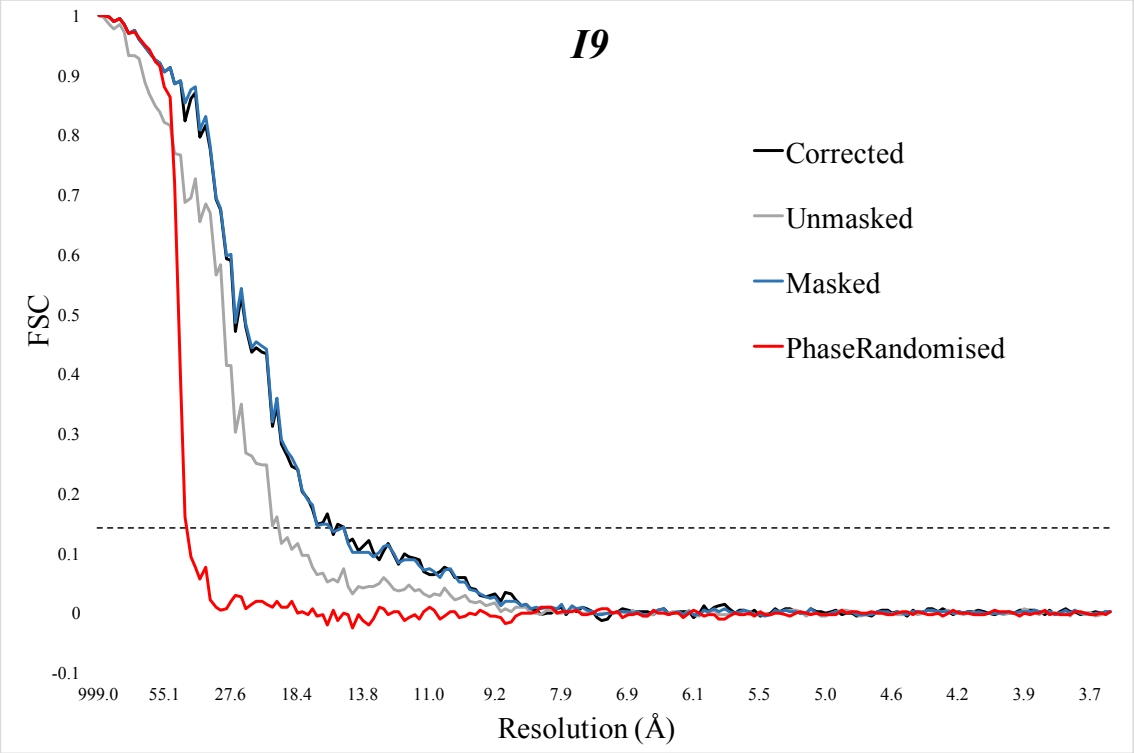

I4
Corrected
Unmasked
Masked
PhaseRandomised
FSC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0 55.1 27.6 18.4 13.8 11.0 9.2 7.9 6.9 6.1 5.5 5.0 4.6 4.2 3.9 3.7
Resolution (Å)


I5
Corrected
Unmasked
Masked
PhaseRandomised
FSC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0 55.1 27.6 18.4 13.8 11.0 9.2 7.9 6.9 6.1 5.5 5.0 4.6 4.2 3.9 3.7
Resolution (Å)

