Structural basis of the complete poxvirus transcription initiation process

Utz Fischer (utz.fischer@biozentrum.uni-wuerzburg.de)
University of Würzburg https://orcid.org/0000-0002-1465-6591
Clemens Grimm
University of Würzburg
Julia Bartuli
University of Würzburg
Bettina Böttcher
University of Würzburg https://orcid.org/0000-0002-7962-4849
Aladar Szalay
University of Würzburg

Article

Keywords: poxviruses, transcription, virus-encoded multi-subunit polymerase (vRNAP)

Posted Date: April 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-402920/v1

Version of Record: A version of this preprint was published at Nature Structural & Molecular Biology on September 23rd, 2021. See the published version at https://doi.org/10.1038/s41594-021-00655-w.

Abstract

Poxviruses express their genes in the cytoplasm of infected cells using a virus-encoded multi-subunit polymerase (vRNAP) and unique transcription factors. We present cryo-EM structures that uncover the complete transcription initiation of the poxvirus vaccinia. In the pre-initiation complex, the heterodimeric early transcription factor VETFs/I adopts an arc-like shape spanning the polymerase cleft and anchoring upstream and downstream promoter elements. VETFI emerges as a TBP-like protein that inserts asymmetrically into the DNA major groove, triggers DNA melting, ensures promoter recognition and enforces transcription directionality. The helicase VETFs fosters promoter melting and the phospho-peptide domain (PPD) of vRNAP subunit Rpo30 enables transcription initiation. An unprecedented upstream promoter scrunching mechanism assisted by the helicase NPH-I likely fosters promoter escape and transition into elongation. Our structures shed light on unique mechanisms of poxviral gene expression and aid the understanding of thus far unexplained universal principles in transcription.

Introduction

Transcription by DNA-dependent RNA polymerases (RNAPs) is the first step in the expression of the genome in all forms of life. Eukarya use three structurally related nuclear multi-subunit RNAPs (Pol I, Pol II and Pol III) to transcribe distinct subset of genes [1]. They cooperate with distinct sets of transcription factors (TFs) that enable recruitment to specific classes of promoters. As an exception, only the TATA-box binding protein (TBP) is common to all three polymerases and has hence been termed the 'universal transcription factor'[2].

The mechanism of Pol II transcription, which accounts for the production of cellular mRNA has been studied in great detail. Transcription starts with the formation of the pre-initiation complex (PIC), which melts the promoter, enforces exact positioning of the transcription machinery, and defines the template strand. In the subsequent initial transcription phase, the transcript length of 5-7 nucleotides marks a decision point, beyond which Pol II either transitions into processive RNA synthesis[3,4] or aborts transcription. Passing of the decision point is referred to as promoter escape[4,5] and aided by an energy-loaded transition state in which melted downstream DNA is 'scrunched'[5,6] into the polymerase. Scrunching is an essential mechanism for start site selection and promoter escape which appears to be conserved from phages[7,8] over bacteria[9,10] to eukaryotes[5,6,11].

Most DNA viruses make use of the Pol II transcription machinery of the host to express their genome. A remarkable exception are poxviruses, which cause smallpox in humans and various zoonoses[12-14]. They multiply exclusively in the cytoplasm of infected cells and thus depend on their own set of factors ensuring gene expression and replication. Studies on the prototypical vaccinia virus identified a virus-encoded multi-subunit RNA polymerase (vRNAP) associated with factors that ensure the production of polyadenylated and m^7G-capped mRNA[15-19]. More recently, cryo-EM structures gave insight into the atomic details of vRNAP complexes and their mechanisms of transcription elongation and transcription-

coupled capping[20,21]. The core vRNAP enzyme, consisting of subunits Rpo147, Rpo132, Rpo35, Rpo30, Rpo22, Rpo19, Rpo18 and Rpo7, is evolutionarily related to eukaryotic RNAPs. Strong idiosyncrasies, however, are apparent for viral factors that are associated with vRNAP during mRNA production[22-25]. How these factors relate to TFs of nuclear RNAPs is largely unknown. *In vivo*, the core vRNAP associates with five additional virus-encoded proteins and one host factor: transcription factor Rap94 with partial homology to TFIIB[3,26,27], the capping enzyme D1/D12[28], the helicase NPH-I[29], the core protein E11, the early transcription factor VETF consisting of subunits VETFl and VETFs and cellular $tRNA^{Gln}$. This unit, termed complete vRNAP, is necessary and sufficient to bind early promoters, which consist of a single, A/T-rich consensus sequence, termed the critical region (CR) located upstream of the transcription start site[30]. Upon early promotor binding, the complete vRNAP enables the entire transcription process, including initiation, elongation and termination.

Here, we have used complete vRNAP to reconstitute a range of complexes that represent vaccinia initial transcription states from the pre-initiation phase to promoter escape. The structural analysis of these complexes uncovers principles of promoter recognition and -melting, template strand capture and an unprecedented helicase-assisted upstream DNA scrunching mechanism prior to promoter escape. Our study hence not only unravels the mechanisms of poxviral gene expression but also helps to understand thus far unexplained universal principles of transcription.

Results

Reconstitution of vaccinia transcription initiation complexes

We used complete vRNAP purified from HeLa cells infected with an engineered vaccinia strain to reconstitute FLAG-tagged[21] vRNAP pre-initiation and initially transcribing vaccinia vRNAP complexes. Transcriptionally active complete vRNAP was used to reconstitute transcription complexes on a vaccinia early promoter scaffold consisting of the critical region (CR), a non-complementary bubble including the transcription start site (+1) and a template cassette lacking G nucleotides (Fig. 1a). Complex formation was observed in the presence of NTPs upon incubation of complete vRNAP (Fig. 1b). A large-scale reconstitution of DNA-bound vRNAP complexes was separated by gradient centrifugation (Supplemental Fig. 1a), transferred onto holey carbon grids and three cryo-EM datasets were collected (Supplemental Figs. 1b,c). After extensive 3D classification, several distinctive particle classes could be separated (Supplemental Fig. 1d, 3a and 4a) that represented vRNAP complexes in transcription stages ranging from pre-initiation to co-transcriptional capping.

Cryo-EM structure of the vaccinia pre-initiation complex

Biochemical studies had previously shown that vRNAP, VETF and Rap94 are required for early transcription initiation[31,32,26]. We identified one particle class in our reconstitution that contained these factors along with the DNA scaffold and thus represented the *bona fide* PIC[31,32]. Its single-particle reconstruction displayed an overall resolution of 3.0 Å and diffuse density for DNA and VETF. Signal

subtraction and focused refinement resolved the VETF-DNA subcomplex (Supplemental Fig. 1e-i, Supplemental Tab. 1). The density was docked with the core vRNAP model and the VETFl and VETFs chains and parts of Rap94 were traced *de novo*, allowing complete modelling of the PIC (Fig. 1c and d). Within the PIC, the promoter is positioned at the distal edge of the polymerase cleft. The upstream DNA contacts the protrusion domain of the polymerase subunit Rpo132, directly adjacent to the C-terminal domain (CTD) of Rap94. The downstream promoter region interacts with the vRNAP core through positions on the clamp head (Supplemental Fig. 2a). The melted promoter region is predominantly disordered but could be visualized with mild Gaussian filtering (Fig. 1e). It localizes centrally above the opening of the cleft forming a second contact zone with the clamp head (Supplemental Fig. 2a). Both DNA strands appear only minimally separated within the bubble region. The latter joins the adjacent double-helical upstream and downstream sections in a 100° angle accompanied by a 25 Å translational shift of the helix axes (Fig. 1e). We thus conclude that the DNA is in the initially melted state.

Of note, neither the B-cyclin, nor the B-homology region of the early transcription factor Rap94 establish direct DNA contacts in the PIC (Fig. 1d and Supplemental Fig. 2a). However, on the opposite side of the core vRNAP, VETFs and VETFl contact the DNA in the distal upstream and downstream promoter regions, respectively (Fig. 1f, Supplemental Figs. 2b,c). Therefore, and due to the absence of contacts in the initially melted region (IMR), the VETF heterodimer appears to be anchored like a bridge on the upstream and downstream region of the promoter (Fig. 1d and Supplemental Fig. 2b).

Unique mode of DNA-binding by the VETF heterodimer

The structure of VETF allowed us to decipher the mechanisms of vRNAP recruitment to the early promoter. VETFl folds into five distinct domains, termed NTD, TBPLD, CRBD, Domain 4 and CTD (Figs. 1c,f). Despite the absence of *a priori* detectable sequence homology, the second domain displays a bi-lobal TBP fold, and hence is a TBP-like domain (TBPLD). It is located centrally above the polymerase cleft and, unlike TBP in other structures of PICs, contacts the promoter in a sequence-independent manner. Sequence-specific DNA binding in the vaccinia PIC is instead facilitated by the neighbouring domain, which recognizes the CR (Figs. 1a,d and f). Based on its fold and binding mode, this module constitutes a novel type of double-stranded DNA binding domain, hence termed Critical Region Binding Domain (CRBD). While holding only a limited content of secondary structure elements, it gains structural rigidity through three disulphide bridges that position a 3_{10}-helix ideally for its insertion into the major groove of the DNA (Fig. 2a). The sidechain-to-base contacts of this helix are the major site for sequence-specific readout of the promoter sequence (Figs. 2b,c). Only weak bending of the DNA helix axis is introduced in this region (Fig. 2a).

The joint structural context of TBPLD and CRBD in VETFl establishes specific contacts to the upstream promoter (Supplemental Fig. 2b). On the core vRNAP this part of the promoter is anchored via the interaction of domain 2 of Rap94 with the NTD of VETFl (Fig. 1d). All other domains of VETFl (NTD, Domain 4 and CTD) contribute to the structural backbone of VETF. Domain 4 and the CTD of VETFl make up the interface to VETFs (Fig. 1f).

