Science
AAAS

Supplementary Materials for

Structure of the human Mediator-bound transcription pre-initiation complex

R. Abdella[1,2]†, A. Talyzina[1,2]†, S. Chen[1,2], C. J. Inouye[3,4,5,6], R. Tjian[3,4,5,6]*, Y. He[1,2,7,8]*.

Correspondence to: yuanhe@northwestern.edu, jmlim@berkeley.edu

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Movies S1 to S4

Materials and Methods

Purification of PIC components

Pol II and TFIIH were purified endogenously from HeLa cells, and TFIIB, TFIIA, TBP, TFIIE, and TFIIF were purified recombinantly, as previously described (*12, 41*). Mediator was purified endogenously from HeLa cells, as previously described (*42*). In short, HeLa cell nuclear extract was run over a phosphocellulose column using 0.1M KCl HEMG (20 mM HEPES, 10 mM EDTA, 2 mM $MgCl_2$, and 10% glycerol). Stepwise elution of protein complexes was performed at 0.1 M, 0.3 M, 0.5, M and 1.0 M HCl HEMG. The 0.5 M and 1.0 M elutions were dialyzed against 0.1 M HEMG before being subjected to further affinity purification using a GST-VP16 fusion protein bound to glutathione Sepharose resin (GE). Following a 3-hour incubation with the affinity resin, the resin was washed 5 times with 50 column volumes of 0.5 M KCl HEGN (20 mM HEPES, 10 mM EDTA, 10% glycerol, 0.1% NP-40), followed by one wash with 50 column volumes of 0.15 M KCl HEGN (0.02% NP-40). Mediator was eluted using 30 mM glutathione in 0.15 M TEGN (20 mM Tris pH 7.9, 0.1 mM EDTA, 10% glycerol, and 0.02% NP-40). Mediator is found in both the P0.5M and P1M fractions, and in our hands, no significant difference in Mediator composition is seen between the two fractions.

Assembly of hMed-PIC

Human Med-PIC complexes were assembled as previously described for the PIC with the following changes to accommodate the incorporation of Mediator into the complex (*12, 41*). For negative staining, three subcomplexes were assembled in parallel. First, 0.25 pmol of a super core promoter DNA template (sense: 5'-GAAGGGCGCCTATAAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACACTCGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3', anti-sense: 5'-/5BiotinTEG/ACTGGGGAATTCCATGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGTTCGATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTATAGGCGCCCTTC-3') was mixed with 1.8 pmol TFIIB, 2 pmol TBP, 1 pmol TFIIA. 0.1 pmol RNA Pol II was mixed with 0.7 pmol TFIIF in a second tube. In a third tube, 1.5 pmol Mediator was mixed with 2.5 pmol TFIIE56, 7.6 pmol TFIIE34, and 1 pmol TFIIH. The salt concentration of each solution was adjusted to 100 mM KCl with the addition of buffers A (12 mM HEPES pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, and 0.05% NP-40) and B (12 mM HEPES pH 7.9, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 1 mM DTT, and 0.05% NP-40). After 30 minutes at room temperature (RT), all components were combined and incubated for an additional 30 minutes at RT before binding to T1 streptavidin beads (Fisher Scientific) at RT for 15 minutes. Assembled complexes were washed with buffer C (10 mM HEPES pH 7.9, 10 mM Tris pH 7.9, 5% glycerol, 5 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, and 0.05% NP-40) and eluted with buffer D (10 mM HEPES pH 7.6, 5% glycerol, 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 0.05% NP-40, and 30 units EcoRI-HF (New England Biolabs)).

Complex assembly for cryo-EM was identical to negative staining samples. Assembled complexes were always used fresh for microscopy and never flash-frozen to maintain the structural integrity of the complex.

Electron Microscopy

Negative stain samples were prepared using 400 mesh copper grids (Electron Microscopy Sciences) with a thin layer of continuous carbon on a nitrocellulose support film that was glow-discharged in air for 10 seconds with 25 W of power using the PELCO easiGlow (TED PELLA).

Purified Med-PIC complexes in buffer D were cross-linked with 0.05% glutaraldehyde for 10 minutes on ice and incubated for 10 minutes on a grid in a homemade humidity chamber at 4 ℃. The grid was sequentially incubated on 4, 40 µL drops of 2% uranyl formate solution for 5, 10, 15, and 20 seconds and blotted dry with #1 filter paper (Whatman). Images were collected on a Jeol 1400 equipped with a Gatan 4k × 4x CCD camera at 30,000X magnification (3.71 Å/pixel), a defocus range of -1.5 to -3 µm, and 20 $e^-/Å^2$ total electron dose using Leginon (*43*).

Cryo-EM samples were prepared using Quantifoil 2/1 300 mesh copper grids (EMS). Grids were glow discharged in air for 10 seconds with 5 W of power using the PELCO easiGlow, and then a thin layer of graphene oxide was applied as described previously (*44*). Eluted Med-PIC samples (~3.5 µL) were incubated with 0.05% glutaraldehyde for 10 minutes on ice in the dark. The sample was applied to a grid suspended in a Vitrobot operating at 4 ℃ with 100% humidity. After 5 minutes, the sample was blotted with 10 force for 4 seconds and immediately plunged into liquid ethane cooled to liquid nitrogen temperatures. A data set of 19,881 images was collected at the Pacific Northwestern Center for Cryo-EM (PNCC). Images were collected using semi-automated data collection in Serial EM (*45*) on a 300 kV Titan Krios-3 microscope (Thermo Fisher) equipped with a Gatan K3 direct detector operating in super-resolution mode at a magnification of 30,000X (0.5295 Å/pixel). Images were collected using a defocus range of -2 to -4 µm with a 45-frame exposure taken over a total of 2.1 seconds using a dose rate of 15 e^-/pixel/second for a total dose of 31.5 $e^-/Å^2$. 19,881 images total were collected.

Image processing

For negative-stained samples, particles were picked using DogPicker, extracted, and 2D classified using iterative MSA/MRA topological alignment within the Appion data processing software (*46-49*). A particle stack of at least 50,000 particles with a box size of 144 x 144 pixels was subjected to iterative, multi-reference projection-matching 3D refinement using libraries from the EMAN2 software package to generate an initial reference for cryo-EM data processing (*50*).

RELION 3.1 was used for all pre-processing, 3D classification, model refinement, post-processing, and local-resolution estimation jobs (*51*). To pre-process the cryo-EM data, movie frames were aligned using RELION's implementation while binning by a factor of 2 (1.059 Å/pixel). After motion correction, micrographs were manually inspected, resulting in the exclusion of 3,903 micrographs from further processing. Particles were automatically picked using Gautomatch (developed by K. Zhang, MRC Laboratory of Molecular Biology, Cambridge, UK), and the local CTF of each micrograph was determined using Gctf or CTFFIND-4.1 (*49*, *51*).

An initial particle stack of 885,514 particles was binned by a factor of 4 (4.236 Å/pixel), extracted, and subjected to an initial round of 3D classification using the negative stain reconstruction (low-pass filtered to 30 Å) as an initial reference. Class 5 (156,383 particles) showed sharp and clear structural features of Med-PIC, so it was selected for further processing. The selected particles were 3D auto-refined, re-centered, and re-extracted without binning (1.059 Å/pixel, box size = 450 pixels). Another round of 3D auto-refinement was performed with a soft mask applied around the whole complex, resulting in a 5.79 Å resolution reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion (*53*). Per-particle CTF refinement was performed by first estimating magnification anisotropy, then per-particle defocus and per-micrograph astigmatism, and finally beam tilt, followed by Bayesian particle polishing. 3D auto-refinement using the polished

particles yielded a 4.8 Å resolution map. The map was post-processed using DeepEMhancer (*54*). This map is the full Med-PIC map used for deposition.

To improve the map quality of distal portions of the complex that showed significant averaging out in the Med-PIC map, we segmented the Med-PIC map into 7 bodies: MedTail, Med14Med24, MedHead, MedMiddle-CAK, cPIC, cTFIIH, and Med1. For each body, a partial soft mask was applied to the corresponding volume, and particles were signal subtracted, re-centered, and re-extracted with a suitable box size (384, 360, 384, 320, 320, 288, and 288 pixels, respectively). Next, the particles are binned by a factor of 2 (2.118 Å/pixel) and 3D auto-refined locally with an initial angular sampling interval of 3.7°. Subsequent three-class 3D classification without alignment was performed, and the class with the best features and resolution (See Supp. Fig. 2 for particle numbers of each selected class) was selected, un-binned (1.059 Å/pixel), auto-refined, and post-processed. Local resolution of the maps was estimated with RELION 3.1.

3D variability analysis was performed on the Med-PIC, MedΔTail-PIC, and Med-CAK maps using CryoSPARC (*55*). For Med-PIC, a soft mask was applied, signal subtraction was performed, and the subtracted stack was binned by 2 (2.118 Å/pixel), re-centered, and re-boxed (280 pixels) in Relion. This stack was transferred to CryoSPARC for masked non-uniform refinement, which resulted in a 4.3 Å resolution map. 3D variability analysis was performed on the aligned stack after filtering to 5 Å resolution, and the first three principal components were selected for analysis. A similar strategy was used for the remaining two maps with box sizes of 180 and 270 pixels, respectively. Both maps gave 4.3 Å resolution maps after non-uniform refinement.

UCSF Chimera and UCSF Chimera X were used for all volume segmentation, figure and movie generation, and rigid-body docking (*56, 57*). In parallel with post-processing done in RELION3.1, DeepEMhancer was applied on the refined maps to better correct local B-factors and yielded cleaner maps for model building and docking (*54*).

Model building

cPIC:

The human cPIC bound to a closed DNA template (PDB:5IYA, *12*) was fit as a rigid body into the cPIC density map as an initial model using UCSF Chimera (*57*). Manual adjustments were made in Coot thanks to the high resolution of the cPIC map. Modifications made to TFIIE were guided by the human TFIIE crystal structure (PDB: 5GPY, *58*). The cPIC model was real space refined in Phenix to the cPIC map (*59*).