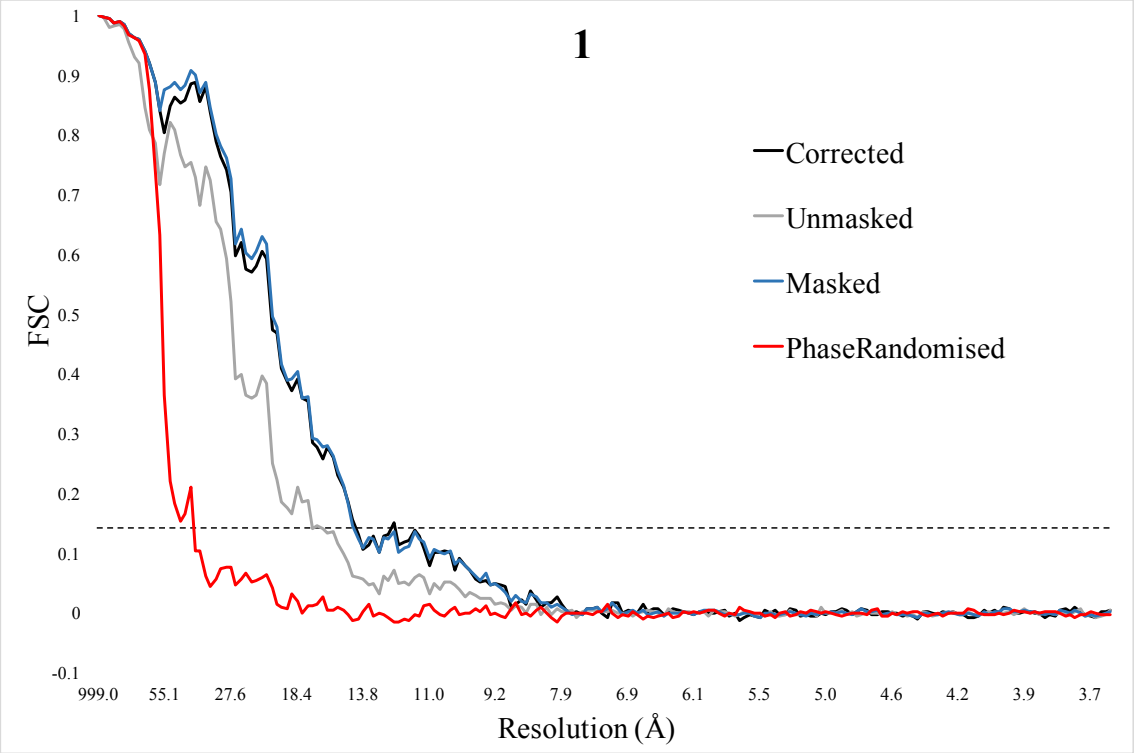

I6
Corrected
Unmasked
Masked
PhaseRandomised
FSC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0 55.1 27.6 18.4 13.8 11.0 9.2 7.9 6.9 6.1 5.5 5.0 4.6 4.2 3.9 3.7
Resolution (Å)

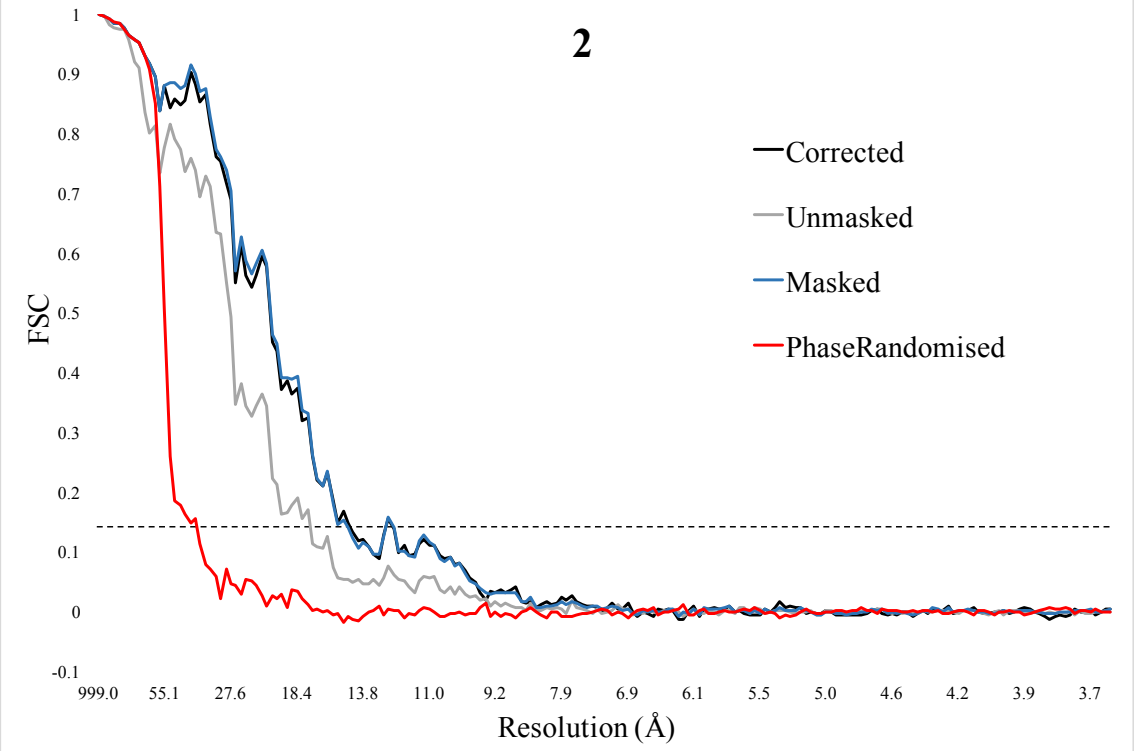

I7
FSC
Resolution (Å)
Corrected
Unmasked
Masked
PhaseRandomised
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0
55.1
27.6
18.4
13.8
11.0
9.2
7.9
6.9
6.1
5.5
5.0
4.6
4.2
3.9
3.7

