Study of the structure and function of the influenza virus RNA-dependent RNA polymerase



Itziar Serna Martín

Lincoln college
University of Oxford

Thesis submitted in fulfilment of the requirements for the degree of *Doctor of Philosophy* at the University of Oxford.

Trinity term 2017

Study of the structure and function of the influenza virus RNA-dependent RNA polymerase

Itziar Serna Martín, Lincoln College, University of Oxford
Trinity term 2017

Abstract

Influenza viruses are segmented, negative-sense RNA viruses responsible for influenza disease. Unlike other negative-sense RNA viruses, influenza virus genome replication occurs in the nucleus of infected cells. The virus encodes a heterotrimeric RNA-dependent RNA polymerase (FluPol) that is responsible for transcribing and replicating the viral genome. In order to transcribe, the FluPol generates capped RNA primers by a process termed "cap-snatching". In it, the polymerase binds host capped RNAs and cleaves them 10-13 nucleotides downstream of the cap structure. The catalytic core of FluPol is found in the PB1 (Polymerase Basic protein 1) subunit, which associates with PB2 (Polymerase Basic protein 2) and PA (Polymerase Acidic protein). PB2 and PA contain a cap binding domain and an endonuclease domain respectively, which enable cap-snatching to take place. Viral transcription and replication occur in the context of a ribonucleoprotein (RNP) complex, where an RNA segment is coated by nucleoprotein and bound by a single copy of FluPol. Regulation of transcription and replication during infection is dependent on multiple viral and host factors, including nucleoprotein, cellular DNA-dependent RNA polymerase II (Pol II) and nuclear import factors. While high-resolution structures of the FluPol heterotrimer and low-resolution structures of RNPs are available, multiple aspects of the mechanisms involved in polymerase activity and its modulation remain poorly understood.

In this DPhil project, I have employed a combination of functional and structural approaches towards exploring FluPol activity. Here I present the work carried out towards optimising the methods for FluPol structural analysis, more specifically, the use of cryo-electron microscopy to explore both the structure of the heterotrimer in complex with modulators of polymerase activity as well as RNPs. I also describe the work carried out on characterisation of the requirements for transcription and replication by FluPol. This includes the role of Pol II in transcription regulation as well as those of FluPol, nucleoprotein and host factor ANP32 in controlling viral replication. Furthermore, I also present the characterisation of a panel of 24 camelid single-chain antibodies, known as nanobodies, as tools to explore polymerase activity. Within these, nanobodies which exhibited robust inhibition of all polymerase activity were found, which opens the possibility of employing these to identify sites on the polymerase that would be suitable targets for small molecule inhibitor design.

Declaration of work

All the work described in my thesis is my own except for that detailed below:

Chapter 3: Influenza ribonucleoprotein structure and function

- Cryo-electron microscopy work was carried out in collaboration with Juha Huiskonen, who devised the methodology discussed for data processing.

Chapter 4: Regulation of transcription in the influenza C virus polymerase

- The structure of the apo form of the influenza C virus polymerase is discussed. This structure was solved by Narin Hengrung. It is described in an article published in Nature which I am also an author in. Figures depicting the structure are my own.
- The original protocol for purification of influenza C polymerase was devised by Narin Hengrung
- The structure of the Serine-5 phosphorylated CTD peptide bound form of the influenza C polymerase is discussed. This structure was solved by Narin Hengrung and Max Renner. Figures depicting the structure are my own.
- A model of the Serine-5 phosphorylated CTD peptide bound form of the influenza C polymerase in an open conformation is discussed. This model was generated by Max Renner. Figures depicting the model are my own.
- Crystals were mounted by or with the help of Kamel el Omari.

Chapter 5: Modulation of influenza A virus polymerase activity by camelid single-chain antibodies

- Some of the purified nanobody preparations were made by Narin Hengrung
- The original protocol for purification of influenza A polymerase was devised by Narin Hengrung
- Cryo-electron microscopy work was carried out with the help and guidance of Abhay Kotecha

Table of contents

Abstract i

Declaration of work ii

Table of contents iii

List of figures vii

List of tables x

Abbreviations xi

1 Introduction 1

1.1 Influenza virus 1

1.1.1 Virus structure 3

1.1.2 Phylogenetics and influenza virus subtypes 4

1.1.3 Influenza virus life cycle 6

1.2 Influenza virus ribonucleoprotein 8

1.3 Influenza virus RNA dependent RNA polymerase 11

1.3.1 *PB1* 14

1.3.2 *PA* 15

1.3.3 *PB2* 17

1.3.4 Promoter RNA 23

1.3.5 Conformational flexibility of FluPol 25

1.3.6 Nuclear import of polymerase subunits 27

1.4 Transcription and replication 29

1.4.1 Transcription initiation 30

1.4.2 Replication initiation 33

1.4.3 Elongation 36

1.4.4 Termination 37

Project aims 39

2 Materials and methods 40

2.1 Materials and reagents 40

2.1.1 DNA oligonucleotides 41

2.1.2 RNA oligonucleotides....41
2.1.3 Plasmids....41
2.1.4 DNA-dependent RNA polymerase II CTD peptides....44
2.1.5 Protein expression constructs....44
2.2 Cell culture....44
2.3 Influenza virus....45
2.3.1 Influenza viruses....45
2.3.2 Influenza virus propagation....45
2.3.3 Plaque assay....46
2.4 Molecular cloning....46
2.4.1 Bacterial culture....46
2.4.2 Competent cells....46
2.4.3 Transformation....47
2.4.4 Preparation of plasmid DNA....48
2.4.5 Polymerase chain reaction and site-directed mutagenesis....48
2.4.6 Baculovirus production....49
2.5 Protein expression and purification....50
2.5.1 Influenza C virus RdRp....50
2.5.2 Influenza A virus RdRp....52
2.5.3 Ribonucleoproteins....53
2.5.4 PB1-PA-RanBP5....54
2.5.5 ANP32....56
2.5.6 NP....57
2.5.7 Nanobodies....58
2.6 Protein analysis....59
2.6.1 SDS polyacrylamide gel electrophoresis....59
2.6.2 Western blots....59
2.7 RNA analysis....60
2.7.1 RNA isolation....60
2.7.2 Oligonucleotide radiolabeling....60
2.7.3 Capped, radiolabelled oligonucleotides....61
2.7.4 Primer extension....61
2.7.5 RNA polyacrylamide gel electrophoresis....61
2.7.6 RNP reconstitution....62

2.7.7 RNP activity assays....63
2.7.8 Polymerase activity assays....63
2.7.9 Capped, radiolabelled oligonucleotide primed assays....63
2.8 Crystallography....64
2.9 Electron microscopy....64
2.9.1 Negative stain EM....64
2.9.2 Cryo-EM sample preparation....65
2.9.3 Cryo-EM data collection....65
2.9.4 Cryo-EM data processing....65
3 Influenza ribonucleoprotein structure and function....66
3.1 Introduction....66
3.2 Purification of vRNPs and cRNPs from infected cells....69
3.3 Cryo-EM analysis of vRNPs and cRNPs....74
3.3.1 Cryo-EM sample preparation and data collection....75
3.3.2 Data processing and the "helix path-walker approach"....75
3.3.3 Conclusions and future directions....79
3.4 *In vitro* analysis of cRNA to vRNA replication....80
3.4.1 RNP *in vitro* activity assays in presence of additional purified FluPol....82
3.4.2 Nucleoprotein....85
3.4.3 ANP32....87
3.4.4 Conclusions and future directions....90
3.5 Summary....91
4 Regulation of transcription in the influenza C virus polymerase....93
4.1 Introduction....93
4.2 Conservation of cap-snatching in influenza C virus polymerase....95
4.3 Purification of $FluPol_C$....98
4.4 Structural analysis of $FluPol_C$ bound to promoter vRNA....100
4.4.1 Crystallisation trials of $FluPol_C$ bound to RNA and CTD....101
4.4.2 Cryo-EM sample preparation and data collection....106
4.4.3 Cryo-EM of $FluPol_C$ bound to promoter vRNA....109
4.4.4 Conclusions and future directions....110
4.5 The effect of Pol II CTD in *in vitro* polymerase activity....111
4.5.1 The effect of Pol II CTD in *in vitro* $FluPol_C$ activity....112
4.5.2 The effect of Pol II CTD in *in vitro* $FluPol_A$ activity....114

4.5.3 Conclusions and future directions 116
4.6 Summary and model of FluPol transcription activation by Pol II 116
5 Modulation of influenza A virus polymerase activity by camelid single-chain antibodies 121
5.1 Introduction 121
5.2 FluPol$_A$ expression and purification 123
5.3 Nanobody screening 126
5.3.1 Polymerase *in vitro* activity assays 127
5.3.2 RNP reconstitutions 129
5.3.3 Conclusions and future directions 132
5.4 Cryo-EM sample preparation and screening 133
5.4.1 Purification of FluPol$_A$-nanobody-promoter vRNA complexes 134
5.4.2 Cryo-EM sample preparation 135
5.4.3 Strategies for reducing NaCl concentrations 136
5.4.4 Strategies for correcting particle preferential orientations 139
5.5 Cryo-EM data collection 143
5.5.1 Dataset 1 - Phosphate buffer 144
5.5.2 Dataset 2 - Sodium thiocyanate 146
5.5.3 Dataset 3 - Ammonium sulfate, Tween 20 148
5.5.4 Data collection on a tilted sample 150
5.5.5 Conclusions and future work 152
5.6 Summary 155
6 Nuclear import of PB1-PA by RanBP5 157
6.1 Introduction 157
6.2 Production of baculovirus for co-expression of PB1-PA-RanBP5 159
6.3 Purification of RanBP5-PB1-PA 161
6.3.1 Purification of co-expressed RanBP5-PB1-PA 161
6.3.2 Reconstitution of PB1-PA-RanBP5 complex from separately purified components 166
6.3.3 Conclusions and future directions 168
6.4 Summary 170
Concluding remarks 172
References 174
Appendix 191

List of figures

Figure 1 – Influenza virion architecture 4

Figure 2 - The influenza virus life cycle 6

Figure 3 – Viral ribonucleoprotein cryo-EM structures 11

Figure 4 – Structure of the influenza virus polymerase 13

Figure 5 – RdRp motif organization on PB1 of $FluPol_A$ 14

Figure 6 – PA subunit 17

Figure 7 – PB2 cap binding domain 19

Figure 8 – PB2 627 and NLS domains 21

Figure 9 – Promoter RNA structure 24

Figure 10 – Multiple configurations of FluPol 26

Figure 11 – Polymerase nuclear import. 28

Figure 12 – Transcription by FluPol 32

Figure 13 – Replication initiation 34

Figure 14 – Models for elongation through an RNA template in an RNP context. 37

Figure 15 – mRNA polyadenylation by FluPol 38

Figure 16 – Comparison of the NA segment influenza A/WSN/33 virus promoter sequence in vRNA and cRNA. 67

Figure 17 – Purification of vRNPs and cRNPs. 71

Figure 18 – RNP purification optimisation 72

Figure 19 – Modified RNP purification 74

Figure 20 – Cryo-EM analysis of RNPs 76

Figure 21 – RNP filament flexibility and particle heterogeneity 78

Figure 22 – RNP *in vitro* activity assay setup 81

Figure 23 – RNP *in vitro* activity assays in presence of free polymerase 83

Figure 24 – Effect of NP addition in *in vitro* RNP activity assays 86

Figure 25 – Effect of ANP32 addition to RNP *in vitro* activity assays 89

Figure 26 – Amino acid sequence alignment of the CTD binding sites on the PA C-terminal domain identified in $FluPol_A$ 94

Figure 27 – Amino acid sequence alignment of influenza A virus, influenza B virus, influenza C virus and Thogoto virus polymerase subunit active sites 96

Figure 28 – RNP reconstitution assay to determine the conservation of cap-snatching in $FluPol_C$ 98

Figure 29 – $FluPol_C$ purification 99

Figure 30 – $FluPol_C$ crystallisation trials and optimisation 103

Figure 31 – Cryo-EM analysis of promoter vRNA-bound $FluPol_C$ 108

Figure 32 – Cryo-EM density of $FluPol_C$ bound to promoter vRNA 110

Figure 33 – Effect of CTD peptide addition in $FluPol_C$ in vitro activity assays 113

Figure 34 – Effect of CTD peptide addition in $FluPol_A$ in vitro activity assays 115

Figure 35 – Crystal structure of $FluPol_C$ bound to S_5P CTD peptide 119

Figure 36 – Modelled $FluPol_C$ bound to S_5P CTD in the "open" conformation 120

Figure 37 – Proposed mechanism for transcription activation of $FluPol_C$ 120

Figure 38 – Schematic diagram comparing the domain arrangement of human IgG with that of camelid single chain antibodies 122

Figure 39 – Purification of $FluPol_A$ 124

Figure 40 – Nanobody purification 126

Figure 41 – Effect of nanobody addition on *in vitro* polymerase activity 127

Figure 42 – RNP reconstitutions in presence of nanobodies 130

Figure 43 – Purification of the $FluPol_A$-8201-promoter vRNA complex 135

Figure 44 – Purification of $nbFP_A$ by size exclusion chromatography 138

Figure 45 – Cryo-EM dataset for $nbFP_A$ in phosphate buffer 144

Figure 46 – Cryo-EM dataset for $nbFP_A$ in buffer containing sodium thiocyanate 147

Figure 47 – Cryo-EM dataset for $nbFP_A$ in buffer containing ammonium sulphate and Tween 20 148

Figure 48 – Tilted cryo-EM dataset for $nbFP_A$ in buffer containing ammonium sulphate and Tween 20 151

Figure 49 – Purification of mOrange-tagged $FluPol_{NT60}$ in complex with promoter vRNA and nanobody 8201 154

Figure 50 – Structure comparison of RanGTP and NLS bound Pse1p 158

Figure 51 – Cloning strategy for the co-expression of RanBP5, PB1 and PA in baculovirus infected insect cells 160

Figure 52 – Western blot against FluPol and RanBP5 on the lysate (SNP) and soluble fraction (SN) of baculovirus infected insect cells 161

Figure 53 – Affinity purification on IgG beads of co-expressed PB1-PA-RanBP5 and PB1-RanBP5 162

Figure 54 – Effect of NaCl concentration on the RanBP5-PB1-PA interaction 164

Figure 55 – Screening of purification conditions to improve yields of PB1-PA-RanBP5 165

Figure 56 – Purification of RanBP5 and PB1-PA 166

Figure 57 – Reconstitution of a PB1-PA-RanBP5 complex from separately purified components .. 168

List of tables

Table 1 – Influenza virus proteins encoded in each genomic segment 6

Table 2 – Plasmids 43

Table 3 – Protein expression constructs and corresponding Genbank accession numbers 44

Table 4 – Influenza viruses 45

Table 5 – Antibiotics 46

Table 6 – Transfected plasmids in RNP reconstitutions 62

Table 7 – PB2 amino acid identity matrix (%) 97

Table 8 – PA amino acid identity matrix (%) 97

Table 9 – Details of some of the conditions of $FluPol_C$ crystallisation hits 105

Table 10 – Summary of some of the sample and grid preparation conditions tested during optimisation 143

List of non-standard abbreviations

Å	Ångström
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ANP32	Acidic leucine-rich nuclear phosphoprotein 32
APS	Ammonium persulphate
ATP	Adenosine triphosphate
CAT	Chloramphenicol acetyltransferase
CBD	Cap binding domain
CCD	Charge-coupled device
CMC	Critical micelle concentration
cRNA	Complementary ribonucleic acid
CTD	C-terminal domain
CTF	Contrast transfer function
CTP	Cytidine triphosphate
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
FCS	Foetal Calf Serum
FISH	Fluorescence *in situ* hybridisation
Fj	influenza A/duck/Fujian/1/2002 (H5N1)
FluPol	Influenza virus polymerase
FRET	Förster resonance energy transfer
GTP	Guanosine triphosphate
HA	Haemagglutinin
HEF	Haemagglutinin-esterase-fusion protein
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISAV	Infectious salmon anaemia virus
kDa	Kilo Dalton

LB	Luria broth
MDBK	Madin-Darby Bovine Kidney
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
NA	Neuraminidase
NDSB	Non-detergent sulfobetaine
NEP	Nuclear export protein
NLS	Nuclear localisation sequence
NMWL	Nominal Molecular Weight Limit
NP	Nucleoprotein
NS1	Non-structural protein 1
nt	Nucleotide
NT60	influenza A/Northern Territory/60/1968 (H3N2)
NTA	Nitrilotriacetic
NTP	Nucleotide triphosphate
OD600	Optical density at 600 nm
OTG	Octylthioglucoside
P3	Polymerase protein 3
PA	Polymerase acidic protein
PAGE	Polyacrylamide Gel Electrophoresis
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethylinemine
Pol II	DNA-dependent RNA polymerase II
PP7 CP	*Pseudomonas aeruginosa* bacteriophage coat protein
RanBP5	Ran binding protein 5
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rpm	Revolutions per minute

S_2P	Serine-2 phosphorylated
S_5P	Serine-5 phosphorylated
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC-MALS	Size exclusion chromatography multi-angle light scattering
Sf9	*Spodoptera frugiperda* 9
svRNA	Small viral RNA
TBE	Tris, Boric acid, EDTA
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
UTP	Uridine triphosphate
vRNA	Viral ribonucleic acid
YT	Yeast Tryptone
β-OG	n-octyl-β-D-glucoside

1 Introduction

1.1 Influenza virus

The influenza virus is the causative agent of influenza in humans, a respiratory tract infection of varying severity that is transmitted via respiratory droplets or fomites. The infection is often symptomatic, with a runny nose, fever, headache, myalgia, coughing etc., usually developing approximately two days after initial exposure. Most cases occur during seasonal epidemics, which take place typically during winter in temperate regions, with a broader temporal distribution in tropical regions (Azziz Baumgartner et al., 2012). Despite most cases being self-limiting and resolving within one to two weeks, individuals with pre-existing conditions, the elderly, children and pregnant women are at risk of complications. These include sinusitis, viral pneumonia, secondary bacterial pneumonia as well as worsening of conditions like asthma and congestive heart failure. Overall, seasonal influenza is still a major cause of hospitalization and death, in addition to an economic burden.

One of the major obstacles towards eradicating or controlling influenza virus is the fact that, being an RNA virus, it accumulates genomic mutations very rapidly, in a process termed “antigenic drift”. This alone is enough to render vaccines and drugs ineffective over relatively short time frames. In addition to this, the fact that it is a segmented virus enables it to undergo “antigenic shift” through segment reassortment. This is of particular concern with influenza A viruses which have the potential of causing pandemics, as it enables the formation of antigenically different viruses that then enter a naïve population. Influenza A viruses (and to a limited extent influenza B and C viruses) also exist in a multitude of animal hosts, although wild water fowl and other shore birds are thought to be the main reservoir in nature (Webster et al., 1992). These animal reservoirs can and have given rise to outbreaks of zoonotic origin (Taubenberger and Kash, 2010). These are often viruses that are poorly adapted to replication in humans and therefore cause more severe disease, but which are also ill-adapted for transmission in humans. The danger of this scenario is that these viruses of zoonotic origin could rapidly evolve to achieve effective airborne transmission between humans while retaining high pathogenicity. This could lead to pandemics like that of 1918, where up to an estimated 50 million deaths resulted from the outbreak (Johnson and Mueller, 2002). The four pandemics of the last

century, all of which were caused by influenza A viruses, are thought to have arisen from reassortment although the possibility of a future pandemic arising solely from antigenic drift cannot be discarded (Herfst et al., 2012; Morens and Taubenberger, 2011; Sorrell et al., 2011).

Vaccines and antiviral drugs exist to target influenza. Seasonal vaccines need to be updated biannually (with one vaccine aimed for northern hemisphere use and one for the southern hemisphere) and usually consist of a combination of two influenza A and one or two influenza B virus strains that are predicted in advance to be prevalent in the upcoming influenza season (Grohskopf, 2016). While this vaccine undoubtedly reduces the rates of disease and the complications that result from it, its effectiveness can be quite poor and varies from year to year, depending on the accuracy of the match between the strains employed in the vaccine (which are decided upon months prior to the influenza season) and the actual circulating strains (McNeil et al., 2016). The elderly in particular display poor protective immunity after vaccination, which has led to the development of high-dose inactivated influenza vaccines and adjuvanted inactivated influenza vaccines (DiazGranados et al., 2014; Reber et al., 2011; Tsai, 2013). These however are not widely available, which leaves one of the most susceptible demographics, who are highly likely to have co-morbidities, vulnerable to infection. A live attenuated influenza vaccine administered intra-nasally is also available, which has been proposed to have improved efficacy over the inactivated influenza vaccine in children (Belshe et al., 2007). A universal influenza vaccine that wouldn't require yearly administration, with high efficacy in all age groups is not yet available, although much effort has gone into its development (He et al., 2016). This is largely due to the rapid rates at which mutations accumulate in the surface glycoproteins which render neutralising antibodies ineffective. Antiviral drug use is limited by arising resistance, as well as the short window of time in which administration is effective. Nevertheless, they are useful in the treatment of severe cases or for prophylaxis. Those currently in clinical use target the neuraminidase (oseltamivir, zanamivir) and M2 proteins (rimantadine, amantadine). Neuraminidase inhibitors work by mimicking a transition state of the neuraminidase reaction catalysed by this protein, which therefore prevent viral release from infected cells. M2 ion channel blockers on the other hand, act upon viral entry into the cell, preventing M2 function by physically blocking the ion channel and

preventing proton translocation. Further polymerase, haemagglutinin and nucleoprotein inhibitors are currently under development (Naesens et al., 2016). Several polymerase inhibitors have been designed that target one of the three polymerase active sites: the cap binding site (Clark et al., 2014), the endonuclease active site (Nakazawa et al., 2008) or the polymerase active site itself (Furuta et al., 2013). Favipiravir, a nucleobase inhibitor, works by causing RNA chain termination and increased rates of mutagenesis by the viral polymerase, and has already been approved for use in Japan, which shows the potential of polymerase inhibitors as treatment options for influenza disease (Nakazawa et al., 2008; Sangawa et al., 2013). Currently, in the event of a new a new pandemic strain arising, production of an adequate vaccine is likely to take several months while the availability and effectiveness of antivirals may be limited. This highlights the urgent requirement for the development of both improved vaccines and novel antivirals that would increase the number of treatment options available in case of arising resistance.

1.1.1 Virus structure

Influenza virions are enveloped, ~100nm in diameter and are either spherical or slightly oval in shape. Filamentous viral particles have also been described in certain strains and clinical isolates, but their exact role remains up for discussion (Chu et al., 1949; Dadonaite et al., 2016; Morgan et al., 1956). There are three membrane-bound influenza proteins in the virions: hemagglutinin (HA), the glycoprotein responsible for interaction with sialic acid containing cellular receptors, neuraminidase (NA), which cleaves sialic acid on cell surface proteins to ensure viral release and the M2 protein, a pH-sensitive ion channel involved in viral genome release from endosomes. The matrix protein (M1) lines the inside of the virus particles and is involved in recruitment of the genome segments into budding virions. Influenza A and B viruses contain 8 genome segments (influenza C viruses contain 7). Each segment is found in the form of ribonucleoproteins (RNPs); in these, a single RNA strand is coated in nucleoprotein, with the terminal ends bound by a single copy of the heterotrimeric RNA dependent RNA polymerase (Eisfeld et al., 2014). These assemble in an ordered bundle, forming specific inter-segment interactions that are thought to ensure a single copy of each segment is packaged (Noda et al., 2006). However, recruitment of genome segments to budding virions is

actually relatively error-prone, with a large proportion of virions not containing a full set (Brooke et al., 2013). Virions also contain NEP (nuclear export protein) and NS1 (non structural protein 1), an interferon antagonist, in low abundance (Hutchinson et al., 2014). A schematic diagram of the viral components is shown in Fig. 1.

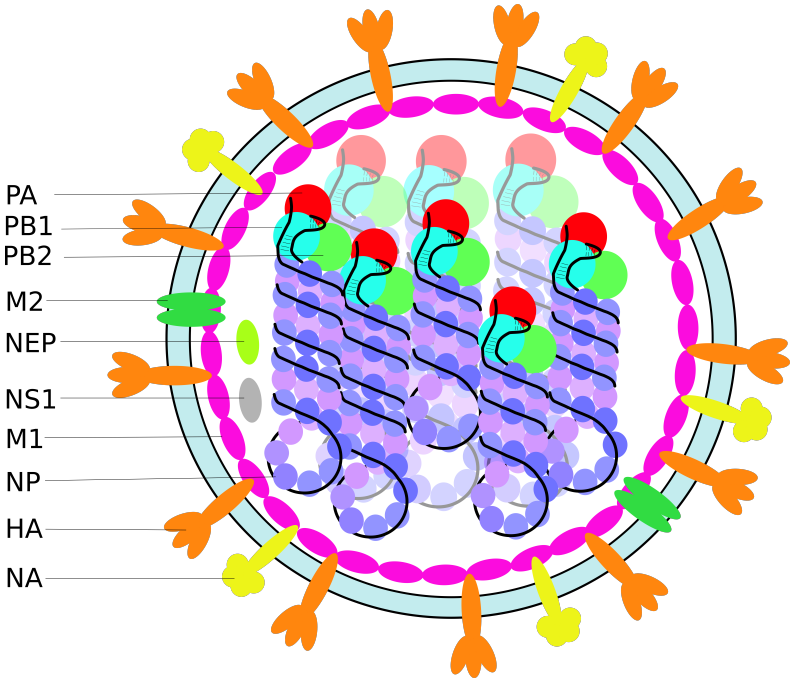


Figure 1 – Influenza virion architecture. RNP complexes assemble into a bundle, interacting with the M1 matrix protein, which lines the membrane envelope. There are three membrane associated proteins, of which HA and NA protrude out from the membrane and are involved in attachment and release from host cells. Non-structural protein NS1 and NEP are also found in the virion in low amounts (Hutchinson et al., 2014).

1.1.2 Phylogenetics and influenza virus subtypes

Influenza viruses are classified under the *Orthomyxoviridae* virus family, which encompasses a set of enveloped viruses with segmented negative sense RNA genomes (Knipe and Howley, 2007). As (-) ssRNA viruses, they belong to Group V in the Baltimore classification. A common feature of viruses belonging to this group is the fact that they encode an RNA dependent RNA polymerase to replicate their viral genome, in which the core is conserved but accessory domains vary depending on the replication/transcription strategy evolved by the virus (Ortin and Martin-Benito, 2015). The *Orthomyxoviridae* include Influenza virus A, B, C and D, as well as Isavirus, Quaranjavirus and Thogotovirus genera. The Quaranjavirus and Thogotovirus genera consist of Quaranfil, Thogoto and Dhori viruses, all of which are arthropod borne viruses while the Isavirus genera has a single member: infectious salmon anaemia virus, which causes disease in salmon (Clerx et al., 1983; Clouthier et al., 2002; Presti et al., 2009; Sandvik et al., 2000).

Influenza A and B viruses are mostly responsible for influenza disease in humans, with influenza C viruses generally causing only mild respiratory illness. Influenza A virus (IAV) has the broadest host range, being found in a wide range of birds, particularly water fowl, as well as land and sea mammals: pigs, horses, dogs, bats, seals and of course, humans. Influenza A viruses are further classified into subtypes based on hemagglutinin (HA) and neuraminidase (NA) antigenicity. Currently, H1N1 and H3N2 viruses are the main circulating seasonal strains in the human population. Influenza B viruses (IBV) exhibit a more restricted host range, having only been found in humans and seals (Osterhaus et al., 2000). Both IAV and IBV genomes contain 8 segments. It is worth noting that, despite causing very similar disease and having a similar composition, influenza A and B viruses are quite divergent in terms of amino acid identity and cannot undergo reassortment with each other during co-infection. Influenza C viruses (ICV) have only been found in humans, pigs and dogs (Ohwada et al., 1986; Yuanji et al., 1983). A related virus has also been found in cattle (and pigs), which was initially thought to belong to the ICV genus but which has since been suggested to form a new, separate genus: influenza D virus (Hause et al., 2014). Unlike, IAV and IBV, ICV contains a single, bifunctional membrane protein: HEF (Hemagglutinin-Esterase Fusion protein) rather than separate HA and NA proteins and accordingly only contains 7 genome segments. Some of the genome segments in influenza viruses encode more than one protein despite giving rise to a single transcript. These arise from alternative splice variants and non-canonical translation (Vasin et al., 2014). The full list of proteins identified to date per segment are detailed in table 1.

Segment	*IAV*	*IBV*	*ICV*
1	PB2, PB2-S1	PB2	PB2
2	PB1, PB1-N40, PB1-F2	PB1	PB1
3	PA, PA-X, PA-N155, PA-N182	PA	P3
4	HA	HA	HEF

5	NP	NP	NP
6	NA	NA, NB	M1, CM2
7	M1, M2, M42	M1, BM2	NS1, NEP
8	NS1, NEP, NS3	NS1, NEP	-

Table 1: Influenza virus proteins encoded in each genomic segment (Chen et al., 2001; Jagger et al., 2012; Vasin et al., 2014; Wise et al., 2012, 2009; Hongo et al., 1994; Betakova et al., 1996; Knipe and Howley, 2007; Yamayoshi et al., 2016)

1.1.3 Influenza virus life cycle

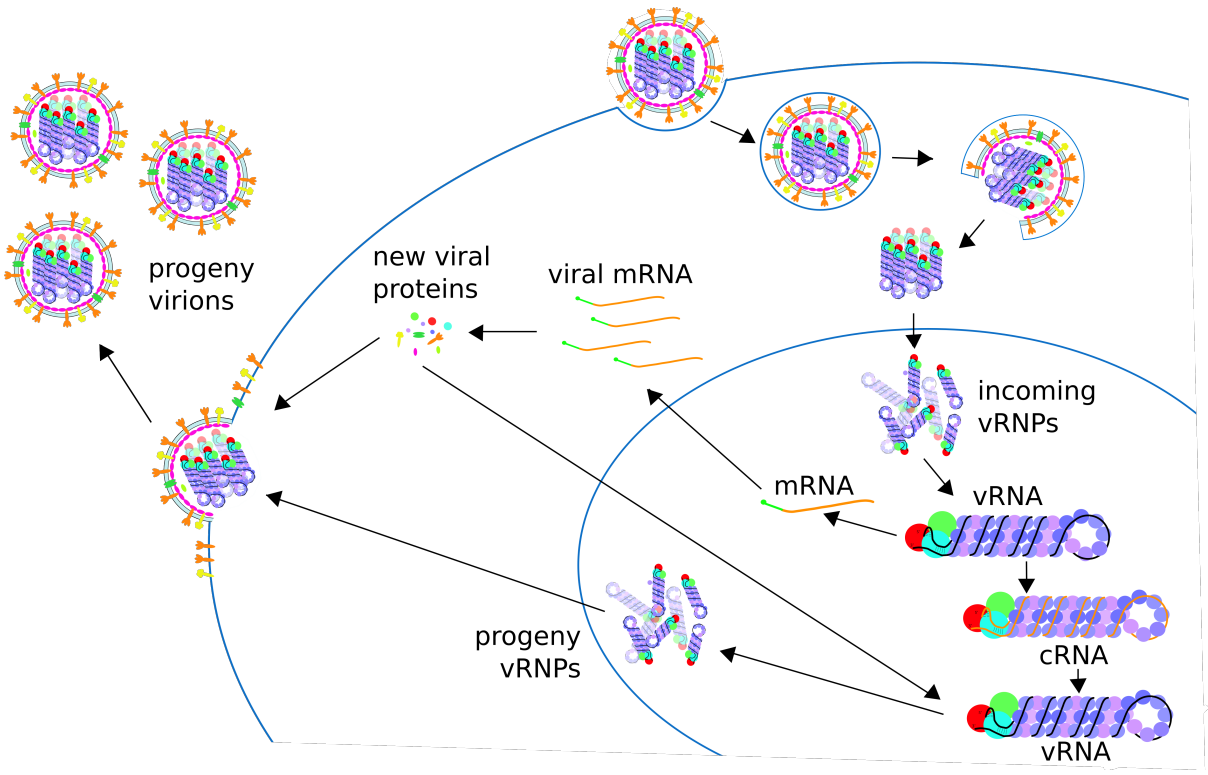


Figure 2 - The influenza virus life cycle.

Influenza virus HA binds sialic acid, commonly found on cell surface glycoproteins, to initiate infection. More specifically, HA recognises NANA (N-acetylneuraminic acid) at the end of sugar chains linked to galactose via either an α2,3 or an α2,6 linkage (influenza C HEF recognises 9-O-acetylated neuraminic acid derivatives) (Herrler et al., 1985; Stegmann, 2000; Wang and Veit, 2016). α2,3 linked sialic acid is found most commonly in the guts of birds (and in the human lower

respiratory tract), while α2,6 linked sialic acid is found in the upper respiratory tract in humans. This therefore presents a barrier for cross-species transmission which must be overcome by mutation of the HA, which has been found to occur quickly in infection models (Connor et al., 1994; Lakdawala et al., 2015). After binding of the virion to the cell surface, it is taken into the cell via receptor-mediated endocytosis (Stegmann et al., 1987). A conformational rearrangement in HA triggered by a drop in pH enables it to fuse the viral envelope with the endosomal membrane, releasing the contents of the virion into the cytoplasm. As endosomal acidification takes place, the inside of the virion also becomes acidified, due to the action of the M2 ion channel; this results in dissociation of the matrix protein and in detachment of the viral RNPs (vRNPs) (Matlin et al., 1981; Pinto and Lamb, 2006). These are then imported into the nucleus, where viral transcription and replication occur. In influenza A viruses, this is dependent on the interaction of an unconventional NLS on the nucleoprotein N-terminus with karyopherin α1/2 (Cros et al., 2005; Hutchinson and Fodor, 2012; O'Neill et al., 1995).

Inside the nucleus, the viral polymerase carries out transcription and replication of the viral genome. These processes will be discussed in detail in section 1.4 but are briefly summarised here. Production of mRNA relies on cap structures which are obtained from host pre-mRNA via a process termed cap-snatching, where the influenza polymerase (FluPol) cleaves host capped RNAs to form short primers for transcription. These transcripts are also polyadenylated and are therefore indistinguishable from cellular transcripts. Replication, which occurs later on in infection, when viral proteins have been produced by the infected cell, requires the generation of a positive sense RNA intermediate or cRNA (complementary RNA), which also forms RNPs (York et al., 2013). These positive sense copies of the viral RNA are distinguishable from mRNA as they are not capped, but have a triphosphate at the 5′ end, product of direct, unprimed, replication. They also lack a polyA tail, making them an exact complementary copy of the vRNA (Hay et al., 1977). cRNPs are then responsible for the generation of more vRNPs, which act as templates for further transcription and replication as well as being packaged into progeny virions (Pflug et al., 2017).

Nuclear export of progeny vRNPs is dependent on the matrix protein (M1) and the nuclear export protein (NEP), which in turn associates with the cellular nuclear export machinery, including Crm1 (Neumann et al., 2000; Paterson and Fodor, 2012). vRNPs are trafficked to the apical plasma membrane of polarised cells and arrange into a bundle of RNPs positioned perpendicular to the tip of the budding virion (Noda et al., 2006). HA, NA and M2 are produced by endoplasmic reticulum associated ribosomes and are transported via the Golgi apparatus to the apical cell membrane where HA and NA accumulate in lipid rafts (Barman et al., 2001). Viral budding preferentially takes place from these lipid rafts. NA has dual activity, cleaving both the sialic acid on cell surface receptors and on the virion, therefore preventing aggregation of the virus particles onto the infected cell and to neighbouring virions (Knipe and Howley, 2007). The full sequence of events in the influenza life cycle is summarised in Fig. 2.

1.2 Influenza virus ribonucleoprotein

Influenza genomic segments in the virion are assembled into ribonucleoprotein complexes. They contain all the core components required for transcription to take place and hence, enable it to occur immediately after nuclear import without the need for further protein expression. Structural studies of vRNPs have shown they are mostly composed of a flexible double helical filament of NP-coated RNA, where the helicity is dictated by NP-NP interactions (Arranz et al., 2012; Compans and Duesberg, 1972; Heggeness et al., 1982; Jennings et al., 1983; Moeller et al., 2012; Pons et al., 1969). Indeed, purified NP is capable of oligomerising and forming RNP-like structures in near absence of RNA (Ruigrok and Baudin, 1995). The NP monomers interact with each other in a head-to-tail fashion via the insertion of the tail loop into the groove in the neighbouring NP molecule, a process that is regulated by phosphorylation (Mondal et al., 2015a; Turrell et al., 2015). NP recognises the RNA in a non-specific manner, binding only to the phosphate backbone and leaving the bases exposed to the solvent. The RNA is also not completely protected by NP as shown by the fact that it can be partially digested with RNases (Baudin et al., 1994; Duesberg, 1969; Kingsbury and Webster, 1969; Pons et al., 1969). The main role of NP in the RNP appears to be to ensure processivity of the FluPol and, in its absence, only short templates can be transcribed or replicated (Turrell et al., 2013). This points

towards a role in linearising the RNA by preventing RNA-RNA interactions, which would enable the polymerase to translocate over the RNA without encountering secondary structures that will impede elongation (Baudin et al., 1994). On the other hand, artificially introduced RNA secondary structures have been shown to form in RNPs (York et al., 2013). It has also been suggested that RNA secondary structure may be responsible for dictating selective packaging of RNPs into progeny virions (Fournier et al., 2012; Kobayashi et al., 2016; Noda et al., 2012).

The exact organisation of nucleoprotein monomers in the RNP remains poorly characterised. Three single particle cryo-electron microscopy studies have aimed to address this issue but have produced conflicting results. Coloma *et al.* determined the structure of a mini-RNP isolated from transfected cells (Coloma et al., 2009). These RNPs contain a short, 248-nucleotide long, template coated by 9 NP monomers. This template is too short to allow nucleoprotein to acquire the characteristic helical arrangement, instead, they form a single large loop connected by the polymerase heterotrimer. The NP particles appear to be formed of two densities, a larger body and a smaller head, with small contacts between neighbouring NP monomers, suggesting flexible interaction points (Fig. 3).

Two studies have been carried out on full length RNPs, on either RNPs isolated from transfected cells or RNPs in virions (Arranz et al., 2012; Moeller et al., 2012). Both studies produced models for the helical portion of RNPs, which showed it consists of two antiparallel NP strands which coil forming major and minor grooves. Interestingly, Arranz *et al.* obtain a left-handed helix (which they confirm by tomography of RNPs in virions) whereas Moeller *et al.* present a right-handed helix, with correspondingly different models of NP fitting into the density (Fig. 3). Arranz *et al.* place NP monomers with the head domain facing towards the outside of the minor groove whereas the Moeller *et al.* model places the head domain towards the inter-strand interaction surface. This consequently affects the modelling of the RNA path on the RNPs. In both cases, the path of the RNA is implied from surface electrostatic charge patterns on NP as the resolution (18 Å) is not sufficient for the single stranded RNA to be visible. The Arranz *et al.* model suggests a large proportion of the RNA is accessible to the solvent. This has implications for the mechanisms of transcription and replication in

the context of an RNP. In order to access the template, the FluPol presumably has to disrupt the interaction between NP and RNA as well as NP inter- and intra-strand interactions. On the other hand, the fact that the RNA is not completely isolated by NP could indicate that full disassembly of the helical structure may not be necessary, but rather that local destabilisation of the NP oligomer would suffice to allow access to the template. The Moeller et al. model shows the RNA being intimately bound by NP, hidden between the antiparallel strands which would perhaps require a larger conformational rearrangement to enable FluPol access. Otherwise, the helical parameters obtained by both studies are comparable, with one reporting a helical rise of 28.4 Å and -57 to -64 º rotation per monomer and the other 32.6 Å helical rise and 73.9 º helical twist (positive values as it is a right handed helix) (Arranz et al., 2012; Moeller et al., 2012). This discrepancy in the structures of the RNP helical region is nevertheless not surprising, as isolated RNP complexes are highly flexible, leading to heterogeneous samples that hamper high resolution structural analysis (Gallagher et al., 2016).

One end of the RNP helical filament contains a loop where the NP oligomer folds back upon itself. The size of the loop appears to vary in different studies, ranging from 3-8 NP protomers. This is indicative of different degrees of unwinding of the helix. The flexibility and lack of symmetry in this area of the RNP has resulted in low resolution models at 27 Å or 40 Å (Arranz et al., 2012; Moeller et al., 2012).

The other end of the RNP helix is capped by the polymerase heterotrimer, which binds the partially complementary terminal ends of the RNA segment (Murti et al., 1988). These are also termed the promoter ends and span 14-15 nucleotides of the 5′ and 3′ ends. Again, due to lack of symmetry, the resolution of the polymerase was limited, revealing detail only to around 20 Å. Nevertheless, Arranz *et al.* can distinguish two polymerase populations in different conformations, both with a large groove running through, but the biological significance of this is unclear (Arranz et al., 2012). Moeller *et al.* go further to identify the polymerase subunits within their density, suggesting a flexible PA C terminal domain, which is responsible for feeding RNA into the active site, and an additional NP monomer which associates with the polymerase and is not part of the helix (Fig. 3). They also propose that the

PB2 627 domain interacts with the NP helix (Moeller et al., 2012). Coloma *et al*. also attempt to identify polymerase subunits in their density, proposing that PB1 and PB2 mediate NP interactions (Coloma et al., 2009).

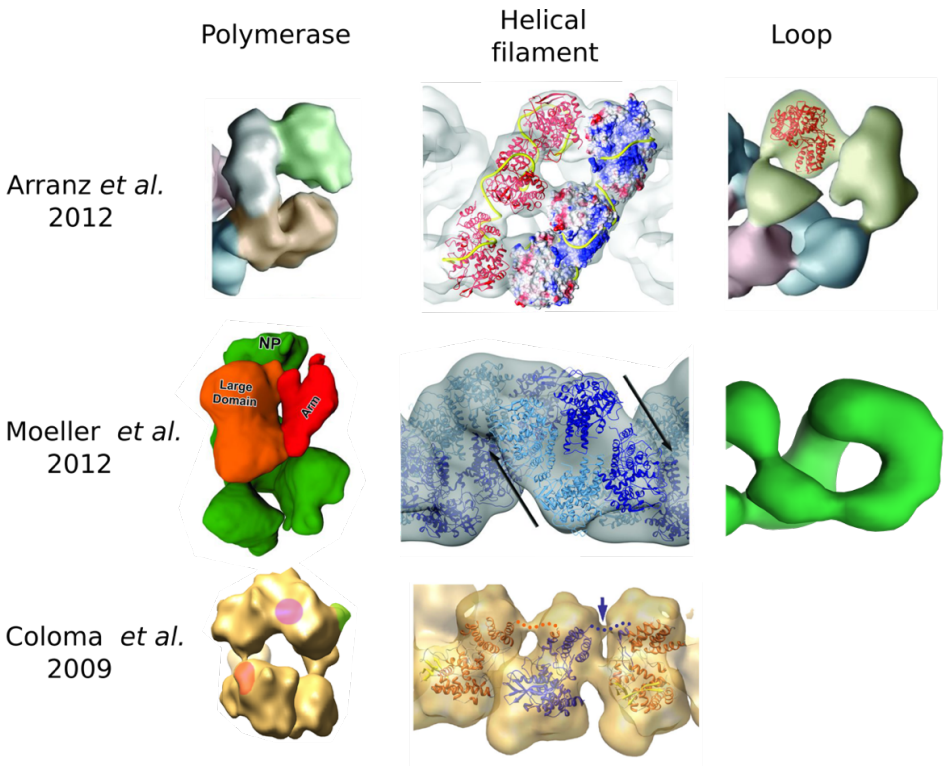


Figure 3 – Viral ribonucleoprotein cryo-EM structures comparing the polymerase, helix and loop densities and the three NP fitting models. Figure panels obtained or adapted from Arranz et al., 2012; Coloma et al., 2009 and Moeller et al., 2012.

Positive sense RNA containing RNPs or cRNPs, have only been analysed by negative stain EM. In 2D class averages from these data they are undistinguishable from vRNPs, displaying the same polymerase-helix-loop arrangement. It is likely, however, that the polymerase ends in cRNPs and vRNPs are not alike, as they have different requirements for replication (York et al., 2013). These are discussed in greater depth in section 1.4.

1.3 Influenza virus RNA dependent RNA polymerase

The influenza RNA dependent RNA polymerase (RdRp) is a 255kDa heterotrimer composed of PB1 (basic protein 1), PB2 (basic protein 2) and PA (acidic protein), known as P3 in influenza C virus. As mentioned, the polymerase is responsible for both transcription and replication of the viral negative

sense RNA. Recently, several crystal structures have become available of influenza A, B and C virus polymerases in different conformations. These have provided significant insight into the mechanisms involved in polymerase activity and yet, many aspects of transcription and replication remain obscure. Overall, the polymerase has a globular structure, where the subunits are intimately associated with each other (Fig 4a). This is particularly true for PB1 and PA. PB1, which contains the polymerase catalytic core, sits between the two globular PA domains: the endonuclease domain and the C terminal domain. These are joined by a long linker that lies along the PB1 surface. The PB2 subunit has been found to be highly dynamic, and packs on one face of PB1, while also making contacts with one or both PA globular domains depending on polymerase conformation.

FluPol has multiple channels that run through the structure to accommodate the RNA template and product, as well as for allowing entry of rNTPs to the catalytic site at the core of the protein (Fig 4b) (Gerlach et al., 2015). Based on homology with other polymerases and on the location of promoter RNA on the structure, the template entry channel has been proposed to reside underneath the PB1 β-ribbon, with the template exit channel located on the same face of the polymerase, next to the PB1 C-terminus and the PB2 N-terminal domain. This latter observation is interesting in terms of the mechanism by which the polymerase gains access to an NP bound template, again supporting the idea that only a small disruption of the local structure may be necessary as only a short stretch of template is threaded through the FluPol at a given time. The rNTP entry channel appears to be located between the PA C terminal domain and the PB1 palm domain. Lastly, the product exit channel has been suggested to run underneath the PB1 palm domain and the PB2 cap binding domain (Gerlach et al., 2015; Pflug et al., 2014; Reich et al., 2014). The FluPol, like other negative sense RNA virus polymerases, lacks helicase activity. Instead, the template and product strands are thought to be separated as they exit the polymerase. In the available structures of FluPol in an “open”, pre-initiation conformation, the PB2 lid domain partially occludes the product exit channel, which has led to suggestions that it is involved in splitting the template-product RNA duplex (Reich et al., 2014; te Velthuis and Fodor, 2016).

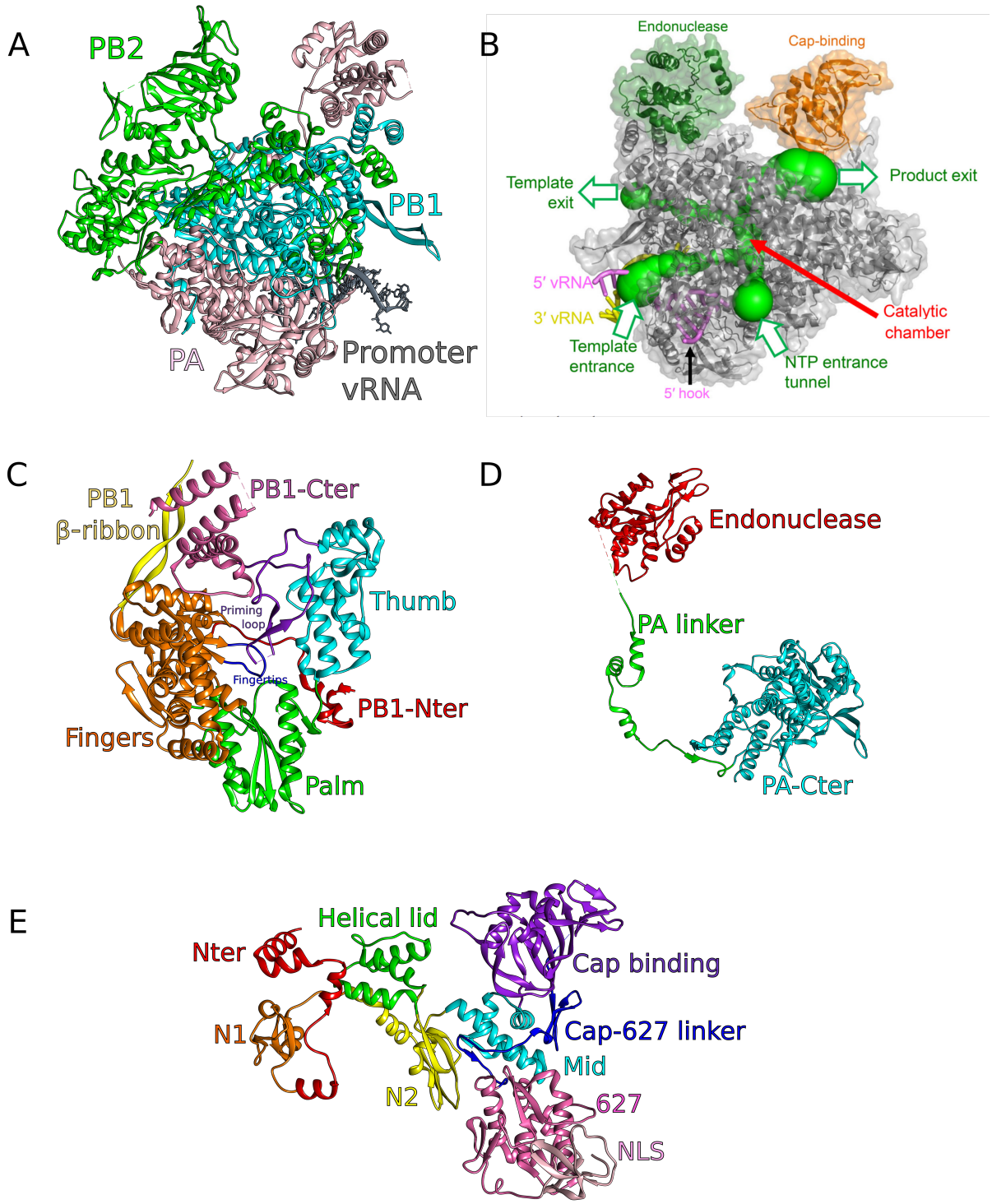


Figure 4 – Structure of the influenza virus polymerase. A) Structure of the full polymerase heterotrimer bound to promoter vRNA. PDB ID 4wsa. B) Influenza polymerase channels showing the template entry and exit channels as well as the product exit channel and the rNTP entry channel, which converge at the PB1 active site. Figure panel obtained from Pflug *et al*., 2017. C) PB1 subunit domain depiction showing the canonical right-hand conformation. PDB ID: 4wsb D) PA subunit domain depiction. PDB ID 4wsb. E) PB2 subunit domain depiction. PDB ID 4wsb.

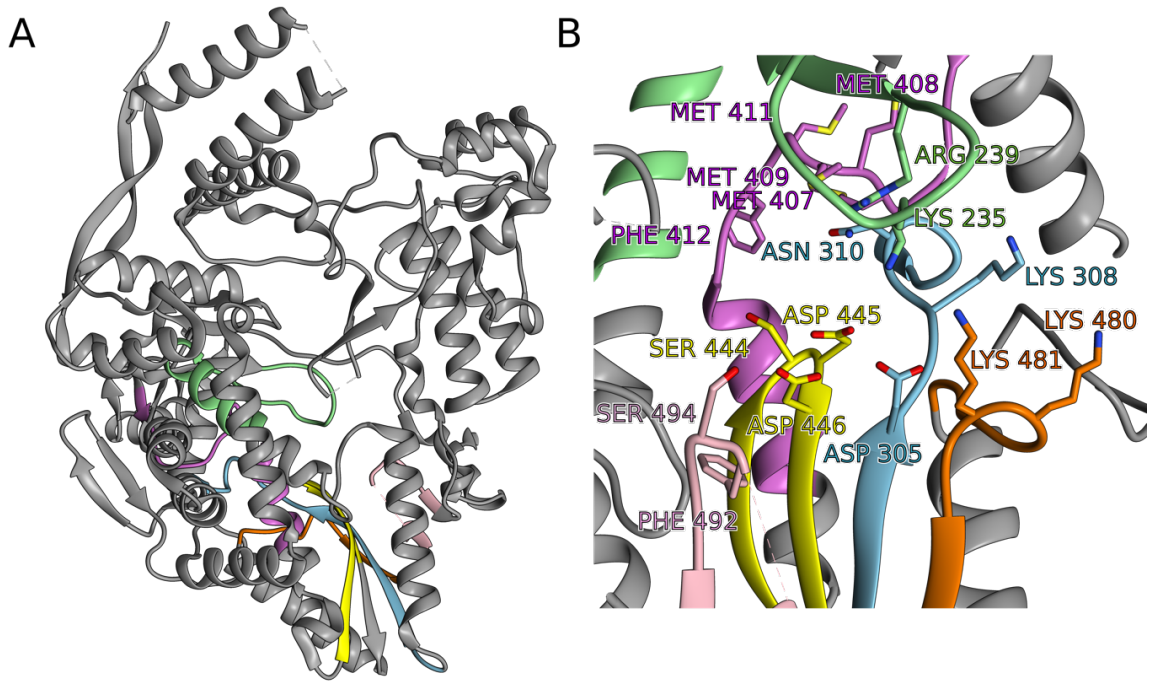


Figure 5 - RdRp motif organization on PB1 of $FluPol_A$ (PDB ID 4wsb). Motif A shown in blue, motif B shown in magenta, motif C shown in yellow, motif D shown in orange, motif E shown in pink and motif F shown in green. A) Overview of PB1 displayed to show the canonical right hand fold with the 6 RdRp motifs highlighted. B) Detailed view of the 6 motifs with key residues for catalytic activity highlighted. In motif A, D305 contributes towards coordinating two magnesium ions required for addition of NTPs to the nascent chain. Motif B shows the highly conserved methionine rich loop, important in stabilizing base pairing of the template and NTPs. Motif C D445 and D446 are also involved in coordinating the metal cations. In motif D positively charged lysines line the NTP entrance channel. Motif E is thought to be involved in stabilizing the template nucleotide at the active site. Motif F positively charged residues bind and stabilize NTPs at the active site. (Bruenn, 2003; Pflug et al., 2014).

1.3.1 PB1

PB1, being the polymerase catalytic core, acquires the canonical right hand fold observed throughout RNA and DNA polymerases with thumb, palm, fingers and fingertips domains (Fig. 4c). It contains the six characteristic motifs for RdRps, A to F, required for polymerization (Fig. 5a) (Biswas and Nayak, 1994; Bruenn, 2003; Müller et al., 1994; Poch et al., 1989). Within motif C, which spans residues 436-449, two aspartate residues D445 and D446 coordinate two magnesium ions together with D305 of motif A. These are essential for polymerase activity as these magnesium ions interact with the phosphate groups of incoming NTPs and enable their nucleophilic attack by the terminal 3´-OH group of the nascent chain (Fig. 5b). Positively charged residues in motifs D and F guide the NTPs into the active site and stabilize their position to enable the reaction to occur (Bruenn, 2003; Pflug et al., 2014). Similar to some other RdRps including HCV, Dengue virus and ϕ6 bacteriophage, FluPol PB1 contains a priming loop (Bressanelli et al., 1999; Butcher et al., 2001; Thueng-in et al.,

2012; Yap et al., 2007). This flexible loop resides at the active site of PB1 at the end of a beta hairpin and is important for terminal initiation in replication (Fig. 4c) (te Velthuis et al., 2016).

A feature that stands out in the PB1 structure is a flexibly hinged β-ribbon that spans residues 353-370 which contains the PB1 NLS and is hence rich in positively charged residues (Fig. 4c). This β-ribbon is also located close to the binding site for promoter RNA and interacts with it in the promoter vRNA bound bat influenza A structure (PDB ID 4wsb) as well as in one of the available vRNA promoter bound influenza B structures (PDB ID 4wrt) (Pflug et al., 2014; Reich et al., 2014).

1.3.2 PA

The PA subunit can be divided into three domains, the endonuclease domain, the C-terminal domain and the long flexible linker which connects both (Fig. 4d) (Guu et al., 2008; Hengrung et al., 2015; Pflug et al., 2014). The endonuclease activity of the polymerase is essential for cap snatching activity, as it is responsible for the cleavage of host capped RNAs to generate primers for transcription. Prior to the availability of structural information on the PA N terminus, the source of endonuclease activity was unknown and was attributed to each of the three polymerase subunits in different studies (Blok et al., 1996; Fodor et al., 2002; Li et al., 2001; Shi et al., 1995).

The endonuclease domain has a globular fold made up of a 5-stranded β-sheet and seven α-helices, and has structural homology with nucleases of the PD-(D/E)XK superfamily, which includes type II restriction endonucleases like *E. coli* EcoRV (Dias et al., 2009; Yuan et al., 2009). The active site contains conserved residues which are essential for activity which include H41, E80, D108 and E119 (Fig. 6a, b, c) (Crepin et al., 2010; Hara et al., 2006). These coordinate two divalent cations, which have been found to be either magnesium or manganese in different studies (Dias et al., 2009; DuBois et al., 2012; Yuan et al., 2009). Indeed, the affinity for manganese is 500 fold higher than that for magnesium, although it has been proposed that magnesium is likely to be the metal occupying the endonuclease active site in the cell due to its higher abundance (Crepin et al., 2010; Zhao et al., 2009).

There is another conserved residue in the active site, K134, essential for endonuclease activity, which could be the catalytic lysine in the PD-(D/E)XK motif, in this case PDLYDYK. Somewhat unusually, this lysine protrudes into the active site from the α_5 α-helix rather than from the β-sheet as is the case in other members of the family (Fig. 6a, b, c) (Crepin et al., 2010; Dias et al., 2009; Hara et al., 2006; Yuan et al., 2009). A structural analysis of the endonuclease domain bound to rNMPs, which should represent the structure after endonucleolytic cleavage, revealed that both E119 and K134 are involved in binding phosphate moieties, although interestingly, one of the Mn^{2+} ions observed in the apo structures (that would interact with E119) is substituted by a solvent molecule in these structures (Fig. 6c). It is suggested that the second metal binding site has a lower affinity and can be occupied by solvent or the E119 side chain (Fig. 6b) (Zhao et al., 2009).

A flexible loop formed by residues 50-73, which is not found in $FluPol_C$ endonuclease, protrudes from the globular structure and appears to have no role in catalytic activity. It was suggested that it may be involved in inter-subunit interactions in the heterotrimer but, in the available $FluPol_A$ and $FluPol_B$ structures, it protrudes into the solvent, so its significance remains unknown (Dias et al., 2009; DuBois et al., 2012; Pflug et al., 2014; Reich et al., 2014).

The PA C-terminus is a ~55kDa, compact globular domain that resembles a dragon's head. The structure can be divided into a "brain" and a "mouth", where the brain is composed of a seven-strand β-sheet surrounded by five α-helices and the mouth, made up completely of α-helices, surrounds two small PB1 N-terminal helices (Fig. 6d). It also contains a highly basic channel close to the "neck" which is essential for 5′ promoter RNA binding (He et al., 2008; Obayashi et al., 2008; Pflug et al., 2014). As well as having a large interaction surface with PB1 and with the promoter RNA, the C-terminal domain has been shown to interact with host Pol II CTD, an interaction which is essential for cap-snatching during infection (Lukarska et al., 2016).

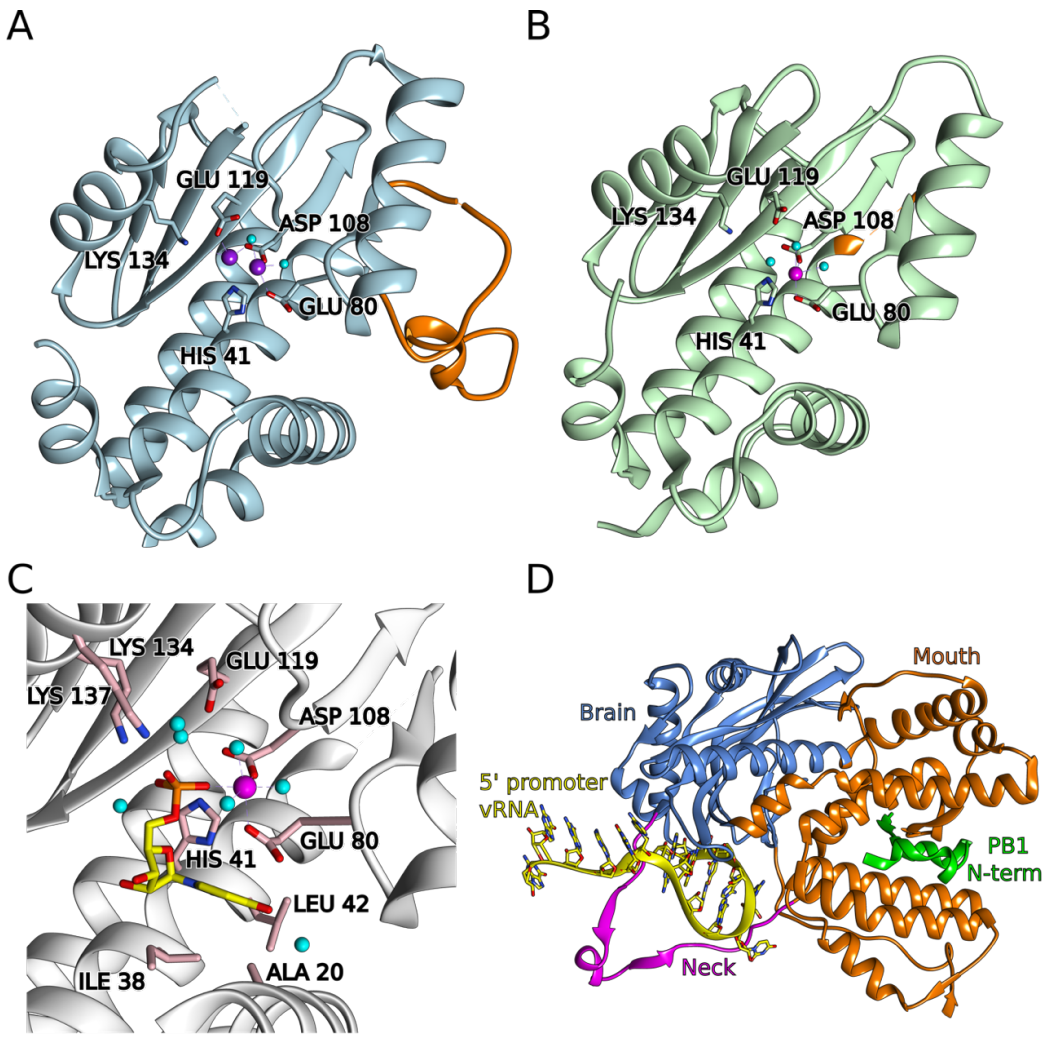


Figure 6 – PA subunit. A) Endonuclease domain bound to two manganese ions with the active site residues highlighted. The loop formed by residues 50-73 is shown in orange. PDB ID 2w69. B) Endonuclease domain bound to a single magnesium ion (magenta) with the active site residues highlighted. Solvent molecules are shown in cyan. The loop formed by residues 50-73 is shown in orange. PDB ID 3ebj. C) Close-up view of the endonuclease active site bound to UMP (yellow) with active site residues highlighted in light pink. The magnesium ion is shown in magenta and solvent molecules are shown in cyan. PDB ID 3hw3. D) C-terminal domain of PA shown with the 5´ promoter vRNA and the PB1 N-terminus. PDB ID 4wsb.

1.3.3 PB2

The PB2 subunit is divided into several subdomains. The N-terminal domains interact closely with PB1 on the opposite side of the PA linker, while the C terminal domains protrude from the core structure and appear to be extremely flexible. The N-terminal domain of PB2 is tightly intertwined with the PB1 C terminal and thumb domains and additionally interacts with the proximal end of the 3´ promoter RNA strand (Fig. 4a, e). This is followed by the N1 and N2 domains as well as the "helical lid" domain (Pflug et al., 2014). This latter one consists of a bundle of 4 α-helices which lie partially occluding the template exit channel. It has been hypothesized that during transcription and replication,

this domain may need to undergo a conformational shift to open the channel like that seen in HCV, where the polymerase thumb domain opens to accommodate the product-template duplex. In FluPol this motion in the thumb domain could drive a rotation of the lid domain (Mosley et al., 2012; Reich et al., 2014).

The C terminus of PB2 can be divided into the mid, cap-binding domain (CBD), CBD-627 linker, 627 and NLS domains (Fig. 4e). These acquire an arc-like shape both in the open and closed polymerase conformations, although the domain arrangement varies greatly between the two. The mid and CBD-627 linker domains form a rigid structure that lies at the interface between the CBD and the 627-NLS domains. The CBD is responsible for recognizing and binding the cap structure in host capped RNA, therefore enabling cap-snatching to occur. The structures of the CBD both in presence and absence of m^7GTP were solved in isolation before full heterotrimeric structures were available. They show the CBD acquires a globular fold comprised of a 5-strand antiparallel β-sheet packed against a small bundle of 4 α-helices, as well as an additional β-ribbon and a C-terminal moiety formed of short β strands (Fig. 7a) (Guilligay et al., 2008; Liu et al., 2013).

The mode of recognition of the cap structure is similar to that of other cap binding proteins, including eIF4E and VP39 of vaccinia virus (Fechter, 2005). The cap is sandwiched by aromatic residues; in the case of influenza A virus, these are F404 and H357. Further aromatic residues F323, F325, F330 and F363 form a hydrophobic patch which also contributes to the interaction with the base moiety, while basic residues R355, K339 as well as H432 interact with the phosphate groups (Fig. 7b, c) (Guilligay et al., 2008; Liu et al., 2013). Interestingly, a further m^7GTP bound structure available, where the substrate was soaked into the crystals rather than co-crystallized with the protein, shows the phosphate groups protruding into the solvent and not making contacts with the previously discussed residues, although this has not been observed again in any of the later structures that have emerged (Tsurumura et al., 2013).

