Structural basis of the complete poxvirus transcription initiation process

Clemens Grimm^{1,4}✉, Julia Bartuli^{1,4}, Bettina Boettcher², Aladar A. Szalay^{1,3} and Utz Fischer^{1,3}✉

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 phase 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 probably 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.

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 subsets of genes¹. 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 messenger RNA (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 is 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 that 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. The poxviruses, which cause smallpox in humans and various zoonoses, are a remarkable exception^{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 m⁷G-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} and essential for targeting vRNAP to the virion²⁸, the capping enzyme D1/D12²⁹, the helicase NPH-I³⁰, the core protein E11, the early transcription factor VETF consisting of subunits VETFI 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 (TSS)³¹. Following early promoter 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 transcription initiation complexes. We used complete vRNAP purified from HeLa cells infected with an engineered vaccinia strain to reconstitute FLAG-tagged²¹ 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 CR, a non-complementary bubble including the TSS (+1) and a template cassette lacking G nucleotides (Fig. 1a). Complex formation was observed in the presence of nucleoside

¹Department of Biochemistry and Cancer Therapy Research Center (CTRC), Theodor Boveri-Institute, University of Würzburg, Würzburg, Germany.

²Department of Biochemistry, Rudolf-Virchow-Center, University of Würzburg, Würzburg, Germany. ³Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany. ⁴These authors contributed equally: Clemens Grimm, Julia Bartuli.

✉e-mail: clemens.grimm@biozentrum.uni-wuerzburg.de; utz.fischer@biozentrum.uni-wuerzburg.de

Fig. 1 | Structure of the vaccinia PIC. **a**, Synthetic DNA scaffold used for complex formation. **b**, Band shift assay (native gel electrophoresis) of complete vRNAP in the presence of the radiolabeled 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 gray as a solvent-accessible surface. usDNA, upstream DNA; dsDNA, downstream DNA. **e**, Transparent isosurface of the DNA cryo-EM density, filtered by Gaussian blur with 1.5σ standard deviation, and the DNA model, 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 IMR is denoted. This view is rotated by 20° relative to **d**. **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, gray transparent isosurface) docked with the VETFI structure and shown along with the complete vRNAP model (PDB 6RFL) in cartoon representation (color code as in Figs. 1–3 and ref. ²¹ for the complete vRNAP-specific factors). The core polymerase is depicted in gray as a solvent-accessible surface. The predominantly disordered interface of VETFI and the tRNA aminoacyl stem is marked with an orange dotted line.

triphosphates (NTPs) upon incubation of complete vRNAP (Fig. 1b). A large-scale reconstitution of DNA-bound vRNAP complexes was separated by gradient centrifugation (Extended Data Fig. 1a), transferred onto holey carbon grids, and three cryo-EM datasets were collected (Extended Data Fig. 1b,c). After extensive three-dimensional (3D) classification, several distinctive particle classes could be separated that represented vRNAP complexes in transcription stages ranging from pre-initiation to co-transcriptional capping.

Cryo-electron microscopy structure of the PIC. Biochemical studies had previously shown that vRNAP, VETF and Rap94 are required for early transcription initiation^{26,32,33}. We identified one particle class in our reconstitution that contained these factors along with the DNA scaffold, and thus represented the bona fide PIC^{32,33}. 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 (Extended Data Fig. 1d–i and Table 1). The density was docked with the core vRNAP model and the VETFI and VETFs chains and parts of Rap94 were traced de novo, allowing complete modeling of the PIC (Fig. 1c,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 (Extended Data 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 (Extended Data Fig. 2a). Both DNA strands appear only minimally separated within the bubble region. The latter joins the adjacent double-helical upstream and downstream

