

Transcription initiation complex structures elucidate DNA opening

C. Plaschka^{1*}, M. Hantsche^{1*}, C. Dienemann¹, C. Burzinski¹, J. Plitzko² & P. Cramer¹

Transcription of eukaryotic protein-coding genes begins with assembly of the RNA polymerase (Pol) II initiation complex and promoter DNA opening. Here we report cryo-electron microscopy (cryo-EM) structures of yeast initiation complexes containing closed and open DNA at resolutions of 8.8 Å and 3.6 Å, respectively. DNA is positioned and retained over the Pol II cleft by a network of interactions between the TATA-box-binding protein TBP and transcription factors TFIIA, TFIIB, TFIIE, and TFIIF. DNA opening occurs around the tip of the Pol II clamp and the TFIIE ‘extended winged helix’ domain, and can occur in the absence of TFIIH. Loading of the DNA template strand into the active centre may be facilitated by movements of obstructing protein elements triggered by allosteric binding of the TFIIE ‘E-ribbon’ domain. The results suggest a unified model for transcription initiation with a key event, the trapping of open promoter DNA by extended protein–protein and protein–DNA contacts.

For transcription initiation, Pol II assembles with the basal transcription factors (TF) IIB, TFIID (or its subunit TBP), TFIIE, TFIIF, and TFIIH^{1–4} on double-stranded promoter DNA to form the closed complex (CC). Upon DNA opening, the template strand slips into the Pol II active centre, and a DNA ‘bubble’ forms ~20–30 base pairs (bp) downstream of the TATA box⁵, leading to the open complex (OC). Efficient DNA opening requires TFIIE and TFIIH^{6–8}, but these factors are not required for low levels of transcription^{9,10} and TFIIE alone can open certain promoters^{11,12}. Subsequent RNA synthesis results in the initially transcribing complex (ITC), which is then converted into an elongation complex for processive RNA synthesis.

The three-dimensional architecture of Pol II initiation complex intermediates was studied in yeast^{13–19} and human²⁰ systems. These studies revealed that the promoter assembly containing TBP and TFIIB resides over the Pol II wall and positions DNA above the polymerase cleft and along the clamp. TFIIE and TFIIF bind to opposite sides of Pol II and flank promoter DNA. Owing to the limited resolution of available structural studies, many questions regarding the molecular basis of initiation remain, including the mechanisms of DNA opening and template-strand loading into the active centre. A differing model of the yeast initiation complex¹⁷ was revised recently²¹, and now agrees, on a topological level, with other studies and results reported here.

Here we report cryo-EM structures of CC and OC assemblies from the yeast *Saccharomyces cerevisiae* at resolutions of 8.8 Å and 3.6 Å, respectively. The structures contain all of the basal transcription factors except TFIIH, which is currently not available at a quality required for high-resolution analysis. We show that DNA opening can occur in the absence of TFIIH, and provide mechanistic insights into DNA opening and template-strand loading. Our results also unveil the exact location and intricate structure of basal factors and their induced folding interactions with each other, with Pol II, and with promoter DNA. The data also demonstrate the high structural conservation between yeast and human initiation systems.

Transcription initiation complex at 3.6 Å

We extended our previous cryo-EM analysis of a Pol II core initiation complex containing TBP, TFIIB, and TFIIF¹⁹ by adding TFIIA and

TFIIE (Methods). Formation of a stable and stoichiometric complex required the presence of core Mediator, which showed low occupancy and high flexibility under cryo-EM conditions, as observed previously¹⁹, and was excluded from further analysis (Methods, Extended Data Fig. 1a–c). We acquired 257,259 cryo-EM single particle images using a K2 direct electron detector (Methods, Extended Data Fig. 1b, c). Unsupervised particle sorting led to an OC structure at an overall resolution of 3.6 Å, revealing the Pol II core at up to 3.1 Å (OC1), compared to 7.8 Å for our core ITC structure¹⁹ (Fig. 1a, b, Extended Data Figs 1d–g). Further particle sorting revealed improved density for TFIIB and TFIIF at around 4 Å resolution (OC2 and OC4), and for TFIIE at 4.4 Å resolution (OC3, Methods, Extended Data Figs 1i, k, l). The final structure mainly consists of atomic models (90%) and contains backbone models for parts of the basal factors (Extended Data Fig. 1f–l).

The structure reveals upstream DNA above the Pol II wall and downstream DNA in the active centre cleft¹⁹. We observed only fragmented density for single strands within the DNA bubble (Fig. 1b, c, Extended Data Fig. 1m), and no density for RNA, which was not retained during cryo-EM sample preparation (not shown). The bubble contains 15 mismatches, begins at the natural distance of ~20 bp downstream of the TATA box⁵, and extends further downstream than when it initially forms, resembling the situation during transcription start-site scanning. For the basal factors, regions that are essential for cell viability are generally observed, whereas non-essential, non-conserved regions^{16,22–26} are often mobile (Fig. 1a, b, Extended Data Fig. 1j–l). TFIIA is located near upstream DNA and TBP as expected²³ (Fig. 1b, Extended Data Fig. 1j). For TFIIB, the B-ribbon and B-core domains are well defined, including the newly modelled carboxy-terminal cyclin domain from yeast (Fig. 2), whereas the B-linker shows weak density, and the B-reader^{15,19} is mobile (Extended Data Fig. 1i).

DNA positioning and retention

TFIIF adopts an intricate fold within the OC. Its dimerization module and charged helix are located on the Pol II lobe domain as in the ITC^{18,19} (Extended Data Fig. 2). The ‘arm’ in the large TFIIF subunit Tfg1 (human RAP74) adds a β-strand to the Pol II protrusion, and projects into the cleft, where it may stabilize the DNA bubble^{18–20}.

¹Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany. ²Max Planck Institute for Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

*These authors contributed equally to this work.

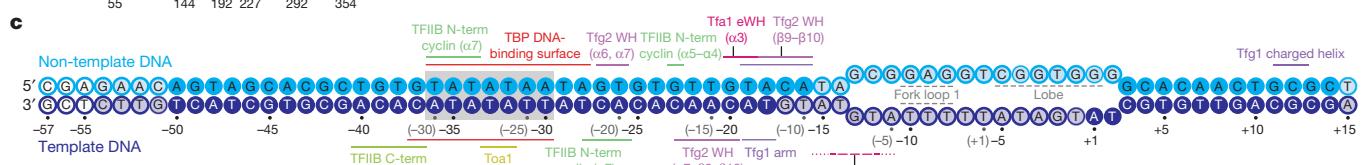
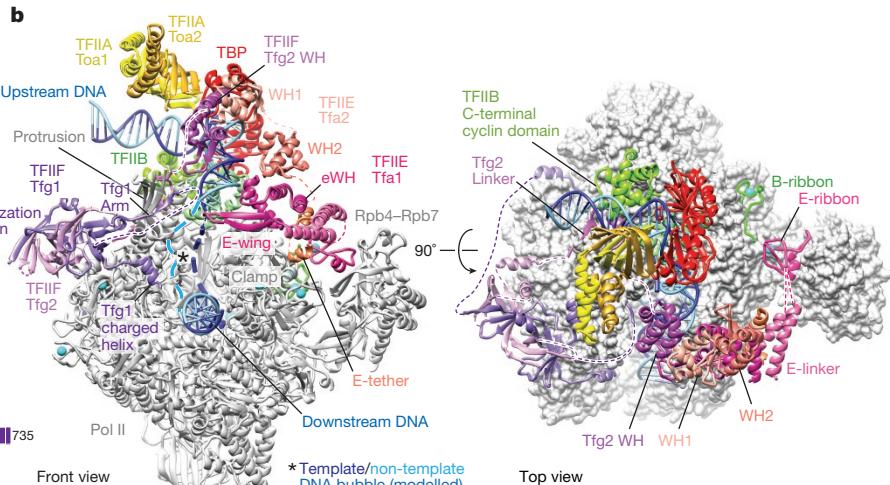
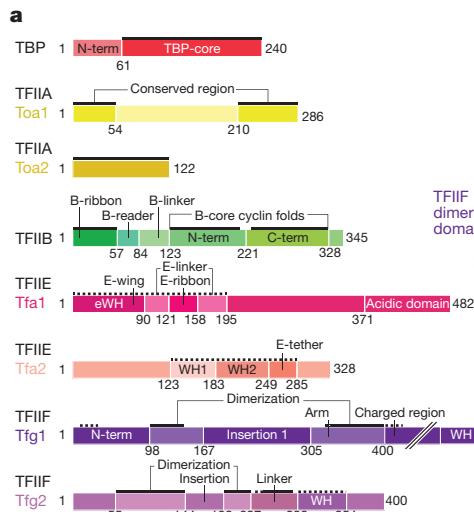


Figure 1 | Open complex structure at 3.6 Å resolution. **a**, Domain organization of yeast basal transcription factors TBP, TFIIA, TFIIB, TFIIE, and TFIIF. Solid and dashed black bars indicate protein regions that are present in the OC structure as atomic and backbone models, respectively. Colour code used throughout. **b**, Two views⁵⁰ of the yeast OC structure. Pol II is in silver. DNA template and non-template strands are in dark blue and cyan, respectively. On the right, Pol II is shown as a surface representation, all other proteins are shown as ribbon models. **c**, Protein–

(Fig. 2). The linker in TFIIF subunit Tfg2 (human RAP30) emanates from the dimerization module, and winds along the base of the protrusion, where it binds a hydrophobic pocket. The Tfg2 linker continues between the protrusion and the TFIIB cyclin domains, and connects to the Tfg2 C-terminal winged-helix (WH) domain on top of the cleft (Fig. 2, Extended Data Fig. 2f). The Tfg2 linker stabilizes TFIIB on the wall of Pol II^{27,28}. The yeast-specific amino-terminal region of Tfg1 binds near the Pol II external 1 domain, according to a

DNA contacts. Promoter DNA nucleotides are depicted with solid, shaded, and empty circles when they were included in the structure, excluded owing to weak density, or excluded owing to a lack of density, respectively. Solid and dashed lines indicate observed and putative protein interactions, respectively. A magenta dashed line indicates the contact between closed DNA and the TFIIE E-wing. The register of promoter DNA is given for analogous yeast (black) and human (grey) positions with respect to the transcription start site (TSS, +1). TATA box is indicated by a grey box.

separate crystallographic analysis (Extended Data Fig. 2d, e, Extended Data Fig. 9c).

TFIIE is located between the clamp and the Rpb4–Rpb7 stalk (Fig. 3a, b, Methods, and Extended Data Fig. 3), consistent with previous topological placement of TFIIE^{16,20,29–31} and its archaeal counterpart³². The TFIIE structure differs from a recent model obtained at 6 Å resolution²¹ (see Extended Data Fig. 1e legend for details, Extended Data Fig. 3, and refs 16, 17). The large TFIIE subunit Tfa1 (human TFIIEα) contains an extended winged helix ('eWH') domain³³ and a zinc ribbon domain³⁴ ('E-ribbon') that are connected by α-helices (called here 'E-linker') (Fig. 3b). The eWH domain uses its helix α3 to contact both the tip of the Pol II clamp helices and the DNA backbone at positions –13–14 upstream of the transcription start site (TSS, position +1) (Fig. 1c, 3a). The E-ribbon binds between the clamp, the Rpb7 oligonucleotide-binding (OB) domain, and the B-ribbon (Fig. 3a). The E-linker^{20,21} and the mobile C-terminal domain of Tfa1 contact TFIIE^{7,35}, which may alter TFIIE conformation. The small TFIIE subunit Tfa2 (human TFIIEβ) contains two WH domains ('WH1' (ref. 36), 'WH2'), and two conserved α-helices (called here 'E-tether') that bind the E-linker (Fig. 3b). Consistent with the structure, the E-tether is essential for TFIIE function^{16,37}.

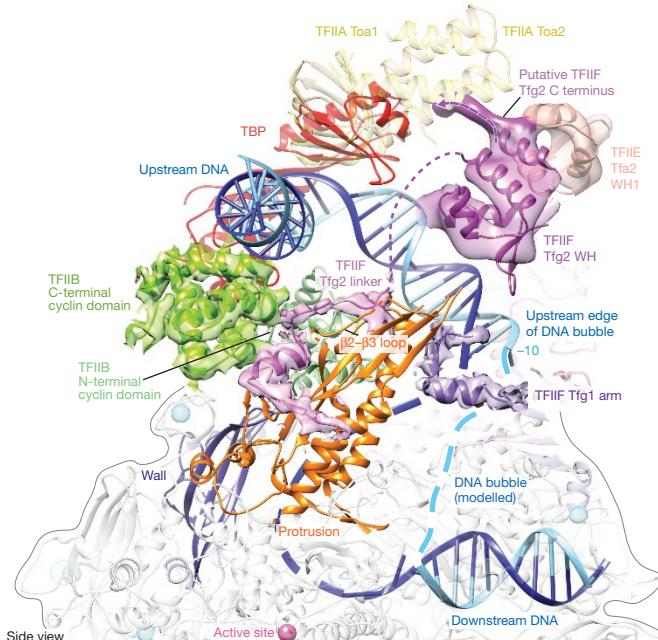


Figure 2 | Basal factors position and retain DNA. Details of the upstream DNA assembly viewed from the side⁵⁰. Highlighted are the locations of the Pol II wall (navy blue), protrusion (orange), TBP (red), TFIIB (green), TFIIF arm (purple), Tfg2 linker (dark magenta) and winged helix (light purple), TFIIE Tfa2 WH1 (light salmon), DNA (template, blue; non-template, cyan), and the active site (magenta sphere). TFIIA is transparent (light yellow). Cryo-EM densities for newly modelled regions of TFIIB, TFIIF, and TFIIE are superimposed on their structural models. Interactions with the Pol II protrusion and upstream edge of the DNA bubble are indicated. TBP contacts a density assigned to the Tfg2 C-terminal region, consistent with interaction of their human counterparts²⁹ (Extended Data Fig. 2f).

