

Materials Design Analysis Reporting (MDAR) Checklist for Authors

The MDAR framework establishes a minimum set of requirements in transparent reporting applicable to studies in the life sciences (see Statement of Task: [doi:10.31222/osf.io/9sm4x](https://doi.org/10.31222/osf.io/9sm4x)). The MDAR checklist is a tool for authors, editors and others seeking to adopt the MDAR framework for transparent reporting in manuscripts and other outputs. Please refer to the MDAR Elaboration Document for additional context for the MDAR framework.

Materials

Antibodies	Yes (indicate where provided: page no/section/legend)	n/a
For commercial reagents, provide supplier name, catalogue number and RRID, if available.		n/a
Cell materials	Yes (indicate where provided: page no/section/legend)	n/a
Cell lines: Provide species information, strain. Provide accession number in repository OR supplier name, catalog number, clone number, OR RRID	Hi5 cells: Expression Systems, Tni insect cells in ESF921 media, item 94-002F Sf9 cells: Thermo Fisher Scientific, cat-12659017, Sf9 cells in Sf-900TM III SFM	
Primary cultures: Provide species, strain, sex of origin, genetic modification status.		n/a
Experimental animals	Yes (indicate where provided: page no/section/legend)	n/a
Laboratory animals: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository OR supplier name, catalogue number, clone number, OR RRID		n/a
Animal observed in or captured from the field: Provide species, sex and age where possible		n/a
Model organisms: Provide Accession number in repository (where relevant) OR RRID		n/a
Plants and microbes	Yes (indicate where provided: page no/section/legend)	n/a
Plants: provide species and strain, unique accession number if available, and source (including location for collected wild specimens)		n/a
Microbes: provide species and strain, unique accession number if available, and source	Escherichia coli XL1 (Agilent), stable cells (NEB)	
Human research participants	Yes (indicate where provided: page no/section/legend)	n/a
Identify authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		n/a
Provide statement confirming informed consent obtained from study participants.		n/a
Report on age and sex for all study participants.		n/a

Design

Study protocol	Yes (indicate where provided: page no/section/legend)	n/a
For clinical trials, provide the trial registration number OR cite DOI in manuscript.		n/a
Laboratory protocol	Yes (indicate where provided: page no/section/legend)	n/a
Provide DOI or other citation details if detailed step-by-step protocols are available.		n/a
Experimental study design (statistics details)	Yes (indicate where provided: page no/section/legend)	n/a
State whether and how the following have been done, or if they were not carried out.		
Sample size determination		n/a
Randomization		n/a
Blinding		n/a
Inclusion/exclusion criteria		n/a
Sample definition and in-laboratory replication	Yes (indicate where provided: page no/section/legend)	n/a
State number of times the experiment was replicated in laboratory	RNA cleavage assays were highly reproducible. We provided two and three representative replicates for assay on stem loop and single-stranded RNA substrates, respectively in Data S2 (separate file).	
Define whether data describe technical or biological replicates	Biological replicates	
Ethics	Yes (indicate where provided: page no/section/legend)	n/a
Studies involving human participants: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		n/a
Studies involving experimental animals: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval.		n/a
Studies involving specimen and field samples: State if relevant permits obtained, provide details of authority approving study; if none were required, explain why.		n/a
Dual Use Research of Concern (DURC)	Yes (indicate where provided: page no/section/legend)	n/a
If study is subject to dual use research of concern, state the authority granting approval and reference number for the regulatory approval		n/a

Analysis

Attrition	Yes (indicate where provided: page no/section/legend)	n/a
State if sample or data point from the analysis is excluded, and whether the criteria for exclusion were determined and specified in advance.		n/a
Statistics	Yes (indicate where provided: page no/section/legend)	n/a
Describe statistical tests used and justify choice of tests.		n/a
Data Availability	Yes (indicate where provided: page no/section/legend)	n/a
State whether newly created datasets are available, including protocols for access or restriction on access.	The cryo-EM maps and atomic coordinates were deposited in the Electron Microscopy Data Bank under accession code EMD-13479, and in the Protein Data Bank under accession code 7PKS. Materials are available from Patrick Cramer upon request under a material transfer agreement with the Max Planck Society.	
If data are publicly available, provide accession number in repository or DOI or URL.	The cryo-EM maps and atomic coordinates were deposited in the Electron Microscopy Data Bank under accession code EMD-13479, and in the Protein Data Bank under accession code 7PKS.	
If publicly available data are reused, provide accession number in repository or DOI or URL, where possible.	The following atomic coordinates from the Protein Data Bank were used as input models for structural modeling: 6GML, 7CUN, 7BFP, 3DW8	
Code Availability	Yes (indicate where provided: page no/section/legend)	n/a
For all newly generated code and software essential for replicating the main findings of the study:		n/a
State whether the code or software is available.		n/a
If code is publicly available, provide accession number in repository, or DOI or URL.		n/a

Reporting

Adherence to community standards	Yes (indicate where provided: page no/section/legend)	n/a
MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.		n/a
State if relevant guidelines (eg., ICMJE, MIBBI, ARRIVE) have been followed, and whether a checklist (eg., CONSORT, PRISMA, ARRIVE) is provided with the manuscript.		n/a



Supplementary Materials for

Structural basis of Integrator-mediated transcription regulation

Isaac Fianu *et al.*

Corresponding author: Patrick Cramer, patrick.cramer@mpibpc.mpg.de

Science **374**, 883 (2021)
DOI: 10.1126/science.abk0154

The PDF file includes:

Materials and Methods
Figs. S1 to S11
Tables S1 and S3
References

Other Supplementary Material for this manuscript includes the following:

Movies S1 to S4
Data S1
MDAR Reproducibility Checklist

Materials and Methods

Molecular Cloning

The open reading frames (ORFs) of human Integrator subunits were amplified by PCR from cDNA clones (Harvard plasmid repository) except INTS1 and INTS2. PCR-amplified ORFs were inserted into the 438 series of MacroLab vectors (modified pFastBac, Addgene #55218 and #55220) by ligation independent cloning (41). The three main vectors employed in this study were 438-A, 438-B and 438-C, which encode no tag, an N-terminal 6xHis tag and an N-terminal 6xHis-MBP tag, respectively. For INTS1, the ORF was divided into four and synthetic DNA for each part was purchased from Integrated DNA Technologies (IDT). The fourth fragment encoded a 10xHis tag at the C-terminus. The four fragments were assembled into the 438-C vector by uracil excision cloning (42). To obtain the ORF for INTS2, two synthetic DNA fragments encoding the full-length protein were purchased and assembled into the 438-A vector using Circular Polymerase Extension Cloning (43). INTS1 and INTS12 truncations as well as INTS11(E203Q) mutant were created using 'round the horn' PCR-based mutagenesis (44). Co-expression vectors containing multiple subunits were created by successive rounds of ligation-independent cloning. Constructs were amplified using XL1 blue chemically competent bacteria (Agilent) except for those larger than 25,000 base pairs which were amplified using NEB® Stable Competent *E. coli* (NEB).

To reconstitute the 14-subunit Integrator after heterologous protein expression in insect cells, we generated four vectors for baculovirus-based overexpression. The first vector contains subunits of the cleavage module (INTS4-9-11) (18) with a 6xHis-MPB tag on INTS11. The second vector contained the INTS10-13-14 heterotrimer (21, 45) with a 6xHis tag on INTS13. Using systematic co-expression of Integrator subunits, affinity purification and crosslinking mass spectrometry, we identified the 8-subunit Integrator core that forms a stable complex. For this subcomplex we created two vectors one containing 6xHis-MPB-INTS1-12 heterodimer and the other containing INTS2-3-5-6-7-8 subcomplex with 6xHis-MBP tag on INT5 and 6xHis tag on INTS6.

To clone the human PP2A complex, the ORF for PP2A-A was amplified using a purchased plasmid (Addgene: #81760) as template and the ORF for PP2A-C was purchased as a synthesized DNA (IDT). PP2A-A and PP2A-C were cloned into 438-A and 438-B vectors, respectively by ligation independent cloning. To increase the protein yield, a catalytically dead mutant of PP2A-C (D88N) (46) was generated by site-directed mutagenesis PCR (Qiagen) using the 438B-PP2A-C plasmid as template. All clones were verified by sanger sequencing.

Protein expression and purification

Production of bacmids and baculoviruses was done essentially as described previously (47, 48). V1 baculoviruses were stored at 4 °C for 1-2 years except for the baculovirus co-expressing 6-subunit subcomplex (INTS2-3-5-6-7-8) which was made every 6 months to ensure protein expression. Protein expression was induced in Hi5 cells by infecting a 300 ml starting culture with 150 µL of V1 viruses. For expression of 8-subunit Integrator core, we co-infected Hi5 cells with the two baculoviruses harboring INTS2-3-5-6-7-8 and INTS1-12 dimer. Careful titration of viruses

was necessary to determine the amount of each virus. The PP2A complex was expressed by co-infection of Hi5 cells with baculoviruses expressing PP2A-A and PP2A-C(D88N).

To prepare the Integrator cleavage module, Hi5 insect cells expressing the cleavage module were harvested ~72 hrs after infection by centrifugation at 238xg for 30 min. The supernatant was discarded and the cell pellet from 1 L culture was resuspended in 35 ml of lysis buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 30 mM imidazole, 10% (v/v) glycerol, 10 mM beta-mercaptoethanol (β ME), 10 μ M ZnCl₂, 0.284 μ g/ml leupeptin, 1.37 μ g/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine). The harvested cells were flash frozen in liquid nitrogen and stored at -80 °C until purification. The frozen pellets were thawed in a water bath at 25 °C and lysed by sonication using a BRANSON Digital sonifier. The lysate was cleared by centrifugation at 87,207xg for 1 hr and filtered through a 0.8 μ m syringe filter. The cleared lysate was applied to a pre-equilibrated 5 ml HisTrap column (GE Healthcare Biosciences) at a flow rate of 1.5 ml/min and the column was washed with 100 ml of wash buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 30 mM imidazole, 10% (v/v) glycerol, 10 mM β ME, 10 μ M ZnCl₂). A self-packed amylose column with a total bed volume of 15 ml pre-equilibrated in wash buffer was connected in tandem to the 5 ml HisTrap column. The bound protein was eluted from the 5 ml HisTrap column using Ni elution buffer (20 mM HEPES pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 10 mM β ME, 10 μ M ZnCl₂, 250 mM imidazole). The HisTrap column was detached and the amylose column was washed with 100 ml of wash buffer before eluting with wash buffer supplemented with 100 mM maltose. The fractions containing the heterotrimeric complex were combined and treated overnight with 6xHis-TEV protease to remove the affinity tag. The protein was applied to 5 ml HisTrap in a reverse nickel affinity step to remove the TEV protease, the affinity tag and proteins with undigested affinity tag. The unbound protein was collected and concentrated using a 100 kDa cut-off Amicon Ultra Centrifugal Filter (MERCK Millipore). The protein was further purified on a Superose 6 10/300 increase size exclusion column (GE Healthcare Biosciences) pre-equilibrated in size exclusion buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, 10 mM β ME, 10 μ M ZnCl₂). Peak fractions were combined, concentrated, flash frozen in liquid nitrogen and stored at -80 °C.

To prepare the INTS10-13-14 module, protein expression and harvesting of cultures were executed as described for the cleavage module using a baculovirus harboring the INTS10-13-14 heterotrimer. The cell pellet from 1 L culture was resuspended in 35 ml of lysis buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 10% (v/v) glycerol, 10 mM β ME, 30 mM imidazole, 0.284 μ g/ml leupeptin, 1.37 μ g/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine). The harvested cells were flash frozen in liquid nitrogen and store at -80 °C until purification. Frozen pellets of cells expressing the subcomplex were thawed in a water bath at 25 °C. Cell lysis and clearance of the lysate were performed as described for the cleavage module. The cleared lysate was applied to a pre-equilibrated 5 ml HisTrap column at a flow rate of 1.5 ml/min and contaminating proteins were removed by washing the column with 100 ml of the lysis buffer without protease inhibitors. Bound protein was eluted from the HisTrap column with a gradient (0-100%) of Ni elution buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 10% (v/v) glycerol, 10 mM β ME, 500 mM imidazole).

Fractions containing the heterotrimeric complex were pooled and treated with 6xHis-tagged TEV protease while dialyzing against the wash buffer to remove the affinity tag and reduce imidazole concentration. A reverse nickel affinity step was done to remove the TEV protease, the affinity tag and proteins with undigested affinity tag. The unbound protein was collected, concentrated using a 100 kDa cut-off Amicon Ultra Centrifugal Filter and applied to a Superose 6 10/300 Increase size exclusion column pre-equilibrated in a size exclusion buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, 10 mM β ME). Peak fractions were pooled, concentrated, flash frozen in liquid nitrogen and stored at -80 °C.

To prepare the 8-subunit Integrator core, we started with 8-10 L of Hi5 cells co-infected with the two baculoviruses harboring INTS1-12 and INTS2-3-5-6-7-8. Cultures were harvested as described for the cleavage module and cell pellets were suspended in lysis buffer (50 mM HEPES 7.4, 200 mM NaCl, 1 mM DTT, 10% (v/v) glycerol, 0.005% (w/v) CHAPS and cOmplete™ Protease Inhibitor cocktail (MERCK)). Resuspended cells were flash frozen in liquid nitrogen and stored at -80 °C until purification. Frozen cells were thawed at 25 °C and lysed by 2 cycles of 60 s sonication with a 2 min cooling of the sample between cycles. Lysate was cleared by centrifugation at 85,000xg for 30 min in a high-speed centrifuge and 45K rpm for 45 min in an ultracentrifuge using a Ti45 rotor (Beckman). The supernatant was filtered with a 0.8 μ m syringe filter and applied to a self-packed amylose column at 1.5 ml/min flow rate. The column was washed with 150 ml of wash buffer (50 mM HEPES 7.4, 200 mM NaCl, 1 mM DTT, 10% (v/v) glycerol, 0.005% (w/v) CHAPS) and eluted with wash buffer supplemented with 100 mM maltose. The peak fractions from amylose affinity purification were pooled and treated with TEV protease overnight. The protein was further purified on a 1 ml HiTrap Q column (GE Healthcare Biosciences) equilibrated in wash buffer. After washing the column with 10 ml of the wash buffer, the bound protein was eluted with a gradient (0-100%) of high salt buffer (20 mM HEPES pH 7.4, 1 M NaCl, 10% (v/v) glycerol, 1 mM DTT, 0.005% (w/v) CHAPS). The peak fractions were pooled and concentrated using a 100 kDa cut-off Amicon Ultra Centrifugal Filter (Millipore). Protein was used immediately for downstream experiments.

To reconstitute the complete Integrator, we mixed the purified Integrator core, cleavage module and INTS10-13-14 module in a 1:2:2 molar ratio and incubated on ice for 1 hour. The resulting complex was separated on a 10-40% (v/v) glycerol gradient that was prepared in complex buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 3 mM $MgCl_2$, 1 mM DTT, 0.005% (w/v) CHAPS). The gradient was spun at 34,000 rpm for 16 hours using a SW60 rotor (Beckman) in a Beckman ultracentrifuge operated at 4 °C. The gradient was manually fractionated into 200 μ l fractions from the top and fractions 4 – 20 were analyzed by SDS-PAGE followed by Coomassie staining. Protein bands were verified by mass spectrometry.

