NORTHWESTERN UNIVERSITY

Structural Studies of RNA Polymerase II and RNA Polymerase III Transcription Initiation

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences

By

Anna Talyzina

EVANSTON, ILLINOIS

September 2023

ABSTRACT

Transcription plays a pivotal role in the transfer of genetic information within living organisms. It serves as the initial step in gene expression, allowing cells to convert the instructions encoded in their DNA into functional molecules. Eukaryotic transcription initiation is a complex and dynamic process that requires joint efforts from many proteins. In this work I explored the intricate mechanisms underlying this process, employing state-of-the-art cryo-electron microscopy (cryo-EM) and single-molecule techniques to visualize the interplay of molecules involved in RNA Polymerase (Pol) II and III transcription initiation.

Pol II synthesizes messenger RNA (mRNA) molecules, ultimately enabling the expression of protein-coding genes. The Pol II coactivator Mediator is recruited by transcription factors, facilitating the assembly of the preinitiation complex (PIC) and stimulating phosphorylation of the Pol II C-terminal domain (CTD) by the TFIIH cyclin-activated kinase (CAK) module. I present the cryo-EM structure of the human Mediator-bound PIC. The precise orientation of the CAK within the Mediator-PIC is revealed, with clear density for the Pol II CTD in the active site. A second CTD binding site between Mediator Head and Middle demonstrates how Mediator positions the rest of the CTD for phosphorylation by CDK7. Many regions of Mediator that interact with transcription factors are flexibly tethered, facilitating its assembly. The structure also provides key insights into the conformational landscape of Mediator relative to the PIC.

Pol III is responsible for transcribing 5S rRNA, tRNAs, and other short non-coding RNAs. Its recruitment to the 5S rRNA promoter requires transcription factors TFIIIA, TFIIIC, and TFIIIB. I use cryo-EM to visualize the *S. cerevisiae* complex of TFIIIA and TFIIIC bound to the promoter. Partial TFIIIB binding further stabilizes the DNA, resulting in the full-length 5S rRNA gene

wrapping around the complex. Single-molecule FRET (smFRET) experiments reveal that the DNA undergoes both sharp bending and partial dissociation on a slow timescale, consistent with the model predicted from our cryo-EM results. These findings provide new insights into the mechanism of how the PIC assembles on the 5S rRNA promoter, a crucial step in Pol III transcription regulation.

ACKNOWLEDGEMENT

I am grateful to so many people that were helping me along this journey. Good science always depends on communication, and research is impossible without exchange of ideas and collaboration.

The Mediator-PIC project was done in collaboration with Tjian lab at the University of California, Berkeley. I want to thank Carla Inouye for purifying Mediator and other proteins, crucial for this project. I want to thank Ryan Abdella, my co-first author on the manuscript, who started the project in our lab and later on guided me through model building and manuscript preparation. I also want to thank Siyu Chen, who helped me with data processing and taught me all the tricks. The TFIIIC project was initiated by Yan Han, a postdoc in our lab. He designed and tested a DNA construct, that helped us capture this interesting step in RNA Pol III transcription initiation. I also want to thank our collaborators Chiranjib Banerjee and Reza Vafabakhsh at Northwestern, who worked on the smFRET part of the project.

Thank you to all present and past members of our lab for your support and advice. Thank you for the warm welcome to the lab, for teaching and mentoring me. I especially want to thank Sue Fishbain for sharing lab and life advice, and chocolate. I very thankful to Alex Vogt, Jingfei Xu, and Noah Bradley for our lunch and lab break conversations. I cannot imagine my grad school experience without you. I am grateful to Yuan He, my PhD advisor, for believing in me and guiding me through my projects. The enthusiasm and support you shared helped me to stay on track and keep solving research puzzles.

Thank you to my thesis committee members, Heather Pinkett, Alfonso Mondragon, Reza Vafabakhsh, for providing valuable feedback on my research progress. I want to thank Cathy

Prullage, Michelle McDonough, and the staff of the IBiS, MBTP, and Molecular Biology department for their support. Many thanks to Jason Pattie and the rest of SBF staff for technical support with the computational cluster and the microscope. Thank you to Janette Myers and the rest of PNCC team for collecting our cryo-EM data and letting me shadow the process.

I want to thank everyone in my IBiS cohort for becoming my first community in the United States and making my time in grad school much more enjoyable. Laura, thank you for being there for me since the day one. Thank you for introducing me to American culture, musicals, and podcasts. Alexis, I am grateful to have so many shared memories from our Indiana, Michigan, and Wisconsin trips. Thank you both for being part of our local wine club ☺ I am also thankful for the opportunity to join the Cook fourth floor teatime group during my first year. Ryan, Chris, Nick, and others, thank you for sharing your grad school experience and advice with me.

Jane and Yana, Anya, Danil, and Maxim, thank you for being my Russian speaking community. You made Chicago feel more like home. Jane and Yana, thank you both for our memorable trips to Zion and Orlando, and our time in Chicago.

My mama, my sisters, my papa, and everyone in my family, I am so grateful to have you. Thank you for believing in me more than I do myself. I would not be here without you. I wish I could visit you more often.

Sergey, my soulmate, my inspiration, my love, thank you. I truly cannot imagine this journey without you. You have been there for all the beautiful and ugly moments, all my wins and losses. Thank you for being ambitious and understanding, thank you for dreaming big and realizing our dreams together. I love you.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER I: Introduction

1.1 Eukaryotic transcription

In eukaryotes, three common DNA-dependent RNA polymerases are responsible for transcription (*1*). Each polymerase is responsible for transcribing different classes of genes and plays a crucial role in various cellular processes. RNA polymerase I (Pol I) primarily transcribes ribosomal RNA (rRNA) genes, contributing about 80% of the RNA mass in the cell (*2*). RNA polymerase II (Pol II) is responsible for transcribing protein-coding genes, producing messenger RNA (mRNA) molecules that serve as templates for protein synthesis. Pol II transcription is a highly regulated process that involves coordinated action of multiple transcription factors (*3*). Pol II also transcribes non-coding RNA genes, such as those encoding small nuclear RNAs (snRNAs) involved in RNA splicing. RNA polymerase III (Pol III) transcribes a variety of non-coding RNA genes, including transfer RNAs (tRNAs), which are essential for protein synthesis, 5S ribosomal RNA (5S rRNA), and small regulatory RNAs such as small nucleolar RNAs (snoRNAs) (*4*). Pol III transcription is characterized by short and highly structured RNA products, that represent the majority of mature RNA molecules in the cell (*2*).

The three polymerases share common composition of the catalytic core and the general mechanism of transcription (*5*). There are three parts of transcription process: initiation, elongation, and termination (Figure 1.1). At the initiation step, the polymerase is being recruited to promoter sequence via a set of general transcription factors (GTFs) (Figure 1.1A). These factors also help the polymerase to open the promoter region and create a transcription bubble. Once the DNA is melted, and the template sequence is exposed, RNA synthesis begins. Nucleotides, one by one, are getting annealed to the template strand and are added to the 3’ end of the synthesized RNA. During this process the polymerase, sometimes accompanied by a set of elongation factors, translocates along the DNA (Figure 1.1B). Eventually, the polymerase encounters a termination signal, which may be conducted by termination factors or termination sequence. During transcription termination, the polymerase and the newly synthesized RNA molecule dissociate from the DNA (Figure 1.1C).

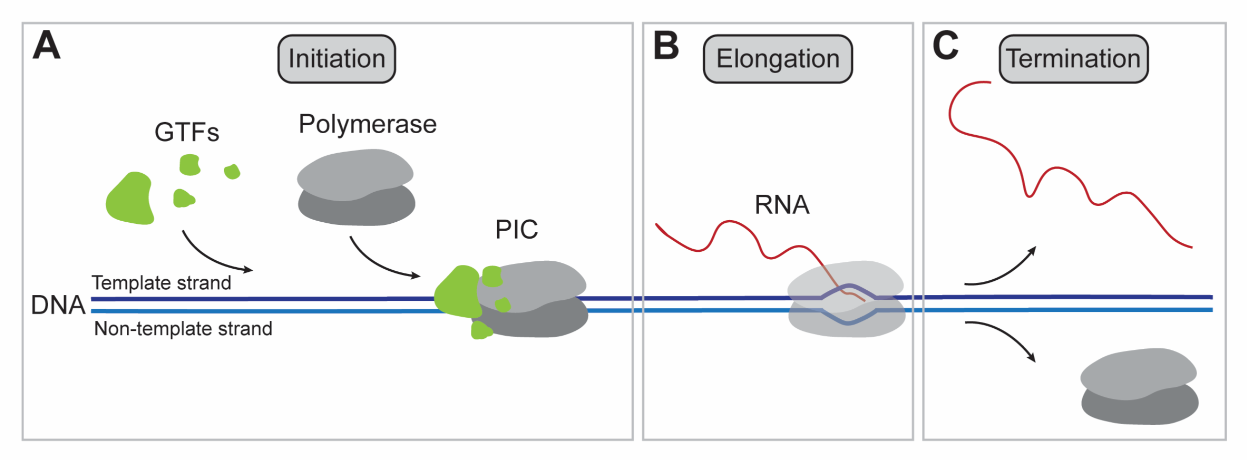

Figure 1.1. Schematic overview of the transcription process. A) Transcription initiation involves recruitment of polymerase and GTFs to promoter. B) During elongation RNA is being synthesized and polymerase translocates along the DNA. C) Polymerase and newly synthesized RNA dissociate from DNA during transcription termination.

Each RNA polymerase uses its own set of transcription factors, promoter sequences, and regulatory elements that dictate their selective transcription. Pol I, Pol II, and Pol III are comprised of 13, 12, and 17 subunits, respectively (*6*). Pol III has the highest number of subunits, while Pol II requires the highest number of GTFs to initiate transcription. This difference represents the two different priorities: Pol III needs lower number of additional factors, which gives the advantage of more rapid and efficient transcription. At the same time, Pol II, which requires a lot of cofactors, can be more precisely controlled, and transcribe more diverse genes. Understanding the distinct roles of RNA polymerases I, II, and III, and their associated transcriptional machinery, provides valuable insights into the regulation of gene expression, cellular differentiation, and organism development. It also helps decipher the molecular basis of diseases linked to transcriptional dysregulation and informs the development of therapeutic strategies targeting transcriptional processes.

1.2 Transcription initiation

Transcription initiation is a highly regulated process, crucial for correct and timely RNA synthesis. All polymerases require a set of GTFs to initiate transcription. Together with the polymerase, these factors bind promoter region and form transcription pre-initiation complex (PIC). PICs are organized differently for each RNA polymerase, reflecting different requirements in terms of transcription rate, localization, and quantity of RNA product. GTFs may assist the polymerase in promoter DNA unwinding. Once the single-strand DNA is exposed, the system transitions into the initially transcribing complex (ITC) and synthesizes a short RNA strand. Then the polymerase escapes the promoter and starts transcription elongation.

Pol I transcription initiation starts with RRN3 factor binding Pol I, which acts as an activator (*7*). RRN3-bound Pol I is being recruited to promoter and interacts with core promoter complex (SL1 in humans and CF in yeast) (*8, 9*). Other essential factors are TATA-binding protein (TBP), UBF in humans, and UAF in yeast (*10, 11*).

In the case of Pol II, initiation is a complex process involving multiple steps. It begins with the recruitment of GTFs and additional factors to promoter (*12*). The PIC then undergoes a series of conformational changes and interactions with additional transcription factors, ultimately leading to the melting of the DNA duplex and the initiation of RNA synthesis (*13*). The details of Pol II transcription initiation will be discussed below.

Pol III initiation also involves the assembly of specific transcription factors (TFIIIA, TFIIIB, and TFIIIC) on DNA. These factors are being recruited by a regulatory sequence, which also represents a part of gene sequence. Another feature of Pol III is ability to be recycled on the same gene and go through multiple rounds of transcription (*14*). Pol III transcription initiation will be further discussed below.

Overall, the organization and composition of PICs reflect the distinct requirements of each RNA polymerase for efficient and specific transcription initiation. Understanding the intricate mechanisms underlying PIC assembly and regulation is essential for unraveling the complexities of gene expression and the precise control of RNA synthesis in eukaryotic cells.

1.3 Pol II and mRNA transcription

Transcription of all mRNA in eukaryotes is carried out by RNA polymerase II (Pol II) (*12*). Pol II cannot by itself locate the transcription start site, open a transcription bubble to expose the template strand, and transition to an elongation state. Instead, DNA-bound transcription factors position the coactivator complex Mediator to facilitate the assembly of the preinitiation complex (PIC), which consists of Pol II and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID (TBP), TFIIE, TFIIF, and TFIIH (*15*). These GTFs help position Pol II to initiate transcription at the correct genomic locus, and TFIIH feeds DNA into the active site of Pol II, generating force against the TBP-TFIIB-TFIIA lobe to unwind the DNA and expose the template strand (*16*). The Mediator complex was first biochemically isolated from crude yeast fraction by Kornberg’s group

(*17*). The Mediator-containing fraction was shown to enable activated transcription and to stimulate basal transcription. The entire Mediator-bound PIC (Med-PIC) is 2.7 MDa in size, contains 56 polypeptides, and represents a particular challenge for structural characterization because of difficulties in obtaining and assembling these complexes, as well as to their inherent flexibility (*18-23*).

The largest subunit of Pol II, RPB1, contains a long, repetitive C-terminal domain (CTD) connected by a flexible linker region to the rest of the subunit (*24*). The CTD consists of 26 repeats in yeast and 52 repeats in humans of the consensus YSPTSPS heptamer sequence. Phosphorylation of the serine at position 5 (Ser5) by cyclin-dependent kinase 7 (CDK7) during transcription initiation leads to the recruitment of the 5′ capping enzymes that are indispensable for growth (*25*). Neither the CTD nor the 80-residue linker is visible in structures of the PIC because of their mobility (*26*). CDK7 is part of the cyclin-activated kinase (CAK) module of TFIIH, together with cyclin-H and Mat1. Mounting evidence indicates that Mediator recruits TFIIH to the PIC and stimulates the CAK module's ability to phosphorylate Ser5 (*27, 28*).

Atomic models of Med-PICs are currently limited to yeast (yMed-PIC) (*20-23*). Mediator is divided into four modules: The head (MedHead) interacts with Pol II, the middle (MedMiddle) primarily serves a structural role, the tail (MedTail) serves as a hub for the binding of transcription factors, and the dissociable kinase module (MedKinase) is found at enhancers and prevents interaction with the PIC (*29, 30*). The existing high-resolution models of yMed-PIC highlight differences in how MedHead interacts with Pol II, suggesting that the interface between Mediator and Pol II is not rigid (*20, 21*). MedHead is capable of binding the CTD, as shown in a cocrystal structure (*31*). Based on the location of the CTD modeled into the full yeast Mediator complex, the CTD also likely serves to stabilize the interface between MedHead and MedMiddle (*22, 23*). Functional and structural studies have identified a minimal core Mediator (cMed), devoid of both MedKinase and MedTail (*32, 33*). Previously solved structures of Mediator have poorly defined density for MedTail, leaving open the question of where most transcription factors bind (*19, 20, 22*). A second highly flexible part of Med-PIC is the CAK module of TFIIH. It has been localized to the end of the hook domain of Mediator, directly above core TFIIH (cTFIIH), to which it remains flexibly tethered (*21, 22*). However, the orientation of the individual subunits within this density has not yet been determined.

1.4 Pol III and 5S rRNA transcription

RNA Polymerase III (Pol III) transcribes various types of short, non-coding, and abundant RNAs from three types of promoters. Only Type I and II promoters are known in yeast. The Type I promoter is found in the 5S ribosomal RNA (rRNA) gene, the Type II promoter is associated with transfer RNA (tRNA) genes, and the Type III promoter is used in U6 small nuclear RNA (snRNA) genes and others (*34-36*). Both Type I and Type II promoters contain internal control regions (ICRs) in the gene body (*37, 38*). The ICRs of the Type I promoter harbor an A-box, an intermediate element (IE), and a C-box, while the Type II promoter consists of an A-box and a B-box (*4*). Generally, Pol III transcription initiation requires transcription factors (TFs), including TFIIIC and TFIIIB (*39*). TFIIIA is a specific TF for Type I promoters and consists of nine zinc-finger (ZF) repeats in *S. cerevisiae* (*40*). TFIIIA was the first gene-specific TF identified in eukaryotes (*37*). It also represents the first cloned and sequenced eukaryotic TF, which led to discovery of ZF motif (*41, 42*). TFIIIA is the first factor that recognizes and binds the 5S rRNA promoter. The large, six-subunit TFIIIC is recruited to the Type I promoter via TFIIIA. In the case of Type II promoters, TFIIIC can directly recognize and bind to A-box and B-box elements, and recruits TFIIIB, positioning it upstream of the transcription start site (TSS) (*43*). TFIIIB, which consists of three subunits – TATA-box binding protein (TBP), TFIIB-related factor (Brf1), and B double prime factor (B”) – may support Pol III transcription alone once stably assembled on the promoter, as shown for *S. cerevisiae* (*43*). Subunits of TFIIIB only form a stable complex when all of them are bound to DNA (*44*). Interestingly, 5S rRNA gene transcription initiation requires both gene-specific (TFIIIA) and general transcription factors (TFIIIC and TFIIIB) (*45*). The six subunits of the TFIIIC complex are organized into two lobes: subunits τ131, τ95, and τ55 form the τA lobe, and subunits τ138, τ91, and τ60 form the τB lobe. The two lobes are proposed to be connected via a flexible linker that helps TFIIIC bind ICRs of different lengths (*46, 47*). While several structures of TFIIIC subcomplexes and domains are solved, including the τA lobe (*48*), the τ131 N-terminal tetra-trico peptide repeats (TPR) array (*46*), the histidine phosphatase domain (HPD) of τ55 (*49*), the τ138 extended winged-helix (eWH) domain (*46*), and a subcomplex of τ60 and τ91 (*50*), the structure of the complete TFIIIC complex remains elusive, possibly due to its high flexibility. Structures of TFIIIB components have been solved as a part of the Pol III transcription pre-initiation complex (PIC) (*51-53*). To date, structures of TFIIIA include ZF 1-3 bound to DNA (*54*), ZF 4-6 bound to 5S rRNA (*55*), and ZF 1-6 bound to 5S rRNA gene (*56*). However, the full-length structure of TFIIIA has not been solved.

Misregulation of Pol III transcription has been linked to cancer (*57-60*), with changes in the expression of TFIIIC subunits being associated with infection and disease (*61*). Several TFIIIC subunits have been found to be overexpressed in ovarian tumors (*62*). Stress conditions have been shown to result in TFIIIC's increased occupancy at tDNAs in yeast and at Alu elements in human cells (*63, 64*). Additionally, research suggests that human TFIIIC bound to extra TFIIIC (ETC) sites may play a role in chromosome organization, as shown for yeast and humans (*65, 66*). Despite the importance of these findings, the mechanism by which TFIIIC recruits Pol III to its promoters is not well understood.

1.5 Using cryo-EM to study transcription complexes

Cryo-electron microscopy (cryo-EM) has emerged as the optimal technique for solving structures of large multisubunit macromolecular machines at near-atomic resolution due to several key advantages it offers. Firstly, cryo-EM allows the study of specimens in their near native, hydrated state, avoiding the need for crystallization and potential structural alterations. This is particularly crucial for multiprotein transcription complexes that are challenging to crystallize or exist in dynamic conformational states (*67*).

Secondly, cryo-EM can handle very large structures as it does not require the formation of well-ordered crystals. Due to the large size and considerable heterogeneity, many components of the transcription machinery pose challenges for structural methods like X-ray crystallography or NMR.

Thirdly, recent advancements in detector technology, imaging software, and computational algorithms have significantly improved the resolution attainable with cryo-EM (*68*). With the advent of direct electron detectors, it is now possible to achieve near-atomic resolution and accurately visualize the intricate details of these macromolecular machines. High-resolution cryo-EM maps can provide evidence to build precise atomic models, enabling the determination of key functional sites within the complexes. This level of detail aids in understanding the molecular mechanisms underlying the functions of these machines and can guide the design of targeted therapeutics.

Finally, cryo-EM enables the visualization of macromolecular machines in different functional states. This allows for the investigation of dynamic processes, enabling the capture of various conformational changes. Transcription machinery is known to undergo conformational rearrangements during its functional cycles, and cryo-EM can provide snapshots of these dynamic states.

CHAPTER 2: Structure of the human Mediator-bound transcription pre-initiation complex

2.1 Introduction

Eukaryotic transcription requires the assembly of a multisubunit preinitiation complex (PIC) composed of RNA polymerase II (Pol II) and the general transcription factors. The coactivator Mediator is recruited by transcription factors, facilitates the assembly of the PIC, and stimulates phosphorylation of the Pol II C-terminal domain (CTD) by the TFIIH subunit CDK7. Here, we present the cryo–electron microscopy structure of the human Mediator-bound PIC at a resolution below 4 angstroms. Transcription factor binding sites within Mediator are primarily flexibly tethered to the tail module. CDK7 is stabilized by multiple contacts with Mediator. Two binding sites exist for the Pol II CTD, one between the head and middle modules of Mediator and the other in the active site of CDK7, providing structural evidence for Pol II CTD phosphorylation within the Mediator-bound PIC.

2.2 Methods

2.2.1 Purification of protein components

Pol II and TFIIH were purified endogenously from HeLa cells, and TFIIB, TFIIA, TBP, TFIIE, and TFIIF were purified recombinantly, as previously described (*26, 69*). Mediator was purified endogenously from HeLa cells, as previously described (*70*). 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 MgCl2, 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.

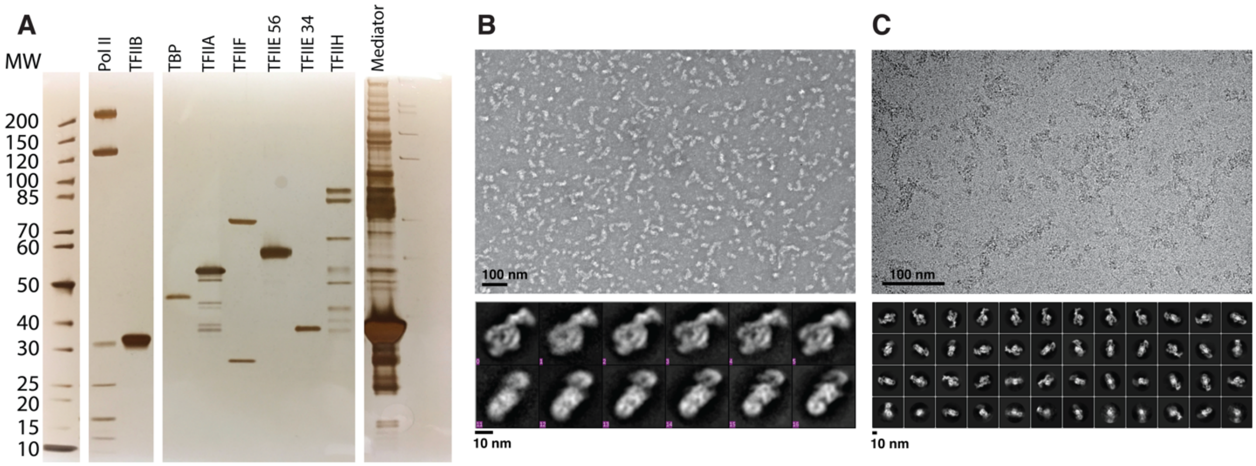

Figure 2.1. Assembly of Med-PIC. A) SDS PAGE gel of purified Med-PIC factors. Lanes were rearranged for clarity. B-C) Representative negative stained (B) and cryogenic (C) electron micrograph and class averages show intact Med-PIC complexes with multiple views.

2.2.2 Assembly of Mediator-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 (Figure 2.1A) (*26, 69*). For negative staining, three subcomplexes were assembled in parallel. First, 0.25 pmol of a super core promoter DNA template (sense: 5'-GAAGGGCGCCTATAAAGGGGGTGGGGGCGCGTTCGTCCTCAGTCGCGATCGAACA

CTCGAGCCGAGCAGACGTGCCTACGGACCATGGAATTCCCCAGT-3’, anti-sense: 5’-/5BiotinTEG/ACTGGGGAATTCCATGGTCCGTAGGCACGTCTGCTCGGCTCGAGTGTTC GATCGCGACTGAGGACGAACGCGCCCCCACCCCCTTTTATAGGCGCCCTTC-3’) was mixed with 1.8 pmol TFIIB, 2 pmol TBP, 1 pmol TFIIA (Figure 2.2). 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 MgCl2, 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 MgCl2, 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 MgCl2, 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 MgCl2, 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.

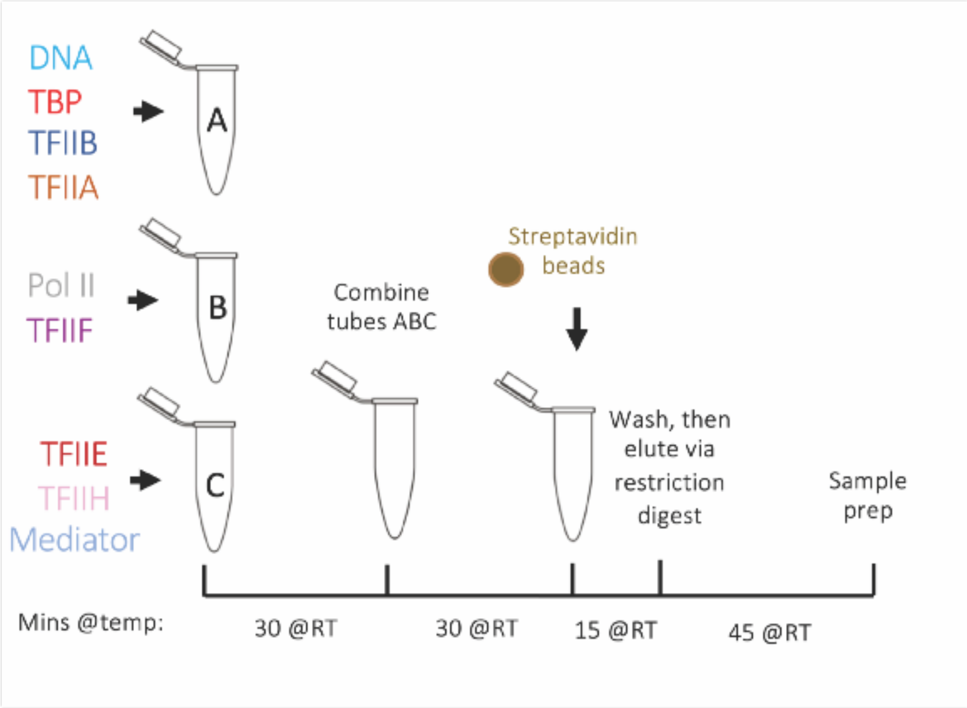

Figure 2.2. Mediator-PIC complex assembly scheme. The proteins were split into three groups, incubated separately, then combined for further incubation, bead binding, washing, and elution steps.