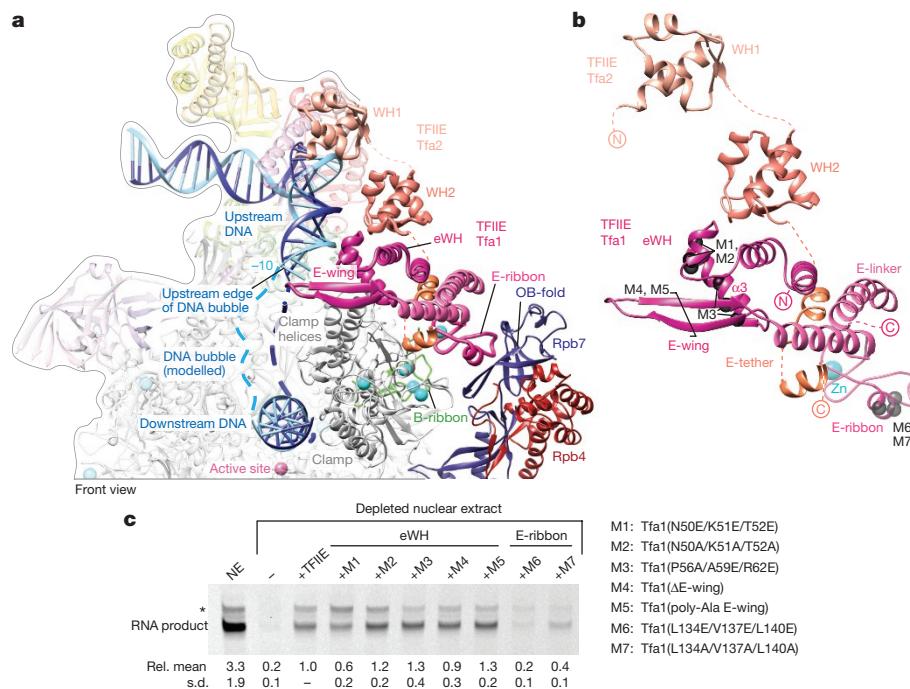


Figure 3 | TFIIE architecture and interactions. **a**, TFIIE interactions within the OC. Depicted are interactions of the TFIIE E-ribbon with the Pol II clamp, stalk subunit Rpb7, and the TFIIB B-ribbon, and interactions of the TFIIE eWH domain with the Pol II clamp helices and upstream DNA. The eWH E-wing lies close to the upstream DNA edge, similar to WH domains involved in DNA strand separation (Extended Data Fig. 3j). Colours as in Fig. 1, except for the Pol II stalk (Rpb4, dark red; Rpb7, dark blue). **b**, TFIIE domain architecture. The TFIIE variants used for functional assays are indicated as α spheres for point mutations, and with a black bracket for E-wing alterations (compare Extended Data Fig. 3g, h, j). Connectivity of the Tfa2 E-tether helices is uncertain. **c**, Selected TFIIE

variants impair transcription from a *HIS4* promoter (Methods, Extended Data Fig. 3g, h). TFIIE-depleted nuclear extract (NE) was reconstituted with recombinant TFIIE or TFIIE variants carrying mutations in the Tfa1 eWH (M1, Tfa1(N50E/K51E/T52E); M2, Tfa1(N50A/K51A/T52A); M3, Tfa1(P56A/A59E/R62E); M4, Tfa1(Δ E-wing); M5, Tfa1(poly-Ala E-wing) and the Tfa1 E-ribbon (M6, Tfa1(L134E/V137E/L140E); M7, Tfa1(L134A/V137A/L140A) (Extended Data Fig. 3g). RNA products were visualized by primer extension and the mean intensity and standard deviation (s.d.) from triplicate experiments are provided, relative (rel.) to the activity of wild-type TFIIE. An asterisk marks RNA products resulting from an alternative upstream transcription start site.

and subunit dimerization (Extended Data Fig. 3a, refs 16, 38). The structure further indicates that TFIIE must be displaced or at least moved before the elongation factor Spt4/5 (human DSIF) can bind to polymerase^{32,39}.

The OC structure thus reveals how TFIIF and TFIIE bind open promoter DNA from opposite sides of the Pol II cleft. First, TFIIF adopts an extended induced structure that allows it to retain the upstream DNA–TBP–TFIIB assembly on the wall and to bind the DNA bubble and downstream duplex in the cleft (Figs 1b, 2). Second, the TFIIF

Tfg2 WH domain and the TFIIE Tfa2 WH1 domain contact each other above upstream DNA to encircle and retain DNA. Third, the eWH domain of TFIIE binds DNA in the region of initial DNA opening and its long β 1– β 2 hairpin³³ (called here the ‘E-wing’) projects to the upstream edge of the bubble (Fig. 3a), suggesting that the eWH domain stabilizes open DNA. Taken together, the highly modular and flexible basal factors TFIIF and TFIIE undergo substantial induced folding transitions to engage in multiple protein–DNA and protein–protein interactions to stabilize the OC.

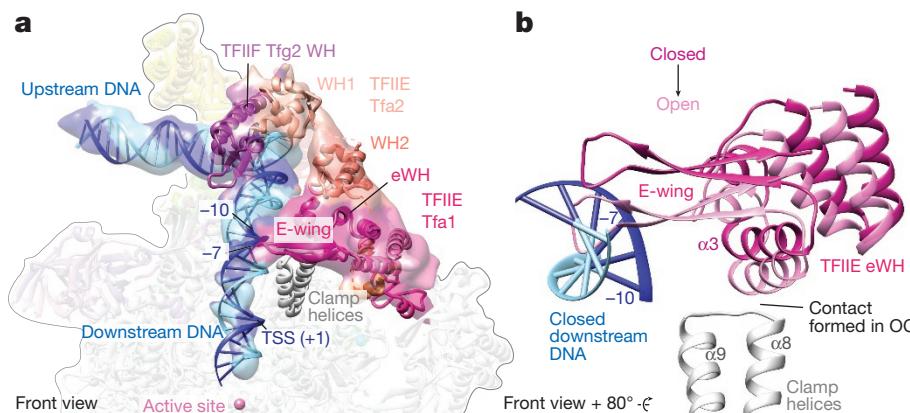


Figure 4 | Closed complex cryo-EM structure. **a**, Details of the closed complex viewed from the front⁵⁰. Highlighted are TFIIE, TFIIF Tfg2 WH, and DNA, superimposed on their density. The promoter DNA displays increased flexibility downstream of the E-wing contact at position –7

upstream of the TSS (+1). **b**, Different positions of the TFIIE eWH in closed (dark magenta) and open (light magenta) complexes, viewed from the top⁵⁰. Compare Extended Data Fig. 4e–g.

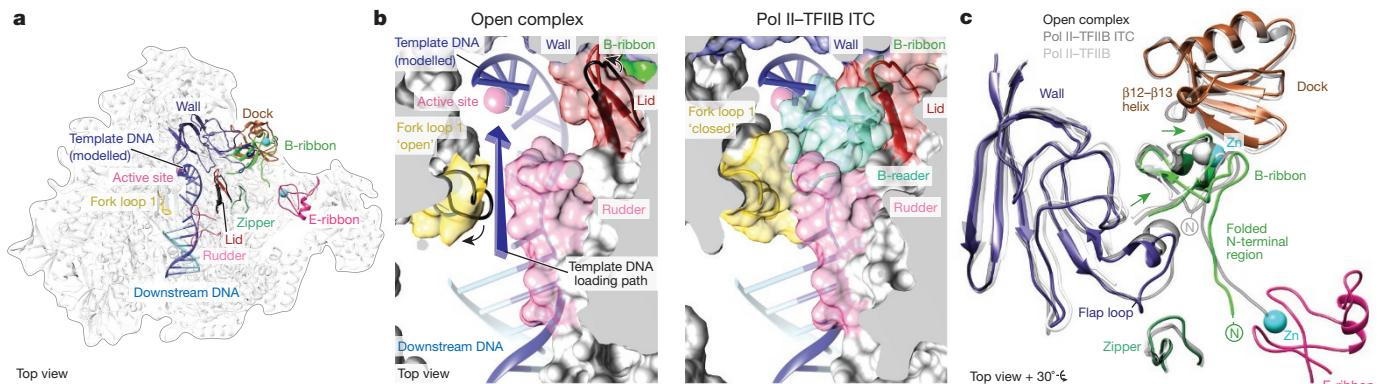


Figure 5 | Cleft clearance and DNA template loading. **a**, OC structure viewed from the top⁵⁰. Highlighted are the Pol II active site (magenta sphere), fork loop 1 (yellow), lid (dark red), rudder (magenta), wall (navy blue), dock (brown), zipper (dark green), TFIIB B-ribbon (green), TFIIE E-ribbon (magenta) and downstream DNA (template, dark blue; non-template, cyan). The template single-strand was modelled using the Pol II-TFIIB ITC¹⁵ crystal structure. **b**, Fork loop 1 and lid assume new positions in the OC compared to the ITC¹⁵ and this opens a path (arrow) for loading of the template DNA strand (blue) into the active site (magenta sphere). Surface representations of Pol II cleft (silver), and cleft elements

fork loop 1, lid, and rudder in the OC (left), and in a Pol II-TFIIB ITC¹⁵ (PDB: 4BBS, right). Movement of the Pol II lid (left, black to dark red) leads to a steric clash with the B-reader (cyan). Compare Extended Data Fig. 5a–c. **c**, Allosteric binding of the TFIIE E-ribbon may lead to an altered position of the TFIIB B-ribbon. Movements in Pol II wall and flap loop (navy blue), dock (brown), zipper (dark green), and B-ribbon (green) are observed in presence of TFIIE compared to the crystal structures of the binary Pol II-TFIIB complex¹³ (dark grey, PDB: 3K1F) and Pol II-TFIIB ITC¹⁵ (light grey, PDB: 4BBS). The altered B-ribbon position may be stabilized by binding to a short helix formed in loop β12–β13 of the dock domain.

DNA opening and loading

Modelling of a closed DNA promoter onto the OC structure shows that closed DNA would clash with the TFIIE eWH domain. This suggests that the eWH domain adopts a different position before DNA opening. To investigate this, and to provide insights into the transition from the CC to the OC, we repeated structure determination with closed DNA instead of pre-opened DNA (Methods, Extended Data Fig. 4). Surprisingly, cryo-EM analysis revealed that about 3 out of 4 particles contained open DNA, although closed DNA was used

for complex preparation. DNA opening occurred in the absence of TFIID (Extended Data Fig. 4h, i). From these particles we obtained an independent reconstruction of the spontaneously formed OC at 6.1 Å resolution (OC5, Extended Data Fig. 4h). Weak density for upstream and downstream DNA segments indicates that DNA bubbles of various sizes formed during DNA opening (Extended Data Fig. 4i). The reconstruction resembles the high-resolution OC structure, suggesting that the latter was not perturbed by the use of pre-opened DNA (Extended Data Fig. 4h).

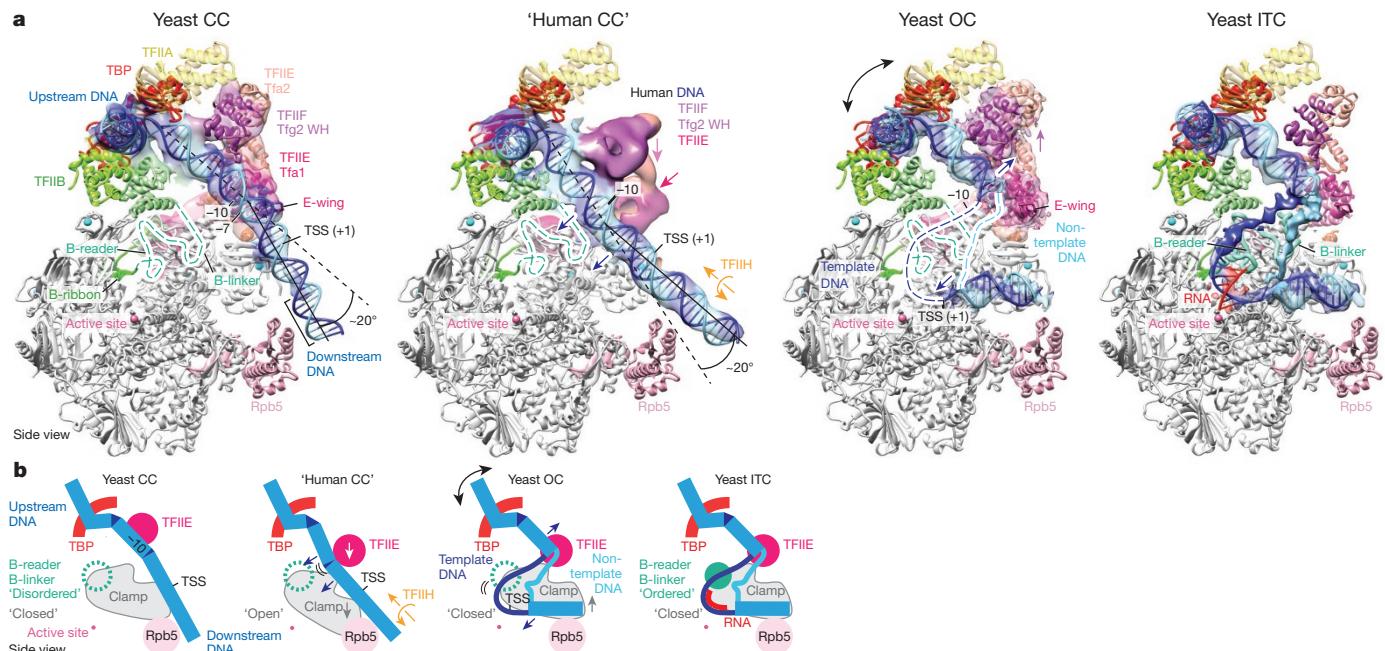


Figure 6 | Model for DNA opening during transcription initiation. **a**, Gallery of initiation complexes depicting proposed movements (arrows) of DNA and basal factors during the transition from the CC to OC to ITC, from left to right, viewed from the side⁵⁰. Yeast CC and OC structures (this work) were complemented with our previous yeast ITC¹⁹ structure (EMD-2785) and an alternative model of the CC ('human CC'), which was obtained by replacing the DNA with that in the human CC²⁰ (EMD-2306),

and adjusting the clamp to the position observed in the human CC. Shown are cryo-EM densities for DNA, Tfg2 WH, and TFIIE. DNA positions –10, –7 (yeast CC) and +1 (TSS) are labelled. DNA was extended by one turn for the yeast CC (black bracket). The locations of TFIID and TFIIF in the ITC were inferred from the yeast OC. Obstructing Pol II and TFIID regions were removed for clarity. **b**, Schematic representation of **a**. Key elements for DNA opening are indicated (compare Extended Data Fig. 6).

From the remaining particle images we obtained a reconstruction of the CC at 8.8 Å resolution (Fig. 4a). Comparison of this CC reconstruction with the OC structure reveals movements mainly in TFIIE (Fig. 4b, Extended Data Fig. 4e–g). In the CC, the E-wing lies on top of the DNA around position –7, in the region where DNA opening begins⁷ (Extended Data Fig. 4e). Consistent with this contact, Tfa1 (or the human counterpart TFIIE α) crosslinks to DNA near this point in the CC^{30,40,41} (Extended Data Fig. 4e). Conversion of the CC to the OC involves movement of upstream DNA and the DNA-associated domains Tfa2 WH1 and Tfg2 WH towards the cleft. DNA opening allows the TFIIE eWH domain to bind the tip of the clamp, and enables the E-wing to move near the upstream edge of the DNA bubble. Consistent with these changes, crosslinks between upstream DNA and the large TFIIE subunit are altered when the CC is converted to the OC⁴⁰. These observations indicate that DNA opening involves TFIIE, and in particular the eWH domain.

TFIIE may also be involved in loading of the DNA template strand into the active centre during the transition from the CC to the OC. In previous structures, the path for loading the template was obstructed by the TFIIB B-reader and the Pol II fork loop 1 and lid. However, the B-reader is mobile in the OC structure, and fork loop 1 and the lid are moved to provide a path for template-strand loading (Fig. 5b, Extended Data Fig. 5a–c). These movements are apparently triggered by TFIIE because binding of the E-ribbon leads to a shift in the B-ribbon that partially withdraws the B-reader from the cleft (Fig. 5c). Thus allosteric binding of the E-ribbon apparently induces ‘clearance’ of the Pol II cleft that may facilitate template-strand loading into the active centre, and transcription start-site scanning in yeast.

To support the proposed functions of structural elements in TFIIE, we prepared recombinant TFIIE variants and tested them for binding to the CC and for promoter-dependent transcription activity in yeast nuclear extract (Methods, Fig. 3c, Extended Data Fig. 3g–i). Mutation of only three surface residues in the E-ribbon that contact the Pol II subunit Rpb7 strongly impaired both binding to the CC and transcription activity, and led to a severe growth defect in yeast (Fig. 3c, Extended Data Fig. 3h, j). Further, the eWH domain is required for TFIIE function¹⁶, and disruption of the eWH contact with DNA by introducing negatively charged glutamate residues in the eWH helix α 3 leads to a transcription defect and impairs binding to the CC (Fig. 3c, Extended Data Fig. 3h). Other mutations in the eWH domain did not show functional defects in our assays. Deletion of the E-wing results in a mild growth phenotype (Extended Data Fig. 3j), and does not impair *in vitro* transcription (Fig. 3c), maybe because TFIIH compensates for the loss of E-wing function in these assays.