To prepare the PP2A phosphatase, Hi5 cells co-infected with baculoviruses expressing PP2A-A and PP2A-C(D88N) were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris 7.9, 100 mM NaCl, 5 mM β ME, 0.284 μ g/ml leupeptin, 1.37 μ g/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine). The resuspended cells were flash frozen and stored at -80 °C until protein purification. Lysis of cells and clearance of the lysate was performed as

described for the cleavage module. The cleared lysate was applied to a 5 ml HisTrap column (GE Healthcare Biosciences), which was pre-equilibrated with lysis buffer. The column was washed with lysis buffer and the complex was eluted from the nickel column with a linear imidazole gradient from 0 to 500 mM. The eluate of nickel column was cleaved with TEV protease to remove the His-tag on PP2A-C, while dialyzing in a dialysis buffer (50 mM Tris 7.9, 100 mM NaCl and 1 mM DTT). Some precipitation was observed during cleavage and dialysis. The cleaved PP2A complex in solution was loaded on tandem 5 ml HisTrap and HiTrap Q columns (GE Healthcare Biosciences), which were pre-equilibrated with Q buffer A (25 mM Tris 7.9, 100 mM NaCl and 1 mM DTT). The columns were washed with 40 ml Q buffer A, followed by removal of HisTrap column. After further washing the HiTrap Q column with 20 ml of Q buffer A, the complex was eluted with a linear salt gradient (100 mM-1M NaCl) and further polished with size exclusion using Superdex 200 10/300 increase column (GE Healthcare Biosciences). The PP2A complex was concentrated in the size exclusion buffer (25 mM Tris 7.9, 100 mM NaCl and 1 mM DTT) and stored at -80 °C.

To prepare proteins for reconstitution of the paused elongation complex (PEC), mammalian RNA Pol II was purified from *S. scrofa* thymus essentially as described (49), with minor modifications as described (50). Cloning and expression of full-length human DSIF (heterodimer of SPT4 and SPT5) was described (23) and purification was carried out according to our published protocol (22). Cloning, expression and purification of the 4-subunit human NELF complex was also done as described (22). All protein work was carried out on ice or at 4 °C unless otherwise stated.

Preparation of PEC-Integrator-PP2A complex

To investigate binding of Integrator and PP2A to the paused elongation complex (PEC), we compared the migration of Integrator, PEC and PEC-Integrator-PP2A on a 10-40% glycerol gradient. The gradient was prepared in complex buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 3 mM MgCl₂, 1 mM DTT, 0.005%(w/v) CHAPS). We first formed PEC as previously described using the same nucleic acid scaffolds (22) and analyzed its migration. Briefly, 100 pmol of Pol II, 200 pmol of annealed RNA-template DNA, 200 pmol non-template DNA were assembled. We then added 400 pmol each of DSIF and NELF in a buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl, 4% (v/v) glycerol, 3 mM MgCl₂, 1 mM DTT, 0.005%(w/v) CHAPS). The sample was incubated at 30 °C for 30 min and applied to the glycerol gradient. The gradient was spun at 34,000 rpm for 16 hrs in an SW60 rotor using a Beckman ultracentrifuge operated at 4 °C. The fractions from the gradient were analyzed as described for Integrator. To form the PEC-Integrator-PP2A complex, 300 pmol each of the purified Integrator subcomplexes and 300 pmol of the PP2A complex were mixed and incubated on ice for 1 hour to pre-form a 16-subunit Integrator-PP2A complex. We used the cleavage module containing INTS11E203Q variant for complex formation to impair RNA cleavage. In parallel, a Pol II elongation complex was assembled using 100 pmol of Pol II, 200 pmol of annealed RNA-template DNA and 200 pmol non-template DNA. We then added the Integrator-PP2A complex and 400 pmol each of DSIF and NELF to the Pol II elongation

complex. The sample was incubated for 30 min at 30 °C in the complex buffer and applied to the glycerol gradient. Fractions were analyzed by SDS-PAGE followed by Coomassie staining. The identity of all proteins in the peak fraction was confirmed by mass spectrometry.

Cryo-EM sample preparation

For cryo-EM analysis, fraction 15 from the PEC-Integrator-PP2A glycerol gradient was stabilized by batch crosslinking using 0.2% (v/v) glutaraldehyde for 10 min on ice. The crosslinker was quenched using 100 mM Tris-HCl (pH 8). The crosslinked sample was dialyzed for 7 hrs at 4 °C against a dialysis buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 1% (v/v) glycerol, 3 mM MgCl₂, 1 mM DTT, 0.01%(w/v) CHAPS) using a 20,000 MWCO Slide-A-Lyzer MINI Dialysis Unit (Thermo Fisher Scientific). For preparation of cryo-EM grids, homemade ~2.7 nm thin continuous carbon was floated on the dialyzed sample for ~2 min. The carbon was fished using a holey-carbon grid (Quantifoil R3.5/1, copper, mesh 200). Grids were blotted for 3 s with a blot force of 5 and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) operated at 4 °C and 100 % humidity.

Cryo-EM data collection and processing

Cryo-EM data collection was performed with SerialEM (51) using a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. Images were acquired in EFTEM mode with a slit width of 20 eV using a Quantum LS energy filter and a K3 direct electron detector (Gatan) at a nominal magnification of 81,000× corresponding to a calibrated pixel size of 1.05 Å/pixel. Images were recorded with -0.5 to -2.0 µm defocus range. About 45,000 exposures were recorded in counting mode for 3 s with a dose rate of 16.97 e⁻/pixel/s resulting in a total dose of 46.18 e⁻/Å² which was fractionated into 50 movie frames. Motion correction, dose weighting, CTF estimation and particle picking were performed on-the-fly using Warp (52). The data was divided into four parts for easier handling. Particles were extracted in a box of 200 pixels and binned to a pixel size of 3.15 Å/pixel using RELION 3.1 (53, 54). A fourth of the data set was subjected to 2D classification in cryoSPARC 2.15 (55). A good *ab initio* 3D model was generated using selected good 2D class averages and three bad *ab initio* models were calculated from selected bad 2D class averages using cryoSPARC. To clean the data from contamination and broken particles, heterogenous refinement was performed on the entire data set using the generated *ab initio* 3D models in cryoSPARC. The selected good particles were subjected to another round of cryoSPARC heterogenous refinement using only the good *ab initio* as input model. This allowed us to select ~610,000 particles which have good density for both Integrator-PP2A and PEC. This set of particles was re-extracted in a box of 480 pixels without binning and subjected to global 3D auto-refinement in RELION 3.1 to yield a 3.7 Å reconstruction of PEC-Integrator-PP2A. The resolution of the reconstruction was improved to 3.6 Å (Map 1) by one round of CTF refinement followed by 3D auto-refinement.

To improve the resolution of various regions of the map and to aid model building we used focused classification without image alignment and refinement of the best class. To obtain a high-

resolution reconstruction for the PEC, we applied a soft mask around Pol II, NELF and DSIF which led to a 2.7 Å reconstruction of the PEC. Parts of DSIF and NELF could not be improved further by focused classification due to conformational flexibility in these regions. In the same way, different regions of the Integrator-PP2A complex were improved by applying masks around regions that appeared to be one moving body or guided by the difference in local resolution. This allowed us to obtain maps better than 3.2 Å for most parts of the Integrator-PP2A complex that were used in modeling the respective subunits. Details of the masks used and the resultant focused maps are provided in **fig. S2**. All focused maps were aligned on Map 1 and a composite map for the overall structure was generated using Warp (52). During cryo-EM data processing, we noted density for the CTD on Integrator in the absence of stably bound PEC. To investigate this, classes lacking stably bound PEC from the cryoSPARC heterogeneous refinement were subjected to 3D refinement in RELION. We then applied a soft mask around the CTD-binding region and performed focused refinement. This revealed a defined density of the CTD in this population of particles despite the absence of PEC density (**fig. S11A-C**).

Model building and refinement

To build a model for PEC, the previously published PEC structure (6GML) (22) was rigid-body fitted into the focused map for PEC in UCSF Chimera (56) and manually adjusted in Coot (57). A model containing Pol II (except the RPB4-RPB7 stalk), SPT5 KOW5 domain, NELF-A, and NELF-C C-terminus (residues 193-590) was refined against this map to good chemical geometry using Phenix Real-Space refinement with standard settings and nonbonded weights of 500 (58, 59). The Pol II stalk (RPB4-RPB7), NELF-B, NELF-E and the N-terminal region of NELF-C (residues 55-192) were rigid-body fitted into the PEC map sharpened with a B factor of -50 Å^2 in RELION 3.1. The KOW2-3 and KOWx-4 domains of DSIF subunit SPT5 were rigid-body docked into a 6 Å-filtered Map 1.

To build a model for Integrator, we modeled the previously unobserved INTS1 N-terminal region based on results from threading using trRosetta (60) and BS3 crosslinking. The models of INTS1(1-950) and INTS1(900-1318) from trRosetta were flexibly fitted into the focused maps corresponding to INTS1(300-800) and INTS1(949-1316), respectively using NAMDINATOR (61) and manually adjusted in Coot. Regions that have weak or no density due to flexibility were removed from the model. We only traced the backbone for INTS1(357-866) because this region lacks clear side-chain density. For modeling of INTS1(1360-1766), chain A from PDB 7CUN (15) was rigid body fitted into the INTS1(1360-1750) focused map. The N-terminal part until residue 1598 was built *de novo* to correct for a register shift in the published model. The regions C-terminal of residue 1610 were manually adjusted in Coot. Region 1770 to 2042 was removed from the model as it was moved according to our density. The very C-terminal part of INTS1 (residue 2043-2190) was adjusted in Coot and a few amino acids were added to the C-terminus using the focused map for INTS5-8 that included this part of INTS1. To build models for INTS2, INTS4 C-terminus (608-959), INTS5, INTS7 and INTS8, chains B, D, E and H of the published Integrator-PP2A structure (PDB 7CUN) (15) were used as input models. INTS2 (chain B), INTS4(608-959) (part

of chain D) and INTS7 (chain G) were fitted into the focused map for INTS1-2-7. INTS5 (chain E) and INTS8 (chain H) were fitted into the focused map for INTS5-8 in UCSF Chimera (56). The input models were manually adjusted using the corresponding maps in Coot. Major rebuilding was performed for INTS2, INTS5, INTS8 and part of INTS4(608-959) due to many differences in residue registers according to our maps. The INTS7 model was extended at the C-terminus after the β -sandwich domain to include the newly observed helix α 40. The INTS6 model was built by placing INTS6 (chain F) from PDB 7CUN into the INTS6 focused map using UCSF Chimera followed by manual adjustments in Coot to fit the density. To build INTS9 and INTS11, we used the cleavage module structure (PDB 7BFP) (19) as input model. Chains were individually fitted into the corresponding focused map and manually adjusted. Minor adjustments were made to INTS9. Due to conformational changes in the β -CASP domain of INTS11 compared to the input model, we manually rebuilt INTS11 residues 208-350. INTS4 (35-585) was modeled by initially placing chain D from PDB 7CUN into the CM focused map. The N-terminal region (residues 35-322) was manually adjusted in Coot. Residues 376-585 were built *de novo* using Coot.

To build the PP2A-A and PP2A-C models, chain D and F of the PP2A holoenzyme (PDB 3DW8) (62) were fitted into the focused map for PP2A and adjusted as described above. The PP2A-A initial model was divided into two parts, an N-terminal (residues 1-375) and a C-terminal part (residues 376-end).

Modelling of the two observed Pol II CTD repeats was carried out *de novo* using the INTS1-2-7 focused map. Due to the pseudosymmetric nature of the CTD heptad repeats, we could not assign the directionality of the polypeptide chain with absolute certainty, but chose the directionality that fits the density better. The CTD repeats was assigned to a new chain, chain M.

Each model was real-space refined against the corresponding focused map using phenix and showed good stereochemistry. We merged the various refined models in the composite map to create the final model which was subjected to manual inspection using Coot and real-space-refinement in phenix. The model was validated by Molprobit (63) and EMringer (64) and showed good stereochemistry (**Table S1**). Figures were prepared using Pymol 2.4.1, UCSF Chimera and UCSF ChimeraX (65).

Crosslinking-mass spectrometry analysis

The peak fractions of the PEC-Integrator-PP2A glycerol density gradient (fractions 13-16) containing ~45 pmol of PEC-Integrator-PP2A complex were pooled and crosslinked with 3 mM BS3 at 30 °C for 30 min. Following quenching by 100 mM Tris-HCl (pH 8.0), the sample was pelleted by ultracentrifugation in an S100AT4 rotor (Thermo Fisher Scientific). The pellet was solubilized in 50 mM ammonium bicarbonate (pH 8.0) supplemented with 4M Urea, reduced with DTT and alkylated with 17 mM iodoacetamide. After dilution to 1 M urea with 50 mM ammonium bicarbonate, crosslinked complex was digested with trypsin (Promega) in a 1:30 enzyme-to-protein ratio (w/w) at 37 °C overnight. Peptides were reverse-phase extracted using SepPak Vac tC18 1cc/50mg (Waters) and eluted with 50% acetonitrile (ACN) / 0.1% trifluoroacetic acid (TFA). The eluate was split in two halves and vacuum dried. Dried peptides were dissolved in

either 40 μ l 10 mM ammonium hydroxide and separated on basic pH Reverse Phase (bRP) using an xBridge C18 3.5 μ m 1x150mm column (Waters) or in 40 μ l 30% ACN / 0.1% TFA and subjected to peptide size exclusion chromatography (pSEC, using a Superdex Peptide PC3.2/300 column, GE Healthcare) at a flow rate of 50 μ l/min. Fractions of 50 μ l were collected. Fractions containing the crosslinked peptides were vacuum dried and dissolved in 5% ACN / 0.1% TFA for subsequent uHPLC-ESI-MS/MS analysis. Crosslinked peptides enriched by pSEC were analysed as technical triplicates on a Thermo Orbitrap Exploris mass spectrometer. bRP fractionated peptides were measured twice on an Orbitrap Exploris with and without FAIMS (Thermo Scientific) installed. In both cases the mass spectrometer was coupled to a Dionex UltiMate 3000 uHPLC system (Thermo Scientific) with a custom 30 cm C18 column (75 μ m inner diameter packed with ReproSil-Pur 120 C18-AQ beads, 1.9 μ m pore size, Dr. Maisch GmbH). MS1 and MS2 resolution were set to 120000 and 30000, respectively. Only precursors with a charge state of 3-8 were included. Protein-protein crosslinks were identified by pLink2.3.9 search engine (pfind.ict.ac.cn/software/pLink) and filtered at false discovery rate (FDR) of 1% according to the recommendations of the developer(66, 67).

RNA cleavage assay

All concentrations refer to the final concentration used in the assay. The PEC template and non-template DNA scaffolds (22) were used in the assay. In addition to the HIV TAR RNA used for structural analysis (**fig. S10A**), we also used a single-stranded RNA that anneals to the PEC DNA template (**Fig. 2C** and **fig. S10B-F**) for the cleavage assay. RNAs were purchased from IDT with a 5' 6-FAM fluorescent label. To test RNA cleavage activity of free Integrator-PP2A complex, the annealed nucleic acid scaffolds (30 nM) were incubated with the wild-type Integrator-PP2A complex (lane 2) or Integrator(E203Q)-PP2A complex (lane 3) that contains the INTS11E203Q mutant. The Integrator-PP2A complexes were formed by incubating 150 nM of the Integrator subcomplexes and 150 nM PP2A on ice for 1 hr.

To test the activity of Integrator in the context of a Pol II elongation complex (EC), we formed a Pol II EC using 150 nM of Pol II, 30 nM of annealed RNA-template DNA and 30 nM non-template DNA. We added Integrator-PP2A complex (lane 4) or Integrator(E203Q)-PP2A complex (lane 5) which were preformed using 150 nM of each Integrator subcomplex and PP2A, and 150 nM of DSIF and NELF. The assay was also performed using only Integrator-PP2A complex (lane 8) or in the absence of DSIF (lane 6) or NELF (lane 7). All cleavage reactions were performed in 20 μ l final volume in a buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, 3 mM MgCl₂, 1 mM DTT, 1 U/ μ l RNasin® plus (Promega). The reactions were incubated at 30 °C for 1 hr. Reactions were stopped by adding 20 μ l of 2x stop buffer (6.5 urea, 50 mM EDTA, 1x TBE) and boiling at 70 °C for 2 minutes. Proteins were digested using Proteinase K (Thermo Scientific). 8 μ l of each reaction was applied to a denaturing gel (8 M urea, 1x TBE, 20% Bis-Tris acrylamide 19:1 gel run in 1x TBE buffer at 300V for 80 min). Products were visualized using the 6-FAM label and a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences). The same gels were stained with SYBR gold to visualize the microRNA marker (NEB catalogue #N102S).