2.2.3 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 °C. 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 (Figure 2.1B) (*71*).

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 (*72*). 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 °C 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 (*73*) 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 (Figure 2.1C).

2.2.4 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 (*74-77*). 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 (*78*).

RELION 3.1 was used for all pre-processing, 3D classification, model refinement, post-processing, and local-resolution estimation jobs (*79*). 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 (*77, 79*).

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 (Figure 2.3). 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 (*80*). 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 (*81*). This map is the full Med-PIC map used for deposition.

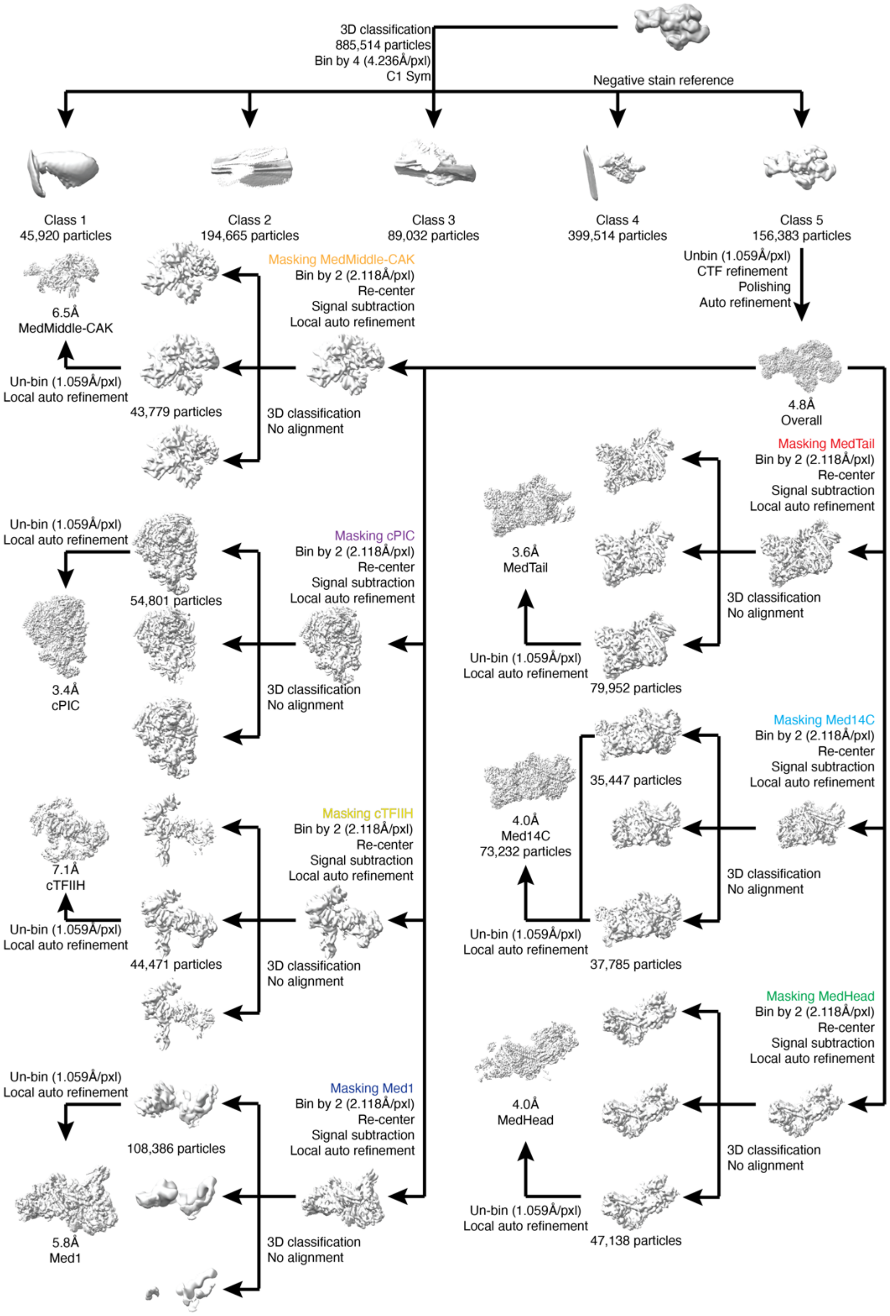

Figure 2.3. Med-PIC cryo-electron microscopy processing pipeline. An initial 3D classification of all particles resulted in a single class that could be refined to an overall resolution of 4.8 Å. Focused local refinements on subcomplexes were performed for cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, 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.

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 Figure 2.3 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 (Figure 2.4).

3D variability analysis was performed on the Med-PIC, MedΔTail-PIC, and Med-CAK maps using CryoSPARC (*82*). 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 (*83, 84*). 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 (*81*).

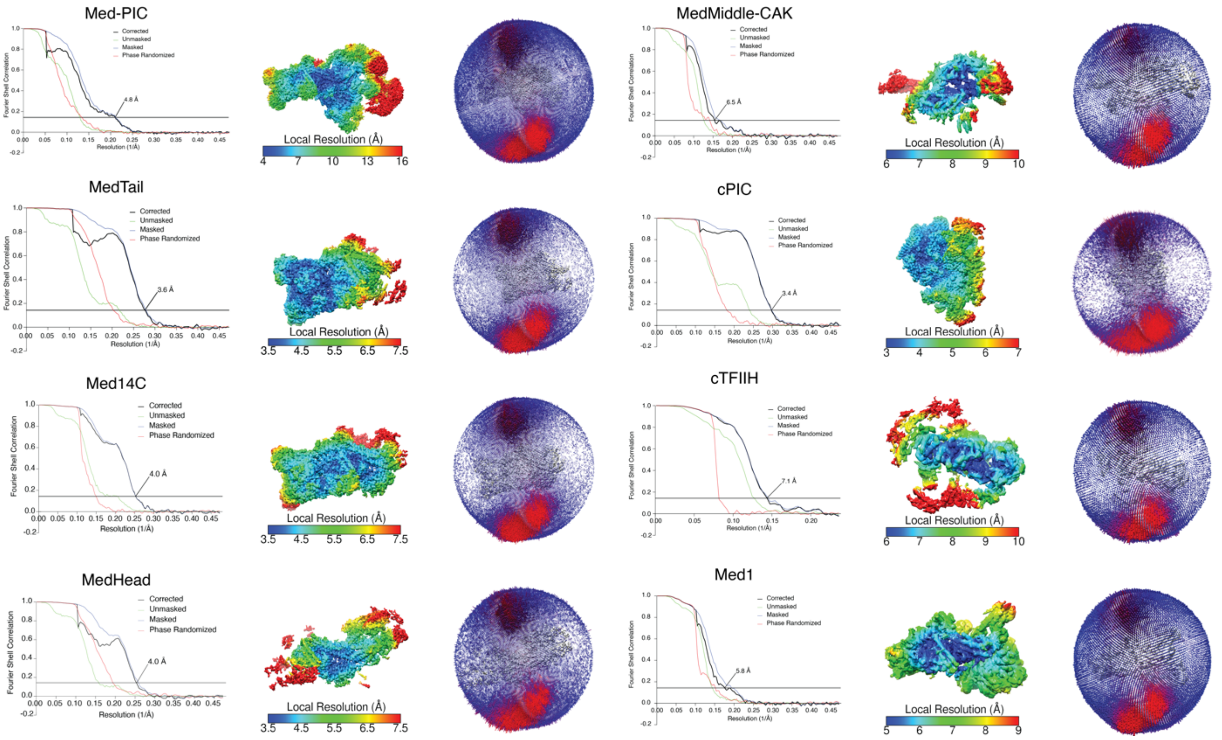

Figure 2.4. Map resolution, local resolution maps, and angular distributions for the full Med-PIC map and cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1 focused refinements. The overall reconstruction shows significant variation in local resolution that improves in each subcomplex following focused refinement.

2.2.5 Model building

cPIC:

The human cPIC bound to a closed DNA template (PDB:5IYA, (*26*)) was fit as a rigid body into the cPIC density map as an initial model using UCSF Chimera (*84*). 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, (*85*)). The cPIC model was real space refined in Phenix to the cPIC map (*86*).

MedHead:

S. pombe subunit structures (PDB:5U0S, (*20*)) of Med6, 8, 17, and 22 or the *S. cerevisiae* (PDB:5OQM, (*21*)) structure of Med11 were used as initial models for building the human structures using the MedHead map in Coot (Figure 2.5) (*87*). 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 (*88*) and Jpred4 (*89*), 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 (Figure 2.6).

Homology models of Med18 and Med20 were built using the MPI Bioinformatics Toolkit and Modeller (*88, 90*). These homology models were aligned to the structure of the Med18-Med20 dimer structure (PDB:2HZM, (*91*)), 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.

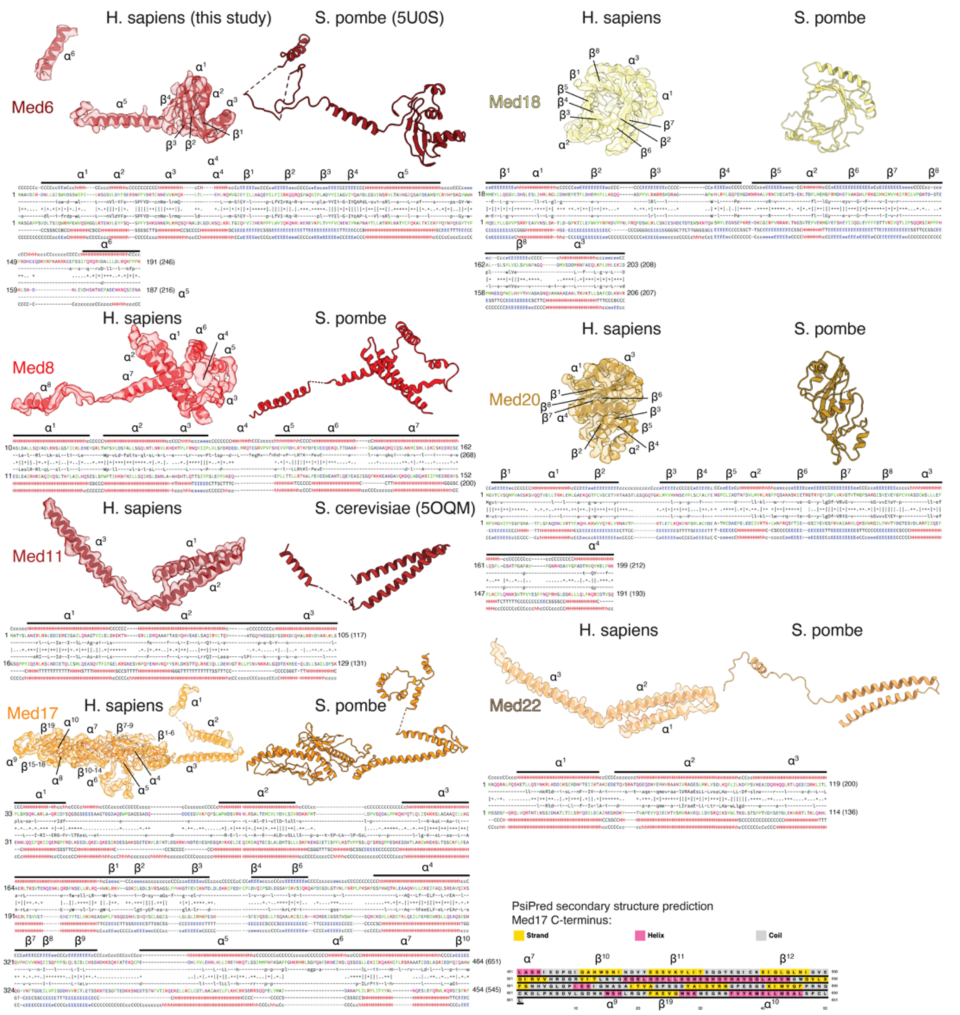

Figure 2.5. Comparison of conserved MedHead subunits. Each MedHead subunit is shown with the corresponding homology model from either *S. pombe* (PDB:5U0S) or *S. cerevisiae* (PDB:5OQM) and the sequence alignment used to build the human model. Colored bars above the

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.

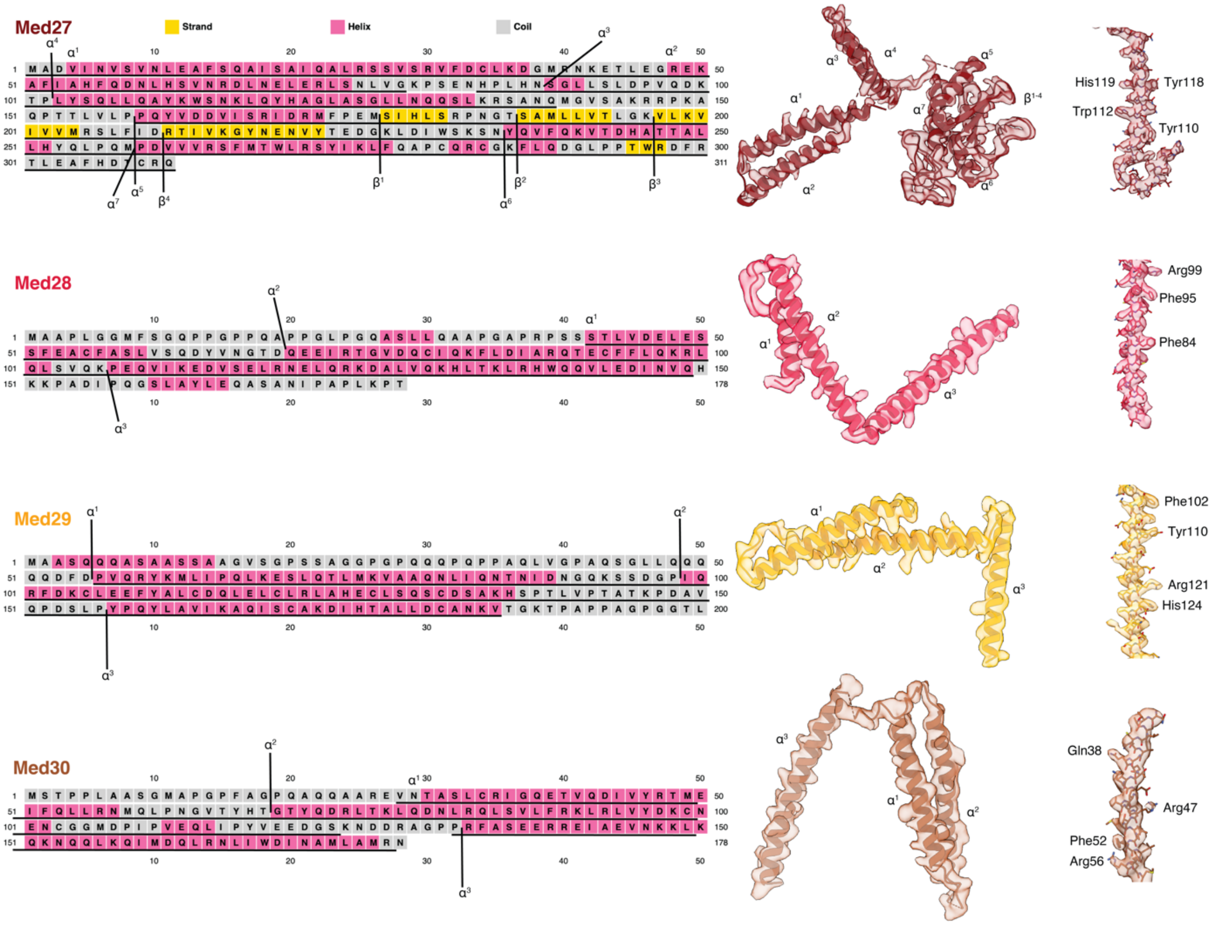

Figure 2.6. Models of MedHead subunits Med27, Med28, Med29, and Med30. Secondary structure prediction using PsiPred closely agrees with secondary structure visible in the density maps and allowed the building of atomic models for each subunit. Secondary structure elements are labeled on both the protein sequence and the models. Representative model-to-map fits (far right) show clear density for bulky side chains that enabled atomic model building.

Med14C:

Yeast Med14 (PDB:5OQM, (*21*)) 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, (*92*)) was fit as a rigid body into the map of MedTail, and manual adjustments were made in Coot (Figure 2.7). 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 (*93*)). 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.

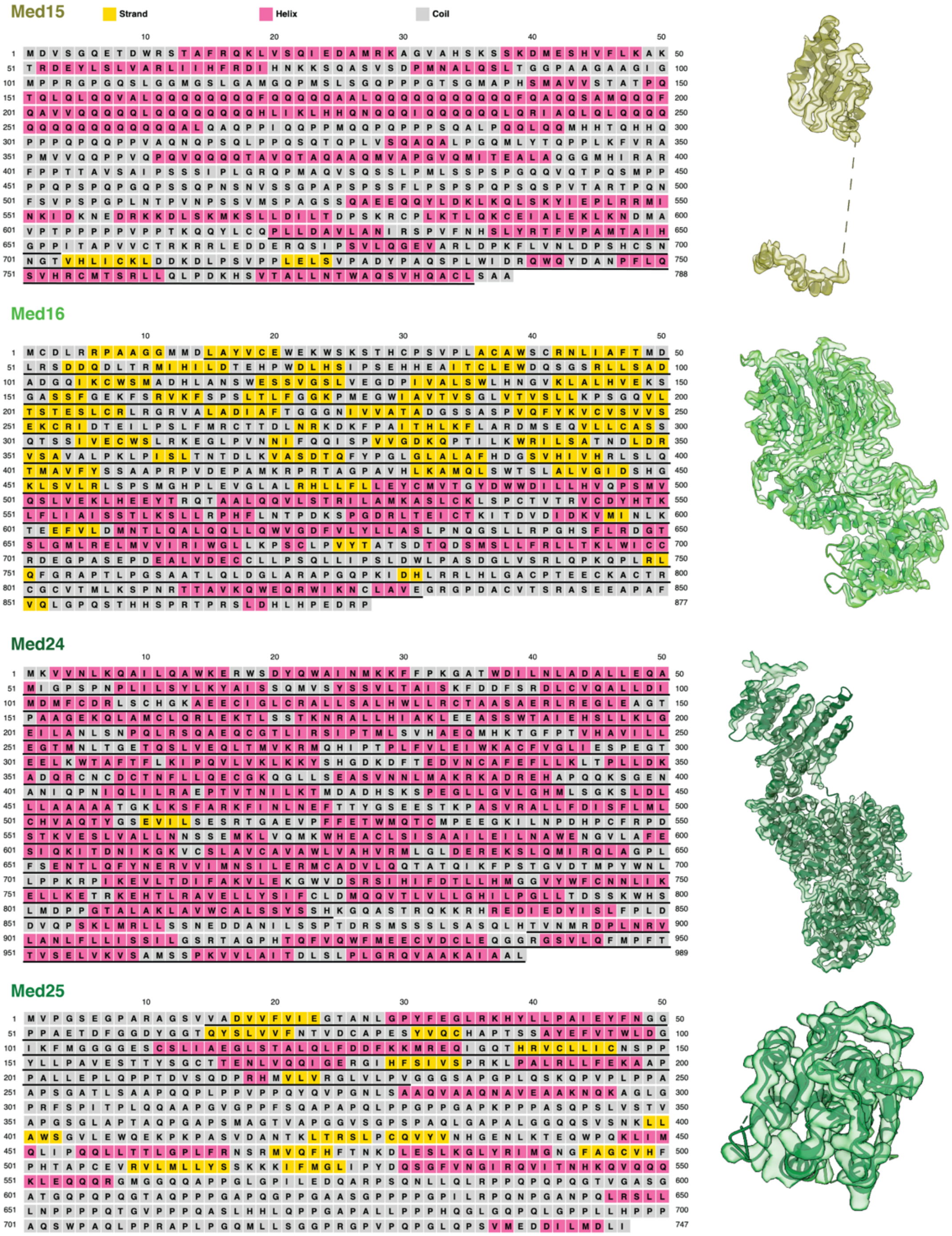

Figure 2.7. Models of MedTail. Secondary structure prediction and model-to-map fit for each subunit of MedTail.

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, (*21*)) in Coot (Figure 2.8). These homology models were flexible fitted into the MedMiddle-CAK density using Namdinator (*94*). 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.

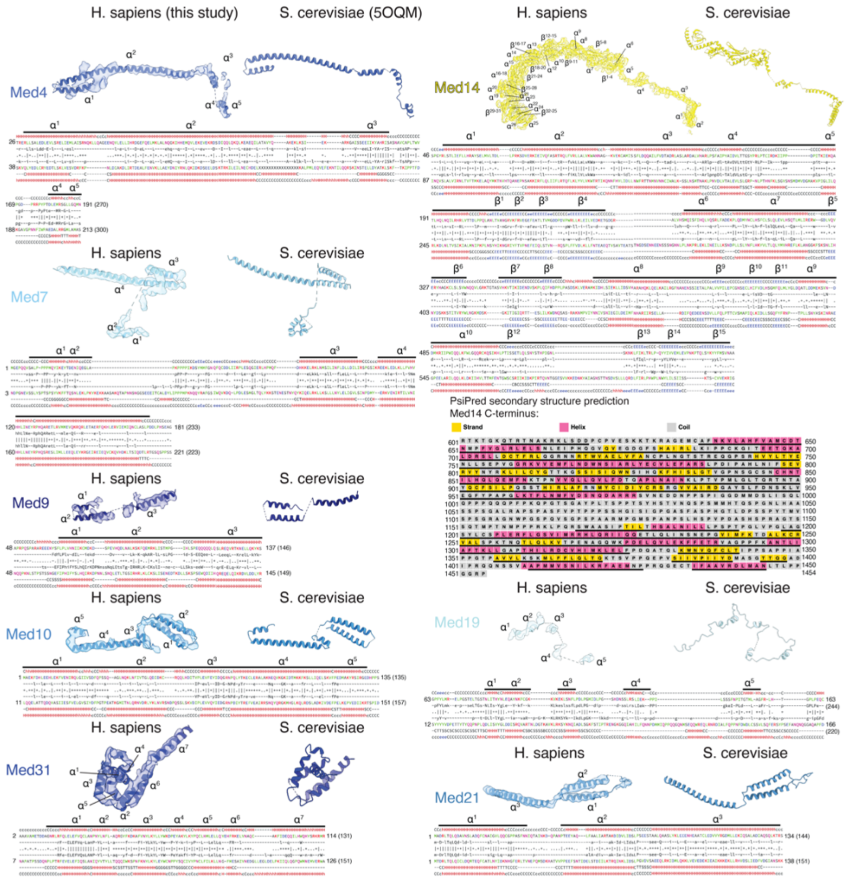

Figure 2.8. Comparison of MedMiddle subunits. Each MedMiddle subunit is shown with the corresponding homology model from *S. cerevisiae* (PDB:5OQM) and the sequence alignment used

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.

The MEDCTD structure was created by first aligning the yeast MedHead-CTD structure (PDB:4GWQ, (*31*)) 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 Y1 residues was visible in the MedHead density, and the remaining model was built using Coot.

TFIIH-CAK:

The human CAK module structure (PDB: 6XBZ, (*95*)) was fit as a rigid body into the MedMiddle-CAK density. The CDK2-cyclin A-peptide substrate structure (PDB: 1QMZ, (*96*)) 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 S5P6 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 (*86*). 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, (*97*)) 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.

Table 2.1. 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							
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							
Micrographs collected (no.)	19,881							
Micrographs used (no.)	15,978							
Total extracted particles (no.)	885,514							
Reconstruction	Med-PIC	cPIC	cTFIIH	Med Head	MedMiddle-CAK	Med14C	MedTail	Med1
Refined 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
Map sharpening B factor ($Å^2$)	-126	-76	-248	-92	-183	-92	-81	-221
Model composition								
Protein	15,877							
Ligands	19							
RNA/DNA	128							
Model Refinement								

Refinement package	Phenix
-real or reciprocal space	Real
Model-Map scores	
-CC	0.46
-Average FSC	
B factors (Å^2)	
Protein residues	80.23
Ligand	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 2.2. 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
MEDCTD	A	MedHead	PDB:4GWQ	Backbone trace
CDKCTD	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 PDB:5U0S	Backbone trace
			-	Atomic level
		MedHead		
				Atomic level
		Med14C		
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:5QOM	Backbone trace
Med20	l	MedHead	PDB:2HZM	Atomic level
Med21	x	MedMiddle-CAK	PDB:5QOM	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:5QOM	Backbone trace

2.3 Results

2.3.1 Structural characterization of the human Mediator-PIC

The Med-PIC complex was assembled by extending our previous protocol for assembling the closed complex PIC from purified factors to accommodate the addition of Mediator (Figure 2.1A) (*26*). In contrast to previous protocols where factors were added in a stepwise manner, three subcomplexes, DNA-TBP-TFIIB-TFIIA, Pol II-TFIIF, and TFIIE-TFIIH-Mediator, were first assembled and were next incubated together. Negative-stain electron microscopy (EM) of assembled complexes indicated that a subset of particles contained all components of Med-PIC and that substantial conformational heterogeneity exists (Figure 2.1B).

