Published in final edited form as:
Nature. 2017 November 09; 551(7679): 204–209. doi:10.1038/nature24282.

Structures of transcription pre-initiation complex with TFIIH and Mediator

S. Schilbach[1], **M. Hantsche**[1], **D. Tegunov**[1], **C. Dienemann**[1], **C. Wigge**[1], **H. Urlaub**[1], and **P. Cramer**[1]

[1]Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany

Abstract

For transcription initiation, RNA polymerase (Pol) II assembles with general transcription factors on promoter DNA to form the pre-initiation complex (PIC). We report cryo-EM structures of the yeast PIC and PIC-core Mediator (cMed) complex at nominal resolutions of 4.7 Å and 5.8 Å, respectively. The structures reveal TFIIH and suggest how the TFIIH modules 'core' and 'kinase' function in promoter opening and Pol II phosphorylation, respectively. The TFIIH core subunit Ssl2 (human XPB) is positioned on downstream DNA by the 'E-bridge' helix in TFIIE, consistent with TFIIE-stimulated DNA opening. The TFIIH kinase module subunit Tfb3 (human MAT1) anchors the kinase Kin28 (human Cdk7) that is mobile in the PIC but preferentially located between the Mediator hook and shoulder in the PIC-cMed complex. Open spaces between the Mediator head and middle modules may allow access of the kinase to its substrate, the C-terminal domain (CTD) of Pol II.

Transcription of protein-coding genes begins with the formation of a pre-initiation complex (PIC) on promoter DNA1. The PIC consists of RNA polymerase (Pol) II and the transcription factors (TF) IIA, -B, -D (or its subunit TBP), -E, -F, and -H (Extended Data Table 1). The coactivator Mediator stabilizes the PIC2 and is globally required for initiation3–5. Structures of the PIC lacking TFIIH (core PIC, cPIC) have been derived for

Correspondence and request of materials should be addressed to PC (patrick.cramer@mpibpc.mpg.de).

Author contributions SS carried out all experiments and data analysis except for the following. MH carried out EDC crosslinking and Mediator modeling. CD carried out TFIIE modeling and established a protocol for cPIC complex formation. DT wrote and applied the WarpCraft software. CW supervised EM data collection. HU conducted mass spectrometry. PC designed and supervised research. SS and PC prepared the manuscript.

Author information 3D cryo-EM density maps of PIC and PIC-cMed have been deposited in the Electron Microscopy Database under the accession numbers EMD-3846, and EMD-3850 respectively. Coordinate files of PIC and PIC-cMed have been deposited in the Protein Data Bank under accession numbers 5OQJ, and 5OQM, respectively. Reprints and permissions information is available at www.nature.com/reprints.

The author declare that they have no competing financial interest.

Code availability statement
The source code for the WarpCraft software is available as Supplementary Data 1 for download at Nature and via GitHub (https://github.com/cramerlab/warpcraft).

Data availability statement
The electron density reconstructions and final models for the PIC and PIC-cMed complex were deposited with the EM Data Base (accession codes EMD-3846 and EMD-3850, respectively) and with the Protein Data Bank (PDB ID 5OQJ and 5OQM, respectively).

Europe PMC Funders Author Manuscripts

the yeast *S. cerevisiae*6 and human7 by cryo-electron microscopy (cryo-EM) at 3.6 Å and 3.9 Å resolution, respectively. The crystal structure of core Mediator (cMed) was obtained for the fission yeast *S. pombe* at 3.4 Å resolution and contains the essential Mediator subunits5. Detailed structural information is lacking for TFIIH, but TFIIH has been located within the PIC7–10 and its subunit topology7,9,11–13 was revealed.

TFIIH is essential for transcription and DNA repair and consists of a 7-subunit core and a 3-subunit kinase module14. Whereas the core suffices for DNA repair, the kinase module is additionally required for transcription15. The core comprises the ATPases Ssl2/XPB (yeast/human) and Rad3/XPD, and subunits Tfb1/p62, Tfb2/p52, Ssl1/p44, Tfb4/p34 and Tfb5/p8. Ssl2/XPB functions in promoter opening16 and escape17,18, but is not universally required for DNA opening6,19. The TFIIH kinase module contains the kinase Kin28/CDK7, the cyclin Ccl1/CycH, and Tfb3/MAT1. Kin28/Cdk7 phosphorylates the C-terminal domain (CTD) of Pol II20, is stimulated by Mediator21, and facilitates promoter escape22.

Here we extend our previous structural studies of cPIC6 and cMed5 to arrive at the structures of the yeast PIC containing TFIIH and of the PIC-cMed complex. The latter structure has a molecular weight of ~2 MegaDalton, includes 46 polypeptides, and contains all transcription initiation-related proteins that are essential in yeast. The structures reveal TFIIH and its interactions with Pol II, TFIIE, DNA, and Mediator.

Structures of PIC and PIC-cMed complex

Thus far TFIIH was purified in small quantities from natural sources. To overcome this limitation, we prepared both TFIIH modules in recombinant form after co-expressing their subunits (Methods, Extended Data Fig. 1). The two modules contained TFIIH subunits in apparently stoichiometric amounts and could be assembled into the complete 10-subunit TFIIH. Reconstituted TFIIH formed a stable complex with cPIC and cMed. The resulting 46-subunit PIC-cMed complex was subjected to cryo-EM data collection (Methods, Extended Data Fig. 1). Unsupervised particle sorting led to cryo-EM reconstructions of the PIC and PIC-cMed complex at nominal resolutions of 4.7 Å and 5.8 Å, respectively (Extended Data Fig. 2).

Secondary structure was visible in maps obtained with RELION23 only after focused refinement on cPIC, TFIIH, or cMed. To reconstruct continuous cryo-EM maps from particles with such flexible regions we developed a computational tool, ‘WarpCraft’ (Methods, Supplementary Data 1). WarpCraft represents maps as pseudo-atomic models and simulates restrained motions between flexible map regions. This avoids the spatial divergence of separate focused refinements, and can make the construction of composite maps obsolete. Thus we obtained cryo-EM maps that revealed highly defined secondary structure throughout the PIC and PIC-cMed complex.

To solve the PIC structure (Fig. 1, Supplementary Video 1), we first fitted our cPIC structure6 to the density and made adjustments to TFIIB, the TFIIE subunits Tfa1 and Tfa2, and the Pol II clamp. The PIC adopts the open promoter state with unwound DNA in the active center as before6. Structures and models for 22 TFIIH domains were unambiguously

fitted to the remaining density (Supplementary Data Table 1). Eleven connections within and between TFIIH domains were traced and the obtained model was refined by flexible real space fitting (Methods). The TFIIH structure is consistent with 153 known protein-protein crosslinks obtained with bis(sulfosuccinimidyl)suberate (BS3) and 1,1'-(suberoyldioxy)bisazabenzotriazol) (SBAT)10,24,25, and with additional 55 crosslinks obtained with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (Extended Data Fig. 3; Supplementary Data Tables 2, 3).

To solve the PIC-cMed structure (Fig. 2, Supplementary Video 2), we placed the generated PIC model into the PIC-cMed cryo-EM map. We then fitted the remaining density with the *S. cerevisiae* cMed model obtained from the *S. pombe* crystal structure5. We obtained a model for the PIC-cMed complex after flexible real space fitting of seven rigid bodies in cMed and manual adjustments (Methods, Supplementary Data Table 4). The DNA path is virtually identical in both new structures, highly similar in the yeast open cPIC6, and similar in the human open PIC7. The obtained PIC and PIC-cMed structures consist of atomic models where high-resolution structures were available (76% and 73%, respectively), and backbone models for other parts of TFIIH and cMed.

TFIIH structure

The PIC structure reveals that the TFIIH core forms a crescent-shaped complex spanning from Ssl2/XPB to Rad3/XPD (Fig. 3, Extended Data Fig. 4). Ssl2 binds downstream DNA as previously observed7,9,26, consistent with its role in DNA opening16. Rad3 is located ~40 Å away from DNA, in agreement with its ATPase activity being dispensable for transcription27. TFIIH subunits Tfb5, Tfb2, Tfb4, and Ssl1 are arrayed in between the two ATPases. The Tfb1 subunit meanders along Tfb4, Ssl1 and Rad3 and its plekstrin homology domain (PHD) protrudes from the crescent towards the Pol II clamp.

The TFIIH core structure shows that the bilobal Ssl2 ATPase contains a C-terminal extension in lobe 2 that contacts Tfb5 in the Tfb2-Tfb5 dimerization module28. Ssl2 and Tfb2 interact via newly observed 'clutch' domains. Tfb2 further contains a region with three helix-turn-helix subdomains that binds Tfb4, which contains a van Willebrandt (vWA) domain with an insertion and an extended zinc finger (eZnF) domain. Like Tfb4, Ssl1 contains a vWA and eZnF domain, and an additional RING domain. Tfb4 and Ssl1 interact intimately and form the backbone of TFIIH. Ssl1 also binds Rad3, which contains a bilobal ATPase with two insertions in lobe 1, an iron-sulfur (FeS) cluster, and an ARCH domain. Whereas the FeS cluster resembles that in known archaeal structures29–31, the ARCH domain contains an additional helix and two helix extensions. Tfb1 comprises an N-terminal PHD, two BTF2-like, synapse-associated and DOS2-like (BSD) domains32, helical regions that anchor Rad3 and Tfb4 (Rad3 anchor and Tfb4 anchor, respectively), and a C-terminal 3-helix bundle that binds the two eZnF domains.

Our TFIIH structure defines the orientation of eight domains in TFIIH subunits that were inferred by previous studies of the PIC7–9. It also reveals 15 additional domains, numerous connections, and details of domain interactions. Regions in TFIIH subunits that are essential for cell viability in yeast33 tend to be ordered in our structure (Extended Data Fig. 5a). The

Europe PMC Funders Author Manuscripts

TFIIH structure also suggests the effect of mutations in human TFIIH subunits p8, XPB and XPD that are associated with the human diseases Xeroderma pigmentosum, Trichothiodystrophy, and Cockayne syndrome14,34,35. Many of the mutated sites are predicted to destabilize the TFIIH core structure (Extended Data Fig. 5b).

TFIIH interactions with cPIC

The PIC structure reveals four sites of interaction between TFIIH and cPIC (Fig. 4). First, the TFIIH kinase module subunit Tfb3 bridges between the Pol II stalk subcomplex Rpb4-Rpb7, TFIIE, and Rad3 (Extended Data Fig. 6a). In particular, the Tfb3 RING domain binds between the Rpb7 OB domain and the TFIIE E-linker helices, and the Tfb3 'ARCH anchor' contacts the Rad3 ARCH domain. This is consistent with the known interaction between the TFIIH kinase module and Rad327,36 and the initiation function of Rpb4-Rpb737, which also binds TFIIE6 and cMed4. The Tfb3 contact with Pol II further explains why the PIC recruits TFIIH that contains the kinase module, rather than only core TFIIH15. A role for Tfb3 in TFIIH recruitment can also explain why the kinase module is required for transcription initiation in a reconstituted system20 although its kinase activity is not38. The C-terminal part of Tfb3 is disordered and connects to the kinase-cyclin pair, which is also mobile in the PIC structure.

The three additional interactions between TFIIH and cPIC involve the mobile C-terminal region of TFIIE subunit Tfa1 (human TFIIEα). This TFIIE region forms three previously unobserved helices that are flexibly connected and named here E-dock (α7), E-bridge (α8), and E-floater (α9) (Extended Data Fig. 6b). The E-dock apparently enables docking of the Tfb1 PHD to the TFIIE extended winged helix (eWH) domain that is located on the Pol II clamp (Extended Data Fig. 6c). The E-bridge extends from Tfb1 domain BSD2 to Ssl2 lobe 2 (Extended Data Fig. 6d, e). The E-floater binds the BSD1 domain in Tfb1 (Extended Data Fig. 6f, g). Taken together, TFIIE contacts TFIIH at four sites, explaining why TFIIE is required for TFIIH recruitment to the PIC39.

TFIIH and DNA opening

The PIC structure shows that the Ssl2 ATPase engages with promoter DNA ~25-30 base pairs (bp) downstream of the putative transcription start site +1 (Fig. 5, Extended Data Fig. 7). This location is consistent with crosslinking data40 and previous cryo-EM studies7,9, and with the translocase model for ATP-dependent DNA opening26,41. According to this model, Ssl2 uses ATP hydrolysis to translocate on DNA away from Pol II. If the Ssl2 location is fixed, Ssl2 action results in a reeling of DNA into the active center. The PIC structure supports a fixed location of Ssl2 and the proposed directionality of translocation. The two ATPase lobes bind the DNA backbones on both sides of the minor groove, similar to the ATPase in the chromatin remodeler Chd142. Comparisons with Chd1 and with ATPase structures of NS3 and Rad3 (Extended Data Fig. 7d, e) indicate that Ssl2 tracks along the DNA template strand in the 3'-5' direction, consistent with biochemical studies43–45. One study suggested that tracking occurs on the non-template strand in 5'-3' direction26, but this would result in the same overall movement.

The PIC structure also suggests how TFIIE may stimulate the ATPase activity of TFIIH46. According to the current model for ATPase translocation42,47, ATP binding induces a ratcheting movement of lobe 2 with respect to lobe 1, and a DNA translocation by one base pair. In our structure we trapped the pre-translocation state of Ssl2 with an empty ATPase active site (Fig. 5b). The C-terminal end of the TFIIE E-bridge contacts Ssl2 lobe 2, suggesting that the E-bridge can influence the conformational ratcheting in the Ssl2 ATPase that occurs during DNA translocation.

TFIIH and Pol II phosphorylation

The PIC-cMed structure provides details on the previously described PIC-Mediator interfaces4, and suggests conformational changes in Mediator upon PIC binding (Fig. 6, Extended Data Fig. 8, Supplementary Video 3). The Mediator head module is largely unchanged48, but the conformation of the middle module differs from that in the cMed structure5 (Extended Data Fig. 8c). The submodules in the middle module apparently undergo concerted movements. Whereas the plank rotates to bind the Pol II foot, the hook and knob undergo swinging motions and the beam moves towards the head module jaws. Comparison with the cMed cryo-EM structure49 also suggests conformational changes in Mediator upon PIC binding.

The PIC-cMed structure further reveals an additional density for the Kin28-Ccl1 (CDK7-cyclin H) kinase-cyclin pair on the outer surface of cMed (Fig. 6). This density is located above one of two openings that flank the knob at the Mediator head-middle interface. The kinase-cyclin pair resides between the Mediator hook, knob, and shoulder, roughly consistent with its previously reported position10. The density for Kin28-Ccl1 is weaker than density for cMed or TFIIH, indicating that the kinase-cyclin pair retains some mobility.

How does the TFIIH kinase reach its phosphorylation substrate, the Pol II CTD? The linker to the mobile CTD extends from Pol II towards the inner surface of Mediator that lines a previously described cradle formed between Mediator and Pol II4 (Fig. 6). To reach the kinase, the CTD may exit the cradle and extend around Mediator or through Mediator10. However, the CTD crosslinks to the inner surface of the cradle5, suggesting that it resides in the cradle, where it can be accommodated if it adopts a compact globular shape50. The TFIIH kinase may access the CTD through the openings at the head-middle interface. Phosphorylation of CTD regions would then lead to repulsion between accumulating negative charges, expansion of the CTD globule in the cradle, a weakening of the Pol II-Mediator interaction and Mediator dissociation. Loss of Mediator destabilizes the PIC and would facilitate Pol II escape from the promoter.

Conclusions

We aimed at detailed structures of the yeast Pol II PIC and its complex with Mediator ever since the structure of the core Pol II enzyme was determined50. Important steps towards this goal included the Pol II-TFIIB crystal structure, which led to minimal models of the closed and open promoter complexes51, and our recent structures of cPIC6 and cMed5. The critical step reported here was to prepare recombinant TFIIH, to derive its structure, and to arrive at

structures of the PIC and the PIC-cMed complex. The PIC-cMed complex lacks TFIID and the Mediator tail module, but their location on the PIC was derived by others in the human52 and yeast10 systems, respectively.