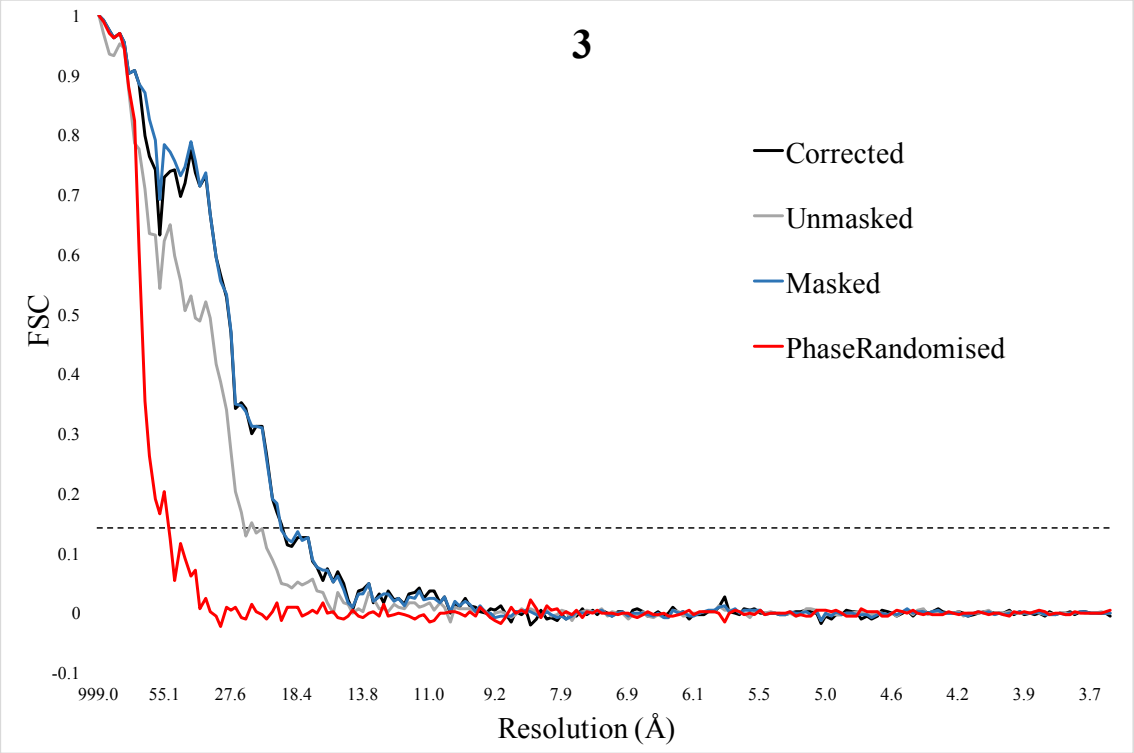

I8
FSC
Resolution (Å)
Corrected
Unmasked
Masked
PhaseRandomised
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0
55.1
27.6
18.4
13.8
11.0
9.2
7.9
6.9
6.1
5.5
5.0
4.6
4.2
3.9
3.7


I9
FSC
Resolution (Å)
Corrected
Unmasked
Masked
PhaseRandomised
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0
55.1
27.6
18.4
13.8
11.0
9.2
7.9
6.9
6.1
5.5
5.0
4.6
4.2
3.9
3.7

