Mechanisms of Genome Segment Assortment in Reovirus

Jack Peter Kelly Bravo

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

The University of Leeds
School of Molecular & Cellular Biology

September 2019

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

Work in Chapter 4 contains data and figures from a jointly authored publication:

Bravo, J. P. K., Borodavka, A., Barth, A., Calabrese, A. N., Mojzes, P., Cockburn, J. J. B., Lamb, D. C., Tuma, R. (2018). Stability of local secondary structure determines selectivity of viral RNA chaperones. *Nucleic Acids Research*, *46*(15), 7924–7937. https://doi.org/10.1093/nar/gky394

σNS protein samples were prepared by the candidate. NSP2 protein samples were prepared by Dr Alex Borodavka. Single-molecule fluorescence data were collected and processed by Drs Alex Borodavka and Anders Barth. Native mass spectrometry data were collected and processed by Dr Anton Calabrese. All SAXS data, binding data and other biophysical and biochemical analysis in this chapter was performed by the candidate.

In all other chapters, the candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Assertion of moral rights (optional):

The right of Jack Peter Kelly Bravo to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

I would first like to thank my supervisor Dr Roman Tuma for guidance, support and generosity. Aside from your constant enthusiasm and broad scientific (and non-scientific) knowledge, the opportunities to travel and present my work that I have been afforded (together with a steady stream of beer and coffee) have made the last few years both productive and incredibly enjoyable.

I would like to thank Alex for your guidance and friendship. You've been a wonderful travel companion – who'd have thought we'd cover so much ground in this much time! It's been a blast. It's been a lot of fun to explore so many new places, all of which probably ended up with some tipsy (albeit still high-quality) discussions. If only we'd started taking notes sooner…

I would also like to thank Dr Anton Calabrese for the time he spent spraying native MS samples for a biological system that we didn't really understand at the time (although we're getting there now). Thank you to Dr Anders Barth and Kira Bartnik for making sure spFRET data is collected and processed (and done properly!).

I would like to thank the ABSL EM team for technical assistance, and for teaching me the importance of patience when it comes to EM sample prep and data collection. We managed to get there in the end though! I would like to thank my office-mate Dr Matt Watson for maintaining a highly dynamic environment, dynamic discussions and for sharing my dynamic excitement every time I managed to shave 0.1 Å off the resolution of one of my blobs. I would also like to thank Diamond Jim Horne for his role as squash-buddy. I

hope that he manages to stop pressing himself up against walls when he has an audience in the future (although what you do at home is nobody else's business).

I would like to thank Dr Chen Davidovich and his group (Vita, Qi, Michael, Emma, Nick, Sarena, Brady and Sam) for their generosity and hospitality. I would like to personally thank Chen for being so incredibly approachable, and his support and enthusiasm during my time with his group. The time I spent at Monash was fantastic, and I could not have asked for a more welcoming group of people. I would like to give extra thanks to Sarena for ~~putting up with~~ appreciating my high-quality puns and inviting me to high-quality gigs, and to Emma for participating in the most memorable karaoke night I've ever attended. It was a ripper good few months.

It would be amiss not to thank to the four amigos. You're a good lot despite being a bunch of bogans. François, I am grateful for the years of pure, unadulterated domestic bliss that we have maintained. I hope you keep North Grange Brews up and running for years to come, and that I don't have to taste the beer. Hopefully we can convince Dan that being in DTMFs is a noble pursuit and all quit our jobs (jobs?!) to do what we all do best (best?!).

I would finally like to thank my parents and family for their unwavering support. Without your patience and encouragement this never would have been possible. I owe so much to you all, and this thesis is dedicated to you.

Abstract

Members of the Reoviridae family of double-stranded (ds)RNA viruses represent a wide range of important human animal and plant pathogens. During replication, each virus must select and package a complete genome, comprised of up to 12 distinct RNA segments. Stoichiometric genome segment selection is achieved through a series of inter-segment RNA-RNA interactions, mediated by virally-encoded proteins with RNA chaperone-like activities.

Two of the most well-characterised examples of such proteins are rotavirus NSP2 and avian reovirus σNS. Although both proteins bind multiple single-stranded (ss)RNAs concurrently with near-identical, low-nanomolar affinities, possess helix destabilising activity and can promote RNA duplex formation, it has long remained unclear whether NSP2 and σNS employ similar mechanisms to promote RNA-RNA matchmaking between RNA segments.

The work presented in this thesis aims to understand the mechanisms underlying the selective formation of inter-molecular RNA duplexes by the viral RNA chaperone proteins NSP2 and σNS. A comparison of the RNA-binding, helix unwinding and RNA-annealing activities of NSP2 and σNS revealed different modes of RNA chaperone activities. The role of the C-terminal domain of NSP2 in its RNA chaperone activities was subsequently investigated. Single-molecule fluorescence, protein-RNA crosslinking binding kinetics measurements revealed a role for the CTD in RNA dissociation, which was further supported by cryoEM structures of NSP2 alone and in complex with RNA (3.9 Å and 3.2 Å resolution, respectively). Finally, σNS oligomerisation and ribonucleoprotein assembly was explored, revealing that σNS assembles into large, filamentous RNP species. These may serve as a platform for RNA remodelling, as they are sensitive to RNA structure and disassemble upon duplex formation.

Together, these data demonstrate the diversity in mechanisms used to facilitate RNA-RNA interactions employed by viral RNA chaperones, and suggest that NSP2 and σNS utilise different auto-regulatory mechanisms to modulate RNA unwinding and annealing.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
5’RNA triphosphate	RTP
Adenosine triphosphate	ATP
Autocorrelation function	ACF
Avian reovirus	ARV
Base pairs	bp
Bluetongue virus	BTV
C-terminal domain	CTD
C-terminal helix	CTH
Charged coupled device	CCD
Circular dichroism	CD
Collision cross-section	CCS
Cross-correlation	CC
Cross-correlation function	CCF
Cross-linking	XL
Cryogenic electron microscopy	CryoEM
Deoxyribonucleic acid	DNA
Detector Quantum Efficiency	DQE
Dithiothreitol	DTT
Double-layered particle	DLP
Double-stranded	ds
Dynamic light scattering	DLS
E Pluribus Unum	EPU
Electron microscopy	EM
Electrospray ionisation	ESI
Ethylenediaminetetraacetic acid	EDTA
Fluorescence Correlation Spectroscopy	FCS
Fluorescence Cross-Correlation Spectroscopy	FCCS
Fluorescence resonance energy transfer	FRET
Fourier Shell Correlation	FSC
Fourier transform	FT
Host factor Qβ	Hfq
Hours post infection	HPI
Human immunodeficiency virus	HIV
Hydrodynamic radius	R_g
Immobilised metal affinity chromatography	IMAC
Influenza A virus	IAV
Inner core particle	ICP
Ion exchange chromatography	IEX
Ion mobility spectrometry	IMS
Isopropyl β-D-1-thiogalactopyranoside	IPTG
Kilobases	kb
Laplacian-of-Gaussian	LoG
Lysogeny broth	LB
Mammalian reovirus	MRV
Mass spectrometry	MS

Maximum likelihood	ML
Messenger RNA	mRNA
Molecular Weight	MW
Nickel-nitrilotriacetic acid	Ni-NTA
Non-structural Protein 2	NSP2
Normalised Spatial Discrepancy	NSD
Nucleocapsid	NC
Nucleocapsid protein	NP
Nucleotide triphosphate	NTP
Open reading frame	ORF
Plaque-forming unit	pfu
PolyA-binding protein	PABP
Polyacrylamide gel electrophoresis	PAGE
Positive transcription elongation factor b	P-TEFb
Primer-binding site	PBS
Protein data bank	PDB
Radius of gyration	R_h
Ribonucleic acid	RNA
Ribonucleoprotein	RNP
RNA-dependent RNA polymerase	RdRp
Rotavirus	RV
Sigma non-structural protein	σNS
Signal : noise ratio	SNR
Single-pair FRET	spFRET
Single-particle averaging	SPA
Single-stranded	ss
Size-exclusion chromatography	SEC
Small angle scattering biological data bank	SASBDB
Small RNA	sRNA
Sodium dodecyl sulphate	SDS
Standard error	SE
Stem loop	SL
Surface plasmon resonance	SPR
Three-dimensional	3D
Triple-layered particle	TLP
Two-dimensional	2D
Transcription enzyme complex	TEC
Transcription response element	TAR
Untranslated region	UTR
Viral RNA	vRNA
Wild-type	WT

Chapter 1: Introduction

1.1 RNA folding and the requirement for RNA chaperone proteins

Although 75% of the human genome is transcribed into RNA, fewer than 2% codes for proteins (Cruz and Westhof, 2009; Cech and Steitz, 2014). The ability of RNA to perform a diverse range of different roles including transcriptional and translational regulation, catalysis and signalling can be attributed to its intrinsic ability to fold into complex 3D structures (Serganov and Patel, 2012; Jaafar and Kieft, 2019). Such RNA structures can undergo highly specific conformational changes in response to an extensive variety of cellular factors, permitting multifaceted regulation at all levels of cellular function (Mustoe, Brooks and Al-Hashimi, 2014; Jones and Ferré-D'Amaré, 2015; Mustoe *et al.*, 2018).

The assembly of 3D RNA structures is dominated by the folding of highly modular, stable, local secondary structural elements (Leontis, Lescoute and Westhof, 2006; Sim, Minary and Levitt, 2012; Somarowthu *et al.*, 2015). The simplicity and limited chemical complexity of the four natural RNA nucleotides results in a plethora of possible alternative base pairings with similarly high thermodynamic stabilities (Weeks, 1997; Noller, 2005). For example, the free energy of forming a 10-base pair (bp) stem loop at 37°C ranges from -10 to -20 kcal mol^{-1}, whereas a partially mismatched helix of the same length may only be 1 – 2 kcal mol^{-1} less thermodynamically stable (Woodson, Panja and Santiago-Frangos, 2018). As RNAs can adopt a wide range of alternative, non-native structures with similar stabilities during folding, the search for native structure becomes highly frustrated, producing an incredibly rugged free energy diagram (**Figure 1.11**) (Woodson, 2010). The plethora of RNA possible RNA conformations allows RNA to follow multiple alternative folding pathways with different transition energy barriers (**Figure 1.11**).

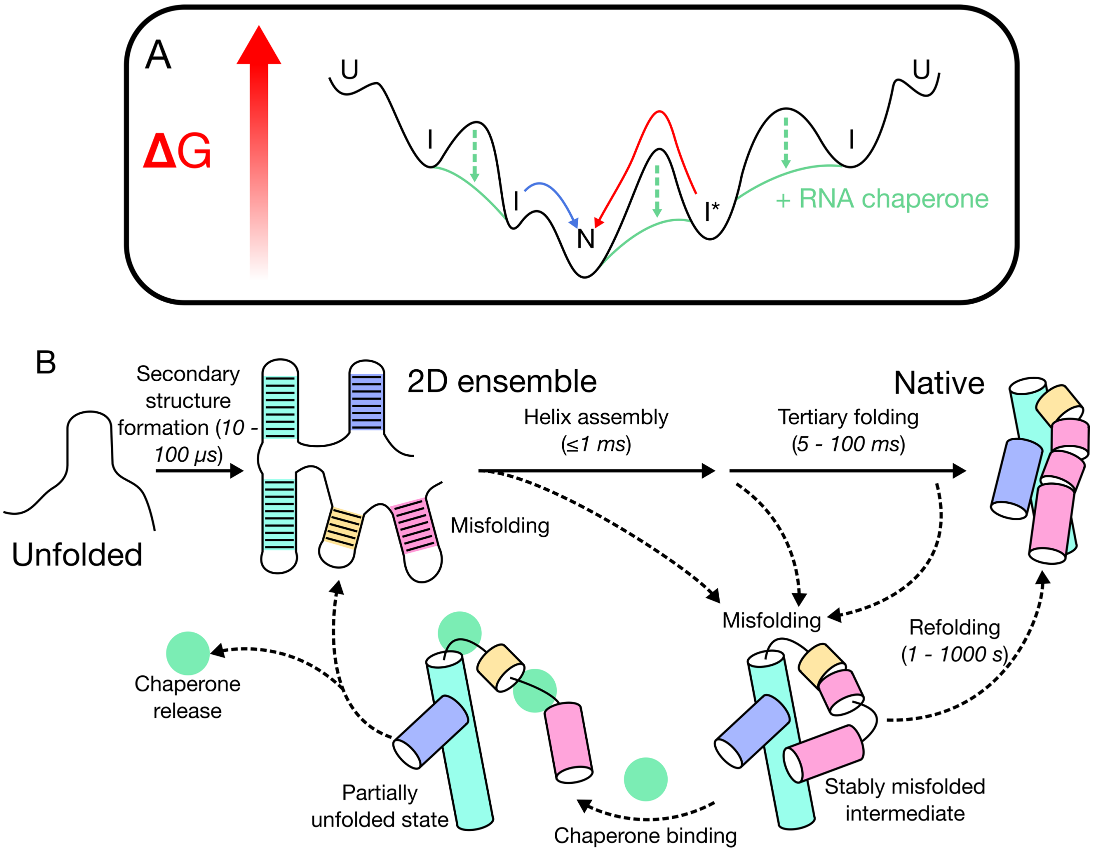


Figure 1.11 General principles of RNA folding. *A – Free energy landscape for RNA folding*. Unfolded (U) RNA molecules are presented with multiple possible folding pathways comprising multiple intermediate (I) structures as they seek out native (N) structures with low free energies. Stably misfolded intermediates (I*) become kinetically trapped as they must cross high energy transition barriers (red arrow), whereas other folding pathways are more rapid (blue arrow). Binding of folding intermediates by RNA chaperone proteins (green) can assist the folding process by reducing energy transition barriers, rescuing misfolded intermediates and promoting native RNA folding. *B – RNA folding pathways involving chaperones*. Kinetics of a RNA folding pathway, illustrating rapid assembly of secondary structural elements followed by a slower search for native tertiary contacts. The reduction of energy transition barriers by RNA chaperones (detailed in *A*) is due to their ability to capture and partially unfold stable non-native intermediates, allowing the RNA to refold correctly. Adapted from (Woodson, Panja and Santiago-Frangos, 2018).

During folding, local RNA stem loops are generally formed within ~10 – 100

μs, driven by stable hydrogen bonding and base stacking (Russell, Millett, *et al.*, 2002; Rangan *et al.*, 2003; Woodson, Panja and Santiago-Frangos, 2018). These modular structures are then able to nucleate the rapid formation of extended helices (at a rate of ~1 μs/bp) (**Figure 1.11**) (Thirumalai and Woodson, 1996; Woodson, Panja and Santiago-Frangos, 2018). Subsequent compaction then allows long-range tertiary contacts to form (Batey, Rambo and Doudna, 1999; Bailor *et al.*, 2011). This process often requires binding of polyvalent cations, notably Mg^{2+}, to neutralise the strong negative charge presented by the phosophodiester backbone and to decrease the entropic penalty of long-range interactions by reducing the disorder of the unfolded (or partially unfolded) conformational ensemble (Draper, 1999, 2004; Draper, Grilley and Soto, 2005; Fiore, Holmstrom and Nesbitt, 2011; Bowman *et al.*, 2012). A compacted state allows conformational sampling and subsequent tertiary interactions to form. Tertiary contact formation is a relatively slow process (≤ 10 ms in the presence of Mg^{2+}) and may require destabilisation and unfolding of pre-formed secondary structures (Wu and Tinoco, 1998; Allewell, 2013) (**Figure 1.11**).

The role of polyvalent cations in RNA structure and folding is not solely limited to overcoming self-repulsive negative charges during compaction, however. Although RNA is believed to be shrouded within an 'ion atmosphere' (consisting of diffusely-bound cations that weakly interact with the RNA), strongly-interacting site-bound ions are often required for RNA to access the native state (Draper, Grilley and Soto, 2005; Trachman and Draper, 2017). Site-bound cations can alter RNA chain flexibility and the local positioning of

RNA backbone phosphate groups, and is likely why Mg^{2+} ions are often found within a close vicinity of RNA catalytic sites (Doudna and Cech, 2002; Suslov *et al.*, 2015; Panja *et al.*, 2017).

Together, the rugged folding energy landscape and hierarchical nature of RNA structure predispose RNA to adopt misfolded states with similar thermodynamic stabilities to the native conformation. Resolution of these misfolded intermediates may require substantial structural rearrangement, resulting in high energy barriers separating different conformational states. Such intermediates can therefore be extremely long-lived and thus represent kinetically trapped states (**Figure 1.11**) (Russell, Zhuang, *et al.*, 2002; Russell, 2008).

The intrinsic tendency of RNA to misfold was first demonstrated for tRNA. During folding, tRNA can adopt two dominant conformations – the native conformation (which is amenable to aminoacylation) and a misfolded conformation (Russell, 2008). The misfolded tRNA conformation contains fewer base pairs than the native state and is less compact (Gartland and Sueoka, 1966). However, it was also demonstrated that *in vitro* RNA misfolding is a reversible process, which can be overcome through heat annealing (i.e. heat-denaturing the misfolded RNA structure), and allowing the RNA to refold in the presence of Mg^{2+} ions (Lindahl, Adams and Fresco, 1966). This method of RNA refolding through heat annealing is obviously not applicable to natively misfolded RNAs *in vivo*, and thus other strategies have been evolved instead.

The existence of proteins with RNA chaperone activity in vivo was first hypothesised by Daniel Hershlag in 1995, although the concept of such a role for proteins had been proposed since the mid-1970s on the basis of in vitro studies (Herschlag, 1995). Such early experiments demonstrated that RNA binding proteins were able to renature non-cognate RNAs (i.e. RNAs and proteins encoded by different species e.g. eukaryotic UP1 protein and *E. coli* 5S rRNA) (Karpel, Miller and Fresco, 1982).

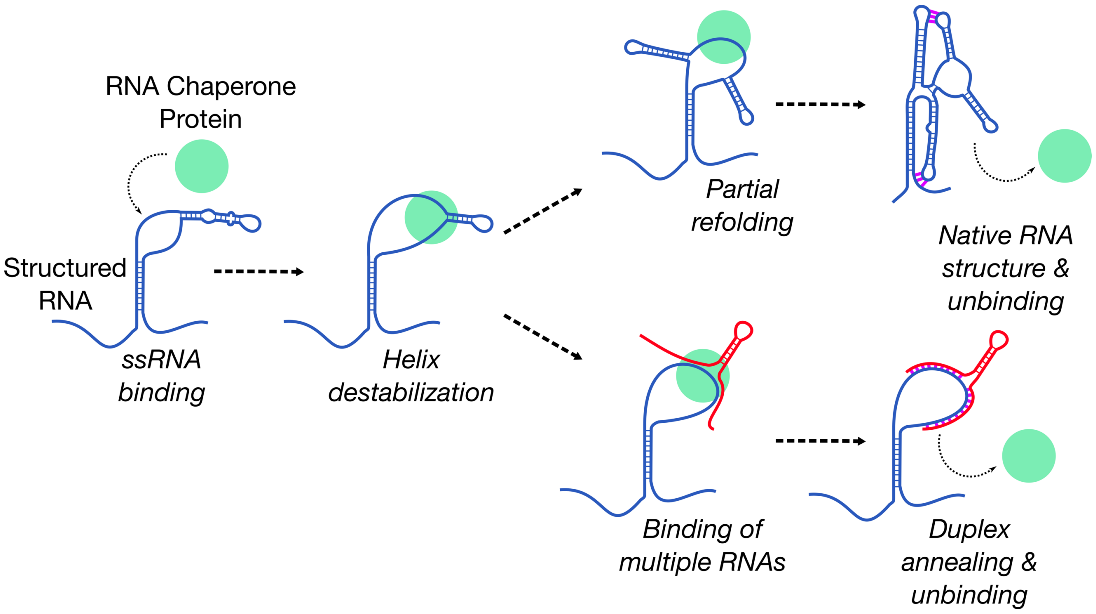


Figure 1.12. RNA chaperone protein-mediated remodelling and duplex formation

A fundamental property of all RNA chaperone proteins is the ability to bind and stabilise single stranded (ss)RNA with higher affinity than they bind double-stranded (ds)RNA. This allows chaperone proteins to bind to folded RNAs via unstructured or metastable (i.e. weakly structured) regions that naturally occur in the vast majority of RNA molecules (Noller, 2005). By binding to ssRNA with high affinity, RNA chaperone proteins are able to effectively compete with RNA structure, ultimately disrupting secondary or tertiary structures and ultimately increasing the likelihood of the RNA reaching its native conformation (**Figures 1.11 & 1.12**).

The ATP-dependent RNA chaperone protein CYT-19 can bind and disrupt both native and misfolded RNA structures (Russell, Jarmoskaite and Lambowitz, 2013). As non-native structures are typically more accessible and less thermodynamically stable than their native counterparts, CYT-19 is able to unfold misfolded conformations preferentially (Tijerina, Bhaskaran and Russell, 2006; Yang *et al.*, 2007; Chen *et al.*, 2008). This is also the case with the ATP-independent RNA chaperone proteins La, StpA and hnRNP A1 (Herschlag *et al.*, 1994; Grossberger *et al.*, 2005; Mayer *et al.*, 2007; Vakiloroayaei *et al.*, 2017). Stably folded, compact, native RNA structures are therefore less favourable targets for RNA chaperone proteins.

By binding and remodelling RNA structures, RNA chaperone proteins modulate the local free energy landscape. The resolution of metastable RNA conformations reduces the free energy barriers between conformational states, allowing the RNA to transiently adopt less stable, kinetically preferred states required for native folding (**Figure 1.11, 1.12**) (Chu and Herschlag, 2008).

Although RNA chaperone proteins are highly heterogenous in terms of overall structure, sequence, function and mechanism, several common features have emerged over the last two decades of their study.

Broadly speaking, RNA chaperone proteins are often highly positively charged, allowing strong interactions with the RNA backbone. However, the preference of these proteins for ssRNA over dsRNA is also indicative of non-electrostatic interactions with base and ribose moieties (Carlomagno, 2014; Bravo *et al.*, 2018). For example, *E. coli* cold-shock protein CspA is able to non-specifically capture unstructured regions of RNA via aromatic side chains within an otherwise positively charged surface. These residues are crucial for RNA chaperone activity and contribute to the destabilisation of the RNA via stacking interactions (Rennella *et al.*, 2017).

Many RNA chaperone proteins bind RNA without sequence specificity (Rajkowitsch and Schroeder, 2007; Woodson, Panja and Santiago-Frangos, 2018). This allows these proteins to interact with and remodel RNAs based upon thermodynamic stability and conformation, rather than through interactions with a specific RNA sequence motif (Herschlag *et al.*, 1994; Schumacher *et al.*, 2006; Doetsch *et al.*, 2010; Jenkins, Malkova and Edwards, 2011; Bravo *et al.*, 2018).

Additionally, RNA chaperone proteins may have the ability to act as matchmakers of intermolecular RNA-RNA interactions (**Figure 1.12**). The action of RNA chaperone proteins results in the exposure of sequences that would be otherwise sequestered by intramolecular base-pairing, allowing them to participate in intermolecular RNA-RNA interactions (Grohman *et al.*, 2013). RNA chaperone proteins can also facilitate duplex formation by holding ssRNAs in unfolded conformations that are amenable for intermolecular base pairing, thereby reducing the energy barrier of the transition from ssRNA and dsRNA (Schumacher *et al.*, 2006).

Annealing can also be achieved by the simultaneous binding of multiple complementary RNAs, drastically increasing their effective local concentration and allowing rapid duplex formation (**Figure 1.12**). Such multivalent RNA binding is a common feature of homo-oligomeric RNA chaperone proteins such as *E. coli* Hfq, *T. brucei* gBP21 and rotavirus NSP2 (Lambert *et al.*, 1999; Schuck *et al.*, 2001; Schumacher *et al.*, 2006; Panja and Woodson, 2012a; Bravo *et al.*, 2018).

One of the most well-studied RNA chaperone proteins is Hfq, a bacterial protein, which will be discussed as a model system in the following section (**Section 1.11**).

1.11 Hfq – a model RNA chaperone protein

Hfq is a highly abundant bacterial RNA binding protein essential for bacterial riboregulation (Vogel and Luisi, 2011; Updegrove, Zhang and Storz, 2016; Gorski, Vogel and Doudna, 2017; Andrade *et al.*, 2018; Kavita, de Mets and Gottesman, 2018). Notably, Hfq is required for small RNA (sRNA)-mediated post-transcriptional gene regulation. In this role, the RNA chaperone activities of Hfq enable it to function as a matchmaker of sRNA-mRNA interactions (Papenfort and Vanderpool, 2015).

The highly conserved ~65-residue Sm-core of Hfq assembles into a homohexameric ring (**Figure 1.13**). The two surfaces (proximal and distal) of the core have distinct RNA-binding sites, separated by an equatorial lateral rim that contains multiple patches of arginine residues (Moller *et al.*, 2002; Brescia *et al.*, 2003) (**Figure 1.13**). The distal face binds to A-rich RNA sequences (specifically, a repeating ARN triplet) that are commonly found with mRNA 5’-UTRs, with each monomer forming contacts with three bases (Link, Valentin-Hansen and Brennan, 2009). The proximal face binds to uridine-rich RNA sequences (such as those within the 3’-end of sRNAs), with each monomer forming contacts with a single RNA base (**Figure 1.13**) (Schumacher *et al.*, 2002; Mikulecky *et al.*, 2004; Dimastrogiovanni *et al.*, 2014; Stanek *et al.*, 2017). Although this confers some degree of sequence-specificity to Hfq-RNA interactions, the ubiquity of such target sequences within bacterial RNAs (together with the high cellular abundance of Hfq (Persson *et al.*, 2013)) results in widespread RNA binding by Hfq in vivo (Holmqvist *et al.*, 2016).

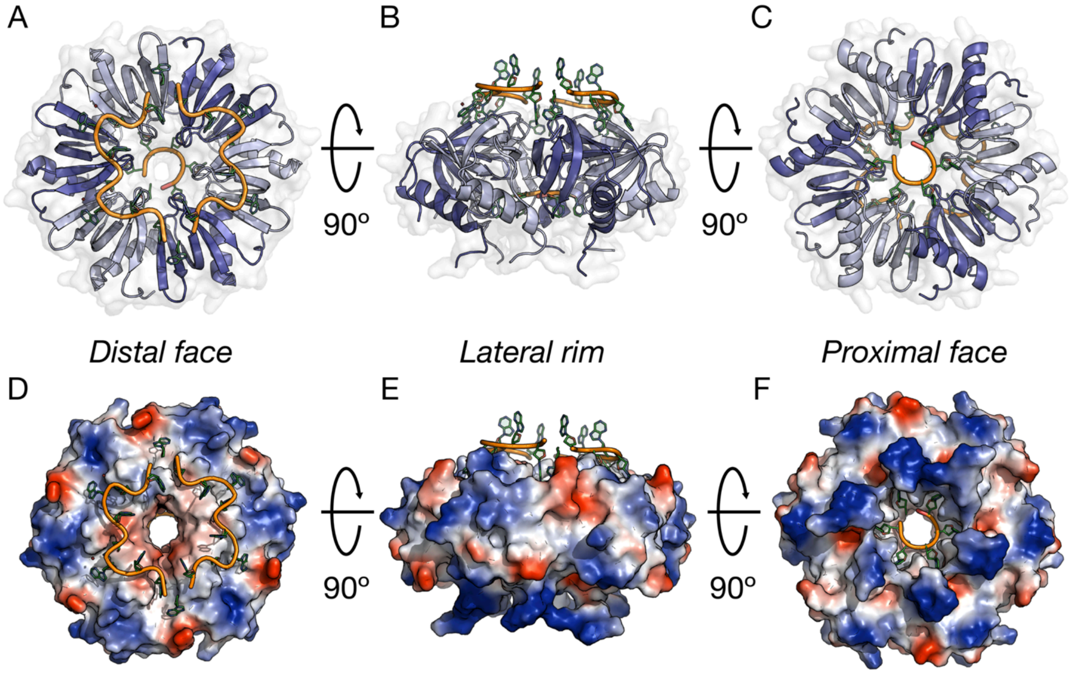


Figure 1.13 Structure and RNA binding sites of Hfq. A chimeric structure of Hfq with multiple RNAs was created by aligning two co-crystal structures of Hfq with RNA bound to either the distal (PDB 4HT8 (Wang *et al.*, 2013) or proximal (PDB 2YLC) faces (Sauer and Weichenrieder, 2011).

Although Hfq-RNA interactions have an electrostatic component (Hopkins *et al.*, 2009), the sequence-specific nature also requires non-electrostatic interactions. Short, unstructured RNAs bind rapidly to Hfq, while structured RNAs bind orders of magnitude more slowly (Hopkins, Panja and Woodson, 2011). The RNA unwinding activity of Hfq, however, is not solely responsible for achieving maximum annealing.

The ability of Hfq to bind multiple RNAs (forming a ternary Hfq-sRNA-mRNA complex) concurrently is essential for RNA-RNA annealing. This allows otherwise transient intermolecular RNA-RNA interactions to be stabilised via interactions with basic patches within the lateral rim of Hfq (Panja and Woodson, 2012b; Panja, Schu and Woodson, 2013). While initial binding of RNAs to Hfq is rapid, nucleation of RNA-RNA interactions is slow (and thus the rate limiting step of Hfq-mediated duplex-formation) (Santiago-Frangos and Woodson, 2018). Hfq rim-RNA interactions stabilise these initial RNA-RNA contacts (which are often fewer than 10 nucleotides long and contain several mismatches), enabling otherwise kinetically and thermodynamically unfavourable RNA duplexes to eventually form stable, extended duplexes. Although initial nucleation is the bottle-neck of this process, formation of an extended duplex is rapid (Hopkins, Panja and Woodson, 2011; Panja *et al.*, 2015).

Hfq dissociation from nascent duplexes is facilitated by natively unstructured C-terminal domains (CTDs) that emanate from the rim of Hfq (Santiago-Frangos *et al.*, 2017; Santiago-Frangos and Woodson, 2018). Unlike the conserved Sm-core (**Figure 1.13**), Hfq CTDs are highly variable between bacterial species, and can vary from fewer than 10 to almost 200 residues (although most commonly ~9 residues) in length. Although the majority of the CTD has little sequence conservation, the tip of the C-terminus is enriched for acidic residues (Santiago-Frangos *et al.*, 2017). Through charge complementation, the acidic tips of the CTD can make contacts with the positively-charged Sm core of Hfq, directly competing for binding with RNA (Santiago-Frangos *et al.*, 2017, 2019). This promotes dsRNA release from Hfq, enabling Hfq to bind novel ssRNA targets and promote duplex formation.

Additionally, Hfq core-CTD interactions play an autoinhibitory role, reducing unstable, transient Hfq-RNA interactions. This is an important aspect of Hfq function, as this ‘stringency filter’ may limit indiscriminate Hfq-DNA interactions in the context of the bacterial cell (Santiago-Frangos *et al.*, 2017).

1.2 Roles of virus-encoded RNA chaperone proteins in the selective packaging of viral genomes

RNA viruses encode all information required for their replication and assembly within their genome both as a linear RNA sequence, and as complex secondary and tertiary structures. Structural elements within viral RNAs (vRNAs) are highly abundant and play regulatory roles in genome replication, translation, packaging in addition to interactions with host-cell factors (Costantino *et al.*, 2008; Kieft, 2008; Hutchinson *et al.*, 2010; Filbin *et al.*, 2012; MacFadden *et al.*, 2018; Steckelberg *et al.*, 2018).

The multifunctionality of vRNAs is also somewhat problematic; vRNAs often serve as both mRNA transcripts that must be actively translated, and as genetic material that must be selectively packaged into progeny viruses. While active translation destabilises mRNA structure in cells (Spitale *et al.*, 2015; Mustoe *et al.*, 2018), many vRNAs are highly structured and compact due to the structural requirements and spatial restraints imposed by packaging (i.e. it is easier to package a highly compact RNA molecule within a capsid than an extended, unstructured RNA molecule) (Gopal *et al.*, 2014; Ben-Shaul and Gelbart, 2015).

In addition to forming highly compact structures suitable for packaging, vRNA often utilise secondary and tertiary structures for a range of functions (Jaafar and Kieft, 2019). For many viruses including Influenza A virus (IAV), dengue virus and rotavirus (RV), cis-acting, long-range intramolecular RNA-RNA interactions between 5’ and 3’ sequences (termed ‘cyclization’) are essential for efficient translation (Tortorici, Shapiro and Patton, 2006; McDonald and

Patton, 2011; Giese, Bolte and Schwemmle, 2016; Boerneke, Ehrhardt and Weeks, 2019). The presence of mutually-exclusive structures adopted by vRNAs therefore necessitates the activity of virally-encoded proteins with RNA chaperone-like activities (Zúñiga *et al.*, 2009; Yang *et al.*, 2014; Xia *et al.*, 2015).

Furthermore, for several viruses, assembly requires the selective packaging of multiple RNA molecules (Giese, Bolte and Schwemmle, 2016; Borodavka, Desselberger and Patton, 2018; Rein, 2019). Although this selectivity is thought to be achieved through a series of inter-segment RNA-RNA interactions, virally-encoded RNA chaperone proteins play a crucial role in this process. This section describes the roles of viral RNA chaperone-like proteins that mediate selective packaging of HIV-1, IAV and RV.

1.2.1 HIV-1 genome packaging and assembly

HIV-1 is a retrovirus, and is the causative agent of AIDS (Diskin *et al.*, 2013). Although there are effective treatments to hinder viral replication, there is no cure meaning that long-term treatment is required. Thus, the search for novel targets for antiviral therapy is essential (Freed, 2015).

The viral particle contains the diploid viral genomic RNA (vRNA), ~2000 copies of virally-encoded Gag protein, surrounded by an envelope containing viral glycoproteins (Briggs *et al.*, 2003). After the release of viral particles from the host cell Gag undergoes autoproteolysis (into multiple stable fragments), resulting in particle re-organisation and maturation. This is essential for infectivity (Schur *et al.*, 2016).

As HIV-1 particle assembly occurs at the plasma membrane, the vRNA must compete for packaging with highly abundant cellular mRNAs (Eckwahl *et al.*, 2016). Indeed, previous studies have demonstrated that in the absence of vRNA, host cell mRNAs are packaged stochastically into viral particles (Eckwahl *et al.*, 2015, 2016). To overcome this, the virus must employ a mechanism of selective genome packaging.

Previous studies have demonstrated that vRNA is packaged as a homodimer (Nikolaitchik *et al.*, 2013; Rein, 2019). This provides several advantages in terms of viral fitness. Firstly, it allows the virus to compensate for detrimental mutations that may occur within one of the copies of the vRNA. Additionally, as cDNA is reverse-transcribed within the viral particle, the virally-encoded reverse transcriptase (RT) is able to 'jump' between the two copies of the vRNA (Larsen *et al.*, 2018). High-frequency recombination provides a source for rapid genome evolution. Genome dimerisation occurs through a series of intermolecular RNA-RNA interactions involving the 5' untranslated regions of the vRNA (**Figure 1.21**) (Keane *et al.*, 2016).

5'UTR contains several distinct RNA structural elements with various roles in virus replication (**Figure 1.21**). TAR (transcription response element) stimulates viral transcription through binding to the viral Tat protein and the host-derived positive transcription elongation factor b (P-TEFb) (Schulze-Gahmen *et al.*, 2016). PBS (primer-binding site) is the sequence at which a host-cell tRNA^{Lys} is annealed, serving as a primer for reverse transcription

(Larsen *et al.*, 2018). Downstream are a series of four stem loop (SL) structures (SL1 – 4). SL1 contains the dimerisation-initiation site (DIS), SL2 contains the major splice donor (SD), SL3 contains Ψ (Packaging signal, Psi), and SL4 contains the translation start codon (AUG) (Johnson and Telesnitsky, 2010; Bieniasz and Telesnitsky, 2018) (**Figure 1.21A**).

In one conformation, the dimerisation-initiation site is sequestered by base pairing to the U5 region (**Figure 1.21B**). This precludes DIS from participating in intermolecular RNA-RNA interactions required for genome dimerisation, thereby impeding virus assembly (Brigham *et al.*, 2019). In this conformation, AUG is exposed, allowing the vRNA to be actively translated. Alternatively, U5 can base pair with AUG, preventing translation (Keane *et al.*, 2015, 2016). In this conformation, DIS is exposed in a dimerisation-competent conformation. Interestingly, the packaging-competent conformation involves secondary structure elements involving SD. This may ensure that the full-length vRNA is packaged, rather than splice variants (which lack SD). This complex conformational switch allows the vRNA to perform multiple roles during viral replication. The free energy required to trigger conformational switching between these two structures, however, must arise from somewhere.

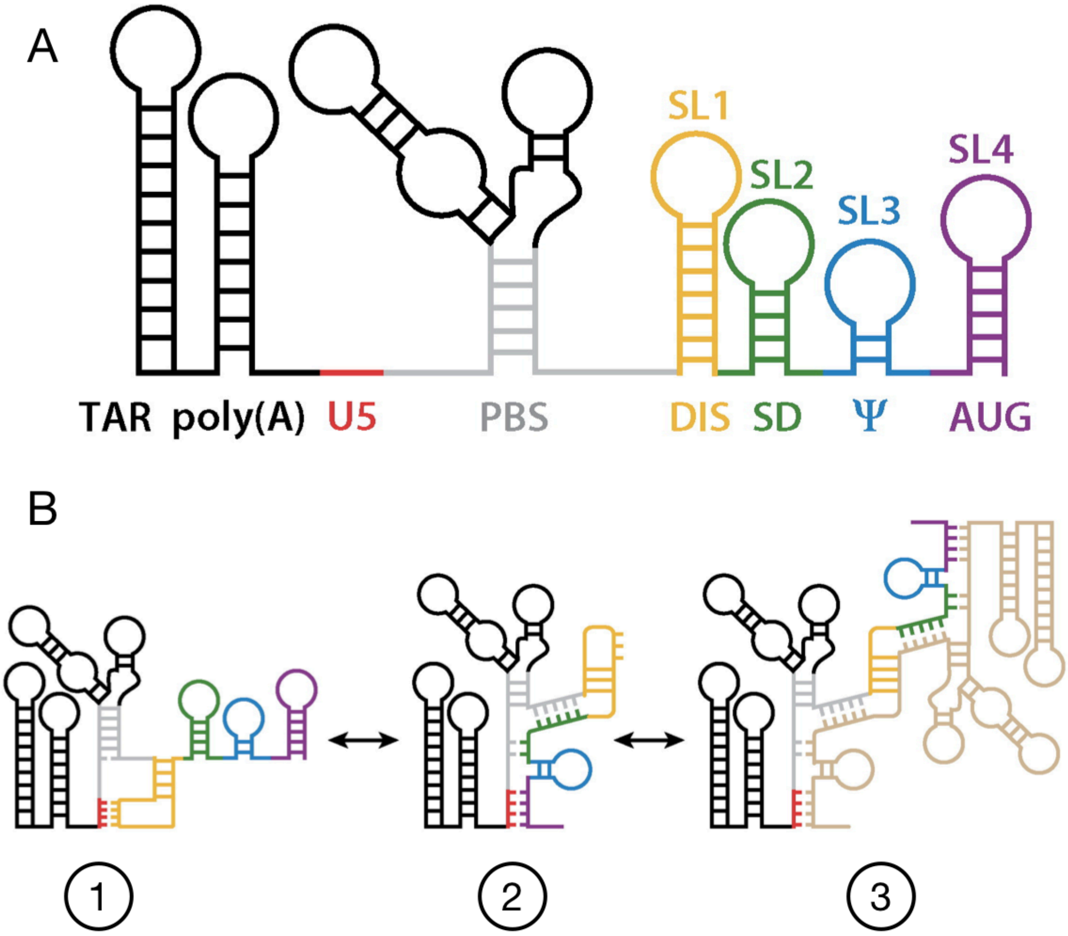


Figure 1.21 Secondary structure re-arrangements within HIV-1 5'UTR. A: Anatomy of 5'UTR. This region contains multiple structural elements including TAR (transcription activation element), PBS (primer-binding site), DIS (dimerisation-initiation site), SD (splice donor), Ψ (packaging signal), and AUG (translation start codon). B: Switchable RNA conformations allow the vRNA to transition from 1: a translation-competent structure (wherein DIS is sequestered and AUG is exposed) to 2: a packaging-competent structure (wherein AUG is sequestered and DIS is exposed). DIS can then participate in RNA-RNA interactions resulting in 3: genome dimerisation and extensive intermolecular base-pairing. Adapted from (Bieniasz and Telesnitsky, 2018)

This region adopts multiple complex, 3D structures that determine vRNA fate.

HIV-1 Gag polyprotein contains a short (~55 residue), highly basic (pI 9.93) nucleocapsid (NC) domain to facilitate this vRNA conformational switch. NC contains two zinc-finger motifs flanked by unstructured, positively-charged extensions (**Figure 1.22**) (Herschlag *et al.*, 1994; Rein, 2010; McCauley *et al.*, 2015). These structural features allow NC to interact with RNA both through non-specific, electrostatic interactions (via positive extensions) and via hydrophobic, sequence-specific interactions (via zinc-fingers). NC preferentially interacts with exposed, unpaired guanosine nucleotides (Grohman *et al.*, 2013), enabling NC to act as an RNA chaperone and destabilise RNA structure. Point mutations that disrupt the zinc fingers abrogate specific genome packaging, and RNA G→U mutants are less efficiently remodelled, reducing the yield of vRNA dimers (Grohman *et al.*, 2013; Rein, 2019). Additionally, the positively charged extensions are thought to reduce charge repulsion between vRNAs, thereby increasing dimerisation efficiency (Rein, 2010, 2019; Holmstrom *et al.*, 2019). These general RNA chaperone-like activities serve to reduce the free-energy barrier between the two alternative vRNA conformations.

Once the palindromic DIS is exposed, it can participate in intermolecular RNA-RNA interactions with DIS from another vRNA. DIS-DIS kissing-loop interactions initiate further vRNA structural re-arrangements, forming extensive intermolecular interactions within the extended dimer (**Figure 1.21B**). This conformation of vRNA exposes Ψ to Gag, nucleating virus particle assembly (Lu, Heng and Summers, 2011).

The basis for specific packaging is still somewhat unclear, although interactions between Gag and Ψ within vRNA dimers are essential. In experiments where two copies of Ψ have been placed in tandem within the same RNA molecule, monomeric vRNA is packaged (Sakuragi, Shioda and Panganiban, 2002).

Gag binds non-specifically to ssRNAs with high (25 – 50 nM) affinity, predominantly through electrostatic interactions under physiological ionic conditions. The selective binding of Ψ, however, is believed to occur through non-electrostatic interactions. Under high ionic strength conditions, Gag-vRNA interactions are maintained, while interactions between Gag and RNAs that do not contain Ψ are lost (Comas-Garcia *et al.*, 2017). Thus, although the source of packaging specificity may not be thermodynamic, the specific Gag-Ψ interactions may provide a kinetic advantage to virus assembly. Binding of multiple copies of Gag to the vRNA dimer may nucleate particle assembly, overcoming a kinetic barrier (D'Souza and Summers, 2005). As the initial particle nucleation is thought to be a rate-limiting step to virus assembly, this gives an advantage to Ψ-containing vRNA over other RNAs and may provide a mechanism for specific packaging of a dimeric viral genome (Rein, 2019).

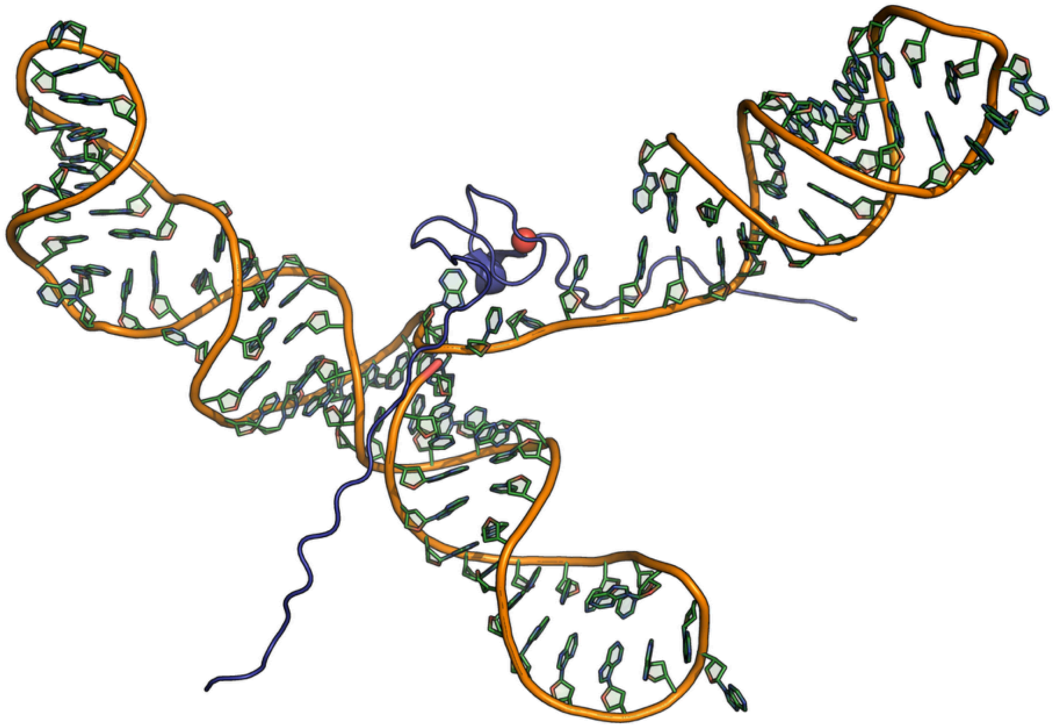


Figure 1.22 Structure of Murine Leukaemia virus (MLV) NC bound to Ψ. MLV is a prototypical retrovirus and is often used as a proxy for HIV-1 in experimental studies. NC (purple) contains unstructured extensions away from a central zinc-finger domain (zinc ion is red). Although HIV-1 NC contains two zinc-fingers, both NC variants are thought to interact with and remodel RNA analogously (D'Souza and Summers, 2004).

1.2.2 Influenza A virus (IAV) genome packaging and assembly

IAVs are the causative agents of influenza epidemics, as well as severe global pandemics (Giese, Bolte and Schwemmle, 2016). The IAV genome is distributed between eight single-stranded negative sense (-)RNA segments ranging from 0.9 – 2.4 kb, all of which must be selected and incorporated into progeny virions in order to produce infectious viral particles (Gerber *et al.*, 2014). The multi-segmented genome provides an evolutionary advantage to IAV through the generation of re-assortment progeny during the co-infection of multiple IAV strains (Gao and Palese, 2009; Lowen, 2017). This was the case during the 2009 pandemic, where the H1N1 strain was a reassortant of bird, swine and human influenza viruses (Gerber *et al.*, 2014).

Each viral RNA segment (vRNA) consists of an open reading frame (ORF), flanked by a 5’ and 3’ UTR. These UTRs are highly conserved and self-complementary and form a panhandle structure, which is captured by the multisubunit viral RNA-dependent RNA polymerase (RdRP) (consisting of the catalytical subunit PB1, and the accessory subunits PB2 and PA) (Kouba, Drncová and Cusack, 2019). The remainder of the vRNA is coated with nucleocapsid protein (NP) – a ssRNA-binding protein – forming the double-helical viral ribonucleoprotein complex (vRNP) (**Figure 1.23**).

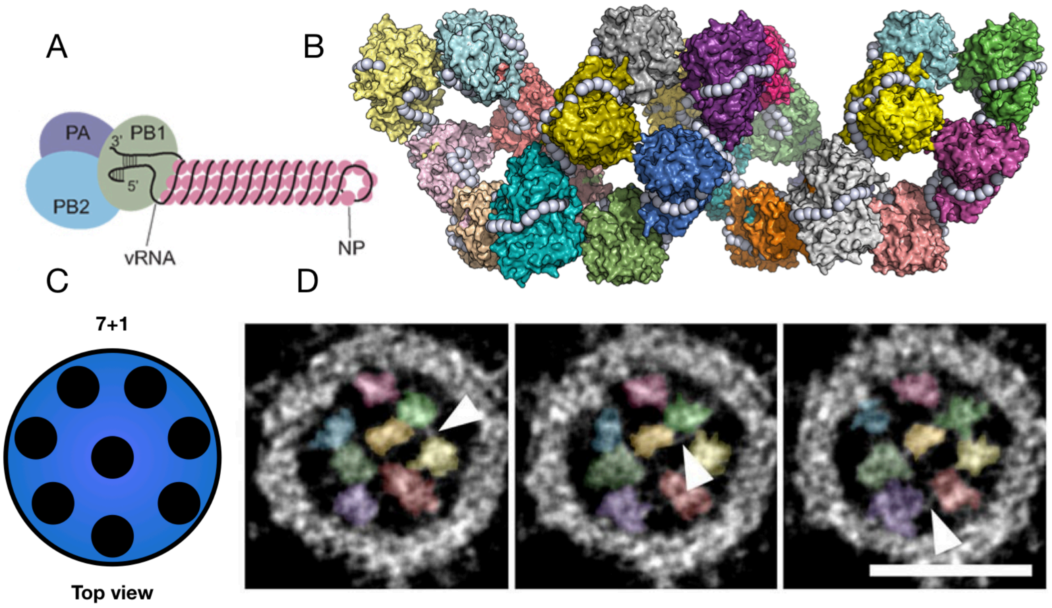


Figure 1.23 Architecture of IAV RNP. A: Schematic of IAV RNP. RdRP (PA, PB1 and PB2) bind the panhandle structure formed by 5' and 3' of viral RNA segment (vRNA). The remainder of the vRNA is bound by nucleocapsid protein (NP). Adapted from (Newburn and White, 2019). B: Atomic model of helical part of IAV RNP (PDB 4BBL) (Moeller *et al.*, 2012). NP monomers are shown as muticoloured surfaces while vRNA is shown as white spheres. C: Schematic of top-down view of IAV virion, showing a 7+1 arrangement. D: Slices through cryoET reconstruction of IAV virion. vRNPs are highlighted in different colours. Arrows indicate putative RNA-RNA interactions. Adapted from (Noda *et al.*, 2012)

Within the virion, the genome segments are arranged in a "7+1" pattern, wherein a single, central segment is flanked by the other seven (**Figure 1.23**) (Noda *et al.*, 2006). This genomic organisation has been implicated to play a role in RNA selection, although the process through which this is achieved is poorly understood (Noda *et al.*, 2018). The majority of viral particles contain a single copy of each RNA segment, supporting the notion of selective, stoichiometric genome packaging (Chou *et al.*, 2012).

NP binds ssRNA with high affinity, (~20 nM), without sequence specificity, with a monomer RNA binding footprint of ~12-nt (Baudin *et al.*, 1994). NP interacts with the RNA phosphodiester backbone through predominantly electrostatic interactions, resulting in exposure of nucleobases and the disruption of RNA secondary structures (Baudin *et al.*, 1994). Although these RNA chaperone-like activities are rarely referred to as such, they play a crucial role in the regulation of viral transcription, replication and genome segment assembly (Moeller *et al.*, 2012; Moreira *et al.*, 2016). The structure of vRNA within the vRNP is critical for selective genome packaging (Dadonaite *et al.*, 2019).

Although the binding of NP to vRNA results in the disruption of RNA structures, highly-stable structures are able to persist (York *et al.*, 2013; Dadonaite *et al.*, 2019), resulting in the non-uniform distribution of NP long the vRNA. This permits exposed sequences to participate in intersegment RNA-RNA interactions, providing a mechanism for selective segment packaging. Within EM-derived structural models of the vRNP, there are numerous regions of solvent-exposed vRNA that may participate in such interactions (**Figure 1.23B**) (Moeller *et al.*, 2012).

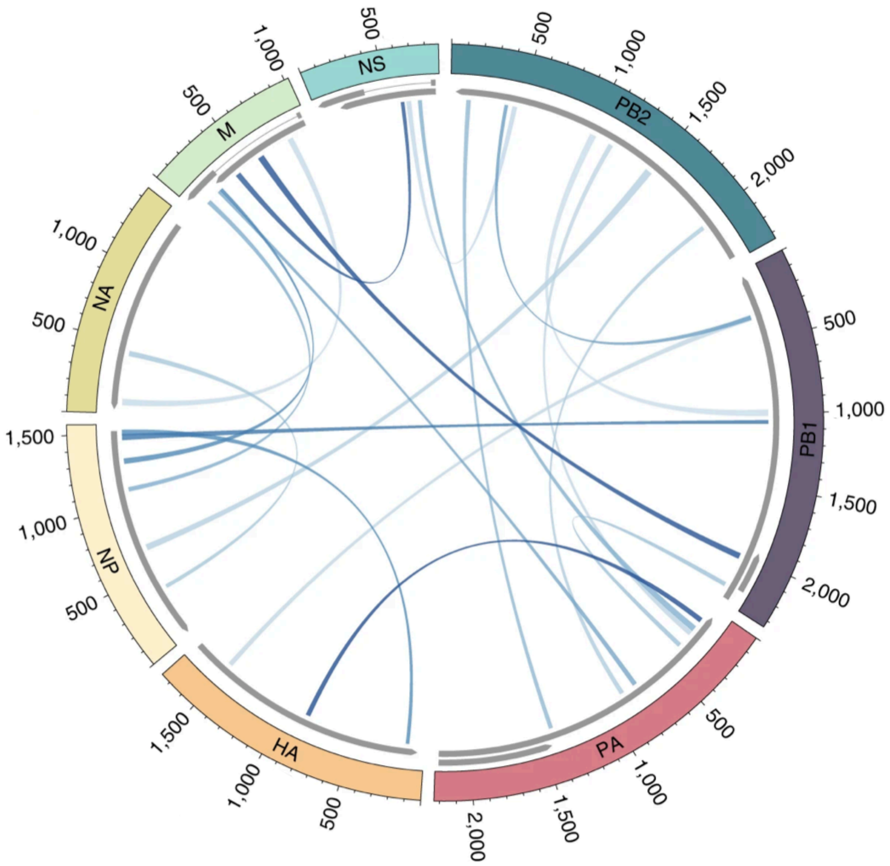


Figure 1.24 Map of IAV intersegment RNA-RNA interactions. RNA-RNA interactions between viral RNA segments for the IAV H1N1 strain, as mapped by SPLASH (Aw *et al.*, 2016). Eight segments are shown aligned within a circle, with inter-segment interactions represented by curved lines. Adapted from (Dadonaite *et al.*, 2019).

There is now significant evidence for direct, intermolecular RNA-RNA interactions between viral RNA segments (Essere *et al.*, 2013; Gavazzi *et al.*, 2013). Within tomographic slices of in-tact virions, bridging densities between vRNPs that may represent such RNA-RNA interactions are clearly visible (**Figure 1.23D**) (Noda *et al.*, 2012). Furthermore, RNA-RNA crosslinking within intact virions has revealed numerous direct RNA-RNA interactions between genome segments (**Figure 1.24**). Interestingly, some sequences are capable of interacting with multiple RNA segments, indicating mutual exclusivity (**Figure 1.24**). This suggests that the intersegment RNA-RNA interactions may be somewhat redundant, and that IAV has a degree of flexibility in terms of genome structure. This flexibility may accommodate for reassortment during co-infection, while still ensuring selective genome packaging (Newburn and White, 2019). This model suggests that there is no conserved hierarchy of genome segment selection, and there is fluidity in terms of the order in which genome segments are selected.

1.3 *Reoviridae* structure, replication and assembly

1.3.1 Rotavirus replication

Rotavirus (RV) currently has the most well-characterised replication cycle of any *Reoviridae* family member. Similar to mammalian orthoreovirus (MRV) and bluetongue virus (BTV), RV has a multi-layered capsid, although the RV capsid consists of three layers as opposed to the double-layered particles of MRV and BTV (Borodavka, Desselberger and Patton, 2018). The inner core particle of all reoviruses (including MRV, BTV and RV) is transcriptionally active and consists of the complete viral genome (9 – 12 segments of genomic RNA, varying between viral species), a layer of coat protein and the viral transcription enzyme complex (TEC). The outer layers of such viral particles are acquired later during virus assembly (Trask, McDonald and Patton, 2012).

Rotavirus genome consists of dsRNA segments (**Figure 1.31**) (McDonald, Martha I. Nelson, *et al.*, 2016). The genomic RNA segments are thought to be bound at the 5’ and 3’ ends by the viral RNA-dependent RNA polymerase (RdRP), VP1. VP1 is in complex with VP3, the polymerase co-factor and capping enzyme, responsible for formation of the 5’ cap on nascent viral mRNAs, increasing their cellular half-life through ribonuclease protection (**Figure 1.31**) (Tortorici, Shapiro and Patton, 2006).

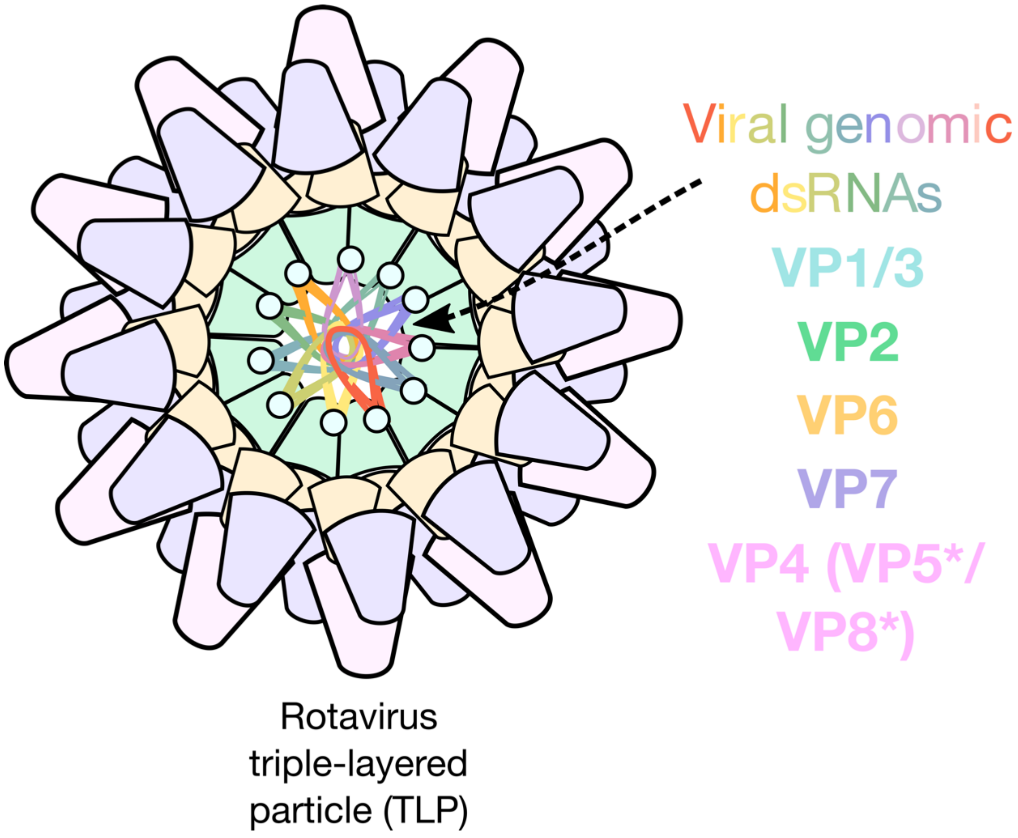


Figure 1.31 Architecture of a rotavirus particle Rotavirus (RV) triple-layered particle (TLP) is composed of multiple concentric protein layers surrounding a complete set of 11 genome segments, each associated with its own transcription enzyme complex (TEC) formed by VP1 and VP3 (cyan). TEC-genome is encapsidated with a T=1 symmetry icosahedral VP2 (green), forming the transcriptionally active inner core particle (ICP). VP6 (orange) assembles a T=13 layer around the ICP, forming the double-layered particle (DLP). Egress through the endoplasmic reticulum results in accumulation of VP7 (purple) and VP4 (magenta), completing TLP assembly. Once released from the cell, VP4 undergoes proteolytic cleavage of into VP5* and VP8* resulting in a mature particle.

RNA-bound VP1/VP3 recruits VP2, nucleating assembly of a VP2 inner capsid. The inner capsid shell (termed ‘core’) is capable of in vitro RdRP activity, with VP2 being essential for this. VP1/VP3 binds to VP2 tetramers at specific, off-centre locations that vary at the equivalent 5-fold symmetry vertices (Trask, McDonald and Patton, 2012; Ding *et al.*, 2019). The precise role of VP2 in RdRP activation remains poorly understood, although it may act as a structural scaffold for VP1/VP3, effectively positioning VP1 in a conformation that is conducive to transcription. Recent structures have demonstrated that VP2 anchors VP1 via an N-terminal amphipathic helix. This prevents ‘unfavourable interactions’ with genomic RNA, effectively promoting RdRP activity (Lu *et al.*, 2008; Estrozi *et al.*, 2013; Ding *et al.*, 2019; Jenni *et al.*, 2019).

The inner core is surrounded by VP6, forming the double-layered particle (DLP). The VP6 layer serves as an adaptor for the outer layer of the viral particle which is composed of the VP7 glycoprotein (**Figure 1.31**). VP7 is stabilised by calcium ions, and attaches to a small protrusion of VP6 via ‘grip arm’ mechanism (Trask, McDonald and Patton, 2012). VP7 is perforated, allowing VP4 attachment protein to protrude from the underlying VP6 layer (Trask, McDonald and Patton, 2012). VP4 undergoes posttranslational cleavage by intestinal trypsin-like proteases after cellular egress, resulting in the two non-covalently attached products termed VP5* and VP8*. VP5* forms a trimeric base, while two VP8* molecules sit at the tip of the spike and bind to sialic acid moieties (in addition to other unidentified host cell surface receptors), thereby mediating host cell attachment. VP8* detaches upon interaction with the host cell, allowing VP5* to undergo a significant conformational change, exposing hydrophobic loops that destabilise the cell membrane, facilitating endosomal escape (Trask, McDonald and Patton, 2012).

Within the early endosome, low calcium concentrations destabilise VP7 and trigger TLP disassembly. Endosomal escape is also thought to be mediated by VP5*, although it is unknown whether this is accomplished while VP5* is still associated with the viral particle, and the mechanism of VP5*-mediated membrane penetration is also poorly understood (Guo and Bartel, 2016).

Once the now-double-layered particle (DLP) enters the cytoplasm, it is transcriptionally active, continuously synthesizing and extruding nascent positive sense, single-stranded (+)RNAs (Periz *et al.*, 2013). These +RNAs are multifunctional, serving as both mRNAs encoding viral proteins and as templates for dsRNA synthesis. The +RNAs are transcribed by the encapsidated TEC, using the packaged negative-sense (-)RNAs as templates. The packaged -RNA, TEC and transient nascent dsRNA (prior to extrusion of +RNAs from the capsid) are thought to remain within the viral particle at all times, possibly as a mechanism of limiting host cell immune response by sequestering viral dsRNA (Periz *et al.*, 2013).

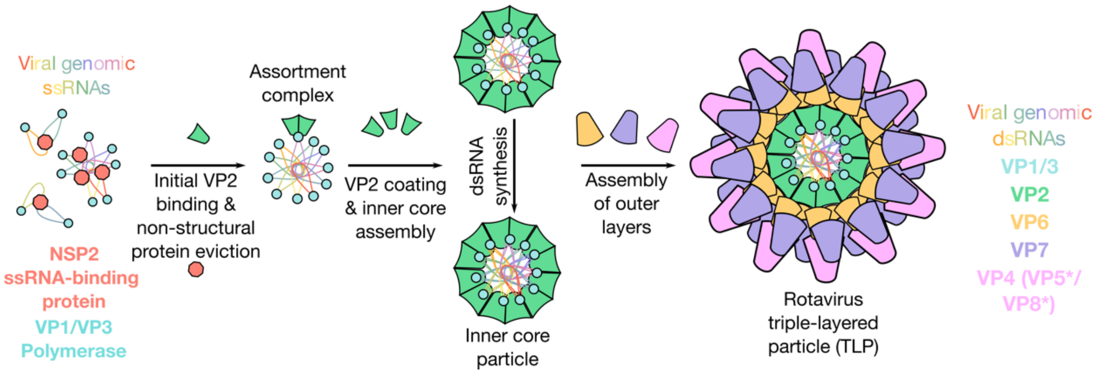


Figure 1.32 Rotavirus assembly pathway RV assembly is initiated through the assembly of an assortment complex containing a single copy of each genomic RNA segment. Selection and assortment of a complete genome occurs through a series of intersegment RNA-RNA interactions, mediated by ssRNA-binding protein NSP2. The inner core subsequently assembles around the assortment complex through VP1-VP2 interactions, forming an actively transcribing inner-core particle. The two outer layers (VP6 and VP4/VP7, respectively) can subsequently assemble around the inner core, producing a TLP.

VP1 is a hollow enzyme, with multiple “tunnels” that allow for entry of NTPs and template -RNA, and the exit of the template and the nascent +RNA, which becomes capped by VP3 during export from the viral particle. RV genome replication is fully conservative (Lu *et al.*, 2008).

Within two to three hours post-infection (HPI), large cytoplasmic inclusions termed viral factories or viroplasms are assembled from the non-structural proteins NSP2 and NSP5. These inclusions serve as the sites of viral replication and assembly, and are highly enriched for both viral proteins (including VP1, 2, 3 & 6) and viral RNA (Fabbretti *et al.*, 1999; Viskovska *et al.*, 2014; Criglar *et al.*, 2018).

NSP5 is a serine and threonine-rich protein that undergoes a hyperphosphorylation cascade, triggered by Casein Kinase 1α which is thought to be essential for viroplasm assembly (Eichwald *et al.*, 2004; Criglar *et al.*, 2014, 2018). NSP5 acts a scaffold of viroplasm assembly, recruiting structural proteins VP1, VP2 and VP6 (Vascotto *et al.*, 2004) in addition to its interaction with NSP2.

NSP2 plays several overlapping roles during RV morphogenesis. It has nucleotide triphosphate activity that stimulates NSP5 phosphorylation, making it essential for viroplasm assembly (Vasquez-Del Carpio *et al.*, 2006a). The role of NSP2 in genome segment assortment is further discussed later (**Section 1.33**).

The fate of viral RNA (whether it is destined for translation or packaging) may be determined by its subcellular localisation. Viral RNAs are recruited to the viroplasm through the specific recognition of a 3'UGUGACC sequence by VP1 and are packaged into nascent core particles (Tortorici *et al.*, 2003; Trask, McDonald and Patton, 2012). These nascent particles are transcriptionally active, producing vast amounts of +RNAs required for subsequent virus assembly (Periz *et al.*, 2013).