MedHead:

S. pombe subunit structures (PDB:5U0S, *6*) of Med6, 8, 17, and 22 or the *S. cerevisiae* (PDB:5OQM, *7*) structure of Med11 were used as initial models for building the human structures using the MedHead map in Coot (*60*). Map quality was sufficient to see side chains of bulky residues what was crucial for determining the register of the sequences. Final models were built by threading the human sequences onto the yeast structures and making any necessary adjustments guided by sequence alignment and secondary structure prediction from the MPI Bioinformatics Toolkit (*61*) and Jpred4 (*62*), respectively. The Med14C map was used for building portions of Med17C, Med27, Med28, Med29, and Med30, which were better resolved in that map than MedHead.

Homology models of Med18 and Med20 were built using the MPI Bioinformatics Toolkit and Modeller (*61, 63*). These homology models were aligned to the structure of the Med18-Med20 dimer structure (PDB:2HZM, *64*), flexible fitted into the post-processed map of MedHead using ISOLDE in UCSF Chimera X and manually adjusted in Coot. Med28 and Med30 were built using secondary structure prediction and their known closer association with MedHead than Med27 and Med29. They could be correctly assigned in the density by initially noting that Med30 has a much longer flexible linker between helices than Med28. Med27 was built by identifying the location of the C-terminus through homology to the *S. pombe* structure. The N-terminus and Med29 were built by identifying the remaining helical density that closely matched secondary structure prediction and identifying the register based on clear bulky side chain density. Med27 was also validated due to the proximity of its N- and C-terminal ends. MedHead subunits were combined and real space refined in Phenix to the MedHead map.

Med14C:

Yeast Med14 (PDB:5OQM, *7*) was fitted into the Med14C density as rigid body and used to guide building the RM1 and RM2 domains guided by sequence alignment and secondary structure prediction. The C-terminus was built guided by secondary structure prediction and the high quality of the density in this area. Residues 968 to 1167, which are not predicted to form common secondary structure elements, were missing in the density, but we were able to build much of the final RM domain, which displays the typical one helix-four strand-two helix-four strand fold. The quality of the map at the very C-terminus was not sufficient to build loops between secondary structure elements or identify the correct register of the final beta-sheets. Med14C was real space refined in Phenix against the Med14C map.

MedTail:

Human Med23 (PDB:6H02, *65*) was fit as a rigid body into the map of MedTail, and manual adjustments were made in Coot. Med16 was built by first locating the seven-stranded WD-40 domain in the map of MedTail. A homology model for this domain was built using PDB:5MZH (*66*). Manual alignment of this model into the density was performed in UCSF Chimera by noting the connectivity of the domain to the C-terminus of the protein. The model was then manually improved in Coot. The C-terminus was built by following the density from the C-terminus of the WD-40 domain. Clear helices were visible for the rest of the density and showed clear side-chain density for bulky residues, allowing manual building for the rest of the protein. Med24 is predicted to be almost entirely helical and was localized above Med23 in the density. This was the only remaining largely helical density where a subunit of this size could be located in MedTail. The register was established by identifying the longest predicted helices using secondary structure prediction, locating possible densities, and identifying bulky side chains.

The von Willebrand factor type A (vWA) domain of Med25 was built by first building a homology model (PDB:2KY6) using Modeller in the MPI Bioinformatics Toolkit and rigid body docking it into any unmodeled density remaining in MedTail. Manual adjustments were made to the final model in Coot.

The RWD of Med15 (residues 677-786) was built by first building a homology model (PDB:2EBK) using Modeller in the MPI Bioinformatics Toolkit and then rigid body fitting it into any unmodeled density remaining in MedTail using UCSF Chimera. Manual adjustments

were then made using Coot. Residues 617-652 were built by using secondary structure prediction and looking in the direction of the N-terminus of the RWD domain. The two helices, one with a large kink in it, showed clear side-chain density that matched the predicted sequence of Med15. MedTail subunits were combined and real space refined in Phenix to the MedTail map.

MedMiddle:

Homology models for Med4, 7, 9, 10, 14 (1-195), 19, 21, and 31 were created using sequence alignment and secondary structure prediction to their *S. cerevisiae* counterparts (PDB:5OQM, *7*) in Coot. These homology models were flexible fitted into the MedMiddle-CAK density using Namdinator (*67*). Manual inspection of the results, including building an additional C-terminal helix in Med31, N-terminal helix of Med17, and C-terminal helix of Med6, was done in Coot.

The $_{MED}$CTD structure was created by first aligning the yeast MedHead-CTD structure (PDB:4GWQ, *17*) to our human structure using Med8. The peptide was used as an initial model to rigid body fit into the MedHead density. Clear density for the sidechains of two Y^1 residues was visible in the MedHead density, and the remaining model was built using Coot.

TFIIH-CAK:

The human CAK module structure (PDB: 6XBZ, *35*) was fit as a rigid body into the MedMiddle-CAK density. The CDK2-cyclin A-peptide substrate structure (PDB: 1QMZ, *36*) was aligned using CDK2 to align to CDK7. The substrate peptide structure didn't need any adjustment to fit into the MedMiddle-CAK density. The sequence was mutated to the consensus sequence of the Pol II CTD, maintaining the SP motif in the substrate with S^5P^6 in the CTD and truncated to match the density visible in the structure. The model was combined with the MedMiddle structure and real space refined using Phenix (*57*). The CAK and MedMiddle subunits were combined, and real space refined in Phenix to the MedMiddle-CAK map.

cTFIIH:

The human apo-TFIIH structure (PDB: 6NMI, *68*) was used as an initial model for building into the TFIIH density. Because of differences in the shape of the horseshoe, individual subunits were docked into the density as rigid body. Portions of p62 and p44 for which there was no density in our structure were removed. XPB undergoes a conformational change between its position in the cTFIIH structure and its structure in the PIC. To model this, we separately rigid body docked residues 34 to 164, 165 to 296, 297 to 502, and 503 to 730 into the density and refined the connections between those domains in Coot. The aligned subunits were combined, and real space refined using Phenix to the cTFIIH map.

Med-PIC:

The cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, and MedTail maps were segmented to remove overlapping segments and fit into the Med-PIC map. The models corresponding to each map were aligned with the maps, combined, and validated using Phenix.

Supplementary Text

<u>Flexibility of Med-PIC</u>

Because of the size of Med-PIC and the number of rigid bodies required, multi-body refinement in Relion-3 was computationally prohibitive. Instead, we performed non-uniform refinement and 3D variability analysis in CryoSparc (*55*), which shows a broad distribution of movement of Mediator relative to the PIC (Fig. S11) (Video S2-4). This observation explains the low resolution or missing density far from the center of the post-processed map. We performed this analysis on three portions of Med-PIC: Med-PIC, MedΔTail-PIC, and Med-CAK (Fig. S11). Analysis of the first three principal components for each complex shows a high degree of similarity of movement with the interface between MedHead and the stalk of Pol II, serving as a pivot point for the rotation of Mediator relative to Pol II. This movement can either be up-and-down as in the case of Med-PIC PCs 1 and 3, MedΔTail-PIC PC 2, and Med-CAK PC 1, side-to-side as in the case of Med-PIC PC 2, MedΔTail-PIC PC 1, and Med-CAK PC 2, or a combination of the two as in MedΔTail-PIC PC 3 and Med-CAK PC 3.

This analysis led us to compare the interface between MedHead and Pol II in our human structure with that in the existing yeast Med-PIC structures. MedHead forms a closer association with the stalk of Pol II (RPB4/7) in the human structure than seen in any previous structure to date (Fig. S12A-B). Helices α^1 and α^2 of Med8 stack on top of helices α^4 and α^7 of RPB4 in all three Med-PIC structures available. However, the extent of that interaction differs significantly between species. In the human structure, these pairs of helices run parallel to each other, forming an extensive interface between MedHead and the RPB4/7 stalk, highlighted by close interactions between Med18 α^3 and RPB7 $\beta^{C1\text{-}C3}$ (Fig. S12A-B). In the scMed-PIC (*7*), scMedHead slides towards scMedTail, resulting in a sinking of the Med18-Med20 flexible jaw away from the stalk and the RNA exit tunnel of Pol II, and a lifting of the shoulder of Mediator (Fig. S12A-B). Due to the stabilization of the CAK module by the shoulder domain, this change would likely result in a lifting of the CAK module or shifting of the interface. In the *S. pombe* Med-PIC (spMed-PIC) structure (*6*), this movement is even more exaggerated with minimal overlap between the Med8 and RPB4 helices, resulting in an even larger gap between the stalk and flexible jaw and a slight rotation of spMedHead away from spMedMiddle. We were unable to identify a prominent principal component in our data set that captured the positions of yeast MedHead relative to Pol II.

Superimposing human MedHead with the scMedHead and spMedHead structures shows that they align very well with just subtle movements of the flexible jaw between species (Fig. S12C-D). The only difference is the position of the mobile jaw, Med18, and Med20. In humans and *S. pombe*, the presence of the Med27 subunit stabilizes the mobile jaw, but its absence in *S. cerevisiae* causes the sinking of the mobile jaw away from the fixed jaw.

Comparison of the *S. pombe* apo-Mediator and Med-cPIC structures shows that Med14 contains a hinge between the RM1 and RM2 domains, which ultimately leads to the raising and lowering of spMedMiddle, relative to spMedHead (*6*). This is very similar to Med-CAK PC 1, suggesting this flexibility remains after engagement with the PIC (Fig. S11C).

