The gammaherpesviral TATA-box-binding protein directly interacts with the CTD of host RNA Pol II to direct late gene transcription

Angelica F. Castañeda[a,1], Allison L. Didychuk[a,1], Robert K. Louder[b,c,2], Chloe O. McCollum[d], Zoe H. Davis[e], Eva Nogales[b,d,f,g], Britt A. Glaunsinger[a,d,f,g,3]

[a]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [b]Molecular Biophysics and Integrative Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Biophysics Graduate Group, University of California, Berkeley, CA 94720; [d]Department of Molecular and Cell Biology, University of California Berkeley, CA 94720; [e]Division of Infectious Diseases and Immunity, School of Public Health, University of California, Berkeley, CA 94720; [f]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720; [g]Howard Hughes Medical Institute, Berkeley, CA 94720

[1]These authors contributed equally to this work

[2]Present address: Department of Biology, Johns Hopkins University, Baltimore, MD 21218

[3]To whom correspondence may be addressed. **Email:** glaunsinger@berkeley.edu

Classification

Biological Sciences (Microbiology)

Keywords

Herpesvirus, KSHV, transcription, RNA polymerase II, late gene

Author Contributions

A.F.C, A.L.D., R.K.L., Z.H.D., E.N., and B.A.G designed research; A.F.C., A.L.D., R.K.L., and C.O.M. performed research; A.F.C., A.L.D., R.K.L., C.O.M., Z.H.D, E.N., and B.A.G. analyzed data; and A.F.C, A.L.D., R.K.L., C.O.M., Z.H.D., E.N., and B.A.G. wrote the paper.

This PDF file includes:

Main Text

Figures 1 to 5

ABSTRACT

β- and γ-herpesviruses include the oncogenic human viruses Kaposi's sarcoma-associated virus (KSHV) and Epstein-Barr virus (EBV), and human cytomegalovirus (HCMV), which is a significant cause of congenital disease. Near the end of their replication cycle, these viruses transcribe their late genes in a manner distinct from host transcription. Late gene transcription requires six virally-encoded proteins, one of which is a functional mimic of host TATA-box-binding protein (TBP) that is also involved in recruitment of RNA polymerase II (Pol II) via unknown mechanisms. Here, we applied biochemical protein interaction studies together with electron microscopy-based imaging of a reconstituted human preinitiation complex to define the mechanism underlying Pol II recruitment. These data revealed that the herpesviral TBP, encoded by ORF24 in KSHV, makes a direct protein-protein contact with the C-terminal domain of host RNA polymerase II (Pol II), which is a unique feature that functionally distinguishes viral from cellular TBP. The interaction is mediated by the N-terminal domain (NTD) of ORF24 through a conserved motif that is shared in its β- and γ-herpesvirus homologs. Thus, these herpesviruses employ an unprecedented strategy in eukaryotic transcription, wherein promoter recognition and polymerase recruitment are facilitated by a single transcriptional activator with functionally distinct domains.

SIGNIFICANCE STATEMENT

The β- and γ-herpesviruses mediate their late gene transcription through a set of viral transcriptional activators (vTAs). One of these vTAs, ORF24 in Kaposi's sarcoma-associated herpesvirus (KSHV), is a mimic of host TATA-box-binding protein (TBP). We demonstrate that the N-terminal domain of ORF24 and its homologs from other β- and γ-herpesviruses directly bind the unstructured C-terminal domain (CTD) of RNA Pol II. This functionally distinguishes the viral TBP mimic from cellular TBP, which does not bind Pol II. Thus, herpesviruses encode a transcription factor that has the dual ability to directly interact with promoter DNA and the

polymerase, a property which is unique in eukaryotic transcription and is conceptually akin to prokaryotic transcription factors.

MAIN TEXT

INTRODUCTION

Eukaryotic transcription begins with the formation of a pre-initiation complex (PIC) at the core promoter, starting with binding of TFIID and deployment of TATA-box-binding protein (TBP) onto the TATA box or pseudo-TATA box region upstream of the transcription start site (TSS). This is followed by recruitment of the other general transcription factors (GTFs), which recruit and position the 12-subunit RNA polymerase II (Pol II) at the core promoter (1). The largest Pol II subunit, Rpb1, has a low-complexity carboxyl terminal domain (CTD) that in humans is composed of 52 heptapeptide repeats with a consensus sequence of YSPTSPS. The CTD is a regulatory hub responsible for coordinating signals throughout the different stages of transcription and RNA processing (2). The phosphorylation state of the CTD controls progression through different states of transcription, as well as interactions with other cellular machinery (2, 3). Pol II with a hypophosphorylated CTD is recruited into the PIC (4), and phosphorylation signals release from the PIC into an elongating complex.

DNA viruses hijack the host transcriptional machinery to direct their own gene expression. Given that the mechanisms governing transcription from viral promoters are often similar to those at host promoters, viruses have been invaluable models for understanding transcription complex assembly and regulation (5-7). A conserved feature of double-stranded DNA (dsDNA) viruses is the temporal cascade of gene expression that begins with the expression of two classes of early genes, followed by viral DNA replication, and ending with the expression of late genes. In the β- and γ-herpesviruses, immediate early and early genes are transcribed in a manner similar to host genes. In contrast, the mechanism underlying the

regulation of late gene transcription remains poorly understood, yet is known to be distinct from host and early viral gene transcription.

Late gene transcription in the β- and γ-herpesviruses is divergent from that of the α-herpesviruses and has a number of unique features. First, β/γ late gene transcription is regulated in part by a core promoter sequence 12-15 base pairs in length that has a TATT motif followed by a RVNYS motif in lieu of the canonical TATA box found in early viral promoters and in some cellular promoters (8-11). Additionally, late gene expression requires at least six viral proteins, called viral transcriptional activators (vTAs), which form a complex at late gene promoters (12-17). Little is known about the functional role these vTAs play in late gene transcription.

The best studied protein in the vTA complex is a virally-encoded TBP mimic (vTBP), encoded by ORF24 in the γ-herpesvirus Kaposi's sarcoma-associated herpesvirus (KSHV). ORF24 is predicted to have a TBP-like domain in the central portion of the protein, which was identified through an *in silico* protein fold threading analysis performed with BcRF1, the homolog of ORF24 from Epstein-Barr virus (18). Indeed, ORF24 replaces TBP at late gene promoters during infection, and a virus with mutations in the predicted DNA-binding residues in the TBP-like domain of ORF24 is unable to transcribe late genes (19). Thus, β- and γ-herpesviruses encode their own vTBP, which promotes efficient transcription from a distinct set of late gene promoters.

While both cellular TBP and ORF24 (vTBP) bind DNA, a unique feature of vTBP is that co-immunoprecipitation experiments revealed it additionally interacts with Pol II in cells (12, 19). In host transcription, recruitment of Pol II to promoters is mediated by TFIIB instead of a direct protein-protein interaction between TBP and Pol II (20, 21). While other viral proteins have been shown to directly or indirectly bind the Pol II CTD (22-24), none are thought to also bind promoter DNA. In contrast, prokaryotic transcription is dependent on sigma factors that facilitate both promoter selection and polymerase recruitment (25). We were therefore intrigued by the

possibility that vTBP is a unique bifunctional eukaryotic transcriptional activator and sought to understand the basis of its ability to recruit Pol II.

Here, we demonstrate that interaction with Pol II is conserved across the ORF24 β- and γ-herpesvirus homologs and is mediated through a conserved motif in the N-terminus. We identified a minimal N-terminal domain (NTD) that is sufficient for interaction with Pol II. Using recombinantly expressed protein, we demonstrate that ORF24 makes a direct protein-protein contact with the CTD of Pol II. Using pulldown assays and an *in vitro* reconstituted PIC assembly assay coupled with negative stain electron microscopy, we determined that ORF24-NTD directly interacts with the Pol II CTD, and demonstrate that at least four heptapeptide CTD repeats are required for this interaction. These findings suggest that vTBP is a fundamentally unique protein when compared to other eukaryotic Pol II-interacting proteins, as it both directly interacts with the Pol II CTD and binds promotor DNA to coordinate late gene expression.

RESULTS

A leucine-rich motif is necessary for interaction with RNA polymerase II across β- and γ-herpesviruses

We previously revealed that KSHV ORF24 co-immunoprecipitates with Pol II in cells, in a manner dependent on three conserved leucine residues (L73-75; the RLLLG motif) in the N-terminus of ORF24 (19). β- and γ-herpesviral homologs of ORF24 from murine gammaherpesvirus 68 (MHV68; mu24), Epstein-Barr virus (EBV; BcRF1), and human cytomegalovirus (HCMV; UL87) have also been reported to interact with Pol II in cells (12, 19, 26). While this interaction requires no other viral proteins, how it is orchestrated remains a central open question. For example, it is unknown whether Pol II binding occurs through an ORF24 domain separable from the region required for binding the ORF34 vTA (which links ORF24 to the rest of the late gene transcription complex (14)) and/or the region required for binding to promoter DNA. Furthermore, it is not known whether ORF24 binds Pol II directly or

138 indirectly via bridging cellular factors - as is the case for all other eukaryotic promoter DNA
139 binding transcription factors.

140 The RLLLG motif is well-conserved in all β- and γ-herpesvirus homologs, despite overall
141 poor sequence identity (**Supplemental Figure S1**), suggesting that this N-terminal region of
142 ORF24 homologs may also be necessary for Pol II recruitment. To test if the three conserved
143 leucine residues are involved in the homolog-Pol II interactions, we generated full-length wild-
144 type or triple leucine mutants (3L_A) of UL87, mu24, and BcRF1 with C-terminal Strep tags,
145 transiently transfected plasmids encoding these constructs into HEK293T cells, and affinity
146 purified using StrepTactinXT beads. Similarly to ORF24, mutation of the RLLLG motif ablated
147 the interactions of all homologs with Pol II (**Figure 1**), suggesting that vTBPs interact with Pol II
148 through their respective N-terminal domains in a manner dependent upon this highly conserved
149 patch of residues.

150

151 **The N-terminal domain of ORF24 is sufficient for interaction with Pol II**

152 To next test if the N-terminal domain is sufficient for the Pol II interaction, we constructed
153 a series of N-terminal ORF24 fragments to identify a minimal region of the protein that retained
154 the Pol II interaction (**Figure 2A**). Plasmids encoding full-length or truncated FLAG-tagged
155 ORF24 were transfected into HEK293T cells and immunoprecipitated using FLAG antibodies to
156 determine which ORF24 segments retained binding to endogenous Pol II. The smallest
157 fragment of ORF24 that retained the Pol II interaction consisted of amino acids (a.a.) 1-201,
158 which we termed the ORF24 N-terminal domain (ORF24-NTD) (**Figure 2B**). Notably, the
159 remainder of the protein (a.a. 202-752), which includes the vTBP domain and a region known to
160 interact with the vTA ORF34 (14), failed to co-immunoprecipitate Pol II. Thus, ORF24 contains
161 an NTD that is both necessary and sufficient for Pol II binding, which is separable from the other
162 two known functions of the protein.

163

Identification of a minimal Pol II-interaction domain in the β- and γ-herpesviruses

We next sought to determine if the NTD of homologs of ORF24 was similarly sufficient for Pol II interaction, with the goal of identifying whether the β- and γ-herpesviruses share a common domain for polymerase recruitment. Based on the boundary of the ORF24-NTD identified in Figure 2 and sequence alignments (**Supplemental Figure S1**), we designed constructs for mu24, BcRF1, and UL87 that encompassed either an analogous NTD, or versions of all the homologs where this domain is truncated by ten a.a. or extended by 25 a.a.. Based on isoelectric point (pI) calculations (27), the full-length proteins are predicted to be electropositive, but the N-terminal protein-protein interaction domains are predicted to be electronegative (**Supplemental Figure S1**). The BcRF1 NTD appears to be an outlier in terms of overall predicted pI, which may change its physical properties when taken out of context of the full-length protein.