The cellular translational machinery, however, does not localise to the viroplasm, suggesting that translation of viral proteins occurs elsewhere within the cell. RV NSP3 is a non-structural protein whose N-terminus binds to the 3'-end of +RNAs and C-terminus binds eIF4G. eIF4G binds to the 5'-RNA cap structure, circularising viral RNAs and promoting translation efficiency (Deo *et al.*, 2002; Groft and Burley, 2002). As viral +RNAs are not polyadenylated, NSP3 is a functional mimic of host cell polyA binding protein (PABP), utilizing a similar mechanism of eIF4G recognition (via hydrophobic interactions). The hijacking of eIF4G inhibits cellular mRNA translation and results in preferential translation of RV gene products. This may provide an elegant solution to whether a given +RNA is destined to be translated or packaged. By spatially separating these two essential aspects of virus replication, viral translation and virus assembly become two distinct and efficient processes.

Once the DLP has assembled, it must exit the viroplasm and enter the endoplasmic reticulum (ER) in order to acquire the outer layer proteins VP4 and VP7. Although this process is poorly understood, it is thought that the non-structural membrane protein NSP4 accumulates within the ER membrane and can bind both VP6 (i.e. the DLP) and VP4, potentially chaperoning their interactions and TLP assembly. In vitro assembly assays indicate that VP4 must assemble around the DLP prior to VP7 acquisition and TLP assembly. The mechanism of TLP egress from the ER is poorly understood, although rotavirus virions are released from cells by direct lysis (Musalem and Espejo,

1985) or an unknown secretion pathway (distinct from the Golgi apparatus and lysosomes) (Jourdan *et al.*, 1997).

1.32 Genome segment assortment in rotavirus

Rotavirus genomes are distributed between 11 distinct genomic RNA segments, all of which are essential for virus assembly and replication. The probability of packaging a complete genome consisting of a single copy of each segment at random is very low ((11!)/(11^11), or 0.014%). In spite of this, the efficiency of infective virus production (as measured by particle : plaque-forming unit, pfu) is 100 : 1, indicating that at least 1% of the viruses produced are infectious (Trask, McDonald and Patton, 2012; McDonald, Martha I Nelson, *et al.*, 2016). The amount of viral particles that contain a complete genome is likely to be even higher than this, as infectivity requires several further assembly and post-translational cleavage steps which may act as further levels of quality control. The observation that RV infection can proceed upon the uptake of a single transcriptionally active viral particle suggests that the efficiency of viral genome packaging is indeed much higher than the particle : pfu suggests (Salgado, Upadhyayula and Harrison, 2017).

This means that the virus must instead rely on an efficient mechanism of high-fidelity, stoichiometric genome assembly. This notion is further supported by the observation that each viral RNA is present in equimolar amounts, and that no strains of rotavirus that contain genome segment duplicates or lack a gene segment have been reported (McDonald, Martha I. Nelson, *et al.*, 2016; Borodavka, Desselberger and Patton, 2018).

Genome segment assortment involves a series of inter-segment, sequence-specific RNA-RNA interactions, leading to the formation of an ‘assortment complex,’ containing a complete set of viral +RNAs (McDonald *et al.*, 2016). Each +RNAs is associated with a dedicated TEC (VP1/VP3 heterodimer), which in turn bind to core protein VP2 (McDonald and Patton, 2011; Trask, McDonald and Patton, 2012). Each genome segment is associated with dedicated TEC, which nucleates VP2 to assemble a shell around the complete assortment complex. As a result, the number of TECs within a viral particle is equal to the number of packaged genomic RNAs. This has been demonstrated for cytoplasmic polyhedrosis virus (10 genomic segments), which has TECs localised to 10 of the 12 five-fold vertices of the viral capsid. The remaining 5-fold vertices are instead occupied by viral RNA (Zhang *et al.*, 2015). Recent cryoEM studies of RV, BTV and aquareovirus have also indicated the presence of unoccupied 5-fold vertices, indicating that they package a stoichiometric set of genomic RNAs each bound to their own TEC (Ding, Nguyen and Zhou, 2018; Ding *et al.*, 2019; He *et al.*, 2019). This assembly mechanism requires the assortment of a complete viral genome as a prerequisite for capsid assembly. Sequence-specific inter-segment RNA-RNA interactions are the key determinants of specific genome packaging. However, due to the highly structured nature of viral RNAs, sequences involved in such interactions may be sequestered by stable RNA structures and therefore not spontaneously interact. This has been observed in vitro, as viral +RNAs often do not spontaneously interact (Borodavka *et al.*, 2017). To overcome this, RV encodes the non-structural (i.e. is not part of the viral capsid) ssRNA-binding protein NSP2. NSP2 has several RNA chaperone-like activities (including ssRNA-binding, helix destabilisation and strand annealing) that allow it to bind and remodel viral RNAs, exposing trans-interacting sequences (Taraporewala and Patton, 2001; Borodavka *et al.*, 2017). By binding to multiple RNAs concurrently, NSP2 is able to act as a matchmaker of RNA-RNA interactions (Schuck *et al.*, 2001; Borodavka *et al.*, 2017). As NSP2 does not bind double-stranded (ds)RNA, it dissociates from the nascent duplex and rebinds elsewhere (Taraporewala and Patton, 2004; Borodavka *et al.*, 2017). This chaperone recycling allows NSP2 to promote multiple RNA-RNA interactions throughout the viral genome. The structure and function of NSP2 is discussed further below (**Section 1.33**).

1.33 Structure and function of NSP2

In vitro studies have demonstrated that NSP2 binds ssRNA without sequence specificity (Taraporewala, Chen and Patton, 2001). RNA structure probing experiments have indicated that NSP2 is able to induce major conformational rearrangements in large RNA structures, although the most stably structured RNA helices remain inaccessible and do not become disrupted in the presence of NSP2 (Borodavka *et al.*, 2017). This indicates that the RNA chaperone activity of NSP2 may be fine-tuned by the stability of RNA structure – only metastable or unstructured regions of +RNAs are targets for remodelling by NSP2.

NSP2 is a homo-octameric protein (35 kDa monomer, 280 kDa octamer) with a toroidal architecture, although no functional roles have been attributed to its central pore. It assembles as two stacked tetramers, with overall four-fold dihedral (D4) symmetry. A notable feature of NSP2 is the equatorial basic groove (**Figure 1.33**) (Jayaram *et al.*, 2002). Based on this observation, it has been proposed that this is the RNA-binding site. A low-resolution cryoEM reconstruction of NSP2 in complex with RNA and mutagenesis studies support this notion, but the mechanism and underlying structural basis of RNA binding and remodelling is somewhat ambiguous (Jiang, *et al.*, 2006; Vasquez-Del Carpio *et al.*, 2006b).

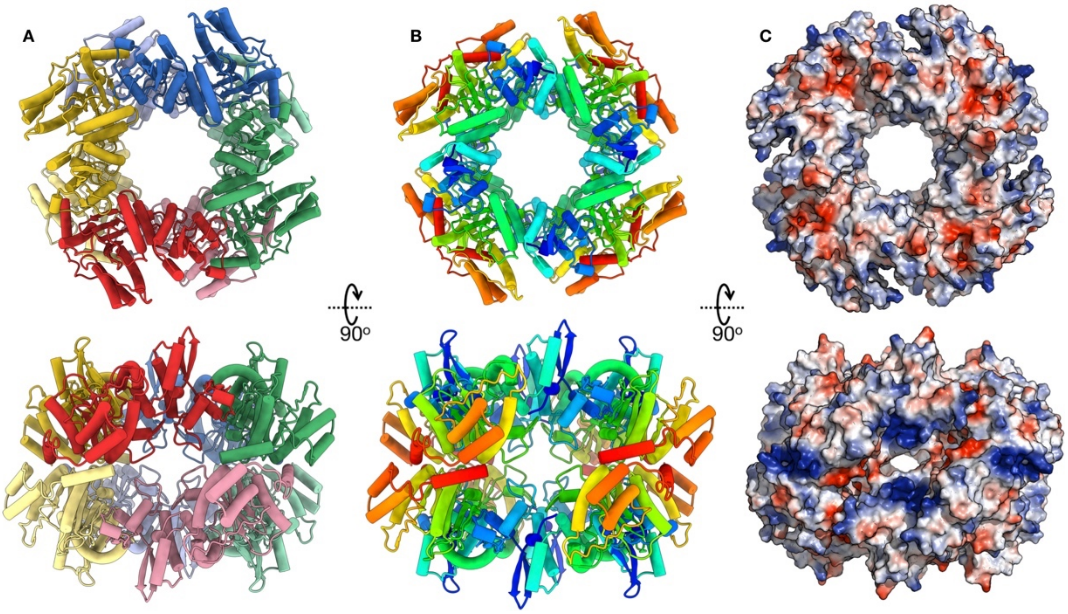


Figure 1.33 Structure of NSP2 octamer NSP2 octamer (PDB 1L9V) rendered in three different ways: A, by monomer, showing the arrangement of two tetrameric rings on top of one another. B: By sequence position, from N-terminus (blue) to C-terminus (red). C: By surface electrostatic potential (as calculated using PyMol). Electropositive regions are blue, whereas electronegative regions are red.

CryoEM analysis of a NSP2 in complex with a short fragment of NSP5 (residues 66 – 188) indicated that NSP5 may also interact with NSP2 via the positively-charged groove, competing with RNA for binding to NSP2 (Jiang *et al.*, 2006). Interestingly, this positively-charged groove has also been reported to bind tubulin (Martin *et al.*, 2010). This particular study suggested that NSP2 plays a role in microtubule (MT) depolymerisation, although the authors fail to provide direct evidence to support this claim. Ultimately, the resolution of all previously published cryoEM reconstructions of NSP2 and its various complexes are limited (ranging from >10 Å to ~25 Å). Thus, the structural interpretation of these maps and the conclusions reached are debatable, and shall be revisited using new, higher resolution reconstructions.

There have been a number of crystal structures of NDP2 determined, although this has led to some potentially incorrect or contradictory hypotheses concerning the role of NSP2 in viroplasm assembly. One particular structure of NSP2 exhibited a 'flipped out' C-terminal helix conformation that participates in domain swapping with neighbouring NSP2 octamers (**Figure 1.34**) (Hu *et al.*, 2012). Based upon this observation, it was proposed that such domain swapping interactions are crucial for viroplasm assembly, Alternatively, this may be a crystallographic artefact, since such domain swapping, octamer-octamer interactions have not occurred in any of the other >10 NSP2 crystal structures available. Although this domain-swapped octamer-octamer interaction has yet to be validated, the C-terminus of NSP2 is essential for RV replication, although its role is as yet unknown (Criglar *et al.*, 2014).

More recently, it was proposed that NSP2 phosphorylation was important for viroplasm assembly. Although this may be true, this was based upon the observation that NSP2 containing a phosphomimetic residue (S313D mutation) rapidly crystallised, and novel octamer-octamer interactions occur within the crystal lattice (Criglar *et al.*, 2018). These interactions, however, do not involve domain swapping, and involve a completely different octamer-octamer interface than had been previously reported. This interface would in fact occlude the surface-exposed positively-charged groove of NSP2 from binding RNA, NSP5 and MTs. Again, one must be cautious not to over-interpret such data and attribute a biological role to interactions that may simply arise from the specific crystallization conditions.

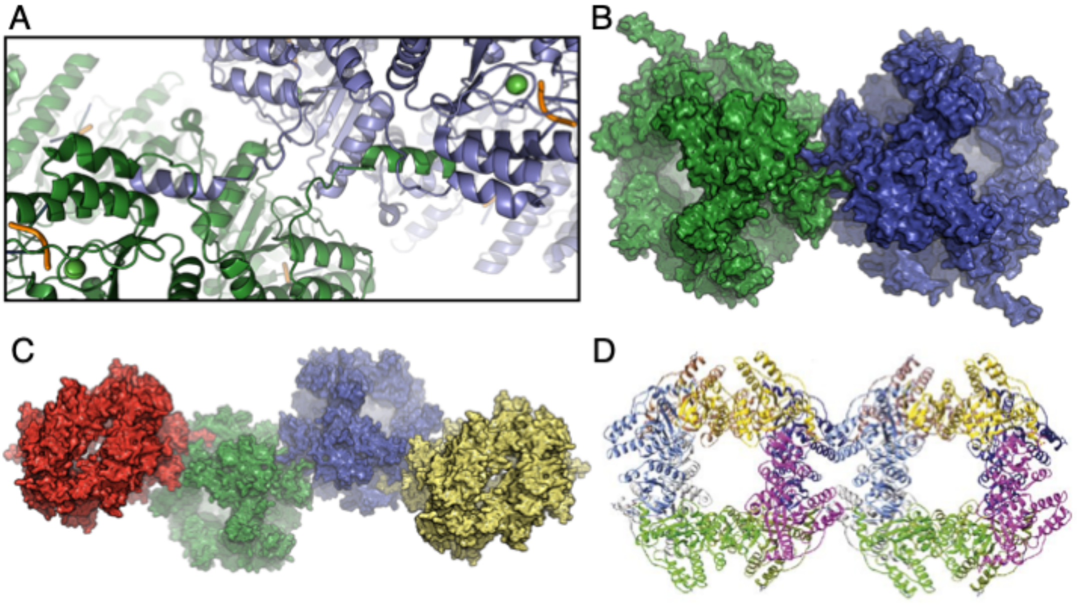


Figure 1.34 NSP2 octamer-octamer interactions A & B: Octamer-octamer interactions involving C-terminus domain swapping (PDB 4G0A) (Hu *et al.*, 2012). A: Zoom-in and cartoon representation of octamer-octamer interactions depicted in B as a surface representation. Green spheres are co-crystallised Cl^- ions. Co-crystallised RNA fragment is orange. C: Long-range octamer interactions present *in crystallo*. D: Alternative octamer-octamer interactions between NSP2 S313D point mutants (i.e. phosphomimetic) (PDB 6CYA) (Criglar *et al.*, 2018). No C-terminus domain swapping occurs, and the orientation of octamer-octamer interactions would preclude the basic groove present on the surface of NSP2 from RNA binding.

NSP2 contains a histidine triad (HIT)-like motif, a small domain that commonly occurs within proteins with nucleotide hydrolysis activity (Jayaram *et al.*, 2002). Indeed, NSP2 possesses nucleotide hydrolysis activity, hydrolysing the γ-phosphate group of nucleoside-triphosphates (NTPs) and 5’RNA triphosphates (RTPs) (Taraporewala *et al.*, 2006; Vasquez-Del Carpio *et al.*, 2006a). This activity appears to play a role in triggering the hyperphosphorylation cascade of NSP5 required for viroplasm assembly (Criglar *et al.*, 2014). The nucleotide hydrolysis activities of NSP2 are therefore more likely to be involved in viroplasm assembly rather than in RNA chaperone activity, as the presence of 1mM ATP has no effect on NSP2-mediated RNA-RNA interactions (Borodavka *et al.*, 2017). Like many other RNA chaperone proteins, the RNA-RNA annealing activity is not coupled to ATP hydrolysis (Rajkowitsch *et al.*, 2007; Woodson *et al.*, 2018).

NSP2 has never been reported to co-package with the viral RNA, indicating that it is somehow evicted from the viral RNA. As assortment complex assembly occurs, there is an increasing amount of dsRNA present (arising from inter-segment RNA-RNA interactions), effectively reducing the amount sites NSP2 is able to bind. Additionally, VP2 binds RNA and may compete with NSP2 as the inner capsid layer is assembled. There may be further unidentified mechanisms of NSP2 release.

1.34 Genome segment assortment in mammalian orthoreovirus and avian orthoreovirus

Mammalian orthoreovirus (MRV) and avian orthoreovirus (ARV) both have genomes distributed between 10 RNA segments. MRV and ARV capsids are highly similar, but differ from RV capsids in that their capping enzyme oligomers form extended ‘turrets’ protruding from the inner core layer, while being connected to the RdRP within the inner core. The use of intersegment

RNA-RNA interactions as a mechanism to achieve stoichiometric genome packaging is believed to be universal to all *Reoviridae* family members (Borodavka *et al.*, 2018).

Reverse genetics studies of MRV have indicated that the 5'-terminal ~200 nucleotides and 3'-terminal ~180 nucleotides are crucial for assortment and packaging (Borodavka *et al.*, 2018). However, these sequence lengths are greater than the UTR, and cover regions of the ORF. The exact sequences involved in this packaging remain unexplored. There have been no studies of the sequences involved in ARV genome segment assortment.

Similarly to RV, MRV and ARV replication occurs in cytoplasmic viral factories (Broering *et al.*, 2004; Tourís-Otero, Martínez-Costas, *et al.*, 2004). In either virus, virus factory assembly is mediated by a single protein, *μ*NS. *μ*NS contains two C-terminal coiled-coil regions that are sufficient for viral factory assembly (Parker *et al.*, 2002). The N-terminus of *μ*NS contains several partially overlapping regions that are involved in recruitment of viral structural and non-structural proteins, resulting in their accumulation within viral factories (Miller *et al.*, 2003, 2010). In electron tomography studies of late MRV viral factories (16 HPI), lattice-like arrays of assembled particles were observed, with ~100% of progeny virus particles containing a packaged genome (Shah *et al.*, 2017).

Although virus replication and assembly occur within viral factories, *μ*NS is thought to act as a scaffold and is not directly involved in genome segment assortment. Instead, the non-structural ssRNA-binding protein σNS is thought to be involved (Borodavka *et al.*, 2015). σNS is highly abundant in viral factories (Tourís-Otero, Cortez-San Martín, *et al.*, 2004; Tourís-Otero, Martínez-Costas, *et al.*, 2004; Miller *et al.*, 2010).

Although MRV σNS and ARV σNS share little sequence similarity, they have several common features. Both proteins are 41 kDa, assemble into hexamers in solution and are highly conserved among their respective orthoreovirus strains (Tourís-Otero *et al.*, 2005; Zamora *et al.*, 2018).

In all previous studies, σNS proteins bind without sequence specificity to ssRNA, but do not bind dsRNA, and have ATP-independent RNA helix destabilisation activity (Gillian and Nibert, 1998; Gillian *et al.*, 2000; Tourís-Otero *et al.*, 2005; Borodavka *et al.*, 2015). ARV σNS has been further characterised, and has been demonstrated to possess RNA strand annealing activity, further supporting its role as an RNA chaperone in genome segment assortment (Borodavka *et al.*, 2015). MRV σNS has been proposed to act as a general RNA stability factor, increasing the half-life of viral RNAs (Zamora *et al.*, 2018). There have been no observations that contradict the notion that MRV σNS may also act as an RNA chaperone.

Like NSP2, σNS is not present inside virions. The similar RNA chaperone-like activities of σNS and NSP2 have led to suggestions that they are functionally analogous, despite a lack of any sequence similarity (Taraporewala and Patton, 2004; Hu *et al.*, 2012). As there have been no structures of σNS solved to date, it is unknown whether these proteins share structural homology. It would therefore be worthwhile to undertake a rigorous, side-by-side comparison of the RNA chaperone activities of these proteins to ascertain whether they share a common mechanism of promoting RNA-RNA interactions, and to pursue a high-resolution structure of σNS.

1.3.5 Genome segment assortment in Bluetongue virus

Bluetongue virus (BTV) is a member of the *Orbivirus* genus of the *Reoviridae* family. Like other *Reoviridae* family members, BTV packages a complete genome prior to capsid assembly and dsRNA synthesis. Efficient core assembly requires the presence of all 10 BTV +RNA segments, and the presence of both RdRP and capping enzyme (forming the TEC). Bioinformatic and reverse genetic analysis have indicated that 5' and 3' UTRs contribute to the process of reassortment. However, the exact size and location of such sequences involved in RNA-RNA interactions remains poorly understood (for example, in BTV segment 9 +RNA, up to 276 5'-terminal and 393 3'-terminal nucleotides are required) (Fajardo *et al.*, 2015). It was observed that segments S7 – S10 (i.e. the smallest segments) significantly contribute to assortment complex assembly (Sung and Roy, 2014), although the roles of the larger segments remain unexplored. This led to a 'sequential packaging' hypothesis, requiring a hierarchical recruitment of viral RNAs to an assortment complex. Unlike the ϕ6 dsRNA bacteriophage which packages its three genome segments in a defined sequence into a preformed procapsid through the use of an ATPase packaging motor (Lísal *et al.*, 2004; Borodavka, Desselberger and Patton, 2018), BTV core particle assembly does not precede genome segment assortment (Lourenco and Roy, 2011). Furthermore, no plausible mechanisms of sequential packaging have been proposed, suggesting that this unsubstantiated hypothesis may be questionable.

BTV encodes non-structural protein NS2, which is somewhat functionally analogous to RV NSP2 and reovirus σNS. Like σNS and NSP2, NS2 binds ssRNA without sequence specificity, but the affinity is sensitive to the secondary structure context of the RNA (Taraporewala *et al.*, 2001; Lymperopoulos *et al.*, 2003; Taraporewala and Patton, 2004). The oligomeric state of NS2 has remained ambiguous, with multimers of 6 ± 2 monomers (Taraporewala *et al.*, 2001) observed in vitro, although phosphorylation and RNA binding are able to trigger further oligomerisation (Modrof *et al.*, 2005; Mumtsidu *et al.*, 2007). Although the role of NS2 in promoting intersegment RNA-RNA interactions has remained unexplored, it is entirely plausible that it plays a similar role to NSP2 or σNS (although possibly with a different molecular mechanism to either protein). Up to this point, all studies of BTV intersegment RNA-RNA interactions have been performed in the absence of NS2 (Sung and Roy, 2014; Fajardo *et al.*, 2015).

NS2 is unique among its functionally homologous ssRNA-binding proteins in that it is capable of assembling viral factories without the aid of other proteins (RV requires NSP2 and NSP5, in MRV and ARV this role is performed by a separate protein, *μ*NS, that is not an RNA chaperone). NS2 is phosphorylated by host cell casein kinase 2 (CK2) which results in the assembly of large inclusions. Akin to RV NSP5 and MRV/ARV *μ*NS, NS2 is capable of recruiting viral protein and RNA components and is essential for virus replication and inner core assembly (Mohl and Roy, 2016).

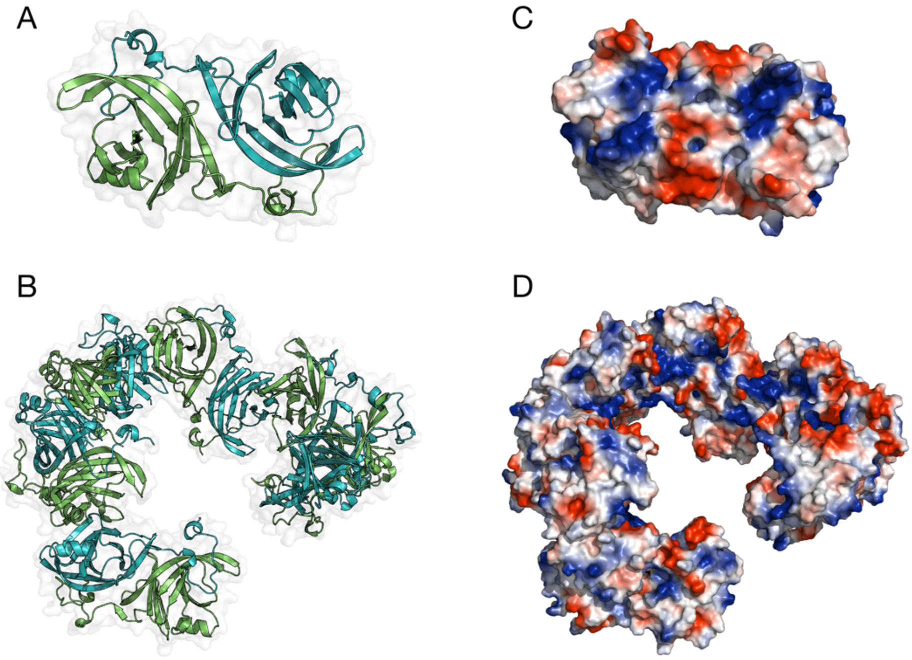


Figure 1.35 Structure of BTV NS2 N-terminal domain (NTD) Crystal structure of BTV NS2 residues 1 – 182 (PDB 1UTY). A & C: Crystal contacts between adjacent NS2 NTD molecules (A) and their surface electrostatic potential (C). B & D: Higher-order NS2 NTD structures present *in crystallo*.

To date, the only structural information available for NS2 is a 2.4 Å resolution crystal structure of the N-terminal domain (residues 1 – 182), which is responsible for ssRNA binding (specifically, residues 2 – 11 and 153 – 166) (**Figure 1.35**) (Butan, Van Der Zandt and Tucker, 2004; Butan and Tucker, 2010). Within the crystal lattice, NS2 assembles into an infinite spiralled chain. RNA-binding residues are localised to the centre of this higher-order structure (**Figure 1.35**), but it is unknown whether this is particularly physiologically relevant. C-terminal domain (CTD) forms an elongated dimer in solution, but may also assemble larger oligomers (Mumtsidu *et al.*, 2007). Collectively, these studies indicate that BTV NS2 is not a particularly amenable system for structural analysis.

1.4 Aim of the thesis

The PhD thesis project aimed to study the structure and function of viral RNA chaperones implicated in RNA selection in multi-segmented viral genomes. Specifically, the molecular mechanisms of rotavirus NSP2 and avian reovirus σNS proteins and their various ribonucleoprotein (RNP) complexes. The purpose of this work was to understand the molecular basis for how these two functionally similar proteins interact with RNA, and to provide a better understanding of the molecular mechanisms that RNA chaperone-mediated genome segment assortment.

This project was designed upon the initial assumption that NSP2 and σNS have similar RNA chaperone-like activities and therefore may be functionally interchangeable with one another. However, these two proteins had never been directly compared, and this hypothesis was based upon observations made by different groups over the span of 15 years and from studies using different substrates, reagents and experimental conditions (Gillian *et al.*, 2000; Taraporewala and Patton, 2001; Jayaram *et al.*, 2002; Tourís-Otero *et al.*, 2005; Borodavka *et al.*, 2015).

Another outstanding question in the field was the nature and mechanisms of RNA binding by NSP2. Although well-studied, NSP2 had only recently been directly demonstrated to act as a chaperone of viral intersegment RNA-RNA interactions. Given that the start of this project coincided with the cryoEM “Resolution Revolution,” (Kuhlbrandt, 2014) it was considered to be potentially worthwhile to pursue a cryoEM structure of NSP2 in complex with RNA.

Additionally, while several structures of NSP2 have been solved (Jayaram *et al.*, 2002; Jiang *et al.*, 2006; Kumar *et al.*, 2007; Hu *et al.*, 2012; Criglar *et al.*, 2018), no structures of σNS had been elucidated. Therefore, it was initially a major focus of this project to solve the crystal structure or obtain a high resolution atomic model based on cryoEM of σNS.

In summary, the goals of this PhD thesis project were to:

1. Perform a rigorous side-by-side functional comparison of the RNA chaperone-like activities of NSP2 and σNS to understand whether these functionally analogous proteins employ similar mechanisms to achieve RNA matchmaking
2. Through functional analysis, understand how these two non-specific RNA binding proteins can achieve specific RNA-RNA interactions.
3. Determine the RNA-binding interface of NSP2 using UV-crosslinking and cryoEM
4. Structurally characterise σNS as both an apoprotein and in complex with RNA.

Chapter 2: Methods and techniques

2.1 Small-angle X-ray scattering (SAXS)

SAXS is a powerful technique for the analysis of biological molecules in solution, providing a diverse range of insights into size, global shape and compactness. Although limited in resolution, SAXS data is often highly complementary to other structural and biophysical techniques. Historically, SAXS has been used to provide crucial insights into early stages of RNA folding, revealing that RNA undergoes a rapid (on the millisecond time-scale) compaction upon interaction with polyvalent cations (Russell *et al.*, 2000). Since then, technical and computational advances have allowed SAXS to become a popular technique for characterizing proteins, RNA and ribonucleoprotein (RNP) complexes (Mallam *et al.*, 2011; Fang *et al.*, 2013; Lapinaite *et al.*, 2013; Peng, Yi *et al.*, 2014; Chen and Pollack, 2016).

The phenomena of SAXS relies on the elastic scattering of photons that occurs when a monochromatic X-ray encounters randomly oriented molecules in solution (Blanchet and Svergun, 2013). As these molecules are randomly oriented, the resulting 2D scattering profile is isotropic, i.e. is rotationally averaged as to give the scattered intensity I as a function of the momentum transfer q:

$$q = \frac{4\pi \, sin \, \theta}{\lambda}$$

where λ is the wavelength of the incident beam, and θ is half the angle between the incident beam and the scattered radiation. The scattering angle

2θ is defined relative to the direction of the incident beam, the scattering intensity typically decreases with increasing angle. In SAXS experiments, the small angles (typically 0.1 – 10º) used can provide structural information in the resolution range of 1 – 25 nm.

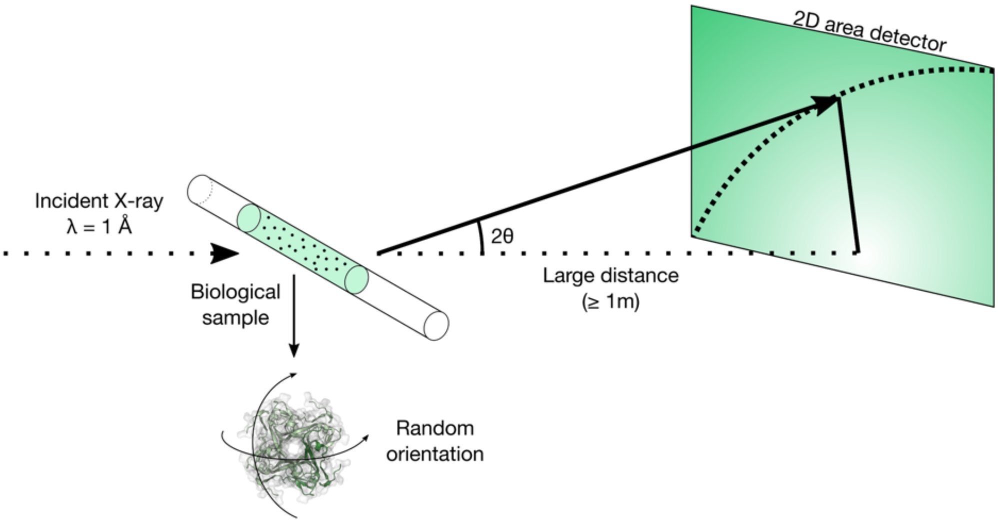


Figure 2.11 Schematic representation of a typical SAXS experiment involving a monochromatic X-ray beam irradiating a biological sample in solution. The large distance between the sample and the detector define the small angle at which the radially averaged data is collected.

SAXS is capable of measuring the size of macromolecules in solution, expressed as the radius of gyration, or R_g, which is defined as the mass distribution of a molecule. The most straightforward method to determine Rg from SAXS is based upon the Guinier approximation, as developed by Andre Guinier in the 1930s (‘Guinier, Fournet, Small Angle Scattering of X-Rays (1955) .pdf’, no date). At small q values (in the range $q\ .\ R_g < 1.3$), this is as follows:

$$I(q) = I(0).\exp\left(-\frac{1}{3}R_g^2q^2\right)$$

or

$$lnI(q) = lnI(0) - \frac{1}{3}R_g^2q^2$$

where $I(0)$ is the scattered intensity at the zero angle. Plotting the scattering curve, $I(q)$ as a function of q^2 on a semi-log scale yields a Guinier plot, which should be linear for monodisperse samples. As it is not possible to measure $I(0)$ experimentally (as elastic scattering resulting from the sample is masked by the direct beam), the linear nature of the Guinier plot allows extrapolation to $q = 0$, (i.e. $I(0)$). $I(0)$ reports on interparticle interactions and has a linear relationship with sample concentration and mass in the absence of aggregation (an increase in $I(0)$ at low angles is indicative of attractive inter-particle interactions).

Further information about the shape of the particle of interest is provided by the pairwise-distance distribution (PDDF, or $P(r)$), the real space representation of the SAXS data, generated by a Fourier transform of the SAXS pattern:

$$P(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2I(q)\frac{\sin{(qr)}}{qr}dq$$

This function describes the paired-set of all distances between all points within an object, and in the case of a SAXS experiment describes all pairwise electron distances within the macromolecule. The $P(r)$ function theoretically is zero at $r = 0$ and $r = D_{\max}$, the maximum dimension (i.e. the maximum linear dimension of the particle), although in practice this may be affected by unfolded sample, protein aggregation of improper background subtraction. Globular particles display bell-shaped PDDF profiles, with a single maximum at ~ $D_{\max}/2$ (**Figure 2.12**). Direct Fourier transformation of the scattering curve from this finite number of points is not possible (as I(q) has a limited range from q_{min} to q_{max}. The I(0) and Rg values estimated for the indirect Fourier transform (i.e. the PDDF) are typically more accurate than those obtained from Guinier analysis as they take into account the entire scattering curve.

Kratky plots ($q^2I(q)$ vs q) provide information about the compactness of molecules (**Figure 2.12**). Globular, folded molecules exhibit a characteristic peak at lower angles (q_{p}), whereas unfolded molecules show a continuous increase in $q^2I(q)$ as a function of q. Flexible, multi-domain biomolecules can potentially be readily identified from their Kratky plots as they display a mixture of folded and unfolded features. As the height of the peak q_{p} is related to the R_{g} value ($R_{\text{g}} = \sqrt{(3/ q_{\text{p}})}$), there are limitations in the ability to compare the compactness of objects of different sizes by their Kratky plots alone. A dimensionless Kratky plot where $(qR_{\text{g}})^2I(q)/I(0)$ is displayed as a function of qR_{g} allows the angular scale of the plot to be displayed independent of the molecular size. Dimensionless Kratky plots of folded molecules exhibit a maximum value at the Guinier-Kratky point ($\sqrt{3}$, 1.104) (**Figure 2.12**), whereas unfolded or flexible molecules will display an upward-right shift from this point, as well as a plateau at higher qR_{g} values.

Further insights into macromolecular structure can be gained through the use of *ab initio* reconstructions methods to generate 3D envelope bead models based upon the SAXS data. The most popular of such programmes is DAMMIN, developed by the Svergun group (Svergun, Petoukhov and Koch, 2001). This programme generates a spherical search volume (as determined by the D_{max}) within which dummy atom 'beads' are placed, and subsequently individually moved at random to ultimately form a compact, inter-connected model that fits the SAXS data.

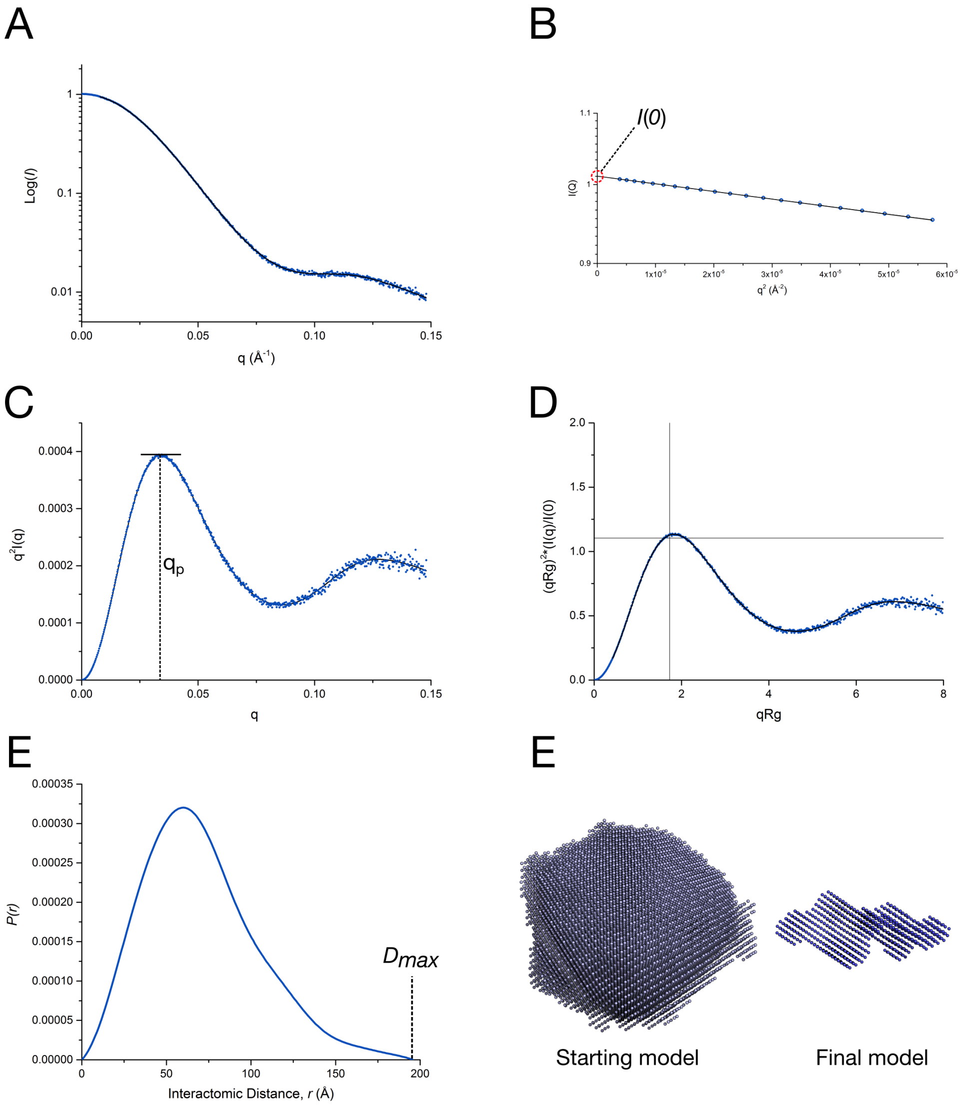


Figure 2.12 Examples of different SAXS plots for the same globular particle, where experimental data is shown in blue, and fits generated from the ATSAS suite are in black. A. Scattering curve. B. Guinier plot extrapolated to q2 = 0, or I(0). C. Kratky plot, where the height of the maximum at the low angle q_p is related to the R_g value ($R_g = \sqrt{[3/q_p]}$). D. Dimensionless Kratky plot, where the anglular scale of plot is displayed independent of maximum size. Cross-hairs denote the Guinier-Kratky point, the theoretical maximum for a globular particle. E. Pair-wise distance distribution function, indicating the maximum dimension of the particle, D_{max}. E. Typical starting and final ab initio bead models as generated by DAMMIN. The starting model is a sphere with a diameter corresponding to D_{max} the volume of which is subsequently reduced using simulated annealing methods to generate a compact final model whose theoretical SAXS profile is in good agreement with the experimental SAXS profile.

A simulated annealing algorithm is used to guide the search function to an optimal bead configuration. The program GASBOR implements an alternative approach by substituting dummy atoms with dummy ‘residues,’ with altered scattering intensity, and a penalty term is applied to ensure that the ensuring model is interconnected and assumes a ‘protein-like’ structure (Tuukkanen et al., 2016).

The unique strength of SAXS is to provide (albeit low-resolution) insights into the structure, function and behaviour of an ensemble population of biological macromolecules in solution, which is not amenable to other biophysical techniques. By combining SAXS data with different modelling approaches and other complementary methods, insights into the assembly and behaviour of dynamic biological complexes can be obtained (Yi *et al.*, 2014; Chen and Pollack, 2016; Schlundt et al., 2017).

2.2 Electron cryomicroscopy (cryoEM)

Electron cryomicroscopy (cryoEM) is a powerful weapon within the arsenal of biomolecular structure determination. Through the recent technical advancements that have enabled high resolution cryoEM structural determination, it is rapidly burgeoning into a viable alternative to X-ray crystallography (Kuhlbrandt, 2014; Glaeser, 2019). The ability of cryoEM to visualise macromolecules such as proteins, nucleic acids and lipids under non-crystalline conditions has resulted in something of a renaissance in single particle analysis (SPA).

In transmission electron microscopy (TEM), images are created by the transmission and scattering of an electron beam through a specimen, resulting in a 2D projection image. The short wavelength of electrons (picometres) allows potentially high-resolution information to be obtained (Henderson, 1995). A transmission electron microscope fundamentally consists of an electron source, a series of electromagnetic lenses (to focus the electron beam), a specimen holder, and a detection system (**Figure 2.21**). These components are held under a high vacuum as gas molecules within the electron path will scatter the beam. Hydrated samples would rapidly dehydrate within a vacuum, meaning that various sample preparation techniques must be utilised to preserve the sample in a native (or native-like) state (Cheng *et al.*, 2015).

Image formation

The projected image is a consequence of electron scattering, which can occur in two main ways; elastically and inelastically. Elastic scattering alters the path of the electron without resulting in energy loss, while elastic scattering results in transfer of energy from the electron to the specimen. Inelastic scattering events occur frequently (3 – 4 inelastic scattering events for every elastic scattering event that contributes to image formation), and result in radiation-induced damage to the sample (**Figure 2.21**) (Glaeser, 2019).

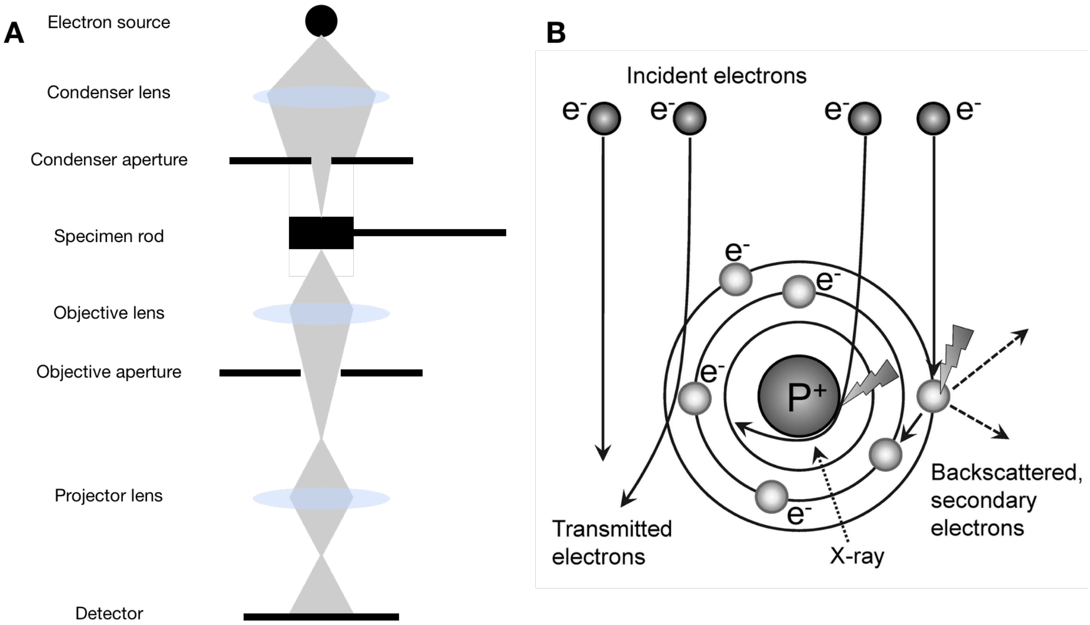


Figure 2.21 Anatomy of a transmission electron microscope (A) and interactions of the electron beam with the sample (B). A: Schematic of electron microscope featuring electron beam (grey) and lenses (blue). B: Incident electrons interacting with the specimen. Elastically scattered electrons interact with the Coulombic field of the atom, and are deflected without energy loss. Inelastically scattered electrons interact with the atom, and energy is lost. Adapted from (Orlova and Saibil, 2011).

Elastic scattering occurs when electrons are deflected by the electrostatic (Coulombic) field of an atom, and depends on the electron cloud and the nuclei. For neutral and positively charged atoms, electron scattering factors (the Fourier transforms of electrostatic potential distributions) are positive at all angles (resolutions), and therefore contribute to image formation (**Figure 2.21**) (Orlova and Saibil, 2011). Negatively charged atoms have negative scattering factor amplitudes at low scattering angles but positive amplitudes at high scattering angles. This phenomenon can result in reduced positive potentials and potentially negative intensities (i.e. holes) in cryoEM density maps acids (Marques et al., 2019). This can, however, provide additional details about the charge state of atoms – ionic oxygen atoms have negative scattering factors at low angles, whereas uncharged oxygen atoms have positive scattering factors. These effects are most common and pronounced around carboxyl groups of protein acidic side chains and phosphodiester groups of nucleic acids (Marques et al., 2019).

Microscope hardware

The current high-end (and most expensive) microscopes operate at 300 kV acceleration, although 200 kV microscopes are also capable of producing high-resolution structures (Herzik, Wu and Lander, 2017, 2019; Sandate *et al.*, 2019). Although lower acceleration voltages provide greater image contrast with similar signal : noise ratios (SNRs) than at higher energies, the amount of inelastic scattering increases at lower energies and thus the electrons are less capable of penetrating the sample (Glaeser, 2019). Furthermore, direct electron detectors (DEDs) perform better at higher energies as higher-energy electrons have reduced a point-spread function

(PSF) resulting in increased detector quantum efficiency (DQE). DQE is defined as follows:

$$DQE = \frac{SSNR_{out}}{SSNR_{in}}$$

Where $SSRN_{out}$ and $SSNR_{in}$ are the squares of the SNR from the output and input of the camera, respectively (Orlova and Saibil, 2011).

One of the major technical advancements in EM has been the development of direct electron detectors (DEDs) (Ruskin et al., 2013; McMullan *et al.*, 2014). Obviously, the method of recording the image is paramount to the quality of the data obtained, and thus DEDs provide several advantages over traditional (and somewhat outdated) methods of image formation (film and charged-coupled device (CCD) cameras). CCDs typically have a point-spread function larger than the physical pixel size, leading to poor performance at high resolution, while DEDs allow electrons to directly impact the pixelated sensor and therefore have higher accuracy and detector quantum efficiencies (DQEs).

The ability of DEDs to record images as multiple-frame movies has been a major technical advancement. Higher frame rates (>20) are capable of accounting for beam-induced movement that occurs at the start of the movie.

The greatest amount of beam-induced movement occurs rapidly after intial electron exposure, rendering the first few frames somewhat redundant. This can be compensated for through the use of detectors with variable frame-rates (such as the Gatan K3 detector) (Glaeser, 2019).

Additionally, fractionating an image into multiple frames allows for increased electron exposure and therefore increased sample contrast. Although this increases the amount of radiation damage, frames towards the end of the movie (which have accumulated the highest electron dose) can be discarded during later stages of image processing, reducing the effects of radiation damage in the final 3D reconstruction while benefiting from increased contrast during particle picking.

An ideal camera would combine high frame rate (exemplified by the Gatan K2 which can collect at 400 frames per second (fps)) with a ‘large’ pixel size (exemplified by the 15 μm pixels of the Falcon II, which are a good match for the PSF of 300-kV accelerated electrons) (Glaeser, 2019). A high frame rate is essential as it allows use of low electron exposure per frame with low probabilities of a given pixel encountering multiple electrons within the same frame. A smaller pixel size can result in the misassignment of electrons encountered by the detector to neighbouring pixels and causes the modulation transfer function (MTF, a description of how different spatial frequencies are recorded by the detector) to decrease at high resolution.

Furthermore, detectors can operate in either counting or integrating modes. Integrating mode (also known as linear mode) detectors combine the signal arising from all electrons encountered by the detector within a frame. This effectively increases the noise in the final image due to the variability of electron energy across the detector. In counting mode, the individual electron events are detected, thus greatly reducing the camera readout noise and Landau noise (statistical noise arising from energy deposition from the encountered electron). Integrating mode detectors have lower DQE but can collect with shorter exposure times and higher dose rates. Data acquisition is therefore more rapid and suffers less from specimen drift (Glaeser, 2019).

Negative stain EM

Negative staining is a common and relatively similar method for sample preservation of macromolecular complexes at room temperature. Typically, sample is adsorbed onto a support film (typically ~10 nm thick carbon) and imbedded into a solution of heavy metal salt and dried. 1 – 2% uranyl acetate is commonly used as a stain. This allows negative visualisation of the sample as the heavy metal atoms strongly absorb electrons, resulting in amplitude contrast between the absence of stain (i.e. the specimen) and the surrounding shell of heavy metal atoms. Although positive stain (where the sample becomes stained) can occur, this typically yields less informative images (Orlova and Saibil, 2011).

The high amplitude contrast typical of negative stain imaging gives a high SNR. Although adsorption of stain can induce artefacts (no interior density variations of the molecular are visualised and sample shape can be distorted), negative stain EM provides an avenue to visualise samples and rapidly generate 2D and 3D data (Orlova and Saibil, 2011; Toropova et al., 2017).

CryoEM

CryoEM involves imaging the sample of interest in a thin layer of vitreous ice, maintained at <180ºC. This allows preservation of biomolecules in native, hydrated states. The preparation of cryoEM samples is non-trivial, as the sample must be frozen rapidly (~10^6K/second) (Dubochet *et al.*, 1988) to prevent the formation of hexagonal crystalline ice within the hydration shell of complexes. Once vitrified, the sample must be maintained at these temperatures to prevent the formation of cubic ice (Adrian *et al.*, 1984).

CryoEM grids are traditionally 3.05 mm in diameter, with a metal mesh made from copper or gold, coated with a film of holey carbon. This gives the appearance (under high magnification) or a grid of squares, each containing an array of circular holes. Prior to sample application, the grid is rendered hydrophilic through treatment with glow discharge and UV irradiation, although the exact nature of grid treatment will vary between samples. Sample is then applied to the grid (~3 *μ*l) and blotted away to ensure that only a thin film of sample (ideally <100 nm) is suspended within holes. The grid is then plunged into a cryogen (typically liquid ethane cooled by liquid nitrogen) to vitrify the sample (Cheng *et al.*, 2015).

As the field of cryoEM has matured and become increasingly widespread, the perils of the air-water interface (i.e. the surface of the vitrified sample) have become clearer. The majority of vitrified samples are adsorbed into the air-water interface, which can damage the sample and induce a preferred sample orientation (Tan *et al.*, 2017; Lyumkis and Baldwin, 2019). The use of surfactants can ameliorate this effect somewhat; the zwitterionic detergent CHAPSO (3-([3-Cholamidopropyl]dimethylammonio) – 2 – hydroxy – 1 - propanesulfonate) has been demonstrated to reduce preferred orientation effects that may arise from sample surface adsorption (Chen *et al.*, 2018). Alternatively, the use of a support film (such as hydrophilised graphene) can reduce protein denaturation at the air-water interface (D’Imprima *et al.*, 2019; Naydenova, Peet and Russo, 2019).

Ultimately, the length of time that the sample remains on the grid between blotting and vitrification is a major contributing factor. The longer particles remain in solution, the greater the opportunity they have to diffuse to (and subsequently get stuck in) the air-water interface (Noble, Dandey, *et al.*, 2018). Although this can be circumvented through the use of rapid plunge-freezing techniques, the use of such methods is not widespread (Noble, Wei, *et al.*, 2018).

Data processing

Motion correction

As mentioned above, during early frames of a movie, the sample undergoes large movements, which can result in image blurring, and the loss of high-resolution information (Li *et al.*, 2013). This is thought to occur as a result of radiation-induced ‘tension relief’ of the thin biological specimens (induced through thermal compaction as the sample is cooled), which can result in motion in the z axis (termed ‘doming’), uniform whole-frame motion perpendicular to the axis of the beam, and local anisotropic motions. Within a movie, these motions can be tracked and accounted and corrected for (**Figure 2.22**). Frames are compared to one another, allowing the total frame stack to be shifted to maximise their correlation, ultimately resulting in ‘unblurring’ and image quality enhancement (Vilas *et al.*, 2018).

It is worth noting that holey carbon support films are unstable, contributing to beam-induced particle motion. This is because carbon films ‘crinkle’ when cooled from ambient temperatures to cryogenic temperatures (Glaeser *et al.*, 2015; Russo and Passmore, 2016). This greatly increases doming effects as the specimen is effectively under increased tension. This can be ameliorated somewhat through the use of metal (e.g. gold) grids that do not experience crinkling (Russo and Passmore, 2014, 2016). However, as the thermal contraction of vitreous ice is much greater than that of the support film, beam-induced motion will still occur.

CTF correction

The contrast transfer function (CTF) is a mathematical description of the point-spread function of the microscope. It is a result of cumulative imperfections of the imaging conditions (aberrations within the lens system, other imperfections resulting from the microscope). A major factor that influences the CTF is the defocus at which the image is taken (Zhang, 2016).

During the CTF correction, the defocus and CTF parameters are estimated through iteratively fitting the Thon rings within the power spectrum of each micrograph to simulated power spectra (**Figure 2.22**). There have been several recent improvements in CTF correction software, notably gCTF (which greatly accelerates CTF estimation using graphics processing units (GPUs)) (Vilas *et al.*, 2018). Further downstream the data processing pipeline, it is possible to perform per-particle CTF correction, whereby the defocus values for each individual particle within a micrograph are calculated. This can compensate for differences in ice thickness or z-height of each particle that often occur within a micrograph, improving the accuracy of CTF determination (Zivanov *et al.*, 2018).

The oscillating nature of the Thon rings means that at any given defocus value, there are multiple spatial frequencies that will have zero (and negative) contrast. While correcting for negative contrast is trivial (through phase flipping), it is necessary to collect data across a range of different defocus values to restore the information lost from CTF zero frequencies. In reality, this is not strictly necessary – no microscope is stable enough, or capable of estimating z-height accurately and consistently enough to ensure that all micrographs taken are at *identical* levels of defocus (i.e. there will be sufficient deviation in defocus to compensate for CTF zero frequencies). Thus, even a dataset collected at a "single" defocus value can still produce a high-resolution cryoEM map (although this is not advisable) (Lyumkis, 2019).

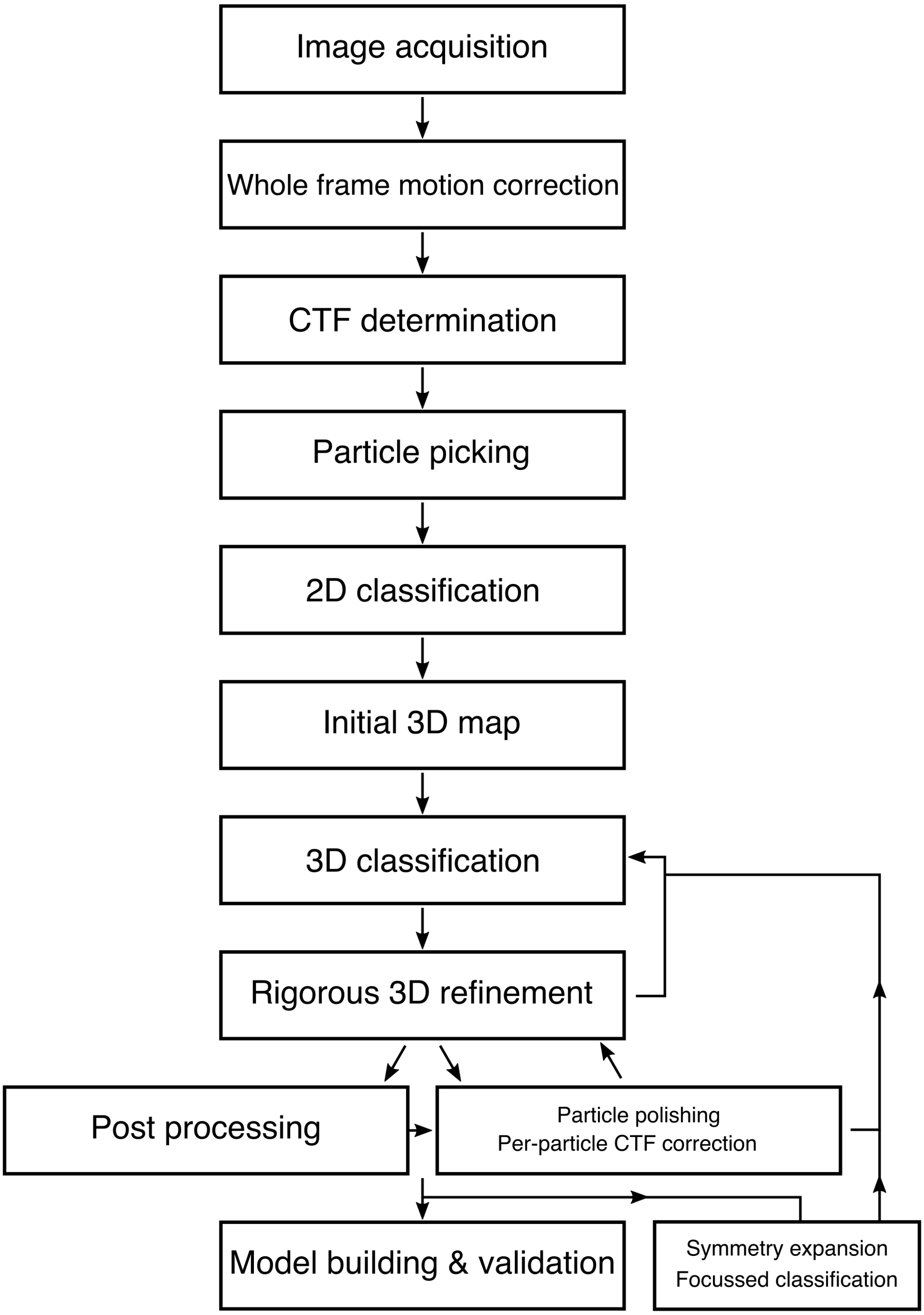


Figure 2.22 – Typical cryoEM SPA data processing workflow

Particle picking

In a traditional cryoEM data processing work-flow, a small subset of particles is manually picked (1000 – 5000) and subjected to 2D classification to generate reference templates for autopicking (**Figure 2.22**). This is a somewhat problematic step as it requires significant user time and can introduce bias at several stages.

Firstly, the user is biased to pick the most obvious or clear particles present within a micrograph. As the particles at the air-water interface of the ice have the highest contrast, they are the easiest to pick but these may represent partially damaged or denatured particles by the air-water interface (Chen *et al.*, 2018; D'Imprima *et al.*, 2019). The use of these potentially damaged particles as references for autopicking is not ideal. Additionally, manual picking a small subset of particles also introduces bias as dominant or preferred orientations will align well during 2D classification to suitable autopicking templates, while less abundant orientations will not (Tan *et al.*, 2017). This biases the autopicking to select preferred orientations, causing less abundant orientations to be overlooked. Preferred specimen orientations result in nonuniform distributions of angular projections and under-sampling of components in Fourier space during 3D reconstruction, producing a cryoEM map with smeared density (Tan *et al.*, 2017). Furthermore, micrographs with greater defocus values have greater contrast and are therefore easier to manually pick. This means that the references generated from these particles will be optimised to autopick from micrographs with the greatest defocus values rather than from micrographs taken closer to focus (which have less contrast but contain higher resolution information). Therefore, the use of these templates can result in the under-picking of micrographs with less defocus. Finally, if there is a structure of a complex already known, the user is biased to pick particles that resemble said known structure. The use of these particles as autopicking templates may prevent the picking of less abundant (and potentially biologically relevant) species or conformations. For these reasons, reference-free autopicking using Laplacian-of-Gaussian (LoG) filtering methods implemented in Relion 3 was employed (Zivanov *et al.*, 2018).

Laplacian filters are used to find areas of strong difference (i.e. the interface between signal (particles) and background (ice)). Application of a Gaussian filter to 'smooth' the micrograph prior to application of the Laplacian filter reduces the spatial frequency of background signal (noise), limiting the picking of false positives. In LoG-based autopicking, particles are selected with user-defined particle dimensions. Pixels with the greatest LoG-filtered values (i.e. sharpest edges) that satisfy these dimensions will be picked, and pixels within a user-specified distance surrounding these picked particles will be disregarded. Iterations of particle picking proceed until there are no suitable pixels remaining that satisfy these criteria. By selecting pixels with the strongest 'edges' rather than greatest signal intensity, LoG-based autopicking is optimised for picking particles within ice rather than on the carbon (as the carbon will have the strongest signal intensity). The use of a user-specified LoG threshold value also prevents selection of pixels at the edge of the hole (i.e. with the strongest edge).

2D classification

As biological samples such as proteins tend to be mainly composed of light atoms (C, H, O, N), which weakly interact with electrons, they ‘deflect’ rather than absorb the incident beam (**Figure 2.21**). The exiting electron wave experiences little intensity modulation and the resulting image has low amplitude contrast. Even with high defocus values, the main drawback of cryoEM SPA is the low contrast and poor SNR. This requires the acquisition of a large number of experimental images which can then be aligned and averaged, generating 2D averages.

Image alignment involves determining the relative orientation of each low SNR particle and calculating its angular relationship with other particles within the dataset. Following 2D alignment, particles are classified into different subsets (i.e. class averages). Reference-free 2D classification performs alignment and classification without user supervision (i.e. the initial references are not defined) and helps to avoid user bias.

Relion (Scheres, 2016) implements a Maximum Likelihood (ML)-based approach, where each particle contributes to all classes in a weighted manner (i.e. a particle that is likely to be within a certain class average is upweighted and therefore contributes more than an ambiguous particle). This provides an advantage over standard cross-correlation (CC) methods as ML methods account for the inherent uncertainty in classification of low SNR data (Scheres, 2012).

Besides improving the SNR, 2D classification can be used as a quality control step, to remove contaminants (edges of the carbon, aggregates) that arise from over-generous autopicking. Additionally, during 2D classification, individual particles are normalised to a common mean and standard deviation, as particles from different micrographs are likely have different SNR and pixel intensity values.

A bottleneck in reference-free 2D classification is the initial population of 2D classes. With large datasets, many different classes (at least 100) are required in order to represent the multiple different particle orientations present, and to separate them from junk particles. Initial sorting of these particles into different classes is slow and computationally expensive as there are no pre-existing references for which to sort the particles into (hence ‘reference-free’). Once the first several iterations of 2D classification have been performed, suitable 2D classes are used as references, and sorting particles between these references is much faster. The ‘Fast subsets’ option within Relion-3.0 (Zivanov *et al.*, 2018) overcomes this bottleneck by using smaller random subsets of particles during initial iterations, making it much faster to increase class population during early iterations. The subset of particles gradually increases through iterations, which are rapidly sorted into the already-established classes. (iterations 1 – 5: #of particles used = number of classes * 100; iterations 6 – 10: #of particles used = number of classes * 300; iterations 11 – 15: 1/3 of total particles; remaining iterations are performed with the entire dataset). Once the classification has reached iterations 16 and onwards, the classes formed using smaller subsets are suitable references in which to sort a large dataset. This process is rapid. Although this process introduces some bias into classes (multiple orientations of the same view can be easily aligned grouped into the same class if there are fewer particles, while false positive particles with strong signal-to-noise will be sorted out into distinct classes), 2D classification with fast subsets provides a quick and effective way to 'clean up' autopicked particles (i.e. separate bona fide particles from junk).

3D classification and refinement

Once a subset of homogenous, high-quality particles has been determined, these can be used to refine a high-resolution 3D reconstruction. 3D reconstruction has three main steps – initial 3D model generation, multiple rounds of 3D classification, and a final high-resolution 3D reconstruction (also referred to as 3D autorefinement). Throughout this procedure, the information within the 2D images must be conferred to the ideal 3D model. This is not a trivial process.

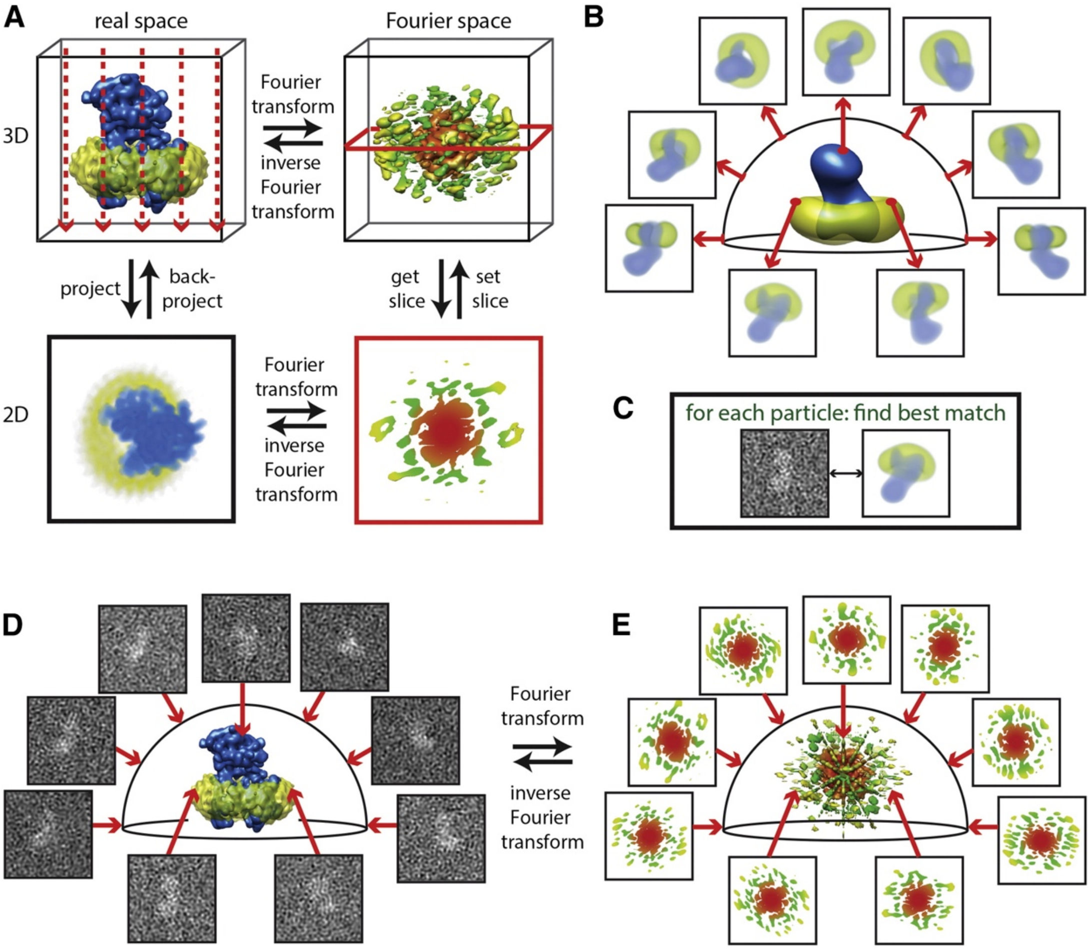


Figure 2.23 Projection-slice theorem and 3D reconstruction. A: A macromolecular complex adopts a defined orientation in real space. TEM imaging provides a corresponding projection. The Fourier transform of the projection is the central slice through the 3D Fourier transform of the 3D structure. B – E: Projection matching. 2D reference projections are calculated from a 3D model (B) which are in turn compared with experimental 2D projections (C). Projection slice theorem implies that the 3D reconstruction can then be calculated through correctly positioning 2D projections in Fourier space (E), from which the inverse transform (real space) reconstruction can be determined (D). Adapted from (Nogales and Scheres, 2015)

3D reconstruction relies on the projection-slice theorem, whereby the 2D images corresponding to multiple different views of the object are combined (**Figure 2.23**) (Nogales and Scheres, 2015). The Fourier transform of the 2D projection represents a central slice through the 3D Fourier transform of the object. By calculating the relative angular and translational relationships between multiple 2D projections, a 3D structure can be determined (**Figure**

2.23). This is achieved through projection matching, where 2D projections are calculated from a reference 3D model, and compared with experimental 2D projections. Projection matching is performed iteratively, whereby relative angles and translations of Fourier transforms of 2D projections are optimised until a consensus has been achieved. Like 2D classification, comparison of images with projections can use either CC or ML (Orlova and Saibil, 2011).

