INFLUENZA POLYMERASE INTERACTION WITH ITS HOST TRANSCRIPTIONAL PARTNERS



Alaa Baazaoui

St John's College
and
Sir William Dunn School of Pathology
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2024

Acknowledgements

It has been a great privilege to embark on this deeply transformative educational journey over the past years. As a traveller on one of the many paths of learning, I am deeply thankful to have had a wonderful mentor and experienced guide, Professor Ervin Fodor, whose enthusiasm and curiosity have been a constant source of inspiration. I thank him for his unwavering support, trust, and guidance in helping me reach this milestone in my journey. I would also like to extend my sincere thanks to my second supervisor, Professor Jonathan Grimes, for his support on the structural side of the project and for always being available when needed. Special thanks are also due to Dr Jeremy Keown for an excellent collaboration, the best coffee breaks, and for helping me grow as a researcher. My "doctoral grandfather," Professor George Brownlee, has been a source of inspiration since my first days in the Fodor lab. His insightful questions and advice will remind me never to stop learning and remain curious.

I am deeply indebted to the support from the Sir William Dunn School of Pathology throughout my DPhil, particularly from Professor Anton van der Merwe and Lucinda Risius, who have played a pivotal role in making my time as a DPhil student a truly wonderful experience in such an incredible department.

I also extend my heartfelt thanks to my academic mentors who supported and guided me at crucial points in my journey — Professor Esther Becker, my first mentor in Oxford, and Professors Günter Schwarz and Stefan Höning who first sparked my interest in medical biochemistry during my early days in Cologne. My gratitude also goes to Frau Brandes, Frau Ringhoff, and Frau Luz, who fostered my love for biochemistry even before I knew it existed as a university subject.

This endeavour would not have been possible without an exceptional research team, composed of both current and former Fodorians and Grimesters. I would like to express my deepest gratitude to Jane Sharps, the pillar of the lab, whose endless knowledge and encouragement have been invaluable. I am also immensely grateful for working with Dr Haitian Fan, whose generous sharing of knowledge and patience has made him not only a remarkable scientist but also a role

model. The “food for thought” and extraordinary cakes from Dr Kuang-Yu Chen nourished both my mind and soul throughout this DPhil.

In preparing this thesis, I also benefited from the use of AI-based tools for editing, proofreading, and paraphrasing, which provided valuable assistance. In addition, my sincere thanks go to my colleague and friend Jack Whitehead for his meticulous proofreading and feedback during the final stages of writing.

In the daily hustle of lab life and during the challenging times of the COVID-19 pandemic, my labmates and friends made all the difference. My deepest thanks go to Linnea Drexhage, Alison Rep, Hana Veler, Jack Whitehead, and Amanda Zihan Zhu for their friendship and support. I am also eternally grateful to my friends from St John’s College and beyond. Thank you for everything, Archana, Björn, Blerina, Charlotte, David, Filippos, Hala, Heba, Lilli, Lisa, Lorika, Luca, Maja, Mona, Munazza, Romain, Sophie, Valeta, Yusra, and Zeynep.

Finally, my deep gratitude goes to my family and childhood friends who have always been there for me, whose unconditional love and support made me who I am today. Thank you, Maryam, Safia, Safa, Takwa, and Zaineb for your sisterhood. I am forever grateful for Tante Claudia whose patience and support were a long-term investment in my future. My siblings Fatima, Ahmed, Youssef, and Younes who are a constant source of fun, dreams, and the occasional reality check. My love and gratitude extend to my grandparents Nanna, Jaddi, Jedda; their good wishes and prayers have always been with me.

Mama and Baba – thank you for all your sacrifices and your boundless love and support throughout my journey of learning and growth.

Declaration

I, Alaa Baazaoui, hereby declare that the research presented in this thesis is a product of my own work, except where specifically stated and acknowledged.

Alaa Baazaoui
Oxford, October 2024

Abstract

Influenza virus infections pose a significant burden on public health systems, leading to seasonal outbreaks and recurring pandemics caused by novel zoonotic strains crossing into humans. As segmented negative-strand RNA viruses, influenza viruses package their genome consisting of negative-sense RNA into viral ribonucleoprotein complexes (vRNPs). To transcribe and replicate their genome, influenza viruses hijack the host cellular machinery, and they rely on the interaction with the host RNA polymerase (Pol II) to produce viral RNA. The viral RNA-dependent RNA polymerase (FluPol) binds to the Ser5 phosphorylated (pS5) form of the C-terminal domain (CTD) of Pol II, enabling the virus to 'steal' RNA primers from nascent host RNA in a process called "cap-snatching." Additionally, FluPol relies on other host factors to facilitate viral RNA synthesis and regulate its activity. This thesis investigates the interaction between FluPol and the host transcriptional machinery, with a particular focus on the FluPol-Pol II CTD interaction that is essential for cap-snatching.

Structural and functional characterisation of 1918 (H1N1) FluPol A bound to a synthetic pS5 CTD peptide mimicking the 52 heptad repeats in mammalian Pol II CTD, revealed a continuous CTD binding site at the C-terminal domain of the viral polymerase subunit PA (PA-C). The structure also uncovered a novel arrangement of the 627 domain of the viral PB2 subunit, suggesting a role of the PB2 subunit in mediating the FluPol-Pol II interaction. Functional analysis of mutations within the CTD binding site of the PA-C demonstrated defects in transcription and replication, indicating the importance of PA-C in viral transcription and replication. To explore additional CTD binding sites on FluPol, recombinant CTD peptide mimics composed of 26 heptad repeats (CTD_{26x}) of the Pol II CTD consensus sequence YSPTSPS were generated. Biochemical experiments showed that phosphorylated CTD_{26x} ($pCTD_{26x}$) peptides bind strongly to FluPol from influenza A, B, and C, promoting FluPol oligomerisation, which may influence viral transcription and replication. Cryo-EM analysis of native Pol II bound to 1918 FluPol A revealed a binding path for the CTD similar to the synthetic pS5 CTD peptide in the PA-C domain. Further analysis of different phosphorylation patterns on the CTD, combining S2, S5, and S7 phosphorylations,

demonstrated that FluPol selectively recognises pS5 CTD and that additional phosphorylations disrupt CTD binding. Moreover, the FluPol-CTD binding interface is also influenced by the presence of host factors such as ANP32B, which share overlapping binding sites with the CTD, raising important questions about the spatiotemporal regulation of viral transcription and replication.

In conclusion, this thesis provides novel insights into how influenza viruses exploit the host transcription machinery by investigating the molecular details of the FluPol-Pol II interaction and opening new avenues for the development of antiviral therapies targeting this critical interface.

Contents

List of Figures **xi**

List of Tables **xiii**

List of Abbreviations **xiv**

1 Introduction **1**

1.1 Overview of influenza viruses 2

1.1.1 General background on influenza viruses 2

1.1.2 Influenza virus classification and phylogenetics 5

1.1.3 Structure of the influenza virion 6

1.1.4 Overview of the influenza virus life cycle 8

1.2 Structure and function of the influenza virus polymerase 12

1.2.1 Structural features of the influenza virus polymerase . . . 12

1.2.2 Influenza virus transcription 17

1.2.3 Influenza virus replication 23

1.3 Interaction of the influenza virus polymerase with the host transcriptional apparatus 28

1.3.1 The role of post-translational modifications on the Pol II CTD 28

1.3.2 Overview of Pol II transcription 31

1.3.3 Interaction of influenza virus polymerase with the host transcription machinery 34

1.3.4 Impact of influenza virus infection on host transcription . . 36

1.4 Aims of this study 38

2 Materials and Methods **39**

2.1 Materials and reagents 40

2.1.1 DNA and RNA oligonucleotides 40

2.1.2 Plasmids 42

2.1.3 Synthetic RNA polymerase II CTD mimic peptides 43

2.1.4 Antibodies 44

2.2 Molecular cloning 45

2.2.1 Competent cells and bacterial culture 45
2.2.2 Transformation 45
2.2.3 Preparation of plasmid DNA 46
2.2.4 PCR and site-directed mutagenesis 46
2.2.5 Gibson Assembly 46
2.2.6 Mutagenesis of pCAGGS plasmids 47
2.3 Cell and virus culture 47
2.3.1 Mammalian cell culture 47
2.3.2 Insect cell culture 48
2.3.3 Influenza virus propagation 48
2.4 Protein expression and purification 49
2.4.1 Insect cell expression and purification of influenza virus polymerase 49
2.4.2 Bacterial expression and purification of Pol II CTD-derived peptides 50
2.4.3 *In vitro* phosphorylation and purification of Pol II CTD-derived peptides 50
2.4.4 Insect cell expression and purification of CDK7 51
2.4.5 Purification of native Pol II 51
2.4.6 Mammalian expression and purification of ALFA-tagged Pol II 52
2.4.7 Bacterial expression and purification of DSIF 53
2.4.8 Bacterial expression and purification of GST-Spt5$_{\mathrm{CTR}}$ . . . 54
2.5 Protein analysis 54
2.5.1 Protein concentration measurement 54
2.5.2 SDS polyacrylamide gel electrophoresis and gel staining . . 55
2.5.3 Western blotting 55
2.6 Protein interaction analysis 56
2.6.1 Size-exclusion chromatography 56
2.6.2 CTD-binding assay 57
2.6.3 GST pulldown assays 57
2.6.4 ANP32 and NEP competition assays 58
2.6.5 Proteomics analysis of FluPol interactors from nuclear extract 58
2.7 Cell-based analysis of influenza viral polymerase 60
2.7.1 Mini genome assay 60
2.7.2 Split-luciferase assay 60
2.8 *In vitro* influenza virus assays 61
2.8.1 Cap-dependent transcription activity assay 61
2.9 RNA analysis 61

2.9.1 Extraction of RNA 61
2.9.2 Radiolabelling of oligonucleotides 62
2.9.3 Primer extension assay 62
2.9.4 Denaturing polyacrylamide gel electrophoresis 63
2.9.5 RNP reconstitution assay 63
2.10 Cryo-electron microscopy 64
2.10.1 Cryo-EM sample preparation 64
2.10.2 Cryo-EM data collection and processing 65

3 Structural and functional characterisation of the Pol II CTD binding to the PA subunit of the 1918 influenza A virus polymerase 68
3.1 Introduction 69
3.2 Results 71
3.2.1 Structure of the 1918 pandemic FluPol A bound to Pol II CTD mimic peptide 71
3.2.2 Arrangement of PB2-C in the Pol II CTD-bound influenza virus polymerase 76
3.2.3 Effect of mutations in the PA-C Pol II CTD binding site on polymerase function 79
3.3 Discussion 87

4 Structural characterisation of Pol II CTD binding to the influenza virus polymerase 93
4.1 Introduction 94
4.2 Results 96
4.2.1 Generation of recombinant phosphomimetic Pol II CTD mimetic peptides 96
4.2.2 Generation of recombinant phosphorylated Pol II CTD mimetic peptides 98
4.2.3 Phosphomimetic and phosphorylated CTD_{26x} peptides bind differently to viral polymerases from IAV, IBV, and ICV . 101
4.2.4 Phosphomimetic S5E CTD_{26x} mimetic peptide forms a complex with the influenza C virus polymerase 107
4.2.5 Phosphorylated Pol II CTD_{26x} mimetic peptide forms complexes with the influenza A, B, and C virus polymerase . . 108
4.2.6 Structural characterisation of the influenza virus polymerase interaction with recombinant Pol II CTD_{26x} mimetic peptide 112
4.2.7 Structural characterisation of the influenza virus polymerase interaction with native Pol II 114
4.3 Discussion 123

5 Functional characterisation of Pol II CTD binding to the influenza virus polymerase 132
5.1 Introduction . . . 133
5.2 Results . . . 135
5.2.1 The influenza virus polymerase strongly recognises pS5 Pol II CTD over alternative phosphorylation patterns . . . 135
5.2.2 CTD binding enhances the transcriptional activity of influenza virus polymerase C *in vitro* . . . 137
5.2.3 The PB2-C domain does not bind to Pol II CTD mimics *in vitro* . . . 142
5.3 Discussion . . . 147

6 Characterisation of host factors modulating the FluPol-Pol II CTD interaction 153
6.1 Introduction . . . 154
6.2 Results . . . 157
6.2.1 ANP32B competes with pS5 Pol II CTD for FluPol A binding 157
6.2.2 NEP does not outcompete pS5 Pol II CTD for FluPol A binding . . . 158
6.2.3 The Spt5 subunit of the DSIF complex interacts with 1918 FluPol A . . . 160
6.2.4 Cells expressing ALFA-tagged Pol II could be a useful tool for studying the interaction of FluPol with its host transcriptional partners . . . 166
6.3 Discussion . . . 168

7 Conclusion and future directions 173

References 179

Appendices

A Appendix 199

List of Figures

1.1 Influenza virus architecture and genome segments 8
1.2 Overview of the influenza virus life cycle 9
1.3 Structural features of the viral polymerase 14
1.4 Viral transcription initiation 19
1.5 Influenza virus polymerase transcription cycle 20
1.6 Structural details of synthetic Pol II CTD peptides bound to FluPol. 22
1.7 Replication initiation of vRNA and cRNA templates. 25
1.8 Post-translational modifications of the Pol II CTD 29
1.9 Overview of Pol II transcription 32

3.1 Structure of pS5 CTD-bound 1918 FluPol A 72
3.2 Cryo-EM processing scheme 73
3.3 Comparison of 1918 FluPol A CTD binding and H7N9 FluPol A . 75
3.4 Structure of pS5 CTD-bound 1918 FluPol A with resolved PB2-C domains 77
3.5 Map of PB2-C domain using DeepEMhancer 79
3.6 Multiple sequence alignment of PA 80
3.7 RNP reconstitution assay on the NA segment 81
3.8 RNP reconstitution assay on the NS segment 83
3.9 Effect of mutation in the PA subunit on ANP32B binding 84
3.10 Effect of mutation in the PA subunit on FluPol activity 86
3.11 Comparison of the Pol II CTD and ANP32 binding sites on the PA-C domain. 92

4.1 Purification of recombinant CTD_{13x} and CTD_{26x} peptides 97
4.2 Purification of CDK7 kinase module 99
4.3 Phosphorylation of CTD_{26x} peptide using a CDK7 kinase module 100
4.4 Pulldown of 1918 FluPol A with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide 102
4.5 Pulldown of FluPol B with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide 103
4.6 Pulldown of FluPol C with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide 104

4.7 SEC analysis of phosphomimetic S5E CTD_{26x} with FluPol C . . . 108
4.8 SEC analysis of $pCTD_{26x}$ with 1918 FluPol A . . . 109
4.9 SEC analysis of $pCTD_{26x}$ with FluPol B . . . 110
4.10 SEC analysis of $pCTD_{26x}$ with FluPol C . . . 111
4.11 Cryo-EM data collection on FluPol A and FluPol C in complex with $pCTD_{26x}$ . . . 113
4.12 Purification of mammalian Pol II . . . 115
4.13 Cryo-EM 2D classification of a 1918 FluPol A-Pol II complex . . . 119
4.14 Cryo-EM analysis of a 1918 FluPol A-Pol II complex . . . 120
4.15 Cryo-EM data collection on cross-linked FluPol-Pol II complex . . 122

5.1 CTD peptide design and CTD binding assays with FluPol A, FluPol B, and FluPol C . . . 136
5.2 Transcription activity of FluPol A in the presence of synthetic and recombinant CTD mimic peptides . . . 138
5.3 Transcription activity of FluPol C in the presence of synthetic and recombinant CTD mimic peptides . . . 139
5.4 Mobility of recombinant CTD mimic peptides in transcription activity assays using different formamide dye ratios . . . 140
5.5 Pulldown of 627-NLS and PB2-C (Viet) with unphosphorylated and $pCTD_{26x}$ peptide . . . 143
5.6 CTD binding assay with GST-627, 627-NLS and PB2-C (Viet) domains . . . 145
5.7 CTD binding assay with IAV GST-PB2-C (NT60) and GST-PB2-C (Viet) domains . . . 145

6.1 Competition assay of ANP32B and pS5 CTD_{4x} binding to 1918 FluPol A . . . 158
6.2 Competition assay of NEP and pS5 CTD_{4x} binding to 1918 FluPol A 160
6.3 Purification of GST-$Spt5_{CTR}$ . . . 161
6.4 Pulldown of 1918 FluPol A with GST-$Spt5_{CTR}$ . . . 162
6.5 Interaction of truncated GST-$Spt5_{CTR}$ with 1918 FluPol A . . . 163
6.6 Purification of the DSIF complex . . . 164
6.7 SEC analysis of FluPol A with the DSIF complex . . . 165
6.8 Purification of ALFA-tagged Pol II . . . 167

List of Tables

2.1 DNA oligonucleotides 40
2.2 RNA oligonucleotides 42
2.3 Plasmids 42
2.4 Synthetic Pol II CTD mimic peptides 44
2.5 Antibodies 44

A.1 Cryo-EM data collection and model refinement statistics. 200

List of Abbreviations

1918 . . . . . Influenza A/Brevig Mission/1/1918 (H1N1) virus

Å . . . . . Ångström

ADP . . . . . Adenosine diphosphate

ALFA . . . . . Artificial short epitope tag

ANP32 . . . . Acidic nuclear phosphoprotein 32

apo . . . . . RNA-free form of FluPol

APS . . . . . Ammonium persulfate

ATP . . . . . Adenosine triphosphate

BSA . . . . . Bovine serum albumin

CBC . . . . . Cap-binding complex

CBD . . . . . Cap-binding domain

CDK . . . . . Cyclin-dependent kinase

cDNA . . . . . Complementary DNA

CE . . . . . Capping enzyme

Ci . . . . . Curie

Crm1 . . . . . Cytoplasm via the exportin 1

cRNA . . . . . Complementary ribonucleic acid

cRNP . . . . . Complementary ribonucleoprotein

cryo . . . . . Cryogenic

CTR . . . . . C-terminal region

CTD . . . . . C-terminal domain

CTD_{4x} . . . . . CTD mimic peptide with 4 heptad consensus repeats

CTD_{13x} . . . . CTD mimic peptide with 13 heptad consensus repeats

CTD_{26x} . . . . CTD mimic peptide with 26 heptad consensus repeats

CTF . . . . . Contrast transfer function

CTP . . . . . Cytidine triphosphate

DKO . . . . . Double knockout

DMEM . . . . Dulbecco's modified Eagle medium

DMSO . . . . . Dimethyl sulfoxide

DNA . . . . . Deoxyribonucleic acid

dNTP . . . . . Deoxynucleoside triphosphate

DSIF . . . . . DRB sensitivity-inducing factor

DTT . . . . . Dithiothreitol

ECL . . . . . Enhanced chemiluminescence

E. coli . . . . . *Escherichia coli*

EDTA . . . . . Ethylenediaminetetraacetic acid

EM . . . . . Electron microscopy

EMDB . . . . Electron Microscopy Data Bank

eV . . . . . Electronvolt

FCS . . . . . Foetal calf serum

FluPol . . . . . Influenza virus RNA-dependent RNA polymerase

g . . . . . Standard gravity

GAPDH . . . Glyceraldehyde 3-phosphate dehydrogenase

GFP . . . . . Green fluorescent protein

GSH . . . . . Glutathione

GTFs . . . . . General transcription factors

GTP . . . . . Guanosine triphosphate

GST . . . . . Glutathione S-transferase

HA . . . . . Hemagglutinin

HEF . . . . . Hemagglutinin-esterase-fusion

HEK293T . . Human Embryonic Kidney 293T cells

HEPES . . . . 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HCT116 . . . . Human Colorectal Carcinoma cell line

HRP . . . . . Horseradish peroxidase

IAV . . . . . Influenza A virus

IBV . . . . . Influenza B virus

ICV . . . . . Influenza C virus

IDV . . . . . Influenza D virus

IPTG . . . . . Isopropyl -d-1-thiogalactopyranoside

kDa . . . . . Kilo Dalton

KOW . . . . . Kyprides Ouzounis Woese motif

LB . . . . . Lysogeny broth

LCAR . . . . . Low-complexity acidic region

LC-MS/MS . Liquid chromatography with tandem mass spectrometry

LRR . . . . . Leucine-rich repeat

MDBK . . . . Madin-Darby Bovine Kidney

MDCK . . . . Madin-Darby Canine Kidney

M1 . . . . . Matrix 1 protein (influenza virus)

M2 . . . . . Matrix 2 ion channel (influenza virus)

m7G . . . . . N7-methyl guanosine

mAu . . . . . Milli absorbance unit

$MgCl_2$ . . . . . Magnesium chloride

min . . . . . Minute

MOI . . . . . Multiplicity of infection

mRNA . . . . Messenger RNA

NA . . . . . Neuraminidase

NaCl . . . . . Sodium chloride

NaOH . . . . . Sodium hydroxide

NELF . . . . . Negative elongation factor

NEP . . . . . Nuclear export protein

Ni-NTA . . . . Nickel-chelated nitrilotriacetic acid

NLS . . . . . Nuclear localisation signal

NLR . . . . . Normalised luminance ratio

NP . . . . . Nucleoprotein

NS1 . . . . . Non-structural protein

nt . . . . . Nucleotide

NT60 . . . . . Influenza A/Northern Territory/60/1968 (H3N2) virus

NTP . . . . . Nucleotide triphosphate

OD_{600} . . . . . Optical density at 600 nm

ORFs . . . . . Open reading frames

OTG . . . . . Octylthioglucoside

P3 . . . . . Polymerase protein 3

PA . . . . . Polymerase acidic protein

PA-C . . . . . C-terminal domain of PA

PA-endo . . . Endonuclease domain of PA

PAGE . . . . . Polyacrylamide Gel Electrophoresis

PAF . . . . . Positive elongation factor PAF1 complex

PARP1 . . . . Poly(ADP)-ribose 1

PB1 . . . . . Polymerase basic protein 1

PB2 . . . . . Polymerase basic protein 2

PB2-C . . . . . C-terminal domain of PB2

PB2-N . . . . . N-terminal domain of PB2

PCR . . . . . Polymerase chain reaction

pCTD26x . . . Phosphorylated CTD26x

PDB . . . . . Protein Data Bank

PIC . . . . . Pre-initiation complex

Pol II . . . . . RNA polymerase II

poly(A) . . . . Polyadenylic acid

poly(U) . . . . Polyuridylic acid

PMSF . . . . . Phenylmethylsulfonyl fluoride

P-TEFb . . . . Positive transcription elongation factor b

PTMs . . . . . Posttranslational modifications

pS2 . . . . . Phosphorylated serine 2

pS5 . . . . . Phosphorylated serine 5

pS7 . . . . . Phosphorylated serine 7

RPAP2 . . . . RNA polymerase associated protein 2

RPB1-12 . . . RNA polymerase II subunit

RNGTT . . . . RNA guanylyltransferase

RLU . . . . . Relative light units

rpm . . . . . Revolutions per minute

RNA . . . . . Ribonucleic acid

RNaseA . . . . Ribonuclease

RNAPII . . . RNA Polymerase II

RNMT . . . . RNA guanine-7 methyltransferase

RPB1-12 . . . RNA polymerase II subunits 1-12

RT . . . . . Reverse transcription or room temperature

S5E CTD . . . Phosphomimetic CTD peptide

SDS . . . . . Sodium dodecyl sulfate

SEC . . . . . Size exclusion chromatography

s.e.m. . . . . . Standard error of the mean

Sf9 . . . . . *Spodoptera frugiperda 9*

snRNAs . . . . Small nuclear RNAs

snoRNAs . . . Small nucleolar RNAs

Spt4 . . . . . Subunit of DSIF

Spt5 . . . . . Subunit of DSIF

SUMO . . . . Small ubiquitin-like modifier

TBE . . . . . Tris, Boric acid, EDTA

TCEP . . . . . Tris(2 carboxyethyl)phosphine buffer

TEV . . . . . Tobacco Etch Virus

TEMED . . . Tetramethylethane-1,2-diamine

TFIIH . . . . . Transcription factor II H

Tris . . . . . Tris(hydroxymethyl)aminomethane

TSS . . . . . Transcription start site

U . . . . . Enzyme unit

UTP . . . . . Uridine triphosphate

UTR . . . . . Untranslated region

UV . . . . . Ultraviolet

Viet . . . . . Influenza A/Viet Nam/1203/2004 (H5N1) virus

vRNA . . . . . Viral ribonucleic acid

vRNP . . . . . Viral ribonucleoprotein

WSN . . . . . Influenza A/WSN/1933 (H1N1) virus

Xrn2 . . . . . 5'-3' exoribonuclease 2

YT . . . . . Yeast tryptone

$ZnCl_2$ . . . . . Zinc chloride

1 Introduction

1.1 Overview of influenza viruses

1.1.1 General background on influenza viruses

Influenza is a highly contagious respiratory disease caused by influenza viruses, leading to seasonal epidemics and occasional pandemics. Symptoms of influenza infection include fever, cough, sore throat, runny or stuffy nose, fatigue, and occasionally non-respiratory complications such as cardiovascular and neurological complications [1, 2]. Severe influenza infections can cause a secondary bacterial infection that may result in pneumonia, which can be fatal for high-risk groups such as the elderly, immunocompromised individuals, young children, or those with chronic health conditions [1, 3]. The impact of influenza viruses on public health and healthcare systems is substantial, globally resulting in approximately 1 billion infections and an estimated 290,000 to 650,000 deaths per year [1, 4].

Influenza A viruses (IAV) lead to repeated, seasonal outbreaks and a regular occurrence of global pandemics both of which impose a substantial burden of illness and death. Over the past century, several pandemics have occurred, including the 1918 H1N1 'Spanish flu', the 1957 H2N2 'Asian flu', the 1968 H3N2 'Hong Kong flu', and the 2009 H1N1 pandemic [1]. The 1918 flu pandemic caused at least tens of million deaths with estimates ranging from 20-100 million deaths and the 1918 IAV strain is the source of all other influenza viruses that have led to pandemics since [5, 6]. To effectively prevent future influenza outbreaks, it is essential to understand the molecular mechanisms that lead to the emergence of pathogenic strains. Influenza viruses pose significant challenges to their eradication due to their strong evolutionary drive caused by antigenic shift and antigenic drift [1, 7, 8].

Influenza viruses can evade detection by the host immune system through a process called antigenic drift, which involves mutations in surface antigens like HA (H) and NA (N) which serve for their classification [1, 8]. The high mutation rate of the viral genome, caused by error-prone RNA synthesis, means that

influenza vaccines need to be updated annually [9]. Despite the slow changes in the genetic makeup of the virus, there is still a possibility of a pandemic occurring due to antigenic drift [10]. The segmented RNA genome of influenza viruses facilitates the swapping of gene segments between different strains, enabling the acquisition of new genetic combinations through reassortment. This phenomenon, known as antigenic shift, is responsible for the emergence of novel influenza A subtypes that have the potential to trigger pandemics [11]. Furthermore, influenza viruses, particularly IAV, have a broad host range which allows them to infect various hosts, including humans, birds, pigs, and bats. While the main reservoir of IAV is thought to be wild waterfowls and shorebirds, influenza viruses have crossed the interspecies barrier and caused zoonotic transmission to humans [12, 13]. H1, H2, and H3 IAV strains are examples of influenza viruses that have evolved to spread among humans [14]. Zoonotic strains usually cause more severe symptoms and are not adapted to replication in human cells leading to low human-to-human transmission [12]. However, the emergence of zoonotic strains poses a threat to global health. In rare cases, these strains can have high transmission rates between humans and cause severe diseases for which humans have no natural immunity. In summary, influenza A viruses are highly adaptable and can undergo reassortment events, which may lead to the emergence of novel strains with increased pathogenic and pandemic risk [15]. In addition, these risks are driven by the rapid evolution of influenza viruses through antigenic drift and shift, posing a significant challenge in their mitigation [1, 11].

Research efforts have contributed to the prevention and containment of influenza outbreaks through the development of vaccinations and antivirals [9, 16]. Influenza vaccines are updated each year based on the antigenic characteristics of circulating strains, to reduce the likelihood of infection and lessen the severity of disease if infection does occur. However, despite the efforts to update vaccines annually, the constantly evolving nature of influenza viruses poses a challenge in maintaining their effectiveness. A universal vaccine is still under development to provide

long-lasting protection against multiple influenza subtypes [17].

In addition to vaccination, antiviral drugs are commonly used for the treatment of influenza and are of particular importance when vaccines prove ineffective. Antivirals target different stages in the viral life cycle. For example, neuraminidase inhibitors like oseltamivir and zanamivir work by inhibiting the neuraminidase enzyme of the virus, thereby preventing the release of viral particles from infected cells [18]. M2 ion channel inhibitors, such as amantadine and rimantadine, disrupt the ability of the virus to create an acidic pH environment necessary for viral uncoating in the early stages of infection. This disruption prevents the release of viral ribonucleoproteins (vRNP) into the host cell cytoplasm [19].

Influenza virus polymerase (FluPol) inhibitors against all three polymerase subunits have been developed and are partly approved for clinical use. Baloxavir Marboxil inhibits the influenza cap-dependent endonuclease domain of the polymerase acidic (PA) subunit to perform cap-snatching, required for the generation of viral messenger RNA (mRNA) [20]. Favipiravir targets the polymerase basic protein 1 transcriptase function. It acts as a purine nucleoside analogue, leading to RNA chain termination and a higher mutation rate of FluPol by acting as a competitive substrate inhibitor for FluPol [21]. Pimodivir , which has been under clinical trials, inhibits the cap-binding function of the polymerase basic protein 2 (PB2) subunit of FluPol by mimicking the 7-methyl GTP (m^7GTP) on the 5'-end of the host pre-mRNA [22]. The narrow time window for drug administration limits the effectiveness of antivirals, and there are concerns about the emergence of drug resistance. Therefore, there is an urgent need to develop and approve effective influenza vaccines and antivirals to combat the imminent threat of influenza outbreaks [15].

1.1.2 Influenza virus classification and phylogenetics

Influenza viruses are RNA viruses with a segmented, negative-strand genome that belong to the *Orthomyxoviridae* virus family. Members of this virus family are characterised by their RNA-dependent RNA polymerase that is required for genome replication. Influenza viruses are categorised into four types, namely A, B, C, and D, which vary in their genome structure, host range, and pathogenicity [23, 24].

The influenza A virus genome consists of 8 gene segments. This influenza virus genus has the broadest host range, capable of infecting a wide range of mammals and birds which leads to the highest risk for zoonotic transmission and future pandemics. The classification of influenza A viruses into various subtypes is determined by the surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). Currently, 18 HA (H1-18) and 11 NA (N1-11) subtypes are known, leading to a wide range of possible HA and NA combinations [24]. For instance, H1N1 and H3N2 strains circulate during influenza season [25].

Influenza B viruses (IBV) have a smaller host range and no pandemic has been attributed to influenza B viruses so far. Even though IBV also have 8 gene segments like IAV, the amino acid identities have diverged during evolution, inhibiting reassortment events between IAV and IBV [26, 27]. IBV have only been isolated from humans and seals so far, indicating a narrower host range than IAV [28]. The classification of IBV is based on the antigenic properties of HA, dividing IBV into the Victoria and Yamagata lineages, the latter reported to have gone extinct during the COVID-19 pandemic [29, 30].

In 1947, the influenza C virus (ICV) was discovered during an epidemic and was subsequently categorised as a distinct genus of the *Orthomyxoviridae* as it did not demonstrate cross-reactivity with antisera against influenza A and B viruses. Humans are the primary reservoir for ICV, but the virus can also infect pigs

and dogs, typically causing mild illness [31]. Influenza D virus (IDV) which was first discovered in 2011 is related to ICV and infects mainly cattle and pigs [23, 32]. Both ICV and IDV exhibit differences from IAV and IBV in their genome structure, containing only 7 gene segments. Influenza C virus is distinct in the fact that it only contains one spike protein, the hemagglutinin-esterase-fusion glycoprotein (HEF). HEF encompasses multiple functions including receptor binding and destroying as well as membrane fusion, replacing the HA and NA gene segments in influenza A and B viruses [31].

1.1.3 Structure of the influenza virion

Influenza virions are enveloped and have diverse morphologies, including spherical, bacilliform, or filamentous forms (Figure 1.1 C-E). These variations can significantly impact their infectious capabilities, with filamentous forms potentially increasing virulence [33]. The diameter of a virion is estimated to be approximately 100 nm (Figure 1.1 E-F) [34].

The outer lipid membrane embeds the surface proteins hemagglutinin (HA), neuraminidase (NA), and the matrix 2 ion channel (M2) (Figure 1.1 A) [35]. HA mediates the initial attachment of the virus to sialic acid receptors on host cells facilitating viral entry. NA facilitates the release of progeny virions from the host cell by cleaving sialic acid residues while the M2 protein serves as a proton channel that enables the endosomal release of the viral ribonucleoprotein (vRNP) complexes into the cytoplasm [36]. The inside of the virion membrane is coated with matrix 1 protein (M1) supporting the stability and integrity of the virus particle [34]. Apart from these structural proteins, virions also contain high levels of proteins packaged within the particle, such as non-structural protein (NS1), a multifunctional protein that inhibits the host immune response as well as nuclear export protein which mediates vRNP trafficking across the cell [37–39]. Both NS1 and NEP, have also been shown to regulate viral transcription and replication [40].

Each virus particle packages the viral RNA genome segments into viral ribonucleoprotein complexes (vRNPs) together with nucleoprotein (NP) and the heterotrimeric viral polymerase (FluPol), consisting of the PA, PB1, and PB2 subunits (Figure 1.1 B) [41]. The vRNA is a single-stranded negative-sense RNA molecule with a length ranging from approximately 0.9 to 2.3 kilobases [42]. Each vRNA has one or two open reading frames (ORFs) and encodes up to two viral proteins per segment that play a crucial role in the virus's life cycle [43]. The ORFs are flanked by untranslated regions (UTRs), with a triphosphorylated 5' terminus and a non-polyadenylated 3' end. The 5' and 3' termini of the vRNA serve as promoter elements that together form a 'panhandle' structure and are selectively bound to FluPol while the rest of the vRNA is non-specifically covered by NP to protect the vRNA from the host cell environment [44, 45].

RNA viruses such as the influenza virus are under strong selective pressure to optimise the usage of the viral genome by producing multiple proteins from their mRNAs [49]. Therefore, alongside viral proteins, various regulatory proteins are expressed from the viral mRNA in the infected host cell by alternative splicing and non-canonical translation mechanisms with a single transcript giving rise to the translation of multiple viral protein variants [49–51]. IAV encodes for segment 1, PB2, and PB2-S1 [52]. It also produces various derivatives of PA for segment 3, including PA-X, PA-N155, and PA-N182 [37, 51]. The IAV and IBV encode up to 17 or 11 proteins, respectively, while ICV encodes 9 proteins [53, 54]. Influenza virus particles also contain hundreds of host-encoded proteins that were identified using proteomics approaches [37, 55–58]. It is assumed that these host proteins may play a role in the structure of influenza virions and the viral life cycle, however, their specific roles remain to be explored further [37].

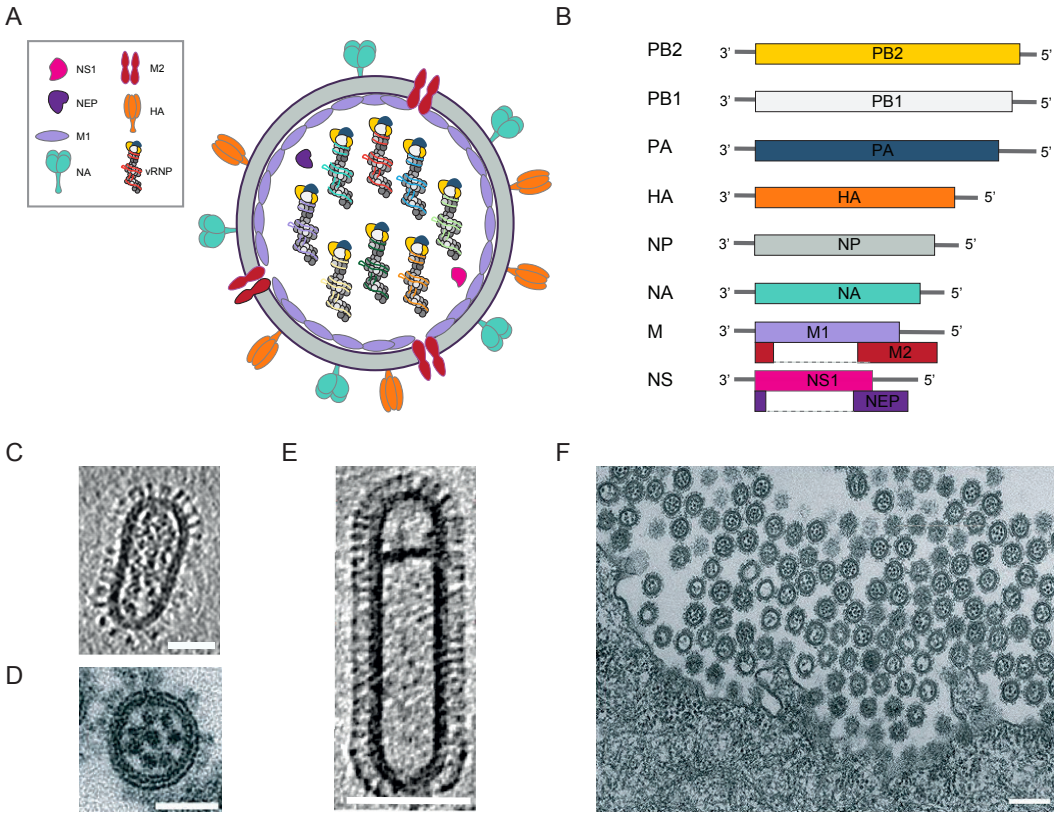

Figure 1.1: Influenza virus architecture and genome segments. **(A)** Schematic of an influenza virion and its protein components. **(B)** Overview of influenza A virus gene segments arranged according to their length in descending order and in 3' to 5' direction of the negative-sense vRNA; rectangles indicate ORFs. **(C-F)** Electron micrographs of influenza virions adopting filamentous, spherical, and bacilliform shapes. **(C)** Elongated influenza A/Aichi/68 (H3N2) virions with the vRNP complexes aligned with the long axis of the virion. Adapted from Harris et al. (2010) [46]. Scale bar: 50 nm. **(D)** Spherical influenza A/WSN/33 (H1N1) virus particle, the large dark circles show the vRNP complexes. Adapted from Noda et al.(2006) [47]. Scale bar: 50 nm. **(E)** Elongated influenza A/Aichi/68 (H3N2) particle. Adapted from Calder et al. (2010) [48]. Scale bar: 100 nm. **(F)** Budding virions of influenza A/WSN/33 (H1N1) strain. Adapted from Noda et al. (2006) [47]. Scale bar: 200 nm.

1.1.4 Overview of the influenza virus life cycle

Influenza virus HA binds to plasma membrane receptors decorated with sialic acid residues. Human influenza viruses specifically recognise sialic acid linked with galactose in an α2,6-bond found on the cell surface of epithelial cells of the upper respiratory tract. In birds, avian influenza viruses recognise surface receptors with α2,3-linked sialic acid on epithelial cells of the intestinal tract.

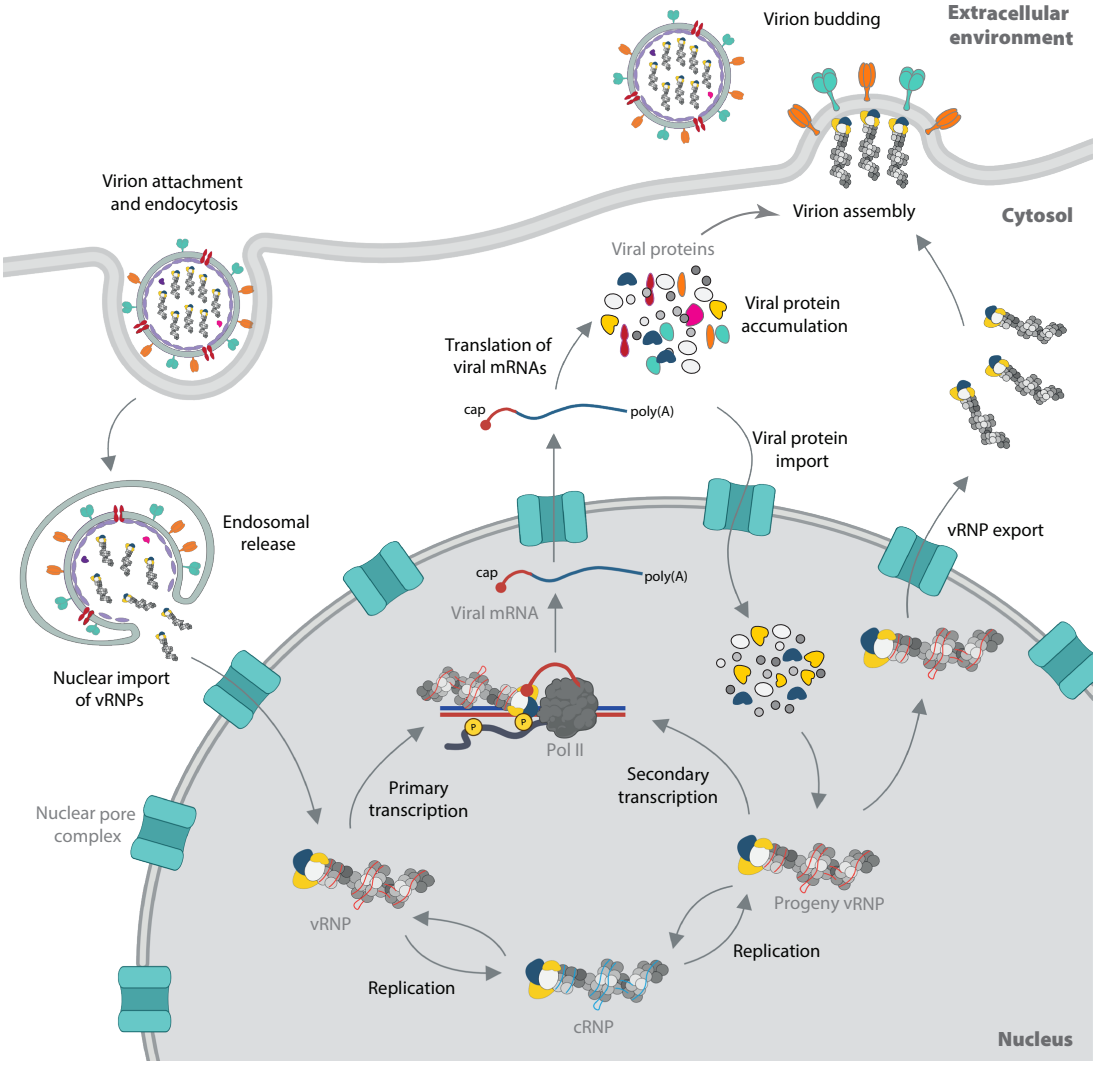

Figure 1.2: Overview of the influenza virus life cycle. Influenza virions bind to the plasma membrane and are endocytosed into the cell. Within a low pH environment, vRNPs are released into the cytosol and trafficked into the nucleus through the nuclear pore complex. Once in the nucleus, various processes occur including primary transcription of viral mRNA using incoming vRNPs as a template. This process involves FluPol associating with serine 5 phosphorylated RNA polymerase II for cap-snatching and production of viral 5' capped and 3' polyadenylated mRNA that is then exported to the cytosol for viral protein synthesis. The synthesised viral proteins are subsequently imported into the nucleus, contributing to a shift from viral transcription to replication. During replication, negative-sense vRNA is replicated into positive-sense cRNA which serves as a template for the synthesis of progeny vRNA. Together, FluPol and NP are imported into the nucleus where progeny vRNPs are assembled before undergoing secondary transcription and replication. In later stages of infection, progeny vRNPs leave the nucleus and travel to be assembled along with viral proteins at the plasma membrane. The fully formed virion buds off from there and enters the extracellular environment upon release. Parts of the figure were created with BioRender.

This tropism presents a barrier for cross-species transmission, which must be overcome by mutation of the HA to allow for infection of a different host [31, 59]. The virus enters the cell through receptor-mediated endocytosis, forming an early endosome which undergoes gradual acidification by endosomal membrane proton ATPase [60]. The low pH environment induces a conformational change in HA, enabling the fusion of the viral envelope with the endosomal membrane [61]. Acidification is further promoted by the M2 ion channel, which mediates the influx of protons into the virus particle, ultimately causing the release of vRNP into the host cell cytoplasm by weakening the M1-vRNP interaction as well as RNA-RNA interaction of the vRNP bundle [62, 63]. In addition, M1 binding to cytoplasmic transportin-1 further enhances the release of vRNPs into the host cell cytoplasm [64]. The released vRNPs are trafficked into the nucleus, a process mediated by the importin α protein family which recognises the nuclear localisation signal (NLS) of NP [39, 43].

In the nucleus, transcription and replication of the viral genome take place sequentially. First, during primary transcription, the vRNPs are transcribed by the FluPol into positive-sense viral mRNAs. The viral mRNA produced contains a 5' terminal N7-methyl guanosine (m^7G) cap and a 3' polyadenylated (polyA) tail, making it structurally similar to host mRNA. As a result, it remains unrecognized by the host's innate immunity, allowing the virus to hijack the host translation machinery for viral protein synthesis in the cytoplasm [41, 65, 66]. The host mRNA is exploited by FluPol as a source of nascent RNA caps to prime its own viral transcription in a process called 'cap-snatching' [41]. During cap-snatching, FluPol directly interacts with the host RNA polymerase II (Pol II) in the early stages of Pol II-mediated host transcription to secure its access to host RNA caps [67].

The produced viral proteins are imported into the nucleus where they shift FluPol from its transcription to replication function. Viral replication is a two-

step process in which positive-sense complementary RNA (cRNA) is used as an intermediate template to synthesise new vRNA [65]. The cRNA forms together with newly synthesised FluPol and NP, a complementary RNP complex (cRNP). In the second step, the cRNP serves as a template for the synthesis of new vRNA that is packaged into progeny vRNPs which in turn can initiate a second round of transcription or replication [44]. Progeny vRNPs are also exported to the cytoplasm via the exportin 1 (Crm1) pathway [43]. To enable vRNP nuclear export, the viral proteins NEP and M1 interact with the cellular nuclear export machinery, however, the molecular details of this process remain under investigation [68, 69].

Once the vRNPs are transported into the cytoplasm, they are trafficked to the apical side of the cell membrane to form vRNP bundles [70]. Lipid rafts in the apical cell membrane facilitate virus budding by accumulating HA and NA which are newly synthesised at the endoplasmic reticulum and further processed in the Golgi apparatus [71]. HA and NA in the lipid rafts and M1 proteins associated with progeny vRNPs recruit M2, driving virion budding. At the budding virion site, vRNPs assemble into bundles of 8 perpendicular to the cell membrane. Selective packaging of the vRNP bundles is facilitated by RNA-RNA interactions between the different viral gene segments [72]. Upon budding, NA cleaves sialic acid residues of the plasma membrane glycoproteins to prevent reattachment with the infected cell or aggregation with other virions [36, 43, 65].

1.2 Structure and function of the influenza virus polymerase

1.2.1 Structural features of the influenza virus polymerase

The influenza virus RNA-dependent RNA polymerase has three subunits - PA, PB1, and PB2- that form a heterotrimeric complex with a molecular weight of approximately 250 kDa. It also has approximate dimensions of 115 × 100 × 75 Å for its height, width and depth, respectively [73]. The structure consists of a globular catalytic core with two protruding domains: the PA endonuclease domain and the PB2 cap-binding domain (Figure 1.3). This multifunctional protein is part of the vRNP complex and undergoes extensive conformational rearrangements to facilitate both replication and transcription of the viral genome. These diverse structural arrangements are stabilised by hydrophobic subunit interfaces while flexible linkers and reversible hydrophilic interactions enable adjustments across various FluPol subunits. Studies using cryo-EM and crystallography have revealed intricate details about this architectural complexity as well as the underlying molecular mechanisms involved in these processes [44, 73, 74].

PB1 forms the catalytic core and features a canonical right-handed RNA-dependent RNA polymerase (RdRP) architecture, including thumb, fingers, and palm subdomains (Figure 1.3 A-C). It also contains motifs A-F typically found in RdRps that form the polymerase's active site necessary for the nucleotide addition reaction [73, 75]. Before nucleotides can be added, motifs D and F guide free nucleotides into the active site through a channel containing basic amino acids to stabilise their position for subsequent catalysis [76–79]. Motif C includes two amino acid residues D445 and D446 (corresponding to ICV aspartate residues D446 and D447) which coordinate two magnesium ions along with D305 (in motif A) for polymerase activity [80]. The magnesium ions act as catalysts for adding nucleotides to the nascent RNA chain by facilitating a nucleophilic attack

of the 3' hydroxyl group on the incoming nucleotide's 5' triphosphate group, resulting in the incorporation of a nucleoside monophosphate while releasing pyrophosphate [81, 82]. PB1 also contains a priming loop, a flexible loop at the active site of PB1 located at the PB1 thumb and crucial for terminal initiation in replication (Figure 1.3 C) [83].

PA consists of two domains connected by a long flexible linker (Figure 1.3 C). This linker connects the endonuclease domain (PA endo) at the N-terminus with the large C-terminal domain of PA (PA-C, P3-C in ICV). In the cell, the PA and PB1 subunits are trafficked into the nucleus together as a heterodimeric complex that associates with Ran-binding protein 5 through the NLS domain of PB1 [84–86]. The N-terminal PA-endo domain mediates cap-snatching with the PB2 cap-binding domain and has a globular fold [73]. As a member of the PD-(D/E)xK nuclease superfamily, it contains in the active site, alongside the catalytic K134 required for endonuclease activity, the four conserved residues H41, E80, D108 and E119 which coordinate two divalent cations, either magnesium or manganese [87–92]. The importance of these divalent cations is underlined by the development of chelators that serve as anti-influenza drugs with baloxavir as a prominent example [73, 93]. The PA-C is a compact globular domain with a molecular weight of 55 kDa and interacts with PB1 as a part of the rigid FluPol core (Figure 1.9 B, E). The domain is also involved in promoter RNA binding and includes a highly basic channel, essential for 5´ promoter RNA binding (Figure 1.3 B, D) [77, 94, 95]. Importantly, it is considered an anchor for both transcription and replication as PA-C binds to the host Pol II CTD and potentially other factors involved in replication [74, 96]. The CTD bipartite interaction site is located either entirely (FluPol A) or partially (FluPol B and FluPol C) on the PA-C domain and will be discussed in more detail later (Figure 1.3 E) [41, 96, 97].

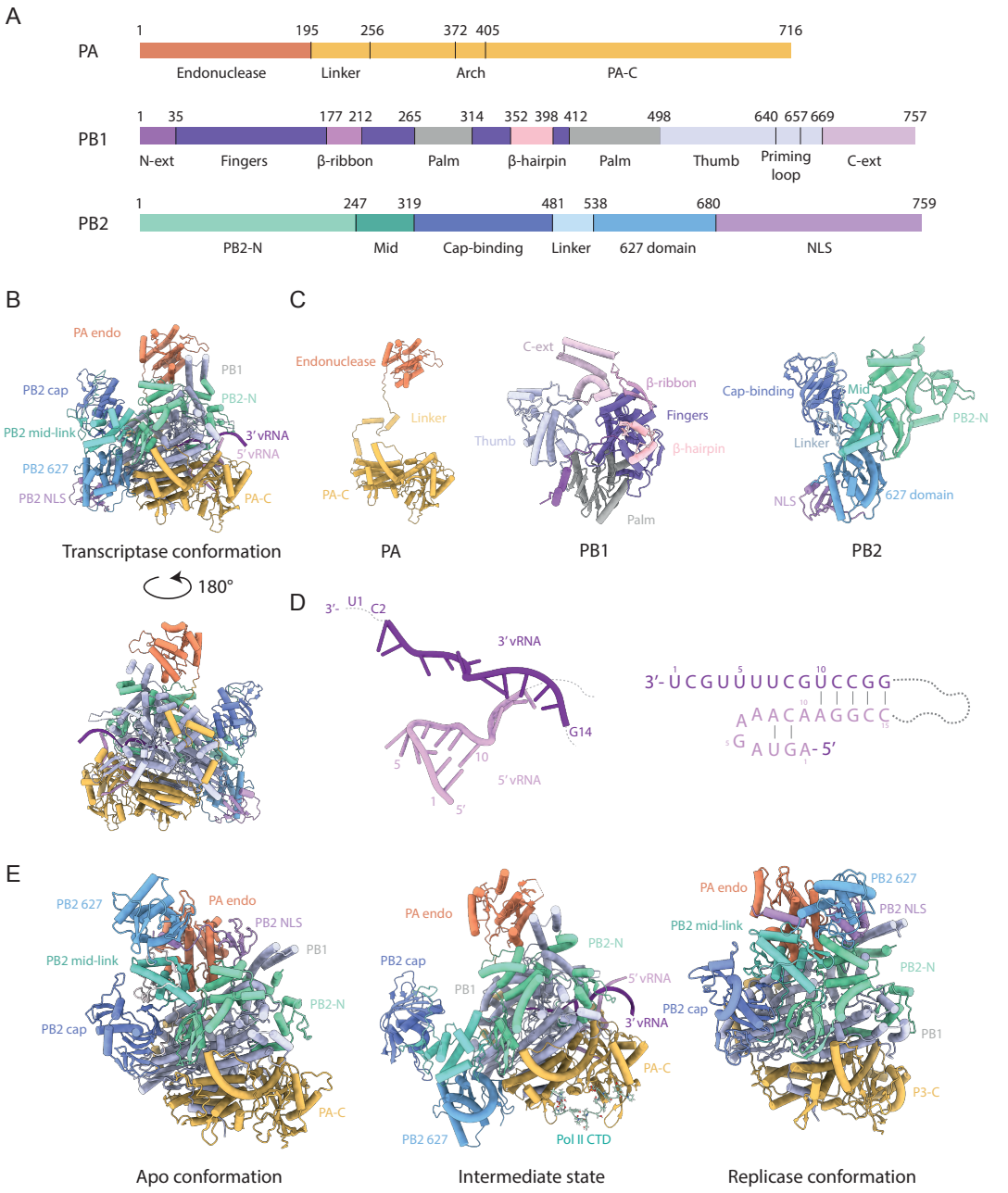

Figure 1.3: Structural features of the viral polymerase. (A) Schematic of the influenza virus polymerase A subunit domains. **(B)** Cartoon diagram of heterotrimeric FluPol A in a transcriptase conformation (PDB: 6RR7) bound to viral RNA promoters. The polymerase domains are coloured and labelled. **(C)** Structural diagrams of the FluPol subunits PA, PB1, and PB2 and their domains (PDB: 6RR7). **(D)** The viral RNA promoter structure in the solved transcriptase conformation (PDB: 6RR7) shows the 5' hook and base pairing with the 3' vRNA terminus. Nucleotides U1, C2, and G14 in the 3' vRNA and C15 in the 5' vRNA are not resolved in the structure (left). A schematic of the full-length 5' and 3' viral promoters of the viral genome segment (right). **(E)** Cartoon diagram of FluPol showing its apo conformation (PDB: 6QNW), intermediate conformation (PDB: 8R3L), and replicase conformation (PDB: 6XZR), and the arrangement of its flexible domains.

PB2 consists of a C-terminal domain (PB2-C), composed of various subdomains, and an N-terminal domain (PB2-N) that forms an interface with PB1 at the core of the polymerase. The flexibly connected PB2-C subdomains contribute to the various functions it has to fulfil [73]. The cap-binding domain (CBD) has a globular fold, crucial for cap-snatching [98, 99]. Like other cap-binding proteins, the CBD traps the host mRNA cap via binding of the m^7G cap structure which is embedded in an aromatic sandwich formed by residues H357 and F404 in influenza A FluPol. Interestingly, FluPol A CBD exhibits a greater preference for m^7GTP, while FluPol B demonstrates an equal affinity for both methylated and unmethylated caps [100]. Structural studies of CBD also found that the flexible 424 loop within the CBD (residues 413-428) might play a role in transcription by leading the cleaved mRNA cap into the polymerase active site to serve as a primer for viral transcription [98, 99, 101–103]. The cap-binding domain in IAV has a relatively low affinity for the cap analogue m^7Gppp [98]. In the heterotrimer context, however, the adjacent PB2 midlink domain enables additional interactions with the capped RNA substrate [73, 77, 104]. Nevertheless, further investigation is needed to understand how the influenza polymerase can selectively hijack capped nascent Pol II transcripts while competing against the host cap-binding complex, which typically binds and sequesters the cap with higher affinity early in transcription [104, 105].

The 627-NLS double domain constitutes a part of PB2-C with the two subdomains 627 and NLS connected by a linker [73]. The double domain has no catalytic function but is pivotal in host adaptation and nuclear import [73]. The name of the 627 subdomain results from the amino acid mutation E627K that allows avian influenza viruses to cross the species barrier to replicate well in mammalian cells [106]. Influenza viruses with glutamate at the 627 position can only replicate in mammalian cells if other substitutions such as G590S and Q591R are present [73, 107–109]. Examination of the surface electrostatic potential indicates a strongly basic area surrounding this particular residue, which is disturbed in the presence

of E627 [110, 111]. While it was shown *in vitro* that the 627 domain is not required for transcription, the *in vivo* context of 627 plays a role in mediating interaction with host factors. The basic patch on the domain surface plays a role in accommodating the binding of the negatively charged phosphates in serine 5 phosphorylated RNA polymerase II CTD and stabilises a transcriptase conformation in FluPol C [97]. The differences between FluPols from different influenza types remain under investigation [40, 41, 112].

The NLS, which has a bipartite nuclear localisation signal, is linked to the 627 domain by a linker on the subdomain 627. The primary function of NLS is to traffick the PB2 subunit via the interaction with the importin α family into the nucleus [73, 113, 114]. Crystal structures of isolated 627-NLS, as well as full-length FluPol, have revealed distinct conformations where in FluPol not bound to vRNA promoters, the so-called apo-conformation, the 627 and NLS domains are spatially separated, resulting in an open conformation that is required for nuclear import [115–117]. More recently, cryo-EM studies showed that PB2 627-NLS interacts with PA-endo in a unique, replicase-like arrangement, required for viral replication [44]. In contrast, a transcription-compatible conformation of the heterotrimer shows that 627 and NLS pack against each other in a closed state [73].