The downstream promoter interacts almost exclusively with VETFs (Figs. 1D, 2D, and S2B). Only one additional pointed contact to the core vRNAP is established by the clamp head close to the TSS (Supplemental Fig. 2a). We observe a striking similarity of the first two domains of VETFs with the canonical helicase fold of chromatin remodelling SNF2-type ATPases [22,33], of which INO80 [34] is the closest homologue. With the latter, VETFs shares, along with the vRNAP-associated transcription factor NPH-I, an extended brace helix that stably bridges the N- and the C-lobe of the helicase fold (Supplemental Fig. 6). The intense DNA interaction of the VETFs helicase module is accompanied by a strong bend of the helix (Supplemental Fig. 2e). At the point of inflection, Phe271 intercalates via the minor groove, effectively disturbing the planar base-stacking over the range of roughly 3 base pairs on either side of the insertion site (Fig. 2d). Although melting of the two DNA strands is not observed at this position, this mechanism bears some similarity to the 'scalpel' method of strand-separating helicases [35] .

Promoter positioning and enforcement of transcription directionality in the PIC

We next asked how the DNA contacts established by the CRBD of VETFl control the initiation process. The 3_{10}-helix of CRBD inserts into the major groove, making it the reader head of VETF (hence termed the CRBD reader, Figs. 2a and b). The CR is essentially a consensus sequence of 15 A nucleotides, interrupted by a TG dinucleotide[30,36] (Figs. 1a and 2c). Arg370 and Gln375 engage in base-specific H-bonding that involves the bases of the TG motif on the non-template strand and the complementary AC dinucleotide on the opposing template strand (Figs. 2b and 2c, Supplemental Video 1). By this means, VETFl anchors the promoter in a defined position relative to the polymerase cleft. The CR displays a high propensity for A nucleotides downstream of the TG motif (Figs. 1a and 2c). Consistent with this, the C5 methyl groups of the corresponding complementary T nucleotides at positions -18 and -17 of the template strand interact cooperatively with the reader head by stacking with Tyr376. Inverse promoter binding would imply an unfavourable contact of Tyr376 with adenine bases (Figs. 2b and 2c) and thus a single promoter direction is coerced. By this means, the CRBD-DNA interaction ensures the i) identification of the CR, ii.) alignment of the CR relative to the polymerase cleft, and iii.) enforcement of transcription directionality. The CRBD is thus is the main regulator of the transcription initiation process.

Asymmetric DNA binding by the TBP-like domain of VETFl

Our structure identified VETFl as a TBP-like protein (TBPLP) whose TBPLD is engaged in an intricate contact network comprising the neighbouring domains of VETFl, VETFs and Rap94 (Fig. 3a). Members of the TBPLD family had previously been identified solely by means of sequence homology. However, VETFl stands apart from previously known TBPLPs because of its extremely divergent sequence that until now had prevented its classification as such. Nevertheless, the structural conservation of the TBPLD is comparably high, resulting in a Z-score of 4.2 determined by PDBeFold [37] when matching it to PDB entry 1TBP. To compare their structures and binding modes, we aligned the TBPLD - upstream DNA module of VETFl (Fig. 3b) with the yeast TBP - TATA-box crystal structure (Fig. 3c). The TBPLD of VETFl features the characteristic saddle structure that was previously described for TBP[38-41], however, the symmetry that is evolutionary conserved in TBP[42,43] appears broken. As a consequence, and unlike TBP, which contacts

the TATA-box symmetrically, VETFl binds the promoter asymmetrically and sequence-independently solely through its C-terminal TBP lobe. Most strikingly, the TBPLD inserts into the DNA major groove, contrary to the canonical binding mode of TBP, which is based on minor groove insertion. In accordance with this observation, the two strictly conserved pairs of DNA-intercalating phenylalanine residues on each lobe of TBP are absent in the TBPLD[38-41]. Still, the TBPLD induces a pronounced DNA bend via intercalation of aliphatic, rather than aromatic, sidechains (Fig. 3b). In agreement with the fundamentally different binding mode of the TBPLD, a consensus TATA-box is absent from vaccinia early promoters[30].

Rearrangement of the complete vRNAP into the PIC

Complete vRNAP is the predominant vRNAP complex found in infected cells and necessary and sufficient to execute viral early transcription. Hence, we previously speculated that this unit becomes incorporated into virions as a pre-assembled unit to promote the restart of transcription in the next infection cycle[21]. To investigate the transformation of complete vRNAP into the PIC, we compared both structures and their cryo-EM reconstructions. The VETF heterodimer is already present in the complete vRNAP, yet defined density could only be observed for the CRBD of VETFl whereas the remaining parts were mobile. Under the assumption that the adjacent TBPLD is flexibly joined to the CRBD, we were able to dock the diffuse residual density in the vRNAP reconstruction with the VETFl coordinates extracted from the PIC model, resulting in reasonable overlap. In the resulting structure (Fig. 1g) VETFl displays a flexible interface to $tRNA^{Gln}$. A comparison with the PIC structure reveals major reconfigurations, including the release of all associated factors from complete vRNAP except for the VETF heterodimer and Rap94 (Supplemental Video 1). This underlines the importance of complete vRNAP as a pre-formed early transcription unit and the high plasticity of vaccinia transcriptional complexes (see also Supplemental Video 1 for a summary of core aspects of the PIC).

Structure of the late pre-initiation complex

The structural transition described above explains how complete vRNAP becomes recruited to the viral early promoter to form the PIC. We next solved the structure of vRNAP particle classes that represent *bona fide* transcription stages following the pre-initiation phase. Based on biochemical evidence, such particles are predicted to be devoid of VETF but contain Rap94. Particles of class 1, subclass 2 (Supplemental Fig. 3a), which yielded a reconstruction at 3.0 Å resolution (Supplemental Figs. 1b-d, Supplemental Tab. 2) fulfilled this criterion. The density could be docked with the complete vRNAP model[21]. Disordered density corresponding to DNA is visible upstream next to the Rap94 CTD and within the downstream DNA channel. These sites roughly coincide with the DNA anchor points on the core vRNAP observed in the PIC (compare Fig. 1d). However, no density for the DNA transcription bubble or nascent RNA was detected in the active cleft (Fig. 4a). Instead, we found well-defined density for the highly phosphorylated stretch within the C-terminus of Rpo30 (termed phospho-peptide domain, PPD, Fig. 4b). It is in a similar conformation as in the complete vRNAP[21] and follows the path of the template- and non-template strand in the elongation complex (EC). This allows its pairing with the B-reader of Rap94 (Figs. 4a,b) and enables single-strand capture at later stages (see below). We therefore conclude that this

particle represents a late state of the PIC (lPIC) in which VETF has been expelled, the melted promoter has been handed over to the core vRNAP, but transcription has not yet been initiated.

PPD assisted single-strand capture and formation of the ITC

Next, we investigated the structural basis of lPIC conversion into an initially transcribing complex (ITC). Three vRNAP particle classes yielded reconstructions that were identified as different conformations of the ITC based on their composition and promotor positioning (lTC1-3, Supplemental Figs. 3a-d, Supplemental Table 3). The exact location of the polymerase on the promoter could be determined because its downstream blunt end was readily visible in the density (Supplemental Fig. 5a). In contrast to the lPIC, we observed ordered density for DNA in the downstream DNA channel and for a DNA/RNA hybrid above the active site (Fig. 5a). The PPD of Rpo30, which occupied the position of the DNA/RNA hybrid in the lPIC has been displaced by the template strand. Consequently, the B-homology region became mobile and is not visible in the density (Fig. 5b). No density for upstream DNA was identified. The three ITC complexes superimposed well but differed in the positioning of the DNA within the downstream DNA channel (Fig. 5a) and the state of the clamp (Fig. 7b). For ITC3, downstream DNA density was located in a shallower position and was comparably less ordered. In the ITC1 particle, the clamp is in a closed conformation with the DNA bound firmly and deep in the downstream DNA channel. ITC2 and ITC3 display an open clamp conformation and the downstream DNA appears mobile and bound in a shallower position. No significant differences between the three ITC complexes were discernible with regard to the DNA/RNA hybrid region. Thus, the three ITC structures inform on the conformational flexibility of the ITC, and, in concert with the lPIC structure, on the template-strand capture mechanism.

Upstream promoter scrunching in the late initially transcribing complex

During 3D classification, one particular class stood out because it comprised particles considerably larger than the ITC (Supplemental Fig. 4a). After a further round of focused classification of these particles on the observed extra density followed by multibody refinement a reconstruction was obtained that allowed the construction of a complete model (Figs. 6a and Supplemental Figure 4b-d, Tab. 4). This complex was classified as a late ITC (lITC), based on the positions of the blunt ends of the upstream and downstream promoter-DNA segments that are visible in the density (Supplemental Fig. 5b) and on the presence of a RNA/DNA hybrid. Except for Rap94, the core vRNAP was in a conformation similar to that observed in the ITC complexes. The path of the downstream DNA fitted best that observed in the ITC3 particle, indicating loose binding. The downstream blunt end of the DNA scaffold had advanced roughly 5 base pairs in downstream direction compared to the ITC (Supplemental Fig. 5a). Massive extra density above the cleft was unambiguously attributed to upstream DNA-bound NPH-I, and the NTD Rap94 and B-cyclin domain of Rap94 (Fig. 6a). Strikingly, the Rap94 B-homology region, the NTD and adjacent linkers appeared entirely reconfigured in comparison to other vRNAP complexes (Supplemental Figs. 5d, e) and the whole path of the Rap94 chain was visible (Fig. 6b). We also note that the path of the upstream DNA in the lITC is fundamentally different from that observed in the vaccinia PIC and in the ITC of Pol II [44].

The blunt ends of the DNA promoter scaffold are visible in the EM density of the IITC (Supplemental Fig. 5b), thus allowing to determine the position of vRNAP relative to, and the size of, the transcription bubble (Supplemental Fig. 5c). Strikingly, the upstream end of the scaffold can only be accommodated within the IITC under the assumption of massive promoter scrunching. This includes 13 base-pairs upstream of the artificial non-complementary region of the promoter scaffold, that have been additionally melted when compared to the ITC (Supplemental Fig. 5c). It is likely that this condition enables promoter escape and hence contributes to the transition of the initiation phase into productive elongation (Supplemental Video 2).

Discussion

In this study, we describe six vRNAP structures that represent snapshots of the poxviral transcription initiation phase. When viewed together, a comprehensive mechanistical picture of the early events during vaccinia transcription emerges.