A cryo-EM data set was collected, and two-dimensional (2D) classification in Relion-3 showed many classes representing the full complex (Figure 2.1C and Table 2.1) (*98*). A subset of 156,383 particles refined to a resolution of 4.8 Å, but because of the high intrinsic flexibility of Med-PIC, distal regions — including MedMiddle, MedTail, and TFIIH — are mostly averaged out in the postprocessed map. Focused refinements on subcomplexes were used to improve the resolution of all portions of the density compared with the full complex (Figures 2.2 and 2.3).

These regions were chosen either because the subcomplex behaves like a rigid body within the full complex, as is the case for the core PIC (cPIC), cTFIIH, MedHead, MedTail, and MedMiddle-CAK, or to center a region within the box to improve its resolution, as in the case of Med1 and Med14C. These refinements improved the resolution of the vast majority of MedTail, Med14, MedHead, and Pol II to 3.5 Å or better (Figure 2.3) and that of flexible regions, including Med1, MedMiddle-CAK, and cTFIIH, to 5.8, 6.5, and 7.1 Å, respectively. These improvements allowed the building, refining, or flexible fitting of atomic models for nearly the entire complex (Figures 2.5 to 2.9, Table 2.2). Overall, the structure of Med-PIC is highly similar to previous human PIC complexes and yeast Med-PIC complexes (Figure 2.10) (*21, 22, 26*). The presence of Mediator does not cause substantial changes in the structures of Pol II or the GTFs, including TFIIB, TBP, TFIIA, TFIIE, and TFIIF. Med-PIC is compatible with the incorporation of TFIID because no clashes are observed upon superimposition of the structure of TFIID-TFIIA-DNA (Figure 2.10D) (*99*).

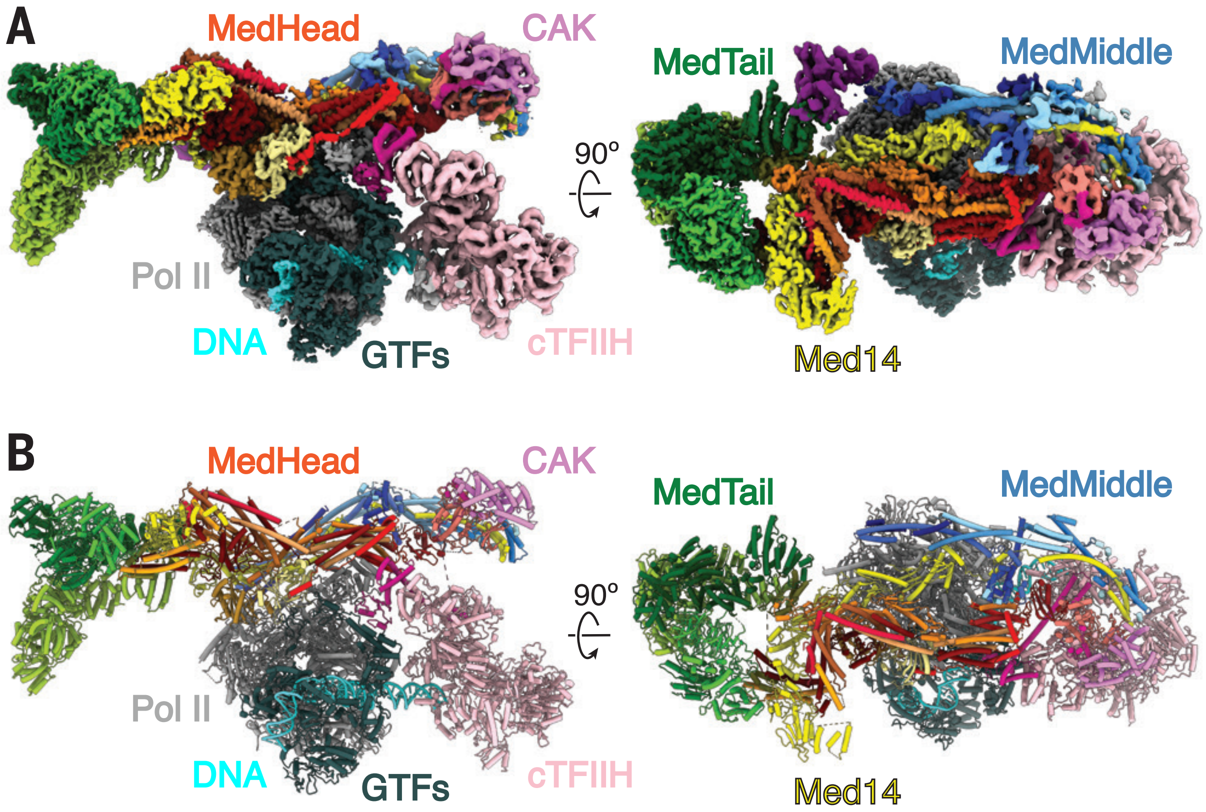

Figure 2.9. Structure of the human Mediator-bound PIC. A) Composite density map for Med-PIC built from the focused refinement maps for cPIC, cTFIIH, MedHead, MedMiddle-CAK, Med14C, MedTail, and Med1. B) Model of the human Mediator-bound PIC. Gray, Pol II; dark gray, GTFs; pink, TFIIH core; salmon, CDK7; violet, cyclin H; medium violet red, Mat1; cyan, DNA; red shades, MedHead; blue shades, MedMiddle; yellow, Med14; green shades, MedTail.

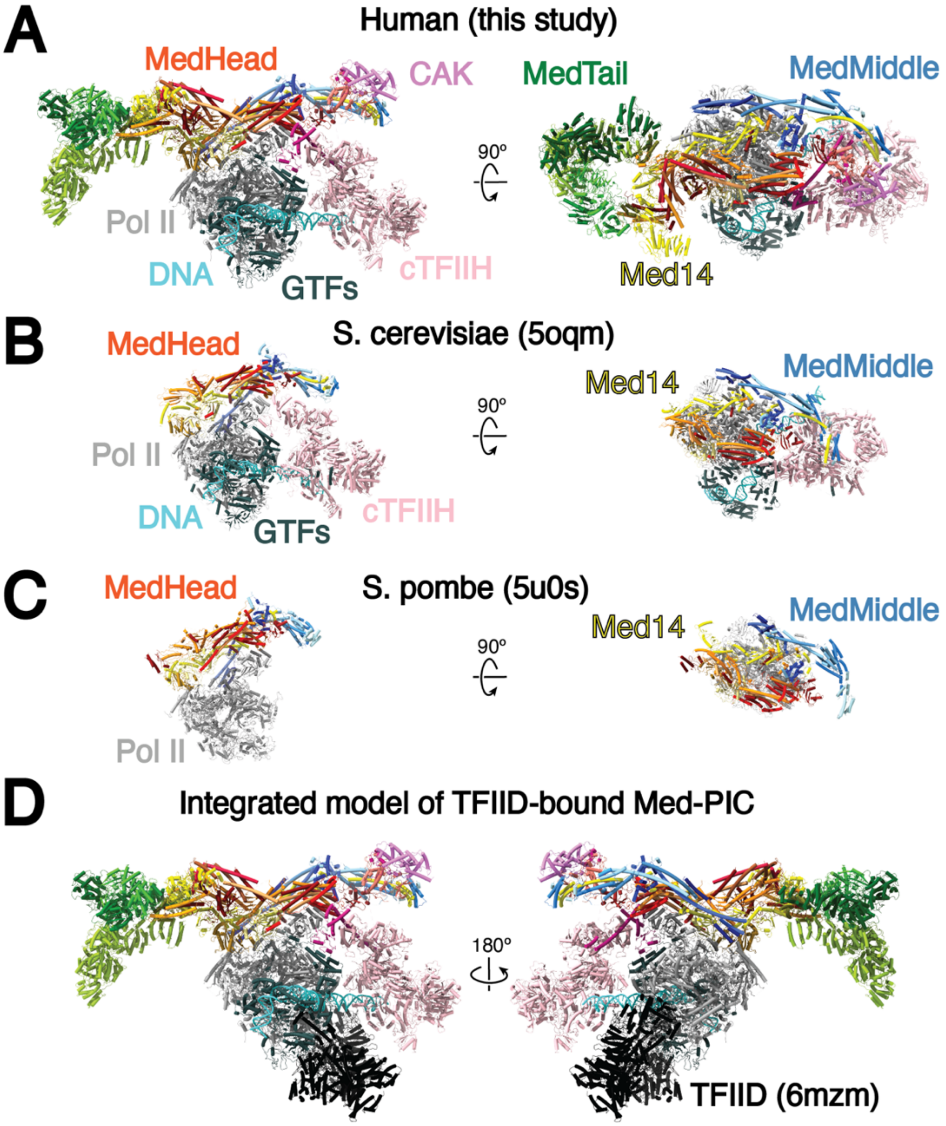

Figure 2.10. Comparison of Med-PICs between yeast and humans and integrated model of a TFIID-bound Med-PIC. A) Structure of the human Med-PIC as shown in Figure 1. B-C) Structure of *S. cerevisiae* and *S. pombe* Med-PIC complexes. The *S. cerevisiae* complex was reconstituted without MedTail, which 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.

2.3.2 Structure of human Mediator

The human Mediator complex within Med-PIC is divided into three modules that are held together by the central Med14 scaffold subunit (Figure 2.11). MedMiddle closely resembles the structure of its yeast counterpart (*21, 22*). Homology models for the human MedMiddle subunits Med4, Med7, Med9, Med10, Med19, Med21, and Med31, based on the *S. cerevisiae* ortholog structures, were built using the MedMiddle-CAK map (Figure 2.11A and Figure 2.8). The N-terminal 200 residues of Med14 were modeled similarly. Additional density near the connector domain of MedMiddle could be assigned to Med26, a metazoan-specific subunit that has been shown to localize in this part of Mediator and interact with Med4, Med7, and Med19 (Figure 2.11A) (*29*). The C terminus of Med26 is sufficient to interact with Mediator, strongly suggesting that the C terminus of Med26 is what can be seen, leaving the N terminus flexibly attached. The N terminus has been shown to interact with the super elongation complex (SEC), which is responsible for the release of paused Pol II through phosphorylation of the Pol II CTD and SPT5 by CDK9 (*100*). Additional unmodeled density attributed to the N terminus of Med1 is located between the plank domain (Med4 and Med9) and MedTail subunit Med24 (Figure 2.11A and Figure 2.12). This is consistent with the location of Med1 shown in both yeast and humans previously (*29*). Density for the plank, Med1, and the N terminus of Med24 is considerably worse than surrounding areas, indicating that this portion of Mediator moves independently of MedMiddle and MedTail. Previous structures of yeast Med-PICs show interactions between Med9 and the foot domain of Pol II (Figure 2.13, A and B) (*20, 21*). In *S. pombe*, Med4 and Med9 also interact with Med1, but there is no change in the overall structure compared with *S. cerevisiae*, where Med1 was not included during complex assembly. The contact between Med9 and the foot domain of Pol II is broken in the human Med-PIC. Instead, Med9 is very close to RPB8, and the interactions between Med4 and Med9 with Med1 are retained (Figure 2.13C). These differences are likely driven by the presence of the larger MedTail in the human Med-PIC, which positions Med1 further away from the plank through interactions with Med24.

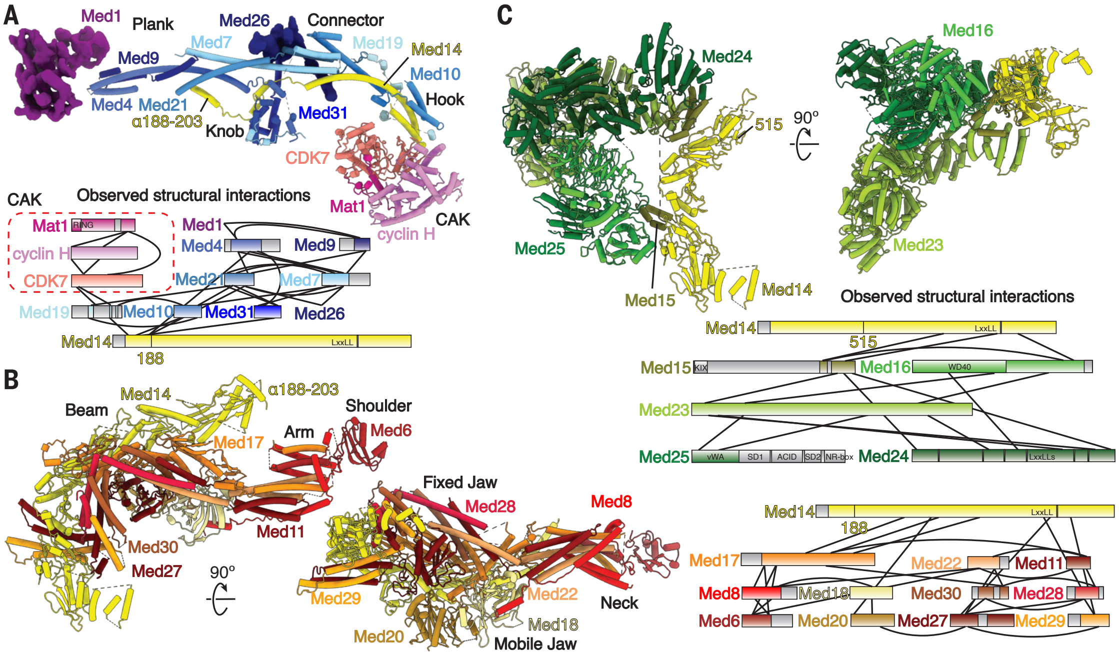

Figure 2.11. Models and observed structural interactions for human Mediator. (A to C) Model and observed structural interaction diagram for MedMiddle and the CAK module of TFIIH (A), MedHead (B), and MedTail (C). The N terminus of the scaffold subunit Med14 extends the length of MedMiddle. Putative densities for Med1 and Med26 are shown and colored purple and dark blue, respectively. The C terminus of Med14 forms extensive interactions with MedHead. MedTail also interacts with the C terminus of Med14 but on the opposite face. Portions for the

models that were built are shown in color; unmodeled sections are shown in gray. Known domains are shown with a light-to-dark (top-to-bottom) gradient. Everything else is shown with a dark-to-light gradient.

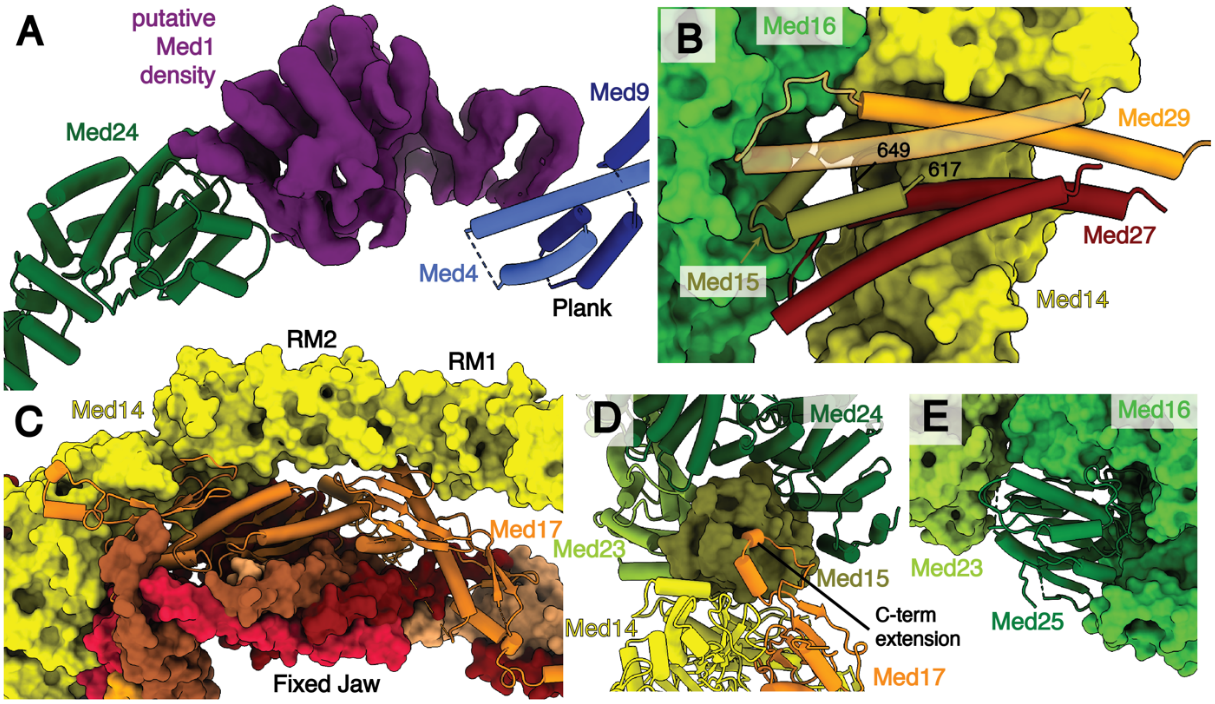

Figure 2.12. Key interfaces in Mediator-PIC. A) Putative density for Med1 is located between the N-terminus of Med24 and the end of the plank domain formed by Med4 and Med9. B) Two helices of Med15, residues 617 to 649 are sandwiched between Med27 and Med29 and together form one of the two main interfaces 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).

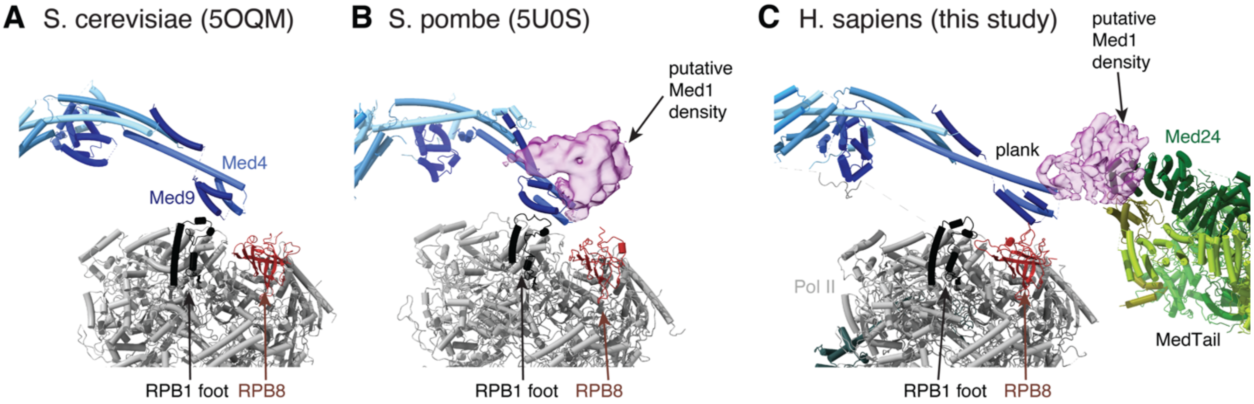

Figure 2.13. Comparison of Mediator plank domain interactions with Pol II. A) The *S. cerevisiae* plank domain interacts with the RPB1 foot (black). The Med1 subunit is not present in the *S. cerevisiae* structure. B) The *S. pombe* plank domain interacts with both the RPB1 foot (black) and putative 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.

MedHead adopts a very similar structure to the yeast model except for the presence of the additional subunits Med27, Med28, Med29, and Med30 (Figure 2.11B and Figure 2.5). These subunits, which were assigned previously to either MedHead or MedTail (*29, 101*), exhibit extensive interactions with the fixed jaw of MedHead and were therefore assigned to MedHead. The C terminus of the scaffold subunit Med14 extends the RM1 and RM2 repeats, which are visible in yeast structures, and wraps around MedHead, serving as a clear divider between MedHead and MedTail. Med17, a scaffold subunit within Med- Head, stabilizes the fixed jaw on one face and interacts with the RM1 and RM2 repeats of Med14 on the other (Figure 2.11B and Figure 2.12B).

MedTail connects to the rest of Mediator through two relatively small interfaces with MedHead and Med14. Two C-terminal domains of Med15 are crucial for forming both interfaces. The first contact site is located near the terminus of Med14. Two helices each from Med27 and Med29 project underneath Med14, with two helices of Med15 (residues 617 to 649) wedged between them (Figure 2.12C). A concave surface on Med16 contacts both this site and Med14. The second site is formed by a C-terminal extension between b18 and b19 (residues 596 to 620) of Med17 that interacts with the Ring-WD40-DEAD (RWD) domain of Med15 (residues 674 to 692) (Figures 2.5 and 2.12D). The RWD domain of Med15 is wedged in a large cavity between Med23 and Med24.

The rest of MedTail is formed by subunits Med16 and Med23 to Med25. Med16 is divided into N-terminal and C-terminal domains, with the N terminus forming a seven-blade WD40 domain and the C terminus forming a mostly helical domain that constitutes much of the first interface with MedHead described above (Figures 2.7 and 2.12B). The N terminus of Med24 interacts with Med1 and is much more flexible than the rest of MedTail. We could only identify a single domain of Med25, the von Willebrand factor type A (vWA) domain, wedged in a pocket formed by Med16 and Med23 (Figure 2.12E).

Almost all domains that are bound by transcription factors in Mediator, including the N terminus of Med15, the N terminus of Med25, and the C terminus of Med1, are flexibly attached to the main body and not visible in the density map (Figure 2.14). The first visible portion of Med15 is located underneath MedTail, near the upstream DNA, allowing its N terminus to easily engage with DNA-bound transcription factors. The C terminus of Med1 contains the NR-boxes important for nuclear receptor (NR) binding (*102*). Many NRs also bind to a C-terminal fragment (residues 1147 to 1454) of Med14 (*103-105*). These two binding interfaces for NRs are quite far from each other (Figure 2.14). The NR AF-1 and AF-2 domains that mediate these interactions are at opposite ends of NR sequences, suggesting that NRs might have to stably associate with the full complex to bridge these two interfaces.

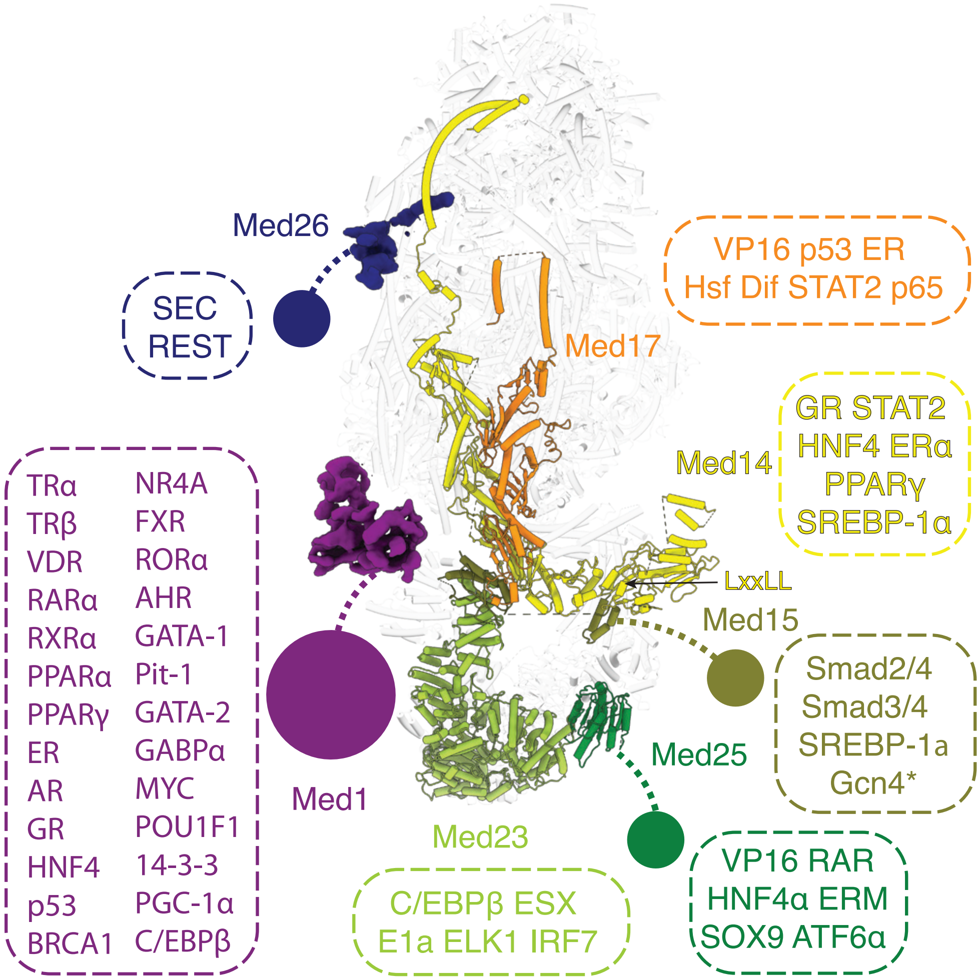

Figure 2.14. Location of Mediator domains and subunits that interact with transcriptional activators or elongation factors. Flexible tethered domains are indicated by solid circles

connected by dashed lines. All interactions shown are between human factors except for Gcn4, which is from yeast and indicated by an asterisk.

The VP16 activation domain (AD) used to purify Mediator for this study binds to the ACID domain located at the N terminus of Med25 (*106, 107*). The VP16 AD appears to stay bound to Mediator during complex assembly. Because of the absence of density of the ACID domain bound to the VP16 AD in this structure, we can conclude that the ACID domain remains flexibly tethered upon activator binding. It has been hypothesized that conformational changes after activator binding to Mediator could lead to the activation of Med-PIC (*108-110*). Given that so many of the activator-binding domains within Mediator are flexibly tethered to the main body, it is unlikely that this is a universal mechanism for activating Med-PIC for transcription.