In the context of RNP, FluPol binds to the highly conserved 5' and 3' viral RNA promoters of the cRNA or vRNA. This binding influences the activity and direction of conformational rearrangement for both transcription and replication functions. Viral promoters can trigger large structural rearrangements through their ability to form secondary structures and bind to FluPol [73]. The apo form, with no RNA promoter bound to FluPol, is similar to the 5' cRNA-bound conformation. In this conformation, PB2-C is rotated by 140 degrees, unlike the vRNA-bound FluPol structure, which adopts a transcriptionally active conformation [41]. The 5' and 3' ends of the promoters are partially complementary and have base pairing with either 4 (cRNA) or 5 (vRNA) Watson-Crick base pairs [118–120]. The 5'

promoters of both vRNA and cRNA can form a stem-loop, a hook-like secondary structure, that is formed by two Watson-Crick base pairs and two mismatch base pairs within the first 10 nucleotides of the 5' end. The hook binds in a PB1 pocket to activate the polymerase by triggering conformational rearrangements [77, 116]. In contrast, the 3' promoters are single-stranded and flexible with structural studies showing that the 3' vRNA promoter binds to FluPol in three distinct modes required for different functional states of the polymerase [121–123]. During viral transcription, the vRNA-bound FluPol adopts different conformations with a flexible PB2 cap-binding domain that rotates by 70 degrees [41]. It was also suggested that during viral transcription, the promoters remain bound to the FluPol throughout the complete cycle of viral mRNA synthesis [73].

1.2.2 Influenza virus transcription

FluPol depends on the host transcriptional machinery to produce its positive-sense viral mRNA [41]. During a process called 'cap-snatching', FluPol steals mRNA caps from nascent host mRNA synthesised by the host RNA polymerase II (Pol II) to initiate viral transcription. This occurs during the initial stages of Pol II-mediated host transcription, enabling FluPol to secure its access to host RNA caps through the interaction of FluPol with the CTD of Pol II [124, 125]. Functional and structural studies support that FluPol selectively recognises the Ser5 phosphorylated (pS5) CTD heptad repeat sequence YSPTpSPS, the initiation signature for Pol II [41, 126, 127]. While FluPol specifically recognises the pS5 form of the CTD, deep sequencing studies of viral mRNA proposed that the most abundant capped RNAs drive the preference of FluPol for host transcripts. These include not only cellular mRNA but also various non-coding RNAs such as small nuclear and nucleolar RNAs (snRNAs, snoRNAs) transcribed by Pol II [128–130]. During cap-snatching, the endonuclease domain recognises the m^7G host mRNA cap and cleaves 10-13 nt downstream of the 5' cap to generate a primer for transcription initiation [41]. The endonuclease domain of FluPol tends to favour cutting at the 3´ terminus of purines, particularly

guanine [91]. For efficient transcription initiation, FluPol also prefers capped primers that are complementary to the 3' vRNA promoter strand at their 3' end through canonical base pairing [129, 131–133].

Molecular mechanism of viral transcription

FluPol undergoes structural rearrangements to perform viral transcription and requires simultaneous binding to vRNA and the C-terminal domain (CTD) of Pol II to be in a transcriptionally active conformation that allows for cap-snatching [41].

The PB2 cap-binding domain of FluPol binds the capped nascent Pol II transcript followed by cleavage of the transcript 10-13 nucleotides downstream of the m^7G cap by the PA endonuclease domain. The distance between the PA endonuclease and the PB2 cap-binding domain, as well as their conformational flexibility during cap-snatching, influences the length of the capped RNA primer leading to slight variations in its total length (Figure 1.4, Figure 1.5 A) [102, 134, 135]. Following cleavage of the nascent transcript, the PB2 cap-binding domain rotates by 70° directing the capped RNA fragment to the FluPol active site for template pairing with the first nucleotides of the viral RNA template (Figure 1.4, Figure 1.5 B) [102, 134, 135]. In the case of weak template pairing during transcription initiation, the capped RNA fragment can undergo multiple realignments [129, 131].

The cap-binding domain releases the 5' end of the nascent viral mRNA. As the capped fragment extends on the 3' vRNA promoter, the newly synthesised viral mRNA exits through the product exit channel, while the vRNA template is directed through a template exit channel (Figure 1.5 C) [102, 104, 134, 135]. As additional nucleotides are incorporated into the nascent viral mRNA, the stability of the template-product duplex increases. This disrupts the interaction between the vRNA promoter duplex, triggering a conformational change that allows further product elongation [134].

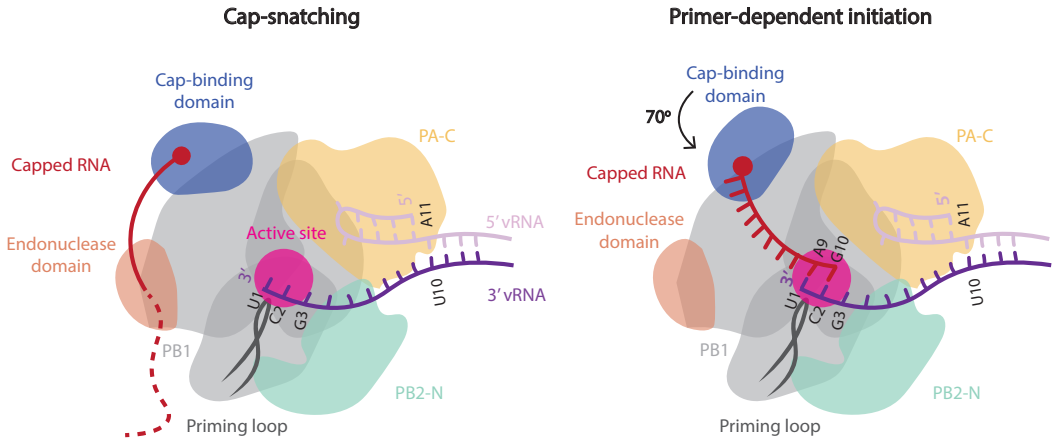

Figure 1.4: Cartoon diagrams depicting viral transcription initiation. The cap-binding domain of the influenza virus polymerase binds to the m7G cap of the host nascent RNA. This is followed by the endonuclease domain-mediated cleavage of an RNA primer and rotating of the cap-binding domain by 70°. The RNA primer forms base pairs with the 3' vRNA terminus to initiate viral transcription. Elements of this figure were adapted from Zhu et al. (2023) [44].

Once FluPol approaches the end of the vRNA template, it polyadenylates the viral mRNA at the 3' end by a process called stuttering where FluPol repeatedly backtracks over an oligo-U sequence of 5-6 U nucleotides located close to the 5' end of the vRNA template (Figure 1.5 D) [102]. The stuttering is caused by the inability of FluPol to translocate the product mRNA further caused by the tightly bound hook structure of the 5' vRNA promoter which remains bound in the promoter binding site throughout viral transcription [104, 136]. After polyadenylation, the produced viral mRNA is masked as host RNA due to its 5' host cap and 3' poly(A) tail which enables it to be further processed. After product release, the vRNA termini remain bound to FluPol, allowing a new round of viral transcription (Figure 1.5 E). Viral mRNAs encoding M1 and NS1 are subject to splicing in the nucleus before they can be exported via the NXF1/p15 pathway like viral mRNA without introns [53, 137, 138]. NS1 also contributes to host shut-off by inhibiting the cellular polyadenylation machinery and the cellular antiviral defence mechanisms [139]. For instance, NS1 sequesters cellular CPSF30,

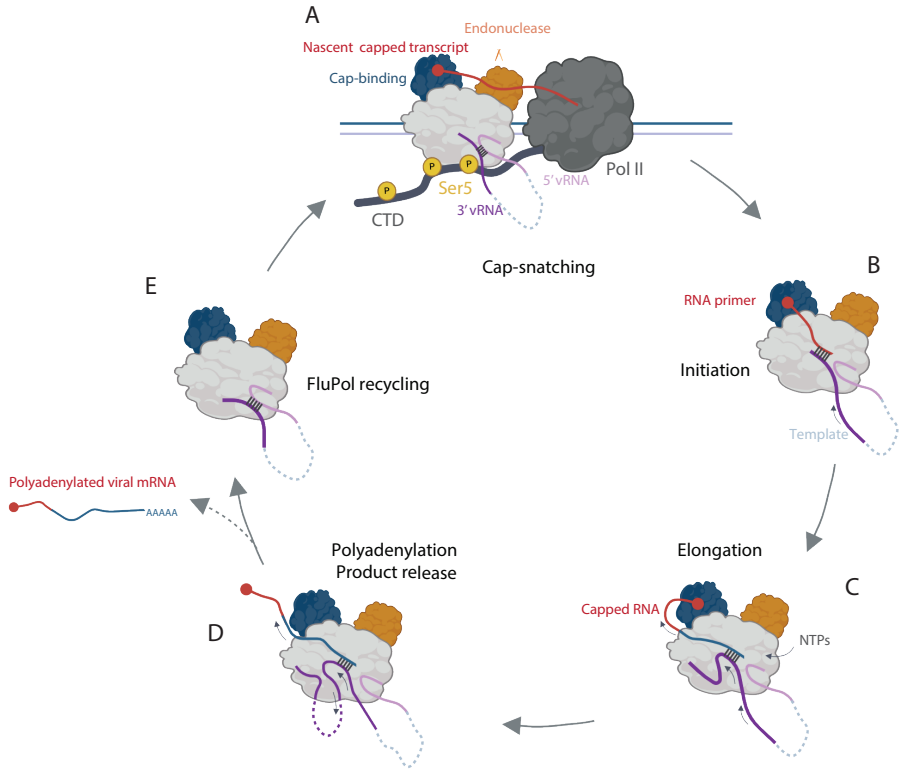

Figure 1.5: Overview of the influenza virus polymerase transcription cycle. **(A)** FluPol performs cap-snatching by binding to the serine 5 phosphorylated (yellow circle) Pol II CTD (grey), positioning it close to the capped nascent host transcript (red) (PDB: 4WSB). The PB2 cap-binding domain (dark blue) binds the host RNA cap, allowing the PA endonuclease domain (orange) to cleave off a capped RNA fragment to prime viral transcription. **(B)** Viral transcription is initiated by rotation of the PB2 cap-binding domain which positions the capped RNA primer in the FluPol active site to form base pairs with the 3' vRNA terminus (PDB: 4WSA). **(C)** FluPol elongates the viral transcript along the vRNA template by incorporating incoming nucleotides (NTPs), resulting in the release of the mRNA cap from the PB2 cap-binding domain. **(D)** Once the 3' vRNA terminus exits the polymerase active site, it interacts with a secondary site (Mode B site) on the polymerase surface that later facilitates recycling of FluPol for another round of transcription after termination (PDB: 6SZU). Since FluPol is unable to release the tightly bound 5' vRNA promoter for further elongation, the viral mRNA is polyadenylated by stuttering on a poly(U) sequence near the 5' vRNA terminus (PDB: 6T0S). The transcription cycle produces a hybrid viral mRNA consisting of a host-derived RNA cap, the viral segment coding sequence (dark blue) and a poly(A) tail. **(E)** After releasing the viral mRNA, the 5' and 3' vRNA termini remain bound to FluPol allowing a new round of transcription (PDB: 6T0U). Parts of this figure were created with BioRender.

a 30-kDa pre-mRNA processing factor, inhibiting cellular 3'-end processing and ultimately the generation of functional host mRNA [140].

Molecular details of the FluPol-CTD interaction

Cap-snatching occurs when FluPol forms a close association with Pol II, especially with the CTD. This process is thought to occur while Pol II transcriptional activity pauses, allowing host RNA transcripts to be capped shortly after the initiation of Pol-II mediated transcription [73]. Functional studies using synthetic CTD mimic peptides have demonstrated that FluPol specifically recognises a pS5 modification rather than unphosphorylated or serine 2 phosphorylated (pS2) Pol II CTD [26, 74, 96, 97, 126]. Although these functional studies suggested a direct interaction between FluPol and the pS5 form of the CTD, the structural mechanism for this binding was only uncovered in recent structural studies on influenza A and B as well as influenza C polymerases [26, 96, 97].

Structural studies using cryo-EM and crystallography approaches revealed in most cases two binding sites of the CTD to FluPol by employing Pol II synthetic peptides containing four repeats of the Pol II heptapeptide with a pS5 modification (pS5 CTD_{4x}) (Figure 1.6) [40, 41, 74, 96, 97]. The bipartite CTD binding site exhibits both shared and unique characteristics across FluPol from different influenza types [40]. For FluPol A, the two binding sites (site 1A and site 2A) are exclusively on the PA-C subdomain (Figure 1.6 A) [40, 73, 96]. Site 1A is formed by multiple alpha helices in the PA-C while site 2A is embedded in a beta-hairpin loop further away from the PB1 core [40, 41]. Basic residues in the binding sites stabilise the negative charge of the phosphate group on the pS5 CTD. As a result, mutation of conserved residues K635 and R638 in site 1A or K289 and R449 in site 2A leads to loss of viral transcription *in vitro* and severely affects viral replication in viruses [40, 73, 96]. Recently, new FluPol A structures bound to synthetic pS5 Pol II CTD mimic peptides were shown to have one long, continuous binding site instead of the commonly observed bipartite binding pattern observed

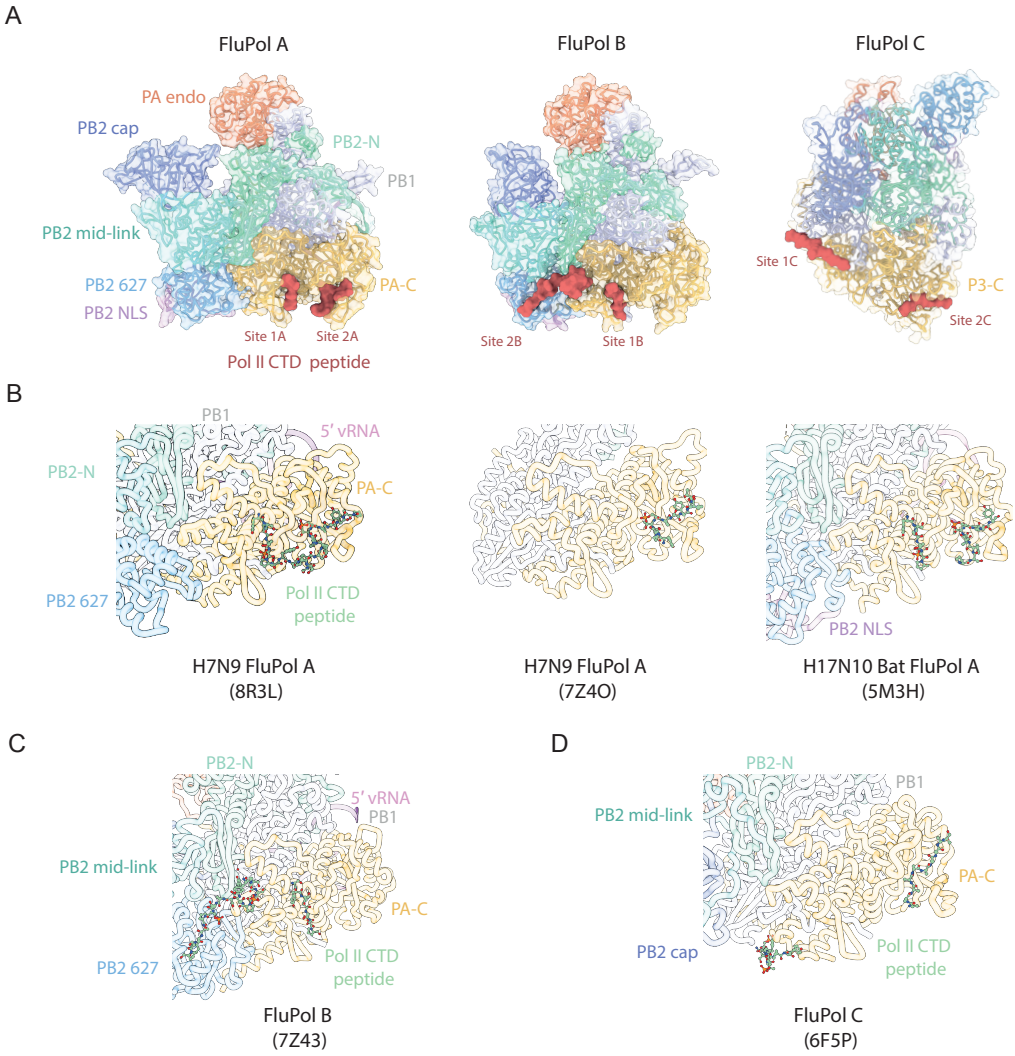

Figure 1.6: Structural details of synthetic Pol II CTD peptides bound to FluPol. **(A)** Structures of FluPol A (bat, PDB: 4WSB), FluPol (B) (human, PDB: 7Z43), and FluPol C (human, PDB: 6F5P) bound to pS5 synthetic Pol II CTD mimic peptides (red) shown as cartoon surface diagrams with coloured domains. **(B)** Comparison of the CTD binding sites of FluPol A of different strains with all models shown in the same orientation (PDB: 8R60, 8R3L, 7Z4O,5M3H). **(C)** CTD binding sites of FluPol B (PDB: 7Z43). **(D)** CTD binding sites of FluPol C (PDB: 6F5P). Panels in B-D are adapted from Keown, Baazaoui et al. (2024).

in previous studies indicating the potential of further Pol II binding sites [74, 112]. In FluPol B, protein sequence alignments and a crystal structure indicate that site 1 (site 1B) is conserved compared to FluPol A (Figure 1.6 A, C). However, the second site (site 2B) does not overlap with site 2A; instead, it extends from the PA-C region to the PB2 627-NLS double domain [26, 40, 96].

The CTD binding sites on FluPol C differ from those found on FluPol A and FluPol B [97]. The first site (site 1C) is located at the boundary between the PB1 and P3 subunits with P3 corresponding to the PA domain in FluPol A (Figure 1.6 A, D). Site 2 in FluPol C (site 2C) lies at a similar position to site 2A but does not have sequence conservation indicating a different mode of binding (Figure 1.6 A) [40, 41, 97]. The presence of multiple CTD binding sites across different influenza virus types is thought to contribute to a high-affinity interaction by avidity and cooperativity of the binding of multiple heptad repeats of the Pol II CTD to FluPol, particularly in the context of mammalian Pol II CTD with 52 heptapeptide repeats [41, 73].

Despite the common feature of CTD binding sites present on the FluPol surface, lack of sequence conservation and different locations on the FluPol domains indicate that the binding of FluPol from different influenza virus types to Pol II CTD types has diverged during evolution and the presence of other potential CTD binding sites remain elusive [26]. Considering that mammalian Pol II CTD contains 52 heptad repeats of the consensus sequence YSPTSPS, which far exceeds the length of CTD mimetic peptides, it appears likely that the FluPol-Pol II CTD interaction takes place beyond the identified binding sites [40, 41].

1.2.3 Influenza virus replication

Once sufficient levels of viral mRNA and viral protein have been produced at 4-6 h post-infection, FluPol increasingly starts replicating the viral genome [141]. The dual function of FluPol involves different mechanisms to be able to perform both

transcription and replication. Replication is a two-step process where initially positive-sense cRNA is produced from the negative-sense vRNA and in the second step cRNA is used as a template to produce more vRNA which then, together with FluPol and NP, can form progeny vRNPs [44, 83]. In contrast to viral mRNA, vRNA and cRNA do not have a 5' cap and a 3' poly(A) tail which can trigger the host's immune response resulting in RNA degradation [142]. Therefore, to minimise direct exposure to the host cell environment, both cRNA and vRNA are thought to be assembled into RNPs during the replication process [44, 141].

Unlike viral transcription initiation, replication initiation is a primer-independent process which involves two distinct mechanisms for cRNA and vRNA replication (Figure 1.7) [44, 83]. For replication initiation, a pppApG dinucleotide is produced by FluPol from ATP and GTP. In the case of cRNA synthesis, pppApG dinucleotide synthesis starts at the first nucleotide of the 3' vRNA promoter, termed terminal initiation [44]. The replication initiation for vRNA synthesis, known as internal initiation, involves producing a dinucleotide that pairs with residues 4 and 5 on the 3' end of the cRNA instead of residues 1 and 2 on the 3' end of the vRNA template (Figure 1.7 A) [44]. The disparity in initiation mechanisms arises from differences in the promoters of cRNA and vRNA [44, 123]. The first ten nucleotides at the 5' end of the cRNA and vRNA fold into a hook-like structure which binds tightly to an interface formed by the PB1 and PA-C [44, 77]. The next 5 to 6 nucleotides after the hook form base pairs with the 3' end cRNA and vRNA template, respectively [44]. Terminal initiation on the vRNA template places U1 and C2 of the 3' end opposite the dinucleotide in the active site, followed by 7 nucleotides in the template entry channel and residue U10 starting base pairing with the 5' end of the vRNA (Figure 1.7 A) [44].

Internal initiation on the cRNA template for vRNA synthesis is more complex and requires the formation of base pairs starting from C12 at the 3' end of the cRNA template with nucleotides adjacent to the hook structure of the 5'

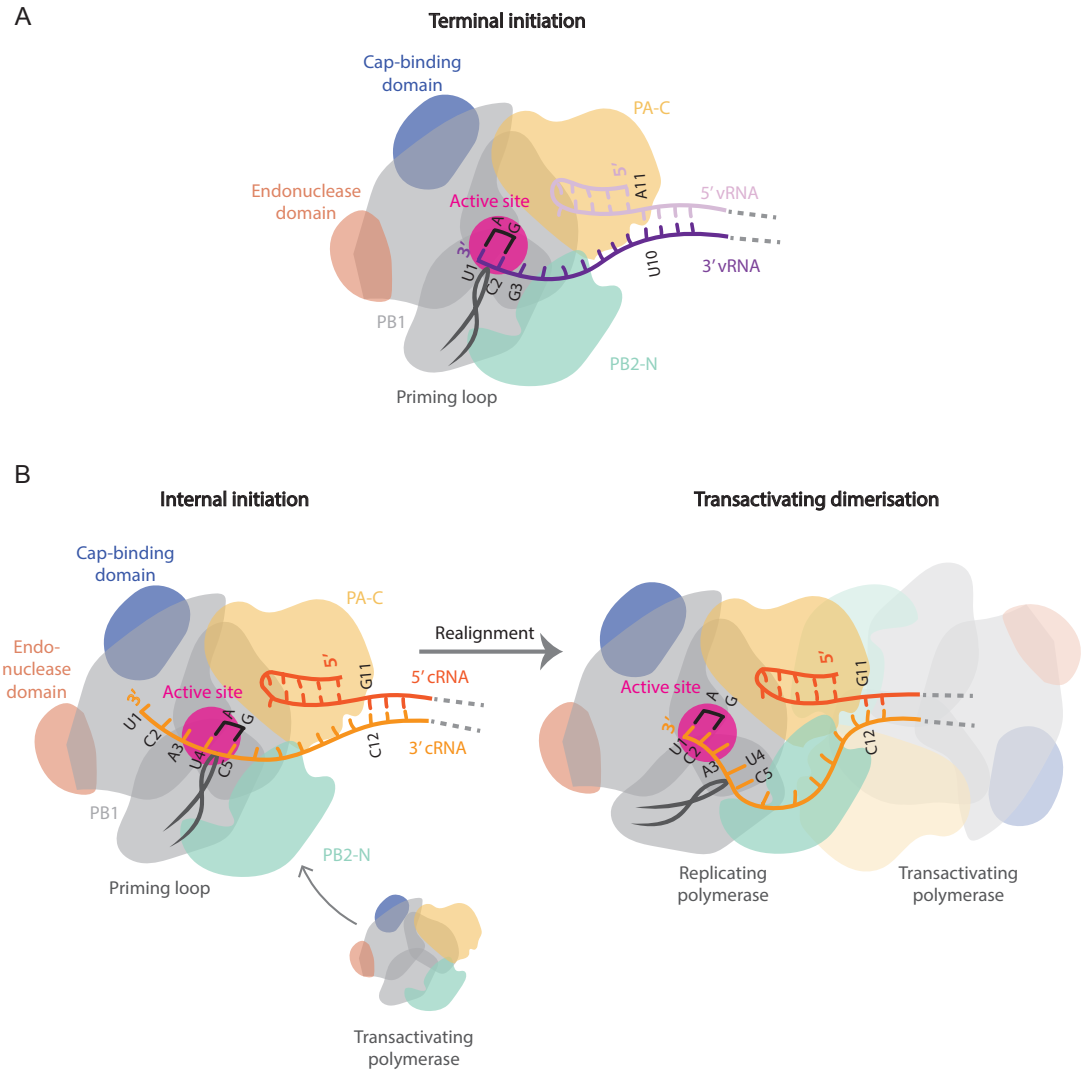

Figure 1.7: Replication initiation of vRNA and cRNA templates. (A) ApG dinucleotide base pairs with U1 and C2 of the 3'vRNA terminus for terminal initiation, stabilised by the priming loop. **(B)** The replication of the cRNA template is a two-step process. First, an ApG dinucleotide is synthesised on the inserted template and forms base pairs with U4 and C5 of the 3' cRNA terminus. A second FluPol molecule, the transactivating polymerase, binds to the replicating polymerase, causing conformational rearrangements. This interaction results in the polymerase core opening which causes a shift in the priming loop that realigns the 3' cRNA terminus. As a result, bases U1 and C2 are positioned opposite to the previously formed ApG dinucleotide. Elements of this figure were adapted from from Zhu et al. (2023) [44].

end (Figure 1.7 B). This results in 11 nucleotides that are threaded through the template entry channel, followed by pppApG dinucleotide synthesis on the template, placing residues U4 and C5 in the active site opposite the pppApG dinucleotide (Figure 1.7 B). Realignment before elongation is crucial for preserving genome integrity, as initiation takes place downstream on the 3' cRNA template [44].

The priming loop, a flexible β-ribbon of the PB1 thumb domain, extends into the FluPol active site where it plays an important role in both replication initiation processes. In terminal initiation, it contributes to the correct positioning of the template and initial nucleotides by stabilising them in the active site. However, a structural model capturing the initiation process remains to be investigated [44, 83, 104]. The priming loop along with another FluPol molecule, termed the transactivating polymerase, plays a pivotal role in internal initiation [44, 143, 144]. Current research suggests that through the binding of the transactivating polymerase to the replicating polymerase, the polymerase core is opened up which triggers the priming loop to move the cRNA template, resulting in the realignment of the dinucleotide to be opposite the first two nucleotides U1 and C2 of the cRNA template (Figure 1.7 B) [44].

Recent structural studies have provided insights into the molecular details of viral replication and the importance of FluPol dimer formation as well as the reliance on host factors [44, 145]. Specifically, proteins of the host ANP32 family (ANP32A and ANP32B in mammalian cells and ANP32A in avian cells) play a pivotal role in replication and host-range specificity [106]. Multiple isoforms of ANP32 family proteins are present in various species, such as chA and chB in chicken, and huA, huB, and huE in humans. However, viral polymerases differ in their capacity to exploit ANP32 family proteins. For instance, avian FluPol is unable to use mammalian ANP32 proteins due to molecular incompatibility,

which underlies host-range specificity [146, 147].

A replication platform is formed by a complex of a FluPol asymmetric dimer consisting of a replicating and an encapsidating polymerase together with ANP32 [44]. The FluPols in the asymmetric dimer are in two different conformations. The replicating polymerase is bound to the RNA template, while the encapsidating polymerase remains RNA-free until it receives the newly synthesised RNA [145]. This prevents degradation and activation of host innate immunity mechanisms through co-replicative assembly [44].

Influenza virus transcription and replication peak at different times in the viral life cycle, but the exact details of how these processes are temporally regulated and coordinated remain to be fully understood. Transcription activity peaks at 4-6 h post-infection while cRNA and vRNA are detected in the later stages of infection when sufficient amounts of viral protein have been synthesised for the co-replicative assembly of vRNPs and cRNPs [67, 148, 149]. Current studies suggest that the switch from transcription to replication is not purely a stochastic process but is also dependent on molecular cues [44]. For instance, it was shown that binding to the Pol II enhances cap-snatching, and the presence of free FluPol in the nucleus promotes vRNA synthesis through the assembly of cRNPs [26, 44, 97]. Newly synthesised, free FluPol molecules that increase in later stages of infection could act as molecular signals to slow down transcription by binding to the host Pol II and RNP-associated FluPol, which limits their availability for each other [44]. This inhibition would prevent the RNP-bound FluPol from performing cap-snatching due to a blocked interface, as was demonstrated in a nanobody-based study [150].

1.3 Interaction of the influenza virus polymerase with the host transcriptional apparatus

1.3.1 The role of post-translational modifications on the Pol II CTD

Eukaryotes encode three RNA polymerases (Pol I-III) of which RNA polymerase II (Pol II) catalyses the transcription of all protein-coding and many non-coding regions of the eukaryotic genome [151]. RNA polymerase II is responsible for transcribing mRNA as well as a variety of non-coding RNAs, including long non-coding RNAs, microRNAs, small nucleolar RNAs, and small nuclear RNAs [40, 151]. Pol II consists of 12 subunits (RBP1-12), with the largest one being RPB1, containing a long disordered C-terminal domain (CTD) that is subject to various post-translational modifications (PTMs) [152, 153].

The RPB1 subunit comprises a globular core connected to the CTD through a linker, followed by the middle region and the distal region of the CTD (Figure 1.8 A). The length of the CTD varies between eukaryotes, with the middle region of the CTD containing 26 repeats of a conserved YSPTSPS heptapeptide in yeast and 52 repeats in mammalian cells [154]. In mammalian cells, the 52 heptad repeats consist of 21 consensus repeats and 31 variants with repeats in the distal CTD where serine 7 is replaced by asparagine, lysine, or threonine, indicating the complexity required to signal various cellular processes (Figure 1.8 B) [155, 156].

The CTD is largely disordered and undergoes a variety of post-translational modifications, primarily phosphorylation, but also glycosylation, methylation, ubiquitination, and acetylation. Together, these form the basis for the "CTD code," a hallmark of various stages during the transcription cycle (Figure 1.8 B) [152, 153]. The highly regulated modification of the CTD orchestrates the spatiotemporal regulation of Pol II-mediated transcription. Furthermore, it allows the CTD to serve as a platform for the recruitment of various host factors involved in transcription and RNA processing and to adopt various conformations depending

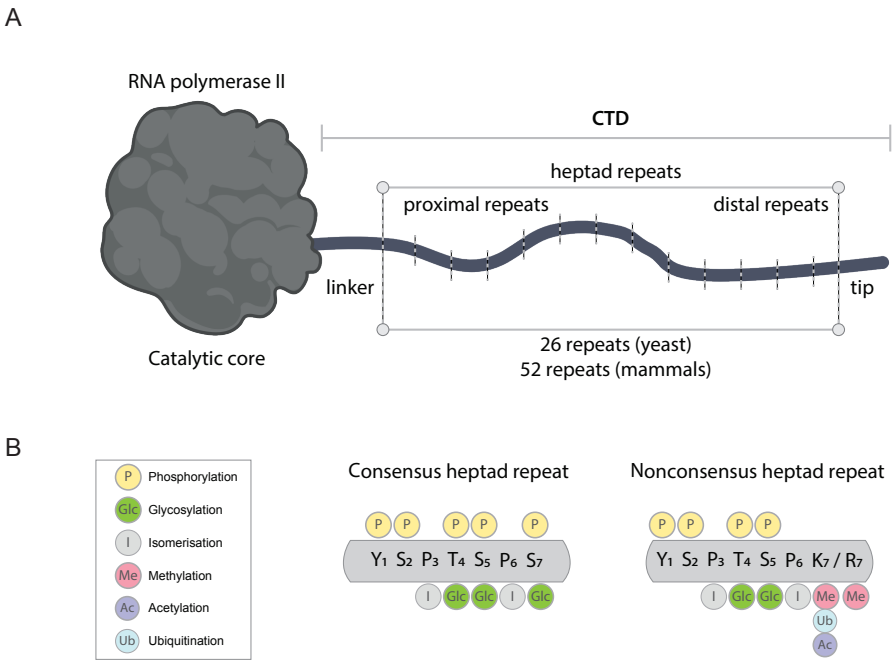

Figure 1.8: Post-translational modifications of the Pol II CTD. (A) RNA polymerase II (Pol II) consists of a globular catalytic core and a long disordered C-terminal domain (CTD). The CTD is conserved and consists of a linker connecting the Pol II core, heptapeptide repeats, and a tip at the distal end. Mammalian CTD contains 52 repeats, while yeast CTD is composed of 26 repeats. **(B)** The heptapeptide is subject to various post-translational modifications outlined above for the consensus and nonconsensus heptad repeat. The seventh residue is modified from Ser7 to Arg or Lys in the nonconsensus heptad repeat. Nonconsesus repeats form the distal part of the CTD. Pol II was illustrated with BioRender.

on its interaction with various host factors [40, 157, 158].

The transcription cycle encompasses initiation and promoter-proximal pausing required for the correct processing of the nascent RNA transcript, followed by elongation and termination (Figure 1.9). The phosphorylation of the CTD is so far the best understood, however, the so-called 'CTD code' has not been fully solved yet [153]. Phosphorylation and glycosylation occur on all three serine residues, tyrosine, and threonine during the transcription process, while the two proline residues are susceptible to cis-trans isomerization (Figure 1.8 B) [152–154]. The non-consensus residues in position 7 are subject to acetylation, methylation, or ubiquitination. Kinases and phosphatases, such as the cyclin-dependent kinases (CDKs) tightly regulate the phosphorylation and dephosphorylation of

these amino acids throughout transcription (Figure 1.8 B) [159]. Despite the theoretically large number of combinations of different PTMs, phosphorylation on the serine residues, in particular serine 2 and serine 5, is the most dominant. Importantly, at most one residue is modified within each heptad repeat [153].

Serine 5 phosphorylation (pS5) peaks during early transcription and decreases sharply after the transition into elongation [153]. Transcriptional and co-transcriptional processes heavily depend on serine 5 phosphorylation, for example for the transition from pause release into elongation and RNA processing such as pre-mRNA capping [153]. The interaction of pS5 CTD with the spliceosome indicates that the CTD acts as a platform for both transcription and co-transcriptional processes such as splicing [160].

In the transcription cycle, various CDKs catalyse the pS5 modification of the CTD including CDK 7, which acts during early transcription as part of the transcription factor II H (TFIIH) complex and phosphorylates both serine 5 and serine 7 residues. Serine 5 phosphorylation leads to the recruitment of capping factors [151, 153]. Serine 2 phosphorylation (pS2) levels are initially low and increase during elongation, catalysed by CDK9 forming part of the positive transcription elongation factor b (P-TEFb) complex [153, 161]. The pS2 modification plays a pivotal role in the transition from promoter-proximal pausing into productive elongation and it is required for the recruitment of elongation factors such as Spt6 and the positive elongation factor PAF1 complex (PAF) as well as the 3' RNA processing machinery [151, 162–164]. The function of the phosphorylation on serine 7, as well as threonine and tyrosine, is still not fully understood [165]. Serine 7 phosphorylation (pS7) is required for recruitment of the pS5 phosphatase RNA polymerase associated protein 2 (RPAP2) and acts together with pS2 [166]. Compared to the continuous presence of serine 2 and serine 5 residues in each repeat, serine 7 residues are often replaced in the distal region of the CTD by non-consensus amino acids, lowering serine 7

presence across the CTD [153]. Mutation of threonine 4 causes elongation defects indicating its role in transcription termination and 3' end processing [154, 157, 167]. Phosphorylation of tyrosine 1 increases during promoter-proximal pausing and affects the termination of transcription, directing P-TEFb to phosphorylate serine 2 for the paused Pol II complex to transition into productive elongation [168–170].

1.3.2 Overview of Pol II transcription

Pol II-mediated transcription is a tightly regulated process, involving various protein factors [171]. Transcription initiation requires the recruitment of general transcription factors (GTFs) to the promoter regions of the genome that together with Pol II form the pre-initiation complex (PIC) (Figure 1.9 A) [151, 172]. The Mediator complex stabilises the PIC and binds to GTFs and the Pol II CTD which is mostly unphosphorylated during the early transcription stages [157, 173–176]. The Mediator complex also recruits the TFIIH complex that phosphorylates serine 5 and serine 7 residues of the CTD through its CDK7 module resulting in the promoter release of Pol II and the dissociation of the Mediator complex (Figure 1.9 B) [40, 177]. Pol II transitions into promoter-proximal pausing by halting its activity 20-80 nucleotides downstream of the transcription start site (TSS), an important control point to ensure the initial capping of the nascent transcript [178]. The recruitment of the negative elongation factor (NELF) complex and DRB sensitivity-inducing factor (DSIF), composed of the Spt5 and Spt4 subunits, stabilises the paused Pol II (Figure 1.9 C)[151]. Structural studies have provided mechanistic details on the mechanism through which NELF and the DSIF complex inhibit Pol II from transitioning into productive elongation [172, 179, 180]. NELF and DSIF both form extensive binding interfaces with Pol II. NELF restricts Pol II in a confirmation in which the DNA-RNA hybrid is tilted, blocking elongation and the addition of NTPs [151].

DSIF forms interactions with Pol II, the DNA template and the RNA product mainly through its large Spt5 subunit composed of various subdomains [40, 172,

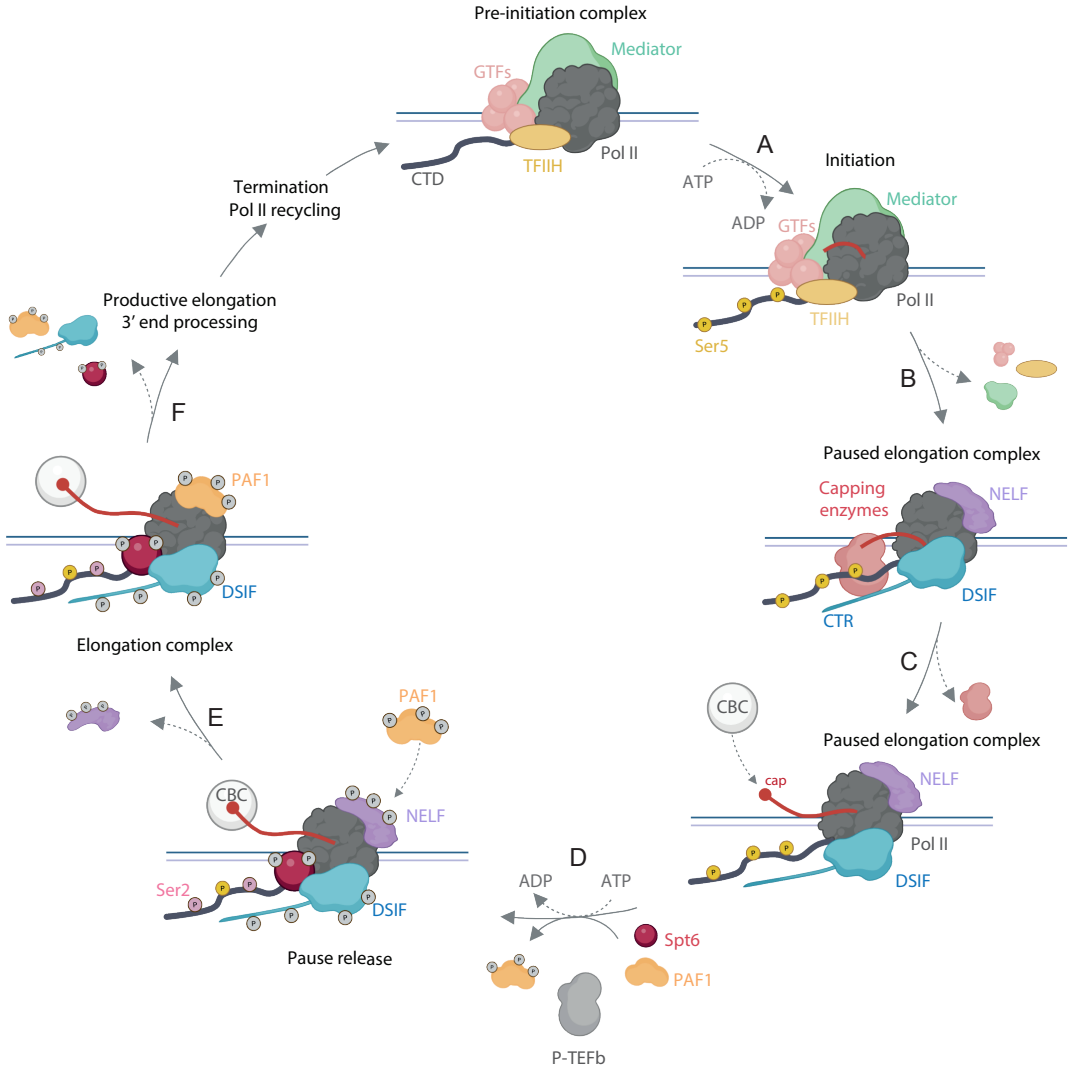

Figure 1.9: Overview of the Pol II transcription cycle. **(A)** General transcription factors (GTFs) in the promoter region recruit the Mediator complex, TFIIH, and unengaged Pol II to form a pre-initiation complex. Ser5 residues on the Pol II CTD are phosphorylated (pS5) by the CDK7 module of TFIIH (yellow circle). **(B)** Phosphorylation of Ser5 induces Pol II promoter escape and recruits capping enzymes, which bind to pS5 CTD and the CTR of the DSIF complex. This leads to the formation of a paused elongation complex which is stabilised by the binding of the pausing factors NELF and unphosphorylated DSIF. **(C)** Promoter proximal pausing shortly after promoter escape facilitates the synthesis of an m^7G cap on the nascent transcript by capping enzymes and the recruitment of the cap-binding complex (CBC) to the nascent capped transcript. **(D)** P-TEFb releases Pol II from its paused state by phosphorylating Ser2 (light red circle) on the Pol II CTD as well as DSIF, NELF, and the elongation factors PAF1 and Spt6 (grey circles). **(E)** Phosphorylated DSIF acts as an elongation factor with phosphorylated Spt6. PAF1 competes with NELF for exclusive binding to Pol II, promoting pause release and the formation of a stable elongation complex. **(F)** Later in transcription, the elongation complex transitions from productive elongation to 3' end processing and transcription termination. Recycled Pol II can engage in a new transcription cycle. Parts of the figure were created with BioRender.

179, 180]. Spt5, like RBP1, also has a C-terminal region (CTR) composed of pentapeptide repeats. It facilitates the recruitment of transcription factors such as capping enzyme (CE) through post-translational modification and provides a scaffold for transcription factors (Figure 1.9 C) [151, 181]. However, unlike the highly conserved CTD of Pol II, this region is not as conserved and can be phosphorylated by P-TEFb at serine and threonine residues [181].

Capping of the nascent RNA transcript takes place co-transcriptionally in the context of the paused elongation complex [151]. A 5' m^7G cap is added to the pre-mRNA by capping enzymes, including RNA 5'-triphosphatase (RT), RNA guanine-7 methyltransferase (RNMT) and RNA guanylyltransferase (RNGTT) [105, 182]. In mammalian cells, capping enzyme (CE) performs the functions of the RT and RNGTT [40]. The position of the RNMT allows simultaneous interaction with the 5' end of the nascent RNA and phosphorylated serine 5 on the Pol II CTD. This leads to the recruitment of RNGTT, followed by addition of a guanosine nucleotide and RNMT-mediated methylation. Methylation is necessary for recognition by the cap-binding complex (CBC) and efficient downstream translation of the mRNA in the cytoplasm [40, 183]. Furthermore, CE forms binding interfaces with Spt5 CTR, the Pol II core, and the pS5 CTD of Pol II that leads to its allosteric activation and positioning near the RNA exit tunnel [40, 184, 185]. Capping is essential for shielding the newly formed pre-mRNA from exoribonuclease degradation and protecting the RNA products from being identified as foreign molecules that could activate host innate immunity pathways [41, 186]. Downstream processes such as splicing are also performed co-transcriptionally with strong involvement of the Pol II CTD [151, 187].

At this early elongation regulatory checkpoint, Pol II has to transition from the paused state into productive elongation to prevent premature transcription termination mediated by the Integrator complex [188]. P-TEFb drives this transition through its CDK9 module by simultaneous phosphorylation of serine 2

residues on the Pol II CTD, NELF, the Spt5 CTR, and PAF (Figure 1.9 D) [40, 151, 189]. Structural studies have revealed that PAF outcompetes NELF for Pol II binding due to overlapping binding sites and steric incompatibility [172, 190]. Spt6 is recruited to the phosphorylated CTD linker which together with the phosphorylation of the Spt5 CTR, triggers a conformational change in DSIF to release Pol II from its paused state [189]. The P-TEFb-driven phosphorylation of the paused elongation complex is essential in the shift from pausing to productive elongation (Figure 1.9 E) [190]. Additionally, Tat-stimulatory factor 1 (Tat-SF1) is also involved later during pause release in concert with PAF and DSIF to aid productive elongation [151, 191–193]. In the later transcription stages, pS2 levels increase further which is required for recruitment of the 3' processing machinery that performs polyadenylation, termination and splicing (Figure 1.9 F) [153]. Details about the later stages of the transcription cycle, including molecular aspects of transcription termination and Pol II recycling, still require further investigation. Despite a substantial collection of approximately 200 Pol II complex structures, our understanding of the transcription cycle remains incomplete. There is also a need to understand the spatiotemporal coupling of Pol II transcription with other processes in the nucleus, such as splicing and 3' end processing of RNA transcripts [151].

1.3.3 Interaction of influenza virus polymerase with the host transcription machinery

RNA viruses such as influenza virus, among others, have been shown to hijack the host transcriptional apparatus by several means [194]. The influenza virus relies on its interaction with the Pol II CTD to perform cap-snatching; an essential process for viral transcription. Various studies established that FluPol specifically recognises and directly binds to the pS5 form of Pol II CTD, strategically placing it to excise the nascent RNA transcript during the early stages of Pol II-mediated transcription [126, 195]. Although structural studies have confirmed the presence of a bipartite CTD binding site on FluPol from influenza viruses A, B, and C,

it remains unclear how FluPol interacts with the host transcriptional apparatus beyond these binding sites [26, 96, 97]. This is due to the extended, disordered nature of the CTD and the numerous transcription factors that have been identified as interactors with FluPol using proteomics approaches [196–198].

The FluPol interaction network expands beyond Pol II to involve multiple host factors linked to transcriptional elongation. High-throughput studies including proteomics and knockdown studies have provided extensive amounts of data that could provide the basis for establishing the links between the viral and host transcription. However, the data sets show little overlap potentially attributed to different experimental conditions [40, 196, 197, 199]. Proteins hypothesised to be involved in the process of cap-snatching, through direct or indirect interaction, have been identified across various studies and reviewed by Krischuns et al. (2021) [40]. Only a small number of the identified hits have been experimentally confirmed, and there is still limited understanding of their specific roles in influenza infection [40, 41].

To date, only a few studies have investigated FluPol nuclear mobility within the context of live cells [40]. Using fluorescence recovery after photobleaching (FRAP) experiments, it was shown that the nuclear mobility of vRNPs is enhanced following inhibition of Pol II with α-amanitin [200]. Analyses of proteins and experiments involving immunoprecipitation have primarily discovered associations with elongation factors. Notably, only a limited few were connected to the Mediator complex or the overall transcription initiation machinery [40]. This indicates that FluPol primarily disrupts transcription after the initiation stage has occurred [40, 41]. The DSIF complex was reported in independent studies to interact with FluPol through both its Spt5 and Spt4 subunits [40, 196–198]. PARP1, which supports transcription elongation through ADP-ribosylation on NELF has been reported in multiple studies as an interactor of FluPol that regulates FluPol activity [196, 201–205]. Other proteins that drive Pol II pause

release and productive elongation such as PAF, Spt6, P-TEFb and Tat-SF1 have been reported to interact with FluPol with P-TEFb impact FluPol activity when overexpressed or depleted in RNAi screens [201, 202]. Further experimental validation of Tat-SF1 suggests that it stimulates viral replication by interacting with NP in the vRNP complex rather than through direct interaction with FluPol [40, 206, 207]. Moreover, various proteins listed in proteomics studies have been linked to the capping process and have been reported to enhance viral replication [40, 208]. As viral genome replication occurs in the nucleus, various nuclear proteins were suggested to interact with FluPol directly or indirectly regulate its functions indirectly [40]. For instance, FluPol has been shown to interact with the nuclear RNA exosome complex, an essential multi-subunit complex that orchestrates RNA homeostasis, to enhance cap-snatching by attracting FluPol to host gene promoter regions [209]. So far, detailed mechanistic insights into these interactions are poorly understood. Biochemical characterisation of the interaction of these host factors with FluPol would allow the precise determination of the time window during which FluPol hijacks Pol II-mediated transcription, providing the basis for the structural elucidation of these interactions [40, 41].

1.3.4 Impact of influenza virus infection on host transcription

FluPol performs cap-snatching and positions itself close to the RNA splicing and export machinery, required for downstream processing of viral mRNAs [41]. Despite the virus's dependence on the host transcription apparatus, viral infection reduces Pol II transcriptional activities, particularly affecting elongation and termination. This dysregulation of host transcription eventually leads during later stages of infection to the degradation of Pol II and host shut-off caused by inhibition of cellular gene expression [67, 210, 211].

Studies demonstrated that the FluPol-Pol II interaction alone is only part of a complex network of interactions between FluPol, Pol II and other host transcriptional factors allowing FluPol to hijack the host transcriptional machinery [40, 41]. Influenza virus infection is detrimental to the cell as the influenza virus exploits host functions, for instance, viral mRNA can make up to 50% of the total mRNA in infected cells [212]. Cellular transcript cleavage disrupts normal transcription, reducing the quantity of elongating Pol II complexes and exposing the decapped nascent transcript to degradation by exonucleases such as Xrn2 [124, 209, 213].

The interaction of viral FluPol with Pol II also drives the degradation of RBP1, the largest Pol II subunit, resulting in dysfunctional Pol II [210]. As transcription initiation depends on the interaction with Pol II, the transition to replication observed later in infection could be due to the ratio of FluPol to Pol II in the cell. As infection advances and more FluPol is generated while Pol II is degraded, a larger population of FluPol may not have access to capped RNA, leading it to prioritise replication [67, 211]. Additionally, the large number of FluPol interactors beyond the CTD indicates that FluPol interaction can inhibit transcription factors resulting in dysregulated or halted Pol II transcription [40, 44]. The viral NS1 protein exacerbates these effects by disrupting RNA processing and elongation factor functions, contributing to a comprehensive suppression of cellular antiviral responses, without hindering viral polymerase functions [139].

In conclusion, the interaction of FluPol with the cellular transcription apparatus demonstrates a complex viral strategy of controlling host cellular processes for efficient viral mRNA synthesis while undermining host cellular transcription. A deeper molecular understanding is needed to inform the development of targeted antivirals and enhance the current understanding of influenza pathogenesis.

1.4 Aims of this study

Influenza viruses pose significant challenges to human health and the global economy. The molecular details of viral genome transcription and replication, key steps in the viral life cycle, are still not fully understood. My DPhil work aimed to understand better the role of host factors, particularly RNA polymerase II, in viral transcription.

In the first project, I characterised the binding of a pS5 Pol II CTD mimic peptide bound to the pandemic 1918 influenza virus polymerase to understand how critical residues in the determined binding site influence viral genome transcription and replication (Chapter 3). Following that, I established the binding of long recombinant CTD mimic peptides containing 26 repeats instead of the four repeats present in synthetic CTD mimics, and I carried out biochemical interaction analyses to perform cryo-EM structural analyses for the long peptide as well as endogenous full-length Pol II (Chapter 4). As previous studies mainly investigated the binding of pS2 and pS5 modification of the Pol II CTD, I was interested in understanding how FluPol binding to the CTD is influenced beyond the determined binding sites and phosphorylation states. First, I designed a combination of various phosphorylation patterns and I performed interaction and functional analyses to assess how these influence the binding of the Pol II CTD to FluPol, given the complexity of the Pol II CTD code. Second, given the negative electrostatic surface potential of the PB2-C domain, I investigated how this domain might contribute to the binding of the Pol II CTD in FluPol A where no CTD binding site has been described so far (Chapter 5). Lastly, I conducted interaction analyses of host factors, hypothesised to modulate the FluPol-Pol II interaction. Specifically, I investigated the role of the DSIF complex and its Spt5 subunit as well as ANP32 and NEP, involved in viral replication and export, respectively (Chapter 6).

2
Materials and Methods

2.1 Materials and reagents

2.1.1 DNA and RNA oligonucleotides

DNA oligonucleotides

For this study, all DNA oligonucleotides were acquired from Sigma-Aldrich and subsequently dissolved and diluted to a concentration of 100 mM using nuclease-free water (Ambion) (Tab. 2.1).

Table 2.1: DNA oligonucleotides used in this study.

Name	Sequence (5' – 3')
5S 100	TCCCAGGCGGTCTCCCATCC
NA 1280	TGGACTAGTGGGAGCATCAT
NA 160	TCCAGTATGGTTTTGATTTCCG
NS1-226	CGCTCCACTATTTGCTTTCC
NS-734	TGATTGAAGAAGTGAGACACAG
pCAGGS PB1a fwd	AGCAAGTCTAGAGCCTCTGCTAACCATGTTCATGCCTTC
pCAGGS PB1a rev	GTCACACTCGAGCATGCCCGGGGGAGGTATATCTTTCG
DSIF-Spt5-rev	TCGCCATCAATACTCAGCAG
DSIF-Spt4-rev	ACCAATCTGTTCTCTGTGAGC
DSIF-Spt4-forw	ATTTGGACATGGAGGATAACG
1918 PA F1	GCCTCTGCTAACCATGTTCATGC
1918 PA F2	CACAAAGGCCGACTACACTC
1918 PA F3	TGCAGGATATTGAGAATGAGGAG
1918 PA F4	GATCCAAGACTTGAACCGCAC
1918 PB1 F1	AGCCTCTGCTAACCATGTTCATGC
1918 PB1 F2	AGTGGACAAGCTGACGCAAG
1918 PB1 F3	GAAGAAAGCCAAGTTGGCAAATG
1918 PB1 F4	GTGGGATGGTCTTCAATCTTCTG
1918 PB1 F5	AGAAATCTCCACATTCCTGAAGTC
P1 XbaI fwd	ATCACAGTATCTCTAGAGCCTCTGCTAACC
P2 XhoI rev	CATCGACTATCTCGAGGCATATGGCATATGTTGC
1918 PA K289A fwd	CCTGCTGATGGATGCCTTGGCATTAAGCAT TGAGGACCCA
1918 PA K289A rev	TGGGTCCTCAATGCTTAATGCCAAGGCATC CATCAGCAGG
1918 PA L290A fwd	CCTGCTGATGGATGCCTTGAAAGCAAGCA TTGAGGACCCAA
1918 PA L290A rev	TTGGGTCCTCAATGCTTGCTTTCAAGGCA TCCATCAGCAGG
1918 PA T313A fwd	CGATCAAGTGCATGAGAGCATTCTTTGGGTGGAAG
1918 PA T313A rev	CTTCCACCCAAAGAATGCTCTCATGCACTTGATCG

1918 PA E416A fwd	GTTCAACAAGGCATGCGCACTGACCGATTCAAGCT
1918 PA E416A rev	AGCTTGAATCGGTCAGTGCGCATGCCTTGTTGAAC
1918 PA D419A fwd	GCATGCGAACTGACCGCTTCAAGCTGGATAGAG
1918 PA D419A rev	CTCTATCCAGCTTGAAGCGGTCAGTTCGCATGC
1918 PA Y445A fwd	ATTGCAAGCATGAGAAGGAATGCTTTCACA GCAGAGGTGTCTC
1918 PA Y445A rev	GAGACACCTCTGCTGTGAAAGCATTCCTT CTCATGCTTGCAAT
1918 PA E449A fwd	GGAATTATTTCACAGCAGCGGTGTC TCATTGCAGAGC
1918 PA E449A rev	GCTCTGCAATGAGACACCGCTGCTGTGAAATA ATTCC
1918 PA R454A rev	CATTATGTATTCTGTGGCTGCGCAATGAGACAC CTCTGCT
1918 PA R454A fwd	AGCAGAGGTGTCTCATTGCGCAGCCACAGAATA CATAATG
1918 PA L527A rev	TTCAAGTCTTGGATCAGTGGCGGAAAACTCC ATGCTCACA
1918 PA L527A fwd	TGTGAGCATGGAGTTTTCCGCCACTGATCCAAGA CTTGAA
1918 PA R551A rev	GGCCTATGGCACTTGCTAGAAGCATATCCCCTAT CTCAAGGA
1918 PA R551A fwd	TCCTTGAGATAGGGGATATGCTTCTAGCAAGT GCCATAGGCC
1918 PA R559A rev	AAGAACATGGGCGCTGACACTTGGCCTATGGCACTTC
1918 PA R559A fwd	GAAGTGCCATAGGCCAAGTGTCAGCGCCCATGTTCTT
1918 PA F612A rev	GGCCATGTTTCTGATTTGTTCTCAGCGAATTCT TTGGTCATGTCTTTCTC
1918 PA F612A fwd	GAGAAAGACATGACCAAAGAATTCGCTGAGAAC AAATCAGAAACATGGCC
1918 PA K635A rev	AACAAAGTCCTGCAGACCGCCCCAATGGAACCTTCCTC
1918 PA K635A fwd	GAGGAAGGTTCCATTGGGGCGGTCTGCAGGACTTTGTT
1918 PA T639A rev	CGATTTTGCCAACAAAGCCCTGCAGACCTTCCCAA
1918 PA T639A fwd	TTGGGAAGGTCTGCAGGGCTTTGTTGGCAAAATCG

RNA oligonucleotides

RNA oligonucleotides, designed by Dr Jeremy Keown and Dr Narin Hengrung and sourced from Dharmacon Inc., were diluted to 4 mM stocks using nuclease-free water (Tab. 2.2).

Table 2.2: RNA oligonucleotides used in this study.

Name	Sequence (5' – 3')
1918 FluA 5' vRNA	AGUAGAAACAAGGCC
1918 FluA 3' vRNA	GGCCUGCUUUUGCU
1918 FluA 3' vRNA (3 nt extension)	GGCCUGCUUUUGCUAUU
FluB 5' vRNA	AGUAGUAACAAGAGG
FluB 3' vRNA	CCUCUGCUUCUGCU
FluC 5' vRNA	AGCAGUAGCAAGGGG
FluC 3' vRNA	CCCCUGCUUCUGCU

2.1.2 Plasmids

Plasmids used in this study and their sources are outlined below (Tab. 2.3).

Table 2.3: List of plasmids used in this study.