The structure of the vaccinia PIC in the initially melted state provides insight into poxvirus early promoter identification and binding. The arc-shaped VETF heterodimer spans the polymerase cleft a upstream and downstream promoter elements and thus allows precise insertion of the TBPLD at the site of initial melting. The upstream contact to the promoter is established by VETFI, and its CRBD is the decisive element for its sequence-specific recognition. The CRBD recognizes the critical region of the promoter through a thus far unknown DNA-binding domain, which is stabilized by three disulphide bridges. Cystine formation in the CRBD may be introduced by vaccinia-encoded enzymes[45] as potential host factors for this task are confined to the endoplasmic reticulum. The TBPLD cooperates with the CRBD in upstream promoter binding and introduces a sharp DNA bend, which likely generates the nucleation site for promoter melting. Strikingly, the TBPLD of VETFI displays an asymmetric DNA binding mode. This sharply contrasts the canonical, symmetric DNA binding mode observed in all TBP-DNA complexes solved so far, including PIC complexes of the nuclear polymerases. Our findings could help the understanding of the dual nature of TBP[46] which, in its canonical binding mode, recognizes the TATA-box. Evidently, TBP is capable of an alternate, sequence-independent mode of action when directing the transcription machinery to TATA-less promoters[47,48]. Since vaccinia early promoters do not contain a TATA-box, an attractive explanation for the deviant binding mode of the TBPLD is that its orientation in the vaccinia PIC mirrors the alternate function of TBP at TATA-less promoters. Asymmetric DNA binding by TBP has been proposed in the context of Pol I [49] and may also occur in other TBPLDs[42,50,51]. Our structure might therefore contribute to the general understanding of this domain family.

Based on our data and prior findings for the Pol II system[52], we propose a mechanism for vaccinia early promoter melting (Fig. 3a): (i) The CRBD of VETFI binds the promoter at the CR, thereby enforcing directionality (Figs. 2a,b). (ii) VETFs pulls the DNA in an ATP-dependent reaction towards the vRNAP clamp and lobe, analogous to the XPB helicase in the Pol II system[53] (compare Supplemental Fig. 2g to Supplemental Fig. 2h). (iii) The clamp closes tightly around the DNA (Fig. 7b), thereby shaping its path[54].

(iv) The promoter DNA becomes underwound and bent by 80° towards the C-lobe of VETFs, exposing bases for an interaction with the latter (Fig. 2d). (v) The tip of the C-terminal lobe of the VETFl TBPLD intercalates upstream of the IMR, inducing a second sharp bend in the promoter (Fig. 3b). (vi) This bend triggers the initial melting event at the transcription start site, and the IMR absorbs the negative twist of the adjacent DNA segments.

A structure-based comparison of vaccinia and eukaryotic transcription systems reveals common principles but also obvious differences in the bound transcription factors. Similar positioning of the promoter relative to the core polymerase is observed in all PICs. Likewise, the positions of the B-homology region of Rap94 in the vaccinia PIC and the corresponding domain of TFIIB in the Pol II PIC overlap[44,55] (compare Supplemental Fig. 2g to Supplemental Fig. 2h). However, whereas TFIIB directly contacts the promoter, the B-homology region in Rap94 does not bind DNA (Fig. 1d). Furthermore, some features in the distal section of the DNA path appear to be conserved. A common principle might be the binding of a helicase transcription factor to the downstream promoter. It appears plausible that the helicase domains of VETFs (Supplemental Figs. 2e and 6) and of the TFIIH subunit XPB (SSl2 in yeast, Supplemental Fig. 2f) are functional counterparts[56].

In contrast to a recent study describing a PIC intermediate of Pol II immediately prior to the initially melted state[52], we do not observe underwinding of the DNA duplex in the vaccinia PIC. A possible explanation for this is that the IMR has absorbed a previous negative twist during the melting process. At the promoter upstream side, we noticed a topological relationship of the VETFl-promoter complex and the positioning of the Rap94 CTD with the TBP/TFIIF module on the DNA in the Pol-II PIC (Supplemental Fig. 2h). This notion is corroborated by the fact that despite their fundamentally different binding modes both, TBP and the VETFl TBPLD, induce a strong bend of the DNA. Thus, although the architecture of the vaccinia PIC differs fundamentally from its nuclear counterparts with respect to the involved transcription factors, basic architectural features are conserved. Our studies further reveal that VETF and Rap94 perform functionalities of TBP and TFIIB. The three conformationally different ITC structures mirror the flexibility of the transcription machinery in the initially transcribing phase and may coincide with non-processive RNA synthesis and TSS search as observed in the in Pol II system[57].

During transition to the lITC, a dramatic reorganisation of the transcription machinery takes place. This includes the recruitment of NPH-I, a re-routing of the upstream DNA path and widening of the transcription bubble, extending from promoter position +12 to -22 (Supplemental Fig. 5c). The only plausible explanation for the latter observation is that the NPH-I helicase motor has actively melted and scrunched upstream DNA duplex into the core vRNAP[58,59]. By this means, NPH-I likely assists promoter escape by adding the free energy of ATP hydrolysis to the generation of an energy-rich transcription intermediate[11]. While for Pol II, only downstream promoter scrunching has so far been observed[5,6,11], vRNAP employs a novel mechanism in which downstream and upstream promoter scrunching are combined. Strikingly, a mechanism in which the helicase transcription factor TFIIH injects free energy from ATP hydrolysis into the ITC during TSS scanning has been postulated for Pol II[10]. We thus reason

that the IITC might be the poxviral analogue of a hitherto undiscovered Pol II complex in which TFIIH fulfils a similar role to that of NPH-I in the IITC. In addition to its function as helicase motor, NPH-I plays an obvious role for the statics of the transcription bubble. Both, the 80° bend of the DNA (Supplemental Fig. 2e) and the insertion of the 'wedge' residue Phe273 (Fig. 2d) stabilize the upstream fork point of the transcription bubble in the IITC. We observed that the helicase Rad26 (CSB in humans) in the structure of yeast Rad26-bound Pol II occupies an equivalent position to NPH-I in the IITC, albeit in a different orientation (compare Fig. 6c to Fig. 6a). Yet, there are important differences between both helicases as the lack of a brace helix in Rad26 (Supplemental Fig. 6) and there is no indication that NPH-I is directly involved in DNA repair. Processive vRNAP elongation complexes can be assembled in absence of Rap94 *in vitro*[20]. *In vivo*, such complexes are found associated with the latter[60,61]. Thus, Rap94 may ensure the efficient recruitment of NPH-I to ECs stalled at pause sites to enable readthrough [59] and the resulting vRNAP complex might be structurally similar to the IITC (Fig. 6a).

After assignment of our structures to the transcription timeline we propose a comprehensive model of initial transcription (Fig. 7, Supplemental Video 2): Complete vRNAP reconfigures to the PIC (step 1). In the PIC, vRNAP-bound VETF has selected, aligned, positioned and melted the promoter DNA and the clamp is in a tight conformation (Fig. 7b). Upon handover of the melted promoter to the core polymerase, VETF leaves the complex, giving rise to the IPIC (step 2). Here, the promoter is supported upstream by the CTD of Rap94 and is anchored in the downstream DNA channel. The single-stranded DNA region is dynamic in this phase and therefore not visible (Fig. 4a). Through the interaction with the PPD of Rpo30 the B-homology domain of Rap94 is kept in an initiation-ready conformation. The template-strand capture goes along with the displacement of the PPD, which might be driven by the pronounced electronegative charge of DNA interacting with the positively charged active site region of vRNAP. After single strand capture (step 3), the B-reader scans the template strand for the TSS in an analogous manner as has been observed for Pol II [4]. Once the TSS is located, the B-homology domain becomes mobile and RNA synthesis commences (step 4). This phase is highly dynamic as documented by three ITC structures deviating in the state of the clamp (Fig. 7b) and positioning of the downstream DNA in the downstream DNA channel (Fig. 5a). The vRNAP promoter escape is accompanied by recruitment of NPH-I, a large-scale remodelling of Rap94, and major changes to the path of the upstream DNA (step 5). In the IITC complex (Fig. 6a), NPH-I acts as a strand-separating helicase, widens the transcription bubble, defines its upstream fork point, and shapes the path of the single-stranded template- and non-template DNA (Fig. 6a,d). Transition to a processive EC (step 6) triggers contraction of the transcription bubble, mobilization of the upstream DNA duplex and loss of NPH-I. Alternatively, abortive initiation might lead to re-initiation via re-recruitment of the Rpo30 PPD (step 6b). We note that all vRNAP complexes of the transcription initiation phase contain the core polymerase in a virtually constant conformation. Still, each transition of the transcription complexes is accompanied by changes of the clamp position (Fig. 7b).

Our study provides detailed mechanistical insights into the initial phase of poxvirus transcription. Some features observed in the presented structure are poxvirus-specific, for example the unique promoter recognition by the CRBD of VETFI. Others as the hitherto unknown behaviour of a TBP-like protein, the

observation of the initial melting event and the discovery of an ATP-dependent scrunching mechanism might be of relevance for the general understanding of multi-subunit RNAPs.

Methods

Generation of recombinant Vaccinia virus GLV-1h439 and vRNAP purification

GLV-1h439 was derived from GLV-1h68 as described previously [21]. For vRNAP purificationHela S3 cells were cultured in Dulbecco's modified Eagle Medium (DMEM), containing 10 % foetal bovine serum at 37° C in presence of 5 % CO_2. Cells were grown to 90 % confluence and infected with purified GLV-1h439 with a multiplicity of infection (MOI) of 1.2. After 24h, the infected cells were pelleted and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1,5 mM $MgCl_2$, 0.5 % [v/v] NP-40, 1 mM DTT), supplemented with complete EDTA-free protease inhibitor cocktail (Sigma-Aldrich). The soluble supernatant of the cellular extract was incubated for 3 h at 4 °C with anti-FLAG Agarose beads (Sigma-Aldrich). Beads were washed four times with buffer containing 50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.1 % [v/v] NP-40, 1 mM DTT, equilibrated with elution buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$ and 1 mM DTT) and eluted with a 200 µg/ml solution of 3xFLAG Peptide (Sigma-Aldrich). The eluate was analysed by SDS-PAGE. Approx. 50 µg of purified vRNAP was obtained from a single 15 cm Petri-dish of infected Hela S3 cells.

