Supplementary Information

Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states Girbig & Misiaszek et al.

Supplementary Note 1 | Cryo-EM data collection and processing and structural model building

Cryo-EM data collection and processing

Automated data acquisition was performed using SerialEM⁷⁶. Grids were screened using a Talos Arctica TEM operated at 200 keV (Thermo Fisher Scientific) and equipped with a Falcon III direct electron detector (Thermo Fisher Scientific). A small dataset of Apo Pol III applied to a 300-mesh Au R1.2/1.3 grid (Quantifoil) was recorded at a magnification of 92,000x corresponding to a pixel size of 1.566 Å/pixel. A total of 753 image stacks (12 frames, exposure time 2.1 s) were collected using a dose rate of 3.91 e/ A^2 /frame (46.92 e/ A^2 total dose) at a defocus range of 1-3.5 µm on an FEI Talos Arctica microscope equipped with a Falcon III detector operated in linear mode. Highresolution cryo-EM data on apo Pol III, plunge frozen on a 200-mesh Cu R2/1 grid (Quantifoil), was collected on a Titan Krios TEM operated at 300 keV (Thermo Fisher Scientific) and equipped with a Quantum energy filter (Gatan) and a K3 direct detector (Gatan). A magnification of 105,000x was used, corresponding to a pixel size of 0.822 Å/pixel. We recorded 9,544 image stacks (40 frames, exposure time 1.2 s) in counting mode using a dose of 0.95 $e/\text{\AA}^2$ /frame (38.11 $e/\text{\AA}^2$ total dose) and a defocus range of 0.75-2.25. High-resolution cryo-EM data on elongating Pol III was collected on a 300 keV Titan Krios TEM equipped with a K2 direct detector (Gatan). We used a magnification of 130,000x, corresponding to a pixel size of 1.05 Å/pixel, and collected 9,488 micrographs (40 frames, exposure time 11.2 s) in counting mode at a dose of 1.01 e/ \AA^2 /frame (40.35 e/ \AA^2 total dose). The defocus range was 0.75-2.00 µm.

For all datasets, initial frame alignment, contrast transfer function (CTF)-estimation, doseweighting and automated particle picking was performed using Warp^7 . The images were further processed with RELION 3.1⁶⁸. In order to run particle polishing at later stages during processing, the micrographs, acquired with a Titan Krios, were also frame-aligned and dose-weighted using RELION's own implementation of the MotionCor2-algorithm⁷⁸ and CTF-corrected using Gctf⁷⁹. For the Talos Arctica dataset, a total of 65,009 particles were picked and extracted with a box size of 230 pixels. The images were processed with cryoSPARC⁶⁹. After 2D classification, 48,473 particles were chosen to generate an initial model using 2 classes. One of the initial models resembled the shape of yeast RNA Pol III and was chosen for homogenous refinement using 37,932 particles. The resulting map of apo human Pol III reached a resolution of 8.0 Å.

In case of the high-resolution apo Pol III dataset, 304,683 particles were picked and extracted with a box size of 328 pixels using Warp. Subsequently, global 3D classification in RELION was performed using 4 classes. As a reference model, we used the 8.0 Å apo human Pol III map obtained from the Talos Arctica dataset, and low-pass filtered it to 60 Å. A single class was selected, and the 111,289 particles were subjected to 3D refinement yielding a 3D reconstruction at 3.8 Å resolution. The same set of particles was re-extracted with a 440 pixels box size from the micrographs that were pre-processed in RELION and subjected to 3D refinement. This was followed by two iterative rounds of particle polishing and CTF refinement, which included the estimation of anisotropic magnification and high-order aberrations⁸⁰. Subsequent 3D refinement and map sharpening yielded a 3D reconstruction at 3.3 Å resolution (Map A), which was further subjected to multi-body refinement⁶⁷ using two masks. Mask 1 covered the core of the polymerase and subunits RPC4/5/10, and mask 2 included the stalk, the clamp and the RPC3/6/7 heterotrimer. The two resulting maps, A1 and A2, were refined to 3.2 and 3.4 Å resolution, respectively, and used for model building of the apo Pol III structure.