Model of transcription initiation

Structural comparisons of the yeast CC, OC, and ITC¹⁹ with each other and with the highly conserved human CC²⁰ reveal differences in the positions of DNA, the clamp, and TFIIE, and lead to an extended model of DNA opening (Fig. 6a, b, Extended Data Fig. 6). In this model, promoter DNA is initially bent away from the active site by ~20° near position –10 at the tip of the clamp helices (yeast CC). The clamp then opens slightly (Extended Data Fig. 6b), allowing promoter DNA to bend in the opposite direction and to enter the upper part of the cleft (‘human CC’). This frees the site at the tip of the clamp helices that can now bind the TFIIE eWH domain. DNA can then no longer swing back to its initial position, because this would result in a steric clash with the repositioned eWH domain. However, when DNA opening occurs around position –10, the upstream DNA can swing back to its original position and this stabilizes the DNA duplex single-strand junction. TFIIH then rotates downstream DNA and pushes the template single strand into the cleft using its ATP-dependent translocase activity, as suggested previously¹⁶. When the bubble extends downstream⁵, the template strand is loaded into the active centre cleft via a path cleared by allosteric binding of TFIIE (yeast OC). Template-strand loading allows the clamp to close again and to trap downstream

DNA in the cleft (Extended Data Fig. 6b). The B-reader then covers the template strand, and helps to detect the transcription start site¹³, triggering RNA synthesis (yeast ITC).

This model explains how DNA opening can be achieved with the use of binding energy alone, at least at some promoters^{11,12}. DNA opening allows for new protein interactions at the upstream DNA duplex single-strand junction involving the eWH domain, the TFIIF arm, and the Pol II clamp. DNA loading into the cleft also enables new interactions of the downstream DNA duplex with the TFIIF charged helix and the Pol II cleft and clamp. These additional contacts may compensate for the energy needed for DNA melting and help to trap open DNA and to prevent its re-closure.

A similar mechanism for ATP-independent DNA opening may be used in other transcription systems. Proteins with homologies to TFIIB, TFIIE, and TFIIF, but not TFIIH, are present in the Pol I and Pol III systems⁴², and counterparts of TBP, TFIIB, and TFIIE are found in archaea³⁸. The bacterial transcription initiation system is structurally unrelated, but conceptually similar. DNA opening occurs spontaneously above the cleft, the open DNA is trapped with the use of binding energy, and this requires clamp opening and closure^{43–47}. DNA opening at a subset of bacterial genes however requires the ATP-dependent initiation factor σ ⁵⁴, and this enables further regulation⁴⁸. Similarly, Pol II regulation can occur at the level of DNA opening⁴⁹, and TFIIH is required to keep DNA open at least at selected promoters⁸. This suggests that the Pol II initiation system evolved to depend on the ATP-consuming factor TFIIH for DNA opening, probably in response to an expanded need for gene regulation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 November 2015; accepted 8 April 2016.

Published online 11 May 2016.

- Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
- Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
- Grünberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* **38**, 603–611 (2013).
- Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 129–143 (2015).
- Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
- Ohkuma, Y. & Roeder, R. G. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368**, 160–163 (1994).
- Maxon, M. E., Goodrich, J. A. & Tjian, R. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIH: a model for promoter clearance. *Genes Dev.* **8**, 515–524 (1994).
- Holstege, F. C., van der Vliet, P. C. & Timmers, H. T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J.* **15**, 1666–1677 (1996).
- Parvin, J. D., Timmers, H. T. & Sharp, P. A. Promoter specificity of basal transcription factors. *Cell* **68**, 1135–1144 (1992).
- Pan, G. & Greenblatt, J. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J. Biol. Chem.* **269**, 30101–30104 (1994).
- Buratowski, S., Sopta, M., Greenblatt, J. & Sharp, P. A. RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *Proc. Natl. Acad. Sci. USA* **88**, 7509–7513 (1991).
- Holstege, F. C., Tantin, D., Carey, M., van der Vliet, P. C. & Timmers, H. T. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* **14**, 810–819 (1995).
- Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
- 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).
- Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437–440 (2013).
- Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* **19**, 788–796 (2012).

17. Murakami, K. et al. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).
18. Mühlbacher, W. et al. Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun.* **5**, 4310 (2014).
19. Plaschka, C. et al. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376–380 (2015).
20. He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
21. Murakami, K. et al. Structure of an RNA polymerase II preinitiation complex. *Proc. Natl. Acad. Sci. USA* **112**, 13543–13548 (2015).
22. Nikolov, D. B. et al. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
23. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830–836 (1996).
24. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–134 (1996).
25. Eichner, J., Chen, H.-T. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* **29**, 706–716 (2010).
26. Deng, W. & Roberts, S. G. E. TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* **116**, 417–429 (2007).
27. Fishburn, J. & Hahn, S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Mol. Cell. Biol.* **32**, 12–25 (2012).
28. Čabart, P., Újvári, A., Pal, M. & Luse, D. S. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proc. Natl. Acad. Sci. USA* **108**, 15786–15791 (2011).
29. Robert, F., Forget, D., Li, J., Greenblatt, J. & Coulombe, B. Localization of subunits of transcription factors IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. *J. Biol. Chem.* **271**, 8517–8520 (1996).
30. Forget, D., Langlier, M.-F. F., Thérien, C., Trinh, V. & Coulombe, B. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell. Biol.* **24**, 1122–1131 (2004).
31. Chen, H.-T. T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat. Struct. Mol. Biol.* **14**, 696–703 (2007).
32. Grohmann, D. et al. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43**, 263–274 (2011).
33. Meinhart, A., Blobel, J. & Cramer, P. An extended winged helix domain in general transcription factor E/IIE alpha. *J. Biol. Chem.* **278**, 48267–48274 (2003).
34. Okuda, M. et al. A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. *J. Biol. Chem.* **279**, 51395–51403 (2004).
35. Okuda, M. et al. Structural insight into the TFIIE-TFIIF interaction: TFIIE and p53 share the binding region on TFIIFH. *EMBO J.* **27**, 1161–1171 (2008).
36. Okuda, M. et al. Structure of the central core domain of TFIIE β with a novel double-stranded DNA-binding surface. *EMBO J.* **19**, 1346–1356 (2000).
37. Okamoto, T. et al. Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIE β subunit. *J. Biol. Chem.* **273**, 19866–19876 (1998).
38. Blombach, F. et al. Archaeal TFE α/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife* **4**, e08378 (2015).
39. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
40. Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000).
41. Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat. Struct. Mol. Biol.* **13**, 603–610 (2006).
42. Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
43. Chakraborty, A. et al. Opening and closing of the bacterial RNA polymerase clamp. *Science* **337**, 591–595 (2012).
44. Feklistov, A. & Darst, S. A. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**, 1257–1269 (2011).
45. Zhang, Y. et al. Structural basis of transcription initiation. *Science* **338**, 1076–1080 (2012).
46. Zuo, Y. & Steitz, T. A. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* **58**, 534–540 (2015).
47. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S. A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* **4**, (2015).
48. Yang, Y. et al. Structures of the RNA polymerase- σ^{54} reveal new and conserved regulatory strategies. *Science* **349**, 882–885 (2015).
49. Kouzine, F. et al. Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988–999 (2013).
50. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).

Acknowledgements We thank C. Bernecke, W. Mühlbacher, S. Neyer, S. Sainsbury, and D. Tegunov for help and discussions; L. Larivière and L. Wenzke for cloning and initial purification of TFIIE; W. Mühlbacher for initial cloning of TFIIA; S. Bilakovic for the modified pET-DUET-1 vector; J. Mahamid for help with data collection for the CC; K. Maier for help with yeast growth assays; M. Raabe and H. Urlaub for protein identification; K. Kinkel in for initial Pol II-TFIIF co-crystallization; and S. Hahn for providing the *TFA1* yeast strain and the shuffle plasmid pSH810. C.P. (SFB860), M.H. (GRK1721), and P.C. were supported by the Deutsche Forschungsgemeinschaft, the Advanced Grant TRANSIT of the European Research Council, and the Volkswagen Foundation.

Author Contributions C.P. designed and carried out high-resolution cryo-EM structure determinations of OC1–OC4. M.H. designed and carried out Pol II-TFIIF crystallographic analysis, and cryo-EM structure determinations of OC5 and CC. C.P. and M.H. designed and carried out functional assays. C.D. cloned and purified full-length TBP and TFIIA. C.D. and C.B. assisted with protein purification. J.P. supervised electron microscopy data collection. P.C. designed and supervised research. C.P., M.H., and P.C. prepared the manuscript.

Author Information Three-dimensional cryo-EM density maps of OC1, OC2, OC2-focused, OC3, OC3-focused, OC4, OC4-focused, OC5, and CC have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-3375, EMD-3376, EMD-3377, EMD-3378, EMD-3379, EMD-3380, EMD-3381, EMD-3382, and EMD-3383, respectively. Coordinate files of the OC and CC have been deposited in the Protein Data Bank under accession numbers 5FYW and 5FZ5. Coordinates and structure factors of the Pol II-TfI peptide and the Pol II-TFIIF crystals have been deposited at the Protein Data Bank under the accession numbers 5IP7 and 5IP9. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (patrick.cramer@mpibpc.mpg.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Vectors and sequences. The open reading frame (ORF) encoding full-length TBP was amplified from *Saccharomyces cerevisiae* (*Sc*) genomic DNA and cloned into a pOPINE vector, containing a C-terminal 6×histidine tag. Codon-optimized Tfg1, Toa1, and Toa2 were commercially obtained for expression in *Escherichia coli* (*E. coli*, Life Technologies). Toa1 (Δ 95–209) was obtained by quick-change PCR and cloned with Toa2 into a pOPINE vector, containing a C-terminal 6×histidine-tag on Toa2. Tfa1 and Tfa2 ORFs were amplified from genomic *Sc* DNA and cloned sequentially into a pET vector. ORFs for Tfg1 and Tfg2 were cloned sequentially into a modified pET-Duet-1 vector, containing an N-terminal 10×histidine-8×arginine-SUMO-tag on Tfg1. Additional ribosomal binding sites were introduced as described⁵¹. For heterologous co-expression of *Sc* 16-subunit core Mediator (cMed)-Med1, 13-subunit cMed lacking Med4-Med9 was cloned with previously described vectors¹⁹. Three-subunit Med1-Med4-Med9 was prepared by cloning *Sc* ORFs of Med1, Med4, and Med9 into a modified pET-Duet-1 vector, containing an additional N-terminal 10×histidine-8×arginine-SUMO-tag on Med1. Sequences are available upon request.

Recombinant proteins. All proteins were expressed in *E. coli* BL21(DE3)RIL cells (Stratagene). The identity of all purified proteins was confirmed by mass spectrometry. All purified proteins and complexes were flash-frozen and stored at –80 °C.

Full-length TBP was expressed in *E. coli* that were grown in lysogeny broth (LB) medium at 37 °C to an optical density (OD) of 0.5 at 600 nm, and induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) for 4 h at 20 °C. Cells were lysed by sonication in buffer A (25 mM HEPES (pH 7.5), 500 mM KCl, 10% glycerol, 2.5 mM dithiothreitol (DTT)) containing protease inhibitors⁵². The soluble fraction was applied to a 5 ml HisTrap HP column, washed with buffer A containing 1 M KCl, and eluted with buffer A containing 350 mM imidazole. The sample was diluted 1:4 with buffer A lacking KCl and applied to anion-exchange chromatography using a MonoS 5/50 (GE Healthcare), and eluted with a linear gradient of buffer A from 100–1000 mM KCl. TBP was further purified by size-exclusion chromatography using a Superose 12 10/300 (GE Healthcare) column, equilibrated in buffer B (10 mM HEPES (pH 7), 200 mM NaCl, 5% glycerol, 2 mM DTT). TBP-containing fractions were pooled and concentrated to 8 mg ml^{−1}.

TFIIA was obtained by co-expression of its subunits Toa1 and Toa2 in *E. coli*. Transformed cells were grown in LB medium at 37 °C to an OD of 0.5 at 600 nm. Expression was induced with 0.5 mM IPTG at 37 °C for 4 h. Cells were lysed in buffer C (25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% glycerol, 2 mM DTT) containing protease inhibitors⁵², and cleared by centrifugation. The supernatant was applied to a 5-ml HisTrap HP column, washed with buffer C containing 1 M NaCl, and eluted with buffer C containing 250 mM imidazole. Fractions containing the complex were pooled, diluted 1:4 with buffer C lacking NaCl, loaded on a MonoQ 5/50 anion-exchange column, and eluted with a linear gradient of buffer C from 100–500 mM NaCl. TFIIA containing fractions were pooled and applied to a Superdex 75 10/300 column (GE Healthcare), in buffer B. TFIIA was concentrated to 8 mg ml^{−1}. TFIIB was prepared as described¹⁵.

Recombinant TFIIE was obtained by co-expression of its subunits Tfa1 and Tfa2 in *E. coli*. Cells were transformed and grown in LB medium at 37 °C to an OD of 0.6 at 600 nm, and expression was induced by 0.5 mM IPTG for 18 h at 18 °C. Cells were lysed by sonication in buffer D (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 0.02% Tween-20, 5 mM DTT) containing protease inhibitors⁵². The lysate was cleared by centrifugation and applied to a 5 ml HisTrap HP column, equilibrated in buffer E (buffer M, lacking Tween-20, containing 10 mM imidazole). The column was washed with 10 column volumes of buffer E and eluted with buffer E containing 250 mM imidazole. TFIIE was then subjected to anion-exchange chromatography using a 5-ml Hi-Trap Heparin column (GE Healthcare), equilibrated in buffer F (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 2% glycerol, 5 mM DTT). The complex was eluted with a linear gradient of buffer F from 100–2000 mM NaCl. To improve purity, TFIIE was further applied to a Superose 12 10/300 size-exclusion column, in buffer G (5 mM HEPES (pH 7.25), 40 mM ammonium sulphate, 10 μM ZnCl₂, 10 mM DTT). TFIIE containing fractions were pooled and concentrated to 9.6 mg ml^{−1}.

Sc TFIIF subunits Tfg1 and Tfg2 were co-expressed in *E. coli* and cells were grown in LB medium at 37 °C to an OD of 0.8 at 600 nm. Expression was induced with 0.2 mM IPTG for 3 h at 37 °C. Cells were lysed by sonication in buffer H (50 mM HEPES (pH 7.0), 350 mM KCl, 10% glycerol, 2 mM DTT) supplemented with 50 mM imidazole and protease inhibitors⁵². Cleared lysate was applied to a 5-ml HisTrap HP column equilibrated in buffer H. The column was washed with 8 column volumes of buffer H containing 1000 mM KCl, and eluted with a linear gradient from buffer H to buffer I (50 mM HEPES (pH 7.0), 250 mM KCl, 800 mM

imidazole, 10% glycerol, 2 mM DTT). The conductivity of the eluate was adjusted to match that of buffer J (50 mM HEPES (pH 7.0), 150 mM KCl, 10% glycerol, 2 mM DTT) and 3C protease cleavage was carried out for 2 h. The complex was then applied to cation-exchange chromatography using a 1-ml HiTrap SP HP column (GE Healthcare), equilibrated in buffer J, and eluted in a linear gradient from 150–1000 mM KCl. TFIIF was further purified by size-exclusion chromatography using a Superdex 200 10/300 Increase column (GE Healthcare), in buffer K (10 mM MES (pH 6.2), 150 mM KCl, 10% glycerol, 2 mM DTT). Purified TFIIF was concentrated to 3.9 mg ml^{−1}. For previous studies of the initiation complex we used the conserved *S. mikatae* Tfg1 (refs 18, 19), owing to difficulties in recombinant expression of its *Sc* homologue. Here, expression of *Sc* Tfg1 was enabled by codon optimization.

