



**HAL**  
open science

# Structural and functional characterization of the interaction between influenza polymerase and the cellular transcription machinery

Mariya Lukarska

► **To cite this version:**

Mariya Lukarska. Structural and functional characterization of the interaction between influenza polymerase and the cellular transcription machinery. Structural Biology [q-bio.BM]. Université Grenoble Alpes, 2018. English. NNT : 2018GREAV031 . tel-02954348

**HAL Id: tel-02954348**

**<https://theses.hal.science/tel-02954348>**

Submitted on 1 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESIS / THÈSE**

To obtain the title of / *Pour obtenir le grade de*

**DOCTEUR DE LA COMMUNAUTE UNIVERSITE  
GRENOBLE ALPES**

Discipline / *Spécialité* : **Structural Biology and Nanobiology**

Arrêté ministériel : 25 mai 2016

Presented by / *Présentée par*

**Maria LUKARSKA**

Thesis supervisor / *Thèse dirigée par* **Dr. Stephen Cusack**

Thesis prepared at / *Thèse préparée au sein du*

**European Molecular Biology Laboratory (EMBL) Grenoble**

In / *Dans l'École Doctorale de Chimie et Sciences du Vivant*

**Structural and functional  
characterization of the interaction  
between influenza polymerase and the  
cellular transcription machinery**

Public defense on / *Thèse soutenue publiquement le* **09/11/18**

Jury members / *devant le jury composé de* :

**Dr. Rob Ruigrok**

Président

**Dr. Ervin Fodor**

Rapporteur

**Dr. Meindert Lamers**

Rapporteur

**Dr. Carrie Bernecky**

Membre

**Dr. Christoph Müller**

Membre



To Donka Nikolova and Svetla Petrova, for sparking the fire.

'I have never known a dull enzyme'  
Arthur Kornberg, 'For the Love of Enzymes'

# Table of Contents

Table of Contents.....	4
List of Abbreviations .....	9
List of Figures .....	13
Abstract.....	16
Résumé en Français .....	17
1. Introduction.....	18
Résumé en Français.....	19
1.1. Influenza viruses – classification, diversity and pathogenicity.....	22
1.2. Virion structure of influenza A and B viruses .....	24
1.3. Viral cycle overview .....	24
1.3.1. Entry, fusion and uncoating of the viral particles.....	24
1.3.2. Transcription and replication of the viral genome .....	25
1.3.3. Transport of the vRNPs and budding of the viral particles .....	25
1.4. Transcription and replication by the influenza polymerase .....	27
1.4.1. Architecture of the influenza polymerase.....	27
1.4.2. Transcription by the influenza polymerase .....	29
1.4.3. Replication by influenza polymerase .....	31
1.5. Interaction of influenza polymerase with the cellular transcription machinery.....	32
1.5.1. Overview of Pol II-dependent transcription.....	32
1.5.1.1. Regulation of cellular transcription by phosphorylation of the C-terminal domain of Pol II.....	33
1.5.1.2. Promoter-proximal pausing.....	34
1.5.2. Association of FluPol with the cellular transcription machinery .....	37
1.5.2.1. Interaction of FluPol with Pol II.....	37
1.5.2.2. Interactions of FluPol with other nuclear factors .....	37

1.5.2.3.	Effect of influenza infection on cellular transcription.....	38
1.5.3.	Targeting the viral transcription machinery .....	39
1.6.	Scientific aims.....	40
2.	Materials and Methods.....	43
	Résumé en Français.....	44
2.1.	Purification of recombinant influenza polymerase .....	45
2.2.	Mutagenesis .....	46
2.3.	<i>In vitro</i> binding assays of FluPol and CTD peptides based on fluorescence anisotropy.....	46
2.4.	Structure determination of FluPol-CTD peptide complexes.....	47
2.4.1.	FluA polymerase-CTD crystallization and structure determination.....	47
2.4.2.	FluB polymerase-CTD crystallization and structure determination.....	48
2.4.3.	FluA polymerase crystallization with modified CTD peptides .....	48
2.5.	Polymerase activity assay with radiolabeled capped RNA .....	48
2.6.	Mini-genome assay .....	49
2.7.	Expression of polymerase in mini-genome assays.....	50
2.8.	RNA quantification in the mini-genome assay .....	50
2.9.	Reverse genetics.....	51
2.10.	<i>In vitro</i> polymerase activity assays .....	51
2.11.	<i>In vitro</i> transcription, purification and capping of RNAs .....	52
2.12.	FluPol – Pol II complex assembly.....	53
2.13.	Functional assays of FluPol – Pol II complex.....	54
2.13.1.	Cleavage assays with different scaffolds.....	55
2.13.2.	Effect of DSIF on the FluPol recognition of capped RNA, pre-bound to Pol II... ..	55
2.13.3.	NTP incorporation by FluPol .....	55
2.13.4.	Pol II effect on FluPol endonuclease activity.....	56
2.13.5.	NTP incorporation by Pol II in presence of FluPol (PA/K134A) mutant .....	56

2.13.6.	Effect of DSIF and Tat-SF1 on FluPol endonuclease activity .....	56
2.14.	Negative stain electron microscopy .....	57
2.14.1.	Sample preparation for FluPol/Pol II complex.....	57
2.14.2.	Negative stain analysis of FluPol/Pol II complex .....	57
2.14.3.	Negative stain of FluPol/Pol II/DSIF complex .....	58
2.15.	Cryo-electron microscopy on FluPol/Pol II complex .....	58
2.15.1.	Sample preparation.....	58
2.15.2.	Data collection.....	59
2.16.	Purification of full-length DSIF .....	59
2.17.	Cloning, expression and purification of DSIF CTR domain.....	60
2.18.	Purification of KOWx-5 domain.....	61
2.19.	Tat-SF1 purification .....	61
2.20.	Analytical SEC studies on the interaction between FluPol and elongation factors ..	62
2.20.1.	Interaction between FluPol and full-length DSIF or DSIF domains .....	62
2.20.2.	SEC analysis of FluPol and Tat-SF1 .....	62
2.21.	Pull-down experiments of FluPol and elongation factors .....	63
2.21.1.	Pull-down experiments on FluPol and DSIF .....	63
2.21.2.	Pull-down experiments on FluPol and Tat-SF1 .....	63
3.	Results.....	64
	Résumé en Français .....	65
3.1.	Interaction between influenza polymerase and the C-terminal domain of RNA polymerase II .....	67
3.1.1.	Peptide design.....	67
3.1.2.	Binding assays of FluPol with CTD peptides.....	67
3.1.2.1.	Binding of FluA polymerase to CTD peptides.....	67
3.1.2.2.	Binding of FluB polymerase to CTD peptides .....	69

3.1.3.	Structural characterization of FluPol and CTD peptides.....	70
3.1.3.1.	Structure of FluA polymerase bound to four-repeat SeP5 CTD peptide.....	70
3.1.3.2.	Structures of SeP5 CTD-bound FluB polymerase.....	76
3.1.4.	Effect of CTD binding on FluPol activity .....	81
3.1.5.	Functional validation of the FluPol – CTD interaction.....	84
3.1.5.1.	FluA polymerase – CTD association.....	84
3.1.5.1.1.	<i>In vitro</i> characterization.....	84
3.1.5.1.2.	Effect of disrupting CTD binding in the cellular context .....	85
3.1.5.1.3.	Reverse genetics of the FluA CTD-binding deficient mutants.....	88
3.1.5.2.	FluB polymerase – CTD interaction.....	97
3.1.5.2.1.	Mini-genome assay of FluB CTD-binding deficient mutants .....	97
3.1.5.2.2.	Recombinant influenza B viruses with CTD-binding deficiencies .....	98
3.1.6.	Structure-based drug design targeting the FluPol – CTD interaction .....	99
3.2.	Assembly of full-length complex between influenza polymerase and elongating RNA polymerase II .....	103
3.2.1.	Reconstitution of a stalled cap-snatching complex .....	104
3.2.1.1.	<i>In vitro</i> transcription and capping of substrate RNAs .....	104
3.2.1.2.	Assembly of FluPol/Pol II complex with different scaffolds .....	105
3.2.1.3.	Role of the CTD in complex assembly.....	107
3.2.1.4.	Stoichiometry of the FluPol/Pol II complex.....	109
3.2.2.	Functional characterization of FluPol-Pol II coupling .....	111
3.2.2.1.	Recognition by FluPol of Pol II-pre-bound RNAs with different lengths .....	111
3.2.2.2.	Effect of Pol II on RNA cleavage by FluPol.....	117
3.2.2.3.	Effect of FluPol binding to capped RNA on Pol II transcription .....	119
3.2.3.	Structural studies on the FluPol/Pol II complex.....	120
3.3.	Interaction between influenza polymerase and cellular transcriptional regulators .....	124

3.3.1.	Association of FluPol with DRB sensitivity-inducing factor (DSIF).....	125
3.3.1.1.	Binding studies of FluPol with full-length DSIF .....	125
3.3.1.2.	Interaction of FluPol with separate domains of DSIF .....	129
3.3.1.3.	Assembly of a ternary complex of elongating Pol II, FluPol and DSIF.....	131
3.3.2.	Association of FluPol with Tat-SF1 .....	134
3.3.2.1.	Interaction of FluPol with Tat-SF1.....	135
3.3.2.2.	Assembly of a ternary complex consisting of FluPol, elongating Pol II and Tat-SF1.....	137
3.3.3.	Effect of DSIF and Tat-SF1 on FluPol endonuclease activity in the context of elongating Pol II .....	138
4.	Discussion.....	142
	Résumé en Français.....	143
4.1.	Recruitment of FluPol to transcribing Pol II by interaction with the CTD.....	145
4.2.	Reconstitution of an active cap-snatching complex.....	153
	Acknowledgements.....	158
	Appendix.....	160
	Appendix Table 1. Sequences of CTD peptides.....	161
	Appendix Table 2. Sequences of promotor RNAs .....	162
	Appendix Table 3. Capped RNAs produced <i>in vitro</i> .....	163
	Appendix Table 4. DNA scaffolds .....	163
	Appendix Table 5. Data collection and refinement statistics for FluA-CTD, FluB-CTD, FluA-tetralinker CTD, FluA-thiophosphate CTD structures .....	164
	References.....	166

## List of Abbreviations

ADP	adenosine diphosphate
ANP32	acidic nuclear phosphoprotein 32-kDa family
ATP	adenosine triphosphate
BM2	influenza B matrix protein 2
BPB	bromophenol blue
BSA	bovine serum albumin
BXM	baloxavir marboxil
CBC	nuclear cap-binding complex
Cdk	cyclin-dependent kinase
CE	capping enzyme
CPSF	cleavage and polyadenylation factor
CRM1	chromosome maintenance 1 (exportin 1)
cRNA	complementary ribonucleic acid
CTD	C-terminal domain of RNA polymerase II
CTP	cytidine triphosphate
CTR	C-terminal region of DRB sensitivity-inducing factor
DBT	desthiobiotin
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
Dpn 1	Diplococcus pneumoniae nuclease I
DSIF	DRB sensitivity-inducing factor
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FAM	fluorescein amidite
FluA	influenza A
FluB	influenza B
FluPol	influenza polymerase
GAPDH	glyceraldehyde 3-phosphate dehydrogenase

GST	glutathione S-transferase
GTF	general transcription factors
GTP	guanosine triphosphate
HA	hemagglutinin
hCLE	C14orf166 component of tRNA ligase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
HRP	horseradish peroxidase
IAV	influenza A virus
IBV	influenza B virus
ICV	influenza C virus
IDV	influenza D virus
$K_D$	dissociation constant
kDa	kilodalton
KOW	Kyprides, Ouzounis, Woese motif
M1	matrix protein 1
M2	matrix protein 2
MCM	mini-chromosome maintenance complex
mRNA	messenger ribonucleic acid
MWCO	molecular weight cut-off
NA	neuraminidase
NELF	negative elongation factor
NEP/NS2	nuclear export/non-structural protein 2
NGS	Next Generation Sequencing
Ni-NTA	nickel-chelated nitrilotriacetic acid
NLS	nuclear export localization signal
NP	nucleoprotein
NS1	non-structural protein 1
ntDNA	non-template deoxyribonucleic acid
NTP	nucleotide triphosphate
PA	polymerase acidic protein
PA-C	C-terminal domain of polymerase acidic protein

PA-endo	endonuclease domain
PAF	PAF1 elongation factor complex
PARP1	poly(ADP)-ribose 1
PB1	polymerase basic protein 1
PB2	polymerase basic protein 2
PB2-C	C-terminal domain of polymerase basic protein 2
PB2-N	N-terminal domain of polymerase basic protein 2
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PIC	pre-initiation complex
Pol II	RNA polymerase II
P-TEFb	positive transcription elongation factor b
RED/SMU1	spliceosomal factors
RNA	ribonucleic acid
RNP	ribonucleoprotein particle
RPB1-12	subunits of RNA polymerase II 1-12
SAM	S-adenosyl methionine
SDS-PAGE	sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SeP2, SeP5	serine 2-phosphorylated, serine 5-phosphorylated
SFPQ	splicing factor proline-glutamine rich
snoRNA	small nucleolar ribonucleic acids
snRNA	small nuclear ribonucleic acids
Spt4, 5	subunits of DRB-sensitivity-inducing factor
Spt6	elongation factor Spt6
Tat-SF1	Tat stimulatory factor 1
TBE	Tris-boric acid
TCEP	Tris(2-carboxyethyl)phosphine
tDNA	template deoxyribonucleic acid
TF	transcription factors
TFIIH	transcription factor of RNA polymerase II H

UTP	uridine triphosphate
V1	amplified virus for insect cell expression
VCE	vaccinia capping enzyme
vRNA	viral ribonucleic acid
XC	xylene cyanol
Xrn2	5'-3' exoribonuclease 2

## List of Figures

Figure 1.	Influenza replication cycle.....	26
Figure 2.	Architecture of the influenza polymerase and vRNA promoter.....	28
Figure 3.	Overview of the transcription cycle of influenza polymerase.....	30
Figure 4.	Schematic representation of the C-terminal domain of Pol II.....	34
Figure 5.	Overview of Pol II-dependent transcription, highlighting protomer-proximal pausing.....	36
Figure 6.	Graphical representation of the thesis aims.....	42
Figure 7.	Binding assays of FluA polymerase and SeP5-CTD peptides.....	68
Figure 8.	Interaction between FluA polymerase and a four-repeat non-phosphorylated CTD peptide and a two-repeat SeP2 peptide.....	69
Figure 9.	Interaction between FluB polymerase and SeP5 two- and four-repeat CTD peptides and non-phosphorylated four-repeat peptide.....	70
Figure 10.	Structure of FluA polymerase bound to four-repeat SeP5 CTD peptide.....	71
Figure 11.	Structure of the SeP5 CTD peptide bound to FluA site 1.....	72
Figure 12.	Structures of various host cell proteins bound to CTD peptides.....	73
Figure 13.	Interactions between SeP5 CTD peptide and FluA polymerase site 2.....	74
Figure 14.	Comparison of CTD-bound conformations of FluA polymerase and Cgt1.....	75
Figure 15.	Sequence alignment of the CTD-binding region from PA of influenza A, compared to B, C and D strains.....	75
Figure 16.	Structures of different crystal forms of FluB polymerase bound to four-repeat SeP5 CTD peptide.....	76
Figure 17.	Details of the interaction between SeP5 CTD peptide and FluB polymerase.....	78
Figure 18.	Sequence alignment of the regions of CTD-interacting site 2' of FluB.....	79
Figure 19.	Effect of SeP5 CTD peptides on endonuclease and transcription activity of FluA and FluB polymerases.....	82-83
Figure 20.	Binding of SeP5 CTD peptides to recombinant FluA polymerase mutants.....	84
Figure 21.	Mini-genome assay of FluA CTD-binding deficient mutants in presence or absence of NS2/NEP protein.....	86
Figure 22.	Quantification of different RNA species in the mini-genome assay (in presence of NS2/NEP) by RT-qPCR.....	87

Figure 23.	Comparison between cap-primed transcription and v- and cRNA-dependent replication activities of wild type and mutant FluA polymerases using fluorescence-based assay.....	88
Figure 24.	Reverse genetics of FluA CTD-binding deficient mutants.....	90
Figure 25.	Plaque assays for PA revertant mutants.....	92
Figure 26.	Analysis of the revertant mutants on PA subunit, rescuing CTD binding.....	94
Figure 27.	Revertant mutant residues on PB1 and PB2 mapped on the polymerase structure.....	96
Figure 28.	Mini-genome assay of FluB CTD-binding mutants.....	97
Figure 29.	Crystal violet staining of cell monolayers infected with recombinant viruses with several CTD-binding mutations in influenza B (B/Brisbane/60/88).....	98
Figure 30.	Structure of FluA polymerase with a tetralinker CTD peptide.....	100
Figure 31.	Co-crystal structure of FluA polymerase with 22mer thiophosphate CTD peptide.....	101
Figure 32.	Schematic representation of the cap-snatching complex.....	104
Figure 33.	Purification and capping of <i>in vitro</i> transcribed RNAs.....	105
Figure 34.	Size-exclusion chromatography profiles (SEC) of proteins alone or the complex between FluPol and Pol II.....	106
Figure 35.	Effect of CTD-phosphorylation on FluPol/Pol II complex assembly.....	108
Figure 36.	Effect of the disruption of CTD binding on the complex formation.....	109
Figure 37.	Stoichiometry of the FluPol/Pol II complex. ....	110
Figure 38.	Sequences of the DNA scaffolds used for elongating the capped RNA with one nucleotide (UTP) at a time.....	112
Figure 39.	Sequential incorporation of NTPs by FluPol, using v3'18 template vRNA.....	113
Figure 40.	Determination of the optimal length of Pol II-pre-bound RNA, recognized by FluPol.....	115
Figure 41.	Comparison of the cleavage of 27 and 28-mer RNA by FluPol.....	116
Figure 42.	Activation of FluPol endonuclease cleavage in the context of Pol II elongation complex.....	118
Figure 43.	Binding of FluPol to capped RNA is not enough to trigger its dissociation from elongating Pol II.....	119
Figure 44.	Negative stain electron microscopy of FluPol/Pol II complex.....	121
Figure 45.	Cryo-EM data collection on FluPol/Pol II complex.....	122
Figure 46.	Domain organization and structure of Pol II/ DSIF elongation complex.....	125
Figure 47.	SEC analysis of the interaction between recombinant DSIF and v5'RNA-bound FluPol.....	126
Figure 48.	SEC analysis of the interaction between FluPol and DSIF, phosphorylated by P-TEFb.....	127
Figure 49.	Pull-down experiment on the interaction between FluPol and DSIF.....	128
Figure 50.	Interaction studies by SEC of different DSIF domains with FluPol.....	130
Figure 51.	SEC of ternary complex of elongating Pol II, FluPol and DSIF-P.....	132

Figure 52.	Micrograph of negative stain EM on the FluPol/PolIII/DSIF-P complex.....	133
Figure 53.	Disorder prediction of Tat-SF1 with the domain organization shown above.....	134
Figure 54.	SEC analysis of Tat-SF1 interaction with FluPol.....	135
Figure 55.	Pull-down experiments of FluPol with Tat-SF1 and a ternary complex formation of FluPol, Tat-SF1 and DSIF.....	136
Figure 56.	Ternary complex assembly with FluPol, Pol II and Tat-SF1 analysed by SEC.....	137
Figure 57.	Effect of DSIF and Tat-SF1 on RNA cleavage by FluPol in the elongating Pol II context.....	138
Figure 58.	Optimal length of Pol II-pre-bound RNA cleaved by FluPol in presence of DSIF.....	140
Figure 59.	Overview of the structures of FluA, FluB and FluC polymerase in complex with SeP5 CTD peptides, illustrating the two separate CTD-binding sites for each polymerase.....	147
Figure 60.	Interaction of FluA site 2 with the CTD peptide highlighting residues involved in host restriction.....	150
Figure 61.	Overlap of the putative functional subdomain of PA-C, identified by fitness profiling with FluA CTD-binding sites.....	151
Figure 62.	Conformational change of the DSIF KOW <sub>x</sub> -KOW <sub>4</sub> , forming the RNA clamp, upon phosphorylation by P-TEFb.....	155
Table 1.	Revertant mutants generated during the passaging of mutant viruses.....	91

## Abstract

Influenza is an infectious disease causing seasonal epidemics and occasional pandemics, which are a significant health burden for the global human population. The causative agent, influenza virus, is a negative-strand segmented RNA virus. Each segment of the viral genome is transcribed and replicated by a virally-encoded RNA-dependent RNA polymerase in the nucleus of the infected cell. In order to produce a functional viral messenger RNA (mRNA) that can be processed by the cellular translation machinery, influenza polymerase employs a mechanism called 'cap-snatching'. The viral polymerase binds to a nascent, capped transcript, produced by the cellular RNA polymerase II, cleaves it shortly after the cap and uses it as a primer for the transcription of its own genomic segments. Viral transcription is therefore functionally dependent on cellular transcription. The studies described in this thesis aimed to investigate the interaction between influenza polymerase and the cellular transcription machinery. The main focus was to characterize structurally and functionally the association of influenza polymerase with the C-terminal domain (CTD) of RNA polymerase II, which serves as a scaffold for coordinating co-transcriptional events during mRNA synthesis and processing. Crystal structures of influenza A and B polymerase bound to phosphorylated peptides mimicking the CTD of RNA polymerase II gave new insight on how the viral polymerase directly recognizes the transcribing cellular polymerase. Moreover, disrupting this interaction was found to be severely detrimental to viral replication, confirming the essentiality of this association. In the second part of this work, an active cap-snatching complex was assembled *in vitro* and characterized functionally. This comprised a reconstituted RNA polymerase II elongation complex with emerging capped transcript and phosphorylated CTD with bound and transcriptionally active influenza polymerase. Additionally, the interaction of the viral polymerase with two factors involved in the regulation of transcription elongation by RNA polymerase II, DRB sensitivity-inducing factor and Tat stimulatory factor 1, was analyzed biochemically. Overall, the work presented here gives insights into the mechanism of recruitment of the influenza polymerase to the cellular transcription machinery and shows that the association of the viral and cellular polymerase is essential for the viral replication. Targeted disruption of this interaction is therefore a promising avenue for the design of novel anti-influenza drugs.

## Résumé en Français

La grippe est une maladie infectieuse qui se traduit par des épidémies saisonnières et pandémies occasionnelles et qui a des conséquences importantes sur la santé publique. Elle est causée par le virus influenza, un virus à ARN négatif segmenté. Chaque segment du génome est transcrit et repliqué dans le noyau de la cellule hôte par la polymérase virale, une ARN polymérase ARN-dépendente. Afin de pouvoir produire un ARN messenger (ARNm) viral qui peut être reconnu par la machinerie de traduction de la cellule, la polymérase de la grippe se sert d'un mécanisme appelé 'cap-snatching' (mécanisme de 'vol de coiffe'). La polymérase virale interagit avec un ARNm coiffé, produit par l'ARN polymérase II de la cellule hôte, coupe cet ARN à proximité de la coiffe et l'utilise comme amorce pour l'initiation de la transcription de son propre matériel génétique. Par conséquent, la transcription virale est fonctionnellement liée à la transcription cellulaire. L'objet des recherches présentées dans cette thèse portait sur l'interaction entre la polymérase de la grippe et la machinerie de transcription de la cellule hôte et plus précisément, la caractérisation structurale et fonctionnelle de l'association entre la polymérase de la grippe et le domain C-terminal de l'ARN polymérase II (CTD), qui sert de support pour la coordination des processus de synthèse et les modifications post-transcriptionnelles de l'ARNm. Les structures cristallographiques obtenues de la polymérase d'influenza A et B, associées aux peptides dérivés du domain CTD de l'ARN polymérase II, ont permis d'élucider comment la polymérase virale reconnaît spécifiquement la polymérase cellulaire. La perturbation de cette interaction mène à l'inhibition de l'amplification du virus, ce qui confirme que cette association est essentielle. Dans la seconde partie de la thèse, un complexe actif de 'cap-snatching' a été assemblé *in vitro* et caractérisé fonctionnellement, constitué du complexe d'élongation contenant l'ARN polymérase II avec un ARNm coiffé et son domain CTD phosphorylé, lié à la polymérase virale elle-même transcriptionnellement active. De plus, l'interaction de la polymérase virale avec deux facteurs impliqués dans la régulation de l'élongation par l'ARN polymérase II, DRB sensitivity-inducing factor et Tat stimulatory factor 1, a été caractérisée. En conclusion, le travail de thèse présenté ici a contribué à élucider le mécanisme de recrutement de la polymérase d'influenza à la machinerie de transcription cellulaire et a montré que l'association entre les polymérases virale et cellulaire est essentielle pour la réplication du virus. La perturbation de cette interaction est une piste prometteuse pour la conception de nouveaux médicaments contre le virus influenza.

## **1. Introduction**

## Résumé en Français

Influenza (la grippe) est une maladie infectieuse qui touche jusqu'à 650 000 personnes par an et qui peut provoquer des effets graves sur la santé. Lors d'une pandémie, la grippe peut affecter 40 % de la population humaine. Le virus d'influenza, responsable de la maladie qui porte le même nom, est un virus à ARN segmenté et de polarité négative. Il existe plusieurs types de virus, A, B, C et D. Les virus du type A infectent principalement les oiseaux sauvages, mais certaines souches circulent dans d'autres organismes, tels que les humains, chiens, chevaux, cochons etc. Les virus du type B et C affectent les humains et ceux du type D, les vaches. L'évolution rapide du virus, nécessite la recherche de composés capable d'inhiber sa réplication.

Le matériel génétique du virus est organisé en huit segments d'ARN (pour les virus du type A et B) enrobés de plusieurs molécules de la nucléoprotéine (NP) et d'une copie de la polymérase virale, responsable de la transcription et la réplication des ARN. Ces processus ont lieu dans le noyau de la cellule hôte. Lors de la transcription de l'ARN, la polymérase produit des ARN messagers (ARNm) qui sont exportés du noyau et traduits par la machinerie de traduction cellulaire. Certaines protéines virales synthétisées, NP et les trois sous-unités de la polymérase, sont importés dans le noyau. La polymérase produit aussi des copies positives de son ARN génomique, qui serviront de matrice pour la synthèse de nouveaux ARN viraux négatifs (ARNv), qui s'associent alors avec les NPs et polymérase nouvellement produites. Les mécanismes de transcription et réplication sont clairement différents et l'ARN qui en résulte est soit un ARNm, qui sera traduit pour la production de protéines virales, soit un ARNv qui servira directement à la formation de nouvelles particules virales.

Pour produire des ARNm viraux qui sont correctement exportés du noyau, modifiés et traduits par la machinerie cellulaire de traduction, les ARNm doivent avoir une coiffe à leur extrémité 5'. Le virus ne possédant pas d'activité de synthèse de coiffe, il se sert de la machinerie de transcription cellulaire, et notamment l'ARN polymérase II, qui est responsable de la production des ARNm cellulaires coiffés. La polymérase de la grippe s'associe à l'ARN polymérase II, interagit avec l'extrémité 5' coiffée de son ARNm et, après l'avoir clivée à proximité de la coiffe, s'en sert pour produire un amorce pour la transcription de son propre matériel génétique ('cap-snatching' ou 'vol

de coiffe'). De cette manière, elle se procure un 5' coiffé qu'elle n'a pas pour que ses ARNm puissent être traduits par la cellule hôte. L'association de la polymérase virale à l'ARN polymérase II cellulaire a été proposée comme une condition nécessaire pour le mécanisme de vol de coiffe.

Le domaine C-terminal de l'ARN polymérase II (CTD) sert de structure pour la coordination des processus de synthèse et modifications des ARNm cellulaires. Il est composé de plusieurs unités (52 dans les cellules mammifères) avec la composition  $Y_1S_2P_3T_4S_5P_6S_7$ . Ces résidues sont fortement modifiés pendant les différentes phases de la transcription et ces modifications sont cruciales pour l'interaction de ce domaine avec des facteurs nécessaires pour la régulation de la transcription. La phosphorylation des résidues en position 2 et 5 (Ser2 et Ser5) est une marque phare pour la phase de la transcription: lors de l'initiation, Ser5 est phosphorylée par une kinase qui fait partie du complexe d'initiation. Pendant les phases d'élongation et de terminaison, la Ser2 est phosphorylée par différentes kinases.

Lorsque l'ARN polymérase II atteint la position entre le 20ème et 100ème nucléotide, elle s'arrête. Ce processus, strictement régulé par plusieurs facteurs d'élongation, est nécessaire pour la vérification de la qualité de l'ARN. Pendant cette phase, l'ARN polymérase interagit avec plusieurs facteurs, tels que DSIF (DRB sensitivity-inducing factor) et NELF (negative elongation factor). La transition entre l'état de pause et l'état d'élongation processive, est régulée par l'action de la kinase P-TEFb (positive transcription elongation factor b) qui phosphoryle le domaine C-terminal de l'ARN polymérase II, tout comme NELF et DSIF. L'état de pause est alors un point important pour le recrutement des enzymes qui modifient l'extrémité 5' de l'ARN, tel que l'enzyme de synthèse de coiffe (capping enzyme).

La polymérase de la grippe étant fonctionnellement dépendante de la machinerie de transcription cellulaire, a été décrite comme interagissant physiquement avec cette dernière. Plusieurs études ont aussi montré l'association avec d'autres protéines impliquées dans les différentes phases de pause et de transition vers l'élongation de la transcription cellulaire. Cette interaction implique d'une part, le rapprochement de la polymérase virale de la source des ARNm coiffés, et d'autre part une potentielle inhibition de la transcription cellulaire. Cela résulte en une diminution de l'expression des protéines cellulaires, et notamment une inhibition de la réponse anti-virale de la cellule.

Le sujet de la thèse présentée ici était d'étudier l'interaction entre la polymérase de la grippe et la machinerie de transcription de la cellule. Dans un premier temps, l'étude s'est focalisée sur l'interaction entre la polymérase virale et le domaine C-terminal de l'ARN polymérase II. L'objectif était de comprendre comment la polymérase grippale s'associe à ce domaine et quelles sont les conséquences de cette interaction sur la réplication du virus. Dans un deuxième temps, la thèse s'est concentrée sur la reconstitution du complexe de 'cap-snatching' actif *in vitro*, formé par des protéines recombinantes et son étude par différentes approches fonctionnelles et structurales. Cette approche a consisté aussi à identifier d'autres protéines impliquées dans la transcription cellulaire et nécessaires au recrutement de la polymérase virale.

### 1.1. Influenza viruses – classification, diversity and pathogenicity

Influenza is a common respiratory disease in humans, which causes mild symptoms or in some cases, severe illness. The annual mortality caused by influenza is estimated to be up to 650 000 <sup>1,2</sup>. Occasionally, it causes global pandemics, which can infect up to 40 % of the population <sup>3</sup>. There have been four major pandemics in the last two centuries (1918, 1957, 1968 and 2009) and it is estimated that the so-called Spanish influenza from 1918 was responsible for over 40 million deaths, highlighting the global health burden caused by influenza <sup>4</sup>.

Influenza is caused by influenza viruses, which are members of the *Orthomyxoviridae* family. They are characterized by a single-stranded, negative-sense segmented RNA genome. Influenza viruses are divided into four different genera: A, B, C and D (IAV, IBV, ICV and IDV, respectively). IAV are further classified according to their surface glycoproteins hemagglutinin (HA) and neuraminidase (NA): currently 18 HA and 11 NA subtypes have been described <sup>5</sup>. IBV now circulate in two divergent lineages, derived from B/Victoria and B/Yamagata <sup>6</sup>.

The main reservoir of influenza A strains is aquatic wild birds, but they can also infect humans, pigs, dogs, horses, poultry, whales, and seals <sup>7,8</sup>. There are 16 HA and 9 NA conventional IAV subtypes which mainly circulate in birds, but can occasionally cross the inter-species barrier and adapt to different hosts. Such examples are H1, H2 and H3 IAV strains, which have adapted to replicate in humans. Recently, new IAV types were discovered in bats (H17, H18, N10 and N11), which can replicate in different mammalian cell lines (e.g. human and canine), suggesting that they could also adapt to new hosts and represent a yet unexplored reservoir of influenza <sup>5,9,10</sup>. IBV co-circulate with IAV (A/H3N2 and A/H1N1) during seasonal epidemics. They infect humans and there is no evidence for zoonotic transmission. The morbidity and mortality due to IBV are as important as those caused by seasonal IAV <sup>6</sup>. ICV are restricted to humans and IDV is exclusively pathogenic to cattle <sup>7,8,11</sup>.

Two major processes drive the evolution of human influenza viruses – antigenic shift and antigenic drift. Antigenic shift consists of drastic changes in the antigenic properties of a circulating virus. Due to the segmented nature of the viral genome (IAV and IBV carry eight segments, while ICV and IDV have seven), viruses from different strains, which infect the same cell, can exchange segments. This process is known as reassortment and the resulting strain can have a novel combination of surface antigens (HA and NA) precluding its recognition by the host immune

system <sup>2,3</sup>. Reassortments between different IAV subtypes have been crucial events in the emergence of pandemics (e.g. in 1957, 1968 and 2009) <sup>4</sup>. The two reassorting viruses must infect the same organism, and pigs are thought to play a role of a ‘mixing vessel’ since they can be infected by both avian and human strains, reassortment of which can create new pandemic strains <sup>2</sup>. Additionally, mutations in the receptor proteins or the polymerase can lead to an adaptation to a new host, therefore, overcoming the species restriction <sup>2</sup>. In such occasions, avian viruses can be directly transmitted to humans (e.g. H5N1, as well as several H7 and H9 strains) <sup>4</sup>.

Antigenic drift consists of continuous changes under the selective pressure of the host immune system. These mutations, caused by the error-prone viral RNA synthesis, are mainly associated with the surface receptors HA and NA <sup>2</sup>. They allow the viruses to evade the immune response and proliferate <sup>3</sup>. Overall, antigenic drift and shift are responsible for the rapid evolution of influenza virus and the unpredictability of this evolution is the reason why influenza is difficult to combat.

The high incidence and serious consequences of influenza infections on human health requires continuous efforts to tackle the disease. Vaccines are the first line of prevention against influenza. However, the high variability of the surface receptors can lead to low cross-reactivity of the vaccines and their inefficiency towards circulating strains. Currently, the development of a universal vaccine is one of the main anti-influenza strategies, but it remains challenging and the efficiency against new pandemic strains is difficult to assess <sup>12</sup>. Together with vaccines, antivirals are essential for influenza treatment, especially in the case of pandemics, when vaccines may not be available (e.g. the 2009 pandemic). There are existing approved drugs, such as oseltamivir and zanamivir (neuraminidase inhibitors, see below), amantadine and rimantadine (M2 channel inhibitors), but emergence of resistance against them remains a major concern <sup>2</sup>. Therefore, there is a need for novel inhibitors with different targets, providing possibilities for combination therapies <sup>13</sup>.

## 1.2. Virion structure of influenza A and B viruses

IAV and IBV are enveloped with a lipid membrane derived from the infected cell and containing the glycoproteins hemagglutinin (HA) and neuraminidase (NA) and the ion channel M2 (BM2 for influenza B) (Figure 1A). The inner side of the membrane is covered by a layer of the matrix protein M1. Inside the virion reside the eight genome segments, each comprising a single-stranded RNA molecule, covered by multiple copies of the nucleoprotein (NP) and bound to a single RNA-dependent RNA polymerase (FluPol). These components form the ribonucleoprotein particle (RNP), which is the functional unit for the transcription and replication of the viral genome. The polymerase consists of three subunits, the polymerase acid (PA), polymerase basic 1 (PB1) and polymerase basic 2 (PB2) proteins. The virion also contains a nuclear export/nonstructural protein 2 (NEP/NS2), which is involved in the trafficking of the RNPs within the cell and regulation of transcription and replication of the viral RNA <sup>2,8</sup>.

Each viral protein is encoded by a separate genome segment, with the exception of the M1/M2 and NS1/NS2 proteins, which are alternatively spliced products of the same segment. In addition to the proteins found within the virion, there are other proteins encoded in the viral genome, which have regulatory functions. These are produced inside the infected cell by alternative splicing, alternative translation initiation sites or frameshifts (e.g. PB1-F2 and PB1-N40 for segment 2, PA-X, PA-N155 and PA-N182 for segment 3, M42 for segment 7 and NS1 and NS3 for segment 8). Overall, the genome of IAV encodes for up to 17 proteins and the IBV for eleven <sup>14</sup>.

## 1.3. Viral cycle overview

### 1.3.1. Entry, fusion and uncoating of the viral particles

Influenza virus HA glycoprotein recognizes sialic acid receptors on the surface of the host cell, which is an epithelial cell of the respiratory tract (in mammals) or the intestinal tract (in birds). This tropism is largely determined by the recognition of a specific linkage on the receptors of the corresponding epithelial cells: human influenza HA binds preferentially to  $\alpha$ 2,6-linked sialic acid, while avian influenza recognizes an  $\alpha$ 2,3 linkage. The virion is internalized via receptor-mediated endocytosis and the low pH in the endosome triggers a conformational change in the HA, which leads to the fusion between the viral and endosomal membranes and the release of the viral RNPs

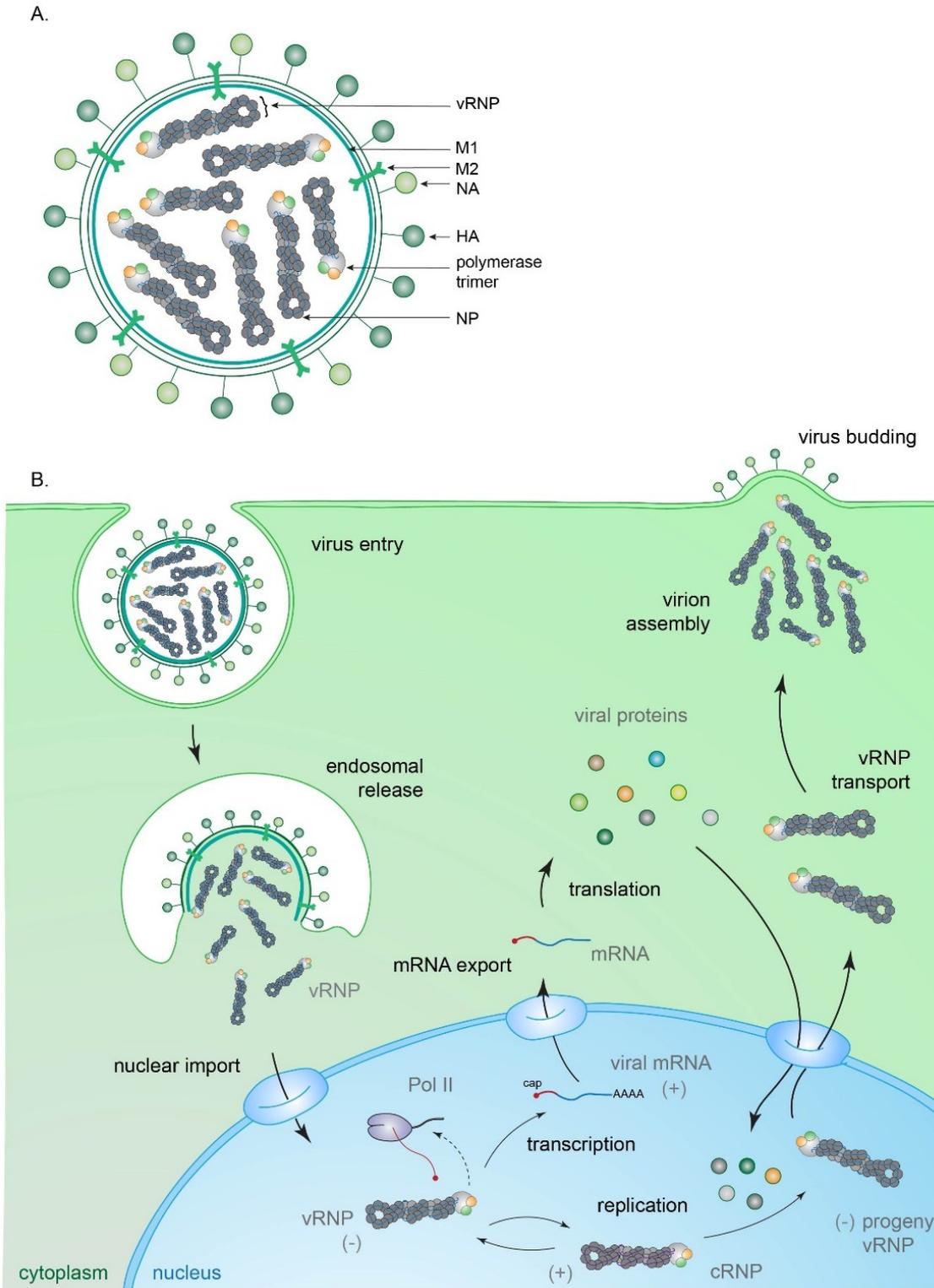
(vRNPs) into the cytoplasm (Figure 1B). The entry of H<sup>+</sup> through the M2 ion channel further acidifies the interior of the virion, which helps the dissociation of the RNPs from the matrix formed by M1<sup>8,15</sup>.

### **1.3.2. Transcription and replication of the viral genome**

The incoming vRNPs are imported into the nucleus, where transcription and replication of the viral RNA (vRNA) take place. The import is mediated through the interaction of the NP with the importin  $\alpha$  family of cellular transporters<sup>8,16</sup>. Once in the nucleus, the vRNPs are transcribed by the viral polymerase (primary transcription) into viral messenger RNAs (mRNAs), which are then exported to the cytoplasm and translated into proteins by the host protein synthesis machinery. The viral polymerase also performs replication, by producing a full complementary RNA (cRNA) copy of each segment, which is then used as a template for a second round of replication, yielding new vRNA. The NP and polymerase subunits, produced as a result of the primary transcription, are imported to the nucleus, where they assemble with the cRNA and the vRNA to form cRNPs and progeny vRNPs, respectively. The vRNPs are used as templates for secondary transcription, leading to production of more viral mRNAs. The vRNPs are eventually exported to the cytoplasm through the exportin 1 (CRM1) pathway with the help of the viral proteins M1 and NEP<sup>15,16</sup>.

### **1.3.3. Transport of the vRNPs and budding of the viral particles**

The newly synthesized membrane-associated HA, NA and M2 are further processed in the endoplasmic reticulum and Golgi apparatus and are directed to the apical membrane of the host cell. At the same time, progeny vRNPs associated with M1 are also trafficked via microtubule networks to the membrane and selectively packaged into virions so that one copy of each RNA segment is incorporated into the new virus particle<sup>17,18</sup>. The molecular details of the assembly of the different segments are not completely understood yet, but it was shown that association between the vRNPs and inter-molecular RNA-RNA interactions play a role in the selective packaging<sup>19,20</sup>. The newly assembled mature virions exit the cell by budding<sup>16,21</sup>. The release of the virus particles is an active process dependent on the enzymatic function of NA, which cleaves the sialic acid-containing receptors that would otherwise bind to HA and trap the virus. This leads to dissociation of the virus from the cell<sup>8</sup>.



**Figure 1. Influenza replication cycle.** (A) Simplified representation of the influenza virion (HA – hemagglutinin, NA – neuraminidase, NP – nucleoprotein, M1 and M2 – matrix proteins). (B) Overview of the replication cycle in the infected cell, the different processes are explained in detail in the text.

## 1.4. Transcription and replication by the influenza polymerase

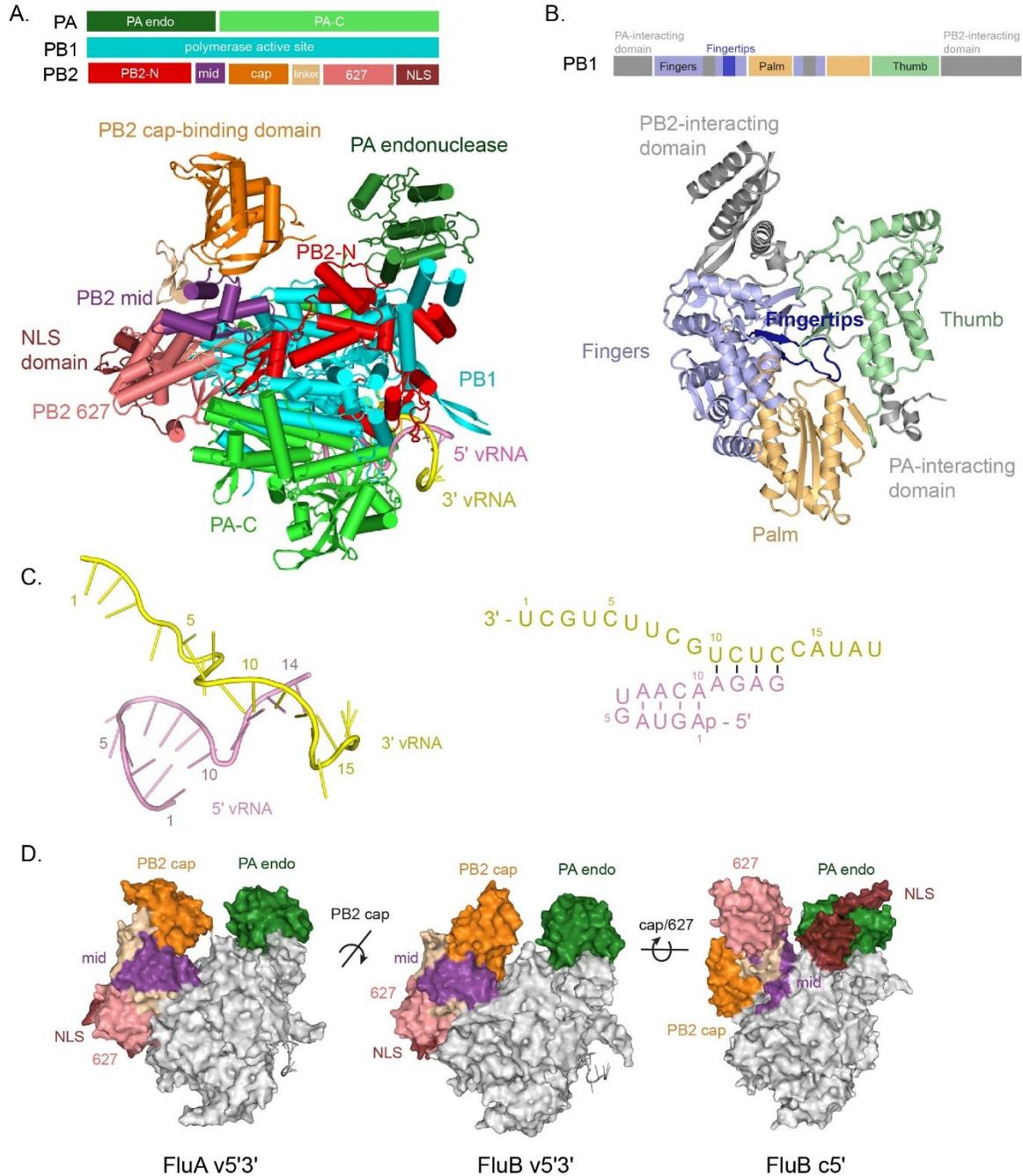
### 1.4.1. Architecture of the influenza polymerase

The polymerase of influenza plays a central role in the viral replication cycle. It has two functionally and mechanistically distinct activities – transcription and replication of the viral RNA. Recent structural data on influenza A, B and C polymerases revealed some of the molecular details of these processes<sup>22–25</sup>. The three subunits of the polymerase, PA, PB1 and PB2, form a U-shape (Figure 2A). PB1 has a right-handed RNA-dependent RNA polymerase fold with fingers, fingertips, palm and thumb domains and carries the conserved A-F motifs, which form the polymerase active site (Figure 2B). PA consists of an N-terminal endonuclease domain (PA-endo) connected by a flexible linker to a large C-terminal domain. PB2 has an N- and C-terminal domain, each divided into smaller folded subdomains. PB2-N wraps on one side of the PB1, while PB2-C forms one arm of the polymerase U-shape. PB2-C comprises the cap-binding, mid, 627, linker and NLS domains. The stable core of the polymerase is formed by PB1, the C-terminal domain of PA (PA-C), the PA linker, and the N-terminal part of PB2 (PB2-N). The flexibly linked peripheral domains – the endonuclease and the PB2 C-terminal region (cap-binding, mid-link, 627 and NLS domains) can undergo large conformational rearrangements with respect to the core, which are essential for the polymerase activity<sup>21,23–26</sup>.

The ends of each RNA segment consist of a highly conserved sequence, called the viral promoter, which can adopt specific secondary structures (Figure 2C). The 5' end forms a stem-loop, or a 'hook', containing two canonical base pairs (G2-C9 and U3-A8) and two mismatch base pairs (A1-A10, A4-A7). This region serves as an allosteric activator of the polymerase – binding of the promoter hook is essential to trigger conformational changes, necessary for activating the polymerase. The partial complementarity between the two ends allows for base-pairing between regions 11-15 of the 5' end and 10-14 of the 3' end, which is important for the initiation of RNA synthesis<sup>27</sup>.

The modular architecture of the polymerase gives it versatility, necessary for performing its functions. Solution studies and crystal structures of different states of FluPol have shown that the presence of bound RNA promoter, as well as its sequence (vRNA versus cRNA) can influence the overall conformation of the polymerase. In the apo- and c5'-end bound structures, the entire PB2

C-terminal domain is rotated by around 140° compared to the vRNA-bound conformation, bringing the NLS domain of PB2 and the PA endonuclease in close proximity (Figure 2D) <sup>24,25</sup>. On the other hand, comparison of several vRNA-promoter bound crystal structure shows that the cap-binding domain can undergo up to 70 ° rotation, which plays a role during transcription (see section 1.4.2) <sup>23</sup>.



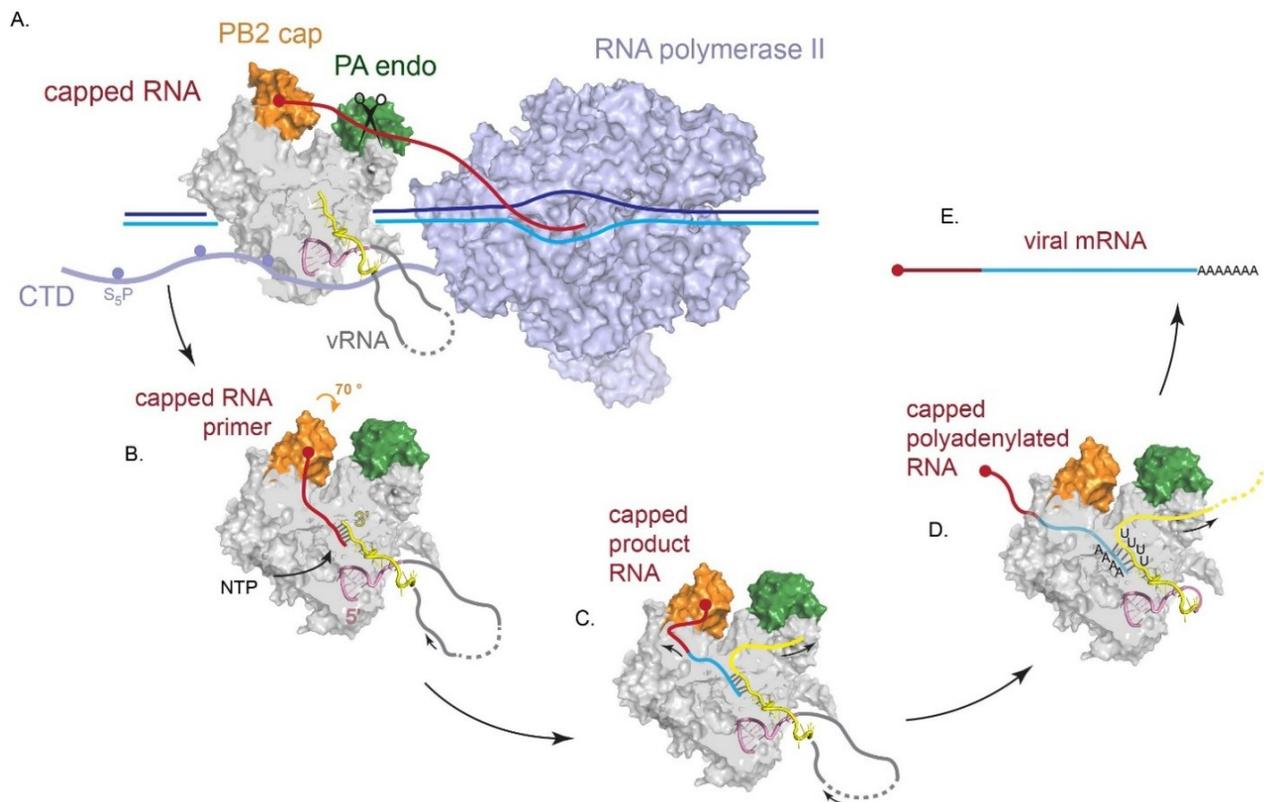
**Figure 2. Architecture of the influenza polymerase and vRNA promoter.** (A) Cartoon representation of FluPol with domain diagram indicating the color code shown above. (B) PB1 motifs characteristic of RNA-dependent RNA polymerases, with the color code shown above. (C) Structure of the viral RNA promoter, in the polymerase-bound conformation, with a schematic diagram of the observed base-pairings. (D) Different conformations of the polymerase, observed in diverse crystal structures. The cap-binding domain can be observed in two different conformations in the vRNA-bound structures of FluA (PDB: 4WSB) and FluB polymerase (PDB: 4WSA), while the whole PB2 C-terminus can adopt an entirely new conformation in the cRNA-bound structure of FluB polymerase (PDB: 5EPI), positioning the NLS domain on the endonuclease. Color code for the PB2 mid, cap, 627, NLS and PA-endo domains is as in (A), the rest of the polymerase is in grey.

