Cryo-EM structure of mammalian RNA polymerase II in complex with human RPAP2

Isaac Fianu1, Christian Dienemann1, Shintaro Aibara1, Sandra Schilbach1 & Patrick Cramer1

Nuclear import of RNA polymerase II (Pol II) involves the conserved factor RPAP2. Here we report the cryo-electron microscopy (cryo-EM) structure of mammalian Pol II in complex with human RPAP2 at 2.8 Å resolution. The structure shows that RPAP2 binds between the jaw domains of the polymerase subunits RPB1 and RPB5. RPAP2 is incompatible with binding of downstream DNA during transcription and is displaced upon formation of a transcription pre-initiation complex.
RNA polymerase II (Pol II) is the enzyme that transcribes protein-coding genes in the nucleus to produce mRNA. A large amount of data has accumulated on Pol II structures that are relevant for understanding transcription. However, the molecular basis of Pol II biogenesis and nuclear import remains poorly understood. Currently, only one structure of Pol II in complex with a biogenesis factor is available, a low-resolution cryo-EM structure of yeast Pol II bound by the nuclear import factor Iwr1.

The RNA polymerase II associated protein 2 (RPAP2) was identified by co-purification with Pol II and its associated factors RPRD1A, RPRD1B, RPRD2, GRINL1A, and RECQL5. RPAP2 shuttles between the cytoplasm and the nucleus and its silencing causes cytoplasmic accumulation of Pol II. The RPAP2 homolog in Arabidopsis thaliana is also involved in nuclear import of Pol II and its knockdown delays cell differentiation. The yeast homolog of RPAP2 is called regulator of transcription 1 (Rtr1) and also shuttles between the cytoplasm and nucleus. Rtr1 accumulates in the nucleus when the Pol II biogenesis factor Gpr1 is mutated and the structure of Rtr1 is known.

RPAP2 and Rtr1 were also described as phosphatases that target the Pol II C-terminal repeat domain (CTD). For transcription of snRNA genes, RPAP2 may be recruited by serine-7 phosphorylation of the CTD, and dephosphorylates serine-5. However, Rtr1 from the yeast Saccharomyces cerevisiae phosphorylates both tyrosine-1 and serine-5 residues. Rtr1 from the yeast S. cerevisiae was reported to have CTD phosphatase activity and can dephosphorylate both tyrosine-1 and serine-5 residues in the CTD. The structure of S. cerevisiae Rtr1 was reported to contain a phosphoryl transfer domain and an active site for an atypical phosphatase. However, Rtr1 from the yeast Kluyveromyces lactis apparently lacks an active site and phosphatase activity. The molecular basis of how RPAP2 (and its homologs) binds to Pol II to perform these functions is not known.

Here we report the cryo-EM structure of mammalian Pol II bound by human RPAP2. The structure shows that the highly conserved N-terminal domain of RPAP2 binds between the jaw domains of Pol II subunits RPB1 and RPB5 in the downstream region of the Pol II active center cleft. RPAP2 location on Pol II is incompatible with binding of the downstream DNA during transcription and it is displaced upon formation of a transcription pre-initiation complex (PIC).

Results

Cryo-EM analysis of Pol II–RPAP2 complex. To investigate how RPAP2 interacts with Pol II, we aimed at localizing this factor on the Pol II surface. We prepared recombinant human RPAP2 after heterologous expression in insect cells and found that it readily bound to Sus scrofa domesticus Pol II that we purified from thymus samples as described. A stable Pol II–RPAP2 complex could be purified by size-exclusion chromatography (Supplementary Fig. 1a). We prepared cryo-EM grids from purified Pol II–RPAP2 complex and obtained a single-particle cryo-EM reconstruction at a resolution of 2.8 Å from 365,000 particles (Fig. 1, Supplementary Fig. 1).