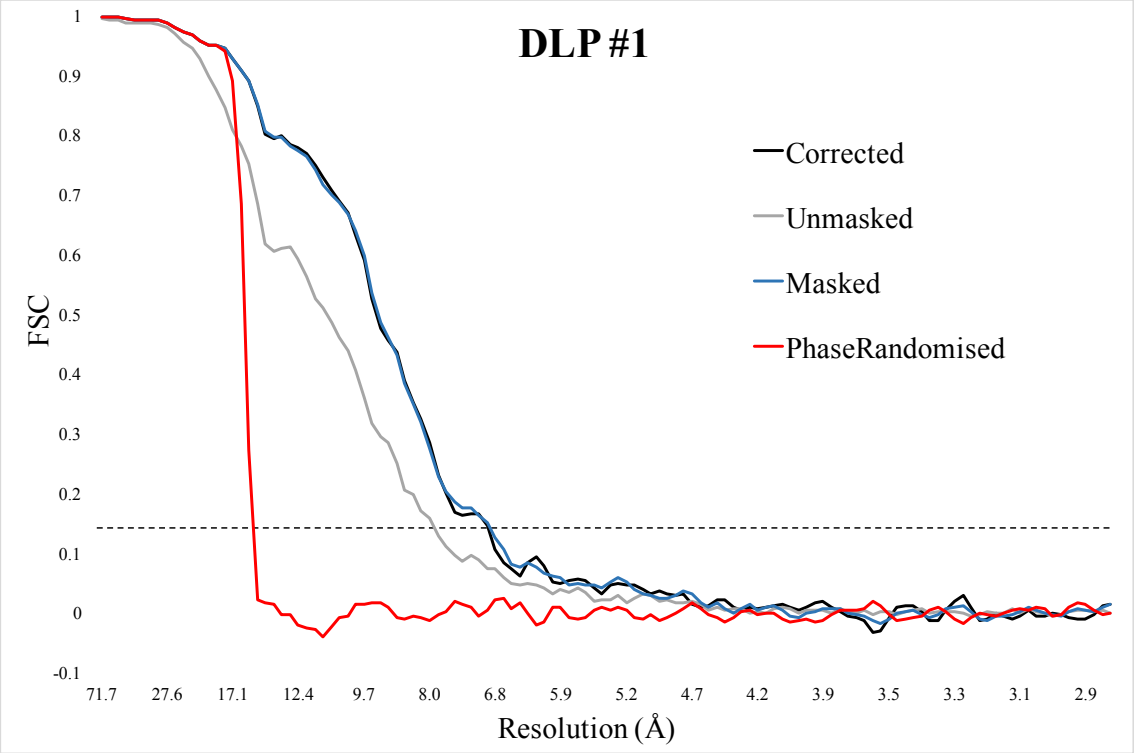
1.7. Peculiar intermediates from ϕ6 DTNC