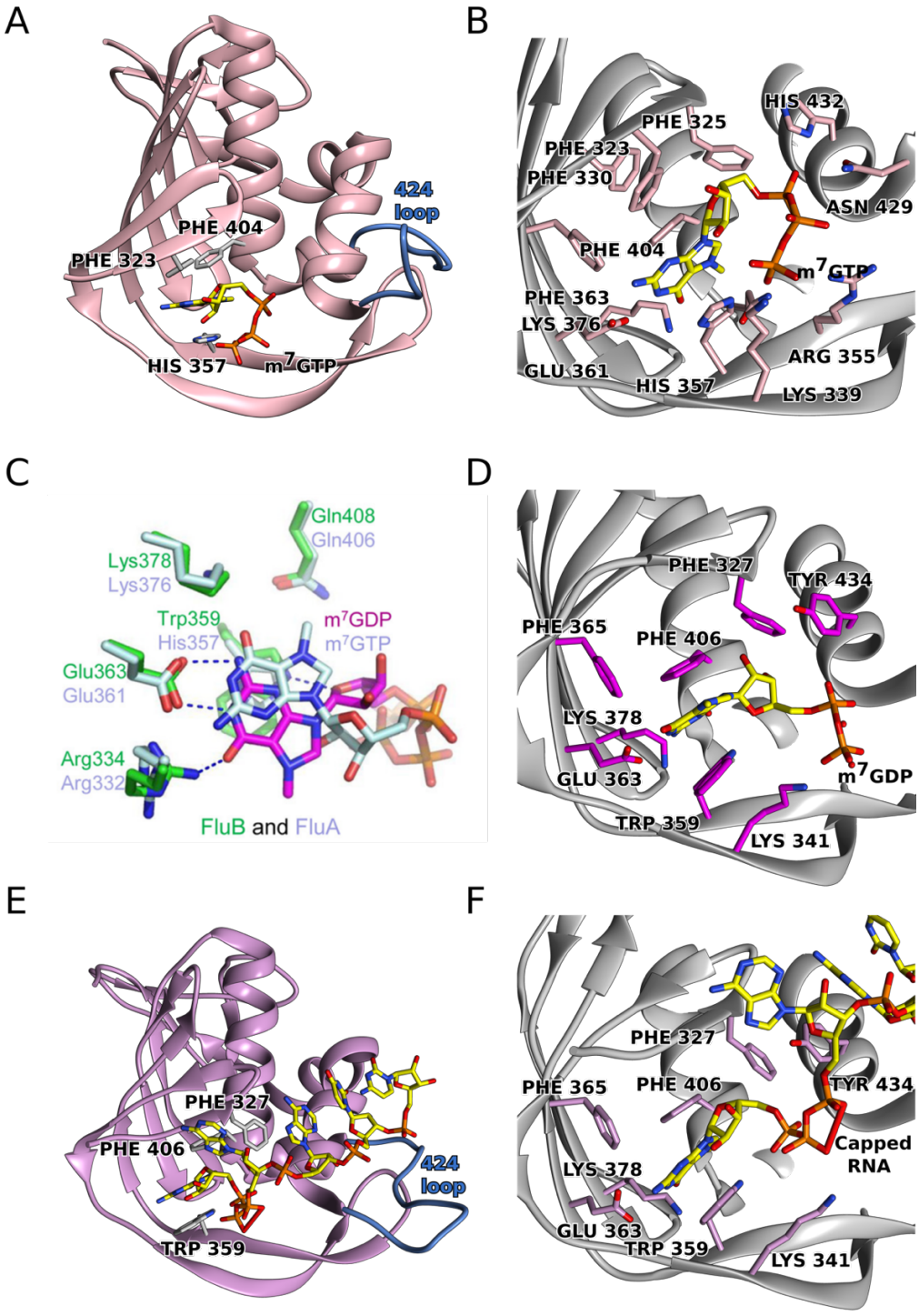


Figure 7 – PB2 cap binding domain. A) Influenza A virus CBD bound to m^7GTP (yellow), with residues involved in cap binding highlighted in grey. The 424 loop is shown in cornflower blue. B) Close-up view of the cap binding site on A), with key residues highlighted in light pink. PDB ID 2vqz. C) Comparison of the positioning of the cap structure in influenza A and B CBD structures showing a 180º flip of the molecule. Figure panel obtained from Liu *et al*., 2015. D) Close-up view of influenza B virus CBD bound to m^7GDP with key residues highlighted in magenta. PDB ID 4or4. E) Influenza A virus CBD bound to capped mRNA (yellow). Key residues are highlighted in grey. The 424 loop is shown in cornflower blue. F) Close-up view of the cap binding site on E), with key residues highlighted in plum. PDB ID 5msg.

The presence of a histidine at position 357 is somewhat unusual; indeed influenza B and C viruses contain a tryptophan at this position. $FluPol_A$ CBD has a higher specificity for m^7GTP whereas $FluPol_B$ recognizes both methylated and unmethylated caps with equal affinity (Wakai et al., 2011). Crystallographic studies later showed that in $FluPol_B$ CBD bound to m^7GDP or GDP, the substrate is flipped 180º along the horizontal axis compared to $FluPol_A$ CBD, with the N^7-methyl group pointing out into the solvent, which could explain the difference in specificity (Fig. 7c, d) (Liu et al., 2015). On the other hand, in the structure of a capped primer-bound $FluPol_B$ heterotrimer, the methylated guanine was modelled as in the $FluPol_A$ CBD structures, with the N7-methyl group pointing into the active site (Fig. 7e, f) (Reich et al., 2017). These discrepancies highlight the need for further high-resolution structures of stalled polymerase initiation complexes in presence of capped primers, particularly as the endonuclease active site is an attractive target for small molecule inhibitor design.

A final feature worth noting in the CBD is a flexible loop formed by residues 413-428, termed the 424 loop, which protrudes out into the solvent in various different conformations in the isolated CBD structures. It was found to be essential for transcription but not cap binding and has since been suggested to play a role in guiding the capped primer generated by cap-snatching into the PB1 active site (Fig. 7e) (Guilligay et al., 2008; Liu et al., 2013, 2015; Reich et al., 2014; Tsurumura et al., 2013).

Linked to the CBD by the CBD-627 linker is the 627 domain. This domain has no catalytic activity and yet is crucial for the *in vivo* function of the polymerase, having been attributed an important role in host adaptation. It is named after residue 627 which is predominantly a glutamate in avian adapted viruses but is mutated to a lysine in mammalian adapted viruses. Polymerases containing a glutamate at this position struggle to replicate in mammalian cells unless there are compensatory mutations present (Subbarao and Murphy, 1993; Yamada et al., 2010). 627K containing polymerases on the other hand can function adequately both in avian and mammalian cells. The structure of the domain reveals an α-helical bundle that is joined to a set of β strands on one face of the bundle by a 10-residue linker. This linker in fact contains the amino acid 627, which protrudes out into the solvent (Fig. 8a).

Analysis of the surface electrostatic potential reveals a highly basic patch surrounding this residue, which is disrupted when 627E is present (Fig. 8b) (Kuzuhara et al., 2009; Tarendeau et al., 2008).

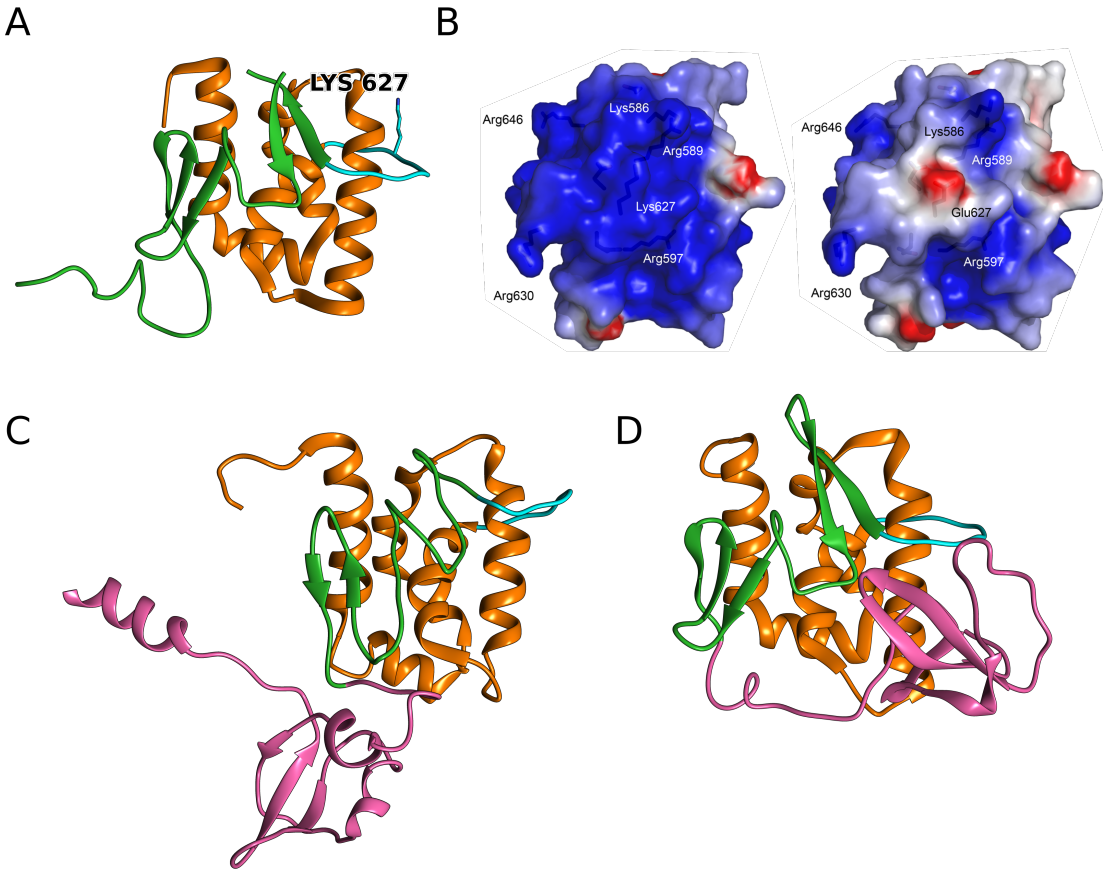


Figure 8 – PB2 627 and NLS domains. A) 627 domain with α-helical bundle shown in orange and β-sheet packed against it shown in green. The loop containing the 627 domain is shown in cyan. PDB ID 2vy7. B) Surface electrostatic potential of the 627 domain showing the disruption of a highly basic patch where glutamate is present at position 627. Figure panel obtained from Tarendeau et al., 2008. C) and D) Relative positioning of the 627 and NLS domains showing a drastic conformational rearrangement. NLS domain is shown in pink. C) is $FluPol_C$, PDB ID 5d98. D) is $FluPol_A$, PDB ID 4wsb.

The fact that it is highly exposed, together with other *in vitro* host adaptation studies, has led to suggestions that it is crucial for the interaction with a viral or host-specific cellular factor responsible for host restriction of polymerase activity (Naffakh et al., 2000). Multiple factors have been proposed to play this role, including nucleoprotein, importins, promoter RNA and DEAD box RNA helicase DDX17 (Bortz et al., 2011; Gabriel et al., 2008; Hudjetz and Gabriel, 2012; Kuzuhara et al., 2009; Labadie et al., 2007; Mänz et al., 2013; Ng et al., 2012; Paterson et al., 2014; Rameix-Welti et al., 2009). Interestingly, both activatory and inhibitory roles have been proposed for the 627 residue interactor in different studies (Mehle and Doudna, 2008; Moncorgé et al., 2010). More recently, a different family of host proteins, ANP32 (Acidic Nuclear Phosphoprotein 32), was identified as

responsible for the host restriction of avian polymerase activity. These proteins have an N-terminus which consists of leucine-rich repeats (LRR) and a low complexity acidic region (LCAR) at the C-terminus. This protein family has several members. Avian ANP32A, unlike mammalian ANP32A or ANP32B (both avian and mammalian), contains an additional stretch of 33 amino acids in the LCAR which is responsible for enabling activity by avian adapted polymerases. This is true both in avian and mammalian cells, such that expression of avian ANP32A in human cells rescues avian-adapted polymerase activity, which otherwise would be highly crippled, and knockdown of this factor in avian cells reduces polymerase activity. Mammalian ANP32A and ANP32B were also found to have a role in supporting mammalian-adapted polymerase activity (Long et al., 2016). A separate study identified ANP32A and ANP32B independently of their role in host adaptation. In it, these two proteins were shown to have a role in enabling cRNA to vRNA replication (Sugiyama et al., 2015). It is possible that these two functions are linked and that avian ANP32A is required to enable cRNA to vRNA replication although this has not yet been shown. The role of ANP32 proteins in replication will be further discussed under section 3.4.3.

An alternative hypothesis to explain FluPol host restriction implicates the 627 residue in temperature adaptation. The virus replicates at 38-41 ºC in the intestinal tract of birds but must adapt to replication at 33 ºC in the mammalian upper respiratory tract for successful infection for which residue 627 appears to be key (Aggarwal et al., 2011; Massin et al., 2001).

An additional role has been attributed to this domain independent of host adaptation. 627 domain-deleted PB2 is still capable of assembling into functional trimers which can transcribe and replicate *in vitro*, but which are defective during infection. The mechanism for this defective phenotype has been suggested to be a lack of cRNA stabilisation during replication leading to its degradation. This activity does not appear to be linked to the identity of residue 627 (Nilsson et al., 2017).

The final domain at the C terminus of PB2 is the NLS domain, which contains a bipartite nuclear localisation signal. The core of this domain forms a small globular fold composed of a three-strand beta sheet and three short helices. It is joined to the 627 domain by a long flexible linker, which enables it to undergo large conformational rearrangements (Fig. 8c, d) The C terminal tail of the NLS domain, which contains the NLS sequences, also appears to be highly flexible. The major NLS was only partially visible in the context of the heterotrimer except in the apo $FluPol_C$ crystal structure where it extends away from the core of the domain and packs against the PA endonuclease domain forming an alpha helix (Hengrung et al., 2015). In the NMR structure of the isolated domain which was initially available, this region appears to be loosely structured (Tarendeau et al., 2007).

1.3.4 Promoter RNA

The terminal ends of genomic viral RNAs in influenza comprise highly conserved sequences that stretch for 9-12 nucleotides at the 3′ end and 10-13 nucleotides at the 5′, although they differ slightly between influenza genera (Robertson, 1979). Both transcription (including cap-snatching activity) and replication rely on adequate promoter binding to occur (Cianci et al., 1995; Hagen et al., 1994; Leahy et al., 2001a; Li et al., 1998). The two strands are partially complementary and their fold has been subject to extensive work. Three models have been proposed: the panhandle, the fork and the corkscrew (Fig. 9a, b, c) (Flick et al., 1996; Fodor et al., 1994; Noble et al., 2011). The structure of the isolated promoter vRNA was determined and showed a panhandle fold, where the RNA forms a non-canonical double helix, whereas several studies supported the corkscrew model for the polymerase bound promoter (Bae et al., 2001; Brownlee and Sharps, 2002; Cheong et al., 1999; Leahy et al., 2001b; Pritlove et al., 1999; Tomescu et al., 2014). However, the promoter vRNA bound structures of $FluPol_A$ and $FluPol_B$ have shown yet another arrangement (Pflug et al., 2014; Reich et al., 2014). The two strands interact at the distal end between nucleotides 10-14 of the 3′ strand and 11-15 of the 5′ strand forming a five base pair duplex, in a manner similar to that predicted by the corkscrew model, while the proximal end of the 5′ strand (nucleotides 1-10) forms a stem loop or “hook” structure (Fig. 9d, g, i). The hook involves more extensive base pairing than that predicted in the corkscrew model, due to two A-A non-canonical base pairs which elongate the stem of the loop (Fig. 9h). The proximal

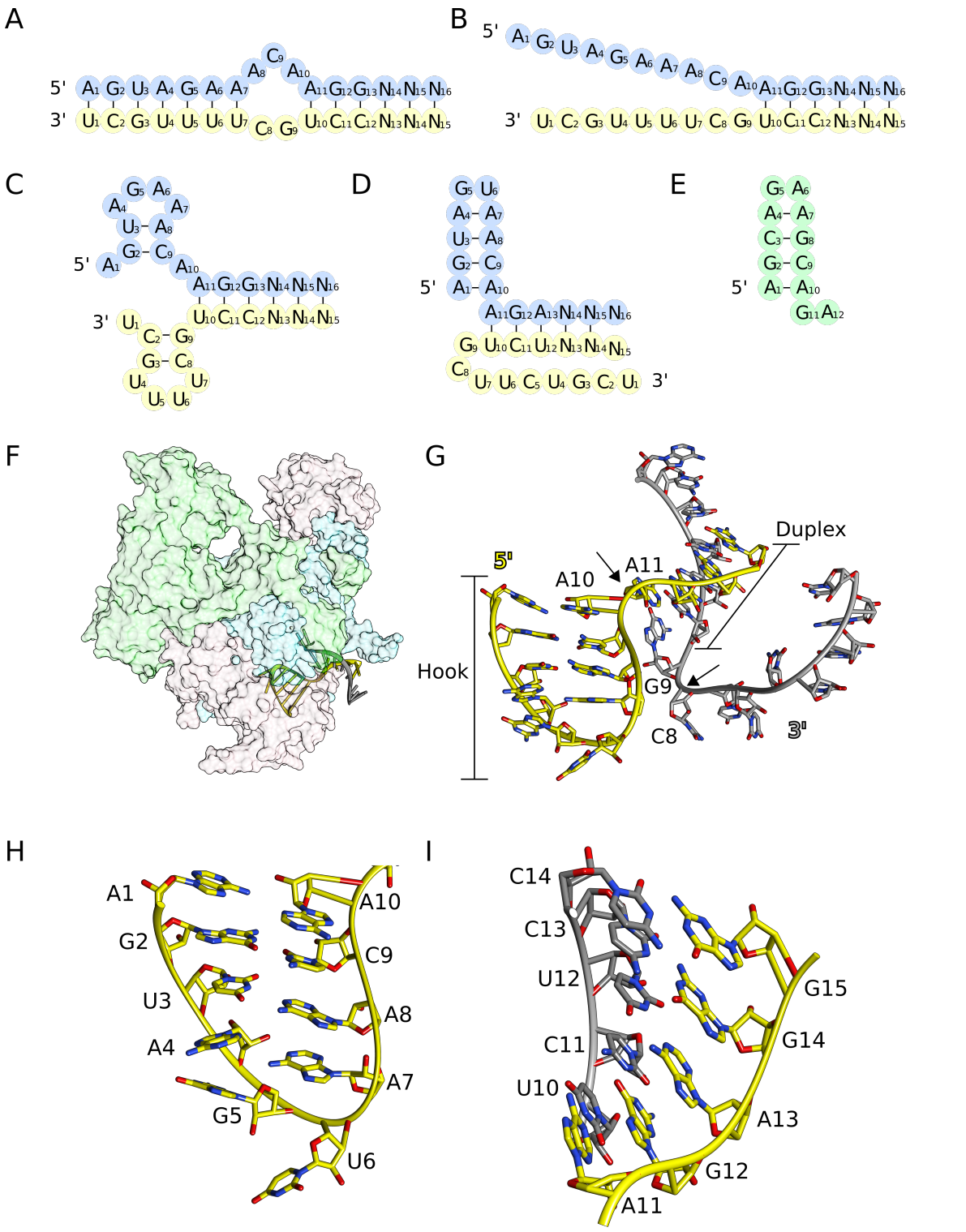


Figure 9 – Promoter RNA structure. A) Panhandle model for promoter vRNA. B) Fork model for promoter vRNA. C) Corkscrew model for promoter vRNA. D) Diagramatic representation of the promoter vRNA observed in the $FluPol_{B2}$ structure (Reich *et al*., 2014). E) Diagramatic representation of the 5′ promoter cRNA found bout to $FluPol_B$ (Thierry *et al*., 2016). F) Position of promoter vRNA in the heterotrimer. 5′ promoter vRNA in yellow, 3′ promoter vRNA in grey. Semi-transparent surface rendering of the protein with PB2 in green, PA in light pink and PB1 in cyan. PDB ID 4wsb. G) Detailed view of the promoter vRNA, highlighting the hook and duplex regions. H) Detailed view of the 5′ hook. I) Detailed view of the duplex region. PBD ID 4wrt.

end of the 3′ strand in these structures is only partially visible (up to nucleotide U7) in most of the available trimeric structures. The exception is the $FluPol_{B2}$ structure, which lacks density for a large portion of PB2 but where the full 3′ promoter strand can be modelled. It is inserted in the template entry channel that leads to the active site, but then turns sharply away between nucleotides G9 and C8 and points towards the solvent. Its overall shape is suggestive of a hairpin, but there is no base pairing within the strand, instead, the bases point towards the protein (Fig. 9g).

The promoter RNA is bound tightly by the polymerase, with a reported affinity for the duplex of 0.4 nM. In terms of the individual strands, the affinity is higher for the 5′ promoter end (Kd of 2.2 ± 0.6 nM) than that for the 3′ promoter end alone (Kd of >1 μM) (Tomescu et al., 2014). The promoter sits at the interface between the PA C-terminal domain and the PB1 subunit, close to the fingers domain and also making contacts with the flexible PB1 β-ribbon and the PB2 N terminus (Fig. 9f). Both strands contact the polymerase through numerous polar interactions, mostly with the phosphate backbone but also with specific bases. The tight bend between the hook and the duplexed region of the 5′ strand, at nucleotides A10-A11 is particularly closely stabilised by protein contacts. The cRNA promoter was also proposed to acquire a corkscrew structure, in a manner similar to the vRNA promoter (Azzeh et al., 2001; Crow et al., 2004). Currently, only a 5′ cRNA bound FluPol structure is available, where the 5′ end acquires a “hook” like conformation and binds to the polymerase at a nearly identical position to the vRNA 5′ strand (Fig. 9e) (Thierry et al., 2016).

1.3.5 Conformational flexibility of FluPol

As more polymerase crystal structures have become available over the last couple of years it has become clear that several of the domains on the heterotrimer are extremely flexible. This is particularly true for the C terminal domains of PB2 as well as the PA endonuclease domain. These differences in conformation appear to be partially dictated by the presence and type of promoter RNA (Fig. 10). In all of the vRNA promoter bound structures, the 627 domain lies close to the PA C terminal domain, with the cap binding domain placed close to the endonuclease domain, forming an

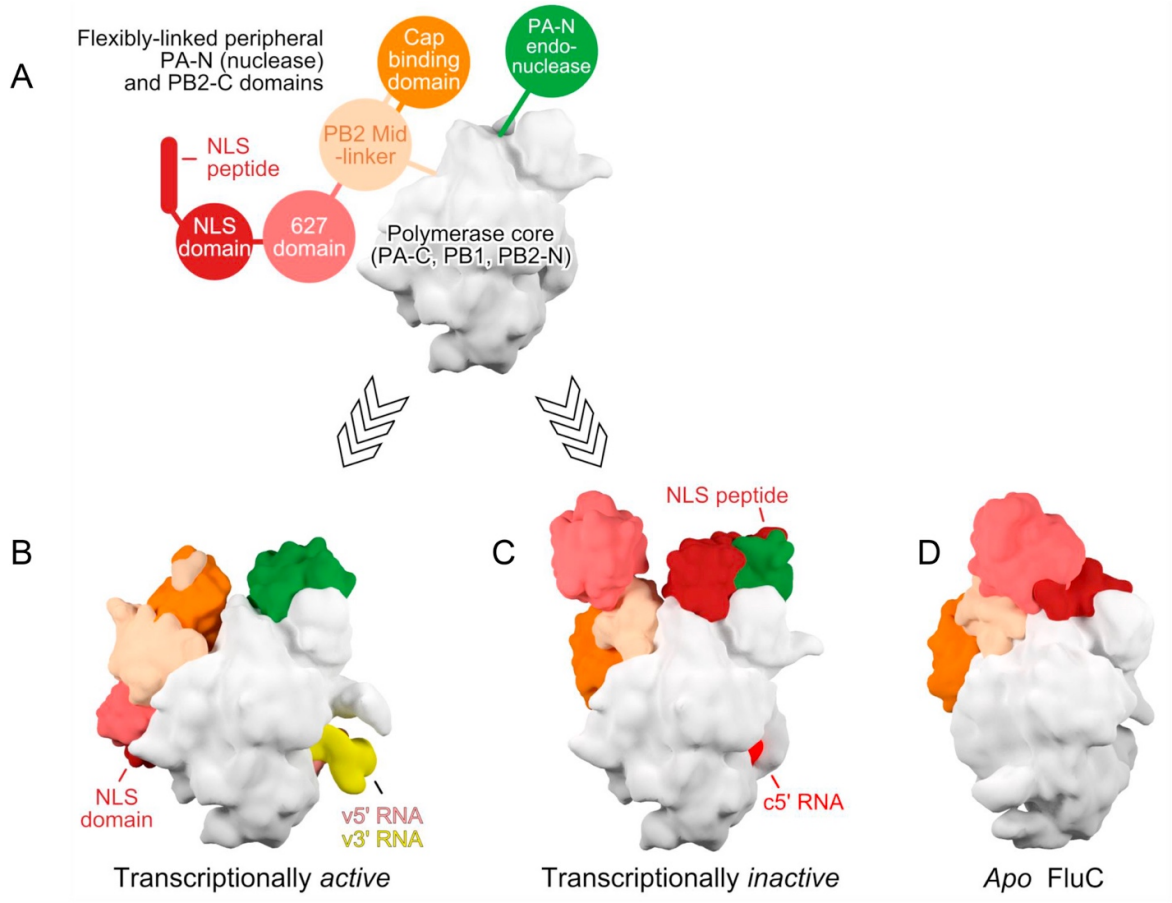


Figure 10 – Multiple configurations of FluPol. A) The heterotrimer has a stable core formed by PB1 plus the PA C terminal domain and the PB2 N terminal domain. The other domains surrounding the core are highly flexible and undergo large rearrangements which appear to be influenced by the presence and type of promoter RNA bound. B) FluPol conformation when bound to promoter vRNA C) FluPol conformation when bound to 5´promoter cRNA D) Apo form of FluPol. Figure obtained from Pflug *et al*., 2017.

overall U-shaped polymerase. In this form, the cap binding domain active site is exposed, allowing capped mRNA binding and this is indeed the conformation observed in the capped primer bound crystal structure (further discussed in section 1.4.1) (Reich et al., 2017). This conformation has been termed the pre-initiation state (Lukarska et al., 2016; Pflug et al., 2014; Reich et al., 2014, 2017). The apo $FluPol_C$ structure on the other hand, acquires a more compact or "closed" conformation, which could represent the form the polymerase acquires after assembly in the nucleus but before RNA association during infection or alternatively, a replication competent conformation (or both). In this form, the 627 domain is tightly packed against the endonuclease domain, which is itself slightly rotated, while the cap binding domain is positioned closer to the PB1 palm domain, occluding the cap binding site. The NLS domain also changes position: the linker between the NLS and the 627 domain unravels, allowing the NLS to pack close to the endonuclease domain, with the helix containing the

major NLS signal resting on it. These rearrangements require a rotation of the C-terminal PB2 subunits of 130º coupled with a translation of 90 Å. Interestingly, this conformational change appears to occur quite readily and, in solution, the heterotrimer can acquire the pre-initiation or "open" conformation even in absence of promoter RNA (Hengrung et al., 2015). The 5′ cRNA promoter bound structure displays a domain arrangement more comparable to that of the apo structure, except for the 627 domain, which protrudes out into the solvent rather than packing against the PA endonuclease domain. Yet again, this conformation would not be compatible with cap-snatching, which may explain why capped, polyadenylated transcripts have not been found to be produced from cRNA during infection (Thierry et al., 2016).

1.3.6 Nuclear import of polymerase subunits

Influenza transcription and replication occur in the nuclei of infected cells. Therefore, incoming vRNPs as well as newly produced polymerase subunits require nuclear import through nuclear pore complexes (NPCs) (Hutchinson and Fodor, 2012). The classical import pathway involves the binding of cargo via an importin α adaptor, which is in turn bound by importin β1. This complex is imported to the nucleus where cargo is released via binding of RanGTP, the concentration of which is kept high in the nucleus by Ran GEF (Ran Guanosine Exchange Factor) (Görlich and Kutay, 1999). In non-classical nuclear import, the cargo is bound directly by an importin β, without requiring an adaptor. The NLSs in the non-classical import pathway are often arginine rich, in contrast to the canonical lysine rich NLSs (Palmeri and Malim, 1999). All three polymerase subunits and NP contain one or more nuclear localisation signals. Import of incoming RNPs is thought to occur via two NLS on NP, particularly the N-terminal NLS, which interact with importin α1 and α5 isoforms (O'Neill et al., 1995; Wu et al., 2007a, 2007b). It remains unclear whether the NLSs on the polymerase subunits play a role in import of RNP complexes.

Polymerase trimer assembly appears to occur inside the nucleus. The individual subunits therefore need to be imported into the nucleus separately. PB2, as discussed previously, has an NLS domain at

its C terminus containing a bipartite NLS at residues 736–739 and 752–755. This NLS domain is capable of interaction with multiple importin α isoforms including importin α1, α3, α5 and α7. Adaptive mutations in this domain that affect importin α binding also appear to be important for host restriction, which include D701N and S714R. Indeed adaptation to achieve efficient interaction with host importin α isoforms is thought to be one of the main barriers for inter-species transmission (Boivin and Hart, 2011; Czudai-Matwich et al., 2014; Delaforge et al., 2015; Gabriel et al., 2008; Resa-Infante and Gabriel, 2013; Tarendeau et al., 2007, 2008). The NLS domain has been observed to be packed tightly against the 627 domain but is also capable of undergoing a large conformational rearrangement where the linker between the two domains extends to generate a much more flexible structure (Fig, 8c, d). It is this latter conformation that appears to be required for the interaction with importin α in order to prevent a steric clash of importin α with the 627 domain (Fig. 11a). This extended conformation is observed in the apo heterotrimer structure, and it shows the C-terminal NLS helix packing against the PA endonuclease domain (Hengrung et al., 2015). This could suggest a direct displacement of importin α occurs upon trimer assembly which doesn't require the NLS domain to fold into a more compact conformation.

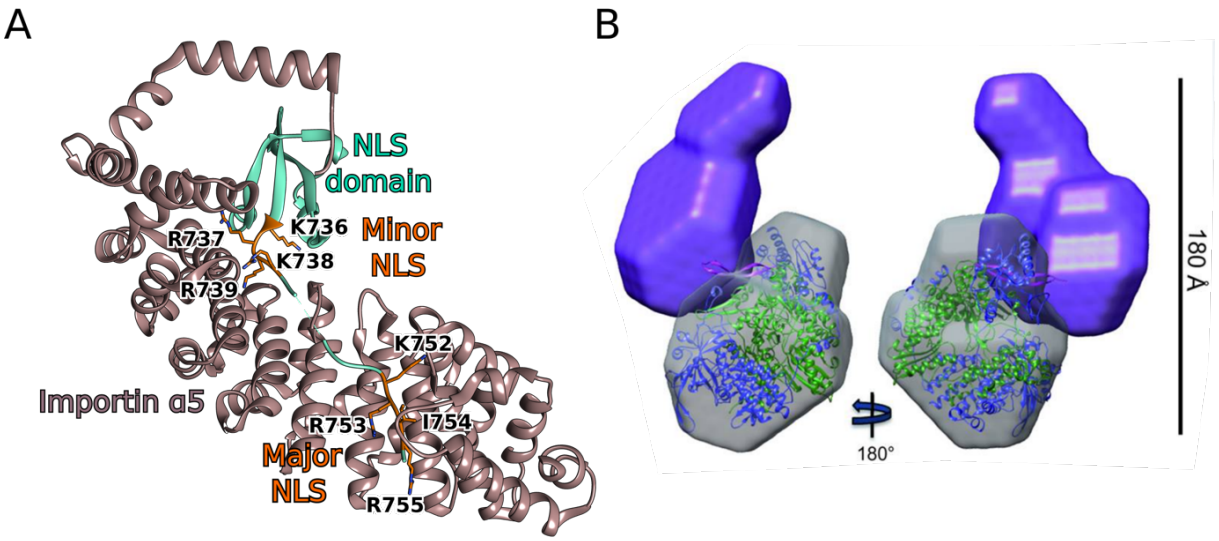


Figure 11 – Polymerase nuclear import. A) Importin α5 bound to the PB2 NLS domain. The NLS sequences are highlighted in orange. PDB ID 2jdq. B) SAXS envelope of the PB1-PA-RanBP5 complex. PB1 in green, PA in blue. PB1-PA envelope in grey, RanBP5 envelope in purple. Figure adapted from Swale et al., 2016.

The other two polymerase subunits, PB1 and PA, have been found to undergo nuclear import as a dimeric complex despite the presence of NLSs in both proteins. PA contains an NLS between residues 124–139 and 186–247 and mutations at residue 154 inhibit import (Nieto et al., 1994). However, during infection, import appears to occur via the bipartite NLS found on the PB1 β-ribbon (Hutchinson et al., 2011). This NLS interacts with RanBP5 (Ran Binding Protein 5 also known as karyopherin β3, importin β3, or importin 5) a member of the importin β family, which interacts with the PB1-PA dimer directly without requiring importin α (Deng et al., 2006a). It has been observed that nuclear accumulation only occurs efficiently for the PB1-PA dimer but not for PB1 alone. Now that the full structures of the heterotrimer are available, this observation is not surprising given how closely associated PB1 and PA are. The PB1-PA-RanBP5 complex has resisted structural characterisation, although there have been attempts to model the complex employing SAXS (Fig. 11b) (Swale et al., 2016). Separate import of PB1-PA and PB2 implies that the heterotrimer must be capable of assembly in the nucleus; this has indeed been observed to occur both *in vitro* and in cell culture (Deng et al., 2005; Huet et al., 2010).

1.4 Transcription and replication

Immediately after import of vRNPs into the infected cell nucleus, these initiate transcription and replication of the viral genome. Accumulation of mRNA has been found to occur before build-up of cRNA or vRNA (Kawakami et al., 2011). As both transcription and replication are carried out by vRNPs, the mechanism that drives the switch between transcription (early in infection) to replication (later in infection) has become a subject of interest. It was initially suggested that NP accumulation drove a conformational change in FluPol that pushed it into a replicative state. The delay in replication would therefore be due to the time needed for sufficient transcription and translation of NP to take place (Beaton and Krug, 1986). Further data has since modified this view, instead suggesting that the role of NP in regulating replication is mostly to support the polymerase in stabilising cRNA, which is in fact also produced early in infection but which is simply degraded (Mullin et al., 2004; Turrell et al., 2013; Vreede and Brownlee, 2007; Vreede et al., 2004). Therefore, in early infection, incoming vRNPs produce capped polyadenylated transcripts, which are taken over by the cellular mRNA

processing machinery to produce new viral proteins, but also cRNA, which cannot accumulate in the cell yet due to the absence of viral proteins to coat it and protect it from nucleolytic attack.

Transcription and replication are mechanistically quite different. In order to produce mRNA, the FluPol must be able to acquire a capped primer and polyadenylate, while for cRNA production it must be able to initiate polymerisation in a primer independent manner and end the transcript, to be a precise complementary copy of the vRNA. vRNA production from cRNA appears to be even more complicated, as the resident polymerase and NP in the cRNP are not sufficient for this process to occur. In addition to this, both transcription and replication need to take place in the context of an RNP, where the template RNA is coated by NP and arranged in a helical structure, which will need to be at least partially disassembled to enable access by the polymerase. While many of the mechanisms involved in these processes remain unknown, the currently available information is summarised below.

1.4.1 Transcription initiation

As previously mentioned, transcription is initiated from a primer generated from host capped RNAs by cap-snatching (Beaton and Krug, 1981; Plotch et al., 1981). Access to this RNA source is ensured by the association of FluPol with the host DNA dependent RNA polymerase II (Pol II). This interaction occurs via the C terminal domain (CTD) of Pol II, an unstructured region consisting of 52 repeats of the heptad sequence $Y_1S_2P_3T_4S_5P_6S_7$ which becomes extensively post-translationally modified, regulating the activity of Pol II (Engelhardt et al., 2005; Jeronimo et al., 2016). More specifically, FluPol binds the initiating form of Pol II, where the CTD is phosphorylated at S_5 (Lukarska et al., 2016; Martínez-Alonso et al., 2016). Association with S_5P CTD ensures the FluPol is present during the short window between pausing of the Pol II and elongation initiation, where the CTD becomes S_2 phosphorylated. During this time, the nascent mRNA becomes capped, and the cap sequestrered by the cellular cap binding complex (CBC). CBC binding is coupled with dephosphorylation of the S_5 and recruitment of kinases responsible for phosphorylation of S_2, which triggers the transition to elongation by Pol II (Lidschreiber et al., 2013). The cleavage of cellular transcripts interferes with

regular cellular transcription and depletes the number of elongating Pol II complexes (Chan et al., 2006). In addition to this, the FluPol interaction with Pol II facilitates inhibition of host gene expression, as degradation of Pol II is induced later in infection (Rodriguez et al., 2007; Vreede et al., 2010). As transcription initiation is reliant on the interaction with Pol II, it may be that the switch to replication seen later in infection is simply a product of the ratio of FluPol to Pol II in the cell. As infection progresses and more FluPol is produced as Pol II is degraded, a much greater population of FluPol won't be able to access a source of capped RNA, which could drive it to carry out replication (Vreede and Fodor, 2010; Vreede et al., 2010).

The structures of $FluPol_A$ and $FluPol_B$ bound to a S_5 phosphorylated CTD peptide have recently become available. These show the presence of two distinct binding sites in each, of which only one is conserved between the two polymerases (Fig. 12a). In $FluPol_A$, both binding sites observed are located at the PA C-terminal domain, whereas in $FluPol_B$, extra density was described perpendicular to the interface between the PA C terminal domain and the 627 domain as well as in the PA C-terminal domain (Lukarska et al., 2016). As capped RNA is produced by Pol II, the FluPol must be able to bind the cap and cleave it to produce a capped primer before Pol II progresses onto elongation, where FluPol would no longer be able to bind CTD as it becomes S_2 phosphorylated. As previously mentioned, the cap binding site is found in the PB2 C terminus in the CBD, while the endonuclease domain is found at the PA N terminus. Cap snatching produces primers that are generally 10-13 nucleotides long. This length is determined by the distance between these two domains, which was found to be about 50 Å in the pre-initiation polymerase conformation. The observed variation in primer lengths could be due to a combination of the differing cap primer sequences and therefore structures, as well as the flexibility of the PB2 C terminus and the PA endonuclease domain (Reich et al., 2014). The endonuclease domain appears to have some preference for cleavage at the 3′ end of purines or, more specifically, guanine where it is available. At the priming step, there is also a certain selectivity for primers where the terminal nucleotides at the 3′ end have some complementarity to the 3′ end of the 3′ promoter strand (Datta et al., 2013; Geerts-Dimitriadou et al., 2011b, 2011a; Hagen et al., 1995; Rao et al., 2003). This ensures efficient initiation from the primers. Overall however, the

FluPol endonuclease does not appear to be highly selective and deep sequencing reveals that cap snatching simply targets the most abundant capped RNAs in the cell, including non-coding RNAs like snRNAs (Gu et al., 2015; Koppstein et al., 2015; Sikora et al., 2014).

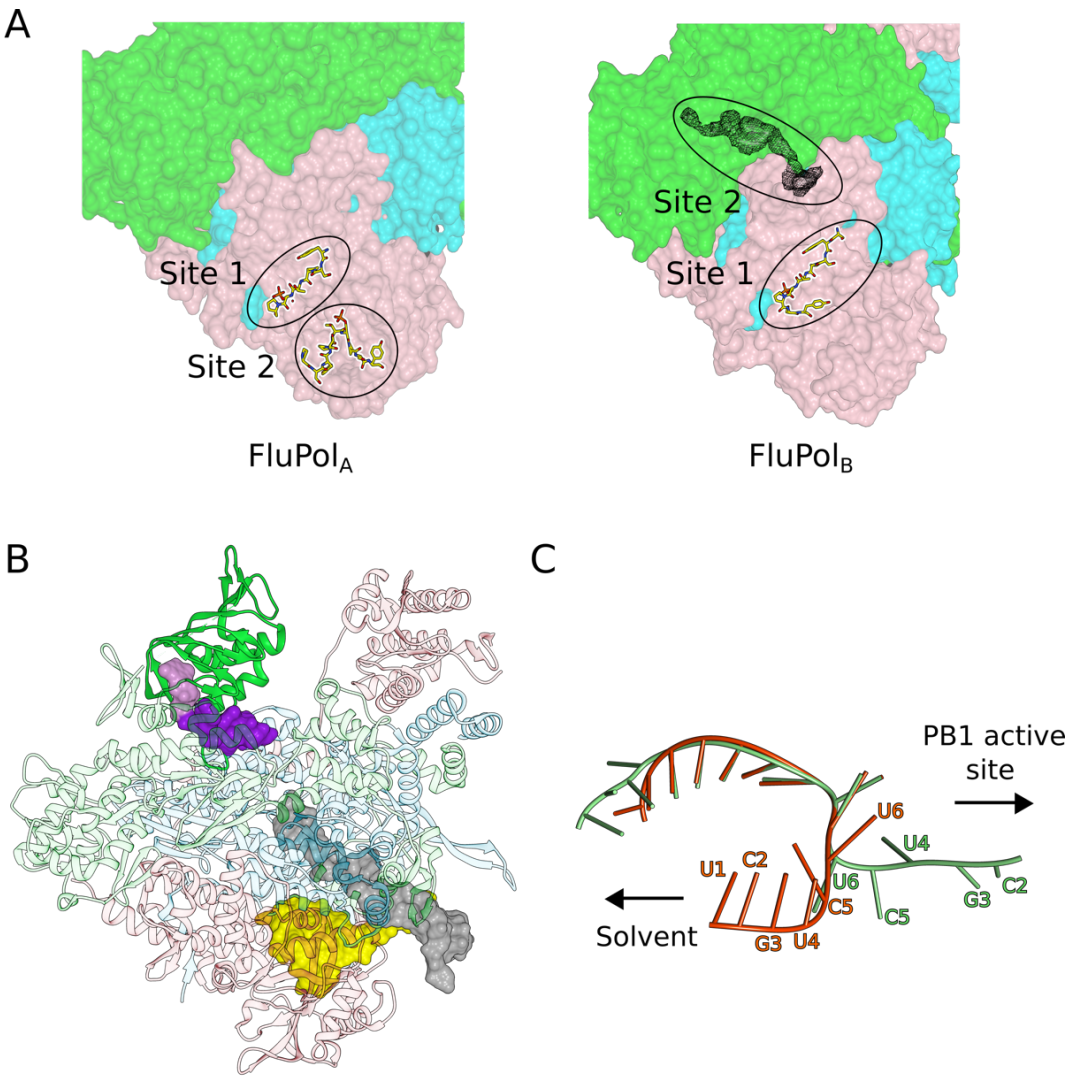


Figure 12 – Transcription by FluPol. A) Pol II S_5P CTD binding sites on FluPol$_A$ and FluPol$_B$. CTD peptide depicted in yellow or density shown in black mesh (omit difference electron density, Fo−Fc, 2.9σ) where no peptide was modelled. PA is shown in light pink, PB2 in green and PB1 in cyan. PDB ID 5m3h for FluPol$_A$, 5m3j for FluPol$_B$. B) Structural overview of FluPol$_B$ bound to capped primer (purple, with the cap structure in plum) and promoter vRNA, showing the 3′ end (grey) entering the PB1 active site. 5′ vRNA promoter shown in yellow. The capped primer is bound to the PB2 CBD in solid green. Other protein domains are depicted in translucent green for PB2, light pink for PA and cyan for PB1. PDB ID 5msg. C) Comparison of the 3′ vRNA end in its "resting" position pointing towards the solvent with the conformation it acquires when redirected towards the PB1 active site. 3′ end in the pre-initiation position depicted in red orange, PDB ID 4wrt. 3′ end entering the active site depicted in green, PDB ID 5msg.

The CBD then must rotate to redirect the capped primer into the PB1 active site, which is possible due to the fact that the domain is flexibly hinged. Two events must follow the generation of a primer: the relocation of the 3′ end of the template into the active site and the threading of the 3′ end of the primer

into the active site, allowing base pairing between the two. This initiation state was captured in a crystal structure where $FluPol_B$ is bound to 5′ and 3′ promoter RNAs as well as a capped primer (Fig. 12b). Even though the full length of the primer cannot be observed inside the binding site, this structure does show the relocation of the proximal end of the 3′ promoter strand, which is flexible at nucleotide U6, allowing the 3′ strand to either bend towards the solvent or enter the active site, which doesn't require disruption of the duplex region between the 5′ and 3′ promoter strands (Fig. 12c). The base pairing between the primer and the template was not visible in this structure, despite the design of the RNAs allowing three base pairs to form, which may suggest that this interaction is only transient during early initiation. Transcription initiation then begins with the addition of G or C to the 3′ end of the capped primer, directed by C2 or G3 of the template strand respectively. Later, the addition of more nucleotides and further base pairing, which would stabilise the template-product duplex, allows the disruption of the promoter duplex and enables elongation to proceed (Reich et al., 2017).

1.4.2 Replication initiation

Replication initiation occurs differently from a vRNA template and from a cRNA template. Interestingly, replication from a vRNA template can occur in isolated vRNPs, requiring only the resident RNP components, whereas cRNPs appear to be unable to replicate *in vitro* unless additional FluPol is provided (Vreede and Brownlee, 2007; York et al., 2013). This led to the suggestion that a transacting polymerase carries out cRNA to vRNA replication while the resident RNP polymerase sits idle (Jorba et al., 2009; Moeller et al., 2012). A role for small viral RNAs (svRNA), short, 22-27 nucleotide long RNAs identical to the 5′ end of the vRNA, has been proposed in this context. The transacting polymerase could be bound to an svRNA which would aid it in sequestering the 3′ end of the RNP-resident polymerase and therefore allow it to initiate cRNA to vRNA replication in a segment specific manner (Perez et al., 2010, 2012). The transacting polymerase hypothesis was challenged by the fact that a PB1 active site mutant polymerase (which is unable to polymerise RNA) could also activate cRNA to vRNA replication. This therefore suggests the extra polymerase has an activatory role, interacting with the resident polymerase, which can then replicate its template in *cis* (York et al., 2013). Indeed, there is some evidence for influenza polymerases being capable of dimerising or

forming higher order oligomers (Chang et al., 2015; Jorba et al., 2008). Host proteins may contribute towards regulating replication by FluPol for instance, ANP32 protein family members have been implicated in stimulating cRNA to vRNA replication both *in vitro* and in infected cells (Sugiyama et al., 2015).

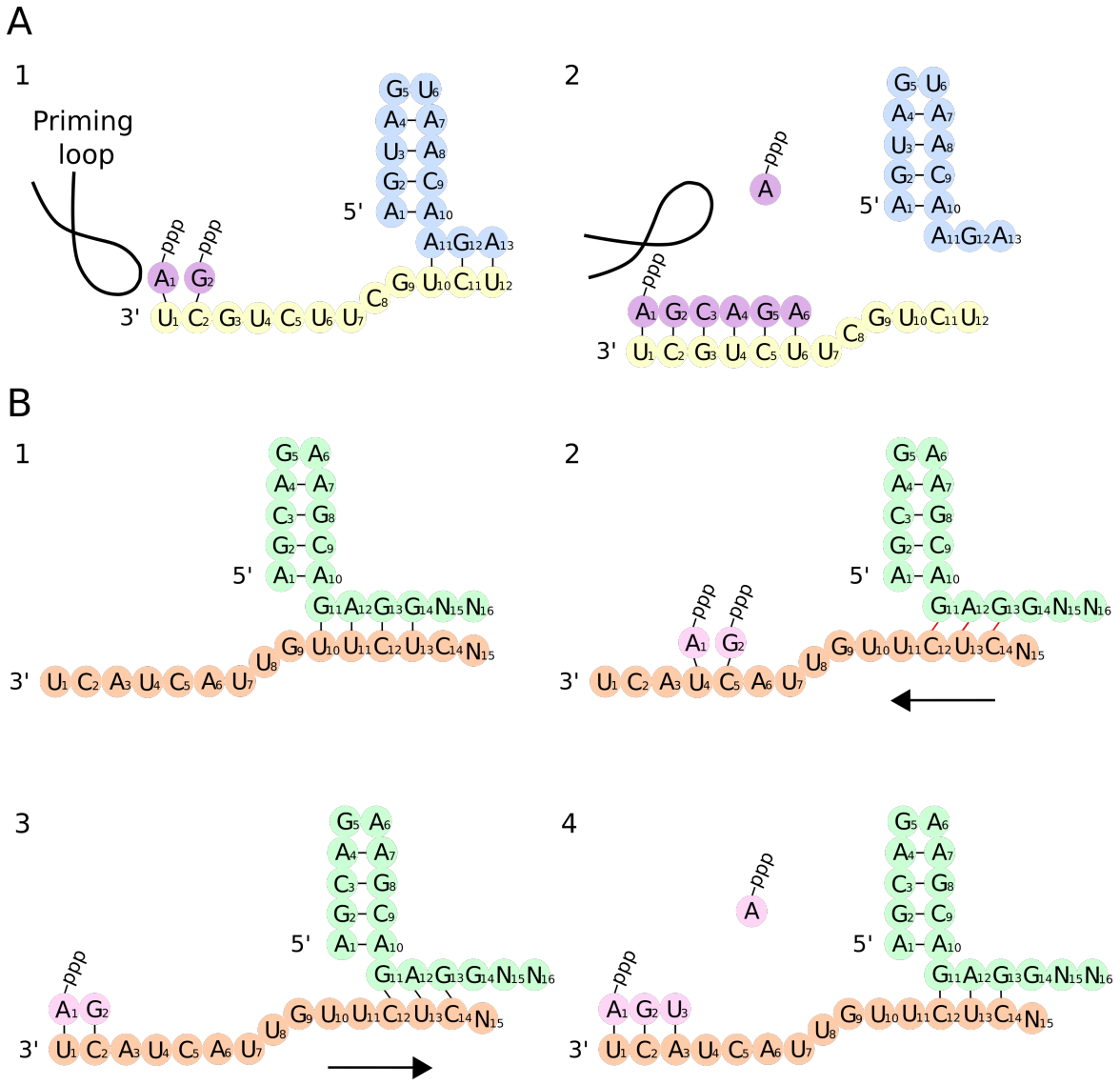


Figure 13 – Replication initiation. A) Model for terminal initiation from a vRNA template. 1- The priming loop stabilises the formation of the initial ApG dinucleotide, which base pairs at positions 1 and 2 of the 3′ end. 2- As elongation takes place, the priming loop must retract from the active site to allow space for the product-template duplex to exit the active site. B) Model for internal initiation from a cRNA template. 1- Promoter cRNA in its pre-initiation state. 2- The 3′ end shifts towards the PB1 active site, forming new duplex base pairs, which places nucleotide positions 4 and 5 at the active site. ATP and GTP base pair with the template and ApG dinucleotide is formed. 3- The strain on the duplex region forces a shift of the 3′ end backwards and the pppApG dinucleotide forms a new base pairing interaction with positions 1 and 2 of the template. 4- ApG now acts as a short primer, allowing elongation to occur.

Replication involves initiation *de novo* to generate a product with perfect complementarity to the template. This makes it a significantly (~300 fold) less efficient process than transcription, as there is no priming mechanism (Reich et al., 2017). Like in transcription, the 3´ promoter end of the RNA segment must enter the active site but in the absence of a primer, the polymerase needs to provide all the stabilising interactions to arrange incoming nucleotides in place to enable polymerisation to occur. The PB1 active site contains a flexible priming loop on a β hairpin (also found in other RNA dependent RNA polymerases) which appears to be particularly important for this purpose, supporting the sugar-base of the first incoming NTP (Butcher et al., 2001; Lescar and Canard, 2009). The priming loop has been found to be entirely unnecessary in transcription and must actually retract from the active site cavity to allow it to occur. It also hinders the rates of extension both in transcription and replication as it obstructs egress of the product-template duplex and therefore must also fold away from the active site during elongation. It is however essential for terminal initiation in vRNA to cRNA replication, where PB1 Pro 651 particularly is required at the tip of the loop to stack against the sugar-base groups of initiating NTPs (Fig. 13a) (te Velthuis et al., 2016). Initiation on a cRNA template does not occur at the terminus of the 3´ end; instead, initiation occurs internally at positions U4 and C5, generating a pppApG pseudo-primer that then realigns to positions U1 and C2 and allows replication to begin (Fig. 13b) (Deng et al., 2006b). Unlike vRNA to cRNA replication, cRNA to vRNA replication does not require the activity of the priming loop, presumably because internal initiation is a more energetically favourable process (te Velthuis et al., 2016). In order to initiate at the 4-5 position, the 3´ template must protrude further into the active site but then retract when the pppApG dinucleotide has formed. The retraction is probably driven by strong base-pairing at the duplex region of the promoter, which is under strain when the 3´ end is inserted far into the PB1 active site. Once the dinucleotide has primed replication from positions U1 and C2, further elongation would be prevented until more extensive base pairing between the product and template strands occurs as more nucleotides are incorporated, which would compensate for the disruption of the promoter duplex (Reich et al., 2017).

1.4.3 Elongation

After successful initiation, the duplex formed by the 5´ and 3´ promoter RNA ends must be disrupted in order to allow the template to freely slide through the polymerase. It has been suggested that this is made possible because base pairing between the template and product strands in the active site compensates for the disruption of base pairing in the duplex (Reich et al., 2017). After initiation of both transcription or replication, the FluPol must be able to access the remainder of the NP coated template. This is not trivial, as the RNA follows the NP helical structure, which will presumably require some degree of modification to release RNA. In addition to this, the RNA ends are likely to require shielding by protein (likely either polymerase or NP) when they dissociate from the resident RNP polymerase, as they otherwise risk being degraded or being recognised by cellular pattern recognition receptors (PRRs) that would lead to interferon induction. Multiple models have been proposed for how this takes place (Fig. 14). One hypothesis is that, as the template is pulled into the polymerase, the NP filament associated with the RNA disassembles entirely, only reforming on the template as it exits the polymerase. An alternative to this is a model that suggests that the polymerase travels over the helical filament, causing only local disruption of the structure and releasing small sections of RNA at a given time. The fact that the template entry and exit channels in the polymerase are located adjacent to each other would be in agreement with this model. It has been suggested that for cRNA to vRNA replication, a transacting polymerase could be operating in this manner, going around the NP helix replicating the template while the resident polymerase remains bound to the template promoter ends. There is some evidence to support this model, from negative stain EM images which potentially show branched RNPs, with a short RNP attached via a polymerase (detected by immunogold labelling) to a full length RNP (Moeller et al., 2012). Of course, both mechanisms could be operating in the RNP, and it has been previously proposed that transcription occurs via the first mechanism where the RNP is fully disassembled, while replication occurs via the second (Eisfeld et al., 2014).

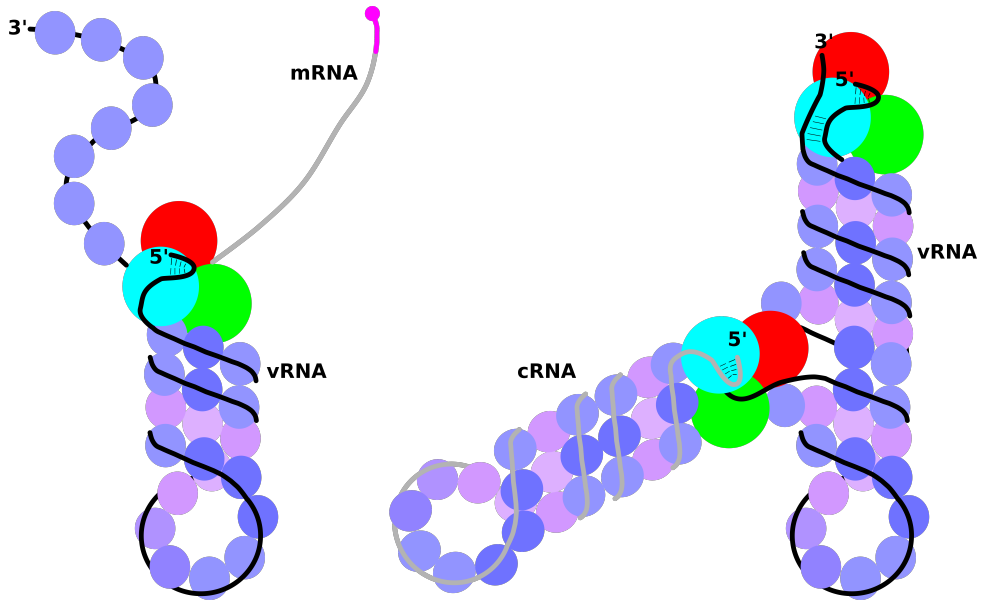


Figure 14 – Models for elongation through an RNA template in an RNP context. The first model involves disassembly of the NP helix as the polymerase pulls the template through, with the NP filament re-forming on the released 3′ end. This model is depicted as if carrying out transcription, with the mRNA produced exiting the polymerase and not associating with NP, as it is sequestered by host mRNA processing machinery. The second model involves local disruption of the RNA-NP interaction, as only a short RNA loop is released from the helix at a time and threaded through the polymerase. In this case, a trans-acting polymerase is depicted which binds the product RNA, which is in turn NP coated generating a new RNP (Moeller et al., 2012).

1.4.4 Termination

For transcription termination, the FluPol polyadenylates its mRNA transcripts, producing viral mRNAs that are effectively indistinguishable from cellular mRNAs. This occurs through stuttering of the polymerase on an oligoU track near the 5′ end of the template (nucleotides 17 to 22 from the 5′ end) caused by the 5′ hook remaining tightly bound to the polymerase (Fig. 15) (Poon et al., 1999; Pritlove et al., 1999; Robertson et al., 1981; Zheng et al., 1999). It may be that the return of the 3′ end to the promoter binding site and base pairing with the 5′ end helps to keep the 5′ hook stably bound (te Velthuis and Fodor, 2016). Eventually, the product strand must get released from the polymerase after polyadenylation produces a tail of 30-180 nt, but whether this is just a stochastic event or regulated in some way remains to be determined (Pritlove et al., 1998). After release, viral mRNAs are bound by cellular cap binding protein and taken up by the pre-mRNA processing machinery for further modification, including splicing for some segments, followed by export and translation (Bier et al., 2011).

Replication products on the other hand must be terminated without the introduction of a polyA tail. For this, the 5′ hook must be released to avoid causing the steric hindrance that leads to polymerase stuttering. How this is regulated remains unknown, but it seems likely that binding of the 5′ hook would need to be weakened. This could be occurring either via a conformational change of the polymerase or by reducing or eliminating base pairing to the 3′ promoter end (for instance, if the 3′ strand was bound by a separate polymerase).

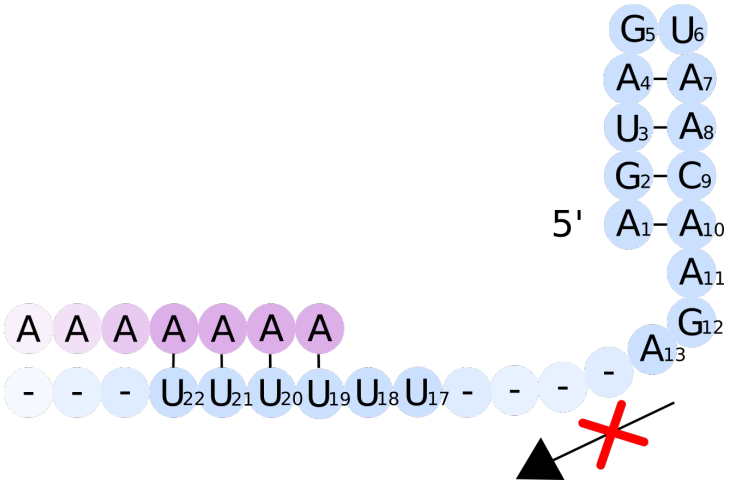


Figure 15 – mRNA polyadenylation by FluPol. As the template strand is copied during transcription, the 5′ hook on the promoter end remains bound by the polymerase. This creates steric hinderance as the polymerase attempts to pull the rest of the template through. This creates stuttering over a poly-U stretch at positions 17 to 22 from the 5′ end. As a result, a polyA tail is introduced at the 3′ end of viral mRNAs.

Project aims

The overall aim of my DPhil work was to further characterise the mechanisms by which the influenza polymerase carries out transcription and replication by employing both structural and functional approaches. In terms of functional assays, I attempted to unravel the role of several viral and host factors in regulating polymerase activity, including viral promoter RNA, nucleoprotein, cellular Pol II and ANP32 family proteins. Structural characterisation of the influenza polymerase has remained challenging despite recent breakthroughs which have produced several crystal structures of the heterotrimer. I therefore aimed to optimise expression and purification protocols and explored the use of cryo-electron microscopy as a tool to analyse the structure of FluPol. This included FluPol in complex with promoter RNA, in RNPs, as well as in complex with camelid single-chain antibodies raised against FluPol.

2 Materials and methods

2.1 Materials and reagents

2.1.1 DNA Oligonucleotides

Lyophilised oligonucleotides were purchased from Life Technologies.

5S 100	5′-TCCCAGGCGGTCTCCCATCC-3′
5S 62	5′-ACCCTGCTTAGCTTCCGAGA -3′
ANP32 Ach F	5′-GCTGAATTCATGGACATGAAGAAGC-3′
ANP32 Ach R	5′-TGGCCTCGAGTCAGTCATCC-3′
ANP32 Ah F	5′-CGAGAATTCATGGAAATGGGCAGACG-3′
ANP32 Ah R	5′-GGCCTCGAGTTAGTCGTCATC-3′
ANP32 Bh F	5′-CGAGAATTCATGGATATGAAACGCAG-3′
ANP32 Bh R	5′-GGCCTCGAGTCAGTCATCCTC-3′
CAT 2	5′-CGCAAGGCGACAAGGTGCTGA-3′
CAT 89	5′-ATGTTCTTTACGATGCGATTGGG-3′
Mut SPB1 pA F	5′-GAACTGCGTCGTCAGAAGGCATCTAGAGAGAACCTGTACTTC-3′
Mut SPB1 pA R	5′-GAAGTACAGGTTCTCTCTAGATGCCTTCTGACGACGCAGTTC-3′
NA 1280	5′-TGGACTAGTGGGAGCATCAT-3′
NA 160	5′-TCCAGTATGGTTTTGATTTCCG-3′
NA cmRNA 2	5′-CCATTGTCTTTGCTGTGTATAGC-3′
Nanobody F	5′-GTAGGTACCCCATGGCCCAGGTGCAG-3′
Nanobody R	5′-GATGGATCCGGTACTAGGCTTCAGGTTC-3′
P3 D93A mut F	5′-GAACCACTACCTTTTCTTTGTGCCATTTTCGACAGAG-3′
P3 D93A mut R	5′-CTCTGTCGAAAATGGCACAAAGAAAAGGTAGTGGTTC-3′
P3 E104A mut F	5′-CAGAGAAGAGAAACAATTCGTTGCAATTGGAATAACAAG-3′
P3 E104A mut R	5′-CTTGTTATTCCAATTGCAACGAATTGTTTCTCTTCTCTG-3′
P3 E65A mut F	5′-GAGTGATAGATTTGTGCTCATTGCAGGAAGAAAAAGAGG-3′
P3 E65A mut R	5′-CCTCTTTTTCTTCCTGCAATGAGCACAAATCTATCACTC-3′
P3 K119A mut F	5′-GATGATAGCTATTTTCAATCCGCGTTTGGTAAACTTGG-3′
P3 K119A mut R	5′-CCAAGTTTACCAAACGCGGATTGAAAATAGCTATCATC-3′
PB1 D446A/D447A mut F	5′-GACTGGACTCCAATCTTCTGCTGCCTTTGTTCTTTTTGC-3′
PB1 D446A/D447A mut R	5′-GCAAAAAGAACAAAGGCAGCAGAAGATTGGAGTCCAGTC-3′

PB2 F416A mut F	5′-CTTCTAGCTATGTGCATGATAGCTTGCAGAGATGGAG-3′
PB2 F416A mut R	5′-CTCCATCTCTGCAAGCTATCATGCACATAGCTAGAAG-3′
PB2 W369A mut F	5′-CAGAGCAGTACAATTTGAATACTGGAGTGAGCAAGAAG-3′
PB2 W369A mut R	5′-CTTCTTGCTCACTCCAGTATTCAAATTGTACTGCTCTG-3′
Protein A F XbaI	5′-GCSTCTAGAGAGAACCTGTACTTCCAAG-3′
Protein A R XbaI	5′-TCTAGATTAGGTGGACTTGCCAG-3′

2.1.2 RNA Oligonucleotides

Influenza A virus 5′ vRNA	5′-AGUAGAAACAAGGCC-3′
Influenza A virus 3′ vRNA	5′-GGCCUGCUUUUGCU-3′
Influenza C virus 5′ vRNA	5′-AGCAGUAGCAAGGGG-3′
Influenza C virus 3′ vRNA	5′-CCCCUGCUUCUGCU-3′
Influenza C virus 3′ vRNA (1-18)	5′-UAUACCCCUGCUUCUGCU-3′
Influenza C virus 3′ vRNA (5-18)	5′-UAUACCCCUGCUUC-3′
11mer	5′-ppGAAUACUCAAG-3′
20mer	5′-ppAAUCUAUAAUAGCAUUAUCC-3′

2.1.3 Plasmids

Plasmid	*Notes*
pUCDM PB1-pA PA	A protein A tag was inserted at the C terminus of PB1 in a pUCDM PB1 PA plasmid (obtained from Narin Hengrung), separated by a TEV cleavage site. Protein A was amplified by PCR from pMA-RQ PB2-TAP (obtained from Narin Hengrung) using primers Protein A R XbaI and Protein A F XbaI, which introduced XbaI restriction sites at both 5′ and 3′ ends. It was then inserted into the pUCDM PB1 PA backbone by restriction digest of both vector and insert with XbaI (Roche), followed by de-phosphorylation of the vector ends with calf intestinal alkaline phosphatase (Promega), and ligation with T4 DNA ligase (Roche). The orientation of the insert was confirmed by Sanger sequencing. The stop codon in PB1 was removed by site-directed mutagenesis using primers Mut SPB1 pA F and Mut SPB1 pA R.
pUCDM PB1-pA	A protein A tag was inserted at the C terminus of PB1 in a pUCDM PB1 plasmid (obtained from Narin Hengrung), separated by a TEV cleavage site. Produced in the same way as pUCDM PB1-pA PA.
pFL RanBP5	A synthetic gene for RanBP5 optimised for expression in insect cells (Life Technologies) flanked by XhoI and NheI restriction sites was inserted into a pFL backbone. A restriction digest of pMA-RQ RanBP5 and pFL with both XhoI (New

	England Biolabs) and NheI (New England Biolabs) was carried out and the products separated by agarose gel electrophoresis. The correct fragments were purified from the gel with a QIAquick gel extraction kit (Qiagen) and the backbone was dephosphorylated with calf intestinal alkaline phosphatase (Promega). The insert and vector were ligated with T4 DNA ligase (Roche).
pHMG PB1	Mammalian expression of influenza C/Johannesburg/1/66 PB1. Kindly provided by Nadia Naffakh (Crescenzo-Chaigne et al., 1999).
pHMG PB1 D446A/D447A	Produced by site-directed mutagenesis of pHMG PB1 using primers PB1 D446A/D447A mut F and PB1 D446A/D447A mut R.
pHMG PB2	Mammalian expression of influenza C/Johannesburg/1/66 PB2. Kindly provided by Nadia Naffakh (Crescenzo-Chaigne et al., 1999).
pHMG PB2 F416A	Produced by site-directed mutagenesis of pHMG PB2 using primers PB2 F416A mut F and PB2 F416A mut R.
pHMG PB2 W369A	Produced by site-directed mutagenesis of pHMG PB2 using primers PB2 W369A mut F and PB2 W369A mut R.
pHMG PB2 W369A/F416A	Produced by site-directed mutagenesis of pHMG PB2 F416A using primers PB2 W369A mut F and PB2 W369A mut R.
pHMG P3	Mammalian expression of influenza C/Johannesburg/1/66 P3. Kindly provided by Nadia Naffakh (Crescenzo-Chaigne et al., 1999).
pHMG P3 E65A	Produced by site-directed mutagenesis of pHMG P3 using primers P3 E65A mut F and P3 E65A mut R.
pHMG P3 D93A	Produced by site-directed mutagenesis of pHMG P3 using primers P3 D93A mut F and P3 D93A mut R.
pHMG P3 E104A	Produced by site-directed mutagenesis of pHMG P3 using primers P3 E104A mut F and P3 E104A mut R.
pHMG P3 K119A	Produced by site-directed mutagenesis of pHMG P3 using primers P3 K119A mut F and P3 K119A mut R.
pHMG NP	Mammalian expression of influenza C/Johannesburg/1/66 NP. Kindly provided by Nadia Naffakh (Crescenzo-Chaigne et al., 1999).
pC/PR CAT	Template RNA in influenza C RNP reconstitutions. Encodes chloramphenicol acetyl transferase gene flanked by C/Johannesburg/66 NS segment NCR. Kindly provided by Nadia Naffakh (Crescenzo-Chaigne et al., 1999).
pQE30 RanBP5	RanBP5 bacterial expression.
pcDNA PB1 (WSN)	Mammalian expression of influenza A/WSN/33 (H1N1) PB1.

pcDNA PB2 (WSN)	Mammalian expression of influenza A/WSN/33 (H1N1) PB2.
pcDNA PA (WSN)	Mammalian expression of influenza A/WSN/33 (H1N1) PA.
pcDNA PB1 (Fj)	Mammalian expression of influenza A/duck/Fujian/1/2002 (H5N1) PB1.
pcDNA PB2 (Fj)	Mammalian expression of influenza A(duck)/Fujian/1/2002 (H5N1) PB2.
pcDNA PA (Fj)	Mammalian expression of influenza A(duck)/Fujian/1/2002 (H5N1) PA.
pPOLI NA	Mammalian expression of influenza A/WSN/33 (H1N1) NA segment vRNA.
pGEX GST-ANP32Ah	The ANP32A human gene was amplified by PCR from pCAGGs ANP32Ah, which was kindly provided by Prof. Wendy Barclay (Long et al., 2016) using primers ANP32Ah F and ANP32Ah R, which introduced an EcoRI restriction site at the 5′ end. The PCR product and a pGEX 4Ti backbone were digested with EcoRI and XbaI, followed by dephosphorylation of the vector with calf intestinal alkaline phosphatase. The insert and backbone were ligated with Mighty Mix ligation kit (Takara).
pGEX GST-ANP32Ach	The ANP32A chicken gene was amplified by PCR from pCAGGs ANP32Ach, which was kindly provided by Prof. Wendy Barclay (Long et al., 2016) using primers ANP32Ach F and ANP32Ach R. Cloning into pGEX 4Ti was done as for pGEX GST-ANP32Ah.
pGEX GST-ANP32Bh	The ANP32B human gene was amplified by PCR from pCAGGs ANP32Bh, which was kindly provided by Prof. Wendy Barclay (Long et al., 2016) using primers ANP32Bh F and ANP32Bh R. Cloning into pGEX 4Ti was done as for pGEX GST-ANP32Ah.
pcDNA PP7-Strep	Mammalian expression of Strep-tagged PP7 coat protein. Obtained from Ashley York (York et al., 2013).
pcDNA PP7-His	Mammalian expression of His-tagged PP7 coat protein. Obtained from Ashley York (York et al., 2013).
pMESy4 nanobody	Bacterial expression of C-terminal 6xHis-tagged nanobodies.
pcDNA 3A nanobody	Mammalian expression of C-terminal 6xHis-tagged nanobodies. Nanobody genes were amplified by PCR from pMESy4 nanobody plasmids using primers nanobody F and nanobody R, the latter of which introduces a BamHI restriction site. The PCR product and a pcDNA 3A backbone were digested with NcoI and BamHI, followed by dephosphorylation of the vector with calf intestinal alkaline phosphatase. The insert and backbone were ligated with Mighty Mix ligation kit (Takara).