The structures presented here define interactions of TFIIH within the PIC and interactions of cMed with the PIC, and provide unexpected insights. First, anchoring of TFIIH to the cPIC involves a subunit of the TFIIH kinase module, ensuring that complete TFIIH is incorporated into the PIC. Second, a mobile extension of TFIIE tethers multiple parts of TFIIH, including the Ssl2/XPB ATPase. Third, the TFIIH kinase is mobile in the PIC, but adopts a preferred location on Mediator when cMed binds the PIC. Finally, PIC-bound Mediator contains two openings at its head-middle interface that may allow access of the TFIIH kinase to the Pol II CTD residing in the cradle. The structures thus provide the basis for future mechanistic studies of TFIIE-stimulated and TFIIH-dependent promoter opening, Mediator-stimulated CTD phosphorylation and promoter escape, and gene regulation at the level of transcription initiation.

Methods

Cloning and protein expression

Full-length subunits of *S. cerevisiae* TFIIH with the exception of Rad3 and Ssl2 were amplified from purified genomic DNA by PCR and transferred into modified pFastBac vectors (derivatives of 438-A and 438-C; Addgene #55218 and #55220) by ligation independent cloning (LIC). The intron in Kin28 was removed by quick-change mutagenesis PCR after initial vector assembly. DNA sequences encoding full-length Rad3 and Ssl2 were obtained as *S. frugiperda* codon-optimized constructs from GeneArt (ThermoFisher Scientific, Waltham, USA), amplified from the vectors by PCR, and transferred into modified pFastBac vectors by LIC. Within the vectors of the 438-series, the TFIIH subunits contain either N-terminal 6xHis- or 6xHis-MBP-tags or remain untagged. N-terminal 6xHis-tags are followed by cleavage sites for either Ulp1 or the rhinovirus protease (3C) whereas the N-terminal 6xHis-MBP-tags are followed by a modified cleavage site for tobacco etch virus (TEV) protease. After separate transfer of each gene into a 438-vector, the single vectors were combined by successive rounds of LIC to generate a 7-subunit construct encoding the genes for core-TFIIH (Rad3, Ssl1, Ssl2, Tfb1, Tfb2, Tfb4, Tfb5) and a 3-subunit construct with the genes for the TFIIH kinase module (encoding Ccl1, Kin28, Tfb3). Each subunit is preceded by a PolH promoter and followed by a SV40 termination site. Within these constructs, the 6xHis-MBP-tags are placed on either Ssl1, Tfb4 or Kin28. Plasmid sequences are available upon request. Preparation of bacmids, production of insect cell virus of the V0 and V1 stage and protein expression in insect cells were performed essentially as described42. Cells were harvested by centrifugation (238x g, 45 min, 4°C) and resuspended in lysis buffer (400 mM KOAc, 25 mM Na•HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine). The cell suspension was flash cooled in liquid nitrogen and stored at -80°C.

Protein purification

Preparation of *S. cerevisiae* Pol II, TBP, TFIIA, TFIIB, TFIIE, TFIIF and 16-subunit cMed was performed as described4. Protein subunits of *S. cerevisiae* core-PIC and cMed were purified and assembled into subcomplexes TFIIA, TFIIE, TFIIF, Pol II and cMed essentially as reported4. Recombinant *S. cerevisiae* core-TFIIH was purified by consecutive steps of affinity chromatography, ion exchange chromatography and size exclusion chromatography. All purification procedures were performed at 4°C unless stated otherwise. Frozen insect cell pellets were thawed at 25°C, supplemented with catalytic amounts of DNAseI and lysed with an EmulsiFlex-C5 cell disruptor (Avestin, Ottawa, Canada) (3 passages, 12 000 psi). The cell lysate was cleared by centrifugation (79,000x g; 60 min) and the protein-containing soluble fraction was filtered through 0.8 μM syringe filters (Merck Millipore, Billerica, USA). The supernatant was then applied to a GE XK 16-20 column (GE Healthcare, Little Chalfont, United Kingdom) containing a bed volume of 25 mL amylose resin (New England Biolabs, Ipswich, US) and pre-equilibrated in buffer M-300 (300 mM KOAc, 25 mM K•HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole). After application of core-TFIIH-containing lysate supernatant, the column was washed with 3 CV of buffer M-300 and the protein was eluted with 2 CV buffer ME (350 mM KOAc, 25 mM K•HEPES pH 7.5, 10% glycerol (v/v), 50 mM maltose and 5 mM β-mercaptoethanole) onto a GE HiTrap Heparin HP (5 mL) column pre-equilibrated in buffer M-350 (350 mM KOAc, 25 mM K•HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole). The column was washed with 3 CV of buffer M-350 and the protein was eluted with a linear gradient of 0-30% buffer M-2000 (2 M KOAc, 25 mM K•HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole) in 20 CV. Peak fractions were pooled, supplemented with 1 mg 6xHis-TEV protease, 0.5 mg 6xHis-3C protease and 0.5 mg 6xHis-Ulp1 protease and kept at 4°C for 6 hours. The cleaved sample was subjected to anion exchange chromatography using a GE HiTrap Q HP (1 mL) column pre-equilibrated in buffer A-400 (400 mM KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole). After sample application the column was washed with 10 CV buffer A-400 and the protein was eluted with a linear gradient from 0-30% buffer A-2000 (2 M KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole) in 80 CV. Fractions containing stoichiometric 7-subunit core-TFIIH were pooled, concentrated using a Vivaspin 6 MWCO 50 000 (GE Healthcare, Little Chalfont, United Kingdom) centrifugal device and applied to a GE Superose12 10/300 GL size exclusion column pre-equilibrated in gel filtration buffer (600 mM KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 2 mM TCEP). Peak fractions were pooled, concentrated to 4 mg/mL using a Vivaspin 500 MWCO 50 000 (GE Healthcare, Little Chalfont, United Kingdom) centrifugal device, aliquoted, flash-cooled in liquid nitrogen and stored at -80°C. Typical yields were in the range of 0.3-0.4 mg per 1 L of insect cell culture.

The TFIIH kinase module was prepared similarly. After cell lysis and lysate clearance, the sample was loaded onto a GE XK 16-20 column containing a bed volume of 25 mL amylose resin pre-equilibrated in buffer M-200 (200 mM KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole). The column was washed with 3 CV buffer M-300 and the protein was eluted with 2 CV buffer ME (200 mM KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 50 mM maltose and 5 mM β-mercaptoethanole). Peak fractions were pooled, supplemented with 1 mg 6xHis-TEV protease and 0.5 mg 6xHis-3C protease

and kept at 4°C for 6h. The cleaved protein sample was subjected to anion exchange chromatography using a GE HiTrap Q HP (1 mL) column pre-equilibrated in buffer M-200. After sample application the column was washed with 10 CV of buffer M-200 and the protein was eluted with a linear gradient from 0-30% buffer A-2000 (2 M KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole) in 80 CV. Fractions containing stoichiometric kinase trimer were pooled, concentrated using a Vivaspin 6 MWCO 10 000 (GE Healthcare, Little Chalfont, United Kingdom) centrifugal device and applied to a GE Superdex200 10/300 GL size exclusion column pre-equilibrated in gel filtration buffer (150 mM KOAc, 25 mM K•HEPES pH 7.5, 5% glycerol (v/v), 2 mM TCEP). Peak fractions were concentrated to 7 mg/mL using a Vivaspin 500 MWCO 10 000 (GE Healthcare, Little Chalfont, United Kingdom) centrifugal device, aliquoted, flash-cooled in liquid N_2 and stored at -80°C. Typical yields were in the range of 1.0 mg per 500 mL of insect cell culture.

Preparation of the PIC-cMed complex

Closed PIC-cMed complex was prepared according to a protocol adapted from the previously reported assembly scheme4, but with a slightly altered nucleic acid scaffold. The 106 nucleotide scaffold is based on the HIS4-promoter sequence (template: 5'-TGACACAGCGCAGTTGTGCTATGATATTTTTATGTATGTACAACACACATCGGAGGTGAATCGAACGTTCCATAGCTATTATATACACAGCGTGCTACTGTTCTC G-3'; non-template: 5'-CGAGAACAGTAGCACGCTGTGTATATAATAGCTATGGAACGTTCGATTCACCTCCGATGTGTGTTGTACATACATAAAATATCATAGCACAACTGCGCTGTGTCA-3') and contains additional downstream DNA. Complete 10-subunit TFIIH was reconstituted from the 7-subunit core and the kinase trimer at 4°C prior to formation of the PIC-cMed complex. The PIC-cMed complex was assembled for cryo-EM according to the order in Extended Data Table 1. Beginning with the formation of a Pol II-IIF complex, the other initiation factors were added to generate a Pol II/IIA-IIB-TBP-IIF-DNA complex. TFIIE was incubated with previously assembled 10-subunit TFIIH for several minutes before being added to the Pol II-containing complex. After incubation for 5 minutes, buffer S (25 mM K•HEPES pH 7.5, 2 mM $Mg(OAc)_2$, 2.5% glycerol (v/v), 1 mM TCEP), with an appropriate amount of AMP-PNP to reach a final concentration of 0.75 mM, and cMed were added. The PIC-cMed complex was incubated for another 120 minutes shaking gently at 400 rpm. Unless stated otherwise, all incubation steps were performed at 25°C.

The PIC-cMed sample was centrifuged at 21,000x g for 10 minutes and subjected to sucrose-gradient centrifugation in a 5 mL centrifugation tube. The gradient was generated from a 15% sucrose light solution (15% (w/v) sucrose, 150 mM KOac, 25 mM K•HEPES pH 7.5, 2 mM $Mg(OAc)_2$, 2.5% glycerol (v/v), 1 mM TCEP, 0.75 mM AMP-PNP) and a 40% sucrose heavy solution (40% (w/v) sucrose, 150 mM KOac, 25 mM K•HEPES pH 7.5, 2 mM $Mg(OAc)_2$, 2.5% glycerol (v/v), 1 mM TCEP, 0.75 mM AMP-PNP) containing 0.13% (v/v) glutaraldehyde crosslinker with a BioComp Gradient Master 108 (BioComp Instruments, Fredericton, Canada). Centrifugation was performed at 175,000x g for 16 hours at 4°C. Subsequently 200 μL fractions were collected and quenched with a mix of 10 mM aspartate and 30 mM lysine for 10 minutes. Fractions containing crosslinked PIC-cMed

complex were dialyzed for 10 hours in dialysis buffer (150 mM KOac, 25 mM K•HEPES pH 7.5, 2 mM $Mg(OAc)_2$, 1 mM TCEP) in Slide-A-Lyzer MINI Dialysis Devices (2 mL, 20 000 MWCO) (ThermoFisher Scientific, Waltham, USA) to remove sucrose and glycerol. The dialyzed sample was concentrated to 0.7 mg/mL using a Vivaspin 500 MWCO 100 000 (GE Healthcare, Little Chalfont, United Kingdom) centrifugal device and applied to cryo-EM grids.

Cryo-electron microscopy

Cryo-EM data collection was performed on R1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany). Grids were glow-discharged for 45 seconds before application of 5 μL concentrated PIC-cMed sample, blotted for 5 seconds and vitrified by plunging into liquid ethane with a Vitrobot Mark IV (FEI Company, Hillsboro, US) operated at 4°C and 100% humidity. Cryo-EM data were acquired on a FEI Titan Krios G2 transmission electron microscope (FEI, Hillsboro, USA) operated in EFTEM mode at 300 kV and equipped with a K2 Summit direct detector (Gatan, Pleasanton, USA). Automated data acquisition was carried out using the FEI EPU software package at a nominal magnification of 105,000x (1.37 Å/pix). A total of 14,000 image stacks were collected at a defocus range from -0.5 μM to -5.0 μM. Each stack contained 40 frames that were acquired over a 10 seconds exposure time window in the counting mode of the camera. A dose rate of 4.2 $e^-/Å^2s$ was applied, resulting in a total dose of 42 $e^-/Å^2$.

Image processing

Cryo-EM image frames were stacked and processed with MotionCor253 and CTF parameter estimation was performed with Gctf54. CTF correction and subsequent image processing were performed with the RELION 2.0.4 package23,55 unless indicated otherwise. Post-processing of refined models was performed with automatic B-factor determination in RELION and resolution was reported based on the gold-standard FSC (0.143 criterion) as described56 unless indicated otherwise. Local resolution estimates were determined using a sliding window of 40^3 voxels as described6. To obtain an initial particle set, coordinates of ca. 15 000 particles were determined semi-automatically with the e2boxer.py tool implemented in EMAN257. The coordinates were imported into RELION and the respective particles were extracted with a 380^2 pixel box and normalized. Reference-free 2D class-averages were calculated and 20 representative 2D classes were selected. These were low-pass filtered to 20 Å and used as templates for automated particle picking on the first 700 micrographs, resulting in ca. 200,000 particles. Particles were extracted with a 380^2 pixel box size, normalized and screened by a combination of manual inspection and iterative rounds of reference-free 2D-classification. From the obtained improved 2D class-averages, 20 representative 2D classes were selected, low-pass filtered to 20 Å and used as templates for automated picking on the remaining micrographs with RELION. Initially ca. 1.6 million particle images were obtained. Particles were extracted with a box size of 350^2 pixel, normalized and screened using a combination of iterative rounds of reference-free 2D- and template-guided 3D-classification with image alignment combined with manual inspection of the images in specific classes. An initial reference (ModelI) for the screening 3D-classifications had been obtained by performing one pre-3D-classification with the initial 200,000 particles using a 60 Å low-pass filtered EM map of the core-PIC-cMed complex

(EMD-2786)4 as reference. Calculation of five 3D classes resulted in one class with the complete PIC-cMed complex. This class was used as 'Model I' for the screening 3D-classifications after low-pass filtering to 60 Å. During the screening process, ca. 60% of the initial 1.6 million particles were discarded, resulting in 650,000 input particles. Using 'Model I' as the initial reference, iterative rounds of hierarchical 3D-classification with image alignment were performed as outlined in Extended Data Fig. 2. After the first round of classification, classes with clearly visible density for cMed were selected. The same procedure was applied for classes with clear TFIIH density but no density for cMed, resulting in a separation of the classification tree in one branch for the PIC-cMed particles and one branch for PIC particles that lacked cMed. Prior to the second round of 3D-classification, new reference models ('Model PIC' and 'Model PIC-cMed') were generated from the best classes of the first round of 3D-classification and low-pass filtered to 60 Å. The second round of template-guided 3D-classification for the PIC-cMed branch was consequently performed with 'Model PIC-cMed' as a reference whereas for the PIC branch 'Model PIC' served as reference. Subjecting the best 3D class of the PIC branch to a focused 3D-refinement with a local mask encompassing only TFIIH resulted in a reconstruction with a resolution of 7.4 Å (after post-processing) from 32,000 particles.

Flexible refinement (WarpCraft)

Both the PIC and the PIC-cMed complexes showed intrinsic flexibility. In particular, TFIIH was flexible with respect to cPIC, and cMed was flexible with respect to PIC. Although such flexibility can be dealt with using local refinement in RELION, this leads to composite density maps. In order to obtain reconstructions with a continuous density throughout the entire maps, we developed and used a flexible refinement tool, WarpCraft. To calculate the reconstructions, the best classes of the second round of 3D-classification of the PIC and PIC-cMed branches were merged as shown in Extended Data Fig. 2. The first 20 normal modes were calculated as described58, using 15,000 pseudo atoms derived from a globally refined map of the complexes, and a distance cut-off of 8 Å. Maps were then automatically divided in 20 regions with the objective to minimize the mean intra-region across all normal modes. The region masks were given a raised cosine fall-off of 8 pixels within the particle boundaries to create a slight overlap, and 16 pixels outside the boundaries. The mask values were normalized to have a sum of 1 in each intra-particle voxel. Separate reference volumes were then generated by multiplying each initial, locally low-pass filtered half-map by each region mask. The local filtering was performed with a 40 pixel window and an FSC threshold of 0.7. The refinement procedure aimed to find the optimal linear combination of normal modes that described the conformation observed in each experimental projection. To achieve this, the squared difference between the experimental projections, and the sum of all region reference projections multiplied by the previously determined CTF was minimized using the L-BFGS algorithm59. The orientation of each region in the projections was defined as the global particle rotation and translation, adjusted by the rigid body transform that best described the shift of pseudo-atoms within that region, as defined by the current linear combination of normal modes for the particle. After 20 optimization steps, the reconstructions were obtained as follows. For each half-map and each region, a reconstruction was calculated using the particle orientations adjusted by that region's rigid body transform determined in the optimization. The region reconstructions were multiplied

by masks identical to those used for optimization, except that the fall-off region outside the particle boundaries was also normalized to have a sum of 1 in each voxel, so as not to create additional masking in the result. The masked reconstructions were added up to form the final half-map volumes, and the local resolution for each region was calculated with a 40 pixel window and an FSC threshold of 0.3. This process was repeated until the resolution values converged, usually after 5-6 iterations. The code for WarpCraft is available as Supplementary Data 1.