For preparation of recombinant 16-subunit cMed-Med1, 13-subunit cMed lacking Med4-Med9 was prepared essentially as described¹⁹. Separately, a plasmid containing Med1-Med4-Med9 was transformed in *E. coli* and cells were grown in LB medium at 37 °C to an OD of 0.6 at 600 nm. Protein expression was induced with 0.5 mM IPTG at 18 °C for 24 h. Cells were collected and lysed by sonication in buffer L (25 mM HEPES (pH 7.5), 400 mM potassium acetate, 10% glycerol, 20 mM imidazole (pH 8), 2 mM DTT) containing protease inhibitors⁵². Lysate was cleared by centrifugation and applied to a 5-ml HisTrap HP column (GE Healthcare), equilibrated in buffer M (25 mM HEPES (pH 7.5), 400 mM potassium acetate, 10% glycerol, 30 mM imidazole (pH 8), 2 mM DTT). The column was washed with 5 column volumes of buffer M and a linear gradient over 10 column volumes from 30–100 mM imidazole. The heterotrimer was eluted with 300 mM imidazole over 10 column volumes. Fractions containing the complex were diluted 1:3 with buffer N (25 mM HEPES (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM EDTA, 5 mM DTT) and incubated with 3C protease for 2 h on ice, to remove the affinity tag. The complex was further purified by anion-exchange chromatography using a MonoQ 5/50 GL column (GE Healthcare), equilibrated in buffer N, and was eluted with a linear gradient from 100–600 mM KCl over 150 column volumes. Fractions containing the complex were pooled and applied to size-exclusion chromatography using a Superose 6 10/600 column (GE Healthcare), equilibrated in buffer O (25 mM HEPES (pH 7.5), 200 mM KCl, 5 mM DTT). Purified Med1-Med4-Med9 complex was concentrated to 2.7 mg ml^{−1}. To prepare 16-subunit cMed-Med1, 13-subunit cMed was incubated with a twofold molar excess of Med1-Med4-Med9 for 30 min at 25 °C and 30 min on ice, and purified by size-exclusion using a Superose 6 10/600 column, in buffer P (25 mM HEPES (pH 7.5), 400 mM KCl, 5% glycerol, 5 mM DTT). Fractions containing cMed-Med1 were pooled and concentrated to 2 mg ml^{−1}.

Preparation of initiation complexes. Yeast 12-subunit Pol II was prepared as described⁵³. Open and closed initiation complexes were prepared with several modifications of the previous cITC-cMed assembly scheme¹⁹ and with a different nucleic acid scaffold for the closed complex. The 72 nucleotide nucleic acid scaffold previously used to prepare the Pol II core ITC (cITC)¹⁹ contains a 15 nucleotide mismatch transcription bubble and six nucleotide RNA, and was used for assembly of the open initiation complex (OC). The closed complex (CC) contained a nucleic acid scaffold with 13 nucleotide longer downstream DNA, based on the *HIS4* promoter (template, 5'-TGATATTTTATGTATGTACA ACACACATCGGAGGTGAATCGAACGTTCCATAGCTATTATACACAGCG TGCTACTGTTCTCG-3'; non-template, 5'-CGAGAACAGTAGCACGCTGTGT ATATAATAGCTATGGAACGTTCGATTACCTCCGATGTGTGTACATA CATAAAAATATCA-3'). The initiation complex was prepared as follows. Pol II (200 μg at 3 mg ml^{−1}) was incubated with a fourfold molar excess of TFIIF. A two-fold molar excess of nucleic acid scaffold over Pol II, tenfold molar excess of TFIIA, fourfold molar excess of TBP and TFIIB were added to buffer Q (25 mM HEPES-KOH (pH 7.5), 150 mM potassium acetate, 5% glycerol, 2 mM MgCl₂, 5 mM DTT) and incubated with pre-formed Pol II-TFIIF complex for 8 min at 25 °C. TFIIE was added in a tenfold molar excess over Pol II and incubated for 5 min at 25 °C. cMed-Med1 was added in a 1.2-fold molar excess over Pol II and incubated for 50 min at 25 °C. Open and closed complexes were purified using a Superose 6 3.2/300 size exclusion column (GE Healthcare), equilibrated in buffer Q. Fractions containing the complex were pooled (0.4–0.8 mg ml^{−1}) and additionally incubated with equimolar amounts of nucleic acid scaffold. The sample was then crosslinked for 30 min on ice using 0.1% glutaraldehyde (Electron Microscopy Sciences), and the reaction was quenched with 50 mM lysine (Sigma). The crosslinked sample was re-purified in a second size-exclusion step using a Superose 6 3.2/300 column, equilibrated in buffer Q lacking glycerol. Fractions containing initiation complexes were pooled (0.2–0.6 mg ml^{−1}) and used for EM grid preparation.

Electron microscopy. Initiation complex samples were applied to R3.5/1 holey carbon grids (Quantifoil). Grids were glow-discharged for 15 s before deposition of 4.5 μl complex, and subsequently blotted and vitrified by plunging into

liquid ethane with a Vitrobot Mark IV (FEI) operated at 4 °C and 100% humidity. Cryo-EM data was acquired on a FEI Titan Krios operated in ETEM mode at 300 keV, and equipped with a K2 Summit direct detector (Gatan). Automated data collection was carried out using the TOM toolbox⁵⁴ to acquire 1756 movies of the OC with a range of defocus values (from $-0.7\text{ }\mu\text{m}$ to $-4.2\text{ }\mu\text{m}$) at a nominal magnification of $37,000\times$ (1.35 Å per pixel). The camera was operated in ‘super-resolution’ mode (0.675 Å per pixel), with a total exposure time of 10 s fractionated into 25 frames, a dose rate of $\sim 5\text{ e}^{-}$ per pixel per second, and total dose of $33\text{ e}^{-}\text{ \AA}^{-2}$ per movie. Cryo-EM data of the CC was collected in the same manner, except that 959 movies were acquired with a defocus range from $-0.8\text{ }\mu\text{m}$ to $-5\text{ }\mu\text{m}$, a total exposure time of 6 s with 20 frames, a dose rate of $\sim 8\text{ e}^{-}$ per pixel per second, and total dose of $40\text{ e}^{-}\text{ \AA}^{-2}$. Movies were aligned as described^{19,55}, except that images were not partitioned into quadrants.

Image processing. For single particle analysis of the OC, an initial set of 10,225 particles was selected semi-automatically using e2boxer.py from EMAN2 (ref. 56). CTF parameters were estimated using CTFFIND4 (ref. 57). CTF correction and subsequent image processing was performed with RELION 1.3 (ref. 58), unless otherwise noted. Resolution was reported on the basis of the gold-standard Fourier shell correlation (FSC) (0.143 criterion) as described previously⁵⁹ and temperature factors were determined and applied automatically in RELION⁵⁸. Selected particles were extracted with a 300^2 pixel box and pre-processed to normalize images. Reference-free two-dimensional (2D) class averages were calculated, and twelve representative classes were low-pass filtered to 25 Å resolution and used as templates for automated picking⁶⁰ of all micrographs. The resulting 415,030 particle images were screened manually and by reference-free 2D classification, yielding 257,259 particle images that were used for subsequent processing. The 7.8 Å cryo-EM map of the yeast cITC¹⁹ (EMD-2785) was low-pass filtered to 50 Å and used as initial model for 3D refinement of the 10,225 particle set. This revealed an OC density to an estimated resolution of 10 Å. This density was low-pass filtered to 50 Å and used for processing of the complete OC cryo-EM data set (Extended Data Fig. 7a–d). A 3D reconstruction of all particles was calculated to 3.75 Å, and subjected to particle polishing using RELION 1.4 beta⁵⁸. This lead to an improved density at 3.58 Å resolution.

Hierarchical 3D classification was carried out without image alignment, to reduce computational requirements and identify homogeneous single particle groups (Extended Data Fig. 7a–d). Soft masks encompassing the complete OC or smaller regions of Pol II and basal factors were generated using the volume eraser in UCSF Chimera⁶¹ and RELION⁵⁸. This included masks for Pol II stalk-TFIIE E-ribbon (Extended Data Fig. 7a, c), upstream DNA-TBP-TFIIA-TFIIB-Tfg2 linker (Extended Data Fig. 7b), TFIIE-Tfg2 WH (Extended Data Fig. 7c), TFIIF (Extended Data Fig. 7d), and TFIIF dimerization domain (Extended Data Fig. 7d). Each class was refined using the 3D auto-refine procedure against the respective particles within that class with a soft reference mask in the shape of the OC (maximum diameter of 270 Å), generated in RELION⁵⁸. The OC1 reconstruction (improved Pol II core, B-ribbon, and E-ribbon density) was determined from 102,876 particles to a resolution of 3.58 Å and with a temperature factor of -111 \AA^2 (Extended Data Fig. 8a). The OC2 reconstruction (improved upstream DNA, TBP, TFIIA, TFIIB, and Tfg2 linker density) was determined from 17,282 particles to a resolution of 3.97 Å and with a temperature factor of -95 \AA^2 (Extended Data Fig. 8b). The OC3 reconstruction (improved TFIIE density) was determined from 11,231 particles to a resolution of 4.35 Å and with a temperature factor of -125 \AA^2 (Extended Data Fig. 8c). The OC4 reconstruction (improved TFIIF dimerization and Tfg1 arm density) was determined from 29,455 particles to a resolution of 3.89 Å and with a temperature factor of -79 \AA^2 (Extended Data Fig. 8d). Focused refinement of the upstream DNA assembly (OC2-focused) was achieved by continuing the auto-refinement (round 1, class 4; Extended Data Fig. 7c) from the first round of local searches using a corresponding soft mask. Focused refinement of TFIIE-Tfg2 WH (OC3-focused) was achieved by continuing the OC3 auto-refinement from the first round of local searches using a soft mask encompassing TFIIE, Tfg2 WH, and the interacting segment of upstream DNA. Focused refinement of TFIIF dimerization domain (OC4-focused) was achieved by continuing the auto-refinement (round 2, class 2; Extended Data Fig. 7d) from the first round of local searches using a soft mask encompassing the TFIIF dimerization domain and the Pol II lobe. The resolution of focused refinements was determined using a soft mask with a 30-pixel soft edge⁶², to 4.7 Å, 7.5 Å, and 4.09 Å for OC2-, OC3-, and OC4-focused refinements respectively (Extended Data Fig. 8b, c, d).

Single particle cryo-EM analysis of the closed complex was carried out essentially as for the open complex, with the following differences. Reference-free 2D classification of 10,591 particles, picked semi-automatically using e2boxer.py from EMAN2⁵⁶, gave three representative 2D class averages that were low-pass filtered to 30 Å resolution and used as templates for automated particle picking in RELION of all micrographs⁶⁰. The resultant 155,079 particles were screened manually and by reference-free 2D classification, yielding 111,625 particles that were used for

subsequent processing. The cITC (EMD-2785), low-pass filtered to 50 Å, was used as the reference model for the initial 10,591 particle set, and the resultant 3D reconstruction (14 Å resolution), again low-pass filtered to 50 Å, was used as initial reference model for 3D refinement using the full cryo-EM data set. This revealed the CC at 7.5 Å, and was improved by particle polishing to 6.5 Å resolution using RELION 1.4 beta⁵⁸. Hierarchical 3D classification without image alignment using a soft mask encompassing the complete CC resulted in two populations, CC and OC5, respectively (Extended Data Fig. 7e). Both classes were refined using the 3D auto-refine procedure against the respective particles within that class with a soft mask in the shape of the CC. The CC soft mask was also suitable for the OC as both complexes have a similar shape. The CC reconstruction was determined from 7,527 particles to a resolution of 8.2 Å. The OC5 reconstruction was determined from 79,797 particles to a nominal resolution of 6.1 Å and a temperature factor of -176 \AA^2 was applied (Extended Data Fig. 8g). To improve the densities TFIIE, Tfg2 WH and downstream DNA in the CC, we carried out focused classification of the Pol II stalk and subsequently TFIIE, Tfg2 WH and downstream DNA using soft masks encompassing these regions. The individual classes were refined as before and yielded a CC at 8.8 Å comprising 5,690 particles and a temperature factor of -300 \AA^2 was applied (Extended Data Fig. 8f). A similar focused classification scheme for OC5 revealed moderately improved density for TFIIE (not shown).

Local resolution estimates were determined using a sliding window of 40^3 voxels as previously described except that a single pair two half-maps was used and resolution estimates were not capped at the nominal resolution¹⁹ as no local filters were applied.

Structural modelling. A composite model of the OC was obtained using cryo-EM densities OC1–OC4 and the focused refinements of OC2, OC3, and OC4 particles. Structural models were built in COOT⁶³ unless indicated otherwise. Models were refined using the real space refinement routine in Phenix⁶⁴ into the respective OC density, as indicated, with secondary structure and rotamer restraints. First, structural models of Pol II (lacking Rpb4–Rpb7), TFIIB B-ribbon (residues 22–59), and downstream DNA (PDB: 4V1N¹⁹) were placed into the OC1 map using UCSF Chimera⁶¹, followed by rigid-body group refinement in Phenix⁶⁴. Models of the Pol II core and TFIIB B-ribbon were adjusted and extended manually, and refined in Phenix⁶⁴ into the OC1 density. The Rpb4–Rpb7 structure (PDB: 4V1N) was fitted, and residues at the base of the stalk (Rpb7 1–81, 150–159) were modified to fit the density and refined into the OC1 map. TFIIE was modelled using structural information on eWH³³, E-ribbon³⁴, and WH1³⁶, and a homology model for WH2 (Fig. 3a, b). A homology model of the human TFIIE α E-ribbon (PDB: 1VD4 (ref. 34)) was generated using MODELLER⁶⁵ (Tfa1 residues 121–158) and rigid-body fitted into the unsharpened OC1 density. Upstream DNA (PDB: 4V1N), TFIIB N-terminal cyclin (PDB: 4BBR, chain M and residues 122–213), the homology model of TFIIB C-terminal cyclin domain (PDB: 4V1N, chain M and residues 233–343) were fitted into the OC2 cryo-EM density. Models of Pol II protrusion, wall and the TFIIB N-terminal cyclin were adjusted and extended manually, and the TFIIB C-terminal cyclin (residues 236–328) and Tfg2 linker regions (residues 249–280) were built into OC2 and OC2-focused maps and refined into the OC2 map in Phenix⁶⁴. X-ray structures of TBP (PDB: 1YT⁶⁶, chain A), and TFIIA (PDB: 1YTF²⁵, chains B–D) were individually fitted into the OC2 map obtained from focused refinement (OC2-focused). The fit of upstream DNA, from the upstream edge of the DNA bubble to TBP, was improved in UCSF Chimera⁶¹ using the OC3 map. Protein models of Tfa1 eWH (residues 1–90) and Tfa2 WH2 (residues 187–249), generated with the I-TASSER prediction server⁶⁷, were fit into the OC3 density and adjusted. The TFIIE linker helices were built *de novo* and modified to match the density, and together with Tfa1 eWH and Tfa2 WH2 models were refined into the unsharpened OC3 map in Phenix⁶⁴ and subjected to the phenix.model_idealization routine to optimize geometry. Published NMR data of the Tfg2 WH domain⁶⁸ (BMRB accession number 17916) was used to calculate a low-energy model with the BMRB CS-ROSETTA server⁶⁹. The Tfg2 WH and an I-TASSER model of Tfa1 WH1 (residues 123–181) were placed into the OC3 density obtained from focused refinement (OC3-focused). The Tfg2 WH positioning is also consistent with NMR data for the Tfg2 WH–DNA interface⁶⁸. A part of the Tfg1 arm domain (residues 327–349) was built into OC3 and OC4 densities and refined into the OC4 density. A previous homology model of the TFIIF dimerization domain¹⁸ was used as a starting template for modelling into OC4 and OC4-focused density maps, and was refined into the OC4 density. The crystal structure of the Tfg1 N-terminal peptide (residues 21–35, see below) was fitted into an OC4 density (round 2 class 2, Extended Data Fig. 7d). The individually refined models showed good stereochemistry and were validated with Molprobity⁷⁰, and the FSC of map versus model (Extended Data Figs 8 and 9).