Table 2 – Plasmids

2.1.4 DNA-dependent RNA polymerase II CTD peptides

Unphosphorylated CTD (UnP)	Biotin - PEG - YSPTSPSYSPTSPSYSPTSPSYSPTSPS
Serine 2 phosphorylated CTD (S_2P)	Biotin - PEG - YS*PTSPSYS*PTSPSYS*PTSPSYS*PTSPS
Serine 5 phosphorylated CTD (S_5P)	Biotin - PEG - YSPTS*PSYSPTS*PSYSPTS*PSYSPTS*PS
Scrambled CTD (Scr)	Biotin - PEG - PSSSTPSSYTPSPSSSPTSYSPYYTSPP

2.1.5 Protein expression constructs

Organism	*Protein*	*Accession number*
Influenza C/Johannesburg/1/66 virus	PB1	AAF89738.2
	PB2	AAF89739.2
	P3	AAF89737.2
Influenza A/Northern Territory/60/1968 (H3N2) virus	PB1	ABF83455.1
	PB2	ABF83457.1
	PA	ABF83454.1
Influenza A/duck/Fujian/1/2002 (H5N1) virus	PB1	ADC97079.1
	PB2	ADC97078.1
	PA	ADC97081.1
	NP	AY585420.1
Influenza A/WSN/33 (H1N1) virus	PB1	J02178.1
	PB2	J02179.1
	PA	X17336.1
	NP	LC333186.1
Homo sapiens	hANP32A	NP_006296.1
	hANP32B	NP_006392.1
Gallus gallus	chANP32A	XP_413932.3

Table 3 – Protein expression constructs and corresponding Genbank accession numbers.

2.2 Cell culture

Human Embryonic Kidney 293 T (HEK293T) cells and Madin-Darby Bovine Kidney (MDBK) cells were grown in Dulbecco's Modified Eagle Medium (Sigma-Aldrich) and Minimum Essential Medium

(Sigma-Aldrich) respectively. Both were supplemented with 10 % foetal calf serum (Sigma-Aldrich), unless otherwise specified. MEM was also supplemented with 1x L-glutamine 200 mM (Gibco). Cells were kept at 37 ºC, 5 % CO_2. Cells were trypsinised with TrypLE™ Express Enzyme (Gibco).

Spodoptera frugiperda 9 (Sf9) insect cells were grown in Insect-XPRESS protein-free insect cell medium (Lonza) supplemented with 1x PenStrep (Gibco). Cell density was maintained between 0.5-2 million cells/ml. Cells were kept at 27 ºC, shaking at 110 rpm.

FreeStyle™ 293-F (HEK293F) cells (ThermoFisher) were grown in FreeStyle™ 293 Expression Medium (ThermoFisher). Cells were grown at 37 ºC, 5 % CO_2, shaking at 120 rpm and kept at densities between 0.3-3 million cells/ml.

2.3 Influenza virus

2.3.1 Influenza viruses

Virus	*Notes*
WSN	Influenza A/WSN/33 (H1N1) wild-type virus
WSN (+)	Influenza A/WSN/33 (H1N1) with a PP7 hairpin tagged NA segment, in the region encoding the glycoprotein stalk. Tagged in the positive RNA sense (mRNA and cRNA). Virus obtained from Ashley York (York et al., 2013).
WSN (-)	Influenza A/WSN/33 (H1N1) with a PP7 hairpin tagged NA segment, in the region encoding the glycoprotein stalk. Tagged in the negative RNA sense (vRNA). Virus obtained from Ashley York (York et al., 2013).

Table 4 – Influenza viruses

2.3.2 Influenza virus propagation

Influenza virus stocks were produced in MDBK cells. Virus stocks were diluted in MEM supplemented with 0.1 % FCS to achieve an MOI of 0.01-0.1. The virus was applied to MDBKs at ~80 % confluency and incubated at room temperature for 1 h. Extra media was then added to the cells, which were then transferred to 37 ºC and incubated for 72 h. The supernatant was then harvested and assayed for viral titres in a plaque assay.

2.3.3 Plaque assay

MDBK cells were grown to ~80 % confluency in 6-well plates. Serial 1/10 dilutions of viral stocks (10^{-1} to 10^{-6}) were made in MEM + 0.5 % FCS. 200 µl of each dilution were applied to the cells and these were incubated for 1 h at room temperature. 2 ml of agarose overlay (2 % agarose in PBS mixed with an equal volume of MEM + 0.5 % FCS) was then applied to the wells and allowed to set. Cells were then incubated for 72 h before being stained with 2 ml/well of Coomasie Brilliant Blue stain (0.1 % Coomassie R250, 10 % acetic acid, 40 % methanol).

2.4 Molecular cloning

2.4.1 Bacterial culture

E. coli bacterial cultures were grown in LB or 2x YT media at 37 ºC, shaking at 180 rpm unless otherwise specified. Where antibiotics were required, they were added in the following concentrations:

Antibiotic	*Working concentration*
Ampicillin	100 µg/ml
Tetracyclin	10 µg/ml
Gentamycin	10 µg/ml
Chloramphenicol	25 µg/ml

Table 5 - Antibiotics

2.4.2 Competent cells

Chemically competent *E. coli* DH5α: a 15 ml overnight pre-culture (2x YT, 10 mM $MgSO_4$, 10 mM $MgCl_2$) inoculated with a single colony was set up. 200 ml culture (2x YT, 10 mM $MgSO_4$, 10 mM $MgCl_2$) were inoculated with enough pre-culture to bring the OD_{550} to ~0.01. This was incubated until the OD_{550} reached 0.45. At this point, the culture was chilled on ice for 15 min and then centrifuged at 2000 rpm for 5 min. The supernatant was discarded and the pellet resuspended in 64 ml sterile transformation buffer 1 (30 mM KAc pH 7.5, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 15 % glycerol, 50 mM $MnCl_2$), followed by a 15 min incubation on ice. The cells were centrifuged as before and the

supernatant discarded. The pellet was resuspended in 16 ml sterile transformation buffer 2 (10 mM MOPS pH 6.8, 10 mM $RbCl_2$, 75 mM $CaCl_2$, 15 % glycerol). This was then aliquoted for use and frozen at -80 ºC.

Chemically competent *E. coli Pir*+ HC: a 5 ml LB overnight pre-culture inoculated with a single colony was set up.100 ml LB were inoculated the following day with enough pre-culture to achieve an OD_{600} of ~0.01. This was incubated until the OD_{600} reached 0.3. The culture was then chilled on ice for 15 min before being centrifuged at 3300 g for 10 min at 4 ºC. The supernatant was discarded and the pellet resuspended in 40 ml 0.1 M sterile $CaCl_2$. The suspension was kept on ice for 30 min and then centrifuged as before. The supernatant was discarded and the pellet resuspended in 6 ml sterile 0.1 M $CaCl_2$ with 15 % glycerol. This was then aliquoted for use and frozen at -80 ºC.

Electrocompetent DH10MultiBac *E. coli*: these cells were produced as described in (Bieniossek et al., 2008). A 2 ml 2x YT overnight pre-culture was inoculated with a single colony. This was then used to inoculate a 400 ml 2x YT culture supplemented with tetracycline. The culture was grown until the $OD_{600} \approx 0.25$. The culture was then chilled on ice for 15 min and centrifuged at 3500 g for 15 min at 4 ºC. The supernatant was discarded and the pellet resuspended in 500 ml sterile 10 % glycerol. The centrifugation and resuspension steps were repeated with decreasing glycerol volumes: 250 ml, 10 ml and 1 ml. This was then aliquoted for use and frozen at -80 ºC.

2.4.3 Transformation

Chemically competent cells: ~10 ng plasmid DNA were added to 50 µl competent *E. coli* and incubated on ice for 15 min followed by a 40 s heat shock at 42 ºC. 300 µl 2x YT were added to the cells, which were then allowed to recover for up to 1 h at 37 ºC. The transformed cells were plated in LB agar plates, supplemented with the antibiotics where necessary, which were incubated at 37 ºC overnight.

Electrocompetent cells: ~10 ng plasmid DNA were added to 50 µl DH10MultiBac *E. coli* and incubated on ice for 15 min. This was followed by electroporation at 2500 V. 800 µl 2x YT were added to the cells, which were allowed to recover for 8 h at 37 ºC. The transformed cells were plated on LB agar plates supplemented with the appropriate antibiotics and incubated at 37 ºC for 16-48 h.

2.4.4 Preparation of plasmid DNA

Plasmid DNA was isolated using QIAprep Spin miniprep kits (Qiagen) or Genopure plasmid maxi kits (Roche) following manufacturers' instructions.

2.4.5 Polymerase chain reaction and site-directed mutagenesis

PCR was performed using either Phusion® High-Fidelity DNA Polymerase (New England Biolabs), Pfu Turbo DNA polymerase (Agilent) or Q5® High-Fidelity DNA Polymerase (New England Biolabs) following manufacturers' guidelines. The reactions were set up with ~10 ng template DNA, 0.5 µM forward primer, 0.5 µM reverse primer and 200 µM dNTPs, in addition to the corresponding polymerase reaction buffer. Primers were designed to have a melting temperature of approximately 60-65 ºC. The reactions were run for 30 cycles.

For site-directed mutagenesis, forward and reverse primers were designed to be completely overlapping, with the desired mutation in the centre and flanked by ~15 nt on each side. First, a PCR reaction was set up with the same conditions as described above, but where the PCR was only allowed to proceed for 18 cycles. This was followed by a 1 h DpnI digest to eliminate template DNA, followed by transformation into *E. coli* DH5α.

Colony PCR was carried out to determine the success of gene insertions into plasmids. Individual colonies were resuspended in 20 µl sterile water. 5 µl of this suspension was added to a reaction mixture containing 1X GoTaq Green Buffer (Promega), 0.25 µM forward primer, 0.25 µM reverse primer, 500 µM dNTPs and 1.25 U GoTaq DNA Polymerase (Promega). The reaction has heated to 95

ºC for 2 min, followed by 30 cycles of heating to 95 ºC for 15 s, 50-60 ºC for 30 s and 72 ºC for 1 min, with a final extension at 72 ºC for 7 min. The remaining 15 µl of colony suspension were used to inoculate cultures for plasmid isolation.

DNA was checked by agarose gel electrophoresis. DNA was stained with Midori Green (Nippon Genetics).

2.4.6 Baculovirus production

Baculoviruses were produced following the MultiBac protocol (Bieniossek et al., 2008). pUCDM PB1-pA, pUCDM PB1-pA PA and pFL RanBP5 were cloned as described in section 2.1.3.

Preparation of the multigene vector: pUCDM and pFL plasmids were merged by Cre-Lox recombination. 500 ng pFL RanBP5 were added to 500 ng of either pUCDM PB1-pA or pUCDM PB1-pA PA, together with 1 µl Cre recombinase and 1 µl Cre buffer in a total reaction volume of 10 µl. This was incubated at 37 ºC for 30 min before being used to transform chemically competent *E. coli* DH5α. The transformed cells were plated on LB agar plates supplemented with chloramphenicol and ampicillin. The plasmid was then isolated from individual colonies by a miniprep DNA kit and analysed by restriction digest with NheI (New England Biolabs) to confirm the presence of all of the desired genes.

Transposition into bacmid: The multigene vector produced by merging the pFL and pUCDM plasmids was then inserted into the baculovirus genome by Tn7 transposition. 50 µl electrocompetent DH10MultiBac *E. coli* were transformed with 10 ng of the multigene vector. The transformed cells were plated on LB agar plates supplemented with tetracycline, gentamycin, ampicillin, 0.5 mM IPTG and 500 µg/ml X-gal. White colonies were used to set up 5 ml overnight 2x YT cultures supplemented with gentamycin, tetracycline and ampicillin for plasmid preparation.

Bacmid isolation: The overnight cultures were centrifuged at 8000 rpm for 3 min. The bacmid was isolated by following the first three steps of the QIAprep Spin miniprep kits (Qiagen), followed by DNA precipitation with 800 µl isopropanol. The precipitated DNA was centrifuged at 13000 rpm for 10 min and the supernatant discarded. The pellet was washed twice with 250 µl 70% ethanol, spinning at 13000 rpm for 5 min each time. The pellet was allowed to air dry and resuspended in 25 µl sterile nuclease free water (Ambion).

V_0 baculovirus production: 100 µl antibiotic free insect cell medium was added to 20 µl isolated bacmid. 20 µl Lipofectin transfection reagent (Invitrogen) were added to 100 µl antibiotic-free insect cell medium and incubated at room temperature for 5 min. The bacmid containing media was added to the Lipofectin containing media and incubated at room temperature for 30 min. Two 35 mm dishes were seeded with 0.5 million Sf9 cells in 2 ml media for every virus, which were allowed to adhere to the dish for 15 min at 27 ºC. The media was removed from the cells and 100 µl of the bacmid-Lipofectin mix were added per dish. Each dish was topped up with 3 ml antibiotic-free media and incubated for 72 h. The supernatant containing the V_0 virus was then harvested.

V_1 and V_2 baculovirus production: For V_1 virus production, 50 ml Sf9 culture at 0.5 million cells/ml were infected with 2.5 ml V_0 virus. The cells were split daily to 0.5 million cells/ml until an arrest in cell division was observed. The cells were incubated for 48 h and then the cells were centrifuged at 800 g for 15 min before harvesting the virus containing supernatant. V_2 virus was produced in the same manner, utilising V_1 virus to infect the cells and infecting larger volumes of cells as needed. Virus stocks were stored at 4 ºC.

2.5 Protein expression and purification

2.5.1 Influenza C virus RdRp

Expression: Influenza C RdRp was purified from 2 L Sf9 insect cells infected with influenza C virus RdRp baculovirus (obtained from Narin Hengrung). This virus encodes the three polymerase genes of

influenza C/Johannesburg /1/66, where the PB2 gene has a C-terminal protein A tag. Cells were grown to a density of 2 million cells/ml and infected with 6 ml V_2 virus stock per litre of culture. The cells were incubated for 72 h before harvesting.

Purification: The cells were harvested at 800 g for 15 min at 4 ºC and resuspended in 50 ml lysis buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG, 1x Complete EDTA-free protease inhibitors (Roche)). The cells were lysed by sonication (15 µm oscillation, 4 times 30 s on – 30 s off, on ice) and then centrifuged at 35000 g for 45 min at 4 ºC. The supernatant was then applied to 4 ml equilibrated IgG sepharose beads and incubated for 3 h. The beads were washed three times with 25 ml wash buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG). For every wash, the beads were incubated for 10 min, rotating at 4 ºC and then centrifuged at 1000 g for 1 min. The complex was eluted by incubation with 40 µl TEV protease in 12 ml cleavage buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG, 2.5 mM reduced glutathione) overnight at 4 ºC.

After cleavage, the beads were removed by centrifugation at 4000 g for 5 min. This was followed by filtration through a 0.22 µm filter (Merck Millipore) to remove any remaining beads before concentrating the protein to 400 µl using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa). The complex was then further purified by size exclusion chromatography on a Hi-Load Superdex 200 16/60 gel filtration column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare) at 4 ºC in gel filtration buffer (25 mM Hepes pH 7.5, 500 mM NaCl). The peak fractions were pooled and concentrated to the desired concentration using Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa).

Addition of RNA: where the polymerase in complex with promoter vRNA was purified, both 5´ and 3´ promoter RNAs were added after gel filtration in 1.2x molar excess.

2.5.2 Influenza A virus RdRp

Expression: Influenza A RdRp was purified from 2 L Sf9 insect cells infected with either NT60, NT60-mOrange or Fj influenza A virus RdRp baculovirus (obtained from Narin Hengrung). All three viruses encode the three polymerase genes of either influenza A/Northern Territory/60/1968 (H3N2) or influenza A/duck/Fujian/1/2002 (H5N1) respectively. Like the C3P baculovirus, the PB2 gene has a C-terminal protein A tag. In addition to the protein A tag, the NT60-mOrange virus has an additional mOrange tag on the PB2 C-terminus, which lies to the N-terminus of the TEV cleavage site. Cells were grown to a density of 2 million cells/ml and infected with 6 ml V_2 virus stock per litre of culture. The cells were incubated for 72 h before harvesting.

Purification: The cells were harvested at 800 g for 15 min at 4 ºC and resuspended in 50 ml lysis buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG, 1 mM DTT, 1x Complete EDTA-free protease inhibitor (Roche)). The cells were lysed by sonication (15 µm oscillation, 4 times 30 s on – 30 s off, on ice) and then centrifuged at 35000 g for 45 min at 4 ºC. The supernatant was then applied to 4 ml equilibrated IgG sepharose beads (GE Healthcare) and incubated for 3 h. The beads were washed three times with 25 ml wash buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG). For every wash, the beads were incubated for 10 min, rotating at 4 ºC and then centrifuged at 1000 g for 1 min. The complex was eluted by incubation with 40 µl TEV protease in 12 ml cleavage buffer (25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG, 2.5 mM reduced glutathione) overnight at 4 ºC.

After cleavage, the beads were removed by centrifugation at 4000 g for 5 min and filtration through a 0.22 µm filter (Merck Millipore). The complex was then concentrated to ~3 ml and diluted in heparin buffer 1 (50 mM Hepes pH 7.5, 10 % glycerol) to bring the [NaCl] to 470 mM before being further purified on a 5 ml HiTrap Heparin HP column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare) at 4 ºC. The protein was eluted in a NaCl gradient created with heparin buffer 1 and heparin buffer 2 (50 mM Hepes pH 7.5, 10 % glycerol, 1 M NaCl). The protein was injected at 47 % buffer 2 and eluted in a gradient up to 100 % buffer 2 over 20 min. The peak fractions were

pooled and concentrated to 400 µl using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa).

The complex was then further purified by size exclusion chromatography on a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare) at 4 ºC in gel filtration buffer (25 mM Hepes pH 7.5, 500 mM NaCl) or ammonium sulphate gel filtration buffer (25 mM Hepes pH 7.5, 100 mM ammonium sulphate). The peak fractions were pooled and concentrated to the desired concentration using Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa).

Addition of RNA or nanobodies:

Where polymerase had to be purified in complex with promoter vRNA or nanobodies, these were added in 1.2x and 1.5x molar excess, respectively, prior to gel filtration.

2.5.3 Ribonucleoproteins

Transfection and infection: HEK293 Freestyle cells were grown to a density of ~1.8 million cells/ml in a total volume of 1200 ml. They were centrifuged at 100 g for 5 min before being resuspended in 600 ml fresh media. 3 µg of pcDNA PP7-Strep were added per ml of cells and incubated for 5 min at 37 ºC, shaking at 120 rpm. Next, a 1:1 mixture of polyethylenimine (PEI – at 1 mg/ml) and media was prepared. 9 ml of this mixture were added per litre of culture. The cells were incubated overnight and then made up to the original total volume (1200 ml) with fresh media. The cells were incubated for a further 8 h and then centrifuged as before. The supernatant was discarded and the cells were resuspended in 30 ml of either WSN, WSN (+) or WSN (-) viruses (section 2.3.1) in MEM + 0.5 %FCS, resulting in an MOI ≈ 4. The cells were kept on ice for 1 h, before being centrifuged again and the excess virus containing media removed. The pellets were resuspended in 1200 ml fresh media and incubated for a further 16 to 24 h before harvesting.

Purification: Cells were harvested by centrifugation at 800 g for 10 min and resuspended in 30 ml lysis buffer (50 mM Tris-HCl pH 8, 200 mM NaCl, 30 % glycerol, 0.1 % OTG, 1 mM DTT, 1 mM $MgCl_2$, 1x cOmplete EDTA-free protease inhibitors (Roche)). The suspension was left rotating at 4 ºC for 1 h and then briefly sonicated (10 μm oscillations, 20 s). The lysate was cleared by centrifugation at 17000 g for 15 min at 4 ºC and the supernatant added to 7 ml equilibrated Strep-Tactin sepharose beads (IBA). This was incubated rotating at 4 ºC for 3-4 h before being washed 4x in wash buffer (20 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM $MgCl_2$, 0.05 % OTG, 10 % glycerol, 1 mM PMSF). In each wash, 20 ml wash buffer were added to the beads, incubated for 10 min rotating at 4 ºC and then centrifuged at 1000 g for 1 min. Elution of the complex from the beads was carried with 10 ml 1x Biotin elution buffer (IBA) incubated for 30 min at 4 ºC.

Where a further purification step was required, the complex was further purified in a step-wise glycerol gradient. The protein sample was concentrated to 1 ml and applied to the top of a glycerol gradient containing four layered solutions of 50 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM $MgCl_2$ and 33 %, 40 %, 50 % and 70 % glycerol. The gradients were ultracentrifuged in an SW55Ti rotor (Beckman Coulter) at 45000 rpm for 4 h at 4 ºC. Fractions (~300 μl each) were collected from the bottom of the gradient and analysed for their RNA content by primer extension to determine the RNP containing fractions. Fractions containing the complex were pooled together and concentrated as appropriate using Amicon Ultra centrifugal filter units (Merck Millipore, NMWL 100 kDa).

2.5.4 PB1-PA-RanBP5

Multiple variations of this protocol were tested which will be further discussed in chapter 4.

Expression in Sf9 cells: PB1-PA-RanBP5 were co-expressed in insect cells infected with the PB1-pA PA RanBP5 baculovirus. 2 L Sf9 insect cells were grown to a density of 2 million cells/ml and infected with 6 ml V_2 virus stock per litre of culture. The cells were incubated for 72 h before harvesting.

Purification: the purification of the complex via the protein A tag was carried out in an identical manner to that of influenza C polymerase (section 2.5.1) but the [NaCl] in the buffers was varied between 500 mM and 150 mM. After cleavage from the IgG sepharose beads, the beads were removed by centrifugation at 4000 g for 5 min. This was followed by filtration through a 0.22 μm filter (Merck Millipore) to remove any remaining beads before concentrating the protein to 1.5 ml using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa). The complex was then further purified by size exclusion chromatography on a Hi-Load Superdex 200 16/60 gel filtration column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare) at 4 ºC in gel filtration buffer (25 mM Hepes pH 7.5, 150 mM NaCl, 5 % glycerol).

The peak fractions were pooled and concentrated to the desired concentration using Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa).

Bacterial expression of RanBP5: chemically competent BL21 (DE3) *E. coli* were transformed with 10 ng pQE30 RanBP5 and plated on LB agar plates supplemented with ampicillin. A 50 ml overnight 2x YT pre-culture was inoculated with a single colony. 10 ml of this pre-culture were used to inoculate a 1 L 2x YT culture supplemented with 47.5 ml ethanol, 19 ml glycerol and ampicillin. The culture was grown at 37 ºC until the OD_{600} reached ~0.68. At that point, expression was induced with IPTG (final concentration of 0.5 mM). The culture was incubated overnight at 20 ºC, shaking at 180 rpm before harvesting.

Purification of bacterial expressed RanBP5: cells were harvested by centrifugation at 4800 g for 10 min at 4 ºC. The pellet was frozen at -80 ºC for at least 3 h before being used for purification. The pellet was resuspended in 9 ml lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 300 mM sucrose, 1x cOmplete mini EDTA-free protease inhibitors (Roche) and 5 μl DNaseI (Promega) and sonicated for 15 min (30 s on – 30 s off) in a waterbath sonicator. Imidazole was added to the lysate to a final concentration of 5 mM and the lysate was then cleared by centrifugation at 17000 g for 45 min at 4 ºC. The supernatant was applied to 1 ml equilibrated Ni^{2+}-NTA agarose beads (Qiagen) and incubated for 4 h. The beads were washed 3x in wash buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 300 mM sucrose, 30 mM imidazole, 1

mM PMSF). For each wash, the beads were incubated with 15 ml wash buffer for 10 min rotating at 4 ºC and then centrifuged at 1000 g for 1 min before discarding the supernatant. RanBP5 was eluted in 4 ml elution buffer (20 mM Tris-HCl pH 8, 100 mM NaCl, 5 mM $MgCl_2$, 250 mM imidazole, 1 mM DTT, 200 mM sucrose, 1x cOmplete mini EDTA-free protease inhibitors (Roche)) rotating at 4 ºC for 30 min.

Where a further purification step was required, the protein was concentrated to 1.5 ml using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 100 kDa) and ran through a Hi-Load Superdex 200 16/60 gel filtration column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare).

2.5.5 ANP32

Expression: chemically competent BL21 (DE3) *E. coli* were transformed with either pGEX GST-ANP32A human, pGEX GST-ANP32A chicken or pGEX GST-ANP32B human and plated on ampicillin LB agar plates. A 10 ml overnight pre-culture was inoculated with a single colony. 1 L LB medium supplemented with ampicillin was inoculated with enough pre-culture to achieve an $OD_{590}\approx$ 0.01. This was incubated at 37 ºC until the $OD_{590}\approx$ 0.5. The culture was then transferred to 30 ºC and induced with IPTG to achieve a final concentration of 1 mM. The culture was incubated for a further 6 h before harvesting.

Purification: the cells were harvested by centrifugation at 4020 g for 10 min and the pellet frozen at -80 ºC for at least 3 h. The pellet was resuspended in 10 ml lysis buffer (50 mM Hepes pH 7.9, 150 mM KCl, 0.25 % NP-40, 1 mM DTT, 1x cOmplete mini EDTA-free protease inhibitors (Roche), 5 µl DNaseI (Promega)). The suspension was sonicated 4x 30 s on-30 s off and cleared by centrifugation at 17000 g for 45 min at 4 ºC. The supernatant was applied to 1 ml equilibrated glutathione sepharose 4B resin (GE Healthcare) and incubated rotating at 4 ºC for 4 h. The beads were washed 3x in 15 ml wash buffer (50 mM Hepes pH 7.9, 100 mM KCl, 1 mM $CaCl_2$, 1 mM DTT). For each wash, the beads were

incubated in wash buffer rotating at 4 ºC for 10 min, followed by centrifugation at 1000 g for 1 min. The protein was eluted by cleavage of the strep tag with thrombin. 10 U thrombin in 0.1 % BSA were added to 3 ml wash buffer, which were applied to the beads and incubated overnight. Where a further elution was required, 3 ml elution buffer (50 mM Hepes pH 7.9, 100 mM KCl, 10 mM reduced glutathione) were added to the beads and incubated for for 30 min rotating at 4 ºC.

2.5.6 NP

Expression and purification were carried out in a similar manner as described in (Turrell et al., 2015).

Expression: 400 ml Sf9 insect cells at a cell density of ~2 million cells/ml were infected with baculovirus encoding Influenza A/NT/60/68 (H3N2) NP, which was obtained from Lauren Turrell, and incubated for 72 h.

Purification: the cells were harvested by centrifugation at 800 g for 15 min at 4 ºC. The pellet was resuspended in 15 ml lysis buffer (50 mM Hepes pH 8, 150 mM NaCl, 25 % glycerol, 0.5 % NP-40, 1 mM DTT, 1x cOmplete EDTA-free protease inhibitors, 100 µg/ml RNaseA) and the suspension lysed by sonication (15 µm oscillations, 4x 30 s on-30 s off). The lysate was cleared by centrifugation at 35000 g for 45 min at 4 ºC. The supernatant was applied to 1 ml equilibrated IgG sepharose beads (GE Healthcare) and incubated rotating at 4 ºC for 3 h. The resin was washed three times with 14 ml wash buffer (50 mM Hepes pH 8, 150 mM NaCl, 10 % glycerol, 0.1 % NP-40, 1 mM PMSF). Each wash was incubated rotating at 4 ºC for 10 min before centrifugation at 1000 g for 1 min. Nucleoprotein was eluted by cleavage of the protein A tag with TEV protease. 12 µl TEV protease were added to 2 ml cleavage buffer (50 mM Hepes pH 8, 150 mM NaCl, 10 % glycerol, 0.1 % NP-40, 1 mM PMSF, 1 mM DTT) and applied to the beads. This was incubated overnight rotating at 4 ºC.

The beads were removed by centrifugation at 4020 g for 5 min followed by filtration through a 0.22 µm filter (Merck Millipore). The sample was concentrated to ~400 µl using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 30 kDa) and further purified by size exclusion chromatography on a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare) connected

to an Äkta chromatography system (GE Healthcare) in gel filtration buffer (50 mM Hepes pH 8, 150 mM NaCl, 10 % glycerol).

2.5.7 Nanobodies

Expression: chemically competent WK6 (su^-) cells were transformed with pMESy4 nanobody expression plasmids and plated on LB agar plates supplemented with 2 % glucose and ampicillin. A 20 ml pre-culture was grown per nanobody in 2x YT medium supplemented with 0.1 % glucose, 2 mM $MgCl_2$ and ampicillin. 9 ml of this were used to inoculate 1 L 2x YT culture supplemented with 0.1 % glucose, 2 mM $MgCl_2$ and ampicillin. The culture was grown to an OD_{600} of 0.7 and then expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was transferred to 28 ºC and incubated overnight.

Purification: the cells were harvested by centrifugation at 7000 rpm for 10 min and then resuspended in 15 ml TES buffer (200 mM Tris-HCl pH 8, 0.5 mM EDTA, 0.5 M sucrose). The suspension was incubated rotating at 4 ºC for 1 h. 30 ml of a ¼ dilution of the TES buffer was then added to the suspension, followed by a further 45 min incubation. This procedure disrupts the outer bacterial membrane by osmotic shock, releasing the contents of the periplasm. This solution was centrifuged at 8000 rpm for 30 min at 4 ºC and the supernatant added to 500 µl Ni^{2+}-NTA agarose resin (Qiagen) equilibrated in phosphate buffer 1 (50 mM Na_2HPO_4:NaH_2PO_4 pH 7, 1 M NaCl). This was incubated at room temperature for 1 h. The resin was washed with 10 ml phosphate buffer 1, followed by 30 ml phosphate buffer 2 (50 mM Na_2HPO_4:NaH_2PO_4 pH 6, 1 M NaCl), centrifuging the beads at 1000 g for 1 min each time before removing the supernatant. The protein was eluted by addition of 7.5 ml acetate buffer (50 mM NaOAc, 1 M NaCl, pH 4.6). The eluate was immediately neutralised by addition of 1.5 ml 1 M Tris-HCl pH 7.5. The eluate was concentrated to 400 µl using an Amicon Ultra centrifugal filter unit (Merck Millipore, NMWL 3 kDa) and further purified on a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare) connected to an Äkta chromatography system (GE Healthcare) in a nanobody gel filtration buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl).

2.6 Protein analysis

2.6.1 SDS polyacrylamide gel electrophoresis

Protein samples to be analysed by SDS-PAGE were mixed with 4X Laemmli buffer (200 mM Tris-HCl pH 6.8, 2 % SDS, 40 % glycerol, 4 % β-mercaptoethanol, 50 mM EDTA, bromophenol blue) and heated to 95 ºC for 10 min. SDS-PAGE was carried out with mini-protean gel apparatus (BioRad). Stacking gel solution contained 125 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED and 3.2 or 4 % acrylamide (29 acrylamide:1 bis-acrylamide). Resolving gel solution contained 375 mM Tris-HCl pH 8.8, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED and 8, 10 or 12 % acrylamide (29 acrylamide:1 bis-acrylamide). Gels were run at 120-150 V for ~1 h 30 min. A Precision plus protein all blue standards' (Bio-Rad) pre-stained protein marker was used as a molecular weight marker. Gels were either stained with SimplyBlue SafeStain (Life Technologies), silver stained with SilverXpress silver staining kit (Invitrogen) following manufacturer's instructions or used for western blotting.

2.6.2 Western blots

Protein samples were separated by SDS-PAGE before being transferred onto a nitrocellulose membrane (Amersham Biosciences) using a Mini Trans-Blot cell (BioRad) at 100 V for 75 min at 4 ºC. The membrane was blocked with 2.5 % milk in PBS + 0.1 % Tween 20 for 1 h at room temperature or overnight at 4 ºC. The blot was then incubated with primary antibody for 1 h at room temperature or overnight at 4 ºC followed by 3 x 5min washes with PBS + 0.1 % Tween 20. This was followed by a 1 h incubation with a horseradish peroxidase conjugated secondary antibody at room temperature and a further three washes in PBS + 0.1 % Tween 20. The protein was visualised by chemiluminescence, using Amersham ECL western blotting detection reagent (GE Healthcare) or Merck Millipore Immobilon chemiluminescent HRP substrate.

2.7 RNA analysis

2.7.1 RNA isolation

From 35 mm dishes: the cell monolayer was resuspended in 1 ml PBS and then centrifuged at 1000 g for 4 min. The supernatant was discarded and the pellet resuspended in 200 µl PBS. 100 µl of this were used for RNA extraction. 500 µl of TRI reagent (Sigma-Aldrich) were added, followed by 100 µl chloroform and mixed. This was centrifuged at 17000 g for 20 min at 4 ºC. The clear supernatant (~410 µl) was removed and added to 400 µl isopropanol to precipitate the RNA. This was centrifuged as before and the supernatant discarded. The pellet was washed in 500 µl 70 % ethanol and centrifuged at 17000 g for 10 min at room temperature. The supernatant was discarded again and the pellet was briefly air dried. The RNA was dissolved in 8-15 µl nuclease-free water (Ambion).

From purified protein samples: 20 µl protein sample were mixed with 180 µl TRI reagent (Sigma-Aldrich) and 40 µl chloroform. The sample was centrifuged at 17000 g for 20 min at 4 ºC and the clear supernatant (~120 µl) was collected. 1 µl 10 mg/ml *E. coli* tRNA (Roche) was added to this, followed by 100 µl isopropanol. This was centrifuged as before and the supernatant was discarded. The pellet was washed in 250 µl 70 % ethanol and centrifuged at 17000 g for 10 min at room temperature. The ethanol was discarded and the pellets briefly air dried. The pellet was dissolved in 8 µl nuclease-free water (Ambion).

2.7.2 Oligonucleotide radiolabelling

1 µl of oligonucleotide at 10 µM was mixed with 1 µl [γ-^{32}P]ATP (~6000 Ci/mmol, 10 mCi/ml, Perkin Elmer), 1 µl PNK enzyme (Fermentas) and 1 µl PNK buffer A in a total reaction volume of 10 µl. This was incubated at 37 ºC for 1 h. The oligonucleotides were isolated using a QiaQuick Nucleotide Removal Kit (Qiagen) following manufacturer's instructions, eluting in 30-50 µl nuclease-free water (Ambion).

2.7.3 Capped, radiolabelled oligonucleotides

1 µl 20mer or 11mer oligonucleotides at 10 µM were mixed with 1.5 µl [α-^{32}P] GTP (~3000 Ci/mmol, 10 mCi/ml, Perkin Elmer), 0.5 µl SAM methyl donor, 1 µl methyltransferase, 1 µl capping enzyme (Epicentre Biotechnologies), 2 µl 10x buffer in a total reaction volume of 20 µl. The reaction was incubated at 37 ºC for 1 h and then heated to 95 ºC for 4 min. 20 µl formamide loading dye were added, followed by a further incubation at 95 ºC for 4 min. The samples were placed on ice and separated by RNA polyacrylamide gel electrophoresis. The products were detected by autoradiography on films (Biomax MR film, Carestream) and the radiolabelled 20mer or 11mer bands were cut from the gel and eluted in 1 ml nuclease-free water (Ambion). The eluate was de-salted on NAP-10 columns, dried and dissolved in 100 µl nuclease-free water (Ambion).

2.7.4 Primer extension

The RNA content of samples was analysed by primer extension. 1-4 µl isolated RNA were mixed with the appropriate radiolabelled primers; usually 0.25 µl vRNA binding primer and 0.25 µl m/cRNA binding primer in a total reaction volume of 5 µl. Where the RNA was isolated from cells, a 5S rRNA binding primer was also included: 0.1 µl radiolabelled 5S primer and 0.9 µl non-radiolabelled 5S primer. This was incubated at 95 ºC for 4 min and then placed on ice. The samples were then heated to 50 ºC and 5 µl of superscript containing master mix (20 mM DTT, 1x First Strand buffer (Invitrogen), 10 U/µl SuperScript III (Invitrogen), 1 mM dNTPs) were added per reaction. This was incubated at 50 ºC for 1 h, followed by a 10 min incubation at 70 ºC. The samples were then analysed by RNA polyacrylamide gel electrophoresis.

2.7.5 RNA polyacrylamide gel electrophoresis

Samples were heated to 95 ºC for 3 min, then mixed with formamide loading dye (9 ml formamide, 200 µl 0.5 M EDTA pH 8, bromophenol blue, xylene blue) and heated again to 95 ºC for 3 min. These were run on 6-20 % acrylamide:bisacrylamide (19:1) gels, containing 7 M urea and 1X TBE buffer

(90 mM Tris base, 90 mM boric acid, 2.5 mM EDTA) polymerised with 0.1 % ammonium persulphate and 0.01 % TEMED. The gels were run at 1250-2000 V for 1 h 30 min to 2 h in 1x TBE. Bands were detected by autoradiography on Biomax MS film (Carestream) using Kodak Biomax intensifying screens where necessary or on phosphor screen imaging plates (Fujifilm), which were scanned on a Fujifilm FLA-5000 scanner for quantitation. Aida Image Analyser software (Raytest) was used to determine band intensities.

2.7.6 RNP reconstitution

Components	*C/Johannesburg/1/66*		*A/duck/Fujian/1/2002 (H5N1)*		*A/WSN/33 (H1N1)*	
PB1	1 µg	pHMG PB1	0.5 µg	pcDNA PB1 (Fj)	0.5 µg	pcDNA PB1 (WSN)
PB2	1 µg	pHMG PB2	0.5 µg	pcDNA PB2 (Fj)	0.5 µg	pcDNA PB2 (WSN)
PA/P3	1 µg	pHMG P3	0.5µg	pcDNA PA (Fj)	0.5 µg	pcDNA PA (WSN)
Nucleoprotein	1 µg	pHMG NP	2 µg	pcDNA NP (Fj)	2 µg	pcDNA NP (WSN)
Template	1 µg	pC/PR CAT	1 µg	pPOLI NA	1 µg	pPOLI NA
Others	-	-	1 µg	pcDNA nanobody	1 µg	pcDNA nanobody

Table 6 - Transfected plasmids in RNP reconstitutions

HEK293T cells were seeded into 35 mm dishes and transfected with a 1:1 mixture (µl:µg) of Lipofectamine 2000 (Invitrogen) and DNA in 500 µl OptiMEM (Gibco). The plasmid quantities transfected are displayed in table 2. Cells were harvested 30-48 h post-transfection and the RNA content analysed by primer extension.

2.7.7 RNP activity assays

RNP complexes purified on a glycerol gradient were used for *in vitro* activity assays. 10 µl RNPs at ~150 ng/µl were incubated at 30 ºC for 16 h with 1 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 0.5 mM UTP, 5 mM $MgCl_2$, 2 mM DTT, 0.5 U/µl RNasin (Promega) and either 0.5 mM ApG (Fischer IBA) or 4.69 ng/µl β-globin mRNA (Sigma) in a total reaction volume of 20 µl. The RNA was then isolated and analysed by primer extension.

2.7.8 Polymerase activity assays

ApG primed, β-globin mRNA primed or unprimed activity assays:

1.5 µl purified influenza virus polymerase at 325 nM were incubated for 2 h at 30 ºC with 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 10 µM GTP, 5 mM $MgCl_2$, 2 mM DTT, 1 U/µl RNasin (Promega), 0.1 µl [α-^{32}P] GTP (~3000 Ci/mmol, 10 mCi/ml, Perkin Elmer), 0.67 µM 5´ promoter RNA and 0.67 µM 3´ promoter RNA in a total reaction volume of 3.435 µl. Where ApG or β-globin mRNA were required, they were added to a final concentration of 0.5 mM and 1.7 ng/µl respectively. Where indicated, purified nanobodies were added to a final concentration of 70 ng/µl. Pol II CTD peptides were added where indicated at a final concentration of 1.42 µM.

Capped, radiolabelled oligonucleotide primed assays:

Capped, radiolabelled 11mer and 20mer were produced as described in section 2.7.3. 0.5 µl of 11mer/20mer were mixed with purified polymerase at a final concentration of 142 nM, in addition to 7.5 mM $MgCl_2$, 2 mM DTT, 1 U/µl RNasin (Promega), 0.5 µM 5´ promoter RNA and 0.5 µM 3´ promoter RNA in a total reaction volume of 3.96 µl. ATP, UTP, CTP and GTP were added at a final concentration of 0.5 mM each in reactions where extension from a capped primer was being assayed. Pol II CTD peptides were added where indicated at a final concentration of 1.42 µM. The reactions were incubated at 30 ºC for 1 h unless otherwise stated.

2.8 Crystallography

Crystallisation trials were carried out by the vapour diffusion method on sitting drops in 96-well plates (CrystalQuick 96-well Greiner plates) or 24-well plates. These were set up either manually for the 24-well plates, or using a Cartesian Technologies robotics system for the 96-well plates (Genomic Solutions). For the 24-well plates, 1 µl protein plus 0.5 µl reservoir and 0.25 µl crystal seeds were placed in a raised well above 500 µl of mother liquor. In the 96-well plates, 200 nl protein solution and 100 nl reservoir were placed in the central shelves of each well. Each well contained 95 µl mother liquor. Multiple crystal screens were tested including: Block 1, Block 2, Block 3, Block 4 and Block 5 matrix screens (Oxford Protein Production Facility; note, these block matrix screens include conditions from a range of other crystallisation screens from Hampton Research and Emerald BioSystems) as well as Morpheus (Molecular Dimensions), PEGRx (Hampton Research), ProPlex (Molecular Dimensions), Nuc-Pro (Jena Bioscience), PGA (Hampton Research) and SaltRX (Hampton Research). Silver Bullets additive screen (Hampton Research) was also included in some of the trials. Different protein to reservoir ratios, seeding concentrations as well as dilutions of the mother liquor were tested. Further details of the conditions used can be found in section 4.4. Cryo-protection of crystals was carried out by adding 0.5 µl of a solution of mother liquor supplemented with 25 % glycerol to the crystal-containing drop. The crystal was mounted and then plunged into liquid nitrogen. Crystal diffraction was analysed at the Diamond Light Source, Didcot, UK. Diffraction data was auto-processed with XIA2 (Winter et al., 2013).

2.9 Electron microscopy

2.9.1 Negative stain EM

Protein samples for negative stain were diluted to a final concentration of 0.01-0.1 mg/ml and applied to glow discharged 200 mesh carbon film coated copper grids (Agar Scientific). The protein was incubated on the grid for 1 min before blotting off the excess. The grid was washed in ddH_2O, followed by a 30 s incubation on 0.75 % uranium formate and allowed to dry. Stained grids were imaged on an FEI Tecnai T12 transmission electron microscope (120 keV).

2.9.2 Cryo-EM sample preparation

All cryo-EM grid preparation was carried out on a Vitrobot Mark IV (FEI). Generally, protein was diluted to 0.3-0.4 mg/ml before applying 3-4 µl onto C-flat holey carbon copper grids (2.0 µm holes, 1.2 µm spacing, 200 mesh, Electron Microscopy Sciences) and incubating for ~30 s before plunging into liquid ethane. The Vitrobot was set to plunge at 20 ºC, 95 % humidity, blot-force -25, 3.5 s blotting time. Other grid types were also employed during optimisation: Quantifoil holey carbon coated gold grids (1.2 µm holes, 1.3 µm spacing, 300 mesh, Agar Scientific), Quantifoil holey gold coated gold grids (1.2 µm holes, 1.3 µm spacing, 300 mesh, Agar Scientific), ultrathin carbon coated lacey support copper grids (300 mesh, Ted Pella), graphene oxide on lacey carbon support copper grids (400 mesh, Agar Scientific), graphene oxide on Quantifoil holey carbon copper grids (2 µm holes, 2 µm spacing, 300 mesh, Agar Scientific).

2.9.3 Cryo-EM data collection

Images were collected on an FEI Tecnai Polara microscope (300 keV) fitted with a K2 direct electron detector (Gatan). Usually, data was collected at defocus values between -2 to -4.5 µm using Serial EM software. Alternatively, data was also collected on an FEI Titan Krios (300 keV) fitted with a K2 direct electron detector (Gatan). FEI EPU software was used for automated data collection at -2 to -4.5 µm defocus values.

2.9.4 Cryo-EM data processing:

Contrast transfer function correction was carried out with CTFFIND 3 or GCTF software (Mindell and Grigorieff, 2003; Zhang, 2016). Particle picking was carried out on EMAN (Ludtke et al., 1999) and further particle processing, 2D classifications, 3D classifications and refinement were performed on Relion (v1.4 or vlion) (Scheres, 2012).

3 Influenza ribonucleoprotein structure and function

3.1 Introduction

Transcription and replication are carried out by ribonucleoproteins during infection. Structural and functional data on the polymerase heterotrimer in isolation has provided valuable insight into the mechanisms involved in polymerase activity. On the other hand, multiple aspects of transcription and replication remain poorly characterised. Many of these would be better addressed by studying the polymerase in the context of a full ribonucleoprotein, one of which is replication initiation from a vRNA template versus a cRNA template. During replication, both template and product RNAs are associated to NP and FluPol. It is clear therefore, that understanding how RNPs are assembled will be an essential step in elucidating the mechanisms involved in this process.

vRNP-associated polymerase is capable of transcribing and replicating the vRNA template in *cis* both *in vitro* and in cells (Vreede and Brownlee, 2007). cRNP-associated polymerase however, was found to be incapable of replicating to produce vRNA unless additional purified polymerase was present (York et al., 2013). The promoter vRNA and cRNA sequences are only different at a few nucleotide positions (Fig. 16), and the cRNA promoter has not one, but two sites which could template pppApG dinucleotide formation for initiation. This makes it difficult to understand why replication doesn't occur from a cRNA template unless other factors are present. The role of free, additional polymerase in replication has been studied by a variety of methods. Incoming vRNPs during infection are known to transcribe immediately, without the requirement for other proteins in addition to those found within the vRNPs. Interestingly, they were found to also produce cRNA, which was only detected when FluPol was pre-expressed in the infected cells. The pre-expressed polymerase could have a PB1 active site mutation which renders it inactive and it would still enable cRNA accumulation. This points towards a role for free polymerase in binding and stabilising replication products, protecting them from degradation (Vreede and Brownlee, 2007). Further evidence for this was found in a genetic trans-complementation assay, where a polymerase containing mutations which prevent it from carrying out

replication could be incorporated into progeny RNPs, suggesting the free polymerase simply acts to receive the product of replication but doesn't enzymatically act upon it (Jorba et al., 2009). This presumably would also be the case when cRNPs replicate as the product vRNA needs to become FluPol and NP associated to make progeny vRNPs.

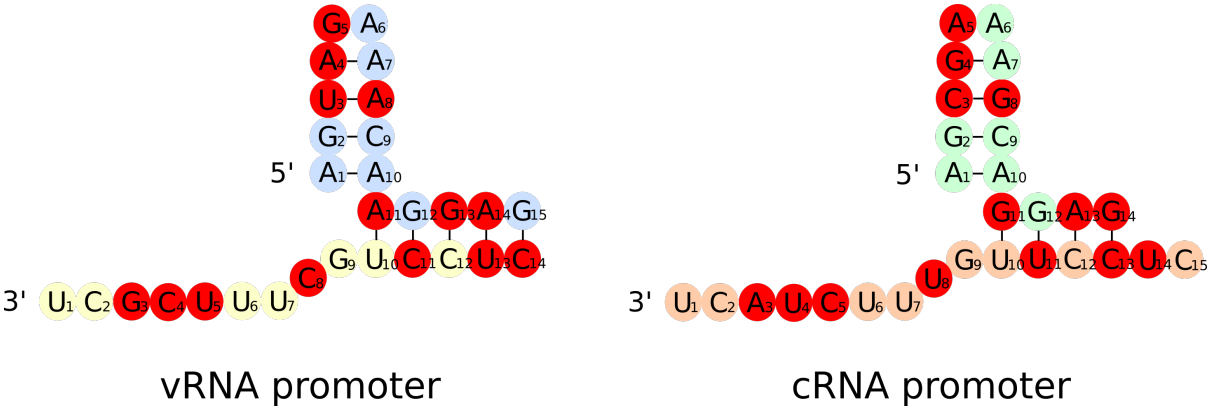


Figure 16 - Comparison of the NA segment influenza A/WSN/33 virus promoter sequence in vRNA and cRNA. Nucleotides which differ between the two are highlighted in red.

However, a further role was proposed for "free polymerase" in replication. An RNP containing a polymerase incapable of carrying out replication could generate a product when additional replication-competent polymerase was co-expressed (Jorba et al., 2009). It is worth noting that an exchange of polymerases doesn't seem to occur in RNPs. For instance, an RNP containing a polymerase with mutations that inhibit transcription cannot produce mRNA even when a transcriptionally-active polymerase is co-expressed (Jorba et al., 2009). This would suggest that the additional, active polymerase is replicating the template in *trans*. This led to the development of the "transacting polymerase" hypothesis where the resident polymerase sits idle while another polymerase carries out replication in *trans*. Negative-stain EM images of vRNPs isolated from transfected cells appear to show branching RNPs, where a core, full-length RNP has a shorter RNP attached to it via a polymerase, which is detected via a 5 nm Nanogold particle tag at the PB2 C-terminus (Moeller et al., 2012). *In vitro*, purified cRNPs are inactive in replication unless additional polymerase is provided. However, an inactive polymerase (PB1 active site mutant) was found to also be able to stimulate replication by cRNPs, which implies that this effect is independent of its catalytic activity (York et al.,

2013). The “transactivating” polymerase model was therefore proposed, where the free polymerase interacts with the resident RNP polymerase and promotes replication in *cis*. Both the transacting and transactivating polymerase models may require an interaction between two polymerase heterotrimers, in one case in order to gain access to the 3´ RNA end for replication and, in the other, for activation of the resident polymerase. There is some evidence of direct FluPol interactions in polymerases isolated from mammalian and insect cells (Chang et al., 2015; Jorba et al., 2008; Swale et al., 2016). This would fit with observations made in polymerases from other positive and negative sense RNA viruses including Poliovirus, HCV and Sendai virus, all of which undergo oligomerisation that is necessary for function, although it is worth noting that none of these display nuclear replication (Çevik et al., 2003; Hobson et al., 2001; Lyle et al., 2002; Wang et al., 2002).

Some structural information has become available over the last couple of years on vRNP complexes (discussed in section 1.2) which show a single polymerase heterotrimer binding to the RNA segment ends, with the rest of the RNA coated by NP and acquiring a helical arrangement (Arranz et al., 2012; Coloma et al., 2009; Moeller et al., 2012). On the other hand, there is very limited structural data available for FluPol in association with a cRNA template. The 5´ cRNA bound FluPol structure offered insight into the mechanism which prevents transcription from occurring from a cRNA template, showing an occluded cap binding site which would prevent cap-snatching from taking place (Thierry et al., 2016). However, it provided very little insight into replication. Structural data on full cRNPs is very scarce, with only negative stain EM images that show an identical overall arrangement to that observed for vRNPs (York et al., 2013). The lack of structural information on cRNPs is likely due to the difficulties involved in isolating them. Being replication intermediates, they are significantly less abundant in infected cells and therefore yields have been a limiting factor. Successful separation of cRNPs from vRNPs for their use in functional assays has also been challenging but was made possible by the development of a protocol which enabled selective purification based on an RNA tag (York et al., 2013).

Study of the mechanisms involved in elongation through an NP-associated RNA template and transcription/replication termination are also aspects of FluPol activity that will likely benefit from further analysis of RNP complexes both from a structural and a functional standpoint. Nevertheless, my focus for this project was to further characterise the requirements for replication initiation, more specifically, to attempt to discern the requirements for initiation from a cRNA template and why these are different to those from a vRNA template, as well as to further explore the role of FluPol dimerisation in replication. My aim was therefore to optimise the procedure for isolation of RNPs from infected cells, particularly that of cRNPs, to enable structural studies by cryo-electron microscopy. Higher resolution structural information on RNPs would enable comparison of the polymerase ends, which likely contain the key features which result in the replication activity differences observed in cRNPs and vRNPs. Analysis of the helical region and its interface with the polymerase would allow us to more definitively identify the FluPol-NP interface, as well as the path the RNA takes through the NP filament. I also attempted to further characterise the requirements for cRNA to vRNA replication through *in vitro* functional studies, particularly the role of host restriction factor ANP32, which was recently identified to play a role in this step of the influenza infection cycle.

3.2 Purification of vRNPs and cRNPs from infected cells

Previous work carried out by Ashley York established a methodology for the isolation of RNP complexes from infected cells (York et al., 2013). In it, an RNA sequence which folds into a hairpin was introduced into the genome segment encoding neuraminidase (NA), more specifically, the stalk region of NA, which was found not to be detrimental for viral growth (as determined by plaque assay). This RNA hairpin is recognised by the PP7 *Pseudomonas aeruginosa* bacteriophage coat protein (PP7 CP), a 14 kDa protein, which was C-terminally His tagged (Lim and Peabody, 2002). Recognition of the hairpin is both sequence and secondary structure dependent. While a similar secondary structure may form with this sequence both in the positive and the negative RNA sense, the identity of the nucleotide bulging at the stalk of the hairpin, as well as those in the hairpin loop are crucial for binding to the coat protein (Fig. 17a) (Lim and Peabody, 2002). These would of course not be the same in a complementary RNA. This means that when the sequence is introduced in the negative

sense (i.e. vRNA tagged), the positive sense RNA won't be efficiently recognised by the PP7 CP and vice versa (Fig. 17b). Two viruses were made in an influenza A/WSN/33 (H1N1) background, one tagged in the positive sense and one in the negative sense, which allowed the separation of vRNPs and cRNPs. In the original protocol described in York *et al*., 2013, the viruses were used to infect a HEK293T cell line stably expressing PP7-His at an MOI of 5. The cells were then harvested and RNPs purified by affinity chromatography on Ni^{2+} NTA beads, followed by a step-wise glycerol gradient (33 %, 40 %, 50 % and 70 %). This yielded relatively pure preparations that could be used for *in vitro* activity assays or negative stain electron microscopy (Fig. 17c). However, the protein concentration was nowhere near sufficient for cryo-electron microscopy, particularly for cRNPs, which are significantly less abundant in infected cells compared to vRNPs. In addition to that, the glycerol gradient purification step leaves the RNPs in a buffer containing high levels of glycerol which are not compatible with cryo-EM sample requirements.

I therefore sought to modify this protocol to improve the yields of the complex, as well as adapt it to make the sample suitable for cryo-EM analysis. As the RNPs are purified from infected cells, there is limited scope for manipulating expression levels. It therefore became clear that the easiest way to increase the quantity of the complex isolated was to simply upscale the numbers of infected cells. This however was not trivial with adherent cells, as we required over a 10-fold improvement in yield, which would make the procedure impractical due to the large surface area of cells that would be needed. Use of HEK293 cells in suspension (HEK293 Freestyle) allowed me to circumvent this issue, as relatively small volumes of media can contain large numbers of cells. That way, I achieved sufficient yields of RNPs from only 1200 ml of culture, which would have been roughly equivalent to 1.65 m^2 (108 x 15 cm dishes) of adherent cells. Rather than making a stable cell line, the cells were transfected with the desired PP7 coat protein encoding plasmid using PEI (polyethylenimine). The transfected cells were allowed to express PP7 for 8 hours prior to infection to ensure PP7 CP could be expressed before RNP production during infection.

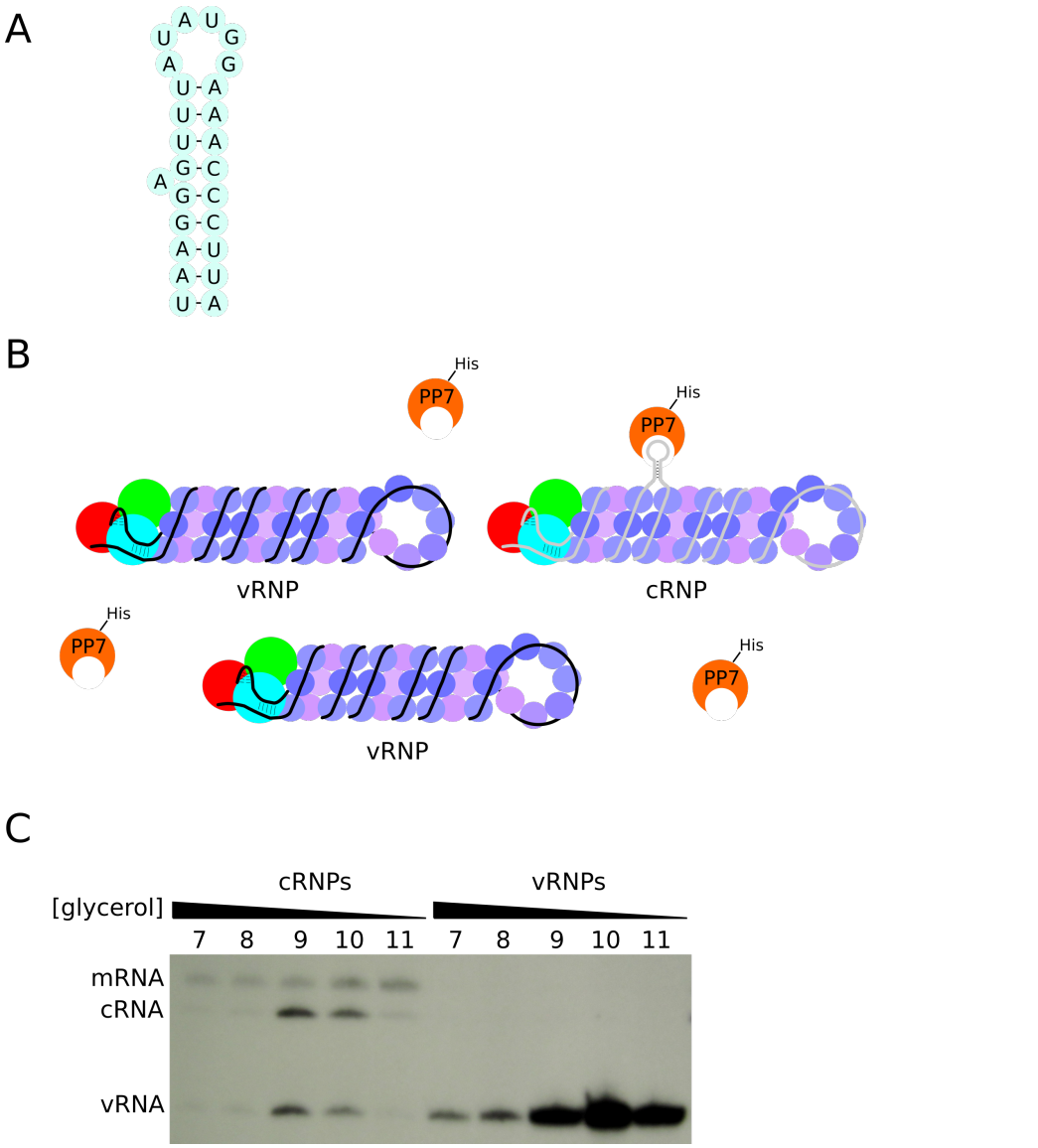


Figure 17 – Purification of vRNPs and cRNPs. A) Diagrammatic representation of the hairpin recognised by the PP7 coat protein. B) Diagram showing the purification of cRNPs via the PP7 hairpin. The hairpin is bound by the PP7 coat protein, which contains a His tag, allowing its purification. C) Primer extension (using NA160 and NA 1280 primers – see Fig 22a) of the RNA isolated from glycerol gradient fractions containing RNPs. These RNPs were isolated from infected HEK293T PP7-His cells.

The infection procedure had to be adapted for suspension cells. Traditionally, influenza virus infects through the apical cellular membrane, which cannot occur in suspension cells as they lack polarity. Where normally a cell monolayer would be covered with a small volume of virus preparation to achieve the desired MOI (in this case, an MOI of 5), suspension cells were centrifuged to form a loose pellet, which was gently resuspended directly in the virus stock. The pellet was incubated on ice for 1 hour to allow the virions to adhere to the cell surface. Excess medium was then removed by centrifugation and the resulting pellet resuspended in pre-warmed media, which would allow infection

to proceed. Successful infection was assessed by analysing the levels of viral vRNA, cRNA and mRNA produced by the cells. RNA was isolated through phenol-chlorophorm extraction and analysed by primer extension using NA segment-specific primers. Cell samples were taken at several time points after infection to determine which one resulted in the highest levels of vRNA and cRNA (Fig. 18a). vRNA quantities were found to gradually increase over time, whereas there was an initial peak of mRNA production 6 hours post-infection followed by a drop at later time points (which correlates with the idea that transcription occurs soon after infection but then halts, with RNPs mostly performing replication at later time points). As all time points appeared to contain robust levels of vRNA and cRNA, cells were harvested 16-24 hours post-infection for convenience.

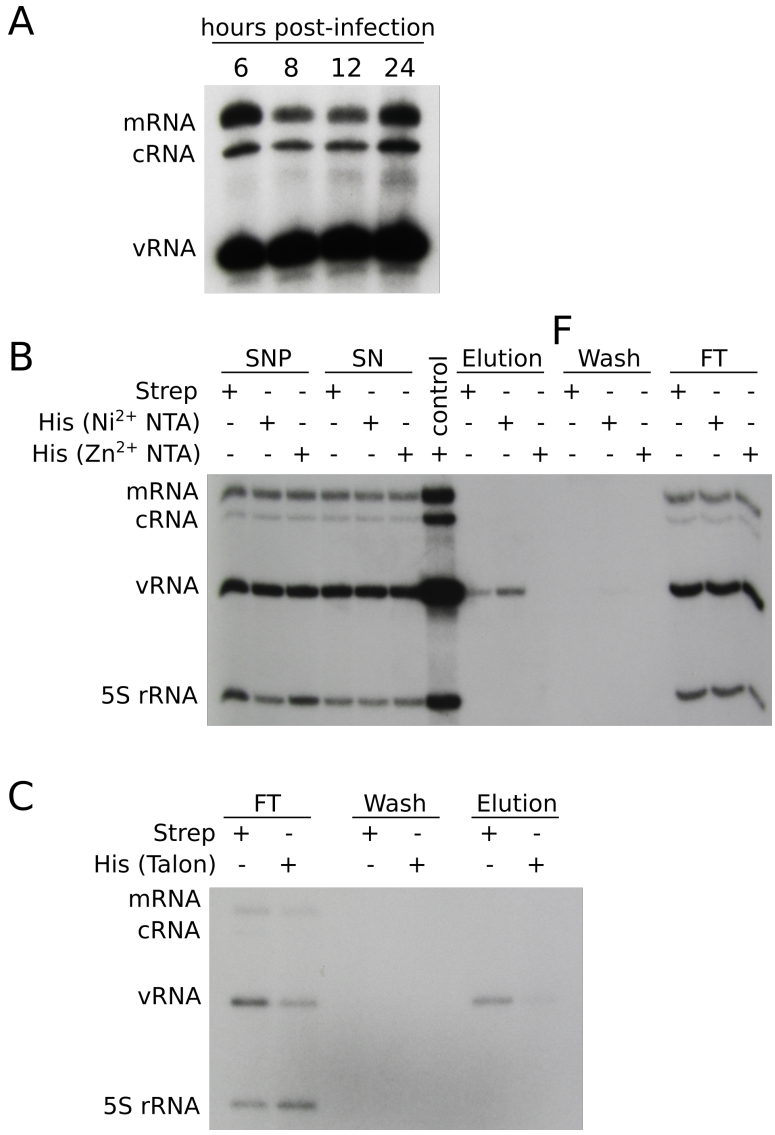


Figure 18 – RNP purification optimisation A) Time-course of influenza virus infection in HEK293 Freestyle cells. Primer extension of total cellular RNA. B) Comparison of purification efficiency of vRNPs with a His tag (on Ni^{2+} NTA and Zn^{2+} NTA) vs a Strep tag. Primer extension of different purification step samples. SNP: cell lysate, SN: soluble fraction of the cell lysate, FT: flow through. C) Comparison of purification efficiency of vRNPs with a His tag (on Talon resin) vs a Strep tag. Primer extension of different purification step samples. FT: flow through. Primer extensions carried out using NA-specific primers NA160 and NA1280 (see Fig. 22a).