Architecture of cTFIIH

cTFIIH undergoes a conformational change from the apo structure (PDB:6NMI) to its structure in Med-PIC that involves an opening of the horseshoe (Fig. S13A-D). Structural changes within the ATP-dependent DNA helicase XPB subunit that contacts downstream DNA result in a large rotation of the damage recognition domain (DRD, residues 195-296) away from Mat1 and

towards the DNA, breaking the interaction between Mat1 and helix 248-261 of XPB (Fig. S13E). A slight shift of the RecA1 domain towards the DNA occurs upon DNA engagement. This conformation of XPB is nearly identical to that of XPB in the TFIIH-XPA-DNA structure (PDB:6RO4) formed during the initial steps of nucleotide excision repair (NER) (Fig. S13F). We did not observe any density for helix 248-261 of XPB even though the rest of the DRD was well structured. The loss of this contact site between XPB and Mat1 results in the C-terminal half of the long Mat1 helix (residues 163-210) being visible only at a much lower map threshold (Fig. S13G). XPD and Med8 sandwich the visible portion of the Mat1 helix, and RPB4/7 and TFIIE also contribute to the stabilization of the rest of the Mat1 N-terminus (Fig. S13H). Thus, assembly of the CAK module into Med-PIC does not require significant structural changes in Mat1. Even if the C-terminal half of the long helix is significantly more flexible in its elongated state, it could still connect to the Mat1 C-terminus bound to the CAK module. The opening of TFIIH also results in loss of density for the N-terminus of p44 (residues 1-50), which bridges across the horseshoe (Fig. S13A-B). This is also seen in the scMed-PIC (Fig. S13C). No density is observed for significant portions of p62 (residues 1-106, 148-371) that are present in both the apo-TFIIH structure and the *S. cerevisiae* Med-PIC structure. In the scMed-PIC, these portions of p62 interact with TFIIE, but this interaction is seemingly not essential for complex assembly (Fig. S13C) (*7*).

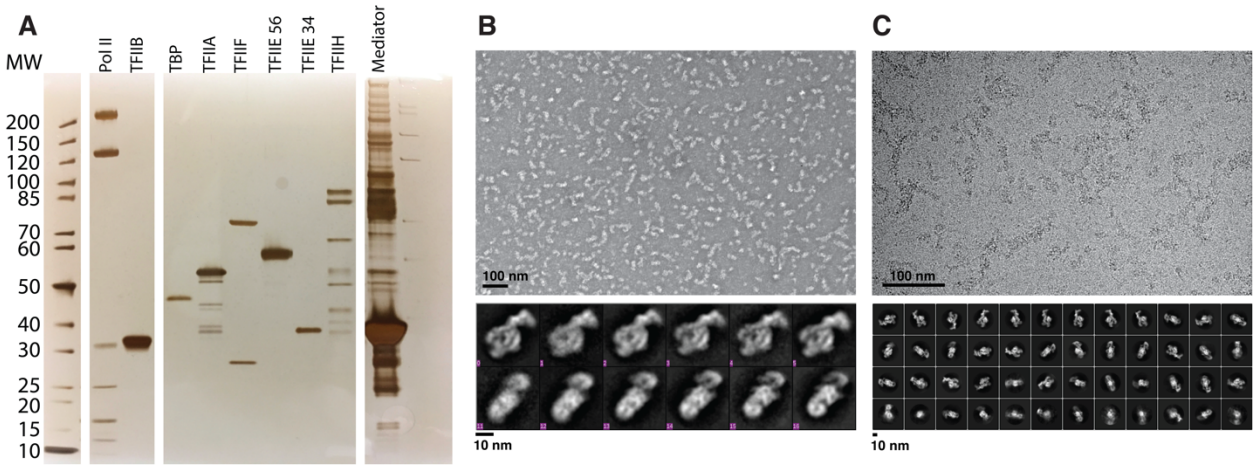

Fig. S1.

Assembly of Med-PIC. A) SDS PAGE gel of purified Med-PIC factors. Lanes were rearranged for clarity. B-C) Representative negative stained (B) and cryogenic (C) electron micrograph and class averages show intact Med-PIC complexes with multiple views.

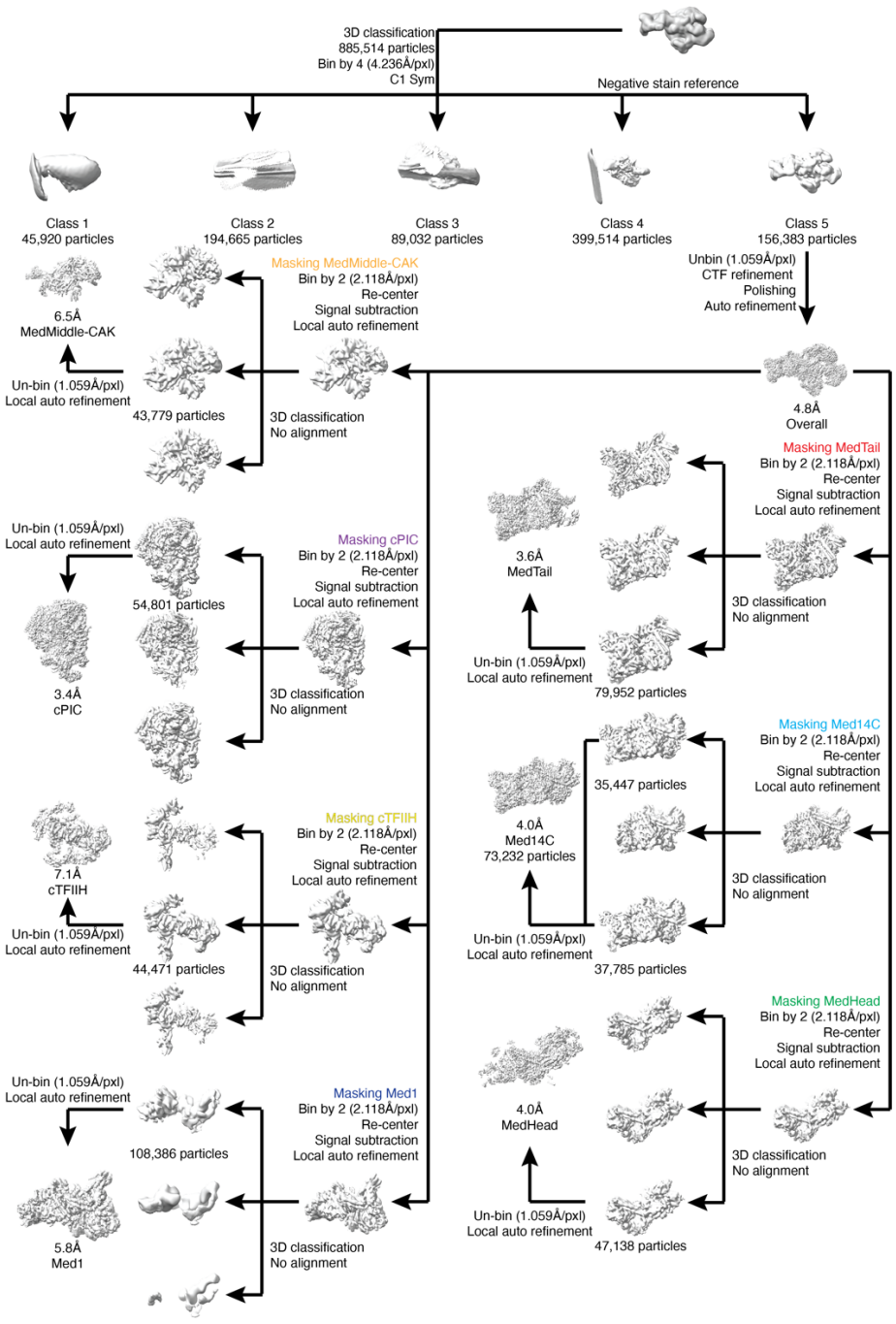

Fig. S2.

Med-PIC cryo-electron microscopy processing pipeline. An initial 3D classification of all particles resulted in a single class that could be refined to an overall resolution of 4.8 Å. Focused local refinements on subcomplexes were performed for cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, Med1, and MedTail by binning by a factor of 2, re-centering, and signal subtracting away the rest of the complex. 3D classification without alignment was performed to select a subset of particles that were unbinned and refined to 3.4, 7.1, 4.0, 6.5, 4.0, 5.8, and 3.6 Å, respectively.

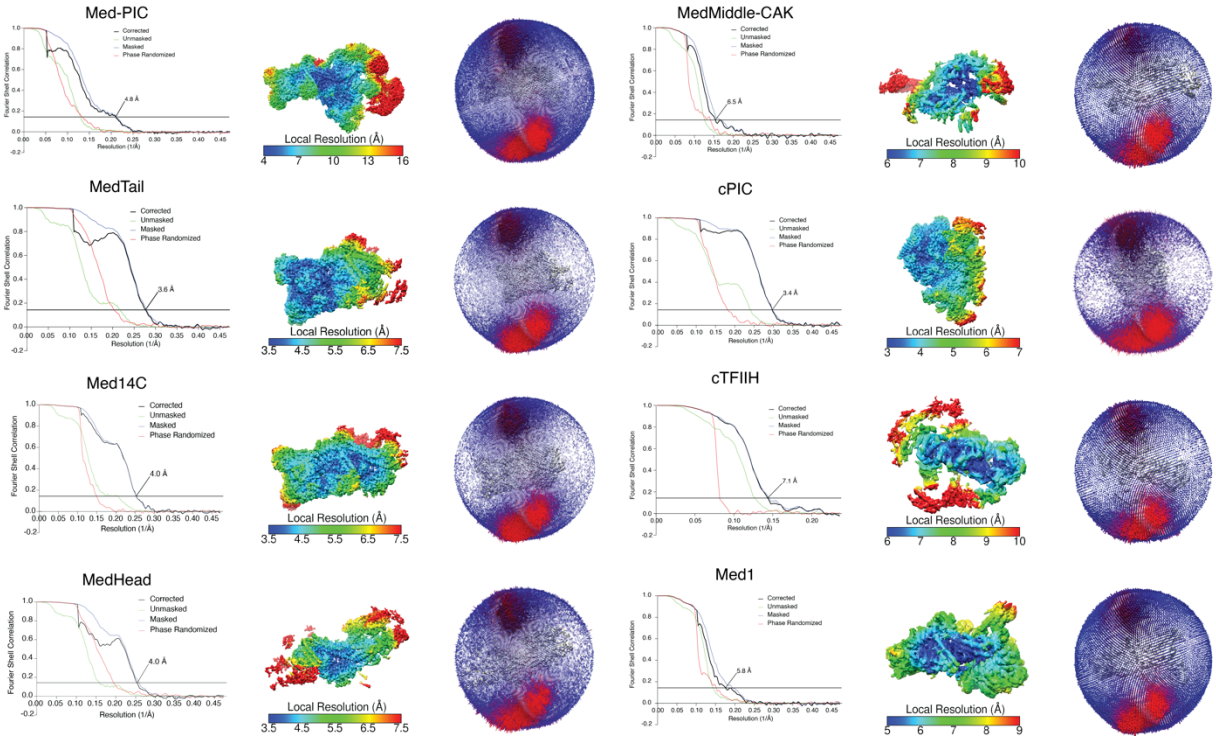

Fig. S3.