Plasmid	Origin
pCAGGS-1918-PB1a	Keown, Zhu, Carrique, Fan et al. (2022) [150]
pCAGGS-1918-PB1	Tumpey et al. (2005) [214]
pcCAGGS-1918-PB2	Tumpey et al. (2005) [214]
pCAGGS-1918-PA	Tumpey et al. (2005) [214]
pCAGGS-1918-NP	Tumpey et al. (2005) [214]
pcDNA3	Fodor et al. (2003) [215]
pPOLI-NA	Fodor et al. (1999) [216]
pPOLI-Luc	Manz et al. (2016) [217]
pcDNA-WSN-PB1	Fodor et al. (2002) [218]
pGEX-6P-1-huANP32B	Wang et al. (2022) [219]
pcDNA-WSN-PB2-luc1	F. Wang
pcDNA-huANP32B-luc2	F. Wang
pCAGGS-1918-PA K289A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA L290A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA T313A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA E416A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA D419A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA Y445A	Generated in this study. Keown, Baazaoui et al. (2024) [112]

pCAGGS-1918-PA E449A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA R454A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA L527A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA R551A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA R559A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA F612A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA K635A	Generated in this study. Keown, Baazaoui et al. (2024) [112]
pCAGGS-1918-PA T639A	Generated in this study. Keown, Baazaoui et al. (20223) [112]
pGEX-6P-Spt5-CTR	J. Keown
pETDuet-1 DSIF	Bernecky et al. (2017) [179]
pET151/D-TOPO GST-Viet PB2-C	Veler et al. (2022) [220]
pET151/D-TOPO GST-Fj PB2-C	Veler et al. (2022) [220]
pET151/D-TOPO GST-NT60 PB2-C	Veler et al. (2022) [220]
pGEX-6P-1 GST-NT60 627-NLS	Veler et al. (2022) [220]
pGEX-6P-1 GST-NT60 627	H. Fan
pGEX4T1-GST-CTD_{26x}-His_{7x}	O. Jasnovidova
pGEX4T1-GST-CTD_{13x}-His_{7x}	O. Jasnovidova
pFastBac 438B His_{6x}-CDK7 His_{6x}-MAT1 His_{6x}-CCNH	M. Šebesta

2.1.3 Synthetic RNA polymerase II CTD mimic peptides

Peptides were produced via solid-phase peptide synthesis (Peptide Protein Research Ltd). These peptides, designed to include four heptad repeats of the Pol II CTD consensus sequence (YSPTSPS), were chemically modified during their synthesis to represent various phosphorylation states at serine residues in positions 2, 5, and 7 of the CTD. Detailed amino acid sequences can be found in (Tab. 2.4). Each peptide was synthesised to include a C-terminal amidation and N-terminal biotinylation, along with a nine-atom polyethylene glycol spacer. Purification was conducted using high-performance liquid chromatography, achieving a minimum purity of 90%. The quality of the peptides was verified through mass spectrometry.

Table 2.4: Synthetic Pol II CTD mimic peptide design. CTD peptides mimic different phosphorylation states of Serine (pS) in each CTD heptad repeat. Serine 2 (pS2), 5 (pS5), and 7 (pS7) were phosphorylated. * Alternating phosphorylation of serine residues per heptad repeat. ** Double phosphorylation of serine residues per heptad repeat.

Peptide	Sequence
pS2	Y(pS)PTSPS Y(pS)PTSPS Y(pS)PTSPS Y(pS)PTSPS
pS5	YSPT(pS)PS YSPT(pS)PS YSPT(pS)PS YSPT(pS)PS
pS7	YSPTSP(pS) YSPTSP(pS) YSPTSP(pS) YSPTSP(pS)
pS(2/5).1*	Y(pS)PTSPS YSPT(pS)PS Y(pS)PTSPS YSPT(pS)PS
pS(2/5).2*	YSPT(pS)PS Y(pS)PTSPS YSPT(pS)PS Y(pS)PTSPS
pS(5/7).1*	YSPT(pS)PS YSPTSP(pS) YSPT(pS)PS YSPTSP(pS)
pS(5/7).2*	YSPTSP(pS) YSPT(pS)PS YSPT(pS)PS YSPT(pS)PS
pS2 pS5**	Y(pS)PT(pS)PS Y(pS)PT(pS)PS Y(pS)PT(pS)PS Y(pS)PT(pS)PS
pS5 pS7**	YSPT(pS)P(pS) YSPT(pS)P(pS) YSPT(pS)P(pS) YSPT(pS)P(pS)
Scrambled	PSSSTPSSYTPSPSSSPTSYSPYYTSPP

2.1.4 Antibodies

Antibodies used in this study are summarised in 2.5.

Table 2.5: Antibodies used in this study.

Antibody	Species	Dilution	Manufacturer
Anti-NP	Rabbit	1:5,000	Genetex (GTX636741)
Anti-PA	Rabbit	1:1,000	Engelhardt et al. (2005) [195]
Anti-PA	Rabbit	1:1,000	Genetex (GTX636828)
Anti-Pol II CTD	Mouse	1:300	Abcam (ab26721)
Anti-Pol II	Mouse	1:500	Millipore
Anti-GAPDH	Rabbit	1:1,000	Cell Signalling Technology
Anti-ALFA	Rabbit	1:500	Nanotag Biotechnologies (N1580)
Anti-Gaussia luciferase	Rabbit	1:2,000	ThermoFisher Scientific (PA1-181)
Anti-rabbit IgG 800CW	Goat	1:10,000	LI-COR Biosciences (926-32211)
Anti-mouse IgG 800CW	Goat	1:10,000	LI-COR Biosciences (926-32210)
HRP anti-mouse IgG	Rabbit	1:10,000	Genetex (GTX26728)
HRP anti-rabbit IgG	Goat	1:10,000	Genetex (GTX213110-01)

2.2 Molecular cloning

2.2.1 Competent cells and bacterial culture

Chemically competent *Escherichia coli (E. coli)* (sourced from The Division of Structural Biology, University of Oxford) were cultured in either lysogeny broth (LB) or 2×YT media at 37°C, 180 rpm (Mixer HC, STARLAB). Culture media was supplemented with the required antibiotics, depending on the antibiotic resistance gene of each plasmid employed. To generate chemically competent DH5α, Rosetta (DE3), and BL21(DE3) *E. coli* cells, overnight cultures of 5 ml were grown, followed by inoculation of 500 ml of antibiotic-free 2xYT medium and incubation at 37°C until an OD_{600} of 0.3-0.4 was reached. The culture was then cooled on ice for 15-20 min to stop growth and cells were harvested by centrifugation at 3,500 g for 10 min at 4°C. After removal of the supernatant, the pellet was gently resuspended in 100 ml of ice-cold 0.1 M $MgCl_2$, followed by centrifugation at 3,500 g at 4°C for 10 min. The pellet was then gently resuspended in 100 ml of ice-cold 0.1 M $CaCl_2$. After another spin, cells were finally resuspended in 20 ml ice-cold 0.1 M $CaCl_2$ in 15% glycerol. Aliquots in sterile tubes were flash-frozen in liquid nitrogen and stored at -80°C.

2.2.2 Transformation

E. coli DH5α, Rosetta (DE3) and BL21(DE3) cells were thawed on ice. An aliquot of 50 µl of bacterial cells was incubated with 2 µl of plasmid DNA for 30 min on ice, followed by a heat shock at 42°C for 60 sec. Cells were then chilled on ice for a further 10 min and recovered in 500 µl LB broth at 37°C, 1,000 rpm (Mixer HC, STARLAB) for 1 h. The transformed cells were then plated on LB agar plates supplemented with 100 µg/ml ampicillin and kept at 37°C overnight. For *E. coli* Rosetta (DE3) cells, LB agar plates were supplemented with 100 µg/ml ampicillin and 50 µg/ml chloramphenicol.

2.2.3 Preparation of plasmid DNA

Plasmid DNA was isolated using either QIAprep Spin Miniprep, QIAprep Midiprep or QIAprep Maxiprep Kit (Qiagen), depending on the required amounts of plasmid DNA using the manufacturer's instructions.

2.2.4 PCR and site-directed mutagenesis

All PCR reactions were run in a ProFlex PCR System Applied Biosystems thermal cycler (ThermoFisher Scientific). PCR reactions were performed using Phusion High-Fidelity (HF) DNA Polymerase (NEB). Reactions were performed in 50 µl reactions with 10 ng plasmid DNA, 2.5 µl of each 10 µM forward, 2.5 µl of 10 µM reverse primer, 1 µl of 10 mM dNTPs, 1.5 µl of 100%(v/v) DMSO, and 10 µl of 5× Phusion HF Reaction Buffer (NEB).

For site-directed mutagenesis, a PCR was carried out using Pfu Turbo DNA Polymerase (Agilent) in 50 µl reactions according to the manufacturer's recommendations. PCR reactions were then digested with Dpn I (NEB) for 1 h at 37°C, followed by transformation into DH5α *E. coli* cells. The presence of the introduced mutations was confirmed by Sanger sequencing (Source BioScience).

2.2.5 Gibson Assembly

Primers for Gibson assembly were designed using the NEBuilder primer design tool (NEB). PCR was performed using the Phusion High-Fidelity DNA Polymerase (NEB) following the manufacturer's guidelines. The PCR products were analysed by agarose gel electrophoresis. Relevant DNA fragments were extracted using the QIAquick Gel Extraction Kit (Qiagen), following the manufacturer's protocol. The NEBuilder HiFi DNA Assembly mix (NEB) was used for Gibson assembly following the manufacturer's instructions. The assembly products were transformed into DH5α *E. coli* cells and plated on LB agar containing 100 µg/ml ampicillin. All sequences were confirmed by Sanger Sequencing (Source BioScience).

2.2.6 Mutagenesis of pCAGGS plasmids

Mutagenesis of pCAGGS plasmids was perfomed by splitting the gene subject to mutagenesis into two segments using two sets of primers with each set having one primer introducing the mutation and one primer outside the gene. For mutagenesis of the pCAGGS_1918_PA plasmid, the following primers where used: P1_XbaI (forward primer outside the PA gene containing the Xba I restriction site) together with M(114)_rev (a set of reverse primers each encoding a mutation in the PA subunit) and P2_XhoI (reverse primer outside the PA gene containing the Xba I restriction site) together with M(1-14)_fw (a set of reverse primers each encoding a mutation in the PA subunit).The mutagenesis PCR was carried out in a reaction volume of 50 µl using Pfu Turbo DNA Polymerase (Agilent) following the manufacturer's instructions. After separation of the PCR products on an agarose gel, the PCR were extracted from the gel using the QIAquick Gel Extraction kit (Qiagen). The segmented mutated gene segments were assembled by PCR using P1_XbaI and P2_XbaI which are a set of primers outside of the gene yielding the entire amplified gene flanked by restriction site overhangs. The PCR products were separated by an agarose gel and relevant band were extracted from the gel as described. The DNA fragment and pCAGGS backbone vector were then digested with Xba I and Xho I restriction enzymes (NEB) and ligated using ligase (Takara Bio) according to the manufacturer's instructions. Ligation products were then transformed into DH5α *E. coli.* cells and plated. Introduction of mutations into the gene were confirmed by Sanger sequencing (Source BioScience).

2.3 Cell and virus culture

2.3.1 Mammalian cell culture

Human Embryonic Kidney 293T (HEK293T), Madin-Darby Bovine Kidney (MDBK) and Madin-Darby Canine Kidney (MDCK) were sourced from the Sir William Dunn School, University of Oxford. HEK293T cells with a double

knockout in human ANP32A and ANP32B (dKO-ANP32) were sourced from Dr Xiaojun Wang, Chinese Academy of Agricultural Sciences, China. HCT116 cells derived from a human colorectal carcinoma cell line were a kind gift from Dr Steven West, University of Exeter. HEK293T and HCT116 were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal calf serum (FCS). MDBK and MDCK cells were grown in Minimum Essential Medium (Sigma Aldrich) supplemented with 10% FCS and 200 mM (1×) L-glutamine (Gibco). Cells were kept at 37°C, 5% CO_2. For cell passaging, cells were washed using Dulbecco's Phosphate-Buffered Saline (Gibco) and trypsinised with TrypLE Express Enzyme (Gibco) for 3-10 min, depending on the cell line.

2.3.2 Insect cell culture

Sf9 insect cells from *Spodoptera frugiperda* were sourced from The Division of Structural Biology, University of Oxford (Zhang Lab). Cells were grown as a suspension culture in Sf-900 II serum-free medium (Gibco) without antibiotics, and cell density was maintained between 0.5-2 $\times 10^6$ cells/ml. Cells were grown in the dark at 27°C in a shaking incubator at 120 rpm.

2.3.3 Influenza virus propagation

Influenza A/WSN/1933 (H1N1) virus was propagated by infecting 80% confluent MDBK or MDCK cells at a Multiplicity of Infection (MOI) of 0.01. The culture medium was collected 48 h post-infection (hpi), and cell debris was removed by centrifugation at 4,000 rpm (Hettich ROTINA 420, rotor 5624) for 10 min. The virus stock was aliquoted and stored at -80°C. Viral titres were measured by a standard plaque assay [221].

2.4 Protein expression and purification

2.4.1 Insect cell expression and purification of influenza virus polymerase

Influenza virus polymerase was purified following previously published work [144]. *Sf9* insect cells were infected with baculovirus (generated by Dr Haitian Fan) to express FluPol from influenza virus types A (Influenza A/Brevig Mission/1/1918 H1N1), B (Influenza B/Panama/45/1990) and C (Influenza C/Johannesburg/1/1966). Along with WT 1918 FluPol A, a mutant FluPol A, containing a PA D108A mutation (generated by Dr Jeremy Keown) in the endonuclease domain, was used to prevent the cleavage of capped host RNA. For purification, the PB2 subunit of FluPol was designed to contain a C-terminal Tobacco Etch Virus (TEV) cleavage site and a protein A tag. Cells were harvested after infection and incubation for 72 h.

Cell pellets from one litre of culture were resuspended in 50 ml lysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) n-octyl beta-D-thioglucopyranoside (OTG), 1 mM DTT, supplemented with cOmplete EDTA-free protease inhibitor (Roche), 0.1 mg/ml ribonuclease A (RNase A, Qiagen)) and lysed by sonication. The cell lysate was clarified by centrifugation at 35,000 g for 45 min at 4°C. The supernatant was then incubated with IgG sepharose beads (GE Healthcare) for 3 h, followed by washing of the beads with wash buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) OTG, 1 mM DTT). TEV protease was added to the beads to elute proteins overnight. The eluate was concentrated using an Amicon Ultra centrifugal filter (Merck Millipore) with a cutoff weight of 100 kDa and purified by SEC (flow rate 0.5 ml/min) on a Superdex Increase S200 10/300 column (Cytiva) equilibrated in SEC buffer (25 mM HEPES/NaOH pH 7.5, 250 mM NaCl, 5%, 1 mM DTT). Fractions containing pure protein were concentrated, aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

2.4.2 Bacterial expression and purification of Pol II CTD-derived peptides

The repeat sequence of the Pol II CTD was recombinantly expressed in a BL21(DE3) RIL *E. coli* strain. The expression constructs (kindly provided by Dr Marek Šebesta) were cloned in a pGEX4T1 plasmid backbone expressing 26 repeats of the Pol II hepatad consensus sequence YSPTSPS (CTD_{26x}). For a phosphomimetic peptide the Ser5 of each heptad repeat was mutated to a Glu residue (S5E CTD_{26x}). The peptide sequence was flanked by a GST tag at the N-terminus and a C-terminal His_{7x} tag to prevent degradation (Dr Richard Štefl, personal communication). Plasmid propagation was carried out at 30°C. After inoculation of an overnight starter culture with LB medium at 30°C, cells were diluted 1:100 in autoinduction media ZYP5052 and grown at 30°C for 6 h, followed by 16 h at 18°C [222]. The pellet was resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 10 mM imidazole pH 8.0, 1 mM DTT, supplemented with protease inhibitors, 3 mg/ml lysozyme, 0.1 mg/ml RNase). Cells were lysed by sonication, followed by centrifugation to remove cellular debris. The supernatant was incubated with Ni-NTA sepharose beads (Qiagen) for 1 h, followed by washing in wash buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 10 mM imidazole pH 8.0, 1 mM DTT, supplemented with protease inhibitors) and elution with elution buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 400 mM imidazole pH 8.0, 1 mM DTT). The eluate was concentrated and purified by SEC on a Superdex Increase S200 10/300 column (Cytiva) equilibrated in SEC buffer (25 mM HEPES/NaOH pH 7.5, 250 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). Aliquots were flash-frozen in liquid nitrogen and stored at -80°C until further use.

2.4.3 *In vitro* phosphorylation and purification of Pol II CTD-derived peptides

CTD_{26x} was phosphorylated using a CDK7 kinase module following previously described reaction conditions [223]. The reaction was performed with 50 µM CTD_{26x} and 2 µM CDK7 kinase module in kinase reaction buffer (25 mM

HEPES/NaOH pH 7.5, 250 mM NaCl, 5% glycerol, 20 mM $MgCl_2$, 20 µM $ZnCl_2$, 2 mM TCEP) and 8 mM ATP were added to start the reaction which was incubated for 1.5 h at 30°C. The reaction was quenched with 40 mM EDTA followed by SEC to purify phosphorylated CTD_{26x} ($pCTD_{26x}$) in a final buffer of (25 mM HEPES/NaOH pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT).

2.4.4 Insect cell expression and purification of CDK7

Sf9 insect cells were infected with baculovirus containing a human CDK7-MAT1-CCNH kinase module construct (kind gift from Dr Marek Šebesta) to mimic the kinase function of the TFIIH complex *in vitro*. Each subunit has a TEV-cleavable C-terminal His-tag. The infected cells were harvested 72 hpi and the pellets were flash-frozen in liquid nitrogen and stored at -80°C until further use.

The cell pellet was resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 10 mM imidazole pH 8.0, 1 mM DTT, supplemented with protease inhibitors, 3 mg/ml lysozyme, 0.1 mg/ml RNase, 1 µl/200 ml culture volume benzonase). After lysis by sonication and centrifugation of the cell lysate, the supernatant was incubated with Ni-NTA beads (Qiagen) for 1 h. Following washing with wash buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 10 mM imidazole pH 8.0, 1 mM DTT), the protein was eluted in elution buffer (50 mM Tris-Cl pH 8.0, 500 mM NaCl, 400 mM imidazole pH 8.0, 1 mM DTT). The eluate was then concentrated and purified using SEC (25 mM HEPES/NaOH pH 7.5, 250 mM NaCl, 10% (v/v) glycerol, 1 mM DTT).

2.4.5 Purification of native Pol II

The mammalian, twelve-subunit Pol II complex was purified by Dr Marek Šebesta. Pol II was purified from calf *(Bos taurus)* thymus (New Zealand Biomaterials) to near homogeneity by successive rounds of ion exchange, immunoaffinity, ion exchange, and size-exclusion chromatography as described with modifications [224]. The pooled sample from the second ion exchange chromatography on a MonoQ

column (Cytiva), was concentrated to ~500 µL, followed by *in vitro* phosphorylation. For phosphorylation of Ser5 and Ser7 residues, CDK7 complex (100 µg) and a mixture of ATP/$MgCl_2$ (at a final concentration of 2 mM and 3.5 mM, respectively) were added to the concentrated sample. The volume was adjusted to 1 mL, and the mixture was incubated at 30°C for 30 min. Subsequently, the Pol II was separated from CDK7 and ATP on a Superose 6 Increase SEC column (Cytiva). Fractions containing Pol II were pooled, concentrated, snap-frozen in liquid nitrogen, and stored at -80°C.

2.4.6 Mammalian expression and purification of ALFA-tagged Pol II

HCT116 cells stably expressing Pol II with a triple ALFA epitope tag on the C-terminus of the RPB1 subunit were generated by Dr Steven West (University of Exter). The ALFA-tagged Pol II was purified with ALFA Selector PE beads (Nanotag Biotechnologies) and the purification protocol followed the manufacturer's instructions. Cells grown in a T75 flask were harvested, resuspended, and lysed in 1.25 ml 5× cell lysis buffer (Promega) by shaking on ice. The lysate was cleared by centrifugation for 10 min at 14,000 g at 4°C and 20 µl ALFA Selector PE beads (Nanotag Biotechnologies) were equilibrated in 2×1 ml lysis buffer, beads were collected by centrifugation at 1,000 g for 1 min and the supernatant was carefully removed. The cleared lysate was then incubated with the equilibrated beads for 1 h at 4°C with head-over-tail rotation. The beads were then washed three times with wash buffer (20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 5% (v/v) glycerol, 1 mM DTT) and once with PBS. The protein was then eluted twice with 50 µl 200 µM ALFA peptide in PBS (Nanotag Biotechnologies) for 15 min at room temperature. Samples taken throughout the purification were analysed on a NuPAGE 4-20% gradient gel (Invitrogen), followed by silver staining to visualise protein bands.

2.4.7 Bacterial expression and purification of DSIF

DSIF was expressed in BL21 (DE3) RIL *E. coli* as previously described [179]. The expression construct (kindly provided by Dr Patrick Cramer) is a modified pETDuet vector for codon-optimised co-expression of the Spt4 and Spt5 subunits. The first cassette contains the sequence of human Spt4, followed by human Spt5 in the second cassette. The N-terminus of Spt4 contains a His_{10}-Arg_8-SUMO-tag and a 3C protease cleavage site.

BL21 (DE3) RIL *E.coli* were transformed with the plasmid expressing DSIF and grown in LB medium containing kanamycin and chloramphenicol at 37°C to an OD_{600} of 0.6. IPTG at 1 mM was used to induce expression, 10 µM $ZnCl_2$ was supplemented, cells were left to grow further for 3-4 h at 37°C and were finally harvested.

Cells were resuspended and lysed in lysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 10 µM $ZnCl_2$, 30 mM imidazole pH 8.0, 1 mM DTT, supplemented with protease inhibitors). The cell lysate supernatant was incubated with Ni-NTA sepharose beads (Qiagen) for 1 h. The beads were washed, and the protein was eluted using a single-step elution for 1 h with elution buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% glycerol, 10 µM $ZnCl_2$, 400 mM imidazole pH 8.0, 1 mM DTT). The eluate was then incubated overnight with 3C protease and the buffer was exchanged by dialysis to remove excess imidazole with dialysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% glycerol, 10 µM $ZnCl_2$, 30 mM imidazole, 1 mM DTT). The dialysed sample was then incubated with Ni-NTA sepharose beads (Qiagen) for 30 min to remove the cleaved tag. The eluate was then concentrated and further purified using SEC. The identity and quality of purified DSIF was further analysed using tryptic digest and whole protein mass spectrometry (Centre of Medicines Discovery, University of Oxford).

2.4.8 Bacterial expression and purification of GST-Spt5$_{CTR}$

GST-tagged Spt5 was expressed in *E. coli* BL21 (DE3) RIL cells. The construct expresses GST at the N-terminus followed by a 3C cleavage site and the 387 amino acid residues of the C-terminal region of Spt5 (Spt5$_{CTR}$), corresponding to the CTR flanked by the KOW5 and KOW6/7 motifs. An overnight LB pre-culture supplemented with 100 µg/ml ampicillin was grown at 37°C and 180 rpm in a shaking incubator. The cells were then expanded, and the culture was grown until an OD_{600} of ~0.6 was reached to induce with 1 mM IPTG. The culture was incubated further for 6 h at 37°C and 180 rpm in a shaking incubator. Cells were harvested by centrifugation and stored at -20°C.

The cell pellet was resuspended in lysis buffer (50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 10% glycerol, 0.05% OTG, 1 mM DTT, protease inhibitor and lysozyme) and lysed by sonication. Cell debris was removed by centrifugation at 35,000 g for 60 min at 4°C and the supernatant was incubated with pre-equilibrated glutathione sepharose beads (GE Healthcare) on a rotor at 4°C for 2 h. After binding, the beads were washed with wash buffer (50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 10% glycerol, 0.05% OTG, 1 mM DTT). The bound protein was eluted overnight with elution buffer at 4°C (50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 10% glycerol, 0.05% OTG, 20 mM GSH, 1 mM DTT, protease inhibitor). The eluate was then concentrated and GSH was removed by diluting in wash buffer (10-20×of the sample volume) and subsequent concentrating. The concentrated eluate was then purified using a Superdex Increase 10/300 S200 column. Fractions containing the desired protein were pooled and concentrated. The sample was then aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C until further use.

2.5 Protein analysis

2.5.1 Protein concentration measurement

The concentration of protein samples was measured with a Nanodrop 2000 spectrophotometer (Thermo Scientific) based on their absorbance at 280 nm. The

measured concentration was adjusted by taking into consideration the extinction coefficient of the protein of interest according to the Beer-Lambert equation. Extinction coefficients were approximated with the Expasy ProtParam tool (Swiss-Prot Group) [225].

2.5.2 SDS polyacrylamide gel electrophoresis and gel staining

Protein samples were diluted with 4× SDS loading dye (500 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 40% (w/v) glycerol, 2 mM DTT, bromophenol blue) and heated at 95°C for 5 min. The samples were then loaded on either self-made SDS-PAGE gels or on a commercial gradient gel (Invitrogen). Non-gradient SDS-PAGE gels consisted of a stacking gel layer (125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.01% (v/v) TEMED and 4% (v/v) acrylamide/bis-acrylamide (29:1) and a resolving gel layer (375 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) APS, 0.01% (v/v) and 8-15% (v/v) acrylamide/bis-acrylamide (29:1)). Proteins were separated by running the gels in a mini-protean gel apparatus (Bio-Rad) at 150-200 V in running buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% (w/v) SDS). Commercial gels were run using NuPAGE running buffers (Invitrogen) according to the manufacturer's instructions. Protein bands were visualised, depending on the detection limit, by either Coomassie staining using PageBlue Protein Staining Solution (ThermoFisher Scientific) or by silver staining using the SilverXpress kit (Invitrogen) following the manufacturer's instructions.

2.5.3 Western blotting

Immunoblotting was performed to determine protein expression levels. Following protein separation by SDS-PAGE, proteins were transferred onto a nitrocellulose membrane (Amersham) by wet or semi-dry transfer using either a mini TransBlot cell (BioRad) or a semi-dry TransBlot system (BioRad) according to the manufacturer's recommendations. At either blocking or incubation of the membrane with an antibody the membrane was washed three times for 10 min

with PBST (PBS + 0.1% Tween-20). Following protein transfer, the membrane was blocked with blocking buffer (3% (w/v) skim milk powder in PBST) for 1 h at RT. The membrane was then incubated with the primary antibody overnight at 4°C (diluted in 3% BSA in PBST and 0.02% NaN_3), followed by incubation with the secondary antibody for 1 h at RT (diluted in 3% BSA in PBST). When a secondary antibody conjugated to horseradish peroxidase was used, protein bands were visualised using Amersham ECL western blotting detection reagent (GE Healthcare), followed by exposure on X-ray film (Super RC Fuji Medical, Kodak). The use of secondary antibodies conjugated to fluorophores required visualisation of protein bands with an Odyssey DLx imaging system (LiCOR). All antibodies used in this work are listed under 2.1.4.

2.6 Protein interaction analysis

2.6.1 Size-exclusion chromatography

SEC was carried out to analyse protein complex formation. In general, protein samples were analysed on a Superdex 200 Increase 10/300 column (Cytiva) operated by an ÄKTA fast protein liquid chromatography system (GE Healthcare) at 4°C. The UV absorbance at 280 nm was plotted against the elution volume (ml). Fractions were collected, analysed by SDS-PAGE, and concentrated using an Amicon Ultra centrifugal filter where required (Merck Millipore).

FluPol (0.3 µg) was incubated with promotor 5' and 3' vRNA (1.5 molar excess, see sequences under 2.1.1) for 10 min on ice before an excess amount of analyte was added (1.2–3× molar excess, depending on analyte, see results section for exact molar ratios). After incubation with the analyte on ice for 10–30 min, each sample was diluted with the appropriate SEC buffer (25 mM HEPES/NaOH pH 7.5, 150–300 mM NaCl, 5–10% glycerol, 1 mM DTT) and run on a Superdex Increase 10/300 S200 column (GE Healthcare).

2.6.2 CTD-binding assay

The binding of FluPol to synthetic four-heptad repeat Pol II CTD mimic peptides (CTD_{4x}) with various phosphorylation patterns (see 2.1.3 for sequence details) was performed as previously described [126, 150]. Biotinylated Pol II CTD peptides (20 µg) were immobilised on 10 µl streptavidin agarose resin (Thermo Scientific) for 2 h at 4°C in binding buffer (10 mM HEPES pH 8.0 [PAA catalogue no. S11-001], 150 mM NaCl, 0.1% (v/v) Igepal, 1× Halt protease inhibitor cocktail [Pierce], 1% (w/v) bovine serum albumin (BSA)), followed by washing three times for 10 min with wash buffer (10 mM HEPES [PAA catalogue no. S11-001], 150 mM NaCl, 0.1% (v/v) Igepal, 1 mM PMSF). After washing the beads, 4 µg of FluPol in binding buffer were incubated with the peptide-bound beads for 1.5-2 h at 4 °C. Beads were washed four times and heated in SDS-PAGE sample buffer at 95 °C for 5 minutes. Samples were analysed by SDS-PAGE analysis, followed by silver staining to visualise low-abundance protein bands according to the manufacturer's instructions (Invitrogen).

2.6.3 GST pulldown assays

Glutathione sepharose beads (GE Healthcare) were equilibrated in binding buffer (25 mM HEPES pH 7.5, 150-300 mM NaCl, 5%-10% glycerol, 1 mM DTT). Approximately 50 µg of the analytes (GST-tagged proteins) and negative controls (GST tag alone, FluPol alone) were immobilised on 50 µl beads (i.e. 100 µl slurry) in 600 µl binding buffer, followed by incubation for 2 h at 4°C. Following washing with binding buffer, the analyte was added at a molar ratio of 0.5-3×to the immobilised bait, depending on protein properties, and incubated in 600 µl binding buffer for 1 h at 4°C. The beads were washed five times and the complex was eluted twice in 200 µl elution buffer (25 mM HEPES pH 7.5, 150-300 mM NaCl, 25 mM GSH, 5% glycerol, 1 mM DTT), followed by overnight incubation at 4°C. The eluate was collected, and the eluted beads were washed with binding buffer before samples were analysed by SDS-PAGE analysis.

2.6.4 ANP32 and NEP competition assays

Unphosphorylated and pS5 Pol II CTD mimic peptide (20 µg; see 2.1.3) were immobilised on 10 µl streptavidin agarose resin (Thermo Scientific) for 2 h at 4°C in binding buffer (10 mM HEPES pH 8.0 [PAA catalogue no. S11-001], 150 mM NaCl, 0.1 % (v/v) Igepal, 1× Halt protease inhibitor cocktail [Pierce], 1% (w/v) BSA, followed by washing three times for 10 min with wash buffer (10 mM HEPES [PAA catalogue no. S11-001], 150 mM NaCl, 0.1% (v/v) Igepal, 1 mM PMSF). After washing of the beads, 4 µg of FluPol together with a serial dilution of either human ANP32B (kindly provided by Fangzheng Wang) or GST-NEP (kindly provided by Alison Rep) from A/Brevig Mission/1/1918 (H1N1) at molar ratios of 0.33×, 1×, 3×, 9×, and 27× in binding buffer were incubated with the peptide-bound beads for 2 h at 4 °C. Beads were washed four times and heated in SDS-PAGE sample buffer at 95 °C for 5 min. Samples were analysed by SDS-PAGE analysis, followed by silver staining to visualise protein bands according to the manufacturer's instructions (Invitrogen). The protein bands were quantified with ImageJ, followed by data analysis using Excel (Microsoft) and Prism 10 (GraphPad). Measurements were normalised to the conditions without the presence of the competitor protein human ANP32B or GST-NEP, respectively.

2.6.5 Proteomics analysis of FluPol interactors from nuclear extract

Sample preparation of the nuclear extraction was performed by Serdar Baysal, the His_{10} and Strep-tagged WSN FluPol was provided by Dr Haitian Fan. Proteomics sample preparation and analysis was carried out by Marjorie Fournier at The Advanced Proteomics Facility (University of Oxford).

HeLa cells stably expressing Pol II with the RPB9 subunit carrying an N-terminal Flag-tag were harvested. Nuclear extracts were prepared by resuspending the pellet in cytoplasmic fraction solution (50 mM HEPES NaOH pH 7.6, 1.5 mM $MgCl_2$,10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF) with a homogeniser using

four times the volume of the pellet. Following centrifugation at 3,500 rpm for 3 min, the supernatant containing the cytoplasmic fraction was discarded and the pellet containing the nuclear fraction was resuspended with twice the volume of the pellet using nuclear fraction solution (500 mM NaCl, 50 mM HEPES NaOH pH 7.6, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF) and incubated on a rotator at 4°C for 45 min, followed by centrifugation at 13,000 rpm for 15 min. The supernatant, the nuclear extract was used for immunoprecipitation.

FluPol was immobilised using either 12 µl of Ni-NTA (Qiagen) or streptavidin agarose (Thermo Scientific) beads per sample. The beads were washed five times in 1 ml BC300 buffer (300 mM NaCl, 40 mM HEPES NaOH pH 7.6, 5% glycerol, 0.4 mM EDTA, 0.5 mM PMSF, 0.1% NP-40, 0.5 mM DTT 0.1% NP-40 at room temperature. Following washing, 150 µl of BC150 buffer (150 mM NaCl, 40 mM HEPES-NaOH pH 7.6, 0.4 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) was added to the beads along with 0.02% NP-40 and 3 µl of purified NT60 FluPol A (provided by Dr Haitian Fan) at 7 mg/ml carrying a C-terminal His_{10}-tag on the PA subunit and a 2xStrep-tag on the PB2-subunit, respectively. The beads were incubated on a rotator for 3 h at 4°C, followed by centrifugation and removal of the supernatant. After washing with BC150, whole nuclear extract supplemented with 0.02% of NP-40 was incubated on a rotor at 4°C for 3 h. After the incubation, the beads were centrifuged, and the supernatant was discarded. The beads were then washed four times with BC150 supplemented with 0.1% NP-40 followed by three times washing in BC150 without detergent and on-bead digestion. Liquid chromatography with tandem mass spectrometry (LC-MS/MS) was then performed at the Advanced Proteomics Facility (University of Oxford). The identified proteins underwent filtering with a false discovery rate of less than 1%. Hits showing a minimum 2-fold change, as determined by the number of spectral counts divided by the protein length (PSM ratio), were then listed [220, 226].

2.7 Cell-based analysis of influenza viral polymerase

2.7.1 Mini genome assay

HEK293T cells were seeded in a 96-well plate with 5×10^4 per well. The cells were transfected with 160 ng of each plasmid encoding WT or mutant PA, PB1, PB2, and NP of influenza A/Brevig Mission/1/1918 virus as well as 80 ng of plasmid encoding a luciferase vRNA segment (see 2.1.2). For control conditions, a PB1 active site mutant (PB1a) construct or an empty pcDNA3 vector was used. After incubation for 24 h, the cells were lysed in 100 µl Luciferase Cell Culture Lysis Reagent (Promega), followed by vigorous shaking for 15 min. Luciferase enzyme expression was determined using the ONE-Glo luciferase-based detection system (Promega) according to the manufacturer's instructions. In a 96-well opaque plate, 20 µl of the cell lysate was mixed with 80 µl of Luciferase Assay Reagent (Promega). The plate was briefly shaken and read in a SpectraMax M3 (Molecular Devices) microplate reader. Three independent transfections with three technical replicates, respectively, were carried out. Data analysis was performed in Excel (Microsoft) and Prism software (versions 8-10, GraphPad).

2.7.2 Split-luciferase assay

HEK293T double-knockout ANP32A and ANP32B (dKO-ANP32) cells were seeded in 24-well plates and transfected in suspension with 200 ng of each plasmid encoding a chimeric 1918/WSN polymerase with the subunits PB1 (WSN), PB2 (1918), PA (1918) and 200 ng of a plasmid expressing human ANP32B (huB) (see 2.1.2). The split *Gaussia princeps* luciferase was placed C-terminally on the huB (luc1) PB1 (luc2). After 24 h, cells were lysed for 45 min in *Renilla* lysis buffer (Promega) by shaking at room temperature and remaining attached cells were resuspended by pipetting. Luminescence was measured for 10 sec after mixing 10 µl of cell lysate with 10 µl of Renilla luciferase substrate (Promega) in a GLOMAX 20/20 Luminometer (Promega). The normalised luminescence

ratio (NLR) was determined by division of the luminescence activity measured in relative light units (RLU) of cells expressing PA, PB2, PB1-luc2 and huB-luc1 with the sum of the following control conditions: 1. cells expressing luc1 backbone with PB1-luc2, PA, and PB2 and 2. cells expressing luc2 backbone with huB-luc1, PA, and PB2 [227]. Three independent transfections with three technical replicates, respectively, were carried out. Data analysis was performed in Excel (Microsoft) and Prism software (versions 8-10, GraphPad).

2.8 *In vitro* influenza virus assays

2.8.1 Cap-dependent transcription activity assay

Reactions were performed in 10 µl reactions containing 5 mM $MgCl_2$, 1 mM DTT, 2 U RNasin, 1× nucleotide mix (0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.1 µM GTP, 0.15 µl α^{32}P-GTP (3,000 Ci/mmol; Perkin Elmer), 2.5 µM 5' and 3' vRNA (see 2.1.1), 1× reaction buffer (25 mM HEPES NaOH pH 7.5, and 100 mM NaCl) and 71 nM influenza virus polymerase from A/Brevig Mission/1/1918 (H1N1) and C/Johannesburg/1/1966. Capped mRNA was added at 10-20 ng/µl (rabbit β-globin mRNA (Sigma Aldrich) or CleanCap Firefly Luciferase mRNA (TriLink Biotechnologies). CTD synthetic and recombinant peptides (see 2.1.3) were added in a 10× molar excess at 7.1 µM. Reactions were incubated for 1.5 h at 30°C before the reaction was stopped with 10 µl formamide loading dye at 95°C for 3 min. Samples were analysed by 18% PAGE.

2.9 RNA analysis

2.9.1 Extraction of RNA

Cells grown as a monolayer in a 6-well plate were resuspended in 500 µl TRI-Reagent (Sigma Aldrich), 150 µl of chloroform were added, and the mixture was centrifuged at 13,000 g for 15 min at 4°C. For RNA precipitation, the supernatant was carefully moved into a fresh tube and combined with an equal volume of isopropanol to precipitate the RNA. The mixture was centrifuged at 13,000 g for

30 min at 4°C and the supernatant was gently removed. The remaining RNA pellet was washed with 1000 µl of 70% ethanol and centrifuged at 13,000 g for 10 min at 4°C. After discarding the supernatant, the pellet was briefly dried and then resuspended in 10 µl nuclease-free water.

2.9.2 Radiolabelling of oligonucleotides

Oligonucleotides at a concentration of 1 µM were labelled in a total reaction volume of 10 µl containing 10 U of T4 Polynucleotide Kinase (NEB), 1× kinase buffer A, and 1 µl of γ^{32}P-GTP (3,000 Ci/mmol; Perkin Elmer). The reaction mix was incubated at 37°C for 1 h and further cleared by the removal of unincorporated nucleotides using the QIAquick Nucleotide Removal Kit (Qiagen), following the manufacturer's instructions. Radio-labelled primers were eluted in a volume of 30-50 µl nuclease-free water and stored at -20°C.

2.9.3 Primer extension assay

A volume of 1 µl total cellular RNA was mixed with 4 µl of primer mix diluted in water. The primer mix contained 0.25 µl radio-labelled NA 160, NA 1280 as well as 0.1 µl 5S rRNA primer and 0.4 µl of the unlabelled 5S rRNA primer (see 2.1.1). The primers were annealed to the RNA by heating to 95°C for 3 min and subsequently cooling on ice. RNA was reverse transcribed in a reaction volume of 10 µl at 50°C for 1 h, followed by heating to 70°C for 10 min. The reaction mix consisted of 50 U SuperScript III Reverse Transcriptase (Invitrogen), 1× first-strand (FS) buffer, 0.5 mM dNTPs and 10 mM DTT. Samples were denatured by adding 10 µl loading dye (90% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol) and boiling at 95°C for 3 min before loading on a 6% denaturing PAGE gel. Following band quantification as described in 2.9.4, measurements were normalised to the 5S rRNA loading control. Additionally, the vRNA signal in the negative control lane was subtracted from all lanes.

2.9.4 Denaturing polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE) was used to separate radiolabelled RNA and DNA. PAGE gels with a percentage of 6-20% were prepared using a 19:1 ratio of acrylamide:bis-acrylamide (AccuGel, Geneflow) and containing 7M urea dissolved in 1×Tris-Borate EDTA (TBE: 90mM Tris, 90 mM boric acid, 2.5 mM EDTA). The polymerisation of the gel mixtures was achieved by adding 0.1% ammonium persulfate (APS) and 0.1% N,N,N,N'-tetramethylethylendiamine (TEMED). PAGE gels were run at 1,250-2,000 V in 1x TBE and radiolabelled products were visualised by exposure to phosphorimaging screens (FujiFilm), followed by scanning on a FujiFilm FLA-5000 scanner. Quantification of the bands was carried out with ImageJ, followed by data analysis using Prism software (versions 8-10, GraphPad).

2.9.5 RNP reconstitution assay

HEK293T grown to 80% confluency in 6-well plates were transfected with plasmids expressing the vRNP components of influenza A/Brevig Mission/1/1918 to compare the activity of mutant viral polymerases in cell culture [214]. Plasmids expressing WT or mutant pCAGGS-PA-1918, pCAGGS-PB1-1918, pCAGGS-PB2-1918, pCAGGS-NP-1918, and pPOLI-NA-RT expressing a segment 6 vRNA under the control of RNA polymerase I (Pol I) as previously described were transfected with Lipofectamine 2000 (Invitrogen) using 0.5 µg per plasmid [216]. An empty vector replaced pCAGGS-PA-1918 for the negative control. Cells were incubated for 24 h and total RNA was extracted using TRI reagent (Sigma) following the manufacturer's protocol. RNA levels were analysed by primer extension assay and the reaction products were separated on a 6% denaturing PAGE and cDNA bands were analysed using ImageJ (Fiji) and Prism software (versions 8-10, GraphPad).

2.10 Cryo-electron microscopy

2.10.1 Cryo-EM sample preparation

All cryo-EM grids were prepared on a Vitrobot Mark IV (FEI) and protein samples were diluted to 0.2-0.4 mg/ml before application on Quantifoil Holey Carbon R2/1, 200 mesh copper grids (Electron Microscopy Sciences). For some optimisation experiments, other grid types were used including Quantifoil holey carbon-coated gold grids (1.2 µm holes, 1.3 µm spacing) or Quantifoil holey gold-coated gold grids (1.2 µm holes, 1.3 µm spacing).

Cryo-EM sample preparation of FluPol bound to synthetic CTD peptide:

Cryo-EM sample preparation was performed as previously described with slight modifications [150]. Purified polymerase of the A/Brevig Mission/1/1918 (H1N1) influenza virus with a D108A mutation in the PA subunit was mixed with 5' and 3' vRNA promoters (see 2.1.1) at 1.2× molar excess and incubated on ice for 20 min. Following that, the sample was further purified with SEC into a buffer containing 25 mM HEPES/NaOH pH 7.5 and 500 mM NaCl and concentrated to 1 mg/ml. A capped RNA primer, with six complementary bases at the 3' to the 3' vRNA promoter (see 2.1.1), was added to this sample to a final concentration of 0.2 mM. Additionally, a pS5 Pol II CTD mimic peptide (see 2.1.3), was added to a final concentration of 0.2 mM. Immediately before the preparation of the grids, the sample was diluted 1:3 with a dilution buffer (HEPES/NaOH, pH 7.5, 37.5 mM NaSCN and 0.0075% (v/v) Tween 20). A volume of 3.5 µl of the sample was applied to a freshly glow-discharged grid which was blotted for 3.5 sec and plunge-frozen in liquid ethane. Preparation of the grids was performed at 100% humidity and 20°C.

Preparation of FluPol-GST-CTD$_{26x}$ complexes:

Purified polymerase of the influenza viruses A/Brevig Mission/1/1918 (H1N1), B/Panama/45/1990, and C/Johannesburg/1/1966 was thawed and mixed with

5' and 3' vRNA promoters at 1.2× molar excess (see 2.1.1) and incubated on ice for 10 min. The sample was then purified with SEC into a high salt buffer (as above, FluPol A) or a buffer containing 25 mM HEPES/NaOH pH 7.5 and 150 mM NaCl (for FluPol B and FluPol C). Purified and p-CTD_{26x} was dosed into the concentrated sample at various concentrations ranging between 0.5-3× molar excess. Grids were prepared as described above.

Preparation of FluPol-Pol II complexes:

For the preparation of the native Pol II sample, pS5 Pol II (provided by Dr Marek Šebesta) was prepared as described above and mixed with a 1.5× molar excess of DNA/RNA scaffold consisting of pre-annealed RNA and template DNA before non-template RNA was added (2.1.1). FluPol from IAV, IBV, and ICV prepared as above without the addition of capped RNA or CTD was added to the sample. The final concentration of Pol II was approximately 0.2 mg/ml and the FluPol 0.1 mg/ml. Cryo-EM grids were prepared as above with no additional sample modification.

For crosslinking of the FluPol-Pol II sample, 0.8 µM pS5 Pol II, 7.8 µM DSIF, 1.2× molar excess RNA/DNA scaffold and 0.1% glutaraldehyde were mixed and incubated for 10 min on ice. The reaction was quenched by 1 M Tris-HCl pH 7.5 and dialysed overnight in a Slide-A-Lyzer MINI dialysis device (ThermoFisher Scientific) into a buffer containing 10 mM HEPES-NaOH, 100 mM NaOH, 2 mM DTT, 2 mM $MgCl_2$, and 10 µM $ZnCl_2$. The sample was then added to vRNA-bound 1918 FluPol D108A and prepared as described above.

2.10.2 Cryo-EM data collection and processing

Cryo-EM grids screening and data acquisition were carried out by Dr Jeremy Keown and Dr Loïc Carrique at the Division of Structural Biology (Oxford Particle Imaging Centre), University of Oxford.

Grid screening and small data collections for optimisation purposes were performed on 200 keV Glacios (Thermo Scientific) microscope with a Falcon 3 detector. Data acquisition was conducted on a 300 kV Titan Krios G3i microscope (Thermo Scientific) with a K2 Summit camera (Gatan) and a GIF Quantum energy filter and data collection was automated with SerialEM or EPU (version 2.3-2.5) software. Data processing was carried out using cryoSPARC versions V3-4.2 and figures were generated using ChimeraX [228].

In general, the movies were aligned by motion correction, followed by contrast transfer function (CTF) estimation, particle picking using a blob picker and 2D classification. Selected 2D classes were used for automatic template picking, followed by 2D classification. After the selection of 2D classes, ab initio models were generated and further processed by heterogenous refinement. Further refinement was performed by using non-uniform refinement on each 2D class. The generated cryo-EM maps were then fitted with a structural model in ChimeraX [228].

Data acquisition for the 1918 FluPol-pS5 CTD_{4x} complex was conducted at the Oxford Particle Imaging Centre using a 300 kV Titan Krios with a K2 Summit camera (Gatan) and a GIF Quantum energy filter [150]. Data collection was carried out using SerialEM and figures were generated using ChimeraX [228, 229]. Detailed data processing methodologies and a summarising illustration are presented in chapter 3. For image processing tasks such as patch motion correction, CTF estimation, and initial particle selection using the blob picker cryoSPARC V4-4.2 was utilised [230, 231]. After pre-processing, low-quality micrographs were discarded manually, followed by 2D classification and creation of a high-resolution refinement that would serve for the training of a topaz model and improvement of another iteration of particle picking. The picked particles were pooled and duplicates were discarded. Subsequent iterative 3D classification cycles of the CTD peptide dataset yielded a consensus refinement with 303,608

particles, achieving a map with a resolution of 3.22 Å. Model building was carried out manually in COOT and was based on the PDB 7NHX model, followed by refinement using PHENIX [232]. An additional iteration of 3D classification revealed a reconstruction with an additional density that corresponded to PB2-C which was refined further with deepEMhancer [233]. Within this enhanced map, the pre-established model, which includes the PA, PB1, and the N-terminal segment of PB2 (PB2-N), was aligned. Following this, the PB2-C domains, derived from the PDB 7NHX, were matched with their respective densities. Owing to the low resolution in this area, manual linkage of the domains was performed in COOT, followed by a constrained rigid-body real-space refinement as well as anisotropic displacement parameter refinement [234].

3

Structural and functional characterisation of the Pol II CTD binding to the PA subunit of the 1918 influenza A virus polymerase

A significant portion of this chapter was published in:

Jeremy Keown*, **Alaa Baazaoui***, Marek Šebesta, Richard Štefl, Loic Carrique, Ervin Fodor, Jonathan Grimes (2024) *Structural and functional characterisation of the interaction between the influenza A virus RNA polymerase and the CTD of host RNA Polymerase II. Journal of Virology* doi: 10.1128/jvi.00138-24

* Joint first authorship

Changes in the main text were made to accommodate unpublished data.

The structural characterisation of the described pS5 CTD_{4x}-bound 1918 FluPol A complex presented in sections 3.2.1 and 3.2.2 as well as the interpretation of the structural data was carried out by Dr Jeremy Keown.

3.1 Introduction

During influenza infection, the vRNPs are trafficked to the cell nucleus where the viral polymerase performs both viral transcription and replication. Replication, a two-step process, is primer-independent and requires *de novo* initiation by the viral polymerase to synthesise new vRNA, using positive-sense complementary RNA (cRNA) as an intermediate template [44, 65]. In contrast, viral transcription depends on the interaction of the viral polymerase with the host transcriptional machinery to produce viral positive-sense mRNA. The synthesised viral mRNA contains a 5' terminal N7-methyl guanosine (m7G) cap and a 3' poly(A) tail which makes viral mRNA structurally identical to the host mRNA and allows the virus to hijack the host translation machinery for viral protein synthesis in the cytoplasm [41, 65]. In a process called "cap-snatching", the viral polymerase steals 10–15 nucleotides long capped RNA fragments from nascent capped host RNA, generated by the host RNA polymerase II (Pol II), to prime its own transcription using the PB2 cap-binding and PA endonuclease domains [40, 41]. Cap-snatching occurs during the early stages of Pol II-mediated transcription, which enables the viral polymerase to secure access to host RNA caps before being bound by high-affinity host cap-binding proteins [209].

To access the cap of nascent host RNA, the viral polymerase physically associates with host Pol II [125, 200]. Functional and structural studies confirmed that the viral polymerase selectively recognises the serine 5 phosphorylated (pS5) CTD heptad repeat sequence YSPTpSPS, the signature for initiating Pol II involved in capping [40, 41]. In structural studies, two distinct binding sites of the CTD to the viral polymerase were observed using polymerase from influenza virus types A, B, and C and synthetic four-repeat pS5 CTD peptides [26, 96, 97]. Importantly, until recently, the only CTD-bound FluPol A structure available was FluPol originating from bats, which, compared to human FluPol A, is expected to behave differently due to reported differences in *in vitro* experiments and its ability to replicate in different mammalian cell lines [235, 236]. Structures of the

CTD-bound polymerase of different influenza types revealed shared and distinct binding features between each type, while binding of the Pol II CTD beyond the identified binding sites remains largely unknown (Figure 1.6). Considering that mammalian Pol II CTD contains 52 heptad repeats of the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$, which far exceeds the length of CTD mimic peptides, it appears likely that further interactions between the viral polymerase and Pol II CTD take place beyond the identified binding sites [40, 41]. In addition, it has been reported that the interaction of the viral polymerase with Pol II involves not only the CTD of the large subunit, RPB1, but also other subunits of the 12-subunit holoenzyme such as RPB4 [26, 237].

This chapter presents the structures of the CTD-bound polymerase of the 1918 pandemic influenza A virus. In these structures, a continuous stretch of the CTD binding site is observed, encompassing 25 of the 28 amino acid residues of the four heptad repeats (repeats a–d) on the C-terminal domain of the PA subunit (PA-C). Additionally, a minor population of particles with a unique arrangement of the PB2 C-terminal (PB2-C) region, likely representing a transcriptase intermediate was found. Based on these structures, residues in the CTD-binding site that are conserved across influenza A virus strains were mutated, to determine their effect on viral polymerase activity using a range of functional assays, including a cell-based vRNP reconstitution primer extension assay, luciferase-reporter assays, and biochemical assays. Findings in this chapter demonstrate that mutagenesis of conserved residues in the CTD binding site results in transcriptional and replicational defects, indicating a potential overlap of binding sites for host factors that facilitate transcription and replication, respectively.

3.2 Results

3.2.1 Structure of the 1918 pandemic FluPol A bound to Pol II CTD mimic peptide

To characterise the binding of the CTD of host Pol II to the 1918 pandemic influenza A virus polymerase, recombinant influenza virus polymerase expressed in insect cells and a four-repeat synthetic pS5 Pol II CTD mimic peptide were used. Cryo-EM analysis of the complex generated a high-quality map to a global resolution of 3.22 Å (Figure 3.1 A, Figure 3.2) comprising the polymerase core (PA-C, PB1, PB2-N domains), the PA endonuclease domain, 15 bases of the 5′ vRNA promoter, 8 bases from the 5′ end of the 3′ vRNA promoter, and continuous density for 25 amino acids of the CTD peptide (Figure 3.1 B). Even though capped RNA was added to the complex, this was not observed either in the influenza polymerase active site or in the PB2 cap-binding domain. The PB2-C domains were not observed in the consensus refinement.

The global polymerase conformation is that of a transcriptase as characterised by the position of the endonuclease domain packing against the PB1 C-terminal helices [77, 104, 122]. The CTD occupies an elongated binding site along the PA-C domain, where we observe 25 of the 28 residues from the peptide (Figure 3.1 C). The phosphorylated peptide covers a total surface area of 1.160 $Å^2$ with the interaction mediated predominantly by the first (designated repeat a), third (designated repeat c), and the N-terminal half of the fourth heptad repeat (designated repeat d). The phosphorylated serine residues pS_{5a} and pS_{5c} of the peptide are highly ordered and are coordinated into binding sites formed by K635/R638 and K289/R454, respectively. E416 and E457 aid in the correct positioning of R454 to coordinate pS_{5c}. Other amino acid residues of the polymerase interact with unmodified residues from the CTD peptide. Specifically, residues Y445, E449, and F612 interact with Y_{1a} and D419 may contact Y_{1b}. At the C-terminal end of the resolved part of the peptide L290 and Y313 contact Y1d, L527 may interact with P_{6c}, and R551/R559 contact the peptide as it leaves

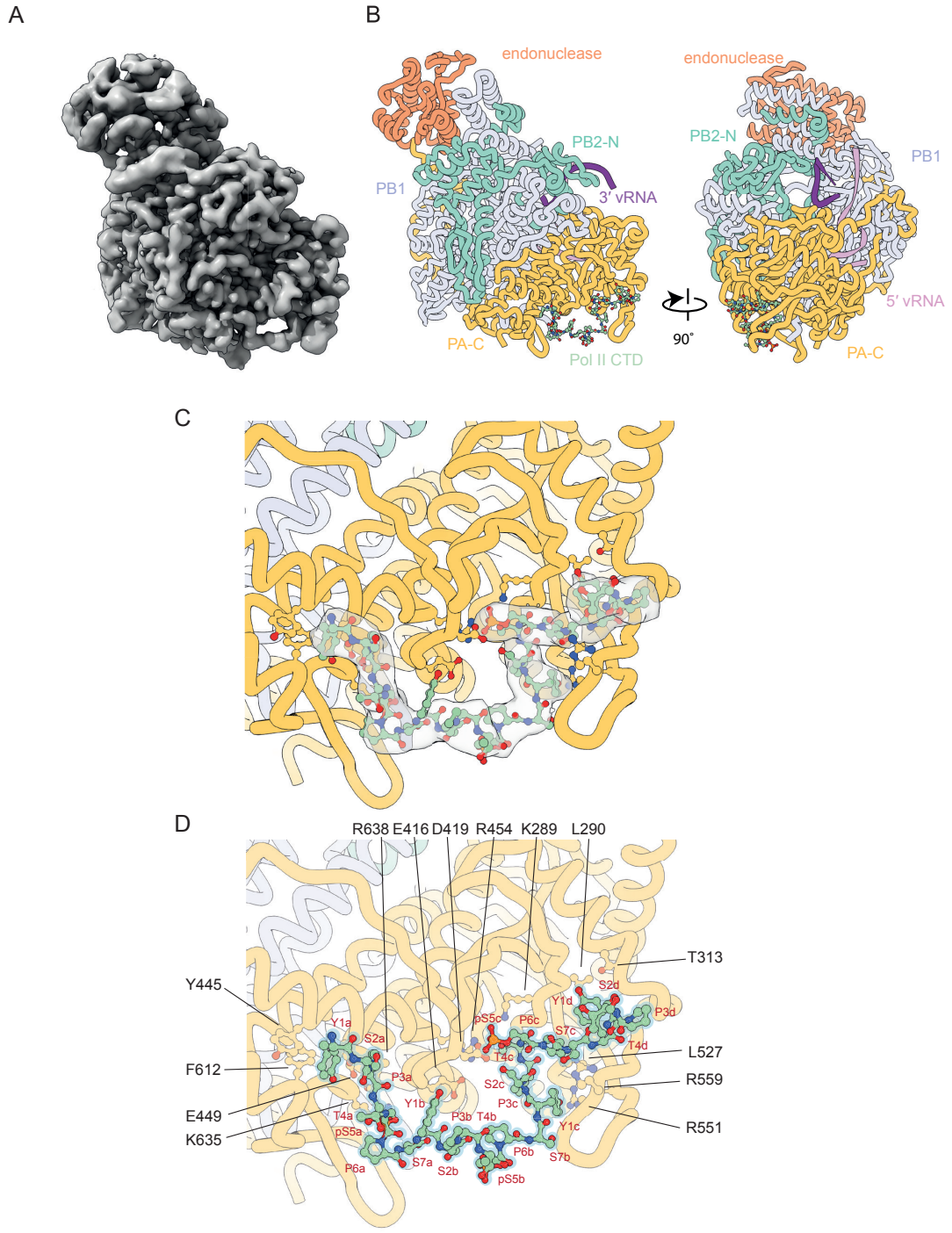

Figure 3.1: Structure of the 1918 pandemic H1N1 influenza A virus polymerase bound to a pS5 Pol II CTD mimic peptide. Structure of the 1918 pandemic H1N1 influenza A virus polymerase bound to a pS5 Pol II CTD mimic peptide. **(A)** DeepEMhancer modified cryo-EM map. **(B)** Structural model in two orientations showing the position of the CTD peptide bound at the PA-C domain. **(C-D)** Close-up views detailing the interaction of the CTD peptide with the PA-C domain. **(C)** Fit of the peptide-bound model. **(D)** Amino acids involved in mediating interactions with the CTD. Figure prepared by Dr Jeremy Keown and adapted from Keown, Baazaoui et al. (2024) [112].