Table 1 | Cryo-EM data collection, refinement and validation statistics

	PIC (EMD-11824)	IPIC (EMD-11843)	ITC1 (EMD-11848)	ITC2 (EMD-11850)	ITC3 (EMD-11851)	IITC (EMD-11844)
	(PDB 7AMV)	(PDB 7AOF)	(PDB 7AOZ)	(PDB 7AP8)	(PDB 7AP9)	(PDB 7AOH)
Data collection and processing						
Magnification	75,000	75,000	75,000	75,000	75,000	75,000
Voltage (kV)						
Electron exposure (e ⁻ /Å ²)	77.5	77.5	77.5	77.5	77.5	77.5
Defocus range (μm)	−1.0 to −2.2	−1.0 to −2.2	−1.0 to −2.2	−1.0 to −2.2	−1.0 to −2.2	−1.0 to −2.2
Pixel size (Å)	1.0635	1.0635	1.0635	1.0635	1.0635	1.0635
Symmetry imposed	C1	C1	C1	C1	C1	C1
Initial particle images (no.) ^a	10,816 / 9,878 / 3,640	10,816 / 9,878 / 3,640	1,513,003 / 5,310,128 / 0	1,513,003 / 5,310,128 / 0	1,513,003 / 5,310,128 / 0	4,766,002 / 5,310,128 / 942,258
Final particle images (no.)	181,788	88,828	133,467	73,035	96,165	278,260
Map resolution (Å) ^b	2.64 (3.15)	3.0	2.9	3.2	3.0	2.5 (3.7)
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.5–5.5	2.8–5.0	2.7–5.0	2.9–5.5	2.8–5.0	2.4–7.0
Refinement						
Initial model used (PDB code)	–	7AOZ	6RIC	7AOZ	7AOZ	NA
Model resolution (Å)	2.6	2.6	2.5	2.7	2.5	2.7
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143
Model resolution range (Å)	2.5–4.2	2.5–5.5	2.4–4.5	2.5–4.8	2.4–4.5	2.5–6.5
Map sharpening B factor (Å ²)	−40	−40	−40	−40	−40	−40
Model composition						
Nonhydrogen atoms	45,175	32,727	31,512	31,399	30,881	39,139
Protein residues	5,348	4,034	3,782	3,782	3,782	4,675
Nucleotides	84	0	42	77	11	77
B factors (Å²)						
Protein	76	110	84	84	110	75
Nucleotides	92	–	183	182	199	143
R.m.s. deviations						
Bond lengths (Å)	0.006	0.004	0.005	0.003	0.009	0.005
Bond angles (°)	0.83	0.90	1.07	0.56	1.08	0.67
Validation						
Clashscore	2.8	4.3	3.0	3.4	4.3	3.5
Poor rotamers (%)	0.41	1.8	2.4	1.8	2.9	1.6
Ramachandran plot						
Favored (%)	97.2	95.0	94.2	94.6	93.1	96.9
Allowed (%)	2.8	5.0	4.8	5.4	6.9	3.1
Disallowed (%)	0.0	0.0	0.0	0.0	0.0	0.0

^aValues for dataset 1 / dataset 2 / dataset 3. ^bValues for consensus refinement. Values in parentheses for two-body multibody refinement, body no. 2 (that is, VETF + DNA for PIC; NPH-I + upstream DNA for IITC).

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 Extended Data Fig. 2a). However, on the opposite side of the core vRNAP, VETFs and VETF1 contact the DNA in the distal upstream and downstream promoter regions, respectively (Fig. 1f and Extended Data Fig. 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 Extended Data 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. VETF1 folds into five distinct domains, termed NTD (N-terminal domain), TBPLD, CRBD, domain 4 and CTD (C-terminal domain) (Fig. 1c,f). Despite the absence of a priori detectable sequence homology, the second domain displays a bilobal TBP fold, and hence is a TBP-like domain (TBPLD). It is located centrally above the polymerase cleft and,

Fig. 2 | Structure of the VETF heterodimer. a, VETF1 CRBD binding to the upstream critical promoter region. Disulfide bridges are depicted as a stick model. **b**, Details of the VETF1 CRBD promoter interaction. The model is depicted in stick representation and base pairs are numbered relative to the TSS. Only bases for the non-template strand are labeled, and the template strand is sequence complementary. Contact between Tyr367 and thymidine bases at positions –18 and –17 are displayed as a transparent van der Waals (vdW) surface. The protein–DNA hydrogen-bond network is depicted as dotted yellow lines. **c**, Schematic of the sequence-specific interactions of the CRBD reader. The CR consensus sequence is depicted according to Yang and others³⁷. The vdW interactions are indicated in gray and hydrogen bonds in yellow. **d**, Details of VETFs binding to the upstream promoter. The intercalating ‘wedge’ residue Phe271 is shown in stick representation.

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 neighboring domain, which recognizes the CR (Fig. 1a,d,f). Based on its fold and binding mode, this module constitutes a novel type of double-stranded DNA binding domain, hence termed the critical region binding domain (CRBD). Although holding only a limited content of secondary structure elements, it gains structural rigidity through three disulfide bridges that position a 3_{10} -helix ideally for its insertion into the major groove of the DNA (Fig. 2a). The side chain-to-base contacts of this helix are the major sites for sequence-specific readout of the promoter sequence (Fig. 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 VETF1 establishes specific contacts to the upstream promoter (Extended Data Fig. 2d). On the core vRNAP, this part of the promoter is anchored via the interaction of domain 2 of Rap94 with the NTD of VETF1 (Fig. 1d). All other domains of VETF1 (NTD, domain 4 and CTD) contribute to the structural backbone of VETF. Domain 4 and the CTD of VETF1 make up the interface to VETFs (Fig. 1f).

The downstream promoter interacts almost exclusively with VETFs (Figs. 1d and 2d and Extended Data Fig. 2b). Only one

additional pointed contact to the core vRNAP is established by the clamp head close to the TSS (Extended Data Fig. 2a). We observe a striking similarity of the first two domains of VETFs with the canonical helicase fold of chromatin remodeling SNF2-type ATPases^{22,34}, of which INO80³⁵ is the closest homolog. With the latter, VETFs shares, along with the vRNAP-associated transcription factor NPH-I, an extended brace helix that stably bridges the N- and C-lobe of the helicase. The intense DNA interaction of the VETFs’ helicase module is accompanied by a strong bend of the helix (Extended Data Fig. 2e). At the point of inflection, Phe271 intercalates via the minor groove, effectively disturbing the planar base-stacking over the range of roughly three 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 helicases³⁶ (see Extended Data Fig. 2f–h for a comparison to the Pol II system).