#### 1.4.2. Transcription by the influenza polymerase

During transcription, FluPol associates with nascent capped transcripts, emerging from RNA polymerase II (Pol II), cleaves them 10-12 nucleotides downstream of the cap, and uses them as primers for the synthesis of its own messenger RNA (mRNA) in a process called ‘cap-snatching’ (Figure 3) <sup>28-30</sup>. High-throughput sequencing studies of the 5’ end of the viral mRNAs suggested that FluPol snatches the most abundant capped RNAs <sup>31</sup>. More recently, two different studies have identified that cellular mRNAs and some non-coding RNAs, such as small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) are most commonly cleaved by FluPol <sup>32,33</sup>. Whether this specificity is due to the preference of particular sequence at the cleavage site by FluPol, or to the recruitment of FluPol to the nascent transcripts by interaction with factors involved in the processing of these specific RNAs, is not well understood. Capped primers that can form canonical base-pairs with the 3’ end of the vRNAs, are preferentially used by FluPol <sup>32,34</sup>.

Transcription involves conformational changes of the flexibly linked cap-binding and possibly endonuclease domain. Initially, the cap-binding domain of FluPol associates with the 5’ cap of the nascent Pol II transcript (Figure 3A). In one of the crystal structures of FluPol, the cap-binding domain faces the endonuclease across a solvent channel, which would position it to allow the cleavage of a bound capped RNA at a distance of 10-15 nt from the cap. The trimmed capped RNA is then directed towards the FluPol active site by the rotation of the cap-binding domain by 70° (Figure 3B) <sup>23</sup>. The capped primer can potentially base-pair with the first nucleotides of the viral template in the polymerase active site. Incorporation of nucleotides leads to the elongation of the capped primer, using the vRNA as a template. The 5’ end of the product is released from the cap-binding domain and exits the polymerase through a putative exit channel, while the template

translocates through a different tunnel (Figure 3C). As the hook structure at the 5' end of the template remains bound to the promoter-binding site throughout transcription, when FluPol reaches the end of the template, it cannot translocate further and stutters along an oligo-U sequence (5-6 Us), located 17 nt before the 5' end, producing a poly(A) tail (Figure 3D)<sup>35,36</sup>. The cap (produced by cap-snatching) and poly(A) tail (resulting from the stuttering) are necessary for the proper processing, export and translation of the viral mRNAs. The RNA synthesized by FluPol is a hybrid since it contains a cell-derived 5' terminal sequence followed by the coding sequence of the corresponding viral segment (Figure 3E). Two important implications of this mechanism are that FluPol depends on the cellular transcriptional machinery for its access to capped transcripts, while it is independent of the cellular 3'-end-processing apparatus for the synthesis of poly(A) tails. Moreover, the cellular polyadenylation machinery is inhibited by NS1, which together with the cap-snatching contributes to the shut-off of cellular transcription and inhibition of the antiviral response<sup>37</sup>.



**Figure 3. Overview of the transcription cycle of influenza polymerase.** FluPol is shown in grey, with endonuclease domain in forest green and cap-binding domain in orange. (A) FluPol (PDB: 4WSB) associates with Pol II in order to access and cleave the capped transcripts (red). Pol II is shown in light blue (PDB: 5FLM), with its phosphorylated C-

terminal domain represented schematically (see section 1.5.1.1). (B) Priming and initiation of transcription, triggered by the rotation of the cap-binding domain. This conformation of the cap-binding domain is observed in the crystal structure of FluB polymerase (PDB: 4WSA). The capped primer (red) base-pairs with the template (yellow). (C) Elongation of the product RNA by incorporation of nucleotides, leading to dissociation of the cap from the cap-binding domain. (D) Polyadenylation by stuttering mechanism over a oligo(U) stretch, while the 5' end (pink) remains bound to the promoter-binding site. (E) The product of the transcription cycle, a hybrid RNA, containing the viral segment coding sequence (blue), a cap and host-derived 5' end sequence (red) and a poly(A) tail.

### 1.4.3. Replication by influenza polymerase

In addition to the transcription of viral mRNA, FluPol also performs replication of the viral RNA genome. This process consists of two rounds: first, the viral segment serves as a template for the synthesis of a full cRNA copy, which is then replicated in a second round of replication, producing the progeny vRNA. Both the intermediate cRNA and the product vRNA are bound by the newly produced polymerase and nucleoprotein, forming cRNPs and vRNPs, respectively<sup>38</sup>. Unlike transcription, replication is primer-independent and the mechanism of replication initiation differs depending on the template. The c-promoter adopts a similar secondary structure to the v-promoter: the c5' end forms the same hook structure as the v5' (shown on Figure 2C) and the c5' and 3' can base-pair, but the distance between the 3' end and the base-pairing region differs, which affect the position of the template in the active site. During cRNA-templated RNA synthesis, the polymerase initiates internally at the fourth nucleotide (U4) and realigns to reposition the 3' end (U1) to the active site, while vRNA-directed replication starts terminally at the U1 of the v3' template<sup>27,39,40</sup>. The elongation is believed to proceed in an identical way to transcriptional elongation, while the termination must involve dissociation of the 5' promoter from the polymerase, so that the end is completely copied. Mechanistic details of the replication process are still not well defined. For instance, replication has been suggested to differ from transcription by the recruitment of a *trans*-activating or *trans*-acting polymerase<sup>41,42</sup>. A *trans*-activating polymerase would induce replication and associate with the emerging 5' end of the product RNA, while a *trans*-acting polymerase, possibly activated by a short 5' viral RNA, is suggested to replicate the 3' terminus bound to the resident polymerase<sup>21,43</sup>. Mechanistic details on the assembly of new RNPs (vRNPs or cRNPs) from the product RNA and newly synthesized polymerase and NP are also still lacking.

The temporal regulation of the transcription and replication activities of FluPol is an important unresolved question<sup>44</sup>. Upon infection, transcription activity peaks around 6 h post infection, while replication products are detected later<sup>45</sup>. Different hypotheses explain the switch between transcription and replication by stabilization of the cRNA and vRNA by the newly synthesized FluPol and NP, products of the viral transcription<sup>46</sup> or by regulation by viral factors, such as NEP/NS2<sup>45</sup> and/or small viral RNAs<sup>43</sup>. The model in which replication requires a newly-synthesized *trans*-polymerase provides a plausible explanation for the switch between replication and transcription, and also further highlights the differences between the two processes<sup>41,44</sup>. Moreover, *in vitro* transcription is much more efficient compared to replication (almost 300-fold difference in kinetic constants)<sup>27</sup>, suggesting that some host factors might play a role in the regulation of FluPol functions. Host proteins that have been shown to directly affect viral replication are the mini-chromosome maintenance complex (MCM) and members of the acidic nuclear phosphoprotein 32-kDa family (ANP32)<sup>47-49</sup>. MCM is proposed to induce unprimed RNA synthesis, while ANP32 is particularly involved in cRNA-dependent replication. On the other hand, one of the important prerequisites for transcription is the access to cellular capped transcripts and the availability of these transcripts could play a role in determining whether FluPol does transcription or replication. The dependence on the cellular transcription machinery is therefore of crucial importance for regulating the activity of the viral polymerase.

## **1.5. Interaction of influenza polymerase with the cellular transcription machinery**

### **1.5.1. Overview of Pol II-dependent transcription**

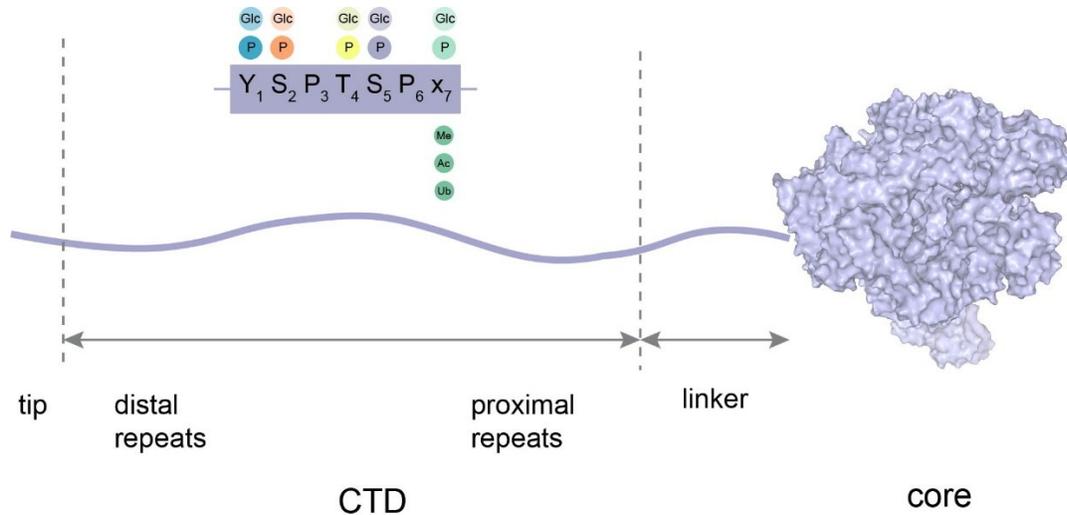
The cellular Pol II is responsible for the synthesis of mRNA, as well as other small RNAs (miRNA, snRNA, snoRNA) and long non-coding RNA (lncRNA). Transcription by Pol II starts with the recruitment of general transcription factors (GTFs) and the polymerase to the gene promoter region, forming the pre-initiation complex (PIC)<sup>50</sup>. At this stage, the mediator of Pol II transcription (Mediator complex) and the transcription factor IIIH (TFIIH) associate with Pol II. The Mediator complex serves as a functional bridge between specific transcription factors (transcription activators and repressors) and the PIC<sup>51</sup>. Within the first 20 to 100 nucleotides proximal to the promoter, Pol II pauses, a process regulated by its association with elongation factors (see section 1.5.1.2). At a length around 25 nucleotides, during the paused state, the

transcript is capped by the action of the capping enzymes. Capping consists of synthesis of a 7-methylguanosine cap at the 5' end of the transcript; it is required for nuclear export, processing, translation and protection of the nascent RNA from nucleases<sup>52,53</sup>. The pausing is released by the action of the positive transcription elongation factor (P-TEFb), which triggers productive elongation<sup>54</sup>. Further processing of the emerging RNA, such as splicing and editing, also happens co-transcriptionally<sup>55</sup>. Termination of mRNA transcription depends on the recruitment of a cleavage and polyadenylation specificity factor (CPSF), which associates with a specific signal at the end of the coding sequence and triggers cleavage and release of the transcript. The 3' end of the RNA is further modified by the addition of a poly(A) tail, which is necessary for its stability, export and translation<sup>56</sup>. After completion of RNA synthesis, Pol II is released from the DNA template and recycled for a new round of transcription.

#### **1.5.1.1. Regulation of cellular transcription by phosphorylation of the C-terminal domain of Pol II**

Pol II consist of 12 subunits (RPB1-RPB12), which form the functional core of the polymerase<sup>57</sup>; RPB1, the largest subunit is of a particular interest, because of its long flexible C-terminal domain (CTD). The CTD is an evolutionary conserved domain, composed of three regions: linker, heptad repeats and tip (Figure 4)<sup>58</sup>. The mammalian CTD contains 52 repeats with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>). About half of the repeats are identical to the consensus sequence, while many of the distal repeats have modifications, often at position 7 (e.g. Lys, Thr or Asn)<sup>58</sup>. The CTD is largely disordered<sup>59,60</sup>, but can adopt a variety of conformations, when bound to its respective interactors (reviewed in<sup>61,62</sup>). It functions as a scaffold for the recruitment of various RNA processing factors and plays a role in the temporal coordination of co-transcriptional RNA modification events<sup>63</sup>. It is subject to multiple post-translational modifications, constituting a 'CTD code': phosphorylations and glycosylations on Tyr1, Ser2, Thr4, Ser5, Ser7; methylations, acetylations and ubiquitinations of the non-consensus Lys and Arg at the seventh position; Pro isomerisations<sup>62</sup>. The phosphorylations of Ser2 and Ser5 are historically the most extensively studied and their implications are well understood: phosphorylation of Ser5 peaks in the promoter-proximal region, while phosphorylated Ser2 is more prominent at later stages and is linked to elongating Pol II and transcription termination. This simplistic view of the phosphorylation pattern gives a general idea of the temporal regulation by

CTD modifications, however, the interplay of the different kinases and phosphatases, their complex substrate recognition and specificity, the non-uniformity of the modifications over the length of the CTD suggest a much more complicated scheme of transcriptional regulation.



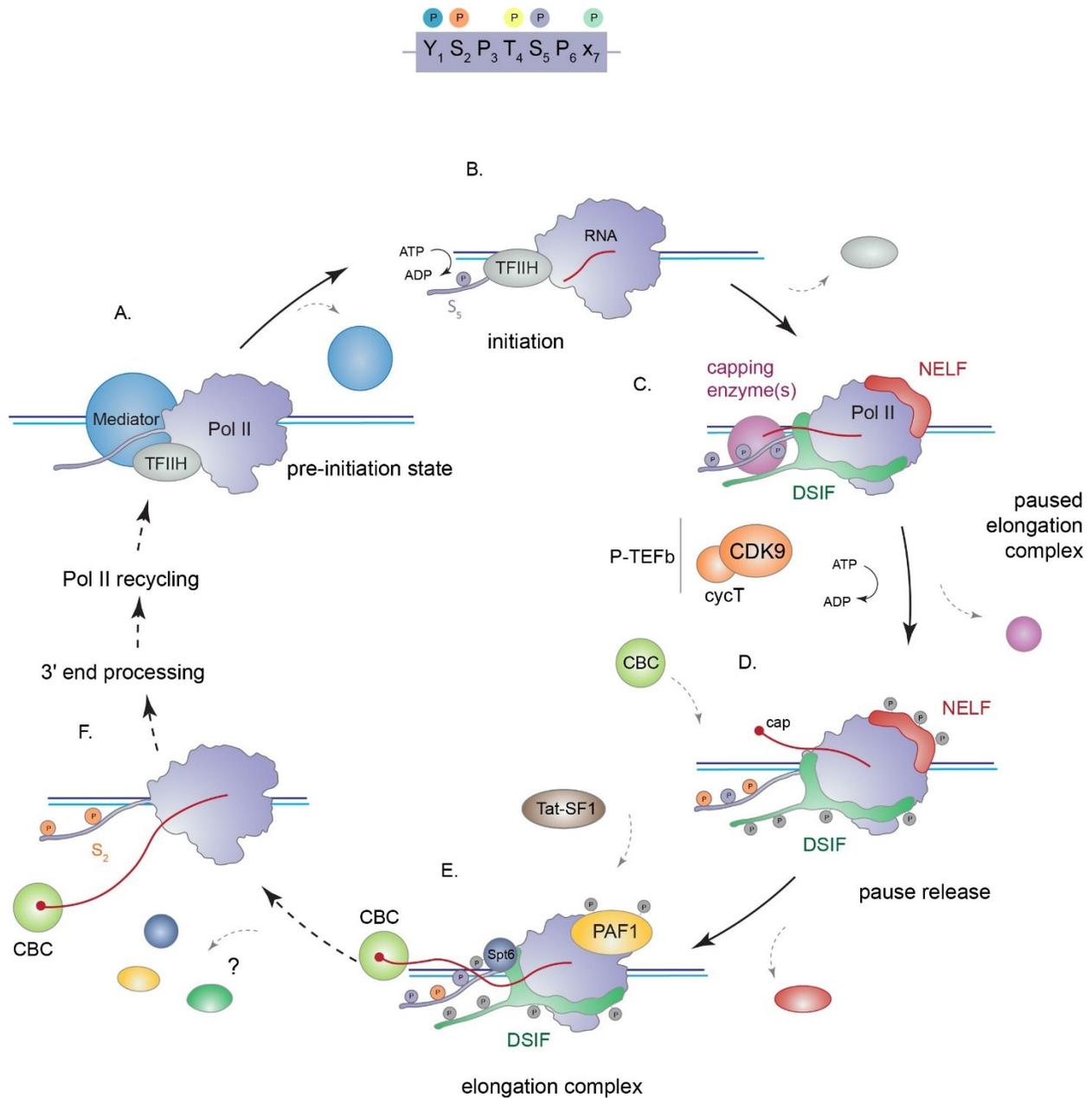
**Figure 4. Schematic representation of the C-terminal domain of Pol II.** The heptad repeat is highlighted, in the proximal repeats the seventh residue is a Ser, while in the distal repeats it is Lys, Thr or Asn.

### 1.5.1.2. Promoter-proximal pausing

At the beginning of transcription, Pol II is recruited to a promoter region with the help of general transcription factors. During initiation, the CTD of Pol II is unmodified and can interact with the Mediator complex, which integrates different regulatory cues by linking Pol II to transcription activators (Figure 5A) <sup>64</sup>. After initiation, Ser5 and Ser7 are phosphorylated by the cyclin dependent kinase 7 (Cdk7) activity of TFIIF, which triggers dissociation from the Mediator complex and promoter escape (Figure 5B) <sup>65</sup>. Subsequently, Pol II pauses 20-100 bp downstream from the transcription start site. This promoter-proximal pausing is a regulatory check-point for the correct processing of the nascent transcript (Figure 5C) <sup>54</sup>. It is dependent on two factors: DRB sensitivity-inducing factor (DSIF), composed of subunits Spt4 and Spt5, and negative elongation factor (NELF), containing four subunits (A, B, C and E) <sup>66,67</sup>. Recent structures of paused elongation complexes, containing Pol II, DSIF and NELF have shed light on the mechanism of pausing. Both proteins have a modular architecture, forming multiple interactions with the Pol II surface <sup>68-70</sup>. NELF binds to Pol II and restricts its conformational changes, stabilizing a particular state of the DNA/RNA scaffold that does not allow elongation. It also restricts the funnel by

probably preventing the diffusion of nucleotide triphosphates (NTPs) to the active site of Pol II. DSIF binds to the upstream DNA, as well as to the exiting transcript, forming a DNA and an RNA clamp, respectively. The Spt5 subunit contains a long disordered C-terminal region (CTR), consisting of multiple repeats (although not as conserved as the Pol II CTD repeats) which can be phosphorylated (e.g. by P-TEFb, see below). It is considered to play a similar scaffolding role as the CTD for the recruitment of RNA processing factors <sup>71</sup>. For instance, Ser5-phosphorylated CTD (SeP5) and the unphosphorylated CTR of Spt5 contribute to the recruitment of the capping enzyme (CE) to the paused polymerase <sup>72-74</sup>. The capping enzyme is also allosterically activated by the SeP5 <sup>73</sup> and properly positioned at the RNA exit channel due to additional interactions with the Pol II core <sup>75</sup>. The capped 5' end of the nascent RNA is later bound by the nuclear cap-binding complex (CBC) <sup>52</sup>. Splicing and further processing also occurs co-transcriptionally, and the CTD plays an important role for the coordination of these processes <sup>55</sup>.

Transition of the paused transcription complex into a productive elongation complex is linked to phosphorylation of Pol II and elongation factors by the positive transcription elongation factor b (P-TEFb) kinase, which consists of a cyclin-dependent kinase 9 (Cdk9) and cyclin T (Figure 5D) <sup>76</sup>. P-TEFb has multiple targets, which are phosphorylated during pause release. It phosphorylates NELF, leading to its dissociation from Pol II <sup>77</sup>. DSIF is phosphorylated at multiple sites, including the CTR and the RNA clamp, which probably leads to the release from the RNA and its conversion into a positive elongation factor. The CTD of Pol II is also subject to phosphorylation by P-TEFb at both Ser5 and Ser2 <sup>78-81</sup>. P-TEFb triggers the exchange of elongation factors - dissociation of NELF allows binding of the PAF1 complex and Spt6 (Figure 5E) <sup>82</sup>. NELF and PAF1 have overlapping binding sites on Pol II, but PAF1, DSIF and Spt6 can simultaneously bind to Pol II in a P-TEFb-dependent manner, triggering productive elongation. Another elongation factor, Tat stimulatory factor 1 (Tat-SF1) has been shown to have a complementary role in the pause release and to function together with DSIF and PAF1 <sup>77,83,84</sup>. At the later stages of transcription, Cdk12 together with P-TEFb increases the levels of Ser2 phosphorylation <sup>85</sup>, which is linked to recruitment of the 3' end processing machinery (Figure 5F) <sup>86,87</sup>.



**Figure 5. Overview of Pol II-dependent transcription, highlighting promoter-proximal pausing.** Details are explained in the text; the color code for the CTD phosphorylation is shown above, for simplicity only Ser5 and Ser2 phosphorylations are indicated.

## **1.5.2. Association of FluPol with the cellular transcription machinery**

### **1.5.2.1. Interaction of FluPol with Pol II**

A functional link between viral and host cell transcription has been known since very early studies showing that inhibitors of the cellular RNA synthesis affect influenza replication<sup>88-90</sup>, and, more importantly, that their effect is not direct but mediated by Pol II<sup>91,92</sup>. This has been rationalized by the mechanism of cap-snatching, which depends on Pol II as the source of FluPol transcription primers<sup>29,30</sup>. A physical association between FluPol and Pol II, and more particularly, to the SeP5-form of its CTD, has been suggested to bring the two polymerases in proximity to facilitate this process<sup>93</sup>. This would position FluPol in the right context for interacting with the capped RNA during the early phase of transcription, before it has been sequestered by the CBC complex<sup>94</sup>. Chromatin immunoprecipitation studies show indeed that FluPol is enriched in proximity to the promoter region<sup>95</sup>. Moreover, experiments on the dynamics of FluPol in the host cell nucleus, have suggested that there is an association with the cellular transcriptional machinery and that additional interactions of FluPol with Pol II, beyond the CTD, are possible<sup>96</sup>. Those early studies, however, did not establish whether the interaction is direct or mediated through other factors. A yeast two-hybrid approach suggested that a direct association exists between the PA subunit and the CTD<sup>97</sup>.

### **1.5.2.2. Interactions of FluPol with other nuclear factors**

Several studies have reported interactions of FluPol with factors involved in transcriptional elongation. Spt5, one of the two subunits of the DSIF complex, has been identified as a polymerase interactor<sup>98</sup>. In the same study, PAF1 also appears to associate with the polymerase. Tat-SF1, another elongation factor, have been suggested to play a role in RNP assembly through interactions with NP, but was also shown to bind to FluPol<sup>98,99</sup>. Three different studies have listed PARP1, which ADP-ribosylates NELF and promotes elongation<sup>100</sup>, as a polymerase interactor<sup>97,98,101</sup>. P-TEFb, the kinase responsible for pause release into productive elongation, has also been shown to bind to FluPol and influence its activity<sup>102</sup>. These factors have been identified through proteomics approaches and, with two exceptions, little is known about their interaction with the polymerase - the binding of the FluPol trimer with Spt5 has been further validated with immunoprecipitation experiments, while Tat-SF1 has been shown to have a stimulatory effect on viral replication<sup>98,99</sup>.

Other nuclear proteins, interacting with FluPol, are proposed to play a role in the regulation of the polymerase activities. Among those are hCLE (component of the tRNA ligase complex), a positive regulator of influenza replication<sup>103-105</sup>; SFPQ (splicing factor) involved in FluPol-mediated polyadenylation<sup>106</sup>; RED-SMU1, a regulator of viral RNA splicing<sup>107</sup>; MCM (DNA helicase), activator of unprimed replication<sup>47,48</sup> and ANP32, activator of cRNA-dependent replication<sup>49</sup>. Genetic screens and high-throughput proteomics studies have provided extensive data on the influenza polymerase-host-factor interactions and functional networks<sup>108, 109</sup>, but the molecular mechanisms underlying the role of these cellular proteins during influenza infection are poorly understood.

### **1.5.2.3. Effect of influenza infection on cellular transcription**

The interaction of FluPol with Pol II brings the viral polymerase in the right context for its cap-snatching activity and in proximity to the splicing and RNA export machineries, necessary for the processing of the viral mRNAs. On the other hand, this interaction is detrimental to the cell and has been suggested to downregulate Pol II transcription<sup>110</sup>. While FluPol does not affect cellular transcription initiation, it has an effect on elongation, decreasing the occupancy of Pol II downstream the promoter, and on termination<sup>95,111</sup>. There are several proposed mechanisms for the host transcription shut-off. First, FluPol endonuclease-dependent cleavage of the nascent transcript exposes the it to 5'-3' exonucleases, such as Xrn2<sup>112</sup>, or the lack of cap structure at the 5' end downregulates the export from the nucleus<sup>113,114</sup>. Second, FluPol can interfere with the Pol II CTD processing factors or elongation factors, therefore affecting CTD-dependent transcription regulation<sup>111</sup>. Finally, association of FluPol with Pol II has been shown to trigger degradation of the largest subunit of Pol II<sup>115</sup>. Additionally, the viral NS1 protein also interacts with RNA processing and elongation factors and contributes to the shut-off of cellular transcription<sup>116,117</sup>. NS1 interferes with the association of the 3'-end processing factor CPSF to the RNA and inhibits the 3' end formation of the cellular transcripts<sup>37</sup>. It also binds to the elongation factor PAF1 and suppresses transcriptional elongation<sup>116</sup>. Influenza infection thus interferes with elongation and termination, without affecting transcription initiation. This is in line with the hypothesis that viral cap-snatching happens early during the Pol II transcription process, plausibly during promoter-proximal pausing. The cellular antiviral response is suppressed as a consequence of host transcription shut-off, without affecting the proper functioning of the viral polymerase.

### **1.5.3. Targeting the viral transcription machinery**

Given the primary functional role of viral transcription during infection, any compound, interfering with this activity is of potential interest. Moreover, the structural peculiarities of the viral transcription machinery make it an attractive target for drug design. In particular, inhibitors of the cap-binding and endonuclease domains have been designed, validated and extensively characterized. VX-787, which targets the cap-binding domain of FluA strains is in clinical trials<sup>118-120</sup>. Baloxavir marboxil (BXM, commercial name Xofluza), an endonuclease inhibitor active against IAV and IBV<sup>121</sup>, was approved for sale in Japan in February 2018. A nucleotide triphosphate analogue, T-705, which targets a broad spectrum of RNA viruses, including influenza, was also approved in Japan in 2014 for treatment in case of pandemics<sup>122</sup>. These compounds, which interfere with the polymerase functions, bring alternatives to the existing neuraminidase and M2 inhibitors, but the fast evolution of the virus and emergence of resistance necessitate the discovery of new potential inhibitors.

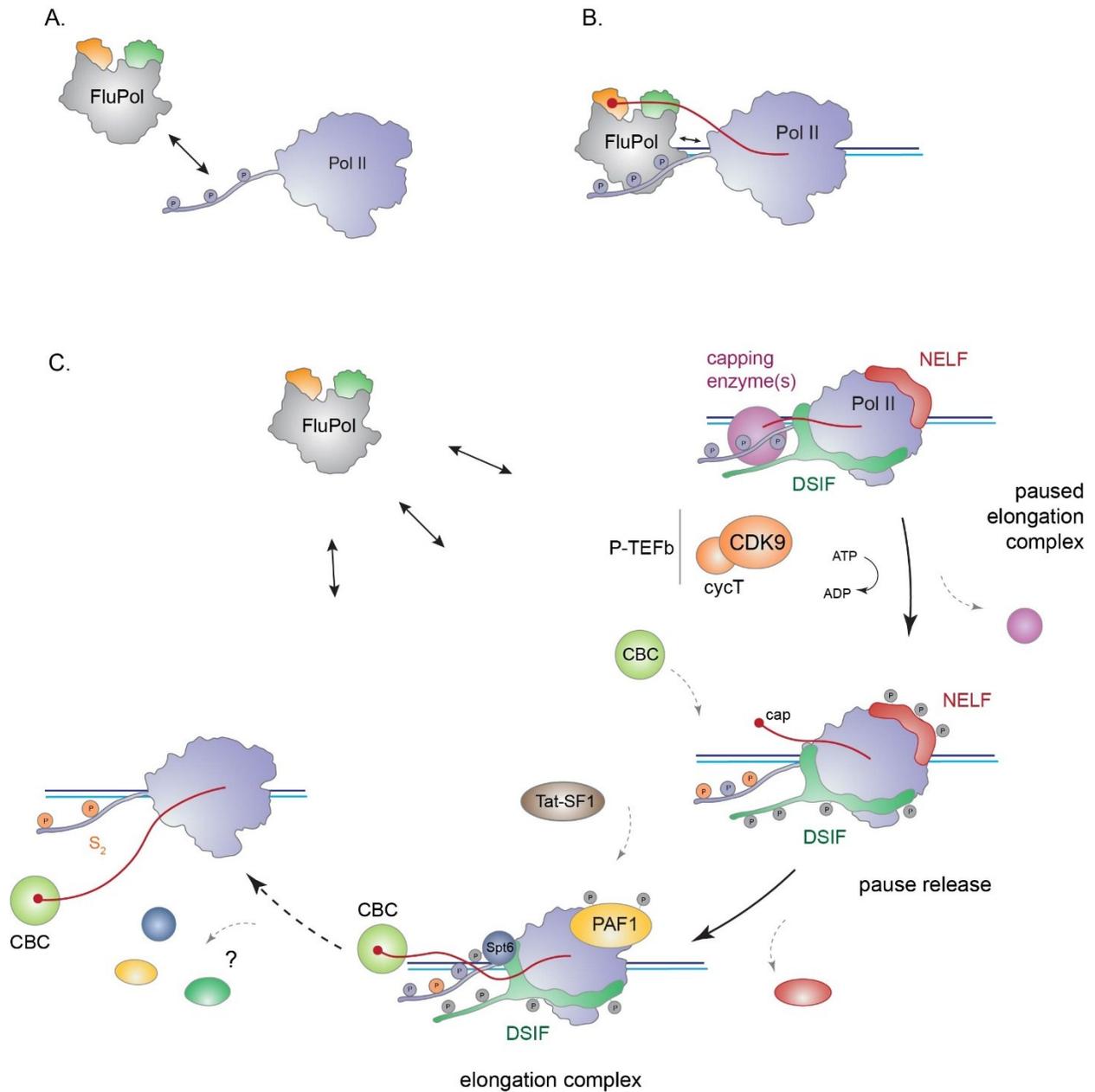
## 1.6. Scientific aims

The aim of this thesis project was to investigate the interaction between the viral and cellular transcription machineries in influenza-infected cells. At the beginning of the work, there was no data showing whether the interaction between FluPol and cellular Pol II is direct or mediated through other factors. I therefore sought to study this interaction *in vitro*. Circumventing the difficulties in expressing recombinant polymerases of human or avian strains, the purification of recombinant polymerase of influenza A/bat/Guatemala/060/2010/H17N10 and influenza B/Memphis/13/2003 was established in our group prior to the beginning of the current project. The polymerase of influenza A/H17N10 can perform its functions in human lung cells <sup>5</sup> and its sequence has around 70 per cent identity with human/avian A strains <sup>22</sup>, suggesting it is a good model for other influenza A polymerases.

Based on the study showing an association of FluPol with the SeP5 form of the CTD of Pol II, the first approach I used was to characterize *in vitro* the interaction of influenza A (H17N10) and influenza B polymerases with CTD-mimic peptides of various lengths and phosphorylations (section 3.1). This would help understanding the mechanism of recruitment of FluPol to the transcriptionally-engaged Pol II (Figure 6A).

The results obtained during the first two years confirming the direct interaction with the CTD, prompted me to investigate whether I could assemble a complex of FluPol with elongating full-length Pol II that would recapitulate the real cap-snatching situation (section 3.2). A previous cryo-electron microscopy structure and cross-linking mass-spectrometry analysis of the yeast capping enzyme, bound to transcribing Pol II, showed that besides associating with the CTD, additional interactions with the polymerase core position the capping enzyme at the right place for binding the emerging RNA. Similar to the capping enzyme, FluPol is recruited to the Pol II by interaction with the SeP5-CTD and must gain access to the nascent transcript, before the latter is tightly bound by the nuclear cap-binding complex. Together with the nuclear dynamics data, suggesting that FluPol can make additional contacts with other domains of Pol II <sup>96</sup>, this prompted me to study whether FluPol uses similar mechanisms to be positioned close to the exiting capped RNA. Therefore, I aimed to reconstitute an active cap-snatching complex from purified components, including transcribing Pol II with a nascent capped transcript emerging from its exit channel and FluPol in a transcriptionally active conformation (Figure 6B).

Finally, I wanted to investigate the time frame of FluPol action regarding the Pol II transcription cycle. The functional dependence on a capped RNA suggests that FluPol associates with a post-initiation state. The promoter-proximal paused state of Pol II, when the exiting RNA is already capped, but its elongation is stalled and the cap is probably not yet bound by the nuclear cap-binding complex, was thus an attractive target (Figure 6C). In light of the recent structures of Pol II elongation complexes, showing tight association of various negative or positive elongation factors to Pol II and extensive contacts with the nucleic acids, and based on the literature on interactions of FluPol with host-factors, I wanted to investigate what other factors might play a role in recruiting FluPol to the nascent transcript. I decided to focus on two elongation factors, DSIF and Tat-SF1, which have both been identified as FluPol interactors, and on the other side, are involved in transcription-proximal pausing of Pol II (section 3.3), and characterize their interaction with FluPol *in vitro*. This would help understanding the cellular context in which viral cap-snatching occurs.



**Figure 6. Graphical representation of the thesis aims.** (A) Investigating the interaction of FluPol with the CTD of Pol II. (B) Reconstituting a cap-snatching complex of FluPol and elongating Pol II. (C) Understanding the cellular context of cap-snatching by studying interaction with host-factors, involved in transcriptional regulation.

## **2. Materials and Methods**

## Résumé en Français

Ce chapitre présente les matériels et méthodes utilisés dans cette étude. La purification des protéines qui ont été utilisées pour les études d'interaction *in vitro*, les études structurales, la reconstitution des complexes de 'cap-snatching' et les analyses d'interaction, est décrite en détails. L'interaction avec les peptides modifiés du domaine C-terminal de l'ARN polymérase II a été suivie par des méthodes de l'anisotropie de fluorescence. Les structures de la polymérase d'influenza A et B en complexe avec ces mêmes peptides ont été obtenues par cristallographie et diffraction aux rayons X. Les tables détaillées de ces structures sont en annexe à la fin du manuscrit. Le suivi des activités de la polymérase a été fait, soit en utilisant des ARN substrats avec un marquage radioactif, soit en utilisant des ARN marqués avec des fluorophores. L'activité des mutants des polymérases a été suivie par des essais de 'mini-génomme'. Les virus récombinants ont été faits par les méthodes de génétique inverse.

Pour la reconstitution d'un complexe actif de 'cap-snatching', nous avons utilisé des protéines purifiées recombinantes ou endogènes (pour ARN polymérase II), assemblées sur des supports d'ADN et ARN pour imiter un contexte de ARN polymérase II en cours d'élongation. L'assemblage du complexe a été suivi par chromatographie d'exclusion de taille et certains complexes ont été étudiés par microscopie électronique. Les tests d'activités dans le contexte du complexe 'cap-snatching' ont nécessité des ARNs produits et coiffés *in vitro* avec un marquage radioactif.

Les interactions entre la polymérase de la grippe avec les facteurs d'élongation DSIF et Tat-SF1 ont été étudiés par chromatographie d'exclusion de taille et expériences de pull-down. Les effets sur l'activité de la polymérase dans le contexte 'cap-snatching' de ces facteurs ont été suivis par les mêmes tests d'activités utilisant des ARNs marqués d'une coiffe radioactive. L'assemblage des complexes ternaires avec les deux polymérases et les facteurs d'élongation a été étudié par chromatographie d'exclusion de taille et le complexe de la polymérase virale, ARN polymérase II et DSIF a été étudié par microscopie électronique.

## 2.1. Purification of recombinant influenza polymerase

The polymerases used in this work are from influenza A/little yellow-shouldered bat/Guatemala/060/2010 strain (FluA) and influenza B/Memphis/13/2003 (FluB) strains.

Influenza polymerase and mutants were expressed and purified as described<sup>22,23</sup>. Briefly, the polymerase was expressed as a self-cleaving polyprotein using the baculovirus system in insect cells (HighFive). The construct contained an N-terminal 8xHis tag, followed by PA-PB1-PB2 coding sequences separated by a cleavable linker, and a C-terminal Strep-tag. Purifications were performed from 3-4 L of HighFive cells. Cells infected with the V1 amplified virus were harvested 72 h post infection by centrifugation at 900 g for 20 min, flash-frozen in liquid nitrogen and stored at - 80 °C or lysed straightaway. The cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 10% (v/v) glycerol and 5 mM  $\beta$ -mercaptoethanol), supplemented with protease inhibitors (cOmplete, Roche) and lysed through sonication. Nucleic acids were removed by ammonium sulfate precipitation. The protein was purified by two-step affinity purification using Ni-NTA (Quiagen) and StrepTactin (IBA) affinity columns. Both affinity chromatography steps were performed using lysis buffer as running buffer. The protein was eluted from Ni-NTA with lysis buffer supplemented with 400 mM imidazole. The elution fraction was loaded on StrepTactin resin (IBA) and eluted with lysis buffer containing 2.5 mM desthiobiotin (DBT). The sample eluted from StrepTactin resin was diluted with 50 mM HEPES pH 7.5, 5 % (v/v) glycerol, 2 mM Tris(2-carboxyethyl)phosphine (TCEP) to a final concentration of 250 mM NaCl prior to loading on a heparin column (HiTrap Heparin HP, GE Healthcare). Gradient to 1 M NaCl was used to elute the protein, free of nucleic acids. The peak fractions were buffer exchanged (50 mM HEPES pH 7.5, 500 mM NaCl, 5 % (v/v) glycerol, 2 mM TCEP) and concentrated (using Amicon Ultra Centrifugal Filters with a molecular weight cut-off (MWCO) of 50 kDa) to 10 mg/mL. For the purification of endonuclease-deficient mutant (PA/K134A), a final size-exclusion chromatography step was performed. The heparin elution peak was concentrated and loaded on Superdex 200 10/300 GL column. The peak corresponding to monomeric protein was concentrated. All proteins were flash-frozen in liquid nitrogen and stored at - 80 °C.

## 2.2. Mutagenesis

All FluPol mutants used in this work were produced by site-directed mutagenesis using the QuickChange protocol (Stratagene). After PCR amplification, DNA was treated with DpnI and transformed in Top10 competent cells. All clones were confirmed by sequencing.

## 2.3. *In vitro* binding assays of FluPol and CTD peptides based on fluorescence anisotropy

The binding of CTD peptides to FluPol was followed by fluorescence anisotropy. The principle of the method is based on the change in the tumbling of a fluorescently-labelled ligand upon binding to a larger molecule. The emitted light of a free labelled ligand is depolarized because of its fast tumbling, and binding to a larger partner results in an increase of polarization and anisotropy (which are different mathematical interpretations of the same phenomenon)<sup>123</sup>. CTD peptides, labelled at the N-terminus with fluorescein amidite (FAM) (purchased from Covalab) were used to quantify binding to FluPol. Sequences of all used peptides can be found in Appendix Table 1. Assays were performed in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 5 mM MgCl<sub>2</sub>, 2 mM TCEP. For a titration experiment, 0.1 μM peptide was equilibrated with increasing concentrations of recombinant FluPol (or mutant variants) and equilibrium fluorescence anisotropy signals were recorded with a fluorescence spectrometer (Photon Technology International, with excitation/emission 495/520 nm at 23 °C). 5 min were enough to reach equilibrium, so all data were recorded at 5 min. In the experiments with FluPol bound to RNA promoter, v5'13 and v3'13 were used for FluA (or corresponding mutants) and v5'14 and v3'13 for FluB (or mutants). For the experiment with the cRNA promoter c5'14 and c3'15 were used (for all promoter sequences, see Appendix Table 2).

Binding isotherms were fitted using KaleidaGraph software to an equation assuming 1:1 binding. Dissociation constants ( $K_D$ ) were obtained by fitting the anisotropy ( $f$ ) as a function of the total protein concentration ( $P$ ):

$$f(P) = \frac{(L + P + K_D - \sqrt{(L + P + K_D)^2 - 4LP})}{2L} A + B$$

where  $L$  is the total concentration of fluorescent peptide,  $A$  is the maximum anisotropy and  $B$  is the offset.

Competition assays were performed in the same conditions but with 0.5  $\mu$ M FAM-labelled SeP5 CTD peptide incubated with equimolar amounts of promoter-bound FluA polymerase. Increasing amounts of non-labelled peptide were titrated to the preformed complex, competing with the labelled peptide. The apparent  $K_D$  was estimated using the same binding model (see above).

Triplicate experiments were averaged and plotted with error bars showing the standard deviation.

## **2.4. Structure determination of FluPol-CTD peptide complexes**

### **2.4.1. FluA polymerase-CTD crystallization and structure determination**

*Crystals of FluA polymerase were produced with the help of Alexander Pflug, who also provided the seeds used for seeding, assisted in data collection and solving the structure.*

Crystals of promoter-bound FluA polymerase were produced by vapor diffusion method at 4 °C with hanging drop. Drops were set at 9 mg/ml polymerase concentration and equimolar amounts of FluB RNA promoter v5'16 and v3'18 (Appendix Table 2) and four-repeat SeP5 CTD peptide (Appendix Table 1). The mother liquor contained 0.8 M sodium/potassium phosphate at pH 5.0. Crystallization was accelerated by seeding with existing RNA-bound crystals. The crystals for the seeding contained the same RNAs (v5'16 and v3'18) and are described in <sup>22</sup>. The obtained crystals after the seeding were soaked with 0.2  $\mu$ L of the peptide stock (5 mM) and flash-frozen in cryo-protectant solution containing mother liquor, supplemented with 25% (v/v) glycerol. Data were collected at the European Synchrotron Radiation Facility (ESRF), beamline ID29. Integration and scaling were performed with XDS suite <sup>124</sup> and structure was solved by molecular replacement at 2.5 Å resolution using PHASER and the FluA polymerase model (PDB: 4WSB). The electron density allowed to manually build the peptide in Coot <sup>125</sup>. Refmac5 and Phenix were used for refinement <sup>126, 127</sup>. Figures of the structures were made with PyMol <sup>128</sup>. Data collection and refinement statistics are shown in Appendix Table 5. Structure factors and co-ordinates are deposited in the Protein Data Bank (PDB) under the accession code 5M3H.

### **2.4.2. FluB polymerase-CTD crystallization and structure determination**

The first crystal form of FluB polymerase bound to CTD was grown in 0.1 M bicine, pH 9.0, 10% MPD. Polymerase at 9 mg/ml was mixed with FluB promoter v5'14 and v3'14 and 15-fold molar excess of four-repeat SeP5 CTD peptide. Crystals grew at 4 °C and were flash-frozen in mother liquor solution supplemented with 25% glycerol. Diffraction data were collected on ID23-1 beamline at the ESRF. The structure was solved at 3.5 Å resolution by molecular replacement with PHASER and the promoter-bound FluB polymerase structure as a search model (PDB:4WSA) and refined with Refmac5 and Phenix<sup>126, 127</sup>. Data collection and refinement statistics can be found in Appendix Table 5. Structure factors and coordinates are deposited in the PDB under the code 5M3J.

*The second crystal form of FluB in complex with the CTD (discussed in the manuscript), from which the model for site 2' was built, was produced by Petra Drncova (EMBL Grenoble).*

### **2.4.3. FluA polymerase crystallization with modified CTD peptides**

*The tetra-linker and thiophosphate peptides were synthesized by Michael Claron (EMBL Chemical Biology Core Facility).*

Crystals of FluA polymerase with both modified peptides (tetralinker and thiophosphate CTD) were obtained in the same way as described above for the initial FluA crystal form. Diffraction data were collected on beamline ID23-1 of the ESRF, integrated and scaled using the XDS suite and solved by molecular replacement using PHASER and the CTD-bound FluA polymerase structure (PDB:5M3H) and refined with Refmac5. Data collection and refinement statistics are shown in Appendix Table 5.

## **2.5. Polymerase activity assay with radiolabeled capped RNA**

*FluB polymerase for the activity assays with CTD peptides was provided by Joanna Wandzik.*

To measure the effect of CTD peptides on polymerase functions, activity assays were performed with radiolabeled capped RNA produced by *in vitro* transcription (see section 2.10). 23 nt RNA was capped with vaccinia capping enzyme and 2'-O-methyltransferase (NEB) using [ $\alpha^{32}\text{P}$ ]-GTP.

Capped RNA was incubated with the respective amounts of FluA and FluB polymerase (25 nM for endonuclease assay with FluA, 50 nM for transcription assay with FluA and 50 nM for both endonuclease and transcription with FluB) in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 2 mM TCEP. 1.2-fold molar excess of v5'14 RNA was added to each polymerase (FluA and FluB v5'14 sequences are shown in Appendix Table 2). Transcription assays were performed in the presence of 150 nM v3'18 vRNA template and 250 μM NTP mix (ThermoFisher). CTD peptides were added at 50 μM in order to be at a concentration of at least 10 times the K<sub>D</sub> of the lowest measured affinity (two-repeat peptide). Two and four-repeat peptides were purchased from Covalab and the six-repeat peptide was synthesized by Michael Claron from the Chemical Biology Core Facility at EMBL Heidelberg. Reactions were incubated at 30 °C for 30 min and quenched with RNA loading dye (formamide, 8 M urea, 0.1 % SDS, 0.01 % bromophenol blue (BPB), 0.01 % xylene cyanol (XC)), supplemented with 50 mM EDTA and boiled at 95 °C. Samples were separated on 20 % denaturing acrylamide gel (containing 8 M urea) in Tris-Borate-EDTA (TBE) buffer, exposed on a Storage Phosphor screen and recorded with a Typhoon reader. DECADE<sup>TM</sup> marker system was used as an RNA ladder. The size of the capped products did not directly correspond to the size of uncapped RNA products from the ladder, due to the different electrophoretic mobility of the cap structure, as reported <sup>129</sup>.

## 2.6. Mini-genome assay

The polymerase activity in a cellular context was determined using a mini-genome assay. Briefly, expression plasmids encoding the three polymerase subunits and nucleoprotein subunits were co-transfected with pPolII-Luciferase containing firefly luciferase reporter in negative polarity and the 5' and 3' promoter region of segment 5 <sup>130</sup>. The luciferase RNA produced by Pol I can assemble with the polymerase and nucleoprotein into vRNPs. The amount of produced reporter, which can be detected with a luciferase assay, is directly dependent on the activity of the polymerase within the vRNPs. The mini-genome assays for FluA were performed using influenza A/WSN/33(H1N1) and A/Victoria/3/75/H3N2 strains. HEK293T cells were transfected in 12-well plates (using Lipo293 transfection reagent) with expression plasmids for polymerase, nucleoprotein and NS2 protein: pcDNA3-PA (5 ng), pcDNA3-PB1 (10 ng), pcDNA3-PB2 (10 ng), pcDNA3-NP (100 ng), pCI-NS2 (40 ng) and 100 ng pPolII-Luciferase. pRenilla-TK from Promega was used as a

transfection control (100 ng). Cells were lysed 24h post transfection and firefly and Renilla luciferase activities were measured according to manufacturer instructions (Promega), using Clariostar microplate reader (BMG Germany). The average activity was calculated from three independent experiments performed in triplicate. Results were plotted with KaleidaGraph, error bars indicate the standard deviation from the three experiments.

For the FluB minireplicon assay, cells were transfected with 10 ng plasmids for each polymerase subunit, 100 ng nucleoprotein of influenza B/Memphis/13/2003 strain and 100 ng pPoll-Luciferase. Measurements were performed as described above for FluA.

## **2.7. Expression of polymerase in mini-genome assays**

*Transfection and Western blot analysis of the polymerase expression was performed by Guillaume Fournier (in the laboratory of Nadia Naffakh, Institut Pasteur).*

The expression of polymerase and nucleoprotein was followed by Western blotting. The expression of FluA/WSN (Figure 21A) was controlled on the sample prepared as for the RNA quantification assay (section 2.8). Rabbit anti-PA (provided by B. Delmas) and mouse anti-nucleoprotein antibody (Kerafast) were used. Total protein amount was controlled by the expression of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (antibody purchased from Genetex).

For FluB, single PA or PB2 subunits were transfected with Lipo293 reagent (0.5  $\mu$ g of each plasmid in a 6-well plate). Custom made rabbit anti-PA and anti-PB2 antibodies were used for detection and total protein amount was controlled by detection of  $\beta$ -actin (Abcam).

## **2.8. RNA quantification in the mini-genome assay**

*Detection of the RNA using RT-qPCR was performed essentially as described <sup>131</sup>. Experiments were performed by Guillaume Fournier.*

In brief, HEK293T were transfected with 150 ng pcDNA3-PA, 150 ng pcDNA3-PB1, 150 ng pcDNA3-PB2, 150 ng pCI-NS2, 300 ng pcDNA3-NP and 60 ng pPoll-NA, encoding the

NA gene in negative polarity. RNA was extracted with RNeasy mini kit (QIAGEN), and treated with DNase (TURBO DNase, Ambion). Strand-specific RT-qPCR for NA vRNA, cRNA and mRNA were performed. Synthetic viral RNAs were used to generate standard curves. Reactions were followed on LightCycler 480 (Roche). RNA copy numbers were normalized to GAPDH mRNA. Experiments were performed in triplicates.

## **2.9. Reverse genetics**

*Reverse genetics experiments were performed as described*<sup>131</sup> *by Guillaume Fournier and Catherine Isel-Griffiths.*

Recombinant viruses were generated according to<sup>132, 133</sup> with modifications. A plaque assay was used to evaluate the efficiency of reverse genetics by titrating the supernatants on MDCK cells. Viral RNA was extracted upon plaque-purification and amplification using QIAamp Viral RNA Mini kit (QIAGEN) and subjected to reverse transcription (using SuperScript), amplification and Sanger sequencing. One-Step RT-PCR kit (Invitrogen) was used to amplify the 5' and 3' termini of the PA segment. Reverse transcription and amplification of all the segments were performed by RT-PCR for the Next-Generation sequencing (NGS)<sup>134</sup>. NGS was done at Institut Pasteur, PIBnet platform using the Nextera XT DNA Library Preparation kit (Illumina), the NextSeq 500 sequencing systems (Illumina), and the CLC Genomics Workbench 9 software (Qiagen) for analysis.

## **2.10. *In vitro* polymerase activity assays**

A fluorescence-based method for RNA synthesis detection previously developed in the laboratory<sup>27</sup>, was used to compare the RNA synthesis kinetics of wild type and mutant polymerases. The reaction mixture contained 250 nM polymerase bound to 5' promoter – v5'14 vRNA or c5'14 cRNA, 150 nM fluorescently-labelled template v3'18 or c3'18, 500 nM capped primer (5'-m7GpppAAUCUAUAAUAG-3') (for transcription reactions). The assay buffer was the same as for the CTD binding assays. Reactions were initiated with addition of 25 or 500  $\mu$ M NTPs (for transcription or replication, respectively) and quenched at indicated reaction times with

4.5 M NaCl. Assays were performed at 24 °C. The resulting change in fluorescence polarization due to association of the product RNA to the fluorescent template was recorded at Clariostar microplate reader (BMG Germany) with excitation/emission of 485/520 nm. A single exponential equation was used to fit the unprimed replication reactions:

$$f(t) = -Ae^{-kt} + B .$$

The transcription time-course was fitted to a double exponential equation, because of the observed bi-phasic behavior:

$$f(t) = (-Ae^{-k_1t} + B) + (-Ce^{-k_2t} + D) .$$

$A$  and  $C$  are observed polarization amplitudes,  $B$  and  $D$  are final polarization values for each phase;  $t$  is the time and  $k$  the observed rate constants.

The average of triplicates were plotted with KaleidaGraph software. Error bars show the standard deviation.