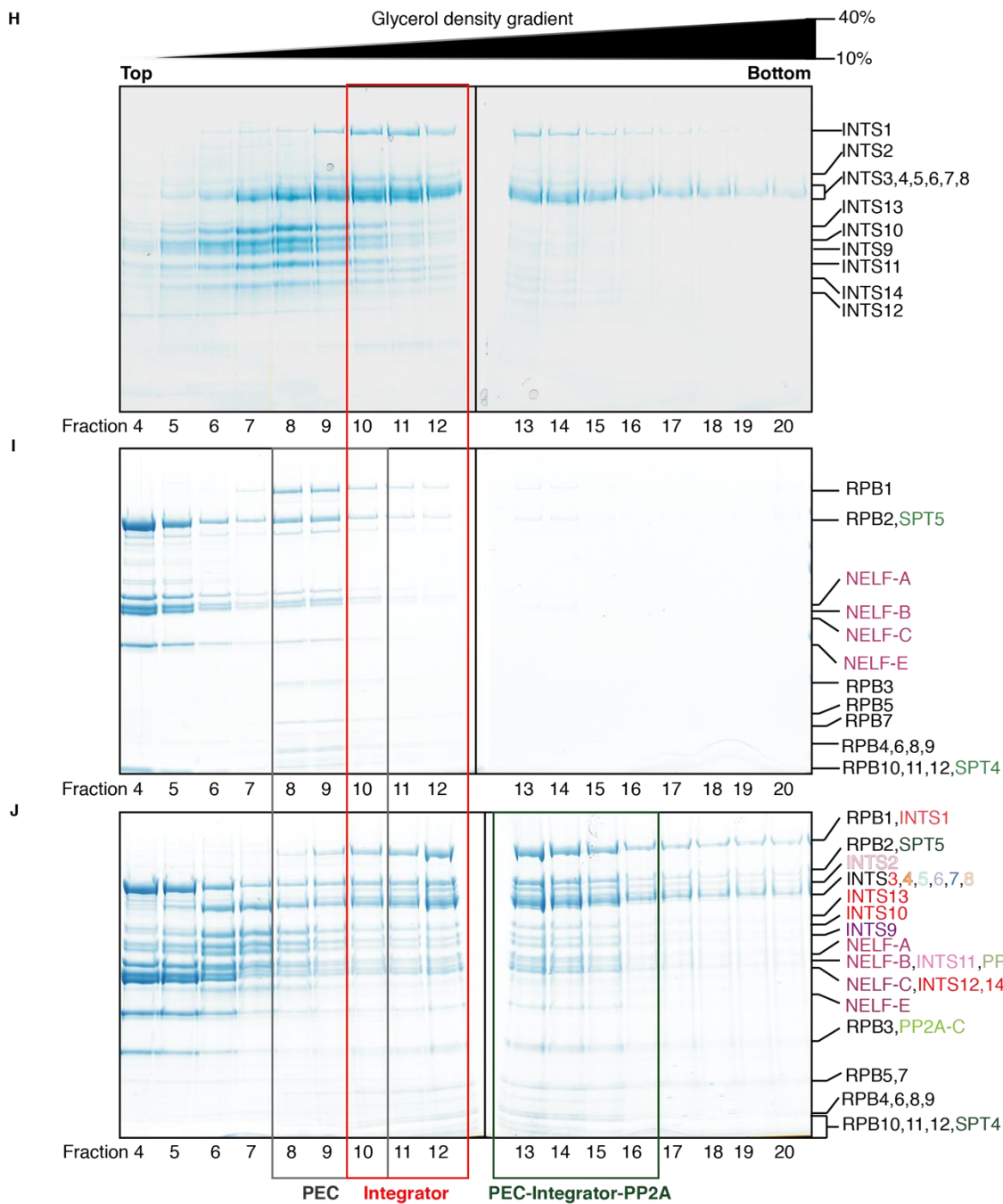
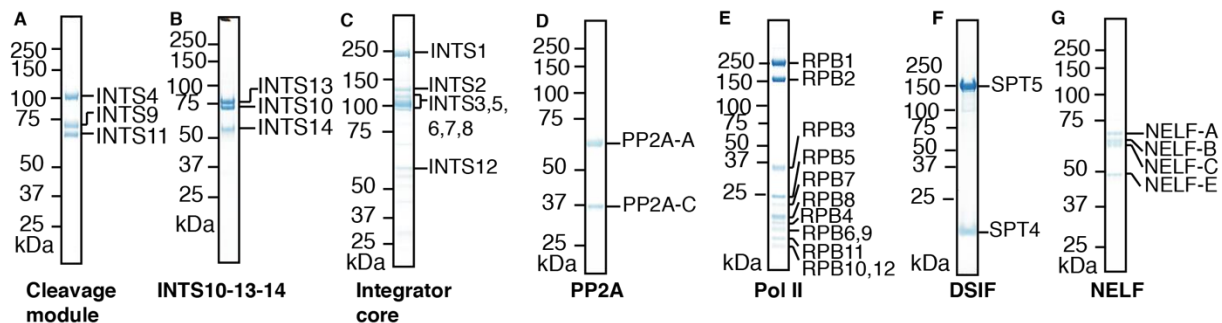


Fig. S1. Preparation of proteins and reconstitution of complexes

Representative Coomassie stained SDS-PAGE analysis of proteins used in this study are shown for (A) the Integrator cleavage module, (B) the INTS10-13-14 module of Integrator, (C) the Integrator core, (D) the PP2A phosphatase, (E) Pol II, (F) DSIF and (G) NELF. (H) SDS-PAGE analysis of fractions from a glycerol density gradient shows the formation of Integrator from the three subcomplexes shown in A-C. Fractions analyzed are indicated at the bottom of the gel. For this gradient, Integrator core, cleavage module and INTS10-13-14 module were mixed in a 1:2:2 molar ratio. (I) SDS-PAGE showing the formation of PEC from Pol II (E), DSIF (F) and NELF (G). (J) SDS-PAGE showing formation of PEC-Integrator-PP2A. Identity of proteins in the peak fraction was confirmed by mass spectrometry.

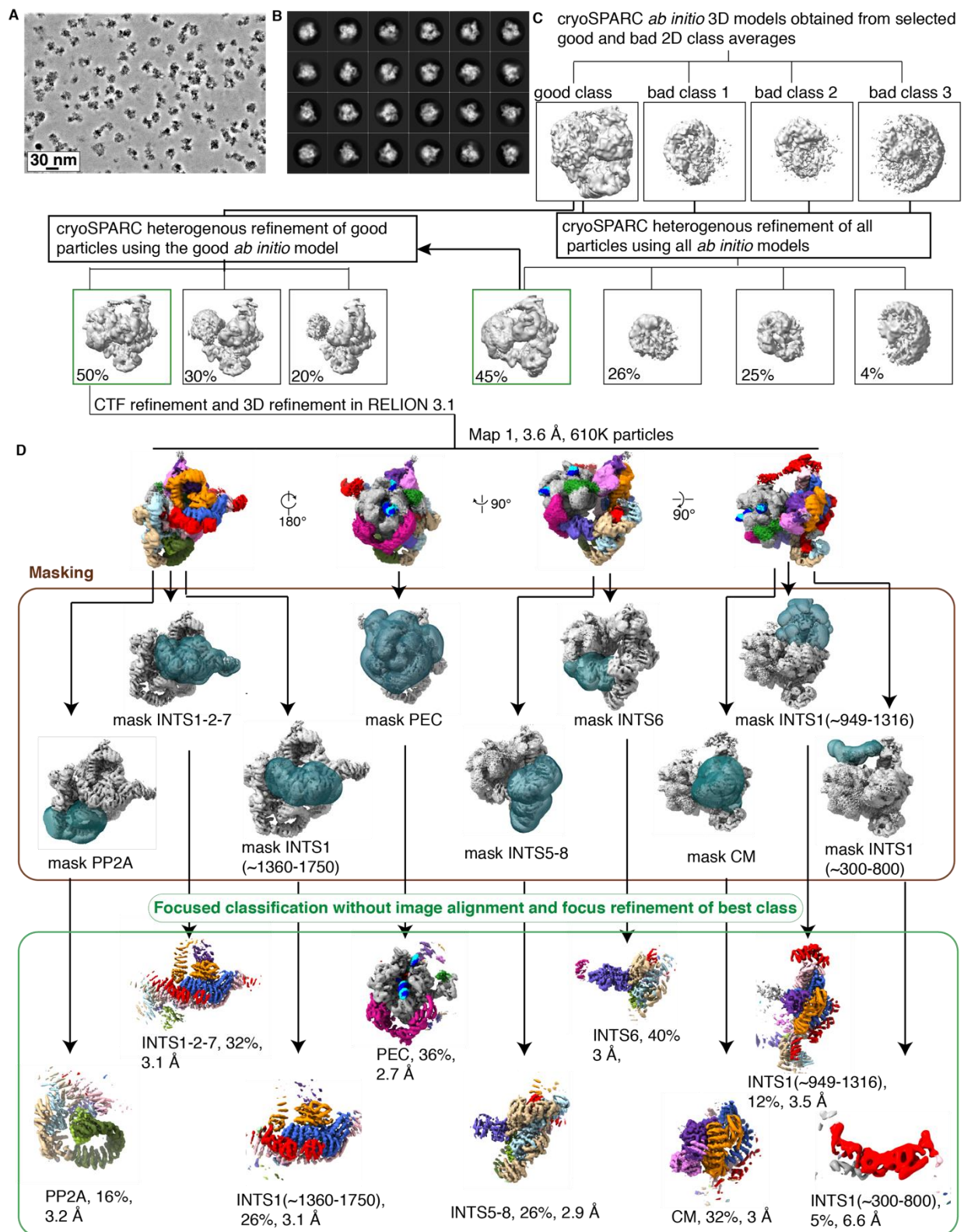


Fig. S2. Cryo-EM data processing

(A) A representative micrograph denoised using Warp (52) showing distribution and concentration of particles per field of view. A 30 nm scale bar is shown. (B) Selected good 2D class averages showing different views of the cryo-EM reconstruction. (C) Initial clean-up of the dataset in cryoSPARC (55). Four *ab initio* 3D models generated from the dataset were used as references for heterogenous refinement. The selected good particles were subjected to another round of heterogenous refinement which yielded ~610K particles that have density for both Integrator-PP2A and PEC. (D) Processing of the selected good particles from cryoSPARC in RELION 3.1 (53, 54). Selected particles were subjected to CTF refinement and 3D refinement resulting in a 3.6 Å reconstruction (Map 1). Various regions of the reconstruction were masked for 3D classification without image alignment and 3D refinement. This produced well-resolved focused maps that were used for initial modeling. Focused map names are based on the dominant subunits present and colored with the same color code as in **Fig. 1**. CM, Integrator cleavage module. Note that only the selected good class of each focused classification is shown. The percentage of the input 610K particles that went into the selected good class is indicated as well as the estimated resolution.

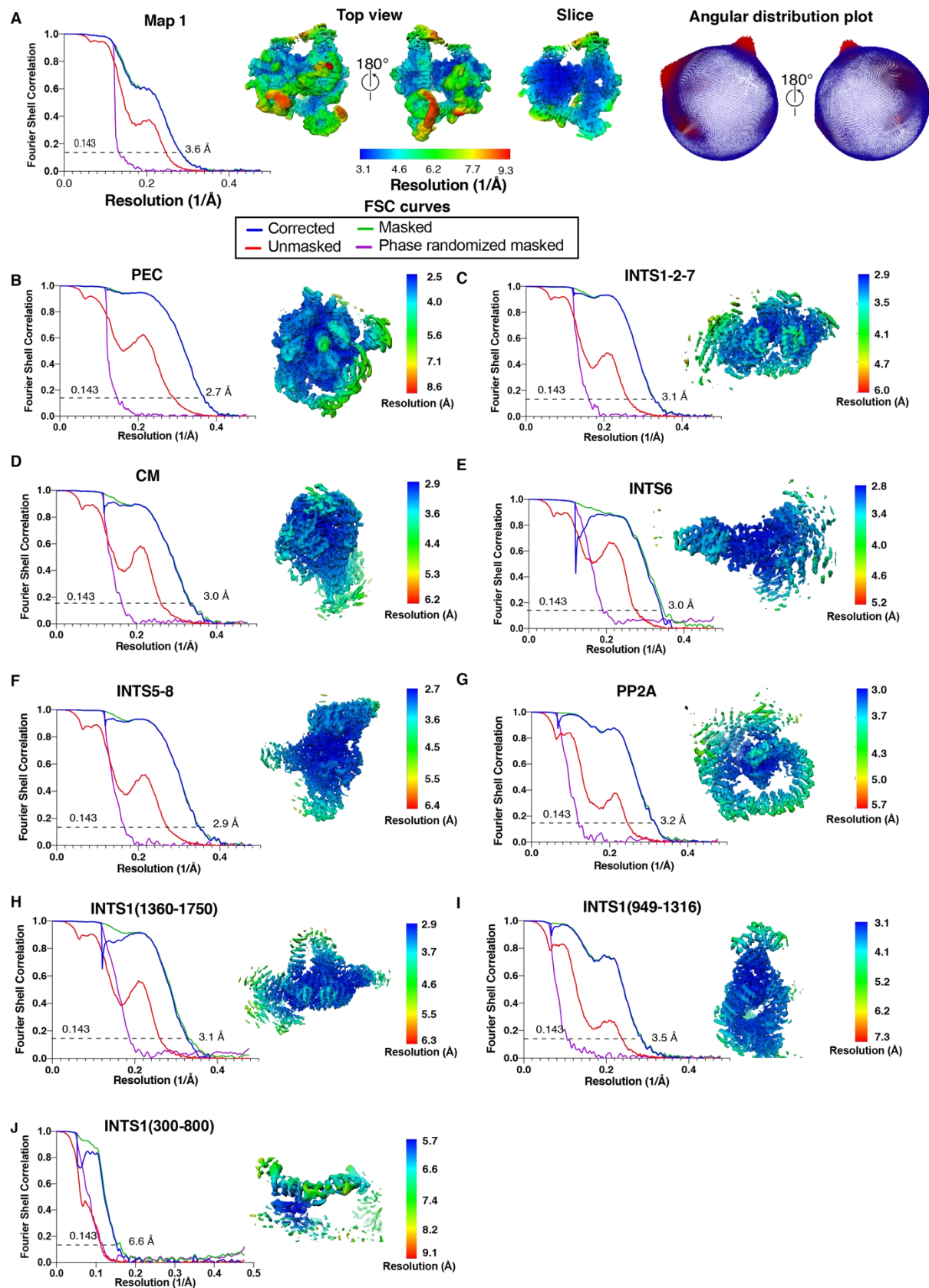


Fig. S3. Resolution of cryo-EM reconstructions

(A) Resolution of Map 1 (**fig. S2**) and angular distribution plot for the reconstruction. FSC plot shows the overall resolution of the reconstruction as determined by the ‘gold standard’ FSC cut-off of 0.143 which is indicated by a dashed line. Two views show the local resolution (as implemented in RELION 3.1) of regions on the surface and a slice through shows resolution inside the core of the reconstruction. A scale bar indicates the resolution represented by the various colors. An angular distribution plot of the reconstruction showing assignment of particles to the various angles is also provided. (B-J) Show FSC plots and local resolutions of the focused maps. CM, Integrator cleavage module