For initial 3D classification and reconstructions, an initial model is needed to provide a low-resolution cryoEM density from which 2D projections can be calculated. This is a somewhat challenging procedure for biomolecules with unknown structures, as this can lead to bias in the final map (Henderson, 2013).

Model bias can be a major pitfall of 3D reconstructions. Due to the low SNR of the individual images, the initial model must be low-pass filtered (typically to ~60 Å) to remove high-resolution features. This means that any resulting high-resolution features within the final reconstruction arise from the experimental data rather than from over-fitting noise to the starting model.

Once a subset of homogenous, high-quality particles has been determined, these can be used to refine a high-resolution 3D reconstruction. Reconstruction in Relion estimates the accuracy of angular assignments and automatically determines when the refinement has converged (Scheres, 2016).

The most widely-used method of global resolution assessment is the ‘gold standard’ Fourier Shell Correlation (FSC) (Scheres and Chen, 2012). This is used to calculate the agreement (i.e. correlation) between two independently-refined half-datasets. The map resolution is defined as the resolution at which the FSC curve falls below 0.143 (Rosenthal and Henderson, 2003). Despite its widespread use, this is a somewhat arbitrary threshold (various studies have used 0.5 or 0.25 (Nguyen *et al.*, 2015; Saathoff *et al.*, 2019)). It is also important to note that FSC analysis does not empirically determine map resolution, but rather the consistency between two half-maps.

The output from 3D refinement is a cryoEM density map that corresponds to the experimental data. A mask can be applied to this volume to remove background noise and (often) improve the FSC-derived resolution, as high-resolution uncorrelated noise is removed from the final reconstruction.

Chapter 3: Materials and Methods

Protein expression and purification

E.coli BL21(DE3)pLysS cells were transformed with a pET15b-σNS plasmid encoding an upstream hexa-histidine tag separated from the N-terminus of σNS by a thrombin cleavage site. Colonies were grown on LB-agar plates supplemented with carbenicillin (50 *µ*g/ml) and chloramphenicol (50 *µ*g/ml) before being grown overnight at 37°C in 5 ml of LB (supplemented with the same amounts of antibiotics). Overnight cultures were transferred to 200 ml antibiotic-supplemented LB and incubated until OD_{600} was between 0.6 – 0.7. 25 ml aliquots were taken and grown up to 1 L in the same manner before induction with 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) and incubated at 18°C with constant aeration (180 rpm) for at least 12 hours. Induced cells were pelleted by centrifugation at 5000 rpm for 30 minutes at 4°C using JLA-8.1 rotor.

Cells were lysed on ice by the addition of Lysis buffer (50 mM Tris pH8, 500 mM NaCl, 1 mM EDTA, 20*µ*g/ml lysozyme, 1 protease inhibitor tablet (Roche) per 25 mL of buffer, 5 mM β-mercaptoethanol) and incubated for 15 mins. 20*µ*g/ml DNase I (Sigma) was then added with 1 M $MgCl_2$ (final concentration of 10 mM), and this was left to incubate for another 30 minutes.

Lysed cells were ultracentrifuged at 20,000 x g for 1 hour using a SW-32 Ti rotor. Supernatant was loaded onto a 5 mL HiTrap HP NiNTA column equilibrated with NiNTA Buffer A (50 mM Tris-HCl pH 8, 1.2 M NaCl, 0.5 % Tween 20). After washing with Buffer A, protein bound to the column was washed with NiNTA Buffer *Int* (50 mM Tris-HCl pH 8, 0.2 M NaCl, 0.5 % Tween 20) for 10 column volumes before being eluted using buffer B (50 mM Tris-HCl pH8, 0.2 M NaCl, 1 M imidazole) with a gradient (0 – 100% over 10 column volumes). The peak eluting at ~35% buffer B contained σNS.

Fractions containing protein were diluted 5-fold with 50 mM Tris-HCl pH 8 to lower the NaCl concentration, then loaded onto a 5 ml HiTrap Q ion exchange (IEX) column equilibrated with IEX Buffer A (50 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA). After washing with Buffer A, protein was eluted using IEX buffer B (50 mM Tris pH 8.0, 1 M NaCl, 1 mM EDTA) with a gradient (0 – 100% over 10 column volumes). The σNS peak eluted at ~45% buffer B.

Pooled fractions were incubated at room temperature with 50 units per 10 ml of bovine thrombin (Sigma) for 4 hours, before being incubated at 4°C overnight to remove N-terminal his-tag. The thrombin-treated σNS was loaded onto nickel affinity column as described above, and flow-though was collected and loaded onto an IEX column as previously described. Fractions were eluted as described above and analyzed by SDS-PAGE and denaturing mass spectrometry for purity. Fractions containing protein were concentrated to a final volume of 0.5 ml and loaded onto a Superdex S200 size-exclusion column (resin volume 24 ml) pre-equilibrated with SEC buffer (150 mM NaCl, 25 mM HEPES pH 7.5) at a flow rate of 0.5 ml/min. 0.5 ml fractions were collected. Freshly purified σNS was stored by flash freezing with liquid nitrogen in 50 – 100 μl aliquots (concentrations ranging from 2 – 7 mg/ml).

Rotavirus NSP2 (RF strain) was overexpressed similarly to σNS, whereby *E.coli* BL21(DE3)pLysS cells were transformed with a pET28a-NSP2 plasmid containing C-terminally hexa-histadine tagged NSP2. Transformation and inoculation were performed as with σNS, although induced cultures were incubated at 24°C for 12 – 15 hours with constant aeration. Following induction, cell harvesting, lysis, clarification and affinity chromatography were done essentially as described above for σNS. As NSP2 has a pI of 9.1, ion exchange chromatography was performed using a 5 ml HiTrap SP column (GE health sciences). All buffers were prepared as for σNS IEX, with pH adjusted to 7.0. Elution was performed as described above. Pooled fractions were concentrated using Amicon Ultra-4 centrigugal filter units with a molecular weight cut-off of 30 kDa to a final volume of 0.5 ml and subjected to size-exclusion chromatography as described for σNS. The C-terminal his-tag was not cleaved.

RNA design and production

AlexaFluor488 dye-labelled 20mer and unlabelled, unstructured 15mer, 20mer and 40mer RNAs were purchased from Integrated DNA technologies (IDT) (see **Table 3.1** for all sequences). Cy3- and Cy5-labelled 17mer RNAs derived from the rotavirus segment S11 (nucleotides 49 – 66, complementary to segment S5 nucleotides 307 – 324) were ordered from IDT. Dual-labelled RNA stem-loop was designed to have a similar minimal free energy (MFE) of folding to that of the stable 20mer hairpin, but with a larger loop to enable efficient protein binding. A 36mer hairpin with MFE = -8.9 kcal mol^{-1} with 5'-Atto532 and 3'-Atto647N dyes was designed and purchased from IBA Life Sciences.

RNA substrate	Sequence (5' – 3')	ΔG of folding (kcal mol^{-1})*
Unstructured 20mer (unlabeled)	CUUUUCAAGACAUGCAACAA	0
Unstructured 20mer (labelled)	Alexa488-CUUUUCAAGACAUGCAACAA	0
Metstable 20mer	Alexa488-CUUCUUUCGAAAUUAAGAAG	-3.8
Stable 20mer	Alexa488-UCUUUGCAAACAUGCAAAGA	-8.1
Unstructured 40mer	CUUUUCAAGACAUGCAACAACUUUUCAAGACAUGCAACAA	-1.6
Complementary 40mer	UUGUUGCAUGUCUUGAAAGUUGUUGCAUGUCUUGAAAAC	-3.5
Unstructured 60mer	CUUUUCAAGACAUGCAACAACUUUUCAAGACAUGCAACAACUUUUCAAGACAUGCAACAA	-1.6
Cy3-17mer	Cy3-UUCCUUCUAUUUCUUCU	0
Cy5-17mer	Cy5-UUCCUUCUAUUUCUUCU	0
Dual-labelled stem loop	ATTO532-AAAUCUUUGCAAACUAUCCAAUCCAUGCAAAGAUAA-ATTO647N	-8.9
S5f / *Stem loop A*	GUAACAAA**AGAAGAUAUAGAAGGAC**UGCUAACCAUGUAUGAAACGUUGUUUC	-4.5
S11f / *Stem loop B*	AGUAUUGACGUGACGAGUC**UUCCUUCUAUUUCUUCU**AGUAUAUACAA	-6.3
Stem loop C	AAAUCUUUGCAAACUAUCCAAUCCAUGCAAAGAUAA	-8.9

Table 3.1: RNAs used in this thesis. *MFE parameters of RNA folding were calculated using the online Vienna RNA server (Gruber *et al.*, 2008). Nucleotide sequences of S5f and S11f in **bold** correspond to complementary sequences used to anneal these two RNAs.

Size-exclusion chromatography (SEC), dynamic light scattering (DLS) and small angle X-ray scattering (SAXS)

σNS-ribonucleoprotein (RNP) complex was prepared by mixing σNS (175 μM monomer) with 20mer unstructured RNA (**Table 3.1**) (70 μM) to ensure complete saturation of σNS with RNA, assuming that a single σNS hexamer binds two RNA molecules. In the case of a higher-order 20mer-RNP complex (i.e. octamer bound to 2 RNAs), this stoichiometry also ensures protein saturation with RNA. For light scattering analysis, 25 μM σNS and σNS-20mer complex were run at a flow-rate of 0.4 ml min^{-1} on a TSKgel G6000PW$_{xl}$ 30 cm x 7.8 mm SEC column (Tosoh) with an AKTA pure system at 4°C in SEC buffer (150 mM NaCl, 25 mM HEPES pH 7.5) (GE Healthcare) connected to a DAWN HELEOS (Wyatt).

SAXS samples were prepared in a similar manner, whereby σNS was incubated with saturating amounts of unstructured 20mer or unstructured 40mer RNA (**Table 3.2**), as described above. SAXS intensity data, $I(q)$ versus momentum transfer q, where q is defined as follows:

$$q = \frac{4\pi \sin\theta}{\lambda}$$

where λ is the wavelength of the incident beam, and θ is half the angle between the incident beam and the scattered radiation, were collected using SEC-SAXS on beamline B21 at Diamond Light Source (Didcot, UK) over a range of q of $0.004 < q < 0.442$ Å^{-1}. 50 μl of each sample (~175 μM) was loaded onto a 2.4 ml Superdex 200 Increase 3.2 column mounted on Agilent HPLC, and the eluent was flowed through the SAXS beam at 0.04 ml/min.

The SEC buffer used as the background was collected after one SEC column volume. SAXS data were collected at 1 second intervals using a Pilatus 2M detector (Dectris, Switzerland) at a distance of 3.9 m and an X-ray wavelength of 1 Å.

Frame-by-frame buffer subtraction of in-line SEC-SAXS data was performed using *DATASW*(Shkumatov and Strelkov, 2015). Individual SAXS .dat files were renamed prior to *DATASW* processing using the following command:

```
a=1
for i in *.dat; do
  new=$(printf "%04d.dat" "$a")
  mv -i -- "$i" "A_$new"
  let a=a+1
done
```

Output from DATASW included I0 and Rg, which were plotted as a function of recorded frame number.

Guinier plot fit and real space inversions were performed using Primus (Konarev *et al.*, 2003) and GNOM from the ATSAS software package v. 2.8.3 (Konarev *et al.*, 2006; Petoukhov *et al.*, 2012). Radii of gyration (R_g) were estimated using AUTORG (Konarev *et al.*, 2006). Low resolution envelopes were determined using the simulated annealing procedures implemented in DAMMIF (Franke and Svergun, 2009) in slow mode, with no symmetry applied. Each scattering curve generated 20 independent models, which were averaged and filtered using DAMAVER (Volkov and Svergun, 2003) and

DAMFILT with a mean normalised spatial discrepancy of 0.820 ± 0.05 (σNS apoprotein) and 0.576 ± 0.03 (σNS-20mer complex). SAXS experimental data together with the relevant experimental conditions and the derived models are available from SASBDB (SASDDU5 and SASDDT5 for σNS apoprotein and 20mer-RNP complex, respectively).

Further modelling of SAXS data was undertaken using the ATSAS suite (Konarev *et al.*, 2006; Petoukhov *et al.*, 2012). MONSA was employed for multiphase bead modelling of σNS apoprotein within the 20mer-RNP complex(Svergun, 1999). Scattering profiles corresponding to both the apoprotein and 20mer-RNP complex were input as individual phases with their corresponding radii of gyration.

The program GASBOR was used to generate ab initio models of both σNS apoprotein and 20mer-RNP complexes as a continuous chain of dummy residues (as opposed to a lattice of unconnected dummy atoms as is the case with DAMMIN)(Svergun, Petoukhov and Koch, 2001). GASBOR dummy residue models were generated without symmetry applied (i.e. in P1 symmetry), unknown anisometry and with appropriate numbers of dummy residues (2200 and 3000, for σNS apoprotein and 20mer-RNP complexes, respectively), and modelling was performed in reciprocal space mode (where $I(q)$ plot is used for fitting, as opposed to the real-space $P(r)$ plot).

Affinity measurements by fluorescence anisotropy

Fluorescence anisotropy measurements with AlexaFluor488 dye-labelled RNAs were performed at 25°C using a POLARstar Omega plate reader (BMG Labtech) in Greiner 384 well black polypropylene plates. Serial 2-fold dilutions of NSP2 and σNS were titrated into 5 nM RNA in 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.05% Tween-20 in a total volume of 50 μl and equilibrated at room temperature for 15 minutes prior to measurements were taken. Where required, buffers were supplemented with 10 mM $MgCl_2$. Raw Anisotropy (*r*) values were calculated as follows:

$$r = \frac{(I \parallel - I \perp)}{(I \parallel + 2I \perp)}$$

Where $I \parallel$ and $I \perp$ are the parallel and perpendicular emission signals, respectively. Anisotropy values were normalised to values corresponding to free RNA and RNA saturated with either protein, and plotted as a function of protein concentration and fitted to a Hill equation using OriginPro 9.0.

Electrostatic contributions to free energies of protein-RNA interactions

The dependence of K_{obs} on buffer ionic strength can be expressed as

$$\log K_{obs} = -m'\psi \log[M^+] + \log K_{obs,1M}$$

where $[M^+]$ is the monovalent counterion concentration (in this case Na^+), *m'* is the number of ion pairs formed, ψ is defined as the thermodynamic extent of counterion binding and $K_{obs,1M}$ is the non-electrostatic contribution to the dissociation constant, defined as the K_{obs} at 1 M NaCl, when the polyelectrolyte effect is minimal (Davidovich *et al.*, 2013). The slope, SK_{obs}, of $\log(K_{obs})$ against $\log[M^+]$ relates to the number of counterions released upon binding as follows:

$$SK_{obs} = -\frac{\Delta \log K_{obs}}{\Delta \log[M^+]} = m'\psi$$

The electrostatic (poly-electrolyte) contribution to free energy of binding, ΔG_{PE}, can therefore be determined as

$$\Delta G_{PE} = \Delta G_{obs} - \Delta G_{nPE} = -RT(SK_{obs} \log[M^{+}])$$

where ΔG_{obs} is the total free energy of binding, and ΔG_{nPE} is the non-electrostatic (non-polyelectrolyte) contribution. The parameter ψ was taken to be between 0.68 and 0.78 based upon previous empirically determined values(Record et al., 1976; Davidovich *et al.*, 2013). As NSP2 octamers disassemble under high ionic strengths (Schuck *et al.*, 2001), a range of 50 – 250 mM NaCl was used to determine RNA binding affinities. σNS, however, aggregates at low (<50 mM) NaCl concentrations, but is not affected by higher ionic strength conditions, so a wider range of ionic strength buffers were used to determine RNA binding affinities (up to 500 mM NaCl).

Electrophoretic Mobility Shift Assays (EMSAs) and in-gel FRET

A dual-labelled RNA stem-loop was designed with a minimum free energy (MFE) of folding of -8.9 kcal mol^{-1}, containing 5'-donor (ATTO532) and 3'-acceptor (ATTO647N) dye fluorophores as a FRET pair (**Table 3.2**). Dual-labelled stem-loop was heat-annealed in binding buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA) at 75^{o}C for 5 minutes and cooled to 4^{o}C prior to incubation with increasing amounts of σNS (final RNA concentration 10 nM). 10 μl of each sample was mixed with 2 μl 6x loading buffer (30% v/v glycerol, 5% Ficoll 400, 50 mM NaCl, 10 mM HEPES pH 8, 5 mM EDTA, 0.002% w/v bromophenol blue). Electrophoresis was carried out on a nondenaturing 1.5% agarose gel for 90 min at 100 V in 1 X Tris-borate-EDTA buffer at 4^{o}C. Gels were imaged using a fluorescence scanner (Fujifilm FLA-

5100) with 532 nm excitation, imaging donor and acceptor wavelengths separately. 2D densitometry was performed using ImageJ. Apparent FRET efficiencies ($E_{FRET(app)}$) were calculated as follows:

$$E_{FRET(app)} = {I_{ATTO647N}}/{(I_{ATTO532} + I_{ATTO647N})}$$

Where $I_{ATTO532}$ and $I_{ATTO647N}$ are donor and acceptor emission intensities, respectively. For probing RNA : RNA interactions between S5f and S11f fragments, 100 nM of each RNA was individually heat-annealed at 85°C for 5 minutes in folding buffer (10 mM HEPES pH 7, 1M NaCl, 10 mM $MgCl_2$) and snap-cooled to 4°C and incubated for 20 minutes in order to stem-loop folding prior to co-incubation. 10 nM of S5f and S11f were incubated at 37°C in annealing buffer (10 mM HEPES pH 7, 100 mM NaCl, 1 mM $MgCl_2$) and immediately analysed by electrophoresis on a native 15% acrylamide-TBE gel, run at 80 V at 4°C, and stained with 0.01% (w/v) SYBR gold. Gels were imaged using a fluorescence scanner as described above, using 488 nm excitation.

Ensemble FRET

Dual-labelled RNA stem-loop (**Table 3.2**) was heat-annealed at 75°C for 5 minutes and cooled to 4°C. Folded RNA-alone and denatured RNA (in 50% v/v formamide) were initially measured in 100 μl volumes at a final RNA concentration of 10 nM. Serial 2-fold dilutions of NSP2 and σNS from 15 μM were incubated with 10 nM RNA at room temperature for 15 minutes prior to measurement. Measurements were performed using a Fluorolog spectrofluorimeter (Horiba Jobin-Yvon). Solution $E_{FRET(app)}$ values were calculated as described for in-gel $E_{FRET(app)}$ values.

Single-pair FRET measurements

Dual-labelled RNA stem-loop used for ensemble FRET measurements was heat-annealed as described above and diluted to 10 pM final concentration. σNS and NSP2 were incubated with the stem-loop (25 nM each). Single-pair FRET measurements were performed on a custom-built confocal microscope with multiparameter fluorescence detection (MFD) and pulsed interleaved excitation (PIE) (Kudryavtsev *et al.*, 2012) as described previously (Voith von Voithenberg *et al.*, 2016). Briefly, two picosecond pulsed lasers (532 nm and 640 nm) were operated at a repetition rate of 26.66 MHz and delayed by 18 ns with respect to each other to achieve rapid alternating excitation of donor and acceptor fluorophore at 100 μW laser power. By diluting the sample to picomolar concentrations, single molecule events were detected as bursts of fluorescence as they diffuse through the confocal volume on the millisecond timescale. Bursts were selected using a sliding time window burst search (Nir *et al.*, 2006) with a count rate threshold of 10 kHz, a time window of 500 μs and a minimum of 100 photons per burst. Using time-correlated single photon counting and polarised detection, one can calculate for every molecule its FRET efficiency, labelling stoichiometry, and the fluorescence lifetime and anisotropy of the donor and acceptor fluorophores. To remove molecules lacking the donor or acceptor dye, we used the ALEX-2CDE filter (Tomov *et al.*, 2012) with a time constant of 100 μs and an upper threshold of 10. Accurate FRET efficiencies E were calculated from background-corrected photon counts in the donor channel and acceptor channel after donor excitation (F_{DD}, F_{DA}) and acceptor channel after acceptor excitation (F_{AA}) by correcting for crosstalk ($\alpha = 0.03$), direct excitation ($\delta = 0.06$) and differences in the quantum yield and detection efficiency ($\gamma = 0.65$):

$$E = \frac{F_{DA} - \alpha F_{DD} - \delta F_{AA}}{\gamma F_{DD} + F_{DA} - \alpha F_{DD} - \delta F_{AA}}$$

Species-selective fluorescence correlation functions were determined as follows: Sub-populations of molecules were selected using FRET efficiency thresholds (NSP2: low-FRET E < 0.4, high-FRET E > 0.6; σNS: low-FRET E < 0.15, medium-FRET 0.3 < E < 0.8, high-FRET E > 0.9). For every burst, photons in a time window of 50 ms around the edges of the burst were added. If another single molecule event was found in the time window, the respective burst is excluded from the analysis. Correlation functions were calculated for each individual burst using acceptor photons after acceptor excitation to ensure that the obtained correlation functions are independent of the FRET efficiency. Species-selective correlation functions were then averaged to obtain the burst-selective correlation function (Eggeling *et al.*, 1998; Laurence *et al.*, 2007). All analysis was performed using the *PAM* software package (Schrimpf, Waldemar, 2018).

Fluorescence cross-correlation spectroscopy (FCCS)

Equimolar amounts of 18-nt Cy3- and Cy5-labelled non-complementary RNAs (10 nM each) were incubated with varying concentrations of NSP2 and σNS in 50 mM NaCl, 20 mM HEPES pH 7.4. Interactions between S5 and S11 RNAs and S5 and Cy5-17mer were measured as previously described

(Borodavka *et al.*, 2017). Briefly, 55 nM of each RNA strand was incubated with 5 – 10 μM NSP2 or σNS at 37°C for 15 minutes. Samples were then diluted into 150 mM NaCl, 20 mM HEPES pH 7.4, 0.05% Tween 20, resulting in a final RNA concentration of 1 nM each labelled RNA. It has been previously observed (Borodavka *et al.*, 2017) that NSP2 removal using proteinase K does not significantly reduce the amplitude of cross-correlation, (therefore suggesting that the observed cross-correlation is due to direct RNA-RNA interactions). Proteinase K treatment was therefore not employed.

Circular Dichroism (CD)

CD experiments were performed on a Chirascan plus spectrometer (Applied Photophysics). Samples were prepared by dialyzing protein solutions against 10 mM phosphate buffer pH 7.4, 50 mM sodium fluoride. Spectra were recorded over a wavelength range of 190 – 260 nm, with a bandwidth of 1 nm, step size of 1 nm, using a 1 mm path-length CD cuvette (Hellma precision cells, synthetic far-UV quartz). An average of 3 scans were used to generate the final spectra, and baseline correction was performed via subtraction of the buffer spectrum. Thermal stability was analysed by monitoring the CD signal at 222 nm during heating from 20°C to 70°C with a heating rate of 1°C min^{-1}. Measured ellipticities, θ, (mdeg) were expressed in terms of mean residue ellipticity (MRE, with units deg cm^2 dmol^{-1}) as follows:

$$MRE = \frac{\theta \cdot MRW}{l \cdot c \cdot 10}$$

Where *MRW* is the mean residue weight (*MRW* = 110.45), *l* is the path length in cm, and *c* is the protein concentration in mg/ml.

CD spectra of unstructured 20mer RNA (**Table 3.2**) (4 μM) was performed as above, over a wavelength range of 240 – 320 nm. Relative intensities at 263 nm and 271 nm were observed as reporters on RNA secondary structure, with a peak at 263 nm being indicative of A-form double-stranded RNA and a peak at 271 nm being typical for unstructured, single-stranded RNA (Borodavka *et al.*, 2015).

Raman spectroscopy

Raman spectra of RNA, σNS and 20mer-RNP (all 25 μM) and their corresponding buffers were acquired on a modular multi-channel Raman spectrograph Jobin Yvon–Spex 270M in 90° scattering geometry using 532 nm line of a frequency doubled continuous-wave solid-state Nd:YVO_4 laser for excitation (power of 240 mW at the sample), as described in detail elsewhere (Palacký, Mojzeš and Bok, 2011). Raman measurements were performed in a temperature-controlled hermetical quartz microcell (4 μL volume) at 20°C and 60°C. Final spectra represent averages of 30 - 720 individually acquired and analysed scans (depending on the sample type) each of 1 min integration time to notice any spectral changes during laser exposure and to increase signal-to-noise ratio without mathematical smoothing. Wavenumber scales were precisely calibrated (± 0.1 cm^{-1}) using the emission spectra of a neon glow lamp taken before and after each Raman measurement. The Raman contribution from corresponding buffer was subtracted, and the spectra were corrected for non-Raman background.

σNS crystallization

Initial trials in previously promising conditions (as found by Alex Borodavka, *Borodavka, A., 2013, Understanding the dynamics of viral RNA genomes using single-molecule fluorescence, PhD thesis*) were performed using thrombin-cleaved (i.e. his-tag removed) σNS (2 mg/ml) and the mother liquor (ML) conditions as below:

Buffer	Precipitant	Salt
0.1 M Sodium cacodylate pH 6.5	5% PEG 8000	0.2 M $MgCl_2$
0.1 M Sodium cacodylate pH 6.5	5% PEG 3000	0.2 M $MgCl_2$
0.1 M Tris-HCl pH 6.5	5% PEG 8000	0.2 M $MgCl_2$
0.1 M HEPES pH 7.0, 5% Tacsimate pH 7.0	10% PEG MME 5000	-

Table 3.2: Pre-established crystallization conditions

Drops were set up using 4 µl : 1 µl, 2 µl : 1 µl and 1 : 1 µl σNS : ML in 24-well plates using the hanging drop vapour diffusion technique. Plates were incubated at 27ºC, 18ºC and 12ºC and were examined for signs of crystal growth every 2 days. Protein crystals appeared after a week at 18ºC and were subjected to diffraction screening using a home-source X-ray generator (Rigaku MicroMax-007 HF microfocus) with a Raxis4++ detector and Oxford Instruments liquid nitrogen cryostream. Further diffraction screening was performed at Diamond Light Source synchrotron (Rutherford Appleton Laboratory, Harwell, UK) on beamline IO4 at 100K after being cryoprotected in 25% glycerol.

Further crystallization trials of σNS were performed using SEC-purified his-tagged σNS apoprotein, thrombin-cleaved (i.e. untagged) σNS apoprotein, and thrombin-cleaved σNS : unstructured 20mer RNA (**Table 3.2**) complex at

2 – 5 mg/ml. Higher protein concentrations were avoided based upon previous unsuccessful crystal trials performed by Alex Borodavka. Trials were performed using the following commercial screens: Wizard 3 & 4 (Emerald BioSystems), Morpheus, MIDAS, PACT, JCSG I, II, III & IV (Molecular Dimensions), Hampton Index screen 1, and SaltRX (Hampton Research). 100, 200 and 400 nl σNS was mixed with 200 nl ML using a Formulatrix NT8 robot in 96-well 3-drop UV- transmissible MRC plates (Swissci) and incubated at 20C.

In situ data was collected at Diamond Light Source, beamline 103-1 with a 1 second exposure time at 100% transmission. Grid scans were performed, oversampling data (100 μm beam size and 50 μm grid size), ensuring that if there was any diffraction it would be seen. However, these crystals did not diffract.

Negative-stain electron microscopy and data processing

For negative-stain grid preparation, 4 μl of sample (at various concentrations ranging from 100 – 500 nM) was incubated on glow-discharged (using PELCO easiGlow) carbon-coated Formvar 300-mesh Cu grid (Agar scientific) for 90 seconds prior to blotting, and stained twice with 20 μl 2% uranyl acetate (first stain immediately blotted, second stain incubated for 20 seconds prior to blotting) and allowed to dry. Micrographs were collected on an FEI Tecnai 12 transmission electron microscope operated at 120 kV and equipped with a Gatan UltraScan 4000 CCD camera. Data processing was done in RELION 2.1. Initial particle picking of ~1,000 complexes was performed manually, and particles were extracted with 200^2 pixel box at a size of 3.74 Å per pixel and subjected to reference-free 2D classication to obtain 2D initial references for auto-picking within RELION. A total of 17261 and 20260 particles were picked for σNS apoprotein and 20mer-RNP complexes, respectively.

Further classification of the σNS apoprotein dataset was carried out in order to visualise low-abundance σNS dimers hypothesised to be present. Autopicked particles were re-extracted with a reduced box size (100^2 pixel) and subjected to 2D classification as described above.

	σNS filaments	**σNS filament ends**	**σNS**	**σNS RNP**	**NSP2-ΔC**
Nominal Magnification	30,000				
Voltage (kV)	120 kV				
Defocus range (μm)	~ -3				
Detector	Gatan US4000/SP				
Pixel size (Å)	3.74				
Symmetry imposed	C1				
Autopicked particles	35,138	8,876	31,407	33,980	14,740
Final particle images (number)	5,113	3,944	N/A		2,864
Map resolution (Å)	17	N/A	N/A		22
FSC threshold	0.143				

Table 3.3: Negative stain EM data collection and reconstruction parameters

Cryogenic electron microscopy (cryoEM) and data processing

CryoEM was performed exclusively with Quantifoil 1.2/1.3 holey carbon grids (i.e. regular ~1.2 μm circular holes with a regular spacing of ~1.3 μm), purchased from Quantifoil. All grids were glow discharged in air using GloQube glow discharge system (Quorum) immediately prior to use. All grids were prepared using a Vitrobot IV (FEI) at 100% humidity and 4°C, with a blotting time of 6 seconds and a nominal blotting force of 6. Samples flash-frozen in liquid nitrogen (LN_2)-cooled liquid ethane were immediately transferred to storage dewars under LN_2.

Vitrified samples were imaged at low temperature in-house (Astbury Biostructure Laboratory, University of Leeds), using Thermo Fisher Titan Krios microscopes equipped with either a Falcon III (NSP2 apoprotein) or a Gatan K2 (NSP2 RNP and σNS) detector. Data was collected with an acceleration voltage of 300 kV and a nominal magnification of 75,000x, resulting in pixel sizes of 1.065 Å per pixel (Falcon III) or 1.07 Å per pixel (K2). Data collection parameters are described in **Table 3.4**.

Image processing was carried out using the Relion 3 pipeline. Movie drift-correction was performed using MOTIONCOR2, and the contrast transfer function of each movie was determined using gCTF. Initial particle picking of a subset of 5 – 10 randomly chosen images was performed with the Laplacian-of-Gaussian (LoG) tool within the Autopicking module within Relion3. Particles were extracted and subjected to initial 2D classification in order to identify particles and assess autopicking success. Following this, the entire dataset was picked using LoG methods, extracted using 256 pixel box size and binned four times (effective box size 64 pixels) and subjected to 2D classification with fast subsets in order to remove false-positive particles that had been erroneously picked. Next, a more rigorous 2D classification was performed (without fast subsets). Particles originating from 2D classes with clear secondary structure features were selected and used to generate an initial model using SGD in Relion 3. Following several rounds of 3D and 2D classification, suitable particles were selected for 3D auto-refinement with various symmetry parameters applied (**Table 3.4**). Following refinement, per-particle CTF and Bayesian Polishing were performed in Relion 3, and 'shiny' particles were re-refined. Post-processing was performed with a soft mask and B-factor estimated automatically in Relion 3.

After particle polishing, NSP2 apoprotein and RNP complex were subjected to further 3D classification into three classes without particle orientations. This yielded three similarly-sized subsets of near-identical particles for NSP2 apoprotein whose resolution did not improve upon the original, larger dataset following 3D auto-refinement and post-processing. For NSP2 RNP complex, this gave a single class with 86% of particles, and two other classes with 6% and 8% of the particles. The class with 6% of input particles had surprisingly well-defined protein and RNA densities and was used for 3D auto-refinement and post-processing. This improved map resolution from 3.5 Å to 3.4 Å with C4 symmetry. A D4 symmetry reconstruction further increased the map resolution from 3.5 Å to 3.2 Å.

For NSP2 RNP complex, symmetry expansion was performed using *relion_particle_symmetry_expand*, generating four symmetry-related orientations for each particle. A mask covering a single basic groove of NSP2 was made using the volume eraser tool in UCSF Chimera and Relion 3, with a soft edge of 15 pixels. The symmetry-expanded dataset was then subjected to focussed classification into 10 classes using this mask, without particle orientations.

	NSP2 apoprotein	**NSP2 RNP**		**σNS**
Nominal Magnification	75,000			
Voltage (kV)	300 kV			
Defocus range (μm)	-1.2 to -2.4	-1.5 to -3.5		
Detector	Falcon III	Gatan K2		
Pixel size (Å)	1.065	1.07		
Electron exposure (e^-/Å^2)	110	55		56
Symmetry imposed	D4	C4	D4	C2
Particles in first 3D classification	955,893	1,245,861		376,032
Final particle images (number)	305,164	635,599	38,252	154,617
Map resolution (Å) (unmasked/masked)	4.4/4.1	3.6/3.5	3.6/3.2	7.2/6.7
FSC threshold	0.143			
Map resolution range (Å)	3.75 – 4.75	3.2 – 4.1	3.1 – 4.1	5.9 – 7.9 Å

Table 3.4: CryoEM data collection parameters

Filament assembly and characterization

10 – 15 μM σNS was incubated with varying RNA substrates maintaining a consistent stoichiometry of 5 σNS monomer : 2 RNA at room temperature prior to imaging using negative stain electron microscopy. For making grids, samples underwent two serial 10-fold dilutions (final σNS concentration 100 – 150 nM) to ensure well dispersed particles. Micrographs were collected at 30,000x magnification.

Filaments assembled using different RNA substrates (unstructured 40mer, S5f, S11f, full-length RV S11) were characterised using semi-automated tracking procedures implemented in the Matlab-based suite *FiberApp* (Usov and Mezzenga, 2015).

Native mass spectrometry

σNS was dialysed into 200 mM ammonium acetate, pH 7.6 overnight at 4°C. 20mer-RNP complexes were assembled as described above. σNS apoprotein and 20mer-RNP complex were diluted to a final protein concentration of 20 μM. NanoESI–IMS–MS spectra were acquired with a Synapt HDMS mass spectrometer (Waters) with platinum/gold-plated borosilicate capillaries prepared in-house. Typical instrument parameters were: capillary voltage, 1.2–1.6 kV; cone voltage, 40 V; trap collision voltage, 6 V; transfer collision voltage, 10 V; trap DC bias, 20 V; backing pressure, 4.5 mbar; IMS gas pressure, 0.5 mbar; traveling wave height, 7 V; and traveling wave velocity, 250 ms^{-1}. Data were processed with MassLynx v4.1, Driftscope 2.5 (Waters) and Massign (Morgner and Robinson, 2012). Collisional Cross-Sections (CCSs) were estimated through a calibration (Ruotolo *et al.*, 2008; Smith *et al.*, 2009; Bush *et al.*, 2010) using arrival-time data for ions with known CCSs (β-lactoglobulin A, avidin, concanavilin A and yeast alcohol dehydrogenase, all from Sigma Aldrich).

The CCS values corresponding to the lowest charge state (i.e. the greatest m/z value) of each oligomer were selected for comparison with SAXS-derived CCS estimates, as these species will be the least affected by Coulombic repulsion (Ruotolo and Robinson, 2006).

Estimated CCSs for the 20 independently generated dummy atom models (of both σNS apoprotein and 20mer-RNP complex) were generated using IMPACT (Marklund *et al.*, 2015), and were in good agreement (standard deviation <1%).

Mapping protein-RNA interactions in vitro using RBDmap

10 *μ*M σNS or NSP2 was incubated with 5'-A_{25}-S11 RNA, in a final volume of 100 *μ*l. RNP complexes were incubated at room temperature for 30 minutes and applied to a single well of a 24-well plate. This 24-well plate was placed on an aluminium block cooled to 4°C within a plastic container of ice and subjected to 6 rounds of UV irradiation (254 nm, 0.83 J cm^{-2} per round) in a UVP CL-1000 Ultraviolet Crosslinker (Scientifix). Crosslinked RNP complexes were digested by LysC (NEB, #P8109S) (500 ng per crosslinked RNP sample) overnight at room temperature.

30 *μ*l of input sample was set aside at 4°C until RNase digestion of all samples as a negative control, to be analysed in parallel with crosslinked samples. The remaining sample was incubated with 1 ml of oligo $d(T)_{25}$ magnetic beads (NEB, # S1416S) that were pre-washed and pre-incubated with 1 ml buffer. After sample application, beads were incubated at room temperature for 30 minutes, then 1 hour at 4°C prior to capture using a magnet. Supernatant was carefully removed, and beads were resuspended with 17 ml of buffer 1 (20 mM Tris-HCl, pH 7.5, 500 mM LiCl, 0.5% lithium dodecyl sulfate (LDS) (w/v) (Sigma Aldrich, #L9781), 1 mM EDTA, 5 mM DTT) with gentle agitation for 10 minutes at 4°C, prior to a second magnet capture step. Two wash/capture steps were performed using buffer 1, followed by another two less stringent

wash/capture steps using buffer 2 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 0.1% LDS, 1 mM EDTA, 5 mM DTT), and two final capture/wash steps with buffer 3 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA, 5 mM DTT). After a final magnet capture, supernatant was removed, and beads were resuspended in 300 *μ*l buffer 3. RNA and cross-linked peptides were eluted from magnetic beads by heating to 60°C for 2 minutes and a subsequent round of magnet capture, allowing the eluate (i.e. supernatant) to be removed.

RNAse A was added to samples and incubated at 37°C for 1 hour, followed by reduction using 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) at pH 8, for 30 minutes at 65°C. Samples were alkylated using 40 mM 2-chloroacetamide (CAA, Sigma, #C0267) at room temperature for 30 minutes in the dark. Samples were then adjusted to pH 8.0 using 1 M Tris-HCl and incubated with trypsin and incubated overnight at 37°C and shaken at 200 r.p.m. Samples were supplemented with additional trypsin and incubated for a further 2 hours as described above. Digestion was quenched by addition of formic acid to pH 3. Tryptic peptides were purified using OMIX C18 pipette tips (Agilent Technologies). Samples were dried using a centrifugal vacuum concentrator prior to storage at -80°C.

Samples were resuspended in 20 *μ*l of 0.1% formic acid and sonicated in a water bath for 10 minutes (Grant Instruments XUBA3 Analog Ultrasonic Bath) and centrifuged for 10 5 minutes at 21,000*g* and subjected to tandem mass spectrometry analysis.

Peptides were analysed by LC-MS/MS using an Orbitrap Fusion Tribrid machine connected to an UltiMate 3000 UHPLC liquid chromatography system (Thermo-Fisher Scientific). Data were collected in positive mode using data-dependent acquisition. Precursor scan covered a range of 375-2000 m/z at a resolution of 120,000 and was followed by up to 12 subsequent MS/MS scans measured at a resolution of 60,000. High energy collisional dissociation was used to fragment precursor ions with a charge state of 2-7 using a collision energy of 32%.

Resulting Thermo Scientific mass spectrometer binary data files were analysed using the Andromeda search engine within the MaxQuant software package. Identified tryptic peptides were further analysed using R scripts (crisscrosslinker) (Zhang *et al.*, 2019).

Mapping protein-protein interactions using chemical crosslinking

10 μM σNS was incubated either alone or in complex with stoichiometric excess (5 σNS monomer : 2 RNA) of either unstructured 20mer or 40mer RNA (**Table 3.2**) with 20 μg Bis(sulfosuccinimidyl)suberate (BS3) in 20 mM HEPES pH 7.5, 150 mM NaCl buffer. Reactions were carried out for 15 minutes at room temperature before quenching with 1 M Tris-HCl pH 8.0 (final concentration 50 mM). Peptides were reduced, alkylated and trypsinised as described for RBDmap peptides. Mass spectrometric identification of peptides was performed as described above. BS3-linked peptides were identified using pLink (Purcell *et al.*, 2007), and two-dimensional representations of σNS crosslinks were generated using xiNET (Combe et al., 2015).

Chapter 4: Stability of RNA secondary structures determines the selectivity of viral RNA chaperones

4.1 Introduction

Members of the Reoviridae family of double-stranded RNA viruses have genomes distributed between nine to twelve (depending on the virus species) distinct RNA segments, all of which are essential for virus replication and assembly. This provides a mechanistic conundrum – how does each assembling virion select and package a complete cohort of RNA segments? As it is improbable that a complete set of genomic RNA segments will be packaged through random incorporation, these viruses must employ specific mechanisms to achieve high-fidelity, stoichiometric genome segment assembly.

This is thought to occur through a series of inter-segment, sequence-specific RNA-RNA interactions. As RNA molecules can adopt many different 3D conformations, not all of which are capable of participating in stable intermolecular contacts, virally-encoded proteins with RNA chaperone-like activities are required to mediate these interactions.

Two of the most well-characterised examples of such proteins are rotavirus (RV) NSP2 and reovirus σNS. Both proteins have similar RNA chaperone like activities that implicate them as matchmakers of RNA-RNA interactions:

i. Binding single-stranded (ss)RNA with low-nM affinity, but no sequence specificity
ii. Multivalent RNA binding

iii. Destabilising RNA structure and disrupting RNA helices

iv. RNA annealing

These activities allow NSP2 and σNS to promote specific, stable RNA-RNA interactions between complementary sequences that may otherwise be sequestered by metastable RNA structures and therefore not spontaneously anneal.

Despite these apparent similarities, NSP2 and σNS do not share any sequence homology. They also differ in oligomeric state – NSP2 is a stable octamer in solution (in both the presence and absence of RNA), while σNS apoprotein is hexameric. The stoichiometry of the σNS ribonucleoprotein (RNP) complex is unknown but has been demonstrated to contain multiple RNA molecules.

This raises several questions: do NSP2 and σNS employ similar mechanisms to promote RNA-RNA interactions, if so, are they functionally interchangeable, and ultimately how do these non-specific ssRNA-binding proteins achieve selectivity?

The work presented in this chapter describes the expression and purification of σNS and NSP2. The abilities of NSP2 and σNS to promote interactions between full-length viral RNA segments were compared, revealing that these functionally analogous proteins are not interchangeable. The mechanistic differences between NSP2 and σNS were further explored.

This chapter provides functional insights into the surprisingly diverse modes of RNA chaperone activity of NSP2 and σNS, and implicate the stability of RNA structure as a key regulator of RNA chaperone activity.

4.2 Results and Discussion

4.2.1 Expression and purification of σNS and NSP2

Details of expression and purification of σNS and NSP2 are given in Chapter 3 (**Materials and Methods**). σNS expression was optimised by lowering the growth temperature after induction from 21°C to 18°C, and by increasing the length of induction from 12 hours to 18 – 20 hours. This resulted in a yield of ~15 mg σNS per litre of *E. coli* culture (after the initial ion exchange chromatography (IEX) step) (**Figure 4.1**).

Hexahistidine-tagged σNS was first purified through immobilised metal affinity chromatography (IMAC), yielding several fractions of relatively-high purity, as monitored by SDS-PAGE (**Figure 4.1**). Purified fractions were diluted (to reduce the salt concentration) and subjected to IEX. IEX fractions were subjected to overnight thrombin cleavage to facilitate his-tag removal, and any uncleaved σNS (i.e. still bearing his-tag) was removed by IMAC. Thrombin-cleaved σNS was further purified by IEX chromatography to remove any minor protein or nucleic acid contaminants and to re-concentrate σNS prior to size-exclusion chromatography (SEC) (**Figure 4.1**).

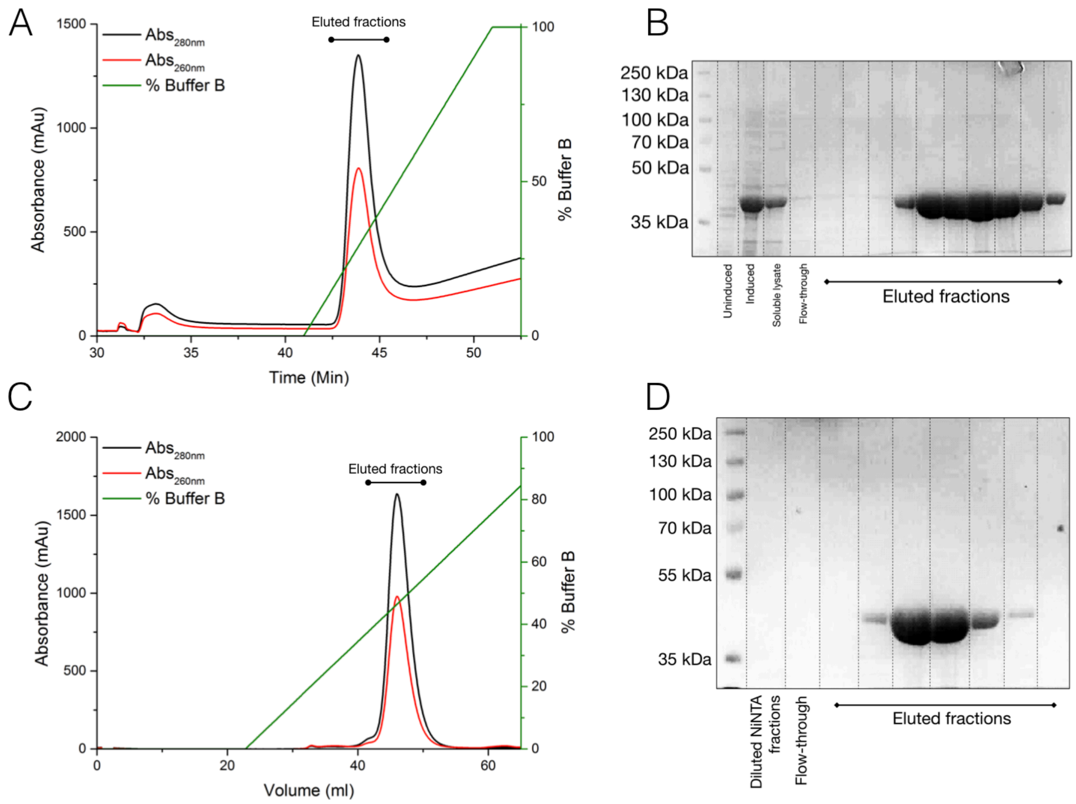


Figure 4.1 Purification of recombinant σNS (A & B) Immobilised metal affinity chromatography (IMAC) purification of σNS. Chromatogram of σNS elution from Ni-NTA column (A) and SDS-PAGE of fractions visualised by staining with Instant Blue, with Molecular weight markers indicated (B). Immobilised σNS was eluted across a gradient of Buffer B (containing 1 M imidazole). Left y axis shows absorbance at 260 nm (red) and 280 nm (black), while right y axis (green) shows % Buffer B (green). (C & D) Ion exchange chromatography (IEX) of peak fractions containing σNS eluted from Ni-NTA column. σNS was loaded onto 5 ml Sepharose Q column, and eluted across a gradient of Buffer B (containing 1 M NaCl) (C). Corresponding fractions visualised on SDS-PAGE gel stained with Instant Blue.

Fractions were subjected to a final 'polishing' step of size-exclusion chromatography (SEC). The elution profile of σNS from a Superdex 200 10/300 increase GL column showed a single, symmetrical peak (**Figure 4.2**) with an A_{260nm}/A_{280nm} ratio of 0.57, indicative of highly purified σNS, lacking nucleic acid contamination. The elution volume of 11.7 ml indicates that σNS forms a large oligomer, as previously described (Borodavka *et al.*, 2015). The purity of σNS was further corroborated by SDS-PAGE analysis. A ~40 kDa band (expected MW = 40.9 kDa) lacking any visible contamination (**Figure 4.2**).

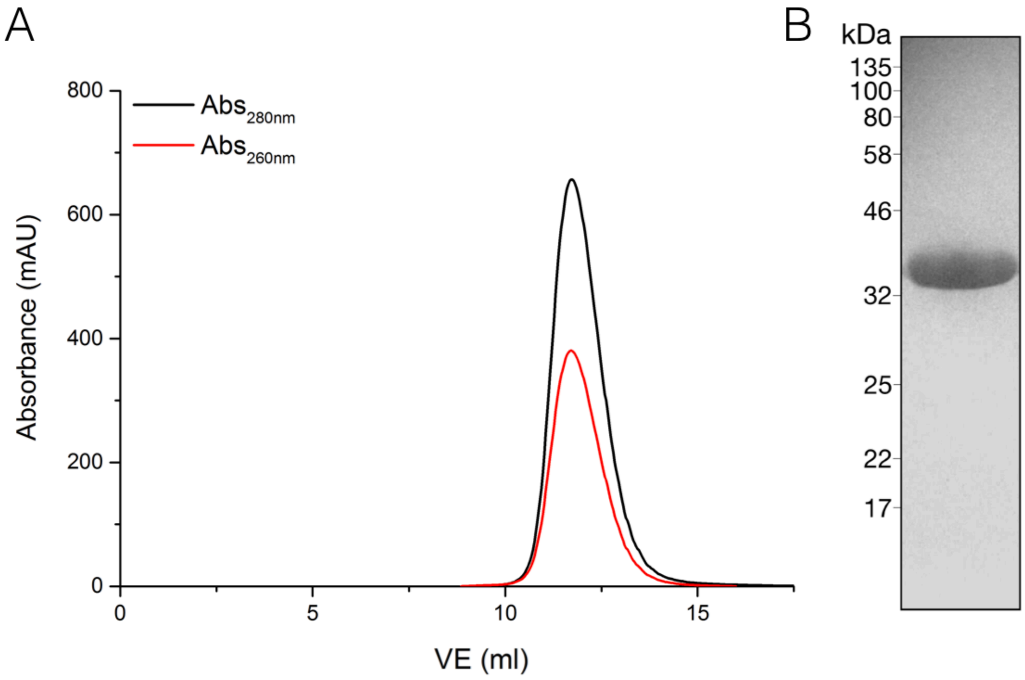


Figure 4.2 Size exclusion chromatography (SEC) of σNS (A) SEC chromatogram of thrombin-cleaved σNS loaded onto Superdex 200 increase column. As 260/280 ratio of peak is ~0.57, purified σNS is free from nucleic acid contamination. (B) Peak fraction of SEC elution visualised on SDS-PAGE gel stained with Instant Blue.

Hexahistidine-tagged NSP2 was purified in a similar manner to σNS, but without his-tag cleavage. Although all attempts to cleave NSP2 affinity tags were unsuccessful, previous studies have indicated that his-tagged NSP2 is still functional (Borodavka *et al.*, 2017). IMAC and IEX purification of NSP2 (**Figure 4.3**) alongside SDS-PAGE analysis indicate that large-scale expression of NSP2 provides lower yields than that of σNS (2 – 5 mg of NSP2 per litre of cell culture). Fractions eluted from IEX chromatography were subjected to SEC (**Figure 4.4**), and a single, symmetrical peak was observed with an elution volume of 11.6 ml, with an A_{260nm}/A_{280nm} ratio of 0.59, confirming a lack of nucleic acid contamination. SDS-PAGE analysis showed a single band at ~35 kDa (expected MW = 36.7 kDa), confirming that NSP2 fractions were highly pure. The similar elution volumes of σNS and NSP2 (11.6 and 11.7, respectively) indicate that both proteins assemble oligomers, as is expected (Schuck *et al.*, 2001; Borodavka *et al.*, 2015).

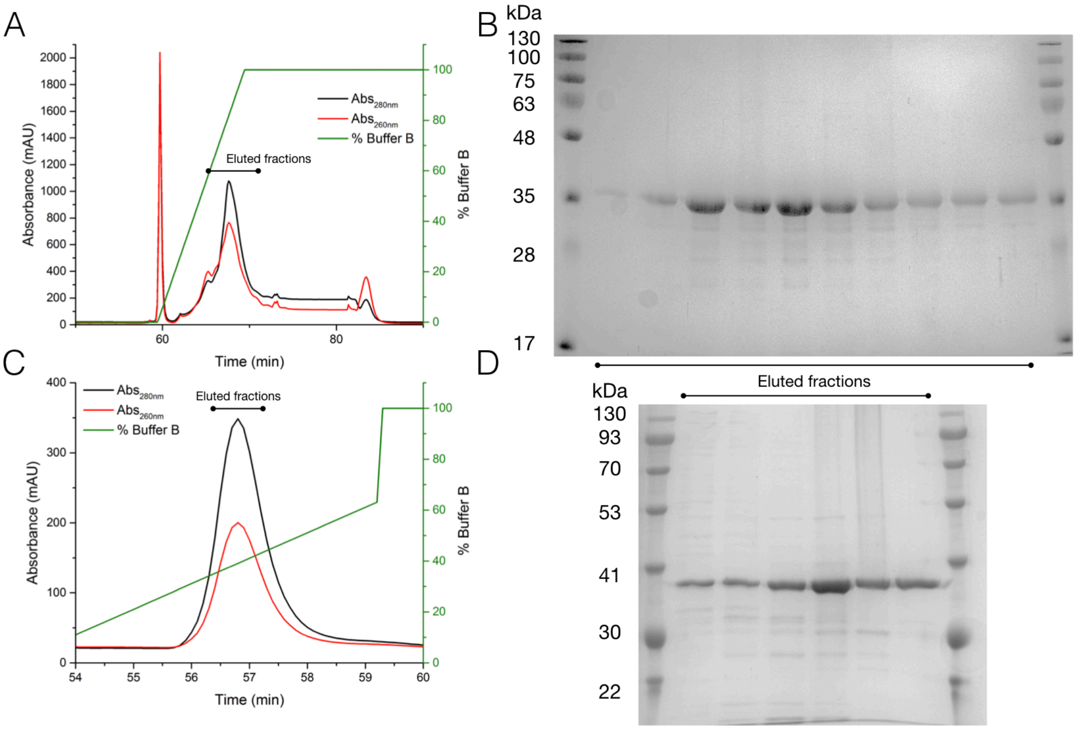


Figure 4.3 IMAC & IEX of NSP2 (A & B) Immobilised metal affinity chromatography (IMAC) purification of NSP2. Chromatogram of NSP2 elution from Ni-NTA column (A) and SDS-PAGE of fractions visualised by staining with Instant Blue, with Molecular weight markers indicated (B). Immobilised NSP2 was eluted across a gradient of Buffer B (containing 1 M imidazole). Left y axis shows absorbance at 260 nm (red) and 280 nm (black), while right y axis (green) shows % Buffer B (green). A sharp peak eluted at ~2% Buffer B, likely corresponding to non-specifically associated nucleic acids (due to the 260/280 ratio of >1.5. (C & D) Ion exchange chromatography (IEX) of peak fractions containing NSP2 eluted from Ni-NTA column. NSP2 was loaded onto 5 ml Sepharose SP column, and eluted across a gradient of Buffer B (containing 1 M NaCl) (C). Corresponding fractions visualised on SDS-PAGE gel stained with Instant Blue.

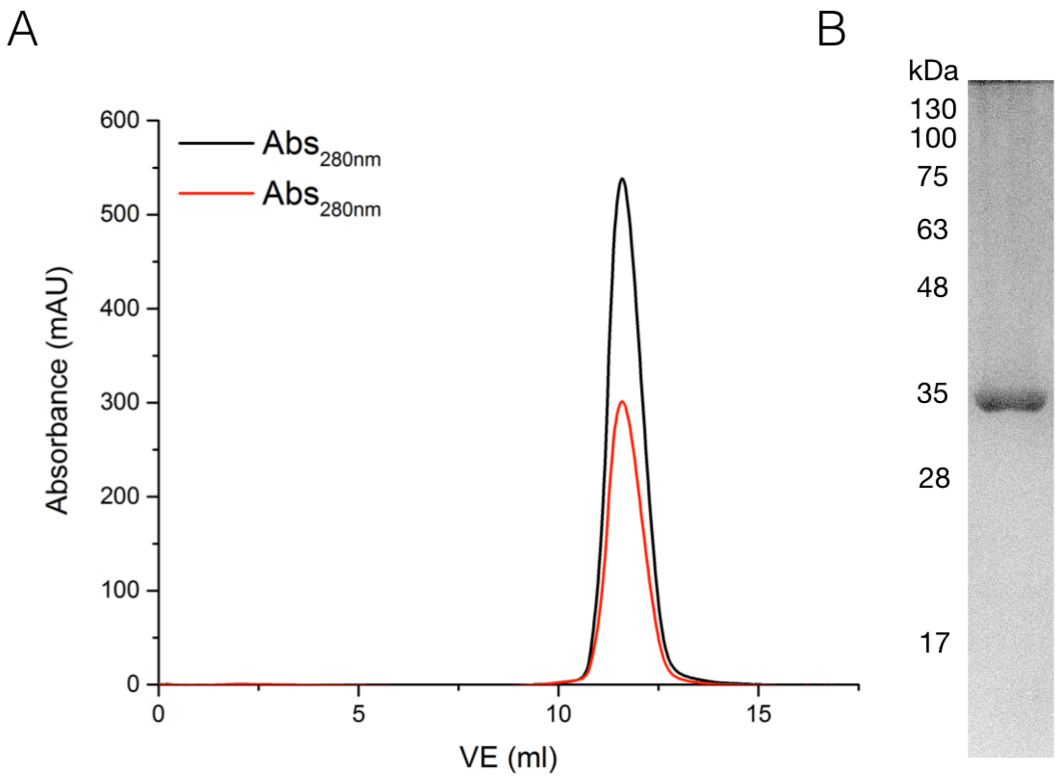


Figure 4.4 SEC of NSP2 (A) SEC chromatogram of NSP2 loaded onto Superdex 200 increase column. As 260/280 ratio of peak is ~0.57, purified NSP2 is free from nucleic acid contamination. (B) Peak fraction of SEC elution visualised on SDS-PAGE gel stained with Instant Blue.

The SEC step yielded σNS and NSP2 fractions of high purity (**Figure 4.5**), with MW of 40.7 kDa and 33.7 kDa for σNS and NSP2, respectively. Both proteins were subsequently characterised by circular dichroism (CD) spectroscopy. The CD spectra of either protein indicates that both σNS and NSP2 are folded, but have different secondary structure components (**Figure 4.5**). In agreement with previous studies, both proteins have an α/β fold (Schuck *et al.*, 2001; Borodavka *et al.*, 2015), with σNS having a high α-helical content, while NSP2 has a higher β-sheet content.

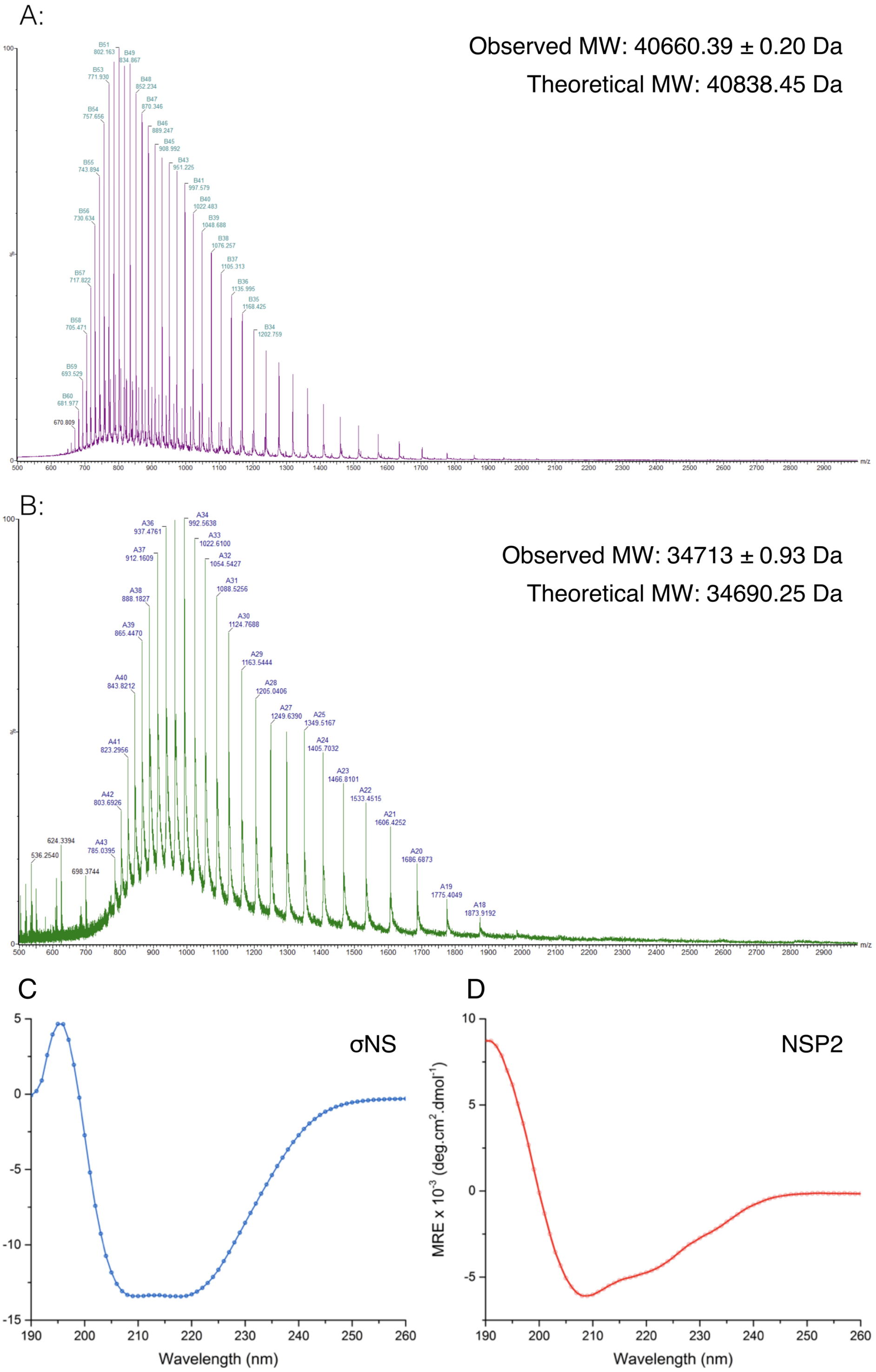


Figure 4.5 Quality control of σNS and NSP2: A & B: Denaturing ESI-MS of purified σNS (A) and NSP2 (B). (C & D) Circular dichroism (CD) spectra of σNS (C, blue) and NSP2 (D, red), indicating that recombinant purified proteins are folded, but have different secondary structure components. This is in agreement with previous studies (Schuck *et al.*, 2001; Borodavka *et al.*, 2015).

4.2.2 σNS is unable to promote inter-segment interactions between rotavirus RNA segments

We first sought to compare the abilities of NSP2 and σNS to mediate intersegment RNA-RNA interactions between full-length viral RNAs. Previous studies have relied on native gel electrophoresis to visualise such interactions, but this is not without several caveats. Firstly, if an RNA-RNA interactions is observed (through a retardation or gel-shift of an RNA band), it remains unclear whether this is arising from an interaction between two *different* RNAs, or whether this simply an artefact of non-specific RNA-RNA duplexes (either as homo-dimers or metastable hetero-dimers) that naturally occur spontaneously through Watson-Crick interactions between RNAs at high concentrations. Instead, a method of detecting RNAs that satisfies the following criteria is essential:

- High sensitivity: Can detect RNA-RNA interactions at low-nM RNA concentrations (i.e. concentrations at which non-specific RNA-RNA interactions are largely avoided).
- Capable of distinguishing between two different RNAs (and therefore homo- and hetero-RNA interactions).

Accordingly, we used a two-colour fluorescence cross-correlation spectroscopy (FCCS)-based assay to interrogate protein-mediated RNA-RNA interactions between full-length viral RNA segments. FCCS measures the correlated co-diffusion of multiple molecules labelled with different fluorophores (here, two different viral RNAs) to detect and quantify stable interactions. If these RNAs interact, they will co-diffuse and therefore result in a non-zero cross-correlation function (CCF) amplitude.

Based on these assumptions, we studied interactions between rotavirus (RV) ssRNA segments S5 and S11, as these have been previously demonstrated to participate in a sequence-specific RNA-RNA interaction facilitated by NSP2. As such interactions between avian reovirus (ARV) genome segments are as of yet unidentified, we used the S5 : S11 interaction as a model to compare the RNA-RNA matchmaking activities of NSP2 and σNS. This also allowed us to test if σNS can substitute for NSP2 in vitro.

In the absence of either protein, S5 and S11 do not spontaneously interact (**Figure 4.6A**, magenta). Co-incubation with their cognate RNA chaperone protein, NSP2, was able to induce a stable interaction as indicated by the increased CCF amplitude (**Figure 4.6**, blue). Co-incubation with σNS, however, was not able to produce a stable S5 : S11 interaction (**Figure 4.6A, black**).

Given that σNS promotes duplex formation between short oligonucleotides, its failure to promote interactions between highly structured RNA segments may be due to the sequestration of complementary sequences within RNA structure. To investigate this, we analysed protein-mediated strand-annealing between full-length S5 and a short, S11-derived complementary RNA (S11f). While co-incubation with NSP2 stabilises this interaction, σNS was unable to promote S5 : S11f (**Figure 4.6B**). Co-incubation of S5 and S11f in the absence of either protein results in a CCF amplitude of ~0.6. This is likely due to the removal of secondary and tertiary structure elements present within S11 that may sequester the sequence that interacts with S5. By presenting the complementary sequence within a short, unstructured RNA (S11f), a higher yield of spontaneous interactions with S5 occurs (**Figure 4.6B**).