### **2.11. *In vitro* transcription, purification and capping of RNAs**

RNAs for the reconstitution and activity assays with Pol II were produced *in vitro* using T7 transcription system. Single-stranded DNA oligonucleotides with double-stranded T7 promoter region were used as templates. Reaction buffer contained 40 mM Tris/HCl pH 8.4, 1 mM spermidine, 0.01 % Triton X-100, 30 mM MgCl<sub>2</sub>, 5 mM dithiothreitol (DTT), 4 mM each NTP, 2 μM DNA template (pre-heated at 95 °C for 3 min), 0.2 mg/mL T7 polymerase. The reactions were incubated at 37 °C for 4 h, magnesium pyrophosphate precipitate was removed by centrifugation and the RNA was extracted by phenol/chloroform extraction and precipitated with iso-propanol. Each RNA was separated on a denaturing acrylamide gel and a single band with the corresponding size was gel-purified, precipitated with ethanol and resuspended in DEPC-treated H<sub>2</sub>O. RNAs were capped with recombinant vaccinia capping enzyme (VCE) (produced by Delphine Guilligay) and commercial 2-O'-methyltransferase (NEB). Buffer for the capping reaction contained: 50 mM Tris/HCl (pH 8), 6 mM KCl, 1.25 mM MgCl<sub>2</sub>, 1 mM DTT, 0.1 mM S-adenosyl-methionine (SAM), 0.5 mM GTP (for radiolabeled capped RNA, [<sup>32</sup>P]-GTP was used), 1 μM VCE and 50 units 2-O'-methyltransferase. Capping reactions were

incubated for 2 h at 37 °C and the product RNAs were precipitated with ethanol and resuspended in DEPC-H<sub>2</sub>O. Sequences of the produced RNAs are shown in Appendix Table 3.

## **2.12. FluPol – Pol II complex assembly**

*Pol II and P-TEFb used throughout this study were provided by Seychelle Vos from Patrick Cramer's group at the Max Planck Institute for Biophysical Chemistry, where the initial complex assembly experiments were performed with her help.*

FluPol was purified as described in section 2.1. FluPol/PA-K134A mutant was purified as described in section 2.1, with an additional size-exclusion chromatography step at the end, using S200 column and collecting the monomeric fraction, running at retention volume of around 12 mL. Pol II is isolated from *S. scrofa* thymus, as described in <sup>135</sup> and P-TEFb is recombinantly expressed in insect cells using the baculovirus system <sup>82</sup>. Complex assembly was performed in a buffer containing 50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 4 % glycerol, 3 mM MgCl<sub>2</sub>, 2 mM TCEP. All complex assembly reactions were performed in 45 µL. For sequences of the capped RNAs, template and non-template DNA, refer to the Appendix Table 3 and 4.

The elongation scaffold, containing 20 µM template DNA, 20 µM capped RNA and 30 µM non-template DNA was pre-annealed by incubation at 95 °C for 4 minutes and gradual cooling down to 30 °C by 1 °C per minute. For complex assembly, 25 pmol Pol II was pre-incubated with 50 pmol of the DNA/RNA scaffold in 45 µL at 30 °C for 20 min. 18 pmol P-TEFb, 1 mM ATP, 60 pmol FluPol(PA/K134A), pre-incubated with 1.2 molar excess of vRNA promoter (v5'15 and v3'14) were added and incubated with the Pol II/nucleic acid complex for additional 25 min at 30 °C. The sample was spun down at 20 000 g at 4 °C to remove aggregates and injected on Superdex 200 Increase 3.2/300 column equilibrated in the complex assembly buffer. Samples were analyzed on 4-20 % gradient SDS-PAGE (NuPAGE, ThermoFisher) and stained with Instant Blue.

Size-exclusion chromatography of FluPol alone was performed using 60 pmol FluPol pre-bound to vRNA promoter.

Size-exclusion chromatography of Pol II was performed in presence of P-TEFb: 25 pmol Pol II was incubated with 18 pmol P-TEFb and 1 mM ATP.

To prevent phosphorylation of Pol II, ATP was omitted from the reaction. For detection of the phosphorylated forms of Pol II, Western blot was performed on the peak fractions, where Pol II was present. Anti-SeP5 (3E8) and anti-SeP2 (3E10) antibodies were incubated in PBS/Tween 0.1 % with the blotted membrane and a secondary antibody anti-rat-HRP was used for detection. For details on the specificity of the antibodies, see <sup>136</sup>.

The stoichiometry of the two polymerases was determined by incubating FluPol and Pol II in different molar ratios, as described in the results section 3.2.1.4 (28 pmol Pol II and 60 pmol (1:2), 140 pmol (1:5) and 507 pmol (1:18) FluPol(PA/K134A)).

The experiment with mutant FluPol(K289A/R449A) was performed in the same conditions as described above for the complex assembly with wild type FluPol.

Tat-SF1 and DSIF for the assembly of ternary complexes were provided by Seychelle Vos.

The ternary complex assembly experiments with DSIF and Tat-SF1 were identical to the assembly experiment described above, with the addition of 50 pmol of the respective factor. In brief, Pol II/scaffold complex was pre-incubated (20 min at 30 °C) and added to 0.4 μM P-TEFb, 1 mM ATP, 50 pmol DSIF/Tat-SF1, 60 pmol FluPol(PA/K134A) in the assembly buffer, incubated for 25 min and analyzed by SEC.

### **2.13. Functional assays of FluPol – Pol II complex**

All activity assays were performed in buffer, containing 50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 5 mM MgCl<sub>2</sub> and 2 mM TCEP. A 10x stock was prepared and 2 μL were added to each reaction of 20 μL. When suitable, master mixes were prepared. DEPC-H<sub>2</sub>O was used for dilutions.

Pol II stock was diluted in the reaction buffer to 1 μM.

FluPol was pre-bound to 1.2x molar excess of 5'vRNA promoter (v5'14) and diluted to 250 nM in buffer, containing 50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 5 % glycerol, 2 mM TCEP.

In general reactions were performed in 20 μL (unless otherwise stated) and stopped with addition of 20 μL 2x loading dye (formamide, 8 M urea, 50 mM EDTA, 0.1% SDS, 0.01% BPB,

0.01% XC). Samples were boiled at 95 °C for 3 min and loaded on denaturing urea gel (20 % acrylamide), run in TBE. Gels were dried for 5 min, exposed to Storage Phosphor screen over night and revealed with Typhoon reader. DECADE ladder was used as an approximative marker (see section 2.5).

### **2.13.1. Cleavage assays with different scaffolds**

The DNA scaffolds used for the Pol II walking experiment are listed in Appendix Table 4 (+1 to +5). 2  $\mu$ L of single-stranded template DNA (tDNA) (15  $\mu$ M) was pre-annealed with 2  $\mu$ L of the radiolabeled 23-mer capped RNA stock (estimated concentration is maximum 7.5  $\mu$ M) (or 28-mer for the experiment on Figure 41) by heating them at 55 °C for 5 min and cooling down gradually to 20 °C by 1 °C/min in a PCR machine. The RNA/tDNA scaffold stock was diluted 50 times in DEPC-H<sub>2</sub>O and 1.5  $\mu$ L (15/7.5 nM of tDNA/capped RNA) was added to 2  $\mu$ L of 1  $\mu$ M Pol II, followed by incubation at 30 °C for 20 min. 1  $\mu$ L of 500 nM ntDNA was added to the Pol II/scaffold (tDNA/RNA) and further incubated for 10 min, then supplemented with 0.4  $\mu$ M P-TEFb, 1 mM ATP (for the kinase) and 1 mM UTP (for elongation by Pol II) in reaction buffer (final volume 20  $\mu$ L). Reactions were allowed to proceed for 30 min at 30 °C before 10  $\mu$ L of the mixture was quenched by addition of 2x loading dye. The rest of the reaction (10  $\mu$ L) was supplemented with 25 nM FluPol/5', incubated for another 30 min at 30 °C and quenched with loading dye.

### **2.13.2. Effect of DSIF on the FluPol recognition of capped RNA, pre-bound to Pol II**

The experiment was performed as described for the Pol II walking (section 2.13.1) with the difference that after elongation by Pol II, an incubation step with 5 pmol DSIF was added for 30 min, before the addition of FluPol.

### **2.13.3. NTP incorporation by FluPol**

The differential incorporation of NTPs by FluPol was followed using 200 nM FluPol/v5'14 (200 nM FluPol and 240 nM v5'14), 150 nM v3'18 vRNA template, 7.5 nM radioactively labeled capped RNA primer (maximal estimated concentration) and 250  $\mu$ M of the respective individual NTP (ATP, UTP, GTP or CTP, ThermoFisher) or a combination of them. The reaction (20  $\mu$ L) was incubated at 30 °C for 30 min and quenched with 20  $\mu$ L loading dye.

#### **2.13.4. Pol II effect on FluPol endonuclease activity**

20-mer and 28-mer radiolabeled capped RNAs were pre-annealed with 43 nt tDNA as described above (section 2.13.1). For the experiments where capped RNA was added *in trans*, the same size of capped RNAs were used as for the *cis* experiments, but the RNA was not pre-bound to tDNA and Pol II. 7.5 nM of the scaffold or capped RNA (maximal estimated concentration) was incubated with 0.5  $\mu$ M Pol II for 10 min at 30 °C. 160 nM ntDNA were added, followed by 10 min incubation. Then, the reactions were initiated by adding 25 nM FluPol/v5'14 in reaction buffer for 30 min. The final concentration of Pol II in the reaction was 25 nM. The reactions with CTD peptide contained 50  $\mu$ M six-repeat SeP5 peptide. Reactions with phosphorylated Pol II were supplemented with 0.4  $\mu$ M P-TEFb and 1 mM ATP.

#### **2.13.5. NTP incorporation by Pol II in presence of FluPol (PA/K134A) mutant**

The experiment with the endonuclease-deficient mutant was performed by annealing Pol II and the 28-mer capped RNA and 43 nt tDNA/ntDNA scaffold, as described in the previous section (2.13.3). For the assay of Pol II-catalyzed elongation, the reaction was initiated with 1 mM UTP for 20 min. 1 mM ATP was added to match the conditions of the assay with P-TEFb. For the reaction following the effect of FluPol mutant on Pol II, Pol II was pre-incubated with 100 nM FluPol(PA/K134A)/v5'14, 0.4  $\mu$ M P-TEFb and 1 mM ATP. After 30 min, 1 mM UTP was added and the reaction was incubated for another 30 min. To test for FluPol(PA/K134A) residual endonuclease activity, a control experiment was set with the capped RNA scaffold and FluPol(PA/K134A) alone.

#### **2.13.6. Effect of DSIF and Tat-SF1 on FluPol endonuclease activity**

The effect of DSIF and/or Tat-SF1 (purified as described in section 2.16 and 2.19 and diluted to 5  $\mu$ M stocks in their corresponding SEC buffers) was followed in the same conditions as described in section 2.13.3. Briefly, 7.5/15 nM pre-annealed scaffold (28-mer capped RNA/43 nt tDNA) were incubated with 0.5  $\mu$ M Pol II (10 min) and 25 nM ntDNA (10 min). 2  $\mu$ L 10x reaction buffer was added, together with 0.4  $\mu$ M P-TEFb, 1 mM ATP and 250 nM DSIF or Tat-SF1 and 25 nM FluPol, followed by 30 min incubation. The final concentration of Pol II was 25 nM.

## **2.14. Negative stain electron microscopy**

### **2.14.1. Sample preparation for FluPol/Pol II complex**

FluPol/Pol II complex for negative stain EM was prepared following a similar protocol as described in section 2.12. Mismatch bubble scaffold was pre-annealed as described in section 2.12, using 28-mer capped RNA, tDNA and ntDNA of 39 nt (for sequences see Appendix Table 3 and 4). A shorter DNA scaffold was chosen compared to the functional assays to avoid unspecific Pol II-Pol II interactions. 58 pmol Pol II was mixed with the 100 pmol of the capped RNA/DNA scaffold for 20 min at 30 °C. 100 pmol of promoter pre-bound FluPol(PA/K134A) (v5'14 and v3'13, see Appendix Table 2), 18 pmol P-TEFb, 1 mM ATP and reaction buffer (50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 4 % glycerol, 3 mM MgCl<sub>2</sub>, 2 mM TCEP) were added and incubated for 25 min at 30 °C in a 45 µL final volume. The complex was applied to size-exclusion chromatography, (Superdex 200 Increase 3.2/300 column, equilibrated in reaction buffer) and the peak fractions were analyzed by SDS-PAGE and stained with Instant Blue. The peak fractions, containing FluPol(PA/K134A) and Pol II were cross-linked with 0.1 % glutaraldehyde for 10 min on ice and quenched by dialysis against 20 mM Tris/HCl pH 8, 50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 3 mM MgCl<sub>2</sub>, 1 mM TCEP using a 20 kDa MWCO Slide-A-Lyzer-MINI Dialysis unit for 4 h, with buffer exchange after 2h.

### **2.14.2. Negative stain analysis of FluPol/Pol II complex**

*Negative stain data was collected and analyzed with the help of Manikandan Karuppasamy (EMBL Grenoble).*

CF300-Cu grids (EMS) were glow-discharged at 25 mA, 0.3 mbar (15 s glow and 60 s hold). 5 µL of the sample at concentration of 24 nM was applied on the grid and incubated for 2 min, washed with 30 µL of H<sub>2</sub>O for 30 sec, and floated consecutively on three drops of 20 µL of 2 % uranyl acetate. After the last staining step, the grid was incubated for 1.5 min and excess stain was removed by filter paper. Images were taken at FEI Tecnai T12 electron microscope at EMBL Grenoble, operated at 120 kV and a magnification of 49000x.

### **2.14.3. Negative stain of FluPol/Pol II/DSIF complex**

*DSIF for this experiment was provided by Seychelle Vos, who also helped with data collection.*

Sample preparation for FluPol/Pol II-DSIF complex followed the protocol as described in section 2.14.1 with the following modifications: 43 nt DNA / 23-mer capped RNA scaffold was used to stabilize DSIF binding to Pol II. The complex was assembled with addition of FluPol(PA/K134A), P-TEFb and ATP, supplemented with 100 pmol DSIF. The cross-linked complex was quenched by addition of 8 mM aspartate and 2 mM lysine and incubation for 10 min on ice prior to dialysis. 5  $\mu$ L of the sample at 85 nM concentration were applied to CF300-Cu grids and stained with uranyl formate (saturated solution) using identical protocol as described above. The grids were imaged on a CM200 microscope, operated at 120 kD with a magnification of 88000x.

## **2.15. Cryo-electron microscopy on FluPol/Pol II complex**

### **2.15.1. Sample preparation**

A 39 nt DNA/28-mer capped RNA scaffold (Appendix Table 3 and 4) was used for the assembly. 100 pmol of Pol II were premixed with 150 pmol DNA/RNA scaffold and incubated for 20 min at 30 °C in 80  $\mu$ L complex assembly buffer (50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 4 % glycerol, 3 mM MgCl<sub>2</sub>, 2 mM TCEP). 120 pmol FluPol(PA/K134A), 32 pmol P-TEFb and 1 mM ATP were added and incubated for 25 min at 30 °C. At these concentrations of the components, the complex could not be eluted from SEC, probably due to unspecific interactions with the column. Instead, the sample was applied on a 4.5 mL 10-30 % sucrose gradient. 0.1 % glutaraldehyde was added to the heavy sucrose solution prior to mixing. Sedimentation was performed using SW60 rotor at 32000 rpm for 16 h at 4 °C. 150  $\mu$ L fractions were collected and quenched with 5 mM aspartate, 5 mM lysine and 5 mM Tris/HCl pH 8. 2  $\mu$ L of each fraction were plotted on nitrocellulose membrane, blocked with 3 % BSA in PBS/Tween 0.05 % for 1 h, incubated with anti-His-HRP antibody (Sigma). The fractions containing FluPol were dialyzed against 20 mM Tris/HCl pH 8, 50 mM HEPES/NaOH, 150 mM NaCl, 3 mM MgCl<sub>2</sub>, 1 mM TCEP with 20 kDa MWCO dialysis cups for 4 hours. Sample was concentrated using Amicon Ultra Centrifugal Filters with a cutoff of 100 kDa to around 110 nM. R2/2 Cu grids (Quantifoil) were

glow-discharged for 15 s (with 30 s hold) at 25 mA and 0.3 mbar. 3  $\mu$ L of the sample were applied to the grids, and blotted for 1 s with blotting force 5 with a Vitroblot (FEI), at 4 °C with 100 % humidity.

### **2.15.2. Data collection**

*Data acquisition was performed by Maria Bacia (Institute of Structural Biology, Grenoble).*

Data were collected on a Tecnai Polara microscope (FEI) with a K2 Summit (Gatan) detector, operated at 300 kV. Automated data acquisition was performed with a nominal magnification of 31000x and a pixel size of 1.21 Å. 1145 image stacks of 40 frames over 8 s were collected.

Frames were processed with MotionCorr 2<sup>137</sup>. CTF correction was performed with CTFFIND4<sup>138</sup>. Image processing was performed with RELION 2.1.<sup>139</sup>. Particles from 100 micrographs were automatically picked with Gautomatch and 6869 particles were extracted with a box size of 336<sup>2</sup> pixels. An initial 2D classification provided a template for auto-picking in RELION. 159282 particles were extracted and subjected to 2D classification with a mask diameter of 300 Å. 52692 particles remained after screening and polishing. 3D classifications were performed with initial model of Pol II alone (EMD-3218), low-pass filtered to 50 Å, using a mask diameter of 300 Å.

### **2.16. Purification of full-length DSIF**

*The DSIF expression construct was provided by Carrie Bernecky (Institute of Science and Technology Austria).*

Recombinant DSIF was expressed in *E.coli*, as described<sup>68</sup>. The expression construct contains the sequence of human Spt4 and Spt5 cloned in a modified pETDuet vector, with a N-terminal His<sub>10</sub>-Arg<sub>8</sub>- SUMO-tag and a 3C protease cleavage site, followed by first cassette with Spt4 and a second cassette with Spt5.

DSIF was produced in BL21(DE3)RIL *E.coli* strain and grown to an OD<sub>600</sub> of 0.6. Expression was induced with addition of 1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and cells were grown at 37 °C for 3 h. The media was supplemented with 10  $\mu$ M of ZnCl<sub>2</sub> during expression and cells were harvested by centrifugation. Purification steps were performed at 4 °C. Cells were lysed

by sonication in lysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10 % glycerol, 10  $\mu$ M ZnCl<sub>2</sub>, 150 mM imidazole pH 8, 1 mM DTT, supplemented with cOmplete protease inhibitors). Insoluble material was removed by centrifugation and the supernatant was loaded on Ni-NTA column. Protein was eluted with a gradient to 400 mM imidazole and the peak fractions were cleaved with 3C protease over night and dialysed against 50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10 % glycerol, 10  $\mu$ M ZnCl<sub>2</sub>, 150 mM imidazole, 1 mM DTT. Prior to loading on a second Ni-NTA, the sample was diluted to a final concentration of 250 mM NaCl. The flow-through fraction, containing cleaved DSIF protein was loaded directly on HiTrap Q column and eluted with gradient to 1 M NaCl. The peak fraction, eluting around 400 mM NaCl, was concentrated and loaded on Superdex 200 10/300 GL SEC column in SEC buffer, containing 20 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10 % glycerol, 10  $\mu$ M ZnCl<sub>2</sub>, 1 mM DTT. The middle fractions of the peak, containing the smallest amount of N-terminal degradation product of Spt5 were concentrated to around 2.9 mg/mL, flash-frozen in liquid nitrogen and stored at - 80 °C.

### **2.17. Cloning, expression and purification of DSIF CTR domain**

The CTR domain was defined based on <sup>68,140</sup>, containing residues 754-950 of Spt5. The coding sequence was generated by PCR amplification of the full-length construct and cloned in pGEX6p1 vector (*gift from Valentina Speranzini, EMBL Grenoble*). The CTR sequence was preceded by a N-terminal GST tag and 3C cleavage site. The plasmid was transformed in *E.coli* Rosetta strain, grown at 37 °C until an OD<sub>600</sub> of around 0.6. Expression was induced with 1 mM IPTG at 18 °C. After 16 h, cells were harvested by centrifugation, lysed by sonication in 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10 % glycerol and 1 mM DTT buffer. The clarified lysate was incubated over night with Gluthatione Sepharose resin and the GST-CTR eluted with 10 mM Gluthatione in lysis buffer. The sample was cleaved with GST-3C protease and applied on a Gluthatione Sepharose. The untagged protein was collected from the flow-through fraction, concentrated and injected on SEC using Superdex 75 10/300 GL column. Peak fractions were concentrated with Amicon concentrator (MWCO 10 kDa) to around 4 mg/mL, flash-frozen and stored at - 80 °C.

## 2.18. Purification of KOW<sub>x</sub>-5 domain

*The construct for the KOW<sub>x</sub>-5 domain was provided by Carrie Bernecky.*

KOW<sub>x</sub>-5 domain covers amino acids 526-754 of Spt5. The open reading frame was cloned in pOPINB vector with a N-terminal His<sub>6</sub>-tag, followed by 3C protease cleavage site. The plasmid was transformed in *E.coli* BL21(DE3)RIL cells, which were grown at 37 °C to an OD<sub>600</sub> of around 0.6. Expression was induced with 1 mM IPTG for 4 h at 37 °C. After harvesting cells were lysed by sonication in 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10 % glycerol buffer. The soluble lysate was loaded on Ni-NTA, eluted with lysis buffer supplemented with 250 mM imidazole. Cleavage with 3C protease was performed over night, in dialysis against lysis buffer with 10 mM imidazole. After a second Ni-NTA, the flow-through fraction containing the protein without tag was diluted to 250 mM NaCl and loaded on heparin column. Elution with gradient of up to 1 M NaCl was performed. The peak (around 750 mM NaCl) was concentrated and applied on Superdex 75 10/300 GL SEC column. The fractions containing the target protein were concentrated (using Amicon concentrator with MWCO of 3 kDa) to around 6.3 mg/mL, flash-frozen in liquid nitrogen and kept at - 80 °C.

## 2.19. Tat-SF1 purification

*Expression construct for Tat-SF1 was provided by Seychelle Vos.*

The construct for recombinant human Tat-SF1 contained an N-terminal His<sub>6</sub>-tag, followed by a TEV cleavage site. The protein was produced in insect cells using the baculovirus system<sup>141</sup>. HighFive cells were used for expression and were harvested 72 h post-infection with V1 amplified virus. Cells were lysed by sonication in 20 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10 % glycerol, 30 mM imidazole pH 8, 1 mM DTT, supplemented with cOmplete protease inhibitor mix (Roche). After centrifugation, the soluble supernatant was loaded on Ni-NTA resin and target the protein was eluted with 500 mM imidazole. The His-tag was cleaved with TEV protease over night in dialysis against lysis buffer and applied on a second Ni-NTA. The untagged protein was collected from the flow-through fraction of a second Ni-NTA affinity chromatography,

concentrated and subjected to SEC (Superdex 200 10/300 GL column). Peak fractions were concentrated to around 6.5 mg/mL, flash-frozen in liquid nitrogen and stored at - 80 °C.

## **2.20. Analytical SEC studies on the interaction between FluPol and elongation factors**

### **2.20.1. Interaction between FluPol and full-length DSIF or DSIF domains**

*KOWx-4 and KOW6-7 proteins, used for SEC experiments, were provided by Carrie Bernecky.*

All analytical gel-filtration and pull-down experiments with DSIF were performed in buffer, containing 50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 4 % glycerol, 3 mM MgCl<sub>2</sub>, 2 mM TCEP.

SEC analysis of FluPol interaction with full-length DSIF or DSIF domains were performed using 2.3 μM FluPol, prebound to v5'13 (Appendix Table 2) and 3-fold molar excess of DSIF or the respective domains. The proteins were incubated for 20 min at 30 °C. For the experiment with P-TEFb, 0.4 μM of the kinase and 1 mM ATP were mixed with the FluPol/DSIF and incubated for 20 min at 30 °C. Samples were clarified by centrifugation and subjected to SEC on Superdex 200 Increase 3.2. Fractions were analyzed by SDS-PAGE (4-20 % NuPAGE) and stained with Instant Blue.

### **2.20.2. SEC analysis of FluPol and Tat-SF1**

Analytical SEC was performed identically to the experiments with DSIF (section 2.20.1).

## **2.21. Pull-down experiments of FluPol and elongation factors**

### **2.21.1. Pull-down experiments on FluPol and DSIF**

For the pull-down experiments, the complexes were incubated as described in section 2.20 and loaded on 30  $\mu$ L StrepTactin resin (IBA). 1.6  $\mu$ M FluPol/v5'13 were mixed with 3-fold molar excess of DSIF with or without P-TEFb. For the assay with cRNA, c5'14 (Appendix Table 2) was added to the polymerase. The sample mixtures were incubated with the resin for 30 min at 4  $^{\circ}$ C, and washed three times with 200  $\mu$ L of the assembly buffer (section 2.20.1), and eluted with 30  $\mu$ L of the same buffer, supplemented with 2.5 mM DBT. Samples were analyzed on SDS-PAGE (4-20 % NuPAGE) and stained with Instant Blue.

### **2.21.2. Pull-down experiments on FluPol and Tat-SF1**

Pull-down experiments of FluPol and Tat-SF1 were performed identically to the ones with DSIF (section 2.21.1). For the FluPol/DSIF/Tat-SF1 pull-down, both factors (DSIF and Tat-SF1) were added in 3-fold molar excess to the FluPol/v5'14.

### **3. Results**

## Résumé en Français

Ce chapitre présente les résultats obtenus au cours de la thèse. La première partie décrit les études sur les interactions entre la polymérase de la grippe et le domaine C-terminal (CTD) de l'ARN polymérase II. Tout d'abord sont présentés les résultats qui ont démontré qu'une interaction directe existe entre les peptides du domaine CTD et la polymérase virale. Cette interaction est dépendante de la phosphorylation de ces peptides, et notamment le Ser5. Deux structures de la polymérase du virus d'influenza A et B, en complexe avec des peptides phosphorylés sur Ser5, ont permis de comprendre les détails de ces interactions, ce qui pourra amener à la conception des molécules capable de l'inhiber. Les détails de l'interaction ont été investigués grâce à une méthode de criblage à haut-débit de peptides de longueurs et modifications variées. Deux structures complémentaires avec des peptides ayant des affinités pour la polymérase encore plus élevées que les peptides originaux, sont présentés également. Les effets de perturbation de l'interaction entre la polymérase d'influenza et le domaine CTD de l'ARN polymérase II ont montré que cette interaction est essentielle. Des mutants de la polymérase virale qui ne peuvent pas reconnaître la modification Ser5, ont des activités fortement affaiblies dans un contexte cellulaire, tout en restant actifs *in vitro*, en présence de substrats nécessaires. Cela suggère que l'interaction entre les deux protéines est nécessaire pour fournir à la polymérase virale ces substrats, les ARN coiffés pour la production des amorces pour la transcription virale. Nous avons aussi produit, en collaboration avec le groupe de Nadia Naffakh (Institut Pasteur), des virus récombinants portant ces mutations, qui sont fortement atténués et génétiquement instables. Cela confirme l'essentialité de l'interaction des deux polymérases pour la réplication du virus et mets en valeurs les résultats obtenus.

Dans la deuxième partie de la thèse, en collaboration avec le groupe de Patrick Cramer (Max Planck Institute for Biophysical Chemistry), nous avons reconstitué un complexe 'cap-snatching' composé de la polymérase de la grippe, l'ARN polymérase II en phase d'élongation, avec un ARN coiffé émergeant sur sa surface. Les analyses de l'assemblage du complexe ont montré que la phosphorylation sur Ser5 du domaine CTD de l'ARN polymérase II est nécessaire pour la formation du complexe. La stoechiométrie du complexe *in vitro* a par ailleurs été étudiée et a montré que le nombre de sites de fixation sur le domaine CTD qui sont occupés par la polymérase de la grippe, est dans l'ordre de un ou deux. Des études fonctionnelles ont montré que la longueur

de l'ARN sortant de l'ARN polymérase II, qui peut être reconnu et clivé par la polymérase d'influenza, est d'environ 27-28 nucléotides. Ces études d'activité ont aussi aidé à étudier le mécanisme d'activation de la polymérase de la grippe dans le contexte du complexe 'cap-snatching', en confirmant les résultats qui avaient montré un effet allostérique sur les activités endonucléase et de synthèse des ARNm. Toutes ces études fonctionnelles ont démontré que l'assemblage d'un complexe actif de 'cap-snatching' est possible. Les études structurales de ce complexe par microscopie électronique ont montré des particules doubles qui peuvent être interprétées comme un complexe des deux polymérases, mais l'instabilité ou la flexibilité intrinsèque du complexe n'ont pas permis une analyse plus profonde par cryo-microscopie électronique.

La troisième partie du manuscrit décrit les analyses biochimiques de l'interaction entre la polymérase de la grippe avec deux facteurs d'élongation – DSIF et Tat-SF1. Des expériences de chromatographie d'exclusion de taille ont montré que les deux facteurs interagissent directement avec la polymérase de la grippe. Ces interactions ont été confirmées par des méthodes de pull-down et ont aussi montré l'indépendance des modifications post-transcriptionnelles pour l'association des protéines. Pour le facteur DSIF, un complexe ternaire avec l'ARN polymérase II et la polymérase virale a été assemblé et des études par microscopie électronique ont été menées, mais le complexe n'a pas été assez stable pour une caractérisation ultérieure. Cependant, le facteur Tat-SF1 ne forme pas un complexe ternaire avec les deux polymérase, suggérant qu'il peut être en compétition avec l'ARN polymérase II pour le site de fixation sur la polymérase d'influenza. Des tests d'activités n'ont pas montré un effet des deux facteurs sur les activités de la polymérase de la grippe dans le contexte du complexe 'cap-snatching', mais il est possible que les conditions spécifiques pour étudier cette activité n'ont pas permis l'interprétation définitive de ce manque d'effet observé.

### **3.1. Interaction between influenza polymerase and the C-terminal domain of RNA polymerase II**

#### **3.1.1. Peptide design**

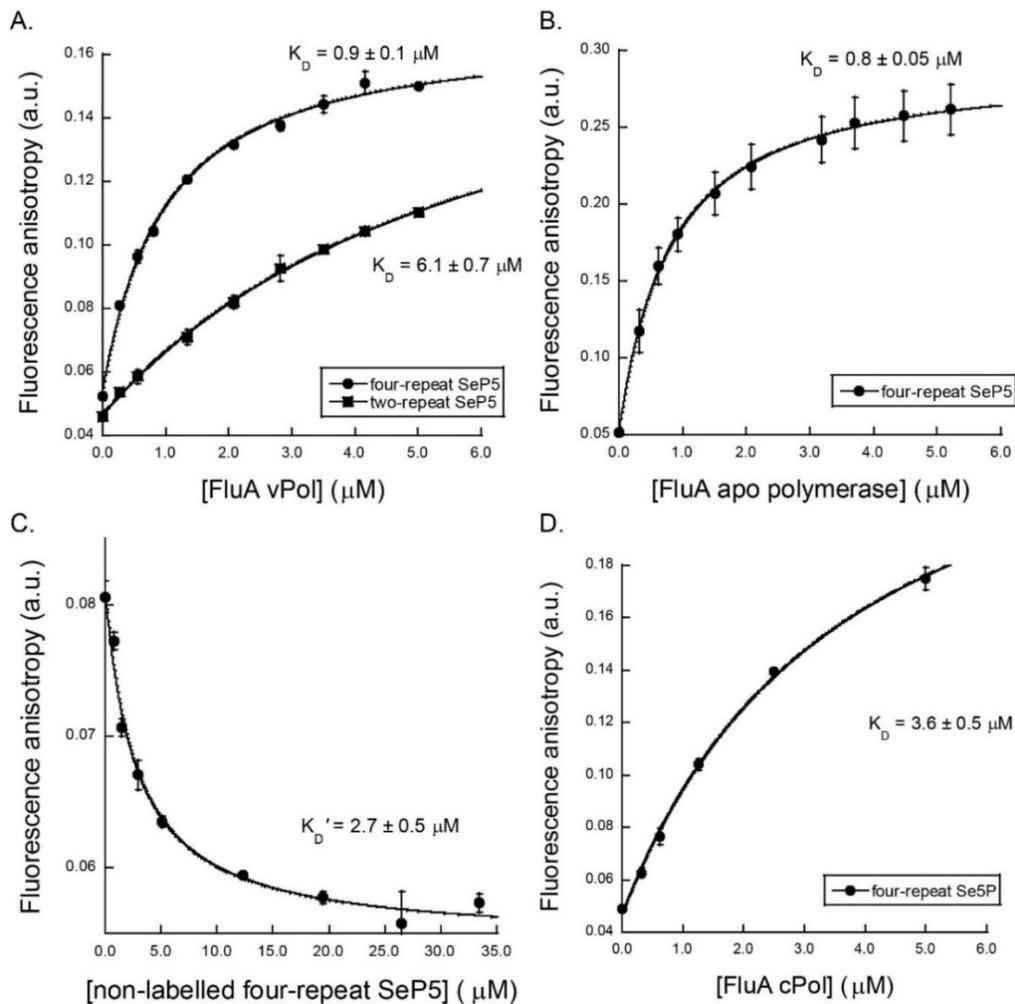
To address the interaction between influenza polymerase (FluPol) and the C-terminal domain (CTD) of RNA polymerase II (Pol II), I designed two different length of peptides, one containing two consensus repeats (14 amino acids) and one of four repeats (28 amino acids). The design was based on the co-crystal structures of the capping enzymes of two different species: *C. albicans* Cgt1 in complex with a four-repeat CTD peptide and the fission yeast Pce1, bound to a three-repeat peptide<sup>142,143</sup>. The consensus sequence of the CTD arbitrarily defines the repeats as Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub><sup>144</sup>, but most available CTD-bound structures show binding to more than one repeat<sup>142,143,145,146</sup>. This explains why the shortest peptide used in this study contains two repeats. Given the repetitive nature of the CTD sequence, it is expected that recruitment of partners could be enhanced due to increased affinity towards multiple connected repeats (avidity effect). Also, in both structures mentioned beforehand, only a part of the peptide was visible in the electron density (17 amino acids for Cgt1 and 8 for Pce1), so for structural studies I used a four-repeat CTD peptide.

#### **3.1.2. Binding assays of FluPol with CTD peptides**

##### **3.1.2.1. Binding of FluA polymerase to CTD peptides**

A fluorescence polarization-based binding assay was developed to characterize the interaction between FluPol and the CTD by using N-terminally FAM-labelled peptides. It was known that FluPol associates with Ser5-phosphorylated (SeP5) CTD *in vivo*<sup>93</sup>. To test this *in vitro*, I performed binding studies with SeP5 two- and four-repeat peptides. Since it is known that FluPol can adopt various conformations during the replication cycle<sup>25</sup>, in order to mimic a transcriptionally active form, binding assays were performed in the presence of the vRNA promoter. Influenza A polymerase (bat/Guatemala/060/2010/H17N10, FluA) interacts with Se5P-CTD peptides and shows a higher affinity towards the four-repeat peptide ( $K_D = 0.9 \mu\text{M}$ ) compared to a two-repeat peptide ( $K_D = 6.1 \mu\text{M}$ ), which can be hypothesized to be due to an avidity effect (increased affinity for multiple identical repeats) (Figure 7A). In absence of the promoter, the affinity to the four-repeat peptide was very similar, indicating that the CTD binding region is

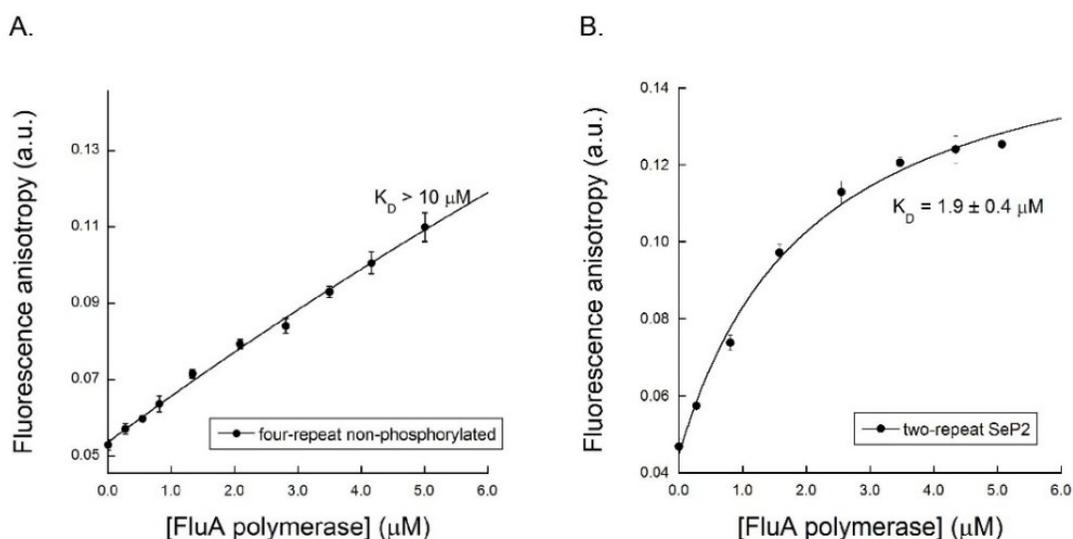
far from the promoter-binding site (Figure 7B). The specificity of the interaction was confirmed by displacement assay using non-labelled SeP5 peptide (Figure 7C). Interestingly, the FluA polymerase bound to cRNA promoter, had weaker affinity to the four-repeat peptide compared to the vPol or the apo-polymerase (Figure 7D).



**Figure 7. Binding assays of FluA polymerase and SeP5-CTD peptides** (A) Interaction of promoter-bound polymerase with two- and four-repeat SeP5 CTD peptides. (B) Binding to four-repeat SeP5 CTD peptide to FluA polymerase in absence of RNA promoter. (C) Displacement assay with non-labelled four-repeat SeP5 CTD peptide. (D) cRNA-bound FluA polymerase interaction with four-repeat SeP5 CTD. Error bars show standard deviation of three experiments,  $K_{DS}$  are shown  $\pm$  the error of the fit.

Non-phosphorylated four-repeat peptide did not show significant binding to FluA polymerase, with an estimated  $K_D$  of more than  $10 \mu\text{M}$  (Figure 8A). Due to limitations of the assay it was not

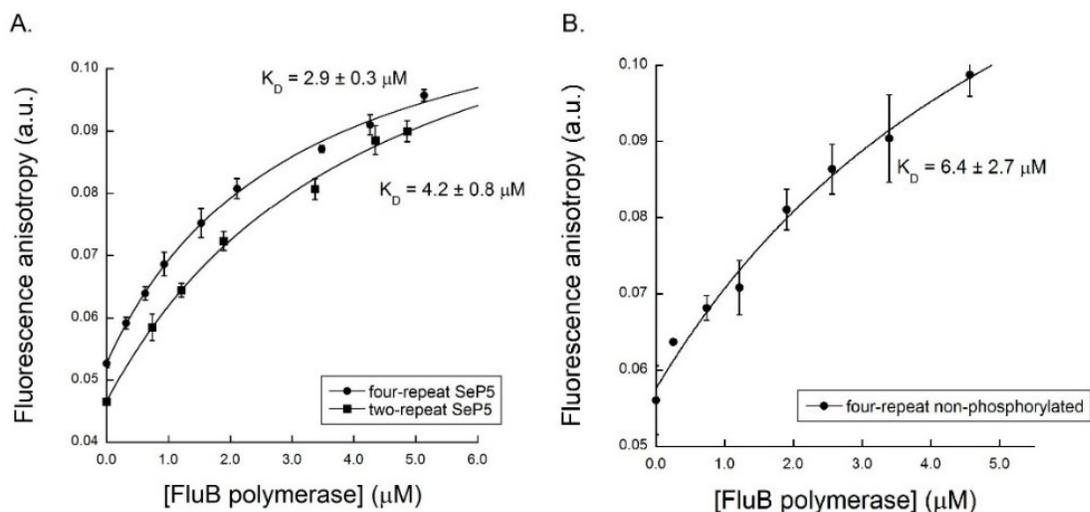
possible to reach saturation and to calculate an exact  $K_D$  value. According to the literature<sup>93,147</sup> FluPol does not associate with SeP2 CTD *in vivo*. Surprisingly, in this assay, two-repeat SeP2 CTD peptide showed binding with 2  $\mu\text{M}$  affinity (Figure 8B). However, this could not be rationalized based on the structural data observed later (section 3.1.3), and I could not confirm the binding by other means, which could possibly be explained by unspecific interaction with this peptide.



**Figure 8. Interaction between FluA polymerase and a four-repeat non-phosphorylated CTD peptide (A) and a two-repeat SeP2 peptide (B).** Error bars show standard deviation of three experiments,  $K_D$ s are shown  $\pm$  the error of the fit.

### 3.1.2.2. Binding of FluB polymerase to CTD peptides

Unlike FluA, FluB polymerase (B/Memphis/13/2003) interacted with SeP5 CTD peptides with similar affinities independently of the length of the peptides – I measured  $K_D$ s of 2.9 and 4.2  $\mu\text{M}$  for a four and two-repeats, respectively (Figure 9A). Non-phosphorylated peptide shows lower affinity, but the difference is not as significant as for FluA, suggesting that FluB discriminates less well the phosphorylation pattern compared to FluA polymerase (Figure 9B).



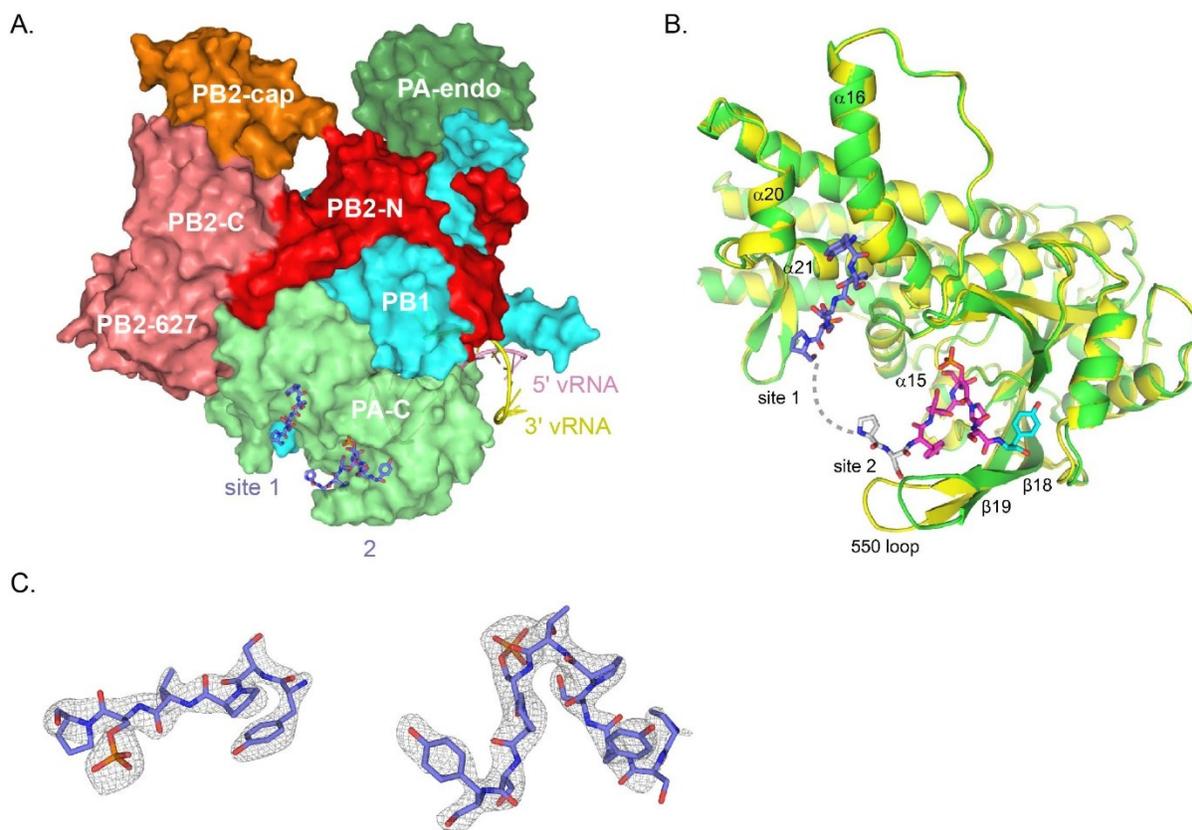
**Figure 9. Interaction between FluB polymerase and SeP5 two- and four-repeat CTD peptides (A) and non-phosphorylated four-repeat peptide (B).** Error bars show standard deviation of three experiments,  $K_D$ s are shown  $\pm$  the error of the fit.

The binding assays with CTD peptides showed that FluA and FluB polymerases can directly interact with the CTD of Pol II with a preference to its phosphorylated form. This first proof *in vitro* of a direct association between the two polymerases suggested this would be a promising target for structural studies.

### 3.1.3. Structural characterization of FluPol and CTD peptides

#### 3.1.3.1. Structure of FluA polymerase bound to four-repeat SeP5 CTD peptide

Influenza A polymerase, bound to the RNA promoter, was co-crystallized in complex with a four-repeat Ser5-phosphorylated CTD peptide. The structure was built at 2.5 Å resolution. 16 residues of the peptide could be unambiguously modelled in the structure, interacting at two separate sites on the polymerase surface, both located on the PA subunit (Figure 10A, C). The interaction regions are far from the polymerase active site, as well as from the other functional domains involved in polymerase activity (cap-binding and endonuclease domain). The binding of the peptide does not induce any long-range conformational changes and a main-chain-atom superposition to the structure of FluA polymerase without peptide gives an overall RMSD of 1.497 Å<sup>2</sup> (Figure 10B).

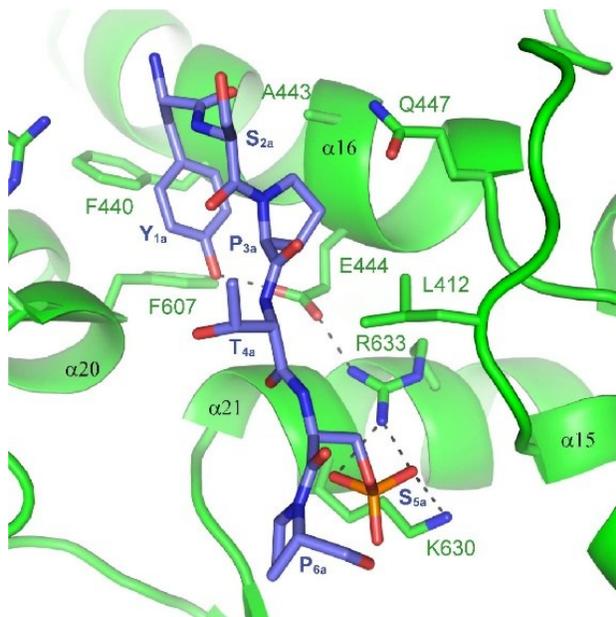


**Figure 10. Structure of FluA polymerase bound to four-repeat SeP5 CTD peptide.** (A) View of the whole polymerase in surface representation and the peptide in stick representation (in blue). (B) Structures of the PA C-terminal domain bound to the CTD peptide (green) (PDB: 5M3H) or in apo form (yellow) (PDB: 4WSA). Consecutive repeats are color coded from the N-terminus to C-terminus: blue, white, magenta and cyan. The possible linker between the two peptides is indicated with dashed line. (C)  $F_o - F_c$  omit map of the CTD peptides, contoured at  $3\sigma$ .

Residues belonging to different repeats are denominated with a corresponding suffix (a, b, c, d). Six residues from one CTD repeat ( $Y_{1a}S_{2a}P_{3a}T_{4a}S_{5Pa}P_{6a}$ ) bind to a conserved groove, formed by helices  $\alpha 16$ ,  $\alpha 20$ ,  $\alpha 21$  and a loop between helices  $\alpha 15$  and  $\alpha 16$ . The total buried surface of this interaction region of PA (denoted site 1) is  $672 \text{ \AA}^2$ . Residues from three repeats ( $P_{6b}S_{7b} - Y_{1c}S_{2c}P_{3c}T_{4c}S_{5pc}P_{6c}S_{7c} - Y_{1d}$ ) are accommodated at the other site (site 2), with an interaction surface area of  $1,168 \text{ \AA}^2$ . The peptide interaction site includes the  $\beta 18 - \beta 19$  ribbon and loop 550<sup>22</sup>. No density for the region connecting the two tightly bound peptides could be observed, but it is plausible that they belong to the same chain. The binding data showing higher affinity for a four-repeat compared to a two-repeat peptide (Figure 7A), suggests a higher likelihood that longer

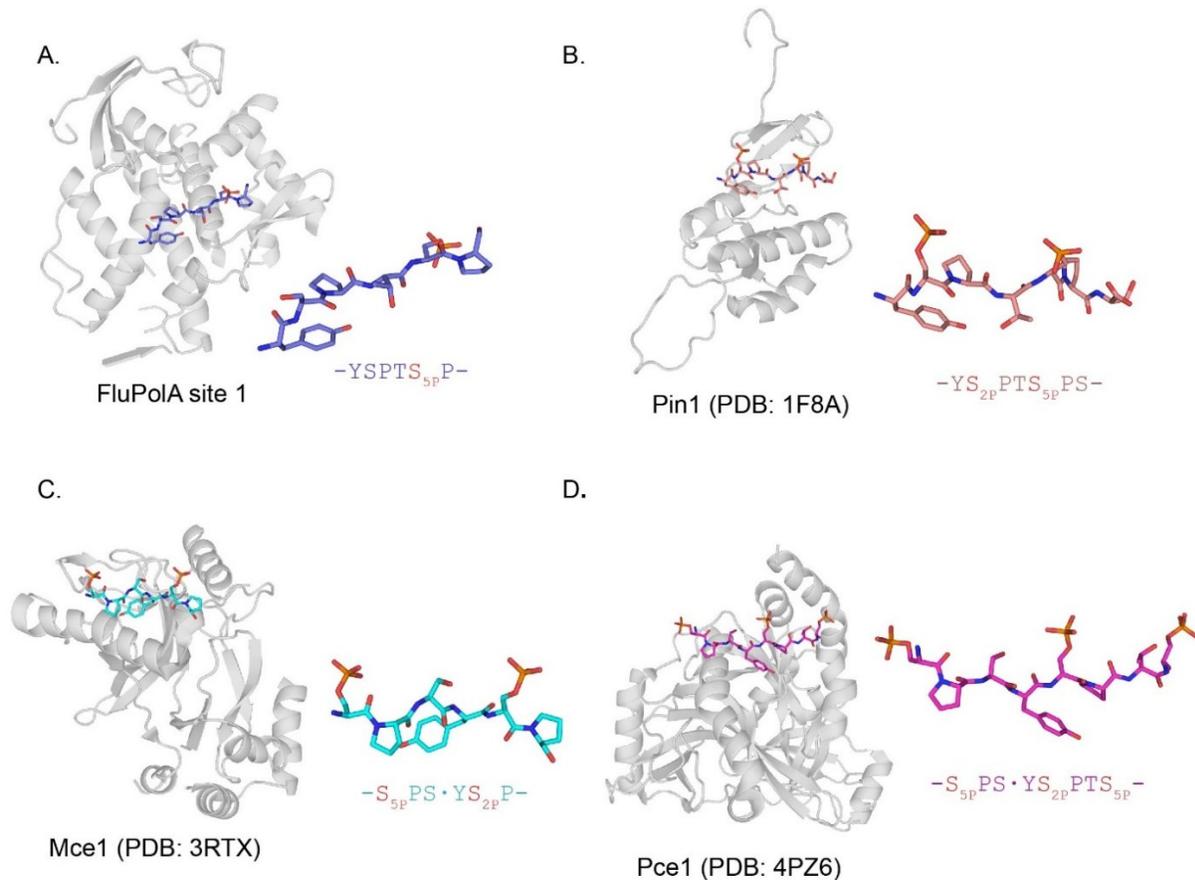
peptide, rather than two non-connected peptides would be bound. The missing residues in the structure (-S<sub>7a</sub>- Y<sub>1b</sub>-S<sub>2b</sub>-P<sub>3b</sub>-T<sub>4b</sub>-S<sub>5pb</sub>-) are probably not making any contacts with the protein and are looped out from the polymerase surface. This ‘looping out’ has already been observed in other CTD-bound structures<sup>142</sup>. It is speculated that this can occur when CTD-interacting proteins are ‘screening’ the CTD sequentially for the right phosphorylation state. It also might have a functional relevance, as the looped region can be associated with other CTD-binding proteins and remodeling factors, as previously suggested<sup>61</sup>.

In site 1, the bound peptide binds in an extended  $\beta$ -like conformation, with every other residue pointing towards the protein surface (Figure 11). This conformation is common for other CTD interactors, but the positioning of the individual residues and the contacts with the proteins are highly specific (Figure 12)<sup>148,145,143,149</sup>.