The overall architecture of Mediator appears highly conserved based on recent structures of mouse and *Chaetomium thermophilum* Mediator (*111, 112*). The putative locations of Med1 and Med26 described earlier are in agreement with the mouse structure. The loss of subunits present in other species, including Med27 to Med30, Med23, and Med24, appears to increase the flexibility of MedTail in the *C. thermophilum* structure.

2.3.3 Flexibility of Mediator-PIC

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 (*82*) which shows a broad distribution of movement of Mediator relative to the PIC (Figure 2.15). 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 (Figure 2.15). 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.

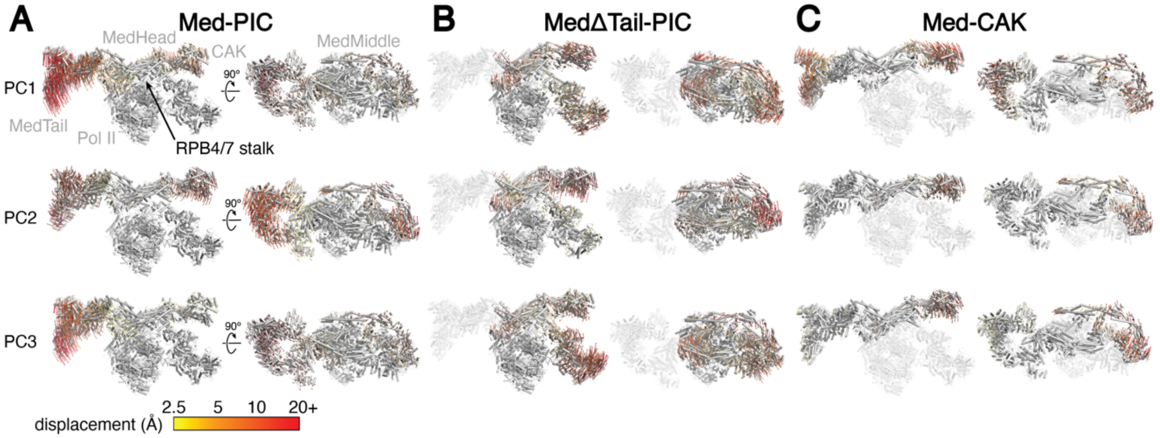

Figure 2.15. 3D variability analysis of Mediator-PIC. A) Top three principal components (PCs) of movement within Mediator-PIC. MedTail and the CAK module undergo the largest displacements in Mediator-PIC. B) Top three PCs of movement within MedΔTail-PIC. When isolating movement from MedTail, the rotation 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).

2.3.4 Mediator stabilizes the CAK module of TFIIH

Although previous structural studies of Med-PICs established that the CAK module of TFIIH occupies a position between the shoulder and hook domains of Mediator, the position and orientation of each CAK module subunit could not be determined (*20-23*). Rigid body docking of the human CAK module structure into our density led to an unambiguous orientation of the CAK module with the active site of CDK7 facing the hook domain of MedMiddle (Figure 2.16A) (*95*). Mediator stabilizes the CAK module through interactions involving Med6, the N terminus of Med14, and a small fragment of Med19 (residues ~133 to 148) with CDK7 (Figure 2.16A). This orientation of the CAK module positions the C terminus of Mat1 ~50 Å from the N terminus bound to cTFIIH, a distance easily spanned by the small fragment of Mat1 (residues 211 to 243) missing in the structure (Figure 2.16B).

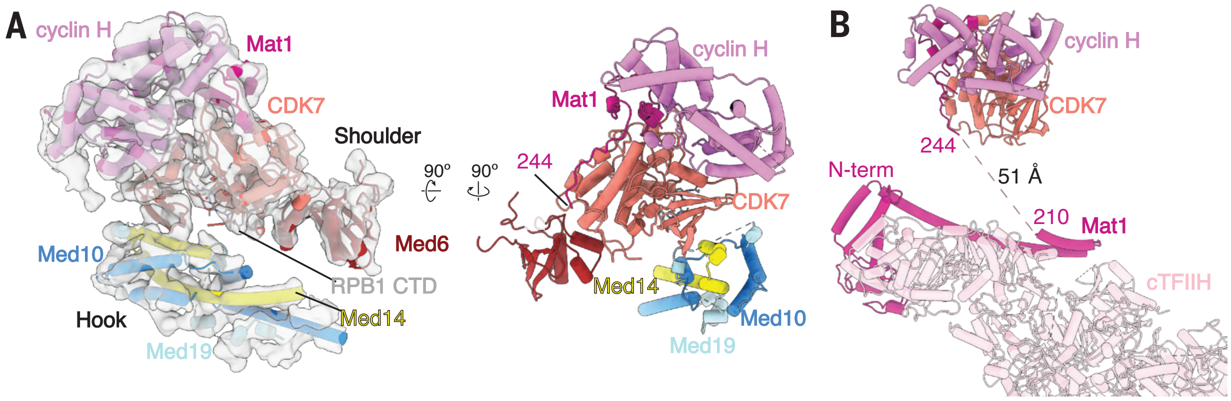

Figure 2.16. Structure of TFIIH within Mediator-PIC. A) Docking of the CAK module (CDK7, cyclin-H, and Mat1) within the MedMiddle-CAK density. The CAK module of TFIIH is stabilized in the Mediator-PIC by interactions between CDK7 and Med6, the N terminus of Med14, and a small fragment of Med19. B) The model of the complete human TFIIH complex places the two

modeled segments of Mat1 (residues 1 to 210 and 244 to 308) close to each other. The missing 34 residues can easily span the 51-Å distance between the termini.

CDK7 adopts the active conformation seen in the human CAK module structure, with the T-loop projecting toward Mat1 and away from the active site (Figure 2.17, A and B) (*95*). Clear density in the CDK7 active site closely matches the location of the substrate peptide in the CDK2–cyclin-A–substrate peptide complex (*96*) (Figure 2.17, A to C). This peptide shares the identical serine-proline sequence that is found in the RPB1 CTD targeted by the kinase. Therefore, we built a model for the RPB1 CTD in the active site that we designate as ${}_{\text{CDK}}\text{CTD}$.

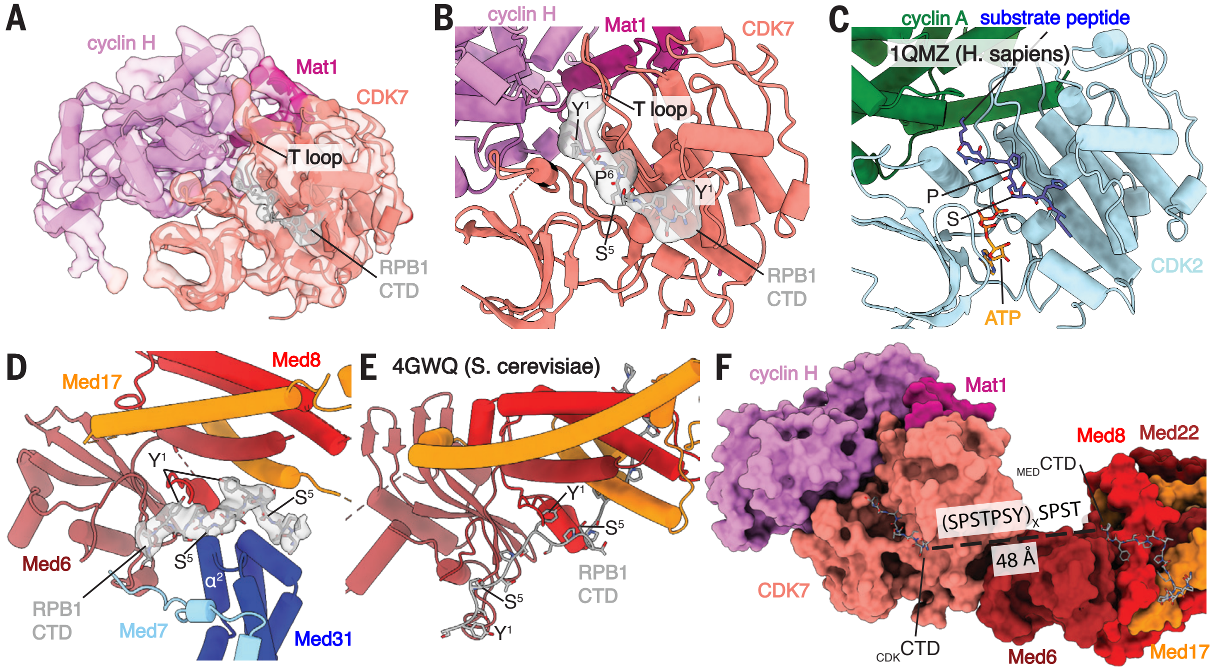

Figure 2.17. Location of RPB1 CTD binding in Mediator-PIC. A) Structure of the TFIIH CAK module. Segmented map of MedMiddle-CAK shows clear density representing an active conformation of the T-loop of CDK7 and density for Pol II CTD in the active site of CDK7. B) Model of the CAK module with density observed for the ${}_{\text{CDK}}\text{CTD}$ in the active site. A consensus sequence of the Pol II CTD is modeled because of limited resolution. The T-loop is in the extended,

active conformation. Y^1, Tyr^1; P^6, Pro^6; S^5, Ser^5. C) Model of the CDK2–cyclin A–substrate peptide structure shows high similarity to the CAK module structure with the conserved SP motif that is common to substrates of both enzymes. ATP, adenosine triphosphate. *H. sapiens*, *Homo sapiens*. D) Model and density of $_{\mathrm{MED}}$CTD with interacting subunits of MedHead and MedMiddle. Ser^5 makes close contacts with α^2 of Med31, preventing binding of phosphorylated repeats in this location. E) Model of $_{\mathrm{MED}}$CTD in the yeast MedHead crystal structure shows a more extensive interface between $_{\mathrm{MED}}$CTD and MedHead than in the Med-PIC, likely because of the presence of MedMiddle in the Med-PIC. F) View of $_{\mathrm{CDK}}$CTD and $_{\mathrm{MED}}$CTD within the human Med-PIC structure. Based on the directionality of the CTD, $_{\mathrm{CDK}}$CTD is C-terminal to $_{\mathrm{MED}}$CTD, and the gap between them would require at least two repeats of the CTD. MedMiddle is hidden for easier visibility. Annotated domains of Mediator are labeled in black. The black dashed line represents the missing residues $(\mathrm{SPSTPSY})_x$SPST.

S. cerevisiae MedHead (scMedHead) was co-crystallized with a short peptide of the RPB1 CTD, which shows slightly more than three full repeats engaged with scMedHead at the shoulder and neck domains (*31*). We observed additional electron density in this same location and used the *S. cerevisiae* structure to build a model for this portion of the CTD that we will refer to as $_{\mathrm{MED}}$CTD (Figure 2.17, D and E). $_{\mathrm{MED}}$CTD is 16 residues long, is slightly more than two full repeats, and adopts a somewhat different path than the yeast structure, likely because of the presence of Med31, which interacts with the other side of $_{\mathrm{MED}}$CTD. In scMedHead, the elongated structure of the N-terminal portion of $_{\mathrm{MED}}$CTD forms extensive interactions with Med17. By contrast, we see clear density for $_{\mathrm{MED}}$CTD starting to wrap around Med31. The C-terminal end of $_{\mathrm{MED}}$CTD also does not form as extensive of an interface with Mediator as it does in scMedHead, because of a clash with the Med7 N terminus. Experiments in *S. pombe* show that the CTD is necessary for interaction between MedHead and Pol II in vitro, suggesting that ${}_{\text{MED}}$CTD is critical for this interaction (*22*). ${}_{\text{MED}}$CTD binding to Mediator would likely be disrupted after phosphorylation of Ser5 because of close interactions between Ser5 and the end of Med31 helix a2 (Figure 2.17D).

The directionality of ${}_{\text{MED}}$CTD and ${}_{\text{CDK}}$CTD is the same, with the N-terminal end of ${}_{\text{MED}}$CTD pointing toward Pol II and the C-terminal end of ${}_{\text{CDK}}$CTD leading away from Med-PIC (Figure 2.17F). This observation strongly suggests that ${}_{\text{MED}}$CTD is N-terminal to ${}_{\text{CDK}}$CTD within the full CTD sequence. The distance between the termini of those two CTD fragments is 48 Å. In an elongated state, one repeat of the CTD can span about 25 Å (*113*), so although two repeats of the CTD may be sufficient to bridge that gap, we would likely see better-defined density for the CTD in that case. Therefore, we suspect that three or more repeats are likely looped out between ${}_{\text{MED}}$CTD and ${}_{\text{CDK}}$CTD.

CHAPTER 3: Structural basis of TFIIIC-dependent RNA Polymerase III transcription initiation

3.1 Introduction

RNA Polymerase III (Pol III) is responsible for transcribing 5S ribosomal RNA (5S rRNA), tRNAs, and other short non-coding RNAs. Its recruitment to the 5S rRNA promoter requires transcription factors TFIIIA, TFIIIC, and TFIIIB. Gene-specific factor TFIIIA interacts with DNA and acts as an adaptor for TFIIIC-promoter interactions. However, structural basis of TFIIIC-dependent Pol III promoter recruitment remained elusive. Here we use cryo-electron microscopy to visualize the *S. cerevisiae* complex of TFIIIA and TFIIIC bound to the promoter. We also visualize DNA binding of TFIIIB subunits, Brf1 and TBP, which results in the full-length 5S rRNA gene wrapping around the complex. Our single-molecule FRET study reveals that the DNA within the complex undergoes both sharp bending and partial dissociation on a slow timescale, consistent with the model predicted from our cryo-EM results. Our findings provide new insights into the transcription initiation complex assembly on the 5S rRNA promoter and allow to directly compare Pol III and Pol II transcription adaptations. Our results constitute a comprehensive structural framework for past and future studies of activator-dependent gene transcription.

3.2 Methods

3.2.1 Purification of protein components

TFIIIA:

Two liters of transformed BL21 (DE3) pRARE cells were grown in LB at 37 °C to an OD600 of 0.6. Then 50 μM $ZnSO_4$ was added, the cells were induced with 1 mM IPTG, and the protein was expressed for two hours at 37 °C. Cells were pelleted for 20 minutes at 4000 x g and resuspended in 40 mL of buffer A (20mM HEPES 7.6, 250 mM NaCl, 5 mM $MgCl_2$, 50 μM ZnSO4, 10% glycerol, 5mM dithiothreitol [DTT], 1 mM phenylmethylsulfonyl fluoride [PMSF]). After cell lysis by sonication, cell pellet was collected by centrifugation for 10 minutes at 4000 x g. The pellet was resuspended in buffer A + 5M urea, then briefly sonicated again and incubated on nutator overnight at 4 °C. Next day, cell debris was removed by centrifugation at 15000 x g at 4 °C for 30 minutes and the supernatant was filtered. The supernatant was loaded twice onto a gravity column with 1 ml of HIS-Select Nickel Affinity Gel (Sigma-Aldrich) equilibrated with buffer A + 5M urea + 20mM imidazole. The resin was washed six times with 2 ml of buffer A + 5M urea + 20mM imidazole. The protein was eluted in 1 ml fractions using buffer A + 5M urea + 300mM imidazole. Fractions containing TFIIIA were pooled together and diluted in three steps over 45 minutes with Buffer A + 1mM PMSF + 5mM DTT. The diluted protein was dialyzed against 1L of Buffer A + 1mM PMSF + 10mM BME in 3.5kDa cut-off snakeskin tubing (Thermo Fisher) overnight at 4 °C. Next day aggregation was removed by centrifugation at 10000 x g for 5 minutes. Supernatant containing refolded TFIIIA was flash-frozen in liquid nitrogen.

TFIIIC:

TFIIIC was purified from an *S. cerevisiae* strain with a TAP tag at the C terminus of τ60 (GE Dharmacon, YSC1178-202233621). Eight liters of yeast were grown overnight. Cells were harvested by centrifugation and resuspended in 200ml of cold TAP extraction buffer (40 mM HEPES pH 8, 250 mM ammonium sulfate, 1 mM EDTA, 10% glycerol, 0.1% Tween-20, 1 mM

DTT, 1 mM PMSF, 2mM benzamidine, 0.3 µg/ml leupeptin, 1.4 µg/ml pepstatin, 2 µg/ml chymostatin). Cells were lysed using BeadBeater (Biospec Products). Cell debris was removed by centrifugation at 15000 x g at 4 °C for two hours. The lysate was incubated with 2ml of IgG Sepharose beads (GE Healthcare) for two hours at 4 °C. The beads were washed and resuspended in 4ml of cold TEV cleavage buffer (10 mM HEPES pH 8, 200 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 10% glycerol). TEV cleavage was performed using 25 µg of TEV protease at room temperature (RT) for one hour. TEV flowthrough was collected and $CaCl_2$ was added for a final concentration of 2 mM. 800 µl of Calmodulin Affinity Resin (Agilent Technologies) was washed with Calmodulin binding buffer (15 mM HEPES pH 8, 1 mM magnesium acetate, 1 mM imidazole, 2 mM CaCl2, 0.1% NP-40, 10% glycerol, 200 mM ammonium sulfate, 1 mM DTT, 1 mM PMSF, 2mM benzamidine, 0.3 µg/ml leupeptin, 1.4 µg/ml pepstatin, 2 µg/ml chymostatin) and incubated with TEV flowthrough overnight at 4 °C. Following incubation, the beads were washed with Calmodulin binding buffer, followed by Calmodulin wash buffer (same as the binding buffer but with 0.05% NP-40), followed by Calmodulin transfer buffer (same as wash buffer but without $CaCl_2$). 400 µl of Calmodulin elution buffer (15 mM HEPES pH 8, 1 mM magnesium acetate, 1 mM imidazole, 5 mM EGTA, 10% glycerol, 0.05% NP-40, 200 mM ammonium sulfate) was added to the beads and incubated for 45 minutes at 4 °C. First, 400 µl fraction was eluted, another 400 µl of elution buffer was added to the beads, and eluted after 5 minutes. The following fractions were eluted immediately. Fractions containing protein were pooled together, concentrated and flash-frozen in liquid nitrogen.

Brf1-TBP:

The chimera protein Brf1N-TBPc-Brf1C (*114*) was purified as follows. Two liters of transformed BL21 (DE3) pRARE cells were grown in LB at 37 °C to an OD600 of 0.6, the cells were induced using 0.5 mM IPTG, and the protein was expressed overnight at 18 °C. Cells were harvested by centrifugation and resuspended in 35 ml of BT lysis buffer (20 mM HEPES pH 7.6, 25 μM EDTA, 1.14 M NaCl, 5% glycerol, 10 mM BME, 0.5 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml pepstatin, 300 μg/ml lysozyme), followed by incubation on ice for one hour and followed by sonication. Cell debris was removed by centrifugation at 13000 x g at 4 °C for one hour. The lysate was loaded onto a gravity column with 500 μl of HIS-Select Nickel Affinity Gel (Sigma-Aldrich) equilibrated with BT lysis buffer. The resin was washed five times with 1 ml of BT wash buffer 1 (20 mM HEPES pH 7.6, 7 mM MgCl2, 0.5 M NaCl, 10 mM imidazole, 5% glycerol, 10 mM BME, 0.5 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml pepstatin) and five times with 1 ml of BT wash buffer 2 (20 mM HEPES pH 7.6, 7 mM MgCl2, 0.5 M NaCl, 20 mM imidazole, 5% glycerol, 10 mM BME, 0.5 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml pepstatin). Protein was eluted in five 1 ml fractions with BT elution buffer (20 mM HEPES pH 7.6, 7 mM MgCl2, 0.5 M NaCl, 200 mM imidazole, 5% glycerol, 10 mM BME, 0.5 mM PMSF, 1 μg/ml leupeptin, 1 μg/ml pepstatin). Fractions containing Brf1-TBP were pooled and dialyzed against 500 ml of dialysis buffer (20mM HEPES pH 7.6, 200mM NaCl, 7mM MgCl2, 0.01% Tween-20, 10% glycerol, 0.2mM PMSF, 10mM BME) in 12 kDa Pur-A-Lyzer Maxi (SigmaAldrich) overnight at 4 °C. Next day the dialysis buffer was replaced with fresh dialysis buffer and the protein was dialyzed for another 5 hours. Fractions containing protein were pooled together and flash-frozen in liquid nitrogen.

3.2.2 Assembly of TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes

First, 2 pmol of 5S rRNA gene DNA template (sense: 5’-/5BiotinTEG/TACGGACCATGGAATTCCCCAGT AACATGTCTGGACCCTGCCCTCATATCACCTGCGTTTCCGTTAAACTATCGGTTGCGGCCATATCTACCAGAAAGCACCGTTTCCCGTCCGATCAACTGTAGTTAAGCTGGTAAGAGCCTGACCGAGTAGTGTAGTGGGTGACCATACGCGAAACTCAGGTGCTGCAATCT -3’, antisense: 5’-AGATTGCAGCACCTGAGTTTCGCGTATGGTCACCCACTACACTACTCGGTCAGGCTCTTACCAGCTTAACTACAGTTGATCGGACGGGAAACGGTGCTTTCTGGTAGATATGGCCGCAACCGATAGTTTAACGGAAACGCAGGTGATATGAGGGCAGGGTCCAGACATGTTACTGGGGAATTCCATGGTCCGTA -3’) was mixed with 100 nmol TFIIIA and incubated at RT for 5 minutes. Then 200 nmol TFIIIC was added and incubated at RT for 5 minutes (Figure 3.1A). For the TFIIIA-TFIIIC-Brf1-TBP complex, the previous step was followed by the addition of 150 nmol Brf1-TBP. The salt concentration was adjusted to 100 mM KCl with the addition of buffer 1 (12 mM HEPES pH 7.6, 0.12 mM EDTA, 12% glycerol, 8.25 mM $MgCl_2$, 1 mM DTT, and 0.05% NP-40). All components were incubated for an additional 5 minutes at RT before binding to T1 streptavidin beads (Fisher Scientific) at RT for 15 minutes. Assembled complexes were washed with buffer 2 (10 mM HEPES pH 7.6, 10 mM Tris pH 7.6, 5% glycerol, 5 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, and 0.05% NP-40) and eluted with buffer 3 (10 mM HEPES pH 7.6, 5% glycerol, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.05% NP-40, and 30 units EcoRI-HF (New England Biolabs) (Figure 3.2A).

3.2.3 Electron Microscopy

Negative stain samples were prepared using 400 mesh copper grids (Electron Microscopy Sciences) coated with continuous carbon on a nitrocellulose support film. Before usage, they were glow-discharged for 10 seconds with 25 W of power using the Solarus plasma cleaner 950 (Gatan). Purified TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes in buffer 4 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 °C. The grid was stained on four 40 µL drops of 2% uranyl formate solution for 5, 10, 15, and 20 seconds sequentially and blotted dry with #1 filter paper (Whatman). Images were collected on a Jeol 1400 microscope 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 (*71*).

Cryo-EM samples were prepared using Quantifoil 2/1 300 mesh copper grids (EMS). Grids were glow discharged for 10 seconds with 25 W of power using the Solarus plasma cleaner 950 (Gatan), and then a thin layer of graphene oxide was applied as described previously (*72*). Purified TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP samples (~3.5 µL) were incubated with 0.05% glutaraldehyde for 10 minutes on ice. The sample was applied to a grid in a Vitrobot Mark IV (Thermo Fisher Scientific) operating at 4 °C with 100% humidity. After 5 minutes of incubation, the sample was blotted with 10 force for four seconds and immediately plunged into liquid ethane cooled to liquid nitrogen temperature.

Cryo-EM data were collected at the Pacific Northwestern Center for Cryo-EM (PNCC). Images were collected using semi-automated data collection in Serial EM (*73*) on a Titan Krios transmission electron microscope (TEM) operated at 300 keV (Thermo Fisher Scientific), equipped with a Quantum energy filter (Gatan), and with a K3 direct detector (Gatan) operating in super-resolution mode.

For the TFIIIA-TFIIIC sample, images were collected at a magnification of 30,000X (super-resolution mode, 0.5395 Å/pixel for raw data, 1.079 Å/pixel after binning) using a defocus range of -1.5 to -3 μm with a dose rate of 1 e^-/pixel/frame for a total dose of 50 e^-/Å^2 (Figure 3.1B). A dataset of 11,645 images was collected. For the TFIIIA-TFIIIC-Brf1-TBP sample, images were collected at a magnification of 30,000X (super-resolution mode, 0.528 Å/pixel for raw data, 1.056 Å/pixel after binning) using a defocus range of -2 to -5 μm with a dose rate of 1 e^-/pixel/frame for a total dose of 60 e^-/Å^2(Figure 3.2B). A dataset of 23,211 images was collected.

Table 3.1. Cryo-EM data collection, refinement, and validation statistics.

	TFIIIA-TFIIIC-Brf1-TBP complex	Focused refinement on the Brf1-TBP-DNA	TFIIIA-TFIIIC complex
PDB code	8FFZ		
EMDB code	29071	29356	29358
Data collection and processing			
Microscope	Titan Krios 3		
Voltage (kV)	300		
Camera	Gatan K3		
Magnification	30k		
Pixel size at detector (Å/pixel)	1.056	1.056	1.079
Total electron exposure (e–/Å^2)	60	60	50
Defocus range (μm)	-2 to -5	-2 to -5	-1.5 to -3
Micrographs used (no.)	23,211	23,211	11,645
Total extracted particles (no.)	5,748,589	5,748,589	2,176,308
Refinement			
Refined particles (no.)	78,512	12,232	109,548
Point-group or helical symmetry parameters	C1	C1	C1
Resolution (global, Å)	3.83	7.14	6.62

FSC threshold	0.143	0.143	0.143
Model composition			
Protein	4045		
Ligands	9		
RNA/DNA	302		
Model Refinement			
Refinement package	Phenix		
-real or reciprocal space	Real		
Model-map CC score	0.66		
B factors (Å^2)			
Protein residues	66.73		
Ligands	134.61		
RNA/DNA	193.74		
R.m.s. deviations from ideal values			
Bond lengths (Å)	0.013		
Bond angles (°)	1.971		
Validation			
MolProbity score	1.01		
CaBLAM outliers (%)	1.48		
Clashscore	0.35		
Poor rotamers (%)	0.38		
C-beta deviations	0.44		
Ramachandran plot			
Favored (%)	94.38		
Outliers (%)	0.13		

3.2.4 Image processing

For negative stain data, particles were picked using DogPicker, extracted, and 2D classified using iterative MSA/MRA topological alignment within the Appion data processing software (*74-77*). A particle stack of ~50,000 particles with a box size of 96 x 96 pixels was subjected to iterative, multi-reference projection-matching 3D refinement using EMAN2 software package to generate an initial reference for cryo-EM data processing (*78*).