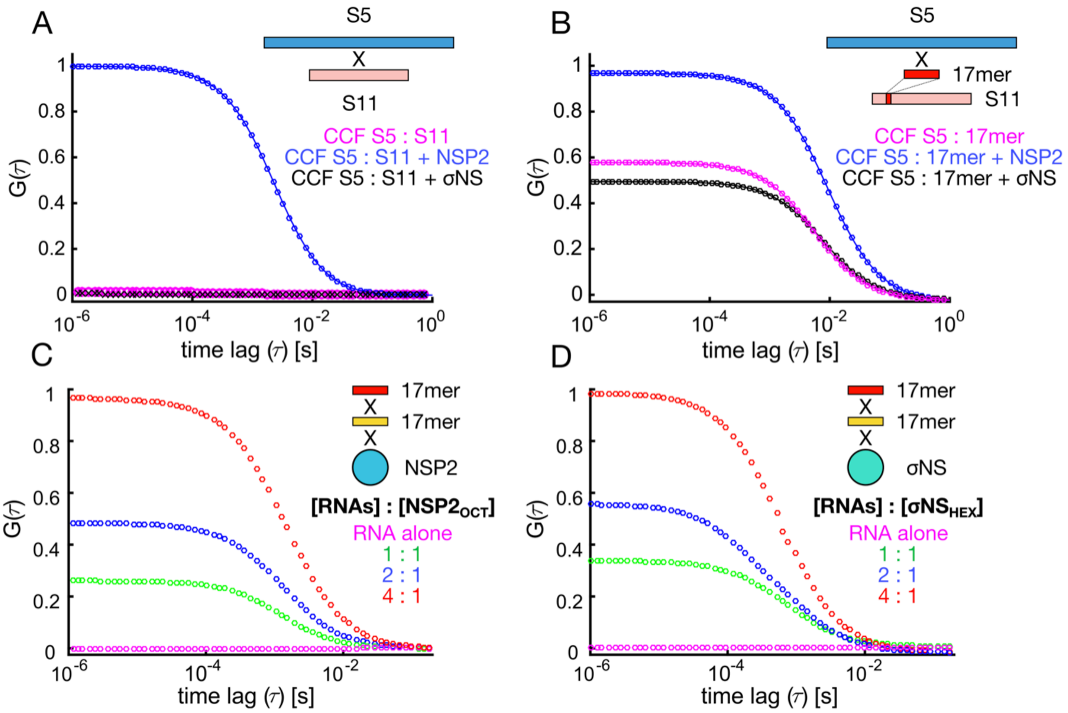


Figure 4.6 Probing intermolecular RNA-RNA interactions using fluorescence cross-correlation spectroscopy (FCCS) A: Interactions between rotavirus (RV) segments S5 and S11 co-incubated in the absence (magenta) or presence of either NSP2 (blue) or σNS (black). B: Interactions between RV S5 and a 17-nt RNA derived from S11 (complementary to a region within S5). RNAs were co-incubated either alone (magenta), or in the presence of NSP2 (blue) or σNS (black). C & D: Both NSP2 and σNS oligomers bind to multiple RNAs concurrently. Non-complementary 17mer RNAs labelled with Cy3 and Cy5 were co-incubated alone (magenta) and in the presence of NSP2 (C) and σNS (D) at increasing RNA : protein ratios. The RNAs alone (magenta) do not interact (hence zero CCF amplitude).

As concurrent binding of multiple RNAs to the same protein oligomer is required for efficient RNA-RNA matchmaking (Rajkowitsch *et al.*, 2007; Woodson, Panja and Santiago-Frangos, 2018), we investigated whether NSP2 and σNS can bind multiple RNAs in solution. We used short, 17-nt non-complementary RNAs with different fluorescent labels and co-incubated them either alone or in the presence of either protein (**Figure 4.6C & D**). As expected, these RNAs do not spontaneously interact. Co-incubation with

NSP2 or σNS at increasing RNA : protein oligomer ratios increases the CCF amplitude, indicating that both NSP2 and σNS can bind multiple unstructured RNAs.

It is therefore a plausible hypothesis that the inability of σNS to promote interactions may be due to an inability to remodel RNAs. RNA remodelling is required for increasing the accessibility of complementary sequences within RNA segments.

4.2.3 σNS undergoes RNA-driven oligomerisation

Having determined that both NSP2 and σNS can bind to multiple RNAs concurrently in solution, we further investigated the oligomeric state of the σNS RNP complex. NSP2 assembles into stable octamers in solution (either as an apoprotein or in complex with RNA) (Schuck *et al.*, 2001). Although σNS apoprotein is thought to be hexameric, the stoichiometry of σNS RNP complexes is unknown.

Unstructured, 20mer RNA was used to assemble σNS-RNP complexes, as this corresponds to the previously-established RNA-binding footprint of σNS (Gillian *et al.*, 2000; Borodavka *et al.*, 2015). A molar ratio of 2 RNA : 5 σNS monomers was employed to assemble RNP complexes, as this should ensure slight stoichiometric excess of RNA bound to σNS (assuming that it binds at least two RNAs per hexamer). In the case of assembly of a higher order oligomer (e.g. octamer with 2 RNAs bound), this stoichiometry still maintains saturation of σNS with RNA. 70μM σNS was incubated with 28 μM RNA and analysed by SEC-DLS. Surprisingly, the RNP complex (henceforth termed ‘RNP’) eluted earlier than the hexameric apoprotein, suggesting that a larger species has assembled. The RNP elution peak has an A_{260nm}/A_{280nm} ratio of 1.4, confirming the presence of RNA. A small peak also occurs at 20 ml (consistent with a small hydrodynamic size of a protein-free 20mer RNA), corresponding to unbound RNA (A_{260nm}/A_{280nm} ratio ~2.0). The observed change in size of the RNP complex is corroborated by the increase in DLS-derived hydrodynamic radius (~8 nm, compared to ~5 nm for σNS apoprotein) (**Figure 4.7**). As an increase of this magnitude cannot be explained by the binding of RNAs alone, indicating that upon RNP complex formation σNS may undergo a change in conformation or oligomeric state.

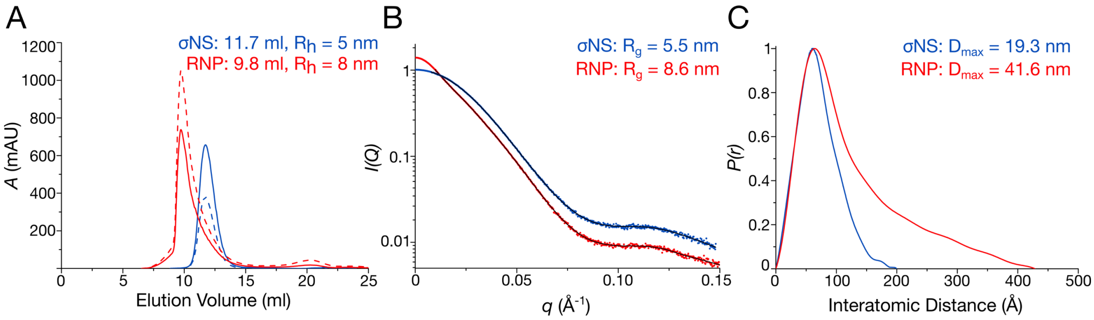


Figure 4.7 RNP complex formation produces a larger σNS oligomer A: SEC chromatograms of σNS (blue) and σNS-RNP complex (red). Continuous and dashed lines correspond to 280 and 260 nm absorbances, respectively. Elution volumes and DLS-derived hydrodynamic radii of either complex is indicated. B: Experimental small-angle X-ray Scattering (SAXS) scattering curves of σNS (blue) and RNP (red) and corresponding fits (black). Radii of gyration (R_g) of either species are indicated. C: Normalised pairwise distance distribution function (PDDF) plot, as calculated from experimental SAXS scattering curves (B). Maximum dimension of the particle (D_{max}) is indicated.

This was further investigated using small-angle X-ray scattering (SAXS). We observed an increase in radius of gyration (R_g) upon RNP complex formation, from 5.5 ± 0.03 nm to 7.6 ± 0.05 nm (**Figure 4.7, Table 4.1**). This increase in particle size and shape is further exemplified by the pairwise-distance distribution (PDDF, or *P(r)*) plot) (**Figure 4.7**). This is the real-space representation of the scattering curve, and describes the distribution of the paired-set of all distances between all points within an object; in the case of a SAXS experiment, it describes all pairwise electron distances within the macromolecule. RNP complex formation results in a 2-fold increase in the maximum distance (D_{max}) dimension of σNS, from 19.9 nm to 41.6 nm, and the shape of the PDDF transitions from a bell-shaped curve (indicative of a globular particle), to a curve with a long shoulder extending from its maxima (61 nm). This indicates that in addition to undergoing a change in particle size (as evidenced by the increase in R_h, R_g and Porod volume (**Table 4.1**)), RNA binding by σNS also induces a change in particle shape (**Figure 4.7**).

Guinier analysis of SAXS data indicates that both σNS and RNP samples are monodisperse in solution (as both σNS and RNP scattering curves are linear in the region of the Guinier plot closest to the zero scattering angle), indicating that the observed increase in particle size is not due to partial σNS aggregation upon RNA binding and resulting heterogeneity (**Figure 4.8**). Dimensionless Kratky analysis indicates that while σNS apoprotein is globular, the RNP complex adopts a flexible, extended structure as is indicated by the shift of its apex towards a higher momentum transfer (*q*) (**Figure 4.8**). These data suggest that σNS undergoes RNA-driven oligomerisation upon RNP complex formation.

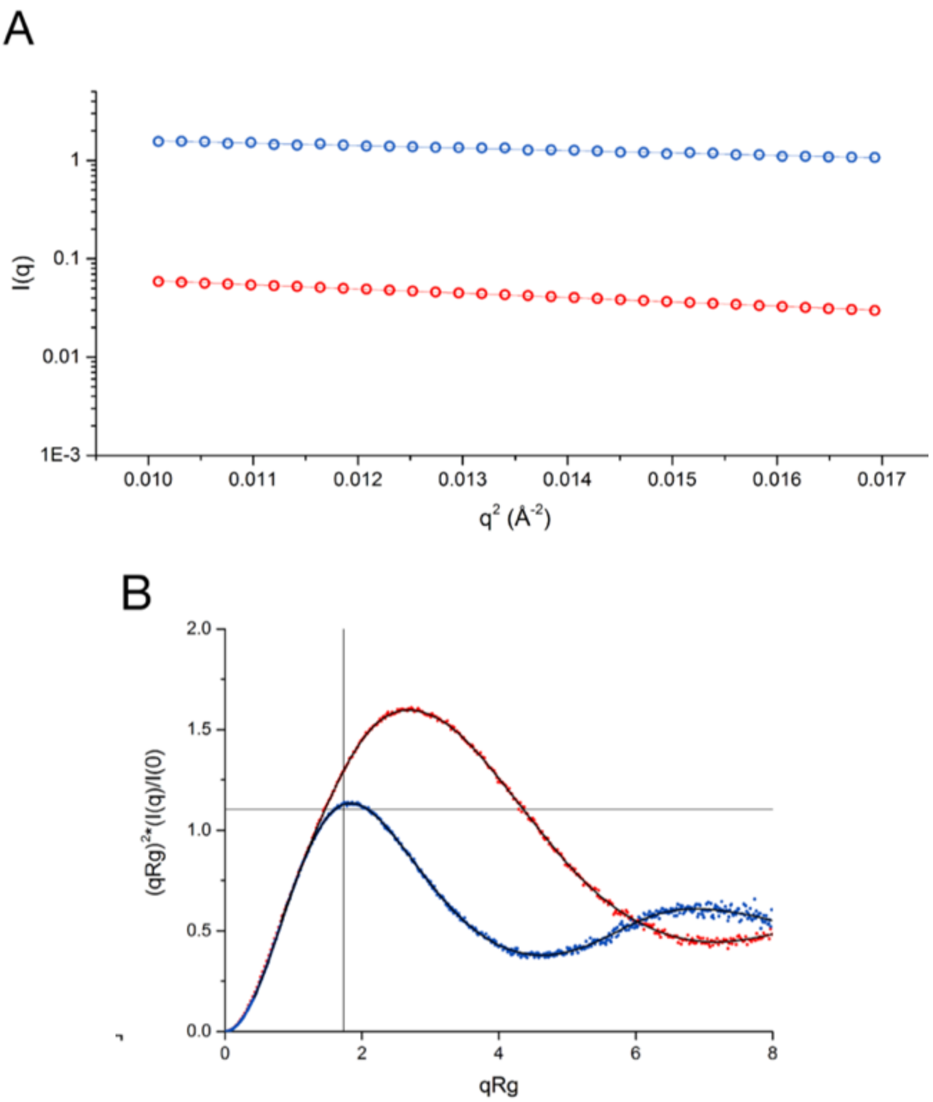


Figure 4.8 Further SAXS analysis of σNS and RNP complex A: Guiner plot of σNS (blue) and RNP (red) complex. A linear relationship between momentum transfer squared (q^2) and scattered intensity (I(q)) at small angles is indicative of particle monodispersity. These Guinier plots indicate that both σNS and NSP2 are monodisperse. B: Dimensionless Kratky plots of σNS (blue) and RNP (red). Guinier-Kratky point (1.732, 1.104) is indicated by a crosshair. Although σNS apoprotein exhibits a peak at this point (and thus is globular), the upward-rightward shift of the RNP complex curve indicates that it is highly extended, although the decrease exhibited after its apex is indicative of a folded particle.

To obtain more specific information on the solution structures of σNS and RNP, we generated ab initio models from the SAXS data using DAMMIN. Reconstructed dummy atom models for both σNS apoprotein and RNP complex were aligned (**Figure 4.9A & B**) and averaged, producing an averaged envelope (from DAMAVER) and a filtered envelope (from DAMFILT)

(**Figure 4.9**). The low normalised special discrepancy (NSD) values for these two models (0.829 ± 0.05 and 0.576 ± 0.03 for σNS and RNP, respectively. NSD values <1 are indicative of highly self-consistent models) (**Table 4.1**). The SAXS envelop of σNS is globular and compact, while the envelope of RNP is elongated, with a thin 'tail' separated from the more compact, globular 'body' (**Figure 4.9**).

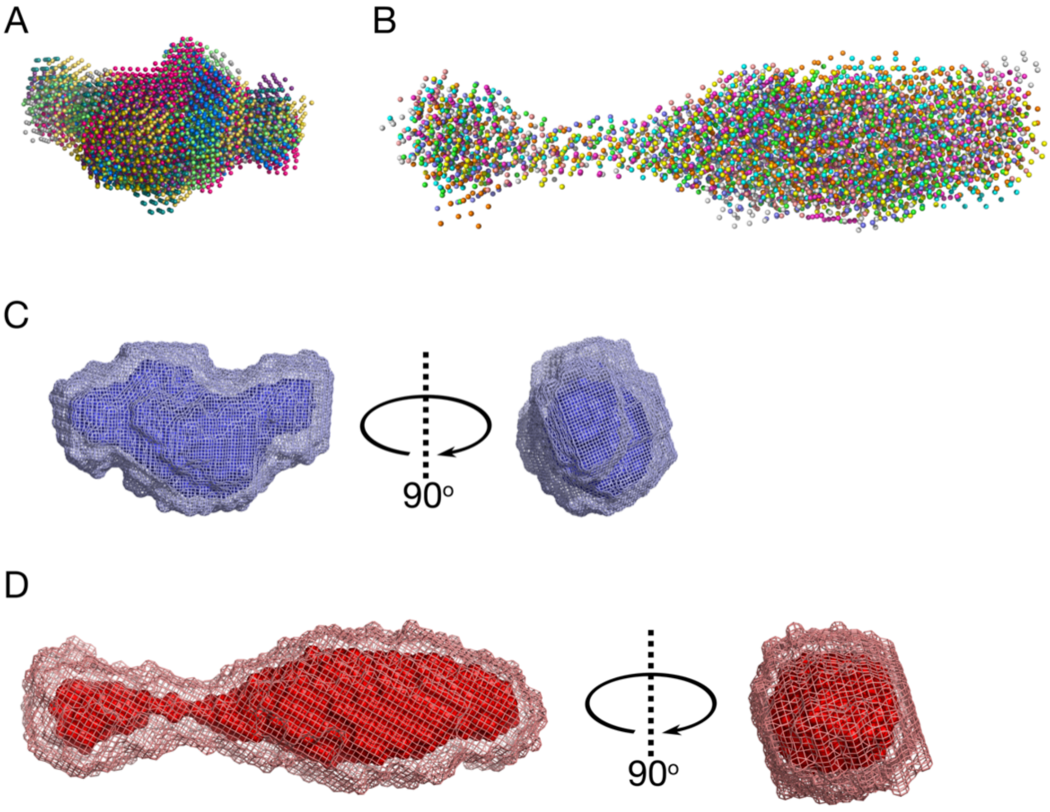


Figure 4.9 Ab initio modelling of σNS and RNP A & B: DAMMIN-derived ab initio models of σNS (A) and RNP (B). 10 best dummy atom models (i.e. those with highest correlation between predicted and experimental scattering curves) for either dataset was overlaid. C & D: DAMFILT and DAMAVER models of σNS (C) and RNP (D). While DAMAVER represents the ensemble-averaged dummy atom model, DAMFILT indicates the similarity between each set of ten models (normalised spatial discrepancy (NSD) of 0.82 ± 0.05 for σNS and 0.58 ± 0.03 for RNP). DAMFILT NSD values <1 suggest that each set of dummy atom models was highly similar (Bingaman *et al.*, 2017). SASBDB (https://www.sasbdb.org/) ID numbers: SASDDT5 (RNP), SASDDU5 (σNS apoprotein).

	σNS apoprotein	**σNS-20mer (RNP)**	**NSP2 apoprotein**
Data collection parameters			
Beamline	B21, Diamond Light Source		
Wavelength	1 Å		
***q* range (Å^{-1})**	$0.004 < q < 0.442$		
Temperature	288 K		
Concentration (mg/ml) / Volume (μl)	7/60		
Structural parameters			
***I(0) (relative)* (from *P(r)*)**	1.01	0.31	0.12
R_g (Å) from P(r)	55.07	86.47	51.26
***I(0) (relative)* (from Guinier)**	1.01	0.31	0.12
R_g (Å) (from Guinier)	55.00	85.68	51.33
D_{max} (Å)	192.80	416.21	141.25
Porod volume (Å^3)	753000	966000	759000
Molecular mass determination (Da)			
From Porod volume (V_p)	345940	715433	426948
From excluded volume (V_{ex})	389434	512312	260059
From SAXS MoW2 ($q$max, Å^{-1})	348272	664180	397175
Bayesian Inference (MW probability %)	318450 (99.85)	585225 (92.28)	318450 (99.89)
From sequence	245895	327860	294522
Modelling parameters			
Shape reconstruction	DAMMIN		
Symmetry	P1	P1	N/A
NSD	0.829 ± 0.05	0.576 ± 0.03	N/A
SAXSDB codes	SASDDT5	SASDDU5	N/A
Software employed	PRIMUS, GNOM, AUTORG, DAMMIF, DAMAVER, DAMFILT, PyMol		

Table 4.1 SEC-SAXS data collection and structural parameters used for ab initio modelling. Note: I(0) does not scale with particle size as the concentration of sample within the frames scaled and merged was not consistent between σNS and RNP.

To analyse the stoichiometry of σNS and its RNP complex, we used native electrospray ionization – ion mobility spectrometry – mass spectrometry (ESI-IMS-MS). The ESI-MS spectrum of σNS apoprotein reveals a range of different oligomers (dimer, tetramer, hexamer and octamer) as is typical of oligomeric assemblies (**Figure 4.10**). This is due to progressive disassembly of larger species within the gas phase due to weakened hydrophobic interactions (Calabrese and Radford, 2018). In agreement with previously published results, the spectrum of σNS contains hexamers, which likely assembled as a trimer-of-dimers (Borodavka *et al.*, 2015). In contrast, the RNP spectrum shows a large population of octameric σNS bound to two RNA molecules, and a small fraction of hexameric σNS bound to a single RNA (**Figure 4.10**). Together with SAXS data, this indicates that binding of multiple RNAs induces a hexamer-to-octamer transition.

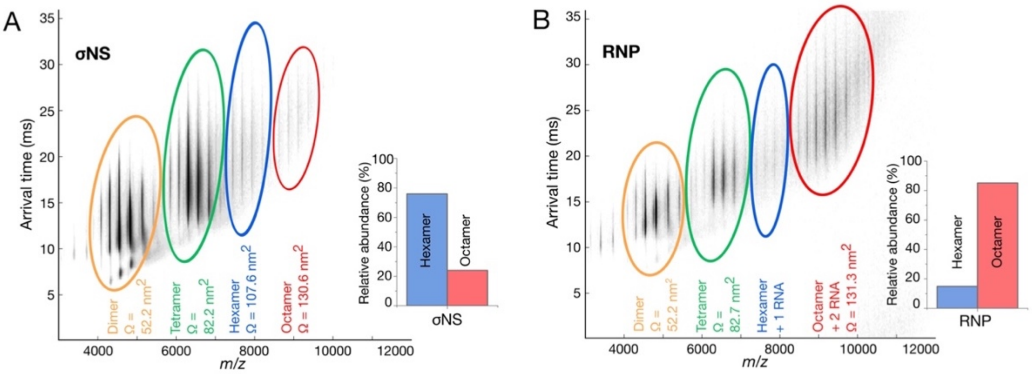


Figure 4.10 σNS undergoes a hexamer-to-octamer transition upon binding of multiple RNA molecules Native ion mobility spectrometry – mass spectrometry (IMS-MS) of σNS (A) and RNP (B). Rotationally-averaged collisional cross-sectional areas (Ω) are indicated for each species. Relative abundance of hexameric and octameric complexes for σNS (A) and RNP (B) are shown as inset bar chart.

Using this IMS-MS data, we estimated rotationally averaged collision cross-sections (CCSs) for each observed oligomer. The CCS for each observed oligomer increased with charge state due to partial unfolding (**Figure 4.10**). Therefore, the least ionised species (i.e. that with the greatest m/z) is the least affected by this and is the 'most native' (i.e. least unfolded). Using the SAXS-derived ab initio models, CCS estimates were generated using the *IMPACT* software (Marklund *et al.*, 2015). This uses the projection approximation (PA) method to consider the 'shadow' that a given molecule would cast on a flat surface as representative of its CCS. CCS estimates were generated for 20 independent ab initio models of both σNS and RNP. We directly compared the observed CCSs for each oligomer with SAXS-derived CCS estimates. The CCS values of the hexameric σNS (103.6 ± 2.8 nm^2) are in agreement with the CCS value estimated for the apoprotein SAXS envelope (102.2 ± 0.5 nm^2) (**Figure 4.11**). Likewise, the measured CCS value for octameric σNS bound to two RNAs (120.5 ± 0.5 nm^2) closely corresponded to the CCS estimated from the RNP SAXS envelope (119.4 ± 1.0 nm^2). This confirms that in solution, concurrent binding of two RNAs results in the assembly of elongated, octameric σNS-RNP complexes.

We next aimed to investigate whether there was a functional aspect to this observed RNA-driven oligomerisation. As σNS is an RNA chaperone with helix-destabilising activity, we decided to explore the relationship between this activity and its oligomeric state. To this end, a dual-labelled RNA stem loop was designed, bearing a 5'-Atto532 (donor, D) and a 3'-Atto647N (acceptor, A) FRET pair. This allowed us to use FRET efficiency ($E_{FRET(app)}$) as a proxy for the RNA conformation; if the RNA was folded, the 5’ and 3’ termini would be expected to be within close proximity of one another, resulting in high $E_{FRET(app)}$. If the RNA was unfolded, a reduced $E_{FRET(app)}$ would occur.

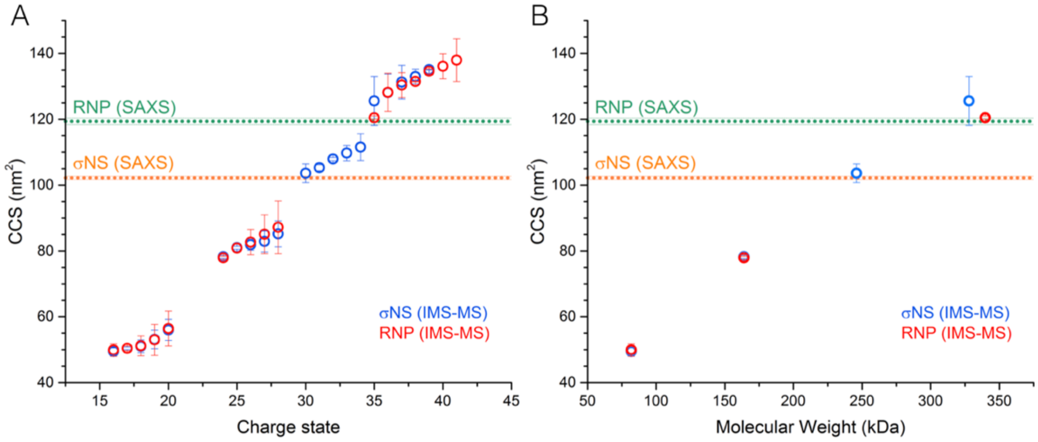


Figure 4.11 Collision cross-section distributions of σNS and RNP. A: Plot of observed CCS of σNS (blue) and RNP (red) oligomers as a function of charge state. B: Plot of smallest observed CCS as a function of oligomer molecular weight. A & B: CCS estimates using SAXS data are displayed as horizontal lines (RNP, green; σNS, orange) with standard deviation shown as thinner, broken lines.

Increasing concentrations of σNS were co-incubated with this stem-loop, and the different oligomers were separated using a native agarose gel EMSA (**Figure 4.12**). Multiple bands were observed as σNS concentration increased, corresponding to free RNA and different RNA-bound σNS oligomers. Given the oligomers observed in the RNP IMS-MS spectrum, the first band was interpreted as hexameric σNS, and the second as octameric σNS. An additional upper band is also observed at high σNS concentrations (>2 μM). Given the length of the RNA used (36-nt) is roughly twice the established RNA-binding footprint of σNS (20-nt), this was interpreted as multiple σNS oligomers simultaneously binding a single RNA molecule (**Figure 4.12**). The fraction of RNA bound by σNS was measured through densitometry, and the apparent dissociation constant (K_D) was 25 nM (**Figure 4.12C, red**).

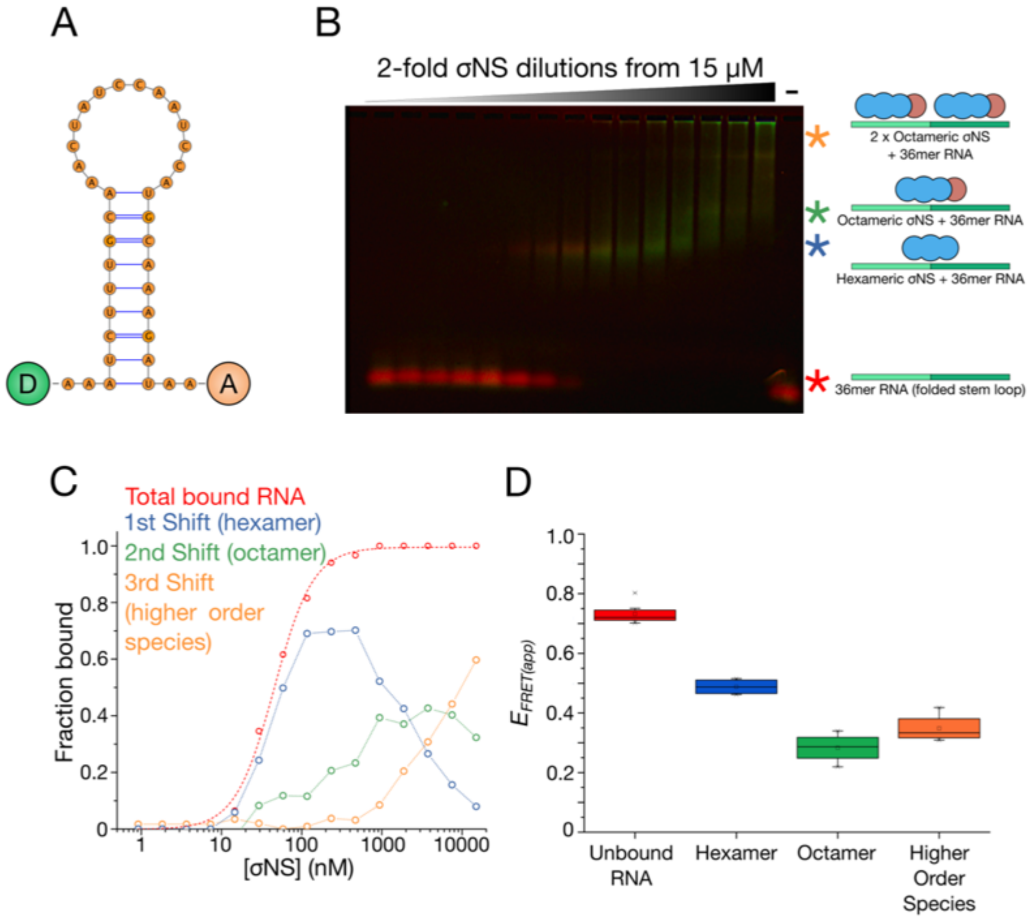


Figure 4.12 Analysis of RNA unfolding activities of different σNS oligomers using in-gel FRET A: Dual-labelled 36-nt stem loop containing 5'-Atto532 (donor, D) and 3'-Atto647N (acceptor, A) fluorophores. This stem loop has an estimated MFE of $\Delta G = -8.9$ kcal mol^{-1}. B: Pseudocoloured electrophoretic mobility shift assay (EMSA) of σNS binding dual-labelled 36mer stem loop. Native agarose gels were visualised using a fluorescence scanner with 532 nm excitation, imaging donor and acceptor channels separately. Different RNA and σNS species are indicated: free RNA – red, hexamer-bound RNA – blue, octamer-bound RNA – green, higher-order species – yellow. C: EMSA-derived binding curve. Total RNA bound by all σNS (red) ($K_{D(app)}$ = 25 nM), with different shifts appearing as σNS concentration increases. D: Apparent in-gel FRET efficiencies of RNA bound to different σNS oligomers.

Through measuring the in-gel $E_{FRET(app)}$ it was determined that the RNA stem loop was folded ($E_{FRET(app)} = 0.73 \pm 0.03$). Hexamer-bound RNA had $E_{FRET(app)} = 0.40 \pm 0.11$, which decreased further for octamer-bound RNA ($E_{FRET(app)}$ =

0.28 ± 0.05) (**Figure 4.12**). This indicates that octameric σNS has a greater capacity for disrupting RNA structure than its hexameric form. This may be due to an increased RNA-binding surface present on octameric σNS.

4.2.4 Functional analysis of NSP2 and σNS

Having investigated the stem-loop unfolding activity of σNS, we investigated the relationship between RNA binding and unwinding by both NSP2 and σNS using single-pair (sp)FRET. This allowed us to visualise discrete FRET populations corresponding to protein-free and oligomer-bound RNAs, using the dual-labelled stem loop as described above (**Figure 4.12**). In the absence of protein, a single, high-FRET population is present (**Figure 4.13**), indicating that this stem loop adopts a stable, hairpin conformation. In the presence of NSP2, this high-FRET population persists, in addition to a low-FRET population corresponding to unfolded RNA.

As protein-bound RNA will diffuse more slowly than unbound RNA (and hence have a rightward shifted autocorrelation function (ACF)), this allows us to investigate the effect of NSP2 binding on RNA conformation. Species-selective analysis revealed that NSP2 is bound to both the high- and low-FRET populations, indicating that it can bind to both folded and unfolded RNA structures (**Figure 4.13**).

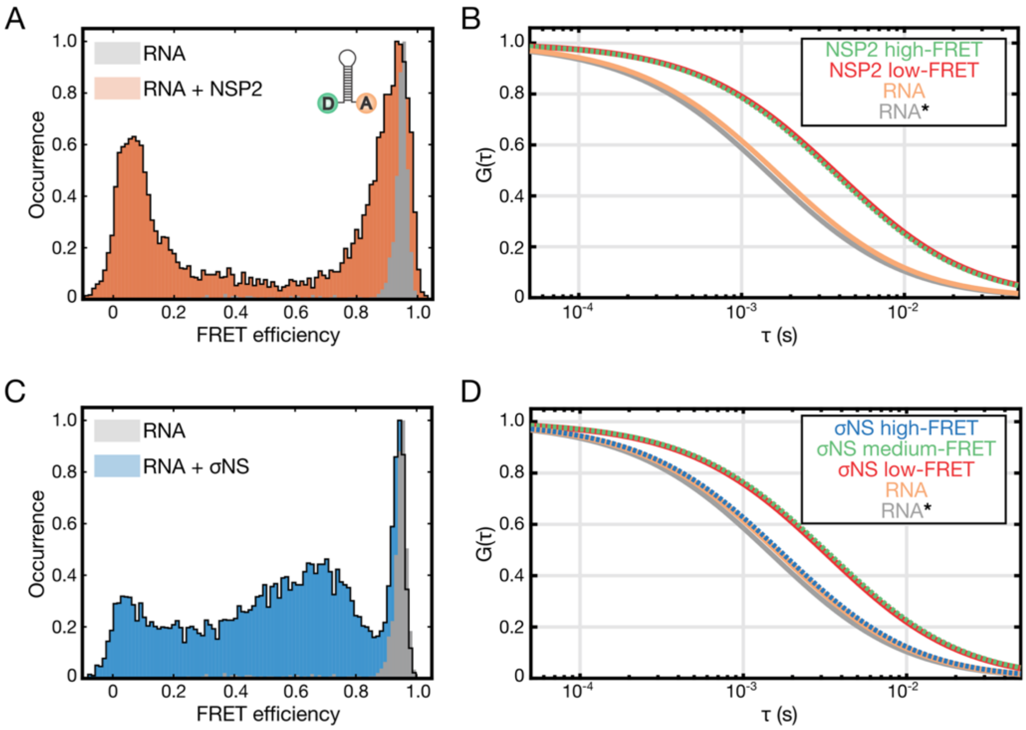


Figure 4.13 spFRET analysis of helix destabilisation by σNS and NSP2 A: Histograms of spFRET efficiencies of dual-labelled RNA stem loop (as described in Figure 4.14) alone (grey) or in the presence of 25 nM NSP2 (orange). B: Species-selective correlation analysis of freely diffusing RNA (orange), and NSP2-bound RNA FRET populations (high-FRET – green, low-FRET red). A typical ACF of freely diffusing RNA is shown in grey. C: Histograms of spFRET efficiencies of dual-labelled RNA stem loop alone (grey) and in the presence of σNS (blue). D: Species-selective correlation analysis of freely diffusing folded RNA (orange), and σNS-bound FRET populations (high-FRET – blue, medium-FRET – green, low-FRET – red). A typical ACF of freely diffusing RNA is shown in grey. Grey and orange ACFs serve as an internal control to confirm the accuracy of species-selective correlation analysis. spFRET data were collected by Dr Alex Borodavka, and analysed by Dr Anders Barth, LMU Munich.

We similarly analysed RNA unfolding by σNS. Although the high-FRET population is still present upon incubation with σNS, a range of mid- and low-FRET populations are produced corresponding to partially- and fully-unfolded RNA structures (**Figure 4.13**). Species-selective correlation analysis indicated that while σNS is bound to the mid- and low-FRET populations, it is not bound to the high-FRET population (**Figure 4.13**). These results indicate that NSP2 is able to bind to both folded and unfolded RNA conformations, although its binding may not necessarily induce RNA unwinding. In contrast, σNS binding induces a range of unfolded RNA intermediates.

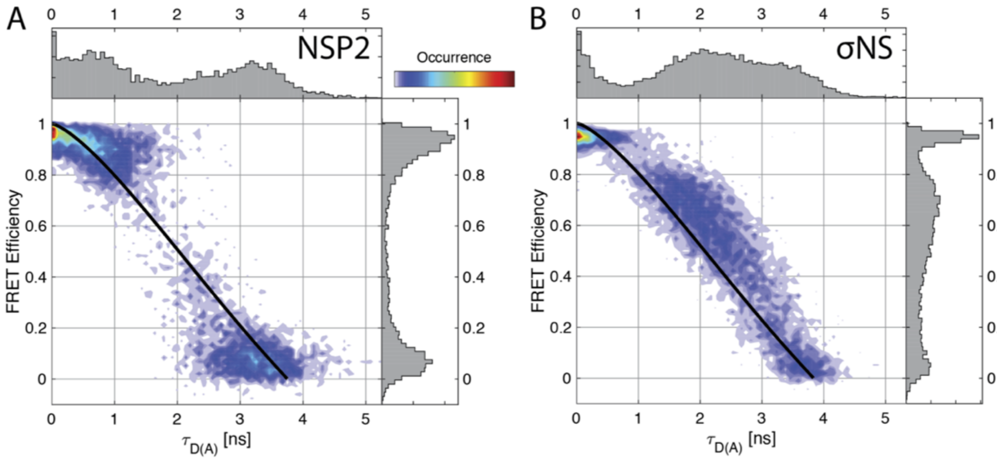


Figure 4.13 Conformational dynamics of RNA stem-loop in the presence of NSP2 and σNS A & B: 2D FRET histograms of FRET efficiency vs fluorescence lifetime of the donor fluorophore, τ, for RNA stem-loop bound to NSP2 (A) and σNS (B). Data were collected by Dr Alex Borodavka, and analysed by Dr Anders Barth, LMU Munich.

To understand the dynamics of σNS-mediated helix destabilisation, we performed burst-wise fluorescence lifetime analysis. If the RNA is undergoing dynamic conformational changes on a time-scale faster than diffusion (i.e. sub-millisecond), the measured FRET efficiency represents an averaged value corresponding to multiple species. By plotting the spFRET efficiency as a function of donor fluorescence lifetime, the presence of these dynamics can be visualised as sub-millisecond conformational transitions induce a shift towards longer fluorescence lifetimes, deviating from the relationship between FRET efficiency and donor lifetime expected for a static FRET species (Barth, Voithenberg and Lamb, 2014; Voith von Voithenberg *et al.*, 2016; Barth *et al.*, 2018). NSP2 did not induce intermediate E_{FRET} values, and no sub-millisecond dynamics were observed (**Figure 4.13**). The mid-FRET population of σNS is dynamic on the sub-millisecond time-scale and likely represents an ensemble of inter-converting σNS-bound partially unwound RNA conformations.

As ssRNA binding is a prerequisite for RNA unfolding, the dependence of RNA binding by both NSP2 and σNS on RNA structure was investigated. Three fluorescently-labelled 20mer RNAs with different thermodynamic stabilities were designed for this purpose: an unstructured RNA, a metastable RNA hairpin (ΔG = -3.8 kcal mol^{-1}), and a stable hairpin (ΔG = -8.1 kcal mol^{-1}). It is worth noting that while the stable hairpin contained a tetraloop (AACA), this is not predicted to further stabilise the RNA stem structure, as is the case with GNRA or UNCG tetraloops (where N is any base and R is purine) (Xia, 2008; D'Ascenzo *et al.*, 2016).

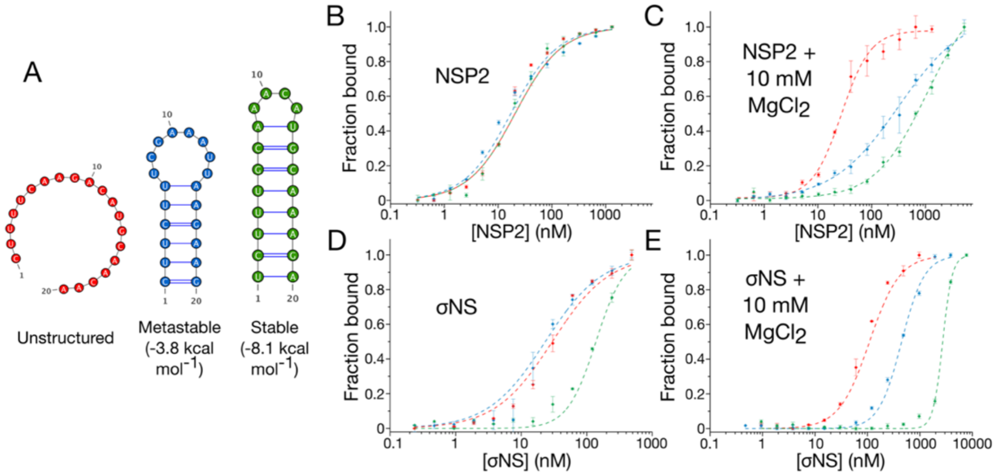


Figure 4.14 Stability of RNA structures determines the binding preferences of NSP2 and σNS A: 20mer model stem loops bearing 5'-Alexafluor488 used for fluorescence anisotropy binding assays. RNA length (20-nt) was maintained, but estimated MFE (i.e. stability) was varied. B & D: NSP2 binds to all RNAs measured with identical affinities (B). σNS binds to both unstructured (red) and metastable (blue) RNAs, but displays reduced binding for stable (green) hairpin. C & E: Mg^{2+}-dependent stabilisation of RNA structures reduces binding of both NSP2 (C) and σNS (E) to RNA stem loops (metastable – blue, stable – green), while there is a negligible (~30%) decrease in binding affinity for the unstructured (red) RNA.

NSP2 binds all three RNAs with near-identical affinities (K_D = 20 ± 3.2 nM), supporting the previous observations of sequence-nonspecific albeit high-affinity RNA binding activity (**Figure 4.14, Table 4.2**). While σNS binds to both the unstructured and metastable RNAs with similar affinities (K_D = 37 ± 1.5 nM and 24 ± 2.6 nM, respectively), it has a reduced affinity for the stable RNA (K_D = 137 ± 1.9 nM) (**Figure 4.14, Table 4.2**).

To further investigate the role of RNA structural stability on protein binding, we co-incubated RNAs with $MgCl_2$ and performed binding affinity measurements. Divalent cations such as Mg^{2+} are able to stabilise RNA secondary and tertiary structures by reducing electrostatic repulsion between closely-packed phosphate groups. Although Mg^{2+} is able to stabilise and compact both structured and unstructured RNAs through diffuse binding (i.e. nonspecific interactions with the anionic field of the nucleic acid backbone), only RNA secondary and tertiary structures are stabilised through site-specific binding (i.e. the ion specifically interacts with the RNA either directly or through a single hydration shell). Although there was a minimal increase in K_D values of NSP2 and σNS for the unstructured RNA (1.4- and 1.5-fold increase, respectively), both proteins exhibited a 10- to 30-fold decrease in affinity for the metastable and stable RNAs (**Figure 4.14, Table 4.2**). This corresponds to a $\Delta\Delta G$ of -1.28 to -1.97 kcal mol^{-1}.

This result indicated that although NSP2 can bind to stable RNA structures more efficiently than σNS, both proteins preferentially bind unstructured RNAs.

	RNA substrate	$K_{d(obs)}$ (nM)	ΔG_{obs} (kcal mol^{-1})	$K_{d(obs)}$ (nM) + 10 mM $MgCl_2$	ΔG_{obs} (kcal mol^{-1}) + 10 mM $MgCl_2$	$\Delta\Delta G_{obs}$ (kcal mol^{-1})
NSP2	Unstructured	20.7±1.7	-10.12 ± 0.65	28 ± 1.5	-9.94 ± 0.67	-0.19
	Metastable	20.6±3.2	-10.12 ± 0.63	212 ± 22	-8.79 ± 0.66	-1.32
	Stable	20.8±1.8	-10.12 ± 0.65	629 ± 47	-8.16 ± 0.68	-1.97
σ NS	Unstructured	37.2±1.5	-9.91 ± 0.67	45.3 ± 0.7	-9.67 ± 0.71	-0.23
	Metastable	24.1± 2.6	-10.03 ± 0.64	420 ± 14	-8.39 ± 0.70	-1.64
	Stable	137± 2	-9.03 ± 0.73	1283 ± 95	-7.76 ± 0.68	-1.28

Table 4.2 Overall free energies of NSP2- and σNS-20mer RNA interactions, calculated from binding curves in **Figure 4.14**. $\Delta\Delta G_{obs}$ corresponds to the difference in free energies of binding in the presence and absence of 10 mM $MgCl_2$.

To test the dependency of K_D on ionic strength (and therefore to infer the electrostatic contributions to RNA binding of either protein), the affinities of NSP2 and σNS for the unstructured 20mer RNA were determined at a range of NaCl concentrations.

As NSP2 octamers disassemble under high ionic strengths, a range of 50 – 250 mM NaCl was used to determine binding affinities. σNS, however, aggregates at NaCl concentrations below 50 mM yet appears to be unaffected by higher ionic strength conditions. A greater range of ionic strength buffers were therefore used to determine σNS-RNA affinities (100 – 500 mM NaCl).

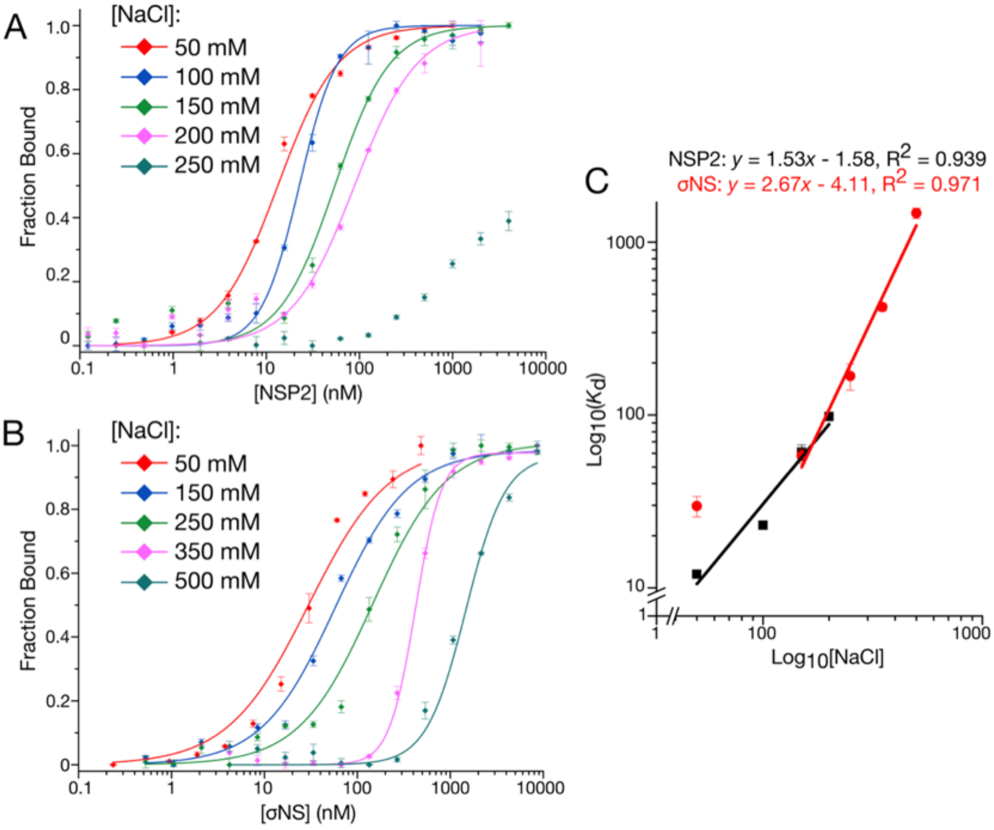


Figure 4.15 Electrostatic contributions to NSP2- and σNS-RNA interactions Affinities of NSP2 (A) and σNS (B) binding 20mer unstructured RNA as measured by fluorescence anisotropy across a range of monovalent cation (Na^+) concentrations. Linear correlation between log (K_D) and log([NaCl)] for both NSP2 and σNS. The apparent number of electrostatic interactions involved in RNA binding is estimated from fitting slopes, and corresponds to <2 for NSP2, and 3 – 4 for σNS.

Strong linear relationships between log([NaCl]) and log(K_D) were observed for both NSP2 and σNS (R^2 = 0.939 and 0.971 for NSP2 and σNS, respectively) (**Figure 4.15**). According to the semi-empirical extension of Manning's counterion concentration theory, the number of ions released from the protein-nucleic acid interface into bulk solution (and thus the number of electrostatic interactions between protein and nucleic acid) is equal to the slope of this plot (Manning, 1969). This indicated that NSP2 binds RNA with at most two salt bridges, while σNS binds RNA by between 3 – 4 electrostatic contacts. It is worth noting that the affinity of σNS binding measured under the lowest ionic strength conditions was excluded from the linear fit. In previous studies, this has been attributed to the maximum number of ionic bonds occurring at this low [NaCl] (Draper and von Hippel, 1978; Müller and Göringer, 2002).

	NaCl (mM)	$K_{d(obs)}$ (nM)	ΔG_{obs} (kcal mol^{-1})	ΔG_{PE} (kcal mol^{-1})	ΔG_{nPE} (kcal mol^{-1})	% Electrostatic contribution
NSP2	50	12.2 ± 0.5	-10.43 ± 0.67	-2.62 ± 0.63	-7.81 ± 0.63	25.1
	100	21 ± 1	-10.05 ± 0.67	-2.01 ± 0.63	-8.04 ± 0.63	20.0
	150	63 ± 6	-9.50 ± 0.68	-1.66 ± 0.63	-7.84 ± 0.63	17.5
	200	98 ± 9	-9.23 ± 0.66	-1.41 ± 0.63	-7.82 ± 0.63	15.3
	1000*	N/A	-7.89*	0	-7.89*	0*
σ NS	50	29.7 ± 1.1	-9.91 ± 0.68	-4.57 ± 0.64	-5.34 ± 0.64	46.1
	150	58 ± 5	-9.53 ± 0.66	-2.90 ± 0.64	-6.63 ± 0.64	30.4
	250	168 ± 29	-8.92 ± 0.49	-2.12 ± 0.64	-6.80 ± 0.64	23.7
	350	421 ± 22	-8.39 ± 0.64	-1.60 ± 0.64	-6.79 ± 0.64	19.1
	500	1470 ± 101	-7.68 ± 0.69	-1.06 ± 0.64	-6.62 ± 0.64	13.8
	1000*	N/A	-6.71*	0	-6.71*	0*

Table 4.3 Electrostatic contributions to free energy of binding of NSP2 and σNS, as calculated from binding curves measured in Figure 8. * denotes values estimated through extrapolation of binding energies to 1 M NaCl, where electrostatic component of free energy of binding is eliminated (Samatanga *et al.*, 2017).

Under physiological ionic strength conditions (150 mM NaCl), both NSP2 and σNS bind unstructured 20mer RNA with near identical apparent free energies

(ΔG = -9.50 ± 0.68 kcal mol^{-1} and -9.53 ± 0.66kcal mol^{-1}). However, NSP2 interacts with a 17.5% electrostatic component, while σNS has a 30.4% electrostatic component. This indicated that although NSP2 and σNS bind RNA with near-identical affinities, σNS has a greater electrostatic contribution to binding than NSP2. This suggests that NSP2 and σNS may employ different modes of RNA binding.

4.3 Discussion

Inter-segment RNA-RNA contacts have been long-proposed to support the selection and packaging of multi-segment genomes of viruses encompassing the Reoviridae family of dsRNA viruses. Although RNA chaperone proteins can remodel RNA structure and thereby facilitate RNA remodelling and RNA-RNA annealing, this is strongly contingent on the stability of RNA structure. In this particular case, we observed that RNA secondary structure stability (along with the mode in which RNA chaperones bind) may provide an avenue on which to regulate the activity of RNA chaperone proteins that would otherwise bind ssRNAs without any sequence specificity. In the cases of NSP2 and σNS, we observed that a significant determinant of protein-mediated intersegment RNA-RNA interactions is the efficiency of protein-mediated RNA remodelling. Although both NSP2 and σNS bind ssRNA with ~20 nM affinity, their RNA unfolding activities differ. As the downstream RNA chaperone activities (RNA-RNA annealing and duplex release) are contingent on successful protein-mediated RNA unfolding, this particular activity could act as a key regulation step for determining the specificity of intersegment RNA-RNA interactions required for genome segment assembly.

This provides a potentially novel regulatory mechanism for RNA wherein the stability of local RNA structure dictates or governs the binding and remodelling activities of RNA chaperone proteins. Other proteins with similar RNA chaperone activities such as E. coli StpA and Hfq, or MLV NC preferentially bind to unstructured ssRNAs (D'Souza and Summers, 2004; Grossberger *et al.*, 2005; Santiago-Frangos and Woodson, 2018). This indicates that chaperone-mediated RNA remodelling may be fine-tuned by the stability of RNA structures.

We observed that upon binding two 20mer RNAs, σNS undergoes a hexamer-to-octamer transition. This is a distinct mechanism of assembly from that of other viral RNP complexes that assemble into non-discrete higher-order species (Butler and Klug, 1971; Reguera, Cusack and Kolakofsky, 2014; Miles *et al.*, 2016). This appears to be unique to σNS, as other functionally analogous proteins encoded by the *Reoviridae* family do not appear to undergo such RNA-driven oligomerisation (Taraporewala and Patton, 2004; Jiang *et al.*, 2006; Akita *et al.*, 2011, 2012; Wu *et al.*, 2013). We observed that RNA binding results in a shift in equilibrium between σNS hexameric assemblies (as a trimer of dimers) to octameric species (as tetramers of dimers, in complex with multiple RNA molecules). Therefore, we propose that σNS apoprotein hexamers are in dynamic equilibrium with low-population σNS octamers, whereby the octameric species is further stabilised by bound RNA. In light of this finding, it may be interesting to explore the effects of binding of longer RNA substrates on σNS oligomerisation.

Altogether, our results provide the first evidence that proteins encoded by different members of the *Reoviridae* family utilise different RNA chaperone mechanisms to achieve genome segment assortment. In conclusion, we propose that the distinct mechanisms of RNA chaperone activity, together with RNA structural stability regulate the selection of a complete cohort of genomic RNAs.

Chapter 5: Role of NSP2 C-terminus in RNA release and chaperone recycling

5.1 Introduction

Rotavirus NSP2 is a non-structural protein essential for virus replication and assembly (Taraporewala, Chen and Patton, 1999). As the RNA chaperone activity of NSP2 plays a key role within the rotavirus replication cycle (Borodavka, Desselberger and Patton, 2018), the RNA-binding site of NSP2 presents an attractive target for structure-based drug design. However, in order to design effective drugs against NSP2, the molecular mechanisms underlying this activity must be understood.

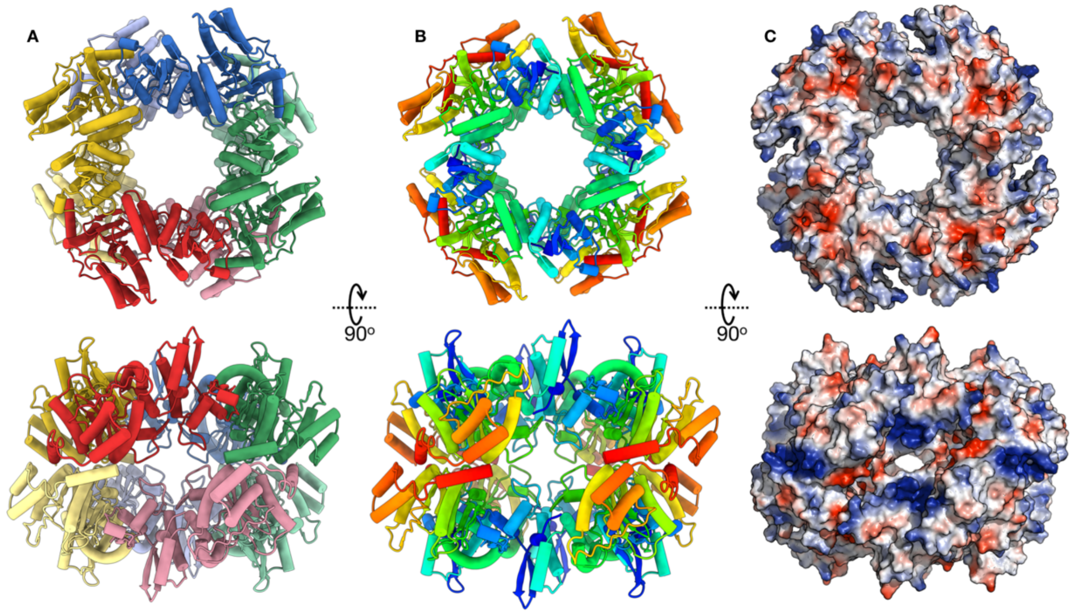


Figure 5.1 Structure of NSP2 octamer A: Top-down and side views of NSP2 octamer (PDB 1L9V), with monomers shown as different colours. B: View of NSP2 as in A, but coloured by sequence (rainbow, blue – N-terminus, red – C-terminus). C: Surface electrostatic potential of NSP2 octamer. Red: negatively charged, blue: positively charged.

Major breakthroughs in understanding the structural basis of NSP2-RNA interactions include a co-crystal structure and a low-resolution cryoEM reconstruction (Jiang *et al.*, 2006; Hu *et al.*, 2012). These two structures also represent the challenges and pitfalls of obtaining a structure of an NSP2 RNP complex.

In the co-crystal structure, only two of the eight nucleotides of the co-crystallised RNA had interpretable electron density (Hu *et al.*, 2012). This is somewhat unsurprising, as the NSP2-RNA interaction lacks sequence specificity. This may permit RNA to adopt multiple conformations and bind in multiple registers on the surface of NSP2, resulting in an averaging of RNA density throughout the crystal. The two nucleotides that could be identified (5'-GG) were positioned within the enzymatic cleft of NSP2, the NTPase active site (Hu *et al.*, 2012). As other studies have demonstrated that the enzymatic activity of NSP2 is not essential for its RNA chaperone activity (Borodavka *et al.*, 2017), it is unlikely that this dinucleotide-binding site represents the major RNA-binding site of NSP2.

A cryoEM reconstruction of NSP2 bound to poly(A)$_8$ RNA was determined to a resolution of 8.5 Å in 2006 (Jiang *et al.*, 2006). A difference map (i.e. the density for the NSP2 apoprotein subtracted from the density for the NSP2 RNP complex) exhibited diffuse RNA density (Jiang *et al.*, 2006). This density was localised not only to positively charged grooves (**Figure 5.1**), but to regions all around the equator of NSP2. However, due to the modest resolution, it was not possible to visualise exactly *how* NSP2 interacts with RNA. Therefore, it remains to be established how NSP2 captures RNA, and whether NSP2 RNP complex formation is associated with conformational change.

Another feature of interest of NSP2 is the conformation of the C-terminal helix (CTH) spanning residues 303 – 316. This helix is connected to the rest of NSP2 by an unstructured, flexible linker region (residues 291 – 302). The CTH and linker together comprise the C-terminal domain (CTD). The CTD has been implicated in multiple different aspects of the rotavirus replication cycle, including the genome segment assortment and viroplasm assembly (Borodavka *et al.*, 2017; Criglar *et al.*, 2018), and *in vivo* experiments have identified that the CTD of NSP2 is essential for virus replication (Criglar *et al.*, 2014). Despite extensive studies, the precise role of NSP2 CTD in the RNA chaperone activities of NSP2 has remained unexplored.

The work in this chapter first describes a functional role for the CTD in the RNA chaperone activity of NSP2 through the use of complementary biophysical techniques (fluorescence cross-correlation spectroscopy and single-pair FRET). Additionally, given the recent technical advances in cryogenic electron microscopy (cryoEM), we aimed to determine a high-resolution structural model of NSP2 as an apoprotein and in complex with RNA. The structural and functional insights presented in this chapter demonstrate that the CTD of NSP2 plays a role in modulating RNA binding kinetics.

5.2 Results

5.2.1 Establishing the role of NSP2 C-terminal domain (CTD) in RNA chaperone activity

We first investigated the role of the CTD in the RNA matchmaking activity of NSP2 using a fluorescence cross-correlation (FCCS)-based assay (**Chapter 4** (Borodavka *et al.*, 2017; Bravo *et al.*, 2018)). FCCS experiments were carried out by Dr Alex Borodavka. Rotavirus segments S6 and S11 were co-incubated either alone, or in the presence of excess full-length NSP2 or C-terminally truncated NSP2 (residues 1 – 293, termed NSP2-ΔC). The yield (i.e. the amount of the resulting dsRNA product produced by the interacting S6 and S11 ssRNAs) was measured using FCCS (**Figure 5.1**).

In the absence of NSP2, S6 and S11 a cross-correlation function (CCF) amplitude of ~0 was observed, indicating that these two RNAs do not spontaneously interact (**Figure 5.1**). It is likely that this is due to complementary sequences within either segment being sequestered within stable RNA secondary structures (Borodavka *et al.*, 2017; Bravo *et al.*, 2018). As expected, co-incubation of S6 and S11 with NSP2 produced a high CCF amplitude, indicative of intermolecular RNA duplex formation. However, co-incubation of NSP2-ΔC with these RNAs results in a reduced CCF amplitude (**Figure 5.1**), indicating that NSP2-ΔC has a reduced RNA annealing activity relative to full-length NSP2. This result suggests a role for the CTD in NSP2's RNA chaperone activity.

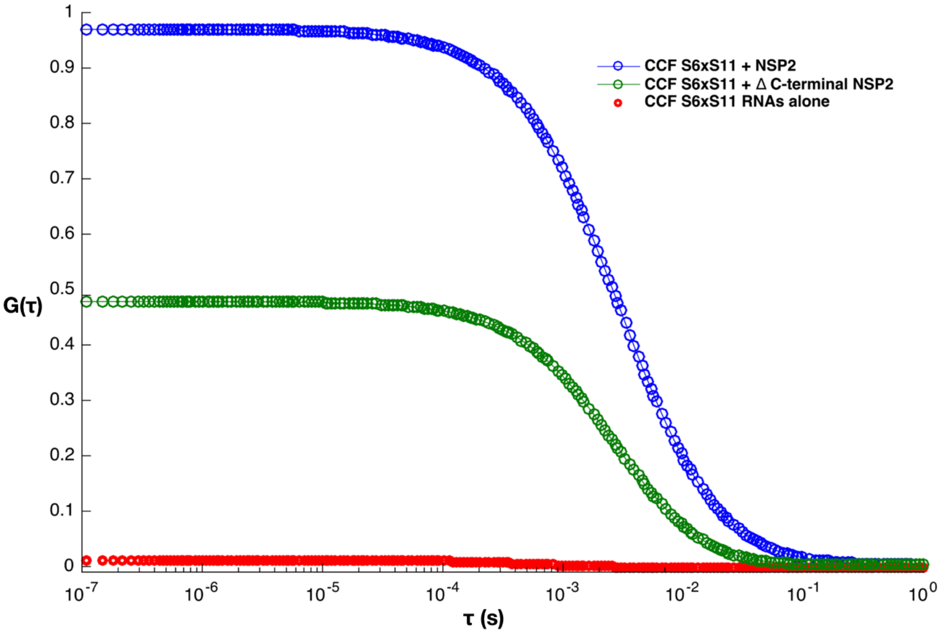


Figure 5.2 NSP2 C-terminus is required for RNA-RNA matchmaking Interactions between rotavirus segments S6 and S11 incubated either alone, or in the presence of NSP2 or NSP2-ΔC were probed by fluorescence cross-correlation spectroscopy (FCCS). Incubation of S6 and S11 alone results in a cross-correlation amplitude of ~0 (red), indicating that these RNAs do not spontaneously interact. Co-incubation with full-length NSP2 (blue) results in a high yield of interacting RNAs, whereas co-incubation with NSP2-ΔC does not (normalised CCF amplitudes of ~1.0 and ~0.5, respectively). Data were collected and processed by Dr Alex Borodavka.

As the ability of NSP2 to unfold and remodel RNA structures is a prerequisite for its RNA matchmaking activity, we next investigated the role of the CTD in helix destabilisation. We used single-pair (sp)FRET to directly compare the abilities of NSP2 and NSP2-ΔC to unfold an RNA stem-loop (ΔG = -8.9 kcal mol^{-1}) bearing a FRET pair on its 5'- and 3'- termini (Atto532 and Atto647N, respectively) (**Chapter 4, Figure 4.11**). spFRET measurements were performed by Kira Bartnik, Ludwig-Maximilian University, Munich.

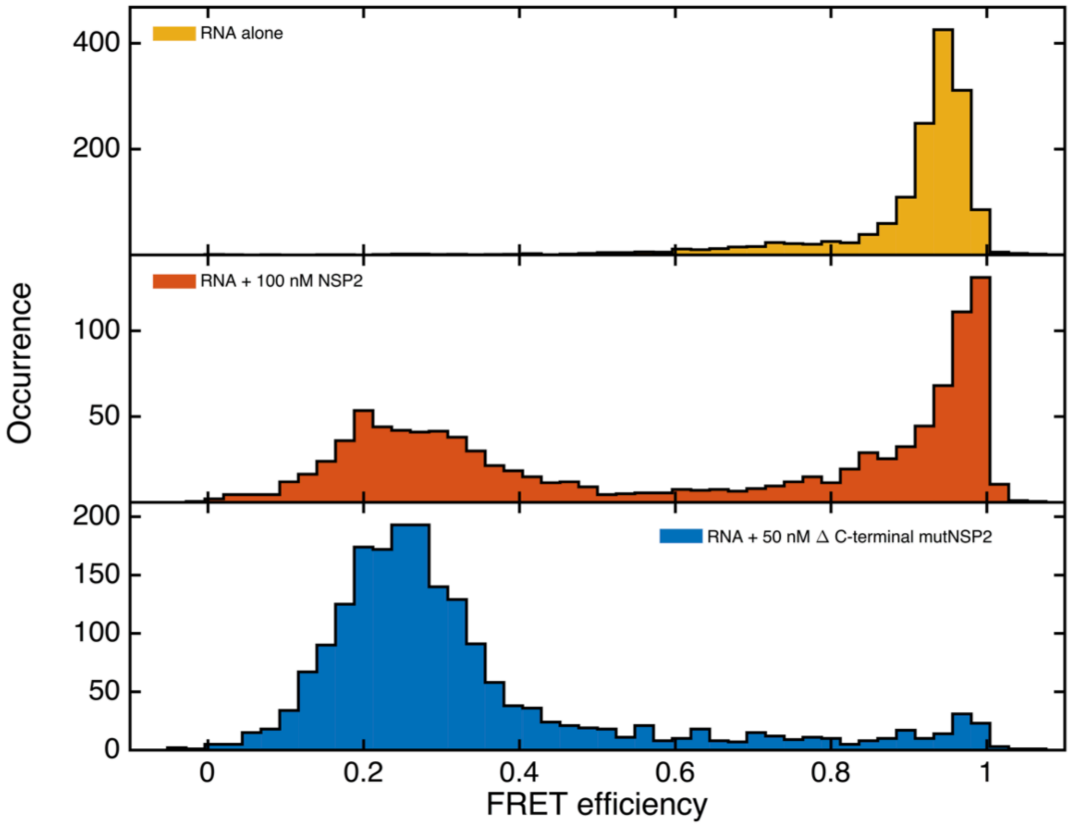


Figure 5.3 NSP2-ΔC has enhanced RNA helix unfolding activity spFRET efficiency histograms of dual-labelled RNA stem loop alone (top, yellow) and in the presence of 100 nM full-length (middle, orange) or 50 nM NSP2-ΔC (bottom, blue). Data was collected and processed by Kira Bartnik.