To generate the CC model, the OC model of Pol II, TFIIA, TFIIB, TBP, TFIIE E-ribbon and TFIIF dimerization domain was rigid-body fitted into the CC density

using an automated global 6D correlation search in Situs⁷¹. The remaining protein part was divided into two rigid bodies containing (i) Tfa1 and Tfa2 WH2, and (ii) Tfa2 WH1 and Tfg2 WH, which were independently fitted into the CC density using Situs⁷¹. To model closed promoter DNA, upstream DNA was extended with canonical duplex B-form DNA in COOT⁶³ and rigid-body fitted in UCSF Chimera⁵⁷ to reflect the density.

All figures were generated using UCSF Chimera⁶¹.

Crystallographic analysis of Pol II–TFIIF complex. The structure of the Tfg1 N-terminal region (residues 21–35) bound to Pol II was determined by X-ray crystallographic analysis. 12-subunit Pol II was prepared and crystallized⁵³, and TFIIF or selenomethionine (SeMet)-substituted Tfg1 peptide (residues 19–41, Peptide Speciality Laboratories GmbH, Heidelberg) was added to the cryo-protectant as described⁷². Diffraction data were collected on a PILATUS 6M detector at the X06SA beamline (SLS, Villigen, Switzerland) at 100 K. Data for the Pol II–Tfg1 SeMet peptide crystal was collected at a wavelength of 0.97972 Å and for the Pol II–TFIIF crystal at 0.91889 Å. Data were processed with XDS and XSCALE⁷³. The structure was phased with the crystal structure of the 12-subunit Pol II (PDB: 3PO2) lacking nucleic acid and refined using BUSTER⁷⁴. Model building of the TFIIF N-terminal residues in COOT⁶³ was guided by a selenium anomalous difference Fourier peak (M27). Subsequent refinement in BUSTER used secondary structure restraints for the Tfg1 peptide helix. The final structures had an R_{free} factor of 18.0% and 19.4% for the Pol II–Tfg1 SeMet peptide crystal and the Pol II–TFIIF crystal, respectively (Extended Data Fig. 9c), and showed good stereochemistry⁷⁰. In the Pol II–Tfg1 SeMet peptide structure 90% of the residues fall in favoured regions of the Ramachandran plot and 3% in disallowed regions. For the Pol II–TFIIF structure 89% of the residues fall in favoured regions of the Ramachandran plot and 3% in disallowed regions.

Yeast and functional assays. TFIIE interfaces with Pol II and DNA were probed by *in vivo* mutagenesis. The yeast strain used for *TFA1* genetic assays¹⁶ was provided by S. Hahn (Fred Hutchinson Cancer Research Center). Mutations were introduced into a plasmid encoding *TFA1* (pSH810, ARS CEN LEU2 3×Flag), as indicated in Extended Data Fig. 3g. The *tfa1*(ΔE-wing) mutant was generated by deletion of *TFA1* residues 71–84, and their replacement by the residues GSG. The *TFA1* constructs were transformed into the shuffle strain, and were streaked twice on –Ura –Leu plates, and subsequently onto yeast extract-peptone-dextrose (YPD) plates. Yeast were freshly grown in YEPD medium and resuspended in water to an OD of 1 at 600 nm, and tenfold dilutions were spotted on 5-fluoroorotic acid (5-FOA) and YEPD plates, and incubated at 30 °C.

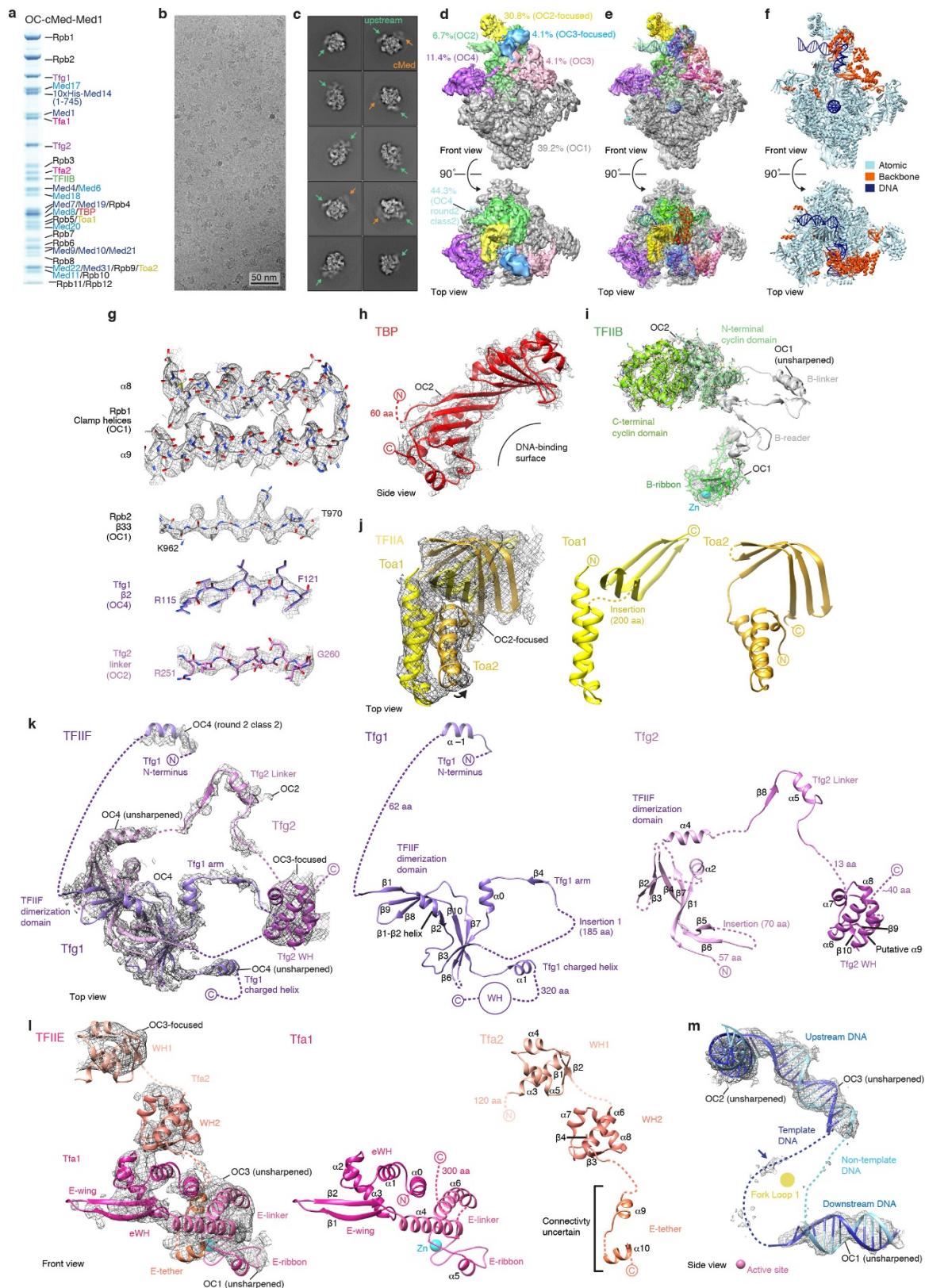
To further characterize the interfaces between TFIIE and the initiation complex, recombinant TFIIE mutants were purified according to the same protocol as wild-type TFIIE (Recombinant proteins, Extended Data Fig. 3g, h). The Tfa1(poly-Ala E-wing) mutant was generated by replacement of *TFA1* residues 71–84 with poly-Alanine.

To assess the interaction of wild-type TFIIE and the TFIIE mutants with the CC by protein pulldown, 3 µg purified Pol II was first biotinylated on the Rpb3 subunit as described¹⁹. The CC, containing biotinylated Pol II, was subsequently prepared as above ('preparation of initiation complexes'), but without TFIIE. This preparation was immobilized on 15 µL Dynabeads M280 streptavidin resin (Life Technologies), equilibrated in buffer Q. 10 µg TFIIE or TFIIE mutant (tenfold molar excess over Pol II) was incubated with the immobilized CC or control beads for 1 h at 4 °C. The beads were washed four times and bound proteins were analysed by SDS-PAGE (Extended Data Fig. 3h).

Promoter-dependent *in vitro* transcription was used to determine the activity of TFIIE and the TFIIE mutants. The yeast strain used for *TFA1* genetic assays¹⁶ was transformed with a plasmid containing 3×Flag-tagged *TFA1* (pSH810, ARS CEN LEU2 3×Flag). Transformants were streaked once onto –Ura –Leu plates, once onto –Leu plates, twice onto 5-FOA plates, and subsequently onto YEPD plates. Nuclear extract from 3 l yeast culture was prepared as described¹⁹. The nuclear extract was immunodepleted of 3×Flag-Tfa1 as described¹⁹ with the following modifications. Before nuclear extract was incubated with anti-Flag M2 agarose beads (Sigma), beads were incubated with 1 mg ml⁻¹ bovine serum albumin (BSA) protein (Sigma) for 1 h at 4 °C on a turning wheel followed by three wash steps. Immunodepleted nuclear extract was separated from beads by Micro Bio-Spin chromatography columns (Biorad). Specificity of the depletion was confirmed by western blot carried out as described previously¹⁹ for 3×Flag-tagged Tfa1 (Macs Miltenyi Biotec, 130-101-572), Rpb3 (Neoclone, WP012), TFIIB (Abcam, sc-274) and Histone H3 (Abcam, ab21054) (Extended Data Fig. 3i). The secondary antibodies anti-rabbit IgG horse-radish peroxidase (HRP; GE Healthcare, NA934) and anti-mouse IgG HRP (Abcam, ab5870) were used (Extended Data Fig. 3i). Activator- and promoter-dependent *in vitro* transcription

and primer extension were carried out as described previously¹⁹. Recombinant TFIIE (5 pmol) and TFIIE mutants (5 pmol) were added to the depleted nuclear extract as indicated in Fig. 3c. Transcripts were visualized on a denaturing 8% polyacrylamide TBE gel with a Typhoon 9500 scanner (GE Healthcare) and quantified with ImageQuant (GE Healthcare). For quantification, the relative activity of each variant compared to TFIIE was calculated for each replicate. The mean intensity and standard deviation of three replicates was calculated from their relative activities. A second RNA product from the HIS4 promoter was observed, apparently resulting from an alternative upstream transcription start site. Although some differences in the relative use of the two TSSs are observed in the assay, we refrained from interpreting these.

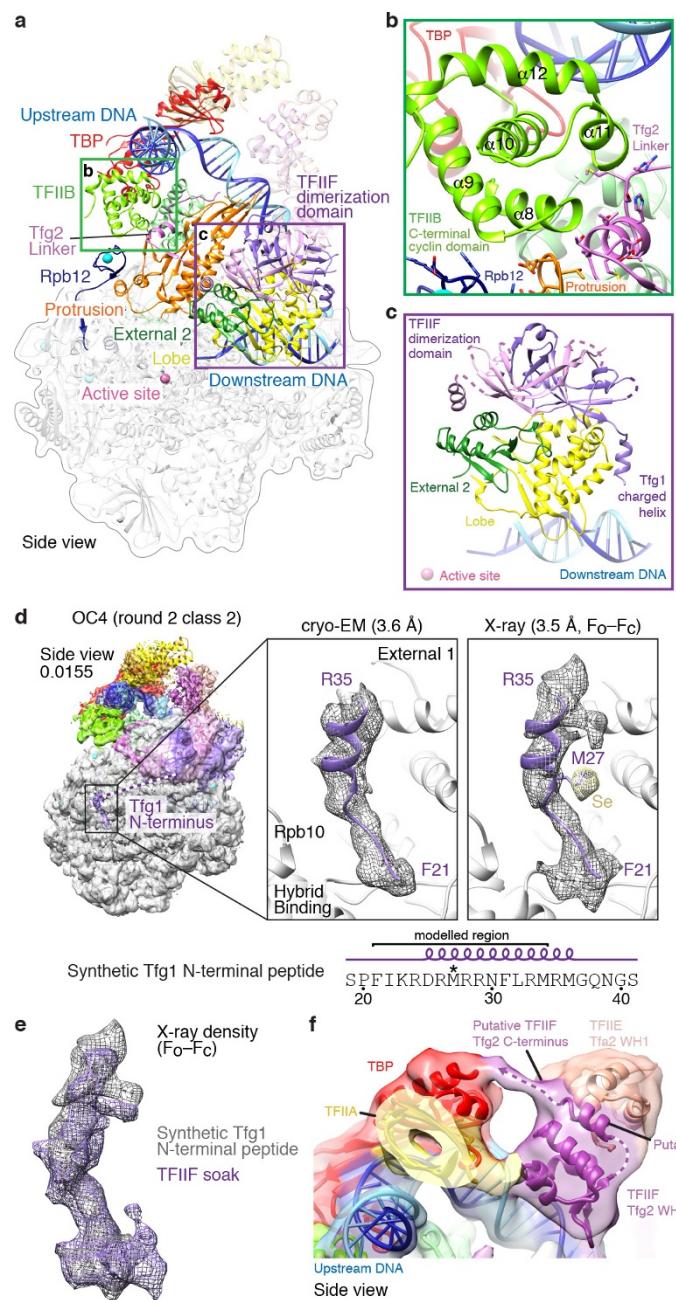
51. Baumli, S., Hoeppner, S. & Cramer, P. A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J. Biol. Chem.* **280**, 18171–18178 (2005).
52. Seizl, M. et al. A conserved GA element in TATA-less RNA polymerase II promoters. *PLoS One* **6**, e27595 (2011).
53. Sydow, J. F. et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34**, 710–721 (2009).
54. Korinek, A., Beck, F., Baumeister, W., Nickell, S. & Plitzko, J. M. Computer controlled cryo-electron microscopy – TOM² a software package for high-throughput applications. *J. Struct. Biol.* **175**, 394–405 (2011).
55. Li, X., Zheng, S. Q., Egami, K., Agard, D. A. & Cheng, Y. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J. Struct. Biol.* **184**, 251–260 (2013).
56. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
57. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
58. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
59. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
60. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
61. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
62. Nguyen, T. H. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
64. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
65. Eswar, N. et al. Comparative protein structure modeling using Modeller. *Current Protoc. Bioinformatics* <http://dx.doi.org/10.1002/0471250953.bi0506s15> (2006).
66. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
67. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
68. Kilpatrick, A. M., Koharudin, L. M., Calero, G. A. & Gronenborn, A. M. Structural and binding studies of the C-terminal domains of yeast TFIIF subunits Tfg1 and Tfg2. *Proteins* **80**, 519–529 (2012).
69. Shen, Y. et al. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* **105**, 4685–4690 (2008).
70. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
71. Wriggers, W. Using Situs for the integration of multi-resolution structures. *Bioophys. Rev.* **2**, 21–27 (2010).
72. Kinkelin, K. et al. Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proc. Natl. Acad. Sci. USA* **110**, 15277–15282 (2013).
73. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
74. Blanc, E. et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
75. Gaiser, F., Tan, S. & Richmond, T. J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J. Mol. Biol.* **302**, 1119–1127 (2000).
76. Ha, I. et al. Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* **7**, 1021–1032 (1993).
77. Hararni, G. M., Gyimesi, M. & Kovács, M. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem. Sci.* **38**, 364–371 (2013).
78. Esnault, C. et al. Mediator-dependent recruitment of TFIIF modules in preinitiation complex. *Mol. Cell* **31**, 337–346 (2008).
79. Cabart, P. & Luse, D. S. Inactivated RNA polymerase II open complexes can be reactivated with TFIIE. *J. Biol. Chem.* **287**, 961–967 (2012).