This protocol produced large quantities of RNPs, however, the purification procedure still required modification. I initially considered whether it would be possible to remove the glycerol gradient step, as this resulted in high concentrations of glycerol in the final protein sample and its removal by buffer exchange could be detrimental for the RNPs. Purification on Ni^{2+}-NTA beads produced preparations with high levels of contaminating proteins so, alone, it would not suffice. I therefore tested purification on Zn^{2+}-NTA beads and Talon beads (Qiagen), both of which have better specificity for His tags than Ni^{2+} (but also lower binding capacities). In addition to that, I also tested the use of Strep-tagged PP7, with purification on StrepTactin beads (IBA). While yields were significantly lower when purifying using a Strep-tag, the purity was better than that achieved with a His-tag and Ni^{2+} NTA beads. Both Zn^{2+}-NTA and Talon beads failed to bind RNPs efficiently (Fig. 18b, c). These tests also highlighted the fact that most of the RNPs were being lost in the flow through regardless of the tag employed. Large quantities of PP7 eluted from the beads however (Fig. 19a), which suggested that maybe binding of PP7 CP to the hairpin RNA in cells was inefficient, perhaps because the hairpin is not always exposed on the NP-associated RNA.

Overall, the best balance of purity and yield was obtained with the Strep-tagged PP7 protein. It allowed me to avoid the use of glycerol gradients in the preparation of samples for cryo-EM. Although the purity was not optimal, the RNPs are sufficiently distinguishable with their long, helical shape, to be identifiable against the background in micrographs (Fig. 20a, b). vRNP contamination of cRNPs was also reduced due to the greater specificity of the tag, which were previously not removed as they migrate on a glycerol gradient in an identical manner (Fig. 19b). mRNA always co-purified to some extent with cRNPs, as they are both positive sense RNA species, but since it is sequestered by the cellular mRNA processing machinery, it should be morphologically very different from cRNPs. This protocol allowed me to purify greater quantities of RNPs which are in a buffer suitable for direct use for cryo-EM grid preparation. The glycerol gradient step was only employed on samples used for *in vitro* activity assays, as higher purity is desirable in these in order to remove any co-purifying cellular proteins that might confound interpretation of results. This is particularly important, as I was assessing the effect of cellular host factors on replication.

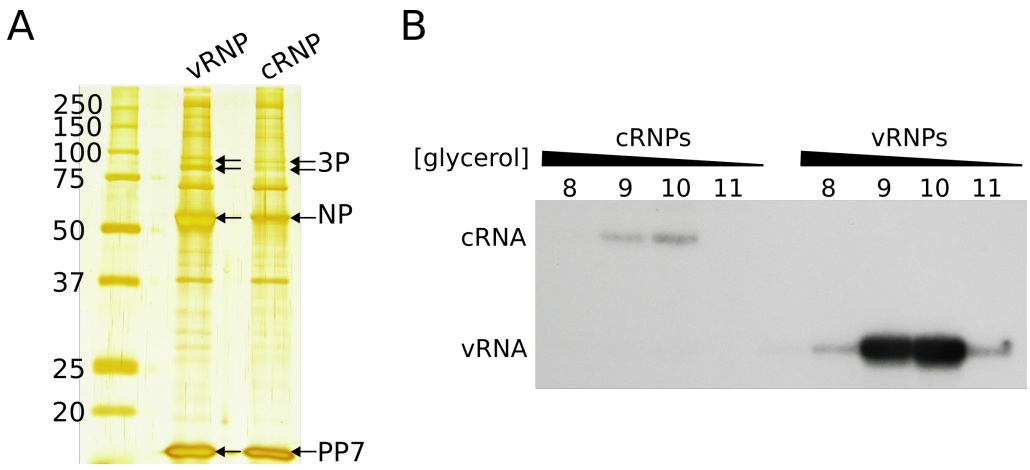


Figure 19 – Modified RNP purification A) Silver stained SDS-PA gel of purified and concentrated vRNPs and cRNPs after a single-step affinity purification on StrepTactin beads. Contaminating protein is present in the sample but NP and FluPol bands are visible, as is PP7. B) Primer extension of the RNA isolated from glycerol gradient fractions containing RNPs. These RNPs were isolated from infected HEK293F cells transfected with PP7-Strep. Primer extension carried out using NA-specific primers NA160 and NA1280 (see Fig. 22a).

3.3 Cryo-EM analysis of vRNPs and cRNPs

Being very large protein-RNA complexes, RNP structural characterisation was clearly best suited for cryo-electron microscopy. There is precedent in the literature for the study of RNPs by this method but they have only yielded low resolution models which have failed to provide detailed information on the protein-protein interfaces or the RNA path along the helix (Arranz et al., 2012; Coloma et al., 2009; Moeller et al., 2012). The relatively recent introduction of direct electron detectors to cryo-electron microscopy has resulted in a dramatic increase in the resolution that can be achieved for biological samples, as it allows data collection at very low electron doses with good signal to noise ratios. Previous studies on RNPs have employed CCD cameras for data collection, which may partially explain their limited resolution. Here I discuss the hurdles encountered in attempting to obtain high-resolution data on RNPs and the methods developed to tackle them. All the cryo-EM work involved in this project was carried out in collaboration with Juha Huiskonen, who was particularly involved in the development of the data processing strategy.

3.3.1 Cryo-EM sample preparation and data collection

RNPs bound to PP7-Strep were eluted from StrepTactin beads (IBA) in a buffer containing 100 mM Tris HCl pH 8, 150 mM NaCl, 1 mM EDTA and 2 mM biotin. The protein samples had to be concentrated to approximately 1-2 mg/ml in order to achieve a good particle distribution in vitreous ice. Grids were prepared on a Vitrobot mark IV (FEI) set to 80-100 % humidity. The sample was applied directly to glow discharged carbon coated (40 nm film) copper C-flat grids (Protochips) with 2 μm holes and 1.2 μm spacing. The grids were blotted for 3.5 seconds before plunging into liquid ethane. The RNPs were found to distribute preferentially in thicker ice close to the edges of the holes. Hence grids with a thicker carbon film (40 nm instead of 20 nm), which produced slightly thicker ice, achieved a better particle distribution. Data for vRNPs and cRNPs was collected on a 300 keV Polara microscope (FEI) fitted with a K2 direct electron detector (Gatan). A total dose of ~40 e^-/A^2 was used over 10 seconds, split over 25 frames. Motion correction was carried out using Motioncorr and CTF correction was carried out using CTFFIND3 (Li et al., 2013; Mindell and Grigorieff, 2003). Further data processing was carried out using Relion (Scheres, 2012).

3.3.2 Data processing and the "helix path-walker" approach

The RNP particles were easily identifiable in vitreous ice, which is not surprising given that RNPs are large mega-Dalton complexes which have a very characteristic morphology (Fig. 20a, b). A total of 2089 cRNP and 680 vRNP micrographs were collected, which resulted in 10073 cRNP and 11272 vRNP particles extracted. The low particle number obtained is partially a direct result of the size and shape of RNPs. Very few particles can be seen per micrograph, with an optimum of around 20 particles. Increasing sample concentration causes the RNPs to overlap with each other, which results in a large number of discarded particles. Achieving the optimum particle concentration of around 20 per image was also not trivial: after a single purification step, it was difficult to accurately determine the RNP concentration by measuring absorbance at 280 nm due to the presence of other contaminating proteins in the sample. Therefore, the final concentration of each protein preparation used to make grids for data collection had to be determined empirically through negative stain EM or cryo-EM. The concentration was particularly difficult to estimate for cRNPs, which produced significantly lower

yields (around 10-fold) than vRNPs. This often translated to low particle numbers on the grids: there was an average of 4.8 cRNPs per micrograph versus 16.6 for vRNPs.

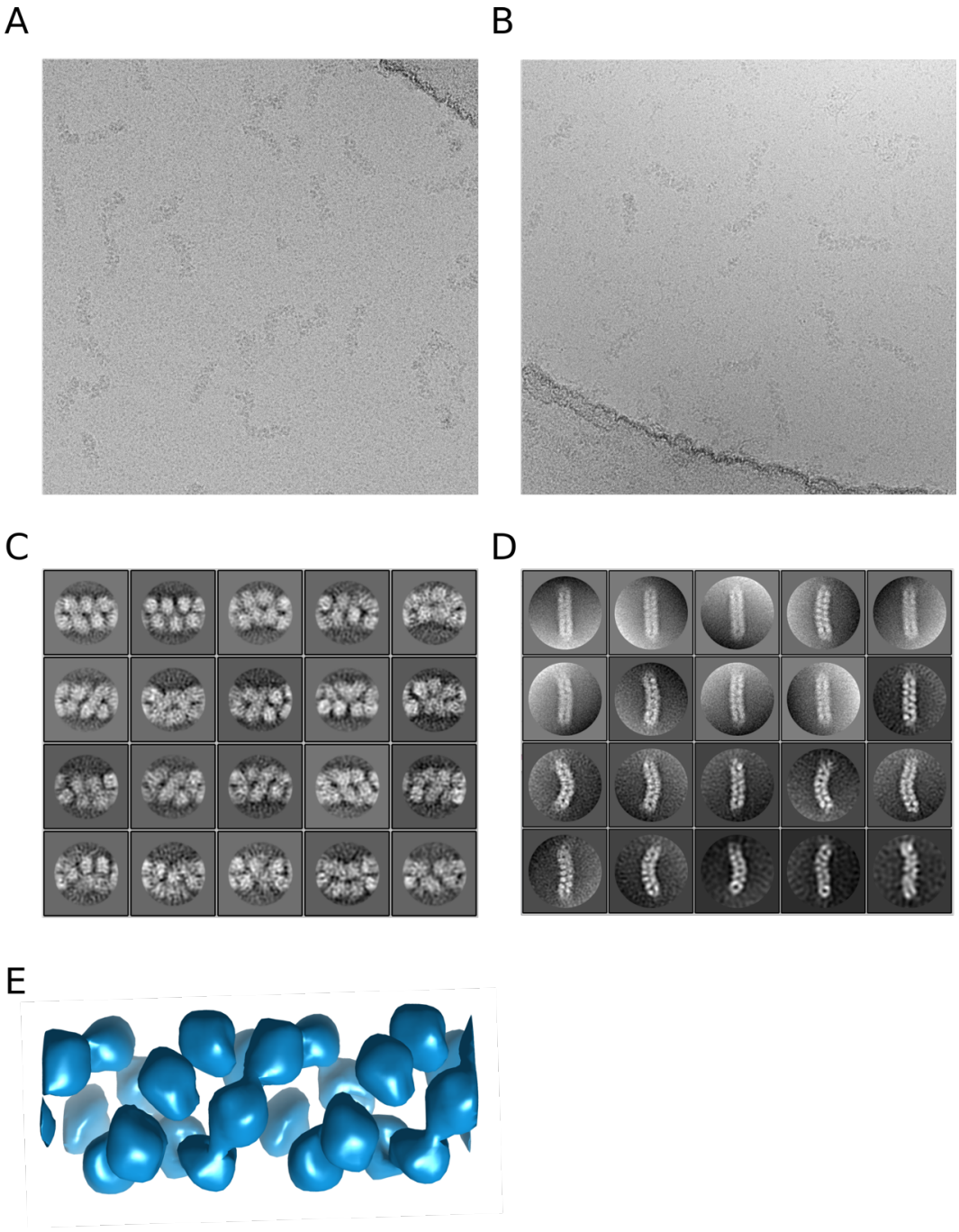


Figure 20 – Cryo-EM analysis of RNPs. A) Motion-corrected micrograph of vRNPs. B) Motion-corrected micrograph of cRNPs. C) 2D class averages of the cRNP helical filament. D) 2D class averages of full cRNPs. E) Refined density for the cRNP helical filament (unmasked).

Previous data processing strategies applied to RNPs involved dividing each RNP particle into three segments: the polymerase, the helix and the loop and classifying these separately (Arranz et al., 2012; Moeller et al., 2012; York et al., 2013). I carried out 2D classifications for the full-length cRNPs as

well as for their central helical region (Fig. 20c, d). The NP helix showed the expected morphology, although alignment was poor in some classes and the two NP strands could not be distinguished, likely a consequence of low particle numbers. The helical region was further processed and refined imposing the helical symmetry parameters described by Arranz et al. (Arranz et al., 2012), resulting in a 19 Å resolution map (without masking) (Fig. 20e). This is comparable to the resolution obtained for the helical regions in the published vRNP datasets (Arranz et al., 2012; Moeller et al., 2012). Predictably, the helix appears no different from that of vRNPs, and the resolution is unfortunately not high enough to be able to gain novel insights into the NP-NP interface or the location of the RNA. The full length RNPs presented similar issues in terms of alignment of the central helical region and the loop/polymerase ends were extremely poorly defined.

This processing protocol where the different RNP regions are picked separately is feasible for small numbers of particles but becomes challenging for the large datasets that will be required in order to achieve better resolutions. There is an additional level of complexity in processing single particle data of RNPs and that is particle heterogeneity. RNPs are extremely flexible structures that can bend and stretch along the helix (Fig. 21a) (Gallagher et al., 2016). The helix as well as the loop end also display variable degrees of twisting, accordingly producing bigger or smaller loop ends. Traditionally, bent RNPs have been discarded from datasets for processing as they result in poor alignments, indeed, most of the poorly aligned 2D classes we obtained contain bent particles (Fig. 21b). This is likely also at least partially responsible for the very poor definition of the ends when 2D classifying full length RNPs. However, given the low density of particles per micrograph, being able to keep as many particles as possible during processing would be beneficial to avoid making the datasets even smaller. It therefore became clear that a novel strategy would be required in order to use our available data in a more efficient manner.

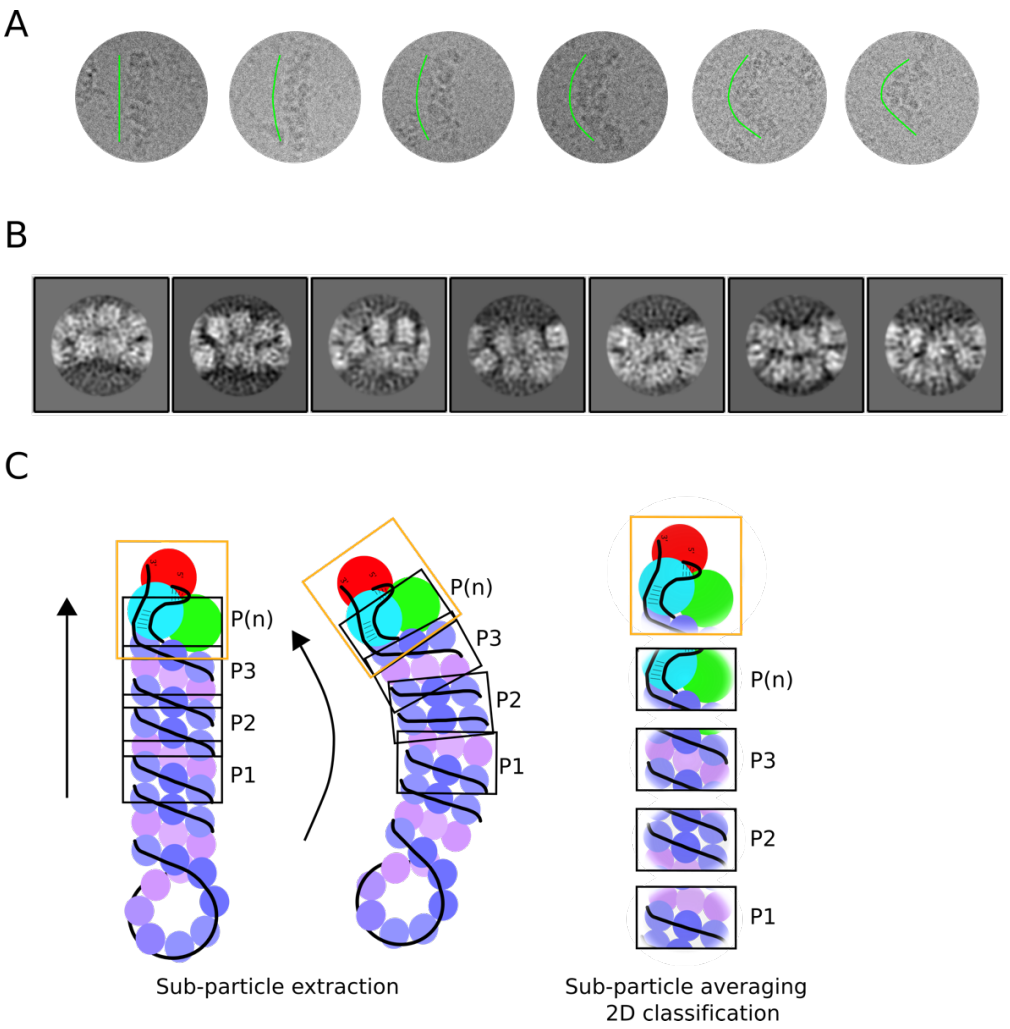


Figure 21 – RNP filament flexibility and particle heterogeneity. A) Individual cRNP particles exhibiting varying degrees of bending. B) 2D classes of the cRNP helical region showing deformed filaments. These classes were poorly aligned compared to classes containing straight filaments. C) The "helix path-walker" approach for sub-particle averaging.

This led us to developing what we later termed the "helix path-walker" approach (Fig. 21c). It is a variation on sub-particle averaging methods that have been employed to achieve better alignments of features within large multi-protein complexes (Ilca et al., 2015). In it, a short central section of the helix (particle 1 or P1) is initially extracted and used for 2D alignment. A new sub-particle (P2) is then produced by shifting the position of the previous box by a short distance towards the ends, so there is an overlap between P1 and P2. The angular search for the alignment of P2 can then be restricted as desired, as P2 is highly unlikely to be rotated over +/- 120º compared to P1. This is then repeated along the filament in both directions, effectively straightening out the RNPs *in silico*, until the 2D classes no longer show the helical filament. At this point, a new particle can be extracted from the filament ends, for which we would now have precise coordinates. The particles extracted from the

ends can then be averaged in another 2D classification where the loop and polymerase ends would fall into different classes and can be separated for further processing. This method would allow us to overcome the fact that most software adapted for helical refinement assumes a straight, rigid helix is present. The polymerase end, which would be challenging to pick manually with precision, would be found experimentally. This would give us the best chance to obtain a good alignment and therefore better resolution. As with the helical segments, the freedom of rotation of the polymerase during alignment could be limited to a certain range. In addition to that, as the sub-particles are still associated to their parent particle, any particles where both "end" sub-particles have been classified as a loop or as a polymerase could be discarded, as this could indicate they originated from overlapping RNPs.

3.3.3 Conclusions and future directions

Unfortunately, the "helix path-walker" method is still under development so I cannot present preliminary data here, as it has not yet been applied to the RNP datasets. Further work on this project is being carried out by Mattia Bertinelli and Serban Ilca. Nevertheless, it is clear that this project will require the acquisition of larger volumes of data due to the low particle numbers per micrograph. This will be particularly necessary in order to resolve the polymerase end, as it is asymmetric and only one copy is present per RNP. The polymerase is of course, the main point of interest for high resolution structural characterisation as it is likely to contain the key differences that would provide insight into replication initiation from a cRNA versus a vRNA template. Having helical symmetry, the NP filament should be easier to solve with lower RNP particle numbers, although their flexibility might still be a problem. Indeed the "helix path-walker" approach addresses the fact that the filament can bend, but will still encounter issues with stretching of the filament or uncoiling. Another problem to contend with is the fact that RNPs always orient themselves parallel to the air-liquid interface on cryo-EM grids. This means end-on views will be lacking for the polymerase, which may cause distortions of the density map, as full coverage of views of a particle is generally required for adequate 3D reconstruction. As RNPs are found preferentially on thicker ice, it is possible that they also bend perpendicularly to the ice surface, which could explain the presence of some RNPs which appear

shorter or where the ends are very poorly defined. This may be overcome through the use of a carbon support on the grid, whether graphene oxide or ultrathin carbon. Overall, while there is still some room for optimisation in terms of grid preparation, the key to obtain high resolution structures of cRNPs and vRNPs in order to compare their polymerase ends will probably involve the acquisition of more data as well as the development of novel methods for data processing.

3.4 *In vitro* analysis of cRNA to vRNA replication

Being able to obtain relatively pure preparations of cRNPs and vRNPs enables us to study their activity *in vitro*. There are two main advantages of using this method versus in-cell assays such as genetic trans-complementation assays described in the literature for investigating the effect of transacting FluPol. First is that, being a simpler assay with less components, we have greater control over the reaction conditions (Jorba et al., 2009). This therefore enables us to more accurately evaluate the requirements for transcription and replication. Second, as we can purify cRNPs and vRNPs separately with very little cross-contamination, we can compare their intrinsic activity and the effect of additional factors on cRNA to vRNA versus vRNA to cRNA replication.

For *in vitro* RNP activity assays, RNPs were purified as described above, including a glycerol gradient step, which left them in a buffer containing 50 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM $MgCl_2$ and ~50% glycerol. RNPs were stored in this buffer at -20ºC. RNPs were incubated at 30 ºC with rNTPs and ApG dinucleotide or β-globin mRNA for priming where required. ApG primed reactions mimic *de novo* initiation, except the addition of ApG overcomes the initial, rate-limiting step of dinucleotide formation. β-globin mRNA primed reactions enable us to analyse the cap-snatching activity of the RNP, as it uses the mRNA to generate short capped primers for transcription. The template RNA is already found within the RNP so no additional RNA is required. In order to analyse the products of the reaction, RNA is isolated by phenol-chloroform extraction and analysed by primer extension (reverse transcription) with NA-specific radio-labelled primers (Fig. 22a). The primers bind close to the ends of the RNA and extend towards the 5´end, producing short ssDNA fragments which can be analysed

by PAGE. This method allows the differentiation between cRNA and mRNA, as mRNA products are 10-13 residues longer due to the capped primer sequence obtained from the host. Depending on the source of capped RNA, this results in 2-3 bands corresponding to viral mRNA due to the variation in cleavage position during primer generation. These often migrate close to each other in PAGE giving the appearance of a single broad band.

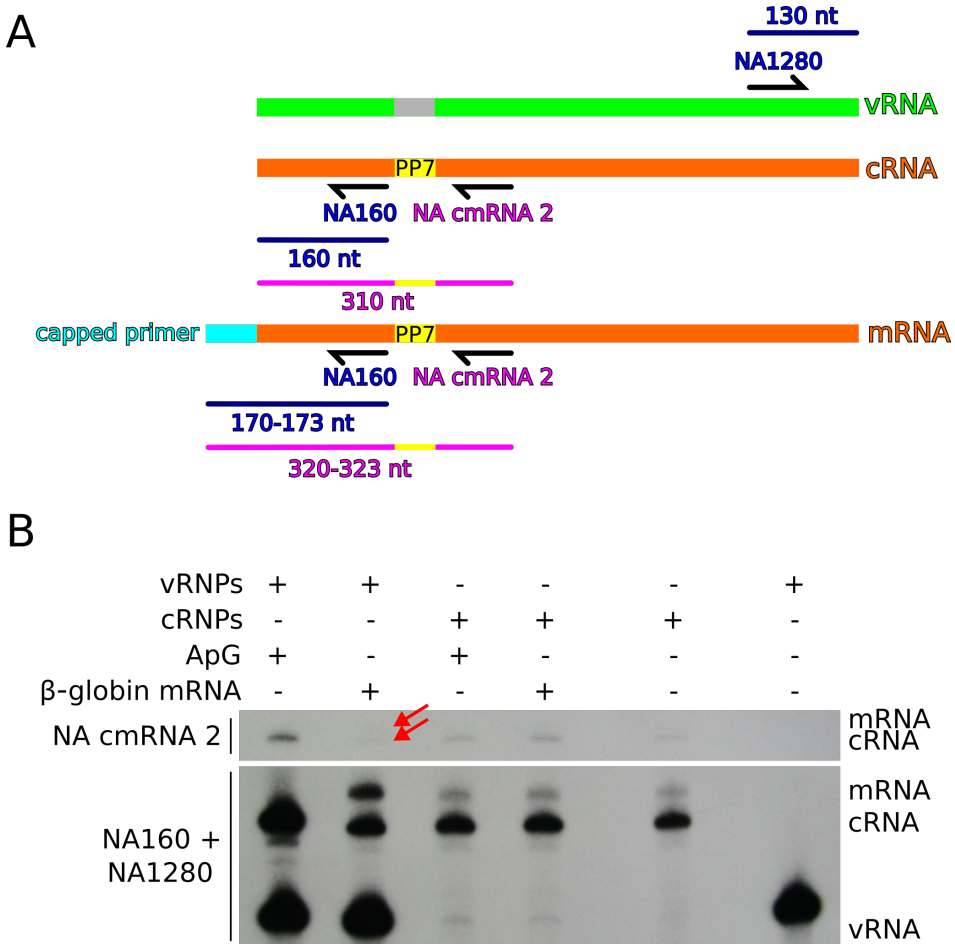


Figure 22 – RNP *in vitro* activity assay setup A) Schematic representation of primer extension assay using NA160, NA1280 and NA cmRNA 2 primers. NA1280 binds negative sense RNA, producing a 130 nt product. NA160 binds positive sense RNA producing a 160 nt product from cRNA and a 170-173 nt product from mRNA. The NA cmRNA 2 primer binds upstream of the PP7 hairpin tag, producing a 310 nt product from cRNA and a 320-323 nt product form mRNA. B) Primer extension to test the ability of FluPol to transcribe/replicate through the PP7 hairpin. NA cmRNA 2 primer binds upstream of the tag. The red arrows point towards the cRNA and mRNA bands (mRNA band particularly is very faint).

This protocol has been extensively used in the analysis of FluPol RNA products (Luo et al., 1992), however I initially had to verify that the presence of the PP7 tag did not interfere with polymerase elongation (Fig. 22b). As these RNPs are purified via the PP7 hairpin, which is in turn tightly bound by the PP7 coat protein, it was important to ensure that the polymerase was able to displace any

remaining PP7 protein as well as disrupt the hairpin in order to replicate or transcribe through it. As the primer extension primers bind near the 5′ ends of the RNAs, the reverse transcription products don't span the region of the PP7 tag. Therefore, the polymerase could potentially be stalling at the tag and producing only short abortive products, which would not be seen when using the NA160 and NA1280 primers (Fig. 22a). In order to test whether this was the case, I designed a primer that would bind positive sense RNAs (i.e. mRNA and cRNA) downstream of the PP7 hairpin and analysed the products of ApG and β-globin mRNA primed reactions. vRNPs displayed robust activity as shown by the products from the NA160 and NA1280 primers. ApG primed reactions produced a product that corresponded to the length expected for cRNA. In the reactions containing β-globin mRNA, vRNPs produced both mRNA and cRNA. The cRNA band in this case is produced via "true" *de novo* initiation (unaided by the ApG dinucleotide). In agreement with previous studies, cRNPs displayed no detectable activity whether with ApG or β-globin mRNA. Wherever cRNA was detected by the NA160 primer, a product, proportional in intensity, was detected by the NA cmRNA 2 primer. However, the products originating from the NA cmRNA 2 primer were significantly weaker, so much so that the bands corresponding to capped mRNA are barely visible. This is likely due to poor primer efficiency, particularly given that it primes reverse transcription of a longer product than NA160. On the other hand, this could indicate a mixture of abortive products and full-length products are generated due to the PP7 hairpin. Nevertheless, this assay indicated that the polymerase is capable of elongation past the PP7 tag, although further analysis of the products (by sequencing for instance) would be required if elongation efficiency past the PP7 tag was to be assessed.

3.4.1 RNP in vitro activity assays in presence of additional purified FluPol

I next confirmed whether I could reproduce the published observations that showed activation of replication by cRNPs when in presence of additional purified polymerase (York et al., 2013). In this assay, purified polymerase expressed in insect cells was included in RNP *in vitro* activity assays primed with ApG, β-globin mRNA as well as in unprimed reactions. I tested the effect of addition of a mammalian-adapted and an avian-adapted influenza virus polymerase: A/Northern Territory/60/1968 (H3N2) and A(duck)/Fujian/1/2002 (H5N1) respectively. As before, the RNA products from the

reaction were analysed by primer extension with NA160 and NA1280 primers (Fig. 23). Addition of free polymerase had the expected effect on cRNA to vRNA replication in ApG primed reactions (comparing lanes 2 and 4 in Fig. 23a and lanes 4 and 9 in Fig. 23b). Where there was no detectable product in absence of additional FluPol, its inclusion in the reaction led to the appearance of two distinct vRNA bands. The presence of two bands may be a result of a mixture of terminal and internal initiation during replication initiation as ApG can base pair at positions 1 and 2 or 4 and 5 on a cRNA template. No product was visible in the unprimed reactions, which suggests that initial dinucleotide formation on a cRNA template is inefficient in our *in vitro* assay. Nevertheless, the increase in cRNP activity in presence of polymerase supports the idea that free polymerase has a role in activating cRNA to vRNA replication. This phenotype was observed both with $FluPol_{NT60}$ (Fig. 23a) and $FluPol_{Fj}$ (Fig. 23b). This indicates that, while the RNP resident polymerase is of mammalian origin, *in vitro*, both mammalian-adapted and avian-adapted polymerases are compatible with activation of cRNA to vRNA replication. This supports the well-established idea that the restriction in activity observed in avian polymerases in mammalian cells is linked to incompatibilities with host proteins, not the polymerase itself.

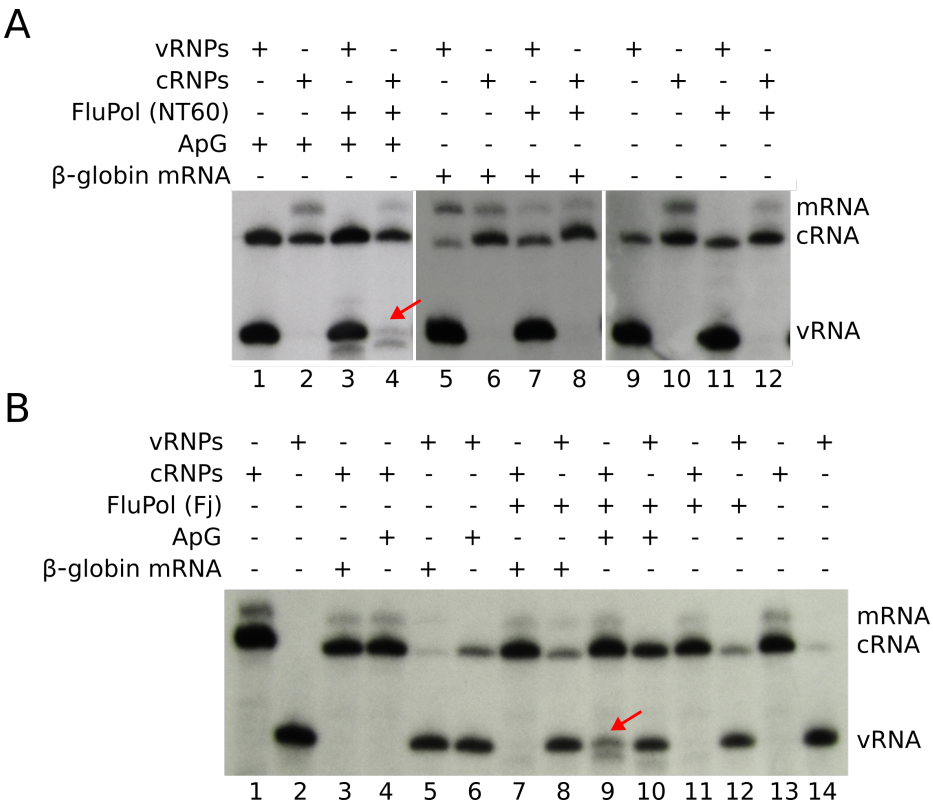


Figure 23 - RNP in vitro activity assays in presence of free polymerase. A) ApG and β-globin mRNA primed activity assay, testing the effect of addition of mammalian-adapted $FluPol_{NT60}$. Red arrow points towards the vRNA product produced by cRNPs in presence of additional polymerase. B) Like A) but testing the addition of avian-adapted $FluPol_{Fj}$.

Interestingly, I also observed a consistent increase in activity in vRNP containing reactions whenever additional FluPol was present. This was true for both ApG primed (comparing lanes 1 to 3 in Fig. 23a or lanes 6 and 10 in Fig. 23b) and unprimed reactions (comparing lanes 9 and 11 in Fig. 23a or lanes 14 and 12 in Fig. 23b). This was even visible in the cRNA band produced in reactions containing β-globin mRNA (lanes 5 and 7 in Fig. 23a or lanes 5 and 8 in Fig. 23b). There has been no prior evidence to suggest that free polymerase is required for vRNA to cRNA replication, or enhancement in activity by its addition. However, our data suggest that this may indeed be the case. This additional polymerase could simply be acting as a recipient for the product cRNA strand. This has been previously proposed in the literature based on cRNA stabilisation assays and genetic trans-complementation assays discussed in the introduction to this chapter (Jorba et al., 2009; Vreede et al., 2004). These show that free polymerase is capable of binding and stabilising replication products, which is independent of its catalytic activity. Binding of free FluPol to cRNA produced by the vRNP could result in an overall increase in activity due to improved product release from the replicating RNPs. Free RNA in the reaction may also hinder replication if it causes displacement of the resident RNA on the RNP or hybridises with other RNAs forming secondary structures that the polymerase struggles to replicate through (RNA bases are exposed in RNP-associated RNA as shown by the fact that FISH probes can by hybridised to it (Lakdawala et al., 2014)). However, we must also consider the possibility that the transactivating polymerase model also has a role in replication by vRNPs. This would mean that in addition to stabilising cRNA produced by vRNPs, newly produced polymerase in an infected cell might also interact with the resident vRNP polymerase to promote replication. This would add another layer of regulation to the switch from transcription, which occurs early in infection, to replication later in infection that is dependent on the concentration of free FluPol in the cell.

No significant effect was observed on transcription upon FluPol addition in reactions containing vRNPs (lanes 5 and 7 in Fig. 23a or lanes 5 and 8 in Fig. 23b). This is consistent with literature suggesting that transcription occurs exclusively *in cis* and doesn't require any additional viral factors (Jorba et al., 2009).

3.4.2 Nucleoprotein

In an infected cell, NP is essential for the replication of full length RNA segments, as it acts as a processivity factor (Turrell et al., 2013) (discussed under section 1.2). However, additional roles have been proposed for this protein (reviewed in Portela and Digard, 2002), among others, in the switch from transcription to replication. It has been proposed that nucleoprotein expression is required for replication in cells. Several temperature-sensitive mutants of NP exhibited decreased ability to support replication at the non-permissive temperature (Medcalf et al., 1999; Shapiro and Krug, 1988). Prior to production of NP in infection, vRNPs would therefore only be capable of template transcription. There have been suggestions that the role of NP in regulating the switch from transcription to replication is dependent on its interaction with polymerase (Biswas et al., 1998; Mena et al., 1999).

I decided to test whether addition of nucleoprotein to RNP activity assays would improve processivity *in vitro* or whether it would have a more profound effect on the regulation of transcription and replication. NP was expressed in Sf9 insect cells and purified via a protein A tag on IgG beads, which was followed by size exclusion chromatography (Fig. 24a, b). NP forms oligomers, particularly at the lower salt concentrations used during gel filtration (150 mM NaCl). Accordingly, multiple peaks were observed on the chromatogram. Based on their elution profile in gel filtration (column calibrated using a GE Healthcare Gel Filtration HMW Calibration Kit), the main peaks appeared to correspond to monomers, trimers (which were the largest population), pentamers and another larger oligomer that could correspond to either heptamers or octamers. These peaks were pooled together for use in *in vitro* RNP activity assays.

The effect of NP on *in vitro* activity was tested both in presence and absence of additional polymerase (Fig. 24c). Surprisingly, NP had a detrimental effect on transcription, both in presence and absence of additional FluPol (comparing lanes 5 and 6 to 7 and 8). Indeed, no mRNA band was visible in vRNP reactions primed with β-globin mRNA where NP was also present. It is possible that, as NP binds RNAs non-specifically it sequesters the β-globin mRNA in the reaction, making it unavailable for cap-

snatching by the polymerase. On the other hand, this observation could imply and active role for NP in regulating the switch from transcription to replication. More specifically, rather than promoting replication, NP could be directly inhibiting transcription by vRNPs. This effect is independent of the presence of free polymerase which would suggest NP acts directly upon the resident RNP polymerase. While the RNP polymerase would be already bound to NP coating the vRNA, FluPol is likely to have additional NP binding sites near the product exit channel, as this needs to be coated with NP during replication. It is possible therefore that as newly produced NP accumulates in the cell during infection, it binds vRNP polymerase and prevents it from carrying out further transcription. Additionally, it would also then be readily available to coat new replication products.

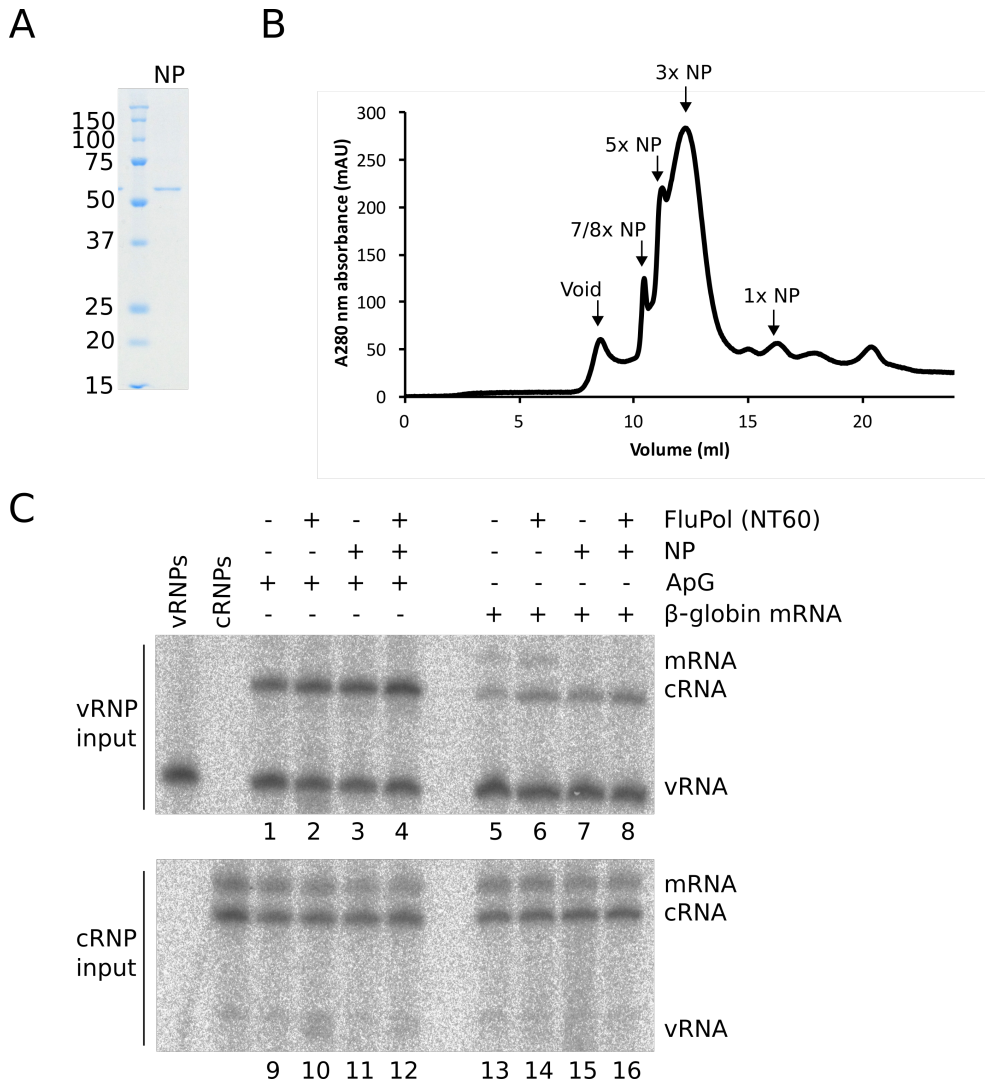


Figure 24 – Effect of NP addition in *in vitro* RNP activity assays. A) Simply Blue SafeStain stained SDS PA gel of purified NP. B) Gel filtration profile of NP showing multiple peaks of oligomeric NP. C) ApG and β-globin mRNA primed RNP activity in presence of purified NP and additional $FluPol_{NT60}$.

ApG primed replication levels remained mostly unchanged, both for cRNPs and vRNPs. Overall, addition of NP did not appear to have an effect on overall processivity. While this role is undoubtedly crucial in cells, as shown by the lack of transcription and replication of long templates in absence of NP (Turrell et al., 2013), it has proven to be difficult to analyse *in vitro*. Oligomerisation of NP in cells is a tightly regulated process, dependent on phosphorylation, which is not trivial to replicate using purified components in an *in vitro* setting (Mondal et al., 2015; Turrell et al., 2015).

3.4.3 ANP32

ANP32 protein family members were recently identified as a key factor in influenza host adaptation (introduced in section 1.3.3). In addition to that, a separate study described a role for mammalian ANP32A and ANP32B in replication by FluPol (Sugiyama et al., 2015). ANP32A and ANP32B were shown to interact with free FluPol specifically i.e. FluPol that was not RNA bound, and promote cRNA to vRNA replication *in vitro*. Knockdown of these two proteins in cells resulted in reduced viral growth, indicating they are essential during the viral life cycle. ANP32 family members have also been previously identified in proteomic analyses of polymerase interactors (Bradel-Tretheway et al., 2011; Watanabe et al., 2014). How these findings are linked to the role of ANP32 protein family members in host adaptation however remains unknown. The data presented in Long *et al*., 2016 showed that ANP32 proteins promoted all polymerase activity, increasing levels of vRNA, mRNA and cRNA. However, it is worth noting that they assay RNA levels in a whole cell system where levels of different viral RNA species are inter-dependent. Their data also supports the idea of mammalian ANP32A and ANP32B promoting activity of a mammalian-adapted FluPol, which would be in agreement with the data presented by Sugiyama et al., which was carried out on mammalian-adapted polymerase. The dependency of avian FluPol on avian ANP32A however was shown to be much greater than that of mammalian-adapted polymerase. Avian ANP32A contains a 33 amino acid insert in the LCAR which was essential for avian-adapted FluPol activity, and isoforms lacking it were unable to support it.

The exact mechanism which drives this increase in polymerase activity (whether specifically in cRNA to vRNA replication or otherwise) is unknown. Previous studies have all been carried out in cells or utilising partially purified protein preparations which means we also currently have no data on whether other proteins are involved in the interaction between FluPol and ANP32 proteins. I therefore sought to further characterise the effect of addition of ANP32 proteins in RNP *in vitro* activity assays. Three ANP32 variants were cloned into a pGEX-4Ti vector for expression in *E. coli*: human ANP32A, human ANP32B and chicken ANP32A (which contains the 33 amino acid insertion in the LCAR). All three isoforms expressed well in this system. The proteins were N-terminally tagged with GST and so were purified on glutathione sepharose beads, yielding relatively pure preparations for use in *in vitro* assays (Fig. 25a). hANP32A, hANP32B and chANP32A purified from mammalian cells were also tested, which were obtained from Benjamin Nilsson.

The *in vitro* RNP activity assay was performed both in presence and absence of additional free $FluPol_{NT60}$ as hANP32A and hANP32B were reported to interact only with free polymerase, which would suggest they are unable to interact with the RNP-resident polymerase. Only ApG primed reactions were tested in this case for simplicity, particularly as strong evidence was provided in the literature that only replication and not transcription was affected by addition of hANP32A and hANP32B (Sugiyama et al., 2015). Surprisingly, no difference in replication rates was observed in presence of any of the ANP32 isoforms whether from a bacterial or a mammalian source (comparing odd numbered lanes which do not contain additional $FluPol_{NT60}$ in Fig. 25b). Addition of $FluPol_{NT60}$ (even numbered lanes) produced the expected increase in replication, both in the cRNA to vRNA sense and vice versa. However, addition of ANP32 proteins in these also had no reproducible effect on replication. Upon seeing no clear effect with addition of bacterially-purified ANP32 proteins, we hypothesised that *E. coli* expression may be unsuitable for these proteins, perhaps resulting in misfolding or a lack of crucial post-translational modifications which would not occur in a prokaryotic expression system. Hence addition of mammalian expressed ANP32 was tested but this also produced no effect on replication rates. It is worth noting that the concentration of ANP32 purified from mammalian cells was much lower than that of the bacterially expressed preparations. However, the

fact that no difference at all was observed suggests that a key factor in this interaction (like other proteins which may bridge FluPol and ANP32 proteins) is likely missing in this assay. This is certainly possible as there is no precedent for studying this interaction using highly purified components. Alternatively, the enhancement in replicative activity may be dependent on a cellular context. The "blockade" that it overcomes in an infected cell may be missing entirely in an *in vitro* system, which would prevent us from seeing an effect.

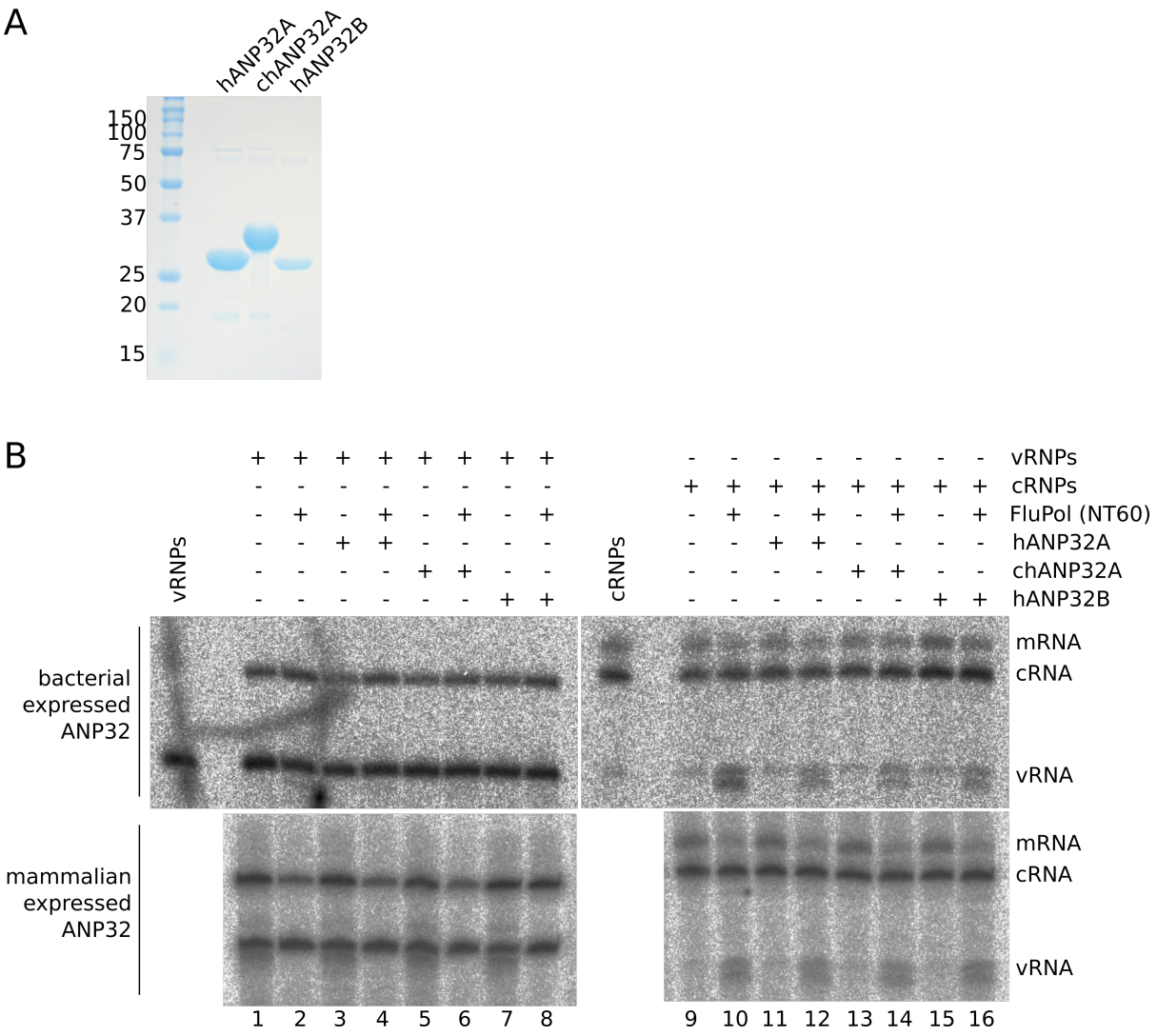


Figure 25 – Effect of ANP32 addition to RNP *in vitro* activity assays. A) Simply Blue SafeStain stained SDS PA gel of purified ANP32 human A and B isoforms as well as chicken A isoform. B) ApG and β-globin mRNA primed RNP activity in presence of purified hANP32A, chANP32A and hANP23B and additional mammalian FluPol.

Very recently, an additional report on the FluPol-ANP32 interaction became available (Domingues and Hale, 2017). It found that the enhancement in polymerase activity caused by ANP32 proteins was

dependent on SUMOylation. The effect in avian-adapted FluPol activity caused by chANP32A was dependent on the C-terminal LCAR (residues 221-235) but also on a SUMO interaction motif (SIM) found in the 33 amino acid insertion specific to ANP32A from avian cells. Presence of this SIM also improved interaction with mammalian-adapted polymerases (containing a lysine at position 627), but it appears as though these are less reliant on ANP32 proteins for polymerase activity (although they do have an effect), while avian-adapted polymerases require a strong and stable interaction. However, co-expression of SUMO proteases in a mini-replicon assay also had a deleterious effect on mammalian-adapted FluPol activity, which suggests SUMOylation has a role in the interaction beyond the SIM in the 33 amino acid insertion. The binding target for this SIM was not identified in this study. They therefore propose a model where SUMOylation of another host factor is essential for bridging chANP32A and FluPol, although this could extend to hANP32 isoforms too. This would be in agreement with our data and support both the idea that there may be additional proteins missing which bridge FluPol and ANP32 proteins, and the fact that post-translational modifications are important for the interaction.

3.4.4 Conclusions and future directions

Being able to assess RNP activity *in vitro* has allowed us to further analyse the factors that regulate and promote replication by FluPol. We have uncovered a role for additional FluPol in stimulating not only cRNA to vRNA replication but also replication by vRNPs (Fig. 23). In addition to that, the inhibitory effect of nucleoprotein on transcription indicates that this protein too is involved in regulation of the switch from transcription to replication (Fig. 24). Together, these data suggest that, during infection, the incoming vRNPs carry out transcription until new polymerase and nucleoprotein become available in the nucleus. At this point, the presence of NP, which could potentially be interacting with the RNP polymerase, would discourage transcription from taking place. Free FluPol would then also interact with the vRNP-resident polymerase and stimulate it to replicate *in cis*. This would be in agreement with observations, which we confirm in our assays, that suggest that vRNPs are capable of replication without additional polymerase (Vreede and Brownlee, 2007). At this point, both NP and FluPol would be available in the proximity of the vRNP to coat the cRNA product strand.

While free polymerase appears to have a moderate stimulatory role on vRNPs, our data confirms it is essential to trigger cRNA to vRNA replication. Use of FluPol containing a PB1 active site mutation shed light on the mechanism by which replication by cRNPs was stimulated by free polymerase, showing that catalytic activity was unnecessary and therefore proposing a "transactivation" mechanism (York et al., 2013). It would now be interesting to test whether the increase in vRNP activity was also observed in assays including an inactive polymerase, as this would confirm that the same mechanism is in place in this step of replication too. Of course, it is possible that both transacting and transactivating mechanisms are at play, with one polymerase interacting with the RNP resident polymerase and activating replication, while another polymerase acts as a recipient for the product RNA strand. While the inclusion of ANP32 family proteins in our RNP in vitro assays did not provide any information on their effect on replication, the negative result is indicative of a more complicated interaction model than had been proposed in the literature (Long et al., 2016; Sugiyama et al., 2015). It now appears likely that interaction between FluPol and ANP32 proteins is mediated by additional host factors. In addition to that, SUMOylation (and perhaps other post-translational modifications) also seem to be involved in the interaction, which could even provide a means for regulation of replication initiation. Further work will be required to determine whether this is the case, including a careful analysis of existing proteomics data which could provide candidates for additional host factors involved in the interaction.

3.5 Summary

Useful steps have been taken towards being able to achieve high resolution data on full RNP structures, including their isolation and preparation for analysis by cryo-EM. With the development of improved data processing pipelines and greater volumes of data, a detailed structural comparison of vRNPs and cRNPs should soon be attainable. This would hopefully provide insight into the mechanism behind the differing replication initiation strategies of vRNPs and cRNPs and how variations at a few nucleotide positions in the promoter result in drastically different initiation activity. The *in vitro* RNP acitivty assay data has allowed us to gain more insight into the influence of free polymerase, nucleoprotein and ANP32 proteins in the regulation of the distinct switch between

transcription and replication that occurs during infection. It suggests a model where both polymerase and nucleoprotein availability are capable of modulating RNP activity. It also points towards a more complex system for ANP32 stimulation of FluPol activity, involving post-translational modifications and perhaps other host factors, than was previously proposed.

4 Regulation of transcription in the influenza C virus polymerase

4.1 Introduction

Influenza C viruses generally only cause mild upper respiratory disease in humans and therefore receive less attention than influenza A and B viruses, which have a significant yearly toll on human health. Like influenza A and B however, influenza C viruses encode a trimeric RNA dependent RNA polymerase, and therefore potentially share the same mechanisms for transcription and replication. Our laboratory has found that expression and purification of influenza C polymerase in insect cells produces large yields of stable protein, making it ideal to study the mechanistic details of polymerase activity of influenza viruses.

At the time when this work was performed, no structural data was available which showed polymerase during transcription initiation (Thierry et al., 2016). The crystal structures of vRNA promoter bound FluPol consistently acquired a conformation that appeared to be compatible with cap-snatching based of the distance between the cap binding and endonuclease domains but whether this was indeed the domain arrangement found in the polymerase during transcription initiation was unknown (Pflug et al., 2014; Reich et al., 2014). This was therefore proposed to be the "pre-initiation" conformation of the polymerase, which will also be referred to as the "open" polymerase conformation throughout this chapter. In contrast, the apo FluPol$_C$ structure, which was solved by Naring Hengrung in our laboratory soon after the beginning of my DPhil (Hengrung et al., 2015), showed the polymerase in a "closed" conformation not compatible with transcription, (a comparison of the different polymerase conformations can be found under section 1.3.5). Interestingly, SAXS data showed that, in solution, the RNA-free FluPol$_C$ trimer was also capable of acquiring the more "open" pre-initiation conformation and that the polymerase conformation was also influenced by buffer composition. This suggests there is a very low energetic threshold for the exchange between the two conformations. Nevertheless, the presence of promoter vRNA appears to have a role in driving the polymerase to

acquire the transcription-competent conformation. On the other hand, given that we had no data on the structure of FluPol$_C$ when bound to RNA, one could claim that the differences in conformation observed between vRNA bound FluPol$_{A/B}$ and the apo FluPol$_C$ are simply due to the different influenza types used and not relevant biologically. We also lack a mechanism for how promoter binding at a site close to the PB1 β-ribbon affects the positioning of the PB2 C-terminal domains, which are located on the opposite side of the polymerase.

In order for FluPol to cap-snatch, it requires a reliable source of host capped RNAs (Beaton and Krug, 1981; Plotch et al., 1981). The polymerase appears to ensure this is the case by associating with host DNA dependent RNA polymerase II (introduced in section 1.4.1). The interaction occurs via the Pol II C-terminal domain (CTD), specifically when the heptapeptide repeat that forms the CTD is serine-5 phosphorylated (Engelhardt et al., 2005; Martínez-Alonso et al., 2016). The structures of FluPol$_A$ and FluPol$_B$ bound to an S_5P CTD peptide mimic show two distinct binding sites in each, of which only one is conserved in both polymerases (Lukarska et al., 2016). These structures highlight how divergent CTD binding sites can be across influenza types. Comparison of the binding site amino acid sequences suggests none of these binding sites is likely to be conserved in FluPol$_C$ (Fig. 26). Reduced access to Pol II in the later stages of infection in a cell, due to Pol II degradation caused by FluPol, is thought to contribute to the switch from transcription to replication (Vreede et al., 2010). However, despite the availability of structural information on the interaction with CTD, there has been no data that addressed whether there is an active involvement of Pol II CTD binding in regulating transcription.

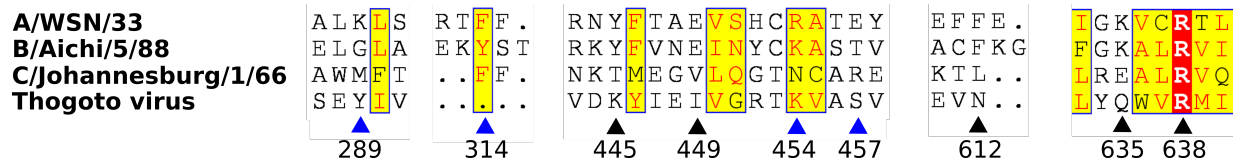


Figure 26 – Amino acid sequence alignment of the CTD binding sites on the PA C-terminal domain identified in FluPol$_A$. Arrows in black highlight key residues in binding site 1 and arrows in blue highlight key residues in binding site 2. Numbering is as for influenza A/WNS/33 (H1N1).

The main aim of this project was therefore to take advantage of the ease of expression and purification of FluPol$_C$ to study the mechanisms involved in transcription. More specifically, I attempted to

structurally characterise the effects of promoter vRNA binding on polymerase conformation both by crystallography and cryo-electron microscopy. In addition to that, ApG dinucleotide and S_5P CTD peptide were also employed in crystallisation trials to try to capture states of the polymerase relevant to the process of transcription initiation. *In vitro* polymerase activity assays were also employed to test the effect of binding to Pol II CTD in cap-snatching and extension from a capped primed.

4.2 Conservation of cap-snatching in influenza C virus polymerase

While studying FluPol$_C$ would offer a great opportunity to investigate the processes of transcription and replication in influenza viruses, the conservation of these has never been assessed. Indeed, while other members of the Orthomyxoviridae encode related heterotrimeric polymerases, the cap-snatching mechanism is not conserved in them. Thogotoviruses, Orthomyxoviruses related to influenza viruses that include Thogoto, Dhori and Jos viruses, appear to have lost the ability to cap-snatch. Despite highly conserved overall cap binding and endonuclease domain folds, the key active site residues have mutated causing a loss of function. Accordingly, the endonuclease domain was found to be unable to bind divalent cations and the cap binding domain had no detectable affinity for cap structures (Guilligay et al., 2014). This is particularly surprising given that Thogotoviruses were thought to acquire cap structures from the cell like influenza viruses, except with cleavage occuring immediately after the first transcribed nucleotide (m^7GpppA_m) (Albo et al., 1996). For this reason, I initially assessed the conservation of cap-snatching in FluPol$_C$.

Alignment of the polymerase subunit amino acid sequences reveals that, overall, sequence conservation between influenza viruses is relatively low, with sequence identity values with FluPol$_A$ around 25 % for PB2 and P3 (Fig. 27 and Table 5, 6). On the other hand, the key amino acid residues in the active sites of the subunits responsible for cap-snatching are identical in influenza C and influenza B. In order to experimentally confirm that these were functional, I generated constructs for expression of the polymerase trimer in HEK293T cells with mutations to alanine in the PB2 and P3 active site residues. These included the aromatic sandwich residues W369 and F416 in PB2 and the

metal coordinating residues E65, D93, E104 and K134 of P3. I addressed their capacity for transcription and replication in an RNP reconstitution assay (Fig. 28). In this assay, cells are transfected with plasmids to express all the necessary components to form functional RNPs, that is, the three polymerase subunits in addition to nucleoprotein and, in this case, a CAT template flanked by C/Johannesburg/66 virus NS segment non-coding regions. The original plasmids for expression of influenza C ribonucleoprotein components for RNP reconstitution were kindly donated by Dr Nadia Naffakh (Crescenzo-Chaigne et al., 1999). The cells were harvested 30 hours post-transfection and the RNA isolated by phenol-chloroform extraction. The RNA was analysed by primer extension, where the RNA products are reverse transcribed using radiolabelled primers that bind m/cRNA or vRNA produced from the CAT template.

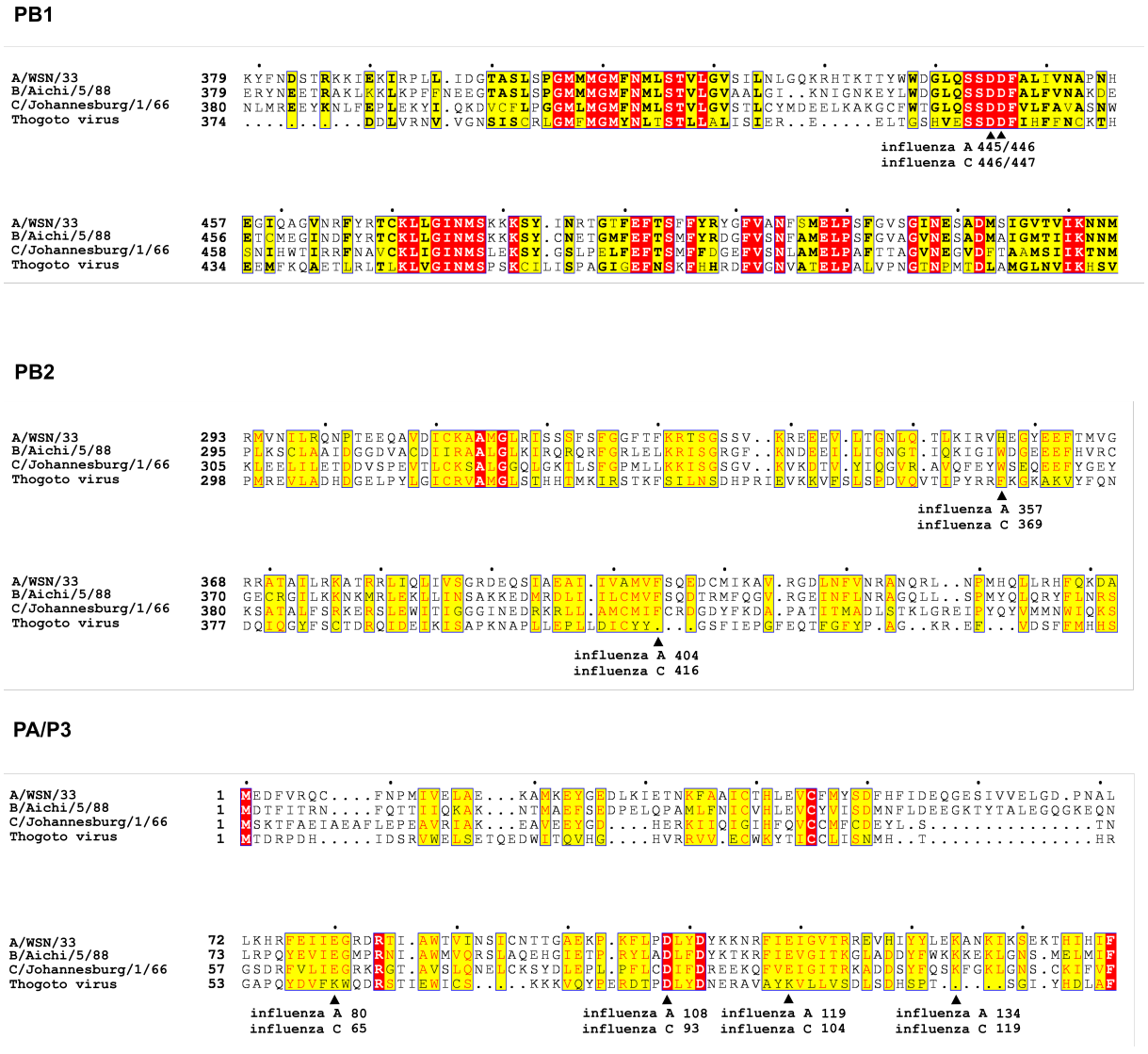


Figure 27 – Amino acid sequence alignment of influenza A virus, influenza B virus, influenza C virus and Thogoto virus polymerase subunit active sites. Highly conserved regions are shown in red. Active site residues are annotated.

	A/WSN/33 (H1N1)	B/Aichi/5/88	C/Johannesburg/1/66	Thogoto virus
A/WSN/33 (H1N1)	100.00	38.04	23.62	17.00
B/Aichi/5/88	38.04	100.00	24.70	16.06
C/Johannesburg/1/66	23.62	24.70	100.00	16.76
Thogoto virus	17.00	16.06	16.76	100.00

Table 7 - PB2 amino acid identity matrix (%)

	A/WSN/33 (H1N1)	B/Aichi/5/88	C/Johannesburg/1/66	Thogoto virus
A/WSN/33	100.00	37.66	26.53	20.27
B/Aichi/5/88	37.66	100.00	26.40	20.47
C/Johannesburg/1/66	26.53	26.40	100.00	18.83
Thogoto virus	20.27	20.47	18.83	100.00

Table 8 - PA amino acid identity matrix (%)

As expected, all the cap binding and endonuclease domain mutations resulted in abolished transcription, suggesting the cap-snatching mechanism is conserved. The PB1 active site mutant was inactive both in transcription and replication. In addition to a defect in transcription, PB2 F416A was also defective in replication. The equivalent mutation, F404A, in influenza A virus polymerase has been found to inhibit transcription specifically but leaves replication unaffected (Fechter et al., 2003). This result could hence be a consequence of misfolding, low expression levels or poor trimer assembly. Alternatively, this residue could have an additional role in polymerase activity in influenza C that is not found in influenza A. Unfortunately, due to a lack of antibodies against FluPol$_C$ I was unable to verify this. Despite this, the fact that the residues tested are conserved in influenza C, in conjunction with the RNP reconstitution assay data, suggest that the cap-snatching mechanism is indeed conserved throughout the influenza viruses, making FluPol$_C$ a good system in which to study transcription. The apo FluPol$_C$ structure which became available after this work was completed confirmed the conservation of the cap binding and endonuclease domains, which were essentially identical in fold to those found in FluPol$_A$ and FluPol$_B$ (Hengrung et al., 2015).