As expected, Strep-tagged ORF24 a.a. 1-201 (ORF24-NTD) or the extended domain containing a.a. 1-226 both interacted with Pol II in Strep affinity purifications from whole cell lysate (**Figure 3A**). Notably, reducing the ORF24-NTD by even ten a.a. eliminated its ability to interact with Pol II, suggesting that ORF24-NTD (a.a. 1-201) is, or is nearly, the minimal domain for the Pol II interaction (**Figure 3A**). The mu24 and UL87 proteins showed a similar pattern, in which the predicted domain equivalent to ORF24-NTD was sufficient for Pol II interaction, but further truncation by 10 a.a. eliminated the interaction (**Figure 3B, D**). In contrast, all truncations of BcRF1 failed to interact with Pol II (**Figure 3C**).

The above findings indicated that at least 3 of the vTBP homologs share a common N-terminal domain that, despite substantial sequence divergence, is necessary and sufficient for Pol II binding but distinct from other known functional regions of the proteins. Previous data demonstrated that the full-length vTBP from one virus cannot complement homologs in other herpesviruses (12). However, we considered the possibility that the specific Pol II recruitment domain might be functionally interchangeable between these vTBP homologs if the primary role

of this domain is to bring Pol II to late promoters. To test this, we generated chimeras of ORF24 wherein the ORF24-NTD (a.a. 1-201) was replaced by the minimal NTD of its homologs (**Supplemental Figure S2A**). These chimeras retain the region of ORF24 that we previously identified as important for the ORF24-ORF34 interaction (14) as well as the ORF24 vTBP domain (18) and C-terminal domain. We noted that the full-length homologs of ORF24 do not interact with KSHV ORF34, despite conservation of an arginine (ORF24 R328) essential for the ORF24-ORF34 interaction (14) (**Supplemental Figure S2B**). However, each of the NTD chimeras interacted with both Pol II and KSHV ORF34, suggesting that the minimal NTD is sufficiently well-folded when fused to ORF24 a.a. 202-752 (**Supplemental Figure S2C**). Interestingly, although the minimal domain of BcRF1 alone failed to interact with Pol II by co-IP (**Figure 3D**), it is capable of interaction with Pol II when fused to ORF24 (**Supplemental Figure S2C**), suggesting that the BcRF1 minimal domain is sufficient for interaction with Pol II, but may have properties not compatible with Pol II interaction when truncated in an *in vitro* setting.

We assessed the ability of the NTD chimeras to functionally complement ORF24 using an established transfection-based late gene transcription assay (13-15). The six vTAs (ORFs 18, 30, 31, 34, 66 and either WT ORF24, its homologs, or chimeras) were co-transfected into HEK293T cells, along with a firefly luciferase reporter controlled by the K8.1 late gene promoter or, as a control, the early ORF57 promoter. A Renilla luciferase reporter was also included to control for transfection efficiency. Consistent with previous observations (12), none of the full-length homologs could functionally complement ORF24 to activate the K8.1 promoter (**Supplemental Figure S3D**). However, the mu24-ORF24 chimera promoted transcription to levels ~40% that of wild-type ORF24. Interestingly, neither the BcRF1-ORF24 or UL87-ORF24 chimeras were functional for late gene transcription, despite the fact that they could interact with both ORF34 and Pol II **(Supplemental Figure S3D**). This may suggest that the N-terminal domain has additional functions or interactions beyond polymerase recruitment that mu24 is able to maintain due to sequence similarity, or it may simply be that the mu24 fusion (but not the

BcRF1 or UL87 fusions) is positioned relative to the promoter and other vTAs in a manner similar enough to ORF24 to be functional. Thus, although the minimal NTD of all vTBPs interacts with Pol II, other contacts or functions may be necessary to successfully promote late gene transcription.

The minimal Pol II interaction domain identified here varies greatly in length (191 a.a. in mu24 vs. 248 a.a. in UL87) and in sequence, as no residues are conserved in both β- and γ-herpesviruses other than the RLLLG motif (**Supplemental Figure S1**). Despite this significant variation, ORF24 and its homologs have evolved a shared mechanism for Pol II recruitment that is primarily mediated by their respective N-terminal domains.

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

We next sought to determine how ORF24-NTD interacts with Pol II within a minimal PIC. To generate purified minimal ORF24-NTD for this purpose, we appended it to an N-terminal glutathione-S-transferase (GST) tag, and achieved robust expression of the protein in *E. coli*, similar to that of GST alone (**Figure 4A**). We confirmed that GST-tagged ORF24-NTD retained the ability to interact with Pol II using a GST pulldown with GST-ORF24-NTD and whole cell lysate from HEK293T cells, suggesting that recombinantly expressed GST-ORF24-NTD is well-folded (**Figure 4B**).

To assemble a minimal PIC that includes ORF24-NTD, we used a streptavidin-immobilized DNA scaffold based on the super core promoter element (SCP) (28), which contains the TATA, BRE, and INR core promoter elements (**Figure 4C**). Human TBP, TFIIA, TFIIB, and TFIIF were purified from *E. coli*, and Pol II was immunopurified from HeLa cell extracts (29). We expected that TBP binding to the TATA motif would initiate formation of the PIC with binding of TFIIA and TFIIB, followed by Pol II and TFIIF recruitment, and then binding of GST-ORF24-NTD. Therefore, any additional density beyond that seen in the well-

characterized minimal PIC (DNA/TBP/TFIIA/TFIIB/Pol II/TFIIF) (28) could be attributed to ORF24-NTD.

The minimal PIC was assembled by sequentially incubating the DNA scaffold with TBP, TFIIA and TFIIB, Pol II and TFIIF, and GST-ORF24-NTD, followed by immobilization on streptavidin-coated beads, and finally washing and elution from the beads (**Figure 4C**). We performed single particle negative stain EM of this assembled complex containing GST-ORF24-NTD, and searched for particles that displayed extra density beyond the well-defined minimal PIC structure. A total of 79,381 PIC particle images were analyzed by two-dimensional and three-dimensional classification, and approximately 25% showed clear extra density in the resulting class averages (**Figure 4D** and **Supplemental Figure S3**). Two of the three-dimensional classes (called Class 1 and 2) exhibit extra density on opposing faces of the Pol II stalk (**Figure 4D**), while a third class (Class 5) is comprised of the majority of particles without extra density near the Pol II stalk. Thus, Class 5 was used to subtract the density corresponding to the minimal PIC from Classes 1 and 2, resulting in difference maps that clearly show the region occupied by bound GST-ORF24-NTD (**Figure 4F**).

By superimposing the three-dimensional difference maps representing the extra density attributed to GST-ORF24-NTD onto the cryo-EM structure of the human minimal PIC (**Figure 4G**) (30), it became evident that the ORF24-NTD is flexibly bound to a region of Pol II near the Rpb4/7 stalk module. An inherently flexible domain of Pol II is the CTD of the largest subunit, Rpb1, which contains a linker followed by 52 heptapeptide repeats (2) that is not visualized by EM due to its disordered structure. Docking of the crystal structure of Rpb1 from *Schizosaccharomyces pombe* (PDB 3H0G) (31), in which the structure of the CTD linker region is partially resolved and can be seen to extend along the surface of the stalk module, further indicates that the ORF24-NTD may be interacting with the flexible CTD of Pol II (**Figure 4H**). Together, these observations suggest that the NTD of the ORF24 vTBP interacts directly with

one or more Pol II subunits, which would be a unique feature distinct from other characterized eukaryotic viral or cellular transcription factors.

ORF24-NTD binds the CTD repeats of Rpb1

Based on the EM results, we hypothesized that ORF24 could be interacting with either Rpb4/7, the Rpb1 linker, or the Rpb1 CTD heptapeptide repeats. We therefore assessed whether recombinantly purified versions of each of these factors bound purified ORF24-NTD. We first performed GST-pulldown assays with GST-tagged Rpb1 CTD or Rpb1 linker 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-3L_A). This mutation renders the protein unable to interact with Pol II in cells (19). Notably, WT ORF24-NTD bound to the Rpb1 CTD repeats, but not to the Rpb1 linker region (**Figure 5A**). This interaction was specific and mimicked the results in mammalian cell lysate, as no binding to either CTD fusion was observed with ORF24-NTD-3L_A (**Figure 5A**). We also performed Strep pulldown assays using ORF24-NTD-Strep and the heterodimeric GST-Rpb4/Rpb7-His complex. Again, ORF24-NTD interacted with GST-CTD, but not with GST-Rpb4/Rpb7-His (**Figure 5B**). Thus, recombinant ORF24-NTD directly interacts with recombinant Pol II CTD repeats *in vitro*, and enrichment using tags on either the Pol II CTD or on ORF24 allows for the isolation of the other. The lack of interaction with the CTD linker domain or the Rpb4/7 stalk suggests that the CTD repeats are likely to be the primary point of Pol II contact with ORF24.

The CTD is a key regulatory component of Pol II, responsible for coordinating signals through interactions with multiple transcriptional modulators (2). CTD binding proteins engage with the heptapeptide repeats in a variety of ways, from recognition of a few residues of a repeat (in the case of kinases that promote phosphorylation at conserved serines, threonines, or tyrosines within a given repeat), to multiple consecutive repeats (in the case of Mediator (32)),

to recognition of the intrinsically disordered entire domain (in the case of newly-appreciated phase-separated interactions (33, 34)). To determine which of these types of interactions occur between the CTD and ORF24, we performed Strep pulldown assays with purified ORF24-NTD-Strep and recombinant GST-CTD constructs containing either 2, 4, 10, or 52 (full length) heptapeptide repeats. The recombinant GST-CTD used in these assays was unphosphorylated, as our previous results demonstrated that ORF24 co-immunoprecipitates only hypophosphorylated Pol II from mammalian cells (19). Notably, ORF24 interacted with 4x, 10x, and 52xCTD, but failed to interact with either 2xCTD repeats or with control GST (**Figure 5C**). Therefore, similar to the Mediator complex (32), ORF24 likely engages four tandem repeats to interact with the Pol II CTD.

DISCUSSION

Late gene transcription in the β- and γ-herpesviruses depends on a set of viral transcriptional activators, including a virally-encoded mimic of host TBP. Here, we demonstrate that the N-terminal domain (NTD) of the ORF24 vTBP from KSHV recruits Pol II through a direct protein-protein interaction with four heptapeptide repeats of the Pol II C-terminal domain (CTD). Conserved residues in the ORF24-NTD are required for this interaction, suggesting that vTBPs in β- and γ-herpesviruses have evolved a shared strategy for recruitment of Pol II to late gene promoters. Domain swapping experiments between homologs suggest that the primary function of the NTD of vTBP in β- and γ-herpesviruses is to recruit Pol II to the late gene promoter and that the NTD is functionally distinct from the remainder of the protein. Our work conclusively demonstrates that vTBP is unique among eukaryotic transcriptional activators in its ability to simultaneously bind promoter DNA and Pol II, and that this strategy is conserved across the β- and γ-herpesviruses.