Structural modeling

For structural modeling we used both the continuous EM maps obtained by WarpCraft and EM maps with focus on specific regions in TFIIH. Model placement and docking of rigid bodies into the EM maps was performed with UCSF Chimera60. The I-TASSER61,62, SWISS-Model63,64 and Rosetta65,66 tools were utilized for the generation of homology models of various PIC-cMed components as indicated in Supplementary Data Table 1 and Supplementary Data Table 4. Manual modification of models and *de-novo* model building procedures were performed with COOT67. The model of the *S. cerevisiae* cPIC6 was placed into the EM map and the Pol II clamp and stalk regions, as well as TFIIA, TFIIF and peripheral regions in Rpb3, Rpb6, Rpb8, Rpb9 and Rpb12 were adjusted as rigid bodies. The model of TFIIB was extended in the B-linker and B-reader regions based on the Pol II-TFIIB crystal structure68 (PDB 4BBR). Homology models for TFIIE subunits Tfa1 and Tfa2 were generated based on the *H. sapiens* crystal structures of TFIIE69 (PDB 5GPY) and flexibly fitted into the TFIIE density, replacing the previous TFIIE model. The *S. cerevisiae* cMed homology model was adapted from the previously generated homology model of the *S. pombe* cMed crystal structure5 (PDB 5N9J). To improve the fit to the EM map, cMed was divided into seven rigid bodies (head module, knob, hook-connector, plank, beam RWD1-UBC1, beam RWD2, beam UBC2) that were placed in the density individually. Downstream DNA was generated by placing three pieces of ideal B-DNA into the density, connecting these in COOT and performing alternating rounds of real space refinement with secondary structure restraints and geometry optimization in PHENIX70. For a summary on structural modeling of proteins refer to Supplementary Data Table 1 and Supplementary Data Table 4.

We generated a conservative model of *S. cerevisiae* TFIIH with the use of available structural information. Models of domains were first derived based on structures of TFIIH homologues from different species and on other structures with regions of partially related sequences. Homology models were generated for the Tfb1 BSD1 and BSD2 domains, for the three Tfb2 HTH motives, for the Tfb3 RING-Finger, for the Tfb4 vWA-fold, for the extended Zn-Finger motifs in Tfb4 and Ssl1, and for the Ssl1 RING-Finger. These models were derived from the *H. sapiens* NMR structure of the BSD1 domain (PDB 2DII), the *S. aureus* CadC crystal structure71 (PDB 1U2W), the *H. sapiens* MUS81 NMR structure72 (PDB 2MC3), the *P. furiosus* TrmBL2 crystal structure73 (PDB 5BOX), the NMR structure of the *H. sapiens* Mat1 RING-Finger74 (1G25), the crystal structure of the *H. sapiens* p34 vWA-fold75 (PDB 4PN7), the crystal structure of a *H. sapiens* E3 ubiquitin ligase (PDB 3LRQ), the crystal structure of *P. furiosus* rubrerythrin76 (PDB 1NNQ) and the NMR structure of the *H. sapiens* p44 RING-Finger77 (1Z60), respectively. In addition, two 3-helix bundle domains, one located at the C-terminus of Tfb1 (residues 543-639) and one located

C-terminally of the Tfb3 RING-Finger (residues 71-145) were modeled *ab-initio* utilizing the QUARK server78. Together with the crystal and NMR structures of the Tfb1 PHD79 (PDB 1Y5O), the Tfb2/Tfb5 dimerization domains28 (PDB 3DGP) and the vWA-fold of Ssl180 (PDB 4WFQ), the homology- and *ab-initio* models listed above were placed into the density and rigid-body adjusted. If the correct position of the models could not be deduced from the electron density directly, placement was performed on the basis of BS3- and SBAT-derived crosslinks that had been published10,24,25 or EDC-derived crosslinks obtained in this study (Extended Data Fig. 3).

Several homology models were subjected to conservative modifications, in particular minor truncations, short α-helical extensions and positional corrections, to improve their fit to the electron density manually. The *S. cerevisiae* crystal and NMR structures exhibited a good fit to the electron density and did not require modification with the exception of the Tfb1 PHD that was C-terminally extended (residues 115-121). Homology models for the ATPases Rad3 and Ssl2 were generated from crystal structures of their *T. acidophilum*29 (PDB 2VSF), *A. fulgidis*81 (PDB 2FWR) and *H. sapiens*82 (PDB 4ERN) homologues. The models were split into their domains (Rad3: Lobe 1, FeS-cluster, ARCH, Lobe 2; Ssl2: Lobe 1 and Lobe 2) and placed individually into the electron density. Lobe 1 and Lobe 2 of Rad3 did not require further adaptation. The FeS-cluster was placed by superpositioning the *T. acidophilum* Rad3 structure onto the TFIIH model in COOT and extracting the coordinates of the Fe and S atoms. A backbone model of the ARCH domain was generated with Gorgon83,84 and used as an additional input to calculate a second homology model, which then was adjusted to the density. It accounted for an evolutionary difference between *S. cerevisiae* and *T. acidophilum* that had resulted in an extension of two α-helices and an insertion of one α-helix and a loop (residues 255-347). In the Ssl2 homology model an additional α-helix (residues 468-481) was placed into well-defined density substituting for an initially unstructured stretch of residues. Two loops (residues 426-451, 692-702) with significant deviation from the EM density were manually adjusted. The location of four Tfb1 α-helices (residues 308-330, 369-394, 465-483, 495-519) and two TFIIE α-helices (residues 267-289, 349-373) was confirmed by XL-MS analysis as described above and the respective α-helices were placed into the corresponding density. Additionally, a few linkers and α-helical regions within TFIIH subunits Tfb1 (residues 219-251, 295-307, 331-353, 484-495), Tfb2 (residues 3-40, 113-159, 195-214, 433-450), Tfb4 (residues 89-97, 103-114, 257-273), and Ssl1 (residues 309-324, 373-386) which could be clearly traced in the EM maps and assigned respectively were built *de-novo*. Lastly, the TFIIE acidic peptide (residues 407-417) interacting with the PHD of Tfb1 was modeled based on the co-NMR structure of the *H. sapiens* TFIIE C-terminus and the PHD of p6285 (PDB 2RNR).

The model fit to the EM maps was further optimized by iterative rounds of flexible fitting with vmd86 and MDFF87. Each flexible fitting procedure was divided in three simulation steps, starting with a simulation at room temperature, followed by a cooling step to 0 K and a third step in which the simulation was performed at 0 K. Flexible fitting was performed without domain restraints for small units and with domain restraints once models had been combined into larger entities.

Europe PMC Funders Author Manuscripts

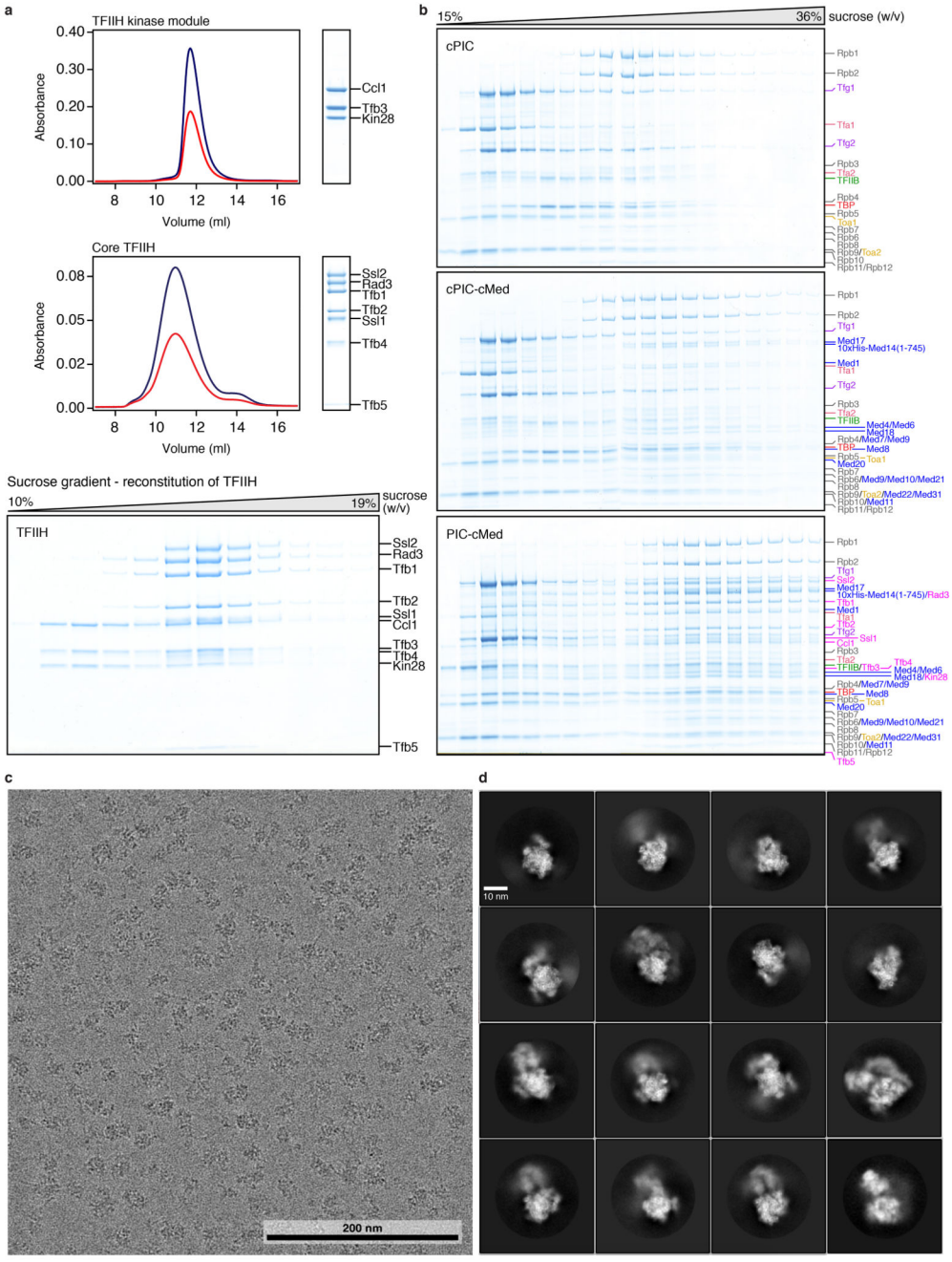
Density-adjusted PIC and PIC-cMed models were refined using the geometry minimization routine in PHENIX70 with applied secondary structure and rotamer restraints. A brief overview of EM-data collection, data processing and model statistics for the final PIC and PIC-cMed models is provided in Extended Data Table 2. Figures were generated using UCSF Chimera60.

Crosslinking analysis

PIC-cMed sample was crosslinked with a final concentration of 200 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (ThermoFisher Scientifc, Waltham, USA) in the sucrose heavy solution during gradient centrifugation. Fractions from the sucrose gradient were quenched with 50 mM ammonium bicarbonate. Fractions were dialyzed as before to remove sucrose and pooled for precipitation. Precipitated sample was dissolved in 50 μl buffer containing 8 M urea and 50 mM ammonium bicarbonate. Crosslinked sample was digested 1:20 (w/w) with trypsin and peptides were enriched by peptide size-exclusion chromatography and analyzed in duplicate on an Dionex UltiMate 3000 RSLCnano HPLC system (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific). MS acquisition was performed as described88 with the exception that peptides were separated on the analytical column using a 63-min linear gradient. The datasets were analyzed with pLink 1.2389 against a database containing the sequences of the protein components in the complex. Database search parameters included mass accuracies of MS1 < 10 ppm and MS2 < 20 ppm, carbamidomethylation on cysteine as a fixed modification and oxidation on methionine as a variable modification. The number of residues of each peptide on a cross-link pair was set between 5 and 40 amino acids. A maximum of two trypsin-missed cleavage sites was allowed. An initial false discovery rate (FDR) cutoff of 1% was set. For simplicity, the crosslink score was represented as a negative logarithm value of the original pLink score and identified spectra with a score larger than three were considered. Results were visualized using the xiNET online server90 and the XLink Analyzer Plugin91 for UCSF Chimera60. New crosslinks are summarized in Extended Data Fig. 3.

Europe PMC Funders Author Manuscripts

Extended Data



Extended Data Figure 1. Preparation of TFIIH and PIC-cMed complex.

a. Preparation of recombinant TFIIH. Analysis of purified TFIIH core and kinase modules by size exclusion chromatography and SDS-PAGE revealed high purity and homogeneity of the complexes with apparently stoichiometric subunits. SDS-PAGE analysis of fractions 1-13 of a sucrose gradient centrifugation after reconstitution of TFIIH from purified core and kinase modules. A shift of the bands originating from the subunits of the kinase module

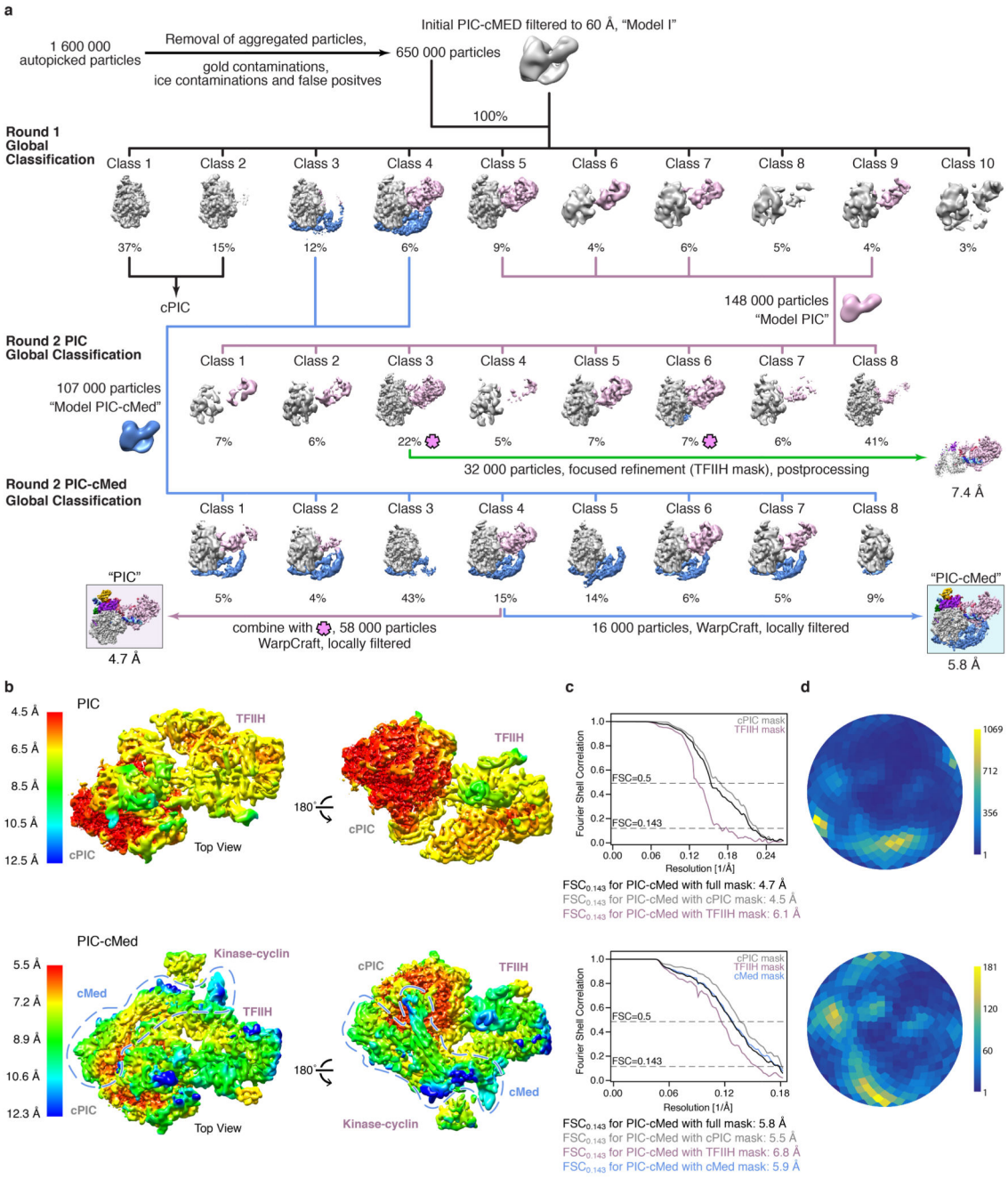
(Ccl1, Kin28 and Tfb3) by four fractions was detected, indicating formation of complete TFIIH. This experiment was repeated multiple times with equivalent results.

b. Assembly of complexes. SDS-PAGE analysis of fractions 1-19 of 15-40% sucrose gradient centrifugations (Methods). Labeling of protein subunits according to the color scheme in Figs. 1 and 2. The analysis demonstrates successful formation of the cPIC, cPIC-cMed and PIC-cMed complexes (top to bottom). Bands originating from Pol II, cMed and TFIIH are shifted by several fractions, indicating formation of higher-order complexes. Subunits are present in apparently stoichiometric amounts. This experiment was repeated multiple times with equivalent results.

c. Representative cryo-EM micrograph of PIC-cMed complex. A scale bar is provided. This experiment was repeated multiple times with equivalent results.

d. 2D-class averages reveal 2D reconstructions from particles with clear signal for TFIIH and/or cMed adjacent to the centrally located Pol II density. A scale bar is provided.