Map resolution, local resolution maps, and angular distributions for the full Med-PIC map and cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1 focused refinements. The overall reconstruction shows significant variation in local resolution that improves in each subcomplex following focused refinement.

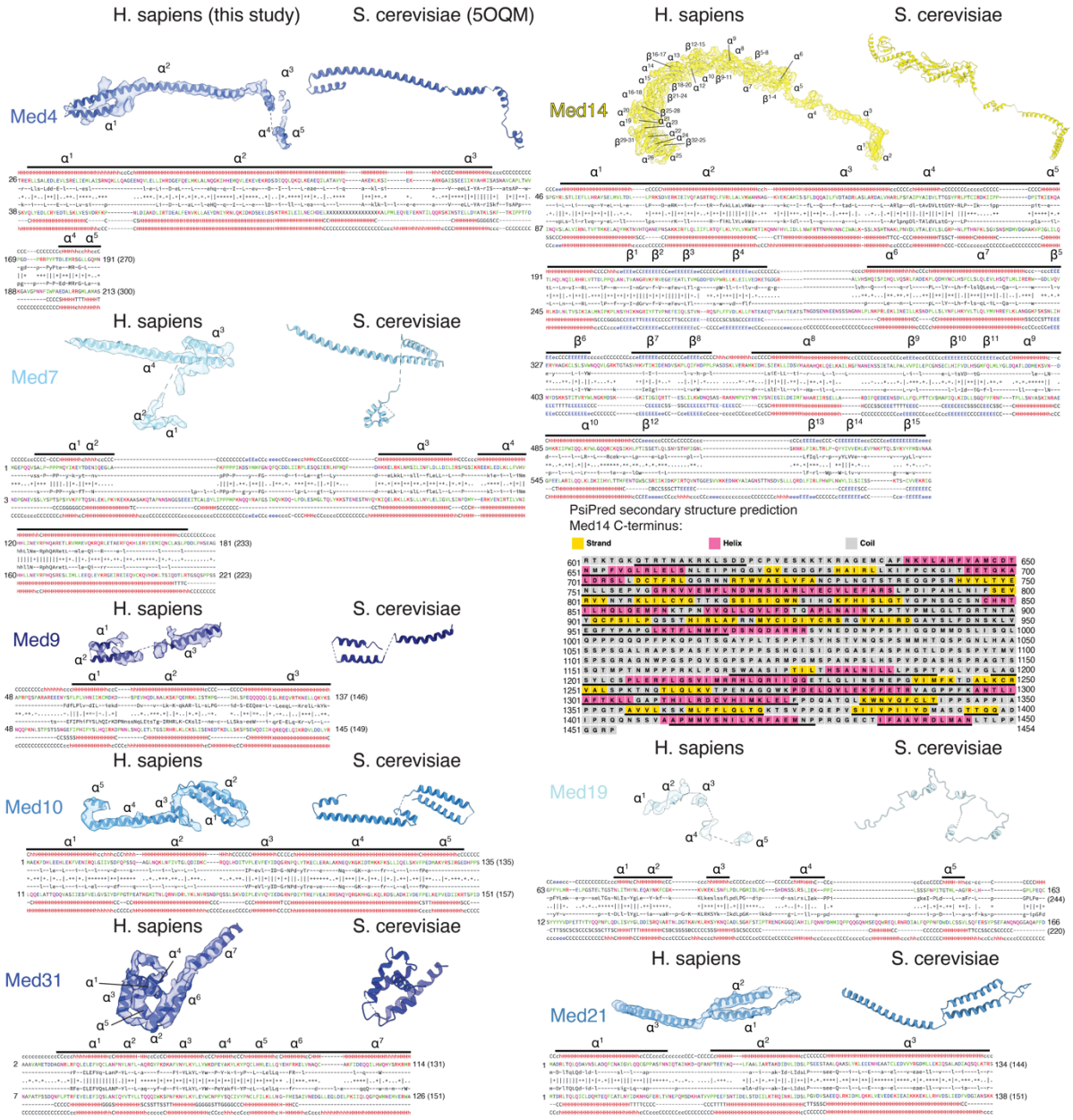

Fig. S4.

Comparison of MedMiddle subunits. Each MedMiddle subunit is shown with the corresponding homology model from *S. cerevisiae* (PDB:5OQM) and the sequence alignment used to build the human model. Colored bars above the sequence alignment show portions for which models were built, excluding any small missing loops. Sequence alignments only include those portions of each subunit for which sequence alignment was successful. Secondary structure prediction is shown for the C-terminal extension of Med14 not found in yeast using PsiPred.

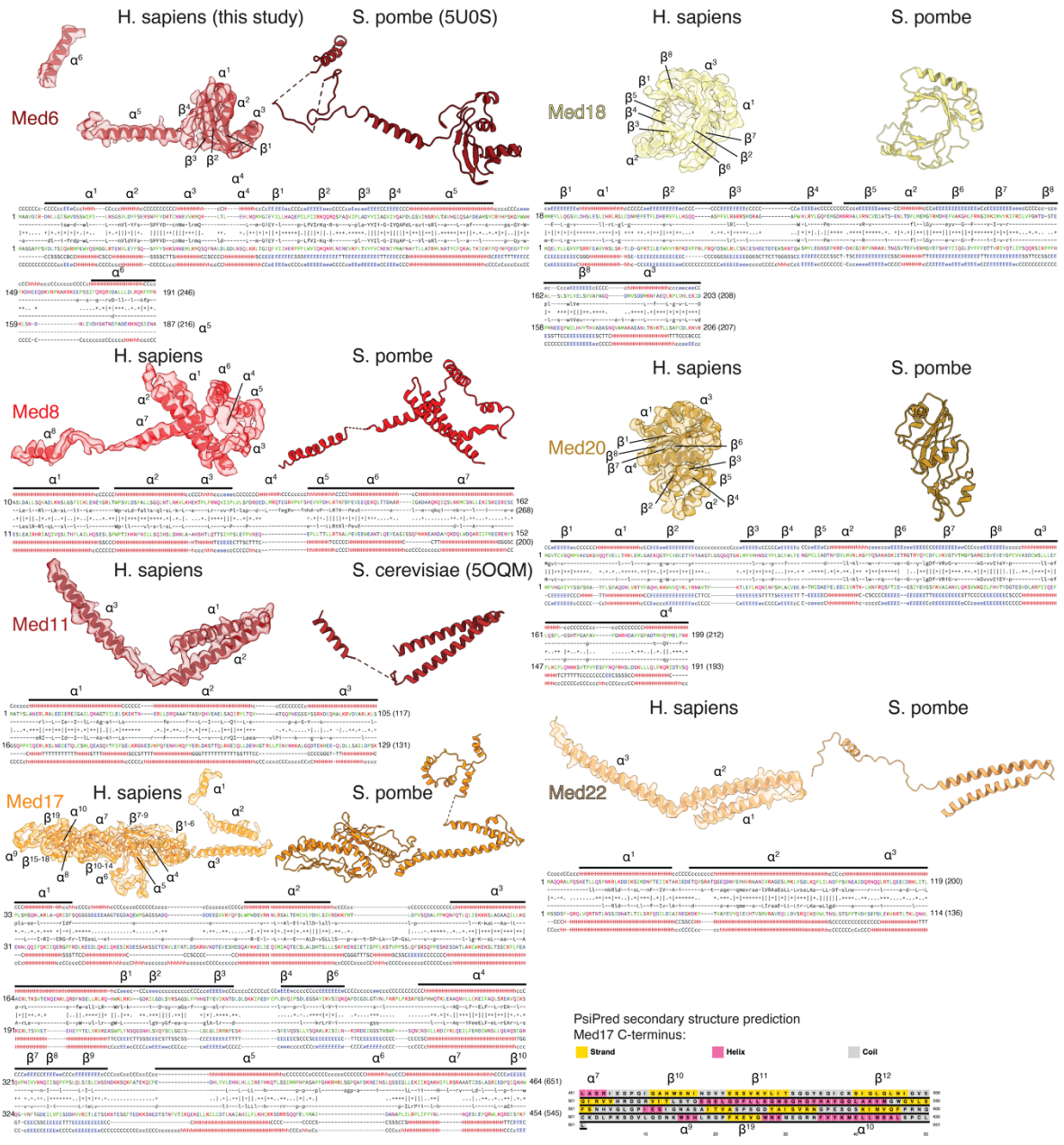

Fig. S5.

Comparison of conserved MedHead subunits. Each MedHead subunit is shown with the corresponding homology model from either *S. pombe* (PDB:5U0S) or *S. cerevisiae* (PDB:5OQM) and the sequence alignment used to build the human model. Colored bars above the sequence alignment show portions for which models were built, excluding any small missing loops. Sequence alignments only include those portions of each subunit for which sequence alignment was successful. Secondary structure prediction is shown for the C-terminal extension of Med17 not found in yeast using PsiPred.

