Structure of the RNA-Dependent RNA Polymerase from Influenza C Virus

Narin Hengrung

Medical Sciences Doctoral Training Centre
Christ Church
University of Oxford

A thesis submitted in fulfillment of the requirements for the degree of *Doctor of Philosophy* at the University of Oxford

Trinity Term 2014

Structure of the RNA-Dependent RNA Polymerase from Influenza C Virus

Narin Hengrung, Christ Church, University of Oxford
DPhil, Medical Sciences Doctoral Training Centre
Trinity Term 2014

Abstract

The influenza virus causes a disease that kills approximately 500,000 people worldwide each year. Influenza is a negative-sense RNA virus that encodes its own RNA-dependent RNA polymerase. This protein (FluPol) carries out both genome replication and viral transcription. Therefore, like the L-proteins of non-segmented negative-sense RNA (nsRNA) viruses, FluPol also contains mRNA capping and polyadenylation functionality. In FluPol, capping is achieved by snatching cap structures from cellular mRNAs, so requiring cap-binding and endonuclease activities. This makes FluPol a substantial machine. It is a heterotrimeric complex, composed of PB1, PB2 and PA/P3 subunits, with a total molecular weight of 255 kDa. PB1 houses the polymerase active site, whereas PB2 and PA contain, respectively, cap-binding and endonuclease domains. Currently, we only have high resolution structural information for isolated fragments of FluPol. This severely hampers our understanding of influenza replication and consequently inhibits the development of therapies against the virus.

In this *DPhil* project, I have determined a preliminary structure for the heterotrimeric FluPol of influenza C/Johannesburg/1/66, solved by x-ray crystallography to 3.6 Å. Overall, FluPol has an elongated structure with a conspicuous deep groove. PB1 displays the canonical right-hand-like polymerase fold. It sits at the centre of the particle, sandwiched between the two domains of P3, and with PB2 stacked against one side of this dimer. In the structure, the polymerase and endonuclease catalytic sites are both ~40 Å away from the cap-binding pocket. This pocket also faces a tunnel leading to the polymerase core. This suggests a mechanism for how capped cellular mRNAs are cleaved and then fed into the polymerase active site to prime transcription. The structure also hints at a unique trajectory for template RNA, in which the RNA exits at an angle ~180º from which it came in. This provides an explanation for how the polymerases of influenza, and other nsRNA viruses, can copy templates that are packaged into ribonucleoprotein complexes.

My work reveals the first molecular structure of any polymerase from an nsRNA virus. It uncovers the arrangement of functional domains within FluPol, illuminating the mechanisms of this and related viral polymerases. This work will help focus future experiments into FluPol biology, and should hopefully spur the development of novel antiviral drugs.

Acknowledgements

This work was only possible because of the efforts of a great number of people. I am immensely thankful to my supervisors: Dr Jonathan Grimes and Professor Ervin Fodor, for allowing me to take on such a challenging and rewarding project, and providing me with the guidance necessary to see it through. They have been generous with their time and have always answered my questions with patience. Their enthusiasm for science has been an inspiration. Huge thanks must also go to Professor Dave Stuart and Professor George Brownlee, who jointly started the influenza polymerase project many years ago, and continue to have major influence on its direction; providing a wealth of useful advice and constant encouragement. I thank Dr Frank Vreede for helping to design the polymerase expression system and for supervising me at the bench, when I first started work on the project. Dr Imre Berger developed the MultiBac system and kindly donated this to us. More specific thanks go to Dr Karl Harlos and Dr Tom Walter, who were very helpful on the crystallisation front, and Dr Kamel El Omari, who's input into the later stages of the project was invaluable. Additionally, I thank the Wellcome Trust, for the generous funding that I have received during my DPhil.

The project required a fairly large amount of synchrotron beamtime, so I especially want to offer thanks to the staff of the MX beamlines at the Diamond Light Source for all of their assistance and for keeping such a wonderful facility running on a day-to-day basis.

It has been a privilege to have spent the last four years in Oxford. I have learnt a great deal and have had a lot of fun from working in such an open and collaborative community, right at the forefront of science. I want to thank everyone in the Fodor Lab and at StruBi, for contributing to this environment. It has also been a pleasure to have been part of the Graduate Common Room at Christ Church. I immensely value the friends that I have made here and elsewhere. I would be lost without their friendship and support. A special thank you to my partner Alys, whose love has brought a happiness and contentment into my life that I have not known before.

Lastly, I want to offer my deepest thanks to my parents: Phongsai and Praphan, for everything they have given me.

Declaration of Work

All of the described herein is my own, except that detailed immediately below:

Grid preparation for electron microscopy and micrograph collection (**section 2.2.8**) was performed with the help of Juha Huiskonen.

The production of nanobodies, against $FluPol_{A/Fj}$ (**section 3.1**), was done in collaboration with the lab of Jan Steyaert at the VIB research institute in Brussels. Els Pardon and colleagues, from this lab, carried out llama immunisations, generation of a nanobody library and nanobody selection.

Crystal freezing (**section 3.2)** was performed by Jonathan Grimes, Karl Harlos and Kamel El Omari.

X-ray data collection (**section 3.2**) was carried out mainly by myself, with the supervision of Jonathan Grimes and Kamel El Omari. However, a few datasets, including that from the vdx_D3 crystal, were collected in my absence by the above two individuals.

Initial x-ray data analysis (data reduction and phasing) and model building (**section 3.2**) was performed jointly by myself and Kamel El Omari. However the latter stages of model building, and model refinement, was performed exclusively by Kamel El Omari.

Several sections of **chapter 4**, both text and figures, are based on a manuscript that has been submitted for review. This was co-written by myself, Ervin Fodor, Jonathan Grimes and Dave Stuart.

Table of Contents

List of Figures

List of Tables

List of Non-Standard Abbreviations

A/Fj	Influenza A/Northern Territory/60/1968 (H3N2)
A/NT	Influenza A (duck)/Fujian/1/2002 (H5N1)
CBP	Calmodulin Binding Peptide
cRNA / RNP	Complementary RNA / ribonucleoprotein
DPBA	2,4-dioxo-4-phenylbutanoic acid
EM	Electron Microscopy
$FluPol_{A/B/C}$	Influenza A/B/C virus RNA-dependent RNA polymerase
FRET	Förster Resonance Energy Transfer
HA	Haemaglutinin
HEF	Haemaglutinin-Esterase-Fusion
I(A/B/C)V	Influenza (A/B/C) Virus
ISAV	Infectious Salmon Anaemia Virus
MALS	Multi-Angle Light Scattering
NA	Neuraminidase
NCS	Non-Crystallographic Symmetry
NDSB-201	Non-Detergent Sulfobetaine 201
NEP	Nuclear Export Protein
NLS	Nuclear Localisation Sequence
NP	Nucleoprotein
nsRNA	Negative-sense RNA
NTP	Nucleotide Triphosphate
OTG	Octyl-β-Thioglucoside
P3	Polymerase Protein 3
PA	Polymerase Acidic Protein
PA_C / $P3_C$	The C-terminal domain of PA/P3.
PA_{Endo} / $P3_{Endo}$	The N-terminal, Endonuclease domain of PA/P3.
PB1	Polymerase Basic Protein 1
PB2	Polymerase Basic Protein 2
$PB2_{Cap}$	The Cap-binding domain of PB2.
$PB2_{NLS}$	The NLS domain of PB2
poly-U	poly-Uridine
PPi	Pyrophosphate
protA	Protein-A
RdRp	RNA-dependent RNA polymerase
RNP	Ribonucleoprotein
RMSD	Root-Mean-Square Deviation
RT	Room Temperature
SAD	Single-wavelength Anomalous Dispersion

S-SAD	Sulphur-SAD
SeMet	Selenomethionine
SFM	Serum-Free Medium
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
svRNA	Small viral RNA
TAP	Tandem Affinity Purification
TEV	Tobacco Etch Virus
vRNA/vRNP	Viral RNA / ribonucleoprotein

Introduction

1.1 Influenza

Influenza (flu) in humans is a virally transmitted acute respiratory disease. Most people will have personally experienced its symptoms (a blocked nose, high fever, muscle pains and so forth), but its very familiarity means that the true severity of flu can be easy to overlook. Seasonal flu epidemics occur annually or biannually in most countries (Azziz Baumgartner *et al*, 2012), causing a considerable disease burden, particularly amongst infants and the elderly (Simonsen, 1999; Molinari *et al*, 2007). For the United States alone, estimates of between roughly 20,000 and 60,000 total deaths per year have been reported (Thompson *et al*, 2003; Foppa & Hossain, 2008; Thompson *et al*, 2009), whilst the annual number of worldwide deaths is estimated to be ~500,000 (Lozano *et al*, 2012). Within these mean figures, a large amount of annual variation is also evident; particular epidemics are several times more severe than others.

On top of this not insignificant seasonal burden, pandemic influenza also occasionally arises (Taubenberger & Morens, 2010). Examples include the 1957 Asian flu (over one million deaths (Potter, 2001)) and the recent 2009 swine flu (approximately 300,000 deaths (Dawood *et al*, 2012)). The exceptionally savage 1918 Spanish flu, in which an estimated 50 – 100 million people died (Johnson & Mueller, 2002), is a sobering reminder of the threat from pandemic flu.

Furthermore, it appears that this pattern of influenza occurrence – seasonal epidemics punctuated by occasional pandemics – has been repeating for many centuries at

least. Evidence exists for epidemics having occurred since 1173 and there is a possible case mentioned by Hippocrates as early as 412 B.C.(Kuszewski & Brydak, 2000; Potter, 2001). The first evidence for a pandemic appears in 1510 (Morens *et al*, 2010b; 2010a) and 11 – 14 pandemics have occurred since, roughly one every 40 years (Potter, 2001; Taubenberger & Morens, 2010). This ongoing association of flu with humans is partly due to the ability of the virus to constantly mutate, sometimes suddenly, through the pathways of antigenic drift and antigenic shift. Influenza viruses are also able to infect a variety of birds and mammals (Morens & Taubenberger, 2010), so maintain large, long-running animal reservoirs. This further complicates viral control.

From the above discussion, it is clear that influenza viruses have and continue to, impose an enormous burden on human health. It should not therefore be a great surprise that they have been at the centre of an extensive research effort ever since Smith, Andrews and Laidlaw first isolated a human influenza virus in the 1930s (Smith *et al*, 1933). This pioneering work, which demonstrated serial passage of human flu in ferrets, along with that of Richard Shope (Shope, 1931a; Lewis & Shope, 1931; Shope, 1931b), Alfonse Dochez (Dochez *et al*, 1933) and their colleagues, helped overturn the prevailing wisdom of the time that flu was caused by a bacterium. Since the 1930s, we have learnt a great deal more about influenza viruses; their structure and composition, their lifecycle, how they interact with the cell and how they cause disease. Our increased understanding has revealed that even this relatively simple biological entity is still fascinatingly complex. As with many things, the more we learn, the more we realise how much we still have to learn.

The subject of my thesis is the RNA-dependent RNA polymerase (RdRp) of the influenza virus. The focus is on solving the three-dimensional molecular structure of this protein and what this teaches us about how the virus replicates. In this introductory

section, I will attempt to set this effort in its appropriate context. The key topics I will address are:

1. The general biology of influenza viruses.
2. Our current knowledge about polymerase structure.
3. Unanswered questions about polymerase mechanism.

1.2 General Biology of the Influenza Virus

1.2.1 Taxonomy and Phylogenetics

Influenza viruses are classified into the *Orthomyxoviridae* family (King *et al*, 2011), based on shared characteristics such as an envelope and a segmented, single-stranded, negative-sense RNA (nsRNA) genome. *Orthomyxoviridae* have not been assigned an order, but it is logical to group them with other segmented nsRNA viruses (Baltimore, 1971). Hence, virus families that are similar to *Orthomyxoviridae* include the *Arenaviridae* (Machupo virus) and *Bunyaviridae* (Hantaan virus) (Ecker *et al*, 2005). On the other side, non-segmented, linear, nsRNA viruses are assigned the order *Mononegavirales*. Amongst this order are the *Filoviridae* (Ebola virus) and the *Rhabdoviridae* (Rabies virus) (King *et al*, 2011). Notably, the only common feature between all RNA viruses is the RNA-dependent polymerase, however, no atomic-resolution structure has yet been reported for the complete polymerase of any nsRNA virus.

Within the *Orthomyxoviridae*, there are three distinct genera of influenza viruses: *influenzavirus A*, *B* and *C* (each made up of one species), as well as three additional genera: *Thogotovirus*, *Quaranjavirus* and *Isavirus* (http://www.ictvonline.org/virusTaxonomy.asp). Influenza A and B viruses (IAV and IBV) are broadly similar, sharing the same genome organisation but slightly different protein compositions. However, IAVs cause the majority of human disease and have a much broader host range. In addition to humans, IAV is known to infect waterfowl, poultry and a variety of land and sea mammals (including pigs, horses, seals and bats), whereas IBV has only been found in humans and seals (Osterhaus *et al*, 2000). Influenza C virus (ICV) stands out from IAV and IBV, as it expresses a bifunctional membrane glycoprotein responsible for both cell-entry and exit instead of two separate proteins. Accordingly, ICV has fewer genome segments (seven instead of eight) (Knipe *et al*,

2013). Influenza C, like B, is also quite host restricted, and is currently only known to infect humans, pigs (Guo *et al*, 1983) and dogs (Ohwada *et al*, 1986). Though ICVs have also been isolated from cattle, it has been proposed that these particular viruses actually represent a distinct genus within the *Orthomyxoviridae* (Hause *et al*, 2013; 2014). Given this recent discovery and our knowledge about the rapid evolution of RNA viruses, it seems likely that we still do not know the full diversity of the influenza viruses.

Of the other three *Orthomyxoviridae* genera, *Isavirus* contains only one known member, infectious salmon anaemia virus (ISAV), and this possesses an eight-segment genome. The remaining genera, *Thogotovirus* and *Quaranjavirus*, both represent tick borne viruses that infect vertebrate hosts and have members with at least six genome segments (Clerx *et al*, 1983; Presti *et al*, 2009; Knipe *et al*, 2013).

1.2.2 Genome Structure and Proteome

Influenza viruses have a genome of approximately 13 kb that is divided into seven or eight segments. Each of these segments gives rise to one transcript, however the number of proteins produced is greater than the number of segments because of transcript splicing and non-canonical translation (Vasin *et al*, 2014). These proteins are summarised in tables 1.1 and 1.2. The larger number of IAV proteins likely reflects the greater scrutiny this virus receives, rather than any biological difference – indeed many of these proteins have only recently been discovered. Most, but not all of the proteins listed are structural components of the virion.

On top of the proteins within table 1.1, the influenza proteome also includes a number of post-translationally modified viral proteins (Hutchinson *et al*, 2012). This should not be a surprise, since viral proteins are made within eukaryotic cells and hence exposed to the same post-translational modification machinery that any eukaryotic protein would also face. Only with the maturation of mass spectrometry techniques has it become possible to explore this proteomic diversity and we are only just starting to uncover the full extent of this. Since few modifications have currently had physiological roles assigned to them (Chenavas *et al*, 2013), it is still unclear how important these modified proteins really are to the biology of influenza.

Table 1.1. Gene segments of influenza viruses and their protein products.
Produced using information from Vasin *et al*, 2014 and Knipe *et al*, 2013.

Influenza A		Influenza B		Influenza C	
Segment	Product	Segment	Product	Segment	Product
1	PB2	1	PB2	1	PB2
2	PB1	2	PB1	2	PB1
	PB1-N40	3	PA	3	P3
	PB1-F2	4	HA	4	HEF
3	PA	5	NP	5	NP
	PA-X	6	NA	6	M1
	PA-N155		NB		CM2
	PA-N182	7	M1	7	NS1
4	HA		M2		NEP
5	NP	8	NS1		
6	NA		NEP		
7	M1				
	M2				
	M42				
8	NS1				
	NEP				
	NS3				

Table 1.2. Proteins of influenza A viruses.
Lengths shown are for influenza A/Anhui/1/2013 (H7N9). Adapted from Vasin *et al*, 2014.

Protein	mRNA	Mechanism of expression	Length (aa)	Function
PB2	Segment 1, non-spliced	Translation from AUG1	759	Subunit of polymerase. Contains cap-binding domain.
PB1	Segment 2, non-spliced	Translation from AUG1	757	Polymerase catalytic subunit.
PB1-N40	Segment 2, non-spliced	Translation from AUG5	718	Maintains balance between PB1 and PB1-F2 expression.
PB1-F2	Segment 2, non-spliced	Translation from AUG4 in +1 ORF	90	Induces mitochondria-associated apoptosis. Influences polymerase activity.
PA	Segment 3, non-spliced	Translation from AUG1	716	Subunit of polymerase. Contains endonuclease domain.
PA-X	Segment 3, non-spliced	Translation from AUG1 and +1 frameshifting at codons 190-193	252	Modulates the host response and viral virulence.
PA-N155	Segment 3, non-spliced	Translation from AUG11	568	?
PA-N182	Segment 3, non-spliced	Translation from AUG13	535	?
HA	Segment 4, non-spliced	Translation from AUG1	560	Viral receptor binding.
NP	Segment 5, non-spliced	Translation from AUG1	498	Major component of RNP complex.
NA	Segment 6, non-spliced	Translation from AUG1	465	Facilitates progeny escape from infected cells.
M1	Segment 7 non-spliced	Translation from AUG1	252	Forms structural layer underneath viral envelope.
M2	Segment 7, alternatively -spliced	Translation from AUG1	97	Proton channel important for genome unpacking during viral entry.
M42	Segment 7, alternatively -spliced	Translation from AUG1	99	Can functionally replace M2 in M2-null viruses
NS1	Segment 8 non-spliced	Translation from AUG1	217	Multifunctional – regulates viral/ host gene expression and inhibits the antiviral response.
NEP	Segment 8, alternatively -spliced	Translation from AUG1	121	Mediates vRNP export from the nucleus to the cytoplasm
NS3	Segment 7, alternatively -spliced	Translation from AUG1	174	Potentially important in adaptation to mouse hosts.

1.2.3 Virion Structure

Influenza viruses have been investigated using several electron microscopy (EM) techniques. The conclusions from these studies are summarised in figure 1.1. Flu virions are pleiomorphic. They form roughly spherical particles 80 – 170 nm in diameter, as well as capsule-like particles with a short diameter of 60 – 100 nm (Waterson *et al*, 1963; Compans *et al*, 1977; Booy *et al*, 1985; Fujiyoshi *et al*, 1994; Harris *et al*, 2006; Calder *et al*, 2010). Influenza A and B virions are nearly identical in appearance (Waterson *et al*, 1963), whereas C virions are often covered by a striking hexagonal reticular structure (Apostolov & Flewett, 1969; Compans *et al*, 1977). Long filamentous particles with a diameter of around 50 nm and reaching upwards of 1 μm in length have also frequently been observed for viruses from all influenza genera (Mosley & Wyckoff, 1946; Waterson *et al*, 1963; Apostolov & Flewett, 1969; Compans *et al*, 1977; Nishimura *et al*, 1990; Fujiyoshi *et al*, 1994; Calder *et al*, 2010).

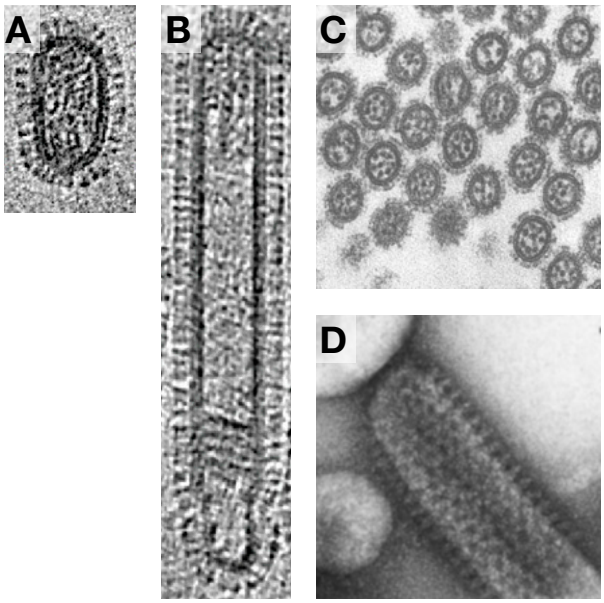
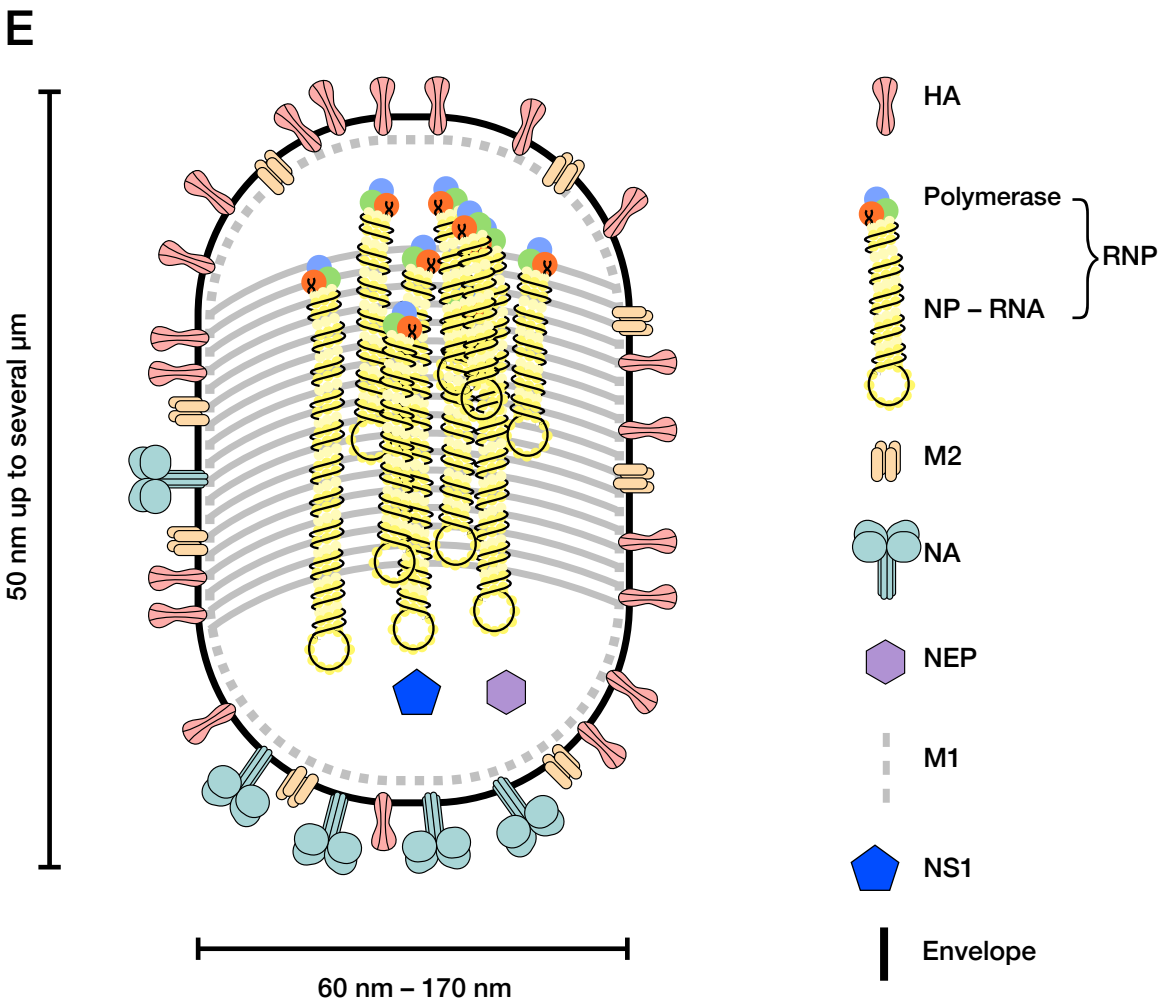


Figure 1.1. Influenza virion structure. (**A – B**) Tomogram sections of frozen-hydrated IAV virions, showing capsule-like and filamentous particles (Calder *et al*, 2010). Reprinted with permission from the National Academy of Sciences of the United States of America. (**C**) Transversely sectioned, stained IAV virions, viewed by EM (Noda *et al,* 2010). The dotted-circle arrangement of RNPs within the virion can be seen. (**D**) ICV virion stained with uranyl acetate and viewed by EM (Hewat *et al*, 1984). This reveals the reticular structure on the surface of ICV virions. Reprinted with permission from Elsevier. (**E**) Summary of the structure of IAV virions.



In IAVs, the viral envelope contains three proteins; haemagglutinin (HA), neuraminidase (NA) and the proton channel M2. In IBVs, these are joined by NB. ICVs do not express an equivalent of NB, and replace HA and NA with a bifunctional haemaglutinin-esterase-fusion (HEF) protein (Knipe *et al*, 2013).

For all influenza viruses, beneath the envelope is a layer of M1, which adopts a helical super-structure with a pitch of 38 Å (Ruigrok *et al*, 1989; Calder *et al*, 2010). This maintains the shape of the virion and is an adaptor between the envelope and the ribonucleoproteins (RNPs) at the core of the virus particle. The genome of influenza is packaged into these RNPs, which are segments of RNA wrapped around a helical oligomer of nucleoprotein (NP) and bound at both ends by a single heterotrimeric polymerase complex (Pons *et al*, 1969; Compans *et al*, 1977; Arranz *et al*, 2012; Moeller *et al*, 2012). Influenza viruses A and B have eight genome segments and therefore their virions contain eight RNPs. Influenza C only has seven segments and hence seven RNPs. The seven/eight RNPs lie in parallel inside the virion, adopting a dotted circle arrangement reminiscent of a microtubule cross-section (Fournier *et al*, 2011; Noda *et al*, 2012).

Non-structural protein 1 and nuclear export protein (NEP) are also associated with the virion (Hutchinson *et al*, 2014).

1.2.4 Ribonucleoprotein Complex Structure

Since the influenza polymerase (FluPol) always functions as part of an RNP and not as an isolated entity, it is useful to develop a deeper understanding of the structure of this complex. The RNPs found within virions contain negative-sense, viral RNA (vRNA) and hence are known more precisely as vRNPs. During replication, a positive-sense version of this vRNP is produced, called the complementary RNP (cRNP). Most of the structural information we have is for the vRNP, because it is easier to purify, though our group have seen that cRNPs have much the same form (York *et al*, 2013). Similarly, RNPs of the different influenza genera are likely indistinguishable (Compans *et al*, 1977). Negative-stain EM work from Pons *et al* as well as Compans *et al* in the late 1960s and early 1970s provided us with the description of the vRNP outlined above (Pons *et al*, 1969; Compans *et al*, 1972); that is of an NP-RNA oligomer resembling a hairpin that is clasped at the free end and twisted into a helix (figure 1.2a).

It is only relatively recently that a more detailed picture of the vRNP has emerged, thanks to EM studies carried out in the lab of Ian Wilson (Moeller *et al*, 2012), and the labs of Jaime Martín-Benito and Juan Ortín (Arranz *et al*, 2012) (figure 1.2b – d). Both groups overcame the problems of vRNP length-heterogeneity and flexibility, which had previously restricted structural analysis to much shorter mini-RNPs (Martín-Benito *et al*, 2001; Area *et al*, 2004; Coloma *et al*, 2009), by analysing straight-central and end sections of the vRNP independently of each other. They confirmed that the central region of the vRNP is a double-helix formed by two antiparallel strands of NP-RNA. The values for per-subunit rise of the helix found by each group are similar: 28.4 Å and 32.6 Å. However, the handedness of the helices are opposite leading to different rotation-per-step angles: -57º to -64º for Arranz *et al* (left-handed helix) but 73.9º for Moeller *et al* (right-handed helix).

Arranz *et al* provide additional tomographic evidence that supports their interpretation and present a more detailed reconstruction.

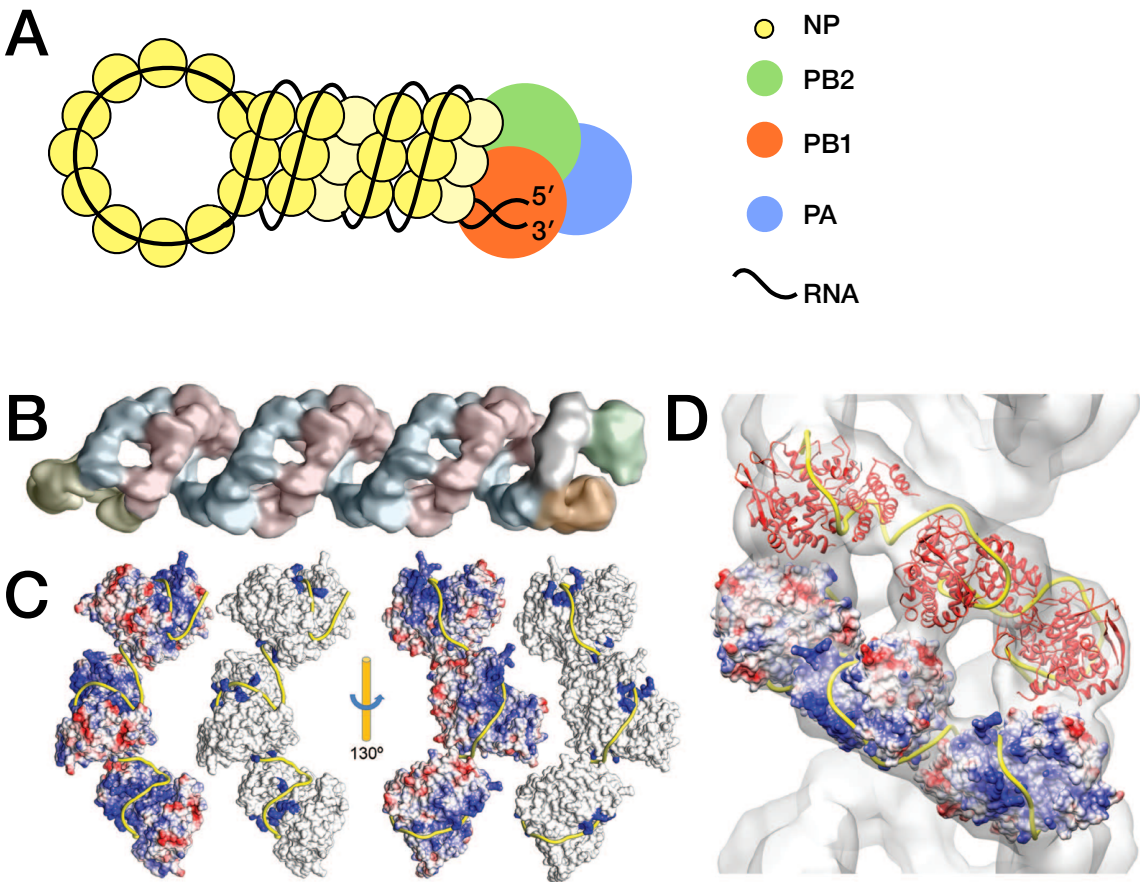


Figure 1.2. RNP structure.
(A) Schematic representation of an influenza RNP. The above subunit colouring will be consistently used throughout all subsequent figures. **(B)** Model for the segment 8 vRNP. Antiparallel strands of NP are coloured red and blue, whilst NP at the looped end is yellow. The polymerase is coloured grey, green and brown. These colours do not correspond to polymerase subunits. **(C and D)** Docking of the structure of NP, coloured by electrostatic potential, into the vRNP reconstruction. The proposed path of RNA is highlighted in yellow. Figures B – D from Arranz *et al*, 2012. Reprinted with permission from AAAS.

NP from several influenza viruses (and ISAV) has been crystallised, revealing the interaction mediating NP oligomer formation: the insertion of a tail-loop at the back of one NP into a groove in the body domain of another (Ye *et al*, 2006; Ng *et al*, 2008a; 2012b; Zheng *et al*, 2013). The NP structure can be docked into both vRNP reconstructions in a manner that is compatible with the required oligomerisation interaction, however, like earlier EM

work with mini-RNPs (Coloma *et al*, 2009), neither reconstruction has sufficient resolution to directly reveal the path of the RNA. Instead, Arranz *et al* offer a reasonable-looking model for the RNA based on connecting the channels of positive-charge on the surface of NP protomers within each strand (figure 1.2c and d). This model shows the RNA taking a serpentine route around each NP strand, but not actually wrapping completely around it (the path is not helical). The topology of the NP-RNA complex is something that must be kept in mind when thinking about the mechanics of polymerase elongation.

At the non-polymerase end of the RNP, NP seems to form a simple, small loop. In the Arranz *et al* model, only three NPs are present in the loop connecting the two strands, though others have shown that the loop can vary in size between three to eight NP protomers (Moeller *et al*, 2012; York *et al*, 2013).

The polymerase end of the RNP has naturally attracted the most interest, but is the most difficult to interpret. This may be due to difficulties in distinguishing the polymerase end from the non-polymerase end on top of conformational flexibility of the polymerase. Indeed, Arranz *et al* find that they can identify two forms of the polymerase in the vRNP (figure 1.3a). Moeller *et al*, in addition to finding an unexpected, additional NP monomer associated with the polymerase (figure 1.3b), also report that the polymerase within a vRNP adopts a different conformation from an isolated polymerase (compare with figure 1.6d) The region of density occupied by the extra NP in the Moeller reconstruction, corresponds to regions in other reconstructions that have been labelled as polymerase (Coloma *et al*, 2009).

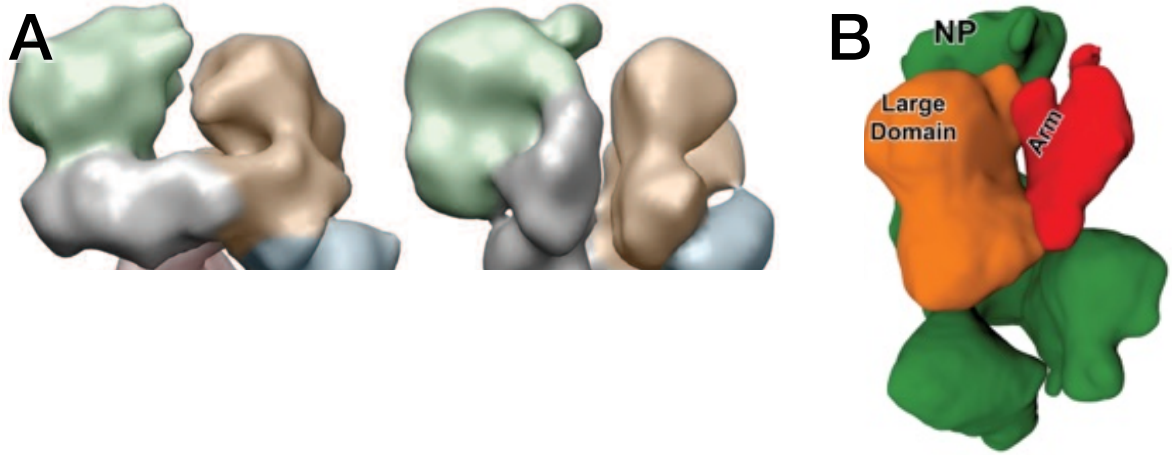


Figure 1.3. Influenza polymerase within the vRNP.
Reconstructions of the polymerase end of the vRNP, from Arranz *et al*, 2012 (**A**) and Moeller *et al*, 2012 (**B**). In panel A, the same colouring is used as in figure 1.2b, whereas in panel B, the polymerase is coloured red and orange. Reprinted with permission from AAAS.

The other major outstanding question about the structure of the vRNP is the arrangement of the vRNA at the polymerase end. This topic will only be touched upon here, but I will go into much more detail in the discussion of FluPol structure below. For now, it suffices to say that the stretch of each gene segment associated with FluPol is known as the promoter. As mentioned above, this comprises both 5′ and 3′ ends of the RNA. The 12 to 13 nucleotides at these ends are highly conserved across gene segments and virus strains, and are partially complementary; forming an imperfect duplex structure in solution (Bae *et al*, 2001; Noble *et al*, 2011). The structure of the polymerase-bound promoter remains unclear, though a mechanistic understanding of polymerase initiation depends absolutely on knowing this.

1.2.5 Viral Lifecycle

Influenza viruses attach to a host cell membrane by binding to sialic acids, which are ubiquitous monosaccharide moieties found in membrane glycoproteins. These are bound by the envelope glycoproteins HEF or HA . Once bound, the virions are internalised by cellular endocytic processes. As the endosome acidifies, HA or HEF undergo a conformational change that causes fusion of the viral envelope and endosome membrane. The environment inside the virion also acidifies, due to the action of the M2 proton channel in the viral envelope. This change in pH disrupts the matrix underlying the viral envelope and its interactions with vRNPs. Together, these processes bring about the release of vRNPs into the cytoplasm (Hutchinson & Fodor, 2013; Knipe *et al*, 2013).

The released individual vRNPs travel into the nucleus. For IAVs, this occurs through interactions between a nuclear localisation signal (NLS) on NP and the import factor, karyopherin α (Martin & Helenius, 1991; O'Neill *et al*, 1995). It is unclear which signals direct import for IBV and ICV vRNPs.

Once inside the nucleus, the influenza polymerase can begin producing viral mRNA. The polymerase initiates transcription close to the 3′ end of each vRNA using a 10 – 13 nt 7-methyl-GTP capped primer cleaved from host mRNAs (Bouloy *et al*, 1978; 1979; 1980; Beaton & Krug, 1981; Plotch *et al*, 1981). This mechanism for generating capped transcripts is known as cap-snatching, and is thought to be used by all *Orthomyxoviridae,* as well as *Bunyaviridae* and *Arenaviridae*. In contrast, other nsRNA viruses, such as *Filoviridae* and *Paramyxoviridae*, have polymerases that are able to directly synthesise cap structures (Decroly *et al*, 2012). One of the advantages of basing transcription within the nucleus is that it provides the cap-snatching polymerase with ready access to host mRNAs. Also, it endows influenza with the ability to exploit nuclear splicing machinery to increase coding

capacity. Transcription termination occurs at a poly-uridine (poly-U) site 16 nt upstream of the 5′ end of each vRNA(Robertson *et al*, 1981; Luo *et al*, 1991; Li & Palese, 1994). Again, the influenza virus polymerase itself catalyses polyadenylation (Poon *et al*, 1998), probably by stuttering repeatedly across this poly-U region(Poon *et al*, 1999; Zheng *et al*, 1999). Thus, influenza mRNAs are nearly complete, capped and polyadenylated positive-sense copies of each genome segment.

Replication of vRNPs requires soluble NP (Beaton & Krug, 1986), and so occurs later in the infection cycle, after some protein synthesis has already taken place (Hay *et al*, 1977; Taylor *et al*, 1977; Barrett *et al*, 1979; Hatada *et al*, 1989). Replication is a two-step process. Firstly, the polymerase copies each vRNA to produce a positive-sense, complementary RNA (cRNA) intermediate. In contrast to viral mRNAs, these are full length copies of each segment and are not capped or polyadenylated (Skehel & Hay, 1978). cRNAs are also encapsidated by NP, forming cRNPs (York *et al*, 2013). Secondly, the polymerase copies these cRNAs back, forming new vRNAs/vRNPs. Both cRNA and vRNA have triphosphates at their 5′ ends, demonstrating replicative processes are primer-independent (Hay *et al*, 1977; 1982). The processes of transcription and replication are summarised in figure 1.4.

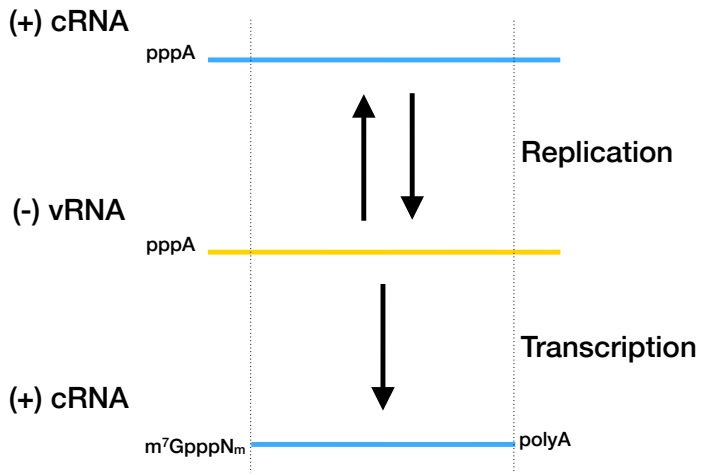


Figure 1.4. Processes catalysed by the influenza RdRp.
pppA indicates a triphosphate at the 5′ end.

To complete the infection cycle, vRNPs are exported from the nucleus and proceed to the apical cell membrane together with other structural proteins for virion assembly and budding. RNPs are coupled to the nuclear export machinery through M1 and NEP (Paterson & Fodor, 2012; Hutchinson & Fodor, 2013). As assembly of a virion takes place, so the membrane around it distorts outward, mainly under the influence of M1 oligomerisation. This "budding" is completed when the membranes at the base of the bud pinch together due to the action of M2. Finally, NA or HEF cleave sialic acid moieties bound by the virion, thus completely freeing it from the surface of the cell (Knipe *et al*, 2013).

1.3 Influenza Polymerase Structure

1.3.1 Overall Structure

FluPol is a 255 kDa heterotrimer, comprised of PB1 (polymerase basic protein 1), PB2 and PA (polymerase acidic protein, known as P3 in ICV). All three subunits are roughly 85 kDa in size, and all are needed for polymerase activity. PB1 contains all six of the canonical motifs of RdRps (Poch *et al*, 1989; Müller *et al*, 1994; Bruenn, 2003), and hence is where the polymerase active site resides. Unsurprisingly then, it is also the backbone subunit of the polymerase, able to independently pull-down PA and PB2 in immunoprecipitation assays (Digard *et al*, 1989; González *et al*, 1996; Zürcher *et al*, 1996; Ohtsu *et al*, 2001). Indeed the only fragments of PB1 that have been crystallised are its N- and C-termini, which are in complex with domains from PA and PB2 respectively (He *et al*, 2008; Obayashi *et al*, 2008; Sugiyama *et al*, 2009). A direct interaction between PA and PB2 can be detected using bi-molecular fluorescence complementation assays, but not immunoprecipitation, indicating that this is much weaker than the PB1 interactions (Hemerka *et al*, 2009). Cellular localisation and subunit purification experiments show that PB1 and PA assemble into a stable dimer before binding PB2 (Fodor & Smith, 2004; Deng *et al*, 2005). This suggests that, though only a short helix from PB1 specifically binds PA (Ohtsu *et al*, 2001; He *et al*, 2008; Obayashi *et al*, 2008), the two subunits nevertheless contact each other reasonably intimately. This is supported by the finding that PB1-binding protects a flexible linker in PA from proteolysis (Guu *et al*, 2008). PB2 forms a looser association, on top of the PB1-PA dimer, but interacting mainly with PB1.

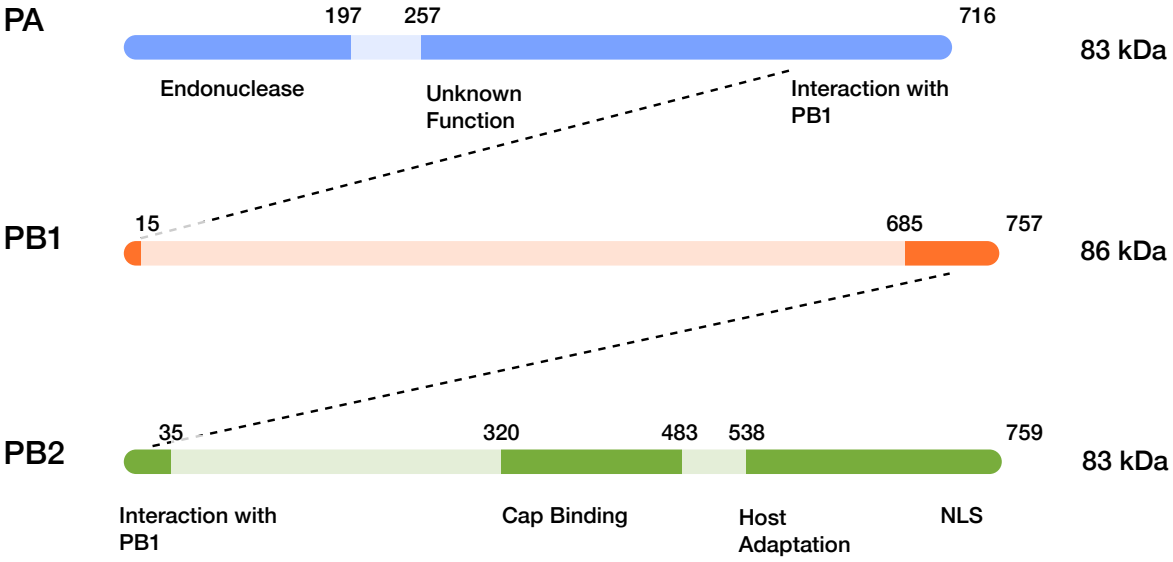


Figure 1.5. Solved fragments of FluPol$_A$ structure.
Each bar represents the primary sequence of a polymerase subunit, with bold sections highlighting regions for which the structure is known. The numbers above the bars indicate the residue boundaries of these regions. Fragments that have been crystallised in complex with each other are joined by a dotted line. The IBV cap-binding domain has also been crystallised (PDB ID: 4ORO, on hold). Figure adapted from Ruigrok *et al*, 2010.

X-ray crystallography has provided the structure of several small fragments of the IAV polymerase (FluPol$_A$) and one of the IBV polymerase (FluPol$_B$). This is summarised in figure 1.5. PB2 contains the cap binding domain, involved in cap-snatching (Fechter *et al*, 2003; Guilligay *et al*, 2008), in addition to a domain of unknown function that is important in host adaptation (Tarendeau *et al*, 2008; Kuzuhara *et al*, 2009; Yamada *et al*, 2010). PA/P3 contains an endonuclease active site (Dias *et al*, 2009; Yuan *et al*, 2009), also part of the cap-snatching machinery. However, we currently only have low resolution (13 Å maximum), EM reconstructions for the entire polymerase heterotrimer (Martín-Benito *et al*, 2001; Area *et al*, 2004; Torreira *et al*, 2007; Coloma *et al*, 2009; Resa-Infante *et al*, 2010; Arranz *et al*, 2012; Moeller *et al*, 2012). These suggest that the polymerase is a roughly elliptical particle; about 100 Å in diameter and containing a large central cavity that is accessible by several openings. An exception is the reconstruction reported by *Moeller et al*, in which a distinct "arm" domain projects out of the main body of the polymerase. Several attempts have been

made to assign particular regions of density to specific subunits, by labelling with antibodies and docking x-ray structures (Area *et al*, 2004; Coloma *et al*, 2009; He *et al*, 2008; Moeller *et al*, 2012). However both the low resolution of these reconstructions and the differences between them hinders the interpretability and usefulness of these assignments (figure 1.6). Importantly, this means that even though we have crystal structures of several isolated domains of FluPol, we still do not know how these domains are arranged within the trimer and therefore lack a holistic picture of how the polymerase functions.

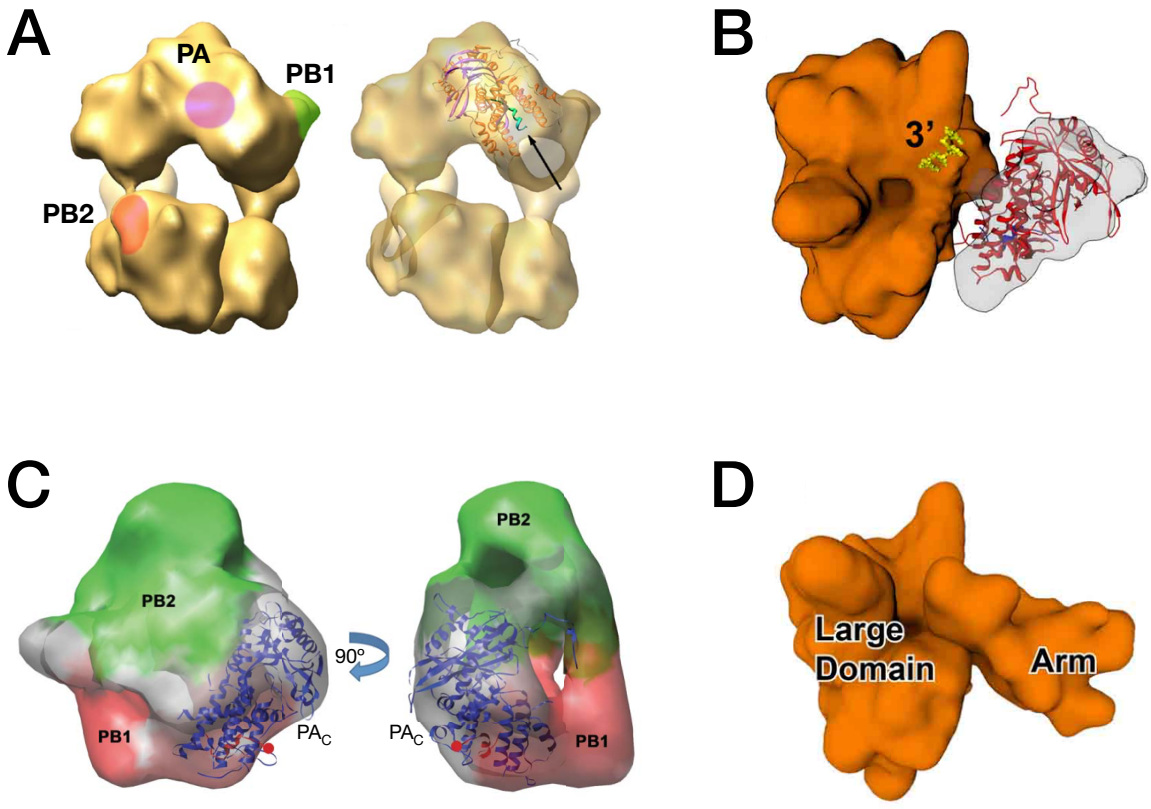


Figure 1.6. Comparison of FluPol EM structures.
(**A – B**) Reconstructions of the polymerase within an RNP. The position of each subunit, inferred by antibody labelling, is indicated. Each panel also shows the docked structure of the C-terminus of PA (PA$_C$). (**C – D**) Reconstructions of isolated polymerase heterotrimers. Again, the docked structure is that of PA$_C$. Figure assembled from Coloma *et al,* 2009, He *et al,* 2008 and Moeller *et al,* 2012. Reprinted with permission from AAAS and Macmillan Publishers Ltd.

1.3.2 PB1 – Interaction Domains

Frustratingly, given that it is the site of polymerisation, PB1 has remained the most intractable subunit to structural investigation. As mentioned above, only the structures of its extreme N- and C-termini are known (figure 1.7a). The first 15 amino acids of PB1 form a small helix that binds, via hydrogen bonds and hydrophobic interactions, the C-terminal domain of PA (He *et al*, 2008; Obayashi *et al*, 2008). Residues 16 – 81 are flexible, because they are not visible in the structure obtained by *Obayashi et al*, but were present in the crystal. At the other end of PB1, the last 72 residues form three helices that interact with the N-terminus of PB2 (figure 1.8) (Sugiyama *et al*, 2009). The large buried surface area at this latter interface (over 1400 Å^2) suggests a strong affinity between the two polypeptides. This structure also revealed four salt bridges linking the two chains: K698 to E2, E6 and R3; and D725 to R3; however these were not verified by mutation. Notably, the mutational work that was carried out did generate some unexpected results, suggesting that this domain is more than a simple interaction site. For instance, the V715S mutation did not disrupt the PB1-PB2 interaction, agreeing with the prediction from the structure, but did inhibit polymerase activity. Moreover, the F699A and I750D mutations weakened PB1-PB2 association, yet seemed to slightly increase overall catalytic activity. Thus the authors conclude that this domain could allow for regulation of polymerase activity by PB2. This is not an unreasonable suggestion, given the need for communication between subunits for the coordination of polymerisation with cap-snatching.

Finally, a bipartite NLS is present in IAV PB1, occupying residues 187 – 190 and 207 – 211 (Nath & Nayak, 1990). This NLS interacts with the importin RanBP5 (Hutchinson *et al*, 2011), but the structure of the complex is not known.