Reconstitution of promoter bound vRNAP complexes

A synthetic double stranded DNA oligonucleotide scaffold mimicking the vaccinia virus early promoter region was generated by annealing of two partially complementary DNA oligonucleotides (Fig. 1A). Annealing was performed in buffer containing 100 mM NaCl, 20 mM HEPES, pH 7.5, and 3 mM $MgCl_2$ by heating to 95 °C for 5 min followed by slow cooling to room temperature. The resulting double-stranded DNA oligonucleotide was precipitated with isopropanol and the pellet was resuspended in resuspension buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

For reconstitution of promoter-bound vRNAP complexes 1 pmol of [^{32}P]-labelled DNA promoter-scaffold was incubated for 30 min at 30 °C with the indicated amount of vRNAP in the presence of 1 mM of the indicated NTPs. Complexes were analysed by native gel electrophoresis (4 % acrylamide and 0.13 % bis-acrylamide, 25 mM Tris-HCl pH 7.4, 25 mM Boric acid and 0.5 mM EDTA) at 4 °C. For large-scale reconstitution of promoter/vRNAP complexes, purified vRNAP was concentrated in a Vivaspin 10kDa cut-off concentrator (Sartorius). A total of 400 µg of vRNAP was incubated with a 60 fold-molar excess of the DNA scaffold in reconstitution buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$ and 1 mM DTT) in the presence of ATP and UTP (1 mM each) for 30 min at 30 °C. The mixture was separated by 10 %-30 % sucrose gradient centrifugation (16h, 35.000 rpm, Beckman 60Ti rotor, 4 °C). Gradient fractions were collected manually and analysed by SDS-PAGE followed by silver- and ethidium bromide staining to visualize the proteins and DNA. The indicated fractions (Fig. S1A) were used for cryo-EM analysis after

buffer exchange against modified reconstitution buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$ and 1mM DTT) in a concentrator.

Cryo-EM and model building of the PIC

Following sucrose gradient purification, the indicated fractions (Supplemental Fig. 1) were pooled, diluted 1:50 with a buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$ and 1 mM DTT, and centrifuged in a Vivaspin concentrator to remove the sucrose. R 1.2/1.3 holey carbon grids (Quantifoil) were glow-discharged for 90 s (Plasma Cleaner model PDC-002. Harrick Plasma Ithaca, NY/USA) at medium power, and 3.5 µl of C2 sample was applied inside a Vitrobot Mark IV (FEI) at 4 °C and 100% relative humidity. Grids were blotted for 3 sec with blot force 5 and plunged into liquid ethane. Cryo-EM datasets comprising 10816 (dataset 1), 9878 (dataset 2), and 3640 (dataset 3) micrographs, respectively, were collected from three different grids with a Thermo Fisher Titan Krios G3 equipped with a Falcon III camera (Thermo-Fischer). Data were acquired with EPU at 300 keV and a primary magnification of 75,000 (calibrated pixel size 1.0635 Å) in movie-mode with 47 fractions per movie and counting of the electron signal. The total exposure was 77.5 e/Å² for 75 sec, with 2 exposures per hole.

Dose-weighted, motion-corrected sums of the micrograph movies were calculated with Motioncor2 [63]. The contrast-transfer function of each micrograph was fitted with Relion3.1 [64]. An initial set of 25,000 particles was picked with the Gaussian picker and subjected to three rounds of 2D classification in Relion[64] to clean up the dataset. Eight class averages were selected as templates for subsequent automated particle picking within Relion and a total of 300,000 particles were picked using the Relion autopicker. After a second round of 2D classification, 3D classification was performed using the vRNAP core structure as template. Particles belonging to the PIC were selected and 2D classes for autopicking were calculated. The resulting three particle stacks, one for each dataset, were cleaned up individually by four rounds of 2D classification each, and contained 1,064,795 (dataset 1), 1,205,746 (dataset 2), and 323,776 (dataset 3) good particles. Each particle stack was then subjected to 3D classification and particles that fell in the defined PIC class were selected. The PIC particle stacks of the three datasets were then united into a single stack, and CTF refinement, followed by a consensus 3D refinement, was performed. This united particle stack was then subjected to a focused 3D classification with a mask that selected for VETF and DNA. Two of the resulting three classes yielded high-resolution reconstructions of VETF and DNA in minimally divergent conformations (Supplemental Fig. 2c). The particles from the two good classes were then forwarded to a Multibody (MB) refinement in Relion, either pooled or separately. MB refinement was performed with two bodies, representing either VETF or DNA and core vRNAP. We noticed that minor variations of the mask pairs resulted in the improvement of particular regions of the reconstruction. We therefore repeated the MB refinement with 11 more mask pairs and combined the resulting maps with Phenix.combine_focused_maps to create a single, optimal map for refinement. To build the PIC model, the vRNAP core excluding the Rpo30 phospho-peptide domain (PPD) was extracted from the complete vRNAP structure (PDB 6RFL) and docked into the cryo-EM density map. Within the residual density, the path of the DNA was identified and manually docked with section-wise stretches of

ideal B-DNA. VETF was then traced *de novo in* COOT 0.9[65]. To this end, the SNF2 helicase core of VETFs was located and built, followed by well-defined regions of VETFl. The resulting partial model was initially refined with Phenix.real_space_refine and forwarded to Phenix.combine_focused_maps to create a stitched, optimal map. The VETF model was then completed manually. The full polypeptide chains of both, VETFs and VETFl, were continuously modelled. Finally, residual density was identified as the relocated Rap94 NTD, and the DNA sequence was assigned. The resulting model was manually optimized with the real-space-refinement routine of COOT 0.9 and subjected again to refinement with Phenix.real_space_refine [66] including ADP refinement steps. During refinement, secondary structure and mild Ramachandran restraints were imposed. After four further cycles of manual inspection and automated refinement, the refinement converged, and a model with excellent stereochemistry and good correlation with the cryo-EM map was obtained (Supplemental Tab. 1).

3D reconstruction and model building of lPIC and ITC complexes

The lPIC particle stack obtained as described above was subjected to two rounds of focused 3D classification with 3 classes in each of the two rounds. The classification was focused with a mask on the cleft, active site and downstream DNA channel as well as the region of the Rap94 cyclin domain. From the resulting set of nine class averages (Supplemental Fig. 3a) four reasonable reconstructions were obtained after a final round of 3D refinement and post-processing, and the associated complexes were identified as the lPIC, and ITC1-3 (Supplemental Fig. 3a). The resolution was determined by fourier-shell correlation (FSC) to 3.0Å for the lPIC and 2.9Å, 3.2Å and 3.0Å for ITC1, ITC2 and ITC3, respectively (Supplemental Fig. 3b-d). To build the lPIC model, the vRNAP core including the Rpo30 PPD was extracted from the complete vRNAP structure (PDB 6RFL) and docked into the cryo-EM density. The positioning of the Rap94 cyclin domain and the adjacent linker regions were adjusted manually with Coot[65] and the model was refined with Phenix.real_space_refine [66] including an ADP refinement step. During refinement secondary structure and Ramachandran restraints were imposed. After two further cycles of manual inspection and automated refinement, the refinement converged and a model with excellent stereochemistry and good correlation with the cryo-EM map was obtained (Suplemental Tabs. 2 and 3).

3D reconstruction and model building of the lITC

The lITC particle stack obtained as described above was subjected to a round of focused 3D classification with a mask on the NPH-I and upstream DNA region. From the three resulting classes, a single one displayed good occupancy and resolution for NPH-I. Particles belonging to this class were subjected to a two-body multibody refinement (MB) in Relion using a mask for NPH-I and upstream DNA and a mask for the core vRNAP. The postprocessed reconstructions for both bodies were then combined with Phenix.combine_focused_maps. To build the lITC model, the ITC1 structure was docked into the density. Within the residual density a characteristic SNF2 helicase fold was recognized that was docked with either VETFs or NPH-I from the complete vRNAP structure (PDB 6RFL). NPH-I unequivocally fitted the density while VETFs did not. Further residual density could then be identified as the relocated Rap94 B

cyclin domain the relocated Rap94 NTD and the NPH-I CTD. After manual adjustments with Coot including rebuilding of remodeled Rap94 linker regions the model was refined with Phenix.real_space_refine including an ADP refinement step. During refinement secondary structure and Ramachandran restraints were imposed. After two further cycles of manual inspection and automated refinement, the refinement converged and a model with excellent stereochemistry and good correlation with the cryo-EM map was obtained (Table S4).

3D reconstruction and model building of lPIC and ITC complexes

The lPIC particle stack obtained as described above was subjected to two rounds of focused 3D classification with 3 classes in each of the two rounds. The classification was focused with a mask on the cleft, active site and downstream DNA channel as well as the region of the Rap94 cyclin domain. From the resulting set of nine class averages (Supplemental Fig. 4a) four reasonable reconstructions were obtained after a final round of 3D refinement and post-processing, and the associated complexes were identified as the lPIC, and ITC1-3 (Supplemental Fig. S4b-d). The resolution was determined by fourier-shell correlation (FSC) to 3.0Å for the lPIC and 2.9Å, 3.2Å and 3.0Å for ITC1, ITC2 and ITC3, respectively. To build the lPIC model, the vRNAP core including the Rpo30 PPD was extracted from the complete vRNAP structure (PDB 6RFL) and docked into the cryo-EM density. The positioning of the Rap94 cyclin domain and the adjacent linker regions were adjusted manually with Coot[65] and the model was refined with Phenix.real_space_refine[66] including an ADP refinement step. During refinement secondary structure and Ramachandran restraints were imposed. After two further cycles of manual inspection and automated refinement, the refinement converged and a model with excellent stereochemistry and good correlation with the cryo-EM map was obtained (Supplemental Tab. 4).

Cryo EM data deposition

Coordinate files for the vRNAP complex structures were deposited with the Protein Data Bank. The associated cryo-EM maps have been deposited with the Electron Microscopy Data Bank (EMDB). The respective entries are available under the following accession codes. PIC: PDB entry ID 7AMV and EMDB entry ID EMD-11824; lPIC: PDB entry ID 7AOF and EMDB entry ID EMD-11843; ITC1: PDB entry ID 7AOZ and EMDB entry ID EMD-11848; ITC2: PDB entry ID 7AP8 and EMDB entry ID EMD-11850; ITC3: PDB entry ID 7AP9 and EMDB entry ID EMD-11851; IITC: PDB entry ID 7AOH and EMDB entry ID EMD-11844.