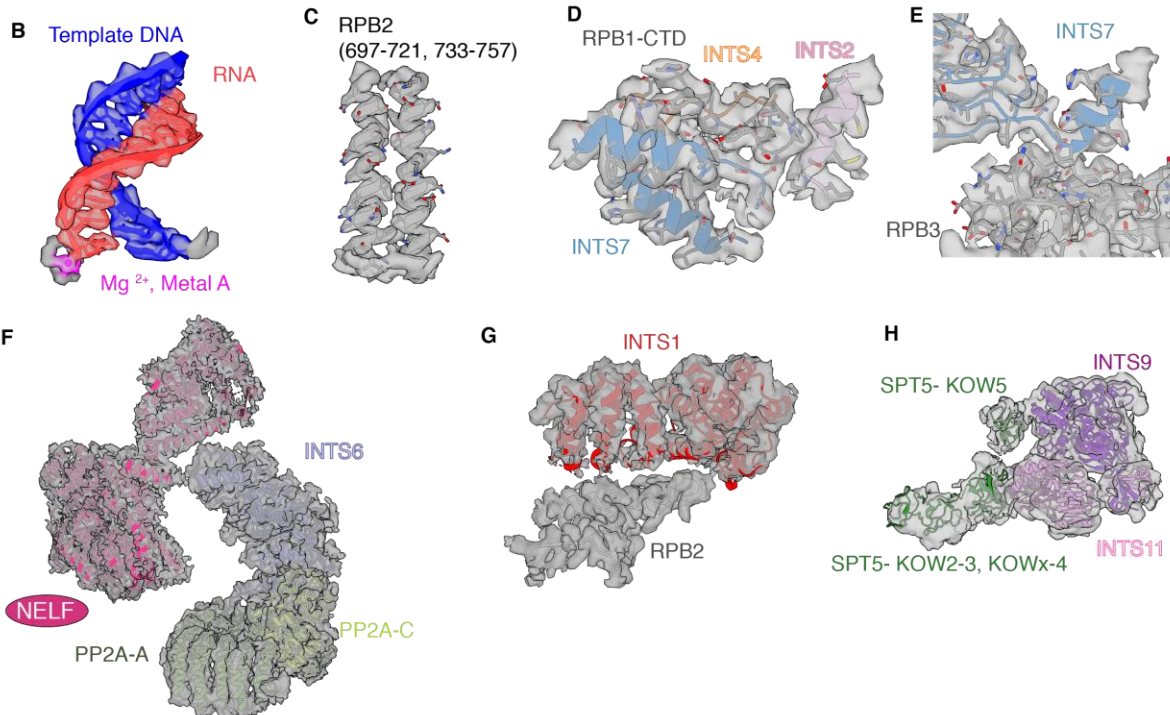
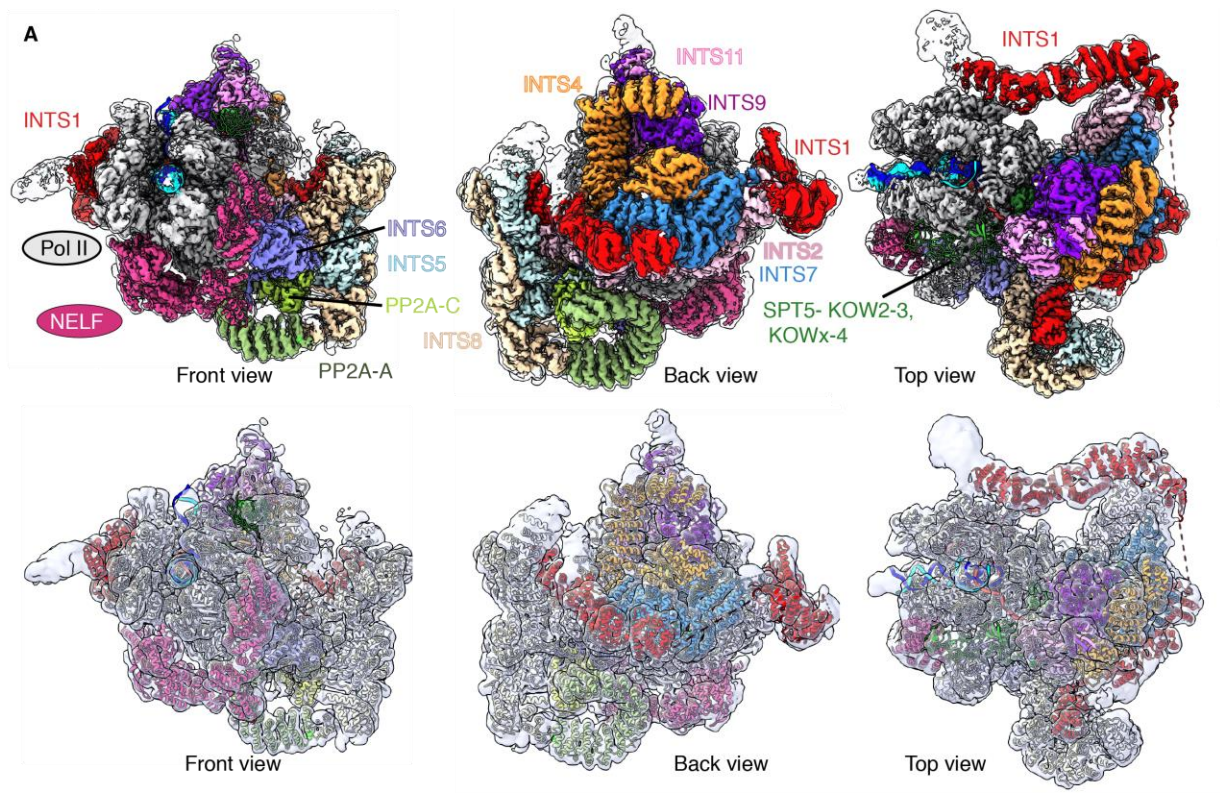


Fig. S4. Quality of the cryo-EM densities.

(A) Overview of the PEC-Integrator-PP2A model relative to the cryo-EM reconstruction. In the top panel the composite map used for modelling is shown fitted into the 6 Å low-pass filtered version used for docking. The lower panel shows our model fitted into the 6 Å low-pass filtered version of the composite map used for docking. Left panel, front view; middle panel, back view; right panel, top view. The views correspond to those in **Fig. 1B-C**. The color code of the model is the same as in **Fig. 1**. (B) Tilted DNA-RNA hybrid in the Pol II active center. The map for the DNA-RNA hybrid and the catalytic metal ion (metal A) is shown as surface representation. DNA template strand (blue) and RNA transcript (red) are shown as sticks. Metal A (magenta) is shown as a sphere. (C) Cryo-EM density for Pol II RPB1 funnel helices. RPB1 is shown as ribbon and sticks (side-chains). (D) Cryo-EM density of the region around the two observed CTD heptad repeats. INTS2, light pink; INTS4, orange; INTS7, sky blue; RPB1-CTD, gray. Proteins are shown as ribbon and sticks (side-chains). (E) Cryo-EM density at the RPB3 (gray)-INTS7 (sky blue) interface. Proteins are shown as ribbon and sticks (side-chains). (F) Cryo-EM density of NELF (warm pink)-INTS6 (slate) interface. INTS6 forms a bridge to PP2A (PP2A-A, smudge; PP2A-C, limon). (G) Cryo-EM density at the RPB2 (gray)-INTS1 (red) interface. (H) Low-pass filtered (6 Å) cryo-EM density for SPT5 (forest green) interfaces with INTS11 (violet) and INTS9 (purple blue).

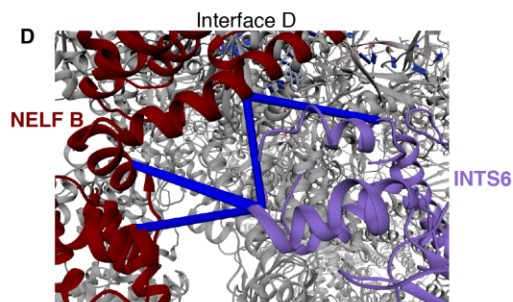
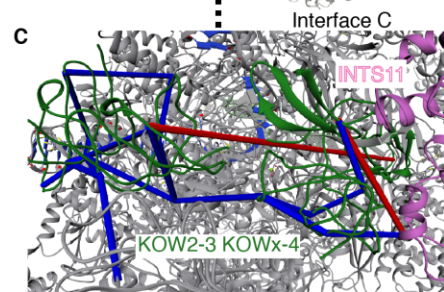
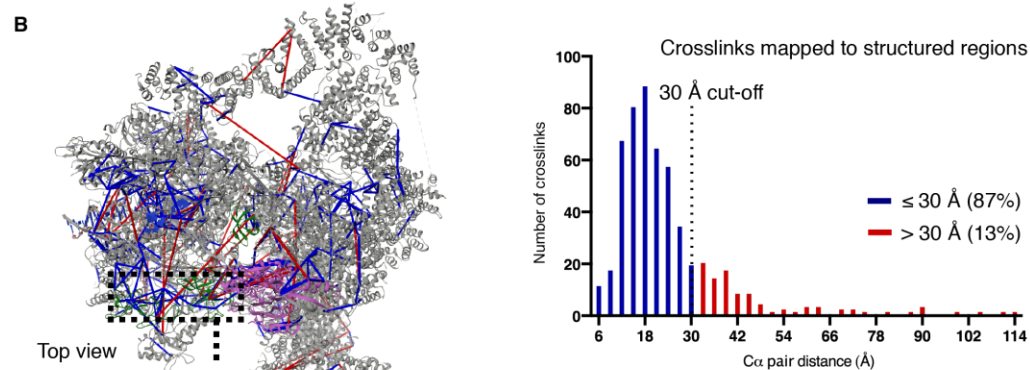
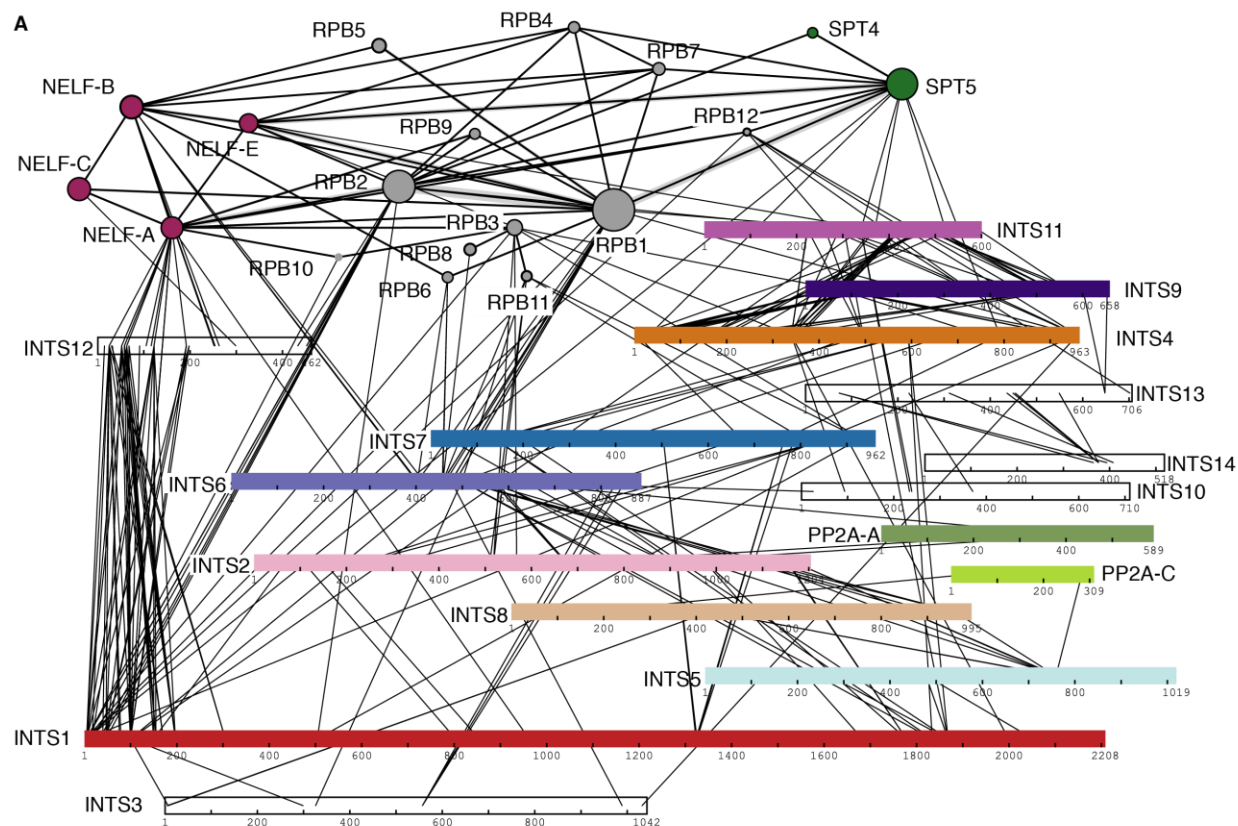


Fig. S5. Crosslinking-mass spectrometric analysis.

(A) Overview of BS3 mediated crosslinks visualized using xiNET (68). Subunits are colored according to Fig.1. For clarity, only inter-protein crosslinks that have at least eight crosslinked peptide-spectrum matches (CSMs) are shown. Note distinctive crosslinks between the N-termini of INTS1 and INTS12, INTS12 and NELF-A, INTS10-13-14 to subunits of the cleavage module.

(B) Mapping of crosslinks on our PEC-Integrator-PP2A structure using the Xlink analyzer (69) plugin in UCSF Chimera. The histogram shows distribution of C α pair distances of unique crosslinks mapped on our structure. The y-axis shows the number of unique crosslinked peptides with the specific C α pair distances shown on the x-axis. Estimation of C α pair distances were done using Xlink analyzer (69) plugin in UCSF Chimera. A dotted line demarcates the 30 Å distance cut-off for BS3 crosslinked C α pair. ~87% of the unique crosslinks are within the favored distance.

(C) Crosslinks from SPT5 KOW2-3 and KOWx-4 domains that support formation of interface C.

(D) Crosslinks between INTS6 and NELF-B that support formation of interface D are shown.

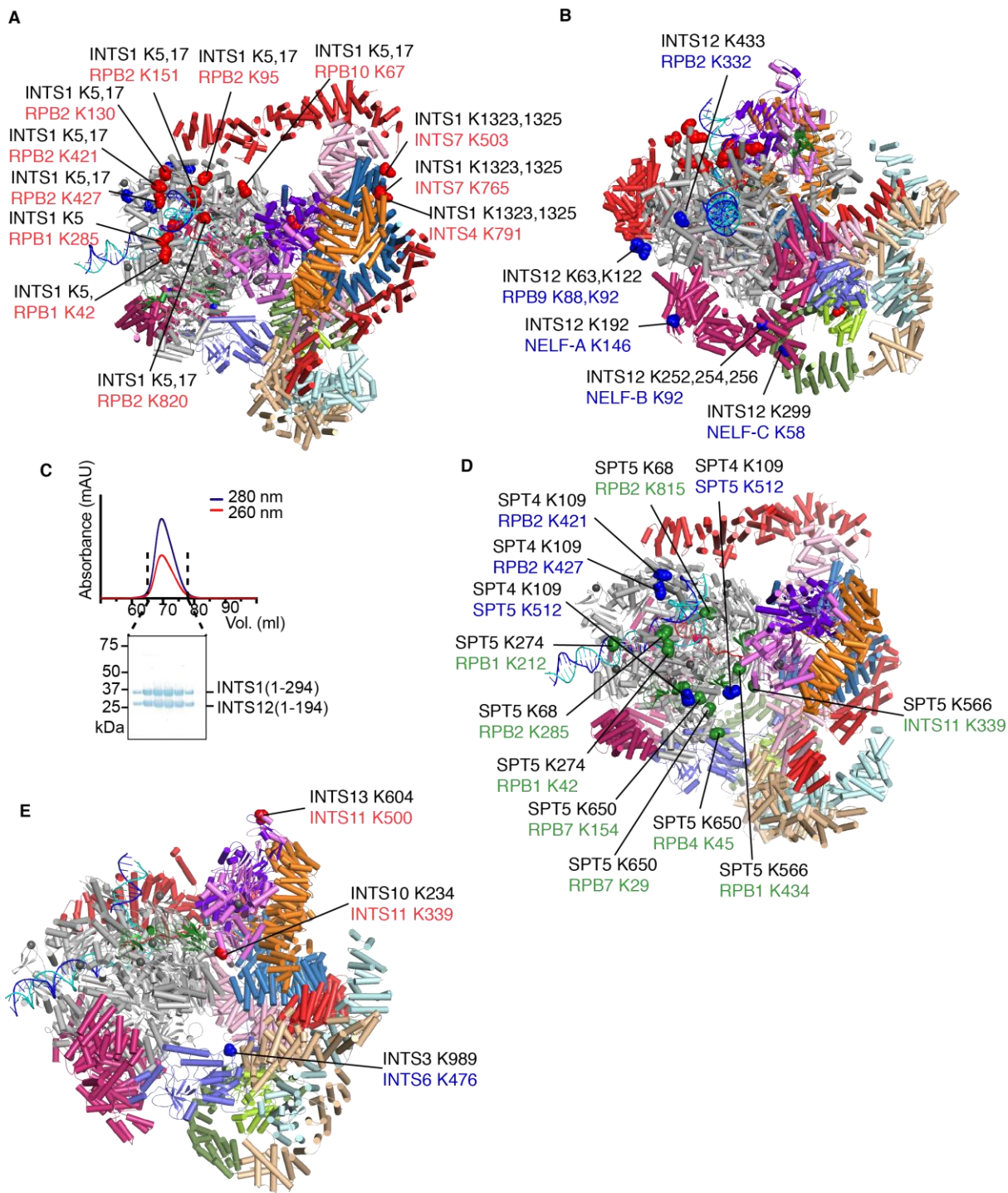


Fig. S6. Localization of mobile subunits/regions using XLMS. The specifically crosslinked lysine residues in our structure are shown as spheres and the crosslinked lysine residues in the mobile region indicated. **(A)** The crosslink sites of the mobile regions of INTS1 on our structure are shown as red spheres. **(B)** The crosslink sites of INTS12 on our structure are shown as blue spheres. **(C)** A size exclusion chromatogram and SDS-PAGE confirms the physical interaction between INTS1(1-292) and INTS12(1-194) as predicted by the XLMS data. **(D)** Mapping of crosslink sites of the flexible parts of DSIF on our structure. SPT4 and SPT5 crosslink sites are shown as blue and green spheres, respectively. **(E)** Mapping of INTS3 (blue sphere) and INTS10-13-14 (red spheres) crosslink sites on our structure.