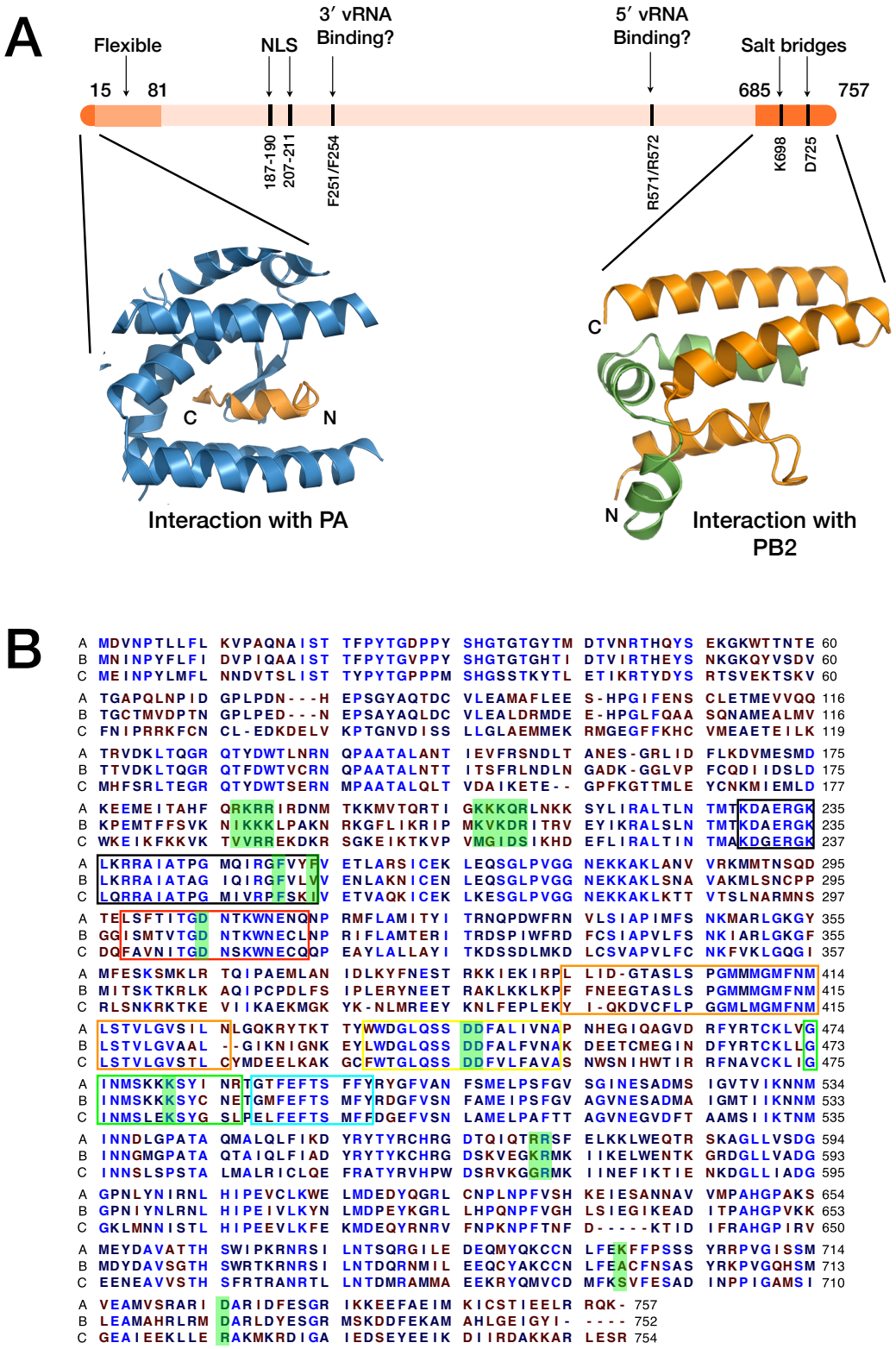


Figure 1.7. PB1 structure summary.
(**A**) Solved fragments of PB1 structure (PDB IDs: 2ZNL and 3A1G), with a selection of other regions annotated. (**B**) PB1 sequence alignments coloured blue to red, according to conservation from high to low respectively. Certain important residues are highlighted in green (see text for details). Polymerase motifs are boxed, using the same colour scheme as in figure 1.10. The alignment was generated using *M-Coffee*, with PB1 sequences of the A(duck)/Fujian/1/2002, B/Panama/45/1990 and C/Johannesburg/1/1966/ viruses (Moretti *et al,* 2007).

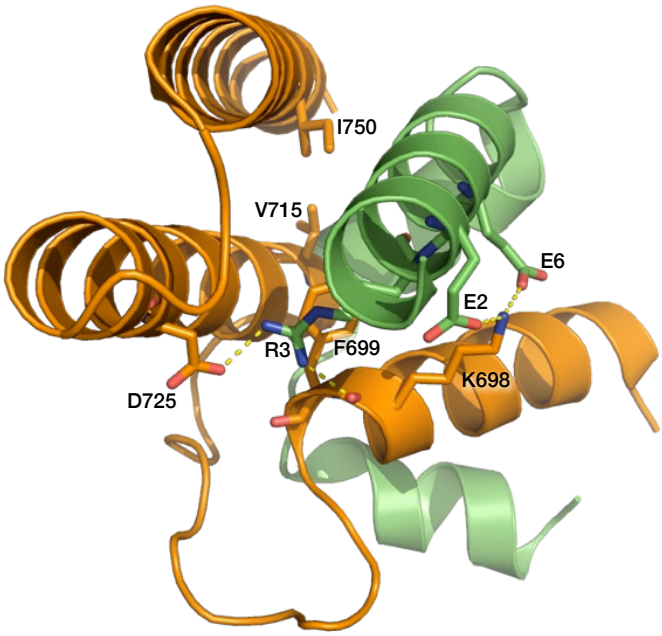


Figure 1.8. PB1-PB2 interaction domain.
Crystal structure of the C-terminus of PB1 (orange) in complex with the N-terminus of PB2 (green). Residues involved in the interaction are marked. Polar contacts are indicated by dotted yellow lines. PDB ID: 3AIG.

1.3.3 PB1 – Inferences from Other Polymerases

As mentioned above, the central portion of PB1 contains all of the motifs required to form the polymerase catalytic site. Though we do not currently know the structure of this region, we can be reasonably confident that it will resemble every other viral RdRp core, since the structures of all such proteins solved to date share a common architecture; namely a fold reminiscent of a cupped right-hand, with distinct "thumb", "fingers" and "palm" domains (figure 1.9a) (Ng *et al*, 2008b). Indeed, this similarity extends to an incredible number of polymerases, comprising DNA/RNA-dependent DNA and RNA polymerases from viruses and all three domains of life (Mönttinen *et al*, 2014). Some of these right-handed polymerases are shown in figure 1.9b. The palm is the most structurally conserved of the core polymerase domains, recognisably consisting of three to four strands of antiparallel β-sheet packing closely with two to three α-helices. In PB1, this will lie

approximately between residues 300 to 500, as this stretch contains the four characteristic viral-RdRp motif sequences, A – D that make up the palm domain (Hansen *et al*, 1997). This is illustrated in figure 1.10, which shows the position of these motifs in the PB1 sequence and their likely structure, based on the hepatitis C virus polymerase. Outside of the palm domain, it is difficult to infer the structure of PB1 from that of other polymerases, since the fingers and thumb domains are less structurally conserved. What we can say is that these regions should be mainly α-helical (though sometimes the fingers contain an antiparallel beta sheet), with the fingers likely being formed by residues N-terminal of the palm (250 – 300 amino acids) and between motifs A and B. The thumb probably folds from residues C-terminal of the palm (200 amino acids) (Hansen *et al*, 1997; Butcher *et al*, 2001; Tao:2002wq Ferrer-Orta *et al*, 2004; Salgado *et al*, 2004). These domains make extensive interactions with upstream template nucleic acids, so in PB1 could well be involved in promoter binding (Li *et al*, 1998), as well as forming the pore through which (d)NTPs travel to the active site. RdRps of positive-sense RNA and double stranded RNA viruses also contain the "fingertips" domain, which is a series of loops connecting the fingers and thumb domains, so maintaining these polymerases in a "closed-hand" conformation (Ferrer-Orta *et al*, 2006; Ferrer-Orta & Verdaguer, 2009). It is unclear whether PB1 will also display this fold because no structure for a nsRNA virus RdRp is yet known.

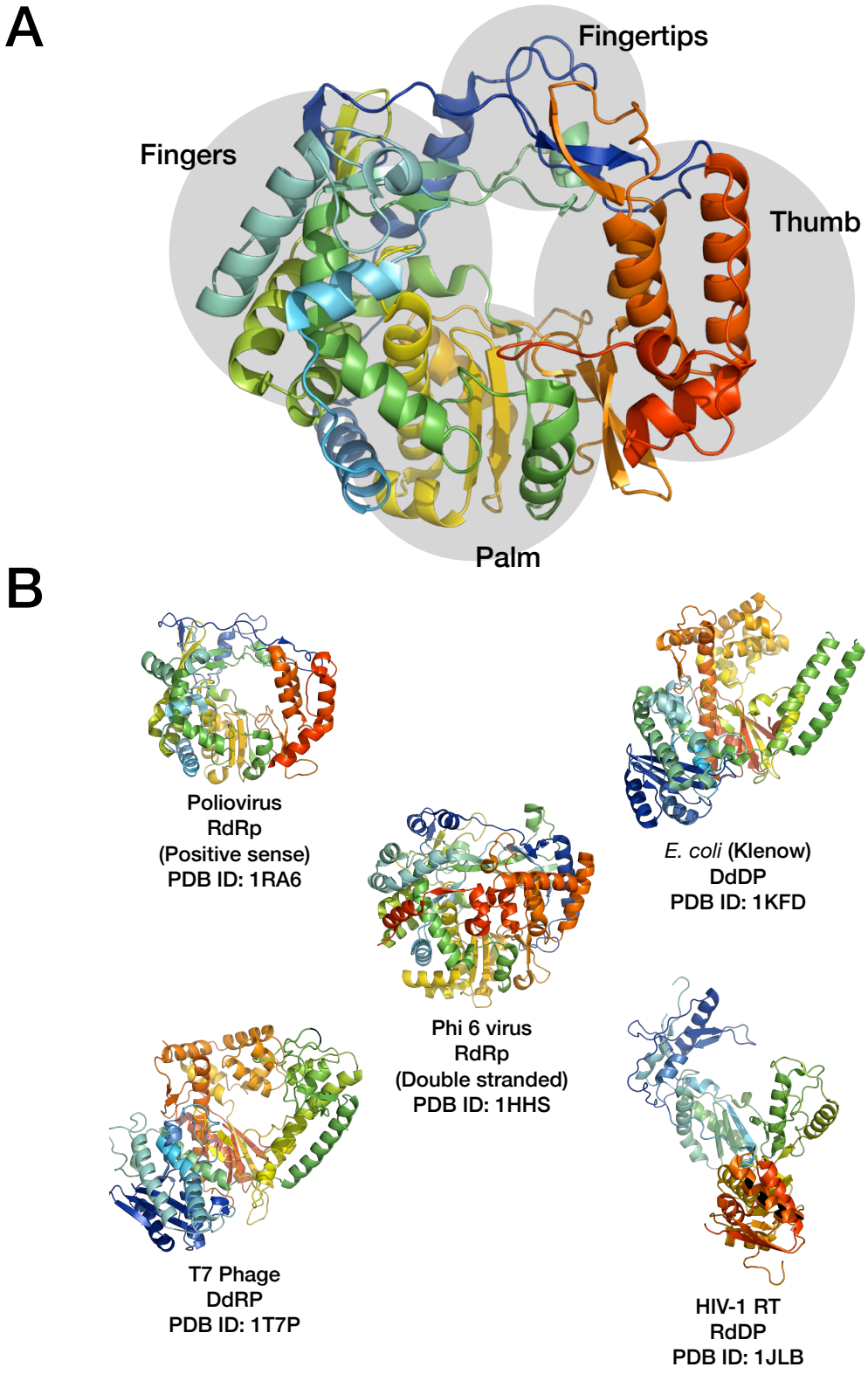


Figure 1.9. The architecture of right-handed polymerases.
(**A**) Crystal structure of Norwalk virus RdRp (PBD ID: 1SH0), highlighting the subdomains found in all RdRps. Residues are coloured according to position in the polypeptide chain, going from the N-terminus (Blue) to the C-terminus (Red). (**B**) Crystal structures of a range of different polymerases (coloured as above), illustrating the similarity between the core domains of all right-handed polymerases. DdDP: DNA-dependent DNA polymerase, DdRP: DNA-dependent RNA polymerase RdDP: RNA-dependent DNA polymerase.

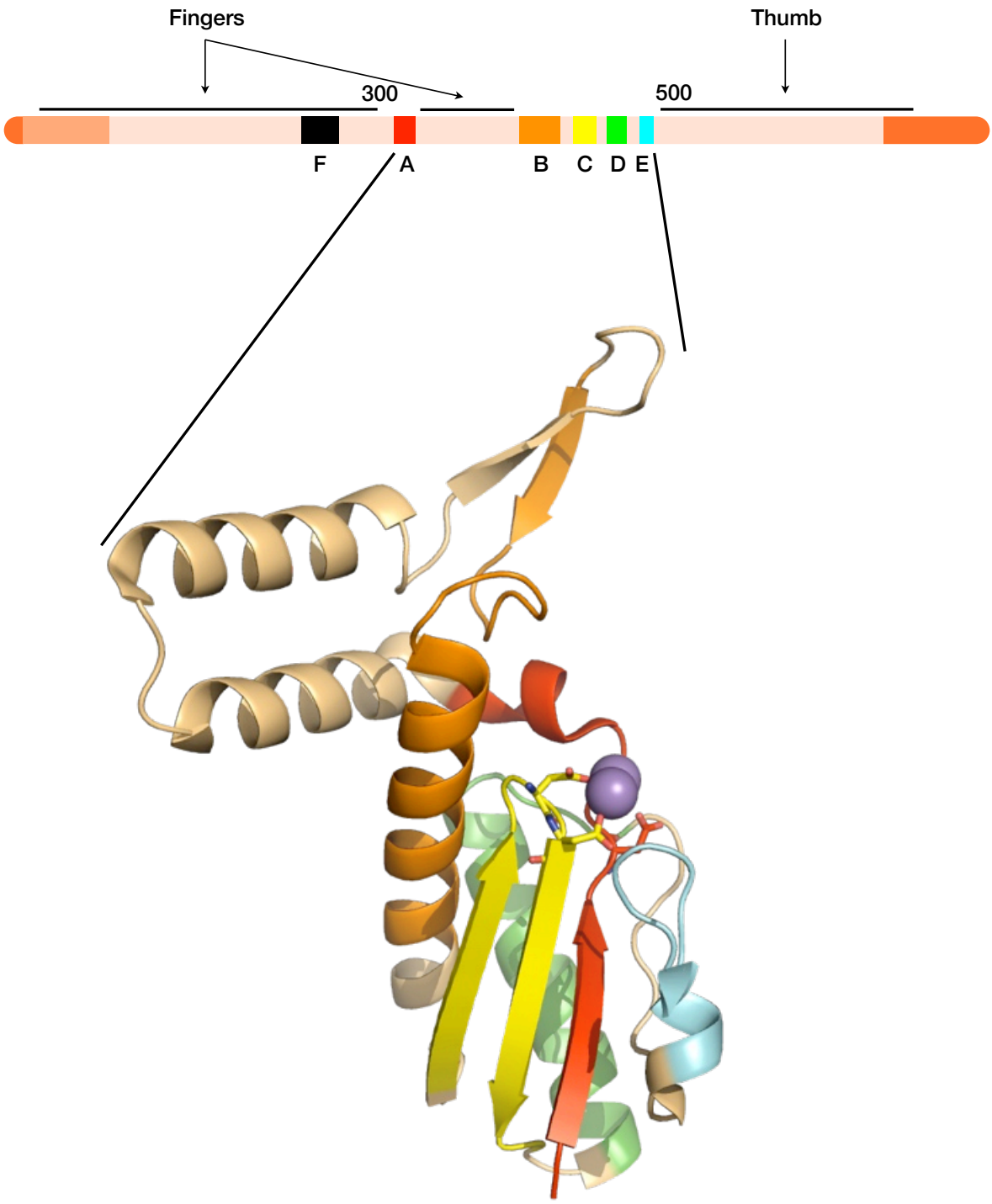


Figure 1.10. Probable structure of the central region of PB1.
Structure of the central region (containing the palm domain) of hepatitis C virus RdRp (PDB ID: 1GX5), coloured, as indicated, according to the corresponding RdRp motif sequence. The central region of PB1, as it has all of these motifs, is likely to adopt a similar conformation. Two metal ions at the active site are indicated by the grey spheres, whilst the active-site aspartates are shown in stick representation.

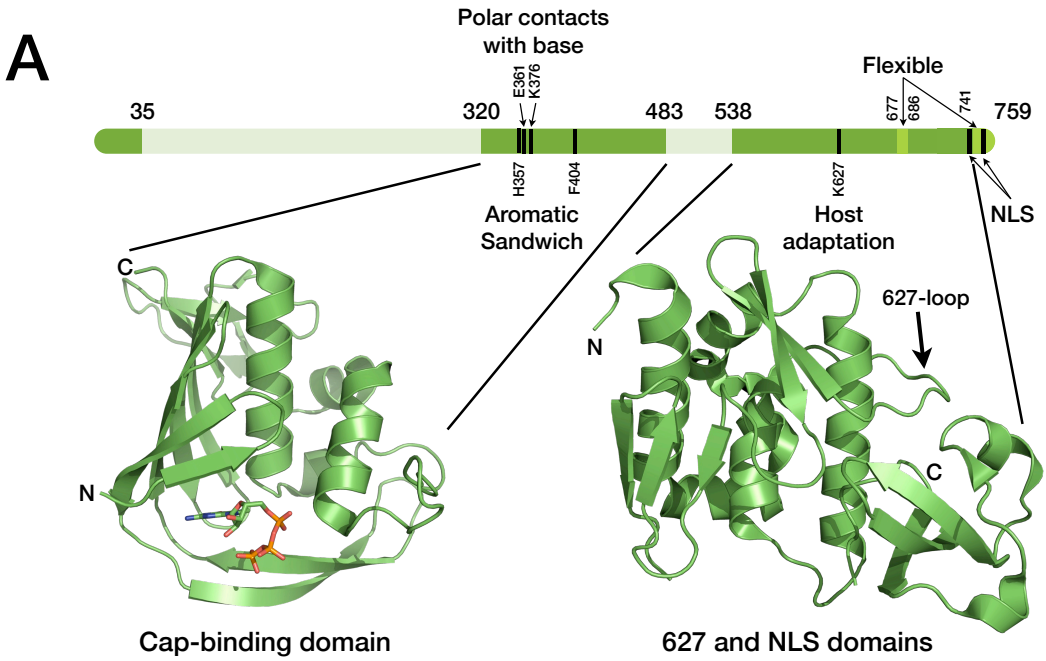
1.3.4 PB1 – Catalytic Mechanism

The palm domain is the site of phosphodiester bond formation, and the chemistry employed here is the same as with all enzymes catalysing nucleotidyl transfer reactions; namely the two-metal-ion mechanism (Steitz & Steitz, 1993; Yang *et al*, 2006). In polymerases, the metals involved are usually Mg^{2+} ions, and these are precisely coordinated in two positions (designated A and B) by strictly conserved aspartates in the A and C motifs together with an adjacent backbone carbonyl (figure 1.10). Mg^{2+} at position A lowers the pKa of the 3′-OH group thus helps activate the end of the nascent strand for S_N2-type nucleophilic attack on the incoming nucleotide triphosphate (NTP). A general base within the enzyme is also likely required for this deprotonation event, however the identity of this base is still elusive (Castro *et al*, 2007). The position B Mg^{2+} interacts with the phosphates of the incoming nucleotide, both stabilising the pentacovalent transition state and, due to differences in coordination geometry, actively favouring product formation (Yang *et al*, 2006). It has also been suggested that the pyrophosphate (PPi) leaving group is protonated by a conserved lysine (around residue 470 in influenza viruses) or histidine in motif D, though very few crystallographic structures are compatible with this (Castro *et al*, 2009; Michielssens *et al*, 2011). This motif is flexible, and does not close to form a catalytically-competent structure around incorrectly base-paired incoming nucleotides (Yang *et al*, 2012). Therefore, motif D also probably contributes to polymerase fidelity.

In addition to the four motifs found in all viral polymerases, RdRps including PB1 contain two additional conserved sequences: motifs E and F. Motif E of RdRps folds into a β-turn-β structure that connects the palm and thumb domains, maintaining the correct relative orientation between these domains (O'Reilly & Kao, 1998). Motif F lies in the fingers domain, N-terminal of motif A, and forms a pre-active-site nucleotide binding pocket, which may allow incoming NTPs to probe the RNA template (Butcher *et al*, 2001).

1.3.5 PB2 – Cap-Binding Domain

As shown in figure 1.11, the structure of most of the C-terminal half of IAV PB2 is known (as well as that of the interaction domain with PB1). Residues 320 to 483 of IAV PB2 fold into a cap-binding domain, which facilitates the use of cellular mRNAs as primers for viral transcription. Up until the first structure was solved in 2008 (Guilligay *et al*, 2008), the location of this domain within PB2 had been controversial because the crosslinking assays used to pinpoint this had given conflicting and incorrect results (Honda *et al*, 1999; Li *et al*, 2001). The only correct assignments of cap-binding-domain residues, F363 and F404, came out of mutagenesis studies carried out by our group, but this still did not reveal the full extent of the domain (Fechter *et al*, 2003). Guilligay *et al* successfully solved this problem using a high-throughput screening method, called ESPRIT (Tarendeau *et al*, 2007; Yumerefendi *et al*, 2010), to identify fragments of PB2 that would both be solubly expressed from *E. coli* and bind to m^7GTP-sepharose. The ESPRIT technique was also instrumental in the solution of the C-terminal domains of PB2 (Tarendeau *et al*, 2007; 2008).



B

```
A ----MERIKE LRDLMSQSRT REILTKTTVD HIAIIKKYTS GRQEKNPALR MKWMMAMKYP 56
B --MTLAKIEL LKQLLRDNEA KTVLKQTTVD QYNIIRKFNT SRIEKNPSLR MKWAMCSNFP 58
C MSLLLTIAKE YKRLCQDAKA AQMMTVGTVS NYTTFKKWTT SRKEKNPSLR MRWAMSSKFP 60

A ITADKRIMEM -IPERNEQGQ TLWSKTNDAG S-DRVMVSSL AVTWWNRNGP -TTSTVHYPK 113
B LALTKGDMAN -RIPLEYKGI QLKTNAEDIG T-KGQMCSIA AVTWWNTYGP -IGDTEGFEK 115
C IIANKRMLEE AQIPKEHNNV ALWEDTEDVS KRDHVLASAS CINYWNFCGP CVNNSEVIKE 120

A VYKTYFEKVE RLKHGTFGPV HFRNQVKIRR RVDVNPGHAD LSAKEAQDVI MEVVFPNEVG 173
B VYESFFLRKM RLDNATWGRI TFGPVERVRK RVLLNPLTKE MPPDEASNVI MEILFPKEAG 175
C VYKSRFGRLE RRKEIMWKEL RFTLVDRQRR RVDTQPVEQR LRTGEIKDLQ MWTLFEDEAP 180

A ARILTSESQL TITKEKKEEL QDCKIAPLMV AYMLERELVR KTRFLPVAGG TSSVYIEVLH 233
B IPRESTWIHR ELIKEKREKL KGTMITPIVL AYMLERELVA RRRFLPVAGA TSAEFIEMLH 235
C LASKFILDNY GLVKEMRSKF ANKPLNKEVV AHMLEKQFNP ESRFLPVFGA IRPERMELIH 240

A LTQGTCWEQM YTPGGEVRND DVDQSLIIAA RNIVRRATV- STDPLASLLE MCHSTQIG-- 290
B CLQGENWRQI YHPGGNKLTE SRSQSMIVAC RKIIRRSIV- ASNPLELAVE IANKTVID-- 292
C ALGGETWIQE ANTAGISNVD QRKNDIRAVC RKVCLAANAS IMNAKSKLVE YIKSTSMRIG 300

A --GIRMVDIL RQNPTEEQAV DICKAAMGLR ISSSFSFGGF TFKRTSGSSV KKEEEVLTGN 348
B --TEPLKSCL TAIDGGDVAC DIIRAALGLK IRQRQRFGRL ELKRISGRGF KNDEEILIGN 350
C ETERKLEELI LETDDVSPEV TLCKSALGGQ LGKTLSFGPM LLKKISGSGV KVKDTVYIQG 360

A LQTLKIRVHE GYEEFTMVGR RATAILRKAT RRLIQLIVSG RDEQSIAEAI IVAMVFSQED 408
B GTIQKIGIWD GEEEFHVRCG ECRGILKKSK MRMEKLLINS AKKEDMKDLI ILCMVFSQDT 410
C VRAVQFEYWS EQEEFYGEYK SATALFSRKE RSLEWITIGG GINEDRKRLL AMCMIFCRDG 420

A CMIKAVRGDL NFVNRAN--- QRLNPMHQLL RHFQKDAKV- --L-FQNWGI EPIDNVMGMI 461
B RMFQGVRGEI NFLNRAG--- QLLSPMYQLQ RYFLNRSND- --L-FDQWGY EESPKASELH 463
C DYFKDAPATI TMADLSTKLG REIPYQYVMM NWIQKSEDNL EALLYSRGIV ETNPGKMG-S 479

A GILPDMTPST EMSLRGVRVS KMGVDEYSST ERVVVSIDRF LRVRDQRGNV LLSPEEVSET 521
B GIN-ELMNAS DYTLKGVVVT KNVIDDFSST ETEKVSITKN LSLIKRTGEV IMGANDVSEL 522
C SMGIDGSKRA IKSLRAVTIQ SGKIDMPESK EKIHLELSDN LEAFDSSGRI VATILDLPSD 539

A QGTERLTITY SSSMMWEING PESVLVNTYQ WVIRNWETVK IQ--WSQD-- -PTMLYNKME 576
B ESQAQLMITY DTPKMWEMGT TKELVQNTYQ WVLKNLVTLK AQ--FLLG-- -KEDMFQWDA 577
C KKVTFQDVSF QHPDLAVLRD EKTAITKGYE ALIKRLGTGD NDIPSLIAKK DYLSLYNLPE 599

A FEPFQSLIPK AARGQQYSGFV RTLFQQMRDV VGTFDTVQII KLLPFTAAPP EQ--SRMQFS 634
B FEAFESIIPQ KMAGQQYSGFA RAVLKQMRDQ -EVMKTDQFI KLLPFCFSPP KLRSNGEPYQ 636
C VKLMAPLIRP NRKGVYSRVA RKLVSTQVTT -GHYSLHELI KVLPFTYFAP K---QGMFEG 655

A SLTVNVRGSG -MRILVRGNS PVFNYNKATK RLTVLG-KDA GALTEDPDE- GTAGVESAVL 691
B FLRLVLKGGG ENFIEVRKGS PLFSYNPQTE VLTICG-RMM SLKGKIEDEE RNRSMGNAVL 695
C RLFF----SN DSFVEPGVNN NVFSWSKADS SKIYCHGIAI RVPLVVGDEH MDTSL--ALL 709

A RGFLILGKED KRYGPALSIN ELSNLAKGEK ANVLIGQGDV VLVMKRKRDS SILTDSQTAT 751
B AGFLVSGKYD PDLGDFKTIE ELEKLKPGEK ANILLYQGKP VKVVKRKRYS ALSNDISQGI 755
C EGFSVC-END PR-APMVTRQ DLIDVGFGQK VRLFVGQGSV RTFKRTASQR AASSDVNKNV 767

A KRIRMAIN-- ----- 759
B KRQRMTVESM GWALS 770
C KKIKMSN--- ----- 774
```

Figure 1.11. PB2 structure summary.
(**A**) Solved fragments of PB2 structure (PDB IDs: 2VQZ and 2VY6), with a selection of other regions annotated. (**B**) PB2 sequence alignments. See figure 1.7b for details.

The cap binding domain folds into a bundle of four α-helices, enclosed by a five-stranded antiparallel β-sheet, a β-hairpin and a C-terminal region of short β-strands (figure 1.11). The domain was co-crystallised in the presence of m^7GTP (the terminal residue of the cap structure), and this revealed the cap-binding site to be in a positively charged region at the intersection of the β-sheet, β-hairpin and α-helical subdomains (figures 1.11a and 1.12a). This fold is novel among cap-binding proteins, and the mode of recognition is familiar, but with interesting differences. Like other cap-binding proteins, such as eIF4E and vaccinia virus VP39, the methyl-guanine base in PB2 is sandwiched between two aromatic residues, H357 and F404 (figure 1.12b). However the involvement of a histidine, instead of a tyrosine or tryptophan, makes this aromatic sandwich unique specifically to IAVs (even IBVs and ICVs employ tryptophan at this position). Indeed the specificity of cap recognition seems to differ between the influenza genera (Wakai *et al*, 2011). Histidine cannot make such extensive π-stacking interactions as tyrosine or tryptophan, and as a result, IAV PB2 has >1000x weaker affinity for m^7GTP compared to eIF4E, and approximately 20x poorer discrimination between the methylated and unmethylated trinucleotides. Also in contrast with other cap-binding proteins, PB2 does not interact with the ribose hydroxyl groups of m^7GTP. The functional reasons for these differences are not yet understood.

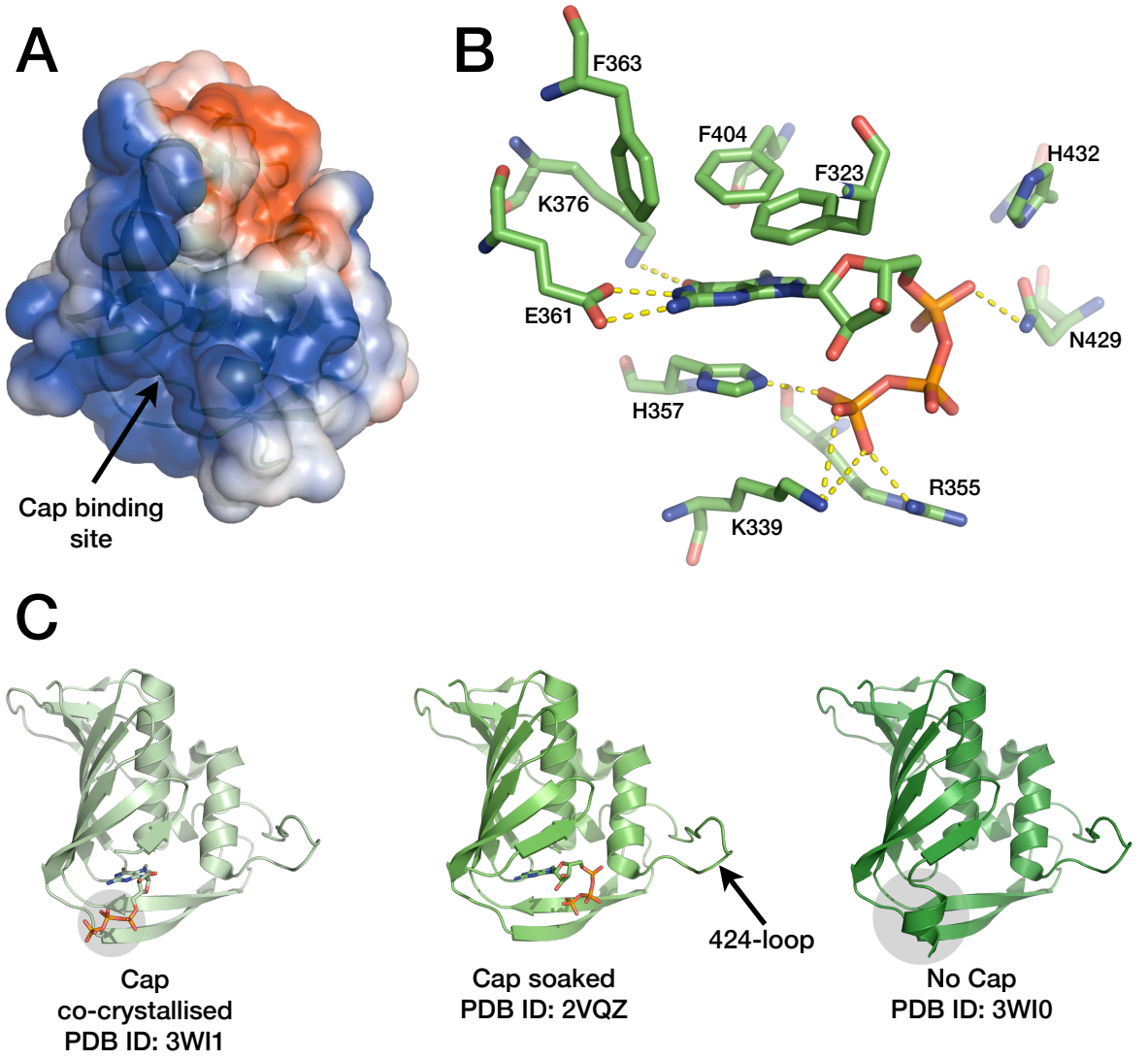


Figure 1.12. PB2 cap-binding domain.
(**A**) A solvent-accessible surface representation of the PB2 cap-binding domain bound to m^7GTP (PDB ID: 2VQZ), coloured by electrostatic potential (red = -2.0 k_bT/e_c, blue = +2.0 k_bT/e_c, where k_b is Boltzmann's constant and e_c is the elementary charge) as calculated using *APBS* (Baker *et al*, 2001). The same colour scheme will be used for all subsequent diagrams of surface electrostatics in this chapter. (**B**) Stick representation of the same structure, showing m^7GTP and the residues around the binding site. (**C**) Crystal structures of the PB2 cap-binding domain with and without bound cap. The slight differences in cap binding and PB2 structure are highlighted.

As well as the aromatic sandwich, the cap is held in place by polar contacts between the methylated guanine, and E361 and K376 (also conserved in the PB2 of IBV and ICV); mutation of either of these residues completely abolishes binding capability of the domain. Interestingly, there is some disagreement about the position of the nucleotide phosphate

groups (figure 1.12c): in two independent structures (both co-crystallisations), the phosphates are bent towards the β-hairpin at the base of the domain, interacting with K339, R355, H357, N429 and H432 (Guilligay *et al*, 2008; Liu *et al*, 2013); however the structure of a m^7GTP soaked crystal shows the phosphates pointing out towards solvent and only interacting with R332, K339, R355 (Tsurumura *et al*, 2013). Neither of these conformations is necessarily physiologically relevant, as these phosphates may bind differently in the context of an actual capped mRNA (the full cap1 structure being m^7GpppN_m). In the absence of bound m^7GTP, the structure of the domain does not change, with the exception of the N-terminus, at which an extra helix can be seen, partially blocking the m^7GTP binding site. This could provide a hint towards an allosteric mechanism for coordinating promoter and cap binding activities (Cianci *et al*, 1995; Li *et al*, 1998), but the change must be interpreted with caution, as three of the seven residues of the helix come from the vector and are not present in PB2. Guilligay *et al* speculate that the region involved in allosteric regulation could be a prominent loop around residue 424, however this is unchanged in apo- and liganded structures.

Remarkably, the equivalent domain in the PB2 of thogoto virus, despite having a similar structure, is unable to bind cap structures as it lacks many of the important features discussed above: the residues forming the aromatic sandwich, the polar residues that contact the methylated base and the 424-loop (Guilligay *et al*, 2014). Sequence alignments suggest that these will also be missing in the PB2s of the other thogotoviruses: dhori virus and jos virus. Similar observations were also made about the PA "endonuclease domains" of these viruses, rendering the capping mechanism of thogotoviruses even more of a mystery than that of the influenza viruses.

1.3.6 PB2 – C-Terminal Domains

Residues 538 to 759 of IAV PB2 fold into two distinct domains (figure 1.11 (Tarendeau *et al*, 2008; Kuzuhara *et al*, 2009; Yamada *et al*, 2010). The first of these (538 to 677) does not have any known function, but is of interest because this region contains a high density of the host-determining residues of PB2 (Ruigrok *et al*, 2010; Mänz *et al*, 2013), notably residue 627, which underpins the adaptation of most avian IAV polymerases to mammalian hosts (Subbarao *et al*, 1993; Herfst *et al*, 2012; Mok *et al*, 2014; Zhang *et al*, 2014). Like the cap-binding domain, this "627-domain" exhibits a fold unique to orthomyxoviruses: a six-helix bundle covered on one side by five short β-strands (Guilligay *et al*, 2014). The α-helical and β-strand halves of the protein are connected by an extended loop that wraps round the fifth helix of the bundle. Residue 627 is in this loop. In mammalian-adapted IAVs, this residue is usually a lysine, contributing to a basic patch on the surface of the domain (figure 1.13a). Avian-adapted viruses usually select a glutamate at position 627, so disrupting the basic patch (Yamada *et al*, 2010). Partly for this reason, many studies have proposed that the 627-domain interacts with a cellular or viral factor. This factor has not been definitively identified, though numerous suggestions have been made: NP (Labadie *et al*, 2007; Rameix-Welti *et al*, 2009; Ng *et al*, 2012a), promoter RNA (Kuzuhara *et al*, 2009; Paterson *et al*, 2014), importins (Hudjetz & Gabriel, 2012), the DEAD-box RNA helicase DDX17/p72 (Bortz *et al*, 2011), and unspecified activatory (Moncorgé *et al*, 2010) and inhibitory (Mehle & Doudna, 2008) cellular factors. Further supporting the idea that the 627-domain is an interaction domain, is the fact that the G590S and Q591K/R mutations, which affect the shape and charge profile of the basic patch, can allow avian viruses without 627K to replicate in mammalian cells (Mehle & Doudna, 2009; Yamada *et al*, 2010; Liu *et al*, 2012). Nonetheless, other groups have proposed that the actual effect of the 627 mutation is to directly influence temperature-dependent polymerase activity (Massin *et*

al, 2001; Aggarwal *et al*, 2011). The abundance of these competing hypotheses underlines our need for fresh insight into this domain.

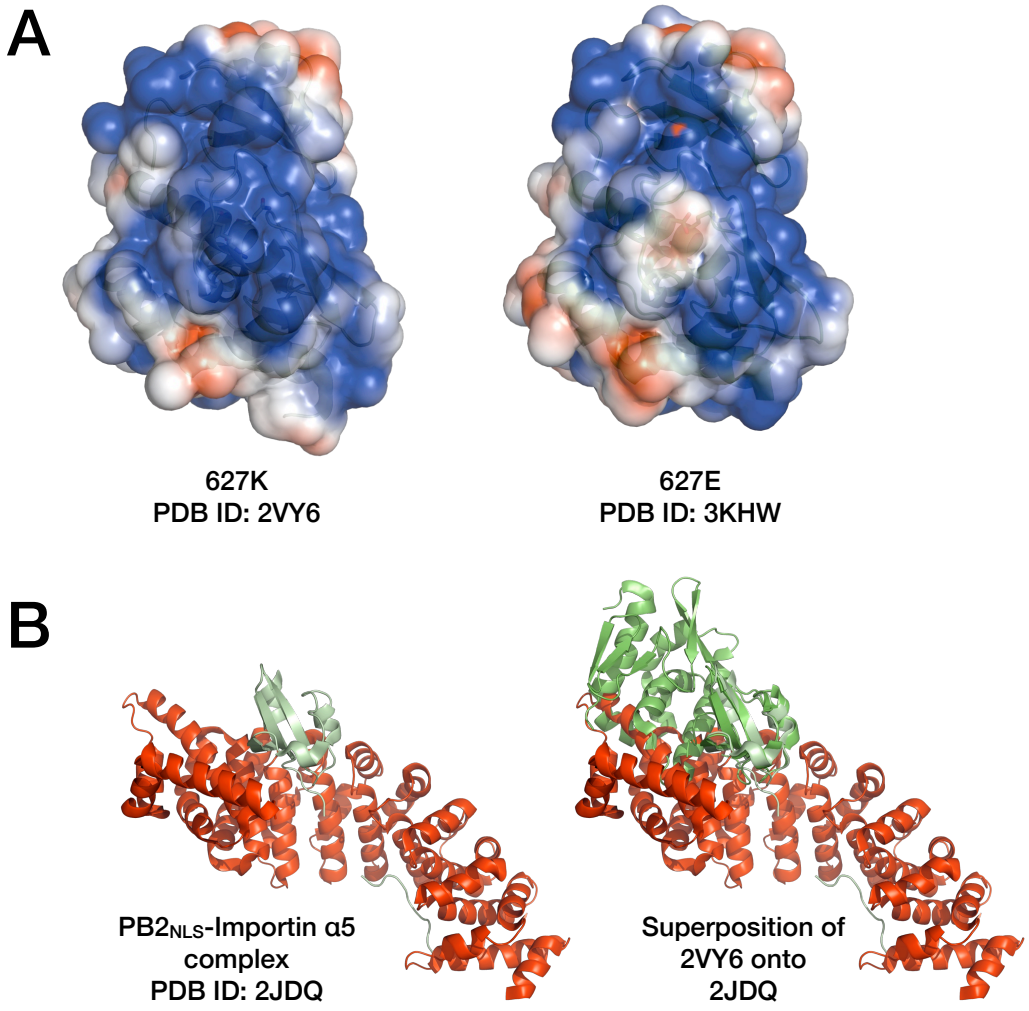
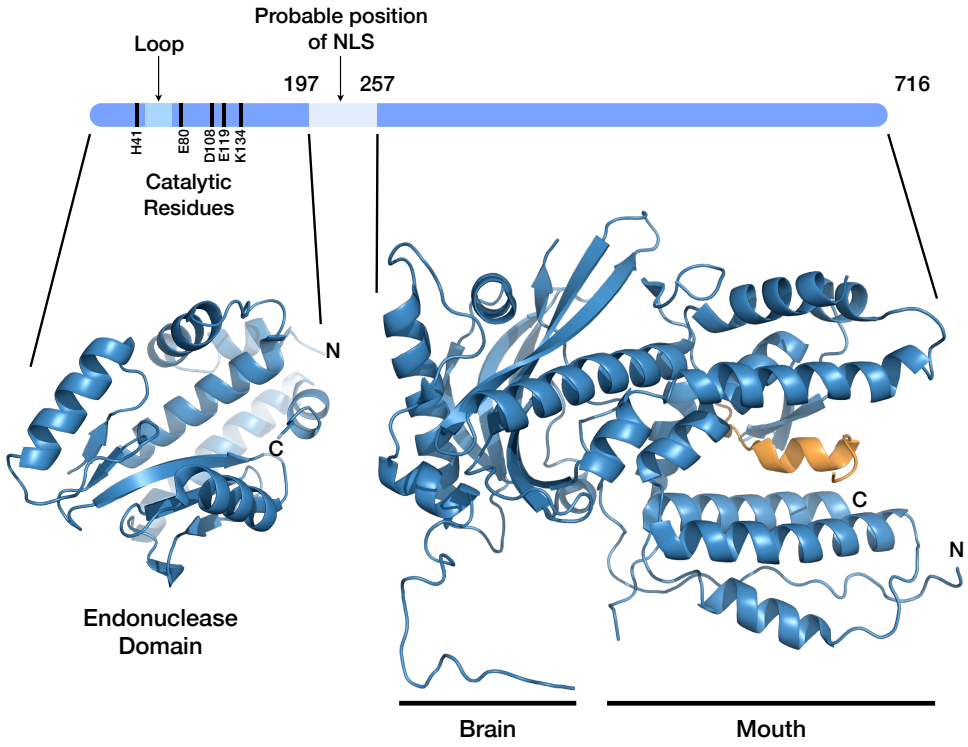


Figure 1.13. PB2 C-terminal domains.
(**A**) Solvent-accessible surface representations of crystal structures of the PB2 627-domain, coloured according to electrostatic potential. These illustrate the disruption to the basic patch on the protein caused by the K627E mutation. (**B**) Left: crystal structure of the NLS domain of PB2 ($PB2_{NLS}$) in complex with importin α5. Right: Superposition of the previous structure with that of the double 627-NLS-domain. Note the clash of the importin and 627-domains.

The second C-terminal domain (residues 692 to 759) was solved by crystallography and nuclear magnetic resonance. It contains a bipartite nuclear localisation sequence (NLS, residues 736 to 739, and 752 to 755) that appears to be conserved in IBV and ICV PB2s (Tarendeau *et al*, 2007). Overall, the domain folds into an α-β arrangement, with two short

α-helices lying either side of a three-stranded antiparallel β-sheet. The first motif of the NLS is on a loop at the C-terminus of the visible domain; the second motif is in a flexible region. Again, this domain has no known structural homologs. The C-terminus of IAV PB2 has also been co-crystallised with the nuclear import factor importin α5 (figure 1.13b) (Tarendeau *et al*, 2007), revealing that the NLS adopts an extended conformation in this context, with the rest of the C-terminal domain remaining unchanged. In contrast, superpositions show that this structure is not compatible with those of the joint 627-NLS domains (Tarendeau *et al*, 2008; Kuzuhara *et al*, 2009), indicating that the linker between these domains must be relatively flexible.

1.3.7 PA – N-terminal Endonuclease




```
A MEDFVRQCFN PMIVELAEKA MKEYGEDPKI ETNKFAAICT HLEVCFMYSD FHFIDERGES 60
B MDTFITRNFQ TTIIQKAKNT MAEFSEDPEL QPAMLFNICV HLEVCYVISD MNFLDEEGKS 60
C MSKTFAEIAE AFLEPEAVRI AKEAVEEYGD HERKIIQIGI HFQVCCMFCD -EYLSTNGSD 59

A IIVESGD-PN ALLKHRFEII EGRDRTMAWT VVNSICNTTG AEKPKFLPDL YDYKENRFIE 119
B YTALEGQGKE QNLRPQYEVI EGMPRTIAWM VQRSLAQEHG IETPKYLADL FDYKTKRFIE 120
C ---------- -----RFVLI EGRKRGTAVS LQNELCKSYD LEPLPFLCDI FDREEKQFVE 104

A IGVTRREVHI YYLEKANKIK SEKTHIHIFP FTGEEMATKA DYTLDEESRA RIKTRLFTIR 179
B VGITKGLADD YFWKKKEKLG -NSMELMIFS YN-QDYSLSN ESSLDEEGKG RVLSRLTELQ 178
C IGITRKADDS YFQSKFGKLG -NSCKIFVFS YD-GRLDKNC EGPMEEQ-KL RIFSFLATAA 161

A QEMASRGLWD SFRQSERGEE TIEERFEITG TMRRLADQSL PPNFSSLENF RAYVDGFEPN 239
B AELSLKNLWQ VLIGEEDVEK GID--FKLGQ TISRLRDISV PAGFSNFEGM RSYIDNIDPK 236
C DFLRKENMFN EIFLPDNEET IIE--MKKGK TFLELRDESV PLPFQTYEQM KDYCEKFKGN 219

A -GCIEGKLSQ MSKEVNARIE PFLKTTPRPL R---LPDGPP CSQRSKFLLM DALKLS---I 292
B -GAIERNLAR MSPLVSATPK KLKWEDLRPI GPHIYNHELP EVPYNAFLLM SD-ELGLANM 294
C PRELASKVSQ MQSNIKLPIK HYEQNKFRQI R---LPKGPM APYTHKFLME EAWMFTK--I 274

A EDPSHEGEGI PLYDAIKCMK TFFGWKEPNI IKPHEKGINP NYLLAWKQVL AELQDIENEE 352
B TEGK---SKK PKTLAKECLE KYSTLRDQTD PILIMKSEKA NENFLWKLWR DCVNTISNEE 351
C SDPERSRAGE ILID------ -FFKKGNLSA IRPKDKPLQG KYPIHYKNLW NQIKAAIAD- 326

A KIPKKKNMKK TSQLKWALGE NMA----PEK VDFEDCKDVS DLKQYDSDEP EPRSLASWIQ 408
B ---MSNELQK TNYAKWATGD GLT----YQK IMKEVAIDDE TMCQEEPKIP NKCRVAAWVQ 404
C ---RTMVINE NDHSEFLGGI GRASKKIPEI SLTQDVITTE GLKQSENKLP EPRSFPRWFN 383

A SEFNKACELT DS-SWIELDE IGEDVAPIEH IASMRRNYFT AEVSHCRATE YIMKGVYINT 467
B TEMNLLSTLT SK-RALDLPE IGPDVAPVEH VGSERRKYFV NEINYCKAST VMMKYVLFHT 463
C AEWMWAIKDS DLTGWVPMAE YPPADNELED YAEHLNKTME GVLQGTNCAR EMGKCILTVG 443

A ALLNASCAAM DDFQLIPMIS KCRTKEGRR- --------KTN LYGFIIKGRS HLRNDTDVVN 519
B SLLNESNASM GKYKVIPITN RVVNEKGES- --------FDM LYGLAVKGQS HLRGDTDVVT 515
C ALMTECRLFP GKIKVVPIYA RSKERKSMQE GLPVPSEMDC LFGICVKSKS HLNKDDGMYT 503

A FVSMEFSLTD PRLEPHKWEK YCVLEIGDML LRTAIGQ--- -VSRPMFLYV RTNGTSKIKM 575
B VVTFEFSSTD PRVDSGKWPK YTVFRIGSLF VSGRE----- ---KSVYLYC RVNGTNKIQM 567
C IITFEFSIRE PNLEKHQ--K YTVFEAGHTT VRMKKGESVI GREVPLYLYC RTTALSKIKN 561

A KWGMEMRRCL LQSLQQIESM IEAESSVKEK DMTKEFFE-- --NKSETWPI GESPKGVEEG 631
B KWGMEARRCL LQSMQQMEAI VEQESSIQGY DMTKACFKGD RVNSPKTFSI GTQEGKLVKG 627
C DWLSKARRCF ITTMDTVETI CLRESAKAEE NLVEKTL--- --NEKQMW-I GKKNGELIAQ 615

A SIGKVCRTLL AKSVFNSLYA SPQLEGFSAE SRRLLLIVQA LRDNLEPGTF DLGGLYEAIE 691
B SFGKALRVIF TKCLMHYVFG NAQLEGFSAE SRRLLLLIQA LKDRKGPWVF DLEGMYSGIE 687
C PLREALRVQL VQQFYFCIYN DSQLEGFCNE QKKILMALEG DKKNKSSFGF NPEGLLEKIE 675

A ECLINDPWVL LNASWFNSFL T-HAL----- ---------R 716
B ECISNNPWVI QSAYWFNEWL G-FEKEGSKV LESVDEIMDE 726
C ECLINNPMCL FMAQRLNELV IEASKR-GAK FF-----KTD 709
```

Figure 1.14. PA structure summary.
(**A**) Solved fragments of PA structure (PDB IDs: 3EBJ and 3CM8), with a selection of other regions annotated. (**B**) PA sequence alignments. See figure 1.7b for details.

PA has proven to be the most structurally amenable polymerase subunit (figure 1.14). The protein can be divided into only two domains, connected by a relatively long and flexible linker (Hara *et al*, 2006; Guu *et al*, 2008). The structures of both the N- and C-terminal domains of IAV PA have been solved by crystallography, with only the flexible linker still outstanding (He *et al*, 2008; Obayashi *et al*, 2008; Dias *et al*, 2009; Yuan *et al*, 2009; Zhao *et al*, 2009; Tefsen *et al*, 2014). The N-terminal domain is a mixture of five β-strands and seven α-helices, in a fold reminiscent of resolvases and type II restriction endonucleases, and containing the characteristic (P)DX_N(D/E)XK active site motif of these enzymes (Dias *et al*, 2009; Yuan *et al*, 2009; Zhao *et al*, 2009; Tefsen *et al*, 2014). Indeed, the recombinant PA N-terminus will bind divalent cations and can cleave RNA (Crépin *et al*, 2010). Thus, this domain is the endonuclease required for cap-snatching (PA_{Endo}). As with the cap-binding domain, the argument as to the location of this endonuclease was only settled when these structures were released; previous suggestions had included sites from all three polymerase subunits (Shi *et al*, 1995; Blok *et al*, 1996; Li *et al*, 2001; Fodor *et al*, 2002; Honda *et al*, 2002; Hara *et al*, 2006). Thogotoviruses are again an exception, as the N-terminus of their PA subunits, despite displaying the same fold, lack key active site residues (Guilligay *et al*, 2014). In IAV PA, these residues are H41, E80, D108, E119 and K134 (figure 1.15a). The first four of these coordinate two catalytic Mg^{2+} ions (Dias *et al*, 2009; Crépin *et al*, 2010; DuBois *et al*, 2012; Kowalinski *et al*, 2012), which presumably participate in the deprotonation of water molecules for nucleophilic attack on phosphodiester bonds. The lysine may be the general base in this deprotonation event (Xiao *et al*, 2014), although it is located in an unusual position in the active site. The active site itself is an acidic cleft in the centre of one face of the domain. Structures of PA_{Endo} bound to soaked nucleotide monophosphates show the substrate in the active site cleft, though the binding site overlaps with one of the proposed catalytic metal ions (Zhao *et al*, 2009). The nucleotide phosphate

can interact directly with K134, and via water molecules with E119, Y130 and K137. H41 also interacts with the ribose and base. These moieties appear to be more flexible than the phosphate. The path of RNA leading to the active site has been modelled by analogy with the binding mode of EcoRV endonuclease and through molecular dynamics simulations (figure 1.15b) (Zhao *et al*, 2009; Xiao *et al*, 2014). These studies suggest that a patch formed by R124, R125 and R192 strongly interacts with the 3′ end of substrate RNA, whereas the 5′ end is more loosely associated. As a side note, early work had identified an NLS in the active site region (residues 124 – 139), however, given the structure, this assignment does not appear to be correct, and the NLS probably lies within the flexible linker region (Nieto *et al*, 1994).