RELION 3.1 was used for all cryo-EM data pre-processing, 3D classification, model refinement, post-processing, and local-resolution estimation jobs (*115*). Particles were picked using Gautomatch (developed by K. Zhang, MRC Laboratory of Molecular Biology, Cambridge, UK) and Laplacian-of-Gaussian (LoG) picking in RELION 3.1. Duplicated particles were removed. Local CTF of each micrograph was determined using CTFFIND-4.1 (*77*).

For the TFIIIA-TFIIIC complex, a stack of 2,176,308 particles was binned by a factor of 2 (2.158 Å/pixel) and extracted with a box size of 132 pixels. First round of 3D classification (10 classes) with the negative stain reconstruction as a reference was used to clean up the particle stack (Figure 3.1; Table 3.1). One class of 159,078 particles showed clear structural features of the TFIIIA-TFIIIC complex and was selected for further processing. The selected particles were 3D auto-refined and used in the second round of 3D classification (5 classes). Three classes (109,548) were selected and 3D auto-refined. Another round of 3D auto-refinement was performed with a soft mask, resulting in a 6.78 Å resolution reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criteria (*80*). Three rounds of per-particle CTF refinement were performed, followed by Bayesian particle polishing. 3D auto-refinement using the polished particles resulted in a 6.62 Å resolution map.

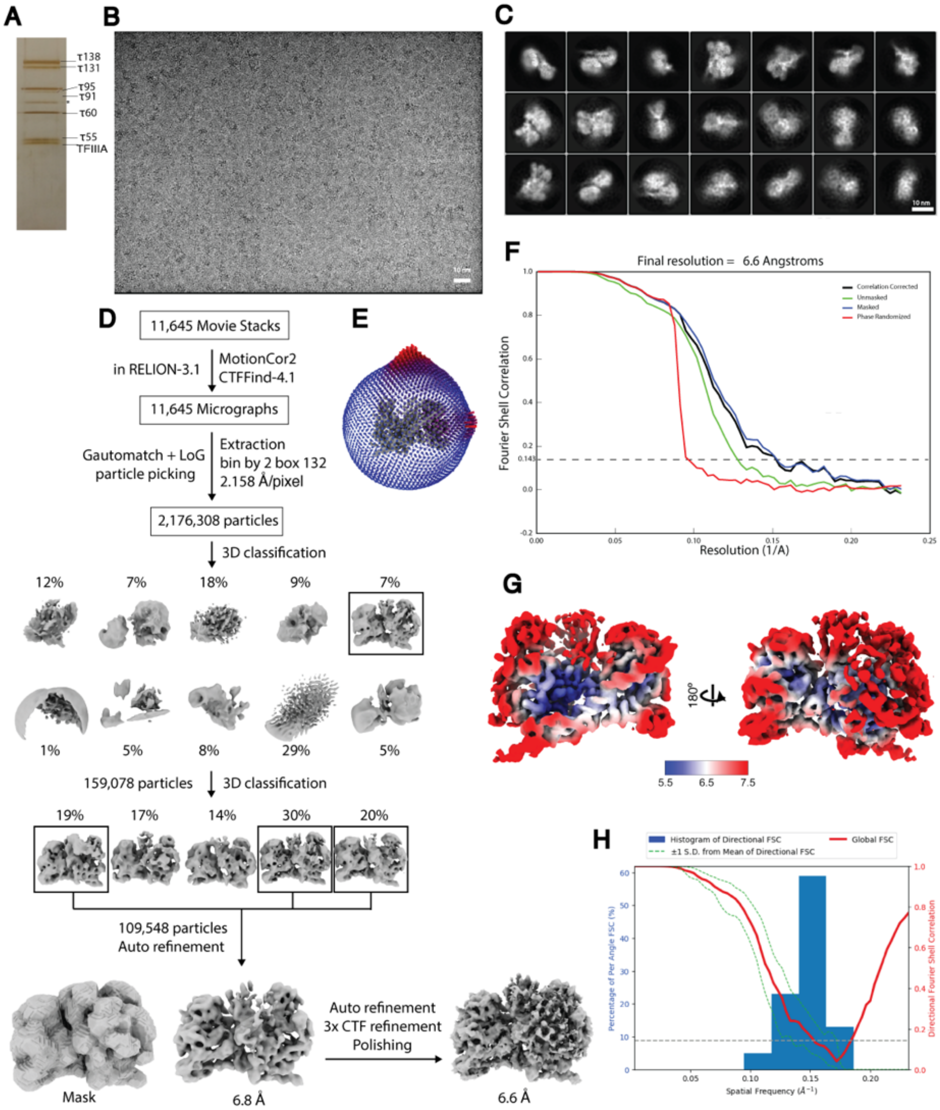

Figure 3.1. TFIIIA-TFIIIC complex data processing. A) SDS PAGE gel of purified TFIIIA-TFIIIC complex on DNA. Individual subunits are labeled on the gel. A band labeled with an asterisk corresponds to τ91 degradation product. B) Cryogenic electron micrograph of TFIIIA-TFIIIC complex. C) Representative 2D class averages of final sorted particles. D) TFIIIA-TFIIIC cryo-electron microscopy processing pipeline. Several rounds of 3D classification resulted in a reconstruction that could be refined to an overall resolution of 6.78 Å. CTF refinement and Bayesian polishing resulted in a map with 6.62 Å resolution. E) Angular distribution for particles

in the final reconstruction of the TFIIIA-TFIIIC complex. Color from blue to red and the height of the bars correlate with the number of particles at a specific orientation. F) Fourier shell correlation (FSC) between the half maps of the final reconstruction. The reported resolution is estimated at the FSC 0.143 cut-off (dashed line). G) TFIIIA-TFIIIC reconstruction colored according to local resolution. The color bar from blue to red indicates the local resolution range in Å. H) 3DFSC plot for TFIIIA-TFIIIC complex on DNA.

For the TFIIIA-TFIIIC-Brf1-TBP complex, a stack of 5,748,589 particles was binned by a factor of 2 (2.112 Å/pixel) and extracted with a box size of 132 pixels. First round of 3D classification (10 classes) with the negative stain reconstruction as a reference was used to clean up the particle stack (Figure 3.2; Table 3.1). One class of 665,210 particles showed clear structural features of the TFIIIA-TFIIIC-Brf1-TBP complex and was selected for further processing. The selected particles were 3D auto-refined and used in the second round of 3D classification (3 classes) without alignment. One class (78,512 particles) was selected. The selected particles were 3D auto-refined, re-centered, and re-extracted without binning (1.056 Å/pixel, box size = 264 pixels). Another round of 3D auto-refinement was performed with a soft mask applied around the whole complex, resulting in a 4.36 Å resolution reconstruction. The particle stack was re-extracted with a bigger box size (384 pixels). Three rounds of 3D auto-refinement, per-particle CTF refinement and Bayesian particle polishing were performed. 3D auto-refinement using the polished particles resulted in a 3.83 Å resolution map. To improve the map quality of the Brf1-TBP part, focused 3D variability analysis was performed in cryoSPARC (*82*). Briefly, the particle stack that resulted in 4.36 Å resolution reconstruction, was imported into cryoSPARC, where it was used for Non-uniform (NU) refinement, local refinement (soft mask applied to focus on Brf1-TBP-DNA density), signal subtraction and 3D variability analysis in cluster mode. This procedure produced a 7.14 Å resolution reconstruction of Brf1-TBP-DNA density. In parallel with post-processing done in RELION 3.1, DeepEMhancer was used to better correct local B-factors and produced cleaner maps for model building and docking (*81*).

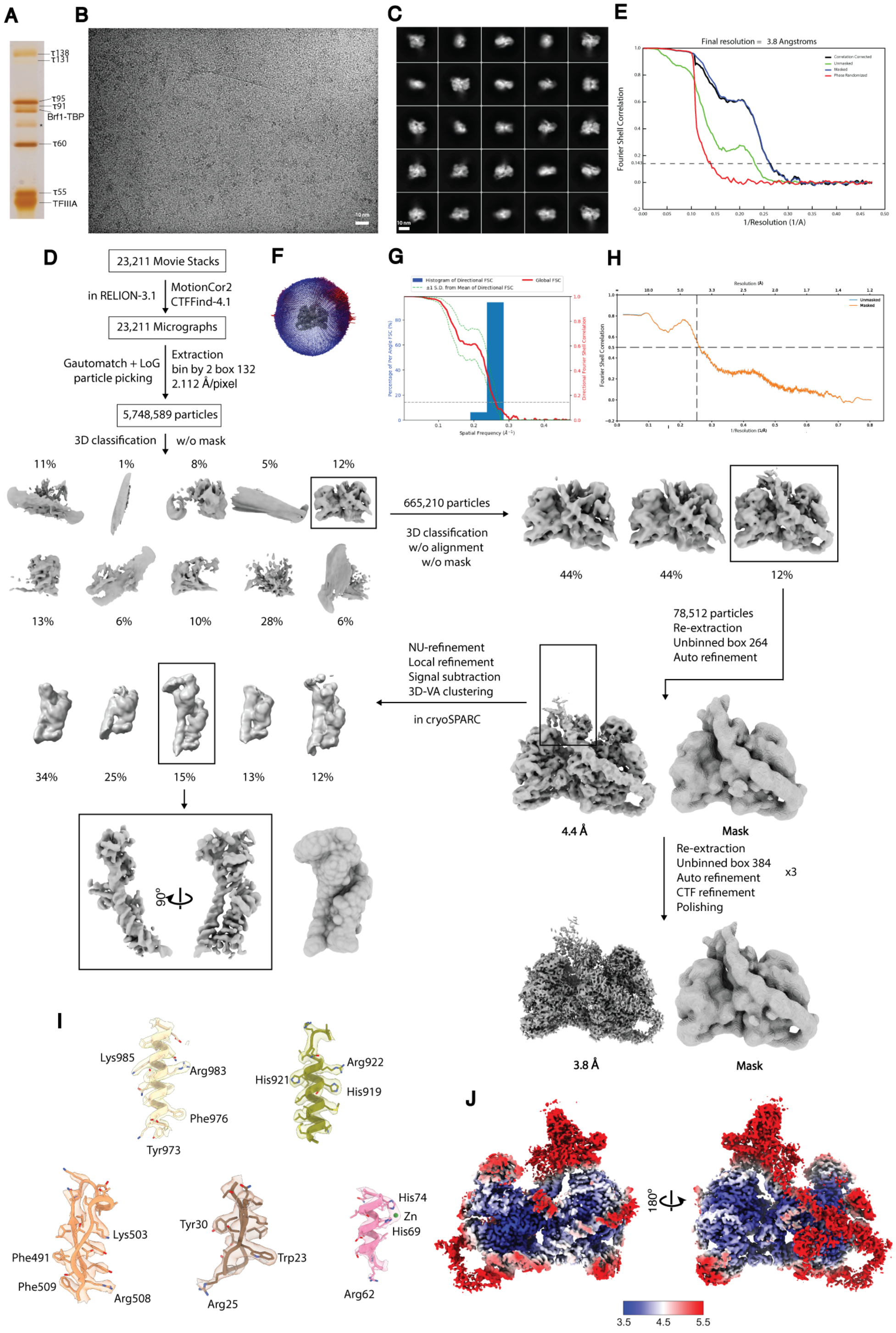

Figure 3.2. TFIIIA-TFIIIC-Brf1-TBP complex data processing. A) SDS PAGE gel of purified TFIIIA-TFIIIC-Brf1-TBP complex on DNA. Individual subunits are labeled on the gel. A band

labeled with an asterisk corresponds to τ91 degradation product. B) Cryogenic electron micrograph of TFIIIA-TFIIIC-Brf1-TBP complex. C) Representative 2D class averages of final sorted particles. D) TFIIIA-TFIIIC-Brf1-TBP cryo-electron microscopy processing pipeline. Several rounds of 3D classification resulted in a reconstruction that could be refined to an overall resolution of 4.36 Å. Three rounds of CTF refinement and Bayesian polishing resulted in a map with 3.83 Å resolution. The Brf1-TBP-DNA part of the density was masked and separately processed in cryoSPARC. E) Fourier shell correlation (FSC) between the half maps of the final reconstruction. The reported resolution is estimated at the FSC 0.143 cut-off (dashed line). F) Angular distribution for particles in the final reconstruction of the TFIIIA-TFIIIC-Brf1-TBP complex. Color from blue to red and the height of the bars correlate with the number of particles at a specific orientation. G) 3DFSC plot for TFIIIA-TFIIIC-Brf1-TBP complex on DNA. H) Model-to-map FSC. The resolution at the FSC=0.5 criterion is indicated by a dashed line. I) Examples of model-to-map fit with several labeled bulky side chains. Model colors correspond to the colors used in the main figures. J) TFIIIA-TFIIIC-Brf1-TBP reconstruction colored according to local resolution. The color bar from blue to red indicates the local resolution range in Å.

3.2.5 Model building

TFIIIA-TFIIIC-Brf1-TBP model:

The resolution of the TFIIIA-TFIIIC-Brf1-TBP complex map (3.8 Å) allowed for AlphaFold-guided model building for the entire complex (*116*). AlphaFold pLDDT score (per-residue estimate of the modeling confidence on a scale from 0 – 100) was used as scoring metric for predictions. Alphafold was used to predict individual domains or full subunits of TFIIIA and

TFIIIC, and the produced models were fitted as a rigid body into the density map using UCSF Chimera (*84*). Published experimental structures were used in rigid body fitting as well (Table 3.2). Side chains were not stubbed. Side chain orientations from Alphafold predictions or experimental models were kept in cases, where map quality did not allow to distinguish a side chain orientation. Manual adjustments were made in Coot (*87*) and ISOLDE (*117*). The final model was refined in Phenix. The refined model was inspected with the help of ISOLDE, where clashes and rotamer outliers were resolved (Table 3.2).

TFIIIA:

X. laevis structure of ZF 1 to 3 bound to DNA (PDB:1TF3) was fitted as rigid body into the TFIIIA density. *S. cerevisiae* TFIIIA Alphafold prediction was aligned with the experimental *X. laevis* structure, and individual ZF was fitted as rigid body into the TFIIIA density. Predicted regions that were not visible in the density were trimmed.

TFIIIC:

The structure of the subcomplex of τ60 and τ91 (PDB: 2J04) was fitted as rigid body into the density. *S. cerevisiae* Alphafold predictions for τ60 and τ91 were aligned with the experimental structure, and flexible loops, not visible in the density, were trimmed. The cryo-EM structure of the τA lobe comprised of τ131, τ95 and τ55 subunits (PDB: 6YJ6) was fitted as rigid body. Alphafold predictions were aligned with subunits and manually adjusted. Predicted regions that were not visible in the density were trimmed. For τ138, individual folded domains from Alphafold prediction were fitted as rigid body into the map, and connecting linkers were manually adjusted.

Brf1-TBP:

Brf1-TBP model from *S. cerevisiae* Pol III PIC structure (PDB:6CNB) was fitted into Brf1-TBP density as rigid body, and flexibly tethered domains of Brf1 that could not be traced in the density (cyclin fold 1 and zinc ribbon), were trimmed. This model was not manually adjusted due to low resolution of the map in this area.

DNA:

Double strand DNA model with 5S rRNA gene sequence was generated using webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp). The DNA register was traced using sequence alignment between *X. laevis* and *S. cerevisiae* ICR and structure of *X. laevis* structure of ZF 1 to 3 bound to DNA (PDB:1TF3), as well as the positioning of Brf1-TBP at the upstream region. The DNA was fitted into the density in ISOLDE.

TFIIIA-TFIIIC model:

The model of the TFIIIA-TFIIIC-Brf1-TBP complex was fitted in the density map of the TFIIIA-TFIIIC complex. TFIIIC subunits τ55, τ95, τ91, and τ60 completely fit in the map. Protein-DNA contacts within the τB model are identical as well. C-terminal TPR and helical domains of τ131 are placed identically to the TFIIIA-TFIIIC-Brf1-TBP complex. N-terminal TPR cannot be resolved, but low-resolution features suggested a conformation of τ131, similar to the one in τA lobe structure (PDB: 6YJ6). TFIIIA zinc fingers 1-5 interact with DNA, while zinc fingers 6-8 protrude away from the DNA. The overall position of TFIIIA is identical to the TFIIIA-TFIIIC-Brf1-TBP complex; however, zinc finger 9 and adjacent helix are not visible in the map. DNA in this complex is also positioned identically to the TFIIIA-TFIIIC-Brf1-TBP complex.

UCSF Chimera and UCSF Chimera X were used for figure and movie generation (*83, 84*). ConSurf web server was used for estimating the evolutionary conservation of residues (*118*).

Table 3.2. Model building starting models and model confidence.

Subunit/domain	Chain ID	Experimental structures	Alphafold Prediction (UniProt ID)	Level of confidence
TFIIIA ZF1-8	A	PDB:1TF3 (Xenopus)	P39933	Atomic level/backbone trace
TFIIIA ZF9	A	-	P39933	Rigid body fit
τ138	B	PDB:5AIM	P34111	Atomic level/backbone trace
τ131 N-term TPR	C	PDB:6YJ6	P33339	Rigid body fit
τ131 C-term TPR	C	PDB:6YJ6	P33339	Atomic level/backbone trace
τ95	D	PDB:6YJ6	P32367	Atomic level/backbone trace
τ91	E	PDB:2J04	Q06339	Atomic level/backbone trace
τ60	F	PDB:2J04	Q12308	Atomic level/backbone trace
τ55	G	PDB:6YJ6	Q12415	Atomic level/backbone trace
Brf1-TBP	H	PDB:6CNB	-	Rigid body fit

3.2.6 Single-molecule FRET studies

The PAGE-purified DNA oligonucleotides containing biotin for immobilization or amine modifications were purchased from Genscript (sense: AACATGTCTGGACCCTGCCCTCATATCACCTGCGTTTCCGTTAAACTATCGGTTGCGGCCATATCTACCAGAAAGCACCGTTTCCCGTCCGATCAACTGTAGTTAAGCTGGTAAGAGCCTGACCGAGTAGTGTAGTGGG/iAmMC6dT/GACCATACGCGAAACTCAGGTGCTGCAATCTGTAGATTCATTGGACTGGTG; antisense: AGATTGCAGCACCTGAGTTTCGCGTATGGTCACCCACTACACTACTCGGTCAGGCTCTTACCAGCTTAACTACAGTTGATCGGACGGGAAACGGTGCTTTCTGGTAGATA/iAmM

C6dT/GGCCGCAACCGATAGTTTAACGGAAACGCAGGTGATATGAGGGCAGGGTCCA GACATGTTACTGGGGAATTCCATGGTCCGTA;

sense, biotinylated: 5’-/5BiotinTEG/TACGGACCATGGAATTCCCCAGT-3’).

Oligonucleotides, carrying an amine group, were labeled with NHS ester-conjugated fluorescent dyes. For the labeling reaction, 6.25 µl of 40 µM amine-modified antisense DNA was mixed with an excess of NHS ester-conjugated Cy3 dye (200:1 ratio of dye to DNA) in 100 mM Na2B4O7 pH8.4 and 40% DMSO. The mix was incubated at 25°C for 6 hours, then at 4°C overnight on a gently shaking mixer in the dark. The excess dye was removed by ethanol precipitation and 70% ethanol wash. The labeling reaction was performed twice to increase labeling efficiency. Amine-modified sense DNA was labeled with NHS ester-conjugated Cy5 dye in the same way. The dsDNA construct (1 µM) was prepared by mixing the biotinylated sense DNA with sense DNA-Cy5 and antisense DNA-Cy3 at a molar ratio of 1.2:1.1:1 and annealed in water in a heat block by incubation at 95°C for 5 minutes followed by gradual cooling to RT.

To achieve consistency between the two methods, we performed the protein-DNA complex assembly in the same way as preparing the cryo-EM samples with a few modifications. The annealed and Cy3-/Cy5-labeled dsDNA construct was used instead of the 5S rRNA gene DNA template. After the complex elution with EcoRI, 1.2 pmol of antisense biotinylated DNA (5’-/5BiotinTEG/CACCAGTCCAATGAATCTAC-3’) was added to the elution and incubated at RT for 10 minutes. The integrity of TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes assembled on fluorescently labeled DNA was confirmed by negative stain EM. The sample was incubated on ice before the smFRET experiment.

Single-molecule experiments were performed in a flow chamber prepared by sandwiching PEG (mPEG and biotin-PEG 1% (w/w), Laysan Bio) passivated glass coverslip (VWR) and slides (ThermoFisher Scientific) (*119*). Before imaging, flow chambers were washed with T-50 buffer (50 mM NaCl, 10 mM Tris, pH 7.4) and incubated with 50 nM NeutrAvidin (ThermoFisher Scientific) for 2 minutes. Then the flow chambers were washed with T-50 to remove unbound NeutrAvidin. Finally, the pre-assembled complex of labeled DNA and proteins, as described above, was injected into the chamber to immobilize on the surface. The complex was diluted and imaged in an imaging buffer consisting of 4 mM Trolox, 10 mM HEPES, 10 mM $MgCl_2$, 50 mM KCl, 5uM $ZnCl_2$, 5 % glycerol, and an oxygen-scavenging system consisting of 4 mM protocatechuic acid (Sigma) and 1.6 Uml^{-1} bacterial protocatechuate 3,4-dioxygenase (rPCO) (Oriental Yeast), pH 7.35. All the chemicals were purchased from Sigma (purity >99.9%).

smFRET data were recorded at 50 ms time resolution using a TIRF microscope with a 100x oil-immersion objective (Olympus, NA 1.49). The donor and acceptor fluorophores were excited using 532 nm and 638 nm lasers, respectively.

Single-molecule fluorescence intensity traces were analyzed using smCamera (http://ha.med.jhmi.edu/resources/), custom-written Matlab (MathWorks) scripts, and OriginPro. smFRET particles were selected based on the Gaussian intensity profile of the spots, acceptor brightness at least 5% above the background, and the acceptor signal upon donor excitation. The donor and acceptor intensity traces were selected based on the following criteria. A single donor and a single acceptor bleaching step during the acquisition time window, stable total intensity ($I_D + I_A$), and anticorrelated intensity profiles of donor and acceptor without blinking. Two authors independently examined all the data and found identical results. Each experiment, unless otherwise noted, was carried out three times to ensure that the findings could be repeated. Apparent FRET efficiency was calculated using $(I_A-0.088 \times I_D)/(I_D + I_A)$, where I_D and I_A are raw donor and acceptor intensities, respectively. smFRET histograms for the DNA, DNA+TFIIIA+TFIIIC, and DNA+TFIIIA+TFIIIC+Brf1-TBP were generated from the 3, 4, and 16 movies from different days.

To determine the FRET value, the smFRET histograms were fitted using OriginPro with 1 or 2 Gaussian distributions.

$$y(x) = \sum_{i=1}^{n} \frac{A_i}{w_i\sqrt{\frac{\pi}{2}}} e^{-2\frac{(x-xc_i)^2}{w_i^2}}$$

where *n* is the number of Gaussians, *A* is the peak area, *xc* is the FRET peak center, and *w* is the full-width half maximum for each peak.

To derive the transition density plots (TDPs), all the time traces were idealized by fitting with hidden Markov model (HMM) using vbFRET software in Matlab (*120*). Traces were fit with 1 to 4 step models and majority of traces fit with a 1 or 2 steps. Then, from the idealized traces, all the transitions were extracted to create a transition density plot.

3.3 Results

3.3.1 Brf1-TBP stabilizes the TFIIIC-TFIIIA complex

The TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes were assembled on a double-stranded (ds) DNA template composed of the *S. cerevisiae* 5S rRNA gene, including the upstream TFIIIB-binding site and the gene body containing ICR (Figure 3.3A). The complexes were assembled in a step-wise manner using individually purified *S. cerevisiae* factors. Cryo-EM datasets were collected for both complexes, with a subset of 109,548 particles refined to 6.6 Å resolution for the TFIIIA-TFIIIC complex (Figures 3.3B and 3.1; Table 3.1), and a subset of 78,512 particles refined to 3.8 Å resolution for the TFIIIA-TFIIIC-Brf1-TBP complex (Figures 3.3C and 3.2; Tables 3.1 and 3.2).

The TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes are assembled on identical DNA templates. The overall conformation of TFIIIC is similar between the two complexes, and the ZF array of TFIIIA interacts with the ICR in a similar manner (Figure 3.3, B and C; Figure 3.4A). All six subunits of TFIIIC are visible in both complexes (Figure 3.3, B to D). The two lobes of TFIIIC, τA and τB, are in close contact with each other, held together by multiple interactions with DNA and the τ138 subunit, which is shared between the two lobes. The τA lobe is comprised of a τ95-τ55 dimer, τ131, and the C-terminal half of τ138. The τB lobe includes subunits τ91, τ60 and the N-terminal half of τ138. The interaction between τA and TFIIIA, as well as those between τB and DNA, are identical in the two complexes.