Europe PMC Funders Author Manuscripts



Extended Data Figure 2. Cryo-EM data processing and quality of the obtained reconstructions.

a. Particle sorting and classification tree used for 3D reconstruction of the PIC and PIC-cMed complex at nominal resolutions of 4.7 Å and 5.8 Å, respectively. The distinct branches of the classification tree (Methods) are highlighted in pink (PIC) and blue (PIC-cMed). In a conventional focused refinement approach in RELION23,55, the best-resolved PIC class was reconstructed with a local TFIIH mask, resulting in a focused map with a nominal resolution of 7.4 Å (green branch) that was not deposited.

Europe PMC Funders Author Manuscripts

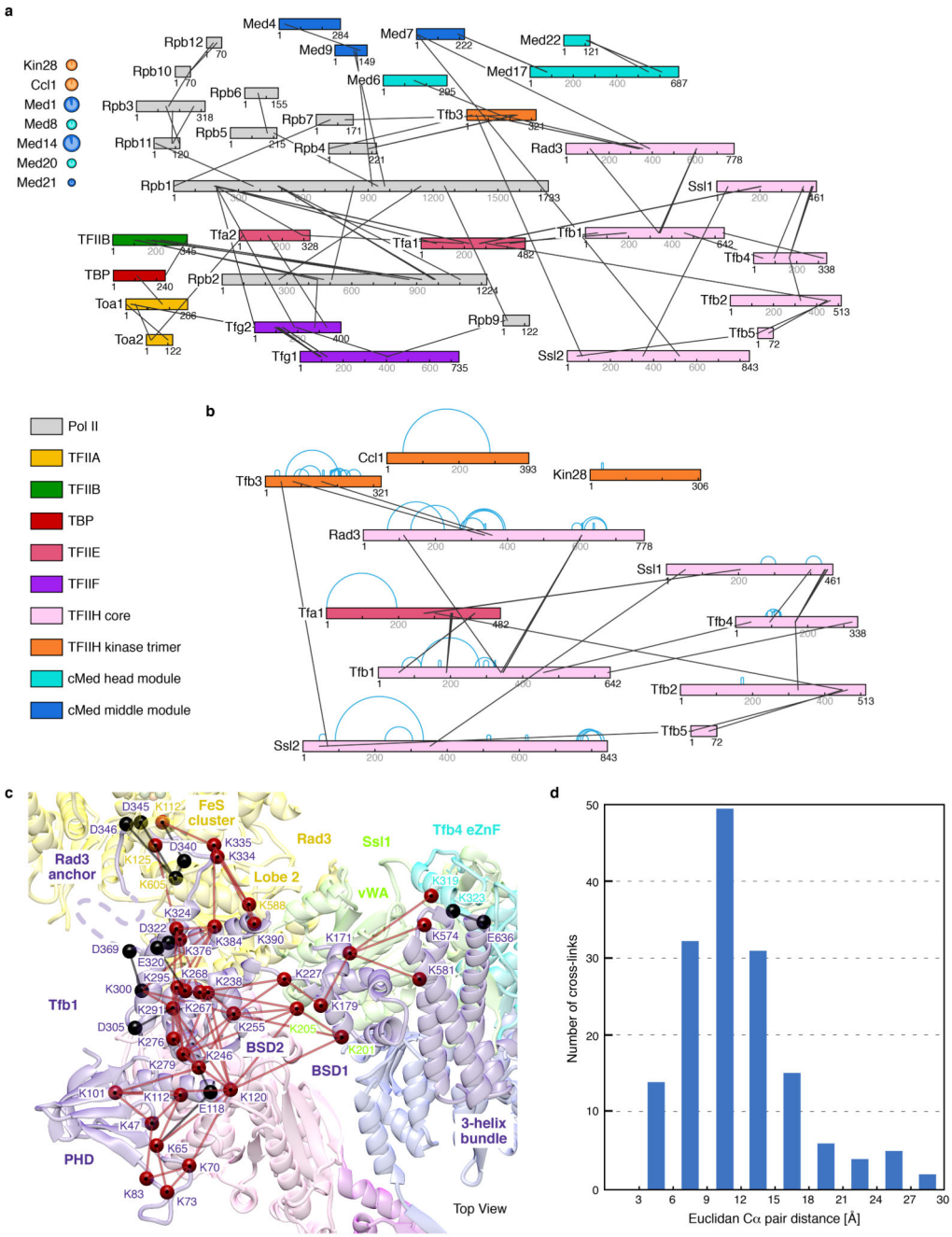
b. Two views of the final reconstructions of PIC and PIC-cMed colored according to local resolution6. The color scheme is indicated.

c. Fourier shell correlation (FSC) between half maps of the final reconstructions of PIC and PIC-cMed. Resolutions for the gold-standard FSC 0.143 criterion are listed. For comparison of distinct regions within PIC and PIC-cMed reconstructions, FSC 0.143 was additionally calculated utilizing local masks.

d. Angular distribution plot for all particles in the final reconstructions of PIC and PIC-cMed. Color shading from blue to yellow correlates with the number of particles at a specific orientation as indicated.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Figure 3. EDC-crosslinking analysis of PIC-cMed.

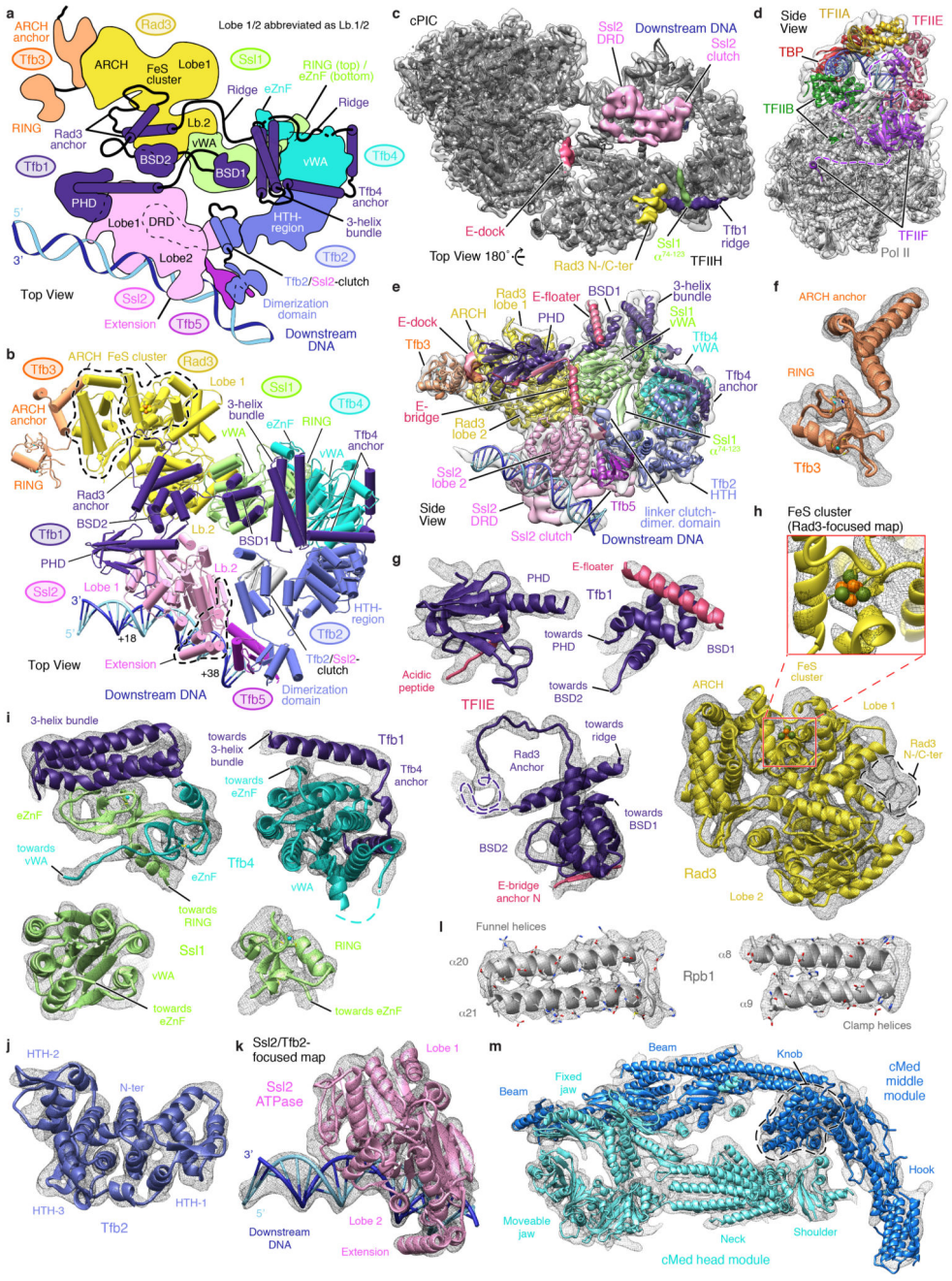
a. EDC-derived inter-subunit crosslinks between selected subunits in the PIC-cMed complex. Observed crosslinks are consistent with the structure of the cPIC and with positions of previously reported BS3- and SBAT-crosslinks. Color code is indicated.

b. EDC-crosslinks observed in TFIIH and between TFIIH and cPIC. Intra- and inter-subunit crosslinks are depicted as blue and black lines, respectively. Crosslinks between the TFIIE Tfa1 C-terminal region and Tfb1, Tfb2 and Ssl1 confirm interactions between TFIIE elements and TFIIH.

c. Crosslinking hub of the Tfb1 N-terminal region. Ribbon representation of Tfb1 (residues 1-353, 369-394, 544-639) and the surrounding domains of Rad3, Ssl1 and Tfb4. BS3-/SBAT- and EDC-derived crosslinks are depicted in red and black, respectively. The displayed crosslinks aided modeling of the Tfb1 PHD, BSD1, BSD2 and Rad3 anchor domains into the cryo-EM density.

d. Statistical analysis of EDC-derived crosslinks. Most observed crosslinks are within a cutoff Cα-distance of 16 Å. Cα-distances of up to 21 Å may be attributed to flexibility of the involved residues and the coordinate error of the model. Some outliers with Cα-distances of 22-30 Å were observed for the well-defined cPIC and Rad3 structures and may have originated from over-crosslinking of particles.

Europe PMC Funders Author Manuscripts



Extended Data Figure 4. TFIIH structure and quality of the cryo-EM density.

a. Schematic of TFIIH subunit and domain architecture with bound dsDNA using the top view. Flexible linkers are depicted as black lines. Prominent helices within the folds of the tethering subunit Tfb1 and in Tfb2 are highlighted.

b. Top view of the TFIIH structure in cylindrical representation. Prominent domains are labeled. The DNA register with respect to the putative transcription start site +1 is indicated.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

c. Overall fit of PIC structure into final WarpCraft PIC reconstruction. Observed density for a few remaining regions that could be clearly assigned but were not modeled are highlighted as indicated in Supplementary Data Table 1.

d. Fit of cPIC structure into final WarpCraft PIC reconstruction at a higher contour level than in **(c)** shows the high resolution of the map in this region.

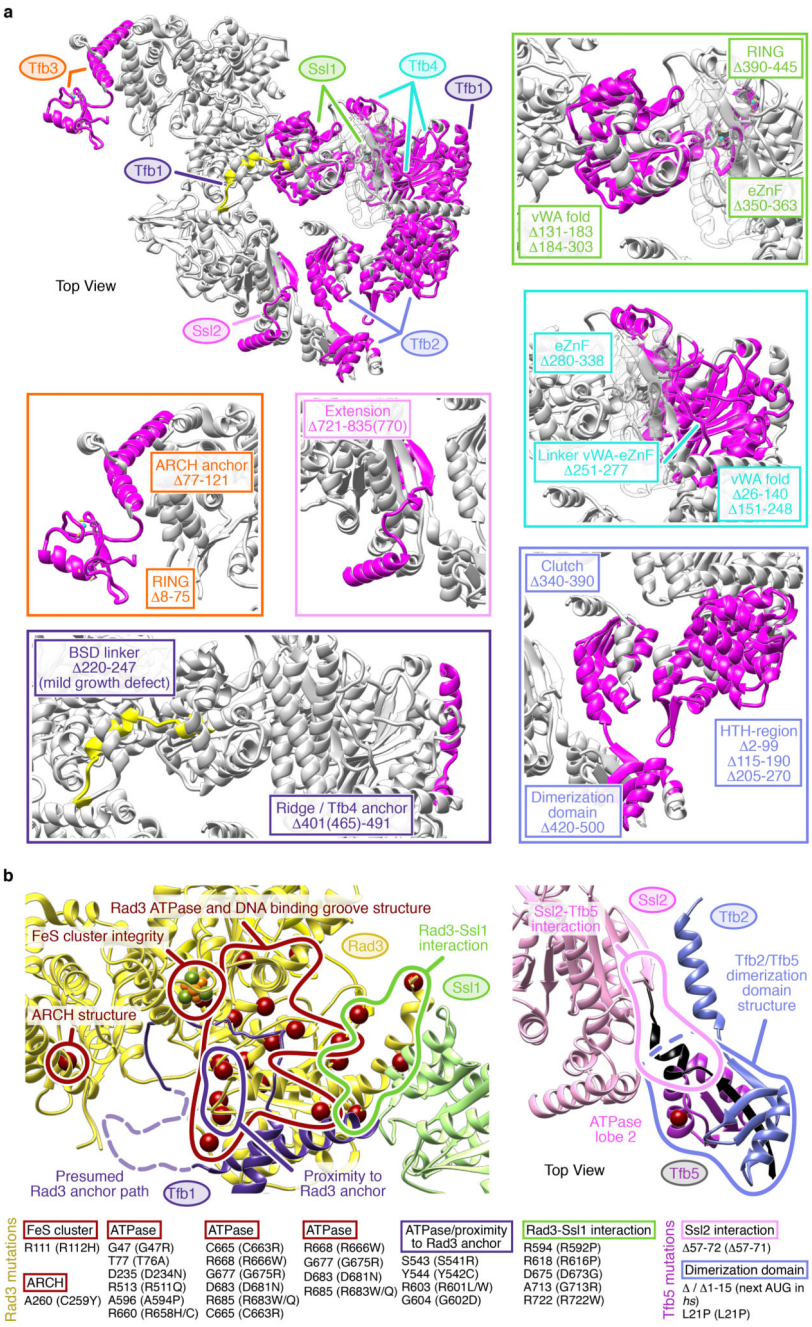
e. Fit of TFIIH model into final WarpCraft PIC reconstruction. EM map reveals secondary structure throughout. Observed density for regions that could be clearly assigned but were not modeled are highlighted (compare Supplementary Data Table 1).

f-k. EM density (black mesh) for domains and subunits of TFIIH reveals secondary structure throughout. Loops and linkers were traced when continuous density between unambiguously placed models was observed. Depicted density is part of either the WarpCraft PIC reconstruction or a focused reconstruction with a local mask on TFIIH core unless indicated otherwise.

l. Cryo-EM reconstruction of the PIC reveals side chain density in well-ordered regions. Depicted are helical regions in the large Pol II subunit Rpb1.

m. Fit of the PIC-cMed model into the final WarpCraft PIC-cMed reconstruction. Structures of cMed head and middle modules account for density within this region.