To build a model for the Pol II–RPAP2 complex, we first placed the structure of Pol II (PDB:5FLM) into the density after nucleic acids in the model were removed. We then deleted the Pol II clamp and the RPB4/RPB7 stalk domains because density for these regions was absent or weak, respectively. After minor adjustments to the Pol II model, we placed a homology model for RPAP2 derived from the structure of yeast Rtr1 into the remaining density. We manually adjusted the homology model, guided by detailed density that revealed side chains (Supplementary Fig. 2). We observed unambiguous density for RPAP2 residues 48–155 and confidently modeled this region including side chains. We extended the RPAP2 model by a few residues C-terminally using continuous density, but did not observe any density for the large C-terminal region of RPAP2 (Fig. 1a). The structure was real-space refined and has good stereochemistry (Table 1).

Structure of Pol II–RPAP2 complex. The structure shows that RPAP2 occupies a large part of the downstream region of the active center cleft of Pol II (Fig. 1b). The conserved N-terminal region of RPAP2 binds between the jaw domains of the Pol II subunits RPB1 and RPB5. This is consistent with the previous finding that it is the N-terminal region of RPAP2 that mediates Pol II binding. This N-terminal region of RPAP2 consists of four helices that are conserved in yeast Rtr1. In addition, the N-terminal region of RPAP2 contains an ‘insertion’ (residues 116–128) that forms two β-strands and thus deviates from the corresponding region in S. cerevisiae Rtr1, which forms an α-helix (Fig. 1b and Supplementary Fig. 3a). Residues in the RPAP2 insertion are conserved from Drosophila to human (Supplementary Fig. 3b), but not in yeast, suggesting it is a metazoan-specific feature of RPAP2. The region of RPAP2 following C-terminally of the conserved N-terminal domain extends to the opposite side of the cleft to the RPB1 jaw-lobe module (Fig. 1b).

Our structure also reveals details of the interaction of RPAP2 with Pol II. The RPAP2 insertion contacts helix α43 in the RPB1

| Table 1 Cryo-EM data collection, refinement and validation statistics. |
|--------------------------|--------------------------|
| Data collection and processing |                          |
| Magnification            | 81,000                   |
| Voltage (kV)             | 300                      |
| Electron exposure (e-/Å²) | 42.5                     |
| Defocus range (µm)       | −0.5 to −3.2             |
| Pixel size (Å)           | 1.05                     |
| Symmetry imposed         | CI                       |
| Initial particle images (no.) | 1,556,776               |
| Final particle images (no.) | 364,771                 |
| Map resolution (Å)       | 2.8                      |
| FSC threshold            | 0.143                    |
| Map resolution range (Å) | 2.58–6.00                |
| Map sharpening B factor (Å²) | −55                    |

Refinement

Initial model (PDB) 5FLM
Model resolution (Å) 3.1
Model composition
Non-hydrogen atoms 25,142
Protein residues 3,112
Ligands 6
B factors (Å²)
Protein 98.13
Ligand 124.63
r.m.s. deviations
Bond lengths (Å) 0.002
Bond angles (°) 0.481
Validation
MolProbity score 1.84
Clashscore 8.72
Poor rotamers (%) 0.07
Ramachandran plot
Favored (%) 94.51
Allowed (%) 5.49
Disallowed (%) 0.0
This variant readily bound Pol II (Fig. 2c, Peak 4). Incubation of a region of RPAP2 observed in our structure, and found that during transcription elongation would displace RPAP2 from Pol II and formation of a core PIC (Fig. 2c, Peak 4). Incubation of a Pol II–RPAP2(1–215) complex with the DNA–RNA scaffold displaced RPAP2(1–215) from Pol II (Fig. 2c, Peak 1). These results indicate that RPAP2 binding to Pol II is incompatible with formation of a Pol II EC and transcription.