The presence of Brf1-TBP in the complex alters the interactions between TFIIIC τA lobe and DNA (Figure 3.3, B and C). In the TFIIIA-TFIIIC-Brf1-TBP complex, the DNA upstream of ICR is visible and is stabilized by multiple interactions with TFIIIC τA lobe. The addition of Brf1-TBP dramatically changes the position of τ131 N-terminal TPR array. This part of τ131 is not well-resolved in the TFIIIA-TFIIIC complex, possibly due to its unconstrained movement relative to the rest of the complex (Figure 3.4, C and D). The residues connecting the N-terminal and C-terminal TPR arrays are not resolved in both structures, suggesting that the two arrays are connected by a flexible ‘hinge’ domain. Interactions between τ131 N-terminal TPR, upstream DNA, and Brf1-TBP lead to the extended conformation of τ131. This state is additionally stabilized by contacts between the 'ring' domain of τ131 (residues 390 - 428) and τ91 (Figure 3.3, B and C). TFIIIA ZF9, bound to the beginning of ICR, becomes visible in this structure. The TFIIIA-TFIIIC-Brf1-TBP complex is resolved to higher resolution, thus it is used to describe protein-protein interactions within TFIIIC.

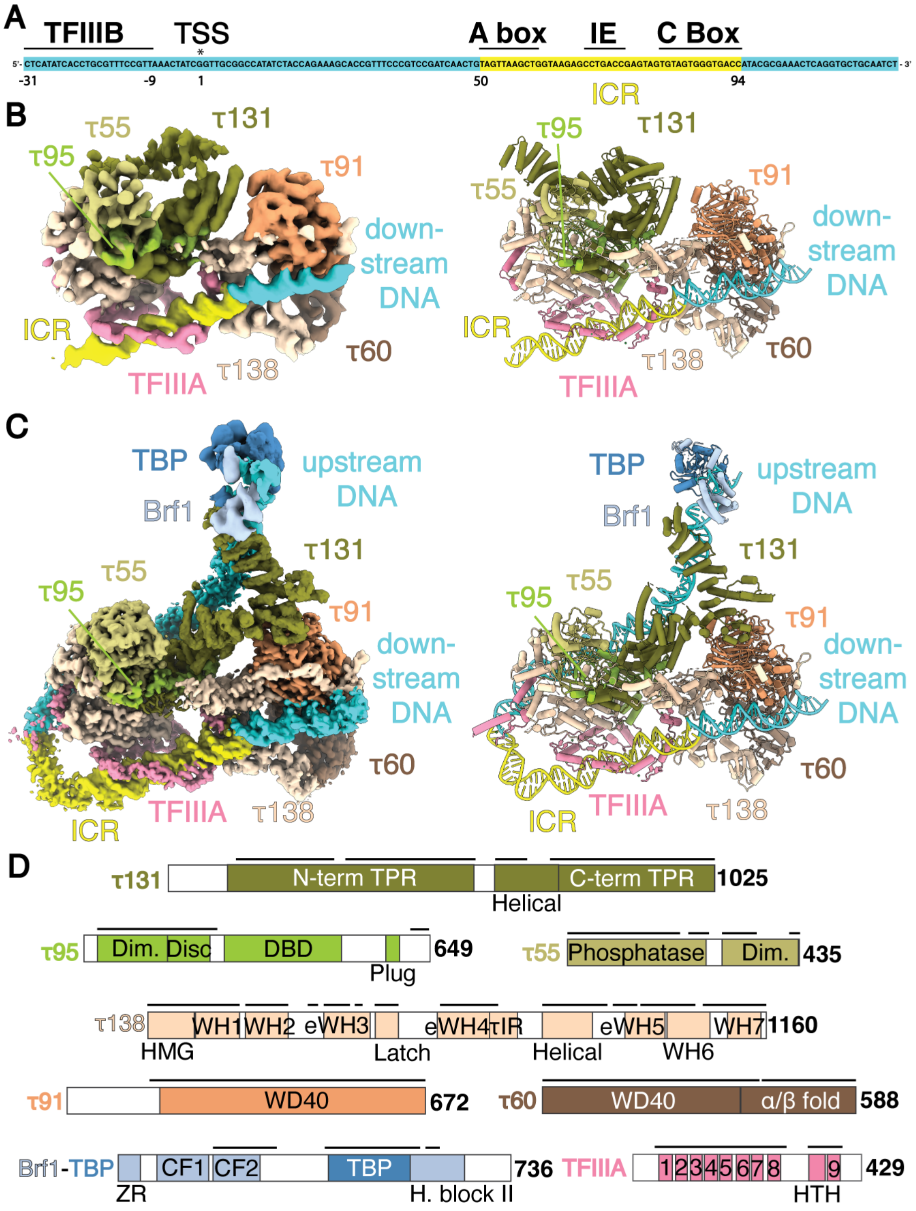

Figure 3.3. Cryo-EM structures of the TFIIIA-TFIIIC complex and the TFIIIA-TFIIIC-Brf1-TBP complex bound to the 5S rRNA gene. A) Schematic representation of the DNA template, with transcription start site (TSS) and intermediate element (IE) indicated. B) Density map of the TFIIIA-TFIIIC complex (left) and the corresponding model of the TFIIIA-TFIIIC complex (right). The internal control region (ICR) is highlighted. C) Composite density map of the TFIIIA-TFIIIC-Brf1-TBP complex (left) and the corresponding model of the TFIIIA-TFIIIC-Brf1-TBP complex (right). D) Schematic domain representation of TFIIIA, TFIIIC, and Brf1-TBP subunits. The amino-acid lengths of the subunits are labeled at the C-termini. Black lines above the bars show the portions for which models were built. Dim., dimerization domains; DBD, DNA-binding domain; HMG, high mobility group box domain; WH, winged helix; eWH, extended winged helix; τIR, τ131-interaction region; ZR, zinc ribbon; CF1,2, cyclin fold 1,2; H. block II, homology block II; TFIIIA 1-9, TFIIIA zinc fingers 1-9; HTH, helix-turn-helix.

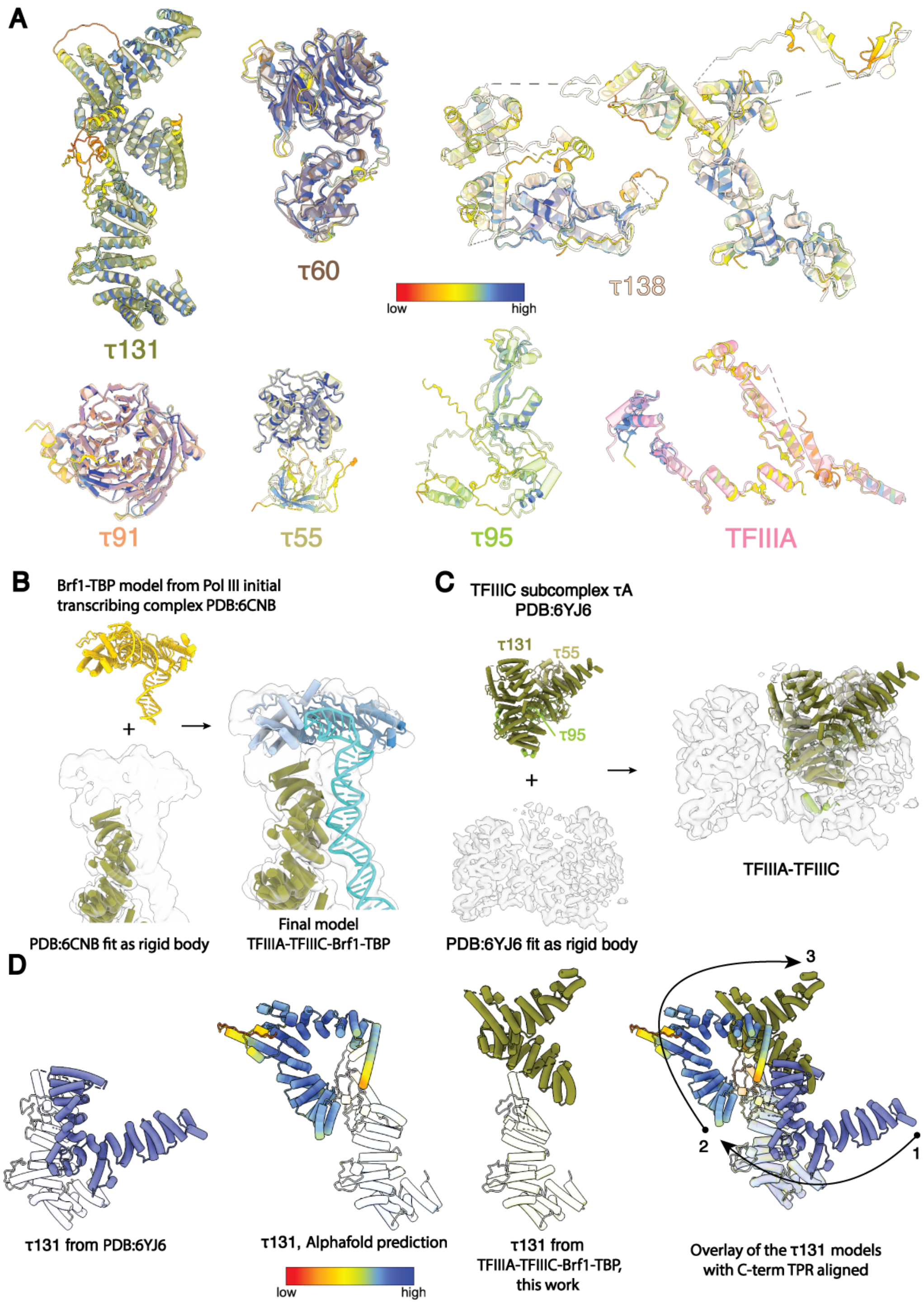

Figure 3.4. Model building details. A) Alphafold structure prediction of individual domains in TFIIIA-TFIIIC-Brf1-TBP complex guided by the density map. The color bar from red (low) to

blue (high) indicates the pLDDT confidence measure. B) Brf1-TBP model on DNA was built using the Brf1-TBP-DNA model from PDB:6CNB (yellow). This model was fit as rigid body, and model fragments out of density were trimmed. C) N-terminal TPR of τ131 in TFIIIA-TFIIIC complex was built using PDB:6YJ6. Subunit τ131 from the published model was aligned with C-terminal TPR from TFIIIA-TFIIIC-Brf1-TBP. The aligned model of τ131 from PDB:6YJ6 was fit as rigid body into the TFIIIA-TFIIIC map. Identical colors are used for τ131, τ95 and τ55 subunits in both models. D) Comparison of possible τ131 conformations. C-terminal TPR is shown as transparent tubes. N-terminal TPR is shown as opaque tubes. The color bar from red (low) to blue (high) indicates the pLDDT confidence measure for the Alphafold prediction. The overlay of aligned τ131 models suggests how the N-terminal TPR may move relative to the C-terminal TPR. The numbers from 1 to 3 indicate the trajectory of N-terminal TPR array movement from the most closed (1) to the most open (3) conformation.

3.3.2 τ138 bridges TFIIIC τA and τB lobes

The two largest subunits of TFIIIC, τ138 and τ131, facilitate interactions between the τA and τB lobes (Figure 3.5, A and B). This is consistent with previous genetic study that identified functional connections between τ138 and τ131 in *S. cerevisiae* (*121*). These two large subunits also play a role in supporting interactions between subunits within each lobe. TFIIIC is known to accommodate varying lengths of DNA within ICR (*47, 122*). The observed interactions between the two lobes may contribute to this intriguing ability of TFIIIC.

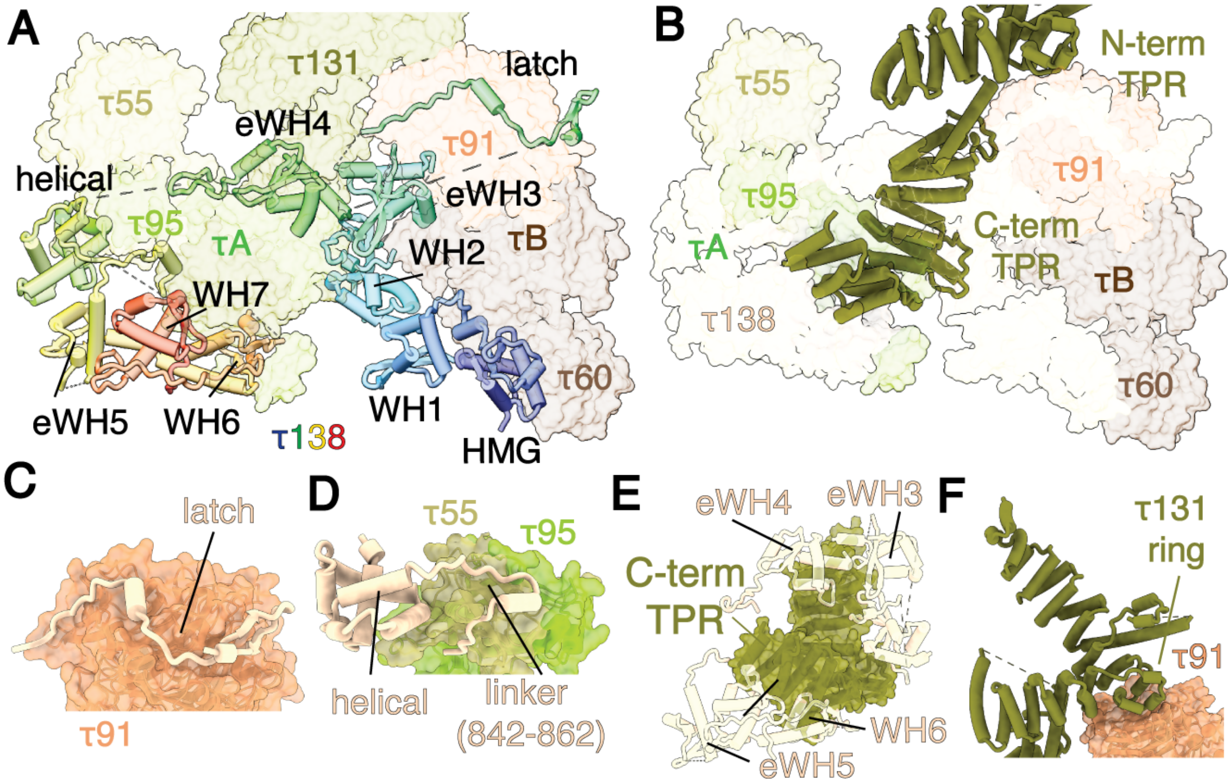

Figure 3.5. Bridging of TFIIIC τA and τB lobes by the two largest subunits, τ138 and τ131. A) The τ138 subunit acts as a scaffold for the body of TFIIIC and mediates the interaction between τA and τB lobes. The τ138 subunit is shown as cartoon, while other TFIIIC subunits are shown as transparent surfaces. The τ138 subunit is colored from N-terminus to C-terminus in a rainbow pattern. B) The τ131 subunit is located in the middle of the complex. The C-terminal TPR array of τ131 is buried in the body of the complex, while the N-terminal TPR array is extended away from the body of the TFIIIA-TFIIIC-Brf1-TBP complex. The τ131 subunit is shown as cartoon, with other TFIIIC subunits depicted as transparent surfaces. C) The latch of τ138 (residues 425-470) is attached to the surface of τ91. D) The helical domain and linker (residues 842-862) of τ138 bind to τ55. E) Contacts between τ138 and τ131. The C-terminal TPR of τ131 interacts with eWH4 (residues 1000-1023) and WH7 of τ138. F) The N-terminal TPR array of τ131 is in the extended

state, protruding away from the C-terminal TPR array and the body of TFIIIC. This position is stabilized through interactions of the τ131 ring domain with the τ91 subunit.

The domains of τ138, the largest TFIIIC subunit, are distributed between the two lobes (Figure 3.5A). The N-terminus of τ138 belongs to the τB lobe, the middle region is situated in the center, and C-terminus is a part of the τA lobe. The compact τA and τB regions of τ138 are connected by a less structured region (residues 418-739) that contains the eWH4 domain in the middle. Part of this region (residues 641-693) has been shown to be the main link between the two lobes of TFIIIC in *S. cerevisiae* (*46*). Subunit τ138 comprises seven WH domains, three of which are eWH (Figures 3.3D, 3.5A). The HMG, WH2, and eWH3 domains bind to τ60 subunit, and eWH3 contacts τ91 (Figure 3.6A), consistent with previous biochemical work in yeast (*50, 123*). The 'latch' domain of τ138 (residues 425-470) is attached to the surface of τ91 WD40 (Figure 3.5C). The eWH4 domain interacts with the helical domain of τ131 (residues 612 - 732), and the linker (residues 643-657) connects eWH4 with the τ55-τ95 dimer (Figure 3.6B). The helical domain of τ138 resides in the τA lobe and interacts with the τ55-τ95 dimer (Figure 3.5D). eWH5, WH6, and WH7 form a compact structure, interacting with the τ55-τ95 dimer and C-terminal region of τ95 (residues 612-647) (Figure 3.6C), consistent with previous crosslinking mass spectrometry results (*46*). Subunit τ138 serves as hub that brings together all other parts of the complex by interacting with five other TFIIIC subunits, TFIIIA and DNA (Figures 3.5A, 3.7C, 3.9B). Additionally, the two largest subunits, τ138 and τ131, also interact with each other (Figure 3.5E).

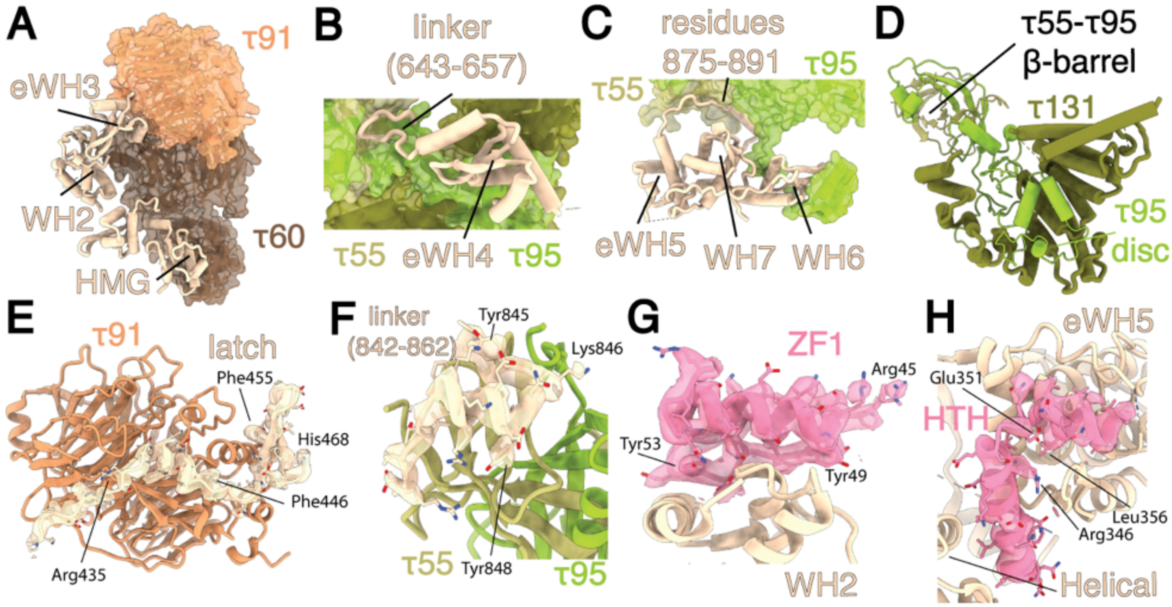

Figure 3.6. Subunit interactions within TFIIIC. A) The WD40 domain of τ60 supports the WH2 and eWH3 domains of τ138, while the α/β fold of τ60 contacts the HMG domain of τ138. The WD40 domain of τ91 supports the eWH3 domain of τ138. B) The eWH4 domain of τ138 contacts the helical domain of τ131, followed by a linker (residues 643-657) binding to the τ55-τ95 dimer. C) The residues 875-891 of τ138 contact the τ55-τ95 dimer. The eWH5, WH6 and WH7 domains form a compact fold, interacting with the τ55-τ95 dimer and the C-terminal region of τ95 (residues 612-647). D) Interactions within τA lobe: the dimerization domains of τ55 and τ95 subunits form a β-barrel. The disc domain of τ95 interacts with the C-terminal TPR array of τ131. (E-H) Regions of the final cryo-EM maps overlaid with their corresponding refined structural models. Cryo-EM densities are depicted as semi-transparent surface and selected residues in the model are denoted for orientation. The shown regions of TFIIIC τ138 include the latch (E) on the surface of τ91 and the linker (842-862) F) contacting the τ55-τ95 dimer. The regions of TFIIIA include ZF1 (G) contacting WH2 of τ138 and HTH (H) contacting eWH5 and Helical domain of τ138.

τ131 is a component of the τA lobe and is located in the middle of the structure (Figure 3.5B). The TPRs of τ131 are divided into two modules: N-terminal TPR and C-terminal TPR. The N-terminal TPR is further subdivided into two 'arms' with a 'ring' domain between them (*46, 124*). The concave surface of the C-terminal τ131 TPR array accommodates the 'disc' domain (residues 161-236) of τ95 (Figure 3.6D), which is in agreement with previous studies in yeast and humans (*46, 48, 125*). The ring domain of N-terminal TPR also contacts τ91 (residues 259-280) (Figure 3.6F). Notably, the area of contact between the lobes is smaller in comparison to the area of contact within each lobe (Figure 3.5, C to F, Figure 3.6, A to D). τB lobe has a buried surface area of 6669 $Å^2$, while τA lobe buries a large area of 15828 $Å^2$. However, the buried surface area between τA and τB is only 675 $Å^2$.

3.3.3 5S rRNA gene wraps around the complex

The DNA construct used in the experiment is composed of the 5S rRNA gene sequence extended to the upstream TFIIIB-binding region (Figure 3.3A) (*126, 127*). The TFIIIB-binding sequence (bp -31 to -9) is followed by the TSS (position 1), the ICR (bp 50 to 94), and the downstream region. The location of Brf1-TBP and TFIIIA are in excellent agreement with the registers of the TFIIIB-binding site and the ICR in the 5S rRNA gene, as determined by tracing the DNA density in the full complex reconstruction. The DNA wraps around the complex and makes multiple contacts with TFIIIA, TFIIIC, and Brf1-TBP (Figure 3.7; Figure 3.9A). The path of the bound DNA appears to make a roughly 180 degrees turn, bringing the upstream and downstream regions closer together. The ICR is the most sharply bent region of the DNA (Figure 3.7A). These findings align with previous research in which both TFIIIA and TFIIIC were shown to introduce bending within DNA in *X. laevis* and *S. cerevisiae* (*128-131*). To our knowledge, Pol III system represents the first eukaryotic visualized complex that requires sharp DNA bending in the middle of promoter, which helps one transcription factor (TFIIIC) to position another transcription factor (TFIIIB). Interestingly, similar sharp DNA bending and wrapping is required for transcription initiation in *E. coli*, which was observed in FRET experiments (*132*).

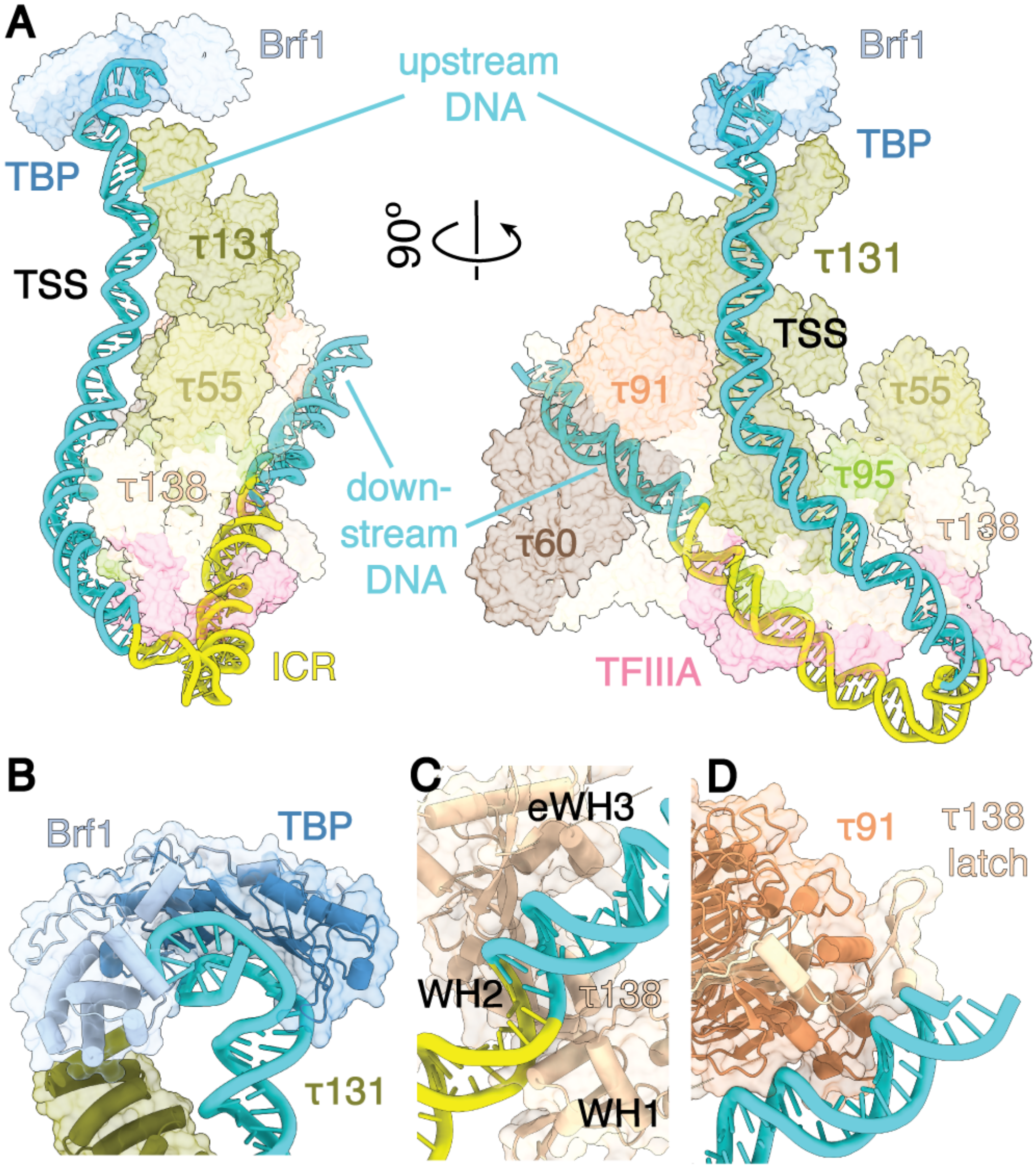

Figure 3.7. Brf1-TBP binding leads to the 5S rRNA gene wrapping around the TFIIIA-TFIIIC body. A) DNA wraps around the TFIIIA-TFIIIC-Brf1-TBP complex. The largest protein surfaces that interact with DNA are τ131 and TFIIIA. DNA is shown as cartoon, and proteins are shown as transparent surfaces. B) TBP bends upstream DNA region. This part of DNA is stabilized by τ131, TBP, and Brf1. C) τ138 WH1, WH2, and eWH3 wrap around downstream DNA. D) The end of the 5S rRNA gene is supported by the τB lobe. Both τ91 and the τ138 latch (residues 425-470) are associated with this part of the DNA.