Europe PMC Funders Author Manuscripts



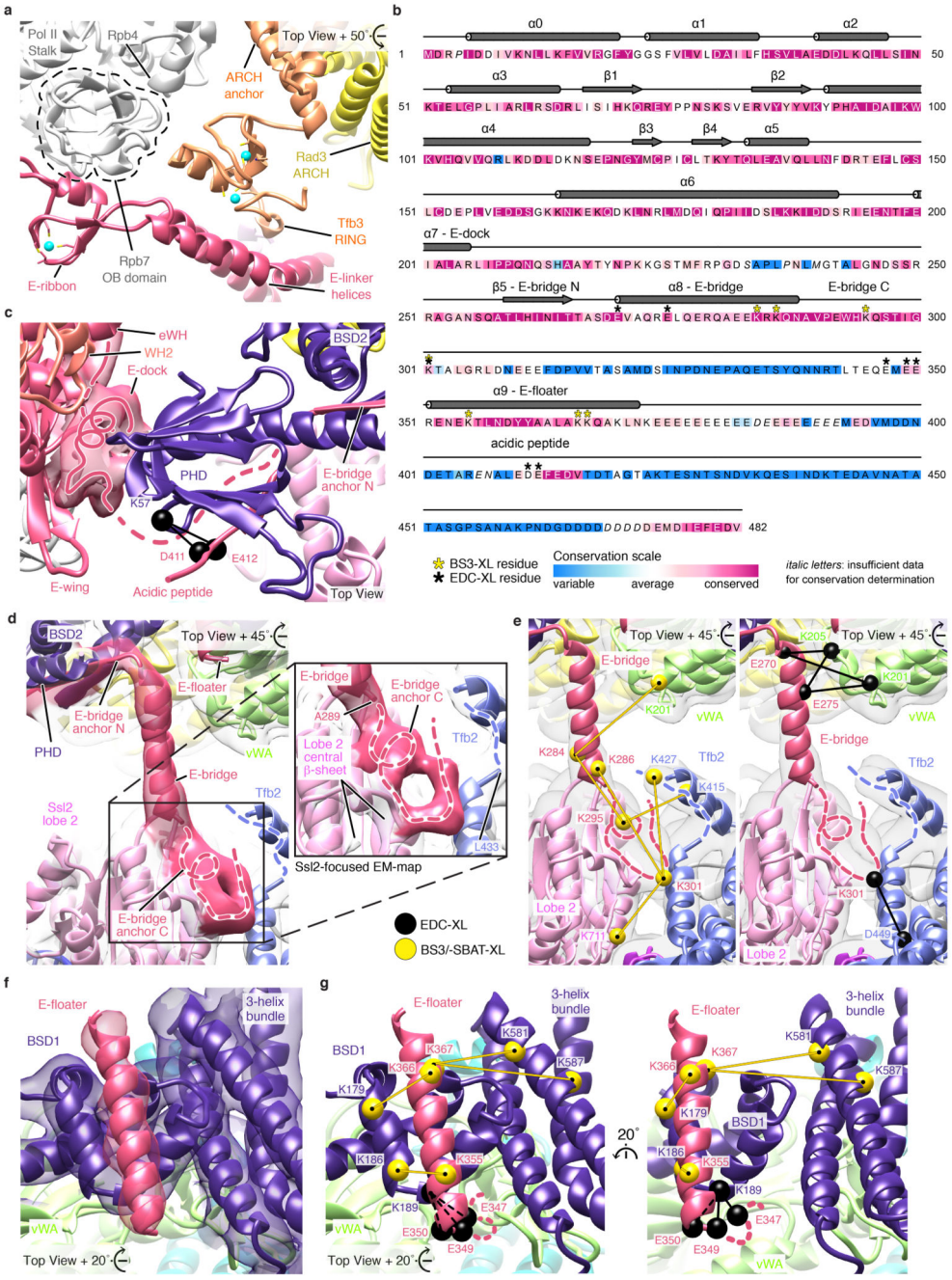
Extended Data Figure 5. Location of mapped essential regions in TFIIH and sites mutated in disease.

a. TFIIH regions essential for cell viability in yeast. Mapping of TFIIH regions identified to be essential in *S. cerevisiae* by *in vivo* deletion analysis33 on the PIC structure revealed that they are generally forming well-ordered regions of the TFIIH core. Structures are viewed from the top (Fig. 1) with regions colored in magenta or yellow if their removal caused cell lethality or growth defects, respectively. Affected TFIIH subunits and ranges of deleted

residues are highlighted in colors according to Fig. 3. For deletions exceeding the modeled residue range, the last modeled residue is indicated in parentheses.

b. Mapping of human disease mutations onto the structures of Rad3 (human XPD) and Tfb5 (human p8). Reported mutations in Xeroderma pigmentosum, Trichothiodystrophy or Cockayne syndrome14,34,35 were included. The sites of point mutations are depicted as red spheres and Tfb5 truncations are colored in black. Color coding of TFIIH subunits as in Fig. 3. A list of yeast residues highlighted in the PIC structure is provided together with the corresponding human mutations in parentheses. Mutation sites are conserved. Rad3 mutations apparently interfere either with the stability and/or the function of the ATPase core or with the Rad3-Ssl1 interaction. Only few mutations target the FeS cluster or ARCH domain. Newly available data on the Rad3 anchor in Tfb1 suggest close proximity to at least four mutation sites that may affect the Rad3-Tfb1 interaction in this region. Tfb5 mutations either abolish Ssl2 binding or the formation of the dimerization domain with the Tfb2 C-terminus, resulting in destabilization of the Ssl2/Tfb2 region. If the clutch domains remain intact, however, a complete disruption of the Ssl2/Tfb2 interaction seems unlikely. We omitted Ssl2 from analysis as our structure does not cover the region in which reported mutations occur.

Europe PMC Funders Author Manuscripts

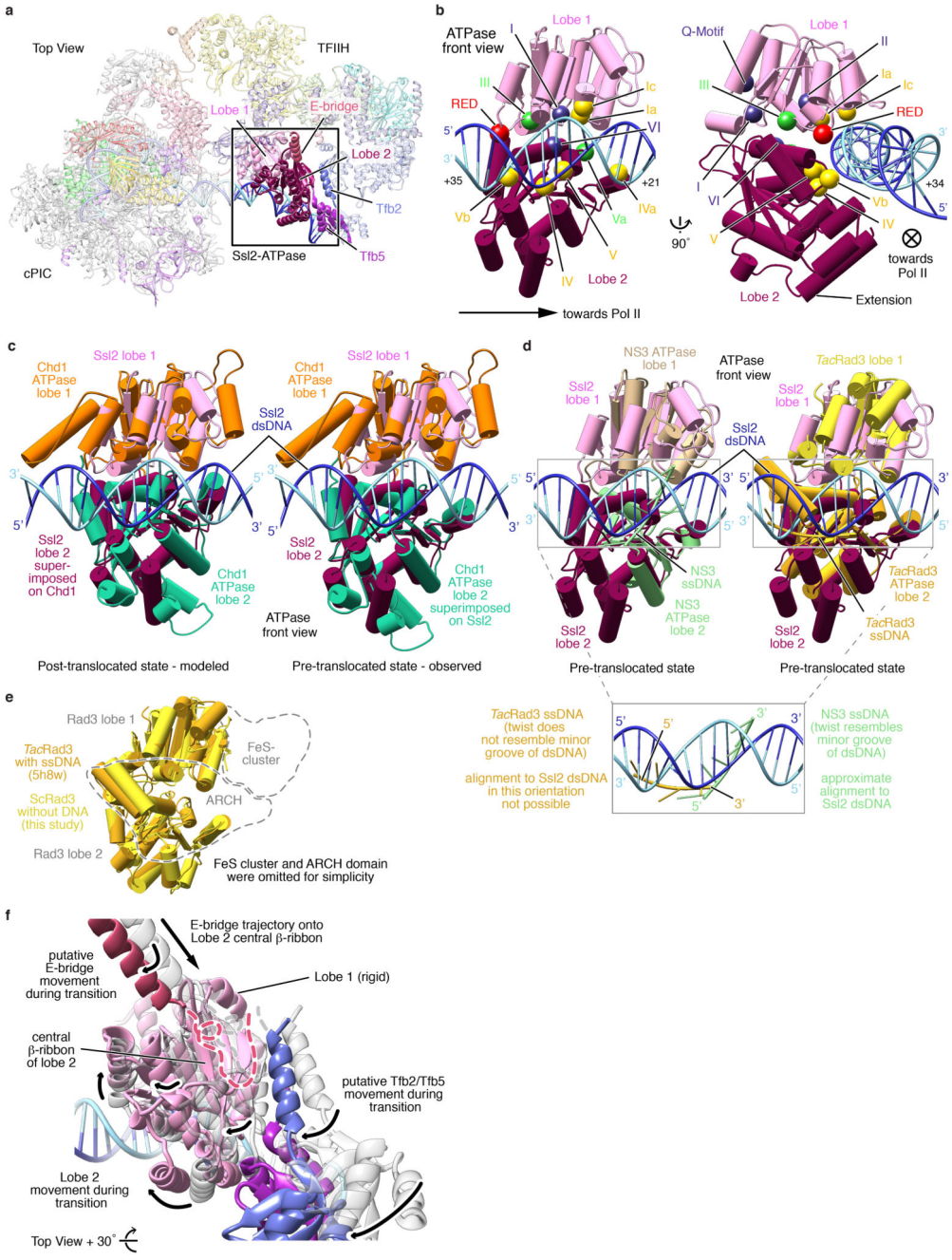


Extended Data Figure 6. TFIIE-TFIIH interactions.

a. Tfb3-Pol II interaction. The TFIIH kinase module subunit Tfb3 (human MAT1) tethers Pol II and the TFIIH core together. Ribbon representation of the Tfb3 N-terminal RING-finger binding in a groove between the Pol II stalk subunit Rpb7 and the TFIIE E-linker helices. The RING-finger is linked to the ARCH anchor which binds the ARCH domain of Rad3.

b. Secondary structure and conservation of TFIIE subunit Tfa1 as determined with CONSURF92. Regions observed in the PIC and PIC-cMed structures are exceptionally well conserved throughout evolution. C-terminal residues with utilized crosslinks are indicated.
c. E-dock. The predicted Tfa1 helix α7 is wedged between the TFIIE eWH domain situated on the Pol II clamp and the PHD of Tfb1 in the TFIIH core. α7 was not modeled due to weak density at the interface of the two major mobile parts of the PIC structure (cPIC and TFIIH) and due to the absence of crosslinks (Methods). The Tfb1 PHD is additionally contacted by the Tfa1 C-terminal acidic region. The identity and directionality of this acidic peptide were unambiguously established by crosslinking (Methods).
d-e. E-bridge. This helix (α8) extends from the Tfb1 BSD2 domain at the center of the TFIIH crescent to the central β-sheet of the Ssl2 ATPase lobe 2. The C-terminal anchor peptide (dashed line) was not modeled into the density due to limited resolution. The identity and directionality of the E-bridge was unambiguously established by independent crosslinking experiments (Methods).
f-g. E-floater. The Tfa1 helix α9 is positioned by the BSD1 domain of Tfb1 and located adjacent to the 3-helix bundle at the center of the TFIIH crescent. The identity and directionality of the E-floater was unambiguously established by independent crosslinking experiments (Methods).

Europe PMC Funders Author Manuscripts



Extended Data Figure 7. Detailed analysis of Ssl2 ATPase conformation and implications for translocase activity.

a. Overview of PIC complex with highlighted Ssl2 (human XPB) ATPase lobes 1 and 2 (in pink and bordeaux, respectively) and interacting domains of Tfb2, Tfb5 and Tfa1.

b. Detailed view on Ssl2 positioned on dsDNA in the presumed pre-translocation state. The ATP analogue AMP-PNP was present in the buffer but was not observed in the active site of the Ssl2 ATPase, supporting the model that we trapped the structure in the pre-translocation state. Register of covered nucleotides with respect to the TSS +1 is indicated. Highlighted

Europe PMC Funders Author Manuscripts

helicase motifs were identified and assigned as described93. Yellow colored motifs are involved in DNA interaction, purple motifs participate in NTP binding and hydrolysis and green motifs are involved in coupling of ATP hydrolysis to DNA binding. Both lobes of the ATPase contact both nucleic acid strands. The highlighted RED motif is essential and strictly conserved throughout the Ssl2/XPB family.

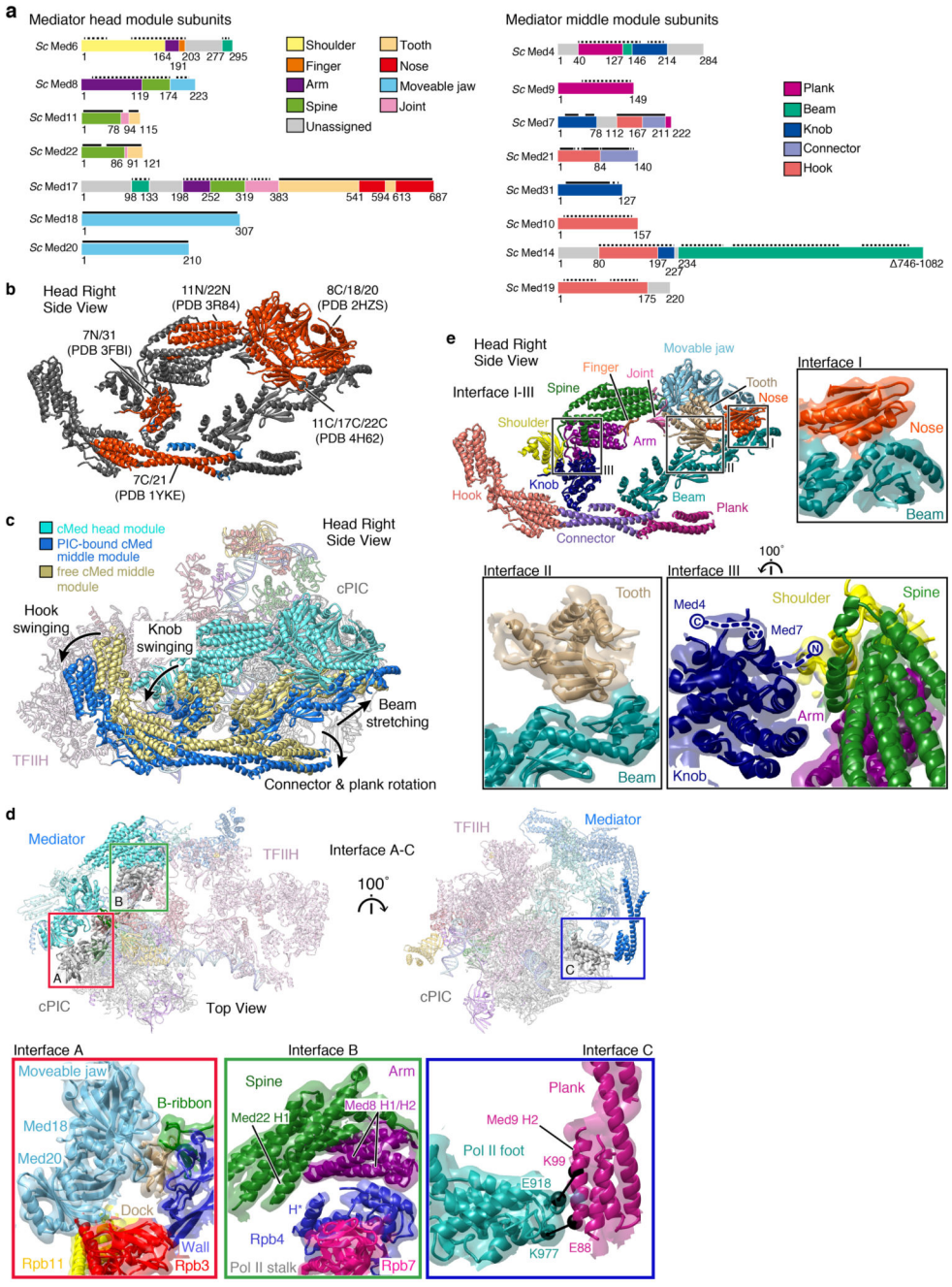
c. Chd1 and Ssl2 ATPases are closely related on a structural level and share the same fold. The presumed post-translocation state of Ssl2 was modeled by separate alignment of ATPase lobe 1 and 2 to the respective lobes in the structure of Chd1 bound to an ATP analogue (PDB 5O9G); the presumed pre-translocated state was modeled vice versa using the Ssl2 structure as reference model. In both states, the structures overlap to a high degree.