There are striking differences between the initial two PA_{Endo} structures (figure 1.15c) (Dias *et al*, 2009; Yuan *et al*, 2009). Firstly, the number and nature of the active site metal ions differed (one Mg^{2+} as opposed to two Mn^{2+} ions). Later studies have tended to favour a two metal model (Crépin *et al*, 2010; DuBois *et al*, 2012; Kowalinski *et al*, 2012; Xiao *et al*, 2014), with Mg^{2+} being the physiologically relevant ion as its concentration within the cell is 1000x higher than Mn^{2+}. More interestingly, the structure from Alexandre Dias and colleagues, featured a prominent loop between residues 50 to 74 inclusive, which projects outward from one PA_{Endo} and binds into the active site of a crystallographically adjacent molecule. In the crystal form of Yuan *et al*, this loop is flexible and therefore missing from their structure. Deletion of this loop region has no effect on the activity of the isolated PA_{Endo} domain (DuBois *et al*, 2012; Kowalinski *et al*, 2012), so it is unclear what its function is. Notably, this loop is missing in ICV P3.

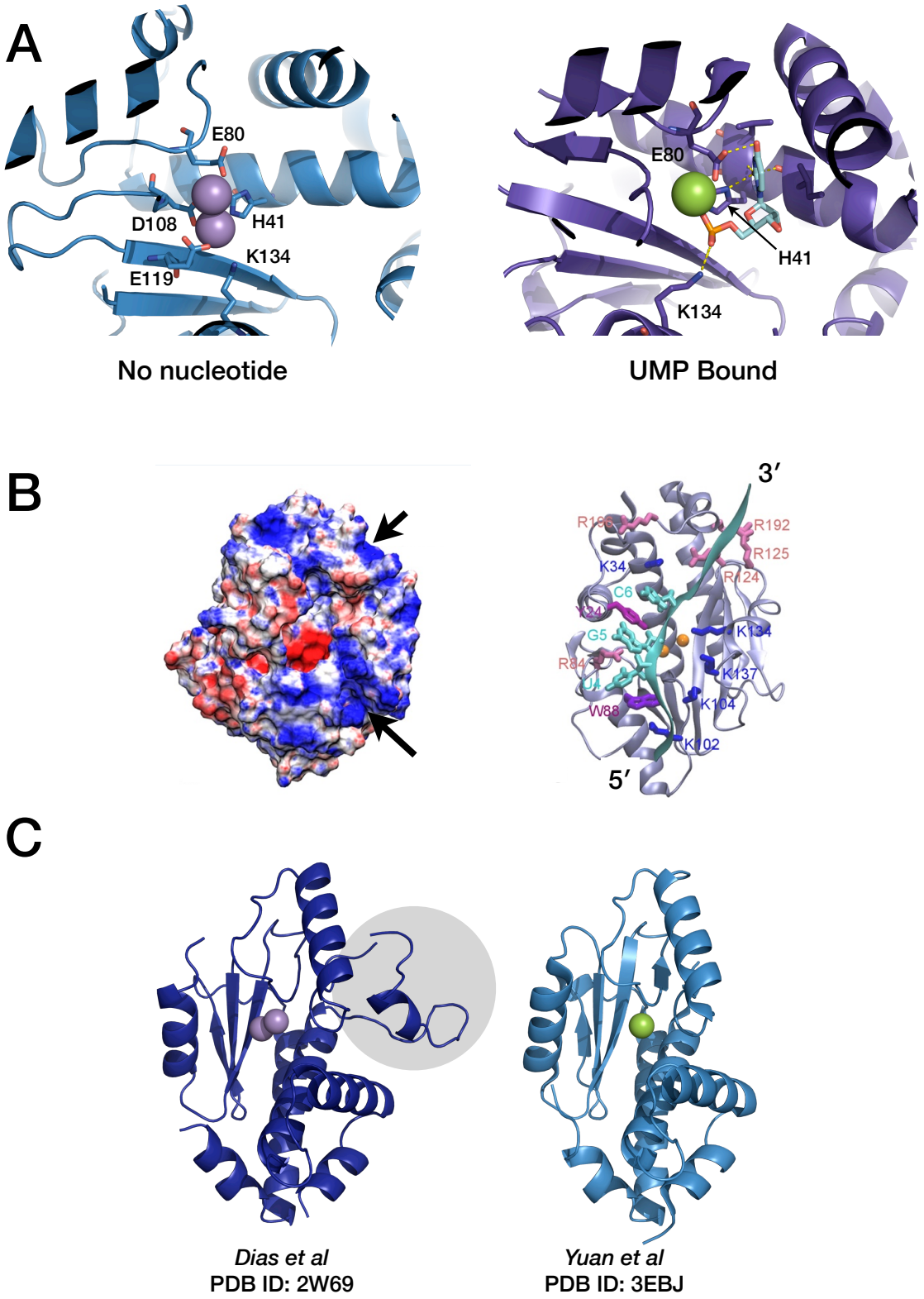


Figure 1.15. PA endonuclease domain.
(**A**) Crystal structures of PA_{Endo} showing the active site with and without UMP bound (PDB IDs 3HW3 and 2W69 respectively). UMP is coloured teal. Catalytically important residues are highlighted. (**B**) Possible path of RNA. Left: surface representation of PA_{Endo}, coloured according to electrostatic charge (±10.0 k_bT/e_c) revealing two basic regions (arrows) that could potentially bind RNA. Right: The modelled RNA path, with interacting residues highlighted. Figure taken from (Xiao *et al*, 2014). (**C**) Crystal structures of PA_{Endo}, solved by two different groups.

1.3.8 PA – C-terminal Domain

The structure of the C-terminal domain of IAV PA (PA_C) in complex with a helix from the N-terminus of PB1 was determined independently by two groups (figure 1.14) (He *et al*, 2008; Obayashi *et al*, 2008). Both structures capture residues 257 – 716 of PA, though there are differing gaps in both. Again, the overall fold is novel. One group colourfully described PA_C as resembling a dragon's head, distinguishing two subdomains: the "brain", consisting of a twisted seven-stranded β-sheet surrounded by five α-helices, and the entirely helical "mouth", biting down on the helix from PB1. Apart from interacting with PB1, any other role of PA_C is unknown. It loosely resembles the N-terminal domain of the reovirus polymerase (figure 1.16a), with a root-mean-square deviation (RMSD) of 4.3 Å over 91 residues (Tao *et al*, 2002; He *et al*, 2008). This N-terminal domain reinforces the contact between the fingers and thumb of the polymerase core, hinting that PA_C could do the same. It has also been suggested that PA_C binds RNA, because it contains a deep and highly basic groove at the base of the brain subdomain, and a basic channel at the intersection of the brain and mouth (figure 1.16b). Mutation experiments with several residues from these features, as well as sequence conservation, provide some support for the above hypothesis. For instance, mutation of the groove residue K539 to alanine, results in specific inhibition of replication (Fodor *et al*, 2002). Similarly, E410A, E524A and the K536A/W537A double mutation (all channel residues except W537) result in defective polymerases (Fodor *et al*, 2002; Regan *et al*, 2006). Moreover, binding of small viral RNAs (svRNAs also known as leRNAs), which match the 5′ ends of genome segments, to a PB1-PA dimer can be disrupted by the R566A mutation and the K328A-K539A double mutation (Perez *et al*, 2012), suggesting that such RNAs interact with these groove-associated residues. This implies that the role of PA_C may be to bind the RNA slightly upstream of the 5′ vRNA promoter, though Perez *et al* suggest instead that PA_C is the site where regulatory

RNAs exert their effect. The K461A and N696 mutations, associated with the channel and groove respectively, have no effect on polymerase activity *in vitro* (Fodor *et al,* 2002).

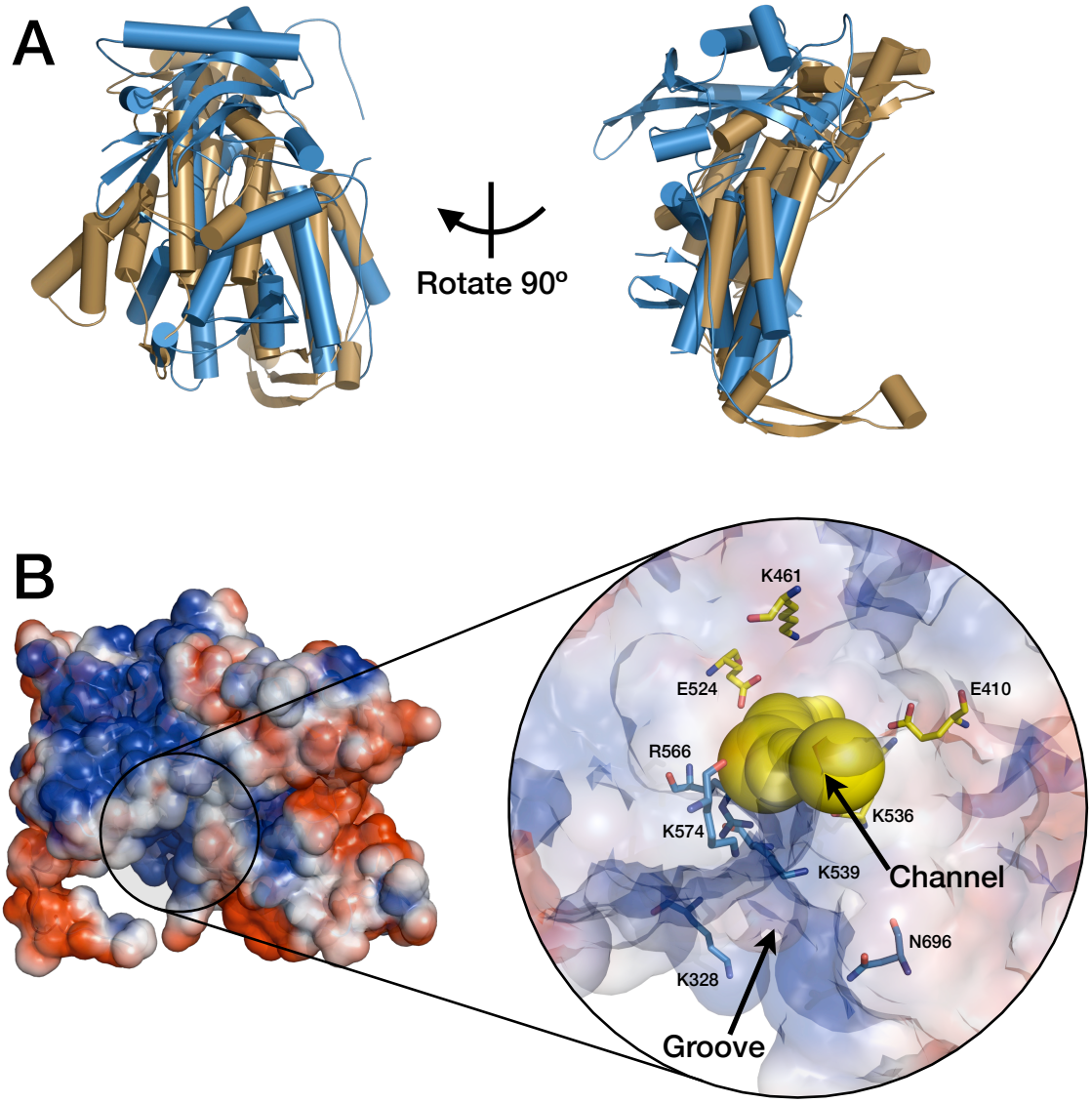


Figure 1.16. C-terminal domain of PA.
(**A**) Two views of the superposition of PA_C (blue, PDB ID: 3CM8) with the N-terminal domain of the reovirus RdRp (brown, PDB ID: 1MUK). (**B**) Surface-electrostatic representation of the crystal structure of PA_C, highlighting the positions of the basic groove and channel (yellow spheres, calculated using *MOLE* (Petrek *et al,* 2007)). Residues lining the groove are shown in blue, whilst those lining the channel are shown in yellow.

1.3.9 Promoter RNA

The influenza polymerase works in the context of an RNP, within which, it will be bound to the 5′ and 3′ ends of the RNA genome segment, known as the promoter. Polymerase function is dependent on correct promoter sequence and the various activities of cap snatching are coordinated with promoter binding (Hagen *et al*, 1994; Cianci *et al*, 1995; Leahy *et al*, 2001a; Lee *et al*, 2003; Velthuis *et al*, 2013). Moreover, differences between the vRNA and cRNA promoters underlie the differences in polymerase initiation from these templates (Honda *et al*, 2001; Deng *et al*, 2006b). Therefore, knowledge about the structure of each bound promoter is a key component of a satisfactory understanding of polymerase mechanism.

The ends of vRNAs are absolutely conserved (13 nt and 12 nt at the 5′ and 3′ ends respectively for IAVs). The promoter is formed from these ends, plus an additional three base pairs (Neumann *et al*, 2004). The various ideas for bound vRNA promoter structure are the panhandle (Tiley *et al*, 1994), fork (Fodor *et al*, 1994) and corkscrew models (figure 1.17) (Flick *et al*, 1996); the latter being the current favourite due to several mutational experiments supporting the intra-strand base-pairing requirements postulated by this model (Flick & Hobom, 1999; Leahy *et al*, 2001a; 2001b; Pritlove *et al*, 1999). Additionally, recent work measuring distances between Förster-resonance-energy-transfer (FRET) pairs in the bound vRNA promoter also corroborates a corkscrew model (Tomescu *et al*, 2014). The cRNA promoter is thought to adopt a similar conformation (Azzeh *et al*, 2001; Crow *et al*, 2004), though differs in the so-called “hinge” region of the corkscrew (Deng *et al*, 2006b). We currently only have definitive evidence for unbound promoter structure (Bae *et al*, 2001; Noble *et al*, 2011). This adopts a duplex conformation.

Precisely how the vRNA and cRNA promoters are bound by the polymerase is still not certain. Crosslinking and mutation experiments have produced inconsistent results. It was initially proposed that the 5′ and 3′ ends of the vRNA promoter bind the thumb and fingers of PB1 respectively (Li *et al*, 1998). Intriguingly, the specific residues implicated in recognising each end (R571 and R572 for the 5′ end and F251 and F254 for the 3′ end) have completely different chemistries, raising the possibility that the two ends are bound in different ways. However, a later study argued against the importance of F251 and F254, as in this case, those residues were not absolutely required for polymerase activity or promoter binding (Jung & Brownlee, 2006). Instead, an additional 5′ promoter binding region was proposed, formed by PB1 residues R233, R238, R239 and R249. Like, F251 and F254, these residues form part of motif F and lie in the fingers subdomain. Additionally, R239 and R249 were said to bind the 5′ promoter in a region that was base-paired with the 3′ strand, while R233 and R238 bind an intra-base-paired region of the promoter corkscrew. Yet another suggestion is that the promoter, as a whole, binds to a domain formed by PB1 residues 1 – 83 and 493 – 757 (González & Ortín, 1999). This would place the binding site near PA_C and the PB1 thumb. One could also speculate that PB1 residues K669, R670 and R672, previously identified as important in viral transcription (Kerry *et al*, 2008), could interact with 5′ vRNA. A potential interaction between PA_C and 5′ vRNA must also be kept in mind (see above). Even less is known about the binding site of the cRNA promoter. One study reports that PA residues 163 – 178 (in the endonuclease domain, towards the linker) are involved, but does not demonstrate a direct interaction (Maier et al, 2008). Clearly, structural information is needed to resolve the situation.

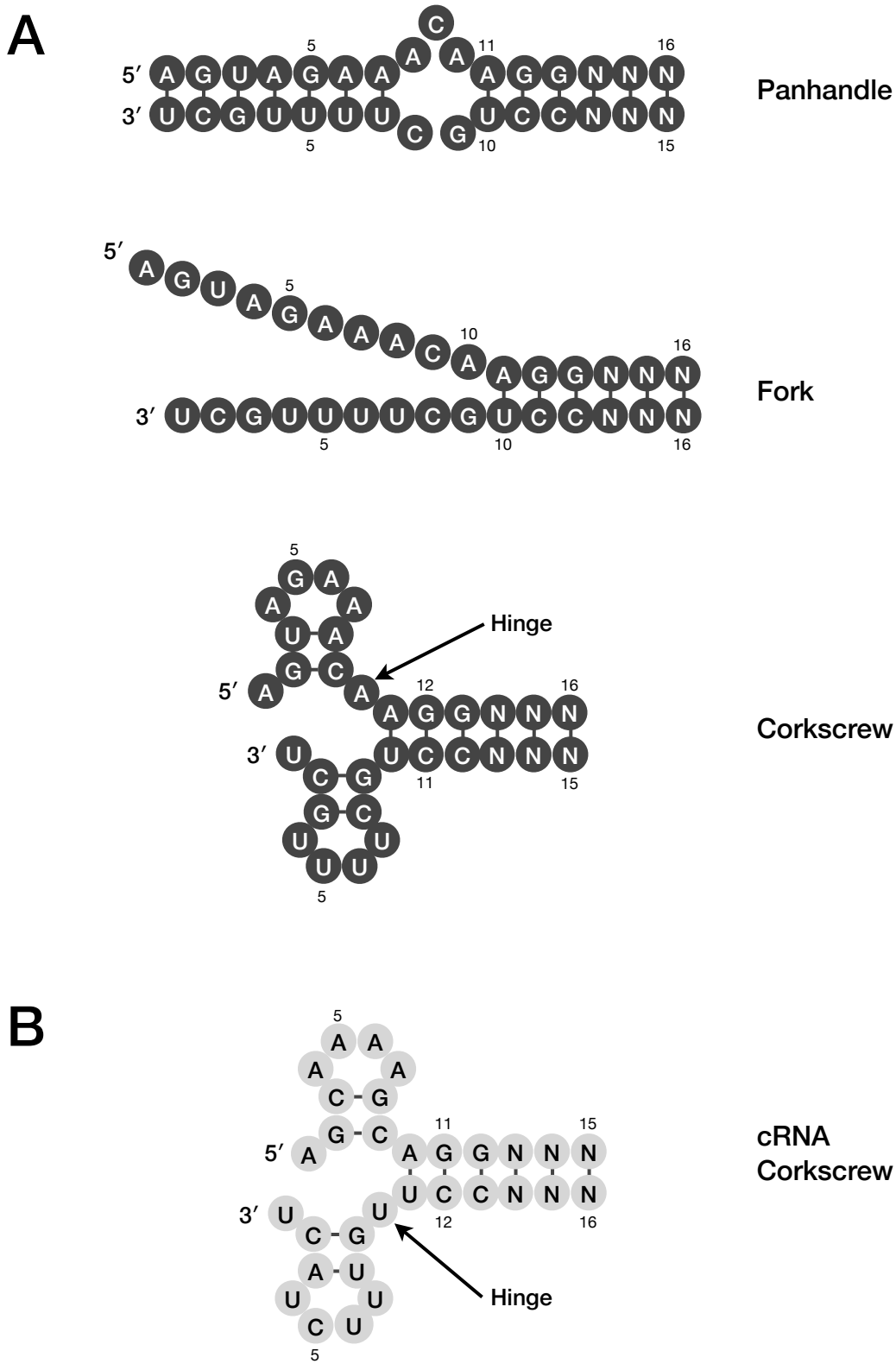


Figure 1.17. Promoter RNA structure.
(**A**) Various models for the structure of the vRNA promoter. (**B**) The cRNA promoter shown in the corkscrew configuration.

1.4 – Influenza Polymerase Mechanism

1.4.1 Primary Transcription Initiation

Having discussed the general biology of influenza viruses and the known structural details about the influenza RdRp, I am now in a position to address the question of how the polymerase actually works. Over the course of this section, I will try to emphasise the unanswered questions we still have about influenza polymerase biology.

An appropriate place to begin seems to be with the events occurring immediately after the arrival of newly infecting vRNPs into the nucleus of a host cell. At this stage, FluPol must exclusively catalyse transcription, as cRNPs and new vRNPs cannot be made in the absence of additional NP (Hay *et al*, 1977; Beaton & Krug, 1986). The trigger to initiate transcription is presumably the availability of capped mRNA within the nucleus. mRNAs are bound at the cap binding domain of PB2 and are then cleaved 10 – 13 nt downstream from the cap by PA_{Endo}. The distance between these two domains is the main determinant of primer length, although intrinsic endonuclease selectivity (preferential cleavage 3′ of guanine residues) (Datta *et al*, 2013) may also be important. After cleavage, the free 3′ end of the primer must move into the PB1 active site at the palm. This corroborates the observation that mRNA 5′ of the endonuclease active site (i.e. the end that must move) is not bound to PA_{Endo} as strongly as mRNA 3′ of it (Xiao *et al*, 2014). However the structural details of primer rearrangement, including whether FluPol itself must undergo any major conformational changes, is unknown. Concurrently, there must also be a reorganisation of the structure of the bound vRNA, because the corkscrew model is not compatible with polymerase elongation. This reorganisation must bring the 3′ strand of the promoter also into the PB1 active site, in a conformation similar to the template strand in other initiation complexes (Butcher *et al*, 2001; Tao *et al*, 2002; Ferrer-Orta *et al*, 2004; Salgado *et al*, 2004).

Again, precisely how this happens is completely unknown. The primers used by FluPol usually have complementarity to the 3′ vRNA promoter (Geerts-Dimitriadou *et al*, 2011b; 2011a), hence in the initiation complex, there are two to three base-pairs between primer and template strands.

1.4.2 Transcription Elongation

The chemistry of elongation and translocation will be the same as for other polymerases, but the larger-scale details are complicated by the involvement of NP. Not only must FluPol be able to use NP-encapsidated RNA, it must ensure that the vRNP is intact after each transcription cycle. This is because new vRNPs cannot be made during primary transcription, so every infecting vRNP needs to template the production of many mRNAs. The closed structure of other RdRps together with the proposed convoluted path of RNA around the vRNP (Arranz *et al*, 2012) means that it is highly unlikely that FluPol can elongate without at least pulling vRNA away from NP. If NP protomers need to be displaced as well, then FluPol must somehow guide these free NPs to rebind into the vRNP once the transcription bubble has passed. Indeed this is what Moeller *et al* speculate in their model for transcription, though there is scant experimental evidence to support this (Moeller *et al*, 2012). The template vRNA must also rejoin the vRNP after passing through the polymerase active site. Additionally, it must be separated from the nascent product strand. The structural basis of this is again not known.

The 5′ end of the vRNA promoter is thought to remain bound to FluPol throughout transcription, because this explains the stuttering which leads to polyadenylation (Poon *et al*, 1999; Zheng *et al*, 1999). However, this requirement makes it hard to visualise how elongation takes place whilst conserving vRNP structure.

A final point regarding elongation is what occurs at the end of each cycle given that the vRNP must be reformed. Does FluPol have to find and rebind the 3′ vRNA promoter, or is this also kept bound at an alternative site throughout transcription? Additionally, what is the structure of the vRNP at this point?

1.4.3 Replication and Secondary Transcription Initiation

Further into infection, as viral proteins are translated, free NP will become available in the nucleus. This will allow new cRNPs and vRNPs to be made (Beaton & Krug, 1986), although it is not clear why cRNA and vRNA are encapsidated by NP, whereas mRNA is not. Replication initiates *de novo*, which means that FluPol will have to provide a "priming loop" or analogous structure to stabilise the NTP at the first position (Butcher *et al*, 2001). Presumably, this structure could not be in the same position for transcription initiation, otherwise it would interfere with primer-dependent initiation. Furthermore, cRNA and vRNA are not polyadenylated, hence during replication, FluPol does not stutter at the 5′ end of the template. The occurrence of 5′ stuttering could be linked to the mode of initiation; e.g. stuttering and polyadenylation could be determined by whether the 5′ vRNA promoter remained bound to FluPol after initiation. Alternatively, stuttering could be entirely independent from initiation. It is difficult to verify this latter proposal, because non-capped but polyadenylated, and capped but non-polyadenylated products, which we would expect FluPol to produce if these processes were independent, are hard to detect in RNP reconstitution assays due to cellular degradation.

The molecular basis for switching between transcription and replication initiation from the vRNA promoter has been contentious. It was initially thought that additional NP affects the conformation of FluPol, directly switching it into replication mode (Beaton &

Krug, 1986; Shapiro & Krug, 1988). However, it is now becoming clear that NP simply stabilises cRNA products that are continually made, but would otherwise be degraded (Mullin *et al*, 2004; Vreede *et al*, 2004; Vreede & Brownlee, 2007; Turrell *et al*, 2013). Instead, a form of switching may be regulated by svRNAs. These are 18 – 27 nt RNAs, which have the same sequence as 5′ vRNA (Perez *et al*, 2010; Umbach *et al*, 2010), and are generated from cRNA template by FluPol under the influence of NEP. svRNAs accumulate to very high levels in the cell, far outnumbering vRNP and cRNP segments, and have been reported to promote full length vRNA production, in a segment specific manner, through a direct interaction with FluPol (Perez *et al*, 2012). However, since svRNAs do not appear to affect cRNA production from vRNA, they still do not explain how FluPol can switch between mRNA and cRNA production from a vRNA template.

Replication initiation from the cRNA promoter appears more complicated. This promoter likely binds in a conformation that prevents FluPol from cap snatching, though no details about this are known. Initiation is again *de novo*, but seems to occur internally via a prime-and-realign mechanism (Deng *et al*, 2006b). In this, an ApG dinucleotide is synthesised from positions 4 and 5 of the 3′ cRNA promoter, and is then realigned with positions 1 and 2 to prime synthesis of the remaining vRNA strand. A separate polymerase, acting *in trans*, may also be required (Jorba *et al*, 2009; York *et al*, 2013), but the molecular reason for this is unclear.

With the generation of new vRNA, NP and FluPol, there are more vRNPs available for mRNA production. FluPol assembles onto the vRNA promoter by binding the 5′ end (probably on a nascent strand as it emerges from a replicating polymerase) followed by the 3′ end. Promoter binding brings about allosteric changes in FluPol. In one model, 5′ vRNA activates cap binding (Cianci *et al*, 1995) and allows the 3′ end to bind; this then activates

the endonuclease domain (Hagen *et al*, 1994; Leahy *et al*, 2001a). In a more recent model, developed from experiments using CA-terminated primers, only 5′ vRNA is required for full FluPol activation; 3′ vRNA is needed simply as a template strand (Rao *et al*, 2003). As the course of infection proceeds still further, so vRNPs are increasingly exported from the nucleus, for packaging into new virions, instead of directing the synthesis of cRNA and mRNA.

1.4.4 Summary

The above discussion should bring home the complexity of the task facing the influenza RdRp. It must:

1. Coordinate cap snatching. This includes binding cellular mRNAs, cleaving their capped ends and then threading the cleaved 3′ end from the endonuclease site into the polymerase active site.
2. Catalyse polymerisation in the context of the RNP. This may require some level of NP disassembly/reassembly. The RNP must be regenerated after each polymerase read-through.
3. Catalyse mRNA polyadenylation by stuttering across a poly-U region on the template. Current theories postulate that this occurs because the polymerase remains bound to the 5′ end of the template during transcription. This is an added complication when thinking about elongation in the context of the RNP.
4. Finish each transcription/replication cycle bound to both ends of the RNA and ready to reinitiate. At the start of an infection, the number of vRNPs is very low, so it seems improbable that the vRNP could reform if the polymerase fell away at the end of a transcription cycle.
5. Initiate in two completely different ways from the same (vRNA) promoter.
6. Bind to two different promoters (vRNA and cRNA), and initiate differently from each one (no mRNA-like molecule is produced from cRNA).

Even though obtaining a molecular structure for the polymerase would not entirely answer all of the questions above, it would greatly clarify the experiments that need to be done to address them. Structural information would also enrich our understanding of intracellular transport of the polymerase subunits, as well as host adaptation. Structures of FluPol with both vRNA and cRNA promoters would precipitate a major advance in our knowledge about the transcription / replication switch.

1.5 Project Aim

The aim of my project was to solve the molecular structure of the complete influenza RdRp, and hence gain insight into its mechanism. The main technique I used to do this was x-ray crystallography. In order to achieve this goal, I developed an expression and purification system that I could reliably use to produce high-quality FluPol for crystallisation trials. I used this to find suitable crystallisation conditions for $FluPol_C$, and then went on to model and refine a 3.6 Å structure for this protein.

Production and Characterisation

2.1 Background

2.1.1 Previous Influenza RdRp Expression Systems

At the time when I was beginning my DPhil, though several groundbreaking structures of fragments of $FluPol_A$ had recently been published, no one had yet reported producing the complete heterotrimeric protein in the quantities necessary for x-ray crystallography (milligram amounts) (Ruigrok *et al*, 2010). This was not for lack of trying. The first reports of FluPol purification date from 1988 (Honda *et al*, 1988; Szewczyk *et al*, 1988), but both are laborious protocols involving isolating and disrupting vRNPs from virus-infected cells. Recombinant FluPol was reportedly first purified in the early 1990s, by using a baculovirus system to express individual subunits, then reconstituting them *in vitro* (Kobayashi *et al*, 1992). However the yield was insufficient to even visualise FluPol on a gel. This was not improved by co-expressing all polymerase subunits, because the purification protocols still required protein denaturation, refolding and precipitation (Shi *et al*, 1995). Over the years, these protocols have been refined, leading to appreciable gains in polymerase yield (Honda *et al*, 2001; 2002; Newcomb *et al*, 2009; Aggarwal *et al*, 2010; Zhang *et al*, 2010).

Other expression systems have been reported, including *Xenopus* oocytes (Digard *et al*, 1989), yeast (Hwang *et al*, 2000) and mammalian cells (Li *et al*, 1998; Deng *et al*, 2005; Torreira *et al*, 2007; Bradel-Tretheway *et al*, 2008; Resa-Infante *et al*, 2010). Systems involving

mammalian cells (either infected with viruses or transfected) are capable of producing polymerase preparations that can easily be seen on silver stained gels. However the yield is still very low, the purity can be uncertain and the scope for scaling-up limited. The authors of the yeast expression system report that they can obtain 23 μg of FluPol from 100 ml of culture (Hwang *et al*, 2000), but the purity is poor and no improved protocols have subsequently been published. FluPol has never been successfully purified from the workhorse of molecular biology: *E. coli.*

2.1.2 Design of Expression System

For these reasons, as well as more general, textbook advantages, such as the presence of eukaryotic chaperones in insect cells, and the relative ease of cell culture and lysis, I decided to use a baculoviral expression system. I hoped that putting together several improvements would allow me to obtain high polymerase yield: using codon-optimised genes, co-expressing a short vRNA, combining all genes into a single baculovirus and employing a protein-A tag for purification. Each of these features, except for the single-virus design, has a basis in published work. The highest yield of FluPol reported for a baculovirus expression system (0.7 mg per litre of starting culture), was obtained using codon-optimised genes (Aggarwal *et al*, 2010; 2011). Our group has shown that FluPol tagged with protein-A, as part of a tandem affinity purification (TAP) tag (Rigaut *et al*, 1999), can be reasonably purified in a single step (Deng *et al*, 2005; 2006a; Kerry *et al*, 2008; Graef *et al*, 2010). Finally, vRNA promoter can stabilise FluPol (Brownlee & Sharps, 2002) and the co-expression of vRNA alongside the polymerase subunits reduces aggregation of purified FluPol (Resa-Infante *et al*, 2010).

It was decided to locate the purification-tag on the C-terminus of PB2. Several reasons contributed to this choice. Firstly, because PB2 is the last subunit to be assembled into FluPol (Fodor & Smith, 2004; Deng *et al*, 2005) placing the affinity-tag on this subunit should make it easier to separate intact polymerase from dissociated polymerase (with the tag on PB2, one would pull-down either isolated PB2 or intact heterotrimer, if PB1 or PA were tagged, one would theoretically pull-down isolated PB1 or PA, the PB1-PA dimer and the intact heterotrimer). Secondly, the C-terminus of PB2 is not the location of a known interaction domain, making it less likely for a tag here to disturb polymerase structure. Additionally, our group has previously verified that a TAP tag in this position does not prevent polymerase activity (Kerry *et al*, 2008). Though this is also true if the tag is placed at the C-terminus of PA (Graef *et al*, 2010), purifications of such FluPol complexes tend to contain excess PA (Aggarwal *et al*, 2011).

The C-terminal TAP tag itself consists of (from N- to C-): a calmodulin-binding peptide, tobacco etch virus (TEV) protease cleavage site and two IgG-binding units of protein A of *Staphyococcus aureus* (protA) (Rigaut *et al*, 1999). Acceptable purification can be achieved using the protein-A unit alone; i.e. the "tandem" element of this tag is not necessary. However, the calmodulin binding peptide (CBP) remains on TAP-tagged proteins post-purification, which could inhibit crystallisation. I therefore decided to also construct PB2 subunits fused with only the protease-site and protein-A elements of the TAP tag.

In order to ensure efficient translation, we wanted to include consensus Kozak sequences (Kozak, 1987) at the start of each of the protein coding genes. These genes would obviously also be bracketed by appropriate restriction sites.

2.1.3 Codon Optimisation

A potential reason for low yield when producing proteins in heterologous hosts is the presence of rare codons in the genes being expressed. Translation of these codons is suboptimal as it requires tRNAs that are not highly abundant in the expression host. The advent of affordable gene synthesis services, and their associated optimisation algorithms, allows for the trivial correction of such codon use mismatches, as well as the avoidance of GC content fluctuation, repetitive sequence (which could cause genetic instability) or unwanted DNA motifs (Raab *et al*, 2010). Such optimisation has proved useful in a number of cases (Burgess-Brown *et al*, 2008; Maertens *et al*, 2010), and as mentioned above, has previously been applied to $FluPol_A$ expression with positive results. Synthetic gene sequences can also be designed such that they already include desirable purification tags, strong translation initiation sequences and restriction enzyme recognition sites; all of which simplifies downstream cloning of the gene. Thus it was decided that all genes used for insect-cell expression in this project would be codon optimised by GeneArt (Life Technologies).

2.1.4 The MultiBac System

For the construction of baculoviruses capable of expressing multiple gene products, I employed the MultiBac system (Bieniossek *et al*, 2008). In this, baculoviruses are generated by transfecting an artificial bacterial chromosome containing the baculovirus genome into insect cells. These chromosomes are called bacmids. Genes of interest, under the control of the baculoviral p10 or polyhedrin promoters (both very-late-stage), can be compatibly inserted into the MultiBac bacmid using *in vivo* Cre-*loxP*-mediated fusion and/or Tn7 transposition. For this, *E. coli* harbouring both the bacmid and a plasmid encoding Cre-recombinase or Tn7-transposase, must be transformed with an appropriate MultiBac

transfer vector containing the gene expression cassette. Multi-gene transfer vectors can be created by conventional ligation-based cloning, or through Cre-*loxP*-mediated fusion of MultiBac "donor" and "acceptor" plasmids; the latter being of particular benefit when working with long genes. This flexibility means a high degree of customisation is possible. A *tour de force* example of the power of the system is the structural work on the 13-subunit anaphase-promoting complex: this was assembled by co-infection with only two MultiBac baculoviruses (Schreiber *et al*, 2011).

The strategy I used to generate baculoviruses is detailed in figure 2.1.

2.1.5 Mini-vRNA Expression Cassette

The co-expression of a mini-vRNA with $FluPol_A$ in human kidney cells has been shown to reduce polymerase aggregation (Resa-Infante *et al*, 2010). Therefore I wanted to see whether this would also be helpful in insect cells.

Expression of such an RNA product requires a specialised system, because vRNA is not capped or polyadenylated. In human cells, a truncated RNA polymerase I promoter can be employed to initiate transcription of vRNA, whilst the production of authentic 3′ ends is achieved by encoding the self-cleaving hepatitis delta virus ribozyme downstream of the vRNA sequence (Pleschka *et al*, 1996). In a baculoviral system, one cannot direct transcription from RNA polymerase I promoters (only baculoviral promoters can be used), hence in the mini-vRNA expression cassette I used, there were self-cleaving ribozymes both upstream and downstream of the vRNA coding region. It was hoped these would direct the production of a 47 nt vRNA. Like that of Resa-Infante *et al*, this vRNA included consensus sequences for both the 5′ and 3′ promoter regions, as well as sufficient connecting nucleotides to form a loop between these ends. Since the expressed FluPol would be used

for crystallography, this loop was kept to a minimum size to minimise its conformational heterogeneity. Additionally, the total mini-vRNA was kept small enough to be replicated in the absence of NP (Turrell *et al*, 2013). This was a requirement because comparable levels of FluPol and mini-vRNA can only be produced if FluPol can also amplify the mini-vRNA (Resa-Infante *et al*, 2010).

2.1.6 Target Polymerases

Since crystallisation is still a relatively unpredictable science, with especially low success rates for large, multi-domain proteins such as the influenza RdRp, I decided to increase my chances of obtaining a favourable outcome by working with several different polymerases. These were chosen to maximise diversity whilst still staying relevant to influenza. Thus, I investigated the polymerases from two strains of IAV as well as IBV, ICV, ISAV, dhori virus and quaranfil virus. IAV is the most medically relevant of these virus species, which is why it accounts for two of the polymerases investigated. The two IAV strains selected represent a well studied human strain (A/Northern Territory/60/1968 (H3N2)) and a highly pathogenic avian strain (A(duck)/Fujian/1/2002 (H5N1)). The latter was chosen because avian influenza viruses replicate at higher temperatures (41ºC) than their human counterparts (33ºC), hence we reasoned that the polymerase from such a virus might be more thermostable and thus a better crystallisation candidate. The A/Fujian polymerase is particularly active in *in vitro* assays (Leung *et al*, 2010).

The polymerases from other species of *Orthomyxoviridae* are still expected to share the same fold and organisation as influenza polymerases, despite the very low sequence similarity between them. This has been proven for the cap-binding and endonuclease domains of Thogotovirus polymerases (Guilligay *et al*, 2014). Their polymerisation

mechanisms should be similar, though the process of cap-snatching in these polymerases may well differ. Nevertheless, the structure of any of these polymerases would be very informative about that of IAV polymerase, in addition to being enormously interesting in its own right.

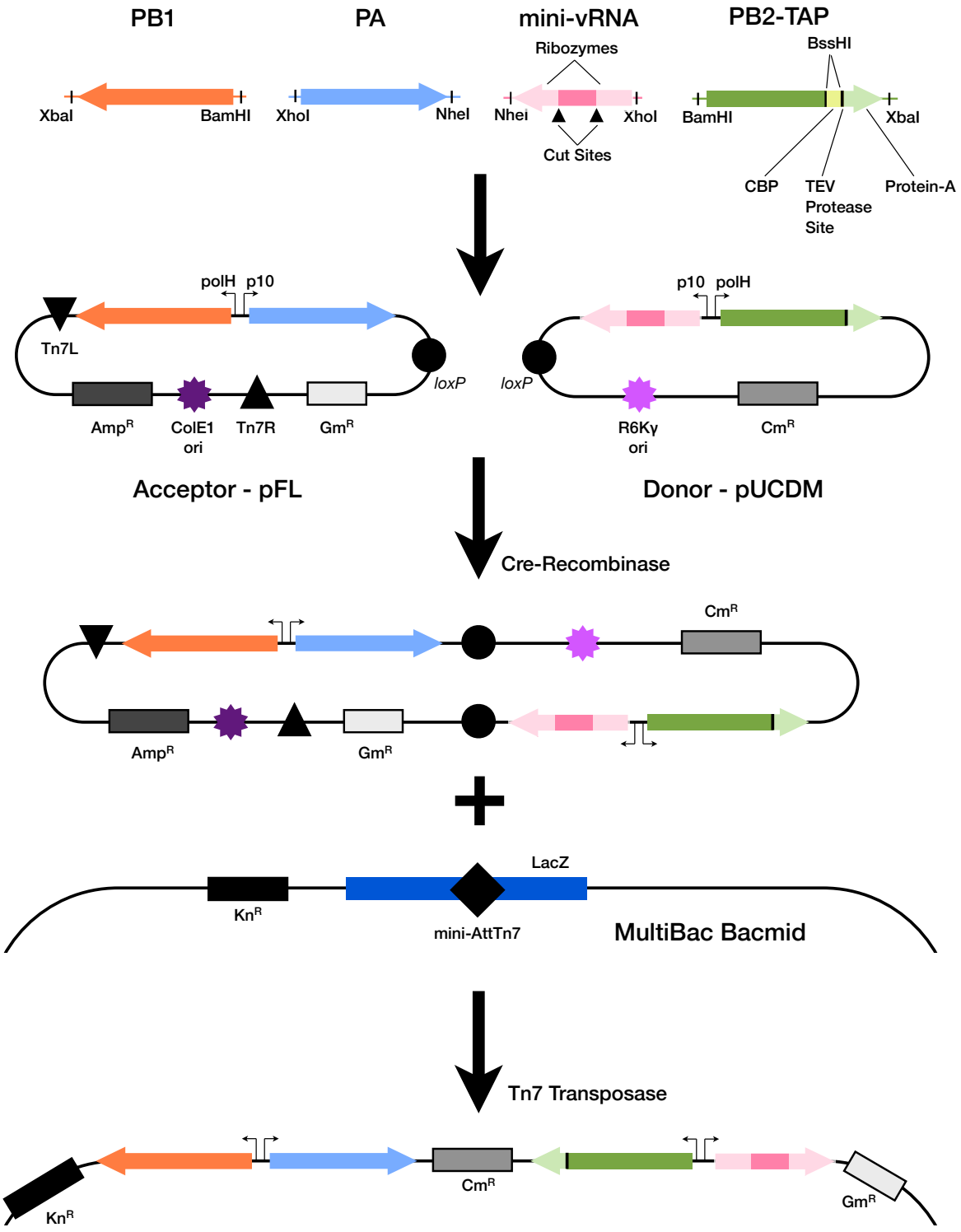


Figure 2.1. MultiBac baculovirus construction.
Top: summary of the genes ordered, showing the restriction sites flanking each gene as well as the design of the tagged PB2 and mini-vRNA expression module. Below: outline of the cloning of these genes into the MultiBac Bacmid. For clarity, some elements are omitted from the diagram of the final integrated product.

2.2 Results

2.2.1 Construction of Bacmids Containing Polymerase Genes

All components of the MultiBac system (plasmids and specific *E. coli* cell lines) were donated to us by Dr Imre Berger. Synthetic genes encoding FluPol were purchased from GeneArt and were supplied in generic cloning vectors. It is worth mentioning that for PB2, I initially worked with a sequence that encoded the full TAP tag at the protein's C-terminus, but which also contained two BssHI sites flanking the CBP element of this tag. This allowed me to easily remove the CBP, so providing me with flexibility about its inclusion. Later synthetic genes did not include CBP, because I found it unnecessary to add an extra affinity purification step after that involving protein-A.

I used conventional molecular cloning techniques to transfer polymerase genes into MultiBac acceptor (pFL) and donor (pUCDM) plasmids, as outlined in figure 2.1. Originally, I planned to clone two genes into each of these plasmids, however early results indicated that mini-vRNA expression was not working (see below), so I did not include this element in most of the constructs I produced. Both acceptor and donor plasmids contained *loxP* sites, which meant that I could fuse these together using Cre recombinase. Indeed, this was how I generated transfer vectors containing all of the desired polymerase genes. I verified each stage of construction by sequencing. Finally, I transformed *E. coli* cells containing the MultiBac bacmid, with each transfer vector. These cells express the Tn7 transposase, which catalyses insertion of the polymerase genes into a mini-attTn7 site, located within a lacZ gene on the bacmid. I used both blue/white selection and PCR to identify clones in which this reaction had successfully occurred. A selection of baculoviruses made is shown in table 2.1.

Table 2.1. Summary of polymerases investigated.
In all cases, the tag is on the C-terminus of PB2. Hence FluPol-TAP indicates PB1, PB2-TAP, PA.

Virus	Abbreviation	Baculoviruses Made	Protein purified?
A/Northern Territory/ 60/1968 (H3N2)	A/NT	FluPol-TAP + mini-vRNA	Yes
		FluPol-protA + mini-vRNA	
		FluPol-TAP	
		FluPol-protA	
A(duck)/Fujian/ 1/2002 (H5N1)	A/Fj	FluPol-protA	Yes
B/Panama45/1990	B	FluPol-protA	Yes
C/Johannesburg/ 1/1966	C	FluPol-protA	Yes
Infectious Salmon Anaemia (taxon: 55987)	ISA	ISAPol-protA	No
Quaranfil (taxon: 688436)	Qu	QuPol-protA	No

2.2.2 Baculovirus Rescue and Expression Verification

Recovery of polymerase-expressing baculoviruses simply involved transfecting bacmid DNA into Sf9 insect cells. I initially determined whether virus had been successfully rescued by testing whether initial V_0 stocks could arrest the growth of healthy Sf9 cells. The V_0 titres were variable, and arrest took between two to seven days. I then confirmed viral rescue by looking for polymerase expression in cells infected with V_0, using western blotting (figure 2.2a, c, d and e). I was able to produce baculoviruses for all of the polymerases I set out to investigate, except that of Dhori virus, for which I recovered baculovirus, but could not verify protein expression. Expression levels of FluPol were often even high enough to easily distinguish baculovirus-infected from uninfected cell-lysate samples on Coomassie-stained SDS-PAGE gels: the polymerase subunits were prominently stained (figure 2.2b). This was already a positive sign that the designed improvements in this system would lead to higher yields over previously published FluPol-production protocols. A caveat was that FluPol could be seen in both soluble and insoluble fractions of the cell lysate (figure 2.2d and e). However, even with losses due to precipitation inside the cell, I could still produce enough soluble FluPol to proceed with purification.

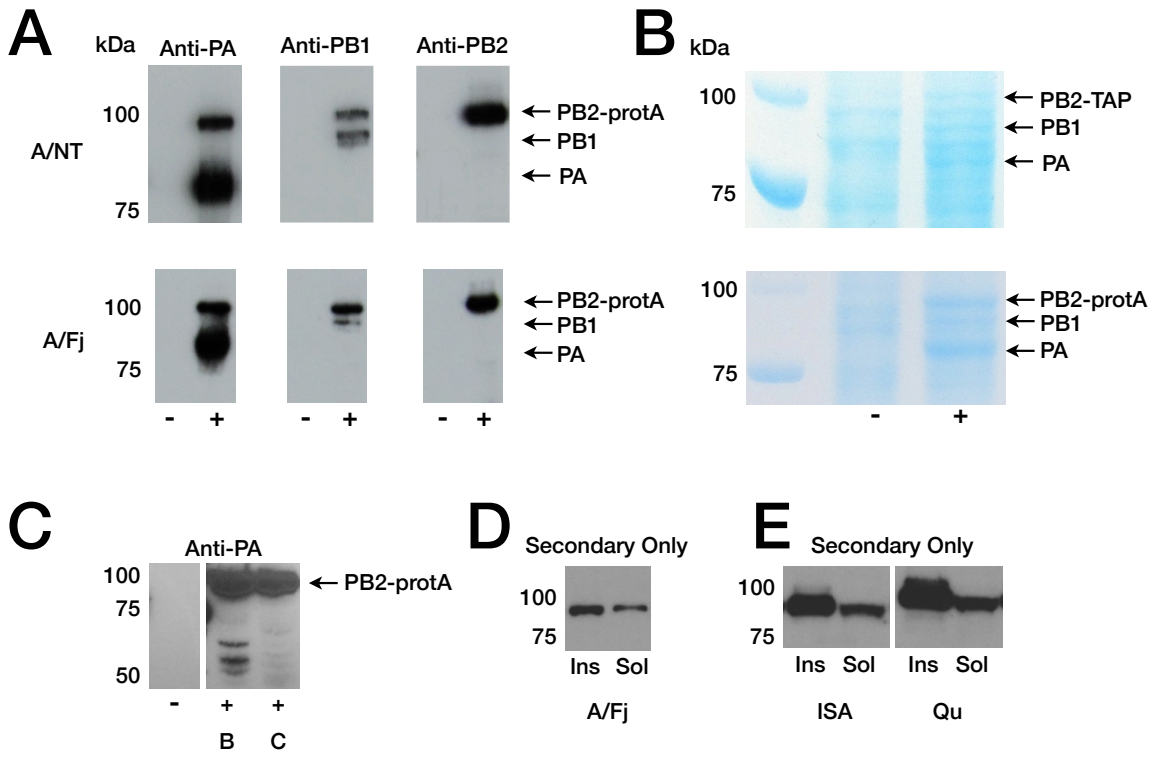


Figure 2.2. Polymerase expression verification.
(A and B) Expression of IAV polymerases. **(A)** Western blots of total cell lysate from Sf9 cells infected with polymerase-expressing V_0 baculoviruses (+) or left uninfected (-). The blots are probed with the indicated antibodies. PB2-protA is detected in all blots because protein-A binds all IgG proteins. **(B)** Coomassie-stained SDS-PAGE gels of similar cell lysates. Bands for each polymerase subunit are clearly visible. **(C – E)** Western blots of total, insoluble (ins) or soluble (sol) cell lysates of cells infected with the indicated polymerase-expressing baculoviruses and stained with the indicated antibodies. In all cases, only PB2-protA is detected (anti-PA does not bind the PA of IBV and ICV).

2.2.3 Evaluation of Mini-vRNA

Mini-vRNA co-expression was designed to reduce polymerase aggregation. However for this to work, the RNA transcript would have to be properly processed by the ribozymes at both of its ends and the mini-vRNA would have to be replicated by FluPol. To test whether this was occurring, I performed a primer extension assay with samples from cells infected with polymerase expressing baculoviruses. This assay uses radiolabelled primers to detect the products of transcription and replication. This assay revealed that Sf9 cells cells infected with a MultiBac virus expressing $FluPol_{A/NT}$ with the mini-vRNA, did

not contain large amounts of correctly processed vRNA (figure 2.3a). A prominent slower-migrating band is visible in the samples from cells infected for 60 h, but this was not characterised further. The same mini-vRNA (together with both ribozymes) was functional in mammalian cells; a product of the correct size is clearly visible in cells transfected with plasmids encoding $FluPol_{A/NT}$ and the mini-vRNA cassette (figure 2.3b). The reason the system did not work in insect cells might have been that FluPol was not transported to the nucleus, or that the mini-vRNA cassette was not expressed or processed properly. It is unlikely that FluPol was simply not folded correctly, because *in vitro* assays with purified FluPol showed that it was active (see section 2.2.6).

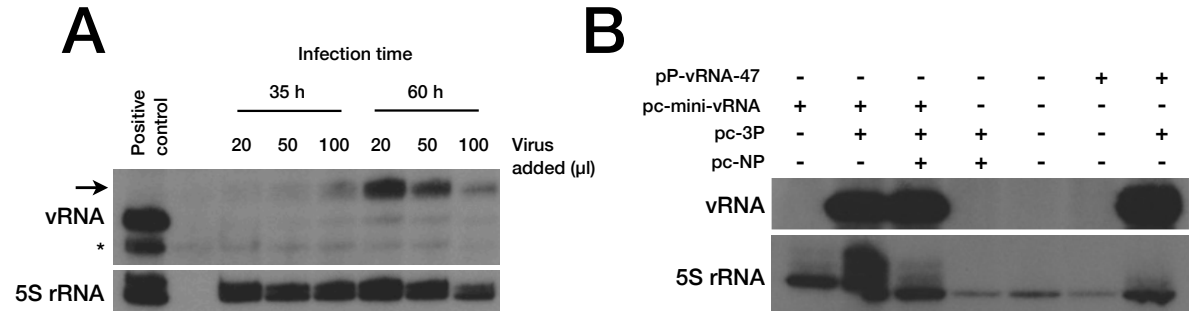


Figure 2.3. Evaluation of mini-vRNA.
(A) Primer extension assay, probing for FluPol replication products from RNA of Sf9 cells infected with the indicated volume of MultiBac $FluPol_{A/NT}$-protA + mini-vRNA V_1 for the indicated amount of time. The positive control is the same sample as lane 2 of (b). 5S rRNA was used as a loading control. The products of the primer extension reaction were separated by PAGE and visualised by autoradiography. Bands corresponding to vRNA (42 nt) and 5S rRNA (100 nt) were identified based on their relative electrophoretic mobility. The arrow indicates an unknown product specific to infected Sf9 cells, while the asterisk indicates a non-specific background product. **(B)** Primer extension assay probing for FluPol replication products from RNA of HEK293T cells transfected with the indicated plasmids. pP-vRNA-47 was used as a positive control, as it expresses the 47 nt RNA (from a pol-I promoter) that matches the desired fully processed product of the mini-vRNA expression cassette. pc-mini-vRNA contains this cassette downstream of the pol-II promoter of the pcDNA3a plasmid. pc-FluPol and pc-NP are separate pcDNA3a plasmids expressing PB1, PB2, PA and NP. Again, the products were identified based on their relative electrophoretic mobility.