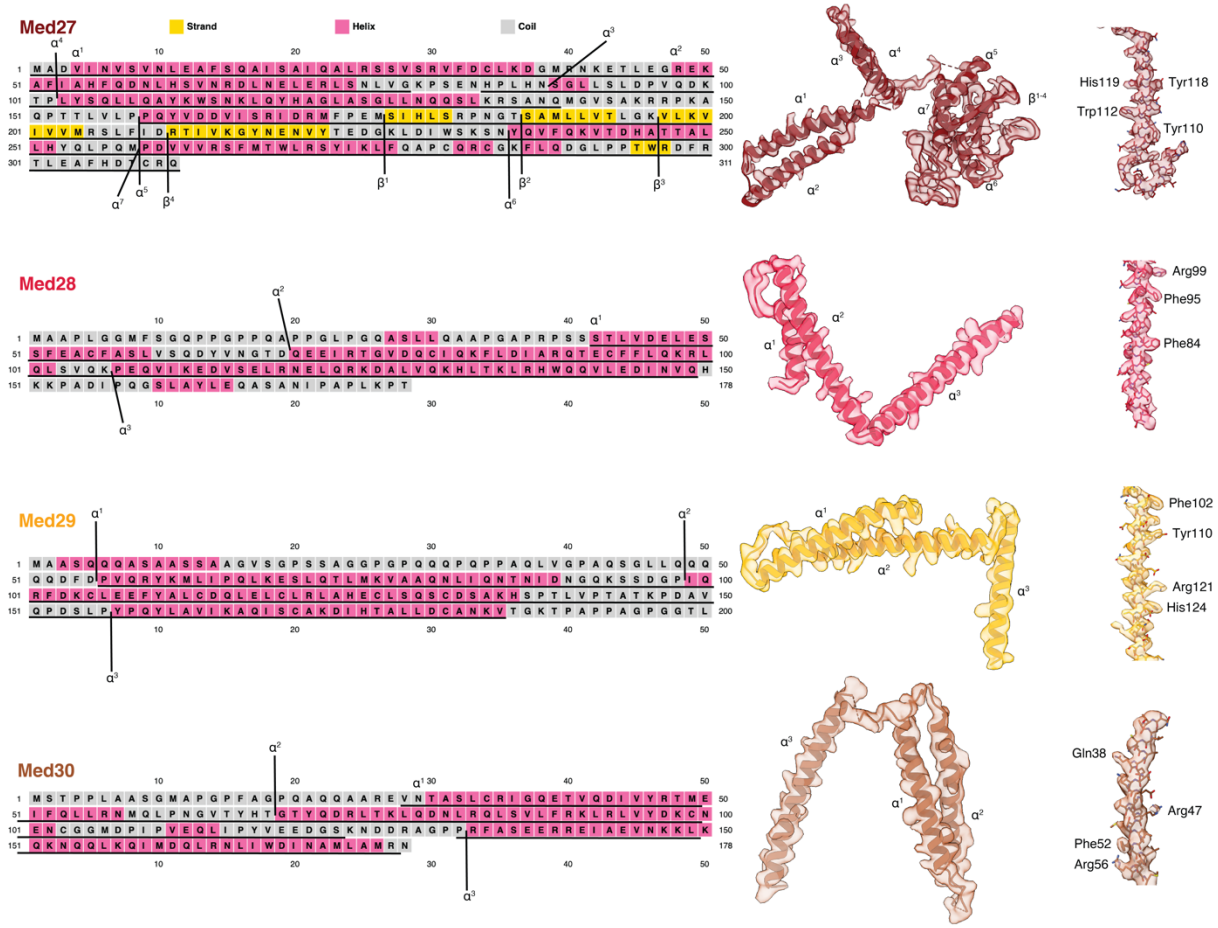

Fig. S6.

Models of MedHead subunits Med27, Med28, Med29, and Med30. Secondary structure prediction using PsiPred closely agrees with secondary structure visible in the density maps and allowed the building of atomic models for each subunit. Secondary structure elements are labeled on both the protein sequence and the models. Representative model-to-map fits (far right) show clear density for bulky side chains that enabled atomic model building.

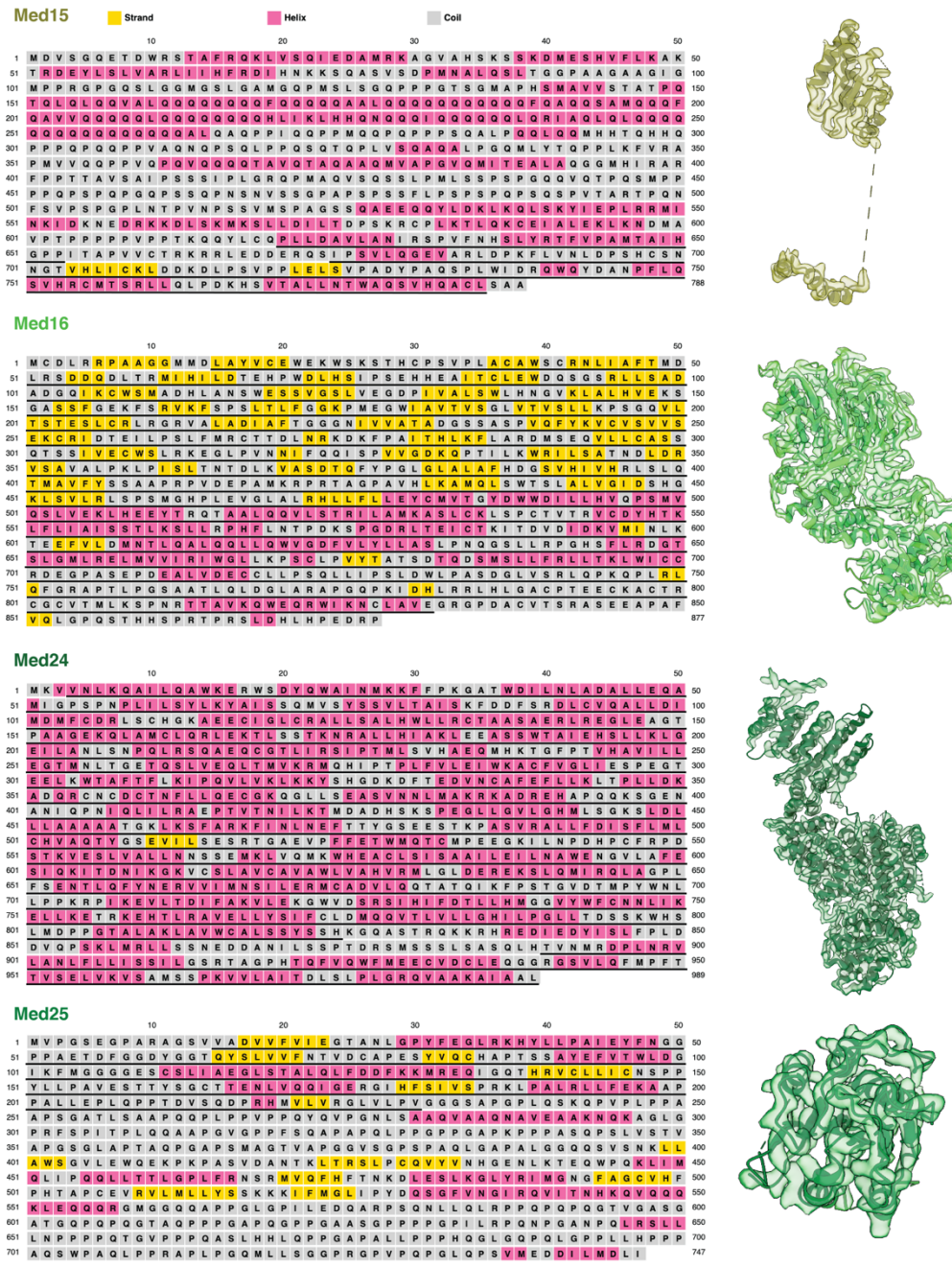

Fig. S7.

Models of MedTail. Secondary structure prediction and model-to-map fit for each subunit of MedTail.

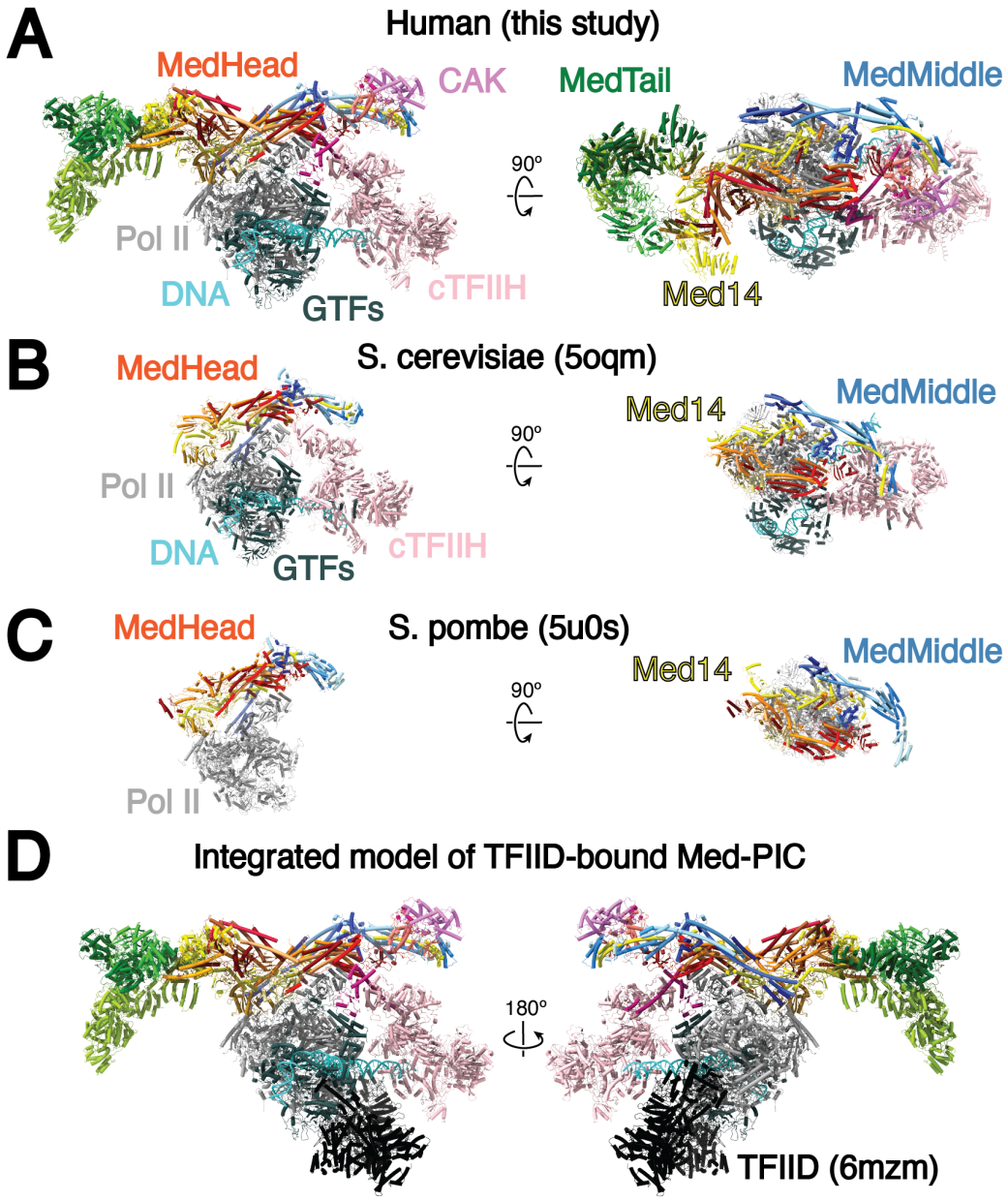

Fig. S8.