The upstream TFIIIB-binding region is recognized by TBP and Brf1 (*127, 133*). The resolution of this part of the map is not sufficient to allow *de novo* model building, so the Brf1-TBP-DNA model (PDB ID 6cnb) was docked in the map as rigid body (Figure 3.7B; Figure 3.4B). The DNA upstream of the ICR is bound to positively charged patches on the τA lobe, primarily τ131 (residues 192 to 252; 667 to 718; 832 to 841; 924 to 931) (Figure 3.7A; Figure 3.8, A and B). The binding of τ131 to the upstream region of the *S. cerevisiae* 5S rRNA and SUP4 tRNA Tyr genes has also been previously shown through site-specific DNA-protein photocrosslinking (*133, 134*). The first three WH domains of τ138 are wrapped around the DNA downstream of the ICR, following the minor groove of DNA (Figure 3.7C). The positively charged regions of WH1 and eWH3 have close contacts with the DNA minor and major groove, and similar examples of WH-DNA interactions can be found in other transcription initiation complexes (Figure 3.8D). The DNA-binding surfaces of τ138 and τ131 have highly conserved residues forming positively charged patches (Figure 3.8, A to C; Figure 3.8, E and F). The downstream region of the DNA is bound by the WD40 domain of τ91 subunit (Figure 3.7D; Figure 3.8C). This subunit has been shown to photo-crosslink to the very end of the 5S rRNA gene (*133*), while the *S. pombe* homolog of τ91, Sfc6p, has been shown to recognize the B-box in Type II promoter (*135*). This part of the DNA is additionally supported by the τ138 latch (residues 449 - 470) (Figure 3.7D).

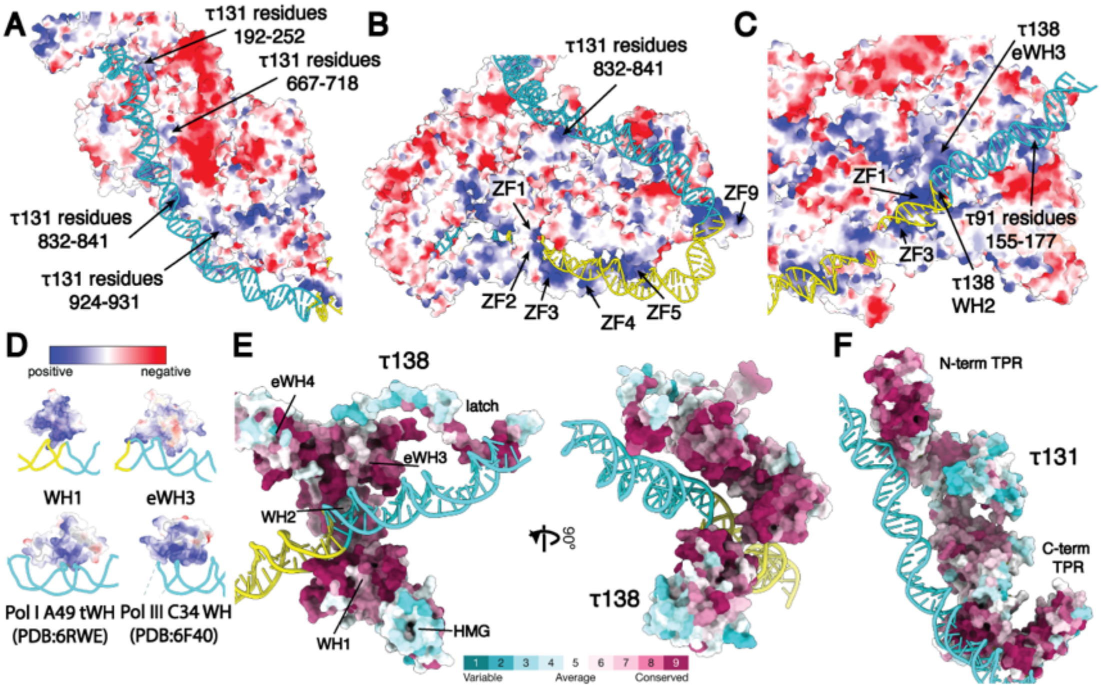

Figure 3.8. TFIIIC interactions with DNA. (A-C) TFIIIA-TFIIIC-Brf1-TBP shown as surface and colored according to coulombic potential. The model is colored in a range from red for negative potential to blue for positive potential. Upstream (A), center (B), and downstream (C) regions of DNA are bound to positively charged patches of the protein complex. (D) Comparison of DNA-binding WH from this work (top row) with several published DNA-binding WH (bottom row). All WH domains are shown as surface and colored according to coulombic potential. The models are colored in a range from red for negative potential to blue for positive potential. (E-F) Regions of τ138 (E) and τ131 (D), which are in proximity of DNA, are colored according to the conservation score (prepared using ConSurf web server). The color bar from cyan (low) to maroon

(high) indicates ConSurf sequence conservation score. DNA-contacting regions have the highest conservation score. DNA colors correspond to the colors used in the main figures.

3.3.4 TFIIIA binds the ICR of the 5S rRNA gene

DNA footprinting assay has revealed that both *X. laevis* and *S. cerevisiae* TFIIIA bind to the ICR of the 5S rRNA gene (*127*). DNA binding of the first three ZF in our model is identical to the *X. laevis* TFIIIA-DNA structures but deviates for the following ZF, likely due to the presence of its binding partner, TFIIIC, in the full complex (Figure 3.10A) (*56, 127, 136*). All nine ZFs share the same fold, but ZF 9 has a longer helix (Figure 3.10B). The ICR is protected by TFIIIA ZF 1 to 5 and ZF 9, while ZF 6 to 8 point away from DNA (Figure 3.7A; Figure 3.9A). Previous study has shown that purified TFIIIA protects the ICR of the 5S rRNA gene (bp 66 to 95) from DNase I cleavage, with enhanced cleavage at bp 50 and 65, consistent with our structural model (*127*). The footprints of the five N-terminal ZF bound to 5S rRNA gene were indistinguishable from binding of full-length TFIIIA, and it was previously suggested that ZF 6 to 9 do not bind DNA tightly (*136-138*). Consistently, ZF 1 to 3 in our model bind in the DNA major groove (Figure 3.9A; Figure 3.10C) (*54, 139*). ZF 4 traverses the minor groove, while ZF 5 and 9 bind the major groove again (Figure 3.9A; Figure 3.10C). The sharpest part of the DNA bend is located between ZF 5 and ZF 9 (Figure 3.9A). All DNA-binding ZF show positive charge and high conservation of their DNA-contacting surface (Figure 3.9, C and D; Figure 3.10D).

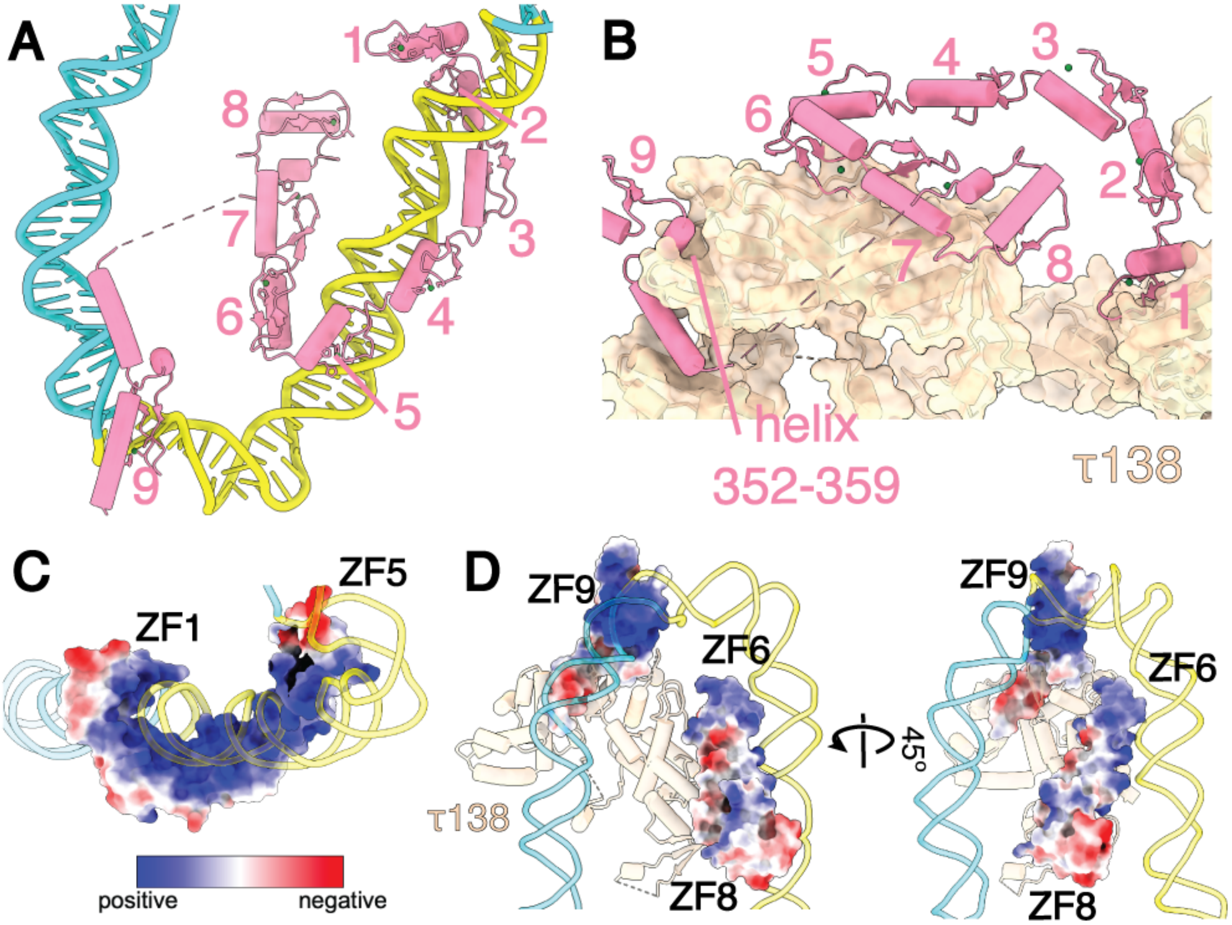

Figure 3.9. TFIIIA binds the ICR of the 5S rRNA gene. A) ZF 1 to 5 and ZF 9 bind DNA in the ICR, while ZF 6-8 protrude away from ICR. The most sharply bent region of DNA is located between ZF 5 and 9. B) TFIIIA has a large interaction surface with τ138. ZF 1, 6-8, and a helix near ZF 9 contribute to the interaction. C) DNA-binding residues of ZF 1 to 5 form a large, extended, positively charged surface. TFIIIA is shown as surface and colored according to coulombic potential. The model is colored in a range from red for negative potential to blue for positive potential. D) ZF 6 to 8 do not interact with DNA as closely as the first five ZFs, but their positively charged surfaces are directed towards DNA and away from τ138. ZF9 DNA binding surface is positively charged. TFIIIA is shown as surface and colored according to coulombic

potential. The model is colored in a range from red for negative potential to blue for positive potential.

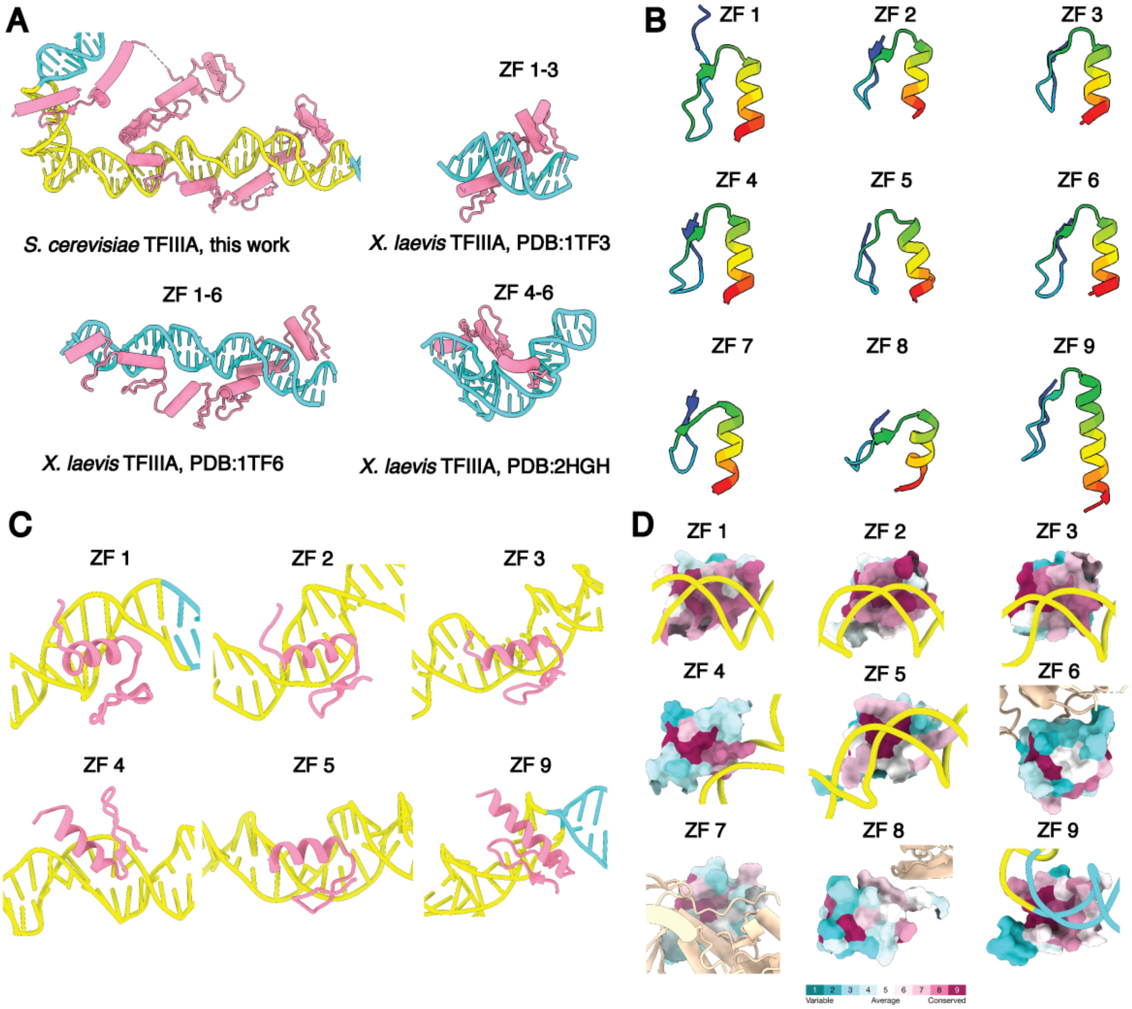

Figure 3.10. Comparison of TFIIIA ZFs. A) Comparison of yeast TFIIIA with published models of *X. laevis* TFIIIA. In all cases, TFIIIA is bound to DNA, but in PDB:2HGH, TFIIIA binds RNA. ZF fold and linkers between ZF are similar between species. ZF 1 to 3 bind DNA similarly. However, the following ZFs interact with DNA in a different mode. B) Aligned TFIIIA ZFs are colored in rainbow from N-terminus to C-terminus. Nine ZFs of *S. cerevisiae* have identical C2H2 fold, and ZF9 has a longer helix. C) Comparison of DNA-binding TFIIIA ZF. ZF 1,2,3,5,9 bind

the major grove and ZF4 binds the minor grove. D) TFIIIA ZFs are aligned and colored according to conservation score (prepared using ConSurf web server). The color bar from cyan (low) to maroon (high) indicates ConSurf sequence conservation score. DNA-contacting ZFs have the highest conservation score on their DNA-contacting surface. DNA colors correspond to the colors used in the main figures.

τ138 was previously suggested to bracket TFIIIA on 5S rRNA gene (*133*). Our observation reveals that the contact between TFIIIA and TFIIIC is maintained through τ138, with a large surface area between ZF 6-8 of TFIIIA and residues 980-1072 of τ138 (Figure 3.9B). Additionally, ZF 1 and helix-turn-helix domain (residues 331 - 363) of TFIIIA also contribute to this interaction. This region within the C-terminus of *Xenopus* TFIIIA has been identified as a TFIIIC binding and non-DNA binding site (*140*). ZF 7 was found to be essential for the assembly of transcriptionally active complex in yeast (*141*). The presence of ZF 7 to 9 has also been shown to be necessary for the transcription activity of the complex, likely due to higher-order interactions in the complex (*137, 138*). A flexible linker between ZF 8 and 9 is not visible in the map, except for the helix-turn-helix domain (residues 331 - 363) right next to ZF 9 (Figure 3.9A). TFIIIA lacking this region (residues 283 - 364) has been shown to be able to recruit TFIIIC but unable to promote transcription in yeast (*137*). Specifically, the deletion of a leucine-rich segment 352-NGLNLLLN-359, a small helix next to ZF 9, resulted in the loss of transcription activity in *S. cerevisiae* (*142*). This helix appears to be an anchor point of ZF 9 on the surface of τ138 (Figure 3.9B).

3.3.5 Single-molecule FRET shows dynamic nature of the complex

To understand the conformational dynamics of the complex and independently verify our structural model, we perform smFRET assay using the full TFIIIA-TFIIIC-Brf1-TBP complex (Figure 3.11A). We purified the complex using a dsDNA molecule labeled with Cy3 and Cy5, following the same protocol as for the cryo-EM sample preparation. The positions of Cy3 and Cy5 were chosen such that, based on our structure, they are 66 Å apart in the assembled complex, which would allow for efficient FRET (Figure 3.7A). The DNA-only sample shows stable FRET close to zero, indicating an extended conformation of the DNA (Figure 3.11B, Figure 3.12A). The TFIIIA-TFIIIC-Brf1-TBP complex assembled on the DNA shows two higher FRET states (Figure 3.11C, Figure 3.12B). The FRET histogram shifts to higher FRET, indicating a decrease in the distance between the fluorophores (Figure 3.11D). Notably, most traces (62%) display slow transitions between the two FRET states at 0.07 and 0.22 before photobleaching (Figure 3.11, C and E; Figure 3.12B). The second FRET peak for the full complex at a FRET value of 0.22 corresponds to a generic distance of 66 Å, which is consistent with our prediction from the structure, confirming that this is the wrapped DNA conformation. To further investigate the nature of the low FRET state, we conduct the experiment on the complex that is assembled without Brf1-TBP (Figure 3.12C). The FRET histogram for this complex shows a single major peak at a FRET value of 0.07, similar to the lower FRET peak for the full complex, suggesting that this represents a dynamic, partially unwrapped intermediate (Figure 3.11D). The presence of this lower FRET state in TFIIIA-TFIIIC-Brf1-TBP complex trajectories suggests that this complex is dynamic and the upstream DNA is not always bound by both TFIIIC and Brf1-TBP. We performed the experiment with the DNA and TFIIIA only (Figure 3.12, D and E). Interestingly, we observed a small percentage of DNA molecules showing very brief and infrequent visits to a higher FRET state, consistent with the possibility that TFIIIA alone can induce transient bending of DNA. The frequency and lifetime of these bent states was much shorter than what we observed in the presence of all the three proteins, suggesting that TFIIIC and Brf1-TBP are essential for supporting the bent DNA.

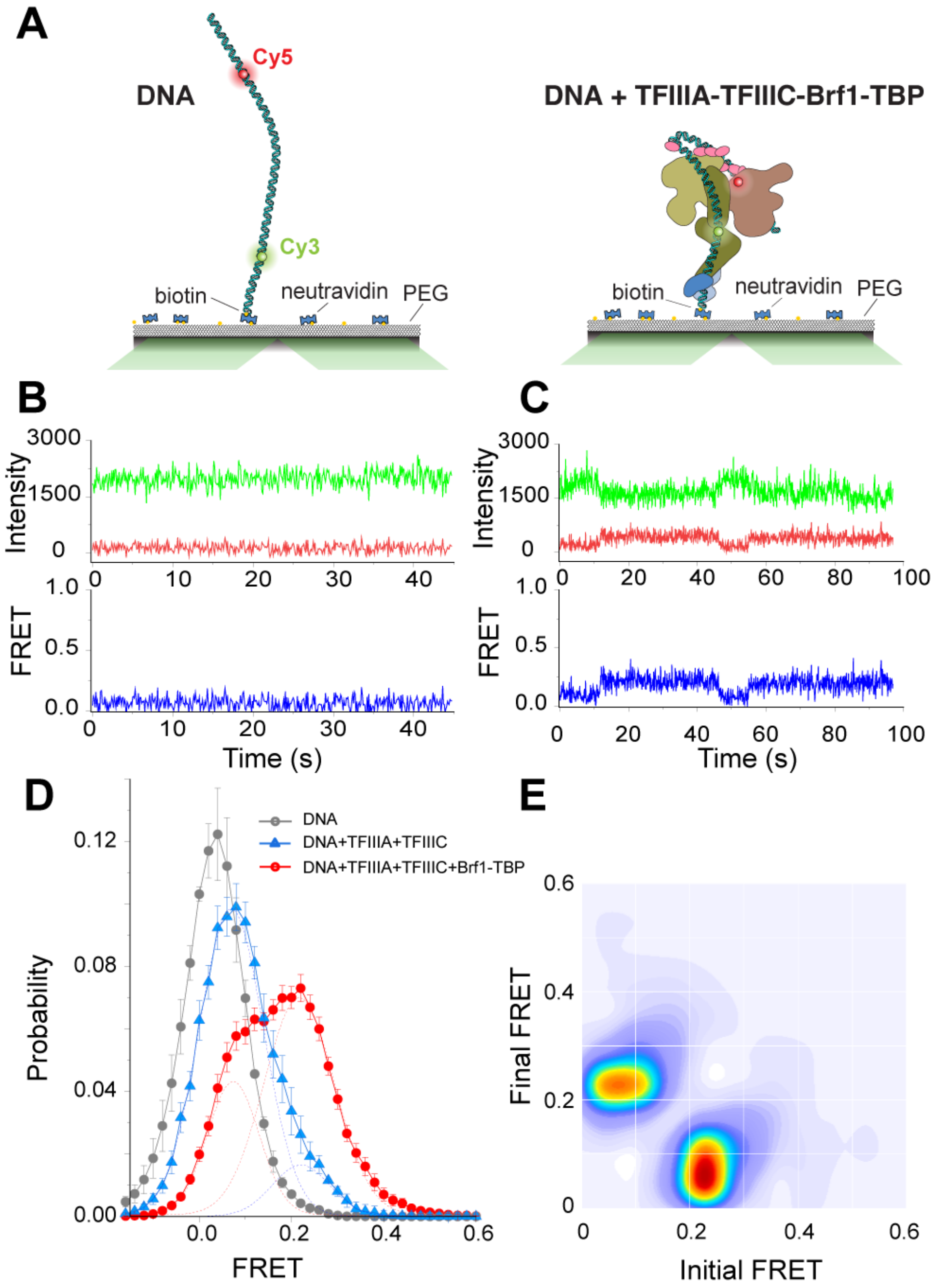

Figure 3.11. smFRET reveals DNA bending that is consistent with the cryo-EM study. A) Schematic representation of the smFRET assay. Left, labeled DNA; right, labeled DNA + TFIIIA-TFIIIC-Brf1-TBP complex. B) Representative single-molecule time traces of DNA only showing donor (green) and acceptor (red) intensities and the corresponding FRET (blue). C) Representative

single-molecule time traces of the TFIIIA-TFIIIC-Brf1-TBP complex showing donor (green) and acceptor (red) intensities and the corresponding FRET (blue). The FRET data shows the presence of two states. D) smFRET population histogram of DNA only, DNA + TFIIIA-TFIIIC, and DNA + TFIIIA-TFIIIC-Brf1-TBP. Data represents mean±s.e.m. The histogram of DNA + TFIIIA-TFIIIC is fitted to two Gaussian distributions (blue dashed line) centered on 0.07 and 0.22, respectively. The histogram of the full complex, DNA + TFIIIA-TFIIIC-Brf1-TBP, is also fitted to two Gaussian distributions (red dashed line) centered on 0.07 and 0.22, respectively. E) Transition density plot of the full complex, DNA + TFIIIA-TFIIIC-Brf1-TBP. Transitions are from two independent measurements.