The stem-loop alone produced a single high-FRET population (E_{FRET} = ~0.95), indicating that it adopts a folded hairpin conformation in solution (**Figure 5.3**). Incubation with 100 nM NSP2 resulted in two distinct FRET populations, corresponding to fully folded (E_{FRET} = ~0.95) and unfolded (E_{FRET} = ~0.25). No intermediate FRET populations were observed (corresponding to partially-disrupted stem loop conformations) (**Chapter 4**).

Similarly, we analysed the stem-loop unfolding activity of NSP2-ΔC. Surprisingly, no high-FRET population was observed in the presence of NSP2-ΔC, and the low-FRET state was more abundant (**Figure 5.3**). This finding was particularly interesting, as it demonstrated that C-terminally truncated NSP2 has enhanced RNA unfolding activity. This result is somewhat paradoxical; NSP2-ΔC is *better* at unfolding RNA, and yet is *worse* at promoting intermolecular RNA-RNA interactions. Although these results directly implicate NSP2 CTD in its RNA chaperone activity, it would appear that CTD may not be exclusively involved in RNA-binding as was previously proposed (Hu *et al.*, 2012). Rather than modulating the RNA binding affinity of NSP2, the CTD may be play a role in the duplex formation or RNA chaperone recycling (i.e. RNA dissociation from NSP2) phase.

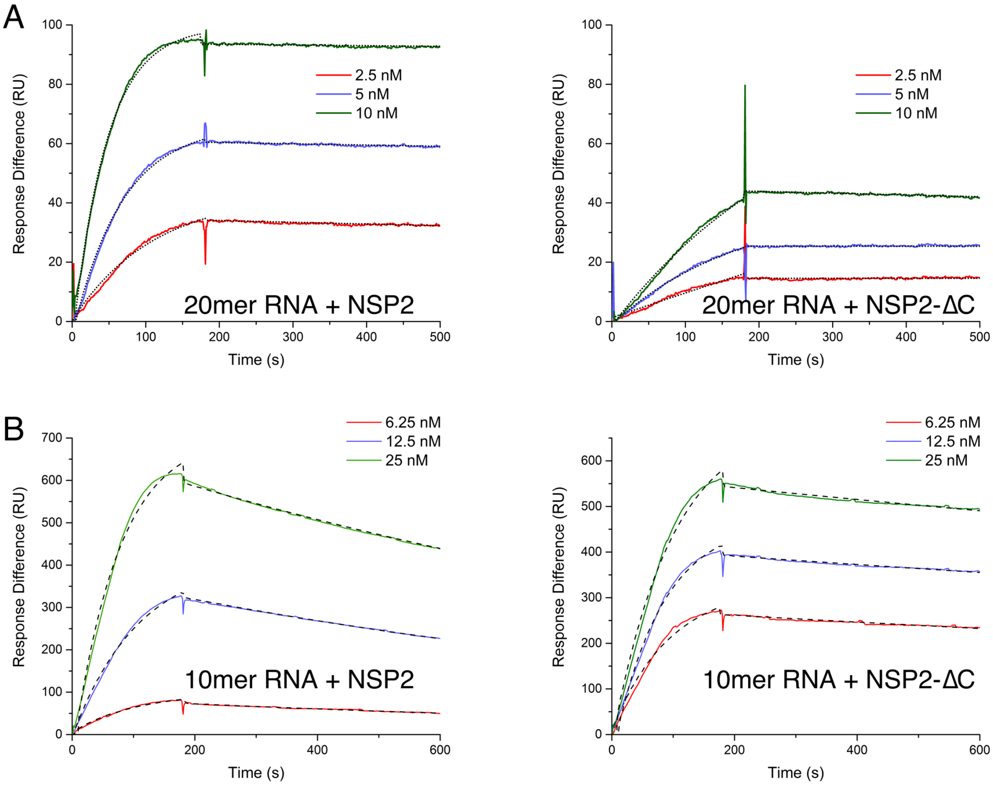


Figure 5.4 Role of C-termini in RNA release. A: SPR sensograms of NSP2 wild-type (left) and NSP2-ΔC (right) binding to 20mer RNA, exhibiting slow dissociation kinetics. B: SPR sensograms of NSP2 wild-type (left) and NSP2-ΔC (right) binding to 10mer RNA.

To investigate this, binding kinetic measurements were performed using surface plasmon resonance (SPR) (**Figure 5.4**). Initial experiments using 5'-biotinylated unstructured 20mer RNA (immobilised to a streptavidin-coated chip) exhibited very slow off-rate and therefore cannot be used to quantify RNA binding kinetics. On a qualitative level, however, this result indicated that NSP2-ΔC forms a highly stable complex with 20mer RNA.

RNA used	Sample	[NSP2] (nM)	K$_{on}$ (M^{-1}s^{-1})	K$_{off}$ (s^{-1})	K$_{D}$ (pM)	χ^2	ΔOFF⚓
20mer RNA	**NSP2 WT**	2.5	3.55x10^{6}	1.76x10^{-4}	49.5	0.215	N/A
		5	3.01x10^{6}	1.77x10^{-4}	58.8	0.071	
		10	2.16x10^{6}	2.50x10^{-5}	11.6	0.729	
	NSP2 ΔC	2.5	2.00x10^{4}	7.21x10^{-7}	36.1	0.428	**244***
		5	2.11x10^{4}	4.61x10^{-6}	218	0.339	**38.4***
		10	1.49x10^{4}	1.14x10^{-8}	0.765	0.457	**2190***
10mer RNA	**NSP2 WT**	6.25	1.37x10^{6}	9.04x10^{-4}	6580	1.742	N/A
		12.5	6.71x10^{5}	8.58x10^{-4}	1280	15.23	
		25	5.12x10^{5}	7.20x10^{-4}	1410	128.7	
	NSP2 ΔC	6.25	8.24x10^{5}	2.92x10^{-4}	354	15.85	**3.10**
		12.5	4.03x10^{5}	2.40x10^{-4}	596	30.43	**3.58**
		25	4.95x10^{5}	2.46x10^{-4}	497	72.05	**2.93**
10mer RNA *(previous study*)*	**NSP2 WT**	N/A	1.5x10^{6}	3.80x10^{-2}	25,300	N/A	7.34
	NSP2 ΔC	N/A	8.26x10^{4}	5.18x10^{-3}	62,700	N/A	

Table 5.1 Binding kinetics of NSP2 and NSP2-ΔC as measured by SPR Association (K$_{on}$) and dissociation (K$_{off}$) rate constants as obtained by SPR. *Previous study (Hu *et al.*, 2012) used different immobilisation strategy (3'-biotinylated RNA rather than 5'-biotinylated RNA, as used in this chapter. No χ^2 values were reported. *these values are unreliable due to mass transport effects and no observable dissociation within the timeframe of the experiment. ⚓ΔOFF corresponds to the fold-difference in dissociation rate (K$_{off}$) between NSP2 WT and NSP2 ΔC.

As affinities of nonspecific protein-RNA interactions increase with RNA length, a shorter 10mer RNA was used to improve NSP2 dissociation. SPR experiments using this shorter substrate show faster dissociation for both NSP2 and NSP2-ΔC, allowing accurate fitting of kinetic parameters (**Figure 5.21, Table 5.1**). Although on-rates remain constant across a range of NSP2 and NSP2-ΔC concentrations, NSP2-ΔC exhibited 3.2 ± 0.34-fold slower dissociation than NSP2, confirming the role of the CTH in dissociation of NSP2 from RNA.

Together, these results indicated that NSP2-ΔC has an enhanced unwinding activity (**Figure 5.3**) and remains stably associated with RNA for longer than WT NSP2 (**Figure 5.4**). As a consequence, NSP2-ΔC has reduced RNA annealing activity compared to WT NSP2 (**Figure 5.2**), indicating that the CTD may play a role in RNA dissociation from NSP2 that is essential for virus replication.

5.2.2 Structural studies of NSP2

Having established that the CTD of NSP2 is involved in both RNA unwinding and RNA-RNA matchmaking activities of NSP2, we sought to visualise its position and orientation within the context of a complete NSP2 ribonucleoprotein (RNP) complex using cryoEM.

In order to understand whether NSP2 is a suitable system for high-resolution structural analysis using cryoEM, we first aimed to determine a structure of NSP2 apoprotein. Although there have been several previous crystal structures of NSP2 solved (Jayaram *et al.*, 2002; Vasquez-Del Carpio *et al.*, 2004; Taraporewala *et al.*, 2006), a high-resolution solution structure of NSP2 would potentially resolve ambiguities regarding octamer-octamer interactions observed in crystallo (Hu *et al.*, 2012; Criglar *et al.*, 2018).

Grids were prepared with a range of NSP2 concentrations, from 5 μM to 50 μM using Quantifoil 1.2/1.3 holey carbon grids. The use of unsupported ice for NSP2 (i.e. holey carbon as opposed to continuous carbon) was chosen to minimise background signal arising from a carbon support. 5 μM NSP2 resulted in poor particle distribution, and 50 μM resulted in highly crowded particle distributions (**Figure 5.5**). 25 μM NSP2 produced an optimal distribution, where particles were abundant in the ice, but not overcrowded (i.e. overlapping particles) (**Figure 5.5**).

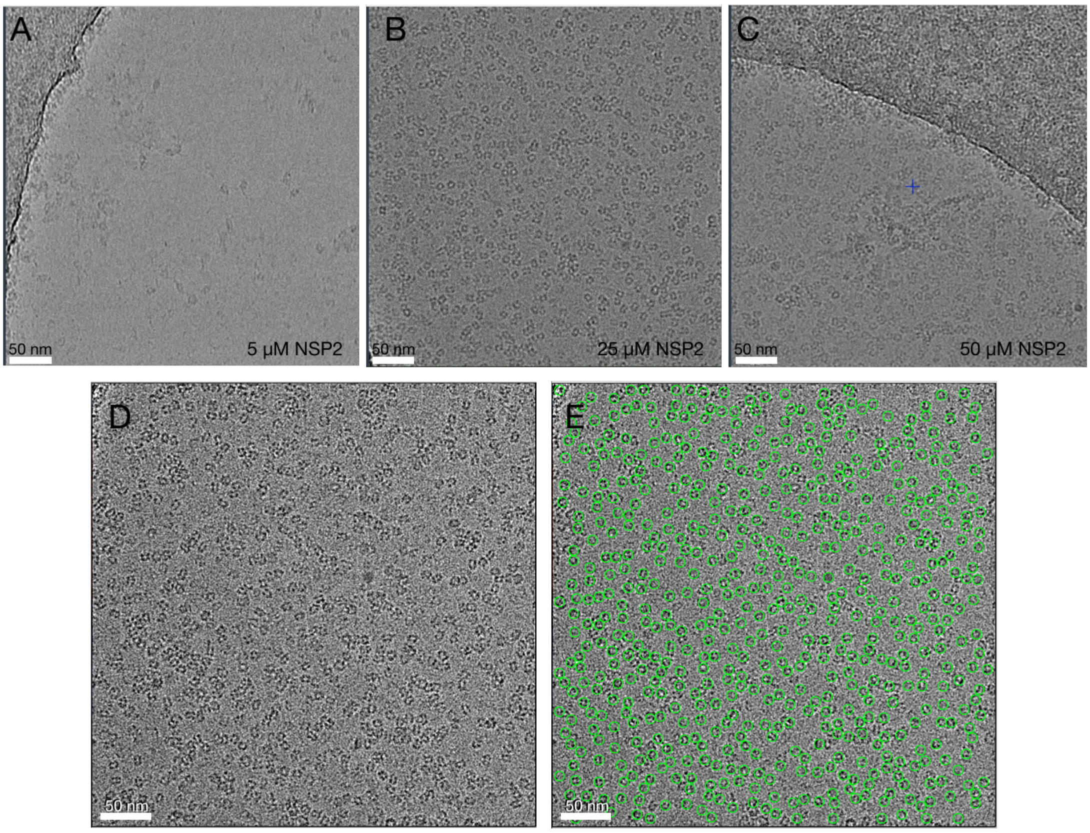


Figure 5.5 Representative NSP2 apoprotein micrographs. A – C: Typical micrographs of NSP2 apoprotein in unsupported ice (Quantifoil 1.2/1.3 grids). At a range of concentrations (5 – 50 μM). Scale bar = 50 nm. D: Example NSP2 (25 μM) micrograph collected on Titan Krios equipped with Falcon III detector. E: Particles picked using Laplacian-of-Gaussian (LoG) filtering and reference-free autopicking implemented in Relion 3 (Zivanov *et al.*, 2018).

High resolution data was subsequently collected in-house (Astbury Biostructure Laboratory) on a Titan Krios microscope operating at 300 kV, equipped with a Falcon III direct electron detector. 2428 micrograph movies were collected using FEI EPU automated data collection software, and pre-processed (micrograph import, motion correction and CTF correction) on-the-fly in Relion 3 (Zivanov *et al.*, 2018). Full NSP2 apoprotein cryoEM data processing workflow is outlined in **Figure 5.7**. LoG-based autopicking allowed rapid and accurate reference-free autopicking of 1,268,983 particles from 2428 micrographs (**Figure 5.5D&E**).

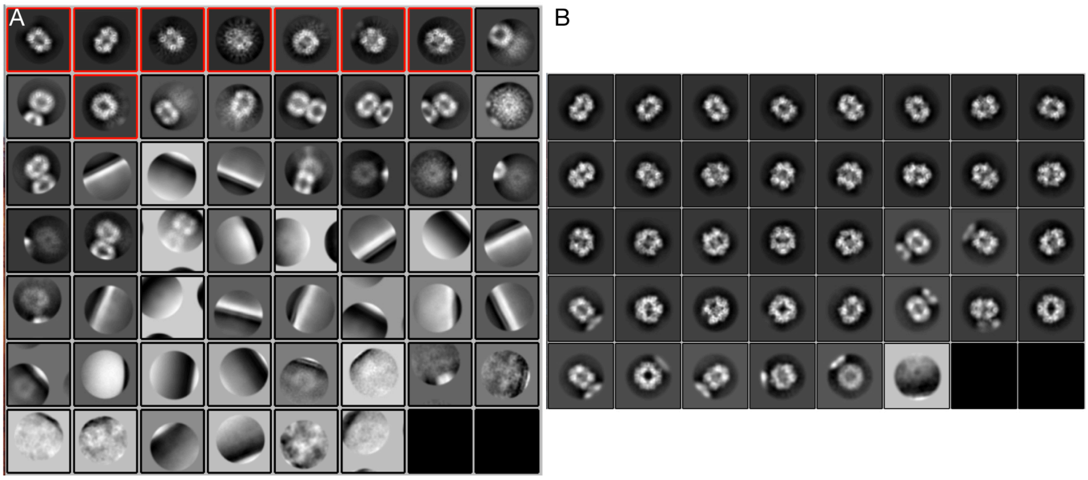


Figure 5.6 Representative 2D class averages of NSP2 apoprotein. A: NSP2 2D class averages of autopicked particles, classified using fast subsets. Subsets of particles corresponding to classes of NSP2 without overlapping particles were selected (red) and used for further downstream processing (955,893 from a total of 1,278,983 input particles). Box size: 272.6 Å B: Further 2D classification (no fast subsets) of selected particles showing multiple different view of NSP2. Box size: 272.6 Å Complete NSP2 apoprotein cryoEM processing workflow is outlined in **Figure 5.6**.

Autopicked particles were subjected to multiple rounds of iterative (≥25) reference-free 2D classification. Despite optimising autopicking parameters, it is impossible to completely eliminate selection of “junk particles” (ice contamination, edges of hole etc). To rapidly sort these false positives from bona fide particles, reference-free 2D classification using ‘fast subsets’ was performed (**Figure 5.6**). Highly populated 2D classes were selected (955,893 particles) and subjected to further, more rigorous (i.e. no fast subsets) reference-free 2D classification (**Figure 5.7**).

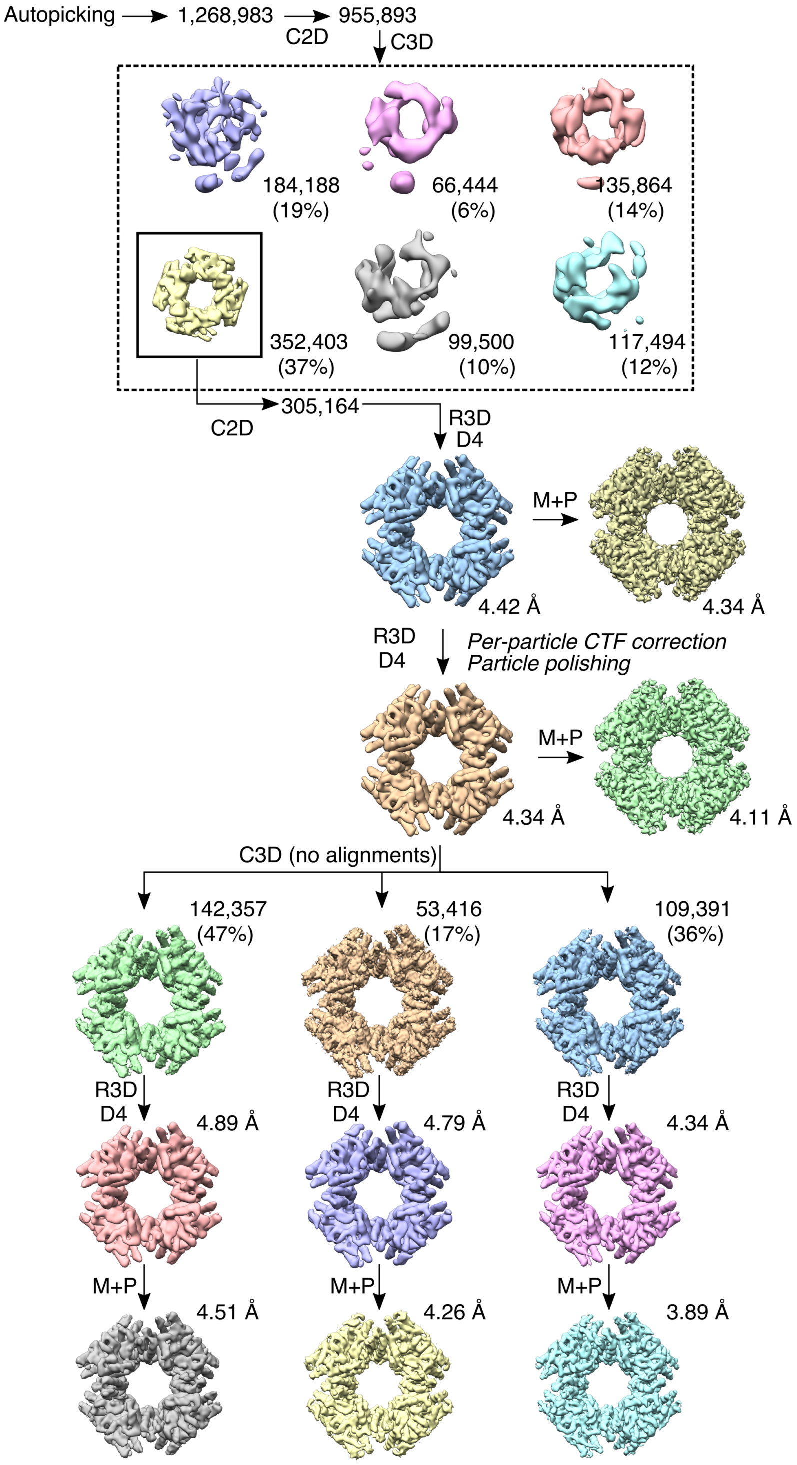


Figure 5.7 NSP2 apoprotein cryoEM image-processing workflow. Image processing workflow used to determine 3D reconstruction of NSP2 apoprotein. Initial rounds of 2D classification (C2D) and 3D classification (C3D) were performed using Fast Subsets (see **Chapter 3.2** for further details), and image alignments with 25 iterations. The highest quality 3D class average was subjected to further 2D classification without fast subsets, and used for 3D refinement (R3D), with D4 symmetry imposed. After per-particle CTF correction and particle polishing, further 3D classification was performed without fast subsets, and without performing image alignment. 3D reconstructions were ultimately subjected to masking and post-processing (M + P).

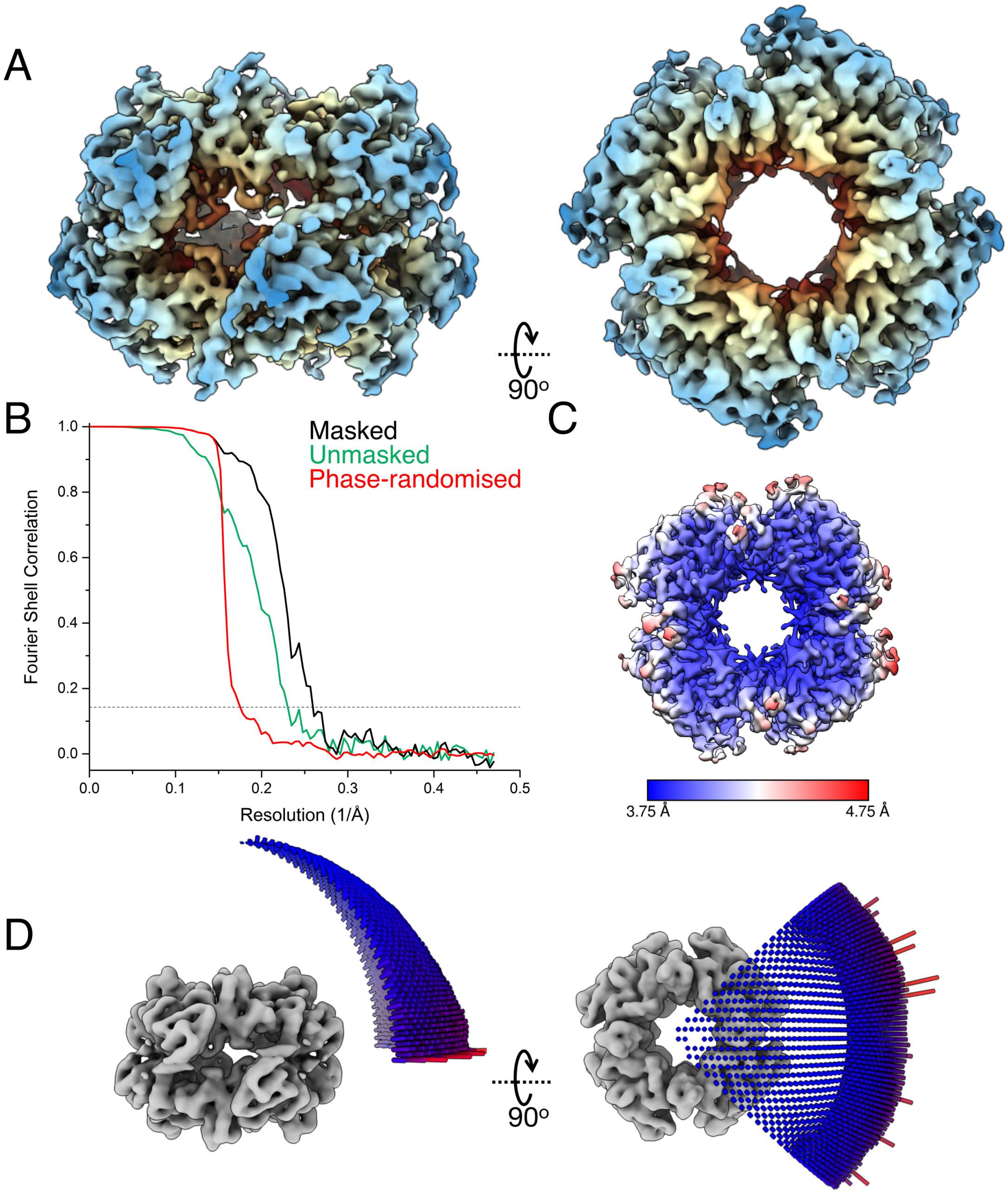


Figure 5.8 3D reconstruction of NSP2 apoprotein. A: 3D reconstruction of NSP2 with D4 symmetry. B: Fourier Shell Correlation (FSC) curve showing FSC amplitude as a function of 1/resolution (Å). “Gold-standard” (0.143) threshold. FSC curves from unmasked and masked pre- and post-“particle polishing” (per-particle CTF correction and Bayesian motion correction) are shown as light green (unmasked, unpolished), dark green (unmasked, polished), grey (masked, unpolished) and black (masked polished). Masked random-phase data (red) and corresponding resolutions are indicated. Final resolution: 3.9 Å. C: Local resolution of NSP2 apoprotein octamer. D: Euler angle distribution.

The angular distribution plot (**Figure 5.8**) indicates that NSP2 does not have strong orientation bias. Although there are highly populated orientations (**Figure 5.8D, red**), this is likely an artefact as these are at the edge of the asymmetric unit. If there was orientation bias, then there would be stretched density, which is clearly not the case when the cryoEM map is compared to the NSP2 crystal structure (**Figure 5.9**). Additionally, if there was preferred orientation, then a cluster of highly populated orientations would be observed rather than ~10 individual, highly represented orientations (Tan *et al.*, 2017).

A crystal structure of NSP2 octamer (PDB 1L9V) was rigid-body fitted into the cryoEM density map using UCSF Chimera (**Figure 5.9**) (Pettersen *et al.*, 2004). As anticipated, the atomic model fitted well into the density map. Interestingly, the density corresponding to the C-terminal domain of NSP2 (**Figure 5.9, red**) is in a 'closed' conformation rather than the flipped out 'open' conformation (**Figure 5.9B & C, red**). The open conformation observed *in crystallo* allows the C-terminus to participate in domain swapping with neighbouring NSP2 octamers. This result indicates that this alternate C-terminal conformation is likely to be a crystallographic artefact.

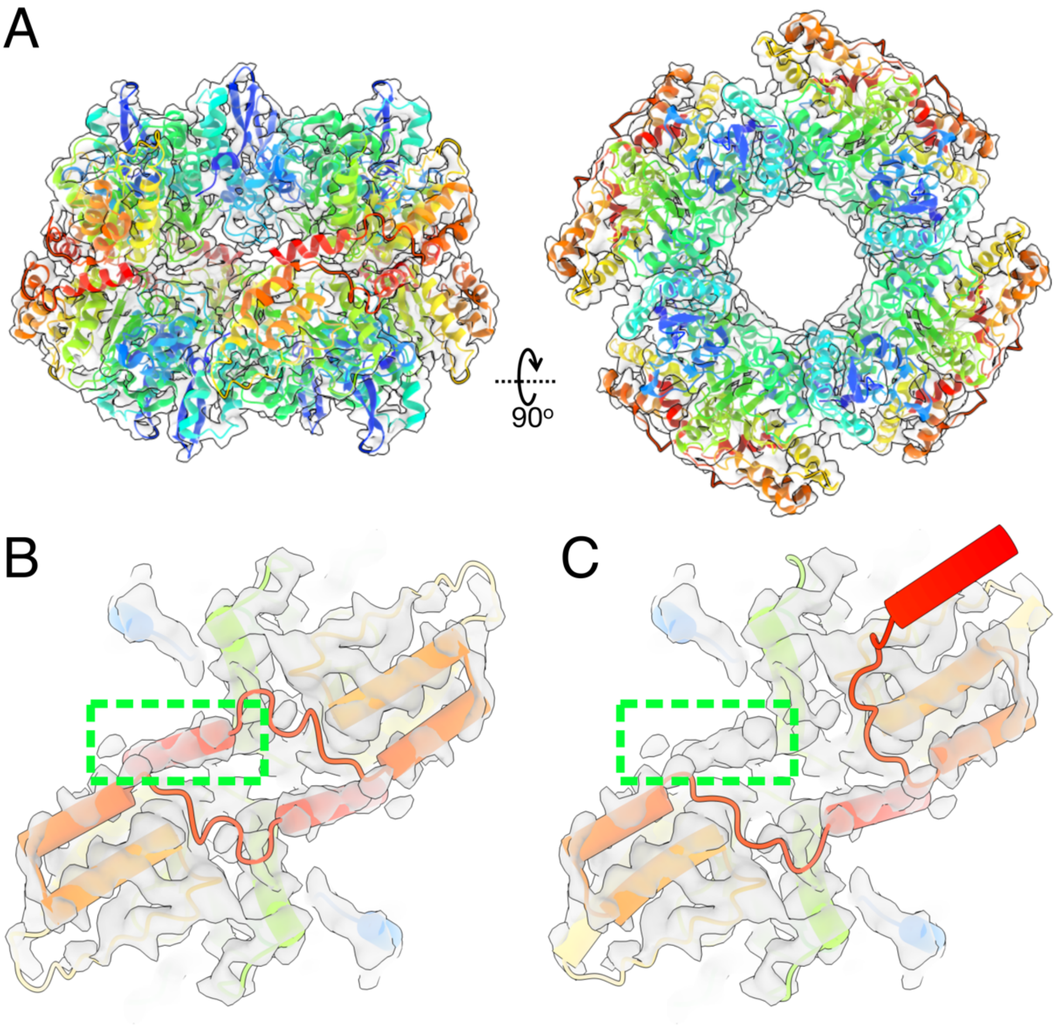


Figure 5.9 Comparison of NSP2 cryoEM map with atomic model. A: Rigid-body fitting of NSP2 octamer crystal structure (PDB 1L9V) into NSP2 apoprotein cryoEM density map using UCSF Chimera, and visualised using UCSF ChimeraX (Pettersen *et al.*, 2004; Goddard *et al.*, 2018). Atomic model of NSP2 is coloured by sequence, from N-terminus (blue) to C-terminus (red). B & C: Comparison of cryoEM density map (grey) to fitted atomic models of C-terminus in closed conformation (B, PDB 1L9V) and open conformation (C, PDB 4G0A). Green box indicates cryoEM density corresponding to C-terminus.

Having determined a 3.9 Å resolution cryoEM reconstruction of NSP2 apoprotein, we next sought to determine a reconstruction of an NSP2 ribonucleoprotein (RNP) complex.

NSP2 RNP complex was assembled by incubating NSP2 with 40mer RNA. A range of different NSP2 : RNA ratios were tested, but ratios above 1 RNA : 2 NSP2 monomer resulted in immediate complex precipitation. As NSP2 contains four putative RNA-binding sites (i.e. positively charged 15Å-deep grooves present on the surface of NSP2 (**Figure 5.1**)), a ratio of 1 RNA : 2 NSP2 monomer (i.e. 4 RNA per NSP2 octamer) should ensure saturation of NSP2 with RNA (Schuck *et al.*, 2001; Jayaram *et al.*, 2002).

To establish optimal cryoEM grid conditions, grids with a range of NSP2 RNP concentrations were made. RNP complex was incubated at room temperature (25°C) for 30 minutes prior to application to Quantifoil 1.2/1.3 holey carbon grids and vitrification. At 25 μM RNP complex (~1 mg/ml), no particles were present in the ice; this was probably due to preferentially resting on the carbon instead of distributing evenly throughout the ice (**Figure 5.10**). This was ameliorated by a 2-fold increase in the RNP concentration (while maintaining a 1 RNA : 2 NSP2 monomer molar ratio), producing a reasonable (albeit somewhat crowded) particle distribution within the ice (**Figure 5.10, 5.11**). This drastic change in particle distribution arising from a small increase in sample concentration suggests that at 50 μM RNP, the carbon had been fully saturated, allowing particles to sit within the ice. High resolution data was subsequently collected in-house (Astbury Biostructure Laboratory) on a Titan Krios microscope equipped with a Gatan K2 direct electron detector. 4050 micrographs were collected using FEI EPU automated data collection software, and pre-processed (micrograph import, motion correction and CTF correction) on-the-fly in Relion 3 (Zivanov *et al.*, 2018).

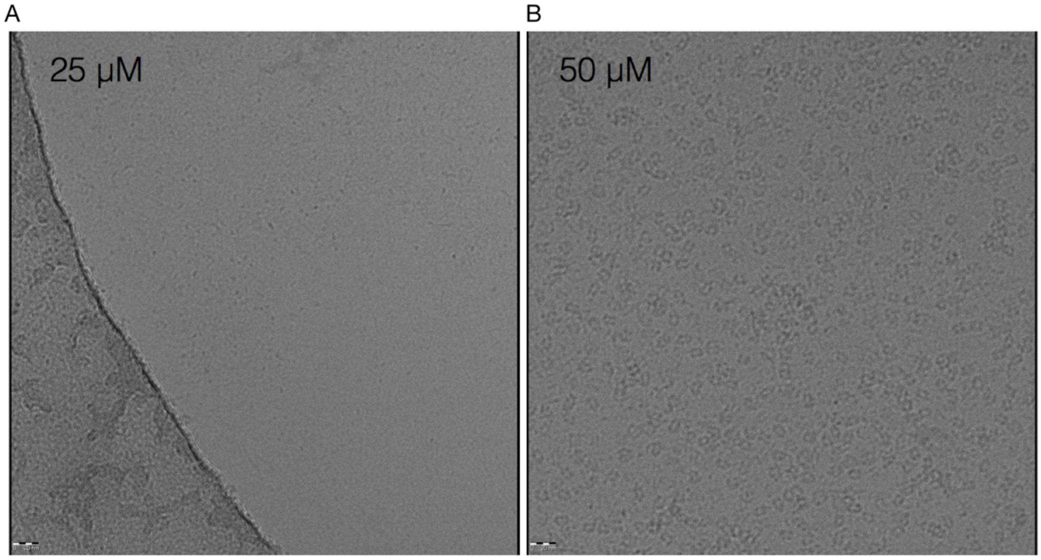


Figure 5.10 Examples of NSP2 RNP complex micrographs. Typical micrographs of NSP2 – 40mer RNP complexes in unsupported ice (Quantifoil 1.2/1.3 grids). A: 25 *µ*M RNP complex showed few particles in ice, as they likely have affinity for the amorphous carbon. B: 50 *µ*M RNP complex shows uniform, albeit somewhat crowded distribution. Scale bar: 20 nm.

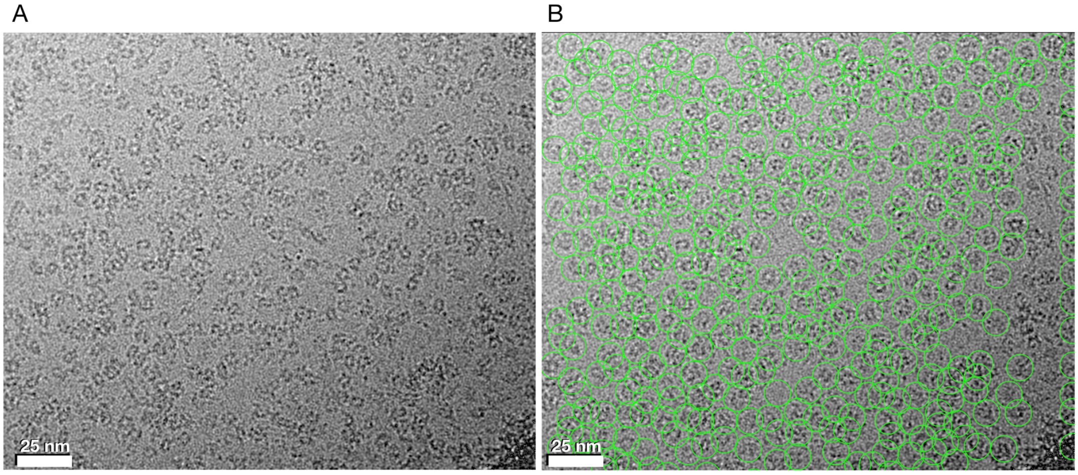


Figure 5.11 Example of NSP2 RNP cryoEM data. A: Example NSP2 RNP micrograph collected on Titan Krios equipped with K2 detector. B: Particles picked using Laplacian-of-Gaussian (LoG) filtering and reference-free autopicking implemented in Relion 3.

Particles were autopicked using LoG methods as described in **Materials and Methods**, resulting in 1,609,482 particles. These particles were subjected to multiple rounds of iterative (≥25) reference-free 2D classification, as described above for NSP2 apoprotein. Highly populated 2D classes corresponding to NSP2 were selected (1,245,861 particles) and subjected to further, more rigorous (i.e. no fast subsets) reference-free 2D classification (**Figure 5.13**). Suitable classes were selected and used to generate an initial 3D model as a reference for 3D classification. 635,599 particles contributing to the 3D class average with the highest resolution (7.93 Å) (which coincidentally also contained the largest fraction of the input particles – 50.2% of the cleaned stack of input particles) were selected and used to generate a 3D reconstruction (**Figure 5.13**).

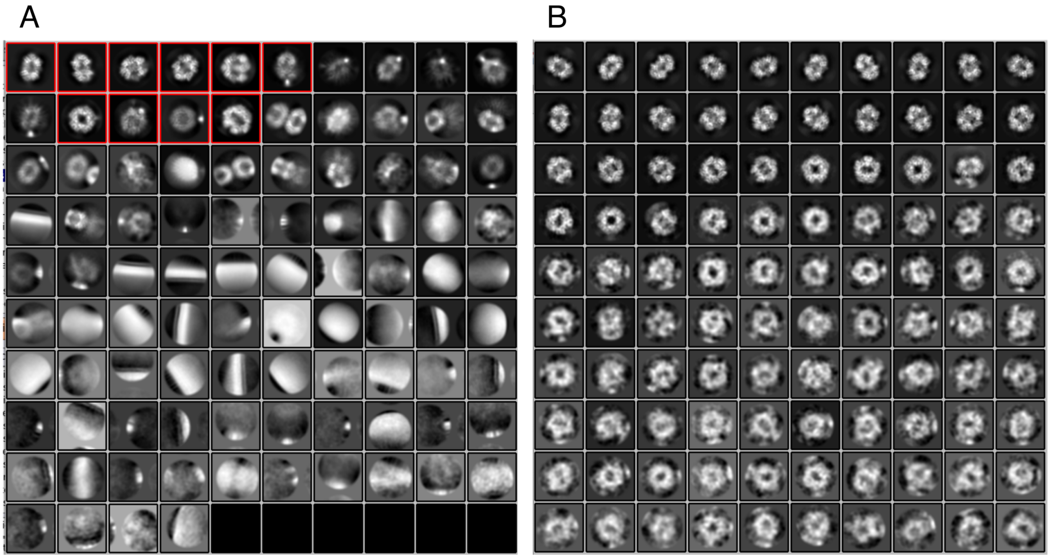


Figure 5.12 Representative 2D class averages of NSP2 RNP. A: NSP2 RNP 2D class averages of autopicked particles, classified using fast subsets. Particles selected (red) were used for further downstream processing (~74% of input particles). Box size: 214 Å B: Further 2D classification (no fast subsets) of selected particles showing multiple different view of NSP2. Box size: 214 Å Complete RNP cryoEM processing workflow is outlined in **Appendix 5.2**.

From these 635,599 particles, two different data processing strategies were used to determine a high resolution cryoEM density map. First, this subset particles were subjected to a further round of C3D (without fast subset), from which the most highly populated class was selected and subjected to C2D (without fast subsets). Poorly aligned particles were removed, producing a subset of 156,095 particles, which were used to determine a 3D reconstruction at 3.8 Å resolution (**Figure 5.13**) (with C4 symmetry applied).

A second strategy was undertaken, where the entire subset of particles selected from the first round of C3D were used for R3D (with C4 symmetry). This larger subset of particles allowed determination of a 3.51 Å reconstruction, which was improved to 3.46 Å after per-particle CTF correction and Bayesian particle polishing (Zivanov *et al.*, 2018) (**Figure 5.13**).

In order to improve the resolution of this map, a final round of C3D was performed without image alignments (**Materials and Methods**). Although the vast majority of the particles were grouped into the same class (86%), refinement and masking did not increase the resolution of this reconstruction beyond what had already been attained. However, a small subset (6%) of the input particles formed a highly homogeneous class that was refined to 3.41 Å resolution (**Figure 5.13**).

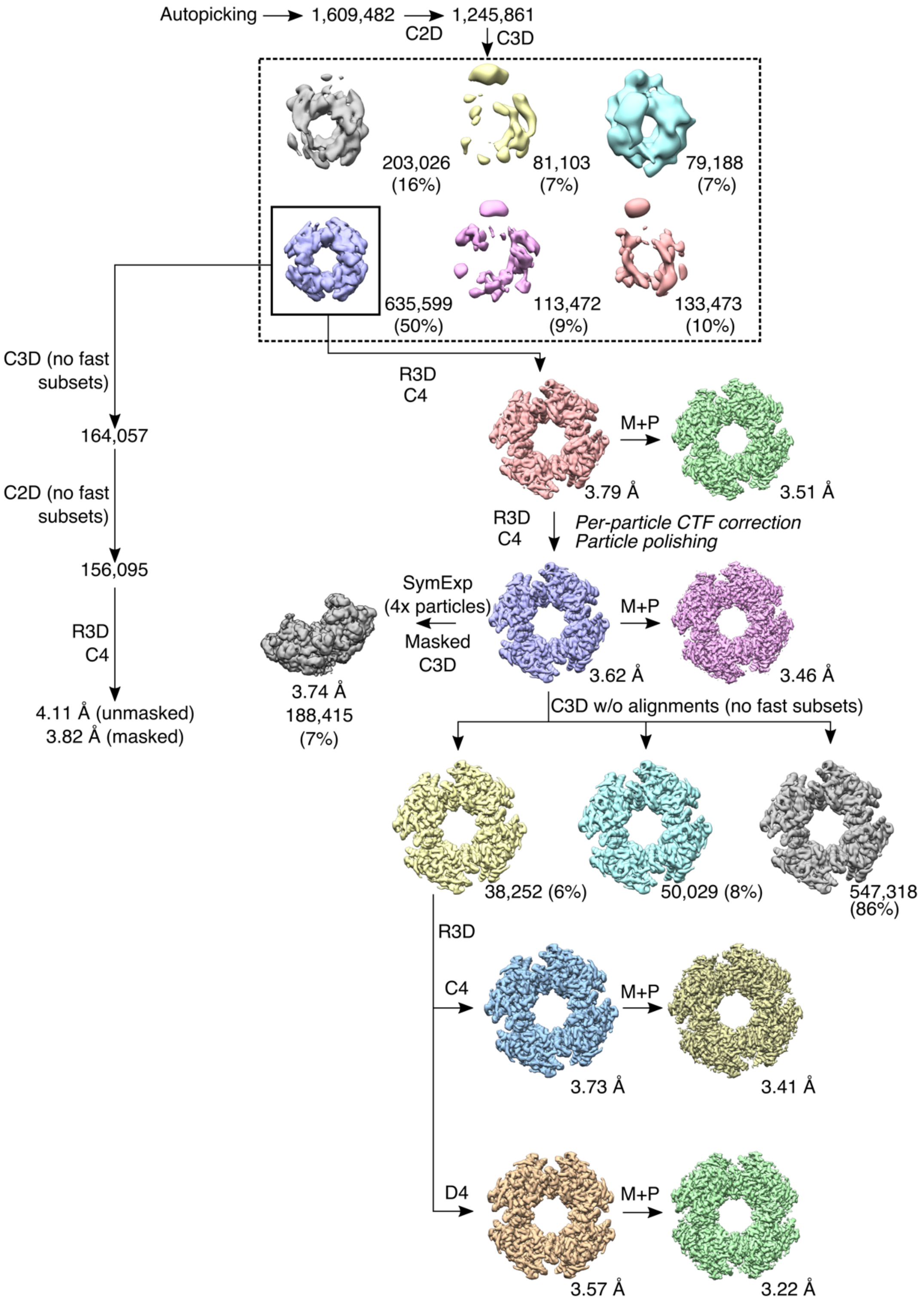


Figure 5.13 NSP2 RNP cryoEM image-processing workflow. Image processing scheme used to determine 3D reconstruction of NSP2 RNP. Initial 2D and 3D classifications (C2D and C3D, respectively) were performed with fast subsets.

Despite the inherent 4-fold dihedral (D4) symmetry of NSP2, density from bound RNA will, at most, have 4-fold cyclic (C4) symmetry (i.e. RNA density arising from each positively-charged groove on NSP2, but without an additional equatorial 2-fold symmetry axis). 3D reconstructions were performed with C1 (i.e. with no symmetry applied), C4 and D4 symmetries to examine the effect of symmetry-related artefacts. Manual inspection did not reveal any structural inconsistencies between reconstructions with different symmetries applied (apart from slight differences in resolution).

However, as the RNA density was weak in all RNP reconstructions (and was not present after masking and post-processing), a final reconstruction was performed with D4 symmetry, improving the global resolution to 3.2 Å (as defined by the FSC threshold of 1.43) (Scheres and Chen, 2012) (**Figure 5.13, 5.14**). The C4 reconstruction was used for symmetry expansion with focused classification, which will be discussed in detail later in this chapter.

Local resolution of the density map was calculated in Relion (Scheres, 2016). This reconstruction has local resolution ranging from 3.2 Å (core) to 4.1 Å (periphery) (**Figure 5.14**). This is a common feature of cryoEM reconstructions of globular particles, as different orientations are most easily aligned by their centre-of-mass (i.e. core).

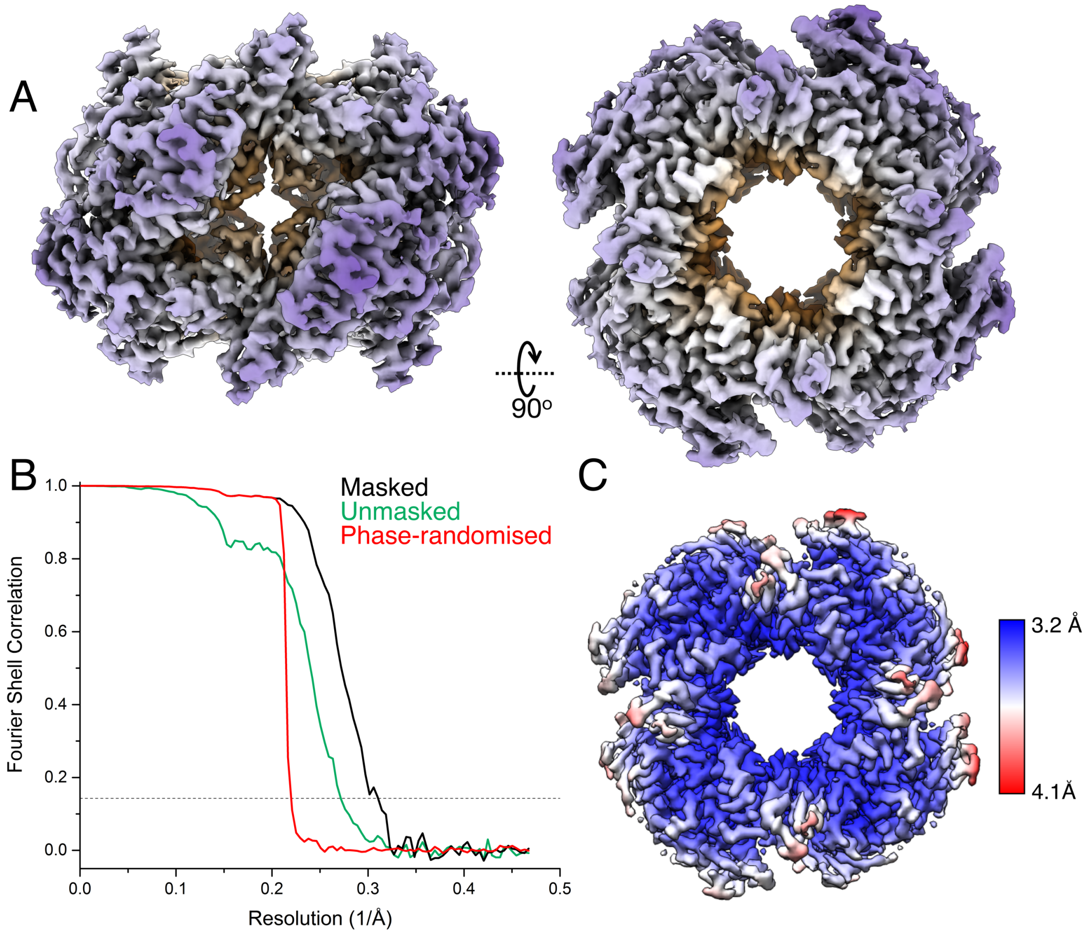


Figure 5.14 3D reconstruction of NSP2 RNP. A: 3D reconstruction of NSP2. B: Fourier Shell Correlation (FSC) curve showing FSC amplitude as a function of 1/resolution (Å). “Gold-standard” (0.143) threshold. FSC curves from unmasked and masked pre- and post-“particle polishing” (per-particle CTF correction and Bayesian motion correction) are shown as light green (unmasked, unpolished), dark green (unmasked, polished), grey (masked, unpolished) and black (masked polished). Final resolution: 3.2 Å. Masked random-phase data (red) and corresponding resolutions are indicated. C: Local resolution of NSP2 RNP octamer. D: Euler angle distribution

The Euler angle distribution indicates that there were no significantly preferred orientations (**Figure 5.15**). This is also clear from the lack of smeared or stretched density of NSP2 that would originate from directional resolution anisotropy (Tan *et al.*, 2017). This is even clearer when comparing the angular distributions between reconstructions performed with D4 and C4 symmetry (**Figure 5.15**). The high degree of similarity between these two reconstructions indicates that there are no artefacts due to enforcing D4 symmetry in the 3.2 Å resolution NSP2 RNP reconstruction.

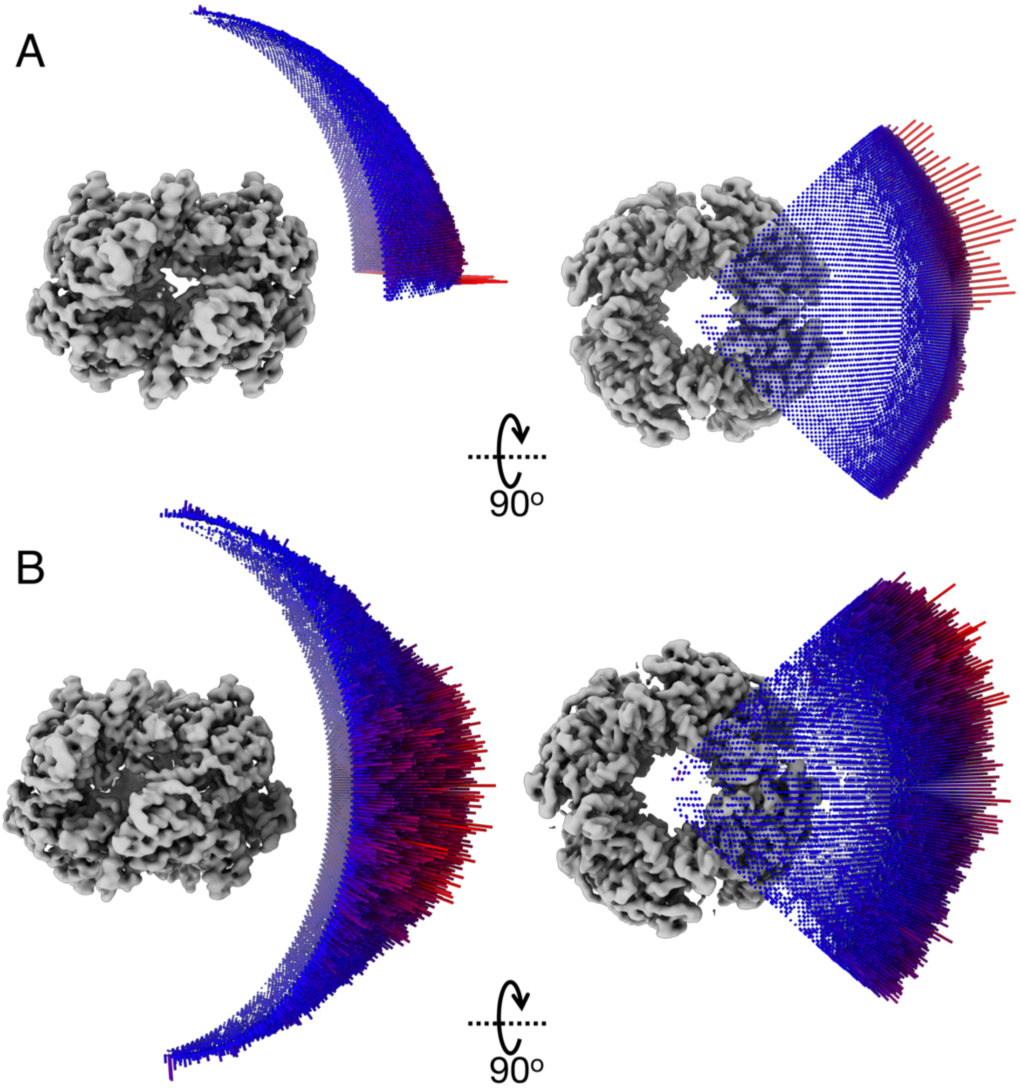


Figure 5.15 Angular distribution plots of NSP2 RNP 3D reconstructions. Highly populated orientations occurred at the edge of the equatorial 2-fold symmetry axis of NSP2 when reconstructed in D4 symmetry (A). These were more dispersed when reconstructed in C4 symmetry (B).

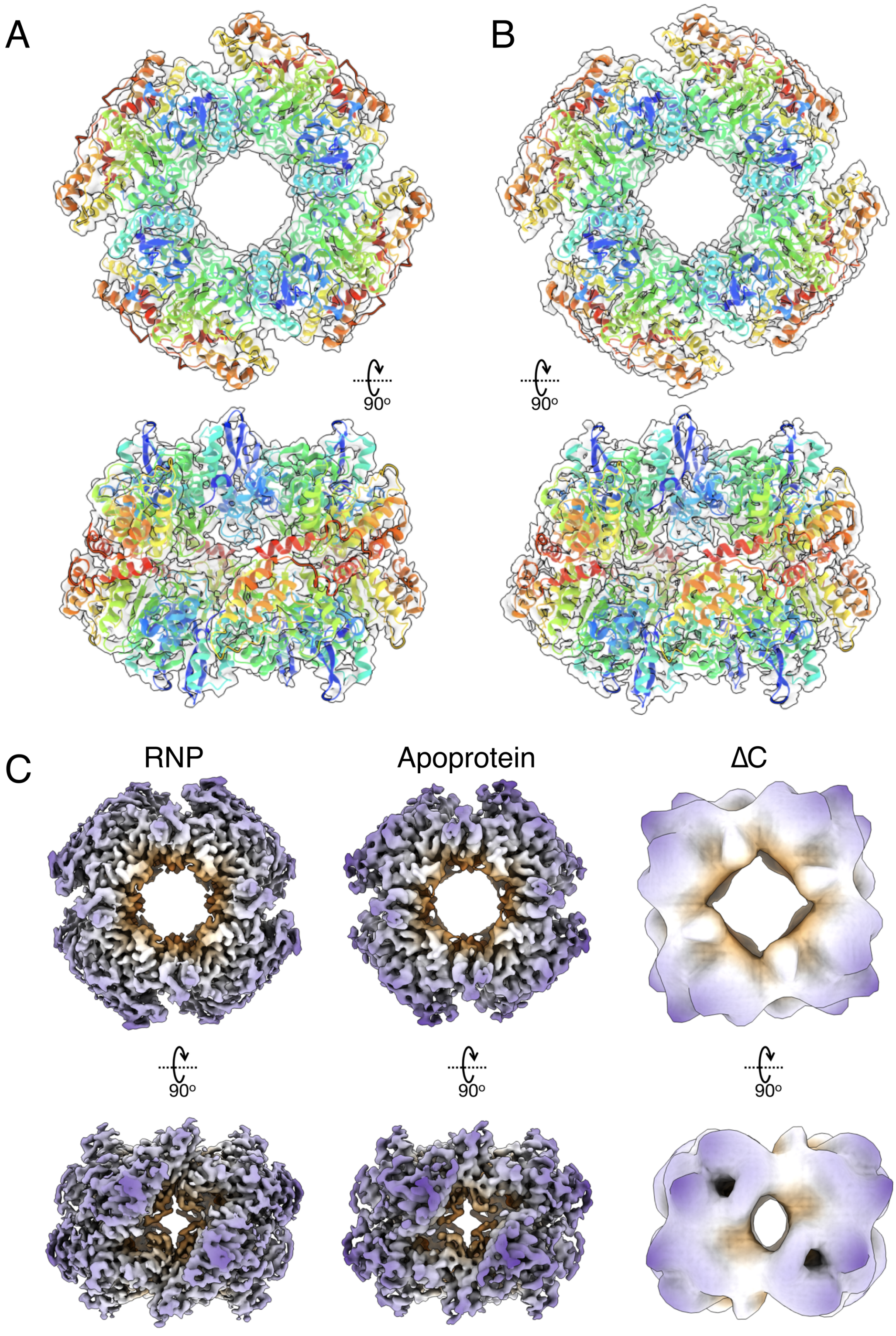


Figure 5.16 Comparison between NSP2 RNP and apoprotein cryoEM density maps NSP2 atomic structure (PDB 1L9V) fitted into NSP2 apoprotein (A) and RNP complex (B) density maps. C: Comparison between NSP2 RNP, apoprotein and NSP2-ΔC (negative stain EM) 3D reconstructions. Despite the inherent low resolution of the NSP2-ΔC reconstruction, it is clearly still octameric and has dimensions consistent with cryoEM reconstructions of WT NSP2 and its RNP complex.

NSP2 crystal structure (1L9V) was rigid-body fitted into the respective density maps (**Figure 5.16**). In both cases, the density maps correspond to the previously determined crystal structures, indicating that NSP2 does not undergo a conformational change upon RNA binding.

A structure of NSP2-ΔC was also attempted. However, this was hindered by substantial aggregation of both NSP2-ΔC and NSP2-ΔC-RNA upon vitrification. As an alternative, a negative stain EM reconstruction of NSP2-ΔC was determined (**Figure 5.16**). While this obviously lacks the spatial resolution of the cryoEM maps, this reconstruction confirms that NSP2-ΔC is folded and octameric. The toroidal architecture and equatorial groove of NSP2-ΔC is also consistent with cryoEM reconstructions (**Figure 5.16**).

A previous cryoEM reconstruction determined in 2006 of NSP2 RNP complex (8.5 Å) and flexible fitting analysis at the time had indicated that NSP2 may undergo "significant conformational changes" upon RNP complex formation (Jiang *et al.*, 2006). Our high-resolution density map (reconstructed from ~40,000 particles from >4000 micrographs taken on a direct electron detector compared to ~5,000 particles from 24 CCD images in the previous study) unequivocally shows that this is not the case (**Figure 5.16**). This artefact has likely arisen from over-fitting a structure into a low-resolution, low-quality map. It is also likely that the resolution of the previous reconstruction was overestimated, as this was estimated using alternative procedures (Scheres and Chen, 2012), and lacks the structural features expected of a reconstruction at this resolution (alpha helices should be resolved at ~8 Å resolution). In reality, the previous reconstruction is likely at 10-20 Å resolution, and thus could only serve for rigid body fitting.

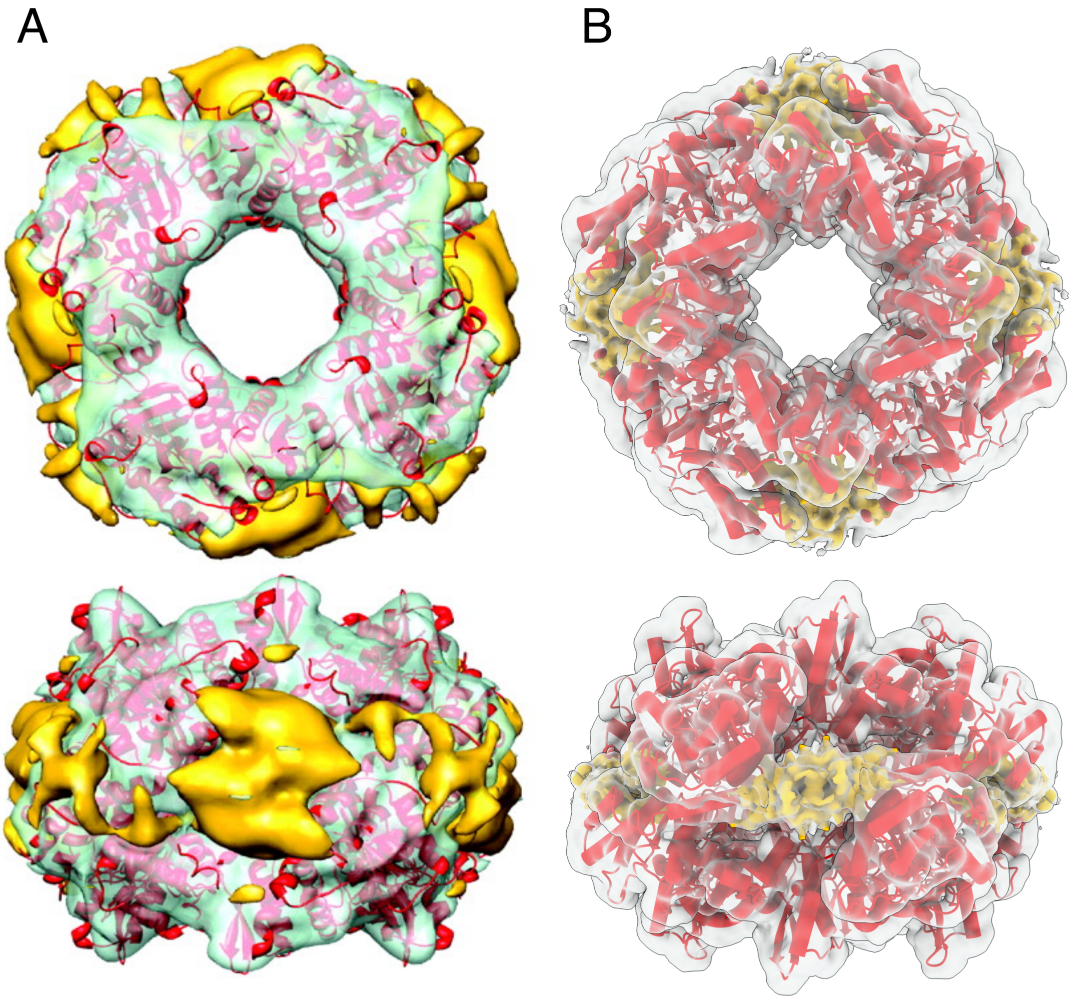


Figure 5.17 Fitting of NSP2 crystal structures into RNP cryoEM density maps. A: Previously determined NSP2 RNP cryoEM reconstruction, taken from (Jiang *et al.*, 2006). RNA difference density is gold. B: NSP2 octamer (PDB 1L9V, red) fitted into RNP density map from this study (transparent grey). RNA difference density is gold, computed by subtracting a 3.7 Å resolution simulated surface of NSP2 apoprotein (from PDB 1I9V) from RNP complex cryoEM density map.

While the localisation of the RNA density is largely similar between the two reconstructions (**Figure 5.17**), the high-resolution reconstruction presented here has RNA density localised exclusively to the basic groove (as opposed to weak, diffuse density around the entire NSP2 equator). Although this is a significant improvement on the previous reconstruction, this density is still rather diffuse, and unsuitable for fitting or building an atomic model of RNA into.

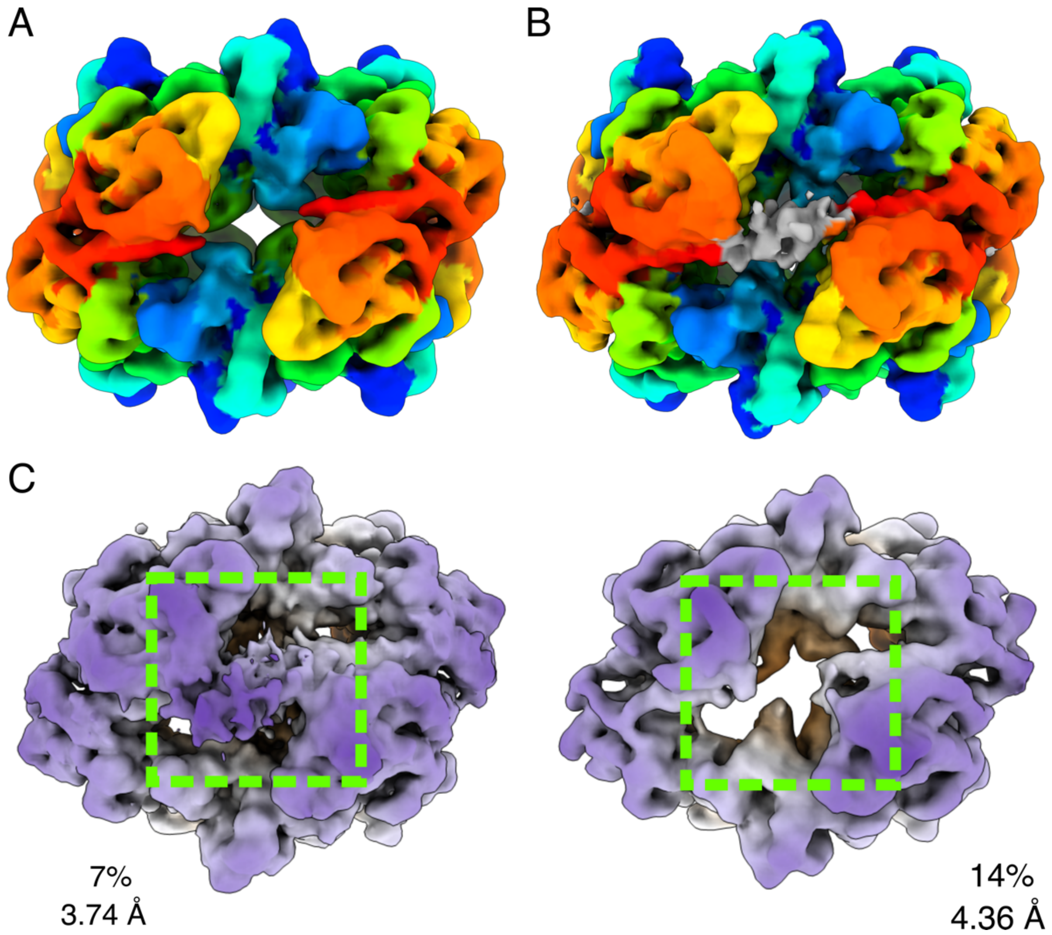


Figure 5.18 NSP2 RNP complex RNA density. A & B: Low-pass filtered (5Å) cryoEM maps of NSP2 apoprotein (A) and RNP (B). Maps are rainbow coloured according to corresponding spequence position (N-terminus: blue, C-terminus: red). Novel RNA density in RNP map (B) corresponding to RNA is grey. C: Examples of 3D-classified symmetry-expanded 3D class averages showing the presence (left) and absence (right) of putative RNA density. Percentage of total symmetry expanded particles and estimated resolutions are shown. Site of putative RNA density is highlighted by a green box.