**Figure 11. Structure of the SeP5 CTD peptide bound to Flua site 1.** PA-C domain of Flua polymerase is depicted in green, peptide is in blue and shown as sticks.

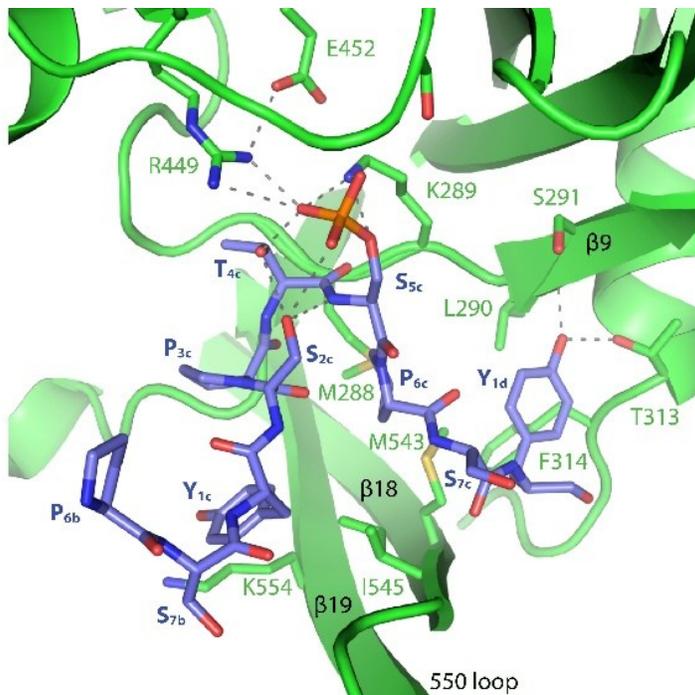
Residues Y<sub>1a</sub>, P<sub>3a</sub>, S<sub>5Pa</sub> intricately interact with the polymerase, while S<sub>2a</sub>, T<sub>4a</sub> and P<sub>6a</sub> point away from the polymerase and do not make any substantial contribution to the recognition. Y<sub>1a</sub> is bound to a hydrophobic pocket of residues F440 and F607 and its hydroxyl group makes a hydrogen bond with E444. The residues L412 and A443 are interacting with P<sub>3a</sub>, which is in *trans* configuration. The phosphorylated S<sub>5Pa</sub> is anchored by multiple hydrogen bonds to a positively charged pocket formed by K630 and R633 (Figure 11).



**Figure 12. Structures of various host cell proteins bound to CTD peptides.** Proteins are depicted with cartoon representation and colored in grey; peptides are in sticks with different colors. (A) FluA polymerase site 1; (B) Peptidyl-prolyl isomerase Pin1 (PDB: 1F8A); (C) Mammalian capping enzyme Mce1 (PDB: 3RTX); (D) *S. pombe* capping enzyme Pce1 (PDB: 4PZ6).

The repeats binding to site 2 of FluA polymerase form a  $\beta$  turn, flanked by extended regions, which clamp the 550 loop (Figure 13). The  $\beta$ 18-  $\beta$ 19 ribbon and 550 loop are shifted compared to the FluA structure without peptide (Figure 10B), with the tip of the loop moving to a maximum distance of 6 Å from the position observed in the apo-structure. The  $\beta$  turn is formed by residues  $S_{2c}P_{3c}T_{4c}S_{5p}$  and is stabilized by two hydrogen bonds (between  $\gamma$ O of  $S_{2c}$  and  $T_{4c}$  amide; and between  $S_{2c}$  carbonyl and  $S_{5p}$  amide). All prolines of the peptide are in *trans* conformation. Unlike the SeP5 peptide bound to the fungal capping enzyme (Figure 14)<sup>142</sup>, in which the  $\beta$  turn loops out the phosphorylated serine so that it does not interact with the protein, in our structure the SeP5 is positioned to form a tight interaction with a positively charged pocket formed by K289 and R449. The two Tyr ( $Y_{1c}$  and  $Y_{1d}$ ), as well as  $P_{6c}$  are bound in hydrophobic pockets, formed by I545

and the backbone of K554 (for Y<sub>1c</sub>); M543 and M288 (P<sub>6c</sub>); M543, L290 and F314 (Y<sub>1d</sub>). Hydrogen bonds with T313 and the main chain amide of S291 are also stabilizing Y<sub>1d</sub>. An interesting point is that residue I545 (corresponding to 550 in H1N1) is known to play a role in the viral pathogenicity and RNA-polymerase II degradation induced by influenza infection (see Discussion section 4.1) <sup>150</sup>. Another feature of this conformation is that S<sub>2c</sub> is pointing towards the S<sub>5Pc</sub>, which would exclude a double Ser2-Ser5 phosphorylation at this position.

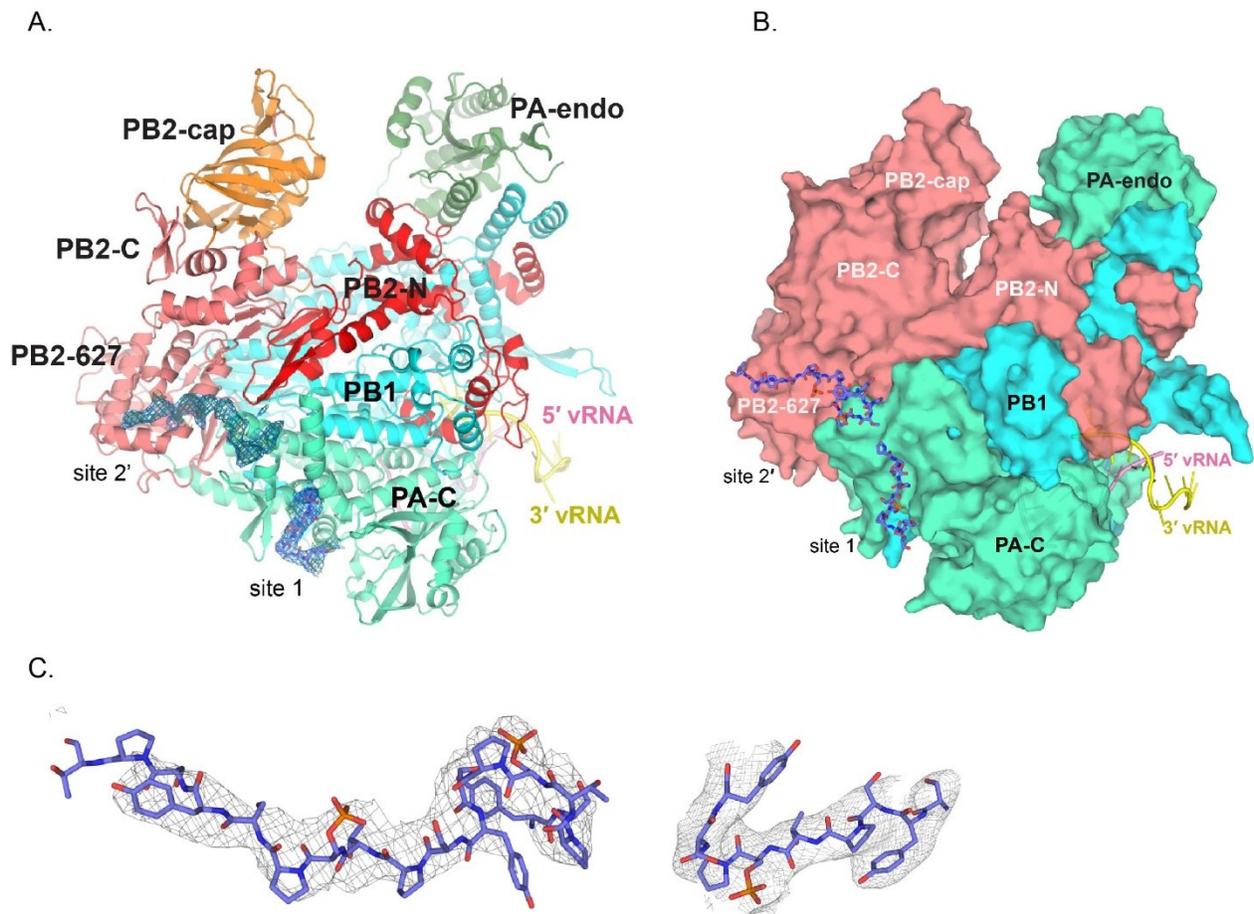


**Figure 13. Interactions between SeP5 CTD peptide and Flua polymerase site 2.** The PA-C domain is in green cartoon, peptide - in blue sticks.



### 3.1.3.2. Structures of SeP5 CTD-bound FluB polymerase

I wanted to understand how FluB binds to the CTD, since the conservation of the residues involved in CTD binding in FluA and the binding data suggested that FluB has only one binding site for the peptide. The structure of FluB polymerase bound to a four-repeat SeP5 CTD peptide was determined, confirming that the peptide binds to the conserved site 1. Interestingly, the structure showed some extra density extending to PB2-627 domain (Figure 16A). Initially, the resolution and quality of the map did not allow to allocate a peptide sequence to this density. This was solved by a different crystal form (produced by Petra Drncova), which diffracted to higher-resolution (2.7 Å) and which allowed to unambiguously assign the peptide chain binding PA-PB2. Figure 16B shows the FluB polymerase structure with both peptide binding sites observed in the two structures.

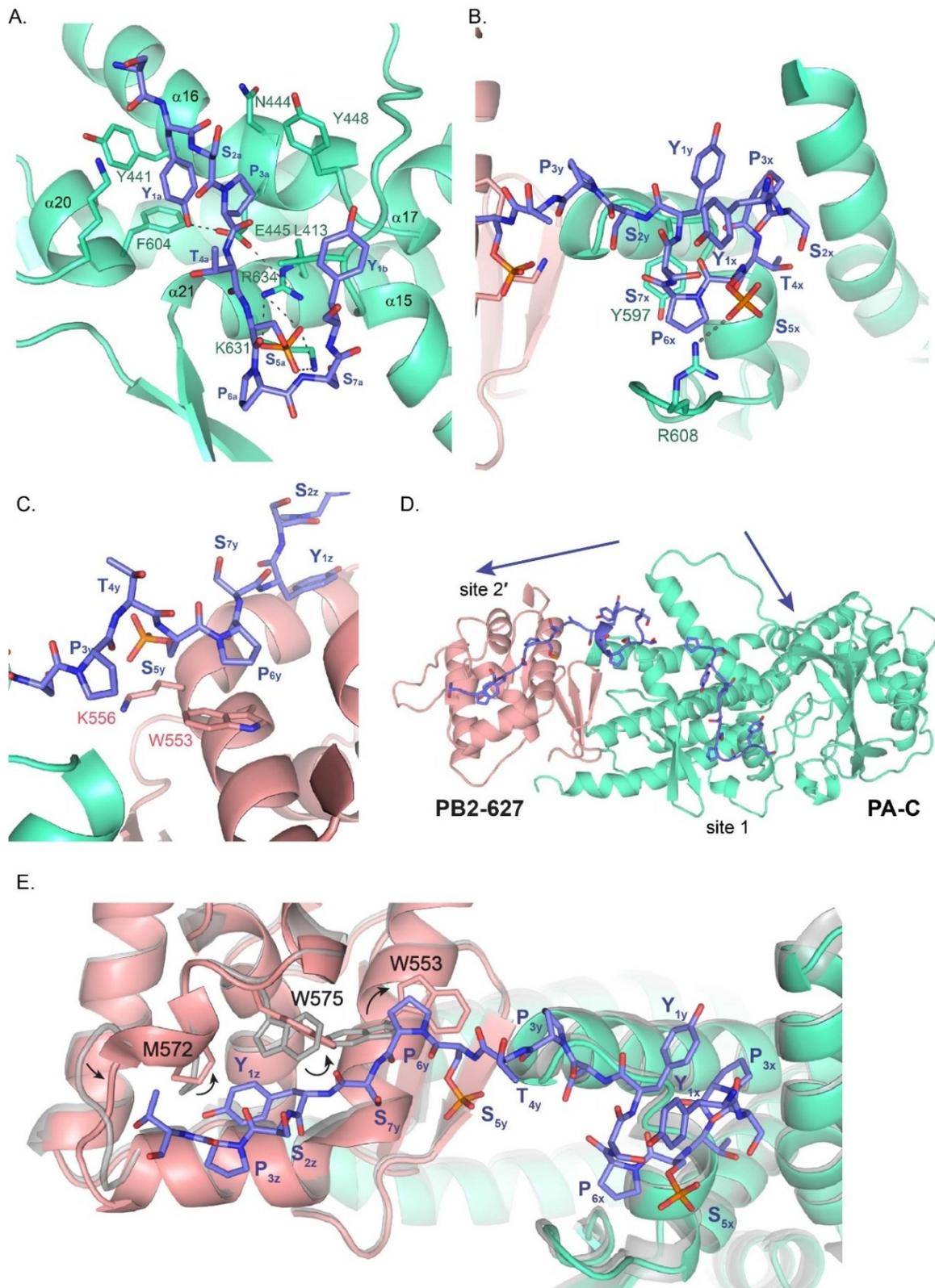


**Figure 16. Structures of different crystal forms of FluB polymerase bound to four-repeat SeP5 CTD peptide.** (A) The first crystal form with clear density for site 1 with the peptide modelled in blue sticks; extra density observed

on the surface is shown by omit difference map ( $F_o - F_c$ ,  $3.0\sigma$ , blue mesh). (B) The peptide-binding sites observed in the two structures shown in blue stick representation. In the second crystal form, site 1 was not occupied. (C)  $F_o - F_c$  omit map of the CTD peptides, contoured at  $3\sigma$ , the model from the new structure was fit into the density from the first crystal form; left – site 2', right – site 1.

The peptide at site 1 adopts very similar extended conformation as in FluA polymerase-CTD structure (Figure 17A, compared to Figure 11). Y441 and F604 are accommodating  $Y_{1a}$ , which is also hydrogen bonding to E445. The conserved positive residues interacting with the phosphate of  $S_{5a}$  are K631 and R634. In the FluB structure, extra residues are seen at this site, including the tyrosine from the next repeat.

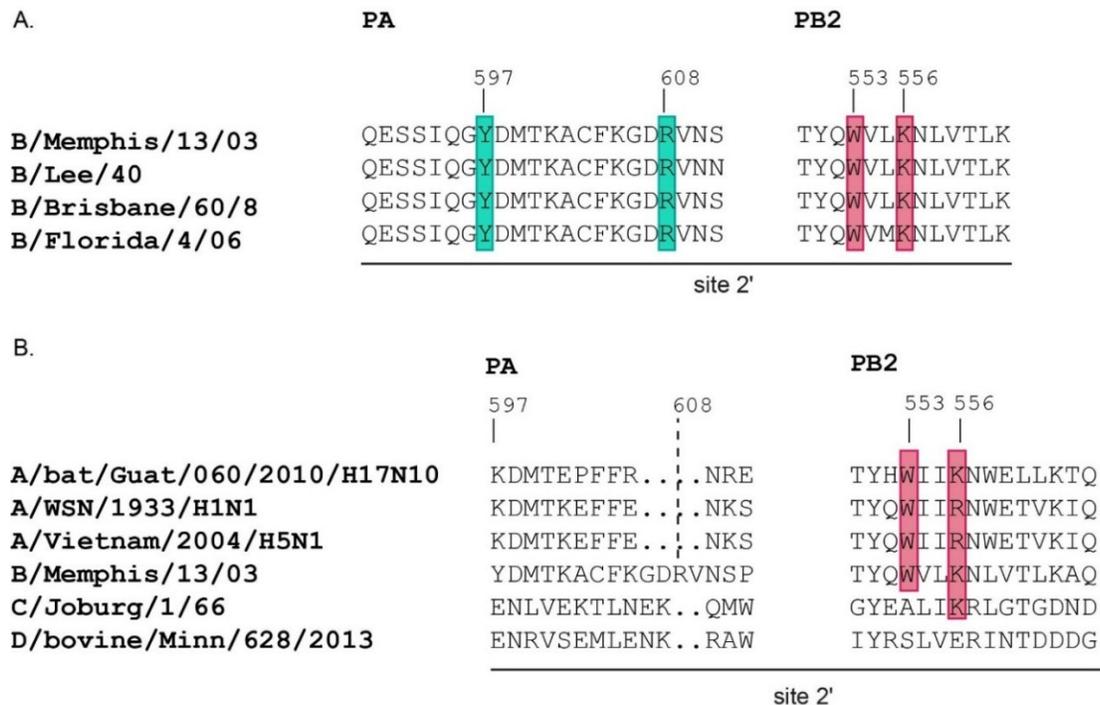
The newly observed binding site (denoted site 2') includes part of PA and the 627 domain of PB2. Unexpectedly, the direction of the chain is opposite to that of the peptide in site 1, suggesting that they are either not connected or a big part of the peptide is looped out. It can be estimated that at least six repeats are needed to bind and link the two sites. This would explain, on one hand, why it is difficult to observe clear density at both sites, since we are using only a four-repeat CTD peptide and, on the other hand, why the binding data using two- or four-repeat peptides showed very similar affinities. The peptide at site 2' makes a helical turn, stabilized by intramolecular interactions (Figure 17B, D). R608 from PA hydrogen bonds to the phosphate at  $S_{5x}$ , ensuring the specificity towards the phosphoserine in this repeat.  $S_{2x}$  points towards  $S_{5x}$ , excluding the possibility of double phosphorylation at these positions. The chain extends to PB2, with few specific interactions, e.g. W553 stacking on  $P_{6y}$  and supposedly K556 in hydrogen bond distance to  $S_{5y}$ , but no density for the K556 side chain is observed (Figure 17C). Figure 17D shows the whole CTD peptide chain with arrows indicating the directions (from N- to C-terminus). Several conformational changes contribute to the accommodation of the peptide by an induced fit mechanism: W553 and W575 side-chains are reoriented compared to the apo-structure; the loop, containing residue M572 shifts, adjusting the latter to form van der Waals interactions with  $Y_{1z}$  (Figure 17E). The total buried area upon CTD binding is  $1207 \text{ \AA}^2$ .



**Figure 17. Details of the interaction between SeP5 CTD peptide and FluB polymerase.** (A) Interaction at site 1, PA-C domain is in green cyan, peptide in blue sticks. (B) Site 2', interaction with PA (residues are labelled with the

corresponding colors, green cyan for PA and salmon for PB2). (C) Site 2', interaction with PB2; the possible position of the K556 side chain is indicated, but the latter is not seen in the electron density. (D) The two sections of peptides bound to PA and PB2 with arrows indicating the directions of the polypeptide chain of the CTD (from N- to C-terminus). (E) Conformational changes induced by the CTD binding, color code for the CTD-bound structure is the same as for the previous panels, the apo-form is in grey (PDB: 4WSA). Arrows indicate the movement of the respective residues or regions.

Sequence alignment of the site 2' between different influenza B strains shows that the interacting residues from both PA and PB2 subunits are conserved (Figure 18A). Interestingly, the respective residues on PB2 are also conserved in A strains (Figure 18B), but not those on PA (e.g. R608 is part of a loop, which is only conserved in FluB). It is therefore possible that the interaction site on PB2 is conserved in influenza A, although we did not observe it in our structure with A/H17N10 polymerase.



**Figure 18. Sequence alignment of the regions of CTD-interacting site 2' of FluB.** (A) Alignment of IBV strains: B/Memphis/13/03 and B/Brisbane/60/8 are part of the Victoria lineage, B/Florida/4/06 is from the Yamagata lineage and B/Lee/40 is before the divergence of the two lineages. (B) Alignment of the PA and PB2 regions from site 2' with other influenza strains (A, C and D). The numbering shown relates to B/Memphis/13/03.

The structures of FluA and FluB polymerases in complex with CTD peptides show that the two proteins share a common binding site (site 1) for recognizing a Ser5-phosphorylated CTD peptide. Interestingly, the polymerases of each strain have an additional interaction site, which is only conserved within the same strain (A or B, respectively). Moreover, despite apparent similarities in the conformation of the CTD peptides in comparison with structures of cellular CTD-binding factors, the specificity of the interactions suggests this region could be an interesting target for anti-viral drug design.

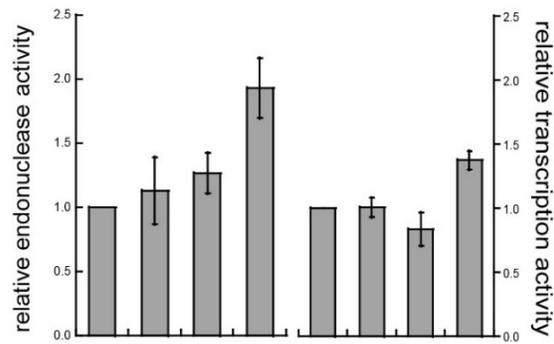
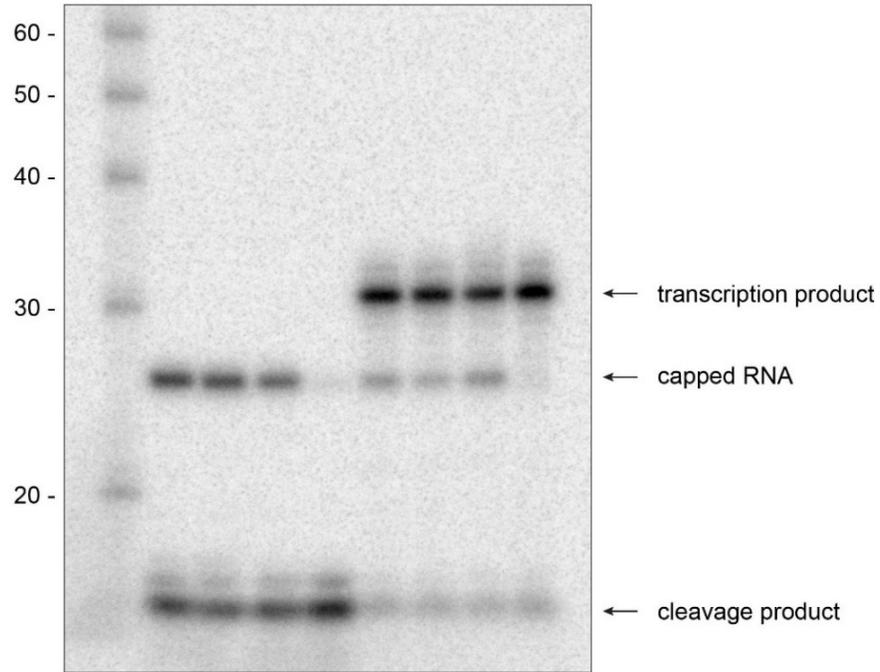
#### 3.1.4. Effect of CTD binding on FluPol activity

The structural data suggest that CTD binding would not have a direct effect on the polymerase enzymatic activities, since the interaction sites are neither close to the promoter-binding site nor to the cap-binding, endonuclease or polymerase active sites. In order to investigate possible allosteric effects on polymerase activities, I performed endonuclease and transcription assays in presence of SeP5 CTD peptides of two-, four- or six-repeats. An increase of the endonuclease and transcription activities of both FluA and FluB polymerases was observed in the presence of six-repeat SeP5 peptide, but not with any shorter peptide (two or four-repeats) (Figure 19). The secondary binding site observed in the latest FluB polymerase structure in complex with the CTD peptide showed that a longer peptide could extend to the PB2 627 domain, and have an allosteric effect on the activity, perhaps by stabilizing a transcriptionally active conformation and positioning the flexible polymerase domains in a conformation compatible with cap-snatching (see Discussion section 4.1). Surprisingly, FluA (H17N10), for which we had only observed interactions between the CTD peptide and the PA-C domain, also showed increased endonuclease and transcription activities in the presence of the six-repeat peptide. Together with the observed decrease in affinity to the CTD in a cRNA-bound conformation (Figure 7D), this data suggests the presence of an additional CTD binding site on FluA, which, however, has not been observed yet structurally.

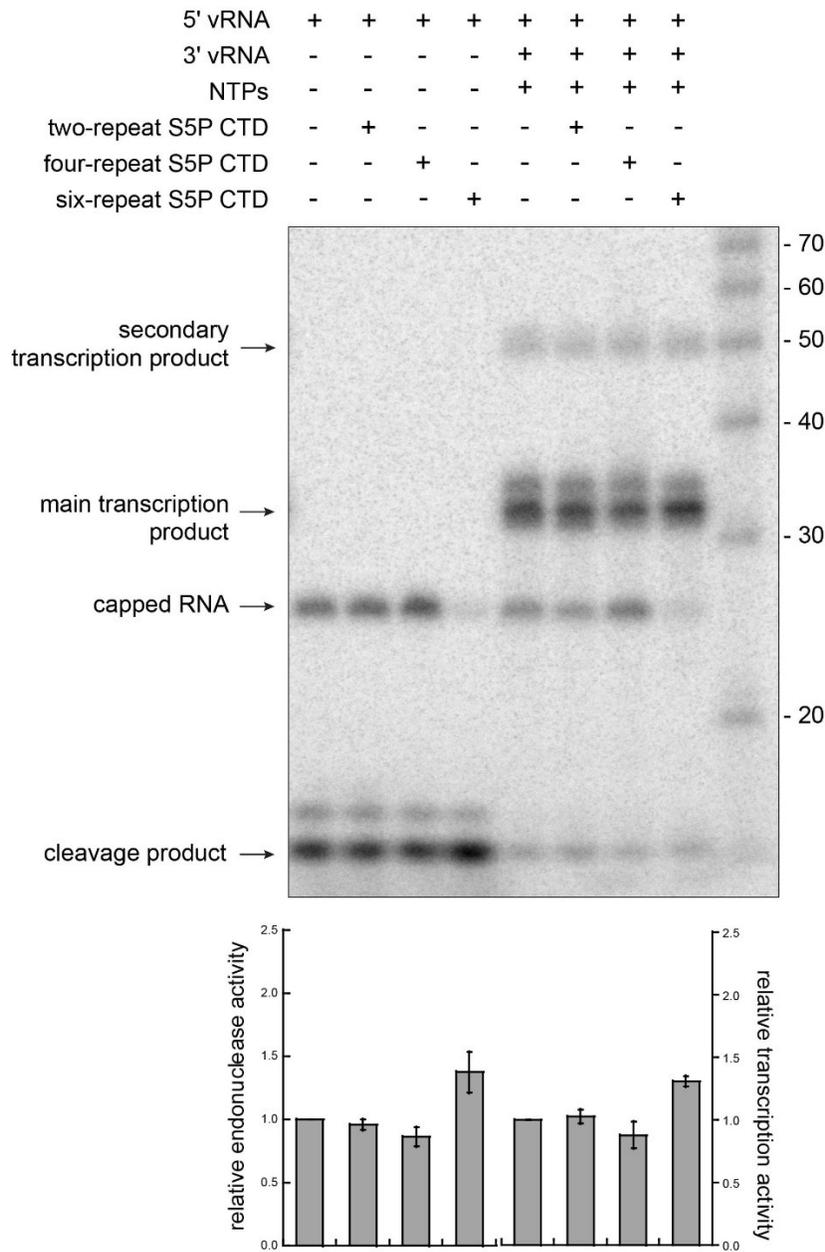
Previous data has shown that a four-repeat SeP5 CTD peptide has an effect on the activity of FluC and A/NT/60/68 (H3N2) polymerases<sup>151</sup>. In the case of FluC, a structure of the FluC polymerase bound to CTD shows that the peptide binding could stabilize the transcriptionally active state (but without binding to PB2), similar to the results presented here for FluB. I did not observe the same effect of a four-repeat peptide on FluA (H17N10) and FluB polymerase activities. However, in the FluC polymerase, the CTD binding site lies at the interface between the subunits (PA and PB1), which would explain why a shorter peptide can have an effect on its activity by keeping the polymerase in transcriptionally active conformation.

A.

5' vRNA	+	+	+	+	+	+	+	+
3' vRNA	-	-	-	-	+	+	+	+
NTPs	-	-	-	-	+	+	+	+
two-repeat S5P CTD	-	+	-	-	-	+	-	-
four-repeat S5P CTD	-	-	+	-	-	-	+	-
six-repeat S5P CTD	-	-	-	+	-	-	-	+



B.



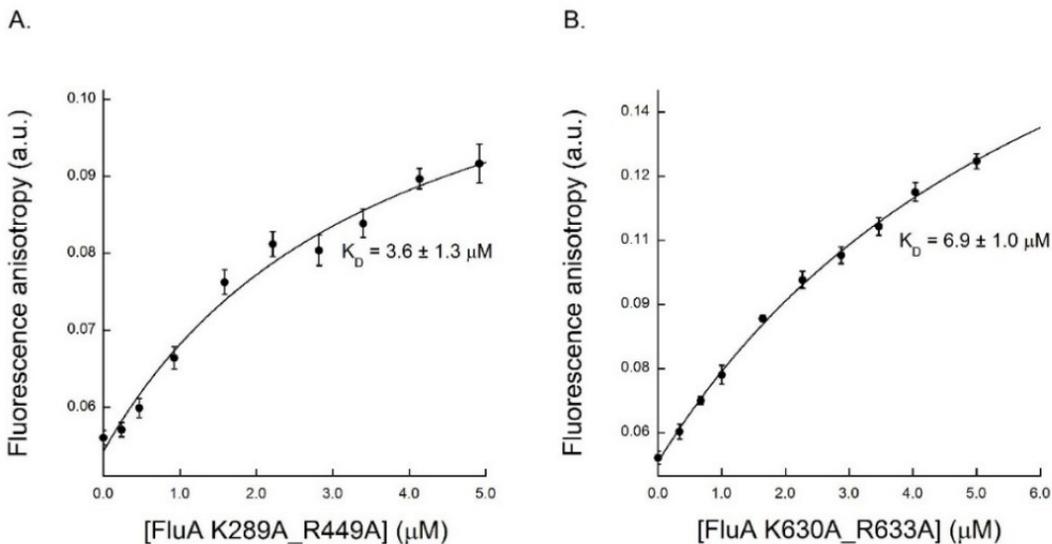
**Figure 19. Effect of SeP5 CTD peptides on endonuclease and transcription activity of FluA (A) and FluB (B) polymerases.** Products of the reactions are run on denaturing acrylamide gels. Quantification of three independent experiments is shown below each corresponding lane, with error bars indicating the standard deviation. The products of the reactions are normalized to the total RNA amount for each reaction. The activity in absence of peptides is used for normalization of the activities.

### 3.1.5. Functional validation of the FluPol – CTD interaction

#### 3.1.5.1. FluA polymerase – CTD association

##### 3.1.5.1.1. *In vitro* characterization

To study the role of the interaction between FluPol and CTD, I mutated the conserved basic residues that are forming the specific interaction with the phosphoserines at site 1 and site 2 in FluA polymerase to alanines (K289, R449 (R454 in human/avian A strains), K630 (K635) and R633 (R638)). A quadruple mutant with all four residues mutated could not be expressed recombinantly. In agreement with this observation, studies with fluorescently labelled polymerase in cells showed that this mutant is not properly localized (data not shown). For this reason I purified recombinant proteins bearing double mutations at each site (K289A/R449A and K630A/R633A). I measured the binding to four-repeat CTD peptide and observed a 4- to 7.5 fold decrease in affinity for the K289A/R449A and K630A/R633A mutants compared to wild type FluA, respectively (Figure 20, compared to Figure 7A).

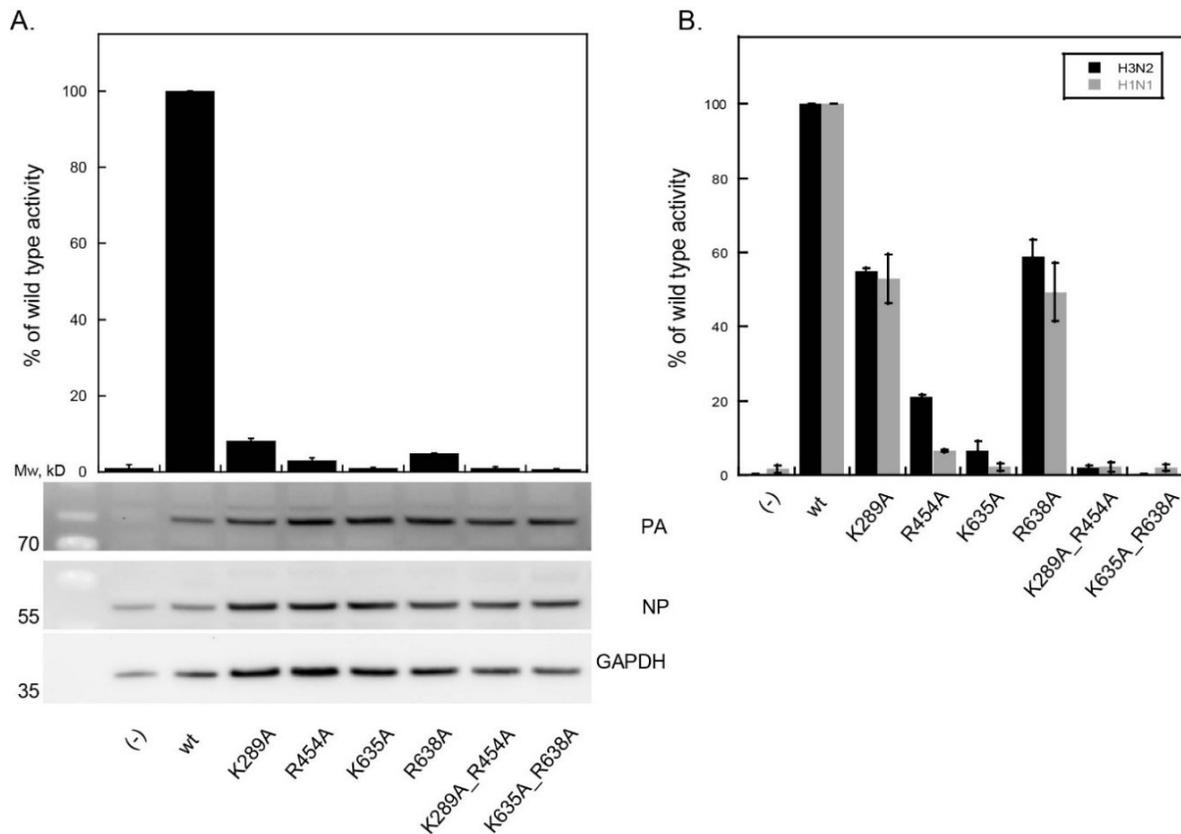


**Figure 20. Binding of SeP5 CTD peptides to recombinant FluA polymerase mutants.** (A) K289A/R449A and (B) K630A/R633A interaction with a four-repeat SeP5 CTD peptide. Error bars show standard deviation of three experiments,  $K_D$ s are shown  $\pm$  the error of the fit. The binding curve for the wild type is shown on Figure 7A with a  $K_D$  of 0.9  $\mu\text{M}$ .

The decrease of affinity of these mutants is not very drastic, as expected from the structural data showing multiple specific interactions of the peptide with the polymerase binding pockets. Nevertheless, the measured difference in affinities to a short peptide probably does not reflect the real effect when multiple repeats are present, as it occurs in the context of a full-length CTD. It is possible that in this case the cumulative effect of the weakened interaction is much more significant. Moreover, these mutations specifically disrupt the recognition of the functionally important for FluPol SeP5 phosphorylation state of the CTD.

#### **3.1.5.1.2. Effect of disrupting CTD binding in the cellular context**

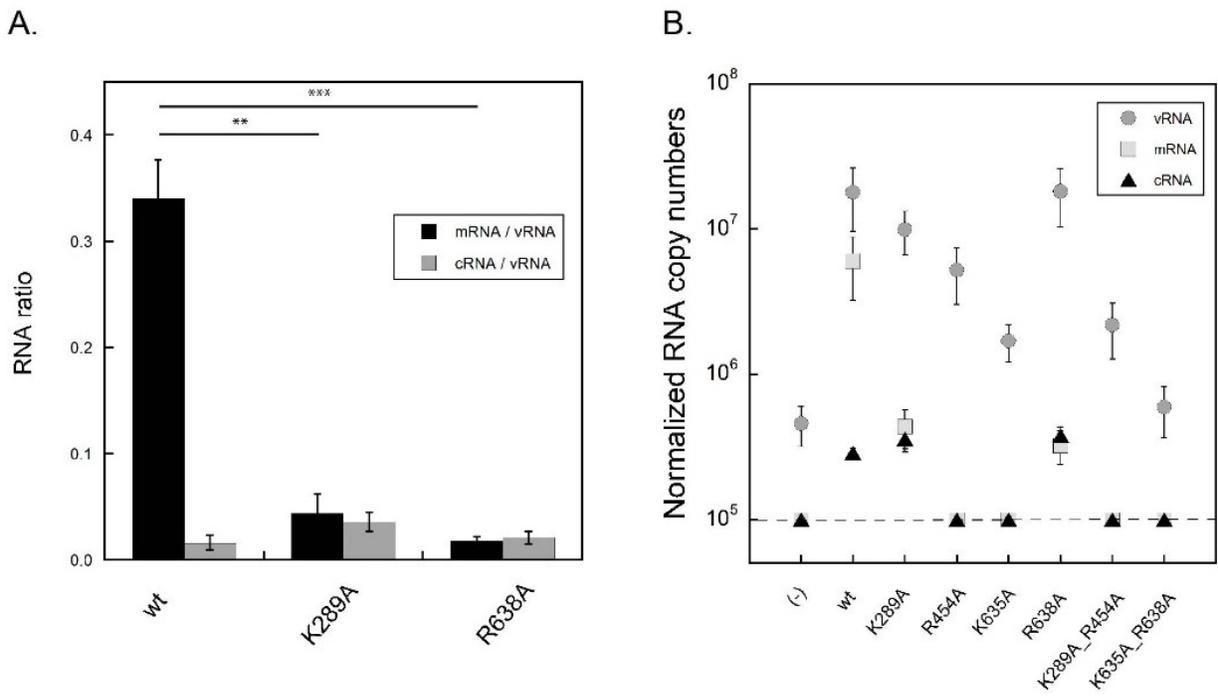
The above mentioned basic residues of FluPol involved in the CTD interaction were mutated in influenza A/WSN/33(H1N1) strain (corresponding residues: K289, R454, K635 and R638), either separately or as a combination of double residues at each site (as described above) to study the effect of impaired FluPol – CTD interaction. The activity of the polymerase was measured using a mini-genome assay. In this assay the polymerase and nucleoprotein are expressed together with a luciferase reporter encoded in negative polarity. The luciferase coding region is flanked by the conserved influenza promoter termini and is under the control of a Pol I promoter. This construct produces a luciferase-containing vRNA-like transcript, which can be expressed only by a functional FluPol. Another viral protein, NS2/NEP, was also added, since it is known to have an effect on transcription and replication. The co-expression of NS2 in a mini-genome assay has been reported to regulate the ratios of the different forms of RNA (vRNA, cRNA and mRNA) and to better mimic real infection conditions <sup>45</sup>. This setup was also used later for quantification of the RNAs produced by the viral polymerase. Each single mutation had a strong effect on the activity of the polymerase in this assay and a combination of mutants almost completely abolished any activity (Figure 21A). A mini-genome assay in the absence of NS2 was also performed in a more minimal system, where only direct effects on the polymerase functions could be followed (Figure 21B). Double mutant combinations similarly abolished the activity, while two of the single mutants (K289A and R638A) retained about 50 % of the wild type activity. This experiment was also performed with another FluA strain (A/Victoria/3/75/H3N2) and gave very similar result, as predicted from the high sequence similarities of the influenza A strains and confirming the conservation of the function of these residues among the influenza A species.



**Figure 21. Mini-genome assay of FluA CTD-binding deficient mutants (A) in presence or (B) absence of NS2/NEP protein.** (A) The activity of CTD-binding mutants in presence of NS2 is strongly attenuated. Western blots confirming the expression of polymerase subunits and NP are not affected by the mutations, are shown below. (B) Experiments without NS2 were performed with H1N1 (black) and H3N2 strains (grey) for comparison. Error bars show standard deviation of three independent experiments performed in triplicate.

The observed expression of luciferase is a cumulative effect of FluPol transcription and replication activities in the cell. To understand the effect of the mutations, in collaboration with the group of Nadia Naffakh in Institut Pasteur we determined the levels of the different RNAs (mRNA, vRNA and cRNA) produced by the polymerase and mutant variants. The RNAs were quantified by quantitative reverse-transcription PCR (RT-PCR)<sup>152</sup>, using segment 6 (NA) as a reporter and specific primers to detect the different RNA species. We found that there was a difference in the levels of replication products between the mutants and the wild type (vRNA), but comparison of the ratios of transcription and replication products (mRNA/vRNA compared to cRNA/vRNA) showed more significant effect on transcription than on replication (Figure 22).

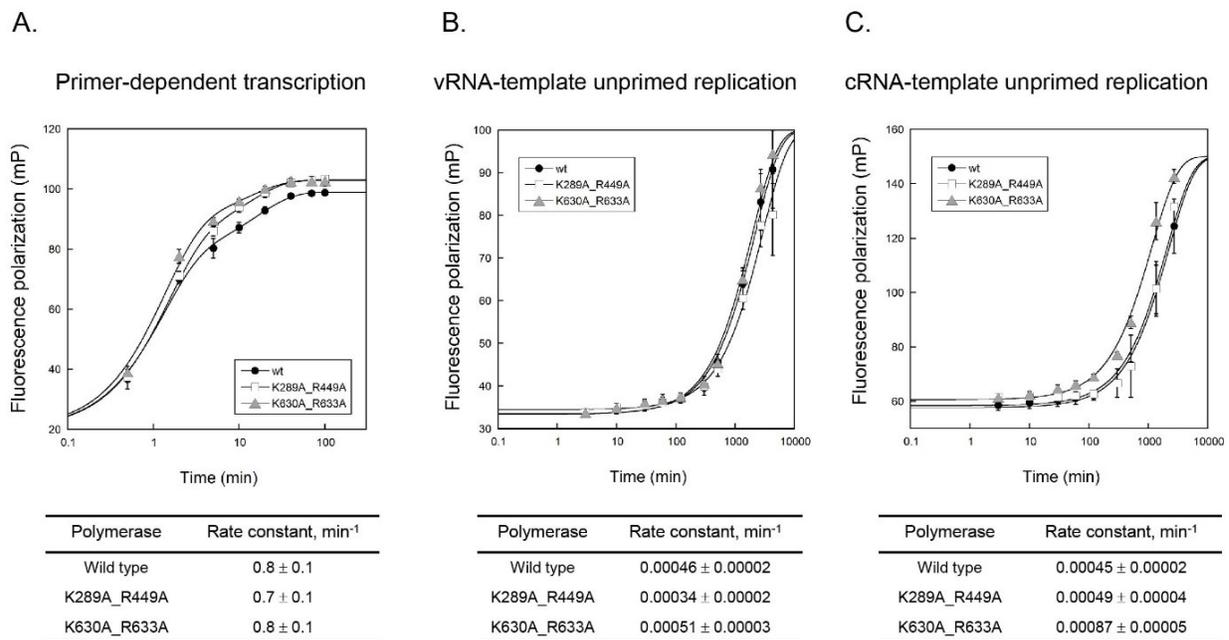
The effect on replication of the CTD binding mutations probably cannot be rationalized with the lack of association with Pol II, since the interaction with the cellular polymerase is expected to play a role only on viral transcription. It is possible that the disruption of transcription leads to replicative defects, but in a mini-genome system, where the two processes are not coupled (unlike in infection, where replication depends on transcription), this does not have a simple explanation. Nevertheless, a functional link between the two processes or some common mode of regulation is plausible (see Discussion section 4.1).



**Figure 22. Quantification of different RNA species in the mini-genome assay (in presence of NS2/NEP) by RT-qPCR.** (A) mRNA/vRNA and cRNA/vRNA ratios of two single mutants; R454A, K635A and double mutants had undetectable levels of cRNAs and are not shown on the graph. Error bars represent standard deviation of three independent experiments (\*\*P < 0.01, \*\*\*P < 0.001, using multivariate linear model with an interaction term between mutation and RNA type). (B) Normalized RNA copy numbers of wild type and mutant polymerases. Dashed line indicates the detection limit (lower than 10<sup>5</sup>). *Experiments were performed by Guillaume Fournier and Nadia Naffakh and they provided the data for the figures.*

The mini-genome assays suggested that mutant polymerase deficient in CTD binding is not functionally active in cells. To exclude the possibility that the intrinsic enzymatic activities of these mutants are impaired, I measured the activity in an *in vitro* assay using recombinant mutant polymerases (Figure 23) <sup>27</sup>. A cap-dependent transcription assay (Figure 23A) and unprimed

replication assays using a vRNA (Figure 23B) or cRNA promoter (Figure 23C) were performed in order to mimic the intrinsic FluPol enzymatic activities *in vitro*. No difference was observed in the activities of the wild type and the mutant polymerases. This confirmed that the double mutants (which had almost abolished activity in a cellular context) are as active as the wild type *in vitro*, when provided with all necessary substrates. The loss of interaction with the CTD is thus likely the reason for the loss of activity in the cell, due to impaired recruitment to transcribing Pol II and hence unavailability of capped transcripts for cap-snatching by influenza. Taken together these results confirm the importance of the FluPol – CTD interaction observed in the structural data for the function of FluPol *in vivo*.



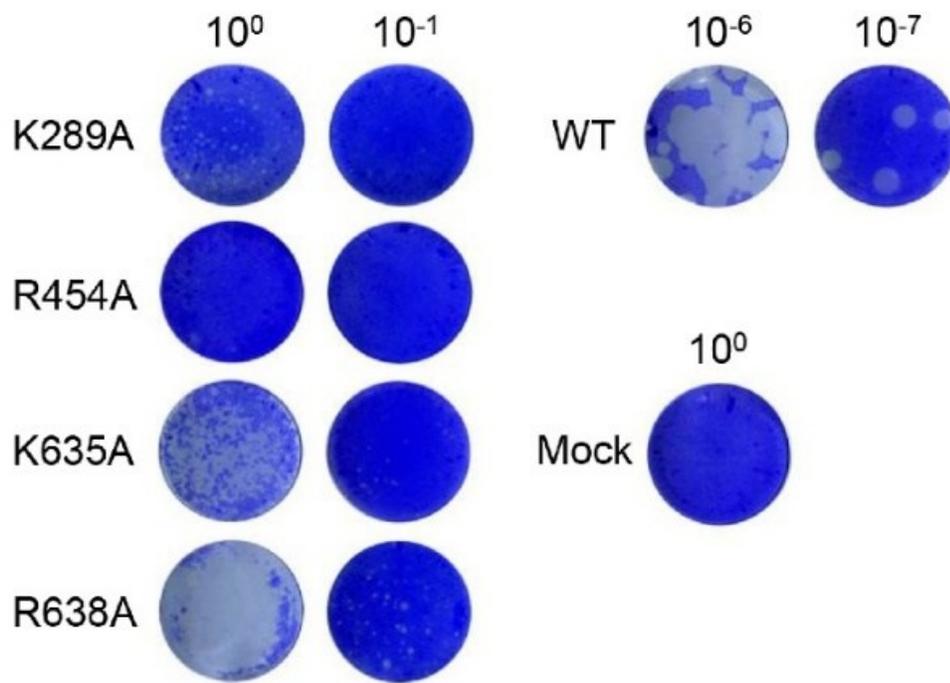
**Figure 23. Comparison between cap-primed transcription (A) and v- (B) and cRNA-dependent (C) replication activities of wild type and mutant polymerases using fluorescence-based assay.** Time courses for each reaction are shown, tables below the graphs indicate the measured rate constants. Error bars represent standard deviation of three independent reactions, rate constants are shown on the table ± the error of the fit.

### 3.1.5.1.3. Reverse genetics of the FluA CTD-binding deficient mutants

*Reverse genetics experiments were performed by Guillaume Fournier and Catherine Isel-Griffiths in the laboratory of Nadia Naffakh at Institut Pasteur, Paris.*

Reverse genetics was used to generate recombinant viruses, bearing mutations impairing the CTD binding in order to follow their effect in an infectious context. The same strain as for the

mini-genome assay (A/WSN/33(H1N1)) was used and the same single and double mutations (K289A, R454A, K635A, R638A and K289A/R454A, K635A/R638A) were generated. The single PA mutants showed titers lower than  $10^3$  plaque-forming units (pfu) per mL, compared to the  $10^7$  pfu/mL for the wild-type virus (Figure 24), while neither of the double mutants could be rescued. The mutant viruses showed small plaque phenotype. The sequence of the PA mutations was confirmed by Next-Generation Sequencing (NGS), but interestingly, additional mutations on PA and/or other subunits (PB1, PB2, NP and M segment) were found (Table 1, Figure 26 and 27). These mutations were observed in almost all of the reads (in two independent amplifications of the same plaque), while not being detected in the wild type. The mutations affecting the CTD binding thus produced strongly attenuated and genetically instable viruses and it is likely that the additional mutations conferred some replicative advantage or compensated the deficiency of CTD binding.