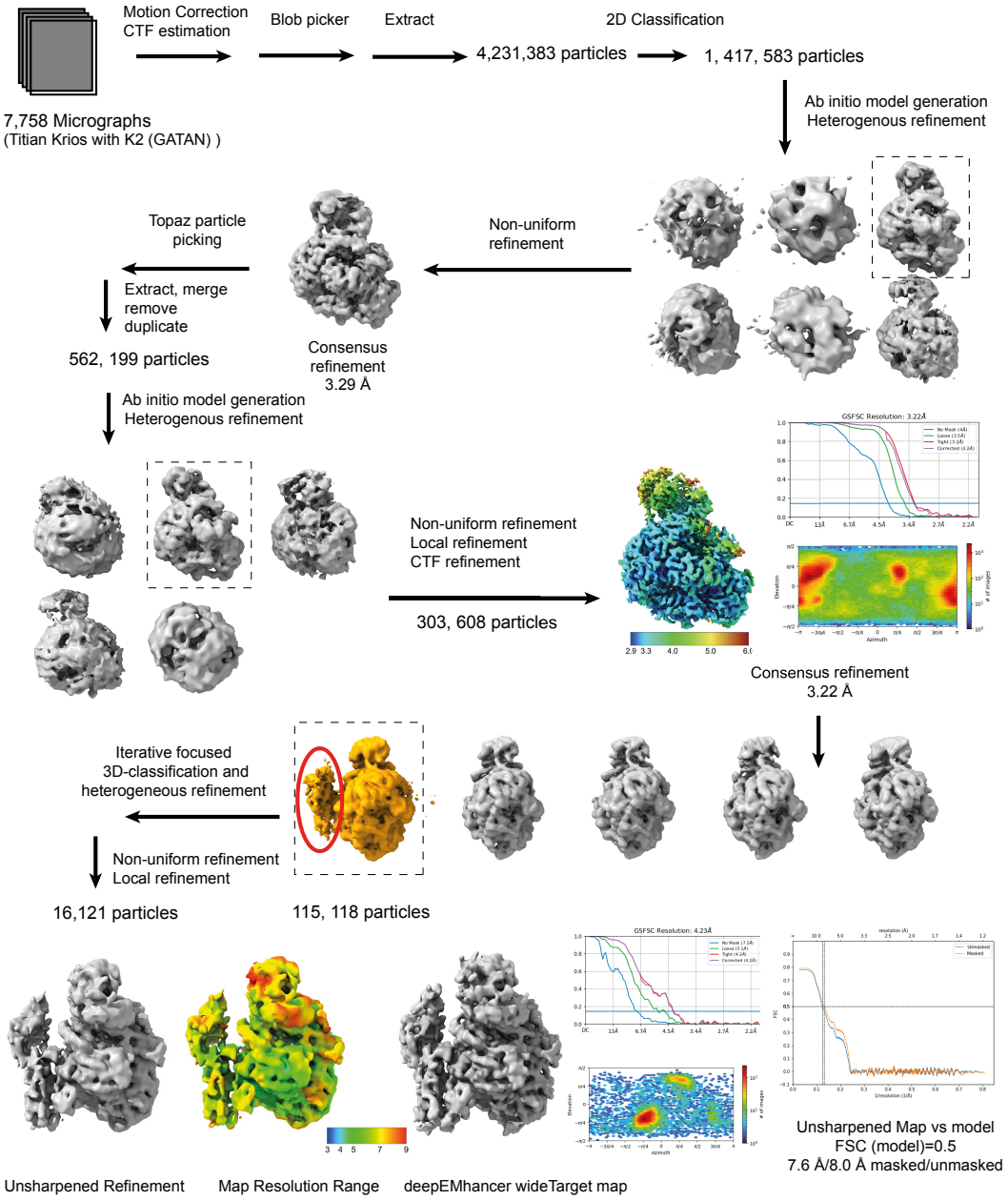

Figure 3.2: Cryo-EM processing of samples containing 1918 pandemic H1N1 influenza A virus polymerase bound to a pS5 Pol II CTD mimic peptide. Processing workflow depicting various steps that resulted in a map of 1918 influenza virus polymerase bound to pS5 CTD. Figure prepared by Dr Jeremy Keown and adapted from Keown, Baazaoui et al. (2024) [112].

the surface of the PA-C over the top of the PA 550-loop. The phosphorylated pS_{5b} was partially ordered, showing a clear main chain, and does not appear to contact the polymerase core (Figure 3.1 D). All residues interacting with the peptide are from the PA subunit.

During the review process of the publication associated with this project, a new cryo-EM structure of FluPol A in complex with a synthetic Pol II CTD_{6x} peptide has been published (PDB ID 8R3L) [74]. Specifically, the viral polymerase was derived from the influenza A/Zhejiang/DTID-ZJU01/2013 (H7N9) virus (H7N9 FluPol A), a human isolate of an avian strain that possesses two adaptation mutations in PB2 [74]. This study used a 6-repeat pS5 CTD peptide (pS5 CTD_{6x}) with designated repeats a-f, of which amino acids S_{7a} to P_{3e} were resolved in the structure [74]. Comparison of our CTD-bound 1918 FluPol A structure with CTD-bound H7N9 FluPol A reveals very similar peptide paths in the PA-C (Figure 3.3). Both structures reveal a single continuous CTD-binding site on the PA-C with 25 resolved amino acids, adopting a very similar peptide path (Figure 3.3 A, B). We observed in our structure CTD peptide residues Y_{1a} to T_{4d} while in the H7N9 FluPol A structure, the resolved peptide starts with S_{7a} and ends with P_{3e} [74]. The critical pS5a and pS5c in our structure overlap with the corresponding pS5b and pS5d in H7N9 FluPol A (Figure 3.3 C, D). These pS5 residues are stabilised in highly basic grooves, while pS_{5b} (pS_{5c} in H7N9 FluPol A) is exposed toward the solvent (Figure 3.3 C, D). PS_{5b}, together with its surrounding amino acids in repeat b (repeat c in H7N9 FluPol A), adopts a slightly different arrangement. For instance, Y_{1b} (Y_{1c} in H7N9 FluPol A), T_{4b} (T_{4c} in H7N9 FluPol A) are shifted (Figure 3.3 C, D). At the end of the peptide path, we also observe that P_{3d} and P_{3e} in H7N9 FluPol A exhibit different orientations (Figure 3.3 C, D).

The observed CTD binding site in our structure and H7N9 FluPol A is similar to that observed previously for other IAV polymerases as well as IBV and ICV

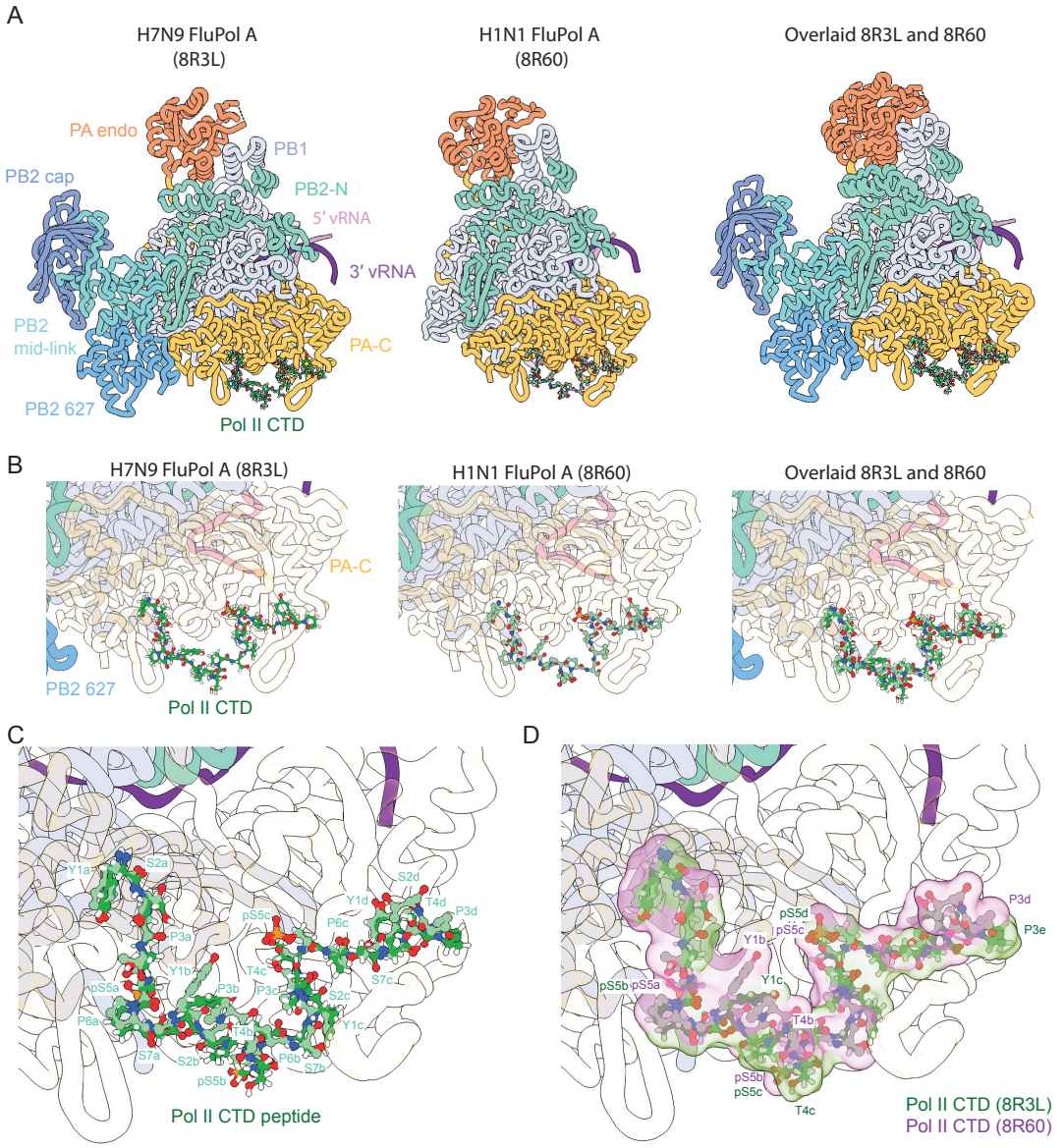

Figure 3.3: Comparison of 1918 FluPol A and H7N9 FluPol structures bound to pS5 Pol II CTD mimic peptide. (A) Structures of FluPol A bound to pS5 Pol II CTD mimic peptide from A/Zhejiang/DTID-ZJU01/2013 (H7N9) (PDB ID 8R3L, left panel) and 1918 FluPol A (PDB ID 8R60, middle panel). Superposition of 8R3L and 8R60, aligned on PB1 and PA-C (right panel). **(B)** Close-up view on the CTD-binding site found in the FluPol A structures 8R3L (left panel) and 8R60 (middle panel) and their superposition (right panel). **(C)** Close-up view on the overlaid CTD peptides path from 1918 FluPol A (light green) and H7N9 FluPol A (dark green). The resolved CTD peptide amino acids are labelled according to the 1918 FluPol A structure (light green). **(D)** Close-up view on the overlaid CTD binding site in the PA-C that includes coloured surface map representations of the peptide paths for 1918 FluPol A (purple) and H7N9 FluPol A (green).

polymerases, underpinning the conserved role of PA-C in binding Pol II CTD (Figure 1.2.2). However, there are clear distinctions in the mode of CTD binding involving different regions of PA-C for the polymerases of the three different influenza types. Additionally, in contrast to FluPol A, in FluPol B, the CTD binding site extends to the PB2 627-NLS domain, while in FluPol C, the binding site also involves amino acid residues from the PB1 subunit. Together, these data show that although polymerases of all three influenza types use PA-C as the main interface for Pol II CTD binding, the mode of binding has diverged between influenza virus types involving distinct regions [26, 40, 41].

3.2.2 Arrangement of PB2-C in the Pol II CTD-bound influenza virus polymerase

While most particles in the used cryo-EM data set lacked density for the flexible PB2-C domains, including the cap-binding domain, mid-link, 627, and nuclear localisation signal (NLS) domains, approximately 5% of the particles from the consensus refinement contained additional electron density, suggesting that in these particles the PB2-C domains had become ordered (Figure 3.4, Figure 3.2) [117]. By aligning the core of previously determined polymerase structures, it was possible to accurately position the cap-binding domain and mid-link domain in the density (Figure 3.4 B). The position of these two domains was essentially identical to that of the recently determined transcription intermediate structure (PDB ID 7NHX) [150]. In this model, the mid-link domain has remained in close contact with the polymerase core (in a position common to transcriptase conformations), while the cap-binding domain has retracted away from the core and primer entry channel. A putative role of this conformation was recently proposed, as an intermediate state between transcription and replication [238]. To aid the unambiguous orienting of the 627 domain and/or the NLS, deepEMhancer was employed to modify the map. From this modified map, the 627 domain could be positioned such that the connectivity to the mid-link was maintained and with a good fit to the modified density (Figure 3.5).

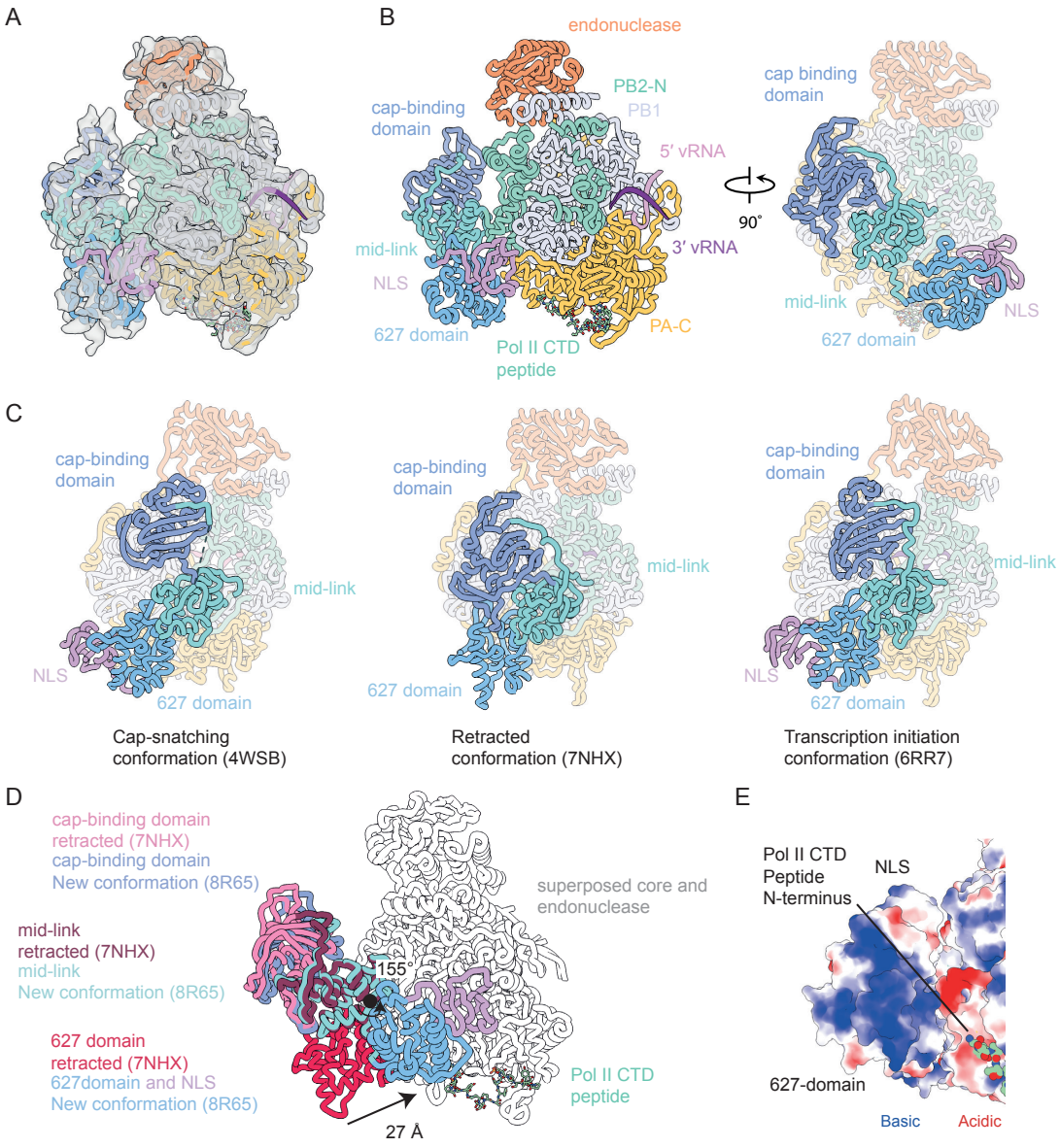

Figure 3.4: Structure of the 1918 pandemic H1N1 influenza A virus polymerase bound to a pS5 Pol II CTD mimic peptide with resolved PB2-C domains. **(A)** DeepEMhancer modified cryo-EM map. **(B)** Model in two orientations showing the domain structure with the arrangement of the PB2-C domains highlighted on the right. **(C)** Previously determined structures of FluPol with different arrangements of the PB2-C domains highlighted. **(D)** Structures of the 1918 CTD-bound polymerase and the polymerase in the retracted conformation superposed on the polymerase core and endonuclease domain (shown in grey). The PB2-C domains are highlighted in different colours and the 27 Å movement of the PB2 627 domain is indicated. **(E)** Electrostatic analysis of the PB2 627 domain that is oriented towards the N-terminus of the CTD peptide shows a highly basic surface. Figure prepared by Dr Jeremy Keown and adapted from Keown, Baazaoui et al. (2024) [112].

The 627 domain and NLS in this model adopt a position that is distinct from their location in either a replicating polymerase, encapsidating polymerase, or previously observed transcription initiation or cap-snatching structures (Figure 3.4 C) [239]. In this model, the 627 domain and NLS have rotated 155° and moved 27 Å from the transcription initiation position (Figure 3.4 D). The movement changes the interaction with the polymerase from contacting the PA and PB1 subunits through a three-stranded beta-sheet formed by residues 640-676 such that it now contacts the PA-C and PB2-N2 domains through residues 643-654. This interface is smaller than observed in the transcription initiation conformation, resulting in poor resolution of this domain. A small interface was formed between the NLS and residues from the PB1 (583-587), PB2 (125-127), and PA (430-432) subunits.

Although the protein domains could be precisely positioned, due to the low resolution, it was not possible to accurately describe residue contacts in these regions. This relocation of the 627 domain and NLS brings them proximal to the N-terminus of the Pol II CTD peptide but no additional peptide density for the Pol II CTD was observed in this reconstruction, which could indicate a direct interaction between the CTD and PB2-C. However, given the low-resolution maps, the possibility of a direct contact cannot be excluded. In fact, electrostatic analysis of the surface of the 627 domain that is oriented toward the N-terminus of the CTD peptide shows a highly basic surface (Figure 3.4 E). This surface would likely be able to bind the negatively charged phosphorylated CTD peptide. Recent structural evidence of Pol II CTD binding to the influenza B virus polymerase demonstrates a large additional CTD-binding site across the surface of the 627 domain [26, 96]. However, the corresponding residues in influenza A virus polymerases are not conserved. In line with this, deletion of the 627 domain from FluPol B dramatically reduced binding to Pol II CTD, while a similar deletion did not affect the ability of the FluPol A to bind CTD [26]. This is suggestive that the

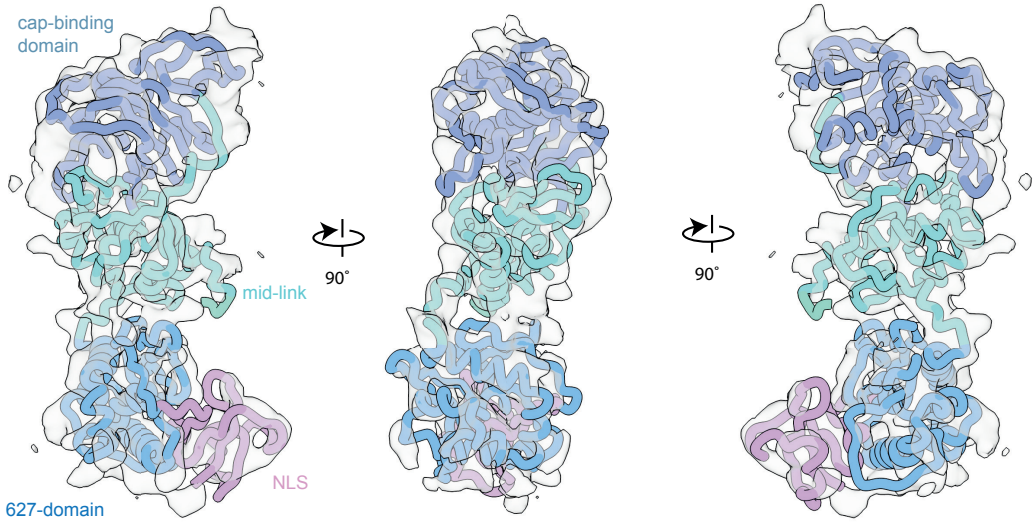

Figure 3.5: DeepEMhancer map showing the fit of the PB2-C domains from the new PB2 conformation. The model shown is based on the CTD-bound 1918 pandemic influenza A virus polymerase structure (PDB 8R65). Figure prepared by Dr Jeremy Keown and adapted from Keown, Baazaoui et al. (2024) [112].

loss of binding from the deletion of the 627 domain is compensated by the much larger binding surface of the CTD peptide on the influenza A virus PA-C domain.

3.2.3 Effect of mutations in the PA-C Pol II CTD binding site on polymerase function

Mutations in the PA-C Pol II CTD binding site cause replication and transcription defects

The structural insights of the CTD-bound 1918 FluPol A structure informed the functional analysis of residues mediating CTD binding. To address the role of the identified Pol II CTD-binding site in PA-C in polymerase function, a vRNP reconstitution assay was used, followed by the analysis of RNA by primer extension. To assemble vRNPs, we co-expressed the three polymerase subunits, nucleoprotein, and either segment 6 (encoding NA vRNA) or segment 8 (encoding NS2 vRNA) in HEK 293T cells. This assembly forms the minimal viral complex required for transcription (mRNA synthesis) and genome replication (cRNA and vRNA synthesis). Based on our structural model and previously determined FluPol structures from IAV strains, 14 conserved residues in the PA-C domain

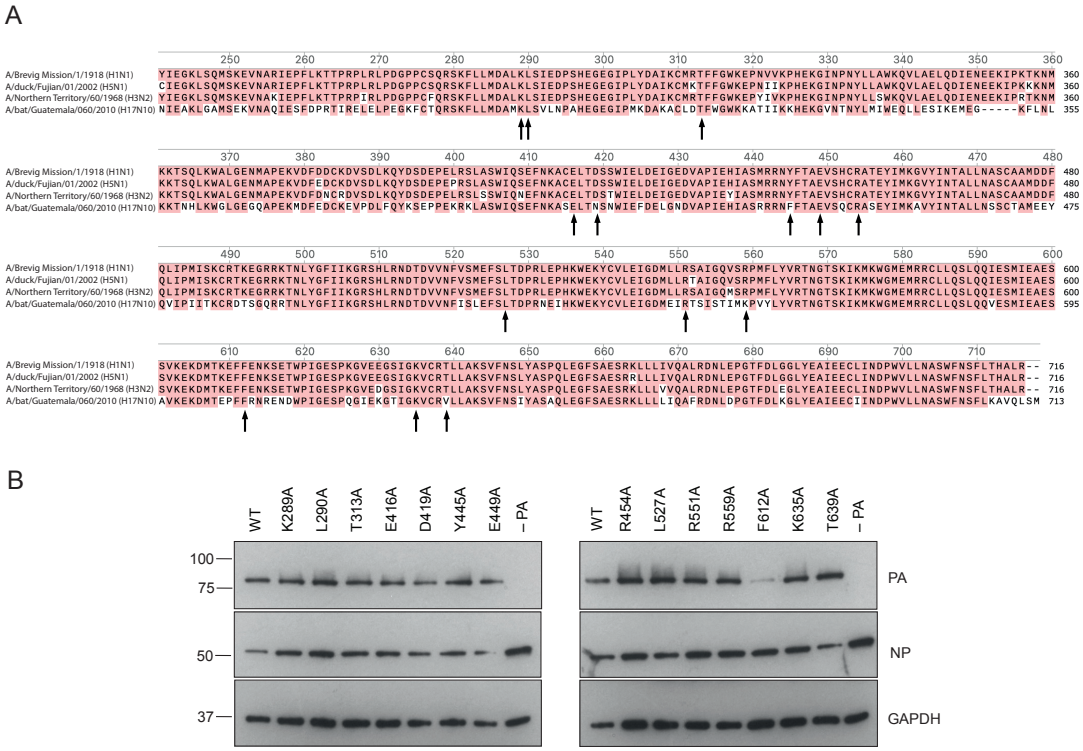

Figure 3.6: Identification of conserved amino acid residues in the CTD-binding site of the influenza A virus PA subunit and expression of PA mutants. (A) Multiple sequence alignments of PA-C from A/Brevig Mission/1/1918 (H1N1) (UniProt Q3HM39), A/duck/Fujian/01/2002 (H5N1) (D3YNB8), A/Northern Territory/60/1968 (H3N2) (P03434), and A/bat/Guatemala/060/2010 (H17N10) (H6QM92) using Clustal-Omega in SnapGene. Identical residues are coloured in red. Based on our structural model of pS5 CTD-bound 1918 pandemic H1N1 influenza A virus polymerase, black arrows depict conserved residues hypothesised to affect CTD binding. **(B)** Expression of the PA (WT and mutants) and NP of the 1918 pandemic H1N1 influenza A virus polymerase. HEK293T cells were transfected to express PB1, PB2 and PA (WT or mutant) polymerase subunits, NP, and segment 6 vRNA and cell lysates were analysed by SDS-PAGE and Western blotting with antibodies against PA, NP and GAPDH (loading control). A representative blot with size markers (kDa) is shown. Figure adapted from Keown, Baazaoui et al. (2024) [112].

were mutated that were hypothesised to influence CTD binding (Figure 3.6 A) [26, 96, 122, 150]. With the exception of residue F612, all mutants were expressed at similar levels to wild-type FluPol following transfection (Figure 3.6 B).

Levels of vRNA were minimal when the PA subunit was omitted from the transfection (Figure 3.7, Figure 3.8). On the other hand, the expression of a complete wild-type FluPol resulted in the detection of mRNA and cRNA, along

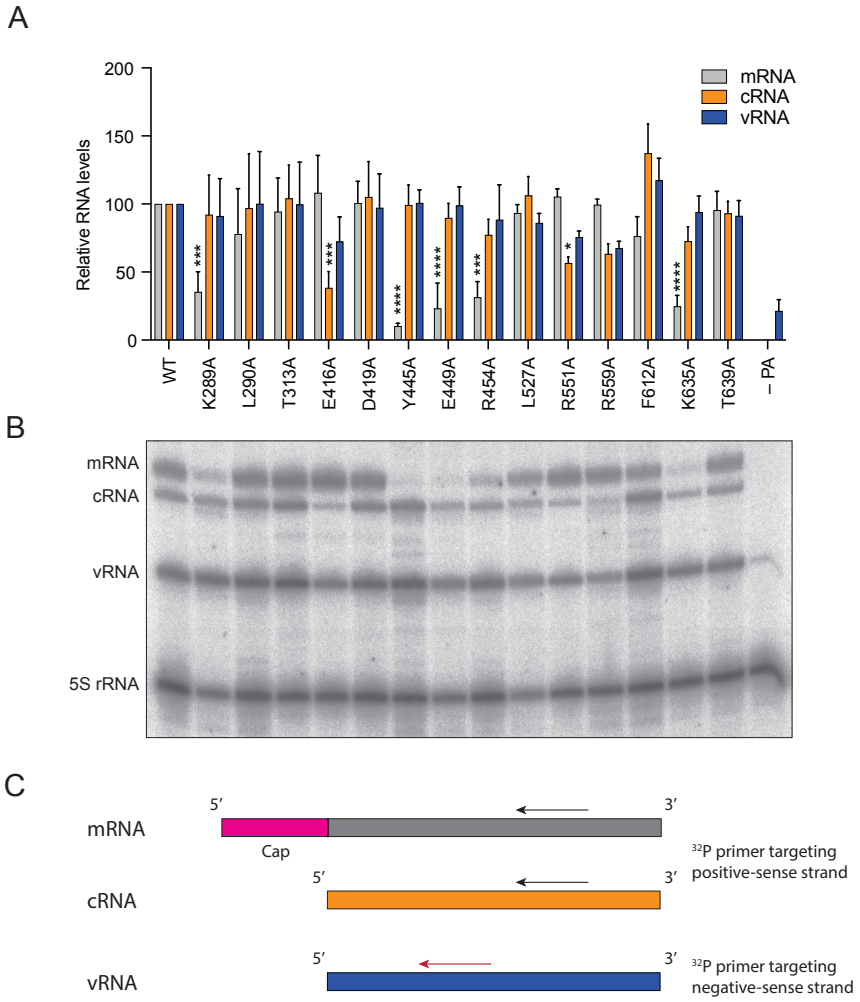

Figure 3.7: The effect of mutations in the CTD binding site of the 1918 influenza A virus polymerase on viral RNA synthesis. HEK 293T cells were transfected to express 1918 FluPol A subunits PA (WT and mutant), PB1, and PB2, as well as segment 6 vRNA (expressing NA). The plasmid expressing PA was replaced with an empty vector for the negative control. Total RNA was isolated 24 hours post-transfection and was analysed by primer extension assay. **(A)** Quantification of mRNA, cRNA, and vRNA levels from n=3 independent vRNP reconstitution assays. **(B)** A representative PAGE of a primer extension analysis, with 5S rRNA serving as a loading control. The signal intensity of the mean is shown relative to the signal intensity from WT 1918 viral polymerase. Radiolabelled products were visualised by exposure to phosphorimaging screens (FujiFilm), followed by scanning on a FujiFilm FLA-5000 scanner. Quantification of the bands was carried out with ImageJ. **(C)** Schematic illustrating the principle of RNA species detection using primer extension analysis. Data are presented as mean ± s.e.m. An ordinary two-way ANOVA was employed to determine significant differences with p set at *p $<$ 0.05; ***p $<$ 0.001; and ****p $<$ 0.0001 (GraphPad Prism). Figure adapted from Keown, Baazaoui et al. (2024) [112].

with a significant increase in vRNA (Figure 3.7, Figure 3.8). This indicates that the vRNA, expressed from the transfected plasmid, is transcribed and replicated by FluPol. When comparing the RNA outputs of the wild-type with that of the mutant polymerases in the NA-based primer extension assay, mutations K289A, Y445A, E449A, R454A, and K635A affected transcription with relatively little if any effect on replication (Figure 3.7). The phenotype of these mutant polymerases suggests that CTD binding is critical for viral transcription, as it enables FluPol access to capped RNA fragments generated from nascent Pol II transcripts. Other mutations such as L290A, T313A, D419A, L527A, R559A, F612A, and T639A had no or only a small effect on transcription, suggesting that these amino acid residues are not involved in essential interactions with the CTD (Figure 3.7 A). Surprisingly, two mutations, E416A and R551A, specifically reduced cRNA synthesis, suggesting that amino acid residues around the Pol II CTD binding site also participate in the replication of the viral RNA genome (Figure 3.7).

In the NS segment-based primer extension assay, residues K289, R454, and K635 exhibited both transcription and replication defects, while residues L290 and Y445 affected only transcription and residues R551 and R559 specifically replication (Figure 3.8 A). Residue L290 affected transcription in the NS-based vRNP reconstitution assay and also led to reduced FluPol activity in a minigenome assay (Figure 3.8, Figure 3.10). Slight observational discrepancies in the mutation phenotypes are likely caused by the lower signal-to-noise ratio in the NS-based primer extension assay compared to the NA-based primer extension assay. The mostly overlapping results for the tested mutants confirm a pivotal role for the PA-C domain as an important site for transcription and replication.

Mutations in the PA-C Pol II CTD binding site interfere with ANP32 binding and overall polymerase function

A split-luciferase assay was performed to investigate the impact of mutations in the PA-C on ANP32B (huB) binding. HEK293T cells with a double knock-out in human ANP32A and ANP32B were transfected to express a chimeric FluPol with

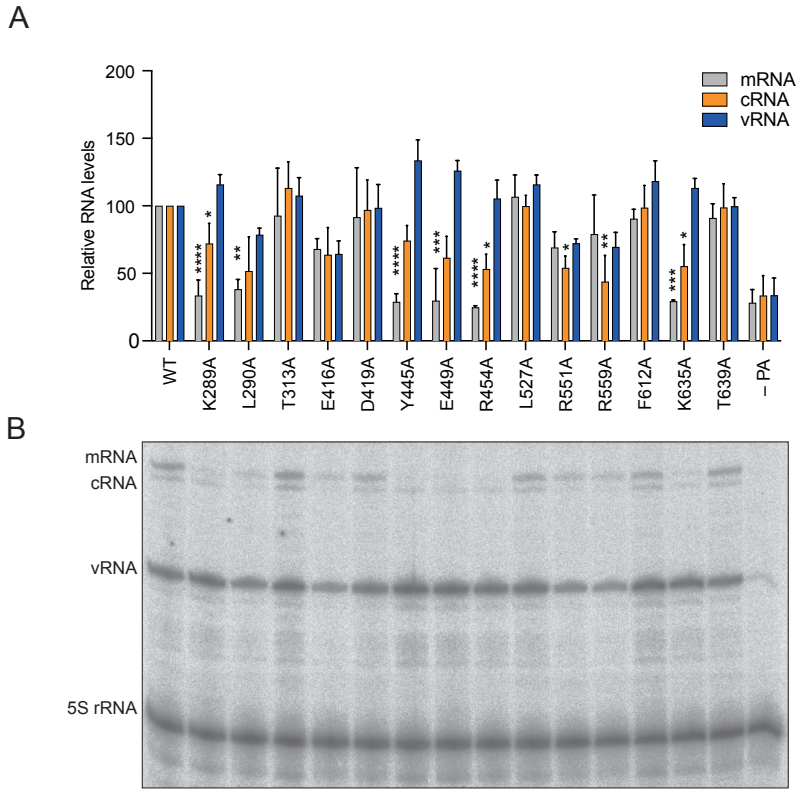

Figure 3.8: The effect of mutations in the CTD binding site of the 1918 pandemic H1N1 influenza A virus polymerase on viral RNA synthesis. HEK 293T cells were transfected to express 1918 FluPol A subunits PA (WT and mutant), PB1, and PB2, as well as the segment expressing NS protein. For the negative control, the plasmid expressing PA was replaced with an empty vector. Total RNA was isolated 24 hours post-transfection and was analysed by primer extension assay. **(A)** Quantification of mRNA, cRNA, and vRNA levels from three independent vRNP reconstitution assays. **(B)** A representative PAGE of a primer extension analysis, with 5S rRNA serving as a loading control. Radiolabelled products were visualised by exposure to phosphorimaging screens (FujiFilm), followed by scanning on a FujiFilm FLA-5000 scanner. Quantification of the bands was carried out with ImageJ. The signal intensity of the mean is shown relative to the signal intensity from WT 1918 viral polymerase. Data are presented as mean ± s.e.m. An ordinary two-way ANOVA was employed to determine significant differences with p set at *$p < 0.05$; ***$p < 0.001$; and ****$p < 0.0001$ (GraphPad Prism).

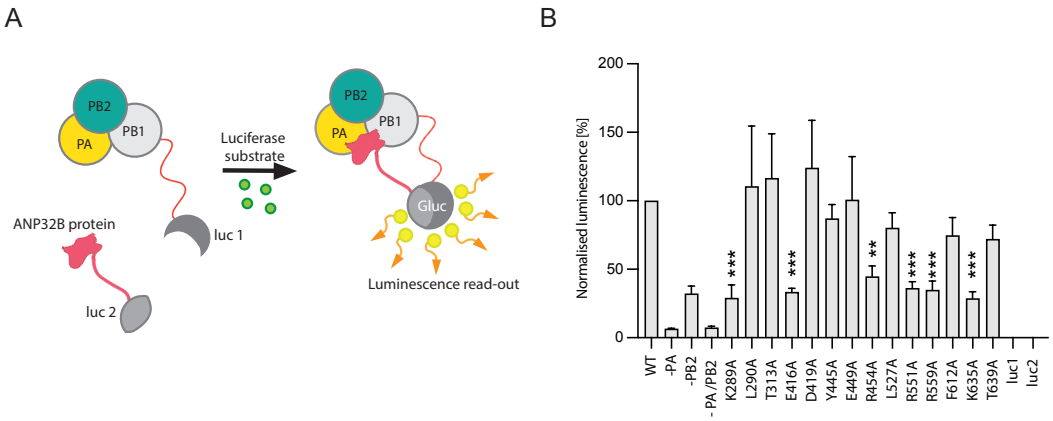

Figure 3.9: The effect of mutations in the CTD binding site of the PA subunit on ANP32 binding by a chimeric (1918/WSN) influenza A virus polymerase. **(A)** Schematic of split luciferase-tagged influenza virus polymerase and the split luciferase assay principle. **(B)** *Gaussia princeps* split luciferase assay to measure the interaction between mutant influenza virus polymerase and human ANP32B (huB). Plasmids were transfected in dKO huB HEK293T cells to express C-terminally tagged split luciferase (luc1 and luc2) on WSN PB1-luc1 (H3N2) and huB-luc2 together with 1918 PA (H1N1), and 1918 PB2 (H1N1). As controls, plasmids expressing PA and PB1 were replaced with empty plasmids, and polymerase subunits co-transfected with either empty luc1 or luc2 plasmids alone. Data are from three independent transfections and two technical replicates. The mean signal intensity is shown relative to the signal intensity from WT PA 1918-WSN chimeric viral polymerase. Data are mean ± s.e.m. Ordinary one-way ANOVA was employed to determine significant differences with *$p < 0.05$ and ***$p < 0.001$.

PB1-luc2, PA, PB2, and huB-luc1. The -PB2 negative control showed relatively high background levels of approximately 30% compared to the -PA and -PA/PB2 negative controls potentially caused by a more stable PA-PB1-luc2 dimer that, together with huB-luc1, can cause stronger background luminescence (Figure 3.9). Mutations in residues K289, E416, R454, R551, R559, and K635 resulted in an average reduction of luminescence levels by approximately 30-50%, indicating perturbed ANP32B binding. This observation is supported by the reduced cRNA levels in the vRNP reconstitution assays (Figure 3.7, Figure 3.8). Overall, mutations in the PA-C domain in chimeric 1918-WSN FluPol can weaken the interaction with ANP32B, potentially resulting in replication defects and reduced cRNA levels.

A luciferase reporter minigenome assay was carried out to assess the impact of the mutations on overall polymerase function (Figure 3.10). All mutants, except for D419A, showed reduced or entirely abolished activity compared to the WT (Figure 3.10 B). Mutations at residues T313, L527, and T639 led to a reduction in activity by up to 70% and at K289 and L290 by up to 90% and at residues E416A, Y445A, R551A, R559A, and K635A viral polymerase activity was entirely abolished (Figure 3.10 B). Interestingly, the minigenome assay revealed a discrepancy to the primer extension assay, exhibiting normal mRNA levels for mutants E416A, R551A, and R559A but complete abolition of viral activity in the minigenome assay (Figure 3.7, Figure 3.10). This discrepancy could be caused by multiple factors, including impaired polyadenylation of the produced viral mRNA resulting from the introduced mutations, poor transport or translation of the produced viral mRNA, or an enzymatically inactive mutant polymerase or luciferase [240].

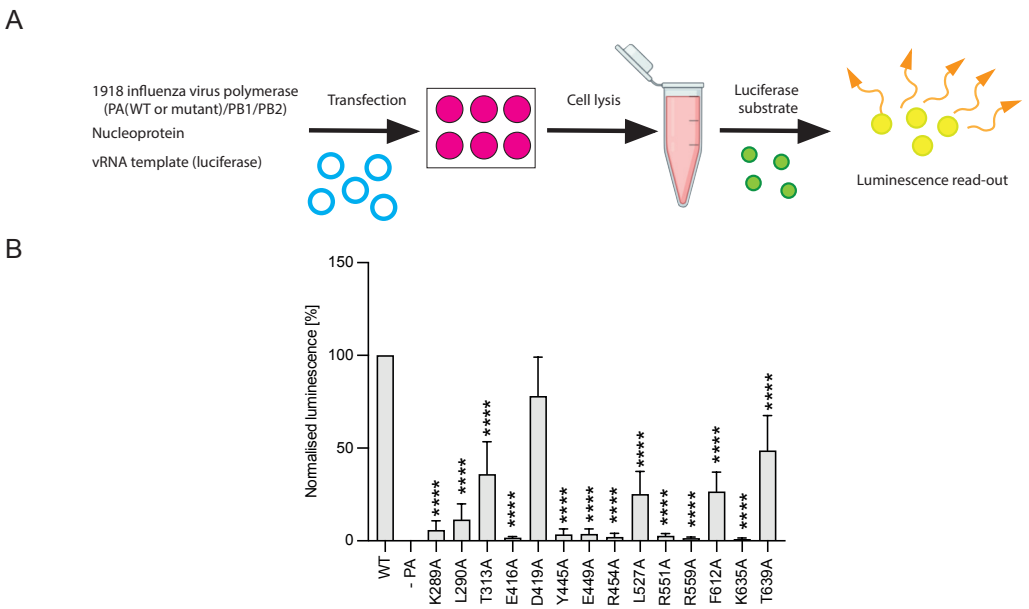

Figure 3.10: The effect of mutations in the CTD binding site of the PA subunit on the H1N1 (1918) pandemic influenza A virus polymerase activity. **(A)** Schematic diagram of the luciferase-based minigenome assay. **(B)** Luciferase reporter minigenome assay with HEK 293T cells transfected to express 1918 FluPol A subunits PA (WT and mutant), PB1, and PB2, as well as luciferase reporter. Data are from three independent transfections and n=3 technical replicates. The mean signal intensity is shown relative to the signal intensity from WT 1918 viral polymerase. Data are from three independent transfections and three technical replicates and are mean ± s.e.m. Ordinary one-way ANOVA was employed to determine significant differences with *$p < 0.05$; ***$p < 0.001$; and ****$p < 0.0001$.

3.3 Discussion

In this study, the solution state structure of the 1918 pandemic H1N1 influenza virus polymerase bound to pS5 Pol II CTD using cryo-EM was solved. The structure revealed a continuous density for the CTD peptide with 25 out of the 28 amino acid residues fully resolved (Figure 3.1). The CTD peptide was found to bind exclusively at the PA-C with two of the phosphorylated serine residues playing prominent roles in guiding the interaction by being accommodated in highly basic grooves. The interaction interface partly overlaps with that previously observed in a study using polymerase of a bat H17N10 IAV [96]. Unlike the study that identified two distinct binding sites for the CTD in a bat influenza model, this work presents a continuous density with the CTD connecting the two binding sites in a medically significant pandemic virus model [96, 241].

Recently, while this work was under review for publication, a new structure of FluPol A bound to a Pol II CTD peptide was reported [74]. The CTD-bound structure of FluPol A, derived from influenza A/Zhejiang/DTID-ZJU01/2013 (H7N9) virus, revealed a very similar CTD-binding site on PA-C as reported here (Figure 3.1). Both CTD-bound structures of 1918 FluPol A and H7N9 FluPol A revealed almost overlapping peptide paths in the PA-C domain (Figure 3.3). Only minor differences were observed in amino acids either exposed toward the solvent or at the N- and C-termini (Figure 3.3 C, D). The critical pS5 residues embedded in the structure are in the same highly basic grooves (Figure 3.3 C, D). Overall, the new CTD-bound FluPol A structures from 1918 FluPol A and H7N9 FluPol A show great similarities in their CTD-binding sites and global arrangement. The common and distinct features of CTD binding across different influenza virus types highlight the divergent evolution of CTD binding modes and the need to investigate the FluPol-Pol II interaction in the context of the full-length Pol II holoenzyme [26, 74, 97, 112].

In our presented structure, the PB2-C domains remained unstructured in most particles, suggesting that binding of four-repeat CTD peptide in solution does not stabilise the cap-snatching or transcription initiation complexes and that a longer CTD peptide and/or other interacting partners might be required (Figure 3.1). However, in a small population of particles, the PB2 cap-binding, mid-link, 627, and NLS domains were structured, revealing an arrangement not previously observed (Figure 3.4, Figure 3.5). In this novel arrangement, the 627 domain is positioned close to the N-terminus of the CTD peptide, providing a potential basic surface for additional interactions with the acidic CTD (Figure 3.4). Binding of the CTD across PA-C and the adjacent 627 domain could stabilise the 627 domain and, consequently, enable the PB2 cap-binding and PA endonuclease domains to assume a cap-snatching competent conformation. Pol II CTD peptides composed of four heptad repeats used in this study might be too short to cover the entire binding surface that might extend from PA-C across the PB2 627-NLS domains. Further studies using Pol II CTD peptides composed of more than four heptad repeats could reveal further insights into the interaction of the influenza virus polymerase with Pol II CTD. In addition, interactions with other Pol II subunits beyond the RBP1 CTD should be structurally investigated as functional assays indicate an interaction of FluPol with other subunits including RPB4 and RBP7 [26, 237]. Interestingly, current data suggest that while the PA-C domain targets the CTD of RPB1, RPB4 interacts with PB2 to position FluPol close the RNA exit tunnel of Pol II. This would place the cap-binding domain near the nascent transcript, facilitating cap-snatching [237].

The PA subunit plays an essential role in polymerase activity and is involved in various other functions, including protein stability and nuclear localisation, endonuclease function, cap binding, promoter binding, and viral packaging [215, 218, 242, 243]. Mutations within the N-terminal region of PA (PA-N) have been shown to significantly impact polymerase stability and function. For instance, mutations

between residues L109 and F117 result in PA degradation via the proteasome-ubiquitin pathway, ultimately affecting polymerase activity [244]. Additionally, point mutations K102A, D108A, and K134A have been observed to severely inhibit polymerase activity. Specifically, mutation K102A affects transcription and replication due to a significant decrease in cap-binding and viral promoter-binding activity, whereas D108A and K134A primarily cause a transcriptional defect due to a loss of endonuclease activity [244]. Moreover, mutation at position 157 leads to a replication defect, impairing cRNA synthesis function without affecting viral transcription [245]. Mutational analyses have also been conducted on the C-terminal region of PA (PA-C), revealing its critical role in viral RNA synthesis. For instance, mutating residue Y393, which is subject to phosphorylation, affects both mRNA and cRNA synthesis by preventing the binding of the 5' promoter of the vRNA, selectively controlling viral polymerase activity [243]. Di-codon mutations in PA-C, including K497A/T498A, N499A/L500A, and others, almost entirely eliminate viral RNA synthesis, significantly affecting mRNA, cRNA, and vRNA levels [242]. Intriguingly, certain mutations within PA-C result in a polymerase that can still synthesise RNA but fails to produce infectious viral particles, highlighting the multifaceted roles of PA in the viral life cycle [242].

Our observations of mutations at the CTD-binding site within PA-C further elucidate its role in both transcription and replication (Figure 3.6, Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10). Mutations such as K289A, Y445A, E449A, R454A, and K635A disrupted transcription in the NA-based primer extension assay while showing minimal impact on genome replication (Figure 3.7). These amino acid residues, alongside the previously extensively studied R638, are involved in direct interactions with pS5 of the first and third heptad repeat of the Pol II CTD (pS_{5a} and pS_{5c}) as well as Y1 of the first repeat (Y_{1a}), highlighting that interactions with these three amino acids are the most critical for CTD binding (Figure 3.1) [96, 215]. These findings suggest, in line with other studies, that the interaction between the CTD and the viral polymerase is crucial for

enabling cap-snatching, which is essential for viral mRNA synthesis [40, 41]. Interestingly, mutations such as E416A, R551A, and R559A affected replication more than transcription. The vRNP reconstitution assays on the NA and NS segments indicate overlapping results with slight differences likely attributed to statistical analysis and image background variations (Figure 3.7, Figure 3.8). The NS-based vRNP reconstitution assay indicates that mutations K289A, R454A, and K635A not only impact transcription but also reduce replication by up to 40% of the total signal, highlighting the role of the PA-C domain for both transcription and replication (Figure 3.8).

Further analysis revealed that a subset of PA mutations also disrupt ANP32B binding in a chimeric FluPol model, as shown in our split-luciferase assay (Figure 3.9). The FluPol subunits have highly conserved sequences, and a chimeric WSN-1918 FluPol was expected to be compatible with ANP32B binding in a split luciferase assay [112]. Results from this experiment indicate that the mutation of residues K289, E416, R454, R551, R559 disrupts ANP32B binding in a chimeric FluPol model (Figure 3.9). This observation aligns with the replication defects observed in the vRNP reconstitution assays. However, a full 1918 FluPol model in a split-luciferase experimental set-up is required to validate these observations. The high background level of the -PB2 control is potentially caused by forming a stable PB1-PA dimer in the absence of PB2, which can provide an interaction site for ANP32B on PA-C (Figure 3.9) [246, 247]. In the absence of PB1, PA would likely be expressed at lower levels, and a stable dimer would be unable to form with PB2 alone. Omitting PA would abolish the binding interface for ANP32B, resulting in a very low signal. In addition, a minigenome assay on chimeric WSN-1918 FluPol could be conducted and compared to pure 1918 FluPol to ensure that the combination of FluPol subunits from different IAV strains does not impact its function. Furthermore, establishing a CTD-based split luciferase assay system for 1918 FluPol would add valuable insights into the role of the CTD in the context of transcription and replication. This would facilitate the performance of *in vivo*

competition binding assays between ANP32 and the CTD [26].

Recent studies of FluPol A, FluPol B, and FluPol C in complex with human and avian ANP32 revealed that the ANP32 binding site overlaps with part of the Pol II CTD-binding site (Figure 3.11) [74, 145, 248, 249]. ANP32 proteins have been shown to be essential factors for the genome replication of all three types of influenza viruses [145, 248–251]. Specifically, FluPol A was shown to interact with ANP32 through its PA-C domain [248, 249]. The involvement of FluPol A PA-C in ANP32 binding has also been demonstrated by data analysing the effect of mutations in PA-C on ANP32 binding and the replication activity of FluPol [74, 248, 249]. In addition, the importance of PA-C in the viral replication cycle has recently been highlighted through the use of nanobodies [150]. Specifically, two nanobodies that bind PA-C at sites adjacent to the CTD peptide binding site have been shown to block CTD binding, inhibiting both mRNA and cRNA synthesis and, consequently, viral growth. Therefore, it is likely that PA mutations E416A, R551A, and R559 might also inhibit genome replication by interfering with ANP32 binding [248].

Further analysis of the impact of mutations in the PA-C on overall polymerase function showed that transcription and replication defects both lead to a reduction or the abolition of polymerase activity (Figure 3.10). Despite normal mRNA levels in primer extension assays, some PA mutants exhibited perturbed polymerase activity in minigenome assays (Figure 3.7, Figure 3.8, Figure 3.10). This observed discrepancy is likely attributed to the experimental framework of the minigenome assay, where polymerase activity is indirectly measured through the activity of the luciferase reporter instead of a direct measurement of the synthesised mRNA in the primer extension assay. While normal mRNA levels might be produced in the minigenome assay, the introduced mutations could impact polymerase fidelity, resulting in the expression of a dysfunctional luciferase. In addition, similar to the NA and NS-based primer extension assays, Western blotting could be conducted

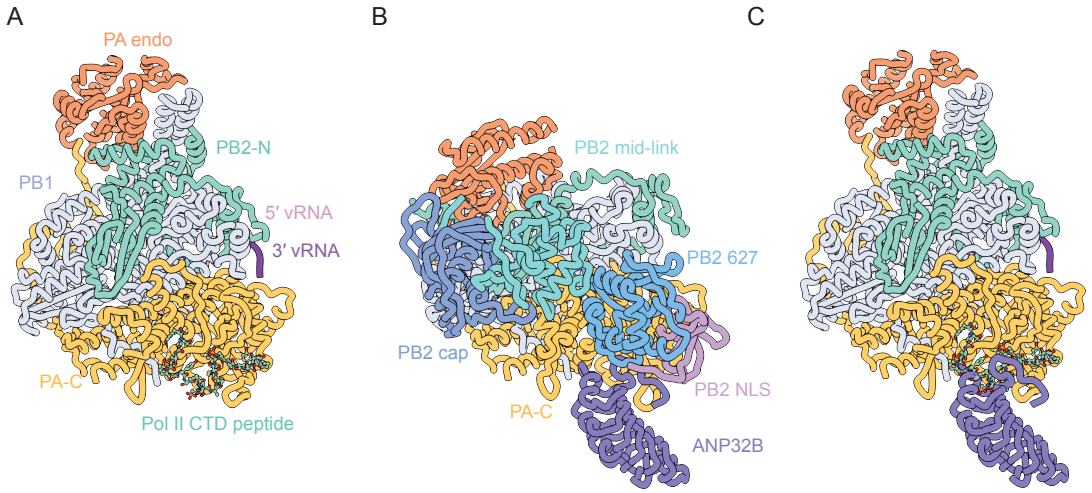

Figure 3.11: Comparison of the Pol II CTD and ANP32 binding sites on the PA-C domain. (A) 1918 FluPol A from H1N1 virus bound to Pol II CTD peptide (PDB ID 8R60). **(B)** FluPol A from H5N1 virus in the encapsidating conformation bound to human ANP32B (PDB ID 8R1J), aligned on PB1 and PA-C to the structure in panel A. **(C)** Positioning of the ANP32B-bound structure on the CTD-bound 1918 FluPol A structure (A) based on the superposition in (B) indicates an overlap of the Pol II CTD and ANP32B binding sites on PA-C.

in future experiments to confirm protein expression levels in the minigenome assay.

Descendants of the 1918 H1N1 influenza A virus have persisted in humans for over a century, contributing genes to new viruses and causing pandemics and epidemics [241]. This work has uncovered the structural basis of the binding of the CTD of Pol II to the polymerase of the 1918 pandemic influenza A virus. Future research aiming to understand cap-snatching in molecular detail in the context of mammalian Pol II will require the inclusion of additional accessory proteins suspected to contribute to the complex formation between Pol II and the viral polymerase. The highly repetitive and intrinsically disordered nature of the CTD drives phase separation and leads to the recruitment of Pol II and transcription factors that together form condensates [223, 252–255]. In the future, CTD-driven phase separation could be investigated within the context of viral infection to facilitate a spatiotemporal understanding of the links between the host transcriptional apparatus and viral transcription and replication beyond the PA-C.

4

Structural characterisation of Pol II CTD binding to the influenza virus polymerase

4.1 Introduction

Influenza viruses hijack the host transcriptional machinery to perform viral transcription and rely on host RNA caps to prime viral mRNA synthesis [41]. The influenza virus polymerase closely associates with RNA polymerase II (Pol II) through direct binding to the pS5-modified C-terminal domain (CTD) of Pol II, facilitating the cap-snatching process [40]. Previous structural studies using synthetic pS5 CTD_{4x} mimetic peptides revealed bipartite binding sites in FluPol from influenza A virus (IAV), influenza B virus (IBV), and influenza C virus (ICV) [26, 96, 97]. Only recently, two studies identified continuous binding sites for the CTD on the FluPol surface in IAV, providing detailed molecular insights into the FluPol-Pol II interaction and highlighting the crucial role of CTD binding in viral transcription and replication [74, 112]. FluPol can adopt distinct conformational states to fulfill its various functions [40, 44, 73]. For instance, in the presence of vRNA, FluPol A and FluPol B have been observed in a transcription pre-initiation state [77, 102]. In this state, the PB2 cap-binding domain and the PA endonuclease domain are positioned in a conformation compatible with cap-snatching [83]. FluPol C bound to vRNA has been observed in a replicase conformation, a transcriptionally inactive conformation, when bound to ANP32 but also in a transcriptase conformation in the presence of Pol II CTD, highlighting the complex structural rearrangements FluPol has to undergo throughout the viral life cycle [97, 145].

To this date, structural details on the interaction of Pol II with other cellular proteins have been primarily explored in the context of cellular gene transcription [151]. Influenza virus research has not yet fully elucidated how the influenza virus hijacks host transcription at a molecular level with a structure of an active cap-snatching complex of FluPol and Pol II still missing. Furthermore, other transcription factors pivotal during promoter-proximal pausing have been associated with viral transcription in proteomics and cell-based experiments but the details of a potential interaction with these factors and the viral polymerase

are not yet available [74]. Given that the mammalian Pol II CTD is significantly longer than the synthetic CTD mimetic peptides used in earlier studies, the FluPol-Pol II interaction likely extends beyond the currently identified binding sites [40, 41]. This interaction is also hypothesised to serve as a platform for recruiting various other host factors and RNA processing factors essential for Pol II transcription. Therefore, understanding the structural and functional dynamics of the interaction of FluPol with host transcriptional factors, especially Pol II, is vital for elucidating the mechanisms underlying cap-snatching and the coordination of viral transcription and replication [74].

The results presented in this chapter offer insights into FluPol-Pol II CTD complex formation. At the initial stages of this project, a phosphomimetic S5E CTD mimetic peptide was used that contained a YSPTEPS heptad repeat sequence. To further investigate the FluPol-Pol II CTD interaction at a structural level, recombinant CTD mimetic peptides containing 26 repeats (CTD_{26x}) of the Pol II CTD consensus sequence YSPTSPS were generated and modified to carry a pS5 modification. Biochemical interaction experiments demonstrated that a phosphomimetic S5E CTD_{26x} interacts with FluPol C but not with FluPol A or FluPol B. In contrast, a phosphorylated $pCTD_{26x}$ peptide interacts with FluPol A, B, and C. These interactions were then further explored in structural studies. Cryo-EM experiments with $pCTD_{26x}$ and FluPol A and FluPol C proved challenging due to technical difficulties related to sample aggregation. Finally, mammalian native Pol II with a CTD consisting of 52 heptad repeats was employed to study complex formation with FluPol using cryo-EM. We found that the native Pol II CTD binding site mostly overlaps with the synthetic pS5 CTD_{4x} binding site in 1918 FluPol A, providing a foundation for determining new structures of FluPol in complex with Pol II and other host transcriptional factors.