On top of this, co-expression of the mini-vRNA module did not reduce aggregation of the polymerase seen in initial purifications. Since, the mini-vRNA was of no discernible benefit, but could still potentially be binding to the polymerase in low amounts (which

would lead to undesirable heterogeneity in the purified polymerase), I subsequently decided to work only with baculoviruses that did not contain the mini-vRNA module.

2.2.4 Influenza Polymerase Purification

The FluPol that I was expressing was tagged with protA on the C-terminus of PB2. This tag contains two high-affinity IgG-binding units, and hence allows the purification of FluPol using IgG-coated Sepharose beads. Eluting FluPol from the beads required cleavage of the tag using AcTEV protease. Purifications with the protein-A tag proceeded well for the influenza virus polymerases and gave high purity FluPol from only one purification step (figure 2.4a). However I was not able to make this work for polymerases from quaranfil virus and ISAV. I did not continue experiments with these last two polymerases.

I developed my own purification protocol over the course of my DPhil. This differed from established protocols (Bradel-Tretheway *et al*, 2008) in three important respects. Firstly, I found that the addition of ribonuclease A (RNase A) during cell lysis was essential to reduce aggregation of FluPol. Gel filtration and A_{260}/A_{280} readings indicated that if RNase had not been added, then FluPol would be purified as large RNA-protein aggregates. Secondly, I discovered that FluPol was only soluble at high concentrations in solutions that had a high ionic strength. Thus, I used buffers containing at least 0.5 M NaCl throughout the purification (compared to around 0.2 M in established protocols). Finally, I included a gel filtration step after the initial IgG pull-down step. This allowed me to remove the AcTEV protease added for elution, as well as separating out aggregated FluPol. Thus I could ensure that I was working with homogenous protein when setting up crystallisation trials.

Figure 2.4 shows the material obtained after each stage of a typical, optimised protein purification, as well as profiles of the gel filtration step. The purity of the final FluPol preparations were very high; certainly over 95%, judging by the staining intensities of FluPol compared to impurities. The highest final yields of FluPol were 3 – 4 mg per litre of cell culture.

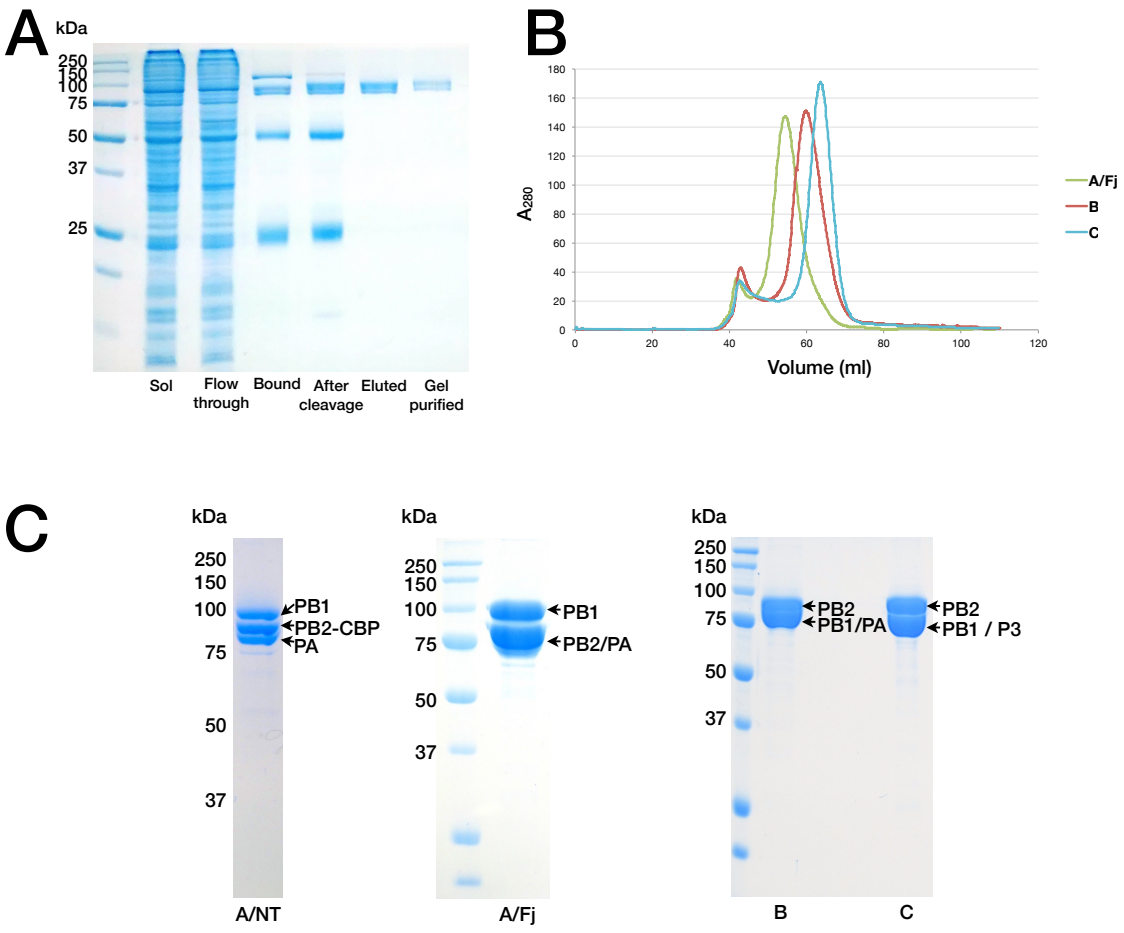


Figure 2.4. FluPol purification.
(A) Coomassie-stained SDS-PAGE analysis of the purification of FluPol$_C$. From left to right, the samples loaded are: soluble infected-cell lysate (sol), protein that did not bind to IgG-Sepharose (flow through), IgG-Sepharose with protein bound (bound), IgG-Sepharose with cleaved protein (after cleavage), protein eluted from IgG-Sepharose (eluted) and protein after gel-filtration (gel purified). **(B)** Gel-filtration profiles of the indicated polymerases. Each of these eluted in one major peak. The profile of FluPol$_{A/NT}$ (not shown) is indistinguishable from that of FluPol$_{A/Fj}$. **(C)** Coomassie-stained SDS-PAGE gels of the indicated purified polymerases.

2.2.5 Effect of Promoter RNA

The elution profile of $FluPol_{C/JHB}$ during gel filtration matched that expected for an isolated polymerase heterotrimer, and that of $FluPol_{B}$ was also close to this (figure 2.4b). However, this was not the case for $FluPol_{A/NT}$ and $FluPol_{A/Fj}$, which eluted at lower volumes, hence were aggregating into higher-order oligomers. This behaviour has previously been observed (Jorba *et al*, 2008), and the solution in that case had been to attempt to mimic the environment of a vRNP by co-expressing a mini-vRNA alongside the polymerase (Resa-Infante *et al*, 2010). Since this strategy had not worked in insect cells, I decided instead to investigate whether addition of RNA during the purification could produce the same beneficial effect. I used 15 nt and 14 nt RNA oligonucleotides representing the 5′ and 3′ ends respectively of the vRNA promoter. These were added after elution of FluPol from the IgG-Sepharose, but before gel-filtration. I found that the procedure which gave the best results was mixing RNA and protein in the presence of 2 M NaCl, then dialysing the salt concentration back down to 0.5 M overnight. The effect of this might have been to reduce binding of RNA to lower affinity off-target sites. The addition of promoter RNA had a marked affect on the elution profile of FluPol, with both ends together causing the polymerase to become monodisperse heterotrimer, as desired (figure 2.5a). This was confirmed by multi-angle light scattering (MALS) experiments (figure 2.5b). The A_{260}/ A_{280} ratio of purified FluPol after RNA-binding was 0.95 – 1.0, which further indicated that RNA was indeed bound (comprising ~5% of the total molecular weight of the complex, compared to a theoretical value of 3.5%).

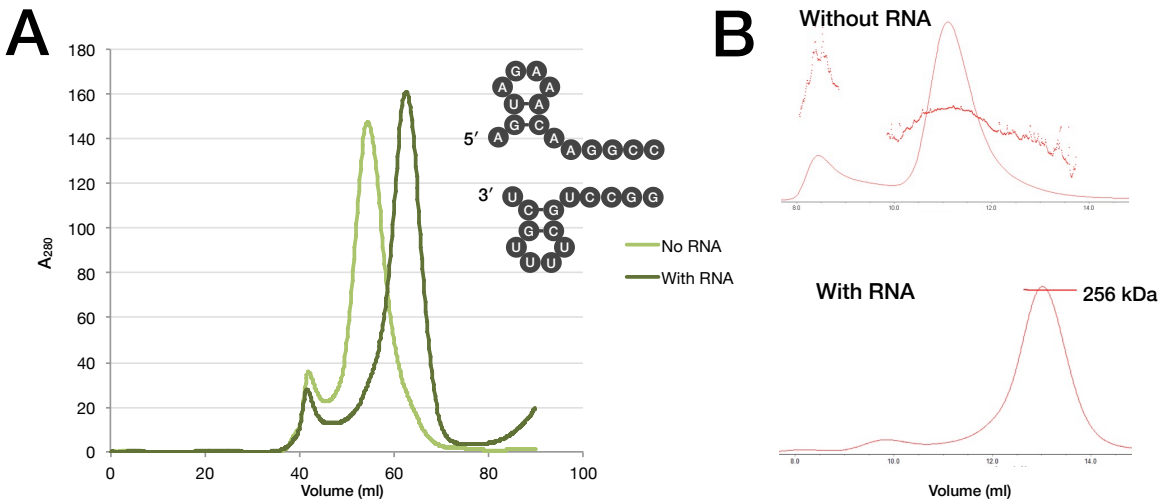


Figure 2.5. Effect of promoter RNA on FluPol$_A$.
(A) Gel-filtration profiles of FluPol$_{A/Fj}$ either on its own or bound to RNAs representing both strands of the vRNA promoter (inset). **(B)** MALS analysis of FluPol$_{A/Fj}$ on its own or bound to the above RNAs. The continuous red trace indicates light scattering, whilst the other red lines indicate molecular mass. The average mass of FluPol$_{A/Fj}$ without bound RNA is 756 kDa, which would correspond to a trimer of trimers, however the shape of the mass profile indicates that aggregation is concentration dependent.

2.2.6 Polymerase Activity

I verified that the polymerase I was producing was intact and correctly folded, by performing activity assays with the purified protein. This was done with $FluPol_{A/NT}$ and $FluPol_C$ (figure 2.6). In these assays, polymerase activity was stimulated in the presence of GTP [α-^{32}P], which allows the products of polymerisation to be visualised by autoradiography. I used two different primers in the assays: ApG dinucleotide, which simply tests for polymerisation; and β-globin mRNA, which additionally probes for functional cap-snatching machinery. $FluPol_{A/NT}$ was fully functional, however $FluPol_C$ did not appear to catalyse cap-snatching. The reason for this is unclear, however is more likely to reflect different reaction condition requirements or specific issues with the purified polymerase rather than genuine lack of cap-snatching activity. This is because mutational experiments on conserved, functional residues (performed by Itziar Serna Martin) confirmed that $FluPol_C$ carried out cap-snatching and polymerisation in the same way as $FluPol_A$. Alternatively, the levels of cap-primed product with $FluPol_C$ were simply below the detection threshold, although this would not explain the lack of primer-independent product. To resolve this issue, further experiments would have to be performed, with size markers included. Despite the inability of $FluPol_C$ to catalyse cap-snatching in this assay, the result still indicated that the purified polymerase was an intact heterotrimer.

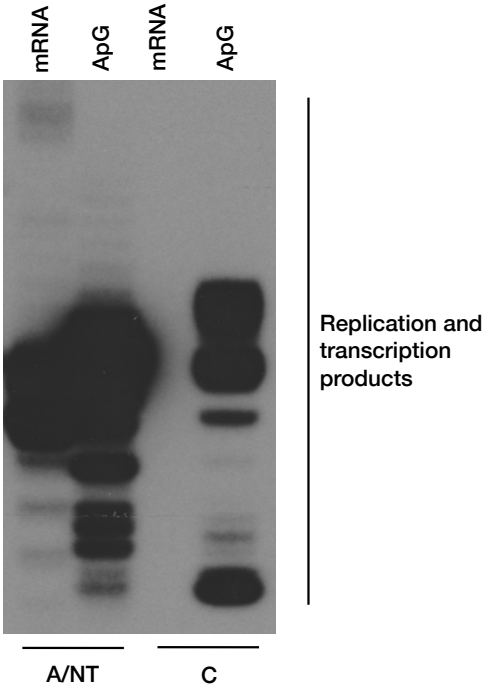


Figure 2.6. *In vitro* polymerase activity assays.
Activity assays were performed with purified FluPol from the indicated viruses. The polymerases were primed with either ApG dinucleotide or beta-globin mRNA. The products of the reactions were separated by PAGE and visualised by autoradiography.

2.2.7 Fluorescence Anisotropy

In order to gain a more quantitative view on the binding of FluPol to promoter RNA, I developed a fluorescence anisotropy assay to measure the affinities of these interactions. Fluorescence anisotropy is related to the degree of polarisation of light emitted from a fluorophore. If a population of static fluorophores is excited by plane polarised light, then the emitted light will also be polarised and the fluorescence anisotropy will be high. In contrast, if the fluorophores can move during the time between excitation and emission, then polarisation, and hence anisotropy, of emitted light will be diminished. Changes in fluorescence anisotropy are therefore related to changes in the tumbling rate of a fluorophore. In this assay, I attached the ATTO647N fluorophore to RNA oligomers representing different promoter strands, and mixed these with varying concentrations of purified FluPol. Binding of FluPol to an RNA causes it to tumble more slowly, so causing an increase in fluorescence anisotropy.

I applied this assay to IAV FluPol, vRNA and cRNA. The results are shown in figure 2.7. I found that the strongest affinity interaction was with the 5′ strand of the vRNA promoter. For $FluPol_{A/NT}$ with 5′ vRNA, my results showed that the K_D was approximately 5 – 10 nM. Fitting of similar results to an 1:1 binding-model gave a K_D estimate of 2 nM (Tomescu *et al*, 2014). I observed no difference between $FluPol_{A/NT}$ and $FluPol_{A/Fj}$. Experiments with the other promoter strands were performed with this latter polymerase. The vRNA 3′ strand bound $FluPol_{A/Fj}$ approximately 10x weaker than the 5′ strand, and with a similar affinity to that of the cRNA 5′ strand. No interaction at all was observed between $FluPol_{A/Fj}$ and the cRNA 3′ strand.

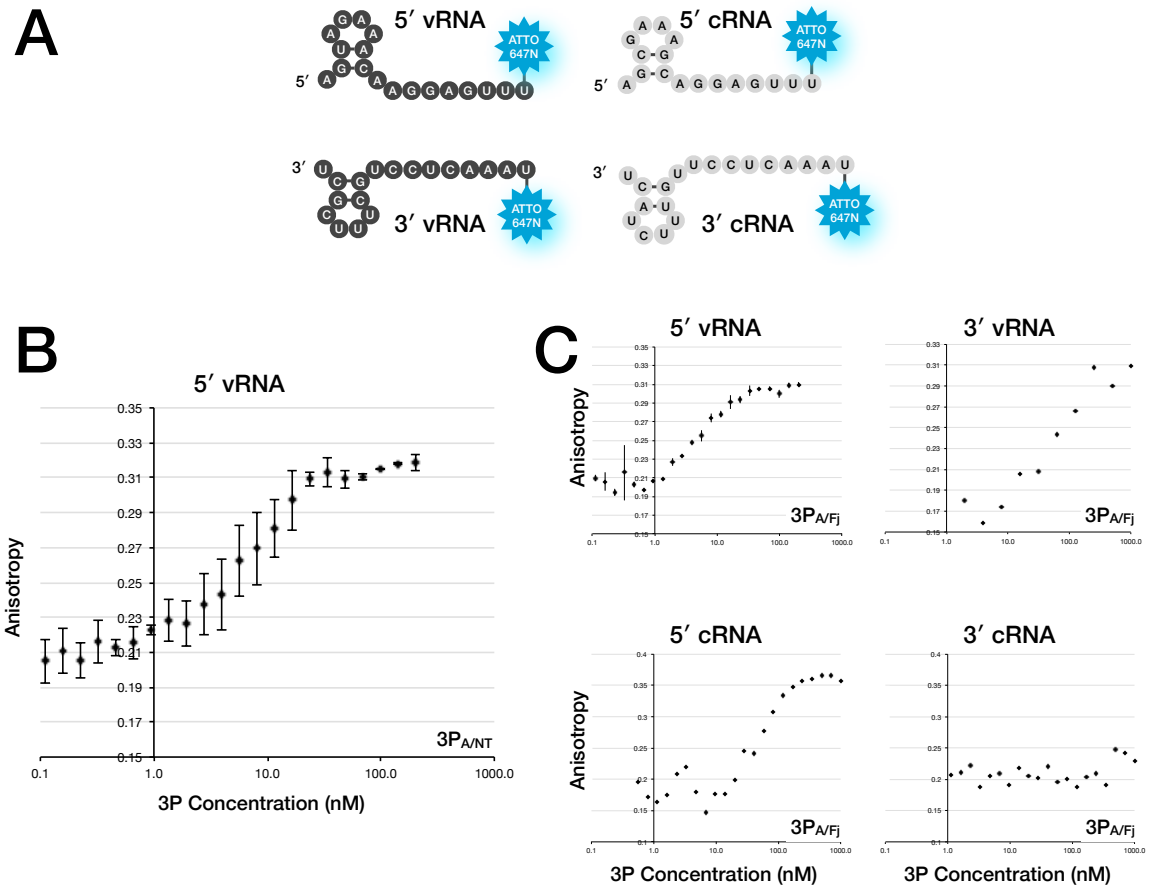


Figure 2.7. Fluorescence-anisotropy assays of promoter binding.
(A) Representations of the labelled RNAs used in the anisotropy experiments. **(B)** Plot showing the fluorescence anisotropy of 1 nM labelled 5′ vRNA with varying concentrations of FluPol$_{A/NT}$. Error bars indicate the standard deviation (n=3). **(C)** Graphs of the fluorescence anisotropy of 1 nM of the indicated RNAs with varying concentrations of FluPol$_{A/Fj}$. This assay was performed twice with 5′ vRNA (error bars indicate the range), but only once with the other RNAs.

2.2.8 Electron Microscopy

Alongside my crystallography efforts, I investigated the use of electron microscopy to study the high-resolution structure of FluPol. EM reconstructions have previously been published for $FluPol_A$, although the detail that can be resolved in these is still modest. No EM work with $FluPol_B$ or $FluPol_C$ has been reported. First off, I looked at negatively stained polymerase particles, in order to quickly determine the suitability of the sample for EM and to help hone optimal sample preparation conditions. As expected from gel-filtration and MALS, preparations of $FluPol_A$ without added RNA appeared to consist of a heterogeneous mix of elongated particles, mostly larger than the polymerase heterotrimer, so suggesting higher-order oligomers, and unsuitable for further structural analysis (figure 2.8a). This contrasted with isolated $FluPol_C$, which appeared predominantly as nearly circular particles probably corresponding to isolated heterotrimer (figure 2.8c). However, there was still evidence for some polymerase aggregation as I could often see two particles lying next to each other, even in relatively sparse areas of the grid. Additionally, the particles were rather heterogeneous in size, ranging between approximately 10 – 20 nm in diameter. This was somewhat concerning, given that it raised the possibility of significant sample contamination (perhaps with free PB2), though no evidence for this was seen from gel-filtration or SDS-PAGE. Alternatively, this heterogeneity could reflect the polymerase having been captured in different orientations on the grid, and I was perhaps seeing end-on and side-on views of an elongated particle. Similar particles were observed for $FluPol_{A/Fj}$ bound to a 37 nt RNA containing both strands of the vRNA promoter, and in the presence of cap analogue (m^7GpppG, figure 2.8b). The most homogeneous particles of $FluPol_{A/Fj}$ were seen when both vRNA and cap analogue were bound, surprisingly indicating that cap analogue was markedly stabilising the polymerase heterotrimer. This was supported by thermofluor assays (not shown), which revealed that the presence of cap analog could

increase the melting temperature of $FluPol_{A/Fj}$ (with vRNA bound) from approximately 43ºC to 46ºC. Unfortunately, I was not able to view particles of $FluPol_B$ during this project.

Obtaining higher resolution information required the use of cryo-EM. I concentrated on investigating $FluPol_C$. The preparation of grids for cryo-EM was complicated by the necessary presence of 0.5 M NaCl in the protein purification buffer, as this high concentration of salt interfered with sample vitrification, resulting in thick and uneven ice across the grid (hence low particle contrast). Dilution of $FluPol_C$ into a solution with less NaCl immediately prior to grid preparation or washing the grids after sample adsorption both caused particle aggregation (not shown). The answer to the problem was to dilute the sample into a solution containing 120 mM NaSCN instead of NaCl (figure 2.8d). The reason for this may be due to the more chaotropic nature of SCN^- compared to Cl^-, so allowing it to exert the same effect as Cl^-, but at lower concentrations. Unfortunately, in the micrographs taken, there was very little contrast between background and $FluPol_C$, even at high defocus, making particle picking and alignment very difficult. This meant that we were not able obtain any useful class averages from the data. However, this ought to be easily rectified by looking at the grids using a lower accelerating voltage.

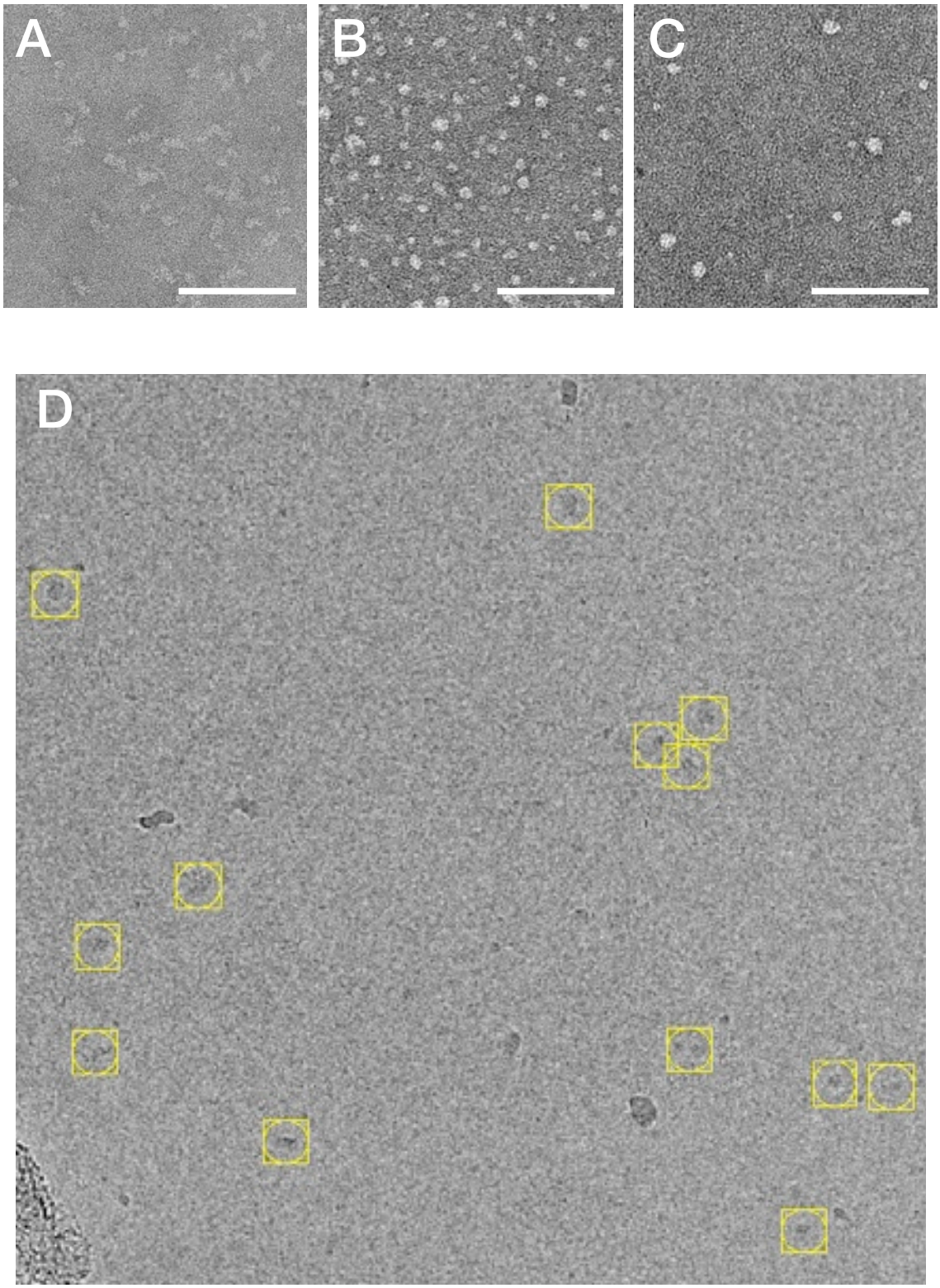


Figure 2.8. Electron microscopy of FluPol.
(A) Micrograph of $FluPol_{A/NT}$-CBP negatively stained with uranyl acetate. The white bar in this and subsequent images indicates 100 nm **(B)** Micrograph of $FluPol_{A/Fj}$ bound to vRNA promoter and in the presence of cap analogue, similarly negatively stained with uranyl acetate. **(C)** Micrograph of negatively stained $FluPol_C$ (uranyl acetate) in the presence of cap analogue. **(D)** Cryo-EM micrograph of $FluPol_C$ in the presence of cap analogue. Particles thought to be polymerase are highlighted with yellow boxes. This part of the figure was prepared by Dr Juha Huiskonen.

2.3 Summary

In this chapter I have demonstrated the development of an baculoviral-expression and purification system capable of producing milligram quantities of >95% pure influenza FluPol. I have used this system to produce polymerases from influenza A, B and C viruses. The purification of $FluPol_A$ from insect cells has previously been accomplished, but that of $FluPol_B$ and $FluPol_C$ is novel. Additionally I am confident that my system outperformed those previously reported for $FluPol_A$, as I could obtain higher yields (3 – 4 mg compared to 0.7 mg per litre of culture (Aggarwal *et al*, 2011)) of good quality polymerase. SDS-PAGE analysis of the purified polymerase revealed even staining of all three polymerase subunits (when these could be separated from each other, figure 2.4b), providing evidence that this was intact, heterotrimeric material. Gel-filtration, as well as allowing me to remove aggregated material from my preparations, provided further evidence for this, as did MALS and *in vitro* polymerase activity assays.

The key features of my insect-cell expression system were codon-optimised genes and co-expression of all three subunits from only one baculovirus. This latter feature is novel for FluPol expression-systems, and probably did help by allowing co-expression with a lower multiplicity of infection. The use of a protein-A tag for purification was also highly advantageous, as the selectivity of this strategy allowed me to minimise the number of purification steps required. This was important for the final yield of FluPol, due to substantial loses during each purification step. The other initial feature of my expression system was the co-expression of a mini-vRNA, but this turned out not to work as hoped. This was not too surprising, since useful expression of this mini-vRNA relied on several processes that had not been tested within insect cells: folding and self-cleavage of the ribozymes at each end of the transcript, nuclear import of FluPol and successful recognition

and amplification of the mini-vRNA by FluPol. Luckily mini-vRNA co-expression was not essential for the purification of isolated polymerase heterotrimers. For $FluPol_C$, and also probably $FluPol_B$, this was because these polymerases did not did not form higher-order oligomers in the absence of bound promoter RNA. For the IAV polymerases, which on their own did aggregate, promoter RNA could be bound during the purification, with a beneficial effect.

This change in the physical behaviour of $FluPol_A$ upon binding promoter RNA is quite interesting, as it provides further evidence that the molecular basis of FluPol regulation by the vRNA promoter is a dramatic conformational change. Moreover, the observation that FluPol without RNA is prone to interacting with itself could be functionally relevant, given the recent models for replication that invoke an additional polymerase heterotrimer acting *in trans* (Jorba *et al*, 2009; York *et al*, 2013).

Another important finding from my purification work was that FluPol at high concentrations would precipitate out of low ionic strength solutions. The basis for this may be the presence of both highly acidic and basic patches on the surface of FluPol, which salt could act to screen, thus preventing their interaction. Indeed, the crystallisation of $FluPol_C$ (see below) must have relied on this salting effect, since crystals formed as the drop size increased.

I was also able to perform some preliminary experiments to measure the binding affinities of $FluPol_A$ for its various promoter RNAs. The higher affinity of FluPol for the 5′ strand of the vRNA promoter (compared to the 3′ strand), supports a model for vRNA assembly in which the first step is binding of free-FluPol to 5′ vRNA as it exits an elongating FluPol. There is no need for the 5′ and 3′ ends to come together to form a duplex before FluPol can bind. The clear differences between cRNA and vRNA promoter binding

affinities are surprising. This suggests that the cRNA promoter strands interact very differently with the polymerase; which might be the underlying reason for the different mode of initiation from cRNA. The finding that the 3′ cRNA is unable to bind FluPol on its own is a direct challenge to the replication model of Jorba *et al*, as this proposes that vRNA synthesis occurs through a *trans*-acting polymerase that initiates off of a free 3′ cRNA end (Jorba *et al*, 2009). However, more work will need to be done to verify this result.

The EM work that I performed was only preliminary, but because I was able to overcome the salt problem, should provide a good foundation to allow further characterisation of FluPol by cryo-EM. This is likely to be a good approach for obtaining high-resolution structural information about complexes that could prove intractable to crystallography, such as FluPol-promoter or FluPol-mRNA complexes, particularly given the advent of direct electron detectors for EM and the ever improving software for single-particle reconstruction. EM could also uncover alternative conformations of the free polymerase that are not captured in the crystal structure.

Crystallography

3.1 Initial Crystallisation Trials

3.1.1 Influenza A Virus Polymerases

Initially, I set up crystallisation trials with $FluPol_{A/NT}$ and $FluPol_{A/FJ}$. Each of these was tested in the sitting-drop vapour-diffusion format (Walter *et al*, 2005; 2008) against several commercial crystallisation screens. In addition to this, I varied several crystallisation parameters, such as temperature, the protein buffer and the presence of promoter RNA or other ligands. A summary of the different screening trials set up is given in table 3.1. The conditions used in these trials were rather variable at first, as I was still testing different protein buffers. Non-detergent sulfobetaine 201 (NDSB-201) or zwitterionic detergent (fos-choline-12 or DDMAB) were included in some crystallisation trials as it was thought that these could reduce protein aggregation. This was also the reason why a protein buffer based around 0.5 M NaCl with 0.5 M $(NH_4)_2SO_4$ was tried. 2,4-dioxo-4-phenylbutanoic acid (DPBA) and cap analog (m^7GpppG) were included in some trials, as these are known ligands of FluPol, and thus might have stabilised the complex. One important factor that I only noticed later on, was that the presence of glycerol and high NaCl concentrations in the FluPol solutions generally caused the crystallisation drops to expand as they equilibrated. Since this indicated that the precipitant and protein were becoming more dilute over the course of the experiment (opposite to the textbook description of an ideal crystallisation experiment), in later trials I tried to ensure that the drops shrunk, by adding extra NaCl (this alone did not work) and glycerol to the reservoir solutions. However, the crystals of $FluPol_C$ that I eventually obtained, grew from drops that

increased in size over time (see section 3.2). Thus crystallisation of $FluPol_C$ was being driven by a decreasing NaCl concentration in the drop. This made sense, given the high NaCl requirement during purification. Since this was also the case with the IAV polymerases, my efforts in most of these trials to ensure that the crystallisation drops shrunk were probably counterproductive.

Table 3.1. Crystallisation trials for $FluPol_{A/NT}$ and $FluPol_{A/Fj}$.

Date	Protein	Conc. (mg/ml)	Temp. (°C)	Protein buffer	Comments
11/11/11	$FluPol_{A/NT}$ (+ CBP)	4	20.5	0.5 M NaCl, 8 mM $MgCl_2$, 25 mM HEPES:NaOH, pH 8.0, 0.5 M NDSB-201	FluPol eluted from gel filtration as an isolated heterotrimer, despite not having RNA bound.
19/11/11	$FluPol_{A/NT}$ (+ CBP)	5	20.5 and 4	0.5 M NaCl, 8 mM $MgCl_2$, 25 mM HEPES:NaOH, pH 8.0, 0.5 M NDSB-201	FluPol eluted from gel filtration as a mixture of higher-order oligomers (mainly trimer of timers?).
30/11/11	$FluPol_{A/NT}$ (+ CBP)	1.6	4	0.5 M NaCl, 8 mM $MgCl_2$, 25 mM HEPES:NaOH, pH 8.0, 0.5 M NDSB-201	FluPol eluted as a mixture of higher-order oligomers. **One drop of crystals grew.** These were not reproducible.
12/3/12	$FluPol_{A/NT}$	4.5	4	0.5 M NaCl, 0.5 M $(NH_4)_2SO_4$, 35 mM Tris:HCl, pH 7.5, 1 mM DTT, 2 mM $MgCl_2$, 0.5 mM EDTA	FluPol eluted from gel filtration at a volume close to that expected for an isolated trimer, but the peak displayed a prominent shoulder.
05/5/12	$FluPol_{A/NT}$	1.5	4	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM DTT, 10% (v/v) glycerol	FluPol eluted as a mixture of different oligomers.
30/5/12	$FluPol_{A/Fj}$ + vRNA	1.7	20.5	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 18% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 1 mM DTT	FluPol eluted as an isolated heterotrimer.
29/06/12 and 04/07/12	$FluPol_{A/Fj}$ + vRNA	4.3	20.5	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol	FluPol eluted as an isolated heterotrimer. Reservoir conditions in some plates was supplemented with 0.5 M NaCl.
01/08/12	$FluPol_{A/Fj}$ + vRNA	5.5	4	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol	FluPol eluted as an isolated heterotrimer. Reservoir conditions in some plates were supplemented with 0.5 M NaCl.
06/08/12	$FluPol_{A/Fj}$ + vRNA	5.5	4	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol, 0.0025 mg/ml trypsin	Same protein as 01/08/12, but flash frozen / thawed. Reservoir conditions were supplemented with 0.5 M NaCl. Testing limited proteolysis to remove flexible loops.

Table 3.1. Crystallisation trials for $FluPol_{A/NT}$ and $FluPol_{A/Fj}$.

Date	Protein	Conc. (mg/ml)	Temp. (°C)	Protein buffer	Comments
28/08/12	$FluPol_{A/Fj}$ + vRNA	5.5	4	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol, 0.0005 mg/ml subtilisin A, 0.5 mM DPBA and 0.5 mM m^7GpppG	Same protein as 01/08/12, but flash frozen / thawed. Reservoir conditions were supplemented with 0.5 M NaCl. Testing limited proteolysis to remove flexible loops.
30/08/12	$FluPol_{A/Fj}$ + vRNA	5.5	4	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol, 0.0005 mg/ml chymotrypsin, 0.5 mM DPBA and 0.5 mM m^7GpppG	Same protein as 01/08/12, but flash frozen / thawed. Reservoir conditions were supplemented with 0.5 M NaCl. Testing limited proteolysis to remove flexible loops.
04/09/12	$FluPol_{A/Fj}$ + vRNA	5.5	20.5	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol, 0.5 mM DPBA and 0.5 mM m^7GpppG	Same protein as 01/08/12, but flash frozen / thawed. Reservoir conditions were supplemented with 0.5 M NaCl. Testing limited proteolysis to remove flexible loops.
04/09/12	$FluPol_{A/Fj}$ + vRNA	5.5	20.5	0.5 M NaCl, 25 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol, 0.5 mM DPBA and 0.5 mM m^7GpppG, 0.0025 mg/ml chymotrypsin	Same protein as 01/08/12, but flash frozen / thawed. Reservoir conditions were supplemented with 0.5 M NaCl. Testing limited proteolysis to remove flexible loops.
31/01/13	$FluPol_{A/NT}$ + vRNA	4.4	4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 10% (v/v) glycerol	Fresh protein. FluPol eluted as an isolated heterotrimer. Reservoir conditions were supplemented with 0.5 M NaCl before setting up drops. 10 µl 70% (v/v) glycerol added to reservoirs after setting up drops.
25/09/13	$FluPol_{A/Fj}$ + vRNA	3.3	20.5	1 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 0.5 mM m^7GpppG and either 14 mM Fos-Choline-12 or 5 mM DDMAB	Fresh protein. FluPol eluted as a mixture of different oligomers. Testing crystallisation with different detergents.
10/10/13	$FluPol_{A/Fj}$	2	20.5 and 4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 10 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP	Fresh protein. FluPol eluted as a mixture of different oligomers.

I was able to grow crystals of $FluPol_{A/NT}$ in only one trial drop (table 3.1). This was set up with the PEGRx (Hampton) A2 condition (0.1 M sodium citrate tribasic dihydrate, pH 5.5, 38% (v/v) polyethylene glycol (PEG) 200). Despite much effort, these crystals were not reproducible. Surprisingly, the crystals grew from polymerase that did not have promoter RNA bound and so purified as a mixture of higher-order oligomers. The crystals appeared after 4/5 days and grew over the next 3 weeks to a size of approximately 10 μm x 10 μm x 20 μm. One of these was tested at beamline I04 at the Diamond Light Source (Didcot, Oxfordshire, UK), but only diffracted to ~60 Å. The number of diffraction spots recorded was insufficient for the determination of space group or unit cell dimensions. No other crystals of $FluPol_{A/NT}$ could be grown. Additionally, attempts to crystallise $FluPol_{A/Fj}$ did not succeed.

3.1.2 Nanobodies

Since these initial trials did not yield reproducible crystals, I set out to explore the use of nanobodies as a crystallisation chaperone. Nanobodies are single-chain, camelid antibodies that have proved to be very useful in the crystallisation of several challenging proteins (Domanska *et al*, 2011; Steyaert & Kobilka, 2011; Baranova *et al*, 2012; Park *et al*, 2012). These proteins are often large, like FluPol, and display conformational heterogeneity, which lowers their propensity to form crystals. Nanobodies appear to aid crystallisation by reducing this heterogeneity, either by binding to flexible loops on the target protein and/or rigidifying the relative orientation of various protein domains. Additionally, they are small, form beneficial crystal contacts, are easily expressed and are easily adapted into high-throughput selection pipelines.

3.1.3 Nanobody Production and Testing

Nanobodies were raised against $FluPol_{A/Fj}$ with vRNA promoter bound. I produced the polymerase complexes, however nanobody selection was performed by Els Pardon and colleagues, in the lab of Jan Steyaert at the VIB research institute in Brussels. This involved immunising two llamas with $FluPol_{A/Fj}$–vRNA complexes. One llama was immunised with polymerase complex that had been stabilised by crosslinking (in order to improve their immunogenicity) while the other was immunised with a non-crosslinked complex. A library of nanobody clones was then generated from these llamas. Phage-display was used to screen this library for nanobodies that would bind the polymerase complex. These hits were also verified by enzyme-linked immunosorbent assays. After this process, I received plasmids for 24 different polymerase-binding nanobodies (table 3.2).

Table 3.2. Nanobody information

Name	Sequence	Family*
CA8189	QVQLQESGGGLVQPGGSLRLSCAASGSDFSLYAMGWYRQVPGKERELVATITS GGTTNFADSARGRFTISRGNAKNTVYLQMNNLKPEDTAVYYCWSRGSYGRYL ETWGQGTQVTVSSHH HHHHEPEA	1
CA8190	QVQLQESGGGLVQAGGSLRLSCAASGRSLSDYTMGWFRQAPGKEREFVTAIT SDGVYANYADSVKGRFTISRDNAKNTAYLQMNSLKPEDTAVYYCAGKTRRSM LMTTGYDYWGQGTQVT VSSHHHHHHEPEA	
CA8191	QVQLQESGGGLVQPGGSLRLSCTASESVASINIVGWYRQISGKERELVARISSGG TITDYSDSVKGRFTITRDNTKNTVYLQMNSLKRGDTAVYYCNAEYQYGSDWF HIWGQGTQVTVSS HHHHHHEPEA	5
CA8192	QVQLQESGGGLVQAGASLRLSCAASGRTFMGWFRQAPGQEREFVAAIDAFGI NTYYADSAKGRFTISRDSAKNTVYLQMNSLKPEDTAVYYCAATLRGTQPGAV MQWRPDYWGQGTQVTV SSHHHHHHEPEA	6
CA8193	QVQLQESGGGLVQPGGSLRLSCAASGFTFKMYAMSWVRQAPGKGLEWVSSIN SAGGSTSYVDSVKGRFTISRDNAKNTLYLQMNSLKPDDTAVYYCVQGRNWPY DYRGQGTQVTVSSHHH HHHEPEA	7
CA8194	QVQLQESGGGLVQPGGSLRLSCAASGFTFSNYYMSWVRQAPGEGREWVSSIN RDGSNTYYADSVKGRFTIARDNVKNTLYLLMDSLKSDDTAVYYCTKGSNWPY DNWGQGTQVTVSSHHH HHHEPEA	2
CA8195	QVQLQESGGGLVQAGGSLRLSCAASTRTFSIYTMGWFRQAPGKEREFVAAIAW NDQTRYADSVKGRFTISRDNAKNTVYLQMNSLKQEDTAVYYCAAGTYSEYEN SYAYWGQGTQVTVSS HHHHHHEPEA	3
CA8196	QVQLQESGGGLVQAGDSLRLSCTTSGSSLSRYGVGWFRQAPGEEREFGASISW NGATYYADSVKGRFTISRDNAKNTVYLQMNSLKAEDTAVYYCAARSDGYLTP YLSPVYWGQGTQVTV SSHHHHHHEPEA	4
CA8197	QVQLQESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSSITS AFGRTSYVSSVKGRFTISRDNAKNTLYLQMNSLKPEDTAVYYCAQGASWPYR YRGQGTQVTVSSHHH HHHEPEA	8
CA8198	QVQLQESGGGLVQTGDSLRLSCAGAERSFNSYTMGWFRQAPGKEREFVAAIA FSGGFTYYAEAVKGRFTISRDSGKNTVYLQMNSLKPEDTAVYYCAARFYGGG YYYQQREYSFWGQGTQ VTVSSHHHHHHEPEA	9
CA8199	QVQLQESGGGLVQAGGSLRLSCAASGRTFSSYAMGWFRQAPGKEREFVGAID LSGLYKHYGDSVKGRFTISRDNAKNTVYLQMSSLKPEDTAVYYCAAAGIYYA GSYSPLLYDYWGQGTQ VTVSSHHHHHHEPEA	11
CA8200	QVQLQESGGGLVQAGDSLRLSCAASGRTFSGYIMAWFRQTPGKERELVARLGP SSSIQYYRDSVKGRFTLSRDNAKNTLYLQMNSLKPEDTAVYYCAFTLKWSSDY RSHNAYDYWGQGTQ VTVSSHHHHHHEPEA	14
CA8201	QVQLQESGGGLVQPGGSLRLSCAASGSDFSLYAMGWYRQVPGKERELVATITI GGTTNFADSARGRFTISRGNAKNTVYLQMNNLKPEDTAVYYCWSRGSYGRYL ETWGQGTQVTVSSHH HHHHEPEA	15
CA8202	QVQLQESGGGLVQAGESLRLSCAASGGTFIYYGMGWFRQAPGKEREFVASISQ TGGDIYAMDFVKGRFTISRDNSKNTLYLQMNSLKPEDTAVYYCAAATSPYASR KLYEYKYWGQGTQV TVSSHHHHHHEPEA	16
CA8203	QVQLQESGGGLVQAGGSLRLSCAAPGRILTLSRYVMGWFRQAPGKEREFVAAI SPGTTDTDYADSVKGRFTISRDNAKNTVYLEMSSLKPEDTAVYVCAAGAASSG TVYRVYGLYNYRGQ GTQVTVSSHHHHHHEPEA	17

Table 3.2. Nanobody information

Name	Sequence	Family*
CA8204	QVQLQESGGGLVQPRGSLRLSCAASGSIAIIYTMGWYRQAPLKERELVATVTY GGNTNYADPVKGRFTISRDNNKNTEYLQMSSLQPEDTAVYYCYMRVRSTAWD TDDYWGQGTQVTVSS HHHHHHEPEA	18
CA8205	QVQLQESGGGMVQPGGSLRLSCLASGFTFSNYAMTWVRQAPGKGPEWVSMV SNNGADTTYTDSVKGRFTISRDNAKNTLYLRMNNVKPEDSAVYYCAKRRYGG IWTGQPTDYDYLGQGTQ VTVSSHHHHHHEPEA	10
CA8206	QVQLQESGGGLVQPGGSLRLSCTASGGIFGISAMGWYRQAPGKERELVARITR SGSTNFADSVKDRFTISRDATENTVVLQMNSLKAEDTAVYYCNGQFWEVDNG NIDLWGQGTQVTVSS HHHHHHEPEA	
CA8207	QVQLQESGGGLVQAGASLRLSCTASGRTSSTYPMGWFRQGPGKEREFVAAIS WAGSSTYYGDFVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCARAARPYA YGLDYSTEASYDYWGQ GTQVTVSSHHHHHHEPEA	12
CA8208	QVQLQESGGGLVQPGGSLRLSCAASGRTFSMYAMAWFRQAPGKQREFVGTIS RSGDYALHADAVKGRFTISRDNAKNTVYLQMNSLKLEDTAVYYCAAGAYHSK DKTLYDYWGQGTQVTV SSHHHHHHEPEA	
CA8209	QVQLQESGGGLVQPGGSLRLSCAASGNIRSINTMSWYRQAPGKERELVASDST GGSTGGSTVYAASVKGRFTISRDNAKNTVYLQMNMLKPEDTAVYYCYVYGM SRFGSTISTLFGQGTQ VTVSSHHHHHHEPEA	13
CA8210	QVQLQESGGGLVQAGESLRLSCAASGGTLIYYGMGWFRQAPGKEREFVASISQ AGGDIYAMDFVKGRFTISRDNSKNTLYLQLNTLKPEDTAVYYCAAATSPYASR KLYDYKYWGQGTQV TVSSHHHHHHEPEA	
CA8211	QVQLQESGGGLVQAGDSLRLSCAASGRTFSGYIMAWFRQTPGKEREFVARLGP SSNMQYYVDSVKGRFTLSRDNAKNTLYLQMNSLKPEDTAVYYCVFTLKWSPD YRSHNAYDYWGQGTQ VTVSSHHHHHHEPEA	19
CA8212	QVQLQESGGGLVQPGGSLRLSCAASGRSFNSYTMGWFRQAPGKEREFVAAIA FSGGFTYYAEAVKGRFTISRDNGEKPVYLRMNSLKPEDTAVYYCAARFYGGG YYYQEREYTDWGQGTQ VTVSSHHHHHHEPEA	20

* A family is a group of nanobodies with high similarity in the complementarity-determining-region-3 (CDR3) sequence.

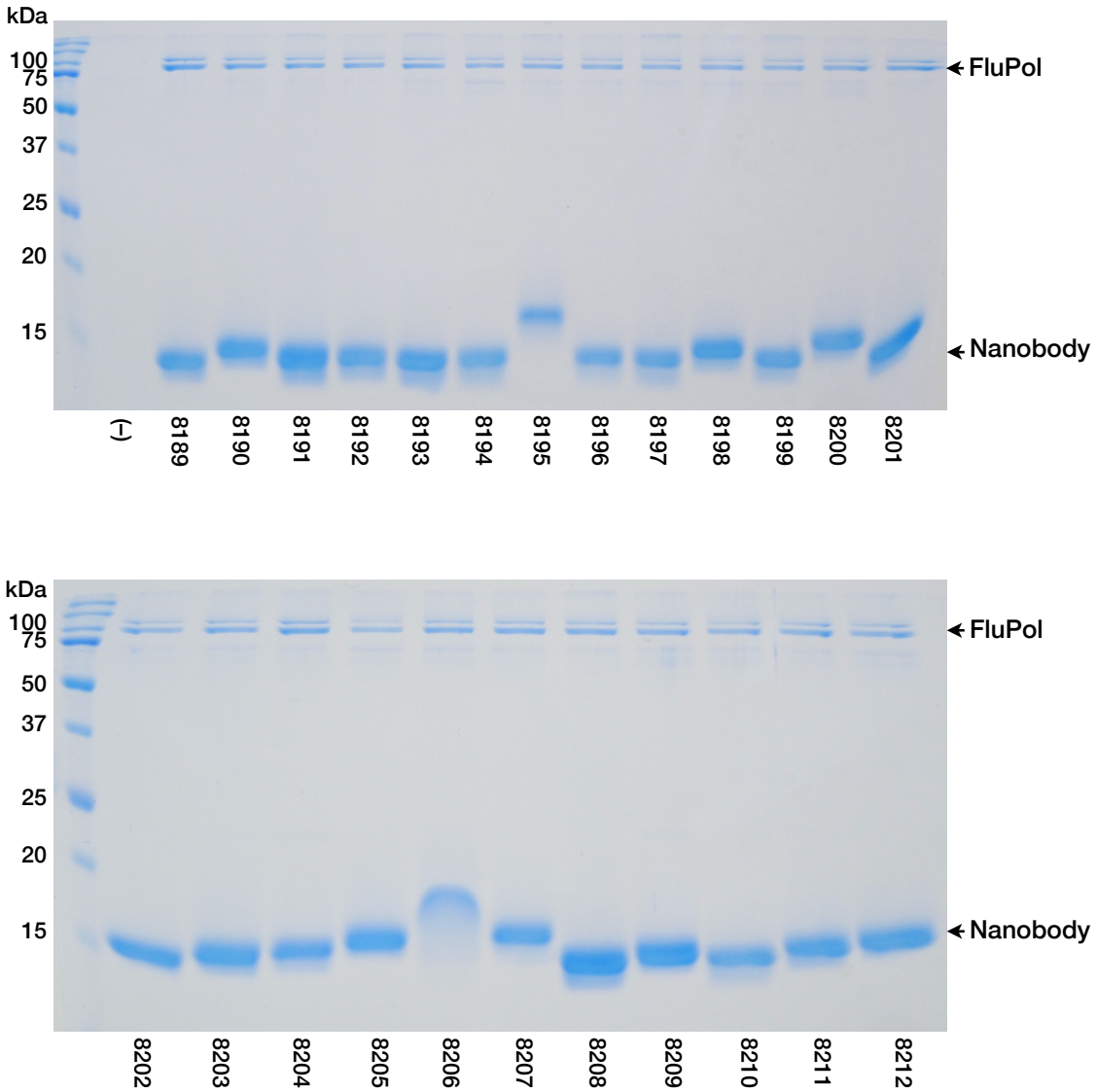


Figure 3.1. Nanobody binding to FluPol$_{A/Fj}$.
Pull-down assays with FluPol$_{A/Fj}$ and Ni-NTA agarose beads alone (–) or coated in the indicated nanobody. The protein pulled down by the beads was analysed by Coomassie-stained SDS-PAGE.

I expressed and purified these nanobodies from *E. coli* cells and further verified polymerase binding using pull-down assays (figure 3.1). These nanobodies were used in crystallisation trials against the index screen (Hampton). This screen was arbitrarily chosen. For each trial plate, a 1.2x molar excess of one of the nanobodies was mixed with the FluPol$_{A/Fj}$-vRNA complex (in 0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 0.5 mM m^7GpppG), and used for

sitting-drop vapour-diffusion crystallisation experiments. After the drops in each plate were pipetted, I added 20 μl 70% (v/v) glycerol to the reservoirs, and then allowed the drops to equilibrate at 20.5ºC. No protein crystals grew from any of these trials. At this point, my focus shifted away from the nanobody work, because of progress with $FluPol_C$. Therefore, no other crystallisation trials were set up with these nanobodies.

3.1.4 Influenza B and C Virus Polymerases

Due to the lack of progress with the IAV polymerases, I broadened my efforts to include $FluPol_B$ and $FluPol_C$ (as well as other related polymerases, though these could not be purified). As above, crystallisation trials for these were set up in the sitting-drop vapour-diffusion format (Walter *et al*, 2005; 2008). These are summarised in table 3.3. Attempts to crystallise $FluPol_B$ did not succeed. On the other hand, $FluPol_C$ proved much more amenable to crystallisation. This will be described in the section below.

Table 3.3. Crystallisation trials for $FluPol_B$ and $FluPol_C$.