Promoter positioning and enforcement of directionality. We next asked how the DNA contacts established by the CRBD of VETF1 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, Fig. 2a,b). The CR is essentially a consensus

Fig. 3 | Comparison of the TBP-like domain from vaccinia VETFI with yTBP. **a**, Schematic of the VETF-promoter interactions. **b**, The TBPLD of VETFI in two orthogonal views. Residues intercalating between the nucleobases are depicted in stick model form. **c**, Structure of the yTBP protein bound to a synthetic TATA-box hairpin DNA oligomer⁶⁴ (PDB 1YTB) in two orthogonal views corresponding to the protein orientation of the VETFI TBPLD as seen in **b**.

sequence of 15 A nucleotides, interrupted by a TG dinucleotide^{31,37} (Figs. 1a and 2c). Arg370 and Gln375 engage in base-specific hydrogen bonding that involves the bases of the TG motif on the non-template strand and the complementary AC dinucleotide on the opposing template strand (Fig. 2b,c and Supplementary Video 1). By this means, VETFI 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 Tyr367. Inverse promoter binding would imply an unfavorable contact of Tyr367 with adenine bases (Fig. 2b,c) and thus a single promoter direction is coerced. By this means, the CRBD-DNA interaction ensures the (1) identification of the CR, (2) alignment of the CR relative to the polymerase cleft and (3) enforcement of transcription directionality. The CRBD is thus the main regulator of the transcription initiation process.

Asymmetric DNA binding by the TBP-like domain of VETFI.

Our structure identified VETFI as a TBP-like protein (TBPLP) whose TBPLD is engaged in an intricate contact network comprising the neighboring domains of VETFI, VETFs and Rap94 (Fig. 3a). Members of the TBPLD family had previously been identified solely by means of sequence homology. However, VETFI stands apart from previously known TBPLPs because of its extremely divergent sequence, which, until now, has 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 PDBFold³⁸ when matching it to PDB entry 1YTB. To compare their structures and binding modes, we aligned the TBPLD, the upstream DNA module of VETFI (Fig. 3b), with the yeast TBP (yTBP), TATA-box crystal structure (Fig. 3c). The TBPLD of VETFI features the characteristic saddle structure that was previously described for TBP^{39–42}. However, the symmetry that is evolutionary conserved in TBP^{43,44}

appears broken. As a consequence, and unlike TBP, which contacts the TATA-box symmetrically, VETFI 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^{39–42}. Still, the TBPLD induces a pronounced DNA bend via intercalation of aliphatic, rather than aromatic, side chains (Fig. 3b). In agreement with the fundamentally different binding mode of the TBPLD, a consensus TATA-box is absent from vaccinia early promoters³¹.

Rearrangement of the complete vRNAP into the PIC. Complete vRNAP is the predominant vRNAP complex found in infected cells and is necessary and sufficient to execute viral early transcription. Accordingly, we have previously speculated that this unit becomes incorporated into virions as a pre-assembled unit to promote the restart of transcription in the next infection cycle²¹. 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; however, defined density could only be observed for the CRBD of VETFI, while 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 VETFI coordinates extracted from the PIC model, resulting in reasonable overlap. In the resulting structure (Fig. 1g) VETFI 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 (Supplementary 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 Supplementary Video 1 for a summary of core aspects of the PIC).

Fig. 4 | Structure of the late PIC. a, Model of the ITC with density for the bound DNA oligomer shown as a blue surface, for the PPD (in transparent gold). **b**, Domain structure of the bound transcription factors. Disordered regions are marked by hatched boxes.

Structure of the late PIC. 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 (Extended Data Fig. 3a), which yielded a reconstruction at a resolution of 3.0 Å (Extended Data Fig. 1b–d and Table 1) fulfilled this criterion. The density could be docked with the complete vRNAP model²¹. 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 the phospho-peptide domain, PPD; Fig. 4b). It is in a similar conformation as in the complete vRNAP²¹ 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 (Fig. 4a,b) and enables single-strand capture at later stages (see ‘PPD assisted single-strand capture and formation of the initially transcribing complex’ below). We therefore conclude that this particle represents a late state of the PIC (IPIC) 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 initially transcribing complex. Next, we investigated the structural basis of IPIC conversion into an initially transcribing complex (ITC). Three vRNAP particle classes yielded reconstructions that were identified

Fig. 5 | Three structures of initially transcribing complexes. a, Model of ITC state 1 shown with an overlay of the downstream DNA from ITC conformation 2 and ITC conformation 3. The core polymerase is depicted in gray as a solvent-accessible surface. Downstream DNA density for ITC conformation 3 was weak, so 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.

as different conformations of the ITC based on their composition and promoter positioning (ITC1–3, Extended Data Fig. 3a–d and Table 1). The exact location of the polymerase on the promoter could be determined, because its downstream blunt end was readily visible in the density (Extended Data Fig. 4a). In contrast to the IPIC, 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 IPIC, has been displaced by the template strand. Consequently, the B-homology region has become 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 substantial 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 IPIC structure, on the template-strand capture mechanism.