Eukaryotic PIC assembly is nucleated by the deployment of TBP onto the promoter DNA by TFIID (35) and is followed by recruitment of other GTFs and Pol II. Unlike the direct

319 interaction that occurs between the ORF24 vTBP and Pol II, the interaction between cellular
320 TBP and Pol II is bridged through GTFs. The β- and γ-herpesviruses have thus adopted a
321 solution reminiscent of bacterial sigma factors, wherein vTBP binds promoter DNA while also
322 recruiting the polymerase directly to the promoter (36). This raises the question of whether
323 vTBP-mediated Pol II recruitment alters the requirement or roles for other GTFs that contact Pol
324 II. Of particular interest is the role or presence of TFIIB, which both interacts with TBP and is
325 necessary for Pol II recruitment early in PIC formation, functions that may not be necessary in
326 the vPIC given the interactions mediated by vTBP (20, 37). In this regard, it has recently been
327 shown that the levels of some components of the Pol II transcriptional machinery, including
328 TFIIB, are significantly decreased at late times during lytic gammaherpesvirus infection (38). It is
329 possible that these viruses assemble alternative transcription complexes in part to compensate
330 for reduced availability of key host factors. The role or presence of other GTFs in vPIC
331 assembly, including TFIIE and TFIIH, which are important for promoter opening and
332 phosphorylation of the Pol II CTD, are also unknown (39). An intriguing possibility is that the
333 other vTAs fulfill some subset of GTF-like functions during formation of the late gene vPIC. In
334 this regard, we have recently demonstrated that the KSHV vTAs ORF30 and ORF66 are
335 essential for stable binding of ORF24 to the K8.1 late gene promoter (15). Determining which of
336 the canonical GTFs are engaged in the vPIC will be key to understanding how this unique
337 hybrid virus-host complex activates late gene transcription.

338 Our work reveals that ORF24 directly binds hypophosphorylated CTD repeats,
339 consistent with a role facilitating viral PIC formation on late gene promoters. CTD-interacting
340 proteins exhibit exceptional diversity in their strategy for recognition of the CTD and in their
341 preference for phosphorylation (40). Relatively few proteins interact with hypophosphorylated
342 CTD; these proteins (TFIID, TFIIE, TFIIF, and Mediator) are all involved in PIC formation, as
343 phosphorylation of the CTD results in release from the PIC into elongating complexes (41). The
344 best characterized hypophosphorylated CTD interacting protein is the Mediator coactivator

complex (42). Mediator is a multi-subunit complex, and its interactions with the Pol II CTD are extensive, involving numerous Mediator subunits (32, 43-45). A crystal structure of the Mediator head module with the CTD revealed coordination of nearly four CTD repeats (32). Given the requirement for ORF24-NTD to bind at least four CTD repeats, we are intrigued by the possibility that vTBP functionally or structurally mimics the Mediator head module. One striking difference is the small size of the ORF24-NTD domain that nevertheless efficiently binds the CTD. Mediator is thought to transmit signals from transcription factors bound at regulatory elements to the basal transcriptional machinery (46). Late gene promoters have exceedingly minimal promoters and lack identified enhancer elements (8); thus, it is unclear if a Mediator-like function is required for transcription, or if recruitment of Pol II to the promoter is sufficient for transcription. The other vTAs may play a role in bridging currently unidentified enhancers or other elements, and may communicate this information to Pol II through the ORF24-ORF34 interaction.

A chimeric vTBP in which the minimal Pol II-interacting domain from ORF24 is replaced with that of mu24 is able to recruit Pol II and maintain interactions with other KSHV vTAs in order to form the vPIC and facilitate transcription. This is the first example of functional interchangeability of any subcomponent of the vTA complex. However, the role of vTBP in late gene transcription extends beyond polymerase recruitment, as vTBP also recruits the remainder of the vTAs to the promoter through protein-protein interactions while also directly binding the late gene promoter DNA.

In summary, ORF24 is a viral transcriptional activator that replaces TBP at late gene promoters and directly recruits Pol II to transcribe viral late genes. That ORF24 binds both the unphosphorylated CTD of Pol II and promoter DNA makes it unique among known eukaryotic CTD-interacting proteins. Since these two functions of ORF24 are genetically separable, they can be characterized independently of one another. Future work to gain atomic-level insight into

how ORF24-NTD coordinates the CTD repeats will advance our understanding of its remarkable role in the regulation of late gene transcription in the β- and γ-herpesviruses.

MATERIALS AND METHODS

Plasmids

All primer sequences are listed in **Table 1**. All plasmids used in this study have been deposited to Addgene. The following ORF24 fragments: residues 1-201 (ORF24-NTD) (Addgene #138420), residues 1-271 (Addgene #138421), and residues 1-133 (Addgene #138422) were PCR amplified from pcDNA4/TO-ORF24-3xFLAG (19) (Addgene #138423) with primers to introduce BamHI and NotI sites and cloned into pcDNA4/TO-3xFLAG (C-terminal tag) using T4 DNA ligase (New England Biolabs). pcDNA4/TO-ORF24 202-752-3xFLAG (Addgene #138424) was generated by inverse site-directed mutagenesis with Phusion DNA polymerase (New England Biolabs) using pcDNA4/TO-ORF24-3xFLAG as a PCR template. PCR products from inverse PCR were DpnI treated, then ligated using T4 PNK and T4 DNA ligase (New England Biolabs).

To generate the plasmid for GST-ORF24-NTD expression (Addgene #138464), ORF24-NTD (residues 1-201) was PCR amplified from pcDNA4/TO-ORF24-3xFLAG 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 WT (Addgene #138465) and 3L_A (Addgene #138466) expression, ORF24-NTD was PCR amplified from pcDNA4/TO-ORF24-3xFLAG WT (Addgene #138423) or 3L_A (Addgene #138425) (19) with primers to introduce SacI and BamHI sites and cloned into pMAL-c2X using T4 DNA ligase. pMAL-c2X encodes an N-terminal MBP tag and the plasmids were cloned to express ORF24-NTD with a noncleavable MBP tag. ORF24-NTD was PCR amplified from pcDNA4/TO-ORF24-3xFLAG WT and cloned into the KpnI and EcoRI sites of

plasmid p6H-SUMO3 using InFusion. A C-terminal Strep tag on ORF24-NTD was added by inverse PCR to generate p6H-SUMO3-ORF24-NTD-Strep (Addgene #138467).

To make the plasmid for GST-Rpb1-linker expression (Addgene #138468), the linker region of Rpb1 (a.a. 1460-1585) was PCR amplified from HEK293T cDNA with primers to introduce BamHI and NotI sites and cloned into pGEX4T1 using T4 DNA ligase. Rpb4 was PCR-amplified from HEK293T cDNA with primers to introduce a BamHI site and cloned into pGEX4T1 using InFusion. The N-terminal BamHI site was regenerated to keep Rpb4 in the same reading frame as the GST fusion tag. Rpb7 was PCR amplified from HEK293T cDNA with primers to introduce a Shine-Delgarno sequence and a C-terminal 6x-His tag and cloned into the EcoRI and NotI sites of pGEX4T1-Rpb4 using T4 DNA ligase to generate pGEX4T1-Rpb4/7 (Addgene #138484).

The 2x and 4x CTD repeat inserts were ordered as a pair of oligonucleotides from IDT, and the primers were annealed by cooling from 90°C to room temperature in a water bath. The 10x CTD repeat insert was ordered as a synthesized gene block from IDT (**Table 2**). All CTD inserts were cloned into the BamHI and NotI sites of pQLink-GST using InFusion cloning (Addgene #138470-138472) (Clontech). pQLink-GST encodes an N-terminal GST tag followed by a TEV protease cleavage site.

The following mutations were introduced into pcDNA4/TO-ORF24-3xFLAG using two primer site-directed mutagenesis with KAPA HiFi polymerase (Roche): L73A (Addgene #138426), L74A (Addgene #138427), L75A (Addgene #138428), L73A_L75A (Addgene #138429), L74A_L75A (Addgene #138430), R68A (Addgene #138431), and R72A (Addgene #138432). The L73A_L74A mutation (Addgene #138433) was introduced using InFusion (Clontech) site-directed mutagenesis.

Full-length ORF24 with a C-terminal Strep tag (pcDNA4/TO-ORF24-2xStrep) (Addgene plasmid #129742) was previously described (13). Full-length ORF24 with the 3L_A mutation was subcloned from pcDNA4/TO-ORF24 3L_A-CFLAG (Addgene #138425) into the BamHI and

XhoI sites of pcDNA4/TO-2xStrep (C-terminal tag) using InFusion cloning (Addgene #138440). Full-length UL87 was PCR amplified from HCMV Towne BAC DNA with primers to introduced EcoRI and XhoI sites and cloned into pcDNA4/TO-2xStrep (C-terminal tag) using T4 DNA ligase (Addgene #138434). Full-length mu24 was PCR amplified from MHV68-infected 3T3 cell cDNA with primers to introduce BamHI and NotI sites and cloned into pcDNA4/TO-2xStrep (C-terminal tag) using T4 DNA ligase (Addgene #138435). Full-length BcRF1 was PCR amplified from pcDNA4/TO-BcRF1-3xFLAG (19) and cloned into the BamHI and XhoI sites of pcDNA4/TO-2xStrep (C-terminal tag) using InFusion cloning (Addgene #138436). Mutations of the RLLLG motif in UL87, mu24, and BcRF1 (3L_A mutations) (Addgene #138437-138439) were generated using inverse PCR site-directed mutagenesis.

The minimal domains, minimal domain - 10 a.a., and minimal domain + 25 a.a. for ORF24, mu24, BcRF1, and UL87 were PCR amplified from these plasmids and cloned into BamHI/XhoI-cut pcDNA4/TO-2xStrep (C-terminal tag) using InFusion cloning (Addgene #138441-138452). Chimeras of the minimal domain (NTD) ORF24 homologs with ORF24 202-752 were generated using two-insert InFusion cloning (Addgene #138453-138455) into BamHI/XhoI-cut pcDNA4/TO-2xStrep (C-terminal tag).

Plasmid K8.1 Pr pGL4.16 (Addgene plasmid #120377) contains the minimal K8.1 promoter and ORF57 Pr pGL4.16 (Addgene plasmid #120378) contains a minimal ORF57 early gene promoter and have been described previously (13). Plasmids pcDNA4/TO-ORF18-2xStrep (Addgene plasmid #120372), pcDNA4/TO-ORF24-2xStrep (Addgene plasmid #129742), pcDNA4/TO-ORF30-2xStrep (Addgene plasmid #129743), pcDNA4/TO-ORF31-2xStrep (Addgene plasmid #129744), pcDNA4/TO-2xStrep-ORF34 (Addgene plasmid #120376) have been previously described (13). Plasmid pRL-TK (Promega) was kindly provided by Dr. Russell Vance.

447 **Tissue Culture and Transfections**

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

451

452 **Immunoprecipitation and Western Blotting**

453 Cell lysates were prepared 24 hours after transfection by washing and pelleting cells in
454 cold PBS, then resuspending the pellets in IP lysis buffer [50 mM Tris-HCl pH 7.4, 150 mM
455 NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor (Roche)] and rotating for 30 minutes at
456 4°C. Lysates were cleared by centrifugation at 21,000 x *g* for 10 min, then 1-2 mg (as indicated)
457 of total protein was incubated with pre-washed M2 α-FLAG magnetic beads (Sigma) or
458 MagStrep “type3” XT beads (IBA) overnight. Beads were washed 3x for 5 min each with IP
459 wash buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.05% NP-40) and eluted by
460 boiling in 2x Laemmli sample buffer (BioRad).