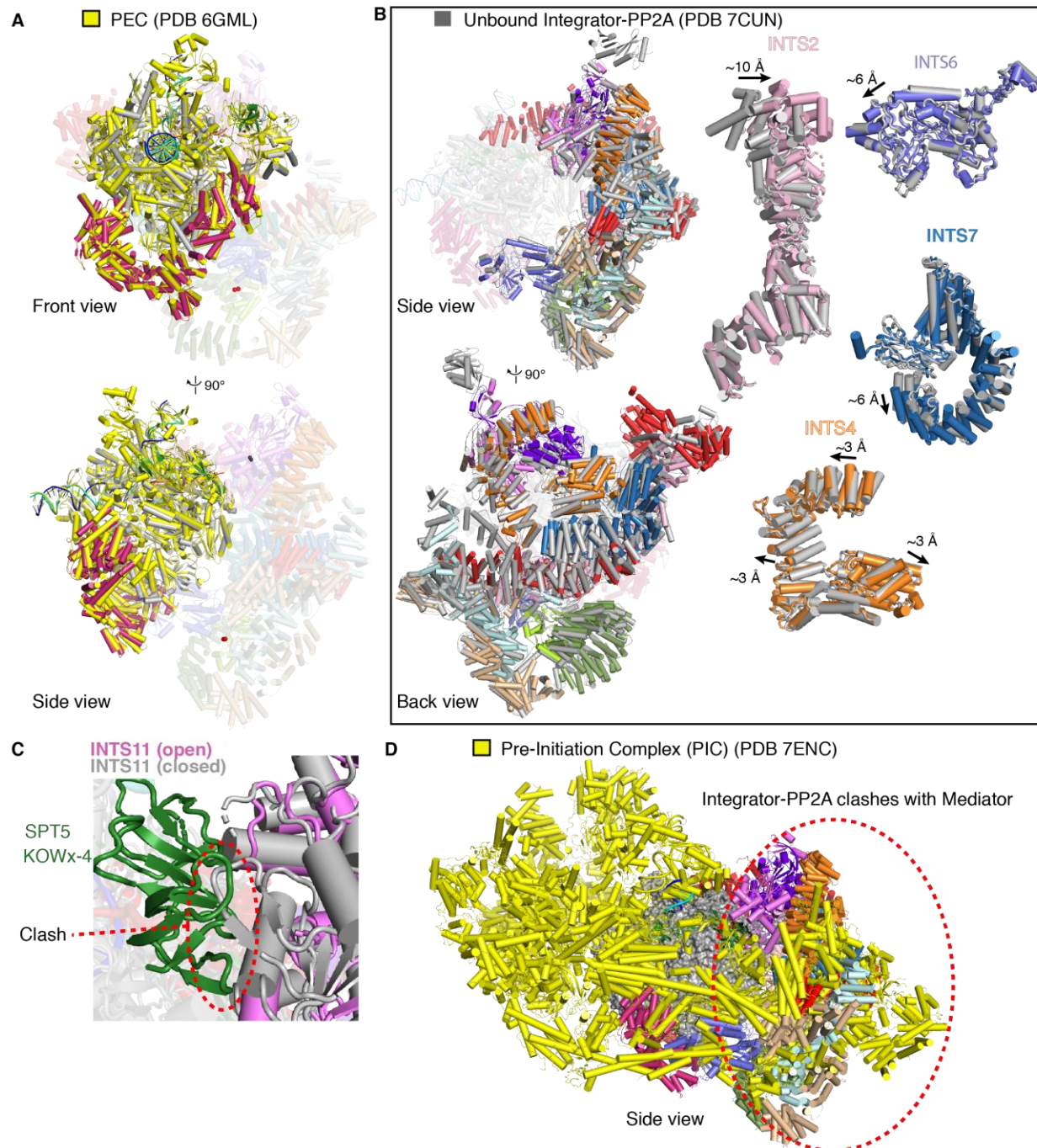


Fig. S7. Structural comparisons. (A) Alignment of the PEC structure (PDB 6GML) (22) onto our PEC-Integrator-PP2A structure reveals no changes in Pol II and only minor conformational changes in NELF regions that bind INTS6. The PEC-Integrator-PP2A structure is colored as in **Fig. 1** and PEC (PDB 6GML) is colored yellow. For clarity, Integrator-PP2A is shown as a transparent cartoon. (B) Alignment of unbound Integrator-PP2A structure (PDB 7CUN) (15) shown as gray onto our PEC-Integrator-PP2A structure shows conformational changes of several elements. Superpositions of INTS2, INTS4, INTS6 and INTS7 subunits of the unbound and PEC-bound Integrator-PP2A are also shown. The arrows show the directions of movements upon PEC

binding. **(C)** A zoom-in view shows that INTS11 in the closed conformation (19) would clash with SPT5 KOWx-4. The view is the same as in **Fig. 2C**. **(D)** Alignment of our PEC-Integrator-PP2A structure onto the structure of Mediator-bound transcription pre-initiation complex (PDB 7ENC; yellow) (70), shows binding of transcription initiation factors and Mediator exclude Integrator binding.

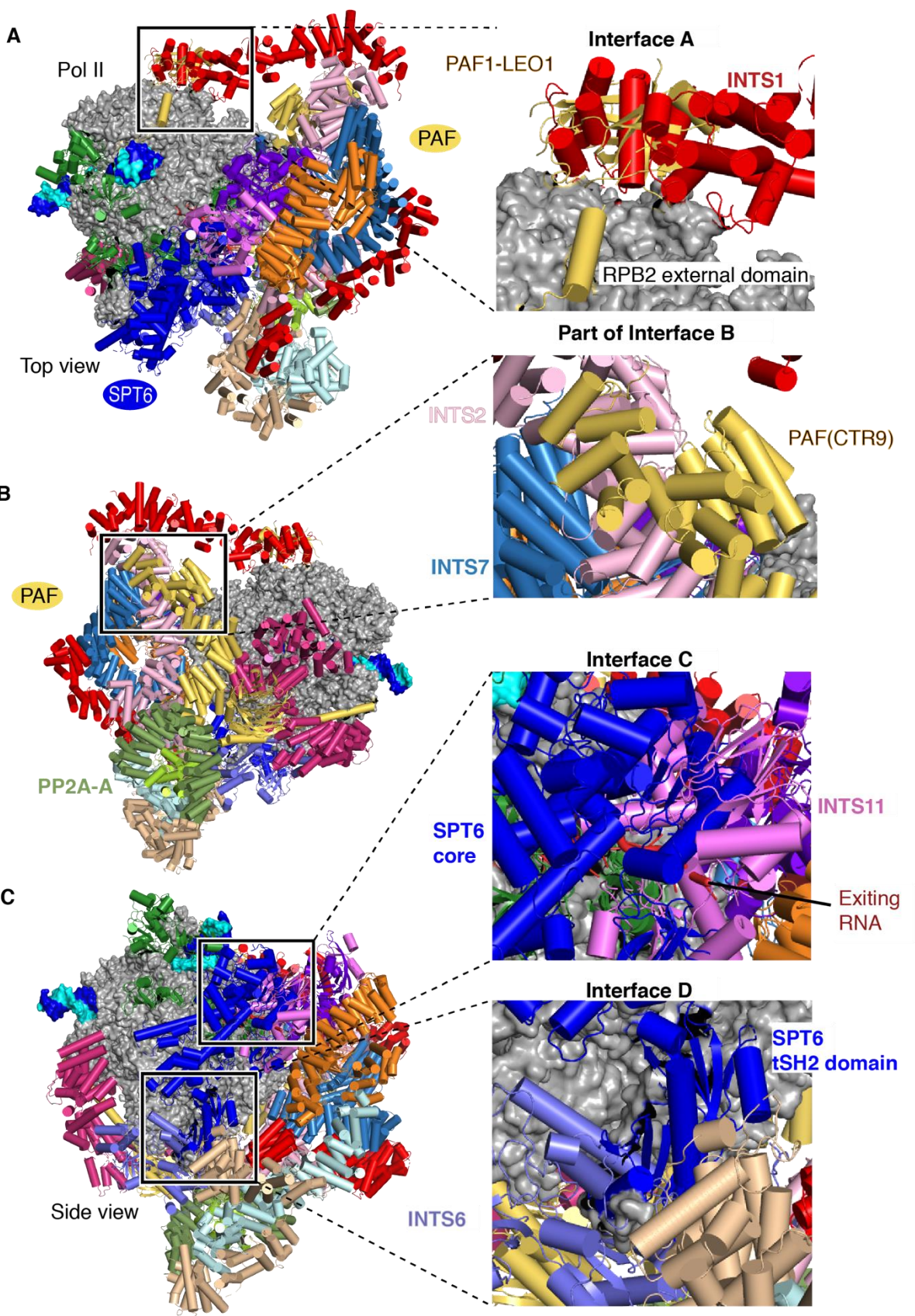


Fig. S8. Integrator excludes elongation factors of the activated elongation complex EC*

(A-C) Comparison of our structure to the activated elongation complex (EC*) structure (PDB 6GMH) (37) shows clashes of PAF and SPT6 with Integrator subunits at all the interfaces between PEC and the Integrator-PP2A complex. Pol II is shown as a gray surface, Integrator-PP2A is colored as in **Fig. 1**, SPT6 is colored blue, and the PAF complex is colored yellow orange. Close-up views showing the clashes are provided in the right panel.

A INTS1

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

343 IDNVSRNLLRLLTSTCGYKEVRLLAVQKLEMMWQONPKLTRPAQDLL
345 IDNVSRNLLRLLTATCGYKEVRLLAVQKLEMMWQONPKLTRPAQDLL
343 MDNVSRNLLRLLTATCGYKEVRQMAVQKLEMMWQONPKLTRPAQDLL
343 MDNVSRNLLRLLTATCGYKEVRLLMAVQKLEMMWQONPKLTRPAQELL
341 MDNVSRNLLRLLTATCGYKEVRLLMAVQKLEMMWQONPKLTRPAQDLL
320 QDNYTRNFKFLCTTSGIAEVRCLISIRLELWIHNGKLVKFAQQLL
302 .EVLPKPLFRALCLLCGYEHIRLFIANRLEIWTLOKFHREISELL

INTS1

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

389 MSVCMNCNTHGSE.DMDVISHLIKI...FLKPKVLIINHFMLCIREL
391 MSVCMNCNTHGSE.DMDVISHLIKI...FLKPKVLIINHFMLCIREL
389 MSVCMNCNTHGSE.DMEVISNLIKI...FLKPKVLIINHFMLCVREL
389 MSVCMNCNTHGPE.DMEVISNLIKI...FLKPKVLIINHFMVREL
387 MSVCMNCNTHGSD.DMEVISNLIKI...FLKPKVLIINHFMVREL
366 SYICFNINIKGRNTQ.DNEVLLVLVKM...FLKPKVLIINHFMVREL
347 LVIGSNINAEENHELDKEIIDTLKLNNTKNMKTRTVVGPFTQAIKTI

INTS1

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

431 LSAHKDNLGTTIKLVIFNELSNARN..PNNMQVLYTALQHSSELAP
433 LNAHKDNLGTTIKLVIFNELSNARN..PNNMQVLYTALQHSSELAP
431 LNAHRDNLGTTIKLVIFNELSNARN..PNNMQVLYTALQHSSELAP
431 LNAHRDNLGTTLVKFVIFNELSNARN..PNNMQVLYTALQHSSELAP
429 LNAHRDNLGTMVKFVIFNELSNARN..PNNMQVLYTALQHSSELAP
408 IFLOPEILSTVMKLVVQNELSNTRN..PNNMGLATMTFQTSADQSA
393 CEVSR.NVEIVLDSLFINETGPAQNRSPTN.QTIYMILQNPTEAT

INTS1

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

475 KFLAMVF.QDLLTNKDDYLRAFRALLRETIKQTKHEIN....FQAF
477 KFLAMVF.QDLLTNKDDYLRAFRALLRETIKQTKHEIN....FQAF
475 KYLAMVF.QDLLTNKDDYLRAFRALLRETIKQTKHEIN....FQAF
475 KFLAMVF.QDLLTNKDDYLRAFRALLRETIKQTKHEIN....FQAF
473 KFLAMVF.QDLLTNKDDYLRAFRALLRETIKQTKHEIN....FQSF
452 ATLAETIY.QEFLQRDDCLRTIRVFLRELVRMLRFQVN....LVKF
438 KALAVIISKELFKEKENCVKMIRAFMREILIRGFYHSKSGDFPFSVF

RPB2

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

545 LEEWSMENLEEISPAAIADATKIFVNGCWVGIHKDPQOLMNTLRKL
545 LEEWSMENLEEISPAAIADATKIFVNGCWVGIHKDPQOLMNTLRKL
545 LEEWSMENLEEISPAAIADATKIFVNGCWVGIHKDPQOLMNTLRKL
545 LEEWSMENLEEISPAAIADATKIFVNGCWVGIHKDPQOLMNTLRKL
546 LEEWSMENLEEISPAAIADATKIFVNGCWVGIHRDPQOLMNTLRKL
551 LEEWSMENLEEVSPSAIADATKIFVNGCWVGIHREPQOLMTTLKKI

RPB2

H.sapiens

H.sapiens
M.musculus
G.gallus
X.laavis
D.rerio
D.melanogaster
C.elegans

637 KRHIDQLKER..EINNYWQDLVASGVVEYIDTLEEETVMLAMTPD
637 KRHIDQLKER..EINNYWQDLVASGVVEYIDTLEEETVMLAMTPD
637 KRHIDQLKER..EINNYWQDLVASGVVEYIDTLEEETVMLAMTPD
637 RRHIDQLKER..EINNYWQDLVASGVVEYIDTLEEETVMLAMTPD
637 RRHIDQLKER..EINNYWQDLVASGVVEYIDTLEEETVMLAMTPD
638 KTHVEMLKER..DINNYWQDLVASGVVEYIDTLEEETVMIAMSPY
643 KRHIDQLKEAADENKYNWSDLVGCGVVELIDSMEEETSMIAMPPE