Extended Data Figure 1 | See next page for caption.

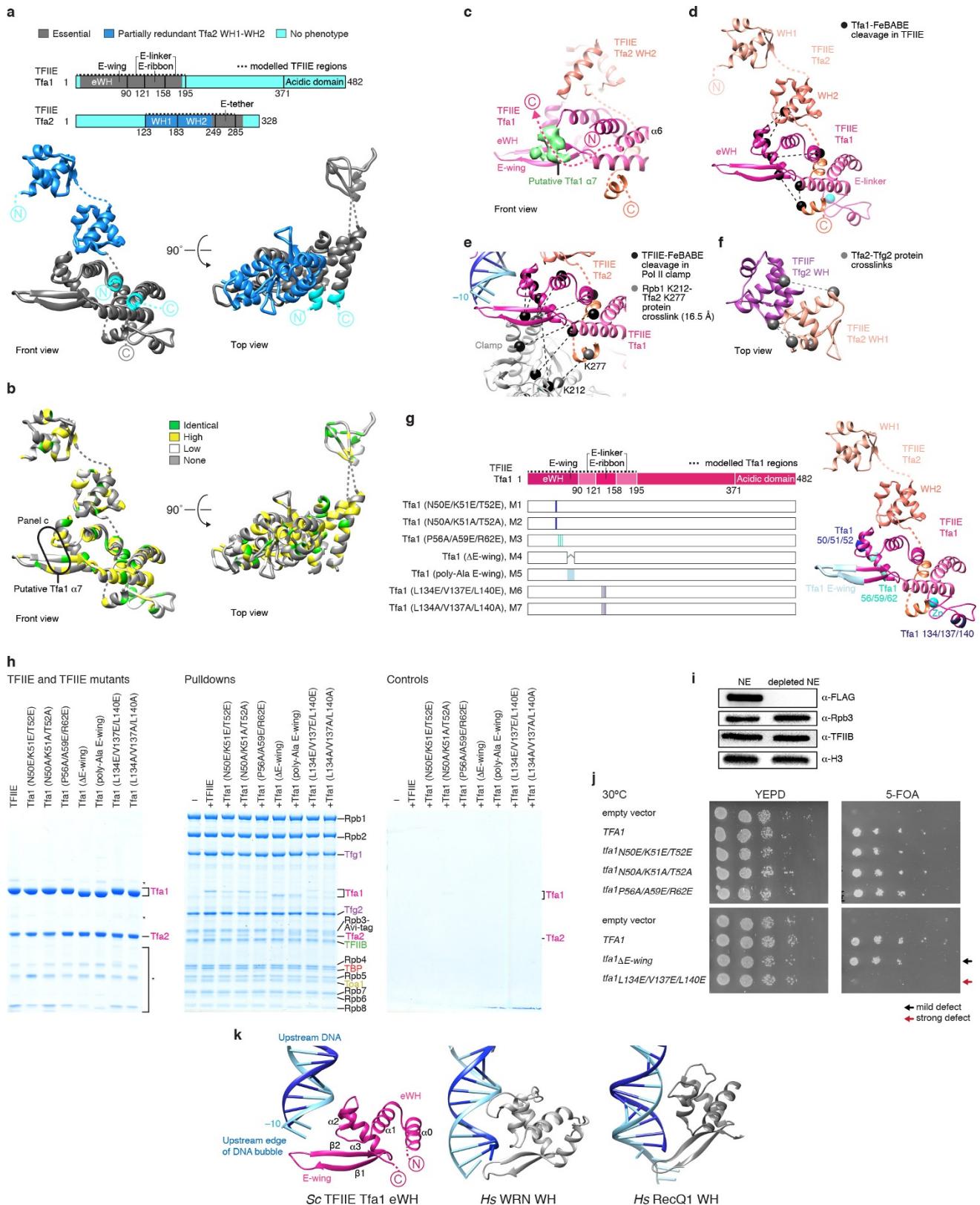
Extended Data Figure 1 | Modelling of open complex cryo-EM densities. **a**, SDS-PAGE analysis of OC-cMed-Med1 complex after size-exclusion chromatography. Protein colours as in Fig. 1. Although core Mediator was required for stable association of TFIIE, it largely dissociated under cryo-EM conditions as observed previously¹⁹. Some remaining core Mediator was flexible and located as described previously¹⁹, but could not be included in further high-resolution analysis. **b**, Cryo-EM micrograph of the OC-cMed-Med1 complex. Scale bar, 50 nm. **c**, Ten representative reference-free 2D class averages of OC-cMed-Med1 reveal flexibility of the upstream DNA assembly including TFIIE (green arrow) and very weak density for core Mediator (orange arrow). Compare Extended Data Fig. 7a, c. **d**, Composite cryo-EM density of the OC shown in front and top views⁵⁰. Colours indicate the cryo-EM densities used for modelling of the open complex (OC1, grey; OC2, green; OC2-focused, yellow; OC3, salmon; OC3-focused, blue; OC4 purple; OC4 round 2 class 2, light blue). Shown are the unsharpened cryo-EM densities. The percentage of particles from the full set of 257,259 that was used for the respective reconstruction is indicated. **e**, Composite cryo-EM density of the OC superimposed on a ribbon model of the OC, coloured as in Fig. 1. The composite cryo-EM density enabled modelling of the initiation factors and DNA. Our structure also enabled correction of the revised yeast initiation complex model obtained by Murakami *et al.* from cryo-EM at 6 Å resolution²¹, and we note the following differences between the structures, superimposed on Rpb1: (1) The TFIIF Tfg2 WH domain is rotated by ~180°, which is further inconsistent with nuclear magnetic resonance (NMR) data on the Tfg2 WH-DNA interface⁶⁸ and fits comparatively worse to protein-protein crosslinking data between the Tfg2 WH and Tfa2 WH1 (ref. 17). (2) Domains of TFIIE, except Tfa2 WH1, were placed incorrectly: Tfa1 eWH (rotation and translation into the E-linker density; 17 Å distance for helix α3 in our CC), Tfa1 E-ribbon (rotation and translation into E-linker density; 35 Å distance between the Zn atoms), and Tfa2 WH2 (~180° rotation). Further, the Tfa2 E-tether region was incorrectly assigned to density belonging to the Tfa1 eWH. The Tfa1 E-linker was not modelled. (3) The TFIIF Tfg1 arm was modelled into an empty space lacking density, and the Tfg1 helix α0 was absent. Our models of the TFIIF dimerization domain, Tfg2 linker, Tfg1 N terminus, and Tfg1 arm fit into densities from a recent study²¹, indicating the electron microscopic reconstruction is correct, but that the modelling was premature at the available resolution. **f**, Ribbon model of the OC coloured according to how different parts of the

OC were modelled into the OC cryo-EM densities (see **d**). Regions with atomic (light blue) and backbone models (orange), and DNA (dark blue) are indicated. Views as in **d**, **g**, Representative regions of the sharpened cryo-EM densities OC1 (3.6 Å), OC2 (4.0 Å), and OC4 (3.9 Å) are shown with the underlying refined coordinate model. The OC1 density shows clear side-chain features for Rpb1 clamp helices α8 and α9 and Rpb2 β33, the OC4 density for Tfg1 β2 that is part of the dimerization domain, and the OC2 density for part of the Tfg2 linker. For OC nomenclature, see Extended Data Fig. 7. **h**, Fit of the TBP crystal structure (PDB: 1YTB)⁶⁶ to the OC2 cryo-EM density, shown in a Pol II side view⁵⁰. **i**, Fit of TFIIB N- and C-terminal cyclin domains, B-linker and B-reader, and B-ribbon elements to OC1 and OC2 cryo-EM densities. The B-linker element displays weak density, and the B-reader is not observed. **j**, Fit of the TFIIA crystal structure (PDB: 1YTF)²³ to OC2-focused cryo-EM density in a Pol II top view⁵⁰ (left). The four-helix bundle undergoes a minor rotation towards the β-barrel, and is apparently flexible (compare Extended Data Fig. 5e). Toa1 (middle) and Toa2 (right) subunit structures are shown. A large non-conserved insertion in Toa1 (Δ95–209), lacking in recombinant TFIIA (Methods), may affect the relative positioning of the four-helix bundle to the β-barrel. **k**, Fit of the TFIIF model to OC cryo-EM densities viewed from the top⁵⁰. TFIIF dimerization domain and Tfg1 N-terminal region, arm, and charged helix elements are superimposed on the OC4 cryo-EM density. Tfg2 linker and WH domains are superimposed on OC2 and OC3-focused cryo-EM densities, respectively. Subunit architectures for Tfg1 (middle) and Tfg2 (right) subunits are shown, indicating disordered regions. Secondary structure elements were labelled according to the crystallographic model of the human RAP30–RAP74 heterodimer⁷⁵. **l**, Fit of the TFIIE model to OC cryo-EM densities shown from the front⁵⁰ (left). Models for Tfa1 eWH, E-linker and E-ribbon are superimposed onto OC1 and OC3 densities. Models for Tfa2 WH1 domain, Tfa2 WH2 and E-tether were fitted into OC3-focused and OC3 densities. Tfa1 (middle) and Tfa2 (right) subunits are shown, indicating disordered regions. The connectivity of the E-tether helices remains uncertain. **m**, Fit of promoter DNA to OC cryo-EM densities is shown in a side view⁵⁰. A weak density for single-stranded template DNA contacts the Pol II fork loop 1, and is indicated by a blue arrow. Upstream and downstream DNA models are superimposed with OC3 and OC1 densities, respectively. The location of the Pol II active site magnesium ion is indicated.



Extended Data Figure 2 | Details of TFIIF and the upstream DNA assembly. **a**, View of the open complex from the side⁵⁰. Pol II elements external 2 (dark green), lobe (yellow), protrusion (orange), Pol II subunit Rpb12 (dark blue) and basal factors TBP, TFIIB, and TFIIF are coloured as in Fig. 1. The remainder of the open complex is transparent. Green and purple boxes indicate the locations of TFIIB C-terminal cyclin and TFIIF dimerization domains, respectively. **b**, Interactions of TFIIB C-terminal cyclin domain with Pol II protrusion, Rpb12, Tfg2 linker and DNA. Colours as in **a**. **c**, Details of TFIIF dimerization domain interactions with Pol II external 2 and lobe⁵⁰. **d**, Crystallographic analysis of the yeast-specific Tfg1 N-terminal region. Weak density for the Tfg1 N-terminal region was observed by cryo-EM (OC4 round 2 class 2) at low contour level (0.0155) close to Pol II elements external 1 and the hybrid binding region⁵⁰ (left). X-ray analysis (right) of the corresponding peptide (Tfg1 F21–R35) enabled modelling and assignment of residue M27 (indicated

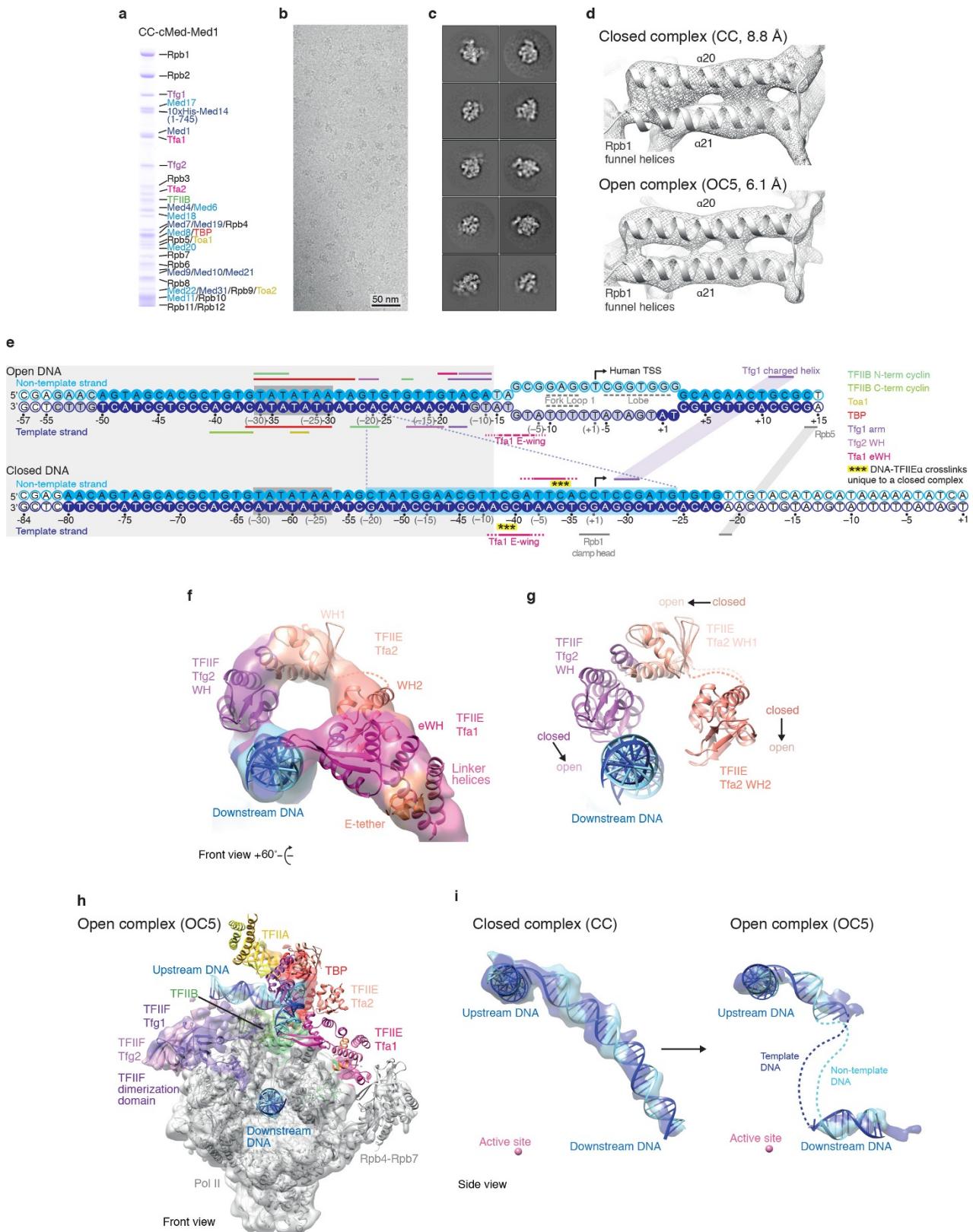
with asterisk) owing to the anomalous signal. The $F_o - F_c$ electron density map (grey, contour level 2.5σ), seleno-methionine anomalous difference Fourier (yellow, contour level 5σ), and final model in ribbon presentation (purple) are shown. The sequence of the synthetic peptide used for soaking into Pol II crystals is shown below. The modified methionine residue and predicted secondary structure are indicated. **e**, The $F_o - F_c$ electron density maps obtained from soaking Pol II crystals with TFIIF (purple) and seleno-methionine labelled peptide (grey), respectively, show similar density in the same location on Pol II. **f**, The putative Tfg2 C terminus contacts TBP. Viewed from the side⁵⁰. A tubular cryo-EM density from the OC3 map, low-pass filtered to 8 Å, emanates from the TFIIF Tfg2 WH–TFIIE Tfa2 WH1 density, and was tentatively assigned to the Tfg2 C-terminal region. The putative Tfg2 density reaches the TBP subunit, consistent with their suggested interaction^{29,76}.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Structure–function analysis of TFIIE and its interactions in the open complex. **a**, The architectural model of TFIIE contains all regions required for viability in yeast¹⁶. A domain schematic (top) indicates the good overlap between modelled (dashed line) and functionally essential regions. Essential (grey), partially redundant Tfa2 WH1 and WH2 domains (blue), and non-essential elements (cyan) are indicated on the TFIIE model, shown in previously defined front and top views⁵⁰ of Pol II. **b**, TFIIE sequence conservation. The sequence conservation among *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Gallus gallus*, and *Homo sapiens* was mapped onto a ribbon representation of the TFIIE model. Highly, strongly, weakly and non-conserved residues are coloured in green, yellow, white, and grey, respectively. The location of a non-modelled helical density in the OC3 cryo-EM map, which may correspond to Tfa1 helix α 7, is indicated. Views as in **a**. **c**, An additional density (green) in the OC3 cryo-EM map on top of the Tfa1 E-wing was tentatively assigned to Tfa1 helix α 7 and this may stabilize the long β -hairpin. A front view is shown⁵⁰. **d**, Tfa1–FeBABE cleavage sites in TFIIE¹⁶ are consistent with the TFIIE architecture. **e**, Tfa1– and Tfa2–FeBABE cleavage sites in the Pol II clamp¹⁶ and a protein–protein crosslink between Rpb1 K212 (Pol II clamp)–Tfa2 K277 (TFIIE E-tether)¹⁷ are consistent with the location of eWH and E-tether. **f**, Tfa2–Tfg2 protein–protein crosslinks¹⁷ are consistent with the Tfg2 WH–Tfa2 WH1 architecture. **g**, The TFIIE mutations used for functional characterization were mapped onto a domain schematic and the model of TFIIE, shown in a front view⁵⁰. **h**, Pulldown assays with recombinant TFIIE variants carrying mutations at the TFIIE–CC interface revealed that the E-ribbon is essential for TFIIE recruitment. For details of the