Date	Protein	Conc. (mg/ml)	Temp. (°C)	Protein buffer	Comments
29/01/13	$FluPol_C$	4.3	20.5 and 4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP	Fresh protein. Reservoir conditions were supplemented with 0.5 M NaCl before setting up drops. 10 µl 70% (v/v) glycerol added to reservoirs of 4ºC plates after setting up drops. **One crystal grew. This was not reproducible.**
07/02/13	$FluPol_B$	2.6	4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 2.5% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 0.02% (w/v) NaN_3	Fresh protein. 10 µl 10% (v/v) glycerol in 5 M NaCl was added to reservoirs before setting up drops. Additional glycerol was added to selected reservoirs after setting up drops.
07/02/13	$FluPol_C$	3.5	4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 2.5% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 0.02% (w/v) NaN_3	Fresh protein. 10 µl 10% (v/v) glycerol in 5 M NaCl was added to reservoirs before setting up drops. Additional glycerol was added to selected reservoirs after setting up drops.
16/10/13	$FluPol_B$	2.5	20.5 and 4	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP	Fresh protein.
25/10/13	$FluPol_C$	3.0	20.5	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 10 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP	Fresh protein. **Some crystals grew.**
12/11/13	$FluPol_C$	3.3	20.5	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 10 mM $MgCl_2$, 0.5 mM TCEP	Fresh protein. **Many crystals grew.**
07/06/14	$FluPol_C$	5	20.5	0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 10% (v/v) glycerol, 10 mM $CaCl_2$, 0.5 mM TCEP	Fresh protein. Screening with added seeds, made from crystals grown in Morpheus G2 condition. Ratio of protein:reservoir:seed in drops was 2:1:1 or 3:1:1. **Many crystals grew.**

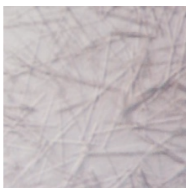
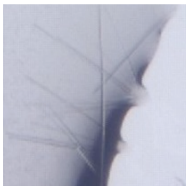
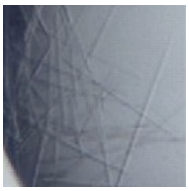
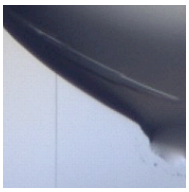
3.2 Crystallisation of Influenza C Virus Polymerase

3.2.1 Crystal Optimisation

The initial crystals of $FluPol_C$ are summarised in table 3.4. These tended to be needle-like and these were often too small to produce measurable diffraction spots. Some crystals were more three-dimensional, however these also appeared layered and multiple. None of the initial hits gave high resolution diffraction. The crystal that grew from the block 4 A6 condition diffracted to ~8 Å, but an indexable dataset could not be collected from this. However, this crystal did provide the first indication of the contents of the unit cell: the point group was *P*222, with *a*, *b* and *c* axes of 109 Å, 219 Å and 595 Å respectively ($\alpha = \beta = \gamma = 90^o$). According to Matthews' analysis (Matthews, 1968), this cell could accommodate four polymerase heterotrimers in the asymmetric unit, with a solvent content of 65%. Later work (section 3.2.2) showed that crystals from the Morpheus conditions (Molecular Dimensions) belonged to the $P4_32_12$ space group, with cell dimensions (*a*, *b*, *c*) of approximately 185 Å, 185 Å, 600 Å. Interestingly, these only accommodated two heterotrimers in the asymmetric unit, with a solvent content of 76%.

The initial crystallisation hits were used as starting points in a more systematic search for optimal conditions. Varying precipitant concentration, protein concentration, protein buffer NaCl concentration, pH, crystallisation temperature, initial protein:precipitant ratio, and the presence of different additives were all tested. These additives included cap analog and promoter RNA (initially IAV vRNA promoter, and subsequently ICV vRNA promoter), as well as those from commercial screens. All of the initial crystals appeared in the presence of $MgCl_2$, which strongly suggested that divalent metal cations promoted crystal formation. Hence, I also tested the effect of varying the nature of the metal cation, on $FluPol_C$ crystallisation.

Table 3.4. Initial crystallisation hits for FluPol$_C$

Conc (mg/ml)	Condition	Morphology	Growth
4.3	Block 4 A6: 20% (w/v) PEG 8000, 10 mM $MgCl_2$, 50 mM MES, pH 5.6, 100 mM $(NH_4)_2SO_4$. Supplemented with 0.5 M NaCl. 20.5ºC	 Longest axis: ~200 µm	Appeared after 2 days. Maximum size after 3 weeks.
4.3	Block 4 C7: 25% (v/v) PEG Monomethyl Ether 550, 5 mM $MgCl_2$, 50 mM HEPES:NaOH pH 7.0. Supplemented with 0.5 M NaCl. 20.5ºC	 Needles, ~50 µm long	Appeared after a few hours. Maximum size after 3 days.
4.3	Block 4 A7: 15% (v/v) 2-Propanol, 20 mM $MgCl_2$, 50 mM MES, pH 6.0. 4ºC	 Needles, ~200 µm long	Appeared after a few hours. Maximum size after 1 day.
3.0	PEGRx A8: 0.1 M MES monohydrate, pH 6.0, 22% (v/v) PEG 400. Protein supplemented with 10 mM $MgCl_2$.	 Needles, ~200 µm long	Appeared after a few hours. Maximum size after 3 days.
3.3	Morpheus H2: 0.1 M amino acids, 0.1 M Imidazole/MES, pH 6.5, 30% Ethylene glycol/PEG 8000. Protein supplemented with 10 mM $MgCl_2$.	 ~500 µm long and ~30 µm wide.	Appeared after 7 days. Maximum size after 18 days.
3.3	Morpheus G2: 0.1 M amino acids, 0.1 M Imidazole/MES, pH 6.5, 30% Ethylene glycol/PEG 8000. Protein supplemented with 10 mM $MgCl_2$.	 ~200 µm long and ~30 µm wide.	Appeared after 3 days. Maximum size after 7 days. Crystals also appeared in other conditions containing the same precipitant mix, but different additives. These other crystals had similar morphologies.

I was not able to obtain any further crystals from conditions based on the block 4 A6 crystallisation hit. Optimisations around the block 4 C7, block 4 A7 and PEGRx A8 conditions, did not lead to any improvement; producing only needle-like crystals. Fortunately, I was able to optimise the Morpheus conditions. In this case, the best crystals were obtained from trials at room temperature, set up by mixing 200 nl $FluPol_C$ (4 – 5 mg/ml in 0.5 M NaCl, 25 mM HEPES:NaOH, 10% (v/v) glycerol, 0.5 mM TCEP and 10 mM $MgCl_2$ or 10 mM $CaCl_2$) with 100 nl reservoir solution (70% (v/v) Morpheus G2 supplemented with 0% – 1% (v/v) 1 M NaOH). These crystals appeared after 1 - 4 weeks, and reached their maximum size after approximately 1 week of growing time. It was also possible to grow well-diffracting crystals in larger, manually pipetted drops (1 μl $FluPol_C$ plus 0.5 μl reservoir). The best crystals were rod-shaped, with lengths of up to 1 mm, and square cross-sections between 10 μm x 10 μm and 50 μm x 50 μm. Those that gave the highest resolution diffraction tended to be shorter but thicker. However, in addition to rods, showers of needle-like crystals frequently grew. These diffracted poorly. Crystals could also be obtained in the absence of added $MgCl_2$ or $CaCl_2$, however the diffraction quality of these crystals was no better than for those obtained with these salts. Screening this condition against the Silver Bullets (Hampton) additive screen did not result in any crystals with different morphologies.

I also investigated crystal dehydration as a means of improving diffraction quality. This was done using the HC1 humidity control device at the Diamond synchrotron, which enabled the systematic testing of a wide range of dehydration parameters. However, all crystals that underwent this dehydration process diffracted very poorly, if at all. Additionally, I carried out soaking experiments with various putative ligands to determine whether any of these could improve the diffraction from these crystals (and to gain information about ligand binding). The substrates tested included: 5′ FluC vRNA, 5′ FluC

vRNA with 3′ FluC vRNA (table 6.3), m^7GpppG, ATP, ATPγS, ApG dinucleotide and DPBA. None of these soaks improved diffraction from the crystals. Indeed the crystals soaked with RNA did not diffract at all. For the crystals that did generate datasets, ligand binding was not observed in the resulting maps.

3.2.2 Analysis of Tetragonal Crystals

I analysed the diffraction of approximately 110 native crystals from the optimised Morpheus condition described above. Most of these produced very poor diffraction patterns that did not contain any measurable reflections at resolutions better than 8 Å, and could not be automatically indexed by *mosflm* (Leslie & Powell, 2007) or *EDNA*, implemented at the beamline. Datasets for these crystals were therefore not collected. However, some of these crystals did diffract, and the details of a selection of these is shown in table 3.5. All of the diffracting crystals were tetragonal and belonged to the same space group, $P4_32_12$, however the precise cell dimensions were variable. The *a* and *b* axes were between 183.6 – 188.5 Å in length, while the *c* axis took values between 595.9 – 601.6 Å. For the best crystals, reflections could be seen out to ~3.6 Å, however the diffraction was extremely anisotropic. The highest resolution reflections were generated along the c^* direction. Crystals that diffracted to 3.6 Å in this direction, still only produced reflections up to 6/7 Å in the a^* and b^* directions. This was reflected by the range of anisotropic B-factors within the merged and scaled datasets. For example, the dataset from the vdx_d3 crystal (table 3.5) had an anisotropic B-factor range of 122.11 Å^2, derived from anisotropic B-factor values of 40.70 Å^2, 40.70 Å^2 and -81.41 Å^2 along the a^*, b^* and c^* axes respectively (as reported by the UCLA Diffraction Anisotropy server: http://services.mbi.ucla.edu/anisoscale/).

Table 3.5. Diffracting tetragonal crystals.

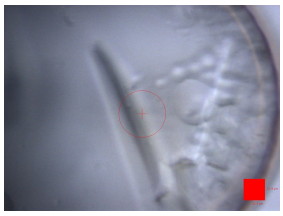
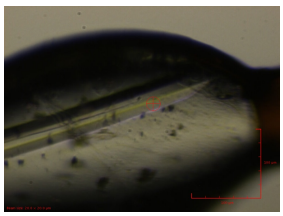
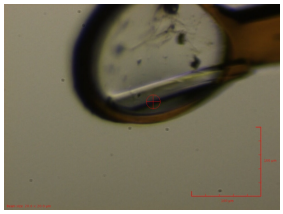
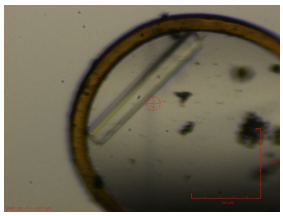
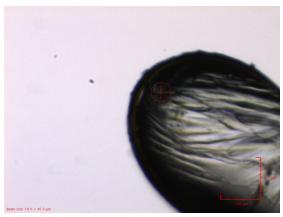
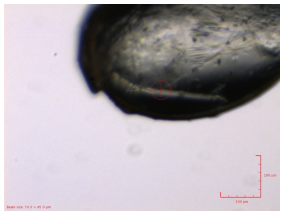
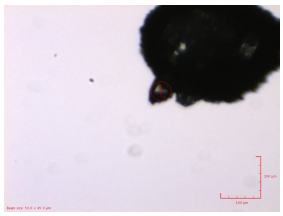
Crystal	Morphology	Resolution[a] (Å)	Cell dimensions [*a*, *b*, *c* (Å)][b]	Completeness (%)
515724_G8-13		7.7	188.5 188.5 599.6	78
518770_F8-7		5.9	185.8 185.8 598.7	100
518770_F8-8		6.6	185.6 185.6 598.2	83
518770_H1-14		5.3	184.6 184.6 595.9	58
518039_F3-6		4.7	187.8 187.8 599.1	88
518039_F2-7		5.5	188.2 188.2 598.0	91
518787_H7_4		7.9	187.4 187.4 601.6	49

Table 3.5. Diffracting tetragonal crystals.

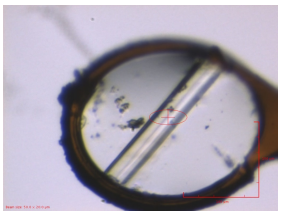
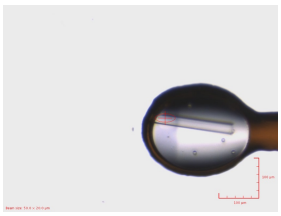
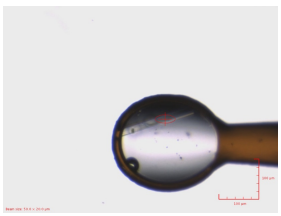
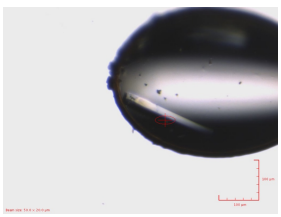
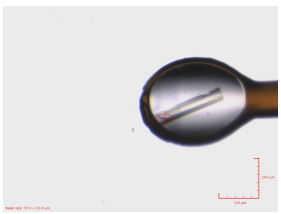
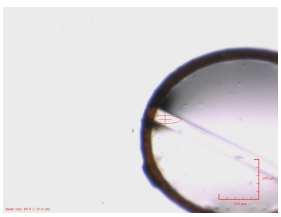
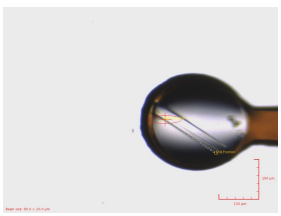
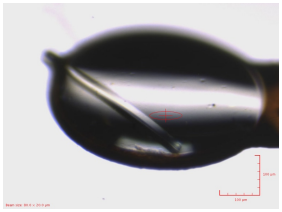
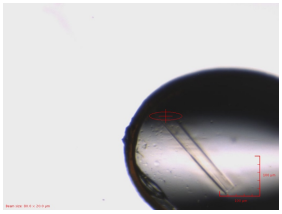
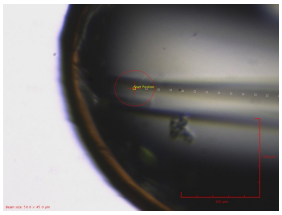
Crystal	Morphology	Resolution[a] (Å)	Cell dimensions [*a*, *b*, *c* (Å)][b]	Completeness (%)
527970_F8		4.8	185.6 185.6 596.2	39
524566_F5-5		6.3	184.4 184.4 599.2	63
524566_F5-6		6.2	184.3 184.3 598.8	49
525679_E4-2		4.9	185.5 185.5 598.9	56
525679_E4-3		6.2	184.0 184.0 598.1	49
525648_F10-11		4.2	183.6 183.6 597.8	73
525648_H2-13		4.6	184.0 184.0 596.2	53

Table 3.5. Diffracting tetragonal crystals.

Crystal	Morphology	Resolution[a] (Å)	Cell dimensions [*a*, *b*, *c* (Å)][b]	Completeness (%)
525648_H7-15		5.3	185.9 185.9 597.2	52
525648_H11-16		4.3	186.1 186.1 598.3	59
vdx_D3		4.4	185.4 185.4 599.0	100

[a] Maximum resolution determined by *xia2* run with default settings, or resolution at which the $CC_{1/2}$ (Karplus & Diederichs, 2012) value drops to below 0.8.
[b] The space group for all crystals was $P4_32_12$, hence $\alpha = \beta = \gamma = 90°$.

It was not usually possible to obtain a complete dataset from a single crystal, due to radiation damage. Therefore, data from multiple crystals was merged, using *xia2* (Winter *et al*, 2013). Not all crystal datasets could be merged with each other, as the variation in the unit cell dimensions was too large. I used *BLEND* (Foadi *et al*, 2013) as well as manual selection, to try to determine the best combination of datasets to merge. However, this often grouped wedges of data collected from the same crystal into different clusters, which did not seem appropriate. All of these datasets processed to resolutions between 4.2 Å – 4.3 Å, with $CC_{1/2} > 0.25$ (Karplus & Diederichs, 2012), though the merging R-factors were very high. The *BLEND*-curated datasets did not have better statistics than the manual datasets. This is summarised in table 3.6.

Table 3.6. Native dataset statistics for tetragonal crystals.

	Manual set 1	Manual set 2	*BLEND* set 1	*BLEND* set 2
Crystals merged	518770_F8-7 518770_F8-8 518770_H1-14	525679_E4-2 525679_E4-3 524566_F5-5 524566_F5-6 524566_H5-7 524566_H7-9 524566_H7-10 525648_F10-11 525648_H2-12 525648_H2-13 525648_H7-15 525648_H11-16	Different combinations of wedges of data from crystals of manual sets 1 and 2.	
X-ray source	Diamond I02	Diamond I03	Diamond I02 and I03	Diamond I02 and I03
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit cell [a, b, c (Å), α = β = γ = 90°]	185.5 185.5 598.1	184.7 184.7 598.1	185.5 185.5 598.6	184.2 184.2 597.2
Resolution range (Å)	100.53 – 4.20 (4.30– 4.20)*	127.62 – 4.30 (4.40 – 4.30)	116.48 – 4.20 (4.30 – 4.20)	127.24 – 4.20 (4.30 – 4.20)
Completeness (%)	99.5 (97.9)	99.1 (96.2)	99.3 (95.7)	97.8 (91.9)
Multiplicity	11.9 (9.7)	10.2 (8.3)	15.4 (12.0)	7.1 (6.2)
$CC_{1/2}$	0.980 (0.277)	0.989 (0.585)	0.985 (0.421)	0.987 (0.270)
$\langle I/\sigma I \rangle$	6.9 (1.2)	5.3 (1.2)	6.9 (0.9)	4.5 (0.8)
R_{merge}	0.450 (2.510)	0.352 (1.581)	0.540 (3.154)	0.307 (1.878)
R_{pim}	0.191 (1.208)	0.157 (0.803)	0.198 (1.339)	0.166 (1.131)

* Highest resolution shell is shown in parenthesis.

In the end, the best native dataset was collected from a single crystal that diffracted unusually well (vdx_D3, see table 3.5 and figure 3.2). This crystal was from a drop that had been set up by hand. The drop therefore had a larger total volume than those made using a cartesian robot, though it is unclear whether or not this was the factor that produced this higher-quality crystal. The data were recorded with a wavelength of 0.86 Å on a PILATUS 6M-F detector (DECTRIS) with the crystal-to-detector distance set to 863.3 mm to cover diffraction to 3.69 Å at the edge. A line scan was used so that the complete length of the crystal was exposed during data collection. The total rotation range was 90°, with an exposure time of 0.75 s per 0.15° (100% beam transmission, ~10^{12} photons/s, 50 μm x 45 μm beam size). A long exposure time was necessary for this and other tetragonal crystals, because these crystals contained a very high solvent content (76%). The data from this crystal were indexed, integrated and scaled using *HKL2000* (HKL Research). The statistics for this are reported in tables 3.7 and 3.8. The maximum resolution was 3.6 Å, with a $CC_{1/2}$ (Karplus & Diederichs, 2012) of 0.365.

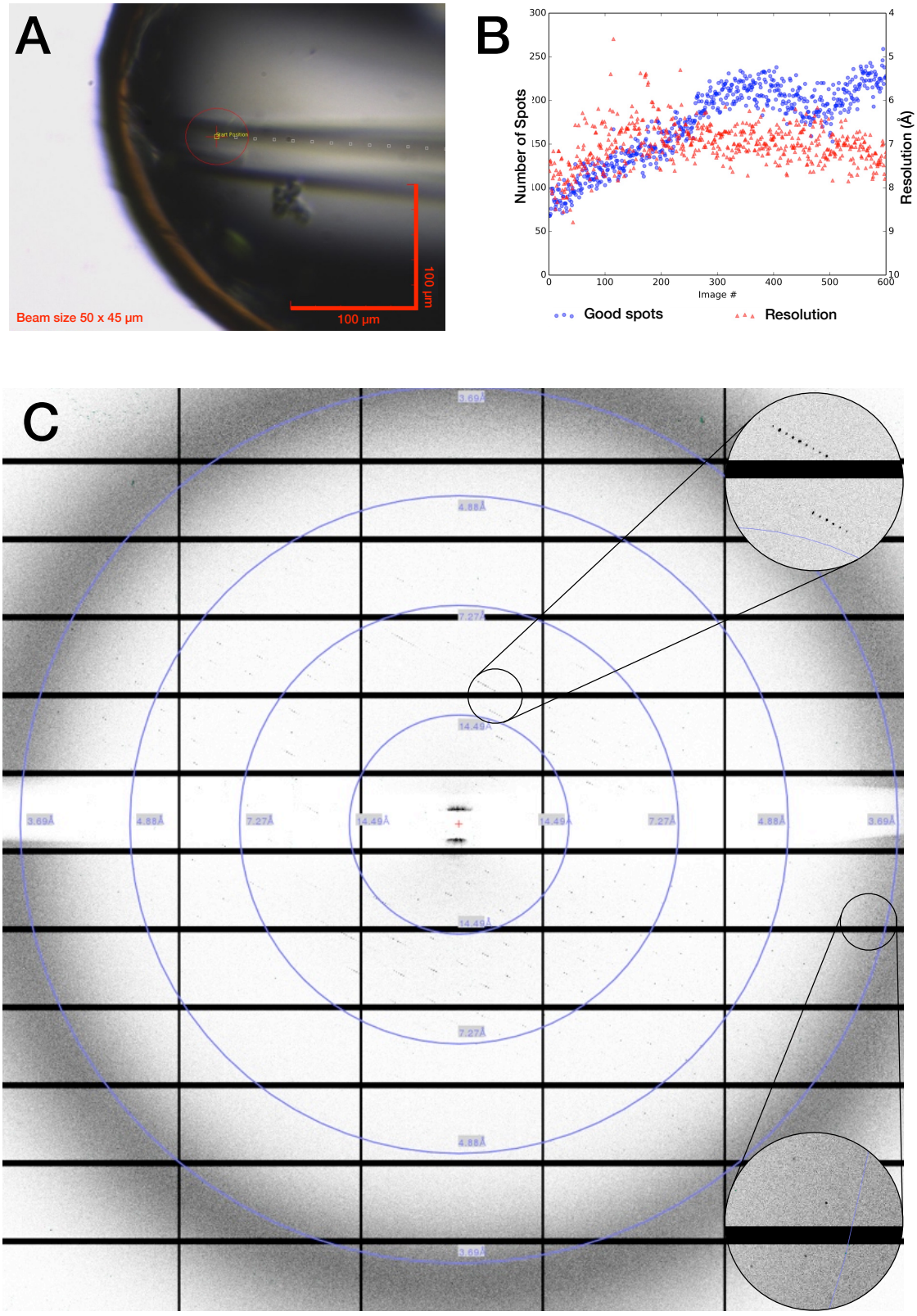


Figure 3.2. Data collection for crystal vdx_D3.
(A) Image of crystal vdx_D3. The red circle indicates the beam size and position. The yellow squares show the path of the line scan collection strategy. **(B)** *DISTL* plot of the number of spots (blue) and estimated resolution (red) for each image collected. **(C)** A representative detector image of the diffraction from crystal vdx_D3.

Table 3.7. Best native dataset for a tetragonal crystal

	Crystal vdx_D3
X-ray source	Diamond I04
Wavelength (Å)	0.8634
Space group	$P4_32_12$
Unit cell [a, b, c (Å), α = β = γ = 90°]	185.27 185.27 598.76
Resolution range (Å)	50 – 3.60 (3.73 – 3.60)
Completeness (%)	99.9 (100)
Multiplicity	6.5 (6.5)
$CC_{1/2}$	0.365
$\langle I/\sigma I \rangle$ ***(%)***	6.0 (0.4)
R_{merge}	0.144 (> 1)
R_{pim}	0.089 (> 1)

Table 3.8. Statistics for vdx_D3, by resolution

Resolution			
Lower	Upper	$CC_{1/2}$	R_{pim}
50.00	7.75	0.999	0.017
7.75	6.15	0.993	0.057
6.15	5.38	0.979	0.123
5.38	4.89	0.967	0.160
4.89	4.54	0.957	0.206
4.54	4.27	0.934	0.305
4.27	4.05	0.873	0.481
4.05	3.88	0.733	0.776
3.88	3.73	0.519	> 1
3.73	3.60	0.365	> 1

3.2.3 Phasing

The native data could not be phased by molecular replacement as there was insufficient prior structural information. Phasing at a low resolution using a pseudo-atomic model based on the highest resolution EM reconstruction of $FluPol_A$ (Moeller *et al*, 2012) was also attempted, but this too did not work. Therefore, experimental phasing had to be employed.

I initially investigated the use of various heavy-metal compounds for derivatisation of the crystals by soaking, which would allow me to obtain an anomalous dataset. This work was all carried out with the tetragonal crystals described above. The first compounds tried were large, covalent clusters, as these would be easiest to locate in the low resolution datasets, and would provide the most phasing power. The clusters used were: tantalum bromide ($[(Ta_6Br_{12})^{2+}].2Br^-$), phosphotungstate ($Na_3[PW_{12}O_{40}].H_2O$), metatungstate ($Na_6[H_2W_{12}O_{40}].H_2O$) and paratungstate ($(NH_4)_{10}[H_2W_{12}O_{42}].H_2O$). Roughly 15 crystals were tested with each tungsten compound, while approximately 25 crystals were tested with tantalum bromide. Metatungstate caused the crystals to become flexible and fail to diffract, even when they were only exposed to this compound for a few seconds. Diffraction from crystals soaked in the other compounds was poor, often extending only to ~10 Å and in the best cases only reaching 6/7 Å. Thus, the clusters appeared to be damaging the crystals. Anomalous signal was only detected in datasets from crystals soaked with tantalum bromide, however this did not extend to useful resolutions.

I then moved on to testing smaller heavy-metal compounds. The compounds tried were: tetrakis(acetoxymercuri)methane, mercury (II) acetate, potassium tetrachloroplatinate (II), gold (I) potassium cyanide and lead (II) acetate trihydrate. The first three of these made

the crystals flexible, and they did not diffract. Crystals soaked in the latter two compounds did diffract.

Datasets from these crystals were recorded at the peak wavelength of the appropriate metal. It was not possible to collect at inflection and remote wavelengths due to the radiation sensitivity of the crystals. Data collection followed an inverse-beam collection strategy, in which the crystal was rotated 180° from the initial position every 3.75° – 5°. This was to minimise the radiation damage occurring between recording each reflection of a Friedel pair. This strategy did allow the collection of datasets that contained anomalous differences, however we also often encountered problems with crystal rotation axis alignment, and with merging data from the 0° and +180° sweeps. Thus in later work (see section 3.2.5), instead of doing inverse-beam collection, I used a mini-κ goniometer to align each crystal along its fourfold axis, so that single images would contain both reflections of each Friedel pair. If I were to repeat the collection of the datasets described below, then this latter strategy would be preferred.

The highest resolution reflections were recorded from lead-soaked crystals. For the lead dataset, an average of 120° of data were collected per crystal, in two wedge series of 4 x 15° each, with a rotation of 0.15° and an exposure of 0.2 s (100% beam transmission, ~10^{12} photons/s, 20 μm x 20 μm beam size) per frame. The data were recorded on a PILATUS 6M-F detector (DECTRIS) with the crystal-to-detector distance set to cover diffraction to 3.5 Å at the edge. This dataset was indexed, integrated, merged and scaled using *xia2* (Winter *et al*, 2013), run with *XDS*, *XSCALE* (Kabsch, 2010) and *aimless* (Evans & Murshudov, 2013). The resolution of the lead dataset was 5.0 Å with a $CC_{1/2}$ (Karplus & Diederichs, 2012) of 0.821 (table 3.9). However, this dataset only contained anomalous signal up to 15.8 Å as determined by *XSCALE* $|\mathbf{F}^+ - \mathbf{F}^-|/\sigma$ of 1.1 with an anomalous correlation of 0.22 (table

3.10). Indeed, it was initially thought that no lead had bound in the crystals at all, and the lead dataset was treated as native (see below). After it was realised that the lead dataset did contain limited anomalous signal, the data could be phased in the same manner as described below; generating surprisingly interpretable electron density maps. However, these were not used during subsequent model building.

Gold-soaked crystals did not diffract to such high resolutions, but anomalous signal was detectable over more of the resolution range. Many gold soaked crystals were examined and an extensive effort was made to obtain a dataset with strong phasing power up to the highest possible resolution. For this, I attempted to merge together data from many crystals, but this was again complicated by the variation in inter-crystal unit cell dimensions. The use of *BLEND*, to try and determine which datasets to merge together, did not produce the best overall dataset, as non-isomorphism between crystals degraded the anomalous signal. Instead, the best gold dataset resulted from merging only two crystals that grew in the same drop. For this dataset, of 180º of data were collected from one crystal and 40º from another, in two wedge series of 5 x 20º each or 1 x 20º each, with a rotation of 0.2º and an exposure of 0.2 s (100% beam transmission, ~10^{12} photons/s, 50 μm x 20 μm beam size) per frame. The data were recorded on a PILATUS 6M detector (DECTRIS) with the crystal-to-detector distance set to cover diffraction to 4.5 Å at the edge. These data were processed using *xia2* as described above for the lead dataset. The statistics for this dataset are summarised in tables 3.9 and 3.10. The resolution of the gold dataset reached 6.9 Å (with a $CC_{1/2}$ (Karplus & Diederichs, 2012) of 0.612 for the highest resolution shell), while the anomalous signal only extended to 8.9 Å, as determined by *XSCALE* $|\mathbf{F}^+ - \mathbf{F}^-|/\sigma$ of 1.1 with an anomalous correlation of 0.36.

Table 3.9. Anomalous dataset statistics.

	Gold	Lead
Crystals	520285_G3 520285_G3-2	518657_F6-10 518657_F6-11 518657_G6-12
X-ray source	Diamond I03	Diamond I04
Wavelength (Å)	1.0350	0.9460
Space group	$P4_32_12$	$P4_32_12$
Unit cell ***[a, b, c (Å),*** ***α = β = γ = 90°]***	184.2 184.2 598.8	185.5 185.5 598.8
Resolution range (Å)	127.29 – 6.91 (7.09 – 6.91)	119.76 – 5.00 (5.13 – 5.00)
Completeness (%)	99.1 (72.7)	91.3 (46.2)
Multiplicity	14.5 (3.7)	15.2 (4.4)
$CC_{1/2}$	0.998 (0.612)	0.996 (0.821)
Anomalous Completeness (%)	97.7 (72.7)	88.9 (46.2)
Anomalous Multiplicity	7.8 (1.7)	8.0 (2.1)
⟨I/σI⟩ (%)	16.5 (1.8)	10.6 (2.1)
R_{merge}	0.144 (0.575)	0.178 (0.597)
R_{pim}	0.051 (0.416)	0.056 (0.408)

Table 3.10. Statistics by resolution for anomalous datasets.

Gold				Lead			
Resolution				Resolution			
Lower	Upper	Anomalous Correlation*	Anomalous signal**	Lower	Upper	Anomalous Correlation*	Anomalous signal**
127.29	30.90	0.88	2.915	119.76	22.36	0.84	1.614
30.90	21.85	0.88	3.151	22.36	15.81	0.22	1.063
21.85	17.84	0.83	2.671	15.81	12.91	0.09	0.939
17.84	15.45	0.84	2.507	12.91	11.18	-0.09	0.814
15.45	13.82	0.76	2.251	11.18	10.00	-0.16	0.738
13.82	12.62	0.72	1.967	10.00	9.13	-0.18	0.741
12.62	11.68	0.67	1.856	9.13	8.45	-0.14	0.730
11.68	10.93	0.61	1.638	8.45	7.91	-0.20	0.690
10.93	10.30	0.55	1.531	7.91	7.45	-0.16	0.679
10.30	9.77	0.48	1.394	7.45	7.07	-0.14	0.659
9.77	9.32	0.46	1.294	7.07	6.74	-0.10	0.657
9.32	8.92	0.36	1.097	6.74	6.46	-0.10	0.645
8.92	8.57	0.28	0.994	6.46	6.20	-0.04	0.662
8.57	8.26	0.20	0.900	6.20	5.98	-0.01	0.663
8.26	7.98	0.24	0.868	5.98	5.77	-0.05	0.643
7.98	7.73	0.09	0.745	5.77	5.59	-0.03	0.623
7.73	7.50	0.08	0.736	5.59	5.42	-0.01	0.668
7.50	7.28	0.02	0.669	5.42	5.27	-0.01	0.669
7.28	7.09	0.02	0.698	5.27	5.13	0.00	0.671
7.09	6.91	0.05	0.650	5.13	5.00	0.05	0.663

* Anomalous correlation = Correlation between random half-sets of anomalous intensity differences
** Anomalous signal = $|F+ - F-|/\sigma$

Phasing efforts proceeded with the gold dataset as this contained the most anomalous signal. The initial gold substructure was determined using the *HKL2MAP* interface (Pape & Schneider, 2004) with *SHELXC/D/E* (Sheldrick, 2010), supplied with unmerged data. Multiple runs of *SHELXD* were performed using different numbers of heavy atom sites, producing solutions with 14 – 48 gold atoms in the asymmetric unit. The site occupancy statistic from these indicated that the true number was ~35 atoms (figure 3.3a). The profile of the correlation coefficients between observed and calculated Bijvoet differences for each *SHELXD* trial demonstrated that a correct solution could be found (figure 3.3b). This was supported by *SHELXE*, which showed a difference in contrast between maps phased using the gold substructure calculated with the original or inverted hand (figure 3.3c). This also confirmed that the space group was $P4_32_12$, not $P4_12_12$. However, the electron density maps produced by *SHELXE* were not interpretable.

Instead, the map initially used for model building was produced by single isomorphous replacement with anomalous scattering (SIRAS) phasing, performed by *phenix.autosol* (Adams *et al*, 2010), with the lead data used as the native dataset (at this point, the vdx_D3 dataset had not yet been recorded) and the gold substructure from *SHELXD*. This incorporated several rounds of density modification by solvent flattening using a solvent content of 76% and twofold non crystallographic symmetry (NCS) averaging, thus extending the resolution of phases from 8.9 Å up to the that of the native dataset. Anisotropy correction and B-factor sharpening were performed automatically for both datasets by the program. The solution contained 34 gold atoms in the asymmetric unit. The map with the most detail was produced when "native" lead data was used up to a resolution of 4 Å. Single-wavelength anomalous dispersion (SAD) phasing with the gold dataset in *phenix.autosol*, followed by phase extension in *phenix.autobuild* (Adams *et al*, 2010) with higher resolution reflections from the lead or manual-set-2 datasets, did not produce

maps with equivalent levels of detail as with SIRAS phasing. Eventually, a more detailed map was generated by SIRAS phasing as above, but with the gold and vdx_D3 datasets and using the refined gold substructure from the initial SIRAS run. The resolution cut-off for the vdx_D3 data was 3.6 Å. This second map was used in the later stages of model building.

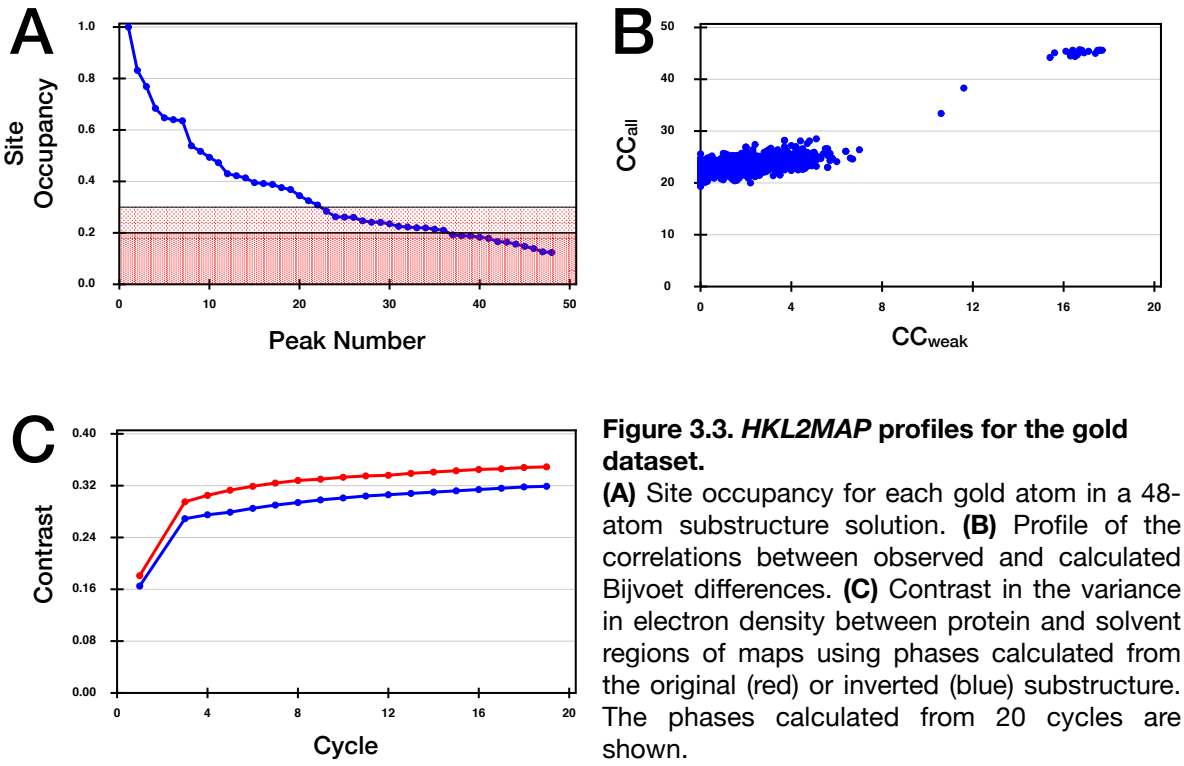


Figure 3.3. *HKL2MAP* profiles for the gold dataset.
(A) Site occupancy for each gold atom in a 48-atom substructure solution. **(B)** Profile of the correlations between observed and calculated Bijvoet differences. **(C)** Contrast in the variance in electron density between protein and solvent regions of maps using phases calculated from the original (red) or inverted (blue) substructure. The phases calculated from 20 cycles are shown.

3.2.4 Model Building

The structure of $FluPol_C$ was built into SIRAS-phased and B-factor sharpened electron density maps using *Coot* (Emsley & Cowtan, 2004). Secondary structure, particularly α-helices, were readily visible in the initial map, made using the "native" lead data to 4 Å. Density corresponding to amino acid side chains could be seen in certain, well-ordered, regions (figure 3.4), however most of the peptide side chains were not visible. This, together with the fact that the map also contained several regions of missing and uninterpretable density, made *de novo* model building extremely challenging. The first stage

of model building was fitting of the available fragments of $FluPol_A$ structure into the $FluPol_C$ electron density. These accounted for ~50% of the structure of the polymerase. Fitting was accomplished manually. We modelled the core of PB1 by fitting the structure of the hepatitis C virus RdRp (PDB ID: 1NB4, (O'Farrell *et al*, 2003)), deleting and modifying regions where appropriate. Sequence alignments were used to guide amino acid assignments for regions that had been fitted with existing structures. The remaining density was filled, as much as possible, with a polyalanine model. This was very difficult, as the electron density emanating from fitted fragments of both PB1 and PB2 led into regions where the connectivity was ambiguous. A later map, produced from the vdx_D3 data, displayed slightly more detail and helped resolve some uncertain regions of the initial model. However this still contained regions of difficult-to-interpret electron density.

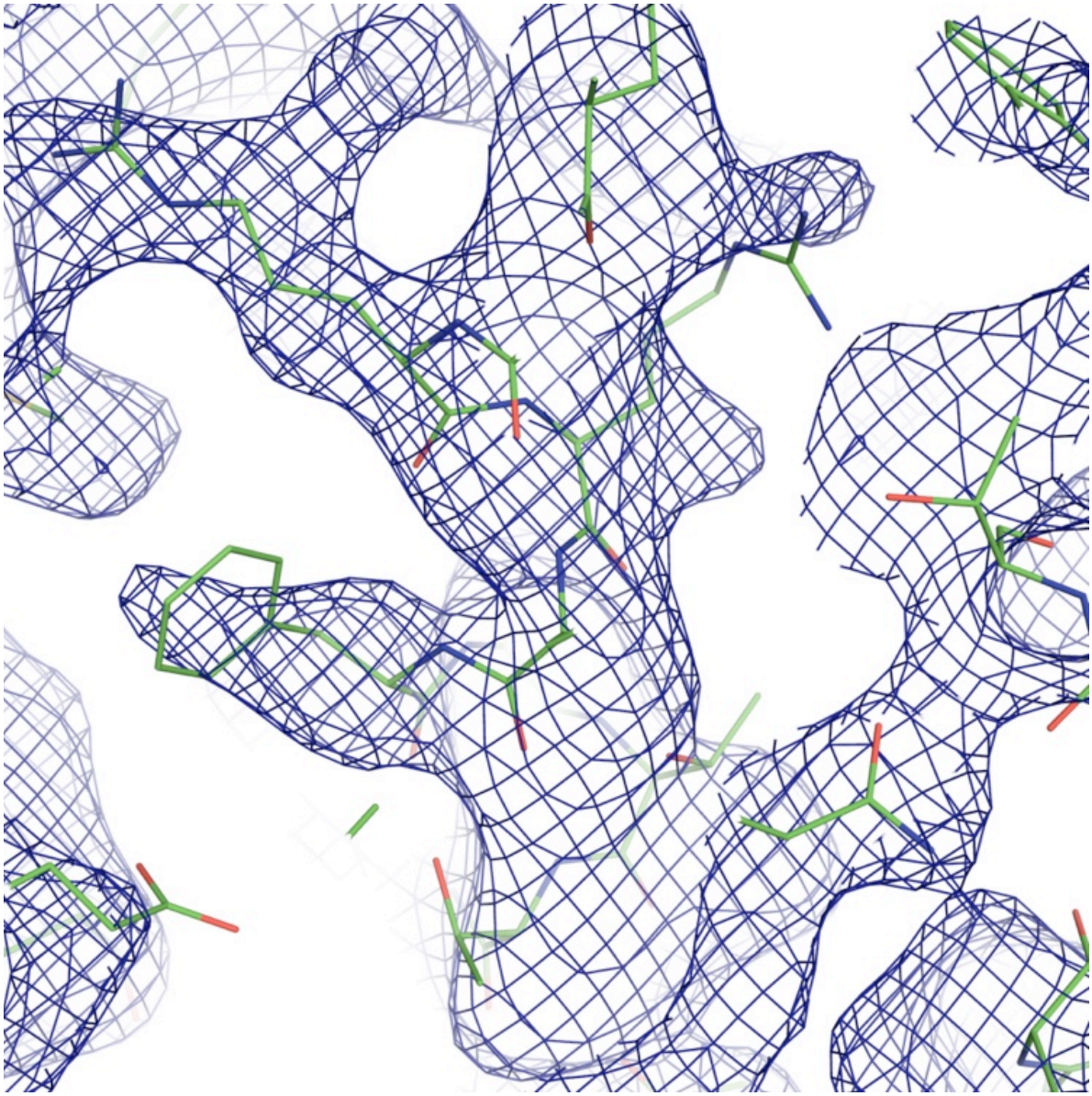


Figure 3.4. Electron density map of tetragonal crystal form.
Model for the $P4_32_12$ structure of $FluPol_C$ (green sticks), showing a well-ordered, helical region around P3 residue 570, overlaid with the corresponding 3.6 Å, unrefined electron density map (blue) contoured at σ1.0.

3.2.5 Selenomethionine and Sulphur Anomalous-Difference Maps

In order to find the positions of sulphur atoms within the electron density (to guide model building at low resolution), I carried out selenomethionine- (SeMet) SAD and sulphur-SAD (S-SAD) experiments. The yields of SeMet-derivatised $FluPol_C$ were much lower than those for the natural protein, and I was only able to produce sufficient quantities of this by using conditions that would likely result in only ~25% SeMet incorporation (Cronin *et al*, 2007). Crystals of SeMet-$FluPol_C$ grew from similar conditions as the un-derivatised protein. The best crystals grew when 50% – 70% (v/v) Morpheus G2 (pH 7.5) was used as the precipitant and seeds were included in the crystallisation drops. The seed-containing drops were set up by dispensing 100 nl seed mixture (see section 6.6.2) on top of 200 nl protein solution (4.5 mg/ml SeMet-$FluPol_C$, 0.5 M NaCl 25 mM HEPES:NaOH, 10% (v/v) glycerol, 0.5 mM TCEP, 10 mM $MgCl_2$ or $CaCl_2$). Various seed dilutions were tested. These drops were equilibrated against reservoirs containing 100% - 56% (v/v) Morpheus G2. Judging from their appearance, the quality of the majority of SeMet-$FluPol_C$ crystals seemed low, because they were often thin and cracked. This was reflected in the diffraction quality of these crystals: 8 were examined and the best crystal only diffracted to 7.36 Å, with $\langle I/\sigma I \rangle$ of 1.1 and $CC_{1/2}$ (Karplus & Diederichs, 2012) of 0.616 in the outer resolution shell. A total of 1080° of data were collected at a wavelength of 0.9795 Å from this crystal, in 6 x 180° wedges at different points along the crystal, aligned on its fourfold axis, with an oscillation of 0.1° and an exposure (100% beam transmission, ~10^{12} photons/s, 20 μm x 20 μm beam size) of 0.05 s per frame. However, due to radiation damage, it was not possible to merge all of this data. Instead the highest anomalous signal was seen in the first 180° wedge, however this only extended to 13.4 Å as determined by *XSCALE* $|\mathbf{F}^+ - \mathbf{F}^-|/\sigma$ of 1.1 with an anomalous correlation of 0.23. Anomalous-difference maps generated using this data with SIRAS-solved phases (section 3.2.4) were very noisy, with some peaks

corresponding to expected methionine positions but many others not corresponding. Therefore, these did not help with interpretation of the electron density maps.

This was similarly the case with the S-SAD experiments. In these, the highest anomalous signal was seen by merging 1312.5° of data from two crystals that grew from the same drop. These data were collected using a wavelength of 1.77 Å in wedges of various sizes at different points along each crystal (each aligned on its fourfold axis), with an oscillation of 0.15° and an exposure (50% beam transmission, ~10^{12} photons/s, 50 μm x 45 μm beam size) of 0.15 s or 0.3 s per frame. The anomalous signal extended to 10.7 Å (from an $|\mathbf{F}^+ - \mathbf{F}^-|/\sigma$ of 1.1 with an anomalous correlation of 0.25). It was not possible to merge more data from other crystals due to the lack of isomorphism between them. Again, the resulting anomalous-difference maps were very noisy and did not aid model building.

3.2.6 Seeding

The low resolution of data from the tetragonal crystals (3.6 Å, with experimentally determined phases up to only 8.9 Å), together with the failure of SeMet-SAD and S-SAD to provide information about sulphur atom positions, meant that model building could not proceed any further. In order to make progress, higher resolution datasets were required. However, an extensive effort had already been made to produce Morpheus condition crystals that diffracted isotropically to resolutions better than 3.5 Å, and this had not succeeded. Therefore, I tried to find different crystal forms of $FluPol_C$ by setting up new crystallisation trials against all of the available screens, but this time including micro-seeds made from crushing crystals grown in the Morpheus condition. This re-screen resulted in crystals growing out of a large number of novel conditions. Many of these crystals were shaped like rods or needles, similar to the Morpheus-derived crystals or to other hits from

the initial screens. However, other conditions produced crystals that were shorter and thicker than previously seen.

3.2.7 Analysis of Orthorhombic Crystals

The diffraction properties of a selection of these seeded-crystals was examined at the Diamond Light Source. Surprisingly, many of the crystals looked at were orthorhombic; belonging to a lower symmetry space group, $P2_12_12_1$, than the crystals that produced nuclei used to seed them. Most only diffracted to poorly (blurred spots, multiple lattices etc.) and to low resolutions, including several crystals that appeared to be both thick and singular when viewed in the drop. A few produced better quality diffraction. The statistics for the best crystal amongst these are summarised in table 3.11 and figure 3.4. A total of 90° of data were collected at a wavelength of 0.9795 Å, in a single wedge, with an oscillation of 0.1° and an exposure (100% beam transmission, ~10^{12} photons/s, 50 μm x 45 μm beam size) of 0.05 s per frame. The data were recorded on a PILATUS 6M-F detector (DECTRIS) with the crystal-to-detector distance set to cover diffraction to 4.5 Å at the edge. This dataset was processed using *HKL2000* (HKL Research).

Table 3.11. Best native dataset for an orthorhombic crystal

	Crystal 535067_D1
X-ray source	Diamond I04
Wavelength (Å)	0.9795
Space group	$P2_12_12_1$
Unit cell [a, b, c (Å), α = β = γ = 90°]	107.05 217.50 597.58
Resolution range (Å)	50 – 4.30 (4.45 – 4.30)
Completeness (%)	99.4 (99.4)
Multiplicity	3.2 (2.9)
$CC_{1/2}$	(0.500)
$\langle I/\sigma I \rangle$ *(%)*	5.1 (1.4)
R_{merge}	0.214 (0.687)
R_{pim}	0.140 (0.478)

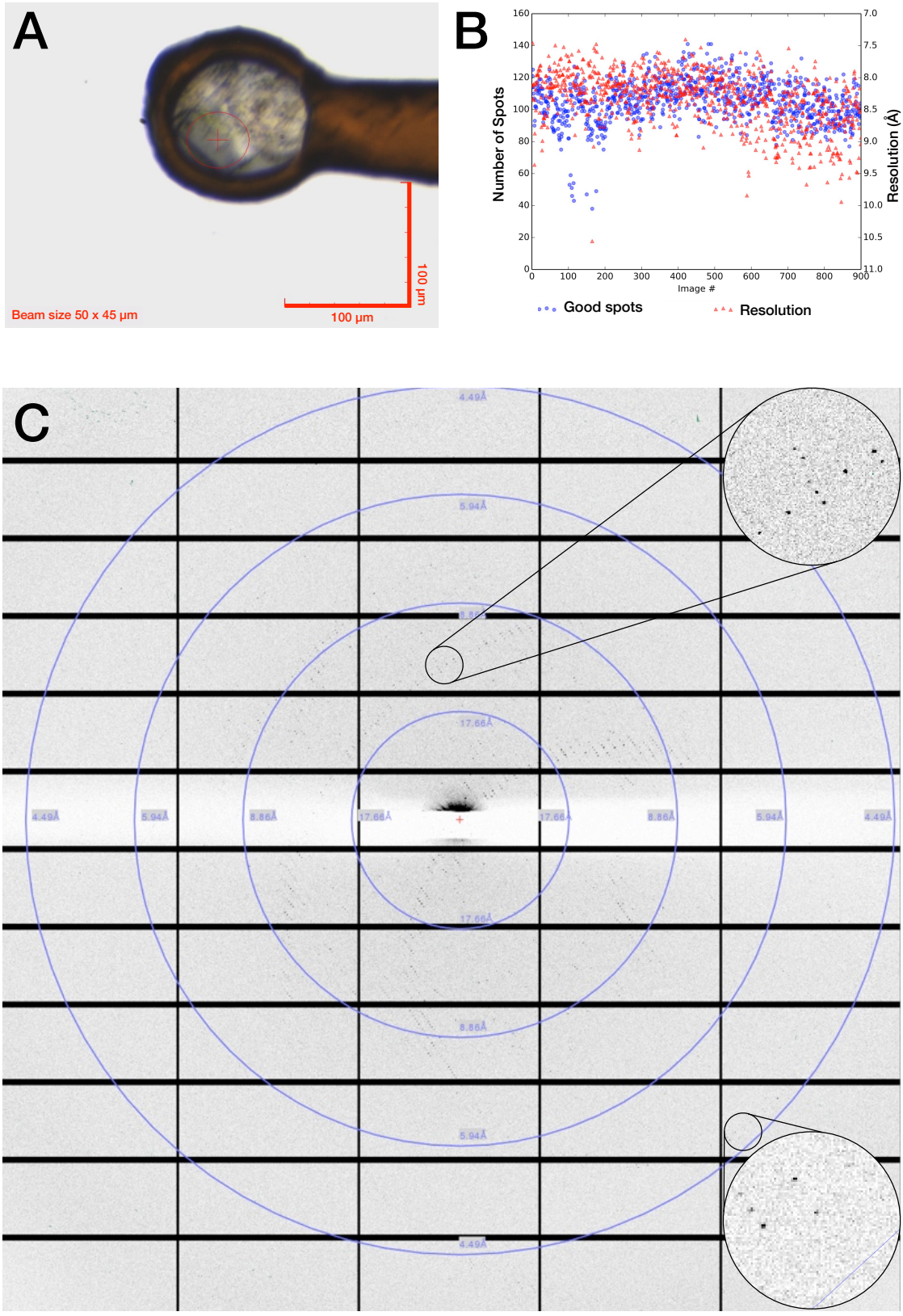


Figure 3.5. Data collection for best orthorhombic crystal.
(A) Image of this crystal. The red circle indicates the beam size and position. **(B)** DISTL plot of the number of spots (blue) and estimated resolution (red) for each image collected. **(C)** A representative detector image of the diffraction from this crystal.