To further validate this putative RNA density, NSP2 apoprotein and RNP reconstructions were low-pass filtered to 5 Å (**Figure 5.18**). The absence of density within the grooves of NSP2 apoprotein further supported the notion that this density corresponded to RNA.

The diffuse nature of the RNA density likely arises from issues relating to differences in RNA occupancy of bound RNA, binding of RNA in multiple registers and orientations poor alignment of RNA density and a dynamic RNA binding mode (Orlova and Saibil, 2011)), no density corresponding to RNA was present after map post-processing with appropriate B-factors applied (**Figure 5.14**) (Rosenthal and Henderson, 2003).

As the protein components of the particle represent the strongest features of the cryoEM map and therefore can be easily aligned during a reconstruction, the weaker RNA density is effectively averaged out. To overcome this, the C4-processed map was subjected to symmetry expansion and subsequent focused classification without alignments (**Figure 5.18**).

Symmetry expansion was used to artificially expand the larger subset of particles 4-fold (from 635,599 particles to 2,542,396 quarter-particles) (**Figure 5.17**). This meant that each particle was replicated and rotated 90° around the 4-fold symmetry axis of NSP2, according to the previously determined Euler angles. Masked 3D classification of a single asymmetric unit (i.e. RNA-binding groove) of NSP2 was performed without particle alignment. By skipping alignments at this stage, the relative orientations of each particle in the reconstruction are taken from the previous, unexpanded reconstruction and therefore the 3D classes will only show heterogeneity within the area of the mask (rather than within the entire octameric RNP). The resulting 3D classes varied in global resolution and showed heterogeneous RNA density, likely due to the non-specific nature of the NSP2-RNA interaction, in addition to differences in occupancies and the multiple different registers and orientations adopted by the bound RNA as described above. Examples of 3D classes with relatively strong and weak RNA densities are shown (**Figure 5.18**). This strategy produced classes with the strongest RNA density features.

Although the quality of the bound RNA density was improved using symmetry expansion and focused classification, the RNA-binding site of NSP2 was still somewhat ambiguous. The diffuse nature of this density prevented the building of atomic models of RNA into the cryoEM reconstruction, and the density from RNA corresponds to (at most) 5-nt of the 40-nt RNA used to assemble this RNP. To validate the RNA-binding site of NSP2, UV crosslinking (UV-XL) and mass spectrometry (MS) were used to map protein-RNA interactions (**Figure 5.19**).

An in vitro adaptation of the RBDmap approach (Castello *et al.*, 2016; Zhang *et al.*, 2019) was used to identify RNA-binding sites. Identification of RNA-crosslinked peptides using mass spectrometry is not a trivial process – the heterogeneity of crosslinked RNA moieties can obfuscate their identification using tandem MS-MS. The RBDmap approach ameliorates this by allowing the identification of peptides adjacent to directly crosslinked peptides, which is a much more straightforward task. In RBDmap, crosslinked RNP complexes are first subjected to proteolysis using a highly specific protease (in this case ArgC, which cleaves at the C-terminus of arginine residues). Non-crosslinked peptides are then removed through several stringent wash steps, where RNA-peptide adducts are trapped using magnetic oligo(dT_{20}) beads via hybridisation of a poly$(A)_{25}$ sequence. Enriched RNA-protein adducts are eluted and peptides are released using RNase A and trypsin digestion, and analysed by MS. The sub-peptides adjacent to the directly crosslinking peptides will be readily identified and used to identify the peptides that directly crosslink to RNA. In this case, NSP2 was crosslinked with in vitro transcribed rotavirus segment 11 ssRNA containing a 25-nt 5'-polyA sequence.

Crosslinked peptides were identified in three independent in vitro RBDmap replicates and were mapped to the sequence of NSP2. Mapping of these peptides to the surface of the octameric NSP2 crystal structure reveal that RNA-binding peptides are clustered around the basic groove of NSP2. Cross-linked peptides were dispersed throughout the entirety of NSP2 with enriched clusters of peptides occurring at positions 39 – 60, 80 – 104, 135 – 210 and 280 – 286. Residues 169 – 210 showed the highest crosslink occurrence (**Figure 5.19**). These UV-XL results strongly corroborate the RNA-binding site identified by cryoEM.

A crystal structure of NSP2 was rigidly fitted (i.e. no flexible remodelling) into the symmetry expanded, focussed reconstruction of NSP2 RNP using UCSF Chimera (Pettersen *et al.*, 2004). This allowed visualisation of RNA-NSP2 contacts without introducing bias from building an NSP2 model into cryoEM density that does not unambiguously correspond to NSP2 (**Figure 5.20**). Interestingly, several RNA densities contacting positively-charged side-chains were observed. Lys58, Lys59, Arg60 and Arg68 were oriented towards RNA density, with residues Arg60 and Arg68 making direct contacts with the RNA density. Interactions between RNA and residues 58 – 60 were identified by RBDmap, and Arg68 is adjacent to crosslinked peptides (69 – 78 and 69 – 104).

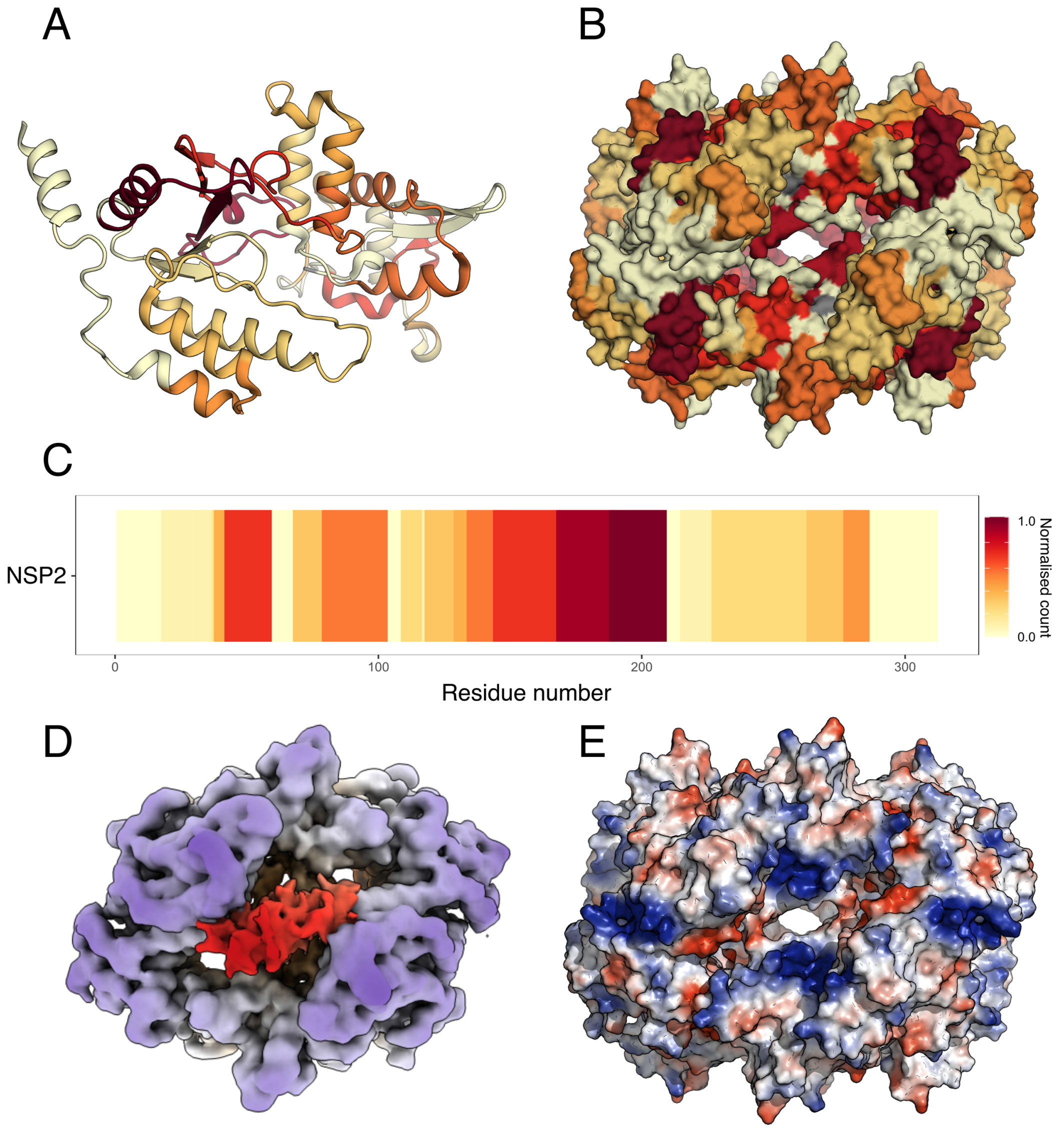


Figure 5.19 Mapping RNA-binding site of NSP2 A: Occurrence of cross-linked peptides as a function of position within NSP2 sequence. B: Peptides as showing in (A) mapped to the surface of NSP2 crystal structure (PDB 1L9V). C: The same RNA-binding face from cryoEM reconstruction. Density is segmented to RNA (red) and protein (purple). D: Symmetry expansion and focussed 3D classification cryoEM density of NSP2, with RNA density segmented and shown in red. Density corresponding to NSP2 is purple. E: Electropositive surfaces are blue, and electronegative surfaces are red, as calculated using PyMol.

The RNA-binding site of NSP2 sits atop a 2-fold symmetry axis (i.e. the interface between the two stacked NSP2 tetramers), and Arg60 and Arg68 residues from both NSP2 monomers both appear to make similar contacts with RNA. These residues are located within an unstructured loop region (60

– 73) of NSP2 (**Figure 5.19**). It is common for RNA-binding proteins to utilise flexible or unstructured regions to capture RNA (D'Souza and Summers, 2004).

Interactions between positively-charged side-chains and RNA is also a common trope of non-specific protein-RNA interactions, and this may provide a molecular basis for the lack of sequence specificity within the NSP2-RNA interaction. Although an exclusively electrostatic NSP2-RNA interaction is a tempting hypothesis, previous studies have demonstrated that this is not the case (under physiological ionic conditions, there is only a ~15% electrostatic contribution to the overall free energy of binding (Bravo *et al.*, 2018)). However, this does not contradict the NSP2-RNA interactions observed within this cryoEM density map, as arginine residues can interact both through electrostatic interaction (with the RNA phosphodiester backbone) and through non-electrostatic hydrogen bonds (with RNA nucleotide moieties).

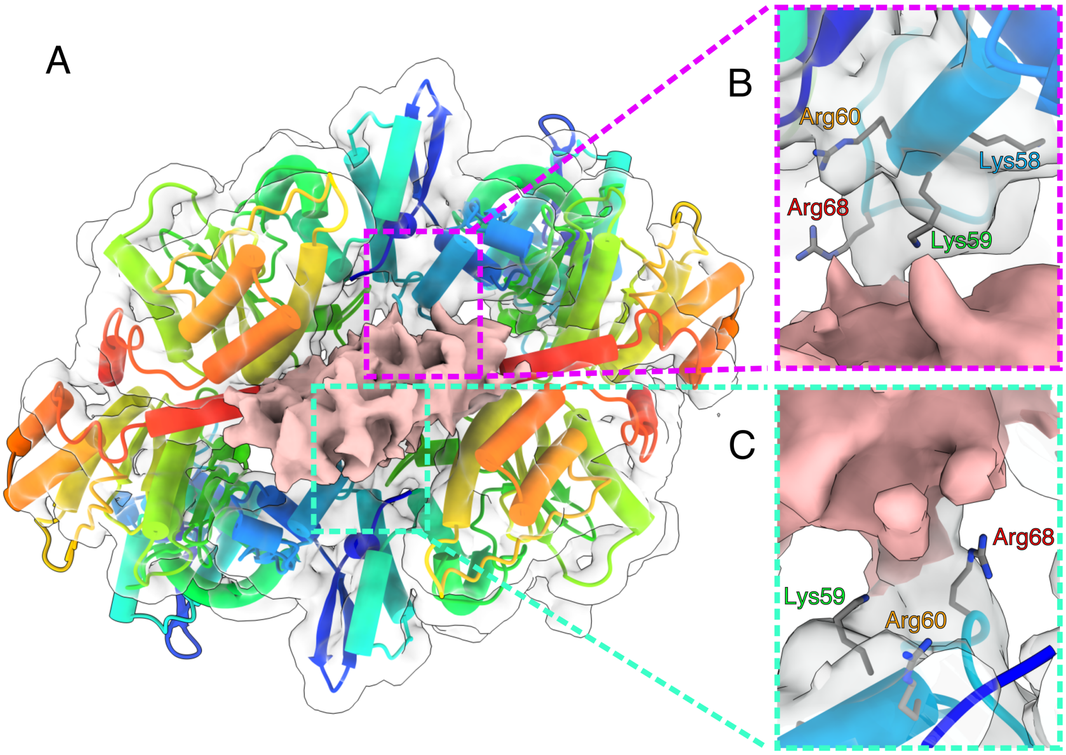


Figure 5.20 NSP2-RNA interactions A: Segmented cryoEM map showing NSP2 density (grey), RNA density (pink) and fitted NSP2 crystal structure (PDB 1L9V) (rainbow progressing from blue N-terminus to red C-terminus). B: NSP2-RNA contacts arising from positively-charged residues (cartoon backbone red, with side chains coloured according to atom identity; grey – carbon, blue – nitrogen). C: RNA contacts arising from a different NSP2 monomer than in (B)

Another purpose for determining the structure of the NSP2 RNP complex was to visualise the orientations of the CTH in relation to RNA and to understand what contacts (if any) that they made. UV-XL experiments (**Figure 5.21**) did not detect any direct interactions between NSP2 CTH and RNA. This is somewhat paradoxical given that FCCS and spFRET experiments have established a role of the CTH in both RNA-RNA matchmaking (**Figure 5.2**) and RNA unwinding (**Figure 5.3**).

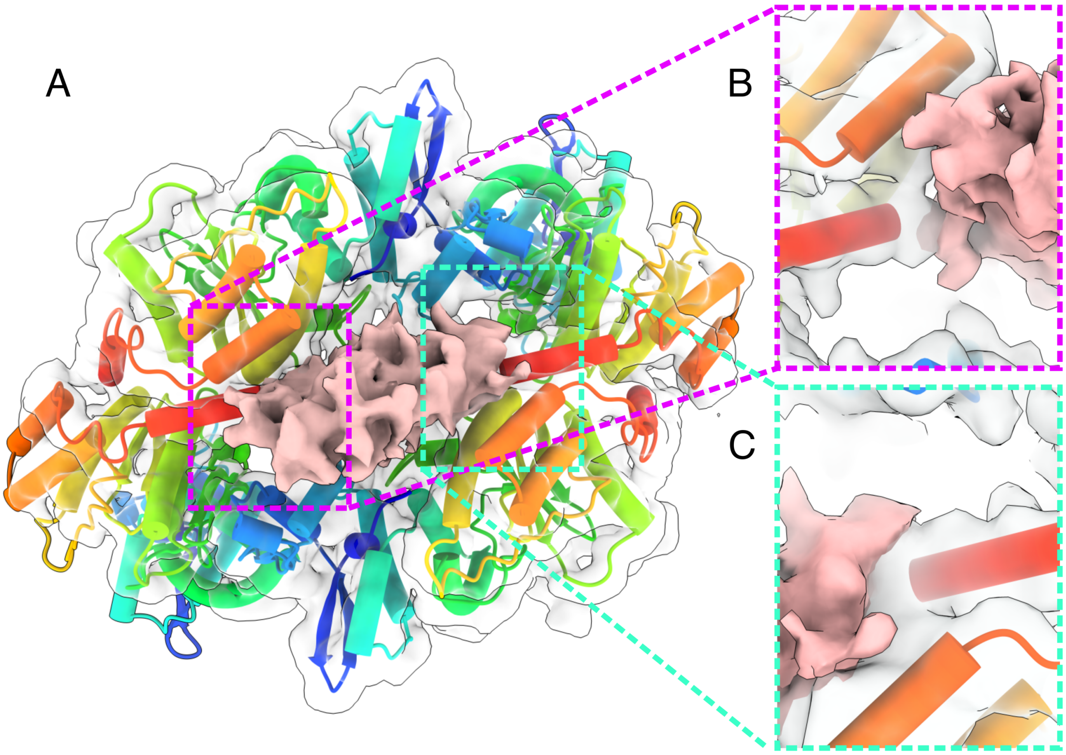


Figure 5.21 Orientation of C-termini within RNP complex. Contacts between NSP2 CTH (red) and RNA density (pink). B & C: CTH-RNA interactions arising from different NSP2 monomers.

When fitted into the cryoEM density map, NSP2 CTH are directly oriented towards the RNA density (**Figure 5.21**). A previous co-crystal structure suggested that binding of RNA may induce a significant conformational change within NSP2 CTH, inducing it to ‘flip out’ (with the flexible linker region (291 – 301) acting as the ‘pivot’) (Hu *et al.*, 2012). However, this conformation is stabilised *in crystallo* through domain-swapping with adjacent NSP2 octamers. This alternative conformation of NSP2 CTH is not supported by the cryoEM density map presented here and is likely a crystallographic artefact.

Putative contacts were observed between the tips of the CTH and the RNA density, although no other CTH-RNA interactions were observed (**Figure 5.20**). In the fitted structure, the CTHs appear ‘poised’ below the RNA density but make few contacts. This suggests that the role of the CTH in the RNA chaperone activity of NSP2 may not be to do with RNA capture (i.e. on-rate) but may instead modulate the affinity of RNA binding via off-rate, as was indicated by the slower dissociation kinetics of NSP2-ΔC as measured by SPR (**Figure 5.4**).

To understand the molecular basis for this, we inspected the electrostatic surface potential of the CTH. Unlike the surrounding, positively-charged RNA-binding interface, the CTH harbours an exposed acidic patch (formed from residues Asp306, Asp310 and Glu311) (**Figure 5.22**). The CTH may promote RNA dissociation from NSP2 through charge repulsion between this acidic patch and the negatively charged RNA phosphodiester backbone.

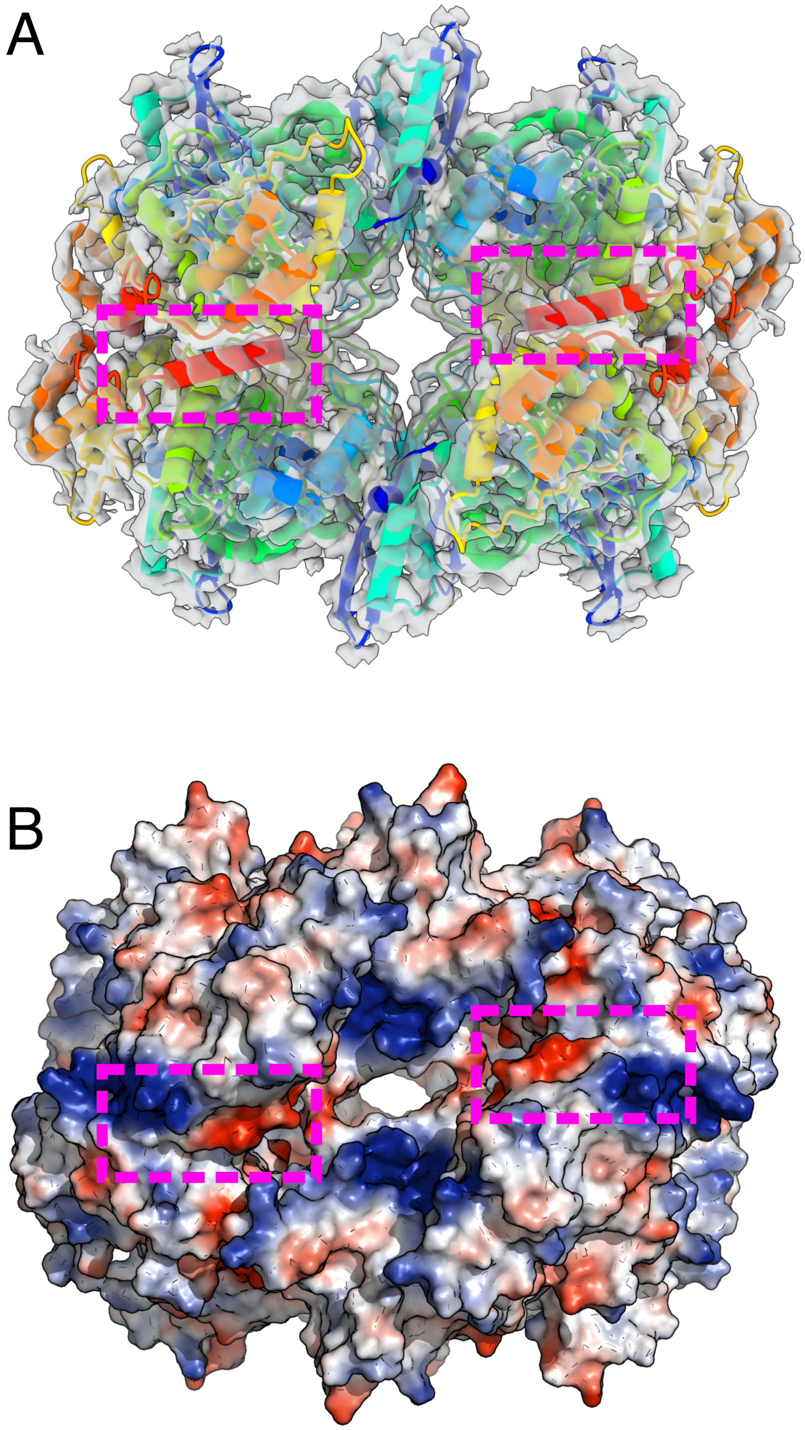


Figure 5.22 Role of C-termini in RNA release. A & B Fitted crystal structure into cryoEM density map (A) and surface electrostatic potential of NSP2 (B) with the positions of the CTH indicated (magenta boxes).

5.3 Discussion

In this chapter, it has been demonstrated that the C-terminus of NSP2 plays an autoregulatory role in NSP2 function by promoting RNA dissociation from the chaperone protein.

This mechanism of RNA dissociation may appear to be remarkably similar to that of the conserved bacterial RNA chaperone protein Hfq. Homohexamers of Hfq bind RNA via three distinct positively-charged sites (termed the proximal, distal and the lateral arginine rim), allowing multivalent RNA binding (Mikulecky *et al.*, 2004; Peng, Yi *et al.*, 2014). Although this core is highly conserved, Hfq possesses an unstructured C-terminal extension whose length varies between species (ranging from fewer than five residues to over 100 residues) (Santiago-Frangos *et al.*, 2016). The tip of the C-terminal domain is enriched with acidic residues, which have been demonstrated to directly interact with the RNA-binding core of Hfq (particularly through interactions with the arginine rim) (Santiago-Frangos *et al.*, 2017, 2019). Through this interaction, the CTD is able to compete for Hfq core binding with nucleic acids, facilitating RNA release from Hfq.

Despite these similarities, there are several important differences in the mechanisms of acidic CTD-mediated chaperone dissociation of Hfq and NSP2. Firstly, NSP2 CTH does not appear to directly compete with RNA for binding to the positively charged core. In fact, the acidic patch within this CTH appears to remain solvent-exposed within the cryoEM structures of both NSP2 apoprotein and NSP2-RNP. Unlike with Hfq, NSP2 CTH does not represent competitive inhibition, but modulates the kinetics of RNP complex formation, presumably via changes to local electrostatic potential. NSP2-ΔC instead remains stably associated with ssRNA, forming a kinetically trapped intermediate (**Figure 5.23**). The high stability and longevity of such an intermediate is not conducive to facilitating inter-segment RNA-RNA interactions (**Figure 5.2**). The work here provides a molecular basis for the lethality of C-terminally truncated NSP2 to rotavirus infection.

Although NSP2 is thought to be incapable of binding double-stranded (ds)RNA, it has been previously demonstrated to be capable of binding structured viral RNAs (Borodavka *et al.*, 2017), both folded and unfolded ssRNAs (Bravo *et al.*, 2018). This raises a question; is the stability of dsRNA duplexes sufficient to promote NSP2 dissociation? If NSP2 is able to bind to RNA structures (albeit less favourably so than to unstructured RNAs (Bravo *et al.*, 2018)), then there may be other factors that promote dissociation of NSP2 from nascent duplexes. This is particularly relevant as it has never been reported that NSP2 is co-packaged with viral RNA segments despite its high affinity for ssRNA. Therefore NSP2 must utilise some kind of molecular mechanism to ensure its eviction from the transient assortment complex during virion assembly and encapsidation.

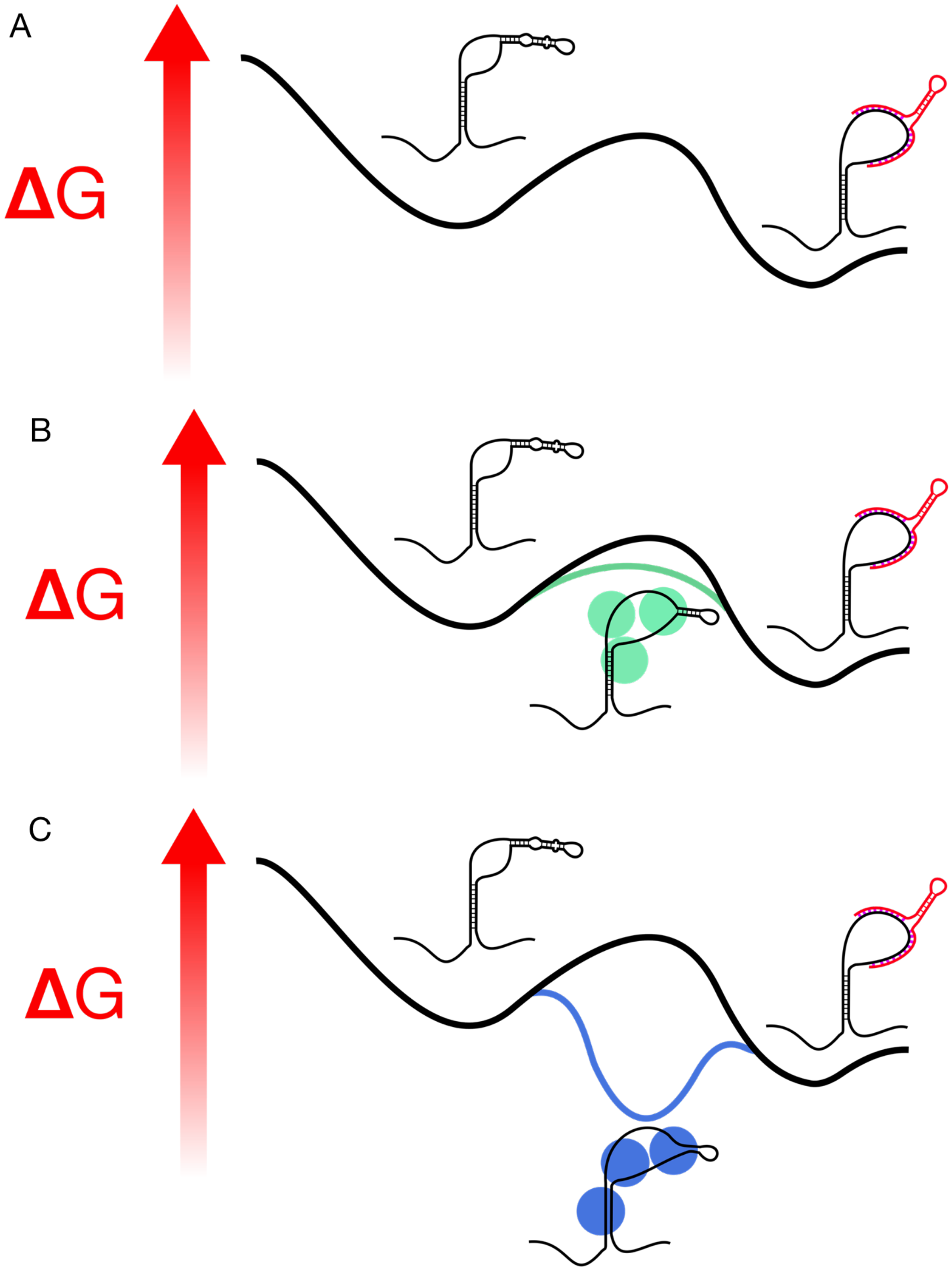


Figure 5.23 Model of NSP2 ΔC RNA chaperone function A – C: Free energy landscape representing two states of RNA: folded ssRNA and dsRNA. A: To venture from ssRNA to dsRNA duplex there is a free energy barrier, representing the free energy required to destabilise ssRNA secondary (and tertiary) structures in order to make sequences involved in intermolecular RNA-RNA interactions exposed. B: The action of an RNA chaperone protein (in this case NSP2 (green)) is able to reduce this energy barrier by destabilising RNA structure, effectively shifting the equilibrium towards RNA duplex formation. C: In the case of NSP2-ΔC, the free energy barrier between ssRNA and dsRNA is further reduced (relative to wild-type NSP2). However, this represents a kinetically-trapped intermediate whereby NSP2-ΔC remains stably associated with unfolded RNA, and thus is not capable of achieving duplex formation.

Chapter 6: Assembly of σNS higher-order RNP complexes

6.1 Introduction

Similarly to rotavirus (RV), avian reovirus (ARV) packages a multi-segmented genome, albeit consisting of 10 (rather than 11) non-identical ssRNAs, and thus likely employs a similar molecular mechanism of RNA selection. As such a mechanism would also require the action of virally-encoded RNA chaperone-like proteins, ARV non-structural protein σNS is thus the best candidate to fulfil this role. Akin to RV NSP2, σNS acts as a matchmaker of sequence-specific RNA-RNA interactions, and is therefore thought to be an essential part of this aspect of virus replication. Indeed, previous studies on the closely-related mammalian reovirus (MRV) have demonstrated that σNS knock-outs result in the production of empty viral particles, highlighting the essential role of σNS in virus assembly (Zamora *et al.*, 2018).

Earlier in this thesis it was demonstrated that σNS undergoes a hexamer-to-octamer transition upon binding of multiple 20-nt RNAs. Although these RNAs correspond to the RNA-binding footprint of σNS, they are far from the physiological substrate of σNS (highly structured viral RNAs, often over a kilobase in length). As such, it is unknown how σNS interacts with longer RNAs.

Unlike NSP2, there is no structural information available regarding σNS. Furthermore, there are no sequence homologues to σNS, preventing the modelling of potential σNS structures based upon related sequences for which structures are already available. Structural characterisation of σNS would provide significant insights into the mechanism used to assemble different oligomers and whether this is coupled to significant conformational change.

Additionally, a hexamer-to-octamer transition would require a source of σNS dimers. While these are not observed in ensemble SAXS measurements (Chapter 4), dissociated dimers are present in the gas phase, as observed in native mass spectrometry analysis (Chapter 4). Whether these smaller oligomers are present in solution remains to be seen.

The work presented in this chapter aims to first establish a structural model for σNS oligomerisation on longer substrates using cryoEM (**Figure 6.9)**. The ability of σNS to assemble larger oligomers upon binding different RNA substrates (including full-length viral RNA) is also investigated. Ultimately, these data indicate that while σNS is in equilibrium between various oligomers in solution, RNA binding is able to stabilise higher-order RNP complexes. This assembly mechanisms may play a role in the RNA chaperone activity of σNS.

6.2 Results

6.2.1 Structural studies of σNS

To understand the structural basis for RNA-driven oligomerization, we sought to determine the crystal structure of σNS. Previous attempts by Dr Alex Borodavka had yielded several conditions suitable for crystal growth, although these crystals did not diffract. Here, we first attempted to perform crystal trials around similar conditions (varying salt concentration, protein : mother liquor ratios). These yielded several crystals with varying morphologies (**Figure 6.1**).

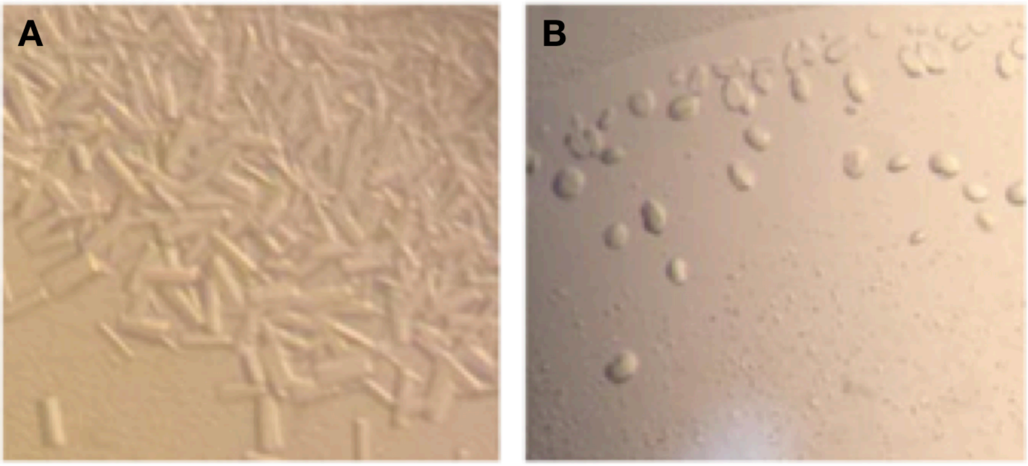


Figure 6.1 Examples of σNS apoprotein crystals. Images of two different crystal morphologies of σNS grown in different buffers (A: 5% PEG 8,000, 0.1 M sodium cacodylate pH 6.8, 0.2 M $MgCl_2$, 18°C, B: 5% PEG 3,000, 0.1 M sodium cacodylate pH 6.4, 0.21 M $MgCl_2$, 12°C). Neither crystals detracted.

Data was collected at Diamond Light Source (DLS) on beamline IO4 at 100K after being cryoprotected in 25% glycerol. However, none of these crystals diffracted. This was repeated on similar crystals at room temperature using the home-source X-ray machine (Rigaku MicroMax-007 HF microfocus), but these crystals did not diffract either. Further optimisation was performed using NT8 crystallisation robot, yielding several similar and new crystal morphologies. An optimal condition (0.1 M Tris pH 7.65, 12% PEG 8,000, 0.2 M $MgCl_2$) was found, and different protein : mother liquor ratios gave different crystal sizes, but a consistent morphology. Data was collected from these crystals *in situ* on Diamond Light Source I03-1 with a 1 second exposure time at 100% transmission. Grid scans were performed, oversampling data (100 µm beam size and 50 µm grid size), ensuring that if there was any diffraction it would be seen. However, these crystals did not diffract at all. Because σNS crystals did not diffract under any of these conditions, it was decided that no further attempts would be made to crystalise the full-length apo-σNS. Crystal trials to attempt to crystalise the σNS in complex with 10mer, 15mer and 20mer RNAs gave several different crystal hits, the majority of which contained 20% PEG 3350 and a magnesium salt. These crystals did not diffract either.

Due to these failures, we attempted to determine a structure of σNS using cryoEM. Preliminary cryoEM studies of sNS had been performed in 2014 by Dr Rebecca Thompson and Dr Alex Borodavka. Grids prepared with 1 mg/ml sNS (~25 µM) appeared highly crowded and were thus deemed not ideal. To overcome this, extensive screening of different σNS concentrations was undertaken. At lower concentrations (<1 µM), particles were present in ice, but were not overly crowded (**Figure 6.2**). A dataset of 1625 micrographs was collected on a Titan Krios equipped with a Falcon III detector operating in integrating mode at a nominal magnification of 75,000x. All data processing was performed using Relion 2.1. Micrographs were motion- and CTF-corrected using Relion’s own motion correction implementation and GCTF, respectively (**Chapter 3 Materials and Methods**). 282,035 particles were autopicked using 2D class averages generated from 3439 manually picked particles as references. Autopicked particles were extracted and subjected to reference-free 2D classification (**Figure 6.2**).

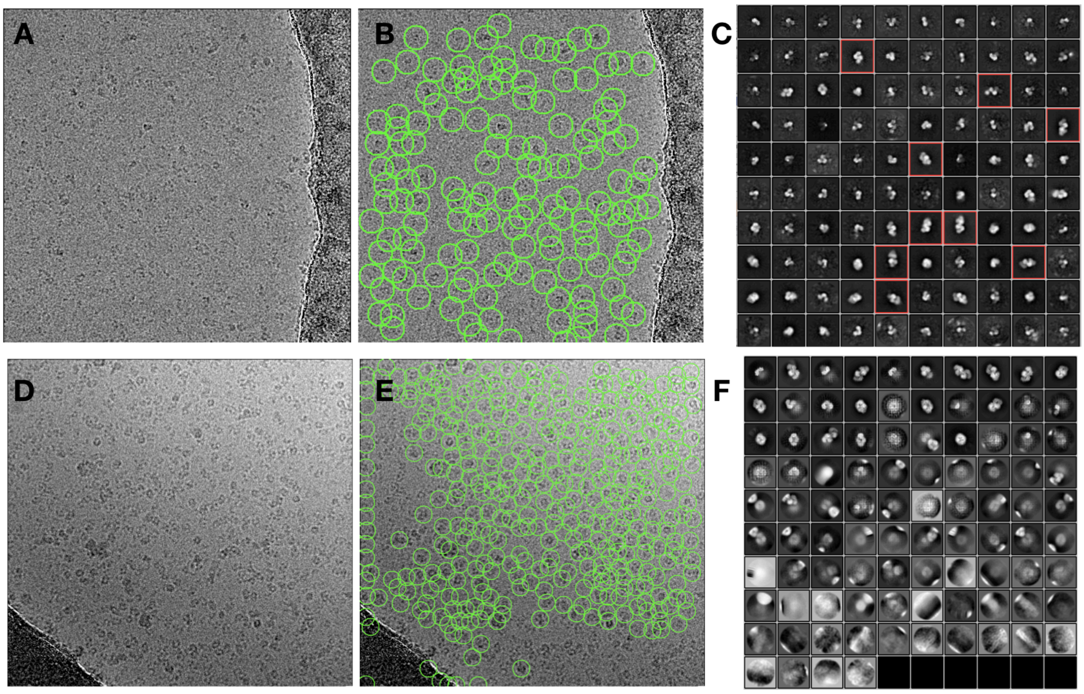


Figure 6.2 CryoEM of σNS apoprotein A & B: Representative cryoEM micrograph of grid prepared with 250 nM σNS, and corresponding autopicked particles. C: Typical 2D classes of σNS from arising from the same dataset as micrograph A. High abundance of smaller (likely dimeric) σNS oligomers are present. Larger, low abundance hexameric and octameric oligomers are indicated by red boxes. D & E: Representative cryoEM micrograph of grid prepared with 25 *μ*M σNS, and corresponding autopicked particles (E). Particles were picked using LoG methods (see **Chapter 3 Materials and Methods**), and some false positives were picked (i.e. at the ice-carbon edge, and along the left edge of the micrograph. F: 2D classes (with fast subsets) of σNS particles picked. Although dimeric σNS is still the most abundant class, a range of different σNS oligomers are more highly represented than in (C).

2D classes indicated that σNS was disassembling into dimers upon vitrification. The most abundant classes likely corresponded to dimeric σNS. Although σNS dimers have been observed in native mass spectrometry and sedimentation velocity experiments, they are likely to be a low abundance population in solution (**Chapter 4**, (Borodavka *et al.*, 2015; Bravo *et al.*, 2018)). Although there were 2D classes corresponding to larger σNS oligomers (**Figure 6.2C, red**), their low abundance relative to dimeric σNS meant that this dataset was not suitable for further analysis.

Serendipitously, additional grids made 1 mg/ml σNS apoprotein in 2014 by Drs Thompson and Borodavka were discovered and screened. Of the eight grids screened, only one had multiple grid squares with unbroken ice. Although the particles were more abundant within the ice, they did not appear to be crowded to such a degree that may be detrimental to data processing. Accordingly, data was collected from this grid using a Titan Krios equipped with a K2 detector operating in counting mode. 3228 micrographs were collected, and from these 1,011,851 particles were autopicked using reference-free autopicking (using Laplacian-of-Gaussian filtering). Extracted particles were binned 4x (particle box size rescaled from 256 pixels (pix) to 64 pixels, with new effective pix size of 4.28 Å/pix) and subjected to reference-free 2D classification (using 'Fast Subsets").

Although Fast Subsets is useful for sorting between bona fide particles and false positives, the most abundant classes do not necessarily reflect the true variation within the particle distributions present (as particles that are broadly similar tend to be classified together). This meant that the most abundant class was, once again, dimeric σNS (27% of input particles, including false positives). However, larger σNS oligomers were far more represented than in the previous dataset. All σNS particles were used for further processing.

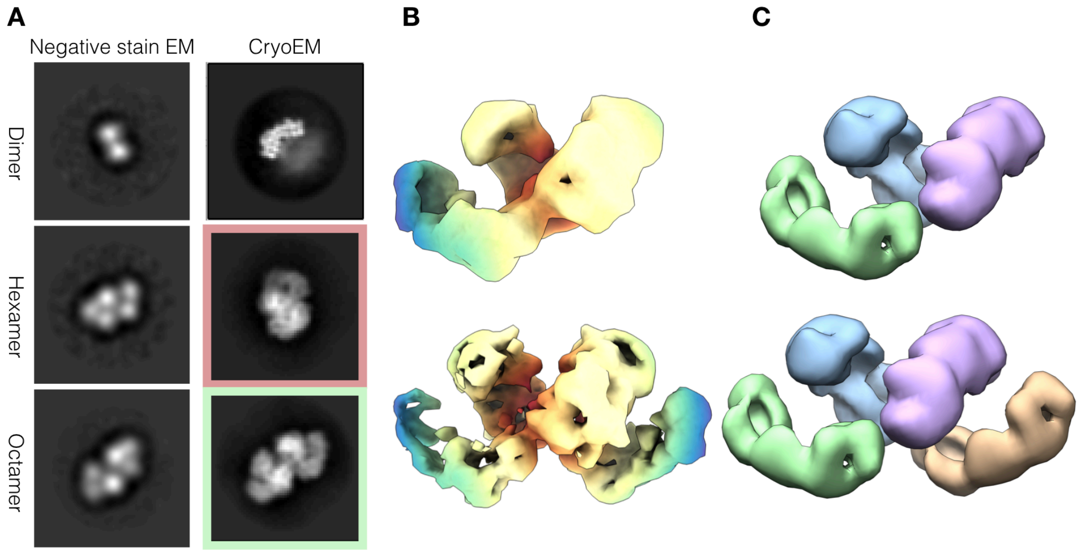


Figure 6.3 Observation of different oligomeric states of σNS A: 2D class averages from negative stain EM (left) and cryoEM (right) of dimeric, hexameric and octameric σNS. The occurrence of these oligomeric states in both negative stain EM indicates that the presence of dimeric σNS in cryoEM class averages is not simply an artefact of larger oligomeric states disassembling in ice. B: 3D classes corresponding to hexameric (A, pink box) and octameric (A, green box) σNS oligomers. C: 3D initial models of σNS dimer fitted into 3D classes of hexameric and octameric σNS to reveal quaternary organisation.

Further 2D classification indicated that there were three major σNS species present: dimers, hexamers and octamers. As it is not uncommon for protein complexes to dissociate upon vitrification (Drulyte *et al.*, 2018), we wished to validate their presence using an orthogonal method. Accordingly, σNS was imaged under negative stain conditions (2% uranyl acetate, see Materials and Methods). ~30,000 autopicked particles were subjected to 2D classification, revealing dimer, hexamer and octamer classes (**Figure 6.3A**). This confirmed that dimeric σNS was not simply an artefact of oligomer dissociation.

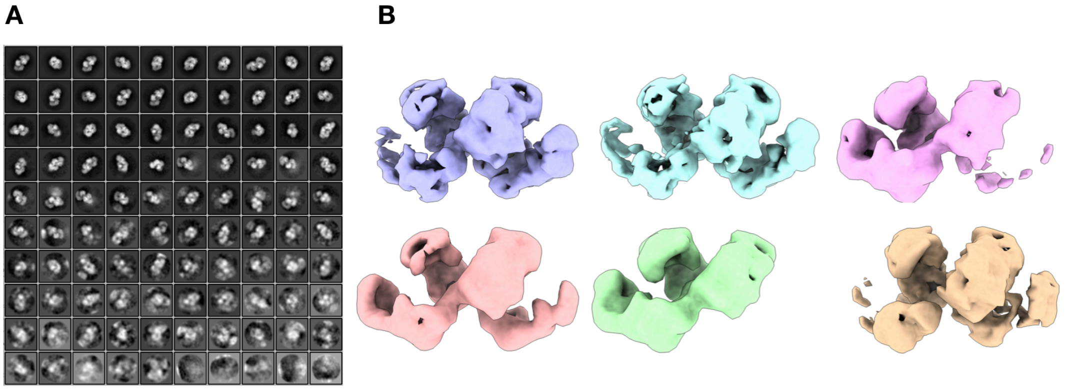


Figure 6.4 Further 2D and 3D classification of σNS A: 2D class averages from selected hexamer and octamer particles. The top 30 classes (i.e. top 3 rows) contain 95% of the total subset of particles. B: 3D classification of these selected particles into 6 classes further establishes a mixture of σNS hexamers and octamers. While classes that appear to be distinct oligomers (i.e. top left, top middle, bottom left, bottom middle), particles with weaker density for some σNS oligomers are also present (top right, bottom right).

2D classes corresponding to hexamer and octamer were subjected to 3D classification, revealing similar global architectures, albeit with additional density for octameric σNS. Particles classified as dimeric σNS were used to generate an initial 3D reconstruction. This density was then fitted into the hexamer and octamer 3D classes using the Fit-to-Map function in Chimera (**Figure 6.3**). Again, the organisation of dimers within hexameric and octameric σNS is remarkably similar, suggesting that σNS is not required to undergo a major conformational change to transition between oligomeric states. This supports the notion of an RNA-driven oligomerisation event, whereby RNA binding shifts the equilibrium from hexamer to octamer.

σNS hexamers and octamers were subjected to further 2D classification (**Figure 6.4A**). From the 100 2D classes used, the most highly populated 30 classes contained 91% of the input particles. These 'clean' particles (i.e classes without obvious false-positive particles, or other "junk") were subjected to further 3D classification (**Figure 6.4B**). These extra classification steps were necessary due to the high amount of heterogeneity within the dataset. These 3D classes again comprised a mixture of hexamers and octamers. Although the central dimers are generally well-resolved, the distal dimers (i.e. those at the periphery of σNS) typically had weaker density. This may be due to flexibility, or a mixture of different dimer occupancies. This supports the notion that these two oligomeric states are present in equilibrium in solution.

From these, a 3D reconstruction of octameric σNS was determined using C2 symmetry, to a global resolution of 6.68 Å (**Figure 6.5**). Local resolution varied from 5.9 Å at the core to 7.9 Å at the peripheries. The angular distribution plot (**Figure 6.6C**) indicates that σNS has preferred orientation. Although this appears to manifest as stretched density in the plane of the preferred orientation (**Figure 6.6**), it is unclear what the overall scale of distortion this is having on the final reconstruction of σNS as there are no other structural models of σNS (or any structurally homologous proteins) available.

Although this preferred orientation may prevent the determination of a high-resolution structure of σNS, this modest resolution density map is still informative. Octameric σNS exists as two interlocked dimers-of-dimers (i.e. tetramers) (**Figure 6.7**). These two tetrameric ‘arms’ brace one another, forming contacts between the core dimers.

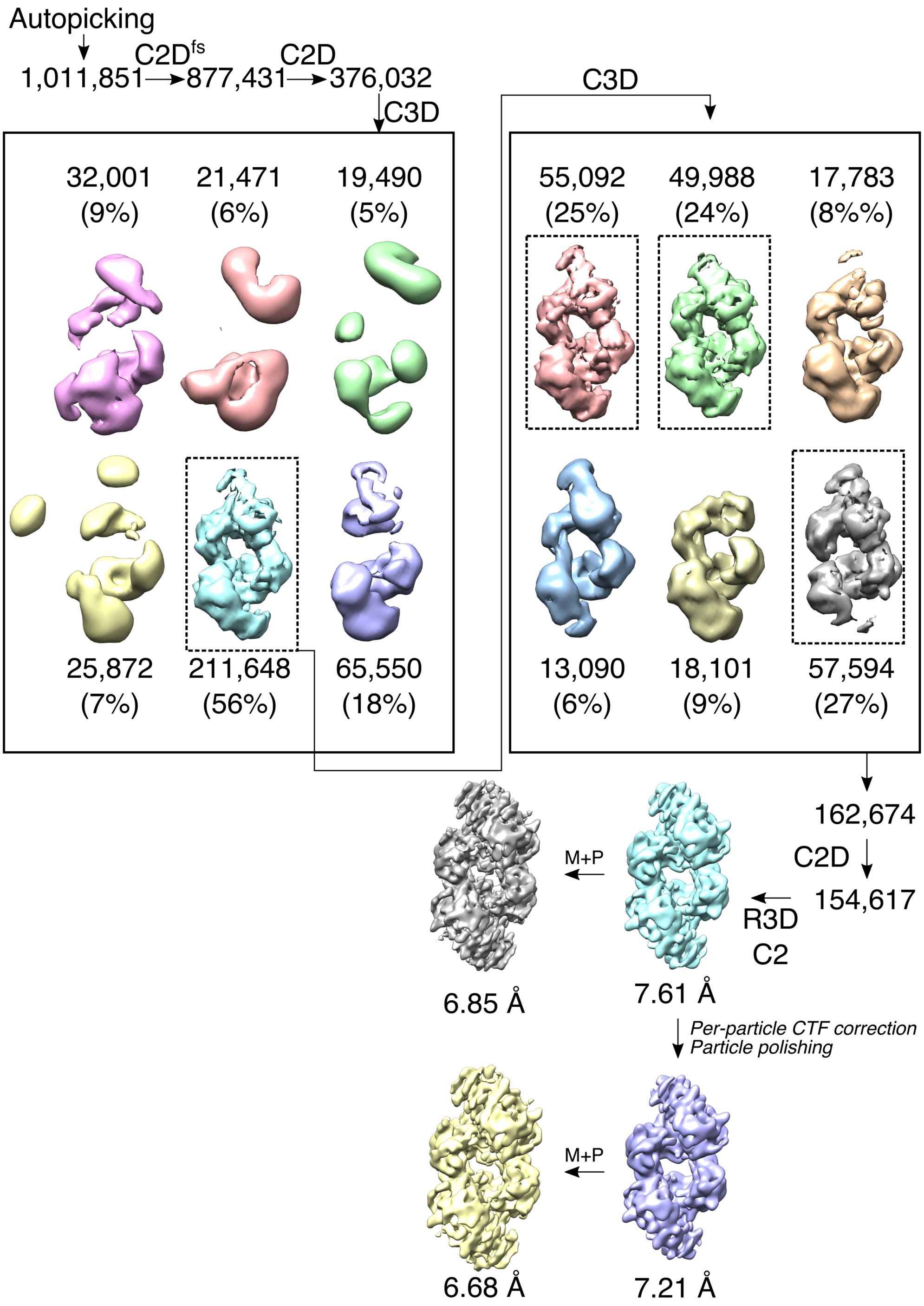


Figure 6.5 σNS cryoEM data processing workflow Image processing workflow used to determine 3D reconstruction of σNS apoprotein octamer. Initial round of 2D classification was performed with fast subsets (C2D^{fs}). A second round of C2D was performed without fast subsets. All rounds of 3D classification (C3D) were performed without fast subsets. 3D reconstructions (R3D) was performed with C2 symmetry. All reconstructions were subjected to masking and post-processing (M + P) to improve map resolution.

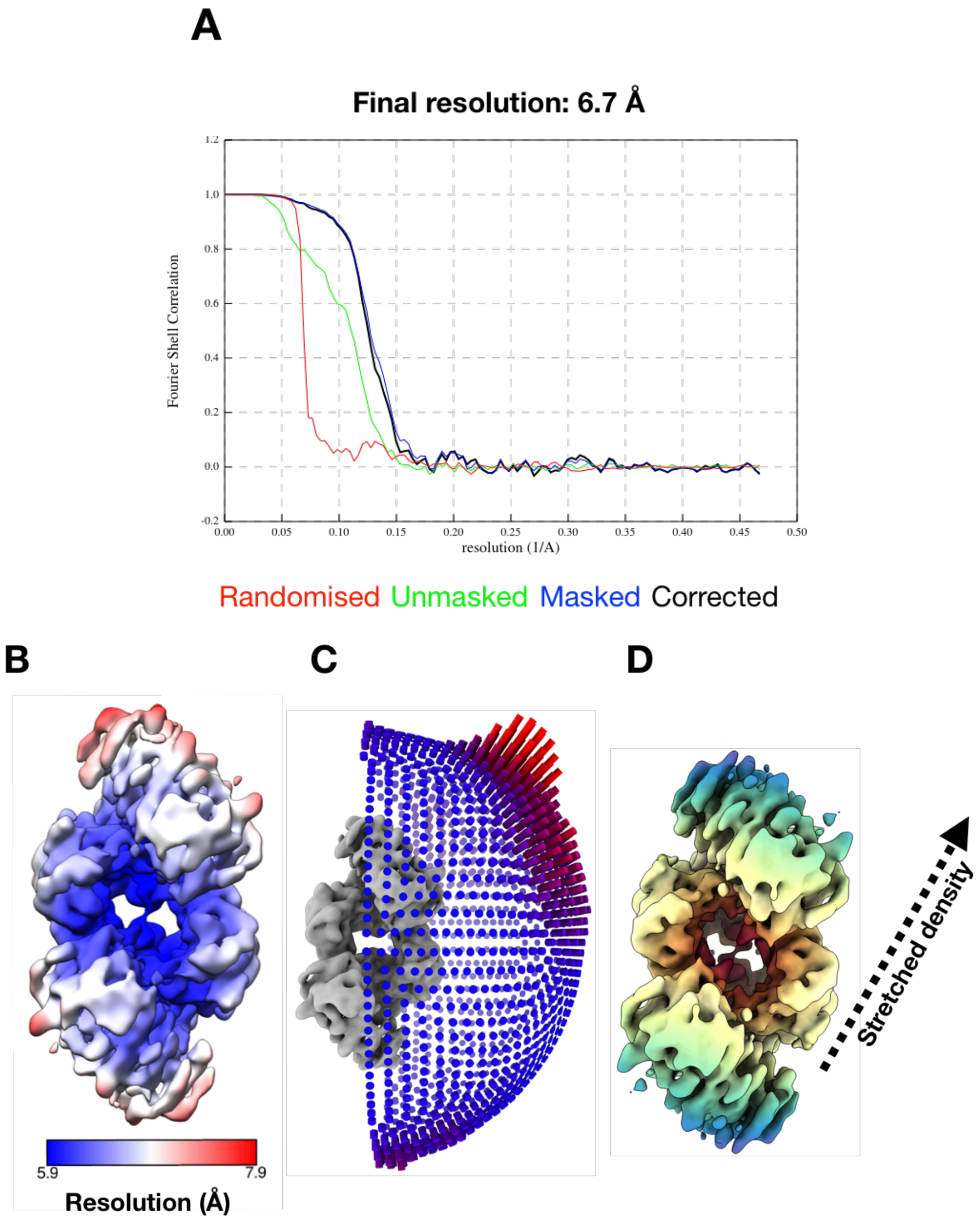


Figure 6.6 Observation of different oligomeric states of σNS. A: Fourier shell correlation (FSC) plot of σNS octamer cryoEM reconstruction. Global resolution was determined to be 6.7 Å according to the "gold standard" FSC of 0.143. B: Local resolution of σNS octamer. Resolution ranges from 5.9 Å in the centre of the particle to 7.9Å at the periphery. C & D: Angular distribution plot of σNS octamer reconstruction (C2 symmetry). Strong preferred orientation is present, resulting in stretched density in the final reconstruction (D) in the plane of the preferred orientation.

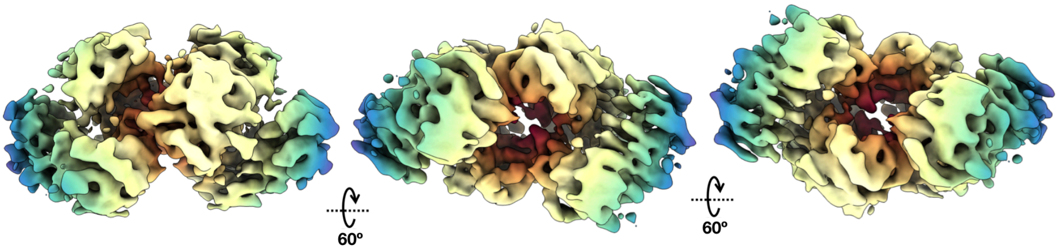


Figure 6.7 CryoEM reconstruction of octameric σNS Side view of σNS octamer, rotates by 60° twice to demonstrate the interlocked nature of the two "dimer-of-dimer" arms.

This density map was compared to a 6.7 Å resolution simulated density map of NSP2 (**Figure 6.8**). Octameric σNS and NSP2 have strikingly different quaternary structures. The interlocked arms of σNS shares little resemblance with the characteristic toroid architecture of NSP2, reflecting the differences in RNA chaperone mechanisms of these two proteins. Despite this, the 'underside' of σNS contains a deep groove across a 2-fold symmetry axis, similar to the grooves on NSP2. As these are the RNA binding sites of NSP2, it is tempting to speculate that σNS may also utilise a similar structural feature to bind RNA.

Attempts to determine a cryoEM structure of a σNS RNP complex were consistently unsuccessful. Although a homogenous σNS RNP complex was observed by negative stain EM (**Figure 6.9A**), with 2D class averages resembling σNS octamers (**Figure 6.9B**), this complex dissociated upon cryoEM preparation (**Figure 6.9B & C**). Although there are similar cases to this in the literature, whereby complex dissociation was ameliorated by addition of beta-octyl glucoside (Larsen *et al.*, 2018), this avenue was not explored due to the relative lack of success in determining a high resolution structure of σNS apoprotein.

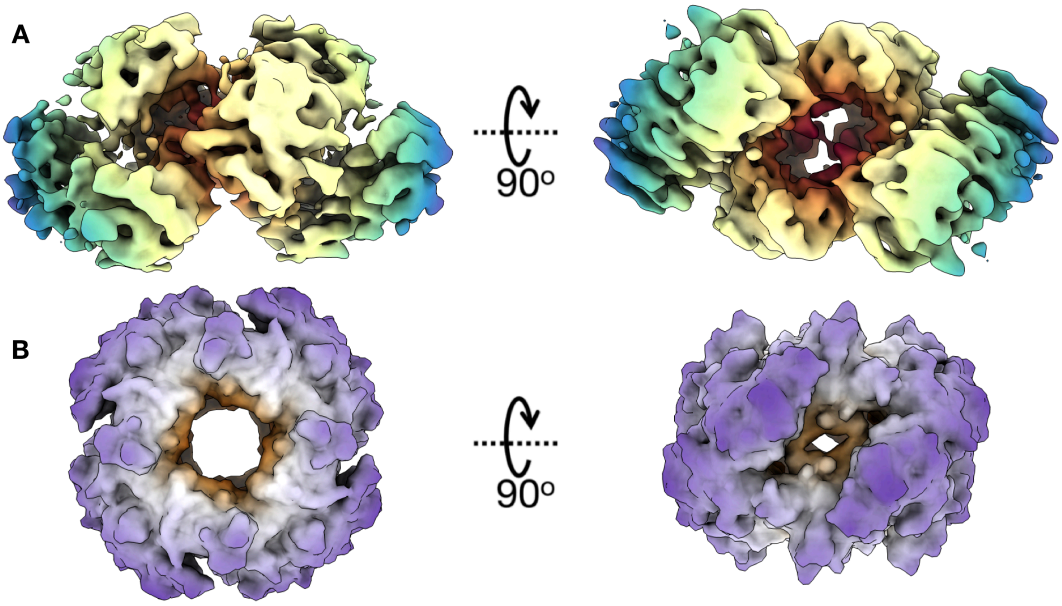


Figure 6.8 Comparison of σNS and NSP2 octamer structures A: 6.7 Å reconstruction of σNS with side and underside views. B: 6.7 Å simulated surface of NSP2 generated in Chimera using PDG 1L9V, viewed in town-down and side orientations.

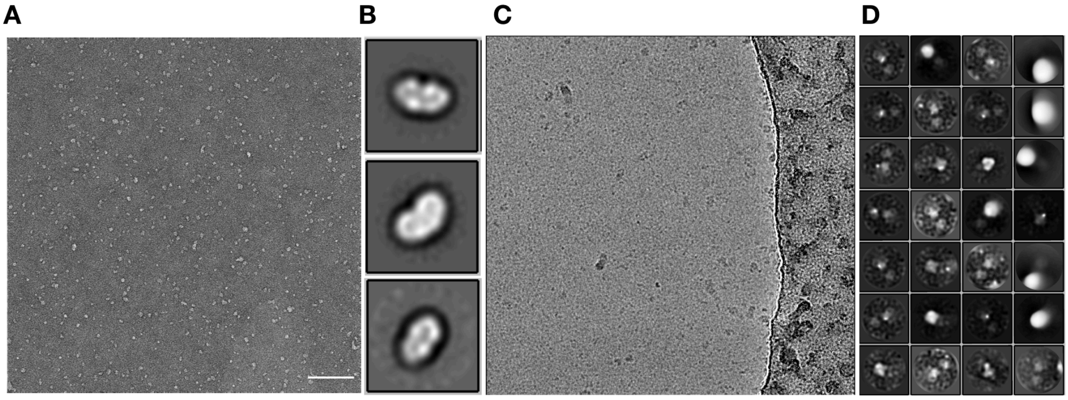


Figure 6.9 Negative stain EM and cryoEM of σNS 20mer-RNP complex. A & B: Negative stain EM micrograph of σNS 20mer-RNP complex (A), and corresponding 2D class averages (B). Particles are monodisperse, and 2D class averages likely corresponding to σNS octamers are present. C & D: CryoEM micrograph of σNS 20mer-RNP (C) and corresponding 2D class averages, indicating that the RNP complex is not intact in ice.

6.2.3 Assembly of higher-order σNS oligomers

The assembly of σNS RNP complexes with short RNA substrates corresponding to the its RNA-binding footprint (~20-nt) has been established (**Chapter 4**) (Bravo *et al.*, 2018). Indeed, the cryoEM density maps of σNS apoprotein provide a structural basis for this oligomeric promiscuity. It is unknown, however, whether σNS assembles into larger oligomers upon binding longer RNA substrates. To understand this, different σNS RNP complexes were assembled and imaged using negative stain EM (see Materials and Methods). σNS apoprotein and σNS-20mer RNP complex exhibit monodisperse populations. (**Figure 6.10**) Surprisingly, σNS-40mer RNP complex assembles into large, filamentous RNP complexes. These filaments are highly flexible and polydisperse, with a range of different lengths of species present (**Figure 6.10**). Interestingly, NSP2 does not assemble into such filaments upon binding 40mer RNA, but rather forms higher-order species that resemble beads-on-a-string (**Figure 6.10D**). This may represent the concurrent binding of multiple discrete NSP2 oligomers to the same RNA molecule(s) as opposed to bona-fide oligomerisation.

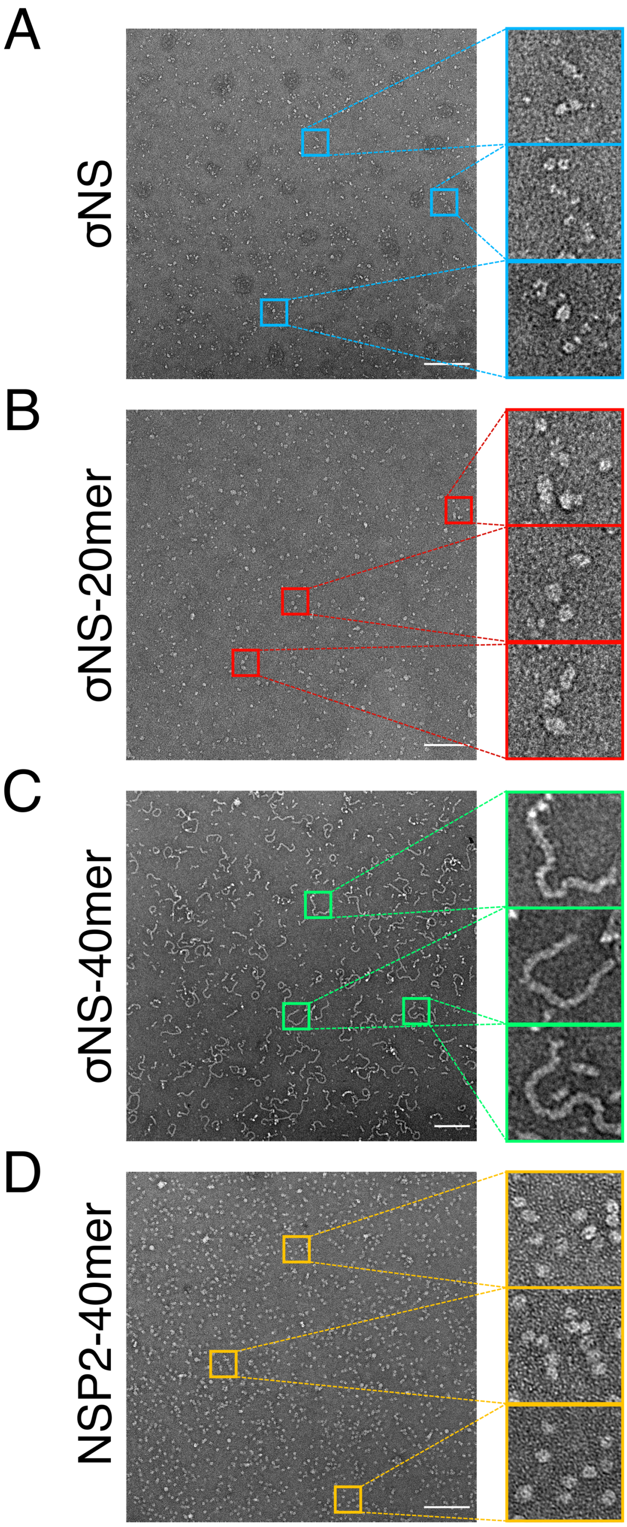


Figure 6.10 Assembly of higher-order σNS RNP complexes. A: Negative stain EM micrographs of A: σNS apoprotein, B: 20mer-RNP and C: 40mer-RNP. As previously described, σNS apoprotein and 20mer-RNP are monodisperse and globular (Chapter 4). σNS-40mer RNP complex, however, assembles into polydisperse filaments. D: NSP2 does not assemble such filaments, but "beads-on-a-string" corresponding to concurrent binding of multiple NSP2 octamers occurs (box).