461 Lysates and elutions were resolved by SDS-PAGE and analyzed by western blot in
462 TBST (Tris-buffered saline, 0.2% Tween 20) using the following primary antibodies: rabbit α-
463 FLAG (Sigma, 1:2500); Strep-HRP (Millipore, 1:2500); rabbit α-Pol II clone N20 (Santa Cruz,
464 1:2500); mouse α-GST clone 8-326 (Pierce, 1:2000); mouse α-MBP (NEB, 1:10000); mouse α-
465 Pol II CTD clone 8WG16 (Abcam, 1:1000), rabbit α-Vinculin (Abcam, 1:1000), mouse anti-
466 GAPDH (1:1,000; Abcam). Following incubation with primary antibodies, the membranes were
467 washed with TBST and incubated with the appropriate secondary antibody. The secondary
468 antibodies used were the following: goat α-mouse-HRP (Southern Biotech, 1:5000) and goat α-
469 rabbit-HRP (Southern Biotech, 1:5000).

470

471

472

Protein Expression and Purification

GST-ORF24-NTD, GST-CTD linker, and GST. Proteins were expressed in Rosetta 2 cells (EMD Millipore) grown in LB at 37°C and induced at an OD_{600} of 0.7 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-2xCTD repeats, GST-4xCTD 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 0.7 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).

ORF24-NTD-Strep. The protein was expressed in NiCo21 (DE3) cells (New England Biolabs) 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 HEPES, pH 7.5, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol, 20 mM imidazole, 1 mM TCEP, 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 HisTrap (GE Healthcare) and step-eluted in wash buffer (100 mM HEPES, pH 7.5, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol, 1 mM TCEP) containing 500 mM imidazole. The fractions containing 6xHis-SUMO-ORF24-NTD-Strep were pooled. The SUMO tag was cleaved overnight at 4°C with 1 mg of SenP2 protease. Following cleavage of the SUMO tag, ORF24-NTD-Strep was purified on an equilibrated StrepTrap (GE Healthcare) and step-eluted in wash buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10 % glycerol, 1 mM TCEP) 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

524 on a 10K Amicon Ultra-15 concentrator (EMD Millipore). Protein aliquots were flash frozen and
525 stored at -70°C.
526
527 **Pulldown Assays**
528 ***GST pulldowns***
529 To test the interaction between GST-ORF24-NTD and Pol II from mammalian cells, 10
530 µg GST-ORF24-NTD or 10 µg GST was added to 20 µl of washed Glutathione Magnetic
531 Agarose beads (Pierce) along with 250 µg of 293T whole cell lysate. IP wash buffer (50 mM Tris
532 pH 7.4, 150 mM NaCl, 0.05% NP-40, 1 mM EDTA) was added to a final volume of 300 µl. The
533 samples were rotated at 4°C for 1 hour. Following the pulldown, the samples were washed with
534 IP wash buffer three times for 5 minutes each time. After the last wash, the protein was eluted
535 by boiling in 2x Laemmli sample buffer (BioRad). The pulldown to test the interaction between
536 GST-CTD repeats or GST-CTD linker and MBP-ORF24-NTD or MBP-ORF24-NTD 3L_A were
537 performed as described above. The elutions were resolved by SDS-PAGE followed by western
538 blot.
539 ***Strep pulldowns***
540 To test the interaction between ORF24-NTD and Rpb4/7 or xCTD repeats, 5 µg of
541 ORF24-NTD-Strep was added to 10 µl of washed MagStrep “type 3” XT Beads (IBA) along with
542 10 µg of GST-Rpb4/Rpb7-6xHis or GST-xCTD repeats. SEC buffer (20 mM HEPES pH 7.4, 100
543 mM NaCl, 1 mM TCEP, 5% glycerol) including 0.05% NP-40 and 5 mM DTT was added to a
544 final volume of 300 µl. The pulldowns were rotated at 4°C for 1 hour followed by three 5 minute
545 washes. The protein was eluted by boiling in 2x Laemmli sample buffer. Elutions were resolved
546 by SDS-PAGE on a Stain-free gel (BioRad) followed by western blot.
547
548
549

Late Gene Reporter Assay

HEK293T cells (1 x 10_6 cells) were plated in 6-well plates and after 24 h each well was transfected with 900 ng of DNA containing 125 ng each of pcDNA4/TO-ORF18-2xStrep, wild-type pcDNA4/TO-ORF24-2xStrep or a homolog of ORF24, pcDNA4/TO-ORF30-2xStrep, pcDNA4/TO-ORF31-2xStrep, pcDNA4/TO-2xStrep-ORF34, pcDNA4/TO-ORF66-2xStrep (or as a control, 750 ng of empty pcDNA4/TO-2xStrep plasmid in lieu of vTA plasmids), with 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 h, cells were rinsed twice with PBS, lysed by rocking for 15 min at room temperature in 500 µL of Passive Lysis Buffer (Promega), and clarified by centrifugation 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. All samples were normalized to the corresponding control containing empty plasmid.

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 (29, 47). The DNA construct was previously described (28) and is an SCP (48) containing a BREu element upstream of the TATA box (21) and an EcoRI restriction enzyme site downstream of the INR element for purification purposes. A biotin tag was coupled to the 5' end of the template strand (Integrated DNA Technologies). The duplex DNA was generated by annealing the single-stranded template strand DNA with equimolar amounts of non-template strand DNA at a final concentration of 50 µM in water. The annealing reaction was carried out by incubating at 98°C for 2 minutes followed by cooling to room temperature at a rate of 1°C per second.

PICs were assembled in assembly buffer (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 10% glycerol, 6 mM $MgCl_2$, 80 mM KCl, 1 mM DTT, 0.05% NP-40). DNA was used as a scaffold and purified TBP, TFIIA, TFIIB, Pol II, TFIIF, and GST-ORF24-NTD were sequentially added into the assembly buffer. 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, 3% trehalose, 8 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 0.025% NP-40). The complex was eluted by incubation at 28°C for 1 hour in digestion buffer (10 mM HEPES, pH 7.9, 3% trehalose, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.01% NP-40, 1 unit μL^{-1} EcoRI-HF (New England Biolabs)). After elution, purified PIC was crosslinked on ice in 0.05% glutaraldehyde for 5 minutes then immediately used for EM sample preparation.

Electron microscopy. Negative stain samples of PIC 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). An aliquot (3.5 µL) of the purified sample was placed onto the grid and allowed to absorb for 5 minutes at 100% humidity. The grid was then placed sample-side-down on five successive 75 µL drops of 2% (w/v) uranyl formate solution for 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 80,000X (1.5 Å/pixel). The data were collected on a 4k X 4k CCD (Gatan) using low-dose procedures (20 e^- $Å^{-2}$ total dose per exposure), using Leginon software to automatically focus and collect exposure images.

Image processing. Data pre-processing was performed using the Appion processing environment (49). Particles were automatically selected from the micrographs using a difference of Gaussians (DoG) particle picker (50). The contrast transfer function (CTF) of each

602 micrograph was estimated using both ACE2 and CTFFind (51, 52). Boxed particle images were
603 extracted using a box size of 256 X 256 pixels from images whose ACE2 confidence value was
604 greater than 0.8, phases were flipped, and images were normalized using the XMIPP to remove
605 pixels which were above or below 4.5σ of the mean value (53). The particle stack was binned by
606 a factor of two and two-dimensional classification was conducted using iterative multireference
607 alignment analysis (MSA-MRA) within the IMAGIC software (54).
608
609 ***Three-dimensional reconstruction*.** Particles belonging to bad two-dimensional classes were
610 thrown out, resulting in a stack of 79381 single particle images that were used for three-
611 dimensional analysis. Three-dimensional classification was performed within RELION (55),
612 using the cryo-EM structural of a minimal PIC (EMD-2305, (28)), low-pass filtered to 50 Å
613 resolution, as an initial reference, and sorted into 6 classes. The resolution of the
614 reconstructions containing GST-ORF24-NTD were estimated to be ~20 Å.
615
616 **ACKNOWLEDGEMENTS**

617 We thank Jie Fang for purification of human GTFs and Pol II, and Patricia Grob and
618 Abhiram Chintangal for microscope and computational support, respectively. We are thankful to
619 all members of the Glaunsinger lab, especially Divya Nandakumar, for helpful discussions and
620 suggestions. This research was supported by NIH R01AI122528 to B.G. and R01GM63072 and
621 R35GM127018 to E.N. A.C. was supported by a National Science Foundation Graduate
622 Research Fellowship (DGE 1752814) and a UC Berkeley Chancellor's Fellowship. A.L.D. is The
623 Rhee Family Fellow of the Damon Runyon Cancer Research Foundation (DRG-2349-18). B.G.
624 and E.N. are investigators of the Howard Hughes Medical Institute.
625
626 **COMPETING INTERESTS**
627 None

REFERENCES

1. M. C. Thomas, C. M. Chiang, The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**, 105-178 (2006).
2. K. M. Harlen, L. S. Churchman, The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* **18**, 263-273 (2017).
3. S. Buratowski, Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**, 541-546 (2009).
4. H. Lu, O. Flores, R. Weinmann, D. Reinberg, The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl Acad Sci U S A* **88**, 10004-10008 (1991).
5. D. L. V. Bauer *et al.*, Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Rep* **23**, 2119-2129 e2113 (2018).
6. T. Chi, M. Carey, Assembly of the isomerized TFIIA--TFIID--TATA ternary complex is necessary and sufficient for gene activation. *Genes Dev* **10**, 2540-2550 (1996).
7. A. Haigh, R. Greaves, P. O'Hare, Interference with the assembly of a virus-host transcription complex by peptide competition. *Nature* **344**, 257-259 (1990).
8. D. Nandakumar, B. Glaunsinger, An integrative approach identifies direct targets of the late viral transcription complex and an expanded promoter recognition motif in Kaposi's sarcoma-associated herpesvirus. *PLoS Pathog* **15**, e1007774 (2019).
9. T. R. Serio, N. Cahill, M. E. Prout, G. Miller, A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. *J Virol* **72**, 8338-8343 (1998).
10. S. Tang, K. Yamanegi, Z. M. Zheng, 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 (2004).
11. E. Wong-Ho *et al.*, Unconventional sequence requirement for viral late gene core promoters of murine gammaherpesvirus 68. *J Virol* **88**, 3411-3422 (2014).
12. V. Aubry *et al.*, Epstein-Barr virus late gene transcription depends on the assembly of a virus-specific preinitiation complex. *J Virol* **88**, 12825-12838 (2014).
13. A. F. Castaneda, B. A. Glaunsinger, The Interaction between ORF18 and ORF30 Is Required for Late Gene Expression in Kaposi's Sarcoma-Associated Herpesvirus. *J Virol* **93** (2019).
14. Z. H. Davis, C. R. Hesser, J. Park, B. A. Glaunsinger, 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 (2016).
15. A. L. Didychuk, A. F. Castaneda, L. O. Kushnir, C. J. Huang, B. A. Glaunsinger, Conserved CxnC motifs in Kaposi's sarcoma-associated herpesvirus ORF66 are required for viral late gene expression and are essential for its interaction with ORF34. *J Virol* 10.1128/JVI.01299-19 (2019).
16. D. Pan *et al.*, Murine Cytomegalovirus Protein pM91 Interacts with pM79 and Is Critical for Viral Late Gene Expression. *J Virol* **92** (2018).
17. H. Gruffat, F. Kadjouf, B. Mariame, E. Manet, 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 (2012).
18. L. S. Wyrwicz, L. Rychlewski, Identification of Herpes TATT-binding protein. *Antiviral Res* **75**, 167-172 (2007).
19. Z. H. Davis *et al.*, Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. *Mol Cell* **57**, 349-360 (2015).
20. H. T. Chen, S. Hahn, 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 (2003).