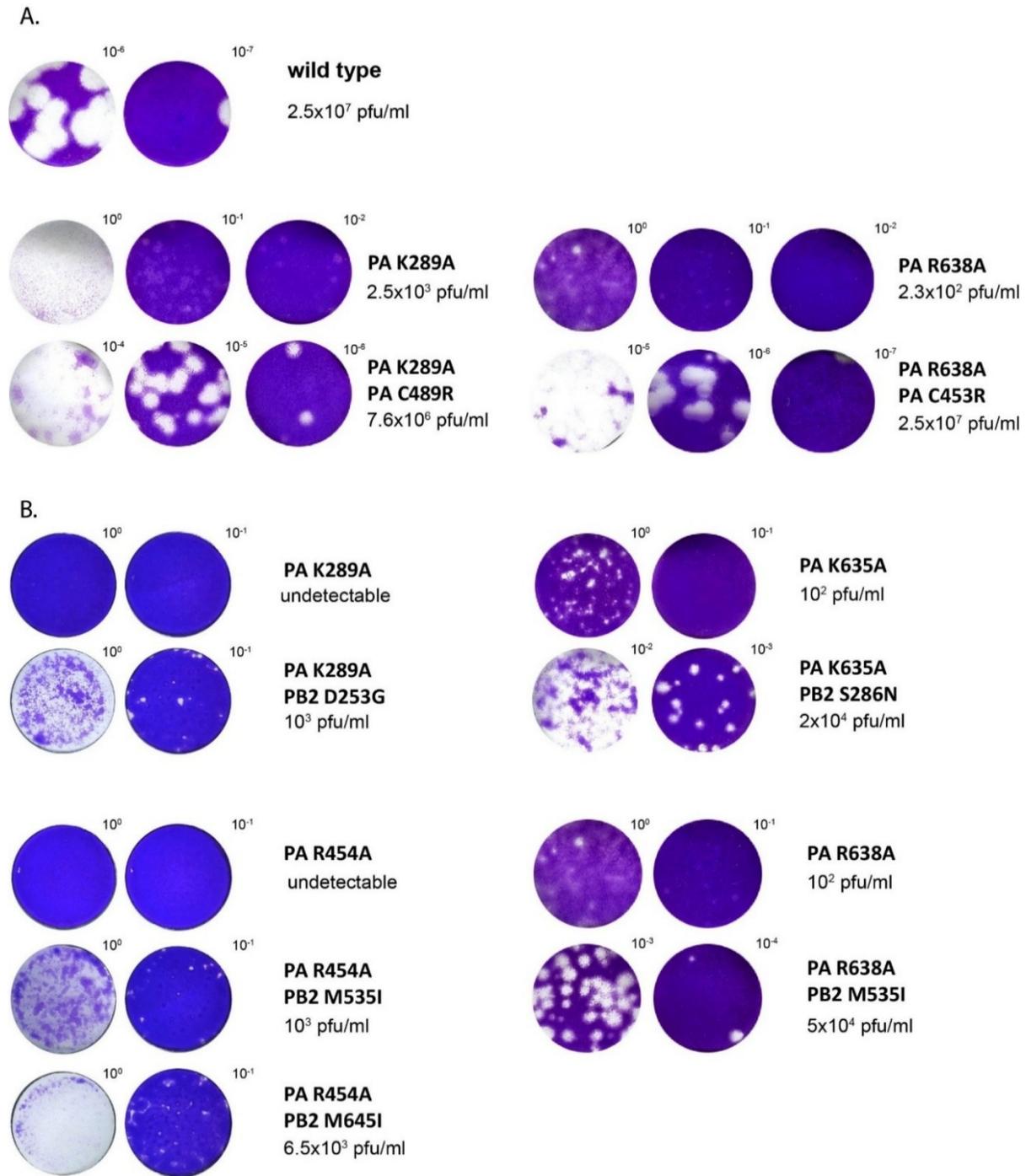


**Figure 24. Reverse genetics of FluA CTD-binding deficient mutants.** A/WSN/33(H1N1) strain was used in for the reverse genetics experiments. Crystal violet staining of cell monolayers infected with reverse genetics supernatants. The dilutions of the supernatants used for infection are indicated above. *Experiments presented on this figure were performed by Guillaume Fournier.*

<b>PA mutation</b>	<b>additional mutations</b>
K289A	PA/C489R PB2/D253G
R454A	PB1/I298M PB2/M535I PB2/M645I
K635A	PA/V90I PB1/D439N PB2/S286N
R638A	PA/I596M PA/R490I PB2/M535I PB2/R375K

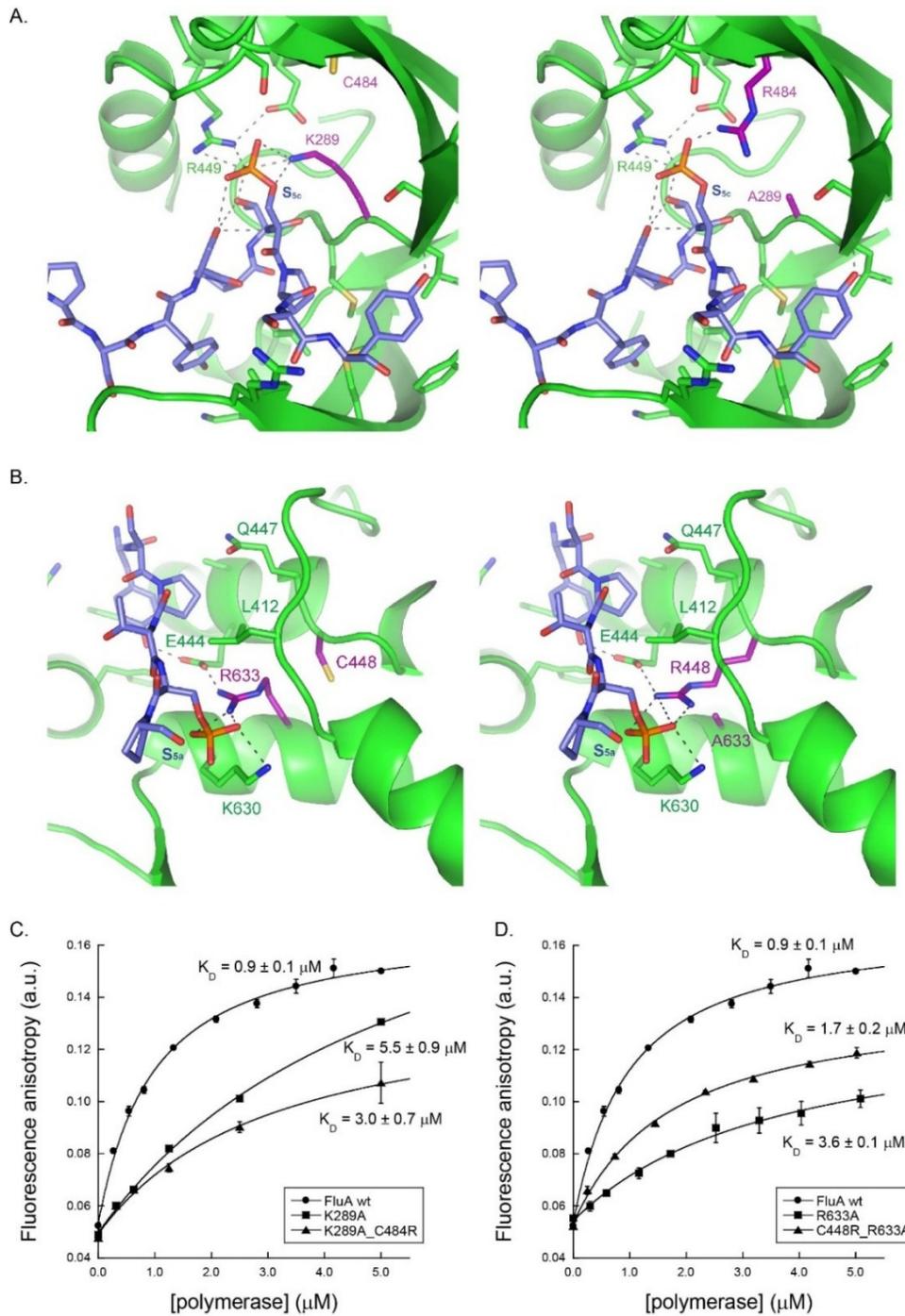
**Table 1. Revertant mutants generated during the passaging of mutant viruses.** Initial mutations are shown on the left, acquired additional mutations on the right. Only mutations affecting the polymerase are listed, since we could not speculate on the role of other mutations affecting the nucleoprotein and M segments.

In order to confirm that these revertant mutations have a real compensatory effect, combinations of the initial CTD-binding deficient mutants with some of the PA and PB2 newly acquired substitutions were produced and the fitness of the respective recombinant viruses was followed. A C453R/R638A mutant which has been previously reported was also tested (see below)<sup>153</sup> in parallel to testing the revertant viruses from our experiments. Both C453R/R638A and K289A/C489R mutants reverted to nearly wild-type infectivity (Figure 25A). The second site PB2 mutations D253G, S286N, M535I and M645I revert their corresponding mutants to intermediate phenotype, creating middle-sized plaques with titers of  $10^4$ - $10^5$ , compared to the small plaques of the initial mutants and titers of  $10^1$ - $10^2$  (Figure 25B).



**Figure 25. Plaque assays for PA revertant mutants.** (A) Revertant mutations on PA: K289A, K289A/C489R, R638A and C453R/R638A. (B) Second site PB2 revertants are showed together with the corresponding PA mutant. Interestingly, M535I mutation appeared during the passage of two different initial mutants (PA R454A and PA R638A). *Experiments shown on this figure were performed by Guillaume Fournier and Catherine Isel-Griffiths, they provided images of the cell monolayers.*

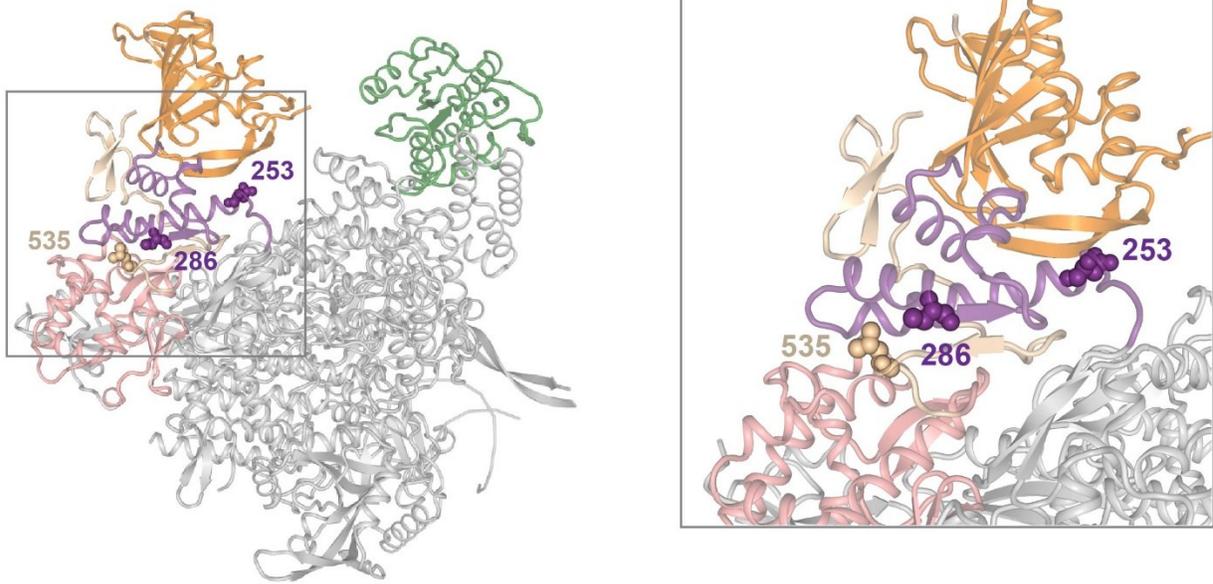
The revertant mutation C489R on PA can be explained by a direct compensation for the reduction of CTD binding caused by the first mutation, since this residue is located in the SeP5 binding pocket. Modelling on the combination of K289A and C484R (corresponding residue in bat/H17N10 strain, for which structural data is available) indicated that it is possible that an arginine at this position could restore the lost binding due to the K289A mutation at site 2, by associating with the phosphoserine (Figure 26A). Another interesting mutation, which has been previously characterized, C453R<sup>153,154</sup> was analogous Cys to Arg mutation at site 1, in proximity to the SeP5, and which could also, according to a model, restore the loss of interaction (Figure 26B) in the corresponding R633A mutant. This mutation was obtained during the characterization of the attenuated R638A virus, whereas the second site revertant C453R restored almost full infectivity. No virus with arginines at both positions (453 and 638) could be generated, consistent with the modelling of the structure, where two long-chain residues would sterically clash. Reverting back to a wild-type virus would require three nucleotide substitutions (Ala to Arg), while replacing Cys with Arg requires only one nucleotide change, but could lead to similar effect by recovering the loss of binding to CTD. In order to confirm this, I produced recombinant polymerases bearing the corresponding mutations in A/H17N10 strain: R633A, C448R/R633A, K289A and K289A/C484R and tested their binding to SeP5 CTD peptide (Figure 26C, D). I could observe an increase of the binding of both C→R mutations compared to their respective K289A and R633A mutants, which supported our hypothesis that these acquired mutations could recover the association of the polymerase to the CTD.



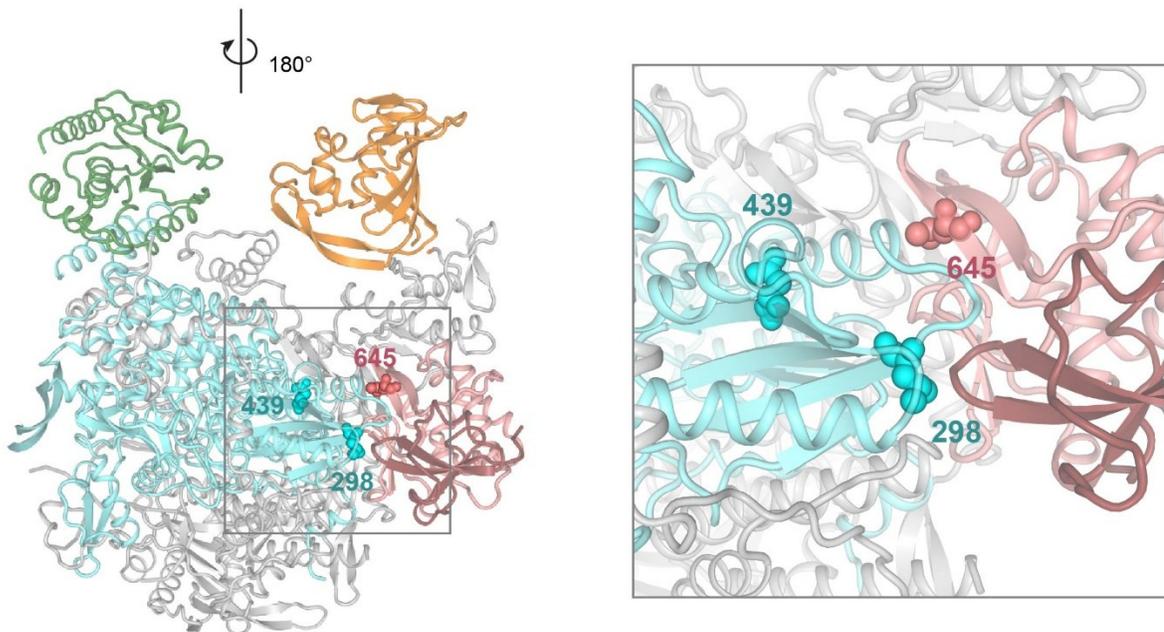
**Figure 26. Analysis of the revertant mutants on PA subunit, rescuing CTD binding.** (A) Comparison between the binding of CTD at site 2, as observed in the structure (left panel) and a model based on this structure, where K289 is substituted with Ala and C484 with Arg (K289A/C484R, right panel). (B) Same as (A) for C448R/R633A (structure of the wild type on the left and model of the mutant on the right). (C), (D) Binding to four-repeat Se5P-CTD peptide of wild type FluA polymerase and corresponding mutants: (C) K289A and K289A/C484R; (D) R633A and C448R/R633A.

Based on the known structure of the influenza polymerase, most of the other compensatory mutations on PB1 and PB2 that partially recover the infectivity of the mutant viruses, could not be explained. They are not directly involved in promoter RNA binding or binding to the capped RNA, so they could not have a direct effect on the enzymatic activity of the polymerase (Figure 27). Nevertheless, the residues map to the same region on PB2 and PB1 and could be involved in interactions between the subdomains: such are S286 and M535, belonging to the mid and cap-627 linker domains, respectively (Figure 27A). Residue I298 is on the interface of PB1 with the PB2 NLS domain, while M645 is on the interface of the 627 domain with PB1 (Figure 27B). Interestingly, these residues are not strictly conserved and differ in H17N10 (PB2 D253, T286, L535, L645; PB1: L298 and D439, respectively). It is possible that they modulate the subunit interactions implicated in the conformational changes of the polymerase, and play a role in the fine-tuning of its activity.

A.



B.

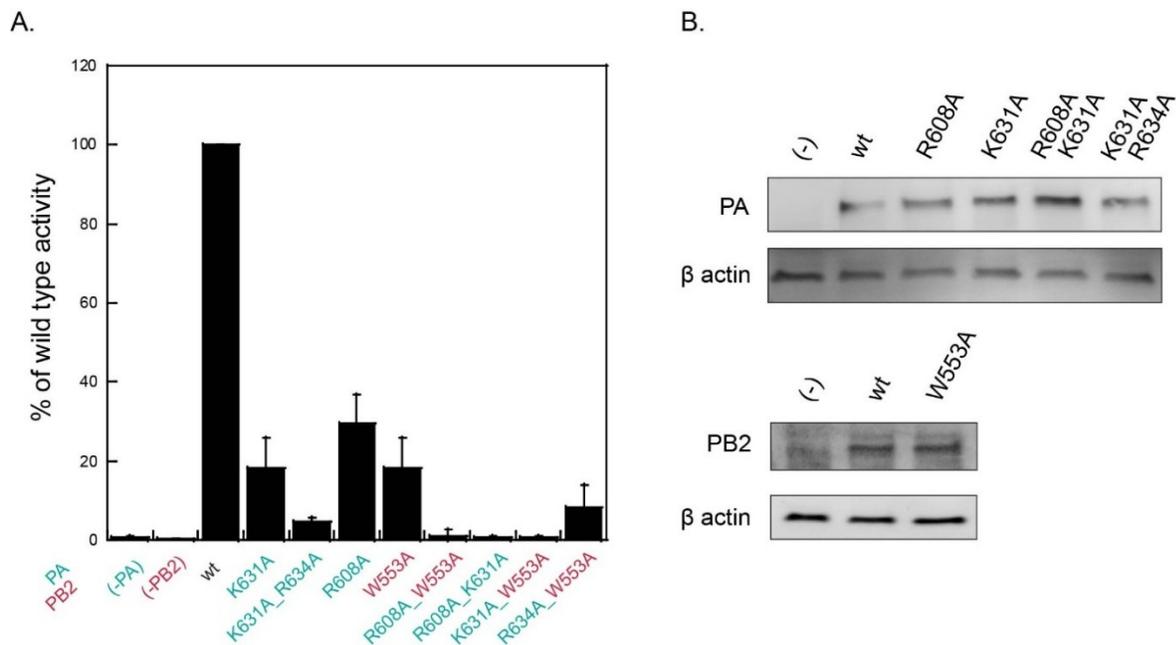


**Figure 27. Revertant mutant residues on PB1 and PB2 mapped on the polymerase structure.** The A/H17N10 residues, corresponding to the revertant mutations in WSN/H1N1 strain, used for reverse genetics experiment, are highlighted with spheres. Cap-binding domain is in orange, endonuclease in forest green (cartoon). (A) PB2 residues mapping on the interface between the mid domain (purple) and cap-627 linker (wheat). 627 domain is colored in salmon (B) Opposite view compared to (A). Residues on PB1 (cyan) and PB2 627 domain (salmon) are highlighted with spheres. NLS domain is colored in ruby brown.

### 3.1.5.2. FluB polymerase – CTD interaction

#### 3.1.5.2.1. Mini-genome assay of FluB CTD-binding deficient mutants

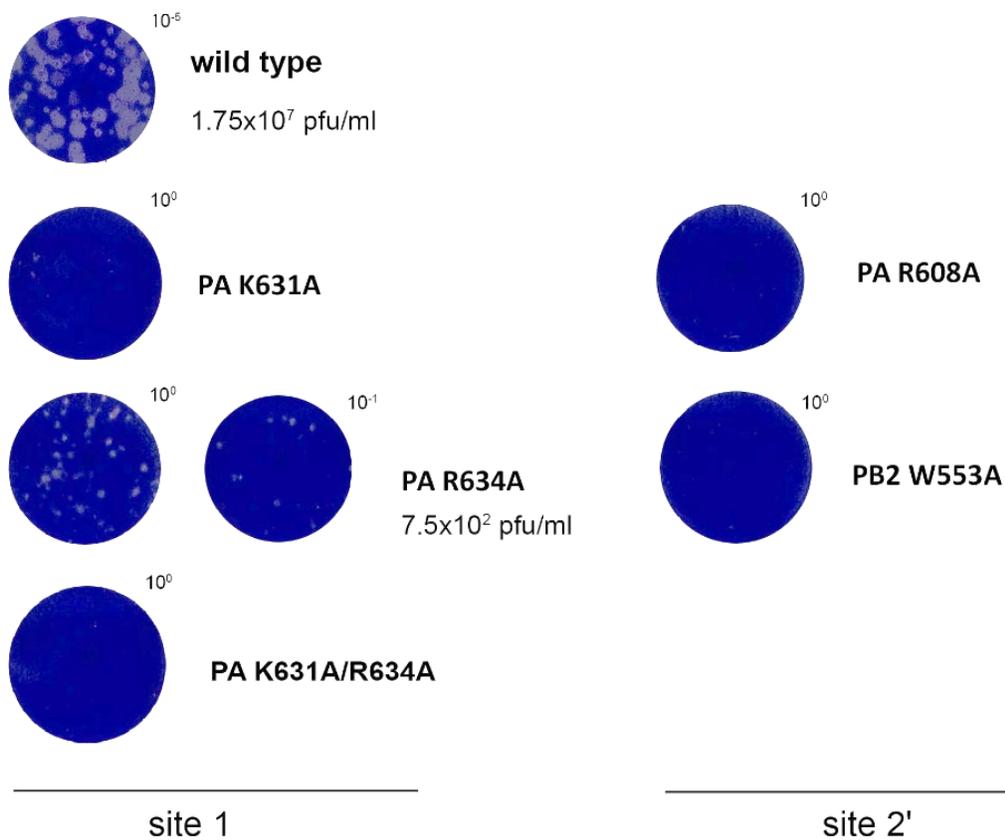
To characterize the interaction between FluB polymerase and CTD, I tested the effect of single mutations at each CTD binding site (for site 1: PA/K631A; site 2': PA/R608A, PB2/W553A) and combination of mutations at one or both sites (site 1: PA/K631A+PA/R634A, site 2': PA/R608A+PB2/W553A; site 1 + site 2': PA/K631A+PA/R608A; PA/K631A+PB2/W553A) in a mini-genome assay. R634A, analogous to R633A in FluA polymerase (R638A in human/avian strains), did not give consistent results in multiple experiments, so this data is not presented. All single mutants had impaired activity (Figure 28), validating the newly discovered binding site. Combination of any two single mutations, either at the same site or at both sites, resulted in a stronger phenotype and almost undetectable activity.



**Figure 28. Mini-genome assay of FluB CTD-binding mutants.** (A) Activity of wild type and corresponding PA and PB2 mutants. Error bars show standard deviation of three independent experiments, performed in triplicate. Labels are colored according to the color-code on Figure 17, with PA in green cyan and PB2 in salmon. (B) Western blot showing the expression levels of the mutants showed in (A).

### 3.1.5.2.2. Recombinant influenza B viruses with CTD-binding deficiencies

Reverse genetics experiments on viruses with mutated CTD-binding site 1 or/and site 2' from influenza B/Brisbane/60/88 strain were performed (for sequence conservation between influenza B strains, refer to Figure 18). Even single mutants at each site (PA: R608A, K631A and PB2: W553A) could not be rescued (Figure 29), suggesting strong attenuation of the viruses due to the impaired CTD binding. Overall, the *in vivo* data strongly implies that disruption of the interaction with Pol II is detrimental to the viral replication during infection.



**Figure 29.** Crystal violet staining of cell monolayers infected with recombinant viruses with several CTD-binding mutations in influenza B (B/Brisbane/60/88). The viruses are attenuated and have smaller-size plaque phenotype. *Experiments presented on this figure were performed by Catherine Isel-Griffiths.*

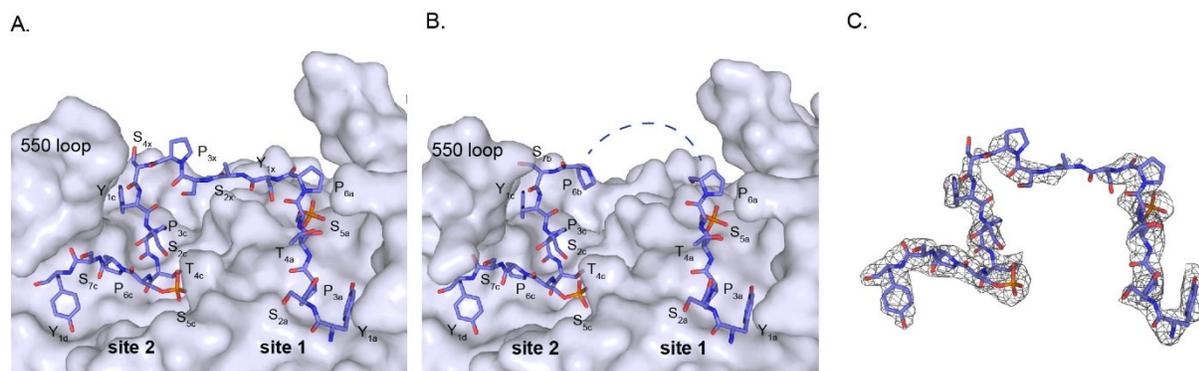
### 3.1.6. Structure-based drug design targeting the FluPol – CTD interaction

In collaboration with the EMBL Heidelberg Chemical Core Facility, we wanted to exploit the possibilities to use the CTD binding site on influenza polymerase as a drug target. In order to understand better the biochemical basis for the recognition by FluA polymerase, Peter Sehr and Michael Claron synthesized CTD-mimic peptides with various modifications and set up a high-throughput binding assay to assess the affinities of the produced peptides. They produced and assayed about 40 different peptides and validated the previously measured affinities. The principle of the assay (AlphaScreen™) is based on an energy transfer between donor and acceptor beads, which are brought into close proximity by the interaction of a protein and a ligand that are bound to each type of bead. Digoxigenin-labelled CTD peptides were immobilized on anti-digoxigenin-coated donor beads while FluPol was attached to a Ni-chelate-coated acceptor bead by its N-terminal His tag. Thus, the interaction of the peptide with the polymerase generates a signal. Measuring the inhibition of this interaction by the addition of a competitor allowed screening peptides with various modifications.

The data suggested that the minimal binding motif contains three consecutive repeats with Ser5 phosphorylation at repeat 1 and 3, and including the Tyr of a fourth repeat (22mer). The lack of phosphorylation of the second repeat did not affect the binding ( $IC_{50}$  of 0.7 versus 0.9  $\mu$ M for the original four-repeat SeP5 peptide). The affinity of a peptide containing three phosphorylated repeats and the following Tyr to FluA polymerase was almost identical to the affinity for the four-repeat SeP5 peptide (0.7  $\mu$ M), while a similar peptide, lacking the Tyr from the fourth-repeat showed a nine-fold difference in the affinity (6.5  $\mu$ M). This is consistent with the structural data, showing extensive interactions with this Tyr (Figure 13).

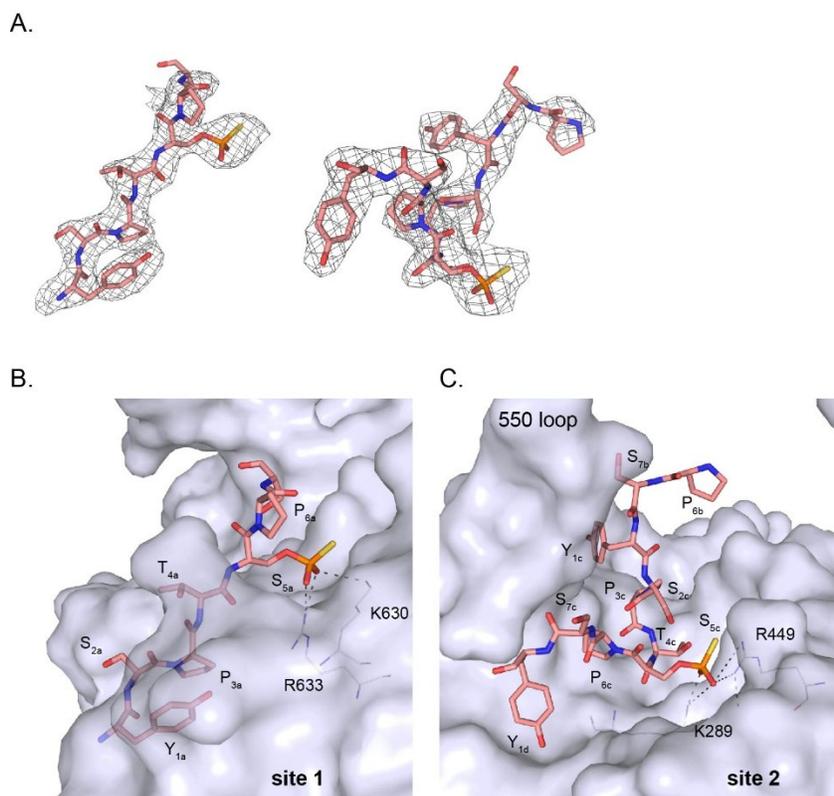
Another insight from the structure was that the middle repeat did not form interactions with the polymerase, so different linkers were designed in order to replace this repeat. None of the modified linkers was bound with affinities close to the originally identified peptide, except of a peptide containing a four amino acid linker (YSPS) instead of a full repeat (YSPTSPS), which has slightly increased affinity of 0.7  $\mu$ M compared to the 0.9  $\mu$ M of the wild type peptide. Since the peptide with the shorter linker bound more tightly, I wanted to see if there are additional interactions formed by the linker residues. I solved the structure of influenza A polymerase in complex with this peptide (Figure 30). The mode of binding of the peptide is very similar to the wild type CTD

peptide (Figure 30A and B), but the density of the linker was not good enough to build side chains of the missing amino acids (Figure 30C). It cannot be excluded that there are additional interactions that contribute to the affinity of the peptide for the polymerase, but they could not be observed in the structure.



**Figure 30. Structure of FluA polymerase with a tetralinker CTD peptide.** (A) Close-up on the interaction between PA-C of FluA polymerase and a peptide with four-amino acid linker, the polymerase is in grey surface representation and peptide in blue sticks. (B) The original four-repeat Se5P-CTD bound-FluA polymerase structure (PDB:5M3H), shown in the same perspective as (A). (C)  $F_o-F_c$  map of the tetralinker peptide, contoured at  $2.5 \sigma$ .

Another identified peptide with increased affinity to FluA polymerase ( $0.4 \mu\text{M}$ ) contains thio-phosphate instead of phosphate at position 5. This modification is thought to improve the stability of phospho-peptides against phosphatases. I also co-crystallized a complex of FluA polymerase with the thiophosphate-peptide (Figure 31). The density allowed to identify the sulfur atoms (Figure 31A), but at both binding sites they did not interact with the protein and were pointing away from the surface (Figure 31B, C).



**Figure 31. Co-crystal structure of FluA polymerase with 22mer thiophosphate CTD peptide.** (A) Omit map ( $F_o - F_c$ ) of the peptide moieties (analogously to the original four-repeat peptide-bound structure, the densities for the two sites are not connected). (B), (C) Binding at site 1 (B) and site 2 (C) of FluA polymerase. The protein is shown in grey surface representation and the peptide in salmon sticks. Hydrogen bonds between the thiophosphate and the protein are indicated with dashed lines.

Taken together, this work provides a detailed biochemical characterization of the requirements for CTD binding to FluPol, and, supported by structures of the promising hits, would allow for the rational design of peptides and eventually compounds that could inhibit the association of influenza polymerase with Pol II. This strategy was successfully used for the identification of several inhibiting compounds, whose characterization is ongoing.

The data presented in this section showed for the first time a direct association of FluPol with the CTD of Pol II and revealed the structural basis for this interaction. These results allowed the design of mutants that disrupt this interaction and to study their effect *in vitro* and in the context of the cell. The impaired interaction had severe consequences on the FluPol functions only in cells, suggesting that the association of the viral and host polymerase is an important prerequisite for accessing the capped primers, necessary for the viral transcription. Moreover, this interaction turned out to be essential during viral infection since recombinant viruses carrying these mutations were highly attenuated and genetically unstable. We further characterized the details of the interaction with the help of modified peptides, designed to mimic the CTD. This helped elucidating the mechanism of binding, which would allow identifying compounds that would inhibit this essential interaction.

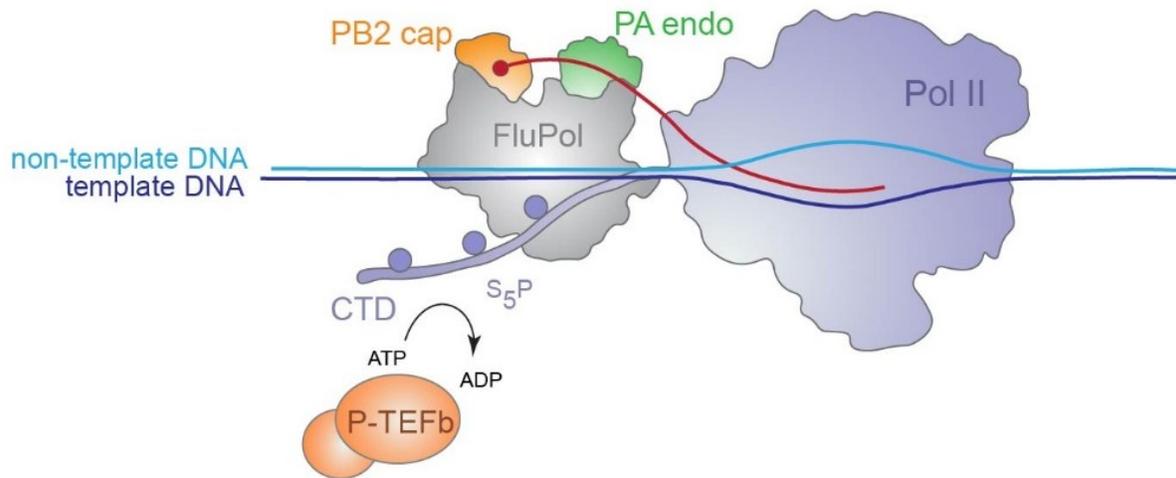
### 3.2. Assembly of full-length complex between influenza polymerase and elongating RNA polymerase II

In order to study the FluPol – Pol II interaction in a system that is closer to the real cellular context, I aimed to reconstitute a cap-snatching complex. It is possible that FluPol might use a similar mechanism to the capping enzyme to get recruited to the exiting nascent RNA from Pol II <sup>75</sup>. Therefore, the aim was to assemble an elongating Pol II complex with FluPol from purified components and characterize it structurally and functionally.

To achieve this, endogeneous Pol II purified from *S. scrofa* thymus was used. This protein has been successfully used for structural characterization of elongation complexes and is almost identical to its human homologue <sup>70,82</sup>. Recombinant FluPol from A/H17N10 strain was used as a model for FluA strains. As discussed in the introduction, FluA/H17N10 polymerase can replicate in human epithelial cells, and has 70 per cent sequence identity to its human counterpart. The use of heterologous proteins was therefore not expected to affect the assay.

The assembly of the complex follows several steps. Initially, an elongating Pol II complex is formed, using endogenously purified swine Pol II bound to a pre-annealed DNA/RNA mismatch-bubble scaffold, as previously described <sup>155,156</sup>. The RNA of the elongation complex must be long enough to exit Pol II and be bound by the FluPol. Additionally, the RNA is pre-capped in order to be recognized by the FluPol cap-binding domain. Then the CTD of Pol II is phosphorylated by a Ser5 kinase (Figure 32). P-TEFb (CDK9) was used for this purpose, as it is known to phosphorylate preferentially Ser5 *in vitro* <sup>78,79</sup>, and FluPol associates with SeP5 CTD <sup>82,93,131,147</sup>. In the real cellular context, it is not known whether FluPol binds to CTD prior to or after its phosphorylation by P-TEFb. Nevertheless, the strong *in vitro* preference of the kinase for Ser5 suggested this would be a suitable kinase for our assay. Plausibly more physiologically relevant CTD kinase would be TFIIH, which is a ten-subunit complex and therefore was not used for practical considerations.

FluPol, pre-bound to a short RNA promoter in order to favour the transcriptionally competent conformation, is then added to the phosphorylated elongating Pol II. Wild type FluPol was used for the functional assays, while for structural studies we used an endonuclease-deficient mutant (PA/K134A) to prevent cleavage of the capped RNA, which could trigger dissociation of the two polymerases.



**Figure 32. Schematic representation of the cap-snatching complex.** The tDNA and ntDNA contain a non-complementary stretch, where the RNA base-pairs with the tDNA, mimicking a synthetic transcription bubble. Phosphorylation of the CTD by P-TEFb is depicted, FluPol binds to the S<sub>5</sub>P CTD. The capped RNA exiting from Pol II is bound by the cap-binding domain of PB2.

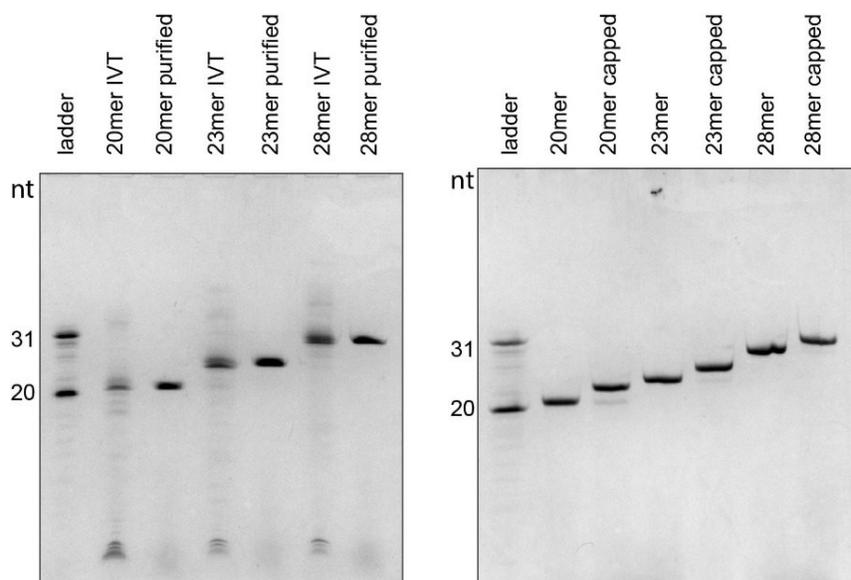
### 3.2.1. Reconstitution of a stalled cap-snatching complex

#### 3.2.1.1. *In vitro* transcription and capping of substrate RNAs

In order to investigate the cap-snatching mechanism, it was required to reconstitute the complex in a state where the capped RNA is part of the Pol II elongation complex and is recognized by the viral polymerase. Since the optimal length of the RNA was not known, I designed the capped RNAs based on the available structural data of the transcribing mammalian Pol II and the complex of the yeast CE with Pol II (the sequences can be found in Appendix Table 3). The minimal length, necessary to reach the Pol II surface is around 14 nt and capping can occur at a length of around 20 nt<sup>75,156</sup>. The structure of the capping enzyme bound to elongating Pol II contains an RNA of 23 nt. I therefore produced three RNAs with different lengths, 20, 23 and 28 nt. A 28 nt RNA would be long enough to be bound simultaneously to the cap-binding domain and be accommodated in the endonuclease domain (a distance of 10-12 nt), while it is still bound to Pol II.

I produced these RNAs by T7 *in vitro* transcription and purified them from a denaturing gel with single-nucleotide resolution. This avoids any heterogeneity at the 3' that could be introduced by

T7, because the sequence at the 3' is crucial for the formation of the mismatch bubble. The RNAs were then capped with purified recombinant vaccinia capping enzyme (Figure 33).

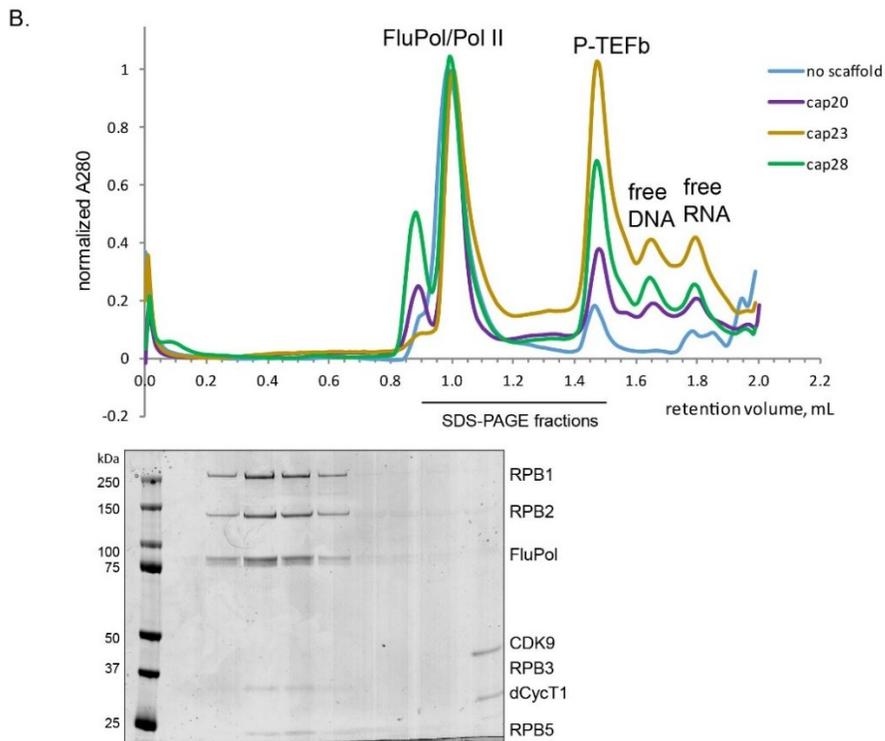
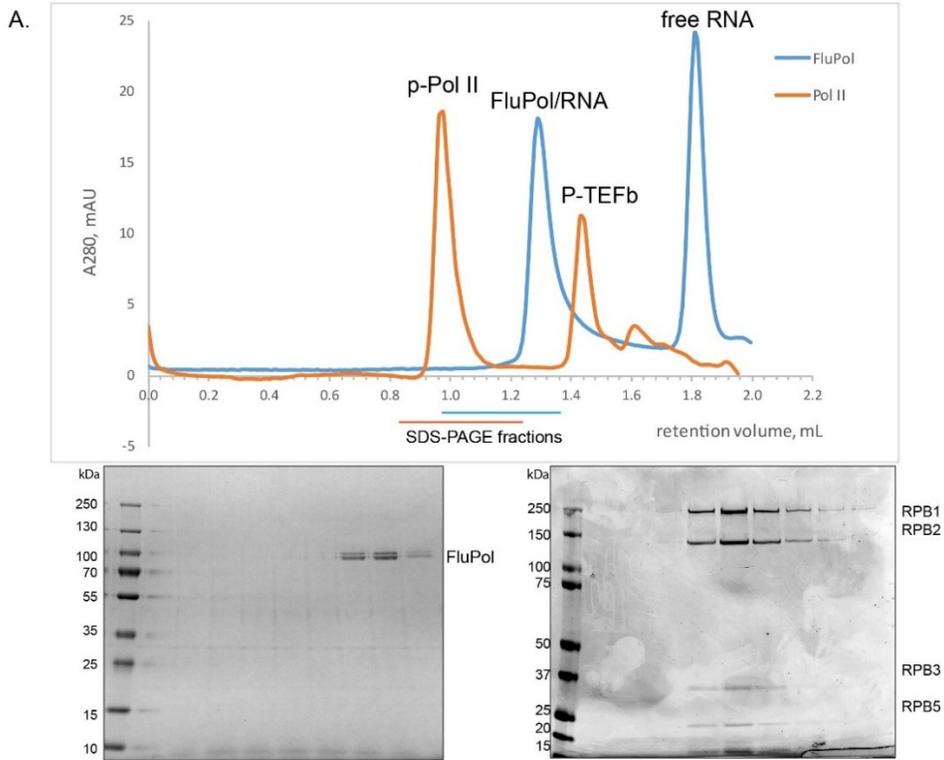


**Figure 33. Purification and capping of *in vitro* transcribed RNAs.** Left: 20, 23 and 28-mer products from *in vitro* transcription and purified from the gel. Right: capping of the purified RNAs, capping results in slower migration of the RNAs<sup>129</sup>. The ladders contain synthetic RNAs of known sizes.

### 3.2.1.2. Assembly of FluPol/Pol II complex with different scaffolds

I first tested if a complex between FluPol and Pol II can be formed using the different capped RNAs that I had produced. DNA/RNA mismatch bubble scaffolds were pre-annealed and incubated with Pol II. Promoter-bound FluPol, P-TEFb and ATP were then added to Pol II and the sample was analyzed by size-exclusion chromatography (SEC) with a Superdex 200 Increase 3.2/300 column (S200). Since the resolution of the S200 size exclusion column that was used does not allow to observe a shift between free Pol II (550 kDa, Figure 34A, orange curve) and Pol II bound to FluPol (around 850 kDa, Figure 34B), the complex formation was analyzed by the disappearance of the free FluPol peak (Figure 34A, blue curve) and the presence of FluPol in the Pol II peak fractions on a SDS-PAGE. Complex formation could be observed in all the conditions used in these experiments. There were no differences in the complex assembly with the different length capped RNAs used for the scaffold, moreover, an association of the two polymerases could also be formed in the absence of the RNA. Therefore, this method did not allow to detect the

minimal required length of capped RNA that can be bound by FluPol, which prompted me to address this question by a functional assay (see section 3.2.2.1).



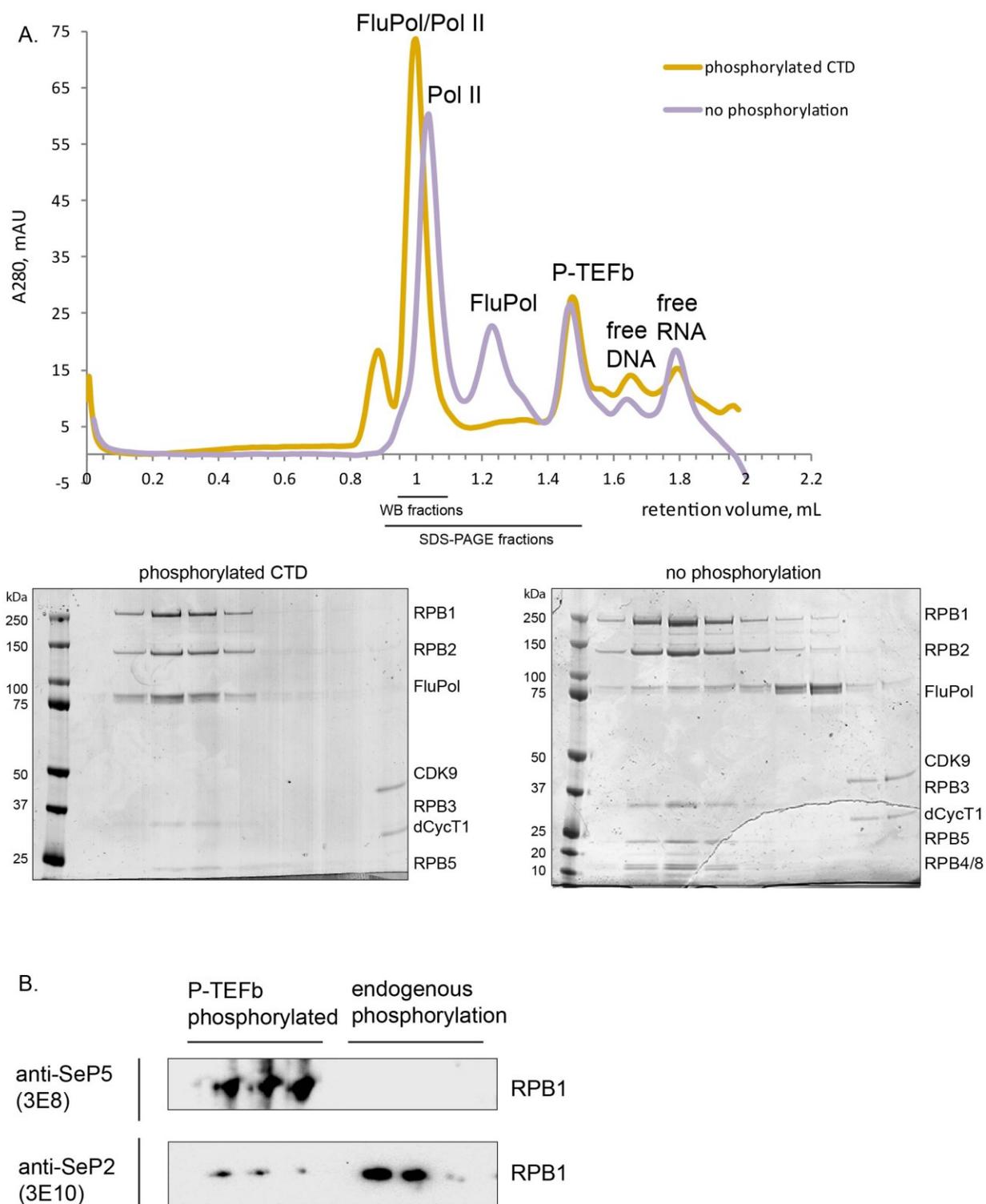
**Figure 34. Size-exclusion chromatography profiles (SEC) of proteins alone or the complex between FluPol and Pol II.** (A) Pol II, phosphorylated by P-TEFb (orange curve) and FluPol, bound to the RNA promoter (blue curve) with corresponding SDS-PAGE fractions, FluPol on the left and Pol II on the right. (B) SEC profile of the complex between phosphorylated Pol II and FluPol with different RNA scaffolds (purple, yellow and green curve) or in the absence of capped RNA (blue). The profiles are normalized to the peak maxima for comparative purposes. Representative gel of the complex containing 23-mer scaffold is shown.

### 3.2.1.3. Role of the CTD in complex assembly

The next step was to address the role of CTD phosphorylation on the interaction of FluPol with the full-length Pol II. To this end we used antibodies against SeP2 and SeP5 to monitor the phosphorylation state of the CTD, and P-TEFb kinase to enrich the Ser5 phosphorylation. To assess the effect of the kinase on complex formation, the complex was assembled in the absence or presence of ATP and analyzed on size-exclusion chromatography (Figure 35A). Western blot with antibodies against SeP2 and SeP5 showed that the purified Pol II used has mainly SeP2 endogenous phosphorylation, whereas P-TEFb preferentially phosphorylates Ser5 *in vitro* (Figure 35B) <sup>136</sup>. In the absence of P-TEFb activity, less FluPol is bound to Pol II and a free FluPol peak is observed. This confirmed that FluPol interacts with the SeP5 CTD also in the context of the full-length Pol II and that this association is required for bringing the two polymerases together, as suggested by our previous data (see section 3.1).

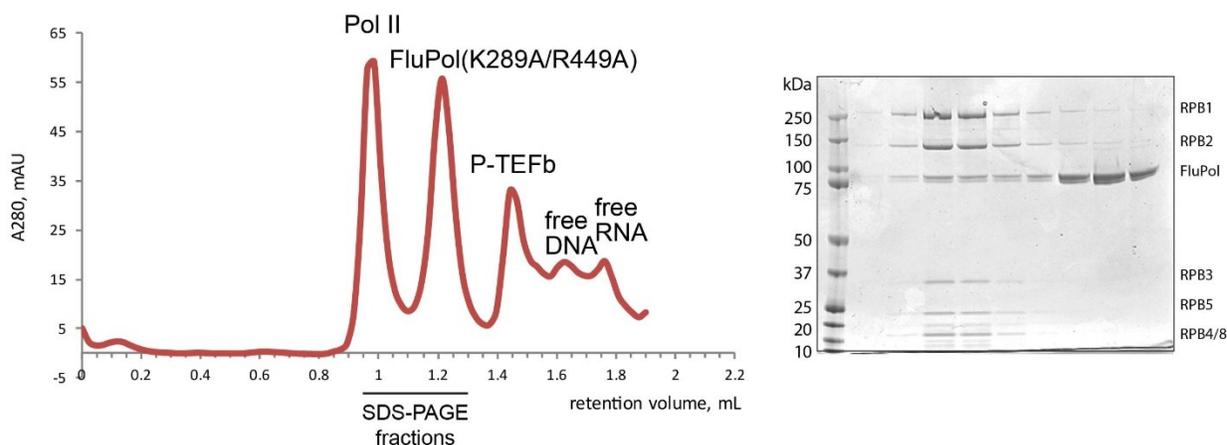
To further test the CTD-dependence of the interaction in the context of our reconstituted system, one of the CTD-binding deficient mutants of FluPol (K289A/R449A) was analyzed for complex assembly by SEC (Figure 36). In line with the previous data (see Figure 20), a much weaker association between the two polymerases was observed and most of the mutant FluPol stayed in its free form.

The conclusion of these experiments is that the Ser5 phosphorylation is a necessary condition for the interaction of FluPol with full-length Pol II and that weaker binding to the CTD due to (1) non-sufficient phosphorylation (Figure 35), or (2) mutation affecting the CTD binding (Figure 36) does not allow complex formation.



**Figure 35. Effect of CTD-phosphorylation on FluPol/Pol II complex assembly.** (A) SEC profile of FluPol/Pol II complex using a 23-mer scaffold, with (orange curve) and without (purple curve) additional Ser5 phosphorylation. Below, SDS-PAGE of the corresponding peaks, phosphorylated sample on the left, non-phosphorylated on the right. (B) Western blot with anti-CTD antibodies on the fractions of the peaks of both chromatograms from (A), showing

only RBP1 subunit. Samples taken for the Western blot are indicated below the chromatogram. 3E8 antibody preferentially recognizes SeP5 and can also detect SeP2-SeP5 CTD; 3E10 mainly recognizes SeP2, but has much weaker reactivity for SeP2-SeP5, which explains why the P-TEFb-phosphorylated sample has weaker SeP2 signal compared to the endogenously-phosphorylated one, which seems to be more enriched in SeP2 alone. For reference on the antibody reactivity towards different CTD peptides, see <sup>136</sup>.