Declarations

Acknowledgements

We thank Christian Kraft for his help during cryo-EM data collection. Cryo-EM of the complete vRNAP was carried out in the cryo-EM facility of the University of Würzburg (DFG - INST 93/903-1 FUGG). This work was supported by the DFG (Fi 573/21-1) and the Hope Realized Medical Foundation.

Author Contributions

J.B. and C.G. designed transcription scaffolds and acquired cryo-EM data. B.B assisted during cryo-EM data collection. J.B. performed functional vRNAP and scaffold binding assays, purified the vRNAP complexes and prepared cryo-EM samples. C.G. processed the cryo-EM data, built and refined the models, analysed the models, and prepared Figs. and the videos. C.G. and U.F. wrote the manuscript. U.F. designed and supervised the project, A.A.S. and U.F. acquired funding.

Declaration of Interests

We have no competing interests to declare.

References

1. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85-98 (2011).
2. Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev* **7**, 1291-308 (1993).
3. Liu, X., Bushnell, D.A., Wang, D., Calero, G. & Kornberg, R.D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206-9 (2010).
4. Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-30 (2009).
5. Revyakin, A., Liu, C., Ebright, R.H. & Strick, T.R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139-43 (2006).
6. Kapanidis, A.N. et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144-7 (2006).
7. Brieba, L.G. & Sousa, R. T7 promoter release mediated by DNA scrunching. *EMBO J* **20**, 6826-35 (2001).
8. Cheetham, G.M. & Steitz, T.A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305-9 (1999).
9. Henderson, K.L. et al. Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. *Proc Natl Acad Sci U S A* **114**, E3032-E3040 (2017).
10. Winkelman, J.T. et al. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* **351**, 1090-3 (2016).
11. Roberts, J.W. Biochemistry. RNA polymerase, a scrunching machine. *Science* **314**, 1097-8 (2006).
12. Shchelkunov, S.N. An increasing danger of zoonotic orthopoxvirus infections. *PLoS Pathog* **9**, e1003756 (2013).
13. Lewis-Jones, S. Zoonotic poxvirus infections in humans. *Curr Opin Infect Dis* **17**, 81-9 (2004).
14. Grant, R., Nguyen, L.L. & Breban, R. Modelling human-to-human transmission of monkeypox. *Bull World Health Organ* **98**, 638-640 (2020).

15. Rohrmann, G., Yuen, L. & Moss, B. Transcription of vaccinia virus early genes by enzymes isolated from vaccinia virions terminates downstream of a regulatory sequence. *Cell* **46**, 1029-35 (1986).
16. Broyles, S.S. & Moss, B. Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. *Proc Natl Acad Sci U S A* **83**, 3141-5 (1986).
17. Ensinger, M.J., Martin, S.A., Paoletti, E. & Moss, B. Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. *Proc Natl Acad Sci U S A* **72**, 2525-9 (1975).
18. Broyles, S.S., Yuen, L., Shuman, S. & Moss, B. Purification of a factor required for transcription of vaccinia virus early genes. *J Biol Chem* **263**, 10754-60 (1988).
19. Yuen, L., Davison, A.J. & Moss, B. Early promoter-binding factor from vaccinia virions. *Proc Natl Acad Sci U S A* **84**, 6069-73 (1987).
20. Hillen, H.S. et al. Structural Basis of Poxvirus Transcription: Transcribing and Capping Vaccinia Complexes. *Cell* **179**, 1525-1536 e12 (2019).
21. Grimm, C. et al. Structural Basis of Poxvirus Transcription: Vaccinia RNA Polymerase Complexes. *Cell* **179**, 1537-1550 e19 (2019).
22. Gershon, P.D. & Moss, B. Early transcription factor subunits are encoded by vaccinia virus late genes. *Proc Natl Acad Sci U S A* **87**, 4401-5 (1990).
23. Niles, E.G., Lee-Chen, G.J., Shuman, S., Moss, B. & Broyles, S.S. Vaccinia virus gene D12L encodes the small subunit of the viral mRNA capping enzyme. *Virology* **172**, 513-22 (1989).
24. Shuman, S., Broyles, S.S. & Moss, B. Purification and characterization of a transcription termination factor from vaccinia virions. *J Biol Chem* **262**, 12372-80 (1987).
25. Broyles, S.S. & Moss, B. Sedimentation of an RNA polymerase complex from vaccinia virus that specifically initiates and terminates transcription. *Mol Cell Biol* **7**, 7-14 (1987).
26. Ahn, B.Y., Gershon, P.D. & Moss, B. RNA polymerase-associated protein Rap94 confers promoter specificity for initiating transcription of vaccinia virus early stage genes. *J Biol Chem* **269**, 7552-7 (1994).
27. Ahn, B.Y. & Moss, B. RNA polymerase-associated transcription specificity factor encoded by vaccinia virus. *Proc Natl Acad Sci U S A* **89**, 3536-40 (1992).
28. Shuman, S. & Hurwitz, J. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme–guanylate intermediate. *Proc Natl Acad Sci U S A* **78**, 187-91 (1981).
29. Paoletti, E. & Moss, B. Two nucleic acid-dependent nucleoside triphosphate phosphohydrolases from vaccinia virus. Nucleotide substrate and polynucleotide cofactor specificities. *J Biol Chem* **249**, 3281-6 (1974).
30. Davison, A.J. & Moss, B. Structure of vaccinia virus early promoters. *J Mol Biol* **210**, 749-69 (1989).
31. Cassetti, M.A. & Moss, B. Interaction of the 82-kDa subunit of the vaccinia virus early transcription factor heterodimer with the promoter core sequence directs downstream DNA binding of the 70-kDa subunit. *Proc Natl Acad Sci U S A* **93**, 7540-5 (1996).

32. Baldick, C.J., Jr., Cassetti, M.C., Harris, N. & Moss, B. Ordered assembly of a functional preinitiation transcription complex, containing vaccinia virus early transcription factor and RNA polymerase, on an immobilized template. *J Virol* **68**, 6052-6 (1994).
33. Broyles, S.S. & Moss, B. DNA-dependent ATPase activity associated with vaccinia virus early transcription factor. *J Biol Chem* **263**, 10761-5 (1988).
34. Eustermann, S. et al. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* **556**, 386-390 (2018).
35. Kitano, K., Kim, S.Y. & Hakoshima, T. Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure* **18**, 177-87 (2010).
36. Yang, Z., Bruno, D.P., Martens, C.A., Porcella, S.F. & Moss, B. Genome-wide analysis of the 5' and 3' ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J Virol* **85**, 5897-909 (2011).
37. Krissinel, E. Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem* **1**, 76-85 (2012).
38. Nikolov, D.B. & Burley, S.K. 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nat Struct Biol* **1**, 621-37 (1994).
39. Kim, J.L. & Burley, S.K. 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat Struct Biol* **1**, 638-53 (1994).
40. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-20 (1993).
41. Kim, J.L., Nikolov, D.B. & Burley, S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-7 (1993).
42. Blombach, F. & Grohmann, D. Same same but different: The evolution of TBP in archaea and their eukaryotic offspring. *Transcription* **8**, 162-168 (2017).
43. Hobbs, N.K., Bondareva, A.A., Barnett, S., Capecchi, M.R. & Schmidt, E.E. Removing the vertebrate-specific TBP N terminus disrupts placental beta2m-dependent interactions with the maternal immune system. *Cell* **110**, 43-54 (2002).
44. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**, 129-43 (2015).
45. Senkevich, T.G., White, C.L., Koonin, E.V. & Moss, B. Complete pathway for protein disulfide bond formation encoded by poxviruses. *Proc Natl Acad Sci U S A* **99**, 6667-72 (2002).
46. Struhl, K. Duality of TBP, the universal transcription factor. *Science* **263**, 1103-4 (1994).
47. Kamenova, I., Warfield, L. & Hahn, S. Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. *Mol Cell Biol* **34**, 2929-43 (2014).
48. Zhao, X., Schramm, L., Hernandez, N. & Herr, W. A shared surface of TBP directs RNA polymerase II and III transcription via association with different TFIIB family members. *Mol Cell* **11**, 151-61 (2003).

49. Bric, A., Radebaugh, C.A. & Paule, M.R. Photocross-linking of the RNA polymerase I preinitiation and immediate postinitiation complexes: implications for promoter recruitment. *J Biol Chem* **279**, 31259-67 (2004).
50. Ravarani, C.N.J. et al. Molecular determinants underlying functional innovations of TBP and their impact on transcription initiation. *Nat Commun* **11**, 2384 (2020).
51. Kramm, K., Engel, C. & Grohmann, D. Transcription initiation factor TBP: old friend new questions. *Biochem Soc Trans* **47**, 411-423 (2019).
52. Dienemann, C., Schwalb, B., Schilbach, S. & Cramer, P. Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Mol Cell* **73**, 97-106 e4 (2019).
53. Murakami, K. et al. Uncoupling Promoter Opening from Start-Site Scanning. *Mol Cell* **59**, 133-8 (2015).
54. Feklistov, A. et al. RNA polymerase motions during promoter melting. *Science* **356**, 863-866 (2017).
55. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209 (2017).
56. Kolesnikova, O., Radu, L. & Poterszman, A. TFIIH: A multi-subunit complex at the cross-roads of transcription and DNA repair. *Adv Protein Chem Struct Biol* **115**, 21-67 (2019).
57. Cheung, A.C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431-7 (2012).
58. Hindman, R. & Gollnick, P. Nucleoside Triphosphate Phosphohydrolase I (NPH I) Functions as a 5' to 3' Translocase in Transcription Termination of Vaccinia Early Genes. *J Biol Chem* **291**, 14826-38 (2016).
59. Deng, L. & Shuman, S. Vaccinia NPH-I, a DExH-box ATPase, is the energy coupling factor for mRNA transcription termination. *Genes Dev* **12**, 538-46 (1998).
60. Yang, Z. & Moss, B. Interaction of the vaccinia virus RNA polymerase-associated 94-kilodalton protein with the early transcription factor. *J Virol* **83**, 12018-26 (2009).
61. Zhang, Y., Ahn, B.Y. & Moss, B. Targeting of a multicomponent transcription apparatus into assembling vaccinia virus particles requires RAP94, an RNA polymerase-associated protein. *J Virol* **68**, 1360-70 (1994).
62. Tan, S., Hunziker, Y., Sargent, D.F. & Richmond, T.J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127-51 (1996).
63. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
64. Kimanius, D., Forsberg, B.O., Scheres, S.H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**(2016).
65. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
66. Adams, P.D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94-106 (2011).