The crystal diffracted to 4.3 Å, with a $CC_{1/2}$ of 0.500. The unit cell dimensions were similar to those of the very first crystal of $FluPol_C$ that I obtained (see section 3.2.1), indicating that there were likely to be four heterotrimers in the asymmetric unit with a solvent content of 65%. This was confirmed after the crystal structure was solved by molecular replacement in *phaser* (McCoy *et al*, 2007), searching using the built model of $FluPol_C$. Phase improvement was carried out using NCS averaging and density modification, in *GAP* (Stuart and Grimes, unpublished). B-factor sharpening was not performed in this case. The resulting electron-density map was good enough to resolve secondary structure and some residue side chains (figure 3.5). Though the overall level of detail was still lower than that for the $P4_32_12$ space group, the $P2_12_12_1$ map did help to resolve some ambiguities in the connectivity of the model. However, this was not sufficient to allow all residues of the polymerase subunits to be built. Indeed the connectivity of the chains toward the C-terminus of PB1 and the N-terminus of PB2 was still not clear. At the time of writing, the best model comprised 640/754 residues of PB1, 669/774 of PB2 and 690/709 of P3.

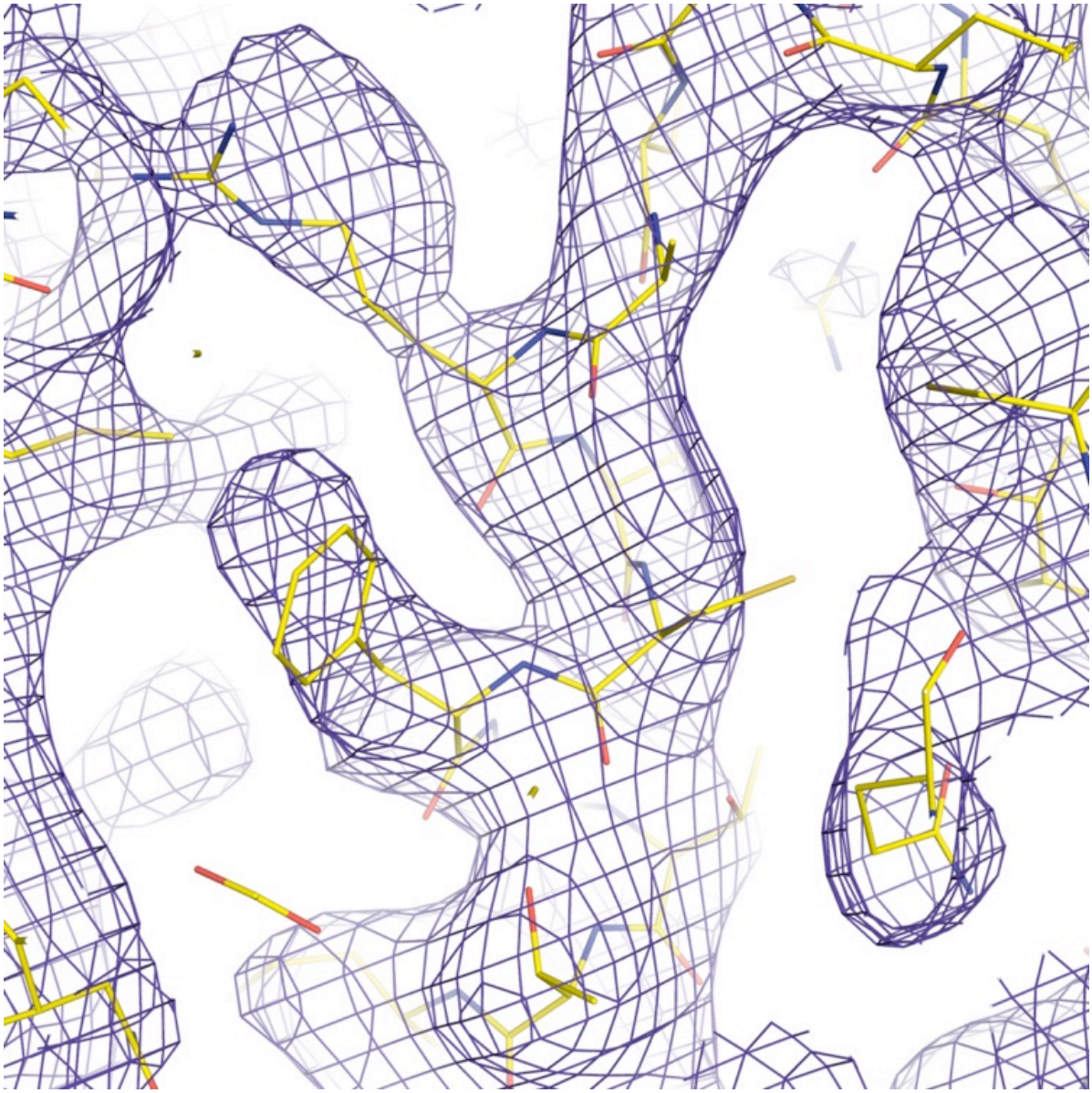


Figure 3.6. Electron density map of orthorhombic crystal form.
Model for the $P2_12_12_1$ structure of $FluPol_C$ (yellow sticks), showing the equivalent region as in figure 3.4, overlaid with the corresponding 4.3 Å, unrefined electron density map (blue) contoured at σ1.0.

3.2.8 Refinement

Given the limited resolution of the data, only partial refinement was possible. The model was refined independently against data from both space groups, using *BUSTER* (Global Phasing). Initially, only rigid-body refinement was performed, but atomic positions and B-factors were refined in subsequent cycles. Tight NCS restraints were employed. This was possible as the polymerase heterotrimer did not adopt different conformations within the two crystal forms; the average RMSD between all pairs of molecules across both space groups was only 0.94 Å, as calculated using *SHP* (Stuart *et al*, 1979). Phase restraints were also employed to prevent model bias dominating the refinement. Ideally, the model would have been simultaneously refined against data from both space groups, but the software currently does not allow this. The refinement statistics are reported in table 3.12. The B-factors are lower for the tetragonal dataset, because the input data were B-factor sharpened. B-factor sharpening was also automatically applied during refinement in *BUSTER*.

Table 3.12. Refinement statistics.

	Tetragonal	Orthorhombic
Space group	$P4_32_12$	$P2_12_12_1$
Resolution Range (Å)	41.7 – 3.6	49.5 – 4.3
Number of reflections	96,160	82,413
R_{work}/R_{free}	0.36/0.38	0.29/0.35
Number of atoms	27,336	54,292
B-factors		
Overall (Å²)	22	63
Wilson (Å²)	15	40
RMSDs		
Bond lengths (Å)	0.011	0.014
Bond angles (°)	1.36	1.56
Total real-space correlation coefficients	over 4014 residues:	over 4858 residues:
Main chain	0.85	0.90
Side chain	0.72	0.76

3.3 Summary

In this chapter I have described my efforts to solve the structure of FluPol, by x-ray crystallography. Extensive crystallisation experiments with two variants of $FluPol_A$ did not yield reproducible crystals. However, I did produce a panel of nanobodies to aid the crystallisation of $FluPol_{A/Fj}$, which I was not able to fully test. Hence, there are still avenues to explore with the crystallisation of this polymerase. Similarly, though I did not obtain any crystals of $FluPol_B$, my efforts were far from exhaustive, so here as well, there is much more that can be done.

Fortunately, with $FluPol_C$, I was successful. The protein crystallised in two different space groups: $P4_32_12$ (tetragonal) and $P2_12_12_1$, These contained two and four heterotrimers in the asymmetric unit respectively, hence provided a total of six views of the polymerase. The best diffraction dataset was recorded from one of the tetragonal crystals, which diffracted to 3.6 Å. A lead-derivatised crystal also diffracted well, but generated a dataset that contained only a weak anomalous signal. The diffraction from all crystals examined was extremely anisotropic. The structure of the tetragonal form was originally solved by SIRAS, using anomalous data from gold-derivatised crystals and "native" data from the lead-derivatised crystals. This was used in the initial rounds of model building, though later stages employed the higher resolution, 3.6 Å map. The anomalous signal of the gold-derivatised crystals was only detectable up to 8.9 Å, hence generating interpretable electron density maps relied heavily on density modification techniques to extend phase information to higher resolutions. In particular, the high solvent content of the crystals (76%), meant that solvent flattening was of great benefit. The structure of the orthorhombic form was solved by molecular replacement, using the the tetragonal structure as the search model. Again, solvent flattening and NCS averaging were employed, to improve the phases

obtained from molecular replacement. As the resolution of the data were low, model building was difficult. This is reflected in the relatively high R-values obtained after refinement. However, key structural features, such as the arrangement of the functional domains and the locations of catalytic residues, were able to be modelled with confidence. This represents a major advance in our understanding of the overall structure of FluPol. Before this work, the only views we had of the complete polymerase hetereotrimer were from much lower resolution EM reconstructions of $FluPol_A$ (Martín-Benito *et al*, 2001; Area *et al*, 2004; Torreira *et al*, 2007; Coloma *et al*, 2009; Resa-Infante *et al*, 2010; Moeller *et al*, 2012; Arranz *et al*, 2012). Attempts to define subunit localisation had been made by fragment docking (He *et al*, 2008; Moeller *et al*, 2012) and antibody-labelling approaches (Area *et al*, 2004; Moeller *et al*, 2012), but the low resolution of the reconstructions severely limited the mechanistic insights that could be gained from this. With the crystallographic structure of $FluPol_C$, one can see in unprecedented detail, the positions of all functional domains within the polymerase. Thus, we are finally able to suggest molecular mechanisms of transcription and elongation for the influenza virus. These will be discussed in the following chapter.

Structure

4.1 Overall Structure

4.1.1 Comparison with EM Reconstructions

Overall, $FluPol_C$ is a relatively compact, elongated particle, with its longest axis measuring ~130 Å (figure 4.1a), and its shortest ~70 Å (figure 4.1b). This explains why $FluPol_C$ particles appeared to be rather heterogeneous when viewed by both cryo- and negative-stain EM. The polymerase particle contains a conspicuous deep groove, running along most of its length. When looking through this groove as in figure 4.1b, the particle has a doughnut-like cross section, however its three-dimensional shape is not simply tubular. Three protuberances emerge from the central tube-like density around the groove. The polymerase therefore has a distinctive appearance (almost rabbit-like from certain perspectives), with features that should be evident even at low resolution.

The highest resolution EM reconstruction of $FluPol_A$, is currently 13 Å (Moeller *et al*, 2012). In this model, the polymerase appears to be split into a large body domain, and a peripheral arm domain (figure 1.6d). These features are not recognisable in the crystallographic model for $FluPol_C$. To gain a more objective assessment of the similarity between these, I used *Phaser* (McCoy *et al*, 2007) to search for molecular-replacement solutions of $FluPol_C$ against structure factors calculated by placing the EM density (EMD-2212, 0.0065 cutoff) into a cubic *P*1 cell ($a = b = c = 279$ Å). *Phaser* did not find a unique solution, confirming that the structural envelopes of the crystallographic and EM models were not similar. Additionally, *MOLREP* (Vagin & Teplyakov, 2010) did not produce

a convincing fit of the crystallographic coordinates into the EM map (figure 4.2a and b). A lower resolution reconstruction of $FluPol_A$, generated from negative-stain EM data (Torreira *et al*, 2007), did initially appear to share similar features with the crystal structure of FluPol; having a hollow centre and several protuberant features. However, on closer inspection, the similarity was again not convincing (figure 4.2c and d). Finally, I tried to fit the crystal structure of $FluPol_C$ into the EM reconstructions of RNP complexes (Arranz *et al*, 2012; Moeller *et al*, 2012), both manually and with *Situs* (Wriggers, 2012). In this case as well, it was not possible to dock the crystal structure in a way that compellingly accounted for the EM density.

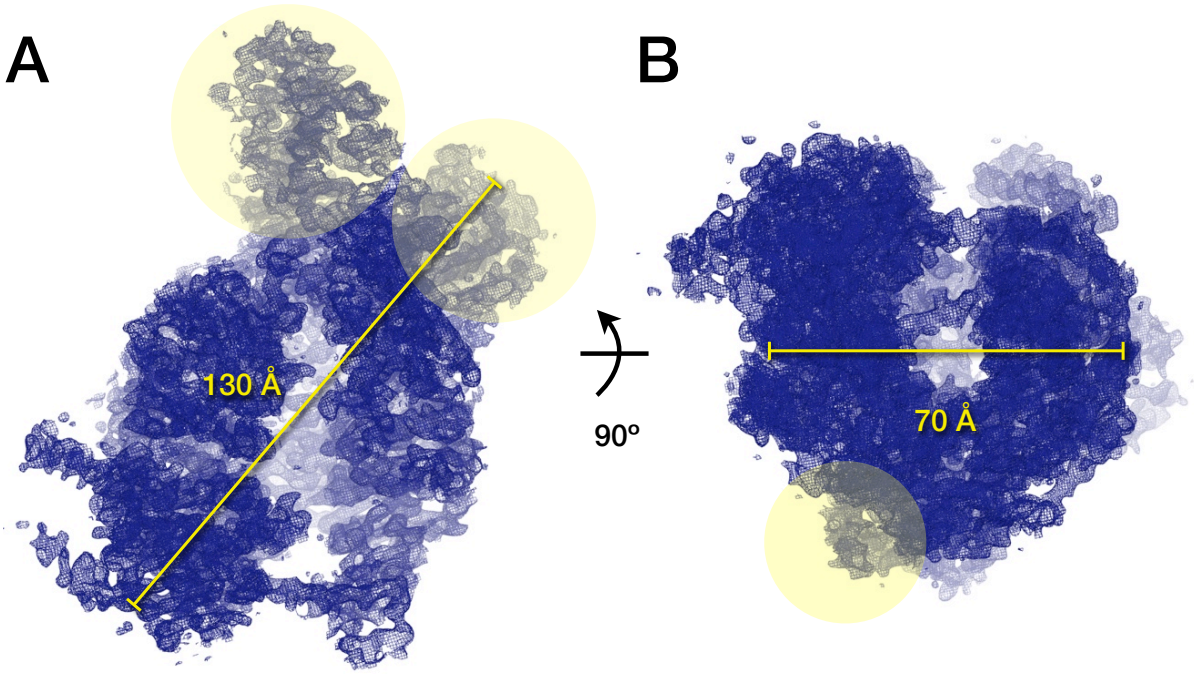


Figure 4.1. Overall structure of $FluPol_C$
(A and B) Two views of the electron density for the $FluPol_C$ heterotrimer. Features protruding from the main body of the particle are highlighted in yellow.

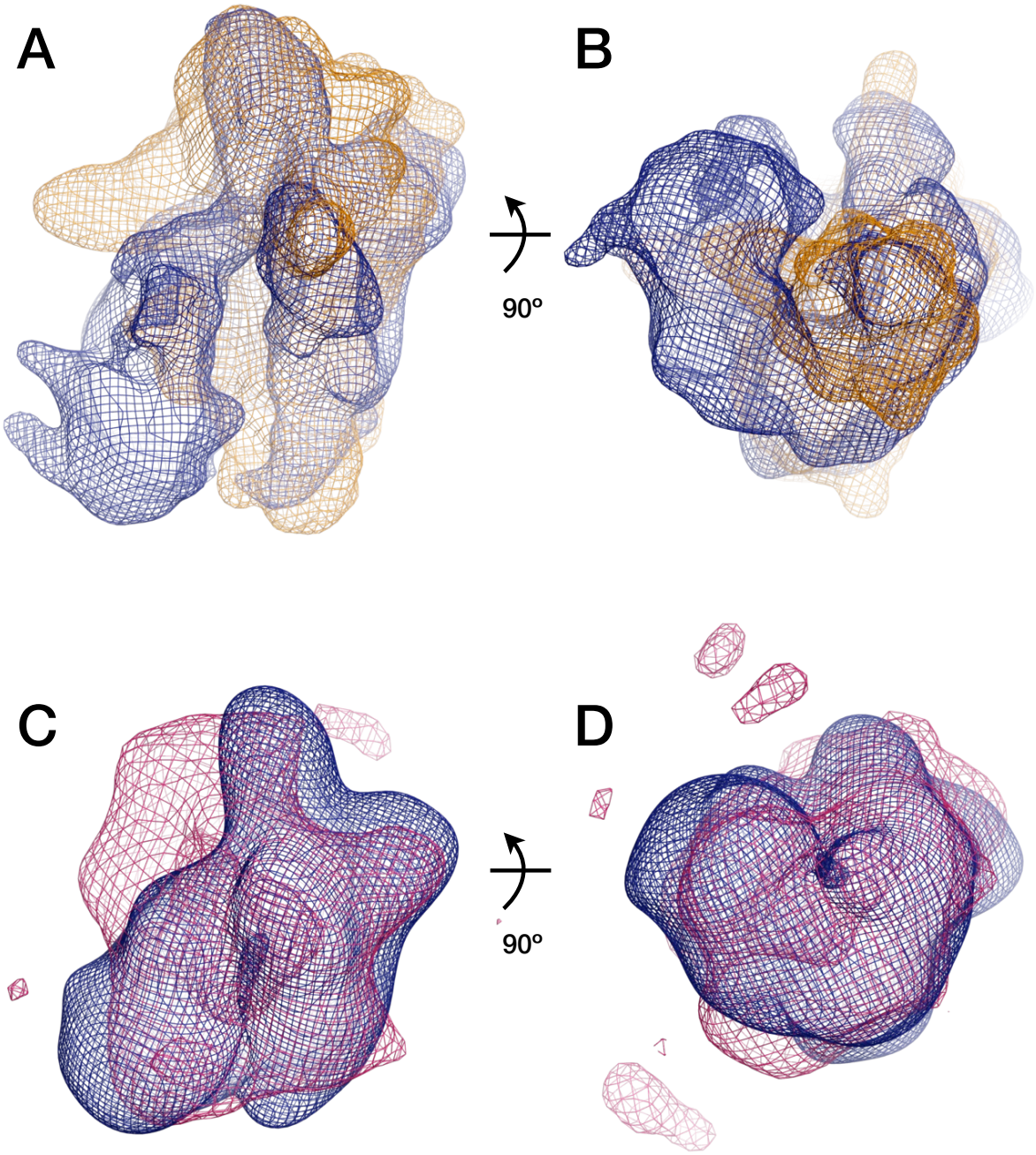


Figure 4.2. Comparisons with EM reconstructions.
(**A and B)** Docking of the crystal structure of $FluPol_C$ (blue mesh) into a 13 Å cryo-EM reconstruction of $FluPol_A$ (orange mesh) (Moeller *et al,* 2012). Both surfaces are contoured at σ1.5. The docking was performed in *MOLREP* (Vagin & Teplyakov, 2010). (**C and D**) Manual fitting of the $FluPol_C$ crystal structure into a 26 Å negative-stain EM reconstruction of $FluPol_A$ (pink mesh) (Torreira *et al,*2007), both contoured at σ2. In all panels, the crystal structure is shown as a Gaussian mesh, calculated at the resolution of the corresponding EM reconstruction. The views are equivalent to those in figure 4.1.

4.1.2 Arrangement of Subunits and Domains

As expected (section 1.3.1), PB1 is the central subunit of the polymerase heterotrimer (figure 4.3a and b). This adopts the canonical right-hand-like polymerase fold, possessing palm, fingers and thumb subdomains. Residues from these subdomains line the middle portion of the deep groove seen in the overall particle, and the catalytic aspartates (R446 and R447) lie at the base of this groove (figure 4.3c). PB1 is sandwiched by the N-terminal endonuclease and C-terminal domains of P3 ($P3_{Endo}$ and $P3_C$, respectively, figure 4.3d). These domains are connected by an extended linker, which runs along the outside of the fingers subdomain of PB1 (figure 4.3b and c). Thus PB1 and P3 form an intimate association, with P3 resembling a pair of headphones wrapped around PB1. This arrangement accounts nicely for the difference in protease sensitivity of PA when it forms a complex with PB1 (Guu *et al*, 2008).

PB2 is a roughly triangular plate that stacks against one face of PB1, and contacts both domains of P3 (figure 4.3b and d). The PB2 cap-binding domain ($PB2_{Cap}$) is located near the base of the palm subdomain of PB1 and protrudes from the central body of the polymerase particle (figure 4.3b). Similarly, the PB2 627-domain ($PB2_{627}$) points outward, with residue 649 (which corresponds to PB2 residue 627 of $FluPol_A$), lying right at the outermost tip of this domain, distant from the catalytic core of the polymerase (figure 4.3b and d). The PB1-PB2 interaction domain is positioned in a relatively exposed location at the top of the fingers subdomain of PB1 (figure 4.3a and b).

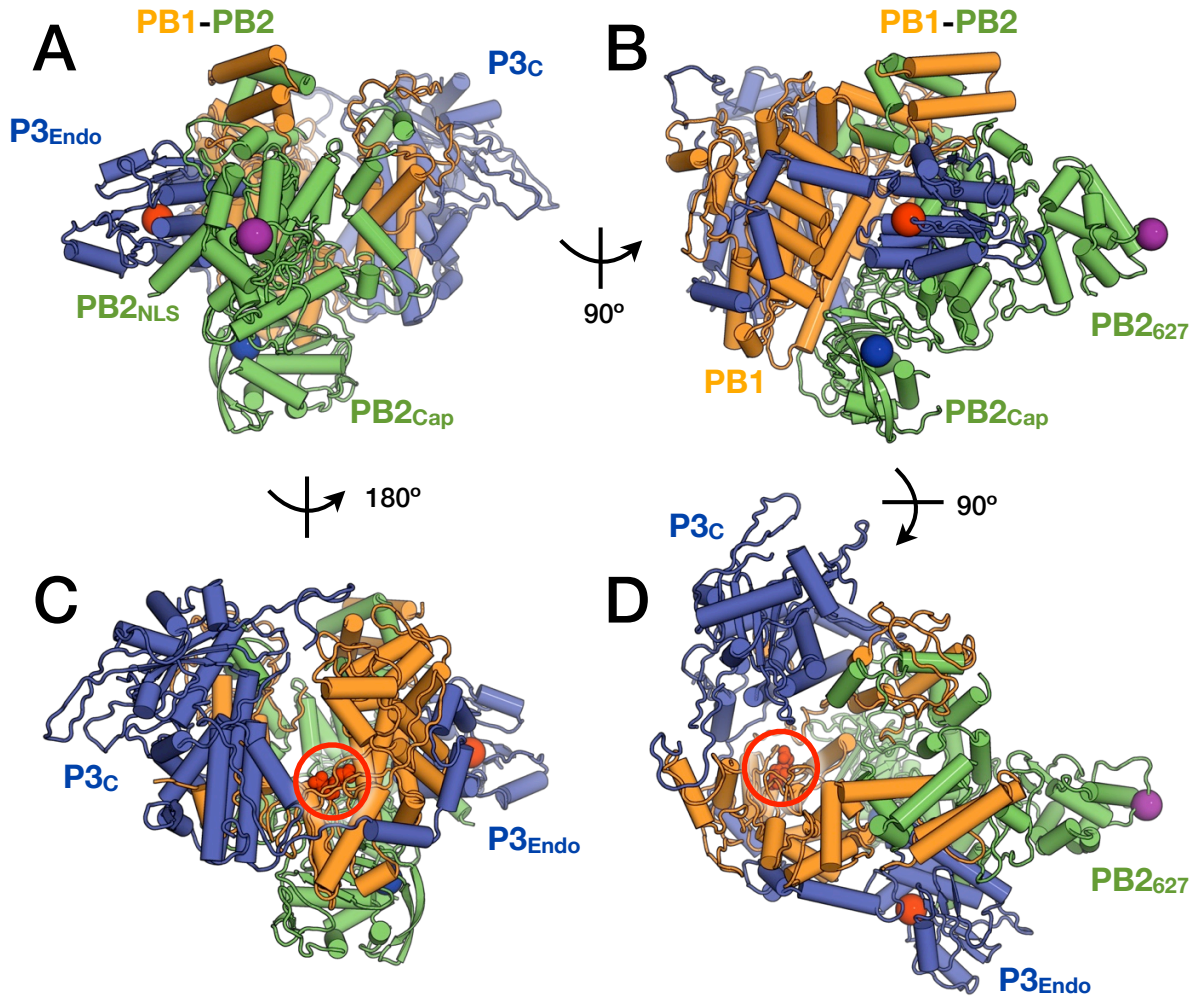


Figure 4.3. Structure of FluPol$_C$.
(A – D) Four views of the structure of FluPol$_C$, coloured according to subunit (PB1, orange; PB2, green; P3, blue). Visible domains are labelled and positions of interest are marked with coloured spheres (endonuclease active site, red; cap-binding pocket, blue; PB2 residue 649 (equivalent to 627 in FluPol$_A$), purple, catalytic aspartate residues, red with red circle).

4.1.3 Subunit Interactions

The interaction domains already known from FluPol$_A$ (He *et al*, 2008; Obayashi *et al*, 2008; Sugiyama *et al*, 2009) are also present in FluPol$_C$. Thus, the N-terminus of PB1 interacts with P3$_C$ and the C-terminus of PB1 interacts with the N-terminus of PB2. The regions of PB1 and PB2 that interact with each other may be more extensive than previously thought, because the relevant termini of these subunits both stretch into an area opposite the fingers subdomain on the other side of the particle groove (where the thumb

subdomain of PB1 is located). Unfortunately, the maps are particularly poor in this region, making it difficult to offer a more precise structural interpretation. In addition, the $PB2_{NLS}$ domain of $FluPol_C$ (approximately from residue 690 onwards), is close to a loop extending from the C-terminus of PB1 (residues 690 – 710, figure 4.4a, red circle). This may account for the observation that both the N- and C-terminal regions of PB2 interact with PB1 (Poole *et al*, 2004). $PB2_{NLS}$ and the C-terminus of PB1 are also close to a helix in $P3_{Endo}$ that contains residues 146 – 160 (figure 4.4a, shaded yellow circle), equivalent to PA residues 163 – 178 that have been reported to modulate cRNA promoter binding (Maier *et al*, 2008). This suggests that these residues act through an indirect mechanism. Unexpectedly, the NLS-domain of PB2 also contains a helix that packs against $P3_{Endo}$, in a position close to where a flexible loop would be, in $FluPol_A$ (figure 4.4b). I cannot yet comment on the chemistry of these potential interaction sites, because the amino acid sequence has not been fully assigned to these regions.

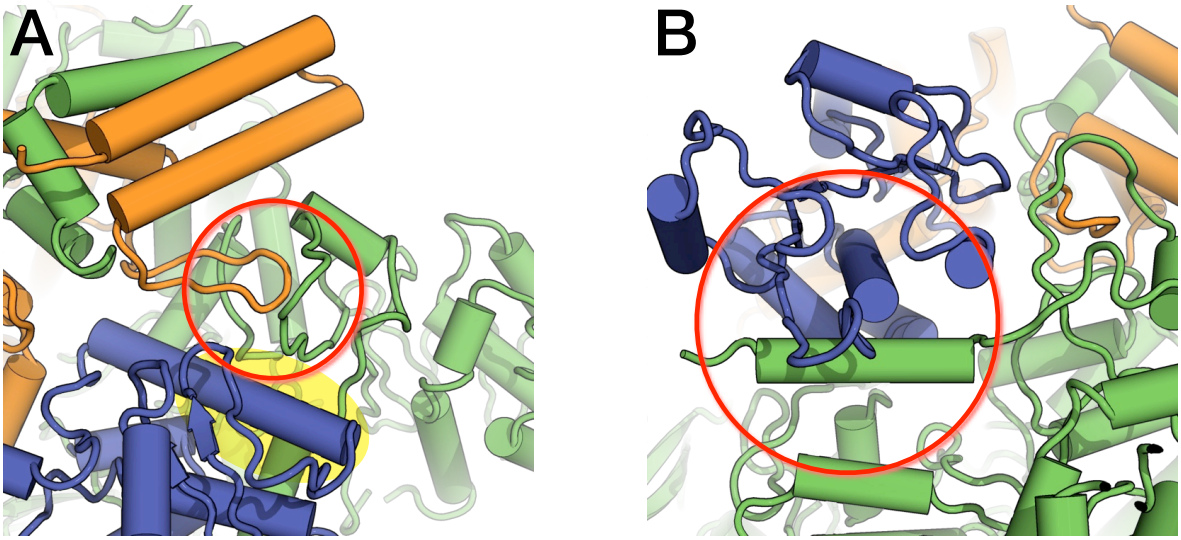


Figure 4.4. Novel subunit contacts.
(**A**) View of potential contact (red circle) between a loop from PB1 (residues 690 – 710) and a length of $PB2_{NLS}$ (approximately residues 710 – 720). P3 residues 146 – 160, which also lie close to this region, are highlighted by a shaded yellow circle. (**B**) View of helix from $PB2_{NLS}$ (approximately residues 730 – 750), which lies against $P3_{Endo}$ (red circle).

The previously described interaction domains appear to act as molecular clamps securing the complex, as they contain interlocking helices from different subunits. However, there are also extensive contact surfaces between the three subunits (table 4.1). The relative sizes of the various contact areas, supports the PB1-PA dimer model of FluPol assembly (Fodor & Smith, 2004; Deng *et al*, 2005), as the area between PB1 and P3 is the largest out of any of the contact surfaces. The structure also confirms the view that PB1 is the backbone subunit of the polymerase, responsible for holding together PB2 and PA/P3 (section 1.3.1). This is because the area between PB2 and P3 is the smallest of the contact surfaces.

Table 4.1. Contact surface areas between subunits in Å^2, calculated with *PISA* (www.ebi.ac.uk/PDBe/PISA).

	PB1	PB2	P3
PB1	–	3200	4900
PB2	3200	–	1400

These values are only qualitative at present, as several sections of the structure are still only modelled with polyalanine.

4.2 Structure of Individual Subunits

4.2.1 PB1

The central portion of PB1 (approximately residues 100 – 650) displays the right-hand-like polymerase fold, while the N- and C-termini form the subunit interaction domains (figure 4.5a). The configuration of the PB1 core broadly matches that predicted from the position of the conserved polymerase motifs (outlined in figure 1.10), with a fingers-palm-thumb arrangement, from N to C. There are no fingertips (loops from the fingers connecting to the thumb) obviously visible, so the polymerase catalytic core is not completely encircled. This may have important consequences regarding the mechanism of elongation, which will be discussed later.

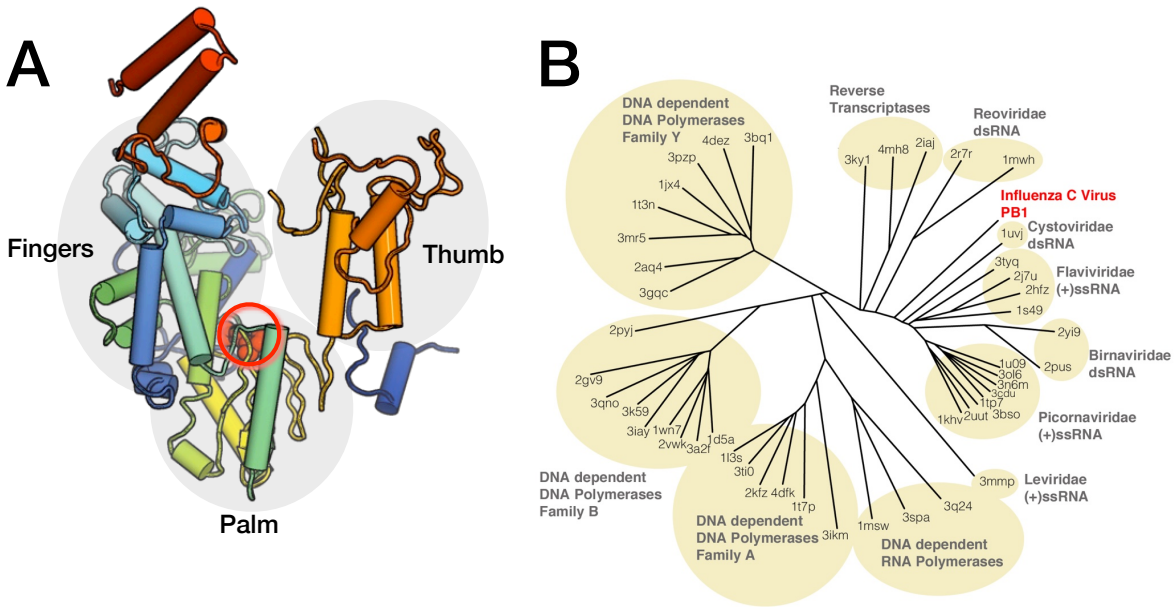


Figure 4.5. PB1 structure.
(**A**) Structure of PB1, in the same orientation as figure 4.3a, coloured from blue to red according to position in sequence. (**B**) Structure-based phylogenetic tree, showing the relationship of FluPol$_C$ to the other right-handed polymerases. Branches are labelled with the PDB accession code of the polymerase. Groupings of viral RdRps are labelled by viral family, while other polymerases are labelled according to their type. Pairwise comparisons were performed using *SHP* (Stuart *et al,* 1979) and the tree was generated using *PHYLIP* (Felsenstein, J., version 3.6.9). The list of polymerases used in the comparison was obtained from (Monttinen *et al,* 2014).

A structural comparison of PB1 against all other right-handed polymerases that have been solved by crystallography, revealed that $FluPol_C$ lies between the RdRps of the *Reoviridae* and *Cystoviridae* viral families (figure 4.5b). Interestingly, these are both families of double-stranded RNA viruses.

4.2.2 PB2

In addition to the previously described domains of PB2 (N-terminal PB1-interaction domain, $PB2_{Cap}$, $PB2_{627}$ and $PB2_{NLS}$), my work reveals the structure of a central body domain ($PB2_{Body}$), formed by residues between the PB1-interaction domain and $PB2_{Cap}$ (figure 4.6a). $PB2_{Body}$ does not seem to have any function beyond a purely structural one. As well as this, there is a possible second PB1-interaction domain lying just N-terminal of $PB2_{Body.}$ However, as mentioned above, it is difficult to get an unambiguous interpretation for this latter region, due to the poor quality of the electron density map close to the thumb subdomain.

$PB2_{Cap}$ and $PB2_{627}$ of $FluPol_C$ are both structurally very similar to their $FluPol_A$ counterparts. Both domains also project out from the polymerase particle, with $PB2_{627}$ seeming particularly exposed. This is consistent with hypotheses that suggest that the 627E/K polymorphism affects an interaction of $PB2_{627}$ with another protein or nucleic acid factor (section 1.3.6). The location of the site of the polymorphism, right a the tip of the domain, seems to rule out suggestions that it directly influences biochemical enzyme kinetics of the influenza polymerase (Aggarwal *et al*, 2011).

$PB2_{NLS}$, in contrast with the previous domains, differs somewhat from the corresponding region of $FluPol_A$. Firstly, the location of the domain, with respect to $PB2_{627}$, is very different (figure 4.6b). However this is not entirely surprising, given that the linker

between $PB2_{627}$ and $PB2_{NLS}$ was not resolved in the $FluPol_A$ structure (Tarendeau *et al*, 2008), indicating that this is relatively flexible. Secondly, the fold of the domain is rather different, most strikingly because of the appearance of a prominent helix in $PB2_{NLS}$ of $FluPol_C$. The basic residues of $FluPol_C$ PB2 that, by comparison with $FluPol_A$, form the bipartite NLS itself (748 – 751/752 and 768 – 771, see figure 1.11b), lie either at the tip of this helix or are just beyond it and not visible in the crystal structure. Therefore, despite the differences, the structure of the $PB2_{NLS}$ domain in $FluPol_C$ is still consistent with it functioning as such.

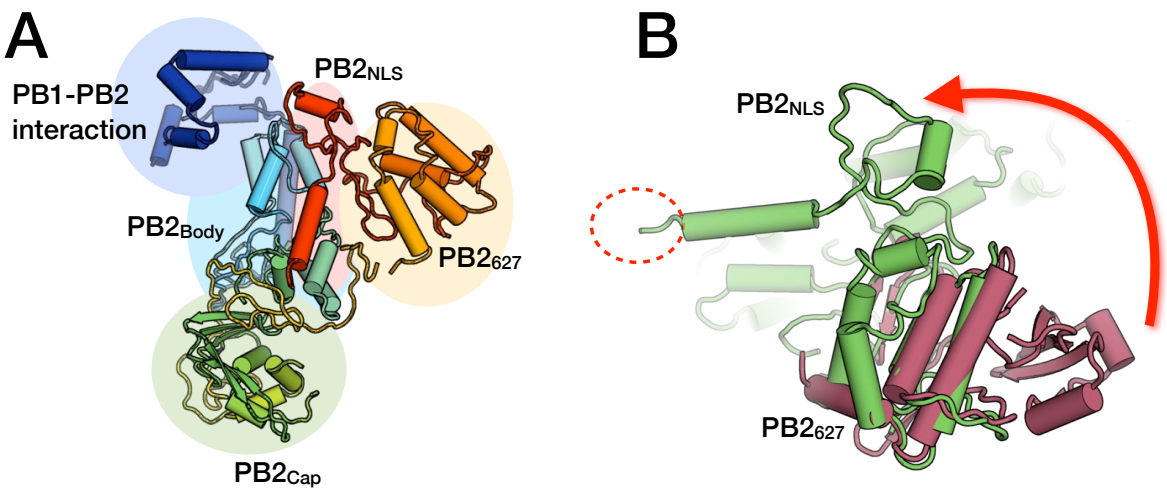


Figure 4.6. PB2 structure.
(**A**) Structure of PB2, in the same orientation as figure 4.3b, coloured from blue to red according to position in sequence, and with its various domains highlighted. (**B**) Closeup view of $PB2_{627}$ and $PB2_{NLS}$, (green) with the equivalent FluPolA structure (PDB ID: 2VY6) superposed on top (maroon). The arrow indicates the difference in position of $PB2_{NLS}$ between the two structures. The dashed red circle shows the likely position of the bipartite NLS.

4.2.3 P3

P3 is strikingly folded into two independent domains, lying on opposite sides of the polymerase particle and connected by a 70 residue, 70 Å extended linker (figure 4.7a and b). Both of these domains resemble their PA equivalents, with $P3_C$ also forming recognisable brain and mouth subdomains. $P3_C$ is located next to the polymerase core, mainly

reinforcing the thumb subdomain, but also contacting the fingers through a loop region. It has been previously noted that PA_C shares similarity with the N-terminal domain of the reovirus λ3 RdRp (He *et al*, 2008), and indeed $P3_C$ is found in a similar position, relative to the polymerase core, as the λ3 N-terminal domain (figure 4.7c and d). This hints at a possible distant evolutionary relationship between the two polymerases, which is supported by the finding that the core domains are also similar (figure 4.7e and 4.5b, PDB ID of λ3: 1MWH). On the other hand, the topology of helices and strands between PA_C/$P3_C$ and the λ3 N-terminal domain do not appear to be very similar, so an evolutionary relationship is by no means certain (see figure 1.16).

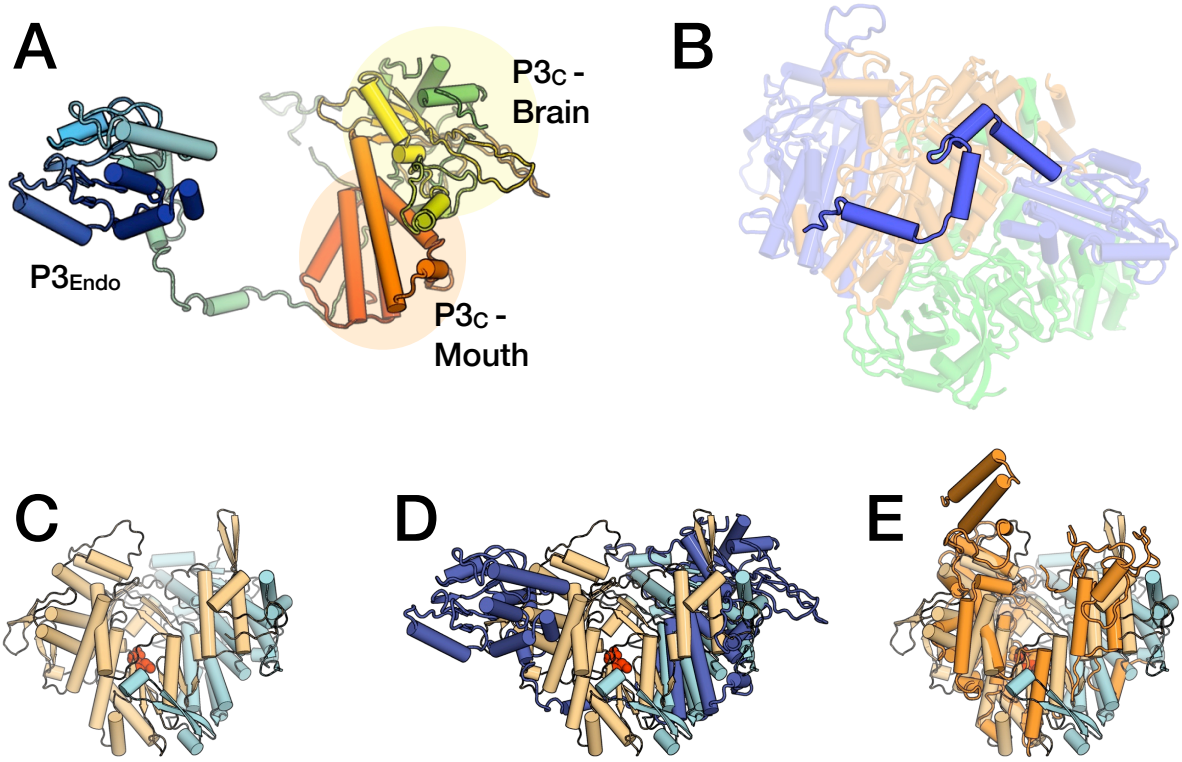


Figure 4.7. P3 structure.
(**A)** Structure of P3, in the same orientation as figure 4.3a, coloured from blue to red according to position in sequence, and with its various domains highlighted. (**B**) View of $FluPol_C$, highlighting the linker between $P3_{Endo}$ and $P3_C$, coloured as in figure 4.3. (**C – E**) Comparison with the structure of λ3 RdRp. (**C**) N-terminal (residues 1 –380) and core (residues 381 – 890) domains of λ3 RdRp (PDB ID: 3MWH), coloured cyan and light orange respectively. The catalytic aspartate residues are shown as red spheres. (**D**) Superposition of P3 (blue) onto the structure in the previous panel. (**E**) Superposition of PB1 onto the structure in panel C.

Like PA_C, $P3_C$ also displays a highly basic region on its surface, which in this case takes the form of a strip running from the top of the brain to a cleft between the two subdomains, and formed by amino acid residues: K262, R299, K301, K303, K314, K321, K521, R552 and K560 (figure 4.8a). This strip is angled towards the polymerase catalytic site, compatible with suggestions that it accommodates the 5′ vRNA promoter, when the polymerase is in a resting state in the vRNP (section 1.3.8). Indeed, several of the PA_C residues which were found to aid binding of svRNAs (Perez *et al*, 2012), have equivalents in the basic strip of $P3_C$ (table 4.2). This table also lists other mutations in PA_C that have known phenotypes, together with the equivalent residues in $P3_C$, and their position within $FluPol_C$. Several of these residue locations suggest a mechanism for the associated mutation. As well as the basic strip, $P3_C$ contains an acidic groove, on the opposite face of the brain subdomain, located between two loops that stick out from the polymerase particle (figure 4.8b). The functional significance of this groove is currently unclear, though some mutations in PA that affect polymerase activity, do map to residues around the groove.

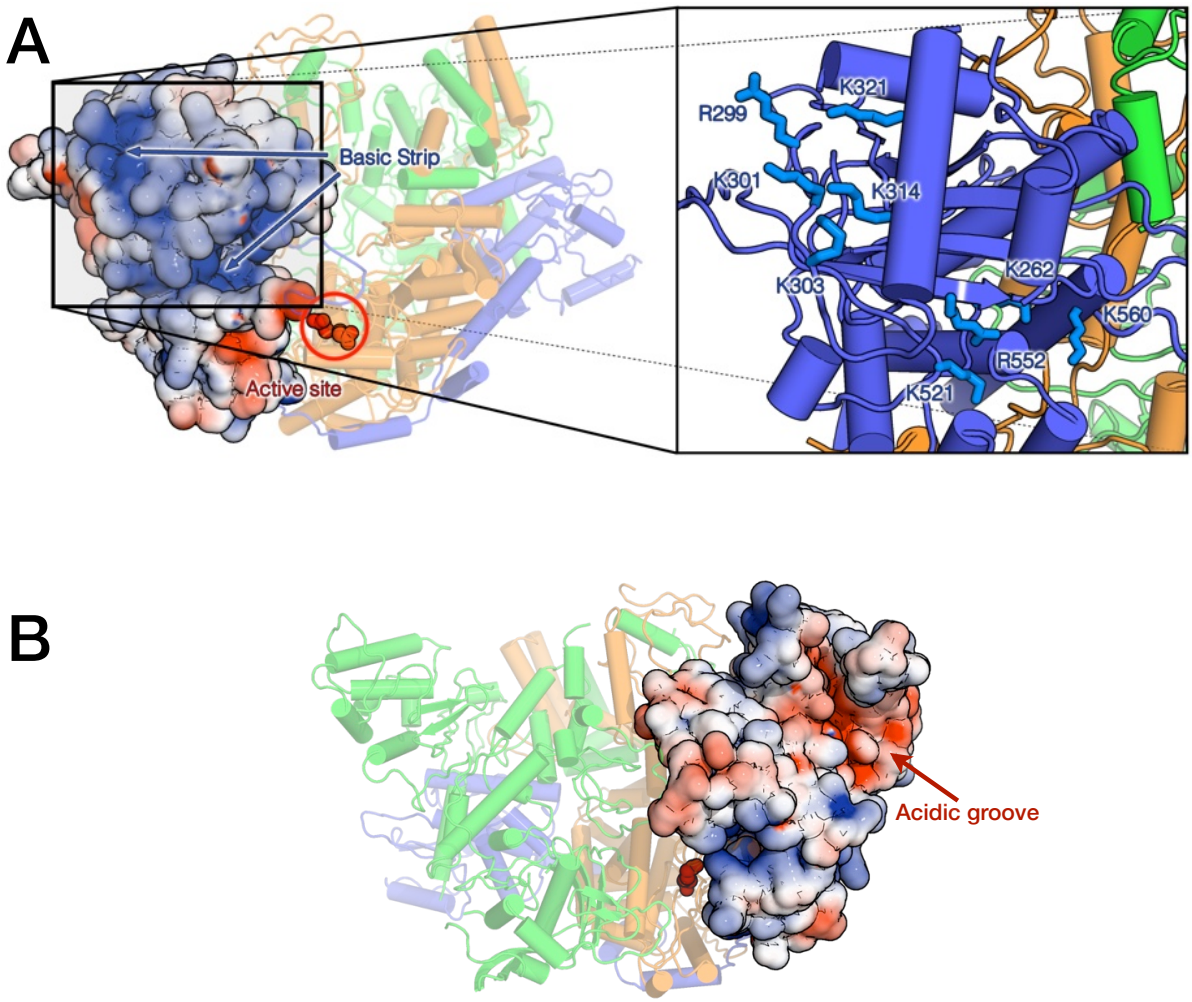


Figure 4.8. Surface electrostatics of P3C.
(**A and B**) Solvent-accessible surface representations of P3C, coloured by electrostatic potential from red (-4 k_bT/e_c) to blue (+4 k_bT/e_c), where k_b is the Boltzmann constant, T is temperature and e_c is the electron charge. Electrostatic potential was calculated using *APBS* (Baker *et al,* 2001). The other polymerase domains are shown in the background, with the catalytic aspartates highlighted, and coloured as in figure 4.3. A close-up of the basic strip is also shown, with the residues lining it, labelled.

Table 4.2 Analysis of some known mutations in PA

Residues found in the $P3_C$ basic strip are **bold**. Residues close to the acidic groove are ***bold italic***. Table adapted from *He et al, 2008*.

PA Mutants	Phenotypes	P3 Equivalent	Notes
E410A[1]	Poor polymerase activity	***E385***	Contributes to the acidic groove on the brain subdomain.
R442A, R443A[2]	Non functional polymerase	L418 N419	Located near the thumb subdomain of PB1, adjacent to a helix from either PB1 or PB2 (the map is poor here). May affect subunit-subunit interactions.
E493A, G494A[2]	Poor polymerase activity	***K469*** ***S470***	Located in a loop emanating from the brain subdomain, bordering the acidic groove and pointing outward from the polymerase particle.
G502A[1]	Non functional polymerase	G486	Located internally, on a β-strand within the brain subdomain.
G507A/ R508A[2]	Virus assembly defect	S491 K492	Located at the top of the brain subdomain, close to the basic strip. Possibly interacts with NP.
H510[1]	Very poor polymerase activity	**H494**	Located in basic strip.
T515A[3]	Found in highly pathogenic H5N1 viruses	D499	Located on a loop from the brain subdomain, pointing towards the polymerase core. Near the basic strip.
E524A[1]	Non functional polymerase	E508	Located internally, on a β-strand within the brain subdomain.
P530A, R531A[2]	Replication impaired polymerase	P514 N515	Located on a small helix, separating the acidic groove and basic strip.
K536A, W537A[2]	Non functional polymerase	K518 H519	Located on a helix, contributing to the basic strip.
K539A (and K328A)[1, 4]	Replication impaired polymerase	**K521** **K303**	Found within the basic strip. Contributes to 5′ vRNA promoter binding.
R566A[4]	svRNA binding defect	**R552**	Found within the basic strip.
K615N[5]	Stimulates polymerase activity, especially vRNA synthesis	E600	Located in the mouth subdomain, towards the base of the polymerase, pointing away from the core and not close to PB1.
E656A, G657A[1, 2]	Non functional polymerase	E640 G641	Located in the mouth subdomain, adjacent to PB1. Probably interferes with PB1 interaction.

[1] (Fodor *et al,* 2002), [2] (Regan *et al,* 2006), [3] (Hulse-Post *et al,* 2007), [4] (Perez *et al,* 2012hk), [5] (Gabriel *et al,* 2007).

The basic patch on PA_{Endo}, proposed to bind mRNA downstream of the endonuclease site (Xiao *et al*, 2014), is also present on $P3_{Endo}$ (figure 4.9). The structure of the complete heterotrimer supports the suggested role, because this patch is located on the side of the endonuclease active site that is furthest away from the PB2 cap-binding pocket. Hence mRNA bound at the cap-binding domain and this basic patch will naturally pass across the endonuclease active site.

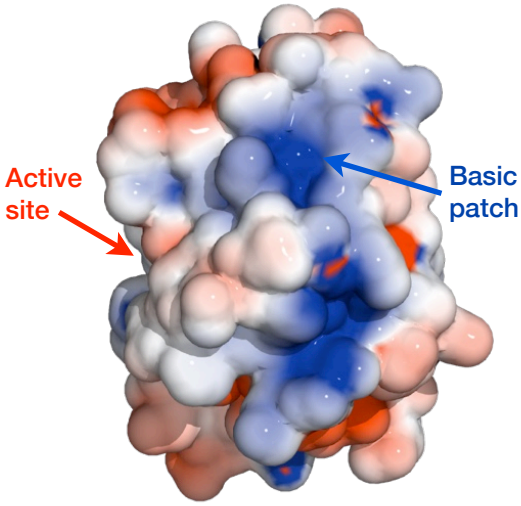


Figure 4.9. Surface electrostatics of $P3_{Endo}$. Solvent-accessible surface representation of $P3_{Endo}$, coloured by electrostatic potential as in figure 4.8.

4.3 From Structure to Mechanism

4.3.1 The Polymerase Catalytic Core

The core of the polymerase is formed by PB1, with $P3_C$ augmenting the thumb subdomain and $PB2_{Body}$ forming a plate that occludes one side of the core (figure 4.3b and d). At the base of the core, near the catalytic aspartates of the palm, there is a small tunnel that leads out to the cap-binding pocket, in PB2 (figure 4.10a). This tunnel provides the shortest connection between the cap-binding and polymerase sites (which are separated by ~40 Å), therefore defining it as the route of entry into the core, for cap-snatched primers bound by PB2 during transcription initiation. It seems likely that product exit also occurs through this tunnel, since the cap-binding site tethers the nascent product. Additionally, the position of the tunnel relative to the catalytic core matches that of the transcription-product-exit channel in the λ3 RdRp (Tao *et al*, 2002). In the crystal structure of $FluPol_C$, the tunnel is slightly too narrow to accommodate single stranded RNA, but only minor structural rearrangements would be required make it sufficiently wide. The channel is almost certainly too narrow to accommodate double stranded RNA, which provides a mechanism for separation of template and product strands.