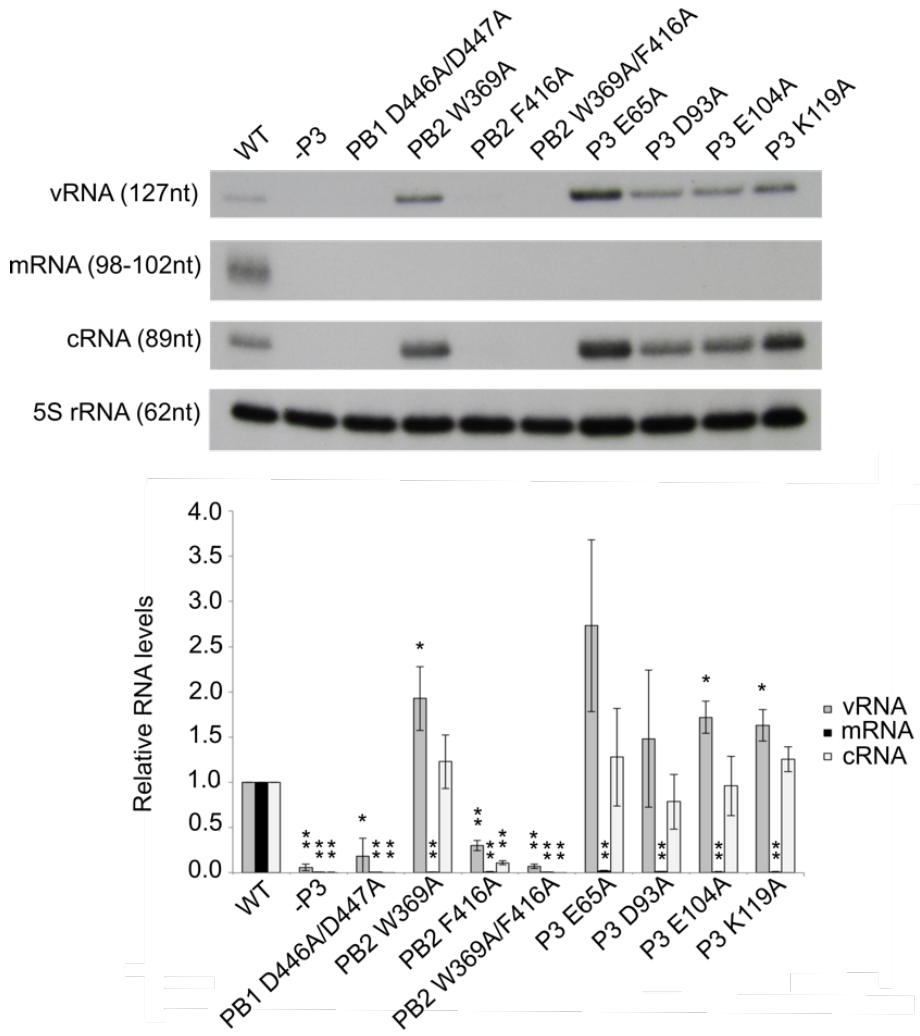


Figure 28 – RNP reconstitution assay to determine the conservation of cap-snatching in FluPol$_C$. Total RNA was extracted from transfected cells and analysed by primer extension using CAT 2, CAT89 and 5S 62nt radiolabelled primers. Quantitation of signal is shown: the mean and standard deviation of three experiments is plotted. Asterisks indicate a significant difference from wild type (WT), which was set to 100% (*P <0.05; **P <0.01, based on a two-sample t-test).

4.3 Purification of FluPol$_C$

Expression and purification of influenza polymerase heterotrimer has been historically challenging. Other laboratories have produced FluPol in insect cells as a self-cleaving polyprotein to achieve expression in equimolar amounts and adequate assembly (Pflug et al., 2014). In our case, FluPol$_C$ expression was carried out in insect cells infected with a baculovirus for co-expression of the three polymerase subunits of influenza C/Johannesburg/66 virus (this baculovirus was designed and initially produced by Dr Narin Hengrung). It includes the codon-optimised sequences of the three polymerase subunits for optimal expression in Sf9 (Spodoptera frugiperda 9) insect cells. The heterotrimer was tagged via the C terminus of the PB2 subunit with a protein A tag separated from PB2 by a TEV

cleavage site (Fig. 29a). For expression, 1 or 2 litres of Sf9 insect cells were infected at an MOI ≈ 1 with the baculovirus and harvested 72 hours later. The complex is initially purified on IgG coated sepharose beads, which have a high specificity for protein A and therefore yield relatively pure preparations after cleavage from the beads with TEV protease. When analysed by SDS-PAGE, these always show apparently equimolar amounts of the three subunits and no degradation (Fig. 29b). The heterotrimer ran as a single peak when analysed by size exclusion chromatography, with a small void volume peak (Fig. 29c).

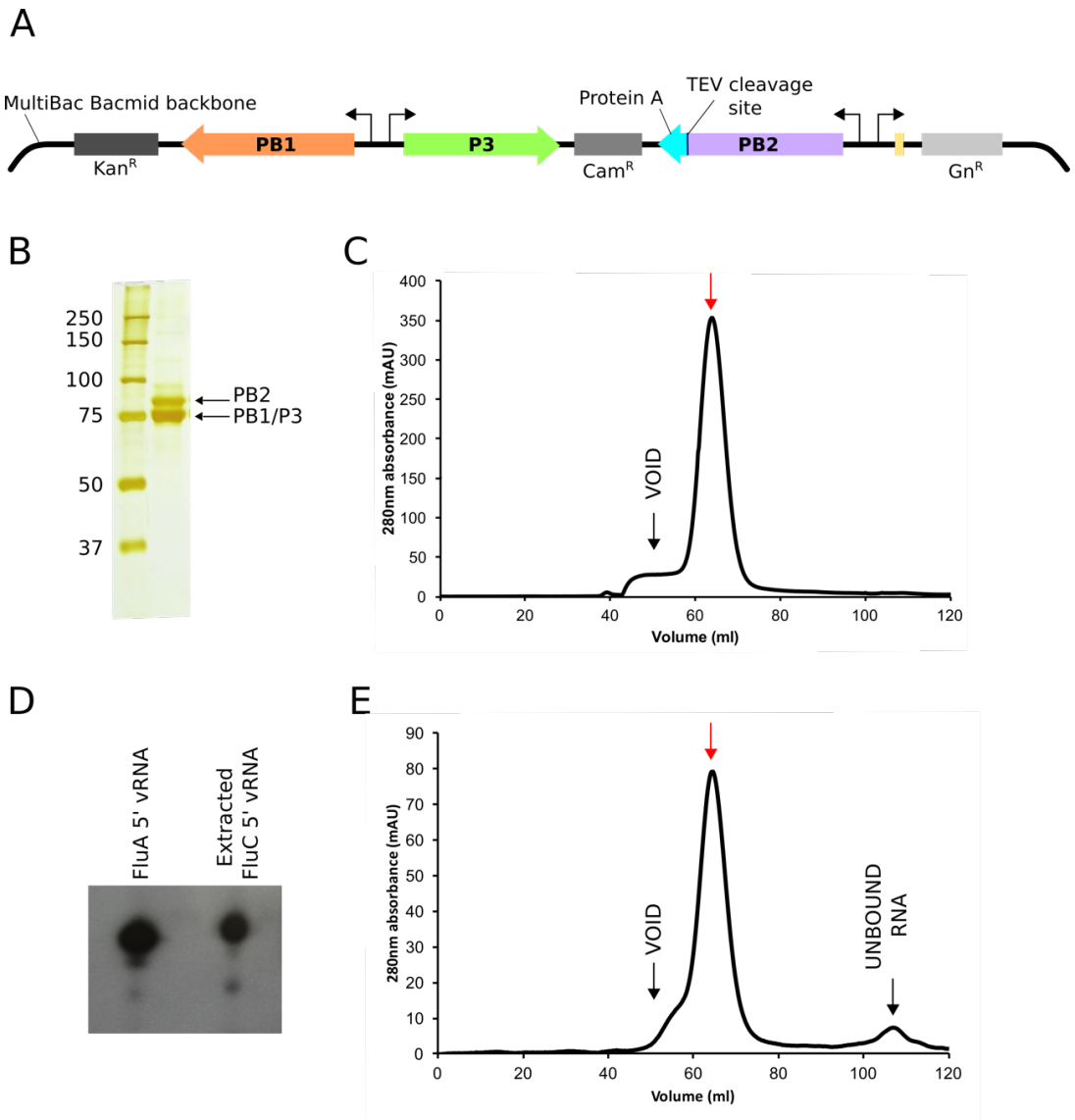


Figure 29 – FluPol$_C$ purification. A) Diagrammatic representation of the baculovirus genome engineered for expression of the three polymerase subunits. PB2 has a C-terminal protein A tag separated by a TEV cleavage site. B) Silver stained SDS-PA gel showing the polymerase after affinity purification. C) Gel filtration profile of the apo FluPol$_C$. D) Radiolabelled RNA isolated by phenol-chloroform extraction from FluPol$_C$ bound to 5′ vRNA (after being run through a size exclusion column to eliminate excess unbound RNA). It is run beside a radiolabelled FluPol$_A$ 5′ vRNA marker. The apparent size difference (they are both 15 nt long) is due to the differing nucleotide composition, which affects migration on PAGE. E) Gel filtration profile of FluPol$_C$ after incubation with 5′ and 3′ promoter vRNA.

Where an RNA-free sample was required, RNase A was added at the lysis step as we found the polymerase tends to co-purify with cellular RNAs. This was avoided where promoter RNA was to be added later to avoid possible RNase cross-contamination. I initially tested binding of RNA under high salt conditions. After purification, the polymerase is left in 500 mM NaCl (high ionic strength is required for FluPol to remain in solution). The NaCl concentration was then adjusted to 2 M and the RNA added in >2 molar excess. The NaCl was then dialysed out to slowly bring the concentration of NaCl back down to 500 mM. The idea behind this was that increasing the ionic strength of the solution would disrupt any existing interactions that could shield RNA binding sites. Then as the NaCl concentration was lowered during dialysis in presence of RNA, new interactions with the RNA would be allowed to form, hopefully resulting in a protein-RNA complex. This was later found to be unnecessary, as RNA appeared to bind when simply added to the purified polymerase. While addition of RNA caused some precipitation of the sample, enough protein remained in solution to be further purified through gel filtration. The gel filtration profile shows a trimer peak identical to that of the apo protein, but a small peak appears at higher volumes which has a high absorbance at 260 nm (as measured by NanoDrop), likely containing the excess unbound RNA (Fig. 29e). The presence of polymerase-bound RNA was confirmed by extracting RNA from the purified sample and 5´ end radiolabelling with polynucleotide kinase. The labelled RNA was analysed by PAGE, which revealed the expected 15mer (or 15mer and 14mer RNA bands where these were used), indicating co-migration in gel filtration of polymerase and RNA and therefore a stable interaction between the two (Fig. 29d).

4.4 Structural analysis of FluPol$_C$

Structural characterisation of FluPol has been mostly carried out by crystallography. Indeed, it was this method which allowed the determination of the apo FluPol$_C$ structure (Hengrung et al., 2015). Nevertheless, reproducibility of good quality, well diffracting crystals has been challenging in our laboratory. There has been some success utilising cryo-electron microscopy for structural analysis of a PB2 truncated form of the FluPol heterotrimer, although it is hard to assess how well this translates to the full length protein (Chang et al., 2015). Below I describe my attempts at analysing the structure of

FluPol$_C$ bound to factors which may allow us to gain insight into the mechanisms involved in transcription. These included ApG dinucleotide, promoter vRNA and S_5P CTD peptide.

The rationale for including these in our structural studies is as follows: through base pairing, ApG could stabilise the 3' promoter strand in the PB1 active site, allowing us to analyse an initiating form of the polymerase. Based on the available crystal structures, promoter RNA binding appears to have drastic effects on the organisation of the PB2 C-terminal domains. However, SAXS data on the apo FluPol$_C$ polymerase suggest this conformational switch can occur even in absence of RNA (Hengrung et al., 2015). Additionally, there has been no direct comparison of RNA bound and unbound FluPol of the same type, as the RNA bound structures were all carried out with FluPol$_{A/B}$, while the only available apo structure is for FluPol$_C$ (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). Obtaining a vRNA-bound FluPol$_C$ structure would therefore enable us to determine whether promoter binding truly stabilises the pre-initiation conformation or whether the conformational differences observed are caused by the use of polymerases from different influenza types. In addition to that, in the published RNA bound structures, the RNA plays an essential role in making crystal contacts, therefore there was a chance its inclusion in crystallisation trials would aid us in achieving higher resolution FluPol$_C$ data (Pflug et al., 2014; Reich et al., 2014). As introduced at the beginning of this chapter, Pol II CTD interacts directly with FluPol and is key during transcription to provide a source of capped RNAs for cap-snatching. However, its interaction with FluPol$_C$ has not yet been characterised and amino acid sequence alignment suggests that the binding sites described in FluPol$_{A/B}$ are not conserved throughout influenza types (Lukarska et al., 2016). I therefore also included a S_5P CTD peptide spanning four heptapeptide repeats in the crystallisation trials (peptide details can be found in section 2.1.4).

4.4.1 Crystallisation trials of FluPol$_C$ bound to RNA and CTD

For crystallisation trials, FluPol$_C$ was purified as described above, but magnesium chloride and manganese chloride were added to the final concentrated protein at 10 mM and 1 mM respectively. These were included as FluPol is known to bind magnesium, both at the active site in PB1 and in the

endonuclease domain which, as discussed in section 1.3.2, can also bind manganese. TCEP was also added to a final concentration of 0.5 mM to prevent aggregation. The interactors (ApG, promoter vRNA or S_5P CTD) were added in a 1.2 molar excess alone or in combination, immediately prior to setting up the crystallisation screens. These were set up in a sitting-drop vapour diffusion format at room temperature (Walter et al., 2005) and the choice of conditions tested was guided by the previous success of Narin Hengrung in crystallising the apo protein . Crystals were tested for diffraction at the Diamond Light Source (Didcot, Oxfordshire, UK) on beamlines I03 or I04. Table 7 gives details of some of the conditions screened which produced crystals, which are discussed further below.

My first target was to increase the reproducibility of generating high quality crystals that would diffract to high resolution, as this had previously been extremely challenging. Crystals appeared consistently in the Morpheus protein crystallisation screen (Molecular Dimensions), more specifically, in conditions containing 10 % w/v PEG 8000 and 20 % v/v ethylene glycol (Fig. 30a). The fastest crystals to appear did so 3 days after setting up the trials, while the slower ones took around 13 days. From that point, they took 4 to 6 days to be fully grown. The crystals produced were occasionally needle shaped (for example condition Morpheus H6) but mostly they appeared as stacked plates. These crystals were quite fragile during handling and unfortunately never diffracted beyond ~6 Å. Further optimisation was attempted through variation of the mother liquor concentration as well as inclusion of additives (Hampton silver bullet additive screen) and seeding. The seeds were produced from crystals grown in the Morpheus G2 condition or in Hampton Crystal Screen Cryo 18 (170 mM ammonium sulfate, sodium cacodylate pH 6.5, 25.5 % w/v PEG 8000, 15 % v/v glycerol). Seeding successfully increased the growth of crystals in Morpheus G2 which appeared less than 24 hours after setting up the drops and ranged in morphology from needles to short thick rods (Fig. 30b). These short rods were sturdier and less layered in appearance but their diffraction was not improved. I also tested whether crystal formation in larger, manually set drops yielded better formed crystals. Some of these included 0.2 % low melting point agarose in the drop to attempt to suspend the crystals in the liquid, as we previously found they had a tendency to sink to the bottom of the drop, which made them difficult to mount due to their fragility. While this method yielded large numbers of crystals, they were

no different in morphology to those appearing in the small droplet screens set up robotically. On the other hand, it does offer a good method for production of large numbers of crystals where a good crystallisation condition is found.

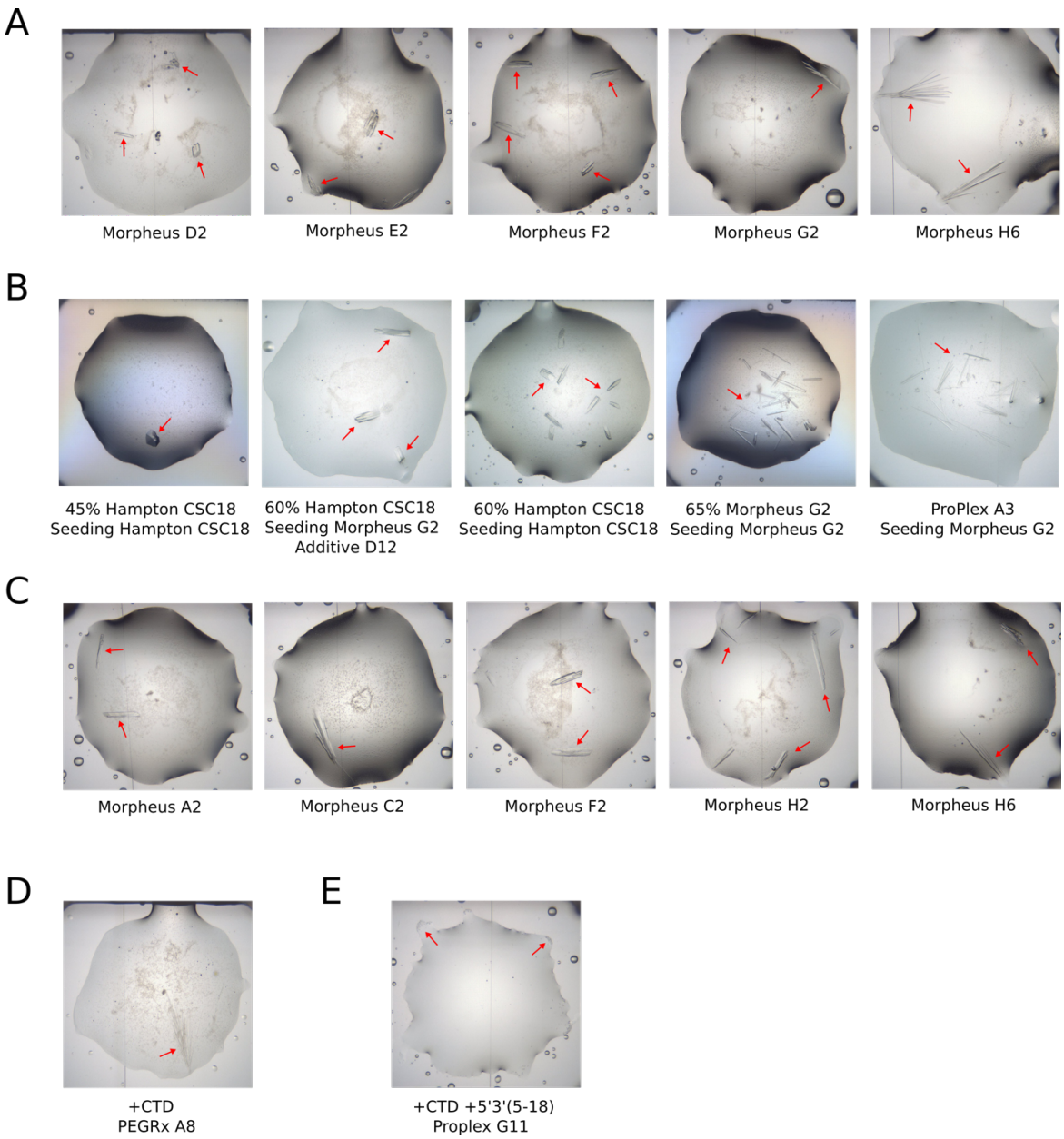


Figure 30 – FluPol$_C$ crystallisation trials and optimisation. A) Crystal drops of apo FluPol$_C$ in the Morpheus crystallisation screen. B) Optimisation of crystals through seeding. C) Crystal drops of FluPol$_C$ in presence of S_5P CTD. The crystals appeared in conditions identical to those of the apo protein. D) Additional hit of co-crystallised FluPol$_C$ with S_5P CTD peptide showing the formation of needles. E) Drop containing micro crystals of FluPol$_C$ in presence of S_5P CTD and 5′ (1-15) and 3′ (5-18) promoter vRNA.

Addition of S_5P CTD peptide for co-crystallisation did not affect which conditions the crystals were formed in (Fig. 30c). They still grew reproducibly in the Morpheus crystal screen with identical morphology to those grown in absence of the peptide. Yet again these crystals did not diffract beyond

6 Å and it was therefore impossible to determine whether the CTD peptide was present. The fact that the crystals grew in the same conditions could indicate that the CTD peptide was not interacting with the protein under those conditions or that addition of CTD did not interfere with the crystal contacts made by the apo polymerase. This second option is also feasible given that diffracting crystals in the Morpheus G2 condition were found to be up to 76 % solvent (Hengrung et al., 2015). Only one other hit appeared in the co-crystallisation of FluPol$_C$ and CTD peptide, in a PegRX screen (Hampton research) condition containing 100 mM MES monohydrate pH 6 and 22 % v/v PEG 400. These were extremely fine needles and could not be reproduced (Fig. 30d).

Addition of promoter RNA completely disrupted the formation of crystals in the conditions in which apo polymerase crystallised. Indeed, it appeared to inhibit crystal formation entirely as no crystals grew in any of the screens tested which included Block 1, Block 2, Block 3, Block 4 and Block 5 matrix screens (Oxford Protein Production Facility) as well as Morpheus (Molecular Dimensions), PEGRx (Hampton Research), ProPlex (Molecular Dimensions), Nuc-Pro (Jena Bioscience), PGA (Hampton Research) and SaltRX (Hampton Research) crystallisation screens among others. I tested two different 3´ promoter RNA lengths covering nucleotides 1-18 or 5-18. These are longer than the 3´ RNAs employed for other assays, which only span nucleotides 1-14, and were used in the hope that we could mimic the RNA-RNA crystal contacts observed in the FluPol$_A$ and FluPol$_B$ crystal structures but this was clearly not successful (Pflug et al., 2014; Reich et al., 2014). The only potential crystals that appeared were a small scatter of micro-crystals at the edge of a drop which also contained CTD peptide in a ProPlex screen condition containing 100 mM Na Hepes pH 7.5 and 1 M sodium acetate (Fig. 30e). They appeared within 24 hours and grew for only 2 days. Despite attempting optimisation these crystals never increased in quantity or size and were difficult to reproduce. I also attempted the co-crystallisation of FluPol$_C$ with promoter RNA and ApG but yet again no crystals appeared in any of the conditions tested. This suggests that addition of RNA is causing a significant change in the surface exposed regions of the polymerase involved in the crystal contacts of the apo protein, which is possibly indicative of a conformational change.

Screen	Protein	Interactors	Crystal morphology	Other
65% Morpheus G2 or 45-100% Morpheus H2	4.7mg/ml 25mM Hepes pH7.5 500mM NaCl 10% glycerol 0.5mM TCEP	-	Layered plates or needles	Seeding with Morpheus G2 crystals P:R:S = 2:1:0.5
60% Hampton CSC18	4.7mg/ml 25mM Hepes pH7.5 500mM NaCl 10% glycerol 0.5mM TCEP	-	Layered plates or needles – some crystals appeared less layered with certain additives e.g. E3 (1% w/v tryptone, 20mM Hepes Na pH 6.8)	Hampton Silver Bullet additive screen Seeding with Morpheus G2 crystals P:R:S:A = 2:1:1:1
48-70% Hampton CSC18	4.7mg/ml 25mM Hepes pH7.5 500mM NaCl 10% glycerol 0.5mM TCEP	-	Layered plates. Some short rods at higher concentrations of CSC18.	Seeding with Hampton CSC18 crystals P:R:S = 2:1:0.5
ProPlex	4.7mg/ml 25mM Hepes pH7.5 500mM NaCl 10% glycerol 0.5mM TCEP	-	Needles	Seeding with Morpheus G2 crystals P:R:S = 2:1:1 Note: these crystals did not appear when seeding with Hampton CSC18 crystals
65% Morpheus H2	4.75mg/ml 25mM Hepes pH7.5 500mM NaCl 10% glycerol 0.5mM TCEP	-	Layered plates	1% 1M NaOH added to reservoir 0.2% agarose in protein droplet Seeding with Morpheus G2 crystals P:R:S = 2:1:0.5 Manually set up
Morpheus	5mg/ml 20mM Hepes pH7.5 500mM NaCl 0.5mM TCEP 10mM $MgCl_2$ 1mM $MgCl_2$	-	Layered plates	P:R = 2:1
Morpheus	5mg/ml 20mM Hepes pH7.5 500mM NaCl 0.5mM TCEP 10mM $MgCl_2$ 1mM $MgCl_2$	S_5P CTD peptide	Layered plates and needles	P:R = 2:1
PEGRx	5mg/ml 20mM Hepes pH7.5 500mM NaCl 0.5mM TCEP 10mM $MgCl_2$ 1mM $MgCl_2$	S_5P CTD peptide	Needles	P:R = 2:1
ProPlex	5mg/ml 20mM Hepes pH7.5 500mM NaCl 0.5mM TCEP 10mM $MgCl_2$ 1mM $MgCl_2$	S_5P CTD peptide, 5´ promoter vRNA, 3´ promoter vRNA (5-18)	Micro-crystals	P:R = 2:1

Table 9 – Details of some of the conditions of FluPol$_C$ crystallisation hits.

It therefore became clear that studying the promoter RNA-bound form of FluPol$_C$ by crystallography would not be trivial. On the other hand, the introduction of direct electron detectors and improved data processing pipelines in cryo-electron microscopy made this method an attractive alternative, which led me to employ it in the study of this complex.

4.4.2 Cryo-EM sample preparation and data collection

Cryo-electron microscopy offers several advantages over crystallography in the study of small protein complexes. Within these is the fact that the protein is imaged in solution, rather than in a crystal lattice, which means there is no need for crystallisation of the sample and the interpretation of the structure is not confounded by the possible influence of crystal contacts, which may stabilise particular conformations or interactions. In addition to that, it is now possible to classify conformationally heterogeneous complexes into separate groups during data processing as long as the different conformations are well defined, so multiple states of the complex can be explored within a single dataset. This may be particularly beneficial in this case as the polymerase is known to be able to acquire multiple conformations. The quantity of sample required is also significantly lower than that for crystallisation which allows lower yield complexes to be studied.

On the other hand, cryo-EM presents its own challenges in terms of sample preparation, particularly for smaller protein complexes. As the protein is suspended in vitreous ice for imaging, the composition of the buffer that the protein is in plays a significant role. In order to achieve good contrast and for small particles to be visible, the buffer must essentially be as close to water as is feasible, as buffer components such as salts, glycerol or detergents can scatter electrons themselves and produce “grainy” looking ice or affect surface tension and generate inadequate ice thickness. This becomes an issue when protein complex solubility is reliant on high ionic strength or the presence of detergents. As well as that, flexible regions of the protein, similarly to crystallography, cannot be resolved if they are highly dynamic.

FluPol$_C$ purification leaves the complex in a buffer containing 500 mM NaCl, which, not surprisingly, did not generate good amorphous ice. For cryo-EM sample preparation therefore, I modified the purification procedure to drop the NaCl concentration. The samples were run through size exclusion chromatography in a buffer containing 400 mM NaCl rather than 500 mM, which was the lowest concentration that would maintain the complex in solution. The protein was then concentrated to ~4mg/ml and diluted 20-40 fold in a buffer containing 25 mM Tris pH 8.5, 150 mM NaSCN and 0.5 mM TCEP. This buffer modification resulted in some aggregate formation but a large proportion of the protein remained in solution and could be imaged, as the particles were visible against the background (Fig. 31a). Promoter 3´ and 5´ vRNA (14mer and 15mer respectively) were added after purification, before grid plunging, at a 1.2 molar excess. Grids were made using a Vitrobot mark IV (FEI) at 80-100 % humidity. 3 µl protein at 0.5 mg/ml were placed on carbon coated (40 nm film) copper C-flat grids (Protochips) with 2 µm holes and 1.2 µm spacing before blotting for 3 seconds and plunging into liquid ethane.

A small dataset of 1822 micrographs was collected on a 300 keV Polara microscope (FEI) fitted with a K2 direct electron detector (Gatan). The total dose used was 46 $e^-/Å^2$ over 22 frames, with a total exposure of 8.8 seconds. The frames were aligned using MotionCor2 and CTF correction was carried out with GCTF (Zhang, 2016; Zheng et al., 2016). Further data processing was carried out on Relion (Scheres, 2012). A total of 48648 particles were extracted from which 15789 remained after 2D classification. Initial attempts at 3D reconstruction revealed the complex acquired certain preferential orientations. As this can result in warped maps influenced by the low-resolution reference model used for 3D classification and refinement, I manually balanced out the particle numbers in the classes, to achieve a more even coverage of views. For this, the particle sorting function in Relion was used to select the particles with the best Z scores (indicating a good match between the particle and the 2D class template) in each preferential view. I grouped the 2D classes based on whether they were "preferential orientation 1", "preferential orientation 2" or "other views" (Fig. 31b). I proceeded to discard the excess of particles from the two groups that contained particles in preferential orientations, which resulted in a final dataset with 9159 particles (Fig. 32c). As this is a very small dataset the

resolution of the density map only reached 9.8 Å after refinement and masking. Nevertheless, this was enough to fit in the available structures in the open and closed conformations into the density.

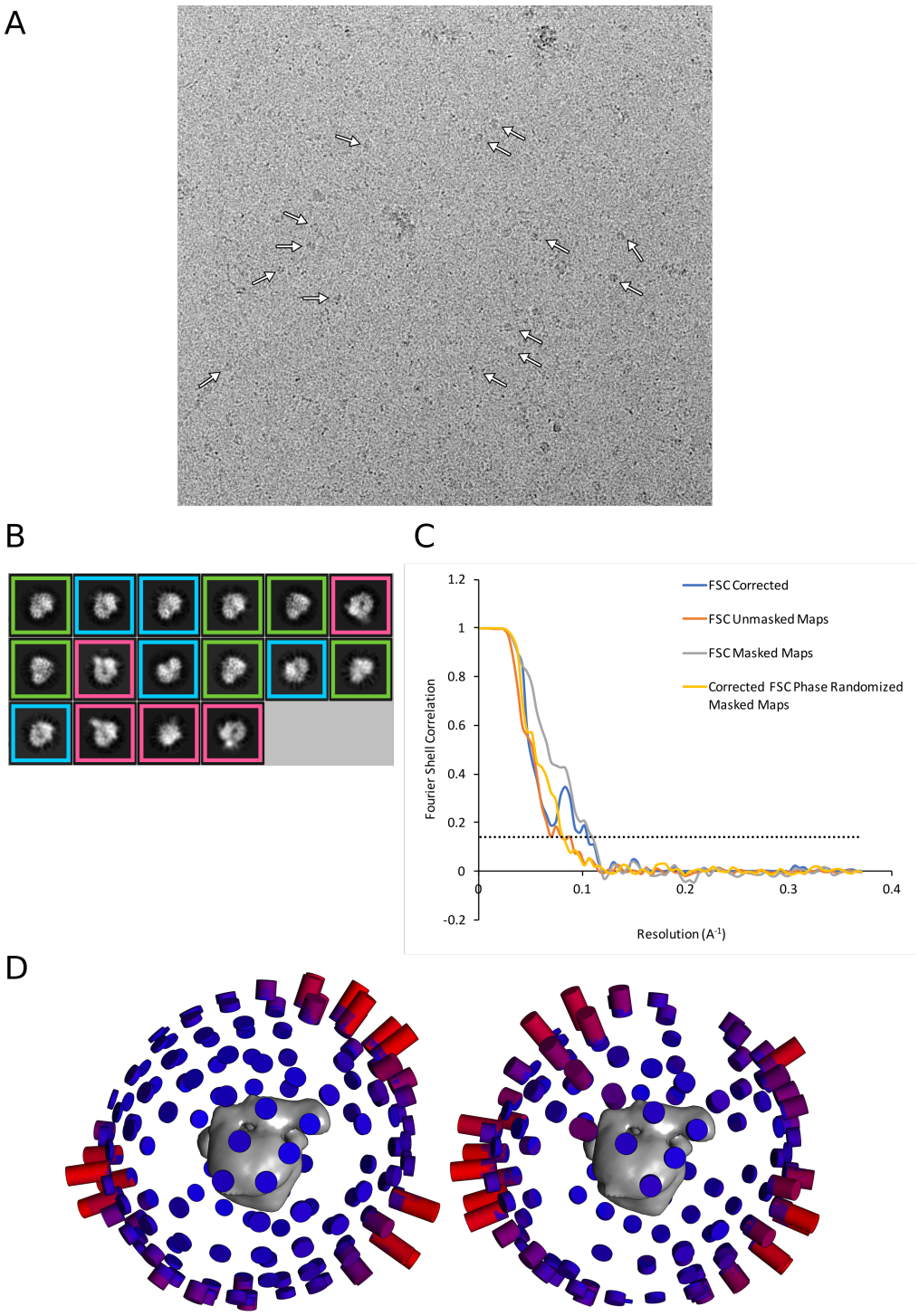


Figure 31 – Cryo-EM analysis of promoter vRNA-bound FluPol$_C$. A) Micrograph of promoter vRNA-bound FluPol$_C$. White arrows point towards individual particles to highlight their position. B) 2D class averages of the RNA-bound FluPol$_C$ dataset. The coloured squares indicate how classes were sorted based on particle orientation for balancing of views. Classes highlighted in green were pooled as "preferential view 1", blue as "preferential view 2" and pink as "other views". A portion of particles from "preferential view 1" and "preferential view 2" were discarded in order to balance the skew in particle orientation. C) Fourier Shell Correlation plot. Dotted line represents the gold standard FSC=0.143 resolution cut-off. D) Particle orientation distribution in 3D classifications before (left) and after (right) balancing particle numbers in the different views.

4.4.3 Cryo-EM of FluPol$_C$ bound to promoter RNA

The cryo-EM map showed a central, slightly flattened globular density with a smaller lobe attached to the side. I proceeded to fit the open, pre-initiation conformation of the polymerase (FluPol$_B$ bound to vRNA PDB ID: 4wsa), and the closed, apo structure (apo FluPol$_C$ PDB ID 5d98) into the density using Chimera) (Fig. 32a, b). This revealed a better overall fit for the open conformation, where the P3 endonuclease domain aligns comfortably with the smaller side density. A significant portion of the PB2 domain region is missing density, as shown by the CBD, which protrudes outside the density. This would indicate that, regardless of the presence of promoter vRNA, the polymerase was still undergoing large conformational rearrangements in solution (Fig, 32a). While most of the vRNA bound FluPol$_B$ structure fits well into the density, a portion of the PA C-terminal domain protrudes out (Fig. 32a). This is unexpected, as this region of the protein interacts closely with PB1 and has been found to be quite stable conformationally when comparing the different available crystal structures (Pflug et al., 2017). However, mapping the local resolution of the density revealed a patch of very poor resolution around this region, which is probably a result of missing views caused by preferential orientations acquired by the polymerase in vitreous ice (Fig. 32c) (Kucukelbir et al., 2014). Fitting of the apo FluPol$_C$ structure was significantly poorer than that of FluPol$_B$. When fitting the structure in Chimera, I consistently obtained two outcomes, one where the PB2 627 domain was placed in the small peripheral density, and one where I guided the P3 endonuclease domain into that density and then allowed the software to adjust the fit. “Fit 1”, with the endonuclease domain in the small density, has multiple regions of the trimer outside the density, including the PB2 627 domain, while also leaving areas of the density unfilled. “Fit 2” has a better overall alignment, but the PB2 627 domain is placed inside the small density, with the P3 endonuclease domain sticking entirely out of the density (Fig. 32b). Given what we know about the high flexibility of the C-terminal PB2 domains and the stability of the endonuclease domain relative to PB1, this scenario where the PB2 627 domain is well defined but the P3 endonuclease domain is so flexible that all of its corresponding density is missing seems unlikely (Pflug et al., 2017).

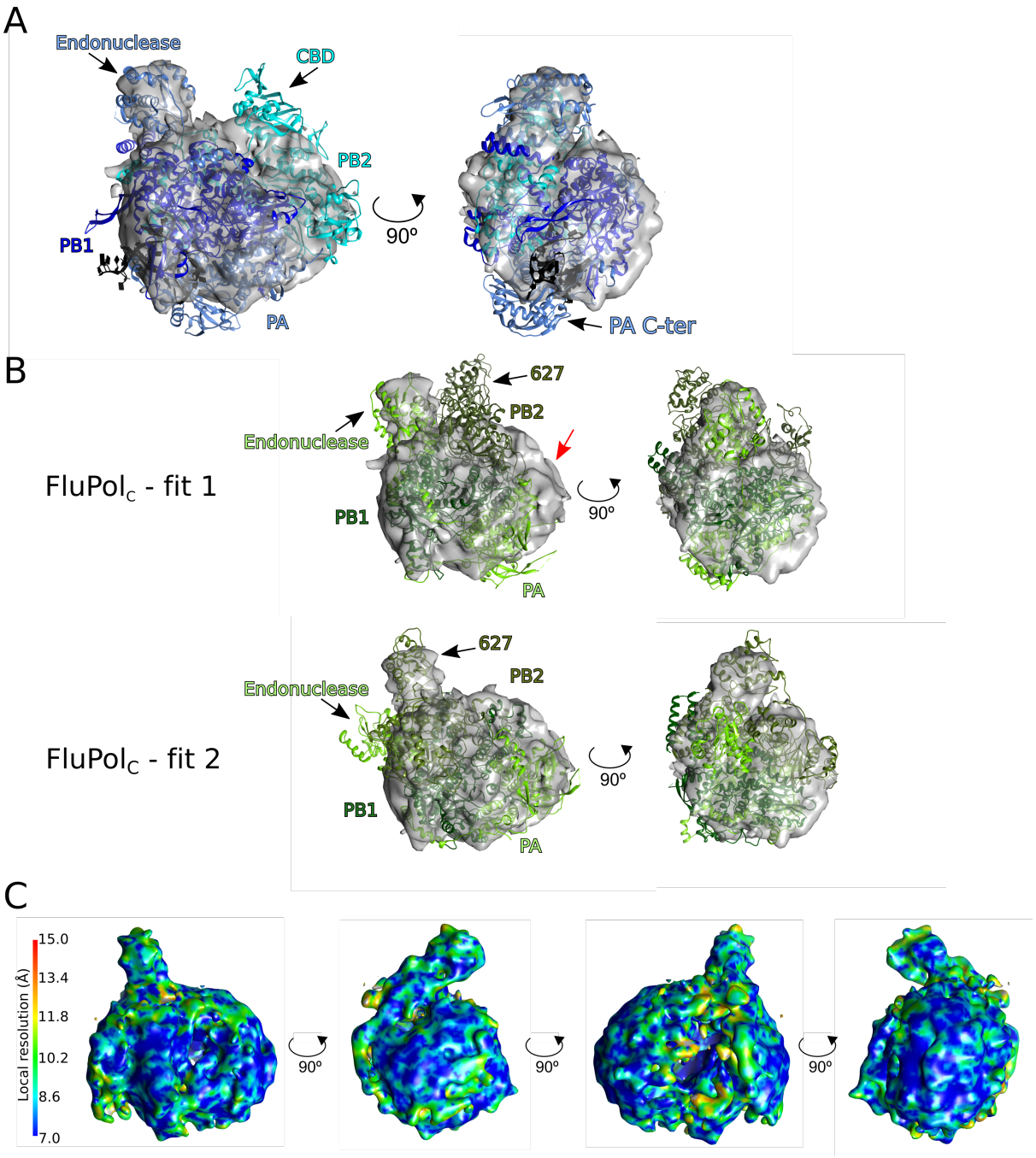


Figure 32 – Cryo-EM density of FluPol$_C$ bound to promoter vRNA. A) Fitting of the vRNA promoter bound FluPol$_B$ crystal structure in the cryo-EM density map of FluPol$_C$ bound to promoter vRNA. PDB ID 4wsa. The position of the endonuclease domain, cap-binding domain (CBD) and PA C-terminal domain are highlighted. Promoter vRNA depicted in black. An area of unfilled density in Fit 1 is highlighted with a red arrow. B) Fitting of the apo FluPol$_C$ crystal structure in the cryo-EM density map of FluPol$_C$ bound to promoter vRNA. PDB ID 5d98. Two different fits are shown. The position of the endonuclease and 627 domains are highlighted. C) Local resolution of the cryo-EM density map. The third panel shows a distinct patch of low resolution.

4.4.4 Conclusions and future directions

Overall, structural characterisation of FluPol$_C$ alone or bound to co-factors involved in transcription remains challenging in our hands. Nevertheless, cryo-EM analysis of vRNA promoter bound FluPol$_C$

has shed light on the effect of promoter addition on polymerase conformational stability. Promoter vRNA binding alone does not appear to be sufficient to keep the polymerase in the pre-initiation conformation. Nevertheless, polymerase structures in the "open" conformation had a better overall fit in the density map obtained than the apo FluPol$_C$ in a closed conformation. Fitting of vRNA bound FluPol$_B$ (Fig. 32a) showed large areas of the PB2 C-terminus outside of the density, which suggests this area of the protein remains highly flexible. Indeed, there were no distinct classes that separated two (or more) conformations, which could indicate very fluid conformational flexibility in this region, with a lack of distinct stable states. On the other hand, this could also simply be a result of having a small dataset; perhaps a large dataset would have yielded clear classes for different conformational states. In any case this suggests that promoter vRNA alone is not sufficient in FluPol$_C$ to achieve a trimer that is stably fixed in an active conformation compatible with transcription. Given the available crystallisation data however, and assuming that FluPol$_C$ behaves like FluPol$_A$ and FluPol$_B$, the promoter RNA does influence the conformation in which the polymerase crystallises. Undoubtedly, addition of RNA to FluPol$_C$ altered the conditions for crystallisation causing no crystals to appear in those conditions where the apo protein had crystallised, which also points towards addition of RNA affecting the polymerase conformation. Together, this would indicate that perhaps addition of vRNA, while not enough to entirely stabilise the trimer in one conformation, does shift the ratio of trimer that is in the open versus the closed conformation in solution at any given time. How this mild stabilisation of the PB2 C terminal domains is achieved by promoter binding is a matter that will require further work, potentially employing techniques like single molecule FRET (Fluorescence Resonance Energy Transfer) to assess changes in conformational flexibility.

4.5 The effect of Pol II CTD in *in vitro* polymerase activity

Our structural data suggests that promoter vRNA binding to polymerase is not sufficient to keep the polymerase in an open conformation compatible with cap-snatching. Since transcription appears to be dependent on interaction with the Pol II CTD, bringing it in close proximity to capped RNA, it seemed plausible that CTD binding would also be involved in regulation of transcription initiation (Engelhardt et al., 2005). Therefore, in order to further investigate the requirements for transcription, I performed a

series of *in vitro* activity assays in presence of both promoter RNA and Pol II CTD peptides. The CTD peptides employed span four heptad repeats and have different phosphorylation states: one is phosphorylated on Ser_5, mimicking the initiating polymerase, another is phosphorylated at Ser_2, which represents the elongating form of the polymerase and finally, the last one is an unphosphorylated peptide. A peptide with a scrambled sequence was used as a control (peptide details can be found in section 2.1.4). FluPol$_C$ expressed in insect cells and purified as described above was used for these *in vitro* assays. In order to allow medium-term storage of the purified protein for use in activity assays, it was placed in a buffer containing 25 mM Hepes pH 7.5, 500 mM NaCl and 40 % glycerol after size exclusion chromatography.

4.5.1 The effect of Pol II CTD in in vitro FluPol$_C$ activity

In order to test the cap-snatching ability of the polymerase, I set up an assay using a capped 20mer oligonucleotide, which the polymerase can bind and cleave to generate a capped primer for transcription. By carrying out the assay in absence of rNTPs, the polymerase can only cleave the oligonucleotide but not extend it, so the only products of the reaction are 10-13 nucleotide long capped RNAs (because of the way this RNA sequence was designed, it will most frequently produce an 11mer with a terminal AG sequence at the 3´ end). Where rNTPs are supplied, the polymerase can then extend this primer using the 3´ promoter 14mer RNA as a template to produce a capped extension product around 24 nucleotides in length. This assay showed that, in absence of promoter vRNA, FluPol$_C$ had no detectable cap cleavage activity (top two panels, lanes 1-5 in Fig. 33). On the other hand, addition of vRNA was not enough to achieve robust activity, producing signal levels only slightly above background in long exposures (top two panels, lane 6 in Fig. 33). Addition of S_5P CTD peptide however, managed to induce an increase in cap cleavage activity, also resulting in a strong band for extension products where rNTPs were included in the reaction (top two panels, lane 7 in Fig. 33). This effect was not observed with any of the other CTD peptides tested (top two panels, lanes 8-10 in Fig. 33).

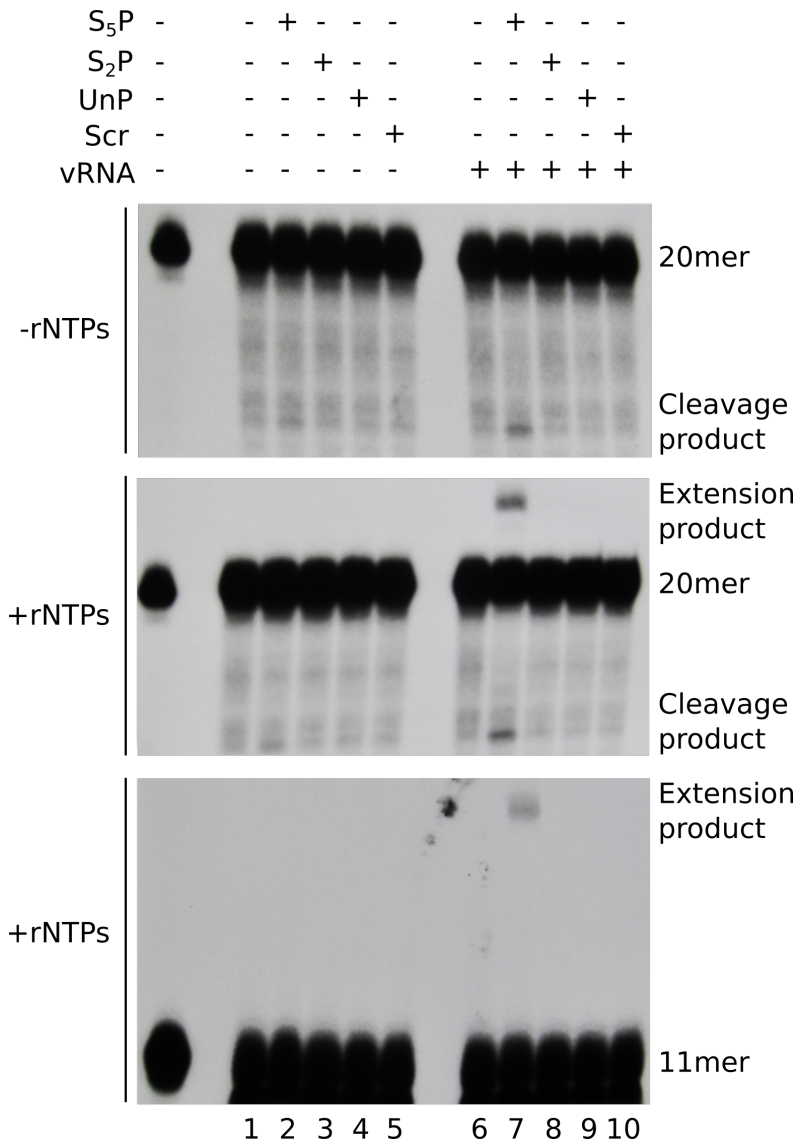


Figure 33 – Effect of CTD peptide addition in FluPol$_C$ *in vitro* activity assays. *In vitro* polymerase assay in presence of S_5P, S_2P, un-phosphorylated and scrambled CTD peptides. From top to bottom: capped RNA cleavage assay; capped RNA cleavage and capped primer extension assay; capped primer extension assay from a synthetic 11mer RNA.

In order to determine whether this enhancement in activity was limited only to the generation of capped primers or whether it also affected extension from such a primer, I carried out a further activity assay using a capped 11mer oligonucleotide primer (bottom panel, Fig. 33). This oligonucleotide mimics the primers generated by cap-snatching and, in presence of rNTPs, will get extended by the polymerase using the 3′ vRNA promoter strand as a template. Therefore, generation of extension products is independent of cap-snatching activity, which allows us to evaluate extension alone. This assay showed that the ability of the polymerase to extend from a capped primer also improved significantly in presence of S_5P CTD peptide, but not any of the other CTD peptides (comparing lane 7

to lanes 8-10, bottom panel, Fig. 33). The polymerase was only capable of extension in presence of promoter vRNA, as it requires a template (comparing lane 2 to lane 7, bottom panel, Fig. 33). Overall, S_5P CTD binding appears to be essential for activation of cap-snatching and elongation from a capped primer in FluPol$_C$.

4.5.2 The effect of Pol II CTD in in vitro FluPol$_A$ activity

Given the effect of adding S_5P CTD in FluPol$_C$ activity assays, I proceeded to test whether a similar effect could be observed when using purified FluPol$_A$. As with FluPol$_C$, FluPol$_A$ was expressed in insect cells and purified following the same protocol. I carried out the *in vitro* activity assay with the capped 20mer as before to analyse capped RNA cleavage and extension from a capped primer. However, FluPol$_A$ did not exhibit the same pattern of activity as FluPol$_C$ (Fig. 34a). In fact, FluPol$_A$ appeared to be highly active in cap cleavage even in absence of promoter vRNA (lanes 1-3 Fig. 34a). Addition of 5′ promoter vRNA greatly enhanced endonuclease activity (lanes 4-6, Fig. 34a), which was increased again upon addition of 3′ promoter vRNA (lanes 7-9, Fig. 34a). Extension from a capped primer only occurred efficiently when both 5′ and 3′ promoters were present, which is explained by the need for a template for extension (lanes 7-9, bottom panel, Fig. 34a). No further enhancement of cap-snatching or extension were observed upon addition of S_5P CTD (lanes 2, 5 and 8, Fig. 34a). As cap-snatching by FluPol$_A$ appeared to be highly efficient, I decided to carry out a time-course for cleavage alone i.e. a reaction in absence of rNTPs, to check whether this would unveil a role for S_5P CTD in the kinetics of the reaction (Fig. 34b). However, at the time points chosen, no clear effect could be observed either. Given how fast the accumulation of cleavage products was, earlier time points may have been more informative. These data highlight a very stark difference in the intrinsic activity of FluPol$_A$ when compared to FluPol$_C$, which is practically unable to cap-snatch or extend a capped primer in absence of RNA and has only background activity in absence of S_5P CTD.

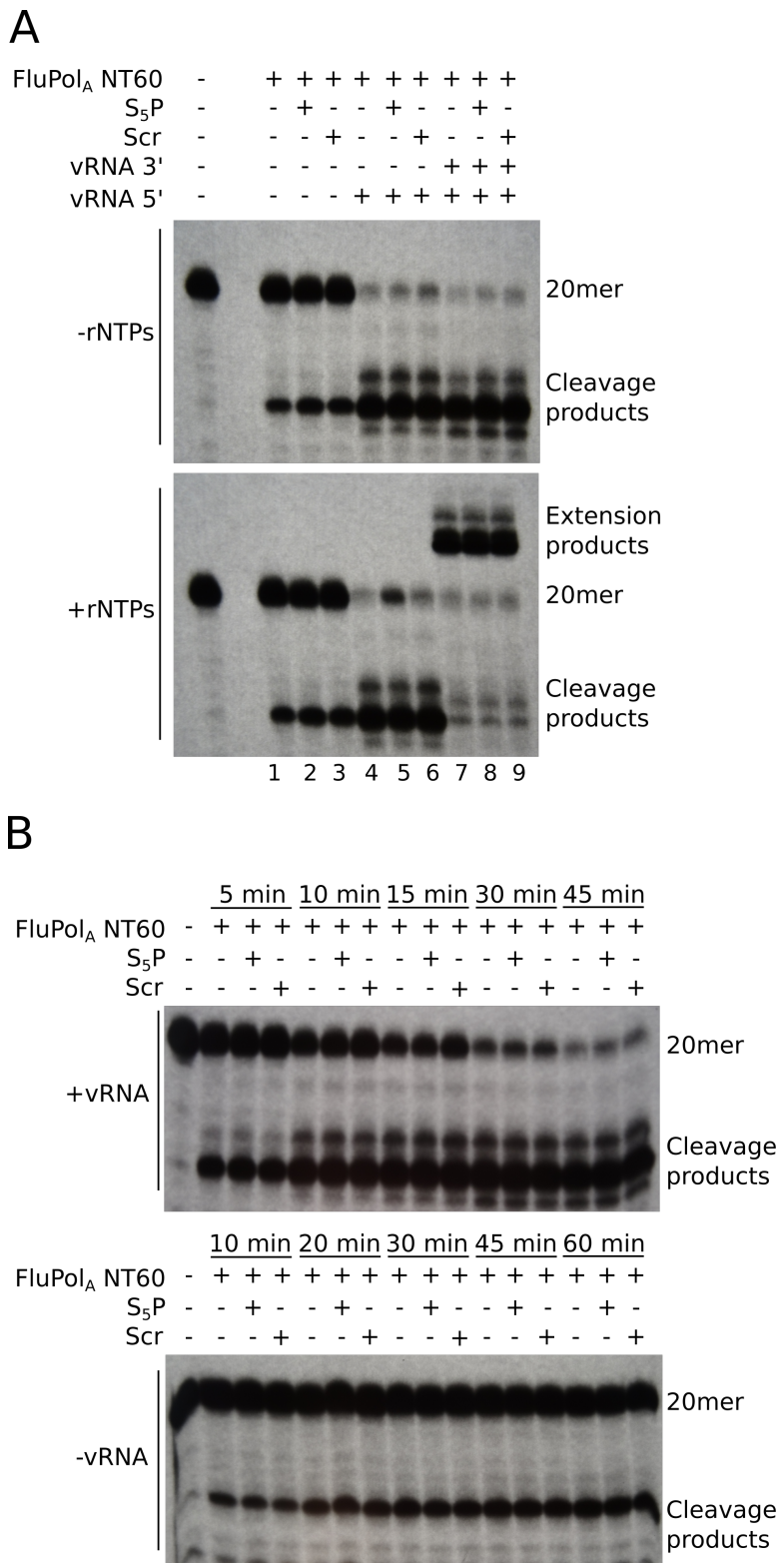


Figure 34 - Effect of CTD peptide addition in FluPol$_A$ *in vitro* activity assays. A) *In vitro* polymerase assay in presence of S$_5$P and scrambled CTD peptides. From top to bottom: capped RNA cleavage assay; capped RNA cleavage and capped primer extension assay. B) Time course of capped RNA cleavage in presence or absence of promoter vRNA.

4.5.3 Conclusions and future directions

Overall, these observations suggest that binding to the S_5P Pol II CTD is crucial to activate transcription by FluPol$_C$. Cap-snatching and extension from a capped primer only occur in the presence of promoter vRNA, which is interesting given that the cap binding and endonuclease domains are located far from the promoter binding site. This is in agreement with the structural data on apo FluPol$_C$ which shows the cap binding site is occluded in the "closed" conformation (Hengrung et al., 2015), as well as with multiple studies showing that cap-snatching only occurs when promoter vRNA is bound to the polymerase (Cianci et al., 1995; Hagen et al., 1994; Leahy et al., 2001; Li et al., 1998). This points towards a two-step activation mechanism, where the polymerase needs to be bound to vRNA (as it would be in a vRNP) to be primed for transcription, but where transcription only occurs efficiently once the polymerase also binds to the initiating Pol II. Transcription initiation is therefore not induced by the presence of capped RNA but by binding to the CTD. However, *in vitro* activity assay data with FluPol$_A$ yielded somewhat conflicting results. The protein was highly active in cap cleavage even in absence of S_5P CTD peptide or even promoter vRNA and we were not able to observe a clear enhancement in activity in presence of S_5P CTD. It is difficult to conclude whether these high levels of polymerase activity also occur in an infected cell or whether they are a bi-product of working with pure preparations of protein in an *in vitro* setting. The FluPol$_A$ and FluPol$_C$ complexes do behave differently during purification (a topic that will be more thoroughly discussed in chapter 5), with the most obvious discrepancy being the fact that FluPol$_C$ behaves as a monomer of the heterotrimer in size exclusion chromatography while both FluPol$_A$ polymerases do so mostly as heterotrimer dimers unless promoter RNA is present. Whether the different rates of activity observed *in vitro* translate to a cellular system and whether this has any implications for pathogenesis remains a mystery but would be an interesting option to consider.

4.6 Summary and model of FluPol transcription activation by Pol II

The structure of FluPol$_C$ bound to S_5P CTD peptide was recently solved by x-ray crystallography in our laboratory by Max Renner and Narin Hengrung to a resolution of 4.1 Å (Fig. 35). The structure was not bound to promoter RNA and preserved the closed conformation previously observed for the

apo form. Two stretches of additional density could be observed on the surface of the polymerase (Fig. 35a), the first, better resolved one, was positioned on the P3 C-terminal domain, close to the tips of helices α24 and α26 and runs towards the P3 linker (Fig. 35b, d). 12 residues of the CTD peptide were modelled into this density: $T_{4a}SeP_{5a}P_{6a}S_{7a}$-$Y_{1b}S_{2b}P_{3b}T_{4b}SeP_{5b}P_{6b}S_{7b}$-$Y_{1c}$. A highly hydrophobic pocket formed by aromatic residues P3 Y241, P3 F663 and PB1 W457 accommodates Y_{1b}, in a manner resembling that observed in other CTD binding proteins (Fig. 35e) (Jasnovidova and Stefl, 2013; Lukarska et al., 2016). The phosphoserine is in turn stabilised by interactions with H461 and R465 of PB1 (Fig. 35f). The second binding site is also found in the P3 C-terminal domain, within a groove formed by P3 β6 and the loop between P3 β11 and β12, at a position resembling the site 2 binding site in FluPol$_A$ (Fig. 35c) (Lukarska et al., 2016). The peptide was not modelled into this density as the resolution was not high enough to determine the position of individual residues, but a 10-residue poly-alanine chain was built in, which is comparable in length to the peptide observed at site 2 in FluPol$_A$ (Lukarska et al., 2016). There is no visible density joining the two CTD binding sites, as was also the case in the published FluPol$_A$ and FluPol$_B$ structures (Lukarska et al., 2016). However, there is a common theme throughout all three polymerases despite the binding sites not being conserved: FluPol has multiple binding sites for Pol II CTD, which suggests that the unstructured CTD could be wrapping around the heterotrimer. The fact that the FluPol$_C$ crystallised in the closed conformation when bound to S_5P CTD is somewhat surprising given the functional data showing that, when bound to S_5P CTD, the protein is capable of cap-snatching. This implies that it acquires a different conformation to that found in the "closed" structure, which is not compatible with cap binding. This may be due to the absence of promoter RNA or due to bias introduced by the crystallisation conditions, which could select for a population of protein in the closed conformation.

Nevertheless, the positioning of the peptide at site 1 suggests it could have a further role in the open polymerase conformation. A model of FluPol$_C$ in the pre-initiation conformation was generated by Max Renner, following the domain positioning observed in the FluPol$_A$ and FluPol$_B$ vRNA promoter bound structures (Fig. 36) (Pflug et al., 2014; Reich et al., 2014). This model showed that, in the open conformation, the 627 domain of PB2 lies above the "site 1" CTD binding site (Fig. 36a, b). We

therefore hypothesised that the S_5P CTD peptide could be stabilising the 627 domain in this position by bridging the P3 C-terminal domain and the 627 domain, therefore allowing the polymerase to remain in an open, transcription-competent conformation. The model supported this hypothesis, showing that the peptide is sufficiently near the 627 domain to interact with it and identifying two residues: H701 on the 627 domain and R748 on the β26 strand of the NLS domain which could be interacting with the peptide (Fig. 36c). The location of the "site 2" binding site observed in the FluPol$_B$ structure bound to S_5P CTD supports this theory as, even though it is positioned differently than "site 1" in FluPol$_C$, it also forms a link between the PB2 627 domain and the PA C-terminal domain. In this case, it runs perpendicular to the interface between these two domains rather than resting under the 627 domain, but functionally, it could be carrying out the same role, suggesting that this mechanism could be conserved across influenza subtypes (Lukarska et al., 2016). Moving forward, structural characterisation of FluPol$_C$ bound to both promoter vRNA and S_5P CTD, even at medium resolution, would be extremely helpful to confirm whether CTD binding does stabilise a transcription-competent "open" conformation of the polymerase and whether the 627 domain interacts with the peptide.

Overall, the functional in vitro data of FluPol$_C$ in presence of S_5P CTD, together with the structure of FluPol$_C$ bound to S_5P CTP, point towards a model where transcription activation in FluPol$_C$ is a two-step event, reliant on the presence of both promoter vRNA and interaction with the Pol II CTD in its S_5 phosphorylated state. In absence of RNA, FluPol$_C$ was found to be essentially unable to cap-snatch. The available crystal structure of the apo form shows that the trimer acquires a "closed" conformation in this state where the cap binding site is occluded, which would explain this observation. On the other hand, SAXS data suggests that, even in absence of RNA, the heterotrimer is highly flexible and oscillates between the closed and open conformations (Hengrung et al., 2015). Addition of RNA induces a slight increase in cap-snatching activity, but the levels remain very low. The cryo-EM data of vRNA promoter bound FluPol$_C$ still exhibited missing density for parts of PB2, which suggests that the trimer remains highly flexible in solution even after addition of vRNA and that this, alone, is not enough to achieve a polymerase that remains stably in the open conformation. On the other hand, the *in vitro* activity assays show a subtle increase in activity, which would suggest the kinetics of this

conformational flexibility are altered in presence of vRNA, which could result in a greater proportion of the polymerase population being able to cap-snatch over a period of time. Binding to S_5P CTD was found to be the key to achieve an actively transcribing polymerase *in vitro*. The structural data on the S_5P CTD peptide-bound FluPol$_C$ provided a clue as to how this activation mechanism may work: binding site 1, which runs from the P3 C-terminal domain towards the P3 linker, rests at an ideal position to interact with the 627 domain in PB2 when the polymerase acquires the open conformation. This would stabilise the 627 domain in that position and allow the polymerase to remain in a transcription-competent conformation. This mechanism would ensure that the FluPol is ready to cap-snatch when in the proximity to Pol II producing capped transcripts. This proposed activation mechanism is summarised in Fig. 37.

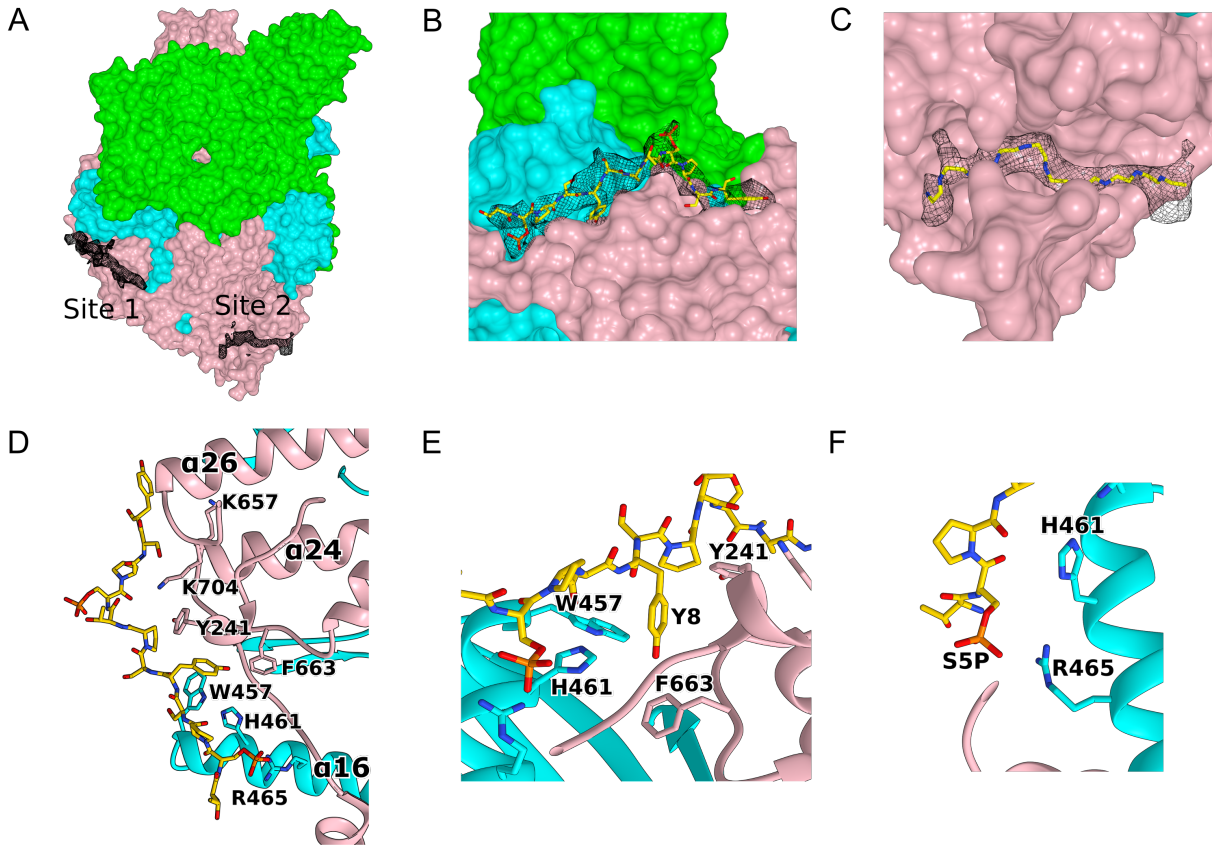


Figure 35 – Crystal structure of FluPol$_C$ bound to S_5P CTD peptide. A) Surface representation of FluPol$_C$ crystal structure with difference electron density maps shown at the two CTD binding sites (Site 1: $2mF_0$-DF_c, 0.9σ, black mesh: site 2: mF_0-DF_c, 2.5σ, black mesh). P3 shown in pink, PB2 in green, PB1 in cyan. B) Binding site 1 close-up with S_5P peptide (yellow) modelled in the electron density map. C) Binding site 2 close-up with poly-alanine chain (yellow) backbone shown in the difference electron density map. D) Detailed view of the S_5-phosphorylated CTD peptide interaction at site 1. E) Detailed view of the hydrophobic pocket that accommodates Y_8. F) Detailed view of the charged pocket that accommodates SeP_{5a}.

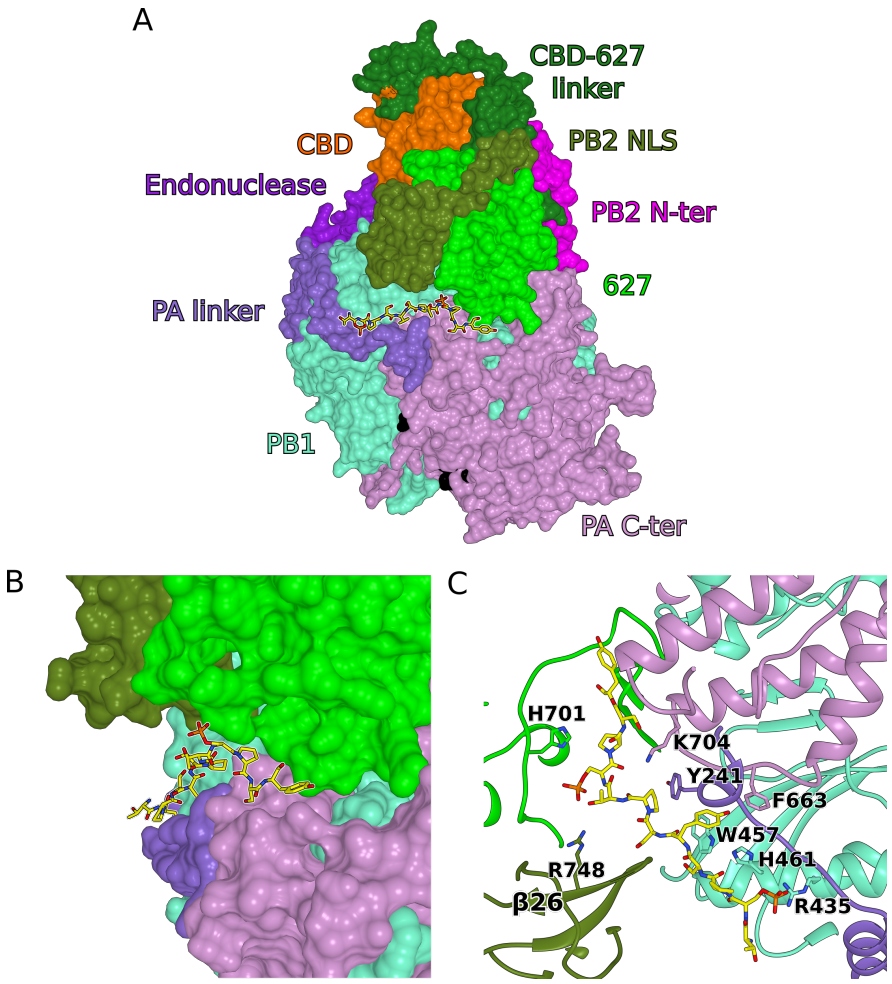


Figure 36 – Modelled FluPol$_C$ bound to S_5P CTD in the "open" conformation. A) Surface representation of the modelled FluPol$_C$ in an open, pre-initiation conformation. CTD peptide is coloured in yellow, promoter vRNA in black. B) Detailed view of the 627 domain packing above the S_5P CTD peptide. C) Detailed view of the S_5P CTD peptide interaction at site 1 in the modelled open conformation FluPol$_C$.