d. The Ssl2-DNA arrangement observed in the PIC structure resembles that of 3'-5'-directed rather than 5'-3'-directed members of the SF2 family. Superposition of the Ssl2-dsDNA structure with models of the NS3 (PDB 3KQK)94 and *T. acidophilum* (*Tac*) Rad3 (PDB 5H8W)30 ATPase domains reveals a closer resemblance of Ssl2 to the 3'-5'-helicase NS3. Additionally, the bound DNA fragment in the NS3 model aligned well to the dsDNA the in the Ssl2 structure whereas the bound fragment in the *Tac*Rad3 structure was positioned differently and did not exhibit a minor groove twist as observed for NS3 and Ssl2 in the respective position.

e. Superposition of structures of *Tac*Rad3 and *Sc*Rad3 ATPase domains indicates very high level of structural homology. ATPase lobes 1 and 2 were superimposed separately to account for the absence of bound DNA in the *Sc*Rad3 structure.

f. Putative movement of E-bridge and the Tfb2-Tfb5 dimerization domain upon Ssl2 transition from the presumed pre- to the presumed post-translocated state (grey and color, respectively). Upon movement of lobe 2, the E-bridge may undergo a rotation-translation movement towards Pol II and against its own trajectory onto the central β-ribbon of Ssl2 ATPase lobe 2. The flexible Tfb2-Tfb5 dimerization domain would swing towards Pol II.

Europe PMC Funders Author Manuscripts



Extended Data Figure 8. Structure and conformational changes of core Mediator (cMed).

a. Schematic representation of cMed subunits. Regions contributing to submodules are colored as in the *Sp* cMed crystal structure5. Solid and dashed black lines refer to protein regions that were modeled as atomic or backbone models, respectively.

b. Ribbon model of cMed colored by type of structural model used for interpreting the cryo-EM density. Regions with backbone models based on the *Sp* cMed structure5, regions with atomic models inclusive of the PDB code, and *de-novo* modeled regions are indicated in grey, orange and blue, respectively.

Europe PMC Funders Author Manuscripts

c. Repositioning of the cMed middle module upon PIC binding. The structures of unbound cMed (khaki, PDB 5N9J) and PIC-cMed complex (blue, this study) were superimposed on the cMed head module. The positions of the cMed middle module domains hook, knob, connector, plank and beam apparently undergo conformational changes upon PIC binding, as indicated by arrows. This may cause or enlarge two observed openings at the head-middle interface.

d. PIC-cMed interactions. Structure of the PIC-cMed complex in two views. The three previously identified interfaces4 between cPIC and cMed are indicated. In interface A the Mediator movable jaw (light blue) contacts the Pol II Rpb3-Rpb11 heterodimer (red/yellow), dock domain (beige) and the TFIIB B-ribbon (green). In interface B the Mediator spine domain (green) contacts helix H* of the Pol II stalk subunit Rpb4 (blue) with its Med22 helix H1, and the Mediator arm domain (violet) contacts Rpb4 with its Med8 helices H1 and H2. In interface C the Mediator plank domain (pink) contacts the Pol II foot region (cyan) with its Med9 helix H2. Two newly observed EDC-crosslinks between Med9 helix H2 and the Pol II foot domain are indicated by black spheres.

e. Mediator head-middle module interfaces. In the unbound *Sp* cMed X-ray structure, four interfaces (I-IV) were observed between the head and middle modules5. Due to stretching of the beam, interfaces I and II are altered in the PIC-bound cMed structure. In the new conformation the Med4 C-terminal region in the Mediator knob is flexible and does not contact the spine region. Interface IV between the shoulder and hook domains is lost. Mediator domains are colored as in panel **(a)**.

Extended Data Table 1
Components of the PIC-cMed complex.

Names of the human counterparts of the yeast subunits are provided. For details about complex assembly and composition also refer to main text and Methods.

Component	Subunit (yeast)	Corresponding human subunit	Construct residues (aa) / scaffold length (nt)	Mass (kDa)	Molarity
Pol II	Rpb1	RPB1	1-1733	191.6	1
	Rpb2	RPB2	1-1224	138.7	1
	Rpb3‡	RPB3	1-318	35.3	1
	Rpb4	RPB4	1-215	25.4	1
	Rpb5	RPB5	1-155	25.1	1
	Rpb6	RPB6	1-171	17.9	1
	Rpb7	RPB7	1-146	19.1	1
	Rpb8	RPB8	1-122	16.5	1
	Rpb9	RPB9	1-70	14.3	1
	Rpb10	RPB10	1-120	8.3	1
	Rpb11	RPB11	1-70	13.6	1
	Rpb12	RPB12		7.7	1
TFIIF	Tfg1‡	Rap74	1-735	82.2	5
	Tfg2	Rap30	1-400	46.6	5
Nucleic acid strands	Template	-	106	32.6	1.5
	Non-template	-	106	32.6	1.5
TFIIA	Toa1	TFIIAα	1-94, 210-286	19.4	10
	Toa2‡	TFIIAβ	1-122	13.4	10
TBP	TBP‡	TBP	1-240	27.0	5

Component	Subunit (yeast)	Corresponding human subunit	Construct residues (aa) / scaffold length (nt)	Mass (kDa)	Molarity
TFIIB	TFIIB‡	TFIIB	1-345	38.2	5
TFIIE	Tfa1‡	TFIIEα	1-482	54.7	2.5
	Tfa2	TFIIEβ	1-328	37.0	2.5
core-TFIIH	Rad3	XPD	1-778	89.8	2.5
	Ssl1	p44	1-461	52.3	2.5
	Ssl2	XPB	1-843	95.3	2.5
	Tfb1	p62	1-642	72.9	2.5
	Tfb2	p52	1-513	58.5	2.5
	Tfb4	p34	1-338	37.5	2.5
	Tfb5	p8/TTD4	1-72	8.2	2.5
TFIIH kinase module	Ccl1	Cyclin-H	1-393	45.2	2.5
	Kin28	Cdk7	1-306	35.2	2.5
	Tfb3	MAT1	1-321	38.1	2.5
cMed	Med1‡	Med1	1-566	64.2	1.5
	Med4	Med4	1-284	32.2	1.5
	Med6	Med6	1-295	32.8	1.5
	Med7	Med7	1-222	25.6	1.5
	Med8	Med8	1-223	25.3	1.5
	Med9	Med9	1-149	17.4	1.5
	Med10	Med10	1-157	17.9	1.5
	Med11	Med11	1-115	13.3	1.5
	Med14‡	Med14	1-745	84.6	1.5
	Med17	Med17	1-687	78.5	1.5
	Med18	Med18	1-307	34.3	1.5
	Med19	Med19	1-220	24.9	1.5
	Med20	Med20	1-210	22.9	1.5
	Med21	Med21	1-140	16.1	1.5
	Med22	Med22	1-121	13.8	1.5
	Med31	Med31	1-127	14.7	1.5
Final	**PIC-cMed**	**46 subunits**	**16,622 aa**	**1948.7**	-

aa: amino acids, nt: nucleotides, kDa: kilodalton

‡constructs contain N- or C-terminal 6xHis or 10xHis tags as described

Extended Data Table 2

Cryo-EM data collection and model statistics for the PIC and the PIC-cMed complex structures.

For details about EM-data collection, data processing and model building refer to main text and Methods.

	PIC (EMD-3846) (PDB 5OQJ)	PIC-cMed (EMD-3850) (PDB 5OQM)
Data collection		
Magnification	105 000	105 000
Particles	58 000	16 000
Pixel size (Å)	1.37	1.37
Defocus range (μm)	-0.5 to -5.0	-0.5 to -5.0
Voltage (kV)	300	300
Electron exposure ($e^-/Å^2$)	42	42
Symmetry imposed	C1	C1
Reconstruction		
Map resolution (Å)	4.70	5.84

	PIC (EMD-3846) (PDB 5OQJ)	PIC-cMed (EMD-3850) (PDB 5OQM)
Map sharpening B-factor (Å^2)	-201	-334
FSC threshold	0.143	0.143
Model composition		
Non-hydrogen atoms	62,931	79,757
Protein residues	8,188	10,825
DNA bases	148	148
Ligand atoms	25	25
Model validation		
MolProbity Score	1.92	1.89
Clashscore	7.83	7.53
Rotamer outliers (%)	0.30	0.37
Cβ-deviations	0	0
Ramachandran Plot		
Favored (%)	91.95	92.21
Allowed (%)	6.66	6.45
Disallowed (%)	1.39	1.34

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Simon Neyer, Carrie Bernecky, Carina Burzinski, Seychelle Vos, Lucas Farnung and other members of the Cramer laboratory for help. We thank Chung-Tien Lee and Iwan Parfentev from the Urlaub group for mass spectrometry. PC was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP1935), the Advanced Grant TRANSREGULON (grant agreement No 693023) of the European Research Council, and the Volkswagen Foundation.

References

1. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci. 1996; 21:327–335. [PubMed: 8870495]
2. Kornberg RD. Mediator and the mechanism of transcriptional activation. Trends Biochem Sci. 2005; 30:235–239. DOI: 10.1016/j.tibs.2005.03.011 [PubMed: 15896740]
3. Takagi Y, Kornberg RD. Mediator as a general transcription factor. J Biol Chem. 2006; 281:80–89. DOI: 10.1074/jbc.M508253200 [PubMed: 16263706]
4. Plaschka C, et al. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature. 2015; 518:376–380. DOI: 10.1038/nature14229 [PubMed: 25652824]
5. Nozawa K, Schneider TR, Cramer P. Core Mediator structure at 3.4 A extends model of transcription initiation complex. Nature. 2017; 545:248–251. DOI: 10.1038/nature22328 [PubMed: 28467824]
6. Plaschka C, et al. Transcription initiation complex structures elucidate DNA opening. Nature. 2016; 533:353–358. DOI: 10.1038/nature17990 [PubMed: 27193681]
7. He Y, et al. Near-atomic resolution visualization of human transcription promoter opening. Nature. 2016; 533:359–365. DOI: 10.1038/nature17970 [PubMed: 27193682]

8. He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. Nature. 2013; 495:481–486. DOI: 10.1038/nature11991 [PubMed: 23446344]
9. Murakami K, et al. Structure of an RNA polymerase II preinitiation complex. Proc Natl Acad Sci U S A. 2015; 112:13543–13548. DOI: 10.1073/pnas.1518255112 [PubMed: 26483468]
10. Robinson PJ, et al. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell. 2016; 166:1411–1422 e1416. DOI: 10.1016/j.cell.2016.08.050 [PubMed: 27610567]
11. Gibbons BJ, et al. Subunit architecture of general transcription factor TFIIH. Proc Natl Acad Sci U S A. 2012; 109:1949–1954. DOI: 10.1073/pnas.1105266109 [PubMed: 22308316]
12. Schultz P, et al. Molecular structure of human TFIIH. Cell. 2000; 102:599–607. [PubMed: 11007478]
13. Chang WH, Kornberg RD. Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. Cell. 2000; 102:609–613. [PubMed: 11007479]
14. Compe E, Egly JM. TFIIH: when transcription met DNA repair. Nat Rev Mol Cell Biol. 2012; 13:343–354. DOI: 10.1038/nrm3350 [PubMed: 22572993]
15. Svejstrup JQ, et al. Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell. 1995; 80:21–28. [PubMed: 7813015]
16. Guzder SN, Sung P, Bailly V, Prakash L, Prakash S. RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. Nature. 1994; 369:578–581. DOI: 10.1038/369578a0 [PubMed: 8202161]
17. Goodrich JA, Tjian R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell. 1994; 77:145–156. [PubMed: 8156590]
18. Moreland RJ, et al. A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. J Biol Chem. 1999; 274:22127–22130. [PubMed: 10428772]
19. Alekseev S, et al. Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Mol Cell. 2017; 65:504–514 e504. DOI: 10.1016/j.molcel.2017.01.012 [PubMed: 28157507]
20. Feaver WJ, Svejstrup JQ, Henry NL, Kornberg RD. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell. 1994; 79:1103–1109. [PubMed: 8001136]
21. Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell. 1994; 77:599–608. [PubMed: 8187178]
22. Wong KH, Jin Y, Struhl K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell. 2014; 54:601–612. DOI: 10.1016/j.molcel.2014.03.024 [PubMed: 24746699]
23. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012; 180:519–530. DOI: 10.1016/j.jsb.2012.09.006 [PubMed: 23000701]
24. Luo J, et al. Architecture of the human and yeast general transcription and DNA repair factor TFIIH. Molecular cell. 2015; 59:794–806. [PubMed: 26340423]
25. Murakami K, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. Science. 2013; 342:1238724. [PubMed: 24072820]
26. Fishburn J, Tomko E, Galburt E, Hahn S. Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc Natl Acad Sci U S A. 2015; 112:3961–3966. DOI: 10.1073/pnas.1417709112 [PubMed: 25775526]
27. Tirode F, Busso D, Coin F, Egly JM. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. Mol Cell. 1999; 3:87–95. [PubMed: 10024882]
28. Kainov DE, Vitorino M, Cavarelli J, Poterszman A, Egly J-M. Structural basis for group A trichothiodystrophy. Nature structural & molecular biology. 2008; 15:980–984.
29. Wolski SC, et al. Crystal structure of the FeS cluster–containing nucleotide excision repair helicase XPD. PLoS biology. 2008; 6:e149. [PubMed: 18578568]