For the elongating Pol III complex, 367,717 particles, extracted with a box size of 220 pixels, were subjected to 3D classification using a 60 Å low-pass filtered reconstruction of apo human Pol III as a reference map. One of the 4 classes, containing 166,071 particles, showed clear density for the bound downstream DNA and was refined to a resolution of 3.2 Å. After particle polishing and CTF refinement, the resolution was further improved to 2.8 Å (Map B). To improve map quality of less well-resolved regions, we performed masked classification in RELION without image alignment and asked for 2 classes per classification procedure. Soft masks were created using the molmap command in UCSF Chimera⁸¹ on the regions of interest followed by mask creation in RELION.

First, we applied a mask covering the downstream DNA and the DNA/RNA hybrid. Using a regularization parameter T of 10 during 3D classification, we obtained a class containing 70,492 particles, which showed improved signal corresponding to the DNA/RNA hybrid in the active site (map C, 3.1 Å). During model building (see below), we observed additional density starting from the linker domain of RPC10 and extending into the polymerase funnel. Therefore, we placed the C-ribbon domain of RPC10 into the funnel using Chimera, created a soft mask and performed masked classification on this area using a regularization parameter of 20. Class 1 contained 30,525 particles and showed clear density inside the funnel (map D, 3.1 Å) and, compared to map B, only

little residual density corresponding to the RPC10 C-ribbon domain outside the funnel. Class 2, containing 100,152 particles, showed no density inside the funnel but stronger signal belonging to the RPC10 ribbon outside the funnel (map E, 2.9 Å). At lower threshold levels, we observed additional density close the cryo-EM density that corresponds to the RPC4/5 heterodimer orthologue C53/C37 in *S. cerevisiae,* leading us to assume that this density belongs to the additional RPC5 WH domains. Thus, we first performed masked classification using a mask that covers the RPC4/5 dimerization module, followed by masked classification on the additional density. We manually place a homology model obtained for the WH1 domain of RPC5 (278-352), created using the Phyre⁸² web server, into this density. The fit of the placement was also inspected using $COOT^{83,84}$, and the C-terminal helix (339-352) of the homology model was re-arranged so that it fitted a close-by density that resembled the shape of a longer helix. The fitted homology model was used for mask creation, and masked 3D classification was performed using a regularization parameter T of 40. One class contained 21,861 particles and showed improved density corresponding to the RPC5 WH 1-2 domains binding the polymerase core (map F, 3.3 Å). We noticed that mostly the polymerase clamp, stalk and heterotrimer change conformation during the transition from apo to elongating Pol III whereas the remaining parts stay rigid. Therefore, we merged the two datasets using RELION in order to improve densities for less well-resolved regions. Pol III apo particles (map A) were re-extracted, binned to a pixel size of 1.05 Å/pixel, merged with the Pol III EC particles (map B) using RELION, and imported into cryoSPARC for further processing steps. In cryoSPARC, the particles were subjected to homogenous refinement, followed by CTF refinement and non-uniform refinement⁸⁵. The resulting 3D reconstruction (map G, 2.9 Å) showed improved density for an area near the RPC4/5 heterodimer that extended towards the polymerase jaw domain. The merged dataset was also used to repeat masked classification on the RPC5 C-terminal extension to resolve a second conformation of the WH domains. We followed the same strategy as described for map F but, this time, focused on the counterpart map, in which additional density was visible at lower threshold levels. A second round of masked classification on the RPC5 WH domains yielded a map containing 31,737 particles that showed density for the WH domains in the alternative conformation. The corresponding density could be slightly improved by non-uniform refinement in cryoSPARC (map H, 3.4 Å). Resolution values reported above are based on the gold-standard Fourier shell correlation (FSC) using the 0.143 cut-off criterion⁸⁶ and were obtained using RELION post-processing. Local resolutions of map A to map

F were estimated using the local resolution estimation tool implemented in RELION. For map G and H, Mono Res^{87} was used to estimate the local resolution range.