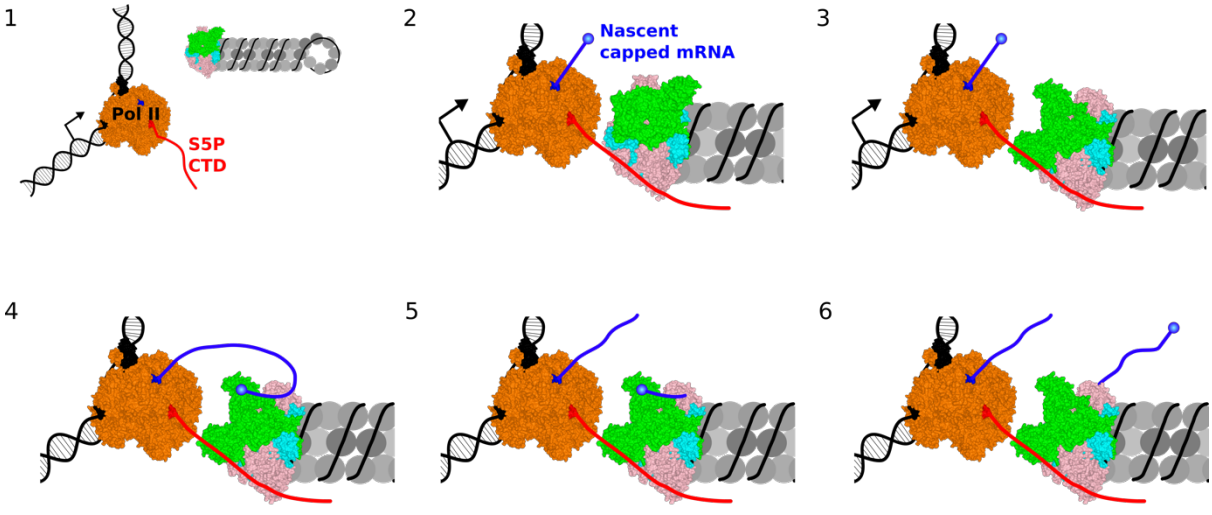


Figure 37 – Proposed mechanism for transcription activation of FluPol$_C$. 1- Prior to association with an initiating Pol II CTD, RNP-associated FluPol is mostly in a closed, inactive conformation, which cannot transcribe efficiently. 2- FluPol binds to the S_5P CTD of Pol II. 3- Binding to the CTD enables the polymerase to remain in the open, transcription-competent conformation as the 627 domain is stabilised by interaction with the CTD. 4. Now in an open conformation, the polymerase binds the nascent pre-mRNA produced by Pol II. 5- Cap-snatching occurs, producing a short, capped primer for transcription. 6- Transcription initiation can take place and the capped primer is elongated.

5 Modulation of influenza A virus polymerase activity by camelid single-chain antibodies

5.1 Introduction

Polymerases of human and avian strains of influenza A virus have resisted structural characterization despite the great advances made in production of the heterotrimeric complex from other influenza types. Indeed, the only high resolution influenza A virus polymerase structures originate from a proposed bat influenza virus (A/little yellow-shouldered bat/Guatemala/060/2010 (H17N10)) (Pflug et al., 2014). Another study employed cryo-EM in the characterization of a polymerase from an avian influenza A strain, but its expression and purification were highly challenging and a heterotrimer missing the C-terminal portion of PB2 had to be employed to improve protein stability (Chang et al., 2015). The C-terminal domains of PB2 being highly flexible, it is not surprising that their deletion made expression and purification easier, on the other hand, these domains (particularly the 627 and NLS domains) contain a high concentration of residues involved in host adaptation and, as described in chapter 4, potentially also have a role in transcription. It is therefore desirable to retain these domains in further structural studies. Overall domain architecture is highly likely to be conserved throughout the influenza viruses, as we have seen from comparison of the available FluPol$_A$, FluPol$_B$ and FluPol$_C$ structures (Pflug et al., 2017). It would nevertheless be interesting to explore transcription and replication directly in the context of a polymerase of avian/human origin as observations made from these are more likely to be relevant to clinically interesting strains.

Multiple aspects of FluPol transcription and replication remain unknown, and it will be difficult to address these without having to painstakingly characterize the structure of the polymerase in multiple states: bound to promoter cRNA, initiating replication, initiating transcription, elongating, terminating etc. We therefore took an alternative approach to exploring polymerase activity. In collaboration with Jan Steyaert, nanobodies were raised against a polymerase from an avian strain, influenza A/duck/Fujian/1/2002 (H5N1), for which Narin Hengrung had generated a baculovirus for its

expression in insect cells. The full list of the 24 nanobodies, their sequences and a sequence alignment for comparison of their hypervariable regions can be found in Appendix 1. Nanobodies are generated from the single-chain antibodies that are naturally produced by camelids and cartilaginous fish (Flajnik et al., 2011). Whereas the epitope-binding region on human antibodies is formed by the variable domains of both the heavy and light chains, in camelids, this is carried out by the heavy chain variable domain alone. These heavy chain-only antibodies consist of three domains with the characteristic immunoglobulin folds from which the variable domain is isolated to produce a nanobody (Fig. 38). Epitope recognition is carried out by three hypervariable loops or complementarity determining regions (CDR1, 2 and 3) on one side of the globular protein. Nanobodies are small (12-15 kDa) and tend to be stable, simple to express and purify, and have high specificity and affinity for their target proteins. In our case, the nanobodies were raised in llamas against apo FluPol$_{Fj}$ which was either crosslinked, biotinylated or unmodified. The nanobodies that had high specificity for the polymerase were selected by phage display, which resulted in a panel of 24 different nanobodies (Pardon et al., 2014). The full set of 24 nanobodies was capable of binding to insect cell expressed FluPol$_{Fj}$ as determined by pull-down assays carried out by Narin Hengrung (not shown).

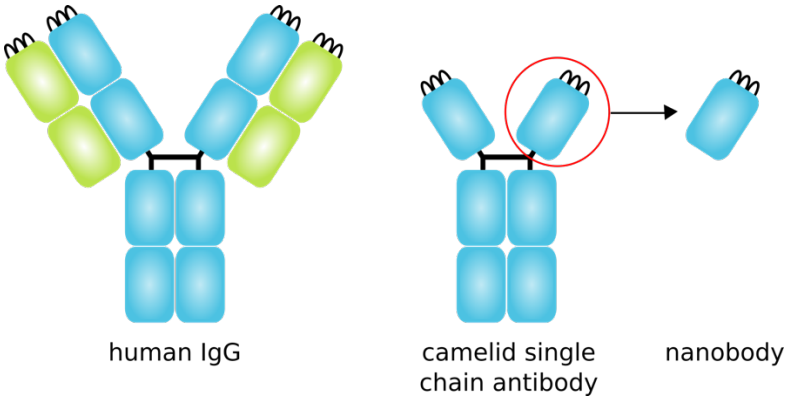


Figure 38 – Schematic diagram comparing the domain arrangement of human IgG with that of camelid single chain antibodies, from which the variable domain is isolated to produce nanobodies.

The rationale behind using nanobodies is that they could enable us to specifically alter the function of the polymerase, for example, by inhibiting capped primer-dependent initiation but not *de novo* initiation, and therefore allow us to study the requirements for these processes. There is also precedent for nanobodies being capable of stabilising transient protein conformations, a function which might be useful in FluPol, as we know it is highly dynamic, which sometimes results in regions of missing

density both in cryo-EM and crystallographic data (Rasmussen et al., 2011; Reich et al., 2014). Nanobodies have been widely used as crystallisation chaperones due to their ability to form crystal contacts while also having a stabilising effect on proteins. Unfortunately, this approach has not been successful with FluPol in our laboratory (work carried out by Narin Hengrung and Haitian Fan). Like promoter vRNA-bound FluPol$_C$, vRNA-bound FluPol$_A$ has resisted all our attempts to crystallise it. Lastly, if any of the 24 nanobodies was found to be efficient at inhibiting all polymerase activity, it would also allow us to pinpoint regions of the polymerase that could be targeted for the design of small molecule inhibitors.

My aim for this project was therefore to first determine the effects of the 24 individual nanobodies on polymerase activity both *in vitro* and in cells. I also attempted to establish a methodology for sample preparation and data collection that would enable us to study the structure of FluPol$_A$ bound to promoter RNA and nanobodies by cryo-electron microscopy.

5.2 FluPol$_A$ expression and purification

Influenza A/duck/Fujian/1/2002 (H5N1) polymerase was expressed in Sf9 insect cells infected with a baculovirus encoding the three polymerase subunits for co-expression (created by Narin Hengrung). The PB2 subunit in this baculovirus is C-terminally tagged with protein A, separated by a TEV cleavage site, which enables the purification of the complex on IgG-coated sepharose beads (Fig. 39a). FluPol$_A$ had a tendency to co-purify with cellular RNAs, but use of RNases in the lysis step risked contaminating the preparation during further purification. As I was interested in the complex bound to promoter vRNA, this was not a feasible option so the protein was purified by affinity chromatography on a HiTrap Heparin HP column. This helped reduce the amount of bound RNA, as determined by measuring the ratio of absorbance at 260 nm:280 nm, as the polymerase preferentially bound to heparin, outcompeting the RNA. The protein was eluted in a NaCl linear gradient up to 1 M NaCl. Interestingly, the polymerase eluted in two, closely overlapping peaks of similar size (Fig. 39b). This is suggestive of two heparin interacting regions with subtly different affinities, which therefore require

slightly different ionic strengths to disrupt. Fractions from both peaks were pooled together for further purification. 5′ and 3′ promoter vRNA was added before size exclusion chromatography in 1.2 molar excess.

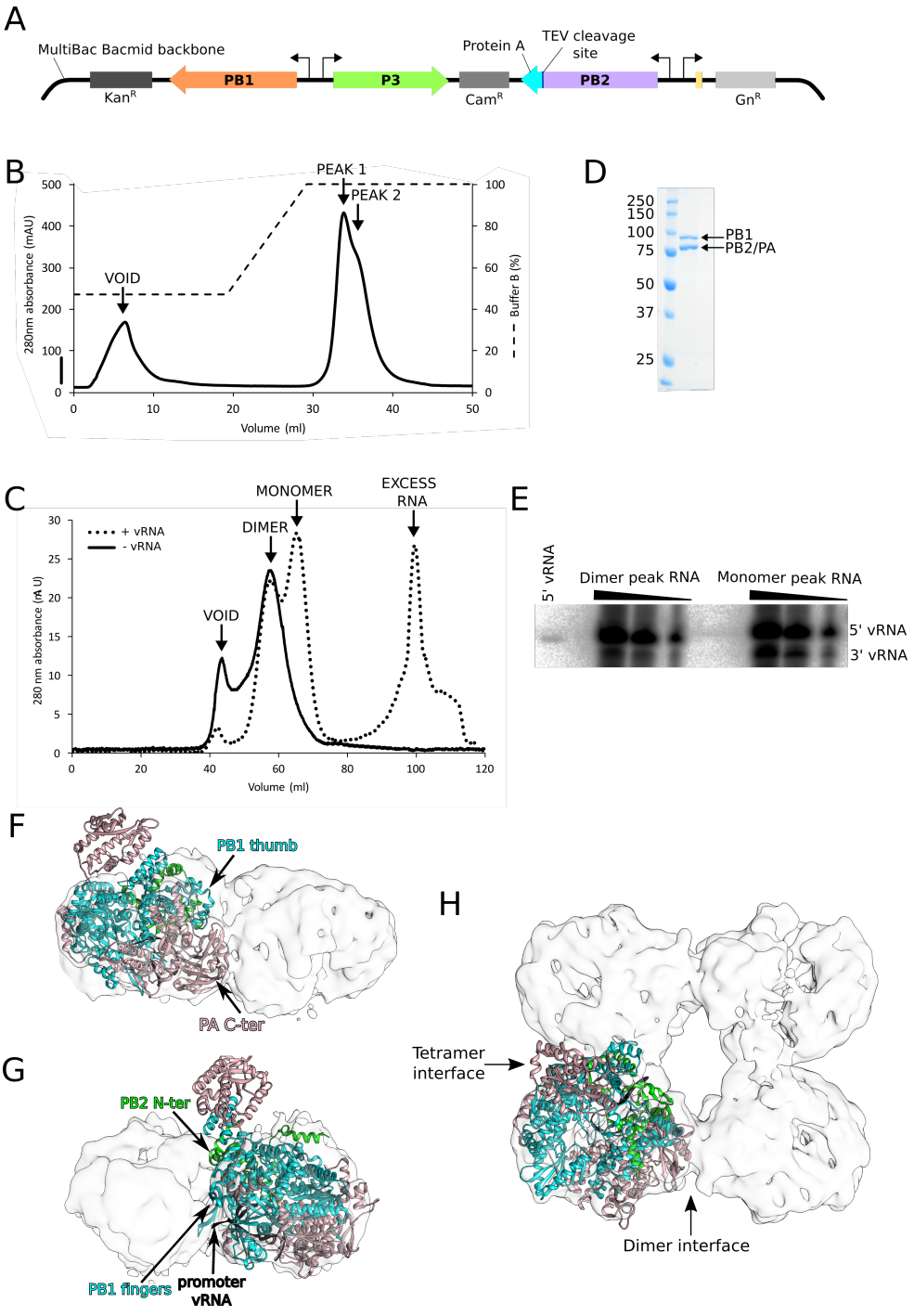


Figure 39 – Purification of FluPol$_A$. A) Diagrammatic representation of the baculovirus genome engineered for expression of the three polymerase subunits. B) Chromatogram of affinity purification on a HiTrap Heparin HP column. FluPol$_A$ elutes in two overlapping peaks on a linear gradient of Buffer B, which contains 1M NaCl. C) Gel filtration profile of FluPol$_A$ in absence and presence of promoter vRNA. Addition of promoter RNA results in a shift towards a monomeric state. D) SDS PA gel of the purified protein stained with SimplyBlue SafeStain. E) End-labelled RNA extracted from dimeric and monomeric gel filtration peaks. Both show the presence of 5′ and 3′ vRNA. F)G)H) Cryo-EM map of FluPol$_A$ (EMDB 6202) with the FluB2 structure fitted into the density (PDB ID 4wrt). FluPol –FluPol interaction interfaces are shown for the dimer (F) and tetramer (G). PB1 is shown in cyan, PB2 in green, PA in light pink and promoter vRNA in black.

The protein was then purified by size exclusion chromatography, which also allowed buffer exchange and removal of excess promoter RNA, producing a highly pure, homogenous sample (Fig. 39c, d). One feature which distinguishes the behaviour of FluPol$_A$ from FluPol$_C$ is that, in absence of RNA, the polymerase elutes as a dimer of the heterotrimer during gel filtration. However, addition of promoter RNA shifts the population towards a mixture of monomer and dimer, both of which contain both strands of promoter RNA in comparable quantities (Fig. 39c). In order to verify this, RNA was isolated by phenol-chlorophorm extraction from purified polymerase after gel filtration, radiolabelled and analysed by PAGE, which show the expected 15mer and 14mer bands (Fig. 39e). Additionally, the A260:A280 ratios for both peaks were always very similar (as measured by NanoDrop), indicating approximately equal RNA content.

Polymerase dimers and tetramers have been previously described for FluPol$_A$ in the cryo-EM study by Chang *et al*., 2015 as well as in overexpressed FluPol purified from HEK293T cells (Jorba et al., 2008) and insect cells (Swale et al., 2016). Chang *et al.* and Swale *et al.* also describe oligomerisation dependency on the presence of RNA. In the Chang et al. study, opposite to what we observe, inclusion of 5´ vRNA induced the formation of higher oligomeric states (tetramer) while 3´ vRNA resulted in monomerisation. Where both were present, tetramers still formed, while the apo protein formed only dimers. From fitting of the FluB$_2$ structure (PDB ID 4wrt) into the EM density map (EMDB 6202), we can see that the dimer interface is formed by residues in the PB1 thumb domain and the PA C-terminal domain (Fig. 39f, h). The dimer-dimer interface that forms the tetramer involves the PB2 N-terminus as well as the PB1 fingers and promoter RNA, which would explain the influence of RNA addition on oligomer formation (Fig. 39g, h). On the other hand, Swale *et al.* also describe dimerization for FluPol containing a PB2 C-terminal truncation, but they find addition of promoter vRNA disrupts dimer formation.

It is difficult to determine whether the dimers we observe are the same as those described in these studies, as we are working with full-length PB2 whereas two out of these three studies worked with

large PB2 deletions. The fact that they find that part of the interaction occurs via the truncated end of PB2 could suggest that their dimer formation is an artefact of working with a truncated protein. On the other hand, the Jorba *et al*., 2008 study also proposes FluPol oligomerisation may occur through PB2 and PB1. This means we essentially have no information on whether the FluPol$_A$ dimer we obtain in gel filtration is biologically relevant or also an artefact of working with purified protein. A consistent population of dimers was also never observed in cryo-EM images or negative stained samples, which points towards a concentration dependency, as samples are diluted before grid preparation (Fig. 43c). As discussed in chapter 3, FluPol-FluPol interaction has been proposed to be part of the transacting/transactivating mechanism for cRNA to vRNA replication but further work will be required to determine whether the dimer formation we observe is linked to this process (Jorba et al., 2008, 2009; York et al., 2013).

5.3 Nanobody screening

Nanobodies were expressed in *E. coli* and isolated from the bacterial periplasm by osmotic shock. They were purified broadly as described in Pardon *et al*., 2014. The nanobodies all contain a C-terminal His tag which enables their purification by affinity chromatography on Ni^{2+} NTA beads. This was followed by size exclusion chromatography, which allows buffer exchange and eliminates any residual contaminants (Fig. 40a). This reproducibly yielded large quantities of pure protein which were used both for *in vitro* activity assays and structural work (Fig. 40b).

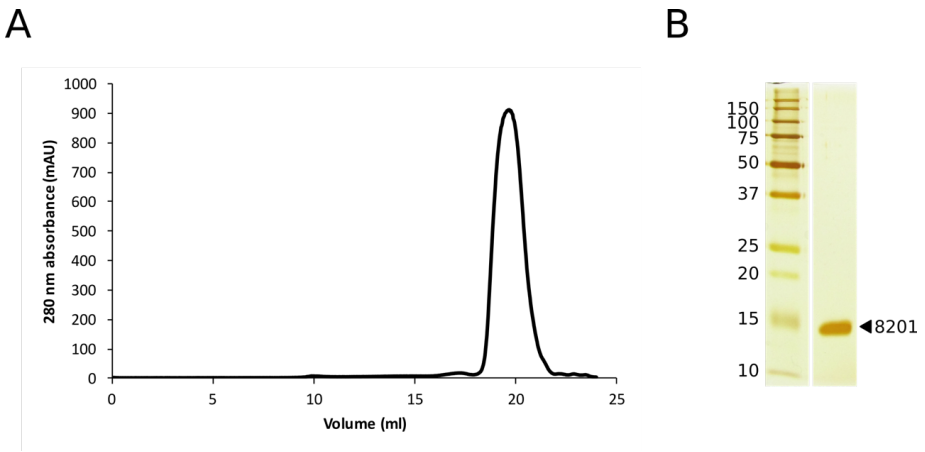


Figure 40 – Nanobody purification. A) Gel filtration profile of nanobody 8201. B) Silver-stained SDS-PA gel of nanobody 8201.

5.3.1 Polymerase in vitro activity assays

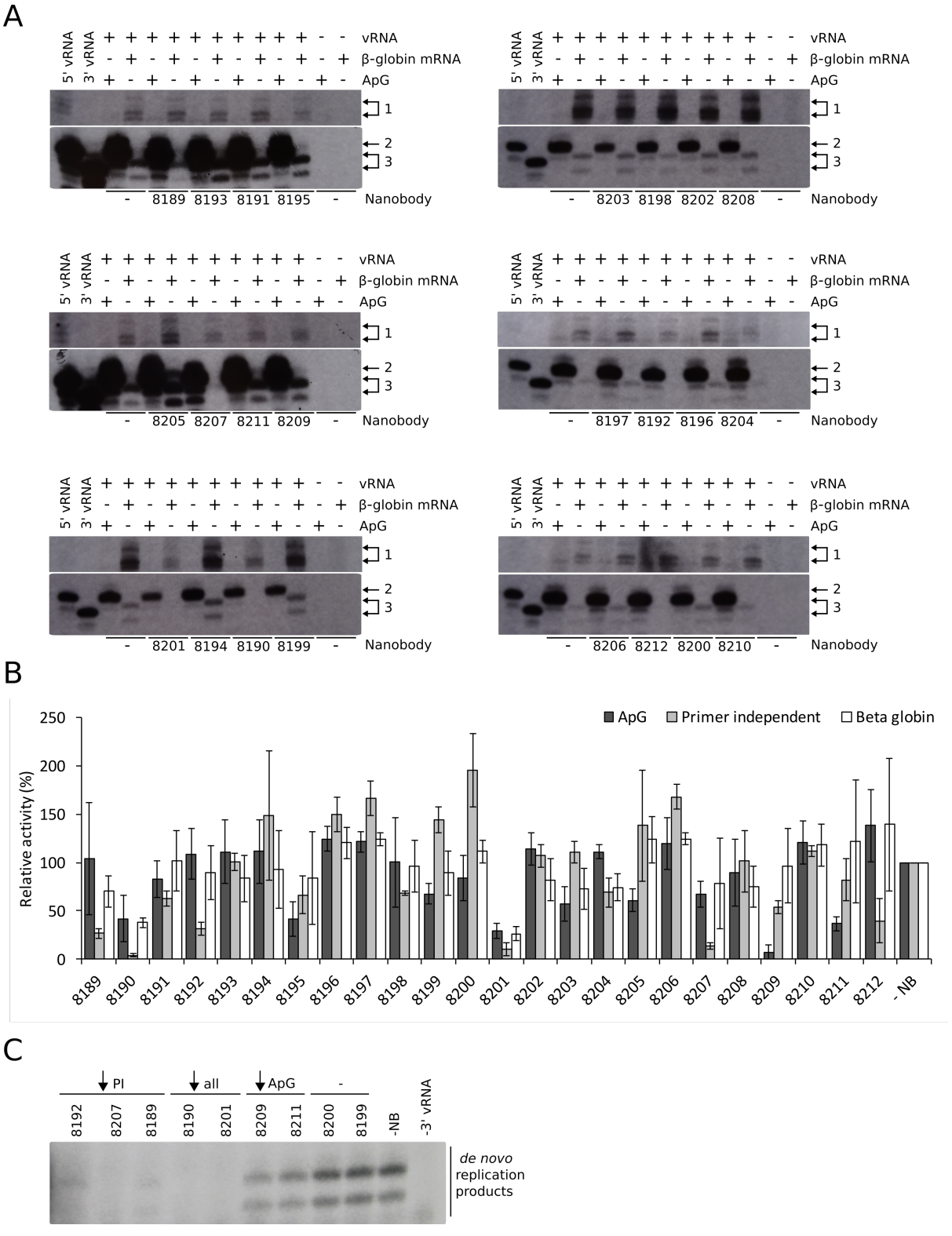


Figure 41 – Effect of nanobody addition on *in vitro* polymerase activity. A) ApG and β-globin mRNA primed *in vitro* polymerase activity assay in presence of nanobodies. 1- β-globin mRNA primed products 2- ApG primed product 3- primer independent products. Note: because of different signal intensities the upper and lower panels have different exposure times. B) Quantitation of A). Mean of three repeats is plotted, error bars represent standard deviation. C) Unprimed *in vitro* polymerase activity assay in presence of nanobodies. ↓PI: nanobodies which reduced primer independent activity, ↓all: nanobodies which reduced all polymerase activity, ↓ApG: nanobodies which reduced ApG dependent activity, -: nanobodies which did not affect polymerase activity.

The set of 24 nanobodies was first tested for their effect on *in vitro* polymerase activity. In this assay, purified FluPol$_{Fj}$ was primed with either ApG or β-globin mRNA to mimic replication and transcription respectively. As previously mentioned, addition of ApG overcomes the rate-limiting step of dinucleotide formation during *de novo* initiation. β-globin mRNA on the other hand serves as a source of capped RNA for cap-snatching and production of short capped primers which are used to initiate transcription. The polymerase is additionally provided with rNTPs as well as promoter vRNA: 15mer 5′ vRNA and 14mer 3′ vRNA, which acts as a template. ApG-primed reactions produce a single, 14mer band which migrates parallel to the end-labelled 15mer 5′ vRNA marker (product labelled 2 in Fig. 41a). The slower migration of this band is due to the presence of a 5′ OH in ApG compared to the single phosphate in the end-labelled RNA markers, which migrate faster due to the extra negative charge. β-globin mRNA primed reactions generated two types of products: the bands with slower migration (labelled 1 in Fig. 41a) correspond to those that result from cap-snatching activity, and therefore contain an extra 10-13 nucleotides at the 5′ end. These products vary slightly in length due to the different cleavage positions during primer generation but were quantitated together. In addition to that, two shorter bands are produced (labelled 3 in Fig. 41a), which could correspond to *de novo* initiation. These migrate faster than the 15mer and 14mer end-labelled markers as the contain a triphosphate at their 5′ end.

Some of the nanobodies tested exhibited interesting effects on polymerase activity (Fig. 41a, b). Addition of 8190 and particularly 8201 resulted in a decrease in all activity, essentially eliminating *de novo* initiation products and leaving only residual amounts of ApG and β-globin mRNA primed products. Somewhat surprisingly, none of the nanobodies exhibited specific inhibition of cap-snatching, which one may have expected to be easily targeted by nanobodies as the PB2 CBD and the PA endonuclease domain are highly exposed. On the other hand, these domains may not have produced tightly binding nanobodies which therefore weren't selected during phage display or perhaps nanobody binding to them is not deleterious, as the domains are highly flexible and may be able to carry out their function with a nanobody bound to them. Several nanobodies exhibited strong specific inhibition of *de novo* initiation: 8189, 8192, 8207 and 8212. Additionally, 8195, 8203, 8211 and

especially 8209 resulted in decreased ApG-dependent activity, which was not coupled with reduced levels of unprimed activity. How ApG primed activity is inhibited without affecting unprimed activity is unclear as they both mimic capped primer independent initiation. Studying these nanobodies may enable us to determine the requirements for replication, as the polymerase-nanobody complexes formed are clearly still capable of polymerizing, which implies that none of the essential features of the polymerase are affected (e.g. active sites or entry/exit channels), but they are strongly impaired in *de novo* initiation. Addition of two nanobodies, 8199 and 8200, resulted in the opposite phenotype, increasing the quantity of unprimed products. Perhaps these stabilize a replication-competent conformation of the polymerase that is more efficient at *de novo* initiation.

In order to verify that what we term "unprimed products" in this assay were truly a result of *de novo* initiation, a selection of nanobodies which displayed interesting phenotypes were tested in an unprimed *in vitro* reaction, where no ApG or β-globin mRNA was added (Fig. 41c). This produced the same banding pattern as well as the same modulation of activity by nanobodies observed in the β-globin mRNA primed reactions, indicating that this is indeed the source of these bands and that evaluating the nanobody effect on *de novo* initiation in reactions containing β-globin mRNA was a valid approach.

5.3.2 RNP reconstitutions

After characterizing the *in vitro* effect of nanobodies on polymerase activity I proceeded to assess their effect on transcription and replication in cells. In order to do this, all 24 nanobodies were cloned into a pcDNA 3A vector for expression in mammalian cells. The effect of nanobodies was assessed in RNP reconstitution assays, where HEK293T cells were transfected with all the necessary components to produce functional RNPs (FluPol subunits, NP and an NA segment template) in addition to nanobodies. This was initially carried out with influenza A/duck/Fujian/1/2002 (H5N1) FluPol and NP. Total RNA was extracted from transfected cells and viral RNA species detected by primer

extension with NA-specific primers NA160 and NA1280 (see Fig. 22a for a schematic diagram of the NA primer binding positions and product sizes).

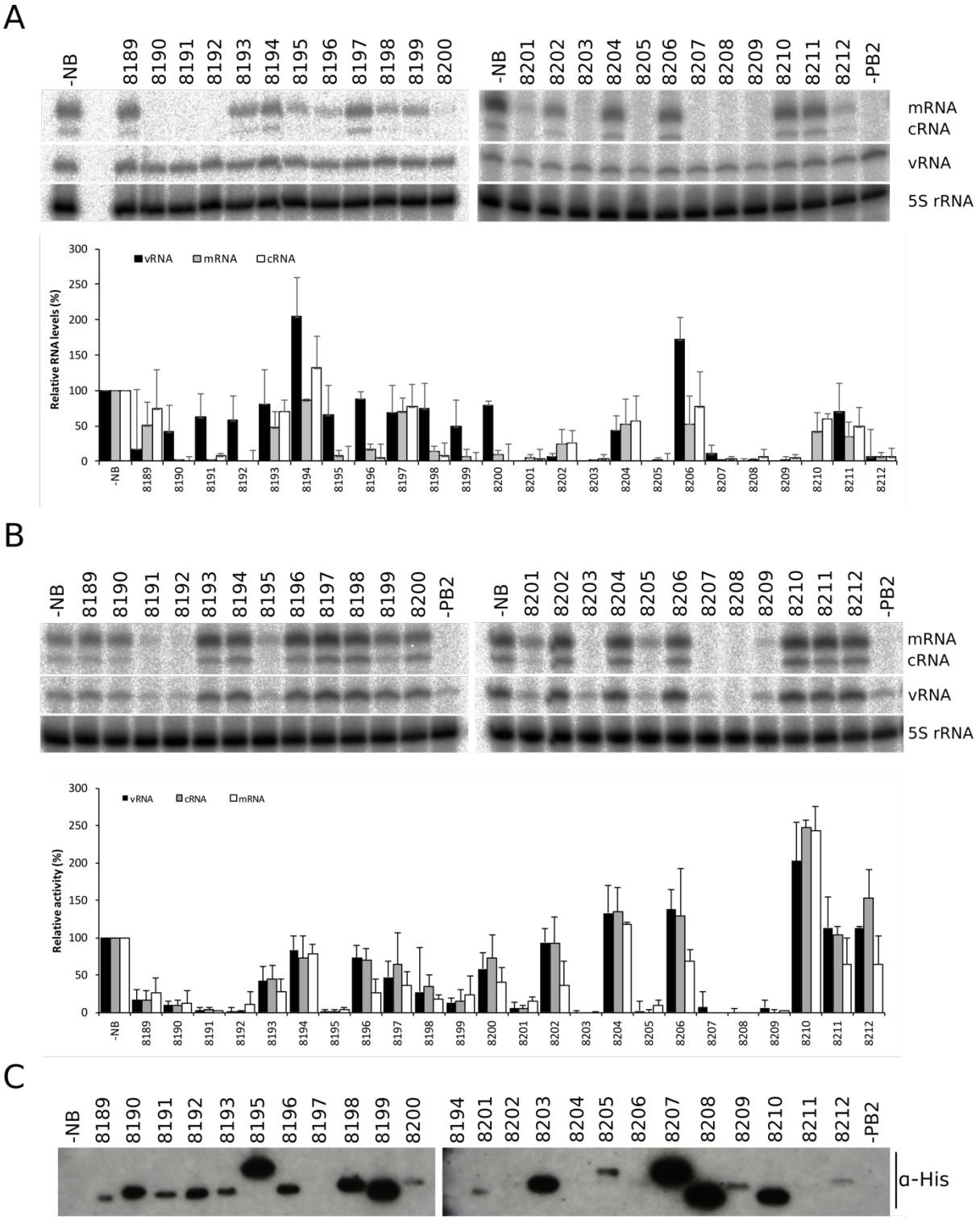


Figure 42 – RNP reconstitutions in presence of nanobodies. A) FluPol$_{Fj}$ RNP reconstitution in presence of nanobodies. Total RNA was analysed by primer extension with primers NA160 and NA1280. Quantitation – mean of 3 repeats is plotted with error bars representing standard deviation. B) FluPol$_{WSN}$ RNP reconstitution in presence of nanobodies. Total RNA was analysed by primer extension with primers NA160 and NA1280. Quantitation – mean of 3 repeats is plotted with error bars representing standard deviation. C) α-His western blot on total cell lysate detecting nanobodies in transfected cells.

In this assay, nanobodies caused more extensive inhibition than that observed in *in vitro* activity assays (Fig. 42a). This is not surprising as polymerase activity in cells is reliant on multiple factors which are not present *in vitro*, for example: nuclear import, trimer assembly, interaction with NP and interaction with host proteins, including Pol II. Over half of the nanobodies caused a strong inhibition of all polymerase activity, however, none of the nanobodies appeared to have a differential effect on replication over transcription or vice versa. Both of the nanobodies (8190 and 8201) which inhibited all activity *in vitro* also had a strong inhibitory effect in RNP reconstitutions. In addition to that, most of the nanobodies which affected primer independent or ApG primed activity *in vitro* also caused a drastic decrease in all activity in this assay. Some nanobodies which had no significant effect *in vitro* however, were also capable of achieving similar levels of inhibition in this assay, which included 8191, 8200, 8204, 8205 and 8196. This may suggest that these nanobodies don't interfere with polymerase activity per se but with one of the other steps required for FluPol to function in a cellular context.

RNP reconstitution with influenza A/duck/Fujian/1/2002 (H5N1) RNP components in HEK293T cells was found to be feasible although not very efficient. This is not surprising given that FluPol$_{Fj}$ comes from an avian influenza virus strain so it is subject to host restriction mechanisms when trying to carry out transcription and replication in a mammalian cell line. This was particularly apparent in the levels of vRNA which were barely above those seen in the negative control lacking PB2. This indicates that multi-cycle replication was inefficient, particularly at the cRNA to vRNA step.

In order to assess cross-reactivity of nanobodies against different influenza A virus strains, I also carried out this assay with FluPol and NP from a mammalian adapted strain: influenza A/WSN/33 (H1N1) (Fig. 42b). This polymerase was capable of significantly more robust levels of transcription and replication in HEK293T cells, as shown by accumulation of vRNA in the nanobody-free reactions compared to the negative control lacking PB2. The phenotypes induced by nanobodies were quite similar to those seen with FluPol$_{Fj}$ except for a couple of nanobodies which were inhibitory against FluPol$_{Fj}$, which were not effective against FluPol$_{WSN}$. This was the case for nanobodies 8200, 8204,

8212 and 8196. On the other hand, nanobody 8210, which didn't have a significant effect on FluPol$_{Fj}$ activity appeared to enhance transcription and replication levels in FluPol$_{WSN}$. Sequence conservation between FluPol$_{Fj}$ and FluPol$_{WSN}$ is relatively high: amino acid identity is 96% for PB1 (728/757 residues), 95% for PB2 (722/759 residues) and 96% for PA (689/716 residues). Many of these amino acid substitutions are also conservative. Nevertheless, our data suggests these variations are sufficient to disrupt binding by some of the nanobodies.

In order to assess nanobody expression levels in the transfected HEK293T cells, I carried out an α-His western blot on the total cell lysate (Fig. 42c). Surprisingly, even though the nanobodies only differ in sequence in the three hypervariable loops, expression levels were extremely variable, with some nanobodies not even reaching detectable levels while others exhibited robust expression. Nevertheless, expression levels did not correlate with inhibition phenotypes. Indeed, some of the more poorly expressed nanobodies were among those that consistently suppressed polymerase activity. This was the case for 8200, 8202, 8211 and 8212, which specifically inhibit FluPol$_{Fj}$ but not FluPol$_{WSN}$, as well as 8201, 8205 and 8209 which inhibited activity by both polymerases in RNP reconstitutions. On the other hand, some of the nanobodies which displayed low expression levels had no effect on polymerase activity including 8194, 8197, 8204 and 8206. As none of these inhibited polymerase activity *in vitro*, it is hard to conclude whether their lack of effect in RNP reconstitutions is down to low expression levels or a genuine phenotype.

5.3.3 Conclusions and future directions

All of the nanobodies in this set of 24 belong to different families. Generally, nanobodies of the same family have similar CDR3 regions and are considered to bind the same epitope. No clear patterns of CDR3 similarity were found in nanobodies that caused the same phenotype (see nanobody sequence alignment in Appendix 1). It is possible however, that some of the nanobodies that have a similar effect on activity *in vitro* or in RNP reconstitution assays bind the same region on the heterotrimer. Overall, the nanobodies displayed some interesting patterns of modulation of polymerase activity, particularly restricting replication in *in vitro* activity assays, which translated as strong inhibition of all

activity in RNP reconstitutions. RNP activity assays using FluPol$_{Fj}$ were somewhat limited by host restriction effects which resulted in lowered overall polymerase activity, particularly replication for the production of progeny vRNA. This assay would probably benefit from the use of a FluPol$_{Fj}$ with an E627K mutation in the PB2 627 domain to improve replication in mammalian cells. Being a single point mutation it is unlikely to affect binding by most nanobodies, and if it did it would be an excellent indication of their binding location.

In terms of further characterizing our panel of 24 nanobodies, it would be interesting to determine the effect that nanobodies have on viral growth. More importantly, it would be key to determine their binding epitopes. This could be carried out in multiple ways the simplest of which would be testing co-precipitation of nanobodies with the individual viral polymerase subunits. On the other hand, some of the nanobodies may bind at subunit interfaces which reduces how informative this assay would be. An alternative would be to carry out competition assays for binding to FluPol, adding two or more nanobodies to purified polymerase at a time and analysing the number of nanobodies which bind by SEC-MALS (Size Exclusion Chromatography – Multi-Angle Light Scattering). This would allow us to identify nanobodies which share the same (or spatially proximal) binding epitopes. Evidently, the most definitive way of identifying the binding sites of nanobodies is structural characterization of FluPol-nanobody complexes. This would of course also enable us to determine their mode of action based on their location and the conformation in which the polymerase is found. Therefore, I proceeded to study FluPol$_A$-nanobody complexes by cryo-electron microscopy. I initially decided to focus my attention on the nanobodies which inhibited all polymerase activity both *in vitro* and in a cellular context, more specifically, on nanobody 8201, which caused the strongest suppression of FluPol activity *in vitro*.

5.4 Cryo-EM sample preparation and screening

In order to determine the mechanism by which nanobody 8201 causes the strong inhibition phenotypes observed both in *in vitro* polymerase assays and in RNP reconstitution assays (both with FluPol$_{Fj}$ and

FluPol$_{WSN}$), I attempted to structurally characterize the complex of FluPol$_A$ bound to nanobody 8201 and promoter vRNA. As we have had no prior success in our laboratory in crystallising promoter RNA bound polymerase (both for FluPol$_A$ and FluPol$_C$) the method of choice for this was cryo-electron microscopy. As discussed in section 4.4.2, recent advances in single particle cryo-EM have made it an attractive alternative for the structural characterisation of small protein complexes to crystallography, particularly for those which are challenging to crystallise or tend to produce poorly diffracting crystals. This project was therefore also an opportunity to develop a protocol for the analysis of FluPol by cryo-EM, which could allow us in future to analyse multiple states of the protein (including FluPol in complex with other nanobodies that display interesting phenotypes) in a more streamlined manner than is currently possibly with crystallography. However, cryo-EM presents its own set of challenges in terms of sample preparation, data collection and analysis. Below I detail my multiple attempts at optimising protocols for these and their outcomes.

5.4.1 Purification of FluPol$_A$-nanobody-promoter vRNA complexes

Formation of FluPolA-8201 complexes proved to be relatively straightforward, undoubtedly due to the high affinity of this nanobody for the polymerase. Addition of nanobody therefore did not require extensive modification of the protocol employed for the purification of FluPol$_A$ in complex with promoter vRNA. Purified nanobody was added in 1.5 molar excess after purification of FluPol$_A$ by heparin affinity chromatography and before gel filtration, immediately after addition of promoter vRNA 5´and 3´strands (15mer and 14mer respectively). The mixture was then run through a gel filtration column, which allowed for buffer exchange and removal of excess RNA and nanobody (Fig. 43a). The gel filtration profile showed a very large void volume when both nanobody and promoter vRNA were added to FluPol$_A$. The complex of FluPol$_A$-8201-promoter RNA (from now on referred to as nbFP$_A$ for simplicity) effectively behaved in an identical manner to FluPol$_A$-promoter RNA, eluting as a mixture of dimers and monomers of the complex. Yet again, these were indistinguishable in terms of protein or RNA content. Nanobody 8201 co-eluted with the polymerase peaks (peak fraction shown in figure 43b), with the excess unbound nanobody running at higher elution volumes at a position similar to the excess RNA. Overall, this protocol produced very pure preparations of the complex,

with yields which were modest (generally around 0.5 mg) but sufficient for analysis by cryo-EM. On negative stain the sample looked relatively homogeneous, with no evident signs of a consistent dimer population, likely due to the fact that samples were diluted to 0.05 mg/ml before being applied to grids (Fig. 43c).

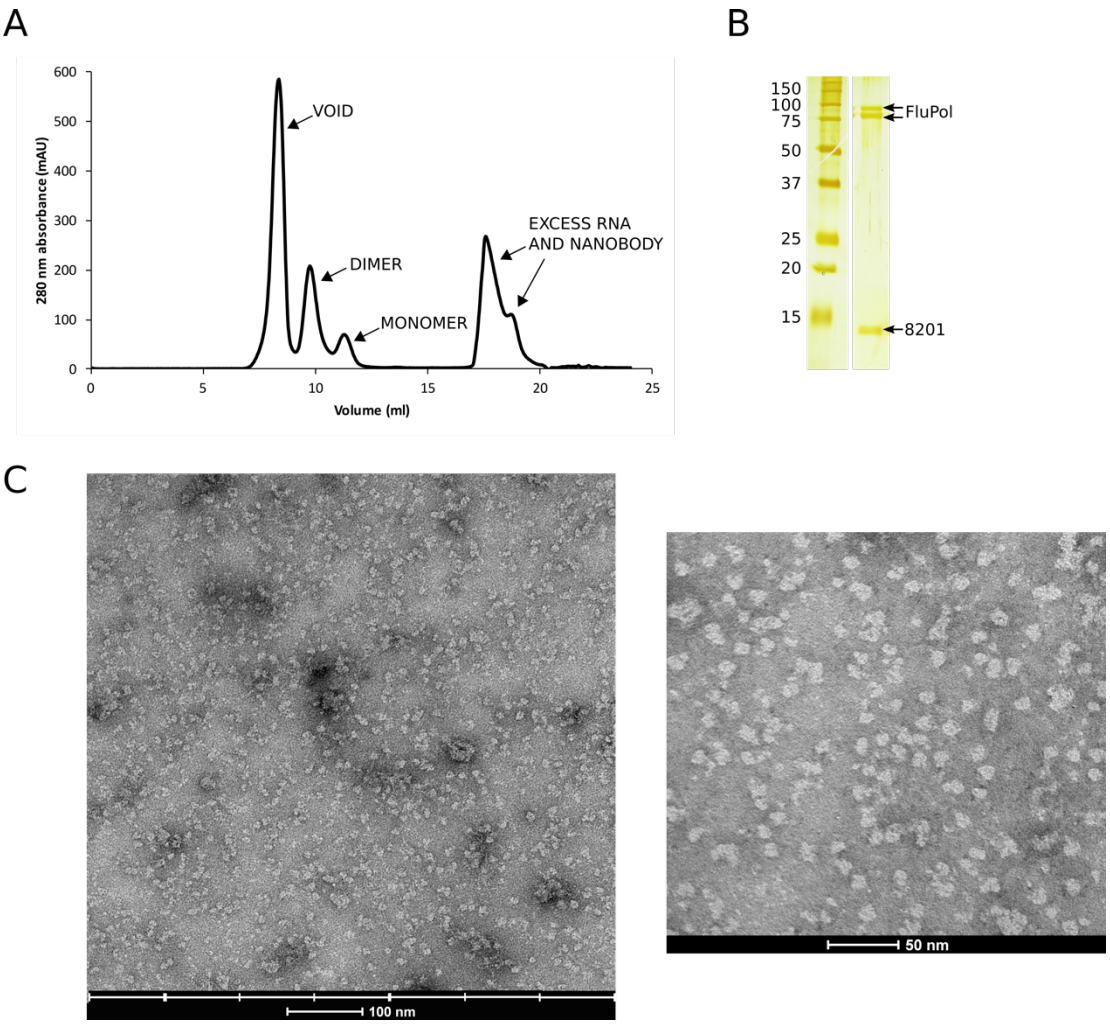


Figure 43 – Purification of the FluPol$_A$-8201-promoter vRNA complex. A) Gel filtration profile of FluPol$_A$ in presence of nanobody 8201 and promoter vRNA. B) Silver-stained SDS PA gel of gel filtration peak fraction containing nbFP$_A$ and nanobody 8201 (as well as promoter RNA as determined by a change in absorbance at 240 nm as measured by NanoDrop). C) Micrographs of negative stained sample of nbFP$_A$ showing individual particles.

5.4.2 Cryo-EM sample preparation

The nbFP$_A$ purification protocol had to be modified in order make the sample suitable for cryo-EM grid preparation. As with FluPol$_C$, the purification of FluPol$_A$ leaves the protein in a buffer containing 500 mM NaCl which, as discussed in section 4.4.1, was not compatible with the formation of good amorphous ice that would allow sufficient contrast between the background and the protein particles.

The main focus was therefore to find alternatives to the buffer employed in gel filtration (25 mM Hepes pH 7.5, 500 mM NaCl) that would allow a reduction in the NaCl concentration while retaining protein stability in solution and preventing aggregation. Unless otherwise indicated, in order to place the protein in a new buffer for testing in grid preparation, the protein was concentrated at least 5x above the concentration required for cryo-EM and then diluted into the desired buffer. Alternatively, buffer exchange was carried out during protein concentration on centrifugal filter units. In terms of preparation of grids for cryo-EM, several grid types were tested in order to optimise ice thickness and the behaviour of the protein on the grid. Within these, multiple variations of blotting conditions (blotting time, humidity, blotting force etc.) were also tested. Grids were prepared on a Vitrobot Mark IV (FEI) and screened either on a 120 keV Tecnai T12 microscope (FEI) or on a 300 keV Tecnai Polara microscope (FEI). A selection of some of the conditions tested detailing buffer compositions and grid types can be found in Table 8, for which I will provide a rationale below.

5.4.3 Strategies for reducing NaCl concentrations

My initial approach for substitution of NaCl in the sample was to replace it with a phosphate based buffer (mixture of dibasic sodium phosphate and monobasic sodium phosphate). As phosphate groups have a higher negative charge than chloride ions, solutions of lower molarity would have an equivalent ionic strength to solutions containing higher concentrations of sodium chloride. Polymerase was found to be stable in a buffer containing 20-80 mM phosphate buffer at pH 7.5 when combined with 100 mM KCl or NaCl. However, grids produced from protein in phosphate buffer had a tendency to contain thick ice or very "grainy" looking ice, which made particle identification difficult. Nevertheless, through screening, a buffer containing 60 mM phosphate buffer pH 7.5 and 100 mM KCl was found to produce areas on the grid with a gradient of ice thickness. On this gradient were areas of thin ice where the protein particles were very clearly visible and where particle concentration was high. A dataset was therefore collected on this sample, the results of which are detailed below under section 5.5.1.

Based on our observations with FluPol$_C$, which showed that the polymerase was stable in buffers containing sodium thiocyanate (section 4.4.2), I decided to test whether this would also be a suitable buffer component to replace NaCl in nbFP$_A$ preparations. This indeed appeared to be the case: all preparations which contained sodium thiocyanate produced grids where particles were clearly visible. I attempted to lower the concentration of NaSCN to determine the minimum quantity required to prevent extensive aggregation of the protein sample. A 1 in 3 dilution of nbFP$_A$ (in gel filtration buffer) into a buffer containing 25 mM Hepes pH 7.5 and 20 mM NaSCN which was then quickly applied to the grid before plunging generated good results, with optimal particle distribution and reasonable contrast of the particles against the background. A dataset was also collected on this sample, the results of which are detailed under section 5.5.2.

The inclusion of several additives in the protein buffer was also tested to determine whether they would improve protein stability and therefore prevent aggregation without having to resort to the use of other salts. Those tested included non-detergent sulfobetaines (NDSBs), glutamate and ATP. Use of NDSBs has been described in the literature for the solubilisation of halophilic proteins and as general aids to protein folding (Expert-Bezançon et al., 2003; Vuillard et al., 1995). I therefore reasoned that they might have a beneficial effect on nbFP$_A$ stability. Glutamate-arginine combinations have been widely employed in buffers as a method to prevent protein aggregation (Golovanov et al., 2004). As the FluPol is an RNA binding protein with multiple channels running through it for RNA template/product entry and exit, it contains multiple regions rich in positively charged amino acids (for interaction with nucleotide phosphate moieties). I therefore included glutamine as an additive in the buffer to supply additional negatively charged molecules in solution, which I hypothesised might stabilise basic regions of the protein. ATP was tested for a similar reason: it has been found to improve protein solubility, in addition to being a negatively charged molecule with which the polymerase naturally interacts (Patel et al., 2017). All of these were employed in buffers with a lowered NaCl concentration of 150 mM. Unfortunately, none of them had a significant effect in reducing polymerase aggregation under these conditions.

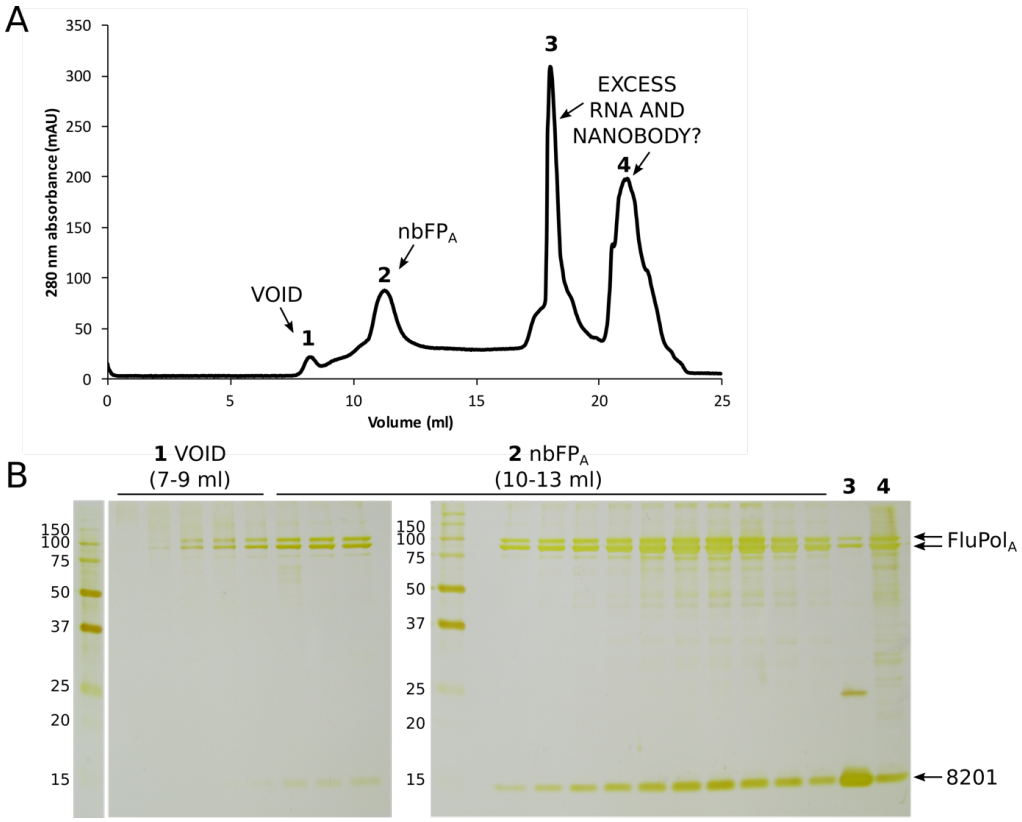


Figure 44 – Purification of nbFP$_A$ by size exclusion chromatography in running buffer containing 25 mM Hepes pH 7.5 and 100 mM ammonium sulphate. A) Gel filtration chromatogram showing 4 distinct peaks. B) Silver stained SDS PA gel of a selection of fractions from gel filtration. 3 and 4 are fractions from the centre of their corresponding peaks.

Lastly, ammonium sulphate was also tested as a substitute to NaCl. Like phosphate groups, sulphate groups also have a stronger negative charge than chloride groups, producing solutions of higher ionic strength at lower concentrations. Buffer exchange into ammonium sulphate after purification of the protein in 500 mM NaCl always resulted in moderate to high levels of aggregation. However, aggregates were intermixed with individual particles. I therefore tested whether I could carry out buffer exchange of the complex into a buffer containing 25 mM Hepes pH 7.5 and 100 mM ammonium sulphate in gel filtration. This was indeed found to be the case. A peak corresponding to a monomer of the complex eluted from the column at the expected size (Peak 2 in Fig. 44a). Interestingly, no distinct dimer peak was observed in this buffer, instead, the monomer peak exhibited a shoulder at the elution volume where the dimer would normally elute. The void volume peak (Peak 1 in Fig. 44a) was also significantly smaller than what had been previously observed using 500 mM NaCl. Instead, two large peaks (labelled 3 and 4 in Fig. 44a) eluted at high volumes which, when analysed by SDS-PAGE, contained mostly excess nanobody as well as some residual polymerase (Fig.

44b). These peaks also had a high absorbance at 260 nm (as measured by NanoDrop), indicating excess unbound RNA eluted here too. nbFP$_A$ purified by this method exhibited significantly less aggregation, likely due to the fact than no buffer exchange was required and the protein had to be concentrated less (only to 0.35 mg/m compared to minimum 1.75 mg/ml that were required when buffer exchange by dilution was carried out). A dataset (described under section 5.5.3) was collected for nbFPA produced by this method to which a small amount of Tween 20 (0.005 %) was also added to address issues regarding particle orientation, which are further detailed in the section below.

5.4.4 Strategies for correcting particle preferential orientations

Generally, in order to obtain high-resolution reconstructions of single particles by cryo-EM, the protein complex in question must acquire random orientations in vitreous ice. However, particles can often acquire preferential orientations on grids due to the way they interact with the air-liquid interface in the very thin liquid layers formed after blotting and before vitrification (Dubochet et al., 1985; Glaeser and Han, 2017; Taylor and Glaeser, 2008), which was found to be the case for nbFP$_A$. The uneven distribution of projections causes a loss of resolution in the direction parallel to the preferential orientation, which distorts or "smears" the resulting density maps (Barth et al., 1989). In the section below (section 5.5) I will discuss issues encountered during data processing regarding preferential orientations of polymerase particles in vitreous ice, and the modifications made to data collection and processing strategies to tackle this issue. For now, I will discuss the screening of buffer compositions, specifically the use of detergents as well as the use carbon supports on grids employed to attempt to improve the distribution of orientations acquired by nbFP$_A$.

Multiple grid types are available for preparation of samples for cryo-EM (Russo and Passmore, 2014). Most of the ones I employed in this project consist of a copper support overlaid with a carbon film with regular arrays of holes (holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing). Once the protein sample is applied and blotted, a very thin layer of liquid remains in the holes, which is then flash-frozen to form vitreous ice. Imaging of small protein complexes in holes is advantageous due to the fact that no additional background signal is obtained from a support. However, some

complications are associated with this approach. For instance, where the ice becomes too thin, the protein can be compressed between the two air-liquid interfaces causing deformation of the particles or segregation of proteins to the rim of the holes, where the ice is thicker. Additionally, protein complexes can become adsorbed to the air-liquid interface resulting in preferential orientations and potentially also denaturation (reviewed in Taylor and Glaeser, 2008). Indeed, nbFP$_A$ was found to display strong preferential orientations when suspended in vitreous ice, which led me to testing other grid types. A common alternative to holey carbon grids are those containing a continuous carbon layer, whether that is ultrathin carbon or graphene oxide (Pantelic et al., 2010). These provide an alternative surface for the protein complexes to adhere to, preventing them from interacting with the air-liquid interface. This results in particles being able to acquire different orientations than they would if they only interacted with the air-liquid interface. On the other hand, despite being extremely thin scaffolds (graphene oxide layers are only 0.34 nm thick), use of a carbon support can increase background signal, which can reduce resolution and cause problems visualising small particles. I initially tested the use of lacey carbon grids coated with ultrathin carbon in combination with the buffers described in section above. Unfortunately, the grids were often very dry or contained large numbers of protein aggregates. I also tested a procedure essentially similar to that employed for negative staining, where the protein in a high salt buffer was applied to the grid, followed by a very gentle blotting step and a wash in water (or 25 mM Hepes pH 7.5) before blotting again and plunging. This eliminates excess salt while the protein is adsorbed onto the carbon surface, in theory preventing aggregation and ensuring good particle distribution. There is precedent for this technique being used for nucleotide-binding protein complexes that require high salt concentrations (Ballandras-Colas et al., 2017). Yet, again, issues with ice thickness and aggregation were encountered in these grids similar to those observed with the protein sample in other buffers. Manually produced graphene oxide coated holey carbon grids were also tested as an alternative to the ultrathin carbon coated grids. However, production of such grids was found to be very challenging, with very uneven graphene oxide coating which made imaging difficult.

Detergents have been widely used to improve the distribution of particles in vitreous ice despite the fact that they can result in artefacts in ice formation during freezing: mainly gradients of ice thickness within grid holes and increased background noise which reduces contrast (Elmlund et al., 2017; Glaeser and Han, 2017). Nevertheless, they are still employed for their ability to alter the surface properties of the thin layer of liquid formed on the grid before plunging into liquid ethane. At concentrations below CMC (Critical Micelle Concentration), detergent molecules do not form micelles but instead accumulate at the air liquid interface with the hydrophobic tails pointing away from the solvent. This effectively shields the air-liquid interface preventing (or at least reducing) its direct interaction with protein complexes in solution. I tested three detergents: n-octyl-β-D-glucoside, octylthioglucoside and Tween 20, at concentrations 10x below CMC, in order to determine whether this would be sufficient to increase the number of views of the complex we were able to acquire. Both n-octyl-β-D-glucoside (β-OG) and octylthioglucoside (OTG) which was used during protein purification) are structurally related, mild, non-ionic detergents. There is precedent for the use of β-OG to improve particle orientation in vitreous ice of membrane proteins (Baker et al., 2015; Zhang et al., 2010). Both β-OG and OTG were tested in buffers containing 150 mM NaCl and 25 mM Hepes pH 7.5, both of which resulted in high levels of aggregation. OTG on the other hand was also tested in a buffer containing 100 mM ammonium sulphate and 25 mM Hepes pH 7.5. This produced lower levels of aggregation and an apparent improvement in the orientations. However, it was addition of Tween 20 to a final concentration of 0.005% to the same buffer (100 mM ammonium sulphate, 25 mM Hepes pH 7.5) that yielded the best results. While small aggregates were still visible, particle distribution was significantly better and consequently, a dataset for this condition was collected, for which further details can be found in section 5.5.3.

Grid type	*[Protein]*	*Buffer*	*Salts*	*Other*	*Outcome*
Holey carbon copper C-flat (40 nm film) (Protochips) 2 µm holes, 1.2 µm spacing	0.2 mg/ml	80 mM Phosphate buffer pH 7.5	75 mM KCl	-	Polymerase accumulates on thick ice, which results in poor contrast
Holey carbon copper C-flat (40 nm film) (Protochips) 2 µm holes, 1.2 µm spacing	0.2 mg/ml	20 mM Phosphate buffer pH 7.5	100 mM NaCl	-	Grainy ice – impossible to identify particles against the background

Holey carbon copper C-flat (40 nm film) (Protochips) 2 μm holes, 1.2 μm spacing	**0.5 mg/ml**	**60 mM Phosphate buffer pH 7.5**	**100 mM KCl**	**-**	**Polymerase clearly visible on some areas of very thin ice. These contained high particle concentrations**
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.3 mg/ml	25 mM Tris HCl pH 8	150 mM NaSCN	-	Particles visible, small amounts of aggregation
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.3 mg/ml	25 mM Hepes pH 7.5	50 mM NaSCN, 50 mM NaCl	-	Particles visible
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.3 mg/ml	25 mM Hepes pH 7.5	Protein was quickly diluted 1/3 in 25 mM Hepes pH 7.5 in order to drop [NaCl] to 167 mM before applying it to the grid	-	Particles were clearly visible but with moderate aggregation. Difficult to reproduce.
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	**0.35 mg/ml double blotted**	**25 mM Hepes pH 7.5**	**Protein was quickly diluted 1/3 in 25 mM Hepes pH 7.5 and 20 mM NaSCN in order to drop [NaCl] to 167 mM before applying it to the grid**	**-**	**Excellent particle distribution and good contrast.**
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.2 mg/ml	25 mM Hepes pH 7.5	100 mM NaCl	20 mM NDSB	Very grainy ice – impossible to identify particles against the background. Other NDSB concentrations were also tested (50 mM, 100 mM, 500 mM), all of which produced very poor ice.
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml double blotted	25 mM Hepes pH 7.5	150 mM NaCl	20 mM glutamate/ 0.07 % *n*-octyl-β-D-glucoside/ 10 mM ATP	Protein aggregation
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml	25 mM Hepes pH 7.5	500 mM NaCl	Water wash	Issues with ice thickness
Ultrathin carbon coated lacey carbon support copper grid	0.35 mg/ml 30 sec incubation	25 mM Hepes pH 7.5	500mM NaCl	25 mM Hepes pH 7.5 wash	Difficult to obtain adequate ice thickness (too dry) and aggregation
Ultrathin carbon coated lacey carbon support copper grid	0.2 mg/ml	25 mM Hepes pH 7.5	200 mM NaCl + 20 mM NaSCN		Very poor particle contrast

Graphene oxide coated holey carbon copper C-flat (40 nm film) (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml	25 mM Hepes pH 7.5	20 mM NaSCN + 150 mM NaCl		Poor graphene oxide coating
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml double blotted	25 mM Hepes pH 7.5	50 mM NaCl	80 mM ammonium sulfate	Partially aggregated, strong preference in particle orientation.
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml	25 mM Hepes pH 7.5	50 mM NaCl, 40 mM ammonium sulfate	0.05 % OTG	Thick ice, aggregation
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	0.35 mg/ml	25 mM Hepes pH 7.5	50 mM NaCl, 40 mM ammonium sulfate	0.005 % Tween 20	Moderate aggregation.
Holey carbon copper C-flat (Protochips) 2 μm holes, 1.2 μm spacing	**0.35 mg/ml**	**25 mM Hepes pH 7.5**	**100 mM ammonium sulfate**	**0.005 % Tween 20**	**Good particle distribution, some aggregation**
Quantifoil holey carbon gold grids 1.2 μm holes, 1.3 μm spacing	0.8 mg/ml	25 mM Hepes pH 7.5	100 mM ammonium sulfate	0.005 % Tween 20	Aggregation, poor ice thickness
Quantifoil holey gold, gold grids 1.2 μm holes, 1.3 μm spacing	0.4 mg/ml	25 mM Hepes pH 7.5	100 mM ammonium sulfate	0.005 % Tween 20	Poor ice thickness

Table 10 – Summary of some of the sample and grid preparation conditions tested during optimisation. Samples for which a full dataset was collected are highlighted in bold.

5.5 Cryo-EM data collection

Data was collected for the nbFP$_A$ complex in conditions that displayed good particle distribution and low amounts of aggregation. After screening, cryo-EM data was collected either on a 300 keV Tecnai Polara microscope (FEI) or a 300 keV Titan Krios microscope (FEI). In this section I will describe the different datasets that were collected and discuss the data processing strategies employed and their outcomes. Micrograph motion correction was carried out either with Motioncorr, for sections 5.5.1 and 5.5.2 or Motioncor2 for sections 5.5.3 and 5.5.4 (Li et al., 2013; Zhang, 2016). All particle picking was carried out using EMAN2 (Tang et al., 2007) and CTF correction was carried out either with CTFFIND3, for sections 5.5.1 and 5.5.2, or GCTF for sections 5.5.3 and 5.5.4 (Mindell and Grigorieff, 2003; Zhang, 2016). Further processing was carried out using the Relion software package (Scheres, 2012).