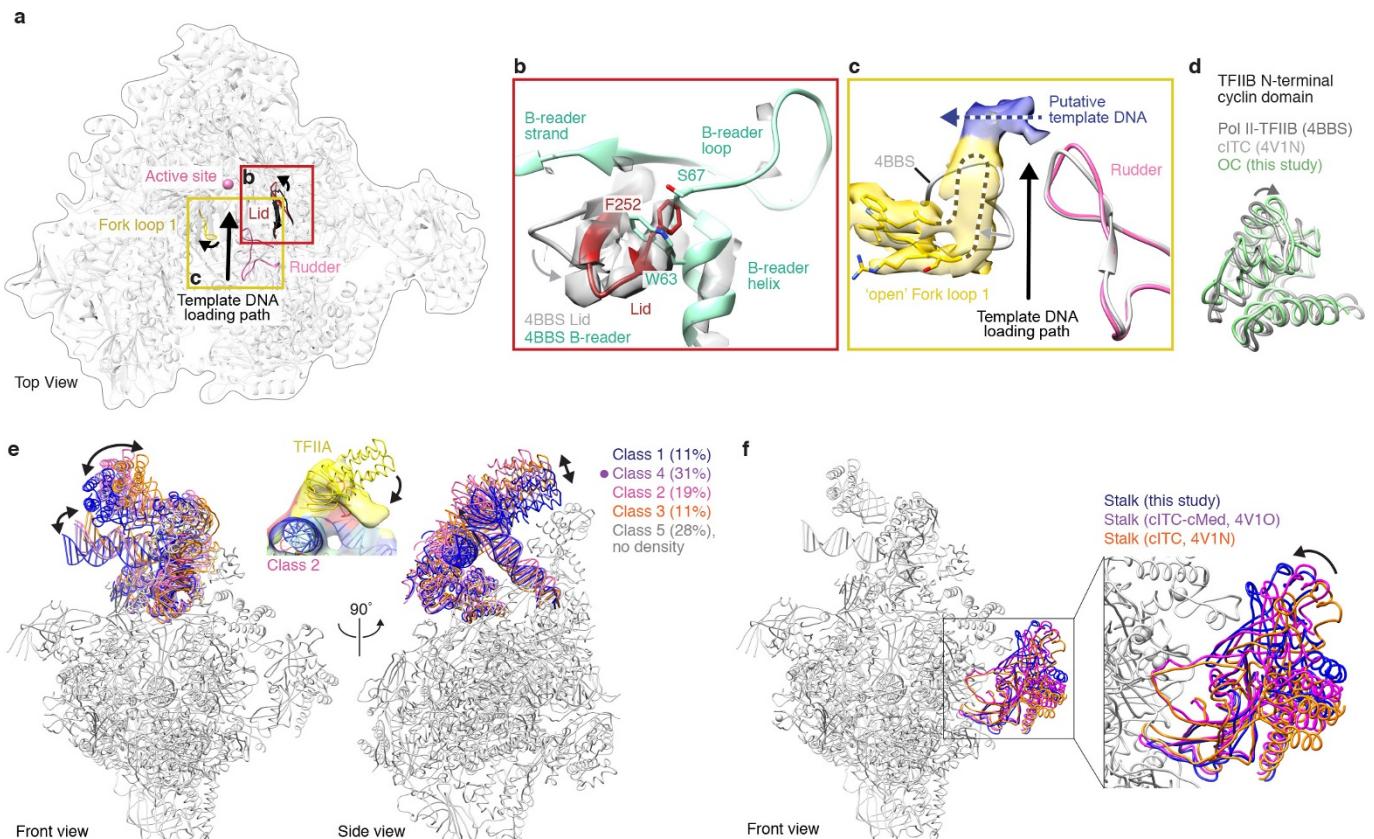
TFIIE mutants, see **g**. Pulldowns were analysed by SDS–PAGE (Coomassie staining). To confirm the integrity of the purified TFIIE variants, 2 μ g were analysed (left). Some minor contaminant and degradation bands of TFIIE are indicated by an asterisk. The bead elution from the pulldown assay is shown (middle), providing negative (no TFIIE) and positive (TFIIE) controls in the two leftmost lanes. The binding of all TFIIE variants to the CC was impaired compared to the wild-type protein, with the exception of the Tfa1(Δ E-wing) mutant, suggesting that all other interfaces contribute to TFIIE binding affinity. The most severe binding defect was observed upon mutation of three residues in the E-ribbon (Tfa1(L134/V137/L140)) to glutamate or alanine. This suggests that the E-ribbon is largely responsible for recruitment of TFIIE to the CC. The bead-only control (right) indicated that TFIIE and TFIIE variants did not show unspecific binding to the beads. **i**, Western blot analysis of the 3 \times Flag-tagged Tfa1 confirms specific immune-depletion of Tfa1 in the nuclear extract (NE), whereas levels of Pol II (Rpb3), TFIIB, and Histone H3 were unaffected. **j**, Yeast complementation assays were performed in triplicate experiments with wild-type *TFA1*, an empty vector, and *TFA1* variants with mutations in the TFIIE eWH domain (N50E/K51E/T52E, N50A/K51A/T52A, and P56A/A59E/R62E in eWH helix α 3, and Δ E-wing), or the E-ribbon(L134E/V137E/L140E) (see Methods). **k**, The long E-wing in the TFIIE subunit Tfa1 eWH is characteristic of WH domains involved in DNA strand separation⁷⁷. The upstream edge of the transcription bubble and eWH domain are shown in a front view⁵⁰ rotated by \sim 20° in the horizontal axis. Corresponding regions of human (*Hs*) Werner syndrome ATP-dependent helicase (WRN) WH (PDB: 2WWY) and RecQ1 WH (PDB: 3AAF) domains are shown.



Extended Data Figure 4 | See next page for caption.

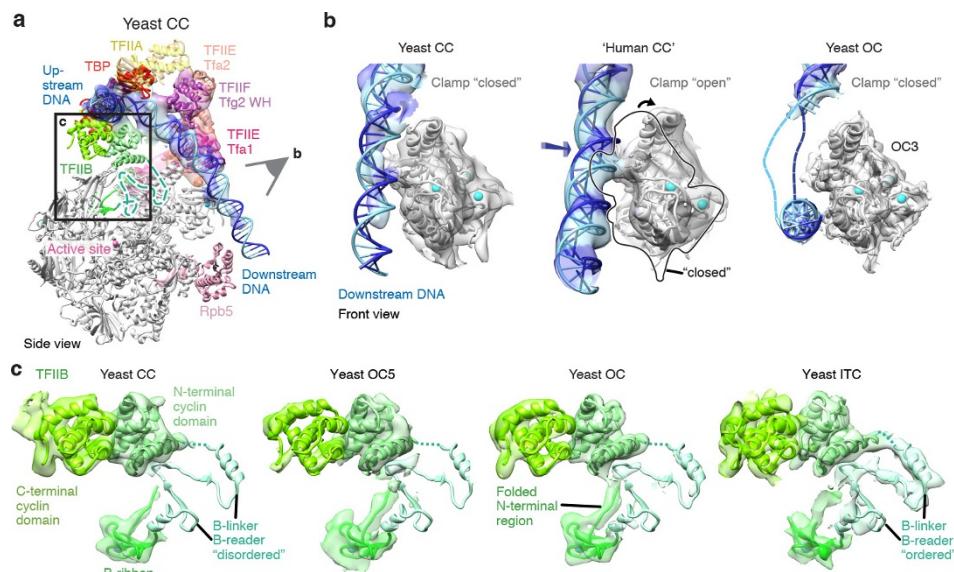
Extended Data Figure 4 | Closed complex and spontaneously formed open complex. **a**, SDS-PAGE analysis of CC–cMed–Med1 complex after size-exclusion chromatography. Protein colours as in Fig. 1. **b**, Cryo-EM micrograph of the CC–cMed–Med1 complex. Scale bar, 50 nm. **c**, Ten representative reference-free 2D class averages of CC–cMed–Med1 reveal flexibility for the upstream complex. Core Mediator was not retained during cryo-EM analysis. **d**, Detailed view of the Pol II funnel helices in the CC (top) and OC5 (bottom) densities. **e**, Promoter sequences and differences in protein–DNA interactions are shown for the two distinct nucleic acid scaffolds used for preparation of closed and open complexes (compare Fig. 1d). Coloured bars indicate DNA–protein interaction. Solid, shaded, and empty circles respectively represent nucleotides included in the structure, excluded owing to weak cryo-EM density, or excluded owing to absence of cryo-EM density. Analogous yeast (black) and human (grey) numbering of promoter DNA is shown. The TATA-box sequence (red box) and *HIS4*-promoter sequence absent in the modified OC nucleic acid scaffold¹⁹ (grey box) are indicated. Protein–DNA interactions in the region covered by the light grey box are unchanged between CC and OC,

and shown only for the OC for clarity. Unique and altered interactions are shown for each complex. DNA–TFIIE α photo-crosslinks, indicated by black asterisks, were observed in a closed but not open promoter state⁴⁰ and are consistent with the CC model. **f**, Fit of TFIIE, Tfg2 WH and downstream DNA into CC density. Two rigid bodies were used for fitting: (i) Tfg2 WH and Tfa2 WH1 and (ii) Tfa2 WH2, eWH, E-linker and E-tether helices. Although the overall fit reflects density well, the eWH domain and its E-wing may be rotated further away from promoter DNA. **g**, Details on the location of downstream DNA (template, blue; non-template, light blue), Tfg2 WH, and Tfa2 WH1 and WH2 in the closed (dark colours) and open (light colours) complexes in the same view as in **f**. **h**, Cryo-EM density of OC5 and the OC ribbon model are shown in a front view⁵⁰. The OC5 map shows weak density in regions of upstream assembly, TFIIE, and DNA that may be caused by increased flexibility owing to the heterogeneous population of spontaneously opened DNA. Colours as in Fig. 1. **i**, Fit of promoter DNA to cryo-EM densities of CC and OC5, shown in a side view⁵⁰.



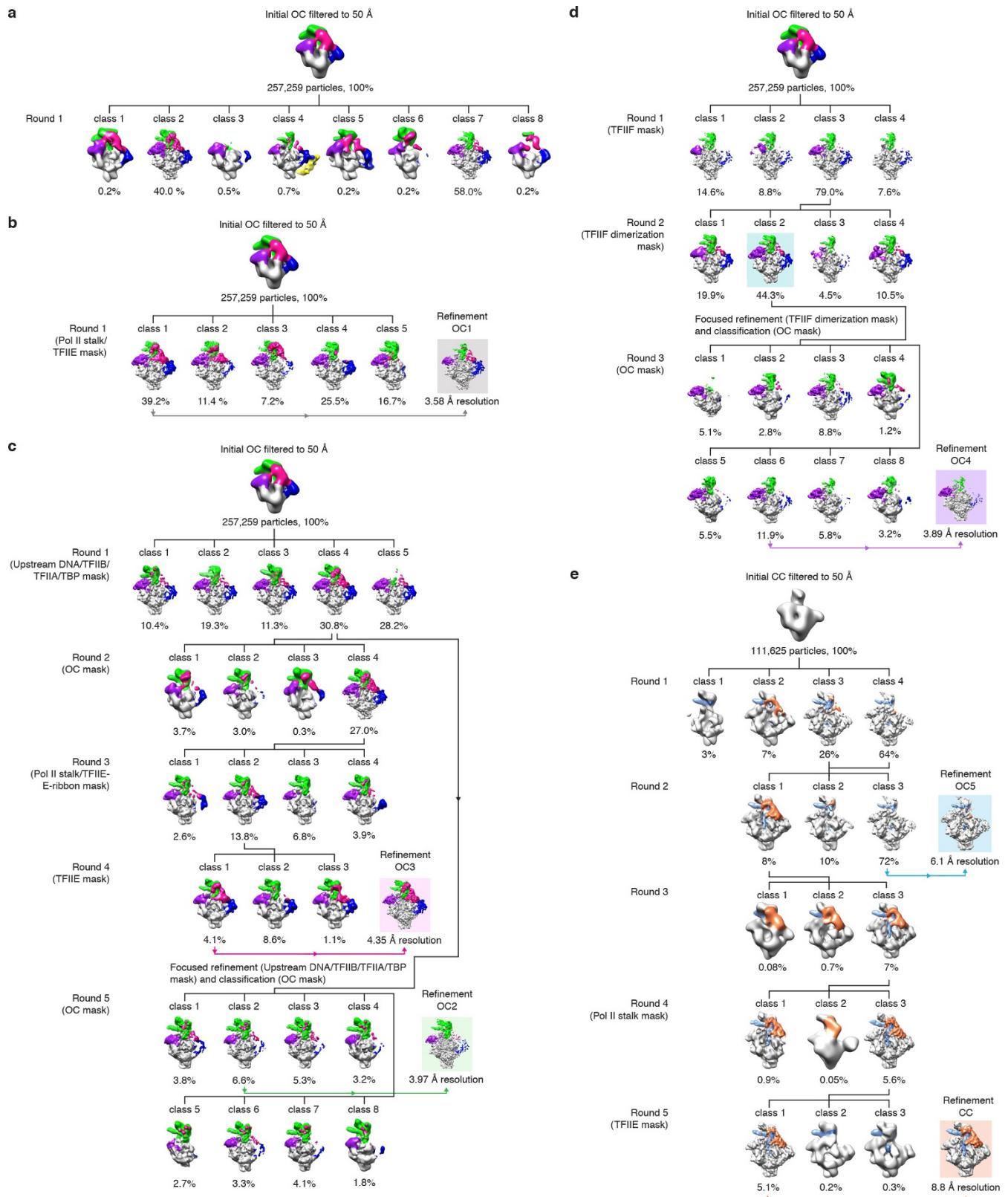
Extended Data Figure 5 | Pol II cleft clearance, structural flexibility and rearrangements in the OC. **a**, Pol II lid and fork loop 1 assume new conformations in the OC, clearing the Pol II cleft for loading of single-stranded template DNA. Arrows indicate the direction of movement of the two Pol II elements, and the template DNA loading path. The lid (dark red) in the open complex is moved in comparison to the lid of a Pol II-TFIIB ITC crystallographic study (PDB: 4BBS). Yellow and red boxes indicate zoomed-in regions of **b** and **c**, respectively. **b**, The movement in the Pol II lid leads to a steric clash with the TFIIB B-reader, observed in a Pol II-TFIIB ITC crystal (PDB: 4BBS), and facilitates its withdrawal in the open complex. In particular the lid residue F252 clashes with W63 and S67 of the B-reader. The OC1 cryo-EM density is shown for both lid and B-reader elements. **c**, The cryo-EM density of the OC1 reveals an 'open' Pol II fork loop 1 and a stably associated fragment of putative template DNA. The 'open' state of fork loop 1 provides additional space for loading of single-stranded template DNA past the Pol II rudder, towards the active site cleft. **d**, The position of the TFIIB N-terminal cyclin domain (light green) is altered in comparison to a Pol II-TFIIB ITC crystal structure¹⁵ (dark grey), but similar to its location in a cITC¹⁹ (light grey), probably owing to the presence of DNA. **e**, Flexibility of the upstream DNA assembly. The cryo-EM data of the OC was sorted on

the basis of structural differences using an upstream assembly mask that included upstream DNA, TFIIA, TBP, and TFIIB cyclin domains (OC2 round 1, compare Extended Data Fig. 7c). Four of five resultant classes revealed different positions of the upstream complex, indicated here by fitted ribbon models of the OC. Previously defined front and side views⁵⁰ are shown. Class 2 (middle) revealed the TFIIA four-helix bundle rotated by 85°, consistent with a high degree of flexibility. Class 4 represents the largest fraction of the data (31%), and gave a more defined density for the upstream complex, which was improved by further classification (Extended Data Fig. 7c). Class 5 presented with no density for the upstream complex or the Tfg2 linker, but did show density for the TFIIB B-ribbon and the TFIIF dimerization domain, suggesting that TFIIB and TFIIF remained bound to the complex. This is consistent with TFIIF-dependent association of the TFIIB-core domain with the Pol II wall²⁷, and this apparently requires an ordered Tfg2 linker. **f**, The Rpb4-Rpb7 stalk adopts different positions in cITC, cITC-cMed, and OC. This suggests that Mediator and TFIIE may bind co-operatively. This is consistent with previous findings⁷⁸ and with pulldowns (Extended Data Fig. 3h), which suggest that the TFIIE E-ribbon–stalk interface, which is important for TFIIE recruitment, is stabilized in the presence of Mediator.



Extended Data Figure 6 | Pol II clamp positions and TFIIB B-reader mobility during DNA opening. **a**, The yeast CC is shown from a side view⁵⁰, indicating the path of DNA and location of TFIIE. The eye symbol (grey) indicates the point of view in **b**. **b**, The Pol II clamp may undergo transitions during DNA opening as indicated. The OC model of the Pol II clamp is shown superimposed on yeast CC (this study), and yeast OC (this study). The OC model Pol II clamp was rigid-body fitted to the human CC cryo-EM density²⁰ (EMD-2306) and is superimposed. The view is from the front⁵⁰. **c**, The TFIIB B-reader element shows strong

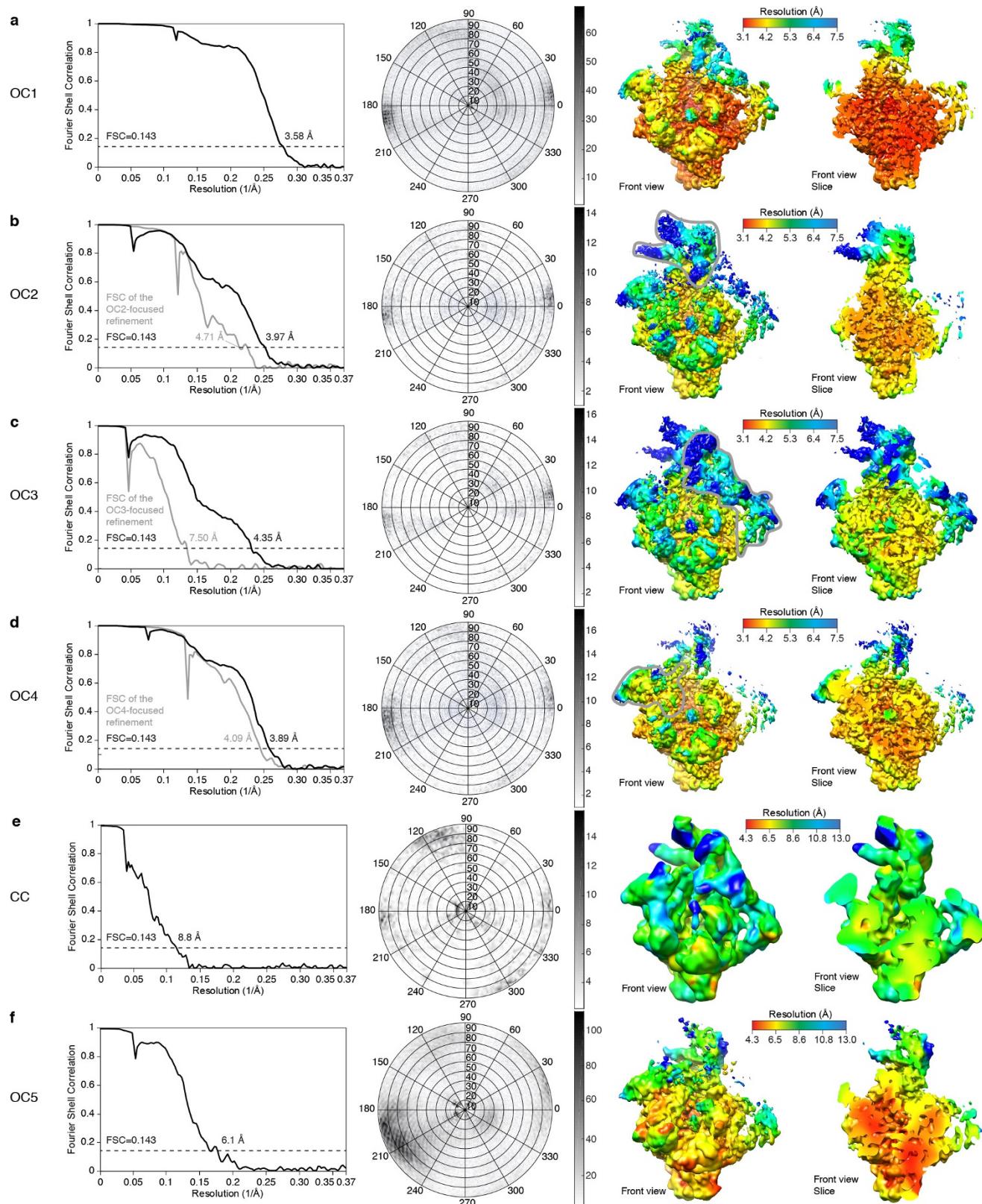
density only in the ITC state, suggesting that its mobility in earlier states may be important for maintaining a cleared path for template DNA loading into the Pol II cleft. Ordering of the B-reader may further lead to stabilization of the upstream promoter assembly that is flexible in the OC (Extended Data Figs 5e, 7c). Cryo-EM densities for yeast CC (this work), OC5 (this work), OC (this work), and ITC (EMD-2785) complexes are superimposed on the TFIIB model (PDB: 4BBS for the B-linker and B-reader). As secondary structure elements could not be resolved in the human CC²⁰, we excluded this cryo-EM density from comparison.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Three-dimensional classification of cryo-EM data. **a**, Three-dimensional image classification of the cryo-EM data set into eight classes using an initial OC reconstruction as the reference model, revealed heterogeneity. The percentage of single particles contributing to each class is provided. To help visualize structural differences, 3D reconstructions of the OC are coloured according to mobile regions: Pol II core, TFIIB B-ribbon (grey); upstream DNA, TFIIA, TBP, TFIIB cyclin domains, Tfg2 linker (green); TFIIF dimerization domain (purple); TFIIE except E-ribbon, Tfg2 WH (magenta); Pol II Rpb4–Rpb7 stalk and E-ribbon (blue); cMed–Med1 (yellow). **b**, Focused classification into five classes using a mask covering the Pol II stalk and E-ribbon. The resultant class 1 (OC1) was subsequently refined to 3.58 Å resolution (grey box) and revealed the location of the TFIIE E-ribbon. Colours as in **a**. **c**, Improvement of densities for Tfg2 linker, TFIIB, and TFIIE, through rounds of focused 3D classification using various masks. First, heterogeneity due to flexibility of upstream DNA and associated factors was overcome by applying a mask around this region (round 1). Focused refinement of the upstream DNA assembly of the resultant class 4 of round 1 (OC2-focused), improved the density quality for TFIIA (Extended Data Fig. 1j). Classification of the OC2-focused density revealed the upstream DNA complex (OC2) at 3.97 Å resolution (green box). Separate classification of class 4 of round 1 using OC, Pol II stalk and TFIIE E-ribbon, and TFIIE masks yielded class 1 of round 4 (OC3, magenta box) that contained a complete TFIIE density at a nominal resolution of 4.35 Å after 3D refinement (see Extended Data Fig. 8c). The small fraction of stably bound TFIIE is consistent with its reduced affinity

to the pre-initiation complex⁷⁹. Focused refinement of OC3 with a TFIIE-stalk mask (OC3-focused) improved density for Tfg2 WH and Tfa2 WH1 domains. Colours as in **a**. **d**, To improve the density of TFIIF dimerization domain and the Tfg1 arm, three rounds of classification using a TFIIF, TFIIF dimerization domain, and OC mask were employed. Class 2 of round 2 (cyan box) enabled fitting of the Tfg1 N-terminal peptide, which was resolved by X-ray analysis (Extended Data Fig. 2d, e). This class was further refined locally using a mask covering the TFIIF dimerization domain, and then classified with an OC mask, revealing class 6 of round 3 at 3.89 Å resolution after 3D refinement (purple box). Colours as in **a**. **e**, 3D classification of the CC cryo-EM data set into four classes, using an initial CC reconstruction as the reference model, revealed heterogeneity. Mobile regions in the reconstructions are highlighted: promoter DNA (blue), TFIIE (except E-ribbon), and Tfg2 WH (orange). Classifying the most populated classes from round 1 into three classes unexpectedly revealed open and closed promoter DNA states in the data set: CC (round 2, class 1) and OC5 (round 2, class 3). Class 3 of round 2 (OC5) was refined to 6.1 Å resolution (blue box). Class 1 from round 2 was further classified into three classes. The resultant class 3 of round 3 revealed density for closed downstream promoter DNA above the Pol II cleft, and TFIIE. The cryo-EM density for downstream DNA and TFIIE was improved by focused classification using two soft-edged masks. A mask covering the Pol II Rpb4–Rpb7 stalk yielded a class with better occupancy for the stalk (round 4, class 3), which was further sorted using a mask covering TFIIE and Tfg2 WH to improve their densities. Class 1 of round 5 was refined to 8.8 Å resolution (CC, orange box).


Extended Data Figure 8 | Resolution of cryo-EM reconstructions.

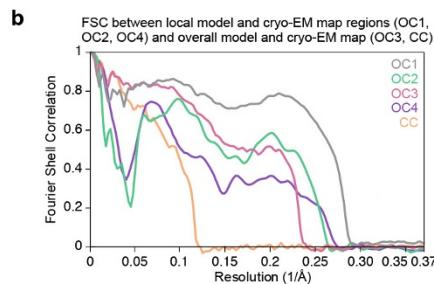
a, Gold-standard FSC (left) of the OC1 cryo-EM single particle reconstruction ($\text{FSC}=0.143$). Orientation distribution plot of all particles that contribute to the OC1 reconstruction (middle). The OC1 cryo-EM map is shown (right) from a front view⁵⁰ and a central slice through the reconstruction, which are coloured by local resolution as described¹⁹. **b**, As in **a** but for the OC2 reconstruction. The gold-standard FSC for the

density obtained from focused refinement (OC2-focused) with a soft mask around the upstream DNA assembly is indicated in grey (see Methods). The region masked for focused refinement is indicated with a grey outline on the cryo-EM map coloured by local resolution (right). **c**, As in **b**, but for the OC3 and OC3 focus-refined reconstructions. **d**, As in **a**, but for the OC4 and OC4 focus-refined reconstructions. **e**, As in **a**, but for the CC reconstruction. **f**, As in **a**, but for the OC5 reconstruction.

a Cryo-EM data collection and refinement statistics of the OC structure

	Pol II-TFIIB ribbon (OC1)	Upstream assembly* (OC2)	TFIIF dimerization (OC4)
Data collection			
Particles	102,876	17,282	29,455
Pixel Size (Å)	1.35	1.35	1.35
Defocus range (μm)	-0.7 to -4.2	-0.7 to -4.2	-0.7 to -4.2
Voltage (kV)	300	300	300
Electron dose (e ⁻ Å ⁻²)	33	33	33
Reconstruction (RELION)			
Accuracy of rotations (°)	1.58	1.11	0.96
Accuracy of translations (pixel)	0.71	0.59	0.50
Resolution (Å)	3.58	3.97	3.89
Map sharpening B-factor (Å ²)	111	95	79
Model composition			
Non-hydrogen atoms	27844	3488	1839
Protein residues	3506	444	221
Refinement (PHENIX)			
Map CC (around atoms)	0.768	0.631	0.539
Rms deviations			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	0.94	1.11	1.25
Validation			
Molprobity score	1.99	2.23	2.51
All-atom clashscore	12.43	18.87	31.30
Rotamer outliers	0.3%	0%	0%
C-beta deviations	0	0	0
Ramachandran plot			
Outliers	0.3%	0.5%	1.0%
Allowed	5.4%	6.8%	8.4%
Favoured	94.3%	92.7%	90.6%

*A model of Upstream assembly components was refined into the OC2 cryo-EM density: Pol II protrusion (residues 44-218 and 405-465), Tfg2 Linker (residues 249-280), TFIIB N-and C-terminal cyclin domains (residues 124-326), and the Tfg1 arm (residues 327-335).



Extended Data Figure 9 | Data collection, refinement statistics, and structure validation. **a**, Cryo-EM data collection and refinement statistics of the OC structure. Different regions of the composite OC structure were refined into OC1, OC2, and OC4 maps as described (see Methods) to obtain an atomic model for 90% of the structure. **b**, Gold-standard FSC

c X-ray crystallographic data collection and refinement statistics

	Pol II-Tfg1 SeMet	Pol II-TFIIF
Data collection		
Space group	C222(1)	C222(1)
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	221.2, 392.7, 282.2	222.8, 392.7, 283.4
Resolution (Å)	50-3.5 (3.6-3.5)*	50-3.9 (4.0-3.9)*
<i>R</i> _{sym} (%)	11.3 (105.4)	13.6 (91.3)
<i>I</i> / <i>σI</i>	15.7 (2.0)	12.8 (2.4)
Completeness (%)	100 (100)	99.8 (100)
Redundancy	7.1 (6.9)	7.5 (7.6)
CC _{1/2} (%)	99.8 (78.9)	99.8 (82.6)
Refinement		
Resolution (Å)	49-3.5	48.9-3.9
No. reflections	151,192	112,717
<i>R</i> _{work} / <i>R</i> _{free} (%)	15.4/18.0	16.2/19.4
No. atoms		
Protein	31,330	31,330
Ligand/ion	9	9
B-factors (Å ²)		
Protein	127.9	143.7
Ligand/ion	125.7	124.4
R.m.s deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.30	1.26

*Highest resolution shell is shown in parenthesis.

between the respective coordinate models and local regions of the OC1, OC2, and OC4 cryo-EM maps used for model refinement and between overall OC and CC models compared to OC3 (best TFIIE density) and CC cryo-EM maps. **c**, X-ray crystallographic data collection and refinement statistics.