B

INTS7

H.sapiens

H.sapiens

M.musculus

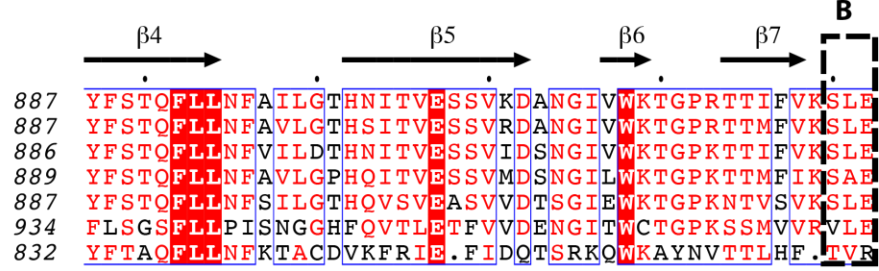
G.gallus

X.laevis

D.rerio

D.melanogaster

C.elegans



INTS7

H.sapiens

H.sapiens

M.musculus

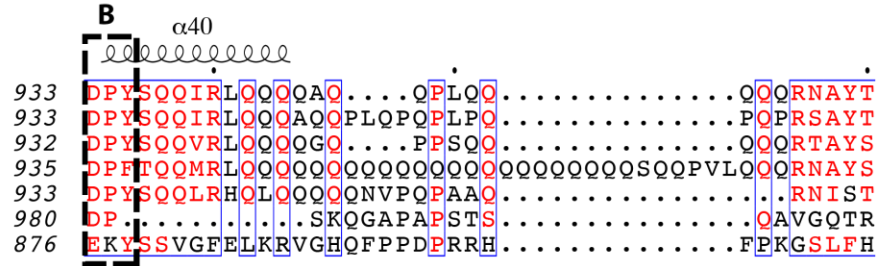
G.gallus

X.laevis

D.rerio

D.melanogaster

C.elegans



RPB3

H.sapiens

H.sapiens

M.musculus

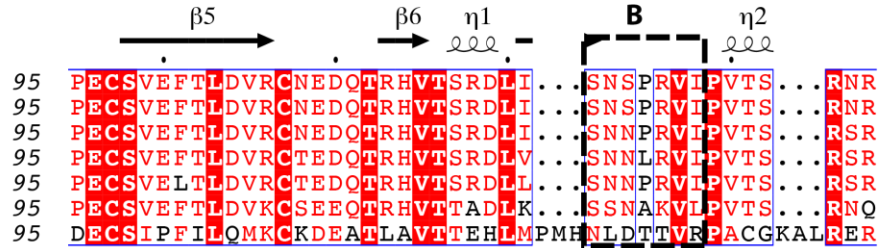
G.gallus

X.laevis

D.rerio

D.melanogaster

C.elegans



C INTS11

H.sapiens

H.sapiens

M.musculus

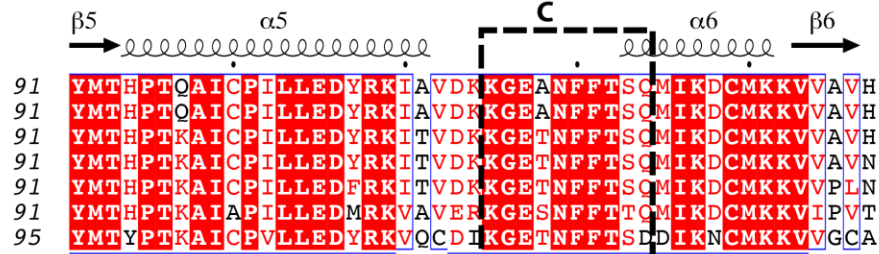
G.gallus

X.laavis

D.rerio

D.melanogaster

C.elegans



INTS11

H.sapiens

H.sapiens

M.musculus

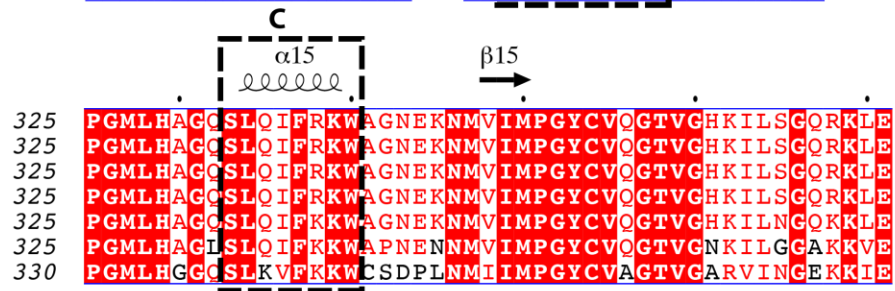
G.gallus

X.laavis

D.rerio

D.melanogaster

C.elegans



SPT5

H.sapiens

H.sapiens

M.musculus

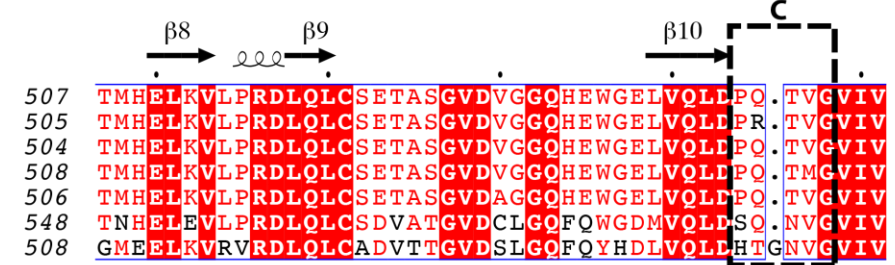
G.gallus

X.laavis

D.rerio

D.melanogaster

C.elegans



SPT5

H.sapiens

H.sapiens

M.musculus

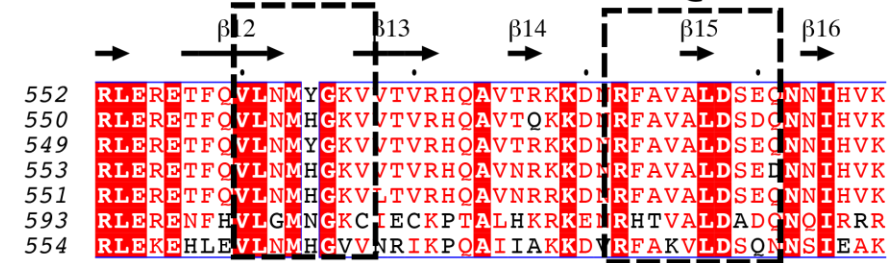
G.gallus

X.laavis

D.rerio

D.melanogaster

C.elegans



D INTS6

H.sapiens

H.sapiens

M.musculus

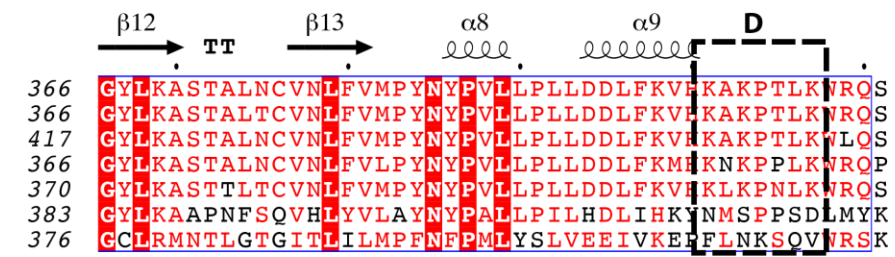
G.gallus

X.laavis

D.rerio

D.melanogaster

C.elegans



NELF-B

H.sapiens

M.musculus

G.gallus

X.laavis

D.rerio

D.melanogaster

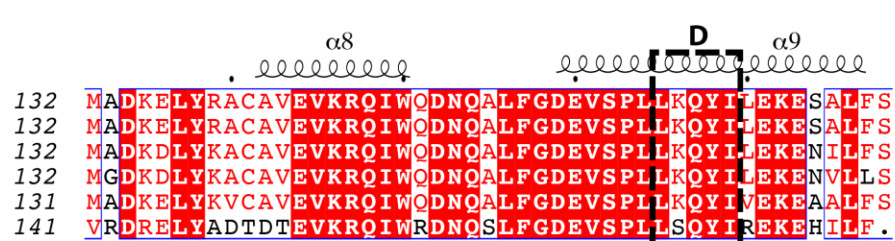


Fig. S9. Sequence alignments of protein regions that for PEC and Integrator interfaces.

The alignments were performed with full-length proteins in Jalview (71) with 'Muscle with defaults' algorithm. Aligned sequences were visualized using ESPript 3.0 (72). Residues with identity above 70% are colored. Regions of each subunit that are involved in contact interfaces are indicated by dashed boxes A-D for interfaces A-D, respectively. **(A)** Sequence alignment of INTS1 and RPB2 regions that interface A. The UniProt IDs for INTS1 are *Homo sapiens* (Q8N201), *Mus musculus* (Q6P4S8), *Gallus gallus* (F1P4X9), *Xenopus laevis* (A0A1L8EXS4), *Danio rerio* (A0A0G2KYX9), *Drosophila melanogaster* (Q9W1C5) and *Caenorhabditis elegans* (O01483). UniProt IDs for RBP2 are *Homo sapiens* (P30876), *Mus musculus* (Q8CFI7), *Gallus gallus* (F1NCB0), *Xenopus laevis* (A0A1L8HW57), *Danio rerio* (A0A0R4IXL0), *Drosophila melanogaster* (A0A0B4LI86) and *Caenorhabditis elegans* (Q10578). **(B)** Sequence alignments of INTS7 and Pol II RPB3 regions that form Interface B. INTS7 UniProt IDs are *Homo sapiens* (Q9NVH2), *Mus musculus* (Q7TQK1), *Gallus gallus* (Q5ZL91), *Xenopus laevis* (A0JMW3), *Danio rerio* (Q8JGR7), *Drosophila melanogaster* (Q9VT41) and *Caenorhabditis elegans* (G5EDW9). RPB3 UniProt IDs are *Homo sapiens* (P19387), *Mus musculus* (P97760), *Gallus gallus* (E1BV97), *Xenopus laevis* (Q6IRR3), *Danio rerio* (Q7ZV11), *Drosophila melanogaster* (O97183) and *Caenorhabditis elegans* (Q93338) **(C)** Sequence alignments of INTS11 and SPT5. INTS11 UniProt IDs for INTS11 are *Homo sapiens* (Q5TA45), *Mus musculus* (Q9CWS4), *Gallus gallus* (Q5ZIH0), *Xenopus laevis* (A0A1L8FM59), *Danio rerio* (Q503E1), *Drosophila melanogaster* (Q9VAH9) and *Caenorhabditis elegans* (Q9U3K2). SPT5 UniProt IDs for INTS11 are *Homo sapiens* (O00267), *Mus musculus* (O55201), *Gallus gallus* (Q5ZI08), *Xenopus laevis* (A1L1H7), *Danio rerio* (Q9DDT5), *Drosophila melanogaster* (Q9V460) and *Caenorhabditis elegans* (Q21338). **(D)** Sequence alignments of INTS6 and NELF-B showing regions that form interface D. The UniProt IDs for INTS6 are *Homo sapiens* (Q9UL03), *Mus musculus* (Q6PCM2), *Gallus gallus* (E1BV07), *Xenopus laevis* (Q5U4W6), *Danio rerio* (Q7SYD9), *Drosophila melanogaster* (Q9W485) and *Caenorhabditis elegans* (B7WN63). The UniProt IDs for NELF are *Homo sapiens* (Q8WX92), *Mus musculus* (Q8C4Y3), *Gallus gallus* (Q5ZLK9), *Xenopus laevis* (Q32NK8), *Danio rerio* (Q6TNV1), and *Drosophila melanogaster* (Q9Y113). Please note the NELF complex is not conserved in *C. elegans* and hence its omission in the sequence alignment

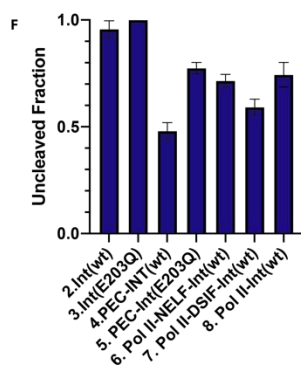
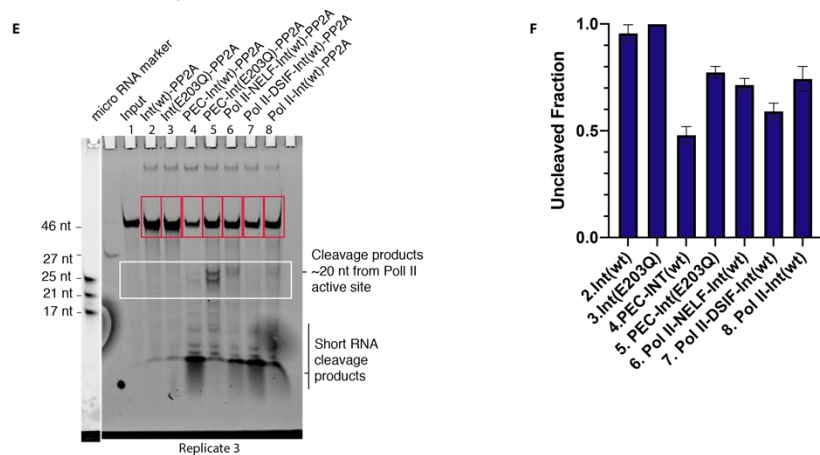
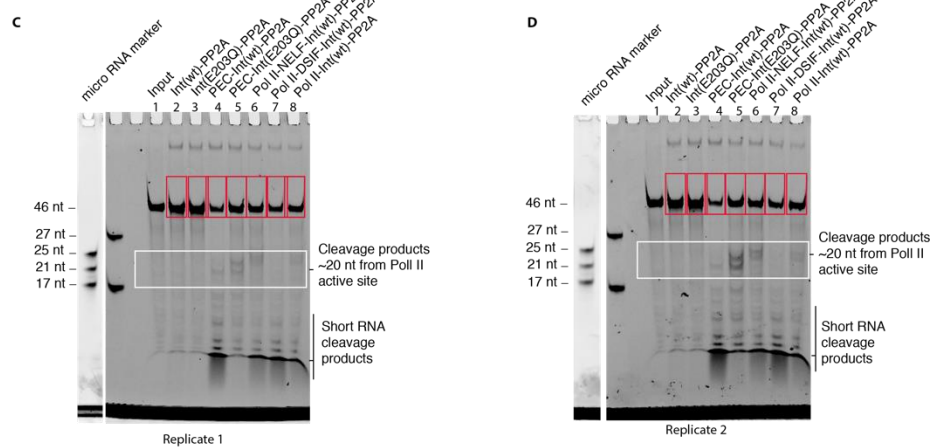
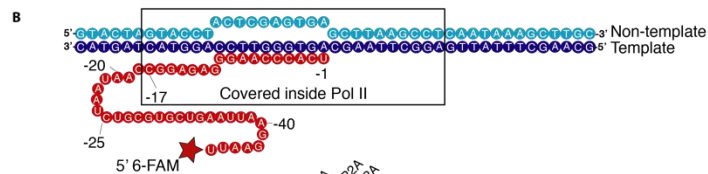
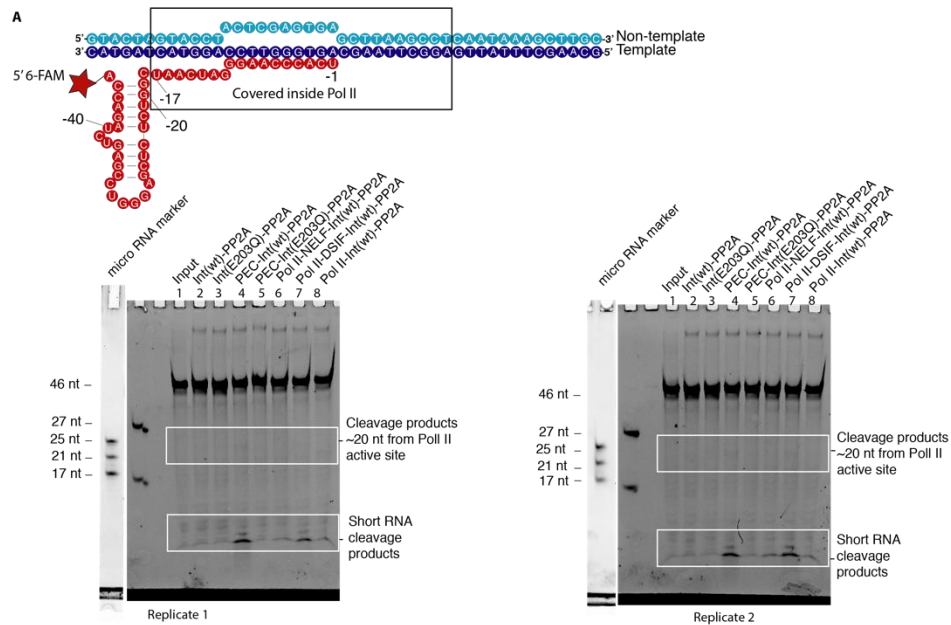


Fig. S10. Integrator RNA cleavage. (A) A schematic representation of the HIV TAR RNA scaffold used for structural analysis and the cleavage assay. Two representative replicates of the RNA cleavage assay performed using this scaffold are provided. After visualization with the 6-FAM label, the same gels were stain with SYBR Gold to visualize the microRNA marker (NEB catalogue #N102S) which is overlaid. A marker containing 6-FAM labeled 27 nt and 15 nt single-stranded RNA was also loaded. (B) Schematic of the single-stranded RNA scaffold that lacks the stem-loop structure. (C-E) representative replicates of the cleavage assay performed using the single-stranded RNA. The red boxes show the regions that were quantified to illustrate Integrator-dependent cleavage of the input RNA. We observe cleavage product of different sizes which is expected because single-stranded RNA that emerges from the RNA exit channel of Pol II is flexible. Therefore, exiting RNA may form loops of different sizes between the point of RNA exit (i.e. the Pol II surface) and the INTS11 endonuclease active site, leading to RNA cleavage products of different lengths. (F) A histogram showing the fraction of the input RNA that was uncleaved for each reaction condition. We quantified the remaining input RNA signal intensities (shown by red boxes) using ImageJ(73) (See table S3 for raw values). After subtracting background signal, we normalized each value by the intensity of the input RNA in the free Int(E203Q) sample that has no detectable activity to determine the uncleaved fraction. Error bars represent standard deviation in the three experiments. Int(wt), wild-type Integrator-PP2A; Int(E203Q), Integrator-PP2A complex with INTS11-E203Q mutation.