Figures

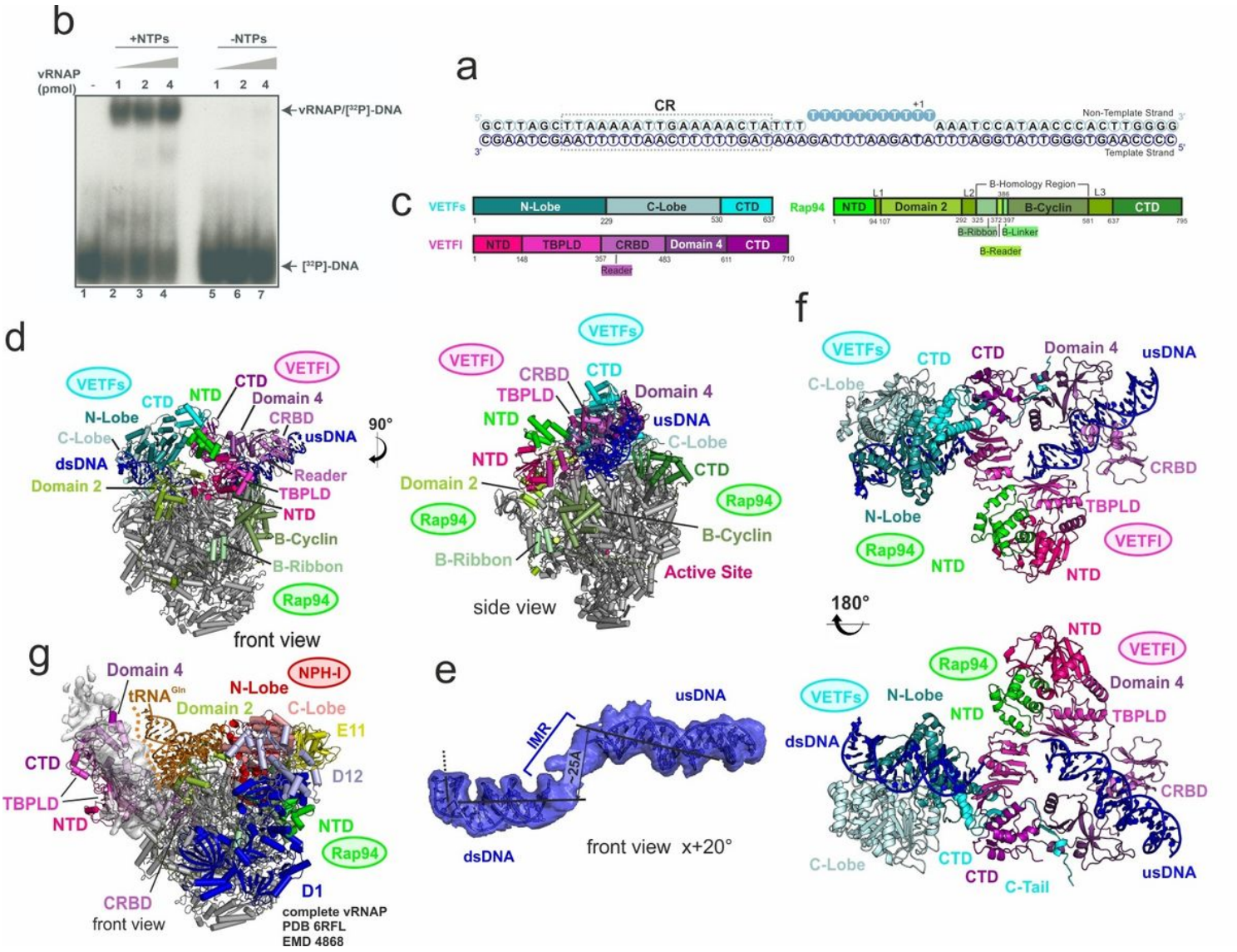

Figure 1

Structure of the vaccinia pre-initiation complex (PIC). (a) Synthetic DNA scaffold used for complex formation. (b) Band shift assay of complete vRNAP in the presence of radiolabelled scaffold as specified in (a). (c) domain structure of VETFs, VETFI and Rap94. (d) Overall structure of the PIC in two orthogonal views. The core polymerase is depicted in grey. (e) Transparent iso-surface of the DNA cryo-EM density, filtered by Gaussian blur with 1.5σ standard deviation, and DNA model are shown in cartoon style. Approximated helix axes of the different duplex DNA sections are indicated, and the translation of the helix axes of the two duplex DNA regions adjacent to the initially melted region (IMR) is denoted. This view is rotated by 20° relative to (b). (f) Detailed view of VETF and promoter DNA. Two views of VETF with the bound promoter within the PIC are displayed. For easier visualization, the core polymerase is hidden. (g) Complete vRNAP residual density (EMD 4868, grey transparent isosurface) docked with the VETFI structure and shown along with the complete vRNAP model (PDB 6RFL) in cartoon representation

(colour code as in Figure 1-3 and 21 for the complete vRNAP-specific factors). The predominantly disordered interface of VETFl to the tRNA aminoacyl stem is marked with an orange dotted line.

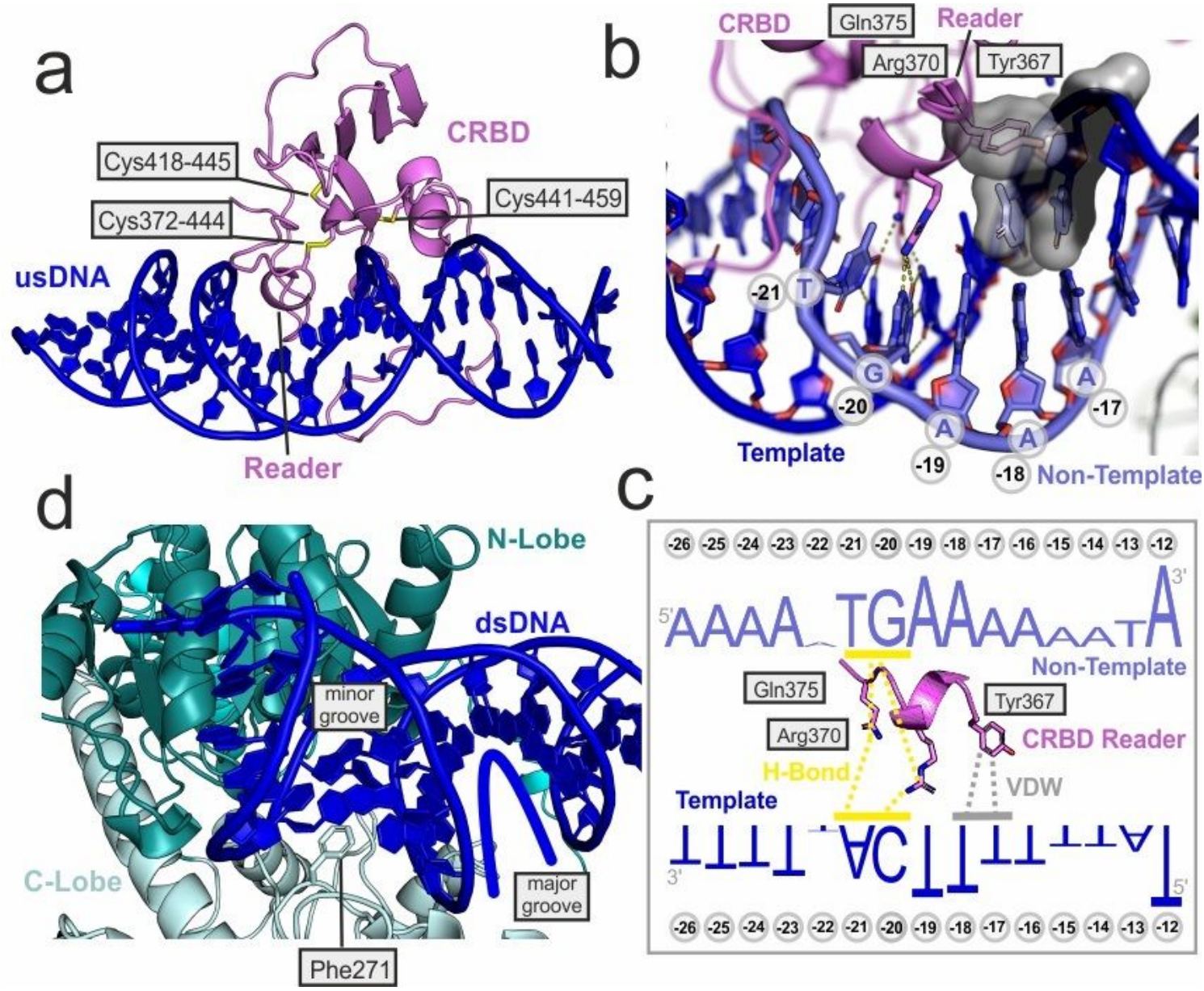

Figure 2

Structure of the VETF heterodimer. (a) VETFl CRBD binding to the upstream critical promoter region. Disulphide bridges are depicted as stick model. (b) Details of the VETFl CRBD promoter interaction. The model is depicted in stick representation, base pairs are numbered relative to the transcription start site (TSS). Only bases for the non-template strand are labelled, the template strand is sequence complementary. Contact between Tyr367 and thymidine bases at positions -18 and -17 are displayed as transparent van-der-Waals surface. The protein-DNA H-bond network is depicted as dotted yellow lines. (c) Schematic representation of the sequence-specific interactions of the CRBD reader. The critical region consensus sequence is depicted according to Yang et al.36. Van der Waals interactions (VDW) are indicated in grey, H-bonds in yellow. (d) Details of VETFs binding to the upstream promoter. The intercalating ‘wedge’ residue Phe271 is shown in stick representation.