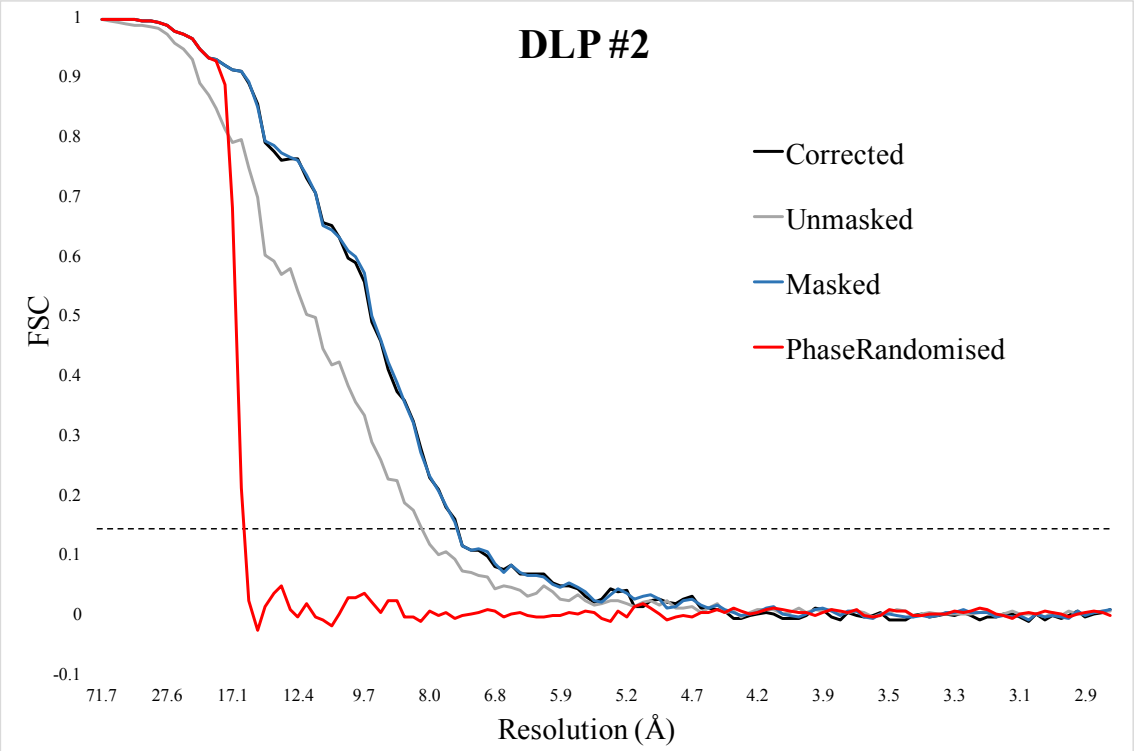




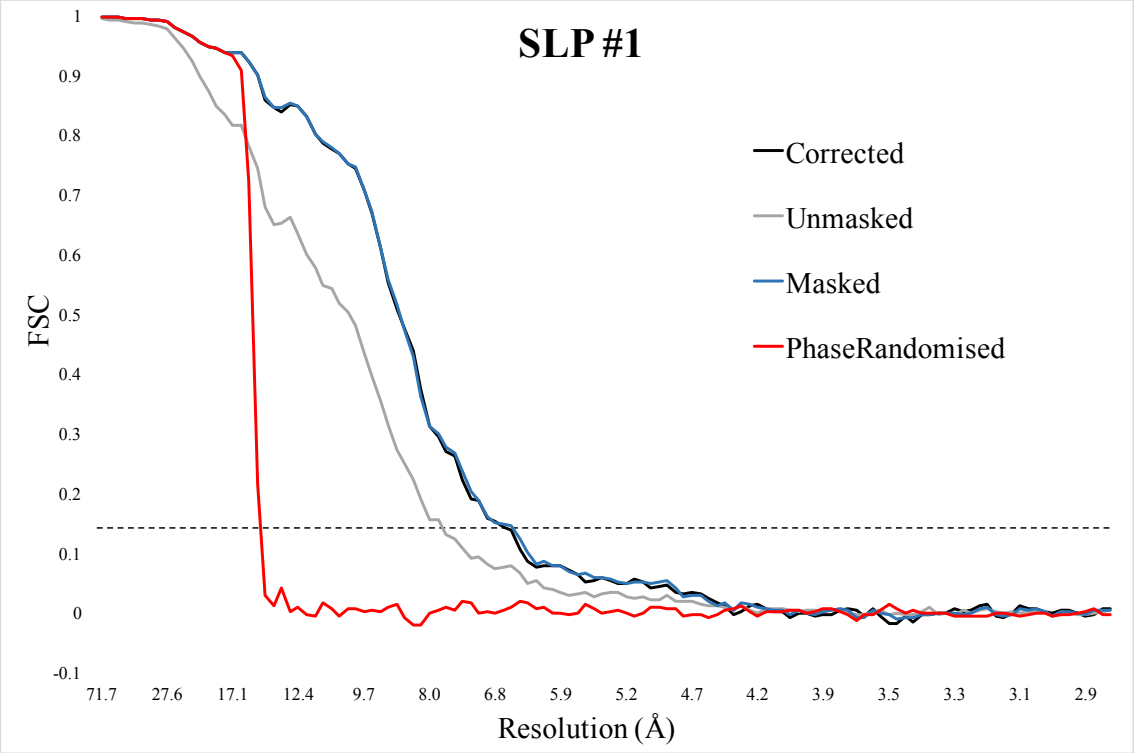

3
Corrected
Unmasked
Masked
PhaseRandomised
FSC
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
999.0
55.1
27.6
18.4
13.8
11.0
9.2
7.9
6.9
6.1
5.5
5.0
4.6
4.2
3.9
3.7
Resolution (Å)