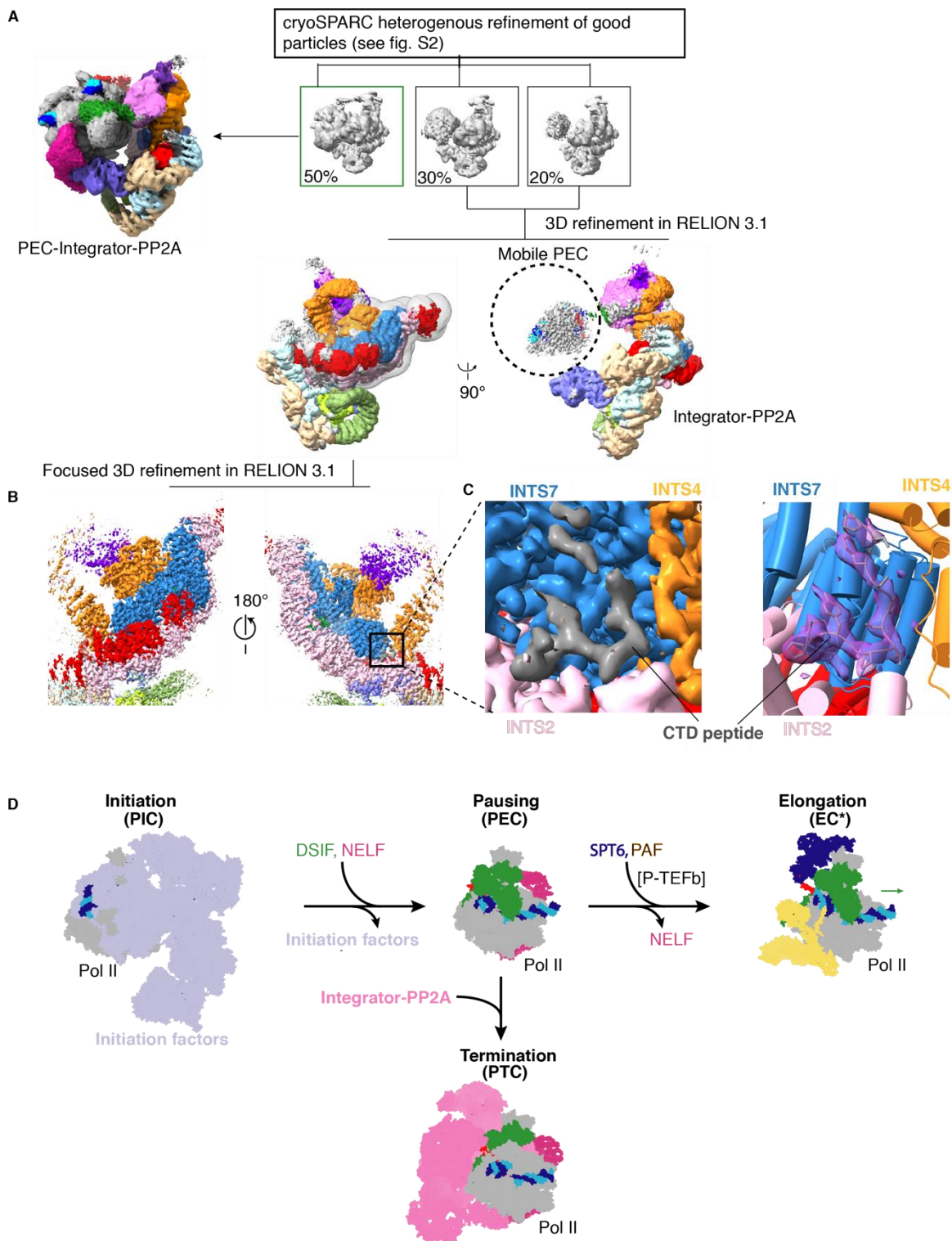


Fig. S11. Integrator recruitment via Pol II CTD and transcription regulation. (A-C) Integrator binds the CTD without stably binding the PEC. **(A)** Particles in 3D classes without stably bound PEC (see fig. S2) were subjected to 3D refinement. Weak density from mobile Pol II is indicated with a dashed circle and the masked used for focused refinement is shown as a

transparent surface. **(B)** Focused refinement map around INTS2, INTS4 and INTS7 regions that bind the CTD. **(C)** Zoom-in showing a density for the CTD peptide fitted with our model for the two CTD heptad repeats. **(D)**. Model for Integrator-mediated transcription regulation. PIC, pre-initiation complex; PEC, paused elongation complex; EC*, activated elongation complex; PTC pre-termination complex. In the promoter-proximal region of protein-coding genes, the PEC can either be converted to EC* or to the PTC (PEC-Integrator-PP2A complex, this study), which is an intermediate during transcription termination. This intermediate likely also underlies 3'-processing of snRNAs and Pol II termination at all non-coding transcription units including enhancers.

Table S1. Statistics of cryo-EM reconstruction and structural model

	#1 Map 1 (PDB 7PKS) (EMD-13479)
Data collection and Processing	
Microscope	FEI Titan Krios
Voltage (kV)	300
Camera	Gatan K3
Magnification	81,000x
Pixel size at detector (Å/pixel)	1.05
Total electron exposure (e ⁻ /Å ²)	46.18
Exposure rate (e ⁻ /pixel/sec)	16.97
Number of frames collected during exposure	50
Defocus range (µm)	-0.5 to -2.0
Automation software	SerialEM
Energy filter slit width	20 eV
Micrographs collected (no.)	45,000
Micrographs used (no.)	45,000
Total extracted particles (no.)	4,437,434
Reconstruction:	
Refined particles (no.)	614283
Final particles (no.)	614283
Point-group or helical symmetry parameters	C1
Resolution (global, Å)	
FSC 0.5 (unmasked/masked)	6.8/4.4
FSC 0.143 (unmasked/masked)	4.1/3.6
Resolution range (local, Å)	3.0-9.3
Map sharpening <i>B</i> factor (Å ²) / (<i>B</i> factor Range)	-90
Map sharpening methods	RELION 3.1
Model composition	
Protein	13,141
Ligands	ZN:10, Mg:1, Mn:2
RNA/DNA	74
Model Refinement	
Refinement package	Phenix (real space)
Model-Map scores (-CC)	0.84
Initial models	6GML, 7CUN, 7BFP, 3DW8
Map used	Composite map [#]
<i>B</i> factors (Å ²)	
Protein residues	90.28
Ligands	147.23
RNA/DNA	208.82
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.04
Bond angles (°)	0.78
Validation	
MolProbity score	1.56
CaBLAM outliers	2.36
Clashscore	5.27
Poor rotamers (%)	0.00

C-beta deviations	0.00
EMRinger score	2.86
Ramachandran plot	
Favored (%)	95.92
Allowed (%)	4.04
Outliers (%)	0.04

A composite map was created by aligning all the high resolution focused maps onto Map 1 using Warp (52). This map was sharpened with a B factor of -70 and used for refinement of the complete structure. We deposited this map as an additional map in the EMDB entry.

Table S2. Input structural models and model confidence

Complex/domain	Chain id(s)	Input model (pdb)	Level of confidence
Pol II	A-C, E, F, H-L, N, T, P	6GML	Atomic
Pol II stalk	D, G	6GML	Rigid body fitting
SPT5 KOW5	Z	6GML	Atomic
SPT5 KOW2-3, KOW _x -4	Z	6GML	Rigid body fitting
NELF-A	U	6GML	Atomic
NELF-B	V	6GML	Rigid body fitting
NELF-C	W	6GML	Atomic
NELF-E	X	6GML	Rigid body fitting
INTS1(357-866)	a	-	Backbone trace
INTS1(949-1316)	a	-	Atomic
INTS1(1360-2190)	a	7CUN	Atomic
INTS2, 4, 5, 6, 7, 8	b, d, e, f, g, h	7CUN	Atomic
INTS9, 11	i, k	7BFP	Atomic
PP2A	p, q	3DW8	Atomic

Table S3: Quantification of input RNA signal intensities in the RNA cleavage assay (related to Fig. S10C-F).

	Area	Mean Intensity	Background subtracted	Uncleaved Fraction
Replicate 1				
background	0.29	22240.595		
2.Int(wt)	0.29	38665.738	16425.143	0.985793238
3.Int(E203Q)	0.29	38902.449	16661.854	1
4.PEC-INT(wt)	0.29	30801.698	8561.103	0.513814549
5. PEC-Int(E203Q)	0.29	35384.689	13144.094	0.788873435
6. Pol II-NELF-Int(wt)	0.29	34103.413	11862.818	0.71197467
7. Pol II-DSIF-Int(wt)	0.29	31710.512	9469.917	0.568359139
8. Pol II-Int(wt)	0.29	33576.192	11335.597	0.680332273
Replicate 2				
background	0.29	22142.326		
2.Int(wt)	0.29	39474.489	17332.163	0.912784503
3.Int(E203Q)	0.29	41130.557	18988.231	1
4.PEC-INT(wt)	0.29	30435.648	8293.322	0.436761171
5. PEC-Int(E203Q)	0.29	36263.024	14120.698	0.743655267
6. Pol II-NELF-Int(wt)	0.29	35184.581	13042.255	0.686859929
7. Pol II-DSIF-Int(wt)	0.29	33034.568	10892.242	0.573631214
8. Pol II-Int(wt)	0.29	36542.306	14399.98	0.75836343
Replicate 3				
background	0.264	39184.911		
2.Int(wt)	0.264	51514.985	12330.074	0.973642249
3.Int(E203Q)	0.264	51848.776	12663.865	1
4.PEC-INT(wt)	0.264	45388.817	6203.906	0.489890409
5. PEC-Int(E203Q)	0.264	49205.817	10020.906	0.791299181
6. Pol II-NELF-Int(wt)	0.264	48649.271	9464.36	0.747351618
7. Pol II-DSIF-Int(wt)	0.264	47232.817	8047.906	0.635501563
8. Pol II-Int(wt)	0.264	49224.132	10039.221	0.792745422

References and Notes

1. D. Baillat, M.-A. Hakimi, A. M. Näär, A. Shilatifard, N. Cooch, R. Shiekhatter, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005). [doi:10.1016/j.cell.2005.08.019](https://doi.org/10.1016/j.cell.2005.08.019) [Medline](#)
2. F. Lai, A. Gardini, A. Zhang, R. Shiekhatter, Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015). [doi:10.1038/nature14906](https://doi.org/10.1038/nature14906) [Medline](#)
3. M. P. Rubtsova, D. P. Vasilkova, M. A. Moshareva, A. N. Malyavko, M. B. Meerson, T. S. Zatsepin, Y. V. Naraykina, A. V. Beletsky, N. V. Ravin, O. A. Dontsova, Integrator is a key component of human telomerase RNA biogenesis. *Sci. Rep.* **9**, 1701 (2019). [doi:10.1038/s41598-018-38297-6](https://doi.org/10.1038/s41598-018-38297-6) [Medline](#)
4. J. Barra, G. S. Gaidosh, E. Blumenthal, F. Beckedorff, M. M. Tayari, N. Kirstein, T. K. Karakach, T. H. Jensen, F. Impens, K. Gevaert, E. Leucci, R. Shiekhatter, J.-C. Marine, Integrator restrains paraspeckles assembly by promoting isoform switching of the lncRNA *NEAT1*. *Sci. Adv.* **6**, eaaz9072 (2020). [doi:10.1126/sciadv.aaz9072](https://doi.org/10.1126/sciadv.aaz9072) [Medline](#)
5. J. Yamamoto, Y. Hagiwara, K. Chiba, T. Isobe, T. Narita, H. Handa, Y. Yamaguchi, DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* **5**, 4263 (2014). [doi:10.1038/ncomms5263](https://doi.org/10.1038/ncomms5263) [Medline](#)
6. A. Gardini, D. Baillat, M. Cesaroni, D. Hu, J. M. Marinis, E. J. Wagner, M. A. Lazar, A. Shilatifard, R. Shiekhatter, Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* **56**, 128–139 (2014). [doi:10.1016/j.molcel.2014.08.004](https://doi.org/10.1016/j.molcel.2014.08.004) [Medline](#)
7. B. Stadelmayer, G. Micas, A. Gamot, P. Martin, N. Malirat, S. Koval, R. Raffel, B. Sobhian, D. Severac, S. Rialle, H. Parrinello, O. Cuvier, M. Benkirane, Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014). [doi:10.1038/ncomms6531](https://doi.org/10.1038/ncomms6531) [Medline](#)
8. J. R. Skaar, A. L. Ferris, X. Wu, A. Saraf, K. K. Khanna, L. Florens, M. P. Washburn, S. H. Hughes, M. Pagano, The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Res.* **25**, 288–305 (2015). [doi:10.1038/cr.2015.19](https://doi.org/10.1038/cr.2015.19) [Medline](#)
9. N. D. Elrod, T. Henriques, K.-L. Huang, D. C. Tatomer, J. E. Wilusz, E. J. Wagner, K. Adelman, The Integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* **76**, 738–752.e7 (2019). [doi:10.1016/j.molcel.2019.10.034](https://doi.org/10.1016/j.molcel.2019.10.034) [Medline](#)
10. D. C. Tatomer, N. D. Elrod, D. Liang, M.-S. Xiao, J. Z. Jiang, M. Jonathan, K.-L. Huang, E. J. Wagner, S. Cherry, J. E. Wilusz, The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* **33**, 1525–1538 (2019). [doi:10.1101/gad.330167.119](https://doi.org/10.1101/gad.330167.119) [Medline](#)
11. F. Beckedorff, E. Blumenthal, L. F. daSilva, Y. Aoi, P. R. Cingaram, J. Yue, A. Zhang, S. Dokaneheifard, M. G. Valencia, G. Gaidosh, A. Shilatifard, R. Shiekhatter, The human Integrator complex facilitates transcriptional elongation by endonucleolytic cleavage of