The start and end coordinates (although there is no apparent polarity) of σNS filaments imaged by negative stain EM were manually picked, extracted and subjected to reference-free 2D classification in Relion 3 (**Figure 6.11**). This strategy ensured that only the straightest segments of these otherwise highly flexible filaments were selected. In spite of this, the resulting 2D classes still exhibit flexibility (i.e. are wiggly and bendy). These filaments exhibit periodic density features, supporting the notion of a helical assembly (**Figure 6.11**). However, due to the inherent heterogeneity of these filaments it was not possible to estimate their helical parameters (pitch and rise).

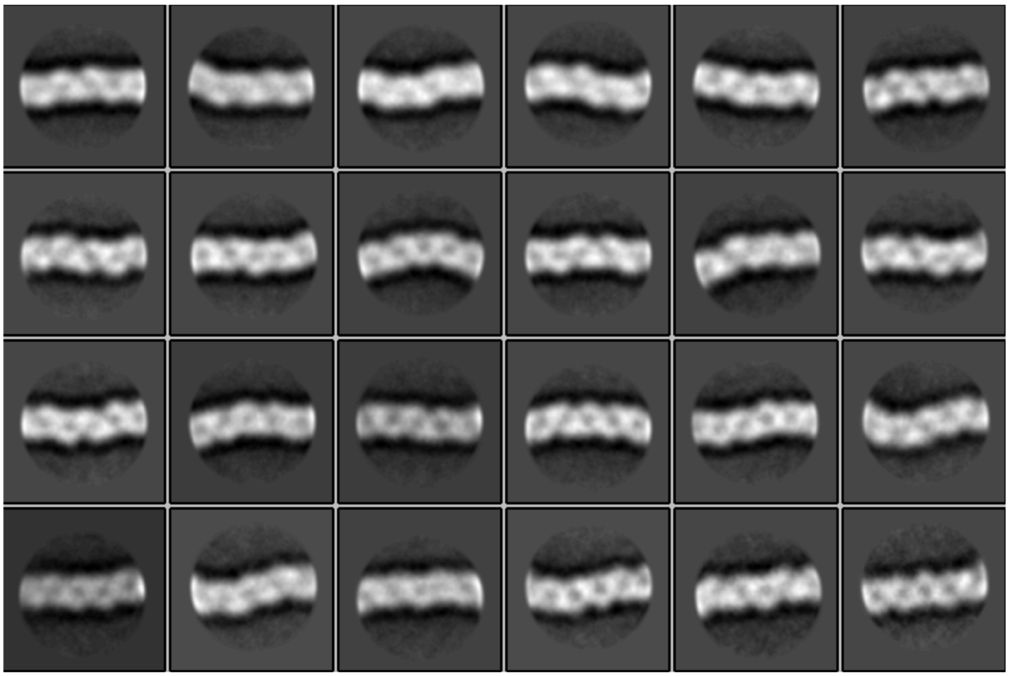


Figure 6.11 Negative stain EM 2D class averages of σNS-40mer filaments. 2D class averages show periodic features, but are still polydisperse and somewhat flexible.

Particles corresponding to well-resolved helices were subjected to 3D classification (**Figure 6.12**). Again, these 3D classes reflect the heterogeneity of the filaments, with all five classes generated exhibiting signs of flexibility (i.e. they are not straight). Particles contributing to the most populated class were selected and used for 3D refinement, resulting in a 17.7 Å reconstruction (**Figure 6.12, box**).

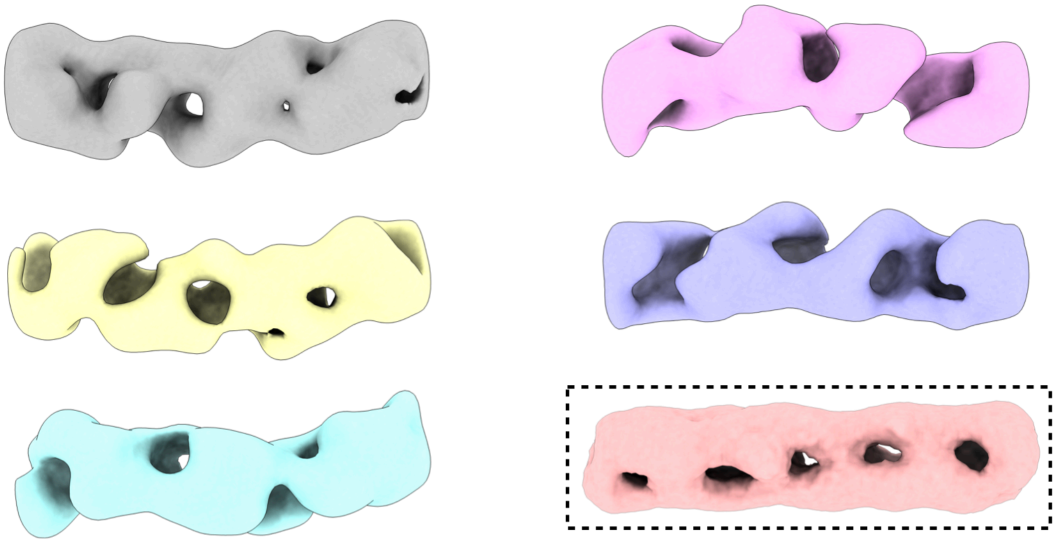


Figure 6.12 Negative stain EM 3D class averages of σNS-40mer RNP filaments. Particles from 2D classification were subjected to 3D classification (25 iterations, 5 classes). The most abundant class was then used for 3D refinement. Refined model (peach) is denoted by dashed box.

The cryoEM density map of σNS octamer was fitted into the refined σNS-RNP filament (**Figure 6.13**). Although negative stain EM reconstructions may contain inherent artefacts as the heavy metal stain may distort the outline of the particle, the octamer density corresponded well with the filament density. In addition to a consistent particle/filament width, the repetitive grooves on the surface of the filament density are a similar size to the groove present on the 'underside' of σNS. This may provide a mechanism for σNS to assemble from discrete, smaller oligomers (dimer, hexamer, octamer) into large, non-discrete RNP complexes.

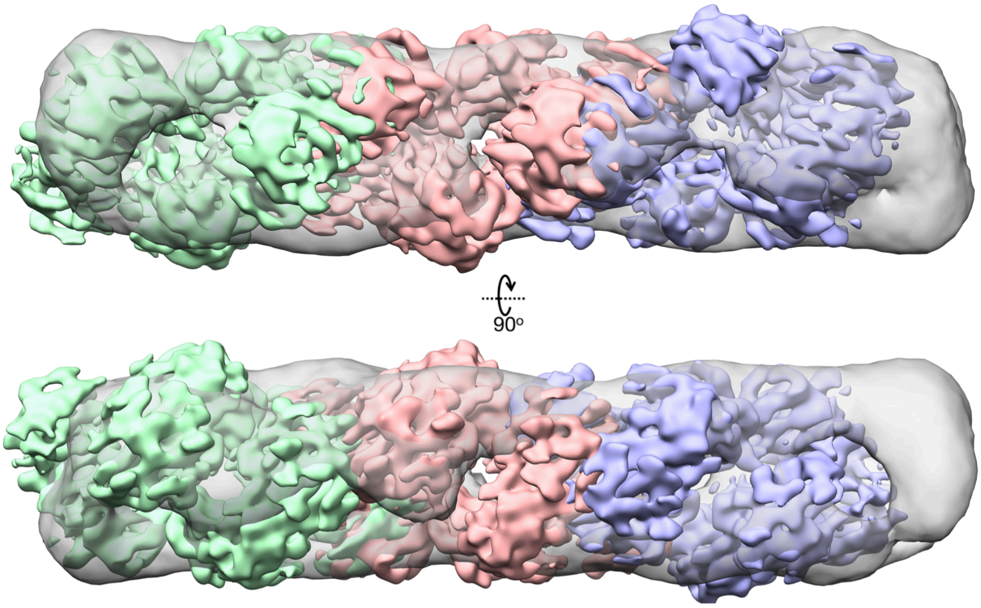


Figure 6.13 Model of σNS assembly into RNP filaments. CryoEM densities of σNS octamer reconstruction were fitted into refined negative stain EM reconstruction using UCSF Chimera (Pettersen *et al.*, 2004), with correlations ranging from 0.967 – 0.972. The underside groove of σNS fits within the periodic grooves present in the refined filament model, and the width of σNS also corresponds to the filament width. Peripheral dimers of σNS overlap between neighbouring octamers, indicating how σNS may assemble a continuous, nondiscrete oligomer upon RNP complex formation.

Reference-free 2D class averages of the ends of the filaments were also determined (**Figure 6.14A**). Examination of these classes did not reveal filament polarity. Several 2D classes showed “fuzzy” density at their tips (**Figure 6.14A, red asterisk**), which may correspond to naked RNA. If this is the case, then this exposed RNA would likely be <20-nt, and therefore not be able to fully accommodate binding of additional σNS.

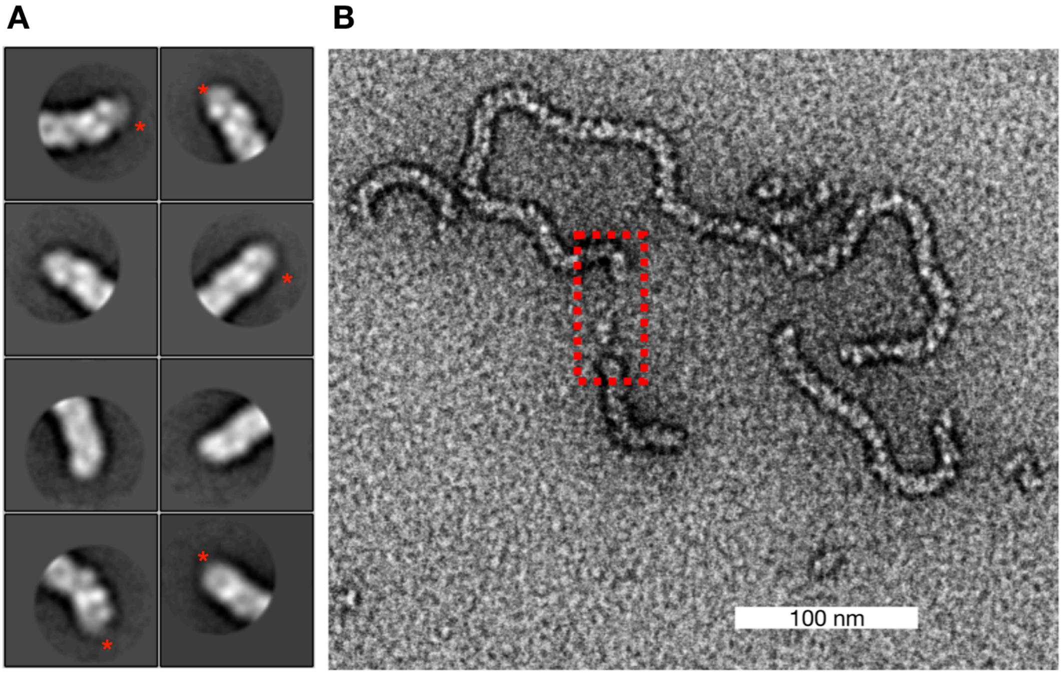


Figure 6.14 Negative stain EM analysis of filament (σNS-40mer) assembly. A: 2D class averages of σNS filament ends. Some classes contain weaker density (red asterisk), which may correspond to uncoated RNA. B: Filament assembly intermediate containing naked RNA (red box) flanked by continuous σNS filament

Within the negative stain EM micrographs collected, a potential filament assembly intermediate was present (**Figure 6.14B**). This consisted of adjacent filaments connected by a regions of narrower density with weaker contrast (not corresponding to protein). This indicates an assembly mechanism where protein-protein interactions are stabilised by RNA. If this was not the case, binding of σNS to RNA may be stochastic, giving the appearance of beads-on-a-string rather than continuous helical assemblies.

The presence of RNA within these RNA assemblies was further validated using SEC. σNS apoprotein, RNP-20 and RNP-40 were assembled and their A260/A280 ratios were analysed. While σNS apoprotein has A260/A280 = 0.57 (indicating that no RNA is present in the sample), the elution peaks of both the 20mer-RNP and 40-RNP complexes have A260/A280 of 1.2 and 1.5, respectively. This indicated that RNA is present in σNS filaments, and is not simply required to nucleate their assembly.

6.2.3 Assembly of higher-order σNS RNP complexes

To unravel the mechanism underlying filament assembly, a range of different RNA substrates with different lengths and structural stabilities were co-incubated with σNS, and the resulting filaments were imaged using negative stain EM. Lengths of filaments were measured using semi-automated procedures implemented in FiberApp (Usov and Mezzenga, 2015). Abundance of filaments of different lengths were normalised to their total measured populations (all >500) and plotted as a histogram of frequency as a function of measured filament length (**Figure 6.15A**).

The dependence of filament assembly on RNA length was first investigated. Filaments were assembled using either unstructured 40-nt or 60-nt RNA, and populations of filaments were quantified (**Figure 6.15A**). Interestingly, populations of filaments assembled using either RNA were highly similar, indicating that RNA length does have an impact on filament assembly (once the RNA length is sufficiently long i.e. somewhere between 20 and 40-nt).

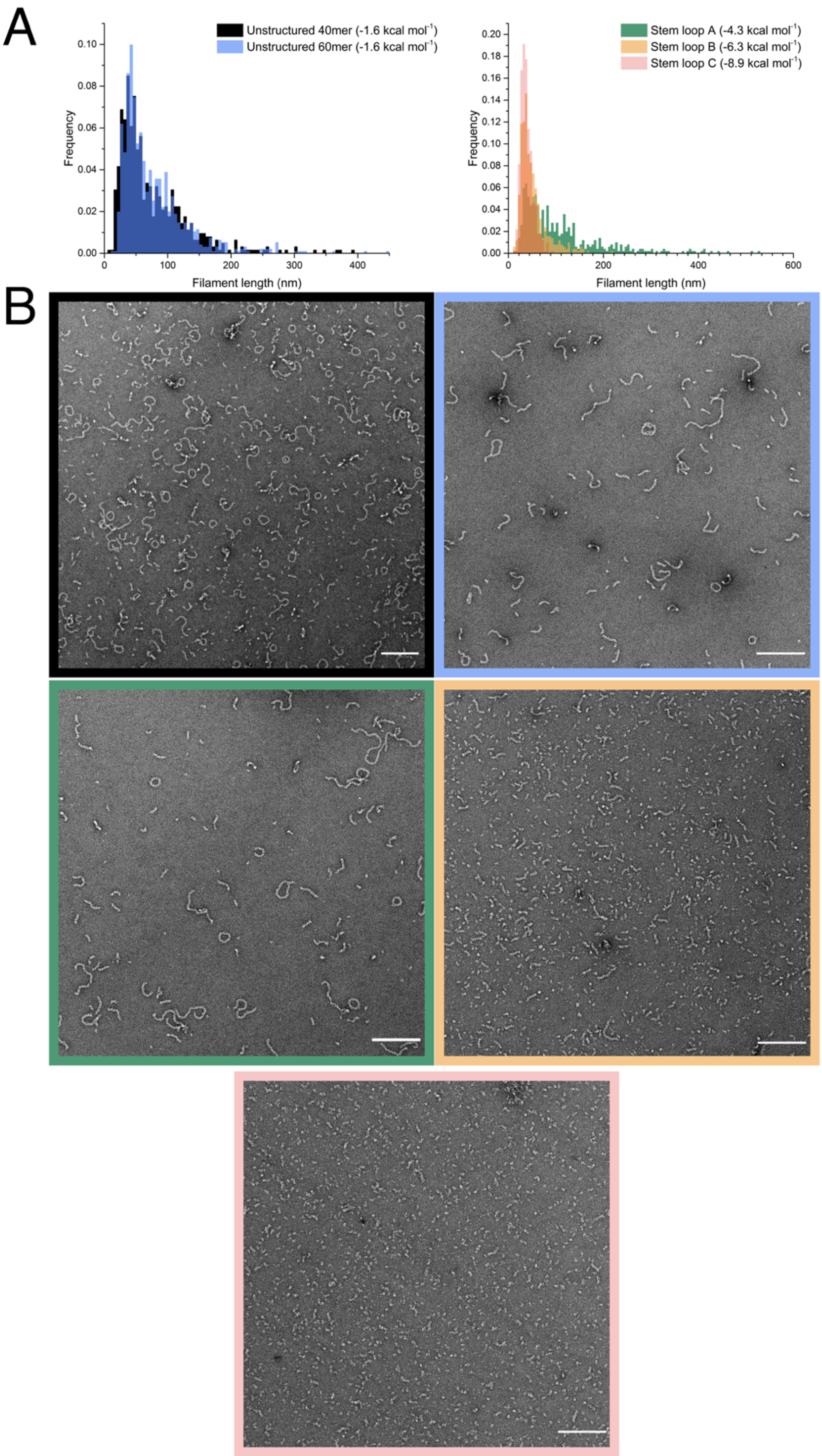
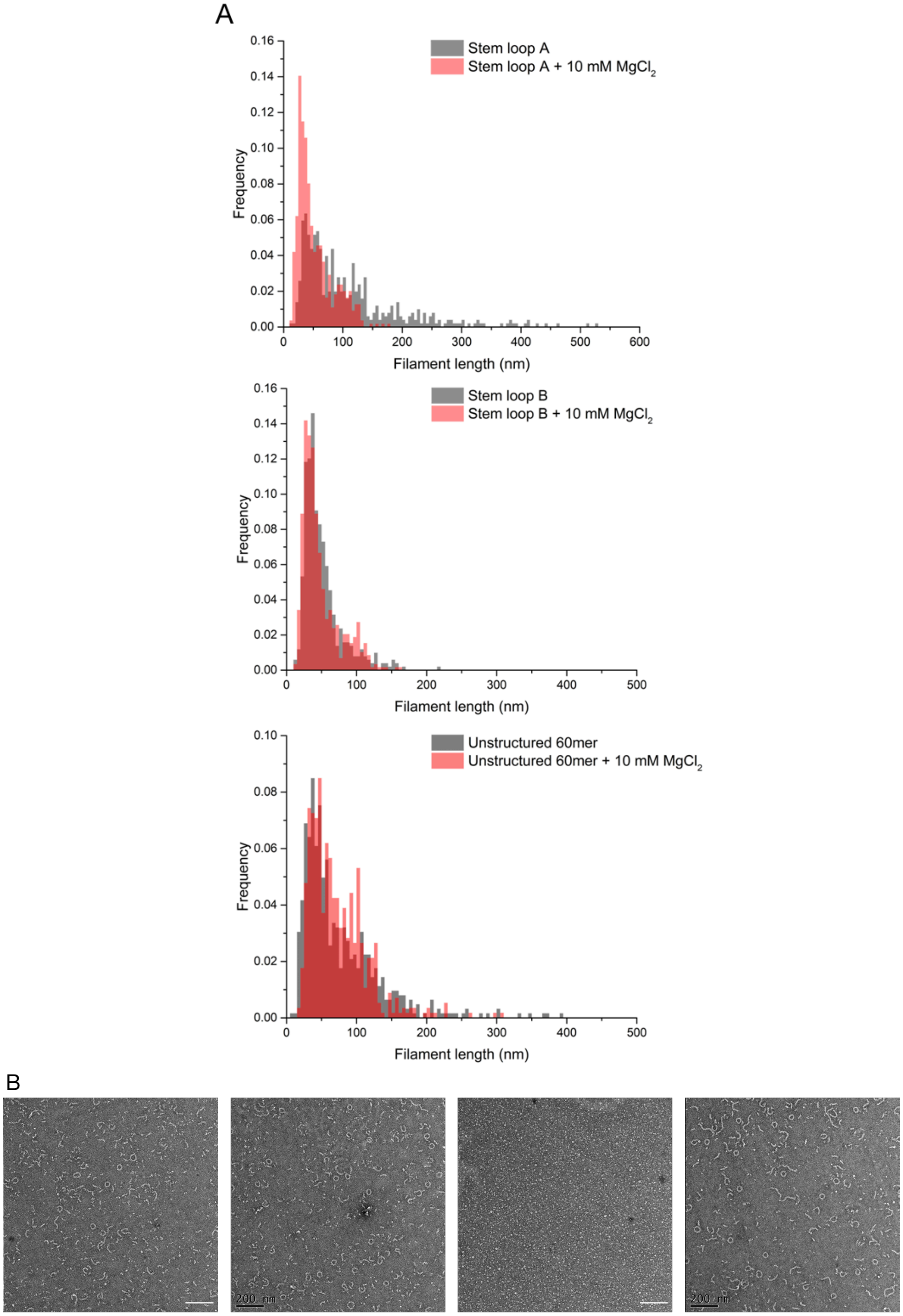


Figure 6.15 Relationship between filament assembly and RNA structural stability. A & B: Histograms of filament lengths assembled using different RNA substrates. Filament lengths were quantified using FiberApp. C: corresponding typical negative stain micrographs of different RNP assembles as in A & B.

As σNS has been demonstrated to assemble larger oligomers in order to increase its stem loop unwinding efficiency, it is plausible that the assembly of these higher-order filamentous RNPs may also be involved in the RNA chaperone activity of σNS. Accordingly, the dependence of filament assembly on RNA structural stability was investigated. Three RNA stem loops of similar length with different secondary structure stabilities (as predicted by the RNA fold webserver) were designed and incubated with σNS – stem loop A (*SLA*, -4.5 kcal mol^{-1}), stem loop B (*SLB,* -6.3 kcal mol^{-1}), and stem loop C (*SLC,* -8.9 kcal mol^{-1}). Filament length populations were subsequently analysed as described above.

Although filaments assembled with all three stem loops had similar modal lengths (32.5 – 37.5, **Table 6.1**), the length histograms are strikingly different. An overall shift towards shorter filaments was observed for *SLB* filament population compared with *SLA* filaments. However, for *SLC* filaments, a further decrease in population length was observed. This data indicated that there is a relationship between RNA structural stability and filament assembly efficiency.



σNS:SLA + 10 mM MgCl$_2$ σNS:SLB + 10 mM MgCl$_2$ σNS:SLC + 10 mM MgCl$_2$ σNS:60mer + 10 mM MgCl$_2$

Figure 6.16 Effects of Mg^{2+} on RNP filament assembly. A: Histograms of filament lengths assembled using different RNA substrates in the absence (black) and presence (red) of 10 mM MgCl$_2$. B: Negative stain EM micrographs of filaments quantified in (A). Although σNS-*SLC* + 10 mM MgCl$_2$ was imaged, no filaments were assembled and thus could not be quantified.

To further test this, RNA stem loops were further stabilised through incubation with 10 mM $MgCl_2$ prior to filament assembly (i.e. incubation with σNS) (**Figure 6.16**). While σNS was able to assemble filaments with Mg^{2+}-stabilised *SLA* and *SLB*, no filaments were observed upon incubation of σNS with Mg^{2+}-stabilised *SLC*. As a control, unstructured 60mer RNA was incubated with $MgCl_2$ prior to filament assembly. No effect on filament population was observed, with 60mer- and 60mer-$MgCl_2$ populations having 25% and 22% of measured filaments longer than 100 nm, respectively. This confirms that σNS filament assembly is dependent on RNA structural stability, and suggests that assembly of such higher-order structures may play a role in σNS-mediated RNA remodelling.

The assembly of σNS higher-order species on full-length viral RNAs was then investigated. σNS was co-incubated with rotavirus S11 RNA (667-nt), as σNS has previously been shown to bind this RNA (Bravo *et al.*, 2018). Structure probing analysis has demonstrated that S11 is highly structured, with several conserved helices (Li *et al.*, 2010; Borodavka *et al.*, 2017). Interestingly, branched filaments were observed for σNS-S11 complexes (**Figure 6.17**), suggesting that the RNA structure may dictate σNS binding and therefore filament assembly.

RNA substrate	**Modal length (nm)**	**Number of filaments measured**
Unstructured 40mer (-1.6 kcal mol^{-1})	42.5	951
Unstructured 60mer (-1.6 kcal mol^{-1})	37.5	642
Stem loop A (-4.5 kcal mol^{-1}) (*SLA*)	37.5	504
Stem loop B (*SLB*) (-6.3 kcal mol^{-1})	37.5	507
Stem loop C (*SLC*) (-8.9 kcal mol^{-1})	32.5	502
SLA + 10 mM $MgCl_2$	27.5	548
SLB + 10 mM $MgCl_2$	27.5	585
SLC + 10 mM $MgCl_2$	N/A	N/A
60mer + 10 mM $MgCl_2$	47.5	565

Table 6.1 Summary of filament parameters assembled using different RNA substrates, as quantified in FiberApp (Usov and Mezzenga, 2015).

While this was not observed for any filaments assembled on <100-nt model RNA substrates as described above, the assembly of branched RNPs may reflect the structured nature of viral RNAs. In general, viral RNAs are highly structured and branched (Ben-Shaul and Gelbart, 2015; Borodavka *et al.*, 2016), and S11 has been demonstrated to contain several stable helices (Li *et al.*, 2010; Borodavka *et al.*, 2017).

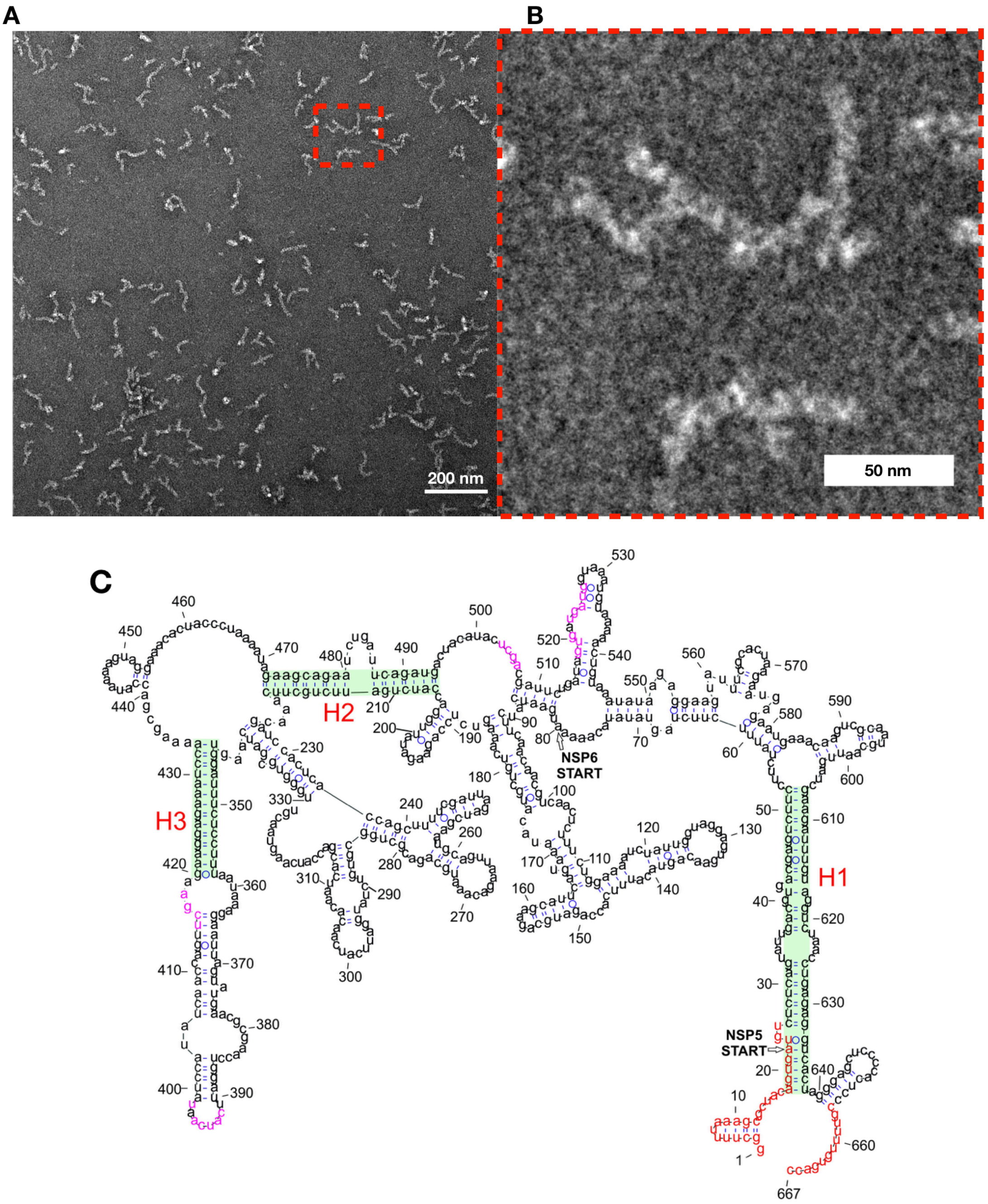


Figure 6.17 Assembly of branched structures with long, viral ssRNA. A & B: Rotavirus segment 11 (S11) : σNS complexes imaged using negative stain EM. Unlike filamentous RNPs assembled using model RNA structures (<100-nt), incubation of σNS with 667-nt S11 results in branched RNP complexes (B: red box, zoom). C: Minimum free energy secondary structure model of S11 adapted from (Borodavka *et al.*, 2017). Sequences predicted to be involved in tertiary contacts are red and magenta. Conserved helices H1 – H3 are highlighted in green.

While this was not observed for any filaments assembled on <100-nt model RNA substrates as described above, the assembly of branched RNPs may reflect the structured nature of viral RNAs. In general, viral RNAs are highly structured and branched (Ben-Shaul and Gelbart, 2015; Borodavka *et al.*, 2016), and S11 has been demonstrated to contain several stable helices (Li *et al.*, 2010; Borodavka *et al.*, 2017).

In this model, the RNP filaments provide a platform for RNA unwinding by σNS as part of its RNA chaperone activity. What is unclear, however, is how (or if) these filaments disassemble.

To understand this, σNS was incubated with 40mer RNA as previously described, and with a fully complementary 40mer RNA (40-C) and imaged using negative stain EM (**Figure 6.18A & B**). Both of these RNA substrates supported filament assembly. Next, 40mer and 40-C filaments were co-incubated at a 1 : 1 ratio for 30 minutes at room temperature prior to imaging. Surprisingly, this resulted in complete filament disassembly. Small, monodisperse particles (most likely corresponding to σNS) were highly abundant (**Figure 6.18C**). As a negative control, 40mer and 40-C RNAs were heat-annealed prior to incubation. This abrogated filament assembly (**Figure 6.18D**), as expected (since σNS does not bind dsRNA (Tourís-Otero *et al.*, 2005; Borodavka *et al.*, 2015)).

This result strongly supports the notion that filament assembly and disassembly are involved in remodelling RNA structure and promoting RNA-RNA annealing.

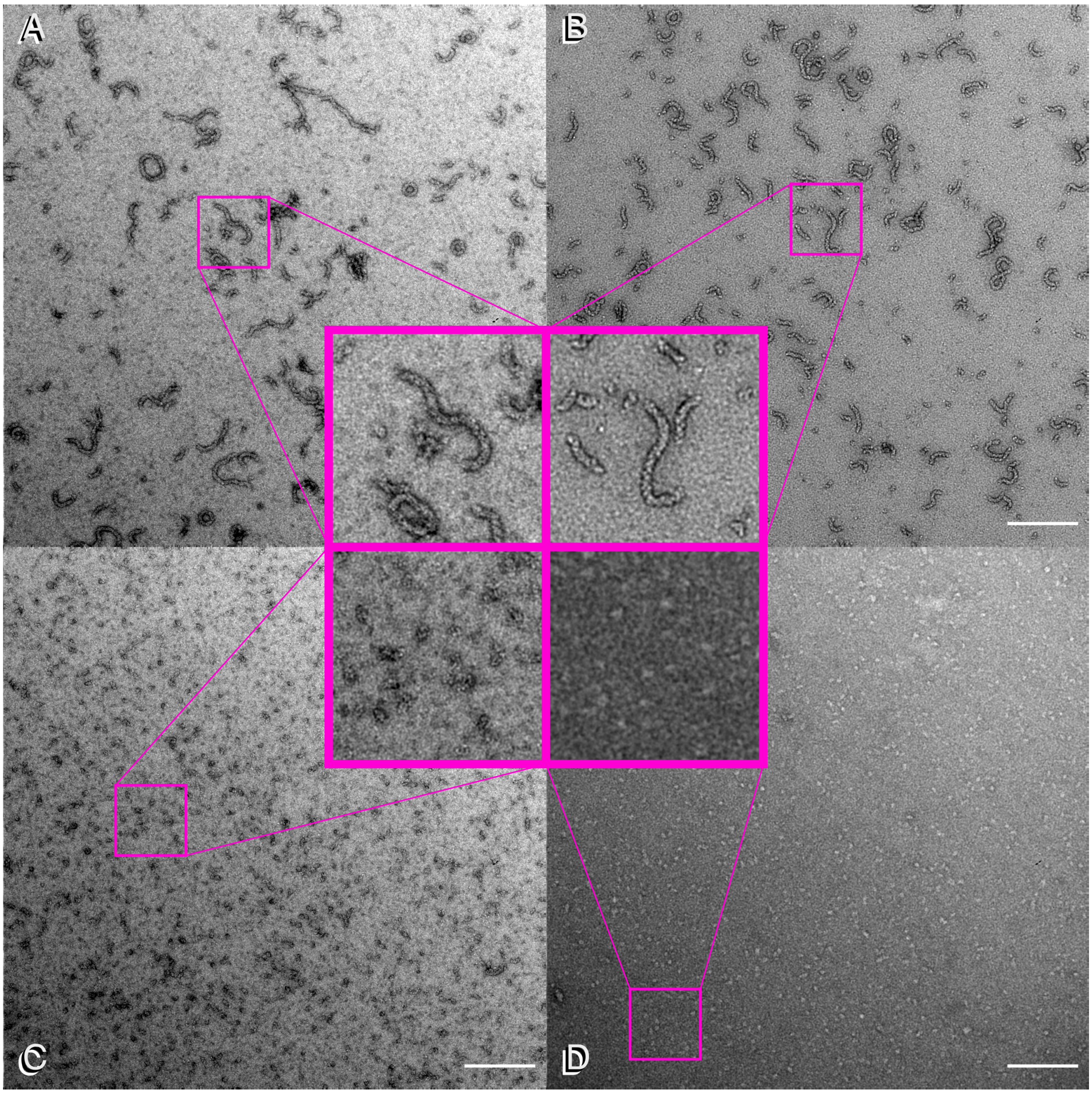


Figure 6.18 Assembly and disassembly of σNS RNP filaments. A & B: Assembly of σNS filaments as described above on 40mer (A) and 40mer-complementary (40-C) RNAs. These two RNAs have complete complementarity. C: Mixing of RNPs from A & B in a 1 : 1 ratio results in filament disassembly. D: Heat-annealing 40mer and 40-C RNAs prior to incubation with σNS prevents filament assembly. Scale bar in A – D: 200 nm

6.2.4 σNS protein-protein and protein-RNA interactions involved in filament assembly

Since RNA binding induces the assembly of different σNS oligomers, we investigated the interfaces between σNS oligomers within different RNP complexes. We mapped protein-protein interactions within σNS, σNS-20mer and σNS-40mer complexes using bis(sulfosuccinimidyl)suberate (BS3) crosslinking with mass spectrometry (BS3-XL-MS) (**Figure 6.19**). BS3 is a bifunctional crosslinker, forming covalent adducts with primary amines (i.e. N-terminus and lysine residues). Of the 8 primary amines within σNS, four participated in protein-protein interactions.

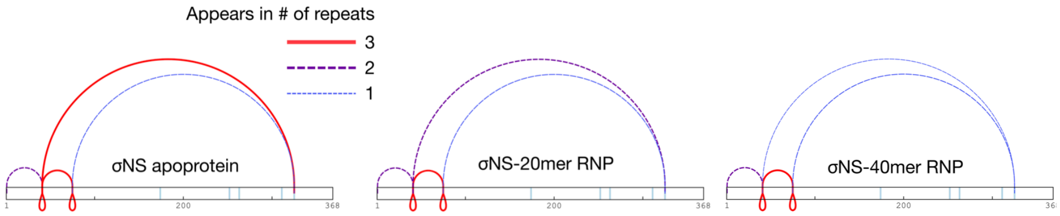


Figure 6.19 σNS protein-protein interactions within different σNS and its RNP complexes. BS3 XL-MS results for σNS, σNS-20mer and σNS-40mer RNPs. Blue horizontal lines represent residues with BS3-reactive primary amines (N-terminus and lysine residues). Red, purple and blue lines represent σNS protein-protein interactions. Loops indicate residues that crosslink to the same residue on a different σNS oligomer.

Strikingly, no differences in crosslink positions were observed between σNS apoprotein and either or the RNP complexes (**Figure 6.19**). This indicated that σNS utilises the same interfaces between monomers to assemble from hexamer to octamer, to higher-order RNP complexes. This is in agreement with the above structural data, where σNS octamer cryoEM densities were fitted into the σNS filament negative stain EM density. Furthermore, this supports the notion that populations of σNS hexamer and octamer are both present for apoprotein in solution as indicated by negative stain and cryoEM analysis.

A feature of interest from these crosslinking data is the loop-link formed between Lys45-Lys45 and Lys75-Lys75. As it is impossible to crosslink a residue to itself with BS3 (each Lys contains a single primary amine), these crosslinks unambiguously correspond to inter-monomer interactions.

The RNA-binding sites of σNS were determined using UV crosslinking (UV-XL). Protein-RNA interactions were detected throughout σNS, indicating that it does not have a single RNA-binding domain (**Figure 6.20**). Comparison of the RNA-binding sites and the BS3-XL-MS-identified protein-protein interactions indicate that some RNA-binding residues may form continuous RNA-binding sites through long-range protein-protein interactions within the context of tertiary or quaternary structure (notably, those within the N- and C-termini) (**Figure 6.20**).

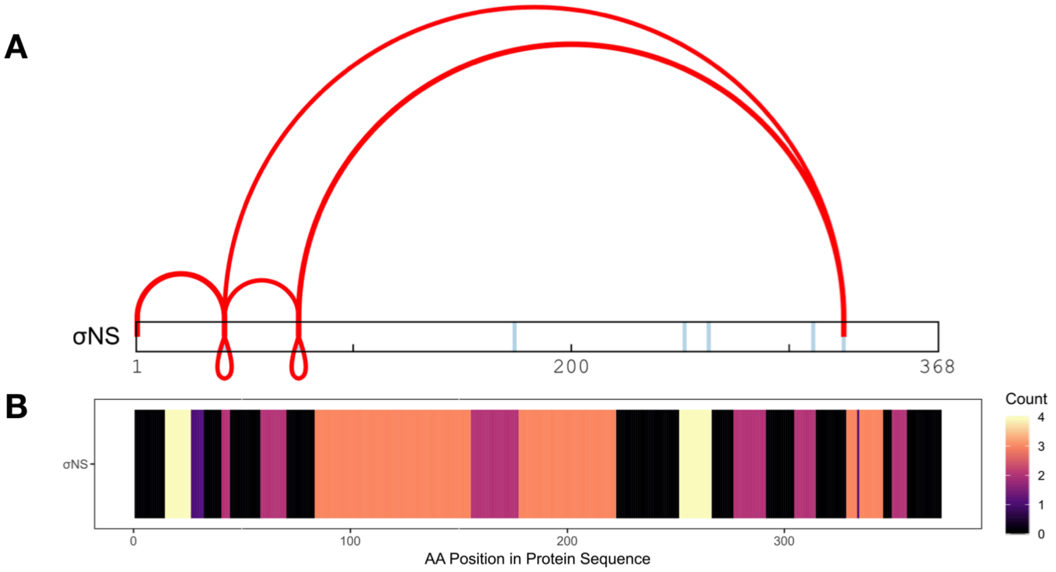


Figure 6.20 Comparison of σNS protein-protein and protein-RNA interaction sites. A: BS3 XL-MS links occurring in σNS apoprotein, σNS-20mer RNP and σNS-40mer RNP mapped to the sequence of σNS. B: UV-XL MS results: amino acids within σNS that directly crosslink to RNA are depicted as a heatmap.

6.2.5 Stability and oligomerisation of σNS

To understand the oligomeric equilibrium of σNS apoprotein, batch-mode SAXS analysis was performed at a wide range of σNS concentrations (~20 μM – 175 μM) (**Figure 6.21**). Guinier analysis indicated that all concentrations measured of σNS were monodisperse. No inter-particle interactions were observed to occur at higher σNS concentrations, indicating that σNS does not transition from hexamer to octamer in a concentration-dependent manner (**Figure 6.21**). This is best exemplified by the dependence of I(0) on [σNS], where a strong linear correlation is observed ($R^2 = 0.9997$). For a sample with a constant molecular weight, the I(0) should be directly proportional to the concentration, which is exactly what is observed. Additionally, there is negligible change of either the radius of gyration or maximum distance dimension across this range of σNS concentrations, indicating that no further oligomerisation has occurred. The prevalence of octamers in cryoEM analysis is therefore likely due to hexamer dissociation upon vitrification (hence the high abundance of dimeric σNS).

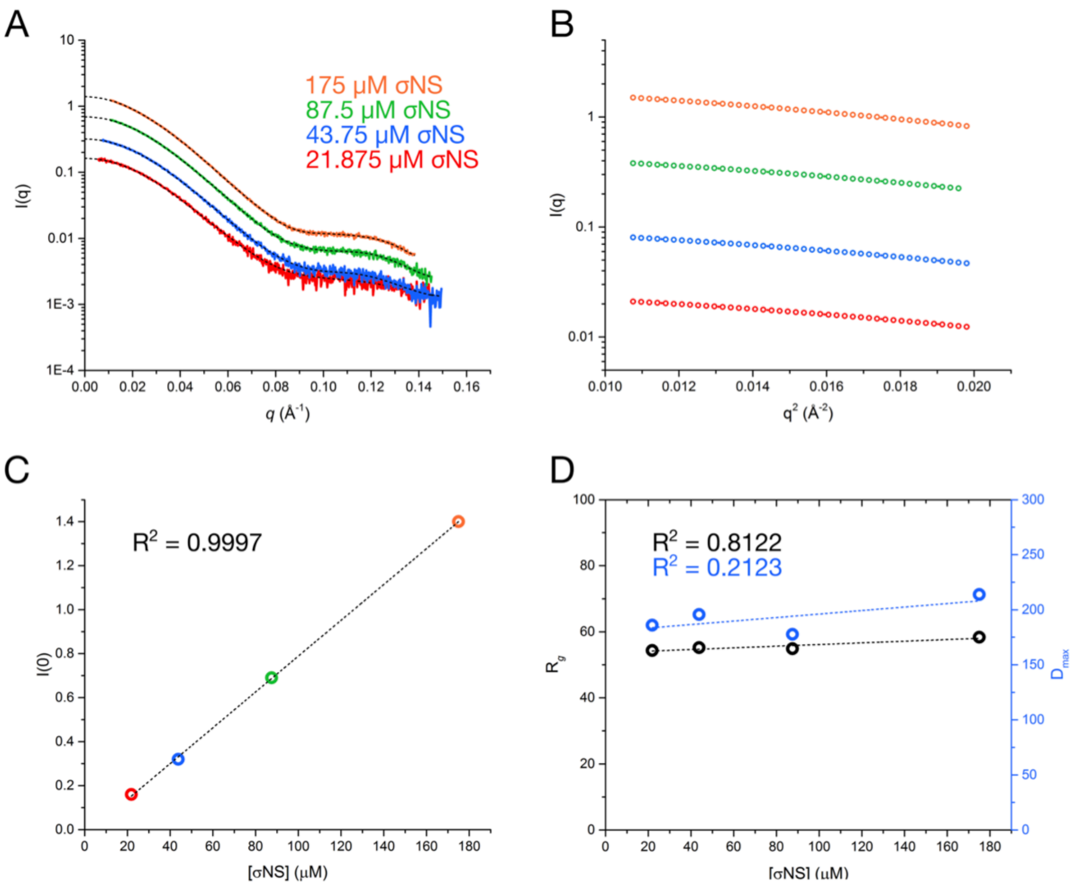


Figure 6.21 Oligomeric state of σNS in solution. A-D: SAXS analysis of σNS apoprotein at a range of concentrations. A: Scattering plot. B: Guiner plot. C: Plot of I(0) as a function of σNS concentration. D: Plot of radii of gyration, R_g (black, left y axis) and maximum distance dimension, D_{max}, (blue, right axis) as a function of σNS concentration.

If σNS is unable to assemble oligomers in the absence of RNA but relies on the same protein-protein interfaces present in apoprotein to assemble higher-order species, it is likely that the RNA is stabilising these larger oligomers. To investigate the influence of RNA on σNS stability, circular dichroism melting experiments were performed on σNS and a series of RNP complexes assembled using increasing lengths of RNA (10mer, 15mer, 20mer and 40mer). The fraction folded (CD signal at 222 nm) was measured during a temperature ramp (20°C – 80°C) (**Figure 6.22**).

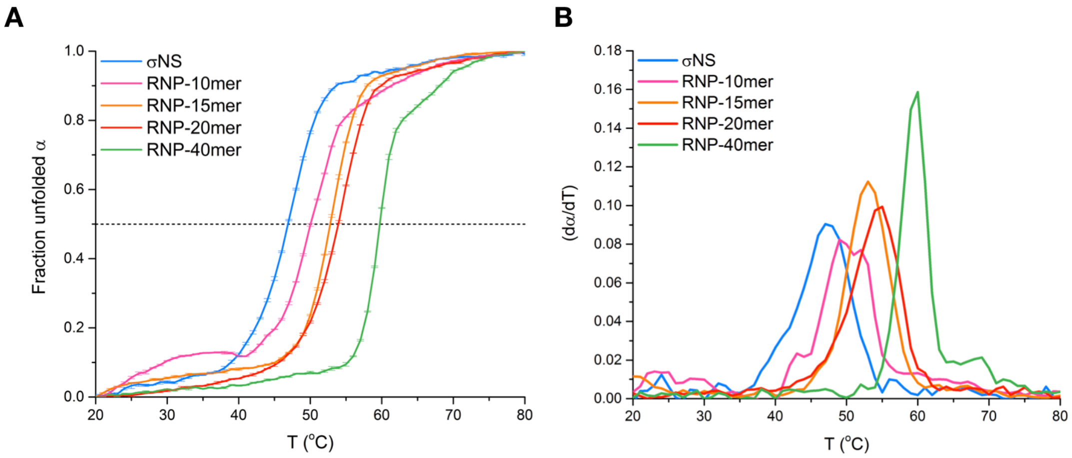


Figure 6.22 Stability of σNS RNP complexes. A: Circular dichroism (CD) melting experiments carried out with σNS in complex with different RNA substrates of increasing length. Θ_{222nm} was measured during a temperature ramp, and normalised signals (fraction unfolded, α) are plotted. B: Negative first derivates of α are plotted as a function of temperature to make otherwise small changes more obvious.

RNP complex formation resulted in increased σNS stability, as indicated by increased melting temperature (T_m). The T_m increases with the length of RNA used. As σNS binds RNA non-specifically, its affinity should increase with RNA length (Davidovich *et al.*, 2013; Davidovich and Cech, 2015). This suggests that σNS is able to more stably associate with longer RNA species. Notably, there is a large increase in the T_m of the RNP-20mer and RNP-40mer complexes (i.e. predominantly octameric σNS and filamentous σNS, respectively). This indicates that σNS filaments are a stable species, and that binding of RNA stabilises larger σNS oligomers, thus promoting filament assembly.

Together with previously published sedimentation velocity analysis and native mass spectrometry (Borodavka *et al.*, 2015; Bravo *et al.*, 2018), this indicates that σNS forms a single, major oligomer population in solution, which is predominantly hexameric (with a small octameric population). As σNS octamer is more stable than the σNS hexamer, vitrification-induced hexamer disassembly provides a source for the high abundance of σNS dimers observed in cryoEM analysis (**Figure 6.22**).

6.3 Discussion

Despite significant efforts to determine an atomic structure of σNS using X-ray crystallography, this was unsuccessful. These attempts were likely thwarted by the inherent oligomeric promiscuity of σNS. Nevertheless, the 6.7Å resolution cryoEM density map of σNS octamer provides unprecedented insights into the quaternary architecture of σNS, and provides a structural basis for the previously-observed hexamer-to-octamer transition (Bravo *et al.*, 2018). Visualisation of σNS density confirms a dimeric "building block" model, whereby transient hexamer-dimer interactions are stabilised by RNA, nucleating the assembly of larger σNS oligomers (**Figure 6.23**).

Interestingly, despite the low sequence similarity between mammalian reovirus (MRV) σNS and avian reovirus σNS, it is plausible that these two proteins share a similar mechanism of RNA chaperone activity (at least, more similar than that of σNS and NSP2). In fact, it was recently observed that MRV σNS can assemble in RNP filaments (Zamora *et al.*, 2018). This particular study shed light on many of the previously ambiguous aspects of MRV σNS function. σNS was implicated as a "stability factor" for viral RNAs by preventing their degradation somehow, although the precise mechanism of how this is achieved remains mysterious. Although the take-home message of this particular study was that "σNS forms RNA-protein complexes in preparation for genome replication," the observations reported are consistent with the role of MRV σNS as an RNA chaperone (Zamora *et al.*, 2018).

The work presented in this chapter further expands on the role of RNA-driven oligomerisation in the RNA chaperone activity of σNS. While it has been demonstrated that the hexamer-to-octamer oligomerisation increases the RNA unwinding activity of σNS (**Chapter 4**), the assembly of larger RNP filament structures may provide a platform to efficiently remodel viral RNA segments. The relationship between filament assembly and RNA structural stability again implicates these higher-order species as within the RNA chaperone activity of σNS.

While σNS assembles linear, non-branched filaments upon binding model RNA substrates, it was observed to form branched RNP complexes when in complex with highly structured viral RNA. The branched RNP structures presented here may represent an interplay between σNS filament assembly and stable RNA structure – while σNS can unwind and coat unstructured or metastable regions of S11, highly stable RNA structures may act as a roadblock, limiting σNS oligomerisation. This model is supported by the notion that although σNS is a non-specific RNA chaperone, its activity is dictated by the stability of RNA structure (**Chapter 4**) (Bravo *et al.*, 2018).

SAXS and CD melting experiments indicate that RNA binding stabilises σNS, and that this is necessary for higher-order RNP complex assembly. As σNS does not binds dsRNA, it dissociates upon formation of stable, extended duplexes (**Figure 6.23**). It remains unclear, however, how the RNA-RNA interaction is initiated, although sequences involved in such interactions may be exposed within the context of structured viral RNAs.

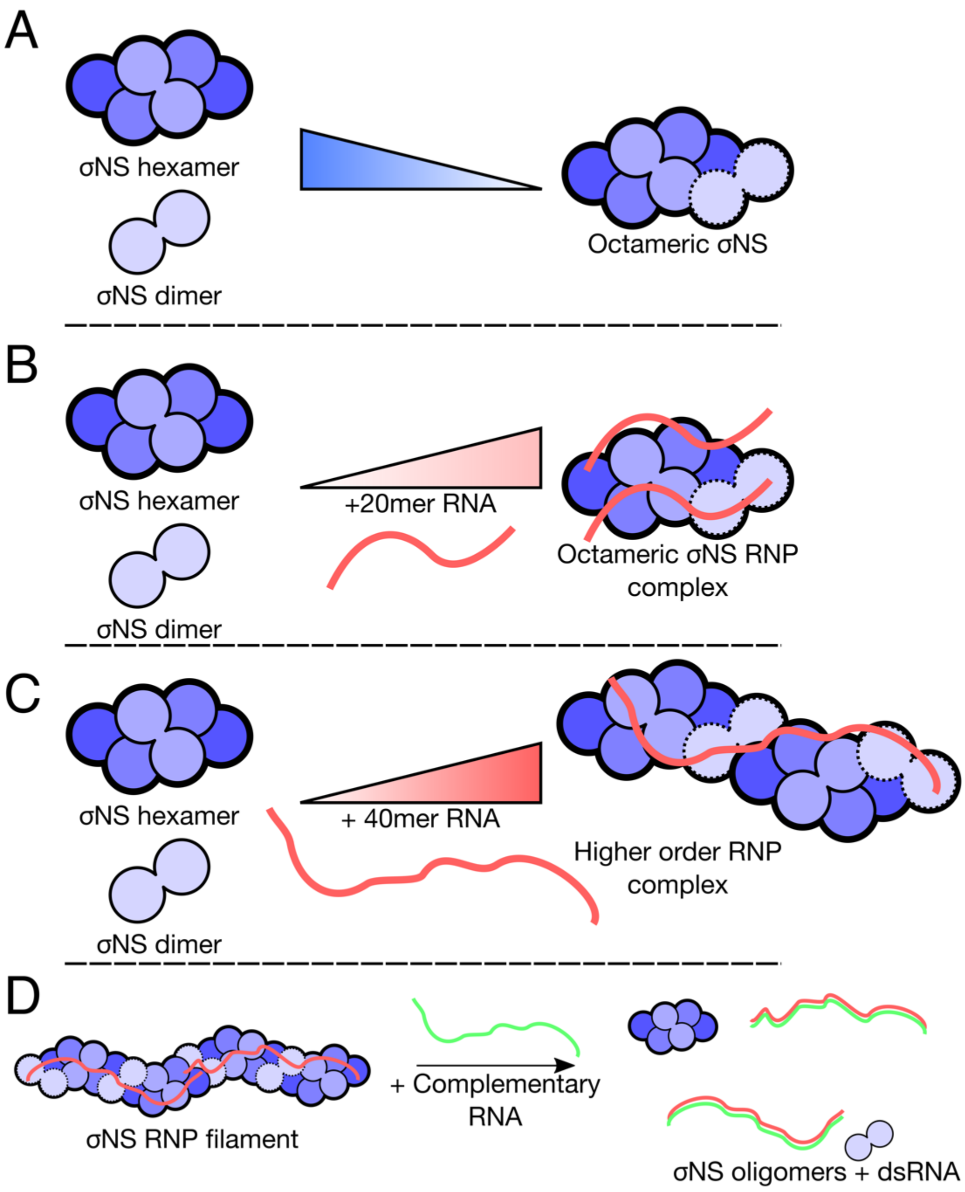


Figure 6.23 Model of σNS oligomerisation. A: σNS apoprotein oligomers are in equilibrium in solution, but in the absence of RNA oligomerisation is unfavourable. Hexamer and dimer may spontaneously associate, but octameric σNS is unlikely to be the dominant oligomeric state. B: Binding of multiple 20mer RNAs (the RNA binding footprint of σNS) is able to stabilise transient σNS octamers, shifting the equilibrium from hexamer/dimer mixture to octameric RNP complexes. C: Binding of longer RNAs further stabilise σNS oligomers, allowing assembly of higher-order RNP species. With unstructured ssRNAs, these RNP complexes are filaments. As the assembly of such species is dependent on RNA structural stability, σNS assembles into branched structures upon binding to full-length RNAs with highly stable RNA structures. D: Addition of complementary RNA results in filament disassembly and dsRNA annealing.

Chapter 7: Conclusions and Future Directions

Chapter 4: Comparison of RNA chaperone mechanisms of NSP2 and σNS

RV NSP2 and ARV σNS share several highly similar RNA chaperone-like activities and have been long postulated to play functionally analogous roles during viral replication and genome segment assortment (i.e. to facilitate sequence-specific intersegment RNA-RNA interactions required for selection and packaging of a complete viral genome). Despite this, it was unclear whether these two proteins utilise similar mechanisms to interact with and remodel RNAs. In this results chapter, it was demonstrated for the first time that NSP2 and σNS exploit different strategies to interact with and remodel RNA. This raises several questions: how do the activities of other RNA chaperone-like proteins from different reoviruses (such as bluetongue virus NS2) compare to NSP2 and σNS? It is therefore worthwhile expanding this investigation in the future to include other functionally analogous proteins to understand whether there are general themes in terms of viral RNA chaperone mechanisms.

The work presented in this chapter has used short, model RNA substrates to investigate the mechanisms with which NSP2 and σNS interact and remodel RNA. Although this has provided valuable insights, there remains a gap in our understanding of how either protein behaves on more complex, highly structured, full-length viral RNAs. It would therefore be worthwhile to directly demonstrate the RNA unfolding activities of NSP2 and σNS on full-length viral ssRNAs using SHAPE-based RNA structure probing (Boerneke, Ehrhardt and Weeks, 2019).

In this chapter, the RNA-RNA annealing activities of NSP2 and σNS were compared using RV segments S5 and S11, as this particular RNA-RNA interaction has been extensively characterised previously (Borodavka *et al.*, 2017). While σNS was not able to promote strand annealing in this particular assay, it was not possible to test its chaperone activities using ARV-derived RNAs as no intersegment RNA-RNA interactions have as yet been demonstrated. It would thus be highly insightful to investigate such intersegment RNA-RNA interactions and directly demonstrate that RNA matchmaking requires the activity of σNS, as has been demonstrated for RV RNAs and NSP2 (Borodavka *et al.*, 2017).

Chapter 5: Role of NSP2 CTD in RNA chaperone autoregulation

Expanding upon the results presented in Chapter 4, NSP2 was further investigated. The main focus of this work was to determine the structural basis for RNA binding by NSP2, and to establish a role for the C-terminal domain (CTD) of NSP2 within the context of its RNA chaperone activity. In previous studies, the C-terminus of NSP2 has been implicated as essential for RV replication, but its precise role in viral replication has remained somewhat elusive. While analysis of protein-mediated RNA-RNA interactions and RNA structure destabilisation implicated the CTD in the RNA chaperone activities of NSP2, its exact functional contributions were still somewhat unclear.

Here, the RNA-binding site of NSP2 was visualised using cryoEM, and directly confirmed through UV-XL. While UV-XL provides a single snapshot of peptides that interact with RNA, it does not provide information about regions of the protein that do not interact with RNA. In many ways hydrogen-deuterium exchange (HDX) is orthogonal to UV-XL, in that it provides peptide-resolution information about regions of proteins that become more or less exposed upon RNP complex formation (Lísal *et al.*, 2005). HDX would not only further corroborate the RNA-binding site of identified by UV-XL and cryoEM. Additionally, it would be worthwhile to perform either UV-XL or HDX on NSP2-ΔC to determine whether additional protein-RNA contacts occur in the absence of the CTD.

While the cryoEM structure of NSP2 RNP complex presented in this chapter provides the most high-resolution cryoEM map of NSP2 bound to RNA to date, it may be worth exploring structures of NSP2 bound to different RNA substrates. Although NSP2 binds to ssRNA, it may be worthwhile to design a substrate with a distinct structured domain (e.g. a tRNA), flanking an unstructured sequence (which would act as a platform for NSP2 to bind. By using an RNA substrate with a larger, well-defined tertiary structure, it may be possible to align RNA density more easily during cryoEM 3D reconstructions, potentially providing in better-resolved RNA density than in the RNP structure presented here.

Despite the work undertaken in this chapter, the molecular mechanism of CTD-mediated RNA dissociation from NSP2 still remains somewhat unclear. The CTD contains a surface-exposed acidic patch which we believe promotes

RNA dissociation from NSP2 due to charge repulsion, but the work here does not as yet rule out the possibility of competition for binding to the positive groove of NSP2 between the CTD and RNA, as is the case for Hfq (Santiago-Frangos *et al.*, 2017). Therefore, it may prove insightful to take the peptide corresponding to the CTD and investigate whether it directly binds to NSP2-ΔC, and if this interaction can compete with NSP2-RNA interactions.

Additionally, the charge repulsion hypothesis could be further investigated by performing further kinetic analysis of NSP2- and NSP2-ΔC – RNA interactions under different ionic conditions. In previous studies of RNA-protein interactions, K_{on} and K_D have a linear relationship with the salt concentration, whereas K_{off} is largely unaffected (Auweter *et al.*, 2006). If the CTD is not contributing to RNA dissociation through charge repulsion, both full-length NSP2 and NSP2-ΔC should display similar changes in binding kinetics at different salt concentrations, and K_{off} should remain unaffected.

Having determined the residues that interact with RNA, it may be possible to use the rotavirus reverse-genetics system to test what affect these mutation of residues would have *in vivo* (Desselberger, 2017).

Chapter 6: σNS structure and function

An initial major aim of this project was to determine a crystal structure of σNS. While ultimately unsuccessful, in hindsight it is somewhat obvious that a protein that exists in equilibrium between dimer, hexamer and octamer forms may not be the most amenable for structure determination through X-ray crystallography. Instead, cryoEM was used to determine a modest-resolution (6.7 Å) structure of σNS octamer. At such a resolution it is not possible to build in an atomic model de novo. In spite of this, this structure offered insights into the quaternary architecture of σNS, and provided a structural basis for the assembly of different oligomers.

There were two major issues that may have prevented the determination of a higher-resolution map: specimen preferred orientation and the low number of non-dimeric particles. To deal with the specimen orientation issue, further cryoEM data could be acquired either with tilted grids (Tan *et al.*, 2017; Beckert *et al.*, 2018) or with σNS vitrified in the presence of the detergent CHAPSO (Chen *et al.*, 2018). To overcome the issue of σNS dissociation in ice and the high abundance of dimeric σNS that occurs as a consequence of this, it may be worthwhile to use a brute force approach – to simply collect a sufficient amount of data so that after all σNS dimers have been removed from the dataset there are still enough hexameric and octameric particles remaining to determine a high-resolution 3D reconstruction.

However, once a high-resolution cryoEM density map has been determined, there would be several other challenges, most notably building a structural model of σNS de novo. This is not a trivial process and would require a high quality, high resolution map and even then, it may prove highly challenging to unambiguously model the entire structure of σNS (given that there are no known proteins with structures determined that are similar to σNS in terms of sequence or structure). However, to understand the molecular basis for RNA-driven oligomerisation in atomic detail, this is essential. With an atomic model, it would also be possible to use the UV-XL data to map the RNA-binding site (or sites) to σNS.

It would also be of interest to determine a high-resolution structure of σNS RNP filaments. Although negative stain analysis has demonstrated that these higher-order species are highly flexible, it may be possible to employ a single-particle analysis-based approach as an alternative to helical reconstruction (Hunkeler *et al.*, 2018).

Interestingly, in a previous study MRV σNS was demonstrated to interact with the pre-initiation complex (PIC) through interactions with eIF3A and the ribosomal protein pS6R (Desmet, Anguish and Parker, 2014). Genome segments from *Reoviridae* do not contain internal ribosome entry sites (IRES) and lack polyA tails, so must therefore utilise alternate strategies of ribosome capture. While rotavirus encodes NSP3 which can hijack host translational machinery through by mimicking polyA-binding protein, it is unknown whether avian reovirus uses a similar mechanism. Ribosomes are an ideal specimen for cryoEM, and the PIC-σNS complex may be a worthwhile pursuit (Vilas *et al.*, 2018).

Chapter 8: References

Adrian, M. *et al.* (1984) ‘Cryo-electron microscopy of viruses’, *Nature*, 308, pp. 32–36. doi: 10.1142/9781848164666_0001.

Akita, F. *et al.* (2011) ‘Viroplasm matrix protein Pns9 from rice gall dwarf virus forms an octameric cylindrical structure’, *Journal of General Virology*, 92(9), pp. 2214–2221. doi: 10.1099/vir.0.032524-0.

Akita, F. *et al.* (2012) ‘Crystallographic Analysis Reveals Octamerization of Viroplasm Matrix Protein P9-1 of Rice Black Streaked Dwarf Virus’, *Journal of Virology*, 86(2), pp. 746–756. doi: 10.1128/JVI.00826-11.

Allewell, N. (2013) *Biophysics for the Life Sciences*, *Springer*. doi: 10.1007/978-1-4614-4954-6.

Andrade, J. M. *et al.* (2018) ‘The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity’, *The EMBO Journal*, p. e97631. doi: 10.15252/embj.201797631.

Auweter, S. D. *et al.* (2006) ‘Molecular basis of RNA recognition by the human alternative splicing factor Fox-1’, *The EMBO Journal*, 25(1), pp. 163–173. doi: 10.1038/sj.emboj.7600918.

Aw, J. G. A. *et al.* (2016) ‘In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation’, *Molecular Cell*. Elsevier Inc., 62(4), pp. 603–617. doi: 10.1016/j.molcel.2016.04.028.

Bailor, M. H. *et al.* (2011) ‘Topological constraints: Using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation’, *Current Opinion in Structural Biology*. Elsevier Ltd, 21(3), pp. 296–305. doi: 10.1016/j.sbi.2011.03.009.

Barth, A. *et al.* (2018) ‘Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of Clostridium thermocellum’. doi: 10.1073/pnas.1809283115.

Barth, A., Voithenberg, L. V. Von and Lamb, D. C. (2014) ‘MFD-PIE and PIE-FI : Ways to Extract More Information with TCSPC’. doi: 10.1007/4243.

Batey, R. T., Rambo, R. P. and Doudna, J. A. (1999) ‘Tertiary motifs in RNA structure and folding’, *Angewandte Chemie - International Edition*, 38(16), pp. 2326–2343. doi: 10.1002/(SICI)1521-3773(19990816)38:16<2326::AID-ANIE2326>3.0.CO;2-3.

Baudin, F. *et al.* (1994) ‘Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent.’, *The EMBO journal*, 13(13), pp. 3158–65. doi: 10.1002/j.1460-2075.1994.tb06614.x.