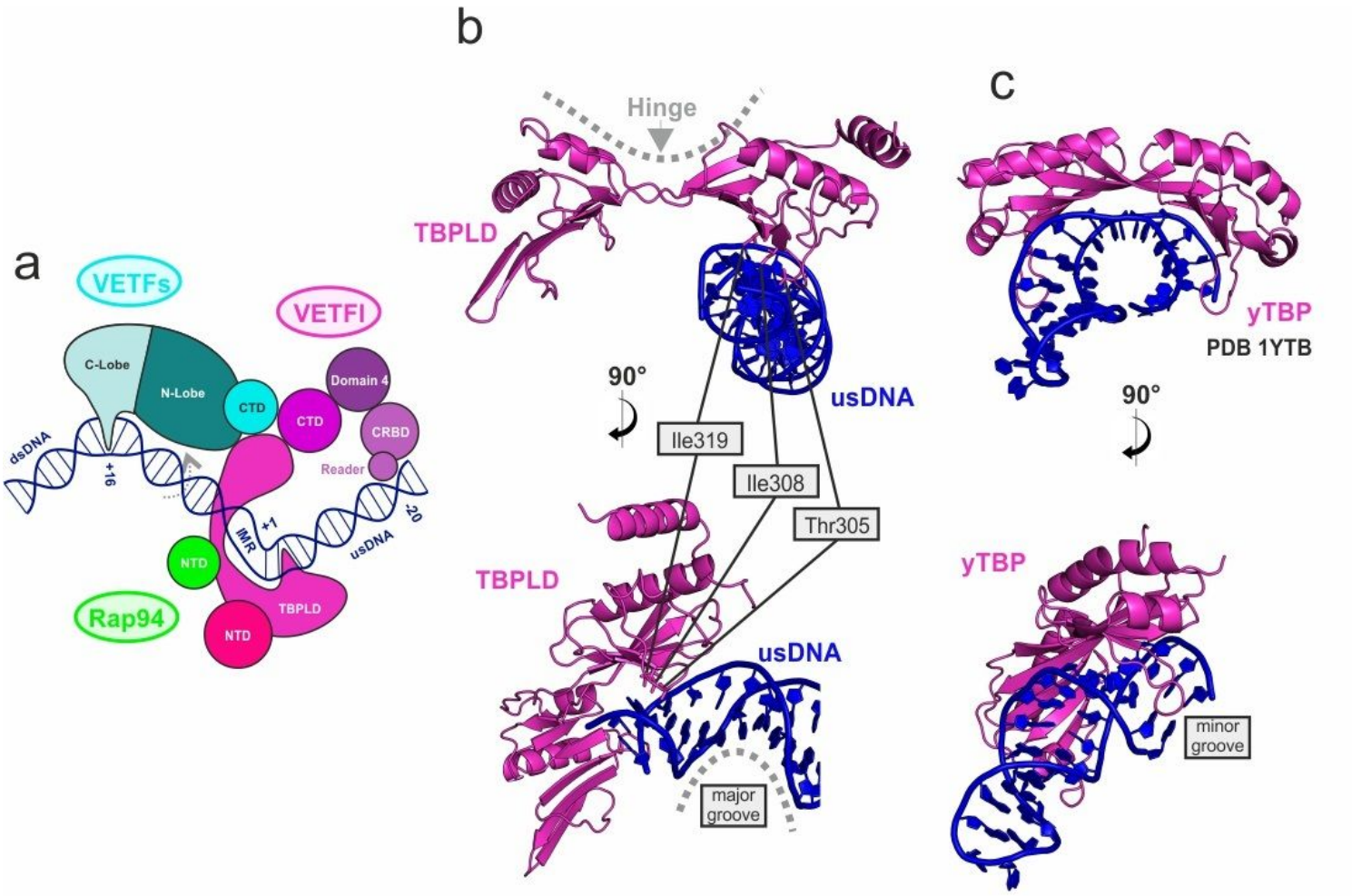

Figure 3

Comparison of the TBP-like domain from vaccinia VETFI with yeast TBP. (a) Schematic representation of the VETF-promoter interactions. (b) The TBPLD of VETFI in two orthogonal views. Residues intercalating between the nucleobases are depicted as stick model. (c) Structure of the yeast TBP protein bound to a synthetic TATA-box hairpin DNA oligomer62 (PDB 1YTB) in two orthogonal views corresponding to the protein orientation of the VETFI TBPLD as seen in (b).

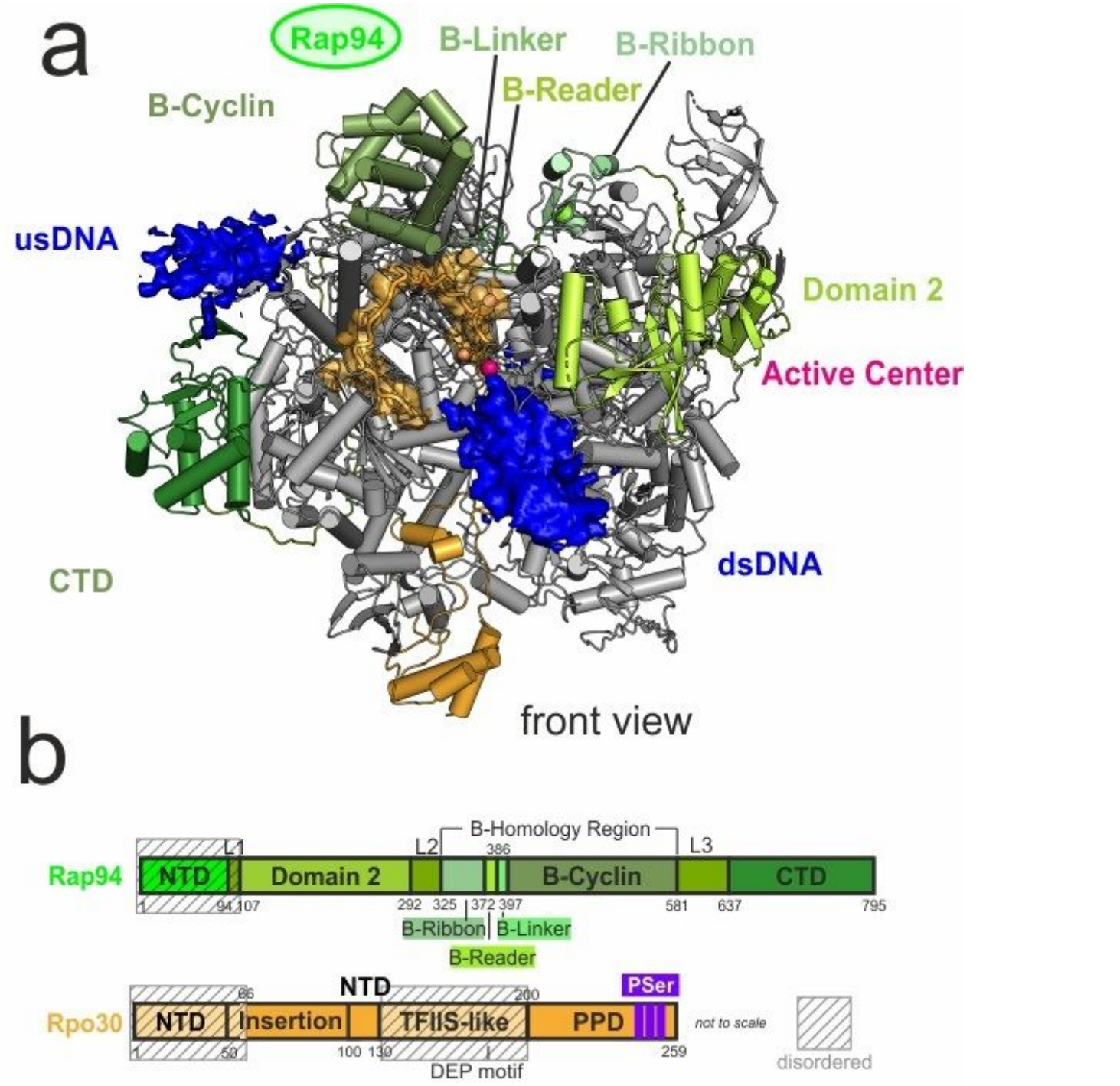

Figure 4

Structure of the late PIC. (a) Model of the lPIC with density for the bound DNA oligomer shown as a blue surface, for the phosphor-peptide domain (PPD) in transparent gold. (b) Domain structure of the bound transcription factors. Disordered regions are marked by hatched boxes.