- nascent transcripts. *Cell Rep.* **32**, 107917 (2020). [doi:10.1016/j.celrep.2020.107917](https://doi.org/10.1016/j.celrep.2020.107917) [Medline](#)
12. S. Lykke-Andersen, K. Žumer, E. Š. Molska, J. O. Rouvière, G. Wu, C. Demel, B. Schwalb, M. Schmid, P. Cramer, T. H. Jensen, Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell* **81**, 514–529.e6 (2021). [doi:10.1016/j.molcel.2020.12.014](https://doi.org/10.1016/j.molcel.2020.12.014) [Medline](#)
 13. K. Kamieniarz-Gdula, N. J. Proudfoot, Transcriptional control by premature termination: A forgotten mechanism. *Trends Genet.* **35**, 553–564 (2019). [doi:10.1016/j.tig.2019.05.005](https://doi.org/10.1016/j.tig.2019.05.005) [Medline](#)
 14. K. L. Huang, D. Jee, C. B. Stein, N. D. Elrod, T. Henriques, L. G. Mascibroda, D. Baillat, W. K. Russell, K. Adelman, E. J. Wagner, Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Mol. Cell* **80**, 345–358.e9 (2020). [doi:10.1016/j.molcel.2020.08.016](https://doi.org/10.1016/j.molcel.2020.08.016) [Medline](#)
 15. H. Zheng, Y. Qi, S. Hu, X. Cao, C. Xu, Z. Yin, X. Chen, Y. Li, W. Liu, J. Li, J. Wang, G. Wei, K. Liang, F. X. Chen, Y. Xu, Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* **370**, eabb5872 (2020). [doi:10.1126/science.abb5872](https://doi.org/10.1126/science.abb5872) [Medline](#)
 16. S. J. Vervoort, S. A. Welsh, J. R. Devlin, E. Barbieri, D. A. Knight, S. Offley, S. Bjelosevic, M. Costacurta, I. Todorovski, C. J. Kearney, J. J. Sandow, Z. Fan, B. Blyth, V. McLeod, J. H. A. Vissers, K. Pavic, B. P. Martin, G. Gregory, E. Demosthenous, M. Zethoven, I. Y. Kong, E. D. Hawkins, S. J. Hogg, M. J. Kelly, A. Newbold, K. J. Simpson, O. Kauko, K. F. Harvey, M. Ohlmeyer, J. Westermarck, N. Gray, A. Gardini, R. W. Johnstone, The PP2A-Integrator-CDK9 axis fine-tunes transcription and can be targeted therapeutically in cancer. *Cell* **184**, 3143–3162.e32 (2021). [doi:10.1016/j.cell.2021.04.022](https://doi.org/10.1016/j.cell.2021.04.022) [Medline](#)
 17. J. Chen, N. Ezzeddine, B. Waltenspiel, T. R. Albrecht, W. D. Warren, W. F. Marzluff, E. J. Wagner, An RNAi screen identifies additional members of the *Drosophila* Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA* **18**, 2148–2156 (2012). [doi:10.1261/rna.035725.112](https://doi.org/10.1261/rna.035725.112) [Medline](#)
 18. T. R. Albrecht, S. P. Shevtsov, Y. Wu, L. G. Mascibroda, N. J. Peart, K.-L. Huang, I. A. Sawyer, L. Tong, M. Dunder, E. J. Wagner, Integrator subunit 4 is a ‘Symplekin-like’ scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic Acids Res.* **46**, 4241–4255 (2018). [doi:10.1093/nar/gky100](https://doi.org/10.1093/nar/gky100) [Medline](#)
 19. M. M. Pfeleiderer, W. P. Galej, Structure of the catalytic core of the Integrator complex. *Mol. Cell* **81**, 1246–1259.e8 (2021). [doi:10.1016/j.molcel.2021.01.005](https://doi.org/10.1016/j.molcel.2021.01.005) [Medline](#)
 20. Z. Dominski, X. C. Yang, W. F. Marzluff, The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. *Cell* **123**, 37–48 (2005). [doi:10.1016/j.cell.2005.08.002](https://doi.org/10.1016/j.cell.2005.08.002) [Medline](#)
 21. K. Sabath, M. L. Stäubli, S. Marti, A. Leitner, M. Moes, S. Jonas, INTS10-INTS13-INTS14 form a functional module of Integrator that binds nucleic acids and the cleavage module. *Nat. Commun.* **11**, 3422 (2020). [doi:10.1038/s41467-020-17232-2](https://doi.org/10.1038/s41467-020-17232-2) [Medline](#)
 22. S. M. Vos, L. Farnung, H. Urlaub, P. Cramer, Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **560**, 601–606 (2018). [doi:10.1038/s41586-018-0442-2](https://doi.org/10.1038/s41586-018-0442-2) [Medline](#)

23. C. Bernecky, J. M. Plitzko, P. Cramer, Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat. Struct. Mol. Biol.* **24**, 809–815 (2017). [doi:10.1038/nsmb.3465](https://doi.org/10.1038/nsmb.3465) [Medline](#)
24. P. Cramer, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001). [doi:10.1126/science.1059493](https://doi.org/10.1126/science.1059493) [Medline](#)
25. J. Chen, B. Waltenspiel, W. D. Warren, E. J. Wagner, Functional analysis of the integrator subunit 12 identifies a microdomain that mediates activation of the *Drosophila* integrator complex. *J. Biol. Chem.* **288**, 4867–4877 (2013). [doi:10.1074/jbc.M112.425892](https://doi.org/10.1074/jbc.M112.425892) [Medline](#)
26. Y. Sun, Y. Zhang, W. S. Aik, X.-C. Yang, W. F. Marzluff, T. Walz, Z. Dominski, L. Tong, Structure of an active human histone pre-mRNA 3'-end processing machinery. *Science* **367**, 700–703 (2020). [doi:10.1126/science.aaz7758](https://doi.org/10.1126/science.aaz7758) [Medline](#)
27. S. Gressel, B. Schwalb, T. M. Decker, W. Qin, H. Leonhardt, D. Eick, P. Cramer, CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6**, e29736 (2017). [doi:10.7554/eLife.29736](https://doi.org/10.7554/eLife.29736) [Medline](#)
28. P. Uguen, S. Murphy, The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO J.* **22**, 4544–4554 (2003). [doi:10.1093/emboj/cdg430](https://doi.org/10.1093/emboj/cdg430) [Medline](#)
29. J. Chen, E. J. Wagner, snRNA 3' end formation: The dawn of the Integrator complex. *Biochem. Soc. Trans.* **38**, 1082–1087 (2010). [doi:10.1042/BST0381082](https://doi.org/10.1042/BST0381082) [Medline](#)
30. N. Hernandez, Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *EMBO J.* **4**, 1827–1837 (1985). [doi:10.1002/j.1460-2075.1985.tb03857.x](https://doi.org/10.1002/j.1460-2075.1985.tb03857.x) [Medline](#)
31. J. Li, X. Ma, S. Banerjee, S. Baruah, N. J. Schnicker, E. Roh, W. Ma, K. Liu, A. M. Bode, Z. Dong, Structural basis for multifunctional roles of human Ints3 C-terminal domain. *J. Biol. Chem.* **296**, 100112 (2021). [doi:10.1074/jbc.RA120.016393](https://doi.org/10.1074/jbc.RA120.016393) [Medline](#)
32. Y. Jia, Z. Cheng, S. R. Bharath, Q. Sun, N. Su, J. Huang, H. Song, Crystal structure of the INTS3/INTS6 complex reveals the functional importance of INTS3 dimerization in DSB repair. *Cell Discov.* **7**, 66 (2021). [doi:10.1038/s41421-021-00283-0](https://doi.org/10.1038/s41421-021-00283-0) [Medline](#)
33. S. Egloff, D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick, S. Murphy, Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777–1779 (2007). [doi:10.1126/science.1145989](https://doi.org/10.1126/science.1145989) [Medline](#)
34. S. Egloff, S. A. Szczepaniak, M. Dienstbier, A. Taylor, S. Knight, S. Murphy, The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J. Biol. Chem.* **285**, 20564–20569 (2010). [doi:10.1074/jbc.M110.132530](https://doi.org/10.1074/jbc.M110.132530) [Medline](#)
35. N. Shah, M. A. Maqbool, Y. Yahia, A. Z. El Aabidine, C. Esnault, I. Forné, T.-M. Decker, D. Martin, R. Schüller, S. Krebs, H. Blum, A. Imhof, D. Eick, J.-C. Andrau, Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol. Cell* **69**, 48–61.e6 (2018). [doi:10.1016/j.molcel.2017.12.009](https://doi.org/10.1016/j.molcel.2017.12.009) [Medline](#)

36. S. Egloff, J. Zaborowska, C. Laitem, T. Kiss, S. Murphy, Ser7 phosphorylation of the CTD recruits the RPA2 Ser5 phosphatase to snRNA genes. *Mol. Cell* **45**, 111–122 (2012). [doi:10.1016/j.molcel.2011.11.006](https://doi.org/10.1016/j.molcel.2011.11.006) [Medline](#)
37. S. M. Vos, L. Farnung, M. Boehning, C. Wigge, A. Linden, H. Urlaub, P. Cramer, Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607–612 (2018). [doi:10.1038/s41586-018-0440-4](https://doi.org/10.1038/s41586-018-0440-4) [Medline](#)
38. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001). [doi:10.1126/science.1059495](https://doi.org/10.1126/science.1059495) [Medline](#)
39. N. J. Proudfoot, Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016). [doi:10.1126/science.aad9926](https://doi.org/10.1126/science.aad9926) [Medline](#)
40. E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018). [doi:10.1002/pro.3280](https://doi.org/10.1002/pro.3280) [Medline](#)
41. S. D. Gradia, J. P. Ishida, M.-S. Tsai, C. Jeans, J. A. Tainer, J. O. Fuss, MacroBac: New technologies for robust and efficient large-scale production of recombinant multiprotein complexes. *Methods Enzymol.* **592**, 1–26 (2017). [doi:10.1016/bs.mie.2017.03.008](https://doi.org/10.1016/bs.mie.2017.03.008) [Medline](#)
42. J. Bitinaite, N. M. Nichols, DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol. Biol.* **Chapter 3**, 21 (2009). [doi:10.1002/0471142727.mb0321s86](https://doi.org/10.1002/0471142727.mb0321s86) [Medline](#)
43. J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways. *PLOS ONE* **4**, e6441 (2009). [doi:10.1371/journal.pone.0006441](https://doi.org/10.1371/journal.pone.0006441) [Medline](#)
44. H. Liu, J. H. Naismith, An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008). [doi:10.1186/1472-6750-8-91](https://doi.org/10.1186/1472-6750-8-91) [Medline](#)
45. E. Barbieri, M. Trizzino, S. A. Welsh, T. A. Owens, B. Calabretta, M. Carroll, K. Sarma, A. Gardini, Targeted enhancer activation by a subunit of the integrator complex. *Mol. Cell* **71**, 103–116.e7 (2018). [doi:10.1016/j.molcel.2018.05.031](https://doi.org/10.1016/j.molcel.2018.05.031) [Medline](#)
46. U. S. Cho, W. Xu, Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445**, 53–57 (2007). [doi:10.1038/nature05351](https://doi.org/10.1038/nature05351) [Medline](#)
47. I. Berger, D. J. Fitzgerald, T. J. Richmond, Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004). [doi:10.1038/nbt1036](https://doi.org/10.1038/nbt1036) [Medline](#)
48. S. M. Vos, D. Pöhlmann, L. Caizzi, K. B. Hofmann, P. Rombaut, T. Zimniak, F. Herzog, P. Cramer, Architecture and RNA binding of the human negative elongation factor. *eLife* **5**, e14981 (2016). [doi:10.7554/eLife.14981](https://doi.org/10.7554/eLife.14981) [Medline](#)
49. C. Bernecky, F. Herzog, W. Baumeister, J. M. Plitzko, P. Cramer, Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551–554 (2016). [doi:10.1038/nature16482](https://doi.org/10.1038/nature16482) [Medline](#)

50. I. Fianu, C. Dienemann, S. Aibara, S. Schilbach, P. Cramer, Cryo-EM structure of mammalian RNA polymerase II in complex with human RPAP2. *Commun. Biol.* **4**, 606 (2021). [doi:10.1038/s42003-021-02088-z](https://doi.org/10.1038/s42003-021-02088-z) [Medline](#)
51. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005). [doi:10.1016/j.jsb.2005.07.007](https://doi.org/10.1016/j.jsb.2005.07.007) [Medline](#)
52. D. Tegunov, P. Cramer, Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019). [doi:10.1038/s41592-019-0580-y](https://doi.org/10.1038/s41592-019-0580-y) [Medline](#)
53. S. H. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). [doi:10.1016/j.jsb.2012.09.006](https://doi.org/10.1016/j.jsb.2012.09.006) [Medline](#)
54. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). [doi:10.7554/eLife.42166](https://doi.org/10.7554/eLife.42166) [Medline](#)
55. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). [doi:10.1038/nmeth.4169](https://doi.org/10.1038/nmeth.4169) [Medline](#)
56. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). [doi:10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) [Medline](#)
57. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). [doi:10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) [Medline](#)
58. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). [doi:10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925) [Medline](#)
59. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018). [doi:10.1107/S2059798318006551](https://doi.org/10.1107/S2059798318006551) [Medline](#)
60. J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, D. Baker, Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020). [doi:10.1073/pnas.1914677117](https://doi.org/10.1073/pnas.1914677117) [Medline](#)
61. R. T. Kidmose, J. Juhl, P. Nissen, T. Boesen, J. L. Karlsen, B. P. Pedersen, Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526–531 (2019). [doi:10.1107/S2052252519007619](https://doi.org/10.1107/S2052252519007619) [Medline](#)

62. Y. Xu, Y. Chen, P. Zhang, P. D. Jeffrey, Y. Shi, Structure of a protein phosphatase 2A holoenzyme: Insights into B55-mediated Tau dephosphorylation. *Mol. Cell* **31**, 873–885 (2008). [doi:10.1016/j.molcel.2008.08.006](https://doi.org/10.1016/j.molcel.2008.08.006) [Medline](#)
63. M. G. Prisant, C. J. Williams, V. B. Chen, J. S. Richardson, D. C. Richardson, New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink “waters,” and NGL Viewer to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020). [doi:10.1002/pro.3786](https://doi.org/10.1002/pro.3786) [Medline](#)
64. B. A. Barad, N. Echols, R. Y.-R. Wang, Y. Cheng, F. DiMaio, P. D. Adams, J. S. Fraser, EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015). [doi:10.1038/nmeth.3541](https://doi.org/10.1038/nmeth.3541) [Medline](#)
65. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021). [doi:10.1002/pro.3943](https://doi.org/10.1002/pro.3943) [Medline](#)
66. Z. L. Chen, J.-M. Meng, Y. Cao, J.-L. Yin, R.-Q. Fang, S.-B. Fan, C. Liu, W.-F. Zeng, Y.-H. Ding, D. Tan, L. Wu, W.-J. Zhou, H. Chi, R.-X. Sun, M.-Q. Dong, S.-M. He, A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019). [doi:10.1038/s41467-019-11337-z](https://doi.org/10.1038/s41467-019-11337-z) [Medline](#)
67. B. Yang, Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He, M.-Q. Dong, Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012). [doi:10.1038/nmeth.2099](https://doi.org/10.1038/nmeth.2099) [Medline](#)
68. C. W. Combe, L. Fischer, J. Rappsilber, xiNET: Cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015). [doi:10.1074/mcp.O114.042259](https://doi.org/10.1074/mcp.O114.042259) [Medline](#)
69. J. Kosinski, A. von Appen, A. Ori, K. Karius, C. W. Müller, M. Beck, Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015). [doi:10.1016/j.jsb.2015.01.014](https://doi.org/10.1016/j.jsb.2015.01.014) [Medline](#)
70. X. Chen, X. Yin, J. Li, Z. Wu, Y. Qi, X. Wang, W. Liu, Y. Xu, Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* **372**, eabg0635 (2021). [doi:10.1126/science.abg0635](https://doi.org/10.1126/science.abg0635) [Medline](#)
71. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009). [doi:10.1093/bioinformatics/btp033](https://doi.org/10.1093/bioinformatics/btp033) [Medline](#)
72. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42** (W1), W320–4 (2014). [doi:10.1093/nar/gku316](https://doi.org/10.1093/nar/gku316) [Medline](#)
73. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012). [doi:10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089) [Medline](#)