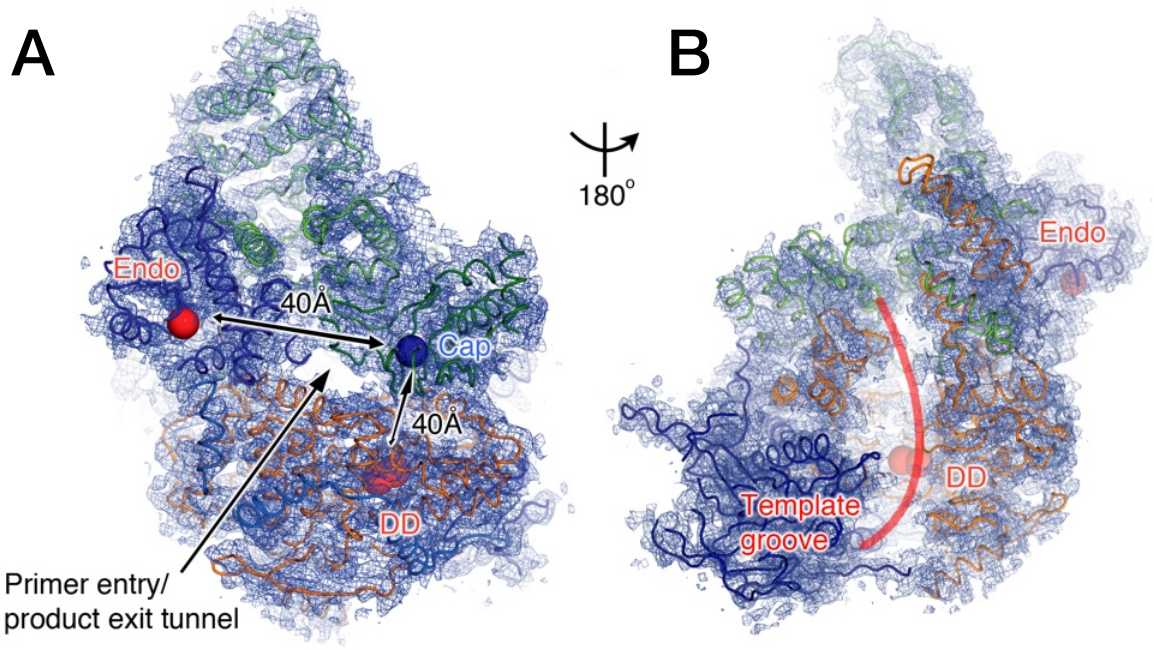


Figure 4.10. Key structural features of FluPol$_C$.
(**A and B**) Two views of FluPolC, with the electron density overlaid. Important structural features are highlighted and labelled.

Thus, while the RNA product leaves through the tunnel at the bottom of the core, the template likely both enters and exits back through the conspicuous groove on the top of the polymerase (figure 4.10b). This proposed trajectory places incoming template along the inside of the fingers subdomain, in common with the arrangement for all other right-handed polymerases that have been solved with template RNA bound (Mönttinen *et al*, 2014). However, in these other polymerases, the template would then exit through a channel in the back of the core, in a direction ~90° from which it entered (figure 4.11a). This is not possible in FluPol$_C$, as PB2 blocks this exit route (figure 4.11b). Unless a large conformational change occurs, which although unlikely, still cannot be ruled out, the only path available for template exit is back up, in the direction that it came in. This path has important consequences for the mechanism of polymerase elongation in the context of the RNP (see below).

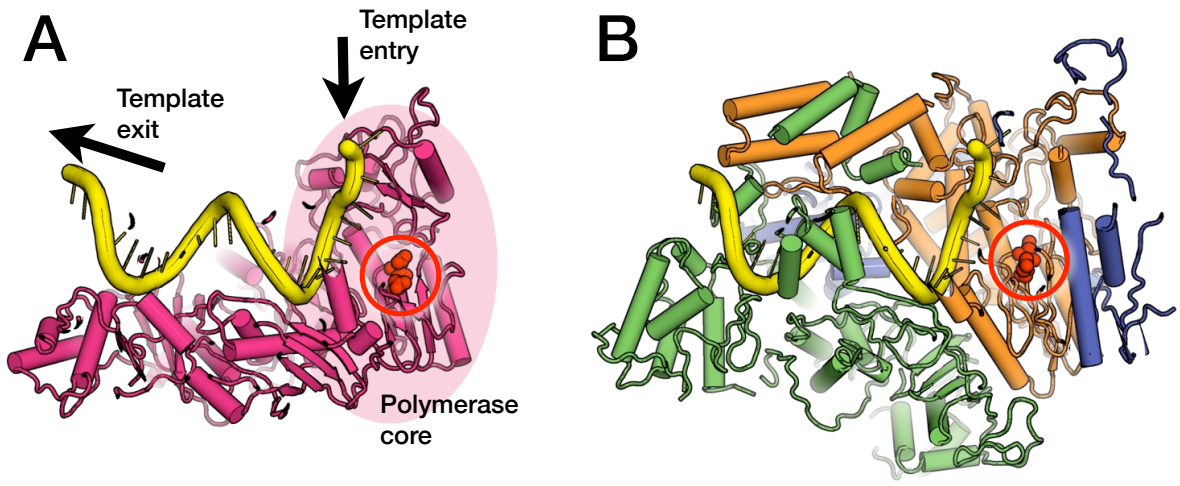


Figure 4.11. RNA template path.
(**A**) Structure of the HIV-I reverse transcriptase (PDB ID: 1RTD), in complex with DNA. The path of the template DNA strand is shown in yellow. This path is typical of right-handed polymerases. The catalytic aspartates are shown in red and circled. (**B**) Structure of FluPol$_C$, in the same orientation, with the template strand from the previous panel, superposed.

Finally, the catalytic aspartates of FluPol$_C$ are also accessible through a large channel formed by the interface of P3$_C$ with PB1 (figure 4.3c). This probably forms a substrate pore (Butcher *et al*, 2001), which allows nucleotides to easily reach the polymerase active site during elongation, when other channels are blocked by template and product RNA. Again, the location of this pore in FluPol$_C$ matches that of the substrate pore in other polymerases (Tao *et al*, 2002; Salgado *et al*, 2004).

4.3.2 Promoter Binding

In the resting state, FluPol will be bound to a double stranded vRNA or cRNA promoter, within an RNP complex (section 1.3.9). These RNAs are thought to initially bind the polymerase in a corkscrew conformation; a partial duplex in which both 5′ and 3′ strands are also internally base paired. As discussed above, the basic strip of P3$_C$ is likely to participate in binding the 5′ vRNA promoter. Additionally, several regions of FluPol$_A$ PB1 have been implicated in 5′ vRNA promoter binding (Li *et al*, 1998; González & Ortín, 1999; Jung & Brownlee, 2006). These encompass suggestions which we can now more confidently

map to both the fingers (R233, R238, R239 and R249) and thumb (R571, R572) subdomains. Binding in these regions would place the promoter somewhere within or near the deep groove in the polymerase particle. It is not clear whether all of the proposed binding sites are mutually compatible, though we can see that $P3_C$ abuts the PB1 thumb, so a continuous binding site formed by these two domains, could easily be envisioned.

The 3′ vRNA strand (which serves as the template during elongation) will be partially base paired to its 5′ partner, in a conformation that is not yet initiation competent. Currently, the only residues implicated in specifically binding 3′ vRNA are $FluPol_A$ PB1 residues F251 and F254 (Li *et al*, 1998). These reside in the fingers subdomain. This makes sense, given that the template strand will be guided into the polymerase core by interactions with the fingers subdomain. However, the importance of these residues has been disputed (Jung & Brownlee, 2006).

Unsurprisingly, without a crystal structure of a FluPol-RNA complex, it is impossible to precisely identify the promoter binding regions. The mode of cRNA binding is also still unclear.

4.3.3 Transcription Initiation

The conformation of $FluPol_C$ captured in the crystal structure suggests a model for primer-dependent transcription initiation (figure 4.12). This begins with resting polymerase bound to the vRNA promoter (i). The first step of initiation is capture of a cellular mRNA, through its cap structure, by $PB2_{Cap}$ (ii). The tethered mRNA may be guided to lie across the endonuclease active site by a basic patch on $P3_{Endo}$ (see figure 4.9). This mRNA is then cleaved, generating a 10–13 nt-long capped RNA (iii). The tethered end of cleaved mRNA, which becomes the primer, is free to dissociate from $P3_{Endo}$, because this stretch of RNA is

not bound by the basic patch (iv). The length of the primer is determined by the distance between the $PB2_{Cap}$ and $P3_{Endo}$ domains (~40 Å, figure 4.10a). This distance equals that between the cap-binding pocket and polymerase catalytic site, hence the primer generated is sufficiently long to span between these two regions, via the primer-entry / product-exit tunnel. Therefore, only minor structural rearrangements (slight opening of the tunnel) are required for the 3′ end of the primer to access the polymerase catalytic site and form the priming platform (v). The capped 5′ end of the primer can still remain tethered to $PB2_{Cap}$ at this stage, but will need to detach to allow elongation. It has been reported that this occurs after 11 – 15 nt of chain growth (Braam et al, 1983). In concert with primer generation, the 3′ vRNA promoter separates from the 5′ end bound in the basic cleft of $P3_C$ and moves into the polymerase catalytic site. With primer and template assembled in the catalytic centre of the polymerase, transcription can initiate (vi).

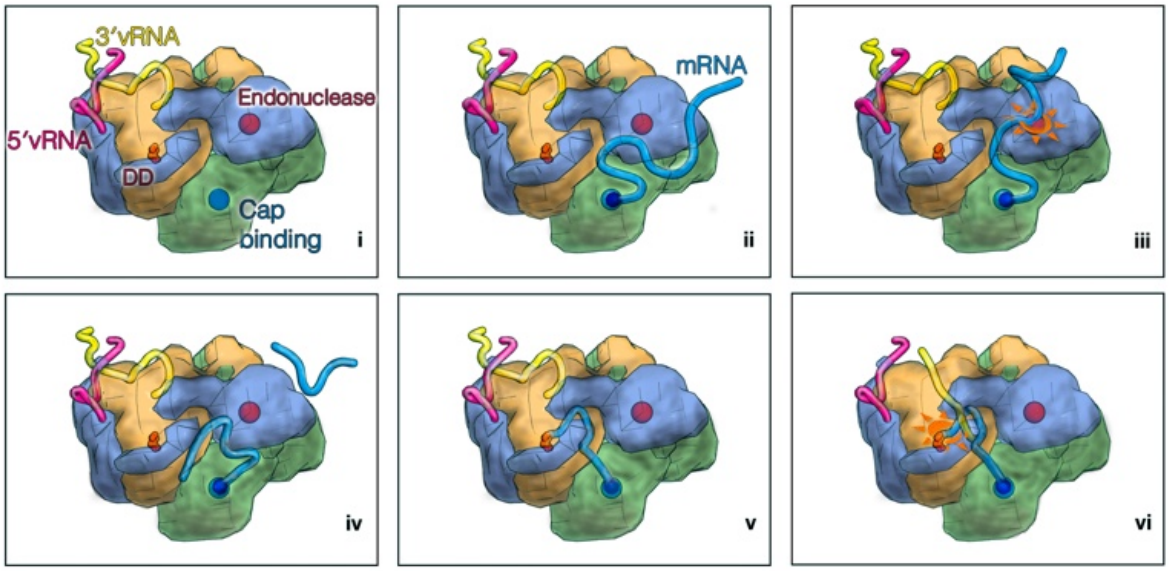


Figure 4.12. Model for transcription initiation.
Cartoon illustrating key points in the transcription initiation reaction mechanism. The proposed RNA paths are shown as tubes, with the mRNA cap structure represented as a blue sphere (the various binding sites are coloured as in figure 4.3). Catalytic activity is shown by an orange star. (**i**) Resting structure, with ends of vRNA bound. (**ii**) Binding of capped-mRNA by PB2. (**iii**) Cleavage of mRNA by $P3_{Endo}$, to generate a capped 10 – 13 nt-long primer. (**iv**) Release of 3′ end of cleaved mRNA. (**v**) Movement of 3′ end of primer into polymerase active site. (**vi**) Movement of 3′ end of vRNA into the polymerase active site, to base pair with the primer and form the catalytically competent initiation complex.

The mechanism for *de novo* initiation remains an enigma, since unlike proteins such as the Φ6 RdRp (Butcher *et al*, 2001), FluPol$_C$ does not possess a protein priming platform. This means that during replication, there is no obvious mechanism to stabilise an initiating nucleotide. The initiation of replication thus requires FluPol to undergo a major conformational change, probably resulting in a portion of PB1 or PB2 (P3 is too far away) being brought into the active site to generate the initiation platform. It is currently not possible to offer any more detail into this process.

4.3.4 Elongation

Having initiated polymerisation, FluPol faces the problem of elongation from a template that is incorporated into an RNP complex. The structure of FluPol$_C$ is suggestive of how this might be accomplished. As discussed above, both upstream and downstream template RNA will likely be accommodated in the groove along the top of the FluPol$_C$ particle. The proposed trajectory places the exiting template RNA adjacent to the PB1-PB2 interaction domain, which could explain why mutations in this region can affect polymerase activity and yet not affect the actual interaction (Sugiyama *et al*, 2009). Thus template entry and exit both occur on the same face of FluPol$_C$. This mechanism, unique amongst viral polymerases, requires only a short length of RNA to be dissociated from the NP oligomer at any one time during elongation (figure 4.13a). Hence, FluPol$_C$ can elongate around the RNP, without disrupting its overall structure (4.13b). This proposal is consistent with the most recent model of the path of RNP-associated RNA (Arranz *et al*, 2012). In this, the RNA adopts a rather meandering route, but does not form a complete helix around each NP strand. This means that sections of RNA are able to be sequentially pulled off from the NP oligomer (in the manner shown in figure 4.13), without individual protomers of NP needing to dissociate from the strand. However, this model does require the 5′ promoter to

dissociate from FluPol after elongation initiation, meaning that polyadenylation must be explained with a separate mechanism. This is in stark contrast with the model proposed by Moeller et al, in which the 5′ promoter remains bound to FluPol, providing a mechanism for polyadenylation through stuttering, but requiring unwinding of the RNP, followed by re-encapsidation on the other side of the elongation bubble (Moeller *et al*, 2012). The Moeller model strikes me as energetically unlikely, because with each step along the template, the polymerase would need to disrupt all inter-strand NP-NP interactions (an idea of the extent of these interactions is visible in figure 4.13b). Indeed, any model in which the 5′ promoter remains bound throughout, will face this problem.

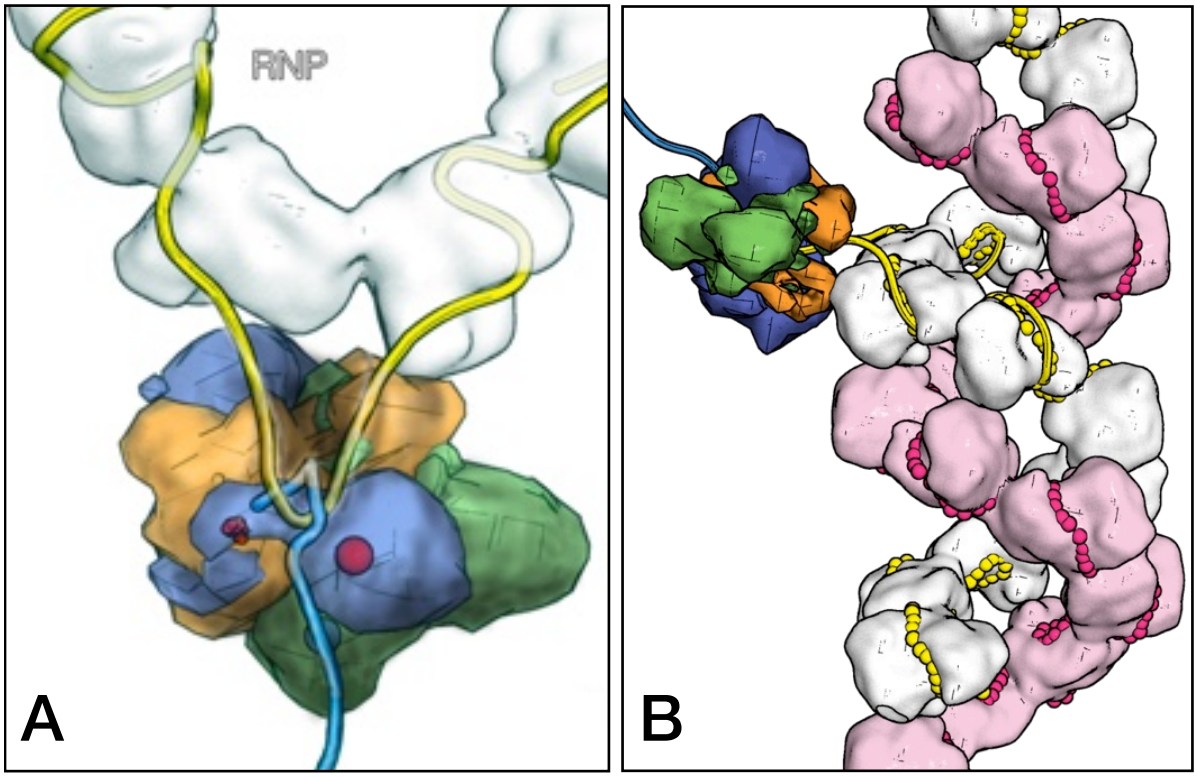


Figure 4.13. Model for FluPol elongation.
(**A and B**) Cartoons illustrating the proposed model for FluPol elongation. The polymerase and RNAs are depicted as in figure 4.12, while NP protomers of the RNP are shown in white or pink. In panel b, both NP strands of the RNP are shown, with the path of RNA also depicted with yellow and pink spheres. The RNP structure drawn here is based of the cryo-EM reconstruction from Arranz *et al,* 2012.

4.3.5 Elongation Termination

I suggest that the problem of transcript polyadenylation could be resolved by postulating that FluPol simply rebinds 5′ vRNA promoter as it re-approaches the top of the RNP. This would then allow stuttering on the poly-U tract near the 5′ end of the vRNA to take place. Supporting this idea is the observation that the 5′ vRNA promoter, on its own, can bind FluPol with high affinity (Tomescu *et al*, 2014). Also, the site of this interaction is separate from the polymerase active site, so both template and 5′ promoter can potentially be accommodated simultaneously. During replication, which occurs later in the infection cycle, polyadenylation does not occur. This could be due to the effect of free FluPol, which perhaps binds to exposed promoters and might prevent rebinding of the 5′ end by the elongating polymerase. However, the FluPol just invoked might itself cause a steric block on the progress of the elongating polymerase, emphasising that these suggestions are extremely speculative at present; much more work will need to be carried out to discover what is really happening.

4.4 Summary

In this chapter, I have outlined the structure of $FluPol_C$ and what this teaches us about the mechanism of the influenza polymerase. The structure solved at present is still not entirely complete, with several sections only able to be modelled as polyalanine and residue numbering not completely secure. This largely prevents the mapping of known functional mutations of $FluPol_A$ onto the structure of $FluPol_C$. Nevertheless, we are able to learn a great deal from the overall shape of the complex and the arrangement of functional domains within it. We see that the position of the cap-binding domain, relative to the polymerase and endonuclease active sites, could allow primer generation and initiation-complex formation to take place without major structural rearrangements. Thus, for the first time, we are able to offer a molecular description of the events of transcription initiation that is actually based on structural evidence. Similarly, I present a detailed model for polymerase elongation that is compatible with the most convincing current picture of RNP structure. This proposal might be generally applicable, because all nsRNA viruses have an NP-encapsidated genome (Ruigrok *et al*, 2011). The model does not require disassembly of the helical NP oligomer, which brings our understanding of influenza elongation into line with the prevailing wisdom for elongation in mononegaviruses such as Rabies (Albertini *et al*, 2006) and respiratory syncytial virus (Tawar *et al*, 2009). Even in viruses where vRNA runs entirely round the centre of a nucleoprotein ring, such as rift valley fever virus, replication and transcription of vRNA does not necessarily require complete disassembly of the ribonucleocapsid (Ferron *et al*, 2011). A caveat of the model is that it requires we abandon our view that 5′ vRNA promoter remains bound to FluPol throughout the elongation cycle. However, it is difficult to see in any case how transcription would proceed with the 5′ promoter attached.

The $FluPol_C$ structure provides supporting evidence for several existing hypotheses regarding FluPol biology. For instance, the values of inter-subunit contact areas reinforce the dimeric intermediate assembly model (Deng *et al*, 2005), while the location of the PB2-627-equivalent residue further shows that this mutation is unlikely to affect intrinsic polymerase activity. The location of $P3_C$ in the overall particle is compatible with suggestions that it binds the 5′ strand of the vRNA promoter, although I was not able to obtain a FluPol-vRNA complex structure to directly verify this. Additionally, as none of the views of $FluPol_C$ that were captured in the two crystal forms displayed any significant conformational changes, I cannot offer any illuminating comments about the structural flexibility that has been ascribed to FluPol (Torreira *et al*, 2007; Arranz *et al*, 2012; Moeller *et al*, 2012). However, since the crystal structure was not consistent with *de novo* replication, FluPol must at some point undergo major conformational changes.

The overall structure of $FluPol_C$ can be described as a compact, tubular polymerase core, formed by PB1, $P3_C$ and $PB2_{Body}$, with the functional domains involved in capping ($PB2_{Cap}$ and $P3_{Endo}$) jutting out from this core. This characterisation squares remarkably well with the EM-derived structural picture of RdRps from other segmented and non-segmented nsRNA viruses (known as L-proteins). EM reconstructions have been reported for Machupo virus L-protein (Kranzusch *et al*, 2010), which cap-snatches, and vesicular stomatitis virus L-protein (Rahmeh *et al*, 2012), which synthesises its own cap structures. In both of these RdRps, the polymerase domain forms a ring-like core, while the various capping domains are appended to this core via flexible hinge regions. Hence, like FluPol, L-proteins display a structural compartmentalisation that reflects the functional dichotomy of the RdRp. This hints that the polymerases of all nsRNA viruses evolved divergently, by attaching different capping machinery onto a common polymerase core (Kranzusch & Whelan, 2012). Confirmation of this hypothesis awaits an atomic-resolution structure of an L-protein core.

Conclusions and Future Work

5.1 Conclusions

At the start of this thesis I outlined the complexity of the task facing FluPol and the outstanding questions about its mechanisms (section 1.4.4). Now, with the molecular structure of $FluPol_C$, I can revisit these issues, highlighting how the structure addresses them and what remains unanswered.

The largest advance made has been in our understanding of the coordination of cap snatching. We now know the arrangement of the domains involved in this and thus can see how capped mRNA, bound by $PB2_{Cap}$, can be cleaved and threaded into the polymerase active site to initiate transcription. We also see that this can be achieved without the need for major conformational rearrangements of the polymerase. However, how these activities are controlled is still unclear, because I was not able to solve structures of FluPol with promoter RNA bound. The form of $FluPol_C$ crystallised should be inactive, as 5′ vRNA, at least, is thought to be required for cap binding and endonuclease activity (Hagen *et al*, 1994; Cianci *et al*, 1995; Leahy *et al*, 2001a; Rao *et al*, 2003). In the $FluPol_C$ structure, $P3_{Endo}$ and $PB2_{Cap}$ display the same fold as their $FluPol_A$ counterparts, which are both active in isolation. Moreover, neither of these domains is obviously occluded in the apo-heterotrimer. Hence, there appears to be no reason why these domains would not be functional, even though vRNA is not bound. Perhaps crystallisation stabilises the polymerase in a promoter-bound-like form, or the relatively low resolution of the current electron density maps hides

subtleties in active-site side chain conformations. Clearly, more work will need to be done to resolve this issue.

The $FluPol_C$ structure has also helped us visualise how polymerisation might take place in the context of the RNP, making it clear that 5′ vRNA is unlikely to remain bound during transcription and that polymerisation can occur without RNP disassembly. These assertions still need to be verified experimentally, but the structure does clarify experiments that could be done. For example, FRET experiments to address elongation can now be designed more easily, because the positions of most surface cysteines (which are potential labelling sites) is now known. The structure also suggests that regions on the tips of the fingers and thumb subdomains of PB1 and the top of the brain subdomain of $P3_C$, may interact with NP during elongation (and in the resting RNP). This could be examined by mutation. The contention that $PB2_{627}$ does not contribute directly to intrinsic polymerase activity could be tested by domain deletion.

The remaining issues I raised in the introduction are still enigmatic. We cannot yet say confidently how polyadenylation takes place, as the events during elongation termination are not known. The details of initiation from vRNA and cRNA promoters is also unclear. How does FluPol produce both mRNA and cRNA from the same template? How is cRNA recognition different from vRNA recognition? What is the mechanism of the *trans*-acting/activating polymerase postulated for vRNA production from cRNA (Jorba *et al*, 2009; York *et al*, 2013)? Why are cRNA and vRNA encapsidated, but mRNA is not? Much of the analysis in this thesis also assumes that the mechanisms of $FluPol_C$ and $FluPol_A$ are identical, and while this may largely be true (mutations of conserved active site residues in $FluPol_C$ have the expected effects, Hengrung *et al*, 2014, under review), it does not exclude possible subtle differences in, for instance, the requirements for initiation.

However, the main findings of this work are still significant. $FluPol_C$, at 255 kDa for the complete heterotrimeric complex, is one of the largest viral polymerases solved by x-ray crystallography to date. It is also the first high-resolution structure of any nsRNA virus RdRp, so provides general insights into these polymerases as they likely have a common ancestor (Kranzusch & Whelan, 2012). The crystallographic structure is far more detailed than the highest resolution EM reconstructions of the complex, enabling unambiguous localisation of functional domains within the overall polymerase particle. The mechanisms suggested by this are novel, advancing our understanding of influenza biology and doubtless precipitating many future experiments to answer unresolved questions.

5.2 Future Work

5.2.1 Improving the Model

One obvious future goal is to obtain a fully built model for $FluPol_C$, containing all of the amino acid residues placed correctly and refined with acceptable bond angles and R-values. This will be pursed by performing further iterative cycles of building and refinement, using tight NCS restraints and cross-crystal averaging. At the same time, other strategies for obtaining higher resolution data will be explored. Additional seeding and crystal optimisation trials could be set up, particularly with the orthorhombic crystals. Another avenue to investigate will be systematic in-drop crystal dehydration, such as that used to great success with the HIV-1 reverse transcriptase (Esnouf *et al*, 1998) and RNA polymerase II (Cramer *et al*, 2000 & 2001).

5.2.2 Promoter RNA Binding and RNP-Associated Structure

Several of the mechanistic ambiguities described above derive from the lack of bound promoter RNA in the $FluPol_C$ structure. Promoter RNA affects polymerase activity and may cause major conformational changes. The physiological environment of FluPol is the RNP, hence the structure of the polymerase in this context is also of great interest.

Ideally, crystallographic structures of FluPol would be determined with all relevant combinations of 5′ and 3′, vRNA and cRNA promoters. I have already unsuccessfully attempted crystal soaking and co-crystallisation experiments with the ICV vRNA promoter. However, far more remains to be tested on the co-crystallisation front. At the same time, lower resolution techniques, such as cryo-EM and small-angle x-ray scattering, can be employed to investigate any large-scale conformational differences between apo- and promoter-bound complexes. With the crystallographic structure solved, these methods

become much more informative, as the molecular structure can be fitted into the structural envelopes generated. Much of the groundwork for cryo-EM of $FluPol_C$ has already been done (section 2.2.8).

With RNP-associated FluPol, further cryo-EM work seems to be the only route forward. The use of direct electron detectors may enable more detail to be obtained than already reported. Additionally, a clever RNA-tagging approach (York *et al*, 2013), could be used to distinguish vRNPs from cRNPs, allowing any differences in polymerase conformation between these to be detected.

5.2.3 Verification of Elongation Mechanism

The model presented for elongation makes at least two claims: that protomers of NP are not disassembled from the RNP and that the 5′ vRNA strand does not remain bound to FluPol. The latter could be tested in a single-molecule FRET experiment, with one tag on an anchored polymerase, its FRET-pair on the 5′ promoter and another fluorophore elsewhere on the vRNA to monitor elongation progress. FRET experiments of this type could also provide insights into polyadenylation and the regeneration of the RNP after each elongation cycle, if suitable labelling schemes were developed. It is difficult to envision how to test whether the RNP remains intact during transcription/replication. It would be useful to gain more quantitative description of the energies involved in RNP assembly and disassembly, perhaps from a computational approach. The RNA path along the RNP also still needs experimental verification; likely via crystallography or EM.

5.2.4 Polymerases of Other Viruses

Comparisons between the polymerases of IAV, IBV and ICV would be very interesting, as this would reveal whether they all shared common mechanisms. The nanobodies produced against $FluPol_{A/Fj}$ (section 3.1.3) still need to be exhaustively tested with this protein in crystallisation trials. Also, much more effort could be put into trying to grow crystals of $FluPol_B$.

Lastly, structural work could be extended to the polymerases of other orthomyxoviruses. Further attempts to produce the RdRps of ISA, Dhori and Quaranfil viruses could be made, or *Thogotovirus* polymerases could be investigated. This would help illuminate the evolutionary relationships of the nsRNA viruses.

Materials and Methods

6.1 Materials

6.1.1 Oligonucleotides

Table 6.1. Sequencing and validation primers.
Primer names follow the convention: target-site (residue number or cloning site name)-direction.

Name	Sequence
pUCDM-MCS1-Fwd	CATACCGTCCCACCATCG
pUCDM-MCS1-Rev	TTGCATTCATTTTATGTTTCAGG
PB1-1522-Rev	TTACGGTTCAGGGTCCAGTC
PB1-1680-Fwd	CGTCGTGTGCGTGACAA
PB1-2448-Fwd	GACGACTTCGCTCTGATCGT
pFL-MCS1-Fwd	TCGATTCGCGACCTACTCC
pFL-MCS1-Rev	GGGAGGTGTGGGAGGTTTT
PB2-5350-Rev	CGACGGATCTTCACTTGGTT
PB2-5444-Fwd	TCGTATCCTGACCTCCGAGT
PB2-6265-Fwd	ACTGGGGTATCGAGCACATC
pUCDM-MCS2-Fwd	TCAGTTAGCCTCCCCCATCT
pUCDM-MCS2-Rev	GGACCTTTAATTCAACCCAACA
PA-1322-Rev	GGGAGAAGTACTGCGTGCTG
PA-1479-Fwd	TAGGGATCAGCTGGAAGTCG
PA-2156-Fwd	GGCTCGATCTTAGCGTTCAC
pFL-MCS2-Fwd	CGGGTTCCTTCCGGTATT
pFL-MCS2-Rev	CTCGACGAAGACTTGATCACC
PB1B-3928-Rev	ACGGTCCTTCACTTTCATGG
PB1B-3906-Fwd	TCCCCATGAAAGTGAAGGAC
PB1B-4788-Fwd	ACGGTTTCGTGTCCAACTTC

Table 6.1. Sequencing and validation primers.
Primer names follow the convention: target-site (residue number or cloning site name)-direction.

Name	Sequence
PB2B-5438-Rev	GAACTCAGCGGAGGTAGCAC
PB2B-5231-Fwd	TTCCAACGTGATCATGGAAA
PB2B-6210-Fwd	GACTTCTCCAGCACCGAGAC
PAB-1071-Rev	AAGGCTCTGCGTGTGATCTT
PAB-1606-Fwd	GGAGGCCTTGCAGTAGTTGA
PAB-2542-Fwd	GCCCAGCTTTTCCTTCTTCT
PB1C-3712-Rev	GGTTTCCTTGATAGCGTCCA
PB1C-4034-Fwd	ATCTGCGAGAAGCTGAAGGA
PB1C-4907-Fwd	GAGTTCCGTGCTACCTACCG
PB2C-4989-Rev	CCACGTTGTTGTGTTCCTTG
PB2C-5503-Fwd	ACACCGCTGGTATCTCCAAC
PB2C-6286-Fwd	ACCTGGAACTGTCCGACAAC
P3C-1444-Rev	GAAGTCCAAGTCCCACCTCA
P3C-1425-Fwd	TGAGGTGGGACTTGGACTTC
P3C-2216-Fwd	GACTGCATCTGGGACACCTT
DhPB1-3399-Rev	TTACGCAGGAAGGACAGGG
DhPB1-3431-Fwd	GCTCCTATCGCTCTGGAAGA
DhPB1-4025-Fwd	CTGTTCAACATCCCATTCCTCA
DhPB2-5254-Rev	AGCGTCCAGGGTAGAGAATT
DhPB2-5192-Fwd	CGTGATCCAAGAAGTCACCC
DhPB2-5853-Fwd	GTGAAGGGCATGATCGTGG
DhPB2-6868-Fwd	TGATGCAGCACAACCGTTAC
DhPA-1154-Rev	ACCATCCGTCGTGTGTACAT
DhPA-1409-Fwd	CCGTGCCATTCCTTCCTAGT
DhPA-1906-Fwd	GTCGGAGTAGTCAGCCAAGA
ISAV_PB1-3194-Rev	TCGAAAGCCTCTTCCATGGA
ISAV_PB1-3587-Fwd	GCTTTCATCTTCGTGCTGGA
ISAV_PB1-4327-Fwd	CTTCGAGAACATCACCGTGC
ISAV_PB2-5111-Rev	TCTTAGCGGAACCCATGGAA
ISAV_PB2-5291-Fwd	CAAGGAAATCGTGACCGCTG
ISAV_PB2-6075-Fwd	ATCTCTCTGCTGACTCGCG

Table 6.1. Sequencing and validation primers.
Primer names follow the convention: target-site (residue number or cloning site name)-direction.

Name	Sequence
ISAV_PA-1181-Rev	CTGAGGGTCGTGAAGTCCAA
ISAV_PA-1321-Fwd	GGTCTTGAACATCATGCCCC
ISAV_PA-1829-Fwd	TCCATCCACAGCACGTTCTC
QuPB1-3550-Rev	CCTCGGTCACTTTGTCGATC
QuPB1-3778-Fwd	GTACAACCGCCGTGTGAAC
QuPB1-4386-Fwd	TGCTGCAGTGGAAGGAAGAT
QuPB2-5111-Rev	TAGTCTCTTCGTTGGGCTCG
QuPB2-5093-Fwd	GTGGCTGGACAAGGAAATCG
QuPB2-5845-Fwd	TTTCCTGGCGCGAGTACA
QuPA-1379-Rev	CTGCTACTGAGTCCGAGCC
QuPA-1379-Fwd	GGCTCGGACTCAGTAGCAG
QuPA-1986-Fwd	GCAGTCCCTCTTCGATACCA
-70 (forward primer for MultiBac verification)	GGCTTTACACTTTATGCTTCCG

Table 6.2. Primer extension oligonucleotides

Name	Sequence	Application
47(+)	TTACTAGTCTACCCTGCTTTTGC	Detection of mRNA and cRNA (replication products of the mini-vRNA expression cassette)
47(-)	AAGCAGGGTAGACTAGTAAAGAAA	Detection of vRNA produced by the mini-vRNA expression cassette

Table 6.3 Promoter RNAs

All unlabelled RNAs were purchased from Dharmacon (now Thermo Fisher Scientific). ATT0647N-labelled RNAs were purchased from IBA (though the 5′ vRNA was a gift from Nicole Robb). All RNAs were PAGE purified by the manufacturer.

Name	Sequence
FluA 5′ vRNA	AGUAGAAACAAGGCC
FluA 3′ vRNA	GGCCUGCUUUUGCU
FluA 37mer vRNA (5′ and 3′ vRNA)	AGUAGAAACAAGGGUACUAGUCUACCC UGCUUUUGCU
FluC 5′ vRNA	AGCAGUAGCAAGGGG
FluC 3′ vRNA	CCCCUGCUUCUGCU
ATTO-FluA 5′ vRNA	AGUAGAAACAAGGAGUUU-***ATTO647N***
ATTO-FluA 3′ vRNA	***ATTO647N***-UAAACUCCUGCUUUCGCU
ATTO-FluA 5′ cRNA	AGCGAAAGCAGGAGUUU-***ATTO647N***
ATTO-FluA 3′ cRNA	***ATTO647N***-UAAACUCCUUGUUUCUACU

6.2 Molecular Cloning

6.2.1 Bacterial Culture

Except where stated, *E. coli* cells were grown in 2xYT medium at 37ºC, with shaking (190 r.p.m.). Working concentrations of the antibiotics used are listed in table 6.1 below:

Table 6.4. Antibiotics used for bacterial culture.

Antibiotic	Working Concentration (μg/ml)
Ampicillin	100
Chloramphenicol	34
Gentamicin	10
Kanamycin	35
Tetracycline	10

6.2.2 Preparation of Competent Cells

Chemically competent PirHC and WK6 (su^-) cells were produced according to the following protocol. 50 ml of 2xYT was inoculated with 0.5 ml overnight culture and grown to an OD_{600} of 0.25 – 0.3. After reaching this, the culture was cooled on ice (15 min), pelleted (3,300 g, 10 min, 4ºC) and then resuspended in 15 ml 0.1 M $CaCl_2$. After incubation (30 min, 4ºC), the cells were pelleted again (3,300 g, 10 min, 4ºC) and resuspended in 3 ml 0.1 M $CaCl_2$, 15% (v/v) glycerol. The final cell suspension was divided into 50 μl aliquots and frozen at -80ºC.

Electrocompetent DH10MultiBac cells were produced according to the protocol in Bieniossek *et al, 2008*. Briefly, 500 ml of 2xYT supplemented with tetracycline was inoculated with 2 ml overnight culture and grown to an OD_{600} of ~0.25. After this, the culture was cooled on ice (15 min), pelleted (3,500 g, 15 min, 4ºC), and washed by

centrifugation in progressively decreasing volumes of 10% (v/v) glycerol (500 ml, 250 ml, 10 ml and 1 ml). The final cell suspension was divided into 100 μl aliquots and frozen at -80°C.

6.2.3 Transformation

Plasmid DNA (generally < 50 ng) was transformed into freshly thawed, competent DH5α (sub-cloning efficiency, Life Technologies), TOP10 (Life Technologies) or PirHC *E. coli* cells by heat-shock (45 s, 42°C) followed by recovery (1 h, 37°C) if appropriate. Transformed cells were selected by growth on 2xYT agar plates, supplemented with appropriate antibiotics.

Competent DH10MultiBac cells (100 μl) were transformed by electroporation (100 ng DNA, 2.5 kV pulse). Transformed cells were supplemented with 1 ml 2xYT and allowed to recover overnight at 37°C. They were then spread onto 2xYT agar plates, supplemented with IPTG (0.4 mM) , X-Gal (40 μg/ml) and appropriate antibiotics.

6.2.4 Preparation of Plasmid DNA

Plasmid DNA was isolated using maxiprep (Roche) or miniprep (Qiagen) kits, following the manufacturer's instructions.

6.2.5 Preparation of Bacmid DNA

Bacmid DNA was isolated through alkaline lysis and alcohol precipitation. Lysis was carried out using the reagents in a miniprep kit, according to the manufacturer's instructions. Bacmid DNA in the resulting solution was precipitated by adding an equal volume of isopropanol. Pellets of bacmid DNA obtained by centrifugation (17,000 g, 20

min, 4ºC) were washed twice with 70% (v/v) ethanol (17,000 g, 10 min) and then resuspended in 10 mM Tris:HCl, pH 8.0.

6.2.6 Recombination Reactions

For cre-*loxP* recombination, 500 ng of each plasmid was mixed with 1 μl Cre buffer and 1 μl Cre recombinase (both from New England Biolabs) in a 10 μl total reaction volume. The reaction was incubated at 37ºC for 0.5 h, then the whole reaction was immediately used to transform 50 μl PirHC cells.

6.2.7 Polymerase Chain Reaction

Integration of the desired genes into the MutliBac genome was confirmed using colony PCR, with the -70 and reverse PB2 primers (table 6.1). This was also used, with appropriate primers, for recombinant construct verification. Each test colony was picked into a 10 μl reaction mix, comprising: 1x GoTaq Green reaction buffer, 0.25 μl GoTaq polymerase (both Promega), 0.2 μM dNTP mix (Fermentas) and 0.25 μM each primer. 30 cycles of the reaction were performed (melting: 95ºC, 30 s; annealing 55ºC, 30 s; extension: 72ºC, 2.5 min) with initial-denaturation (95ºC, 2 min) and final-extension (72ºC, 10 min) steps.

6.3 Baculovirus Production

6.3.1 Insect Cell Culture

All insect cell work was performed with Sf9 cells, from *Spodoptera frugiperda*. Except where indicated, these were grown in suspension culture (vented Erlenmeyer flasks at least 5x larger than volume of culture, 27°C, shaking at 110 r.p.m. in the dark) with Sf900-II or Sf900-III serum-free medium (SFM, Life Technologies), supplemented with 1x penicillin-streptomycin (Life Technologies). Cultures were passaged by dilution and maintained at cell densities of 0.5 x 10^6 – 10 x 10^6 cells/ml (manually counted using a haemocytometer).

New cultures were started from frozen stocks (kindly generated by Dr Weixian Lu) as necessary, usually after approximately 4 months. Frozen cell aliquots (2 ml) were thawed under warm running water, pelleted (1000 r.p.m, 5 min), and resuspended in fresh medium before being transferred into a T25 flask (Corning). Extra medium was added to bring the total volume to ~4 ml. The cells were allowed to recover for 7 days before being transferred into suspension culture.

6.3.2 Transfection

Sf9 cells were transfected using the Lipofectin reagent (Life Technologies), following a similar protocol to that outlined in Bieniossek *et al*, 2008. For every transfection 4 μg of bacmid DNA and 6 μl of Lipofectin were each mixed with 100 μl Sf900-III SFM (no antibiotics), combined and incubated at room temperature for 20 – 40 min. This mixture was then added to Sf9 cells (100 μl per well), which had been seeded into 6-well plates ~15 min earlier (0.5 x 10^6 cells in 3 ml antibiotic-free medium per well). The plates were sealed with parafilm and incubated at 27°C for 4 days before harvesting the resulting virus (V_0 stock). All virus stocks were filter sterilised (0.22 μm) and stored in the dark at 4°C

supplemented with 2% inactivated foetal calf serum. The titre of these was stocks not determined.

6.3.3 Virus Amplification

For V_1 stocks, 0.1 – 1.0 ml V_0 was added to 50 ml Sf9 cells at a density of 0.5 x 10^6 cells/ml. The cells were grown as above, passaging if necessary, and the virus was harvested 2 days after proliferation arrest, from the supernatant after centrifugation (800 g, 5 min). A similar protocol was followed to produce V_2 (50 or 400 ml), using sufficient infecting V_1 to induce growth arrest after a few days.

6.4 Protein Production

6.4.1 FluPol Expression

The expression and purification protocols were refined throughout the project, and I will give the optimised version below. Virus was added to Sf9 cells that had been passaged the previous day to a density of 1.0 – 1.5 x 10^6 cells/ml. The appropriate volume of infecting virus was the lowest that gave maximum expression, determined by western blotting of soluble lysates from small-scale expression cultures. This was around 2 – 6 ml of V_1 or V_2 per litre of culture. The cells were grown as above and harvested by centrifugation (800 g, 15 min, 4°C) 3 days post-infection.

6.4.2 FluPol Selenomethionine Incorporation

For the production of $FluPol_C$ derivatised with SeMet, the following protocol was used. A 1 l culture of Sf9 cells that had been passaged the previous day to a density of 1.5 x 10^6 cells/ml was infected with 6 ml V_2 stock of MultiBac $FluPol_C$-TEVprotA. 20 h post-infection, the cells were pelleted (500 g, 5 min) and resuspended in 1 l Sf900-II SFM, no methionine, no cysteine (Life Technologies), supplemented with 10% (v/v) dialysed foetal bovine serum and 150 mg/l L-cysteine. After growing for a further 4 h, 40 mg/l L-selenomethionine was added to the culture. The cells were harvested 3 days post-infection, and purification proceeded as described below.

6.4.3 FluPol Purification

Cells were lysed by sonication (15 μm oscillation, 5x(30 s on/ 30 s off), on ice) in ~30 ml lysis buffer (0.5 M NaCl, 50 mM HEPES:NaOH pH 7.5, 0.05% (w/v) octyl-β-thioglucoside (OTG), 10% (v/v) glycerol, 1x cOmplete™ protease inhibitor cocktail (Roche),

100 μg/ml RNase A) per litre of culture. Lysates were cleared by centrifugation (35,000 g, 45 min, 4ºC), and then applied onto washed (wash buffer is the same as lysis buffer, but without protease inhibitors or ribonuclease) IgG Sepharose 6 fast flow beads (GE Healthcare, 2 ml per litre of expression culture). This mixture was left shaking at 4ºC for 3 – 4 h. Following this, the beads were washed by repeatedly (5 – 6x) pelleting (4000 g, 1 min) and resuspending them in wash buffer (up to 15 ml). For the cleavage reaction, the beads were resuspended in ~8 ml wash buffer supplemented with 2.5 mM reduced glutathione, and AcTEV protease (Life Technologies, 100 units per millilitre of beads) was added. The reaction was left shaking at 4ºC overnight.

After cleavage, the beads were removed by centrifugation (4000 g, 1 min, 4ºC) followed by filtration (0.22 μm) of the supernatant. The resulting solution was concentrated using an Amicon Ultra centrifugal filter (Merck Millipore, 100 kDa MWCO, 3500 g, 4ºC) to a volume of 2 – 4 ml and flowed through an equilibrated Hi-Load Superdex 200 16/60 gel filtration column (GE Healthcare, 0.8 ml/min, 4ºC, running buffer: 0.5 M NaCl, 25 mM HEPES:NaOH pH 7.5, 10% (v/v) glycerol) connected to an ÄKTA fast protein liquid chromatography system (GE Healthcare). The peak fractions containing FluPol were combined and concentrated as above to the desired concentration. Finally, a solution of TCEP (Thermo Fisher Scientific, 0.5 M, pH 7.0) was added to a final concentration of 0.5 mM. For crystallisation, FluPol was stored at 4ºC, and used within 2 weeks.

6.4.4 Promoter-RNA Binding

Certain preparations of $FluPol_A$ were purified bound to vRNA promoter. For these, expression and purification was carried out as outlined above, up until elution of FluPol from IgG Sepharose. After this step, a high salt buffer was added to the eluate, such that the

final buffer composition was: 2 M NaCl, 10% (v/v) glycerol, 20 mM HEPES:NaOH, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$ and 0.5 mM TCEP. A 3 – 5x molar excess of each RNA to be bound was then added. The solution was transferred into SnakeSkin dialysis tubing (7K MWCO, Thermo Fisher Scientific) and dialysed at 4ºC against an identical buffer solution, which was gradually diluted over a few hours such that the final NaCl concentration was 0.5 M. The protein was left to dialyse overnight. Unbound RNA was removed by gel filtration, as described above, but using a running buffer of 0.5 M NaCl, 10% (v/v) glycerol, 25 mM HEPES:NaOH, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$ and 0.5 mM TCEP.

6.4.5 FluPol Production for Nanobody Production

For immunisation, $FluPol_{A/Fj}$ bound to vRNA promoter (separate 5′ and 3′ strands) was produced as described above. The non-crosslinked protein was concentrated to 0.6 mg/ml and supplemented with 0.5 mM m^7GpppG, 0.5 mM 2,4-dioxo-4-phenylbutanoic acid (DPBA, a known inhibitor of influenza endonuclease) and an extra 10% (v/v) glycerol. The protein was flash frozen and stored in aliquots at -80ºC until required.

Crosslinking of the protein-RNA complex was performed by adding 5 μl of a 10 mg/ml stock of the R200 crosslinker (CovalX, dissolved in DMSO) to 14 ml of 0.17 mg/ml $FluPol_{A/Fj}$–vRNA (purified as above) in 0.5 M NaCl, 10% (v/v) glycerol, 25 mM HEPES:NaOH, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$ and 0.5 mM TCEP. The reaction was left at RT for 2 h, and quenched by adding 750 μl 1 M Tris:HCl pH 7.5. Unreacted crosslinker was removed by dialysis (4ºC, overnight) against the above buffer. After dialysis, the solution was filtered (0.22 μm) and concentrated to 1.2 mg/ml. The protein was flash frozen and stored in aliquots at -80ºC until required. This reaction produced a

very low proportion of crosslinked heterotrimer, however higher concentrations of crosslinker resulted in protein precipitation.

6.4.6 FluPol Biotinylation for Nanobody Selection

Biotinylated $FluPol_{A/Fj}$ bound to vRNA promoter (separate 5′ and 3′ strands) was required for the nanobody selection assays. This was made by mixing a freshly prepared 10 mM solution of EZ-Link sulfo-NHS-LC-biotin reagent (Thermo Fisher Scientific) with FluPol purified as described in section 6.4.3 and 6.4.4, in a molar ratio of 20 biotinylation reagent : 1 protein. A total of 1 mg FluPol was biotinylated, in a reaction volume of 2.5 ml. The reaction was incubated on ice for 3.5 h, then excess biotin was removed by dialysing the solution overnight at 4ºC against 1 l 0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP. These conditions were designed to produce ~1 biotin moiety per FluPol heterotrimer. The biotinylated protein was flash frozen in the above buffer supplemented with 20% (v/v) glycerol at a concentration of 0.2 mg/ml, and stored at -80ºC until use.

6.4.7 Nanobody Expression and Purification

Nanobody expression plasmids were transformed into competent WK6 ($su^{-)}$ cells (Zell & Fritz, 1987) by heat shock. Transformed WK6 (su^-) cells were grown on LB agar plates, supplemented with 2% (w/v) glucose and 100 μg/ml amplicillin. For each expression, 1 l terrific broth supplemented with 0.1% (w/v) glucose, 2 mM $MgCl_2$ and 100 μg/ml ampicillin was inoculated with 9 ml overnight culture (grown in 20 ml LB, supplemented with 2% (w/v) glucose, 1 mM $MgCl_2$ and 100 μg/ml ampicillin). This was grown in 3x 1 l baffled flasks at 37ºC, with shaking (190 r.p.m.), until the OD_{600} was 0.7.

Expression was then induced by adding IPTG (1 mM) and incubating the culture overnight at 28ºC with shaking (190 r.p.m.).

Nanobodies were obtained from the periplasm of the expressing cells by osmotic shock. For each 1 l culture, the cells were pelleted by centrifugation (7000 r.p.m., 10 min) and then resuspended in TES buffer (15 ml for a 1 l culture at OD_{600} = 25, buffer contained 0.2 M Tris:HCl, pH 8.0, 0.5 mM EDTA, 0.5 M sucrose). This mixture was left stirring at 4ºC for 1 h to fully resuspend the cells. Osmotic shock was performed by adding 30 ml (per 15 ml of TES) of 4x diluted TES buffer, and leaving stirring at 4ºC for 45 min. After this, the mixture was centrifuged (8000 r.p.m., 30 min, 4ºC), and the supernatant added to 0.5 ml settled Ni-NTA agarose (Qiagen) that had been equilibrated with phosphate buffer 1 (50 mM Na_2HPO_4:NaH_2PO_4, 1 M NaCl, pH 7.0). This mixture was shaken for 1 h at RT, then the beads were washed with 10 ml phosphate buffer 1, followed by 30 ml phosphate buffer 2 (50 mM Na_2HPO_4:NaH_2PO_4, 1 M NaCl, pH 6.0). Nanobodies were eluted in 7.5 ml acetate buffer 3 (50 mM NaOAc, 1 M NaCl, pH 4.6). The eluate was neutralised by adding 1 M Tris:HCl, pH 7.5, and then the protein was buffer-exchanged into nanobody buffer (150 mM NaCl, 20 mM Tris:HCl, pH 7.5) using a PD-10 desalting column (GE Healthcare).

6.4.8 Protein Quantification

The concentration of protein solutions was determined by measuring their absorbance at 280 nm using a NanoDrop 2000c spectrophometer (Thermo Fisher Scientific). The Beer-Lambert equation, supplied with an estimated molar extinction coefficient obtained from Richard's Protein Calculator (http://www.mrc-lmb.cam.ac.uk/ms/methods/proteincalculator.html), was used to calculate the concentration from this absorbance measurement.

6.4.9 Electrophoresis

SDS-PAGE was used to separate proteins prior to analysis by staining or western blotting. Gels contained 8 – 15 % polyacrylamide (29 acrylamide : 1 bis-acrylamide) in tris-glycine buffer and were cast and run using Bio-Rad Mini-PROTEAN apparatus (200 V, tris-glycine buffer). Precision Plus Protein all blue size markers (Bio-Rad) were run alongside samples being analysed. Proteins were visualised by staining with SimplyBlue SafeStain (Life Technologies), following the manufacturer's instructions.

6.4.10 Western Blotting

Proteins separated using SDS-PAGE were transferred onto a nitrocellulose membrane by electrophoresis in Bio-Rad Mini-PROTEAN apparatus (100 V, 1 h, 4°C). These membranes were blocked (5% milk powder in 1x PBS with 0.5 % (v/v) Tween-20) and then probed with the desired antibodies. The antibodies used and their working concentrations are listed in table 6.5. Protein bands were visualised by chemiluminescence, using horseradish-peroxidase-conjugated secondary antibodies and Amersham ECL western blotting detection reagents (GE Healthcare).