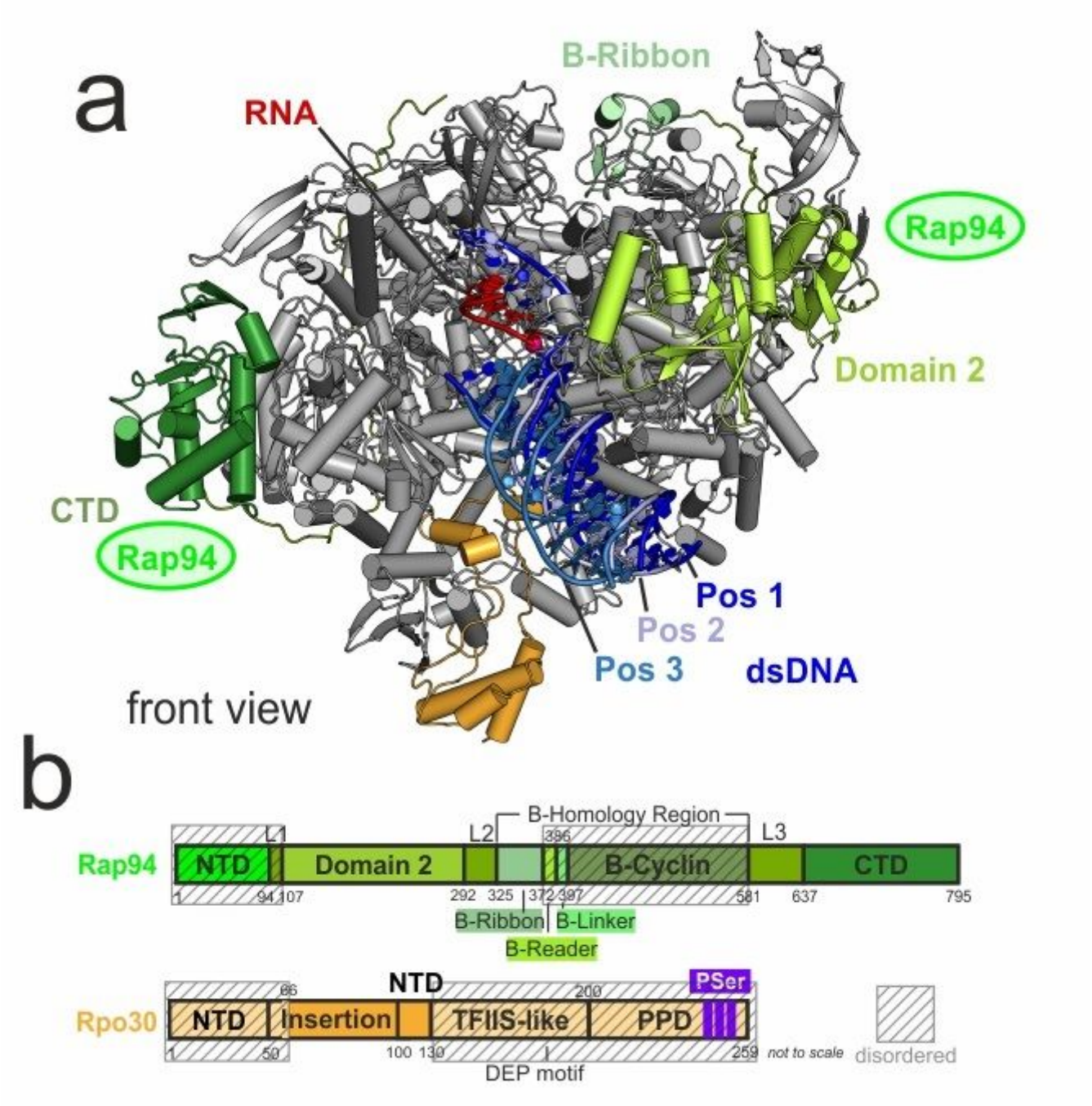

Figure 5

Three structures of initially transcribing complexes. (a) Model of the ITC state 1 shown with overlay of the downstream DNA from ITC conformation 2 and ITC conformation 3. Downstream DNA density for ITC conformation 3 was weak, therefore the DNA model at this position was omitted in the corresponding PDB entry. (b) Domain structure of the bound transcription factors. Disordered regions are marked by hatched boxes.

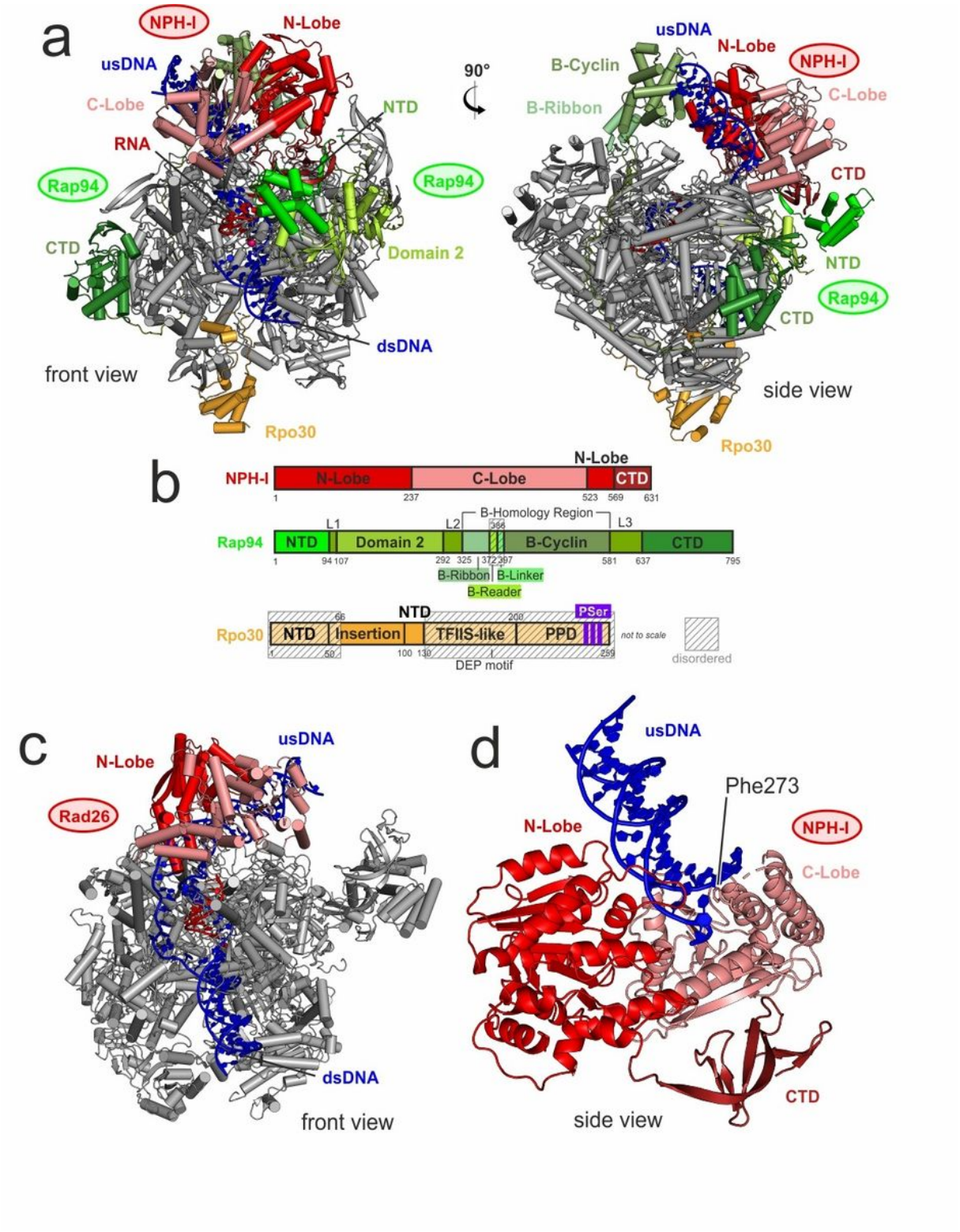

Figure 6

Structure of the late ITC. (a) Model of the IITC in two orthogonal views. (b) Domain structure of the bound transcription factors. Disordered regions are marked by hatched boxes. (c) Structure of the eukaryotic transcription-coupled repair (TCR) initiation complex, orientation as in a., left view. (d) Detailed view of NPH-I bound to the upstream promoter DNA. The ‘wedge’ residue Phe273 at the center of the strand-separating mechanism is indicated.

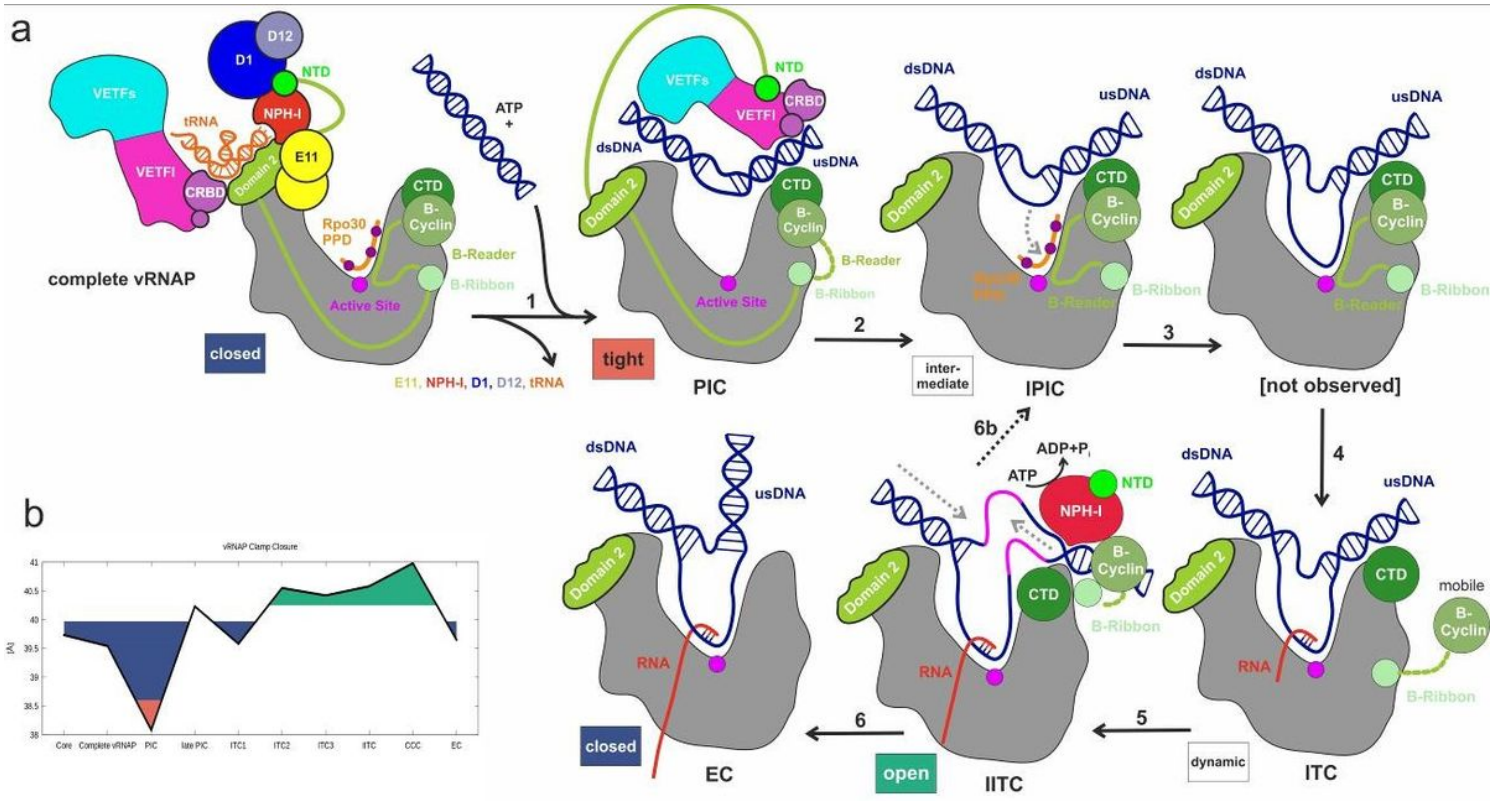

Figure 7

Transition of complete vRNAP to the PIC, and a model for early promoter recognition and opening: (a) Schematic representation of vaccinia early promoter recognition and opening mechanism (colour code as in Figure 1). (b) Plot of clamp closure versus transcription state.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialsNSMB.docx