5.5.1 Dataset 1 - Phosphate buffer

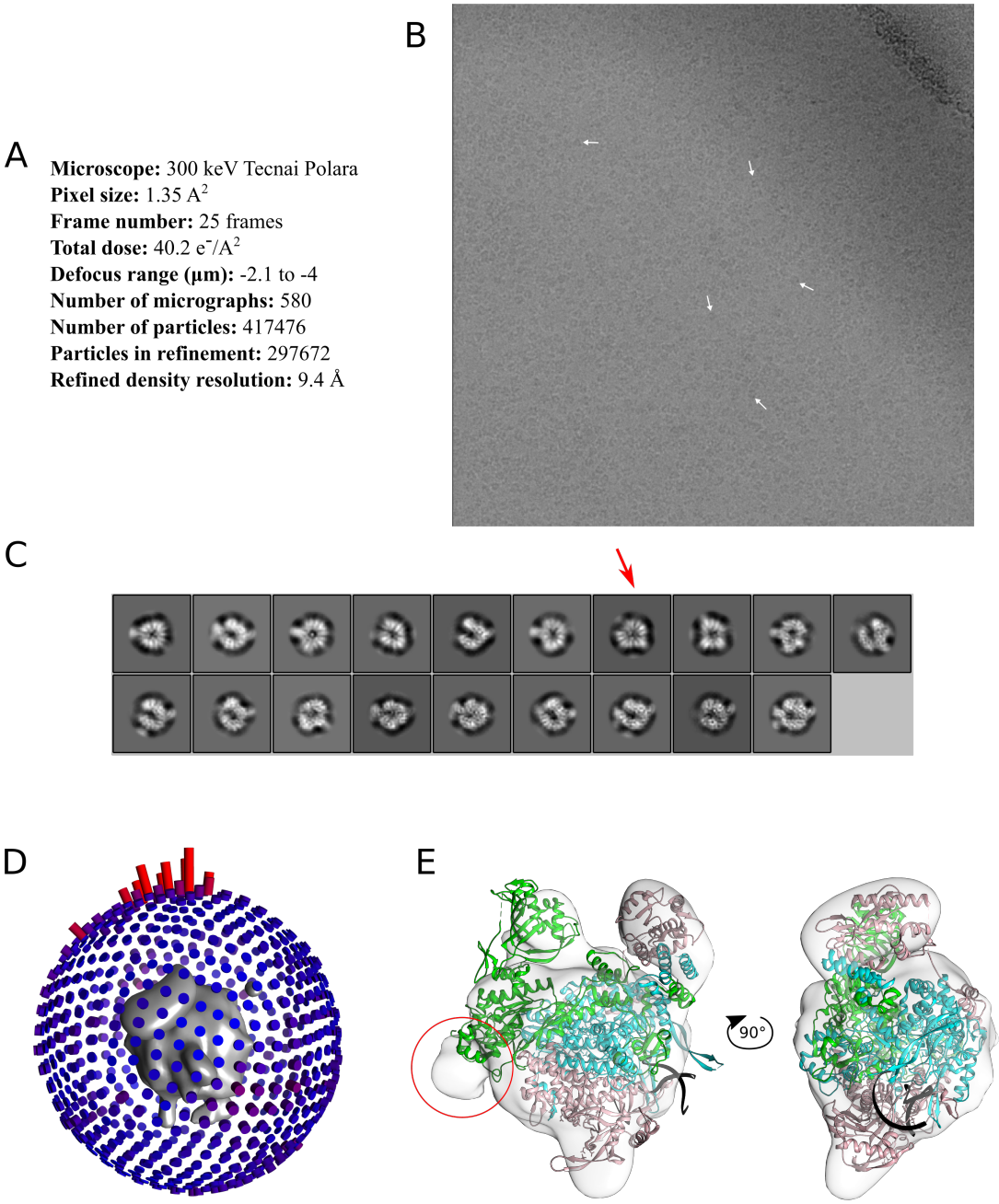


Figure 45 – Cryo-EM dataset for nbFP$_A$ in phosphate buffer. A) Data collection and processing details. Note: resolution estimate is likely highly inaccurate due to particle preferential orientations. B) Motion corrected micrograph. Particles highlighted with white arrows. C) 2D classes selected for 3D classification and refinement. Example of a poorly aligned class highlighted with a red arrow. D) Angular distribution plot of particle orientation. E) Refinement density with bat FluPol$_A$ structure fitted into the density. PDB ID 4wsb. Unfilled density highlighted with a red circle.

In the first dataset collected on this complex, the protein was placed in a phosphate based buffer supplemented with potassium chloride (Table 8). Data collection parameters can be found in Fig. 45b. As can be seen on the micrograph displayed in Fig. 45a, particle concentration in the grid holes was

very high, with particles laying in close proximity to each other. This could be a result of using a slightly higher concentration of protein than was employed for other datasets later on (0.5 mg/ml) but also due to a concentrating effect during "thinning". Thinning describes the time after blotting and before vitrification where the protein resides in a very thin layer of liquid that is subject to fast evaporation (although this is reduced by the use of plunging equipment with a controlled humidity environment). Where the liquid film is very thin, evaporation can cause an artificial concentrating effect. Additionally, if particles adsorb to the air-liquid interface, this may also cause local concentration of the protein (Glaeser and Han, 2017; Taylor and Glaeser, 2008). While these effects, on their own, are not necessarily indicative of an interaction between the protein and the air-liquid interface that is detrimental, in retrospect, they were certainly symptoms of it. No particles were visible in areas of thick ice, likely due to poor contrast of particles against the surrounding background, which was generally found to be quite grainy, except in the areas of very thin ice where data was collected. 2D classification revealed classes with cashew shaped particles which contained an additional small density to the side of the core density (Fig. 45c). The classes appeared relatively well aligned except for a few which displayed star-shaped streaking patterns suggestive of very poor alignment (e.g. class number 7 highlighted with a red arrow in Fig. 45c). Something that stood out in 2D classification was how similar all the classes were. Indeed, upon further 3D classification and refinement, it became apparent that particles were acquiring a single preferential orientation, as can be seen in the angular distribution plot in Fig. 45d. This plot shows there is a very large population of polymerase particles acquiring a single orientation in vitreous ice, with very small populations of what we might call "side views". The resulting unmasked density map (Fig. 45e), which had an estimated resolution of 9.4 Å is therefore unreliable as is this resolution estimate. Nevertheless, I attempted to fit the bat FluPol$_A$ structure (PDB ID 4wsb) into the refined density obtained by simple volume fitting in Chimera. The polymerase structure fit reasonably well in the density map, with a small additional side density which appeared to reside proximal to the PB2 C-terminus. We cannot however draw any conclusions from this map, as it is likely to contain artefacts caused by the particle orientation problem.

As this data was collected from grids where the particles were extremely close together in a layer of very thin ice, we reasoned that the orientation problem could be a consequence of the interactions occurring between nbFP$_A$ particles as well as with the upper and lower air-liquid interfaces (which might contact the protein as a result of the very thin ice). Particles had only been visible in this buffer in areas of very thin ice which meant that collecting data on thicker ice to determine whether this would improve particle orientation was not an option. My focus was therefore to screen buffer conditions to replace the phosphate based buffer for another which would also allow me to reduce the overall NaCl concentration (as described above in section 5.4.2).

5.5.2 Dataset 2 - Sodium thiocyanate

Buffer screening led to the identification of another buffer in which the complex could be imaged. This time, it contained 167 mM NaCl supplemented with 20 mM sodium thiocyanate in a Hepes pH 7.5 buffer (Table 8). Data collection details for this dataset can be found in Fig. 46b. In this buffer, particle distribution was greatly improved, with good spacing between the particles, which were now visible in slightly thicker ice (Fig. 46a). This buffer also produced minimal aggregation, indicating that it had a beneficial effect in terms of keeping the complex stable in solution. 2D classification exhibited what appeared to be a greater variety of views, as judged by the different morphologies of the averaged particles for example, comparing class 1 to class 3 and class 11 (highlighted with green arrows in Fig. 46c). Classes with very poorly aligned particles like those that had been observed in the previous dataset were also apparent in this dataset e.g. class 2 (highlighted with a red arrow). Unfortunately, upon refinement, it became apparent that this dataset did not in fact contain an improved particle orientation distribution (Fig. 46d). There were two main populations of particle orientations only separated by a ~45º tilt. The density map that resulted was extremely distorted in the direction parallel to the preferential views (Fig. 46e) and was hence essentially uninterpretable. As for the previous dataset, I attempted to fit the bat FluPol$_A$ structure (PDB ID 4wsb) into the density, which was possible, but large areas of density appeared to be missing (PB2 C-terminus, PA endonuclease domain and PA C-terminus) and no additional density that could be attributed to a nanobody could be seen, which is not surprising as the density map quality is very poor (Fig. 46e). It therefore became

clear that the skew in particle orientations was unlikely to be due to thin ice or inter-particle interactions but purely due to interaction with the air-liquid interface. I attempted to tackle this issue through screening of grids containing carbon supports as well as the addition of detergents, as described under section 5.4.3.

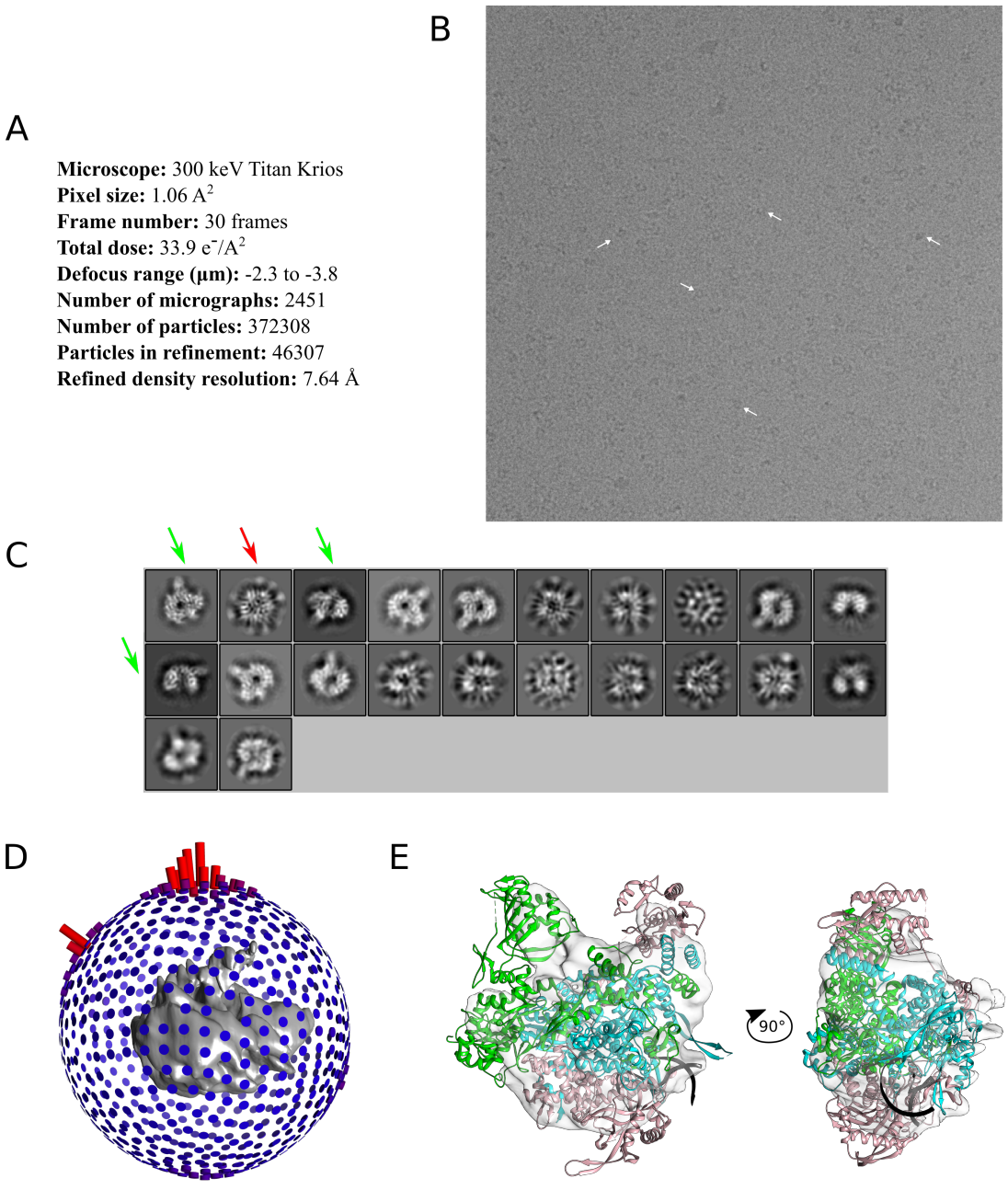


Figure 46 - Cryo-EM dataset for nbFP$_A$ in buffer containing sodium thiocyanate. A) Data collection and processing details. Note: resolution estimate is likely highly inaccurate due to particle preferential orientations. B) Motion corrected micrograph. Particles highlighted with white arrows. C) 2D classes selected for 3D classification and refinement. Example of a poorly aligned micrograph highlighted with a red arrow. Classes possibly exhibiting different views highlighted with green arrows. D) Angular distribution plot of particle orientation. E) Refinement density with bat FluPol$_A$ structure fitted into the density. PDB ID 4wsb.

5.5.3 Dataset 3 - Ammonium sulphate, Tween 20

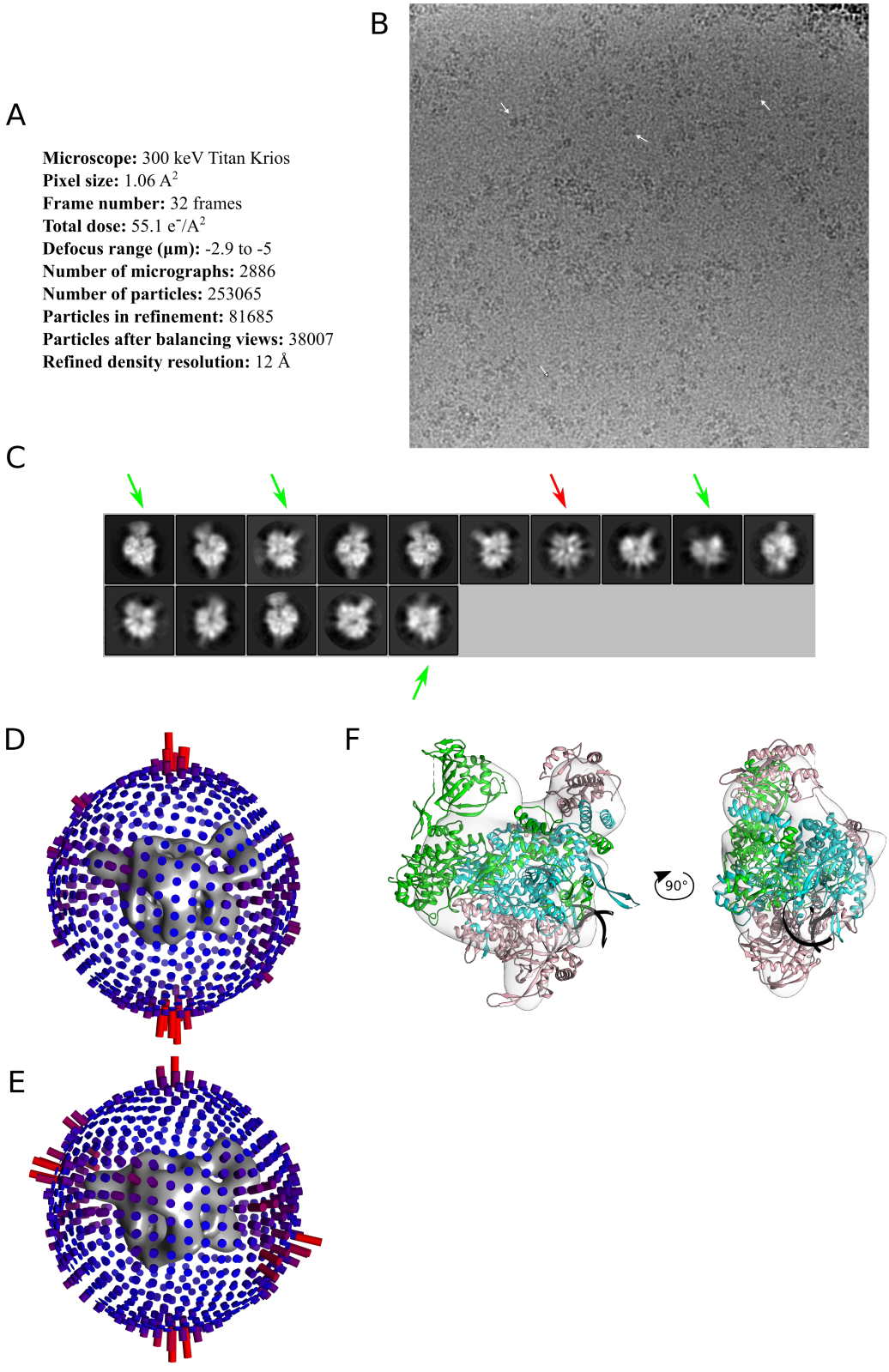


Figure 47 - Cryo-EM dataset for nbFP$_A$ in buffer containing ammonium sulphate and Tween 20. A) Data collection and processing details. Note: resolution estimate is likely highly inaccurate due to particle preferential orientations. B) Motion corrected micrograph. Particles highlighted with white arrows. C) 2D classes selected for 3D classification and refinement. Example of a poorly aligned micrograph highlighted with a red arrow. Classes possibly exhibiting different views highlighted with green arrows. D) Angular distribution plot of particle orientation. E) Angular distribution plot of particle orientation after balancing particle numbers in different views. F) Refinement density with bat FluPol$_A$ structure fitted into the density. PDB ID 4wsb.

Addition of Tween 20 at concentrations 10 times below CMC appeared to have the least detrimental effect on ice morphology and particle aggregation compared to the other detergents and concentrations tested. It was used in conjunction with ammonium sulphate, which was found to be a good substitute for sodium chloride in the buffer, even allowing gel filtration directly in a buffer containing 25 mM Hepes pH 7.5 and 100 mM ammonium sulphate, which meant posterior buffer exchange was not required. A dataset was collected for this condition (Table 8), the details of which can be found in Fig. 47b. As shown in the micrograph displayed in Fig. 47a, the grid on which data was collected contained a moderate amount of diffuse particle aggregation, which greatly limited the number of particles that could be picked per micrograph. Any particles which were closer than 120 Å from each other were discarded, which resulted in a relatively low particle number for the number of micrographs acquired. Nevertheless, 2D classification hinted towards the presence of side views, e.g. comparing classes 1, 3, 9 and 15 (highlighted with green arrows in Fig. 47c). This was confirmed after obtaining the angular distribution plot of the refined density, which showed an improved distribution of orientations (Fig. 47d). In addition to the two views which predominated in the previous two datasets, additional smaller but significant populations were shown to exist in the data. As with the cryo-EM dataset of FluPol$_C$ in complex with promoter vRNA (section 4.4.2), particle numbers in the different views were manually balanced using the particle sorting feature in Relion. This has been shown to drastically improve the resolution in datasets with strong preferential particle orientations (Urnavicius et al., 2015). After this correction, the angular distribution plot showed a more balanced coverage of orientations (Fig. 47e). However, this distribution is still far from ideal, with large areas still lacking coverage, which means the density obtained is still unreliable. The refined density was not obviously distorted and the vRNA bound bat FluPol$_A$ structure could be comfortably fit into the density. Upon inspection of the refined density, no additional densities that could be attributed to the nanobody were found. This is of course a matter of concern, as it could indicate a variety of problems: 1- the particle orientation distribution is still so poor that the map is distorted and unreliable or heavily influenced by the low resolution reference model used during 3D classification and refinement, 2- the nanobody dissociates from FluPol$_A$ in the grid environment, potentially as a result of its interaction with the air-liquid interface or 3- addition of detergent causes dissociation of the nanobody. All our functional and biochemical data on the interaction between nanobody 8201 and FluPol$_A$ point towards

a strong, stable interaction. It seems unlikely therefore that the very low concentrations of detergent employed here (only 0.005%) would disrupt the interaction. However, we certainly have issues regarding the interaction of the complex with the air-liquid interface and the angular distribution shows improved but still poor coverage, which indicated this issue required further work to be resolved.

5.5.4 Data collection on a tilted sample

An alternative to obtain "side views" of protein complexes in single particle cryo-EM is to physically tilt the sample in the microscope (Fig. 48e) (Tan et al., 2017). Data collection for a tilted grid occurs essentially in the same manner as that of a regular single particle dataset. There is no range in tilt as there would be in data acquisition for tomography. Instead, all that changes is that the goniometer on the microscope stage is set to a fixed tilt for the entire data collection. Data from tilted and un-tilted datasets can then be pooled to produce a larger dataset with better coverage of particle views. At low tilt angles (generally below 40º), particles are still visible in vitreous ice, although contrast may be poor as the beam effectively has to travel through a thicker sample when the grid is tilted (Fig 48d). This technique is best suited for samples where some coverage of side views can be obtained, rather than a single preferential view (as observed in the first two datasets described here). For this reason, I collected a dataset at a 30º tilt for the complex purified in the ammonium sulphate and Tween 20 containing buffer. Details on data collection and processing can be found in Fig. 48b. One of the challenges of working with a tilted datset is CTF correction. As the grid is tilted, particles within the same micrograph have different defocus values and consequently different CTF values. This issue was overcome with the use of GCTF software for CTF correction, which can calculate a CTF value for the individual particles within micrographs (Zhang, 2016).

The grid on which data was collected had a lower protein concentration than that of the previous dataset. This was selected on purpose as this ensured there were less diffuse aggregates which made particle picking for the previous dataset challenging. Nevertheless, identifying particles in vitreous ice was not straightforward. As can be seen in Fig. 48a, the contrast of particles against the background

was quite poor. This was probably due to a combination of slightly "grainy" ice (which is invariably present in samples in this buffer), in addition to the fact that the beam has to travel through a thicker layer of vitreous ice in a tilted grid. nbFP$_A$ particles are also relatively small (just under 300 kDa) and were always challenging to identify against the background in all the datasets collected (but especially in this one). Accordingly, in 2D classification the classes exhibit extremely poor alignment (Fig. 48c), which unfortunately meant this data could not be further processed.

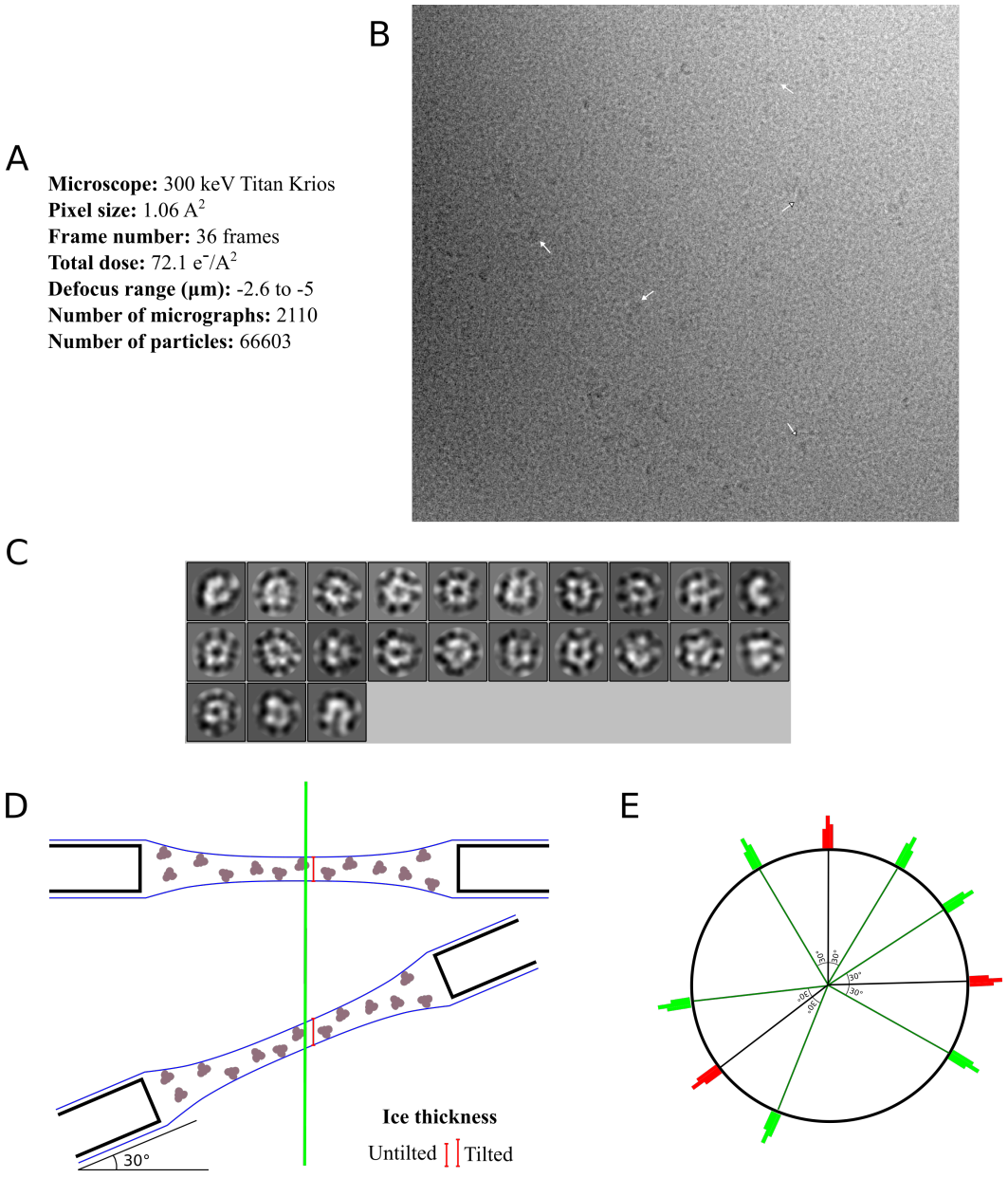


Figure 48 – Tilted cryo-EM dataset for nbFP$_A$ in buffer containing ammonium sulphate and Tween 20. A) Motion corrected micrograph. Particles highlighted with white arrows. B) Data collection and processing details. C) 2D classification. D) Diagrammatic representation of sample on a grid hole during tilting. The difference in distance travelled by the beam in the sample is highlighted. Because of tilting, the beam effectively encounters "thicker" ice, which can negatively affect contrast. E) Diagram showing how tilting increases the number of side views. For every population of a particular orientation in the untilted dataset (red), multiple views are obtained in the tilted dataset (green) around it.

5.5.5 Conclusions and future work

Despite efforts to optimise buffer compositions and data acquisition for the nbFP$_A$ complex, the way it interacts with the air-liquid interface on the grid has prevented us from determining the conformation of the polymerase or the site at which inhibitory nanobody 8201 binds. Nevertheless, there is a large range of approaches that can be tried in order to overcome particle orientation, contrast and protein stability problems. In terms of further optimising buffer compositions, it appears as though the addition of Tween 20 offered a significant improvement upon particle distribution by shielding the air-water interface. However, high background likely caused by other buffer components (like phosphate buffer or ammonium sulphate) would need to be improved. The lowest background was probably obtained with the buffer containing sodium thiocyanate, therefore perhaps a buffer containing it in addition to Tween 20 could produce better grids for data collection. There are alternative methods of shielding the air-water interface beyond the use of detergents. A more "protein-friendly" method has been described where proteins are adsorbed onto a lipid monolayer via an affinity tag, which not only prevents the protein from reaching the air-water interface but also provides a way to carry out on-grid protein purification (Kelly et al., 2008).

Alternatively, further screening of grid preparation conditions could be carried out, particularly of those containing carbon supports. As mentioned, protein complexes can then adhere onto the carbon instead of to the air-water interface, preventing preferential orientations. On the other hand, it is yet to be determined how adhering to a carbon support can affect protein complexes, and use of carbon supports has the disadvantage of increasing background signal (Glaeser and Han, 2017). This would certainly not be desirable for analysis of the nbFP$_A$ complex, as poor contrast is already an issue, even when imaging in holey carbon. It would therefore perhaps be beneficial to optimise the use of holey carbon grids to prevent additional background. Normally, prior to applying the protein sample to the grid, the grid is glow discharged in a plasma cleaner in order to make it hydrophilic. This process makes the surface of the grid negatively charged. However, proteins can be affected by this negative charge and will often preferentially attach to the carbon film rather than enter the holes. This was observed during data collection for nbFP$_A$, which is not surprising as, being a nucleotide binding

protein, it contains highly basic regions. Glow discharging the grids in presence of pentylamine makes them acquire a positive charge, indeed, this technique was initially developed in order to allow nucleic acids (negatively charged) to adhere to EM grids (Dubochet et al., 1971; Miyazawa et al., 1999). Alternatively, one can treat grids with poly-L-lysine prior to incubation with the protein. Having a positively charged carbon film can modify the way particles distribute in the liquid and can encourage particles to acquire different orientations. Lastly, in terms of optimising the choice of grids for analysis of the complex, it would perhaps be beneficial to use gold grids overlaid with a holey gold film. The use of these was briefly tested during this project (Table 8), but further screening of grid preparation conditions (particularly the duration of glow discharge) will be required to obtain good ice thickness and particle distribution. The benefit of using these grids is twofold. First, gold is less reactive than copper and is therefore less likely to interact with the protein of interest applied to the grid. Secondly, and more importantly, grids made entirely out of gold are significantly more stable under the electron beam than copper-carbon grids, inducing less motion, which can improve the resolution obtained (Russo and Passmore, 2014). This would be particularly beneficial if further tilted datasets were to be collected, as greater motion tends to occur on tilted grids, likely due to the uneven way in which the beam interacts with the grid.

nbFP$_A$ particles tend to be relatively poorly defined, even in negative stained samples (Fig. 43c), and result in poor signal to noise ratios that make data processing difficult. This was particularly obvious in the extremely misaligned 2D classes of the tilted dataset, but could also be seen in several classes of the other three datasets discussed (red arrows in Fig. 45c, 46c and 47c). These poorly aligned classes have to be discarded from further data processing, which results in a large number of particles being lost, which can be detrimental to the resolution obtained. The constant presence of these classes that contain large numbers of particles but are not aligned could also be due to the presence of certain views of the polymerase which are essentially featureless (they basically look circular). In combination with poor contrast this would make alignment extremely challenging for data processing software. In terms of data acquisition, poor particle contrast could be tackled by data collection using a

Volta phase plate (Danev et al., 2014). These enable data collection at focus by introducing a phase shift which provides contrast to the sample.

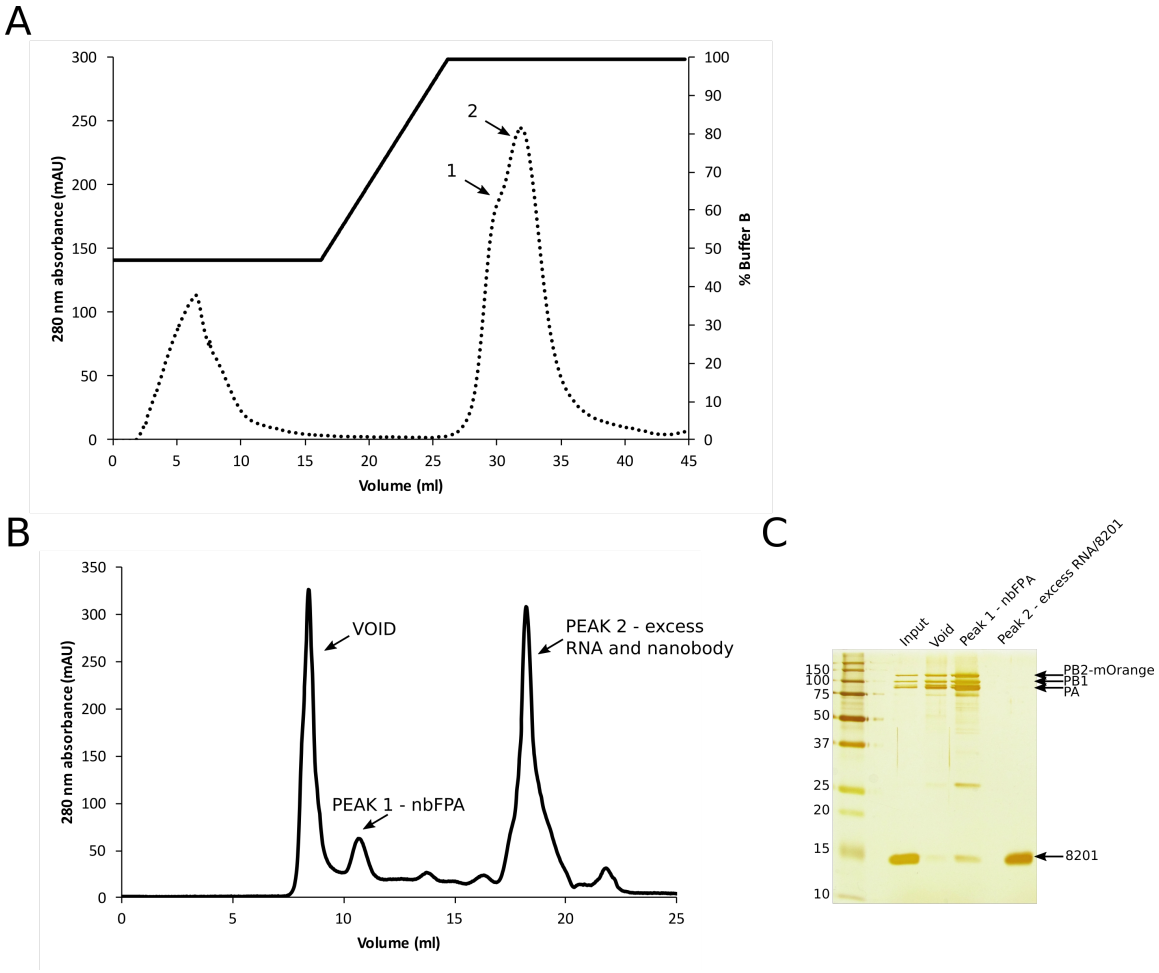


Figure 49 – Purification of mOrange-tagged FluPol$_{NT60}$ in complex with promoter vRNA and nanobody 8201. A) Heparin affinity purification profile showing 2 highly overlapping peaks. B) Gel filtration chromatogram of nbFP$_A$. C) Silver-stained SDS-PA gel of peak fractions from gel filtration.

Alternatively, one could address this issue by modifying the complex itself. One option is to partially cross-link the complex in order to stabilise it while simultaneously modifying its surface properties, which would also address the issue of adherence to the air-water interface. Implementation of the GraFix protocol seems particularly attractive for this purpose, as it has been well characterised for preparation of samples for cryo-EM and allows mild cross-linking conditions on a glycerol/sucrose gradient (Kastner et al., 2007; Stark, 2010). I briefly attempted this protocol for the nbFP$_A$ complex but the choice of gradient and glutaraldehyde concentration will require optimisation to prevent aggregation of the complex. Another option is the addition of protein tags to the protein complex: this would add mass to the particles, potentially making them easier to identify on micrographs. If the tag

is large enough, it would also have the potential of improving particle alignment (as long as it is not highly flexible) or even shielding the surface that tends to adhere to the air-water interface. In this line, I started to analyse the possibility of employing a FluPol$_{NT60}$ construct for expression in baculovirus where the PB2 C-terminus is tagged with mOrange, a 26 kDa globular protein, in addition to the protein A tag used for purification. This mOrange tag resides to the N-terminus of the protein A tag and the TEV cleavage site and therefore remains attached to PB2 after affinity purification. Initial analysis of this complex indicated that it behaved in an identical manner to untagged FluPol$_A$ in terms of its behaviour in heparin affinity purification (Fig. 49a) and binding to nanobody 8201. Nanobody 8201 co-eluted with FluPol$_A$-mOrange in gel filtration (Fig. 49b, c), which suggests the formation of a stable complex. Use of this construct therefore appears to be a viable option to attempt to improve particle contrast and alignment issues. An interesting observation on FluPol$_A$-mOrange is the absence of a dimer peak eluting from size exclusion chromatography, unlike that observed for untagged FluPol$_A$ (Fig. 43a), which suggests the dimer interface is disrupted by the presence of the tag at the PB2 C terminus.

Overall, multiple viable options remain that can be tested in order to solve the issues that have so far prevented us from obtaining a high-resolution structure of the nbFP$_A$ complex. The field of single particle cryo-EM is rapidly expanding and new methods are constantly becoming available which could also be applied in future, including novel methods for application of protein to the grid and scaffolds that regulate ice thickness and control protein orientation (Arnold et al., 2017; Coscia et al., 2016; Martin et al., 2016).

5.6 Summary

By screening a panel of 24 llama nanobodies we have determined that these are an excellent tool for modulation of FluPol activity. A range of interesting phenotypes, from selective inhibition of replication to inhibition of all polymerase activity were found, which will have multiple applications. As binding of nanobodies to their target epitopes relies only on three hypervariable regions,

particularly the CDR3, the interaction surface should be reasonably small. That means that structural characterisation of nanobodies that inhibit polymerase activity has the potential of uncovering sites on the polymerase that can be targeted by small molecule inhibitors. Beyond their therapeutic potential, these nanobodies could be a great tool for analysis of other polymerase-related processes, such as dimerization of polymerase (which we observed in FluPol$_A$), nuclear import and association with other viral or host proteins. Where a nanobody is found to inhibit any one of these functions, the location of its binding site would offer crucial information on the regions or conformations of the polymerase important to that process. This of course, would require a way to easily structurally characterise nanobody-FluPol complexes. In order to develop a method that would enable analysis of such complexes, I attempted to structurally characterise the binding of nanobody 8201 to FluPol. Nanobody 8201 caused strong inhibition of all polymerase activity both in *in vitro* activity assays and in RNP reconstitutions. Unfortunately, issues relating to signal to noise ratios and preferential orientations have prevented us from solving the structure of this complex. Even though use of cryo-EM has not yet yielded the desired results, multiple options remain to be explored in terms of improving sample preparation and data collection. It is likely that a combination of these techniques will eventually allow analysis of FluPol$_A$ in complex with nanobodies, RNA or other factors in a more high-throughput manner than has been previously possible by crystallography. The main advantage of using cryo-EM that has so far become apparent in this project is the possibility of working with very low sample concentrations around 0.3 mg/ml. With the current yields of nbFP$_A$ complex, its structural characterisation by crystallography would have been highly challenging and would have required upscaling or further optimisation of the expression and purification protocol. As a way of obtaining preliminary data that would allow us to roughly determine where nanobodies are binding, nanobody-FluPol$_A$ complexes could be analysed by negative stain EM, which should provide a low-resolution model to further guide our cryo-EM efforts.

6 Nuclear import of PB1-PA by RanBP5

6.1 Introduction

RanBP5 is a member of the importin β superfamily, which consists of over 20 members in humans (Harel and Forbes, 2004). As briefly discussed under section 1.3.6, RanBP5, like other importin β family members, doesn't require an importin α adaptor to mediate interaction with its cargo. Instead, RanBP5 directly binds an arginine-rich NLS on the cargo (in contrast with NLSs in the classical import pathway which are usually lysine-rich) (Palmeri and Malim, 1999). Release of RanBP5 bound cargo is caused by interaction of RanBP5 with RanGTP in the nucleus of the cell. The gradient of RanGDP to RanGTP across the nuclear envelope is maintained by the chromatin associated Ran GEF (Ran GTP exchange factor) in the nucleus, which catalyses the exchange of GDP for GTP in Ran, in conjunction with RanGAP (Ran GTPase activating protein) in the cytoplasm, which stimulates Ran to hydrolyse GTP (Clarke and Zhang, 2001). In an uninfected cell, RanBP5 is involved in the nuclear import of highly abundant proteins such as ribosomal subunits and core histones (Baake et al., 2001; Jakel and Gorlich, 1998). It has also been proposed to have an additional role as a chaperone, more specifically, shielding exposed basic regions on proteins which would otherwise interact non-specifically with cytoplasmic polyanions (Harel and Forbes, 2004; Jakel et al., 2002).

However, during influenza infection, this protein is hijacked for the import of the polymerase subunit heterodimer PB1-PA, which is too large to freely diffuse through nuclear pore complexes, the threshold being around 20-30 kDa. Even though PA has an NLS, import appears to occur in conjunction with PB1 via a bipartite NLS on the PB1 flexible β-ribbon: residues 188–189 (NLS1) and 208–209 (NLS2). However, import of PB1 alone was found to be inefficient. Now that full-length FluPol structures are available, we can hypothesise this is likely due to misfolding of PB1 in absence of PA, as these polymerase subunits interact very closely in the heterotrimer (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). Mutations of the PB1 NLS recognised by RanBP5 reduce

import of PB1-PA and attenuate viral growth. Knockdown of RanBP5 results in reduced accumulation of viral RNAs as transcription and replication are compromised due to poor import of the polymerase subunits. These observations therefore indicate that this import pathway is key for efficient viral replication (Deng et al., 2006; Hutchinson et al., 2011). Once the PB1-PA dimer is imported into the nucleus by RanBP5, the full heterotrimer is formed by association with PB2, which is imported via the classical import pathway (Deng et al., 2005; Huet et al., 2010; Hutchinson and Fodor, 2012).

The structure of RanBP5 has not yet been characterised, however, structures of its yeast homolog Pse1p (30% sequence identity over 1027 residues with RanBP5) are available in its apo form as well as RanGTP and cargo NLS-bound states (Kobayashi and Matsuura, 2013). Pse1p acquires an extended right-handed solenoid-like structure made up of HEAT and Armadillo repeats, with the RanGTP and cargo binding sites located in the inner groove of the solenoid. Residues D353, E396 and D438 within the groove were identified in the structure as key to NLS binding. Comparison of the RanGTP and NLS bound structures (Fig. 50) revealed that RanGTP binds at a site partially overlapping with the cargo binding site. Therefore, binding of RanGTP in the nucleus would directly cause displacement and release of the cargo. In addition to that, RanGTP binding also induces a conformational change in Pse1p that is incompatible with NLS binding (Kobayashi and Matsuura, 2013).

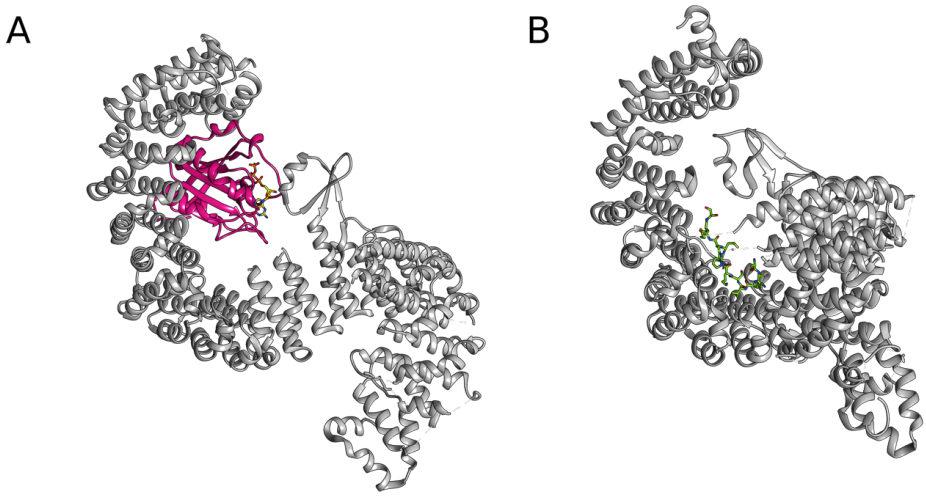


Figure 50 – Structure comparison of RanGTP and NLS bound Pse1p. A) Structure of Pse1p bound to RanGTP. Pse1p shown in light grey, RanGTP in magenta, GTP in yellow. PDB ID 3w3z. B) Structure of Pse1p bound to NLS peptide. Pse1p shown in light grey, NLS peptide in light green. PDB ID 3w3w.

While the interaction between RanBP5 and PB1-PA is clearly key for influenza virus replication, no structural data is available for the complex. My aim in this project was therefore to develop a method for reconstitution of this complex that would provide sufficient yield and purity to enable its structural characterisation, whether by crystallography or cryo-EM.

6.2 Production of baculovirus for co-expression of PB1-PA-RanBP5

Given the success of employing a baculovirus expression system to produce FluPol heterotrimeric complexes, this became the method of choice in our initial attempts at reconstituting a PB1-PA-RanBP5 complex. As RanBP5 is thought to interact exclusively with the PB1 NLS, two viruses were produced, one which would co-express only PB1 and RanBP5 and another which would co-express PB1, PA and RanBP5. Preparation of the baculoviruses was carried out following the MultiBac protocol (Bieniossek et al., 2008). A pUCDM donor plasmid containing genes for influenza A/Northern Territory/60/1968 (H3N2) PB1 and PA expression (obtained from Narin Hengrung) was modified to introduce a protein A tag at the C-terminus of PB1 separated by a TEV cleavage site. A RanBP5 synthetic gene with optimised codons for expression in insect cells was obtained and cloned into a pFL acceptor plasmid (Fig. 51a). Three RanBP5 isoforms exist in human cells, which vary at their N-terminus. Isoform 2 is generally considered the “canonical” sequence, with isoform 1 being a truncated form, 60 amino acids shorter, and isoform 3 being 19 residues longer than isoform 2. As we had no prior information regarding the involvement of the N-terminus of RanBP5 in the interaction with the PB1-PA dimer, we decided to use isoform 3, in order to avoid removing residues which may be necessary.

pUCDM and pFL contain LoxP sites which allow the fusion of donor and acceptor plasmids by Cre-LoxP recombination to produce a multi-gene vector containing all the genes for co-expression (Fig. 51b). This multi-gene vector contains Tn7 transposition sites which are originally encoded in the pFL acceptor plasmid. This allows the integration of the multigene vector into a MultiBac baculoviral genome. This is carried out by electroporation of the multi-gene vector into DH10MultiBac *E. coli*, which harbour the baculoviral genome as well as Tn7 transposition factors. Successful integration is

detected by blue-white selection as insertion of the sequence in the multi-gene vector between the Tn7 sites disrupts a *lacZ* gene in the MultiBac baculoviral genome. This produces a baculoviral genome that contains all the desired genes (Fig. 51c) which can be isolated and transfected into insect cells. The virus can then be recovered from the supernatant of infected cells.

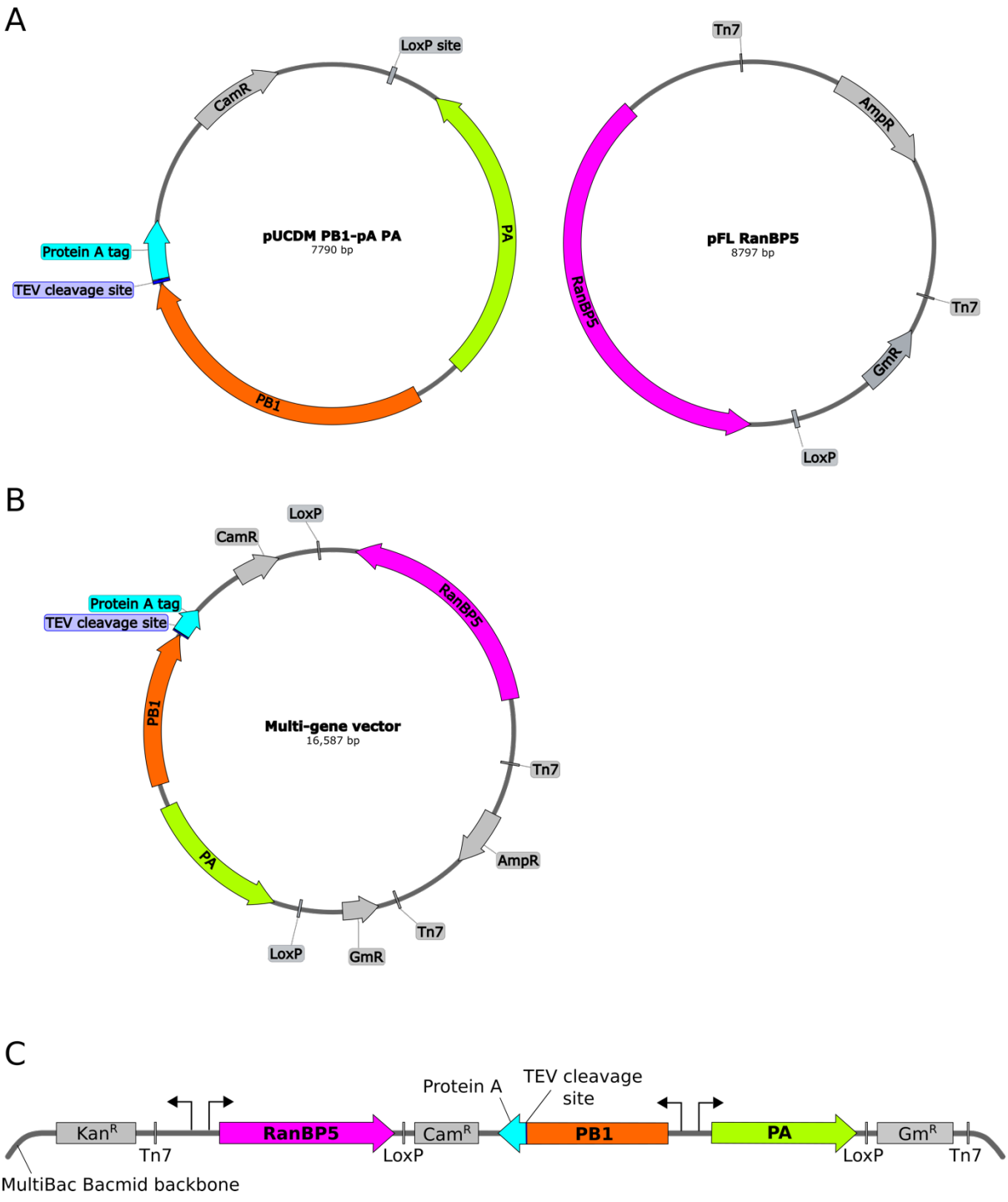


Figure 51 – Cloning strategy for the co-expression of RanBP5, PB1 and PA in baculovirus infected insect cells. A) Donor (pUCDM) and acceptor (pFL) plasmids encoding protein A tagged PB1, PA and RanBP5 respectively. B) Multi-gene transfer vector formed after fusion of the donor and acceptor plasmids by Cre-LoxP recombination. C) Schematic of the gene arrangement in the baculovirus genome after insertion of the multi-gene vector into it by Tn7 transposition.

Both baculoviruses were successfully produced, one for expression of PB1-RanBP5 and one for expression of PB1-PA-RanBP5. For expression, Sf9 insect cells were infected with enough virus to achieve an MOI of approximately 1, which were then harvested 72 hours post-infection. Initial expression and solubility of the overexpressed proteins was tested by western blotting with α-FluPol and α-RanBP5 antibodies (Fig. 52). These showed robust expression of PB1 and PA when they were co-expressed, however, when co-expressed only with RanBP5, PB1 levels were lower. Good levels of RanBP5 could also be detected in the lysate. Cells were lysed following the protocol used for cell lysis in $FluPol_A$/$FluPol_C$ purifications (in a buffer containing 500 mM NaCl). All three proteins could be found in the soluble fraction after lysis and centrifugation, which indicated that a significant proportion of the protein was stable in solution and not aggregating during cell lysis. Note that a band of slightly faster migration on PAGE is visible in the α-RanBP5 western blot; this is likely protein A tagged PB1. As the protein A tag binds IgG, tagged PB1 is detected regardless of the antibody used.

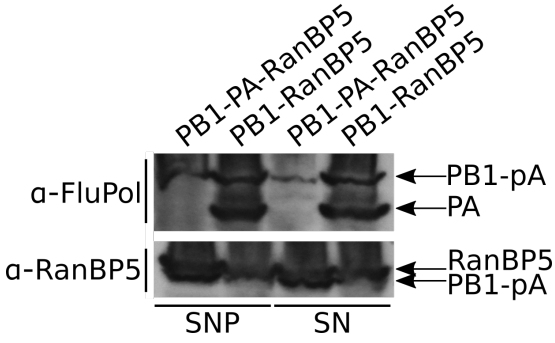


Figure 52 – Western blot against FluPol and RanBP5 on the lysate (SNP) and soluble fraction (SN) of baculovirus infected insect cells. Cells were infected with either PB1-PA-RanBP5 or PB1-RanBP5 baculoviruses. Antibodies: primary rabbit α-FluPol (Eurogentec Ltd), primary mouse α-RanBP5 3(Santa Cruz Biotechnology), secondary α-rabbit IgG or α-mouse IgG HRP-conjugated antibodies.

6.3 Purification of RanBP5-PB1-PA

6.3.1 Purification of co-expressed RanBP5-PB1-PA

As the complex components appeared to be well expressed and soluble, I proceeded to attempt to purify the complex from insect cells. Given that for the purification of FluPol a high concentration of salt is required to keep the protein from aggregating, I initially followed the same protocol for PB1-RanBP5 and PB1-PA-RanBP5 (for details of the FluPol purification protocol see section 2.5.1). I first tested whether the dimeric/heterotrimeric complexes could be recovered after affinity chromatography on IgG beads, which were bound via the protein A tag on the PB1 C terminus. The bound protein was eluted from the beads by cleavage with TEV protease. Samples were collected throughout the

purification for analysis by SDS-PAGE and western blotting. The elution sample for the PB1-RanBP5 expressing baculovirus exhibited a very faint PB1 band with levels of RanBP5 being co-purified that were barely detectable. On the other hand, the purification from PB1-PA-RanBP5 baculovirus infected cells produced higher amounts of the PB1 and PA (molecular weight 86.5 kDa and 83 kDa respectively), which are clearly visible with small amounts of RanBP5 (molecular weight 125.5 kDa) (Fig. 53a). While it was encouraging to be able to obtain the PB1-PA dimer, RanBP5 was not present in equimolar quantities. Analysis of the samples collected during purification by western blotting confirmed that the complex components were soluble after lysis but revealed that large quantities of RanBP5 were being lost in the flow through and the first wash of the IgG beads (Fig. 53b). Indeed, very little RanBP5 bound to the beads (note that the main band that can be seen in the "BB" and "BA" lanes is PB1-pA with only a very faint RanBP5 band above it). Accordingly, very little RanBP5 could be detected in the elution, particularly in proportion to how much PB1-PA dimer could be purified. The PB1-RanBP5 interaction is thought to be a one-to-one interaction, not involving multiple PB1 subunits per RanBP5. The observation that a large amount of RanBP5 was being lost throughout the purification suggests that under the conditions employed for this purification the PB1-PA interaction with RanBP5 was not stable.

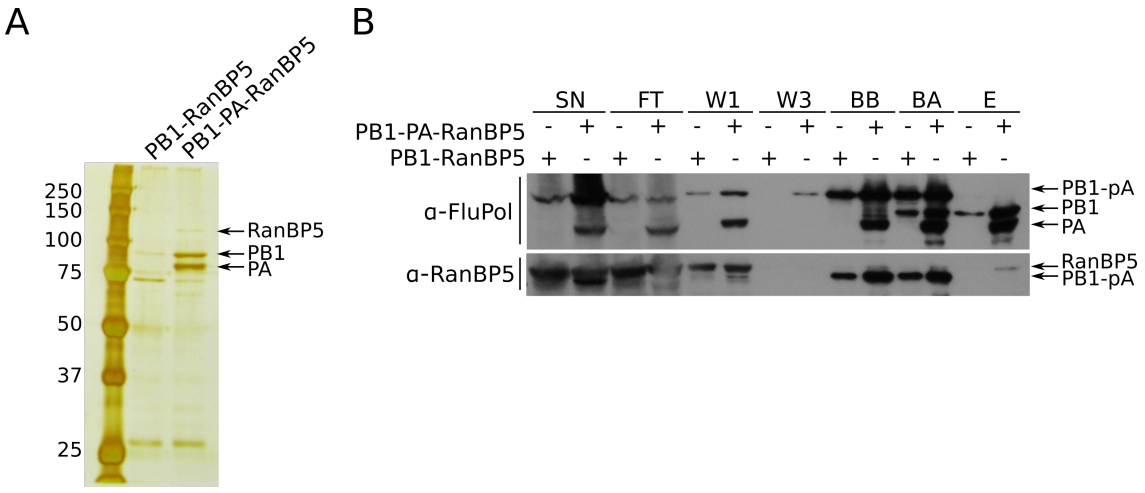


Figure 53 – Affinity purification on IgG beads of co-expressed PB1-PA-RanBP5 and PB1-RanBP5. A) Silver stained SDS-PA gel of elution sample after TEV cleavage. B) Western blot against FluPol and RanBP5 on the soluble lysate fraction (SN), flow through (FT), wash 1 (W1), wash 3 (W3), protein bound to beads before TEV cleavage (BB), beads after TEV cleavage (BA) and cleaved fraction eluted from the beads (E). Antibodies: primary rabbit α-FluPol (Eurogentec Ltd), primary mouse α-RanBP5 (Santa Cruz Biotechnology), secondary α-rabbit IgG or α-mouse IgG HRP-conjugated antibodies.

The fact that the NLS consists of positively charged amino acids and the key residues in the NLS binding site on Pse1p were, not surprisingly, found to be mostly negatively charged residues, suggests that the PB1 NLS - RanBP5 interaction is largely electrostatic (Kobayashi and Matsuura, 2013). We therefore hypothesised that the use of a buffer with high ionic strength (500 mM NaCl) could be disrupting the formation of the complex. I hence carried out the same affinity purification on IgG beads comparing the use of lysis, wash and cleavage buffers which contained 500 mM NaCl versus 150 mM NaCl. The elution sample obtained from the PB1-PA-RanBP5 purification showed a dramatic improvement in the ratio of RanBP5 to PB1-PA in the buffer containing 150 mM NaCl (Fig. 54a). An improvement was also observed in the PB1-RanBP5 complex purification although, yet again, the overall protein quantities were significantly lower than those obtained where PA was also co-expressed. Analysis of the purification samples by western blotting revealed that less RanBP5 was being lost in the flow through where a lower NaCl concentration was employed in the buffer (comparing lanes 5 and 6 in Fig. 54b). More RanBP5 was also detected associated with the beads before TEV cleavage (lanes 23 and 24 in Fig. 54b) and consequently also in the elution (lanes 15 and 16 in Fig. 54b). Interestingly, levels of PB1-PA were unaffected by the drop in NaCl concentration in the buffer, purifying at equal levels in both buffer conditions (lanes 15 and 16, upper panel in Fig. 54b). This suggests that, unlike FluPol, the PB1-PA dimer does not require high ionic strength to remain in solution.

Even though the yields of the complex were very low, I tested whether it would remain stable when further purified by size exclusion chromatography in a running buffer containing 25 mM Hepes pH 7.5 and 150 mM NaCl (Fig. 54c). While the chromatogram is quite noisy, likely due to the low quantities of the complex and the presence of contaminating proteins (which can be seen in Fig. 54a), two distinct peaks eluted in the range where a PB1-PA-RanBP5 complex, or its components, would be expected to elute. Fractions from these peaks were analysed by SDS-PAGE, which showed that peak 1 contained the PB1-PA-RanBP5 complex, which displayed equimolar ratios of its components (Fig. 54d). Peak 2 consisted mostly of RanBP5 with small quantities of PB1 and PA. These data suggest

that under low salt conditions the PB1-PA-RanBP5 complex has improved stability. Nevertheless, the complex dissociated partially during gel filtration and overall yields remained very low.

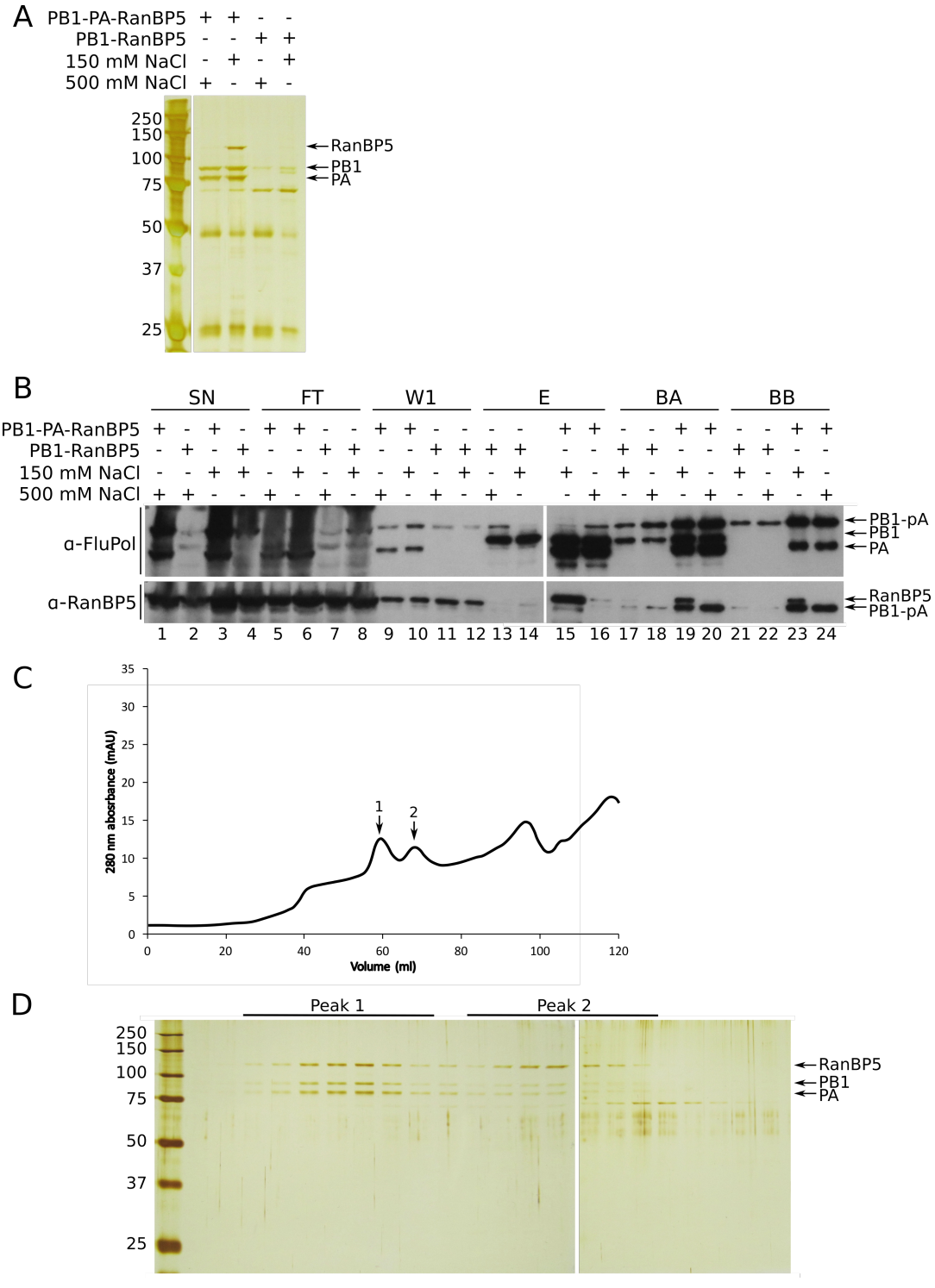


Figure 54 – The effect of NaCl concentration on the RanBP5-PB1-PA interaction. A) Silver stained SDS-PA gel of protein cleaved off IgG beads in buffers containing 150 mM NaCl versus 500 mM NaCl. B) Western blot against FluPol and RanBP5 analysing purification samples. Soluble lysate fraction (SN), flow through (FT), wash 1 (W1), wash 3 (W3), protein bound to beads before TEV cleavage (BB), beads after TEV cleavage (BA) and cleaved fraction eluted from the beads (E). Antibodies: primary rabbit α-FluPol (Eurogentec Ltd), primary mouse α-RanBP5 (Santa Cruz Biotechnology), secondary α-rabbit IgG or α-mouse IgG HRP-conjugated antibodies. C) Gel filtration profile of the RanBP5-PB1-PA elution sample in buffer containing 150 mM NaCl. D) Silver stained SDS-PA gels of fractions from gel filtration depicted in C).

I therefore tested whether modification of other buffer components could improve the quantity of the PB1-PA-RanBP5 we obtained, namely, the use of detergents, glycerol, glutathione and variations in pH. As expression of PB1-RanBP5 had been consistently poor compared to PB1-PA-RanBP5, only the PB1-PA-RanBP5 complex was employed in these assays. Fig. 55a shows the elution samples from several small-scale affinity purifications on IgG coated beads. The buffer composition was kept as before (with 150 mM NaCl) except for the specified components. Variations in pH appeared to have little effect on the total amounts of protein obtained and did not affect the ratios of the PB1-PA-RanBP5. This last observation is not surprising for an interaction that is mostly reliant on basic/acidic residues, as the charge of these remains the same in the pH range rested (pKa values are beyond the range tested). Removal of glycerol did not have a significant effect either. However, removal of glutathione (present in cleavage buffer previously) appeared to improve the quantities of RanBP5 which co-precipitated with PB1-PA but, overall, it didn't have a significant effect on total yields of the complex. Surprisingly, lowering the NaCl concentration below 150 mM had no deleterious effects on total yield and retaining very good levels of PB1-PA and RanBP5. The addition of different detergents during purification (Fig. 55c) had no effect on protein solubility or yields. However, it became apparent that detergent is required during cleavage of the complex from IgG beads, as the control sample in absence of detergent showed almost no RanBP5 eluting, possibly indicating precipitation (which would be removed during centrifugation to sediment the beads before collecting the elution sample).

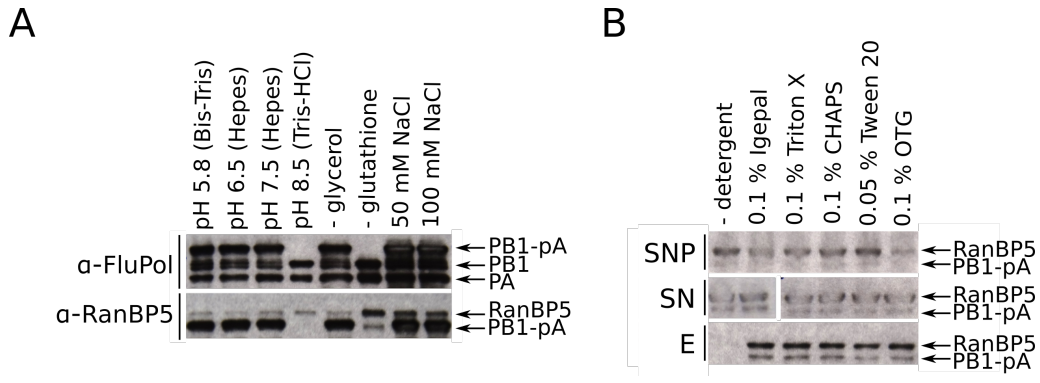


Figure 55 – Screening of purification conditions to improve yields of PB1-PA-RanBP5. A) Western blot against FluPol and RanBP5 of elution samples from IgG beads. Samples purified in varying buffer conditions, lower NaCl concentrations or in absence of glutathione/glycerol. Note: these elutions display some contamination with PB1-pA, likely due to small amounts of beads which were collected with the elution. B) Western blot against RanBP5 testing the effect of different buffer conditions and particularly detergents in recovery of RanBP5 during purification of PB1-PA-RanBP5. Antibodies: primary rabbit α-FluPol (Eurogentec Ltd), primary mouse α-RanBP5 (Santa Cruz Biotechnology), secondary α-rabbit IgG or α-mouse IgG HRP-conjugated antibodies.

6.3.2 Reconstitution of PB1-PA-RanBP5 complex from separately purified components

The protein quantities that were being obtained through this co-expression protocol were very low. Even though we had determined that low ionic strength buffers were key to obtaining equimolar ratios of the three complex components, the yields obtained after IgG purification from 1 litre of insect cells were too low to reproducibly obtain a detectable signal in gel filtration. The complex also appeared to partially dissociate during gel filtration as shown by elution of RanBP5 and PB1-PA in a separate peak at higher elution volumes. I therefore decided to test whether the PB1-PA-RanBP5 complex could be reconstituted from separately purified components. N-terminally His-tagged RanBP5 (isoform 2) was expressed in *E. coli* and purified by affinity chromatography on Ni^{2+}-NTA agarose beads. This yielded large quantities of pure protein (Fig. 56a). When further purified by size exclusion chromatography, the protein eluted as a single peak with a shoulder at lower volumes (these fractions were discarded) (Fig. 56b). Taking advantage of the discovery that high NaCl concentrations effectively disrupt the interaction between PB1-PA and RanBP5, the same baculovirus for expression of PB1-PA-RanBP5 was used to purify PB1-PA alone. After purification (with extensive washing of IgG beads) in 500 mM NaCl containing buffers, a sample containing PB1-PA dimer with no visible RanBP5 contamination could be obtained (Fig. 56c).