ATP-dependent upstream promoter scrunching. During 3D classification, one particular class stood out because it comprised particles considerably larger than the ITC (Extended Data Fig. 5a). After a further round of focused classification of these particles on the

Fig. 6 | Structure of the IITC. a, Model of the IITC in two orthogonal views. The core polymerase is depicted in gray as a solvent-accessible surface. **b**, Domain structure of the bound transcription factors. Disordered regions are marked by hatched boxes. **c**, 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.

observed extra density, followed by multibody refinement, a reconstruction was obtained that allowed the construction of a complete model (Fig. 6a, Extended Data Fig. 5b–d and Table 1). This complex was classified as a late ITC (IITC), based on the positions of the blunt ends of the upstream and downstream promoter-DNA segments that are visible in the density (Extended Data Fig. 4b) as well as 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 five base pairs in the downstream direction compared to the ITC (Extended Data Fig. 4a). 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 (Extended Data Fig. 4d,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 IITC is fundamentally different from that observed in the vaccinia PIC and in the ITC of Pol II⁴⁵.

The blunt ends of the DNA promoter scaffold are visible in the EM density of the IITC (Extended Data Fig. 4b), thus allowing us to determine the position of (and the size of) vRNAP relative to the transcription bubble (Extended Data Fig. 4c). 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 (Extended Data Fig. 4c). It is likely that this condition enables promoter escape and hence contributes to the transition of the initiation phase into productive elongation (Supplementary 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 mechanistic 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 and 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 VETF1, 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 disulfide bridges. This is in agreement with two early studies observing DNase protection of the –15 to –30 promoter region⁴⁶ as well as crosslinking of VETF1 to the upstream and of VETFs to the downstream promoter region³². Cysteine formation in the CRBD

Fig. 7 | Transition of complete vRNAP to the PIC, and a model for early promoter recognition and opening. a, Schematic of vaccinia early promoter recognition and the opening mechanism (color code as in Fig. 1). **b**, Plot of clamp closure versus transcription state. CCC, co-transcriptional capping complex.

may be introduced by vaccinia-encoded enzymes⁴⁷ 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 probably generates the nucleation site for promoter melting. Strikingly, the TBPLD of VETFl displays an asymmetric DNA binding mode. This sharply contrasts with 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 TBP⁴⁸, 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^{49,50}. Because 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 I⁵¹ and may also occur in other TBPLDs^{43,52,53}. To the best of our knowledge, no other structures of TBP-like domains exist in the database, and our structure might therefore contribute to the general understanding of this domain family. In this regard, it is interesting to note that a recent cryo-EM study reports a binding mode of TBP on a TATA-less promoter similar to that on TATA-containing promoters⁵⁴. The behavior of the VETFl TBPLD might therefore be fundamentally different from that of genuine TBP.

Although the order of events cannot definitely be determined given the current state of knowledge, we propose the following mechanism for vaccinia early promoter melting based on our data and prior findings for the Pol II system⁵⁵ (Fig. 3a). (1) The CRBD of VETFl binds the promoter at the CR, thereby enforcing directionality (Fig. 2a,b). (2) VETFs pulls the DNA in an ATP-dependent reaction towards the vRNAP clamp and lobe, analogous to the XPB

helicase in the Pol II system⁵⁶ (compare Extended Data Fig. 2g to Extended Data Fig. 2h). (3) The clamp closes tightly around the DNA (Fig. 7b), thereby shaping its path⁵⁷. (4) 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). (5) 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). (6) This bend triggers the initial melting event at the TSS, 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^{45,58}; compare Extended Data Fig. 2g to Extended Data 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 (Extended Data Figs. 2e and 6) and of the TFIIB subunit XPB (SSI2 in yeast, Extended Data Figs. 2f and 6) are functional counterparts⁵⁹.

In contrast to a recent study describing a PIC intermediate of Pol II immediately prior to the initially melted state⁵⁵, 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/TFIIB module on the DNA in the Pol II PIC (Extended Data Fig. 2h). This notion is corroborated by the fact that, despite their fundamentally

different binding modes, both TBP and the VETFI TBPLD induce a strong bend of the DNA. Thus, the architecture of the vaccinia PIC differs fundamentally from its nuclear counterparts. Although the catalytic cores of all multi-subunit polymerases are largely homologous, only basic architectural features are conserved with respect to the positioning of the early transcription factor. The arrangement of the VETFI TBPLD is so far unprecedented and unexpected. 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 Pol II system⁶⁰.

