The Role of the HSV-1 ICP22 Protein in Regulation of Expression of Host Cell Genes

Nur Firdaus Isa



Linacre College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term, 2017

Abstract

The eukaryotic RNA Polymerase II (Pol II) is responsible for the synthesis of all mRNA and many small regulatory RNAs. The Herpes Simplex Virus 1 (HSV-1) relies exclusively on the host Pol II to synthesise viral RNA during productive infection. A repeating structure on the carboxy terminal domain (CTD) of the Pol II largest subunit consisting of consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, has to be phosphorylated to drive Pol II-mediated transcription. The ICP22 protein of HSV-1 is capable of interacting with P-TEFb, a kinase phosphorylating the Ser2 position of the CTD. Ser2 CTD phosphorylation is a mark associated with Pol II productive elongation. ICP22 has been shown to cause loss of Ser2 CTD phosphorylation.

In the first part of this thesis, I show that transient expression of ICP22 affects the level of transcribing Pol II at the beginning and end of a model gene, which mirrors the effect caused by CDK9 inhibitors. This result suggests that ICP22 possibly inhibits Pol II elongation due to inhibition of the Ser2 CTD kinase, P-TEFb.

In the second part of the thesis, I describe the ICP22 interactome. I performed ICP22 pull-down assays with HeLa cell extract and the cellular protein complexes bound to ICP22 were analysed by mass spectrometry. This work confirmed the interaction between ICP22 and CDK9 and identifies novel proteins that interact, directly or indirectly, with ICP22.

In the last part of the thesis, the interaction of ICP22 and P-TEFb was further characterised. Using recombinant protein of subunits of P-TEFb; CDK9 and Cyclin T1, I demonstrated that ICP22 interacts with both subunits of P-TEFb. In addition, I validated the ICP22 proteomic hits using immunoblots. Here, I reveal that ICP22 possibly interacts with transcriptional kinases other than P-TEFb.

The *in vitro* experiments presented in this thesis involved transiently expressed ICP22 or recombinant proteins ICP22 and VP16 thus eliminating the effects of viral proteins other than ICP22 or VP16. Based on these findings, I have extended our understanding that ICP22 induces alteration of host cellular transcription by inhibiting P-TEFb and possibly other transcriptional kinases. These works serve as a platform for genome-wide analyses of the effect of ICP22 on host gene transcription and structural biology to characterise the interaction of ICP22 and P-TEFb.

Acknowledgements

First, I would like to sincerely thank my supervisor, Professor Shona Murphy for the opportunity to work on this project in her group and for supporting me at every stage of it. She has always been there to talk about everything – work and life – and her support, patience and guidance from my first day to the last have been invaluable. She has been a very influential mentor for my own professional development and I am deeply grateful and honoured for having the chance to pursue a DPhil at the University of Oxford, and in particular in Murphy's Lab. I would also like to thank all the past and present members of the Murphy group; Justyna Zaborowska, Ivan Ferrer-Vicens, Dawn O'Reilly, Michael Tellier; Joana Guiro; Clelia Laitem, and Olga Kuznetsova, for their mentorship, company and laughter. It has been a great joy working together and these lab members have shown me that science is fun.

I am very grateful for the scientific support and materials I have received from various scientists from Dunn School; Professor Susan Lea/Yuhan Fong and Dr Ivan Ahel group for DNA plasmids and technical guidance, Professor Oresto Acuto group for many instruments and reagents and Dr Natalia Gromak group for valuable support. I am also thankful to many scientists outside Dunn School; Svenja Hester/Professor Shabaz Mohammed for the technical advices for my mass spectrometry project and our collaborator Professor Peter O'Hare (Imperial College). My graduate advisor Professor Quentin Sattentau has always been there to provide encouragement and I am grateful for that. I am also grateful for having met all fantastic people during my time in Oxford, both inside and outside the Sir William Dunn School of Pathology. I would like to thank different sources of funding including the MOHE, IIUM and Linacre College for making this work possible.

All my sincerest thanks and appreciation to my parents (Isa Yunus and Siti Halimah Daud) and siblings (Fattah, Faizal, Fathihah, Farhan, Firzanah, Fathurrahman and Farwizah) without whom I would not have come this far. Special gratitude to my own little family; my dear husband Hanis Zulhusmie Mohd Jali and our beautiful daughter Shafiah, that have been with me through thick and thin and stood by me even when they understood nothing about my work. I love you to the moon and back!

Table of Contents

Abstract i

Acknowledgements .....

List of Figures iii

List of Tables iv

Abbreviations v

Chapter 1| Introduction 1

1.1 Chapter abstract 1

1.2 The mechanism of eukaryotic transcription 1

1.2.1 RNA Polymerase II 1

1.2.2 Overview of eukaryotic transcription. 2

1.2.3 The Pol II CTD code 6

1.2.4 Transcriptional cyclin-dependent kinases (CDKs) 9

1.2.5 Dysregulation of CTD kinases 17

1.3 Herpes Simplex Virus 1 20

1.3.1 The human herpesviruses 20

1.3.2 Molecular events during HSV-1 infection 23

1.3.3 Immediate early infected cell protein 22 (ICP22) 27

1.4 HSV-1 infection and changes in host transcriptome 32

1.5 Aims of the thesis 34

Chapter 2| The effect of ICP22 on elongating RNA polymerase II (Pol II) 36

2.1 Chapter abstract 36

2.2 Overview of the *KPNB1* model protein coding gene. 36

2.3 Validation of CDK9 inhibitor treatment on HEK293 cells 37

2.4 Histone occupancy increases in regions where Pol II prematurely terminates 39

2.5 Effect of transient expression of ICP22 on elongating Pol II 41

2.6 Ectopic expression of ICP22 does not affect the KPNB1 protein level 44

2.7 Chapter summary 45

Chapter 3| Characterisation of the interaction proteome of ICP22 47

3.1 Chapter abstract 47

3.2 Proteomic analysis of proteins pulled down by ICP22 47

3.3 Molecular functions associated with ICP22-bound cellular proteins 49

3.4 Confirmation of interactions between ICP22 and transcriptional kinases using mass spectrometry 51

3.5 Chapter summary 54

Chapter 4| Characterisation of the interaction between ICP22 and cellular proteins 56

4.1 Chapter abstract 56

4.2 Expression and purification of His-tag ICP22 recombinant proteins using a bacterial expression system 56

4.3 Expression and purification of GST-CDK9 and Flag-Cyclin T1 recombinant proteins from a bacterial expression system 63

4.4 Direct interactions of ICP22 and CDK9 or Cyclin T1 65

4.5 Sequence description and secondary structure analysis of ICP22 69

4.6 Expression and purification of His-tag VP16 recombinant protein using bacterial expression system 71

4.7 ICP22 and VP16 interact with the active form of P-TEFb 71

4.8 ICP22 but not VP16 interacts with the inactive form of P-TEFb 73
4.9 ICP22 and VP16 interact with the P-TEFb-associated complex PAF1c 74
4.10 ICP22 and VP16 associate with transcriptionally-engaged Pol II 75
4.11 The interaction of ICP22 with factors recruiting P-TEFb to the promoter 76
4.12 ICP22 interacts with kinases other than P-TEFb 77
4.13 Chapter summary 79

Chapter 5| Discussion 82
5.1 The effect of ICP22 on elongating RNA polymerase II (Pol II) (Chapter 2) 82
5.1.1 Future directions 83
5.2 Characterisation of the interaction proteome of ICP22 (Chapter 3) 85
5.2.1 Future directions 86
5.3 Characterisation of the interaction between ICP22 and cellular proteins (Chapter 4) 87
5.3.1 Future directions 89
5.4 Proposed model for the role of ICP22 protein in the regulation of expression of host cell genes 90
5.5 Conclusions 91

Chapter 6| Materials and Methods 92
6.1 Materials 92
6.1.1 Bacteria and mammalian cell lines 92
6.1.2 Plasmids 92
6.1.3 DNA oligonucleotides 93
6.1.4 Non-commercial buffers 93
6.1.5 Commercial buffers, enzymes and kits 93
6.1.6 Antibodies 93
6.2 Bacterial techniques 99
6.2.1 Transformation of competent bacterial cells 99
6.2.2 Growth of bacterial starter culture 100
6.2.3 Preparation of plasmid DNA 100
6.3 Mammalian cell culture techniques 100
6.3.1 Propagation of mammalian cell lines 100
6.3.2 Pharmacological treatments of mammalian cell culture 101
6.3.3 Plasmid transfections 101
6.3.4 Freezing cells 101
6.4 DNA analysis methods 101
6.4.1 Quantitation of DNA 101
6.4.2 Agarose gel electrophoresis 101
6.4.3 DNA sequencing 101
6.4.4 Polymerase chain reaction (PCR) 102
6.4.5 Real-time quantitative PCR 102
6.5 Protein analysis methods 103
6.5.1 Preparation of whole cell extract 103
6.5.2 Quantitation of protein concentration 103
6.5.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE). 104
6.5.4 Coomassie staining 104
6.5.5 Immunoblotting 104
6.5.6 Protein co-immunoprecipitation 105
6.5.7 Chromatin immunoprecipitation (ChIP) 105
6.5.8 Protein expression and purification 106
6.5.9 Pull-down of HeLa cell extract 108
6.5.10 Mass spectrometry analysis 108

Appendix 1 109

Appendix 2 ..... **110**
Appendix 3 ..... **111**
Appendix 4 ..... **112**
Appendix 5 ..... **115**
Appendix 6 ..... **118**
Appendix 7 ..... **119**
Appendix 8 ..... **122**
Appendix 9 ..... **123**
Appendix 10 ..... **124**
Appendix 11 ..... **125**
Appendix 12 ..... **126**
References **129**

List of Figures

Figure	Figure title	Page
Chapter 1		
1.1	The pattern of phosphorylation of the Pol II CTD across mammalian protein-coding genes.	8
1.2	The transcriptional CDKs.	11
1.3	Virion architecture of herpesviruses.	22
1.4	The life cycle of HSV.	25
1.5	Sequential expression of HSV genomes.	26
1.6	The genomic layout of HSV-1 and ICP22.	28
Chapter 2		
2.1	Schematic of the *KPNB1* gene.	37
2.2	The results of qRT-PCR analysis of Pol II ChIP performed on HEK293 cells treated with DMSO as control or treated with CDK9 inhibitor KM.	38
2.3	The qRT-PCR analysis of H3 and H2B ChIPs performed on HEK293 cells treated with DMSO or treated with the CDK9 inhibitor KM.	40
2.4	Immunoblot analysis of HEK293 cell extract after transfection of a vector expressing Myc-tag residues 193-256 of ICP22 (Myc-SCR).	41
2.5	The results of qPCR analysis of Pol II ChIP after 7 h of transfection of pcDNA3 or Myc-SCR of ICP22 in HEK293.	43
2.6	The results of qPCR analysis of Pol II ChIP after 6 h of transfection or 8 h of transfection of pcDNA3 or Myc-SCR of ICP22 in HEK293.	44
2.7	Immunoblot analysis of KPNB1.	45
Chapter 3		
3.1	Qualitative analysis of ICP22 proteomics.	49
3.2	Graphical representations of bioinformatics analysis of ICP22-interacting proteins classified by GO terms in molecular function.	51
3.3	Visual representation of the transcriptional kinases complexes interacting with ICP22.	53
Chapter 4		
4.1	Expression, purification and identification of His-SCR and His-SCR-5/11.	58
4.2	Coomassie stained gels of the different stages of purification.	60
4.3	Expression of the *S. flexneri* protein IpgC .	62
4.4	Expression, detection and purification of GST-CDK9 and Flag-Cyclin T1 (residues 1-266).	64
4.5	ICP22 interacts directly with CDK9.	65
4.6	ICP22 residues 198-245 (His-SCR-5/11) fail to pull-down Cyclin T1.	67
4.7	Sequence alignment of core conserved region of ICP22 of HSV-1 and other herpesviruses, and HIV-1 Tat and HSV-1 VP16 proteins.	69
4.8	Secondary structure analysis of ICP22.	70
4.9	Identification of expressed VP16 residues 177-224 (His-vCon).	71
4.10	ICP22 and VP16 interact with P-TEFb.	73
4.11	ICP22 but not VP16 interacts with inactive form of P-TEFb.	74
4.12	Both ICP22 and VP16 interact with subunits of PAF1c.	75
4.13	ICP22 and VP16 interact with SPT5 but not with Pol II.	76
4.14	ICP22 and VP16 does not interact with MED1 and Tat-SF1.	77
4.15	ICP22 and VP16 are able to interact with transcriptional kinases other than CDK9.	79
Chapter 5		
5.1	Model of the effect of ICP22 on transcription of host cell genes.	91

List of Tables

Table	Table title	Page
Chapter 1		
1.1	List of transcriptional kinases.	9
1.2	Subfamilies of human herpesvirus (HHV).	20
1.3	Known ICP22 cellular binding protein.	31
Chapter 3		
3.1	Table listing the number of ICP22-interacting proteins (gene count) found in a molecular function as classified by GO terms.	50
3.2	List of ICP22-interacting proteins of interest categorised into functional groups.	52

Abbreviations

bp	Base pair
BSA	Bovine serum albumin
ChIP	Chromatin immunoprecipitation
CTD	Carboxy terminal domain
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage Stimulation Factor
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
DRB	5,6-dicloro-1-β-D- ribofuranosylbenzimidazole
DSIF	DRB-sensitivity-inducing factors
E.coli	*Escherichia coli*
EEC	Early elongation checkpoint
e.g.	Example give
EDTA	Ethylenediaminetetraacetic acid
g	g-force
GFP	Green fluorescence protein
GST	Glutathione S-transferase
h	Hour(s)
H2B	Histone H2B
H3	Histone H3
i.e.	That is
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
kDa	Kilodalton
KM	KM05283
nm	Nanometre
min	Minute(s)
mRNA	Messenger RNA
NELF	Negative elongation factor
NLS	Nuclear localisation signal
nt	Nucleotides
PAF1c	RNA polymerase-associated factor 1 complex
PAS	Polyadenylation signal
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PMSF	Phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
P-TEFb	Positive transcription elongation factor
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid

RNase	Ribonuclease
s	Seconds
SEC	Super elongation complex
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
Tat	Trans-Activator of Transcription, HIV protein
TSS	Transcription start site

Chapter 1| Introduction

1.1 Chapter abstract

The last decade has witnessed growing interest in the molecular biology of viral proteins in order to understand how they interact and interfere with the host cell functions to promote viral replication. The work presented in this thesis relates to the Herpes Simplex Virus-1 (HSV-1) ICP22 protein and its role in downregulation of host cell gene expression at the level of transcription. The first part of the introduction chapter describes the mechanism of eukaryotic gene transcription, highlighting the role of transcriptional kinases. Next, I review the general properties of herpesvirus and its life cycle. Finally, the literature relating to the multiple functions of ICP22 is reviewed, emphasizing the role of ICP22 in the inhibition of host cell gene transcription.

1.2 The mechanism of eukaryotic transcription

1.2.1 RNA Polymerase II

The three different eukaryotic RNA polymerases, designated Pol I, II and III, each transcribe a specific set of genes: RNA Pol I synthesizes precursor ribosomal RNAs (pre-rRNAs) that are processed into 28S, 5.8S and 18S rRNAs (Goodfellow and Zomerdijk 2013). RNA Pol II transcribes protein-coding genes, and also small regulatory RNAs particularly small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and micro RNA (miRNA) (Borchert et al., 2006;Egloff and Murphy 2008). RNA Pol III transcribes non-protein coding (nc)RNA genes such as transfer RNAs (tRNAs) and 5S rRNA (Dieci et al., 2013). The work in this thesis focuses on the transcription of protein-coding genes by Pol II.

Human Pol II is a 12-subunit complex composed of subunits Rpb1-Rpb12. The largest subunit of Pol II, Rpb1 has an unique structurally-disordered carboxy terminal domain (CTD), comprising heptapeptide repeats of Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, which are highly conserved in eukaryotes (Corden 1990;McCracken et al., 1997;Corden 2013). The number of repeats varies from one organism to another and correlates with the genomic complexity of the organism. The human Pol II CTD contains 52 repeats, the *Drosophila* CTD has 44-46 repeats, while yeasts contains approximately 26 (different in *Saccharomyces cerevisiae* and *Schizosaccharomyces*

pombe) (Corden 1990). The Pol II CTD is essential to support life. Studies in *S. cerevisiae, S. pombe* and mammalian cells indicate that cells with mutated residues within CTD heptad repeats are viable but experience growth defects (Egloff et al., 2007;Schwer and Shuman 2011;Zhang et al., 2012). The CTD is not necessary for gene transcription *in vitro*, but deletion of CTD in cells is lethal (Zehring et al., 1988;Corden 2013).

Early studies of the Pol II CTD were focused on identifying the factors interacting with the Pol II CTD as the CTD was proposed to act as scaffold for the dynamic interaction of factors during the transcription cycle (Laybourn and Dahmus 1990;Yuryev et al., 1996). The idea that post-translational modifications of the CTD determine the regulatory role of Pol II CTD in transcription was developed after the Pol II CTD was revealed to undergo phosphorylation *in vivo* (Bell et al., 1976). The presence of five out of the seven residues (Tyr1, Ser2, Thr4, Ser5, and Ser7) as potential phosphate acceptors raised the possibility of complex post-translational modification (Buratowski 2009;Hsin et al., 2011;Egloff et al., 2012). Here, phosphorylated residues of CTD are indicated by a P, for example Ser2P denotes phosphorylated Ser2 of the Pol II CTD. All the potential phospho acceptors in the CTD repeats have now been shown to undergo phosphorylation *in vivo* (Zaborowska et al., 2016). Pol II exists in two different forms as a result of post-translational modification; Pol IIa is a hypophosphorylated form and Pol IIo is a hyperphosphorylated form (Bartholomew et al., 1986;Corden 1990). Both can be distinguished on SDS-PAGE with running size of ~250 kDa for Pol IIa and ~260 kDa for Pol IIo.

1.2.2 Overview of eukaryotic transcription.

The regulation of gene transcription was classically viewed as simplified stages starting with the loading of Pol II onto the DNA template. Pol II is then cleared from the promoter and the transcript elongates as the DNA double helix unwinds. Finally, Pol II pauses upon encountering terminator sequences resulting in the release of the nascent transcripts and Pol II. However, the regulation of Pol II-mediated gene transcription is more complex than previously thought. Each individual stage of the Pol II transcription is highly regulated and involves a variety of factors to ensure efficient transcription and co-transcriptional RNA processing leading to successful

completion of RNA synthesis. An overview of eukaryotic transcription is given in this section while the post-translational modifications of Pol II CTD are reviewed Section 1.2.3.

Pre-initiation complex formation (PIC) and initiation

The first stage of transcription in higher eukaryotes is the formation of pre-initiation complex (PIC) at the promoter. Pol II alone is unable to recognise promoter sequence and therefore requires a range of accessory factors known as General Transcription Factors (GTF). These comprise TFIIA, TFIID, TFIIB, TFIIF, TFIIE, and TFIIH that form a PIC together with Pol II, which allows Pol II to bind to the core promoter elements. The core promoter elements include the TATA box, initiator element (Inr), TFIIB recognition element (BRE), downstream core promoter element (DPE) and the motif ten element (MTE). PIC formation is controlled by a combination of *cis*-acting DNA elements (such as enhancers, promoter-proximal elements and silencers) that recruit a variety of activators and repressors that either directly or indirectly through chromatin modification increase or decrease PIC formation and thereby control the amount of initiation that occurs. These activators and repressors allow a spatial and temporal control of transcription initiation which varies between genes (Heintzman and Ren 2009;Baumann et al., 2010). Following the formation of the PIC, TFIIH and TFIIE melt the DNA duplex, which allows access of Pol II to the single-stranded DNA. PIC assembly can also be stimulated by the Mediator complex which has additional roles in transcription elongation. Transcription initiation usually first entails the production of short, 'abortive' transcripts up to 15 nucleotides long, until Pol II clears the promoter and enters productive elongation as Pol II firmly associates with the DNA template (Kugel and Goodrich 2000;Saunders et al., 2006).

Promoter-proximal pausing and elongation

The next phase is transcription elongation, which can be divided into three stages; Pol II pausing, pause release and productive elongation. After Pol II has escaped the promoter and before Pol II begins transcribing the body of the gene, it is subjected to a promoter-proximal pause approximately ~30-100 bp downstream of the transcription start site (TSS), often referred as an early elongation checkpoint (EEC). Pol II is arrested at the EEC by the recruitment of two negative transcription elongation factors (N-TEFs); the DRB-sensitivity-inducing factors (DSIF), which is

composed of two subunits, SPT4 and SPT5, and negative elongation factor (NELF), which is composed of four subunits, A, B, C/D, E (Shilatifard 1998;Sims et al., 2004). During the pause, SPT5 interacts with the capping enzyme, which adds the 7-methylguanylate cap to the 5' end of the nascent RNA (Lidschreiber et al., 2013). To release the pause, the CDK9 subunit of positive transcription elongation factor (P-TEFb) phosphorylates SPT5 and NELF. Phosphorylation of the NELF-E subunit allows dissociation of NELF-E from Pol II, whereas phosphorylation of SPT5 and Ser2P CTD Pol II by P-TEFb permits the release of Pol II from the checkpoint to resume productive elongation (Yamada et al., 2006).

Although previously it was thought that the transcription is regulated primarily at the stage of initiation, accumulating evidence suggests that transcription elongation is also a major point of regulation. Productive elongation is a tightly-regulated process that requires a multitude of co-factors to ensure the smooth transition of Pol II past the EEC. Pol II promoter-proximal pausing is likely to be the rate limiting step in transcription of many genes in Drosophila and mammals (Core and Lis 2008;Boettiger et al., 2011;Jonkers et al., 2014;Day et al., 2016). Pol II pausing at the EEC is also likely to be critical to ensure the precise coordination of factors necessary for productive elongation (Levine 2011;Adelman and Lis 2012). This mechanism may also be a way to allow the rapid and synchronous expression of genes when stimulated, as can be seen in developmental control genes (Levine 2011;Adelman and Lis 2012;Gaertner and Zeitlinger 2014). Cellular factors are required to recruit P-TEFb to relieve the effect of N-TEFs and activate productive elongation. For example, c-myc, NF-κB, and HIV-1 Tat in the case of HIV infection regulate transcription by P-TEFb recruitment (Kwak and Lis 2013). P-TEFb is therefore essential to overcome the influence of these negative elongation factors at the EEC and promote productive elongation. Elongation of Pol II through the first few kilobases of a gene is rather slow, approximately ~0.5 kb per min, perhaps due to the rate of assembly of elongation factors or the incomplete removal of NELF, but Pol II speeds up to ~2-5 kb per min through the gene body (Jonkers and Lis 2015). The elongation rate of Pol II affects the level of mRNA produced and varies between genes (Jonkers and Lis 2015).

Termination and 3'end formation

At the end of transcription cycle, once the Pol II has transcribed the polyadenylation (poly(A)) signal (PAS), the pre-mRNA undergoes endonucleolytic cleavage and is polyadenylated (Shi et al., 2009;Millevoi and Vagner 2010;Proudfoot 2011). The termination process begins with the recognition of the AAUAAA sequence by the Cleavage and Polyadenylation Specificity Factor (CPSF) and the GT-rich downstream sequence element (DSE) by the Cleavage Stimulation Factor (CstF). Cleavage occurs ~10-35 bp downstream of the A(A/U)UAAA sequence in the nascent RNA and requires two additional proteins; Cleavage Factor I and II (CFI and CFII). The pre-mRNA is cleaved by the 73 kDa subunit of CPSF (CPSF73) (Shi et al., 2009;Millevoi and Vagner 2010). The free 3' end is modified with ~200-300 adenosines catalysed by the polyA polymerases (PAPs) (Proudfoot 2004;Bentley 2005). In some cases, transcription termination occurs when the exonuclease Xrn2, recruited by CPSF after nascent RNA cleavage at the poly(A) site, degrades the uncapped RNA, progresses towards the elongating Pol II and causes dissociation of Pol II from the DNA template (Shi et al., 2009;Millevoi and Vagner 2010). Pol II eventually dissociates from the DNA template and is recycled for the next round of transcription.

By treating human HeLa cells with CDK9 pharmacological inhibitors, we have recently shown that there is another point of elongation control near the PAS, termed the poly(A)-associated checkpoint. Pol II is unable to go to through this checkpoint and prematurely terminates when cells are treated with CDK9 inhibitors (Laitem et al., 2015) (Appendix 1). The poly(A)-associated checkpoint will be further described in Chapter 2.

Co-transcriptional RNA processing

While most of this thesis focuses on transcription, co-transcriptional RNA processing is also critical for the generation of a mature mRNA. The discovery of mRNA processing factors as CTD binding proteins suggests that the role of the Pol II CTD goes beyond regulating the Pol II transcription cycle. RNA processing involves 5'capping of the nascent RNA, intron splicing, 3' cleavage and polyadenylation and RNA editing (McCracken et al., 1997;Laurencikiene et al., 2006;Shi et al., 2009;Millevoi and Vagner 2010;Bentley 2014). Shortly after transcription initiation, the synthesized nascent RNA molecule of about ~20-30 nucleotides is capped with a 7-methylguanosine cap by the capping enzyme to protect the nascent RNA from

degradation and activate translation (Bentley 2014). Introns are excised in a process called splicing by the spliceosome, a large complex comprising five small nuclear RNAs (snRNAs); U1, U2, U3, U4 and U6, and a large number of proteins (Kornblihtt et al., 2004;Hoskins and Moore 2012). Through alternative splicing, the spliceosome can produce a number of different mRNAs from one nascent transcript. At the end of the transcription cycle, the PAS of the pre-mRNA is recognised by a large cleavage/polyadenylation machinery to cleave the pre-mRNA and add a poly(A) tail to the 3' end of the transcript (Proudfoot 2011). The poly(A) tail is critical for mRNA stability, cytoplasmic translocation and translation (Ahn et al., 2004;Shi et al., 2009;Millevoi and Vagner 2010;Tellier et al., 2016). The Pol II CTD plays an essential role in coordinating the recruitment of different regulatory factors efficiently during the coupling of transcription and co-transcriptional RNA processing (Phatnani and Greenleaf 2006;Hsin and Manley 2012).

1.2.3 The Pol II CTD code

The five phospho acceptor residues of the Pol II CTD repeat are phosphorylated during transcription (Tyr1P, Ser2P, Thr4P, Ser5P, and Ser7P). The role of the CTD kinases is presented in Section 1.2.4. The CTD code hypothesis postulates that the phosphorylation pattern of residues of Pol II CTD, which undergoes a series of changes throughout the transcription cycle is read as a code to bring about downstream events (Buratowski 2009). The pattern of CTD phosphorylation not only reflects the Pol II activity on the gene body but also recruits specific factors required for RNA processing at the right stage of transcription. Figure 1.1 depicts a simplified model of how the CTD code works on human protein-coding genes. During the early stage of transcriptional initiation, the Pol II with a hypophosphorylated form of the CTD is recruited into the PIC at the promoter. Pol II CTD is then phosphorylated at the Ser5P and Ser7P by the CDK7 subunit of TFIIH. This CTD phosphorylation stops the interaction of Pol II with Mediator and TBP, which may help Pol II to escape the promoter (Sogaard and Svejstrup 2007). The Ser5P Pol II CTD is also recognised by the capping factors, which helps the recruitment of the capping machinery to the 5' end of nascent RNA (McCracken et al., 1997;Nojima et al., 2015). Soon after leaving the promoter, Pol II encounters a promoter-proximal pause at the early-elongation checkpoint (EEC), located approximately ~30-100 bp downstream of the TSS, as a result of the DSIF and NELF (Shilatifard 1998;Sims et al., 2004). The CDK9 subunit

of P-TEFb phosphorylates the SPT5 subunit of DSIF, the NELF-E subunit of NELF and Ser2 of the Pol II CTD, which all together enables the transition of Pol II into the elongation phase. Recent mass spectrometry and *in vitro* phosphorylation experiments suggest that Ser7P is required to prime the CTD to be recognised by P-TEFb to perform its kinase activity and that CDK9 also phosphorylates Ser5P of the Pol II CTD (Czudnochowski et al., 2012). Both Ser5P and Ser2P in the early productive elongation steps are critical for recruitment of important splicing factors and play major role in splicing (Gu et al., 2013;Nojima et al., 2015). As the elongating Pol II approaches the 3' end of the gene, Ser5P is reduced while Ser2P and Thr4P remain high. High Ser2P in addition to the PAS are recognised by several factors involved in 3' end processing (Ahn et al., 2004;Davidson et al., 2014;Tellier et al., 2016). In addition to the recruitment of cleavage and polyadenylation factors to the nascent RNA, the elevated levels of Ser2P at the 3' end of genes also promotes recruitment of polyadenylation factors to Pol II to cleave the nascent RNA and termination factors that then release Pol II from the gene (Kornblihtt et al., 2004;Buratowski 2009;Harlen and Churchman 2017). The Ser2P and Thr4P patterns may be similar towards the 3' end of the genes because P-TEFb can phosphorylate Thr4P as well as Ser2P; Thr4P may also facilitate 3' end processing, as observed on histone mRNA in chicken cell line DT40 (Hsin et al., 2011). Recent structural and proteomics analyses performed in *S. cerevisiae* suggest that Thr4P and Ser2P facilitate termination by recruiting termination factor Rtt103 (Nemec et al., 2017). Tyr1P is implicated in the Pol II transcription cycle but is the least understood of the CTD marks in mammals (Hsin et al., 2014). Loss of Tyr1P marks may be associated with the completion of a transcription cycle, and Tyr1P is associated with antisense promoter transcription and active enhancers (Heidemann and Eick 2012;Descostes et al., 2014;Harlen and Churchman 2017). After transcription termination, the CTD is thought to be dephosphorylated and the hypophosphorylated form of CTD is ready for the next round of transcription (Hsin and Manley 2012).

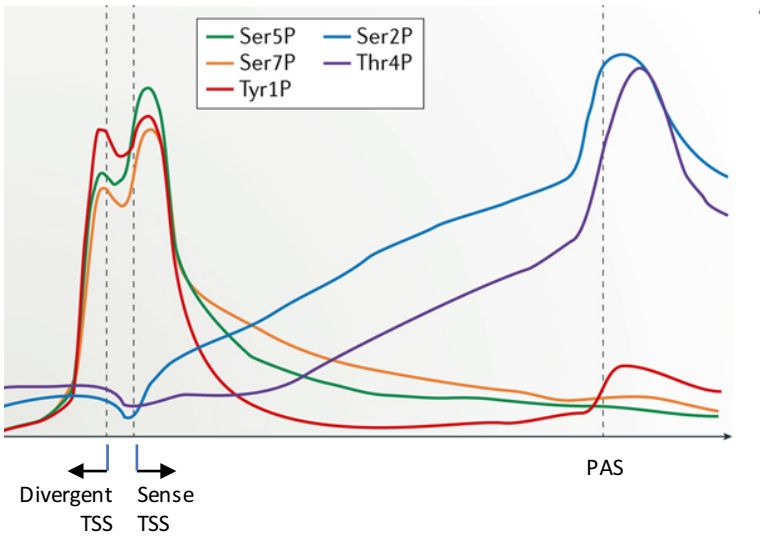


Figure 1.1. The pattern of phosphorylation of the Pol II CTD across mammalian protein-coding genes. An increase in Ser5P, Ser7P and Tyr1P occurs as transcription progresses from the transcription start sites (Sense TSS and Divergent TSS). As Pol II negotiates through the early-elongation checkpoint, which is located approximately ~30-100 bp downstream of TSS, and enters productive elongation Ser2P and Thr4P increase, and Ser5P, Ser7P and Tyr1P decrease as Pol II actively elongates through the gene body. As Pol II reaches the poly(A) signal (PAS), which signals transcription termination at the 3' end of gene, the Ser2P and Thr4P peak. Figure adapted from (Harlen and Churchman 2017).

The dramatic changes in CTD phosphorylation pattern during the transcription cycle illustrate that CTD modification is a highly dynamic process, involving multiple kinases and phosphatases perfectly orchestrated to give the necessary transitions from one stage to another. Studying the CTD code presents many challenges to researchers, because the specific pattern of phosphorylation within the multiple CTD repeats during each transcription cycle is not entirely clear. This is further complicated by the use of different methods such as *in vitro* or *in vivo* studies or different antibodies among the different laboratories studying the CTD. However, although this model is highly simplified, it is generally accepted as a representative pattern of CTD phosphorylation for the majority of mammalian genes. Now, we have seen how the pattern of Pol II CTD phosphorylation changes as transcription progresses, but the key players involved in modifying the CTD are yet to be introduced. A multitude of CTD kinases, phosphatases and isomerases along with many other enzymes are essential for the control of Pol II activity. The role of phosphatases and many other CTD-modifying enzymes are not explored in this study. Below, the role of transcriptional kinases that help to write the CTD code is reviewed.

1.2.4 Transcriptional cyclin-dependent kinases (CDKs)

CTD phosphorylation events can be investigated through the activity of transcription-associated cyclin-dependent kinases (CDKs). The transcriptional CDKs that are able to phosphorylate residues at specific positions within the consensus CTD heptad repeat are referred to as CTD kinases. Most, but not all, transcriptional CDKs known to date are the CTD kinases. Pol II CTD phosphorylation by the CTD kinases creates sophisticated modification states of the CTD as previously described. The CDKs are a family of serine–threonine kinases, which are particularly studied for their essential roles in cell proliferation and gene transcription (Malumbres 2014). The CDK family members 1-6 and 14-16 drive various phases of the cell cycle while CDK family 7-13 and 19-20 which are the focus of this thesis, are involved in the regulation of transcription. The CDKs require binding to their cyclin partners for activation of their kinase activity. Table 1.1 describes each transcriptional kinase, its cyclin-binding motifs, cyclin partner and its kinase activity on the Pol II CTD. The transcriptional CDKs, unlike their cell cycle counterparts, form complexes with cyclin family members that are not expressed periodically in a cell-cycle-dependent manner. The transcription-associated cyclins are therefore active throughout the cell cycle (Malumbres 2014). In addition to the cyclin, the CDK is regulated by phosphorylation and this modification can lead either to activation or inhibition. The activating or inhibitory phosphorylation sites are shown in Figure 1.2.

Table 1.1. List of transcriptional kinases

CDKs	Cyclin binding domains	Activating cyclin	CTD modifications
CDK7	NRTALRE	H (Morgan and De Bondt 1994)	Ser5, Ser7 (Akhtar et al., 2009)
CDK8	SMSACRE	C (Tassan et al., 1995)	Ser5, Ser2 (Rickert et al., 1999)
CDK9	PITALRE	T1, T2a, T2b (Peng et al., 1998) K*(Baek et al., 2007)	Ser2, Thr4, Ser5, Ser7 (review (Zaborowska et al., 2016)
CDK10	PISSLRE	M (Guen et al., 2017)	?
CDK11	PITSLRE	L1, L2 (Loyer et al., 2008) D3 (Zong et al., 2007)	Ser2, Ser5 (Pak et al., 2015)
CDK12	PITAIRE	K (Kohoutek and Blazek 2012)	Ser2, Ser5 (Bartkowiak et al., 2010)

CDK13	PITAIRE	K (Kohoutek and Blazek 2012)	Ser2, Ser5 (Greifenberg et al., 2016)
CDK19	SMSACRE	C (Li et al., 2014)	?
CDK20	PNQALRE	H (Malumbres 2014)	?

Note: *Although a crystal structure of P-TEFb/Cyclin K has been determined, whether Cyclin K can act as a bona fide CDK9 activating cyclin is uncertain. Cyclin K could bind to CDK9 under certain conditions e.g. during developmental stages (Baek et al., 2007;Kohoutek and Blazek 2012).

CTD kinases are usually part of larger complex for example, CDK7 is a kinase component of TFIIH, CDK8 is part of the Mediator complex, and the CDK9 is a part of the Super Elongation Complex (SEC). The structures of CDK9/Cyclin T (Baumli et al. 2008), the inactive confirmation of CDK7 (Lolli et al. 2004), CDK8/Cyclin C (Schneider et al., 2011), CDK12/Cyclin K (Bosken et al., 2014) and CDK13/Cyclin K (Greifenberg et al., 2016) have been solved. Most transcriptional CDKs are very similar in sequence and structure arrangement, having their ATP-binding sites in between the N- and C-terminal lobes, a hinge region that connects the two lobes, cyclin-binding domain, and a similar activation segment as their general features (Figure 1.2). The cyclin binds with the N-terminal lobe of the CDK while the C-terminal lobe of the CDK contains the motif recognising the substrates (Malumbres 2014). In general, the activation of transcriptional CDKs requires phosphorylation of a threonine on the activation segment which induces conformational changes. The conformational changes allow residues within the ATP binding site to bind ATP, and creates a binding platform for the substrate to bind around the site for phospho transfer reactions (Johnson et al., 1998;Brown et al., 1999). The ATP binding site of the activated kinase contains two different clefts (front and back clefts) and a gatekeeper residue. For example the back cleft of CDK9 has a bulky Phe-103 that makes the active site inaccessible to anything bigger than ATP. The front cleft has three main pockets; the A pocket (adenine binding pocket) which is often targeted by small molecule inhibitors, the R pocket (ribose binding pocket) and P pocket (site for phosphate) (Zhang et al. 2009).

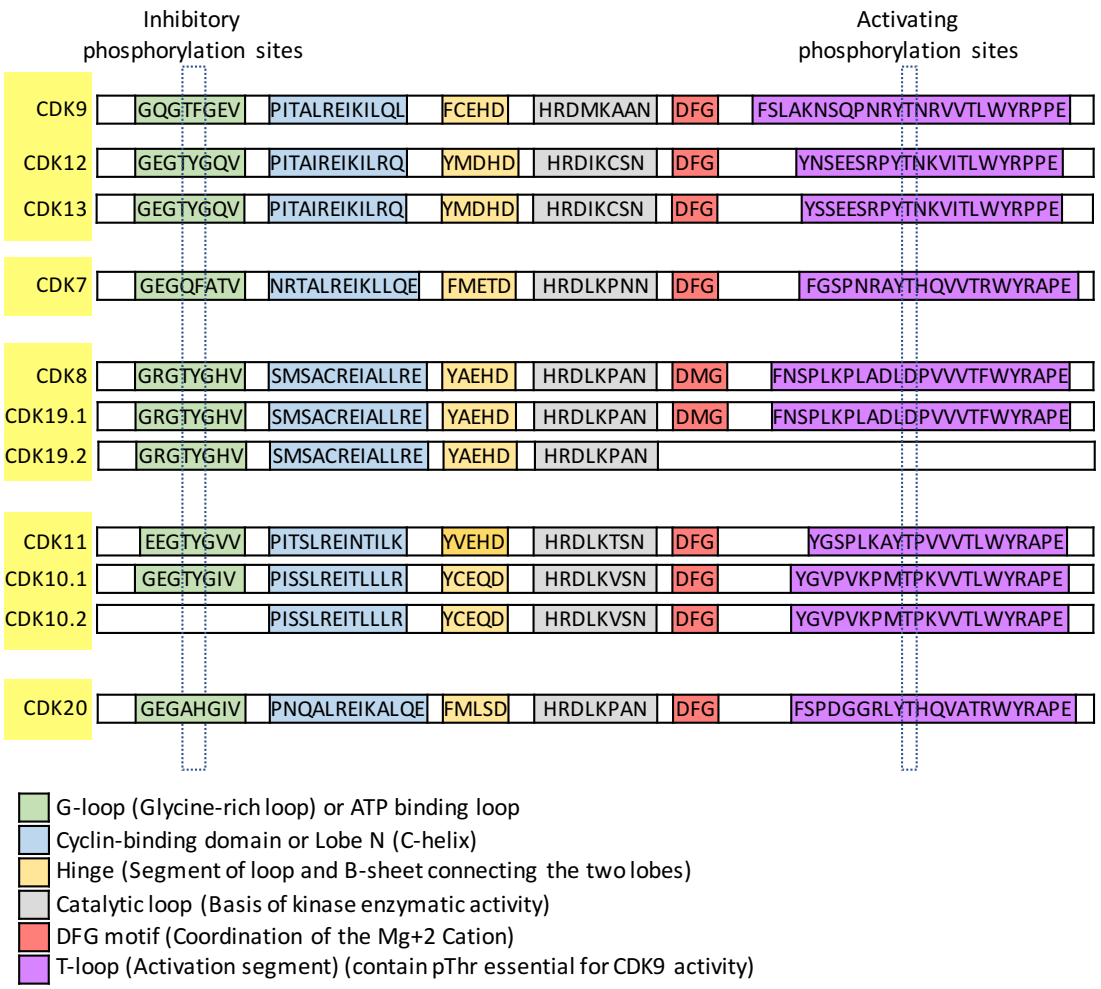


Figure 1.2. The mammalian transcriptional CDKs. The transcriptional CDKs are divided into five subfamilies (CDK9, CDK7, CDK8, CDK11, CDK20 subfamilies; marked in yellow) on the basis of the evolutionary divergence. Each transcriptional CDK is aligned corresponding to the functional domain and amino acid sequence of CDK9. The presence of principal structural elements is represented in colour as indicated in the legend. The start of the T-loop begins with HRD domain in catalytic loop (grey) and ends with PPE domain (purple). The aspartate in the DFG motif is required for the coordination of Mg+2/ATP. The residues within the dotted boxes are targets of phosphorylation, which causes either inhibition or activation. The length of the transcriptional CDKs is not to scale. Figure adapted from (Lim and Kaldis 2013;Malumbres 2014;Paparidis et al., 2017)).

Most transcriptional CDKs have been shown to be capable of phosphorylating the Pol II CTD. CDK7, CDK8 and CDK9 are the three transcriptional CDKs with the best understood roles as CTD kinases; CTD Ser5 and Ser7 phosphorylation by CDK7 is often associated with initiation of transcription, while CTD Ser2 phosphorylation by CDK9 is linked to the ability of P-TEFb to overcome Pol II promoter-proximal pausing and promote Pol II productive elongation (Shilatifard 2004;Sims et al., 2004;Laitem et al., 2015). Since the phosphorylation of the Pol II CTD is involved in

nearly every stage of transcription, it is likely that transcriptional CDKs other than CDK7 and CDK9 have equally important roles in the various steps of transcription.

The role of CDK7 in transcription initiation

CDK7 was originally described as a CTD Ser5 kinase at mRNA gene promoters (Roy et al., 1994;Akoulitchev et al., 1995). A number of recent experiments suggest that CDK7 is able to phosphorylate Ser7 at both mRNA and snRNA genes in budding yeast (Akhtar et al., 2009;Glover-Cutter et al., 2009). As a subunit of the general transcription factor TFIIH, CDK7 dominates the control of transcription at initiation through Ser5 and Ser7 phosphorylation, which marks the early phase of transcription initiation, as shown in the previous model (Figure 1.1). During the early phase of transcription initiation, the hypophosphorylated CTD Pol II binds factors of the PIC such as the TATA-binding protein (TBP) and the Mediator complex. The Mediator complex is involved in the assembly and coordination of the transcription machinery at the promoter. Shortly after this, CDK7 phosphorylates Ser5 and Ser7, disrupting Pol II interaction with the PIC and the Mediator complex. This results in Pol II initiation while the scaffold complex remains at the promoter. Phosphorylation of Ser5 by CDK7 is recognised by the capping enzyme that adds the 7-methylguanisine cap structure to the nascent RNA shortly after initiation (when the RNA is about ~25-30 nt in length) (McCracken et al., 1997;Nojima et al., 2015). CDK7 is able to phosphorylate and activate other CDKs, including CDK9 (Ganuza and Santamaria 2012;Larochelle et al., 2012;Schachter and Fisher 2013). In addition to CDK7, the CDK8 subunit of Mediator is also an important kinase at the stage of transcription initiation.

CDK8 (and its paralogue CDK19)

During the initial phase of transcription initiation, Ser5 Pol II CTD can be phosphorylated by another component of PIC scaffold, the CDK8 subunit of Mediator *in vitro* (Tietjen et al., 2010). However, to what extent CDK8 contributes to the phosphorylation of CTD *in vivo* is unclear. Human core Mediator is a 26-subunit complex that is the main interface between PIC and Pol II during the initial stages of transcription initiation (Allen and Taatjes 2015). CDK8/Cyclin C combined with MED12 and MED13 forms a four-subunit kinase (known as CDK8 module) that is associated with the core Mediator, which makes contact with hypophosphorylated Pol

II and recruits it to the promoter (Allen and Taatjes 2015). The CDK8-Mediator complex may also promote Pol II pause release and elongation via recruitment of the Super Elongation Complex (SEC) (Takahashi et al., 2011). Thus, CDK8 is implicated not only during transcription initiation but also in telongation. CDK8 seems to function in context-specific ways and can have both positive and negative effects on transcription (Akoulitchev et al., 2000;Galbraith et al., 2010;Allen and Taatjes 2015). The dual functionality of CDK8 is attributed to cell-type specificity and its function is thought to vary in response to distinct stimuli (Galbraith et al., 2010;Allen and Taatjes 2015). Some cancer cells express low levels of CDK8, while others express high levels of CDK8. Osteosarcoma cells (SJSA) lack endogenous CDK8 (Clark et al., 2015). Little is known about the paralogue of CDK8, CDK19 and its importance in the regulation of transcription.

The CDK9 subunit of P-TEFb as a stimulator of Pol II pause release

CDK9, a subunit of positive transcription elongation factor b (P-TEFb), is the kinase that allows progression of Pol II from promoter-proximal pausing to productive elongation by phosphorylating the negative transcription elongation factors NELF and DSIF. The ability of CDK9 to promote productive elongation is further substantiated by studies demonstrating the critical role of P-TEFb in Tat-mediated transactivation of HIV-1 gene expression (Tahirov et al., 2010;Gu et al., 2014). HIV-1 Tat forms a trimeric complex of Tat-CDK9-Cyclin T1 that is recruited to the transactivation response element (TAR). The Super Elongation Complex (SEC) subsequently assembles. CDK9 stimulates Pol II elongation from the HIV-1 promoter by phosphorylating NELF, DSIF and Ser2 Pol II CTD, thus allowing production of viral transcripts (Tahirov et al., 2010;Gu et al., 2014).

In mammalian cells, the *CDK9* gene gives rise to two isoforms, $CDK9_{42}$ and $CDK9_{55}$, with $CDK9_{55}$ having a longer N-terminal domain (Shore et al., 2003;Liu and Herrmann 2005). The ratio of $CDK9_{42}$/$CDK9_{55}$ is cell type-dependent and to date it is not clear if there are functional differences between the two isoforms (Liu and Herrmann 2005;Shore et al., 2005). The association of these isoforms with Cyclins T1, T2a, T2b and K gives rise to the formation of eight structurally and perhaps functionally distinct P-TEFb complexes (Baek et al., 2007). However, mass spectrometry has failed to show that Cyclin K is associated with CDK9 (Bezstarosti et al., 2010). Therefore, Cyclin T1 and T2 may be the major cyclins associated with

CDK9. The activity of CDK9 is regulated in several ways including phosphorylation, autophosphorylation, acetylation and by forming a complex with other factors. CDK9 autophosphorylates at several sites including within its activation segment, in particular Thr-186 (Baumli et al., 2008). Thr-186 of CDK9 can also be phosphorylated by CDK7 (Larochelle et al., 2012). The phosphorylation of Thr-186 of CDK9 results in an increase in the CDK kinase activity towards its substrates, for example the SPT5 subunit of DSIF (Larochelle et al., 2012).

Cellular P-TEFb exists in inactive and active forms (Nguyen et al., 2001;Krueger et al., 2010). A large amount of cellular P-TEFb is sequestered in an inactive complex with the noncoding 7SK small nuclear RNA (7SK snRNA), hexamethylene bisacetamide-induced protein 1 (Hexim1) and/or Hexim2, La-related protein 7 (LARP7) and MEPCE (Nguyen et al., 2001;Barboric et al., 2009). The small catalytically active form of P-TEFb is instead generally associated with bromodomain-containing protein 4 (Brd4) and the Super Elongation Complex (SEC) (Jang et al., 2005;Yang et al., 2005). The transition between active and inactive forms is dynamic and tightly regulated. In a situation where active P-TEFb is required to activate genes, such as high stress conditions and induced differentiation, the inactive complex is disrupted and the smaller active complex of P-TEFb/Brd4 is formed (Jang et al. 2005; Yang et al. 2005; He et al. 2006). Brd4 is a chromatin bound protein that is recruited to TSS and enhancer sites by interaction with acetylated histones, Mediator and coactivators (Jang et al. 2005; Yang et al. 2005; Zippo et al. 2009). Genome-wide profiling ChIP-seq revealed that depletion of the Brd4 or AFF1/4, the two subunits that make up the two known activating P-TEFb-containing complexes results in the failure of Pol II negotiating the early-elongation checkpoint (Lu et al., 2016).

In vitro studies have shown that P-TEFb is capable of interacting with Brd4 or being incorporated into the Super Elongation Complex (SEC) (He et al., 2011;Luo et al., 2012). The SEC is a general regulator of Pol II elongation and contains ELL family members (ELL1, ELL2, ELL3), AFF1 (also known as AF4), AFF4, ENL, AF9 and P-TEFb. However, the composition of the SEC varies in different cell types and the assembly of SEC components is gene specific (Luo et al., 2012;Luo et al., 2012). The factors that make up P-TEFb-containing complexes affect P-TEFb recruitment to Pol II during the regulation of Pol II pause release. For example, P-TEFb can be recruited by Brd4 (via MED21/MED23/Tat-SF1), SEC (via AFF1 or AFF4 and MED26), or

PAF1c (via AFF1/ENL(AF9)) (Lu et al., 2016). While it is not known how these recruitment networks are coordinated, it is clear that CDK9 is part of different P-TEFb-containing complexes and is fundamentally important for Pol II pause release.

Chromatin immunoprecipitation (ChIP) experiments indicate that CDK9 is recruited to the 3′ ends of protein-coding genes (Laitem et al., 2015). Brief treatment of HeLa cells with CDK9 inhibitors KM05283 (KM) and 5,6-dichlorobenzimidazone-1-β-d-ribofuranoside (DRB) drastically reduces the level of Pol II near the poly(A) site, indicating CDK9 inhibitors cause Pol II to prematurely terminate (Laitem et al., 2015) (Appendix 1). A similar profile of reduction of Pol II levels is observed when Pol II fails to negotiate the promoter-proximal pausing near EEC following treatment with CDK9 inhibitors and in this case it is likely that Pol II has failed to make the transition to productive elongation and is held at the EEC. The drug treatment also induces the loss of CDK9 near the poly(A) site, indicating that the role CDK9 is critical for the Pol II to negotiate a transcription checkpoint close to the poly(A) site, which we have termed the poly(A)-associated checkpoint (Laitem et al., 2015). The pharmacological CDK9 inhibitors used in this study may be affecting kinases other than CDK9. For example, CDK12 is a Ser2 kinase that is recruited to the 3′ ends of protein-coding genes and is inhibited by the CDK9 inhibitor flavopiridol *in vitro* (Bartkowiak et al., 2010;Blazek et al., 2011;Bosken et al., 2014;Davidson et al., 2014).

The role of CDK12 (and CDK13 as a member of the CDK12 subfamily)

A very early *in vitro* study of CDK12 suggested that CDK12 has CTD kinase activity (Ko et al., 2001). However, due to the presence of an arginine/serine (RS)-rich domain that is predominantly found in factors regulating pre-mRNA splicing, subsequent research focused on the role of CDK12 and CDK13 in RNA processing and specifically in splicing. CDK12 colocalizes *in vivo* with components of the spliceosome, and interacts *in vitro* with splicing factors SF2 and ASF2, as well as the Pol II CTD. CDK13 interacts *in vitro* with p32, an ASF/SF2-associated protein (Ko et al., 2001). CDK13 also interacts with HIV-1 Tat and regulates viral mRNA splicing (Berro et al., 2008). CDK12 and CDK13 are unusually large for CDKs with molecular weights of ~160 and ~180 kDa respectively (most CDKs are about ~50 kDa), and contain large N- and C-terminal extensions that flank the kinase domain (Even et al., 2006; Ko et al., 2001). Both are associated with Cyclin K (Kohoutek and Blazek

2012;Greifenberg et al., 2016). CDK13 bears 92% sequence identity in its kinase domain to CDK12 (Gopinathan et al., 2011). There are two CDK9-like kinases in *S. cerevisiae*; CDK9 is the orthologue of the yeast Bur1, which phosphorylates yeast Spt5, whereas CDK12 is the orthologue of yeast Ctk1, which phosphorylates Ser2 of the yeast Pol II CTD. CDK12 was therefore postulated to be a second Ser2 CTD kinase in higher eukaryotes (Bartkowiak et al., 2010). CDK12 was then shown to phosphorylate Ser2 CTD *in vitro* (Bartkowiak et al., 2010) and depletion of CDK12/Cyclin K in human cells reduces the Ser2P level and expression of selected genes, particularly DNA damage response genes (Blazek et al., 2011).

CDK12 has a similar structural conformation as CDK9 and can also phosphorylate Ser5 of the Pol II CTD *in vitro* (Bosken et al., 2014). It is known that Brd4 regulates active CDK9. However, no known CDK12 or CDK13 regulating factors have been identified yet. Structural studies also indicate that CDK12/Cyclin K has substrate specificity for Ser5 of the CTD (Bosken et al., 2014;Bartkowiak and Greenleaf 2015) but the precise *in vivo* function of CDK12 in CTD phosphorylation is not yet clear.