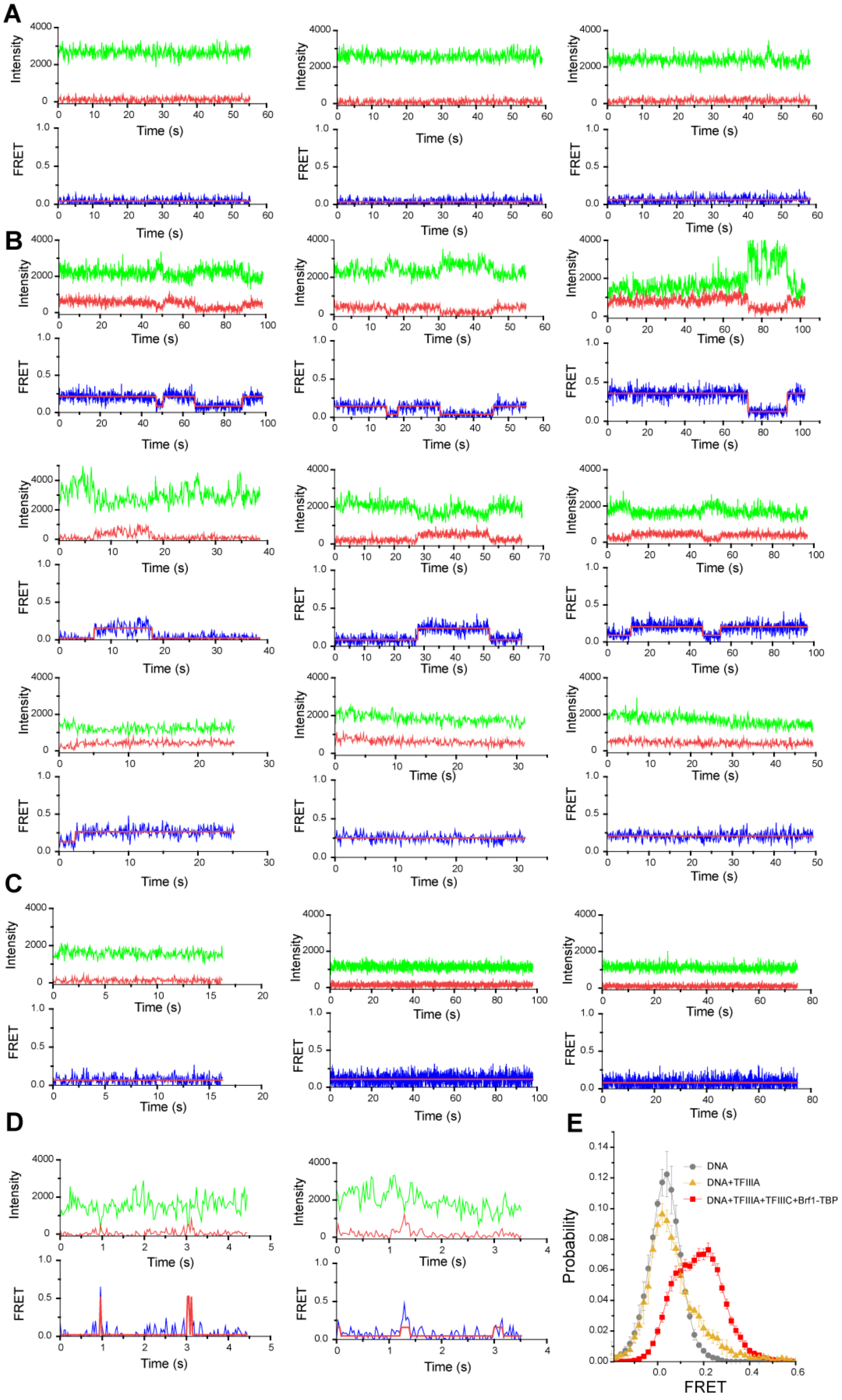

Figure 3.12. smFRET traces of DNA+TFIIIA+TFIIIC+Brf1-TBP. (A-D) Single-molecule time traces of different species showing donor as a green line and acceptor as a red line. The

corresponding FRET profile of the complex shows one or two states depicted as a blue line. A hidden Markov model (HMM) is applied to the time traces of the FRET efficiency to estimate the states, depicted as a red line over the FRET profile. The representative traces include DNA only (A), DNA+TFIIIA+TFIIIC complex (B), full complex DNA+TFIIIA+TFIIIC+Brf1-TBP (C), and DNA + TFIIIA (D). E) smFRET population histogram of DNA only, DNA + TFIIIA, and DNA + TFIIIA-TFIIIC-Brf1-TBP. Data represents mean±s.e.m.

CHAPTER 4: Conclusions and Discussion

4.1 Summary of findings

The process of eukaryotic transcription involves the formation of PIC, which consists of Pol II and a set of general transcription factors. Transcription factors may recruit the coactivator Mediator, which aids in PIC assembly and promotes the phosphorylation of the Pol II CTD by the TFIIH subunit CDK7. In this study, we have visualized the structure of the human Mediator-bound PIC. The binding sites for transcription factors within Mediator are mainly connected to the main body in a flexible manner. Mediator establishes several contacts with the CAK module to stabilize CDK7 kinase. We observe two distinct binding sites for the Pol II CTD: one between the head and middle modules of Mediator and the other within the active site of CDK7. This structural evidence provides mechanism of the Mediator-aided Pol II CTD phosphorylation within the Mediator-bound PIC.

Historically, biochemical studies of Pol III transcription machinery have shaped our understanding of eukaryotic transcription. However, structural basis of TFIIIC-dependent Pol III promoter recruitment remained elusive. In this study, we have visualized TFIIIA-TFIIIC and TFIIIA-TFIIIC-Brf1-TBP complexes assembled on 5S rRNA gene. The overall conformation of TFIIIC in both complexes is similar, but the N-terminal TPR of τ131 is poorly resolved in the TFIIIA-TFIIIC complex. This part of τ131 is stabilized by interactions with Brf1-TBP and upstream DNA in our second structure. The largest TFIIIC subunit τ138 links the two lobes of TFIIIC and interacts with the other five TFIIIC subunits, TFIIIA, and DNA. TFIIIA ZF1-5 and 9 are bound to the ICR of 5S rRNA gene, while ZF6-8 interact with τ138, representing a major contact between TFIIIA and TFIIIC. The DNA makes a 180° turn within the TFIIIA-TFIIIC-Brf1-TBP complex. We used smFRET to investigate the conformational dynamics of the DNA within

TFIIIA-TFIIIC-Brf1-TBP complex. The DNA-only control exhibited stable FRET close to zero, while the TFIIIA-TFIIIC-Brf1-TBP complex assembled on the DNA showed two higher FRET states at 0.07 and 0.22, with the latter corresponding to a 66Å distance of the two fluorophores, consistent with our cryo-EM structure. The FRET value of 0.07 was also observed in the TFIIIA-TFIIIC-DNA complex, indicating a partially unwrapped intermediate.

4.2 Mechanism of CTD phosphorylation in Mediator-PIC

The role of $_{\text{MED}}$CTD binding is likely to capture the CTD and position it in the correct direction and close to the active site of CDK7 to facilitate phosphorylated Ser5 (pSer5) formation. Mass spectrometry experiments with both yeast and human complexes show that pSer5 can be found within any repeat of the CTD except the final repeat (*143, 144*). However, the phosphorylation patterns of individual CTD peptides and the direction in which sequential phosphorylation can occur remain unknown. Two possibilities exist for the direction of sequential phosphorylation that generate different outcomes (Figure 4.1). If the CTD is phosphorylated in a C- to N-terminal direction, binding at $_{\text{MED}}$CTD precedes phosphorylation, and it is not clear how Pol II would dissociate from Mediator given that the CTD is threaded through a hole in Mediator formed by the hook, knob, and shoulder domains and the CAK module of TFIIH. Phosphorylated repeats would also be located far from the nascent RNA that needs to be capped.

If the CTD is phosphorylated in an N- to C-terminal direction, C-terminal phosphorylated repeats would not be able to bind at $_{\text{MED}}$CTD because of steric clashes that would arise with the added phosphates. Given that the CTD is important for Pol II–Mediator interaction and that phosphorylation of the CTD leads to dissociation of Pol II and Mediator, we find this mechanism more likely (*22, 145*). Separation of MedHead and Pol II would place the phosphorylated CTD close to the nascent RNA for capping to occur.

Given the large movements of MedMiddle and the CAK module of TFIIH relative to the PIC, we speculate that these conformational changes play an important role in the sequential phosphorylation of the CTD. The intrinsic flexibility of Mediator has been linked to the opening and closing of the $_{\text{MED}}$CTD binding site on Mediator (5, 6), and if this movement is tied to binding and release of the CTD at $_{\text{MED}}$CTD, it could also facilitate the progression of CDK7 along the CTD.

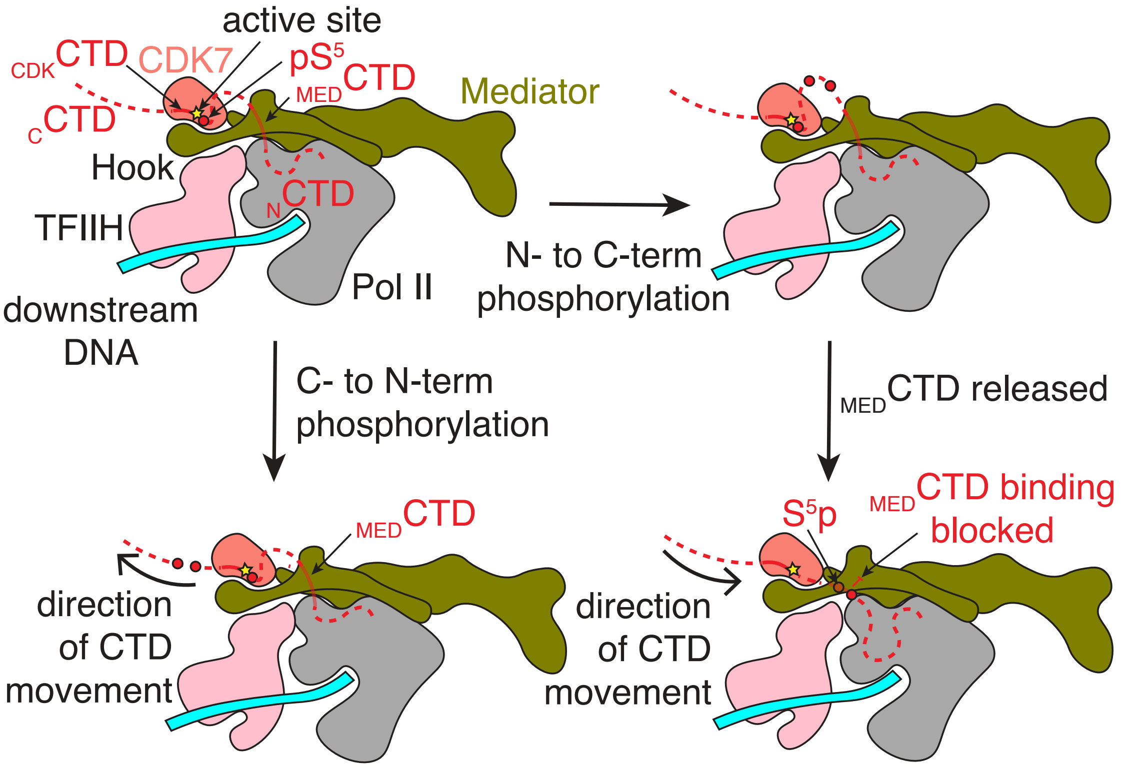

Figure 4.1. Model for phosphorylation of the Pol II CTD by CDK7. $_{\text{MED}}$CTD binding positions the rest of the CTD in the CDK7 active site. After phosphorylation, which is indicated by a red circle, translocation of the CTD toward the N terminus (bottom) would place phosphorylated repeats further from the nascent RNA emerging from Pol II. Separation of Mediator and Pol II

would be difficult without separation of the CAK module and Mediator. Translocation of the CTD toward the C terminus would position phosphorylated repeats to block binding of the CTD at $_{\text{MED}}$CTD, a possible way to favor disassembly of Med-PIC. Phosphorylated repeats would also be substantially closer to the RNA exit tunnel of Pol II to recruit the capping complex properly. CTD, C-terminal domain of RPB1; pS^5, phosphorylated Ser^5 residue (red circle).

4.3 Recent structural studies of Mediator-PIC and TFIIIC

Several other structures of Mediator in complex with Pol II PIC were recently published. Together, these structures provide better understanding of the Mediator – PIC interactions and the role of Mediator in the process of PIC assembly and transcription initiation. Cryo-EM studies of mammalian Mediator include one murine structure (*111*) and two more structures of the human complex (*146, 147*). The Mediator complex in these studies was either endogenously or recombinantly purified. The overall architecture of Mediator in the available structures is nearly identical. Another common feature of the structures includes unresolved parts of Mediator subunits, that are known to interact with activators and may represent intrinsically disordered regions. These subunits often contain intrinsically disordered regions (IDRs) and were suggested to form transcription hubs, or phase-separated condensates (*148, 149*). It was shown that Mediator tends to form such condensates (*150*).

Fragments of Pol II Rpb1 CTD bound to the Mediator surface were visualized for human and yeast Mediator (*146, 151*). Interestingly, Chen et. al. (*146*) also report two CTD fragments bound to the surface of Mediator, however, the CTD regions are different from fragments visualized in our structure. The longest stretch of CTD (11 out of 26 heptapeptide repeats) bound to Mediator was visualized for yeast protein (*151*).

Another recent work presented the structure of Mediator-PIC on a nucleosome DNA template (*152*). This structure showed how the Mediator hook domain interacts with +1 nucleosome.

The structure of human TFIIIC complex unbound and bound to tRNA gene was published recently (*153*). The overall architecture and subunit composition and conformation closely resembles its yeast counterpart. In this structure, only τB lobe of TFIIIC binds the B box region of tRNA gene, which is different from 5S rRNA gene, that interacts both with τA and τB lobes of TFIIIC. The human TFIIIC-tRNA gene structure does not capture A box interaction with the complex, and the τA lobe is not tightly connected to the τB lobe.

4.4 Mechanism of Pol III transcription initiation complex assembly

Based on the cryo-EM structures and smFRET data, we propose a model for the TFIIIA-TFIIIC-mediated assembly of the Pol III PIC. Before the complex assembles, the two lobes of TFIIIC, connected by the τ138 linker, may move relative to each other (Figure 4.2, state 1). This relative flexibility of the lobes have been previously observed by EM in yeast (*47*). We also observe different positions of the TFIIIC lobes by negative stain EM (Figure 4.3). The assembly of the Pol III PIC is thought to begin with TFIIIA locating the 5S rRNA gene (*127, 154, 155*), likely introducing initial bending that is mostly localized within the ICR (*129, 130*). In our structure, only ZF 1 to 5 and ZF 9 are bound to DNA. However, when TFIIIA first binds DNA, all nine ZF may make DNA contacts, similar to *X. laevis* TFIIIA (*156, 157*). In our structure, ZF 1 to

5 bind 29 bp, while ZF 9 protects 5 bp of the ICR. This leaves about one turn of DNA (11 bp) that, in principle, can be occupied by ZF 6 to 8 in this initial phase. The binding of ZF 6 to 8 may be similar to the DNA-binding mode of ZF 1 to 3. Compared to the first five ZF, ZF 6 to 8 have smaller positively charged areas, which may make them easier to be peeled off from DNA when the TFIIIA-TFIIIC complex forms (Figure 4.2, C and D). Alternatively, ZF 6 to 8 may be in search of TFIIIC instead of binding to DNA. This first 'searching' state can be characterized by the flexibility of the TFIIIA-DNA complex and within the unrestrained TFIIIC.

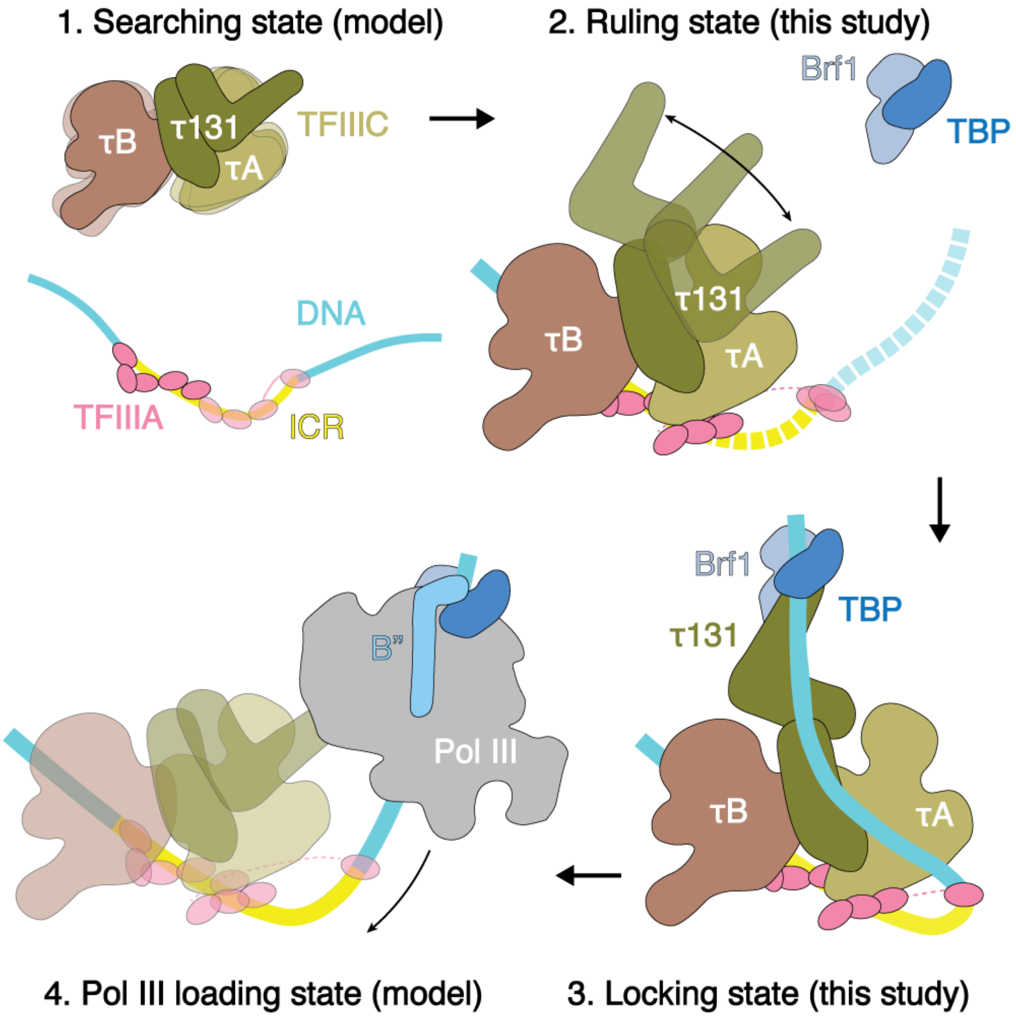

Figure 4.2. Model of TFIIIC-dependent Pol III PIC assembly on 5S rRNA gene.

1. Searching state (model): The two lobes of TFIIIC are not restrained by DNA binding and may move relative to each other. TFIIIA recognizes the ICR of the 5S rRNA gene. ZF 6 to 9 are shown as semi-transparent cartoons to reflect the uncertainty in their position.

2. Ruling state (formation of the TFIIIA-TFIIIC-DNA complex, visualized in this study): TFIIIC binds TFIIIA and DNA downstream of ICR. These interactions restrain mobility between the τA and τB lobes of TFIIIC. The 'hinge' region between the N-terminal TPR and C-terminal TPR of τ131 allows the N-terminal TPR (shown as semi-transparent surface) to transition from the closed state (N-terminal TPR contacts C-terminal TPR) to the fully open state (N-terminal TPR array is turned by 180 degrees relative to its position in the closed state). This sampling movement may help τ131 N-terminal TPR to search for Brf1 and/or DNA for binding. The ICR and downstream DNA is fixed via interactions with ZF 1-5 of TFIIIA and the τB lobe. The flexibility within the upstream DNA helps to search for τ131 and/or Brf1 binding. This complex may represent the 'ruling' state of τ131 because the distance between the N-terminal and C-terminal TPRs of τ131 is variable, but strong τ131-DNA interaction is only possible when N-terminal TPR is in the extended conformation. This allows τ131 to measure the distance from ICR to the TFIIIB-binding region. Upstream DNA, not visible in the structure, is depicted as a dash line.

3. Locking state (formation of the TFIIIA-TFIIIC-Brf1-TBP-DNA complex, visualized in this study): N-terminal TPR array of τ131 is in the extended state. DNA is bent within the ICR, and upstream DNA is bound to the surface of the τA lobe. DNA upstream of TSS is bound by TBP, Brf1, and the N-terminus of τ131.

4. Pol III loading state (model): Once Brf1 and TBP locate the TFIIIB-binding site and bind it, B" and Pol III can be recruited. Pol III may then initiate transcription, while TFIIIC and TFIIIA (both shown as semi-transparent surfaces) have to be displaced from DNA to allow Pol III to transcribe the full length of the gene.

Once TFIIIC binds TFIIIA and downstream DNA, the mobility between the τA and τB lobes becomes limited (Figure 4.2, state 2). The ICR and downstream DNA is fixed via interactions with ZF 1-5 of TFIIIA and τB lobe. The upstream DNA may still move due to the presence of a flexible linker within TFIIIA. This movement of upstream DNA can assist in the search for τ131 and/or Brf1-TBP binding. The lower FRET state of 0.07 may correspond to this complex (Figure 3.11D). The C-terminal TPR array of τ131 is locked in the body of the complex, while the N-terminal TPR array is not restrained. The 'hinge' region between the N-terminal and C-terminal TPRs allows for the potential movement of the N-terminal TPR array from the closed state, where it contacts the C-terminal TPR (*48*), to the fully open state, where it is rotated by approximately 180 degrees relative to its position in the closed state (Figure 3.4D). In this 'ruling' phase, the conformational sampling of the τ131 N-terminus may help to search for Brf1-TBP and DNA. The variable distance between the N-terminal and C-terminal TPR arrays of τ131 can act as a ruler: simultaneous interactions of τ131 with TFIIIB and the TFIIIB-binding region of DNA are possible only when τ131 N-terminus is located within a certain distance from the C-terminal TPR. We propose this model of TFIIIC-aided TFIIIB recruitment as an extension of the previously communicated models (*48, 158-160*).

The TFIIIA-TFIIIC-Brf1-TBP complex, which we observe after adding Brf1-TBP to TFIIIA-TFIIIC complex, corresponds to 'locking' state (Figure 4.2, state 3). In this state, the N-terminal TPR array of τ131 is fixed in the open conformation. This position of τ131 opens a positively charged patch on its surface (Figure 3.8, A to C). Yeast two-hybrid assays have shown that Brf1 interacts with the N-terminus of τ131 subunit, and this result has been supported by mutagenesis analysis and binding assays (*124, 161-163*). The sharpest bend in DNA is located between TFIIIA ZF 5 and 9 within the ICR region. This bent DNA is stabilized through interactions with TFIIIA and τ131 (Figure 3.7A; Figure 3.9A; Figure 3.8, A to C). Upstream DNA is bound by TBP, Brf1, and the N-terminus of τ131. In the smFRET assay, this state is represented by the 0.22 FRET efficiency state (Figure 3.11D). DNAse I footprinting has shown that TFIIIA protects the ICR of the 5S rRNA gene, and the binding of TFIIIC to TFIIIA-DNA can extend the footprint in two ways: a 'core' footprint on the downstream DNA or 'extended' footprint up to upstream DNA region bp -20 (*127*). The addition of TFIIIB extends DNA protection up to bp -45, and the TFIIIA-TFIIIC-TFIIIB complex protects the DNA from bp -45 to bp 120 (*127*). Similar footprinting patterns have been observed for tRNA genes as well (*164*).

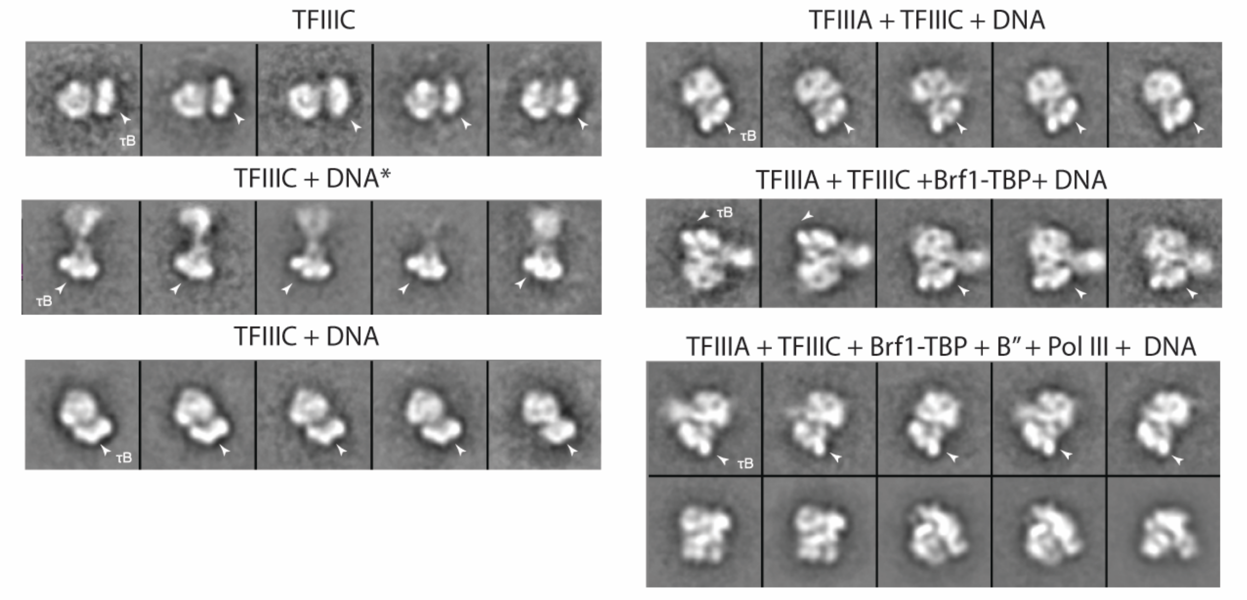

Figure 4.3. TFIIIC interactions with DNA, TFIIIA, Brf1-TBP, and Pol III. A) