Title
Biochemical characterization of the KSHV late gene transcriptional activator complex

Author
Castaneda, Angelica

Publication Date
2019

Peer reviewed|Thesis/dissertation

Biochemical characterization of the KSHV late gene transcriptional activator complex

By

Angelica Flores Castaneda

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Britt A. Glaunsinger, Chair
Professor Laurent Coscoy
Professor Eva Nogales

Spring 2019

Abstract

Biochemical characterization of the KSHV late gene transcriptional activator complex

By

Angelica Flores Castaneda

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Britt A. Glaunsinger, Chair

In the beta- and gammaherpesviruses, a specialized complex of viral transcriptional activators (vTAs) coordinate to direct expression of virus-encoded late genes, which are critical for the production of infectious virions. The vTAs in Kaposi's sarcoma-associated herpesvirus (KSHV) are ORF18, ORF24, ORF30, ORF31, ORF34, and ORF66. While the general organization of the vTA complex has been mapped, the individual roles of these proteins, and how they coordinate to activate late gene promoters, remains largely unknown. Thus, we set out to determine the roles of ORF18, which is a highly interconnected vTA component and of ORF24, which is a putative structural and functional TATA-binding protein (TBP) mimic.

We first performed a comprehensive mutational analysis of the conserved residues in ORF18. The mutants were largely selective for disrupting the interaction with ORF30 but not the other three ORF18 binding partners: ORF31, ORF34, and ORF66. Furthermore, disrupting the ORF18-ORF30 interaction weakened the vTA complex as a whole, and an ORF18 point mutant that failed to bind ORF30 was unable to complement an ORF18 null virus. Thus, contacts between individual vTAs are critical, as even small disruptions in this complex result in profound defects in KSHV late gene expression. These findings underscore how individual interactions between the late gene transcription components are critical for both the stability and function of the complex.

Next, we examined the mechanism by which ORF24 binds RNA polymerase II (Pol II). Previous work in our lab found that ORF24 is a modular protein that binds to viral late gene promoters through its central TBP-like domain. Residues in the N-terminus of ORF24 bind and recruit human Pol II to activate the expression of late genes. We found that ORF24 interacts directly with the heptapeptide repeats of the carboxy terminal domain of Pol II, suggesting that ORF24 may be involved in bringing Pol II to sites of active viral late gene transcription. Collectively, our data highlight how KSHV (and likely other gamma and betaherpesviruses) direct robust transcription of late genes via a unique and streamlined mechanism of pre-initiation complex assembly.

Dedication

To my Mom and Aunt,
Virginia and Maurilia,

I will be forever grateful for the sacrifices you made,
so that your children could have a better life.

To my siblings, John and Alex,
Thank you for your endless
encouragement and support.

TABLE OF CONTENTS

List of Figures and Tables

Acknowledgements

I would like to begin by thanking Britt Glaunsinger for giving me the opportunity to perform my graduate studies in her lab. Britt, you are truly the best mentor anyone could ask for. Every time that I had doubts about my abilities, you believed in me even when I didn't believe in myself, which to be honest was probably 99% of the time. Your endless encouragement allowed me to believe that I could excel as a scientist, which is truly the one thing I've wanted to do since I was a kid. Thank you for the countless scientific and academic opportunities you provided throughout my graduate career. I will be forever grateful for everything you taught me and consider myself very fortunate to have done my graduate training with you.

The first people I worked with in Britt's lab were Mandy Muller and our lab manager, Jennifer Blancas. I remember my first days in the Glaunsinger lab, and the thing that stands out the most is that Mandy and Jennifer were always willing to answer questions and did so without judgement even though I had no idea what I was doing. Mandy and Jennifer, I always looked forward to our daily post-lunch coffee breaks, where you always provided me with much needed sanity-checks. Thank you for listening to me complain about my many experimental failures. Jennifer, I don't know how the lab would function without you. Mandy, I have no doubt that you will be an amazing PI and your students/postdocs are incredibly lucky to work with you.

Shortly after passing my qualifying exam, I moved to what has come to be known as the Protein Bay and sat next to Matthew Gardner for the majority of my time in the Glaunsinger lab. Matt and I became close friends through our shared love of immature humor and comradery over many failed protein purifications. One of the things that stands out the most during my time in the Protein Bay is that although there were many passersby, no one committed to sitting with us longer than necessary. I choose to believe the reason for this is that Matt and I shared a special bond. Matt, thank you for always being willing to help and for (almost always) providing excellent suggestions for protein purifications. I miss our Monday morning conversations about GoT, especially now that Season 8 has gone off the rails.

The one passerby who was up for the challenge of becoming a member of the Protein Bay is Allison Didychuk. Of the qualities that stand out the most about Allison are her intelligence, witty sense of humor, and excellent ideas. In the short amount of time I was lucky enough to spend with Allison in the Glaunsinger lab, she was always willing to share her expertise on structural biology and never made me feel bad for not knowing what I was talking about. Lady Arbitress of Rules, I will forever consider myself lucky to have had the opportunity to work with you. Thank you for always being willing to talk about science (and countless other topics) and for listening to me complain about my doubts and insecurities, without judgement. I hope that I can be half as good a friend to you, as you have been to me.

To everyone in the Glaunsinger and Coscoy labs, thank you for being a supportive group of people who made coming to work enjoyable. I have no doubt that you will go on to do amazing things.

Last, but not least, I would like thank Elena Garcia, Justin Lee, Virginia Ng, Melissa Moore, Bridget Hansen, and Noah Swanson. I consider myself incredibly lucky to have you all in my life. Your friendship means the world to me.

Chapter 1: Introduction

Human Herpesviruses

Herpesviruses are large, enveloped, double-stranded DNA (dsDNA) viruses that infect a wide range of organisms and establish lifelong infections in their hosts. The structure of these viruses consists of linear DNA packed into an icosahedral capsid surrounded by a tegument layer composed of proteins and RNA. The tegument layer is protected by a lipid envelope decorated with glycoproteins. Of the over 90 different herpesviruses discovered thus far, only eight infect humans (Davison, 2010). The human herpesviruses are divided into three subfamilies: alpha, beta, and gammaherpesviruses, based on their genome sequence and tissue tropism. The members of the alphaherpesvirus subfamily are neurotropic viruses that infect mucoepithelial cells and establish latency in sensory neurons. Within the alphaherpesvirus subfamily are varicella zoster virus (VZV), which causes chicken pox and shingles, herpes simplex virus-1 (HSV-1) and herpes simplex virus-2 (HSV-2) which are the causative agents of oral and genital herpes, respectively. Compared to the other two subfamilies of human herpesviruses, alphaherpesviruses exhibit faster replication kinetics. Members of the betaherpesvirus subfamily include human cytomegalovirus (HCMV), human herpesviruses 6 (HHV-6), and HHV-7, which infect epithelial cells and establish latency in monocytes and lymphocytes. Of these viruses, HCMV is the only medically relevant member as it can be vertically transmitted during pregnancy and cause severe congenital birth defects (Carlson et al., 2010). The gammaherpesvirus subfamily includes the opportunistic pathogens Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV), both of which infect epithelial cells and establish latency in lymphocytes. EBV is highly prevalent in the adult population and is the cause of most cases of infectious mononucleosis, while KSHV is the etiologic agent of Kaposi's sarcoma (KS).

Kaposi's Sarcoma-associated Herpesvirus

Prior to the AIDS epidemic, Kaposi's sarcoma was thought of as a passive disease endemic to the Mediterranean basin and Africa affecting only the skin and present most often in elderly Mediterranean men (Ganem, 2006). With the rise of HIV and AIDS, Kaposi's sarcoma became more prevalent and is considered the AIDS-defining illness of the 1980's. The etiologic agent of Kaposi's sarcoma, KSHV (HHV-8) was eventually discovered in KS tissue as DNA sequences with distant homology to EBV (Chang et al., 1994). Today, KSHV remains a leading cause of cancer in sub-Saharan Africa and in immunocompromised individuals, including patients with untreated AIDS. In addition to Kaposi's sarcoma, KSHV is the etiologic agent of the B-cell cancers Multicentric Castleman's disease and primary effusion lymphoma.

The KSHV genome consists of a long unique coding region, which generates approximately 90 open reading frames (ORFs), the majority of which are expressed during lytic replication, flanked on either side by GC-rich terminal repeats. The coding region and terminal repeats combined generate a genome that is approximately 165 kb. Each terminal repeat contains a minimal replication element (MRE) required for replication of the viral genome during latency (Ballestas and Kaye, 2001; Shrestha and Sugden, 2014). Two additional DNA elements on the viral genome are the origins of lytic replication (OriLyt) designated as the left (OriLyt-L) and right (OriLyt-R) origins (AuCoin

et al., 2002; Lin et al., 2003), which are required for DNA replication during the lytic phase of the viral lifecycle. Since KSHV is the newest human herpesvirus to be identified, the function of the ORFs has been mostly inferred from their homology to proteins in other herpesviruses (Chang et al., 1994). However, studies focused on identifying the interactions of the KSHV proteome with host cell proteins have begun to elucidate the potential function of previously uncharacterized KSHV ORFs (Davis et al., 2015).

The KSHV Lifecycle

All herpesviruses establish lifelong latent infections and sporadically reactivate to maintain a reservoir of latently infected cells. As with all herpesviruses, the KSHV lifecycle begins with entry of the virus into target cells. While the infected cell type and tissue tropism differs for each virus, the steps that mediate infection are similar. Entry into cells is initiated by glycoproteins on the viral envelope binding cell surface receptors which leads to endocytosis of the virion (Kumar and Chandran, 2016). In KSHV, gB is the major envelope glycoprotein involved in initiating binding and entry by interacting with heparin sulfate on the surface of the host cell (Akula et al., 2001a, 2001b). Upon viral entry and endosomal release of the tegument and capsid into the cytoplasm, tegument components begin to modulate the host cell, while the capsid is trafficked to the nucleus. At the nuclear membrane, the linear genome is released into the nucleus where replication and transcription will occur.

Once in the nucleus, viral DNA is circularized by host enzymatic machinery and the resulting plasmid is rapidly chromatinized leading to the silencing of most viral gene expression (Tempera and Lieberman, 2010). Silencing of the viral genome causes latency, which is generally the default pathway for KSHV, particularly in tissue culture. The chromatinized viral episome is tethered to the host genome through the latency-associated nuclear antigen (LANA) (Uppal et al., 2014). LANA plays an important role in ensuring the episome is passively replicated by host DNA replication machinery and in the distribution of replicated episomes to daughter cells. During latency, only a handful of proteins and 12 miRNAs are expressed. Their primary purpose is to modulate host cell signaling pathways and maintain the viral genome in dividing cells.

As with all DNA viruses, during the lytic phase of the viral lifecycle, genes are expressed in an ordered temporal cascade that begins with the expression of early genes, followed by delayed early genes, and ends with late genes. The switch from latency to lytic reactivation is not entirely understood, but requires the expression of ORF50, also referred to as the replication and transcriptional activator (RTA) (Lukac et al., 1998; Sun et al., 1998). RTA is a master switch that transactivates its own promoter and induces the transcription of genes containing an RTA-responsive element. Generally, immediate early gene products are involved in modulating the host cell response to lytic infection and are required for the expression of delayed early genes. The products of delayed early genes further shape the host cell response as demonstrated by ORF37, also referred to as shut-off and exonuclease (SOX). SOX cleaves mRNAs, which are then degraded by cellular exonucleases (Covarrubias et al., 2011; Glaunsinger and Ganem, 2004). This activity is important for transcriptional inhibition of the cellular genome while allowing for continued transcription of the viral genome (Abernathy et al., 2015). Other examples of delayed early gene products are proteins required for the replication of the viral genome during the lytic phase such as the viral polymerase (ORF9), processivity factor (ORF59), primase

(ORF40, ORF41), helicase (ORF56, ORF44), and single-stranded DNA binding protein (ORF6).

During DNA replication, histones are removed from the viral genome and newly synthesized genomes are not chromatinized. Furthermore, DNA replication generates viral replication compartments, which push the cellular genome to the periphery of the nucleus (Schmid et al., 2014). Their purpose may be to sequester proteins needed for efficient replication and transcription of the viral genome. The expression of late genes requires viral DNA replication as demonstrated by the late gene defect in all instances where DNA replication is inhibited. Late genes are mainly structural components of the viral particle such as capsid proteins and membrane glycoproteins, which are required to make infectious virions.

The assembly of viral particles begins once sufficient late gene products accumulate in the nucleus. To begin this process, capsid proteins assemble around scaffolding proteins. After the capsid is made, the terminase brings newly replicated DNA to the capsid and threads the DNA into the capsid until it encounters the terminal repeats at which point the terminase cleaves the genome (Heming et al., 2017). This generates a capsid with one linear copy of the viral genome. After DNA is packaged, the nucleocapsid acquires tegument proteins and buds through the nucleus into the cytoplasm where it is trafficked to the Golgi. In the Golgi, the nucleocapsid acquires an envelope and membrane-bound tegument proteins. The virion exits the Golgi in a secretory vesicle which is brought to the cell surface and exits the cell by fusion of the vesicle with the cell membrane (Aneja and Yuan, 2017).

Despite the large coding capacity of KSHV and other herpesviruses, they do not encode a viral RNA polymerase and instead utilize host transcriptional machinery for expression of all viral genes. Since transcription initiation is at the core of this thesis, the following section summarizes the steps in eukaryotic transcription pre-initiation that are co-opted by all members of the human herpesviruses.

Transcription Initiation of Protein Coding Genes

In eukaryotes, transcription is an intricate process that involves the coordination of several protein complexes to activate the expression of genes at specific times during the cell cycle. The activation of transcription is dependent on the formation of a pre-initiation complex (PIC), which is composed of RNA polymerase II (Pol II), Mediator, and general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Thomas and Chiang, 2006). The PIC assembles at the core promoter, which contains *cis*-regulatory elements that serve to both regulate the function of the promoter and for the proper assembly and orientation of the PIC. Among the core promoter elements is the TATA box, which is an A/T rich sequence located 25-30 base pairs (bp) upstream of the transcription start site (**Figure 1.1A**). Surrounding the transcription start site is the initiator (Inr) element, which contains a pyrimidine rich sequence. An additional core promoter element is the downstream promoter element (DPE), which in humans is located at +29 to +35 relative to the transcription start site. Two additional elements located downstream of the transcription start site are the motif ten element (MTE) and the downstream core element (DCE), which are mutually exclusive and serve to enhance Pol II-directed transcription. Finally, the TFIIB recognition element can be located immediately upstream (BRE^u) or downstream (BRE^d) of the TATA box. It should be noted that promoters do not contain all

of the core promoter elements, instead they have different combinations depending on the mode of gene regulation. Furthermore, a bioinformatic study of human genes found that less than 22% of genes contain a TATA box and are regulated by a combination of other core promoter elements (Gershenzon and Ioshikhes, 2005).

To initiate PIC formation, TFIID, which is composed of TATA-binding protein (TBP) in complex with 13 TBP-associated factors (TAFs), binds the TATA box at the core promoter (Thomas and Chiang, 2006). In promoters without a TATA box, TAFs can bind the Inr, DPE, and DCE which allows TFIID to recognize the core promoter (**Figure 1.1A**). Binding of TFIID recruits additional GTFs beginning with TFIIA and TFIIB, which stabilize the TBP-TATA complex (Orphanides et al., 1996). An additional role of TFIIB is to recruit Pol II, which is brought to the PIC with TFIIF. Following this, TFIIF recruits TFIIE and finally TFIIH to form the basal transcription PIC (**Figure 1.1B**). The GTFs have distinct roles in PIC formation and transcription initiation. TFIIF functions with TFIIB and Pol II in start site selection while TFIIE and TFIIH are involved in promoter opening (Holstege et al., 1996; Sainsbury et al., 2015). The enzyme at the core of the PIC, Pol II, is composed of 12 subunits, Rpb1-Rpb12, designated by the decreasing order of their molecular mass (Young, 1991). The catalytic site of Pol II performs two-metal ion catalysis and is located at the interface between Rpb1 and Rpb2 (Wang et al., 2006). The largest subunit of Pol II, Rpb1, has a long carboxy terminal domain (CTD) which, in humans, is composed of 52 heptapeptide repeats, the consensus of which is YSPTSPS. The primary form of CTD modification during transcription is phosphorylation of the serine residues to regulate transcription initiation, elongation, and termination (Harlen and Churchman, 2017).

An additional protein complex required for expression of most Pol II transcripts is Mediator, which in humans is composed of 26 subunits. One of the main functions of Mediator is to communicate signals from gene specific transcription factors to the PIC at core promoters (Allen and Taatjes, 2015). Mediator subunits were initially identified in a screen for factors essential for yeast viability (Myers and Kornberg, 2000) and in screens to identify proteins interacting with the CTD of Pol II (Nonet and Young, 1989; Thompson et al., 1993). Mediator has roles in PIC assembly, Pol II initiation, Pol II pausing, Pol II elongation, and Pol II re-initiation. During PIC assembly, Mediator regulates recruitment of TFIIA, TFIIB, TFIIE, and TFIIF within promoter-bound PICs (Baek et al., 2006; Jishage et al., 2012; Johnson and Carey, 2003).

Once the PIC is fully formed, Pol II is directed to initiate transcription by phosphorylation of its CTD by TFIIH. The ability of TFIIH to phosphorylate the CTD is in part dependent on Mediator (Boeing et al., 2010). The role of Mediator in Pol II pausing and elongation is not as well understood as it is in other transcriptional events. Nevertheless, Mediator physically or functionally interacts with proteins known to regulate Pol II pausing and elongation, such as POLR2M, which stabilizes Pol II pausing, and the super elongation complex (SEC), which is a general regulator of Pol II elongation (Adelman and Lis, 2012; Luo et al., 2012). A key regulatory stage at active genes is transcription re-initiation (Larson et al., 2013). One benefit of transcription re-initiation is that is occurs faster than the initial round of transcription (Sandaltzopoulos and Becker, 1998). PIC scaffolding assemblies, which retain most of the PIC components, including the DNA-binding GTFs and Mediator, remain at the core promoter following promoter escape and facilitate the recruitment of Pol II and TFIIF to re-initiate transcription at active genes (Hahn et al., 2000). Finally, Mediator is involved in the formation of long-range

looping interactions between enhancer and promoter sequences that are important for driving high-level and cell type-specific gene expression (Deng et al., 2012; Sanyal et al., 2012).

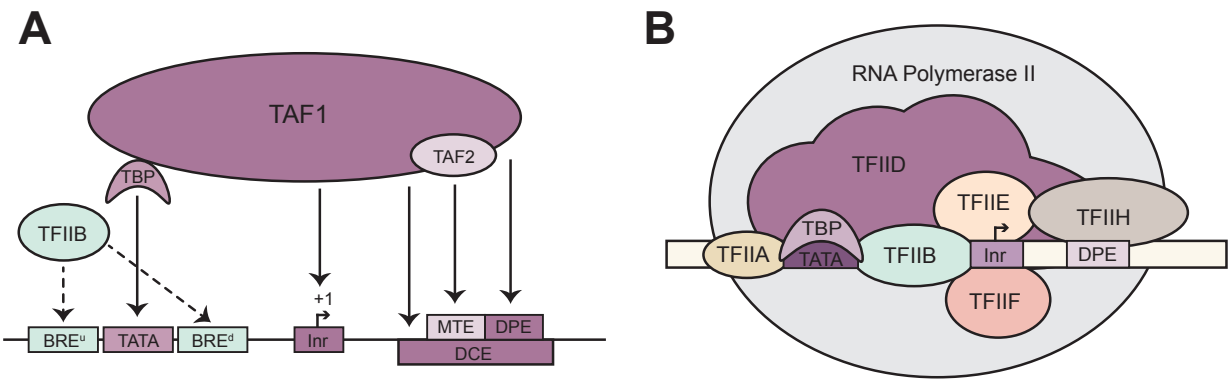


Figure 1.1. Eukaryotic core promoter elements and the basal transcription pre-initiation complex. (A) Schematic depicting eukaryotic core promoter elements and the transcription factors that bind these elements. Among the core elements are BRE^u and BRE^d (TFIIB recognition element, upstream and downstream), TATA box, Inr (Initiator), MTE (motif ten element), DPE (downstream promoter element), and DCE (downstream core element). **(B)** Diagram of the Pol II transcription pre-initiation complex. The stepwise assembly of the complex begins with binding of TFIID at the core promoter. TFIIA and TFIIB stabilize the TFIID-TATA complex and allow recruitment of Pol II associated with TFIIF. Finally, TFIIE and TFIIH are recruited to the core promoter to form a pre-initiation complex for basal transcription.

Late Gene Transcription in dsDNA Viruses

In all dsDNA viruses, late genes are expressed only after replication of the viral genome. While this feature is conserved, viruses have evolved different mechanisms to regulate the temporal regulation of late gene expression. For example, in human papillomavirus (HPV) late gene transcription is regulated, in part, by *cis*-acting elements in the viral genome and chromatin remodeling of the late promoter that occurs during cell differentiation (Bodily and Meyers, 2005; Grassmann et al., 1996; Del Mar Pena and Laimins, 2002). Primate polyomaviruses encode a large T-antigen, which binds the viral origin of replication and also inactivates cell cycle control proteins causing cells to enter S phase where replication of both the cellular and viral genomes occurs (Chapa et al., 2017; Lynch and Frisque, 1990). Following viral genome replication, T-antigen promotes late gene transcription by blocking expression of early genes (Myers et al., 1981) and by stabilizing TFIIA-TBP complexes bound at late promoters thus taking the role of a TAF protein (Damania and Alwine, 1996; Damania et al., 1998)

These examples illustrate a common theme in the regulation of late gene transcription, which is the requirement of early viral proteins, *cis*-acting elements in the viral genome, or both. The next section summarizes what is known about late gene

regulation in herpesviruses and illustrates the differences in late gene transcription between the human herpesvirus subfamilies.

Late Gene Transcription in Alphaherpesviruses

Gene regulation in alphaherpesviruses is best characterized in HSV-1. Studies in HSV-1 revealed that late gene promoters lack distal and far upstream *cis*-regulatory elements; however, the origin of lytic replication is required in *cis* for late gene transcription (Johnson and Everett, 1986). HSV-1 late gene promoters contain a TATA box along with a sequence that resembles an Inr element (Guzowski and Wagner, 1993; Huang and Wagner, 1994). In HSV-1, the main protein responsible for the activation of transcription from late promoters is infected-cell polypeptide 4 (ICP4), which has roles in both early and late transcription. ICP4 contains two transactivation domains, one in the N-terminus and the other in the C-terminus, both of which are required for late gene expression (DeLuca and Schaffer, 1988); however, the C-terminal transactivation domain is dispensable for early gene expression. The N-terminal transactivation domain of ICP4 interacts with TBP and Mediator (Lester and DeLuca, 2011), while the C-terminal domain interacts with TAF1 (Carrozza and DeLuca, 1996). Both of these interactions stabilize the PIC by enhancing the binding of TFIID to the promoter (Grondin and DeLuca, 2000). During the transcription of early genes and in eukaryotic transcription, the interaction between TBP and the promoter is stabilized by TFIIA. However, in HSV-1 the expression of TFIIA decreases significantly during the lytic cycle (Zabierowski and DeLuca, 2004), and the TFIID-promoter interaction is stabilized by ICP4 instead of TFIIA.

Another protein shown to be involved in the transcription of late genes in HSV-1 is ICP27. The exact function of ICP27 in late gene transcription is not fully understood; however, ICP27 interacts with the CTD of Pol II (Dai-Ju et al., 2006). The interaction between ICP27 and the CTD is mediated by both the N-terminus and C-terminus of ICP27. Early studies suggest ICP27 interacts with ICP4 (Panagiotidis et al., 1997), and this interaction may serve as a mechanism by which to ensure that Pol II is brought to sites of viral gene transcription.

Late Gene Transcription in Beta and Gammaherpesviruses

In human herpesviruses, the OriLyt is required in *cis* for robust late gene transcription (Djavadian et al., 2016; Nandakumar and Glaunsinger, 2019). Furthermore, as the viral replication cycle progresses from early to late gene expression, promoters become less complex. The reduction in complexity leads to a decrease in the length of core promoters, which in KSHV has been mapped to 12-15 base pairs (bp) (Tang et al., 2004). In the short core promoter of late genes, beta and gammaherpesviruses have a modified TATA box where the adenine at the fourth position is a thymine (TATT) (Djavadian et al., 2018; Nandakumar and Glaunsinger, 2019). This promoter motif is located ~30 bp upstream from the transcription start site and was first discovered in EBV (Serio et al., 1998) and was later identified in murine gammaherpesvirus 68 (MHV68) and KSHV (Tang et al., 2004; Wong-Ho et al., 2014). In the characterizations of this promoter sequence in EBV and KSHV, mutation of the nucleotide at the fourth position from T to A led to an inability to activate the transcription of late genes (Davis et al., 2016; Gruffat et al., 2012).

Insights into late gene regulation in beta and gammaherpesviruses began with the identification of essential genes in MHV68 (Song et al., 2005). Among these were genes later determined to be required for late gene expression in MHV68, which are conserved in both beta and gammaherpesviruses and are referred to as viral transcriptional activators (vTAs): ORF18, ORF24, ORF30, ORF31, and ORF34 (Arumugaswami et al., 2006; Jia et al., 2004; Wong et al., 2007; Wu et al., 2009). Later studies confirmed the requirement of these vTAs for late gene transcription in KSHV (Brulois et al., 2015; Davis et al., 2015; Gong et al., 2014; Nishimura et al., 2017). The set of conserved viral proteins needed for late gene expression was expanded when the EBV protein BFRF2 (homolog of KSHV ORF66) was identified and determined to be necessary for late gene expression (Aubry et al., 2014). In a transfection based luciferase reporter assay, BFRF2 and the EBV homologs of the other vTAs activated an EBV late gene promoter (Aubry et al., 2014). The same was true in a transfection based assay with the KSHV homologs of the vTAs (Davis et al., 2016). The discovery that a reporter gene under the control of a late gene promoter could be activated in a transfection based system was important because it provided a tool to investigate the effect of specific vTA mutations on the expression of late genes.

While the proteins required for late gene transcription are known, their functions are poorly understood. Experiments that examined the pairwise interactions between the vTAs by either co-immunoprecipitation or split-luciferase demonstrated that the proteins were interconnected (Brulois et al., 2015; Davis et al., 2016; Nishimura et al., 2017). Additionally, co-immunoprecipitation experiments in EBV and murine cytomegalovirus (MCMV) revealed that the vTAs form a complex, the organization of which is different between KSHV (gammaherpesvirus) and MCMV (betaherpesvirus) (**Figure 1.2A**) (Aubry et al., 2014; Davis et al., 2016; Pan et al., 2018). This suggests that although the vTAs are conserved in both beta and gammaherpesviruses, their functions may be different.

Elucidating the individual roles of these proteins has been challenging. The most well studied protein in the complex is ORF24. An *in silico* analysis aimed at identifying proteins that resembled TATA-binding protein (TBP) predicted that EBV BcRF1 and HCMV UL79, both of which are homologs of ORF24, contain a TBP-like domain (Wyrwicz and Rychlewski, 2007). In KSHV, mutations to the predicted DNA-binding residues in ORF24 led to a decrease in the chromatin immunoprecipitation (ChIP) signal of ORF24 at late gene promoters and a failure to transcribe late genes (Davis et al., 2015). In addition to DNA-binding, a co-immunoprecipitation mass spectrometry (IP-MS) based approach to identify interactions between viral and cellular proteins determined that ORF24 interacts with Pol II (Davis et al., 2015). This interaction requires three leucine residues in the N-terminus of ORF24 and is necessary for late gene transcription (Davis et al., 2015). Furthermore, ChIP experiments to determine the GTF composition at late gene promoters in KSHV only identified TFIIB and TFIIH (**Figure 1.2B**) (Davis et al., 2015).

As for the proposed functions of the other vTAs, the CMV homolog of KSHV ORF18, UL79, interacts with Pol II and is required for the accumulation of viral transcripts during CMV infection (Perng et al., 2011). The KSHV protein ORF34 interacts with four other proteins in the vTA complex and is predicted to be a hub for the assembly of ORF18, ORF66, and ORF31 thus mediating the interaction between these vTAs and ORF24 (Davis et al., 2016; Nishimura et al., 2017). In KSHV, when the interaction between

ORF34 and ORF24 is disrupted by a single amino acid mutation in ORF24, the virus fails to transcribe late genes (Davis et al., 2016) suggesting that the ability of ORF34 to interact with ORF24 is crucial for its function. These findings suggest that the vTAs act together to mediate the expression of late genes in beta and gammaherpesvirus; however, the exact mechanism by which they activate late gene transcription is poorly understood.

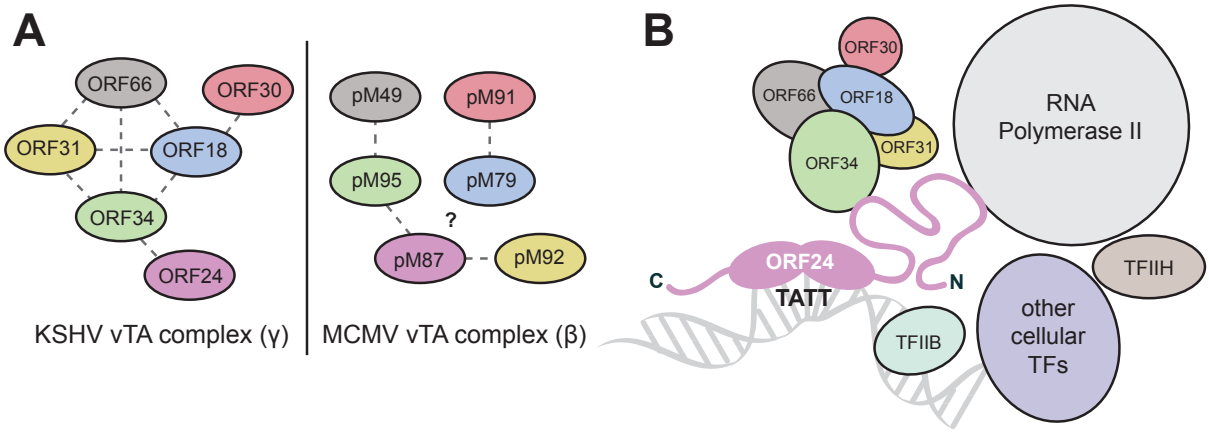


Figure 1.2. Interactions between late gene transcriptional activators in beta and gammaherpesviruses. (A) Diagram of the vTA interactions in KSHV (left) and MCMV (right). The homologs are shown in the same color to illustrate the differences in the predicted organization of the two complexes. **(B)** Illustration of the proteins predicted to localize to a late gene pre-initiation complex in KSHV.

Thesis Overview

This thesis aims to elucidate how the intermolecular interactions between the vTAs are mediated to bring together a set of viral proteins that are at least in part responsible for the transcription of late genes in KSHV.

In the first part, we focus on ORF18 and identify its vTA binding partners. We determine that ORF18 interacts with the majority of the proteins in the vTA complex, and we define residues that are critical for its interaction with ORF30. We find that the interaction between ORF18 and ORF30 contributes to the stability of the complex as a whole and is essential for the late gene expression in KSHV.

The second part of this thesis describes the work done to characterize the interaction between the TBP-mimic ORF24 and Pol II. With the use of single particle negative stain electron microscopy, pulldown assays, and NMR spectroscopy we determine that ORF24 interacts directly with the CTD repeats of Pol II. We discuss some of the implications of this interaction and propose ways to determine the structure of the ORF24 domain that mediates the interaction with the CTD.

Collectively, this work advances our understanding of the intermolecular interactions mediated by the vTA complex both between members of the complex and with cellular proteins.

Chapter 2: The interaction between ORF18 and ORF30 is required for late gene expression in Kaposi's sarcoma-associated herpesvirus

Introduction

A broadly conserved feature of the lifecycle of dsDNA viruses is that replication of the viral genome licenses transcription of a specific class of viral transcripts termed late genes. There is an intuitive logic behind this coupling, as late genes encode proteins that participate in progeny virion assembly and egress, and thus are not needed until newly synthesized genomes are ready for packaging. Additionally, late gene transcription requires ongoing DNA replication, and in the gammaherpesviruses Kaposi's sarcoma-associated herpesvirus (KSHV) and Epstein-Barr virus (EBV) the increase in template abundance appears insufficient to explain the robust transcription of late genes whose products are required in large amounts (Li et al., 2018).

While the mechanisms underlying late gene activation can vary across viral families, in the beta- and gammaherpesviruses, late gene promoters are strikingly minimalistic and primarily consist of a modified TATA box (TATT) and ~10-15 base pairs of variable flanking sequence (Djavadian et al., 2018; Serio et al., 1998; Tang et al., 2004; Wong-Ho et al., 2014). Despite this sequence simplicity, their transcription requires a dedicated set of at least six conserved viral transcriptional activators (vTAs) whose precise roles are only beginning to be uncovered. In KSHV, the vTAs are encoded by open reading frames (ORFs) 18, 24, 30, 31, 34, and 66 (Arumugaswami et al., 2006; Aubry et al., 2014; Brulois et al., 2015; Davis et al., 2015, 2016; Gong et al., 2014; Jia et al., 2004; Nishimura et al., 2017; Wong et al., 2007; Wu et al., 2009). The best characterized of the vTAs is a viral TATA-binding protein (TBP) mimic, encoded by ORF24 in KSHV, which binds both the late gene promoter and RNA polymerase II (Pol II) (Davis et al., 2015; Gruffat et al., 2012; Wong-Ho et al., 2014; Wyrwicz and Rychlewski, 2007). Beyond the viral TBP mimic, the only other vTA with a documented transcription-related function is pUL79 (homologous to KSHV ORF18) in human cytomegalovirus (HCMV), which promotes transcription elongation at late times of infection (Perng et al., 2014). Roles for the remaining vTAs remain largely elusive, although the KSHV ORF34 protein and its murine cytomegalovirus (MCMV) homolog pM95 may function as hub proteins, as they interact with numerous other vTAs (Davis et al., 2016; Nishimura et al., 2017; Pan et al., 2018). In addition to the six conserved vTAs, in EBV, the kinase activity of BGLF4 (homologous to KSHV ORF36) also contributes to the expression of late genes (El-Guindy et al., 2014; McKenzie et al., 2016).

Studies in both beta-and gammaherpesviruses indicate that the vTAs form a complex, the general organization of which has been mapped in MCMV and KSHV (Aubry et al., 2014; Davis et al., 2016; Nishimura et al., 2017; Pan et al., 2018) (**Figure 2.1A**). Notably, several recent reports demonstrate that specific interactions between the vTAs are critical for late gene transcription. In MCMV, mutation of conserved residues in pM91 (homologous to KSHV ORF30) that disrupt its interaction with pM79 (homologous to KSHV ORF18) renders the virus unable to transcribe late genes (Pan et al., 2018). Similarly, the interaction between KSHV ORFs 24 and 34 can be abrogated by a single amino acid mutation in ORF24 which prevents late gene transcription (Davis et al., 2016). Further delineating these contacts should provide foundational information relevant to understanding vTA complex function.

The precise role of KSHV ORF18 in late gene transcription remains unknown, however it is predicted to interact with four of the five other vTAs (ORFs 30, 31, 34, and 66), suggesting that—like ORF34—it may play a central role in vTA complex organization (Davis et al., 2016; Nishimura et al., 2017). Here, we performed an interaction screen of ORF18 mutants to comprehensively evaluate the roles of its conserved residues in mediating pairwise vTA binding. We reveal that ORF30 is particularly sensitive to mutation in ORF18, enabling isolation of mutants that selectively abrogate this interaction while retaining the contacts between ORF18 and the other vTAs. Disrupting the ORF18-ORF30 interaction not only prevents KSHV late gene transcription as measured by K8.1 expression, but also appears to weaken assembly of the remaining vTA complex. These findings underscore the key role that ORF18 plays in late gene transcription and suggest that disrupting just one of its interactions has a destabilizing effect on the vTA complex as a whole.

Results

ORF18 interacts with ORF30, ORF31, ORF34, and ORF66

Previous work using a split luciferase-based interaction screen suggested that ORF18 is highly interconnected with other viral late gene activators, as it interacted with the majority of the proteins in the viral transcription pre-initiation complex (vPIC) (Davis et al., 2016). To independently confirm the binding of ORF18 to ORFs 30, 31, 34, and 66 in the absence of other viral factors, we assessed its ability to co-immunoprecipitate (co-IP) with each of these vTAs in transfected HEK293T cells. Consistent with the screening data, ORF18-3xFLAG interacted robustly with C-terminal 2xStrep-tagged versions of ORF30, ORF31, and ORF66 (**Figure 2.1B-D**). Although ORF18-3xFLAG did not interact with ORF34 tagged on its C-terminus, the interaction was recovered upon moving the 2xStrep tag to the N-terminus of ORF34 (**Figure 2.1E**). Furthermore, we consistently observed that the expression of ORF30 was higher when co-expressed with ORF18 than when transfected with vector control, suggesting that ORF18 may stabilize ORF30 (**Figure 2.1B**). To determine whether ORF18 had a stabilizing effect on ORF30, the half-life of ORF30 was measured in both the presence and absence of ORF18. As can be seen in **Figure 2.1F**, ORF30 stability increased significantly upon co-expression with ORF18.

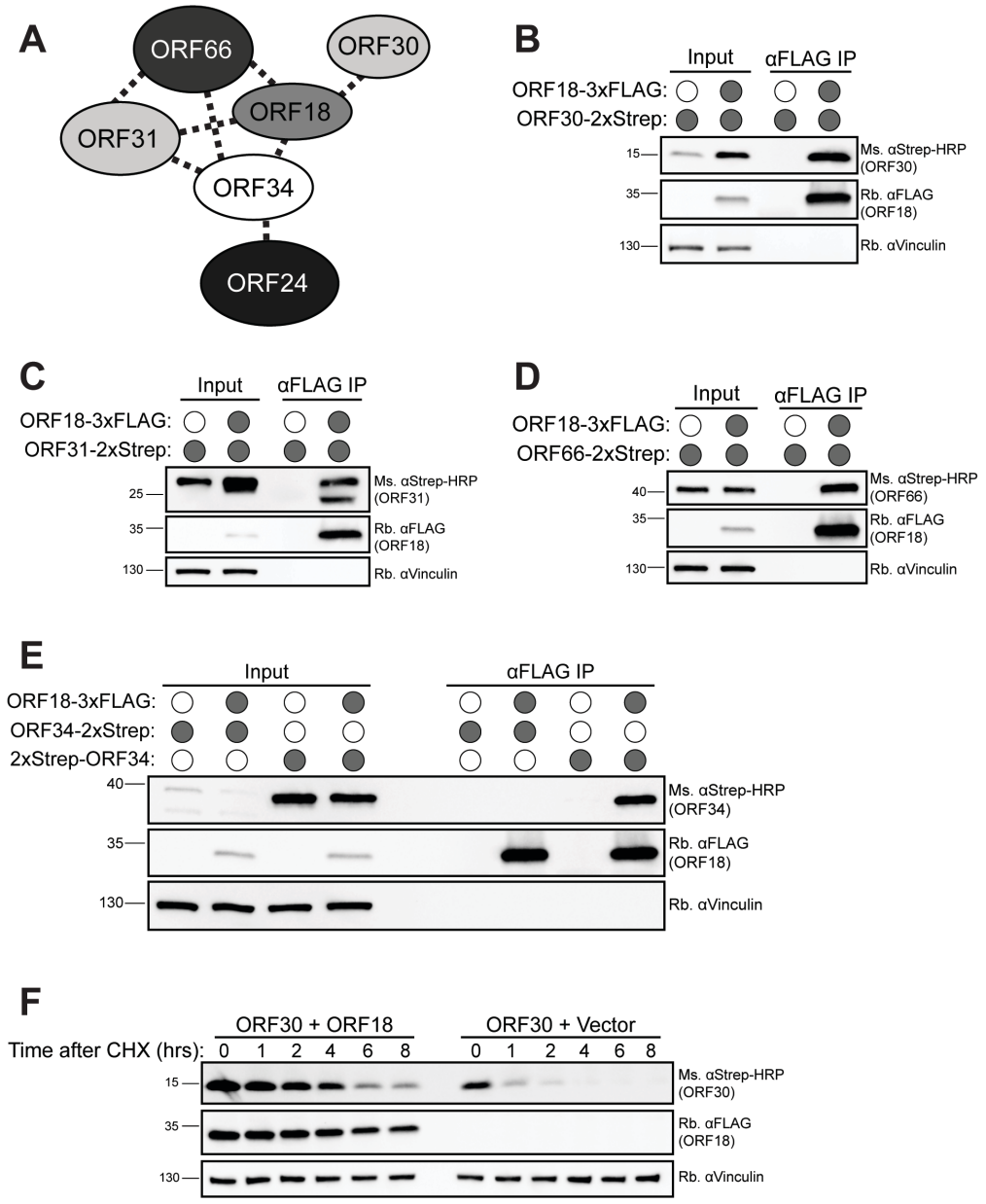


Figure 2.1. ORF18 interacts with ORFs 30, 31, 34, and 66. (A) Diagram of vTA interactions in KSHV from (Davis et al., 2016). **(B-E)** HEK293T cells were transfected with the indicated vTA plasmids, then subjected to co-IP using α–FLAG beads followed by western blot analysis with the indicated antibody to detect ORF18 and either ORF30 (*B*), ORF31 (*C*), ORF66 (*D*), and ORF34 (*E*). Input represents 2.5% of the lysate used for co-IP. Vinculin served as a loading control. **(F)** HEK293T cells were transfected with the indicated vTA plasmids. 24 h post transfection, cycloheximide was added to a final concentration of 100 μg/mL, and samples were collected at the indicated time points after the addition of cycloheximide. 25 μg of whole cell lysate was resolved using SDS-PAGE followed by western blot with the indicated antibodies. Vinculin served as a loading control.

An interaction screen of ORF18 mutants reveals the role of conserved residues in interactions with the other vTAs

To evaluate the importance of the interaction between ORF18 and its individual vTA contacts, we aimed to identify point mutations that disrupted binding to individual vTAs but did not destroy the integrity of the complex. Since the late gene vTA complex is conserved across the beta- and gammaherpesviruses we reasoned that the individual points of contact might depend on conserved amino acid residues. We performed a multiple sequence alignment between KSHV ORF18 and its homologs in five other beta- and gamma herpesviruses (MHV68 ORF18, HCMV pUL79, MCMV pM79, EBV BVLF1, and BHV ORF18) using MUSCLE (Edgar, 2004). The sequence alignment revealed 25 single conserved residues, including six pairs of adjacent conserved residues, which are depicted in **Figure 2.2A** as a schematic of the primary structure of ORF18 showing the positions of conserved residues. We mutated each of the 25 conserved residues to alanine in ORF18-3xFLAG and made double alanine mutations in the six cases of adjacent conserved residues (**Figure 2.2A**). Each of these 31 mutants was screened individually for the ability to interact with ORFs 30, 31, and 66 by co-IP followed by western blot (data not shown). To account for differences in expression between the ORF18 mutants, we calculated the co-IP efficiency of each of the mutants, as described in the methods. These data were used to generate a heat map, which displays the pair-wise interaction efficiencies on a color scale where lighter blocks represent reduced binding and darker blocks represents increased binding relative to WT (**Figure 2.2B**). Overall, the data revealed that ORF30 was the most sensitive to mutations in ORF18, as 24 out of the 31 mutants displayed reduced or no binding. ORF31 showed variable sensitivity to ORF18 mutation (with some mutants even increasing the interaction efficiency), while the ORF66-ORF18 interaction was relatively refractory to the ORF18 point mutations (**Figure 2.2B**).

We focused on the six ORF18 mutants that exhibited <10% co-IP efficiency, relative to WT, with any vTA in our initial screen (L29A, E36A, L151A, W170A, E36A_L37A, and W170A_G171A; highlighted in red in **Figure 2.2A**). These were re-screened 3-4 independent times in co-IP assays with vTA components ORFs 30, 31, 34, and 66, and the co-IP efficiencies were calculated as described in the methods then plotted relative to values obtained for WT ORF18 (**Figure 2.3A-D**). All six of these ORF18 mutants had severe defects in their ability to co-IP ORF30, but none were consistently different than WT ORF18 for interaction with ORFs 31, 34, and 66 (**Figure 2.3B-D**). Among the six mutants, ORF18^{E36A_L37A} and ORF18^{W170A_G171A} showed no detectable binding to ORF30, with ORF18^{E36A_L37A} retaining near-WT levels of interaction with the remaining vTAs.

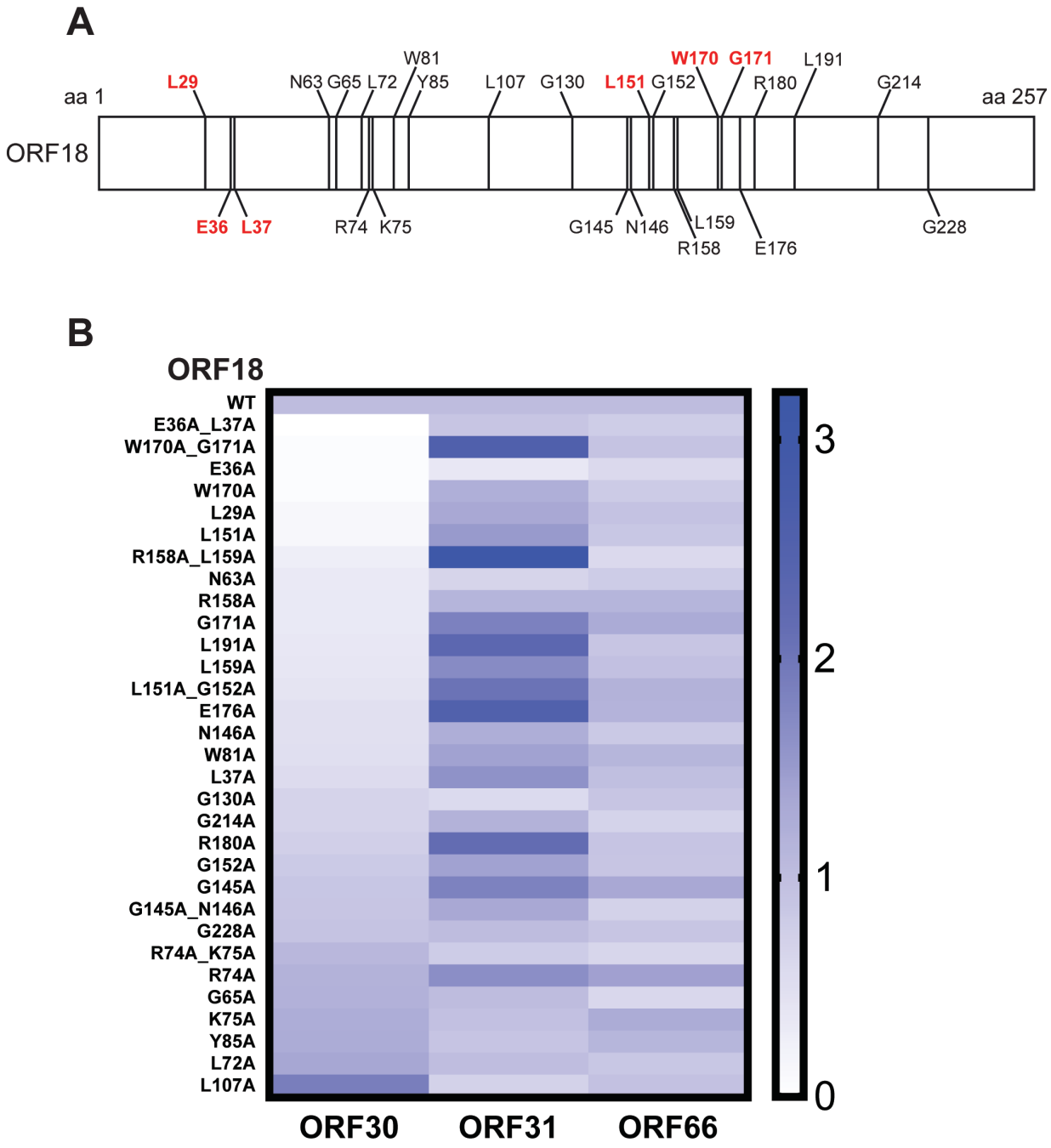


Figure 2.2. ORF18 mutant screen for interactions with ORFs 30, 31, and 66. (A) Diagram depicting the conserved residues in KSHV ORF18 in a MUSCLE alignment with MHV68 ORF18, EBV BVLF1, HCMV pUL79, MCMV pUL79, and BHV4 ORF18. Red color denotes amino acids found to have interactions <10% of WT with any of the vTAs. **(B)** Heat map of the co-IP efficiency of each ORF18 mutant against ORFs 30, 31, and 66.

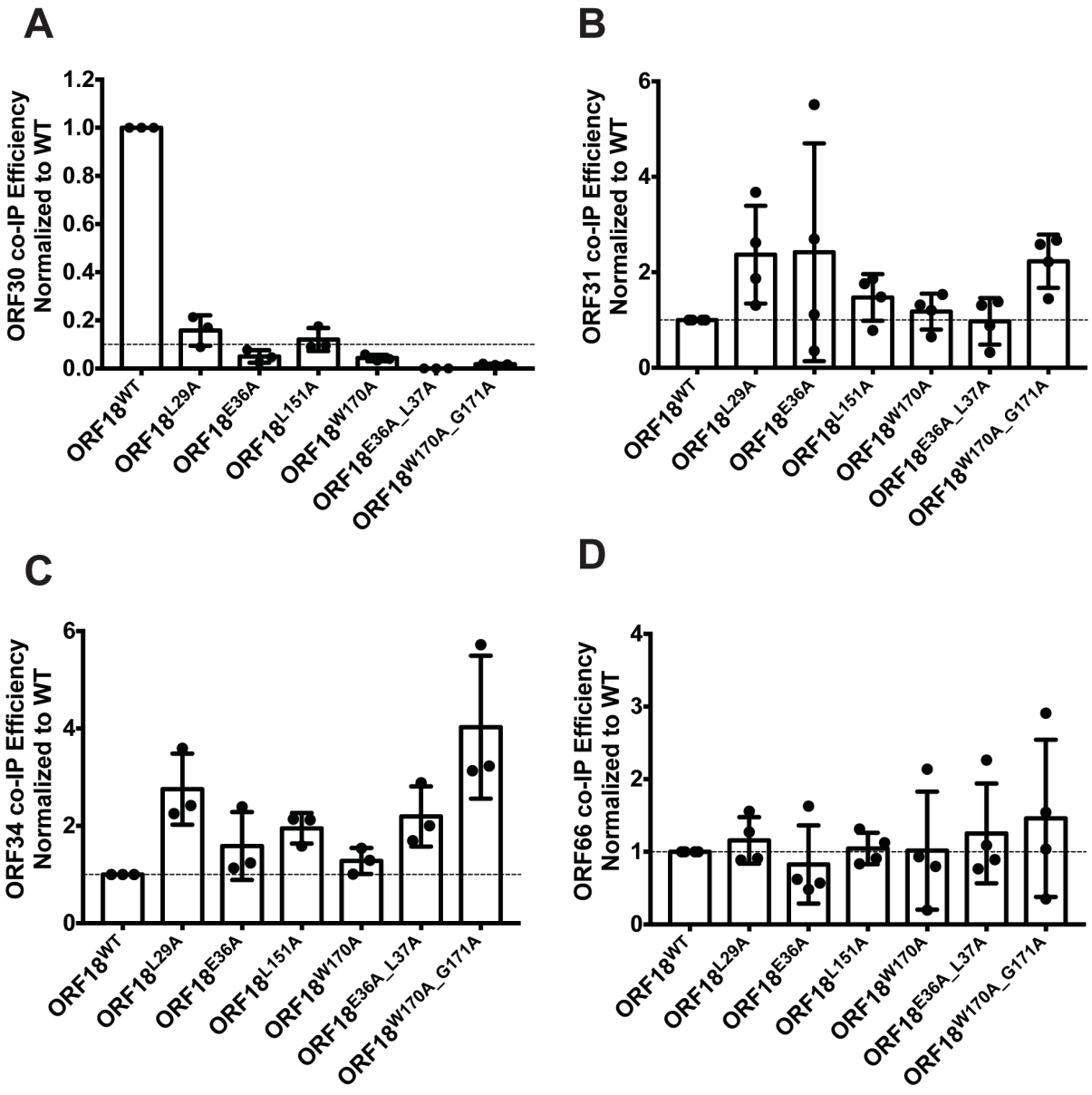


Figure 2.3. Six ORF18 mutants are consistently defective for interaction with ORF30. (A-D) HEK293T cells were transfected with the indicated vTA plasmids, then subjected to co-IP using α-FLAG beads followed by western blot analysis to detect the ability of WT or mutant ORF18 to interact with ORF30 (*A*), ORF31 (*B*), ORF34 (*C*), and ORF66 (*D*). The co-IP efficiency was calculated as described in the text for 3-4 independent experimental replicates and plotted as bar graphs. In *(A)*, the dotted line represents Y = 0.1, and in *(B-D)*, the dotted line represents Y = 1.0.

ORF18 point mutants that weaken the interaction between ORF18 and ORF30 have a reduced capacity to activate the K8.1 late gene promoter

A reporter assay has been developed in which the co-expression of the six individual vTAs can specifically activate a KSHV (or EBV) late gene promoter in transfected HEK293T cells (Aubry et al., 2014; Davis et al., 2016). We used this assay as an initial proxy for the ability of the six ORF18 mutants described above to activate late gene transcription. HEK293T cells were co-transfected with each of the vTAs, including either WT or mutant ORF18, and firefly luciferase reporter plasmids driven by either the late K8.1 promoter or, as a control, the early ORF57 promoter (**Figure 2.4A**). A plasmid containing constitutively expressed Renilla luciferase was co-transfected with each sample to normalize for transfection efficiency. As expected, inclusion of WT ORF18 with the remaining vTA complex resulted in specific activation of the K8.1 late promoter, but not of the early ORF57 promoter (**Figure 2.4B**). ORF18 mutants L29A, E36A, and L151A modestly reduced activation of the late promoter, whereas more significant defects were observed with mutants W170A, E36A_L37A, and W170A_G171A (**Figure 2.4B**). Although ORF18^{W170A_G171A} had the most pronounced transcriptional defect, this mutant showed somewhat more variability than ORF18^{E36A_L37A} in its interactions with the other vTA components (**Figures 2.3 & 2.4**). Thus, we considered ORF18^{E36A_L37A} to be the top candidate for selectively analyzing the importance of the ORF18-ORF30 interaction for KSHV late gene transcription.

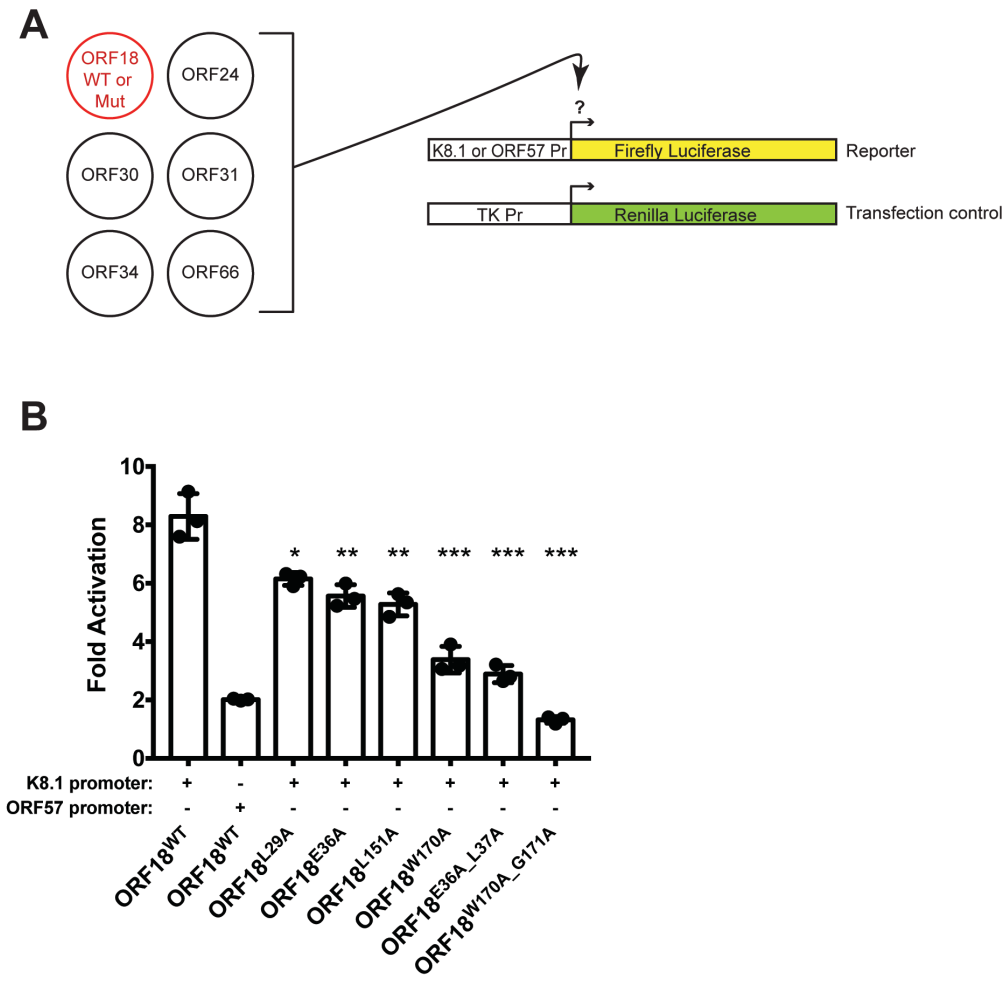


Figure 2.4. Impairing the interaction between ORF18 and ORF30 reduces activation of the late K8.1 promoter. (A) Diagram depicting the vector combinations that were transfected for the late gene reporter assay. **(B)** HEK293T cells were transfected with the vTA plasmids including WT or mutant ORF18, the K8.1 or ORF57 promoter reporter plasmids, and the pRL-TK Renilla plasmid as a transfection control. 24 h post-transfection the lysates were harvested and luciferase activity was measured. Data shown are from 3 independent biological replicates, with statistics calculated using an unpaired t-test where (*) $p < 0.05$, (**) $p < 0.005$, and (***) $p < 0.0007$.

The interaction between ORF18 and ORF30 affects the assembly of the vTA complex

The transcriptional defect of the ORF18^{E36A_L37A} mutant in the reporter assay could be due to a defect in assembly of the complex or due to defects in downstream events. To distinguish between these possibilities, WT or mutant ORF18-3xFLAG was co-transfected into HEK293T cells with each of the other Strep-tagged vTAs. We then performed an α-FLAG IP, revealing that purification of WT ORF18 led to co-IP of the complete vTA complex including Pol II, which has been shown to interact with ORF24 in KSHV (Davis et al., 2015) (**Figure 2.5**). We noted that in this assay ORF18^{E36A_L37A} was more weakly expressed than WT ORF18, so to compare complex formation with equivalent amounts of each protein, we titrated down the amount of WT ORF18 to match the levels of ORF18^{E36A_L37A}. Similar to our observation in a pairwise co-IP (**Figure 2.1B**), the ORF30 protein abundance decreased as the expression of ORF18 was reduced (**Figure 2.5**); however, the complete vTA complex still co-purified even with reduced levels of WT ORF18 (**Figure 2.5**). Notably, the vTA complex was recovered at lower levels in the presence of ORF18^{E36A_L37A}. When imaged at a longer exposure, all of the vTAs, with the exception of ORF30, remained associated with ORF18^{E36A_L37A} (**Figure 2.5**, far right panel). Thus, the selective loss of the ORF18-ORF30 interaction may reduce the overall stability of the vTA complex.

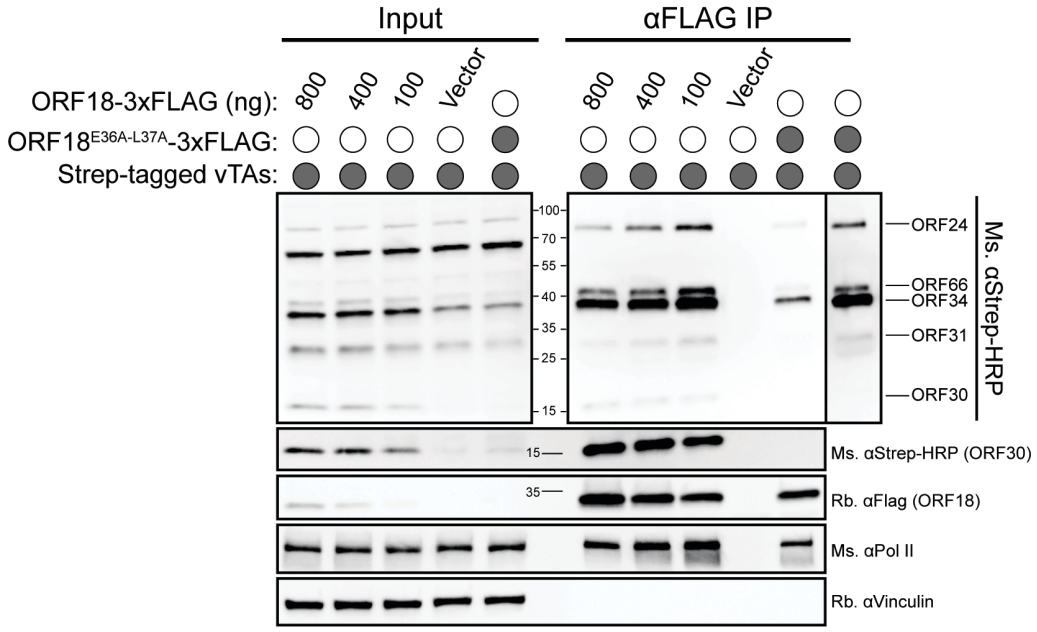


Figure 2.5. Disrupting the interaction between ORF18 and ORF30 weakens the assembly of the vTA complex. HEK293T cells were transfected with the indicated vTA plasmids then subjected to co-IP using α-FLAG beads followed by western blot analysis with the indicated antibody. Far right boxed lane is a longer exposure of the ORF18^{E36A_L37A} IP lane. Input represents 2.5% of the lysate used for co-IP. Vinculin was used as a loading control.

The interaction between ORF18 and ORF30 is crucial for expression of the model late gene K8.1

Next, to characterize the effect of ORF18^{E36A_L37A} on the viral replication cycle, we tested the ability of this mutant to complement the late gene expression defect of a KSHV mutant lacking ORF18 (18.stop) (Gong et al., 2014). The renal carcinoma cell line iSLK harbors the virus (either WT or 18.stop) in a latent state, which can be reactivated upon expression of the doxycycline-inducible major lytic transactivator RTA and treatment with sodium butyrate. Using lentiviral transduction, we generated stable, doxycycline-inducible versions of the 18.stop iSLK cells expressing either ORF18-3xFLAG or ORF18^{E36A_L37A}-3xFLAG (18.stop.ORF18^{WT} and 18.stop.ORF18^{E36A_L37A}, respectively). The cells were assayed 72 hours post lytic reactivation for their ability to replicate DNA, express early and late proteins, and produce progeny virions. Although we observed a modest decrease of viral DNA replication in the 18.stop cells, as measured by qPCR, upon complementation with either WT ORF18 or ORF18^{E36A_L37A}, the levels of DNA replication were not significantly different from iSLK cells infected with WT KSHV (**Figure 2.6A**). Notably, the 18.stop.ORF18^{E36A_L37A} cell line expressed more ORF18 than the 18.stop.ORF18^{WT} cell line, in contrast to the reduced expression of the mutant in HEK293T cells **(Figure 2.6B**, compare to levels in **Figure 2.5**). However, while both reactivated 18.stop.ORF18^{WT} and 18.stop.ORF18^{E36A_L37A} cell lines expressed equivalent levels of the ORF59 early protein, only 18.stop.ORF18^{WT} was able to rescue expression of the model late gene K8.1 (**Figure 2.6B**).

We then evaluated the level of KSHV virion production from the parental WT KSHV-infected iSLK cells, as well as from the 18.stop, 18.stop.ORF18^{WT}, and 18.stop.ORF18^{E36A_L37A} cell lines using a supernatant transfer assay. KSHV produced from iSLK cells contains a constitutively expressed GFP, enabling quantitation of infected recipient cells by flow cytometry (Brulois et al., 2012). Consistent with its late gene expression defect, neither the 18.stop nor the 18.stop.ORF18^{E36A_L37A} cell lines were able to produce progeny virions, whereas virion production in the 18.stop.ORF18^{WT} cells was indistinguishable from the WT KSHV-infected iSLK cells (**Figure 2.6C**). Collectively, these data demonstrate that the specific interaction between ORF18 and ORF30 is essential for K8.1 late gene expression and virion production during KSHV infection.

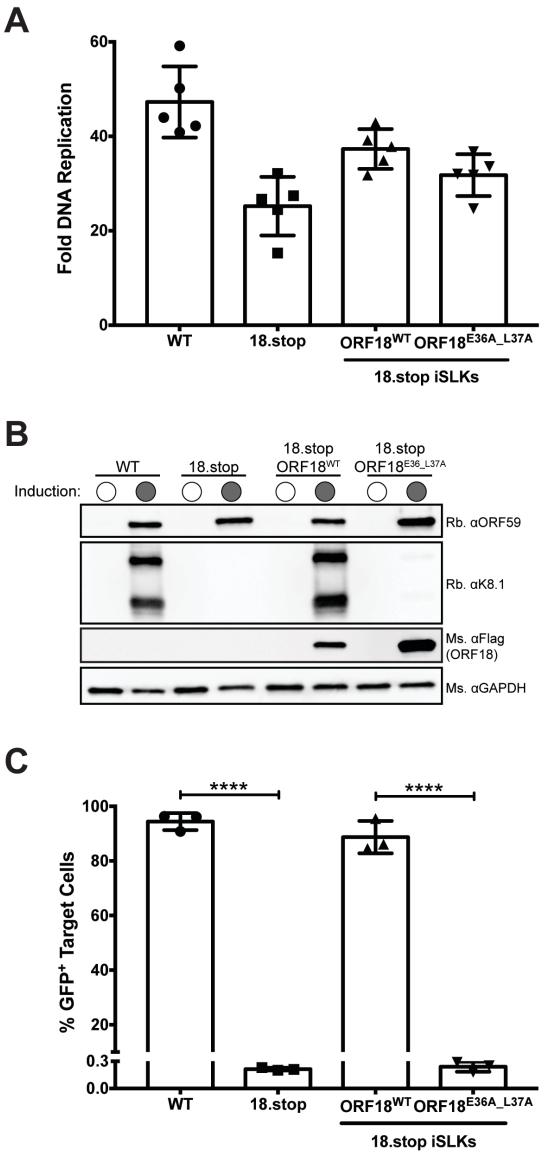


Figure 2.6. Characterizing the effect of the E36A_L37A ORF18 mutation on the virus. (A) iSLK cells latently infected with WT KSHV, 18.Stop KSHV, or 18.Stop complemented with ORF18^{WT} or ORF18^{E36A_L37A} were induced to enter the lytic cycle with 1 μg/mL doxycycline and 1 mM sodium butyrate. 72 h post induction, DNA was isolated and fold viral DNA replication was measured by qPCR before and after induction of the lytic cycle. Data shown are from 5 independent biological replicates. **(B)** Western blots of the expression of the early protein ORF59, the late protein K8.1, ORF18^{WT}, and ORF18^{E36A_L37A} in the indicated cell lines induced as described in (*A*). GAPDH was used as a loading control. **(C)** HEK293T target cells were spinfected with filtered supernatant from induced cells. Progeny virion production was measured 24 h after supernatant transfer by flow cytometry of GFP+ target cells. Data shown are from 3 independent biological replicates with statistics calculated using an unpaired t-test where, (****) $p < 0.0001$.

The analogous mutation in HCMV pUL79 disrupts its interaction with pUL91

As shown in **Figure 2.7A**, the E36_L37 residues are conserved across the beta- and gammaherpesvirus ORF18 homologs. To determine whether these amino acids are similarly important in a betaherpesvirus, we engineered the corresponding double mutation in HCMV pUL79 (pUL79^{E48A_L49A}-3xFLAG). Similar to our observation with KSHV ORF30, HCMV pUL91 protein expression was significantly decreased in the absence of its WT pUL79 binding partner (**Figure 2.7B**). This is consistent with the idea that pUL79 binding stabilizes pUL91. Furthermore, in co-IP assays we detected a robust interaction between pUL79 and pUL91, which was impaired in the presence of pUL79^{E48A_L49A}, even when we accounted for the expression level differences of pUL91 (**Figure 2.7B**). This suggests that the overall protein-protein interface may be conserved in the analogous ORF18-ORF30 interaction across the beta- and gammaherpesviruses.

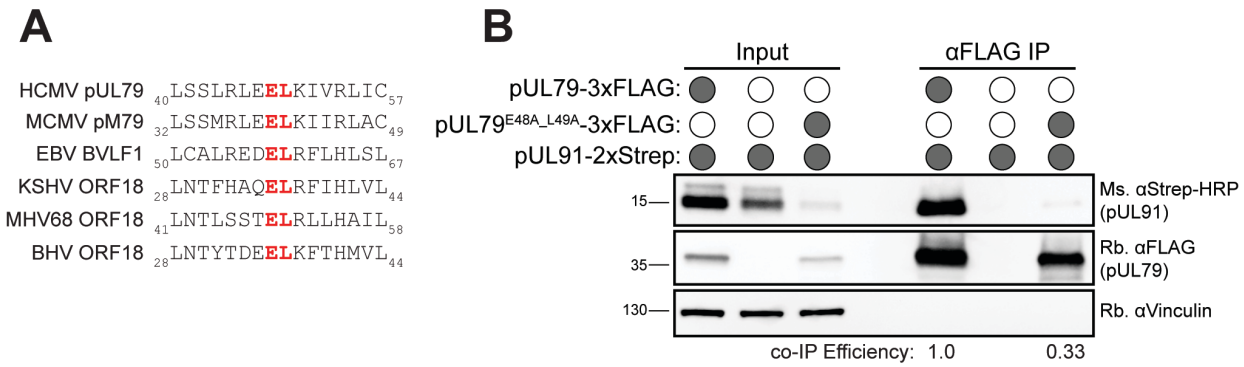


Figure 2.7. Mutating E48_L49 in pUL79 disrupts its interaction with pUL91. (A) MUSCLE multiple sequence alignment for HCMV pUL79 and homologs showing the location of conserved amino acids that correspond to E36_L37 in KSHV ORF18. (B) HEK293T cells were transfected with the indicated plasmids, then subjected to co-IP using α-FLAG beads followed by western blot analysis with the indicated antibody. To normalize for the difference in expression of pUL91 in the different transfection conditions, the co-IP efficiency was multiplied by the fold expression of pUL91 in the presence of WT pUL79 versus E48A_L49A pUL79. Input represents 2.5% of lysate used for co-IP. Vinculin was used as a loading control.

Discussion

Elucidating the architecture of the six-member vPIC complex is central to understanding the mechanism underlying viral late gene expression in beta- and gammaherpesviruses. Although their functions are largely unknown, each of these viral transcription regulators is essential for late gene promoter activation and evidence increasingly suggests that their ability to form a complex is crucial for transcriptional activity (Aubry et al., 2014; Davis et al., 2016; Pan et al., 2018). Here, we reveal that selective disruption of an individual protein-protein contact between KSHV ORF18 and ORF30 within the vPIC is sufficient to abrogate K8.1 late gene expression and virion production in infected cells, emphasizing the sensitivity of the complex to organizational perturbation.

We selected ORF18 for mutational screening due to its ability to form pairwise interactions with the majority of other vPIC components, which suggested that it might serve as an organizational hub for vPIC assembly, similar to what has been proposed for ORF34 (Davis et al., 2016; Nishimura et al., 2017). However, the 31 tested mutants of ORF18 revealed that conserved residues across the length of the protein are extensively—and largely selectively— required for ORF30 binding. This is in contrast to the interaction between the vPIC components ORF24 and ORF34, where the interaction can be localized to a 17 amino acid stretch of ORF24 (Davis et al., 2016). The observation that the majority of the ORF18 point mutations that disrupt the interaction with ORF30 do not affect its binding to ORFs 31, 34, and 66 indicates that these mutations do not significantly alter the overall folding or structure of ORF18. In MCMV, the organization of the vTA complex is similar to KSHV, except that pM92 (homologous to KSHV ORF31) interacts with pM87 (homologous to KSHV ORF24), whereas in KSHV this interaction is bridged through ORF34 (Pan et al., 2018). Our data complement recent findings in MCMV, in which mutations of the ORF30 homolog (pM91) that perturb the interaction with the ORF18 homolog (pM79) similarly cause a defect in the expression of late genes (Pan et al., 2018). Thus, mutations that disrupt the ORF18-ORF30 protein-protein interface in either protein cause the same phenotype.

The ORF18-ORF30 interaction appears sensitive to single amino acid changes in the protein-protein interface. The ORF18^{E36A_L37A} mutation does not impair binding to its other vPIC partners in pairwise co-IP experiments and ORF30 does not engage in pairwise interactions with vPIC components other than ORF18 (Davis et al., 2016; Nishimura et al., 2017). It is therefore notable that the efficiency of the vPIC complex assembly is reduced in HEK293T cells in the presence of the ORF18^{E36A_L37A} mutant, suggesting that the ORF18-ORF30 interaction contributes to the stability of the complex as a whole. The interaction between ORF18 and ORF30 may change the conformation of ORF18, allowing it to interact more strongly with the other vTAs. Alternatively, the E36A_L37A mutation may contribute to binding defects between ORF18 and the other vTAs that are not observed with pairwise interactions, but are enhanced in the presence of all the ORF18 binding partners.

We observed that ORF30 protein expression was consistently higher in the presence of WT ORF18 or ORF18 mutants that retained ORF30 binding, suggesting that ORF18 helps stabilize ORF30. *In silico* protein stability prediction studies have suggested that protein stability is in part affected by protein length, where proteins that are less than 100 amino acids tend to be less stable (Dill, 1985; White, 1992). One explanation for the

higher expression of ORF30 in the presence of ORF18 could therefore be that the 77 amino acid ORF30 is protected from degradation by ORF18. Another possibility is that ORF18 helps keep ORF30 correctly folded—this has been proposed as a mechanism that stabilizes proteins which have interaction partners (Dixit and Maslov, 2013). The interaction-induced stability of a protein often correlates with the relative concentration of its binding partners (Dixit and Maslov, 2013), as we observed when we titrated down the amount of ORF18 in the context of the complete vTA complex. A similar observation has been made between KSHV proteins ORF36 and ORF45, where ORF36 was dependent on the interaction with ORF45 for stabilization (Avey et al., 2016). We did not observe a similar correlation with levels of the other ORF18-associated vPIC proteins, and thus their stability may not require protective interactions. We also observed a stabilizing effect of pUL79 on pUL91, indicating this may be a conserved feature of this interaction in other beta and gammaherpesviruses.

The fact that the late gene expression defect of the ORF18^{E36A_L37A} mutant is exacerbated in the context of the virus, compared to in the plasmid promoter activation assay, likely reflects the fact that the plasmid assay measures basal promoter activation but misses other regulatory components of this cascade. For example, the origin of lytic replication is required *in cis* for late gene expression in related gammaherpesviruses (Djavadian et al., 2016; Tang et al., 2004) and the reporter assay does not capture the important contribution of viral DNA replication towards late gene expression. This may explain why some mutants that are defective for ORF30 binding (e.g. L29A, E36A, L151A) retain partial plasmid promoter activity; perhaps some weak binding between ORF18 and ORF30 enables basal activation of the promoter in a context where the vPIC components are overexpressed. Alternatively, some of the mutations may cause ORF18 to bind to ORF30 more transiently, but their vPIC interaction becomes stabilized in the presence of a late gene promoter.

In summary, the absence of K8.1 late gene expression in the KSHV ORF18.stop-infected cells complemented with ORF18^{E36A_L37A} may derive from a cascade of phenotypes: the failure to recruit ORF30 to the vPIC, the ensuing reduction in the efficiency of overall vPIC complex assembly, and the reduced stability of ORF30 (if it also has additional vPIC-independent functions). Ultimately, generating antibodies that recognize the endogenous KSHV vPIC components will enable these phenotypes to be explored further during infection. In addition, information about the 3-dimensional structure of the vPIC would significantly enhance our understanding of this unique transcription complex, as it is becoming increasingly clear that even small disruptions to the complex dramatically impact completion of the viral lifecycle.

Materials and Methods

Plasmids and Plasmid Construction

To generate ORF18-3xFLAG pCDNA4, ORF18 was subcloned into the BamHI and NotI sites of pCDNA4-3xFLAG. The point mutations in ORF18 were generated using two primer site-directed mutagenesis with Kapa HiFi polymerase (Roche) with primers 1-62 listed in Table 1. All subsequent plasmids described below were generated using InFusion cloning (Clontech) unless indicated otherwise. To generate plasmid pLVX-TetOneZeo, zeocin resistance was PCR amplified out of plasmid pLJM1-EGFP-Zeo with

primers 63/64 (Table 1) and used to replace the puromycin resistance in pLVX-TetOne^{TM}-Puro (Clontech) using the AvrII and MluI restriction sites. To generate pLVX-TetOneZeo-ORF18^{WT}-3xFLAG and pLVX-TetOneZeo-ORF18^{E36A_L37A}-3xFLAG, ORF18^{WT}-3xFLAG and ORF18^{E36_37A}-3xFLAG were PCR amplified from each respective pCDNA4 plasmid using primers 71/72 and inserted into the EcoRI and BamHI sites of pLVX-TetOne-Zeo. To generate UL79-3xFLAG pCDNA4 and UL91-2xStrep pCDNA4, UL79 was PCR amplified with primers 67/68 (Table 1) and UL91 was PCR amplified with primers 69/70 (Table 1) from HCMV Towne strain, which was kindly provided by Dr. Laurent Coscoy, and cloned into the BamHI and NotI sites of 3xFLAG (Cterm) pCDNA4 or 2xStrep (Cterm) pCDNA4. UL79^{E48A_L49A} was generated with two primer site-directed mutagenesis using Kapa HiFi polymerase with primers 73/74 (Table 1). To make 2xStrep-ORF34 pCDNA4, ORF34 was PCR amplified from ORF34-2xStrep pCDNA4 with primers 65/66 and cloned into the NotI and XbaI sites of 2xStrep (Nterm) pCDNA4. Plasmid K8.1 Pr pGL4.16 contains the minimal K8.1 promoter and ORF57 Pr pGL4.16 contains a minimal ORF57 early gene promoter and have been described previously (Davis et al., 2016). Plasmids ORF18-2xStrep pCDNA4, ORF24-2xStrep pCDNA4, ORF30-2xStrep pCDNA4, ORF31-2xStrep pCDNA4, ORF34-2xStrep pCDNA4, and ORF66-2xStrep pCDNA4 have been previously described (Davis et al., 2015). Plasmid pRL-TK (Promega) was kindly provided by Dr. Russel Vance. Lentiviral packaging plasmids psPAX2 (Addgene plasmid # 12260) and pMD2.G (Addgene plasmid # 12259) were gifts from Dr. Didier Trono.

Cells and Transfections

HEK293T cells (ATCC CRL-3216) were maintained in DMEM supplemented with 10% FBS (Seradygm). The iSLK renal carcinoma cell line harboring the KSHV genome on the bacterial artificial chromosome BAC16 and a doxycycline-inducible copy of the KSHV lytic transactivator RTA (iSLK-BAC16) has been described previously (Brulois et al., 2012). iSLK-BAC16-ORF18.stop cells that contain a stop mutation in ORF18 were kindly provided by Dr. Ting-Ting Wu (Gong et al., 2014). iSLK-BAC16 and iSLK-BAC16-ORF18.stop were maintained in DMEM supplemented with 10% FBS, 1 mg/mL hygromycin, and 1 μg/mL puromycin (iSLK-BAC16 media). iSLK-BAC16-ORF18.stop cells were complemented by lentiviral transduction with ORF18^{WT}-3xFLAG or ORF18^{E36A_L37A}-3xFLAG. To generate the lentivirus, HEK293T cells were co-transfected with pLVX-TetOneZeo-ORF18^{WT}-3xFLAG or pLVX-TetOneZeo-ORF18^{E36A_L37A}-3xFLAG along with the packaging plasmids pMD2.G and psPAX2. After 48h, the supernatant was harvested and syringe-filtered through a 0.45 μm filter (Millipore). The supernatant was diluted 1:2 with DMEM and polybrene was added to a final concentration of 8 μg/mL. 1 x 10^6 iSLK-BAC16-ORF18.stop freshly trypsinized cells were spinfected in a 6-well plate for 2 hours at 500 x *g*. After 24 hours the cells were expanded to a 10 cm tissue culture plate and selected for 2 weeks in iSLK-BAC16 media supplemented with 325 μg/mL zeocin (Sigma). For DNA transfections, HEK293T cells were plated and transfected after 24 hours at 70% confluency with PolyJet (SignaGen).

Immunoprecipitation and Western Blotting

Cell lysates were prepared 24 hours after transfection by washing and pelleting cells in cold PBS, then resuspending the pellets in IP lysis buffer [50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1mM EDTA, 0.5% NP-40, and protease inhibitor (Roche)] and rotating for

30 minutes at 4°C. Lysates were cleared by centrifugation at 21,000 x *g* for 10 min, then 1 mg (for pairwise interaction IPs) or 2 mg (for the entire late gene complex IPs) of lysate was incubated with pre-washed M2 α-FLAG magnetic beads (Sigma) overnight. The beads were washed 3x for 5 min each with IP wash buffer [50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1mM EDTA, 0.05% NP-40] and eluted with 2x Laemmli sample buffer (BioRad).

Lysates and elutions were resolved by SDS-PAGE and western blotted in TBST (Tris-buffered saline, 0.2% Tween 20) using the following primary antibodies: Strep-HRP (Millipore, 1:2500); rabbit α-FLAG (Sigma, 1:3000); mouse α-FLAG (Sigma, 1:1000); rabbit α-Vinculin (Abcam, 1:1000); mouse α-GAPDH (Abcam, 1:1000); mouse α-Pol II CTD clone 8WG16 (Abcam, 1:1000); rabbit α-K8.1 (1:10000); rabbit α-ORF59 (1:10000). Rabbit α-ORF59 and α-K8.1 was produced by the Pocono Rabbit Farm and Laboratory by immunizing rabbits against MBP-ORF59 or MBP-K8.1 [gifts from Denise Whitby (Labo et al., 2014)]. Following incubation with primary antibodies, the membranes were washed with TBST and incubated with the appropriate secondary antibody. The secondary antibodies used were the following: goat α-mouse-HRP (Southern Biotech, 1:5000) and goat α-rabbit-HRP (Southern Biotech, 1:5000).

The co-IP efficiency for the pairwise interactions was quantified from the western blot images using Image Lab software (BioRad). The band intensity for the both the Strep-tagged ORF and ORF18^{WT}-3xFLAG or ORF18^{Mu}-3xFLAG was calculated for the IP lanes of the western blot. The ratio of the band intensity of Strep-tagged ORF to ORF18^{Mu}-3xFLAG was divided by the ratio of Strep-tagged ORF to ORF18^{WT}-3xFLAG to generate a co-IP efficiency for each ORF18^{Mu} relative to the co-IP efficiency of ORF18^{WT}.

ORF30 Protein Stability

Translation was inhibited 24 hours after transfection by the addition of 100 μg/mL cycloheximide for 0-8 hours. Cells were washed once in cold PBS, and cell pellets were frozen until all samples were collected. The pellets were lysed in IP lysis buffer by rotating for 30 minutes at 4°C. Lysates were cleared by centrifugation at 21,000 x *g* for 10 min, then resolved by SDS-PAGE followed by western blot.

Virus Characterization

For reactivation studies, 1 x 10^6 iSLK cells were plated in 10 cm dishes for 16 hours, then induced with 1 μg/mL doxycycline and 1 mM sodium butyrate for an additional 72 hours. To determine the fold DNA induction in reactivated cells, the cells were scraped and triturated in the induced media, and 200 μl of the cell/supernatant suspension was treated overnight with 80 μg/mL proteinase K (Promega) in 1x proteinase K digestion buffer (10 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EDTA, 0.5% SDS) after which DNA was extracted using a Quick-DNA Miniprep kit (Zymo). Viral DNA fold induction was quantified by qPCR using iTaq Universal SYBR Green Supermix (BioRad) on a QuantStudio3 Real-Time PCR machine with primers 75/76 (Table 1) for the KSHV ORF59 promoter and normalized to the level of GAPDH promoter with primers 77/78 (Table 1).

Infectious virion production was determined by supernatant transfer assay. Supernatant from induced iSLK cells was syringe-filtered through a 0.45 μm filter and diluted 1:2 with DMEM, then 2 mL of the supernatant was spinfected onto 1 x 10^6 freshly

trypsinized HEK293T cells for 2 hours at 500 x *g*. After 24 hours, the media was aspirated, the cells were washed once with cold PBS and crosslinked in 4% PFA (Ted Pella) diluted in PBS. The cells were pelleted, resuspended in PBS, and a minimum of 50,000 cells/sample were analyzed on a BD Accuri 6 flow cytometer. The data were analyzed using FlowJo (Gardner and Glaunsinger, 2018).

Late Gene Reporter Assay

HEK293T cells were plated in 6-well plate and each well was transfected with 900 ng of DNA containing 125 ng each of pcDNA4 ORF18-3xFLAG or ORF18^{Mu}-3xFLAG, ORF24-2xStrep, ORF30-2xStrep, ORF31-2xStrep, 2xStrep-ORF34, ORF66-2xStrep, and either K8.1 Pr pGL4.16 or ORF57 Pr pGL4.16, along with 25 ng of pRL-TK as an internal transfection control. After 24 hours, the cells were rinsed once with PBS, lysed by rocking for 15 min at room temperature in 500 μl of Passive Lysis Buffer (Promega), and clarified by centrifuging at 21,000 x *g* for 2 min. 20 μl of the clarified lysate was added in triplicate to a white chimney well microplate (Greiner bio-one) to measure luminescence on a Tecan M1000 microplate reader using a Dual Luciferase Assay Kit (Promega). The firefly luminescence was normalized to the internal Renilla luciferase control for each transfection.

Acknowledgements

We thank all members of the Glaunsinger and Coscoy labs, in particular Matthew R. Gardner, for their helpful suggestions and critical reading of the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1752814 and UC Berkeley Chancellor's Fellowship awarded to A.C. B.G. is an investigator of the Howard Hughes Medical Institute. This research was also supported by NIH R01AI122528 to B.G.

Table 2.1. List of synthetic DNA oligonucleotides used in this study.

Primer Number	Purpose	Sequence 5' - 3'	Orientation
1	ORF18 (L29A)	CATGTGGCGCTTTTTGCAAAATAAGAATGCAAATACATTCCACGCCCAAG	F
2	ORF18 (L29A)	CTTGGGCGTGGAATGTATTTGCATTCTTATTTTGCAAAAAGCGCCACATG	R
3	ORF18 (E36A)	AATACATTCCACGCCCAAGCGCTGCGTTTTATTCATTTG	F
4	ORF18 (E36A)	CAAATGAATAAAACGCAGCGCTTGGGCGTGGAATGTATT	R
5	ORF18 (L37A)	GAACCAAATGAATAAAACGCGCCTCTTGGGCGTGGAATGTAT	F
6	ORF18 (L37A)	ATACATTCCACGCCCAAGAGGCGCGTTTTATTCATTTGGTTC	R
7	ORF18 (N63A)	GGGAGGCTACTGCCGCTGCCGGGACCTACG	F
8	ORF18 (N63A)	CGTAGGTCCCGGCAGCGGCAGTAGCCTCCC	R
9	ORF18 (G65A)	CTCGTCGTAGGTCGCGGCATTGGCAGT	F
10	ORF18 (G65A)	ACTGCCAATGCCGCGACCTACGACGAG	R
11	ORF18 (L72A)	GAACCTTGCGTCCCGCGACCACCTCGTCGT	F
12	ORF18 (L72A)	ACGACGAGGTGGTCGCGGGACGCAAGGTTC	R
13	ORF18 (R74A)	CGCAGGAACCTTGGCTCCCAGGACCACC	F
14	ORF18 (R74A)	GGTGGTCCTGGGAGCCAAGGTTCCTGCG	R
15	ORF18 (K75A)	GGTCCTGGGACGCGCGGTTCCTGCGGAG	F
16	ORF18 (K75A)	CTCCGCAGGAACCGCGCGTCCCAGGACC	R
17	ORF18 (W81A)	TCGTACACGAGCTTCGCCACCTCCGCAGGAAC	F
18	ORF18 (W81A)	GTTCCTGCGGAGGTGGCGAAGCTCGTGTACGA	R
19	ORF18 (Y85A)	TCCTCGAGCCCATCGGCCACGAGCTTCCACAC	F
20	ORF18 (Y85A)	GTGTGGAAGCTCGTGGCCGATGGGCTCGAGGA	R
21	ORF18 (L107A)	GTTCAAGTGCATCCAGGCGCTGTCCCGGTATGCC	F
22	ORF18 (L107A)	GGCATACCGGGACAGCGCCTGGATGCACTTGAAC	R
23	ORF18 (G130A)	GTCGTGGGTGACCGCTAGCCGGTGAAA	F
24	ORF18 (G130A)	TTTCACCGGCTAGCGGTCACCCACGAC	R
25	ORF18 (G145A)	CAGATTAAACAAAAAGTTTGCGTCCACCAGGTTTTCCG	F
26	ORF18 (G145A)	CGGAAAACCTGGTGGACGCAAACTTTTTGTTTAATCTG	R
27	ORF18 (N146A)	TTCCCAGATTAAACAAAAAGGCTCCGTCCACCAGGTTTTCCG	F
28	ORF18 (N146A)	CGGAAAACCTGGTGGACGGAGCCTTTTTGTTTAATCTGGGAA	R
29	ORF18 (L151A)	GAGCACACTTCCCGCATTAAACAAAAAGTTTCCGTCCACC	F
30	ORF18 (L151A)	GGTGGACGGAAACTTTTTGTTTAATGCGGGAAGTGTGCTC	R
31	ORF18 (G152A)	TGCAGGGGAGCACACTTGCCAGATTAAACAAAAAG	F
32	ORF18 (G152A)	CTTTTTGTTTAATCTGGCAAGTGTGCTCCCCTGCA	R
33	ORF18 (R158A)	CGCAAGGAGCAGCGCGCAGGGGAGCACA	F
34	ORF18 (R158A)	TGTGCTCCCCTGCGCGCTGCTCCTTGCG	R
35	ORF18 (L159A)	CGCCGCAAGGAGCGCCCTGCAGGGGAGC	F
36	ORF18 (L159A)	GCTCCCCTGCAGGGCGCTCCTTGCGGCG	R
37	ORF18 (W170A)	ATCGCTGCCCGCAAAGGCGAGGCAGTAGCCC	F
38	ORF18 (W170A)	GGGCTACTGCCTCGCCTTTGCGGGCAGCGAT	R
39	ORF18 (G171A)	CGTGTTCATCGCTGGCCCAAAAGGCGAGG	F
40	ORF18 (G171A)	CCTCGCCTTTTGGGCCAGCGATGAACACG	R
41	ORF18 (E176A)	GCGCACCCAGCGTGCGTGTTCATCGCT	F
42	ORF18 (E176A)	AGCGATGAACACGCACGCTGGGTGCGC	R
43	ORF18 (R180A)	CTGGGCGAAGAAGGCCACCCAGCGTTCG	F
44	ORF18 (R180A)	CGAACGCTGGGTGGCCTTCTTCGCCCAG	R
45	ORF18 (L191A)	AAGACGCCCGGAGACTATCGCGTAGCAAATGAAAAGCTTC	F
46	ORF18 (L191A)	GAAGCTTTTCATTTGCTACGCGATAGTCTCCGGGCGTCTT	R
47	ORF18 (G214A)	CCTCCACCGGAGCGGGATAGCCC	F
48	ORF18 (G214A)	GGGCTATCCCGCTCCGGTGGAGG	R
49	ORF18 (G228A)	GCATACGTTCGTATGGCGTACATGGAGCGGA	F
50	ORF18 (G228A)	TCCGCTCCATGTACGCCATACGAACGTATGC	R
51	ORF18 (E36A_L37A)	CAGAGAACCAAATGAATAAAACGCGCCGCTTGGGCGTGGAATGTATTTAAAT	F

52	ORF18 (E36A_L37A)	ATTTAAATACATTCCACGCCCAAGCGGCGCGTTTTATTCATTTGGTTCTCTG	R
53	ORF18 (R74A_K75A)	CCTCCGCAGGAACCGCGGCTCCCAGGACCACCTCGTC	F
54	ORF18 (R74A_K75A)	GACGAGGTGGTCCTGGGAGCCGCGGTTCCTGCGGAGG	R
55	ORF18 (G145A_N146A)	CTTCCCAGATTAAACAAAAAGGCTGCGTCCACCAGGTTTTCCGGGG	F
56	ORF18 (G145A_N146A)	CCCCGGAAAACCTGGTGGACGCAGCCTTTTTGTTTAATCTGGGAAG	R
57	ORF18 (L151A_G152A)	AGGGGAGCACACTTGCCGCATTAAACAAAAAGTTTCCGTCCACCA	F
58	ORF18 (L151A_G152A)	TGGTGGACGGAAACTTTTTGTTTAATGCGGCAAGTGTGCTCCCCT	R
59	ORF18 (R158A_L159A)	CCGCCGCAAGGAGCGCCGCGCAGGGGAGCACAC	F
60	ORF18 (R158A_L159A)	GTGTGCTCCCCTGCGCGGCGCTCCTTGCGGCGG	R
61	ORF18 (W170A_G171A)	GTGTTCATCGCTGGCCGCAAAGGCGAGGCAGTAGCC	F
62	ORF18 (W170A_G171A)	GGCTACTGCCTCGCCTTTGCGGCCAGCGATGAACAC	R
63	pLVX-TetOneZeo	TTTTTGGAGGCCTAGGCTTTTGCAAAACGCGACCATGGCCAAGTTGACCAGTGC	F
64	pLVX-TetOneZeo	ATTGTTCCAGACGCGTTCAGTCCTGCTCCTCGGC	R
65	2xStrep-ORF34 pCDNA4	GAGAAGGGGGCGGCCTTTGCTTTGAGCTCGCTCGTGT	F
66	2xStrep-ORF34 pCDNA4	AAACGGGCCCTCTAGTTAGAGTTGGTTGAGTCCATTCTCCTTGATC	R
67	UL79-3xFLAG pCDNA4	TACCGAGCTCGGATCATGATGGCCCGCGACG	F
68	UL79-3xFLAG pCDNA4	CTCCCTCGAGCGGCCCCACGTCGTTAGCCAGCGT	R
69	UL91-2xStrep pCDNA4	TACCGAGCTCGGATCATGAACTCGTTGCTGGCGG	F
70	UL91-2xStrep pCDNA4	CTCCCTCGAGCGGCCCTGTCACAGGCGCCCGAG	R
71	$ORF18^{WT}$ and $ORF18^{E36\_L37A}$ pLVX TetOneZeo	CCCTCGTAAAGAATTATGCTCGGAAAATACGTGTGTGAGACC	F
72	$ORF18^{WT}$ and $ORF18^{E36\_L37A}$ pLVX TetOneZeo	GAGGTGGTCTGGATCTTAAACGGGCCCCTTGTCGTCG	R
73	$UL79^{E48_L49A}$_3xFLAG pCDNA4	ATGAGGCGTACGATCTTGGCTGCTTCCAAACGCAGCGAGC	F
74	$UL79^{E48_L49A}$_3xFLAG pCDNA4	GCTCGCTGCGTTTGGAAGCAGCCAAGATCGTACGCCTCAT	R
75	ORF59 promoter qPCR	AATCCACAGGCATGATTGC	F
76	ORF59 promoter qPCR	CACACTTCCACCTCCCCTAA	R
77	GAPDH promoter qPCR	TACTAGCGGTTTTACGGGCG	F
78	GAPDH promoter qPCR	TCGAACAGGAGGAGCAGAGACCGA	R

Chapter 3: ORF24 interacts directly with the carboxy terminal domain of RNA polymerase II

The work presented in this chapter was performed in collaboration with Robert K. Louder and Eva Nogales.

Introduction

DNA viruses generally commandeer the host RNA polymerase II (Pol II) transcriptional machinery to direct their gene expression. Given that the mechanisms governing transcription from viral promoters are often similar to those at host promoters, viruses have been valuable models for understanding transcription complex assembly and regulation (Bauer et al., 2018; Chi and Carey, 1996; Haigh et al., 1990). Eukaryotic and viral transcription begins with the formation of a pre-initiation complex (PIC) at the core promoter. The general assembly pathway for the PIC starts with binding of the TATA box by TATA-binding protein (TBP) together with TBP-associated factors (TAFs), which form TFIID, followed by binding of other general transcription factors (GTFs). The GTFs recruit and position the 12 subunit Pol II at the core promoter (Thomas and Chiang, 2006). The largest Pol II subunit, Rpb1, has a carboxy terminal domain (CTD) that in humans is composed of 52 heptapeptide repeats, whose differential phosphorylation impacts subsequent stages of transcription (Harlen et al., 2016). Together, these factors are sufficient for basal transcription, although genes with *cis*-regulatory elements require additional activators and cofactors for robust promoter activation (Malik and Roeder, 2010).

A conserved feature of double-stranded DNA (dsDNA) viruses is that their genes are expressed in an ordered temporal cascade that begins with the expression of two classes of early genes, followed by viral DNA replication, and ends with the expression of late genes. In herpesviruses, the mechanisms underlying the regulation of late gene expression are poorly understood; however, a common theme is the requirement for an early gene product, which acts as a transactivator or helps stabilize the PIC (Gruffat et al., 2016). In beta and gammaherpesviruses, late gene transcription is regulated in part by the core promoter sequence which is 12-15 base pairs (bp) in length and has a TATT-box instead of the canonical TATA box found in cellular and early viral promoters (Serio et al., 1998; Tang et al., 2004; Wong-Ho et al., 2014). Additionally, late gene expression requires at least six viral proteins, called viral transcriptional activators (vTAs), which form a complex at late gene promoters (Aubry et al., 2014; Castañeda and Glaunsinger, 2019; Davis et al., 2016; Pan et al., 2018). Overall, little is known about the mechanism by which the vTAs activate late gene expression.

The best studied protein in the vTA complex is an apparent TATA-binding protein (TBP) mimic, termed ORF24 in Kaposi's sarcoma-associated herpesvirus (KSHV). ORF24 is predicted to have a TBP-like domain in the central portion of the protein (**Figure 3.1A**), which was identified through an *in silico* protein fold threading analysis performed with its Epstein-Barr virus (EBV) homolog BcRF1 (Wyrwicz and Rychlewski, 2007). Indeed, KSHV ORF24 replaces TBP at late gene promoters during infection, and a virus with mutations to the predicted DNA-binding residues in the TBP-like domain of ORF24 is unable to transcribe late genes (Davis et al., 2015). While both ORF24 and TBP bind DNA, a potentially unique feature of this viral TBP mimic is that it also interacts with Pol

II (Davis et al., 2015); whereas, the interaction between TBP and Pol II is bridged through TFIIB (Chen and Hahn, 2003; Tsai and Sigler, 2000).

At present it is unknown whether the ORF24-Pol II interaction is direct or bridged by other proteins as is the case for TBP. To answer this question, we identify a fragment of ORF24 that interacts with Pol II and can be expressed in and purified from *Escherichia coli* (*E. coli*). With this protein fragment we use single particle negative stain electron microscopy (EM) and pulldown assays to determine that ORF24 interacts directly with the CTD of Pol II. We define the minimum number of CTD repeats involved in the interaction, as well as identify a single leucine residue in ORF24 that is critical for Pol II-binding. These findings suggest that ORF24 is different from other Pol II-interacting proteins in that it not only interacts with Pol II directly, it also binds core promoter DNA.

Results

ORF24^{NTD} interacts with Pol II from HEK293T cells

We sought to determine whether KSHV ORF24 directly interacts with Pol II using purified proteins. However, full-length versions of ORF24 proved refractory to recombinant expression in multiple systems (data not shown). We previously showed that the N-terminus of ORF24 contained three essential conserved leucine residues (L73-75) critical for the interaction with Pol II (Davis et al., 2015). We therefore constructed a series of five N-terminal ORF24 domain fragments to map a more minimal region of the protein sufficient to retain the Pol II interaction, which might be more amenable to purification. Three of the fragments included L73-75 required for Pol II binding, while two did not and were used to determine whether any other regions of the N-terminus contributed to Poll II binding (**Figure 3.1B**). Of the five fragments, three were well expressed in HEK293T cells and retained Pol II binding in co-immunoprecipitation assays; all of these included L73-75 (**Figure 3.1C**). The best interacting fragment of ORF24 with Pol II was the first 201 most N-terminal amino acids, which we termed ORF24 N-terminal domain (ORF24^{NTD}) (**Figure 3.1B, C**).

We appended an N-terminal glutathione-S-transferase (GST) tag to ORF24^{NTD}, and achieved robust expression of the protein in *E. coli*, similar to that of GST alone (**Figure 3.1D**). Furthermore, we confirmed that GST-tagged ORF24^{NTD} retained the ability to interact with Pol II using a GST pulldown with purified GST or GST-ORF24^{NTD} and whole cell lysate from HEK293T cells (**Figure 3.1E**). These results indicated that the NTD of ORF24 was properly folded and could be used in biochemical assays to characterize the ORF24-Pol II interaction.

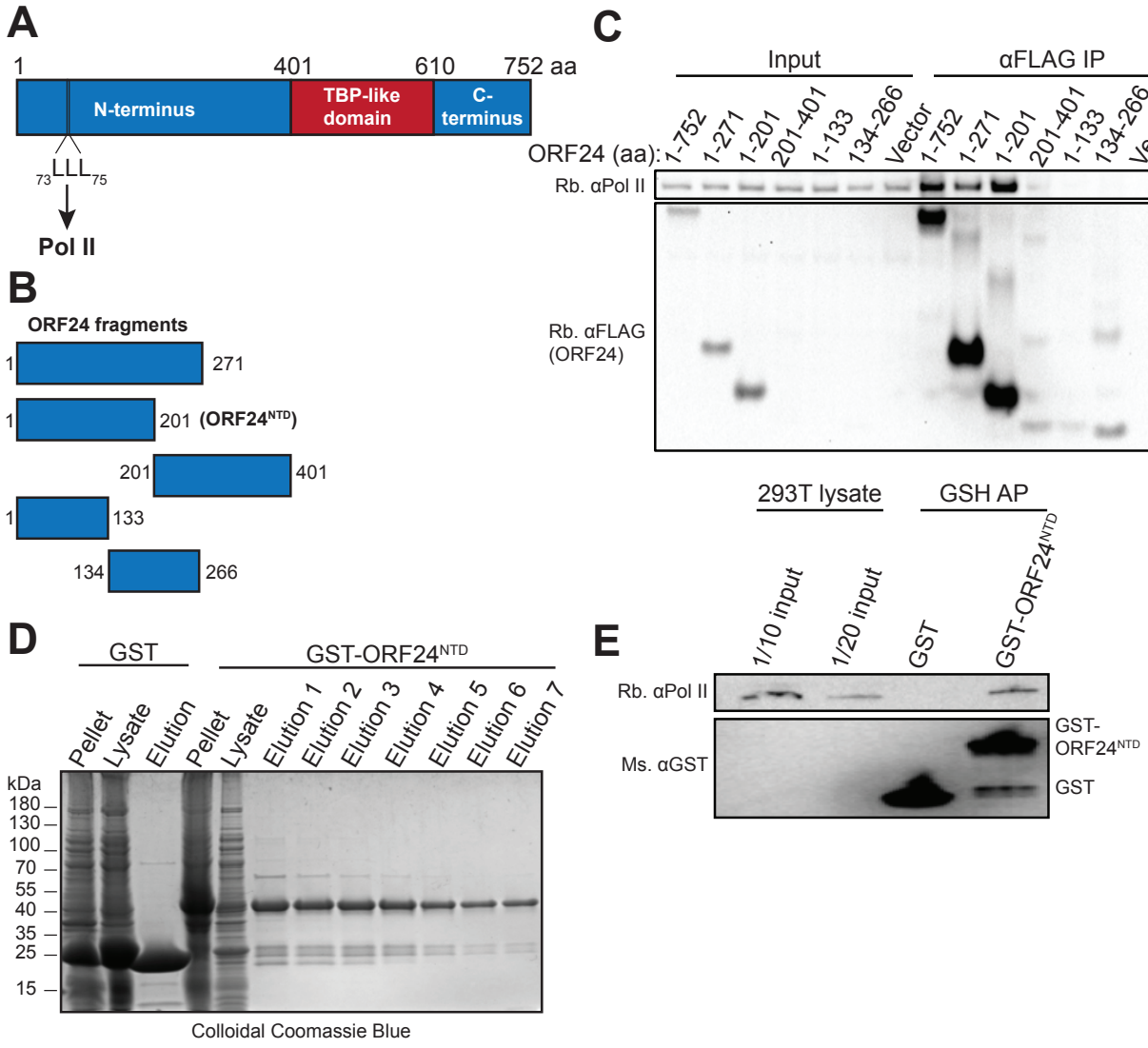


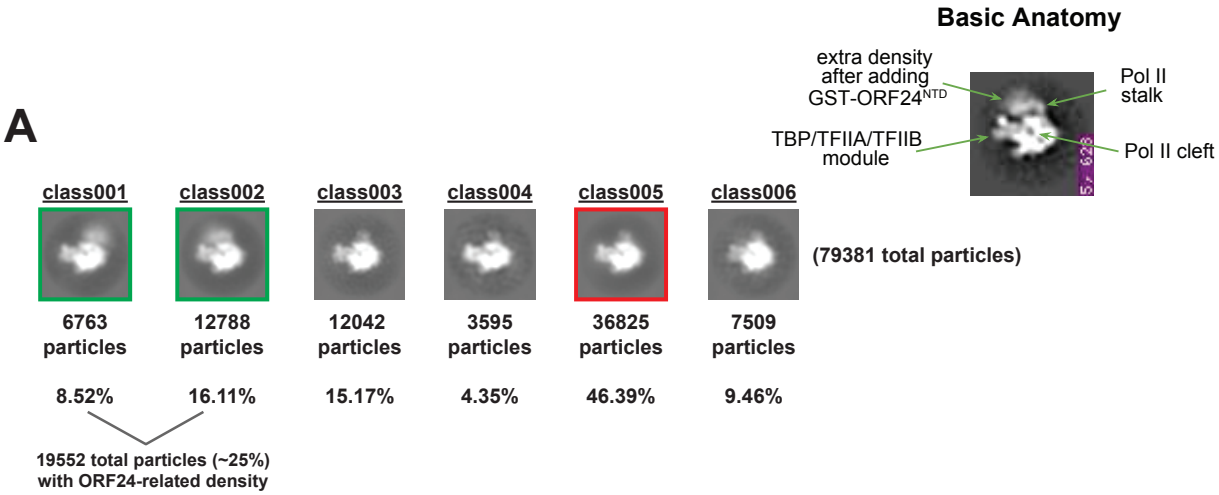
Figure 3.1. GST-ORF24^{NTD} interacts with Pol II from mammalian cell lysate. (A) Diagram of ORF24 showing the predicted boundaries for the N-terminal domain, the TBP-like domain, and the C-terminal domain, including the location of the three leucine residues required for Pol II binding. **(B)** Diagram of ORF24 fragments tested for expression and ability to co-immunoprecipitate (co-IP) with Pol II. Numbers indicate the amino acid residues in each fragment. **(C)** HEK293T cells were transfected with the indicated ORF24 plasmids, then subjected to co-IP using α–FLAG beads followed by western blot with the indicated antibodies to detect ORF24 and Pol II. **(D)** Colloidal Coomassie blue gel of samples collected from the purification of GST-ORF24^{NTD} and GST. **(E)** Western blot of glutathione pulldown performed with purified GST or GST-ORF24^{NTD} and whole cell lysate from HEK293T cells.

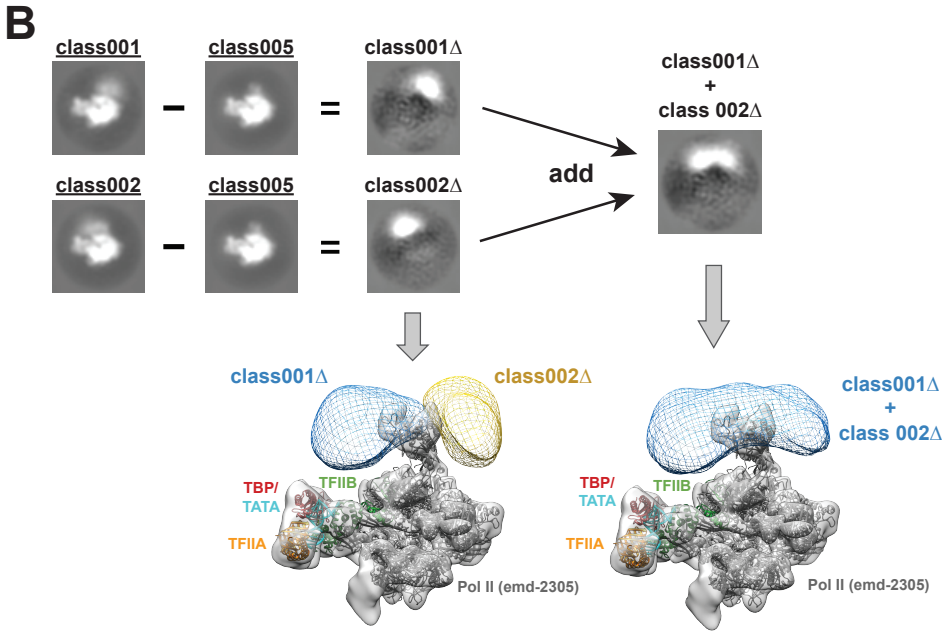
Negative stain electron microscopy of PICs with GST-ORF24^{NTD} suggests interaction with the Pol II stalk

To gain structural insight into how ORF24^{NTD} interacts with Pol II, we performed single particle negative stain EM of GST-ORF24^{NTD} bound to a minimal PIC, composed of Pol II, TBP, and the GTFs TFIIA and TFIIB. A total of 79,381 particles were collected and approximately 25% of them had extra density related to GST-ORF24^{NTD} (**Figure 3.2A**). The defining characteristics of a minimal PIC are the Pol II stalk, composed of Rpb4/7, the Pol II cleft, and the TBP/TFIIA/TFIIB module (**Figure 3.2A**, far right). The particles with extra density were sorted into two classes for further analysis and comparison. Based on a projection of the minimal PIC with the Pol II cleft facing outward, class001 had extra density to the right of the Pol II stalk and class002 had extra density to the left of the Pol II stalk (**Figure 3.2A**). The majority of particles without extra density are represented by class005, which was chosen as the reference class against which to conduct further analysis and comparison.

To amplify the signal to noise ratio of the density related to GST-ORF24^{NTD}, the signal from the class005 average was subtracted from both the class001 and class002 averages (**Figure 3.2B**). The three-dimensional difference maps representing the extra density related to GST-ORF24^{NTD} were mapped on the cryo-EM structure of the human PIC (PDBe ID code emd-2305) (He et al., 2013). The two orientations of GST-ORF24^{NTD} were combined to generate a combined difference map, which illustrates the entire density encompassed by both class001 and class002 averages (**Figure 3.2B**, bottom right). On the combined difference map, GST-ORF24^{NTD} is represented by blue mesh, and the two different orientations for the ORF24^{NTD}-related density suggested that ORF24^{NTD} could be binding to a flexible region of Pol II.

An inherently flexible domain of Pol II is the CTD of the largest subunit, Rpb1, which contains a flexible linker followed by 52 heptapeptide repeats, the consensus of which is YSPTSPS (Harlen and Churchman, 2017). Since the CTD does not adopt a homogenous conformation, it is not visualized by EM. However, an x-ray crystallography structure of Pol II from *Schizosaccharomyces pombe* (*S. pombe*), visualized a portion of the linker to the CTD and shows that it extends near Rpb4/7 (PDB ID code 3H0G) (Spahr et al., 2009). To determine the proximity of the CTD linker in relation to the GST-ORF24^{NTD}-related density, the *S. pombe* CTD linker was docked on the PIC with the extra density corresponding to GST-ORF24^{NTD} (**Figure 3.2C**). In the full PIC, GST-ORF24^{NTD} rests above the stalk of Pol II (**Figure 3.2C**) and the zoomed in visualization shows that the CTD linker localized to the GST-ORF24^{NTD}-related density (**Figure 3.2D**).


Basic Anatomy
extra density after adding GST-ORF24NTD
Pol II stalk
TBP/TFIIA/TFIIB module
Pol II cleft
A
class001
class002
class003
class004
class005
class006
(79381 total particles)
6763 particles
12788 particles
12042 particles
3595 particles
36825 particles
7509 particles
8.52%
16.11%
15.17%
4.35%
46.39%
9.46%
19552 total particles (~25%) with ORF24-related density


B
class001
class005
class001Δ
class002
class005
class002Δ
add
class001Δ + class 002Δ
class001Δ
class002Δ
TBP/ TATA
TFIIB
TFIIA
Pol II (emd-2305)
class001Δ + class 002Δ
TBP/ TATA
TFIIB
TFIIA
Pol II (emd-2305)

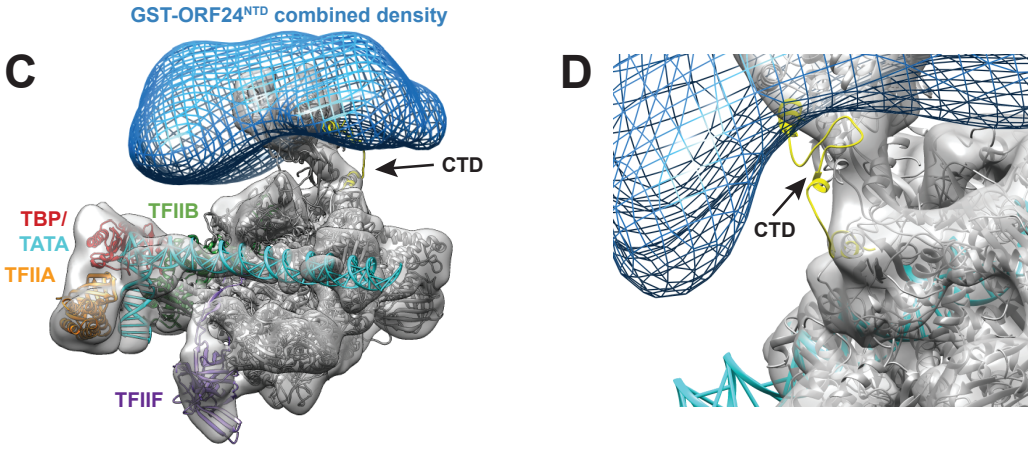

GST-ORF24NTD combined density
C
CTD
TBP/ TATA
TFIIB
TFIIA
TFIIF
D
CTD

Figure 3.2. GST-ORF24^{NTD} binds Pol II in minimal PICs. (A) Representative micrographs of the five different class averages annotated with the number of particles in each class and the percent of total particles in each class. Micrograph annotated with the basic anatomy of a particle with extra density related to GST-ORF24^{NTD} (far right). **(B)** Schematic showing the classes that were used to generate the two-dimensional difference maps for GST-ORF24^{NTD} (top). Three-dimensional difference maps for the extra density attributed to GST-ORF24^{NTD} modeled on the human PIC (PDBe ID code emd-2305). **(C)** Structure of the human PIC depicting a short linker portion of the CTD from *Schizosaccharomyces pombe* (PDB ID code 3H0G) docked on Pol II. **(D)** Zoom in of the interface between GST-ORF24^{NTD} and the CTD linker of Pol II showing the proximity of the two proteins.

ORF24^{NTD} binds the CTD repeats of Rpb1

Based on the EM results, we hypothesized that ORF24 could be interacting with the CTD of Pol II. We therefore tested for a direct interaction between each purified protein by GST pulldown assays followed by western blot, using GST-tagged versions of the Rpb1 CTD linker or repeat regions, and maltose-binding protein (MBP) tagged ORF24^{NTD}. As a control, we also purified an MBP-tagged mutant version of ORF24^{NTD}, in which the three conserved leucines at positions L73-75 were mutated to alanines (ORF24$^{NTD}\Delta$LLL). This mutation renders the protein unable to interact with Pol II in cells (Davis et al., 2015). As shown in **Figure 3.3A**, WT ORF24^{NTD} bound Rpb1 through the CTD repeats but not the linker region. This interaction was specific, as no binding to either CTD fusion was observed with ORF24$^{NTD}\Delta$LLL.

We next sought to identify a length of CTD repeats that could be used in fluorescence polarization to determine the affinity of the CTD-ORF24^{NTD} interaction. We also wanted to obtain structural information about ORF24^{NTD} to understand how it mediates the interaction with the CTD. However, both the GST- and MBP-ORF24^{NTD} fusion proteins were too large for structural determination by NMR. Therefore, we began to optimize the purification of ORF24^{NTD} to generate untagged protein. We found that when cleaving the MBP tag and attempting to separate ORF24^{NTD} from MBP using size exclusion chromatography, some amount of MBP co-sized with ORF24^{NTD} (data not shown). Next, we attempted to optimize the purification of ORF24^{NTD} by adding a C-terminal Strep tag II. We predicted that adding a Strep purification step following cleavage of the MBP tag would entirely remove any remaining MBP tag, but this was not the case (**Figure 3.3B**, ORF24^{NTD}-Strep input). Nevertheless, a pulldown with ORF24^{NTD}-Strep suggested that a solubility tag on ORF24^{NTD} was not necessary for its interaction with the CTD repeats (**Figure 3.3B**). While we further optimized the purification of ORF24^{NTD}, we tested whether there was a specific CTD repeat length required for the interaction. A Strep pulldown with ORF24^{NTD}-Strep and GST-tagged 1x, 2x, 5x, 10x, or 52x CTD repeats showed that ORF24^{NTD} interacts with 5x and 10x CTD repeats, but does not interact with 2x and 1x repeats (**Figure 3.3C**).

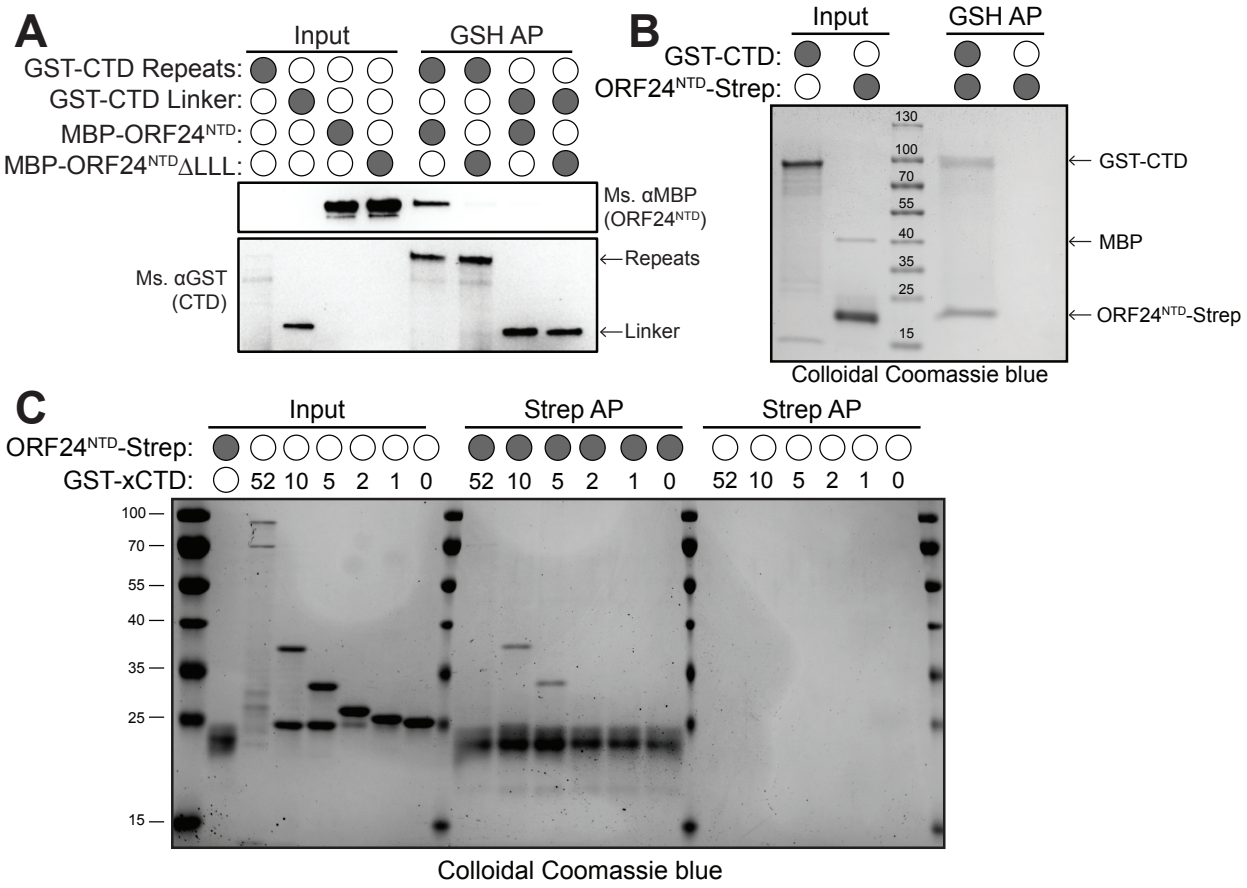


Figure 3.3. ORF24^{NTD} interacts directly with the CTD repeats of Pol II. (A) Western blot of glutathione pulldown performed with the following purified proteins: GST-CTD repeats, GST-CTD linker, MBP-ORF24^{NTD}, MBP-ORF24^{NTD}ΔLLL. The inputs and elutions were resolved on SDS-PAGE and analyzed by western blot with the indicated antibodies. **(B)** Colloidal Coomassie blue gel of glutathione pulldown performed with recombinantly purified GST-CTD repeats and ORF24^{NTD}-Strep. **(C)** Colloidal Coomassie blue gel of Strep pulldown performed with recombinantly purified GST-xCTD repeats and ORF24^{NTD}-Strep.

A single leucine residue in ORF24 is essential for interaction with Pol II

To further characterize the interaction between ORF24 and Pol II, we sought to determine if all three of the leucine residues contributed equally to the interaction. Each of the leucine residues was mutated to alanine individually and in pairs to generate all possible combinations of leucine to alanine mutations. The effect of the mutations was tested by co-immunoprecipitation (co-IP) followed by western blot. Each of the individual leucine mutants disrupted the ORF24-Pol II interaction to varying degrees. L73A had the least effect, followed by L74A, and L75A almost completely disrupted the interaction to the same extent as the three leucine to alanine mutation (**Figure 3.4A**). This was in contrast to the double mutants where in all cases, the interaction between ORF24 and Pol II was completely abolished (**Figure 3.4A**). These results indicated that the leucine residues have an additive effect on the Pol II interaction and do not contribute to the interaction equally.

Given that the structure of the N-terminal domain of ORF24 has not been solved, it is difficult to predict the role of the leucine residues in the ORF24^{NTD}-Pol II interaction. One possibility is that the leucine residues could be directly involved in mediating the interaction between the two proteins. Another possibility is that the leucine residues are required for proper folding of ORF24^{NTD} and a mutation in this amino acid stretch disrupts the secondary structure involved in CTD binding. Without the structure of ORF24^{NTD}, one way to discern whether the mutation of the leucine residues affects the fold is to purify any of the ORF24^{NTD} leucine mutants in the same way that WT ORF24^{NTD} is purified. If the structure of the protein remains unperturbed, we would expect to obtain a similar amount of protein harboring the leucine mutations. However, if the leucine mutations disrupt the fold of ORF24^{NTD}, we would expect to purify a smaller amount of protein.

The current purification scheme for ORF24^{NTD} involves a Strep purification following the removal of the SUMO3 tag with SenP2 protease (**Figure 3.5A**). With WT ORF24^{NTD}, the Strep purification yields an elution peak with absorbance at 280 nm of approximately 850 mAu (**Figure 3.4B**). However, for both the three-leucine mutant (3L_A) and the L75A mutant of ORF24^{NTD}, the absorbance at A280 for the elution peak is approximately 25 mAu (**Figure 3.4B**). Based on the extinction coefficient of ORF24^{NTD}-Strep, the difference in absorbance corresponds to an 18-fold reduction in yield from approximately 9 mg for WT ORF24^{NTD} to 0.5 mg for each of the mutants. Given that the expression and purification conditions are the same, this difference in yield suggests that the leucine to alanine mutations disrupt the structure of ORF24^{NTD}.

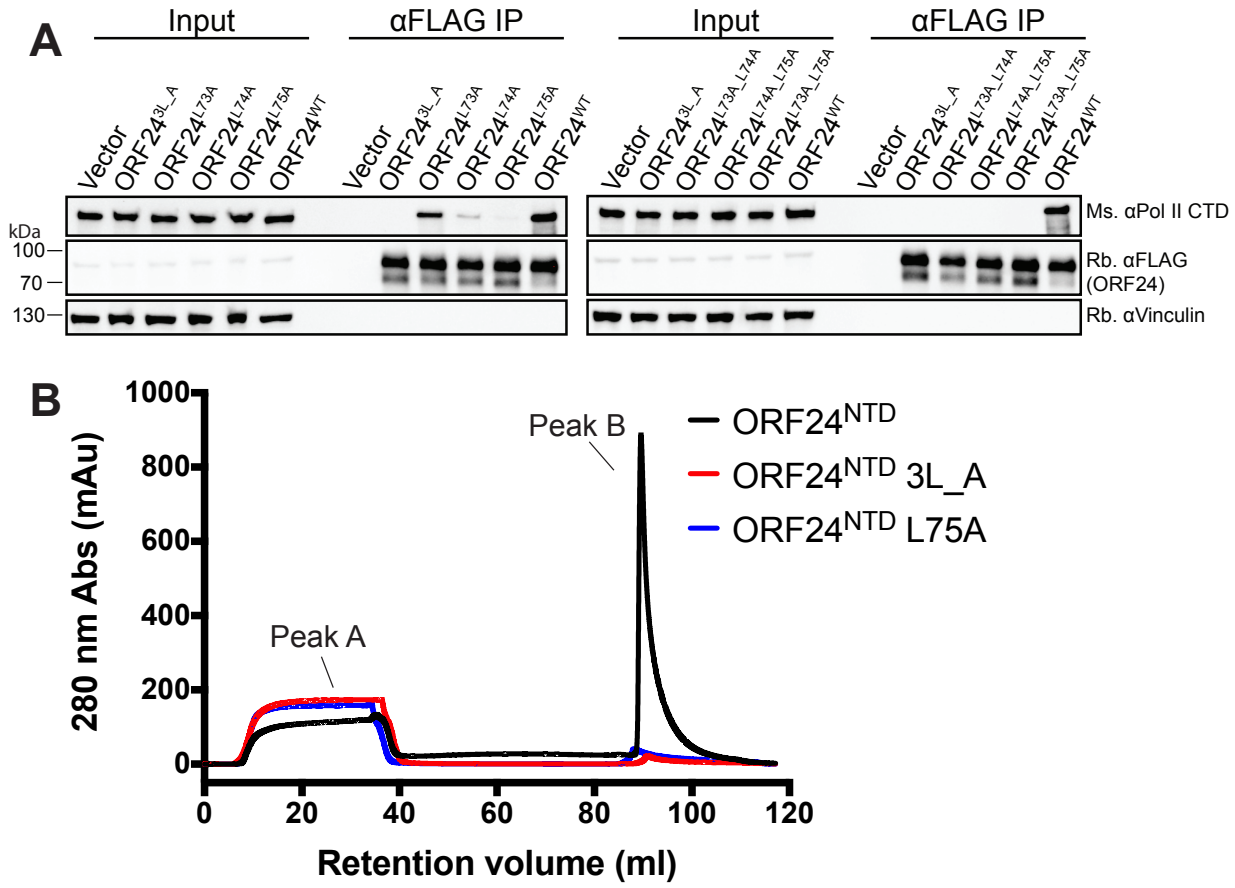


Figure 3.4. Effect of mutations to the leucine residues in the N-terminus of ORF24 are additive. (A) HEK293T cells were transfected with the indicated ORF24 plasmids, then subjected to co-IP using α–FLAG beads followed by western blot with the indicated antibody to detect ORF24 and Pol II. Input represents 2.5% of the lysate used for co-IP. Vinculin served as a loading control. **(B)** Absorbance trace of StrepTrap purification for ORF24^{NTD}-Strep, ORF24^{NTD} 3L_A-Strep, and ORF24^{NTD} L75A-Strep, measured at 280 nm. Peak A represents the flow-through for each run, and peak B represents the elution of each protein.

ORF24^{NTD} may be amenable to structural determination by NMR

Thus far, our attempts at expression and purification of ORF24 have led to the development of an expression construct in plasmid pE-SUMO3 which contains an N-terminal 6xHis-SUMO3 solubility tag. In order to observe the purification of ORF24^{NTD} by western blot, a C-terminal Strep-tag II was appended to ORF24^{NTD} (**Figure 3.5A**). Having two affinity tags on the protein allows us to perform a nickel purification (HisTrap) followed by cleavage of the SUMO3 tag and a Strep purification on a StrepTrap (**Figure 3.5 B, C**). The most concentrated fractions from the Strep purification can be further purified and buffer exchanged using gel filtration chromatography (**Figure 3.5D**). This purification scheme yields highly purified protein, which can be subjected to 2D ^{1}H-^{15}N HSQC NMR.

While we are still working toward optimizing the experimental conditions for NMR, our initial 2D HSQC spectrum looks promising. Ideally, the peaks should be well dispersed throughout the spectrum, which would suggest that the protein domain is folded. The ORF24^{NTD} spectrum has areas that are well dispersed; however, there are peaks that show a significant amount of overlap (**Figure 3.5E**). A possible reason for these results is that the protein is not well-folded in the current buffer conditions. Alternatively, since secondary structure prediction programs struggle with assigning domains for viral proteins, it is possible that the fragment is missing an essential region required for proper folding. In either case, these are the first data to suggest that fragments of ORF24 can be recombinantly purified and may be amenable to structural determination by NMR.

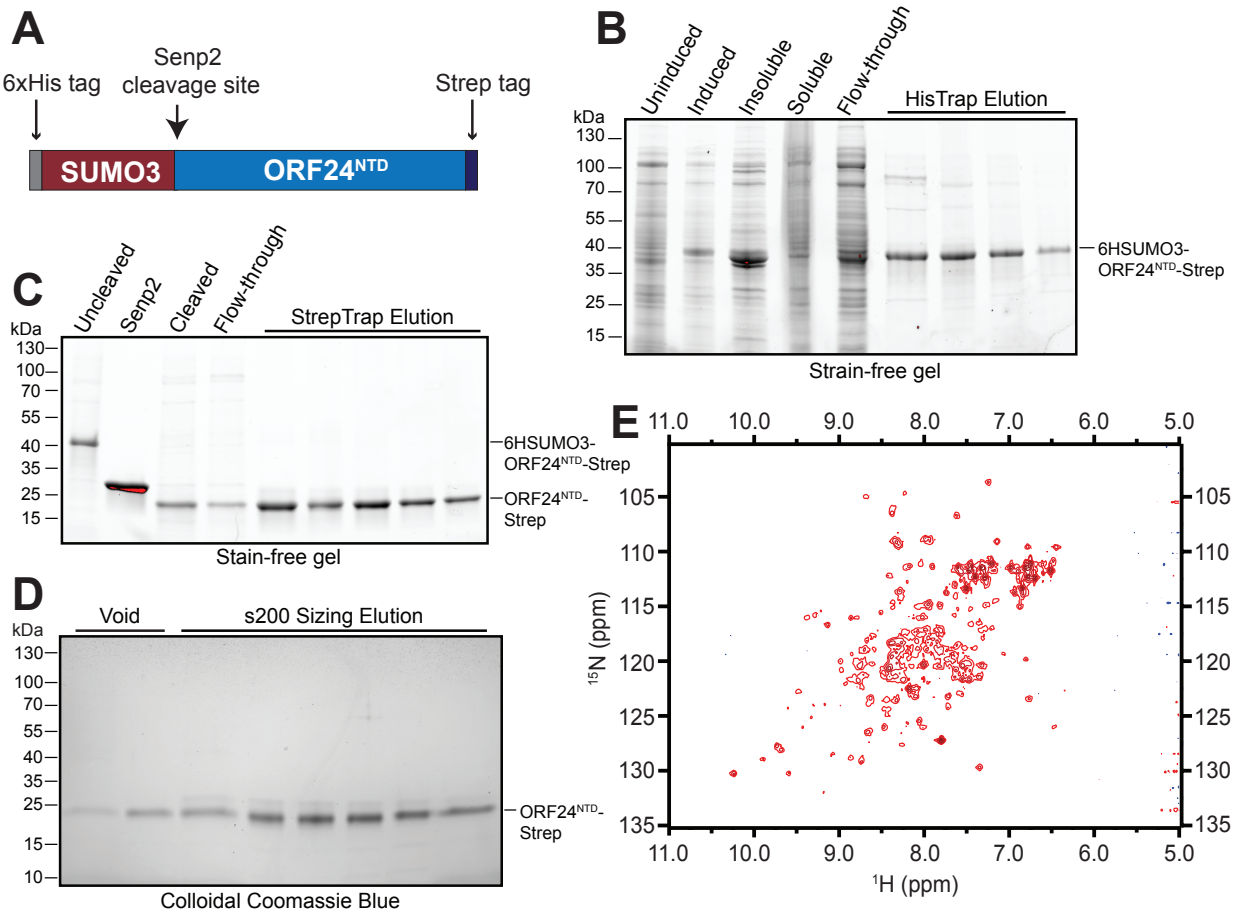


Figure 3.5. ORF24^{NTD} may be amenable to structural determination by NMR. (A) Diagram of the construct used for robust recombinant expression and purification of ^{15}N-labeled ORF24^{NTD}-Strep. **(B)** Stain-free gel demonstrating the purity of 6xHis-SUMO3-ORF24^{NTD}-Strep following nickel affinity purification. **(C)** Stain-free gel showing the effectiveness of SUMO3 tag removal by SenP2 protease and subsequent Strep purification. **(D)** Colloidal Coomassie blue gel showing the purity and molecular weight (~23 kDa) of ORF24^{NTD}-Strep eluted from the Superdex 200 gel filtration column. **(E)** The 2D-^{1}H-^{15}N HSQC spectrum at 900 mHz showing regions of well dispersed peaks suggesting that portions of the protein fragment are folded and may be amenable to structural determination by NMR.

Discussion

Determining the precise role of ORF24 in transcription will improve our understanding of late gene regulation in beta and gammaherpesviruses. One of the obstacles in learning more about ORF24 (and the other vTAs) has been the nearly insurmountable challenge posed by the inability to express and purify these proteins recombinantly. Here, we provide an example where purification of a domain of one of the vTAs provides invaluable *in vitro* information about how the protein co-opts mammalian transcriptional machinery for late gene expression. We show that recombinantly expressed ORF24^{NTD} interacts with Pol II from mammalian whole cell lysate and use negative stain EM to gain insight into the ORF24-Pol II interaction. With the structural information obtained with the NTD of ORF24 we determine that ORF24^{NTD} interacts with the CTD repeats of Pol II. Finally, we describe our ongoing efforts toward gaining a better understanding of this essential interaction.

Single particle EM has greatly advanced our understanding of transcription initiation (Haberle and Stark, 2018; He et al., 2013; Sainsbury et al., 2015). Applying this technique to the KSHV late gene PIC has the potential to reveal the conserved mechanisms of late gene regulation in beta and gammaherpesviruses. We reasoned that since ORF24 interacted strongly with Pol II from mammalian whole cell lysate, it should interact with purified Pol II and we could use negative stain EM to determine how the ORF24-Pol II interaction is mediated. For our negative stain EM experiments, we chose to include the GST affinity tag on ORF24^{NTD} to allow for easier detection of any extra density that could be attributed to GST-ORF24^{NTD}. Based on the molecular weight of GST-ORF24^{NTD}, we expected it to have a diameter of ~50 Angstroms; however, the diameter of the GST-ORF24^{NTD} density was ~65 Angstroms. In addition, the density appeared blurred-out and less intense than the density for Pol II, which indicated the protein interacted with a dynamic region of Pol II. Due to the observed location and flexibility, our initial impression was that the fragment interacted with the CTD. This was somewhat surprising given that CTD-interacting proteins tend to recognize a specific phosphorylation pattern and bind the CTD through well-defined protein domains (Meinhart et al., 2005). Since ORF24^{NTD} binds to a highly flexible region of Pol II, without the TBP-like domain of ORF24, it will be difficult to improve upon the current density map of the PIC with GST-ORF24^{NTD}. With full length ORF24, we anticipate that the TBP-like domain would bind the TATA box and stabilize the complex to decrease the amount of heterogeneity in the data.

That ORF24 replaces TBP at late gene promoters and interacts directly with Pol II makes it reminiscent of bacterial sigma factors. During eukaryotic transcription initiation, the interaction between TBP and Pol II is bridged through GTFs; whereas, in bacteria, sigma factors bind both the DNA and polymerase directly (Decker and Hinton, 2013). In eukaryotes, TFIID binds the TATA box and recruits the other GTFs. Previous work has shown that ORF24 replaces TBP at late gene promoters and co-immunoprecipitates with unphosphorylated Pol II (Davis et al., 2015). ORF24 may bind the polymerase directly as a means to recruit Pol II to late gene promoters in the absence of TFIID. Unlike ORF24, during PIC complex formation, TFIIB, TFIIE, and TFIIF bind Pol II at regions aside from the CTD (Bushnell et al., 1996; Sainsbury et al., 2015). Known CTD-binding proteins in humans include the homolog of the yeast coactivator Mediator complex (Näär et al., 2002). Studies have shown that phosphorylation of the CTD will inhibit the interaction of

Mediator with Pol II (Søgaard and Svejstrup, 2007). We predict that if ORF24 behaves similarly to a CTD-interacting coactivator, phosphorylation of the CTD would block the ORF24-Pol II interaction.

Both mutants of ORF24^{NTD} that do not interact with Pol II in HEK293T cells, 3L_A and L75A, could not be purified despite using the same purification pipeline as is used for ORF24^{NTD}. While all three proteins (WT, 3L_A, and L75A) show similar levels of recombinant expression, for 3L_A and L75A the majority of the protein is found in the insoluble fraction of the lysed bacteria. This was despite the use of a SUMO3 solubility tag that is often appended to aid with the expression and solubility of difficult-to-express proteins (Marblestone et al., 2006). Since mutations in amino acids 73-75 render ORF24^{NTD} insoluble, it is likely that the leucine to alanine mutations destabilized the protein. However, without a structure of ORF24^{NTD}, it will be difficult to determine how ORF24 interacts with the CTD. Studies of CTD-interacting proteins have identified domains that bind phosphorylated CTD (Fabrega et al., 2003; Meinhart and Cramer, 2004; Verdecia et al., 2000). In these examples, the CTD binds to a conserved groove through van der Waals and hydrophobic interactions where the phosphorylation pattern of the CTD contributes to binding (Meinhart et al., 2005).

The best characterized interaction between a protein and unphosphorylated CTD is that of Mediator-Pol II. Mediator is a multi-subunit complex, and its interactions with the Pol II CTD are extensive involving numerous Mediator subunits (Harper and Taatjes, 2018; Plaschka et al., 2015; Robinson et al., 2016, 2012). Recent findings suggest that alpha helical domains in Mediator subunits are involved in the CTD interactions (Plaschka et al., 2015; Robinson et al., 2016). Based on this information, it is likely that the leucine residues in ORF24^{NTD} are involved in a fold that is required for CTD-binding, and mutation of the residues disrupts that fold. One way to examine this further would be to generate leucine to isoleucine or valine mutations at positions 73-75. These amino acids are closer to leucine in their side chain bulkiness and may allow for proper folding of ORF24^{NTD}. If these amino acids recover the structure of the protein, we expect that the Pol II-interaction will be recovered and ORF24^{NTD} harboring valines or isoleucines at positions 73-75 will be soluble when expressed recombinantly.

While mutating the leucine residues to similar amino acids might rescue Pol II binding, structural determination of ORF24^{NTD} would unambiguously identify their role in CTD binding. Our initial attempts at solving the structure of ORF24^{NTD} by NMR indicated that our current sample conditions are not ideal. Since several factors, such as salt concentration, buffer, and pH can affect the solubility and conformation of a protein in solution, it will be necessary to test the effect of different experimental conditions on the 2D HSQC of ORF24^{NTD}. In addition, protein secondary structure prediction programs have difficulty with viral proteins. Thus, it is possible that the fragment we designated as ORF24^{NTD} is truncating an important secondary structure and causing ORF24^{NTD} to be misfolded. Identifying a better construct for structural studies can be accomplished by performing limited proteolysis on a larger N-terminal fragment of ORF24. This will aid in defining a truncation of the protein that is well-folded, stable in solution, and still interacts with the CTD. With such a fragment of ORF24, NMR could be used to determine both the structure of ORF24^{NTD} and how it mediates the interaction with the Pol II CTD.

In summary, ORF24 is a viral transcriptional activator that replaces TBP at late gene promoters and interacts with Pol II to direct transcription of viral late genes. That it

binds both the unphosphorylated CTD of Pol II and promoter DNA makes it unique among known CTD-interacting proteins. Since the two functions of ORF24 are genetically separable, they can be characterized independently of one another. This is advantageous because there is a greater likelihood of successful expression and purification of ORF24 domains than of the full-length protein. Lastly, obtaining structural information about any region of ORF24 will vastly improve our understanding of its distinct role in the regulation of late gene transcription.

Materials and Methods

Plasmids and Plasmid Construction

The following ORF24 fragments: residues 1-201 ($ORF24^{NTD}$), residues 1-271, residues 202-401, residues 134-266, and residues 1-133 were PCR amplified from ORF24-3xFLAG pCDNA4 (Davis et al., 2015) with primers to introduce BamHI and NotI sites and cloned into 3xFLAG pCDNA4 using T4 DNA ligase (New England Biolabs). All primer sequences are listed in **Table 3.1**. To generate the plasmid for GST-$ORF24^{NTD}$ expression, $ORF24^{NTD}$ (residues 1-201) was PCR amplified from ORF24-3xFLAG pCDNA4 with primers to introduce BamHI and NotI sites and cloned into pGEX4T1 using T4 DNA ligase. pGEX4T1 encodes an N-terminal GST tag followed by a thrombin cleavage site. To generate the plasmid for MBP-$ORF24^{NTD}$ expression, $ORF24^{NTD}$ was PCR amplified from ORF24-3xFLAG pCDNA4 with primers to introduce SacI and BamHI sites and cloned into pMAL-c2X using T4 DNA ligase. To generate the plasmid for MBP-$ORF24^{NTD}\Delta$LLL expression, $ORF24^{NTD}\Delta$LLL was PCR amplified from plasmid $ORF24^{AAA}$-3xFLAG pCDNA4 with primers to introduce SacI and BamHI sites and cloned into pMAL-c2X. Plasmid $ORF24^{AAA}$-3xFLAG pCDNA4 was previously described (Davis et al., 2015). pMAL-c2X encodes an N-terminal MBP tag and the plasmids were cloned to express $ORF24^{NTD}$ with an uncleavable MBP tag.

To make the plasmid for GST-CTD linker expression, the linker region of Rpb1 (aa 1460-1585) was PCR amplified from HEK293T cell cDNA with primers to introduce BamHI and NotI sites and cloned into pGEX4T1 (described above) using T4 DNA ligase. The 1x CTD and 2x CTD repeat inserts were ordered as oligonucleotides from IDT and were annealed by cooling from 90°C to room temperature in a water bath. The 5x CTD and 10x CTD repeat inserts were ordered as synthesized gene blocks from IDT. The sequence of both of these inserts is in **Table 3.2**. All CTD inserts were cloned into the BamHI and NotI sites of pQLink-GST using InFusion cloning (Clontech). pQLink-GST encodes an N-terminal GST tag followed by a TEV protease cleavage site.

The following mutations were introduced into ORF24-3xFLAG pCDNA4 using two primer site-directed mutagenesis with KAPA HiFi polymerase (Roche): L73A, L74A, L75A, L73A_L75A, and L74A_L75A. The L73A_L74A mutation was introduced using InFusion (Clontech) site-directed mutagenesis. To generate the plasmid for tag-less $ORF24^{NTD}$ expression, $ORF24^{NTD}$ (residues 2-201) were PCR amplified from ORF24-3xFLAG pCDNA4 and cloned into the BamHI and NotI sites of pQLink-MBP using InFusion cloning. The plasmid pQLink-MBP encodes an N-terminal MBP tag followed by a TEV protease cleavage site. For MBP-$ORF24^{NTD}$-Strep expression, a 1x Strep-tag II was appended to the C-terminus of $ORF24^{NTD}$ in the pQLink-MBP plasmid using inverse PCR with Phusion DNA polymerase (New England Biolabs). To make the plasmid for

robust expression of ORF24^{NTD}-Strep, ORF24^{NTD}-Strep was PCR amplified from MBP-ORF24^{NTD}-Strep pQLink-MBP and cloned into the KpnI and EcoRI sites of plasmid 6HSUMO3 using InFusion cloning. Plasmid 6HSUMO3 encodes a 6xHis affinity tagged N-terminal human SUMO3 tag, which can be cleaved by SenP2 protease.

Tissue Culture and Transfections

HEK293T cells (ATCC CRL-3216) were maintained in DMEM supplemented with 10% FBS (Seradygm). For DNA transfections, HEK293T cells were plated and transfected after 24 hours at 70% confluency with PolyJet (SignaGen).

Immunoprecipitation and Western Blotting

Cell lysates were prepared 24 hours after transfection by washing and pelleting cells in cold PBS, then resuspending the pellets in IP lysis buffer [50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor (Roche)] and rotating for 30 minutes at 4°C. Lysates were cleared by centrifugation at 21,000 x *g* for 10 min, then 1 mg of lysate was incubated with pre-washed M2 α-FLAG magnetic beads (Sigma) overnight. The beads were washed 3x for 5 min each with IP wash buffer [50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.05% NP-40] and eluted with 2x Laemmli sample buffer (BioRad).

Lysates and elutions were resolved by SDS-PAGE and analyzed by western blot in TBST (Tris-buffered saline, 0.2% Tween 20) using the following primary antibodies: rabbit α-FLAG (Sigma, 1:2500); rabbit α-Pol II clone N20 (Santa Cruz, 1:2500); mouse α-GST clone 8-326 (Pierce, 1:2000); mouse α-MBP (NEB, 1:10000); mouse α-Pol II CTD clone 8WG16 (Abcam, 1:1000), rabbit α-Vinculin (Abcam, 1:1000). Following incubation with primary antibodies, the membranes were washed with TBST and incubated with the appropriate secondary antibody. The secondary antibodies used were the following: goat α-mouse-HRP (Southern Biotech, 1:5000) and goat α-rabbit-HRP (Southern Biotech, 1:5000).

Protein Expression and Purification

GST-ORF24^{NTD}, GST-CTD linker, and GST. These proteins were expressed in Rosetta 2 cells (EMD Millipore) grown in LB at 37°C and induced at an OD_{600} of 0.700 with 0.5 mM IPTG for 16 hours at 18°C. The cells were harvested by centrifugation at 6500 x *g* for 10 minutes. The cell pellets were either frozen or immediately resuspended in lysis buffer [50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, protease inhibitors (Roche)] and lysed by sonication. The insoluble fraction was removed by centrifugation at 21,000 x *g* for 30 minutes. GST-OR24^{NTD}, GST-CTD linker, and GST were purified on Glutathione Sepharose (GE Healthcare) by batch purification. The proteins were eluted in wash buffer (50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10 mM reduced glutathione and dialyzed into storage buffer (50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol].

GST-xCTD repeats. GST-1xCTD repeat, GST-2xCTD repeats, GST-5xCTD repeats, and GST-10xCTD repeats were expressed in BL21 Star (DE3) cells grown in Overnight Express Instant TB Medium (EMD Millipore) at 37°C and induced at an OD_{600} of 1.0 by decreasing the temperature to 18°C and growing for an additional 16 hours. The cells

were harvested by centrifugation at 6500 x g for 10 minutes. The cell pellets were either frozen or immediately resuspended in lysis buffer [50 mM HEPES, pH 7.4, 300 mM NaCl, 5 mM DTT, 5% glycerol, protease inhibitors (Roche)] and lysed by sonication. The insoluble fraction was removed by centrifugation at 50,000 x *g* for 30 minutes. The proteins were purified as described above and eluted in wash buffer (50 mM HEPES, pH 7.4, 300 mM NaCl, 5 mM DTT, 5% glycerol) containing 10 mM reduced glutathione.

MBP-ORF24^{NTD}. The protein was expressed in Rosetta 2 cells grown in LB at 37°C and induced at an OD_{600} of 0.700 with 0.5 mM IPTG for 16 hours at 18°C. The cells were harvested by centrifugation at 6500 x *g* for 10 minutes. The cell pellets were either frozen or immediately resuspended in lysis buffer [50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, protease inhibitors (Roche)] and lysed by sonication. The insoluble fraction was removed by centrifugation at 21,000 x *g* for 30 minutes. MBP-ORF24^{NTD} was purified by gravity column chromatography with Amylose Resin (New England Biolabs). The protein was eluted in wash buffer (50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10 mM maltose and dialyzed into storage buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 1mM DTT, 10% glycerol).

MBP-ORF24^{NTD}-Strep. The protein was expressed in Rosetta 2 cells grown in Overnight Express Instant TB Medium (EMD Millipore) at 37°C and induced at an OD_{600} of 1.0 by decreasing the temperature to 18°C and growing for an additional 16 hours. The cells were harvested by centrifugation at 6,500 x *g* for 10 minutes. The cell pellets were either frozen or immediately resuspended in lysis buffer [100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1 mM EDTA, 0.1% CHAPS, 5% glycerol, protease inhibitors (Roche)] and lysed by sonication. The lysate was cleared by centrifugation at 50,000 x *g* for 30 minutes. The clarified lysate was filtered through a 0.45 µm PES filter (Foxx Life Sciences). The protein was purified on an equilibrated MBPTrap (GE Healthcare) and step-eluted in wash buffer (100 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1 mM EDTA, 0.1% CHAPS, 5% glycerol) containing 10 mM maltose. The fractions containing MBP-ORF24^{NTD}-Strep were pooled and the MBP tag was cleaved overnight at 4°C with 1 mg of TEV protease. Following cleavage of the MBP solubility tag, ORF24^{NTD}-Strep was purified on a StrepTrap (GE Healthcare) and step-eluted in wash buffer containing 2.5 mM desthiobiotin (IBA). The Strep elution fractions containing ORF24^{NTD}-Strep were pooled and sized on a Superdex 200 (GE Healthcare) size exclusion chromatography (SEC) column in SEC buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM TCEP, 5% glycerol). The fractions containing ORF24^{NTD}-Strep were pooled and concentrated on a 10K Amicon Ultra-15 concentrator (EMD Millipore). Protein aliquots were flash frozen and stored at -70°C.

^{15}N-labeled 6xHis-SUMO-ORF24^{NTD}-Strep. The protein was expressed in NiCo21 (DE3) cells (New England Biolabs). 3 mL of LB with 1% glucose containing 50 µg/mL kanamycin was grown at 37°C for 3 hours. After 3 hours of growth, 1 mL of the culture was added to 50 mL of non-inducing MDAG media [2 mM MgSO4, 0.2 mL of 5000x metals mix (Blommel et al., 2007), 0.5% glucose, 0.25% aspartate, 1x NPS (50 mM NH_4Cl, 5 mM Na_2HPO_4, 25 mM KH_2PO_4, 5 mM Na_2SO_4), 0.1 mg/mL methionine, 0.1 mg/mL 17A (no C, Y, M) mix, 1 mL of 1000x B12 vitamin cocktail (Blommel et al., 2007)] containing 50 µg/mL kanamycin and incubated at 25°C for 24 hours. The following day, 20 mL of the

overnight MDAG culture was added to 1 L of M9-^{15}N media [1x M9 salts (50 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl), 0.2 mL of 5000x metals mix, 1 mL of 1000x B12 vitamin cocktail, 2 mM $MgSO_4$, 30 µg/mL thiamine, 4 g/L glucose, 1 g/L $^{15}NH_4Cl$, 0.1 mM $CaCl_2$] containing 50 µg/mL kanamycin resulting in an O.D. of ~ 0.150. The culture was incubated at 37°C and induced at an OD_{600} of 1.0 with 1 mM IPTG at 18°C for 16 hours. The cells were harvested by centrifugation at 6,500 x *g* for 10 minutes. The cell pellets were either frozen or immediately resuspended in IMAC buffer (50 mM HEPES base, 50 mM HEPES acid, pH 7.5, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP, 10% glycerol) containing 1% Triton X-100 and protease inhibitors (Roche) and lysed by sonication. The protein was purified on a HisTrap (GE Healthcare) and step-eluted with IMAC buffer containing 500 mM imidazole. The fractions were pooled and dialyzed into Strep running buffer (50 mM Tris pH 8.0, 200 mM NaCl, 1mM TCEP, 10% glycerol) and cleaved with SenP2 to remove the SUMO3 solubility tag. The following day, ^{15}N-labeled ORF24^{NTD}-Strep was purified on a StrepTrap (GE Healthcare) and step-eluted with Strep running buffer containing 2.5 mM desthiobiotin (IBA). The Strep elution fractions containing ORF24^{NTD}-Strep were pooled and sized on a Superdex 200 (GE Healthcare) size exclusion chromatography (SEC) column in SEC buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM TCEP, 5% glycerol). The fractions containing ORF24^{NTD}-Strep were pooled and concentrated on a 10K Amicon Ultra-15 concentrator (EMD Millipore). Protein aliquots were flash frozen and stored at -70°C.

Pulldown Assays

To test the interaction between GST-ORF24^{NTD} and Pol II from mammalian cells, 10 µg GST-ORF24^{NTD} or 10 µg GST was added to 20 µl of washed Glutathione Magnetic Agarose (Pierce) along with 250 µg of 293T whole cell lysate. IP wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% NP-40, 1 mM EDTA) was added to a final volume of 300 µl. The samples were rotated at 4°C for 1 hour. Following the pulldown, the samples were washed with IP wash buffer three times for 5 minutes each time. After the last wash, the protein was eluted with 2x Laemmli sample buffer (BioRad). The pulldown to test the interaction between GST-CTD repeats or GST-CTD linker and MBP-ORF24^{NTD} or MBP-ORF24$^{NTD}\Delta$LLL were performed as described above.

To test the interaction between ORF24^{NTD} and the CTD, 10 µg of GST-CTD repeats (Sigma) or 10 µg of BSA (negative control) were added to 20 µl of washed Glutathione Magnetic Agarose along with 10 µg of ORF24^{NTD}-Strep. IP wash buffer was added to a final volume of 300 µl. The samples were rotated at 4°C for 1 hour then washed three times for 5 minutes each time. The protein was eluted with 2x Laemmli sample buffer.

To test the interaction between ORF24^{NTD} and xCTD repeats, 10 µg of ORF24^{NTD}-Strep or 10 µg BSA (negative control) were added to 20 µl of washed MagStrep “type 3” XT Beads (IBA) along with 10 µg of GST-xCTD repeats. IP wash buffer was added to a final volume of 300 µl. The pulldowns were rotated at 4°C for 1 hour then washed three times for 5 minutes each time. The protein was eluted with 2x Laemmli sample buffer.

Nuclear Magnetic Resonance Spectroscopy

NMR data were collected at the Central California NMR Facility. Spectra were collected on a 900 MHz Bruker Avance spectrometer. Data were processed with NMRPipe and Nova was used to visualize processed data. 2D ^{1}H-^{15}N HSQC was used as an initial

screen to determine whether structural determination of ORF24^{NTD} may be achieved by NMR.

Negative Stain Electron Microscopy

PIC assembly and purification. TBP, TFIIA, and TFIIB were recombinantly expressed and purified from *E. coli*. Pol II was immunopurified from HeLa cell nuclear extracts following previously published protocols (Knuesel et al., 2009; Revyakin et al., 2012). The DNA construct was described in (He et al., 2013) and is a SCP (Juven-Gershon et al., 2006) containing a BREu element upstream of the TATA box (Tsai and Sigler, 2000) and a SalI restriction enzyme site downstream of the INR element for purification purposes. A biotin tag was engineered at the 5' end of the template strand (Integrated DNA Technologies). The duplex DNA was generated by annealing the template strand with equimolar amounts of single-stranded non-template DNA at a final concentration of 50 µM in water. The annealing reaction was carried out at 100°C for 5 minutes and gradually cooled down to room temperature within 1 hour.

PICs were assembled in assembly buffer (12 mM HEPES, pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 60 mM KCl, 1 mM DTT, 0.05% NP-40). Purified proteins and nucleic acid were sequentially added to into the assembly buffer: Pol II, TFIIB, TBP/TFIIA, DNA, GST-ORF24^{NTD}. Following assembly of the PICs, the reaction was incubated at 28°C for 15 minutes using a 1:10 dilution of magnetic streptavidin T1 beads (Invitrogen), which had been previously equilibrated in assembly buffer. The beads were washed three times with wash buffer (10 mM HEPES, 10 mM Tris, pH 7.9, 5% glycerol, 5 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.05% NP-40). The complex was eluted by incubation at 28°C for 1 hour in digestion buffer (10 mM HEPES, pH 7.9, 5% glycerol, 10 mM MgCl2, 50 mM KCl, 1 mM DTT, 0.05% NP-40, 1 unit µL^{-1} BSA-free SalI-HF (New England Biolabs). After elution, purified PIC complexes were crosslinked on ice in 0.05% glutaraldehyde for 5 minutes then immediately used for EM sample preparation.

Electron microscopy. Negative stain samples of PIC complex were prepared on a 400 mesh copper grid containing a continuous carbon supporting layer. The grid was plasma-cleaned for 10 seconds using a Solarus plasma cleaner (Gatan) equipped with 75% argon/25% oxygen. An aliquot (3 µL) of the purified sample was placed onto the grid and allowed to absorb for 5 minutes at 100% humidity. The sample was stained with five successive 75 µL drops of 2% (w/v) uranyl acetate solution rocking 10 seconds on each drop followed by blotting to dryness. Data collection was performed on a Tecnai F20 Twin transmission electron microscope operating at 120 keV at a nominal magnification of X80,000. The data were collected on a Gatan 4k X 4k CCD camera using low-dose procedures (20 e^- Å $^{-2}$ exposures).

Image processing. Data pre-processing was performed using the Appion processing environment (Lander et al., 2009). Particles were automatically selected from the micrographs using a difference of Gaussians (DoG) particle picker (Voss et al., 2009). The contract transfer function (CTF) of each micrography was estimated using both ACE2 and CTFFind programs during data collection (Mallick et al., 2005; Mindell and Grigorieff, 2003), the phases were flipped using CTFFind, and particle stacks were extracted using a box size of 256 X 256 pixels from images whose ACE2 confidence value was greater

than 0.8 followed by normalization using the XMIPP program to remove pixels which were above or below 4.5σ of the mean value (Sorzano et al., 2004). The particle stack was binned by a factor of two and two-dimensional classification was conducted using iterative multireference alignment analysis (MSA-MRA) within the IMAGIC software (Van Heel et al., 1996). Class averages were manually selected to create a new particle stack for reconstruction.

***Three-dimensional reconstruction*.** The three-dimensional reconstruction was conducted using an iterative multi-reference projection-matching approach containing libraries from the EMAN2 (Tang et al., 2007). The resolution of the reconstruction was estimated to be 20 Å for GST-ORF24^{NTD}.

Acknowledgements

We thank all members of the Glaunsinger and Coscoy labs, in particular Allison Didychuk and Zoe Davis for their helpful suggestions and guidance during this study.

Table 3.1. List of synthetic DNA oligonucleotides used in this study.

Primer name	Primer Sequence 5' → 3'
ORF24 aa 1-201_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG
ORF24 aa 1-201_pCDNA_R	TCGAGCGGCCGCCCCTCCAGGAGTGCAAAATAATTTTGATAGATTG
ORF24 aa 1-271_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG
ORF24 aa 1-271_pCDNA_R	TCGAGCGGCCGCTTCTTGACGTCCTGGTGCTTACTCT
ORF24 aa 202-401_pCDNA_F	GCTCGGATCCATGAGCCTGAAGCATCTCTCGTTTTCA
ORF24 aa 202-401_pCDNA_R	TCGAGCGGCCGCTTGCTGCCCAGAGTCCGC
ORF24 aa 134-266_pCDNA_F	GCTCGGATCCATGGATAAACCCTGCGAATTCAGGGAG
ORF24 aa 134-266_pCDNA_R	TCGAGCGGCCGCTTCTTACTCTGTTTCGACAGGTTCTTCAG
ORF24 aa 1-133_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG
ORF24 aa 1-133_pCDNA_R	TCGAGCGGCCGCCCACAATCATCGGTAAGTTCCCATGATC
ORF24 aa 1-201_pGEX_F	GCGTGGATCCATGGCAGCGCTCGAGG
ORF24 aa 1-201_pGEX_R	CGATGCGGCCGCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGTG
ORF24 aa 1-201_pMAL_F	ATTCGAGCTCAATGGCAGCGCTCGAGGG
ORF24 aa 1-201_pMAL_R	TAGAGGATCCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGT
Rpb1_Linker aa 1460-1585_F	GCGTGGATCCCTGGGCCAGCTGGCTC
Rpb1_Linker aa 1460-1585_R	CGATGCGGCCGCTGGTGAAGGGATGTAGGGGCT
1xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATTCCCCCACTTCACCCTCCTAGGCGGCCGCCTAGGACCC
1xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGAGGGTGAAGTGGGGGAATAGGATCCCTGAAAATAA
2xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATTCTCCGACCTCTCCATCATACAGCCCTACCTCCCCGTCCTAGGCGGCCGCCTAGGACCC
2xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGACGGGGAGGTAGGGCTGTATGATGGAGAGGTCGGAGAATAGGATCCCTGAAAATAA
ORF24 (L73A) _F	CCCCAGGAGTGCTCTGCGTATCGCTCTGCAG
ORF24 (L73A) _R	CTGCAGAGCGATACGCAGAGCACTCCTGGGG
ORF24 (L74A) _F	GCGTTCCCCCAGGGCTAGTCTGCGTATCGC
ORF24 (L74A) _R	GCGATACGCAGACTAGCCCTGGGGGAACGC
ORF24 (L75A) _F	GGAGGCGTTCCCCCGCGAGTAGTCTGCGTA
ORF24 (L75A) _ R	TACGCAGACTACTCGCGGGGGAACGCCTCC
ORF24 (L73A_L75A) _F	GGCGTTCCCCCGCGAGTGCTCTGCGTATCGCTCTGCAGGC
ORF24 (L73A_L75A) _R	GCCTGCAGAGCGATACGCAGAGCACTCGCGGGGGAACGCC
ORF24 (L74A_L75A) _F	GGAGGCGTTCCCCCGCGGCTAGTCTGCGTATCGCTCTGC
ORF24 (L74A_L75A) _R	GCAGAGCGATACGCAGACTAGCCGCGGGGGAACGCCTCC
ORF24 (L73A_L74A) _F	GCAGAGCAGCCCTGGGGGAACGCCTCCACC
ORF24 (L73A_L74A) _R	CCAGGGCTGCTCTGCGTATCGCTCTGCAGG
ORF24 aa 2-201_pQLink_F	GTATTTCCAGGGATCCGCAGCGCTCGAGGGC
ORF24 aa 2-201_pQLink_R	GGGTCCTAGGCGGCCGCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGTG
ORF24 aa 2-201_Strep_pQLink_F	TGGAGCCATCCGCAGTTTGAAAAATAAGCGGCCGCCTAGGACCCAGCTTTCTTG
ORF24 aa 2-201_Strep_pQLink_R	CTCCAGGAGTGCAAAATAATTTTGATAGATTG
ORF24 aa 2-201_Strep_6HSUMO3_F	CAGCAGACGGGAGGGGCAGCGCTCGAGGGC
ORF24 aa 2-201_Strep_6HSUMO3_R	GACGGAGCTCGAATTTTATTTTTCAAACTGCGGATGGCTCC

Table 3.2. Nucleotide sequence of synthetic gene blocks.

Gene block Name	Sequence 5' → 3'
5xCTD_repeat_pQLink_insert	TTATTTTCAGGGATCCTATAGCCCCACGAGCCCTTCTTACTC TCCGACCTCACCATCCTATTCGCCTACTAGCCCGAGTTACAG TCCCACATCTCCGTCCTACAGTCCAACAAGCCCCTCCTAGG CGGCCGCCTAGGACCC
10xCTD_repeat_pQLink_insert	TTATTTTCAGGGATCCTATAGCCCGACGTCGCCGAGTTATTC ACCTACGTCCCCATCATACTCCCCCACGAGCCCTAGTTATTC GCCAACTTCCCCGAGCTATTCCCCAACATCACCCAGCTATA GTCCCACTTCACCCTCCTATTCACCTACGAGTCCATCTTATT CTCCAACCAGTCCTTCGTACTCACCCACGTCCCCATCGTATT CTCCTACTTCCCCTAGCTAGGCGGCCGCCTAGGACCC

Chapter 4: Conclusions and Future Directions

Conclusions

Here, we revealed the extent of the interactions between members of the viral transcriptional activator (vTA) complex with a focus on the KSHV proteins ORF18 and ORF24. Specifically, we determined that disruptions to the organization of the complex cause a defect in the expression of late genes. We found that all members of the KSHV vTA complex co-purify with ORF18, suggesting that the interactions between the vTAs are quite stable. We identified residues in ORF18 that disrupt its interaction with ORF30 and render an ORF18-null virus complemented with the ORF18 mutant unable to express late genes. In addition, we determined that ORF24 interacts directly with the carboxy terminal domain (CTD) of Pol II and began to discern how the ORF24-CTD interaction is mediated. Lastly, we identified a fragment of ORF24 that may be amenable to structural determination by NMR spectroscopy. Collectively, this work elucidates previously unappreciated interactions between members of the vTA complex and reveals the challenges posed by the biochemical characterization of this unique mechanism of transcriptional regulation.

Future Directions

The last several years have seen many advances in the field of late gene regulation in beta and gammaherpesviruses. Among them was the confirmation that the origin of lytic replication (OriLyt) is required in *cis* to license the robust expression of late genes (Djavadian et al., 2016; Nandakumar and Glaunsinger, 2019). We also learned that continued viral DNA replication is required for late gene expression and the formation of viral replication compartments (Li et al., 2018). Furthermore, we expanded our understanding of the minimum set of viral proteins required to activate a model late gene promoter absent an OriLyt in *cis* (Aubry et al., 2014; Davis et al., 2016). Finally, we determined that the vTAs form a complex, and the intermolecular interactions between members of the complex and with Pol II are required to activate the transcription of late genes (Aubry et al., 2014; Castañeda and Glaunsinger, 2019; Davis et al., 2015, 2016; Nishimura et al., 2017; Pan et al., 2018). Despite these advances, with the exception of pUL79 (homolog of ORF18 in KSHV) (Perng et al., 2014), ORF24 (Davis et al., 2015; Gruffat et al., 2012; Wyrwicz and Rychlewski, 2007), and ORF34 (Davis et al., 2016; Nishimura et al., 2017), the functions of the vTAs remain unknown. Additionally, the link between viral DNA replication and late gene transcription along with the requirement for the OriLyt in *cis* are poorly understood. Since we have just begun to uncover the complexities of this mechanism of late gene transcriptional regulation, there are many questions that remain unanswered. Among them are:

1. What is the link between viral DNA replication and late gene transcription?
2. Do the vTAs mimic general transcription factors (GTFs)?
3. What is the stoichiometry of the individual components of the vTA complex?
4. What GTFs are recruited to the late gene transcription pre-initiation complex (PIC), and how does the recruitment mechanism differ from the proposed PIC assembly models?
5. Are other viral proteins required for robust transcription of late genes?

Link Between DNA Replication and Late Gene Transcription

The link between viral DNA replication and late gene expression is poorly understood. Studies in EBV (Djavadian et al., 2016) and KSHV (Nandakumar and Glaunsinger, 2019) suggest that the origin of lytic replication (OriLyt) is required in *cis* for the activation of late gene promoters. Recently, in a chromatin immunoprecipitation (ChIP) experiment, our lab found that ORF24 and ORF34 ChIP to both origins of lytic replication: OriLyt-R and OriLyt-L (Nandakumar and Glaunsinger, 2019). This suggests that there is likely a direct link between DNA replication and late gene transcription. Since we can detect late gene promoter activation in a transfection based assay where the vTAs and a firefly luciferase reporter under the control of a late gene promoter are transfected into HEK293T cells (Aubry et al., 2014; Castañeda and Glaunsinger, 2019; Davis et al., 2016), there may be some amount of basal transcription that can occur from late gene promoters absent the OriLyt. This raises the question of what function may be attributed to the OriLyt during late gene transcription.

One possibility is that the OriLyt is a *cis*-acting enhancer element that is required for the robust transcription of late gene. Enhancer elements are DNA sequences that are bound by activator proteins that transmit information from the enhancer to the core promoter. There are examples of enhancer elements in viruses, such as simian virus 40 (SV40) and HCMV, the enhancers of which are added to plasmids to increase gene expression. In addition, enhancer elements have been found in a large number of human genes and play important roles at various stages of development (Shlyueva et al., 2014). In KSHV, when the viral genome is replicated during the lytic phase of the viral lifecycle, repressive chromatin marks are removed and newly replicated genomes are not chromatinized. Therefore, it is possible that the enhancer function of the OriLyt is silenced until the viral genome is replicated.

Another possibility is that there is a direct physical connection between the DNA replication machinery and the viral transcription pre-initiation complex (vPIC). An interaction between these two complexes might ensure late gene transcription is not licensed until DNA replication has begun and viral genomes are ready for packaging. To identify DNA replication proteins that interact with vTA components, one could perform immunoprecipitation-mass spectrometry (IP-MS) with any of the affinity tagged vTA iSLK cell lines. Any potential hits could be confirmed by co-immunoprecipitation (co-IP) followed by western. It would then be necessary to determine whether the proteins in the vTA complex interact directly with DNA replication machinery. This could be accomplished by identifying mutations that disrupt specific interactions. A late gene expression defect resulting from the perturbation of direct interactions between components of the vTA complex and the DNA replication machinery would suggest that late gene expression requires a direct physical connection between these two complexes.

Role of vTAs in Late Gene Transcription

While we know that at least six viral proteins are required for late gene transcription, we do not understand their function at late gene promoters. Work in CMV suggests that pUL79 (ORF18 homolog in KSHV) interacts directly with Pol II and is involved in transcription elongation of all classes of viral genes (Perng et al., 2014). However, KSHV ORF18 does not interact with Pol II (unpublished data) and an ORF18 null virus shows a decrease in the mRNA levels of late genes only (Gong et al., 2014),

which suggests that the function of these homologues is different in these two viruses. In addition, ORF18 co-purifies with the other vTAs (Castañeda and Glaunsinger, 2019), which suggests that in KSHV, ORF18 is involved in transcription initiation rather than elongation. One reason for this is that none of the other ORF18-interacting partners have been implicated in the transcriptional elongation of viral genes. These disparate findings highlight the extent to which we do not fully understand the functions performed by the vTAs. One of the obstacles in identifying their functions is that they are difficult to express and purify recombinantly. With recombinant protein, we would be able to perform experiments such as limited proteolysis to identify fragments of the vTAs that are well folded and stable in solution. These fragments could then be expressed and purified for use in structural studies to identify folded domains and determine whether any of the vTAs share structural similarity with the GTFs.

Determine the Stoichiometry of the vTA Complex

In addition to determining the function of the vTAs, it will be important to define the stoichiometry of the complex. Our data suggests that the complex contains at least one polypeptide of each vTA as shown by western blot when the complex is co-purified using ORF18 as bait (Castañeda and Glaunsinger, 2019). However, we do not know whether any of the vTAs form multimers and if they do, whether these multimers are present in the vTA complex or if they have a different role altogether. One way to determine if the vTAs multimerize is to perform co-immunoprecipitation experiments where the vTAs are differentially tagged. A caveat of this experiment is that the differently tagged vTAs could interact with the same cellular protein and lead to a false positive result. The best way to determine whether any of the vTAs form multimers is to perform size exclusion chromatography on purified protein as was done for the KSHV protein ORF68 (Gardner and Glaunsinger, 2018). This method conclusively determines whether a protein exists in a multimeric state in solution (Hong et al., 2012).