During transition to the IITC, a dramatic reorganization 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 (Extended Data Fig. 4c). To accommodate the re-routed upstream DNA, the cyclin domain of Rap94 undergoes a conformational change (Extended Data Fig. 4d) accompanied by other changes in the Rap94/core enzyme interaction (Extended Data Fig. 4e). The only plausible explanation for the widening of the transcription bubble is that the NPH-I helicase motor has actively melted and scrunched upstream DNA duplex into the core vRNAP^{61,62}. By this means, NPH-I probably assists promoter escape by adding the free energy of ATP hydrolysis to the generation of an energy-rich transcription intermediate¹¹. Although, 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 TFIIF injects free energy from ATP hydrolysis into the ITC during TSS scanning has been postulated for Pol II¹⁰. 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 (Extended Data Fig. 2e) and the insertion of the 'wedge' residue Phe273 (Fig. 2d) stabilize the upstream fork point of the transcription bubble in the IITC. Processive vRNAP elongation complexes can be assembled in the absence of Rap94 *in vitro*²⁰. *In vivo*, such complexes are found associated with the latter^{28,63}. Thus, Rap94 may ensure the efficient recruitment of NPH-I to ECs stalled at pause sites to enable readthrough⁶², 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 and Supplementary Video 2). First, 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. Template-strand capture proceeds 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 a manner analogous to that which has been observed for Pol II⁴. 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 the 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 remodeling 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,c). Transition to a processive EC (step 6) triggers contraction of the transcription bubble, mobilization of the upstream DNA duplex and loss of NPH-I. Alternately, 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 mechanistic insights into the initial phase of poxvirus transcription. Some features observed in the presented structure are poxvirus-specific, such as the unique promoter recognition by the CRBD of VETFI. Others, such as the hitherto unknown behavior 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-021-00655-w>.

Received: 8 April 2021; Accepted: 28 July 2021;

Published online: 23 September 2021

References

- Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).
- Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes Dev.* **7**, 1291–1308 (1993).
- 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**, 206–209 (2010).
- Kostrewa, D. et al. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
- Kapanidis, A. N. et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
- Brieba, L. G. & Sousa, R. T7 promoter release mediated by DNA scrunching. *EMBO J.* **20**, 6826–6835 (2001).
- Cheetham, G. M. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).
- Henderson, K. L. et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **114**, E3032–E3040 (2017).
- Winkelman, J. T. et al. Multiplexed protein–DNA cross-linking: scrunching in transcription start site selection. *Science* **351**, 1090–1093 (2016).
- Roberts, J. W. Biochemistry. RNA polymerase, a scrunching machine. *Science* **314**, 1097–1098 (2006).
- Shchelkunov, S. N. An increasing danger of zoonotic orthopoxvirus infections. *PLoS Pathog.* **9**, e1003756 (2013).
- Lewis-Jones, S. Zoonotic poxvirus infections in humans. *Curr. Opin. Infect. Dis.* **17**, 81–89 (2004).
- Grant, R., Nguyen, L. L. & Breban, R. Modelling human-to-human transmission of monkeypox. *Bull. World Health Organ.* **98**, 638–640 (2020).
- 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**, 1029–1035 (1986).
- 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. USA* **83**, 3141–3145 (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. USA* **72**, 2525–2529 (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**, 10754–10760 (1988).
19. Yuen, L., Davison, A. J. & Moss, B. Early promoter-binding factor from vaccinia virions. *Proc. Natl Acad. Sci. USA* **84**, 6069–6073 (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 (2019).
22. Gershon, P. D. & Moss, B. Early transcription factor subunits are encoded by vaccinia virus late genes. *Proc. Natl Acad. Sci. USA* **87**, 4401–4405 (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**, 513–522 (1989).
24. Shuman, S., Broyles, S. S. & Moss, B. Purification and characterization of a transcription termination factor from vaccinia virions. *J. Biol. Chem.* **262**, 12372–12380 (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**, 7552–7557 (1994).
27. Ahn, B. Y. & Moss, B. RNA polymerase-associated transcription specificity factor encoded by vaccinia virus. *Proc. Natl Acad. Sci. USA* **89**, 3536–3540 (1992).
28. 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**, 1360–1370 (1994).
29. Shuman, S. & Hurwitz, J. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme–guanylate intermediate. *Proc. Natl Acad. Sci. USA* **78**, 187–191 (1981).
30. Paoletti, E. & Moss, B. Two nucleic acid-dependent nucleoside triphosphate phosphohydrolases from vaccinia virus. Nucleotide substrate and polynucleotide cofactor specificities. *J. Biol. Chem.* **249**, 3281–3286 (1974).
31. Davison, A. J. & Moss, B. Structure of vaccinia virus early promoters. *J. Mol. Biol.* **210**, 749–769 (1989).
32. 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. USA* **93**, 7540–7545 (1996).
33. 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**, 6052–6056 (1994).
34. Broyles, S. S. & Moss, B. DNA-dependent ATPase activity associated with vaccinia virus early transcription factor. *J. Biol. Chem.* **263**, 10761–10765 (1988).
35. Eustermann, S. et al. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature* **556**, 386–390 (2018).
36. 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**, 177–187 (2010).
37. 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**, 5897–5909 (2011).
38. Krissinel, E. Enhanced fold recognition using efficient short fragment clustering. *J. Mol. Biochem.* **1**, 76–85 (2012).
39. Nikolov, D. B. & Burley, S. K. 2.1 Å resolution refined structure of a TATA box-binding protein (TBP). *Nat. Struct. Biol.* **1**, 621–637 (1994).
40. Kim, J. L. & Burley, S. K. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1**, 638–653 (1994).
41. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
42. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
43. Blombach, F. & Grohmann, D. Same same but different: the evolution of TBP in archaea and their eukaryotic offspring. *Transcription* **8**, 162–168 (2017).
44. 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).
45. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
46. Broyles, S. S., Li, J. & Moss, B. Promoter DNA contacts made by the vaccinia virus early transcription factor. *J. Biol. Chem.* **266**, 15539–15544 (1991).
47. 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. USA* **99**, 6667–6672 (2002).
48. Struhl, K. Duality of TBP, the universal transcription factor. *Science* **263**, 1103–1104 (1994).
49. 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**, 2929–2943 (2014).
50. 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**, 151–161 (2003).
51. 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**, 31259–31267 (2004).
52. Ravarani, C. N. J. et al. Molecular determinants underlying functional innovations of TBP and their impact on transcription initiation. *Nat. Commun.* **11**, 2384 (2020).
53. Kramm, K., Engel, C. & Grohmann, D. Transcription initiation factor TBP: old friend new questions. *Biochem. Soc. Trans.* **47**, 411–423 (2019).
54. Chen, X. et al. Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, eaba8490 (2021).
55. Dienemann, C., Schwalb, B., Schilbach, S. & Cramer, P. Promoter distortion and opening in the RNA polymerase II cleft. *Mol. Cell* **73**, 97–106 (2019).
56. Murakami, K. et al. Uncoupling promoter opening from start-site scanning. *Mol. Cell* **59**, 133–138 (2015).
57. Feklistov, A. et al. RNA polymerase motions during promoter melting. *Science* **356**, 863–866 (2017).
58. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIB and Mediator. *Nature* **551**, 204–209 (2017).
59. Kolesnikova, O., Radu, L. & Poterszman, A. TFIIB: a multi-subunit complex at the cross-roads of transcription and DNA repair. *Adv. Protein Chem. Struct. Biol.* **115**, 21–67 (2019).
60. Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* **149**, 1431–1437 (2012).
61. 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**, 14826–14838 (2016).
62. Deng, L. & Shuman, S. Vaccinia NPH-I, a DEXH-box ATPase, is the energy coupling factor for mRNA transcription termination. *Genes Dev.* **12**, 538–546 (1998).
63. Yang, Z. & Moss, B. Interaction of the vaccinia virus RNA polymerase-associated 94-kilodalton protein with the early transcription factor. *J. Virol.* **83**, 12018–12026 (2009).
64. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIB/TBP/DNA complex. *Nature* **381**, 127–151 (1996).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Generation of recombinant vaccinia virus GLV-1h439 and vRNAP purification.