To get insights into conformational dynamics, we used the 3D variability analysis $(3DVA)^{41}$ tool implemented into cryoSPARC. Particles belonging to map B, obtained via processing in RELION, were imported into cryoSPARC, subjected to homogenous refinement followed by 3DVA using a soft mask covering the complete map.

Structural model building and refinement

An initial model of apo human RNA Pol III was constructed using homology models and available structures. Homology models of RPC1, RPC2, RPC4 and RPAC2 were obtained from the SWISS-MODEL Repository⁸⁸. Homology models of RPC5, RPC6, RPC7, RPC8, RPC9, RPC10 and RPAC1 were generated using Phyre2. The homology models and the structures of RPC3-RPC7 (PDB: 5AFQ) and of RPABC1, RPABC2, RPABC3, RPABC4, RPABC5 (PDB: 5IY6) were first aligned to their homologues counterparts in the Pol III apo structure from S. cerevisiae (PDB: 5FJ9). The constructed 17-subunit model was fitted as a rigid body into the cryo-EM density map of apo human Pol III using Chimera, and the individual polypeptide chains were fitted individually as rigid bodies followed by manual model building and real-space refinement using COOT.

For the elongating RNA Pol III structure, the model of the apo human Pol III was first fitted into the cryo-EM density of the elongating Pol III map B in Chimera, and the chains were adjusted manually in COOT. Cryo-EM density, corresponding to the upstream DNA, was observed in the DNA-binding cleft, and the DNA/RNA-duplex molecule from the elongating *S. cerevisiae* RNA Pol III structure (PDB: 5FJ8) was placed into the corresponding density. Interatomic distance restraints of the DNA/RNA molecule were generated using ProSMART⁸⁹, implemented in the most recent version of COOT⁸⁴, and the nucleic acids were first fitted automatically into the density and refined using the all-molecule-refine option in COOT. The nucleotide bases were then mutated to match the sequences of the used nucleotides, followed by manual fitting and refinement against map C to obtain a more accurate fit.

When building the model of RPC10, we observed that the density corresponding to the linker and the C-ribbon domain needs to be traced differently compared to the cryo-EM apo structure of *S. cerevisiae* RNA Pol III (PDB: 5fja). From residue E48 ongoing, we observed continuous density that split-up and either extended into the Pol III funnel or towards the polymerase periphery, which could be resolved by masked classification (see above). For the building of the RPC10-C-ribbon

domain, we used map D (inside funnel) and map E (outside funnel). In both cases, a homology model (67-107) was placed into the corresponding densities using Chimera, followed by manual fitting and refinement using COOT. In addition, $ISOLDE⁹⁰$ (implemented into the ChimeraX software⁹¹) was used to reduce the clash score and to optimise fitting with applied secondarystructure restraints on predicted beta-sheets based predictions obtained via the PSIPRED webserver⁹². Placement of the zinc atom in the C-ribbon domain was partially guided by the crystal structure of the C-ribbon of *S. cerevisiae* TFIIS bound to *S. cerevisiae* RNA Pol II (PDB: 3PO3). For building the FeS domain of RPC6, we placed the coordinates of a cubane 4Fe-4S cluster (obtained from PDB: 2B3Y) manually into a strong density signal within the heterotrimer region that was even visible at very high threshold levels. The FeS binding domain could be built d*e novo* by extending the RPC6 homology model into visible density extending towards the 4Fe-4S signal. For the C-terminal extension of RPC5, homology models of WH1 (278-332) and WH2 (349-433), generated via Phyre2, were first placed manually into the density (map F) using Chimera and manually adjusted and extended in COOT. The model of the WH domains was further placed in map H, in which the C-terminal extension adopts an alternative conformation 2. The WH1-WH2 domain was placed as a rigid body into the 'conformation 2' density in Chimera and flexibly fitted using the chain-refine command in COOT in two iterations, first with and second without distance restraints. Manual fitting of individual side-chains was not performed for the RPC5 conformation 2, and the model was trimmed by 25 peripheral residues. The DM extension and the jaw-binding domains of RPC4 were built using map G.

For model building, all cryo-EM density maps were, if not otherwise stated, locally sharpened using LocalDeblur⁹³, integrated into the software framework Scipion⁹⁴. At a later stage during model building, we also used the density modification tool⁹⁵ implemented into PHENIX⁹⁶ to further improve the map qualities.