4.2 Results

4.2.1 Generation of recombinant phosphomimetic Pol II CTD mimetic peptides

So far, published structures of CTD-bound FluPol employed short synthetic pS5 CTD_{4x} mimetic peptides since a full-length Pol II-FluPol complex has been technically difficult to obtain [26, 74, 96, 97]. Considering that the Pol II CTD contains in mammals 52 heptad repeats and in yeast 26 heptad repeats, I aimed to study the FluPol-Pol II CTD interaction using a longer CTD model to investigate potential new binding sites on the viral polymerase [40]. At the start of the project, native Pol II was not yet available and only later became available in very small amounts that were only suitable for small cryo-EM pilot studies. Synthetic long CTD peptides generated by solid-phase peptide synthesis (SPPS) are not available since the yield of the desired peptide decreases exponentially in a growing peptide chain [256]. Therefore, I recombinantly expressed and purified CTD mimetic peptides containing 13 and 26 repeats of the heptad consensus sequence in *E. coli*. All recombinantly expressed CTD mimetic peptides were originally designed by Dr Olga Jasnovidova with an N-terminal GST-tag and a C-terminal His_{7x}-tag that protects the peptide from degradation (Dr R. Štefl, personal communication). The tags flank a CTD sequence of 13 or 26 repeats of the Pol II CTD consensus sequence YSPTSPS. Initially, I generated recombinant phosphomimetic peptides with 13 and 26 repeats, respectively, and purified all constructs using SEC purification (Figure 4.1 A-D). First, I expressed and purified a 26-repeat peptide GST-CTD_{26x}-His_{7x} (CTD_{26x}) in *E. coli* with an expected size of 47 kDa. Samples collected during the purification using Ni-NTA beads and the final SEC were analysed by SDS-PAGE. A main band at approximately 47 kDa indicates the presence of CTD_{26x} with degradation products visible as a smear underneath the main product band (Figure 4.1 A). The SEC trace of CTD_{26x} shows one peak at a retention volume of 8.8 ml, a shoulder at 11.3 ml and the main product peak at 13.3 ml (Figure 4.1 B). The purification of a 13-repeat phosphomimetic S5E GST-CTD_{13x}-His_{7x} with a molecular weight of 37 kDa (S5E CTD_{13x}) construct shows a

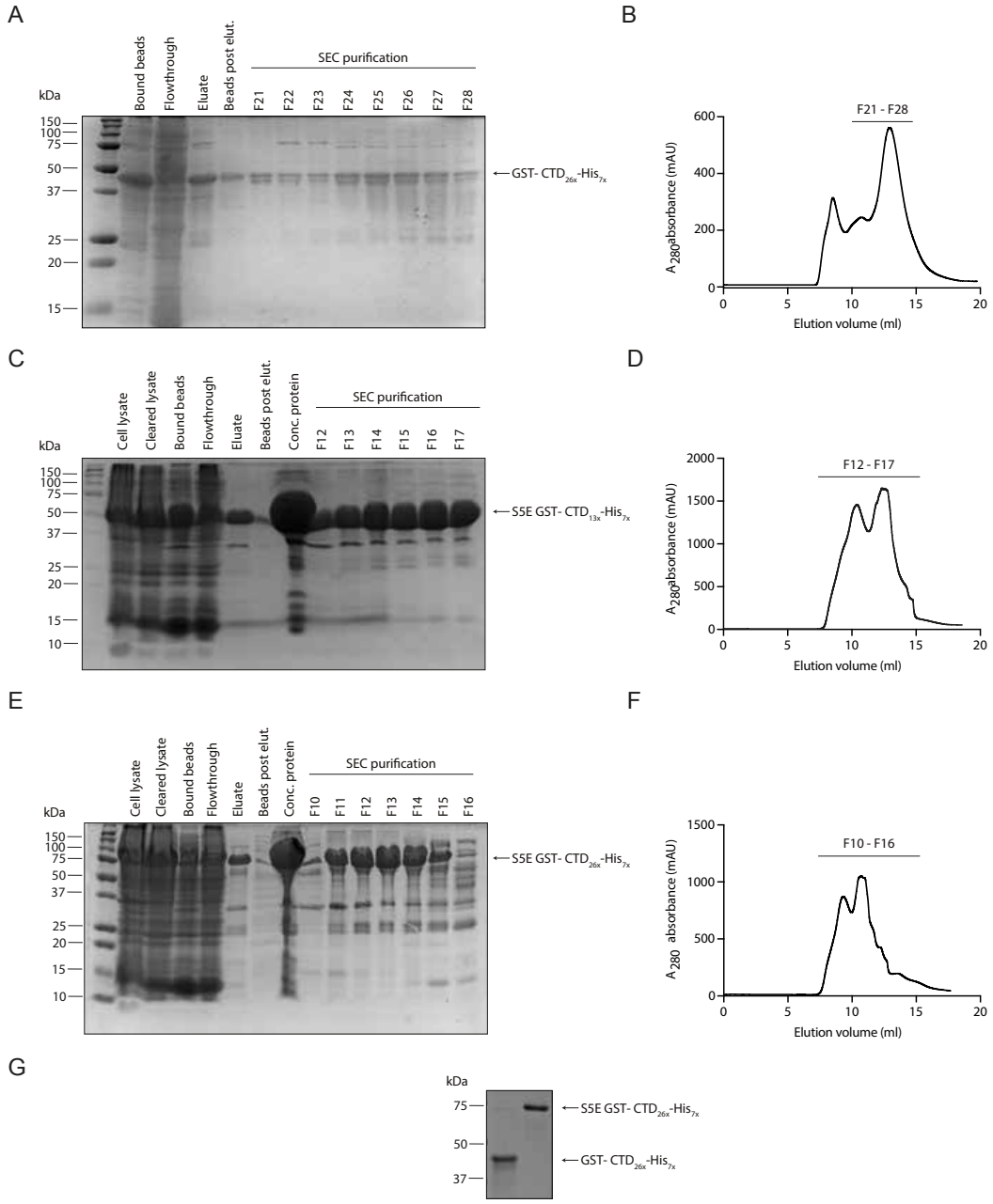

Figure 4.1: Expression and purification of recombinant CTD_{13x} and CTD_{26x} peptides. The Pol II CTD peptides were expressed in *E. coli* and samples taken during the purification process were analysed by SDS-PAGE analysis. Ni-NTA sepharose beads were incubated with cleared cell lysate and then centrifuged and washed (bound beads). Protein was eluted with imidazole (eluate, beads post elution) followed by concentration of the sample (conc. protein) prior to SEC purification. **(A)** Purification of a 26-repeat unphosphorylated GST-CTD_{26x}-His_{7x} peptide with lower amounts of protein loaded than in **C** and **E**. **(B)** Elution profile of the SEC purification of unphosphorylated GST-CTD_{26x}-His_{7x}. **(C)** Purification of a 13-repeat phosphomimetic S5E GST-CTD_{13x}-His_{7x} peptide. **(D)** Elution profile of the SEC purification of GST-CTD_{13x}-His_{7x}. **(E)** Purification of a 26-repeat phosphomimetic S5E GST-CTD_{26x}-His_{7x} peptide. **(F)** Elution profile of the SEC purification of GST-CTD_{26x}-His_{7x}. **(G)** SDS-PAGE analysis showing gel migration of the unphosphorylated peptide with a band around 45 kDa and a band around 75 kDa for the S5E peptide.

strong band at approximately 50 kDa, corresponding to the peptide (Figure 4.1 C). Background bands are visible above and below the main band due to the high amount loaded on the SDS-PAGE (Figure 4.1 C). The observed discrepancy between the theoretical molecular weight of the peptide and the higher molecular weight indicated in the SDS-PAGE analysis might be attributed to differences in gel migration during SDS-PAGE analysis. The SEC trace of S5E CTD_{13x} reveals two main peaks at 10.6 ml and 12.8 ml, with an intensity of approximately 1500-1600 mAu for each peak. The double peak is likely caused by GST dimerisation, leading to dimerised peptides in the first peak and monomeric peptides in the second peak (Figure 4.1 D). Similar observations were made for the longer 26-repeat S5E GST-CTD_{26x}-His_{7x} (S5E CTD_{26x}) peptide (Figure 4.1 E, F). S5E CTD_{26x} runs at approximately 75 kDa, which is higher than unphosphorylated CTD_{26x} (Figure 4.1 A, E, G). In a side-by-side comparison, the S5E modifications appear to result in strong mobility shifts (Figure 4.1 G). Overall, both the unphosphorylated and S5E CTD_{26x} as well as S5E CTD_{13x} peptide could be purified to a high yield. As similar results were expected in using either of the S5E peptides, the S5E CTD_{13x} peptide was not analysed further. The subsequent interaction experiments were performed using unphosphorylated and S5E CTD_{26x}, which will be presented in the following sections.

4.2.2 Generation of recombinant phosphorylated Pol II CTD mimetic peptides

In the transcription cycle, pS5 modification of the Pol II CTD is a key marker of transcription initiation, which FluPol recognises by directly binding to the phosphorylated CTD [41, 73, 153]. Phosphomimetic substitutions to imitate Ser phosphorylations can produce different binding properties than the phosphorylated analyte [257]. Given the preference for using a pS5-modified CTD_{26x} over a phosphomimetic model, I was able to generate a phosphorylated CTD_{26x} with pS5 modifications once I established the *in vitro* phosphorylation of CTD_{26x} using

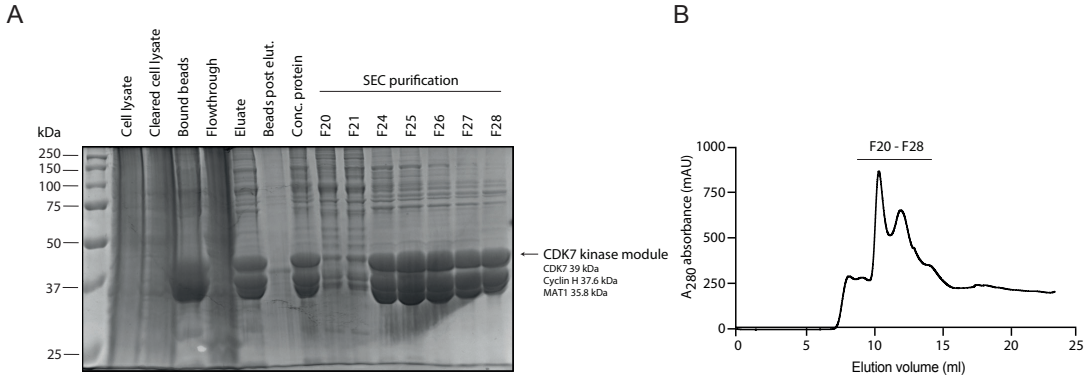

Figure 4.2: Expression and purification of CDK7 kinase module. (A) CDK7 kinase module, consisting of MAT1, CDK7, and cyclin H, was expressed in *Sf9* insect cells and purified using Ni-NTA beads, followed by size-exclusion chromatography (SEC). Samples of the purification were then analysed by SDS-PAGE. **(B)** Elution profile of the SEC purification of CDK7 kinase module.

the Ser5-specific kinase CDK7 [258].

The CDK7 kinase module employed in this study to phosphorylate CTD_{26x} consists of CDK7, Cyclin H, and MAT1, forming a reversibly dissociable 3-subunit kinase module in the TFIIH complex. TFIIH is a 10-subunit transcription factor with various biological roles, including transcription, DNA repair, and cell cycle regulation [259]. The CDK7 kinase activity and substrate specificity are regulated by the Cyclin H and MAT1 subunits. Specifically, Cyclin H controls the selectivity of CDK7 towards cell cycle regulatory substrates and the Pol II CTD. Furthermore, MAT1 guides CDK7-mediated phosphorylation towards DNA-binding transcription factors and anchors the TFIIH core [259]. Initially, I expressed and purified a CDK7 kinase module (CDK7) in *Sf9* insect cells using a baculovirus system (Figure 4.2). Each subunit contained a C-terminal His-tag suitable for purification using Ni-NTA agarose beads. SDS-PAGE analysis of the CDK7 kinase module purification process exhibits three bands at approximately 37 kDa, corresponding to the CDK7 kinase module subunits. These bands are mainly observed in samples expected to contain the desired protein, including the eluate and SEC purification fractions (Figure 4.2 A). The SEC elution profile of CDK7 reveals two major peaks at a retention volume of 10.6 ml and 12.4 ml

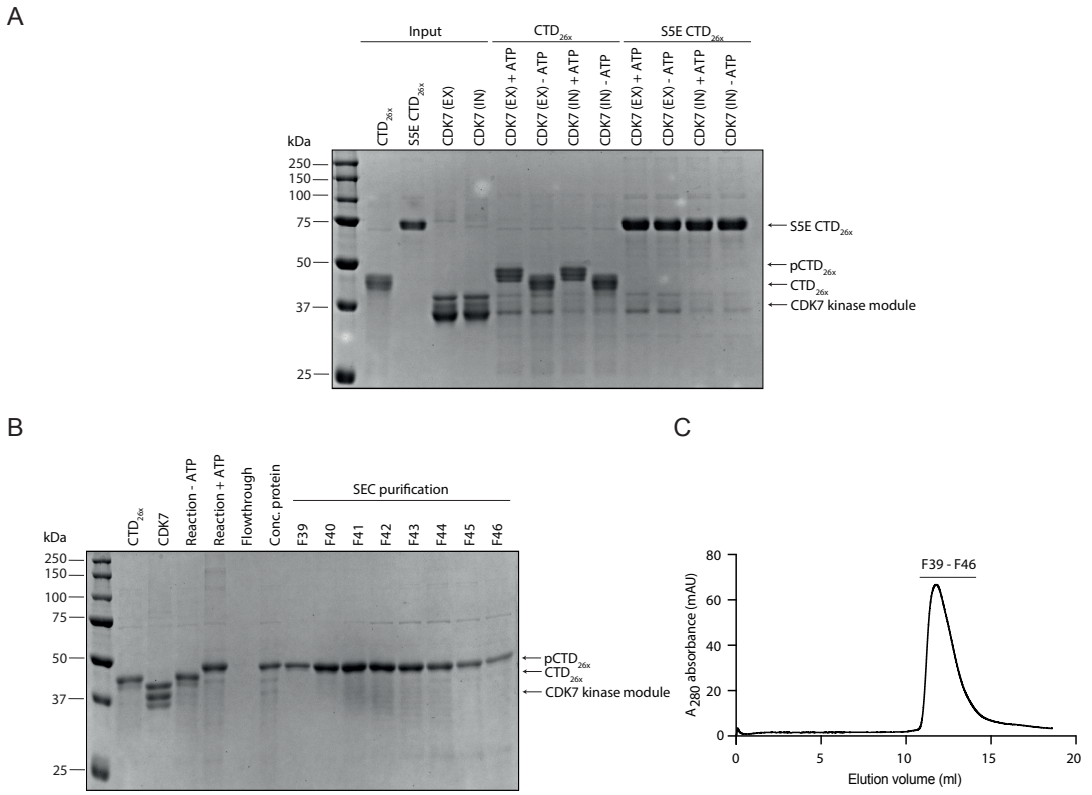

Figure 4.3: Phosphorylation of CTD_{26x} peptide using a CDK7 kinase module. **(A)** Externally provided CDK7 kinase module (EX) module from our collaborators and an in-house expressed CDK7 kinase module (IN) were used to phosphorylate CTD_{26x} peptide, the phosphomimetic S5E CTD_{26x} peptide served as a negative control. CDK7 was mixed with the CTD mimetic peptides with (+ATP) and without (-ATP) ATP, the phosphate group donor in this reaction. A mobility shift can be observed for the unphosphorylated CTD_{26x} peptide after incubation with ATP and the CDK7 kinase module. No visible shift can be observed in reactions with S5E CTD_{26x} peptide. **(B)** SDS-PAGE analysis of samples from the purification of phosphorylated CTD_{26x} peptide ($pCTD_{26x}$). **(C)** Elution profile of the SEC purification of phosphorylated CTD_{26x} peptide.

with an absorbance ranging from 700-900 mAU (Figure 4.2 B). CDK7 kinase module is present in all analysed fractions, and the observed peaks might result from oligomerisation of the three-subunit complex (Figure 4.2 A, B).

To test the activity of the purified CDK7 kinase module, I performed a small-scale phosphorylation reaction of CTD_{26x} using ATP as a phosphate group donor and comparing externally sourced CDK7 kinase module from our collaborators with in-house expressed CDK7 (Figure 4.3 A). CDK7 kinase module, purified by our

collaborators as well as by us, reliably phosphorylated unmodified CTD_{26x}, visible as a mobility shift in reactions supplemented with ATP (Figure 4.3 A). The S5E CTD_{26x}, here used as a negative control, did not exhibit a visible mobility shift after incubation in the presence of ATP and CDK7 kinase (Figure 4.3 A). Finally, the phosphorylation of CTD_{26x} was also performed in a large batch reaction followed by SEC purification to produce pure $pCTD_{26x}$ for interaction studies with FluPol (Figure 4.3 B, C). A mobility shift from approximately 45 kDa to 50 kDa can be observed after completion of the phosphorylation reaction for $pCTD_{26x}$ relative to unmodified CTD_{26x}. Minor levels of degradation are visible as a laddering pattern below the main product band (Figure 4.3 A, B). After SEC purification, fractions containing $pCTD_{26x}$ were pooled and used in interaction experiments and cryo-EM studies described in the following sections.

4.2.3 Phosphomimetic and phosphorylated CTD_{26x} peptides bind differently to viral polymerases from IAV, IBV, and ICV

Since past studies only used short synthetic CTD_{4x} peptides to assess the interaction with FluPol, I aimed to characterise the interaction of long CTD mimetic peptides with FluPol, hypothesising that they would enable more detailed insights beyond previously identified binding sites [40, 41]. For the interaction experiments, I purified recombinant FluPol A, FluPol B, and FluPol C expressed in *Sf9* insect cells using established protocols in our research group (data not shown). Given that distinct CTD binding sites were identified across the three FluPol types using synthetic CTD_{4x} peptides, I aimed to explore how longer, recombinant CTD_{26x} peptides might interact with the three FluPol types. To ensure that FluPol is in a transcription-compatible conformation, short 5' end (15 bases) and 3' end (14 bases) vRNA promoter sequences were pre-bound to FluPol A, FluPol B, FluPol C (see sequences under 2.1.1).

The interaction of FluPol A, FluPol B, and FluPol C with S5E CTD_{26x} and $pCTD_{26x}$ was analysed in pulldown assays (Figure 4.4, Figure 4.5, Figure 4.6). All

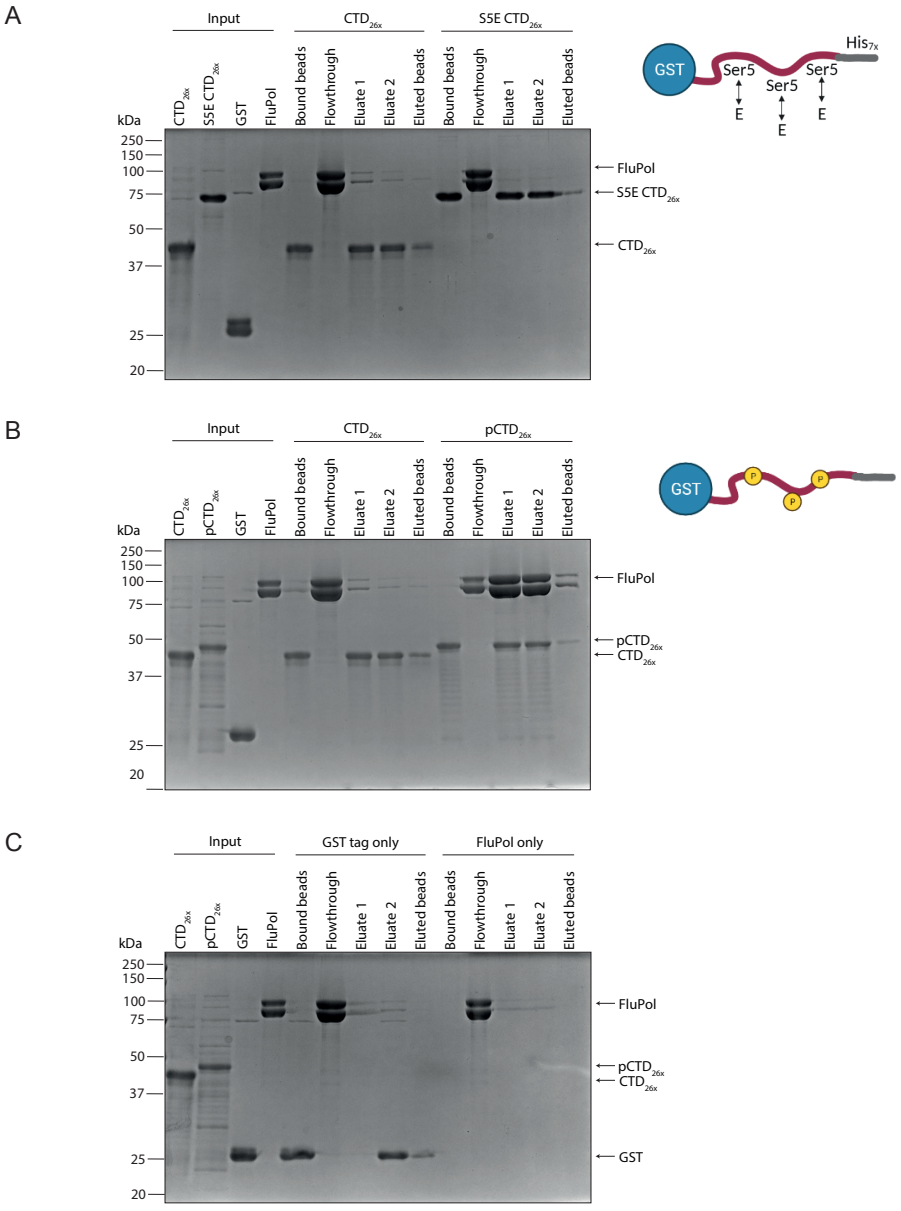

Figure 4.4: Pulldown of 1918 FluPol A with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide. (A, B) The interaction of phosphomimetic S5E CTD_{26x} peptide and phosphorylated $pCTD_{26x}$ peptide was tested with 1918 FluPol A and compared to the interaction with the unphosphorylated CTD_{26x} peptide. CTD peptides were incubated with GST sepharose beads (bound beads), followed by incubation with vRNA-bound FluPol (flowthrough) and elution of the formed complexes with GSH. Elution samples were taken after 2 hours and overnight incubation (eluate 1, eluate 2) along with the remaining beads (eluted beads). **(C)** Incubation with the GST tag or the viral polymerase only served as a negative control.

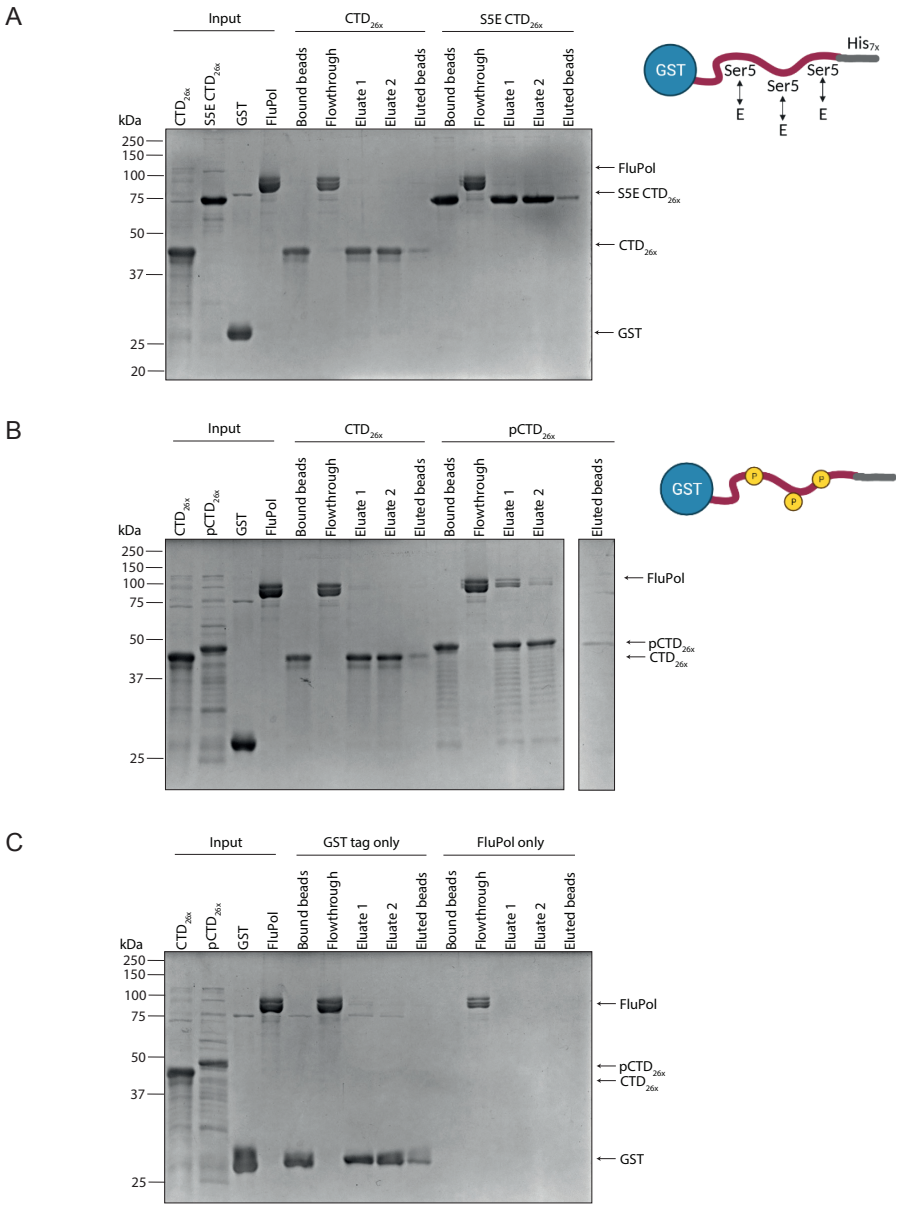

Figure 4.5: Pulldown of FluPol B with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide. (A, B) The interaction of phosphomimetic S5E CTD_{26x} peptide and phosphorylated $pCTD_{26x}$ peptide was tested with FluPol B and compared to the interaction with the unphosphorylated CTD_{26x} peptide. CTD peptides were incubated with GST sepharose beads (bound beads), followed by incubation with vRNA-bound FluPol (flowthrough) and elution of the formed complexes with GSH. Elution samples were taken after 2 hours and overnight incubation (eluate 1, eluate 2) along with the remaining beads (eluted beads). **(C)** Incubation with the GST tag or the viral polymerase only served as a negative control.

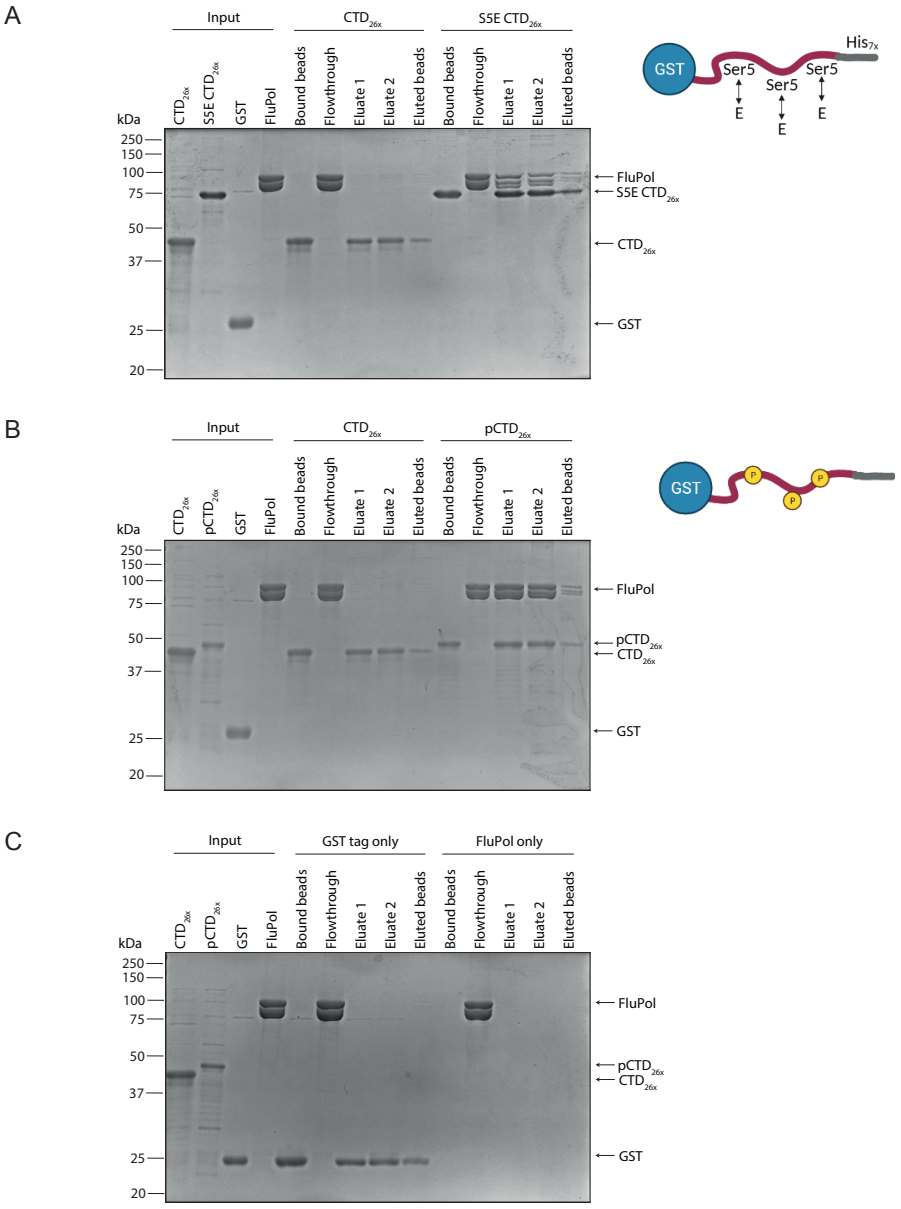

Figure 4.6: Pulldown of FluPol C with phosphomimetic S5E CTD_{26x} and phosphorylated $pCTD_{26x}$ peptide. (A, B) The interaction of phosphomimetic S5E CTD_{26x} peptide and phosphorylated $pCTD_{26x}$ peptide was tested with FluPol C and compared to the interaction with the unphosphorylated CTD_{26x} peptide. CTD peptides were incubated with GST sepharose beads (bound beads), followed by incubation with vRNA-bound FluPol (flowthrough) and elution of the formed complexes with GSH. Elution samples were taken after 2 hours and overnight incubation (eluate 1, eluate 2) along with the remaining beads (eluted beads). **(C)** Incubation with the GST tag or the viral polymerase only served as a negative control.

pulldown assays were carried out in medium salt (300 mM NaCl) conditions due to FluPol aggregation on the beads at low salt (150 mM NaCl) conditions (data not shown). The negative controls consisted of a GST tag to replace the GST-tagged CTD_{26x} peptides and incubation of FluPol in the absence of GST-tagged CTD_{26x} to control for unspecific binding to the GST tag or aggregation of FluPol on the beads in the used buffer conditions (Figure 4.4 C). This was done to control for non-specific binding to the GST tag or aggregation of FluPol on the beads under the buffer conditions used.

In the pulldown assay with FluPol A, the interaction of FluPol was tested with S5E CTD_{26x} and $pCTD_{26x}$ (Figure 4.4 A, B). Input proteins for the pulldown assay, comparing the interaction of unphosphorylated CTD_{26x} and S5E CTD_{26x} with FluPol A appear on the SDS-PAGE at the expected sizes accompanied by minor impurities for CTD_{26x}, S5E CTD_{26x}, $pCTD_{26x}$, and the GST tag (Figure 4.4 A). The $pCTD_{26x}$ peptide exhibits a laddering pattern underneath the main band, corresponding to degradation products. The GST tag shows a doublet (Figure 4.4 A), which is likely a gel artefact that was not observed in other SDS-PAGE analyses using the same GST tag protein batch (Figure 4.4 B, C). Unphosphorylated CTD_{26x} and S5E CTD_{26x} were incubated on glutathione sepharose and washed before FluPol A was incubated with the peptides to compare its interaction with these peptides. FluPol A mostly remained in the flowthrough fractions with background bands visible in both the control condition with unphosphorylated CTD_{26x} and the test condition with S5E CTD_{26x} (Figure 4.4 A). Additionally, some spillover from the flowthrough samples could be detected in adjacent gel lanes (Figure 4.4 A-C). In contrast, incubation of FluPol A with $pCTD_{26x}$ resulted in complex formation and strong bands corresponding to both FluPol and $pCTD_{26x}$ in the eluate samples and a reduction of the FluPol band intensity in the flowthrough sample of the $pCTD_{26x}$ pulldown was observed (Figure 4.4 B). The negative control with GST alone indicated minor background bands corresponding to FluPol in the eluate (Figure 4.4 C).

These bands were comparable to the detected levels in the control condition using unmodified CTD_{26x} to pulldown FluPol A (Figure 4.4 A, B).

In a pulldown experiment of FluPol B with S5E CTD_{26x} and $pCTD_{26x}$, similar observations were made as in the previously described pulldown with FluPol A (Figure 4.5 A). FluPol B did not indicate an interaction with S5E CTD_{26x} in the eluate (Figure 4.5 A). In contrast, immobilised $pCTD_{26x}$ was able to pulldown FluPol B, albeit not at the same levels observed in the pulldown with FluPol A (Figure 4.4 B, Figure 4.5 B). FluPol B exhibited the lowest band intensities in the eluate fractions of the test condition, likely due to the low concentration of the available FluPol B batch used in the pulldown experiment (Figure 4.5 B). Lastly, the negative controls indicated no unspecific interactions were formed (Figure 4.5 B, C).

A pulldown experiment assessing the interaction of FluPol C with S5E CTD_{26x} and $pCTD_{26x}$ demonstrated that both the phosphomimetic and the phosphorylated CTD_{26x} peptides bind to FluPol C (Figure 4.5 A, B). In comparison to the control condition with CTD_{26x}, FluPol and S5E CTD_{26x} bands are present in the eluate, indicating the formation of a complex (Figure 4.5 A, B). Similarly, $pCTD_{26x}$ co-eluted with FluPol C in the eluate samples and exhibited stronger FluPol bands compared to the pulldown with S5E CTD_{26x} (Figure 4.5 A, B). In conclusion, FluPol revealed an interaction with $pCTD_{26x}$ in pulldown experiments, irrespective of the FluPol type and FluPol C showed additional capability to interact with phosphomimetic S5E CTD_{26x}.

4.2.4 Phosphomimetic S5E CTD_{26x} mimetic peptide forms a complex with the influenza C virus polymerase

In pulldown experiments, FluPol C was bound to the phosphomimetic and the phosphorylated CTD_{26x} peptide. I sought to validate these interactions further using SEC analysis, aiming to structurally characterise the formed complexes using cryo-EM (Figure 4.7). Initially, when the generation of $pCTD_{26x}$ was not fully established, I performed SEC analysis on S5E CTD_{26x} and vRNA-bound FluPol C that adopts a cap-snatching compatible conformation to analyse protein complex formation under low salt conditions (150 mM NaCl) [97]. FluPol C was incubated with promotor 5' and 3' vRNA at 1.2 molar excess before adding S5E CTD_{26x} or unphosphorylated CTD_{26x} at 1.5 molar excess (see sequences under 2.1.1). The mixtures were then incubated on ice, separated on an SEC column, and collected fractions were analysed by SDS-PAGE analysis (Figure 4.7).

The SEC trace of unphosphorylated CTD_{26x} peptide with FluPol C showed a combination of individual peaks, and CTD_{26x} was not found in fractions at a smaller retention volume, which would correspond to a formed complex (Figure 4.7 A). CTD_{26x} was not observed in SEC fractions earlier than its typical retention volume alone, and FluPol C exhibited a characteristic dimer and monomer peak pattern (Figure 4.7 A) [122]. In contrast, the S5E CTD_{26x} peptide showed a peak shift and stochiometric band intensities for FluPol C and S5E CTD_{26x} peptides at a smaller retention volume (Figure 4.7 B). S5E CTD_{26x} co-eluted with FluPol C in SEC fractions when incubated together, indicating a formed complex of FluPol C with S5E CTD_{26x} (Figure 4.7 B). In summary, an interaction of FluPol C was observed with S5E CTD_{26x} in interaction experiments, demonstrating the possibility of complex formation between FluPol and a phosphomimetic CTD construct.

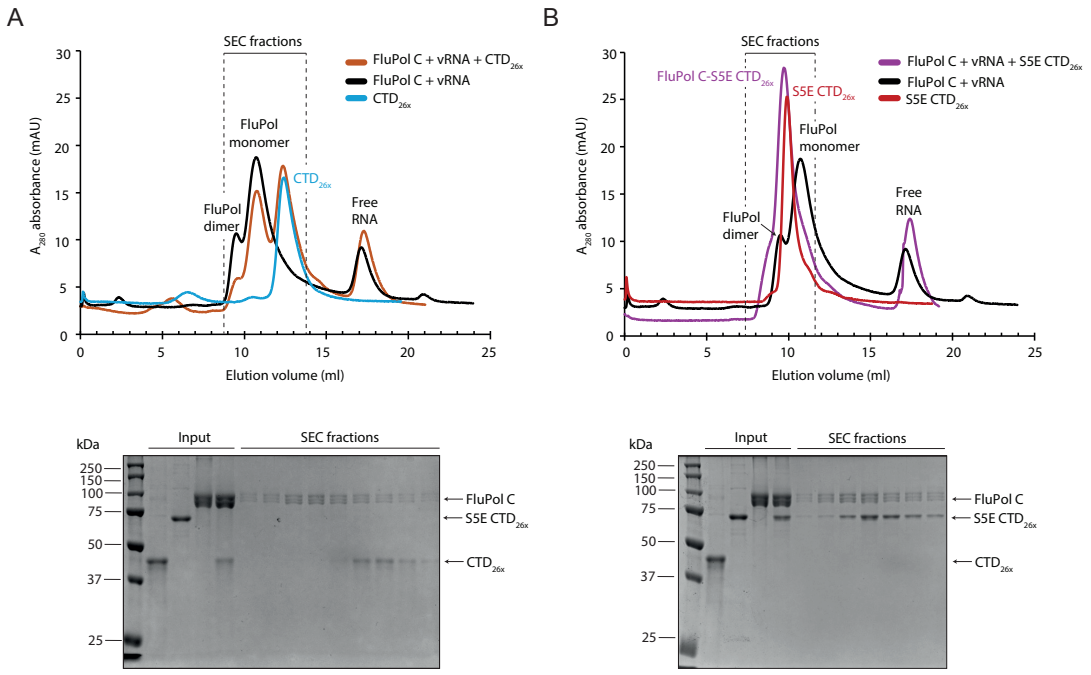

Figure 4.7: SEC analysis of phosphomimetic S5E CTD_{26x} with FluPol C. **(A)** SEC analysis elution profile of influenza C viral polymerase with unphosphorylated CTD_{26x} peptide (top) and SDS-PAGE analysis of collected samples (bottom). **(B)** SEC analysis elution profile of influenza C viral polymerase with phosphomimetic S5E CTD_{26x} peptide (top) and SDS-PAGE analysis of collected samples (bottom).

4.2.5 Phosphorylated Pol II CTD_{26x} mimetic peptide forms complexes with the influenza A, B, and C virus polymerase

Pulldown experiments revealed a potential interaction with the $pCTD_{26x}$ and FluPol A, FluPol B, and FluPol C. Similar to the SEC analysis of S5E CTD_{26x} and FluPol C, I conducted SEC analyses to confirm my observations from pulldown assays, comparing the interaction of CTD_{26x} and $pCTD_{26x}$ with the three FluPol types.

Incubation of 1918 FluPol A D108A with the unphosphorylated CTD_{26x} did not result in a peak shift, and the SEC trace was composed of peaks representing the individual components at their characteristic retention volume (Figure 4.8 A). SDS-PAGE analysis of the collected SEC fractions further confirmed that the two analytes did not co-elute. In contrast, when $pCTD_{26x}$ and FluPol A D108A were

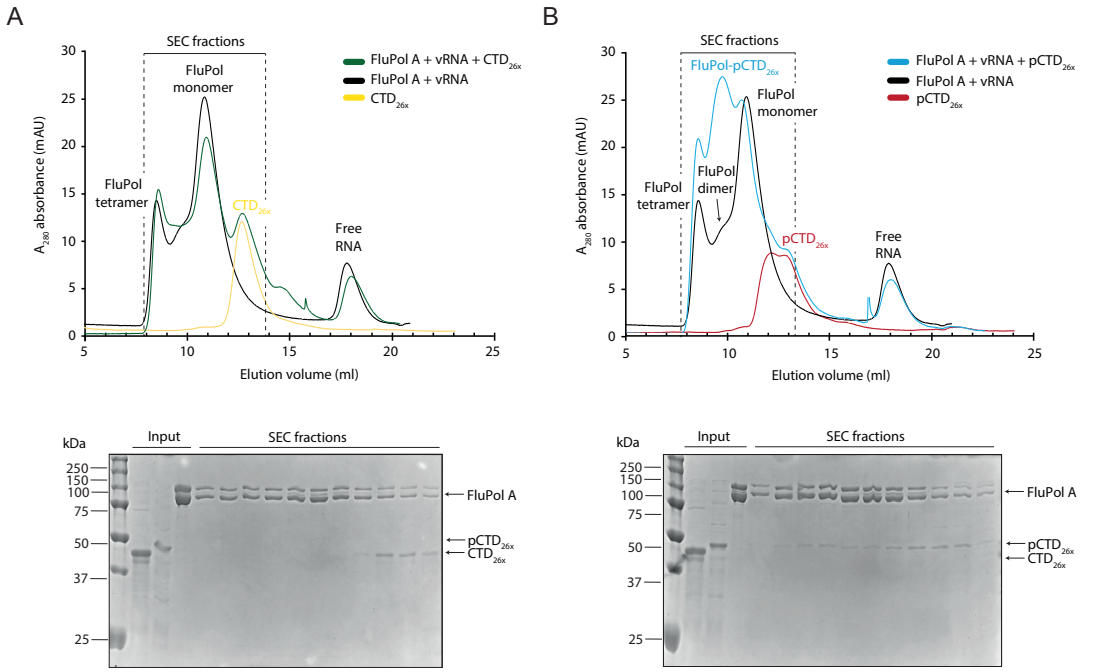

Figure 4.8: SEC analysis of pCTD$_{26x}$ with 1918 FluPol A D108A. (A) SEC analysis elution profile of 1918 influenza A viral polymerase with unphosphorylated CTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom). **(B)** SEC analysis elution profile of 1918 influenza A viral polymerase with pCTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom).

combined, I observed partial co-elution of both analytes and a peak shift towards a smaller retention volume, indicating the formation of a complex (Figure 4.8 B). The detection of pCTD$_{26x}$ bands in SEC fractions collected earlier than the typical retention volume of pCTD$_{26x}$ further supported the potential interaction between pCTD$_{26x}$ and FluPol A D108A (Figure 4.8 B). Interestingly, the SEC control run for FluPol A D108A alone showed a peak pattern with a shoulder at 10 ml preceding the monomer peak of vRNA-bound FluPol A (Figure 4.8 B). The sample of FluPol A D108A and pCTD$_{26x}$ exhibited a different SEC trace shape, with a stronger signal intensity for both the FluPol tetramer and FluPol dimer, the latter yielding the strongest signal out of the three FluPol oligomeric states (Figure 4.8 B).

SEC analysis of FluPol B incubated with pCTD$_{26x}$ demonstrated complex formation, as evidenced by a peak shift in the respective SEC trace compared

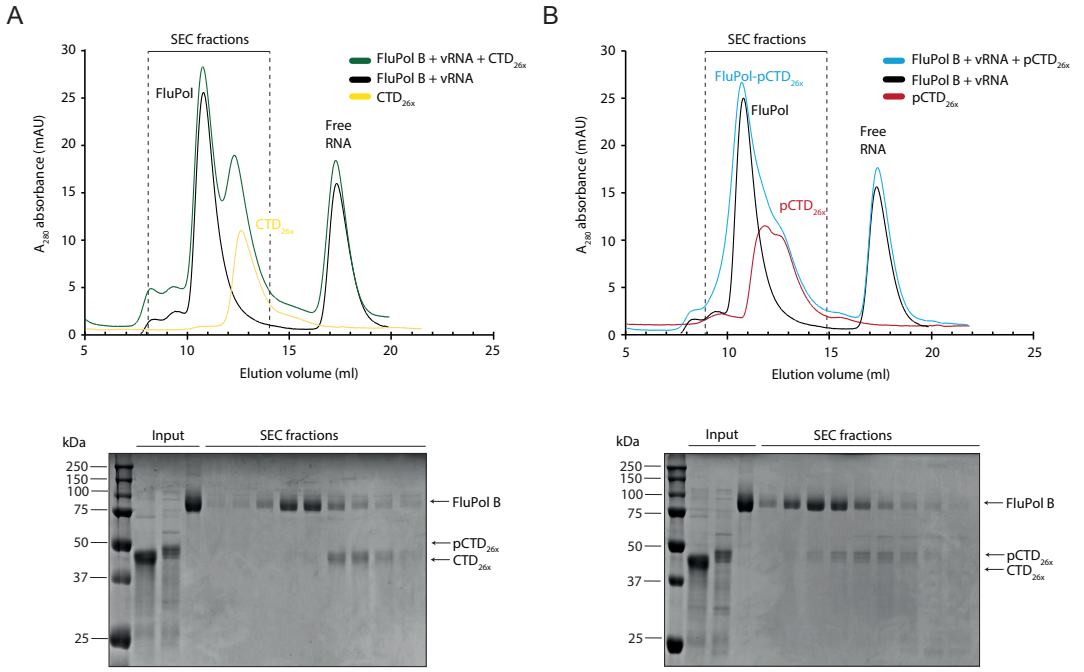

Figure 4.9: SEC analysis of pCTD$_{26x}$ with FluPol B. (A) SEC analysis elution profile of influenza B viral polymerase with unphosphorylated CTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom). **(B)** SEC analysis elution profile of influenza B viral polymerase with pCTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom).

to the control condition using CTD$_{26x}$ without a peak shift (Figure 4.9 A). Incubation of FluPol B with the unphosphorylated CTD$_{26x}$ resulted in SEC traces composed of peaks representing the individual components, with retention volumes similar to the control runs of CTD$_{26x}$ and FluPol B alone (Figure 4.9 A). SDS-PAGE analysis of SEC fractions revealed the presence of pCTD$_{26x}$ in earlier retention volumes compared to CTD$_{26x}$ (Figure 4.9 A, B). Unlike the SEC analysis with FluPol A D108A, the SEC analysis of FluPol B did not show a change in the peak shape (Figure 4.9 B, Figure 4.8 B). The observed peak, indicative of monomeric FluPol, suggests a potential monomeric FluPol B-pCTD$_{26x}$ complex formation (Figure 4.9 B).

Incubation of FluPol C with CTD$_{26x}$ exhibited similar patterns to those observed with FluPol A and FluPol B. When incubated with CTD$_{26x}$, FluPol C did not form a complex, as indicated by the individual peaks corresponding to the control

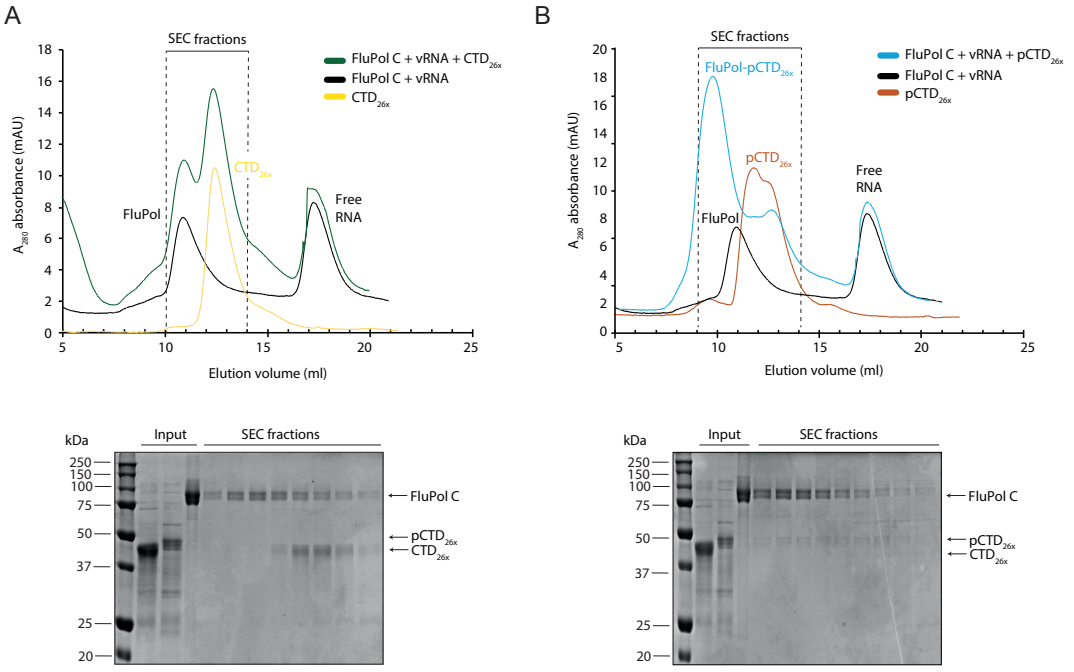

Figure 4.10: SEC analysis of pCTD$_{26x}$ with FluPol C. (A) SEC analysis elution profile of influenza C viral polymerase with unphosphorylated CTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom). **(B)** SEC analysis elution profile of influenza C viral polymerase with pCTD$_{26x}$ peptide (top) and SDS-PAGE analysis of collected samples (bottom).

peptide and FluPol C in the SEC analysis (Figure 4.10 A). In contrast, a complex was formed when pCTD$_{26x}$ was combined with FluPol C, as demonstrated by a peak shift to a smaller retention volume and the detection of pCTD$_{26x}$ bands in earlier fractions during SDS-PAGE analysis (Figure 4.10 A).

In summary, these findings demonstrate the complex formation of FluPol A, FluPol B, and FluPol C with pCTD$_{26x}$ but not with unphosphorylated CTD$_{26x}$, in both pulldown experiments and SEC analyses.

4.2.6 Structural characterisation of the influenza virus polymerase interaction with recombinant Pol II CTD_{26x} mimetic peptide

Cryo-EM data collection was kindly performed by Dr Jeremy Keown and Dr Loïc Carrique at The Division of Structural Biology, University of Oxford.

After pulldown experiments and SEC analyses confirmed the interaction of $pCTD_{26x}$ with FluPol, I aimed to characterise these complexes in a structural approach using cryo-EM. Past research with four-repeat peptides led to the hypothesis that additional CTD binding sites might be discovered using a longer CTD construct, potentially enhancing our understanding of the cap-snatching mechanism [26, 74, 96, 97].

Cryo-EM samples were prepared with FluPol from influenza A, B, and C viruses, with 3' and 5' vRNA promoters. For sample preparation, $pCTD_{26x}$ was added to FluPol in a molar excess ranging from 0.5 to 3 times, followed by grid preparation, screening, and data collection. Data acquisition on $pCTD_{26x}$-FluPol A and $pCTD_{26x}$-FluPol C samples was performed using a 300 kV Titan Krios G3i microscope (Thermo Scientific) with a K2 Summit camera (Gatan). FluPol B-$pCTD_{26x}$ samples were not further analysed due to time constraints. Screening of a 1918 FluPol A sample produced micrographs exhibiting a characteristic particle distribution comparable to a previous 1918 FluPol A dataset collected in the presence of synthetic pS5 CTD_{4x} (Figure 4.11 A) [112]. Despite consistent screening of various FluPol samples, the prepared grids exhibited significant variations in ice thickness across different regions of the sample grid and between individual grids, which hindered our ability to collect a suitable dataset for structural characterisation of the FluPol-$pCTD_{26x}$ complex. Analysis of the 2D classes generated from 548,000 particles revealed the presence of monomeric, dimeric, and tetrameric FluPol A populations (Figure 4.11 B). The FluPol monomers

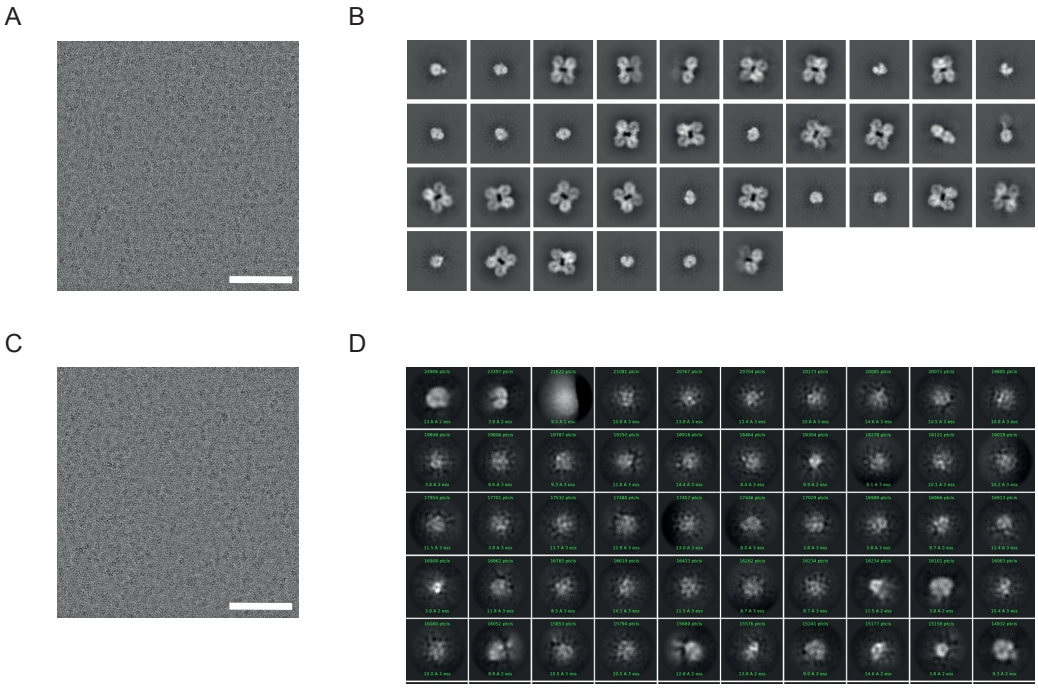

Figure 4.11: Cryo-EM data collection on FluPol A and FluPol C in complex with CTD_{26x}. Data acquisition was conducted on a 300 kV Titan Krios G3i microscope (Thermo Scientific) with a K2 Summit camera (Gatan). **(A)** A micrograph of the FluPol A-$pCTD_{26x}$ sample screening. Scale bar: 100 nm. **(B)** Processed 2D classes generated from 548,000 particles showing monomeric, dimeric, and tetrameric FluPol A populations. **(C)** A micrograph of the FluPol C-$pCTD_{26x}$ sample screening. Scale bar: 100 nm. **(D)** Processed 2D classes generated from 880,000 particles selected from a FluPol C-$pCTD_{26x}$ sample.

appeared spherical, while the dimers and tetramers adopted a symmetrical, cage-like arrangement (Figure 4.11 B). The tetramer 2D class showed a symmetrical organisation with a central hole, and this tetrameric form was the predominant oligomerisation state observed (Figure 4.11 B). Further processing of the 1918 FluPol A-$pCTD_{26x}$ dataset to generate a cryo-EM map included heterogeneous refinement and ab initio reconstruction. Processing of the dataset did not yield a cryo-EM map that could be reliably fitted due to the dominance of preferentially oriented FluPol tetramers. Consequently, these experimental conditions could not confirm the presence of $pCTD_{26x}$ bound to FluPol A.

Preparation of a FluPol C-$pCTD_{26x}$ revealed similar technical challenges during sample preparation and screening. The majority of the prepared grids exhibited

inadequate quality due to varying ice thickness and aggregation. However, a dataset was collected on a grid with satisfactory particle distribution, indicating the presence of FluPol particles (Figure 4.11 C). Data processing of this dataset generated 2D classes from 880,000 particles, which indicated the presence of FluPol with no observed oligomerisation (Figure 4.11 D). Nonetheless, most of the generated 2D classes corresponded to noise, rendering them insufficient for further data processing. Overall, the structural characterisation of $pCTD_{26x}$-FluPol complexes proved technically challenging, frequently involving aggregates observed in both cryo-EM and negative staining approaches (data not shown) which did not allow us to determine a structure of a $pCTD_{26x}$-FluPol complex.

4.2.7 Structural characterisation of the influenza virus polymerase interaction with native Pol II

This work was conducted collaboratively. Pol II used in the following experiments was prepared by Dr Marek Šebesta at Masaryk University, Czech Republic. Dr Jeremy Keown carried out the structural characterisation of the Pol II-FluPol interaction and supported me with grid preparation and data processing. Cryo-EM data collection was performed by Dr Jeremy Keown and Dr Loïc Carrique at the Division of Structural Biology, University of Oxford.

Purification of native Pol II

Cap snatching is a crucial step in the influenza viral life cycle, enabling viral genome transcription. In the cellular context, the viral polymerase binds to the Pol II pS5 CTD, which contains 52 heptad repeats. Further interactions between the viral polymerase (FluPol) and the polymerase core have been described in cell-based experiments, suggesting that FluPol interacts with the RPB4 and RPB7 subunits beyond the CTD [26, 74, 237]. However, the structural and molecular details of these interactions have not been fully elucidated, and an active FluPol-Pol II cap-snatching complex has not yet been presented [40, 41]. In this project, I aimed to investigate the Pol II-FluPol interaction further using

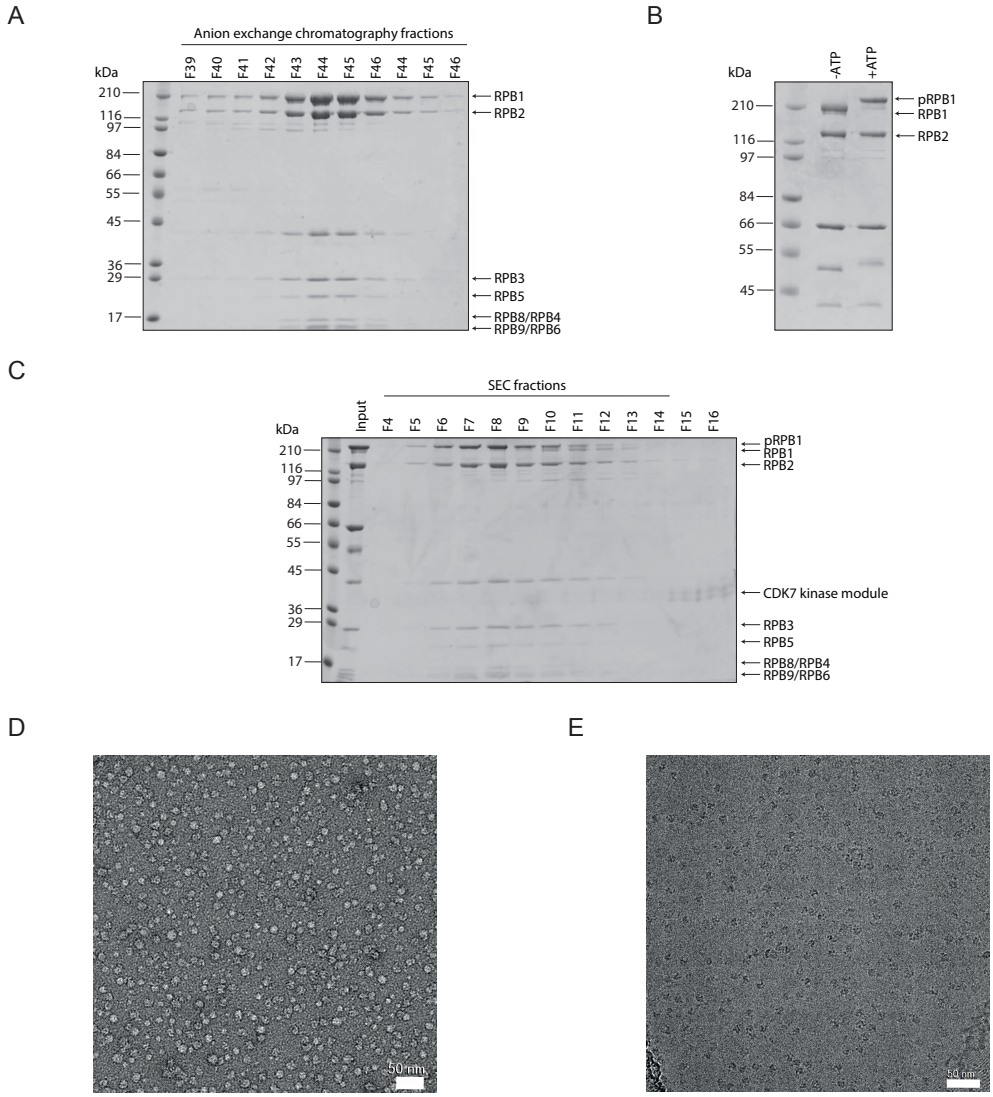

Figure 4.12: Purification of endogenous mammalian Pol II. Pol II was isolated from calf *(Bos taurus)* thymus tissue and purified by ion exchange chromatography, immunoaffinity chromatography, and SEC, followed by pS5 phosphorylation using a CDK7 kinase module and a final SEC purification. The quality of the phosphorylated Pol II was finally analysed using electron microscopy. **(A)** SDS-PAGE analysis of the SEC purification of Pol II. **(B)** SDS-PAGE analysis of phosphorylated Pol II phosphorylation reactions using CDK7 with (+ATP) and without (-ATP). A mobility shift for the phosphorylated RPB1 (pRPB1) can be observed. **(C)** SEC purification after the phosphorylation of Pol II. **(D)** Negative staining EM micrograph of purified pS5 Pol II. Data acquisition was conducted on a Talos F200C microscope with a Ceta-D camera (Thermo Scientific). Scale bar: 50 nm. **(E)** Cryo-EM micrograph of purified pS5 Pol II after a freeze-thaw cycle. Data acquisition was conducted on a Titan Krios microscope (Thermo Scientific) with a K3 camera (Gatan). Scale bar: 50 nm. Data for this figure were provided by Dr Marek Šebesta.

structural approaches.