GLV-1h439 was derived from GLV-1h68 as described previously²¹. For vRNAP purification, HeLa S3 cells were cultured in DMEM, containing 10% FBS at 37 °C in the presence of 5% CO₂. Cells were grown to 90% confluence and infected with purified GLV-1h439 with a multiplicity of infection of 1.2. After 24 h, the infected cells were pelleted and resuspended in lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 0.5% (vol/vol) 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). The beads were washed four times with buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, 0.1% (vol/vol) NP-40, 1 mM DTT, equilibrated with elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1.5 mM MgCl₂, and 1 mM DTT) and eluted with a 200 µg ml⁻¹ solution of 3× FLAG peptide (Sigma-Aldrich). The eluate was analyzed by SDS polyacrylamide gel electrophoresis (SDS–PAGE). Approximately 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 MgCl₂ 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 [³²P]-labeled DNA promoter scaffold was incubated for 30 min at 30 °C with the indicated amount of vRNAP in the presence of 1 mM indicated NTPs. Complexes were analyzed by native gel electrophoresis (4% acrylamide and 0.13% bis-acrylamide, 25 mM Tris-HCl pH 7.4, 25 mM boric acid, 0.5 mM EDTA) at 4 °C. For large-scale reconstitution of promoter/vRNAP complexes, purified vRNAP was concentrated in a Vivaspins 10-kDa cutoff 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 MgCl₂, 1 mM DTT) in the presence of ATP and UTP (1 mM each) for 30 min at 30 °C. The mixture was separated by a 10–30% sucrose gradient centrifugation (16 h, 35,000 r.p.m., Beckman 60Ti rotor, 4 °C). Gradient fractions were collected manually and analyzed by SDS–PAGE followed by silver- and ethidium-bromide staining to visualize the proteins and DNA. The indicated fractions (Extended Data Fig. 1a) 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 MgCl₂, 1 mM DTT) in a concentrator.

Cryo-electron microscopy and model building of the PIC. Following sucrose gradient purification, the indicated fractions (Extended Data Fig. 1) were pooled, diluted 1:50 with a buffer containing 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM MgCl₂ and 1 mM DTT, and centrifuged in a Vivaspins concentrator to remove the sucrose. R1.2/1.3 holey carbon grids (Quantifoil) were glow-discharged for 90 s (Plasma Cleaner model PDC-002; Harrick Plasma) at medium power, and 3.5 µl of C2 sample was applied inside a Vitrobot Mark IV instrument (FEI) at 4 °C and 100% relative humidity. Grids were blotted for 3 s with blot force 5 and plunged into liquid ethane. Cryo-EM datasets comprising 10,816 (dataset 1), 9,878 (dataset 2) and 3,640 (dataset 3) micrographs, respectively, were collected from three different grids with a Thermo Fisher Titan Krios G3 set-up equipped with a Falcon III camera (Thermo Fisher). Data were acquired with EPU (Thermo Fisher) at 300 keV and a nominal magnification of ×75,000 (calibrated pixel size of 1.0635 Å) in video mode with 47 fractions per video and counting of the electron signal. The total exposure was 77.5 e⁻/Å² for 75 s, with two exposures per hole.