PHENIX real-space refinement was used to refine the built models against the sharpened cryo-EM maps. All refinement steps were performed using Ramachandran- and rotamer restraints. For finetuning of model building, combining partial models and refinement, the following strategy was used: partial maps of elongating Pol III derived from RELION multi-body refinement (map B1, B2) were sharpened via density modification and used to fine-tune the model of elongating Pol III. These partial models were refined against map B1 and B2, combined and refined with secondarystructure and reference-model restraints, derived from the partial models, against map B. The model was subjected to ISOLDE and Ramachandran, and rotamer outliers were adjusted, following refinement against map B using reference-model restraints. For refinement of more flexible regions, the following partial-model and map pairs were used for refinement: DNA/RNA – map C; RPC10-C-ribbon inside funnel – map D; RPC-10-C-ribbon outside funnel – map E; RPC5 C-terminal extension (conformation 1) – map F; RPC4 DM extension and jaw-binding domain – map G: RPC5 C-terminal extension (conformation 2) – map H. Model-based restraints for partial models were applied when refining against other maps. For the complete elongating Pol III model, partial models were combined and refined against map B using reference-model restraints. Two disease-associated residues (RPC1 W310, RPC2 W27) were manually adjusted to better fit the cryo-EM density without an additional real-space refinement step in PHENIX, and the resulting final elongating Pol III model (EC-1 Pol III) was subjected to the PHENIX comprehensive validation tool to obtain final refinement statistics. For the apo Pol III model, nucleic acids and the RPC5 C-terminal extension of elongating Pol III model were removed, and chains were manually readjusted to fit the density of apo Pol III (map A), and the model was refined against map A using reference-model restraints.

Supplementary Table 1 | Phylogenomic analysis of Pol III RPC5 subunits and Pol III basal transcription factors. Listed are the analysed species, proteome identifiers and UniProt identifiers for homologs of RPC5 (Pol III), BRF1 and BRF2 (basal transcription factor TFIIIB) and SNAPc1, SNAPc2, SNAPc3, SNAPc4, SNAPc5 (basal transcription factor SNAPc). Homologs were retrieved using the HMMER⁶⁰ tool. See Extended Data Fig. 7e for the corresponding phylogenomic tree.

Species	Proteome ID	RPC ₅	BRF1	BRF ₂	SNAPc1	SNAPc2	SNAPc3	SNAP _{c4}	SNAPc5
Takifugu rubripes	UP00000 5226	A0A3B5 KDH ₁	H ₂ UWY 0	A0A3B5 JWS9	H2S607	H ₂ SIN ₉	H _{2TJN2}	A0A3B5 KEC1	A0A3B5 KKF9
Danio rerio	UP00000 0437	F1R866	Q5TZ89	A8KBY2	Q7ZUP7	E7F685	F1QWM 0	F ₁ QVS ₅	Q1LXI0
Rattus norvegicus	UP00000 2494	D3ZP45	D4A8W8	Q4V8D6	F7ES94	Q68FX5	Q5BK68	D4A3C9	D3ZTK9
Mus musculus	UP00000 0589	Q9CZT4	Q8CFK2	Q3UAW 9	Q8K0S9	Q91XA5	Q9D2C9	Q8BP86	Q8R2K7
Felis catus	UP00001 1712	A0A2I2V 2B4	A0A337 S677	M3X9N0	M3W8U 2	M3WLN 3	M3VZV1	A0A5F5 XJV4	M3WGN
Canis lupus	UP00000 2254	E2RHN7	F1PRK4	E2R245	E2QXT2	J9NTT0	E2REJ9	J9NS15	E2RR56
Homo sapiens	UP00000 5640	Q9NVU0	Q92994	Q9HAW 0	Q16533	Q13487	Q92966	Q5SXM2	O75971
Pan troglodytes	UP00000 2277	K7BW50	H2Q908	H2QW11	H2Q8F4	H _{2RHY8}	H2R1P5	H2QY63	H2Q9N3

Supplementary Table 1 (continued)

Supplementary Table 2 | Comparison between sequence identities of *H. sapiens* **(Hs) RPC7α/β and** *S. cerevisiae* **(Sc) C31.** Listed are sequence identities comparisons of: HsRPC7α vs. ScC31 and of Hs RPC7β vs. Sc C31. Full length RPC7a: 1-223; Full length RPC7β: 1-218.

Supplementary Table 3 | Disease-associated point mutations in Pol III. In bold mutations have been reported as homozygous. SASA – Solvent Accessible Surface Area. Average for whole complex: SASA - 42.4 Å²; degree - 15.

Supplementary Table 4 | Fit of disease-associated Pol III mutations into the cryo-EM density. Model: Human Pol III EC-1, Cryo-EM density: map B. Threshold level: 0.04. Residues are displayed with flanking residues (-1,1). Only 1 residue (RPC5 275) shows weak side-chain density.

Supplementary References

- 67. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, e36861 (2018).
- 68. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
- 69. 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).
- 70. Hwang, S., Gou, Z. & Kuznetsov, I. B. DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins. *Bioinformatics* **23**, 634– 636 (2007).
- 71. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
- 72. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
- 73. Kim, M. K. *et al.* Assembly of SNAPc, Bdp1, and TBP on the U6 snRNA Gene Promoter in Drosophila melanogaster. *Mol. Cell. Biol.* **40**, e00641-19 (2020).
- 74. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
- 75. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
- 76. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
- 77. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
- 78. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- 79. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
- 80. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).
- 81. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
- 82. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
- 83. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
- 84. Casañal, A., Lohkamp, B. & Emsley, P. Current Developments in Coot for Macromolecular Model Building of Electron Cryo‐microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069-1078 (2019).
- 85. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
- 86. Rosenthal, P. B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
- 87. Vilas, J. L. *et al.* MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. *Structure* **26**, 337-344.e4 (2018).
- 88. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).
- 89. Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC 5. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 404–417 (2012).
- 90. Croll, T. I. ISOLDE : a physically realistic environment for model building into lowresolution electron-density maps. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 519–530 (2018).
- 91. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
- 92. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
- 93. Ramírez-Aportela, E. *et al.* Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* **36**, 765-772 (2020).
- 94. de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
- 95. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
- 96. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861– 877 (2019).
- 97. Timmons, M. *et al.* Peripheral and central hypomyelination with hypogonadotropic hypogonadism and hypodontia. *Neurology* **67**, 2066–2069 (2006).
- 98. Daoud, H. *et al.* Mutations in *POLR3A* and *POLR3B* are a major cause of hypomyelinating leukodystrophies with or without dental abnormalities and/or hypogonadotropic hypogonadism. *J. Med. Genet.* **50**, 194–197 (2013).
- 99. Shimojima, K. *et al.* Novel compound heterozygous mutations of POLR3A revealed by whole-exome sequencing in a patient with hypomyelination. *Brain Dev.* **36**, 315–321 (2014).
- 100. Wolf, N. I. *et al.* Leukoencephalopathy with ataxia, hypodontia, and hypomyelination. *Neurology* **64**, 1461–1464 (2005).
- 101. Wolf, N. I. *et al.* Ataxia, delayed dentition and hypomyelination: A novel leukoencephalopathy. *Neuropediatrics* **38**, 64–70 (2007).
- 102. Terao, Y. *et al.* Diffuse central hypomyelination presenting as 4H syndrome caused by compound heterozygous mutations in POLR3A encoding the catalytic subunit of polymerase III. *J. Neurol. Sci.* **320**, 102–105 (2012).
- 103. Potic, A., Brais, B., Choquet, K., Schiffmann, R. & Bernard, G. 4H syndrome with lateonset growth hormone deficiency caused by POLR3A mutations. *Arch. Neurol.* **69**, 920– 923 (2012).
- 104. Kraoua, I. *et al.* Novel *POLR1C* mutation in RNA polymerase III-related leukodystrophy with severe myoclonus and dystonia. *Mol. Genet. Genomic Med.* **7**, e914 (2019).