The role of CDK11 as a CTD kinase

In humans CDK11 is encoded by two highly homologous genes CDC2L1 (CDK11B) and CDC2L2 (CDK11A), both localized on chromosome 1 in a region spanning about ~140 kb (Gururajan et al., 1998). Each CDK11 gene expresses several CDK11 isoforms including CDK11-p110, CDK11-p58 and CDK11-p46. Due to the cell cycle-regulated internal ribosome entry site (IRES) that results in alternative initiation of translation on the mRNA for CDK11-p110, it also produces the CDK11-p58 isoform (Cornelis et al., 2000). Both isoforms can be cleaved by caspases to produce a smaller CDK11-p46 isoform. CDK11-p110 is ubiquitously expressed throughout the cell cycle, but CDK11-p58 is predominantly expressed in G2/M (Cornelis et al., 2000). CDK11 is associated with Cyclin L1 and L2 (Loyer et al., 2008). Like CDK12 and CDK13, CDK11 has an RS domain characteristic of splicing factors and CDK11/Cyclin L has been associated with pre-mRNA splicing (Hu et al., 2003). It was recently suggested that CDK11 regulates the recruitment of Pol II into the PIC by phosphorylating subunits of the Mediator complex (Drogat et al., 2012). Early *in vitro* studies showed that CDK11/Cyclin L can phosphorylate the Pol II CTD, and recent *in vitro* experiments showed that CDK11 can phosphorylate both Ser2 and Ser5 of the Pol II CTD (Dickinson et al., 2002;Pak et al., 2015). Knockdown of CDK11 reduces

the level of Ser2P, but not Ser5P, suggesting that Ser2 is the preferred target of CDK11 in cells (Pak et al., 2015).

CDK10 and CDK20 as transcriptional kinases

The *CDK10* gene generates two isoforms as a result of alternative splicing; a longer isoform, CDK10.1 and a shorter isoform, CDK10.2; both are likely to produce a number of splice variants uncharacterised to date (Brambilla and Draetta 1994;Grana et al., 1994;Guen et al., 2017). Compared to the shorter isoform CDK10.2, CDK10.1 has additional 29 amino acids in the amino-terminus harbouring a glycine-rich or ATP-binding loop and an additional 15 amino acids harbouring a functional bipartite nuclear localization sequence (NLS) in the carboxy-terminus (Sergere et al., 2000;Guen et al., 2017). CDK10.1 interacts with Cyclin M, which has strong sequence homology to that of the L-type cyclin of CDK11, CDK12 and CDK13. However, the shorter isoform, CDK10.2 only weakly binds Cyclin M (Guen et al., 2017). Although CDK10 is a transcriptional CDK, its ability to phosphorylate the Pol II CTD is not yet known. To date, there are two cellular proteins that have been shown to be phosphorylated by CDK10/Cyclin M; ETS2 and PKN2 (Guen et al., 2013;Guen et al., 2017). Studying CDK10 phosphorylation substrates and interacting partners is challenging due to the multiple splice forms of CDK10.

CDK20 was first thought to be a CDK-activating kinase, p42, due to its ability to phosphorylate the cell-cycle kinase, CDK2 (Liu et al., 2004). Later evidence showed that p42, described as PNQALRE/CCRK, co-purifies with CDK7, and it was CDK7 that had robust kinase activity towards CDK2 (Wohlbold et al., 2006). PNQALRE/CCRK was later renamed CDK20 and is considered to be a CDK7 paralogue due to its ability to also phosphorylate CDK2 (Malumbres et al., 2009). Evolutionary analysis of CDK20 and CDK7 shows that they have a distant evolutionary relationship. CDK20 has therefore been placed in a separate CDK subfamily (Malumbres 2014). To date there is no evidence that CDK20 can phosphorylate the Pol II CTD.

1.2.5 Dysregulation of CTD kinases

The involvement of deregulated kinase activity in promoting oncogenic events has made this group of proteins a target for anti-cancer drug discovery and development (Whittaker et al., 2017). CDK inhibitors are designed to selectively target the

hydrophobic pockets such as Glu-99 and Val-100 in CDK7 and Ala-111 and Gly-112 in CDK9 (Lolli and Johnson 2005). Several small molecule inhibitors against CDK9 have successfully been approved for clinical use such as flavopiridol and many more are currently in clinical trials (Krystof et al., 2012). Currently, DRB and flavopiridol are described as CDK9-specific inhibitors and are often used to probe the cellular function of P-TEFb. Flavopiridol, CR8 and DRB are three known small molecule inhibitors that been shown structurally in complex with P-TEFb (Baumli et al., 2008;Baumli et al., 2010;Bettayeb et al., 2010). DRB is a purine nucleoside analogue and a potent ATP-competitive inhibitor of CDK9. As a purine nucleoside analogue, DRB is likely to interfere with any enzyme that requires purine nucleotides and thus more than one cellular kinase may be DRB-sensitive (Mancebo et al., 1997). Treatment of cells with DRB results in the production of shortened transcripts, diagnostic of the inability of Pol II to transcribe through the early elongation checkpoint in the absence of P-TEFb activity (Marshall and Price 1992). Other than CDK9, several studies suggest that DRB inhibits CDK7 and CDK8. Through inhibition of the CTD kinases, DRB inhibits an elongation step during Pol II transcription which can trigger p53-dependent apoptosis of human colon adenocarcinoma cells without inducing genotoxic stress to healthy cells (Zandomeni et al., 1986;Yamaguchi et al., 1998;te Poele et al., 1999). Structure-based analyses provide greater understanding of the mechanism of inhibition of the kinase inhibitors. They also help us to understand drug-resistant mutations and enable development of second generation and more potent inhibitors.

The roles of the transcriptional kinases have been studied not only by using pharmacological inhibitors, but also in the context of viral infections. Viral proteins have been reported to modify the role of cellular kinases, including CDK9. The critical role of P-TEFb in transcription elongation makes this kinase a good target for viruses to subvert to promote viral replication. At least nine viruses are known to hijack the role of P-TEFb (Zaborowska et al., 2015) (Appendix 2). The hijacking of P-TEFb by the HIV-1 Tat protein dissociates P-TEFb from the inactive complex and recruits active P-TEFb to the HIV-LTR to promote its transcription and this process has been extensively studied. Structural studies have shown how HIV-1 Tat complexes with P-TEFb (Tahirov et al., 2010;Gu et al., 2014). P-TEFb is therefore a possible drug target against HIV-1 and potentially other viruses. Understanding the requirements of CDK9 for viral infection and how viral infection alters Pol II CTD

phosphorylation patterns not only advances our knowledge of viral pathogenesis but also provides potential new antiviral drug targets. For this reason, my thesis focuses on the role of viral proteins, in particular, the HSV-1 ICP22 protein, in the regulation of expression of cellular genes at the level of transcription.

1.3 Herpes Simplex Virus 1

1.3.1 The human herpesviruses

As it stands today eight herpesviruses have been identified as human pathogens which routinely infect large populations. These are further classified into three subfamilies; alpha-, beta-, and gamma herpesviruses based on their biological properties such as growth characteristics and tissue tropism, and comprehensive genomic information (Davison 2007;Liu and Zhou 2007) (Table 1.2). Structural analyses in the last several years were mainly focused on the hallmark architecture of Herpes Simplex Virus (HSV), Cytomegalovirus (CMV) and Epstein-Barr Virus (EBV). These viruses are regarded as the experimental prototype representing members of each herpesvirus subfamily. The thesis is particularly focused on HSV.

Table 1.2. Subfamilies of human herpesvirus (HHV)

Subfamily	Type	Common name	Unique properties
Alpha	HHV-1	Herpes Simplex Virus 1 (HSV-1)	Short replicative cycle, rapid destruction of the host cells, wide host range, establish latency in sensory nerve ganglia, replicate efficiently in cell culture.
	HHV-2	Herpes Simplex Virus 2 (HSV-2)	
	HHV-3	Varicella-Zoster Virus (VZV)	
Beta	HHV-5	Cytomegalovirus (CMV)	Long replicative cycle, restricted host range, establish latency in monocytes and spleen cells, replication in tissue culture is slower than alpha-herpesviruses.
	HHV-6	Human herpesvirus 6	
	HHV-7	Human herpesvirus 7	
Gamma	HHV-4	Epstein-Barr Virus (EBV)	Lymphotropic virus (EBV), very limited host range, establish latency in lymphoid tissue.
	HHV-8	Kaposi's sarcoma associated herpesvirus (KSHV)	

Pathogenesis of the Herpes Simplex Virus (HSV)

Both HSV-1 and HSV-2 are equally capable of causing initial infection at any anatomical site, although HSV-1 causes more recurrent oral infections and HSV-2 more recurrent genital infections (Whitley et al., 2007). The infection is mild and self-limiting in healthy individuals but may be severe and life threatening in

immunocompromised individuals and neonates (Herget et al., 2005;Cherpes et al., 2012). HSV-1 and HSV-2 have a similar genetic composition. However, amino acid differences in the glycoprotein G genes allow serologic tests based on the glycoprotein G to distinguish between patients with either HSV-1 or HSV-2 (Wald and Ashley-Morrow 2002). HSV infection can give rise to encephalitis, corneal blindness, peripheral nervous system-related diseases, meningitis in adults and meningoencephalitis in neonates (Whitley et al., 2007;Steiner and Benninger 2013), Infection is also associated with increased spread of HIV (Tsigrelis et al., 2006). HSV infection can be a primary infection that occurs in individuals who do not possess neutralising antibodies in their blood or recurrent infections in individuals previously exposed to the virus and who already possess antibodies against it (Whitley et al., 2007). Several factors determine the fate of the viral life cycle as either a productive infection or a state of latency – these will be described in later sections.

Features of HSV-1

All herpesviruses that have been identified to date infecting both vertebrates and non-vertebrates, share some structural features (Figure 1.3A) (Bowman et al., 2003;Liu and Zhou 2007). These viruses have a highly ordered icosahedral-shape nucleocapsid of about ~125-130 nm in diameter, which encases the linear double-stranded viral DNA (Bowman et al., 2003). The nucleocapsid is surrounded by the tegument; a partially ordered proteinaceous layer consisting of polyamines, lipids and glycoproteins (Bowman et al., 2003). The tegument is enclosed within a polymorphic lipid bilayer containing viral glycoproteins that are responsible for viral attachment and entry into host cells (Liu and Zhou 2007).

The HSV-1 genome structure is complex with two unique regions U_L (unique long region) and U_S (unique short region), that are flanked by repeat sequences (a, b, c) and inverted repeat sequences (a', b', c') (Figure 1.3B) (McGeoch et al., 1988;Frampton et al., 2005). The a sequence can be found at both termini and an inverted copy of a' is located at the U_L/U_S junction; the b sequence and its inverted copy b' flank the U_L; and the c sequence and its inverted copy c' flank the U_S (McGeoch et al., 1988). Replication of the HSV-1 genome begins at three origins of replication found within the genome; one within U_L (oriL) and the other two copies, duplicated oriSs, present within the U_S (Zhong and Hayward 1997). The length of the double-stranded DNA genome of HSV-1 is approximately ~152-162 kb encoding ~84 viral genes (Whitley

1996;Kristie 2007). Half of the genes are essential for viral replication in cell culture (Figure 1.3B, bottom list) while others are non-essential for viral replication (Figure 1.3B, top list) (Frampton et al., 2005). It is noted that the ICP22 open reading frame, which is the focus of this thesis, maps to the first open reading frame of the U_S and was categorised as non-essential for viral replication (red arrow, Figure 1.3B) (Post and Roizman 1981). However, later studies suggest that ICP22 is required for efficient viral growth in some cell types. ICP22 will be discussed in a later section 1.3.3. The viral genes have TATA-box containing promoters and downstream poly(A) signal and therefore are very similar to host cell Pol II-transcribed genes (Oakes et al., 1976;Roizman and Zhou 2015).

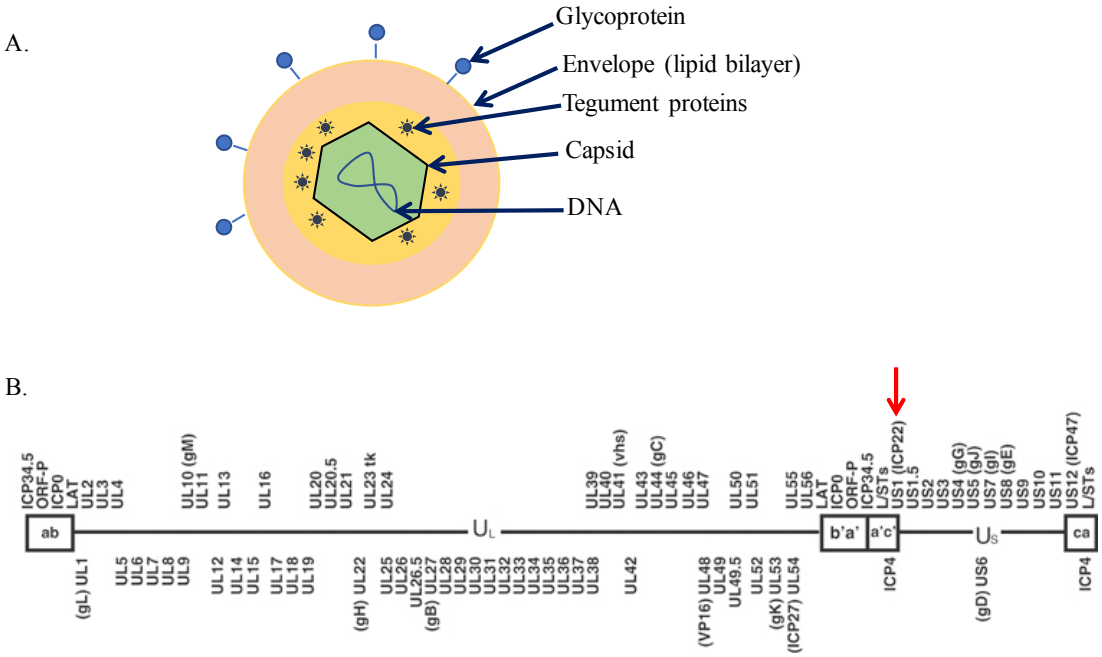


Figure 1.3. A) Virion architecture of herpesviruses. The schematic diagram shows the multilayer organization of the human herpesviruses, adapted from (Liu and Zhou 2007). **B) Layout of HSV-1.** The direction of open reading frames is from left to right. The location of viral genes is indicated. The gene encoding ICP22, the focus of this study, is marked with a red arrow. The layout of HSV-1 was adapted from (Frampton et al., 2005).

1.3.2 Molecular events during HSV-1 infection

The life cycle of HSV

The life cycle of HSV-1 is illustrated in Figure 1.4. HSV-1 infection is initiated by cell surface attachment with a subset of cellular receptors (Davison and Bhella 2007). Viral glycoproteins (gD, gB, gH and gL) are four of approximately 12 viral envelope proteins that induce cell entry (Davison and Bhella 2007). The nucleocapsid traverses the cytoplasm to the nuclear pore, and the viral DNA is released into the nucleus. The viral genes are then sequentially and co-ordinately expressed in a cascade where three classes of genes are regulated; the immediate-early, early, and late genes (Honess and Roizman 1974;Kristie 2007). These three group of genes were originally categorised as α, β, and γ groups respectively, or sometimes referred as IE, E, and L genes respectively (Honess and Roizman 1974). In general, the immediate-early genes encode viral proteins required for the regulation of early and late gene expression, while the early genes encode viral proteins required for viral DNA replication and late genes for virion structural components (Kristie 2007). The capsid which is rich in polyamines, is believed to provide positive charges to neutralize the viral DNA genome during genome packaging and replication (Gibson and Roizman 1971;Bowman et al., 2003). At the end of lytic infection, tegument proteins and structural components required to form infectious virions are assembled followed by viral egress releasing a mature virion from the cell (Britt 2007).

HSV-1 latency

A unique hallmark of herpesvirus infection is the ability of the virus to move back and forth between lytic and latent cycles (Figure 1.4) (Roizman and Knipe 2001;Knipe et al., 2013;Roizman and Zhou 2015). As a human neurotropic virus, HSV establishes lifelong latent infection in peripheral neurons (Knipe and Cliffe 2008;Whitley 2015). In latent infection, the virus reaches the sensory nerves innervating the primary site of infection where the capsid is docked at the nuclear pore and viral genome is released into the nucleus (Stevens 1989;Steiner and Benninger 2013). Throughout the course of latency, the viral genome is organized into repressive heterochromatin and is extensively silenced as shown by the presence of repressive methylation marks on the lysine residues of the histones associated with their promoters (Wang et al., 2005;Cliffe et al., 2009;Roizman and Zhou 2015). Limited viral gene expression is observed during latency; to date only latency-associated transcripts (LATs) and a

subset of viral miRNAs have been detected in latency (Stevens et al., 1987;Zwaagstra et al., 1990;Umbach et al., 2008;Jurak et al., 2010). LAT expression is crucial for the latency process, including maintaining the viral genome chromatin in a repressive state, helping survival of the infected neuron via the anti-apoptotic properties of LAT, and promoting efficient establishment of reactivation (Wang et al., 2005;Steiner and Benninger 2013). In order to produce progeny virus for dissemination, the latent virus enters the lytic phase of replication periodically and this process is called reactivation. The molecular mechanisms of how reactivation occurs are unclear (Roizman and Zhou 2015). The virus can be reactivated either spontaneously, by systemic infection or due to neuronal stress. A recent study suggests that a neuron-specific stress pathway involving JNK mediates viral reactivation by overriding the repressive marks on the chromatin (Cliffe et al., 2015). The infected host cells develop innate anti-HSV responses that can limit viral replication and propagation. Programmed cell death is one of the antiviral countermeasures. The common forms of programmed cell death are apoptosis which is executed by caspases, and necroptosis which can be activated in the absence of caspase activity (Aubert et al., 1999;Mocarski et al., 2011). HSV has evolved sophisticated strategies to overcome programmed cell death. For example, a single viral protein ICP6 can bind to caspase 8 and inhibit apoptosis, and can bind to subunits of the necrosome to inhibit necroptosis (Mocarski et al., 2011;Huang et al., 2015).

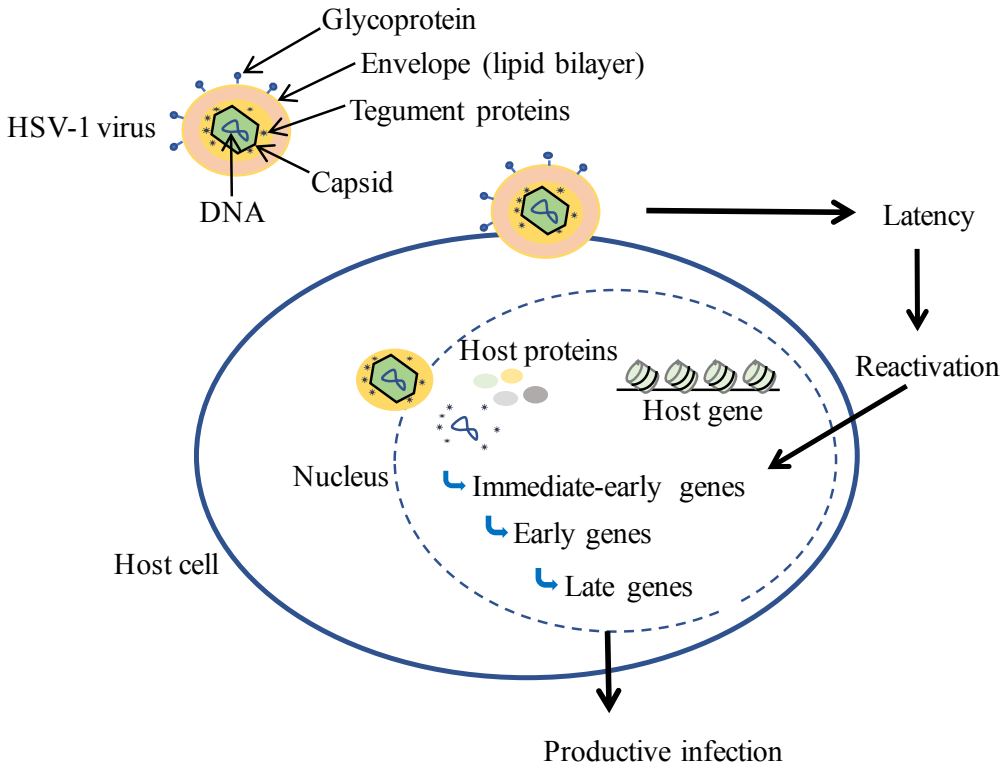


Figure 1.4. The life cycle of HSV. Once the virus reaches a susceptible host cell, the virion particle encounters appropriate cell surface receptor on the host cell membranes. After fusion, the viral nucleocapsid and tegument proteins are uncoated, internalised into the cytoplasm and migrate towards the nuclear membrane via microtubules and reach the host's nucleus where the linear double stranded viral DNA is injected through the nuclear pores and circularizes to form the viral episome. The viral genomes are programmed for transcription at different time periods post-infection, starting with immediate-early, then the early, and the late genes (Honess and Roizman 1974;Honess and Roizman 1975). HSV is able to establish latency in sensory neurons innervating the site of primary infection and can be periodically reactivated (Stevens 1989;Steiner and Benninger 2013).

Regulation of HSV-1 viral genes

Upon entry into the cell, viral DNA is repressed as a result of early host defences and such repression can be relieved by the activities of appropriate transcriptional factors such as tegument protein VP16, which enhances the expression of immediate-early genes (Whitley et al., 2007;Kelly et al., 2009). The HSV-1 viral genes are transcribed by the cellular Pol II and are largely dependent on the host transcription mechanisms for their expression. The transcription of more than 80 HSV genes is tightly co-ordinated in a sequential manner (Figure 1.5). The immediate-early genes which are essential for productive infection, are transcribed immediately within 2-5 h post-infection (Honess and Roizman 1974). These genes are expressed early in infection

and therefore do not rely on other viral protein for transcriptional activation except the viral tegument protein VP16. To activate the transcription of immediate-early genes, VP16 interacts with host cell factor-1 (HCF-1) and octamer-binding transcription factor-1 (Oct-1), forming a *trans*-activator complex that binds to the *cis*-regulatory consensus sequence TAATGARAT in the promoter of immediate-early genes (La Boissiere et al., 1999;Babb et al., 2001). There are five immediate-early gene products which primarily function as regulatory proteins and have immune evasion functions; four of them (ICP0, ICP4, ICP22 and ICP27) localise in the nucleus and are responsible for regulating the expression of the early and late genes, while the fifth (ICP47) localises in the cytoplasm. The early genes are expressed from 5-7 h post-infection, while the late genes are expressed from 7-8 h until at least 12 h post-infection (Honess and Roizman 1974). The early and late genes are dependent on the presence of immediate-early proteins in order to efficiently accomplish their gene expression programme (Honess and Roizman 1974). This thesis focuses on ICP22, the product of an immediate-early gene. The properties of ICP22 are described in the next section 1.3.3 and its proposed role in the alteration of host Pol II CTD modification is described in section 1.4.

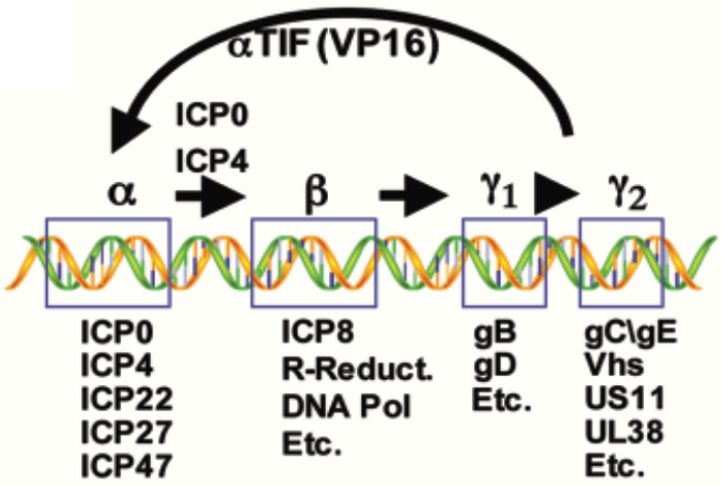


Figure 1.5. Sequential expression of HSV genomes. HSV genomes are sequentially expressed during productive infection. The VP16 (αTIF) viral tegument protein activates the immediate-early (α) genes encoding ICP0, ICP4, ICP22, ICP27 and ICP47 immediately after infection. Subsequently, ICP0 and ICP4 activate early (β) genes while ICP22 and ICP27 maintain the optimal expression of the late (γ) genes. The late genes can be divided into two groups; the γ1 and γ2. The early genes encode viral proteins required for viral DNA replication while late genes encode virion structural components. Figure adapted from (Roizman et al., 2005).

1.3.3 Immediate early infected cell protein 22 (ICP22)

ICP22 is encoded by the α22 or US1 gene of ~4.8 kb in size (Figure 1.6). The US1 gene also encodes an in-frame N-terminally truncated version of ICP22, termed US1.5 due to the presence of an independent promoter (Carter and Roizman 1996;Ogle and Roizman 1999). US1.5 is conserved among alpha-herpesviruses but is not found in beta- or gamma-herpesviruses (Roizmann and Pellet, 2001). US1.5 forms a dispersed pattern in the nucleus unlike ICP22, which is concentrated in small dense nuclear bodies. This suggests that there are differences in their biological functions despite having the same amino acid sequences at the C-terminus (Cun et al., 2006). Due to the co-expression of the two proteins, the functional differences between ICP22 and US1.5 are not well defined (Li et al., 2013). ICP22 comprises 420 amino acids with a theoretical molecular weight of ~68 kDa, while US1.5 begins at Met-147 and comprises 274 amino acids with theoretical molecular weight of ~37 kDa. Reports on the translation start sites for US1.5 differ. US1.5 has been reported to begin at Met-90, Met-147 or Met-171, possibly due to the use of different strains of virus (Figure 1.6A) (Bowman and Schaffer 2009). ICP22 has two known independent nuclear localization signals (NLS) which have been confirmed in cellular studies (Stelz et al., 2002). NLS1 resembles a classical NLS of bipartite nucleoplasmin while NLS2 resembles the NLS of Polyomavirus large T antigen (Stelz et al., 2002).

Based on the amino acid identity and sequence conservation of ICP22 and its homologues in other herpesviruses, Schwyzer *et al.* proposed that ICP22 has four distinctive domains (Figure 1.6B) (Schwyzer et al., 1994). The first 160 residues of the N-terminal region of ICP22 is in the first domain (Schwyzer et al., 1994), and this and the C-terminal residues 398-420 of ICP22 are unique to HSV-1 and possibly HSV-2 (Ogle and Roizman 1999). The second domain of ICP22 from residues 161 to 292 is described as the core conserved region. The residues in this region have high similarity between the human (VZV, HSV-2) and animal herpesviruses (Bovine herpesvirus 1 (BHV-1), Equine herpesvirus 1 (EHV-1), Equine herpesvirus 4 (EHV-4), and Pseudorabies (PRV)) (Schwyzer et al., 1994;Zaborowska et al., 2014). The third region comprises residues 293-386 and is deemed to be an acidic domain with Asp and Glu and hydroxyl (Ser and Thr) residues. The last region in the C-terminus is deemed to be a basic domain with Arg and Lys residues and proline residues (Schwyzer et al., 1994). In general, the N-terminal residues and the C-terminal basic and acidic domains of ICP22 and its homologues vary in length, and these domains

therefore give rise to unique ICP22 proteins within the herpesviruses.

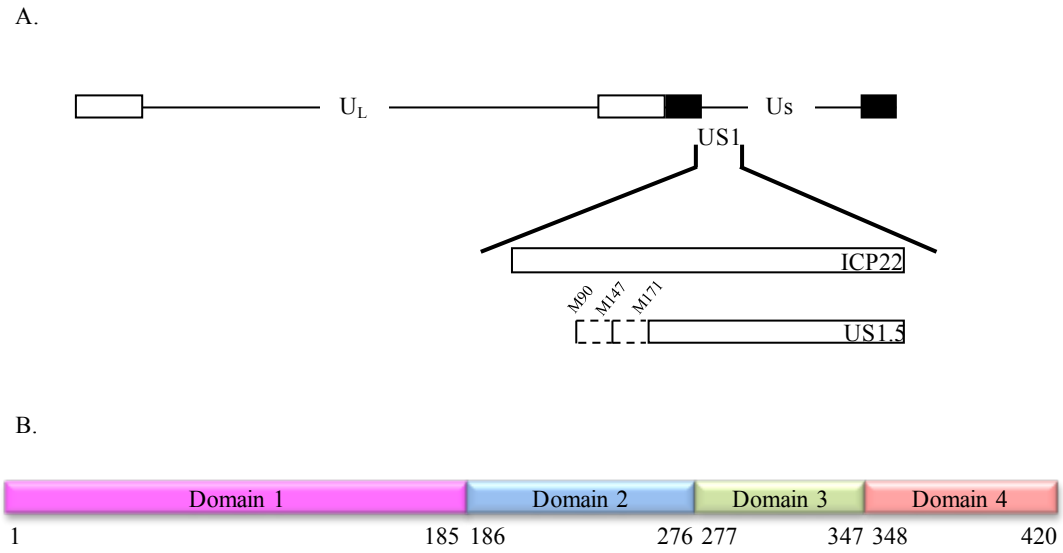


Figure 1.6. A) The genomic layout of HSV-1 and ICP22. The US1 gene, which encodes ICP22 and US1.5 maps to the first open reading frame (ORF) of the U_S region (Carter and Roizman 1996;Bowman and Schaffer 2009). **B) ICP22 domains.** ICP22 is divided into four distinct domains based on amino acid identity and sequence conservation (Schwyzer et al., 1994).

Immediately after synthesis, ICP22 adopts an intranuclear localisation, forming small dense nuclear bodies (Roizmann 1997). Several hours post-infection, ICP22 co-localises with UL4, a late viral protein (Jahedi et al., 1999). During HSV-1 infection, ICP22 is extensively post-translationally modified; primarily by phosphorylation on serine/threonine residues by two HSV-1 protein kinases; UL13 and US3 (Poon et al., 2000;Hagglund et al., 2002;Poon and Roizman 2005). The UL13 protein kinase is readily available from the virion upon cellular invasion (Ogle et al., 1997). ICP22 is also phosphorylated on tyrosine residues but the kinase(s) involved has yet to be identified (O'Toole et al., 2003).

As depicted in Figure 1.3B, ICP22 is listed as a non-essential protein for viral replication. In the 1980s, ICP22 was identified as non-obligatory for cell growth and dispensable for viral replication in cell culture (Post and Roizman 1981;Sears et al., 1985). These studies used Vero cells, a type of cell mostly used to study viral infection *in vitro* (Post and Roizman 1981;Sears et al., 1985). However, later studies showed that ICP22 is required for efficient HSV-1 cell culture growth in other cell lines, for example, human embryonic lung (HEL), mouse embryonic fibroblast

(MEFs), primary human fibroblasts and rabbit skin (RAB-9) cells (Post and Roizman 1981;Sears et al., 1985;Poffenberger et al., 1994;Orlando et al., 2006;Fraser and Rice 2007). The cell lines which require ICP22 for efficient viral growth are termed restrictive cell lines, while cell lines which allow growth of viruses lacking ICP22, for example Vero, HeLa and Hep-2 cells, are termed permissive cell lines (Sears et al., 1985;Poffenberger et al., 1993;Poffenberger et al., 1994). In the restrictive cell lines, ICP22 is necessary for the efficient expression of late genes, particularly the group γ1 late genes (Sears et al., 1985;Poffenberger et al., 1993;Rice et al., 1995). The group γ1 genes encode the proteins required for viral DNA synthesis. The restrictive cell lines infected with a virus without ICP22 therefore have severely reduced viral replication (~100-fold less than in wild-type) (Sears et al., 1985;Poffenberger et al., 1993;Rice et al., 1995). There is also evidence that ICP22 is required for expression of the group γ2 late genes US11, UL38, and UL41 (Poon et al., 2000). In a mouse model, infection is severely attenuated when HSV-1 is null for ICP22, indicating that ICP22 is required to express the viral genes essential for HSV-1 pathogenicity (Sears et al., 1985;Poffenberger et al., 1994;Orlando et al., 2006). ICP22 is also required for efficient expression of the late gene US5, which encodes glycoprotein J (gJ) (Aubert et al., 2008). HSV-1 gJ has an anti-apoptotic function, and the ICP22 null virus was shown to have a reduced anti-apoptotic effect, suggesting that ICP22 may have a role in preventing apoptosis during infection (Aubert et al., 1999).

Whilst these findings demonstrate that ICP22 is required to stimulate the expression of early and late viral genes, a number of findings also suggest that ICP22 is a transcriptional repressor of a subset of viral and cellular genes (Prod'hon et al., 1996;Cun et al., 2009;Guo et al., 2012). By using a transient expression system, ICP22 was shown to repress the transcription of all immediate-early, early and late genes by interacting with P-TEFb and to block P-TEFb recruitment to the viral promoter (Guo et al., 2012). The tegument protein VP16 is able to interact with P-TEFb (Kurosu and Peterlin 2004;Guo et al., 2012) and could increase P-TEFb recruitment to the viral promoter through recognition of the TAATGARAT motif in association with HCF-1 and Oct1 (Guo et al., 2012). VP16 is able to relieve the repressive action of ICP22 by overcoming the inhibition of P-TEFb by ICP22 and thereby promote transcription of viral genes (Guo et al., 2012).

In permissive cells lines, the lack of ICP22 slightly affects viral replication (2 to 10-fold less than wild-type), but the expression and activities of the cellular cell cycle

regulatory proteins are repressed (Advani et al., 2000;Advani et al., 2003;Orlando et al., 2006). A series of reports on the effect of viral infection on cell cycle regulation suggest that ICP22 can modify the expression of cellular genes. A later study suggests that ICP22 is also capable of repressing the antiviral effect of the type I interferons. The presence of HSV-1 genomes in a cell triggers the cellular innate immune responses. Once the sensors are activated, the cells produce type I interferons (IFNs) IFN-α and IFN-β, as a mean of cellular defence (Haller et al., 2006;Kawai and Akira 2006). HSV-1 has evolved strategies to alter expression of the IFNs or interfere with the IFN signalling pathway (Haller et al., 2006;Kawai and Akira 2006). Early experiments showed that ICP22 is not required to counteract the type I IFN response (Mossman et al., 2000). However, these experiments were executed in IFN-treated Vero cells, before it became clear that these cells are permissive for the ICP22 null mutant. When tested in HEL cells restrictive for the ICP22 mutation, the absence of ICP22 results in a reduction of viral yields as shown by a standard plaque assay (Mostafa and Davido 2013). When HEL cells are pre-treated with human IFN-β and then transfected with the ICP22 null mutant, the viral yield is further reduced, as shown by plaque assays (Mostafa and Davido 2013). These results suggest that replication of the ICP22 mutant virus is sensitive to IFN-β in restrictive cell lines, and therefore ICP22 could help to counteract the antiviral effect of type I interferons.

A recent study into the regulation of the interferon-inducible protein IFI16, which acts as a sensor of viral DNA in the host nucleus, showed that ICP22 is enriched 6-fold in the IFI16 proteome (Diner et al., 2015). Although the relationship between ICP22 and IFI16 is not yet clear, the enrichment of ICP22 in the IFI16 proteome suggests that ICP22 may play a role in IFI16-mediated immune signalling (Diner et al., 2015). An *in vivo* experiment has shown that ICP22-null HSV-1 is incapable of invading neurons in mouse and guinea pig, thus indicating that ICP22 plays roles in the establishment of latent infection (Poffenberger et al., 1994). Several studies have indicated that ICP22 physically interacts with cellular proteins, or co-localises with several cellular proteins that either help promote viral gene expression or antagonize host gene expression (Table 1.3).

Table 1.3. Known ICP22 cellular binding proteins

ICP22 known interacting partner			
Cellular proteins	**Cellular properties**	**Proposed effect of ICP22 binding**	**References**
CDK9	Subunit of P-TEFb	Phosphorylates Pol II CTD and promotes transcription on viral genes. Inhibits transcription elongation on host genes.	(Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014)
Pol II	Polymerase	Downregulate expression of host cell genes and facilitate viral gene expression.	(Zaborowska et al., 2014)
p53	Transcription factor	p53 promotes HSV-1 replication.	(Maruzuru et al., 2013)

Cellular proteins that are co-localised or enriched with ICP22			
Cellular protein	**Cellular properties**	**Study**	**References**
IFI16 (Interferon, gamma-inducible protein 16)	Interferon-inducible proteins	IFI16 has antiviral properties. ICP22 is enriched in IF116 co-immunoprecipitation-proteomics in HSV-1 infected in primary human fibroblast. Association of ICP22 and IFI16 is not yet characterised.	(Diner et al., 2015)
MCRS1	Microspherule protein 1, associated with TAF		(Bruni et al., 1999)
Cdc25C	Phosphatase		(Smith-Donald and Roizman 2008)
p60 and p78	Cell-cycle regulated protein		(Bruni and Roizman 1998;Bruni et al., 1999)
EAP	Nucleolar protein		(Leopardi et al., 1997)
Subunits of casein kinase II	Kinase	ICP22 is nucleotidylylated by casein kinase II.	(Mitchell et al., 1994)
Hsc70	Chaperone protein		(Bastian et al., 2010)

GLTSCR2	Glioma tumor suppressor candidate region gene 2		(Bruni et al., 1999)
CLOCK	Histone acetyltransferase		(Kalamvoki and Roizman 2011)

1.4 HSV-1 infection and changes in host transcriptome

HSV strategies to inhibit Pol II-mediated transcription

HSV-1 takes over and maximises the host cell resources to support its life cycle. HSV-1 infection alters the regulation of host gene transcription, as shown by accumulating high-throughput evidence (Abrisch et al., 2015;Rutkowski et al., 2015;Abrisch et al., 2016;Dembowski et al., 2017). Pol II alteration during HSV-1 infection occurs in two ways; first, Pol II is found present on the entire HSV-1 genome, suggesting that Pol II is hijacked and redirected for robust viral transcriptional activity and second, Pol II occupancy on host genomes is reduced, suggesting repression of host gene transcription (Abrisch et al., 2015;Rutkowski et al., 2015;Abrisch et al., 2016;Dembowski et al., 2017).

The first strategy is to maximise transcription of the HSV-1 genome by Pol II. The essential role of Pol II in the activation of viral genes is supported by the fact that Pol II is recruited into the viral replication compartments and the viral genes mimic host cell Pol II-transcribed genes by having TATA boxes in their promoter regions and downstream polyadenylation signal (Oakes et al., 1976;Rice et al., 1994;Dembowski et al., 2017). The early stage of productive infection is regulated at the level of transcription. Pol II transcribes the immediate-early genes and the gene products then activate the cascade of transcription events of early and late genes. A recent study on HSV-1 genome replication-coupled late gene transcription showed that Pol II and Pol II-associated factors such as Mediator, TFIID, Integrator, SPT5 and SPT6 co-purify with the viral replication fork (Dembowski et al., 2017). This observation suggests that HSV-1 hijacks Pol II and cellular factors involved in the transcription initiation and elongation of host genes to activate its own viral gene expression. As the infection progresses, HSV-1 gene expression surpasses the expression of the cellular genes, as a result of repression of host gene transcription (Spencer et al., 1997).

The second strategy of HSV-1 infection results in inhibition of actively engaged Pol II on cellular genes leading to the repression of host cell gene transcription (Abrisch et al., 2015;Abrisch et al., 2016). Such events can be observed from the modification of Pol II CTD phosphorylation by HSV-1 infection (Fraser and Rice 2005). HSV-1-infected Vero cells (later shown in HeLa, HEL cells and infection with HSV-2) produce a distinct form of Pol II, referred as Pol IIi, in contrast to the hyperphosphorylated form of Pol IIo (Rice et al., 1995;Spencer et al., 1997;Long et al., 1999). By using an antibody specific to large subunit of Pol II, it was observed that the modified species Pol IIi migrates faster than Pol IIo, indicating that HSV-1 infection induces a partial loss of phosphorylation of Pol II. This observation was clarified when the Pol IIi was shown to have less Ser2P than in Pol IIo, by using phosphospecific anti-CTD antibody H5, which is specific for Ser2P CTD (Rice et al., 1995;Fraser and Rice 2005;Rice 2011). Interestingly, Pol IIi is detected by the H14 antibody, which is specific for Ser5P CTD (Fraser and Rice 2005;Rice and Fraser 2006). Thus, HSV-1 induces a modified species of Pol II with predominant Ser5P Pol II CTD but a lack of Ser2P Pol II CTD. However, the Ser2P Pol II CTD is required for HSV-1 gene transcription (Ou and Sandri-Goldin 2013). Since the Ser2P Pol II CTD is associated with productive transcription elongation, it is possible that the alteration of this CTD mark following HSV-1 infection reduces the efficiency of transcription elongation on cellular genes, resulting in downregulation of host cell gene expression.

The effect of ICP22 on Pol II phosphorylation

Molecular virologists have attempted to address the question of what viral factor(s) are responsible for downregulating cellular gene transcription since the 1980s (Kemp and Latchman 1988;Spencer et al., 1997). Early evidence suggested that ICP22 is responsible for the modification of the phosphorylation state of the host cell Pol II CTD (Rice et al., 1995). ICP22 mediates two distinct effects on Pol II phosphorylation; first, ICP22 induces Pol IIi, and this effect requires its viral protein kinase UL13 (Long et al., 1999); second, ICP22 triggers loss of Ser2P Pol II CTD, as shown both in infected and transfected cells and this effect is independent of viral protein kinase UL13 (Durand et al., 2005;Fraser and Rice 2007;Zaborowska et al., 2014). Since Ser2P Pol II CTD is required for HSV-1 viral transcription, it was suggested that the observed loss of Ser2P Pol II CTD possibly occurs specifically on

host cell genes (Ou and Sandri-Goldin 2013;Zaborowska et al., 2014). It is possible that ICP22 induces loss of Ser2P Pol II CTD by directly inhibiting the Ser2 CTD kinase, P-TEFb. It has been shown that ICP22 interacts with the CDK9 subunit of P-TEFb, and the large subunit of Pol II (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014). The fact that ICP22 is recruited to cellular genes further strengthens the earlier findings that ICP22 is able to downregulate host cell gene expression (Zaborowska et al., 2014). ICP22 is therefore implicated in the inhibition of host Pol II-dependent gene transcription. ICP22 expression has been shown to cause a reduction in the steady-state level of mRNA from host cell genes and a reduction in the transcriptionally engaged Pol II and CTD phosphorylation on host cell genes (Kemp and Latchman 1988;Zaborowska et al., 2014). Because the level of Pol II close to the TSS was not adversely affected, ICP22 is implicated in the inhibition of transcription elongation rather than initiation (Zaborowska et al., 2014). Since ICP22 triggers the loss of Ser2P Pol II CTD, a mark associated with productive elongation, interacts with P-TEFb, and is recruited to host genes, it is likely that inhibition of P-TEFb kinase activity by ICP22 causes a defect in the transcriptional elongation on cellular genes. A short conserved region of ICP22 (residues 193-256) is sufficient for the interaction with CDK9 and causes loss of Ser2P Pol II CTD (Zaborowska et al., 2014).

1.5 Aims of the thesis

P-TEFb is an important regulator of transcription and it is functionally integrated into the expression of many human viruses, including HSV-1 (Zaborowska et al., 2015). Either by modifying its activity, regulating the amount of active and inactive complexes of P-TEFb and/or physical interaction with components of P-TEFb, viruses have evolved strategies to hijack this key factor via their own regulatory proteins. P-TEFb therefore functions as a central player in virus-host interactions to facilitate the viral replication cycle. Since its discovery as a CTD kinase, much effort has been made towards understanding the role of P-TEFb in Pol II elongation and CTD phosphorylation. Yet, there are still many unresolved questions about its cellular function and its role in pathogenic processes. Importantly, studies of viruses that use P-TEFb as a specific factor for efficient transcription will also help to elucidate the molecular mechanisms controlling transcription in uninfected cells. The ICP22 protein of HSV-1 is thought to act as transcription repressor rather than an activator of

host cell and viral genes, by directly interacting with CDK9 and causing a defect in transcription elongation (Sandri-Goldin 2007;Bowman et al., 2009;Guo et al., 2012). However, VP16, an HSV-1 tegument protein, can interact with P-TEFb in the presence of ICP22 (Cun et al., 2009;Guo et al., 2012) and may overcome CDK9 inhibition for transcription of viral genes. Clearly ICP22 is an important viral factor as its interaction with CDK9 could benefit the HSV-1 by, for example, downregulating expression of the host genome while maximizing expression of the viral genes.

The goal of the project was to investigate the role of ICP22 in the regulation of expression of cellular genes, specifically to provide a better understanding of how ICP22 inhibits P-TEFb, and perhaps transcriptional kinases other than P-TEFb to induce the loss of Pol II activity on host cell genes.

Chapter 2 aims to investigate the role of ICP22 in the transcription of host cell protein-coding genes, by closely examining the effect of transiently expressed ICP22 on transcriptionally engaged Pol II on a model gene.

Chapter 3 aims to identify the ICP22 interactome. This chapter presents the results of a proteomic study of the host cell proteins interacting with ICP22. Recombinant ICP22 and its short conserved regions (SCR), residues 193-256, were used in a pull-down assay with HeLa cell extract. Bound proteins were submitted for mass spectrometry. This chapter summarises the ICP22-bound cellular proteins that are possibly targeted by ICP22.

Chapter 4 aims to validate the ICP22 interactome. This chapter presents the results of an analysis of the interaction between ICP22 and subunits of P-TEFb, and validation of the proteomics hits from Chapter 3 by immunoblotting. This chapter gives an insight into the interaction of ICP22 with P-TEFb and inhibition of Pol II transcription elongation.

Chapter 5 (Discussion) sums up the finding of this work and points to some possible future directions.

Chapter 6 lists the materials and methods used for the experimental work.

Chapter 2| The effect of ICP22 on elongating RNA polymerase II (Pol II)

2.1 Chapter abstract

In the early 1990s HSV infection was shown to cause a dramatic change in the phosphorylation of Pol II (Rice et al., 1994). About a decade later, studies provided evidence that ICP22 is responsible for this effect, and that modification of Pol II phosphorylation can be observed when ICP22 is ectopically expressed in uninfected cells (Durand et al., 2005;Fraser and Rice 2007;Zaborowska et al., 2014). Using phosphospecific anti-CTD antibodies, it was determined that Ser2P Pol II CTD is drastically diminished (Fraser and Rice 2005). ICP22 interacts directly with P-TEFb, which is composed of CDK9 kinase and Cyclin T1, and is a Ser2 CTD kinase (Baumli et al., 2008;Buratowski 2009). Residues 193 to 256 of ICP22 are sufficient for the interaction with CDK9 and reduction of Ser2P Pol II CTD (Zaborowska et al., 2014). This work suggests that ICP22 inhibits the activity of CDK9 and therefore inhibits the transcription of host cell genes. A precise picture of how ICP22 functions to downregulate host cell gene transcription is not yet available. I have therefore further characterised the effect of ectopic expression of ICP22 on Pol II actively transcribing host cell genes.

2.2 Overview of the *KPNB1* model protein coding gene.

To explore the transcription profile of protein-coding genes, I have determined the occupancy profiles of Pol II across a protein-coding model gene *KPNB1* (the karyopherin subunit beta 1 gene). *KPNB1* is selected as a protein-coding model gene because it is well-expressed and a long gene (~36 kb long), which allows Pol II along the gene to be readily analysed during drug treatments. Pol II would 'run off' shorter gene models even with a short treatment, due to the block at the EEC. The occupancy of Pol II on this gene, before and after a short treatment of HeLa cells with known CDK9 inhibitors KM and DRB has already been determined (Laitem et al., 2015) (Appendix 1). These CDK9 inhibitors cause failure of Pol II to negotiate the early elongation checkpoint (EEC) close to the TSS and abrupt termination of Pol II transcription close to the poly(A) site, implicating CDK9 in regulating checkpoints at the beginning and end of protein-coding genes (Laitem et al., 2015). We have termed

the checkpoint close to the poly(A) site the poly(A)-associated checkpoint (Laitem et al., 2015).

The entire transcription unit of *KPNB1*, starting from the TSS to the cleavage and poly(A) site is represented in Figure 2.1. *KPNB1* is located on chromosome 17 and encodes an importin subunit beta-1 protein. It is a member of the importin beta family, which facilitates nuclear protein import either by acting as nuclear transport receptor or by forming heterodimers with an importin alpha subunit (Chook and Blobel 2001). *KPNB1* contains 22 exons. The gene encodes two transcript variants, with the shorter variant having an alternative exon replacing the first two exons of longer variant (Chook and Blobel 2001).

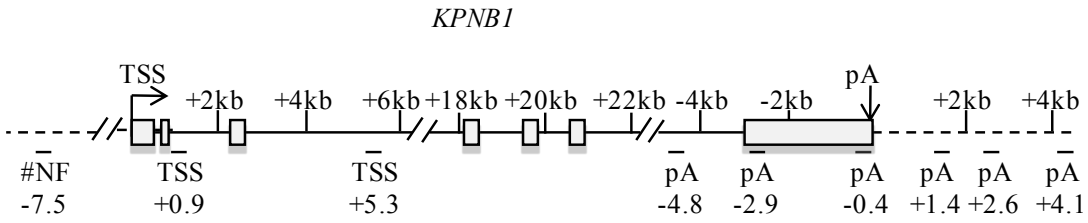


Figure 2.1. Schematic of the *KPNB1* gene. The schematic of the *KPNB1* gene indicated in kilobase pairs (kb), with the transcription start site (TSS) and poly(A) site (pA) marked by arrows, exons by boxes and introns by lines. The region of *KPNB1* amplified in quantitative PCR after chromatin immunoprecipitation (ChIP) assay in the subsequent figures is indicated underneath the gene. The ChIP-qPCR value of each region is normalized to #NF, an upstream region with no detectable Pol II to facilitate comparison of the level of factor associated with the different regions of the *KPNB1* gene.

2.3 Validation of CDK9 inhibitor treatment on HEK293 cells

The effect of CDK9 inhibitors on Pol II associated with *KPNB1* has been shown in HeLa cells (Laitem et al., 2015). Here, I present the results of a short treatment of HEK293 cells with the CDK9 inhibitor KM on Pol II across *KPNB1* (Figure 2.2).

In DMSO-treated cells, the Pol II levels are elevated at the TSS (TSS+0.9) suggesting that Pol II is paused as the results of N-TEFs. Following the switch to productive elongation, lower levels of Pol II are found throughout the body of *KPNB1*. From the poly(A) site (pA-0.4), the Pol II levels are increased. This may be due to pausing, which could facilitate recruitment of factors involved in RNA processing.

The Pol II profile changes in KM-treated cells, particularly at the beginning and end of the *KPNB1* gene. Treatment of HEK293 cells with KM for 30 min causes an increase in the Pol II level at the beginning of *KPNB1* (TSS+0.9), suggesting an

increased pausing of Pol II near the EEC, followed by a drop of Pol II (TSS+5.3) suggesting that the newly-initiated Pol II is unable to make the transition to productive elongation. Pol II that passed the EEC before the drugs took effect continues elongating along the gene body (pA-4.8, pA-2.9). The level of Pol II is significantly reduced downstream of the poly(A) site (pA-0.4, pA+1.4, pA+2.6, pA+4.1), suggesting that Pol II has prematurely terminated at the end of the gene. This suggests that inhibition of CDK9 causes failure of Pol II to negotiate the early elongation checkpoint (EEC) at the beginning of *KPNB1* gene and terminate prematurely at a checkpoint close to the poly(A) site at the end of *KPNB1* gene. Therefore the effect of KM treatment in HEK293 on the level of transcribing Pol II on a model gene is similar to that seen in HeLa cells (Laitem et al., 2015).

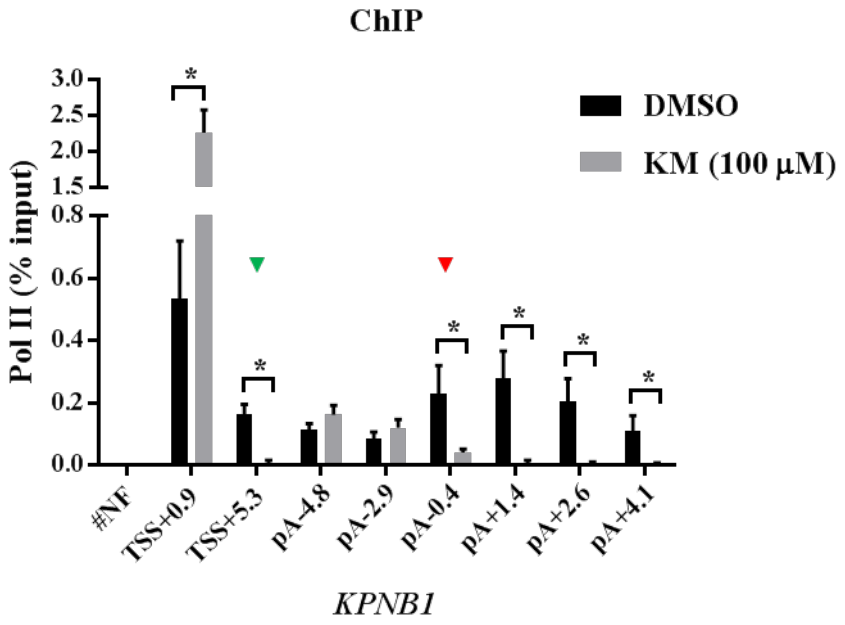


Figure 2.2. The results of qRT-PCR analysis of Pol II ChIP performed on HEK293 cells treated with DMSO as control or treated with CDK9 inhibitor KM. The green arrow indicates the early elongation checkpoint (EEC) and the red arrow indicates the poly(A)-associated checkpoint. Error bars indicate the SEM of n = 3 independent experiments. Asterisks indicate the statistical significance compared with DMSO-treated control cells (p<0.05).

2.4 Histone occupancy increases in regions where Pol II prematurely terminates

The eukaryotic genomic DNA is wrapped around histone proteins in nucleosome core called chromatin. Compared to the 'naked' DNA, the chromatin provides general barrier to the gene expression. The nucleosome-free region at the promoter region is maintained by Mediator complex to promote transcription activation following PIC assembly (Allen and Taatjes 2015;Ansari et al., 2016). Chromatin structure can pose a physical barrier that Pol II must overcome for productive elongation. Nucleosomes can be transiently removed from genes, or remodelled, or undergo modification of histone tails during Pol II elongation (Petesch and Lis 2012;Venkatesh and Workman 2015). To determine the histone occupancy of *KPNB1* following treatment of cells with the CDK9 inhibitor KM, I analysed the profile of H3 and H2B core histones by ChIP-qPCR.

Like many other protein-coding genes (Egloff et al., 2009) where the level of Pol II is highest at the promoter region, *KPNB1* has a nucleosome-free region at the TSS (Figure 2.3). The nucleosome-depleted region on *KPNB1* extends for approximately 900 bp downstream from the TSS, while a high density of nucleosomes is detected within the rest of the gene body. The nucleosome-depleted region could possibly extend beyond 900 bp from the TSS on *KPNB1*, if probed further downstream. In KM-treated cells, the occupancy of H3 and H2B in a region upstream of the TSS (#NF) where Pol II is absent, is unaffected. Similarly, the already low level of both histones in the promoter-proximal region is not affected following the increased pausing of Pol II in response to KM treatment. However, in the region where the inhibitor causes Pol II to prematurely terminate, the level of histone occupancy increases (TSS+5.3, pA-0.4, pA+1.4, pA+2.6). This result may indicate that when CDK9 is inhibited, nucleosomes are assembled and positioned on the region that is no longer transcribed. Alternatively, the increase in nucleosomes in this region could play a role in prematurely terminating Pol II.

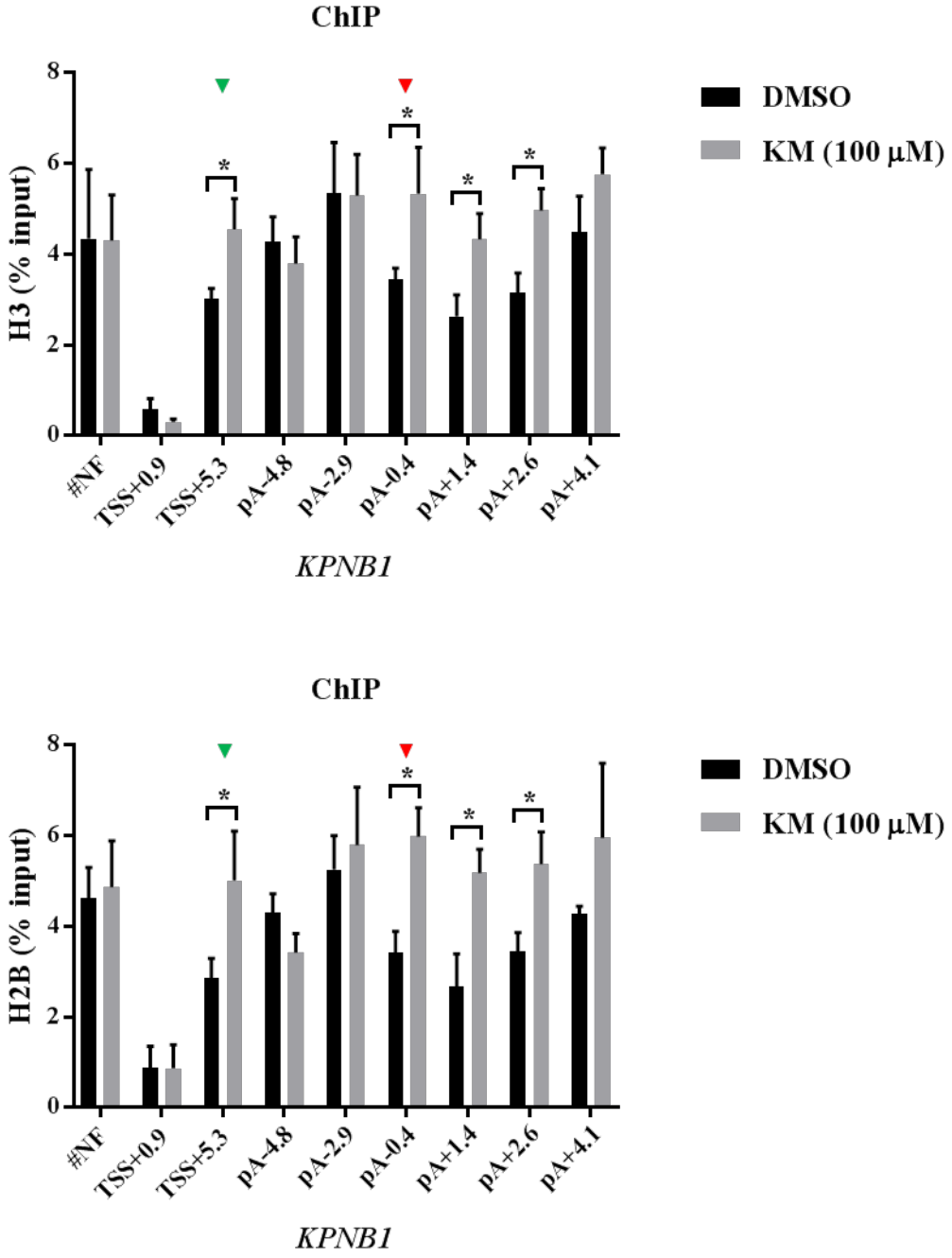


Figure 2.3. qRT-PCR analysis of H3 (top panel) and H2B (bottom panel) ChIPs performed on HEK293 cells treated with DMSO as control or treated with the CDK9 inhibitor KM. The green arrow indicates the early elongation checkpoint (EEC) and the red arrow indicates the poly(A)-associated checkpoint. Error bars indicate the SEM of n = 3 independent experiments. Asterisks indicate the statistical significance compared with DMSO-treated control cells ($p<0.05$).

2.5 Effect of transient expression of ICP22 on elongating Pol II

The inhibition of CDK9 by KM induces premature termination as evidenced by loss of the Pol II ChIP signal downstream of the poly(A) site of the *KPNB1* gene. Since this effect on actively-transcribing Pol II following CDK9 inhibition with a pharmacological inhibitor is now validated in HEK293, I have further investigated the effect of ectopic expression of Myc-tag ICP22 residues 193-256 (Myc-SCR) on the Pol II levels across *KPNB1*.

HEK293 cells were transfected with a vector encoding the short conserved region of ICP22 (residues 193-256) (Myc-SCR) with three tandem Myc epitope tags at the C-terminus and harvested at different time points, followed by immunoblotting (Figure 2.4). By using an anti-Myc antibody, the expression of Myc-SCR can be detected as early as 6 h post-transfection. The level of Ser2P Pol II CTD decreases with the increasing expression of Myc-SCR, which corresponds to the previous evidences showing that ICP22 induces loss of Ser2P Pol II CTD (Fraser and Rice 2005;Zaborowska et al., 2014).

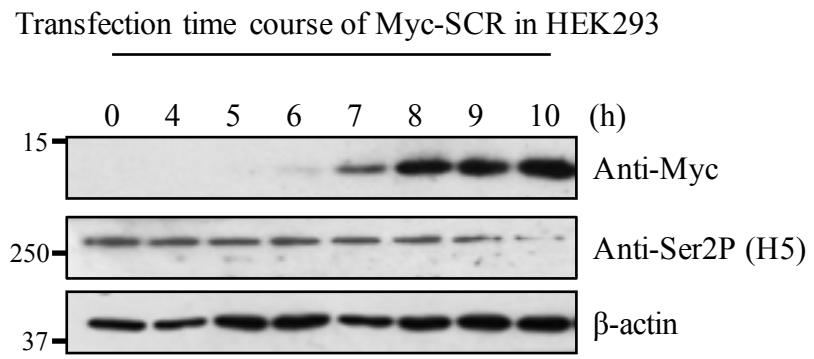


Figure 2.4. Immunoblot analysis of HEK293 cell extract after transfection of a vector expressing Myc-tag residues 193-256 of ICP22 (Myc-SCR). β-actin is used as loading control.

In untreated HEK293 and pCDNA3-transfected HEK293 cells (Figure 2.5), the levels of Pol II peak at the TSS of *KPNB1* (TSS+0.9), as a result of promoter-proximal pausing. The level of Pol II then decreases in the gene body suggesting that Pol II is elongating efficiently. A second peak of Pol II accumulation is observed downstream of the poly(A) sites, suggesting that Pol II pauses before transcription termination. Following 7 h of transient expression of Myc-SCR, the level of Pol II at the beginning of *KPNB1* is significantly elevated at the TSS of *KPNB1* (TSS+0.9) compared to that

of control and mock, suggesting increased of Pol II pausing near the EEC. In addition, the level of Pol II is significantly decreased downstream of the TSS (TSS+5.3), suggesting that the newly-initiated Pol II is unable to make the transition to productive elongation. The level of Pol II in untreated, mock and transfected cells in the middle of gene body (TSS+20.2, pA-4.8 and pA-2.9) remains the same. This suggests that Pol II that has passed the EEC before the expression of Myc-SCR is able to continue transcription elongation. The level of Pol II is significantly reduced downstream of the poly(A) (pA-0.4, pA+1.4, pA+2.6 and pA+4.1) only when Myc-SCR of ICP22 is expressed, suggesting that this region of ICP22 affects the level of transcribing Pol II at the end of *KPNB1* in a similar way to small molecule CDK9 inhibitors (Laitem et al., 2015).

It is noted that at 7 h post-transfection of Myc-SCR, the level of Ser2P Pol II CTD remains unchanged (Figure 2.4). It is possible the global level of Ser2P Pol II CTD is unaffected at this time point but affects the level of Ser2P Pol II CTD locally on a subset of genes. To assess the effect of ICP22 on the level Ser2P Pol II CTD on a single gene, it is necessary to perform ChIP experiments using anti-Ser2P Pol II CTD antibody on *KPNB1* at 7 h post-transfection of Myc-SCR. This result leads me to propose that ICP22 causes failure of Pol II to negotiate the early elongation checkpoint (EEC) at the beginning of *KPNB1* gene and premature termination at a checkpoint close to the poly(A) site at the end of *KPNB1* gene, due to inhibition of the Ser2 CTD kinase, P-TEFb.

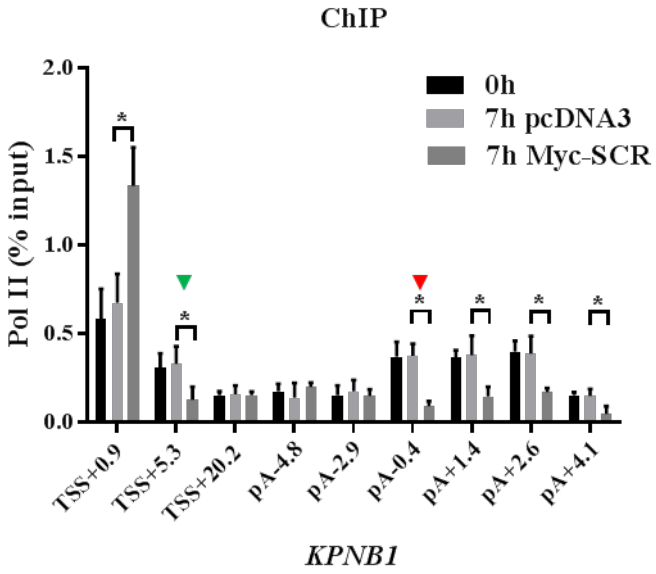


Figure 2.5. The results of qPCR analysis of Pol II ChIP after 7 h of transfection of pcDNA3 or Myc-SCR of ICP22 in HEK293. The green arrow indicates the early elongation checkpoint (EEC) and the red arrow indicates the poly(A)-associated checkpoint. Error bars indicate the SEM of n = 3 independent experiments. Asterisks indicate the statistical significance compared with 7 h pcDNA3 (p<0.05).

At 6 h post-transfection of Myc-SCR (Figure 2.6; top panel), the level of Pol II remains comparable to that in the pCDNA3-transfected samples, while at 8 h post-transfection (Figure 2.6; bottom panel) a significant rise in the Pol II signal is observed at the beginning of the gene (TSS+0.9) and loss of Pol II is observed on all probes in the transcription unit and downstream of the poly(A) site.

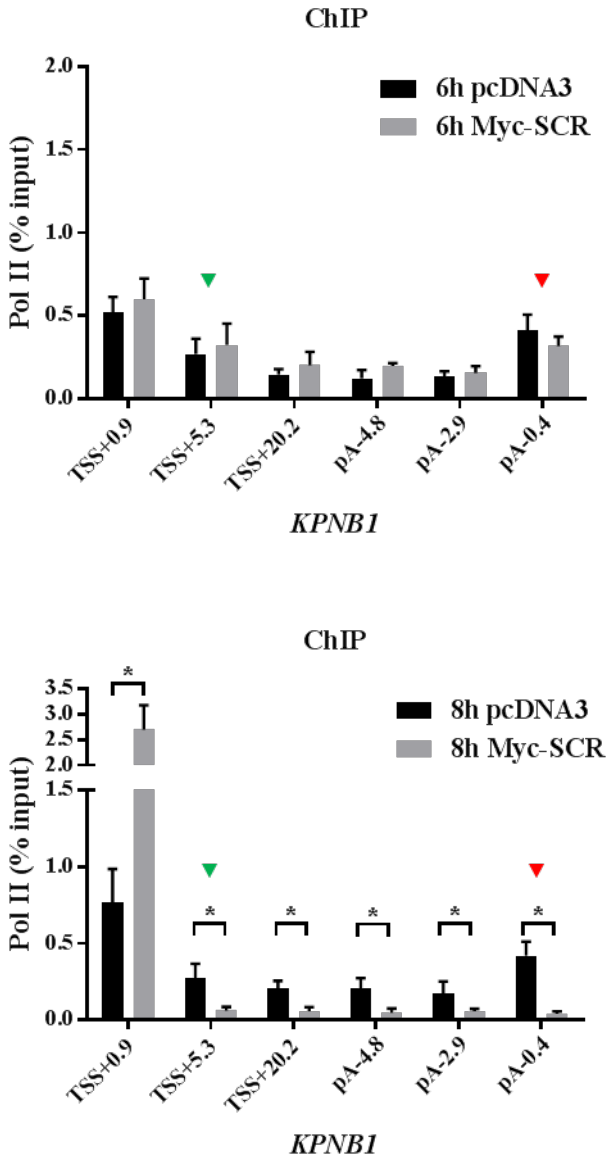


Figure 2.6. The results of qPCR analysis of Pol II ChIP after 6 h of transfection (top panel) or 8 h of transfection (bottom panel) of pcDNA3 or Myc-SCR of ICP22 in HEK293. The red arrow indicates the poly(A)-associated checkpoint. Error bars indicate the SEM of n = 3 independent experiments. Asterisks indicate the statistical significance compared with pcDNA3 (p<0.05).

2.6 Ectopic expression of ICP22 does not affect the KPNB1 protein level

Because Ser2P CTD Pol II marks the productive elongation state of Pol II, and ICP22 causes a global reduction of Ser2P, I asked whether the transient expression of ICP22 affects the expression of *KPNB1* at the protein level. The lysates after 24 h of transfection with Myc-FL and Myc-SCR were immunoblotted with an antibody recognising KPNB1 (Figure 2.7). The results indicate that the transient expression of

ICP22 does not affect the level of KPNB1 protein. Although the actively transcribing Pol II on *KPNB1* gene is prematurely terminated at both ends of the *KPNB1* gene following transient expression of ICP22, this result suggests that KPNB1 protein or its mRNA is stable over the time scale of ICP22 expression.

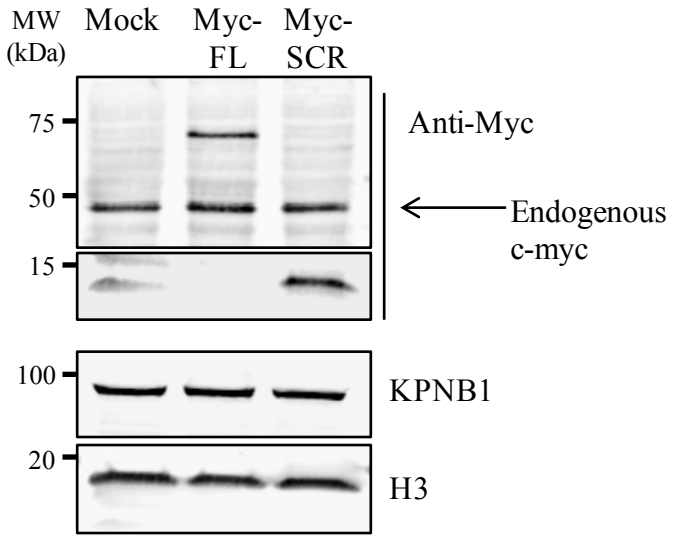


Figure 2.7. Immunoblot analysis of KPNB1 in HEK293 cell extract after transfection of a vector expressing Myc-tag full-length of ICP22 (Myc-FL) and Myc-tag ICP22 residues 193-256 (Myc-SCR). H3 is used as loading control.

2.7 Chapter summary

Downregulation of host gene expression is one of the strategies for regulating the HSV-1 life cycle. In order to understand the contribution of viral proteins in this process, I have studied the role of ICP22 and downregulation of host gene transcription using transiently-expressed ICP22. Since no other viral proteins were transiently expressed, any effects can be solely attributed to ICP22. Whether similar effects would be observed in infected cells has yet to be determined. I conclude that ICP22 affects the level of transcribing Pol II on the model protein-coding gene *KPNB1* in a similar way to the small molecule CDK9 inhibitor KM (Laitem et al., 2015). In summary, the ChIP-qPCR profile of Pol II on *KPNB1* 7 h post-transfection of the short conserved region (SCR) of ICP22 indicates that transient ectopic expression ICP22 residues 193-256 affects the level of transcribing Pol II at the beginning and end of the gene, which mirrors the effect caused by the CDK9 inhibitors, DRB and KM (Laitem et al., 2015). This result suggests that ICP22 possibly inhibits Pol II elongation at the beginning and end of genes due to inhibition

of the Ser2 CTD kinase, P-TEFb. However, this result does not rule out the possibility that the effects are due to the inhibition of alternative/additional CTD kinases. This possibility is explored in the next chapters.

The nucleosome plays a role in the Pol II transcription elongation. The increase in nucleosome occupancy is thought to impose a transcriptional block and obstruct the elongation of Pol II or to reinforce premature termination Pol II at the end of the gene (Egloff et al., 2009;Jonkers and Lis 2015). Cell treatment with CDK9 inhibitors reduces the level of H3K36me3 on a protein-coding gene, a histone mark associated with transcription of active euchromatin (Egloff et al., 2009;Wagner and Carpenter 2012). In this experiment, in cells treated with CDK9 inhibitor the level of H3 and H2B near the EEC and poly(A)-associated checkpoint increased, which supports the notion that nucleosomes are possibly assembled and positioned on regions that are no longer transcribed by Pol II, or an increase in nucleosomes near the poly(A)-associated checkpoint could play a role in prematurely terminating Pol II.

Since the global protein level of KPNB1 or its mRNA was unaffected by the transient expression of ICP22, I conclude that ICP22 is only affecting transcription of this gene and that the time scale (24 h) is too short to have a significant knockdown effect on the level of the protein. However, it would be expected that the protein level would be affected at longer times after transfection.

Chapter 3| Characterisation of the interaction proteome of ICP22

3.1 Chapter abstract

CTD kinases other than CDK9 are implicated in Pol II transcription. The CTD kinase CDK7/Cyclin H is part of the general transcription factor TFIIH and has a well-established role in promoter clearance and capping of pre-mRNA through phosphorylation of Ser5 of the CTD Pol II (McCracken et al., 1997;Nojima et al., 2015). CDK7 also activates CDK9 (Larochelle et al., 2012). CDK9 is associated with mainly Cyclin T1, T2a or T2b to form P-TEFb, a known Ser2 CTD kinase (Paparidis et al., 2017). CDK12 and CDK13, both associated with Cyclin K, are members of the CDK9 subfamily (Fu et al., 1999). CDK12 is capable of phosphorylating Ser2 Pol II CTD (Bartkowiak et al., 2010). CDK8 in association with Cyclin C is part of the Mediator complex that phosphorylates Ser5 of the CTD during transcription initiation (Tietjen et al., 2010). The roles of CDK10, CDK11, CDK19 (CDK8 paralogue) and CDK20 as CTD kinases are not yet clearly defined. All of these transcriptional kinases share structural similarities, cyclin subunits and most importantly the CTD of Pol II as their substrates.

Gene transcription involves the orchestrated action of many protein complexes in highly dynamic protein-protein interaction networks. To understand the biological roles of ICP22 in the regulation of cellular gene transcription, I performed ICP22 pull-down assays with HeLa cell extract and the cellular protein complexes bound to ICP22 were analysed by mass spectrometry. This work was expected to confirm the interaction between ICP22 and CDK9 and identify novel proteins that interact, directly or indirectly with ICP22.

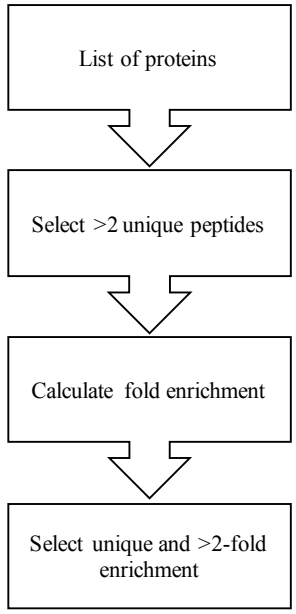
3.2 Proteomic analysis of proteins pulled down by ICP22

To identify ICP22-interacting partners, three pull-down samples were submitted for proteomic analysis to the Advanced Proteomics Facility, Department of Biochemistry Oxford. The three samples were; proteins from HeLa cell extract bound to His-tag IpgC of *Shigella flexneri* as a negative control, proteins from HeLa cell extract bound to His-tag full length of ICP22 (His-FL ICP22), and proteins from HeLa cell extract bound to His-SCR of ICP22 (His-SCR). Briefly, His-FL ICP22, His-SCR and His-IpgC were expressed in Rosetta^{TM} (DE3) *E.coli* and these recombinant proteins were

purified from *E.coli* using nickel-coated beads before incubating with HeLa cell extracts. The interactomes were eluted from the beads and subjected for proteomic analysis. Details of the expression and purification of these recombinant proteins are presented in Chapter 4.

Protein hits obtained were analysed further (Figure 3.1A). To define the most significant hits, two criteria were taken into consideration; 1) the number of unique peptides and, 2) the intensity, which is the sum of the top three most abundant unique peptides per sample protein. Proteins with more than two unique peptides identified were selected, as they were less likely to be artefacts. Fold enrichment of each proteins was calculated using the intensity value of the protein (bait) over the intensity value of the proteins of the negative control. Only proteins that show unique and greater than two-fold enrichment were selected for further analysis. The 'shortlisted' protein hits were further analysed (Figure 3.1B). For His-FL ICP22, a total of 666 proteins were selected, with the bait protein ranked seventh with seven unique peptides identified with an intensity value of 2.75 $x10^9$. The His-SCR of ICP22 pulled-down 503 proteins, with the bait protein is ranked second with two unique peptides identified with an intensity value of 2.75 $x10^9$. There was a considerable protein overlap between proteins interacting with full-length and the SCR of ICP22 as shown by the intersection sets of the Venn diagrams (Figure 3.1C). A total of 374 cellular proteins are commonly found in both full-length and SCR pull-downs, while 292 proteins were exclusively identified as proteins binding only to full-length of ICP22 and 129 proteins were exclusively identified as proteins binding only to the SCR. As the SCR is part of the full-length ICP22, it is possible that these 129 proteins uniquely identified from SCR pull-down are non-specifically bound proteins. The full list of 374 proteins is presented in alphabetical order (Appendix 3), by rank of uniqueness or fold enrichment for the His-FL of ICP22 pull down (Appendix 4), and by rank of uniqueness or fold enrichment for the His-SCR of ICP22 pull-down (Appendix 5). The list of 292 proteins unique to His-FL ICP22 is presented in alphabetical order (Appendix 6) or by rank of uniqueness or fold enrichment (Appendix 7). The list of 129 proteins unique to His-SCR of ICP22 is presented in alphabetical order (Appendix 8) or by rank of uniqueness or fold enrichment (Appendix 9).

A.



B.

Sample				Number of interacting proteins
Name	Rank in the list	Unique peptide	Intensity	
His-FL ICP22	7	7	2.75E+09	666
His-SCR ICP22	2	2	1.38E+10	503

C.

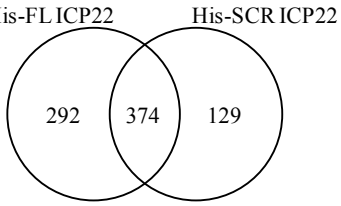


Figure 3.1. Qualitative analysis of ICP22 proteomics. A) Schematic of filtering criteria for the hits. Fold enrichment of the interacting protein was calculated using the intensity of the protein over the intensity of the negative control. **B) The summary table of the hits for the baits of His-FL and His-SCR**. A number of hits have been identified from each His-FL and His-SCR proteomes. **C) Venn-diagram comparing the proteins identified in His-FL and His-SCR pull-downs with HeLa cell extract.** The overlap includes proteins present in both samples. The full list of proteins can be found in Appendix 3 to 9.

3.3 Molecular functions associated with ICP22-bound cellular proteins

To understand the molecular function of proteins bound to ICP22, all 374 pulled-down proteins were analysed based on gene ontology (GO) annotations using the DAVID database (Huang da et al., 2009). From 374 proteins, 364 proteins were filtered by DAVID as proteins of Homo sapiens and 317 proteins were considered for the database search. Only 295 proteins were associated with a list of molecular functions. Each molecular function was presented with a p-value that is readily analysed by DAVID using the method of Benjamini-Hochberg where the p-value was obtained from multiple statistic tests (Benjamini and Hochberg 1995). The molecular function was selected using a corrected p-value of 0.01 as the cut-off for statistical significance thus indicating the number of genes that are significantly more enriched than random chance for a particular molecular function. The analysis classified 23 molecular functions of ICP22-bound cellular proteins and its gene count (Table 3.1). Details of each gene can be found in Appendix 10.

The list of the molecular function of ICP22-bound cellular proteins is visualized in

Figure 3.2. Each molecular function is colour-coded according to the fold-enrichment. Here, the fold enrichment is defined as the ratio of the ICP22 count genes involved in a particular molecular function divided by the expected genes by chance in the total population ((ICP22 count gene/ICP22 total gene) / (count gene/population total gene) of a particular molecular function). Based on the confidence level and fold enrichment value, Figure 3.2 shows that ICP22-bound cellular proteins are clustered in molecular function related to transcriptional activities.

Table 3.1. Table listing the number of ICP22-interacting proteins (gene count) found in a molecular function as classified by GO terms

No	GO molecular function	Gene count
1	RNA binding	66
2	RNA polymerase II transcription factor activity	40
3	general RNA polymerase II transcription factor activity	22
4	transcription regulator activity	95
5	RNA polymerase II transcription mediator activity	17
6	transcription activator activity	45
7	transcription coactivator activity	33
8	transcription cofactor activity	39
9	DNA binding	110
10	transcription factor binding	42
11	vitamin D receptor binding	11
12	chromatin binding	21
13	thyroid hormone receptor binding	11
14	structural constituent of ribosome	21
15	ligand-dependent nuclear receptor transcription coactivator activity	10
16	nuclear hormone receptor binding	13
17	hormone receptor binding	13
18	zinc ion binding	85
19	transcription repressor activity	22
20	RNA splicing factor activity, transesterification mechanism	7
21	helicase activity	12
22	cyclin-dependent protein kinase activity	6
23	transcription factor activity	39

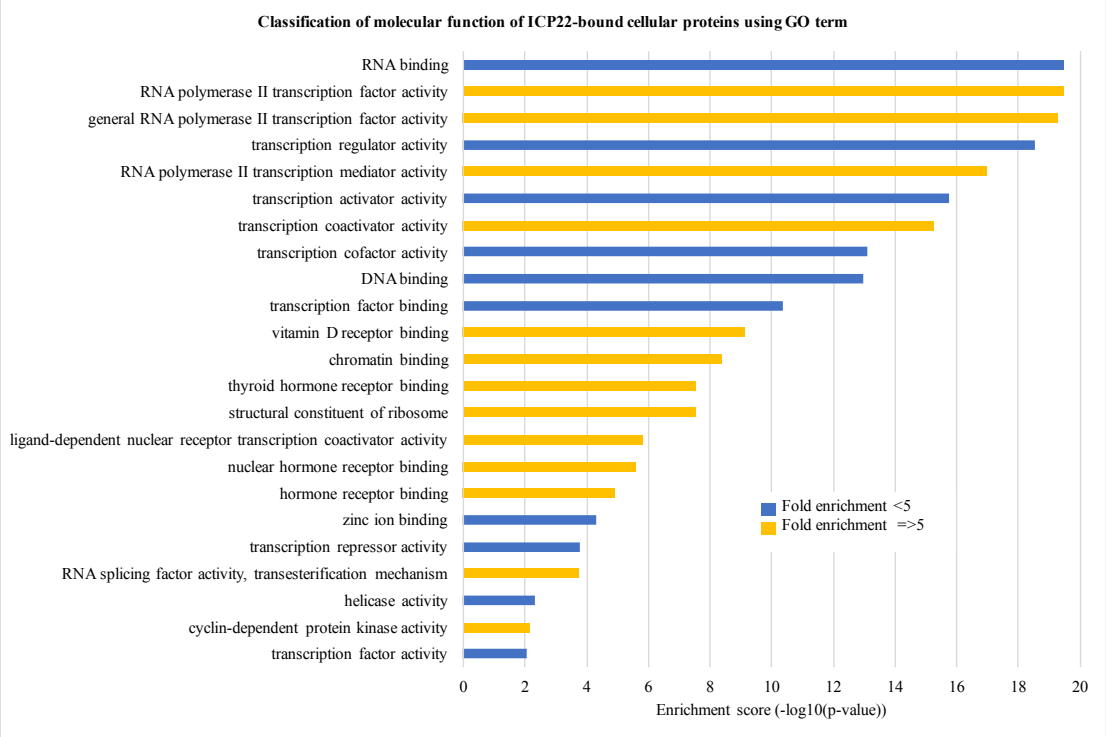


Figure 3.2. Graphical representation of bioinformatic analysis of ICP22-interacting proteins classified by GO terms in molecular function. Enrichment score is –log10-transformed of p-value.

3.4 Confirmation of interactions between ICP22 and transcriptional kinases using mass spectrometry

A list of relevant hits to the current study categorised into functional groups is presented in Table 3.2. Among the 374 significant protein hits, CDK9 which is a previously-identified ICP22-interacting protein, was found. Previous study has identified that Cyclin T1 (CCNT1) and Cyclin T2 (CCNT2) are the cyclin partner of CDK9 (Peng et al., 1998;Baumli et al., 2008). Both cyclins were also present in the list of hits. Several CDK9-associated proteins such as its inhibitory complex LARP7 and MEPCE, and its regulatory factors Brd4, PAF1c and SEC are also present. The mass spectrometry analysis therefore confirms the interaction of ICP22 with CDK9, its cyclin partners and has identified several other CDK9-associated proteins. In addition, ICP22 also pulled-down CDK12 and CDK13, as well as CDK8/CDK19 and CDK11B. As the proteomic analysis indicates that ICP22 interacts with several CDK9 or P-TEFb associated complexes, this could also mean that ICP22 also pulled down cellular factors associated with CDK11B, CDK12, CDK13 and CDK8/CDK19. When compared with a list of interactomes from protein-protein interaction database

BioGrid, several known interactors of these transcriptional kinases are present in the list of proteins pulled down by ICP22 (Figure 3.3).

The technique of using recombinant ICP22 for pull-down with HeLa cell extract has therefore given many relevant hits, including CDK9/P-TEFb and its associated factors, as well as other transcriptional kinases and their associated factors. This suggests that ICP22 is capable of binding to several large complexes. This study has therefore been particularly useful for the identification of novel interacting partners.

Table 3.2. List of ICP22-interacting proteins of interest categorised into functional groups. These are ranked by uniqueness and fold enrichment in Appendix 3-5**.**

No	Gene names	Unique peptide	ICP22 FL			ICP22 SCR			Group
			Unique peptide	Intensity	Fold Enrichme	Unique peptide	Intensity	Fold Enrichme	
1	CDK11B	6	4	1.65E+10	24.1	6	2.17E+10	31.8	CDKs
2	CDK12	32	23	9.28E+09	12.7	29	1.67E+10	22.9	
3	CDK9	15	14	2.70E+09	81.1	15	4.57E+09	137.6	
4	CDK8;CDK19	6	5	2.78E+08	Unique	4	1.26E+08	Unique	
5	CDK13	11	3	7.69E+07	Unique	2	3.41E+07	Unique	
6	CCNK	17	12	3.81E+09	7.4	13	4.75E+09	9.2	Cyclins
7	CCNT1	14	9	2.02E+09	38.9	14	4.94E+09	95.2	
8	CCNT2	11	7	5.61E+08	Unique	10	8.21E+08	Unique	
9	MEPCE	25	17	3.02E+09	8.3	22	4.82E+09	13.3	Inactive P-TEFb
10	LARP7	20	16	1.10E+09	18.2	18	7.54E+08	12.4	
11	MED1	50	39	5.02E+09	10.8	42	4.74E+09	10.2	Mediator
12	MED26	12	8	5.78E+08	19.2	10	5.13E+08	17.0	
13	MED23	15	9	2.76E+08	Unique	6	9.66E+07	Unique	
14	BRD4	41	34	2.01E+10	8.3	36	1.60E+10	6.6	P-TEFb associated
15	HTATSF1	20	19	2.85E+09	Unique	3	2.39E+08	Unique	
16	CTR9	34	33	7.68E+09	41.5	16	9.05E+08	4.9	PAF1c
17	PAF1	23	22	4.12E+09	17.6	12	4.76E+08	2.0	
18	AFF4	32	21	1.70E+09	4	28	2.27E+09	5.4	SEC

Notes: The proteins were identified from both His-FL (blue box) and His-SCR (green box) of ICP22 pull-downs. The first unique peptide column shows the total number of peptides that could potentially be identified. The other unique peptide columns show the total number of peptides identified in each of the indicated pull-down with His-FL or His-SCR of ICP22. The intensity value is the sum of the top three most abundant unique peptides. Fold enrichment was calculated using the intensity value of the protein over the intensity value of the proteins of negative control.

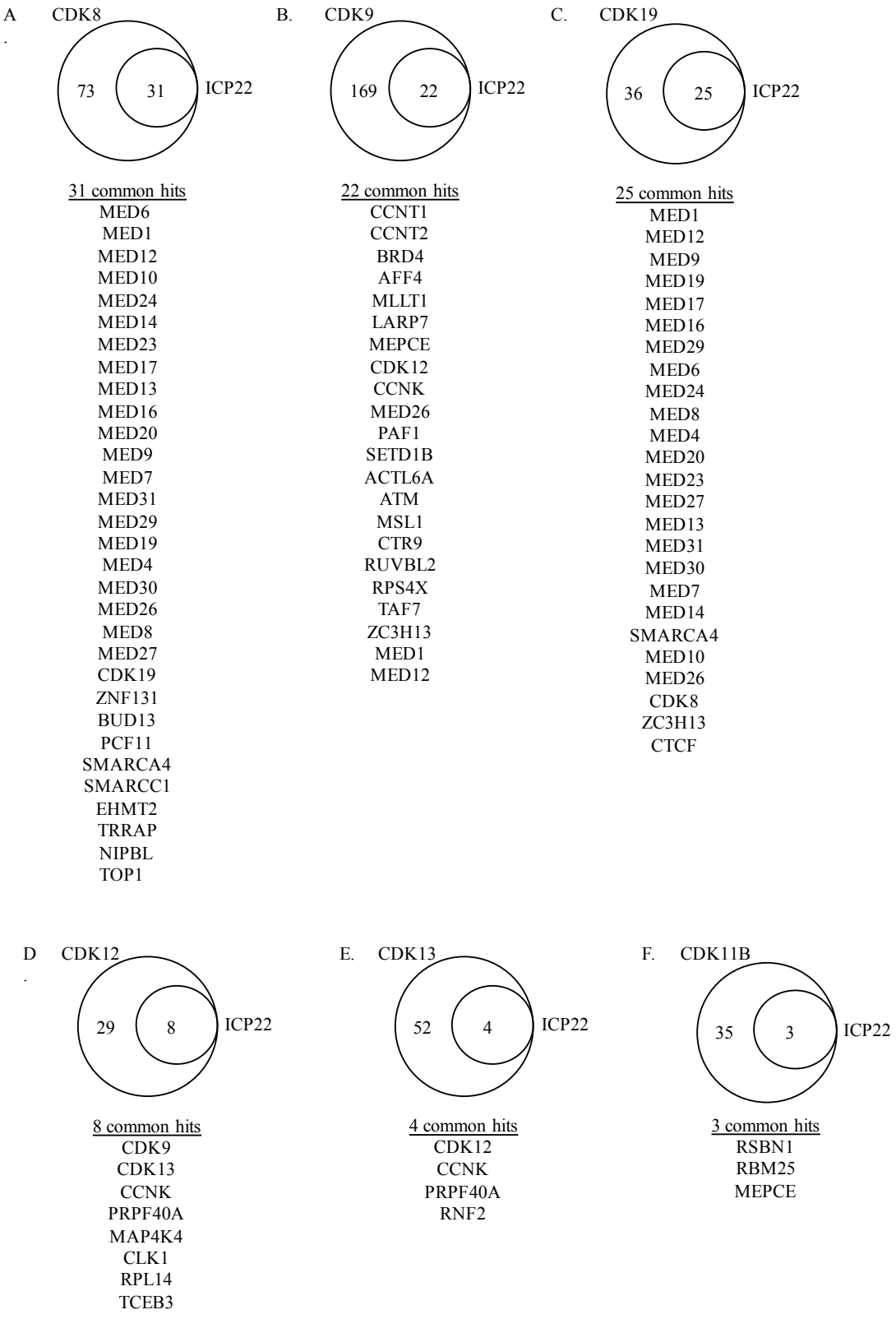


Figure 3.3. Visual representation of the transcriptional kinases complexes interacting with ICP22. Proteins bound to CDK8, CDK9, CDK19, CDK12, CDK13, CDK11B were listed in the BioGrid, a protein-protein interaction database. The overlap includes proteins present in the ICP22 pull-downs from HeLa cell extract.

3.5 Chapter summary

The HSV-1 viral proteins have different activities at different time points post-infection. As a result, the profile of virus-cellular protein-protein interaction may be different depending on the course of the HSV-1 infection. The list of viral interactomes may differ in studies using cells where the viral proteins are over-expressed, or using the purified viral protein as a bait for pull-down using human cell extract. The latter strategy is the basis of this study. Using the purified viral protein from a bacterial expression system offers several advantages. It eliminates the involvement of other viral proteins and the interaction between the viral and cellular protein is therefore exclusive. The lack of post-translational modification of the purified viral protein from a bacterial expression system indicates that its interaction with the cellular protein requires very minimal or no post-translational modification.

In this study, I performed proteomic analysis to examine the cellular proteins from HeLa cell extract interacting with purified ICP22 protein of HSV-1, produced in a bacterial expression system. The analysis identified CDK9 and confirms its interaction with ICP22 as shown by immunoblot (Chapter 4) and the previously published interaction of ICP22 and CDK9 (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014). Moreover, the CDK9 cyclin-partners Cyclin T1 and Cyclin T2 were also shown to co-purify with ICP22, which strongly suggests that ICP22 binds to P-TEFb. Several P-TEFb-associated factors such as P-TEFb inhibitory factors (LARP7 and MEPCE as part of the inactive form of P-TEFb, and P-TEFb regulatory factors (Brd4, PAF1c and SEC) were also pulled down by ICP22. This suggests that ICP22 interaction could identify binding partners or novel regulatory factors of P-TEFb.

The interaction of ICP22 and transcriptional kinases other than CDK9 has never been investigated before. Combinatorial mass spectrometry and immunoblotting (Chapter 4) of the cellular proteins pull-down by purified ICP22 used in this study resulted in the identification of transcriptional kinases other than CDK9, namely CDK11B, CDK12, CDK13 and CDK8/CDK19 as ICP22-interacting transcriptional kinases. The novel regulatory factors of CDK11B, CDK12, CDK13 and CDK8/CDK19 that are pulled down using this technique warrant further examination.

The results presented here by proteomic analyses of ICP22 binding partners therefore provide insights regarding the function of ICP22 as part of the framework of host transcription regulation, and represent a first step towards developing small molecule

inhibitors targeting specific motifs or domains within ICP22 to break the interaction with host cell proteins for HSV-1 therapeutic purposes.

Chapter 4| Characterisation of the interaction between ICP22 and cellular proteins

4.1 Chapter abstract

The kinase activity of the CDK9 subunit of P-TEFb is essential for Pol II transcription elongation and co-transcriptional processing. Inhibition of CDK9 results in failure of Pol II to negotiate the early elongation checkpoint (EEC) and premature termination of Pol II at the newly-described poly(A)-associated checkpoint (Laitem et al., 2015). In Chapter 2, I have shown that a similar effect on elongating Pol II is observed when HEK293 cells are transiently expressing ICP22. As ICP22 mimics the effect of CDK9 inhibitors on the elongating Pol II, an important question now is how ICP22 inhibits CDK9. To address this question, I have explored the interaction between ICP22 and P-TEFb using purified recombinant proteins produced using a bacterial expression system. It has already been shown that residues 193-256 of ICP22 are sufficient for interaction with CDK9 (Zaborowska et al., 2014). My next step was to further define the interaction between ICP22 and P-TEFb. The data presented here serve as a basis for more detailed structural analysis.

4.2 Expression and purification of His-tag ICP22 recombinant proteins using a bacterial expression system

The starting point for this work was the finding that the short conserved region (SCR) of ICP22 comprising residues 193-256 interacts directly *in vitro* with P-TEFb comprising full-length CDK9 and residues 1-266 of Cyclin T1 produced using a Baculovirus expression system (Baumli et al., 2012;Zaborowska et al., 2014). However, whether ICP22 interacts with CDK9 or Cyclin T1 or both subunits of P-TEFb was unclear. To further examine the relationship between ICP22 and subunits of P-TEFb, I have used tagged recombinant proteins produced using a bacterial expression system. His-SCR and His-SCR-5/11 were made using the pET28a plasmid (Figure 4.1A). His-SCR-5/11 is a truncated version of the SCR missing five amino acids from the N-terminus and 11 amino acids from the C-terminus and with an N-terminal His-tag. Full-length CDK9 has an N-terminal GST-tag (GST-CDK9) and Cyclin T1 residues 1-266 has an N-terminal Flag tag (Flag-Cyclin T1 residues 1-266). His-SCR and His-SCR-5/11 were expressed in Rosetta^{TM} (DE3) *E.coli* and purified

from *E.coli* using nickel-coated beads before being evaluated in a His-tag pull-down assay for their ability to bind to subunits of P-TEFb. The poly-histidine tag (His-tag) is a short affinity tag consisting of six or more histidine residues. The amino acid histidine exhibits a strong interaction with immobilized metal ion matrices. The short, minimal charge and unstructured characteristics of the His-tag are ideal for the assay, as they eliminate possible interference with the function of the recombinant proteins compared to larger and more structured protein tags. During overnight expression of His-SCR in *E.coli* at temperatures lower than 30˚C in the presence of IPTG as inducer, the expression of His-SCR appears to impart a metabolic burden and is toxic to the cells as observed from the reduction of bacterial cells growth and the recombinant proteins are severely degraded. However, a shorter induction time (7 h) of His-SCR gives a yield of recombinant proteins sufficient for pull-down experiments. To confirm the expression of the recombinant proteins in *E.coli*, IPTG-induced bacteria were lysed in Laemmli buffer and subjected to SDS-PAGE followed by protein identification using Coomassie staining and immunoblots probed with anti-His tag antibody. The Coomassie stains and immunoblots shown in Figure 4.1B indicate that the His-SCR migrates on SDS-PAGE roughly as an ~12 kDa protein and His-SCR-5/11 as an ~10 kDa protein. In the lanes where IPTG is not added to the culture but the bacterial cells were harvested at the same time as the IPTG-induced culture, the immunoblots show faint protein bands that correspond to His-SCR and His-SCR-5/11. This is because at the time of harvest, the bacterial cells carrying pET28a are expressing the recombinant proteins at low levels in the absence of IPTG inducer due to the 'leaky' T7 promoter of pET28a in combination with the Rosetta™ *E.coli* strain used in this study.

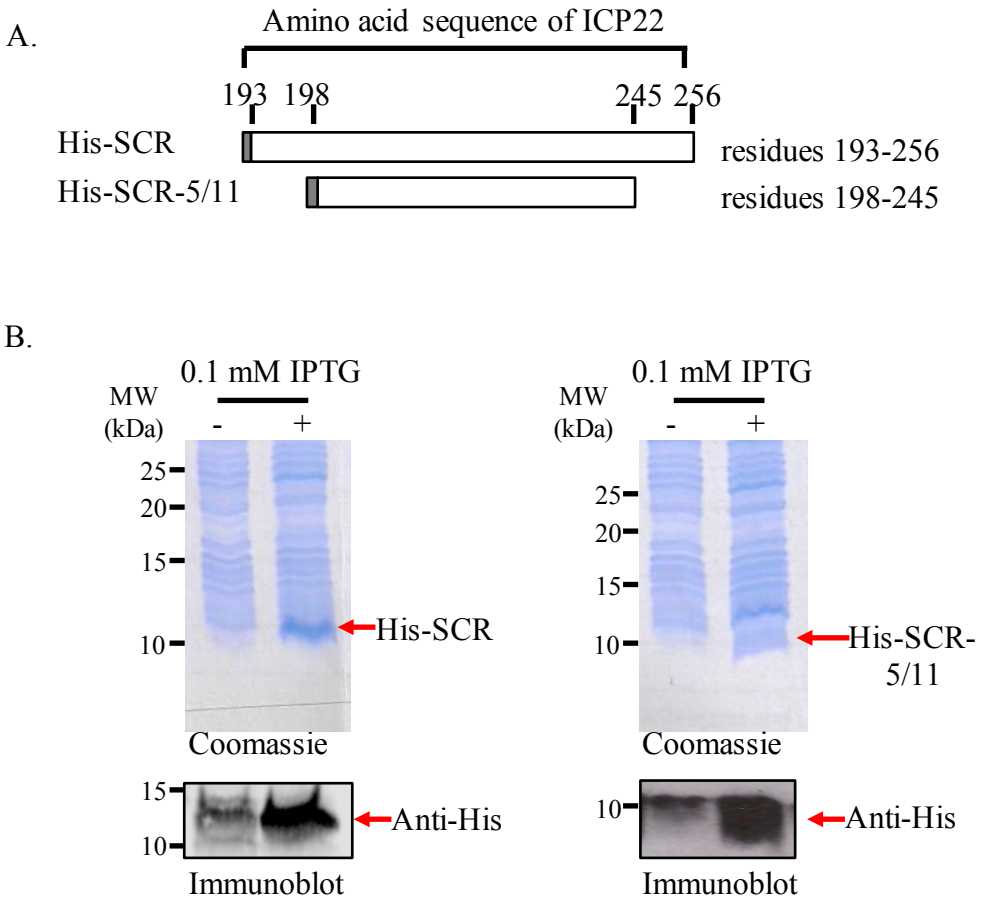
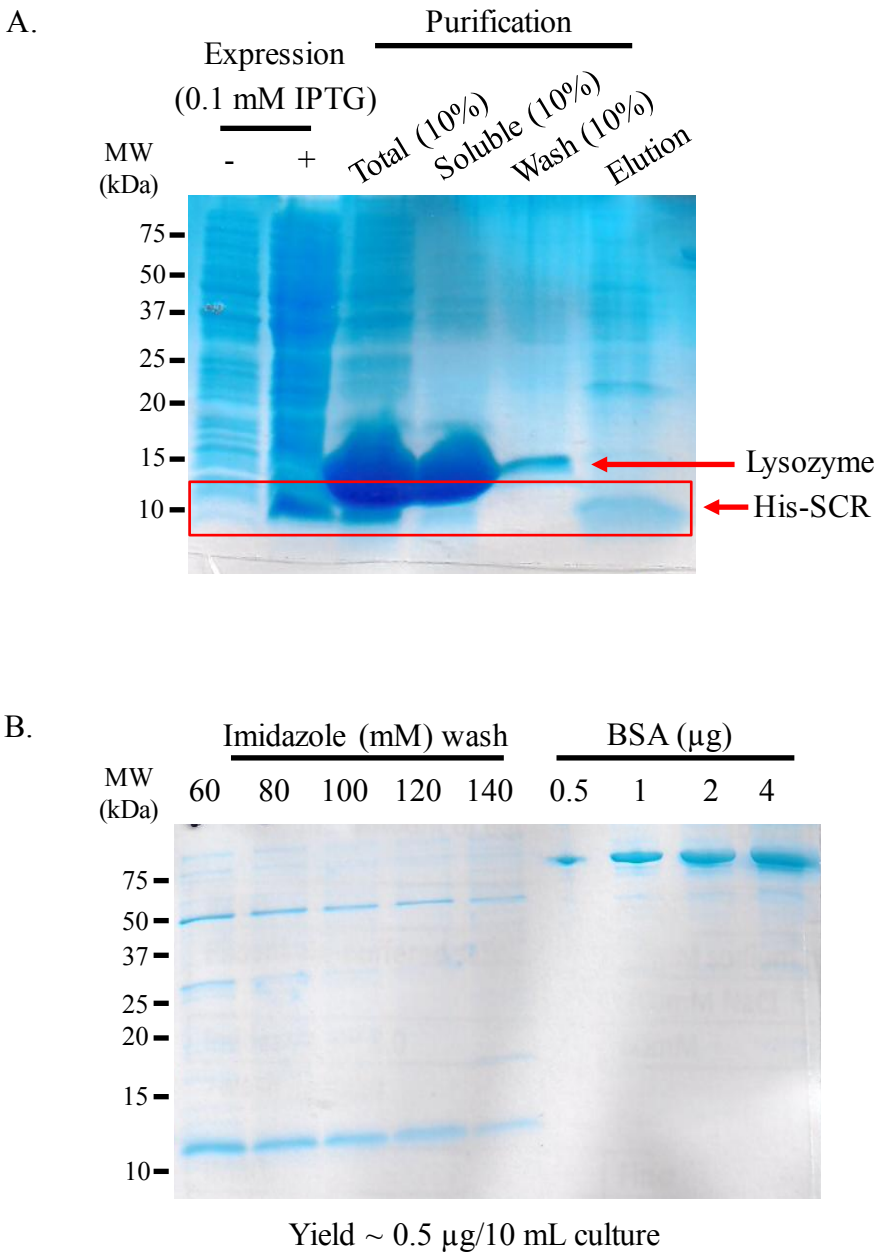


Figure 4.1. Expression, purification and identification of His-SCR and His-SCR-5/11. A) Line drawing of His-SCR and His-SCR-5/11. The grey box indicates the His-tag fused to the N-terminal of His-SCR (residues 193-56) and His-SCR-5/11 (residues 198-245) of ICP22. **B) Identification of the expressed His-SCR and His-SCR-5/11.** Coomassie stain (top) and immunoblot (bottom) of expressed His-SCR (left panel) and His-SCR-5/11 (right panel) in the presence of 0.1 mM IPTG inducer for 7 h at 30˚C. Red arrows point to the His-SCR at approximately ~12 kDa and His-SCR-5/11 at approximately ~10 kDa. Both immunoblots show a band in the left hand lanes where IPTG inducer is not added to the bacterial culture, indicating that a low level of recombinant protein is expressed in the absence of IPTG inducer.

To prepare the recombinant proteins for pull-down experiments, the induced His-tag recombinant proteins were purified. Figure 4.2A shows the purification steps of His-SCR as an example, using lysis buffers optimised to obtain the maximum yield of soluble His-SCR. After harvesting the IPTG-induced bacterial culture, the bacterial cells were lysed with lysozyme-containing lysis buffer, and then sonicated before removal of the bacterial debris by centrifugation. The soluble His-SCR was then incubated with nickel-coated beads and the His-SCR isolated from the bacterial proteins. The bound His-SCR bound to the nickel-coated beads were then washed

with increasing concentrations of imidazole to remove non-specifically bound bacterial proteins and finally eluted with a high concentration of imidazole (500 mM). Figure 4.2B shows the eluted His-SCR after washing the bound His-SCR with increasing concentration of imidazole. Washing buffer with 100 mM Imidazole was used, as higher amounts of imidazole reduce the amount of the bound His-SCR but do not completely eliminate the non-specifically bound bacterial protein observed at ~50 kDa. This unknown bacterial protein perhaps contains stretches of histidine that bind strongly to the nickel-coated beads. The amount of recombinant protein required for a His pull-down assay varies from 1 to 5 µg. The resulting yield of purified His-SCR was approximately 0.5 µg from 10 mL cell culture. Thus, upscaling the cell culture is necessary to produce sufficient protein for pull-down experiments.



Yield ~ 0.5 µg/10 mL culture

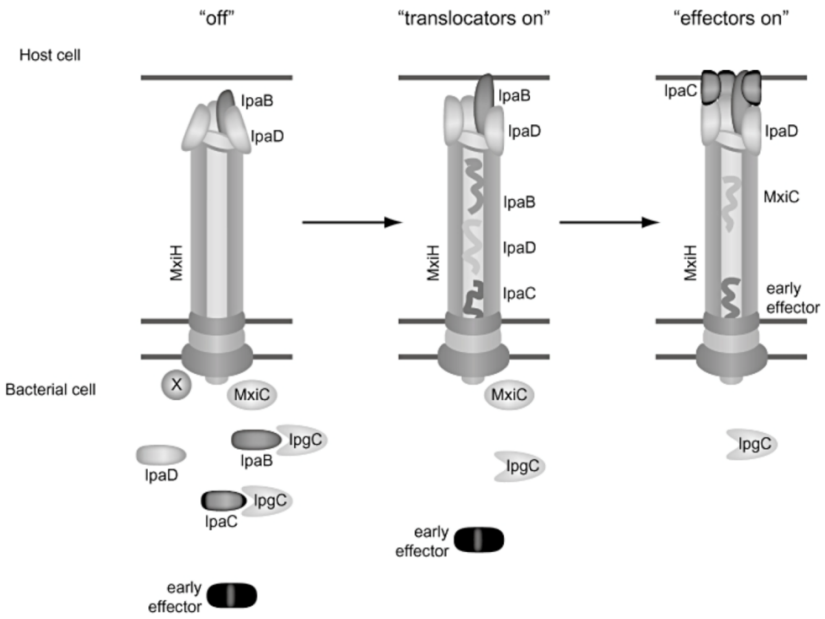
Figure 4.2. Coomassie stained gels of the different stages of purification. A) Purification steps of His-SCR (red box and red arrow). The red arrow indicates the lysozyme in the lysis buffer. The lanes, from left to right, show proteins from Rosetta™ *E.coli* transformed with His-SCR expression vector without IPTG, proteins from Rosetta™ *E.coli* transformed with the His-SCR expression vector after induction with IPTG, supernatant containing expressed His-SCR pre-sonication (total), supernatant containing His-SCR post-sonication and centrifugation (soluble), wash of the beads with 50 mM imidazole (wash), and eluted His-SCR from the nickel-coated beads (elution). **B) Bound His-SCR on the nickel-coated beads were washed with wash buffer containing increasing concentrations of imidazole (60-140 mM).** The titration of BSA serves as reference to the amount of His-SCR. The wash buffer with 100 mM imidazole produces approximately 0.5 µg of purified His-SCR from 10 mL cell culture.

Since the contaminant bacterial protein remains bound on the beads, a good negative control is required for pull-down experiments. The unbound nickel-coated beads itself is one appropriate negative control. I have also used the *Shigella flexneri* protein IpgC as a negative control.

Shigella flexneri causes shigellosis, characterised by frequent passage of stools containing blood and mucus accompanied by painful abdominal cramps. The bacteria invade intestinal epithelial cells and multiply inside these cells. The infected epithelial cells undergo necrosis and dying host cells release bacteria which spread from cell to cell. Briefly, the onset of shigellosis requires the bacteria to invade the host colonic epithelial cells. *Shigella* has a type III secretion system (TTSS) to deliver its translocator and effector proteins to the host cell membrane and cytoplasm needed to colonize the colonic epithelium (Figure 4.3A). The TTSS inject 'effector' proteins through a 'needle' protruding from the bacterial surface directly into host cells after assembly of a 'translocator' pore in the host plasma membrane. Injection and secretion is a tightly regulated process, which is blocked until physical contact with a host cell takes place. The *Shigella* translocator proteins, IpaB and IpaC, prior to their secretion are partitioned in the bacterial cytoplasm by association with the cytoplasmic chaperone IpgC. Upon host cell contact, the IpaB/IpgC complex dissociates from IpgC and then the *Shigella* TTSS inserts IpaB and IpaC into the host cell membrane to form a pore that permits passage of other *Shigella* effectors into the host cell cytoplasm. As a chaperone protein, IpgC remains in the bacterial cells (Figure 4.3A). To date it is known that IpgC only interacts with *Shigella* proteins IpaB and IpaC, and IpgC may therefore be a good negative control. The sequence alignment of IpgC and SCR of ICP22 shows little identity between the two proteins (Appendix 11).

The expression plasmid for N-terminal His-tag IpgC was kindly provided by Professor Susan Lea (Sir William Dunn School of Pathology Oxford). The His-IpgC expression was induced overnight at 25˚C with 1 mM IPTG, as suggested by Professor Lea. The Coomassie stain and immunoblot in Figure 4.3B indicate that the His-IpgC migrates on SDS-PAGE as an ~17 kDa protein.

A.



B.

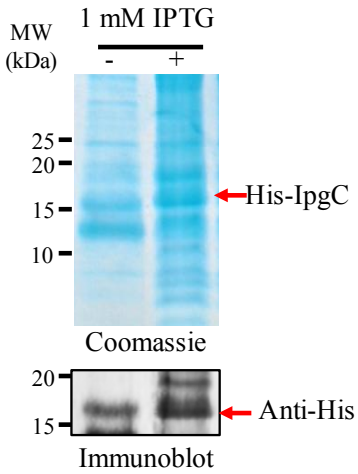


Figure 4.3. Expression of the *S. flexneri* protein IpgC as a negative control for the His pull down. A) *S. flexneri* has a type III secretion system (TTSS) to release the bacterial protein during infection into the host cell. IpgC serves as intrabacterial chaperone of IpaB and IpaC. Upon physical contact with host cells, IpgC releases IpaB and IpaC and they are translocated into the host cells via the TTSS (Martinez-Argudo and Blocker 2010). As a chaperone, IpgC is thought to only interact with its bacterial protein counterparts. **B) Identification of expressed His-IpgC.** Coomassie stain (top) and immunoblot (bottom) of expressed His-IpgC in the presence of 1 mM IPTG at 25˚C overnight. The red arrows point to the His-IpgC at approximately ~17 kDa.

4.3 Expression and purification of GST-CDK9 and Flag-Cyclin T1 recombinant proteins from a bacterial expression system

Interaction between ICP22 and CDK9 so far has been shown using GST-CDK9 incubated with HSV-1 infected cell lysates and immunoblotted for ICP22 (Durand et al., 2005), or cellular proteins immunoprecipitated with ICP22 from CHO-K1 transiently expressing ICP22 or from HSV-1 infected HEp-2 cells, that were then immunoblotted for CDK9 (Guo et al., 2012). Both of these studies used full-length ICP22 and cellular P-TEFb. Our lab used both the GST-ICP22 full-length and GST-ICP22 residues 193-256 to study interaction with CDK9 in HeLa cell extract or with P-TEFb comprising full-length CDK9 and residues 1-266 of Cyclin T1 produced in a Baculovirus system (Zaborowska et al., 2014). To date none of the works cited show whether ICP22 interacts with either or both subunits of P-TEFb. To further characterise which subunit of P-TEFb interacts with ICP22, CDK9 and Cyclin T1 were expressed individually and purified separately. The full length CDK9 fused with GST on its N-terminal was first expressed and purified on glutathione beads, while Cyclin T1 residues 1-266 fused with Flag-tag on its N-terminal was expressed and purified on M2 beads.

CDK9 expressed from a pGEX-6P-1 vector and containing an N-terminal GST tag was chosen for expression of full-length CDK9 (GST-CDK9). The empty pGEX-6P-1 vector construct expressing GST alone was used as a negative control. A cDNA encoding Cyclin T1 residues 1-266 carrying N-terminal 3x Flag peptides was subcloned into the bacterial expression vector pET28a (Flag-Cyclin T1 (residues 1-266)). The plasmids were transformed into Rosetta^{TM} (DE3) *E.coli* competent cells, and protein expression was induced with IPTG once the bacteria reached the logarithmic growth phase with the OD_{600} measurement approximately ~0.6-0.8. After harvesting, the bacterial cells were lysed in Laemmli buffer and subjected to SDS-PAGE followed by protein identification of GST-CDK9 and Flag-Cyclin T1 (residues 1-266) using Coomassie staining and immunoblots probed with anti-GST and anti-Flag antibodies, respectively (Figure 4.4A). Once the expression of both recombinant proteins was confirmed, the recombinant proteins were isolated and purified accordingly (Figure 4.4B). Glutathione agarose beads were used to bind GST-CDK9 and GST, while agarose beads coated with anti-Flag M2 monoclonal antibody were used to bind Flag-Cyclin T1 residues 1-266. GST-CDK9 and GST were dissociated from glutathione agarose beads with 10 mM reduced glutathione and 3x Flag peptide

is used for the competitive elution of Flag-Cyclin T1 residues 1-266 from anti-Flag M2 agarose beads.

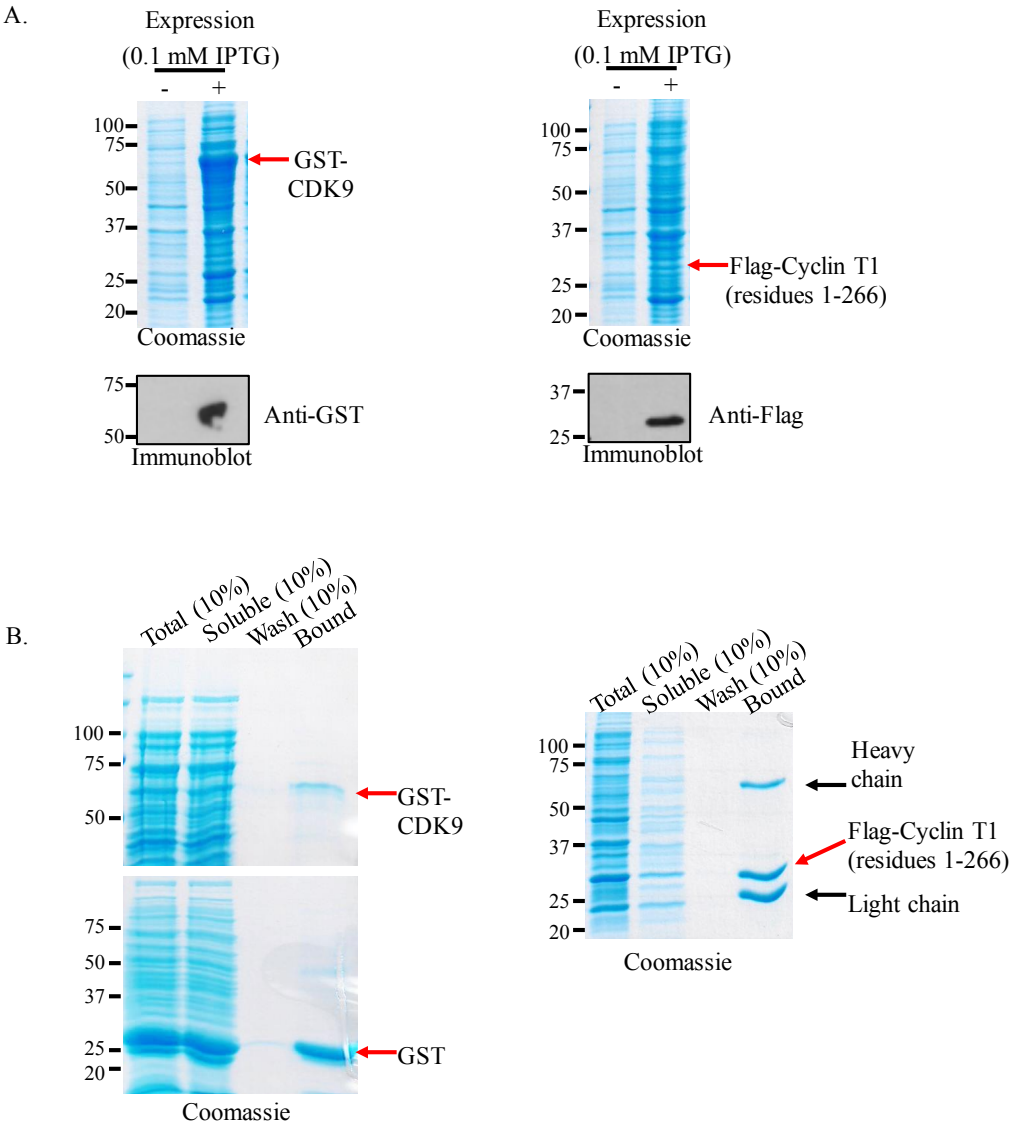


Figure 4.4. Expression, detection and purification of GST-CDK9 and Flag-Cyclin T1 (residues 1-266). A) Coomassie stain (top) and immunoblot (bottom) of expressed GST-CDK9 (left panel) and Flag-Cyclin T1 residues 1-266 (right panel). Culture samples were taken immediately upon IPTG induction and 7 h after induction by IPTG at 30˚C. The red arrows point to the GST-CDK9 at approximately ~55-60 kDa and Flag-Cyclin T1 residues 1-266 at approximately ~30 kDa. **B) Purification steps of GST-CDK9 (top left panel), GST (top bottom panel) and Flag-Cyclin T1 residues 1-266 (right panel).** The lanes from left to right show supernatant containing the recombinant proteins pre-sonication (total), supernatant containing the recombinant proteins post-sonication and centrifugation (soluble), bead washes (wash), and the beads with bound recombinant proteins were boiled in the Laemmli buffer (bound).

4.4 Direct interactions of ICP22 and CDK9 or Cyclin T1

After the recombinant proteins of SCR and SCR-5/11 of ICP22 and subunits of P-TEFb were successfully expressed and purified, the next step was to evaluate the ability of SCR and SCR-5/11 of ICP22 to bind CDK9 or Cyclin T1. To start with, GST-CDK9 alongside GST as a negative control were eluted from the glutathione beads with reduced glutathione, and the reduced glutathione was then removed from the eluate through buffer exchange using a Vivaspin spin column. A total of 0.5% of supernatant containing GST-CDK9 or GST were kept as input. Two tubes of equal amount of His-SCR or His-SCR-5/11 were prepared. Each set of His-SCR and His-SCR-5/11 was incubated with equivalent amounts of GST-CDK9 and GST. After overnight incubation and extensive washing, the bound materials were separated by SDS-PAGE and then analysed by immunoblots using anti-His and anti-GST antibodies. As shown in Figure 4.5, the His-SCR did not interact with GST as no bands appeared at the size of ~25 kDa but pulled down GST-CDK9, shown by the presence of a protein band at approximately ~55-60 kDa. Similarly, the His-SCR-5/11 did not bind GST but pulled down GST-CDK9 (Figure 4.5). This evidence suggests that ICP22 interacts directly with CDK9 and residues 198-245 of ICP22 (His-SCR-5/11) are sufficient for the binding with CDK9.

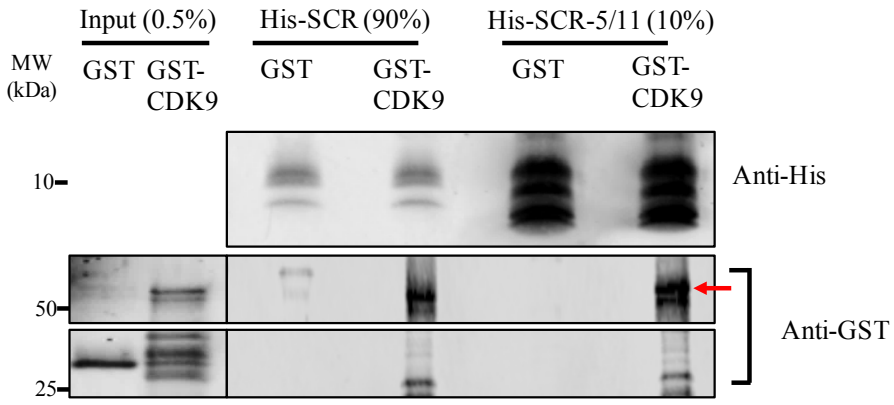


Figure 4.5. ICP22 interacts directly with CDK9. From left to right: A total of 0.5% of the purified GST as a negative control and GST-CDK9 were separated by the SDS-PAGE and immunoblotted with an anti-GST antibody. GST migrates at the size of ~26 kDa while GST-CDK9 migrates at approximately ~55-60 kDa. An equal amount of His-SCR or His-SCR-5/11 were then incubated with GST and GST-CDK9. A total of 90% of bound proteins from His-SCR and 10% from His-SCR-5/11 were separated by the SDS-PAGE, followed by immunoblotting against anti-His and anti-GST antibodies. The red arrows points to the GST-CDK9.

The next step was to evaluate the interaction of His-SCR and His-SCR-5/11 with Flag-Cyclin T1 residues 1-266. The purified Flag-Cyclin T1 residues 1-266 was eluted from anti-Flag M2 agarose beads with 3x Flag peptides. In this experiment, the *S. flexneri* protein IpgC (His-IpgC) was used as negative control. A total of 1% of supernatant containing Flag-Cyclin T1 residues 1-266 were kept as input. An equal amount of Flag-Cyclin T1 residues 1-266 were equally distributed to tubes containing His-IpgC as negative control, His-SCR and His-SCR-5/11. After overnight incubation and extensive washing, the bound complexes were separated by SDS-PAGE followed by immunoblot against anti-His and anti-Flag antibodies. As shown in Figure 4.6, the negative control His-IpgC shows some interaction with Flag-Cyclin T1 residues 1-266, although this is less than with His-SCR. As to my knowledge, IpgC only binds to *Shigella* proteins, it is difficult to conclude from this observation that the IpgC itself can interact specifically with Cyclin T1. Also, His-SCR appears to pull-down Flag-Cyclin T1 residues 1-266 while the His-SCR-5/11 does not. Although only 10% of bound complexes of His-SCR-5/11 were loaded onto the gel, the amount of His-SCR-5/11 as a bait for pull-down is higher than the 90% of His-SCR, as observed from the intensity of the bands detected by anti-His antibody. Based on this pull-down, it can be suggested that His-SCR interacts with Flag-Cyclin T1 residues 1-266. The interaction was lost when five amino acids on its N-terminus and 11 amino acids on its C-terminus were deleted (the His-SCR-5/11). However, this would have to be repeated with, for example, GST-Cyclin T1 and GST alone as negative control to confirm this interaction.

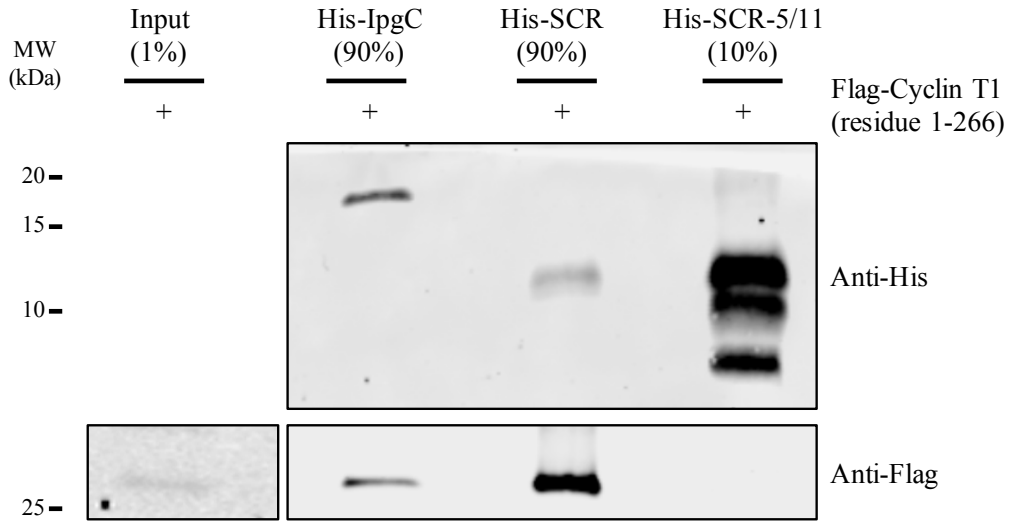


Figure 4.6. ICP22 residues 198-245 (His-SCR-5/11) fail to pull-down Cyclin T1. From left to right: A total of 1% of the purified Flag-Cyclin T1 residues 1-266 was separated by the SDS-PAGE and immunoblotted with anti-Flag antibody. An equal amount of Flag-Cyclin T1 residues 1-266 were distributed into individual tubes containing His-IpgC, His-SCR and His-SCR-5/11. A total of 90% of bound proteins from His-IpgC and His-SCR, and 10% of bound proteins from His-SCR-5/11 were separated by the SDS-PAGE, followed by immunoblotting against anti-His and anti-Flag antibodies.

Since His-SCR-5/11 did not pull-down Flag-Cyclin T1 residues 1-266 effectively, I decided to analyse the deleted residues. Figure 4.7 shows the alignment of SCR and SCR-5/11 of ICP22 of HSV-1 and the core conserved region of the ICP22 homologue in other alphaherpesviruses, based on the previously published sequence analysis (Zaborowska et al., 2014). SCR-5/11 has lost the 193-YMRQCI-197 residues at the N-terminus of the SCR and 246-VSGGTWGMHLR-256 residues on the C-terminus of the SCR. The 193-YMRQCI-197 does not have any highly conserved residues within alphaherpesviruses but the 246-VSGGTWGMHLR-256 shares the last two residues within alphaherpesviruses. Whether Leu-255 and Arg-256 residues of HSV-1 ICP22 are responsible for the interaction with Cyclin T1 is yet to be determined. The VZV homolog of ICP22 does not trigger loss of Ser2P Pol II CTD (Fraser and Rice 2007). To date there are very limited studies on the interaction of the VZV ICP22 homolog with P-TEFb, but it may not interact with P-TEFb. Assuming that VZV ICP22 homolog is not interacting with P-TEFb hence fail to induce loss of Ser2P Pol II CTD, and given that Leu-255 and Arg-256 are present in ICP22 of both HSV-1 and VZV, I postulate that these two residues are not required for the interaction with Cyclin T1.

Next I looked at two viral proteins that have been shown to interact with P-TEFb; the HIV-1 Tat and HSV-1 VP16 viral proteins (Cun et al., 2009;Tahirov et al., 2010;Guo et al., 2012;Gu et al., 2014). When aligned against the ICP22 of HSV-1 and HSV-2, all four viral proteins appear to share a common motif named here as Y-R-Q-R motif. The first two residues (Y-R) of the motif are absent in the SCR-5/11. It is not known whether Tyr-193 and Arg-195 are responsible for the interaction with Cyclin T1. The published interaction between VP16 and CDK9 was analysed by using the full-length VP16 whereby the cellular proteins from CHO-K1 transiently expressing full-length VP16 or the VP16 from HSV-1 infected HEp-2 cell were co-immunoprecipitated and then immunoblotted for CDK9 (Guo et al., 2012). To date, there is no indication of which residues of VP16 are responsible for the interaction with P-TEFb. However, a recent study suggests that 'the HIV-1 Tat Arginine-Rich Motif (ARM) region is N-terminally anchored by Tyr-47 of HIV-1 Tat, that binds to a pocket on the Cyclin T1 surface and interacts with Cyclin T1 residues Leu-44, Asp-47, and Cys-111' (Schulze-Gahmen et al., 2016). This piece of evidence is very intriguing, as the Tat Tyr-47 mentioned corresponds to the first residues of the Y-R-Q-R motif within the HIV-1 Tat amino acid sequence. The Tyr-47 of HIV-1 Tat within the Y-R-Q-R motif is equivalent to Tyr-193 of HSV-1 ICP22, Tyr-190 of HSV-2 ICP22 and Tyr-182 of HSV-1 VP16. Indeed, the three residues of Cyclin T1 (Leu-44, Asp-47, and Cys-111) (Schulze-Gahmen et al., 2016) that serve as binding sites for HIV-Tat are within the expressed Flag-Cyclin T1 residues 1-266 used in this study. It is possible that the same binding sites of Cyclin T1 are required for the binding with the Tyr-193 of the Y-R-Q-R motif within the 193-YMRQCI-197 region of ICP22. However, this needs to be validated with protein-protein interaction assay.

Taken together, the minimal domain of ICP22 that facilitates interaction with CDK9 can be refined to 198 to 245 residues of ICP22, while residues 193 to 197 of ICP22 that contains the residues of the so-called Y-R-Q-R motif are possibly responsible for the interaction with Cyclin T1. Whether the residues of 246 to 256 of ICP22 that are absent in the N-terminus of His-SCR-5/11 are required for binding to Cyclin T1 is not clear.

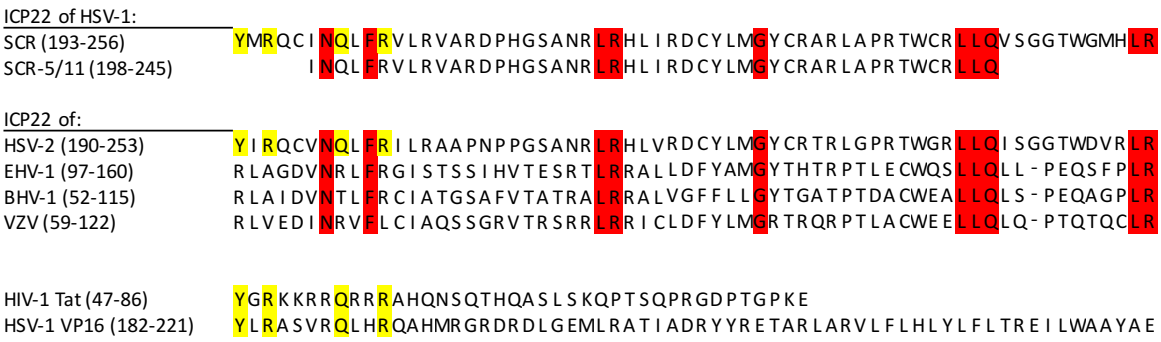


Figure 4.7. Sequence alignment of core conserved region of ICP22 of HSV-1 and other herpesviruses, and HIV-1 Tat and HSV-1 VP16 proteins. The number denotes the position of the amino acid residues is indicated in bracket. Residues within ICP22 marked in red are highly conserved among alphaherpesviruses (Zaborowska et al., 2014). Residues marked in yellow, named as Y-R-Q-R motif are found in ICP22 of HSV-1 and HSV-2, and HIV-1 Tat and HSV-1 VP16 viral proteins.

4.5 Sequence description and secondary structure analysis of ICP22

The structure of ICP22 is yet to be determined but its possible structure can be predicted computationally. The structural prediction tools may be useful as a first step to understand structure-function of ICP22. The ICP22 protein sequence (CAA32287) was used to predict the secondary structure of ICP22 using several open access tools available on the World Wide Web. SMART and RaptorX predict ICP22 to be highly hydrophilic, with 73% of the residues exposed to the cellular environment, while the remaining 27% form a hydrophobic core (Figure 4.8A). This hydrophobic core is probably required for protein folding and stability. The N-terminal domain of ICP22 is a disordered domain while the core conserved region (residues 186-276), and C-terminal acidic (residues 277-347) and basic domains (residues 348-420) are non-disordered or structured (Figure 4.8A). The unstructured region of ICP22 has several proline residues; the proline residues are often found in the N-terminal part of proteins which can give rise to low complexity and unstructured domains. Further analysis using RaptorX on domain 2 (residues 186-276) of the core conserved sequences suggests that the domain can be divided into four alpha-helixes (Figure 4.8B). Domain 3 (residues 348-420) is predicted to contain alpha-helixes without a turn that appear to look like a long arm, whereas the C-terminal domain residues of 348-420 were suggested to adopt either α-helixes (RaptorX) or β-sheets (PredictProtein). Several templates were scanned by the SMART and RaptorX to predict the secondary structure of the core conserved region of ICP22. As indicated in Figure 4.8C, the

conserved region of the ICP22 (residues 199-233) appears to adopt a helix-turn-helix structure like VP16 residues 182-221. Full-length VP16 was used to show interaction with CDK9, and VP16 overcomes the inhibition that ICP22 imposes on CDK9 (Guo et al., 2012). In vitro binding assays using GST-VP16 incubated with purified ICP22 showed that these proteins do not bind each other (Guo et al., 2012). Based on this study and to test whether the region of VP16 that has similar sequence to the SCR of ICP22 can interact with CDK9, I decided to perform His pull-down with HeLa cell extract using VP16 residues 177-224 (vCon) alongside the SCR of ICP22 and IpgC of *S. flexneri* as negative control.

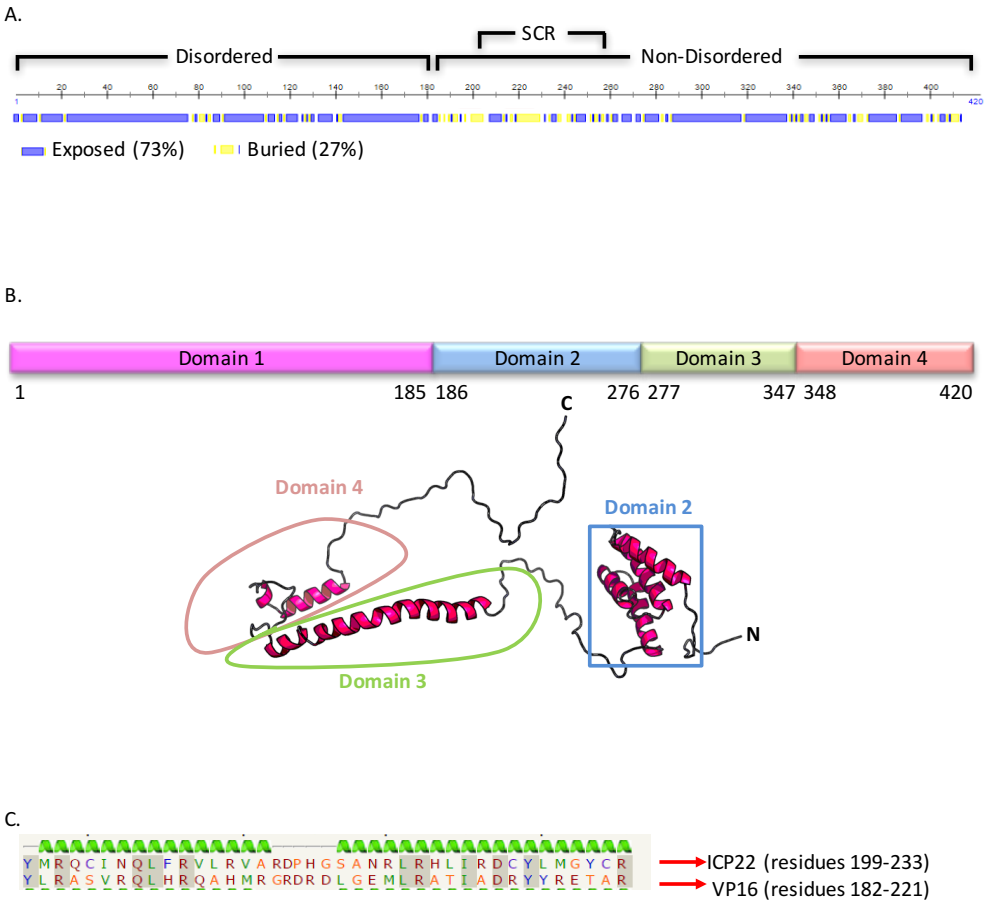


Figure 4.8. Secondary structure analysis of ICP22. **A)** Structure prediction tools SMART and RaptorX indicate that ICP22 is highly hydrophilic with 73% of the residues exposed to cellular environment while the remaining 27% forms the hydrophobic core. The N-terminal region of ICP22 is predicted disordered or unstructured while the C-terminal region of ICP22 where the short conserved region (SCR) is located, is predicted non-disordered (possibly structured). **B)** Based on the amino acid identity and sequence conservation, the ICP22 is divided into four distinct domains (Schwyzer et al., 1994). The RaptorX prediction indicates that residues within the second domain (blue box) have secondary structure that can be divided into four α-helixes. **C)** The residues 199-233 of ICP22 as predicted by Phyre2, are two α-helixes and it roughly resembles the helices observed in the core conserved region of VP16 (residues 182-221).

4.6 Expression and purification of His-tag VP16 recombinant protein using bacterial expression system

In order to determine whether the core conserved region of VP16 is involved in binding to CDK9, a cDNA encoding VP16 residues 177-224 (His-vCon) was cloned into the pET28a plasmid. His-vCon was expressed and purified from Rosetta^{TM} *E.coli* strain using nickel-coated beads and the expressed His-vCon was analysed by SDS-PAGE, followed by Coomassie blue staining and immunoblot (Figure 4.9). The purified His-vCon, alongside His-SCR of ICP22 and His-IpgC of *S. flexneri* as negative control were tested for interaction with CDK9 in a pull-down assay with HeLa cell extract.

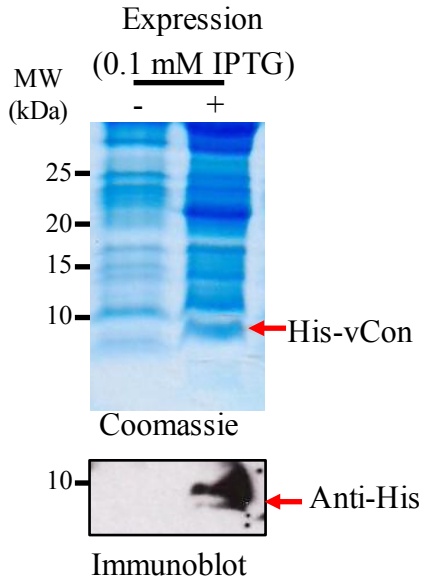


Figure 4.9. Identification of expressed VP16 residues 177-224 (His-vCon). Coomassie stain (top) and immunoblot (bottom) of expressed His-vCon in the presence of 0.1 mM IPTG at 30˚C for 7 h. The red arrows point to the His-vCon at approximately ~9 kDa.

4.7 ICP22 and VP16 interact with the active form of P-TEFb

ICP22 has been shown in the literature and here, to bind to P-TEFb (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014). To further validate the interaction of ICP22 and P-TEFb demonstrated by the pull-down using the recombinant proteins ICP22, CDK9 and Cyclin T1, and to examine whether the purified VP16 can bind to cellular P-TEFb, I tested the binding ability of purified His-SCR of ICP22 and His-vCon of VP16 to bind P-TEFb in HeLa cell extract. It should be noted that the binding buffer used for the pull-down with His-IpgC, His-SCR of ICP22 and His-

vCon of VP16 here contains higher salt content (125 mM NaCl), as compared to the pull-down samples for mass spectrometry (75 mM NaCl) (Chapter 3). The choice of stringent binding buffer was decided because of the presence of 129 proteins unique to His-SCR (Figure 3.1, Chapter 3) that were not pulled down by its full-length counterpart. It is uncertain whether those 129 proteins are true His-SCR interactors, but at this point they are considered as non-specific binding proteins. Thus incubation of bait proteins with HeLa cell extract in a more stringent binding buffer could prevent the binding of non-specific proteins.

The purified His-IpgC, His-SCR of ICP22 and His-vCon of VP16 were incubated with HeLa cell extract, and after extensive washing, the bound complexes were separated by SDS-PAGE and then analysed by immunoblots using anti-His antibody and antibodies against subunits of P-TEFb. As shown in Figure 4.10, the three protein baits are visibly detected using the anti-His antibody. The prey protein CDK9 was probed with two available CDK9 antibodies (H169 and C20), and an antibody recognising CDK9 phosphorylated at Thr-186. Phosphorylation of CDK9 at the Thr-186 is required for the P-TEFb activation to promote Pol II pause-release (Li et al., 2005;Guo and Price 2013). The three different CDK9 antibodies detect CDK9 after pull down of His-SCR and His-vCon. Moreover, the His-SCR and His-vCon interact with the cyclin partner of CDK9, Cyclin T1. It should be noted that the faint bands of CDK9 C20 and Cyclin T1 observed in the lanes of negative control His-IpgC are considered as background. The intensity of the probed prey proteins in the His-SCR and His-vCon are above background indicating that the CDK9 and Cyclin T1 interact with both baits. The interaction between the bait proteins and other cellular proteins, in this instance the H3 and α-tubulin, was also tested and resulted in none of these cellular proteins binding to the bait proteins. From this assay, it can be concluded that the purified His-tag ICP22 residues 193-256 interacts with cellular P-TEFb, and this observed interaction agrees with the proteomic analysis (Chapter 3) and previous study using recombinant CDK9 and Cyclin T1 and also previously published interaction with the GST-ICP22 residues 193-256 (Zaborowska et al., 2014). Interestingly, the purified VP16 residues 177-224 pulled down the cellular P-TEFb. This study for the first time provides evidence that residues 177-224 of VP16 are sufficient for the interaction with cellular P-TEFb.

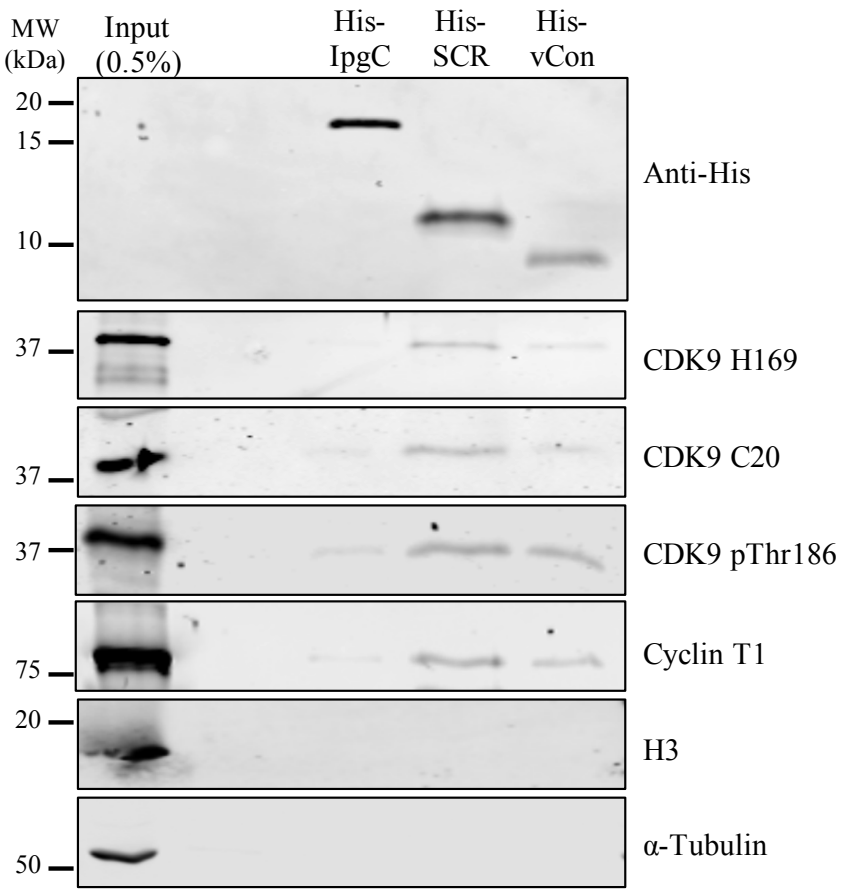


Figure 4.10. ICP22 and VP16 interact with P-TEFb. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against anti-His antibody for purified bait proteins and antibodies recognising subunits of P-TEFb, H3 and α-tubulin.

4.8 ICP22 but not VP16 interacts with the inactive form of P-TEFb

P-TEFb co-exists in both active and inactive forms in mammalian cells, and 70-80% of cellular P-TEFb is catalytically inactive (Krueger et al., 2010). The inactive P-TEFb is sequestered by the inhibitory 7SK small nuclear ribonucleoprotein (7SK snRNP) complex composed of the 7SK small nuclear RNA (snRNA), MEPCE, LARP7 and Hexim1 (Nguyen et al., 2001;AJ et al., 2016). Interactions between ICP22 and the inactive form of P-TEFb have not yet been reported. Using His pull-down with HeLa cell extract, I addressed the question of whether ICP22 and VP16 can interact with the inactive form of P-TEFb by probing for subunits of 7SK snRNA using antibodies against Hexim1, MEPCE, and LARP7.

When blotting for Hexim1, a faint band at the size corresponding to Hexim1 was detected with His-IpgC and this was regarded as background. Interestingly, the His-

SCR pulled-down Hexim1, MEPCE and LARP7 (Figure 4.11). I have therefore shown, for the first time, that the SCR of ICP22 interacts with the inactive form of P-TEFb. Interestingly, it could be observed from the immunoblot that His-vCon weakly binds to Hexim1 but does not bind to MEPCE or LARP7. Due to the high background of Hexim1 with the negative control His-IpgC, it was difficult to determine for His-vCon whether Hexim1 was a binding partner. Because P-TEFb is sequestered by 7SK snRNP consisting of Hexim1, MEPCE and LARP7 as a complex, and His-vCon is not able to bind to MEPCE and LARP7, I conclude that VP16 residues 177-224 do not interact with the inactive form of P-TEFb.

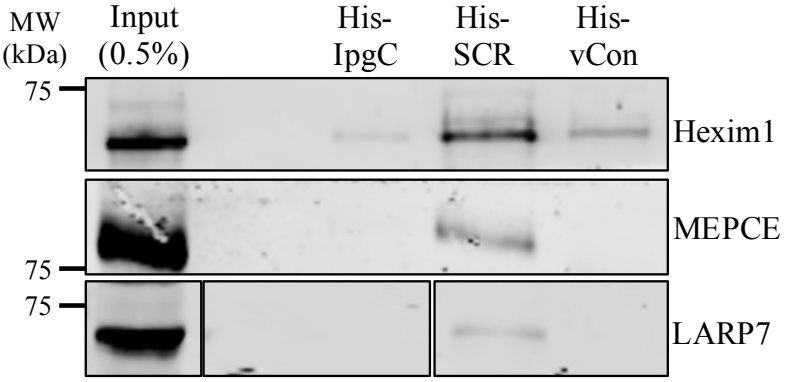


Figure 4.11. ICP22 but not VP16 interacts with inactive form of P-TEFb. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against 7SK snRNP-associated proteins Hexim1, MEPCE and LARP7.

4.9 ICP22 and VP16 interact with the P-TEFb-associated complex PAF1c

P-TEFb interacts with several cellular complexes such as the SEC and the polymerase-associated factor 1 complex (PAF1c), which consists of several subunits (PAF1, Leo1, CDC73/Parafibromin, Ctr9, Rtf1 and WDR61/Ski8) (Zhang et al., 2009;Chen et al., 2015). PAF1c is a critical regulator of promoter–proximally paused Pol II. PAF1c can either enhance or repress elongation efficiency by regulating Pol II pause-release (Chu et al., 2013;Chen et al., 2015;Yu et al., 2015). Because the PAF1c complex is involved in Pol II pause-release, and is associated with P-TEFb, I decided to probe for association of the PAF1c complex components PAF1, Leo1 and CDC73 with ICP22, using His pull-down with HeLa cell extract. As shown in Figure 4.12, both ICP22 and VP16 are associated with subunits of PAF1c. This may not be a

surprise due to the evidence shown earlier that ICP22 as well as VP16 strongly interacts with P-TEFb, and thus possibly indirectly pull-down PAF1c. At this stage, based on the positive interaction with PAF1c that is associated with elongating Pol II, it can be concluded that ICP22 interacts with P-TEFb involved in Pol II pause-release into the gene and ICP22 therefore controls productive elongation. To investigate this further, my next step was to probe for the substrates of P-TEFb; Pol II and the SPT5 subunit of the transcription elongation factor DSIF.

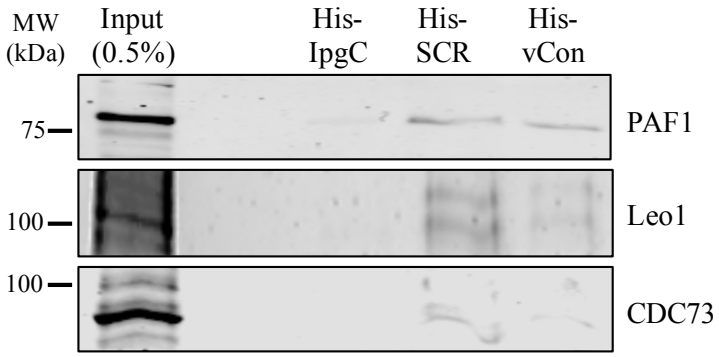


Figure 4.12. Both ICP22 and VP16 interact with subunits of PAF1c. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against PAF1, Leo1 and CDC73.

4.10 ICP22 and VP16 associate with transcriptionally-engaged Pol II

It has been widely accepted that P-TEFb phosphorylation of Ser2 of Pol II CTD, and two negative elongation factors that are recruited to Pol II (the SPT5 subunit of DSIF and the NELF-E of NELF) is a prerequisite for the release of promoter–proximally paused Pol II. Once Pol II negotiates promoter–proximal pausing, the phosphorylated NELF-E is released from Pol II while the phosphorylated SPT5 is switched from being a negative factor to a positive factor and remains bound to elongating Pol II. As P-TEFb substrates, the Pol II CTD, the C-terminal region (CTR) of SPT5 (Yamada et al., 2006), and NELF-E are associated with P-TEFb.

To answer the question of whether ICP22 interacts with P-TEFb substrates, I have probed for Pol II and SPT5 interaction with ICP22 and VP16 using His pull-down with HeLa cell extract. As shown in Figure 4.13, neither ICP22 nor VP16 pull down Pol II, but both interact with SPT5. The negative interaction between the His-SCR and Pol II are in line with the previous published observation where full-length ICP22

has been shown to interact with Pol II but not the ICP22 residues 193-256 (Zaborowska et al., 2014).

The N-terminal region of VP16 known as the transactivation domain (TAD) interacts with the MED25 subunit of the Mediator complex, implying an indirect association with the Pol II machinery (Vojnic et al., 2011). The VP16 residues 177-224 used in this study are located within the middle region of VP16. The result indicates that His-vCon or the VP16 residues 177-224 are not interacting, whether directly or indirectly, with Pol II.

As for the P-TEFb substrate SPT5, it was observed that both ICP22 and VP16 pulled down SPT5. It is not known whether this interaction is direct, but it is probable that His-SCR directly interacts with P-TEFb that is bound to its substrate SPT5. Meanwhile, VP16 residues 177-224 also showed interaction with SPT5. While it is not known whether VP16 directly interacts with SPT5, it is possible that the interaction observed here is indirect, because VP16 residues 177-224 are directly interacting with P-TEFb that is bound to its substrate SPT5.

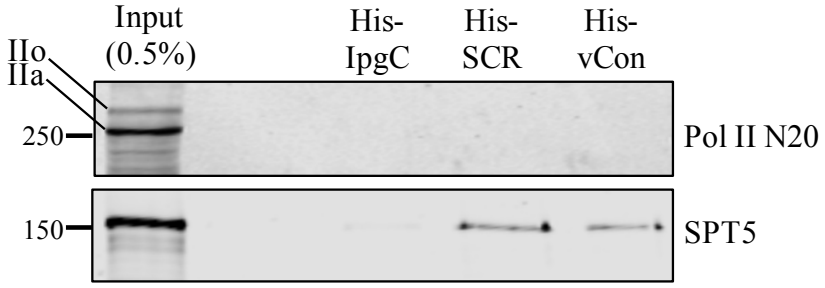


Figure 4.13. ICP22 and VP16 interact with SPT5 but not with Pol II. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against P-TEFb substrates Pol II and SPT5.

4.11 The interaction of ICP22 with factors recruiting P-TEFb to the promoter

Upon activation, P-TEFb is released from the 7SK snRNP and recruited to the promoter. The recruitment of active P-TEFb is a critical step in controlling gene expression, and is regulated at multiple stages. There are a variety of ways for P-TEFb to be recruited to paused Pol II, leading to Pol II pause-release into the gene. Several studies have indicated that many factors are able to recruit P-TEFb to the promoter-proximal region, including the transcriptional activators c-myc and the RelA subunit of NF-κB (Kwak and Lis 2013;Lu et al., 2016). The co-activator Brd4 and Super

Elongation Complex (SEC) are two factors or complexes that are associated with the recruitment of P-TEFb. Brd4 recognises acetylated histone tails, and can bind to Cyclin T1 once P-TEFb is released from the 7SK snRNP. The SEC is a multi-protein complex that mediates the regulation of Pol II pause-release by recruiting elongation factors such as P-TEFb and PAF1c. P-TEFb can also be recruited through the Mediator complex notably by MED26 (Takahashi et al., 2011).

Because there are many factors involved in recruiting P-TEFb to the promoter, using the His pull-down with HeLa cell extract, I decided to probe for MED1 subunit of Mediator complex and Tat-specific factor 1 (Tat-SF1). These two factors were present in the ICP22 interactome (Chapter 3). MED1 and Tat-SF1 have been implicated in the recruitment of Brd4/P-TEFb complex (Lu et al., 2016). As shown in Figure 4.14, neither ICP22 nor VP16 pulled down MED1 or Tat-SF1. This is in contrast with the ICP22 pull down analysed by mass spectrometry. It is possible that the stringent binding buffer used for this pull-down prevented the binding of MED1 and Tat-SF1 either directly to the ICP22, or indirectly by interacting with P-TEFb. Using less stringent binding buffer and having a positive control that does pull down MED1 and Tat-SF1 could further demonstrate their interactions with ICP22.

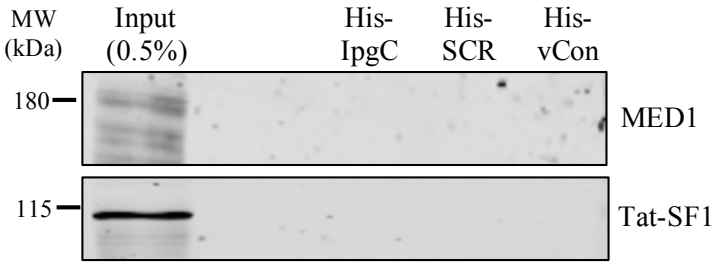


Figure 4.14. ICP22 and VP16 does not interact with MED1 and Tat-SF1. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against factors implicated in P-TEFb recruitment to the promoter MED1 and Tat-SF1.

4.12 ICP22 interacts with kinases other than P-TEFb

Up to this stage, the data show that ICP22 interacts with P-TEFb, and factors associated with P-TEFb such as 7SK-associated proteins and subunits of PAF1c and SPT5. VP16 also pulls down subunits of PAF1c and SPT5. The evidence presented so far suggests that the His pull-down assay with HeLa cell extract used in this study is a

good system to study the protein-protein interaction network. Therefore, using the same bound proteins, I extended the investigation to study the interaction of ICP22 and VP16 with transcriptional kinases other than P-TEFb (Figure 4.15).

The interaction between ICP22 and VP16 with CDK12 was first examined and demonstrated (Figure 4.15A). Two anti-CDK12 antibodies were used, one from Sigma-Aldrich (Sigma) and one from Bethyl Laboratories (Bethyl). Anti-CDK12 Sigma cross-reacted with His-IpgC, however this faint band here is regarded as background. His-SCR consistently pulled down CDK12 regardless of which antibody was used. Thus it can be concluded that His-SCR interacts with CDK12. Meanwhile, His-vCon showed a faint band when probed with the Sigma anti-CDK12 antibody, the same anti-CDK12 antibody that presented a faint band in the negative control His-IpgC, but no band appeared when probed with Bethyl anti-CDK12 antibody. I therefore conclude that His-vCon does not interact with CDK12, although this observation requires further validation.

As well as CDK12, His-SCR is able to interact with CDK11 (Figure 4.15B) and CDK8 (Figure 4.15C). His-vCon possibly interacted with CDK11 as shown by a faint band that appeared with His-vCon (Figure 4.15B). As revealed by the immunoblot on Figure 4.15D, none of the protein baits interacted with transcriptional kinases CDK13, CDK10, and CDK7. The lack of interaction between ICP22 and CDK7 has been shown before (Durand et al., 2005;Durand and Roizman 2008). CDK13 is present in the ICP22 interactome, which contradicts the observation in the immunoblot. This could be attributed to the stringent binding buffer used in this pull-down that prevents the binding of CDK13 to the ICP22.

In conclusion, ICP22 residues 193-256 (His-SCR) are able to interact with transcriptional kinases other than CDK9. Whether the interaction is direct or indirect remains to be determined. These transcriptional kinases are CDK12, CDK11 and CDK8, as shown by the immunoblots of pull-down samples carried out using a stringent binding buffer. The VP16 residues 177-224 (His-vCon) is able to interact with CDK9 and possibly CDK11, however due the poor signal for CDK11 this result requires further validation.

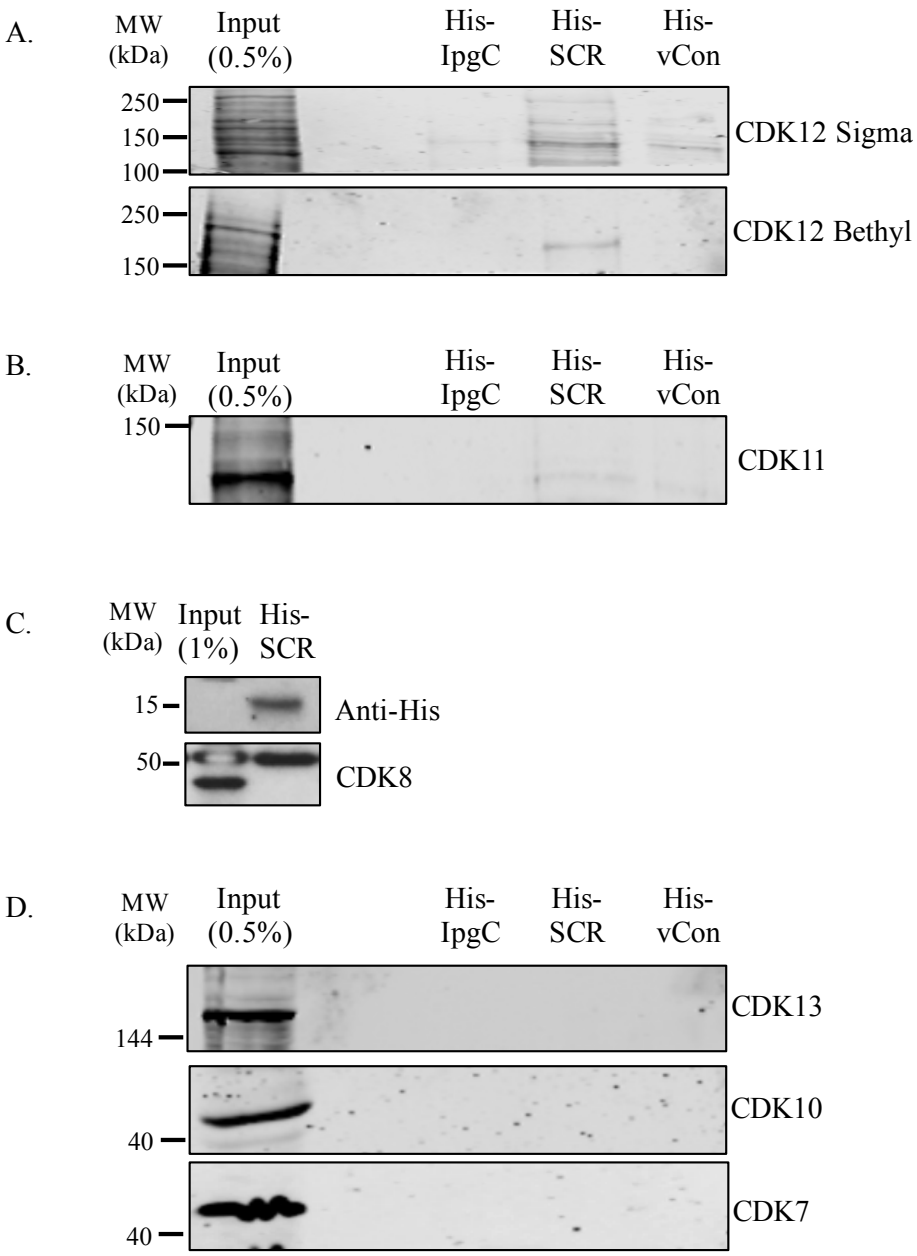


Figure 4.15. ICP22 and VP16 are able to interact with transcriptional kinases other than CDK9. Purified ICP22 residues 193-256 (His-SCR) and VP16 residues 177-224 (His-vCon) were incubated with HeLa cell extract. Input and bound samples were analysed by SDS-PAGE, followed by immunoblot against transcriptional kinases (A) CDK12, (B) CDK11, (C) CDK8 and (D) CDK13, CDK10, and CDK7.

4.13 Chapter summary

The interaction between ICP22 and P-TEFb was first shown by Roizman's group (Durand et al., 2005), followed by Li's group (Guo et al., 2012) where both groups were using lysates from HSV-1 infected human cells. More recently, our laboratory showed that recombinant ICP22 interacts with P-TEFb using a GST-ICP22 pull-down

assay with HeLa cell extract (Zaborowska et al., 2014). This interaction has been confirmed and extended in this thesis using recombinant ICP22, CDK9 and Cyclin T1 produced in bacterial expression systems. The interaction between ICP22 and P-TEFb was also seen in a cellular context, by expressing Flag-tag full-length ICP22 transiently in HEK293 cells and CDK9 was co-immunoprecipitated with anti-Flag M2 beads for the analysis by immunoblot (Appendix 12).

In this chapter I have validated that ICP22 interacts with both subunits of P-TEFb *in vitro*. More specifically, six residues of ICP22 (193 to 197) interact with Cyclin T1 (1 to 266) while the next 48 residues of ICP22 (198 to 245) interact with CDK9. It is possible that the Tyr-193 residue within the Y-R-Q-R motif of the SCR is responsible for the interaction with Cyclin T1. *In vitro*, the interaction with Cyclin T1 appears to be stronger than the interaction with CDK9 but contacts with both partners may be necessary for a strong *in vivo* interaction.

This chapter also documented the interaction of ICP22 with a range of P-TEFb interacting partners such as the 7SK-associated proteins that are part of the inactive form of P-TEFb, PAF1c, which is associated with P-TEFb at the proximal-promoter and productive elongation, and the SPT5 subunit of DSIF, which is a substrate of P-TEFb. Since PAF1c has roles in P-TEFb recruitment and is associated with transcribing Pol II, this observation might suggest ICP22 interacts with active P-TEFb on host genes. Specifically, ICP22 interferes with the role of P-TEFb in Pol II pause-release and therefore controls productive elongation.

VP16 does not interact with the inactive form of P-TEFb, but with the active form of P-TEFb, and with the P-TEFb-associated complex PAF1c and the SPT5 subunit of DSIF. I have shown here, for the first time, that VP16 residues 177-224 interact with PAF1c and SPT5. There is no evidence that VP16 directly interacts with PAF1c or SPT5, and it is likely that the interaction observed here is indirect.

The results from the pull-down followed by immunoblot indicate that neither bait protein ICP22 nor VP16 was associated with MED1 or Tat-SF1. These data contrast with the ICP22 interactome data, perhaps because the stringent binding buffer used prevented interaction of MED1 and Tat-SF1 with P-TEFb to a level sufficient to be detected in ICP22 pull-down.

Other transcriptional kinases pulled down by ICP22 (residues 193-256) in this assay were CDK12, CDK11 and CDK8. The lack of interaction between ICP22 and CDK7 from HeLa cell extract is in agreement with the findings using ICP22 from lysates of

HSV-1 infected human epithelial type 2 cells (HEp-2) that demonstrated that ICP22 binds to GST-CDK9, but not GST-CDK7 (Durand et al., 2005;Durand and Roizman 2008). CDK13 interacts with ICP22 as shown by the proteomic analysis (Chapter 3), but failed to show an interaction in the validation study with the immunoblot, perhaps due to the difference of the stringency of the binding buffer used. The sequence alignment of CDK12, CDK11 and CDK8 versus ICP22 non-binding transcriptional kinases may provide a clue to which residues or binding sites that are critical to form a complex with ICP22.

Chapter 5| Discussion

5.1 The effect of ICP22 on elongating RNA polymerase II (Pol II) (Chapter 2)

HSV-1 is an important human pathogen and remains a major cause of severe morbidity and mortality. The molecular biology of HSV, particularly the ability to switch between lytic and latent infection, reactivation and the mechanisms developed to evade the immune response are still poorly understood and requires intensive studies. Efforts to understand the biology of HSV infection with a view to preventing latency and neuro-invasion could provide opportunities for new treatment approaches. Many studies in the past that aimed to understand the mechanism of transcriptional regulation of viral and cellular genes during HSV-1 infection used HSV-1 infected cells. However, the *in vivo* system is unable to characterise the effect of a single viral protein. In order to address this issue, I have investigated the effect of ICP22 on Pol II-mediated transcription of host cell genes by using transiently-expressed ICP22. Therefore, any effects noted in these experiments can be attributed to ICP22 alone.

To explore Pol II-mediated transcription of cellular genes, I have used the protein-coding model gene *KPNB1* (Figure 2.1). The effect of CDK9 inhibitors on Pol II associated with *KPNB1* has already been shown in HeLa cells (Appendix 1)(Laitem et al., 2015), and this effect is comparable in HEK293 cells (Figure 2.2). In CDK9 inhibitor-treated cells, the profile of Pol II appears to be affected at both ends of the gene. After treatment with CDK9 inhibitors, the levels of Pol II are elevated at the TSS, followed by a drop just downstream, suggesting an increase in Pol II pausing, and that the newly-initiated Pol II is unable to proceed into productive elongation. Pol II that had entered productive elongation before the CDK9 inhibitors took effect remains elongating along the gene. A decreased level of Pol II is observed downstream of the poly(A) site, suggesting that Pol II has prematurely terminated at the end of *KPNB1*.

I postulated that in regions where Pol II is absent, there is an increase of nucleosome occupancy. To test this hypothesis, I performed ChIP-qPCR of H3 and H2B in HEK293 cells treated with the CDK9 inhibitor (Figure 2.3). Here, an increased level of H3 and H2B was shown near the EEC where Pol II is unable to make the transition into productive elongation and downstream of poly(A)-associated checkpoint where Pol II has prematurely terminated. This could suggest that when CDK9 is inhibited,

nucleosome occupancy is increased in regions that are no longer transcribed, or that nucleosome occupancy itself could play a role in prematurely terminating Pol II.

Based on the effect of CDK9 inhibition using small molecule inhibitors, I set out to analyse the effect of ectopic expression of ICP22 on the Pol II level on *KPNB1*. The expression of the short conserved region (SCR) of ICP22 (residues 193-256) induces loss of Ser2P Pol II CTD (Figure 2.4), as previously noted (Durand et al., 2005;Fraser and Rice 2007;Zaborowska et al., 2014). The profile of Pol II on *KPNB1* after 7 h of ectopic expression of Myc-SCR in HEK293 (Figure 2.5) resembles that after treatment of cells with CDK9 inhibitors (Figure 2.2) (Laitem et al., 2015). Ectopic expression of Myc-SCR affects the transition of Pol II past the EEC and the level of transcribing Pol II at the end of *KPNB1*, in a similar fashion to CDK9 inhibitors (Figure 2.2) (Laitem et al., 2015). Thus, ICP22 has the potential to de-regulate the transition of Pol II past the EEC into productive elongation and the elongating Pol II past the poly(A)-associated checkpoint. The level of Pol II seems unaffected 6 h after ectopic expression of Myc-SCR (Figure 2.6; top panel) while 8 h after ectopic expression of Myc-SCR (Figure 2.6; top panel) there is significant pausing of Pol II at the TSS and loss of Pol II in the gene body.

The role of CDK9 in regulating efficient transcription elongation has been well characterised. Using small molecule CDK9 inhibitors is one approach to inhibit its kinase activity. However the target specificity of CDK9 inhibitors is arguable and they may be inhibiting other transcriptional kinases. Treatment of cells with CDK9 inhibitors cause a reduction of Pol II CTD Ser2P while Ser5P was less affected (Laitem et al., 2015). The ectopic expression of ICP22 also causes a reduction of Ser2P while Ser5P was unaffected (Zaborowska et al., 2014). Here I show that the ectopic expression of SCR of ICP22 causes a reduction of Ser2P. Although ICP22 affects global CTD modification, the effect of the ectopic expression of ICP22 does not affect the expression of the model gene *KPNB1* at the protein level and its mRNA or protein stability (Figure 2.7), most likely due to the short treatment times.

Overall, the results demonstrate that ICP22 affects the level of transcribing Pol II on the model protein-coding gene *KPNB1*, in a mechanism similar to small molecule CDK9 inhibitors (Laitem et al., 2015).

5.1.1 Future directions

From the results presented in Chapter 2, it was postulated that ICP22 affects the level

of transcribing Pol II at the beginning and end of the model gene, *KPNB1*, in a similar way to the CDK9 inhibitors. ICP22 has been shown to be recruited to close to the TSS of cellular genes (Zaborowska et al., 2014), but whether ICP22 can be recruited to the end of cellular genes is not known. The next step would be to perform ChIP-qPCR of Flag-ICP22 with anti-Flag antibody (since anti-Myc recognises endogenous c-myc) to analyse whether ICP22 is recruited to both ends of the *KPNB1* gene. Furthermore, ChIP-qPCR of CTD marks, particularly anti-Ser2P Pol II CTD on the *KPNB1* gene is necessary to confirm the inhibitory effect of ICP22 on P-TEFb. The effect of CDK9 inhibitors has been shown to increase nucleosome occupancy, but the histone marks were not analysed. It would therefore be interesting to analyse the histone marks in the Pol II-free regions following cell treatments with CDK9 inhibitors and ectopic expression of ICP22.

Although the results in Chapter 2 were obtained using transfected cells and may not exactly reflect the activity of ICP22 in HSV-1 infected cells, the effect of ectopic expression of ICP22 on the level of transcribing Pol II on *KPNB1*, suggests that ICP22 plays an important role in the transcriptional regulation of cellular genes in infected cells. To extend our understanding of ICP22 inhibition on host genes beyond a representative host gene model, a genome-wide analysis should be carried out to precisely map whether changes in the pattern of transcription and Pol II modification following CDK9 inhibition with ICP22 are universally observed across different classes of genes or are specific to a distinct subset of genes. Whether similar effects would be observed in infected cells is not clear. It will be necessary to study the effect of ICP22 in the context of HSV-1 infection in the future. Nonetheless, genome-wide analyses using transient expression of ICP22 may be biologically relevant to HSV-1 infected cells, as in early HSV-1 infection, ICP22 inhibits transcription of host cell genes in order to favour transcription of viral genes. It is not known which host cell genes are downregulated by ICP22. The role of host Pol II in transcription of viral genes is substantial during the earlier stage of HSV-1 infection. Since HSV-1 has evolved strategies to evade the antiviral host interferon response by blocking the activation of type I interferons or the expression of IFN-stimulated genes (Kawai and Akira 2006;Mostafa and Davido 2013), it is likely that one subset of genes that are suppressed are the IFN-inducible genes.

5.2 Characterisation of the interaction proteome of ICP22 (Chapter 3)

Ectopic expression of ICP22 affects the level of transcribing Pol II and alters the phosphorylation of Ser2 Pol II CTD. The phosphorylation pattern of Pol II CTD is largely mediated by transcriptional kinases, or CTD kinases. Pol II is committed to productive elongation upon phosphorylation of Ser2, primarily by the CDK9 subunit of P-TEFb. It is possible that transcriptional kinases additional to P-TEFb can act as CTD elongation-phase kinases. For example CDK12 is also a Ser2 CTD kinase (Bartkowiak et al., 2010).

An interesting aspect of the molecular virology of HSV is that ICP22 is multifunctional and able to perform diverse functions to provide a favourable environment for viral replication (Rice et al., 1995;Poon et al., 2000;Orlando et al., 2006;Mostafa and Davido 2013). Based on the evidence that ICP22 interacts with CDK9 (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014), and because CDK9 shares highly conserved kinase domains with other transcriptional kinases (Figure 1.2; Chapter 1), ICP22 could interact with transcriptional kinases other than CDK9.

Therefore Chapter 3 focused on the ICP22 interactome. Purified recombinant ICP22 (full length and SCR) were used to pull down interacting proteins from HeLa cell extract and bound proteins were analysed with mass spectrometry (Figure 3.1). The common hits between both full length and SCR of ICP22 were analysed for their molecular functions, as classified by GO terms. A total of 23 molecular functions were associated with ICP22-bound cellular proteins, and mostly are related to transcriptional activities (Table 3.1 and Figure 3.2). Because CDK9 is an important factor in transcription and has been shown to interact with ICP22, the lists of hits was searched for CDK9 (Table 3.2). CDK9 is pulled down by ICP22, which confirms the previously published interaction (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014). Interestingly, ICP22 pulled down several other transcriptional kinases, cyclins and other P-TEFb-associated proteins (Table 3.2). When the lists of ICP22-bound cellular hits are compared with the transcriptional kinases proteome from BioGrid (a protein-protein interaction database), known interactors of the respective transcriptional kinases are present in the list of proteins pulled down by ICP22 (Figure 3.3). This result not only confirms the documented interaction between ICP22 and CDK9, but also indicates that ICP22 interacts with other transcriptional kinases.

Proteomic analysis also identifies cellular proteins that are directly or indirectly

interacting with ICP22 and this analysis can therefore provide new information about the function of ICP22. The interaction of ICP22 and P-TEFb, a vital factor in Pol II transcription, further validates the role of ICP22 in Pol II-mediated gene transcription. ICP22 could have opposite effect on host and viral gene expression during HSV-1 infection. While the function of ICP22 is associated with promoting the expression of HSV-1 late viral genes, ICP22 is thought to act as a transcriptional repressor on both host and viral genes (Bowman et al., 2009;Guo et al., 2012). The identification of ICP22-interacting cellular proteins may help to explain the differential role of ICP22 in viral and host gene transcription.

Given the success of the strategy used in this study to pull down the known ICP22 binding protein P-TEFb, together with P-TEFb-associated factors, and novel factors such as CDK11B, CDK12, CDK13 and CDK8/CDK19, this study might provide a clue to the role of ICP22 that will have broader implications for the outcome of HSV-1 infection. ICP22 appears to be a sophisticated viral protein with the ability to interact with several transcriptional kinases. These interactions may all result in the inhibition of the kinase activity, resulting in the suppression of host gene transcription, possibly suppressing the IFN-inducible genes and/or promoting viral gene transcription. ICP22 therefore appears to play a major role in regulating transcription of both cellular and viral genes. The effect of ICP22 on the activity of transcriptional kinases in addition to P-TEFb is therefore worth further investigation. The ICP22 interactome presented here appears to be convincing, as it gives many relevant hits such as CDK9, P-TEFb-associated complexes, and other transcriptional kinases as well as their known interacting proteins.

5.2.1 Future directions

The ICP22 proteomics can be improved in many ways. First, whilst the small, unstructured His-tag chosen for the recombinant ICP22 is favourable because small tags least interfere with the function of recombinant proteins, the nickel-coated beads that are used to capture His-tagged recombinant proteins can interact non-specifically with cellular proteins. The use of different protein tags, for example GST or HA, could further validate my findings. The HeLa cell extract used for pull down lacks p53, which is a known ICP22 interacting partner (Table 1.3; Chapter 1) (Maruzuru et al., 2013). Using different cellular extracts for ICP22 pull-down perhaps could identify different sets of interactors and hence the network of ICP22 protein-protein

interaction could be mapped better.

Based on the list of relevant hits from proteomics data and validation with immunoblotting, I conclude that ICP22 interacts directly with CDK9 (or other transcriptional kinases), that are part of larger complexes. Thus, ICP22 is probably interacting indirectly with P-TEFb-interacting complexes. To test this, the HeLa cell extract could be incubated with anti-CDK9 antibody and the immunoprecipitated CDK9-bound complexes analysed by mass spectrometry or immunoblotting, alongside the ICP22-bound proteins from the pull-down assay. This approach could identify the subset of P-TEFb bound complexes (or other transcriptional kinases) that are potentially targeted by ICP22.

5.3 Characterisation of the interaction between ICP22 and cellular proteins (Chapter 4)

So far, the ICP22 proteomics indicate that ICP22 interacts with CDK9, in agreement with previous results (Durand et al., 2005;Guo et al., 2012;Zaborowska et al., 2014). Chapter 4 first highlights ICP22's interaction with subunits of P-TEFb, and secondly validates the proteomics hits by immunoblotting.

The exact mode of ICP22 binding to P-TEFb, i.e. whether it binds to either or both subunits of P-TEFb, is not clear. The purified recombinant proteins SCR and SCR-5/11 (residues 198-245) of ICP22 (Figure 4.1 and Figure 4.2), and purified recombinant CDK9 and Cyclin T1 (residues 1-266) (Figure 4.4) were prepared for protein-protein interaction studies. The interaction assay suggests that both SCR and SCR-5/11 of ICP22 are able to interact directly with CDK9 and that the region of ICP22 interacting with CDK9 can be refined to residues 198-245 (Figure 4.5). When incubated with Cyclin T1, the SCR interacts with Cyclin T1 but SCR-5/11 does not (Figure 4.6). When both SCR and SCR-5/11 were compared with several proteins (ICP22 homologues and viral proteins known to interact with P-TEFb; HIV-1 Tat and HSV-1 VP16), there are two patterns worth highlighting (Figure 4.7).

First, the highly conserved residues Leu-255 and Arg-256. The VZV homologue of ICP22 does not trigger loss of Ser2P Pol II CTD, and the ability of ICP22 to cause loss of this mark is not conserved in alphaherpesvirus homologues (Fraser and Rice 2007). There are very limited studies on the interaction of the VZV ICP22 homologue with P-TEFb but it may not interact with P-TEFb. With that in mind, given that Leu-255 and Arg-256 are present in ICP22 of both HSV-1 and VZV, I postulate that these

two residues are not required for the interaction with Cyclin T1.

The second pattern is the Y-R-Q-R motif, which is present in ICP22 of HSV-1 and HSV-2, and in two viral proteins that have been shown to interact with P-TEFb, HIV-1 Tat and HSV-1 VP16 (Cun et al., 2009;Tahirov et al., 2010;Guo et al., 2012;Gu et al., 2014). The tyrosine residue of the motif is of particular interest. The Tyr-47 of HIV-1 Tat interacts with Cyclin T1 residues (Schulze-Gahmen et al., 2016), and that residue corresponds to Tyr-193 of SCR and is absent in SCR-5/11. Based on the studies using Tyr-47 of HIV-1 Tat, it is possible that Tyr-193 of HSV-1 ICP22 is required for the interaction with Cyclin T1. The conclusion here is also based on the assumption that ICP22 of VZV does not interact with P-TEFb and that Tyr-193 of ICP22 could interact with Cyclin T1, however these findings will serve as a basis for biochemical analysis in the future.

The secondary structure of ICP22 was predicted (Figure 4.8), and since the protein-protein interaction assays using recombinant ICP22 and subunits of P-TEFb were successful, the final goal is to co-crystallize the ICP22 protein in complex with P-TEFb. Based on the secondary structure prediction, the SCR of ICP22 adopts an alpha-helical structure similar to that of VP16. VP16 interacts with P-TEFb and the ability of VP16 to overcome the inhibition of P-TEFb imposed by ICP22 has been shown (Cun et al., 2009;Guo et al., 2012). Recombinant VP16 (residues 177-224) (vCon) similarly to the SCR of ICP22 (residues 193-256) was then purified (Figure 4.9) to perform pull-down assay with HeLa cell extract alongside the SCR of ICP22. Both vCon of VP16 and SCR of ICP22 pulled down CDK9 and Cyclin T1. The region of VP16 required for the interaction with cellular P-TEFb is therefore further refined to residues 177-224 (Figure 4.10). ICP22 appears to interact with the inactive form of P-TEFb, while VP16 generally does not (Figure 4.11). Both ICP22 and VP16 are associated with the P-TEFb-associated complex PAF1c (Figure 4.12), and P-TEFb substrate SPT5 (Figure 4.13). ICP22 and VP16 does pull down MED1 and Tat-SF1 (Figure 4.14). When probed for transcriptional kinases other than CDK9, ICP22 pulled down CDK12, CDK11 and CDK8 while the interaction of VP16 with these kinases require further validation (Figure 4.15). The ICP22 pull-down here failed to show interaction with CDK7, which corresponds to the previously shown lack of interaction between ICP22 and CDK7 (Durand et al., 2005;Durand and Roizman 2008). The interaction of ICP22 and CDK13 was not detected in the immunoblotting experiment, perhaps due to the stringent binding buffer used in this particular assay.

My strategy to study the ICP22 and VP16 interaction with transcriptional kinases here is to make use of the recombinant proteins expressed and purified from bacterial expression systems. It should be noted that unlike post-translational modifications of recombinant proteins expressed in more complex organisms, the purified ICP22 and VP16 expressed in the bacterial expression systems used here lack post-translational modifications. It has been shown that post-translational modification of ICP22 is not necessary for the interaction with CDK9 (Zaborowska et al., 2014). Likewise, the interaction observed between ICP22 with CDK12, CDK11 and CDK8 shown in this work requires ICP22 with no or very minimal post-translational modifications. In terms of HSV-1 infection, perhaps it could be advantageous to the HSV-1's infection strategy that the viral proteins require few post-translational modifications. In addition, the purified ICP22 and VP16 used in this study for pull-down assays with HeLa cell extract eliminate the direct or indirect influence of other viral gene products or the context of viral-infected cells, unlike the viral proteomics studies conducted in HSV-1 infected cells.

Overall, ICP22 pull-downs from HeLa cell extract indicate that ICP22 interacts with P-TEFb and P-TEFb-associated factors such as PAF1c and the SPT5 subunit of DSIF that participate in productive elongation. In addition, ICP22 may have multiple targets, as indicated by its interaction with transcriptional kinases other than CDK9.

5.3.1 Future directions

ICP22 could interact with both subunits of P-TEFb; perhaps the contacts with both subunits are required for robust interaction *in vivo.* The region of ICP22 interacting with CDK9 is further refined to residues 198 to 245. The Tyr-193 residue of ICP22 that is proposed to be responsible for the interaction with Cyclin T1 must be considered tentative and requires more biochemical validation, for example, by point mutation of this residue to prevent the interaction with Cyclin T1. The structure of P-TEFb, in complex with HIV-1 Tat or CDK9 inhibitors, is already known (Baumli et al., 2008;Baumli et al., 2010;Tahirov et al., 2010;Gu et al., 2014). It would be ideal to make several truncated ICP22 constructs, ideally containing the Tyr-193, and test each of them for co-crystal formation with recombinant P-TEFb. Although the pull-down assay using recombinant proteins is an ideal approach to detect and analyse protein-protein interaction, this method suffers from a degree of false negatives or false positives. Other good approaches to determine direct ICP22 and P-TEFb

interaction are biophysical methods, such as the Surface Plasmon Resonance (SPR)-based analytical methods or Fluorescence Resonance Energy Transfer (FRET) microscopy imaging of GFP-tagged proteins in live cells.

While ICP22 is capable of interacting with transcriptional kinases other than CDK9 and they require cyclins as regulatory subunits, it would be of interest to investigate further the interaction between ICP22 and other transcriptional cyclins.

5.4 Proposed model for the role of ICP22 protein in the regulation of expression of host cell genes

Productive infection of HSV-1 can be achieved through physical interaction between viral and cellular proteins, which results in the modification of cellular proteins at the cost of cellular events to fulfil the needs of the virus. HSV-1 utilises the host Pol II and its associated machinery to regulate viral gene expression. Alteration of host cell gene expression provides advantages for the viral replication. This enables viral genes to have unlimited access to and control of the essential components of the cellular Pol II transcription machinery.

My current model is that, in uninfected cells (Figure 5.1, left panel), the negative transcription elongation factors (N-TEFs) cause Pol II stalling. Subsequent recruitment of P-TEFb allows phosphorylation of N-TEFs and Ser2 Pol II CTD, which leads to productive transcription elongation. In the context of HSV-1-infected cells (Figure 5.1, right panel), ICP22 affects transcription by either inhibiting P-TEFb or helping to recruit P-TEFb to viral genes where viral tegument VP16 reverses P-TEFb inhibition. Consequently, transcription of cellular genes is inhibited, whereas the viral genes become activated. However, the function of other proteins involved in transcription, including additional CTD kinases, might also be affected by ICP22.

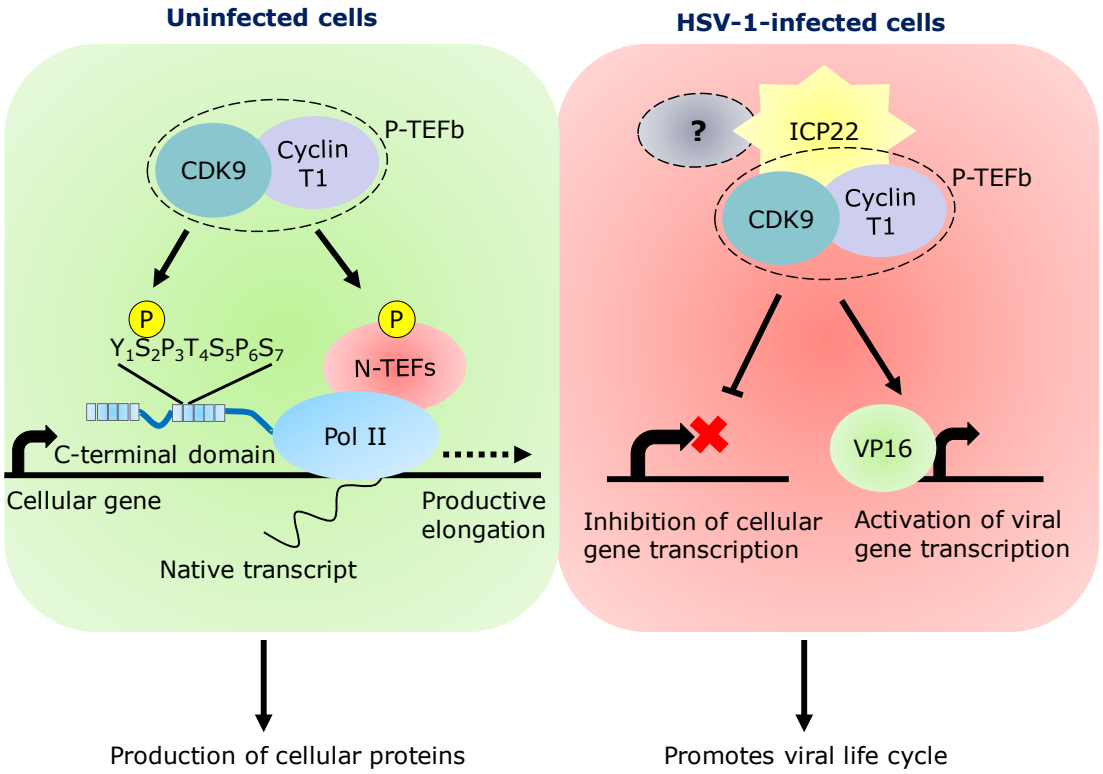


Figure 5.1. Model of the effect of ICP22 on transcription of host cell genes. ICP22 interacts with P-TEFb and possibly other transcriptional kinases and inhibits elongation of transcription of host cell genes.

5.5 Conclusions

In conclusion, the studies presented in this thesis demonstrate that inhibition of CDK9 upon transient expression of ICP22 results in failure of Pol II to negotiate the EEC and premature termination of Pol II near the newly-described poly(A)-associated checkpoint, in a similar way to small molecule CDK9 inhibitors (Laitem et al., 2015). As ICP22 can bind to transcriptional kinases other than CDK9, as shown by a mass spectrometry approach and validation with immunoblots, this suggests that ICP22 is a sophisticated viral protein, which may have multiple targets. Taken together, these finding have implications for Pol II transcription regulation by ICP22, especially during negotiation of the promoter-proximal checkpoint to enter productive elongation and at the newly-described poly(A)-associated checkpoint. Whether ICP22 facilitates the use of transcriptional kinases other than CDK9 to promote the viral gene expression is currently not known and is of great interest.

Chapter 6| Materials and Methods

6.1 Materials

6.1.1 Bacteria and mammalian cell lines

For amplification of plasmid DNA, chemicompetent *E.coli* strains NM554 (Genotype *recA13 araD139 Δ(ara-leu)7696 Δ(lac)l7A galU galK hsdR rpsL (Strr)mcrA mcrB*) was obtained from Stratagene. Chemically competent cells were prepared as previously described in (Green and Rogers 2013).

For recombinant protein expression and purification, chemicompetent *E.coli* Rosetta^{TM} cells (Genotype *F-ompT hsdSB(rB- mB-) gal dcm pRARE (CamR)* was obtained from Millipore.

The human embryonic kidney cells HEK293 for ChIP and transient transfection experiments was obtained from Professor Shona Murphy and have been used before (Laitem et al., 2015).

The HeLa nuclear extract (in 20 mm Hepes, pH 7.9, 100 mm KCl, 20% glycerol, 0.5 mm PMSF, 0.5 mm DTT, 0.2 mm EDTA) was provided by Professor Shona Murphy and prepared as previously described in (Dignam et al., 1983).

6.1.2 Plasmids

pcDNA3.1 (Invitrogen) was used to generate Myc-SCR (ICP22 residues 193-256) for ChIP-qPCR.

pET28a (kindly provided by Dr. Ivan Ahel, Sir William Dunn School of Pathology Oxford) was used to generate His-FL (ICP22 full-length), His-SCR (ICP22 residues 193-256), His-SCR-5/11 (ICP22 residues 198-245), His-vCon (VP16 residues 177-224). pET28a was used to clone Flag-Cyclin T1 (residues 1-266) carrying N-terminal 3x Flag peptides. pET28a used to generate His-IpgC of *S. flexneri* was kindly provided by Professor Susan Lea (Sir William Dunn School of Pathology Oxford). All constructs were used for pull-down with HeLa cell extract or single protein-protein interaction.

pGFP-VP16 (kindly provided by Professor Peter O'Hare, Imperial College London) was used as a template to generate sequences for His-vCon (VP16 residues 177-224)

pGEX-6P-1 (GE Healthcare) was used to generate GST-CDK9 for single protein-protein interaction.

6.1.3 DNA oligonucleotides

All DNA oligonucleotides were synthesised and desalted by Sigma-Aldrich. Lyophilised DNA oligonucleotides were resuspended in TE buffer to final concentration of 100 μM and stored at -20˚C. All DNA oligonucleotides used in this work are listed in Table 6.1.

6.1.4 Non-commercial buffers

All buffers prepared in house were using standard RNase- and DNase-free sterile techniques, and using double-distilled H_2O (18.2 MΩ). These non-commercial buffers are listed in Table 6.2. Chemicals used to prepare the buffers were of molecular biology grade and obtained from Sigma-Aldrich, Thermo-Fisher or VWR International. Materials purchased from supplier other than these three are listed in Table 6.2.

6.1.5 Commercial buffers, enzymes and kits

All the commercial solutions and reagents purchased were used according to the manufacturer's protocol and are listed in Table 6.3.

6.1.6 Antibodies

Antibodies were obtained commercially as listed in Table 6.4, and stored at 4˚C or at -20˚C.

Table 6.1 List of DNA oligonucleotides

For qPCR

Name	Sequence (5' to 3')
KPNB1 #NF	GGATCACATCACTGCACTCC
KPNB1 #NR	GTCAGGCTGGTCTCGAACT
KPNB1 TSS+0.9 F	AGGGAAGGGAAGTTAGGTTGCG
KPNB1 TSS+0.9 R	GAGCTTCCACATGGCCCCTA
KPNB1 TSS+5.3 F	GCAAAGCAGGATACTAAGTGATCCGA
KPNB1 TSS+5.3 R	GAAGCCACAATTGACCTAGAGC
KPNB1 pA-4.8 F	CTGAGGAAACTGAAGAACCAAG
KPNB1 pA-4.8 R	GAAGGCAGTGCTTGCCAGAAT
KPNB1 pA-2.9 F	GAGGAGTGTGCACGGATGCTGAA
KPNB1 pA-2.9 R	CCAAGATGGCCGATGTTATGG
KPNB1 pA-0.4 F	TAGTTACCGTCTGCTTGGGAAGATG
KPNB1 pA-0.4 R	CCTCTGACAGCAAGTCCAACATT
KPNB1 pA+1.4 F	GACTCATCACACCAAGGTCAC
KPNB1 pA+1.4 R	GATAGTGCTGGGAAGGAAATGG
KPNB1 pA+2.6 F	GTACATCTCAGCTTTGGCATATG
KPNB1 pA+2.6 R	GCCCAGAACATAGCAGGCATTGC
KPNB1 pA+4.1 F	GTTTCACCGTGTTAGCCAGGATGG
KPNB1 pA+4.1 R	CCACAGCCATGTTCATTTCTGC

For PCR

Name	Sequence (5' to 3')
His-FL F	TATACCATGGCCCATCATCATCATCATCAC ATGGCCGACATTTCCCCAGGCGCTTTTGCG CCTTGTGTAAAAGC
His-FL R	ATCCGGATCCTCACGGCCGGAGAAACGTGT CGCTGCACGGATAGGGGCAGGCGGTGGAGA AGCGCATTTTCC
His-SCR F	CCCCTGGGATCCATGTACATGCGCCAGTGTAT CAATCAGCTGTTTCGG
His-SCR R	CTGACAGCGGCCGCTCAGCGCAGGTGTA TGCCCCAGGTTCC
His-SCR-5/11 F	CCCCTGGGATCCATGATCAATCAGCTGTT TCGGGTCCTGCGG
His-SCR-5/11 R	CTGACAGCGGCCGCTCACTGCAGCAAGC GGCACCACGTGCGCGG
GST-CDK9 F	CGCGGATCCGCAAAGCAGTACGACTCGG TGGAGTGC
GST-CDK9 R	CTTGCGGCCGCTCAGAAGACGCGCTCAAA CTCCGTCTGGTT
Flag-Cyclin T1 (1-266) F	ACCATGGGCGACTACAAAGACGATGACGA CAAGGAGGGAGAGAGGAAGACAAC
Flag-Cyclin T1 (1-266) R	CAGCGGCCGCTCATTTCTTGGCA GCCTCGCATGCCCTCCAATTC

Table 6.2 List of buffers

General	
TE	10 mM Tris pH 8.0, 0.1 mM EDTA
Coomassie stain solution	0.1% Coomassie Blue R250, 10% acetic acid, 50% methanol, 40% H_2O
Coomassie de-staining solution	10% acetic acid, 50% methanol, 40% H_2O
RIPA buffer (Cell lysis other than for ChIP)	50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS, 5 mM EDTA
2xTY medium	1.6% w/v bacto-tryptone, 1% w/v yeast extract, 0.5% w/v NaCl
LB broth	1% w/v bacto-tryptone, 0.5% w/v yeast extract, 1% w/v NaCl
TAE (1x)	40 mM Tris-HCl pH 7.6, 20 mM acetic acid, 1 mM EDTA
TBE (1x)	89 mM Tris-HCl pH 7.6, 89 mM boric acid, 2 mM EDTA
Laemmli buffer	63 mM Tris-HCl pH 6.8, 2% SDS, 20% glycerol, 0.005% bromophenol blue, 100 mM DTT
For ChIP	
ChIP lysis buffer	10 mM Tris-HCl pH 8.0, 0.25% Triton X-100, 10 mM EDTA, 1x Protease inhibitor cocktail
ChIP wash buffer	10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 1x Protease inhibitor cocktail
ChIP sonication buffer	10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.1% SDS, 1x Protease inhibitor cocktail
ChIP RIPA buffer	10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% Sodium deoxycholate
High Salt wash buffer	10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% Triton X-100, 0.1% Sodium deoxycholate
LiCl wash buffer	10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 1% Sodium deoxycholate
Elution buffer	100 mM Sodium bicarbonate, 1% SDS, 10 mM DTT
Proteinase K buffer (5x)	50 mM Tris-HCl pH 7.5, 25 mM EDTA, 1.25% SDS
For protein co-immunoprecipitation	
Lysis buffer	50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100

TBS buffer	50 mM Tris HCl pH 7.4, 150 mM NaCl, pH 7.4

For His purification	
Lysis buffer (E1)	50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mg/ml lysozyme, 0.2% Triton X-100, 1x Protease inhibitors
Equilibration buffer	Phosphate-buffered saline (20mM sodium phosphate, 500mM NaCl), 5 mM Imidazole pH 8.0
Wash buffer	Phosphate-buffered saline (20mM sodium phosphate, 500mM NaCl), 80 mM Imidazole pH 8.0

For GST purification	
NETN-DPI	20 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.5 mM DTT, 0.2 mM PMSF, 1x Protease inhibitors

For Flag purification	
Lysis buffer (E1)	50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA, 1 mg/ml lysozyme, 0.2% Triton X-100, 1x Protease inhibitors
TBS buffer	50 mM Tris HCl, 150 mM NaCl, pH 7.4

For pull-down of HeLa cell extract	
Pull down buffer	3.25 mM PBS pH7.4, 70 mM NaCl, 0.01% Tween-20 (stringent buffer contains 125 mM NaCl)
Binding/Wash buffer	50 mM PBS pH7.4, 500mM NaCl
Elution buffer (designed for His-tag)	50 mM PBS pH7.4, 300mM NaCl, 500 mM Imidazole

For SDS-PAGE (Tricine gel was based on (Haider et al., 2012))	
Resolving gel	375 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 8-15% acrylamide mix
Stacking gel	125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED, 4% acrylamide mix
Running buffer	250 mM Tris, 2 M Glycine, 1% SDS
Resolving gel (Tricine)	1 M Tris-HCl pH 8.8, 0.3% APS, 0.06% TEMED, 15-16% acrylamide mix
Stacking gel (Tricine)	380 mM Tris-HCl pH 8.8, 0.9% APS, 0.1% TEMED, 4% acrylamide mix
Running buffer (Tricine)	25 mM Tris, 100 mM Tricine, 0.1% SDS

Table 6.3 List of commercial materials

Materials	Supplier
0.2 μm Nitrocellulose membrane (10600128)	VWR International
0.45 μm Nitrocellulose membrane (88018)	Life Technologies
3x Flag peptide (F4799)	Sigma-Aldrich
30% 37.5:1 acrylamide/bisacrylamide (J61505.K2)	Alfa Aesar
Agarose powder (NBS-AG500)	NBS Bio.
Ampicillin (A-9518)	Sigma-Aldrich
Beads - His Mag Ni magnetic beads (28-9673-90)	VWR International
Beads - Pierce Glutathione Magnetic beads (78601)	Fisher Scientific
Beads - Anti-Flag M2 Magnetic beads (M8823)	Sigma-Aldrich
Bradford dye reagent (5000006)	Bio-Rad
BSA Fraction V (10735086001)	Sigma-Aldrich
Cellulose sheets (CSL-GDVHCS)	Cleaver Scientific
Dimethylsulfoxide (DMSO) (D8418)	Sigma-Aldrich
Dithiothreitol (R0861)	Life Technologies
DNA Hyperladder (100bp - 33056) (1 kb - 33053)	Bioline
Dulbecco's Modified Eagle's Medium (DMEM) (D6429)	Sigma-Aldrich
Dynabeads Protein G (10004D)	Life Technologies
Fetal bovine serum	Invitrogen
Formaldehyde (F8775)	Sigma-Aldrich
Glycine (G7126)	Sigma-Aldrich
Herring Sperm DNA (D1815)	Promega
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich
Kanamycin (60615)	Sigma-Aldrich
KM05283 (Maybridge)	Fisher Scientific
L-glutamine (25030-024)	Gibco
Lipofectamine 2000 (11668027)	Life Technologies
Lysozyme (L6876)	Sigma-Aldrich
Dried semi-skimmed milk	Tesco
PBS (BE17-516F)	Lonza
Penicillin / streptomycin (15140-122)	GIBCO
Platinum Pfx DNA polymerase (11708013)	Life Technologies
Ponceau S staining (K793)	VWR International
Precision Plus Dual Color Protein Marker (1610374)	Bio-Rad
Plasmid Maxi Prep Kit (12162)	QIAGEN
Plasmid Mini Prep Kit (27104)	QIAGEN
QIAquick PCR Purification Kit (28104)	QIAGEN
QuantiTect SYBR Green PCR Master Mix	QIAGEN
RNase A (12091021)	Life Technologies
SOC Outgrowth medium (B9020S)	New England Biolabs
Visualizer WB detection kit (64-201)	Millipore
X-ray Fujifilm RX X100 (12735325)	Fisher Scientific

Table 6.4. List of antibodies. The dilutions for immunoblotting and the amount of antibody used per sample for ChIP experiments are indicated.

Primary antibodies	Manufacturer	Immunoblot	ChIP
Pol II (N-20) (sc-899-)	Santa-Cruz	1:1000	-
Pol II (N-20) X (sc-899-X)	Santa-Cruz	-	2 µg
Normal Rabbit IgG (sc-2027)	Santa-Cruz	-	1 µg
Histone H2B antibody (ab1790; rabbit polyclonal)	Abcam	-	1 µg
Histone H3 antibody (ab1791; rabbit polyclonal)	Abcam	1:10000	1 µg
Myc epitope (ab9132; goat polyclonal)	Abcam	1:5000	-
ß-actin (sc-1615; goat polyclonal)	Santa Cruz	1:5000	-
Ser2P H5 (ab5095; rabbit polyclonal)	Abcam	1:1000	-
KPNB1 (A300-482A; rabbit polyclonal)	Bethyl Lab.	1:1000	-
GST (27-4577-01; goat polyclonal)	GE Healthcare		
Anti-Flag M2 (F1804; mouse monoclonal)	Sigma-Aldrich	1:10000	-
6xHis (ab18184; mouse monoclonal)	Abcam	1:5000	-
6xHis-tag with HRP (ab1187; rabbit polyclonal)	Abcam	1:5000	-
CDK9 H169 (sc-8338)	Santa Cruz	1:1000	-
CDK9 C20 (sc-484; rabbit polyclonal)	Santa Cruz	1:1000	-
CDK9 pThr-186 (25495; rabbit polyclonal)	Cell Signalling	1:200	-
Cyclin T1 H-245 (sc-10750; rabbit polyclonal)	Santa Cruz	1:1000	-
α-Tubulin (00-301-880l; mouse monoclonal)	Rockland	1:10000	-
Hexim1 (ab25388; rabbit polyclonal)	Abcam	1:5000	-
MEPCE (14917-1-AP; rabbit polyclonal)	Proteintech	1:1000	-
LARP7 E5 (sc-515209; mouse monoclonal)	Santa Cruz	1:200	-
PAF1 (NB600-274; rabbit polyclonal)	Novus Biologicals	1:1000	-
Leo1 (A300-175A; rabbit polyclonal)	Bethyl Lab.	1:1000	-
CDC73 (A300-170A; rabbit polyclonal)	Bethyl Lab.	1:1000	-
SPT5 D10 (sc-390961; mouse monoclonal)	Santa Cruz	1:500	-
MED1 (17-10530; rabbit polyclonal)	Millipore	1:1000	-
Tat-SF1 (T27750; mouse monoclonal)	BD Transduction	1:1000	-
CDK12 2F6 (WH0051755M2; mouse monoclonal)	Sigma-Aldrich	1:1000	-

CDK12 (A301-679A; rabbit polyclonal)	Bethyl Lab.	1:1000	-
CDK11 (A300-310A; rabbit polyclonal)	Bethyl Lab.	1:1000	-
CDK8 C19 (sc-1521; goat polyclonal)	Santa Cruz	1:500	-
CDK13 CDC2L5 (A301-458A; rabbit polyclonal)	Bethyl Lab.	1:500	-
CDK7 C19 (sc-529; rabbit polyclonal)	Santa Cruz	1:500	-
CDK10 C19 (sc-51266; goat polyclonal)	Santa Cruz	1:200	-

Secondary antibodies	**Manufacturer**	**Dilutions**
IRDye 800CW Goat Anti-Rabbit IgG (926-32211)	Li-Cor	1:10000
IRDye 800CW Goat Anti-Mouse IgG (926-32210)	Li-Cor	1:10000
IRDye 800CW Donkey Anti-Goat IgG (926-32214)	Li-Cor	1:10000
IRDye 680CW Goat Anti-Rabbit IgG (926-32221)	Li-Cor	1:10000
IRDye 680CW Goat Anti-Mouse IgG (926-32220)	Li-Cor	1:10000
IRDye 680CW Donkey Anti-Goat IgG (926-32224)	Li-Cor	1:10000
Goat anti - mouse HRP (1858413)	Pierce	1:10000
Donkey anti-goat HRP (sc-2020)	Santa Cruz	1:10000
Goat anti-rabbit HRP (12-348)	Millipore	1:10000
Anti-rabbit IgG HRP (A0545-1ML)	Sigma-Aldrich	1:10000
Donkey Anti-Goat IgG HRP (ab6885)	Abcam	1:10000

6.2 Bacterial techniques

6.2.1 Transformation of competent bacterial cells

A total of 10-50 ng of plasmid DNA was added to the chemically competent cells (*E.coli* NM554 strain and Rosetta™ *E.coli*), mixed gently, followed by incubation on ice for 10 min. Heat-shock was performed for 5 min at 37˚C, after which cells were placed back on ice for a further 2 min. To grow *E.coli* on agar plates, ampicillin was added to final concentration of 100 μg/ml. Kanamycin was added to final concentration of 30 μg/ml. Transformed cells were then plated on a LB-agar plate containing appropriate selective antibiotics and incubated overnight at 37˚C. For pET28a vector-transformed Rosetta™ *E.coli*, recovery steps is required before plating

on LB-agar containing kanamycin. Recovery steps were performed by adding 100 µl pre-warmed SOC outgrowth medium to the cells and incubated at 37˚C for 1 h with gentle shaking.

6.2.2 Growth of bacterial starter culture

For small-scale growth of bacteria, the colonies were picked from the LB-agar plate grown overnight at 37˚C, inoculated into 5 to 10 ml of LB broth or 2xTY medium containing appropriate selective antibiotic and grown overnight at 37˚C in a horizontal shaker. To grow *E.coli* in a liquid medium, ampicillin was added to final concentration of 100 µg/ml, while kanamycin was added to final concentration of 50 µg/ml.

For medium-scale growth of bacteria, the small-scale starter culture was first grown for 6-8 h at 37˚C in a horizontal shaker. This starter culture was then transferred into 100 mL LB or 2xTY containing appropriate selective antibiotics and incubated overnight at 37˚C.

6.2.3 Preparation of plasmid DNA

For small-scale preparation of plasmid DNA, 2 mL of the overnight small-scale cultures were pelleted and processed according to manufacturer's protocol (Plasmid Mini Prep Kit, QIAGEN).

For large-scale preparation of plasmid DNA, 100 mL of the medium-scale overnight cultures were pelleted at 6000 g for 15 min at 4˚C and processed according to manufacturer's protocol (Plasmid Maxi Prep Kit, QIAGEN).

Plasmid DNA was quantified using a Nanodrop 1000 (Thermo Fisher Scientific), and its concentration was adjusted to 1 µg/µl. The purified plasmid DNA was sent to DNA sequencing and remaining DNA was stored at -20˚C until further use.

6.3 Mammalian cell culture techniques

6.3.1 Propagation of mammalian cell lines

Human embryonic kidney HEK293 was maintained in growth medium Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum, 2 mM L-glutamine, 50 U/mL penicillin and 50 µg/mL streptomycin, and cultured at 37˚C in 5% CO_2. HEK293 cells were split regularly every three days with PBS containing 1 mM EDTA at approximately 80% confluency.

6.3.2 Pharmacological treatments of mammalian cell culture

For ChIP experiments, HEK293 cells were grown overnight in 10 cm dishes. On the following day, at approximately 80% confluency, the HEK293 cells were treated with 100 μM KM05283 for 30 min before ChIP experiments.

6.3.3 Plasmid transfections

For immunoblotting, the HEK293 cells were grown overnight in 6-well plate. For ChIP experiments, the HEK293 cells were grown overnight in 10 cm dishes. On the following day, at approximately 80% confluency, the HEK293 was transfected with 10 μg of plasmid DNA using Lipofectamine2000, following the manufacturer's protocols. Cells were harvested at designated time points.

6.3.4 Freezing cells

HEK293 to be stored at -80˚C were lifted from a T75 flask with PBS containing 1 mM EDTA. Cell pellet were gently resuspended in 1 mL of complete DMEM media supplemented with 20% fetal bovine serum and 10% DMSO. The cells were then transferred into a cryotube and placed in a pre-cooled box filled with isopropanol in a -80˚C freezer.

6.4 DNA analysis methods

6.4.1 Quantitation of DNA

Small amount of plasmid DNA (1 μl) was used to calculate the concentration of the DNA using a Nanodrop 1000 (Thermo Fisher Scientific), at the absorbance of 260 nm. The ratio of 260/280 nm was used to assess the purity of the DNA samples.

6.4.2 Agarose gel electrophoresis

The agarose gel electrophoresis was used to measure sizes of plasmid constructs and PCR products. The DNA was mixed with 1x orange DNA loading dye, and separated at 80V using 0.8-1.5% agarose gels containing 1xTAE buffer and 0.005% ethidium bromide. DNA sizes were estimated by comparison with a 100 bp or 1 kb DNA ladder loaded alongside the samples. DNA bands were visualized under UV light.

6.4.3 DNA sequencing

DNA sequencing was carried out to confirm sequences of all plasmid constructs. A

total of 100 ng/µl of purified plasmid DNA and 3.2 pmol/µl of appropriate primers were required for a single Sanger sequencing reaction. The Sanger sequencing reactions were carried out by Source BioScience UK, Department of Biochemistry, Oxford.

6.4.4 Polymerase chain reaction (PCR)

The tagged sequences of ICP22 (Myc-tag and His-tag) were PCR-amplified using ICP22 template provided by Dr Justyna Zaborowska/Professor Shona Murphy. The His-vCon was PCR-amplified using VP16 template provided by Professor Peter O'Hare (Imperial College, London). The Flag-Cyclin T1 (residues 1-266) and GST-CDK9 were PCR-amplified using templates provided by Professor Shona Murphy.

A total of 20 ng of DNA template was used in a 25 µl PCR reaction containing 1 U Platinum *Pfx* DNA polymerase, 3 mM $MgCl_2$, 0.4 mM dNTPs, 0.4 µM of each appropriate primers listed in Table 6.1.

PCR was set at 94˚C for 5 min, followed by 10 cycles of 94˚C for 20s, 65˚C for 30s and 72˚C for 5 min. This was followed by 20 cycles of 94˚C for 20s, 65˚C for 30s and 72˚C for 5 min with the 72˚C becoming 20s longer in each cycle. After a final step at 72˚C for 10 min, PCR products were resolved on a 0.8 to 1.5% agarose gel.

6.4.5 Real-time quantitative PCR

The real-time qPCR measures the quantity of amplified DNA after each PCR cycle. The fluorescent dye SYBR screen intercalates into double-stranded DNA, and as PCR cycle ensues more SYBR green molecules intercalates into more double-stranded DNAs, resulting in an increase in fluorescence signal.

In this work, ChIP assays were evaluated by real-time qPCR using QuantiTect SYBR Green PCR master mix in a Rotor-Gene 3000 cycler (Corbett Research), equipped with Rotor-Gene 3000 Software analysis program. The immunoprecipitated and control DNA samples were assessed in triplicate reactions.

Each reaction contained: 5 µl of 2x QuantiTect SYBR Green PCR, 3 µl RNase/DNase-free water, 10 pmol/µl of each oligonucleotides and 1 µl of purified ChIP DNA template. Oligonucleotides for *KPNB1* are listed in Table 6.1

The thermal cycles were: 95˚C for 15 min; 40 cycles of 94˚C for 15 sec, 57˚C for 15 sec, 72˚C for 15 sec. Last cycle ended with 10 min melt from 50˚C to 99˚C, with 1˚C increase each 5s.

The analysis program set a threshold value representing fluorescent signal level for all samples at the end of the thermal cycling and converted the value into threshold cycle (Ct). The Ct value (copies/ µl) for each standard was used to generate a standard curve. For ChIP analysis, a standard curve was produced for each primer set using serial dilutions of the input DNA sample (1:5, 1:25, 1:125 and 1:625 dilutions). The Ct value (copies/ µl) for each samples were quantified relative to the standard, resulting in 'ChIP signal'. This ChIP signal is presented as a percentage of total DNA after removal of the background signal from the immunoprecipitation with the IgG antibody. Experiments were replicated three times to ensure reproducibility, and each ChIP sample was measured in triplicate by qPCR. In ChIP experiments, the p-values were calculated by two-tailed unpaired t-test. Data represent the mean ± SEM (standard error of the mean) of three independent experiments and $p<0.05$ was considered to be statistically significant.

6.5 Protein analysis methods

6.5.1 Preparation of whole cell extract

HEK293 cells were washed twice with ice-cold PBS and cell pellets were resuspended in RIPA buffer. After incubation for 20 min on ice, samples were sonicated 3x for 30s on, 30s off at high power setting (Bioruptor, Diagenode) to reduce viscosity. Cell lysates were then separated by centrifugation at 16000 xg for 10 min at 95˚C. The supernatant was stored at -20˚C until further use.

6.5.2 Quantitation of protein concentration

Total protein concentration was determined using Bradford method (Bradford 1976). For each sample, 200 µl 1x Bradford dye reagent were mixed with 795 µl PBS and 5 µl of protein sample. This mixture without any protein sample was used as blank control. The mixtures were transferred to cuvettes and after 5 min incubation at room temperature, the absorbance at 595nm was measured. A standard curve generated from a series of BSA dilutions of ranging from 0.2 mg/ml to 1 mg/ml was used to calculate the protein concentration of the samples. Each protein samples were measured twice.

6.5.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

Protein samples were first separated by SDS-PAGE, and depending on the purpose, the gel was either directly stained with Coomassie or proteins were transferred to a nitrocellulose membrane for further identification by immunoblotting.

Depending of the size of the protein of interest, the resolving gels prepared ranged from 8% to 15% of acrylamide mix, while the stacking gels were 4% of acrylamide mix (Table 6.2). For very small proteins (less than 12 kDa), the gels were cast using separate buffer designed for Tricine-SDS-PAGE (Table 6.2) (Haider et al., 2012).

Proteins were denatured in Laemmli buffer by heating at 95˚C for 5 min. Electrophoresis was run in 1x running buffer, first at 100V until the proteins moved to separating gel, and then at 150V until the proteins reached bottom of the gels.

6.5.4 Coomassie staining.

SDS-PAGE gels were stained in Coomassie staining solution for at least 1 h followed by destaining with the de-staining solution. After running the samples, the gel was washed with H_2O, and dried in between cellulose sheets.

6.5.5 Immunoblotting

Proteins in the gel were transferred onto a nitrocellulose membrane soaked in the 1x TBE, by semi-dry blotter for 1 h at 10V using the Trans-blot SD Semi-Dry transfer cell (Bio-Rad). After semi-dry transfer, the nitrocellulose membrane was stained with Ponceau S stain for 5 min on a shaker at room temperature. The stain can be removed by extensive washing with ddH2O.

The de-stained membranes were blocked in PBS containing 5% milk and 0.1% Tween-20 for 1 h at room temperature. The blocking solution was removed and the membranes were incubated for 1 h at room temperature or overnight at 4˚C with the appropriate primary antibodies, diluted in PBS containing 1% milk and 0.1% Tween-20. After incubation, the membranes were washed for 3x 10 min with PBS containing 1% milk and 0.1% Tween-20 and incubated for 1 h at room temperature with HRP-conjugated secondary antibodies, appropriately diluted in PBS containing 1% milk and 0.1% Tween-20. After another 3x 10 min washes, the proteins signal on the membrane were detected by ECL (Visualizer WB detection kit), by developing X-ray films in a compact X4 X-ray film processor (Xograph imaging system, UK). Alternatively, the immunoblot analysis was performed using the Odyssey infrared

scanner using the secondary antibodies from Li-Cor. Instead of PBS and milk, the buffers used for blocking and diluting antibodies for Li-Cor were replaced with TBS and BSA.

6.5.6 Protein co-immunoprecipitation

Components of buffers are listed in Table 6.2. HEK293 cells transfected overnight with Flag-FL and Flag-SCR of ICP22 were washed with PBS, and lysed in lysis buffer, followed by sonication for 3x for 30s on, 30s off at high power setting (Bioruptor, Diagenode). After removal of insoluble materials, 1 mg of protein extracts were incubated with 10 µl of Anti-Flag M2 Magnetic beads overnight. The beads were washed extensively with TBS buffer, and eluted with 3x Flag peptide for analysis of co-immunoprecipitated proteins using immunoblot.

6.5.7 Chromatin immunoprecipitation (ChIP)

ChIP was performed as described in (Laitem et al., 2015), with minor modifications.

Cross-linking. HEK293 cells were grown overnight in 10 cm dishes in 10 ml DMEM complete media. The formaldehyde was added to a final concentration of 1% to the treated or transfected cells were and mixed for 10 min on a horizontal shaker at room temperature. The reaction was quenched with 652 µl of 2 M glycine incubated for 5 min at room temperature. Cells were washed twice with pre-chilled PBS, and then collected by scrapping. All subsequent steps were carried out at 4˚C. Cells were pelleted by centrifugation at 1500 g for 10 min. The cell pellet can be kept at -80˚C until use.

Cell lysis and sonication. The cells were resuspended in 0.5 mL ChIP lysis buffer and incubated for 10 min. Nuclei were spun down at 1500 g for 5 min and the supernatant was discarded. The pelleted nuclei were resuspended in 0.5 mL of ChIP wash buffer, and again spun down at 1500 g for 5 min. The pelleted nuclei were resuspended in 300 µl ChIP sonication buffer, and left rotating for 15 min. The samples were sonicated for 1 h for 30s on, 30s off at high power setting in an icy water-bath (Bioruptor, Diagenode). After sonication, the samples were spun down to pellet debris. To pre-clear the sheared DNA, the supernatant was transferred to another tube containing 10 µl Dynabeads Protein G and incubated for 30 min on rotator.

Immunoprecipitation (IP). The supernatant was recovered and split into 100 µl fractions, and incubated with appropriate amount of antibody against the proteins of interest. Per each IP samples, 15 µl of Dynabeads Protein G were blocked with 15 µl RIPA containing 4 mg/ml Herring Sperm DNA and 4 mg/ml BSA. All samples were left rotating at 4˚C overnight.

The blocking solution was removed from Dynabeads Protein G, and the sonicated samples pre-incubated with antibody were added to the beads and left rotating for 1 h at 4˚C. Beads were then separated on magnetic rack DynaMag-2 Magnet (Thermo Fisher) and washed as follows: 3x with 300 µl RIPA, 3x with 300 µl high salt buffer, 2x with 300 µl LiCl was buffer and 2x with 300 µl TE buffer.

The immunoprecipitated DNA-protein complexes were then eluted with 2x 15 min steps with 50 µl of elution buffer containing 10 mM DTT at room temperature. The input samples were diluted 1/10 with elution buffer.

Reverse crosslinking and DNA purification. Each eluted samples was incubated with 10 µg RNaseA for 30 min at 37˚C to degrade the RNA. The samples were then incubated for 4-5 h at 65˚C with a final concentration of 200 mM NaCl to reverse the cross-linking. The DNA was then precipitated with 350 µl ethanol at -20˚C overnight.

The precipitated DNA was centrifuged at 20000 g for 20 min, and the supernatant was discarded. Each DNA pellet was resuspended in 100 µl TE buffer with 20 µg of proteinase K in 25 µl of 5x proteinase K, and incubated for 1-2 h at 42˚C to digest DNA-bound proteins. The DNA was then purified using the QIAGEN QIAquick PCR purification kit and eluted in 50 µl for qPCR analysis.

6.5.8 Protein expression and purification

Recombinant protein expression. The overnight small-scale starter culture was diluted 1:10 into fresh 100 ml LB or 2xTY, with appropriate selective antibiotics. The culture was grown to an $OD_{600} \approx 0.6\text{-}0.8$ at 37˚C. Once sufficient cell density had been achieved, the recombinant protein expression was induced by the addition of 1 mM IPTG for 7-8 h or overnight at 25˚C.

Cells expressing recombinant proteins were harvested by centrifugation and resuspended in appropriate lysis buffer. All purification steps were performed on ice or at 4˚C. Bacterial pellet can be stored at -20˚C until use.

His-tag recombinant protein purification. Buffer compositions are listed in Table 6.2. Cell pellet were thawed and resuspended in 5 ml Lysis buffer E1 for 10 min at 4˚C. Cells were lysed by sonication on ice using a Soniprep 150 sonicator (MSE) for several cycles of 10s on, 1 min off until the suspension was cleared. The bacterial lysates was centrifuged at 10000 g for 1 h at 4˚C to separate soluble protein from cell debris. Lysates containing soluble His-tag recombinant protein was added with equal amount of equilibration buffer, before loaded onto pre-equilibrated His Mag Ni magnetic beads, and left rotating overnight at 4˚C. Beads were then separated on magnetic rack DynaMag-2 Magnet (Thermo Fisher), unbound materials were discarded and beads were washed 5x 100 µl with wash buffer. The bound His-tag recombinant protein can be eluted in wash buffer containing 500 mM Imidazole, flash frozen and stored at -80˚C.

GST-tag recombinant protein purification. Buffer compositions are listed in Table 6.2. For purification of GST-CDK9, cell pellet were thawed and resuspended in 5 ml NETN-DPI buffer for 10 min at 4˚C. Cells were lysed by sonication on ice using a Soniprep 150 sonicator (MSE) for several cycles of 10s on, 1 min off until the suspension was cleared. The bacterial lysates was centrifuged at 10000 g for 1 h at 4˚C to separate soluble protein from cell debris. Lysates containing soluble GST-CDK9 were incubated with appropriate quantity of glutathione magnetic beads overnight rotating at 4˚C. Beads were then separated on magnetic rack DynaMag-2 Magnet (Thermo Fisher), unbound materials were discarded and beads were washed 4x 100 µl with NETN-DPI containing 0.5 M NaCl, 0.25 M NaCl, 0.1 M NaCl, 0 M NaCl in order to remove non-specifically bound proteins from the beads. The bound GST-CDK9 can be eluted in last buffer containing 25 mM reduced glutathione adjusted to pH 7, flash frozen and stored at -80˚C.

Flag-tag recombinant protein purification. Buffer compositions are listed in Table 6.2. For purification of Flag-Cyclin T1 (residues 1-266), cell pellet were thawed and resuspended in 5 ml Lysis buffer E1 for 10 min at 4˚C. Cells were lysed by sonication on ice using a Soniprep 150 sonicator (MSE) for several cycles of 10s on, 1 min off until the suspension was cleared. The bacterial lysates was centrifuged at 10000 g for 1 h at 4˚C to separate soluble protein from cell debris. Lysates containing soluble

Flag-Cyclin T1 (residues 1-266) was loaded onto anti-Flag M2 Magnetic beads and left rotating overnight at 4˚C. Beads were then separated on magnetic rack DynaMag-2 Magnet (Thermo Fisher), unbound materials were discarded and beads were washed extensively with TBS buffer. The bound Flag-Cyclin T1 (residues 1-266) can be eluted in last buffer containing 3x Flag peptide, flash frozen and stored at -80˚C.

6.5.9 Pull-down of HeLa cell extract

A total of 500 µl of extracts consisting of 250 µl HeLa cell extract and 250 µl of pull down buffer were pre-cleared with His Mag Ni magnetic beads for 1 h rotating at 4˚C. An equal amount of His-tagged recombinant proteins already bound (non-eluted) to the beads were incubated with the extracts for 1 h rotating at 4˚C. Non-incubated extracts were added with Laemmli buffer and labelled as input sample. The incubated beads were washed 3x 500 µl in binding/wash buffer. Bound proteins were eluted with 100 µl elution buffer for mass spectrometry analysis or with 100 µl Laemmli buffer for SDS-PAGE and immunoblotting analyses.

6.5.10 Mass spectrometry analysis

Eluted bound proteins from pull-down assay were sent to Advanced Proteomics Facility, Department of Biochemistry Oxford. The analysis was performed by the facility, utilising trypsin digestion and Ultimate 3000 RSLCnano HPLC (Dionex, Camberley, UK) coupled to a QExactive Orbitrap mass spectrometer (Thermo Electron, Hemel Hempstead, UK). The database search algorithm used was Mascot v2.2.04, MS/MS Ion Search.

nature structural & molecular biology

CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes

Clélia Laitem[1], Justyna Zaborowska[1], Nur F Isa[1,2], Johann Kufs[3], Martin Dienstbier[4] & Shona Murphy[1]

© 2015 Nature America, Inc. All rights reserved.

Transcription through early-elongation checkpoints requires phosphorylation of negative transcription elongation factors (NTEFs) by the cyclin-dependent kinase (CDK) 9. Using CDK9 inhibitors and global run-on sequencing (GRO-seq), we have mapped CDK9 inhibitor–sensitive checkpoints genome wide in human cells. Our data indicate that early-elongation checkpoints are a general feature of RNA polymerase (pol) II–transcribed human genes and occur independently of polymerase stalling. Pol II that has negotiated the early-elongation checkpoint can elongate in the presence of inhibitors but, remarkably, terminates transcription prematurely close to the terminal polyadenylation (poly(A)) site. Our analysis has revealed an unexpected poly(A)-associated elongation checkpoint, which has major implications for the regulation of gene expression. Interestingly, the pattern of modification of the C-terminal domain of pol II terminated at this new checkpoint largely mirrors the pattern normally found downstream of the poly(A) site, thus suggesting common mechanisms of termination.

RNA pol II elongation is blocked soon after initiation of transcription of many protein-coding genes, owing to the concerted action of the NTEFs, DRB sensitivity–inducing factor (DSIF) and negative elongation factor (NELF)[1–3]. The switch to productive elongation requires the activity of positive elongation factor b (P-TEFb), comprising CDK9 and a cyclin T. P-TEFb phosphorylates NELF, causing it to be released, and DSIF, converting it into a positive elongation factor[1–3]. P-TEFb also phosphorylates Ser2 of the 1-Tyr-Ser-Pro-Thr-Ser-Pro-Ser-7 repeat of the C-terminal domain (CTD) of the large subunit of pol II, to help recruit RNA-processing factors to emerging transcripts[4,5]. This potentially rate-limiting elongation 'checkpoint' can therefore hold back pol II to create a kinetic 'window of opportunity' for the recruitment of factors needed for RNA 5′ capping, productive elongation and RNA processing. In addition, this checkpoint facilitates the rapid and synchronous expression of genes during development[6], and c-myc, NF-κB and HIV Tat recruit P-TEFb to activate transcription elongation[1].

Pol II pausing also occurs at intron-exon junctions to facilitate recruitment of factors involved in splicing[7–9], and a transcription checkpoint is associated with yeast prespliceosome formation[8]. In addition, pol II pausing at the end of the transcription unit is thought to couple 3′ processing to transcription termination[10,11].

We set out to elucidate the contribution of CDK9 to the control of transcription elongation genome wide in human HeLa cells. Our data indicate that productive elongation of the majority of pol II–transcribed human genes requires negotiation of a CDK9-dependent checkpoint early in the transcription cycle. Early-elongation checkpoints are generally (but not exclusively) located within ~200 bp of a transcription start site (TSS), and they occur whether or not pol II is paused close to the TSS. These results are consistent with the effect of CDK9 inhibitors in HEK293 cells[12], mouse embryonic stem (mES) cells[13] and *Drosophila*[14], and they emphasize that CDK9 is a critical kinase for transcription of metazoan pol II–transcribed genes.

Surprisingly, our analysis also revealed an additional major elongation control point close to the terminal poly(A) site, beyond which pol II cannot effectively proceed in the presence of CDK9 inhibitors. CTD phosphorylation associated with premature termination induced by the inhibitors is similar to the CTD phosphorylation pattern associated with normal termination downstream of the poly(A) site. The inhibitors also cause loss of association of CDK9, Spt5 and the poly(A) factors Ssu72 and CstF64, whereas association of the termination factor Pcf11 increases.

Kinase-dependent elongation checkpoints at terminal poly(A) sites provide a point late in the transcription cycle to regulate production of poly(A) mRNA and could, for example, facilitate rapid control of mRNA production from long genes in response to environmental signals and developmental cues.

RESULTS

KM and DRB inhibit transcription of *GAPDH*

To investigate the contribution of CDK9 to transcription elongation, we treated HeLa cervical adenocarcinoma cells with two drugs known to inhibit CDK9: KM05283 (KM) and 5,6-dichlorobenzimidazone-1-β-D-ribofuranoside (DRB)[15] for 15, 30, 45 and 60 min. In accordance with previous findings[16,17], these inhibitors caused loss of nascent transcripts from *GAPDH*, and inhibition was complete after 30 min of

[1]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. [2]Department of Biotechnology, International Islamic University Malaysia, Pahang, Malaysia. [3]Faculty of Science, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg, Germany. [4]Computational Genomics Analysis and Training Programme, Medical Research Council Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK. Correspondence should be addressed to S.M. (shona.murphy@path.ox.ac.uk) or M.D. (martin.dienstbier@cantab.net).

Received 30 July 2014; accepted 6 March 2015; published online 6 April 2015; doi:10.1038/nsmb.3000

Appendix 2

Prospects & Overviews

P-TEFb goes viral

*Justyna Zaborowska[1], Nur F. Isa[1,2] and Shona Murphy[1]**

Positive transcription elongation factor b (P-TEFb), which comprises cyclin-dependent kinase 9 (CDK9) kinase and cyclin T subunits, is an essential kinase complex in human cells. Phosphorylation of the negative elongation factors by P-TEFb is required for productive elongation of transcription of protein-coding genes by RNA polymerase II (pol II). In addition, P-TEFb-mediated phosphorylation of the carboxyl-terminal domain (CTD) of the largest subunit of pol II mediates the recruitment of transcription and RNA processing factors during the transcription cycle. CDK9 also phosphorylates p53, a tumor suppressor that plays a central role in cellular responses to a range of stress factors. Many viral factors affect transcription by recruiting or modulating the activity of CDK9. In this review, we will focus on how the function of CDK9 is regulated by viral gene products. The central role of CDK9 in viral life cycles suggests that drugs targeting the interaction between viral products and P-TEFb could be effective anti-viral agents.

Keywords:
CDK9; CTD; EBV; HIV; HSV; HTLV; P-TEFb.

Introduction

Transcription of many human viruses is dependent on host cell factors such as RNA polymerase II (pol II) [1]. Pol II is made up of 12 subunits (Rpb1–Rpb12) [2]. The carboxyl-terminal domain (CTD) of the largest subunit of pol II (Rpb1) comprises tandemly repeated heptapeptides with the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 [3, 4]. The CTD plays a central role in transcriptional and co-transcriptional RNA processing [4–7] by mediating the recruitment of transcription and processing factors at different steps of the transcription cycle through reversible modification of the residues within the heptapeptide repeats [8]. CTD modification generates a code that regulates the interaction with transcription and RNA processing factors [9–12]. Among the CTD modifications, the phosphorylation of Ser2 (Ser2P) of pol II CTD, is catalyzed by the cyclin-dependent kinase 9 (CDK9) subunit of the positive transcription elongation factor b (P-TEFb). P-TEFb also phosphorylates two negative transcription elongation factors: DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF), which are recruited to the elongation complex in the promoter proximal region to induce pol II pausing [13, 14]. Phosphorylation of NELF promotes its dissociation from the elongation complex, whereas phospho-DSIF remains associated with pol II as an elongation activator [14]. The role of P-TEFb in stimulating productive elongation was elucidated using 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), which is a purine nucleoside analog that inhibits CDK9 activity [15]. Treatment of cells with DRB results in the production of shortened transcripts, diagnostic of the inability of pol II to transcribe through the early elongation checkpoint in the absence of P-TEFb activity [16]. In addition to Ser2 of the pol II CTD, NELF and DSIF, several serine residues in the key transcription factor p53 are phosphorylated by CDK9, which results in p53 activation [17, 18].

Positive transcription elongation factor b comprises the CDK9 kinase and a cyclin partner [15]. CDK9 is an ubiquitously expressed Ser/Thr proline-directed kinase [19, 20]. It was identified in the early 1990s and initially named PITALRE [19, 21]. It exists in two isoforms, CDK9-42 and CDK9-55, which arise from two transcription start sites [22]. The ratio of CDK9-42/CDK9-55 is cell type-dependent [23], and it is unclear if there are functional differences between the two isoforms. CDK9 interacts with T-type cyclins – T1, T2a, T2b and cyclin K [24, 25]. In this heterodimer, the CDK9 subunit provides the enzymatic activity, while the cyclin has a regulatory role [24, 26]. However, mass spectrometry has failed to demonstrate that cyclin K is associated with CDK9 [27]. In addition, a recent study revealed that cyclin K associates with CDK12 and CDK13 [28]. Therefore, cyclins T1 and T2 may be the major cyclins associated with CDK9. In the cell, P-TEFb exists in two major forms. A large amount of cellular P-TEFb is sequestered in an inactive complex with the noncoding 7SK small nuclear RNA (7SK snRNA), hexamethylene bisacetamide-induced protein 1 (HEXIM1) and/or HEXIM2 and the La-related protein 7 [29–35]. The

DOI: 10.1002/bies.201670912

[1] Sir William Dunn School of Pathology, University of Oxford, Oxford, UK
[2] Department of Biotechnology, Kulliyyah of Science, IIUM, Kuantan, Pahang, Malaysia

*Corresponding author:
Shona Murphy
E-mail: shona.murphy@path.ox.ac.uk

Received 7 August 2015; revised 23 September 2015; accepted 26 September 2015 (Originally published as DOI: 10.1002/icd3.1037)

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Appendix 3

A total of 374 proteins bound to ICP22 sorted in alphabetical order

No	Gene names
1	ACTL6A
2	ACTR5
3	AFF4
4	AKAP17A
5	ALKBH2
6	ANKRD11
7	ARID1B
8	ARID5B
9	ASB7
10	ASH2L
11	ATM
12	ATN1
13	ATXN2
14	AXIN1
15	BAZ1B
16	BMS1
17	BPTF
18	BRD4
19	BUD13
20	BZW1
21	CAAP1
22	CACTIN
23	CALML5
24	CBL
25	CBLL1
26	CBX4
27	CCNK
28	CCNL1
29	CCNL2
30	CCNT1
31	CCNT2
32	CDK11B
33	CDK12
34	CDK13
35	CDK8;CDK19
36	CDK9
37	CDYL
38	CHAMP1
39	CHD1
40	CHD2
41	CHD8
42	CHERP
43	CIR1
44	CITED2
45	CLK1
46	CLK3
47	CPSF1
48	CPSF2
49	CPSF3
50	CPSF6
51	CPSF7
52	CSNK1D
53	CTCF
54	CTIF
55	CTR9
56	CXXC1
57	DCAF13
58	DCAF7
59	DCD
60	DDX23
61	DDX42
62	DDX46
63	DHX15
64	DNTTIP1
65	DOCK7
66	DOCK9
67	DPF2
68	DPY30
69	DSP
70	DUSP11
71	DYRK1A
72	E4F1
73	EEFSEC
74	EFTUD2
75	EHMT2
76	ELMSAN1
77	FAM120B
78	FAM76A
79	FAM76B
80	FAM83G
81	FBL
82	FBRS
83	FBXO21
84	FIP1L1
85	FIZ1
86	FOXC2
87	FOXJ3
88	FOXL2
89	FTSJ1
90	FTSJD2
91	GATAD1
92	GLCCI1
93	GNL3
94	GNL3L
95	GPATCH11
96	GPATCH8
97	GPBP1L1
98	GRK5
99	GTPBP4
100	HTATSF1
101	ICP22
102	INO80
103	INO80B
104	INTS1
105	IRS2
106	KANSL1
107	KCTD1
108	KCTD15
109	KIAA1429
110	KLF13
111	KLF16
112	KLF4
113	KLF5
114	LAMA5
115	LARP1
116	LARP7
117	LRCH1
118	LSG1
119	LSM12
120	LUC7L
121	LUC7L2
122	MAP4K4
123	MAP7
124	MAZ
125	MCRS1
126	MEAF6
127	MECP2
128	MED1
129	MED10
130	MED12
131	MED13
132	MED14
133	MED16
134	MED17
135	MED19
136	MED20
137	MED23
138	MED24
139	MED26
140	MED27
141	MED29
142	MED30
143	MED31
144	MED4
145	MED6
146	MED7
147	MED8
148	MED9
149	MEPCE
150	MFAP1
151	MIF4GD
152	MIOS
153	MLLT1
154	MMTAG2
155	MRPS23
156	MSL1
157	MSL3;MSL3P1
158	MTM1
159	MUS81
160	NCOA5
161	NFRKB
162	NFX1
163	NIPBL
164	NONO
165	NSMAF
166	NUDT21
167	NUFIP2
168	NUMA1
169	NUSAP1
170	PAF1
171	PAPOLG
172	PCF11
173	PDCD7
174	PDZD11
175	PELO
176	PHC3
177	PHF3
178	PHF5A
179	PHF6
180	PHGDH
181	PHRF1
182	PIP5K1A
183	PIP5K1C
184	PLEKHA5
185	POGZ
186	POLRMT
187	POP1
188	POP4
189	PPIG
190	PPIL4
191	PPIP5K2
192	PPP1R10
193	PRPF31
194	PRPF38A
195	PRPF38B
196	PRPF40A
197	PRPF4B
198	PRPF6
199	PRPF8
200	PTS
201	PWP1
202	PWP2
203	RBBP5
204	RBBP6
205	RBM15B
206	RBM17
207	RBM25
208	RBM27
209	RBM33
210	RBM7
211	RBMX2
212	RCL1
213	REPIN1
214	RERE
215	REST
216	RING1
217	RNF2
218	RPL10
219	RPL13
220	RPL14
221	RPL15
222	RPL17
223	RPL18
224	RPL19
225	RPL23A
226	RPL24
227	RPL27
228	RPL27A
229	RPL8
230	RPP40
231	RPS11
232	RPS13
233	RPS23
234	RPS24
235	RPS29
236	RPS3A
237	RPS4X
238	RPS6
239	RPS8
240	RSBN1
241	RSBN1L
242	RSRC2
243	RUVBL2
244	SAFB2
245	SAP30BP
246	SART1
247	SEMA3A
248	SEMA3C
249	SENP6
250	SETD1B
251	SF3A1
252	SF3A2
253	SF3A3
254	SF3B1
255	SF3B14
256	SF3B2
257	SF3B5
258	SFPQ
259	SFSWAP
260	SLTM
261	SMARCA4
262	SMARCA5
263	SMARCC1
264	SMARCD1
265	SMARCD2
266	SMURF2
267	SNIP1
268	SNRNP200
269	SNRNP27
270	SNRNP35
271	SNRNP40
272	SNRNP48
273	SNRPA1
274	SNRPB2
275	SOGA2
276	SOWAHC
277	SP1
278	SP4
279	SPATS2L
280	SPTY2D1
281	SUB1
282	SYMPK
283	TAF1;TAF1L
284	TAF10
285	TAF12
286	TAF2
287	TAF3
288	TAF4
289	TAF5
290	TAF6
291	TAF7
292	TAF8
293	TAF9
294	TBL3
295	TCEB3
296	TCF7L2;LEF1;TCF7
297	TEX30
298	TFIP11
299	THAP11
300	THOC1
301	THOC2
302	THOC3
303	THOC5
304	THOC6
305	THOC7
306	TMA16
307	TOP1
308	TOX4
309	TRA2A
310	TRMT2A
311	TRPT1
312	TRRAP
313	TSR1
314	TXNL4A
315	UCHL5
316	USP34
317	USP42
318	USP46;USP12
319	UTP14A
320	VEZF1
321	VIM
322	VPS72
323	WAC
324	WARS2
325	WDR20
326	WDR24
327	WDR3
328	WDR33
329	WDR36
330	WDR5
331	WDR59
332	WDR6
333	WDR82
334	WHSC1L1
335	WTAP
336	WWP2
337	YTHDC1
338	YY1
339	YY2
340	ZBTB14
341	ZBTB17
342	ZBTB40
343	ZBTB43
344	ZBTB7A
345	ZBTB7B
346	ZC3H13
347	ZCCHC8
348	ZEB1
349	ZFP64
350	ZFP91
351	ZGPAT
352	ZIC2
353	ZKSCAN1
354	ZMAT5
355	ZNF131
356	ZNF148
357	ZNF219
358	ZNF239
359	ZNF24
360	ZNF316
361	ZNF574
362	ZNF593
363	ZNF609
364	ZNF629
365	ZNF638
366	ZNF668
367	ZNF691
368	ZNF692
369	ZNF746
370	ZNF76
371	ZNF768
372	ZNF787
373	ZNF8
374	ZRSR2;ZRSR1

Appendix 4

A total of 374 proteins bound to full length of ICP22 sorted in rank based on uniqueness and fold enrichment

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity beads
1	CHERP	28	26	1.59E+10	Unique
2	PRPF4B	52	32	1.29E+10	Unique
3	ATM	81	61	4.59E+09	Unique
4	SYMPK	41	35	4.01E+09	Unique
5	THOC1	35	26	2.95E+09	Unique
6	HTATSF1	20	19	2.85E+09	Unique
7	ICP22-FL	7	7	2.75E+09	Unique
8	WDR36	24	24	2.40E+09	Unique
9	CCNL1	13	9	2.25E+09	Unique
10	LSG1	30	23	2.23E+09	Unique
11	MED14	34	23	2.19E+09	Unique
12	SMARCA5	34	30	2.17E+09	Unique
13	RPL17	11	7	2.05E+09	Unique
14	PRPF38A	11	8	2.03E+09	Unique
15	POGZ	26	22	1.99E+09	Unique
16	RBBP6	45	26	1.91E+09	Unique
17	MED4	11	10	1.83E+09	Unique
18	KIAA1429	28	23	1.77E+09	Unique
19	DOCK9	58	25	1.64E+09	Unique
20	YTHDC1	16	10	1.60E+09	Unique
21	THOC5	20	15	1.60E+09	Unique
22	CHD1	39	25	1.56E+09	Unique
23	CLK3	15	10	1.50E+09	Unique
24	PWP2	22	22	1.47E+09	Unique
25	PPIG	9	7	1.28E+09	Unique
26	MED12	41	27	1.23E+09	Unique
27	RERE	22	13	1.19E+09	Unique
28	THOC6	14	10	1.13E+09	Unique
29	ZBTB7B	9	9	1.05E+09	Unique
30	GTPBP4	26	12	1.02E+09	Unique
31	PWP1	16	12	9.75E+08	Unique
32	RPS23	6	6	9.62E+08	Unique
33	SNRPB2	6	5	9.26E+08	Unique
34	CCNL2	13	6	9.01E+08	Unique
35	MED17	20	10	8.88E+08	Unique
36	RPS4X	20	6	8.76E+08	Unique
37	MTM1	15	9	8.60E+08	Unique
38	TRRAP	37	24	8.35E+08	Unique
39	MIOS	29	13	7.90E+08	Unique
40	EHMT2	19	18	7.89E+08	Unique
41	RPL10	11	5	7.70E+08	Unique
42	ZCCHC8	17	13	7.65E+08	Unique
43	CLK1	14	9	7.40E+08	Unique
44	CHD2	29	17	7.28E+08	Unique
45	GPATCH11	6	5	6.90E+08	Unique
46	SPTY2D1	19	13	6.89E+08	Unique
47	TFIP11	12	10	6.88E+08	Unique
48	MED24	24	11	6.55E+08	Unique
49	TEX30	7	6	6.44E+08	Unique
50	BPTF	18	18	6.31E+08	Unique
51	GNL3L	19	13	6.30E+08	Unique
52	DOCK7	21	13	6.29E+08	Unique
53	MAP7	4	2	6.17E+08	Unique
54	REPIN1	9	3	5.79E+08	Unique
55	DYRK1A	19	12	5.65E+08	Unique
56	FBRS	10	6	5.62E+08	Unique
57	CCNT2	11	7	5.61E+08	Unique
58	RBM15B	24	12	5.48E+08	Unique
59	RBBP5	7	5	5.16E+08	Unique
60	PAPOLG	15	13	4.97E+08	Unique
61	MED9	4	4	4.80E+08	Unique
62	SOGA2	33	7	4.65E+08	Unique
63	TAF6	25	9	4.63E+08	Unique
64	PDZD11	4	3	4.60E+08	Unique
65	CDYL	9	9	4.46E+08	Unique
66	ASH2L	12	9	4.46E+08	Unique
67	MED13	23	13	4.37E+08	Unique
68	CXXC1	15	8	4.32E+08	Unique
69	RBMX2	8	4	4.31E+08	Unique
70	USP34	24	14	4.26E+08	Unique
71	PLEKHA5	12	9	4.22E+08	Unique
72	MED8	3	3	4.15E+08	Unique
73	TAF8	10	5	4.15E+08	Unique
74	SOWAHC	9	8	4.12E+08	Unique
75	AKAP17A	9	8	4.09E+08	Unique
76	MED16	16	5	3.96E+08	Unique
77	VPS72	6	6	3.93E+08	Unique
78	TAF9	7	5	3.92E+08	Unique
79	MED27	9	6	3.89E+08	Unique
80	RBM7	6	6	3.68E+08	Unique
81	TRMT2A	15	11	3.66E+08	Unique
82	ZBTB40	18	8	3.65E+08	Unique
83	CACTIN	8	8	3.58E+08	Unique
84	TAF5	26	7	3.56E+08	Unique
85	PDCD7	8	5	3.39E+08	Unique
86	GPATCH8	16	7	3.38E+08	Unique
87	BMS1	21	8	3.23E+08	Unique
88	MLLT1	13	5	3.08E+08	Unique
89	KCTD15	7	6	3.02E+08	Unique
90	MED7	6	5	2.97E+08	Unique
91	RPS6	5	3	2.96E+08	Unique
92	ZNF638	11	9	2.88E+08	Unique
93	ZNF131	5	4	2.85E+08	Unique
94	CDK8;CDK19	6	5	2.78E+08	Unique
95	MED23	15	9	2.76E+08	Unique
96	RPL13	8	4	2.73E+08	Unique
97	FAM83G	10	2	2.71E+08	Unique
98	MIF4GD	5	4	2.67E+08	Unique
99	ZBTB14	9	5	2.64E+08	Unique
100	ZBTB43	8	4	2.63E+08	Unique
101	SPATS2L	12	3	2.62E+08	Unique
102	FIZ1	9	5	2.62E+08	Unique
103	DNTTIP1	6	5	2.57E+08	Unique
104	SETD1B	22	8	2.57E+08	Unique
105	ZEB1	14	7	2.57E+08	Unique
106	MED6	5	3	2.50E+08	Unique
107	WDR24	16	4	2.49E+08	Unique
108	ZNF8	12	5	2.36E+08	Unique
109	WDR59	19	7	2.18E+08	Unique
110	SNRNP48	11	4	2.16E+08	Unique
111	SEMA3A	11	4	2.15E+08	Unique
112	RPS24	4	3	2.14E+08	Unique
113	ZRSR2;ZRSR1	12	4	1.97E+08	Unique
114	ZNF316	12	4	1.96E+08	Unique
115	LRCH1	9	4	1.91E+08	Unique
116	INO80	22	4	1.86E+08	Unique
117	ZNF692	12	4	1.84E+08	Unique
118	WWP2	10	4	1.79E+08	Unique
119	ACTR5	16	7	1.73E+08	Unique
120	CBX4	12	5	1.72E+08	Unique
121	KCTD1	3	3	1.71E+08	Unique
122	WHSC1L1	7	4	1.70E+08	Unique
123	NUMA1	19	5	1.66E+08	Unique
124	SEMA3C	8	4	1.59E+08	Unique
125	PIP5K1C	5	2	1.55E+08	Unique
126	SNRNP27	4	3	1.53E+08	Unique
127	MED29	4	3	1.51E+08	Unique
128	FTSJ1	3	3	1.43E+08	Unique

Continue Appendix 4:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity beads
129	ZNF574	12	6	1.39E+08	Unique
130	TRPT1	4	2	1.39E+08	Unique
131	INO80B	10	2	1.34E+08	Unique
132	LAMA5	28	5	1.34E+08	Unique
133	RPL18	3	3	1.33E+08	Unique
134	CAAP1	8	4	1.33E+08	Unique
135	BAZ1B	10	5	1.33E+08	Unique
136	RCL1	8	3	1.31E+08	Unique
137	TAF1;TAF1L	33	6	1.28E+08	Unique
138	NSMAF	8	5	1.28E+08	Unique
139	GPBP1L1	7	3	1.26E+08	Unique
140	CITED2	2	2	1.25E+08	Unique
141	TAF3	8	3	1.22E+08	Unique
142	UCHL5	11	3	1.21E+08	Unique
143	ASB7	10	4	1.21E+08	Unique
144	ANKRD11	23	3	1.20E+08	Unique
145	SMURF2	7	3	1.16E+08	Unique
146	POP4	5	4	1.15E+08	Unique
147	RPS13	11	2	1.15E+08	Unique
148	TAF10	4	2	1.01E+08	Unique
149	SP4	3	2	9.98E+07	Unique
150	MED19	5	2	9.83E+07	Unique
151	ZNF76	3	2	9.45E+07	Unique
152	INTS1	7	4	9.30E+07	Unique
153	REST	6	2	8.66E+07	Unique
154	ZFP64	3	2	8.55E+07	Unique
155	FAM76A	5	2	8.52E+07	Unique
156	E4F1	5	2	8.38E+07	Unique
157	MSL3;MSL3P	5	3	8.33E+07	Unique
158	PTS	2	2	8.29E+07	Unique
159	DCAF13	13	3	8.18E+07	Unique
160	MED30	3	2	7.97E+07	Unique
161	ZNF746	3	2	7.96E+07	Unique
162	GLCCI1	6	3	7.78E+07	Unique
163	USP42	11	2	7.73E+07	Unique
164	CDK13	5	3	7.69E+07	Unique
165	AXIN1	4	2	7.66E+07	Unique
166	MED10	5	2	7.61E+07	Unique
167	TAF12	5	2	7.55E+07	Unique
168	TXNL4A	4	2	7.47E+07	Unique
169	GRK5	4	2	7.42E+07	Unique
170	RPL8	2	2	7.04E+07	Unique
171	UTP14A	5	2	6.91E+07	Unique
172	ZBTB17	4	2	6.80E+07	Unique
173	ZKSCAN1	5	2	6.70E+07	Unique
174	FOXL2	3	2	6.51E+07	Unique
175	PRPF31	11	2	6.42E+07	Unique
176	PHC3	7	2	6.32E+07	Unique
177	KLF4	3	2	5.79E+07	Unique
178	ZNF239	4	2	5.77E+07	Unique
179	MCRS1	5	3	5.71E+07	Unique
180	RPS29	4	2	5.67E+07	Unique
181	DUSP11	6	2	5.65E+07	Unique
182	RPL14	3	2	5.65E+07	Unique
183	ARID5B	5	3	5.56E+07	Unique
184	WARS2	5	2	5.56E+07	Unique
185	TCF7L2;LEF1	3	2	5.43E+07	Unique
186	MUS81	6	2	5.11E+07	Unique
187	ZIC2	4	2	5.08E+07	Unique
188	ZNF668	5	2	5.04E+07	Unique
189	DSP	23	2	5.03E+07	Unique
190	PHGDH	7	2	4.88E+07	Unique
191	POLRMT	7	2	4.87E+07	Unique
192	TAF7	11	2	4.80E+07	Unique

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity beads
193	ZMAT5	3	2	4.57E+07	Unique
194	NFRKB	25	2	4.35E+07	Unique
195	ZNF219	4	2	4.12E+07	Unique
196	SNRNP35	3	2	3.91E+07	Unique
197	USP46;USP1	5	2	3.69E+07	Unique
198	SENP6	3	2	3.46E+07	Unique
199	BZW1	4	2	1.32E+07	Unique
200	PRPF6	50	47	1.69E+10	647.2
201	TBL3	21	20	2.00E+09	610.6
202	ZNF629	25	10	8.55E+08	498.5
203	CHD8	85	58	8.84E+09	426.7
204	MED20	7	7	5.85E+08	220.9
205	SNRNP40	21	19	9.53E+09	119.4
206	DDX23	34	30	1.21E+10	112.3
207	WDR3	24	23	2.02E+09	112.2
208	NCOA5	12	11	9.47E+08	87.5
209	CDK9	15	14	2.70E+09	81.1
210	RBM25	36	32	2.39E+10	74.8
211	SMARCD1	19	17	2.81E+09	71.6
212	RSBN1	25	16	2.22E+09	71.6
213	RPL24	6	5	1.86E+09	62.3
214	RBM17	22	21	5.87E+09	61.6
215	SFSWAP	23	19	1.90E+09	59.2
216	NIPBL	44	28	1.48E+09	56.5
217	ARID1B	47	19	8.99E+08	53.9
218	BUD13	16	11	9.40E+08	51.1
219	RPS11	14	8	1.22E+09	47.4
220	THOC2	37	29	4.31E+09	43.7
221	WDR6	23	19	9.09E+08	43.6
222	CTR9	34	33	7.68E+09	41.5
223	CTIF	10	7	3.34E+08	40.6
224	RPL27A	8	5	8.50E+08	38.9
225	CCNT1	14	9	2.02E+09	38.9
226	ZC3H13	18	12	9.73E+08	35.9
227	FAM120B	33	21	1.16E+09	34.8
228	PRPF8	116	107	4.40E+10	34.5
229	CBLL1	13	12	1.11E+09	33.1
230	SF3B1	66	64	9.23E+10	31.2
231	CPSF2	48	34	1.23E+10	30.9
232	SNRPA1	11	9	1.67E+09	29.2
233	WTAP	17	15	2.72E+09	29.1
234	CBL	27	18	1.11E+09	27.4
235	SMARCD2	11	9	6.59E+08	27.0
236	GNL3	21	14	1.45E+09	26.5
237	WDR20	11	5	3.29E+08	25.1
238	CDK11B	6	4	1.65E+10	24.1
239	THOC3	12	10	1.29E+09	23.6
240	PHF5A	8	8	7.71E+09	22.7
241	SF3B2	53	50	8.46E+10	22.2
242	NFX1	29	16	7.09E+08	21.8
243	MAP4K4	10	8	8.63E+08	20.3
244	LARP1	16	16	1.55E+09	19.5
245	MED26	12	8	5.78E+08	19.2
246	LARP7	20	16	1.10E+09	18.2
247	ATXN2	21	15	1.65E+09	17.7
248	PAF1	23	22	4.12E+09	17.6
249	SART1	61	52	3.42E+10	17.6
250	PHRF1	13	8	5.67E+08	17.5
251	RPS8	11	7	7.31E+08	17.1
252	TAF4	26	13	8.87E+08	16.1
253	SNRNP200	118	105	4.63E+10	15.8
254	ZNF148	24	20	2.50E+09	15.3
255	CPSF1	64	53	2.17E+10	15.2
256	ATN1	16	10	2.02E+09	15.0

Continue Appendix 4:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity beads
257	ZFP91	12	8	2.51E+09	14.9
258	ZNF24	16	12	2.35E+09	14.8
259	TAF2	49	26	3.11E+09	14.5
260	RSBN1L	32	24	3.36E+09	14.4
261	PPIP5K2	22	18	9.38E+08	14.1
262	RPL23A	12	8	2.43E+09	14.0
263	SMARCA4	27	24	5.09E+09	12.9
264	IRS2	30	21	2.42E+09	12.8
265	CDK12	32	23	9.28E+09	12.7
266	ELMSAN1	20	17	1.49E+09	12.1
267	EFTUD2	56	52	2.89E+10	12.0
268	TSR1	39	29	1.11E+10	11.9
269	NUFIP2	17	12	1.82E+09	11.7
270	ZBTB7A	9	7	1.39E+09	11.6
271	ZNF691	8	5	3.93E+08	11.3
272	KLF16	7	3	4.16E+08	11.2
273	TOP1	63	45	6.89E+10	11.1
274	SF3B14	6	6	5.24E+09	11.0
275	MSL1	12	8	4.94E+08	11.0
276	MED1	50	39	5.02E+09	10.8
277	SF3A1	46	43	5.67E+10	10.8
278	SF3A3	27	25	2.84E+10	10.6
279	FTSJD2	42	38	7.28E+09	10.3
280	CIR1	8	5	4.05E+08	10.1
281	PCF11	41	33	7.58E+09	10.1
282	TOX4	17	14	3.55E+09	9.9
283	CHAMP1	47	42	8.81E+09	9.8
284	PRPF40A	2	2	2.79E+10	9.4
285	RPL27	4	4	1.40E+08	9.0
286	PPP1R10	42	28	8.44E+09	8.9
287	DCD	4	2	2.24E+08	8.9
288	KANSL1	21	11	3.31E+08	8.7
289	MEPCE	25	17	3.02E+09	8.3
290	PHF3	59	35	4.05E+09	8.3
291	RPL19	6	3	2.74E+08	8.3
292	BRD4	41	34	2.01E+10	8.3
293	CALML5	5	2	3.73E+08	8.2
294	CTCF	17	15	2.13E+09	8.1
295	ZNF768	11	6	2.13E+08	7.9
296	PELO	13	9	1.26E+09	7.4
297	CCNK	17	12	3.81E+09	7.4
298	CPSF3	35	29	8.77E+09	7.2
299	ZNF609	25	20	1.89E+09	7.2
300	MEAF6	7	5	1.34E+09	7.1
301	SF3A2	12	10	7.40E+09	7.1
302	SMARCC1	31	23	7.03E+09	7.1
303	SF3B5	5	4	3.09E+09	6.9
304	RNF2	5	2	1.80E+08	6.9
305	ZNF593	6	4	2.26E+08	6.8
306	VIM	16	6	2.68E+08	6.7
307	GATAD1	5	4	2.73E+08	6.6
308	MECP2	19	12	1.02E+09	6.5
309	PRPF38B	14	11	5.29E+09	6.5
310	SAP30BP	23	20	9.32E+09	6.4
311	ALKBH2	9	5	4.68E+08	6.3
312	THAP11	6	5	4.21E+08	6.3
313	WAC	17	14	4.41E+09	6.2
314	DPY30	3	3	4.88E+08	6.1
315	VEZF1	4	2	2.88E+08	6.0
316	CSNK1D	6	5	1.28E+09	5.9
317	FBXO21	5	5	2.01E+08	5.8
318	PIP5K1A	8	6	9.57E+08	5.7
319	MFAP1	22	18	6.82E+09	5.6
320	POP1	25	24	2.40E+09	5.6
321	SAFB2	30	23	4.36E+10	5.5
322	WDR33	53	41	1.34E+10	5.4
323	WDR82	22	19	7.69E+09	5.3
324	TRA2A	4	4	1.38E+08	5.1
325	WDR5	14	12	1.81E+09	5.1
326	SUB1	10	8	1.54E+09	5.0
327	EEFSEC	28	23	4.61E+09	5.0
328	RPS3A	17	11	1.63E+09	4.7
329	KLF5	5	5	1.73E+09	4.7
330	ACTL6A	15	11	2.80E+09	4.7
331	PHF6	6	4	1.14E+10	4.6
332	DHX15	51	45	8.05E+10	4.5
333	LUC7L	10	7	1.60E+09	4.4
334	SP1	10	8	3.46E+09	4.3
335	CPSF6	21	18	5.56E+10	4.2
336	FOXJ3	6	4	4.01E+08	4.2
337	RPL15	4	3	1.34E+08	4.1
338	AFF4	32	21	1.70E+09	4.0
339	DPF2	13	10	7.91E+08	4.0
340	MMTAG2	13	7	6.57E+08	4.0
341	DDX42	9	5	3.78E+10	3.8
342	FBL	11	9	1.94E+09	3.8
343	FIP1L1	21	15	1.50E+10	3.7
344	NUDT21	23	18	4.39E+10	3.6
345	RBM33	29	22	5.96E+09	3.6
346	LUC7L2	14	12	3.30E+10	3.6
347	SLTM	32	23	9.55E+09	3.6
348	RBM27	34	21	2.90E+09	3.5
349	MAZ	11	6	2.97E+09	3.5
350	ZNF787	11	8	2.47E+09	3.5
351	THOC7	12	8	7.52E+08	3.4
352	LSM12	5	4	7.34E+08	3.3
353	CPSF7	21	18	3.14E+10	3.3
354	NONO	51	37	4.37E+11	3.3
355	ZGPAT	16	13	8.42E+08	3.3
356	SNIP1	19	13	2.21E+09	3.2
357	RING1	5	3	6.33E+08	3.0
358	MRPS23	3	2	7.96E+07	2.9
359	DDX46	15	4	1.04E+08	2.9
360	YY1	13	10	1.39E+10	2.9
361	RSRC2	13	10	6.43E+09	2.7
362	SFPQ	60	44	4.35E+11	2.7
363	FOXC2	6	3	7.07E+07	2.7
364	PPIL4	27	22	8.59E+09	2.6
365	RPP40	10	8	1.18E+09	2.6
366	DCAF7	13	12	1.18E+09	2.6
367	YY2	2	2	7.95E+08	2.4
368	RUVBL2	30	16	2.03E+09	2.4
369	KLF13	6	2	1.43E+08	2.3
370	FAM76B	8	2	2.90E+08	2.3
371	MED31	5	2	7.71E+07	2.3
372	TMA16	10	6	7.06E+08	2.3
373	TCEB3	18	9	5.80E+08	2.3
374	NUSAP1	22	10	1.01E+09	2.1

Appendix 5

A total of 374 proteins bound to SCR of ICP22 sorted in rank based on uniqueness and fold enrichment

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
1	PRPF4B	52	46	2.03E+10	Unique
2	ICP22-SCR	7	2	1.38E+10	Unique
3	RPL17	11	11	8.24E+09	Unique
4	RPS4X	20	20	6.61E+09	Unique
5	ATM	81	75	4.86E+09	Unique
6	RPL10	11	11	4.67E+09	Unique
7	CCNL1	13	10	3.40E+09	Unique
8	TAF6	25	24	3.35E+09	Unique
9	RPS23	6	6	3.19E+09	Unique
10	CHD1	39	34	3.14E+09	Unique
11	TAF5	26	24	3.02E+09	Unique
12	SYMPK	41	34	2.70E+09	Unique
13	GTPBP4	26	25	2.62E+09	Unique
14	DOCK9	58	46	2.53E+09	Unique
15	CLK3	15	15	2.46E+09	Unique
16	LSG1	30	25	2.28E+09	Unique
17	SOGA2	33	29	2.24E+09	Unique
18	THOC1	35	28	2.23E+09	Unique
19	RPS13	11	10	2.01E+09	Unique
20	CHD2	29	26	1.91E+09	Unique
21	TAF9	7	7	1.88E+09	Unique
22	RBBP6	45	27	1.85E+09	Unique
23	TAF8	10	10	1.73E+09	Unique
24	MAP7	4	4	1.67E+09	Unique
25	TAF1;TAF1L	33	30	1.63E+09	Unique
26	PWP1	16	14	1.59E+09	Unique
27	MED14	34	20	1.53E+09	Unique
28	CLK1	14	12	1.51E+09	Unique
29	PRPF38A	11	9	1.47E+09	Unique
30	MIOS	29	23	1.45E+09	Unique
31	NFRKB	25	24	1.44E+09	Unique
32	RBM15B	24	20	1.42E+09	Unique
33	THOC5	20	17	1.34E+09	Unique
34	RPS24	4	4	1.22E+09	Unique
35	CCNL2	13	8	1.16E+09	Unique
36	INO80	22	21	1.13E+09	Unique
37	RERE	22	15	1.10E+09	Unique
38	ACTR5	16	14	1.06E+09	Unique
39	MED4	11	8	1.04E+09	Unique
40	CHERP	28	14	9.79E+08	Unique
41	RPS29	4	4	9.71E+08	Unique
42	DYRK1A	19	17	9.30E+08	Unique
43	MED17	20	15	9.19E+08	Unique
44	RPL13	8	8	9.10E+08	Unique
45	TAF10	4	4	8.79E+08	Unique
46	THOC6	14	9	8.59E+08	Unique
47	CBX4	12	11	8.25E+08	Unique
48	CCNT2	11	10	8.21E+08	Unique
49	REPIN1	9	8	8.06E+08	Unique
50	GNL3L	19	17	7.81E+08	Unique
51	ANKRD11	23	20	7.69E+08	Unique
52	MED12	41	23	7.54E+08	Unique
53	LAMA5	28	21	7.20E+08	Unique
54	MLLT1	13	11	7.18E+08	Unique
55	SPTY2D1	19	12	6.84E+08	Unique
56	INO80B	10	10	6.78E+08	Unique
57	ZNF692	12	11	6.35E+08	Unique
58	SETD1B	22	18	6.34E+08	Unique
59	ZNF316	12	11	6.32E+08	Unique
60	TAF7	11	11	5.99E+08	Unique
61	POGZ	26	17	5.83E+08	Unique
62	MED8	3	3	5.69E+08	Unique
63	PRPF31	11	11	5.66E+08	Unique
64	CXXC1	15	12	5.64E+08	Unique
65	FIZ1	9	9	5.36E+08	Unique
66	RPS6	5	5	5.30E+08	Unique
67	MED9	4	4	5.30E+08	Unique
68	NUMA1	19	14	5.08E+08	Unique
69	RBMX2	8	8	4.98E+08	Unique
70	FAM83G	10	8	4.98E+08	Unique
71	SPATS2L	12	9	4.97E+08	Unique
72	TEX30	7	6	4.92E+08	Unique
73	ZNF8	12	10	4.84E+08	Unique
74	SNRNP48	11	9	4.76E+08	Unique
75	MED19	5	5	4.75E+08	Unique
76	BMS1	21	13	4.68E+08	Unique
77	UCHL5	11	9	4.65E+08	Unique
78	WDR59	19	16	4.34E+08	Unique
79	MED13	23	16	4.29E+08	Unique
80	TAF3	8	8	4.28E+08	Unique
81	USP42	11	10	4.21E+08	Unique
82	WDR24	16	14	4.02E+08	Unique
83	DCAF13	13	11	3.93E+08	Unique
84	PPIG	9	6	3.77E+08	Unique
85	YTHDC1	16	6	3.70E+08	Unique
86	ZBTB14	9	6	3.52E+08	Unique
87	E4F1	5	4	3.45E+08	Unique
88	PHC3	7	7	3.42E+08	Unique
89	ASH2L	12	11	3.39E+08	Unique
90	ZNF574	12	11	3.32E+08	Unique
91	MED10	5	5	3.21E+08	Unique
92	MED24	24	15	3.12E+08	Unique
93	SEMA3A	11	10	3.09E+08	Unique
94	USP34	24	14	3.07E+08	Unique
95	MED7	6	6	3.06E+08	Unique
96	KIAA1429	28	11	3.04E+08	Unique
97	LRCH1	9	6	2.96E+08	Unique
98	DUSP11	6	6	2.87E+08	Unique
99	WWP2	10	7	2.84E+08	Unique
100	SNRPB2	6	4	2.59E+08	Unique
101	TRMT2A	15	11	2.58E+08	Unique
102	ZRSR2;ZRSR1	12	9	2.57E+08	Unique
103	PDCD7	8	5	2.52E+08	Unique
104	HTATSF1	20	3	2.39E+08	Unique
105	PLEKHA5	12	7	2.39E+08	Unique
106	ZEB1	14	10	2.39E+08	Unique
107	MED29	4	4	2.39E+08	Unique
108	KCTD15	7	5	2.38E+08	Unique
109	MED27	9	7	2.37E+08	Unique
110	ZCCHC8	17	6	2.35E+08	Unique
111	SMARCA5	34	10	2.23E+08	Unique
112	FBRS	10	5	2.19E+08	Unique
113	ZBTB40	18	7	2.16E+08	Unique
114	TXNL4A	4	4	2.14E+08	Unique
115	TAF12	5	4	2.14E+08	Unique
116	SP4	3	3	2.06E+08	Unique
117	RBBP5	7	5	1.93E+08	Unique
118	SOWAHC	9	7	1.92E+08	Unique
119	SNRNP27	4	4	1.86E+08	Unique
120	REST	6	6	1.83E+08	Unique
121	WARS2	5	5	1.81E+08	Unique
122	ASB7	10	6	1.79E+08	Unique
123	KLF4	3	3	1.68E+08	Unique
124	GPATCH11	6	5	1.61E+08	Unique
125	DOCK7	21	4	1.52E+08	Unique
126	ZBTB17	4	4	1.49E+08	Unique
127	MED16	16	6	1.48E+08	Unique
128	TRRAP	37	8	1.47E+08	Unique