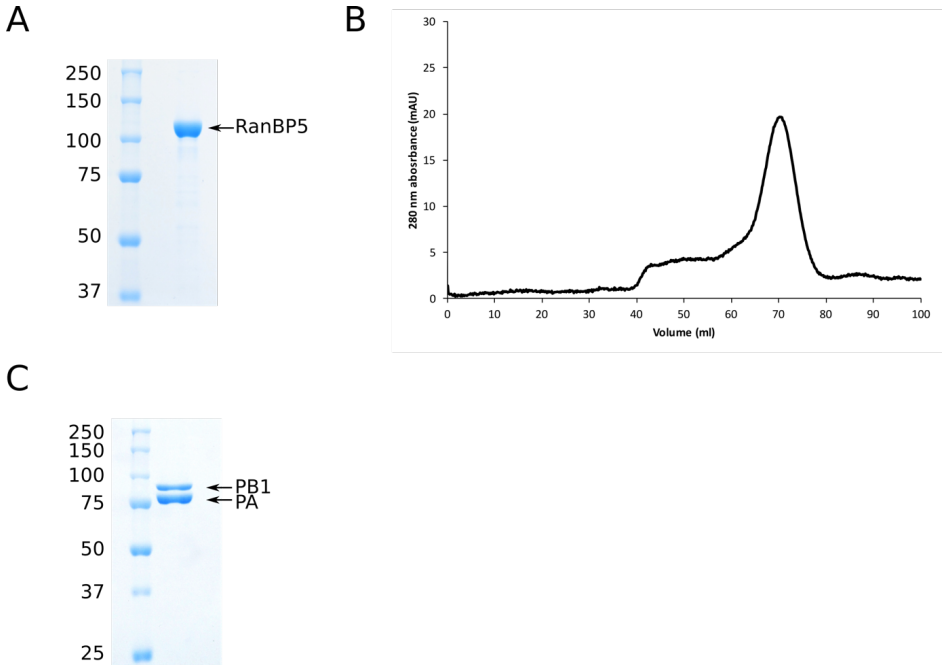


Figure 56 – Purification of RanBP5 and PB1-PA. A) SDS-PA gel of purified bacterially expressed RanBP5. Gel stained with SimplyBlue SafeStain. B) Gel filtration profile of bacterially expressed RanBP5. C) SDS-PA gel of purified insect cell expressed PB1-PA after IgG purification. Gel stained with SimplyBlue SafeStain.

Initial attempts to directly mix these two purified samples resulted in extensive protein precipitation, potentially due to mixing of the very different buffers that the two samples were in (PB1-PA cleavage buffer contained 25 mM Hepes pH 7.5, 500 mM NaCl, 10 % glycerol, 0.05 % OTG, 2.5 mM reduced glutathione whereas RanBP5 elution buffer contained 25 mM Tris pH 8, 100 mM NaCl, 5 mM $MgCl_2$, 250 mM imidazole, 1 mM DTT and 200 mM sucrose). As this was evidently not the right approach, I attempted to determine whether binding between these two components occurred in a pull-down assay (Fig. 57). I tested whether PB1-PA immobilised on IgG beads via the protein A tag on PB1 would co-precipitate purified RanBP5 and vice versa, with RanBP5 immobilised on Ni^{2+}-NTA beads and purified PB1-PA being added (schematic of the assay can be found in Fig. 57a). This was essentially carried out by completing the steps of purification prior to elution (lysis, binding to the beads and washing of the beads) followed by a 1 hour incubation with the binding partner. After incubation, beads were sedimented, the supernatant discarded and the beads thoroughly washed with binding buffer (30 mM Tris pH 8, 125 mM NaCl, 2.5 mM glutathione, 5 mM $MgCl_2$, 0.05 % OTG and 10 % glycerol). Elution from the beads (or cleavage in the case of PB1-PA) was carried out as before (except the cleavage buffer for PB1-PA contained 125 mM NaCl instead of 500 mM NaCl to avoid disrupting the complex). The eluted samples were analysed by SDS-PAGE (Fig. 57b). Sample 1, where PB1-PA was immobilised on IgG beads and affinity purified RanBP5 added, showed a distinct RanBP5 band co-eluting with it. This suggests that binding between the two can take place, although not efficiently, as shown by the excess of PB1-PA compared to RanBP5. The opposite binding assay (sample 2), with RanBP5 immobilised on the beads and affinity purified PB1-PA being added, showed high levels of contaminating proteins in addition to RanBP5. Within these contaminating proteins, bands for PB1 and PA could not be identified.

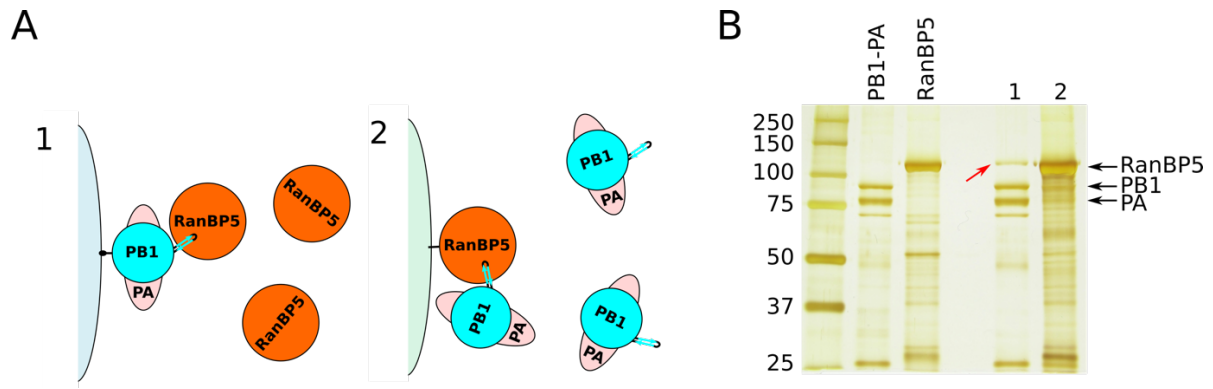


Figure 57 – Reconstitution of a PB1-PA-RanBP5 complex from separately purified components. A) Binding assay schematic diagram B) Binding assay to determine whether reconstitution of the RanBP5-PB1-PA complex is possible from separately purified components. Silver-stained SDS-PA gel of elutions from binding assay.

6.3.3 Conclusions and future directions

Our attempts at purifying a complex containing RanBP5 and the polymerase subunits PB1 or PB1 and PA have not yet been sufficiently optimised to produce the yields required for structural characterisation. On the other hand, our work has provided some valuable information which can be applied to further work in optimising the reconstitution of this import complex (work now being carried out by Mattia Bertinelli). Expression of PB1 in absence of PA is significantly poorer than that in presence of PA (Fig. 52, 53 and 54a, b). As mentioned in the introduction to this chapter, the available FluPol heterotrimer structures which are now available (and which were not available at the time when this work was started), give a possible explanation for this (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). PB1 and PA interact very closely, with the PA endonuclease and C-terminal domains placed on opposite sides of PB1 and joined by a long linker. PB1 and the PA linker and C-terminus have an extensive interaction surface, which may be required to stabilise PB1. Indeed, solubility of PB1 has been linked to the presence of the PA residues 200-260 in the linker (Swale et al., 2016). This observation may also explain why import of PB1 alone was inefficient, as it is possible that the protein was misfolding (Deng et al., 2005; Huet et al., 2010; Hutchinson et al., 2011). It appears therefore that the best option to study the RanBP5-PB1 interaction is to reconstitute a complex that also contains PA. PB1-PA expression was relatively robust in insect cells, however, obtaining a stable interaction with RanBP5, whether co-expressed or through reconstitution of the complex from purified components was challenging. It quickly became apparent that the interaction could be easily

disrupted by high salt concentrations (Fig. 54a, b) and that small amounts of detergent were required during elution of RanBP5 to keep it stably in solution (Fig. 55c). Bearing these factors in mind we could reconstitute a complex with apparently equimolar amounts of PB1, PA and RanBP5 from co-expression in insect cells. However, this complex was still fragile and would partially dissociate during gel filtration (Fig. 54c, d). Our attempts to reconstitute the complex from separately purified components showed that we could purify RanBP5 and PB1-PA in good amounts, better than the levels we could obtain through co-expression (Fig. 56). However, forming a complex from these components was not trivial. We therefore carried out a pull-down assay to determine whether interaction between the two could take place. RanBP5 co-precipitated with PB1-PA immobilised on IgG beads, which suggests the two can interact, although the ratio of RanBP5 to PB1 was low, indicating that only a fraction of PB1-PA stably bound RanBP5 (Fig. 57b). In retrospect addition of the tag at the PB1 C-terminus may not have been the best choice, as it is positioned relatively close to the β-ribbon and may interfere with binding, particularly where PB1-PA was bound to the beads. A better approach in future might involve moving the protein A tag to the C terminal end of PA, which would be less likely to interfere with binding.

A study was published while this work was being carried out which also attempted structural characterisation of this complex (Swale et al., 2016). Similar to our findings, they struggled to obtain a stable complex between RanBP5 and PB1-PA, which they attempted to express as a self-cleaving polyprotein in order to achieve adequate stoichiometry. This led them to deleting the C-terminal 71 amino acids in PB1, after which they could obtain a stable heterotrimeric complex. Why this C-terminal truncation of PB1 aided complex formation is unclear, but it supports the idea that tagging the complex via this region could be having a detrimental effect on complex stability. Even though they obtained a homogeneous sample, the complex proved difficult to structurally characterise, so SAXS was performed on the complex instead. This showed an extended interaction surface, which may suggest there are interaction regions beyond the NLS. They also determined that binding of RanBP5 to the PB1-PA dimer prevented it from associating with 5′ vRNA, which supports the view that RanBP5

can act as a chaperone, shielding proteins and preventing them from binding polyanions (in this case viral RNA) prematurely (Jakel et al., 2002).

For future work on optimising the isolation of this complex for high resolution structural characterisation the protein A tag should be repositioned at the PA C-terminus, as it has become clear that the PB1 C-terminus has a role in the interaction. Once this is done, reconstitution of the complex from purified components should be re-attempted, as this produces greater starting yields than co-expression. On the other hand, yields of the co-expressed complex may improve once the position of the tag is changed so this should be tested too. As the PB1 C-terminus appears to have a role in the interaction, its deletion may not be the best option if we want to obtain a biologically relevant structure. Therefore, in order to further stabilise the complex, mild crosslinking protocols such as GraFix could be employed (Kastner et al., 2007; Stark, 2010).

6.4 Summary

Influenza virus hijacks a pathway that is built for efficient import of very abundant proteins for nuclear import of two of the polymerase subunits. Reconstitution of this import complex responsible for delivering the PB1-PA dimer to the nucleus of infected cells remains challenging. Nevertheless, based on the structures of its yeast homolog Pse1p and the SAXS envelope of the PB1(1-686)-PA-RanBP5 complex, we can speculate that in a PB1-PA-RanBP5 complex the β-ribbon in PB1 would bind on the inner groove of the coil-like structure that is probably formed by RanBP5, as it is also predicted to consist mostly of HEAT and Armadillo repeats. Further interactions beyond the PB1 β-ribbon could be further stabilising this interaction. However, this might not be the full picture: it remains unclear why PA contains an NLS signal when its import appears to occur in complex with PB1, and why mutation of the PA NLS is deleterious for its nuclear import (Nieto et al., 1994). In fact, the presence of PA appears to be more important for nuclear import of PB1 than RanBP5 itself and PB1 containing mutations in the NLS still displayed some nuclear accumulation (Deng et al., 2006; Hutchinson et al., 2011). Further work will therefore be required not only to characterise the PB1-PA-

RanBP5 complex but also to determine whether other import pathways are also exploited by the virus for import of these polymerase subunits.

Concluding remarks

The work presented here has allowed us to improve our understanding of the mechanisms involved in transcription and replication by the influenza virus polymerase. When the influenza virus first enters a host cell, vRNPs are imported into the nucleus, where they carry out transcription without requiring additional protein expression (Vreede and Brownlee, 2007). Our functional assays exploring the effect of Pol II CTD binding to FluPol, in conjunction with the $FluPol_C$ crystal structure bound to a S_5P CTD peptide mimic, have revealed a role for Pol II in transcription activation. Apo $FluPol_C$ is found in a "closed" conformational state not compatible with transcription (Hengrung et al., 2015). vRNA binding, or the availability of capped RNA are not sufficient to enable the polymerase to efficiently initiate transcription. Instead, binding to the S_5-phosphorylated form of Pol II CTD appears to be the key in stabilising the heterotrimer in the "open", pre-initiation conformation, where the cap binding domain in PB2 and the endonuclease domain in P3 acquire the right arrangement to cap-snatch. Based on the available $FluPol_C$ and $FluPol_B$ crystal structures (Lukarska et al., 2016), we hypothesise that this occurs via a mechanism where the S_5P CTD bridges the P3 C-terminal domain and the PB2 627 domain.

As infection progresses, there is a switch from transcription to replication of vRNA. Our work with purified vRNPs and cRNPs has allowed us to gain insight into how this might occur. There are two core aspects that change during infection that might drive a switch towards replication. The first is degradation of Pol II, which is induced by FluPol (Rodriguez et al., 2007; Vreede et al., 2010). At the same time, newly produced viral proteins accumulate in the cell, some of which are imported into the nucleus, including the polymerase subunits and nucleoprotein. A decrease in Pol II availability would reduce the population of FluPol that is capable of transcribing. However, our data suggests that newly produced FluPol and NP also have an active role in this switch from transcription to replication, with FluPol promoting replication and NP suppressing transcription. This multipronged approach to regulation would provide tight control over replication during infection, ensuring it only occurs at the adequate time. This observation also highlights how nuclear import of the viral replication machinery components is key, not only for the formation of progeny vRNPs but also for regulation of RNP activity.

Our data showing that addition of FluPol also enhances vRNP replicative activity suggests that non-RNP-associated polymerase has a role in activating the resident polymerase in vRNPs. This had been previously suggested for cRNPs, which are incapable of replication in absence of additional polymerase, but is a novel observation for vRNPs. The precise role of additional polymerase in these two scenarios remains unknown, as does the reason behind the inability of cRNP-resident polymerases to replicate. However, the methods described here for analysis of cRNPs and vRNPs by cryo-EM should soon allow us to compare these two complexes and determine which structural features are responsible for the differences observed in activity.

Further structural characterisation of FluPol in multiple states and in complex with host and viral proteins will undoubtedly be required to fully understand how it carries out its function. While crystallography has been responsible for most of the full heterotrimeric FluPol structures available, recent advances in cryo-EM should make it a suitable method by which to study FluPol structure, despite it being a small asymmetric molecule. However, issues with sample preparation have so far prevented us from achieving this. Nevertheless, multiple options remain available to address the issues encountered with preferential orientations and particle contrast, which should make it possible to obtain high-resolution structures by this method in the near future. In terms of exploring multiple states of the polymerase, nanobodies have proved to be excellent modulators of polymerase activity. In addition to that, the discovery of nanobodies that provide robust inhibition of all polymerase activity could offer a means to identify sites on the polymerase that can be targeted for inhibitor design. The development of novel influenza inhibitors to use as antivirals is of course essential in the efforts to control influenza disease, particularly for use in times where an adequate vaccine is not available (e.g. novel pandemics of zoonotic origin) or for individuals who have increased risk of exposure, co-morbidities or who cannot acquire adequate protective immunity through vaccination.

References

Aggarwal, S., Dewhurst, S., Takimoto, T., and 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.

Albo, C., Martín, J., and Portela, A. (1996). The 5' ends of Thogoto virus (Orthomyxoviridae) mRNAs are homogeneous in both length and sequence. J. Virol. *70*, 9013–9017.

Arnold, S.A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R.A., Goldie, K.N., Stahlberg, H., and Braun, T. (2017). Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. J. Struct. Biol. *197*, 220–226.

Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J., and Martín-Benito, J. (2012). The Structure of Native Influenza Virion Ribonucleoproteins. Science *338*, 1634–1637.

Azzeh, M., Flick, R., and 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, C.N., Nasreen, S., Bhuiyan, M.U., Mah-E-Muneer, S., Mamun, A.A., Sharker, M.A.Y., Zaman, R.U., Cheng, P.-Y., Klimov, A.I., et al. (2012). Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide. J. Infect. Dis. *206*, 838–846.

Baake, M., Bäuerle, M., Doenecke, D., and Albig, W. (2001). Core histones and linker histones are imported into the nucleus by different pathways. Eur. J. Cell Biol. *80*, 669–677.

Bae, S.-H., Cheong, H.-K., Lee, J.-H., Cheong, C., Kainosho, M., and Choi, B.-S. (2001). Structural features of an influenza virus promoter and their implications for viral RNA synthesis. Proc. Natl. Acad. Sci. *98*, 10602–10607.

Baker, M.R., Fan, G., and Serysheva, I.I. (2015). Single-particle cryo-EM of the ryanodine receptor channel in an aqueous environment. Eur. J. Transl. Myol. *25*, 35–48.

Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jónsson, S.R., Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., et al. (2017). A supramolecular assembly mediates lentiviral DNA integration. Science *355*, 93–95.

Barman, S., Ali, A., Hui, E.K.-W., Adhikary, L., and Nayak, D.P. (2001). Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly of influenza viruses. Virus Res. *77*, 61–69.

Barth, M., Bryan, R.K., and Hegerl, R. (1989). Approximation of missing-cone data in 3D electron microscopy. Ultramicroscopy *31*, 365–378.

Baudin, F., Bach, C., Cusack, S., and Ruigrok, R.W. (1994). Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. EMBO J. *13*, 3158–3165.

Beaton, A.R., and Krug, R.M. (1981). Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. Nucleic Acids Res. *9*, 4423–4436.

Beaton, A.R., and Krug, R.M. (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. *83*, 6282–6286.

Belshe, R.B., Edwards, K.M., Vesikari, T., Black, S.V., Walker, R.E., Hultquist, M., Kemble, G., and Connor, E.M. (2007). Live attenuated versus inactivated influenza vaccine in infants and young children. N. Engl. J. Med. *356*, 685–696.

Betakova, T., Nermut, M.V., and Hay, A.J. (1996). The NB protein is an integral component of the membrane of influenza B virus. J. Gen. Virol. *77 (Pt 11)*, 2689–2694.

Bieniossek, C., Richmond, T.J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al *Chapter 5*, Unit 5.20.

Bier, K., York, A., and Fodor, E. (2011). Cellular cap-binding proteins associate with influenza virus mRNAs. J. Gen. Virol. *92*, 1627–1634.

Biswas, S.K., and Nayak, D.P. (1994). Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J. Virol. *68*, 1819–1826.

Biswas, S.K., Boutz, P.L., and Nayak, D.P. (1998). Influenza Virus Nucleoprotein Interacts with Influenza Virus Polymerase Proteins. J. Virol. *72*, 5493–5501.

Blok, V., Cianci, C., Tibbles, K.W., Inglis, S.C., Krystal, M., and 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.

Boivin, S., and Hart, D.J. (2011). Interaction of the Influenza A Virus Polymerase PB2 C-terminal Region with Importin α Isoforms Provides Insights into Host Adaptation and Polymerase Assembly. J. Biol. Chem. *286*, 10439–10448.

Bortz, E., Westera, L., Maamary, J., Steel, J., Albrecht, R.A., Manicassamy, B., Chase, G., Martínez-Sobrido, L., Schwemmle, M., and García-Sastre, A. (2011). Host- and Strain-Specific Regulation of Influenza Virus Polymerase Activity by Interacting Cellular Proteins. MBio *2*, e00151-11.

Bradel-Tretheway, B.G., Mattiacio, J.L., Krasnoselsky, A., Stevenson, C., Purdy, D., Dewhurst, S., and Katze, M.G. (2011). Comprehensive Proteomic Analysis of Influenza Virus Polymerase Complex Reveals a Novel Association with Mitochondrial Proteins and RNA Polymerase Accessory Factors. J. Virol. *85*, 8569–8581.

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R.L., Mathieu, M., De Francesco, R., and Rey, F.A. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc. Natl. Acad. Sci. *96*, 13034–13039.

Brooke, C.B., Ince, W.L., Wrammert, J., Ahmed, R., Wilson, P.C., Bennink, J.R., and Yewdell, J.W. (2013). Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. J. Virol. *87*, 3155–3162.

Brownlee, G.G., and Sharps, J.L. (2002). The RNA Polymerase of Influenza A Virus Is Stabilized by Interaction with Its Viral RNA Promoter. J. Virol. *76*, 7103–7113.

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

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., and Stuart, D.I. (2001). A mechanism for initiating RNA-dependent RNA polymerization. Nature *410*, 235–240.

Çevik, B., Smallwood, S., and Moyer, S.A. (2003). The L–L oligomerization domain resides at the very N-terminus of the sendai virus L RNA polymerase protein. Virology *313*, 525–536.

Chan, A.Y., Vreede, F.T., Smith, M., Engelhardt, O.G., and Fodor, E. (2006). Influenza virus inhibits RNA polymerase II elongation. Virology *351*, 210–217.

Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., et al. (2015). Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. Mol. Cell.

Chen, W., Calvo, P.A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., et al. (2001). A novel influenza A virus mitochondrial protein that induces cell death. Nat. Med. *7*, 1306–1312.

Cheong, H.-K., Cheong, C., Lee, Y.-S., Seong, B.L., and Choi, B.-S. (1999). Structure of influenza virus panhandle RNA studied by NMR spectroscopy and molecular modeling. Nucleic Acids Res. *27*, 1392–1397.

Chu, C.M., Dawson, I.M., and Elford, W.J. (1949). Filamentous forms associated with newly isolated influenza virus. Lancet Lond. Engl. *1*, 602.

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

Clark, M.P., Ledeboer, M.W., Davies, I., Byrn, R.A., Jones, S.M., Perola, E., Tsai, A., Jacobs, M., Nti-Addae, K., Bandarage, U.K., et al. (2014). Discovery of a Novel, First-in-Class, Orally Bioavailable Azaindole Inhibitor (VX-787) of Influenza PB2. J. Med. Chem. *57*, 6668–6678.

Clarke, P.R., and Zhang, C. (2001). Ran GTPase: a master regulator of nuclear structure and function during the eukaryotic cell division cycle? Trends Cell Biol. *11*, 366–371.

Clerx, J.P.M., Fuller, F., and Bishop, D.H.L. (1983). Tick-borne viruses structurally similar to orthomyxoviruses. Virology *127*, 205–219.

Clouthier, S.C., Rector, T., Brown, N.E.C., and Anderson, E.D. (2002). Genomic organization of infectious salmon anaemia virus. J. Gen. Virol. *83*, 421–428.

Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., Ortín, J., and Martín-Benito, J. (2009). The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog. *5*, e1000491.

Compans, R.W., and Duesberg, P.H. (1972). Structure of the ribonucleoprotein of influenza virus. J. Virol. *10*, 795–800.

Connor, R.J., Kawaoka, Y., Webster, R.G., and Paulson, J.C. (1994). Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates. Virology *205*, 17–23.

Coscia, F., Estrozi, L.F., Hans, F., Malet, H., Noirclerc-Savoye, M., Schoehn, G., and Petosa, C. (2016). Fusion to a homo-oligomeric scaffold allows cryo-EM analysis of a small protein. Sci. Rep. *6*, srep30909.

Crepin, T., Dias, A., Palencia, A., Swale, C., Cusack, S., and Ruigrok, R.W.H. (2010). Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit. J. Virol. *84*, 9096–9104.

Crescenzo-Chaigne, B., Naffakh, N., and van der Werf, S. (1999). Comparative Analysis of the Ability of the Polymerase Complexes of Influenza Viruses Type A, B and C to Assemble into Functional RNPs that Allow Expression and Replication of Heterotypic Model RNA Templates In Vivo. Virology *265*, 342–353.

Cros, J.F., García-Sastre, A., and Palese, P. (2005). An Unconventional NLS is Critical for the Nuclear Import of the Influenza A Virus Nucleoprotein and Ribonucleoprotein. Traffic *6*, 205–213.

Crow, M., Deng, T., Addley, M., and Brownlee, G.G. (2004). Mutational Analysis of the Influenza Virus cRNA Promoter and Identification of Nucleotides Critical for Replication. J. Virol. *78*, 6263–6270.

Czudai-Matwich, V., Otte, A., Matrosovich, M., Gabriel, G., and Klenk, H.-D. (2014). PB2 Mutations D701N and S714R Promote Adaptation of an Influenza H5N1 Virus to a Mammalian Host. J. Virol. *88*, 8735–8742.

Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D., and Hutchinson, E.C. (2016). Filamentous influenza viruses. J. Gen. Virol. *97*, 1755–1764.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. Proc. Natl. Acad. Sci. *111*, 15635–15640.

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

Delaforge, E., Milles, S., Bouvignies, G., Bouvier, D., Boivin, S., Salvi, N., Maurin, D., Martel, A., Round, A., Lemke, E.A., et al. (2015). Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin α. J. Am. Chem. Soc. 151016093853008.

Deng, T., Sharps, J., Fodor, E., and Brownlee, G.G. (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., Engelhardt, O.G., Thomas, B., Akoulitchev, A.V., Brownlee, G.G., and 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., Vreede, F.T., and Brownlee, G.G. (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, A.A., Hart, D.J., Baudin, F., Cusack, S., and Ruigrok, R.W.H. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature *458*, 914–918.

DiazGranados, C.A., Dunning, A.J., Kimmel, M., Kirby, D., Treanor, J., Collins, A., Pollak, R., Christoff, J., Earl, J., Landolfi, V., et al. (2014). Efficacy of High-Dose versus Standard-Dose Influenza Vaccine in Older Adults. N. Engl. J. Med. *371*, 635–645.

Domingues, P., and Hale, B.G. (2017). Functional Insights into ANP32A-Dependent Influenza A Virus Polymerase Host Restriction. Cell Rep. *20*, 2538–2546.

Dubochet, J., Ducommun, M., Zollinger, M., and Kellenberger, E. (1971). A new preparation method for dark-field electron microscopy of biomacromolecules. J. Ultrastruct. Res. *35*, 147–167.

Dubochet, J., Adrian, M., Lepault, J., and McDowall, A.W. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. Trends Biochem. Sci. *10*, 143–146.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Q. Rev. Biophys. *21*, 129.

DuBois, R.M., Slavish, P.J., Baughman, B.M., Yun, M.-K., Bao, J., Webby, R.J., Webb, T.R., and White, S.W. (2012). Structural and Biochemical Basis for Development of Influenza Virus Inhibitors Targeting the PA Endonuclease. PLoS Pathog. *8*.

Duesberg, P.H. (1969). Distinct subunits of the ribonucleoprotein of influenza virus. J. Mol. Biol. *42*, 485–499.

Eisfeld, A.J., Neumann, G., and Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. Nat. Rev. Microbiol. 13, 28.

Elmlund, D., Le, S.N., and Elmlund, H. (2017). High-resolution cryo-EM: the nuts and bolts. Curr. Opin. Struct. Biol. *46*, 1–6.

Engelhardt, O.G., Smith, M., and Fodor, E. (2005). Association of the Influenza A Virus RNA-Dependent RNA Polymerase with Cellular RNA Polymerase II. J. Virol. *79*, 5812–5818.

Expert-Bezançon, N., Rabilloud, T., Vuillard, L., and Goldberg, M.E. (2003). Physical-chemical features of non-detergent sulfobetaines active as protein-folding helpers. Biophys. Chem. *100*, 469–479.

Fechter, P. (2005). Recognition of mRNA cap structures by viral and cellular proteins. J. Gen. Virol. *86*, 1239–1249.

Flajnik, M.F., Deschacht, N., and Muyldermans, S. (2011). A Case Of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels? PLoS Biol. *9*, e1001120.

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

Fodor, E., Pritlove, D.C., and Brownlee, G.G. (1994). The influenza virus panhandle is involved in the initiation of transcription. J. Virol. *68*, 4092–4096.

Fodor, E., Crow, M., Mingay, L.J., Deng, T., Sharps, J., Fechter, P., and Brownlee, G.G. (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.

Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., et al. (2012). A supramolecular assembly formed by influenza A virus genomic RNA segments. Nucleic Acids Res. *40*, 2197–2209.

Furuta, Y., Gowen, B.B., Takahashi, K., Shiraki, K., Smee, D.F., and Barnard, D.L. (2013). Favipiravir (T-705), a novel viral RNA polymerase inhibitor. Antiviral Res. *100*, 446–454.

Gabriel, G., Herwig, A., and Klenk, H.-D. (2008). Interaction of Polymerase Subunit PB2 and NP with Importin α1 Is a Determinant of Host Range of Influenza A Virus. PLoS Pathog. *4*, e11.

Gallagher, J.R., Torian, U., McCraw, D.M., and Harris, A.K. (2016). Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. J. Struct. Biol.

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

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

Gerlach, P., Malet, H., Cusack, S., and Reguera, J. (2015). Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter. Cell *161*, 1267–1279.

Glaeser, R.M., and Han, B.-G. (2017). Opinion: hazards faced by macromolecules when confined to thin aqueous films. Biophys. Rep. *3*, 1–7.

Golovanov, A.P., Hautbergue, G.M., Wilson, S.A., and Lian, L.-Y. (2004). A Simple Method for Improving Protein Solubility and Long-Term Stability. J. Am. Chem. Soc. *126*, 8933–8939.

Görlich, D., and Kutay, and U. (1999). Transport Between the Cell Nucleus and the Cytoplasm. Annu. Rev. Cell Dev. Biol. *15*, 607–660.

Grohskopf, L.A. (2016). Prevention and Control of Seasonal Influenza with Vaccines. MMWR Recomm. Rep. *65*.

Gu, W., Gallagher, G.R., Dai, W., Liu, P., Li, R., Trombly, M.I., Gammon, D.B., Mello, C.C., Wang, J.P., and Finberg, R.W. (2015). Influenza A virus preferentially snatches noncoding RNA caps. RNA *21*, 2067–2075.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortin, J., Hart, D.J., et al. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. *15*, 500–506.

Guilligay, D., Kadlec, J., Crepin, T., Lunardi, T., Bouvier, D., Kochs, G., Ruigrok, R.W.H., and Cusack, S. (2014). Comparative Structural and Functional Analysis of Orthomyxovirus Polymerase Cap-Snatching Domains. PLoS ONE *9*.

Guu, T.S.Y., Dong, L., Wittung-Stafshede, P., and Tao, Y.J. (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, T.D., Butcher, J.A., and Krystal, M. (1994). Recombinant influenza virus polymerase: requirement of both 5'and 3'viral ends for endonuclease activity. J. Virol. *68*, 1509–1515.

Hagen, M., Tiley, L., Chung, T.D.Y., and Krystal, M. (1995). The role of template-primer interactions in cleavage and initiation by the influenza virus polymerase. J. Gen. Virol. *76*, 603–611.

Hara, K., Schmidt, F.I., Crow, M., and Brownlee, G.G. (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.

Harel, A., and Forbes, D.J. (2004). Importin Beta: Conducting a Much Larger Cellular Symphony. Mol. Cell *16*, 319–330.

Hause, B.M., Collin, E.A., Liu, R., Huang, B., Sheng, Z., Lu, W., Wang, D., Nelson, E.A., and 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.

Hay, A.J., Lomniczi, B., Bellamy, A.R., and Skehel, J.J. (1977). Transcription of the influenza virus genome. Virology *83*, 337–355.

He, F., Leyrer, S., and Kwang, J. (2016). Strategies towards universal pandemic influenza vaccines. Expert Rev. Vaccines *15*, 215–225.

He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., et al. (2008). Crystal structure of the polymerase PAC–PB1N complex from an avian influenza H5N1 virus. Nature *454*, 1123–1126.

Heggeness, M.H., Smith, P.R., Ulmanen, I., Krug, R.M., and Choppin, P.W. (1982). Studies on the helical nucleocapsid of influenza virus. Virology *118*, 466–470.

Hengrung, N., El Omari, K., Serna Martin, I., Vreede, F.T., Cusack, S., Rambo, R.P., Vonrhein, C., Bricogne, G., Stuart, D.I., Grimes, J.M., et al. (2015). Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. Nature *527*, 114–117.

Herfst, S., Schrauwen, E.J.A., Linster, M., Chutinimitkul, S., Wit, E. de, Munster, V.J., Sorrell, E.M., Bestebroer, T.M., Burke, D.F., Smith, D.J., et al. (2012). Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets. Science *336*, 1534–1541.

Herrler, G., Rott, R., Klenk, H.D., Müller, H.P., Shukla, A.K., and Schauer, R. (1985). The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. EMBO J. *4*, 1503–1506.

Hobson, S.D., Rosenblum, E.S., Richards, O.C., Richmond, K., Kirkegaard, K., and Schultz, S.C. (2001). Oligomeric structures of poliovirus polymerase are important for function. EMBO J. *20*, 1153–1163.

Hongo, S., Sugawara, K., Nishimura, H., Muraki, Y., Kitame, F., and Nakamura, K. (1994). Identification of a second protein encoded by influenza C virus RNA segment 6. J. Gen. Virol. *75 (Pt 12)*, 3503–3510.

Hudjetz, B., and Gabriel, G. (2012). Human-like PB2 627K Influenza Virus Polymerase Activity Is Regulated by Importin-α1 and -α7. PLoS Pathog. *8*, e1002488.

Huet, S., Avilov, S.V., Ferbitz, L., Daigle, N., Cusack, S., and Ellenberg, J. (2010). Nuclear Import and Assembly of Influenza A Virus RNA Polymerase Studied in Live Cells by Fluorescence Cross-Correlation Spectroscopy. J. Virol. *84*, 1254–1264.

Hutchinson, E.C., and Fodor, E. (2012). Nuclear import of the influenza A virus transcriptional machinery. Vaccine *30*, 7353–7358.

Hutchinson, E.C., Orr, O.E., Man Liu, S., Engelhardt, O.G., and 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.

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

Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I., and Huiskonen, J.T. (2015). Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat. Commun. *6*, 8843.

Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.-A., Wills, N.M., Xiao, Y.-L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L., et al. (2012). An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. Science *337*, 199–204.

Jakel, S., and Gorlich, D. (1998). Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. EMBO J. *17*, 4491–4502.

Jakel, S., Mingot, J.-M., Schwarzmaier, P., Hartmann, E., and Gorlich, D. (2002). Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. EMBO J. *21*, 377–386.

Jasnovidova, O., and Stefl, R. (2013). The CTD code of RNA polymerase II: a structural view. Wiley Interdiscip. Rev. RNA *4*, 1–16.

Jennings, P.A., Finch, J.T., Winter, G., and Robertson, J.S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? Cell *34*, 619–627.

Jeronimo, C., Collin, P., and Robert, F. (2016). The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. J. Mol. Biol. *428*, 2607–2622.

Johnson, N.P.A.S., and 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., and Ortin, J. (2008). Oligomerization of the influenza virus polymerase complex in vivo. J. Gen. Virol. *89*, 520–524.

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

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2007). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods *5*, 53–55.

Kawakami, E., Watanabe, T., Fujii, K., Goto, H., Watanabe, S., Noda, T., and Kawaoka, Y. (2011). Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. J. Virol. Methods *173*, 1–6.

Kelly, D.F., Dukovski, D., and Walz, T. (2008). Monolayer purification: A rapid method for isolating protein complexes for single-particle electron microscopy. Proc. Natl. Acad. Sci. *105*, 4703–4708.

Kingsbury, D.W., and Webster, R.G. (1969). Some Properties of Influenza Virus Nucleocapsids. J. Virol. *4*, 219–225.

Knipe, D.M., and Howley, P.M. (2007). Fields' Virology (Lippincott Williams & Wilkins).

Kobayashi, J., and Matsuura, Y. (2013). Structural Basis for Cell-Cycle-Dependent Nuclear Import Mediated by the Karyopherin Kap121p. J. Mol. Biol. *425*, 1852–1868.

Kobayashi, Y., Dadonaite, B., Doremalen, N. van, Suzuki, Y., Barclay, W.S., and Pybus, O.G. (2016). Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. RNA Biol. *13*, 883–894.

Koppstein, D., Ashour, J., and Bartel, D.P. (2015). Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res. *43*, 5052–5064.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Kuzuhara, T., Kise, D., Yoshida, H., Horita, T., Murazaki, Y., Nishimura, A., Echigo, N., Utsunomiya, H., and 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., Dos Santos Afonso, E., Rameix-Welti, M.-A., van der Werf, S., and 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.

Lakdawala, S.S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A.J., Lamirande, E.W., Fodor, E., Altan-Bonnet, N., Shroff, H., and Subbarao, K. (2014). Influenza A Virus Assembly Intermediates Fuse in the Cytoplasm. PLoS Pathog. *10*, e1003971.

Lakdawala, S.S., Jayaraman, A., Halpin, R.A., Lamirande, E.W., Shih, A.R., Stockwell, T.B., Lin, X., Simenauer, A., Hanson, C.T., Vogel, L., et al. (2015). The soft palate is an important site of adaptation for transmissible influenza viruses. Nature *526*, 122–125.

Leahy, M.B., Dobbyn, H.C., and Brownlee, G.G. (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, M.B., Pritlove, D.C., Poon, L.L.M., and Brownlee, G.G. (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.

Lescar, J., and Canard, B. (2009). RNA-dependent RNA polymerases from flaviviruses and Picornaviridae. Curr. Opin. Struct. Biol. *19*, 759–767.

Li, M.-L., Ramirez, B.C., and Krug, R.M. (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, M.-L., Rao, P., and Krug, R.M. (2001). The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. EMBO J. *20*, 2078–2086.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Lidschreiber, M., Leike, K., and Cramer, P. (2013). Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II. Mol. Cell. Biol. *33*, 3805–3816.

Lim, F., and Peabody, D.S. (2002). RNA recognition site of PP7 coat protein. Nucleic Acids Res. *30*, 4138–4144.

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

Liu, Y., Yang, Y., Fan, J., He, R., Luo, M., and Zheng, X. (2015). The Crystal Structure of the PB2 Cap-binding Domain of Influenza B Virus Reveals a Novel Cap Recognition Mechanism. J. Biol. Chem. *290*, 9141–9149.

Long, J.S., Giotis, E.S., Moncorgé, O., Frise, R., Mistry, B., James, J., Morisson, M., Iqbal, M., Vignal, A., Skinner, M.A., et al. (2016). Species difference in ANP32A underlies influenza A virus polymerase host restriction. Nature *529*, 101–104.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. J. Struct. Biol. *128*, 82–97.

Lukarska, M., Fournier, G., Pflug, A., Resa-Infante, P., Reich, S., Naffakh, N., and Cusack, S. (2016). Structural basis of an essential interaction between influenza polymerase and Pol II CTD. Nature *541*, 117–121.

Luo, G., Bergmann, M., Garcia-Sastre, A., and Palese, P. (1992). Mechanism of attenuation of a chimeric influenza A/B transfectant virus. J. Virol. *66*, 4679–4685.

Lyle, J.M., Bullitt, E., Bienz, K., and Kirkegaard, K. (2002). Visualization and Functional Analysis of RNA-Dependent RNA Polymerase Lattices. Science *296*, 2218–2222.

Mänz, B., Schwemmle, M., and Brunotte, L. (2013). Adaptation of Avian Influenza A Virus Polymerase in Mammals To Overcome the Host Species Barrier. J. Virol. *87*, 7200–7209.

Martin, T.G., Bharat, T.A.M., Joerger, A.C., Bai, X., Praetorius, F., Fersht, A.R., Dietz, H., and Scheres, S.H.W. (2016). Design of a molecular support for cryo-EM structure determination. Proc. Natl. Acad. Sci. *113*, E7456–E7463.

Martínez-Alonso, M., Hengrung, N., and Fodor, E. (2016). RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II. J. Virol. *90*, 6014–6021.

Massin, P., Werf, S. van der, and 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.

Matlin, K.S., Reggio, H., Helenius, A., and Simons, K. (1981). Infectious entry pathway of influenza virus in a canine kidney cell line. J Cell Biol *91*, 601–613.

McNeil, S.A., Hatchette, T., Andrew, M.K., Ambrose, A., Boivin, G., Diaz-Mitoma, F., Bowie, W., Chit, A., Santos, G.D., Elsherif, M., et al. (2016). Influenza Vaccine Effectiveness in the Prevention of Influenza-Related Hospitalization in Canadian Adults Over the 2011/12 Through 2013/14 Season: A Pooled Analysis From the Serious Outcomes Surveillance (SOS) Network of the Canadian Influenza Research Network (CIRN). Open Forum Infect. Dis. *3*.

Medcalf, L., Poole, E., Elton, D., and Digard, P. (1999). Temperature-Sensitive Lesions in Two Influenza A Viruses Defective for Replicative Transcription Disrupt RNA Binding by the Nucleoprotein. J. Virol. *73*, 7349–7356.

Mehle, A., and Doudna, J.A. (2008). An Inhibitory Activity in Human Cells Restricts the Function of an Avian-like Influenza Virus Polymerase. Cell Host Microbe *4*, 111–122.

Mena, I., Jambrina, E., Albo, C., Perales, B., Ortín, J., Arrese, M., Vallejo, D., and Portela, A. (1999). Mutational Analysis of Influenza A Virus Nucleoprotein: Identification of Mutations That Affect RNA Replication. J. Virol. *73*, 1186–1194.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. *142*, 334–347.

Miyazawa, A., Fujiyoshi, Y., Stowell, M., and Unwin, N. (1999). Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel11Edited by J. Karn. J. Mol. Biol. *288*, 765–786.

Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B., and Wilson, I.A. (2012). Organization of the Influenza Virus Replication Machinery. Science *338*, 1631–1634.

Moncorgé, O., Mura, M., and Barclay, W.S. (2010). Evidence for Avian and Human Host Cell Factors That Affect the Activity of Influenza Virus Polymerase. J. Virol. *84*, 9978–9986.

Mondal, A., Potts, G.K., Dawson, A.R., Coon, J.J., and Mehle, A. (2015a). Phosphorylation at the Homotypic Interface Regulates Nucleoprotein Oligomerization and Assembly of the Influenza Virus Replication Machinery. PLoS Pathog. *11*, e1004826.

Mondal, A., Potts, G.K., Dawson, A.R., Coon, J.J., and Mehle, A. (2015b). Phosphorylation at the Homotypic Interface Regulates Nucleoprotein Oligomerization and Assembly of the Influenza Virus Replication Machinery. PLoS Pathog. *11*, e1004826.

Morens, D.M., and Taubenberger, J.K. (2011). Pandemic influenza: certain uncertainties. Rev. Med. Virol. *21*, 262–284.

Morgan, C., Rose, H.M., and Moore, D.H. (1956). Structure and development of viruses observed in the electron microscope. J. Exp. Med. *104*, 171–182.

Mosley, R.T., Edwards, T.E., Murakami, E., Lam, A.M., Grice, R.L., Du, J., Sofia, M.J., Furman, P.A., and Otto, M.J. (2012). Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA. J. Virol. *86*, 6503–6511.

Müller, R., Poch, O., Delarue, M., Bishop, D.H.L., and 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.

Mullin, A.E., Dalton, R.M., Amorim, M.J., Elton, D., and Digard, P. (2004). Increased amounts of the influenza virus nucleoprotein do not promote higher levels of viral genome replication. J. Gen. Virol. *85*, 3689–3698.

Murti, K.G., Webster, R.G., and Jones, I.M. (1988). Localization of RNA polymerases on influenza viral ribonucleoproteins by immunogold labeling. Virology *164*, 562–566.

Naesens, L., Stevaert, A., and Vanderlinden, E. (2016). Antiviral therapies on the horizon for influenza. Curr. Opin. Pharmacol. *30*, 106–115.

Naffakh, N., Massin, P., Escriou, N., Crescenzo-Chaigne, B., and van der Werf, S. (2000). Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. J. Gen. Virol. *81*, 1283–1291.

Nakazawa, M., Kadowaki, S., Watanabe, I., Kadowaki, Y., Takei, M., and Fukuda, H. (2008). PA subunit of RNA polymerase as a promising target for anti-influenza virus agents. Antiviral Res. *78*, 194–201.

Neumann, G., Hughes, M.T., and Kawaoka, Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. EMBO J. *19*, 6751–6758.

Ng, A.K.-L., Chan, W.-H., Choi, S.-T., Lam, M.K.-H., Lau, K.-F., Chan, P.K.-S., Au, S.W.-N., Fodor, E., and Shaw, P.-C. (2012). Influenza Polymerase Activity Correlates with the Strength of Interaction between Nucleoprotein and PB2 through the Host-Specific Residue K/E627. PLoS ONE *7*, e36415.

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

Nilsson, B.E., te Velthuis, A.J.W., and Fodor, E. (2017). Role of the PB2 627 Domain in Influenza A Virus Polymerase Function. J. Virol. *91*, e02467-16.

Noble, E., Mathews, D.H., Chen, J.L., Turner, D.H., Takimoto, T., and Kim, B. (2011). Biophysical Analysis of Influenza A Virus RNA Promoter at Physiological Temperatures. J. Biol. Chem. *286*, 22965–22970.

Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H., and Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. Nature *439*, 490–492.

Noda, T., Sugita, Y., Aoyama, K., Hirase, A., Kawakami, E., Miyazawa, A., Sagara, H., and Kawaoka, Y. (2012). Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. Nat. Commun. *3*, ncomms1647.

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

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

O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P., and 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.

Ortin, J., and Martin-Benito, J. (2015). The RNA synthesis machinery of negative-stranded RNA viruses. Virology *479–480*, 532–544.

Osterhaus, A.D.M.E., Rimmelzwaan, G.F., Martina, B.E.E., Bestebroer, T.M., and Fouchier, R. a. M. (2000). Influenza B Virus in Seals. Science *288*, 1051–1053.

Palmeri, D., and Malim, M.H. (1999). Importin beta Can Mediate the Nuclear Import of an Arginine-Rich Nuclear Localization Signal in the Absence of Importin alpha. Mol. Cell. Biol. *19*, 1218–1225.

Pantelic, R.S., Meyer, J.C., Kaiser, U., Baumeister, W., and Plitzko, J.M. (2010). Graphene oxide: A substrate for optimizing preparations of frozen-hydrated samples. J. Struct. Biol. *170*, 152–156.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G.F., Wohlkönig, A., Ruf, A., Muyldermans, S., Hol, W.G.J., Kobilka, B.K., and Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. Nat. Protoc. *9*, 674–693.

Patel, A., Malinovska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., and Hyman, A.A. (2017). ATP as a biological hydrotrope. Science *356*, 753–756.

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

Paterson, D., te Velthuis, A.J.W., Vreede, F.T., and 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, J.T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., García-Sastre, A., and tenOever, B.R. (2010). Influenza A virus-generated small RNAs regulate the switch from transcription to replication. Proc. Natl. Acad. Sci. *107*, 11525–11530.

Perez, J.T., Zlatev, I., Aggarwal, S., Subramanian, S., Sachidanandam, R., Kim, B., Manoharan, M., and tenOever, B.R. (2012). A Small-RNA Enhancer of Viral Polymerase Activity. J. Virol. *86*, 13475–13485.

Pflug, A., Guilligay, D., Reich, S., and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. Nature 516, 355.

Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S., and Cusack, S. (2017). Structural insights into RNA synthesis by the influenza virus transcription-replication machine. Virus Res.

Pinto, L.H., and Lamb, R.A. (2006). The M2 Proton Channels of Influenza A and B Viruses. J. Biol. Chem. *281*, 8997–9000.

Plotch, S.J., Bouloy, M., Ulmanen, I., and Krug, R.M. (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., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. *8*, 3867–3874.

Pons, M.W., Schulze, I.T., and Hirst, G.K. (1969). Isolation and characterization of the ribonucleoprotein of influenza virus. Virology *39*, 250–259.

Poon, L.L.M., Pritlove, D.C., Fodor, E., and Brownlee, G.G. (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.

Portela, A., and Digard, P. (2002). The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. J. Gen. Virol. *83*, 723–734.

Presti, R.M., Zhao, G., Beatty, W.L., Mihindukulasuriya, K.A., Travassos da Rosa, A.P.A., Popov, V.L., Tesh, R.B., Virgin, H.W., and Wang, D. (2009). Quaranfil, Johnston Atoll, and Lake Chad Viruses Are Novel Members of the Family Orthomyxoviridae. J. Virol. *83*, 11599–11606.

Pritlove, D.C., Poon, L.L.M., Fodor, E., Sharps, J., and Brownlee, G.G. (1998). Polyadenylation of Influenza Virus mRNA Transcribed In Vitro from Model Virion RNA Templates: Requirement for 5′ Conserved Sequences. J. Virol. *72*, 1280–1286.

Pritlove, D.C., Poon, L.L.M., Devenish, L.J., Leahy, M.B., and Brownlee, G.G. (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.

Rameix-Welti, M.-A., Tomoiu, A., Afonso, E.D.S., Werf, S. van der, and 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.

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

Rasmussen, S.G.F., Choi, H.-J., Fung, J.J., Pardon, E., Casarosa, P., Chae, P.S., DeVree, B.T., Rosenbaum, D.M., Thian, F.S., Kobilka, T.S., et al. (2011). Structure of a nanobody-stabilized active state of the β2 adrenoceptor. Nature *469*, 175–180.

Reber, A.J., Chirkova, T., Kim, J.H., Cao, W., Biber, R., Shay, D.K., and Sambhara, S. (2011). Immunosenescence and Challenges of Vaccination against Influenza in the Aging Population. Aging Dis. *3*, 68–90.

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H., et al. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature *516*, 361–366.

Reich, S., Guilligay, D., and Cusack, S. (2017). An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. Nucleic Acids Res. gkx043.

Resa-Infante, P., and Gabriel, G. (2013). The nuclear import machinery is a determinant of influenza virus host adaptation. BioEssays *35*, 23–27.

Robertson, J.S. (1979). 5′ and 3′ terminal nucleotide sequences of the RNA genome segments of influenza virus. Nucleic Acids Res. *6*, 3745–3758.

Robertson, J.S., Schubert, M., and Lazzarini, R.A. (1981). Polyadenylation sites for influenza virus mRNA. J. Virol. *38*, 157–163.

Rodriguez, A., Pérez-González, A., and Nieto, A. (2007). Influenza Virus Infection Causes Specific Degradation of the Largest Subunit of Cellular RNA Polymerase II. J. Virol. *81*, 5315–5324.

Ruigrok, R.W.H., and Baudin, F. (1995). Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. J. Gen. Virol. *76*, 1009–1014.

Russo, C.J., and Passmore, L.A. (2014). Ultrastable gold substrates for electron cryomicroscopy. Science *346*, 1377–1380.

Sandvik, T., Rimstad, E., and Mjaaland, S. (2000). The viral RNA 3'- and 5'-end structure and mRNA transcription of infectious salmon anaemia virus resemble those of influenza viruses. Arch. Virol. *145*, 1659–1669.

Sangawa, H., Komeno, T., Nishikawa, H., Yoshida, A., Takahashi, K., Nomura, N., and Furuta, Y. (2013). Mechanism of Action of T-705 Ribosyl Triphosphate against Influenza Virus RNA Polymerase. Antimicrob. Agents Chemother. *57*, 5202–5208.

Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Shapiro, G.I., and Krug, R.M. (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.

Shi, L., Summers, D.F., Peng, Q., and Galarza, J.M. (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.

Sikora, D., Rocheleau, L., Brown, E.G., and Pelchat, M. (2014). Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. Sci. Rep. *4*, 6181.

Sorrell, E.M., Schrauwen, E.J.A., Linster, M., De Graaf, M., Herfst, S., and Fouchier, R.A.M. (2011). Predicting "Airborne" Influenza Viruses: (Trans-) mission Impossible? Curr. Opin. Virol. *1*, 635–642.

Stark, H. (2010). GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. In Methods in Enzymology, (Elsevier), pp. 109–126.

Stegmann, T. (2000). Membrane Fusion Mechanisms: The Influenza Hemagglutinin Paradigm and its Implications for Intracellular Fusion. Traffic *1*, 598–604.

Stegmann, T., Morselt, H.W., Scholma, J., and Wilschut, J. (1987). Fusion of influenza virus in an intracellular acidic compartment measured by fluorescence dequenching. Biochim. Biophys. Acta BBA-Biomembr. *904*, 165–170.

Subbarao, E.K., and Murphy, B.R. (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., Kawaguchi, A., Okuwaki, M., and Nagata, K. (2015). pp32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. ELife *4*.

Swale, C., Monod, A., Tengo, L., Labaronne, A., Garzoni, F., Bourhis, J.-M., Cusack, S., Schoehn, G., Berger, I., Ruigrok, R.W., et al. (2016). Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. Sci. Rep. *6*, 24727.

Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat. Methods.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: An extensible image processing suite for electron microscopy. J. Struct. Biol. *157*, 38–46.

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W.H., Daigle, N., Ellenberg, J., et al. (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., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S., and Hart, D.J. (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, J.K., and Kash, J.C. (2010). Influenza Virus Evolution, Host Adaptation, and Pandemic Formation. Cell Host Microbe *7*, 440–451.

Taylor, K.A., and Glaeser, R.M. (2008). Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. J. Struct. Biol. *163*, 214–223.

Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F., et al. (2016). Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. Mol. Cell *61*, 125–137.

Thueng-in, K., Thanongsaksrikul, J., Srimanote, P., Bangphoomi, K., Poungpair, O., Maneewatch, S., Choowongkomon, K., and Chaicumpa, W. (2012). Cell Penetrable Humanized-VH/VHH That Inhibit RNA Dependent RNA Polymerase (NS5B) of HCV. PLoS ONE *7*, e49254.

Tomescu, A.I., Robb, N.C., Hengrung, N., Fodor, E., and Kapanidis, A.N. (2014). Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. Proc. Natl. Acad. Sci. *111*, E3335–E3342.

Tsai, T.F. (2013). Fluad®-MF59®-Adjuvanted Influenza Vaccine in Older Adults. Infect. Chemother. *45*, 159.

Tsurumura, T., Qiu, H., Yoshida, T., Tsumori, Y., Hatakeyama, D., Kuzuhara, T., and 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, J.W., Tiley, L.S., Fodor, E., and Vreede, F.T. (2013). The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. Nat. Commun. *4*, 1591.

Turrell, L., Hutchinson, E.C., Vreede, F.T., and Fodor, E. (2015). Regulation of Influenza A Virus Nucleoprotein Oligomerization by Phosphorylation. J. Virol. *89*, 1452–1455.

Urnavicius, L., Zhang, K., Diamant, A.G., Motz, C., Schlager, M.A., Yu, M., Patel, N.A., Robinson, C.V., and Carter, A.P. (2015). The structure of the dynactin complex and its interaction with dynein. Science *347*, 1441–1446.

Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., Plotnikova, M.A., and Kiselev, O.I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. Virus Res. *185*, 53–63.

te Velthuis, A.J.W., and Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nat. Rev. Microbiol. *14*, 479–493.

te Velthuis, A.J.W., Robb, N.C., Kapanidis, A.N., and Fodor, E. (2016). The role of the priming loop in influenza A virus RNA synthesis. Nat. Microbiol. *1*, 16029.

Vreede, F.T., and Brownlee, G.G. (2007). Influenza Virion-Derived Viral Ribonucleoproteins Synthesize both mRNA and cRNA In Vitro. J. Virol. *81*, 2196–2204.

Vreede, F.T., and Fodor, E. (2010). The role of the influenza virus RNA polymerase in host shut-off. Virulence *1*, 436–439.

Vreede, F.T., Jung, T.E., and Brownlee, G.G. (2004). Model Suggesting that Replication of Influenza Virus Is Regulated by Stabilization of Replicative Intermediates. J. Virol. *78*, 9568–9572.

Vreede, F.T., Chan, A.Y., Sharps, J., and Fodor, E. (2010). Mechanisms and functional implications of the degradation of host RNA polymerase II in influenza virus infected cells. Virology *396*, 125–134.

Vuillard, L., Madern, D., Franzetti, B., and Rabilloud, T. (1995). Halophilic protein stabilization by the mild solubilizing agents nondetergent sulfobetaines. Anal. Biochem. *230*, 290–294.

Wakai, C., Iwama, M., Mizumoto, K., and 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, T.S., Diprose, J.M., Mayo, C.J., Siebold, C., Pickford, M.G., Carter, L., Sutton, G.C., Berrow, N.S., Brown, J., Berry, I.M., et al. (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.

Wang, M., and Veit, M. (2016). Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. Protein Cell *7*, 28–45.

Wang, Q.M., Hockman, M.A., Staschke, K., Johnson, R.B., Case, K.A., Lu, J., Parsons, S., Zhang, F., Rathnachalam, R., Kirkegaard, K., et al. (2002). Oligomerization and Cooperative RNA Synthesis Activity of Hepatitis C Virus RNA-Dependent RNA Polymerase. J. Virol. *76*, 3865–3872.

Watanabe, T., Kawakami, E., Shoemaker, J.E., Lopes, T.J.S., Matsuoka, Y., Tomita, Y., Kozuka-Hata, H., Gorai, T., Kuwahara, T., Takeda, E., et al. (2014). Influenza Virus-Host Interactome Screen as a Platform for Antiviral Drug Development. Cell Host Microbe *16*, 795–805.

Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. Microbiol. Rev. *56*, 152–179.

Winter, G., Lobley, C.M.C., and Prince, S.M. (2013). Decision making in xia2. Acta Crystallogr. D Biol. Crystallogr. *69*, 1260–1273.

Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S., and Digard, P. (2009). A Complicated Message: Identification of a Novel PB1-Related Protein Translated from Influenza A Virus Segment 2 mRNA. J. Virol. *83*, 8021–8031.

Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., et al. (2012). Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain. PLoS Pathog. *8*, e1002998.

Wu, W.W.H., Sun, Y.-H.B., and Panté, N. (2007a). Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. Virol. J. *4*, 49.

Wu, W.W.H., Weaver, L.L., and Panté, N. (2007b). Ultrastructural Analysis of the Nuclear Localization Sequences on Influenza A Ribonucleoprotein Complexes. J. Mol. Biol. *374*, 910–916.

Yamada, S., Hatta, M., Staker, B.L., Watanabe, S., Imai, M., Shinya, K., Sakai-Tagawa, Y., Ito, M., Ozawa, M., Watanabe, T., et al. (2010). Biological and Structural Characterization of a Host-Adapting Amino Acid in Influenza Virus. PLoS Pathog. *6*, e1001034.

Yamayoshi, S., Watanabe, M., Goto, H., and Kawaoka, Y. (2016). Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. J. Virol. *90*, 444–456.

Yap, T.L., Xu, T., Chen, Y.-L., Malet, H., Egloff, M.-P., Canard, B., Vasudevan, S.G., and Lescar, J. (2007). Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution. J. Virol. *81*, 4753–4765.

York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T., and Fodor, E. (2013). Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. Proc. Natl. Acad. Sci. U. S. A. *110*, E4238–E4245.

Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., et al. (2009). Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. Nature *458*, 909–913.

Yuanji, G., Fengen, J., Ping, W., Min, W., and Jiming, Z. (1983). Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. J. Gen. Virol. *64*, 177–182.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J. Struct. Biol. *193*, 1–12.

Zhang, J., Baker, M.L., Schröder, G.F., Douglas, N.R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C.J., Levitt, M., Ludtke, S.J., et al. (2010). Mechanism of folding chamber closure in a group II chaperonin. Nature *463*, 379–383.

Zhao, C., Lou, Z., Guo, Y., Ma, M., Chen, Y., Liang, S., Zhang, L., Chen, S., Li, X., Liu, Y., et al. (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, H.A., Palese, P., and 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, S., Palovcak, E., Armache, J.-P., Cheng, Y., and Agard, D. (2016). Anisotropic correction of beam-induced motion for improved single-particle electron cryo-microscopy. BioRxiv 061960.

Appendix – Nanobody information

Nanobody	*Antigen selection on solid phase, coated with:*	*Positive in ELISA against:*	*Family*	*Occurrence*
8189	$FluPol_{Fj}$ crosslinked protein or unmodified	$FluPol_{Fj}$ unmodified	1	2
8201				
8193			5	2
8194			6	2
8197			7	1
8190	$FluPol_{Fj}$ crosslinked		2	1
8191			3	1
8192			4	1
8195	$FluPol_{Fj}$ unmodified		8	2
8196			9	1
8199			11	1
8203			14	1
8204			15	1
8205			16	1
8206			17	1
8207			18	1
8198	$FluPol_{Fj}$ unmodified or biotinylated		10	2
8212				
8200			12	2
8211				
8202			13	2
8210				
8208	$FluPol_{Fj}$ biotinylated		19	5
8209			20	1

Sequences

8195
QVQLQESGGGLVQAGGSLRLSCAASTRTFSIYTMGWFRQAPGKEREFVAAIAWNDQTRYAD
SVKGRFTISRDNAKNTVYLQMNSLKQEDTAVYYCAAGTYSEYENSYAYWGQGTQVTVSSH
HHHHHEPEA

8192
QVQLQESGGGLVQAGASLRLSCAASGRTFMGWFRQAPGQEREFVAAIDAFGINTYYADSAK
GRFTISRDSAKNTVYLQMNSLKPEDTAVYYCAATLRGTQPGAVMQWRPDYWGQGTQVTVS
SHHHHHHEPEA

8190
QVQLQESGGGLVQAGGSLRLSCAASGRSLSDYTMGWFRQAPGKEREFVTAITSDGVYANYA
DSVKGRFTISRDNAKNTAYLQMNSLKPEDTAVYYCAGKTRRSMLMTTGYDYWGQGTQVTV
SSHHHHHHEPEA

8199
QVQLQESGGGLVQAGGSLRLSCAASGRTFSSYAMGWFRQAPGKEREFVGAIDLSGLYKHYG
DSVKGRFTISRDNAKNTVYLQMSSLKPEDTAVYYCAAAGIYYAGSYSPLLYDYWGQGTQVT
VSSHHHHHHEPEA

8203
QVQLQESGGGLVQAGGSLRLSCAAPGRILTLSRYVMGWFRQAPGKEREFVAAISPGTTDTDY
ADSVKGRFTISRDNAKNTVYLEMSSLKPEDTAVYVCAAGAASSGTVYRVYGLYNYRGQGTQ
VTVSSHHHHHHEPEA

8202
QVQLQESGGGLVQAGESLRLSCAASGGTFIYYGMGWFRQAPGKEREFVASISQTGGDIYAMD
FVKGRFTISRDNSKNTLYLQMNSLKPEDTAVYYCAAATSPYASRKLYEYKYWGQGTQVTVS
SHHHHHHEPEA

8210
QVQLQESGGGLVQAGESLRLSCAASGGTLIYYGMGWFRQAPGKEREFVASISQAGGDIYAM
DFVKGRFTISRDNSKNTLYLQLNTLKPEDTAVYYCAAATSPYASRKLYDYKYWGQGTQVTV
SSHHHHHHEPEA

8207
QVQLQESGGGLVQAGASLRLSCTASGRTSSTYPMGWFRQGPGKEREFVAAISWAGSSTYYG
DFVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCARAARPYAYGLDYSTEASYDYWGQG
TQVTVSSHHHHHHEPEA

8208
QVQLQESGGGLVQPGGSLRLSCAASGRTFSMYAMAWFRQAPGKQREFVGTISRSGDYALHA
DAVKGRFTISRDNAKNTVYLQMNSLKLEDTAVYYCAAGAYHSKDKTLYDYWGQGTQVTVS
SHHHHHHEPEA

8198
QVQLQESGGGLVQTGDSLRLSCAGAERSFNSYTMGWFRQAPGKEREFVAAIAFSGGFTYYAE
AVKGRFTISRDSGKNTVYLQMNSLKPEDTAVYYCAARFYGGGYYYQQREYSFWGQGTQVT
VSSHHHHHHEPEA

8212
QVQLQESGGGLVQPGGSLRLSCAASGRSFNSYTMGWFRQAPGKEREFVAAIAFSGGFTYYAE
AVKGRFTISRDNGEKPVYLRMNSLKPEDTAVYYCAARFYGGGYYYQEREYTDWGQGTQVT
VSSHHHHHHEPEA

8200
QVQLQESGGGLVQAGDSLRLSCAASGRTFSGYIMAWFRQTPGKERELVARLGPSSSIQYYRD
SVKGRFTLSRDNAKNTLYLQMNSLKPEDTAVYYCAFTLKWSSDYRSHNAYDYWGQGTQVT
VSSHHHHHHEPEA

8211
QVQLQESGGGLVQAGDSLRLSCAASGRTFSGYIMAWFRQTPGKEREFVARLGPSSNMQYYV
DSVKGRFTLSRDNAKNTLYLQMNSLKPEDTAVYYCVFTLKWSPDYRSHNAYDYWGQGTQV
TVSSHHHHHHEPEA

8196
QVQLQESGGGLVQAGDSLRLSCTTSGSSLSRYGVGWFRQAPGEEREFGASISWNGATYYADS
VKGRFTISRDNAKNTVYLQMNSLKAEDTAVYYCAARSDGYLTPYLSPVYWGQGTQVTVSSH
HHHHHEPEA

8193
QVQLQESGGGLVQPGGSLRLSCAASGFTFKMYAMSWVRQAPGKGLEWVSSINSAGGSTSYV
DSVKGRFTISRDNAKNTLYLQMNSLKPDDTAVYYCVQGRNWPYDYRGQGTQVTVSSHHHH
HHEPEA

8197
QVQLQESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSSITSAFGRTSYVS
SVKGRFTISRDNAKNTLYLQMNSLKPEDTAVYYCAQGASWPYRYRGQGTQVTVSSHHHHH
HEPEA

8194
QVQLQESGGGLVQPGGSLRLSCAASGFTFSNYYMSWVRQAPGEGREWVSSINRDGSNTYYA
DSVKGRFTIARDNVKNTLYLLMDSLKSDDTAVYYCTKGSNWPYDNWGQGTQVTVSSHHHH
HHEPEA

8205
QVQLQESGGGMVQPGGSLRLSCLASGFTFSNYAMTWVRQAPGKGPEWVSMVSNNGADTTY
TDSVKGRFTISRDNAKNTLYLRMNNVKPEDSAVYYCAKRRYGGIWTGQPTDYDYLGQGTVT
VSSHHHHHHEPEA

8209
QVQLQESGGGLVQPGGSLRLSCAASGNIRSINTMSWYRQAPGKERELVASDSTGGSTGGSTV
YAASVKGRFTISRDNAKNTVYLQMNMLKPEDTAVYYCYVYGMSRFGSTISTLFGQGTQVTV
SSHHHHHHEPEA

8189
QVQLQESGGGLVQPGGSLRLSCAASGSDFSLYAMGWYRQVPGKERELVATITSGGTTNFADS
ARGRFTISRGNAKNTVYLQMNNLKPEDTAVYYCWSRGSYGRYLETWGQGTQVTVSSHHHH
HHEPEA

8201
QVQLQESGGGLVQPGGSLRLSCAASGSDFSLYAMGWYRQVPGKERELVATITIGGTTNFADS
ARGRFTISRGNAKNTVYLQMNNLKPEDTAVYYCWSRGSYGRYLETWGQGTQVTVSSHHHH
HHEPEA

8206
QVQLQESGGGLVQPGGSLRLSCTASGGIFGISAMGWYRQAPGKERELVARITRSGSTNFADSV
KDRFTISRDATENTVVLQMNSLKAEDTAVYYCNGQFWEVDNGNIDLWGQGTQVTVSSHHH
HHHEPEA

8204
QVQLQESGGGLVQPRGSLRLSCAASGSIAIIYTMGWYRQAPLKERELVATVTYGGNTNYADP
VKGRFTISRDNNKNTEYLQMSSLQPEDTAVYYCYMRVRSTAWDTDDYWGQGTQVTVSSHH

HHHHEPEA

8191
QVQLQESGGGLVQPGGSLRLSCTASESVASINIVGWYRQISGKERELVARISSGGTITDYSDSV
KGRFTITRDNTKNTVYLQMNSLKRGDTAVYYCNAEYQYGSDWFHIWGQGTQVTVSSHHHH
HHEPEA

Sequence alignment