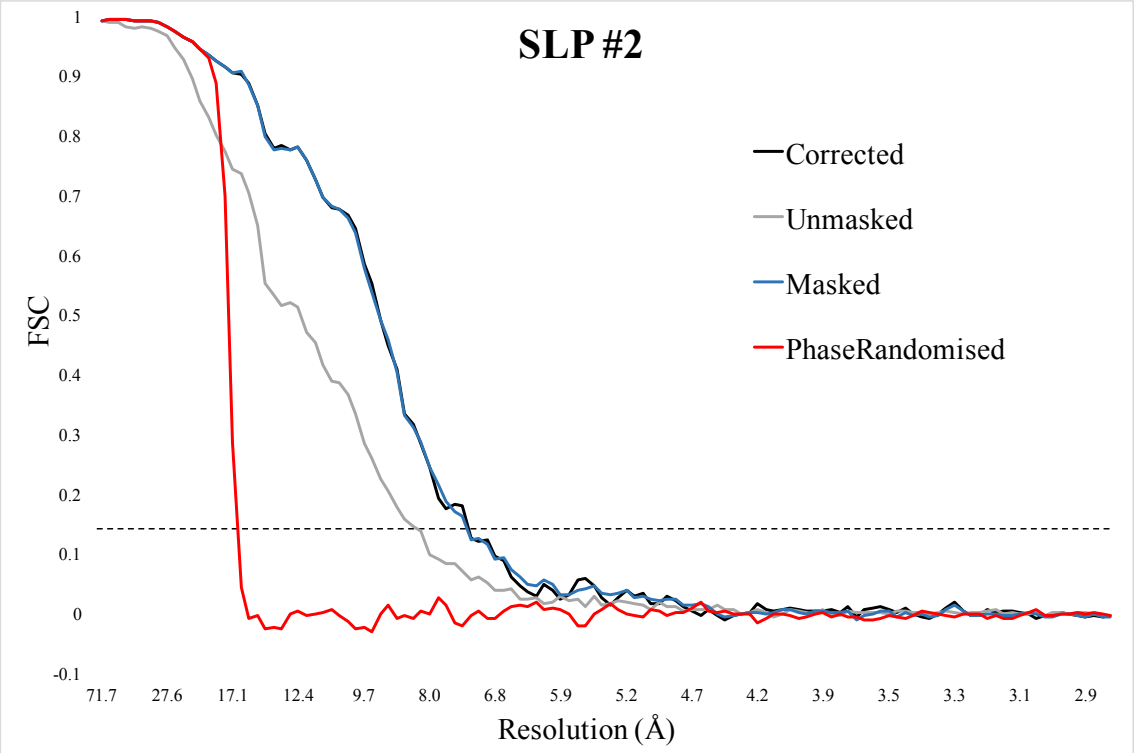
1.8. DLP RdRP arrangements from ϕ6 DTNC



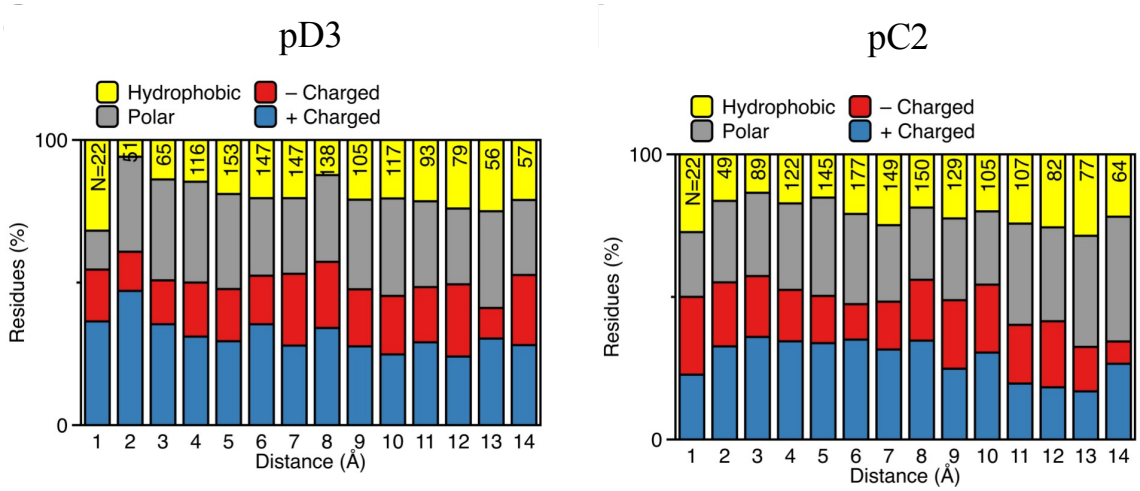


1.9. SLP RdRP arrangements from ϕ6 DTNC





2. ϕ6 P1 residue distribution in dsRNA vicinity



The proportion of P1 residues types (hydrophobic, polar, negatively charged and positively charged) at distances between 1 and 14 Å from the dsRNA in the pD3 and pC2 maps. Adapted from [130].