Comparison of Med-PICs between yeast and humans and integrated model of a TFIID-bound Med-PIC. A) Structure of the human Med-PIC as shown in Figure 1. B-C) Structure of *S. cerevisiae* and *S. pombe* Med-PIC complexes. The *S. cerevisiae* complex was reconstituted without MedTail, which the *S. pombe* complex is missing the GTFs. The overall architecture of the complexes does not differ dramatically between species. D) Integrated model of a TFIID-bound Med-PIC complex created by aligning the DNA from the TFIID-TFIIA-DNA complex (PDB: 6MZM) with the Med-PIC complex. No clashes are observed in this complex, suggesting no changes in Med-PIC architecture would be necessary to accommodate TFIID binding. TFIIA and the DNA from the TFIID-TFIIA-DNA complex are hidden for simplicity.

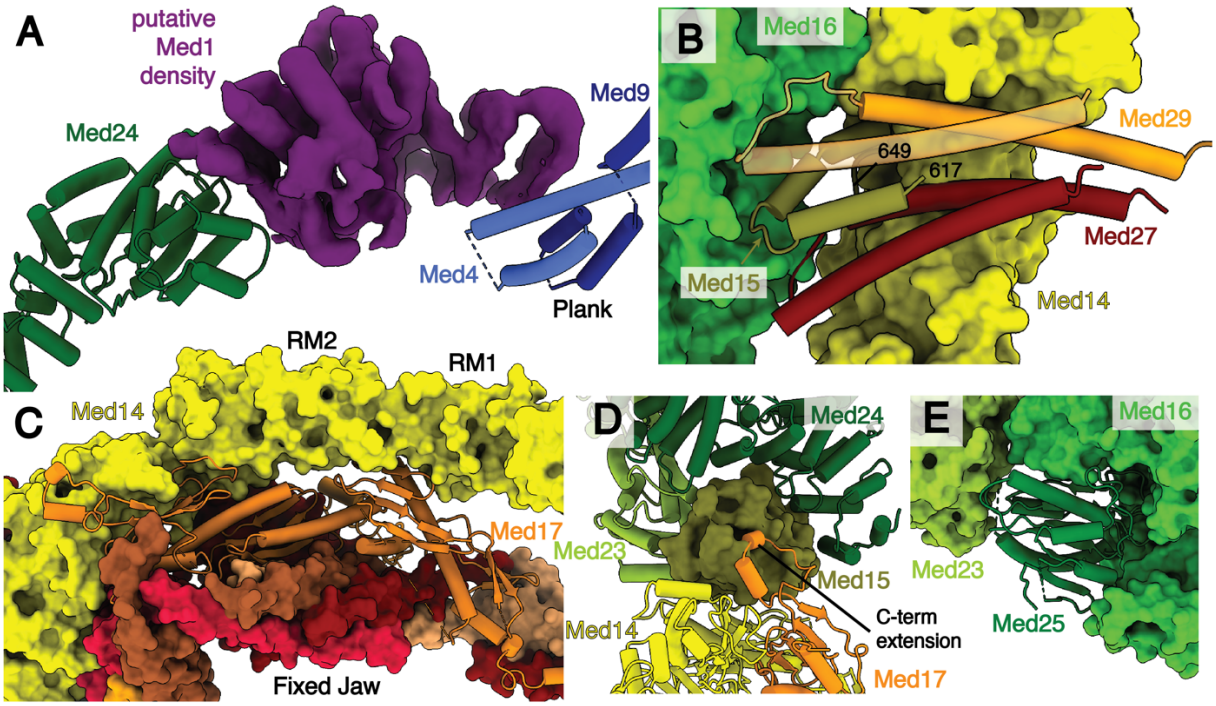

Fig. S9.

Key interfaces in Med-PIC. A) Putative density for Med1 is located between the N-terminus of Med24 and the end of the plank domain formed by Med4 and Med9. B) Two helices of Med15, residues 617 to 649 are sandwiched between Med27 and Med29 and together form one of the two main interfaces between MedHead and MedTail. Models for Med14 and Med16 are shown as surface representations. C) Med17 stabilizes the fixed jaw on one face and interacts with the RM1 and RM2 domains of Med14 on the other face. Models are shown as either ribbon (Med17) or surface (all other subunits). D) A C-terminal extension of Med17 interacts with the RWD domain of Med15 (surface), which is located in a pocket formed by the MedTail subunits Med23 and Med24. E) The vWA domain of Med25 is located in a pocket formed by Med16 and Med23 (surface representations).

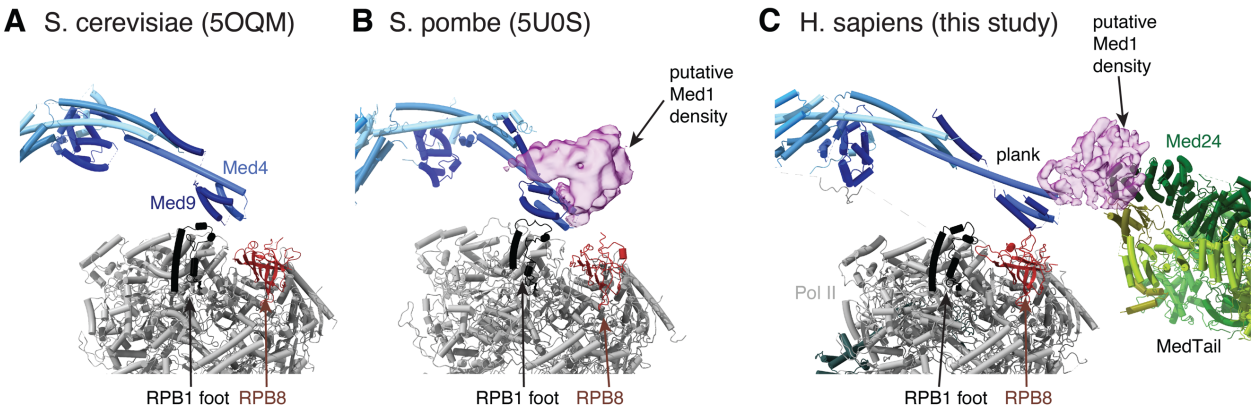

Fig. S10.

Comparison of Mediator plank domain interactions with Pol II. A) The *S. cerevisiae* plank domain interacts with the RPB1 foot (black). The Med1 subunit is not present in the *S. cerevisiae* structure. B) The *S. pombe* plank domain interacts with both the RPB1 foot (black) and putative Med1 density, suggesting that the presence of Med1 is not sufficient to break plank-foot interactions. C) The human Mediator plank domain does not interact with the RPB1 foot (black). Instead, Med1 is stabilized by interactions with Med24 of MedTail, which pulls the plank along with it. Models are colored as in Figure 1.

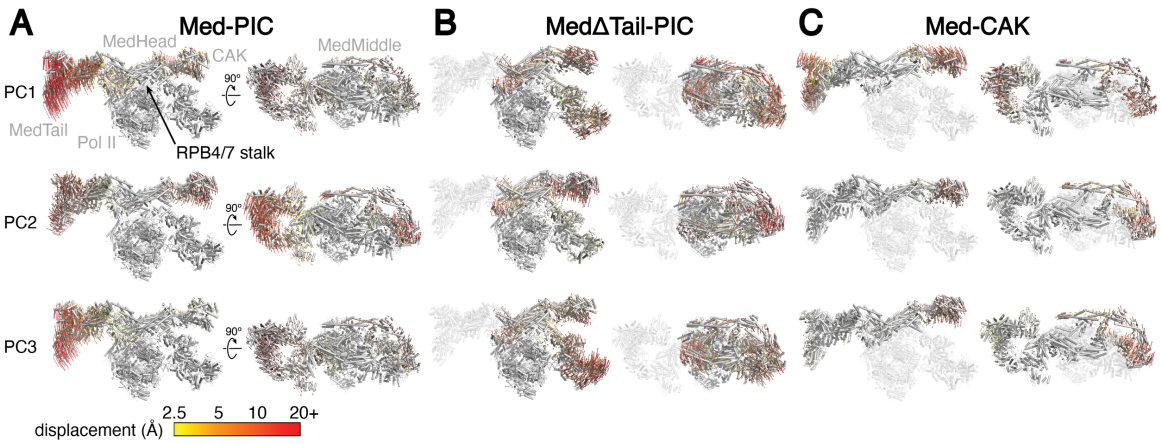

Fig. S11. 3D variability analysis of Med-PIC. A) Top three principal components (PCs) of movement within Med-PIC. MedTail and the CAK module undergo the largest displacements in Med-PIC. B) Top three PCs of movement within MedΔTail-PIC. When isolating movement from MedTail, the rotation of MedHead-MedMiddle-CAK and TFIIH relative to the cPIC is more readily visible. C) Top three PCs of movement within Med-CAK. The movement of MedTail and MedMiddle-CAK is largely independent of each other. PC1 shows that the interface between MedHead and MedTail can act as a hinge, which is reasonable given how small the interface is between the two modules. Movements are colored from yellow (small) to red (large).