By employing long recombinant CTD_{26x} peptides, I successfully demonstrated an interaction between these peptides and three distinct FluPol types. However, further structural characterisation of this interaction through cryo-EM proved technically challenging, and I could not obtain a structural model. As an alternative approach, I sought to investigate the interaction between full-length Pol II and FluPol in collaboration with Dr Jeremy Keown, as this would provide a more physiologically relevant model compared to utilising CTD fragments alone. Unfortunately, the limited availability of purified and pS5-modified Pol II provided by our collaborators constrained the number of cryo-EM trials that could be performed.

Endogeneous mammalian Pol II was isolated and purified from calf (*Bos taurus*) thymus tissue. After preparing and filtering the cell lysate, proteins were precipitated and purified through successive rounds of ion exchange chromatography, immunoaffinity chromatography, and SEC. SDS-PAGE analysis of fractions from anion exchange chromatography following protein precipitation and immunoaffinity chromatography (data not shown) revealed two dominant bands around 216 kDa and 130 kDa, corresponding to the Pol II subunits RPB1 and RPB2, respectively (Figure 4.12 A). Fractions 43 to 46 also exhibited the presence of additional bands corresponding to other subunits of the Pol II 12-subunit holoenzyme, including RPB3 (31 kDa) and RPB5 (25 kDa) (Figure 4.12 D). The RPB8 and RPB4 subunits, with similar molecular weights of 17 kDa and 16 kDa, respectively likely migrated together. Similarly, RPB9 and RPB6, both with a molecular weight of 14 kDa, were observed at the bottom of the SDS-PAGE. However, the remaining subunits RPB7, RPB10, RPB11, and RPB12, each with molecular weight smaller than 14 kDa were not captured in the SDS-PAGE analysis (Figure 4.12 A). Importantly, the RPB7 and RPB4 subunits that together form the heterodimeric stalk have been shown to dissociate from the Pol II core

under various experiment conditions [260]. In addition to the Pol II subunits, some background bands were observed around 55 kDa, as well as a dominant band below 45 kDa (Figure 4.12 A).

Following anion exchange chromatography, the relevant fractions were pooled, and Pol II was phosphorylated using CDK7 and ATP as previously described (see section 4.2.2). After the phosphorylation reaction, samples of the reaction with (+ATP) and without (-ATP) ATP were analysed by SDS-PAGE (Figure 4.12 B). The band corresponding to RPB1, which typically runs at 210 kDa, exhibited a mobility shift after incubation with ATP and CDK7 (Figure 4.12 B). Additionally, a faint band was observed below the shifted main band at a similar position to unphosphorylated RPB1 (Figure 4.12 B). Another band running above 116 kDa, corresponding to the RPB2 subunit, showed only a very minor mobility shift, likely a gel artefact (Figure 4.12 B). The additional bands observed, which did not correspond to other subunits, are likely contaminants introduced during the phosphorylation reaction (Figure 4.12 B). After further purification by SEC, fractions 6 to 9 showed bands corresponding to Pol II with phosphorylated RPB1, and later fractions indicated the presence of an additional band below the phosphorylated RPB1 band (Figure 4.12 C). No contaminants were observed in the analysed SEC fractions compared to the input (Figure 4.12 C). Fractions 15 and 16 contained faint bands migrating at the size of the CDK7 kinase module subunits used in the phosphorylation reaction (Figure 4.12 C). For downstream cryo-EM applications, fractions 6 to 9 that contained pure pS5 Pol II were pooled and concentrated.

The quality of the purified protein was finally assessed through negative staining and cryo-EM (Figure 4.12 D, E). Negative staining micrographs revealed evenly distributed, spherical Pol II particles across the grid (Figure 4.12 D). Cryo-EM imaging of the sample also confirmed a homogeneous distribution of Pol II

particles after a freeze-thaw cycle, indicating that the sample was of sufficient quality for downstream cryo-EM experiments.

Structural analysis of a native Pol II-FluPol complex

After the generation and quality control of the purified native pS5 Pol II, the formation of a complex between FluPol and Pol II was explored using cryo-EM. As previous research, which employed a short synthetic pS5 CTD_{4x}, mainly identified a bipartite CTD binding site, we were interested in how native Pol II CTD binding sites compare to those described in the literature and whether in the context of a full-length CTD, more binding sites could be found. Furthermore, a Pol II-FluPol complex could provide details on how CTD binding directs FluPol conformation, given that FluPol adopts distinct conformations depending on the functions it has to fulfil [40, 44, 102].

For cryo-EM analysis, pS5 Pol II was mixed with a 1.3 molar excess of pre-annealed transcription bubble scaffold consisting of a DNA-RNA hybrid and non-template DNA [172]. Pol II was then mixed with vRNA promoter-bound 1918 FluPol A D108A endonuclease mutant at a molar excess ranging from 0.5 to 3 times and equimolar DSIF complex and incubated on ice to form a complex. As DSIF has been identified to form interactions with FluPol and is required in promoter-proximal pausing of Pol II, we decided to include DSIF in the complex [179, 196, 261]. To express DSIF, *E. coli* transformed to express DSIF were cultured, followed by purification using Ni-NTA beads and SEC (described later in detail in chapter 6.2.3). The sample was then applied on cryo-EM grids, followed by sample screening. A 12,000 movie dataset was collected with a K2 camera on a Krios microscope. Using cryoSPARC, 2D classification was carried out, and classes of interest were chosen to serve as templates for particle picking on the dataset. In total, 180,000 particles were picked, selecting particles whose size and spherical, compact core corresponded to previously observed Pol II particles in the micrographs (Figure 4.13 A). A subset of the 2D classes revealed a small

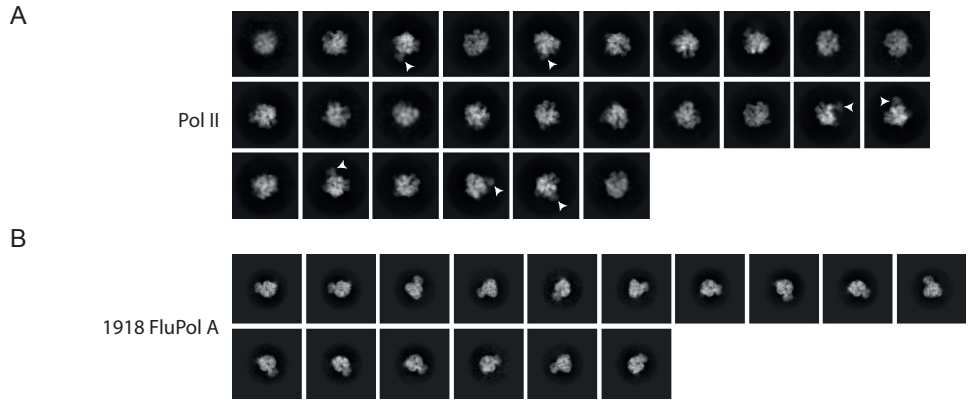

Figure 4.13: Cryo-EM classification of a 1918 FluPol A-Pol II complex. Native pS5 Pol II at a final concentration of 0.2 mg/ml was mixed with a 1.3x molar excess of pre-annealed transcription bubble scaffold composed of a DNA-RNA hybrid and non-template DNA. Pol II was then mixed with vRNA promoter-bound 1918 FluPol A D108A endonuclease mutant and DSIF complex and incubated on ice to form a complex. The sample was then applied on cryo-EM grids, followed by sample screening. A 12,000 movie dataset was collected with a K2 camera on a KRIOS microscrope (ThermoFisher Scientific). **(A)** Selected 2D classes of Pol II extracted from 378,000 particles. White arrows indicate protrusions observed on the Pol II core. **(B)** Selected 2D classes of FluPol from 126,000 particles. Data for this figure was provided by Dr Jeremy Keown.

protrusion on the particle (Figure 4.13 A). A 2D classification was also performed by selecting particles corresponding to FluPol on the micrographs, and 2D class averages were generated from a total of 126,000 particles (Figure 4.13 A). The generated 2D classes of FluPol particles exhibited an elongated shape with three observable lobes. In a few 2D classes, a blurred protrusion can be seen as part of the elongated, spherical shape (Figure 4.13 B).

Based on the obtained 2D class averages, we generated an initial 3D model followed by further processing, including heterogeneous refinement and non-uniform refinement algorithms, which allowed us to derive an electron density map of the respective Pol II and FluPol particles (Figure 4.14 A, B). The FluPol consensus map was refined to a resolution of 4 Å, and the Pol II model yielded a consensus map with a resolution of 3.6 Å. We could observe the polymerase core in a transcriptase conformation by fitting the FluPol density map to a FluPol A CTD-bound structure (PDB ID 8R3L) (Figure 4.14 A). The fitting of the Pol II map to

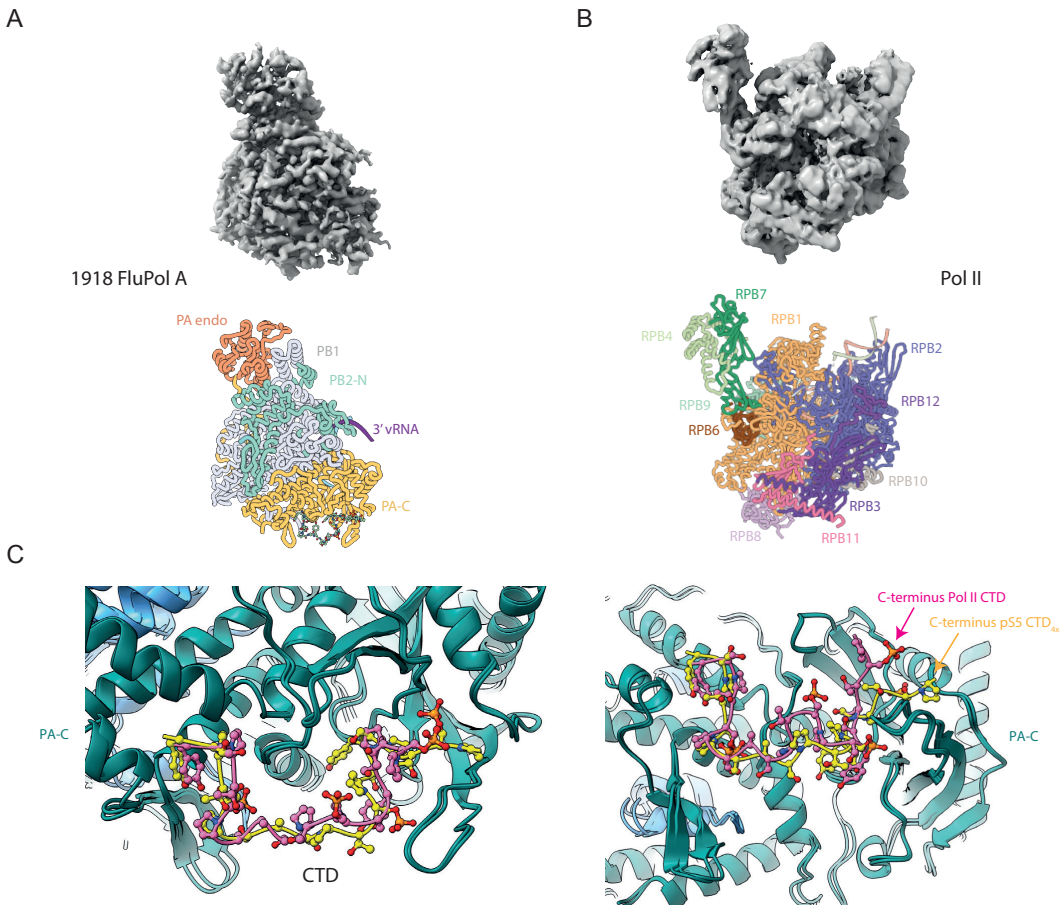

Figure 4.14: Cryo-EM analysis of a 1918 FluPol A-Pol II complex. Cryo-EM maps were generated from a dataset on native pS5 Pol II and 1918 FluPol A D108A. **(A)** Cryo-EM map of FluPol (4.0 Å, top panel) and a structure of the FluPol core in a transcriptase conformation with an ordered endonuclease domain and a resolved Pol II CTD in the PA-C domain that occupies the same path as a synthetic pS5 CTD_{4x} (bottom panel). **(B)** Cryo-EM map of Pol II (3.6 Å, top panel) and Pol II structure with coloured subunits (bottom panel). **(C)** Close-up view of native Pol II CTD (purple) and synthetic pS5 CTD_{4x} mimic peptide (yellow) in complex with 1918 FluPol A. The backbones of both the Pol II CTD and the synthetic peptide show an overall overlap on PA-C. However, the C-terminus of the endogenous peptide occupies a slightly different binding site in the PA-C domain. Data for this figure was provided by Dr Jeremy Keown.

PDB 5OIK revealed a density for the nucleic acid and a globally overlapping Pol II structure (Figure 4.14 A). No density was observed for DSIF in the analysed maps.

Next, we investigated whether the native Pol II CTD binds to FluPol similarly to the synthetic peptide and how the binding sites might differ. The native Pol II CTD-1918 FluPol A structure was overlaid with the 1918 FluPol A bound to the synthetic pS5 CTD peptide$_{4x}$ (PDB ID 8R3L). The backbones of the Pol II CTD and the synthetic peptide showed an overall overlap on the PA-C domain of FluPol (Figure 4.14 C). However, the C-terminal region of the native Pol II CTD occupied a subtly distinct binding site within the PA-C domain, differing from the positioning of the synthetic peptide.

To further explore the interaction between native Pol II, DSIF, and 1918 FluPol A, we used a crosslinking approach to stabilise Pol II and DSIF following Vos et al. (2018) [172]. For crosslinking of the FluPol-Pol II sample, pS5 Pol II, DSIF, and transcription bubble scaffold were incubated and crosslinked with glutaraldehyde, followed by quenching and dialysis overnight. The crosslinked Pol II sample was then added to vRNA-bound 1918 FluPol D108A and incubated on ice before grids were prepared. Electron micrographs of the sample indicated a heterogenous distribution of differently sized particles and the presence of some ice crystals, visible as dark high-contrast particles (Figure 4.15 A). An 11,000 movie data set was collected with a K2 camera on a Krios microscope (eBIC, Diamond Light source). Processing of the dataset generated initial 2D classes, which were selected for complete template picking. The dataset generated 2D classes from 440,000 particles with a characteristic shape indicative of monomeric FluPol particles (Figure 4.15 B). For Pol II, 2D classes were generated from 195,000 particles (Figure 4.15 C). Pol II particles were observed that exhibited a compact spherical shape in the 2D classification, similar to Pol II particles observed in the previous dataset (Figure 4.13 A, Figure 4.15 C). However, compared to the previous dataset, we observed a lack of diversity in the identified 2D

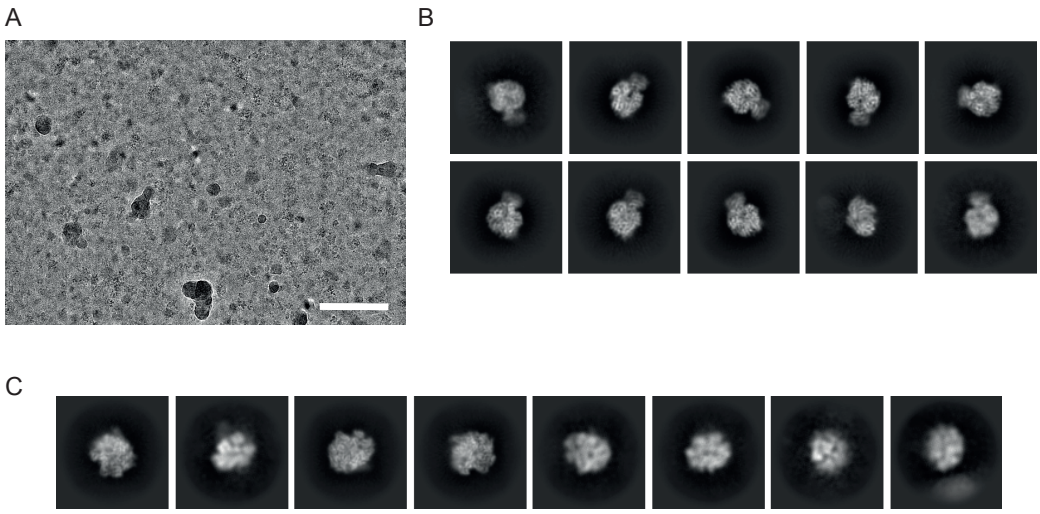

Figure 4.15: Cryo-EM data collection on cross-linked FluPol-Pol II complex. Data acquisition was conducted on a 300 kV Titan Krios G3i microscope (Thermo Scientific) with a K2 Summit camera (Gatan). **(A)** Representative micrograph of the FluPol-Pol II sample screening. Scale bar: 100 nm. **(B)** Processed 2D classes generated from 440,000 particles showing FluPol. **(C)** Processed 2D classes generated from 195,000 particles depicting Pol II.

class orientations. Further data processing confirmed the preferential particle orientation, did not allow the generation of a high-quality 3D model, and resulted in lack of a high-resolution cryo-EM map.

In conclusion, a native Pol II CTD-bound FluPol A structure could be determined to a resolution of 4.0 Å with the CTD binding site located in the PA-C domain and following an almost identical path as a pS5 synthetic CTD_{4x} peptide bound to FluPol A. Additionally, further interactions of FluPol with Pol II were not observed in the dataset and more optimisation is required to obtain an active cap-snatching complex.

4.3 Discussion

The work presented here aimed to study the interaction of the Pol II CTD with the viral polymerase using long, recombinantly expressed CTD mimetic peptides and endogenous mammalian Pol II. In this chapter, I demonstrated the expression of phosphomimetic S5E CTD_{26x} and the generation of phosphorylated $pCTD_{26x}$ peptide using a CDK7 kinase module. The mobility shift observed in the SDS-PAGE analysis of S5E and $pCTD_{26x}$ is likely due to the negative charge of Glu and phosphate residues (Figure 4.1, Figure 4.3). Since the S5E CTD_{26x} contains an S5E mutation in all 26 heptad repeats, a significant mobility shift was observed compared to the unphosphorylated WT construct. The S5E mutation contributes to a higher negative charge, altering migration behaviour on the SDS-PAGE gel. The $pCTD_{26x}$ is likely not phosphorylated at every Ser5 residue of the 26 heptad repeats, given that a more minor mobility shift was observed than the unphosphorylated CTD_{26x} construct. Indeed, *in vivo*, the CDK kinase family does not heavily phosphorylate the CTD with approximately one-third of the CTD repeats estimated to be phosphorylated, and most repeats containing monophosphorylations [262].

CDK7 phosphorylates Ser5 residues and was employed to batch phosphorylate CTD_{26x} and endogenous mammalian Pol II [258]. The *in vitro* batch phosphorylation of CTD_{26x} indicates that the employed CDK7 kinase module modifies the unphosphorylated CTD_{26x}, which resulted in a mobility shift (Figure 4.3 B). The S5E CTD_{26x} did not exhibit a visible mobility shift in SDS-PAGE analysis when incubated with CDK7 kinase and ATP (Figure 4.3 A). However, this observation would require further analysis since mobility shift changes are not linear across SDS-PAGE gels, with mobility shift differences being less prominent at higher molecular weights. The potential for off-target phosphorylation of the S5E CTD_{26x} construct by CDK7 could be investigated through Western blot analysis using antibodies specific to the phosphorylated Ser5 and Ser7 residues. This would confirm whether CDK7 does not modify the S5E CTD_{26x} substrate as expected.

Overall, the observed mobility shifts of S5E CTD_{26x} and $pCTD_{26x}$ are consistent with data from our collaborators (data not shown) and previous studies [223].

After establishing the generation of the phosphomimetic and the phosphorylated CTD_{26x} peptides, their interaction with various FluPol types was investigated. Interaction analysis of FluPol A, FluPol B, and FluPol C with the phosphomimetic S5E CTD_{26x} revealed an interaction with FluPol C in both pulldown assays and SEC analyses (Figure 4.6 A, Figure 4.7). In contrast, FluPol A and FluPol B did not bind to the S5E CTD_{26x} peptide in pulldown assays (Figure 4.4, Figure 4.5). The mode of FluPol binding to the Pol II CTD appears to have diverged among different influenza virus types, with distinct CTD binding interfaces [40, 41]. Specifically, the CTD binding sites of FluPol C differ from those in FluPol A and FluPol B [26, 97]. Amino acid sequence alignments indicate that critical residues in the CTD binding sites of FluPol C are not conserved in FluPol A or FluPol B, despite partial overlaps between the identified CTD binding regions [41, 97]. For example, site 2C, located in a similar region to site 2A, shows no conservation of CTD binding residues, suggesting that different CTD binding modes have evolved in different influenza virus types [40, 41].

It is possible that the binding sites of FluPol A and FluPol B might not accommodate glutamates replacing pS5 residues in the S5E CTD_{26x} peptide, leading to the observed lack of interaction and indicating that the phosphomimetic S5E CTD_{26x} peptide is likely a poor mimic for a pS5-modified CTD. In contrast, the binding sites of FluPol C might either accommodate the phosphomimetic S5E CTD_{26x} peptide, or the peptide might bind to other as yet unidentified CTD binding interfaces where the presence of S5E mutations is tolerated. These hypotheses require structural characterisation to determine the molecular details of the S5E CTD_{26x}-FluPol interaction. It is also important to note that phosphomimicking mutations may not accurately replicate the impact of phosphorylation on protein structure and function [257]. Therefore, phosphomimicking protein variants

should be validated and compared to genuinely phosphorylated controls [257].

After establishing the phosphorylation and purification of the $pCTD_{26x}$ peptide, I studied its interaction in FluPol derived from influenza virus types A, B, and C. In contrast to the interaction experiments with the S5E CTD_{26x} peptide, an interaction of $pCTD_{26x}$ peptide was observed with FluPol A, FluPol B, and FluPol C in both pulldown assays and SEC analyses (see section 4.2.3 and 4.2.4). The N-terminal GST-tag of $pCTD_{26x}$ does not interfere with FluPol CTD binding or cause unspecific off-target interactions in pulldown assays (see section 4.2.3). Furthermore, the interaction with $pCTD_{26x}$ appears stronger than with the phosphomimetic peptide, given that only FluPol C revealed an interaction with S5E CTD_{26x}. As the pulldown experiments had to be conducted in medium salt conditions (300 mM NaCl) due to the aggregation of FluPol on the agarose beads, it can be expected that lower salt concentrations, which better mimic physiological conditions, may result in a stronger interaction between FluPol and both the phosphomimetic and phosphorylated forms of CTD_{26x}. To further characterise the interaction between the recombinant CTD_{26x} peptide and the different FluPol variants, the specific binding affinities for both the phosphomimetic and phosphorylated forms of the CTD_{26x} could be measured using quantitative techniques such as isothermal titration calorimetry or surface plasmon resonance spectroscopy. Overall, the observed interaction of FluPol with $pCTD_{26x}$ aligns with synthetic pS5 CTD_{4x} peptide binding data with FluPol from IAV, IBV, and ICV [96, 97]. Moreover, these pulldown experiments provide evidence for the interaction between FluPol and modified CTD_{26x} peptides, allowing downstream structural analyses.

SEC analyses were carried out to examine further complex formation between the three FluPol types (A, B, and C) and $pCTD_{26x}$, since the phosphomimetic S5E CTD_{26x} only interacted with FluPol C (see section 4.2.3 and 4.2.4). SEC analyses indicated an interaction of FluPol with $pCTD_{26x}$, confirming the observations

from pulldown experiments and previous interaction studies employing synthetic CTD_{4x} peptides [96, 97]. The long, disordered CTD drives phase separation of Pol II during transcription, and it is hypothesised that the long, disordered nature of the CTD promotes the formation of CTD-anchored "viral factories" [74, 252]. The CTD would act as a platform for vRNPs to bind to the CTD and recruit factors involved in transcription and replication [74]. Interestingly, the SEC analysis of 1918 FluPol A with $pCTD_{26x}$ showed that upon $pCTD_{26x}$ binding, the SEC elution profile exhibited a shift to smaller retention volumes and an increase of the FluPol A dimer peak signal (Figure 4.8). FluPol C also exhibited a significant peak shift at a retention volume corresponding to two FluPol C molecules interacting with the long $pCTD_{26x}$ (Figure 4.10). In contrast, FluPol B exhibited the smallest peak shift, which indicates that FluPol B likely does not bind to $pCTD_{26x}$ as strongly as FluPol A or FluPol C, an observation in line with previous interaction data of phosphorylated synthetic CTD mimetic peptides with FluPol B [261]. It can be hypothesised that $pCTD_{26x}$ might drive either FluPol oligomerisation, particularly dimerisation, or it could recruit multiple FluPol molecules to form a "beads on a string" structure [74]. This hypothesis would need to be confirmed further using more quantitative methods, such as mass photometry, to determine the precise molecular weights of the formed complexes.

Biochemical and structural investigations have demonstrated the ability of FluPol to adopt diverse conformations and assemble into oligomeric complexes [122, 145, 263]. However, the functional relevance of the available oligomeric structures has remained unclear [264]. Furthermore, biochemical analyses revealed a FluPol A subcomplex consisting of the PA, PB1, and N-terminal regions of PB2 [263]. This subcomplex was found to exist in a dimeric state in solution, and it could further assemble into a tetrameric conformation by adding viral RNA promoters [263]. This observation is in accordance with the SEC and cryo-EM analysis of the vRNA-bound FluPol A-$pCTD_{26x}$ complex (Figure 4.8). The tetrameric FluPol A structure exhibited a distinctive cage-like arrangement, and the residues

involved in this oligomeric state formation were crucial for viral viability [263]. Interestingly, the tetrameric FluPol conformation was consistently observed across diverse influenza A virus strains, including A/Brevig Mission/1/1918, A/PR/8/34, A/Victoria/3/1975, and A/goose/Guangdong/1/96 [263]. Similarly, in our structural analysis of a FluPol A-pCTD$_{26x}$ we observed a dominance of FluPol A tetramers and dimers (Figure 4.11). Interestingly, split-luciferase-based experiments indicate that overexpression of the CTD induces FluPol oligomerisation. This observed oligomerisation is distinct from symmetric FluPol dimers as a mutant which abolishes symmetric dimerisation still produces an oligomerisation signal in split-luciferase assays [74]. Krischuns et al. (2024) suggest that the CTD can facilitate the association of multiple FluPols nearby, enhancing replication in cooperation with ANP32 [74]. Therefore, it can be hypothesised that the observed dimers in both the SEC and cryo-EM analysis of FluPol A with pCTD$_{26x}$ might form a dimer, distinct from the symmetric dimer state [44, 74]. FluPol B and FluPol C that also formed complexes with pCTD$_{26x}$ might form similar oligomeric states, which need to be further investigated. In summary, pCTD$_{26x}$ interacts with FluPol and drives its oligomerisation.

Using long recombinant CTD$_{26x}$ mimetic peptides instead of synthetic CTD$_{4x}$ peptides, we aimed to structurally characterise the FluPol-CTD interaction using cryo-EM, which proved technically challenging. Cryo-EM data analysis of FluPol A bound to pCTD$_{26x}$ showed a dominance of FluPol tetramers in 2D classes, which is in line with the SEC analyses of pCTD$_{26x}$ with FluPol A (Figure 4.8 C). The presence of tetrameters hampered the downstream processing of the dataset due to preferential orientation problems causing a non-random particle distribution which ultimately did not produce a cryo-EM map with sufficiently high resolution to solve a structure (Figure 4.11 B). Sample analysis of pCTD$_{26x}$ with FluPol C, as well as S5E CTD$_{26x}$ with FluPol C (data not shown), resulted in a majority of 2D classes corresponding to noise, which did not allow further data analysis due to the low number of picked particles (Figure 4.11 C). In addition, during

sample screening, the formation of aggregates was frequently observed, which could be caused by the large number of heptad repeats and the disordered nature of the CTD. Multiple FluPol molecules could recognise pS5 motifs on different repeats of a single CTD peptide, resulting in a non-uniform particle distribution during sample preparation and a high background during data screening and collection. Therefore, future experiments could employ a shorter recombinant CTD_{13x} that was successfully purified during this project but was not further employed due to time constraints (data not shown). Another cause could be that the formed complex might not have remained stable during sample preparation. For instance, damaged particles and protein aggregates can form in very thin ice layers [265]. The thin sample distribution, essential for cryo-EM data collection, can promote protein denaturation at the air-water interface, resulting in different sample behaviour in aqueous solutions [265, 266]. Various sample optimisation steps could be undertaken to optimise the sample preparation of FluPol with $pCTD_{26x}$. For instance, optimising using detergents and other additives such as glycerol can prevent proteins from adsorption to the air-water interface and aggregation [265, 267]. Other techniques, such as affinity grid immobilisation and axisymmetric draining, could be employed in future studies to optimise structural characterisation of the FluPol-$pCTD_{26x}$ interaction [265]. GST-CTD constructs with up to 53 heptad repeats have been routinely used in functional studies as substrates for CTD kinase assays to establish the role of the various kinases in the Pol II transcription cycle [268–270]. However, to our knowledge, no studies have been reported that have employed recombinant CTD constructs in structural studies. Despite various efforts to improve the quality of the grids, we could not solve a FluPol structure bound to a recombinant CTD_{26x} construct.

Native Pol II purified from mammals provided an opportunity to study the interaction with FluPol and the potential to identify interactions beyond the described binding sites with the CTD. FluPol A bound to vRNA and the native Pol II CTD adopted a transcriptase conformation, overlapping with our

1918 FluPol A structure bound to synthetic CTD_{4x} (Figure 4.14 A) [112]. A density for DSIF could not be observed in our structure which is likely caused by the insufficient quality of the protein preparation which was not stable during freeze-thaw cycles (see chapter 6). Furthermore, we mainly observed overlapping peptide paths for the synthetic and the native Pol II CTD in the PA-C domain (Figure 4.14 C). However, the C-terminal end of the resolved native Pol II CTD in the FluPol binding site occupied a subtly distinct path, potentially caused by different binding dynamics of a long 52-repeat peptide in comparison to a synthetic 4-repeat peptide (Figure 4.14 C). It has been demonstrated that each protein factor has its minimal requirement for the functional CTD peptide unit, ranging from as few as four residues to as many as three repeats [271]. Extensive structural analyses have revealed various interactions involving the CTD and diverse protein complexes which are stabilised by steric constraints along with van der Waals forces, intramolecular and intermolecular hydrogen bonds, and electrostatic and stacking interactions [271]. Furthermore, various factors determine the specificity of CTD-protein interactions, including the length of bound CTD peptides and the cis or trans conformations of the phosphoserine-proline peptidyl-prolyl bonds [271]. Overall, these factors might contribute to the slightly different binding observed in the C-terminal region of the native Pol II CTD peptide path observed in our structure.

Recently, functional studies have provided evidence that FluPol forms interactions with other Pol II subunits beyond its interaction with the CTD. However, the lack of a structure of a cap-snatching complex has not allowed us to uncover the molecular details underlying these interactions [26, 237]. The large multimeric Pol II has been shown to interact with the PB2-N domain via its RBP4 subunit in functional experiments, and further auxiliary interactions between PB2-C and the RPB5 and RPB11 subunits have been suggested to stabilise a FluPol-Pol II complex further [237]. Here, we attempted cryo-EM analysis of a FluPol A-Pol II sample using a crosslinking approach to stabilise Pol II during cryo-EM

sample preparation [172]. We observed the presence of both proteins in separate 2D classes but could not identify 2D classes that included both FluPol and Pol II (Figure 4.15). The technical improvements suggested above for the cryo-EM analysis of the FluPol-pCTD$_{26x}$ could also address the dataset's preferential orientation problems that we encountered. In addition, varying scanning angles during data collection could prevent the generation of preferentially oriented 2D classes in future studies. More importantly, given previous unsuccessful attempts to generate a FluPol-Pol II structure from another research group, additional accessory factors might be required to stabilise a cap-snatching complex [40, 261]. The DSIF complex, particularly its Spt5 subunit, has been identified as a key interactor of FluPol and is required in promoter-proximal pausing of Pol II [196]. Furthermore, the capping enzyme is likely involved in this interaction as it binds to the Spt5 CTR and the pS5-modified Pol II CTD [272, 273]. Recent structural studies have revealed important insights into the co-transcriptional capping mechanisms during the Pol II transcription cycle [105, 274]. These findings offer a crucial understanding of the temporal and spatial coordination governing the early phases of Pol II transcription, which can guide future investigations into how the influenza virus exploits this process [105, 274]. Particularly, elucidating structural details of viral transcription using vRNP systems could provide crucial details of viral transcription and new avenues to validate the hypothesis of viral factories present on the CTD [74, 275]. Overall, the interaction data of FluPol with the long recombinant CTD peptides and native Pol II provides a new foundation for the structural and functional characterisation of the FluPol-Pol II interaction.

This chapter presented the generation of long, recombinant Pol II CTD mimetic peptides and their use in studying the interaction of FluPol with Pol II *in vitro*. The phosphomimetic peptide interacted with only FluPol C, whereas pCTD$_{26x}$ was shown to bind to FluPol from IAV, IBV, and ICV. The structural study of these complexes was attempted and will require further optimisation in future studies. Our data also presents structural details between the native Pol II CTD

and FluPol A, revealing that the native CTD also binds to PA-C and follows a mostly overlapping binding path, similar to synthetic pS5 CTD_{4x}. In conclusion, the data provides a basis for further structural and functional characterisation of FluPol cap-snatching and its interaction with Pol II.

5

Functional characterisation of Pol II CTD binding to the influenza virus polymerase

Parts of this chapter were published in:

Jeremy Keown*, **Alaa Baazaoui***, Marek Šebesta, Richard Štefl, Loic Carrique, Ervin Fodor, Jonathan Grimes (2024) *Structural and functional characterisation of the interaction between the influenza A virus RNA polymerase and the CTD of host RNA Polymerase II. Journal of Virology* doi: 10.1128/jvi.00138-24

* Joint first authorship

5.1 Introduction

The direct interaction between the viral polymerase and the pS5 CTD of Pol II plays a pivotal role in cap-snatching, a mechanism required for viral mRNA synthesis [41]. Specifically, pS5 modification within the CTD heptad repeat is a hallmark of Pol II mRNA capping during transcription initiation, when the host RNA cap is synthesised [40, 41, 151]. Previous functional and structural studies have demonstrated that FluPol selectively recognises the pS5-modified CTD, promoting viral transcription initiation by facilitating cap-snatching *in vitro* [41]. Structural studies suggest that pS5 CTD binding induces conformational rearrangements in FluPol, shifting it into a 'transcriptionally active' conformation, which enhances host mRNA cleavage through its PA endonuclease domain [97].

The Pol II CTD exhibits remarkable functional flexibility as many of the amino acids in the heptapeptide can undergo various posttranslational modifications (PTMs), and combinations of these covalent marks are selectively recognised by different protein partners [153, 158, 165, 276]. Among these amino acids composing a heptad repeat, serines at positions 2, 5, and 7 (abbreviated as S2, S5, and S7, respectively) along with tyrosine at position 1 (Y1) and the threonine at position 4 (T4) can be phosphorylated and are important for Pol II transcription [153]. Notably, mass spectrometry analyses demonstrated that S2 and S5 phosphorylations are the most prominent PTMs during the transcription cycle, while recent studies have linked S7 phosphorylation to early events of transcription initiation [262, 277, 278].

Although the Pol II CTD undergoes various post-translational modifications (PTMs) throughout the transcription cycle, the impact of these modifications on the spatiotemporal regulation of the Pol II transcription cycle is still under investigation [153]. In addition, combinations of these main phosphorylation patterns and the effect of pS7 remain poorly understood [112]. To date, functional studies involving viral polymerase have mainly utilised synthetic CTD mimic peptides

with monophosphorylated heptad repeats, limited to a maximum total length of six repeats [26, 74, 97, 112, 261]. These experiments comparing unphosphorylated CTD to pS2 or pS5 CTD peptides have shown that the viral polymerase has a strong preference for the pS5 CTD over pS2 and unphosphorylated CTD [126]. Additionally, increasing the number of identical repeats enhanced overall CTD binding affinity, indicating an avidity effect [96, 97]. In the context of influenza virus transcription, only pS2 and pS5 modifications have been studied utilising simplified CTD models [40]. In this chapter, I aim to address how various phosphorylation combinations of the three Ser residues — S2, S5, and S7 — in the CTD heptad repeat consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ impact the FluPol-CTD interaction. Furthermore, this work investigates the role of the highly flexible PB2 domain in mediating CTD binding through interaction experiments since PB2 undergoes significant movements in various FluPol conformations. Recent evidence from our lab suggests that the PB2 domain may directly contact the N-terminus of the Pol II CTD in the transcriptase conformation [112].

In summary, this chapter explored the interaction of FluPol from IAV, IBV, and ICV with various Ser-phosphorylated CTD mimic peptides and it presents functional assay data using these peptides in comparison to monophosphorylated synthetic CTD_{4x} peptides and recombinant CTD_{26x} peptides. Additionally, the binding of the FluPol A PB2-C domain to Pol II CTD is examined through interaction experiments.

5.2 Results

5.2.1 The influenza virus polymerase strongly recognises pS5 Pol II CTD over alternative phosphorylation patterns

The influenza virus polymerase specifically recognises pS5 CTD over pS2 and unphosphorylated CTD [126]. Combinations of pS2 and pS5 phosphorylation patterns, as well as the role of the pS7, have not yet been investigated in the context of the viral polymerase [112]. To further characterise the binding preference of the influenza virus polymerase to Pol II CTD, I used biotinylated Pol II CTD mimic peptides containing four repeats (designated repeats a–d) of the conserved heptapeptide repeat ($Y_1S_2P_3T_4S_5P_6S_7$) with a combination of phosphorylation patterns (Figure 5.1 A). Peptides with four heptapeptide repeats were used in these binding assays as it has been demonstrated previously that these CTD mimic peptides can bind the influenza virus polymerase specifically [96, 97, 126]. The interaction of this CTD peptide panel was tested with 1918 FluPol A, FluPol B, and FluPol C (Figure 5.1).

Streptavidin-coated beads were bound to the biotinylated peptides before their incubation with FluPol. Eluate samples were collected, followed by SDS-PAGE analysis and silver staining to visualise protein bands. The band intensities were quantified to measure the relative binding of FluPol to the tested CTD peptides. The CTD binding assay with 1918 FluPol A and the CTD peptide panel showed that 1918 FluPol A bound to the pS5 CTD peptide, and low but above background levels of binding to pS7 were detected (Figure 5.1 B, C). Interestingly, none of the other peptides were bound by the 1918 FluPol A above background levels (Figure 5.1 B, C). A similar observation was made for FluPol C, which bound strongly to pS5 CTD and exhibited above background levels of binding to a double-phosphorylated pS5 pS7 CTD peptide (Figure 5.1 E). In contrast, experiments with FluPol B showed higher backgrounds and a less striking binding preference for pS5 CTD. Low binding was also observed with alternately

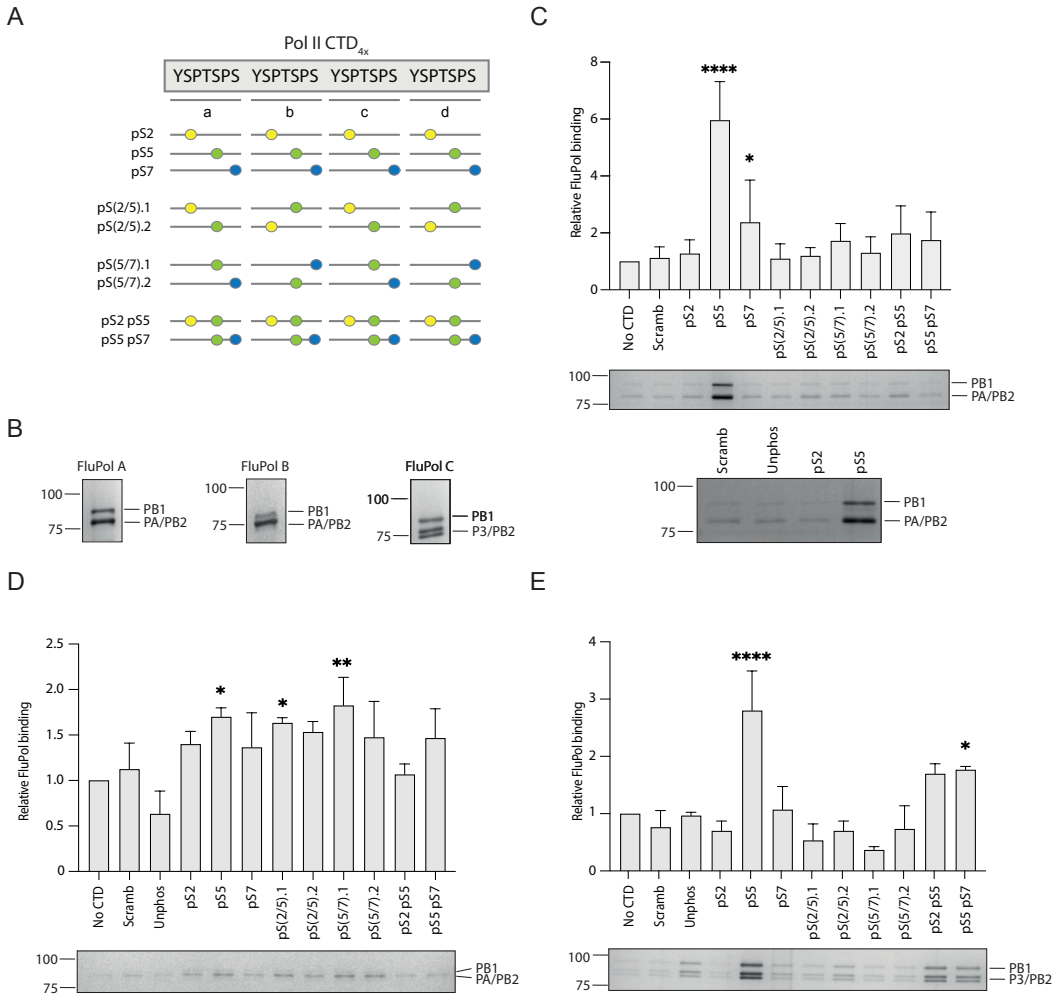

Figure 5.1: Effect of serine phosphorylations of Pol II CTD on binding the influenza virus polymerase. **(A)** Schematic of peptides used in the binding assay with the four heptad repeats (designated repeats a, b, c, and d) depicted as grey lines. Serine phosphorylations (pS) are shown as coloured circles (pS2 (yellow), pS5 (green), pS7 (blue)). **(B)** Input samples of 1918 FluPol A, FluPol B, and FluPol C used in the Pol II CTD binding assays. **(C)** Top panel, quantification of 1918 FluPol A binding to Pol II CTD mimic peptides. Middle panel, a representative silver-stained gel. Bottom panel, controls for the Pol II CTD binding assay, including a scrambled (Scramb), unphosphorylated (Unphos), serine 2 phosphorylated (pS2), and serine 5 phosphorylated (pS5) Pol II CTD mimic peptides. **(D, E)** Top panel, binding of FluPol B **(D)** and FluPol C polymerase **(E)** to serine phosphorylated Pol II CTD mimic peptides. Top panel, quantification of FluPol binding. Bottom panel, a representative silver-stained gel. All quantifications are based on three independent binding assays and data are mean ± s.e.m. An ordinary one-way ANOVA test was employed to compare the relative binding of FluPol in the presence and absence of CTD. Significance levels are denoted as *, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$. Parts of the figure were adapted from Keown, Baazaoui et al. (2024) [112].

monophosphorylated pS(2/5).1 CTD and pS(5/7).1 CTD (Figure 5.1 A, B, D). Importantly, the CTD binding assay with FluPol B showed the lowest signal-to-noise ratio compared to the data with 1918 FluPol A and FluPol C, requiring further validation of the interaction of FluPol B with differentially phosphorylated CTD peptides.

Overall, these experiments show that the influenza virus polymerase has a striking preference for pS5 peptide compared to pS2 and pS7, and additional phosphorylation in the pS5 peptide interferes with the binding of 1918 FluPol A but not necessarily for FluPol C. The observed binding preference of FluPol to pS5 CTD clearly links the influenza virus polymerase to the initiating form of Pol II engaged in capping its nascent RNA transcript [194].

5.2.2 CTD binding enhances the transcriptional activity of influenza virus polymerase C *in vitro*

As the transcriptional activity of FluPol C is enhanced through the binding of synthetic pS5 CTD_{4x} peptides, I aimed to test how the binding of long recombinant CTD_{26x} as well as alternatively phosphorylated CTD_{4x} peptides might impact viral transcription [97].

Initially, I tested various conditions to optimise a viral transcription assay that detects viral transcription products by incorporating radiolabelled GTP and subsequent PAGE analysis of the reaction products. For this GTP incorporation assay, purified recombinant 1918 FluPol A and FluPol C expressed in *Sf9* insect cells were used together with purified rabbit globin mRNA as well as 5' and 3' vRNA promoters with a length of 15 nt and 14 nt, respectively (see table 2.2). As controls for the *in vitro* transcription reaction, peptides were replaced either with water or GST. Key components of the assay, such as capped RNA, purified polymerase or the vRNA promoters, were systematically omitted for the control conditions. These optimisation trials resulted in using low salt concentrations

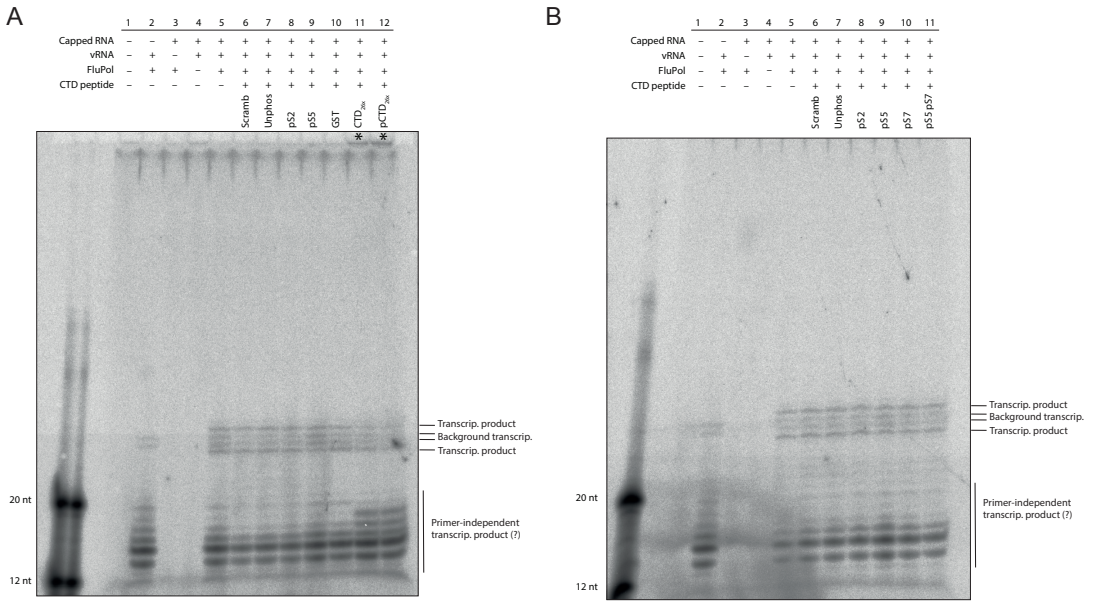

Figure 5.2: Effect of serine phosphorylations of Pol II CTD on the transcription activity of 1918 pandemic H1N1 influenza A virus polymerase. Effect of Pol II CTD mimic peptides on the transcription activity of FluPol A. Products of the GTP incorporation reactions were separated on denaturing PAGE. **(A)** Comparison of synthetic CTD_{4x} mimic peptides (scramb, unphos, pS2, and pS5) with recombinant GST-tagged unphosphorylated and phosphorylated CTD_{26x} peptides (CTD_{26x}, $pCTD_{26x}$). The asterisk (*) marks a high signal detected in the wells. **(B)** Transcription activity of FluPol A in the presence of synthetic CTD_{4x} mimic peptides with serine 7 phosphorylation (pS7) and serine 5/serine 7 double phosphorylation (pS5 pS7) in comparison to pS2 and pS5 CTD_{4x} synthetic peptides (pS2 and pS5).

of 50-150 mM NaCl in downstream experiments and adding CTD peptides in a 10-fold molar excess to FluPol (data not shown).

After optimising the reaction conditions, the transcriptional activities of FluPol A and FluPol C were tested by comparing a selection of monophosphorylated, double-phosphorylated pS5 pS7 CTD_{4x}, and recombinant $pCTD_{26x}$ peptides (Figure 5.2, Figure 5.3). It is expected that a transcription product would have a size of approximately 24 to 29 nt, as the vRNA template has a length of 14 nt and the capped RNA is cleaved to generate a cap ranging from 10-15 nt [41]. In the absence of capped RNA and CTD (well 2), multiple bands were observed, forming a laddering pattern between 12 to 20 nt and two bands could be observed

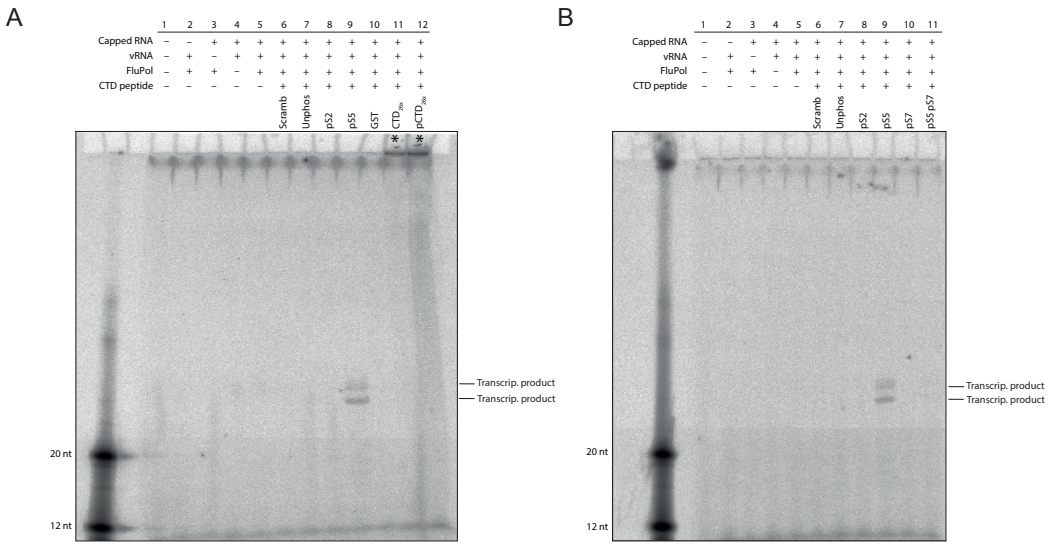

Figure 5.3: Effect of serine phosphorylations of Pol II CTD on the transcription activity of influenza C virus polymerase. Effect of Pol II CTD mimic peptides on the transcription activity of FluPol C. Products of the GTP incorporation reactions were separated on denaturing PAGE. **(A)** Comparison of synthetic CTD_{4x} mimic peptides (scramb, unphos, pS2, and pS5) with recombinant GST-tagged unphosphorylated and phosphorylated CTD_{26x} peptides (CTD_{26x}, $pCTD_{26x}$). The asterisk (*) marks a high signal detected in the wells. **(B)** Transcription activity of FluPol C in the presence of synthetic CTD_{4x} mimic peptides with serine 7 phosphorylation (pS7) and serine 5/serine 7 double phosphorylation (pS5 pS7) in comparison to pS2 and pS5 CTD_{4x} synthetic peptides (pS2 and pS5).

above 20 nt at around 25 nt (Figure 5.2 A, B). Primer-independent transcription products potentially cause the observed laddering pattern. At the same time, the two bands around 25 nt could result from capped RNA priming with capped RNA fragments present in the purified FluPol sample potentially containing remaining RNA fragments from the insect cells (Figure 5.2 A, B). Upon addition of capped RNA, two additional bands appear with a visibly stronger intensity above and below the two bands (wells 5 to 12 in Figure 5.2 A, wells 5-11 in Figure 5.2 B). The addition of pS5 CTD_{4x} as well as the tested recombinant CTD_{26x} and $pCTD_{26x}$ peptides did not show visible differences in band product intensities compared to the control conditions (Figure 5.2 A). Furthermore, two additional strong bands in wells 11 and 12 with a size around 17 to 19 nt could be observed when recombinant CTD_{26x} was added (Figure 5.2 A). Notably, the addition of recombinant CTD_{26x} peptides to the reactions led to high signals

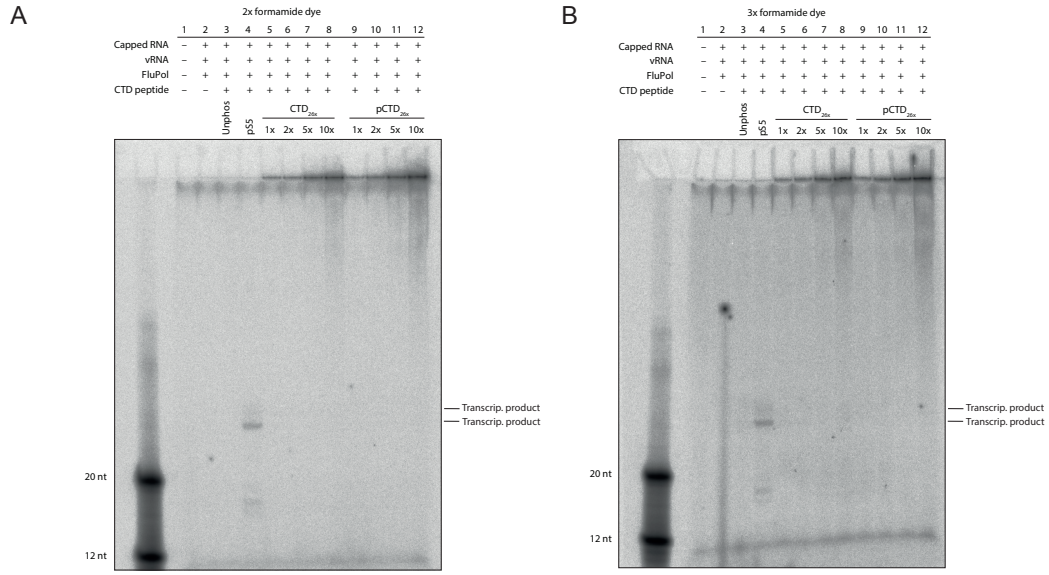

Figure 5.4: Effect of different formamide dye ratios on transcription assay sample mobility of FluPol C with recombinant CTD mimic peptides. Products of the GTP incorporation assay were run on denaturing acrylamide gels. The sample mobility of GTP incorpartion assay reactions of unphosphorylated and pS5 synthetic CTD_{4x} (unphos and pS5) was compared to increasing concentrations of recombinant GST-tagged unphosphorylated and phosphorylated CTD_{26x} peptides (CTD_{26x}, $pCTD_{26x}$). Samples were denatured with 2x **(A)** or 3x formamide dye **(B)**. Increasing peptide concentrations cause higher signals in the PAGE gel wells 5 to 12.

in the wells of the PAGE gel in reactions with both FluPol A and FluPol C, potentially causing lower signals that resulted in the lack of observed signals for recombinant CTD_{26x} peptides (Figure 5.2 A, Figure 5.3 A). Since in the CTD binding assay 1918 FluPol A bound to pS7 CTD_{4x} above background levels, the transcriptional activity of FluPol in the presence of pS7 CTD_{4x} was compared to double-phosphorylated pS5 pS7 CTD_{4x} and pS5 CTD_{4x}. Bands for the conditions containing pS7 and pS5 pS7 CTD_{4x}, respectively, exhibited no visible differences to other test conditions (Figure 5.2 B). Overall, the signal intensities of reactions with 1918 FluPol A comparing CTD_{26x} peptides as well as synthetic pS2, pS5, pS7, and double phosphorylated pS5 pS7 CTD_{4x} peptides resulted in comparable intensities to reactions with scrambled or unphosphorylated CTD_{4x} peptides (Figure 5.2).

The transcriptional activity of FluPol C was also tested using the previously described experimental set-up (Figure 5.3 A). FluPol C produced a strong signal

for pS5 CTD with two main bands migrating around 25 nt and a weak signal around 18 nt (Figure 5.3 A, B). All other tested peptides and control conditions did not produce a detectable transcription product (Figure 5.3 A, B). Furthermore, it was noted that the addition of recombinant CTD_{26x} peptides to the reactions led to high signal in the wells of the PAGE gel, similar to reactions conducted with FluPol A (Figure 5.2 A, Figure 5.3 A). The potential accumulation of reaction product in wells 11 to 12 might cause lower signals, resulting in the lack of observed signals for recombinant CTD_{26x} peptides on the PAGE gel (Figure 5.3 A). When pS7 and pS5 pS7 CTD_{4x} peptides were compared with pS5 CTD_{4x}, only reactions containing pS5 CTD_{4x} resulted in a product band, while control conditions did not produce a signal (Figure 5.3 A). Overall, these experiments show that pS5 CTD_{4x} enhances FluPol C transcriptional activity, while FluPol A does not seem to distinguish between the various peptides under the tested conditions.