In an effort to define the stoichiometry of the vTA complex, we have begun to optimize the purification of the complex with the goal of performing single-particle electron microscopy. Given that purifying these proteins individually has been challenging, it is likely that the vTAs will only be amenable to recombinant expression and purification when co-expressed with an interaction partner or in the context of the complex as a whole. Thus, it will be necessary to combine structural techniques, such as crystallography, NMR spectroscopy, and single-particle electron microscopy to determine the stoichiometry and organization of the vTA complex.

Define the GTF Composition at Late Gene Promoters

In KSHV TATA-binding protein (TBP) and TBP-associated factors (TAFs) do not ChIP to late gene promoters (Davis et al., 2015). However, this same study found that at least two GTFs, TFIIB and TFIIH, localize to late promoters (Davis et al., 2015). The presence of TFIIB is not entirely surprising given that the model late gene promoter, K8.1, contains a BRE^{d} element adjacent to the TATT box, which is one of the sites bound by TFIIB (Thomas and Chiang, 2006). Since TFIIB is involved in recruiting Pol II-TFIIF to the core promoter, it is likely that it performs a similar role at late gene promoters. The functions of TFIIH at the PIC include promoter opening by an ATP-dependent helicase and transcription initiation through phosphorylation of the Pol II CTD by cyclin-dependent

kinase 7 (CDK7) (Greber et al., 2019). Unless any of the vTAs can perform these functions, TFIIB and TFIIH would have to be part of the vPIC. Since the vPIC is composed of both viral and cellular proteins, one important question is what cellular transcription factors are found at late gene promoters and how are they recruited in the absence of TFIID?

Ultimately, the ideal experiment to unambiguously determine whether the exact composition of the vPIC has been identified would be an *in vitro* transcription assay with purified proteins. This may work with only partially purified proteins provided the entire vTA complex can be purified from mammalian cells. While we may not be able to detect robust levels of transcription, even basal transcription would indicate that the minimum set of viral and cellular proteins required for late gene promoter activation have been identified.

Determine whether Other Viral Proteins are Required for Late Gene Transcription

In EBV, late gene expression requires DNA replication, the six vTAs, and the viral kinase BGLF4 (homolog of ORF36 in KSHV) (El-Guindy et al., 2014). During EBV infection, knock down of BGLF4 down-regulates the expression of BGLF3 (homolog of ORF34 in KSHV). In addition, the kinase activity of BGLF4 is necessary for the expression of late genes (El-Guindy et al., 2014). While the same has not been reported in KSHV, it is possible that ORF36 has a similar function. To determine if ORF36 upregulates the transactivation of late promoters, ORF36 could be included in the late gene reporter assay. If ORF36 does not increase reporter activation, it may be required only in the context of viral infection. The functions of ORF36 during viral infection could be tested by characterizing an ORF36-null virus for early gene expression, DNA replication, and late gene expression. One of the functions of ORF36 could be to regulate vTA activity by phosphorylation. We currently do not know whether any of the vTAs undergo posttranslational modifications; however, this could be determined by performing IP-MS experiments from induced cell lysate with antibodies targeting posttranslational modifications.

Collectively, these experiments would enhance our understanding of late gene regulation in beta and gammaherpesviruses and may reveal the molecular mechanisms these viruses utilize to co-opt host transcriptional machinery.

Chapter 5: References

Abernathy, E., Gilbertson, S., Alla, R., and Glaunsinger, B. (2015). Viral nucleases induce an mRNA degradation-transcription feedback loop in mammalian cells. Cell Host Microbe *18*, 243–253.

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

Akula, S.M., Pramod, N.P., Wang, F.Z., and Chandran, B. (2001a). Human herpesvirus 8 envelope-associated glycoprotein B interacts with heparan sulfate-like moieties. Virology *284*, 235–249.

Akula, S.M., Wang, F.Z., Vieira, J., and Chandran, B. (2001b). Human herpesvirus 8 interaction with target cells involves heparan sulfate. Virology *282*, 245–255.

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

Aneja, K.K., and Yuan, Y. (2017). Reactivation and lytic replication of Kaposi's sarcoma-associated herpesvirus: An update. Front. Microbiol. *8*, 1–23.

Arumugaswami, V., Wu, T.-T., Martinez-Guzman, D., Jia, Q., Deng, H., Reyes, N., and Sun, R. (2006). ORF18 is a transfactor that is essential for late gene transcription of a gammaherpesvirus. J. Virol. *80*, 9730–9740.

Aubry, V., Mure, F., Mariame, B., Deschamps, T., Wyrwicz, L.S., Manet, E., and Gruffat, H. (2014). Epstein-Barr Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation Complex. J. Virol. *88*, 12825–12838.

AuCoin, D.P., Colletti, K.S., Xu, Y., Cei, S.A., and Pari, G.S. (2002). Kaposi's Sarcoma-Associated Herpesvirus (Human Herpesvirus 8) Contains Two Functional Lytic Origins of DNA Replication. J. Virol. *76*, 7890–7896.

Avey, D., Tepper, S., Pifer, B., Bahga, A., Williams, H., Gillen, J., Li, W., and Ogden, S. (2016). Discovery of a Coregulatory Interaction between Kaposi ' s Sarcoma- Associated Herpesvirus ORF45 and the Viral Protein Kinase ORF36. J. Virol. *90*, 5953–5964.

Baek, H.J., Kang, Y.K., and Roeder, R.G. (2006). Human mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. J. Biol. Chem. *281*, 15172–15181.

Ballestas, M.E., and Kaye, K.M. (2001). Kaposi's Sarcoma-Associated Herpesvirus Latency-Associated Nuclear Antigen 1 Mediates Episome Persistence through cis-Acting Terminal Repeat (TR) Sequence and Specifically Binds TR DNA. J. Virol. *75*, 3250–3258.

Bauer, D.L.V., Tellier, M., Martínez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep. *23*, 2119–2129.

Blommel, P.G., Becker, K.J., Duvnjak, P., and Fox, B.G. (2007). Enhanced Bacterial Protein Expression During Auto-Induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition. Biotechnol. Prog. *23*, 585–598.

Bodily, J.M., and Meyers, C. (2005). Genetic Analysis of the Human Papillomavirus Type 31 Differentiation-Dependent Late Promoter. J. Virol. *79*, 3309–3321.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. J. Biol. Chem. *285*, 188–196.

Brulois, K., Wong, L.-Y., Lee, H.-R., Sivadas, P., Ensser, A., Feng, P., Gao, S.-J., Toth, Z., and Jung, J.U. (2015). Association of Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is Critical for Late Gene Expression. J. Virol. *89*, 6148–6154.

Brulois, K.F., Chang, H., Lee, a. S.-Y., Ensser, a., Wong, L.-Y., Toth, Z., Lee, S.H., Lee, H.-R., Myoung, J., Ganem, D., et al. (2012). Construction and Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial Artificial Chromosome Clone. J. Virol. *86*, 9708–9720.

Bushnell, D.A., Bamdad, C., and Kornberg, R.D. (1996). A minimal set of RNA polymerase II transcription protein interactions. J. Biol. Chem. *271*, 20170–20174.

Carlson, A., Norwitz, E.R., and Stiller, R.J. (2010). Cytomegalovirus infection in pregnancy: should all women be screened? Rev. Obstet. Gynecol. *3*, 172–179.

Carrozza, M.J., and DeLuca, N.A. (1996). Interaction of the viral activator protein ICP4 with TFIID through TAF250. Mol. Cell. Biol. *16*, 3085–3093.

Castañeda, A.F., and Glaunsinger, B.A. (2019). The interaction between ORF18 and ORF30 is required for late gene expression in Kaposi's sarcoma-associated herpesvirus. J. Virol. *93*.

Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M., and Moore, P.S. (1994). Identification of Herpesvirus-Like DNA Sequences in AIDS-Associated Kaposi's Sarcoma. Science. *266*, 1865–1869.

Chapa, T.J., Du, Y., Sun, R., Yu, D., and French, A.R. (2017). Proteomic and phylogenetic coevolution analyses of pM79 and pM92 identify interactions with RNA polymerase II and delineate the murine cytomegalovirus late transcription complex. J. Gen. Virol. *98*, 242–250.

Chen, H.-T., and Hahn, S. (2003). Binding of TFIIB to RNA Polymerase II: Mapping the Binding Site for the TFIIB Zinc Ribbon Domain within the Preinitiation Complex. Mol. Cell *12*, 437–447.

Chi, T., and Carey, M. (1996). Assembly of the isomerized TFIIA-TFIID-TATA ternary complex is necessary and sufficient for gene activation. Genes Dev. *10*, 2540–2550.

Covarrubias, S., Gaglia, M.M., Kumar, G.R., Wong, W., Jackson, A.O., and Glaunsinger, B.A. (2011). Coordinated destruction of cellular messages in translation complexes by the gammaherpesvirus host shutoff factor and the Mammalian Exonuclease Xrn1. PLoS Pathog. *7*.

Dai-Ju, J.Q., Li, L., Johnson, L.A., and Sandri-Goldin, R.M. (2006). ICP27 Interacts with the C-Terminal Domain of RNA Polymerase II and Facilitates Its Recruitment to Herpes Simplex Virus 1 Transcription Sites, Where It Undergoes Proteasomal Degradation during Infection. J. Virol. *80*, 3567–3581.

Damania, B., and Alwine, J.C. (1996). TAF-like function of SV40 large T antigen. Genes Dev. *10*, 1369–1381.

Damania, B., Lieberman, P., and Alwine, J.C. (1998). Simian Virus 40 Large T Antigen Stabilizes the TATA-Binding Protein–TFIIA Complex on the TATA Element. Mol. Cell. Biol. *18*, 3926–3935.

Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., VonDollen, J., Maher, M.C., Johnson, T., Newton, W., et al. (2015). Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. Mol. Cell *57*, 349–360.

Davis, Z.H., Hesser, C.R., Park, J., and Glaunsinger, A. (2016). Interaction between ORF24 and ORF34 in the Kaposi's Sarcoma- Associated Herpesvirus Late Gene Transcription Factor Complex Is Essential for Viral Late Gene Expression. J. Virol. *90*, 599–604.

Davison, A.J. (2010). Herpesvirus systematics. Vet. Microbiol. *143*, 52–69.

Decker, K.B., and Hinton, D.M. (2013). Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. Annu. Rev. Microbiol. *67*, 113–139.

DeLuca, N.A., and Schaffer, P.A. (1988). Physical and functional domains of the herpes simplex virus transcriptional regulatory protein ICP4. J. Virol. *62*, 732–743.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Dill, K.A. (1985). Theory for the Folding and Stability of Globular Proteins. Biochemistry *24*, 1501–1509.

Dixit, P.D., and Maslov, S. (2013). Evolutionary Capacitance and Control of Protein Stability in Protein-Protein Interaction Networks. PLoS Comput. Biol. *9*.

Djavadian, R., Chiu, Y.F., and Johannsen, E. (2016). An Epstein-Barr Virus-Encoded Protein Complex Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. PLoS Pathog. *12*, 1–25.

Djavadian, R., Hayes, M., and Johannsen, E. (2018). CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. PLoS Pathog. *14*, 1–26.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

El-Guindy, A., Lopez-Giraldez, F., Delecluse, H.J., McKenzie, J., and Miller, G. (2014). A Locus Encompassing the Epstein-Barr Virus bglf4 Kinase Regulates Expression of Genes Encoding Viral Structural Proteins. PLoS Pathog. *10*, e1004307.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell *11*, 1549–1561.

Ganem, D. (2006). KSHV infection and the pathogenesis of Kaposi's sarcoma. Annu. Rev. Pathol. *1*, 273–296.

Gardner, M.R., and Glaunsinger, B.A. (2018). Kaposi's Sarcoma-Associated Herpesvirus ORF68 Is a DNA Binding Protein Required for Viral Genome Cleavage and Packaging. J. Virol. *92*, 1–13.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics *21*, 1295–1300.

Glaunsinger, B., and Ganem, D. (2004). Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. Mol. Cell *13*, 713–723.

Gong, D., Wu, N.C., Xie, Y., Feng, J., Tong, L., Brulois, K.F., Luan, H., Du, Y., Jung, J.U., Wang, C. -y., et al. (2014). Kaposi's Sarcoma-Associated Herpesvirus ORF18 and ORF30 Are Essential for Late Gene Expression during Lytic Replication. J. Virol. *88*, 11369–11382.

Grassmann, K., Rapp, B., Maschek, H., Petry, K.U., and Iftner, T. (1996). Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of

episomal HPV-16 DNA. J. Virol. *70*, 2339–2349.

Greber, B.J., Toso, D.B., Fang, J., and Nogales, E. (2019). The complete structure of the human TFIIH core complex. Elife *8*, 1–29.

Grondin, B., and DeLuca, N. (2000). Herpes Simplex Virus Type 1 ICP4 Promotes Transcription Preinitiation Complex Formation by Enhancing the Binding of TFIID to DNA. J. Virol. *74*, 11504–11510.

Gruffat, H., Kadjouf, F., Mariame, B., and Manet, E. (2012). The Epstein-Barr Virus BcRF1 Gene Product Is a TBP-Like Protein with an Essential Role in Late Gene Expression. J. Virol. *86*, 6023–6032.

Gruffat, H., Marchione, R., and Manet, E. (2016). Herpesvirus late gene expression: A viral-specific pre-initiation complex is key. Front. Microbiol. *7*, 1–15.

Guzowski, J.F., and Wagner, E.K. (1993). Mutational analysis of the herpes simplex virus type 1 strict late UL38 promoter/leader reveals two regions critical in transcriptional regulation. J. Virol. *67*, 5098–5108.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. 1.

Hahn, S., Yudkovsky, N., and Ranish, J.A. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225–229.

Haigh, A., Greaves, R., and O'Hare, P. (1990). Interference with the assembly of a virus-host transcription complex by peptide competition. Nature *344*, 257–259.

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

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.

Harper, T.M., and Taatjes, D.J. (2018). The complex structure and function of Mediator. J. Biol. Chem. *293*, 13778–13785.

He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481–486.

Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Heming, J.D., Conway, J.F., and Homa, F.L. (2017). Herpesvirus Capsid Assembly and DNA Packaging. In Cell Biology of Herpes Viruses, pp. 119–142.

Holstege, F.C.P., van der Vliet, P.C., and Timmers, H.T.M. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J. *15*, 1666–1677.

Hong, P., Koza, S., and Bouvier, E.S.P. (2012). A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. J. Liq. Chromatogr. Relat. Technol. *35*, 2923–2950.

Huang, C.J., and Wagner, E.K. (1994). The herpes simplex virus type 1 major capsid protein (VP5-UL19) promoter contains two cis-acting elements influencing late expression. J. Virol. *68*, 5738–5747.

Jia, Q., Wu, T.-T., Liao, H.-I., Chernishof, V., and Sun, R. (2004). Murine gammaherpesvirus 68 open reading frame 31 is required for viral replication. J. Virol. *78*, 6610–6620.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. Mol. Cell *45*, 51–63.

Johnson, K.M., and Carey, M. (2003). Assembly of a Mediator/TFIID/TFIIA Complex Bypasses the Need for an Activator. Curr. Biol. *13*, 772–777.

Johnson, P.A., and Everett, R.D. (1986). The control of herpes simplex virus type-1 late gene transcription: a 'TATA-box'/cap site region is sufficient for fully efficient regulated activity. *14*, 8247–8264.

Juven-Gershon, T., Cheng, S., and Kadonaga, J.T. (2006). Rational design of a super core promoter that enhances gene expression. Nat. Methods *3*, 917–922.

Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes Dev. *23*, 439–451.

Kumar, B., and Chandran, B. (2016). KSHV entry and trafficking in target cells—Hijacking of cell signal pathways,actin and membrane dynamics. Viruses *8*.

Labo, N., Miley, W., Marshall, V., Gillette, W., Esposito, D., Bess, M., Turano, A., Uldrick, T., Polizzotto, M.N., Wyvill, K.M., et al. (2014). Heterogeneity and Breadth of Host Antibody Response to KSHV Infection Demonstrated by Systematic Analysis of the KSHV Proteome. PLoS Pathog. *10*.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., et al. (2009). Appion: An integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. *166*, 95–102.

Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. Elife *2*, 1–20.

Lester, J.T., and DeLuca, N.A. (2011). Herpes Simplex Virus 1 ICP4 Forms Complexes with TFIID and Mediator in Virus-Infected Cells. J. Virol. *85*, 5733–5744.

Li, D., Fu, W., and Swaminathan, S. (2018). Continuous DNA replication is required for late gene transcription and maintenance of replication compartments in gammaherpesviruses. PLoS Pathog. *14*, 1–25.

Lin, C.L., Li, H., Wang, Y., Zhu, F.X., Kudchodkar, S., and Yuan, Y. (2003). Kaposi's Sarcoma-Associated Herpesvirus Lytic Origin (ori-Lyt)-Dependent DNA Replication: Identification of the ori-Lyt and Association of K8 bZip Protein with the Origin. J. Virol. *77*, 5578–5588.

Lukac, D.M., Renne, R., Kirshner, J.R., and Ganem, D. (1998). Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. Virology *252*, 304–312.

Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nat. Rev. Mol. Cell Biol. *13*, 543–547.

Lynch, K.J.J., and Frisque, R.J.J. (1990). Identification of critical elements within the JC virus DNA replication origin. J.Virol. *64*, 5812–5822.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

Mallick, S.P., Carragher, B., Potter, C.S., and Kriegman, D.J. (2005). ACE: Automated CTF estimation. Ultramicroscopy *104*, 8–29.

Del Mar Pena, L., and Laimins, L.A. (2002). Differentiation-Dependent Chromatin Rearrangement Coincides with Activation of Human Papillomavirus Type 31 Late Gene Expression. J. Virol. *75*, 10005–10013.

Marblestone, J.G., Edavettal, S.C., Lim, T., Lim, P., Zuo, X., and Butt, Tauseef, R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. Protein Sci. *15*, 182–189.

McKenzie, J., Lopez-Giraldez, F., Delecluse, H.J., Walsh, A., and El-Guindy, A. (2016). The Epstein-Barr Virus Immunoevasins BCRF1 and BPLF1 Are Expressed by a

Mechanism Independent of the Canonical Late Pre-initiation Complex. PLoS Pathog. *12*, 1–32.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223–226.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev. *19*, 1401–1415.

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

Myers, L.C., and Kornberg, R.D. (2000). Mediator of Transcriptional Regulation. Annu. Rev. Biochem. *69*, 729–749.

Myers, R.M., Rio, D.C., Robbins, A.K., and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell *25*, 373–384.

Näär, A.M., Taatjes, D.J., Zhai, W., Nogales, E., and Tjian, R. (2002). Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. Genes Dev. *16*, 1339–1344.

Nandakumar, D., and Glaunsinger, B.A. (2019). An integrative approach identifies direct targets of the late viral transcription complex and an expanded promoter recognition motif in Kaposi's sarcoma-associated herpesvirus. BioRxiv 550772.

Nishimura, M., Watanabe, T., Yagi, S., Yamanaka, T., and Fujimuro, M. (2017). Kaposi's sarcoma-associated herpesvirus ORF34 is essential for late gene expression and virus production. Sci. Rep. *7*, 1–12.

Nonet, M.L., and Young, R.A. (1989). Intragenic and Extragenic Suppressors of Mutations in the Heptapeptide Repeat Domain of Saccharomyces cerevisiae RNA Polymerase II. Genetics *123*, 715–724.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription machinery of RNA polymerase II. Genes Dev. *10*, 2657–2683.

Pan, D., Han, T., Tang, S., Xu, W., Bao, Q., Sun, Y., Xuan, B., and Qian, Z. (2018). Murine Cytomegalovirus Protein pM91 Interacts with pM79 and Is Critical for Viral Late Gene Expression. J. Virol. 92:e00675-18.

Panagiotidis, C.A., Lium, E.K., and Silverstein, S.J. (1997). Physical and functional interactions between herpes simplex virus immediate-early proteins ICP4 and ICP27. J. Virol. *71*, 1547–1557.

Perng, Y.-C., Qian, Z., Fehr, A.R., Xuan, B., and Yu, D. (2011). The Human

Cytomegalovirus Gene UL79 Is Required for the Accumulation of Late Viral Transcripts. J. Virol. *85*, 4841–4852.

Perng, Y.-C., Campbell, J.A., Lenschow, D.J., and Yu, D. (2014). Human cytomegalovirus pUL79 is an elongation factor of RNA polymerase II for viral gene transcription. PLoS Pathog. *10*, e1004350.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376–380.

Revyakin, A., Zhang, Z., Coleman, R.A., Li, Y., Inouye, C., Lucas, J.K., Park, S.R., Chu, S., and Tjian, R. (2012). Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. Genes Dev. *26*, 1691–1702.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422.e16.

Robinson, P.J.J., Bushnell, D.A., Trnka, M.J., Burlingame, A.L., and Kornberg, R.D. (2012). Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. Proc. Natl. Acad. Sci. *109*, 17931–17935.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.

Sandaltzopoulos, R., and Becker, P.B. (1998). Heat Shock Factor Increases the Reinitiation Rate from Potentiated Chromatin Templates. Mol. Cell. Biol. *18*, 361–367.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schmid, M., Speiseder, T., Dobner, T., and Gonzalez, R.A. (2014). DNA Virus Replication Compartments. J. Virol. *88*, 1404–1420.

Serio, T.R., Cahill, N., Prout, M.E., and Miller, G. (1998). A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. J. Virol. *72*, 8338–8343.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Shrestha, P., and Sugden, B. (2014). Identification of Properties of the Kaposi's Sarcoma-Associated Herpesvirus Latent Origin of Replication That Are Essential for the Efficient Establishment and Maintenance of Intact Plasmids. J. Virol. *88*, 8490–8503.

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

Song, M.J., Hwang, S., Wong, W.H., Wu, T.-T., Lee, S., Liao, H.-I., and Sun, R. (2005). Identification of viral genes essential for replication of murine gammaherpesvirus 68 using signature-tagged mutagenesis. Proc. Natl. Acad. Sci. *102*, 3805–3810.

Sorzano, C.O.S., Marabini, R., Velázquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H.W., Carazo, J.M., and Pascual-Montano, A. (2004). XMIPP: A new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. *148*, 194–204.

Spahr, H., Calero, G., Bushnell, D.A., and Kornberg, R.D. (2009). Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. Proc. Natl. Acad. Sci. *106*, 9185–9190.

Sun, R., Lin, S.F., Gradoville, L., Yuan, Y., Zhu, F., and Miller, G. (1998). A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. Proc. Natl. Acad. Sci. U. S. A. *95*, 10866–10871.

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

Tang, S., Yamanegi, K., and Zheng, Z.-M. (2004). Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. J. Virol. *78*, 2609–2614.

Tempera, I., and Lieberman, P.M. (2010). Chromatin organization of gammaherpesvirus latent genomes. Biochim. Biophys. Acta - Gene Regul. Mech. *1799*, 236–245.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell *73*, 1361–1375.

Tsai, F.T.F., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human Pol II promoters. EMBO J. *19*, 25–36.

Uppal, T., Banerjee, S., Sun, Z., Verma, S.C., and Robertson, E.S. (2014). KSHV LANA—The Master Regulator of KSHV Latency. Viruses *6*, 4691–4998.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat. Struct. Biol.

7, 639–643.

Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., and Carragher, B. (2009). DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. *166*, 205–213.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. Cell *127*, 941–954.

White, S.H. (1992). Amino acid preferences of small proteins. J. Mol. Biol. *227*, 991–995.

Wong-Ho, E., Wu, T.T., Davis, Z.H., Zhang, B.Q., Huang, J., Gong, H., Deng, H.Y., Liu, F.Y., Glaunsinger, B., and Sun, R. (2014). Unconventional Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus 68. J. Virol. *88*, 3411–3422.

Wong, E., Wu, T.-T., Reyes, N., Deng, H., and Sun, R. (2007). Murine gammaherpesvirus 68 open reading frame 24 is required for late gene expression after DNA replication. J. Virol. *81*, 6761–6764.

Wu, T.-T., Park, T., Kim, H., Tran, T., Tong, L., Martinez-Guzman, D., Reyes, N., Deng, H., and Sun, R. (2009). ORF30 and ORF34 are essential for expression of late genes in murine gammaherpesvirus 68. J. Virol. *83*, 2265–2273.

Wyrwicz, L.S., and Rychlewski, L. (2007). Identification of Herpes TATT-binding protein. Antiviral Res. *75*, 167–172.

Young, R.A. (1991). RNA Polymerase II. Annu. Rev. Biochem. *60*, 689–715.

Zabierowski, S., and DeLuca, N.A. (2004). Differential Cellular Requirements for Activation of Herpes Simplex Virus Type 1 Early (tk) and Late (gC) Promoters by ICP4. J. Virol. *78*, 6162–6170.