30. Constantinescu-Aruxandei D, Petrovic-Stojanovska B, Penedo JC, White MF, Naismith JH. Mechanism of DNA loading by the DNA repair helicase XPD. Nucleic acids research. 2016; 44:2806–2815. [PubMed: 26896802]
31. Kuper J, Wolski SC, Michels G, Kisker C. Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. The EMBO journal. 2012; 31:494–502. [PubMed: 22081108]
32. Doerks T, Huber S, Buchner E, Bork P. BSD: a novel domain in transcription factors and synapse-associated proteins. Trends in biochemical sciences. 2002; 27:168–170. [PubMed: 11943536]
33. Warfield L, Luo J, Ranish J, Hahn S. Function of Conserved Topological Regions within the Saccharomyces cerevisiae Basal Transcription Factor TFIIH. Molecular and cellular biology. 2016; 36:2464–2475. [PubMed: 27381459]
34. Stefanini M, Botta E, Lanzafame M, Orioli D. Trichothiodystrophy: from basic mechanisms to clinical implications. DNA repair. 2010; 9:2–10. [PubMed: 19931493]
35. Oh KS, et al. Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. Human mutation. 2006; 27:1092–1103. [PubMed: 16947863]
36. Rossignol M, Kolb-Cheynel I, Egly JM. Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. EMBO J. 1997; 16:1628–1637. DOI: 10.1093/emboj/16.7.1628 [PubMed: 9130708]
37. Edwards AM, Kane C, Young R, Kornberg RD. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. Journal of Biological Chemistry. 1991; 266:71–75. [PubMed: 1985924]
38. Serizawa H, Conaway JW, Conaway RC. Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. Nature. 1993; 363:371. [PubMed: 8497323]
39. Maxon ME, Goodrich JA, Tjian R. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev. 1994; 8:515–524. [PubMed: 7926747]
40. Kim TK, Ebright RH, Reinberg D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science. 2000; 288:1418–1422. [PubMed: 10827951]
41. Grünberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nature structural & molecular biology. 2012; 19:788–796.
42. Farnung L, Vos SM, Wigge C, Cramer P. Structure of nucleosome-Chd1 complex and implications for chromatin remodeling. Nature. (in revision).
43. Schaeffer L, et al. The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. The EMBO journal. 1994; 13:2388. [PubMed: 8194528]
44. Lin YC, Choi WS, Gralla JD. TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nature structural & molecular biology. 2005; 12:603–607.
45. Hwang JR, et al. A 3'→ 5' XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription. Journal of Biological Chemistry. 1996; 271:15898–15904. [PubMed: 8663148]
46. Ohkuma Y, Roeder RG. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. Nature. 1994; 368:160. [PubMed: 8166891]
47. Wigley DB, Bowman GD. A glimpse into chromatin remodeling. Nature Structural & Molecular Biology. 2017; 24:498–500.
48. Larivière L, et al. Structure of the Mediator head module. Nature. 2012; 492:448–451. [PubMed: 23123849]
49. Tsai KL, et al. Mediator structure and rearrangements required for holoenzyme formation. Nature. 2017; 544:196–201. DOI: 10.1038/nature21393 [PubMed: 28241144]
50. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. Science. 2001; 292:1863–1876. [PubMed: 11313498]
51. Kostrewa D, et al. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature. 2009; 462:323–330. DOI: 10.1038/nature08548 [PubMed: 19820686]

52. Louder RK, et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature. 2016; 531:604–609. [PubMed: 27007846]
53. Zheng SQ, et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature Methods. 2017
54. Zhang K. Gctf: Real-time CTF determination and correction. Journal of Structural Biology. 2016; 193:1–12. [PubMed: 26592709]
55. Kimanius D, Forsberg BO, Scheres SH, Lindahl E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife. 2016; 5:e18722. [PubMed: 27845625]
56. Chen S, et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy. 2013; 135:24–35. [PubMed: 23872039]
57. Tang G, et al. EMAN2: an extensible image processing suite for electron microscopy. Journal of structural biology. 2007; 157:38–46. [PubMed: 16859925]
58. Suhre K, Sanejouand Y-H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic acids research. 2004; 32:W610–W614. [PubMed: 15215461]
59. Matthies H, Strang G. The solution of nonlinear finite element equations. International journal for numerical methods in engineering. 1979; 14:1613–1626.
60. Pettersen EF, et al. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004; 25:1605–1612. [PubMed: 15264254]
61. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010; 5:725–738. [PubMed: 20360767]
62. Yang J, et al. The I-TASSER Suite: protein structure and function prediction. Nature methods. 2015; 12:7–8. [PubMed: 25549265]
63. Biasini M, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research. 2014; 42:W252–W258. [PubMed: 24782522]
64. Bordoli L, et al. Protein structure homology modeling using SWISS-MODEL workspace. Nature protocols. 2009; 4:1–13. [PubMed: 19131951]
65. Song Y, et al. High-resolution comparative modeling with RosettaCM. Structure. 2013; 21:1735–1742. [PubMed: 24035711]
66. Raman S, et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins: Structure, Function, and Bioinformatics. 2009; 77:89–99.
67. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography. 2010; 66:486–501. [PubMed: 20383002]
68. Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature. 2013; 493:437–440. DOI: 10.1038/nature11715 [PubMed: 23151482]
69. Miwa K, et al. Crystal Structure of Human General Transcription Factor TFIIE at Atomic Resolution. J Mol Biol. 2016; 428:4258–4266. DOI: 10.1016/j.jmb.2016.09.008 [PubMed: 27639436]
70. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography. 2010; 66:213–221. [PubMed: 20124702]
71. Ye J, Kandegedara A, Martin P, Rosen BP. Crystal structure of the Staphylococcus aureus pI258 CadC Cd (II)/Pb (II)/Zn (II)-responsive repressor. Journal of bacteriology. 2005; 187:4214–4221. [PubMed: 15937183]
72. Fadden AJ, et al. A winged helix domain in human MUS81 binds DNA and modulates the endonuclease activity of MUS81 complexes. Nucleic acids research. 2013; 41:9741–9752. [PubMed: 23982516]
73. Ahmad MUD, et al. Structural insights into nonspecific binding of DNA by TrmBL2, an archaeal chromatin protein. Journal of molecular biology. 2015; 427:3216–3229. [PubMed: 26299937]

74. Gervais V, et al. Solution Structure of the N-terminal Domain of the Human TFIIH MAT1 Subunit NEW INSIGHTS INTO THE RING FINGER FAMILY. Journal of Biological Chemistry. 2001; 276:7457–7464. [PubMed: 11056162]
75. Schmitt DR, Kuper J, Elias A, Kisker C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. PloS one. 2014; 9:e102389. [PubMed: 25013903]
76. Tempel W, et al. Structural genomics of Pyrococcus furiosus: X-ray crystallography reveals 3D domain swapping in rubrerythrin. PROTEINS: Structure, Function, and Bioinformatics. 2004; 57:878–882.
77. Kellenberger E, et al. Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. Journal of Biological Chemistry. 2005; 280:20785–20792. [PubMed: 15790571]
78. Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins: Structure, Function, and Bioinformatics. 2012; 80:1715–1735.
79. Di Lello P, et al. NMR structure of the amino-terminal domain from the Tfb1 subunit of TFIIH and characterization of its phosphoinositide and VP16 binding sites. Biochemistry. 2005; 44:7678–7686. [PubMed: 15909982]
80. Kim JS, et al. Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. Journal of Biological Chemistry. 2015; 290:8321–8330. [PubMed: 25681444]
81. Fan L, et al. Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Molecular cell. 2006; 22:27–37. [PubMed: 16600867]
82. Hilario E, Li Y, Nobumori Y, Liu X, Fan L. Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. Acta Crystallographica Section D: Biological Crystallography. 2013; 69:237–246. [PubMed: 23385459]
83. Baker ML, et al. Modeling protein structure at near atomic resolutions with Gorgon. Journal of structural biology. 2011; 174:360–373. [PubMed: 21296162]
84. Baker ML, Baker MR, Hryc CF, Ju T, Chiu W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. Biopolymers. 2012; 97:655–668. [PubMed: 22696403]
85. Okuda M, et al. Structural insight into the TFIIE–TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. The EMBO journal. 2008; 27:1161–1171. [PubMed: 18354501]
86. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996; 14:33–38. [PubMed: 8744570]
87. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure. 2008; 16:673–683. [PubMed: 18462672]
88. Jakhanwal S, Lee CT, Urlaub H, Jahn R. An activated Q-SNARE/SM protein complex as a possible intermediate in SNARE assembly. The EMBO Journal. 2017; 36:1788–1802. [PubMed: 28483813]
89. Yang B, et al. Identification of cross-linked peptides from complex samples. Nature methods. 2012; 9:904–906. [PubMed: 22772728]
90. Combe CW, Fischer L, Rappsilber J. xiNET: cross-link network maps with residue resolution. Molecular & Cellular Proteomics. 2015; 14:1137–1147. [PubMed: 25648531]
91. Kosinski J, et al. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. Journal of structural biology. 2015; 189:177–183. [PubMed: 25661704]
92. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic acids research. 2010; 38:W529–W533. [PubMed: 20478830]
93. Fairman-Williams ME, Guenther U-P, Jankowsky E. SF1 and SF2 helicases: family matters. Current opinion in structural biology. 2010; 20:313–324. [PubMed: 20456941]

94. Gu M, Rice CM. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. Proceedings of the National Academy of Sciences. 2010; 107:521–528.

Europe PMC Funders Author Manuscripts

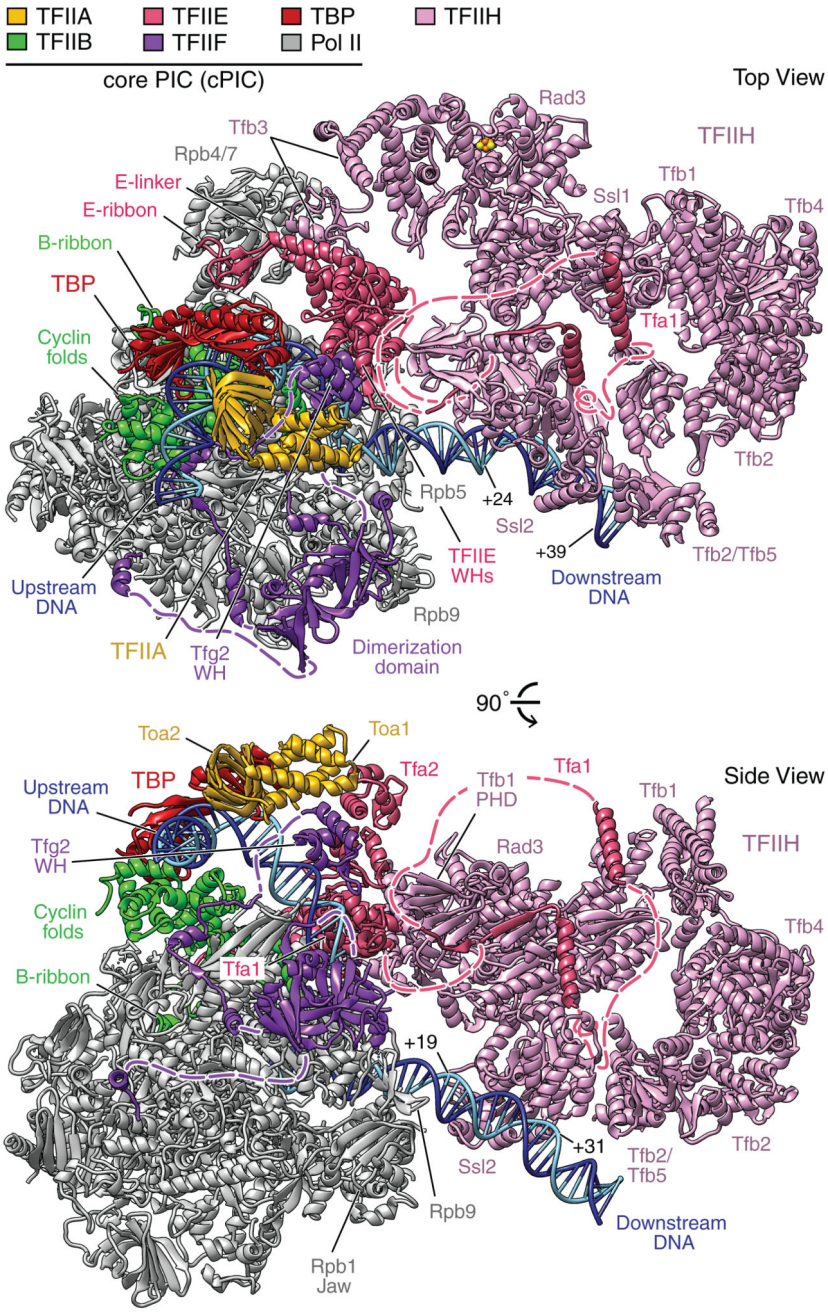


Figure 1. Structure of Pol II pre-initiation complex (PIC).

Two views50 of the yeast PIC cryo-EM structure. The DNA template and non-template strands are in dark and light blue, respectively. Positions of TFIIH subunits are indicated. Dashes lines represent flexible linkers in TFIIE and TFIIF. The color code is used throughout.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

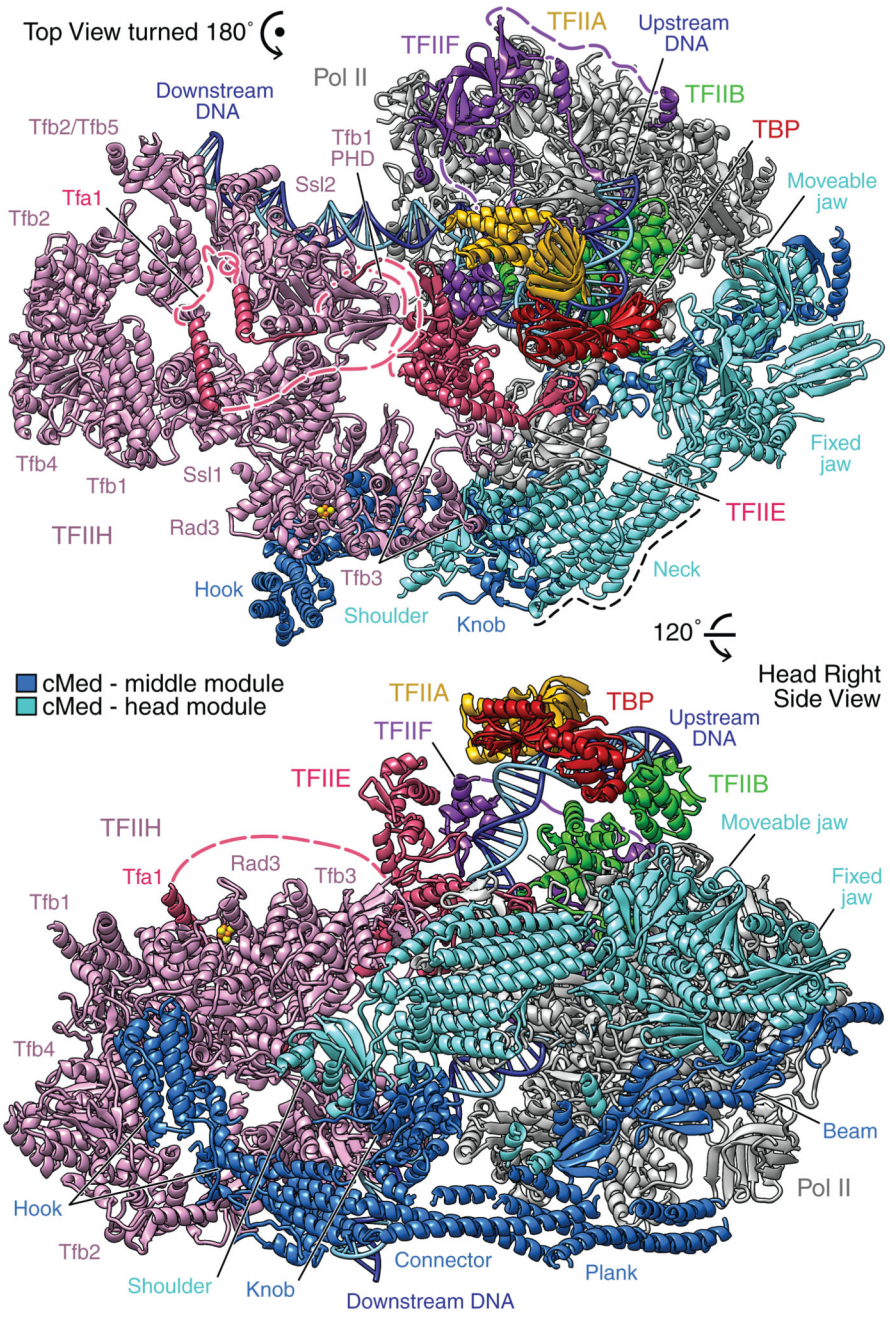


Figure 2. Structure of PIC-core Mediator (PIC-cMed) complex.

Two views of the PIC-cMed cryo-EM structure. The first view is rotated by 180º compared to the top view in Fig. 1. The second view is obtained by a 120º rotation around a horizontal axis. Mediator submodules in the head (blue) and middle modules (cyan) are indicated.