Dose-weighted, motion-corrected sums of the micrograph videos were calculated with Motioncor2⁶⁵. The contrast-transfer function (CTF) of each micrograph was fitted with RELION 3.1⁶⁶ using the built-in CTFIND algorithm. An initial set of 25,000 particles was picked with the Gaussian picker and subjected to three rounds of 2D classification in RELION⁶⁶ 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, respectively. 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 (Extended Data 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 improvement of particular regions of the reconstruction. We therefore repeated the MB refinement with 11 more mask pairs. The resulting 12 map pairs were then combined with Phenix.combine_focused_maps to create a single, optimal map and the procedure was repeated for several selected subsets. The combined map of all 12 map pairs was compared to the different combined maps based on subsets. The combined map based on all 12 map pairs showed comparably better richness of detail and connectivity for VETF and was therefore used for automated model refinement. To build the PIC model, the vRNAP core excluding the Rpo30 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 0.9⁶⁷. To this end, the SNF2 helicase core of VETFs was located and built, followed by well-defined regions of VETF. The resulting partial model was initially refined with Phenix.real_space_refine and forwarded to Phenix.combine_focused_maps to create a stitched map, and the VETF model was completed manually. The full polypeptide chains of both VETFs and VETF were continuously modeled. 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_refine⁶⁸, including ADP refinement steps. During refinement, secondary structure and 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 (Table 1).

Three-dimensional reconstruction and model building of IPIC and ITC complexes. The IPIC particle stack obtained as described above was subjected to two rounds of focused 3D classification with three 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 (Extended Data 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 IPIC and ITC1–3 (Extended Data Fig. 3a). The resolution was determined by Fourier-shell correlation (FSC) to 3.0 Å for the IPIC and 2.9 Å, 3.2 Å and 3.0 Å for ITC1, ITC2 and ITC3, respectively (Extended Data Fig. 3b–d). To build the IPIC 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 was adjusted manually with Coot⁶⁷, and the model was refined with Phenix.real_space_refine⁶⁸ 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 1).

Three-dimensional reconstruction and model building of the IITC. The IITC 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 two-body MB refinement 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 IITC 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.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

Crystallographic structure factors and atomic coordinates have been deposited in the Protein Data Bank (PDB) under accession nos. 7AMV, 7AOE, 7AOZ, 7AP8, 7AP9 and 7AOH. Cryo-EM maps have been deposited with the Electron Microscopy Data Bank (EMDB) under accession nos. EMD-11824, EMD-11843,

EMD-11848, EMD-11850, EMD-11851 and EMD-11844. Source data are provided with this paper.

References

65. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
66. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
67. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
68. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).

Acknowledgements

We thank C. Kraft for 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, analyzed the models and prepared the figures and videos. C.G. and U.F. wrote the manuscript. U.F. designed and supervised the project. A.A.S. and U.F. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41594-021-00655-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-021-00655-w>.

Correspondence and requests for materials should be addressed to Clemens Grimm or Utz Fischer.

Reprints and permissions information is available at www.nature.com/reprints.

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Reconstitution and preparative purification of vRNAP-promoter complexes. Approx. 400 μg affinity-purified complete vRNAP was incubated with a 60-fold molar excess of the synthetic DNA scaffold in the presence of 1 mM NTPs and separated by gradient centrifugation. Fractions 13-16 were pooled and used for cryo-EM studies. **(b)** Representative micrograph and **(c)** representative 2D classes of the dataset. If applicable, the images are annotated with the names of the predominant particle type. **(d)** Classification and refinement scheme. **(e)** Local resolution mapped to the consensus reconstruction density iso-surface (only a mild B-factor sharpening of $-10 \text{ \AA}^2$ was applied). **(f)** Masked VETF and DNA region after multibody refinement. **(g)** FSC curves for consensus and multibody refinements. **(h)** Angular distribution plots referring to the consensus reconstruction in E. **(i)** Selected views of the final, B-factor sharpened ($-60 \text{ \AA}^2$) cryo-EM density isosurface overlaid with the model.