Continue Appendix 5:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
129	PHGDH	7	5	1.42E+08	Unique
130	ZNF746	3	2	1.42E+08	Unique
131	MED6	5	3	1.40E+08	Unique
132	RBM7	6	4	1.35E+08	Unique
133	AKAP17A	9	5	1.35E+08	Unique
134	SEMA3C	8	5	1.35E+08	Unique
135	ZFP64	3	3	1.35E+08	Unique
136	PDZD11	4	2	1.32E+08	Unique
137	CDK8;CDK19	6	4	1.26E+08	Unique
138	GPBP1L1	7	4	1.24E+08	Unique
139	CITED2	2	2	1.23E+08	Unique
140	VPS72	6	4	1.21E+08	Unique
141	ZBTB43	8	4	1.20E+08	Unique
142	SMURF2	7	4	1.19E+08	Unique
143	PIP5K1C	5	2	1.18E+08	Unique
144	POLRMT	7	5	1.18E+08	Unique
145	BZW1	4	4	1.17E+08	Unique
146	MSL3;MSL3P	5	5	1.15E+08	Unique
147	FTSJ1	3	3	1.11E+08	Unique
148	ZNF76	3	2	1.11E+08	Unique
149	NSMAF	8	4	1.09E+08	Unique
150	GLCCI1	6	3	1.09E+08	Unique
151	FOXL2	3	2	1.09E+08	Unique
152	CDYL	9	6	1.07E+08	Unique
153	MIF4GD	5	4	1.05E+08	Unique
154	FAM76A	5	2	1.04E+08	Unique
155	GPATCH8	16	5	1.01E+08	Unique
156	AXIN1	4	4	9.83E+07	Unique
157	MED23	15	6	9.66E+07	Unique
158	SENP6	3	2	9.58E+07	Unique
159	RPL8	2	2	9.10E+07	Unique
160	KCTD1	3	3	9.07E+07	Unique
161	ZNF131	5	4	8.56E+07	Unique
162	ZNF668	5	4	8.49E+07	Unique
163	DNTTIP1	6	4	8.48E+07	Unique
164	ZIC2	4	3	8.02E+07	Unique
165	PTS	2	2	7.86E+07	Unique
166	EHMT2	19	4	7.80E+07	Unique
167	MUS81	6	3	7.61E+07	Unique
168	ZNF219	4	4	7.59E+07	Unique
169	BAZ1B	10	3	7.34E+07	Unique
170	WDR36	24	4	7.17E+07	Unique
171	DSP	23	5	7.01E+07	Unique
172	TFIP11	12	3	6.93E+07	Unique
173	RPL18	3	3	6.83E+07	Unique
174	RCL1	8	3	6.77E+07	Unique
175	CAAP1	8	4	6.76E+07	Unique
176	MED30	3	3	6.49E+07	Unique
177	ZNF239	4	4	6.47E+07	Unique
178	UTP14A	5	4	6.20E+07	Unique
179	RPL14	3	3	6.07E+07	Unique
180	MCRS1	5	2	5.97E+07	Unique
181	ZBTB7B	9	2	5.59E+07	Unique
182	MTM1	15	2	5.54E+07	Unique
183	TRPT1	4	4	5.45E+07	Unique
184	TCF7L2;LEF1	3	2	5.34E+07	Unique
185	ZMAT5	3	2	5.08E+07	Unique
186	CACTIN	8	2	4.04E+07	Unique
187	ZNF638	11	2	3.89E+07	Unique
188	USP46;USP1	5	2	3.73E+07	Unique
189	BPTF	18	3	3.52E+07	Unique
190	SNRNP35	3	2	3.49E+07	Unique
191	CDK13	5	2	3.41E+07	Unique
192	ARID5B	5	3	3.35E+07	Unique

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
193	PAPOLG	15	2	2.82E+07	Unique
194	POP4	5	2	2.68E+07	Unique
195	GRK5	4	3	2.63E+07	Unique
196	ZKSCAN1	5	2	2.60E+07	Unique
197	WHSC1L1	7	2	2.32E+07	Unique
198	PWP2	22	3	1.98E+07	Unique
199	INTS1	7	2	1.59E+07	Unique
200	ZNF629	25	20	2.59E+09	1507.8
201	CHD8	85	79	1.55E+10	748.0
202	PRPF6	50	47	1.15E+10	439.5
203	RPS11	14	14	6.20E+09	240.5
204	RPL27A	8	8	4.76E+09	217.7
205	ARID1B	47	42	2.43E+09	145.5
206	CDK9	15	15	4.57E+09	137.6
207	RSBN1	25	24	3.28E+09	106.0
208	CCNT1	14	14	4.94E+09	95.2
209	MED20	7	4	2.44E+08	92.2
210	RPS8	11	11	3.79E+09	88.5
211	SNRNP40	21	19	6.00E+09	75.2
212	RPL24	6	6	2.11E+09	70.6
213	TAF4	26	25	3.13E+09	56.9
214	RPL23A	12	12	7.65E+09	44.3
215	FAM120B	33	28	1.45E+09	43.7
216	RPL19	6	5	1.16E+09	34.9
217	NFX1	29	21	1.10E+09	33.7
218	TBL3	21	5	1.09E+08	33.3
219	CPSF2	48	43	1.29E+10	32.3
220	CDK11B	6	6	2.17E+10	31.8
221	BUD13	16	14	5.67E+08	30.8
222	WDR20	11	6	3.67E+08	28.0
223	TAF2	49	42	5.95E+09	27.8
224	CBL	27	25	1.10E+09	27.3
225	DDX23	34	26	2.83E+09	26.3
226	RBM25	36	29	8.39E+09	26.2
227	NUFIP2	17	16	3.76E+09	24.3
228	SART1	61	55	4.56E+10	23.4
229	GNL3	21	17	1.28E+09	23.4
230	THOC2	37	29	2.27E+09	23.0
231	CDK12	32	29	1.67E+10	22.9
232	TOX4	17	17	8.12E+09	22.6
233	ZFP91	12	10	3.72E+09	22.0
234	ZNF148	24	23	3.57E+09	21.8
235	CTIF	10	5	1.71E+08	20.8
236	PPP1R10	42	40	1.80E+10	18.9
237	PRPF8	116	103	2.36E+10	18.5
238	KLF16	7	6	6.67E+08	18.0
239	ZNF593	6	6	5.73E+08	17.4
240	TSR1	39	37	1.62E+10	17.3
241	CBLL1	13	10	5.75E+08	17.2
242	ATN1	16	15	2.31E+09	17.1
243	MED26	12	10	5.13E+08	17.0
244	SMARCD1	19	12	6.63E+08	16.9
245	WDR6	23	15	3.49E+08	16.7
246	ZNF768	11	8	4.49E+08	16.6
247	RSBN1L	32	30	3.82E+09	16.4
248	THOC3	12	9	8.82E+08	16.2
249	TOP1	63	59	1.00E+11	16.1
250	MSL1	12	9	6.89E+08	15.3
251	RPS3A	17	16	5.14E+09	14.8
252	GATAD1	5	4	6.06E+08	14.7
253	ZNF24	16	15	2.28E+09	14.4
254	ZNF691	8	6	4.85E+08	13.9
255	CPSF1	64	60	1.90E+10	13.4
256	MEPCE	25	22	4.82E+09	13.3

Continue Appendix 5:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
257	SMARCD2	11	6	3.23E+08	13.2
258	SFSWAP	23	9	4.22E+08	13.1
259	LARP7	20	18	7.54E+08	12.4
260	PHF3	59	47	5.62E+09	11.5
261	ZC3H13	18	9	3.07E+08	11.3
262	SNRPA1	11	7	6.35E+08	11.1
263	VIM	16	12	4.35E+08	10.9
264	MED1	50	42	4.74E+09	10.2
265	PHRF1	13	8	3.17E+08	9.8
266	CCNK	17	13	4.75E+09	9.2
267	ALKBH2	9	6	6.77E+08	9.1
268	SF3B2	53	45	3.44E+10	9.1
269	NUSAP1	22	18	4.30E+09	9.0
270	CIR1	8	6	3.56E+08	8.9
271	MAZ	11	11	7.46E+09	8.7
272	ZBTB7A	9	8	1.03E+09	8.6
273	MECP2	19	16	1.34E+09	8.5
274	MED31	5	4	2.77E+08	8.2
275	THAP11	6	5	5.44E+08	8.1
276	SF3B1	66	56	2.39E+10	8.1
277	ZNF609	25	23	2.14E+09	8.1
278	SAP30BP	23	23	1.18E+10	8.1
279	SNRNP200	118	98	2.36E+10	8.1
280	FOXC2	6	6	2.11E+08	8.0
281	DDX46	15	12	2.84E+08	7.9
282	LARP1	16	11	5.87E+08	7.4
283	PIP5K1A	8	6	1.17E+09	7.0
284	RPL15	4	3	2.27E+08	6.9
285	VEZF1	4	2	3.32E+08	6.8
286	EFTUD2	56	48	1.64E+10	6.8
287	RBM17	22	14	6.50E+08	6.8
288	FOXJ3	6	5	6.38E+08	6.6
289	FAM76B	8	8	8.19E+08	6.6
290	BRD4	41	36	1.60E+10	6.6
291	FTSJD2	42	34	4.55E+09	6.4
292	DCD	4	2	1.59E+08	6.3
293	WDR82	22	20	9.15E+09	6.3
294	EEFSEC	28	25	5.75E+09	6.3
295	WTAP	17	12	5.73E+08	6.1
296	RUVBL2	30	25	5.10E+09	6.0
297	CPSF3	35	30	7.19E+09	5.9
298	SP1	10	9	4.73E+09	5.9
299	PELO	13	12	9.80E+08	5.8
300	FBXO21	5	5	1.99E+08	5.7
301	ATXN2	21	12	5.27E+08	5.7
302	SUB1	10	8	1.69E+09	5.5
303	DDX42	9	9	5.43E+10	5.5
304	AFF4	32	28	2.27E+09	5.4
305	DPF2	13	13	1.05E+09	5.4
306	NCOA5	12	3	5.75E+07	5.3
307	SMARCA4	27	17	2.09E+09	5.3
308	RBM33	29	27	8.61E+09	5.2
309	CHAMP1	47	40	4.50E+09	5.0
310	WDR5	14	12	1.76E+09	5.0
311	CTR9	34	16	9.05E+08	4.9
312	PHF6	6	5	1.15E+10	4.7
313	WDR33	53	43	1.15E+10	4.6
314	YY1	13	13	2.23E+10	4.6
315	TCEB3	18	15	1.18E+09	4.6
316	NONO	51	48	6.17E+11	4.6
317	SF3B14	6	6	2.16E+09	4.5
318	FBL	11	10	2.24E+09	4.3
319	FIP1L1	21	19	1.70E+10	4.2
320	PHF5A	8	6	1.42E+09	4.2
321	RPL27	4	3	6.47E+07	4.1
322	MFAP1	22	19	5.02E+09	4.1
323	KLF5	5	5	1.45E+09	3.9
324	PPIP5K2	22	10	2.60E+08	3.9
325	SLTM	32	27	1.02E+10	3.8
326	CTCF	17	12	1.00E+09	3.8
327	TMA16	10	8	1.18E+09	3.8
328	RBM27	34	27	3.11E+09	3.8
329	LUC7L2	14	12	3.41E+10	3.7
330	ACTL6A	15	13	2.23E+09	3.7
331	THOC7	12	11	8.20E+08	3.7
332	YY2	2	2	1.22E+09	3.7
333	PRPF40A	2	2	1.09E+10	3.7
334	SF3A3	27	20	9.65E+09	3.6
335	MMTAG2	13	8	5.65E+08	3.5
336	SFPQ	60	55	5.49E+11	3.5
337	PPIL4	27	24	1.12E+10	3.4
338	LUC7L	10	8	1.25E+09	3.4
339	SF3A1	46	39	1.75E+10	3.3
340	MEAF6	7	6	6.01E+08	3.2
341	SNIP1	19	17	2.22E+09	3.2
342	PCF11	41	26	2.34E+09	3.1
343	SMARCC1	31	27	3.06E+09	3.1
344	ZNF787	11	11	2.20E+09	3.1
345	PRPF38B	14	12	2.49E+09	3.1
346	MAP4K4	10	3	1.28E+08	3.0
347	MRPS23	3	3	8.14E+07	3.0
348	RNF2	5	3	7.57E+07	2.9
349	NUDT21	23	20	3.48E+10	2.9
350	IRS2	30	15	5.42E+08	2.9
351	NIPBL	44	5	7.54E+07	2.9
352	WDR3	24	2	5.14E+07	2.9
353	CALML5	5	2	1.29E+08	2.8
354	SF3B5	5	5	1.24E+09	2.8
355	CPSF7	21	17	2.55E+10	2.7
356	DPY30	3	3	2.14E+08	2.7
357	TRA2A	4	2	6.93E+07	2.6
358	KLF13	6	3	1.58E+08	2.6
359	ELMSAN1	20	10	3.14E+08	2.5
360	SAFB2	30	23	2.00E+10	2.5
361	POP1	25	18	1.04E+09	2.4
362	DHX15	51	47	4.33E+10	2.4
363	CPSF6	21	20	3.15E+10	2.4
364	ZGPAT	16	12	6.19E+08	2.4
365	RSRC2	13	11	5.48E+09	2.3
366	DCAF7	13	13	1.07E+09	2.3
367	RING1	5	4	4.58E+08	2.2
368	CSNK1D	6	5	4.65E+08	2.1
369	SF3A2	12	7	2.22E+09	2.1
370	LSM12	5	5	4.69E+08	2.1
371	RPP40	10	8	9.51E+08	2.1
372	KANSL1	21	3	7.82E+07	2.1
373	PAF1	23	12	4.76E+08	2.0
374	WAC	17	12	1.42E+09	2.0

Appendix 6

A total of 292 proteins exclusively bound to full length of ICP22 sorted in alphabetical order

No	Gene names
1	ABCF2
2	ABL1
3	ADRBK1
4	AMMECR1
5	ANKHD1
6	AQR
7	ARHGEF5
8	ARHGEF7
9	ARID2
10	ARID4A
11	ARID4B
12	ASCC1
13	ASCC3
14	ATXN2L
15	AURKA
16	BAIAP2L1
17	BAP18
18	BAZ1A
19	BBX
20	BCKDK
21	BCL7C
22	BHLHE40
23	BLM
24	BOD1L1
25	BRD7
26	BRMS1
27	BRMS1L
28	C16ORF80
29	C17ORF85
30	C20ORF27
31	C2ORF68
32	CAMSAP1
33	CAMSAP2
34	CASC3
35	CBX3
36	CCDC85C
37	CCDC97
38	CD2BP2
39	CDC5L
40	CDC73
41	CDK11A
42	CDYL2
43	CFH
44	CHD3
45	CHD7
46	CIC
47	CIRH1A
48	CIT
49	CLASP1
50	CLASRP
51	CLP1
52	COBL
53	CORO2A
54	CRTC3
55	CSNK1E
56	CXORF56
57	DBR1
58	DBT
59	DCP1A
60	DDX47
61	DGKZ
62	DHX35
63	DHX40
64	DHX8
65	DIAPH1
66	DIAPH2
67	DIAPH3
68	DKC1
69	DLGAP5
70	DOCK6
71	ECT2
72	EDC3
73	EHMT1
74	EMSY
75	EP400
76	ERC1
77	ERCC3
78	ERH
79	ESCO1
80	EXOSC10
81	EXOSC3
82	FAM111B
83	FAM195A
84	FAM195B
85	FAM60A
86	FAM83A
87	FAM98A
88	FOXK1
89	FOXK2
90	FRA10AC1
91	GID8
92	GIT1
93	GIT2
94	GPATCH1
95	GPS2
96	GTF2H1
97	GTF2H4
98	GTF3C1
99	GTPBP2
100	HDAC1
101	HDAC2
102	HELZ2
103	HLTF
104	HMGXB4
105	HMMR
106	HNRNPK
107	HNRNPL
108	HSPG2
109	ING1
110	ING3
111	INIP
112	INTS10
113	INTS2
114	INTS3
115	INTS4
116	INTS5
117	INTS6
118	JMJD1C
119	KANSL2
120	KANSL3
121	KAT7
122	KCTD10
123	KCTD3
124	KDM1A
125	KDM1B
126	KIAA1967
127	KIF1C
128	KIF23
129	KNOP1
130	KPNB1
131	KPRP
132	KRT75
133	LENG8
134	LEO1
135	LEPRE1
136	LLGL1
137	LMNA
138	LRWD1
139	LUC7L3
140	LUZP1
141	MAD2L2
142	MAML1
143	MAPRE2
144	MARK2
145	MARK3
146	MAU2
147	MID1
148	MINK1
149	MLLT3
150	MNT
151	MPHOSPH6
152	MRPS2
153	MRPS9
154	MXI1
155	MYBBP1A
156	NAF1
157	NAT10
158	NCOR2
159	NF1
160	NFATC3
161	NFIB
162	NFIC
163	NKRF
164	NOP9
165	NPM1
166	NRDE2
167	NUDCD1
168	OGT
169	ORC5
170	OSBPL10
171	OSBPL11
172	OSBPL9
173	PARP14
174	PATL1
175	PAXBP1
176	PBRM1
177	PCGF1
178	PCGF5
179	PDCD6
180	PEG10
181	PHF10
182	PHF20
183	PHF21A
184	PITHD1
185	PKP3
186	PLEKHA6
187	POLDIP3
188	POLR2A
189	POLR2B
190	PPFIA1
191	PPHLN1
192	PRKRIR
193	PUS7L
194	RALGAPA1
195	RALGAPB
196	RANBP2
197	RANGAP1
198	RAPH1
199	RAVER1
200	RBBP4
201	RBBP7
202	RBM12B
203	RBM15
204	RBM19
205	RBM26
206	RCOR3
207	RECQL5
208	RFC1
209	RFC2
210	RFC4
211	RFC5
212	RNF20
213	RNF40
214	RPP25
215	RPP30
216	RPP38
217	RREB1
218	RSL1D1
219	RXRB
220	SAFB
221	SAP130
222	SAP30
223	SATB2
224	SCAF4
225	SEPSECS
226	SEPT8
227	SEPT9
228	SETD2
229	SF3B3
230	SF3B4
231	SHKBP1
232	SIN3A
233	SKIV2L2
234	SMAD4
235	SMAD5
236	SMARCA2
237	SMARCC2
238	SMARCD3
239	SMG8
240	SMG9
241	SMOC1
242	SNX8
243	SPTAN1
244	SQSTM1
245	SRBD1
246	SREK1
247	SRGAP1
248	SRRM1
249	SUDS3
250	TANC1
251	TAOK1
252	TBPL1
253	TDP1
254	TDRD3
255	TFAP2A;TFAP2B
256	TGIF2LX;TGIF2LY
257	THRAP3
258	TIMM13
259	TJAP1
260	TJP1
261	TJP2
262	TLE3
263	TOP2A
264	TOP2B
265	TOP3B
266	TRA2B
267	TRIM28
268	TRIM33
269	TRIP12
270	TRIP4
271	TTF2
272	U2SURP
273	UBN1
274	UBP1
275	USP22
276	UTP20
277	VAC14
278	WDR61
279	WDR83
280	WIZ
281	XAB2
282	XRN1
283	YLPM1
284	YTHDC2
285	ZBTB1
286	ZCCHC4
287	ZFC3H1
288	ZNF217
289	ZNF414
290	ZNF451
291	ZNF512B
292	ZNF592

Appendix 7

A total of 292 proteins exclusively bound to full length of ICP22 sorted in rank based on uniqueness and fold enrichment

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
1	DIAPH1	29	27	1.85E+09	Unique
2	BAZ1A	26	26	1.37E+09	Unique
3	BAIAP2L1	15	13	9.48E+08	Unique
4	NKRF	15	13	8.82E+08	Unique
5	BRD7	13	12	8.33E+08	Unique
6	PARP14	21	21	7.92E+08	Unique
7	GIT2	16	14	7.49E+08	Unique
8	SUDS3	7	6	7.44E+08	Unique
9	ASCC3	20	20	6.86E+08	Unique
10	RBM19	14	14	6.70E+08	Unique
11	BLM	8	6	6.03E+08	Unique
12	DHX35	13	13	5.37E+08	Unique
13	ERCC3	10	10	4.88E+08	Unique
14	COBL	12	11	4.87E+08	Unique
15	RECQL5	12	12	4.87E+08	Unique
16	TOP3B	14	14	4.70E+08	Unique
17	RBM12B	10	10	4.63E+08	Unique
18	NF1	17	16	4.51E+08	Unique
19	RALGAPB	10	10	4.39E+08	Unique
20	TOP2A	8	8	4.36E+08	Unique
21	INTS6	12	12	4.35E+08	Unique
22	BBX	9	9	4.27E+08	Unique
23	ARID4B	12	12	4.17E+08	Unique
24	SHKBP1	8	6	4.00E+08	Unique
25	SQSTM1	8	7	3.66E+08	Unique
26	ZNF451	9	9	3.46E+08	Unique
27	BRMS1L	5	5	3.39E+08	Unique
28	MINK1	9	8	3.33E+08	Unique
29	INTS5	5	5	3.33E+08	Unique
30	PCGF5	4	3	3.21E+08	Unique
31	CRTC3	6	6	3.10E+08	Unique
32	ZNF217	8	8	3.04E+08	Unique
33	FAM111B	7	7	3.00E+08	Unique
34	ZBTB1	5	4	2.85E+08	Unique
35	TOP2B	6	6	2.79E+08	Unique
36	BAP18	4	4	2.70E+08	Unique
37	MYBBP1A	7	6	2.65E+08	Unique
38	HELZ2	12	10	2.60E+08	Unique
39	CIRH1A	6	6	2.54E+08	Unique
40	RALGAPA1	6	6	2.52E+08	Unique
41	SAP30	2	2	2.52E+08	Unique
42	TRIP4	7	7	2.46E+08	Unique
43	KIF1C	6	6	2.41E+08	Unique
44	CHD3	7	6	2.40E+08	Unique
45	PLEKHA6	10	9	2.37E+08	Unique
46	CLASP1	8	8	2.35E+08	Unique
47	MARK3	5	4	2.32E+08	Unique
48	ZFC3H1	4	4	2.28E+08	Unique
49	DHX8	11	11	2.27E+08	Unique
50	SNX8	8	6	2.27E+08	Unique
51	TGIF2LX;TGIF	4	3	2.17E+08	Unique
52	RANGAP1	8	6	2.09E+08	Unique
53	FRA10AC1	3	3	2.09E+08	Unique
54	KRT75	2	2	2.00E+08	Unique
55	EXOSC3	4	4	1.97E+08	Unique
56	TTF2	6	6	1.95E+08	Unique
57	INTS10	6	6	1.92E+08	Unique
58	FAM60A	4	4	1.91E+08	Unique
59	OSBPL11	3	3	1.87E+08	Unique
60	KAT7	6	6	1.87E+08	Unique
61	AMMECR1	4	3	1.79E+08	Unique
62	UTP20	7	7	1.76E+08	Unique
63	SRBD1	4	4	1.68E+08	Unique
64	TANC1	7	7	1.67E+08	Unique
65	LENG8	5	5	1.57E+08	Unique
66	GTF3C1	5	5	1.56E+08	Unique
67	DBT	2	2	1.54E+08	Unique
68	EP400	12	6	1.46E+08	Unique
69	SRGAP1	5	5	1.44E+08	Unique
70	FAM195B	3	2	1.41E+08	Unique
71	PAXBP1	5	5	1.40E+08	Unique
72	PHF20	7	3	1.37E+08	Unique
73	PHF10	5	4	1.36E+08	Unique
74	BCKDK	4	3	1.30E+08	Unique
75	HMGXB4	2	2	1.26E+08	Unique
76	NAF1	3	2	1.25E+08	Unique
77	KNOP1	6	4	1.14E+08	Unique
78	ZNF512B	5	4	1.12E+08	Unique
79	BCL7C	3	3	1.10E+08	Unique
80	INTS4	3	3	1.03E+08	Unique
81	MARK2	3	3	1.02E+08	Unique
82	BHLHE40	3	3	1.00E+08	Unique
83	SEPSECS	4	3	8.87E+07	Unique
84	RSL1D1	4	4	8.67E+07	Unique
85	CXorf56	2	2	8.47E+07	Unique
86	GTF2H4	3	3	8.46E+07	Unique
87	TFAP2A;TFAP	4	3	8.31E+07	Unique
88	RPP25	2	2	8.30E+07	Unique
89	DGKZ	3	3	8.22E+07	Unique
90	WDR83	2	2	8.16E+07	Unique
91	CDYL2	3	3	8.07E+07	Unique
92	ESCO1	3	3	8.06E+07	Unique
93	INTS2	3	3	7.94E+07	Unique
94	KANSL3	11	4	7.74E+07	Unique
95	GPATCH1	4	3	7.70E+07	Unique
96	GTF2H1	3	3	7.63E+07	Unique
97	MRPS9	4	4	7.52E+07	Unique
98	FAM83A	3	3	7.49E+07	Unique
99	MLLT3	4	3	7.16E+07	Unique
100	TIMM13	2	2	7.01E+07	Unique
101	MXI1	2	2	6.93E+07	Unique
102	PCGF1	2	2	6.68E+07	Unique
103	SMARCD3	3	2	6.60E+07	Unique
104	XAB2	3	3	6.57E+07	Unique
105	ING3	4	2	6.51E+07	Unique
106	MAD2L2	2	2	6.47E+07	Unique
107	BRMS1	3	2	6.39E+07	Unique
108	ASCC1	3	3	6.27E+07	Unique
109	DCP1A	2	2	6.25E+07	Unique
110	MRPS2	2	2	5.66E+07	Unique
111	DOCK6	3	2	5.31E+07	Unique
112	ECT2	2	2	5.30E+07	Unique
113	CSNK1E	2	2	5.21E+07	Unique
114	CFH	2	2	5.16E+07	Unique
115	KANSL2	6	2	5.07E+07	Unique
116	ZCCHC4	2	2	5.05E+07	Unique
117	KPRP	8	2	5.04E+07	Unique
118	ARHGEF5	8	4	4.22E+07	Unique
119	ARID4A	2	2	4.05E+07	Unique
120	EMSY	2	2	3.81E+07	Unique
121	CDK11A	2	2	3.55E+07	Unique
122	ZNF414	2	2	3.22E+07	Unique
123	PEG10	2	2	3.10E+07	Unique
124	GID8	2	2	3.03E+07	Unique
125	SATB2	3	2	2.94E+07	Unique
126	C2orf68	3	2	2.78E+07	Unique
127	HSPG2	4	2	2.29E+07	Unique
128	CIC	4	3	2.22E+07	Unique

Continue Appendix 7:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity beads
129	KDM1B	2	2	1.73E+07	Unique
130	UBN1	5	2	1.18E+07	Unique
131	PHF21A	12	12	9.02E+08	261.3
132	LEPRE1	9	8	3.97E+08	116.0
133	GIT1	13	10	9.36E+08	56.2
134	PBRM1	24	21	1.47E+09	44.7
135	CCDC97	14	14	3.03E+09	32.6
136	CIT	22	21	1.05E+09	31.5
137	TRIP12	3	3	7.17E+07	30.5
138	DIAPH3	14	14	7.61E+08	29.6
139	RFC1	19	19	7.09E+08	24.4
140	PKP3	12	11	4.70E+08	20.9
141	NAT10	19	19	8.71E+08	19.3
142	YTHDC2	31	23	1.29E+09	19.3
143	SMG8	19	18	1.44E+09	18.0
144	CAMSAP2	19	17	7.47E+08	17.9
145	TBPL1	3	3	5.01E+07	16.2
146	ARHGEF7	25	22	3.62E+09	15.4
147	MAU2	4	3	1.98E+08	15.3
148	RANBP2	31	31	2.25E+09	15.3
149	CAMSAP1	16	15	5.92E+08	15.0
150	EHMT1	17	14	8.92E+08	14.9
151	KCTD3	14	14	1.37E+09	14.9
152	NFATC3	10	10	5.14E+08	14.7
153	U2SURP	42	38	1.62E+10	14.7
154	XRN1	25	17	9.01E+08	14.7
155	SIN3A	45	44	9.79E+09	14.0
156	OSBPL10	8	6	7.11E+08	13.5
157	WIZ	27	26	1.88E+09	12.3
158	TJP2	30	30	1.89E+09	11.4
159	TDRD3	10	9	6.13E+08	11.0
160	ABL1	10	10	4.40E+08	10.9
161	UBP1	7	7	4.13E+08	9.7
162	CD2BP2	10	8	1.36E+09	9.7
163	HNRNPL	32	31	1.50E+11	9.6
164	CHD7	41	33	1.40E+09	9.2
165	SMG9	11	10	9.24E+08	9.1
166	GTPBP2	3	2	4.89E+07	8.9
167	SAP130	15	12	1.50E+09	8.5
168	CASC3	6	4	2.67E+08	8.4
169	MNT	7	6	8.48E+08	8.4
170	RREB1	17	13	5.75E+08	8.2
171	LEO1	19	16	3.39E+09	8.2
172	KPNB1	4	3	2.25E+08	8.0
173	NOP9	6	5	3.87E+08	8.0
174	RFC5	10	10	9.99E+08	7.8
175	JMJD1C	36	30	1.37E+09	7.6
176	ARID2	21	20	1.02E+09	7.5
177	PPHLN1	12	12	1.19E+09	7.0
178	SETD2	9	8	2.06E+08	6.9
179	TRIM33	6	6	3.43E+08	6.8
180	SF3B4	9	9	2.20E+10	6.8
181	ANKHD1	2	2	3.57E+08	6.7
182	SMOC1	4	4	1.44E+08	6.7
183	NFIC	19	17	1.08E+10	6.6
184	ERH	4	4	1.12E+09	6.6
185	RPP38	2	2	2.27E+08	6.5
186	SRRM1	13	11	5.37E+09	6.4
187	SMARCC2	2	2	1.53E+10	6.3
188	EXOSC10	27	27	2.46E+09	6.2
189	DIAPH2	9	7	2.05E+08	6.2
190	MID1	13	12	8.20E+08	6.0
191	CDC73	24	24	2.07E+09	5.9
192	RBM15	28	28	7.68E+09	5.8
193	PUS7L	11	11	3.62E+08	5.7
194	SMAD4	21	20	8.37E+09	5.7
195	HDAC2	10	9	3.72E+09	5.7
196	DDX47	9	8	7.29E+08	5.7
197	BOD1L1	101	76	1.34E+10	5.6
198	RFC2	11	10	8.74E+08	5.6
199	INTS3	32	30	5.34E+09	5.3
200	HDAC1	11	11	4.61E+09	5.2
201	SF3B3	57	54	8.30E+10	5.1
202	MPHOSPH6	5	4	3.01E+08	5.0
203	PPFIA1	9	8	2.36E+08	4.9
204	USP22	5	4	3.61E+08	4.9
205	RPP30	13	13	2.00E+09	4.9
206	KDM1A	24	24	2.76E+09	4.8
207	KIAA1967	31	31	8.61E+09	4.8
208	EDC3	9	8	9.90E+08	4.7
209	NRDE2	3	2	6.82E+07	4.7
210	RAVER1	21	21	8.98E+09	4.5
211	TDP1	21	21	2.44E+09	4.4
212	DLGAP5	2	2	5.81E+09	4.4
213	NUDCD1	24	19	2.40E+09	4.4
214	NFIB	8	8	1.99E+09	4.4
215	SEPT8	9	9	1.54E+09	4.3
216	FOXK2	13	12	8.33E+08	4.3
217	SMARCA2	10	8	5.03E+08	4.2
218	C20orf27	7	5	6.86E+08	4.2
219	SAFB	25	23	2.17E+10	4.1
220	SPTAN1	39	36	1.91E+09	4.1
221	POLR2A	15	10	3.79E+08	4.0
222	GPS2	5	5	7.74E+08	4.0
223	RNF40	42	38	8.42E+09	4.0
224	PATL1	3	2	5.07E+07	3.9
225	RNF20	45	40	1.07E+10	3.9
226	TAOK1	18	14	1.18E+09	3.7
227	OSBPL9	15	15	1.40E+09	3.7
228	NCOR2	33	25	1.65E+09	3.6
229	HLTF	33	33	3.55E+09	3.6
230	ZNF592	4	4	1.34E+08	3.6
231	THRAP3	20	18	6.91E+09	3.6
232	DKC1	17	13	1.31E+09	3.5
233	WDR61	14	14	2.92E+09	3.5
234	ERC1	9	7	4.32E+09	3.4
235	CORO2A	2	2	7.14E+07	3.4
236	KIF23	21	18	1.61E+09	3.4
237	VAC14	2	2	9.43E+07	3.3
238	RBBP4	10	10	2.18E+09	3.3
239	YLPM1	33	29	2.68E+09	3.3
240	FOXK1	18	15	1.33E+09	3.3
241	DBR1	22	21	3.94E+09	3.2
242	ADRBK1	14	13	9.23E+08	3.2
243	SKIV2L2	31	25	1.48E+09	3.1
244	PITHD1	10	8	1.46E+09	3.1
245	CBX3	5	5	1.75E+09	3.1
246	TJAP1	20	12	5.51E+08	3.1
247	RCOR3	3	3	6.97E+07	3.1
248	ATXN2L	31	25	5.04E+09	3.0
249	POLR2B	23	19	1.05E+09	3.0
250	RBM26	2	2	1.13E+10	3.0
251	ING1	4	4	2.27E+08	3.0
252	HMMR	9	4	1.54E+08	2.9
253	RBBP7	12	11	7.19E+09	2.9
254	PRKRIR	16	11	1.22E+09	2.9
255	LMNA	4	4	2.26E+11	2.9
256	INIP	5	5	1.81E+09	2.8

Continue Appendix 7:

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
257	TJP1	23	20	1.12E+09	2.8
258	POLDIP3	9	9	5.19E+08	2.8
259	OGT	27	20	1.35E+09	2.8
260	SCAF4	7	7	1.38E+09	2.8
261	ORC5	11	9	6.30E+08	2.7
262	NPM1	4	4	1.05E+08	2.7
263	RXRB	9	9	8.49E+08	2.7
264	MAML1	4	3	1.59E+08	2.6
265	SEPT9	5	4	5.38E+08	2.6
266	TLE3	14	13	6.74E+08	2.6
267	SREK1	10	9	2.52E+09	2.5
268	CLP1	16	12	1.53E+09	2.5
269	SMAD5	4	3	6.95E+08	2.5
270	MAPRE2	12	12	3.62E+09	2.5
271	C17orf85	6	5	2.01E+08	2.5
272	ABCF2	25	22	4.67E+09	2.4
273	RFC4	9	8	5.45E+08	2.4
274	CCDC85C	8	5	1.57E+08	2.4
275	FAM98A	10	8	2.27E+09	2.4
276	TRA2B	7	7	3.85E+08	2.3
277	FAM195A	4	3	4.27E+08	2.3
278	LRWD1	20	19	3.52E+09	2.2
279	PDCD6	3	2	1.07E+08	2.2
280	LUZP1	7	4	1.57E+08	2.2
281	HNRNPK	23	17	3.41E+09	2.1
282	CLASRP	2	2	1.08E+08	2.1
283	C16orf80	3	3	3.38E+08	2.1
284	AURKA	13	13	2.21E+09	2.1
285	AQR	32	28	1.92E+09	2.1
286	CDC5L	37	36	1.15E+10	2.1
287	DHX40	11	9	2.63E+08	2.0
288	KCTD10	6	4	2.10E+08	2.0
289	RAPH1	2	2	7.01E+07	2.0
290	TRIM28	30	28	4.87E+09	2.0
291	LLGL1	11	8	2.59E+08	2.0
292	LUC7L3	15	10	3.18E+09	2.0

Appendix 8

A total of 129 proteins exclusively bound to SCR of ICP22 sorted in alphabetical order

No	Gene names
1	ACTR8
2	ALB
3	API5
4	ARF4
5	ARGLU1
6	ASTE1
7	BMP2K
8	BZW2
9	C20orf27
10	CLK4
11	CREB5
12	CWC25
13	DEK
14	DNAJC9
15	DSG1
16	DYNC1H1
17	DYNLL2
18	ECHS1
19	EIF4A3
20	EME1
21	ETF1
22	FAM117B
23	FAU
24	GATA6
25	GOLGB1
26	H1FX
27	H2AFZ;H2AFV
28	HIST1H2BB
29	HIST1H3A;H3F3
30	HNRNPC
31	INO80C
32	INO80E
33	IQGAP1
34	ISG20L2
35	KLF3
36	KRT4
37	KRT78
38	KRT8
39	LAMB1
40	LAMC1
41	LGALS3BP
42	LRRC16A
43	LSM2
44	LSM3
45	LSM4
46	LSM6
47	MED11
48	MED21
49	MED22
50	MED28
51	MRPS28
52	MSL2
53	MYNN
54	NAA38
55	NHP2L1
56	NOL12
57	NOP16
58	NR4A3
59	NRBP1
60	ODF2
61	PARD6B
62	PLCD3
63	PLIN3
64	POLR2E
65	PPP1CB
66	PSPC1
67	RBM39
68	RNF169
69	RNPC3
70	RPL11
71	RPL18A
72	RPL23
73	RPL29
74	RPL31
75	RPL32
76	RPL35
77	RPL36
78	RPL36AL;RPL36A
79	RPL38
80	RPL4
81	RPL7A
82	RPS15
83	RPS15A
84	RPS18
85	RPS19
86	RPS20
87	RPS26;RPS26P11
88	RPS3
89	RPS5
90	RTN3
91	RTN4
92	RUVBL1
93	S100A12
94	SLC39A10
95	SLC39A6
96	SMR3B
97	SP2
98	TAF11
99	TAF13
100	TAF9B
101	TBRG1
102	TCTEX1D2
103	TFPT
104	TFRC
105	TNKS
106	TSC22D1
107	UBE2M
108	UBR5
109	Unknown
110	WDR60
111	ZBTB11
112	ZBTB24
113	ZBTB9
114	ZFX;ZFY
115	ZKSCAN4
116	ZKSCAN8
117	ZNF276
118	ZNF281
119	ZNF324;ZNF324B
120	ZNF384
121	ZNF460
122	ZNF579
123	ZNF652
124	ZNF710
125	ZNF740
126	ZNF784
127	ZSCAN21
128	ZSCAN26
129	ZSCAN31

Appendix 9

A total of 129 proteins exclusively bound to SCR of ICP22 sorted in rank based on uniqueness and fold enrichment

No	Gene names	Unique peptides	Unique peptides ICP22FL	Intensity ICP22FL	Intensity ICP22FL /Intensity
1	SOGA2	33	29	2.24E+09	Unique
2	RPS19	9	9	2.20E+09	Unique
3	KRT4	8	7	1.07E+09	Unique
4	RPS15	3	3	1.04E+09	Unique
5	FAU	3	3	7.42E+08	Unique
6	RNPC3	2	2	6.03E+08	Unique
7	RPL36AL;RPL	2	2	4.89E+08	Unique
8	INO80C	6	5	3.99E+08	Unique
9	RPS20	3	3	3.66E+08	Unique
10	HIST1H3A;H3	3	3	3.50E+08	Unique
11	S100A12	2	2	3.46E+08	Unique
12	TAF9B	5	5	3.41E+08	Unique
13	RNF169	9	9	2.86E+08	Unique
14	ODF2	10	8	2.72E+08	Unique
15	CREB5	5	5	2.67E+08	Unique
16	MED28	4	4	2.53E+08	Unique
17	GATA6	8	8	2.49E+08	Unique
18	LAMC1	7	7	2.20E+08	Unique
19	UBR5	5	4	2.05E+08	Unique
20	ZNF281	11	6	2.05E+08	Unique
21	MED21	2	2	1.95E+08	Unique
22	RPS18	3	3	1.95E+08	Unique
23	RPS15A	3	3	1.87E+08	Unique
24	RPL35	2	2	1.77E+08	Unique
25	ZKSCAN8	7	6	1.72E+08	Unique
26	ZNF460	4	4	1.67E+08	Unique
27	ZNF579	7	6	1.66E+08	Unique
28	ZNF740	3	3	1.64E+08	Unique
29	NAA38	3	3	1.56E+08	Unique
30	NOP16	2	2	1.56E+08	Unique
31	INO80E	2	2	1.55E+08	Unique
32	NHP2L1	3	3	1.47E+08	Unique
33	TNKS	7	6	1.42E+08	Unique
34	PARD6B	3	3	1.42E+08	Unique
35	MED11	2	2	1.29E+08	Unique
36	ZNF784	5	5	1.28E+08	Unique
37	ZBTB24	4	4	1.24E+08	Unique
38	NR4A3	4	3	1.20E+08	Unique
39	WDR60	3	2	1.17E+08	Unique
40	RPL23	3	3	1.14E+08	Unique
41	CLK4	4	3	1.09E+08	Unique
42	TFPT	2	2	1.08E+08	Unique
43	LSM2	3	3	1.08E+08	Unique
44	ZSCAN21	4	3	1.03E+08	Unique
45	RPL38	2	2	1.02E+08	Unique
46	RPS5	2	2	9.74E+07	Unique
47	RTN4	2	2	9.58E+07	Unique
48	LAMB1	5	4	9.36E+07	Unique
49	ISG20L2	4	4	8.50E+07	Unique
50	ZNF652	5	3	8.45E+07	Unique
51	TBRG1	2	2	8.35E+07	Unique
52	RPL4	5	4	8.16E+07	Unique
53	ZNF384	4	4	8.08E+07	Unique
54	ZNF710	2	2	7.87E+07	Unique
55	FAM117B	3	2	7.83E+07	Unique
56	NOL12	4	3	7.78E+07	Unique
57	KLF3	3	3	7.54E+07	Unique
58	SLC39A10	3	3	7.52E+07	Unique
59	API5	6	3	7.50E+07	Unique
60	TSC22D1	4	3	7.26E+07	Unique
61	EIF4A3	5	3	6.33E+07	Unique
62	SP2	5	3	6.31E+07	Unique
63	RTN3	2	2	6.17E+07	Unique
64	RPL18A	3	2	6.08E+07	Unique
65	PLIN3	3	3	5.77E+07	Unique
66	HNRNPC	4	2	5.74E+07	Unique
67	MRPS28	3	3	5.70E+07	Unique
68	ALB	12	2	5.69E+07	Unique
69	LGALS3BP	5	4	5.45E+07	Unique
70	ZSCAN31	2	2	5.34E+07	Unique
71	ZSCAN26	3	3	5.32E+07	Unique
72	GOLGB1	2	2	5.15E+07	Unique
73	H1FX	2	2	5.11E+07	Unique
74	UBE2M	2	2	4.99E+07	Unique
75	RPL7A	4	3	4.94E+07	Unique
76	ZFX;ZFY	2	2	4.66E+07	Unique
77	BMP2K	4	3	4.63E+07	Unique
78	ZKSCAN4	3	2	4.60E+07	Unique
79	TAF11	4	4	4.59E+07	Unique
80	TAF13	2	2	4.59E+07	Unique
81	MSL2	3	2	4.40E+07	Unique
82	SOGA2	2	2	4.39E+07	Unique
83	LRRC16A	2	2	4.38E+07	Unique
84	TFRC	3	3	4.33E+07	Unique
85	PLCD3	3	3	4.33E+07	Unique
86	EME1	3	2	4.12E+07	Unique
87	ZBTB9	4	2	4.08E+07	Unique
88	RPL11	2	2	4.06E+07	Unique
89	BZW2	3	2	4.04E+07	Unique
90	ZBTB11	4	2	4.01E+07	Unique
91	RPL32	3	3	3.90E+07	Unique
92	ETF1	2	2	3.83E+07	Unique
93	MYNN	2	2	3.69E+07	Unique
94	RPL36	2	2	3.58E+07	Unique
95	ZNF276	2	2	3.03E+07	Unique
96	DEK	2	2	2.82E+07	Unique
97	NRBP1	3	3	2.81E+07	Unique
98	LSM6	2	2	2.78E+07	Unique
99	MED22	4	2	2.75E+07	Unique
100	SLC39A6	2	2	2.68E+07	Unique
101	DSG1	8	2	2.38E+07	Unique
102	TCTEX1D2	2	2	1.83E+07	Unique
103	SMR3B	2	2	1.65E+07	Unique
104	ASTE1	2	2	1.62E+07	Unique
105	ZNF324;ZNF3	2	2	1.40E+07	Unique
106	ECHS1	2	2	1.04E+07	1507.8
107	KRT78	2	2	4.88E+06	748.0
108	RPL31	7	7	2.48E+09	28.0
109	DYNC1H1	13	11	2.41E+08	18.9
110	ARF4	4	3	1.59E+08	11.3
111	RPS3	4	3	5.95E+07	9.1
112	RPS26;RPS26	3	3	1.74E+09	8.7
113	RPL29	2	2	3.59E+08	8.6
114	ACTR8	16	16	1.65E+09	8.5
115	CWC25	11	8	3.80E+08	8.1
116	LSM4	3	3	1.56E+08	7.9
117	IQGAP1	20	17	5.24E+08	6.8
118	KRT8	6	5	1.20E+08	5.7
119	ARGLU1	9	7	5.15E+09	5.0
120	H2AFZ;H2AF	3	3	2.21E+09	4.3
121	PSPC1	30	29	2.82E+10	4.2
122	LSM3	2	2	4.93E+07	3.8
123	RBM39	16	16	2.01E+09	3.7
124	DNAJC9	11	10	1.04E+09	3.7
125	RUVBL1	28	26	6.38E+09	3.2
126	PPP1CB	6	4	4.63E+08	2.8
127	POLR2E	4	3	1.36E+08	2.6
128	C20orf27	7	6	3.65E+08	2.1
129	DYNLL2	3	2	1.50E+08	2.0

Appendix 10

A total of 45 molecular functions for ICP22-bound cellular proteins are classified

GO_Term	Count	old Enrichmen	Benjamini	(-log10)	Genes
RNA binding	66	4.0	3.34E-20	19.476	RPL18, CHERP, RPL19, RPL14, RPL13, TRMT2A, RBM15B, RPL15, CCNT1, RBM7, ZNF638, LARP1, NONO, MAZ, RPS3A, ZRSR1, ZRSR2, PRPF31, SNRPA1, SF3B14, CDK9, RPS4X, SLTM, LARP7, ZNF239, RBMX2, CPSF7, RPS13, CPSF6, RPS11, EEFSEC, THOC2, CPSF3, CPSF2, THOC3, CPSF1, THOC1, FIP1L1, RBM33, SETD1B, TRA2A, SNRPB2, RPL27A, DUSP11, NUFIP2, DDX46, HTATSF1, PRPF8, NUDT21, RPL8, PPIL4, SNRNP35, RBM27, RBM25, DDX42, PPP1R10, RPL24, RPL23A, SF3A1, FBL, SAFB2, ATXN2, SFPQ, POP4, PAPOLG, RBM17
RNA polymerase II transcription factor activity	40	7.2	3.30E-20	19.481	MED24, TCF7L2, MED20, VPS72, ZNF148, HTATSF1, MED26, CDK12, MLLT1, VEZF1, ZBTB7B, KLF5, TAF4, TCF7, TAF6, KLF13, NFRKB, TAF7, MED12, LEF1, MED14, MED13, MED10, MED6, MED31, MED19, MED7, TAF10, MED4, TAF12, MED30, SP1, MED17, MED8, MED9, SP4, SMARCA5, FOXC2, KLF4, MED1
general RNA polymerase II transcription factor activity	22	19.4	5.07E-20	19.295	TAF4, TAF6, TAF7, MED12, MED24, MED14, MED13, MED20, MED10, MED19, MED31, MED6, MED7, TAF10, MED4, MED30, TAF12, MED17, MED8, MED26, MED9, MED1
transcription regulator activity	95	2.8	2.92E-19	18.534	CBX4, MED24, CTCF, MED23, ZKSCAN1, ZEB1, REST, MED20, CITED2, ZGPAT, ASH2L, SMARCD2, ZNF148, MED26, MED27, SMARCD1, CDK12, E4F1, GATAD1, TAF1L, NFX1, ZNF593, YY1, ARID5B, RING1, MECP2, MED12, LEF1, MED14, FOXJ3, MED13, ARID1B, MED10, PRPF6, MED6, MED19, MED7, TAF10, MED4, TAF12, BPTF, BAZ1B, MED16, MED17, MED8, MED9, SMARCA5, FOXC2, MED1, SMARCA4, THOC1, ZNF131, TRRAP, ZBTB17, TCF7L2, VPS72, CXXC1, CHD8, CIR1, ATN1, HTATSF1, MLLT1, TAF9, VEZF1, KLF5, ZBTB7B, ZBTB7A, TAF1, TAF4, TCF7, FOXL2, TAF6, KLF13, TAF5, SUB1, TAF8, NFRKB, CBL, TAF7, ZNF24, AFF4, KLF16, PPP1R10, MED31, MSL3, MED30, SP1, SMARCC1, RNF2, SP4, TCEB3, PHF5A, ZNF219, RERE, KLF4
RNA polymerase II transcription mediator activity	17	25.8	9.82E-18	17.008	MED12, MED24, MED14, MED13, MED20, MED10, MED19, MED6, MED31, MED7, MED4, MED30, MED17, MED26, MED8, MED9, MED1
transcription activator activity	45	4.8	1.70E-16	15.769	MED24, MED23, CTCF, TRRAP, ZEB1, CXXC1, CITED2, CHD8, SMARCD2, ZNF148, MED26, MED27, SMARCD1, TAF9, E4F1, TAF1L, TAF1, TAF4, SUB1, TAF8, YY1, TAF7, RING1, MED12, LEF1, MED14, ARID1B, MED13, PRPF6, MED6, TAF10, MED7, MED4, TAF12, MED30, SP1, MED16, MED17, SMARCC1, SP4, SMARCA5, FOXC2, KLF4, MED1, SMARCA4
transcription coactivator activity	33	6.8	5.19E-16	15.285	MED24, MED23, ZEB1, TRRAP, CITED2, SMARCD2, MED26, SMARCD1, MED27, TAF9, E4F1, TAF1, TAF4, SUB1, YY1, TAF7, RING1, MED12, MED14, ARID1B, MED13, PRPF6, MED6, TAF10, MED7, MED4, MED30, TAF12, MED16, MED17, SMARCC1, SP4, MED1
transcription cofactor activity	39	4.7	7.94E-14	13.100	CBX4, MED24, MED23, CTCF, TRRAP, ZEB1, CITED2, CIR1, ATN1, SMARCD2, MED26, SMARCD1, MED27, TAF9, E4F1, TAF1, ZNF593, TAF4, SUB1, YY1, TAF7, RING1, MECP2, MED12, MED14, ARID1B, MED13, PRPF6, MED6, TAF10, MED7, MED4, MED30, TAF12, MED16, MED17, SMARCC1, SP4, MED1
DNA binding	110	2.1	1.08E-13	12.968	INO80, REST, ZNF638, ZIC2, CITED2, NONO, ZGPAT, TOP1, ZFP91, MED26, ZNF574, ZNF593, YY2, YY1, MECP2, ZNF8, ARID1B, TOX4, ZNF787, BAZ1B, ZNF239, SMARCA5, SMARCA4, GPBP1L1, ZNF76, ZNF131, VPS72, CXXC1, TAF9, VEZF1, THAP11, KLF5, TAF2, TAF1, TAF4, TCF7, TAF3, TAF6, TAF5, KLF13, SUB1, TAF8, TAF7, KLF16, ZBTB40, ZBTB43, ATM, SAFB2, ZNF629, MSL3, POLRMT, ZNF316, SMARCC1, ZNF219, KLF4, RERE, CCNT1, CTCF, ZKSCAN1, ZEB1, PHC3, MAZ, ASH2L, ZNF148, BRD4, E4F1, GATAD1, TAF1L, NFX1, POGZ, ARID5B, LEF1, NUSAP1, CDK9, FOXJ3, TAF10, TAF12, BPTF, ZNF691, ZNF692, FOXC2, RUVBL2, THOC1, MED1, ZFP64, ZBTB17, TCF7L2, CHD8, CIR1, CHD2, MUS81, MLLT1, CHD1, ZBTB7B, ZBTB7A, FOXL2, NFRKB, CBL, ZNF24, AFF4, PPP1R10, ZNF668, SP1, SFPQ, SP4, TCEB3, PHF5A, ALKBH2, ZNF768, REPIN1
transcription factor binding	42	3.6	4.51E-11	10.346	CBX4, MED24, MED23, CTCF, TRRAP, ZEB1, TCF7L2, CITED2, CIR1, ATN1, SMARCD2, MED26, MED27, SMARCD1, TAF9, E4F1, TAF1, ZNF593, TAF4, SUB1, YY1, TAF7, RING1, MECP2, MED12, MED14, ARID1B, MED13, PRPF6, MED6, TAF10, MED7, MED4, MED30, TAF12, BPTF, MED16, MED17, SMARCC1, SP4, MED1, SMARCA4
vitamin D receptor binding	11	21.0	7.35E-10	9.134	MED4, MED30, MED16, MED17, TAF7, RING1, MED12, MED24, MED14, MED13, MED1
chromatin binding	21	6.2	4.25E-09	8.372	RING1, CBX4, LEF1, CTCF, REST, ZEB1, MSL3, SF3B1, TOP1, CHD8, CDYL, BAZ1B, SMARCC1, RNF2, SMARCA5, CHD2, CHD1, FOXC2, ACTL6A, SMARCA4, MED1
thyroid hormone receptor binding	11	15.1	2.93E-08	7.534	MED4, MED30, MED16, MED17, TAF7, RING1, MED12, MED24, MED14, MED13, MED1
structural constituent of ribosome	21	5.5	2.82E-08	7.549	RPL18, RPL17, RPL19, RPL14, RPL13, RPL15, RPL27A, RPL27, RPL24, RPL23A, RPS6, RPS4X, RPS8, RPS29, RPS3A, RPL8, RPL10, RPS13, RPS11, RPS23, RPS24
ligand-dependent nuclear receptor transcription coactivator activity	10	12.2	1.53E-06	5.815	MED4, MED30, MED16, MED17, RING1, MED12, MED24, MED14, MED13, MED1
nuclear hormone receptor binding	13	7.4	2.70E-06	5.568	TAF10, MED4, MED30, MED16, MED17, TAF7, RING1, MED12, MED24, MED14, MED13, MED1, PRPF6
hormone receptor binding	13	6.5	1.14E-05	4.944	TAF10, MED4, MED30, MED16, MED17, TAF7, RING1, MED12, MED24, MED14, MED13, MED1, PRPF6
zinc ion binding	85	1.6	4.87E-05	4.313	ZMAT5, CTCF, ZKSCAN1, ZEB1, ZNF638, REST, ZIC2, PHC3, ZGPAT, ZFP91, MAZ, ASH2L, ZNF148, LUC7L2, ZRSR1, E4F1, ZRSR2, GATAD1, ZNF574, FIZ1, INO80B, ZCCHC8, NFX1, PHRF1, ZNF593, POGZ, YY2, YY1, RING1, ZNF8, RBBP6, ZNF787, SNRNP48, BAZ1B, ZNF691, BPTF, ZNF692, ZNF239, ZNF746, CPSF3, DPF2, ZFP64, ZNF76, ZNF131, CBLL1, ZBTB17, CXXC1, RPS29, ZNF609, PTS, RBM27, VEZF1, THAP11, KLF5, ZBTB7B, TAF2, ZBTB7A, ZC3H13, TAF3, KLF13, CBL, ZNF24, KLF16, PPP1R10, ZBTB40, ZNF668, SF3A2, EHMT2, ZBTB43, SF3A3, ZNF629, PHF3, SP1, ZNF316, RNF2, WHSC1L1, SP4, LUC7L, ZNF219, GPATCH8, RERE, KLF4, PHF6, REPIN1, ZNF768
transcription repressor activity	22	3.1	1.66E-04	3.780	ZBTB7A, ZNF593, YY1, ARID5B, TAF7, RING1, ZNF24, MECP2, CBX4, CTCF, ZEB1, REST, TCF7L2, CITED2, CHD8, CIR1, ATN1, ZNF148, RNF2, E4F1, ZNF219, KLF4
RNA splicing factor activity, transesterification mechanism	7	12.8	1.84E-04	3.735	SF3B1, DDX23, PRPF8, SNRNP40, PRPF6, TXNL4A, SF3A3
helicase activity	12	3.8	4.85E-03	2.314	CHD8, DDX46, DDX23, SNRNP200, SMARCA5, DHX15, CHD2, CHD1, INO80, RUVBL2, DDX42, SMARCA4
cyclin-dependent protein kinase activity	6	8.8	7.03E-03	2.153	CDK19, CDK8, CDK12, CDK9, CDK11B, CDK13
transcription factor activity	39	1.8	9.04E-03	2.044	ZNF131, ZKSCAN1, CTCF, ZEB1, ZBTB17, TCF7L2, VPS72, CITED2, ZGPAT, CIR1, ZNF148, E4F1, GATAD1, NFX1, KLF5, ZBTB7B, FOXL2, TAF4, TAF6, TAF5, YY1, TAF7, CBL, KLF16, AFF4, ZNF24, LEF1, FOXJ3, MSL3, TAF10, TAF12, SP1, BAZ1B, SP4, FOXC2, PHF5A, ZNF219, KLF4, RERE

Appendix 11

The IpgC of *S.flexneri* (Query) and ICP22 of HSV-1 (Subject) sequence alignment

```
Query  16   IDAINSGATLKDINAIPDDMMDDIYSYAYDFYNKGRIEEAEVFFRFLCIYDFYNVDYIMG  75
            +D      T  + +A   D  DD   ++YD     R + A V  R     D  +
Sbjct  88   LDMSAESTTGTETDASVSDDPDDTSDWSYDDIPP-RPKRARVNLRLTSSPDRRDGVIFPK  146

Query  76   LAAIYQIKEQFQQAADLYAVA  96
            +  +   +E   +A    A +
Sbjct  147  MGRVRSTRETQPRAPTPSAPS  167

Query  90  ADLYAVAFA  98
           AD+   AFA
Sbjct  2   ADISPGAFA  10
```

Appendix 11: Sequence alignment of full-length of IpgC of *S.flexneri* (Query) and full-length of ICP22 of HSV-1 (Subject) protein sequences showing minimal homology of amino acid residues.

Appendix 12

Co-IP of transfected ICP22 and P-TEFb from human cell extract

To reproduce the interaction of ICP22 and P-TEFb in the cellular environment, co-immunoprecipitation (Co-IP) was performed using the Flag-tag full-length (Flag-FL) and the short conserved region (Flag-SCR) of ICP22. For Co-IP, the Flag tag is better than the Myc tag that was used in the chromatin immunoprecipitation experiments (Chapter 2). This is because the anti-myc antibody was raised against the epitope located within the amino acid sequence of the human proto-oncogene c-myc, and the anti-myc antibody therefore recognises endogenous c-myc. To start with, the HEK293 cells were transfected with plasmids expressing full length (Flag-FL) and short conserved region of ICP22 (Flag-SCR). It should be noted that the SCR does not contain a classical bona fide nuclear localisation signal (NLS), unlike the Flag-FL within its N-terminus. The proteins interacting with the Flag-SCR could therefore be cytoplasmic. The soluble extracts were prepared and precipitated using anti-Flag M2 magnetic beads designed to capture fusion proteins containing the Flag peptide sequence. As shown in Figure Appendix 12A, the HEK293 cells express the Flag-FL and Flag-SCR of ICP22. However, the recombinant proteins are detected at different levels, with the Flag-SCR expressed at a lower level. Unlike the Myc-SCR that was used in the chromatin immunoprecipitation experiments (Chapter 2) where the Myc tag was placed at the C-terminus, the Flag-tag here was placed at the N-terminus of the SCR. In addition, the secondary structure prediction shows that SCR is a hydrophobic region. The N-terminal Flag-tag SCR may be poorly expressed and/or turned over because of either of these things. Despite an unequal level of cellular Flag-FL and Flag-SCR, both were immunoprecipitated with the anti-Flag M2 magnetic beads and after beads washing, the protein complexes were analysed by SDS-PAGE, followed by immunoblotting against anti-Flag and anti-CDK9. As shown in Figure Appendix 12B, the Flag-FL migrates at the approximate size of ~75 kDa, while Flag-SCR that should migrates at the size of ~12 kDa was not detected. Not only is the Flag-SCR poorly expressed and/or turned over, but the expressed Flag-SCR does not seem to be immunoprecipitated by the anti-Flag M2 magnetic beads. CDK9 co-precipitates with Flag-FL, but not the Flag-SCR. This may be because the Flag-SCR was not immunoprecipitated by the anti-Flag M2 magnetic beads, or the amount of the immunoprecipitated Flag-SCR is not sufficient to detect an interaction with the CDK9 in HEK293 cells. The interaction between the transiently expressed

ICP22 and cellular CDK9 supports and confirms the previous interaction using the recombinant ICP22 and CDK9.

Although the Co-IP using Flag-FL of ICP22 is successful it was difficult to upscale the assay for the detection of interaction between ICP22 and other cellular proteins, or for the proteomic analysis. The strategy of using the recombinant proteins provides a higher amount of bait proteins compared to the protein overexpressed in cell culture, thus providing a higher chance of novel protein discovery. This co-immunoprecipitation can be extended in the future by immunoprecipitating the cellular CDK9 after transiently expressing ICP22.

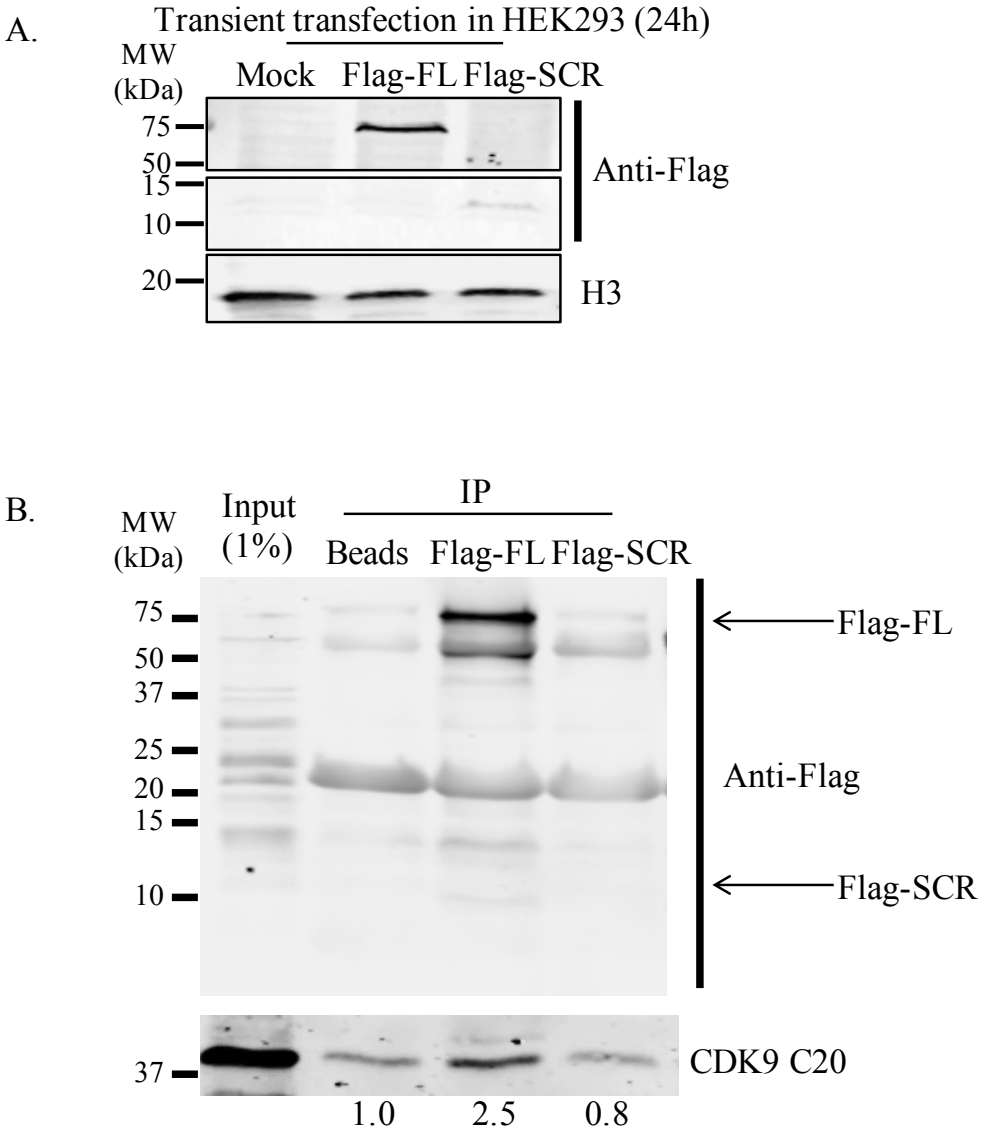


Figure Appendix 12. Confirmation of the ICP22 and CDK9 interaction by co-immunoprecipitation. A) HEK293 cells were transfected with Flag-tag full length (Flag-FL) and short conserved region (Flag-SCR) of ICP22 expressing plasmids. 24 h

post-infection, the cells were harvested in lysis buffer. A total of 20 µg of the lysates was subjected for SDS-PAGE analysis followed by immunoblot with anti-Flag and anti-H3 as loading control. **B)** The remaining lysates were incubated with anti-Flag M2 magnetic beads for 1 h rotating at 4˚C. After washing the beads, the bound proteins were subjected for SDS-PAGE analysis followed by immunoblot against anti-Flag and anti-CDK9. The values below the blot indicate the intensity of the CDK9 protein bands as measured by Image Studio Lite Ver 5.2

References

Abrisch, R. G., et al. (2015). "Infection by Herpes Simplex Virus 1 Causes Near-Complete Loss of RNA Polymerase II Occupancy on the Host Cell Genome." J Virol 90(5): 2503-2513.

Abrisch, R. G., et al. (2016). "Correction for Abrisch et al., Infection by Herpes Simplex Virus 1 Causes Near-Complete Loss of RNA Polymerase II Occupancy on the Host Cell Genome." J Virol 90(24): 11279.

Adelman, K. and J. T. Lis (2012). "Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans." Nat Rev Genet 13(10): 720-731.

Advani, S. J., et al. (2000). "The disappearance of cyclins A and B and the increase in activity of the G(2)/M-phase cellular kinase cdc2 in herpes simplex virus 1-infected cells require expression of the alpha22/U(S)1.5 and U(L)13 viral genes." J Virol 74(1): 8-15.

Advani, S. J., et al. (2003). "Herpes simplex virus 1 activates cdc2 to recruit topoisomerase II alpha for post-DNA synthesis expression of late genes." Proceedings of the National Academy of Sciences of the United States of America 100(8): 4825-4830.

Ahn, S. H., et al. (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing." Mol Cell 13(1): 67-76.

AJ, C. Q., et al. (2016). "Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb." Nucleic Acids Res 44(16): 7527-7539.

Akhtar, M. S., et al. (2009). "TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II." Mol Cell 34(3): 387-393.

Akoulitchev, S., et al. (2000). "TFIIH is negatively regulated by cdk8-containing mediator complexes." Nature 407(6800): 102-106.

Akoulitchev, S., et al. (1995). "Requirement for TFIIH kinase activity in transcription by RNA polymerase II." Nature 377(6549): 557-560.

Allen, B. L. and D. J. Taatjes (2015). "The Mediator complex: a central integrator of transcription." Nat Rev Mol Cell Biol 16(3): 155-166.

Ansari, S. A., et al. (2016). "Mediator, TATA-binding protein, and RNA polymerase II contribute to low histone occupancy at active gene promoters in yeast." J Biol Chem 291(19): 9938.

Aubert, M., et al. (2008). "The antiapoptotic herpes simplex virus glycoprotein J localizes to multiple cellular organelles and induces reactive oxygen species formation." J Virol 82(2): 617-629.

Aubert, M., et al. (1999). "Induction and prevention of apoptosis in human HEp-2 cells by herpes simplex virus type 1." J Virol 73(12): 10359-10370.

Babb, R., et al. (2001). "DNA recognition by the herpes simplex virus transactivator VP16: a novel DNA-binding structure." Mol Cell Biol 21(14): 4700-4712.

Baek, K., et al. (2007). "Crystal structure of human cyclin K, a positive regulator of cyclin-dependent kinase 9." J Mol Biol 366(2): 563-573.

Barboric, M., et al. (2009). "7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development." Proc Natl Acad Sci U S A 106(19): 7798-7803.

Bartholomew, B., et al. (1986). "RNA contacts subunits IIo and IIc in HeLa RNA polymerase II transcription complexes." J Biol Chem 261(30): 14226-14231.

Bartkowiak, B. and A. L. Greenleaf (2015). "Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex." J Biol Chem 290(3): 1786-1795.

Bartkowiak, B., et al. (2010). "CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1." Genes Dev 24(20): 2303-2316.

Bastian, T. W., et al. (2010). "Herpes simplex virus type 1 immediate-early protein ICP22 is required for VICE domain formation during productive viral infection." J Virol 84(5): 2384-2394.

Baumann, M., et al. (2010). "Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview." Mol Biotechnol 45(3): 241-247.

Baumli, S., et al. (2010). "Halogen bonds form the basis for selective P-TEFb inhibition by DRB." Chem Biol 17(9): 931-936.

Baumli, S., et al. (2012). "The CDK9 tail determines the reaction pathway of positive transcription elongation factor b." Structure 20(10): 1788-1795.

Baumli, S., et al. (2008). "The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation." Embo j 27(13): 1907-1918.

Bell, G. I., et al. (1976). "Phosphorylation of yeast RNA polymerases." Nature 261: 429-431.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." J. Roy. Stat. Soc. B Met 57: 289–300.

Bentley, D. L. (2005). "Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors." Curr Opin Cell Biol 17(3): 251-256.

Bentley, D. L. (2014). "Coupling mRNA processing with transcription in time and space." Nat Rev Genet 15(3): 163-175.

Berro, R., et al. (2008). "CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing." J Virol 82(14): 7155-7166.

Bettayeb, K., et al. (2010). "CDK Inhibitors Roscovitine and CR8 Trigger Mcl-1 Down-Regulation and Apoptotic Cell Death in Neuroblastoma Cells." Genes Cancer 1(4): 369-380.

Bezstarosti, K., et al. (2010). "Differential proteomics based on 18O labeling to determine the cyclin dependent kinase 9 interactome." J Proteome Res 9(9): 4464-4475.

Blazek, D., et al. (2011). "The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes." Genes Dev 25(20): 2158-2172.

Boettiger, A. N., et al. (2011). "Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability." PLoS Comput Biol 7(5): e1001136.

Borchert, G. M., et al. (2006). "RNA polymerase III transcribes human microRNAs." Nat Struct Mol Biol 13(12): 1097-1101.

Bosken, C. A., et al. (2014). "The structure and substrate specificity of human Cdk12/Cyclin K." Nat Commun 5: 3505.

Bowman, B. R., et al. (2003). "Structure of the herpesvirus major capsid protein." Embo j 22(4): 757-765.

Bowman, J. J., et al. (2009). "Transient expression of herpes simplex virus type 1 ICP22 represses viral promoter activity and complements the replication of an ICP22 null virus." J Virol 83(17): 8733-8743.

Bowman, J. J. and P. A. Schaffer (2009). "Origin of expression of the herpes simplex virus type 1 protein U(S)1.5." J Virol 83(18): 9183-9194.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem 72: 248-254.

Brambilla, R. and G. Draetta (1994). "Molecular cloning of PISSLRE, a novel putative member of the cdk family of protein serine/threonine kinases." Oncogene 9(10): 3037-3041.

Britt, B. (2007). Maturation and egress. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. A. Arvin, G. Campadelli-Fiume, E. Mocarski et al. Cambridge.

Brown, N. R., et al. (1999). "The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases." Nat Cell Biol 1(7): 438-443.

Bruni, R., et al. (1999). "A novel cellular protein, p60, interacting with both herpes simplex virus 1 regulatory proteins ICP22 and ICP0 is modified in a cell-type-specific manner and Is recruited to the nucleus after infection." J Virol 73(5): 3810-3817.

Bruni, R. and B. Roizman (1998). "Herpes simplex virus 1 regulatory protein ICP22 interacts with a new cell cycle-regulated factor and accumulates in a cell cycle-dependent fashion in infected cells." J Virol 72(11): 8525-8531.

Buratowski, S. (2009). "Progression through the RNA polymerase II CTD cycle." Mol Cell 36(4): 541-546.

Carter, K. L. and B. Roizman (1996). "The promoter and transcriptional unit of a novel herpes simplex virus 1 alpha gene are contained in, and encode a protein in frame with, the open reading frame of the alpha 22 gene." J Virol 70(1): 172-178.

Chen, F. X., et al. (2015). "PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II." Cell 162(5): 1003-1015.

Cherpes, T. L., et al. (2012). "Neonatal herpes simplex virus infection." Clin Obstet Gynecol 55(4): 938-944.

Chook, Y. M. and G. Blobel (2001). "Karyopherins and nuclear import." Curr Opin Struct Biol 11(6): 703-715.

Chu, X., et al. (2013). "Structural insights into Paf1 complex assembly and histone binding." Nucleic Acids Res 41(22): 10619-10629.

Clark, A. D., et al. (2015). "Mediator kinase module and human tumorigenesis." Crit Rev Biochem Mol Biol 50(5): 393-426.

Cliffe, A. R., et al. (2015). "Neuronal Stress Pathway Mediating a Histone Methyl/Phospho Switch Is Required for Herpes Simplex Virus Reactivation." Cell Host Microbe 18(6): 649-658.

Cliffe, A. R., et al. (2009). "Transcription of the herpes simplex virus latency-associated transcript promotes the formation of facultative heterochromatin on lytic promoters." J Virol 83(16): 8182-8190.

Corden, J. L. (1990). "Tails of RNA polymerase II." Trends Biochem Sci 15(10): 383-387.

Corden, J. L. (2013). "RNA polymerase II C-terminal domain: Tethering transcription to transcript and template." Chem Rev 113(11): 8423-8455.

Core, L. J. and J. T. Lis (2008). "Transcription regulation through promoter-proximal pausing of RNA polymerase II." Science 319(5871): 1791-1792.

Cornelis, S., et al. (2000). "Identification and characterization of a novel cell cycle-regulated internal ribosome entry site." Mol Cell 5(4): 597-605.

Cun, W., et al. (2009). "Transcriptional regulation of the Herpes Simplex Virus 1alpha-gene by the viral immediate-early protein ICP22 in association with VP16." Science in China. Series C, Life sciences / Chinese Academy of Sciences 52(4): 344-351.

Cun, W., et al. (2006). "Structural and functional characterization of herpes simplex virus 1 immediate-early protein infected-cell protein 22." Journal of biochemistry 140(1): 67-73.

Czudnochowski, N., et al. (2012). "Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition." Nat Commun 3: 842.

Davidson, L., et al. (2014). "3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells." Genes Dev 28(4): 342-356.

Davison, A. J. (2007). Comparative analysis of the genomes: Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge, Cambridge University Press.

Davison, A. J. and D. Bhella (2007). Comparative genome and virion structure. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. A. Arvin, G. Campadelli-Fiume, E. Mocarski et al. Cambridge.

Day, D. S., et al. (2016). "Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types." Genome Biol 17(1): 120.

Dembowski, J. A., et al. (2017). "Replication-Coupled Recruitment of Viral and Cellular Factors to Herpes Simplex Virus Type 1 Replication Forks for the Maintenance and Expression of Viral Genomes." PLoS Pathog 13(1): e1006166.

Descostes, N., et al. (2014). "Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells." Elife 3: e02105.

Dickinson, L. A., et al. (2002). "Cyclin L is an RS domain protein involved in pre-mRNA splicing." J Biol Chem 277(28): 25465-25473.

Dieci, G., et al. (2013). "Identification of RNA polymerase III-transcribed genes in eukaryotic genomes." Biochim Biophys Acta 1829(3-4): 296-305.

Dignam, J. D., et al. (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." Nucleic Acids Res 11(5): 1475-1489.

Diner, B. A., et al. (2015). "Interactions of the Antiviral Factor Interferon Gamma-Inducible Protein 16 (IFI16) Mediate Immune Signaling and Herpes Simplex Virus-1 Immunosuppression." Mol Cell Proteomics 14(9): 2341-2356.

Drogat, J., et al. (2012). "Cdk11-cyclinL controls the assembly of the RNA polymerase II mediator complex." Cell Rep 2(5): 1068-1076.

Durand, L. O., et al. (2005). "The carboxyl-terminal domain of RNA polymerase II is phosphorylated by a complex containing cdk9 and infected-cell protein 22 of herpes simplex virus 1." J Virol 79(11): 6757-6762.

Durand, L. O. and B. Roizman (2008). "Role of cdk9 in the optimization of expression of the genes regulated by ICP22 of herpes simplex virus 1." J Virol 82(21): 10591-10599.

Egloff, S., et al. (2009). "Chromatin Structure Is Implicated in "Late" Elongation Checkpoints on the U2 snRNA and β-Actin Genes." Mol Cell Biol 29(14): 4002-4013.

Egloff, S. and S. Murphy (2008). "Role of the C-terminal domain of RNA polymerase II in expression of small nuclear RNA genes." Biochem Soc Trans 36(Pt 3): 537-539.

Egloff, S., et al. (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." Science 318(5857): 1777-1779.

Egloff, S., et al. (2012). "Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes." Mol Cell 45(1): 111-122.

Frampton, A. R., Jr., et al. (2005). "HSV trafficking and development of gene therapy vectors with applications in the nervous system." Gene Ther 12(11): 891-901.

Fraser, K. A. and S. A. Rice (2005). "Herpes simplex virus type 1 infection leads to loss of serine-2 phosphorylation on the carboxyl-terminal domain of RNA polymerase II." J Virol 79(17): 11323-11334.

Fraser, K. A. and S. A. Rice (2007). "Herpes simplex virus immediate-early protein ICP22 triggers loss of serine 2-phosphorylated RNA polymerase II." J Virol 81(10): 5091-5101.

Fu, T. J., et al. (1999). "Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription." J Biol Chem 274(49): 34527-34530.

Gaertner, B. and J. Zeitlinger (2014). "RNA polymerase II pausing during development." Development 141(6): 1179-1183.

Galbraith, M. D., et al. (2010). "CDK8: a positive regulator of transcription." Transcription 1(1): 4-12.

Ganuza, M. and D. Santamaria (2012). "Cdk7: open questions beyond the prevailing model." Cell Cycle 11(19): 3519-3520.

Gibson, W. and B. Roizman (1971). "Compartmentalization of spermine and spermidine in the herpes simplex virion." Proc Natl Acad Sci U S A 68(11): 2818-2821.

Glover-Cutter, K., et al. (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." Mol Cell Biol 29(20): 5455-5464.

Goodfellow, S. J. and J. C. Zomerdijk (2013). "Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes." Subcell Biochem 61: 211-236.

Gopinathan, L., et al. (2011). "Established and novel Cdk/cyclin complexes regulating the cell cycle and development." Results Probl Cell Differ 53: 365-389.

Grana, X., et al. (1994). "PISSLRE, a human novel CDC2-related protein kinase." Oncogene 9(7): 2097-2103.

Green, R. and E. J. Rogers (2013). "Chemical Transformation of E. coli." Methods Enzymol 529: 329-336.

Greifenberg, A. K., et al. (2016). "Structural and Functional Analysis of the Cdk13/Cyclin K Complex." Cell Rep 14(2): 320-331.

Gu, B., et al. (2013). "CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo." Nucleic Acids Res 41(3): 1591-1603.

Gu, J., et al. (2014). "Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4." Cell Cycle 13(11): 1788-1797.

Guen, V. J., et al. (2013). "CDK10/cyclin M is a protein kinase that controls ETS2 degradation and is deficient in STAR syndrome." Proc Natl Acad Sci U S A 110(48): 19525-19530.

Guen, V. J., et al. (2017). "The awakening of the CDK10/Cyclin M protein kinase." Oncotarget.

Guo, J. and D. H. Price (2013). "RNA polymerase II transcription elongation control." Chem Rev 113(11): 8583-8603.

Guo, L., et al. (2012). "Herpes simplex virus 1 ICP22 inhibits the transcription of viral gene promoters by binding to and blocking the recruitment of P-TEFb." PloS one 7(9): e45749.

Gururajan, R., et al. (1998). "Duplication of a genomic region containing the Cdc2L1-2 and MMP21-22 genes on human chromosome 1p36.3 and their linkage to D1Z2." Genome Res 8(9): 929-939.

Hagglund, R., et al. (2002). "U(S)3 protein kinase of herpes simplex virus 1 blocks caspase 3 activation induced by the products of U(S)1.5 and U(L)13 genes and modulates expression of transduced U(S)1.5 open reading frame in a cell type-specific manner." J Virol 76(2): 743-754.

Haider, S. R., et al. (2012). "Tricine-SDS-PAGE." Methods Mol Biol 869: 81-91.

Haller, O., et al. (2006). "The interferon response circuit: induction and suppression by pathogenic viruses." Virology 344(1): 119-130.

Harlen, K. M. and L. S. Churchman (2017). "The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain." Nat Rev Mol Cell Biol 18(4): 263-273.

He, N., et al. (2011). "Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin." Proc Natl Acad Sci U S A 108(36): E636-645.

Heidemann, M. and D. Eick (2012). "Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code." RNA Biol 9(9): 1144-1146.

Heintzman, N. D. and B. Ren (2009). "Finding distal regulatory elements in the human genome." Curr Opin Genet Dev 19(6): 541-549.

Herget, G. W., et al. (2005). "Generalized herpes simplex virus infection in an immunocompromised patient--report of a case and review of the literature." Pathol Res Pract 201(2): 123-129.

Honess, R. W. and B. Roizman (1974). "Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins." J Virol 14(1): 8-19.

Honess, R. W. and B. Roizman (1975). "Regulation of herpesvirus macromolecular synthesis: sequential transition of polypeptide synthesis requires functional viral polypeptides." Proc Natl Acad Sci U S A 72(4): 1276-1280.

Hoskins, A. A. and M. J. Moore (2012). "The spliceosome: a flexible, reversible macromolecular machine." Trends Biochem Sci 37(5): 179-188.

Hsin, J. P., et al. (2014). "RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells." Elife 3: e02112.

Hsin, J. P. and J. L. Manley (2012). "The RNA polymerase II CTD coordinates transcription and RNA processing." Genes Dev 26(19): 2119-2137.

Hsin, J. P., et al. (2011). "RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing." Science 334(6056): 683-686.

Hu, D., et al. (2003). "CDK11 complexes promote pre-mRNA splicing." J Biol Chem 278(10): 8623-8629.

Huang da, W., et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc 4(1): 44-57.

Huang, Z., et al. (2015). "RIP1/RIP3 binding to HSV-1 ICP6 initiates necroptosis to restrict virus propagation in mice." Cell Host Microbe 17(2): 229-242.

Jahedi, S., et al. (1999). "Colocalization of the herpes simplex virus 1 UL4 protein with infected cell protein 22 in small, dense nuclear structures formed prior to onset of DNA synthesis." J Virol 73(6): 5132-5138.

Jang, M. K., et al. (2005). "The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription." Mol Cell 19(4): 523-534.

Johnson, L. N., et al. (1998). "The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases." FEBS Lett 430(1-2): 1-11.

Jonkers, I., et al. (2014). "Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons." Elife 3: e02407.

Jonkers, I. and J. T. Lis (2015). "Getting up to speed with transcription elongation by RNA polymerase II." Nat Rev Mol Cell Biol 16(3): 167-177.

Jurak, I., et al. (2010). "Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2." J Virol 84(9): 4659-4672.

Kalamvoki, M. and B. Roizman (2011). "The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and ICP22." J Virol 85(18): 9472-9477.

Kawai, T. and S. Akira (2006). "Innate immune recognition of viral infection." Nat Immunol 7(2): 131-137.

Kelly, B. J., et al. (2009). "Functional roles of the tegument proteins of herpes simplex virus type 1." Virus Res 145(2): 173-186.

Kemp, L. M. and D. S. Latchman (1988). "Induction and repression of cellular gene transcription during herpes simplex virus infection are mediated by different viral immediate-early gene products." Eur J Biochem 174(2): 443-449.

Knipe, D. M. and A. Cliffe (2008). "Chromatin control of herpes simplex virus lytic and latent infection." Nat Rev Microbiol 6(3): 211-221.

Knipe, D. M., et al. (2013). "Snapshots: chromatin control of viral infection." Virology 435(1): 141-156.

Ko, T. K., et al. (2001). "CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles." J Cell Sci 114(Pt 14): 2591-2603.

Kohoutek, J. and D. Blazek (2012). "Cyclin K goes with Cdk12 and Cdk13." Cell Div 7: 12.

Kornblihtt, A. R., et al. (2004). "Multiple links between transcription and splicing." RNA 10(10): 1489-1498.

Kristie, T. M. (2007). Early events pre-initiation of alphaherpes viral gene expression. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. A. Arvin, G. Campadelli-Fiume, E. Mocarski et al. Cambridge, Cambridge University Press 2007.

Krueger, B. J., et al. (2010). "The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK." PLoS One 5(8): e12335.

Krystof, V., et al. (2012). "Perspective of cyclin-dependent kinase 9 (CDK9) as a drug target." Curr Pharm Des 18(20): 2883-2890.

Kugel, J. F. and J. A. Goodrich (2000). "A kinetic model for the early steps of RNA synthesis by human RNA polymerase II." J Biol Chem 275(51): 40483-40491.

Kurosu, T. and B. M. Peterlin (2004). "VP16 and ubiquitin; binding of P-TEFb via its activation domain and ubiquitin facilitates elongation of transcription of target genes." Curr Biol 14(12): 1112-1116.

Kwak, H. and J. T. Lis (2013). "Control of transcriptional elongation." Annu Rev Genet 47: 483-508.

La Boissiere, S., et al. (1999). "HCF-dependent nuclear import of VP16." Embo j 18(2): 480-489.

Laitem, C., et al. (2015). "CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes." Nat Struct Mol Biol 22(5): 396-403.

Larochelle, S., et al. (2012). "Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II." Nat Struct Mol Biol 19(11): 1108-1115.

Laurencikiene, J., et al. (2006). "RNA editing and alternative splicing: the importance of co-transcriptional coordination." EMBO Rep 7(3): 303-307.

Laybourn, P. J. and M. E. Dahmus (1990). "Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription." J Biol Chem 265(22): 13165-13173.

Leopardi, R., et al. (1997). "Association of herpes simplex virus regulatory protein ICP22 with transcriptional complexes containing EAP, ICP4, RNA polymerase II, and viral DNA requires posttranslational modification by the U(L)13 proteinkinase." J Virol 71(2): 1133-1139.

Levine, M. (2011). "Paused RNA polymerase II as a developmental checkpoint." Cell 145(4): 502-511.

Li, M. L., et al. (2013). "Molecular cloning and characterization of the pseudorabies virus US1 gene." Genetics and molecular research : GMR 12(1): 85-98.

Li, N., et al. (2014). "Cyclin C is a haploinsufficient tumour suppressor." Nat Cell Biol 16(11): 1080-1091.

Li, Q., et al. (2005). "Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186." J Biol Chem 280(31): 28819-28826.

Lidschreiber, M., et al. (2013). "Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II." Mol Cell Biol 33(19): 3805-3816.

Lim, S. and P. Kaldis (2013). "Cdks, cyclins and CKIs: roles beyond cell cycle regulation." Development 140(15): 3079-3093.

Liu, F. and Z. H. Zhou (2007). Comparative virion structures of human herpesviruses. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. A. Arvin, G. Campadelli-Fiume, E. Mocarski et al. Cambridge.

Liu, H. and C. H. Herrmann (2005). "Differential localization and expression of the Cdk9 42k and 55k isoforms." J Cell Physiol 203(1): 251-260.

Liu, Y., et al. (2004). "p42, a novel cyclin-dependent kinase-activating kinase in mammalian cells." J Biol Chem 279(6): 4507-4514.

Lolli, G. and L. N. Johnson (2005). "CAK-Cyclin-dependent Activating Kinase: a key kinase in cell cycle control and a target for drugs?" Cell Cycle 4(4): 572-577.

Long, M. C., et al. (1999). "ICP22 and the UL13 protein kinase are both required for herpes simplex virus-induced modification of the large subunit of RNA polymerase II." J Virol 73(7): 5593-5604.

Loyer, P., et al. (2008). "Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors: influence of cyclin L isoforms on splice site selection." J Biol Chem 283(12): 7721-7732.

Lu, X., et al. (2016). "Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II." Nucleic Acids Res 44(14): 6853-6867.

Luo, Z., et al. (2012). "The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output." Mol Cell Biol 32(13): 2608-2617.

Luo, Z., et al. (2012). "The super elongation complex (SEC) family in transcriptional control." Nat Rev Mol Cell Biol 13(9): 543-547.

Malumbres, M. (2014). "Cyclin-dependent kinases." Genome Biol 15(6): 122.

Malumbres, M., et al. (2009). "Cyclin-dependent kinases: a family portrait." Nat Cell Biol 11(11): 1275-1276.

Mancebo, H. S., et al. (1997). "P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro." Genes Dev 11(20): 2633-2644.

Marshall, N. F. and D. H. Price (1992). "Control of formation of two distinct classes of RNA polymerase II elongation complexes." Mol Cell Biol 12(5): 2078-2090.

Martinez-Argudo, I. and A. J. Blocker (2010). "The Shigella T3SS needle transmits a signal for MxiC release, which controls secretion of effectors." Mol Microbiol 78(6): 1365-1378.

Maruzuru, Y., et al. (2013). "Roles of p53 in herpes simplex virus 1 replication." J Virol 87(16): 9323-9332.

McCracken, S., et al. (1997). "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." Genes Dev 11(24): 3306-3318.

McCracken, S., et al. (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." Nature 385(6614): 357-361.

McGeoch, D. J., et al. (1988). "The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1." J Gen Virol 69 (Pt 7): 1531-1574.

Millevoi, S. and S. Vagner (2010). "Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation." Nucleic Acids Res 38(9): 2757-2774.

Mitchell, C., et al. (1994). "Casein kinase II specifically nucleotidylylates in vitro the amino acid sequence of the protein encoded by the alpha 22 gene of herpes simplex virus 1." Proc Natl Acad Sci U S A 91(25): 11864-11868.

Mocarski, E. S., et al. (2011). "Viral infection and the evolution of caspase 8-regulated apoptotic and necrotic death pathways." Nat Rev Immunol 12(2): 79-88.

Morgan, D. O. and H. L. De Bondt (1994). "Protein kinase regulation: insights from crystal structure analysis." Curr Opin Cell Biol 6(2): 239-246.

Mossman, K. L., et al. (2000). "Herpes simplex virus ICP0 mutants are hypersensitive to interferon." J Virol 74(4): 2052-2056.

Mostafa, H. H. and D. J. Davido (2013). "Herpes simplex virus 1 ICP22 but not US 1.5 is required for efficient acute replication in mice and VICE domain formation." J Virol 87(24): 13510-13519.

Nemec, C. M., et al. (2017). "Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner." Proc Natl Acad Sci U S A 114(20): E3944-e3953.

Nguyen, V. T., et al. (2001). "7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes." Nature 414(6861): 322-325.

Nojima, T., et al. (2015). "Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing." Cell 161(3): 526-540.

O'Toole, J. M., et al. (2003). "Mutation of the protein tyrosine kinase consensus site in the herpes simplex virus 1 alpha22 gene alters ICP22 posttranslational modification." Virology 305(1): 153-167.

Oakes, J. E., et al. (1976). "Genome location of polyadenylated transcripts of herpes simplex virus type 1 and type 2 DNA." Virology 75(1): 145-154.

Ogle, W. O., et al. (1997). "The UL13 protein kinase and the infected cell type are determinants of posttranslational modification of ICP0." Virology 235(2): 406-413.

Ogle, W. O. and B. Roizman (1999). "Functional anatomy of herpes simplex virus 1 overlapping genes encoding infected-cell protein 22 and US1.5 protein." J Virol 73(5): 4305-4315.

Orlando, J. S., et al. (2006). "The products of the herpes simplex virus type 1 immediate-early US1/US1.5 genes downregulate levels of S-phase-specific cyclins and facilitate virus replication in S-phase Vero cells." J Virol 80(8): 4005-4016.

Orlando, J. S., et al. (2006). "ICP22 is required for wild-type composition and infectivity of herpes simplex virus type 1 virions." J Virol 80(19): 9381-9390.

Ou, M. and R. M. Sandri-Goldin (2013). "Inhibition of cdk9 during herpes simplex virus 1 infection impedes viral transcription." PloS one 8(10): e79007.

Pak, V., et al. (2015). "CDK11 in TREX/THOC Regulates HIV mRNA 3' End Processing." Cell Host Microbe 18(5): 560-570.

Paparidis, N. F., et al. (2017). "The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE." Mol Biosyst 13(2): 246-276.

Peng, J., et al. (1998). "Identification of multiple cyclin subunits of human P-TEFb." Genes Dev 12(5): 755-762.

Petesch, S. J. and J. T. Lis (2012). "Overcoming the nucleosome barrier during transcript elongation." Trends Genet 28(6): 285-294.

Phatnani, H. P. and A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." Genes Dev 20(21): 2922-2936.

Poffenberger, K. L., et al. (1994). "A herpes simplex virus type 1 ICP22 deletion mutant is altered for virulence and latency in vivo." Archives of virology 139(1-2): 111-119.

Poffenberger, K. L., et al. (1993). "In vitro characterization of a herpes simplex virus type 1 ICP22 deletion mutant." Virus Genes 7(2): 171-186.

Poon, A. P., et al. (2000). "Posttranslational processing of infected cell protein 22 mediated by viral protein kinases is sensitive to amino acid substitutions at distant sites and can be cell-type specific." J Virol 74(23): 11210-11214.

Poon, A. P. and B. Roizman (2005). "Herpes simplex virus 1 ICP22 regulates the accumulation of a shorter mRNA and of a truncated US3 protein kinase that exhibits altered functions." J Virol 79(13): 8470-8479.

Post, L. and B. Roizman (1981). "A generalized technique for deletion of specific genes in large genomes: alpha gene 22 of herpes simplex virus 1 is not essential for growth." Cell 25(1): 227 - 232.

Prod'hon, C., et al. (1996). "Characterization of regulatory functions of the HSV-1 immediate-early protein ICP22." Virology 226(2): 393-402.

Proudfoot, N. (2004). "New perspectives on connecting messenger RNA 3' end formation to transcription." Curr Opin Cell Biol 16(3): 272-278.

Proudfoot, N. J. (2011). "Ending the message: poly(A) signals then and now." Genes Dev 25(17): 1770-1782.

Rice, S. A. (2011). Multiple Roles of Immediate-Early Protein ICP22 in HSV-1 Replication. Alphaherpesviruses: Molecular Virology. S. K. Weller. UK, Caister Academic Press.

Rice, S. A. and K. A. Fraser (2006). The Modification of Cellular RNA Polymerase II During HSV-1 Infection. Alpha Herpesviruses: Molecular and Cellular Biology. R. M. Sandri-Goldin. UK, Caister Academic Press.

Rice, S. A., et al. (1995). "Herpes simplex virus immediate-early protein ICP22 is required for viral modification of host RNA polymerase II and establishment of the normal viral transcription program." J Virol 69(9): 5550-5559.

Rice, S. A., et al. (1994). "RNA polymerase II is aberrantly phosphorylated and localized to viral replication compartments following herpes simplex virus infection." J Virol 68(2): 988-1001.

Rickert, P., et al. (1999). "Cyclin C/CDK8 and cyclin H/CDK7/p36 are biochemically distinct CTD kinases." Oncogene 18(4): 1093-1102.

Roizman, B., et al. (2005). "The first 30 minutes in the life of a virus: unREST in the nucleus." Cell Cycle 4(8): 1019-1021.

Roizman, B. and D. Knipe (2001). "Herpes Simplex Viruses and Their Replication." Fundamental Fields Virology: p2399 - 2460.

Roizman, B. and G. Zhou (2015). "The 3 facets of regulation of herpes simplex virus gene expression: A critical inquiry." Virology 479-480: 562-567.

Roy, R., et al. (1994). "The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor." Cell 79(6): 1093-1101.

Rutkowski, A. J., et al. (2015). "Widespread disruption of host transcription termination in HSV-1 infection." Nat Commun 6: 7126.

Sandri-Goldin, R. M. (2007). Initiation of transcription and RNA synthesis, processing and transport in HSV and VZV infected cells. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. A. Arvin, G. Campadelli-Fiume, E. Mocarski et al. Cambridge.

Saunders, A., et al. (2006). "Breaking barriers to transcription elongation." Nat Rev Mol Cell Biol 7(8): 557-567.

Schachter, M. M. and R. P. Fisher (2013). "The CDK-activating kinase Cdk7: taking yes for an answer." Cell Cycle 12(20): 3239-3240.

Schneider, E. V., et al. (2011). "The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder." J Mol Biol 412(2): 251-266.

Schulze-Gahmen, U., et al. (2016). "Insights into HIV-1 proviral transcription from integrative structure and dynamics of the Tat:AFF4:P-TEFb:TAR complex." Elife 5.

Schwer, B. and S. Shuman (2011). "Deciphering the RNA polymerase II CTD code in fission yeast." Mol Cell 43(2): 311-318.

Schwyzer, M., et al. (1994). "BICP22 of bovine herpesvirus 1 is encoded by a spliced 1.7 kb RNA which exhibits immediate early and late transcription kinetics." The Journal of general virology 75 (Pt 7): 1703-1711.

Sears, A. E., et al. (1985). "Herpes simplex virus 1 mutant deleted in the alpha 22 gene: growth and gene expression in permissive and restrictive cells and establishment of latency in mice." J Virol 55(2): 338-346.

Sergere, J. C., et al. (2000). "Human CDK10 gene isoforms." Biochem Biophys Res Commun 276(1): 271-277.

Shi, Y., et al. (2009). "Molecular architecture of the human pre-mRNA 3' processing complex." Mol Cell 33(3): 365-376.

Shilatifard, A. (1998). "The RNA polymerase II general elongation complex." Biol Chem 379(1): 27-31.

Shilatifard, A. (2004). "Transcriptional elongation control by RNA polymerase II: a new frontier." Biochim Biophys Acta 1677(1-3): 79-86.

Shore, S. M., et al. (2005). "Characterization of Cdk9(55) and differential regulation of two Cdk9 isoforms." Gene 350(1): 51-58.

Shore, S. M., et al. (2003). "Identification of a novel isoform of Cdk9." Gene 307: 175-182.

Sims, R. J., 3rd, et al. (2004). "Elongation by RNA polymerase II: the short and long of it." Genes Dev 18(20): 2437-2468.

Smith-Donald, B. A. and B. Roizman (2008). "The interaction of herpes simplex virus 1 regulatory protein ICP22 with the cdc25C phosphatase is enabled in vitro by viral protein kinases US3 and UL13." J Virol 82(9): 4533-4543.

Sogaard, T. M. and J. Q. Svejstrup (2007). "Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator." J Biol Chem 282(19): 14113-14120.

Spencer, C. A., et al. (1997). "Repression of host RNA polymerase II transcription by herpes simplex virus type 1." J Virol 71(3): 2031-2040.

Steiner, I. and F. Benninger (2013). "Update on herpes virus infections of the nervous system." Curr Neurol Neurosci Rep 13(12): 414.

Stelz, G., et al. (2002). "Identification of two nuclear import signals in the alpha-gene product ICP22 of herpes simplex virus 1." Virology 295(2): 360-370.

Stevens, J. G. (1989). "Human herpesviruses: a consideration of the latent state." Microbiol Rev 53(3): 318-332.

Stevens, J. G., et al. (1987). "RNA complementary to a herpesvirus alpha gene mRNA is prominent in latently infected neurons." Science 235(4792): 1056-1059.

Tahirov, T. H., et al. (2010). "Crystal structure of HIV-1 Tat complexed with human P-TEFb." Nature 465(7299): 747-751.

Takahashi, H., et al. (2011). "Human mediator subunit MED26 functions as a docking site for transcription elongation factors." Cell 146(1): 92-104.

Tassan, J. P., et al. (1995). "Identification of human cyclin-dependent kinase 8, a putative protein kinase partner for cyclin C." Proc Natl Acad Sci U S A 92(19): 8871-8875.

te Poele, R. H., et al. (1999). "RNA synthesis block by 5, 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) triggers p53-dependent apoptosis in human colon carcinoma cells." Oncogene 18(42): 5765-5772.

Tellier, M., et al. (2016). "The point of no return: The poly(A)-associated elongation checkpoint." RNA Biol 13(3): 265-271.

Tietjen, J. R., et al. (2010). "Chemical-genomic dissection of the CTD code." Nat Struct Mol Biol 17(9): 1154-1161.

Tsigrelis, C., et al. (2006). "Viral opportunistic infections in HIV-infected adults." J Med Liban 54(2): 91-96.

Umbach, J. L., et al. (2008). "MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs." Nature 454(7205): 780-783.

Venkatesh, S. and J. L. Workman (2015). "Histone exchange, chromatin structure and the regulation of transcription." Nat Rev Mol Cell Biol 16(3): 178-189.

Vojnic, E., et al. (2011). "Structure and VP16 binding of the Mediator Med25 activator interaction domain." Nat Struct Mol Biol 18(4): 404-409.

Wagner, E. J. and P. B. Carpenter (2012). "Understanding the language of Lys36 methylation at histone H3." Nat Rev Mol Cell Biol 13(2): 115-126.

Wald, A. and R. Ashley-Morrow (2002). "Serological testing for herpes simplex virus (HSV)-1 and HSV-2 infection." Clin Infect Dis 35(Suppl 2): S173-182.

Wang, Q., et al. (2005). "Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection." Proc Natl Acad Sci USA 102(44): 16055 - 16059.

Whitley, R., et al. (2007). Pathogenesis and disease. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. . C.-F. G. Arvin A, Mocarski E. Cambridge, Cambridge University Press.

Whitley, R. J. (1996). Herpesviruses. Medical Microbiology. S. Baron. University of Texas Medical Branch, Galveston (TX).

Whitley, R. J. (2015). "Herpes Simplex Virus Infections of the Central Nervous System." Continuum (Minneap Minn) 21(6 Neuroinfectious Disease): 1704-1713.

Whittaker, S. R., et al. (2017). "Inhibitors of cyclin-dependent kinases as cancer therapeutics." Pharmacol Ther.

Wohlbold, L., et al. (2006). "The cyclin-dependent kinase (CDK) family member PNQALRE/CCRK supports cell proliferation but has no intrinsic CDK-activating kinase (CAK) activity." Cell Cycle 5(5): 546-554.

Yamada, T., et al. (2006). "P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation." Mol Cell 21(2): 227-237.

Yamaguchi, Y., et al. (1998). "Interplay between positive and negative elongation factors: drawing a new view of DRB." Genes Cells 3(1): 9-15.

Yang, Z., et al. (2005). "Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4." Mol Cell 19(4): 535-545.

Yu, M., et al. (2015). "RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II." Science 350(6266): 1383-1386.

Yuryev, A., et al. (1996). "The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins." Proc Natl Acad Sci U S A 93(14): 6975-6980.

Zaborowska, J., et al. (2014). "Herpes Simplex Virus 1 (HSV-1) ICP22 Protein Directly Interacts with Cyclin-Dependent Kinase (CDK)9 to Inhibit RNA Polymerase II Transcription Elongation." PLoS One 9(9): e107654.

Zaborowska, J., et al. (2016). "The pol II CTD: new twists in the tail." Nat Struct Mol Biol 23(9): 771-777.

Zaborowska, J., et al. (2015). "P-TEFb goes viral." Inside the Cell: Accepted Review.

Zandomeni, R., et al. (1986). "Casein kinase type II is involved in the inhibition by 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole of specific RNA polymerase II transcription." J Biol Chem 261(7): 3414-3419.

Zehring, W. A., et al. (1988). "The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro." Proc Natl Acad Sci U S A 85(11): 3698-3702.

Zhang, D. W., et al. (2012). "Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination." J Biol Chem 287(11): 8541-8551.

Zhang, Y., et al. (2009). "The Paf1 complex is required for efficient transcription elongation by RNA polymerase I." Proc Natl Acad Sci U S A 106(7): 2153-2158.

Zhong, L. and G. S. Hayward (1997). "Assembly of complete, functionally active herpes simplex virus DNA replication compartments and recruitment of associated viral and cellular proteins in transient cotransfection assays." J Virol 71(4): 3146-3160.

Zong, H., et al. (2007). "Cyclin D3/CDK11p58 complex is involved in the repression of androgen receptor." Mol Cell Biol 27(20): 7125-7142.

Zwaagstra, J. C., et al. (1990). "Activity of herpes simplex virus type 1 latency-associated transcript (LAT) promoter in neuron-derived cells: evidence for neuron specificity and for a large LAT transcript." J Virol 64(10): 5019-5028.