The GTP incorporation assays with both FluPol A and FluPol C exhibited a high signal in the wells of the PAGE gels for samples containing recombinant CTD_{26x} that might have negatively impacted the signal intensity of the product bands. For this reason, *in vitro* transcription reactions of FluPol C with a gradient of 1-10x recombinant CTD_{26x} were tested using double and triple the amount of formamide dye to denature the samples (Figure 5.4). While the controls with synthetic unphosphorylated and pS5 CTD_{4x} did not show any product accumulation in the wells, the recombinant CTD_{26x} peptides, irrespective of their concentration or phosphorylation state accumulated in the wells (Figure 5.4).

In summary, the reaction conditions used could replicate previous observations concerning the transcription activity of FluPol in the presence of pS5 CTD_{4x}. Further optimisation, such as a post-reaction protease treatment and subsequent RNA purification, is required to reliably characterise the effect of recombinant CTD_{26x} peptides on viral transcription *in vitro*, which could not be pursued due to time constraints in this project.

5.2.3 The PB2-C domain does not bind to Pol II CTD mimics *in vitro*

The long nature of the mammalian Pol II CTD has led to the hypothesis that it may harbour binding sites at other locations on the FluPol surface. Furthermore, the CTD was proposed to either wrap around the FluPol core or to loop out to serve as a platform that recruits host factors involved in transcription and transcript processing [40, 41]. In our 1918 FluPol A structure bound to synthetic pS5 CTD_{4x}, we observed a continuous stretch of peptide density in the PA-C domain. Interestingly, the PB2-C domain was located close to the N-terminal end of the Pol II CTD peptide [112]. The positive electrostatic potential of the modelled PB2-C surface, located close to the CTD binding site, suggested the potential for a novel CTD binding interface that could interact with the negatively charged pS5 CTD. This proposed binding site would extend the previously characterised CTD binding region on the FluPol A PA-C domain [74, 96, 112].

To assess a potential interaction of the PB2-C domain with the Pol II CTD, I carried out *in vitro* interaction experiments between PB2-C and Pol II CTD mimics. In a pulldown experiment, I assessed the binding of IAV NT60 627-NLS domain (derived from A/Northern Territory/60/1968 (H3N2) virus) and IAV Viet PB2-C (derived from A/Vietnam/1203/2004 (H5N1) virus) with CTD_{26x} and $pCTD_{26x}$ peptides, respectively (Figure 5.5 A, B) [220]. In the pulldown with 627-NLS, strong bands were detected in the flowthrough samples after incubation of 627-NLS with immobilised CTD_{26x} or $pCTD_{26x}$ peptide (Figure 5.5 A). Furthermore, the eluate samples contained a strong band corresponding to the respective recombinant CTD_{26x} peptide and only a faint background band corresponding to 627-NLS around 25 kDa was detected (Figure 5.5 A). In the control conditions, GST tag was incubated together with 627-NLS to test for unspecific binding. Due to the similar molecular weight, the bands corresponding to the GST and 627-NLS could not be distinguished (Figure 5.5 A). However, a comparison of the

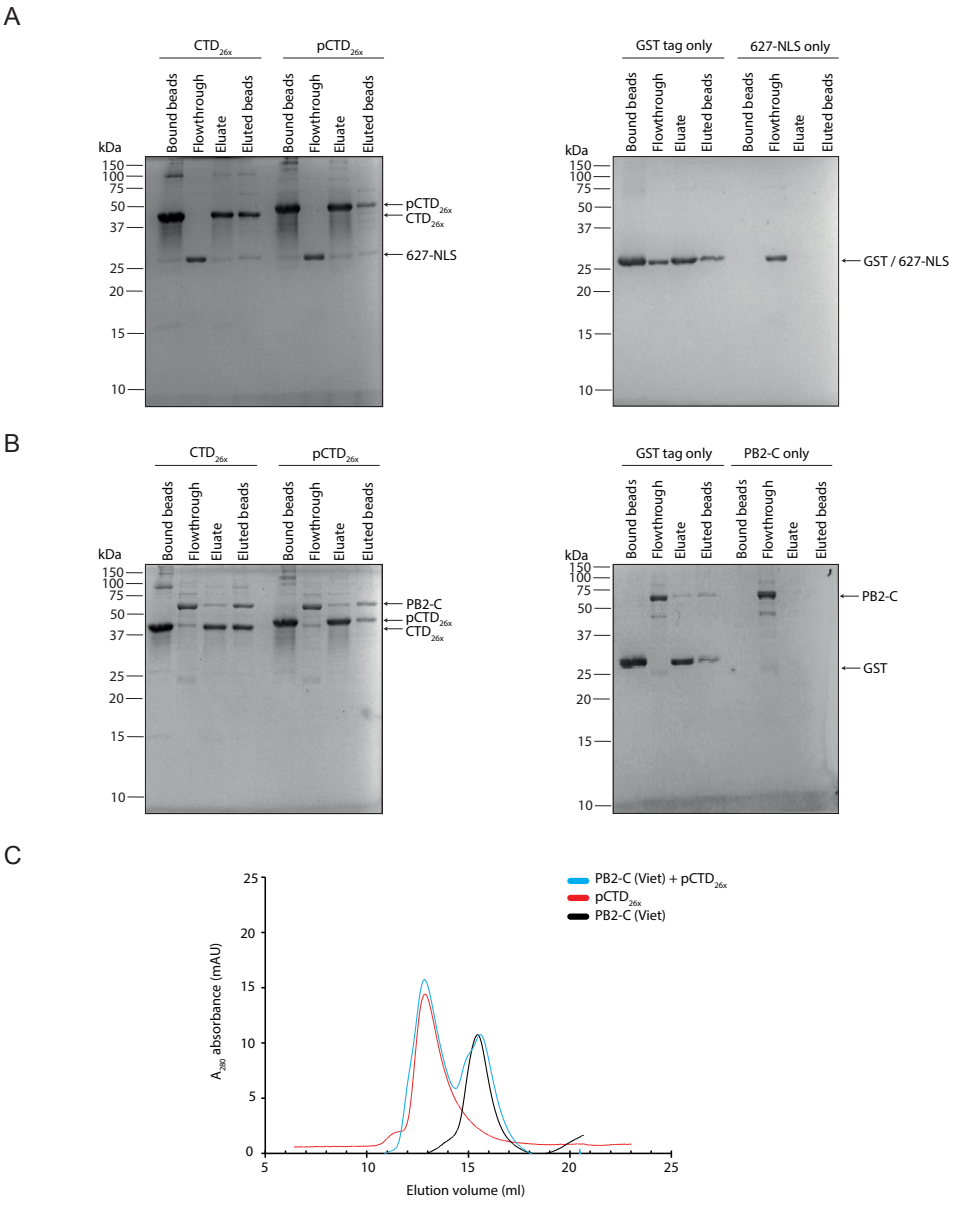

Figure 5.5: Pulldown of 627-NLS (NT60) and PB2-C (Viet) with unphosphorylated and pCTD$_{26x}$ peptide. The interaction of phosphorylated CTD$_{26x}$ peptide with IAV NT60 627-NLS domain (kind gift from Haitian Fan) **(A)** and IAV (Viet) His-tagged PB2-C (kind gift from Alison Rep) **(B)** was compared to the interaction with the unphosphorylated CTD$_{26x}$ peptide (left column). Incubation with the GST tag or the PB2 domain only served as a negative control (right column). Peptides were incubated with GST sepharose beads (bound beads), followed by incubation with vRNA-bound FluPol (flowthrough) and elution of the formed complexes with GSH. Elution samples were taken (eluate) along with the remaining beads (eluted beads). **(C)** SEC analysis of His-tagged PB2-C (Viet) with pCTD$_{26x}$ peptide at 1:1 molar ratio.

band intensities shows a strong band in the flowthrough fraction and no visible change in the band intensities when comparing the bound beads sample and the eluate (Figure 5.5 A). In the other control, where 627-NLS was incubated with beads, no aggregation on the beads was observed in the eluate and post-elution beads sample, indicating that the buffer conditions did not cause any unspecific interactions with the beads (Figure 5.5 A).

In a pulldown experiment and SEC analysis, the interaction of PB2-C was tested with CTD_{26x} and $pCTD_{26x}$ peptides (Figure 5.5 B, C). Similar observations were made with PB2-C as with 627-NLS (Figure 5.5 A). In the pulldown experiment, only background levels of PB2-C around 55 kDa could be detected in eluates of CTD_{26x} and $pCTD_{26x}$ peptides (Figure 5.5 B). Most of the analyte remained in the flowthrough fractions indicated by a strong band corresponding to PB2-C (Figure 5.5 B). Other observed bands around the main PB2-C, CTD_{26x}, and $pCTD_{26x}$ bands are likely minor impurities resulting from the purification process (Figure 5.5 B). The controls with immobilised GST tag or PB2-C alone show no unspecific interactions (Figure 5.5 B). In the eluate fraction with immobilised GST tag, only a faint background band corresponding to PB2-C was observed and no bands were visible in the eluate sample of PB2-C incubated with the beads alone (Figure 5.5 B).

SEC analysis of PB2-C incubated with $pCTD_{26x}$ peptide was carried out to assess whether a complex is formed using a different experimental approach. As controls, SEC traces were recorded for $pCTD_{26x}$ and PB2-C separately (Figure 5.5 C). The SEC trace of $pCTD_{26x}$ alone showed a peak at around 13 ml, while the SEC trace of PB2-C contained a peak at 15.5 ml (Figure 5.5 C). When PB2-C and $pCTD_{26x}$ were incubated together, the resulting SEC trace consisted of two peaks corresponding to the retention volumes of the individual protein components and confirming the lack of complex formation (Figure 5.5 C). In conclusion, pulldowns and a SEC analysis using recombinant CTD_{26x} peptides indicate that

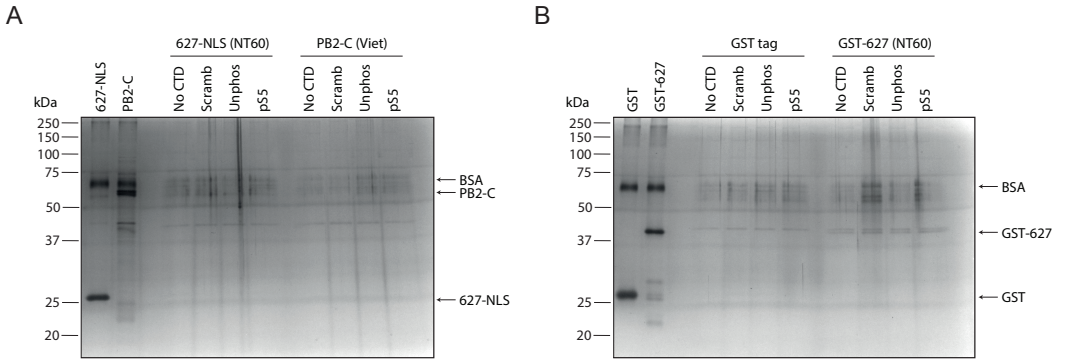

Figure 5.6: Effect of pS5 phosphorylation of Pol II CTD on the binding of IAV (NT60) 627-NLS and IAV (Viet) PB2-C. Streptavidin beads were incubated without peptide (no CTD), or with scrambled CTD_{4x} (scramb), unphosphorylated CTD_{4x} (unphos), and pS5 CTD_{4x} (pS5). **(A)** CTD binding assay with untagged 627-NLS (NT60) and His-tagged PB2-C (Viet). **(B)** CTD binding assay with GST-627 (NT60) and GST tag (control).

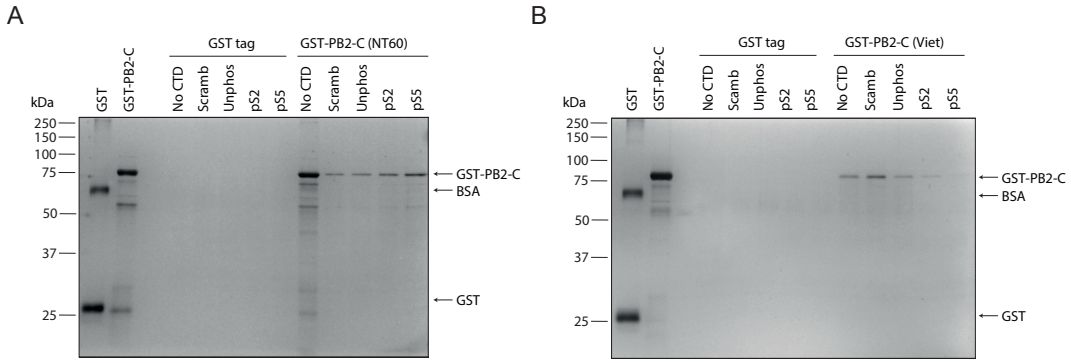

Figure 5.7: Effect of pS5 phosphorylation of Pol II CTD on the binding of IAV PB2-C from Viet and NT60 strains. Streptavidin beads were incubated without peptide (no CTD) or with scrambled CTD_{4x} (scrambled), unphosphorylated CTD_{4x} (unphos), and pS5 CTD_{4x} (pS5). **(A)** CTD binding assay with GST-tagged PB2-C (NT60) (kind gift from Dr Haitian Fan). **(B)** CTD binding assay with GST-tagged PB2-C (Viet) (kind gift from Dr Hana Veler).

no complexes are formed between PB2-C or 627-NLS and the tested peptides.

As synthetic pS5 CTD_{4x} has been shown to interact with FluPol, I used this peptide to assess whether an interaction is formed between pS5 CTD_{4x} and various PB2-C constructs [126]. CTD binding assays of 627-NLS, PB2-C, and GST-627 were conducted with a peptide panel of scrambled, unphosphorylated, pS2 CTD_{4x},

and pS5 CTD_{4x}. Streptavidin beads were coated with the peptides and blocked with BSA before incubation with the analytes. Analysis of 627-NLS (NT60) and PB2-C (Viet) showed that neither of the analytes bound to pS5 CTD_{4x} beyond the background levels observed in the controls (Figure 5.6 A). The same observation was made when GST-627 (NT60) was compared to GST tag alone. Only faint background bands corresponding to GST-627 around 40 kDa were observed across all samples, with all detected bands remaining at background levels (Figure 5.6 B).

In another experiment, GST-tagged PB2-C from two different IAVs, NT60 (H3N2) and Viet (H5N1), were also tested in a CTD binding assay (Figure 5.7). In the control condition with GST tag alone, no background could be observed in either experiment (Figure 5.7 A, B). For GST-PB2-C (NT60), the band signal for pS5 CTD_{4x} appeared stronger compared to pS2 CTD or the control peptides. However, the negative control showed an unusually strong signal for GST-PB2-C, resulting in an inconclusive result (Figure 5.7 A). The CTD binding assay with GST-PB2-C (Viet) did not indicate binding to any of the tested peptides (Figure 5.7 B). Overall, interaction experiments of PB2-C and subdomains of PB2-C, including 627 and 627-NLS, do not seem to show an interaction under the tested conditions.

5.3 Discussion

This chapter explored the influence of varying serine phosphorylation combinations within the Pol II CTD consensus heptad repeats on FluPol binding. The Pol II CTD can be phosphorylated at S2, S5, and S7 as well as on Y1 and T4 of the heptad repeat [153]. Binding assays of 1918 FluPol A to Pol II CTD peptides with diverse phosphorylation patterns revealed a striking preference for pS5 CTD peptides over those phosphorylated at S2 or S7 (Figure 5.1 A-C). Interestingly, combining pS5 with either pS2 or pS7 in different arrangements within the same peptide led to a significant reduction of polymerase binding, suggesting that the viral polymerase interacts with Pol II at a very specific point in the transcription cycle, coinciding with the recruitment of the capping enzyme to Pol II (Figure 5.1 C) [161, 279].

Analysis of our recently published structure that revealed a CTD binding site in the PA-C domain of 1918 FluPol A explains how additional phosphorylation at S2 and S7 would disrupt CTD binding, even for peptides that carry phosphorylation at the most critical S_{5a} and S_{5c} positions (Figure 3.1) [112]. While phosphorylation at S_{2a} and S_{2b} would likely be well accommodated due to their solvent-exposed position, phosphorylation at S_{7a}, S_{7b}, S_{7c}, and S_{2d} could be tolerated but may require some rearrangement of the peptide (Figure 3.1, Figure 5.1 A) [112]. Phosphorylation of S_{2c} would likely be incompatible with polymerase binding, given the tight contact with the polymerase and the highly negative charge on S_{5c} in close proximity. Furthermore, phosphorylation of additional sites beyond the critical pS_{5a} and pS_{5c} could induce conformational changes in the peptide incompatible with polymerase binding. Interestingly, we observed weak but significant binding to the pS7 peptide. While the functional significance of S7 phosphorylation remains poorly understood, previous studies have linked it to events early in the Pol II transcriptional cycle [278]. Given the weak binding observed, further studies using different methodologies are needed to assess pS7

peptide binding to the FluPol A and its potential functional significance.

In contrast to the strong binding of FluPol A, FluPol B exhibited overall weak interactions with the Pol II CTD peptide panel in binding assays. While pS5 CTD_{4x} was recognised above background levels, the difference between pS5 CTD_{4x} and the control peptides, including scrambled, unphosphorylated and pS2 CTD_{4x}, was the smallest compared to FluPol A and FluPol C (Figure 5.1 C-E). Eluate samples from peptides with alternate phosphorylation patterns, specifically p(S2/5) as well as p(S5/7) peptides, exhibited similar band intensities to pS5 CTD (Figure 5.1 D). However, the band intensities were overall lower than those observed in the experiment with FluPol A or FluPol C (Figure 5.1 C, E). These findings align with previous interaction experiments of FluPol B with CTD mimic peptides. In contrast to FluPol A or FluPol C, SEC experiments conducted with $pCTD_{26x}$ and FluPol B (see chapter 4.2.5), as well as previous interaction experiments of synthetic pS5 CTD with FluPol B, indicate that FluPol B does not distinguish well between pS5 CTD and other tested CTD modifications [261]. Given that FluPols from IAV and IBV have evolved different CTD binding modes, it is likely that a broader network of host transcriptional proteins influences the FluPol B-Pol II CTD interaction beyond the identified CTD binding sites [26]. Future studies should assess the different CTD binding modes and affinities across influenza virus types using quantitative approaches and considering the differences in CTD binding and transcriptional factors required to mediate the FluPol-Pol II interaction.

Similar to the results observed with 1918 FluPol A, FluPol C exhibited a strong preference for pS5 CTD peptides over other phosphorylation patterns (Figure 5.1 E). Alternate pS2, pS5, and pS7 combinations abolished CTD binding, indicating that FluPol likely requires the neighbouring pS5 CTD heptad repeat phosphorylation for effective binding. Interestingly, double-phosphorylated pS5 pS7 CTD was recognised by FluPol C above background levels, as indicated by

the quantification of the band intensities (Figure 5.1 E). Meanwhile, SDS-PAGE analysis of the eluate samples showed stronger band intensities above background levels for both the pS2 pS5 CTD and pS5 pS7 CTD double-phosphorylated peptides (Figure 5.1 E). This discrepancy between the quantification and the SDS-PAGE analysis might be due to minor statistical variations. An analysis of the CTD-bound FluPol C structure suggests that FluPol C may potentially tolerate additional phosphorylations present on either the S2 or S7 residues [97, 280]. The CTD binding sites, site 1C and site 2C, found on FluPol C are not conserved in IAV and IBV [40, 41]. In site 1C, 12 CTD peptide residues from three heptad repeats were modelled into the density, namely $T_{4a}pS_{5a}P_{6a}S_{7a}$-$Y_{1b}S_{2b}P_{3b}T_{4b}pS_{5b}P_{6b}S_{7b}$-$Y_{1c}$ [97, 280]. The residues in site 2C could not be modelled into the density due to a low resolution that did not allow building individual residues into the structure [97, 280]. In site 1C, the pS_{5a} is stabilised by interactions with H461 and R465 of PB1 while the other resolved pS_{5b} is exposed toward the solvent [97, 280]. In the first resolved repeat, phosphorylated S_{7a} would be facing toward the polymerase surface but could potentially be accommodated, depending on the peptide arrangement and polymerase conformational rearrangements. A S_{2b} phosphorylation would be exposed to the solvent, while S_{7b} might cause steric clashes. Since site 2C could not be fully resolved, it remains unclear how additional phosphoserines might impact the binding dynamics of the CTD [97]. In conclusion, pS5 CTD shows the strongest binding to FluPol C, while additional phosphorylation either diminishes or completely abolishes CTD binding.

After testing differential CTD phosphorylation patterns, I investigated the impact of CTD binding on FluPol transcriptional activity. Optimisation experiments indicated that salt concentrations around 150 mM are optimal to measure a transcription signal reproducibly (data not shown). These conditions are similar to previously published transcription activity experiments using comparable experimental approaches with recombinantly expressed FluPol, Pol II CTD mimic peptides, and a cap donor [26, 96, 97]. After optimising the assay, the

impact of long recombinant CTD_{26x} as well as synthetic pS7 CTD_{4x} and pS5 pS7 CTD_{4x} on transcriptional activity was examined (Figure 5.2, Figure 5.3). The unphosphorylated CTD_{4x} and pS5 CTD_{4x} could reproduce experimental data from previous research [96, 97]. Additionally, consistent with a previous study, 1918 FluPol A displayed elevated transcriptional activity in the absence of CTD compared to FluPol C, as demonstrated by the background band laddering pattern generated in transcription assays using FluPol A (Figure 5.2, Figure 5.3) [97]. The observed laddering pattern is potentially caused by *de novo* primer-independent replication initiation, which appears to be particularly enhanced in the presence of long recombinant CTD_{26x} peptides (Figure 5.2 A) [44, 280]. Indeed, a recent study proposed that the Pol II CTD is required for both transcription and replication [74]. Experiments with 1918 FluPol A showed a modest increase in signal for pS5 CTD_{4x}, but no visible elevation in signal for pS7 CTD_{4x}, which had previously shown binding above background levels in a CTD binding assay (Figure 5.1 C, Figure 5.2). This suggests that while FluPol can recognise pS7 CTD_{4x}, it may not enhance transcriptional activity. Further replicates would be needed to quantify and statistically analyse the observed signals in future studies.

In the GTP incorporation assay with FluPol C, only the addition of pS5 CTD_{4x} resulted in a visible transcription signal (Figure 5.3). This observation is in line with previous data where FluPol C exhibits a high transcriptional activity after the addition of pS5 CTD FluPol C but not with other CTD peptides [26, 96, 97]. Double-phosphorylated pS5 pS7 CTD_{4x}, which was recognised by FluPol C in CTD binding assays, did not produce a transcription signal (Figure 5.3, Figure 5.1 E). Similar to the pS7 CTD_{4x}, FluPol C might recognise a pS5 pS7 CTD peptide but, in the context of transcription, additional phosphoserines negatively impact viral transcription *in vitro*. In conclusion, the enhanced transcriptional activity caused by pS5 CTD_{4x} could be reproduced for FluPol C, similar to previous work. Additional phosphorylation likely abolishes the transcriptional activity of FluPol

in the preliminary tests conducted in this project.

The transcriptional activity of FluPol, in the presence of recombinant CTD_{26x} peptides, could not be reliably measured due to technical difficulties. In all conducted experiments, an accumulation of radioactive signal was detected within the PAGE gel wells and no differences were detected between unphosphorylated and $pCTD_{26x}$ (Figure 5.2, Figure 5.3, Figure 5.4). Given that the reactions containing recombinant CTD_{26x} had a significantly higher protein content, it is possible that FluPol, the recombinant peptides and the reaction product aggregated, preventing the reaction product from migrating into the PAGE gel. The addition of formamide to varying amounts of recombinant CTD_{26x}, ranging from 1 to 10-fold molar excess of CTD peptide to FluPol, did not increase a detectable signal. In fact, a 10-fold molar excess of CTD_{26x} peptides only resulted in a higher background smear in the PAGE gel (Figure 5.4). Due to time constraints, this technical challenge could not be resolved within the scope of this project. With additional time, approaches such as sample proteolysis before loading or further purification of the reactions through RNA sample extraction could be attempted to improve signal detection.

Past studies proposed that due to the large size of the Pol II and its long CTD, the FluPol surface may harbour additional binding sites that would stabilise the FluPol-Pol II interaction [40, 41]. This project investigated the interaction between PB2-C and phosphorylated CTD mimics, based on the observation of a positive electrostatic surface potential near the N-terminal end of the Pol II CTD peptide [112]. We hypothesised that the positive electrostatic surface potential near the N-terminus of the Pol II CTD peptide might facilitate a binding interaction with a negatively charged pS5 CTD, potentially extending the recently observed continuous CTD binding site on PA-C and stabilising the FluPol-Pol II interaction [112]. To investigate this, the interaction of IAV PB2-C was tested using PB2-C and variations of its subdomains, including 627 and 627-NLS, with

the CTD peptides. Due to the lack of a 1918 IAV PB2-C construct, PB2-C from other IAV strains (Viet and NT60), which share a high sequence conservation, was initially used in pilot experiments [112]. Interaction experiments of PB2-C and 627-NLS, a PB2-C subdomain, with $pCTD_{26x}$, did not reveal any interaction in both pulldown experiments and an SEC analysis (Figure 5.5). In addition, 627-NLS, GST-627 and PB2-C did not reveal an interaction with synthetic pS5 CTD_{4x} in CTD binding assays (Figure 5.6, Figure 5.7). Despite the positive surface of the PB2-C domain, only weak interactions, if any, may exist between PB2-C and the Pol II CTD, which could not be detected using pulldown and SEC experiments. Alternatively, functional studies have described an interaction of PB2-N with the RPB4 subunit of the FluPol core, positioning FluPol close to the RNA exit channel where the nascent Pol II transcript emerges [237]. In this suggested arrangement, PB2-C might be redundant in mediating the FluPol-Pol II CTD interaction but remains important in vRNP export through its direct interaction with the cellular GTPase Rab11 [220, 237]. More structural and functional studies will be required to elucidate the molecular details by which FluPol hijacks Pol II transcription and which particular FluPol domains and Pol II subunits are involved in this process.

This chapter explored how various phosphorylation combinations of the three Ser residues — S2, S5, and S7 — in the CTD heptad repeat consensus sequence influence FluPol-CTD binding, showing that phosphorylation of S5 residues is recognised specifically. The transcriptional activity of FluPol was enhanced by pS5 CTD but not by other modifications. Finally, interaction experiments showed that PB2-C and Pol II CTD mimics do not interact, highlighting the need for future structural work to reveal the molecular details of cap-snatching.

6

Characterisation of host factors modulating the FluPol-Pol II CTD interaction

6.1 Introduction

The influenza virus polymerase interacts with various host factors to support its functions throughout the viral life cycle [40, 41, 44]. Upon entry into the nucleus, the vRNP-associated FluPol carries out transcription and replication [41]. In the early stages of infection, FluPol predominantly performs transcription, but as free viral proteins accumulate in the nucleus, FluPol increases its replication rate [44].

Viral transcription depends on the interaction of FluPol with Pol II and its associated transcription machinery. Specifically, FluPol binds directly to the pS5 CTD of Pol II to perform cap-snatching [40, 41]. Other transcription factors, including the DSIF complex, are believed to support viral transcription and have been shown to interact with FluPol in proteomics and systems biology studies [58, 196, 281–283]. DSIF is a heterodimeric complex consisting of the Spt4 and Spt5 subunits, with Spt5 being significantly larger than Spt4 [284]. Spt5 comprises an acidic domain at the N-terminus, followed by a NusG N-terminal domain, and finally, KOW motifs that flank an extended, flexible, and disordered C-terminal region (CTR) [179]. The Spt5 CTR domain undergoes post-translational modifications, including phosphorylation by the P-TEFb complex, which triggers the release of Pol II from promotor-proximal pausing and the transition to productive elongation [172, 179]. Structural studies have shown that DSIF, when bound to elongating Pol II, interacts extensively with Pol II and the associated DNA and RNA, involving Spt4, the Spt5 NGN domain, and several of the KOW motifs [179].

Viral replication requires the oligomerisation of FluPol and interaction with the host ANP32 (acidic nuclear phosphoprotein 32) protein family [143, 144, 250, 251]. ANP32 proteins consist of a leucine-rich repeat domain at the N-terminus and a C-terminal low complexity acidic region (LCAR), with differences between avian and mammalian species that contribute to the host range of influenza viruses. An asymmetric dimer of FluPol, bridged by ANP32, forms the replication

complex, which comprises a replicase and an encapsidase [44]. The RNA-bound replicase synthesises *de novo* viral RNA, while the RNA-free encapsidase accepts the replication product [145]. NP is recruited to the forming progeny RNP to package the growing replication product via its interaction with the ANP32 LCAR domain [219].

FluPol adopts various conformations to perform its functions depending on the host factors and ligands it associates with by rearrangement of its flexible domains around the polymerase core [44, 73]. For instance, nuclear export protein (NEP), also termed NS2, is a multifunctional protein in the viral life cycle. NEP forms with M1 and the vRNP, a complex that is exported from the nucleus to the plasma membrane for the assembly and packaging of progeny virions [285, 286]. Independent of its export function, NEP supports the transition from viral transcription to replication by supporting the dissociation of FluPol from the Pol II CTD and stabilising FluPol in a replicase conformation [148, 287].

The molecular details underlying the transitions from transcription to replication and vRNP export are still under investigation, and various models have been proposed to address this question [44]. Recently, it was proposed that the Pol II CTD might act as a platform for both transcription and replication, promoting viral transcription in the early stages of infection and later enhancing viral genome replication [74]. Another model suggested that the transition between transcription, replication, and vRNP export is a stochastic process that depends on the availability and concentration of various molecular cues such as free viral proteins, ANP32, NEP and the progression of Pol II degradation [44, 73]. Although the interactions between viral proteins and host factors are likely essential for spatiotemporal regulation of viral transcription, replication, and vRNP export, the precise molecular mechanisms remain elusive [44].

This work aimed to elucidate how host proteins such as ANP32, NEP, and Spt5 influence the interaction between FluPol and the Pol II CTD. In addition, the purification of ALFA-tagged Pol II was pursued for future applications in functional and structural experiments that explore viral transcription and replication.

6.2 Results

6.2.1 ANP32B competes with pS5 Pol II CTD for FluPol A binding

ANP32 and Pol II CTD have overlapping binding sites in the PA-C domain, as was demonstrated in structural and functional experiments [74, 112, 145, 248]. Due to this overlap, we hypothesised that both host factors might compete for FluPol binding. Therefore, I performed a competition assay to assess how ANP32B and Pol II CTD binding to FluPol might interfere with one another (Figure 6.1).

In this assay, either a scrambled CTD_{4x} or pS5 CTD_{4x} was immobilised on streptavidin resin. ANP32B, which is functionally redundant with ANP32A but required for mammalian-adapted IAV and IBV replication, was co-incubated with 1918 FluPol A and added at a 0 to 27-fold molar excess to 1918 FluPol A [250]. SDS-PAGE analysis of the input proteins indicated a band at around 28 kDa, corresponding to ANP32B, and a doublet band around 80 kDa, corresponding to FluPol (Figure 6.1 A). In the control, using scrambled CTD_{4x}, only background bands around 80 kDa were visible across all tested conditions (Figure 6.1 A). However, pS5 CTD_{4x} was able to pulldown FluPol, evident as a strong doublet band around 80 kDa on the SDS-PAGE gel (Figure 6.1 A). As the amount of ANP32B increased, the FluPol bands decreased in intensity (Figure 6.1 A). Quantification of these bands showed that a 0.33-fold molar addition of ANP32B already resulted in a 50% decrease in FluPol binding to pS5 CTD_{4x} (Figure 6.1 B). A gradual reduction in FluPol signal was observed as the concentration of ANP32B was increased. At a 0.33-fold molar addition of ANP32B, a band around 60 kDa corresponding to BSA was detected, likely caused by a contamination during loading of the SDS-PAGE gel. A 27-fold molar excess resulted in a near-complete abolition of FluPol binding, with only background levels of FluPol detected, similar to the control samples containing scrambled CTD_{4x} (Figure 6.1 A, B). In conclusion, increasing concentrations of ANP32B resulted in a reduction of FluPol A binding to pS5 CTD_{4x}.

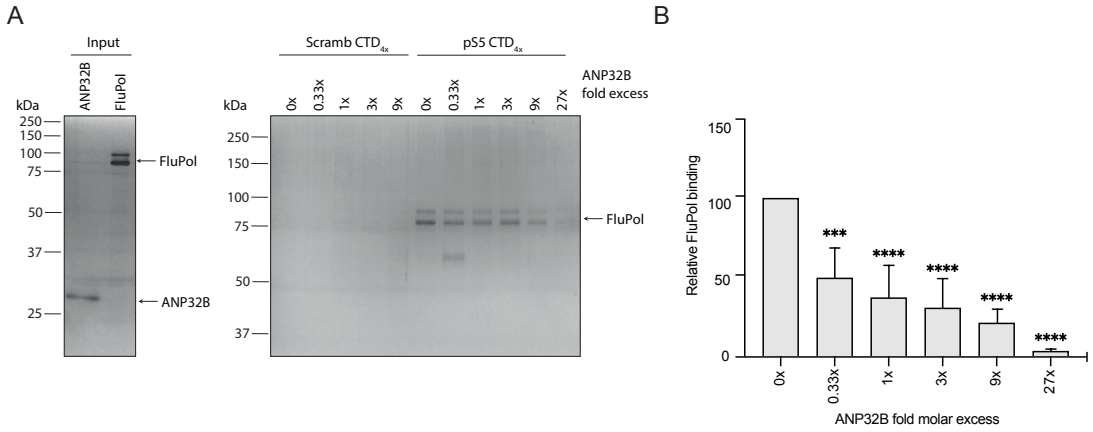

Figure 6.1: Competition assay examining the interaction between 1918 FluPol A and pS5 CTD_{4x} peptide in the presence of ANP32B. (A) Input proteins ANP32B (kind gift from Fangzheng Wang) and 1918 FluPol A (left). Representative silver-stained SDS-PAGE gel. ANP32B was added at a molar excess of 0x to 27x to 1918 FluPol A with beads pre-incubated with scrambled or pS5 CTD_{4x} peptide (right). **(B)** Quantitation of the competition assay, data are mean ± s.e.m from three independent experiments. An ordinary one-way ANOVA test was employed to compare relative FluPol binding in the presence and absence of ANP32B with significant differences denoted as ***, $p < 0.0005$; ****, $p < 0.0001$.

6.2.2 NEP does not outcompete pS5 Pol II CTD for FluPol A binding

The viral proteins NEP and M1 interact with the cellular nuclear export machinery to facilitate vRNP nuclear export [1]. Studies have demonstrated that NEP regulates viral transcription and replication [148]. To investigate its potential role in regulating the FluPol-Pol II CTD interaction, an essential interaction for viral transcription, I conducted competition assays similar to the competition assays with ANP32B (Figure 6.2.1). Increasing amounts of NEP were co-incubated with 1918 FluPol A in its vRNA-bound and RNA-free (apo) form to assess whether different confirmations of FluPol, induced by the presence or absence of vRNA, affected CTD binding in the presence of NEP (Figure 6.2.2). SDS-PAGE analysis of the input proteins, NEP and 1918 FluPol, indicated a band around 37 kDa, corresponding to GST-tagged NEP, and a doublet around 80 kDa corresponding to FluPol (Figure 6.2.2 A). In the condition using a scrambled CTD_{4x} peptide and apo FluPol, FluPol was detected above background levels likely caused by

protein precipitation. This signal was observed for samples with 0.33-fold, 9-fold, and 27-fold molar excess of NEP which exhibited a signal above the -CTD control lane (Figure 6.2.2 A, top panel). In the presence of pS5 CTD_{4x}, FluPol bound to the beads at all conditions. The bands corresponding to FluPol are all similar in intensity to the -NEP control (Figure 6.2.2 B, top panel). The same experiment was then repeated using vRNA-bound FluPol (Figure 6.2.2 B, middle panel). The scrambled CTD_{4x} together with vRNA-bound FluPol only exhibited background staining, while in the presence of pS5 CTD_{4x} bands around 80 kDa, corresponding to FluPol, were visible (Figure 6.2.2 B, middle panel). Compared to the condition without viral RNA (-vRNA) in the top panel, the band intensities were reduced in the experiment with vRNA-bound FluPol, potentially due to technical issues with silver staining (Figure 6.2.2 B, middle panel). Quantification of both conditions, in the absence and presence of FluPol showed that NEP does not appear to reduce CTD binding to FluPol, with all test conditions indicating similar binding as the condition without NEP (-NEP) (Figure 6.2.2 B, bottom panel). In conclusion, this result suggests that NEP does not compete with FluPol for CTD binding under the tested experimental conditions.

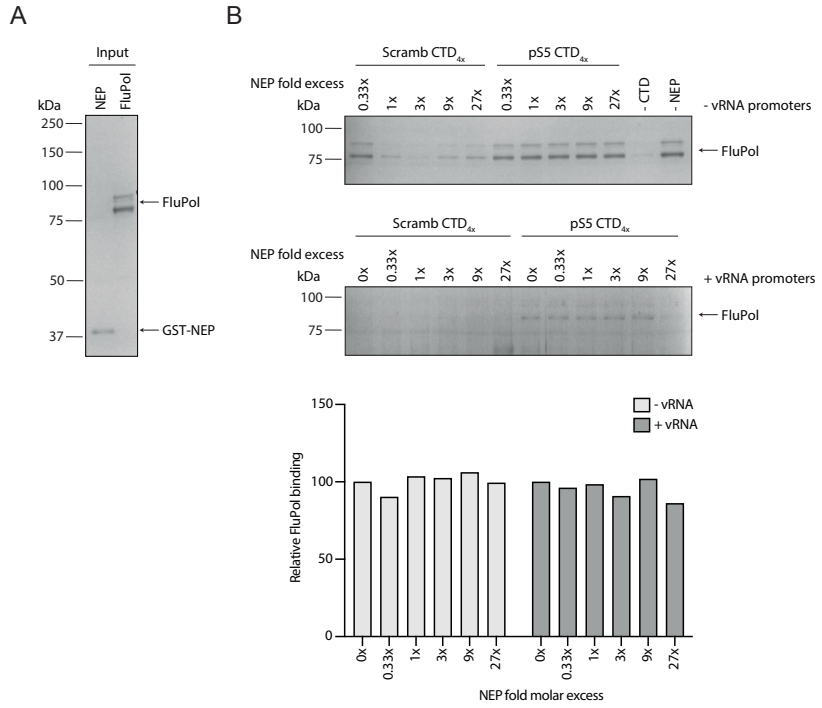

Figure 6.2: Competition assay examining the interaction between 1918 FluPol A and pS5 CTD_{4x} peptide in the presence of NEP. (A) Input proteins GST-tagged 1918 NEP (kind gift from Alison Rep) and 1918 FluPol A. **(B)** Representative silver-stained SDS-PAGE gels of competition assays. 1918 NEP was added at a molar excess of 0x to 27x to apo (top panel) or vRNA-bound (middle panel) 1918 FluPol A with beads pre-incubated with scrambled or pS5 CTD_{4x} peptide. The bands of the competition assay were quantified (bottom panel). Data are normalised to the 0x condition from n=1 experiments per condition (± vRNA).

6.2.3 The Spt5 subunit of the DSIF complex interacts with 1918 FluPol A

FluPol has been reported to interact with various cellular proteins. In the context of transcription, the DSIF complex has been shown to interact with FluPol through its Spt4 and Spt5 subunits [40, 196–198]. As we aimed to characterise an active cap-snatching complex structurally, we hypothesised that DSIF, particularly its Spt5 subunit, might stabilise these interactions since DSIF stabilises Pol II during promoter proximal pausing [40, 172, 179, 180].

Therefore, when a full-length DSIF construct was unavailable at the project's initial stages, I purified an Spt5 construct from *E. coli* that expresses 387 amino acids of Spt5 containing the flexible C-terminal region (CTR) flanked by the

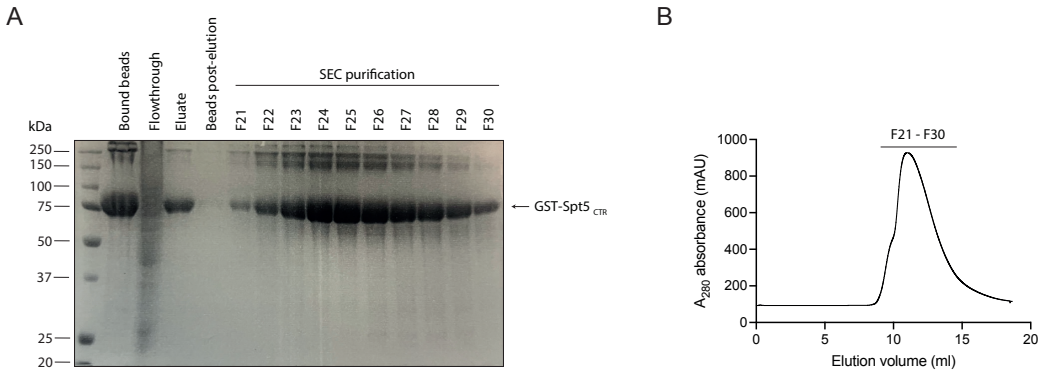

Figure 6.3: Expression and purification of GST-Spt5$_{\mathbf{CTR}}$. Truncated Spt5, a subunit of the DSIF complex, was expressed as a GST-Spt5$_{\mathrm{CTR}}$ protein in *E. coli* and samples taken during the purification process were analysed by SDS-PAGE analysis. Cell lysates were incubated with Ni-NTA sepharose beads (bound beads). Protein was eluted with GSH (eluate, beads post-elution) and concentrated before SEC purification. **(A)** Expression and purification of GST-Spt5$_{\mathrm{CTR}}$. **(B)** Elution profile of the SEC purification of GST-Spt5$_{\mathrm{CTR}}$.

KOW5 and KOW6/7 motifs (Figure 6.3) [172]. The GST-Spt5$_{\mathrm{CTR}}$ construct was purified on glutathione sepharose, and samples taken throughout the purification were analysed by SDS-PAGE (Figure 6.3 A). A strong band around 75 kDa, corresponding to GST-Spt5$_{\mathrm{CTR}}$ was visible in the bound beads, eluate, and SEC fraction samples (Figure 6.3 A). Other bands around 250 kDa were visible, indicating some contaminants (Figure 6.3 A). The SEC purification of GST-Spt5$_{\mathrm{CTR}}$ exhibited a peak with an intensity of 950 mAu around 12 ml (Figure 6.3 B).

To test a potential interaction of FluPol and the GST-Spt5$_{\mathrm{CTR}}$ construct *in vitro*, I conducted a pulldown assay with 1918 FluPol A (Figure 6.4). Input samples indicated bands at 25 kDa, 75 kDa, and 80 kDa, corresponding to GST, GST-Spt5$_{\mathrm{CTR}}$ and FluPol (Figure 6.4). In the control condition, using GST tag instead of GST-Spt5$_{\mathrm{CTR}}$ to control for unspecific binding, a band around 25 kDa was observed corresponding to GST and some background bands, which are likely the result of minor spillovers from the neighbouring wells (Figure 6.4). In the flowthrough, a strong band corresponding to FluPol was detected, and in the eluate, a faint band at the same position was observed together with GST, which

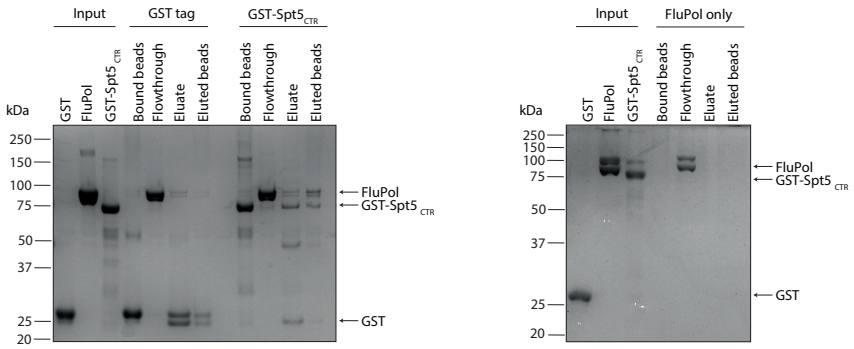

Figure 6.4: Pulldown of 1918 FluPol A with GST-Spt5$_{\mathbf{CTR}}$. The interaction of GST-Spt5$_{\mathrm{CTR}}$ was tested with 1918 FluPol A in pulldown experiment using glutathione sepharose beads. Incubation with the GST tag or the viral polymerase only served as a negative control. Analytes were incubated with the beads (bound beads), followed by incubation with vRNA-bound FluPol (flowthrough) and elution with GSH. Elution samples were taken (eluate) along with the remaining beads (eluted beads).

migrated as a double band due to a gel artefact (Figure 6.4). The eluate sample indicated background binding of 1918 FluPol A to GST. The bound beads sample for the pulldown with GST-Spt5$_{\mathrm{CTR}}$ indicated a main band corresponding to GST-Spt5$_{\mathrm{CTR}}$ and two bands around 50 kDa as well as a band around 25 kDa, corresponding to degradation products of GST-Spt5$_{\mathrm{CTR}}$ (Figure 6.4). A strong FluPol band was observed in the flowthrough, while only a faint FluPol signal was observed with GST-Spt5$_{\mathrm{CTR}}$ in the eluate, similar to the control condition with GST alone (Figure 6.4). However, in the post-elution beads, a signal for both FluPol and GST-Spt5$_{\mathrm{CTR}}$ was detected, indicating either incomplete elution or potential aggregation (Figure 6.4). A negative control where FluPol was incubated with the beads without a bait indicated that FluPol did not aggregate on the beads under the used buffer conditions, evident in FluPol only detected in the flowthrough (Figure 6.4). In conclusion, the pulldown of 1918 FluPol A with GST-Spt5$_{\mathrm{CTR}}$ did not reveal a clear interaction.

Since the pulldown experiment had to be conducted in medium salt conditions using 300 mM NaCl to prevent FluPol aggregation on the beads, an SEC analysis was carried out using low salt conditions (150 mM) to detect a potential interaction between FluPol and GST-Spt5$_{\mathrm{CTR}}$ (Figure 6.5). Input samples on the SDS-PAGE

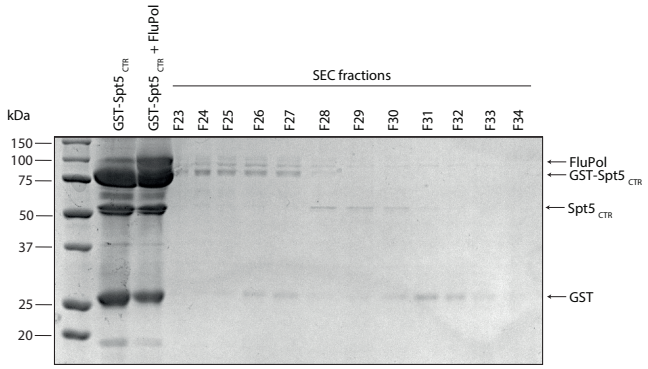

Figure 6.5: Interaction of truncated GST-Spt5$_{\mathbf{CTR}}$ with 1918 FluPol A. SEC analysis of FluPol A, pre-bound to 3' and 5' vRNA promotors (1.2x molar excess) and incubated with truncated Spt5 (1:1 molar ratio). Fractions were collected and analysed with SDS-PAGE. Freshly thawed GST-Spt5$_{\mathrm{CTR}}$ showed signs of degradation to GST and Spt5$_{\mathrm{CTR}}$.

gel revealed a strong band around 75 kDa corresponding to the GST-Spt5$_{\mathrm{CTR}}$ and two bands at 25 kDa and 50 kDa, indicating degradation products of the construct (Figure 6.5). SDS-PAGE analysis of the collected SEC fractions, showed three bands in fractions 23 to 27, corresponding to a potential complex of FluPol and GST-Spt5$_{\mathrm{CTR}}$ (Figure 6.5). However, elution profiles are required to confirm a mobility shift of the analytes. In conclusion, GST-Spt5$_{\mathrm{CTR}}$ potentially formed a complex with 1918 FluPol A and further interaction experiments are needed to confirm this observation.

The DSIF complex binds to the upstream Pol II-bound DNA and forms a DNA-RNA clamp during promotor-proximal pausing, which serves as a control point for the proper processing of nascent RNA during transcription at which FluPol is most likely to first interfere with Pol II-mediated transcription [41, 172]. A proteomics experiment in our research group performed by Serdar Baysal indicated that FluPol interacts with the Spt5 subunit of the DSIF complex when FluPol was incubated with whole cell lysate from HeLa cells (data not shown). To confirm a potential interaction between FluPol and the DSIF complex, I purified DSIF, followed by SEC analysis (Figure 6.6, Figure 6.7). DSIF was expressed in *E. coli* using a construct that co-expressed an N-terminal His$_{10}$-Arg$_8$-SUMO tag on the

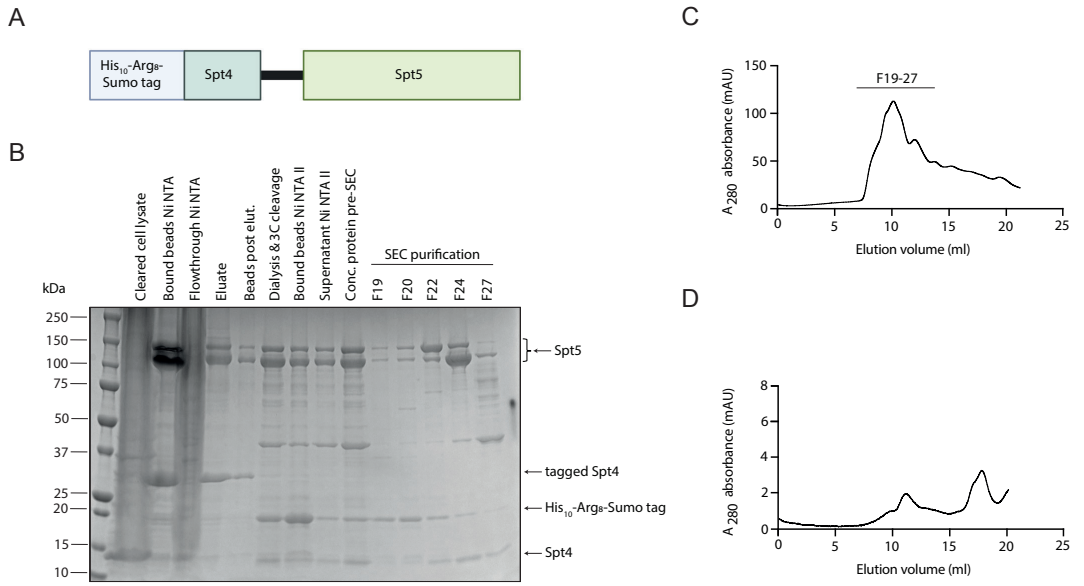

Figure 6.6: Expression and purification of the DSIF complex. The DSIF complex, consisting of the Spt4 and Spt5 subunits was expressed in *E. coli* and samples taken during the purification process were analysed by SDS-PAGE analysis. **(A)** Construct design for expression of DSIF (kind gift from Dr Patrick Cramer). The expression construct is a modified pETDuet vector for co-expression of the Spt4 and Spt5 subunits. The first cassette expresses human Spt4, followed by human Spt5 in the second cassette. The N-terminus of Spt4 contains a His_{10}-Arg_8-SUMO-tag and a 3C protease cleavage site. **(B)** SDS-PAGE analysis of a DSIF purification. Cell lysates were incubated with Ni-NTA sepharose beads (bound beads Ni NTA). Protein was eluted with imidazole (eluate, beads post-elution). After 3C cleavage of the tag on the Spt4 subunit, the tag was removed by incubation with Ni NTA beads (bound beads Ni NTA II) and concentrated before SEC purification. **(C)** Elution profile of the DSIF SEC purification. **(D)** SEC elution profile of freshly thawed DSIF after flash-freezing and storage at -80°C.

Spt4 subunit and untagged Spt5 [172]. After incubation of the cleared cell lysate with Ni-NTA beads, two bands between 100 kDa and 150 kDa and another band around 30 kDa were detected, corresponding to Spt5 and its degradation product as well as tagged Spt4 (Figure 6.6 B). The protein tag on Spt4 was removed through 3C cleavage and a second incubation with Ni-NTA beads, followed by SEC purification (Figure 6.6 B, C). The SEC elution profile showed a main peak at 10 ml with an intensity of approximately 100 mAu and a shoulder peak around 12 ml (Figure 6.6 C). SDS-PAGE analysis detected full-length Spt5 with a smaller band corresponding to a degradation product in the main

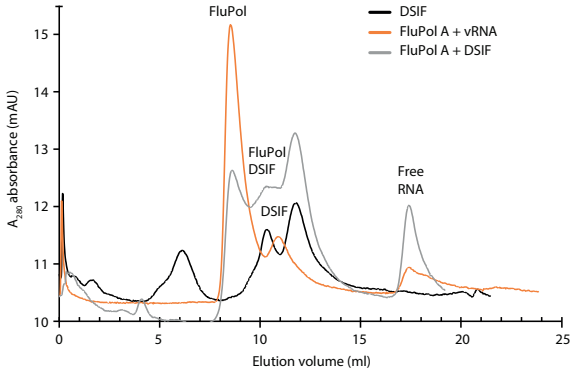

Figure 6.7: SEC analysis of FluPol A with the DSIF complex. SEC elution profile of 3' and 5' vRNA-bound (1.2x molar excess) 1918 FluPol A with DSIF (3x molar excess). The DSIF profile indicates potential degradation or instability, resulting in no complex formation.

peak around fraction 22 (Figure 6.6 C). The shoulder peak corresponding to fraction 24 contained a dominant band for the degradation product and smaller amount of full length Spt5 (Figure 6.6 C). After flash-freezing and thawing of concentrated purified DSIF, the SEC elution profile revealed protein degradation of the sample (Figure 6.6 D). This was also observed when I attempted an SEC analysis of 1918 FluPol A and DSIF where the DSIF elution profile only exhibited a very weak signal compared to the FluPol elution profile. Despite theoretically using a 1.5-fold molar excess of DSIF to FluPol in this experiment, the DSIF elution profile remained at background levels (Figure 6.7). The instability of DSIF was persistent in all protein purification attempts, indicating that the purified protein was not stable enough for freeze-thaw cycles. In conclusion, a stable DSIF complex purification could not be reproduced in our laboratory, which hampered the structural and functional characterisation of a potential FluPol-Pol II-DSIF complex.

6.2.4 Cells expressing ALFA-tagged Pol II could be a useful tool for studying the interaction of FluPol with its host transcriptional partners

FluPol interacts with various cellular factors upon viral infection and requires a close association with Pol II to perform cap-snatching in the nucleus [40, 41]. During this project, access to purified native Pol II was restricted, which constrained our ability to conduct comprehensive cryo-EM studies. To investigate alternative approaches for obtaining native mammalian Pol II for structural and functional experiments, I tested the purification of Pol II from HCT116 cells that stably express ALFA-tagged Pol II (ALFA-Pol II HCT116). The ALFA tag is an artificial short epitope tag consisting of 15 amino acids that form a small and stable α-helix, with the ability to refold even after contact with harsh chemicals [288]. This versatile tag has been be applied in many applications targeting eukaryotic and prokaryotic proteins [288–290]. For applications that require a high-affinity purification of ALFA-tagged protein, a nanobody was engineered that recognises the ALFA tag (NbALFA) and, when conjugated to agarose resin, can be used for biochemical purifications (ALFA Selector PE beads, Nanotag Biotechnologies). This engineered nanobody variant has a higher affinity to ALFA peptides than ALFA-tagged proteins, enabling the elution of the target protein after binding [288].

To purify ALFA-tagged Pol II (ALFA-Pol II), I grew ALFA-Pol II HCT116 cells to 80% confluency and harvested a cell pellet of approximately 0.2 ml. The cells were then lysed, and purification was carried out following the manufacturer's instructions for ALFA Selector PE beads (Nanotag Biotechnologies). During the purification process, HCT116 cells that do not stably express ALFA-Pol II were used as a control. Purification samples were analysed by SDS-PAGE (Figure 6.8).

Eluate samples and the post-elution beads of the purification from ALFA-Pol II HCT116 cells indicate the presence of Pol II (Figure 6.8). A band around

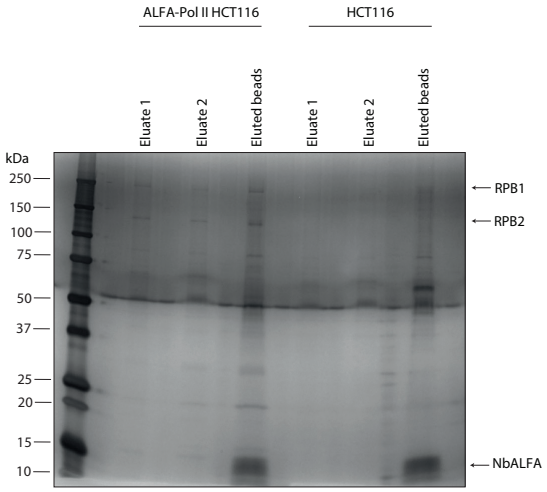

Figure 6.8: Expression and purification of ALFA-tagged Pol II from mammalian cells. ALFA-tagged Pol II was stably expressed in HCT116 cells (kind gift from Dr Steven West) and purified on ALFA Selector PE beads (Nanotag Biotechnologies) and eluted with ALFA peptide (eluate 1, eluate 2, eluted beads). Regular HCT116 cells were used as a negative control.

250 kDa and 130 kDa can be seen, corresponding to the RPB1 and RPB2 subunits (Figure 6.8). In HCT116 cells, only background staining could be observed in the eluates at the same position on the gel (Figure 6.8). Silver staining of the SDS-PAGE gel exhibited a dye front around 50 kDa and a higher background staining of the upper half of the gel, which is likely a staining artefact (Figure 6.8). Two strong bands in the eluates of ALFA-Pol II HCT116 purification were detected around 30 kDa and 25 kDa, corresponding to the RPB3 and RPB5 subunits of Pol II (Figure 6.8 A). The other bands corresponding to RPB4 (16 kDa), RPB8 (19 kDa), and RPB7 (17 kDa) were only weakly detected due to their small size and the amount of protein loaded (Figure 6.8). The remaining Pol II subunits could not be reliably detected in the eluates and might have overlapped with the NbALFA band in the post-elution beads sample (Figure 6.8). Western Blot analysis of samples taken during the purification process will be required to confirm the identity of the detected bands on the SDS-PAGE using antibodies against the Pol II subunits.