21. F. T. Tsai, P. B. Sigler, Structural basis of preinitiation complex assembly on human pol II promoters. *EMBO J* **19**, 25-36 (2000).
22. M. Lukarska *et al.*, Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **541**, 117-121 (2017).
23. C. Zhou, D. M. Knipe, Association of herpes simplex virus type 1 ICP8 and ICP27 proteins with cellular RNA polymerase II holoenzyme. *J Virol* **76**, 5893-5904 (2002).
24. J. Q. Dai-Ju, L. Li, L. A. Johnson, R. M. Sandri-Goldin, 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 (2006).
25. A. Feklistov, B. D. Sharon, S. A. Darst, C. A. Gross, Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-376 (2014).
26. L. V. Nobre *et al.*, Human cytomegalovirus interactome analysis identifies degradation hubs, domain associations and viral protein functions. *Elife* **8** (2019).
27. L. P. Kozlowski, IPC - Isoelectric Point Calculator. *Biol Direct* **11**, 55 (2016).
28. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-486 (2013).
29. A. Revyakin *et al.*, Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev* **26**, 1691-1702 (2012).
30. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365 (2016).
31. H. Spahr, G. Calero, D. A. Bushnell, R. D. Kornberg, Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proc Natl Acad Sci U S A* **106**, 9185-9190 (2009).
32. P. J. Robinson, D. A. Bushnell, M. J. Trnka, A. L. Burlingame, R. D. Kornberg, Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A* **109**, 17931-17935 (2012).
33. H. Lu *et al.*, Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* **558**, 318-323 (2018).
34. M. Boehning *et al.*, RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* **25**, 833-840 (2018).
35. A. B. Patel *et al.*, Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362** (2018).
36. K. B. Decker, D. M. Hinton, Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* **67**, 113-139 (2013).
37. X. Liu, D. A. Bushnell, D. Wang, G. Calero, R. D. Kornberg, Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206-209 (2010).
38. E. Hartenian, S. Gilbertson, J. D. Federspiel, I. M. Cristea, B. A. Glaunsinger, RNA decay during gammaherpesvirus infection reduces RNA polymerase II occupancy of host promoters but spares viral promoters. *PLoS Pathog* **16**, e1008269 (2020).
39. E. Nogales, R. K. Louder, Y. He, Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **46**, 59-83 (2017).
40. C. Jeronimo, A. R. Bataille, F. Robert, The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem Rev* **113**, 8491-8522 (2013).
41. D. Eick, M. Geyer, The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* **113**, 8456-8490 (2013).
42. J. Soutourina, Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* **19**, 262-274 (2018).
43. T. M. Harper, D. J. Taatjes, The complex structure and function of Mediator. *J Biol Chem* **293**, 13778-13785 (2018).

44. C. Plaschka *et al.*, Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376-380 (2015).
45. P. J. Robinson *et al.*, Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422 e1416 (2016).
46. B. L. Allen, D. J. Taatjes, The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**, 155-166 (2015).
47. M. T. Knuesel, K. D. Meyer, C. Bernecky, D. J. Taatjes, The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev* **23**, 439-451 (2009).
48. T. Juven-Gershon, S. Cheng, J. T. Kadonaga, Rational design of a super core promoter that enhances gene expression. *Nat Methods* **3**, 917-922 (2006).
49. G. C. Lander *et al.*, Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* **166**, 95-102 (2009).
50. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166**, 205-213 (2009).
51. S. P. Mallick, B. Carragher, C. S. Potter, D. J. Kriegman, ACE: automated CTF estimation. *Ultramicroscopy* **104**, 8-29 (2005).
52. J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-347 (2003).
53. C. O. Sorzano *et al.*, XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**, 194-204 (2004).
54. M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17-24 (1996).
55. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

FIGURES

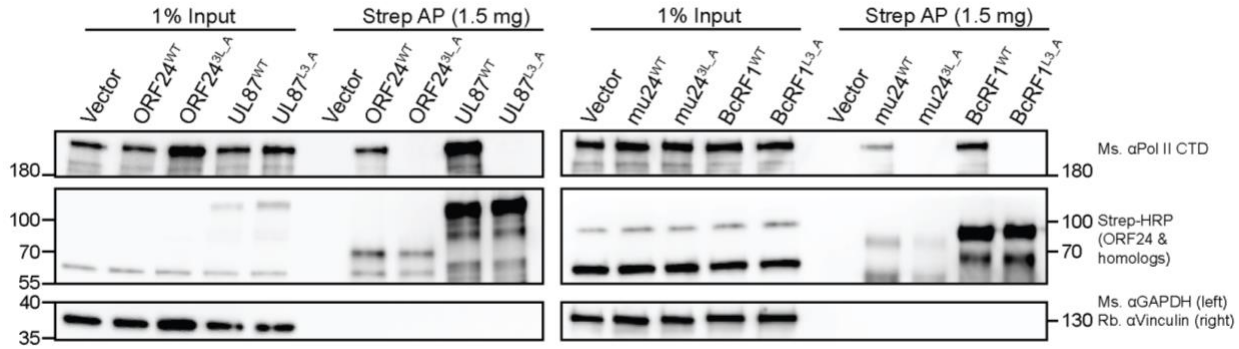


Figure 1. The RLLLG motif in ORF24 homologs from other β- and γ-herpesviruses is required for interaction with Pol II.

Full-length WT or 3L_A mutants, which mutates the conserved RLLLG motif, of Strep-tagged ORF24 or homologs from MHV68 (mu24), EBV (BcRF1), and HCMV (UL87) were transiently transfected into HEK293T cells then co-affinity purified with StrepTactinXT beads followed by western blotting.

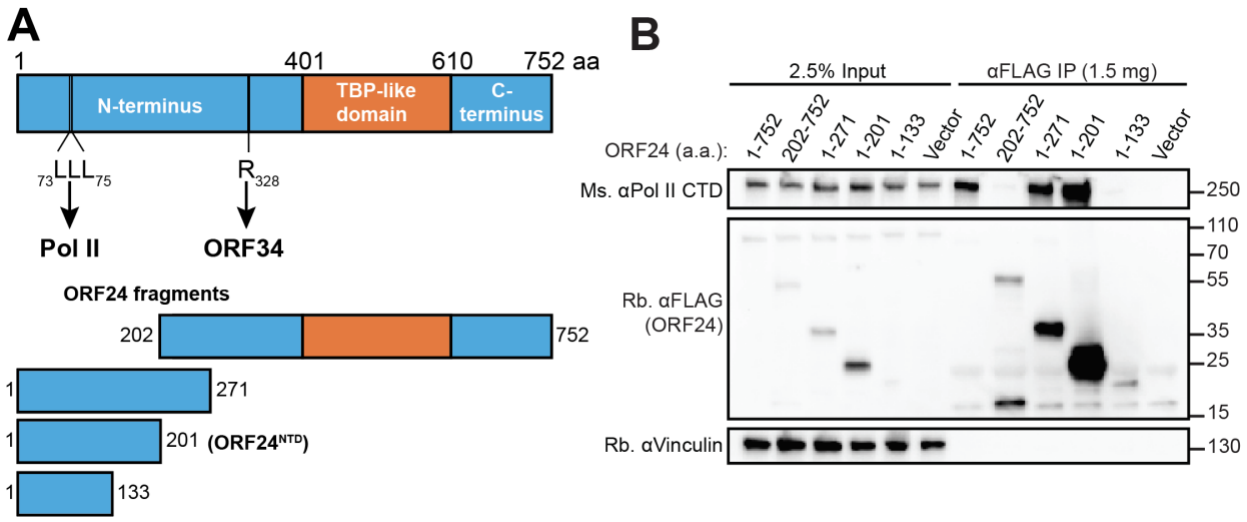


Figure 2. The N-terminal domain of ORF24 (ORF24-NTD) binds Pol II.

(A) Schematic of constructs used to identify a minimal N-terminal domain of ORF24 showing the predicted boundaries for the N-terminal domain, the TBP-like domain, and the C-terminal domain, including residues known to be required for Pol II binding (amino acids 73-75) and interaction with ORF34 (amino acid 328).

(B) HEK293T cells were transiently transfected with full-length or truncated FLAG-tagged ORF24 and co-immunoprecipitated with anti-FLAG beads followed by western blotting with the indicated antibodies to detect ORF24 and Pol II.

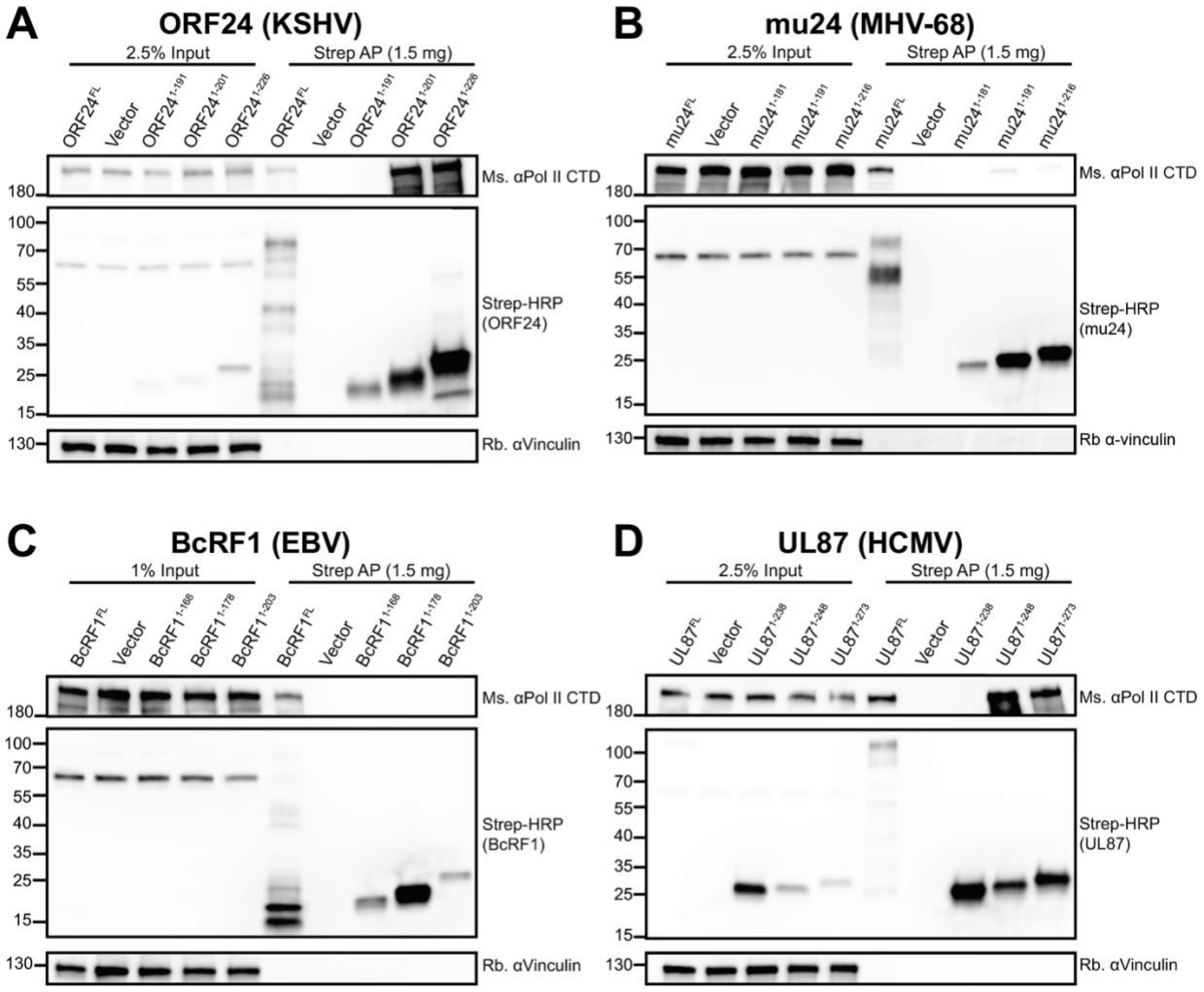


Figure 3. The N-terminal domain of ORF24 homologs from other β- and γ-herpesviruses is sufficient for interaction with Pol II.