We next asked whether RPAP2 may already be displaced from Pol II upon formation of a transcription PIC. Consistent with this idea, superposition of our Pol II–RPAP2 structure onto the structure of the Pol II core PIC showed strong clashes of RPAP2 with downstream promoter DNA (Fig. 2b). To test whether core PIC formation displaces RPAP2 from Pol II, we incubated the Pol II–RPAP2(1–215) complex with the transcription initiation factor (TF)-IIF and preformed upstream promoter complex containing promoter DNA, TFIIA, TFIIB, and the TATA box-binding protein (TBP). This led to displacement of RPAP2(1–215) from Pol II and formation of a core PIC (Fig. 2c, Peak 2). In contrast, double-stranded promoter DNA alone could not displace RPAP2 (1–215) from Pol II (Fig. 2c, Peak 3). Together, these results show that PIC formation displaces RPAP2 from Pol II before transcription starts. The results are consistent with the published observation that RPAP2 does not copurify with the Mediator complex, which is known to bind the PIC.

RPAP2 is displaced by downstream DNA upon transcription. The location of RPAP2 between the jaws of the Pol II active center cleft suggested that RPAP2 interferes with binding of downstream DNA during transcription. Indeed, superposition of our Pol II–RPAP2 structure with the structure of the Pol II elongation complex (EC) revealed severe clashes between RPAP2 and downstream DNA (Fig. 2a). We therefore tested whether a DNA–RNA scaffold that mimics the nucleic acids during transcription elongation would displace RPAP2 from Pol II. We prepared a RPAP2 variant, RPAP2(1–215), which includes the region of RPAP2 observed in our structure, and found that this variant readily bound Pol II (Fig. 2c, Peak 4). Incubation of a Pol II–RPAP2(1–215) complex with the DNA–RNA scaffold displaced RPAP2(1–215) from Pol II (Fig. 2c, Peak 1). These results indicate that RPAP2 binding to Pol II is incompatible with formation of a Pol II EC and transcription.

Discussion
In conclusion, our work provides the structure of human RPAP2 and unambiguously localizes it on the Pol II surface between the jaw domains of RPB1 and RPB5. The RPAP2 location is incompatible with the presence of downstream DNA during transcription and is consistent with RPAP2 being a factor involved in Pol II biogenesis and nuclear import. In addition, our results suggest that formation of the PIC liberates RPAP2 for its recycling by nuclear export. In summary, our findings and published data converge on the following model. RPAP2 enters the nucleus in association with Pol II, and its conserved N-terminal domain is displaced upon assembly of the PIC for transcription initiation. After dissociation from Pol II, RPAP2 returns to the
cytoplasm in a manner that depends on the GTPase GPN1 and remains to be studied.

Our structure of the Pol II–RPAP2 complex and complementary biochemical data thus is consistent with RPAP2 functioning primarily as a Pol II biogenesis factor that is displaced when Pol II engaged with DNA for transcription. It is however not excluded that RPAP2 remains bound to a transcription EC under certain circumstances because the C-terminal region of RPAP2, which is not involved in the interactions observed here structurally, may make additional contacts with Pol II21. Furthermore, RPAP2 may indirectly interact with hyperphosphorylated Pol II via the proteins RPRD1A and RPRD1B, allowing it to dephosphorylate the Pol II CTD. Unfortunately, our structure does not provide new insights into the phosphorylation-related functions of RPAP2.

Methods

Molecular cloning and protein expression. A synthetic DNA fragment encoding the full-length Homo sapiens RPAP2 codon optimized for insect cell expression was purchased from Integrated DNA Technologies (IDT). The DNA fragment was inserted into insect cell expression vector 438-C (Addgene: Plasmid #55220) by ligation independent cloning (LIC)22. Similarly, a DNA fragment encoding RPAP2 (1–215) codon optimized for Escherichia coli expression was purchased from IDT and cloned into vector 1-B (Addgene: Plasmid #29653) by LIC. Both fragments were designed to contain LIC compatible overhangs for cloning into the stated vectors. Homo sapiens TFIIA-α variant (TFIIA-αΔ59-302) in which the unstructured insertion (residues 59–302) was replaced by a linker (1–58—LEVLFQGP—303–376–6xHis) was cloned in the first multiple cloning site (MCS1) of pET-Duet vector and full-length TFIIA-β-6xHis in MCS2. Both genes were codon-optimized for E. coli expression. All clones were verified by Sanger sequencing.