In conclusion, this pilot experiment indicates the presence of ALFA-Pol II in the ALFA-Pol II HCT116 cell line and further optimisation is required to validate the quality and amount of purified Pol II for the use in downstream applications.

6.3 Discussion

This chapter explored the role of host factors ANP32B and Spt5 and viral NEP in modulating the FluPol-Pol II CTD interaction. The binding of 1918 FluPol A to immobilised pS5 CTD_{4x} was assessed under the presence of ANP32B in a CTD binding assay (Figure 6.1). Increasing amounts ANP32B caused reduced binding of FluPol A to pS5 CTD_{4x}, suggesting that ANP32B and the CTD compete for FluPol binding. This observation agrees with similar studies that used FluPol A from other strains, including an H5N1 or WSN (H1N1) virus [248, 291]. A competition assay of pS5 CTD_{4x} with FluPol A from H5N1 demonstrated that ANP32B outcompetes Pol II for FluPol binding in a concentration-dependent manner [248]. In addition, the observed overlap of the CTD and ANP32 binding sites in the PA-C suggests that the binding of these host factors is mutually exclusive [112, 145, 248, 249].

Based on these observations, a model was suggested where newly synthesised free FluPol A binds to ANP32 proteins via its PA-C upon entry to the nucleus. This interaction prevents binding to Pol II CTD, and the competition with vRNP-associated FluPol [248, 275]. In addition, the interaction of ANP32 with FluPol would stabilise it in a conformation compatible with replication, allowing the assembly of a replication platform through binding to another vRNP-associated FluPol [248]. In addition, another study suggested that the CTD acts as a platform for both transcription and replication and that the CTD, together with ANP32, increases replication activity generating a full-length cRNA from a vRNA template [74]. This study also proposed that CTD promotes ANP32 binding and polymerase dimerisation [74]. While ANP32 would bind to the encapsidating polymerase, the vRNP-associated polymerase (replicase) would bind to the Pol II

CTD via its free PA-C domain [74, 248, 249]. However, this suggested model requires further structural evidence.

NEP is a multifunctional protein that influences viral replication and, together with M1, is required for vRNP export [292]. The regulatory role of NEP on viral RNA synthesis is independent of its export function [148]. Moreover, the influence of NEP on replication is nuanced with experiments indicating that low levels of NEP stimulate viral replication while high NEP concentrations lead to the inhibition of transcription and replication [68]. NEP has a binding site in FluPol A that overlaps with that of ANP32 and the Pol II CTD at PA-C [287]. Here, a competition between NEP and pS5 CTD_{4x} for FluPol A binding was tested (Figure 6.2). The experiments showed that even at increasing concentrations, NEP could not outcompete the phosphorylated pS5 CTD_{4x} peptide for binding to FluPol A, whether in its apo or vRNA-bound form. Similar studies conducted by colleagues in our research group also failed to detect NEP outcompeting the Pol II CTD (data not shown).

Our observations in NEP competition experiments contradict a recently published study that tested the interference of NEP with the FluPol A-Pol II CTD binding using a luciferase complementation approach [287]. In this study, co-expression of NEP with FluPol A and Pol II CTD caused a significant reduction of the FluPol-Pol II CTD interaction and a decrease in viral transcription, suggesting that NEP can sterically hamper Pol II CTD binding. Furthermore, the authors of this study suggest that NEP prevents FluPol A from adopting a transcriptase conformation [287]. It is possible that we did not observe a competition due to the experimental set-up that utilised recombinantly purified FluPol instead of reconstitued vRNP in a cellular context. In addition, in the cell, other unknown host factors or cellular processes might contribute to the destabilisation of the FluPol-Pol II CTD interaction. In conclusion, elucidating the role of NEP in orchestrating the switch from transcription and replication to vRNP export will

require further experimental work, potentially involving competition studies using RNP-based approaches and detailed structural insights.

For viral transcription, various host factors have been suggested to be involved in cap-snatching [40]. The DSIF complex, a heterodimeric transcription factor plays an essential role during promoter-proximal pausing and has been investigated as a binding partner of bat FluPol A [261]. We aimed to explore this interaction further to expand on our 1918 FluPol A pS5 CTD_{4x}-bound structure. Initially, an $Spt5_{CTR}$ construct was purified to a high yield and employed in interaction experiments (Figure 6.3). A pulldown experiment of 1918 FluPol with GST-$Spt5_{CTR}$ indicated no difference between the GST tag alone and GST-$Spt5_{CTR}$ (Figure 6.4). However, in the post-elution beads, strong bands for FluPol were detected, and no unspecific binding of FluPol to the beads could be observed in the control containing FluPol alone, indicating that a potential complex was not fully eluted with GSH (Figure 6.4). Another factor that might have contributed to this inconclusive result is the use of a medium salt concentration of 300 mM NaCl, which was required to keep 1918 FluPol A stable in the buffer.

An SEC analysis of GST-$Spt5_{CTR}$ with 1918 FluPol A conducted at 150 mM NaCl, indicated potential co-elution of the analytes, pointing to a possible complex of FluPol and $Spt5_{CTR}$ (Figure 6.5). However, additional experiments are required to confirm this preliminary finding using appropriate controls. This result and proteomics data suggesting that NT60 FluPol A and Spt5 interact with each other (data not shown) prompted us to purify a full-length DSIF complex from an externally sourced construct (kind gift from Dr Patrick Cramer) for use in structural and functional experiments. SDS-PAGE and mass spectrometry analysis (data not shown) of samples of the purification process detected the Spt4 and Spt5 subunits of DSIF (Figure 6.6). However, the protein was unstable and could not withstand a freeze-thaw cycle or application on cryo-EM grids (data not shown) (Figure 6.6 D). For this reason, an SEC analysis of DSIF

with 1918 FluPol A did not reveal an interaction and indicated degradation of DSIF (Figure 6.7). A similar SEC analysis from another research group showed an interaction of DSIF with bat FluPol A [261]. However, we could not reproduce this in our laboratory. With more time, new constructs could be designed where Spt5 and Spt4 form a fusion protein that could stabilise the heterodimeric complex. Alternatively, a different expression system, such as insect cells, could be investigated to produce DSIF [293]. After the successful purification of DSIF, the structural analysis of an active cap-snatching complex will likely be more feasible and together with the expression of NELF, another transcription factor required for promoter-proximal pausing, might be the key to capturing the structure of an active cap-snatching complex [172, 179, 293].

FluPol hijacks Pol II transcription to perform cap-snatching and requires a direct interaction with Pol II [41]. Since access to purified native Pol II was a limiting factor in the functional and structural characterisation of a full-length Pol II-FluPol complex, I sought to explore the purification of mammalian Pol II from a cell line that stably expresses ALFA-Pol II (Figure 6.8). Indeed, bands corresponding to the Pol II subunits RPB1 and RPB2 could be detected using silver-staining (Figure 6.8). However, due to time constraints, Western blotting to confidently detect other Pol II subunits, a yield optimisation of the purification and further pilot experiments using this cell line could not be performed within the scope of this thesis. Due to the versatility of the ALFA-tag system, a myriad of experiments could be pursued in the context of influenza, ranging from cryo-EM applications to pulldown experiments or microscopy studies, as demonstrated by various studies successfully using this system [294–297]. Experiments using the ALFA system could help to further elucidate the role of other host factors that are required for viral transcription and replication, for example, using a proteomics approach with the highly specific NbALFA-conjugated resins. For cryo-EM application, vRNPs or FluPol could be tagged with NbALFA scaffolds [298, 299].

In summary, this work investigated the interaction of various host factors with 1918 FluPol A. ANP32B, an essential host factor for viral replication, outcompeted pS5 CTD_{4x} for FluPol binding while NEP did not lead to a similar competition effect. In addition, we found that FluPol interacts with Spt5. However, the structural details of this interaction remain elusive. Finally, native mammalian ALFA-tagged Pol II was explored as a tool for future studies addressing the FluPol-Pol II interaction and its role in viral genome transcription and replication.

7

Conclusion and future directions

The influenza virus has evolved intricate ways to hijack cellular processes, exploiting the host transcriptional machinery for viral transcription. Cap-snatching, a key step in viral transcription, requires direct interaction of FluPol with the host Pol II [41]. This work extended the understanding of the molecular mechanism by which influenza viruses exploit host proteins for viral transcription.

Our structural analysis of 1918 FluPol A bound to a short four-repeat synthetic CTD mimic peptide provided the first CTD-bound structure of a medically relevant pandemic strain, revealing a continuous binding site in the PA-C domain. In addition, a novel conformation with the PB2-C domain positioned close to the CTD peptide N-terminus was detected, suggesting that the PB2-C might contribute to the FluPol-Pol II interaction, extending beyond influenza B viruses to influenza A viruses as well [112]. This finding contributes to our understanding of how FluPol domains interact with Pol II.

Building on these structural insights, I characterised the CTD-FluPol interaction by studying residues in this binding site predicted to affect CTD binding. Functional analyses of the CTD-binding site of 1918 FluPol A revealed that conserved residues in this binding site affect both transcription and replication. Mutations in these residues resulted in reduced mRNA and cRNA levels and disrupted binding to ANP32B, demonstrating the role of the PA-C domain for transcription and replication [112].

Notably, the structure also demonstrated that phosphorylation at S5 is essential for binding of FluPol, while additional phosphorylation at S2 and S7 residues disrupts this interaction, linking cap-snatching to early Pol II transcription initiation [112, 179, 279]. Furthermore, our study and other structural and functional studies revealed that the PA-C is a key domain that orchestrates fundamental processes of the viral life cycle, including viral transcription, replication, and vRNP export [40, 74, 96, 112, 248, 249, 287]. Structural analyses demonstrate that the binding

of host factors to PA-C is mutually exclusive, opening questions about the spatiotemporal regulation of these processes [112, 248, 249, 287].

Expanding on these structural findings, I investigated the interaction of FluPol with Pol II using recombinant long CTD mimics, comprising 26 repeats, and native Pol II in cryo-EM experiments. Biochemical experiments revealed that CDK7-phosphorylated CTD_{26x} formed complexes with FluPol from influenza A, B, and C viruses. Interestingly, interaction of FluPol with $pCTD_{26x}$ promoted oligomerisation of FluPol, suggesting a potential role of CTD-driven oligomerisation in viral transcription and replication. However, cryo-EM experiments to structurally investigate these complexes faced technical challenges, including particle aggregation and preferential orientation. Further investigation of these oligomeric states in the context of vRNPs, which may form viral factories on the long Pol II CTD, could reveal how the long Pol II CTD might serve as a platform for viral transcription and replication [74].

To investigate a FluPol-Pol II complex, we used native Pol II and 1918 FluPol A in cryo-EM studies with the aim to potentially capture an active cap-snatching complex. Cryo-EM analysis of FluPol A bound to vRNA and the native Pol II CTD revealed a structure closely resembling our first structure of 1918 FluPol A bound to a synthetic CTD peptide. The native CTD followed a highly similar binding path to the synthetic CTD peptide in the PA-C domain. However, a slight divergence of the binding path was observed, likely attributed due to the longer nature of the native CTD peptide containing 52 repeats.

Early on, we hypothesised, based on proteomics and biochemical data, that accessory factors such as DSIF and NELF would stabilise the cap-snatching complex due to their crucial role in promoter-proximal pausing [40]. Although capturing the structure of a complete FluPol-Pol II or FluPol-Pol II-DSIF complex was not achieved due to various challenges, a recent study by Rotsch et al.(2024)

published in the final stages of writing this thesis confirmed our hypothesis that a cap-snatching complex requires an interaction of Pol II and the DSIF complex and takes place during promoter-proximal pausing [300]. The study presents a structure of FluPol bound to a transcribing Pol II-DSIF complex and reveals that this FluPol-Pol II-DSIF complex is further stabilised by the presence of NELF [300]. In addition, new binding sites were observed such as the PA endonuclease domain of FluPol contacting Pol II close to the RNA exit channel and the KOWx-4-KOW5 linker of Spt5. Phosphorylation of this linker by CDK9 would release it into active elongation and be sterically incompatible with FluPol binding, indicating that FluPol might prevent Pol II from transitioning into pause release and active elongation [300, 301]. The PB2 cap-binding domain was shown to interact with the RPB1, RPB3, and RPB11 subunits of Pol II with the interaction site located below the Pol II RNA exit channel. Additionally, mutation of critical residues in the interface between FluPol and the Pol II-DSIF complex resulted in reduced FluPol activity in cell-based minigenome assays [300].

Overall, this breakthrough opens new opportunities to study cap-snatching in the context of vRNPs and to answer the question of the state of FluPol after cap-snatching, including whether it remains bound to the Pol II surface and the CTD, how it might dissociate from Pol II and how it transitions from viral transcription to replication.

Further experiments focused on host factors such as Pol II and ANP32 and viral NEP have been suggested to finetune the spatiotemporal regulation of FluPol throughout the viral life cycle. Functional assays showed that ANP32B and the CTD have an overlapping binding site on FluPol, directing FluPol to perform either transcription or replication. In contrast, NEP, essential for vRNP export, did not compete with the Pol II CTD in binding assays, indicating that its role in influencing viral transcription and replication may be independent of the FluPol-Pol II interaction. The Spt5 subunit of DSIF was also shown to

bind to FluPol. However, this could not be confirmed using full-length DSIF in this project. Assessing our $Spt5_{CTR}$ interaction experiments in light of the recent FluPol-DSIF-Pol II structure by explains our inconclusive results and lack of binding to FluPol in interaction experiments. Spt5 was shown to interact with FluPol via a previously unresolved linker between the KOWx-4 and KOW5 domains of Spt5 (residues 647-703), which binds directly to the 5' end of the vRNA and the PB2 cap-binding domain [300]. Our truncated Spt5 construct missed this linker and only started at the KOW5 motif at its N-terminus. Given the multidomain nature of the DSIF subunits, particularly Spt5, and the various interactions of these domains with the Pol II core, an interaction of FluPol and DSIF would be more stable in the presence of Pol II. However, previous work has shown that FluPol and DSIF alone can form a complex in SEC experiments [261]. These findings suggest a complex regulatory network that requires various host factors to modulate FluPol function during different stages of the viral life cycle.

Additionally, I investigated how different phosphorylation combinations on the Pol II CTD affect FluPol binding. While FluPol has been shown to directly bind to pS5 Pol II CTD, the influence of other phosphorylated residues in the CTD heptad in combination with pS5 remained unclear [41]. As the CTD is subject to various post-translational modifications that modulate the function of Pol II and its associated machinery throughout transcription, the experiments in this work provided a more nuanced understanding of the specificity of FluPol to pS5 CTD [112, 153]. CTD binding experiments demonstrated that while phosphorylation at S5 of the CTD is essential for strong FluPol binding, additional phosphorylation at S2 or S7 of the heptad consensus sequence significantly reduced the ability of FluPol to bind the CTD due to steric clashes. These experiments reinforced that FluPol binds to Pol II at a specific point in the transcription cycle, coinciding with promoter proximal pausing [40, 151]. So far, it remains unknown how CTD phosphorylation patterns might be disturbed due to influenza virus infection, particularly as Pol II degradation progresses during later stages

of infection [67, 210, 211]. In addition, as a study suggested that the CTD might promote viral replication by providing a platform for both transcription and replication, it would be interesting to study whether replicating and transcribing FluPol have different affinities to different phosphorylation combinations and post-translational modifications of the Pol II CTD [74]. Disturbed Pol II transcription and the progressive degradation of Pol II might promote, together with the increase in newly synthesised viral proteins, the release of vRNPs from chromatin and their export from the nucleus to the plasma membrane for packaging [44, 74]. Further development and optimisation of cryo-electron tomography and other high-resolution imaging techniques will allow insights into the function of FluPol in the context of a full vRNP as well as in infected cells.

In summary, this thesis has investigated the molecular mechanisms by which influenza viruses exploit the host transcription machinery to transcribe their viral genome. The structural and biochemical analyses presented here provide new insights into the FluPol-Pol II interaction and how this interaction is influenced by host factors and the mutation of conserved residues in the binding interface. While questions remain regarding the spatiotemporal regulation of viral transcription and replication, especially in the context of vRNPs, this work has provided the groundwork for future investigations in viral transcription of influenza viruses and the development of antivirals targeting the FluPol-Pol II binding interface.

References

1. Krammer, F. *et al.* Influenza. *Nature Reviews Disease Primers* **4** (2018).
2. Sellers, S. A., Hagan, R. S., Hayden, F. G. & Fischer, W. A. The hidden burden of influenza: A review of the extra-pulmonary complications of influenza infection. *Influenza and Other Respiratory Viruses* **11,** 372–393 (2017).
3. Flerlage, T., Boyd, D. F., Meliopoulos, V., Thomas, P. G. & Schultz-Cherry, S. Influenza virus and SARS-CoV-2: pathogenesis and host responses in the respiratory tract. *Nature Reviews Microbiology* **19,** 425–441 (2021).
4. Tyrrell, C. S., Allen, J. L. Y. & Gkrania-Klotsas, E. Influenza: epidemiology and hospital management. *Medicine* **49,** 797–804 (2021).
5. Taubenberger, J. K. & Morens, D. M. 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases* **12,** 15–22 (2006).
6. Spreeuwenberg, P., Kroneman, M. & Paget, J. Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic. *American Journal of Epidemiology* **187,** 2561–2567 (2018).
7. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiological reviews* **56,** 152–179 (1992).
8. Treanor, J. Influenza Vaccine — Outmaneuvering Antigenic Shift and Drift. *New England Journal of Medicine* **350,** 218–220 (2004).
9. Wei, C.-J. *et al.* Next-generation influenza vaccines: opportunities and challenges. *Nature Reviews Drug Discovery* **19,** 239–252 (2020).
10. Herfst, S. *et al.* Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets. *Science* **336,** 1534–1541 (2012).
11. Abdelwhab, E.-S. M., Veits, J. & Mettenleiter, T. C. Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry. *Virulence* **4,** 441–452 (2013).
12. Taubenberger, J. K. & Kash, J. C. Influenza Virus Evolution, Host Adaptation, and Pandemic Formation. *Cell Host & Microbe* **7,** 440–451 (2010).
13. Tscherne, D. M. & García-Sastre, A. Virulence determinants of pandemic influenza viruses. *Journal of Clinical Investigation* **121,** 6–13 (2011).
14. Horimoto, T. & Kawaoka, Y. Influenza: lessons from past pandemics, warnings from current incidents. *Nature Reviews Microbiology* **3,** 591–600 (2005).
15. Mostafa, A., Abdelwhab, E., Mettenleiter, T. & Pleschka, S. Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses* **10,** 497 (2018).

16. Yuan, S., Wen, L. & Zhou, J. Inhibitors of Influenza A Virus Polymerase. *ACS Infectious Diseases* **4,** 218–223 (2018).

17. McMillan, C. L., Young, P. R., Watterson, D. & Chappell, K. J. The Next Generation of Influenza Vaccines: Towards a Universal Solution. *Vaccines* **9,** 26 (2021).

18. Gubareva, L. V., Kaiser, L. & Hayden, F. G. Influenza virus neuraminidase inhibitors. *The Lancet* **355,** 827–835 (2000).

19. Mtambo, S. E. *et al.* Influenza Viruses: Harnessing the Crucial Role of the M2 Ion-Channel and Neuraminidase toward Inhibitor Design. *Molecules* **26,** 880 (2021).

20. Dufrasne, F. Baloxavir Marboxil: An Original New Drug against Influenza. *Pharmaceuticals* **15,** 28 (2021).

21. Hayden, F. G. & Shindo, N. Influenza virus polymerase inhibitors in clinical development. *Current Opinion in Infectious Diseases* **32,** 176–186 (2019).

22. Clark, M. P. *et al.* Discovery of a Novel, First-in-Class, Orally Bioavailable Azaindole Inhibitor (VX-787) of Influenza PB2. *Journal of Medicinal Chemistry* **57,** 6668–6678 (2014).

23. Hause, B. M. *et al.* Characterization of a Novel Influenza Virus in Cattle and Swine: Proposal for a New Genus in the Orthomyxoviridae Family. *mBio* **5** (2014).

24. Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology* **16,** 47–60 (2017).

25. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology* **17,** 67–81 (2018).

26. Krischuns, T. *et al.* Type B and type A influenza polymerases have evolved distinct binding interfaces to recruit the RNA polymerase II CTD. *PLOS Pathogens* **18,** e1010328 (2022).

27. Baker, S. F. *et al.* Influenza A and B Virus Intertypic Reassortment through Compatible Viral Packaging Signals. *Journal of Virology* **88,** 10778–10791 (2014).

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

29. Paget, J., Caini, S., Del Riccio, M., van Waarden, W. & Meijer, A. Has influenza B/Yamagata become extinct and what implications might this have for quadrivalent influenza vaccines? *Eurosurveillance* **27** (2022).

30. Lei, N. *et al.* Molecular evolution of influenza B virus during 2011–2017 in Chaoyang, Beijing, suggesting the free influenza vaccine policy. *Scientific Reports* **9** (2019).

31. Wang, M. & Veit, M. Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. *Protein & Cell* **7,** 28–45 (2015).

32. Salem, E. *et al.* Pathogenesis, Host Innate Immune Response, and Aerosol Transmission of Influenza D Virus in Cattle. *Journal of Virology* **93** (2019).
33. Badham, M. D. & Rossman, J. S. Filamentous Influenza Viruses. *Current Clinical Microbiology Reports* **3,** 155–161 (2016).
34. Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D. & Hutchinson, E. C. Filamentous influenza viruses. *Journal of General Virology* **97,** 1755–1764 (2016).
35. Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H. & Barman, S. Influenza virus morphogenesis and budding. *Virus Research* **143,** 147–161 (2009).
36. Rossman, J. S. & Lamb, R. A. Influenza virus assembly and budding. *Virology* **411,** 229–236 (2011).
37. Hutchinson, E. C. *et al.* Conserved and host-specific features of influenza virion architecture. *Nature Communications* **5** (2014).
38. Hale, B. G., Randall, R. E., Ortín, J. & Jackson, D. The multifunctional NS1 protein of influenza A viruses. *Journal of General Virology* **89,** 2359–2376 (2008).
39. Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* **26,** D49–D53 (2008).
40. Krischuns, T., Lukarska, M., Naffakh, N. & Cusack, S. Influenza Virus RNA-Dependent RNA Polymerase and the Host Transcriptional Apparatus. *Annual Review of Biochemistry, Vol 90, 2021* **90,** 321–348 (2021).
41. Walker, A. P. & Fodor, E. Interplay between Influenza Virus and the Host RNA Polymerase II Transcriptional Machinery. *Trends Microbiol* **27,** 398–407 (2019).
42. Jakob, C., Paul-Stansilaus, R., Schwemmle, M., Marquet, R. & Bolte, H. The influenza A virus genome packaging network — complex, flexible and yet unsolved. *Nucleic Acids Research* **50,** 9023–9038 (2022).
43. Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nature Reviews Microbiology* **13,** 28–41 (2014).
44. Zhu, Z., Fodor, E. & Keown, J. R. A structural understanding of influenza virus genome replication. *Trends in Microbiology* **31,** 308–319 (2023).
45. Luo, G. X., Luytjes, W., Enami, M. & Palese, P. The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. *Journal of Virology* **65,** 2861–2867 (1991).
46. Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences* **103,** 19123–19127 (2006).
47. Noda, T. *et al.* Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439,** 490–492 (2006).
48. Calder, L. J., Wasilewski, S., Berriman, J. A. & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proceedings of the National Academy of Sciences* **107,** 10685–10690 (2010).
49. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *Journal of General Virology* **93,** 1385–1409 (2012).

50. Vasin, A. *et al.* Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Research* **185,** 53–63 (2014).

51. Pinto, R. M., Lycett, S., Gaunt, E. & Digard, P. Accessory Gene Products of Influenza A Virus. *Cold Spring Harbor Perspectives in Medicine* **11,** a038380 (2020).

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

53. Dubois, J., Terrier, O. & Rosa-Calatrava, M. Influenza Viruses and mRNA Splicing: Doing More with Less. *mBio* **5** (2014).

54. Ferguson, L. *et al.* Pathogenesis of Influenza D Virus in Cattle. *Journal of Virology* **90,** 5636–5642 (2016).

55. Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E. & Palese, P. Cellular Proteins in Influenza Virus Particles. *PLoS Pathogens* **4,** e1000085 (2008).

56. Matsuoka, Y. *et al.* A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology* **7** (2013).

57. York, A., Hutchinson, E. C. & Fodor, E. Interactome Analysis of the Influenza A Virus Transcription/Replication Machinery Identifies Protein Phosphatase 6 as a Cellular Factor Required for Efficient Virus Replication. *Journal of Virology* **88,** 13284–13299 (2014).

58. Haas, K. M. *et al.* Proteomic and genetic analyses of influenza A viruses identify pan-viral host targets. *Nature Communications* **14** (2023).

59. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology* **9** (2018).

60. Stegmann, T., Morselt, H. W., Scholma, J. & Wilschut, J. Fusion of influenza virus in an intracellular acidic compartment measured by fluorescence dequenching. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **904,** 165–170 (1987).

61. Loregian, A., Mercorelli, B., Nannetti, G., Compagnin, C. & Palù, G. Antiviral strategies against influenza virus: towards new therapeutic approaches. *Cellular and Molecular Life Sciences* **71,** 3659–3683 (2014).

62. Fodor, E. & Brownlee, G. G. in *Influenza* 1–29 (Elsevier, 2002).

63. Manzoor, R., Igarashi, M. & Takada, A. Influenza A Virus M2 Protein: Roles from Ingress to Egress. *International Journal of Molecular Sciences* **18,** 2649 (2017).

64. Miyake, Y. *et al.* Influenza virus uses transportin 1 for vRNP debundling during cell entry. *Nature Microbiology* **4,** 578–586 (2019).

65. Te Velthuis, A. J. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology* **14,** 479–493 (2016).

66. Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S. & Cusack, S. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Research* **234,** 103–117 (2017).

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

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

69. Neumann, G. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *The EMBO Journal* **19,** 6751–6758 (2000).

70. Amorim, M. J. *et al.* A Rab11- and Microtubule-Dependent Mechanism for Cytoplasmic Transport of Influenza A Virus Viral RNA. *Journal of Virology* **85,** 4143–4156 (2011).

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

72. Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nature Microbiology* **4,** 1781–1789 (2019).

73. Wandzik, J. M., Kouba, T. & Cusack, S. Structure and Function of Influenza Polymerase. *Cold Spring Harbor Perspectives in Medicine* **11** (2021).

74. Krischuns, T. *et al.* The host RNA polymerase II C-terminal domain is the anchor for replication of the influenza virus genome. *Nature Communications* **15** (2024).

75. Kouba, T., Drncová, P. & Cusack, S. Structural snapshots of actively transcribing influenza polymerase. *Nature Structural & Molecular Biology* **26,** 460–470 (2019).

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

77. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516,** 355–360 (2014).

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

79. Chu, C. *et al.* Functional Analysis of Conserved Motifs in Influenza Virus PB1 Protein. *PLoS ONE* **7,** e36113 (2012).

80. Chaturvedi, U. C. & Shrivastava, R. Interaction of viral proteins with metal ions: role in maintaining the structure and functions of viruses. *FEMS Immunology & Medical Microbiology* **43,** 105–114 (2005).

81. Carvalho, A. T. P., Fernandes, P. A. & Ramos, M. J. The Catalytic Mechanism of RNA Polymerase II. *Journal of Chemical Theory and Computation* **7,** 1177–1188 (2011).

82. Svetlov, V. & Nudler, E. Basic mechanism of transcription by RNA polymerase II. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829,** 20–28 (2013).

83. Te Velthuis, A. J. W., Robb, N. C., Kapanidis, A. N. & Fodor, E. The role of the priming loop in influenza A virus RNA synthesis. *Nature Microbiology* **1** (2016).

84. Deng, T. *et al.* Role of Ran Binding Protein 5 in Nuclear Import and Assembly of the Influenza Virus RNA Polymerase Complex. *Journal of Virology* **80,** 11911–11919 (2006).

85. Gabriel, G. *et al.* Differential use of importin- isoforms governs cell tropism and host adaptation of influenza virus. *Nature Communications* **2** (2011).

86. Hutchinson, E. C., Orr, O. E., Man Liu, S., Engelhardt, O. G. & Fodor, E. Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *Journal of General Virology* **92,** 1859–1869 (2011).

87. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–918 (2009).

88. DuBois, R. M. *et al.* Structural and Biochemical Basis for Development of Influenza Virus Inhibitors Targeting the PA Endonuclease. *PLoS Pathogens* **8,** e1002830 (2012).

89. Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. *Nature* **458,** 909–913 (2009).

90. Crépin, T. *et al.* Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit. *Journal of Virology* **84,** 9096–9104 (2010).

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

92. Kowalinski, E. *et al.* Structural Analysis of Specific Metal Chelating Inhibitor Binding to the Endonuclease Domain of Influenza pH1N1 (2009) Polymerase. *PLoS Pathogens* **8,** e1002831 (2012).

93. Noshi, T. *et al.* In vitro characterization of baloxavir acid, a first-in-class cap-dependent endonuclease inhibitor of the influenza virus polymerase PA subunit. *Antiviral Research* **160,** 109–117 (2018).

94. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454,** 1127–1131 (2008).

95. He, X. *et al.* Crystal structure of the polymerase PAC–PB1N complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).

96. Lukarska, M. *et al.* Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **541,** 117–121 (2017).

97. Serna Martin, I. *et al.* A Mechanism for the Activation of the Influenza Virus Transcriptase. *Mol Cell* **70,** 1101–1110.e4 (2018).

98. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology* **15,** 500–506 (2008).

99. Liu, Y. *et al.* Structural and Functional Characterization of K339T Substitution Identified in the PB2 Subunit Cap-binding Pocket of Influenza A Virus. *Journal of Biological Chemistry* **288,** 11013–11023 (2013).

100. Wakai, C., Iwama, M., Mizumoto, K. & Nagata, K. Recognition of Cap Structure by Influenza B Virus RNA Polymerase Is Less Dependent on the Methyl Residue than Recognition by Influenza A Virus Polymerase. *Journal of Virology* **85,** 7504–7512 (2011).

101. Liu, Y. *et al.* The Crystal Structure of the PB2 Cap-binding Domain of Influenza B Virus Reveals a Novel Cap Recognition Mechanism. *Journal of Biological Chemistry* **290,** 9141–9149 (2015).

102. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516,** 361–366 (2014).

103. Tsurumura, T. *et al.* Conformational Polymorphism of m7GTP in Crystal Structure of the PB2 Middle Domain from Human Influenza A Virus. *PLoS ONE* **8,** e82020 (2013).

104. Wandzik, J. M. *et al.* A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase. *Cell* **181,** 877–893.e21 (2020).

105. Garg, G. *et al.* Structural insights into human co-transcriptional capping. *Molecular Cell* **83,** 2464–2477.e5 (2023).

106. Long, J. S. *et al.* Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* **529,** 101–104 (2016).

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

108. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of Virology* **67,** 1761–1764 (1993).

109. Yamada, S. *et al.* Biological and Structural Characterization of a Host-Adapting Amino Acid in Influenza Virus. *PLoS Pathogens* **6,** e1001034 (2010).

110. Kuzuhara, T. *et al.* Structural Basis of the Influenza A Virus RNA Polymerase PB2 RNA-binding Domain Containing the Pathogenicity-determinant Lysine 627 Residue. *Journal of Biological Chemistry* **284,** 6855–6860 (2009).

111. Tarendeau, F. *et al.* Host Determinant Residue Lysine 627 Lies on the Surface of a Discrete, Folded Domain of Influenza Virus Polymerase PB2 Subunit. *PLoS Pathogens* **4,** e1000136 (2008).

112. Keown, J. *et al.* Structural and functional characterisation of the interaction between the influenza A virus RNA polymerase and the CTD of host RNA Polymerase II (2023).

113. Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Structural & Molecular Biology* **14,** 229–233 (2007).

114. Pumroy, R. A., Ke, S., Hart, D. J., Zachariae, U. & Cingolani, G. Molecular Determinants for Nuclear Import of Influenza A PB2 by Importin Isoforms 3 and 7. *Structure* **23,** 374–384 (2015).

115. Delaforge, E. *et al.* Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin. *Journal of the American Chemical Society* **137,** 15122–15134 (2015).

116. Thierry, E. *et al.* Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Molecular Cell* **61,** 125–137 (2016).

117. Hengrung, N. *et al.* Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527,** 114–117 (2015).

118. Tomescu, A. I., Robb, N. C., Hengrung, N., Fodor, E. & Kapanidis, A. N. Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proceedings of the National Academy of Sciences* **111** (2014).

119. Pritlove, D. C., Poon, L. L. M., Devenish, L. J., Leahy, M. B. & Brownlee, G. G. A Hairpin Loop at the 5 End of Influenza A Virus Virion RNA Is Required for Synthesis of Poly(A) + mRNA In Vitro. *Journal of Virology* **73,** 2109–2114 (1999).

120. Leahy, M. B., Pritlove, D. C., Poon, L. L. M. & Brownlee, G. G. Mutagenic Analysis of the 5 Arm of the Influenza A Virus Virion RNA Promoter Defines the Sequence Requirements for Endonuclease Activity. *Journal of Virology* **75,** 134–142 (2001).

121. Peng, Q. *et al.* Structural insight into RNA synthesis by influenza D polymerase. *Nature Microbiology* **4,** 1750–1759 (2019).

122. Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* **573,** 287–290 (2019).

123. Robb, N. C., te Velthuis, A. J. W., Fodor, E. & Kapanidis, A. N. Real-time analysis of single influenza virus replication complexes reveals large promoter-dependent differences in initiation dynamics. *Nucleic Acids Research* **47,** 6466–6477 (2019).

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

125. Engelhardt, O. G., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *Journal of virology* **79,** 5812–5818 (2005).

126. Martínez-Alonso, M., Hengrung, N. & Fodor, E. 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 (2016).

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

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

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

130. Gu, W. *et al.* Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21,** 2067–2075 (2015).

131. Te Velthuis, A. J. W. & Oymans, J. Initiation, Elongation, and Realignment during Influenza Virus mRNA Synthesis. *Journal of Virology* **92** (2018).

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

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

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

135. Pflug, A. *et al.* Capped RNA primer binding to influenza polymerase and implications for the mechanism of cap-binding inhibitors. *Nucleic Acids Research* **46,** 956–971 (2018).

136. Poon, L. L. M., Pritlove, D. C., Fodor, E. & Brownlee, G. G. 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. *Journal of Virology* **73,** 3473–3476 (1999).

137. Amorim, M.-J., Read, E. K., Dalton, R. M., Medcalf, L. & Digard, P. Nuclear Export of Influenza A Virus mRNAs Requires Ongoing RNA Polymerase II Activity. *Traffic* **8,** 1–11 (2006).

138. Read, E. K. C. & Digard, P. Individual influenza A virus mRNAs show differential dependence on cellular NXF1/TAP for their nuclear export. *Journal of General Virology* **91,** 1290–1301 (2010).

139. Nemeroff, M. E., Barabino, S. M., Li, Y., Keller, W. & Krug, R. M. Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3 End Formation of Cellular Pre-mRNAs. *Molecular Cell* **1,** 991–1000 (1998).

140. Noah, D. L., Twu, K. Y. & Krug, R. M. Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3 end processing of cellular pre-mRNAS. *Virology* **307,** 386–395 (2003).

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

142. Jensen, S. & Thomsen, A. R. Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *Journal of Virology* **86,** 2900–2910 (2012).

143. Jorba, N., Coloma, R. & Ortín, J. Genetic trans-Complementation Establishes a New Model for Influenza Virus RNA Transcription and Replication. *PLoS Pathogens* **5,** e1000462 (2009).

144. York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T. & Fodor, E. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the National Academy of Sciences* **110** (2013).

145. Carrique, L. *et al.* Host ANP32A mediates the assembly of the influenza virus replicase. *Nature* **587,** 638–643 (2020).

146. Long, J. S. *et al.* Species specific differences in use of ANP32 proteins by influenza A virus. *eLife* **8** (2019).

147. Peacock, T. P., Sheppard, C. M., Staller, E. & Barclay, W. S. Host Determinants of Influenza RNA Synthesis. *Annual Review of Virology* **6,** 215–233 (2019).

148. Robb, N. C., Smith, M., Vreede, F. T. & Fodor, E. NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *Journal of General Virology* **90,** 1398–1407 (2009).

149. Phan, T. *et al.* Segment-Specific Kinetics of mRNA, cRNA, and vRNA Accumulation during Influenza Virus Infection. *Journal of Virology* **95** (2021).

150. Keown, J. R. *et al.* Mapping inhibitory sites on the RNA polymerase of the 1918 pandemic influenza virus using nanobodies. *Nature Communications* **13** (2022).

151. Osman, S. & Cramer, P. Structural Biology of RNA Polymerase II Transcription: 20 Years On. *Annual Review of Cell and Developmental Biology* **36,** 1–34 (2020).

152. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* **18,** 263–273 (2017).

153. Zaborowska, J., Egloff, S. & Murphy, S. The Pol II CTD: new twists in the tail. *Nat Struct Mol Biol* **23,** 771–7 (2016).

154. Eick, D. & Geyer, M. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chemical Reviews* **113,** 8456–8490 (2013).

155. Buratowski, S. The CTD code. *Nature Structural amp; Molecular Biology* **10,** 679–680 (2003).

156. Corden, J. L. RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. *Chemical Reviews* **113,** 8423–8455 (2013).

157. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes amp; Development* **19,** 1401–1415 (2005).

158. Jasnovidova, O. & Stefl, R. The CTD code of RNA polymerase II: a structural view. *WIREs RNA* **4,** 1–16 (2013).

159. Pluta, A. J., Studniarek, C., Murphy, S. & Norbury, C. J. Cyclin-dependent kinases: Masters of the eukaryotic universe. *WIREs RNA* **15** (2023).

160. Nojima, T. *et al.* RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell* **72,** 369–379.e4 (2018).

161. Heidemann, M., Hintermair, C., Voß, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829,** 55–62 (2013).

162. Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3 End Processing. *Molecular Cell* **13,** 67–76 (2004).

163. Komarnitsky, P., Cho, E.-J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes amp; Development* **14,** 2452–2460 (2000).

164. Cho, E.-J., Kobor, M. S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes amp; Development* **15,** 3319–3329 (2001).

165. Yurko, N. M. & Manley, J. L. The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. *Transcription* **9,** 30–40 (2017).

166. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T. & Murphy, S. Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. *Molecular Cell* **45,** 111–122 (2012).

167. Buratowski, S. Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell* **36,** 541–546 (2009).

168. Mayfield, J. E. *et al.* Tyr1 phosphorylation promotes phosphorylation of Ser2 on the C-terminal domain of eukaryotic RNA polymerase II by P-TEFb. *eLife* **8** (2019).

169. Descostes, N. *et al.* Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* **3** (2014).

170. Mayer, A. *et al.* CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science* **336,** 1723–1725 (2012).

171. Cramer, P. Organization and regulation of gene transcription. *Nature* **573,** 45–54 (2019).

172. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560,** 601–606 (2018).

173. Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences* **30,** 235–239 (2005).

174. Eychenne, T. *et al.* Functional interplay between Mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. *Genes amp; Development* **30,** 2119–2132 (2016).

175. Nozawa, K., Schneider, T. R. & Cramer, P. Core Mediator structure at 3.4 Å extends model of transcription initiation complex. *Nature* **545,** 248–251 (2017).

176. Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proceedings of the National Academy of Sciences* **94,** 6075–6078 (1997).

177. Luse, D. S. Promoter clearance by RNA polymerase II. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829,** 63–68 (2013).

178. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics* **13,** 720–731 (2012).

179. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nature Structural & Molecular Biology* **24,** 809–815 (2017).

180. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357,** 921–924 (2017).

181. Hartzog, G. A. & Fu, J. The Spt4–Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829,** 105–115 (2013).

182. Gonatopoulos-Pournatzis, T. & Cowling, V. H. Cap-binding complex (CBC). *Biochemical Journal* **457,** 231–242 (2013).

183. Shimotohno, K., Kodama, Y., Hashimoto, J. & Miura, K. I. Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proceedings of the National Academy of Sciences* **74,** 2734–2738 (1977).

184. Mandal, S. S. *et al.* Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proceedings of the National Academy of Sciences* **101,** 7572–7577 (2004).

185. Schneider, S., Pei, Y., Shuman, S. & Schwer, B. Separable Functions of the Fission Yeast Spt5 Carboxyl-Terminal Domain (CTD) in Capping Enzyme Binding and Transcription Elongation Overlap with Those of the RNA Polymerase II CTD. *Molecular and Cellular Biology* **30,** 2353–2364 (2010).

186. Ramanathan, A., Robb, G. B. & Chan, S.-H. mRNA capping: biological functions and applications. *Nucleic Acids Research* **44,** 7511–7526 (2016).

187. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics* **15,** 163–175 (2014).

188. Fianu, I. *et al.* Structural basis of Integrator-dependent RNA polymerase II termination. *Nature* **629,** 219–227 (2024).

189. Decker, T.-M. Mechanisms of Transcription Elongation Factor DSIF (Spt4–Spt5). *Journal of Molecular Biology* **433,** 166657 (2021).

190. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560,** 607–612 (2018).

191. Lu, X. *et al.* Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. *Nucleic Acids Research* **44,** 6853–6867 (2016).

192. Chen, Y. *et al.* DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes amp; Development* **23,** 2765–2777 (2009).

193. Zhou, Q. Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages. *The EMBO Journal* **17,** 3681–3691 (1998).

194. Whelan, M. & Pelchat, M. Role of RNA Polymerase II Promoter-Proximal Pausing in Viral Transcription. *Viruses* **14,** 2029 (2022).

195. Engelhardt, O. G., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *Journal of virology* **79,** 5812–5818 (2005).

196. Bradel-Tretheway, B. G. *et al.* Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *Journal of virology* **85,** 8569–8581 (2011).

197. Heaton, N. S. *et al.* Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* **44,** 46–58 (2016).

198. Hubel, P. *et al.* A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nature immunology* **20,** 493–502 (2019).

199. Tripathi, S. *et al.* Meta-and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell host & microbe* **18,** 723–735 (2015).

200. Loucaides, E. M. *et al.* Nuclear dynamics of influenza A virus ribonucleoproteins revealed by live-cell imaging studies. *Virology* **394,** 154–163 (2009).

201. Bortz, E. *et al.* Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *mBio* **2,** e00151–11 (2011).

202. Westera, L. *et al.* Poly-ADP Ribosyl Polymerase 1 (PARP1) Regulates Influenza A Virus Polymerase. *Advances in virology* **2019,** 8512363–8512363 (2019).

203. Tafforeau, L. *et al.* Generation and Comprehensive Analysis of an Influenza Virus Polymerase Cellular Interaction Network. *Journal of Virology* **85,** 13010–13018 (2011).

204. Gibson, B. A. *et al.* Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science* **353,** 45–50 (2016).

205. Mayer, D. *et al.* Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches. *Journal of proteome research* **6,** 672–682 (2007).

206. Naito, T. *et al.* An influenza virus replicon system in yeast identified Tat-SF1 as a stimulatory host factor for viral RNA synthesis. *Proceedings of the National Academy of Sciences* **104,** 18235–18240 (2007).

207. Wang, Z. *et al.* Establishment of a high-throughput assay to monitor influenza A virus RNA transcription and replication. *PLoS One* **10,** e0133558 (2015).

208. Li, B. *et al.* Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nature communications* **11,** 164 (2020).

209. Bauer, D. L. V. *et al.* Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Reports* **23,** 2119–2129 (2018).

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

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

212. Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *eLife* **7** (2018).

213. Levene, R. & Gaglia, M. Host Shutoff in Influenza A Virus: Many Means to an End. *Viruses* **10,** 475 (2018).

214. Tumpey, T. M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310,** 77–80 (2005).

215. Fodor, E., Mingay, L. J., Crow, M., Deng, T. & Brownlee, G. G. A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA Polymerase Promotes the Generation of Defective Interfering RNAs. *Journal of Virology* **77,** 5017–5020 (2003).

216. Fodor, E. *et al.* Rescue of Influenza A Virus from Recombinant DNA. *Journal of Virology* **73,** 9679–9682 (1999).

217. Mänz, B. *et al.* Multiple Natural Substitutions in Avian Influenza A Virus PB2 Facilitate Efficient Replication in Human Cells. *Journal of Virology* **90,** 5928–5938 (2016).

218. Fodor, E. *et al.* A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA Polymerase Inhibits Endonucleolytic Cleavage of Capped RNAs. *Journal of Virology* **76,** 8989–9001 (2002).

219. Wang, F. *et al.* The C-terminal LCAR of host ANP32 proteins interacts with the influenza A virus nucleoprotein to promote the replication of the viral RNA genome. *Nucleic Acids Research* **50,** 5713–5725 (2022).

220. Veler, H. *et al.* The C-Terminal Domains of the PB2 Subunit of the Influenza A Virus RNA Polymerase Directly Interact with Cellular GTPase Rab11a. *Journal of Virology* **96** (2022).

221. Tannock, G. A., Paul, J. A. & Barry, R. D. Relative immunogenicity of the cold-adapted influenza virus A/Ann Arbor/6/60 (A/AA/6/60-ca), recombinants of A/AA/6/60-ca, and parental strains with similar surface antigens. *Infection and Immunity* **43,** 457–462 (1984).

222. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41,** 207–234 (2005).

223. Boehning, M. *et al.* RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural & Molecular Biology* **25,** 833–840 (2018).

224. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529,** 551–554 (2016).

225. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005).

226. Florens, L. *et al.* Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40,** 303–311 (2006).

227. Cassonnet, P. *et al.* Benchmarking a luciferase complementation assay for detecting protein complexes. *Nature Methods* **8,** 990–992 (2011).

228. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30,** 70–82 (2021).

229. Mastronarde, D. N. SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microscopy and Microanalysis* **9,** 1182–1183 (2003).

230. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* **14,** 290–296 (2017).

231. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature methods* **17,** 1214–1221 (2020).

232. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology* **74,** 531–544 (2018).

233. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications biology* **4,** 874 (2021).

234. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography* **60,** 2256–2268 (2004).

235. Zhu, X. *et al.* Hemagglutinin homologue from H17N10 bat influenza virus exhibits divergent receptor-binding and pH-dependent fusion activities. *Proceedings of the National Academy of Sciences* **110,** 1458–1463 (2013).

236. Yang, W., Schountz, T. & Ma, W. Bat Influenza Viruses: Current Status and Perspective. *Viruses* **13,** 547 (2021).

237. Morel, J. *et al.* The Influenza Virus RNA-Polymerase and the Host RNA-Polymerase II: RPB4 Is Targeted by a PB2 Domain That Is Involved in Viral Transcription. *Viruses* **14,** 518 (2022).

238. Li, H. *et al.* An intermediate state allows influenza polymerase to switch smoothly between transcription and replication cycles. *Nature Structural amp; Molecular Biology* **30,** 1183–1192 (2023).

239. Te Velthuis, A. J. W., Grimes, J. M. & Fodor, E. Structural insights into RNA polymerases of negative-sense RNA viruses. *Nature Reviews Microbiology* **19,** 303–318 (2021).

240. Yángüez, E. & Nieto, A. So similar, yet so different: Selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell. *Virus Research* **156,** 1–12 (2011).

241. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. The Persistent Legacy of the 1918 Influenza Virus. *New England Journal of Medicine* **361,** 225–229 (2009).

242. Liang, Y., Danzy, S., Dao, L. D., Parslow, T. G. & Liang, Y. Mutational Analyses of the Influenza A Virus Polymerase Subunit PA Reveal Distinct Functions Related and Unrelated to RNA Polymerase Activity. *PLoS ONE* **7,** e29485 (2012).

243. Liu, L. *et al.* Phosphorylation of the PA subunit of influenza polymerase at Y393 prevents binding of the 5-termini of RNA and polymerase function. *Scientific Reports* **13** (2023).

244. Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. 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. *Journal of Virology* **80,** 7789–7798 (2006).

245. Perales, B. *et al.* The Replication Activity of Influenza Virus Polymerase Is Linked to the Capacity of the PA Subunit To Induce Proteolysis. *Journal of Virology* **74,** 1307–1312 (2000).

246. Fodor, E. & Smith, M. The PA Subunit Is Required for Efficient Nuclear Accumulation of the PB1 Subunit of the Influenza A Virus RNA Polymerase Complex. *Journal of Virology* **78,** 9144–9153 (2004).

247. Deng, T., Sharps, J., Fodor, E. & Brownlee, G. G. 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. *Journal of Virology* **79,** 8669–8674 (2005).

248. Staller, E. *et al.* Structures of H5N1 influenza polymerase with ANP32B reveal mechanisms of genome replication and host adaptation. *Nature Communications* **15** (2024).

249. Arragain, B. *et al.* Structures of influenza A and B replication complexes give insight into avian to human host adaptation and reveal a role of ANP32 as an electrostatic chaperone for the apo-polymerase. *Nature Communications* **15** (2024).

250. Staller, E. *et al.* ANP32 Proteins Are Essential for Influenza Virus Replication in Human Cells. *Journal of Virology* **93** (2019).

251. Zhang, H. *et al.* Fundamental Contribution and Host Range Determination of ANP32A and ANP32B in Influenza A Virus Polymerase Activity. *Journal of Virology* **93** (2019).

252. Flores-Solis, D. *et al.* Driving forces behind phase separation of the carboxy-terminal domain of RNA polymerase II. *Nature Communications* **14** (2023).

253. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361,** 412–415 (2018).

254. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572,** 543–548 (2019).

255. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361** (2018).

256. Chandrudu, S., Simerska, P. & Toth, I. Chemical Methods for Peptide and Protein Production. *Molecules* **18,** 4373–4388 (2013).

257. Kozeleková, A. *et al.* Phosphorylated and Phosphomimicking Variants May Differ—A Case Study of 14-3-3 Protein. *Frontiers in Chemistry* **10** (2022).

258. Ramani, M. K. V. *et al.* Structural Motifs for CTD Kinase Specificity on RNA Polymerase II during Eukaryotic Transcription. *ACS Chemical Biology* **15,** 2259–2272 (2020).

259. Rimel, J. K. & Taatjes, D. J. The essential and multifunctional TFIIH complex. *Protein Science* **27,** 1018–1037 (2018).

260. Mosley, A. L. *et al.* Quantitative Proteomics Demonstrates That the RNA Polymerase II Subunits Rpb4 and Rpb7 Dissociate during Transcriptional Elongation. *Molecular amp; Cellular Proteomics* **12,** 1530–1538 (2013).

261. Lukarska, M. *Structural and functional characterization of the interaction between influenza polymerase and the cellular transcription machinery* Thesis (Université Grenoble Alpes, 2018).

262. Schüller, R. *et al.* Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular Cell* **61,** 305–314 (2016).

263. Chang, S. *et al.* Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. *Molecular Cell* **57,** 925–935 (2015).

264. Chen, K.-Y., Santos Afonso, E. D., Enouf, V., Isel, C. & Naffakh, N. Influenza virus polymerase subunits co-evolve to ensure proper levels of dimerization of the heterotrimer. *PLOS Pathogens* **15,** e1008034 (2019).

265. Han, B.-G., Avila-Sakar, A., Remis, J. & Glaeser, R. M. Challenges in making ideal cryo-EM samples. *Current Opinion in Structural Biology* **81,** 102646 (2023).

266. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophysics Reports* **3,** 1–7 (2016).

267. Glaeser, R. M. Preparing Better Samples for Cryo–Electron Microscopy: Biochemical Challenges Do Not End with Isolation and Purification. *Annual Review of Biochemistry* **90,** 451–474 (2021).

268. Rickert, P., Seghezzi, W., Shanahan, F., Cho, H. & Lees, E. H. Cyclin C/CDK8 is a novel CTD kinase associated with RNA polymerase II. *Oncogene* **12,** 2631–2640 (1996).

269. Rickert, P., Corden, J. L. & Lees, E. Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases. *Oncogene* **18,** 1093–1102 (1999).

270. Ramanathan, Y. *et al.* Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences. *Journal of Biological Chemistry* **276,** 10913–10920 (2001).

271. Jasnovidova, O. & Stefl, R. The CTD code of RNA polymerase II: a structural view. *WIREs RNA* **4,** 1–16 (2012).

272. Doamekpor, S. K., Sanchez, A. M., Schwer, B., Shuman, S. & Lima, C. D. How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes amp; Development* **28,** 1323–1336 (2014).

273. Pei, Y. & Shuman, S. Interactions between Fission Yeast mRNA Capping Enzymes and Elongation Factor Spt5. *Journal of Biological Chemistry* **277,** 19639–19648 (2002).

274. Li, Y., Wang, Q., Xu, Y. & Li, Z. Structures of co-transcriptional RNA capping enzymes on paused transcription complex. *Nature Communications* **15** (2024).

275. Zhu, Z., Fan, H. & Fodor, E. Defining the minimal components of the influenza A virus replication machinery via an in vitro reconstitution system. *PLOS Biology* **21,** e3002370 (2023).

276. Dieci, G. Removing quote marks from the RNA polymerase II CTD 'code'. *Biosystems* **207,** 104468 (2021).

277. Suh, H. *et al.* Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular Cell* **61,** 297–304 (2016).

278. Lyons, D. E., McMahon, S. & Ott, M. A combinatorial view of old and new RNA polymerase II modifications. *Transcription* **11,** 66–82 (2020).

279. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes amp; Development* **33,** 960–982 (2019).

280. Serna Martin, I. *Study of the structure and function of the influenza virus RNA-dependent RNA polymerase* Thesis (University of Oxford, 2017).

281. Watanabe, T., Watanabe, S. & Kawaoka, Y. Cellular Networks Involved in the Influenza Virus Life Cycle. *Cell Host amp; Microbe* **7,** 427–439 (2010).

282. Watanabe, T. *et al.* Influenza Virus-Host Interactome Screen as a Platform for Antiviral Drug Development. *Cell Host amp; Microbe* **16,** 795–805 (2014).

283. Bogdanow, B. *et al.* The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. *Nature Communications* **10** (2019).

284. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes amp; Development* **12,** 343–356 (1998).

285. O'Neill, R. E. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *The EMBO Journal* **17,** 288–296 (1998).

286. Martin, K. & Heleniust, A. Nuclear transport of influenza virus ribonucleoproteins: The viral matrix protein (M1) promotes export and inhibits import. *Cell* **67,** 117–130 (1991).

287. Sun, J. *et al.* NS2 induces an influenza A RNA polymerase hexamer and acts as a transcription to replication switch. *EMBO Reports* (2024).

288. Götzke, H. *et al.* The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nature Communications* **10** (2019).

289. Akhuli, D., Dhar, A., Viji, A. S., Bhojappa, B. & Palani, S. ALIBY: ALFA Nanobody-Based Toolkit for Imaging and Biochemistry in Yeast. *mSphere* **7** (2022).

290. Jin, F. *et al.* Fluorescence-detection size-exclusion chromatography utilizing nanobody technology for expression screening of membrane proteins. *Communications Biology* **4** (2021).

291. Zhu, Z. *Towards Understanding the Influenza Virus Ribonucleoprotein Complex: Novel Structural and Functional Approaches* Thesis (University of Oxford, 2024).

292. Brunotte, L. *et al.* The Nuclear Export Protein of H5N1 Influenza A Viruses Recruits Matrix 1 (M1) Protein to the Viral Ribonucleoprotein to Mediate Nuclear Export. *Journal of Biological Chemistry* **289,** 20067–20077 (2014).

293. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology* **5,** 172–172 (2014).

294. Jedlitzke, B. & Mootz, H. D. A Light-Activatable Photocaged Variant of the Ultra-High Affinity ALFA-Tag Nanobody. *ChemBioChem* **23** (2022).

295. Wu, X. & Rapoport, T. A. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proceedings of the National Academy of Sciences* **118** (2021).

296. Rudd-Schmidt, J. A., Laine, R. F., Noori, T., Brennan, A. J. & Voskoboinik, I. ALFA-PRF: a novel approach to detect murine perforin release from CTLs into the immune synapse. *Frontiers in Immunology* **13** (2022).

297. Choi, W., Wu, H., Yserentant, K., Huang, B. & Cheng, Y. Efficient tagging of endogenous proteins in human cell lines for structural studies by single-particle cryo-EM. *Proceedings of the National Academy of Sciences* **120** (2023).

298. Guo, Q. *et al.* A method for structure determination of GPCRs in various states. *Nature Chemical Biology* **20,** 74–82 (2023).

299. Stover, L. *et al.* Grafting the ALFA tag for structural studies of aquaporin Z. *Journal of Structural Biology: X* **9,** 100097 (2024).

300. Rotsch, A. H. *et al.* Mechanism of Co-Transcriptional Cap-Snatching by Influenza Polymerase. *bioRxiv.* eprint: `https://www.biorxiv.org/content/early/2024/08/11/2024.08.11.607481.full.pdf` (2024).

301. Fong, N., Sheridan, R. M., Ramachandran, S. & Bentley, D. L. The pausing zone and control of RNA polymerase II elongation by Spt5: Implications for the pause-release model. *Molecular cell* **82,** 3632–3645 (2022).

Appendices

A
Appendix

Table A.1: Cryo-EM data collection and model refinement statistics.

	1918 vRNA 4rpt peptide	**1918 vRNA 4rpt peptide Ordered PB2**
EMDB ID	EMD-18945	EMD-18947
PDB ID	PDB 8R60	PDB 8R65
Data collection		
Microscope	Titan Krios G3i (OPIC)	
Voltage (kV)	300	
Detector	Gatan K2 with EF	
Recording mode	Counting	
Magnification	130,000	
Movie/micrograph pixel size (Å)	1.05	
Dose rate (e-/px/sec)	9.6	
Number of frames per movie	40	
Movie exposure time (s)	4.2	
Total dose (e-/Å2)	40	
Defocus range (um)	-1.2 to -2.5	
EM data processing		
Number of movies/micrographs	7,921	
Box size (px)	256	
Particle number (total)	4,231,383	
Particle number (used in final map)	303,608	16,121
Symmetry	C1	
Map resolution (Å, FSC 0.143)	3.23	4.23
Map sharpening B-factor (Å2)	167.9	133.5
Model Building and Validation		
Initial model used	7NHX	7NHX/8R60
Non-hydrogen protein atoms	27,951	35,542
Protein residues	1707	2196
Ligands	23	23
B factors (Å2) - mean		
Protein	77	317
Nucleotide	59	169
RMSD from ideal		
Bond length (Å)	0.004	0.005
Bond angles (°)	0.501	0.621
Validation		

Continued on next page

Table A.1 – continued from previous page

	1918 vRNA 4rpt peptide	**1918 vRNA 4rpt peptide ordered PB2**
Molprobity score	2.01	2.23
Clashscore	8.1	11.5
FSC (0.5) model-vs-map	3.5	7.6
CC model-vs-map (masked)	0.79	0.66
Ramachandran plot		
Favored (%)	93.9	92.8
Allowed (%)	6	6.7
Outliers (%)	0.1	0.5