(A-D) Full-length or truncated Strep-tagged constructs of ORF24 (A) or homologs from MHV68 (mu24; B), EBV (BcRF1; C), and HCMV (UL87; D) were transiently transfected into HEK293T cells then co-affinity purified with StrepTactinXT beads followed by western blotting.

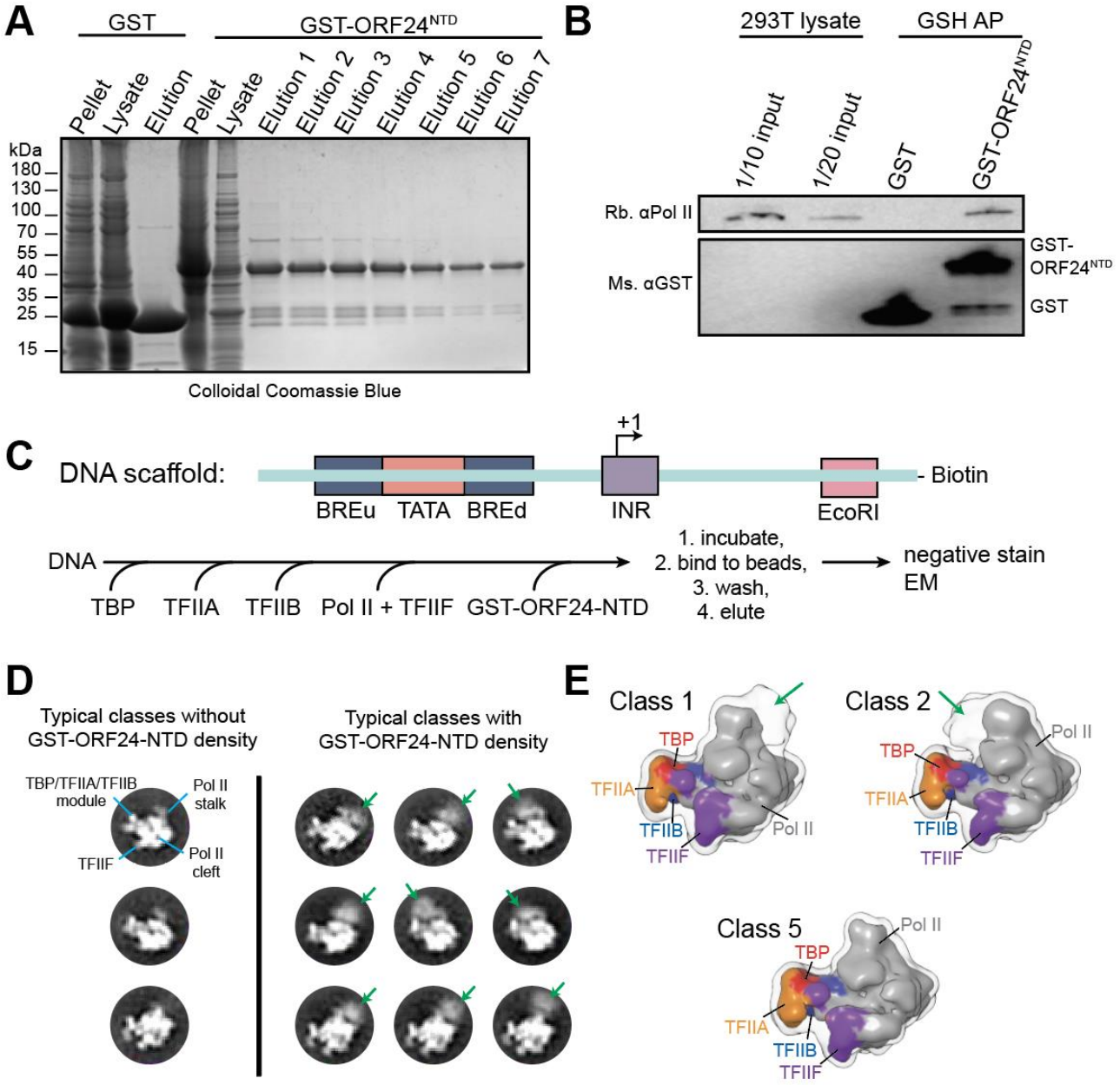


Figure 4. ORF24-NTD binds Pol II in minimal PICs.

(A) Colloidal Coomassie gel demonstrating that GST and GST-ORF24-NTD can be recombinantly expressed in *E. coli* and purified by glutathione sepharose purification.

(B) GST or GST-ORF24-NTD was incubated in HEK293T whole cell lysate, then subjected to affinity purification using glutathione magnetic beads followed by western blotting.

(C) Sequential reconstitution strategy for a minimal PIC containing GST-ORF24-NTD.

(D) Representative reference-free two-dimensional class averages of negatively stained minimal PICs (TBP/TFIIA/TFIIB/TFIIF/Pol II/DNA) assembled in the presence of GST-ORF24-NTD. Three classes on the left show different views of the minimal PIC alone, with the class average in the upper-left annotated with the main features of a minimal PIC particle. The nine class averages on the right show diffuse density in various positions around the Pol II stalk attributed to bound GST-ORF24-NTD (green arrows).

(E) Representative three-dimensional class averages of negatively stained minimal PICs assembled in the presence of GST-ORF24-NTD. Classes 1 and 2 exhibit two major areas occupied by bound GST-ORF24-NTD proximal to the Pol II stalk, while class 5 does not exhibit any such density near the Pol II stalk. Solid surfaces are colored by subunit, while a lower intensity iso-surface is shown in transparency to reveal the weaker density attributed to bound GST-ORF24-NTD (green arrows).

(F) Difference mapping of the densities attributed to bound GST-ORF24-NTD. Shown on the left are two-dimensional projections of class 1 (top) and 2 (bottom) from (E), and on the right are the difference maps, called “Class 1 Δ” and “Class 2 Δ”, calculated by subtracting Class 5 from each of the respective classes.

(G) Three-dimensional difference maps corresponding to the extra density attributed to bound GST-ORF24-NTD, mapped onto the structure of the minimal PIC (PDB 5IYA).

(H) Zoomed in view of (G) with the structure of *Schizosaccharomyces pombe* Rpb1 (PDB 3H0G) superposed onto the human structure to show the location of the beginning portion of the Rpb1 CTD within the Pol II stalk. Note that only the very N-terminal portion of the Rpb1 CTD is visible in this structure, with >450 amino acids following this sequence in the CTD of human Rpb1.