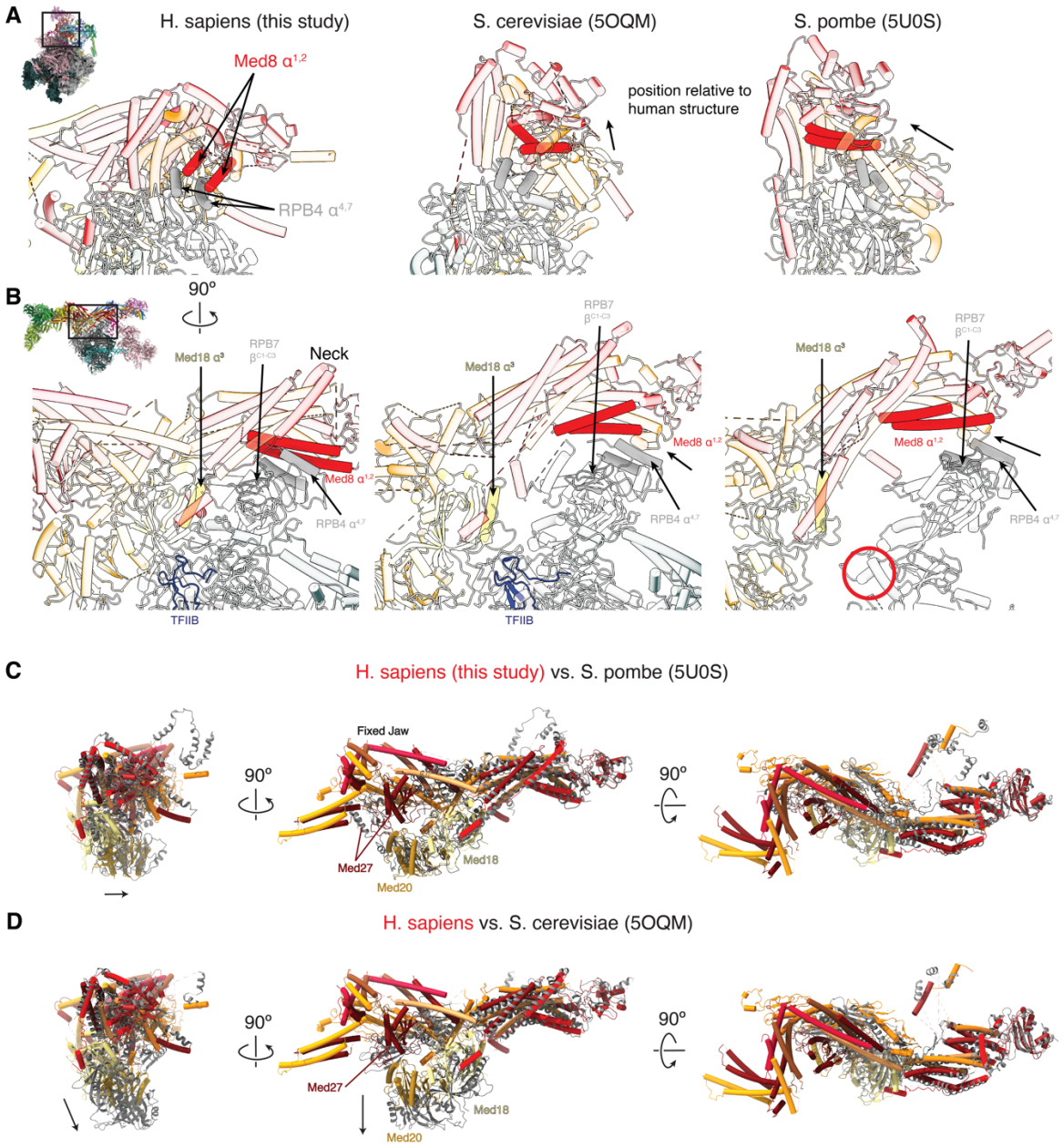
A
H. sapiens (this study)
S. cerevisiae (5OQM)
S. pombe (5U0S)
Med8 α1,2
RPB4 α4,7
position relative to human structure
B
90°
Med18 α3
RPB7 βC1-C3
Neck
Med8 α1,2
RPB4 α4,7
TFIIB
C
H. sapiens (this study) vs. S. pombe (5U0S)
90°
Fixed Jaw
Med27
Med20
Med18
D
H. sapiens vs. S. cerevisiae (5OQM)

Fig. S12.

Comparison of MedHead positions relative to Pol II. A) Human MedHead makes extensive contacts with the stalk of Pol II with helices α^1 and α^2 from Med8 stacking on top of helices α^4 and α^7 of RPB4 and forming an extensive interface between MedHead and the RPB4/7 stalk. In *S. cerevisiae* and *S. pombe,* MedHead slides towards MedTail. B) This movement of MedHead from human to *S. cerevisiae* to *S. pombe* results in an uncovering of the RNA exit channel, occupied by TFIIB (blue ribbon), in the first two structures. The distance between Med18 α^3 and RPB7 $\beta^{C1\text{-}C3}$, which define this gap, is highlighted as opaque ribbon. A red circle denotes where the exit channel is in the *S. pombe* structure. The view of A and B relative to the full complex is shown on the far left. C-D) Superimposing human MedHead (colored/tube) with the spMedHead (C, gray/ribbon, PDB:5U0S) and scMedHead (D, gray/ribbon, PDB:5OQM) structures show that they align very well with just subtle movements of the mobile jaw (Med18 and Med20) between species. In humans and *S. pombe*, the presence of the Med27 subunit stabilizes the mobile jaw, but its absence in *S. cerevisiae* likely allows the sinking of the mobile jaw away from the fixed jaw. Models are colored as in Figure 1.

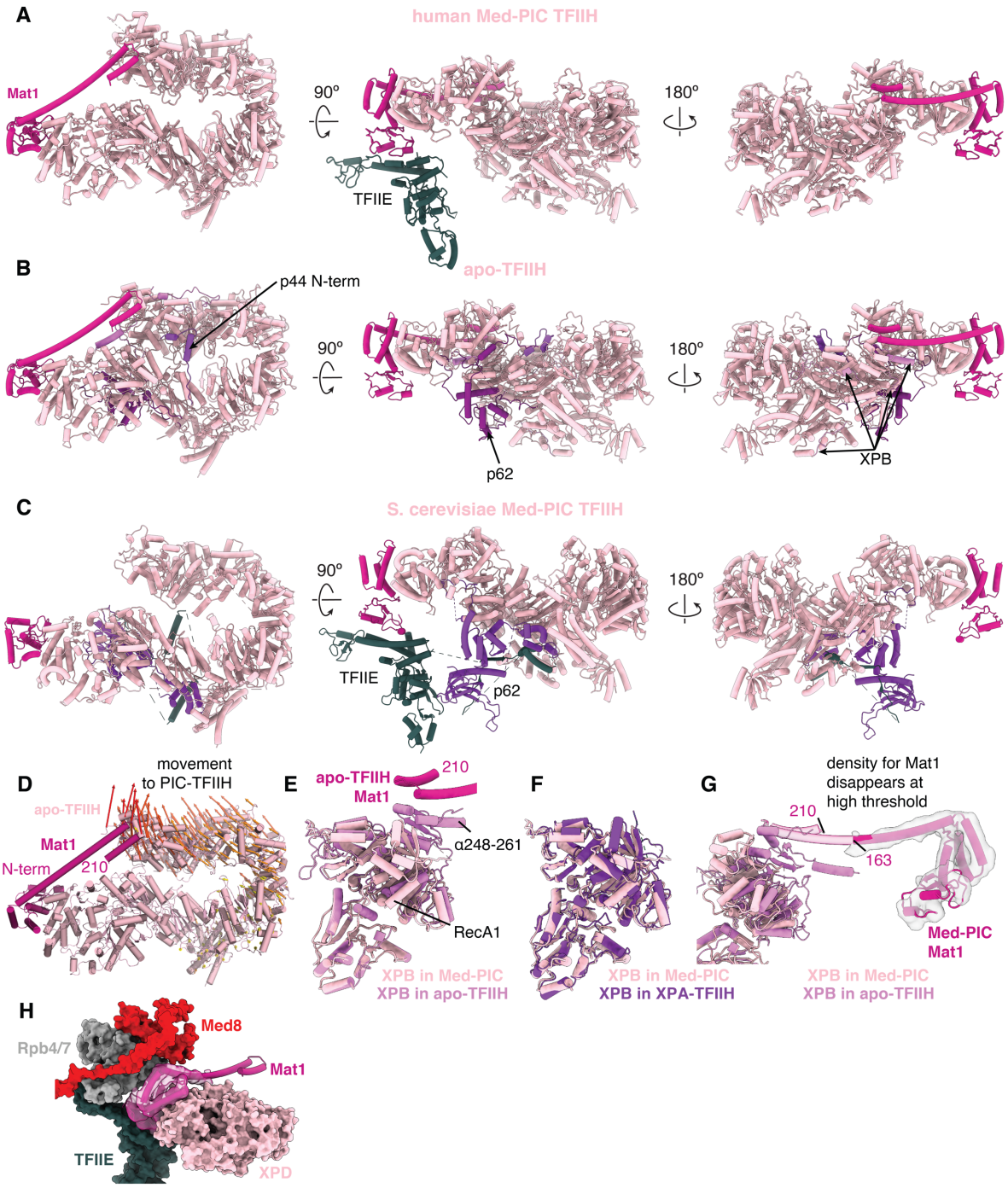
A
human Med-PIC TFIIH
Mat1
90°
TFIIE
180°
B
apo-TFIIH
p44 N-term
90°
p62
180°
XPB
C
S. cerevisiae Med-PIC TFIIH
90°
TFIIE
p62
180°
D
movement
to PIC-TFIIH
apo-TFIIH
Mat1
N-term
210
E
apo-TFIIH
Mat1
210
α248-261
RecA1
XPB in Med-PIC
XPB in apo-TFIIH
F
XPB in Med-PIC
XPB in XPA-TFIIH
G
density for Mat1
disappears at
high threshold
210
163
Med-PIC
Mat1
XPB in Med-PIC
XPB in apo-TFIIH
H
Med8
Rpb4/7
Mat1
TFIIE
XPD

Fig. S13.