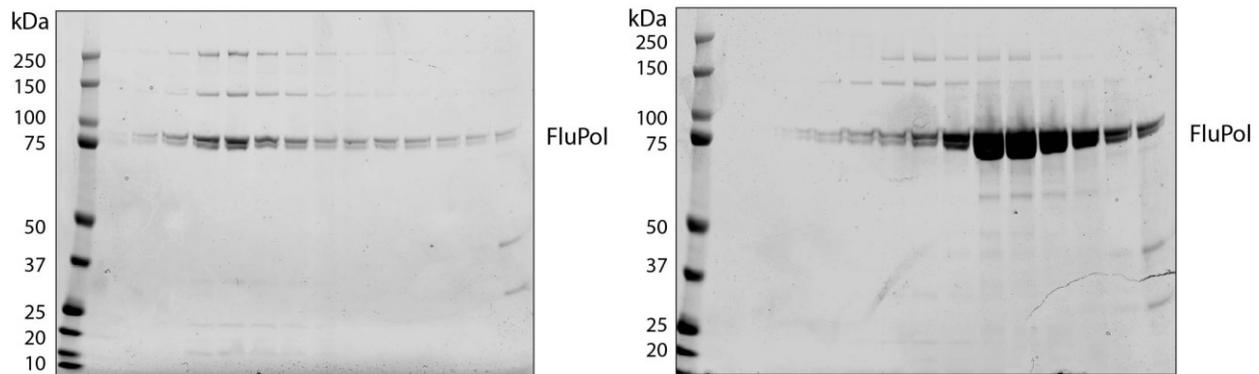
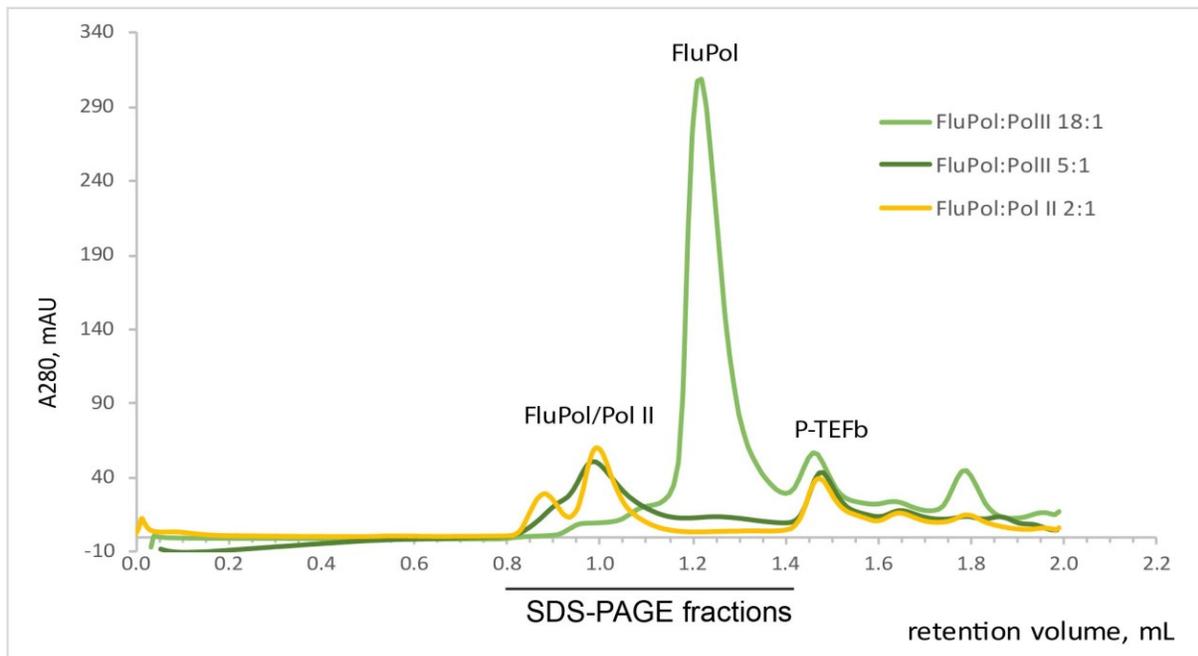


**Figure 36. Effect of the disruption of CTD binding on the complex formation.** Left, SEC profile of FluPol (K289A/R449A) with Pol II, SDS-PAGE of the corresponding peaks on the right.

### 3.2.1.4. Stoichiometry of the FluPol/Pol II complex

To investigate the stoichiometry of the FluPol/Pol II complex, assembly assays were performed at different ratios of the two polymerases. The CTD of the mammalian Pol II has 52 repeats and the structure of FluPol bound to the CTD-peptide shows that it binds to at least four repeats, suggesting that theoretically up to 13 monomers of FluPol can be accommodated on one Pol II. The two polymerases were incubated in different ratios and complex formation was followed by SEC analysis. For the complex assembly, a 2-fold molar excess of FluPol was used and all the FluPol was in the Pol II-bound fraction (Figure 37, in yellow). Increasing the ratio towards a higher excess of FluPol (5:1 FluPol : Pol II) already shows that some unbound FluPol is present in the sample (Figure 37, dark green), and mixing an excess of FluPol in respect to the theoretical binding sites on the CTD (18:1) indicates that most of the FluPol is not complexed with Pol II (Figure 37, light green). These data suggest that the stoichiometry of FluPol to Pol II is close to 2:1 and that not all

the possible binding sites on the CTD can be simultaneously occupied by the viral polymerase, not even *in vitro*, when the CTD is free of any other factors. It is plausible that too many FluPol could not be accommodated on the CTD because of steric hindrance or to the CTD forming transient structures (e.g. turns) when it is not associated with other factors<sup>60</sup> and therefore it would not be fully accessible for as many interactions as theoretically expected.



**Figure 37. Stoichiometry of the FluPol/Pol II complex.** SEC profile of the three different ratios of FluPol to Pol II (2:1, 5:1, 18:1). SDS-PAGE samples for the 5:1 and 18:1 ratio samples. SDS-PAGE analysis of the 2:1 sample can be seen on Figure 34B.

### **3.2.2. Functional characterization of FluPol-Pol II coupling**

Having the fully assembled FluPol/Pol II complex in hand, I aimed to investigate if the activity of FluPol is affected in this context. In particular, I wanted to characterize the length of the RNA necessary for FluPol to perform cap-snatching on the elongating Pol II.

In order to follow the activity of FluPol, I established an assay using radiolabeled capped RNAs, which were pre-bound to elongating Pol II in a mismatch bubble (as described in the previous section) and followed the endonucleolytic cleavage by FluPol. The assays were set up at a limiting concentration of FluPol so that the cleavage of a capped substrate is not complete, which allowed me to observe modulations of the activity. In order to form a stable complex, I phosphorylated Pol II with P-TEFb, therefore all scaffolds were designed in a way that there is no elongation in presence of ATP alone.

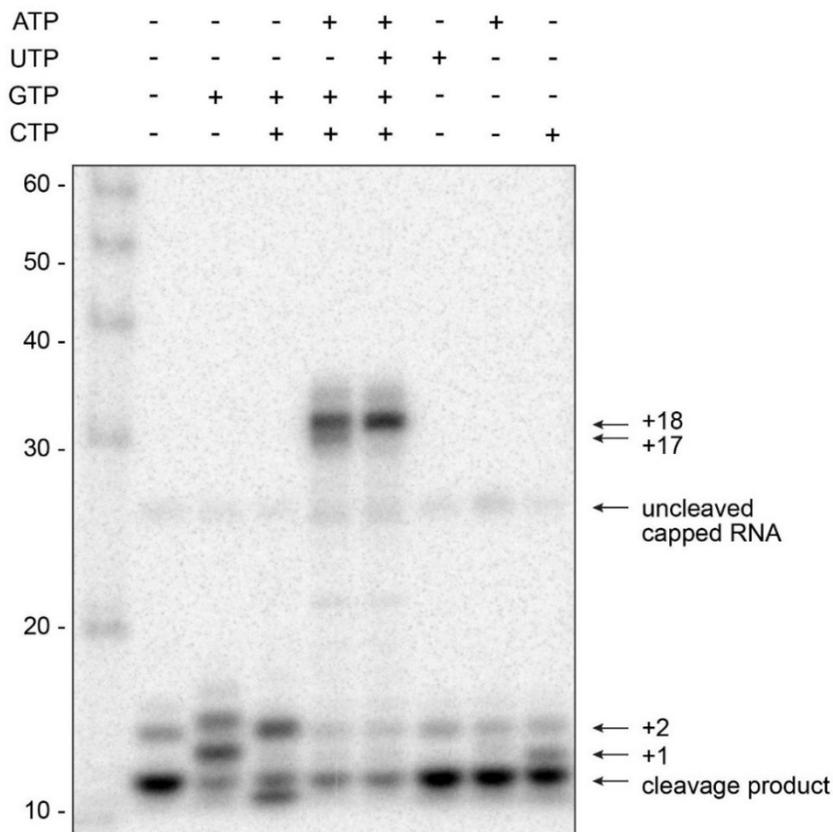
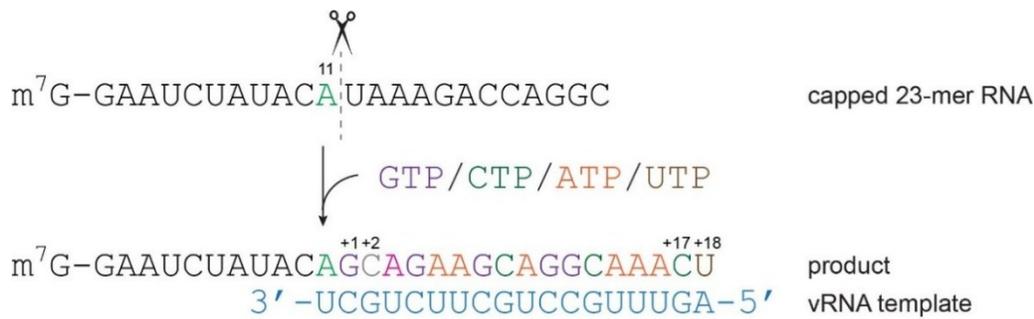
#### **3.2.2.1. Recognition by FluPol of Pol II-pre-bound RNAs with different lengths**

In order to study the length of the RNAs, recognized by FluPol in the context of the elongation complex, I set up an assay to produce several RNAs differing in length by 1 nt. For this purpose, I designed DNA scaffolds containing a various number of As in the template DNA, which would allow incorporation of 1 to 5 Us in the RNA by Pol II (Figure 38).



**Figure 38. Sequences of the DNA scaffolds used for elongating the capped RNA with one nucleotide (UTP) at a time.**

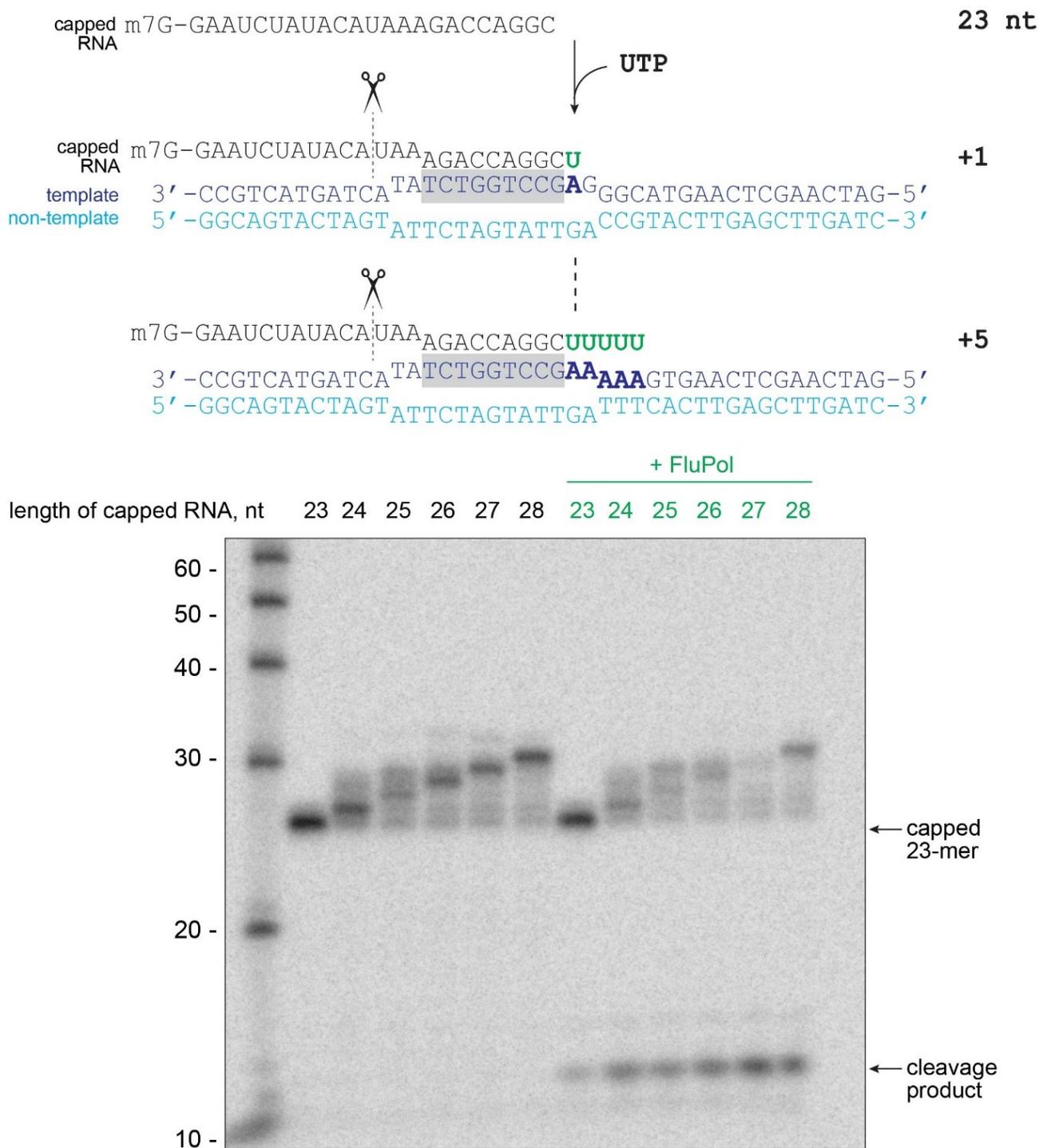
I first tested whether the capped RNA (23-mer) could be recognized by the FluPol endonuclease and used as a primer for vRNA-template-dependent transcription (Figure 39). The sequence of the vRNA template is highly conserved and if the primer base-pairs with the first residue from the 3', the expected incorporation pattern is: GTP, CTP and ATP. This was the pattern I observed, which confirmed that (1) cleavage occurs at the 11<sup>th</sup> residue after the cap (A), which base-pairs with the first U of the template; (2) unless there is GTP, CTP and ATP present, UTP is not incorporated. Knowing the length of the capped RNA substrate and the exact position of the cleavage site allowed to determine the distance between the cleavage and the 3' end of the primer, which would be inside Pol II in the following assay.



**Figure 39. Sequential incorporation of NTPs by FluPol, using v3'18 template vRNA.** The capped primer can only be elongated in presence of GTP (lanes 2-5). The cleavage happens mainly after the A11, which base-pairs with the first U from the 3' end of the template. Some secondary shorter cleavage products are also observed (lane 1 after the ladder). RNA ladder is indicated with respective length in nucleotides. Capped RNA products migrate higher than uncapped RNAs used for ladder (see Figure 33).

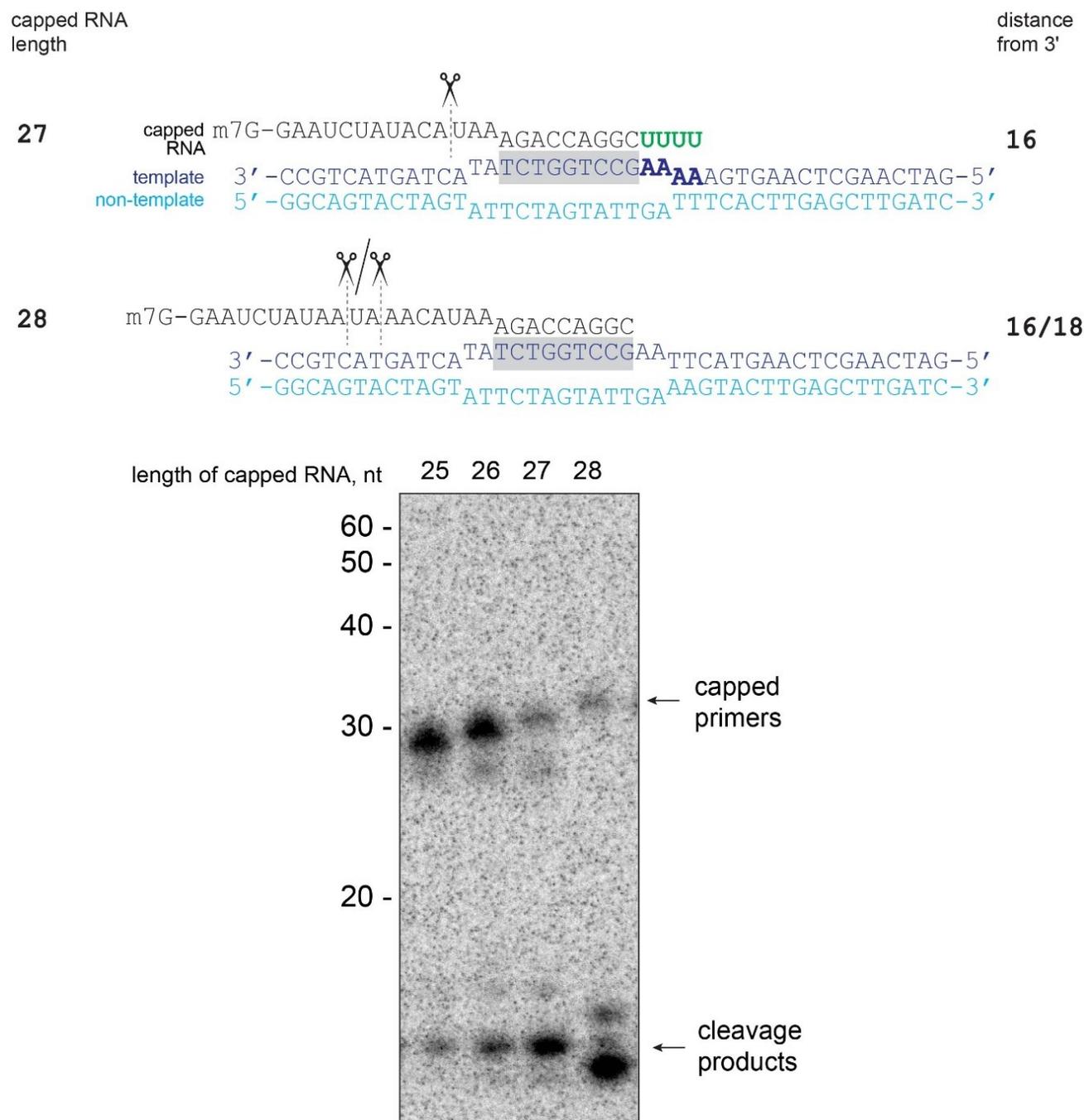
UTP was then added to the Pol II elongation complex, assembled using the different scaffolds. The capped RNA (23-mer) was elongated with one nucleotide at a time in order to obtain lengths from 23 to 28 nt. Since the cleavage by FluPol happens at a fixed position 11 nucleotides downstream of the cap, the distance between the cleavage site of FluPol and the 3' end (located in the Pol II active site) would vary by 1 nt with the different scaffolds. The elongation by Pol II was first followed in absence of FluPol and then FluPol was added (Figure 40). I observed that cleavage was most prominent at a length of 27 nucleotides, leaving 16 nt between the Pol II active site and the FluPol cleavage site. Either longer or shorter capped RNAs were not cleaved as efficiently, suggesting that there is an optimal length recognized by FluPol.

The capped RNA of 28 nt produced by *in vitro* transcription (section 3.2.1.1) had two sites recognized by FluPol endonuclease (10<sup>th</sup> and 12<sup>th</sup> nucleotide), but the shortest distance between the cleavage site and the 3' end was the same (16 nt), therefore it was also a good substrate (Figure 41). In conclusion, these experiments showed that the optimal distance between the Pol II active site and the FluPol cleavage site is 16 nt, suggesting that the endonuclease of FluPol must be in contact with Pol II close to the RNA exit.



**Figure 40. Determination of the optimal length of Pol II-pre-bound RNA, recognized by FluPol.** Two examples of DNA/RNA scaffolds, which allows incorporation of 1 or 5 Us, are shown above; all scaffolds are shown on Figure 38. Below, denaturing urea gel showing elongation by one nucleotide by Pol II (first six lanes after the ladder), then FluPol is added and cleavage of the different lengths of substrate RNAs is observed.

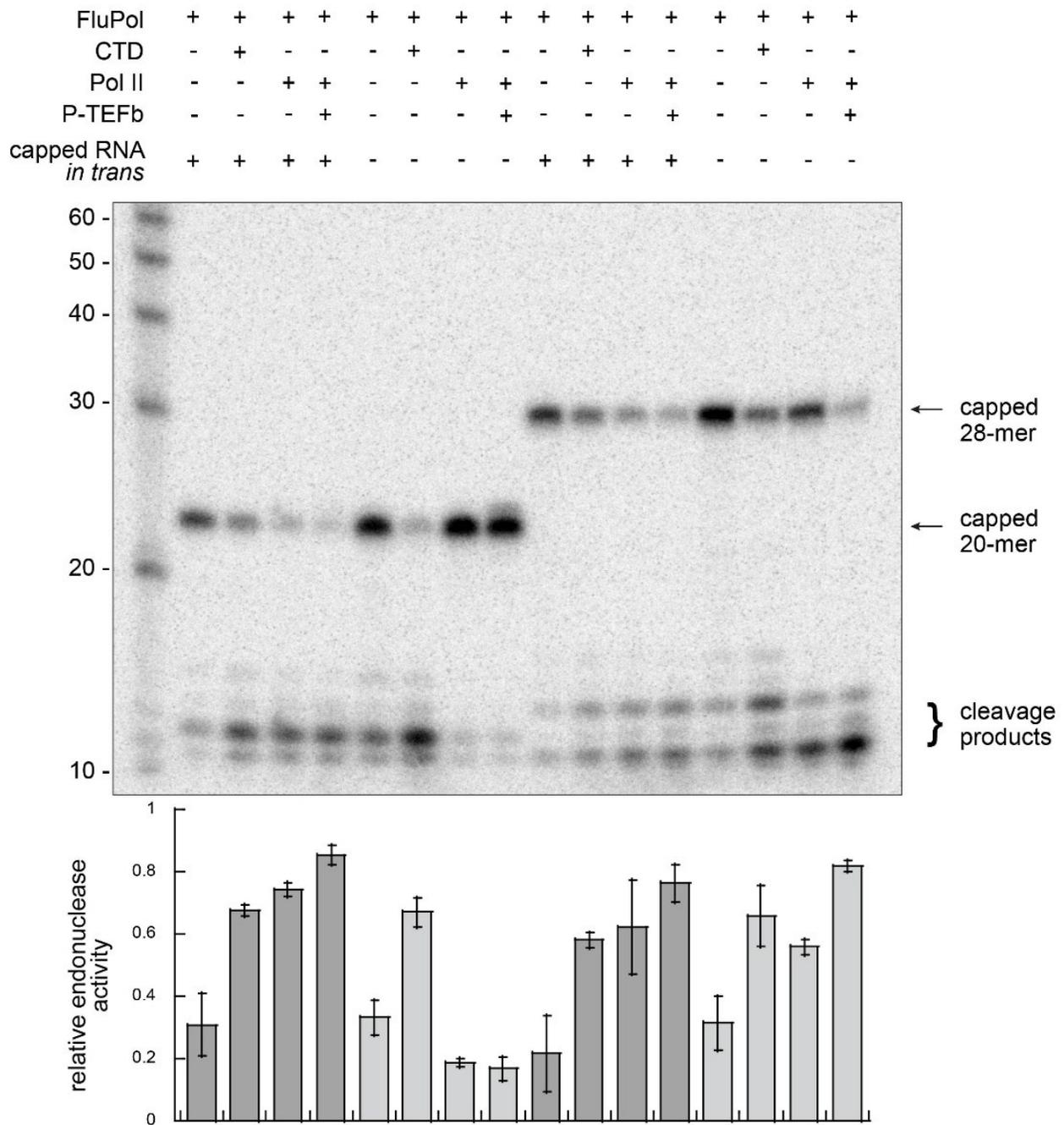
B.



**Figure 41. Comparison of the cleavage of 27 and 28-mer RNA by FluPol.** 25, 26 and 27 nt products are produced as in Figure 40, purified 28 nt RNA is annealed with the same length of scaffold, the sequences of the RNA and DNA are shown above.

### 3.2.2.2. Effect of Pol II on RNA cleavage by FluPol

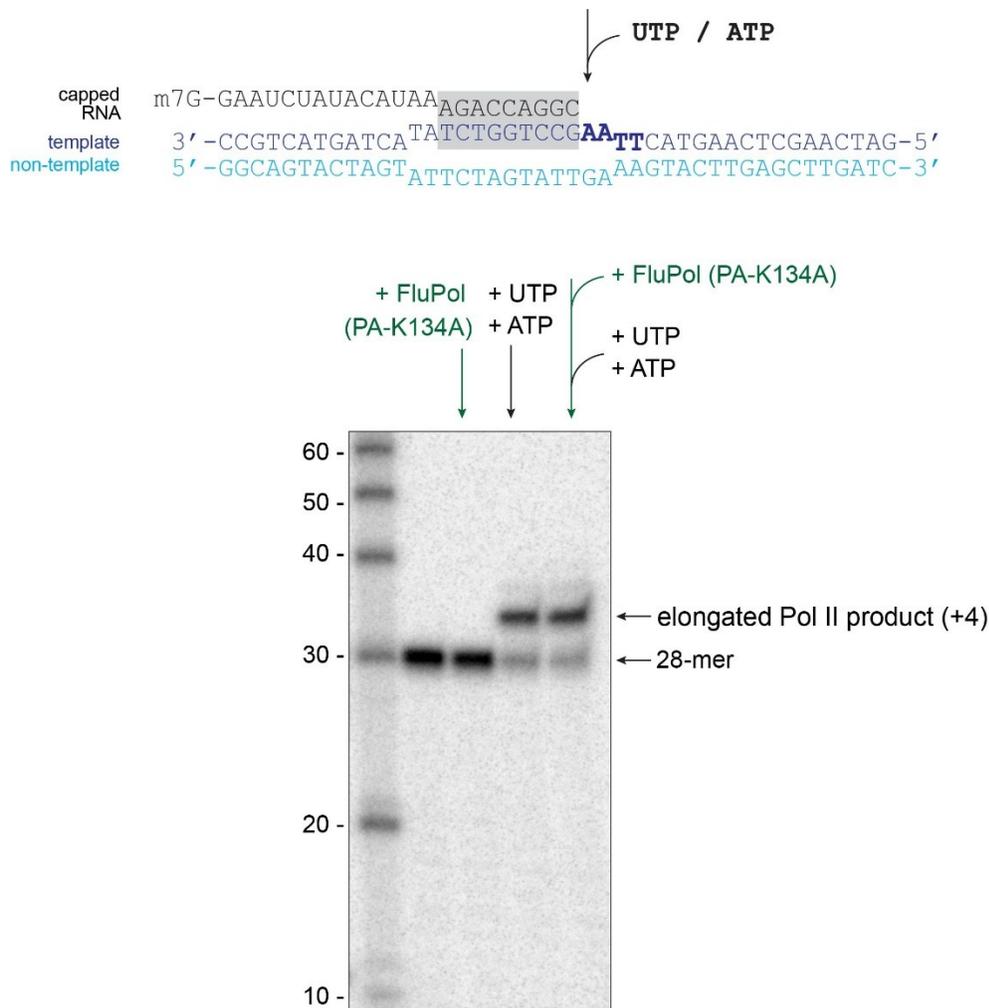
Recent *in vitro* studies on the capping enzyme (CE) showed that several distinct mechanisms exist for activating co-transcriptional capping: allosteric activation by Ser5-phosphorylated CTD binding, increased local concentration of CE in proximity of the RNA (tethering effect) and additional interactions with the core of Pol II. The enhancement of activity due to CTD binding that I observed using CTD-mimic peptides (see section 3.1.4) suggested that an allosteric activation plays a role in the RNA cleavage by FluPol. To investigate a possible tethering effect, I tested the cleavage reaction in the coupled transcription assay using two different lengths of RNA: 20 and 28 (Figure 42). The shorter RNA would not be long enough to exit on the Pol II surface and be cleaved by FluPol, while the 28-mer would be a good substrate, as shown in the previous section. Radiolabeled RNAs were either part of the transcription bubble (forming the DNA:RNA hybrid) or added in *trans*. Addition of a long enough CTD peptide to activate allosterically the polymerase (six-repeats, as determined in section 3.1.4) increased endonuclease cleavage of both short and long transcripts, either when they were part of the transcription bubble or when they were added in *trans*. The effect was enhanced when full-length CTD was present, in absence or presence of additional phosphorylation by P-TEFb (given that endogenous phosphorylation is always present). The short capped RNA (20-mer) had only residual cleavage when it was part of the elongating complex, confirming that it is not long enough to be recognized by FluPol. No significant difference was observed in the assays with full-length Pol II when RNA was part of the complex or added in *trans*. In conclusion, in this experimental setup I could not determine if tethering or additional interactions with the Pol II core play a role on FluPol cleavage of capped transcripts *in vitro*. However, in the case of infection, the tethering to Pol II is likely to be important for recruiting FluPol to the nascent transcript in order to compete with the nuclear cap-binding complex.



**Figure 42. Activation of FluPol endonuclease cleavage in the context of Pol II elongation complex.** Capped RNA substrates are part of a transcription bubble, unless indicated that capped RNA is added *in trans*. A representative gel is shown, with quantification from three independent experiments below each lane. The activity is normalized to the overall amount of input RNA for each individual reaction. Error bars indicate standard deviation from three independent experiments.

### 3.2.2.3. Effect of FluPol binding to capped RNA on Pol II transcription

I wanted to investigate further if binding of FluPol to capped RNA triggers dissociation of the transcript from the elongating Pol II complex. I incubated Pol II with the endonuclease-deficient mutant PA/K134A FluPol in presence of P-TEFb (in conditions to form a FluPol/Pol II complex) and then added nucleotides to follow elongation by Pol II. I did not observe a difference in Pol II elongation after incubation with FluPol, suggesting that no active ‘pulling’ of the RNA by the viral polymerase happens after binding, but rather the cleavage triggers dissociation from Pol II (Figure 43) and eventually further degradation of the decapped and trimmed transcripts.



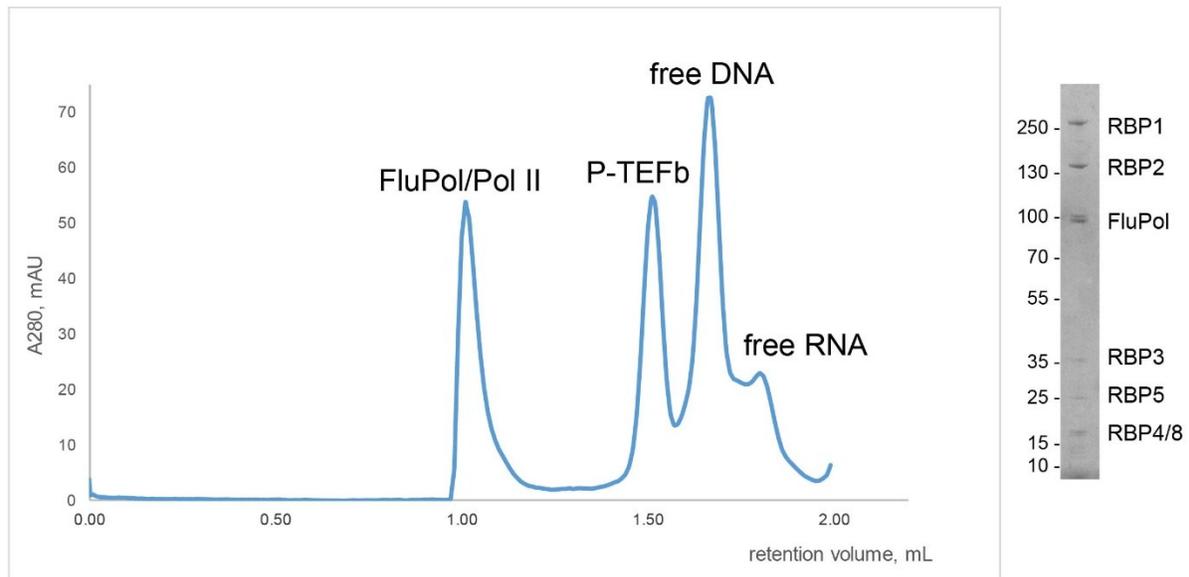
**Figure 43. Binding of FluPol to capped RNA is not enough to trigger its dissociation from elongating Pol II.** Elongation of the capped RNA by Pol II is not affected by the presence of pre-bound endonuclease-deficient FluPol (compare last two lanes). The scaffold used, allowing elongation by Pol II, is indicated above.

In conclusion, I could determine the optimal RNA length recognized by FluPol endonuclease in the context of the Pol II elongation complex (16 nt distance between Pol II active site and FluPol cleavage site). I showed that CTD-dependent allosteric activation of FluPol endonuclease activity is further enhanced by a full-length phosphorylated Pol II, but I could not show if additional interactions with the Pol II core played a role in positioning FluPol closer to the nascent transcript. Structural insights on the complex would help elucidate if additional contacts of FluPol with Pol II can occur. Finally, I observed that binding of FluPol to the emerging cap *per se* does not trigger dissociation of the RNA from elongating Pol II, suggesting that the FluPol-dependent cleavage must happen while the transcript is emerging from the Pol II exit channel. This highlights the functional relevance of the FluPol/Pol II complex for viral transcription. Therefore, I aimed to characterize the reconstituted cap-snatching complex structurally.

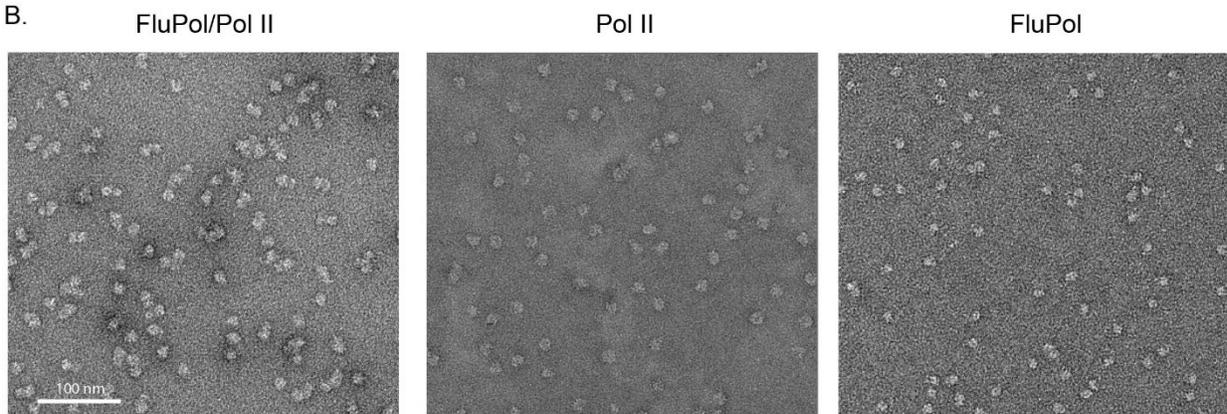
### **3.2.3. Structural studies on the FluPol/Pol II complex**

Size-exclusion chromatography experiments showed that the complex of the two polymerases is stable and homogenous. Furthermore, the functional assays confirmed that the two polymerases are transcriptionally coupled in the context of the complex. In order to understand the mechanistic details of the cap-snatching of the RNA emerging from Pol II, I aimed to structurally characterize this sample. I first assembled the complex as described in section 3.2.1.2, using an endonuclease-deficient mutant of FluPol (PA/K134A), so that the polymerase is still bound to the emerging RNA, but does not cleave it (see Figure 43). The 28-mer RNA was chosen for complex assembly, since I previously determined it to be an optimal length for FluPol recognition (discussed in section 3.2.2.1). I cross-linked the peak SEC fractions containing Pol II and FluPol and performed negative stain electron microscopy (Figure 44). The micrographs showed a significant proportion of ‘double’ particles that could be a complex of the two polymerases, which looked distinct from any of the polymerases alone (Figure 44B).

A.

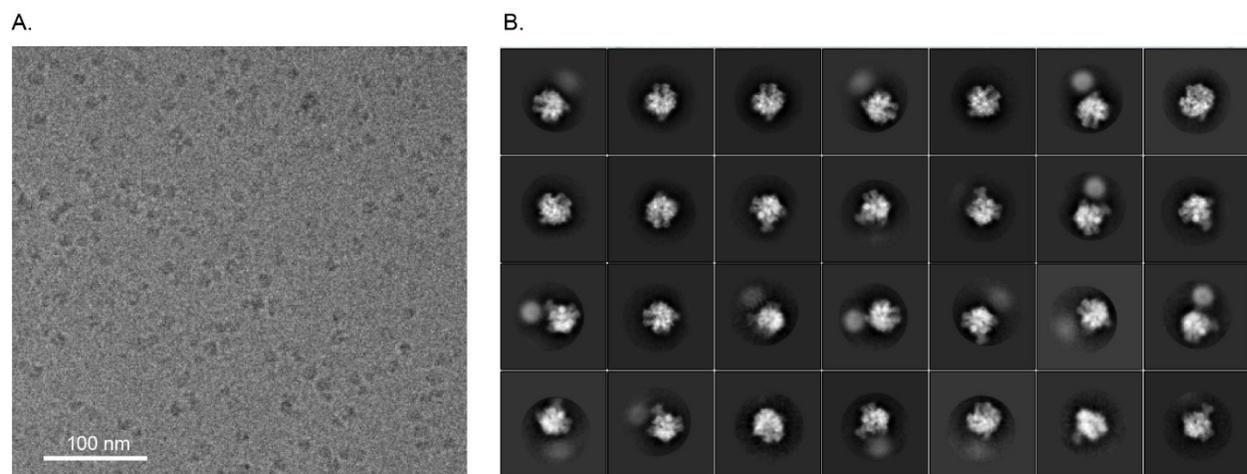


B.



**Figure 44. Negative stain electron microscopy of FluPol/Pol II complex.** (A) Size-exclusion chromatography of the complex, containing 28-mer RNA with a SDS-PAGE sample of the main peak. (B) Representative negative stain micrograph, containing ‘double’ particles (left panel). Pol II and FluPol micrographs are shown for comparison; the samples contain Pol II bound to a DNA/RNA scaffold and FluPol bound to vRNA promoter. *The micrographs of Pol II and FluPol alone were collected by Carrie Bernecky (Pol II) and Manikandan Karuppasamy (FluPol).*

The next step was to investigate the complex using cryo-electron microscopy (cryo-EM). Compared to negative stain microscopy, significantly higher particle concentrations are necessary to obtain good quality data in cryo-EM (see Materials and Methods, section 2.14 and 2.15). Aggregation effects were observed on the size-exclusion column at higher concentrations of the polymerase complexes. Therefore, the complex was purified by sedimentation on a sucrose gradient using the GraFix protocol <sup>157</sup>. The presence of the two polymerases in the respective gradient fractions was detected by dot blot using antibodies against the Strep-tag of FluPol and against Pol II SeP5 CTD (data not shown). Cryo-EM data was collected on this sample with a K2 direct detector (Gatan) on a Polara microscope (Figure 45). Initial 2D classification was performed on 110836 particles and after manual sorting of the 2D classes, 52692 particles seemed to contain particular features of Pol II (compared to 2D classification of Pol II alone or in elongation complexes <sup>68,156</sup>) (Figure 45B). Some of the classes showed some extra blurred density, which we hypothesized could be FluPol bound to Pol II. Selection of the classes that had supposedly the two polymerases yielded 24494 particles. As seen in the 2D classification, the position of the FluPol relative to Pol II (if Pol II is fixed based on some feature) varies in the different classes, so further 3D classification of this subset did not converge to a plausible model of the complex (data not shown).



**Figure 45. Cryo-EM data collection on FluPol/Pol II complex.** (A) Representative micrograph of the cryo-EM sample. (B) 2D classification of 52693 particles, selected for the presence of Pol II. Pol II was identified by comparison with 2D classes from published cryo-EM data on Pol II <sup>156</sup>.

Further efforts to obtain a stable complex suitable for structural characterization by cryo-EM did not show any substantial difference, so we suggest that the heterogeneity of the complex is due to the binding of FluPol to the flexible CTD of Pol II.

The results presented in this section showed that a functional cap-snatching complex between Pol II in an elongation complex with a capped RNA and FluPol could be assembled. The Ser5 phosphorylation of the CTD is essential for complex formation, which validates the results with CTD-mimic peptides presented in section 3.1. I showed that FluPol is able to recognize and cleave the pre-bound capped RNA and the optimal length of the capped RNA substrate in the elongating Pol II context for FluPol cleavage was determined to be 27-28 nt. Cryo-EM experiments suggest that the complex does not adopt a preferential conformation that is stable enough for further structural investigation. This prompted me to study the role of other factors for the recruitment of FluPol to the transcribing Pol II that could potentially stabilize the coupled-polymerases complex.

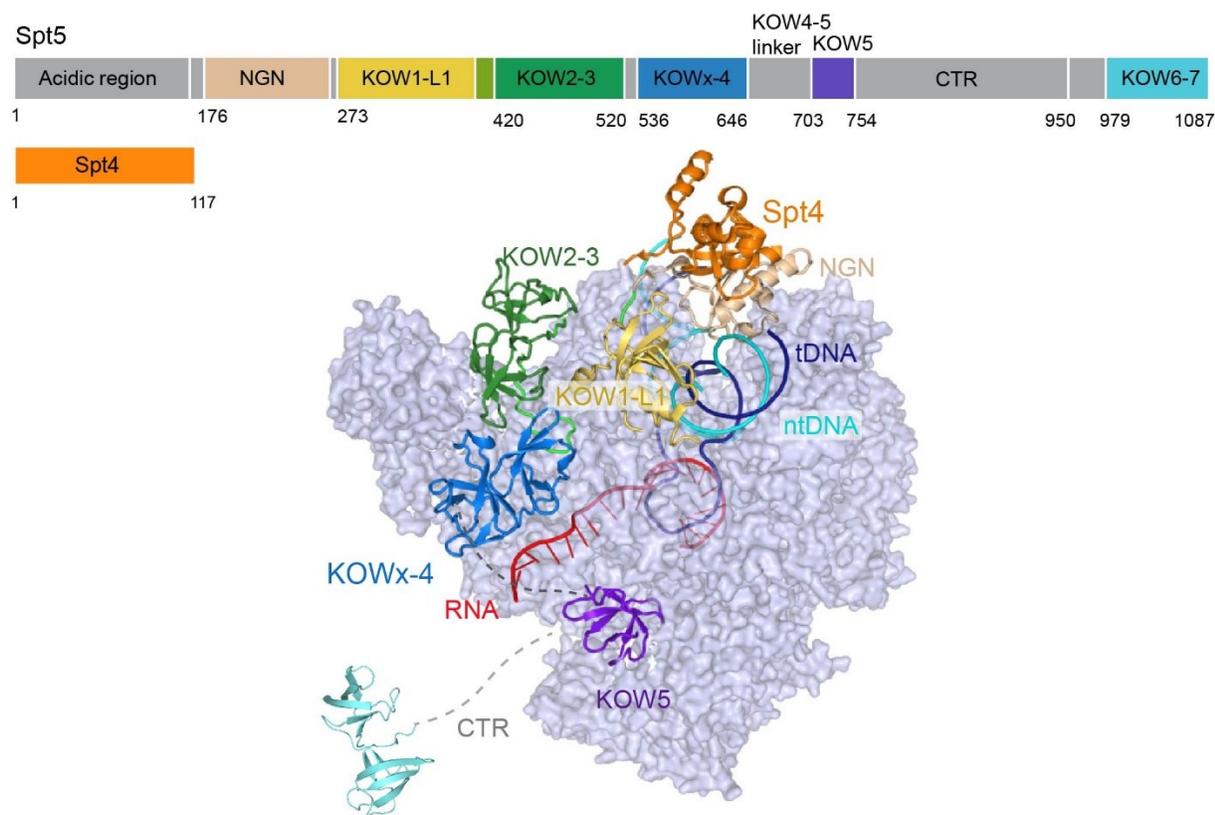
### **3.3. Interaction between influenza polymerase and cellular transcriptional regulators**

The recruitment of FluPol to active Pol II during the early stages of transcription, after capping but before productive elongation, prompted me to investigate possible interactions with factors involved in transcriptional pausing. Proteomics studies of influenza-infected cells have identified two factors involved in promoter-proximal pausing, DRB sensitivity-inducing factor (DSIF) and Tat-stimulatory factor 1 (Tat-SF1), as direct interactors of the trimeric FluPol <sup>98</sup>. Another study has also shown that Tat-SF1 has a stimulatory function on viral replication, even though it has been suggested to happen through interaction with NP, rather than the polymerase <sup>99</sup>. The two cellular factors are functionally related and are known to play a role in transcriptional pausing <sup>83</sup>. Moreover, one of the subunits of DSIF, Spt5, helps recruit the capping enzyme and enhances its activity <sup>158</sup>. The C-terminal region (CTR) of Spt5 plays a similar scaffolding role as the CTD of Pol II and is phosphorylated by P-TEFb during the transition to productive elongation <sup>71</sup>. Recently, several structures of paused Pol II have shown that DSIF is tightly bound to Pol II <sup>68,70,82</sup>. Based on the functional relevance and suggested physical link between FluPol and these factors, I decided to study their possible interaction with FluPol *in vitro*.

### 3.3.1. Association of FluPol with DRB sensitivity-inducing factor (DSIF)

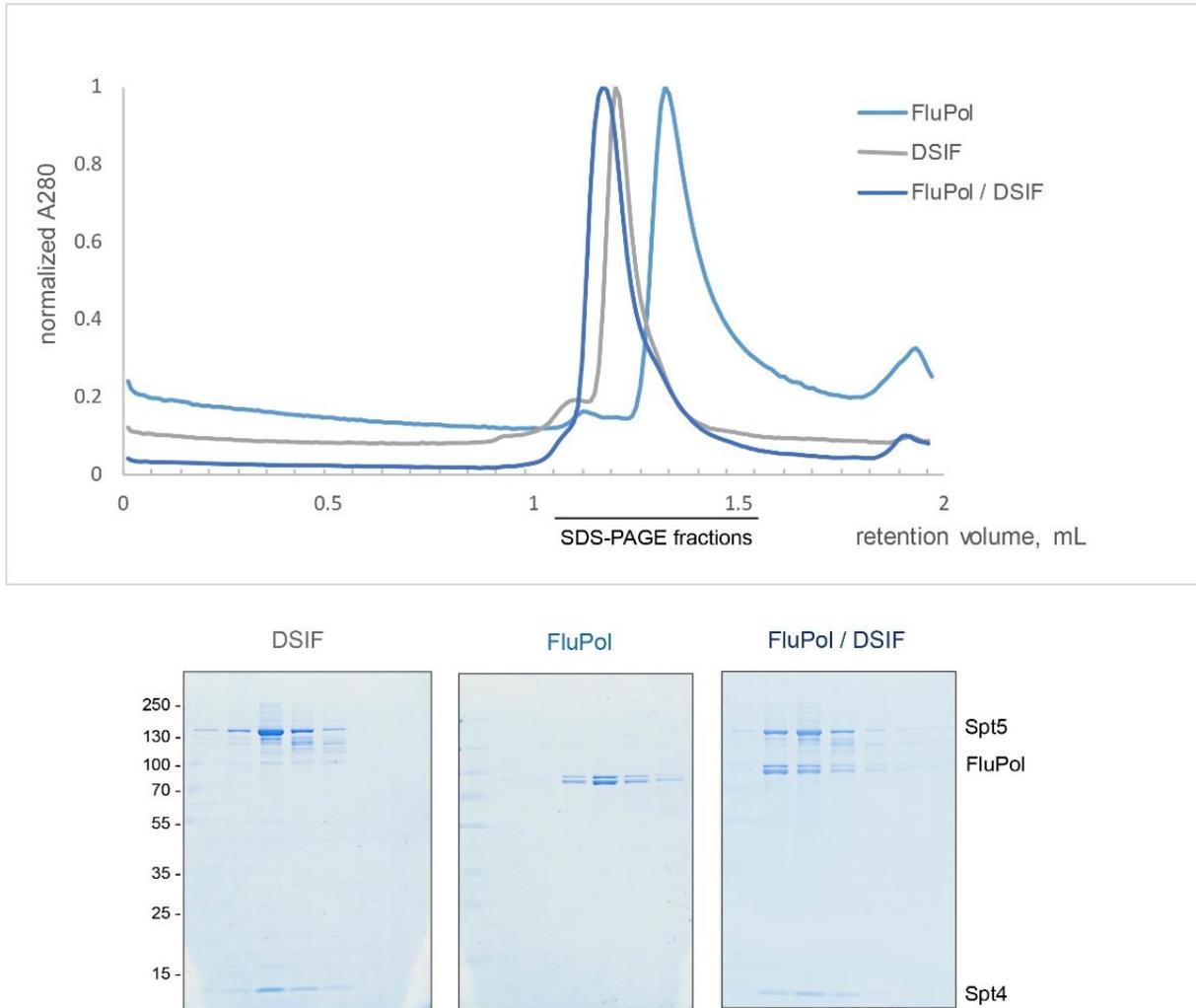
#### 3.3.1.1. Binding studies of FluPol with full-length DSIF

DSIF comprises a heterodimer of Spt5 and Spt4. Spt5 has a modular structure, containing an N-terminal acidic domain, NusG amino-terminal (NGN) domain, several KOW domains (Kyrpides, Ouzonis, Woese motifs) and a flexible C-terminal region (CTR) (Figure 46). Spt4 consists of a NGN-binding domain and a Zn-binding domain<sup>159</sup>. The structure of the DSIF bound to Pol II elongation complex shows intricate interactions with the polymerase and both DNA and RNA. Spt4, the NGN domain, KOW1-L1, KOW2-3, KOWx-4 and KOW5 domains of Spt5 bind to the surface of Pol II. KOWx-4 and KOW5 form a RNA clamp with the protruding nascent transcript<sup>68</sup>.



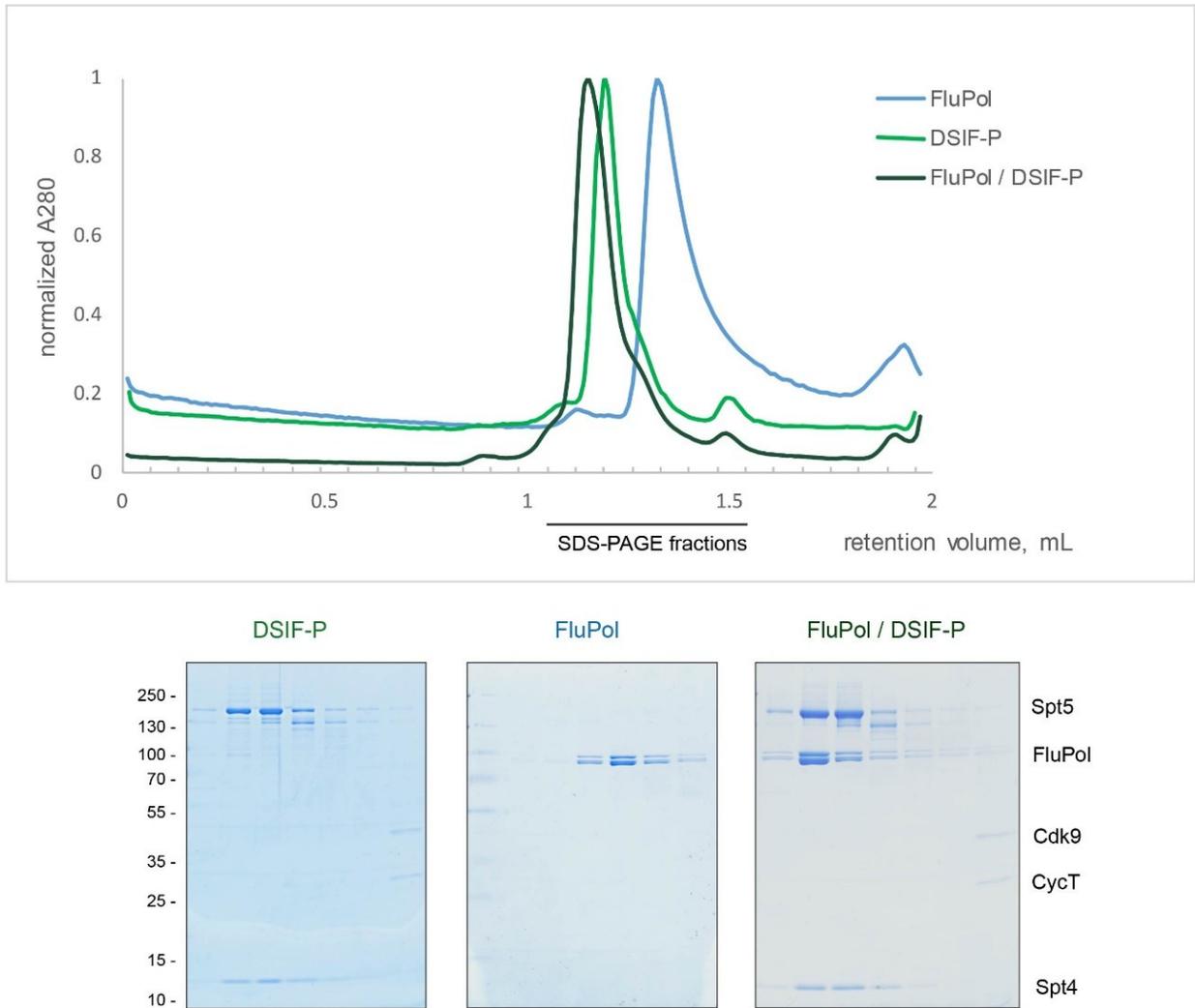
**Figure 46. Domain organization and structure of Pol II/DSIF elongation complex.** Schematic representation of the domain organization of DSIF, modified from<sup>68</sup> and structure of DSIF, bound to Pol II (lightblue), showing the modular structure of the different domains, color coded as above (PDB: 5OIK). The flexible CTR domain (light grey) and KOWx4 - 5 linker (dark grey) are indicated with dashed lines.

I purified recombinant heterodimeric DSIF expressed in *E.coli* and tested its interaction with purified v5'RNA-bound FluPol (A/H17N10 strain) by size-exclusion chromatography (SEC) (Figure 47). DSIF migrates at a lower retention volume than expected from its size, probably due to the long acidic tail at the N-terminus of Spt5, as previously reported<sup>160</sup>. When DSIF and FluPol were injected together, there was a clear shift compared to FluPol alone and SDS-PAGE analysis confirmed that they migrate together, suggesting a direct interaction.



**Figure 47. SEC analysis of the interaction between recombinant DSIF and v5'RNA-bound FluPol.** The same fractions from the three experiments (FluPol alone, DSIF alone and FluPol with DSIF) were loaded on SDS-PAGE for comparison, as indicated on the chromatogram. When the two proteins are mixed, the FluPol is found exclusively in the DSIF-bound fractions.

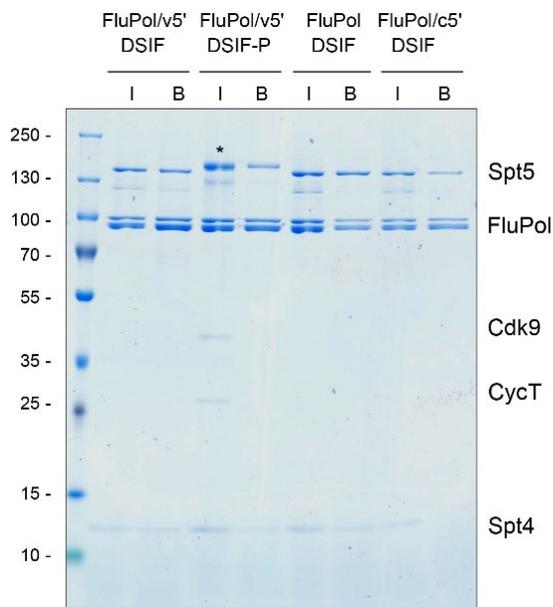
DSIF is one of the substrates of the P-TEFb kinase and its phosphorylation helps trigger the transition from pausing to elongation <sup>161</sup>. I wanted to check if FluPol also recognizes the phosphorylated form of DSIF, so I pre-incubated the complex with P-TEFb (Figure 48). SEC analysis showed that there is also an interaction of FluPol with phosphorylated DSIF (DSIF-P), which migrates slightly higher than the non-phosphorylated form <sup>162</sup> (see fractions of DSIF and DSIF-P on Figure 47 and 48 and first four lanes on Figure 49).



**Figure 48. SEC analysis of the interaction between FluPol and DSIF, phosphorylated by P-TEFb.** Similarly to the non-phosphorylated form of DSIF, FluPol associates with the P-TEFb-phosphorylated DSIF, judging by the peak shift and SDS-PAGE analysis (the fractions for the SDS-PAGE are indicated below the chromatogram, the same fractions are shown for all gels).

To validate the interaction, I performed a pull-down experiment using the Strep-tag on the C-terminus of FluPol PB2 subunit. Both phosphorylated and non-phosphorylated forms of DSIF co-eluted with v5'RNA-bound FluPol, confirming the SEC analysis (Figure 49, first 4 lanes). FluPol in absence of RNA or bound to the c5'RNA also showed an interaction. However, due to their increased flexibility, these conformations of FluPol have lower retention volumes on SEC<sup>25</sup> and were not analyzed by this method because they run at very similar volumes to DSIF.

These experiments showed that there is a direct interaction of FluPol with the DSIF, which is independent of the phosphorylation state of DSIF and of the conformation of FluPol.

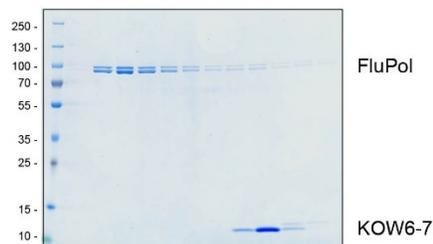
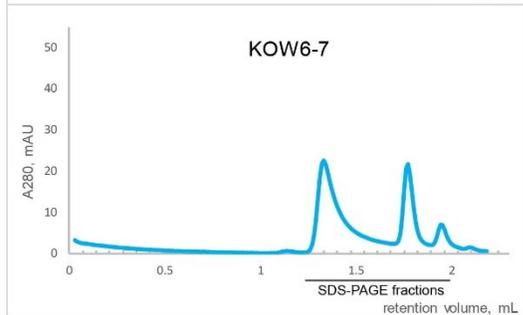
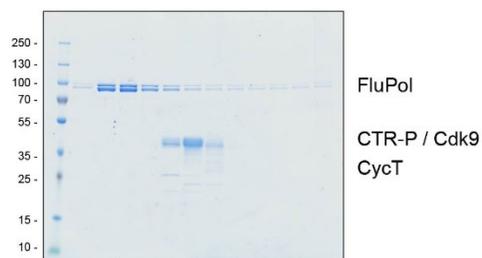
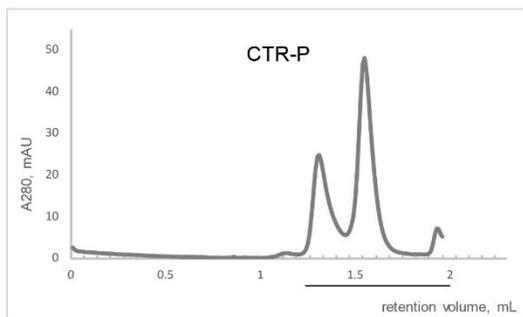
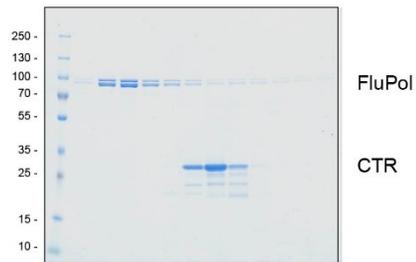
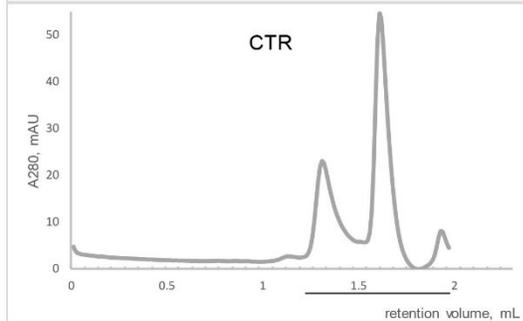
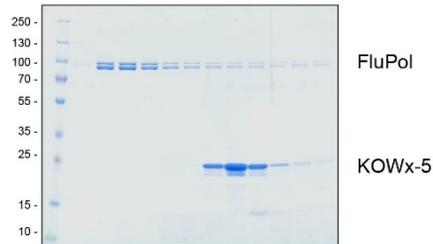
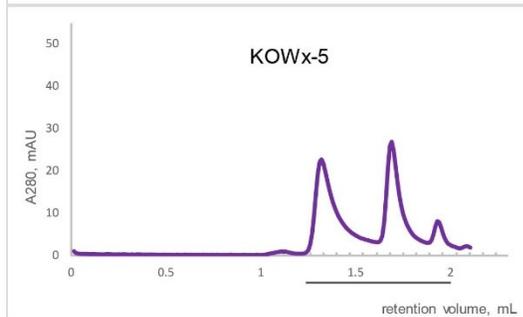
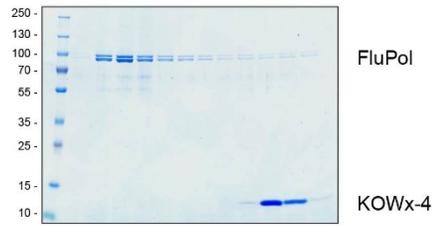
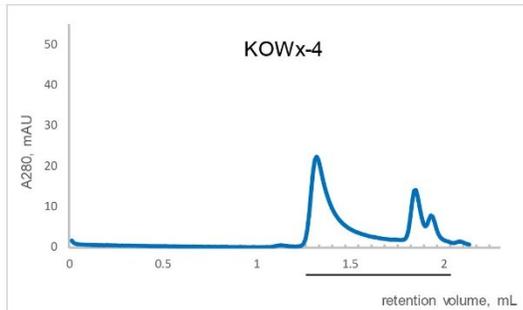
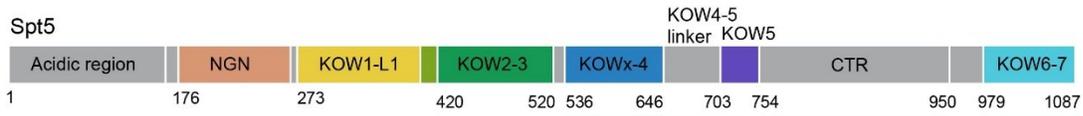


**Figure 49. Pull-down experiment on the interaction between FluPol and DSIF.** Input (I) and bound (B) fractions are shown. Different conformations of FluPol (v5' bound, no RNA or c5'-bound) as well as the effect of DSIF phosphorylation by P-TEFb were assayed. A star indicates the phosphorylated form of DSIF, which migrates higher than the non-phosphorylated one.

### 3.3.1.2. Interaction of FluPol with separate domains of DSIF

I wanted to identify which region of DSIF is responsible for the association with FluPol. Based on the structure of DSIF, bound to Pol II elongation complex, I first tested the domains that are in closest proximity to the RNA exit channel, which are the KOW<sub>x</sub>-4, KOW5 and the linker between them, the flexible CTR domain (which is known to have a scaffolding role similar to Pol II CTD) and the KOW6-7 domain, located after the CTR. (*KOW<sub>x</sub>-4 and KOW6-7 proteins, as well as the construct for KOW<sub>x</sub>-5 were provided by Carrie Bernecky*). I also phosphorylated the purified CTR domain with P-TEFb (the phosphorylation was confirmed by a significant shift in the SDS-PAGE of the CTR). I did not observe interaction with FluPol of any of the domains, suggesting that the binding site might be located on the N-terminal region of the Spt5 or that the affinities of the separate domains are lower than the affinity to the full-length complex, but this has not been addressed yet (Figure 50).

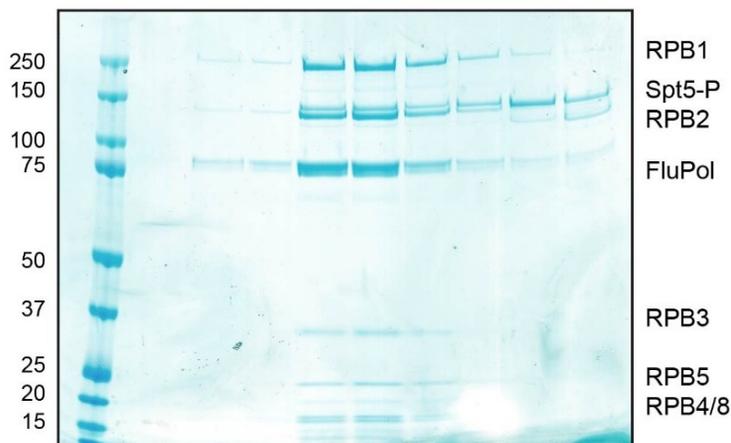
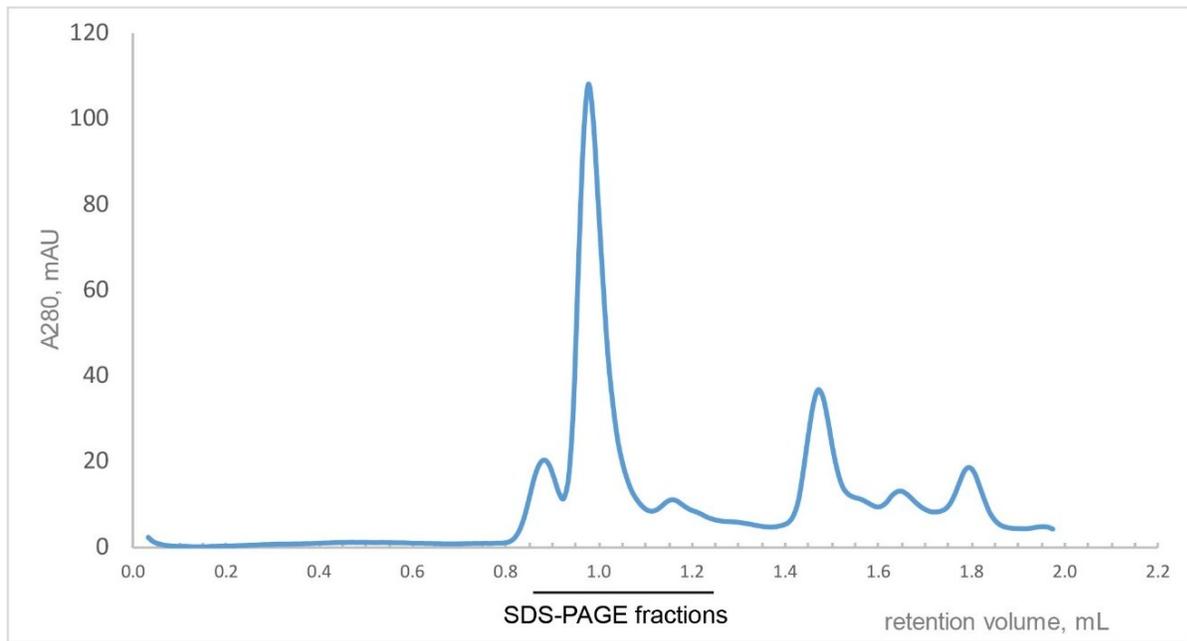
Spt5



**Figure 50. Interaction studies by SEC of different DSIF domains with FluPol.** The domain organization, as described in <sup>68</sup> is shown above. Lines with corresponding color codes indicate the borders of the constructs. No co-elution was observed for any of the tested domains.

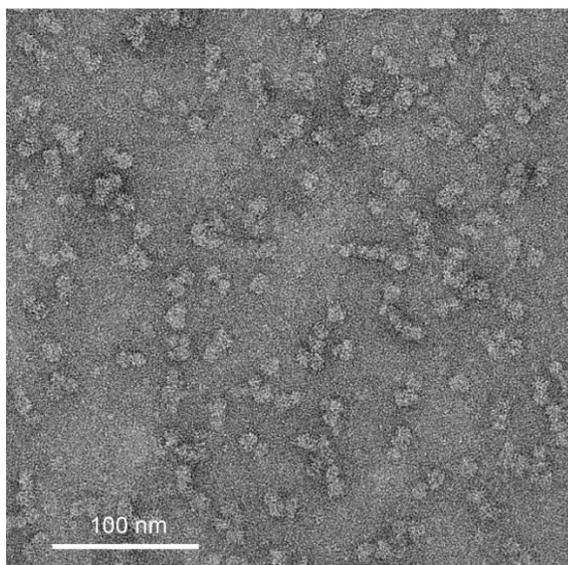
### **3.3.1.3. Assembly of a ternary complex of elongating Pol II, FluPol and DSIF**

To investigate the effect of DSIF on the interaction of FluPol in the context of promoter-proximal pausing, I next tried to assemble a complex with DSIF, bound to the Pol II elongation complex, and FluPol. I pre-formed the elongation complex (as described in section 3.2.1) and I added P-TEFb, which phosphorylates Pol II CTD and DSIF and analyzed the complex by SEC. P-TEFb is necessary for the association of FluPol to Pol II (as shown in section 3.2.1.3), and as described previously, FluPol binding to DSIF is independent of its phosphorylation state (section 3.3.1.1). To avoid cleavage of the capped RNA, an endonuclease-deficient FluPol mutant (PA/K134A) was used (as described for the FluPol/Pol II complex assembly (section 3.2)). A single peak, containing Pol II, FluPol and DSIF was observed (Figure 51).



**Figure 51. SEC of ternary complex of elongating Pol II, FluPol and DSIF-P.** Spt5-P migrates slightly higher than RPB2. Both FluPol and Spt5-P are observed in the Pol II-bound fractions on the SDS-PAGE. Spt4 co-elutes with Spt5 (see Figures 47-49), but is difficult to distinguish on the SDS-PAGE from the small subunits of Pol II, which migrate at similar molecular weights (RPB6, 9, 11).

This complex consisting of FluPol, Pol II and DSIF-P was analyzed by negative staining electron microscopy (Figure 52). As seen for the FluPol/Pol II complex (see Figure 44), the particles formed elongated shapes indicating the close interaction of the two polymerases. Unfortunately, further trials to characterize the complex structurally by cryo-EM were impeded by the sample heterogeneity.

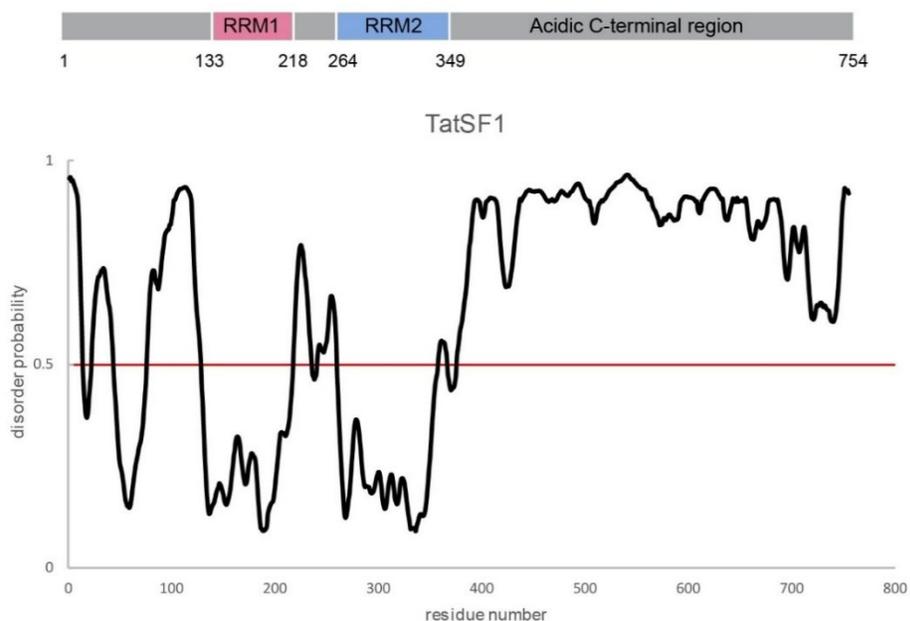


**Figure 52. Micrograph of negative stain EM on the FluPol/Pol II/DSIF-P complex.** Similar double particles as for the FluPol/Pol II complex were observed.

In conclusion, the presented data shows that FluPol directly interacts with the elongation factor DSIF. The domain mapping did not identify which domains of DSIF are involved in the binding, suggesting that the interaction might be with the N-terminal region of Spt5 and/or Spt4, but this was not experimentally addressed yet. A ternary complex of FluPol, elongating Pol II and DSIF-P could be formed, showing that the binding of DSIF and Pol II to FluPol is not mutually exclusive. Therefore, DSIF might act as an additional bridging factor, stabilizing the FluPol/Pol II complex. However, the complexity of the complex formation regarding the phosphorylation states of the components requires further studies to find optimal conditions for structural characterization. It is known that phosphorylation of DSIF alters its conformation on Pol II <sup>82</sup> and this could affect complex formation with FluPol, even though DSIF-P still showed interaction with FluPol. Overall, these results suggest that DSIF is a promising factor for further structural investigation of the cap-snatching mechanism.

### 3.3.2. Association of FluPol with Tat-SF1

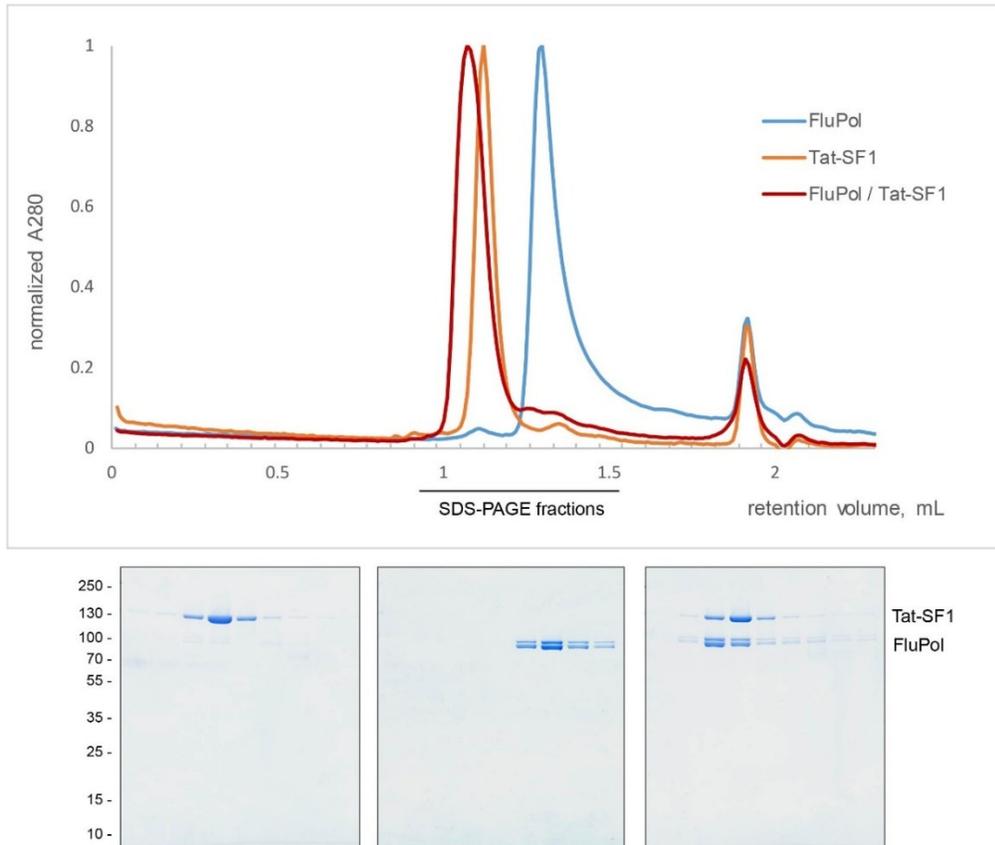
Tat-SF1 has been identified as a cofactor for HIV Tat in transcriptional control <sup>163</sup>. Its function is not well understood, except that it plays a role in elongation, and no structural data is available for the full-length protein. Disorder prediction suggests that the C-terminal region is highly unstructured (Figure 53), while the N-terminal region contains two RRM domains (the structure of the second RRM domain has been solved, PDB ID: 2DIT). Nevertheless, it is known that Tat-SF1 is functionally related to DSIF and P-TEFb and plays a role in the transition from pausing to elongation <sup>83,84</sup>. Besides, proteomics studies have identified Tat-SF1 as an interactor of FluPol <sup>98</sup> and as a factor, involved in the regulation of FluPol functions <sup>99</sup> making it an interesting candidate for further study.



**Figure 53. Disorder prediction of Tat-SF1 with the domain organization shown above.** The plot is generated with the Protein DisOrder prediction System (PrDOS) server, which predicts natively disordered regions from the amino acid sequence (<http://prdos.hgc.jp/cgi-bin/top.cgi>).

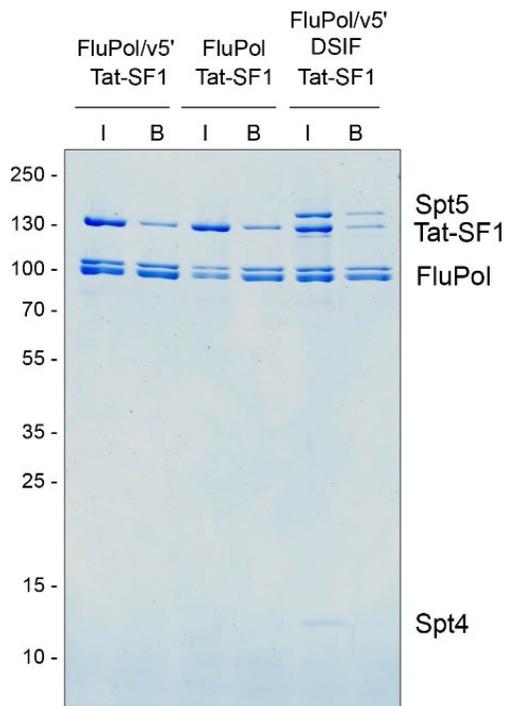
### 3.3.2.1. Interaction of FluPol with Tat-SF1

Recombinant Tat-SF1 was purified from insect cells expression (*construct provided by Seychelle Vos*). The interaction of FluPol with Tat-SF1 was analyzed by SEC. Similar to DSIF, because of its long acidic domain, Tat-SF1 migrated at lower retention volumes in SEC. Interaction with FluPol was assessed by the shift of the free FluPol peak (Figure 54) which showed that FluPol in its v5'RNA-bound form associated with Tat-SF1.



**Figure 54. SEC analysis of Tat-SF1 interaction with FluPol.** The small peak at around 1.8 mL retention volume is the excess vRNA. FluPol is exclusively in the Tat-SF1-bound fractions, confirmed by SDS-PAGE analysis (the corresponding fractions of the three gels are indicated under the chromatogram).

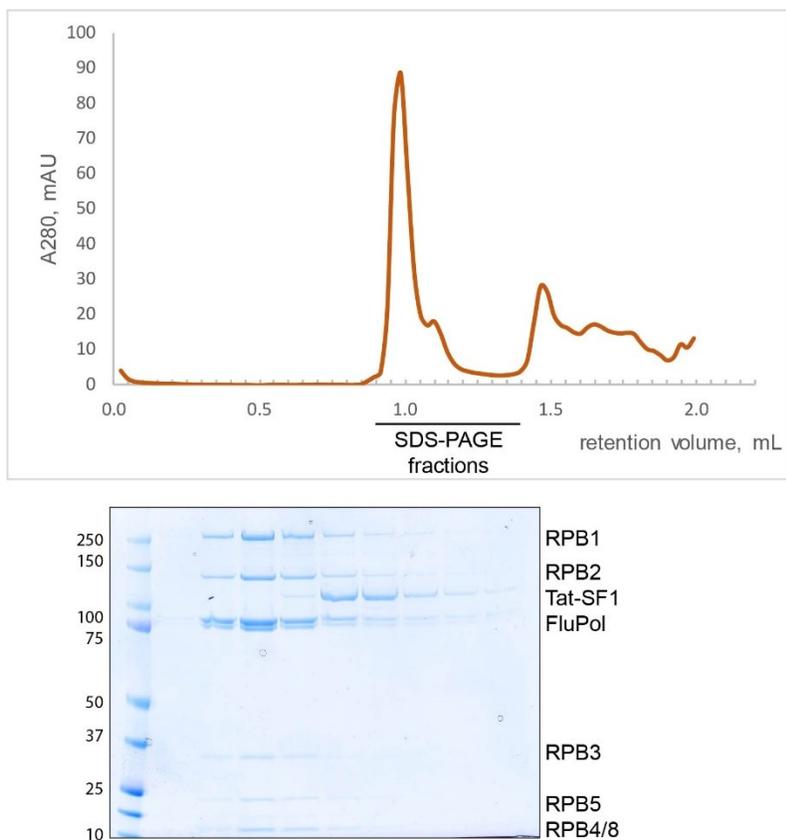
I confirmed the interaction of FluPol with Tat-SF1 with a pull-down experiment (Figure 55). Tat-SF1 also showed binding to apo-FluPol, hinting that it does not only recognize the transcriptionally active conformation. To check if DSIF and Tat-SF1 bind to different sites on FluPol or compete for the same site, FluPol was incubated with both proteins and a pull-down experiment was performed using the Strep-tag of FluPol. The three proteins were observed in the bound fraction, suggesting that Tat-SF1 and DSIF could have independent binding sites on FluPol (Figure 55, last lane). Due to the very close retention volumes of the proteins alone, SEC analysis was not performed on the ternary complex.



**Figure 55. Pull-down experiments of FluPol with Tat-SF1 and a ternary complex formation of FluPol, Tat-SF1 and DSIF.** Promoter-bound and apo-FluPol associate with Tat-SF1 (first four lanes). Tat-SF1 and DSIF both elute with the bound fraction, suggesting that they can simultaneously bind to FluPol.

### 3.3.2.2. Assembly of a ternary complex consisting of FluPol, elongating Pol II and Tat-SF1

I then attempted to assemble a complex with elongating Pol II, FluPol and Tat-SF1. Tat-SF1 did not co-elute with the Pol II-FluPol complex in SEC (Figure 56), therefore a ternary complex could not be formed. I have shown that FluPol can bind to Pol II and Tat-SF1 separately, however, in conditions when both proteins are present, FluPol associated with Pol II and not with Tat-SF1. It is possible that Pol II and Tat-SF1 compete for the same binding site on FluPol. Alternatively, steric hindrance could preclude similar binding of the two factors.

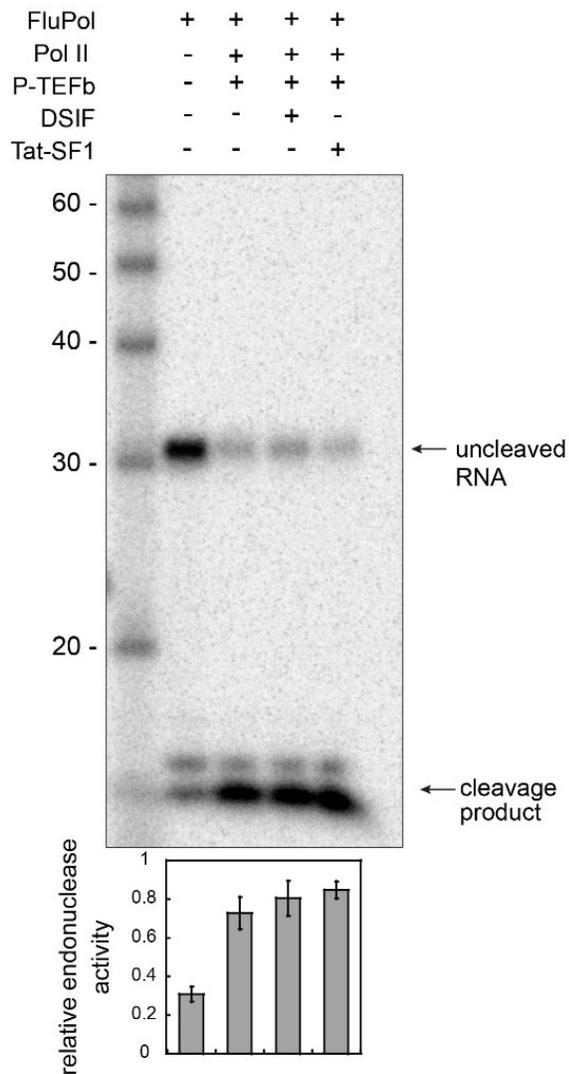


**Figure 56. Ternary complex assembly with FluPol, Pol II and Tat-SF1 analysed by SEC.** SDS-PAGE shows that Tat-SF1 does not co-elute with the FluPol/Pol II complex.

### **3.3.3. Effect of DSIF and Tat-SF1 on FluPol endonuclease activity in the context of elongating Pol II**

DSIF and Tat-SF1 are involved in the regulation of transcription. Therefore, I was interested in studying their role in cap-snatching in the context of a transcribing Pol II. I used the assay that I have established for measuring the endonuclease activity of FluPol in the context of Pol II elongation complex (section 3.2.2). Addition of either DSIF or Tat-SF1 did not have a significant effect on the cleavage of the emerging RNA, compared to the cleavage induced by phosphorylated Pol II alone (Figure 57). However, even though DSIF phosphorylated by P-TEFb can still associate to FluPol, in the context of the Pol II elongation complex DSIF-P might not be functionally relevant for the viral polymerase. It is also possible that DSIF has a kinetic effect on the FluPol cap-snatching reaction, but does not affect the amount of end product of the reaction. The assays presented above are only measuring the steady state and a more detailed kinetic experiment could address this question. Additionally, an assay using pre-phosphorylated Pol II, which would not necessitate the presence of P-TEFb and therefore DSIF would not be phosphorylated, can help investigating the possible effects of DSIF on cap-snatching.

On the other hand, Tat-SF1, FluPol and Pol II did not form a ternary complex, probably due to competition between Pol II and Tat-SF1 for FluPol binding. Therefore, in the experimental conditions used in this assay (which are very similar to the conditions for assembling the complex on Figure 56) Tat-SF1 probably cannot associate with FluPol and therefore cannot exercise an effect on its activity.

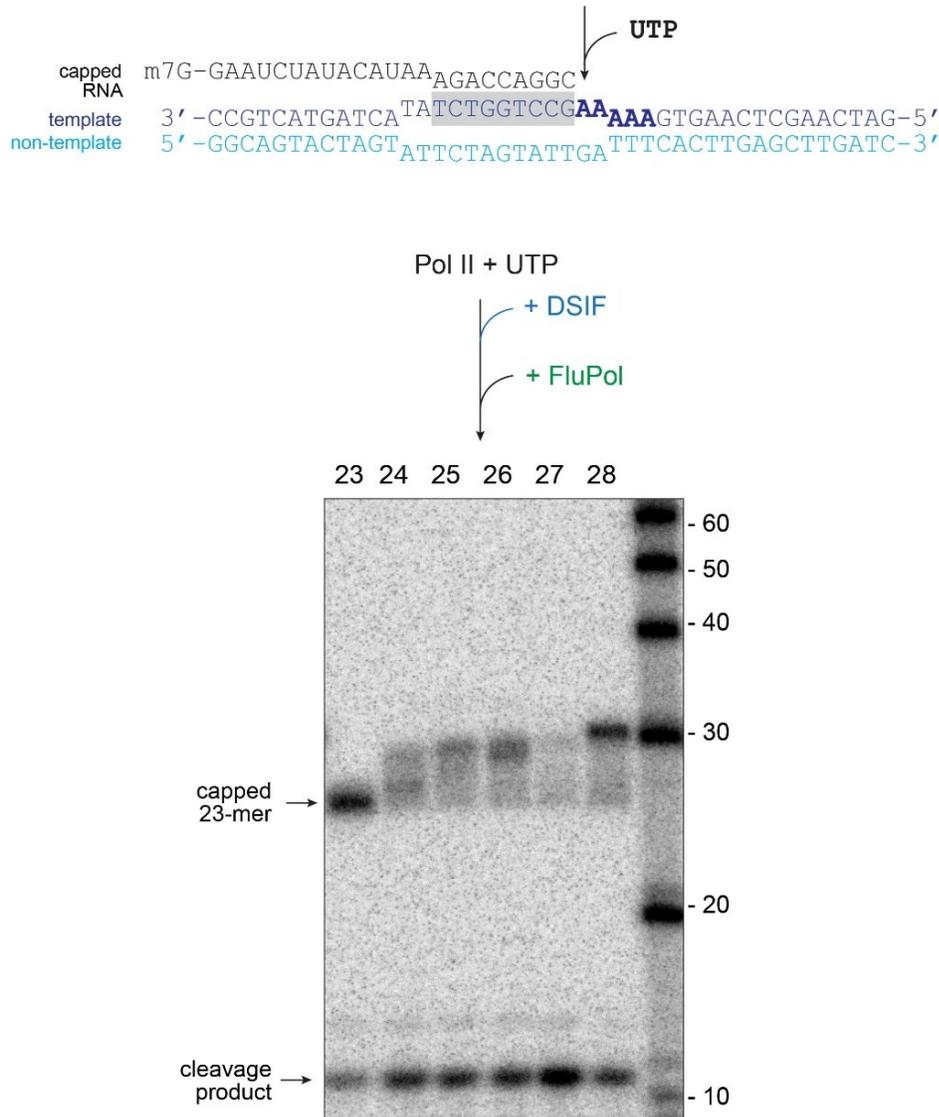


**Figure 57. Effect of DSIF and Tat-SF1 on RNA cleavage by FluPol in the elongating Pol II context.** DSIF and Tat-SF1 were pre-bound to Pol II before addition of FluPol. The assay follows endonucleolytic activity in a Pol II-dependent context. A representative gel is shown with quantification of three independent experiments (below). The activity is normalized to the total RNA for each reaction. Error bars indicate standard deviation of the three experiments.

Recent structural data on the Pol II elongation complexes containing DSIF<sup>68,82</sup> shows that in the paused state, the KOW<sub>x-4</sub> and KOW<sub>5</sub> domains of DSIF form a clamp on the exiting RNA, protecting several additional nucleotides after the end of the Pol II exit tunnel (as previously shown<sup>162</sup>). In the elongating form, when DSIF is phosphorylated and the PAF complex is bound, this clamp is released and one of the possible factors for this dissociation is the phosphorylation

of the KOW<sub>x</sub>-4 – KOW5 linker by P-TEFb. Therefore, I wanted to test if FluPol recognizes the same length of RNA exiting Pol II when the DSIF is pre-bound to the elongation complex.

I did not observe a difference in the optimal length of capped RNA cleaved by FluPol (Figure 58), compared to the same assay in the absence of DSIF. This suggests that the KOW<sub>x</sub>-4 – KOW5 domains no longer protect the RNA in the presence of P-TEFb and FluPol can cleave at the same position as in the absence of DSIF (compare to Figure 40).



**Figure 58. Optimal length of Pol II-pre-bound RNA cleaved by FluPol in presence of DSIF.** No difference of the recognized length was observed. The setup of the experiment is the same as explained in section 3.2.2.1, and data in absence of DSIF can be seen on Figure 40. A DNA scaffold containing a template used for addition of 5 Us is shown above for illustration, all used scaffolds are shown on Figure 38.

With my work on additional transcription factors that influence the FluPol/Pol II complex, I have shown that there is a direct interaction between FluPol and the elongation factors DSIF and Tat-SF1. Together with the data showing an interaction with the SeP5 CTD, this suggests that FluPol might associate with a paused Pol II complex. Moreover, I could form a ternary complex containing FluPol, DSIF and Tat-SF1, showing that the two factors might bind to separate sites on FluPol. FluPol also associates with a Pol II elongation complex with DSIF, but not with Tat-SF1. I hypothesize that there could be a competition between Pol II and Tat-SF1 for binding to FluPol, but more detailed biochemical and structural insights into the Tat-SF1 interaction with FluPol are required.

I was not able to identify the FluPol-interacting domains of DSIF, which remains an object for further studies. Also, I could not observe any effect on DSIF or Tat-SF1 on the cap-snatching activity in the Pol II transcription context. Further investigation of the role of these host factors and the recruitment mechanism of FluPol to the paused Pol II complex are necessary to understand the cellular context of the cap-snatching process.

## **4. Discussion**

## Résumé en Français

Dans la cellule infectée la polymérase de la grippe interagit avec l'ARN polymérase II pour se procurer des ARN coiffés nécessaires à sa propre transcription. Les structures présentées ici des polymérase d'influenza A et B en complexe avec le domaine C-terminal (CTD) de l'ARN polymérase II montrent les bases structurales de cette interaction et expliquent comment la polymérase virale peut distinguer les différents états phosphorylés du domaine CTD. Une comparaison entre les structures des polymérases de différents types (A, B et C) permet de spéculer sur le mode de recrutement de la polymérase de la grippe à proximité de la sortie de l'ARN coiffé par des effets d'avidité. De plus, nous avons observé des effets allostériques sur les activités enzymatiques de la polymérase d'influenza, induits par la fixation du domaine CTD qu'on suggère être causée par une stabilisation d'une conformation active pour la transcription. L'analyse des virus récombinants qui portent des mutations ne leur permettant pas d'interagir avec l'ARN polymérase II, a montré que ces virus sont fortement atténués et génétiquement instables, ce qui suggère l'essentialité de l'interaction. Nous avons aussi identifié des mutations compensatoires qui permettent à ces virus de se multiplier et on peut expliquer ces effets compensatoires sur la base des structures obtenues. Les données présentées ici suggèrent aussi un mécanisme de régulation entre transcription et réplication par la polymérase virale par une compétition pour l'accès aux ARN coiffés entre les polymérases actives qui font partie d'un complexe ribonucléoprotéique et les polymérases virales nouvellement synthétisés, qui ne sont pas fixées au promoteur viral et sont donc inactives.

La reconstitution d'un complexe actif de cap-snatching a été l'objectif de la deuxième partie. Nous avons montré quels sont les éléments nécessaires à la formation d'un tel complexe, la phosphorylation Ser5 du domaine CTD, la longueur de l'ARN coiffé de 27-28 nucléotides et une stoechiométrie proche de 1:1. En même temps on a aussi observé que la polymérase virale est activée allostériquement par l'interaction d'une ARN polymérase II phosphorylée par la kinase P-TEFb *in vitro*, mais *in vivo*, cette interaction peut être dépendante d'autres kinases, telles que TFIIH. Les interactions complémentaires avec les facteurs d'élongation DSIF et Tat-SF1 peuvent aussi aider à recruter la polymérase virale dans le contexte de l'ARN polymérase II en cours d'élongation.

The cap-snatching mechanism by influenza polymerase is essential for the synthesis of viral mRNAs that can be processed, exported and translated by the host cell. The functional coupling of the viral and cellular transcription machineries had been shown to be mediated through a physical association of FluPol with the Ser5-phosphorylated CTD of Pol II, but at the inception of this thesis, no detailed biochemical characterization of this interaction was available. Therefore, I sought to investigate the molecular details of the binding of FluPol to the CTD of Pol II.

Based on a functional comparison with the capping enzyme, which acts co-transcriptionally to process the transcript emerging from Pol II, it can be hypothesized that FluPol exploits similar mechanism to be brought in vicinity to its capped RNA substrate. Two models have been proposed to be driving co-transcriptional capping: a tethering model, which claims that the capping enzyme is recruited to the nascent RNA by the interaction with the phosphorylated CTD, and an allosteric activation model, according to which binding of the SeP5-CTD peptide enhances the capping activity <sup>164</sup>. These two effects enhance the capping reaction in the context of the emerging transcript on the surface of Pol II. In my thesis work, I studied the mechanisms of recruitment of FluPol to the capped RNA in the transcribing Pol II context. In the first part, I investigated how FluPol binds to CTD peptides and characterized this interaction and its effect on FluPol activities. In the second part, I aimed to reconstitute a functional cap-snatching complex and study the recruitment of FluPol to the transcribing Pol II.

#### 4.1. Recruitment of FluPol to transcribing Pol II by interaction with the CTD

In the first part of my thesis I characterized the interaction between FluPol and peptides mimicking the CTD of different lengths (two and four repeats) and phosphorylations (no phosphorylation, SeP2, SeP5). The binding studies suggested a strong preference for phosphorylated peptides. However, using two-heptad repeat peptides, I could not demonstrate a preference for SeP5 compared to SeP2 (Figure 7A and 8B), contrary to what was previously shown. The co-crystal structures of FluA and FluB polymerases in complex with SeP5 CTD-mimic peptides provided a basis for understanding the observed specificity. For both FluA and FluB, there is one conserved binding site on the PA-C domain (denoted site 1), where SeP5 is recognized by two basic residues (K630 and R633 in A/H17N10). Sequence analysis shows that these residues are conserved in influenza A and B strains. In the bound conformation, the Ser2 in site 1 points away from the protein, excluding a possibility that it could be specifically recognized, but could be bound as SeP2 (Figure 11). In the structure of FluA polymerase with the CTD, a second binding site on PA-C makes extensive interactions with another part of the CTD peptide (site 2). The SeP5 at this site is also bound by two basic residues (K289 and R449) absolutely conserved among FluA strains. At this site, the Ser2 points towards the SeP5, showing that a steric clash would occur if both residues are simultaneously phosphorylated (Figure 13). Both Ser7 observed in this region are pointing away from the protein.

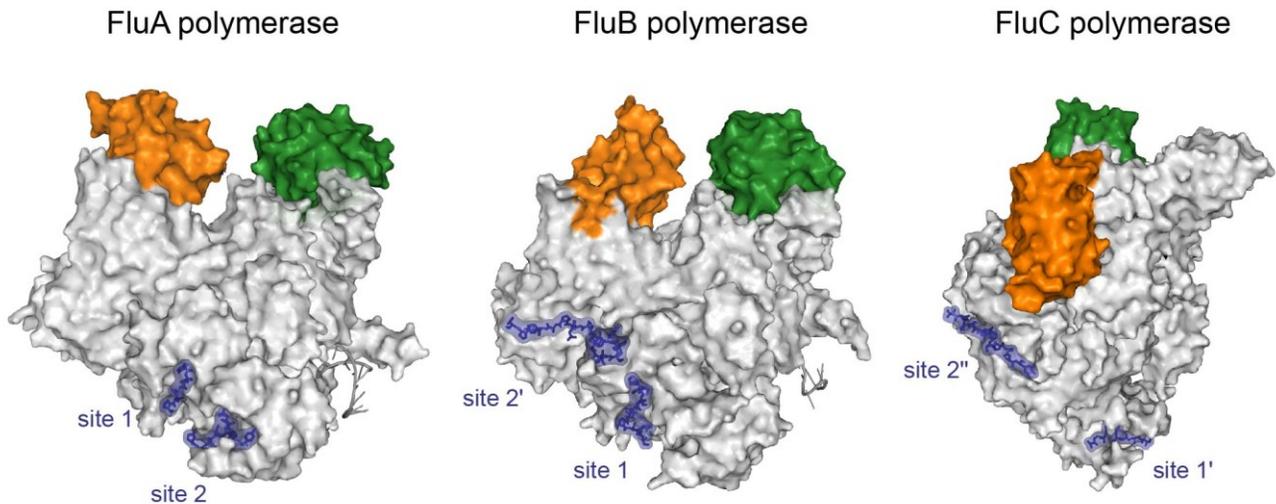
The co-crystal structure of FluB polymerase with CTD peptide revealed a different binding site (site 2'), extending in the opposite direction of the site 2 observed in FluA and reaching the PB2 627 domain. A helical turn of the peptide in this region brings the Ser2 and Ser5 of the same repeat in a close proximity, and similarly to the FluA site 2, excludes a possibility that both residues are phosphorylated at the same time (Figure 17). In the FluB structure, two of the observed Ser7 residues point towards the SeP5 and would clash if they were phosphorylated, and one is pointing away from the polymerase.

Our data provide a structural basis for the recognition of the Ser5-phosphorylated form of CTD. The SeP5-binding pockets are highly conserved between influenza A and/or B strains, implying a functional relevance for this recognition for the virus. The observed peptide conformations either exclude a Ser2 phosphorylation (FluA site 2 and FluB site 2') or suggest that a SeP2 could simply not form specific interactions (site 1). Similarly for SeP7, it would either not be recognized or

would clash with the SeP5. It cannot be excluded that a conformational rearrangement of the flexible CTD could reposition Ser2 or Ser7 so they occupy the SeP5-binding pocket. This could explain the similar affinities that I observed for two-repeat phosphorylated peptides, which are likely to be more flexible. Nevertheless, this kind of structural rearrangement has not been commonly observed and there is only one available structure of a CTD phosphatase (Ssu72) that can recognize both SeP5 and SeP7 in the same binding pocket <sup>146</sup>.

Recent in-depth analysis of the CTD phosphorylation pattern by mass-spectrometry showed that double phosphorylations within the same repeat are not common <sup>165</sup>. The same study also shows that SeP2 and SeP5 are the prevalent phosphorylations. Therefore, the specific recognition of SeP5, and not of SeP2 and SeP7, is physiologically relevant for recruiting FluPol to the initiating Pol II.

The two structures presented in this work, together with the recently published FluC polymerase in complex with the CTD <sup>151</sup> show that the three polymerases have each two separate binding sites for the CTD peptides (Figure 59). In none of the structures is density for the peptide between the two tightly bound regions observed. It is possible that the polymerase may bind to two remote parts of the CTD (which in mammals consist of 52 repeats), thus looping out a long stretch in between. It has been discussed that such looping out can promote sequential recruitment of factors, by inducing specific turns in the CTD <sup>61</sup>. However, this potential role is not so plausible, since FluPol has been suggested to have negative effect on Pol II-dependent transcription through competition with the CTD processing factors <sup>111</sup>. Moreover, the observed avidity effect of binding to multiple repeats shows that interaction at one site increases the affinity to the other. Therefore, it is possible that an interaction with a distal stretch of the CTD could stimulate the binding also to the proximal region, increasing the likelihood of bringing FluPol in closer proximity to the RNA exit channel.



**Figure 59. Overview of the structures of FluA, FluB and FluC polymerase in complex with SeP5 CTD peptides, illustrating the two separate CTD-binding sites for each polymerase.** Cap-binding and endonuclease domain are colored in orange and forest green, respectively. The CTD peptides are shown in blue. Note that for FluC (PDB: 6F5P), the polymerase is in a conformation incompatible with cap-snatching.

In contrast to site 1, which is common for FluA and FluB polymerase, the other interacting regions are only conserved within the same strain (A, B or C), showing that these viruses have evolved divergent strategies to associate with the CTD. A parallel can be made with the recruitment of the capping enzymes from different species to the CTD. The available structures of a fungal, mammalian or fission yeast CE show that they all interact with SeP5 CTD peptides, but the binding interfaces are not conserved between lower and higher eukaryotes<sup>142,143,166</sup>.

The CTD binding site spanning over two domains (PA-C and PB2-627) in the FluB structure or interacting at the interface of PA and PB1 in FluC, suggests a mechanism for allosteric activation of the polymerase by the CTD by stabilization of the transcriptionally active state. This is corroborated by data showing an increase in endonuclease and transcription activity in the presence of a long stretch of CTD (six-repeats) for FluA and FluB (Figure 19). It is surprising that the same effect was also observed for FluA, as there is no obvious structural basis for such an activation, since both binding sites are on the same domain (PA-C). Comparison of the CTD-bound structures with those without the peptide shows that the interaction does not induce any long-range conformational changes and the RNA promoter-binding site remains unchanged upon peptide binding. Therefore, the most plausible explanation of the allosteric effect is that binding of the

peptide would stabilize the conformation in which the PB2 C-terminal domain, which can undergo significant conformational rearrangements, is stabilized in a state, where cap-snatching of the mRNA substrate can occur. The enhancement of activity was even more pronounced when full-length Pol II phosphorylated at Ser5 was used (Figure 42), further confirming that a long sequence spanning the different domains of FluPol has a more significant effect on the activation, probably due to the avidity effect when more repeats are present.

The functional analysis of the CTD-binding mutations presented here, strongly suggests that the interaction with the CTD is essential for the viral transcription. Impairment of the association of the two polymerases results in a strong transcriptional defects in a cellular context, as shown by the mini-genome assays for both FluA and FluB. The experiments showing that the CTD-binding deficient mutants are as active as the wild type in an *in vitro* setup, in which a capped substrate RNA is provided (Figure 23), supports the hypothesis that the CTD interaction is necessary to bring FluPol in proximity to the source of capped primers in the cell. Moreover, recombinant viruses deficient in CTD binding were highly attenuated and tended to revert by acquiring compensatory mutations.

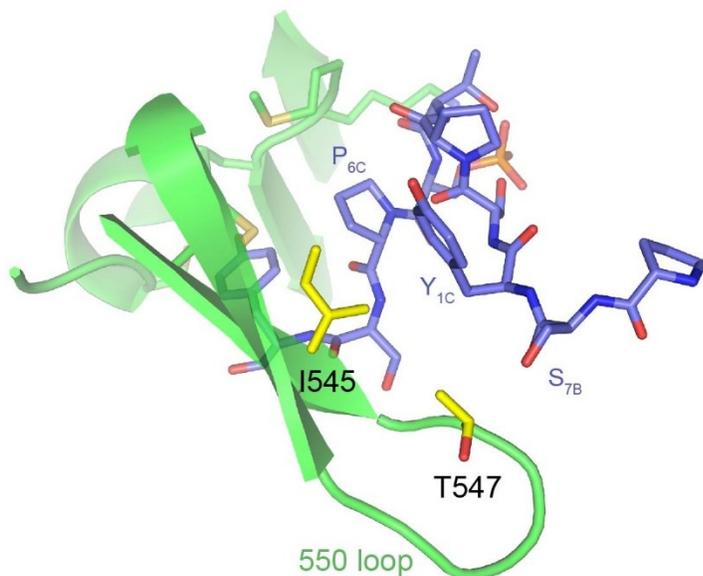
Two examples of revertants that rescued wild type infectivity of the mutant viruses came from this and a previous study<sup>153</sup>. The R→A mutations (R454A and R638A), which disrupt the recognition of SeP5 in both binding sites on FluA polymerase, were rescued by a C→R substitution in the same binding pocket (C489R and C453R, respectively) that could restore the lost interaction (Figure 26). While reverting the A back to R would require three nucleotide changes, the C to R reversion necessitates only one nucleotide change, but would result in the same phenotype. This hypothesis was confirmed by the binding data, showing increased affinities for the CTD peptides of the revertant mutants. The R638A mutant virus had been previously identified in another study, where it was shown to have transcriptional defects<sup>154</sup>. A further characterisation (which led to the discovery of the C453R revertant mutant) explained the attenuation of this virus by the generation of high proportion of defective interfering particles from the R638A virus<sup>153</sup>. Defective interfering particles (DI) are viral RNA segments, which have large internal deletions, but contain the conserved 5' and 3' ends and therefore can be packaged into vRNPs and be incorporated into virions<sup>167</sup>. The mechanisms of generation of DI particles are not well understood, but they must involve translocation or jumping of the polymerase on the vRNA template without replication. A

plausible hypothesis is that an impairment of transcription due to the loss of CTD binding by the R638A mutant could result in an insufficient expression of the viral nucleoprotein, necessary for the faithful replication of the long viral segments, and therefore, the production of DI particles. The generation of DI particles was reduced in the rescued C453R/R638A mutant, further corroborating this model <sup>153</sup>.

The revertant mutations affecting PB1 and PB2 subunits map to a similar region, and many of the involved residues participate in inter-domain interactions (Figure 27). It is possible that they have an indirect effect, by stabilizing particular conformations of the polymerase or affecting the transition between different conformations.

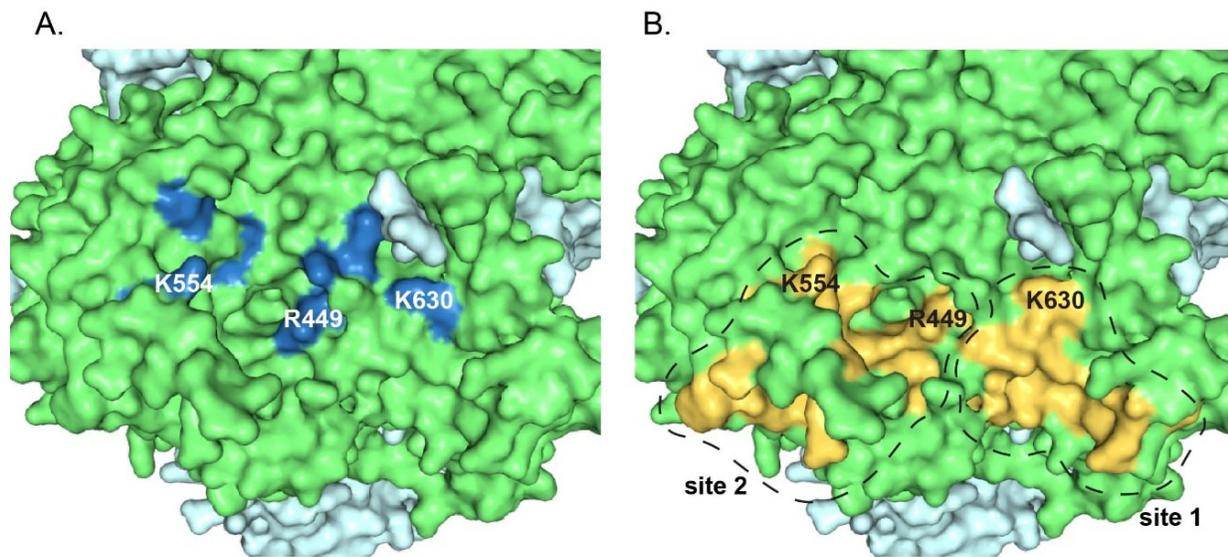
Other residues, which are part of the CTD-binding region, have also been previously characterized in several studies. The 550 loop, which contributes to the binding of the CTD peptide at site 2 of FluA polymerase, contains two residues that have been linked to virulence and host specificity <sup>150,168,169</sup>. Residue 550 (corresponding to 545 in H17N10) is most commonly Leu in human and avian strains and a substitution to Ile leads to attenuation of the virus <sup>168</sup>. Moreover, this residue play a role in inducing Pol II degradation, and this effect is alleviated by the L→I mutation <sup>150</sup>. In the FluA-CTD structure from A/H17N10 strain, the corresponding I545 makes hydrophobic interactions with Y<sub>1</sub> from the CTD (Figure 60). The 550 loop binds to the CTD by an induced fit mechanism, which makes it difficult to speculate on whether such a substitution could affect the affinity to the CTD, but it is plausible that the observed differences in Pol II degradation are dependent on modulations of the direct association of this region to the Pol II CTD. Another residue in this region, 552 (547 in H17N10) confers host-specificity: in avian strains this residue is a Thr, while in human it is a Ser. A T552S mutation strongly enhances the avian polymerase activity in human, but not in avian cells, and it is plausible that this adaptation is dependent on host-factor interactions. Sequence alignment of human and avian Pol II CTD (the only available complete sequence of the RPB1 of an avian species is from Bengalese finch) shows only two amino acid substitutions for the 52 repeats, confirming the high conservation of the CTD between metazoans, but would not explain the host specificity of a T552S mutant. Nevertheless, it is possible that subtle changes in the affinity to the CTD could modulate the activity of the polymerase in the cellular context and combined with other mutations, could explain the observed host restriction. We have observed that single basic residue mutations in the CTD binding pocket,

which lead to only four to seven-fold decrease of the affinity to the phosphorylated peptide, are sufficient to highly attenuate the virus, suggesting that this interaction is very sensitive to small changes (Figure 20, 24). It can be expected that species-dependent affinities to the CTD by various factors modulate a subtle balance with possibly important effect on host restriction.



**Figure 60. Interaction of FluA site 2 with the CTD peptide highlighting residues involved in host restriction.** Residues I545 and 547 (550 and 552 in human/avian FluA strains, respectively), discussed in the text, are shown in yellow sticks.

A systematic fitness profiling study on the PA subunit has identified a putative functional subdomain on the PA-C, which partially overlaps with the CTD-binding sites of FluA polymerase (Figure 61) <sup>170</sup>. This approach was based on evaluating functional constraints, while excluding structural or sequence conservation constraints, providing a measure of the essentialness of the investigated region. Furthermore, two of the SeP5-interacting residues on the PA-C were mutated (R454I and K635E) in the same study and showed strong impairment of the polymerase activity in a mini-genome assay, confirming the results presented here (Figure 21) and supporting the significance of the FluPol-CTD interaction for the virus.



**Figure 61. Overlap of the putative functional subdomains of PA-C, identified by fitness profiling with FluA CTD-binding sites.** (A) The PA-C subdomain, as defined in <sup>170</sup>, with functional residues highlighted in blue. (B) Same view as (A) with residues interacting with the CTD colored in yellow.

Recent studies have suggested that throughout the transcription cycle Pol II undergoes liquid-liquid phase separation, driven by modifications of the CTD <sup>171–173</sup>. The unphosphorylated CTD segregates through hydrophobic interactions and possibly by clustering with the Mediator complex. Phosphorylation by TFIIF leads to the disruption of the CTD-CTD association and allows formation of new electrostatic interactions with processing factors. Hyperphosphorylation by P-TEFb enhances these new interactions and leads to new phase separation, dependent on associations with the cyclin T1, one of the components of the kinase. These segregation processes add a supplementary layer of complexity of the transcriptional regulation through CTD modifications. In the context of influenza infection, FluPol interacts with the CTD as part of a vRNP and it is possible that the association of this vRNP to the sites of transcription is driven by similar clustering mechanisms. It is also plausible that the binding of the vRNP to CTD interferes with the regulation of transcription by disruption of different phase separation processes. However, a recent study of a herpesvirus-infected cells shows that viruses can also recruit the transcriptional machinery (transcription of this virus depends entirely on Pol II) through mechanisms distinct from phase separation, but by direct recruitment of Pol II to the viral DNA <sup>174</sup>. This highlights that

understanding the link between FluPol-CTD association and clustering through phase-separation requires investigating these processes in influenza-infected cells.

The binding data showing that FluPol interacts with the CTD in the presence or absence of the vRNA promoter (Figure 7A, B) suggests a possible mechanism for the switch from transcription to replication. The unprimed RNA synthesis activity of FluPol is lower compared to its transcriptional activity, when provided with capped primers<sup>175</sup>, but in the cell the transcription is dependent on the access of capped RNAs that can be snatched. During infection, the actively transcribing polymerase is tightly associated to the RNA template, constituting the vRNPs. They are the functional units of the virus, and the polymerase performs transcription and replication exclusively in the context of the RNPs. The incoming vRNPs from the virus enter the nucleus and associate with Pol II for initiating the primary transcription by cap-snatching. After being synthesized in the cytoplasm, the newly produced polymerase subunits and nucleoprotein are imported into the nucleus. These apo-polymerases could then compete with the incoming vRNPs for association with the CTD (since the interaction is independent on the presence of the viral promoter), limiting the available Pol II for cap-snatching. Deficiency of capped primers would then promote unprimed replication by the polymerase by default. The replication product vRNAs are packaged with the imported newly synthesized polymerase and nucleoprotein into vRNPs, which can perform a next round of transcription by associating with the Pol II (secondary transcription). This model proposes an essential role of transcription and more precisely, of the interaction of FluPol with the Pol II, for the regulation of the entire viral cycle.

The structural details of the interaction with the CTD and CTD-mimic modified peptides provides a basis for design of molecules, which could inhibit the polymerase function during influenza infection. A comparison with other CTD-bound structures of various cellular partners shows highly specific CTD binding sites, even though the conformation of the CTD can be similar (Figure 12, 14). The functional data presented here shows that inhibiting this interaction even partially would have detrimental effect on the multiplication of the virus, suggesting that targeting this association is a good approach for designing new drugs against influenza. Designing drugs that could inhibit a protein-protein interaction is nevertheless a challenging task due to the large interface surfaces. The high-throughput assay developed based on the interaction data presented in this work was used to identify compounds from screens of small molecules and several

candidate inhibitors were found. Their specificity has to be further assayed in a cell-based assay to verify that they would not have off-site effects by binding to some of the cellular CTD-binding factors. Furthermore, an attractive drug candidate would recognize both FluA and FluB polymerases, since the two strains are co-circulating during seasonal epidemics. The high conservation of CTD binding site 1 suggests this could be an appropriate target.

#### **4.2. Reconstitution of an active cap-snatching complex**

In the second part of this thesis I investigated the association of the viral and cellular transcription machineries in a larger context. The goal was to reconstitute a functional cap-snatching complex with FluPol and actively transcribing Pol II. First, the complex of the two polymerases was characterized in order to understand how FluPol associates with a full-length elongating Pol II, stalled with a capped RNA emerging from the exit channel. Then, two elongation factors, which had been reported to interact with FluPol were investigated for their role in the recruitment of the viral polymerase to the paused Pol II

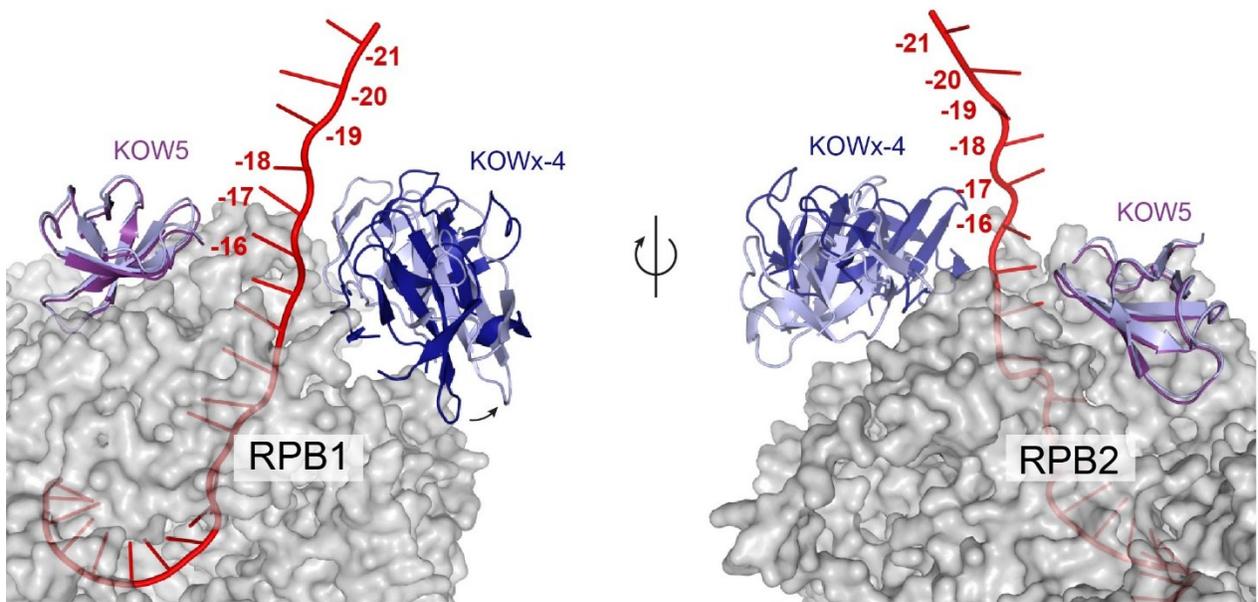
The presented results show that in the context of the full-length Pol II and CTD, a Ser5-phosphorylation is necessary to assemble a complex of the two polymerases, confirming the data obtained using SeP5 CTD-mimic peptides (section 3.1). P-TEFb was the kinase of choice due to its known specificity for Ser5 *in vitro*. It is not well understood what is the functional link between FluPol and P-TEFb during infection. According to one study, P-TEFb interacts with FluPol and promotes the binding of the viral polymerase to the CTD of Pol II<sup>102</sup>. A different study shows that inhibiting P-TEFb with DRB in cells, where polymerase subunits and NP are pre-expressed, does not interfere with viral transcription<sup>95</sup>, suggesting that the FluPol interacts with Pol II before the productive elongation phase. I did not observe phosphorylation of FluPol by P-TEFb, excluding this as an explanation for a functional dependence on the kinase (data not shown). *In vivo* FluPol probably interacts with the CTD after it has been phosphorylated by TFIIF, before the establishment of the pausing state. Nevertheless, in the conditions of *in vitro* reconstitution, the phosphorylation by P-TEFb enhanced the complex formation, and was preferred as it could easily be separated by size-exclusion chromatography from the FluPol/Pol II complex. P-TEFb did not co-elute with FluPol in any of the assembly experiments (Figure 34, 35), including

the experiment with phosphorylated DSIF (see also Figure 48, 49), suggesting that the kinase does not associate with FluPol.

In the context of transcribing Pol II the nascent RNA is capped shortly after it appears on the surface of the polymerase, at a length of around 20 nt in both yeast and human cells <sup>75,176</sup>. The capping enzyme is recruited to the paused Pol II. The nuclear cap-binding complex (CBC) binds to the capped RNA at around 40 nt, supposedly also during the promoter-proximal pausing or early elongation <sup>177</sup>. Therefore, the FluPol cap-snatching, which depends on the presence of a cap, but is most likely to happen before the capped RNA has been bound by CBC, must occur in a very short time-window during the Pol II pausing. The functional assays presented in section 3.2 confirmed that a capped RNA, pre-bound to an elongating Pol II complex, is a suitable substrate for the FluPol endonuclease activity. The optimal length of capped transcript recognized and cleaved by FluPol *in vitro* was determined to be around 27-28 nt (depending on the sequence of the FluPol endonuclease cleavage site), which corresponds to a length of the transcript that in the cellular context would be already capped, but still not bound by the CBC.

The recruitment of FluPol during this narrow time window of transcription is also likely to be regulated with the help of additional factors. This is the case for the capping enzyme, for instance, which besides interacting with the SeP5-form of the CTD, also binds to the unphosphorylated CTR domain of Spt5 (one of the subunits of the DSIF complex) <sup>143</sup>. The recruitment and activation of the capping enzyme and FluPol by the CTD, as well as the similar lengths of the nascent RNA, recognized by the two proteins, prompted us to look for more similarities. Spt5, which had been identified as a FluPol-interacting protein, was therefore a strong candidate for playing a role in the assembly of the functionally relevant cap-snatching complex. DSIF is tightly bound to Pol II and the nascent transcript throughout the promoter-proximal pausing, as well as during productive elongation. Binary interactions between DSIF and FluPol were tested and confirmed a direct association between them, independent of the DSIF phosphorylation by P-TEFb. Moreover, a complex of elongating Pol II, phosphorylated DSIF and FluPol was formed. In this experimental setup P-TEFb activity was necessary for the binding of the viral polymerase to Pol II, and concomitant phosphorylation of DSIF could not be avoided. It is possible that in these conditions conformational changes triggered by the P-TEFb-dependent phosphorylation induce a state characteristic of later stages of transcription (e.g. productive elongation), since it is known that

phosphorylation of DSIF by P-TEFb causes significant rearrangements of its domains<sup>82</sup>. This could be therefore irrelevant for FluPol cap-snatching. One experiment that corroborates this hypothesis, is that the optimal length of pre-bound RNA, cleaved by FluPol, was not affected by DSIF (Figure 40 and 58). This result was not anticipated, since the RNA clamp formed by DSIF would cover a short stretch of the exiting RNA, thus increasing the distance between the Pol II active site and the cleavage site of FluPol. The very recent structure of a phosphorylated DSIF<sup>82</sup> suggests that the conformational changes driven by the P-TEFb phosphorylation loosen the RNA clamp, formed by DSIF (Figure 62). This could explain why it did not affect the length of the cleaved RNA by FluPol. This also could be why a stable complex of elongating Pol II, DSIF and FluPol, suitable for structural studies could not be produced in this experimental setup.



**Figure 62. Conformational change of the DSIF KOW<sub>x</sub>-KOW<sub>4</sub>, forming the RNA clamp, upon phosphorylation by P-TEFb.** Opposite views of the exiting RNA (red) on the Pol II surface (grey). DSIF KOW<sub>x</sub>-4 domain of the non-phosphorylated elongation complex (PDB: 6GMH) is shown in dark blue and the KOW5 domain in purple. The corresponding domains in the phosphorylated elongation complex (PDB: 6GML) are in lightblue.

Based on the structural data of the DSIF-Pol II complex, a rational approach was used to identify the DSIF domains involved in the interaction with FluPol. The CTR domain of Spt5 plays a role in recruiting RNA processing factors and is complementary in function to the CTD of Pol II. This domain was tested in its unphosphorylated (characteristic of the pausing state) or phosphorylated by P-TEFb (elongation pattern) form, but no interaction with FluPol was observed. Other domains, which bind in proximity to the RNA exit channel (KOWx-4 and KOW5), also did not show association with FluPol in the experimental conditions described here. However, it cannot be excluded that in the context of transcribing Pol II, FluPol can interact with one of the domains tested in isolation. Since the CTD plays an allosteric role in FluPol endonuclease activity, it is plausible that the binding to the CTD could also promote interaction with any of these domains. The CTD of Pol II and CTR of Spt5 are known to play a cooperative effect for the recruitment of the capping enzyme *in vivo*<sup>143</sup>, so it is possible that cooperative binding of the two domains to FluPol could occur. Further studies will have to be performed in order to map the interaction and understand the role of DSIF for FluPol cap-snatching.

The elongation factor Tat-SF1, which has also been identified as an interactor of FluPol, is much less characterized, therefore its role remains more speculative. It is involved in the regulation of elongation together with DSIF and the PAF complex, but it does not form a stable complex with transcribing Pol II (Seychelle Vos, personal communication). Even though binding of Tat-SF1 to FluPol was observed, in the context of elongating Pol II, Tat-SF1 did not associate with the FluPol/Pol II complex. Plausible explanations are that either Tat-SF1 and Pol II have overlapping binding sites on FluPol or that steric hindrance would prevent simultaneous binding. In order to understand the binding incompatibility it would be interesting to test whether FluPol, bound to a CTD-mimic peptide, would still be able to associate with Tat-SF1.

The mechanism of action of Tat-SF1 as an elongation factor is not known. It is possible that it is recruited to the nascent transcript by interaction through its RRM domains (Figure 53). An increased local concentration of Tat-SF1 could then compete with Pol II for FluPol binding and lead to dissociation of FluPol. As seen from the complex assembly experiments, FluPol associates with the phosphorylated Pol II even in absence of capped RNA (Figure 34B), suggesting that there must be a mechanism to release it after cap-snatching in order to recycle the abortive Pol II, even though it is not known if the transcribing FluPol remains bound to the CTD.

The binding of DSIF, however, did not seem to interfere with the association between FluPol and Tat-SF1. Whether the ternary complex between FluPol, Tat-SF1 and DSIF has a functional significance and is an intermediate within the FluPol transcription cycle, or it just shows that two elongation factors (DSIF and Tat-SF1) bind to separate sites, remains to be investigated. Furthermore, many open questions have to be addressed: the role of P-TEFb in the assembly of the different ternary complexes (FluPol/Pol II/DSIF, FluPol/Tat-SF1/DSIF); whether it phosphorylates Tat-SF1; the role of P-TEFb-dependent DSIF phosphorylation for FluPol binding in the Pol II context. The data presented here provides a starting point for a more comprehensive investigation of the functional coupling of the viral and cellular transcription machineries.

This thesis presents a detailed structural and functional characterization of the interaction between FluA and FluB polymerases and the CTD of Pol II. The structural basis for the SeP5 specificity is provided, which explains how the viral polymerase is recruited to the initiating phase of transcription. Disruption of the interaction results in impairment of the polymerase transcriptional activity and strong attenuation of the virus, showing that the interaction is essential for influenza replication. The detailed analysis of the structural recognition of the CTD peptides provides a basis for the design of compounds that would prevent this association and therefore inhibit influenza replication. To understand the recruitment of the viral polymerase to the source of capped primers, a cap-snatching complex, containing FluPol and Pol II was reconstituted and characterized. Furthermore, a direct interaction with two elongation factors was demonstrated, providing additional information on the recruitment of the viral polymerase to the right cellular context for cap-snatching. The current work therefore provides a basis for investigation of the physical and functional link between the viral and cellular transcription machineries.

## Acknowledgements

I thank my parents and my brother for their love and support and for always being there for me, even when they are far away.

I would like to thank Stephen Cusack for guiding me through this experience, for being an unlimited source of ideas and for being excited about my ideas; for being supportive when I needed it, for teaching me how to pursue my scientific goals; for his trust and mentorship.

I would like to thank Rob Ruigrok, Ervin Fodor, Meindert Lamers, Carrie Bernecky and Christoph Müller for evaluating my thesis.

I want to acknowledge all the members of the Cusack group, and especially Alexander Pflug, for his great contribution for solving the structure of FluA polymerase with the CTD; Stefan Reich, for his help with establishing the binding and activity assays, the purification of the polymerase, for advice and critical reading of the manuscript and for his moral support when it was most needed; Delphine Guilligay, for her help with insect cell expression and purification, as well as for sharing her expertise on the influenza polymerase and for final corrections on all the French part of the thesis; Petra Drncova for pursuing and succeeding in solving the structure of FluB polymerase with the CTD and for help with baculovirus expression and mutagenesis for the FluB mini-genome assays; Patricia Resa-Infante for her help in setting up the mini-genome assay; Joanna Wandzik, for providing FluB polymerase for the activity assays, but more importantly, for constant discussions, support and friendship; Valentina Speranzini, for advice on protein purifications and for sharing all the sad gels and lonely weekends at the end; Manikandan Karuppasamy and Tomas Kouba, for all their help with electron microscopy; Erika Pellegrini, for help and advice and to all the other members who have been helpful throughout my thesis work.

All my collaborators have been of great help, to begin with Nadia Naffakh, Guillaume Fournier and Catherine Isel-Griffiths who did all the reverse genetics experiments. I would like to acknowledge Jan Ellenberg, Antonio Politi and Nathalie Daigle for their help with fluorescence microscopy experiments. I am extremely thankful to Patrick Cramer for hosting me in his lab for establishing the cap-snatching complex assembly experiments, and most of all to Seychelle Vos and Carrie Bernecky, for providing purified Pol II and constructs and/or proteins for all the elongation factors that have been used in this study and for their constant help and advice.

I would like to acknowledge all the support I have received from everyone around the EMBL Grenoble outstation throughout the years. Ramesh Pillai and his group have been very helpful for setting up the experiments in mammalian cells. The Panne group, with their late night and weekends discussions, were of a great moral support, as well as the Marcia group whose positiveness and friendship have been extremely helpful. Daniel Peter was so kind to (very) critically read the manuscript, for which I am very thankful. Wojtek Galej, Kyle Muir and Paul Sauer provided a lot of advice on the electron microscopy data analysis. I would also like to acknowledge all the facilities of EMBL, IBS and PSB for providing infrastructure for the experiments.

I would have not been at this position if it was not for two inspirational women who have been my scientific and personal mentors through the years. I want to deeply thank Donka Nikolova, for putting the basis on which I am building my knowledge in biology and Svetla Petrova who opened the magical world of enzymology to me. I also don't want to miss to acknowledge my two former labmates who have taught me to enjoy science, Rado and Andi.

Finally, I would like to thank all my friends, far away or close, who have been with me through the whole time and whose love and support have kept me going. I want to thank Nelly, for sharing the road with me and Hervé and Sébastien, for being here when I needed it most.

## Appendix

## Appendix Table 1. Sequences of CTD peptides

<b>FAM-labelled peptides</b>	
two-repeat SeP5	5FAM-YSPT <sub>p</sub> SPSYSP <sub>Tp</sub> SPS
two-repeat SeP2	5FAM-Y <sub>p</sub> SPTSPSY <sub>p</sub> SPTSPS
four-repeat nonP	5FAM-YSPTSPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPS
four-repeat SeP5	5FAM-YSPT <sub>p</sub> SPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPS
<b>non-labelled peptides</b>	
two-repeat SeP5	YSPT <sub>p</sub> SPSYSP <sub>Tp</sub> SPS
four-repeat SeP5	YSPT <sub>p</sub> SPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPSYSP <sub>Tp</sub> SPS
modified peptides	
tetra-linker	Ac-YSPT <sub>p</sub> SPSYSPSYSP <sub>Tp</sub> SPSY
thio-phosphate SeP5	Ac-YSPT <sub>ps</sub> SPSYSP <sub>Tps</sub> SPSYSP <sub>Tps</sub> SPSY

**Appendix Table 2. Sequences of promotor RNAs**

<b>A/little yellow-shouldered bat/Guatemala/060/2010</b>	
v5' 13	pAGUAGAAACAAGG
v5' 14	pAGUAGAAACAAGGC
v5' 15	pAGUAGAAACAAGGCA
c5' 14	pAGCAGAAGCAGGCA
v3' 13	GCCUGCUUCUGCU
v3' 14	UGCCUGCUUCUGCU
v3' 18	AGUUUGCCUGCUUCUGCU
c3' 15	UGCCUUGUUUCUACU
c3' 18	AAAUGCCUUGUUUCUACU
<b>B/Memphis/13/2003</b>	
v5' 14	pAGUAGUAACAAGAG
v5' 16	pAGUAGUAACAAGAGGG
v3' 13	CUCUGCUUCUGCU
v3' 18	UAUACCUCUGCUUCUGCU

All sequences are shown from 5' to 3'.

**Appendix Table 3. Capped RNAs produced *in vitro***

<b>20-mer</b>	m7G-GAAUCACAUAAGACCAGGC
<b>23-mer</b>	m7G-GAAUCUAUACAUAAGACCAGGC
<b>28-mer</b>	m7G- GAAUCUAUAUAACAUAAGACCAGGC

**Appendix Table 4. DNA scaffolds**

<b>template DNA</b>	
<b>39 nt</b>	GGCAGTACTAGTATTCTAGTATTGAAAGTACTTGAGCTT
<b>43 nt</b>	GGCAGTACTAGTATTCTAGTATTGAAAGTACTTGAGCTTGATC
<b>+1</b>	GGCAGTACTAGTATTCTAGTATTGACCGTACTTGAGCTTGATC
<b>+2</b>	GGCAGTACTAGTATTCTAGTATTGACCGTACTTGAGCTTGATC
<b>+3</b>	GGCAGTACTAGTATTCTAGTATTGATCGTACTTGAGCTTGATC
<b>+4</b>	GGCAGTACTAGTATTCTAGTATTGATTCCACTTGAGCTTGATC
<b>+5</b>	GGCAGTACTAGTATTCTAGTATTGATTTCACTTGAGCTTGATC
<b>non-template DNA</b>	
<b>39 nt</b>	AAGCTCAAGTACTTAAGCCTGGTCATTACTAGTACTGCC
<b>43 nt</b>	GATCAAGCTCAAGTACTTAAGCCTGGTCTATACTAGTACTGCC
<b>+1</b>	GATCAAGCTCAAGTACGGGAGCCTGGTCTATACTAGTACTGCC
<b>+2</b>	GATCAAGCTCAAGTACGGAAGCCTGGTCTATACTAGTACTGCC
<b>+3</b>	GATCAAGCTCAAGTACGAAAGCCTGGTCTATACTAGTACTGCC
<b>+4</b>	GATCAAGCTCAAGTGGAAAAGCCTGGTCTATACTAGTACTGCC
<b>+5</b>	GATCAAGCTCAAGTGAAAAAGCCTGGTCTATACTAGTACTGCC

All sequences are shown from 5' to 3'.

**Appendix Table 5. Data collection and refinement statistics for FluA-CTD, FluB-CTD, FluA-tetralinker CTD, FluA-thiophosphate CTD structures**

	FluA polymerase vRNA 28-mer SeP5	FluB polymerase vRNA 28-mer SeP5	FluA polymerase vRNA 22mer thiophosphate SeP5	FluA polymerase vRNA tetralinker SeP5
Data collection				
Space group	C2	$P3_221$	C2	C2
Cell dimensions				
$a, b, c$ (Å)	269.49 147.52 88.48	200.16 200.16 250.38	267.77 147.81 88.41	268.74 147.52 88.54
$\alpha, \beta, \gamma$ (°)	90.00 97.22 90.0	90.0 90.0 120.0	90.0 97.08 90.0	90.0 96.63 90.0
Number of crystals	1	1	1	1
Resolution (Å) Outer shell*	50.0-2.50 2.60-2.50	50.0-3.50 3.60-3.50	50.0-2.70 (2.77-2.70)	50.0-2.80 (2.87-2.80)
$R_{\text{merge}}$	0.118 (1.25)*	0.191 (2.22)*	0.049 (0.500)	0.102 (0.589)
$I/\sigma I$	9.45 (1.21)*	8.38 (1.02)*	14.34 (1.87)	6.35 (1.43)
Completeness (%)	99.6 (99.6)*	99.9 (99.9)*	98.6 (96.7)	96.1 (96.4)
Redundancy	5.84 (5.90)*	8.54 (8.63)*	2.92 (2.69)	2.19 (2.06)
Refinement				
Resolution (Å)	50.0-2.50 (2.53-2.50)*	50.0-3.50 (3.55-3.50)*	50.0-2.80 (2.87-2.80)*	50-2.70 (2.77-2.70)*
No. reflections (work/free)	117997/5756	69950/3523	87532 (4534)	77141 (3923)
$R_{\text{work}}/ R_{\text{free}}$	21.0/24.5 (32.6/35.7)*	24.0/27.1 (43.7/43.3)*	23.0/26.9 (34.0/35.4)*	21.5/25.7 (32.8/36.1)*
No. atoms (total)	18529	18013	18457	18364
Protein	17474	17351	17510	17552

RNA	600	589	600	600
Ligand (pS5)	380	73	135	146
Solvent	75	-	212	65
B-factors ( $\text{\AA}^2$ )				
Overall	77.1	170.1	92.4	83.4
Protein	78.2	170.3	93.2	83.9
RNA	65.9	155.0	81.7	69.7
Ligand (pS5)	61.3	174.0	78.3	75.0
Solvent	85.8	-	76.3	104.3
R.m.s deviations				
Bond lengths ( $\text{\AA}$ )	0.008	0.003	0.007	0.007
Bond angles ( $^\circ$ )	1.279	0.500	1.090	1.013
Validation**				
Ramachandran (%)				
Favoured	94.9	94.0	95.2	95.7
Outliers	0.7	0.32	0.6	0.4
Clash-score	1.27	0.67	1.51	0.77
MolProbity score	1.59	1.29	1.48	1.19

\*High resolution shell in brackets

\*\* Molprobity: <http://molprobity.biochem.duke.edu/>

## References

1. World Health Organization, News Release. (2017).
2. Krammer, F. *et al.* Influenza. *Nat. Rev. Dis. Prim.* **4**, 3 (2018).
3. Nelson, M. I. & Holmes, E. C. The evolution of epidemic influenza. *Nat. Rev. Genet.* **8**, 196–205 (2007).
4. Horimoto, T. & Kawaoka, Y. Influenza: Lessons from past pandemics, warnings from current incidents. *Nat. Rev. Microbiol.* **3**, 591–600 (2005).
5. Tong, S. *et al.* New World Bats Harbor Diverse Influenza A Viruses. *PLoS Pathog.* **9**, (2013).
6. van de Sandt, C. E., Bodewes, R., Rimmelzwaan, G. F. & de Vries, R. D. Influenza B viruses: not to be discounted. *Future Microbiol.* **10**, 1447–1465 (2015).
7. Donis, R. O. Orthomyxoviridae. in *Fenner's Veterinary Virology 5th edition* (eds. MacLachlan, J. & Dubovi, E.) 389–410 (2017).
8. Palese, P., Shaw, M. L., Wright, P., Neumann, G. & Kawaoka, Y. Orthomyxoviridae. in *Fields Virology 6th edition* (eds. Knipe, D. & Howley, P.) 1151–1243 (2013).
9. Tong, S. *et al.* A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. USA* **109**, 4269–4274 (2012).
10. Moreira, É. A. *et al.* Synthetically derived bat influenza A-like viruses reveal a cell type-but not species-specific tropism. *Proc. Natl. Acad. Sci. USA* **113**, 12797–12802 (2016).
11. Hause, B. *et al.* Characterization of a novel influenza virus strain in cattle and swine: proposal for a new genus in the Orthomyxoviridae family. *MBio* **5**, 1–10 (2014).
12. Paules, C. I., Marston, H. D., Eisinger, R. W., Baltimore, D. & Fauci, A. S. The Pathway to a Universal Influenza Vaccine. *Immunity* **47**, 599–603 (2017).
13. Shaw, M. L. The Next Wave of Influenza Drugs. *ACS Infect. Dis.* **3**, 691–694 (2017).
14. Dubois, J., Terrier, O. & Rosa-Calatrava, M. Influenza viruses and mRNA splicing: Doing more with less. *MBio* **5**, e00070-14 (2014).
15. Fodor, E. & Brownlee, G. G. Influenza Virus Replication. in *Influenza* (ed. C.W.Potter) 1–29 (2002).
16. Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: Influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol.* **13**, 28–41 (2015).
17. Fujii, Y., Goto, H., Watanabe, T., Yoshida, T. & Kawaoka, Y. Selective incorporation of influenza virus RNA segments into virions. *Proc. Natl. Acad. Sci. USA* **100**, 2002–2007 (2003).
18. Noda, T. *et al.* Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490–492 (2006).

19. Noda, T. *et al.* Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3**, 636–639 (2012).
20. Bernadeta, D., Barilaite, E., Fodor, E., Laederach, A. & Bauer, D. L. V. The structure of the influenza A genome. *bioRxiv* (2017).
21. te Velthuis, A. J. W. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat. Rev. Microbiol.* **14**, 479–493 (2016).
22. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
23. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).
24. Hengrung, N. *et al.* Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527**, 13034–13039 (2015).
25. Thierry, E. *et al.* Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Mol. Cell* **61**, 125–137 (2016).
26. Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S. & Cusack, S. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Res.* **234**, 103–117 (2017).
27. Reich, S., Guilligay, D. & Cusack, S. An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. *Nucleic Acids Res.* **45**, 3353–3368 (2016).
28. 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 Res.* **41**, 8289–8299 (2013).
29. Bouloy, M., Plotch, S. J. & Krug, R. M. Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc. Natl. Acad. Sci. USA* **75**, 4886–4890 (1978).
30. Plotch, S. J., Bouloy, M. & Krug, R. M. Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc. Natl. Acad. Sci. USA* **76**, 1618–22 (1979).
31. 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. *Sci. Rep.* **4**, 6181 (2014).
32. 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 Res.* **43**, 5052–5064 (2015).
33. Gu, W. *et al.* Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21**, 2067–75 (2015).
34. te Velthuis, A. J. W. & Oymans, J. Initiation, Elongation and Realignment during Influenza Virus. *J. Virol.* **92**, 1–12 (2018).

35. Fodor, E., Pritlove, D. C. & Brownlee, G. G. The Influenza-Virus Panhandle Is Involved in the Initiation of Transcription. *J Virol* **68**, 4092–4096 (1994).
36. Tiley, L. S., Hagen, M., Matthews, J. T. & Krystal, M. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J Virol* **68**, 5108–5116 (1994).
37. Nemeroff, M. E., Barabino, S. M. L., 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. *Mol. Cell* **1**, 991–1000 (1998).
38. 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. *Proc. Natl. Acad. Sci. USA* **110**, E4238–E4245 (2013).
39. Deng, T., Vreede, F. T. & Brownlee, G. G. Different De Novo Initiation Strategies Are Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral RNA Replication. *J. Virol.* **80**, 2337–2348 (2006).
40. Oymans, J. & Te Velthuis, A. J. W. A Mechanism for Priming and Realignment during Influenza A Virus Replication. *J. Virol.* **92**, e01773-17 (2018).
41. Jorba, N., Coloma, R. & Ortín, J. Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog.* **5**, e1000462 (2009).
42. 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. *Proc. Natl. Acad. Sci. USA* **110**, E4238–E4245 (2013).
43. Perez, J. T. *et al.* Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc. Natl. Acad. Sci. USA* **107**, 11525–11530 (2010).
44. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine. *RNA Biol.* **8**, 207–215 (2011).
45. Robb, N. C., Smith, M., Vreede, F. T. & Fodor, E. NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *J. Gen. Virol.* **90**, 1398–1407 (2009).
46. Vreede, F. T., Jung, T. E. & Brownlee, G. G. Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *J. Virol.* **78**, 9568–72 (2004).
47. Kawaguchi, A., Momose, F. & Nagata, K. Replication-Coupled and Host Factor-Mediated Encapsidation of the Influenza Virus Genome by Viral Nucleoprotein. *J. Virol.* **85**, 6197–6204 (2011).
48. Kawaguchi, A. & Nagata, K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *EMBO J.* **26**, 4566–4575 (2007).
49. Sugiyama, K., Kawaguchi, A., Okuwaki, M. & Nagata, K. PP32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. *Elife* **4**, 1–19 (2015).
50. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.*

- 1819**, 391–400 (2012).
51. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).
  52. Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* **2**, 277–298 (2011).
  53. Ghosh, A. & Lima, C. D. Enzymology of RNA cap synthesis. *Wiley Interdiscip. Rev. RNA* **1**, 152–172 (2010).
  54. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
  55. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
  56. Eckmann, C. R., Rammelt, C. & Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA* **2**, 348–361 (2011).
  57. Cramer, P. Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science* **292**, 1863–1876 (2001).
  58. Corden, J. L. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem. Rev.* **113**, 8423–8455 (2013).
  59. Cagas, P. M. & Corden, J. L. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins* **21**, 149–60 (1995).
  60. Kumaki, Y., Matsushima, N., Yoshida, H., Nitta, K. & Hikichi, K. Structure of the YSPTSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. *Biochim. Biophys. Acta* **1548**, 81–93 (2001).
  61. Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19**, 1401–1415 (2005).
  62. Jasnovidova, O. & Stefl, R. The CTD code of RNA polymerase II: A structural view. *Wiley Interdiscip. Rev. RNA* **4**, 1–16 (2013).
  63. Harlen, K. M. & Churchman, L. S. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18**, 263–273 (2017).
  64. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. USA* **88**, 10004–10008 (1991).
  65. Wong, K. H., Jin, Y. & Struhl, K. TFIIF Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Mol. Cell* **54**, 601–612 (2014).
  66. Yamaguchi, Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).

67. 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 Dev.* **12**, 343–356 (1998).
68. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017).
69. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921–924 (2017).
70. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018).
71. Hartzog, G. A. & Fu, J. The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 105–115 (2013).
72. McCracken, S. *et al.* 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306–3318 (1997).
73. Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* **3**, 405–411 (1999).
74. 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. *Mol. Cell. Biol.* **30**, 2353–2364 (2010).
75. Martinez-Rucobo, F. W. *et al.* Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol. Cell* **58**, 1079–1089 (2015).
76. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *Journal of Biological Chemistry* **270**, 12335–12338 (1995).
77. Lu, X. *et al.* Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II. *Nucleic Acids Res.* **44**, 6853–6867 (2016).
78. Ramanathan, Y. *et al.* Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences. *J. Biol. Chem.* **276**, 10913–10920 (2001).
79. Ramanathan, Y., Reza, S. M., Young, T. M., Mathews, M. B. & Pe'ery, T. Human and rodent transcription elongation factor P-TEFb: interactions with human immunodeficiency virus type 1 tat and carboxy-terminal domain substrate. *J. Virol.* **73**, 5448–58 (1999).
80. Czudnochowski, N., Böskén, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3**, 842 (2012).
81. McNamara, R. P., Bacon, C. W. & D'Orso, I. Transcription elongation control by the 7SK snRNP complex: Releasing the pause. *Cell Cycle* **15**, 2115–2123 (2016).

82. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560**, 607–612 (2018).
83. Chen, Y. *et al.* DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev.* **23**, 2765–2777 (2009).
84. Zhou, Q., Chen, D., Pierstorff, E. & Luo, K. Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages. *EMBO J.* **17**, 3681–3691 (1998).
85. Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316 (2010).
86. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430**, 223–226 (2004).
87. 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. *Mol. Cell* **13**, 67–76 (2004).
88. Barry, R. D., Ives, D. R. & Cruickshank, J. G. Participation of deoxyribonucleic acid in the multiplication of influenza virus. *Nature* **194**, 1139–1140 (1962).
89. Rott, R. & Scholtissek, C. Specific inhibition of influenza virus by alfa-amanitin. *Nature* **228**, 56 (1970).
90. Tamm, I., Folkers, K., Shunk, C. H. & Horsfall Jr., F. L. Inhibition of influenza virus multiplication by N-glycosides of benzimidazoles-N. *J Exp Med* **99**, 227–250 (1954).
91. Lamb, R. & Choppin, P. W. Synthesis of influenza virus polypeptides in cells resistant to alpha-amanitin: evidence for the involvement of cellular RNA polymerase II in virus replication. *J. Virol.* **23**, 816–819 (1977).
92. Spooner, L. & Barry, R. Participation of DNA-dependent RNA polymerase II in replication of influenza viruses. *Nature* **268**, 650–2 (1977).
93. Engelhardt, O., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* **79**, 5812–5818 (2005).
94. Lidschreiber, M., Leike, K. & Cramer, P. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Mol. Cell. Biol.* **33**, 3805–3816 (2013).
95. 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).
96. Loucaides, E. M. *et al.* Nuclear dynamics of influenza A virus ribonucleoproteins revealed by live-cell imaging studies. *Virology* **394**, 154–163 (2009).
97. Tafforeau, L. *et al.* Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *J. Virol.* **85**, 13010–13018 (2011).
98. 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. *J. Virol.* **85**, 8569–8581 (2011).
99. Naito, T. *et al.* An influenza virus replicon system in yeast identified Tat-SF1 as a stimulatory host factor for viral RNA synthesis. *Proc. Natl. Acad. Sci. USA* **104**, 18235–18240 (2007).
  100. 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).
  101. Mayer, D. *et al.* Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomics-based approaches. *Proteome* **6**, 672–682 (2007).
  102. Zhang, J., Li, G. & Ye, X. Cyclin T1/CDK9 Interacts with Influenza A Virus Polymerase and Facilitates Its Association with Cellular RNA Polymerase II. *J. Virol.* **84**, 12619–12627 (2010).
  103. Huarte, M., Sanz-Ezquerro, J. J., Roncal, F., Ortín, J. & Nieto, A. PA subunit from influenza virus polymerase complex interacts with a cellular protein with homology to a family of transcriptional activators. *J. Virol.* **75**, 8597–604 (2001).
  104. Popow, J. *et al.* HSPC117 Is the Essential Subunit of a Human tRNA Splicing Ligase Complex. *Science* **331**, 760–764 (2011).
  105. Rodriguez, A., Perez-Gonzalez, A. & Nieto, A. Cellular Human CLE/C14orf166 Protein Interacts with Influenza Virus Polymerase and Is Required for Viral Replication. *J. Virol.* **85**, 12062–12066 (2011).
  106. Landeras-Bueno, S., Jorba, N., Pérez-Cidoncha, M. & Ortín, J. The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. *PLoS Pathog.* **7**, e1002397 (2011).
  107. Fournier, G. *et al.* Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Ciral mRNA Splicing. *PLoS Pathog.* **10**, e1004164 (2014).
  108. Watanabe, T., Watanabe, S. & Kawaoka, Y. Cellular networks involved in the influenza virus life cycle. *Cell Host Microbe* **7**, 427–439 (2010).
  109. Matsuoka, Y. *et al.* A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* **7**, 97 (2013).
  110. Vreede, F. T. & Fodor, E. The role of the influenza virus RNA polymerase in host shut-off. *Virulence* **1**, 436–439 (2010).
  111. Bauer, D. L. V *et al.* Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Rep.* **23**, 2119–2129 (2018).
  112. West, S., Gromak, N. & Proudfoot, N. J. Human 5' → 3' exonuclease Xm2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525 (2004).
  113. Boelens, W. C. & Izaurralde, E. Nuclear Export of Different Classes of RNA Is Mediated by Specific Factors Preparation of RNA for Injection Quantitation of Export Data. *Cell* **124**,

- 627–635 (1994).
114. Izaurralde, E. *et al.* A cap-binding protein complex mediating U snRNA export. *Nature* **376**, (1995).
  115. Rodriguez, A., Perez-Gonzalez, A. & Nieto, A. Influenza Virus Infection Causes Specific Degradation of the Largest Subunit of Cellular RNA Polymerase II. *J. Virol.* **81**, 5315–5324 (2007).
  116. Marazzi, I. *et al.* Suppression of the antiviral response by an influenza histone mimic. *Nature* **483**, 428–433 (2012).
  117. Das, K. *et al.* Structural basis for suppression of a host antiviral response by influenza A virus. *Proc. Natl. Acad. Sci. USA* **105**, 13093–8 (2008).
  118. Clark, M. P. *et al.* Discovery of a novel, first-in-class, orally bioavailable azaindole inhibitor (VX-787) of influenza PB2. *J. Med. Chem.* **57**, 6668–6678 (2014).
  119. Byrn, R. A. *et al.* Preclinical activity of VX-787, a first-in-class, orally bioavailable inhibitor of the influenza virus polymerase PB2 subunit. *Antimicrob. Agents Chemother.* **59**, 1569–1582 (2015).
  120. Pflug, A. *et al.* Capped RNA primer binding to influenza polymerase and implications for the mechanism of cap-binding inhibitors. *Nucleic Acids Res.* **46**, 956–971 (2018).
  121. Omoto, S. *et al.* Characterization of influenza virus variants induced by treatment with the endonuclease inhibitor baloxavir marboxil. *Sci. Rep.* **8**, 9633 (2018).
  122. Furuta, Y. *et al.* Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Res.* **100**, 1–23 (2013).
  123. Lakowicz, J. R. Fluorescence anisotropy. in *Principles of fluorescence spectroscopy* 353–381 (Springer, 2006).
  124. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 133–144 (2010).
  125. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
  126. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).
  127. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
  128. Delano, W. L. The PyMOL Molecular Graphics System. *Schrodinger* <http://www.pymol.sourceforge.net> (2002).
  129. Mandal, S. S. *et al.* Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **101**, 7572–7577 (2004).

130. Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. USA* **102**, 18590–18595 (2005).
131. Lukarska, M. *et al.* Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **541**, 117–121 (2017).
132. Crescenzo-Chaigne, B. & van der Werf, S. Rescue of influenza A virus from recombinant DNA. *J. Virol.* **81**, 11282–11289 (1999).
133. Dos Santos Afonso, E., Escriou, N., Leclercq, I., Van Der Werf, S. & Naffakh, N. The generation of recombinant influenza A viruses expressing a PB2 fusion protein requires the conservation of a packaging signal overlapping the coding and noncoding regions at the 5' end of the PB2 segment. *Virology* **341**, 34–46 (2005).
134. Watson, S. J. *et al.* Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120205 (2013).
135. Hu, X. *et al.* A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **103**, 9506–9511 (2006).
136. Hintermair, C. *et al.* Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J.* **31**, 2784–2797 (2012).
137. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
138. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
139. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
140. Ivanov, D., Kwak, Y. T., Guo, J. & Gaynor, R. B. Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**, 2970–83 (2000).
141. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004).
142. Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **11**, 1549–1561 (2003).
143. 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 Dev.* **28**, 1323–1336 (2014).
144. Dahmus, M. E. Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim Biophys Acta* **1261**, 171–182 (1995).
145. Ghosh, A., Shuman, S. & Lima, C. D. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Mol. Cell* **43**, 299–310 (2011).

146. Xiang, K., Manley, J. L. & Tong, L. An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. *Genes Dev.* **26**, 2265–2270 (2012).
147. 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).
148. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **7**, 639–643 (2000).
149. Xiang, K. *et al.* Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733 (2010).
150. Llompart, C. M., Nieto, A. & Rodriguez-Frandsen, A. Specific Residues of PB2 and PA Influenza Virus Polymerase Subunits Confer the Ability for RNA Polymerase II Degradation and Virus Pathogenicity in Mice. *J. Virol.* **88**, 3455–3463 (2014).
151. Serna Martin, I. *et al.* A Mechanism for the Activation of the Influenza Virus Transcriptase. *Mol. Cell* **70**, 1101–1110 (2018).
152. Kawakami, E. *et al.* Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. *J. Virol. Methods* **173**, 1–6 (2011).
153. 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. *J. Virol.* **77**, 5017–5020 (2003).
154. 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. *J. Virol.* **76**, 8989–9001 (2002).
155. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIIS. *Mol. Cell* **16**, 955–965 (2004).
156. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016).
157. Kastner, B. *et al.* GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
158. Wen, Y. & Shatkin, A. J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13**, 1774–1779 (1999).
159. Klein, B. J. B. *et al.* RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. USA* **108**, 546–550 (2011).
160. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648 (2002).
161. Kim, J. B. & Sharp, P. A. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent

- Kinase-activating Kinase. *J. Biol. Chem.* **276**, 12317–12323 (2001).
162. Crickard, J. B., Fu, J. & Reese, J. C. Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *J. Biol. Chem.* **291**, 9853–9870 (2016).
  163. Parada, C. A. & Roeder, R. G. A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. *EMBO J.* **18**, 3688–3701 (1999).
  164. Gonzalez, M. N. *et al.* CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. *Nat. Commun.* **9**, 3392 (2018).
  165. Schuller, R. *et al.* Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol. Cell* **61**, 305–314 (2016).
  166. Ghosh, A., Shuman, S. & Lima, C. D. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Mol. Cell* **43**, 299–310 (2011).
  167. Akkina, R. K., Chambers, T. M. & Nayak, D. P. Mechanism of interference by defective-interfering particles of influenza virus: Differential reduction of intracellular synthesis of specific polymerase proteins. *Virus Res.* **1**, 687–702 (1984).
  168. Rolling, T. *et al.* Adaptive Mutations Resulting in Enhanced Polymerase Activity Contribute to High Virulence of Influenza A Virus in Mice. *J. Virol.* **83**, 6673–6680 (2009).
  169. Mehle, A., Dugan, V. G., Taubenberger, J. K. & Doudna, J. A. Reassortment and Mutation of the Avian Influenza Virus Polymerase PA Subunit Overcome Species Barriers. *J. Virol.* **86**, 1750–1757 (2012).
  170. Wu, N. C. *et al.* Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. *PLoS Genet.* **11**, 1–27 (2015).
  171. Lu, H. *et al.* Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318–323 (2018).
  172. Boehning, M. *et al.* RNA polymerase II clustering through CTD phase separation. *Nat. Struct. Mol. Biol.* **25**, 833–840 (2018).
  173. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
  174. McSwiggen, D. *et al.* Transient DNA Binding Induces RNA Polymerase II Compartmentalization During Herpesviral Infection Distinct From Phase Separation. *BioRxiv* **17**, 2009 (2018).
  175. Reich, S., Guilligay, D. & Cusack, S. suppl An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. *Nucleic Acids Res.* **45**, 3353–3368 (2016).
  176. Coppola, J. A., Field, A. S. & Luse, D. S. Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc. Natl. Acad. Sci. USA* **80**, 1251–1255 (1983).
  177. Listerman, I., Sapra, A. K. & Neugebauer, K. M. Cotranscriptional coupling of splicing

factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* **13**, 815–822 (2006).