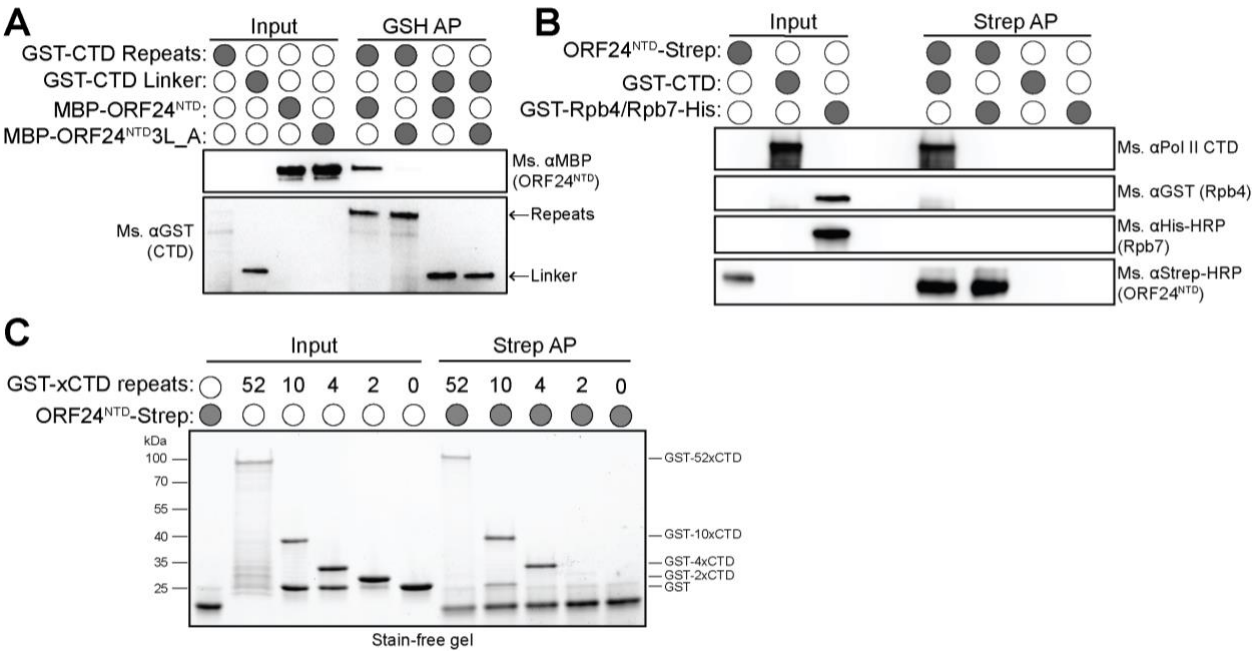
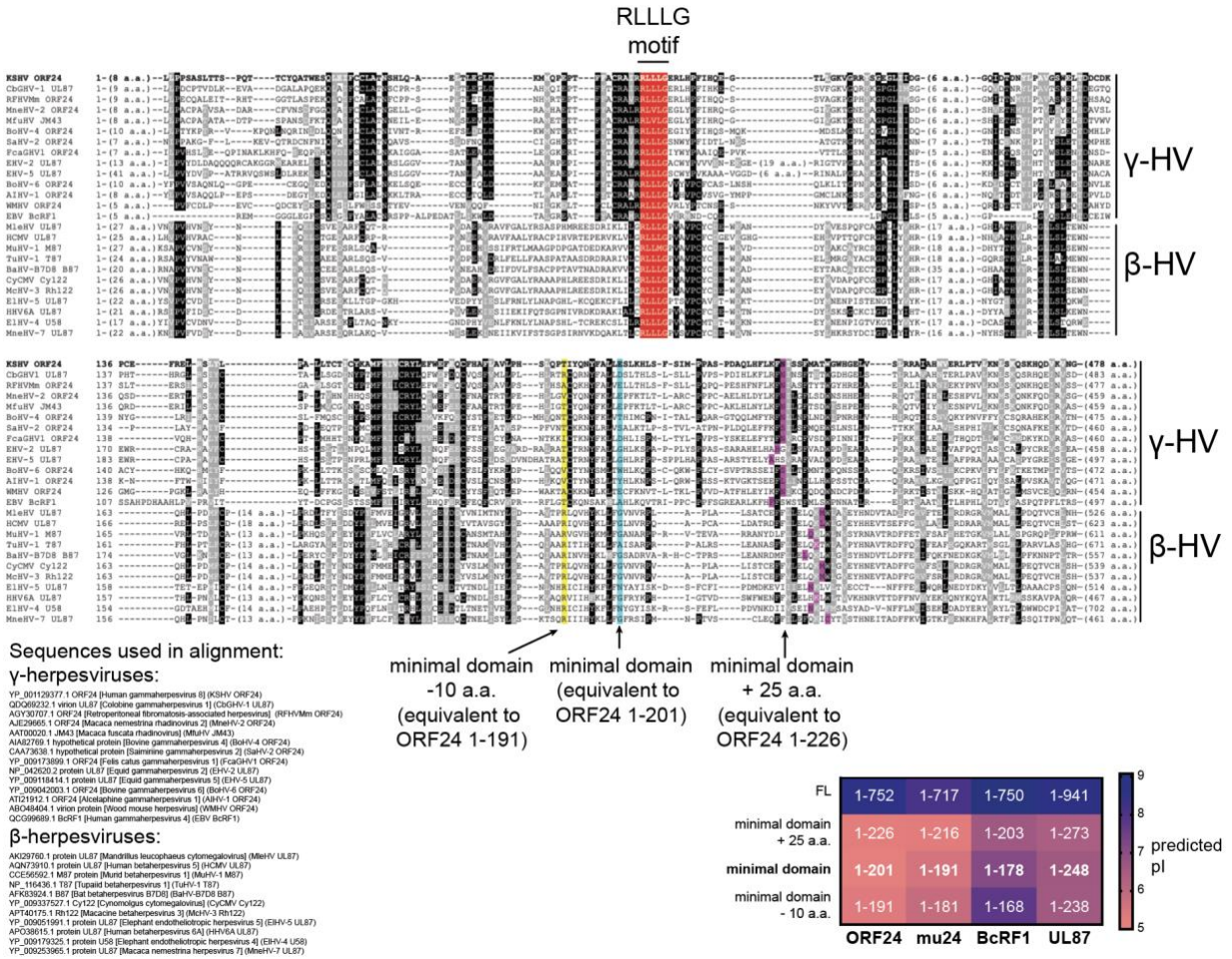
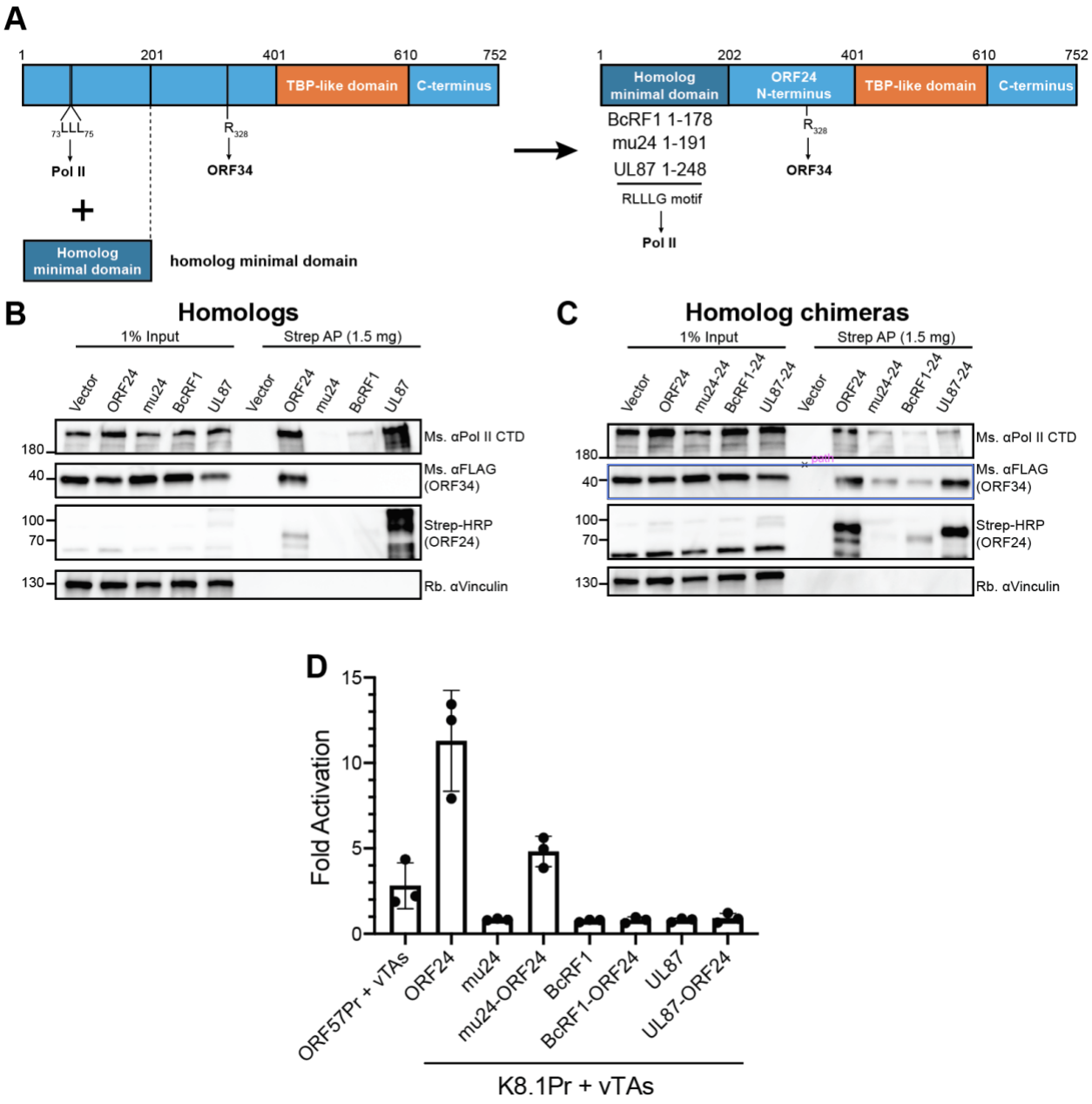


Figure 5. ORF24-NTD directly interacts with four heptapeptide repeats of the Pol II CTD.
(A) Recombinantly purified GST-CTD repeats or the GST-CTD linker were incubated with either purified MBP-ORF24-NTD or MBP-ORF24-NTD-3L_A protein, then subjected to a glutathione pulldown. Samples were resolved by SDS-PAGE and analyzed by western blot.
(B) Recombinantly purified GST-CTD repeats or GST-Rpb4/His-Rpb7 heterodimer were incubated with recombinantly purified Strep-tagged ORF24-NTD, then subjected to a StrepTactinXT pulldown. Samples were resolved by SDS-PAGE and analyzed by western blot.
(C) Recombinantly purified full-length GST-CTD repeats (52 repeats) shorter GST-CTD constructs (10, 4, or 2 repeats), or GST itself were incubated with recombinantly purified Strep-tagged ORF24-NTD, then subjected to a StrepTactinXT pulldown. Samples were resolved by SDS-PAGE and stained with colloidal Coomassie.



Supplemental Figure S1.

Multiple sequence alignment of homologs of ORF24 from other β- and γ-herpesviruses. The conserved triple leucine motif is highlighted in red. The location of truncations for the constructs used in Figure 3 are highlighted in yellow, teal, and purple. Sequences used to construct the alignment are listed in the bottom left. The boundaries for the truncated constructs and their relative predicted isoelectric point based on (1) is shown in the bottom right.



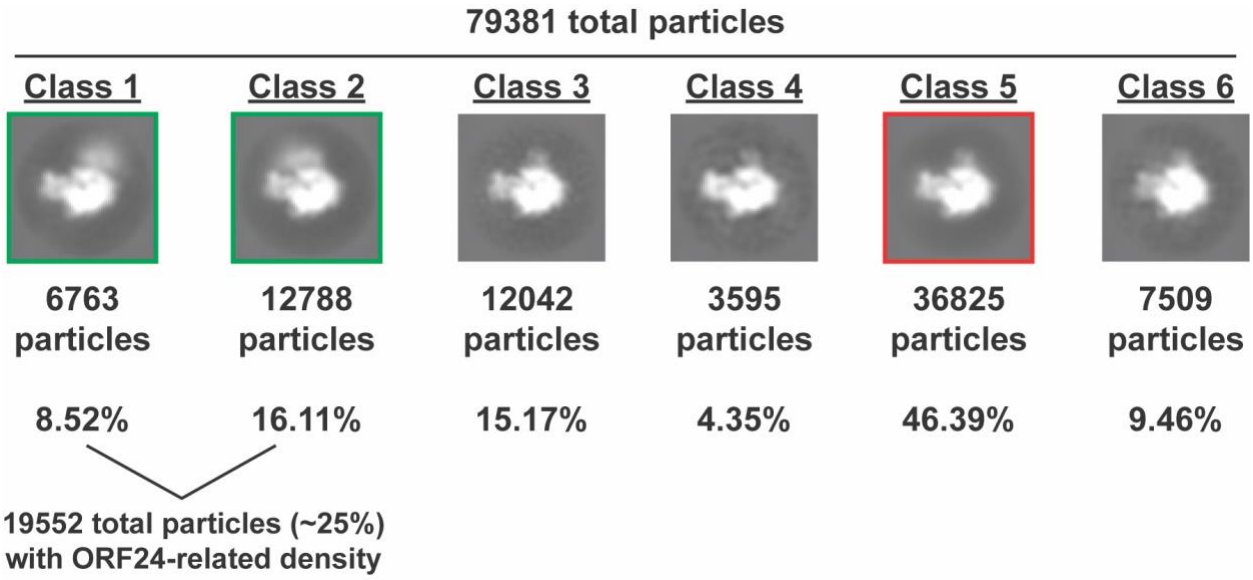
Supplemental Figure S2.

(A) Schematic of construct design for ORF24 chimeras. The ORF24-NTD (a.a. 1-201) was replaced with the experimentally identified minimal domain of mu24, BcRF1, and UL87. These chimeric constructs retain the N-terminal ORF24-ORF34 interaction region, the ORF24 vTBP domain, and ORF24 C-terminal tail.

(B) Full-length Strep-tagged homologs of ORF24 were transiently transfected into HEK293T cells along with FLAG-tagged ORF34, then co-affinity purified with StrepTactinXT beads followed by western blotting.

(C) Full-length Strep-tagged chimeras of ORF24 were transiently transfected into HEK293T cells along with FLAG-tagged ORF34, then co-affinity purified with StrepTactinXT beads followed by western blotting.

(D) HEK293T cells were transiently transfected with a pGL4.16 firefly luciferase plasmid driven by either the ORF57 (early gene) or K8.1 (late gene) promoter. Plasmids encoding either ORF24, its homologs, or the chimeras, along with the five remaining KSHV vTAs (ORFs 18, 30, 31, 34, and 66) and a pRL-TK renilla luciferase plasmid (as a transfection control) were also transfected. After 24 h, cell lysates were harvested and luciferase activity was measured. Fold activation was calculated by normalizing to the firefly/renilla signal in the absence of vTAs.



Supplemental Figure S3.

Two-dimensional projections of the three-dimensional classes resulting from sorting the single particle EM images of negatively stained minimal PICs (TBP/TFIIA/TFIIB/TFIIF/Pol II/DNA) assembled in the presence of GST-ORF24-NTD. The number of particles assigned to each class and the percent of total particles are indicated. Classes 1, 2, and 5 were used for the difference mapping shown in Figure 4.