Full-length RPAP2 was expressed in insect cells. Production of bacmid, V0 and V1 baculoviruses, and protein expression were performed as previously described23. Hi5 cells were harvested by centrifugation, cells were suspended in a lysis buffer (35 mM Tris–HCl pH 8, 300 mM NaCl, 30 mM
imidazole pH 8, 10% glycerol, 10 mM beta-mercaptoethanol (BME), 10 µM ZnCl₂, 0.284 µg/mL leupeptin, 1.37 µg/mL pepstatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine, and stored at −80 °C until purification. Proteins were applied to a 5 mL HisTrap HP column (GE Healthcare) equilibrated in lysis buffer. The column was washed with 50 mL of wash buffer containing 250 mM imidazole with 500 µM IPTG for 18 h at 18 °C. Harvested cells were resuspended in buffer H, 30–250 mM HEPES pH 7.4, 100 mM KCl, 10% glycerol (v/v), 30 mM imidazole, 5 mM BME, 0.284 µg/mL leupeptin, 1.37 µg/mL peptatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine, and stored at −80 °C until purification.

The TFIIA variant was expressed in E. coli as described32 without addition of 0.1 mM ZnCl₂ in the expression medium. The TFIIA variant was expressed in BL21 E. coli by induction at OD600 of 50 with 0.02% of 3-[[3-chloro-4-fluorophenyl]methyl]-[1,2,4]triazolo[1,5-a][1,3,5]triazine at 18 °C. Harvested cells were resuspended in buffer H, 30–250 mM HEPES pH 7.4, 100 mM KCl, 10% glycerol (v/v), 30 mM imidazole, 5 mM BME, 0.284 µg/mL leupeptin, 1.37 µg/mL peptatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine, and stored at −80 °C until purification. Purification of RPAP2–RPAP2(1-215) was expressed in E. coli as described48 without addition of 0.1 mM ZnCl₂ in the expression medium.

Protein purification.

All protein purification steps were performed at 4 °C unless stated otherwise. Frozen pellets of H5 cells expressing full-length RPAP2 were thawed at room temperature and lyzed by sonication. The lysate was cleared by centrifugation. Cat lysates were filtered through 0.8 µm syringe filters (Merck Millipore) and applied to a 5 mL HisTrap HP column (GE Healthcare) equilibrated in lysis buffer. The column was washed with 50 mL of wash buffer containing 250 mM imidazole with 500 µM IPTG for 18 h at 18 °C. Harvested cells were resuspended in buffer H, 30–250 mM HEPES pH 7.4, 100 mM KCl, 10% glycerol (v/v), 30 mM imidazole, 5 mM BME, 0.284 µg/mL leupeptin, 1.37 µg/mL peptatin A, 0.17 mg/mL PMSF, and 0.33 mg/mL benzamidine, and stored at −80 °C until purification. Cloning and expression of human TBP, TFIIH, and TFIIF will be described elsewhere.

Protein purifications were performed in cold rooms at 4 °C and washed with washing buffer. The washed proteins were applied to a 5 mL HisTrap column (GE Healthcare) equilibrated in complex buffer (20 mM HEPES pH 7.4, 0.4% (v/v) glycerol, 3.5 mM MgCl₂, and 1 mM TCEP) and incubated for 2 min at RT. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole.

Protein purifications were performed in cold rooms at 4 °C and washed with washing buffer. The washed proteins were applied to a 5 mL HisTrap column (GE Healthcare) equilibrated in complex buffer (20 mM HEPES pH 7.4, 0.4% (v/v) glycerol, 3.5 mM MgCl₂, and 1 mM TCEP) and incubated for 2 min at RT. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole.