Extended Data Fig. 2 | DNA contacts in the PIC. Transparent blue iso-surface of the cryo-EM density for the bound DNA, filtered by a Gaussian blur to 1.5σ standard deviation. The model is shown in cartoon style and the initially melted region (IMR) is indicated. **(a)** Top view of the PIC with VETF removed (top view) and vRNAP core shown as solvent accessible surface. The clamp head and lobe are marked on the molecular surface by a rose dotted line, respectively. **(b)** Front view of the PIC with core removed (front view) and VETF shown in cartoon representation. **(c)** PIC with vRNAP removed shown in cartoon view turned by roughly 90° relative to **(b)**, and slightly optimized for clarity. **(d)** Upstream promoter contacts to the core vRNAP. Detailed view of the upstream promoter contacts to the core vRNAP in cartoon representation. The lobe region contacting the DNA is indicated with a rose dotted line. **(e)** Cartoon model of VETFs and downstream DNA with superposed ideal B-DNA in transparent grey is shown. The respective helix axes are indicated and Phe 271 is depicted in stick representation. **(f)** A depiction of the promoter-bound yeast XPB homologue SSL2 from the yeast PIC bound to TFIIH and core mediator (Schilbach et al., 2017) (PDB:5oqm) analogous to **(a)**. The axis of the bent, bound DNA (blue) is indicated. Both arms of the DNA helix axis bend angles in **(e)** and **(f)** lie approximately in the paper plane. **(g)** Vaccinia PIC model in cartoon representation as shown in Fig. 1a, front view. **(h)** Pol II core PIC model (PDB 5IY6) in cartoon representation and oriented by superposition of the Pol II core polymerase with the core vRNAP of the vaccinia PIC in **(g)**. Elements identified as functionally, architecturally or structurally corresponding are colored according to the scheme used for the vaccinia PIC throughout this manuscript.

Extended Data Fig. 3 | Cryo EM reconstruction of IPIC and ITC. (a) classification and refinement scheme. Areas used for local classification is encircled in magenta. **(b)** Local resolution mapped to the reconstruction density isosurface. **(c)** FSC plots for consensus refinement and the separate bodies of the multibody (MB) refinement. **(d)** Angular orientation plots referring to the reconstructions in (b).

Extended Data Fig. 4 | Cryo EM reconstruction of the IITC. (a) classification and refinement scheme. **(b)** Local resolution mapped to the reconstruction density isosurface. **(c)** FSC plots for consensus refinement and the two-body MB refinement, respectively. **(d)** Orientation plot referring to the consensus reconstruction in (b).

Extended Data Fig. 5 | Details of the DNA in the ITC and IITC. (a) Cryo EM density for DNA in the ITC. (b) Cryo EM density for DNA in the IITC. The blunt ends of the synthetic DNA scaffold are recognizable in the density. Note that the position of the downstream end of the scaffold has advanced roughly 5 bp when comparing the ITC (a) to the IITC (b) structure. The visible upstream blunt end indicates extensive promoter scrunching. (c) Transcription bubble in the IITC. A zoomed view into the active site region is depicted. Disordered regions of the template and non-template strand are shown as dotted lines. The start and end positions of the melted promoter and the base at the active site are numbered relative to the TSS. (d) Illustration of major rearrangements during ITC formation. The IITC model is shown in cartoon representation. The B-cyclin as in the IPIC structure is shown as a molecular surface. In this orientation it would obstruct the upstream DNA path. The domain rearrangement is indicated by a magenta arrow. (e) Analogous to (d), the B ribbon as in the IPIC structure is shown as a molecular surface, its rearrangement indicated by a magenta arrow. The stabilization of transiently flexible regions of Rap94 in the IITC by formation of a β strand to the core vRNAP is highlighted with a magenta box.

Extended Data Fig. 6 | Comparison of Vaccinia NPH-I and VETFs with structurally related helicases. Color code according to common structural elements. The helicase ATPase modules were extracted from structures with the following PDB ID codes: Complete vRNAP: 6RFL, INO80: 6HTS, Rad26: 5VVR, XPB: 6NMI.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection EPU 2 Software for Single Particle Analysis, Thermo Fisher, Germany (<https://www.thermofisher.com/de/de/home/electron-microscopy/products/software-em-3d-vis/epu-software.html>)

Data analysis Motioncor2 (<https://emcore.ucsf.edu/ucsf-software>), (Relion 3.1 <https://github.com/3dem/relion/releases>), CTFFIND algorithm as implemented in Relion 3.1, Models were built with COOT 0.9.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

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; IPIC: 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	1,064,795 particles (dataset 1), 1,205,746 particles (dataset 2), and 323,776 particles (dataset 3)
Data exclusions	Particles yielding bad 2D classes were excluded from the reconstructions.
Replication	Three Cryo EM datasets were collected and merged in the final phase of the structure determination. Band shift assays were repeated three times.
Randomization	2D classification in Relion 3.1 was performed with random subsets of only K*100 particles, with K being the number of classes for the first five iterations; the next 5 iterations were carried out with K*300 particles, the next 5 with 30% of the data set; and the final ones with all data.
Blinding	No blinding was performed in this work, as is common practice in cryo electron microscopic structure determination.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	ANTI-FLAG® M2 affinity gel, Sigma Aldrich, Germany, Product number A2220, Lot No. SLCH0130
Validation	Certificate of analysis published online at: https://www.sigmaaldrich.com/certificates/COFA/A2/A2220/A2220-BULK____SLCH0130_.pdf

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The HeLa S3 cell line was bought from Thermo Fisher, Germany
Authentication	The cell line was not authenticated.
Mycoplasma contamination	The cell line tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.