Beckert, B. *et al.* (2018) ‘Structure of a hibernating 100S ribosome reveals an inactive conformation of the ribosomal protein S1’, *Nature Microbiology*. Springer US, 3(10), pp. 1115–1121. doi: 10.1038/s41564-018-0237-0.

Ben-Shaul, A. and Gelbart, W. M. (2015) ‘Viral ssRNAs are indeed compact’, *Biophysical Journal*, 108(1), pp. 14–16. doi: 10.1016/j.bpj.2014.11.010.

Bieniasz, P. and Telesnitsky, A. (2018) ‘Multiple, Switchable Protein:RNA Interactions Regulate Human Immunodeficiency Virus Type 1 Assembly’, *Annual Review of Virology*, 5(1), p. annurev-virology-092917-043448. doi: 10.1146/annurev-virology-092917-043448.

Bingaman, J. L. *et al.* (2017) ‘The GlcN6P cofactor plays multiple catalytic roles in the glmS ribozyme’, *Nature Chemical Biology*. Nature Publishing Group, 13(4), pp. 439–445. doi: 10.1038/nchembio.2300.

Blanchet, C. E. and Svergun, D. I. (2013) ‘Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution.’, *Annual review of physical chemistry*, 64(November 2012), pp. 37–54. doi: 10.1146/annurev-physchem-040412-110132.

Boerneke, M. A., Ehrhardt, J. E. and Weeks, K. M. (2019) ‘Physical and Functional Analysis of Viral RNA Genomes by SHAPE’, *Annual Review of Virology*, 6(1), p. annurev-virology-092917-043315. doi: 10.1146/annurev-virology-092917-043315.

Borodavka, A. *et al.* (2015) ‘Evidence that avian reovirus σNS is an RNA chaperone: implications for genome segment assortment’, *Nucleic Acids Research*. Oxford University Press (OUP), 43(14), pp. 7044–7057. doi: 10.1093/nar/gkv639.

Borodavka, A. *et al.* (2016) ‘Sizes of Long RNA Molecules Are Determined by the Branching Patterns of Their Secondary Structures’, *Biophysical Journal*. Biophysical Society, 111(10), pp. 2077–2085. doi: 10.1016/j.bpj.2016.10.014.

Borodavka, A. *et al.* (2017) ‘Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments’, *eLife*. eLife Sciences Publications, Ltd, 6, pp. 1–22. doi: 10.7554/eLife.27453.

Borodavka, A., Desselberger, U. and Patton, J. T. (2018) ‘Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes’, *Current Opinion in Virology*. Elsevier B.V., 33, pp. 106–112. doi: 10.1016/j.coviro.2018.08.001.

Bowman, J. C. *et al.* (2012) ‘Cations in charge: Magnesium ions in RNA folding and catalysis’, *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 262–272. doi: 10.1016/j.sbi.2012.04.006.

Bravo, J. P. K. *et al.* (2018) ‘Stability of local secondary structure determines selectivity of viral RNA chaperones’, *Nucleic Acids Research*, 46(15), pp. 7924–7937. doi: 10.1093/nar/gky394.

Brescia, C. C. *et al.* (2003) ‘Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure.’, *RNA (New York, N.Y.)*, 9(1), pp. 33–43. doi: 10.1261/rna.2570803.

Briggs, J. A. G. *et al.* (2003) ‘Structural organization of authentic , mature HIV-1 virions and cores’, 22(7), pp. 1707–1715.

Brigham, B. S. *et al.* (2019) ‘Intrinsic conformational dynamics of the HIV-1 genomic RNA 5′UTR’, *Proceedings of the National Academy of Sciences*, p.

201902271. doi: 10.1073/pnas.1902271116.

Broering, T. J. *et al.* (2004) ‘Reovirus Nonstructural Protein uNS Recruits Viral Core Surface Proteins and Entering Core Particles to Factory-Like Inclusions’, *Journal of Virology*, 78(4), pp. 1882–1892. doi: 10.1128/JVI.78.4.1882–1892.2004.

Bush, M. F. *et al.* (2010) ‘Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology’, *Analytical Chemistry*, 82(22), pp. 9557–9565. doi: 10.1021/ac1022953.

Butan, C. and Tucker, P. (2010) ‘Insights into the role of the non-structural protein 2 (NS2) in Bluetongue virus morphogenesis’, *Virus Research*, pp. 109–117. doi: 10.1016/j.virusres.2010.05.014.

Butan, C., Van Der Zandt, H. and Tucker, P. a (2004) ‘Structure and assembly of the RNA binding domain of bluetongue virus non-structural protein 2.’, *The Journal of biological chemistry*, 279(36), pp. 37613–37621. doi: 10.1074/jbc.M400502200.

Butler, P. J. G. and Klug, A. (1971) ‘Assembly of the particle of tobacco mosaic virus from RNA and disks of protein’, *Nature New Biology*, 229(2), pp. 47–50. doi: 10.1038/newbio229047a0.

Calabrese, A. N. and Radford, S. E. (2018) ‘Mass spectrometry-enabled structural biology of membrane proteins’, *Methods*, 147, pp. 187–205.

Carlomagno, T. (2014) ‘Present and future of NMR for RNA-protein complexes: A perspective of integrated structural biology’, *Journal of Magnetic Resonance*. Elsevier Inc., 241(1), pp. 126–136. doi: 10.1016/j.jmr.2013.10.007.

Castello, A. *et al.* (2016) ‘Comprehensive Identification of RNA-Binding Domains in Human Cells’, *Molecular Cell*. The Author(s), 63(4), pp. 696–710. doi: 10.1016/j.molcel.2016.06.029.

Cech, T. R. and Steitz, J. A. (2014) ‘The noncoding RNA revolution - Trashing old rules to forge new ones’, *Cell*. Elsevier Inc., 157(1), pp. 77–94. doi: 10.1016/j.cell.2014.03.008.

Chen, J. *et al.* (2018) ‘Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO’, *bioRxiv*, p. 457267. doi: 10.1101/457267.

Chen, Y. *et al.* (2008) ‘DEAD-box proteins can completely separate an RNA duplex using a single ATP.’, *Proceedings of the National Academy of Sciences of the United States of America*, 105(51), pp. 20203–8. doi: 10.1073/pnas.0811075106.

Chen, Y. and Pollack, L. (2016) ‘SAXS studies of RNA: structures, dynamics, and interactions with partners’, *Wiley Interdisciplinary Reviews: RNA*, 7(4), pp. 512–526. doi: 10.1002/wrna.1349.

Cheng, Y. *et al.* (2015) ‘A primer to single-particle cryo-electron microscopy’, *Cell*, 161(3), pp. 438–449. doi: 10.1016/j.cell.2015.03.050.

Chou, Y.-Y. *et al.* (2012) ‘One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis.’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(23), pp. 9101–6. doi: 10.1073/pnas.1206069109.

Chu, V. B. and Herschlag, D. (2008) ‘Unwinding RNA’s secrets: advances in the biology, physics, and modeling of complex RNAs’, *Current Opinion in Structural Biology*, 18(3), pp. 305–314. doi: 10.1016/j.sbi.2008.05.002.

Comas-Garcia, M. *et al.* (2017) ‘Dissection of specific binding of HIV-1 Gag to the ’packaging signal’ in viral RNA’, *eLife*, 6, pp. 1–27. doi: 10.7554/eLife.27055.

Combe, C. W., Fischer, L. and Rappsilber, J. (2015) ‘xiNET: Cross-link network maps with residue resolution’, *Molecular and Cellular Proteomics*, 14(4), pp. 1137–1147. doi: 10.1074/mcp.O114.042259.

Costantino, D. A. *et al.* (2008) ‘tRNA-mRNA mimicry drives translation initiation from a viral IRES’, *Nature Structural and Molecular Biology*, 15(1), pp. 57–64. doi: 10.1038/nsmb1351.

Criglar, J. M. *et al.* (2014) ‘A Novel Form of Rotavirus NSP2 and Phosphorylation-Dependent NSP2-NSP5 Interactions Are Associated with Viroplasm Assembly.’, *Journal of Virology*, 88(2), pp. 786–98. doi: 10.1128/JVI.03022-13.

Criglar, J. M. *et al.* (2018) ‘Phosphorylation cascade regulates the formation and maturation of rotaviral replication factories’, *Proceedings of the National Academy of Sciences of the United States of America*, 115(51), pp. E12015–E12023. doi: 10.1073/pnas.1717944115.

Cruz, J. A. and Westhof, E. (2009) ‘The Dynamic Landscapes of RNA Architecture’, *Cell*, pp. 604–609. doi: 10.1016/j.cell.2009.02.003.

D’Ascenzo, L. *et al.* (2016) ‘Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops’, *Rna*, 23(3), pp. 259–269. doi: 10.1261/rna.059097.116.

D’Imprima, E. *et al.* (2019) ‘Protein denaturation at the air-water interface and how to prevent it’, *eLife*, 8, pp. 1–18. doi: 10.7554/eLife.42747.

D’Souza, V. M. and Summers, M. F. (2005) ‘How retroviruses select their genomes’, *Nature Reviews Microbiology*, pp. 643–655. doi: 10.1038/nrmicro1210.

D’Souza, V. and Summers, M. F. (2004) ‘Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus’, *Nature*, 431(7008), pp. 586–590. doi: 10.1038/nature02944.

Dadonaite, B. *et al.* (2019) ‘The structure of the influenza A virus genome’, *Nature Microbiology*. Springer US. doi: 10.1038/s41564-019-0513-7.

Davidovich, C. *et al.* (2013) ‘Promiscuous RNA binding by Polycomb repressive complex 2.’, *Nature Structural & Molecular Biology*, 20(11), pp. 1250–1257. doi: 10.1038/nsmb.2679.

Davidovich, C. and Cech, T. R. (2015) ‘The recruitment of chromatin modifiers by long noncoding RNAs: Lessons from PRC2', *RNA (New York, N.Y.)*, 21(12), pp. 2007–2022. doi: 10.1261/rna.053918.115.

Deo, R. C. *et al.* (2002) 'Recognition of the rotavirus mRNA 3′ consensus by an asymmetric NSP3 homodimer', *Cell*, 108(1), pp. 71–81. doi: 10.1016/S0092-8674(01)00632-8.

Desmet, E. A., Anguish, L. J. and Parker, J. S. L. (2014) 'Virus-mediated compartmentalization of the host translational machinery', *mBio*. American Society for Microbiology, 5(5), pp. 1–11. doi: 10.1128/mBio.01463-14.

Desselberger, U. (2017) 'At last : a fully tractable , plasmid only based reverse genetics system for rotavirus', 12, pp. 519–524.

Dimastrogiovanni, D. *et al.* (2014) 'Recognition of the small regulatory RNA RydC by the bacterial Hfq protein.', *eLife*, 3, pp. 1–19. doi: 10.7554/eLife.05375.

Ding, K. *et al.* (2019) 'In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release', *Nature Communications*. Springer US, 10(1), p. 2216. doi: 10.1038/s41467-019-10236-7.

Ding, K., Nguyen, L. and Zhou, Z. H. (2018) ' In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating ', *Journal of Virology*, 92(21). doi: 10.1128/jvi.00774-18.

Diskin, R. *et al.* (2013) 'Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies.', *The Journal of experimental medicine*, 210(6), pp. 1235–49. doi: 10.1084/jem.20130221.

Doetsch, M. *et al.* (2010) 'Mechanisms of StpA-mediated RNA remodeling', *RNA Biology*, 7(6), pp. 735–743. doi: 10.4161/rna.7.6.13882.

Doudna, J. A. and Cech, T. R. (2002) 'The chemical repertoire of natural ribozymes', 418(July).

Draper, D. E. (1999) 'On the Role of Magnesium Ions in RNA Stability', 48, pp. 113–135.

Draper, D. E. (2004) 'A guide to ions and RNA structure A guide to ions and RNA structure', *RNA*, 10(3), pp. 335–343. doi: 10.1261/rna.5205404.and.

Draper, D. E., Grilley, D. and Soto, A. M. (2005) 'Ions and RNA Folding', *Annual Review of Biophysics and Biomolecular Structure*, 34(1), pp. 221–243. doi: 10.1146/annurev.biophys.34.040204.144511.

Draper, D. E. and von Hippel, P. H. (1978) 'Nucleic acid binding properties of Escherichia coli ribosomal protein S1', *Journal of Molecular Biology*, 122(3), pp. 321–338. doi: 10.1016/0022-2836(78)90193-6.

Drulyte, I. *et al.* (2018) 'Approaches to altering particle distributions in cryo-electron microscopy sample preparation', *Acta Crystallographica Section D Structural Biology*. International Union of Crystallography, 74(6), pp. 1–12. doi: 10.1107/S2059798318006496.

Dubochet, J. *et al.* (1988) 'Cryo-electron microscopy of vitrified specimens', *Quarterly Reviews of Biophysics*, 21(2), pp. 129–228. doi: 10.1017/S0033583500004297.

Eckwahl, M. J. *et al.* (2015) ‘A retrovirus packages nascent host noncoding RNAs from a novel surveillance pathway’, *Genes and Development*, 29(6), pp. 646–657. doi: 10.1101/gad.258731.115.

Eckwahl, M. J. *et al.* (2016) ‘Analysis of the human immunodeficiency virus-1 RNA packageome’, *RNA*, 22(8), pp. 1228–1238. doi: 10.1261/rna.057299.116.

Eggeling, C. *et al.* (1998) ‘Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy’, *Proc. Natl. Acad. Sci. U. S. A.*, 95(4), pp. 1556–1561. doi: 10.1073/pnas.95.4.1556.

Eichwald, C. *et al.* (2004) ‘Uncoupling substrate and activation functions of rotavirus NSP5: phosphorylation of Ser-67 by casein kinase 1 is essential for hyperphosphorylation.’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), pp. 16304–9. doi: 10.1073/pnas.0406691101.

Essere, B. *et al.* (2013) ‘Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses’, *Proceedings of the National Academy of Sciences*, 110(40), pp. E3840–E3848. doi: 10.1073/pnas.1308649110.

Estrozi, L. F. *et al.* (2013) ‘Location of the dsRNA-Dependent Polymerase, VP1, in Rotavirus Particles’, *Journal of Molecular Biology*. Elsevier Ltd, 425(1), pp. 124–132. doi: 10.1016/j.jmb.2012.10.011.

Fabbretti, E. *et al.* (1999) ‘Two non-structural rotavirus proteins, NSP2 and NSP5, form viroplasm-like structures in vivo’, *Journal of General Virology*, 80(2), pp. 333–339. doi: 10.1099/0022-1317-80-2-333.

Fajardo, T. J., Sung, P. Y. and Roy, P. (2015) ‘Disruption of Specific RNA-RNA Interactions in a Double-Stranded RNA Virus Inhibits Genome Packaging and Virus Infectivity’, *PLoS Pathogens*, 11(12), pp. 1–22. doi: 10.1371/journal.ppat.1005321.

Fang, X. *et al.* (2013) ‘An unusual topological structure of the HIV-1 rev response element’, *Cell*. Elsevier Inc., 155(3), pp. 594–605. doi: 10.1016/j.cell.2013.10.008.

Filbin, M. E. *et al.* (2012) ‘HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation’, *Nature Structural & Molecular Biology*, 20(2), pp. 150–158. doi: 10.1038/nsmb.2465.

Fiore, J. L., Holmstrom, E. D. and Nesbitt, D. J. (2011) ‘Entropic origin of Mg 2þ -facilitated RNA folding’, *Pnas*, 109(8), pp. 2902–2907. doi: 10.1073/pnas.1114859109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1114859109.

Franke, D. and Svergun, D. I. (2009) ‘DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering’, *Journal of Applied Crystallography*. International Union of Crystallography, 42(2), pp. 342–346.

doi: 10.1107/S0021889809000338.

Freed, E. O. (2015) 'HIV-1 assembly, release and maturation', *Nature Reviews Microbiology*. Nature Publishing Group, 13(8), pp. 484–496. doi: 10.1038/nrmicro3490.

Gao, Q. and Palese, P. (2009) 'Rewiring the RNAs of influenza virus to prevent reassortment', *Proceedings of the National Academy of Sciences of the United States of America*, 106(37), pp. 15891–15896. doi: 10.1073/pnas.0908897106.

Gartland, W. J. and Sueoka, N. (1966) 'Two interconvertible forms of tryptophanyl sRNA in E. coli.', *Proceedings of the National Academy of Sciences of the United States of America*, 55(4), pp. 948–956. doi: 10.1073/pnas.55.4.948.

Gavazzi, C. *et al.* (2013) 'A functional sequence-specific interaction between influenza A virus genomic RNA segments', *Proceedings of the National Academy of Sciences*, 110(41), pp. 16604–16609. doi: 10.1073/pnas.1314419110.

Gerber, M. *et al.* (2014) 'Selective packaging of the influenza A genome and consequences for genetic reassortment', *Trends in Microbiology*, pp. 446–455. doi: 10.1016/j.tim.2014.04.001.

Giese, S., Bolte, H. and Schwemmle, M. (2016) 'The feat of packaging eight unique genome segments', *Viruses*, 8(6), pp. 1–11. doi: 10.3390/v8060165.

Gillian, A. L. *et al.* (2000) 'Reovirus protein sigma NS binds in multiple copies to single stranded RNA and shares properties with single stranded DNA binding proteins.', *Journal of Virology*, 74(13), pp. 5939–5948.

Gillian, A. L. and Nibert, M. L. (1998) 'Amino terminus of reovirus nonstructural protein sigma NS is important for ssRNA binding and nucleoprotein complex formation', *Virology*, 240(1), pp. 1–11. doi: S0042682297989054 [pii].

Glaeser, R. M. *et al.* (2015) 'Factors that Influence the Formation and Stability of Thin, Cryo-EM Specimens.', *Biophysical journal*. doi: 10.1016/j.bpj.2015.07.050.

Glaeser, R. M. (2019) 'How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits?', *Annual Review of Biophysics*, 48(1), pp. 45–61. doi: 10.1146/annurev-biophys-070317-032828.

Goddard, T. D. *et al.* (2018) 'UCSF ChimeraX: Meeting modern challenges in visualization and analysis', *Protein Science*, 27(1), pp. 14–25. doi: 10.1002/pro.3235.

Gopal, A. *et al.* (2014) 'Viral RNAs are unusually compact', *PLoS ONE*, 9(9). doi: 10.1371/journal.pone.0105875.

Gorski, S. A., Vogel, J. and Doudna, J. A. (2017) 'RNA-based recognition and targeting: sowing the seeds of specificity', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(4), pp. 215–228. doi:

10.1038/nrm.2016.174.

Groft, C. M. and Burley, S. K. (2002) ‘Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization’, *Molecular Cell*, 9(6), pp. 1273–1283. doi: 10.1016/S1097-2765(02)00555-5.

Grohman, J. K. *et al.* (2013) ‘A Guanosine-Centric Mechanism for RNA Chaperone Function’, *Science (New York, N.Y.)*, 340(6129), pp. 190–195. doi: 10.1126/science.1230715.

Grossberger, R. *et al.* (2005) ‘Influence of RNA structural stability on the RNA chaperone activity of the Escherichia coli protein StpA’, *Nucleic Acids Research*, 33(7), pp. 2280–2289. doi: 10.1093/nar/gki515.

Gruber, A. R. *et al.* (2008) ‘The Vienna RNA websuite.’, *Nucleic acids research*, 36(Web Server issue), pp. 70–74. doi: 10.1093/nar/gkn188.

‘Guinier, Fournet, Small Angle Scattering of X-Rays (1955) .pdf’ (no date).

Guo, J. U. and Bartel, D. P. (2016) ‘Rotaviruses’, *Science (New York, N.Y.)*. Elsevier BV, 353(6306). doi: 10.1016/j.virusres.2014.06.016.

He, Y. *et al.* (2019) ‘In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating’, *Proceedings of the National Academy of Sciences*, 116(33), pp. 16535–16540. doi: 10.1073/pnas.1905849116.

Henderson, R. (1995) ‘The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules’, *Quarterly Reviews of Biophysics*, 28(2), pp. 171–193. doi: 10.1017/S003358350000305X.

Henderson, R. (2013) ‘Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(45), pp. 18037–18041. doi: 10.1073/pnas.1314449110.

Herschlag, D. *et al.* (1994) ‘An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis.’, *The EMBO journal*, 13(12), pp. 2913–24. doi: 10.1016/j.jim.2013.11.004.

Herschlag, D. (1995) ‘RNA chaperones and the folding problem’, *Journal of Biological Chemistry*, 270, pp. 20871–20874. doi: 10.1074/jbc.270.36.20871.

Herzik, M. A., Wu, M. and Lander, G. C. (2017) ‘Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV’, *Nature Methods*, 14(11), pp. 1075–1078. doi: 10.1038/nmeth.4461.

Herzik, M. A., Wu, M. and Lander, G. C. (2019) ‘High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM’, *Nature Communications*. Springer US, 10(1), pp. 1–9. doi: 10.1038/s41467-019-08991-8.

Holmqvist, E. *et al.* (2016) ‘Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo.’, *The EMBO journal*, p. e201593360. doi: 10.15252/embj.201593360.

Holmstrom, E. D. *et al.* (2019) ‘Disordered RNA chaperones can enhance nucleic acid folding via local charge screening’, *Nature Communications*. Springer US, 10(1), p. 2453. doi: 10.1038/s41467-019-10356-0.

Hopkins, J. F. *et al.* (2009) ‘Effect of salt and RNA structure on annealing and strand displacement by Hfq’, *Nucleic Acids Research*, 37(18), pp. 6205–6213. doi: 10.1093/nar/gkp646.

Hopkins, J. F., Panja, S. and Woodson, S. A. (2011) ‘Rapid binding and release of Hfq from ternary complexes during RNA annealing’, *Nucleic Acids Research*, 39(12), pp. 5193–5202. doi: 10.1093/nar/gkr062.

Hu, L. *et al.* (2012) ‘Crystallographic Analysis of Rotavirus NSP2-RNA Complex Reveals Specific Recognition of 5’ GG Sequence for RTPase Activity’, *Journal of Virology*, 86(19), pp. 10547–10557. doi: 10.1128/JVI.01201-12.

Hu, Liya *et al.* (2012) ‘Rotavirus non-structural proteins: Structure and function’, *Current Opinion in Virology*, 2(4), pp. 380–388. doi: 10.1016/j.coviro.2012.06.003.

Hunkeler, M. *et al.* (2018) ‘Structural basis for regulation of human acetyl-CoA carboxylase’, *Nature*. Springer US, 116, pp. 470–464. doi: 10.1073/pnas.1816417116.

Hutchinson, E. C. *et al.* (2010) ‘Genome packaging in influenza A virus’, *Journal of General Virology*, 91(2), pp. 313–328. doi: 10.1099/vir.0.017608-0.

Jaafar, Z. A. and Kieft, J. S. (2019) ‘Viral RNA structure-based strategies to manipulate translation’, *Nature Reviews Microbiology*. Springer US, pp. 110–123. doi: 10.1038/s41579-018-0117-x.

Jayaram, H. *et al.* (2002) ‘Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold.’, *Nature*, 417(6886), pp. 311–5. doi: 10.1038/417311a.

Jenkins, H. T., Malkova, B. and Edwards, T. A. (2011) ‘Kinked β-strands mediate high-affinity recognition of mRNA targets by the germ-cell regulator DAZL.’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(45), pp. 18266–71. doi: 10.1073/pnas.1105211108.

Jenni, S. *et al.* (2019) ‘In situ Structure of Rotavirus VP1 RNA-Dependent RNA Polymerase’, *Journal of Molecular Biology*. The Authors, 431(17), pp. 3124–3138. doi: 10.1016/j.jmb.2019.06.016.

Jiang, X., Jayaram, H., Kumar, M., Ludtke, S. J., Estes, M. K., Prasad, B. V. V., *et al.* (2006) ‘Cryoelectron Microscopy Structures of Rotavirus NSP2-NSP5 and NSP2-RNA Complexes: Implications for Genome Replication’, *Journal of Virology*. American Society for Microbiology, 80(21), pp. 10829–10835. doi: 10.1128/JVI.01347-06.

Jiang, X., Jayaram, H., Kumar, M., Ludtke, S. J., Estes, M. K. and Venkataram Prasad, B. V (2006) ‘Cryoelectron Microscopy Structures of Rotavirus NSP2-NSP5 and NSP2-RNA Complexes: Implications for

Genome Replication', *Journal of Virology*, 80(21), pp. 10829–10835. doi: 10.1128/JVI.01347-06.

Johnson, S. F. and Telesnitsky, A. (2010) 'Retroviral RNA Dimerization and Packaging : The What , How , When , Where , and Why', 6(10), pp. 10–13. doi: 10.1371/journal.ppat.1001007.

Jones, C. P. and Ferré-D'Amaré, A. R. (2015) 'Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains', *Nature Structural and Molecular Biology*, 22(9), pp. 679–685. doi: 10.1038/nsmb.3073.

Jourdan, N. *et al.* (1997) 'Rotavirus is released from the apical surface of cultured human intestinal cells through nonconventional vesicular transport that bypasses the Golgi apparatus.', *Journal of virology*, 71(11), pp. 8268–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9343179%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC192285.

Karpel, R. L., Miller, N. S. and Fresco, J. R. (1982) 'Mechanistic studies of ribonucleic acid renaturation by a helix-destabilizing protein', *Biochemistry*, 21(9), pp. 2102–2108. doi: 10.1021/bi00538a019.

Kavita, K., de Mets, F. and Gottesman, S. (2018) 'New aspects of RNA-based regulation by Hfq and its partner sRNAs', *Current Opinion in Microbiology*. Elsevier Ltd, pp. 53–61. doi: 10.1016/j.mib.2017.10.014.

Keane, S. C. *et al.* (2015) 'Structure of the HIV-1 packaging signal', *Science*, 348(6237), pp. 917–921.

Keane, S. C. *et al.* (2016) 'NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome.', *Proceedings of the National Academy of Sciences of the United States of America*, 113(46), pp. 13033–13038. doi: 10.1073/pnas.1614785113.

Kieft, J. S. (2008) 'Viral IRES RNA structures and ribosome interactions', *Trends in Biochemical Sciences*, 33(6), pp. 274–283. doi: 10.1016/j.tibs.2008.04.007.

Konarev, P. V. *et al.* (2006) 'ATSAS2.1, a program package for small-angle scattering data analysis', *Journal of Applied Crystallography*. International Union of Crystallography (IUCr), 39(2), pp. 277–286. doi: 10.1107/s0021889806004699.

Konarev, P. V *et al.* (2003) 'PRIMUS: a Windows PC-based system for small-angle scattering data analysis', *Journal of Applied Crystallography*. International Union of Crystallography (IUCr), 36(5), pp. 1277–1282. doi: 10.1107/s0021889803012779.

Kouba, T., Drncová, P. and Cusack, S. (2019) 'Structural snapshots of actively transcribing influenza polymerase', *Nature Structural and Molecular Biology*. Springer US, 26(6), pp. 460–470. doi: 10.1038/s41594-019-0232-z.

Kudryavtsev, V. *et al.* (2012) 'Combining MFD and PIE for accurate single-pair förster resonance energy transfer measurements', *ChemPhysChem*, 13(4), pp. 1060–1078. doi: 10.1002/cphc.201100822.

Kuhlbrandt, W. (2014) ‘The resolution revolution’, *Science*, 343(6178), pp. 1443–1444. doi: 10.1126/science.1251652.

Kumar, M. *et al.* (2007) ‘Crystallographic and biochemical analysis of rotavirus NSP2 with nucleotides reveals a nucleoside diphosphate kinase-like activity.’, *Journal of Virology*, 81(22), pp. 12272–84. doi: 10.1128/JVI.00984-07.

Lambert, L. *et al.* (1999) ‘The involvement of gRNA-binding protein gBP21 in RNA editing-an in vitro and in vivo analysis’, *Nucleic acids research*, 27(6), pp. 1429–1436. doi: gkc267 [pii].

Lapinaite, A. *et al.* (2013) ‘The structure of the box C/D enzyme reveals regulation of RNA methylation.’, *Nature*. Nature Publishing Group, 502(7472), pp. 519–23. doi: 10.1038/nature12581.

Larsen, K. P. *et al.* (2018) ‘Architecture of an HIV-1 reverse transcriptase initiation complex’, *Nature*. Springer US, p. 1. doi: 10.1038/s41586-018-0055-9.

Laurence, T. A. *et al.* (2007) ‘Correlation spectroscopy of minor fluorescent species: Signal purification and distribution analysis’, *Biophysical Journal*. Elsevier, 92(6), pp. 2184–2198. doi: 10.1529/biophysj.106.093591.

Leontis, N. B., Lescoute, A. and Westhof, E. (2006) ‘The building blocks and motifs of RNA architecture’, *Current Opinion in Structural Biology*, 16(3), pp. 279–287. doi: 10.1016/j.sbi.2006.05.009.

Li, W. *et al.* (2010) ‘Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting functions’, *Nucleic Acids Research*, 38(21), pp. 7718–7735. doi: 10.1093/nar/gkq663.

Li, X. *et al.* (2013) ‘Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.’, *Nature methods*. Nature Publishing Group, 10(6), pp. 584–90. doi: 10.1038/nmeth.2472.

Lindahl, T., Adams, A. and Fresco, J. R. (1966) ‘Renaturation of transfer ribonucleic acids through site binding of magnesium.’, *Proceedings of the National Academy of Sciences of the United States of America*, 55(4), pp. 941–948. doi: 10.1073/pnas.55.4.941.

Link, T. M., Valentin-Hansen, P. and Brennan, R. G. (2009) ‘Structure of Escherichia coli Hfq bound to polyriboadenylate RNA.’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(46), pp. 19292–7. doi: 10.1073/pnas.0908744106.

Lísal, J. *et al.* (2004) ‘Enzymatic Mechanism of RNA Translocation in Double-stranded RNA Bacteriophages’, *Journal of Biological Chemistry*, 279(2), pp. 1343–1350. doi: 10.1074/jbc.M309587200.

Lísal, J. *et al.* (2005) ‘Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states.’, *Nature structural &*

molecular biology, 12(5), pp. 460–466. doi: 10.1038/nsmb927.

Lourenco, S. and Roy, P. (2011) ‘In vitro reconstitution of Bluetongue virus infectious cores.’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), pp. 13746–51. doi: 10.1073/pnas.1108667108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1108667108.

Lowen, A. C. (2017) ‘Constraints, Drivers, and Implications of Influenza A Virus Reassortment.’, *Annual review of virology*, 4(1), pp. 105–121. doi: 10.1146/annurev-virology-101416-041726.

Lu, K., Heng, X. and Summers, M. F. (2011) ‘Structural determinants and mechanism of HIV-1 genome packaging’, *Journal of Molecular Biology*. Elsevier Ltd, pp. 609–633. doi: 10.1016/j.jmb.2011.04.029.

Lu, X. *et al.* (2008) ‘Mechanism for Coordinated RNA Packaging and Genome Replication by Rotavirus Polymerase VP1’, *Structure*. Elsevier Ltd, 16(11), pp. 1678–1688. doi: 10.1016/j.str.2008.09.006.

Lymperopoulos, K. *et al.* (2003) ‘Sequence specificity in the interaction of Bluetongue virus non-structural protein 2 (NS2) with viral RNA.’, *The Journal of biological chemistry*. in Press, 278(34), pp. 31722–31730. doi: 10.1074/jbc.M301072200.

Lyumkis, D. (2019) ‘Challenges and opportunities in cryo-EM single-particle analysis’, *Journal of Biological Chemistry*, 294(13), pp. 5181–5197. doi: 10.1074/jbc.REV118.005602.

Lyumkis, D. and Baldwin, P. (2019) ‘Non-Uniformity of Projection Distributions Attenuates Resolution in Cryo-EM’, *bioRxiv*, p. 635938. doi: 10.1101/635938.

MacFadden, A. *et al.* (2018) ‘Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs’, *Nature Communications*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41467-017-02604-y.

Mallam, A. L. *et al.* (2011) ‘Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail.’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(30), pp. 12254–9. doi: 10.1073/pnas.1109566108.

Manning, G. S. (1969) ‘Limiting laws and counterion condensation in polyelectroyte solutions. I. Colligative properties’, *Journal of Chemical Physics*, 51(3), pp. 924–933. doi: 10.1063/1.1672157.

Marklund, E. G., Degiacomi, Matteo T, *et al.* (2015) ‘Collision cross sections for structural proteomics’, *Structure*. Elsevier Ltd, 23(4), pp. 791–799. doi: 10.1016/j.str.2015.02.010.

Marklund, E. G., Degiacomi, Matteo T., *et al.* (2015) ‘Collision cross sections for structural proteomics’, *Structure*. Elsevier Ltd, 23(4), pp. 791–799. doi: 10.1016/j.str.2015.02.010.

Marques, M. A., Purdy, M. D. and Yeager, M. (2019) ‘CryoEM maps are full of potential', *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 1–10. doi: 10.1016/j.sbi.2019.04.006.

Martin, D. *et al.* (2010) 'Sequestration of free tubulin molecules by the viral protein NSP2 induces microtubule depolymerization during rotavirus infection.', *Journal of Virology*, 84(5), pp. 2522–2532. doi: 10.1128/JVI.01883-09.

Mayer, O. *et al.* (2007) 'RNA chaperone activity and RNA-binding properties of the E. coli protein StpA', *Nucleic Acids Research*, 35(4), pp. 1257–1269. doi: 10.1093/nar/gkl1143.

McCauley, M. J. *et al.* (2015) 'Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA', *Proceedings of the National Academy of Sciences*, 112(44), p. 201510100. doi: 10.1073/pnas.1510100112.

McDonald, S., Nelson, Martha I., *et al.* (2016) 'Reassortment in segmented RNA viruses: mechanisms and outcomes', *Nature Reviews Microbiology*. Nature Publishing Group, 14(May), pp. 448–460. doi: 10.1038/nrmicro.2016.46.

McDonald, S., Nelson, Martha I, *et al.* (2016) 'Reassortment in segmented RNA viruses: mechanisms and outcomes', *Nature Reviews Microbiology*. Springer Nature, 14(7), pp. 448–460. doi: 10.1038/nrmicro.2016.46.

McDonald, S. and Patton, J. T. (2011) 'Assortment and packaging of the segmented rotavirus genome', *Trends in Microbiology*, pp. 136–144. doi: 10.1016/j.tim.2010.12.002.

McMullan, G. *et al.* (2014) 'Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy', *Ultramicroscopy*. Elsevier, 147, pp. 156–163. doi: 10.1016/j.ultramic.2014.08.002.

Mikulecky, P. J. *et al.* (2004) 'Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs', *Nature Structural & Molecular Biology*, 11(12), pp. 1206–1214. doi: 10.1038/nsmb858.

Miles, S. *et al.* (2016) 'Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence-Dependence', *Angewandte Chemie - International Edition*, 128. doi: 10.1016/j.coviro.2014.01.003.

Miller, C. L. *et al.* (2003) 'Reovirus sigmaNS Protein Localizes to Inclusions through an Association Requiring the μNS Amino Terminus', *J virol*, 77(8), pp. 4566–4576. doi: 10.1128/JVI.77.8.4566.

Miller, C. L. *et al.* (2010) 'Localization of mammalian orthoreovirus proteins to cytoplasmic factory-like structures via nonoverlapping regions of μNS.', *Journal of Virology*. American Society for Microbiology, 84(2), pp. 867–882. doi: 10.1128/JVI.01571-09.

Modrof, J., Lymperopoulos, K. and Roy, P. (2005) 'Phosphorylation of Bluetongue Virus Nonstructural Protein 2 Is Essential for Formation of Viral Inclusion Bodies', *Journal of Virology*, 79(15), pp. 10023–10031. doi:

10.1128/JVI.79.15.10023–10031.2005.

Moeller, A. *et al.* (2012) ‘Organization of the influenza virus replication machinery.’, *Science (New York, N.Y.)*, 338(6114), pp. 1631–4. doi: 10.1126/science.1227270.

Mohl, B.-P. and Roy, P. (2016) ‘Cellular casein kinase 2 and protein phosphatase 2A modulate replication site assembly of Bluetongue virus’, *Journal of Biological Chemistry*, 291(28), p. jbc.M116.714766. doi: 10.1074/jbc.M116.714766.

Moller, T. *et al.* (2002) ‘Hfq: A Bacterial Sm-like protein that mediates RNA-RNA interaction’, *Molecular Cell*, 9(1), pp. 23–30. doi: 10.1016/S1097-2765(01)00436-1.

Moreira, É. A. *et al.* (2016) ‘A conserved influenza A virus nucleoprotein code controls specific viral genome packaging’, *Nature Communications*, 7, p. 12861. doi: 10.1038/ncomms12861.

Morgner, N. and Robinson, C. V. (2012) ‘Massign: An assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies’, *Analytical Chemistry*, 84(6), pp. 2939–2948. doi: 10.1021/ac300056a.

Müller, U. F. and Göringer, H. U. (2002) ‘Mechanism of the gBP21-mediated RNA / RNA annealing reaction : matchmaking and charge reduction’, *Nucleic acids research*, 30(2), pp. 447–455.

Mumtsidu, E. *et al.* (2007) ‘Structural features of the Bluetongue virus NS2 protein’, *Journal of Structural Biology*, 160(2), pp. 157–167. doi: 10.1016/j.jsb.2007.07.013.

Musalem, C. and Espejo, R. T. (1985) ‘Release of progeny virus from cells infected with simian rotavirus SA11’, *Journal of General Virology*, 66(12), pp. 2715–2724. doi: 10.1099/0022-1317-66-12-2715.

Mustoe, A. M. *et al.* (2018) ‘Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing’, *Cell*. Elsevier Inc., 173(1), pp. 181-195.e18. doi: 10.1016/j.cell.2018.02.034.

Mustoe, A. M., Brooks, C. L. and Al-Hashimi, H. M. (2014) ‘Hierarchy of RNA Functional Dynamics.’, *Annual review of biochemistry*, 83, pp. 441–466. doi: 10.1146/annurev-biochem-060713-035524.

Naydenova, K., Peet, M. J. and Russo, C. J. (2019) ‘Multifunctional graphene supports for electron cryomicroscopy’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), pp. 11718–11724. doi: 10.1073/pnas.1904766116.

Newburn, L. R. and White, K. A. (2019) ‘Trans-Acting RNA–RNA Interactions in Segmented RNA Viruses’, *Viruses*, 11(8), p. 751. doi: 10.3390/v11080751.

Nguyen, T. H. D. *et al.* (2015) ‘The architecture of the spliceosomal U4/U6.U5 tri-snRNP’, *Nature*, 523(7558), pp. 47–52. doi: 10.1038/nature14548.

Nikolaitchik, O. A. *et al.* (2013) ‘Dimeric RNA Recognition Regulates HIV-1 Genome Packaging’, *PLoS Pathogens*, 9(3). doi: 10.1371/journal.ppat.1003249.

Nir, E. *et al.* (2006) ‘Shot-Noise Limited Single-Molecule FRET HIstograms: Comparison between Theory and Experiments’, *Journal of Physical Chemistry B*, 110(44), pp. 22103–22124. doi: 10.3174/ajnr.A1256.Functional.

Noble, A. J., Wei, H., *et al.* (2018) ‘Reducing effects of particle adsorption to the air–water interface in cryo-EM’, *Nature Methods*. Springer US, 15(10), pp. 793–795. doi: 10.1038/s41592-018-0139-3.

Noble, A. J., Dandey, V. P., *et al.* (2018) ‘Routine single particle CryoEM sample and grid characterization by tomography’, *eLife*, 7, pp. 1–42. doi: 10.7554/eLife.34257.

Noda, T. *et al.* (2006) ‘Architecture of ribonucleoprotein complexes in influenza A virus particles’, *Nature*, 439(7075), pp. 490–492. doi: 10.1038/nature04378.

Noda, T. *et al.* (2012) ‘Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus’, *Nature Communications*. Nature Publishing Group, 3, p. 639. doi: 10.1038/ncomms1647.

Noda, T. *et al.* (2018) ‘Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging’, *Nature Communications*, 9(1), pp. 1–10. doi: 10.1038/s41467-017-02517-w.

Nogales, E. and Scheres, S. H. W. (2015) ‘Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity’, *Molecular Cell*. Elsevier Inc., 58(4), pp. 677–689. doi: 10.1016/j.molcel.2015.02.019.

Noller, H. F. (2005) ‘RNA structure: Reading the ribosome’, *Science*, pp. 1508–1514. doi: 10.1126/science.1111771.

Orlova, E. V and Saibil, H. R. (2011) ‘Structural Analysis of Macromolecular Assemblies by Electron Microscopy’, *Chemical Reviews*, 111, pp. 7710–7748.

Palacký, J., Mojzeš, P. and Bok, J. (2011) ‘SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations’, *Journal of Raman Spectroscopy*, 42(7), pp. 1528–1539. doi: 10.1002/jrs.2896.

Panja, S. *et al.* (2015) ‘Light-Triggered RNA Annealing by an RNA Chaperone’, *Angewandte Chemie - International Edition*, 54(25), pp. 7281–7284. doi: 10.1002/anie.201501658.

Panja, S. *et al.* (2017) ‘Metals induce transient folding and activation of the twister ribozyme’, *Nature Chemical Biology*. Nature Publishing Group, 13(10), pp. 1109–1114. doi: 10.1038/nchembio.2459.

Panja, S., Schu, D. J. and Woodson, S. A. (2013) ‘Conserved arginines on the rim of Hfq catalyze base pair formation and exchange’, *Nucleic Acids Research*, 41(15), pp. 7536–7546. doi: 10.1093/nar/gkt521.

Panja, S. and Woodson, S. A. (2012a) ‘Hexamer to monomer equilibrium of E. coli Hfq in solution and its impact on RNA annealing’, *Journal of Molecular Biology*. Elsevier Ltd, 417(5), pp. 406–412. doi: 10.1016/j.jmb.2012.02.009.

Panja, S. and Woodson, S. A. (2012b) ‘Hfq proximity and orientation controls RNA annealing’, *Nucleic Acids Research*, 40(17), pp. 8690–8697. doi: 10.1093/nar/gks618.

Papenfort, K. and Vanderpool, C. K. (2015) ‘Target activation by regulatory RNAs in bacteria’, *FEMS Microbiology Reviews*, 39(3), pp. 362–378. doi: 10.1093/femsre/fuv016.

Parker, J. S. L. *et al.* (2002) ‘Reovirus Core Protein μ2 Determines the Filamentous Morphology of Viral Inclusion Bodies by Interacting with and Stabilizing Microtubules’, *Journal of Virology*, 76(9), pp. 4483–4496. doi: 10.1128/JVI.76.9.4483–4496.2002.

Peng, Yi *et al.* (2014) ‘Structural model of an mRNA in complex with the bacterial chaperone Hfq’, *Proc Natl Acad Sci U S A*, 111(48), pp. 17134–9. doi: 10.1073/pnas.1410114111.

Periz, J. *et al.* (2013) ‘Rotavirus mRNAS are released by transcript-specific channels in the double-layered viral capsid.’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(29), pp. 12042–7. doi: 10.1073/pnas.1220345110.

Persson, F. *et al.* (2013) ‘Extracting intracellular diffusive states and transition rates from single-molecule tracking data’, *Nature Methods*, 10(3), pp. 265–269. doi: 10.1038/nmeth.2367.

Petoukhov, M. V. *et al.* (2012) ‘New developments in the ATSAS program package for small-angle scattering data analysis’, *Journal of Applied Crystallography*. International Union of Crystallography, 45(2), pp. 342–350. doi: 10.1107/S0021889812007662.

Pettersen, E. F. *et al.* (2004) ‘UCSF Chimera - A visualization system for exploratory research and analysis’, *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. doi: 10.1002/jcc.20084.

Purcell, S. *et al.* (2007) ‘PLINK: A tool set for whole-genome association and population-based linkage analyses’, *American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Rajkowitsch, L. *et al.* (2007) ‘RNA chaperones, RNA annealers and RNA helicases’, *RNA Biology*, pp. 118–130. doi: 10.4161/rna.4.3.5445.

Rajkowitsch, L. and Schroeder, R. (2007) ‘Dissecting RNA chaperone activity.’, *RNA (New York, N.Y.)*, 13(12), pp. 2053–2060. doi: 10.1261/rna.671807.

Rangan, P. *et al.* (2003) ‘Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme’, *Proc Natl Acad Sci U S A*, 100(4), pp. 1574–1579. doi: 10.1073/pnas.0337743100.

Record, M. T., Lohman, T. M. and Haseth, P. de (1976) ‘Ion effects on ligand-nucleic acid interactions’, *Journal of Molecular Biology*, 107(2), pp.

145–158. doi: 10.1016/S0022-2836(76)80023-X.

Reguera, J., Cusack, S. and Kolakofsky, D. (2014) ‘Segmented negative strand RNA virus nucleoprotein structure’, *Current Opinion in Virology*. Elsevier B.V., 5(1), pp. 7–15. doi: 10.1016/j.coviro.2014.01.003.

Rein, A. (2010) ‘Nucleic acid chaperone activity of retroviral Gag proteins’, *RNA Biology*, 7(6), pp. 700–705. doi: 10.4161/rna.7.6.13685.

Rein, A. (2019) ‘RNA Packaging in HIV’, *Trends in Microbiology*, 27(8), pp. 715–723. doi: 10.1016/j.tim.2019.04.003.

Rennella, E. *et al.* (2017) ‘RNA binding and chaperone activity of the E. coli cold-shock protein CspA’, *Nucleic Acids Research*, 45(7), pp. 4255–4268. doi: 10.1093/nar/gkx044.

Rosenthal, P. B. and Henderson, R. (2003) ‘Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy’, *Journal of Molecular Biology*, 333(4), pp. 721–745. doi: 10.1016/j.jmb.2003.07.013.

Ruotolo, B. T. *et al.* (2008) ‘Ion mobility–mass spectrometry analysis of large protein complexes’, *Nature Protocols*, 3(7), pp. 1139–1152. doi: 10.1038/nprot.2008.78.

Ruotolo, B. T. and Robinson, C. V. (2006) ‘Aspects of native proteins are retained in vacuum’, *Current Opinion in Chemical Biology*, pp. 402–408.

Ruskin, R. S., Yu, Z. and Grigorieff, N. (2013) ‘Quantitative characterization of electron detectors for transmission electron microscopy’, *Journal of Structural Biology*. Elsevier Inc., 184(3), pp. 385–393. doi: 10.1016/j.jsb.2013.10.016.

Russell, R. *et al.* (2000) ‘Small angle X-ray scattering reveals a compact intermediate in RNA folding’, *Nature Structural Biology*, 7(5), pp. 367–370. doi: 10.1038/75132.

Russell, R., Zhuang, X., *et al.* (2002) ‘Exploring the folding landscape of a structured RNA.’, *Proceedings of the National Academy of Sciences of the United States of America*, 99(1), pp. 155–60. doi: 10.1073/pnas.221593598.

Russell, R., Millett, I. S., *et al.* (2002) ‘Rapid compaction during RNA folding.’, *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), pp. 4266–4271. doi: 10.1073/pnas.072589599.

Russell, R. (2008) ‘RNA misfolding and the action of chaperones.’, *Frontiers in bioscience : a journal and virtual library*, 13, pp. 1–20. doi: 10.2741/2557.

Russell, R., Jarmoskaite, I. and Lambowitz, A. M. (2013) ‘Toward a molecular understanding of RNA remodeling by DEAD-box proteins’, *RNA Biology*, pp. 44–55. doi: 10.4161/rna.22210.

Russo, C. J. and Passmore, L. A. (2014) ‘Ultrastable gold substrates for electron cryomicroscopy’, *Science*, 346(6215), pp. 1377–1381.

Russo, C. J. and Passmore, L. A. (2016) ‘Progress towards an optimal specimen support for electron cryomicroscopy’, *Current Opinion in Structural*

Biology. Elsevier Ltd, 37, pp. 81–89. doi: 10.1016/j.sbi.2015.12.007.

Saathoff, J. *et al.* (2019) ‘Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex’, *Molecular cell*, pp. 1–13. doi: 10.1016/j.molcel.2019.07.035.

Sakuragi, J. -i., Shioda, T. and Panganiban, A. T. (2002) ‘Duplication of the Primary Encapsidation and Dimer Linkage Region of Human Immunodeficiency Virus Type 1 RNA Results in the Appearance of Monomeric RNA in Virions’, *Journal of Virology*, 75(6), pp. 2557–2565. doi: 10.1128/jvi.75.6.2557-2565.2001.

Salgado, E. N., Upadhyayula, S. and Harrison, S. C. (2017) ‘Single-Particle Detection of Transcription following Rotavirus Entry’, *Journal of Virology*, 91(18), pp. 1–18. doi: 10.1128/jvi.00651-17.

Samatanga, B. *et al.* (2017) ‘Comparative analyses of the thermodynamic RNA binding signatures of different types of RNA recognition motifs’, *Nucleic Acids Research*, 45(10), pp. 6037–6050. doi: 10.1093/nar/gkx136.

Sandate, C. R. *et al.* (2019) ‘An allosteric network in spastin couples multiple activities required for microtubule severing’, *Nature Structural and Molecular Biology*. Springer US, 26(August). doi: 10.1038/s41594-019-0257-3.

Santiago-Frangos, A. *et al.* (2016) ‘C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA’, *Proceedings of the National Academy of Sciences*, 113(41), pp. E6089–E6096. doi: 10.1073/pnas.1613053113.

Santiago-Frangos, A. *et al.* (2017) ‘Acidic C-terminal domains autoregulate the RNA chaperone Hfq’, *eLife*, 6, pp. 1–25. doi: 10.7554/eLife.27049.

Santiago-Frangos, A. *et al.* (2019) ‘Caulobacter crescentus Hfq structure reveals a conserved mechanism of RNA annealing regulation.’, *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1–10. doi: 10.1073/pnas.1814428116.

Santiago-Frangos, A. and Woodson, S. A. (2018) ‘Hfq chaperone brings speed dating to bacterial sRNA’, *Wiley Interdisciplinary Reviews: RNA*, 9(4), pp. 1–16. doi: 10.1002/wrna.1475.

Sauer, E. and Weichenrieder, O. (2011) ‘Structural basis for RNA 3’-end recognition by Hfq.’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(32), pp. 13065–70. doi: 10.1073/pnas.1103420108.

Scheres, S. H. W. (2012) ‘RELION: Implementation of a Bayesian approach to cryo-EM structure determination’, *Journal of Structural Biology*. Elsevier Inc., 180(3), pp. 519–530. doi: 10.1016/j.jsb.2012.09.006.

Scheres, S. H. W. (2016) *Processing of Structurally Heterogeneous Cryo-EM Data in RELION*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc. doi: 10.1016/bs.mie.2016.04.012.

Scheres, S. H. W. and Chen, S. (2012) ‘Prevention of overfitting in cryo-EM structure determination’, *Nature Methods*. Nature Publishing Group, 9(9), pp.

853–854. doi: 10.1038/nmeth.2115.

Schlundt, A., Tants, J.-N. and Sattler, M. (2017) ‘Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition’, *Methods*. Elsevier Inc., 118–119, pp. 119–136. doi: 10.1016/j.ymeth.2017.03.015.

Schrimpf, Waldemar (2018) ‘PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data’, *Biophysical Journal*.

Schuck, P. *et al.* (2001) ‘Rotavirus Nonstructural Protein NSP2 Self-assembles into Octamers that Undergo Ligand-induced Conformational Changes’, *Journal of Biological Chemistry*, 276(13), pp. 9679–9687. doi: 10.1074/jbc.M009398200.

Schulze-Gahmen, U. *et al.* (2016) ‘Insights into HIV-1 proviral transcription from integrative structure and dynamics of the tat:AFF4:P-TEFB:TAR complex’, *eLife*, 5(OCTOBER2016), pp. 1–21. doi: 10.7554/eLife.15910.

Schumacher, M. A. *et al.* (2002) ‘Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein’, *EMBO Journal*, 21(13), pp. 3546–3556. doi: 10.1093/emboj/cdf322.

Schumacher, M. A. *et al.* (2006) ‘Crystal Structures of T. brucei MRP1/MRP2 Guide-RNA Binding Complex Reveal RNA Matchmaking Mechanism’, *Cell*, 126(4), pp. 701–711. doi: 10.1016/j.cell.2006.06.047.

Schur, F. K. M. *et al.* (2016) ‘An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation’, 9620(July), pp. 1–8.

Serganov, A. and Patel, D. J. (2012) ‘Molecular recognition and function of riboswitches’, *Current Opinion in Structural Biology*. Elsevier Ltd, 22(3), pp. 279–286. doi: 10.1016/j.sbi.2012.04.005.

Shah, P. N. M. *et al.* (2017) ‘Genome packaging of reovirus is mediated by the scaffolding property of the microtubule network’, *Cellular Microbiology*, 19(12). doi: 10.1111/cmi.12765.

Shkumatov, A. V and Strelkov, S. V (2015) ‘DATASW, a tool for HPLC-SAXS data analysis’, *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 71, pp. 1347–1350. doi: 10.1107/S1399004715007154.

Sim, A. Y. L., Minary, P. and Levitt, M. (2012) ‘Modeling nucleic acids’, *Current Opinion in Structural Biology*. Elsevier Ltd, 22(3), pp. 273–278. doi: 10.1016/j.sbi.2012.03.012.

Smith, D. *et al.* (2009) ‘Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies’, *European Journal of Mass Spectrometry*, 15(5), p. 113. doi: 10.1255/ejms.947.

Somarowthu, S. *et al.* (2015) ‘HOTAIR Forms an Intricate and Modular Secondary Structure’, *Molecular Cell*, 58(2), pp. 353–361. doi: 10.1016/j.molcel.2015.03.006.

Spitale, R. C. *et al.* (2015) ‘Structural imprints in vivo decode RNA regulatory mechanisms’, *Nature*. Nature Publishing Group, 519(7544), pp. 486–90. doi:

10.1038/nature14263.

Stanek, K. A. *et al.* (2017) ‘Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: Conservation of the lateral RNA-binding mode’, *Acta Crystallographica Section D: Structural Biology*. International Union of Crystallography, 73(4), pp. 294–315. doi: 10.1107/S2059798317000031.

Steckelberg, A.-L. *et al.* (2018) ‘A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure’, *Proceedings of the National Academy of Sciences*, 115(25), pp. 6404–6409. doi: 10.1073/pnas.1802429115.

Sung, P.-Y. Y. and Roy, P. (2014) ‘Sequential packaging of RNA genomic segments during the assembly of bluetongue virus’, *Nucleic Acids Research*. Oxford University Press (OUP), 42(22), pp. 13824–13838. doi: 10.1093/nar/gku1171.

Suslov, N. B. *et al.* (2015) ‘Crystal structure of the Varkud satellite ribozyme’, *Nature Chemical Biology*. Nature Publishing Group, 11(11), pp. 840–848. doi: 10.1038/nchembio.1929.

Svergun, D. I. (1999) ‘Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing’, *Biophysical Journal*. Elsevier, 76(6), pp. 2879–2886. doi: 10.1016/S0006-3495(99)77443-6.

Svergun, D. I., Petoukhov, M. V and Koch, M. H. (2001) ‘Determination of domain structure of proteins from X-ray solution scattering.’, *Biophysical journal*, 80(6), pp. 2946–2953. doi: 10.1016/S0006-3495(01)76260-1.

Taraporewala, Z. F. *et al.* (2006) ‘Structure-function analysis of rotavirus NSP2 octamer by using a novel complementation system.’, *Journal of Virology*, 80(16), pp. 7984–7994. doi: 10.1128/JVI.00172-06.

Taraporewala, Z. F., Chen, D. and Patton, J. T. (1999) ‘Multimers Formed by the Rotavirus Nonstructural Protein NSP2 Bind to RNA and Have Nucleoside Triphosphatase Activity’, *Journal of Virology*, 73(12), pp. 9934–9943.

Taraporewala, Z. F., Chen, D. and Patton, J. T. (2001) ‘Multimers of the bluetongue virus nonstructural protein, NS2, possess nucleotidyl phosphatase activity: similarities between NS2 and rotavirus NSP2.’, *Virology*, 280(2), pp. 221–31. doi: 10.1006/viro.2000.0764.

Taraporewala, Z. F. and Patton, J. T. (2001) ‘Identification and Characterization of the Helix-Destabilizing Activity of Rotavirus Nonstructural Protein NSP2’, *Journal of Virology*, 75(10), pp. 4519–4527. doi: 10.1128/JVI.75.10.4519–4527.2001.

Taraporewala, Z. F. and Patton, J. T. (2004) ‘Nonstructural proteins involved in genome packaging and replication of rotaviruses and other members of the Reoviridae’, *Virus Research*, pp. 57–66. doi: 10.1016/j.virusres.2003.12.006.

Thirumalai, D. and Woodson, S. A. (1996) ‘Kinetics of Folding of Proteins and RNA’, *Accounts of Chemical Research*, 29(9), pp. 433–439. doi: 10.1021/ar9500933.

Tijerina, P., Bhaskaran, H. and Russell, R. (2006) ‘Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone.’, *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), pp. 16698–16703. doi: 10.1073/pnas.0603127103.

Tomov, T. E. *et al.* (2012) ‘Disentangling subpopulations in single-molecule FRET and ALEX experiments with photon distribution analysis’, *Biophysical Journal*. Biophysical Society, 102(5), pp. 1163–1173. doi: 10.1016/j.bpj.2011.11.4025.

Toropova, K., Mladenov, M. and Roberts, A. J. (2017) ‘Intraflagellar transport dynein is autoinhibited by trapping of its mechanical and track-binding elements’, *Nature Structural and Molecular Biology*. Nature Publishing Group, 24(5), pp. 461–468. doi: 10.1038/nsmb.3391.

Tortorici, M. A. *et al.* (2003) ‘Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus’, *Journal of Biological Chemistry*, 278(35), pp. 32673–32682. doi: 10.1074/jbc.M305358200.

Tortorici, M. A., Shapiro, B. A. and Patton, J. T. (2006) ‘A base-specific recognition signal in the 5’ consensus sequence of rotavirus plus-strand RNAs promotes replication of the double-stranded RNA genome segments.’, *RNA (New York, N.Y.)*, 12, pp. 133–146. doi: 10.1261/rna.2122606.

Tourís-Otero, F., Cortez-San Martín, M., *et al.* (2004) ‘Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of σNS and λA to *μ*NS inclusions’, *Journal of Molecular Biology*. Elsevier BV, 341(2), pp. 361–374. doi: 10.1016/j.jmb.2004.06.026.

Tourís-Otero, F., Martínez-Costas, J., *et al.* (2004) ‘Avian reovirus nonstructural protein μNS forms viroplasm-like inclusions and recruits protein σNS to these structures’, *Virology*. Elsevier BV, 319(1), pp. 94–106. doi: 10.1016/j.virol.2003.10.034.

Tourís-Otero, F. *et al.* (2005) ‘Characterization of the nucleic acid-binding activity of the avian reovirus non-structural protein σNS’, *Journal of General Virology*, 86(4), pp. 1159–1169. doi: 10.1099/vir.0.80491-0.

Trachman, R. J. and Draper, D. E. (2017) ‘Divalent ion competition reveals reorganization of an RNA ion atmosphere upon folding’, *Nucleic Acids Research*, 45(8), pp. 4733–4742. doi: 10.1093/nar/gkw1327.

Trask, S. D., McDonald, S. and Patton, J. T. (2012) ‘Structural insights into the coupling of virion assembly and rotavirus replication’, *Nature Reviews Microbiology*, pp. 165–177. doi: 10.1038/nrmicro2673.

Tuukkanen, A. T., Kleywegt, G. J. and Svergun, D. I. (2016) ‘Resolution of ab initio shapes determined from small-angle scattering’, *IUCrJ*. International Union of Crystallography, 3, pp. 440–447. doi:

10.1107/S2052252516016018.

Updegrove, T. B., Zhang, A. and Storz, G. (2016) 'Hfq: The flexible RNA matchmaker', *Current Opinion in Microbiology*. Elsevier Ltd, 30, pp. 133–138. doi: 10.1016/j.mib.2016.02.003.

Usov, I. and Mezzenga, R. (2015) 'FiberApp: An open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects', *Macromolecules*, 48(5), pp. 1269–1280. doi: 10.1021/ma502264c.

Vakiloroayaei, A. *et al.* (2017) 'The RNA chaperone La promotes pre-TRNA maturation via indiscriminate binding of both native and misfolded targets', *Nucleic Acids Research*. Oxford University Press, 45(19), pp. 11341–11355. doi: 10.1093/nar/gkx764.

Vascotto, F. *et al.* (2004) 'Effects of intrabodies specific for rotavirus NSP5 during the virus replicative cycle', *Journal of General Virology*, 85(11), pp. 3285–3290. doi: 10.1099/vir.0.80075-0.

Vasquez-Del Carpio, R. *et al.* (2004) 'Role of the Histidine Triad-like Motif in Nucleotide Hydrolysis by the Rotavirus RNA-packaging Protein NSP2', *Journal of Biological Chemistry*, 279(11), pp. 10624–10633. doi: 10.1074/jbc.M311563200.

Vasquez-Del Carpio, R. *et al.* (2006a) 'Histidine Triad-like Motif of the Rotavirus NSP2 Octamer Mediates both RTPase and NTPase Activities', *Journal of Molecular Biology*, 362(3), pp. 539–554. doi: 10.1016/j.jmb.2006.07.050.

Vasquez-Del Carpio, R. *et al.* (2006b) 'Histidine Triad-like Motif of the Rotavirus NSP2 Octamer Mediates both RTPase and NTPase Activities', *Journal of Molecular Biology*, 362(3), pp. 539–554. doi: 10.1016/j.jmb.2006.07.050.

Vilas, J. L. *et al.* (2018) 'Advances in image processing for single-particle analysis by electron cryomicroscopy and challenges ahead', *Current Opinion in Structural Biology*, 52, pp. 127–145. doi: 10.1016/j.sbi.2018.11.004.

Viskovska, M. *et al.* (2014) 'Probing the sites of interactions of rotaviral proteins involved in replication.', *Journal of Virology*, 88(21), pp. 12866–81. doi: 10.1128/JVI.02251-14.

Vogel, J. and Luisi, B. F. (2011) 'Hfq and its constellation of RNA', *Nature Reviews Microbiology*, pp. 578–589. doi: 10.1038/nrmicro2615.

Voith von Voithenberg, L. *et al.* (2016) 'Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift', *Proceedings of the National Academy of Sciences*, 113(46), pp. E7169–E7175. doi: 10.1073/pnas.1605873113.

Volkov, V. V and Svergun, D. I. (2003) 'Uniqueness of ab initio shape determination in small-angle scattering', in *Journal of Applied Crystallography*. International Union of Crystallography, pp. 860–864. doi: 10.1107/S0021889803000268.

Wang, W. *et al.* (2013) 'Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA’, *Nucleic Acids Research*, 41(11), pp. 5938–5948. doi: 10.1093/nar/gkt276.

Weeks, K. M. (1997) ‘Protein-facilitated RNA folding’, *Current Opinion in Structural Biology*, pp. 336–342.

Woodson, S. A. (2010) ‘Taming free energy landscapes with RNA chaperones.’, *RNA biology*, 7(6), pp. 677–686. doi: 10.4161/rna.7.6.13615.

Woodson, S. A., Panja, S. and Santiago-Frangos, A. (2018) ‘Proteins That Chaperone RNA Regulation’, *Microbiology Spectrum*, 113(41), pp. E6089–E6096. doi: 10.1073/pnas.1613053113.

Wu, J. *et al.* (2013) ‘Viroplasm protein P9-1 of Rice black-streaked dwarf virus preferentially binds to single-stranded RNA in its octamer form, and the central interior structure formed by this octamer constitutes the major RNA binding site.’, *Journal of virology*, 87(23), pp. 12885–99. doi: 10.1128/JVI.02264-13.

Wu, M. and Tinoco, I. (1998) ‘RNA folding causes secondary structure rearrangement’, *Proceedings of the National Academy of Sciences*, 95(20), pp. 11555–11560. doi: 10.1073/pnas.95.20.11555.

Xia, H. *et al.* (2015) ‘Human Enterovirus Nonstructural Protein $2C^{ATPase}$ Functions as Both an RNA Helicase and ATP-Independent RNA Chaperone’, *PLoS Pathogens*, 11(7), pp. 1–29. doi: 10.1371/journal.ppat.1005067.

Xia, T. (2008) ‘Taking femtosecond snapshots of RNA conformational dynamics and complexity’, *Current Opinion in Chemical Biology*, 12(6), pp. 604–611. doi: 10.1016/j.cbpa.2008.08.033.

Yang, J. *et al.* (2014) ‘A cypovirus VP5 displays the RNA chaperone-like activity that destabilizes RNA helices and accelerates strand annealing’, *Nucleic Acids Research*, 42(4), pp. 2538–2554. doi: 10.1093/nar/gkt1256.

Yang, Q. *et al.* (2007) ‘DEAD-Box Proteins Unwind Duplexes by Local Strand Separation’, *Molecular Cell*, 28(2), pp. 253–263. doi: 10.1016/j.molcel.2007.08.016.

York, A. *et al.* (2013) ‘Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus’, *Proceedings of the National Academy of Sciences*, 110(45), pp. E4238–E4245. doi: 10.1073/pnas.1315068110.

Zamora, P. F. *et al.* (2018) ‘Reovirus Nonstructural Protein σNS Acts as an RNA-Stability Factor Promoting Viral Genome Replication.’, *Journal of Virology*, 92(15), p. JVI.00563-18. doi: 10.1128/JVI.00563-18.

Zhang, K. (2016) ‘Gctf: Real-time CTF determination and correction’, *Journal of Structural Biology*. Elsevier Inc., 193(1), pp. 1–12. doi: 10.1016/j.jsb.2015.11.003.

Zhang, Q. *et al.* (2019) ‘RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2’, *Nature Structural and Molecular Biology*. Springer US, 26(3), pp. 237–247. doi: 10.1038/s41594-019-0197-y.

Zhang, X. *et al.* (2015) ‘In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus’, *Nature*. Nature Publishing Group, 527(7579), pp. 531–534. doi: 10.1038/nature15767.

Zi Tan, Y. *et al.* (2017) ‘Addressing preferred specimen orientation in single-particle cryo-EMthrough tilting’, *Nature Methods*, 14(8), pp. 793–796. doi: 10.1038/nmeth.4347.

Zivanov, J. *et al.* (2018) ‘RELION-3: new tools for automated high-resolution cryo-EM structure determination’, *eLife*, p. 421123. doi: 10.1101/421123.

Zúñiga, S. *et al.* (2009) ‘Role of RNA chaperones in virus replication’, *Virus Research*, 139(2), pp. 253–266. doi: 10.1016/j.virusres.2008.06.015.