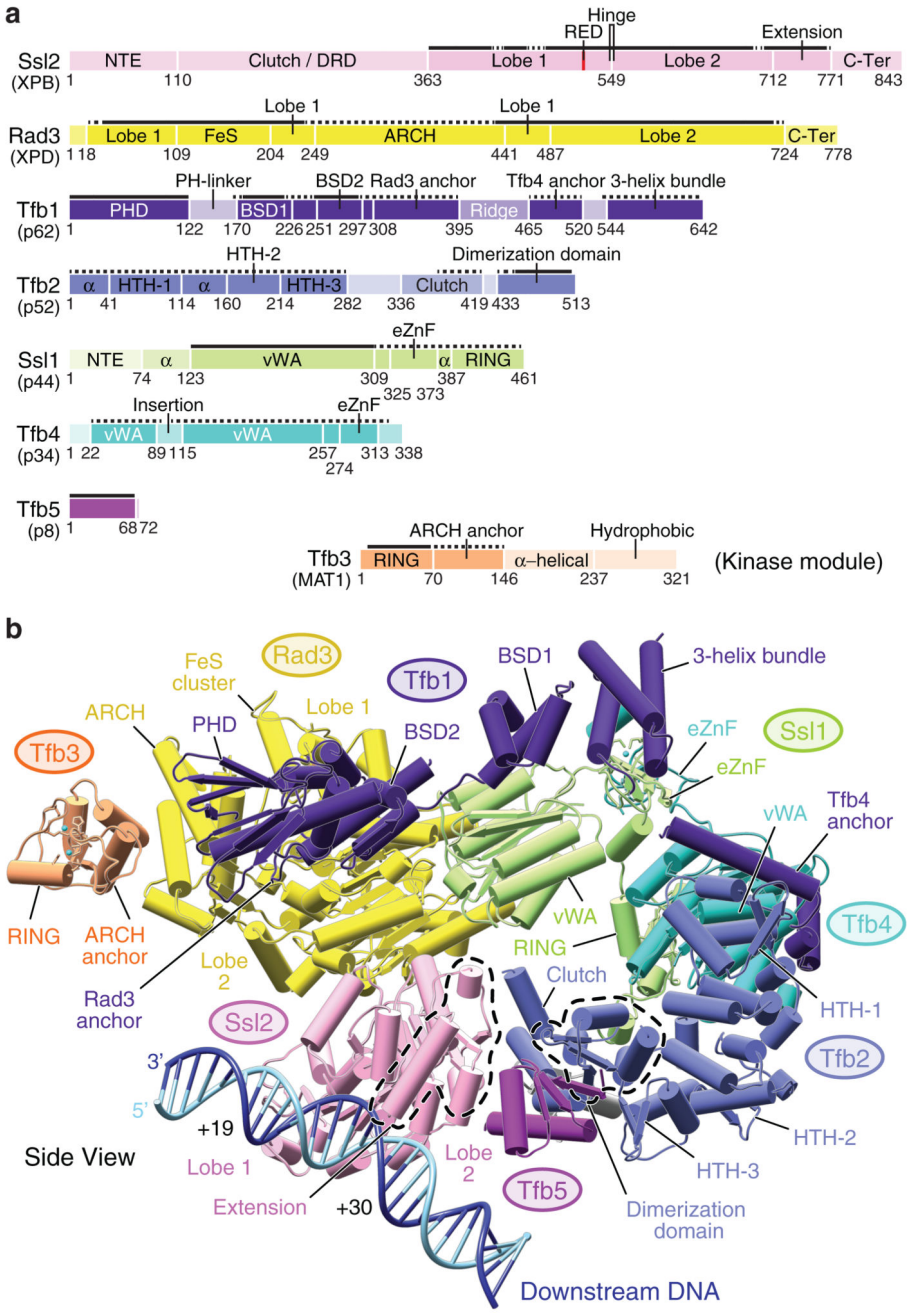


Figure 3. Structure of TFIIH.

a. Domain organization of yeast TFIIH subunits except Kin28 (CDK7) and Ccl1 (CyclinH). Names of corresponding human subunits are in parenthesis. Residue numbers are given for domain borders. Color saturation scales with the percentage of residues modeled as atomic or backbone structures (solid and dashed black bars, respectively).

b. TFIIH structure in cylindrical representation viewed from the side (Fig. 1). The DNA register with respect to the putative transcription start site +1 is indicated.

Europe PMC Funders Author Manuscripts

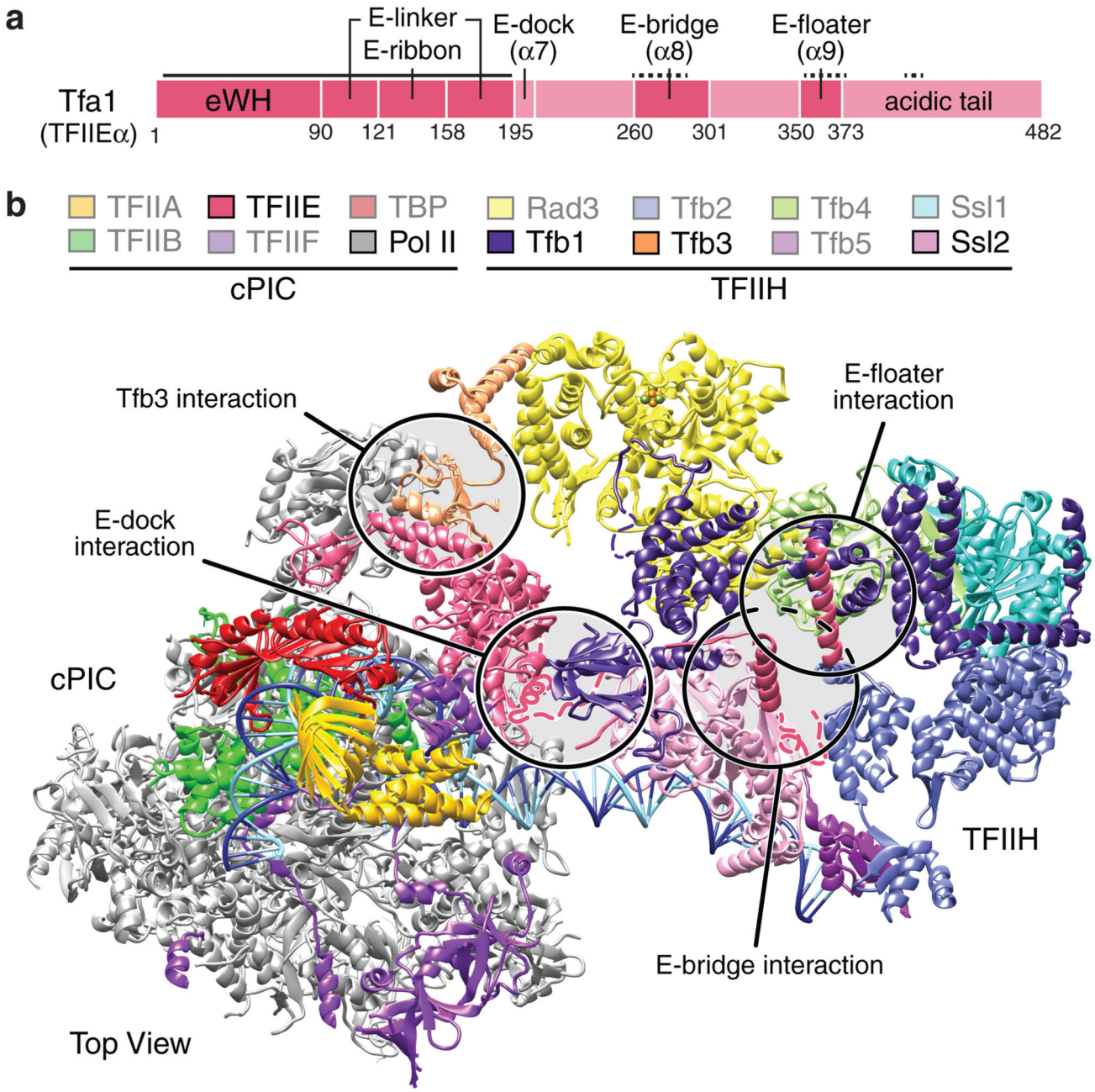


Figure 4. Interactions of TFIIH with core PIC (cPIC).

a.Domain organization of TFIIE subunit Tfa1 (human TFIIEα) including the previously unassigned helices α7 (E-dock), α8 (E-bridge) and α9 (E-floater). Solid and dashed black bars refer to protein residues modeled as atomic or backbone structures, respectively.

b. TFIIH-cPIC interactions. PIC is viewed from the top (Fig. 1). Regions involved in the formation of the four interfaces are encircled. The color code of cPIC and TFIIH subunits highlights components participating in the interaction.

Europe PMC Funders Author Manuscripts

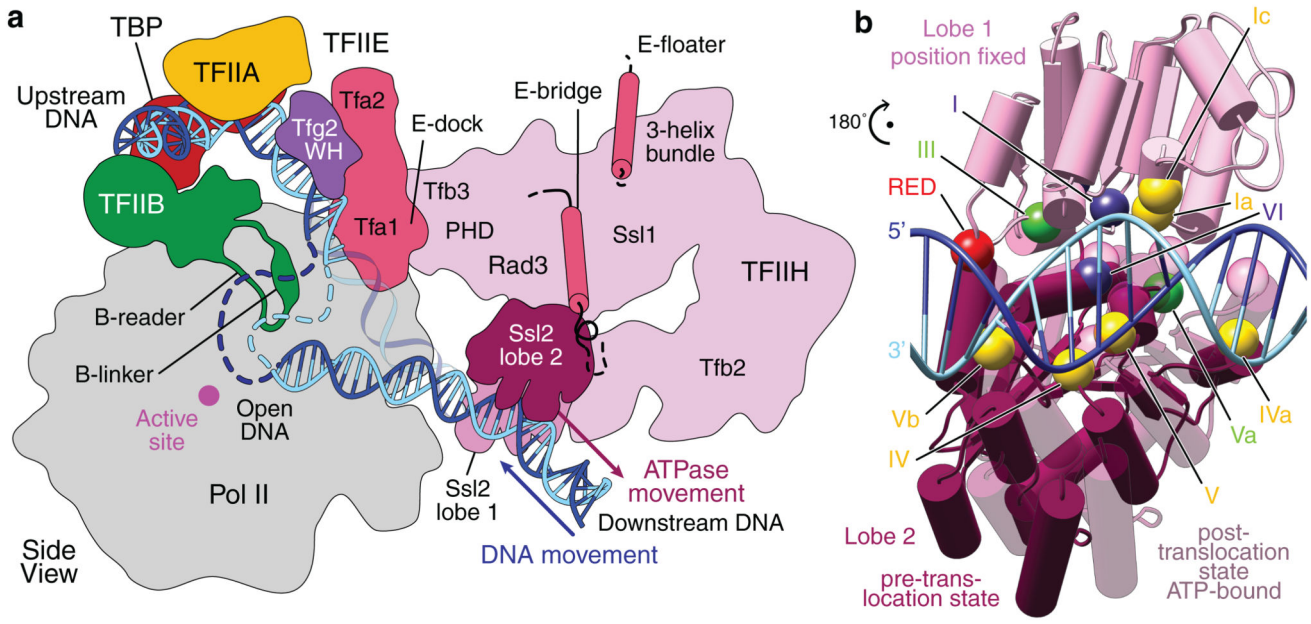


Figure 5. TFIIH and DNA opening.

a. Schematic cross-section of the PIC with open and closed DNA viewed from the side. PIC elements involved in DNA opening are depicted. Color coding as in Fig. 1 except for Ssl2 (human XPB) lobe 1 (pink) and lobe 2 (bordeaux). The Ssl2 ATPase translocates to the right and DNA moves to the left during DNA opening.

b. Putative ratcheting of lobe 2 in the Ssl2 ATPase with respect to lobe 1. The PIC structure reveals the pre-translocation conformation (no ATP bound). The post-translocation conformation of lobe 2 was modeled by superposition of Chd1 (PDB 5O9G). Helicase motifs are indicated (Extended Data Fig. 7).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

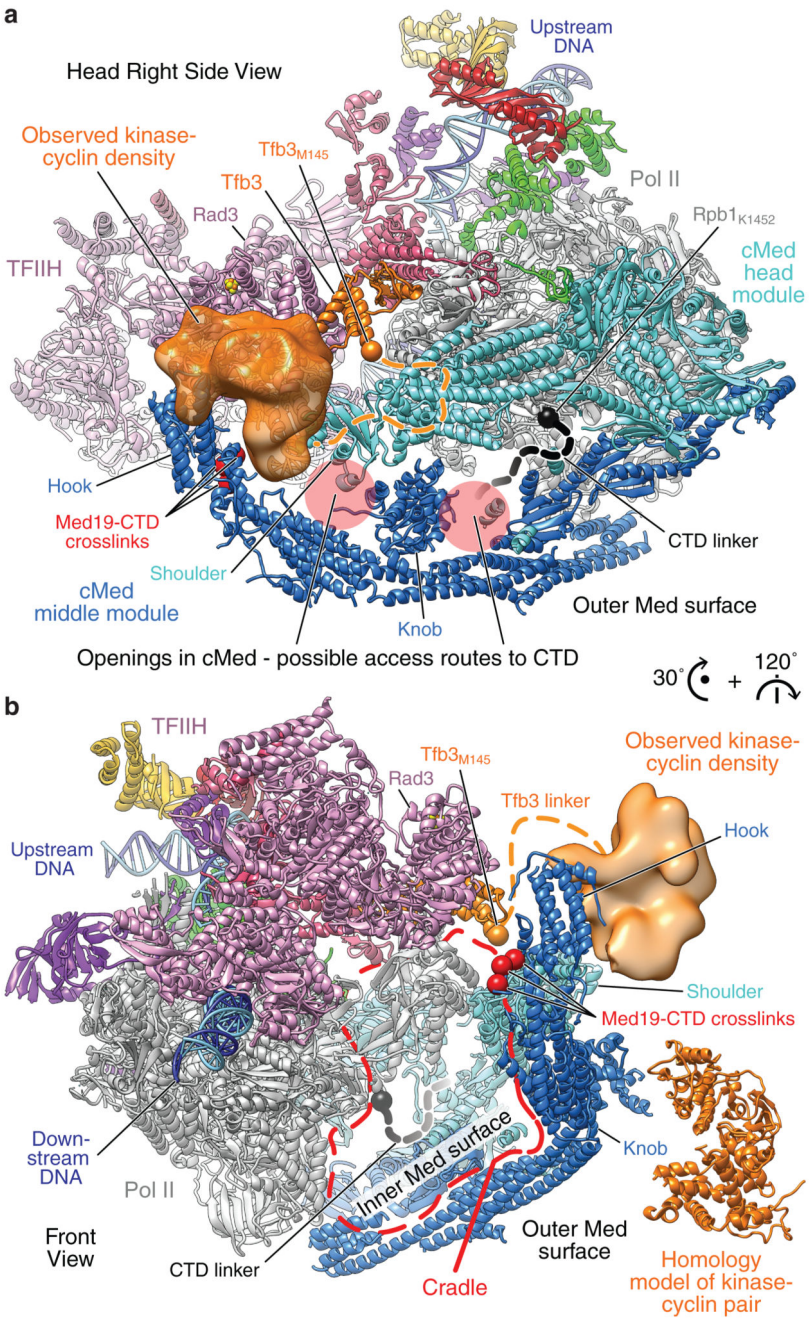


Figure 6. TFIIH and Pol II phosphorylation.

a. PIC-cMed structure as in Fig. 2 but with additional cryo-EM density for the mobile TFIIH Kin28-Ccl1 (human CDK7-CycH) kinase-cyclin pair (orange, filtered to 15 Å). An orange sphere depicts the last modeled residue in the Tfb3 linker to the kinase-cyclin pair. A black sphere depicts the last ordered residue in the Rpb1 linker to the CTD. Red spheres depict Med19 residues that crosslink to the CTD C-terminal end. Two openings at the Mediator head-middle interface are indicated with filled red circles.

b. The same structure viewed from the front into the cradle between Pol II and Mediator (red outline). A model for the kinase-cyclin pair is shown for size comparison in an arbitrary position.

Europe PMC Funders Author Manuscripts