Protein purifications were performed in cold rooms at 4 °C and washed with washing buffer. The washed proteins were applied to a 5 mL HisTrap column (GE Healthcare) equilibrated in complex buffer (20 mM HEPES pH 7.4, 0.4% (v/v) glycerol, 3.5 mM MgCl₂, and 1 mM TCEP) and incubated for 2 min at RT. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole. The HisTrap column was detached and the amylose column was washed with 100 mL of wash buffer supplemented with 250 mM imidazole.
ab initio 3D model was generated from particles selected from good 2D classes. 

--- 1.6 million particles were extracted in Relion 3.0 in a box of 180 pixels and binned to a pixel size of 2.1 Å/pixel. Extracted particles were subjected to 3D classification using a 60 Å low-pass filtered ab initio model generated in cryo-EM as a reference to eliminate bad particles. For further cleaning of the data, selected good particles were subjected to 2D classification in Relion limiting the resolution to 10 Å. The resulting 1.2 million good particles were re-extracted without binning and refined using the ab initio map scaled and low-pass filtered to 60 Å as a reference. To improve the resolution of density corresponding to RPAP2, we applied a soft mask encompassing RPBS, RPAP2 density, and the RPBI jaw-lobe module (Supplementary Fig. 1f) and performed 3D classification without image alignment. We performed a 3D refinement of the best class containing 365,000 particles to obtain a 2.8 Å reconstruction (map 1). We then performed a focused 3D refinement with mask to obtain a 3.0 Å reconstruction of the region around RPBS, RPAP2, and the RPBI jaw-lobe module (map 2). Density corresponding to the Pol II clamp and RP48/RPB7 stalk domains were very noisy and weak suggesting mobility and/or flexibility of these domains. We therefore excluded them in our modeling.

Model building and refinement. To build an atomic model of the Pol II–RPAP2 complex, we used the structure of transcribing Pol II (PDB: 5FLM) after nucleic acid interactions. We focused 3D refinement and density fitting using a 60 Å low-pass filtered map. We then performed a 3D refinement of the best class containing 365,000 particles to obtain a 2.8 Å reconstruction (map 1). We then performed a focused 3D refinement with mask to obtain a 3.0 Å reconstruction of the region around RPBS, RPAP2, and the RPBI jaw-lobe module (map 2). Density corresponding to the Pol II clamp and RP48/RPB7 stalk domains were very noisy and weak suggesting mobility and/or flexibility of these domains. We therefore excluded them in our modeling.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The cryo-EM reconstructions and final model were deposited with the Electron Microscopy Data Base (EMDB) under accession code EMD-12087 and with the Protein Data Bank (PDB) accession 7BU. Source data for Fig. 2c are available with the paper online.

Received: 3 December 2020; Accepted: 8 April 2021;
Published online: 21 May 2021

References
1. Osman, S. & Cramer, P. Structural biology of RNA polymerase II transcription: 20 years on. Annu. Rev. Cell Dev. Biol. 36, 1–34 (2020).
2. Wild, T. & Cramer, P. Biogenesis of multisubunit RNA polymerases. Trends Biochem. Sci. 37, 99–105 (2012).
3. Czeko, E., Seizl, M., Augsberger, C., Mielke, T. & Cramer, P. Iwr1 directs RNA polymerase II nuclear import. Mol. Cell 37, 797–809 (2007).
4. Jeronimo, C. et al. Systematic analysis of the protein interaction network for transcription: 20 years on. Annu. Rev. Cell Dev. Biol. 34, 290–317 (2020).
5. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods 14, 290–296 (2017).
6. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination. Nat. Methods 14, 290–296 (2017).
7. Chen, L. et al. RNA polymerase II associated proteins regulate stomatal development through directly interacting with the stomatal transcription factors in Arabidopsis thaliana. New Phytol. 230, np.h17004 (2020).
8. Mochida, M. et al. RNA-dependent nuclear accumulation of IVO targets auxin-responsive cell differentiation in arabidopsis. Plant Cell 29, 575–588 (2017).
9. Gibney, P. A., Fries, T., Bailer, S. M. & Morano, K. A. Rtr1 is the Saccharomyces cerevisiae homolog of a Novel family of RNA polymerase II-binding proteins. Eukaryot. Cell 7, 938–948 (2008).
10. Guerrero-Serrano, G. et al. Npa3/ScGpn1 carboxy-terminal tail is dispensable for cell viability and RNA polymerase II nuclear targeting but critical for microtubule stability and function. Biochem. Biophys. Acta-Mol. Cell. Res. 1864, 451–462 (2017).
11. Xiang, K., Manley, J. L. & Tong, I. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. Nat. Commun. 3, 946 (2012).
12. Hsu, P. L. et al. Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr2 and Ser5 on the RNA polymerase II CTD. J. Mol. Biol. 426, 2970–2981 (2014).
13. Irahi, S. et al. Structure of Saccharomyces cerevisiae Rtr1 reveals an active site for an atypical phosphatase. Sci. Signal. 9, 1–12 (2016).
14. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T. & Murphy, S. Ser7 phosphorylation of the CTD recruits the RPAP2 Serphosphate to snRNA genes. Mol. Cell 45, 111–122 (2012).
15. Ni, Z. et al. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. Nat. Struct. Mol. Biol. 21, 686–695 (2014).
16. Mosley, A. L. et al. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol. Cell 34, 168–178 (2009).
17. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. Nature 529, 551–554 (2016).
18. Kristinell, E. & Henrich, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797 (2007).
19. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate RNA–DNA clamp. Nat. Struct. Mol. Biol. 24, 809–815 (2017).
20. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. Nature 533, 359–365 (2016).
21. Wani, S., Hirose, Y. & Ohkuma, Y. Human RNA polymerase II-associated protein RPB6 and participates in pre-mRNA 3’ end formation. Drug Discov. Ther. 8, 255–261 (2014).
22. Gradia, S. D. et al. Macroc8ac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes. Methods Enzymol. 592, 1–26 (2017).
23. Vos, S. M. et al. Architecture and RNA binding of the human negative elongation factor elf5. e14981 (2016).
24. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. Nature 560, 601–606 (2018).
25. Mastronarde, D. N. Automated electron microscope tomography using robust large-scale particle picking and refinement. J. Struct. Biol. 152, 36–51 (2006).
26. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods 16, 1146–1152 (2019).
27. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296 (2017).
28. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. elife 7, e42168 (2018).
29. Potters, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).
30. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. Sect. D 66, 486–501 (2010).
31. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and cryocrystallography. Acta Crystallogr. Sect. D 74, 531–544 (2018).
32. Waterhouse, A. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303 (2018).

Acknowledgements
We thank F. Grabe for purifying initiation factors, U. Neef and P. Rus for maintaining the EM facility. We thank Gaurika Garg and Srini-vasan Rengachari for nucleic acids used in EC and cPIC formation respectively. We thank members of the Cramer lab for general discussion of the data. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, SPP2191, EXC 2067/1-390729940) and the ERC Advanced Investigator Grant CHROMATRANS (grant agreement No 693023). S.A. was supported by H2020 Marie Curie Individual Fellowship (894862).

Author contributions
I.F. conceived, designed, and executed all experiments and analyzed data, unless stated otherwise. I.F. and C.D. collected cryo-EM data. I.F. built model with help from S.A. and S.S.S. established human initiation factors preparations and purified TFB1a-h, TFB1b. I.F. and C.D. interpreted the data and wrote the manuscript with input from all authors.

Funding
Open Access funding enabled and organized by Project DEAL.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02088-z.

Correspondence and requests for materials should be addressed to P.C.

Reprints and permission information is available at http://www.nature.com/reprints

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