1006 **Table S1. List of oligonucleotides used in this study**

#	Name	Sequence 5'-3'	Construct
1	ORF24 1-133_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	pCDNA4.TO-ORF24 1-133-CFLAG
2	ORF24 1-133_pCDNA_R	TCGAGCGGCCGCCCACAATCATCGGTAAGTTCCCATGATC	pCDNA4.TO-ORF24 1-133-CFLAG
3	ORF24 1-201_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	pCDNA4.TO-ORF24 1-201-CFLAG
4	ORF24 1-201_pCDNA_R	TCGAGCGGCCGCCCCTCCAGGAGTGCAAAATAATTTTGATAGATTG	pCDNA4.TO-ORF24 1-201-CFLAG
5	ORF24 1-271_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	pCDNA4.TO-ORF24 1-271-CFLAG
6	ORF24 1-271_pCDNA_R	TCGAGCGGCCGCTTCTTGACGTCCTGGTGCTTACTCT	pCDNA4.TO-ORF24 1-271-CFLAG
7	ORF24 202-752_pCDNA_F	ATGAGCCTGAAGCATCTCTCG	pCDNA4.TO-ORF24 202-752-CFLAG
8	ORF24 202-752_pCDNA_R	GGATCCGAGCTCGGTACCAAG	pCDNA4.TO-ORF24 202-752-CFLAG
9	ORF24 1-201_pGEX_F	GCGTGGATCCATGGCAGCGCTCGAGG	pGEX4T1-ORF24-NTD
10	ORF24 1-201_pGEX_R	CGATGCGGCCGCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGTG	pGEX4T1-ORF24-NTD
11	ORF24 1-201_pMAL_F	ATTCGAGCTCAATGGCAGCGCTCGAGGG	pMAL-c2x-ORF24-NTD, WT and ΔLLL
12	ORF24 1-201_pMAL_R	TAGAGGATCCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGT	pMAL-c2x-ORF24-NTD, WT and ΔLLL
13	Rpb1_Linker aa 1460-1585_F	GCGTGGATCCCTGGGCCAGCTGGCTC	pGEX4T1-Rpb1-linker
14	Rpb1_Linker aa 1460-1585_R	CGATGCGGCCGCTGGTGAAGGGATGTAGGGGCT	pGEX4T1-Rpb1-linker
15	Rpb4 pGEX4T1_F	GGTTCCGCGTGGATCCATGGCGGCGGGTGG	pGEX4T1-hsRpb4
16	Rpb4 pGEX4T1_R	GGAATTCCGGGGATCTTAATACTGAAAGCTGCGCT	pGEX4T1-hsRpb4
17	Rpb7 pGEX4T1_F	CCCGGAATTCaggaggTAATTAAATatgttctaccatatctccctagagcac	pGEx4T1-hsRpb4/7
18	Rpb7 pGEX4T1_R	CGATGCGGCCGCttagtgatggtggtgatggtggcttacaagccccaagtaatcgt	pGEX4T1-hsRpb4/7
19	ORF24_2-201_p6H-SUMO3_F	cagcagacgggagggGCAGCGCTCGAGGGC	p6H-SUMO3-ORF24-NTD-Strep
20	ORF24_2-201_p6H-SUMO3_R	gacggagctcgaattTTATTTTTCAAACTGCGGATGGCTCCACTCCAGGAGTGCAAAATAATTTTGATAGATTG	p6H-SUMO3-ORF24-NTD-Strep
21	2xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATTCTCCGACCTCTCCATCATACAGCCCTACCTCCCCGTCCTAGGCGGCCGCCTAGGACCC	pQLink-GST-2xCTD
22	2xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGACGGGGAGGTAGGGCTGTATGATGGAGAGGTCGGAGAATAGGATCCCTGAAAATAA	pQLink-GST-2xCTD
23	4xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATAGCCCCACGAGCCCTTCTTACTCTCCGACCTCACCATCCTATTCGCCTACTAGCCCGAGTTACAGTCCCACATCTCCGTCCTAGGCGGCCGCCTAGGACCC	pQLink-GST-4xCTD
24	4xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGACGGAGATGTGGGACTGTAACTCGGGCTAGTAGGCGAATAGGATGGTGAGGTCGGAGAGTAAGAAGGGCTCGTGGGGCTATAGGATCCCTGAAAATAA	pQLink-GST-4xCTD
24	ORF24_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	pCDNA4.TO-ORF24 3L_A-CSTREP
25	ORF24_R	caccgcctccctcgaggacCAGCGGACG	pCDNA4.TO-ORF24 3L_A-CSTREP
26	BcRF1_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	pCDNA4.TO-BcRF1-CSTREP
27	BcRF1_R	CACCGCCTCCCTCGAGCACTTGAGCATCACGGCAGTGG	pCDNA4.TO-BcRF1-CSTREP
28	BcRF1 (3L_A)_F	GCCGCGGCAggcgttatccgaataaatgactgccaggag	pCDNA4.TO-BcRF1 3L_A-CSTREP
29	BcRF1 (3L_A)_R	acgacgcatagcccgacacgcgtaaaaggtg	pCDNA4.TO-BcRF1 3L_A-CSTREP
30	mu24_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTCTGTGATTTGC	pCDNA4.TO-mu24-CSTREP
31	mu24_R	CTCCCTCGAGCGGCCGCCCGGAGTCTGGTTGGCAAGG	pCDNA4.TO-mu24-CSTREP
32	mu24 (3L_A)_F	GCCGCGGCAggcataagactgtacccaacctgcaatactg	pCDNA4.TO-mu24 3L_A-CSTREP
33	mu24 (3L_A)_R	tcttcttaaagctctacataaacagaatgt	pCDNA4.TO-mu24 3L_A-CSTREP
34	UL87_F	AAAAGAATTCATGGCCGGCGCTGC	pCDNA4.TO-UL87-CSTREP
35	UL87_R	AAAACTCGAGTCGTGATGCAAACCGCAC	pCDNA4.TO-UL87-CSTREP
36	UL87 (3L_A)_F	GCCGCGGCAGGACCCGTGGCCGTACCCTGTTTTTGCGAC	pCDNA4.TO-UL87 3L_A-CSTREP
37	UL87 (3L_A)_R	ACGACCCAGTACCAGCTTGACACGCTCGGA	pCDNA4.TO-UL87 3L_A-CSTREP
38	ORF24_1-191_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	pCDNA4.TO-ORF24 1-191-CSTREP
39	ORF24_1-191_pCDNA_R	CACCGCCTCCCTCGAGTGTGGGTTGTAGACTATGGGGC	pCDNA4.TO-ORF24 1-191-CSTREP
40	ORF24_1-201_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	pCDNA4.TO-ORF24 1-201-CSTREP
41	ORF24_1-201_pCDNA_R	CACCGCCTCCCTCGAGCTCCAGGAGTGCAAAATAATTTTG	pCDNA4.TO-ORF24 1-201-CSTREP
42	ORF24_1-226_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	pCDNA4.TO-ORF24 1-226-CSTREP
43	ORF24_1-226_pCDNA_R	CACCGCCTCCCTCGAGGTTAAACTTTAAAAAATGTAGC	pCDNA4.TO-ORF24 1-226-CSTREP
44	BcRF1_1-168_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	pCDNA4.TO-BcRF1 1-168-CSTREP
45	BcRF1_1-168_pCDNA_R	CACCGCCTCCCTCGAGGGTCCCTAGGAAGCGACGC	pCDNA4.TO-BcRF1 1-168-CSTREP

46	BcRF1_1-178_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	pCDNA4.TO-BcRF1 1-178-CSTREP
47	BcRF1_1-178_pCDNA_R	CACCGCCTCCCTCGAGAGCCAGCATCTTTGCAGAGTTTTGC	pCDNA4.TO-BcRF1 1-178-CSTREP
48	BcRF1_1-203_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	pCDNA4.TO-BcRF1 1-203-CSTREP
49	BcRF1_1-203_pCDNA_R	CACCGCCTCCCTCGAGGAAGTGGAACTTGAGTCTGGCC	pCDNA4.TO-BcRF1 1-203-CSTREP
50	mu24_1-181_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	pCDNA4.TO-mu24 1-181-CSTREP
51	mu24_1-181_pCDNA_R	CACCGCCTCCCTCGAGAGCAGACCTGGCCCAGGG	pCDNA4.TO-mu24 1-181-CSTREP
52	mu24_1-191_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	pCDNA4.TO-mu24 1-191-CSTREP
53	mu24_1-191_pCDNA_R	CACCGCCTCCCTCGAGCTCAGTAAGCTTCAAGTAATTTTTT	pCDNA4.TO-mu24 1-191-CSTREP
54	mu24_1-216_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	pCDNA4.TO-mu24 1-216-CSTREP
55	mu24_1-216_pCDNA_R	CACCGCCTCCCTCGAGATTGAATTTGATATATTCCAAATGA	pCDNA4.TO-mu24 1-216-CSTREP
56	UL87_1-238_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	pCDNA4.TO-UL87 1-238-CSTREP
57	UL87_1-238_pCDNA_R	CACCGCCTCCCTCGAGGCGCGGCGCCGCGGCCTCG	pCDNA4.TO-UL87 1-238-CSTREP
58	UL87_1-248_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	pCDNA4.TO-UL87 1-248-CSTREP
59	UL87_1-248_pCDNA_R	CACCGCCTCCCTCGAGTCCAAAGAGCAGCTTGTAGTGAACC	pCDNA4.TO-UL87 1-248-CSTREP
60	UL87_1-273_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	pCDNA4.TO-UL87 1-273-CSTREP
61	UL87_1-273_pCDNA_R	CACCGCCTCCCTCGAGAAGCTTTTGCAGCTCCAGC	pCDNA4.TO-UL87 1-273-CSTREP
62	BcRF1-ORF24_1_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	pCDNA4.TO-BcRF1-ORF24-CSTREP
63	BcRF1-ORF24_1_R	TTCAGGCTagccagcatctttgc	pCDNA4.TO-BcRF1-ORF24-CSTREP
64	BcRF1-ORF24_2_F	gctggctAGCCTGAAGCATC	pCDNA4.TO-BcRF1-ORF24-CSTREP
65	BcRF1-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	pCDNA4.TO-BcRF1-ORF24-CSTREP
66	mu24-ORF24_1_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	pCDNA4.TO-mu24-ORF24-CSTREP
67	mu24-ORF24_1_R	TTCAGGCTctcagtaagcttcaagtaattttt	pCDNA4.TO-mu24-ORF24-CSTREP
68	mu24-ORF24_2_F	tactgagAGCCTGAAGCATCTCT	pCDNA4.TO-mu24-ORF24-CSTREP
69	mu24-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	pCDNA4.TO-mu24-ORF24-CSTREP
70	UL87-ORF24_1_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	pCDNA4.TO-UL87-ORF24-CSTREP
71	UL87-ORF24_1_R	TTCAGGCTtccaaagagcagct	pCDNA4.TO-UL87-ORF24-CSTREP
72	UL87-ORF24_2_F	AGCCTGAAGCATCTCTCGTTTTCAAT	pCDNA4.TO-UL87-ORF24-CSTREP
73	UL87-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	pCDNA4.TO-UL87-ORF24-CSTREP

Table S2. Nucleotide sequence of synthetic gene blocks used in this study.

Geneblock Name	Sequence 5'-3'	Construct
10xCTD_repeat_pQLink_insert	TTATTTTCAGGGATCCTATAGCCCGACGTCGCCGAGTTATTCACCTACGTCCCCATCATACTCCCCCACGAGCCCTAGTTATTCGCCAACTTCCCCGAGCTATTCCCCAACATCACCCAGCTATAGTCCCACTTCACCCTCCTATTCACCTACGAGTCCATCTTATTCTCCAACCAGTCCTTCGTACTCACCCACGTCCCCATCGTATTCTCCTACTTCCCCTAGCTAGGCGGCCGCCTAGGACCC	pQLink-GST-10xCTD

Supplemental References

1. L. P. Kozlowski, IPC - Isoelectric Point Calculator. *Biol Direct* **11**, 55 (2016).