Table 6.5. Antibodies used for western blotting.

Code	Target	Source / Notes	Species	Working Concentration
118d	PB1	Santa Cruz Biotechnology, Polyclonal	Goat	1:250
150v4	PB2	Made by Simon Carr, Polyclonal, 2nd bleed	Rabbit	1:250
161	PA	Made by Tao Deng, Polyclonal, 2nd bleed	Rabbit	1:1000
1001	Rabbit IgG	Sigma-Aldrich, Conjugated to horseradish peroxidase	Goat	1:12500
1003	Goat IgG	Sigma-Aldrich, Conjugated to horseradish peroxidase	Rabbit	1:12500

6.5 Protein Characterisation

6.5.1 Primer Extension Reaction

The primers used in these reactions were radiolabelled by mixing 1 μl 10 μM primer with 1 μl T4 Polynucleotide kinase, 1x kinase buffer (both Promega) and 1 μl EasyTide ATP [γ-^{32}P] (6000 Ci/mmol, 10 mCi/ml PerkinElmer) in total volume of 10 μl. These reactions were incubated at 37ºC for 1 h, and the radiolabelled primers were purified using the QIAquick nucleotide removal kit (Qiagen) according to the manufacturer's instructions. Radiolabelled primers were stored at -20ºC until use.

RNA was extracted from cells transfected with the indicated plasmids, using the TRIzol reagent (Life Technologies), according to the manufacturer's instructions. Extracted RNA pellets were resuspended in 30 μl RNase-free water.

In the first step of the primer extension reaction, 1 μl of each RNA sample was mixed with: 0.25 μl each radiolabelled test primer, 0.1 μl radiolabelled 5S rRNA primer and 1 μl 10 μM unlabelled 5S rRNA primer, in a total volume of 4 μl. The primers were then annealed to the sample RNA by incubating these mixtures at 95ºC for 5 min, cooling on ice, then warming to 50ºC. Primer extension was initiated by adding 5 μl pre-warmed (50ºC) reverse transcription mix (2x FS buffer, 20 mM DTT, 1 mM dNTP mix (Fermentas), 0.25 μl/reaction Superscript III reverse transcriptase) to each reaction. These were incubated at 50ºC for 1 h. The products were denatured by boiling (98ºC, 5 min) in formamide loading buffer (8 μl), before separation by PAGE on a denaturing 12% polyacrylamide (19 acrylamide : 1 bis-acrylamide) gel, and visualisation by autoradiography.

6.5.2 Fluorescence Anisotropy

For each reading, 1 nM of the indicated RNA (purchased ready-labelled with ATTO647N from IBA), was mixed with the indicated concentration of FluPol, in a buffer containing: 0.5 M NaCl, 10 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM DTT, 0.5 units/μl RNasin (Promega), 100 μg/ml BSA and 50 mM Tris:HCl, pH 8.0. The total volume of the FluPol-RNA solution was 100 μl. The polymerase used (stock concentration ~4.5 mg/ml, in 0.5 M NaCl, 50 mM Tris:HCl, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$, 0.5 mM TCEP, 20% (v/v) glycerol) had been flash-frozen and stored at -80ºC, and was freshly thawed for each experiment. Any protein precipitates present were removed by centrifugation (17,000 g, 10 min) after preparing the FluPol dilutions, but before adding RNA. FluPol-RNA solutions were incubated for at least 10 min at RT before being transferred into a quartz cuvette for anisotropy measurement.

Fluorescence anisotropy readings were taken using a scanning fluorimeter (Photon Technology International), with excitation and emission wavelengths of 644 nm and 669 – 674 nm respectively, and all slits open by 10x full turns.

6.5.3 Multi-Angle Light Scattering

These experiments were performed by passing polymerase either with or without bound promoter RNA, purified as described in section 6.4.3 and 6.4.4 (50 μl at 2.5 mg/ml) through a Superdex 200 10/300 column connected to a DAWN HELEOS detector (Wyatt Technology). The running buffer was 0.5 M NaCl, 20 mM Tris:HCl, pH 7.5, 1 mM $MgCl_2$, 0.1 mM $MnCl_2$. The data were analysed using the *ASTRA* software package (Wyatt Technology).

6.5.4 Electron Microscopy

For negative-stain EM, FluPol was adsorbed onto glow-discharged (30 s) C-flat grids, by applying 3 μl of protein solution (~0.05 mg/ml in protein storage buffer supplemented with 0.25 mM m^{7}GpppG) onto a grid, leaving for 30 – 60 s then blotting away. The grids were washed twice with water and then stained with 0.75% (w/v) uranyl acetate (30 s). The grids were viewed using a Tecnai T12 transmission electron microscope (FEI) operated at 120 kV.

For cryo-EM, FluPol$_C$ was expressed and purified as above, except the gel filtration buffer used was: 0.4 M NaCl, 20 mM Tris:HCl, pH 8.5. After gel filtration, the protein was not concentrated. Instead, a peak fraction containing ~0.2 mg/ml FluPol$_C$ was selected. Protein from this fraction was diluted 10x in: 120 mM NaSCN, 16 mM Tris:HCl, pH 8.5, 0.5 mM TCEP, 1 mM m^{7}GpppG. After dilution, 3 μl of the solution was applied onto glow-discharged (30 s) C-flat grids and vitrified using a Vitrobot (FEI). These grids were viewed at the Netherlands Centre for Electron Nanoscopy, Leiden, using a Titan Krios transmission electron microscope (FEI) equipped with a Falcon II direct electron detector (FEI) and operated at 300 kV.

6.5.5 *In Vitro* Transcription

The indicated FluPol (1 μg) was incubated (2 h, 30^{o}C) with either ApG dinucleotide (0.5 mM) or β-globin mRNA (1.7 ng/μl) in the presence of: 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 7.5 mM $MgCl_2$, 0.5 mM TCEP, 1 unit/μl RNasin, 10 μM GTP, 0.1 μl GTP [α-^{32}P] (3000 Ci/mmol, 10 mCi/ml, PerkinElmer) and 0.7 μM of each strand of the appropriate promoter RNA. The total reaction volume was 4 μl. Reaction products were denatured by boiling (98^{o}C, 5 min) in formamide loading buffer (5 μl) and separated by PAGE (1300 V, 3.5 h), on

a denaturing 20% polyacrylamide (19 acrylamide : 1 bis-acrylamide) gel. These products were visualised by autoradiography.

6.5.6 Nanobody Pull-Down Assays

Ni-NTA agarose beads (Qiagen) were coated with the indicated nanobody by applying 0.1 mg nanobody onto 10 μl beads in 500 μl nanobody buffer (section 6.4.6), and incubating at RT for 0.5 h. The beads were then washed twice with 1 ml nanobody buffer, and mixed with 0.02 mg $FluPol_{A/Fj}$ bound to FluA 37mer vRNA (table 6.3), in 0.5 ml pull-down buffer (0.5 M NaCl, 25 mM HEPES:NaOH, pH 7.5, 0.1% (v/v) Igepal CA-630, 25 mM imidazole). This complex was incubated with each nanobody for 0.5 h at RT, before the beads were washed in 2x 1 ml pull-down buffer. Pulled-down protein was then analysed by SDS-PAGE.

6.6 Crystallography

6.6.1 Crystallisation

Sitting-drop vapour-diffusion format crystallisation trials against the full range of screens available in the lab were set up in CrystalQuick 96-well plates (Greiner), using a cartesian robot (Genomic Solutions). A summary of these trials is provided in table 3.2. The appropriate protein concentration to use in these trials was determined using the PCT pre-crystallisation test (Hampton Research). Each of the drops in these screening plates was set up by mixing 100 nl protein solution with 100 nl reservoir (protein dispensed first). These were allowed to equilibrate against 95 μl reservoir solution. The plates were automatically imaged on a periodic basis.

Crystallisation conditions for specific crystal forms are described in the text of chapter 3.

6.6.2 Micro-Seeding

Seeds were prepared using the seed bead kit (Hampton). Crystals to be made into seeds were transferred into a seed bead eppendorf tube containing 50 μl stabilising solution, and vortexed (1 min). For the seeds used to make crystals of SeMet-FluPol$_C$, the stabilising solution was 70% (v/v) Morpheus G2, 5% (v/v) FluPol$_C$ solution (4.6 mg/ml). For the seeds used in other crystallisation trials, the stabilising solution contained 50% instead of 70% (v/v) Morpheus G2.

6.6.3 Heavy-Metal Derivitisation

Soaking experiments with heavy-metal clusters (listed in section 3.2.3) were performed by adding grains of the test compound directly to crystal-containing drops. The crystals were generally left soaking for 3 – 7 days before being flash-frozen. Much shorter soaking times were tested with the metatungstate cluster, by briefly passing crystals through drops containing this compound, then immediately flash-freezing.

Soaking experiments with other heavy-metal compounds, including gold (I) potassium cyanide and lead (II) acetate trihydrate, were carried out by adding 0.5 μl reservoir solution to crystal-containing drops, followed by a further 0.25 μl heavy-metal compound mixed with reservoir solution. The concentration of the heavy-metal solutions were not known; they were prepared by mixing a very small spatula's worth of compound powder with 10 μl reservoir solution. Crystals were left soaking with the test compounds for 2 – 3 h before being flash frozen. Much longer soaking times (up to 1.5 weeks) were also tested with lead (II) acetate, though the best diffraction was seen with crystals that had only been soaked for 2 h with this compound.

6.6.4 Crystal Cryo-Protection

Fully grown crystals were transferred briefly into freshly made drops of reservoir solution supplemented with 25% (v/v) glycerol or 20% (v/v) ethylene glycol, although glycerol eventually became the preferred cryoprotectant. They were then flash-frozen by plunging into liquid nitrogen.

6.6.5 Data Collection, Reduction and Analysis

X-ray diffraction datasets were collected and processed as described in section 3.2. Data analysis also used programs from the *Collaborative Computational Project Number 4* suite (Winn *et al*, 2011).

6.6.6 Figure Preparation

All structural figures presented in this thesis, unless stated otherwise, were prepared using *PyMOL* (Schrödinger).

List of References

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221

Aggarwal S, Bradel-Tretheway B, Takimoto T, Dewhurst S & Kim B (2010) Biochemical characterization of enzyme fidelity of influenza A virus RNA polymerase complex. *PLoS ONE* **5:** e10372

Aggarwal S, Dewhurst S, Takimoto T & Kim B (2011) Biochemical impact of the host adaptation-associated PB2 E627K mutation on the temperature-dependent RNA synthesis kinetics of influenza A virus polymerase complex. *J Biol Chem* **286:** 34504–34513

Albertini AAV, Wernimont AK, Muziol T, Ravelli RBG, Clapier CR, Schoehn G, Weissenhorn W & Ruigrok RWH (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313:** 360–363

Apostolov K & Flewett TH (1969) Further observations on the structure of influenza viruses A and C. *J Gen Virol* **4:** 365–370

Area E, Martín-Benito J, Gastaminza P, Torreira E, Valpuesta JM, Carrascosa JL & Ortin J (2004) 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc Natl Acad Sci USA* **101:** 308–313

Arranz R, Coloma R, Chichón FJ, Conesa JJ, Carrascosa JL, Valpuesta JM, Ortin J & Martín-Benito J (2012) The Structure of Native Influenza Virion Ribonucleoproteins. *Science* **338:** 1634–1647

Azzeh M, Flick R & Hobom G (2001) Functional analysis of the influenza A virus cRNA promoter and construction of an ambisense transcription system. *Virology* **289:** 400–410

Azziz Baumgartner E, Dao CN, Nasreen S, Bhuiyan MU, Mah-E-Muneer S, Mamun Al A, Sharker MAY, Zaman RU, Cheng P-Y, Klimov AI, Widdowson M-A, Uyeki TM, Luby SP, Mounts A & Bresee J (2012) Seasonality, timing, and climate drivers of influenza activity worldwide. *J Infect Dis* **206:** 838–846

Bae SH, Cheong HK, Lee JH, Cheong C, Kainosho M & Choi BS (2001) Structural features of an influenza virus promoter and their implications for viral RNA synthesis. *Proc Natl Acad Sci USA* **98:** 10602–10607

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037–10041

Baltimore D (1971) Expression of animal virus genomes. *Bacteriol Rev* **35:** 235–241

Baranova E, Fronzes R, Garcia-Pino A, Van Gerven N, Papapostolou D, Péhau-Arnaudet G, Pardon E, Steyaert J, Howorka S & Remaut H (2012) SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature* **487:** 119–122

Barrett T, Wolstenholme AJ & Mahy BW (1979) Transcription and replication of influenza virus RNA. *Virology* **98:** 211–225

Beaton AR & Krug RM (1981) Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. *Nucleic Acids Res* **9:** 4423–4436

Beaton AR & Krug RM (1986) Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. *Proc Natl Acad Sci USA* **83:** 6282–6286

Bieniossek C, Richmond TJ & Berger I (2008) MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr Protoc Protein Sci* **Chapter 5:** Unit 5.20

Blok V, Cianci C, Tibbles KW, Inglis SC, Krystal M & Digard P (1996) Inhibition of the influenza virus RNA-dependent RNA polymerase by antisera directed against the carboxy-terminal region of the PB2 subunit. *J Gen Virol* **77:** 1025–1033

Booy FP, Ruigrok RW & van Bruggen EF (1985) Electron microscopy of influenza virus. A comparison of negatively stained and ice-embedded particles. *J Mol Biol* **184:** 667–676

Bortz E, Westera L, Maamary J, Steel J, Albrecht RA, Manicassamy B, Chase G, Martínez-Sobrido L, Schwemmle M & García-Sastre A (2011) Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *MBio* **2:** e00151–11–e00151–11

Bouloy M, Morgan MA, Shatkin AJ & Krug RM (1979) Cap and internal nucleotides of reovirus mRNA primers are incorporated into influenza viral complementary RNA during transcription in vitro. *J Virol* **32:** 895–904

Bouloy M, Plotch SJ & Krug RM (1978) Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc Natl Acad Sci USA* **75:** 4886–4890

Bouloy M, Plotch SJ & Krug RM (1980) Both the 7-methyl and the 2'-O-methyl groups in the cap of mRNA strongly influence its ability to act as primer for influenza virus RNA transcription. *Proc Natl Acad Sci USA* **77:** 3952–3956

Braam J, Ulmanen I & Krug RM (1983) Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34:** 609–618

Bradel-Tretheway BG, Kelley Z, Chakraborty-Sett S, Takimoto T, Kim B & Dewhurst S (2008) The human H5N1 influenza A virus polymerase complex is active in vitro over a broad range of temperatures, in contrast to the WSN complex, and this property can be attributed to the PB2 subunit. *J Gen Virol* **89:** 2923–2932

Brownlee GG & Sharps JL (2002) The RNA polymerase of influenza a virus is stabilized by interaction with its viral RNA promoter. *J Virol* **76:** 7103–7113

Bruenn JA (2003) A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res* **31:** 1821–1829

Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U & Gileadi O (2008) Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr Purif* **59:** 94–102

Butcher SJ, Grimes JM, Makeyev EV, Bamford DH & Stuart DI (2001) A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410:** 235–240

Calder LJ, Wasilewski S, Berriman JA & Rosenthal PB (2010) Structural organization of a filamentous influenza A virus. *Proc Natl Acad Sci USA* **107:** 10685–10690

Castro C, Smidansky E, Maksimchuk KR, Arnold JJ, Korneeva VS, Götte M, Konigsberg W & Cameron CE (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci USA* **104:** 4267–4272

Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, Uchida A, Götte M, Konigsberg W & Cameron CE (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* **16:** 212–218

Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, Baudin F, Li X, Crépin T & Ruigrok RWH (2013) Monomeric Nucleoprotein of Influenza A Virus. *PLoS Pathog* **9:** e1003275

Cianci C, Tiley L & Krystal M (1995) Differential activation of the influenza virus polymerase via template RNA binding. *J Virol* **69:** 3995–3999

Clerx JP, Fuller F & Bishop DH (1983) Tick-borne viruses structurally similar to Orthomyxoviruses. *Virology* **127:** 205–219

Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortin J & Martín-Benito J (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog* **5:** e1000491

Compans RW, Bishop DH & Meier-Ewert H (1977) Structural components of influenza C virions. *J Virol* **21:** 658–665

Compans RW, Content J & Duesberg PH (1972) Structure of the ribonucleoprotein of influenza virus. *J Virol* **10:** 795–800

Cramer P, Bushnell DA & Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292:** 1863–1876

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR & Kornberg RD (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288:** 640–649

Crépin T, Dias A, Palencia A, Swale C, Cusack S & Ruigrok RWH (2010) Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J Virol* **84:** 9096–9104

Cronin CN, Lim KB & Rogers J (2007) Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci* **16:** 2023–2029

Crow M, Deng T, Addley M & Brownlee GG (2004) Mutational analysis of the influenza virus cRNA promoter and identification of nucleotides critical for replication. *J Virol* **78:** 6263–6270

Datta K, Wolkerstorfer A, Szolar OHJ, Cusack S & Klumpp K (2013) Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. *Nucleic Acids Res* **41:** 8289–8299

Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, Bandaranayake D, Breiman RF, Brooks WA, Buchy P, Feikin DR, Fowler KB, Gordon A, Hien NT, Horby P, Huang QS, Katz MA, Krishnan A, Lal R, Montgomery JM, et al (2012) Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect Dis* **12:** 687–695

Decroly E, Ferron F, Lescar J & Canard B (2012) Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol* **10:** 51–65

Deng T, Engelhardt OG, Thomas B, Akoulitchev AV, Brownlee GG & Fodor E (2006a) Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J Virol* **80:** 11911–11919

Deng T, Sharps J, Fodor E & Brownlee GG (2005) In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *J Virol* **79:** 8669–8674

Deng T, Vreede FT & Brownlee GG (2006b) Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J Virol* **80:** 2337–2348

Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S & Ruigrok RWH (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458:** 914–918

Digard P, Blok VC & Inglis SC (1989) Complex formation between influenza virus polymerase proteins expressed in Xenopus oocytes. *Virology* **171:** 162–169

Dochez AR, Mills KC & Kneeland Y (1933) Studies of the Etiology of Influenza. *Experimental Biology and Medicine* **30:** 1017–1022

Domanska K, Vanderhaegen S, Srinivasan V, Pardon E, Dupeux F, Marquez JA, Giorgetti S, Stoppini M, Wyns L, Bellotti V & Steyaert J (2011) Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant. *Proc Natl Acad Sci USA* **108:** 1314–1319

DuBois RM, Slavish PJ, Baughman BM, Yun M-K, Bao J, Webby RJ, Webb TR & White SW (2012) Structural and biochemical basis for development of influenza virus inhibitors targeting the PA endonuclease. *PLoS Pathog* **8:** e1002830

Ecker DJ, Sampath R, Willett P, Wyatt JR, Samant V, Massire C, Hall TA, Hari K, McNeil JA, Büchen-Osmond C & Budowle B (2005) The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC Microbiol.* **5:** 19

Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Esnouf RM, Ren J, Garman EF, Somers DO, Ross CK, Jones EY, Stammers DK & Stuart DI (1998) Continuous and discontinuous changes in the unit cell of HIV-1 reverse transcriptase crystals on dehydration. *Acta Crystallogr D Biol Crystallogr* **54:** 938–953

Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69:** 1204–1214

Fechter P, Mingay L, Sharps J, Chambers A, Fodor E & Brownlee GG (2003) Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J Biol Chem* **278:** 20381–20388

Ferrer-Orta C & Verdaguer N (2009) RNA Virus Polymerases. In *Viral Genome Replication*

Ferrer-Orta C, Arias A, Escarmís C & Verdaguer N (2006) A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol* **16:** 27–34

Ferrer-Orta C, Arias A, Perez-Luque R, Escarmís C, Domingo E & Verdaguer N (2004) Structure of foot-and-mouth disease virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J Biol Chem* **279:** 47212–47221

Ferron F, Li Z, Danek EI, Luo D, Wong Y, Coutard B, Lantez V, Charrel R, Canard B, Walz T & Lescar J (2011) The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog* **7:** e1002030

Flick R & Hobom G (1999) Interaction of influenza virus polymerase with viral RNA in the 'corkscrew' conformation. *J Gen Virol* **80:** 2565–2572

Flick R, Neumann G, Hoffmann E, Neumeier E & Hobom G (1996) Promoter elements in the influenza vRNA terminal structure. *RNA* **2:** 1046–1057

Foadi J, Aller P, Alguel Y, Cameron A, Axford D, Owen RL, Armour W, Waterman DG, Iwata S & Evans G (2013) Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **69:** 1617–1632

Fodor E & Smith M (2004) The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J Virol* **78:** 9144–9153

Fodor E, Crow M, Mingay LJ, Deng T, Sharps J, Fechter P & Brownlee GG (2002) A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J Virol* **76:** 8989–9001

Fodor E, Pritlove DC & Brownlee GG (1994) The influenza virus panhandle is involved in the initiation of transcription. *J Virol* **68:** 4092–4096

Foppa IM & Hossain MM (2008) Revised estimates of influenza-associated excess mortality, United States, 1995 through 2005. *Emerg Themes Epidemiol* **5:** 26

Fournier E, Moules V, Essere B, Paillart J-C, Sirbat J-D, Isel C, Cavalier A, Rolland J-P, Thomas D, Lina B & Marquet R (2011) A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res* **40:** 2197–2209

Fujiyoshi Y, Kume NP, Sakata K & Sato SB (1994) Fine structure of influenza A virus observed by electron cryo-microscopy. *EMBO J* **13:** 318–326

Gabriel G, Abram M, Keiner B, Wagner R, Klenk H-D & Stech J (2007) Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J Virol* **81:** 9601–9604

Geerts-Dimitriadou C, Goldbach R & Kormelink R (2011a) Preferential use of RNA leader sequences during influenza A transcription initiation in vivo. *Virology* **409:** 27–32

Geerts-Dimitriadou C, Zwart MP, Goldbach R & Kormelink R (2011b) Base-pairing promotes leader selection to prime in vitro influenza genome transcription. *Virology* **409:** 17–26

González S & Ortín J (1999) Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *J Virol* **73:** 631–637

González S, Zürcher T & Ortín J (1996) Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure. *Nucleic Acids Res* **24:** 4456–4463

Graef KM, Vreede FT, Lau Y-F, McCall AW, Carr SM, Subbarao K & Fodor E (2010) The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *J Virol* **84:** 8433–8445

Guilligay D, Kadlec J, Crépin T, Lunardi T, Bouvier D, Kochs G, Ruigrok RWH & Cusack S (2014) Comparative structural and functional analysis of orthomyxovirus polymerase cap-snatching domains. *PLoS ONE* **9:** e84973

Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crépin T, Sehr P, Lewis J, Ruigrok RWH, Ortin J, Hart DJ & Cusack S (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15:** 500–506

Guo YJ, Jin FG, Wang P, Wang M & Zhu JM (1983) Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. *J Gen Virol* **64:** 177–182

Guu TSY, Dong L, Wittung-Stafshede P & Tao YJ (2008) Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. *Virology* **379:** 135–142

Hagen M, Chung TD, Butcher JA & Krystal M (1994) Recombinant influenza virus polymerase: requirement of both 5′ and 3′ viral ends for endonuclease activity. *J Virol* **68:** 1509–1515

Hansen JL, Long AM & Schultz SC (1997) Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **5:** 1109–1122

Hara K, Schmidt FI, Crow M & Brownlee GG (2006) Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J Virol* **80:** 7789–7798

Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM & Steven AC (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc Natl Acad Sci USA* **103:** 19123–19127

Hatada E, Hasegawa M, Mukaigawa J, Shimizu K & Fukuda R (1989) Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. *J Biochem* **105:** 537–546

Hause BM, Collin EA, Liu R, Huang B, Sheng Z, Lu W, Wang D, Nelson EA & Li F (2014) Characterization of a novel influenza virus in cattle and Swine: proposal for a new genus in the orthomyxoviridae family. *MBio* **5:** e00031–14–e00031–14

Hause BM, Ducatez M, Collin EA, Ran Z, Liu R, Sheng Z, Armien A, Kaplan B, Chakravarty S, Hoppe AD, Webby RJ, Simonson RR & Li F (2013) Isolation of a novel swine influenza virus

from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS Pathog* **9:** e1003176

Hay AJ, Lomniczi B, Bellamy AR & Skehel JJ (1977) Transcription of the influenza virus genome. *Virology* **83:** 337–355

Hay AJ, Skehel JJ & McCauley J (1982) Characterization of influenza virus RNA complete transcripts. *Virology* **116:** 517–522

He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z & Liu Y (2008) Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454:** 1123–1126

Hemerka JN, Wang D, Weng Y, Lu W, Kaushik RS, Jin J, Harmon AF & Li F (2009) Detection and characterization of influenza A virus PA-PB2 interaction through a bimolecular fluorescence complementation assay. *J Virol* **83:** 3944–3955

Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, Munster VJ, Sorrell EM, Bestebroer TM, Burke DF, Smith DJ, Rimmelzwaan GF, Osterhaus ADME & Fouchier RAM (2012) Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336:** 1534–1541

Hewat EA, Cusack S, Ruigrok RW & Verwey C (1984) Low resolution structure of the influenza C glycoprotein determined by electron microscopy. *J Mol Biol* **175**: 175–193

Honda A, Endo A, Mizumoto K & Ishihama A (2001) Differential roles of viral RNA and cRNA in functional modulation of the influenza virus RNA polymerase. *J Biol Chem* **276:** 31179–31185

Honda A, Mizumoto K & Ishihama A (1999) Two separate sequences of PB2 subunit constitute the RNA cap-binding site of influenza virus RNA polymerase. *Genes Cells* **4:** 475–485

Honda A, Mizumoto K & Ishihama A (2002) Minimum molecular architectures for transcription and replication of the influenza virus. *Proc Natl Acad Sci USA* **99:** 13166–13171

Honda A, Uéda K, Nagata K & Ishihama A (1988) RNA polymerase of influenza virus: role of NP in RNA chain elongation. *J Biochem* **104:** 1021–1026

Hudjetz B & Gabriel G (2012) Human-like PB2 627K influenza virus polymerase activity is regulated by importin-α1 and -α7. *PLoS Pathog* **8:** e1002488

Hulse-Post DJ, Franks J, Boyd K, Salomon R, Hoffmann E, Yen HL, Webby RJ, Walker D, Nguyen TD & Webster RG (2007) Molecular changes in the polymerase genes (PA and PB1) associated with high pathogenicity of H5N1 influenza virus in mallard ducks. *J Virol* **81:** 8515–8524s

Hutchinson EC & Fodor E (2013) Transport of the influenza virus genome from nucleus to nucleus. *Viruses* **5:** 2424–2446

Hutchinson EC, Charles PD, Hester SS, Thomas B, Trudgian D, Martínez-Alonso M & Fodor E (2014) Conserved and host-specific features of influenza virion architecture. *Nat Commun* **5:** 4816

Hutchinson EC, Denham EM, Thomas B, Trudgian DC, Hester SS, Ridlova G, York A, Turrell L & Fodor E (2012) Mapping the Phosphoproteome of Influenza A and B Viruses by Mass Spectrometry. *PLoS Pathog* **8:** e1002993

Hutchinson EC, Orr OE, Man Liu S, Engelhardt OG & Fodor E (2011) Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J Gen Virol* **92:** 1859–1869

Hwang JS, Yamada K, Honda A, Nakade K & Ishihama A (2000) Expression of functional influenza virus RNA polymerase in the methylotrophic yeast Pichia pastoris. *J Virol* **74:** 4074–4084

Johnson NPAS & Mueller J (2002) Updating the accounts: global mortality of the 1918-1920 'Spanish' influenza pandemic. *Bull Hist Med* **76:** 105–115

Jorba N, Area E & Ortin J (2008) Oligomerization of the influenza virus polymerase complex in vivo. *J Gen Virol* **89:** 520–524

Jorba N, Coloma R & Ortin J (2009) Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog* **5:** e1000462

Jung TE & Brownlee GG (2006) A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *J. Gen. Virol.* **87:** 679–688

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125–132

Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* **336:** 1030–1033

Kerry PS, Willsher N & Fodor E (2008) A cluster of conserved basic amino acids near the C-terminus of the PB1 subunit of the influenza virus RNA polymerase is involved in the regulation of viral transcription. *Virology* **373:** 202–210

King A, Lefkowitz E & Adams MJ (2011) *Virus Taxonomy*

Knipe DM & Howley PM (2013) *Fields Virology*

Kobayashi M, Tuchiya K, Nagata K & Ishihama A (1992) Reconstitution of influenza virus RNA polymerase from three subunits expressed using recombinant baculovirus system. *Virus Res.* **22:** 235–245

Kowalinski E, Zubieta C, Wolkerstorfer A, Szolar OHJ, Ruigrok RWH & Cusack S (2012) Structural analysis of specific metal chelating inhibitor binding to the endonuclease domain of influenza pH1N1 (2009) polymerase. *PLoS Pathog* **8:** e1002831

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15:** 8125–8148

Kranzusch PJ & Whelan SPJ (2012) Architecture and regulation of negative-strand viral enzymatic machinery. *RNA Biol* **9:** 941–948

Kranzusch PJ, Schenk AD, Rahmeh AA, Radoshitzky SR, Bavari S, Walz T & Whelan SPJ (2010) Assembly of a functional Machupo virus polymerase complex. *Proc Natl Acad Sci USA* **107:** 20069–20074

Kuszewski K & Brydak L (2000) The epidemiology and history of influenza. *Biomed Pharmacother* **54:** 188–195

Kuzuhara T, Kise D, Yoshida H, Horita T, Murazaki Y, Nishimura A, Echigo N, Utsunomiya H & Tsuge H (2009) Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding

domain containing the pathogenicity-determinant lysine 627 residue. *J Biol Chem* **284:** 6855–6860

Labadie K, Santos Afonso Dos E, Rameix-Welti M-A, van der Werf S & Naffakh N (2007) Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. *Virology* **362:** 271–282

Leahy MB, Dobbyn HC & Brownlee GG (2001a) Hairpin loop structure in the 3' arm of the influenza A virus virion RNA promoter is required for endonuclease activity. *J Virol* **75:** 7042–7049

Leahy MB, Pritlove DC, Poon LL & Brownlee GG (2001b) Mutagenic analysis of the 5' arm of the influenza A virus virion RNA promoter defines the sequence requirements for endonuclease activity. *J Virol* **75:** 134–142

Lee M-TM, Klumpp K, Digard P & Tiley L (2003) Activation of influenza virus RNA polymerase by the 5′ and 3′ terminal duplex of genomic RNA. *Nucleic Acids Res* **31:** 1624–1632

Leslie AGW & Powell HR (2007) Processing diffraction data with mosflm. *Evolving Methods for Macromolecular Crystallography* **245:** 41–51

Leung BW, Chen H & Brownlee GG (2010) Correlation between polymerase activity and pathogenicity in two duck H5N1 influenza viruses suggests that the polymerase contributes to pathogenicity. *Virology* **401:** 96–106

Lewis PA & Shope RE (1931) Swine Influenza II. A Hemophilic Bacillus from the Respiratory Tract of Infected Swine. *J Exp Med* **54:** 361–371

Li ML, Ramirez BC & Krug RM (1998) RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *EMBO J* **17:** 5844–5852

Li ML, Rao P & Krug RM (2001) The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J* **20:** 2078–2086

Li X & Palese P (1994) Characterization of the polyadenylation signal of influenza virus RNA. *J Virol* **68:** 1245–1249

Liu Q, Qiao C, Marjuki H, Bawa B, Ma J, Guillossou S, Webby RJ, Richt JA & Ma W (2012) Combination of PB2 271A and SR Polymorphism at Positions 590/591 Is Critical for Viral Replication and Virulence of Swine Influenza Virus in Cultured Cells and In Vivo. *J Virol* **86:** 1233–1237

Liu Y, Qin K, Meng G, Zhang J, Zhou J, Zhao G, Luo M & Zheng X (2013) Structural and Functional Characterization of K339T Substitution Identified in the PB2 Cap-binding Pocket of Influenza A Virus. *J Biol Chem* **288:** 11013–11023

Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, et al (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380:** 2095–2128

Luo GX, Luytjes W, Enami M & Palese P (1991) The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. *J Virol* **65:** 2861–2867

Maertens B, Spriestersbach A, Groll von U, Roth U, Kubicek J, Gerrits M, Graf M, Liss M, Daubert D, Wagner R & Schäfer F (2010) Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in Escherichia coli. *Protein Sci* **19:** 1312–1326

Maier HJ, Kashiwagi T, Hara K & Brownlee GG (2008) Differential role of the influenza A virus polymerase PA subunit for vRNA and cRNA promoter binding. *Virology* **370:** 194–204

Martin K & Helenius A (1991) Transport of incoming influenza virus nucleocapsids into the nucleus. *J Virol* **65:** 232–244

Martín-Benito J, Area E, Ortega J, Llorca O, Valpuesta JM, Carrascosa JL & Ortín J (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep* **2:** 313–317

Massin P, van der Werf S & Naffakh N (2001) Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J Virol* **75:** 5398–5404

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491–497

Mänz B, Schwemmle M & Brunotte L (2013) Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. *J Virol* **87:** 7200–7209

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658–674

Mehle A & Doudna JA (2008) An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host Microbe* **4:** 111–122

Mehle A & Doudna JA (2009) Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc Natl Acad Sci USA* **106:** 21312–21316

Michielssens S, Moors SLC, Froeyen M, Herdewijn P & Ceulemans A (2011) Structural basis for the role of LYS220 as proton donor for nucleotidyl transfer in HIV-1 reverse transcriptase. *Biophys. Chem.* **157:** 1–6

Moeller A, Kirchdoerfer RN, Potter CS, Carragher B & Wilson IA (2012) Organization of the Influenza Virus Replication Machinery. *Science* **338:** 1631–1634

Mok CKP, Lee HHY, Lestra M, Nicholls JM, Chan MCW, Sia SF, Zhu H, Poon LLM, Guan Y & Peiris JSM (2014) Amino Acid Substitutions in Polymerase Basic Protein 2 Gene Contribute to the Pathogenicity of the Novel A/H7N9 Influenza Virus in Mammalian Hosts. *J Virol* **88:** 3568–3576

Molinari N-AM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E & Bridges CB (2007) The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* **25:** 5086–5096

Moncorgé O, Mura M & Barclay WS (2010) Evidence for avian and human host cell factors that affect the activity of influenza virus polymerase. *J Virol* **84:** 9978–9986

Morens DM & Taubenberger JK (2010) Historical thoughts on influenza viral ecosystems, or behold a pale horse, dead dogs, failing fowl, and sick swine. *Influenza Other Respir Viruses* **4:** 327–337

Morens DM, North M & Taubenberger JK (2010a) Eyewitness accounts of the 1510 influenza pandemic in Europe. *Lancet* **376:** 1894–1895

Morens DM, Taubenberger JK, Folkers GK & Fauci AS (2010b) Pandemic influenza's 500th anniversary. *Clin. Infect. Dis.* **51:** 1442–1444

Moretti S, Armougom F, Wallace IM, Higgins DG, Jongeneel CV & Notredame C (2007) The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic acids research* **35**: W645–8.

Mosley VM & Wyckoff R (1946) Electron micrography of the virus of influenza. *Nature* **157:** 263

Mönttinen HAM, Ravantti JJ, Stuart DI & Poranen MM (2014) Automated Structural Comparisons Clarify the Phylogeny of the Right-Hand-Shaped Polymerases. *Mol Biol Evol* **31:** 2741–2752

Mullin AE, Dalton RM, Amorim MJ, Elton D & Digard P (2004) Increased amounts of the influenza virus nucleoprotein do not promote higher levels of viral genome replication. **85:** 3689–3698

Müller R, Poch O, Delarue M, Bishop DH & Bouloy M (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J Gen Virol* **75:** 1345–1352

Nath ST & Nayak DP (1990) Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol Cell Biol* **10:** 4139–4145

Neumann G, Brownlee GG, Fodor E & Kawaoka Y (2004) Orthomyxovirus replication, transcription, and polyadenylation. *Curr Top Microbiol Immunol* **283:** 121–143

Newcomb LL, Kuo R-L, Ye Q, Jiang Y, Tao YJ & Krug RM (2009) Interaction of the influenza a virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication. *J Virol* **83:** 29–36

Ng AK-L, Chan W-H, Choi S-T, Lam MK-H, Lau K-F, Chan PK-S, Au SW-N, Fodor E & Shaw P-C (2012a) Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. *PLoS ONE* **7:** e36415

Ng AK-L, Lam MK-H, Zhang H, Liu J, Au SW-N, Chan PK-S, Wang J & Shaw P-C (2012b) Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J Virol* **86:** 6758–6767

Ng AK-L, Zhang H, Tan K, Li Z, Liu J-H, Chan PK-S, Li S-M, Chan W-Y, Au SW-N, Joachimiak A, Walz T, Wang J-H & Shaw P-C (2008a) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J* **22:** 3638–3647

Ng KKS, Arnold JJ & Cameron CE (2008b) Structure-function relationships among RNA-dependent RNA polymerases. *Curr Top Microbiol Immunol* **320:** 137–156

Nieto A, la Luna de S, Bárcena J, Portela A & Ortín J (1994) Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *J Gen Virol* **75:** 29–36

Nishimura H, Hara M, Sugawara K, Kitame F, Takiguchi K, Umetsu Y, Tonosaki A & Nakamura K (1990) Characterization of the cord-like structures emerging from the surface of influenza C virus-infected cells. *Virology* **179:** 179–188

Noble E, Mathews DH, Chen JL, Turner DH, Takimoto T & Kim B (2011) Biophysical analysis of influenza a virus RNA promoter at physiological temperatures. *J Biol Chem* **286:** 22965–22970

Noda T, Sugita Y, Aoyama K, Hirase A, Kawakami E, Miyazawa A, Sagara H & Kawaoka Y (2012) Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat Commun* **3:** 639

O'Farrell D, Trowbridge R, Rowlands D & Jäger J (2003) Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de-novo initiation. *J Mol Biol* **326:** 1025–1035

O'Neill RE, Jaskunas R, Blobel G, Palese P & Moroianu J (1995) Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *J Biol Chem* **270:** 22701–22704

O'Reilly EK & Kao CC (1998) Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* **252:** 287–303

Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, Nagata K, Tame JRH & Park S-Y (2008) The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454:** 1127–1131

Ohtsu Y, Honda Y & Toyoda T (2001) Fine mapping of the subunit binding sites of influenza virus RNA polymerase. *International Congress Series* **1219:** 463–469

Ohwada K, Kitame F & Homma M (1986) Experimental infections of dogs with type C influenza virus. *Microbiol Immunol* **30:** 451–460

Osterhaus AD, Rimmelzwaan GF, Martina BE, Bestebroer TM & Fouchier RA (2000) Influenza B virus in seals. *Science* **288:** 1051–1053

Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELXprograms. *J Appl Crystallogr* **37:** 843–844

Park Y-J, Budiarto T, Wu M, Pardon E, Steyaert J & Hol WGJ (2012) The structure of the C-terminal domain of the largest editosome interaction protein and its role in promoting RNA binding by RNA-editing ligase L2. *Nucleic Acids Res* **40:** 6966–6977

Paterson D & Fodor E (2012) Emerging Roles for the Influenza A Virus Nuclear Export Protein (NEP). *PLoS Pathog* **8:** e1003019

Paterson D, Velthuis te AJW, Vreede FT & Fodor E (2014) Host restriction of influenza virus polymerase activity by PB2 627E is diminished on short viral templates in a nucleoprotein-independent manner. *J Virol* **88:** 339–344

Perez JT, Varble A, Sachidanandam R, Zlatev I, Manoharan M, García-Sastre A & tenOever BR (2010) Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc Natl Acad Sci USA* **107:** 11525–11530

Perez JT, Zlatev I, Aggarwal S, Subramanian S, Sachidanandam R, Kim B, Manoharan M & tenOever BR (2012) A small-RNA enhancer of viral polymerase activity. *J Virol* **86:** 13475–13485

Petrek M, Kosinová P, Koca J & Otyepka M (2007) MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15:** 1357–1363

Pleschka S, Jaskunas R, Engelhardt OG, Zürcher T, Palese P & García-Sastre A (1996) A plasmid-based reverse genetics system for influenza A virus. *J Virol* **70:** 4188–4192

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

Poch O, Sauvaget I, Delarue M & Tordo N (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* **8:** 3867–3874

Pons MW, Schulze IT, Hirst GK & Hauser R (1969) Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology* **39:** 250–259

Poole E, Elton D, Medcalf L & Digard P (2004) Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology* **321:** 120–133

Poon LL, Pritlove DC, Sharps J & Brownlee GG (1998) The RNA polymerase of influenza virus, bound to the 5' end of virion RNA, acts in cis to polyadenylate mRNA. *J Virol* **72:** 8214–8219

Poon LL, Pritlove DC, Fodor E & Brownlee GG (1999) Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J Virol* **73:** 3473–3476

Potter CW (2001) A history of influenza. *J Appl Microbiol* **91:** 572–579

Presti RM, Zhao G, Beatty WL, Mihindukulasuriya KA, da Rosa APAT, Popov VL, Tesh RB, Virgin HW & Wang D (2009) Quaranfil, Johnston Atoll, and Lake Chad viruses are novel members of the family Orthomyxoviridae. *J Virol* **83:** 11599–11606

Pritlove DC, Poon LL, Devenish LJ, Leahy MB & Brownlee GG (1999) A hairpin loop at the 5' end of influenza A virus virion RNA is required for synthesis of poly(A)+ mRNA in vitro. *J Virol* **73:** 2109–2114

Raab D, Graf M, Notka F, Schödl T & Wagner R (2010) The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst Synth Biol* **4:** 215–225

Rahmeh AA, Morin B, Schenk AD, Liang B, Heinrich BS, Brusic V, Walz T & Whelan SPJ (2012) Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proc Natl Acad Sci USA* **109:** 14628–14633

Rameix-Welti M-A, Tomoiu A, Santos Afonso Dos E, van der Werf S & Naffakh N (2009) Avian Influenza A virus polymerase association with nucleoprotein, but not polymerase assembly, is impaired in human cells during the course of infection. *J Virol* **83:** 1320–1331

Regan JF, Liang Y & Parslow TG (2006) Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J Virol* **80:** 252–261

Resa-Infante P, Recuero-Checa MA, Zamarreño N, Llorca O & Ortin J (2010) Structural and functional characterization of an influenza virus RNA polymerase-genomic RNA complex. *J Virol* **84:** 10477–10487

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M & Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17:** 1030–1032

Rao P, Yuan W & Krug RM (2003) Crucial role of CA cleavage sites in the cap-snatching mechanism for initiating viral mRNA synthesis. *EMBO J* **22:** 1188–1198

Robertson JS, Schubert M & Lazzarini RA (1981) Polyadenylation sites for influenza virus mRNA. *J Virol* **38:** 157–163

Ruigrok RW, Calder LJ & Wharton SA (1989) Electron microscopy of the influenza virus submembranal structure. *Virology* **173:** 311–316

Ruigrok RWH, Crépin T & Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* **14:** 504–510

Ruigrok RWH, Crépin T, Hart DJ & Cusack S (2010) Towards an atomic resolution understanding of the influenza virus replication machinery. *Curr Opin Struct Biol* **20:** 104–113

Salgado PS, Makeyev EV, Butcher SJ, Bamford DH, Stuart DI & Grimes JM (2004) The structural basis for RNA specificity and Ca2+ inhibition of an RNA-dependent RNA polymerase. *Structure* **12:** 307–316

Schreiber A, Stengel F, Zhang Z, Enchev RI, Kong EH, Morris EP, Robinson CV, da Fonseca PCA & Barford D (2011) Structural basis for the subunit assembly of the anaphase-promoting complex. *Nature* **470:** 227–232

Shapiro GI & Krug RM (1988) Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. *J Virol* **62:** 2285–2290

Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66:** 479–485

Shi L, Summers DF, Peng Q & Galarz JM (1995) Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. *Virology* **208:** 38–47

Shope RE (1931a) Swine Influenza I. Experimental Transmission and Pathology. *J Exp Med* **54:** 349–359

Shope RE (1931b) Swine Influenza III. Filtration Experiments and Etiology. *J Exp Med* **54:** 373–385

Simonsen L (1999) The global impact of influenza on morbidity and mortality. *Vaccine* **17 Suppl 1:** S3–10

Skehel JJ & Hay AJ (1978) Nucleotide sequences at the 5' termini of influenza virus RNAs and their transcripts. *Nucleic Acids Res* **5:** 1207–1219

Smith W, Andrewes CH & Laidlaw PP (1933) A Virus Obtained from Influenza Patients. *The Lancet* **222:** 66–68

Steitz TA & Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci USA* **90:** 6498–6502

Steyaert J & Kobilka BK (2011) Nanobody stabilization of G protein-coupled receptor conformational states. *Curr Opin Struct Biol* **21:** 567–572

Stuart DI, Levine M, Muirhead H & Stammers DK (1979) Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J Mol Biol* **134:** 109–142

Subbarao EK, London W & Murphy BR (1993) A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J Virol* **67:** 1761–1764

Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JRH, Nagata K & Park S-Y (2009) Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *EMBO J* **28:** 1803–1811

Szewczyk B, Laver WG & Summers DF (1988) Purification, thioredoxin renaturation, and reconstituted activity of the three subunits of the influenza A virus RNA polymerase. *Proc Natl Acad Sci USA* **85:** 7907–7911

Tao Y, Farsetta DL, Nibert ML & Harrison SC (2002) RNA synthesis in a cage--structural studies of reovirus polymerase lambda3. *Cell* **111:** 733–745

Tarendeau F, Boudet J, Guilligay D, Mas PJ, Bougault CM, Boulo S, Baudin F, Ruigrok RWH, Daigle N, Ellenberg J, Cusack S, Simorre J-P & Hart DJ (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14:** 229–233

Tarendeau F, Crépin T, Guilligay D, Ruigrok RWH, Cusack S & Hart DJ (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4:** e1000136

Taubenberger JK & Morens DM (2010) Influenza: the once and future pandemic. *Public Health Rep* **125 Suppl 3:** 16–26

Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagné N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, Eléouët J-F & Rey FA (2009) Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326:** 1279–1283

Taylor JM, Illmensee R, Litwin S, Herring L, Broni B & Krug RM (1977) Use of specific radioactive probes to study transcription and replication of the influenza virus genome. *J Virol* **21:** 530–540

Tefsen B, Lu G, Zhu Y, Haywood J, Zhao L, Deng T, Qi J & Gao GF (2014) The N-terminal domain of PA from bat-derived influenza-like virus H17N10 has endonuclease activity. *J Virol* **88:** 1935–1941

Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ & Fukuda K (2003) Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* **289:** 179–186

Thompson WW, Weintraub E, Dhankhar P, Cheng P-Y, Brammer L, Meltzer MI, Bresee JS & Shay DK (2009) Estimates of US influenza-associated deaths made using four different methods. *Influenza Other Respir Viruses* **3:** 37–49

Tiley LS, Hagen M, Matthews JT & Krystal M (1994) Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J Virol* **68:** 5108–5116

Tomescu AI, Robb NC, Hengrung N, Fodor E & Kapanidis AN (2014) Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proc Natl Acad Sci USA* **111:** E3335–42

Torreira E, Schoehn G, Fernández Y, Jorba N, Ruigrok RWH, Cusack S, Ortin J & Llorca O (2007) Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res* **35:** 3774–3783

Tsurumura T, Qiu H, Yoshida T, Tsumori Y, Hatakeyama D, Kuzuhara T & Tsuge H (2013) Conformational Polymorphism of m7GTP in Crystal Structure of the PB2 Middle Domain from Human Influenza A Virus. *PLoS ONE* **8:** e82020

Turrell L, Lyall JW, Tiley LS, Fodor E & Vreede FT (2013) The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nat Commun* **4:** 1591

Umbach JL, Yen HL, Poon LLM & Cullen BR (2010) Influenza A virus expresses high levels of an unusual class of small viral leader RNAs in infected cells. *MBio* **1:** e00204–10–e00204–10

Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66:** 22–25

Vasin AV, Temkina OA, Egorov VV, Klotchenko SA, Plotnikova MA & Kiselev OI (2014) Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. *Virus Res.* **185:** 53–63

Velthuis te AJW, Turrell L, Vreede FT & Fodor E (2013) Uncoupling of influenza A virus transcription and replication through mutation of the unpaired adenosine in the viral RNA promoter. *J Virol* **87:** 10381–10384

Vreede FT & Brownlee GG (2007) Influenza virion-derived viral ribonucleoproteins synthesize both mRNA and cRNA in vitro. *J Virol* **81:** 2196–2204

Vreede FT, Jung TE & Brownlee GG (2004) Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *J Virol* **78:** 9568–9572

Wakai C, Iwama M, Mizumoto K & Nagata K (2011) Recognition of cap structure by influenza B virus RNA polymerase is less dependent on the methyl residue than recognition by influenza A virus polymerase. *J Virol* **85:** 7504–7512

Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM, Stewart-Jones GBE, Grimes JM, Stammers DK, Esnouf RM, Jones EY, Owens RJ, Stuart DI & Harlos K (2005) A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **61:** 651–657

Walter TS, Mancini EJ, Kadlec J, Graham SC, Assenberg R, Ren J, Sainsbury S, Owens RJ, Stuart DI, Grimes JM & Harlos K (2008) Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr F Struct Biol Cryst Commun* **64:** 14–18

Waterson AP, Hurrell JM & Jensen KE (1963) The fine structure of influenza A, B and C viruses. *Arch Gesamte Virusforsch* **12:** 487–495

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67:** 235–242

Winter G, Lobley CMC & Prince SM (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69:** 1260–1273

Wriggers W (2012) Conventions and workflows for using Situs. *Acta Crystallogr D Biol Crystallogr* **68:** 344–351

Xiao S, Klein ML, LeBard DN, Levine BG, Liang H, MacDermaid CM & Alfonso-Prieto M (2014) Magnesium-dependent RNA binding to the PA endonuclease domain of the avian influenza polymerase. *J Phys Chem B* **118:** 873–889

Yamada S, Hatta M, Staker BL, Watanabe S, Imai M, Shinya K, Sakai-Tagawa Y, Ito M, Ozawa M, Watanabe T, Sakabe S, Li C, Kim JH, Myler PJ, Phan I, Raymond A, Smith E, Stacy R, Nidom CA, Lank SM, et al (2010) Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog* **6:** e1001034

Yang W, Lee JY & Nowotny M (2006) Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. *Mol Cell* **22:** 5–13

Yang X, Smidansky ED, Maksimchuk KR, Lum D, Welch JL, Arnold JJ, Cameron CE & Boehr DD (2012) Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. *Structure* **20:** 1519–1527

Ye Q, Krug RM & Tao YJ (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444:** 1078–1082

York A, Hengrung N, Vreede FT, Huiskonen JT & Fodor E (2013) Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proc Natl Acad Sci USA* **110:** E4238–45

Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z & Liu Y (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458:** 909–913

Yumerefendi H, Tarendeau F, Mas PJ & Hart DJ (2010) ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J Struct Biol* **172:** 66–74

Zell R & Fritz HJ (1987) DNA mismatch-repair in Escherichia coli counteracting the hydrolytic deamination of 5-methyl-cytosine residues. *EMBO J* **6:** 1809–1815

Zhang H, Li X, Guo J, Li L, Chang C, Li Y, Bian C, Xu K, Chen H & Sun B (2014) The PB2 E627K mutation contributes to the high polymerase activity and enhanced replication of H7N9 influenza virus. *J Gen Virol* **95:** 779–786

Zhang S, Weng L, Geng L, Wang J, Zhou J, Deubel V, Buchy P & Toyoda T (2010) Biochemical and kinetic analysis of the influenza virus RNA polymerase purified from insect cells. *Biochem Biophys Res Commun* **391:** 570–574

Zhao C, Lou Z, Guo Y, Ma M, Chen Y, Liang S, Zhang L, Chen S, Li X, Liu Y, Bartlam M & Rao Z (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. *J Virol* **83:** 9024–9030

Zheng H, Lee HA, Palese P & García-Sastre A (1999) Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *J Virol* **73:** 5240–5243

Zheng W, Olson J, Vakharia V & Tao YJ (2013) The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. *PLoS Pathog* **9:** e1003624

Zürcher T, la Luna de S, Sanz-Ezquerro JJ, Nieto A & Ortín J (1996) Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J Gen Virol* **77:** 1745–1749

Appendix

List of Publications

Hengrung N, Omari EL, Serna-Martin I, Harlos K, Vreede FT, Stuart DI, Grimes JM & Fodor E (2015) Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* (Under review)

Tomescu AI, Robb NC, **Hengrung N**, Fodor E & Kapanidis AN (2014) Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proc Natl Acad Sci USA* **111:** E3335–42

York A, **Hengrung N**, Vreede FT, Huiskonen JT & Fodor E (2013) Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proc Natl Acad Sci USA* **110:** E4238–45