Comparison of cTFIIH structure between human Med-PIC (A), apo-TFIIH (B), and scMed-PIC (C). Structure of TFIIH within Med-PIC exhibits a much more open structure than apo-TFIIH. Models built in apo-TFIIH that are absent in the Med-PIC map are shown in light violet (XPB), medium violet (p44), and dark violet (p62). Mat1 and cTFIIH are colored as in Fig. 1. TFIIE is shown in dark slate gray and interacts with the N-terminus of Med1. The structure of TFIIH within *S. cerevisiae* Med-PIC also adopts the more open shape seen in the human Med-PIC but has more extensive interactions between TFIIE and p62 that form a second stabilizing interface that is absent in the human structure. D) Structure of apo-TFIIH showing movements within the complex that accompany incorporation into Med-PIC. Length of movement is colored from yellow to red. E) Comparison of the structure of XPB in Med-PIC versus apo-TFIIH (PDB:6NMI) shows a rotation of the DRD of XPB towards the DNA, breaking contacts with Mat1. Helix 248-261, which forms contacts with Mat1 in apo-TFIIH, is not visible in the Med-PIC structure. F) Comparison of the structure of XPB in Med-PIC versus the XPA-TFIIH-DNA structure. No notable difference is seen between the two structures. G) Density for the Mat1 long helix disappears at a high threshold in the cTFIIH map due to the loss of Mat1-XPB contacts. H) The N-terminus of Mat1 is stabilized through interactions with XPD, RPB4/7, Med8, and TFIIE. When shown at a realistic threshold, density for Med1 (transparent surface) is only visible for portions stabilized by these subunits. Models are colored as in Figure 1.

Table S1. Cryo-EM data collection, refinement, and validation statistics.

	#1 Med-PIC (EMDB: 23255, PDB: 7LBM)	#2 cPIC (EMDB: 23256)	#3 cTFIIH (EMDB: 23257)	#4 MedHead (EMDB: 23258)	#5 MedMiddle-CAK (EMDB: 23259)	#6 Med14C (EMDB: 23260)	#7 MedTail (EMDB: 23261)	#8 Med1 (EMDB: 23262)
Data collection and processing								
Microscope	Titan Krios-3							
Voltage (kV)	300							
Camera	Gatan K3							
Magnification	30k							
Pixel size at detector (Å/pixel)	1.059							
Total electron exposure (e–/Å^2)	~31							
Exposure rate (e$^-$/pixel/sec)	15							
Number of frames collected during exposure	45							
Defocus range (μm)	-2.0 to -4.0							
Automation software	SerialEM							
Energy filter slit width	N/A							
Micrographs collected (no.)	19,881							
Micrographs used (no.)	15,978							
Total extracted particles (no.)	885,514							
Model composition								

Reconstruction	Med-PIC	cPIC	cTFIIH	MedHead	MedMiddle-CAK	Med14C	MedTail	Med1
Refined particles (no.)	156,383	54,801	44,471	47,138	43,779	35,447	79,952	108,383
Final particles (no.)	156,383	54,801	44,471	47,138	43,779	35,447	79,952	108,383
Point-group or helical symmetry parameters	C1	C1	C1	C1	C1	C1	C1	C1
Resolution (global, Å)								
FSC 0.5 (unmasked/masked)	9.82/7.4	7.1/3.82	9.22/8.25	7.92/4.45	9.28/7.82	7.8/4.5	7.7/4.12	8.28/7.25
FSC 0.143 (unmasked/masked)	7.68/4.8	4.15/3.4	7.9/7.1	6.33/4.0	7.9/6.5	5.25/4.0	4.6/3.6	6.98/5.8
Resolution range (local, Å)	3.4 to >10	3 to 7	6 to >10	3.6 to 8	6 to >10	3.6 to 8	3.3 to 7.5	5 to 9
Map sharpening B factor (Å^2)	-126	-76	-248	-92	-183	-92	-81	-221
Map sharpening methods	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
Model composition								
Protein	15,877							
Ligands	19							
RNA/DNA	128							
Model Refinement								
Refinement package -real or reciprocal space	Phenix Real							
Model-Map scores								
-CC	0.46							
-Average FSC								
B factors (Å^2)								
Protein residues	80.23							
Ligands	85.15							
RNA/DNA	71.22							
R.m.s. deviations from ideal values								
Bond lengths (Å)	0.009							
Bond angles (°)	1.236							
Validation								
MolProbity score	2.57							

CaBLAM outliers	5.7
Clashscore	30.04
Poor rotamers (%)	0.71
C-beta deviations	0.03
Ramachandran plot	
Favored (%)	87.88
Outliers (%)	0.65

Table S2. Model building starting models and model confidence.

Domain	Chain ID	Map	Prior knowledge	Level of confidence
cPIC	A-T	cPIC	human cPIC PDB:5IYA human TFIIE PDB: 5GPY	Atomic level
$_{MED}$CTD	A	MedHead	PDB:4GWQ	Backbone trace
$_{CDK}$CTD	A	MedMiddle-CAK	PDB: 1QMZ	Backbone trace
TFIIH-CAK	d-f	MedMiddle-CAK	PDB: 6XBZ	Backbone trace
cTFIIH	W-c	MedMiddle-CAK	PDB: 6NMI	Backbone trace
Med1		Med1	-	-
Med4	s	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med6	g	MedHead	PDB:5U0S	Atomic level
Med7	t	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med8	h	MedHead	PDB:5U0S	Atomic level
Med9	u	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med10	v	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med11	i	MedHead	PDB:5U0S	Atomic level
Med14	r	MedMiddle-CAK	PDB:5OQM	Backbone trace
		MedHead	PDB:5U0S	Atomic level
		Med14C	-	Atomic level
Med15	z	MedTail	PDB:2EBK	Atomic level
Med16	0	MedTail	PDB:2MZH	Atomic level
Med17	j	MedHead	PDB:5U0S	Atomic level
Med18	k	MedHead	PDB:2HZM	Atomic level
Med19	w	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med20	l	MedHead	PDB:2HZM	Atomic level
Med21	x	MedMiddle-CAK	PDB:5OQM	Backbone trace
Med22	m	MedHead	PDB:5U0S	Atomic level
Med23	1	MedTail	PDB:6H02	Atomic level
Med24	2	MedTail	-	Atomic level
Med25	3	MedTail	PDB: 2KY6	Atomic level
Med26		-	-	-
Med27	n	Med14C	-	Atomic level
Med28	o	Med14C	-	Atomic level
Med29	p	Med14C	-	Atomic level
Med30	q	Med14C	-	Atomic level
Med31	y	MedMiddle-CAK	PDB:5OQM	Backbone trace

Movie S1.

Structure of the human Mediator-bound transcription pre-initiation complex.

Movie S2.

3D variability analysis of Med-PIC. Top three principal components (rows 1-3) shown of the front (left) and top (right) views of Med-PIC. Mediator and the CAK module of TFIIH largely move as a rigid body relative to the PIC.

Movie S3.

3D variability analysis of MedΔTail-PIC. Top three principal components (rows 1-3) shown of the front (left) and top (right) views of MedΔTail-PIC. MedHead, MedΔTail rotates relative to the PIC, and that movement can correlate with the movement of cTFIIH relative to the cPIC.

Movie S4.

3D variability analysis of Med-CAK. Top three principal components (rows 1-3) shown of the front (left) and top (right) views of Med-CAK. Mediator can flex, either causing lifting of MedHead and MedMiddle or bringing MedMiddle and MedTail closer together.

References

41. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481--486 (2013).
42. D. J. Taatjes, R. Tjian, Structure and function of CRSP/Med2: a promoter-selective transcriptional coactivator complex. *Molecular Cell* **14**, 675--683 (2004).
43. C. Suloway *et al.*, Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41--60 (2005).
44. A. Patel, D. Toso, A. Litvak, E. Nogales, Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids. *bioRxiv*, 2021.2003.2008.434344 (2021).
45. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
46. T. Ogura, K. Iwasaki, C. Sato, Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *Journal of Structural Biology* **143**, 185--200 (2003).
47. G. C. Lander *et al.*, Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology* **166**, 95--102 (2009).
48. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *Journal of Structural Biology* **166**, 205--213 (2009).
49. A. Rohou, N. Grigorieff, "CTFFIND4: Fast and accurate defocus estimation from electron micrographs," (2015).
50. G. Tang *et al.*, EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38--46 (2007).
51. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).

52. K. Zhang, Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1--12 (2016).
53. R. Henderson *et al.*, Outcome of the first electron microscopy validation task force meeting., 205--214 (2012).
54. R. Sanchez-Garcia *et al.*, DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv*, 2020.2006.2012.148296 (2020).
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).
56. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).
57. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
58. K. Miwa *et al.*, Crystal Structure of Human General Transcription Factor TFIIE at Atomic Resolution. *Journal of molecular biology* **428**, 4258-4266 (2016).
59. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
60. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
61. L. Zimmermann *et al.*, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of molecular biology* **430**, 2237-2243 (2018).
62. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: a protein secondary structure prediction server. *Nucleic acids research* **43**, W389-394 (2015).
63. B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5 6 1-5 6 37 (2016).
64. L. Lariviere *et al.*, Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* **13**, 895-901 (2006).
65. D. Monte *et al.*, Crystal structure of human Mediator subunit MED23. *Nat Commun* **9**, 3389 (2018).
66. M. Taschner, A. Mourao, M. Awasthi, J. Basquin, E. Lorentzen, Structural basis of outer dynein arm intraflagellar transport by the transport adaptor protein ODA16 and the intraflagellar transport protein IFT46. *J Biol Chem* **292**, 7462-7473 (2017).
67. R. T. Kidmose *et al.*, Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526-531 (2019).
68. B. J. Greber, D. B. Toso, J. Fang, E. Nogales, The complete structure of the human TFIIH core complex. *Elife* **8**, (2019).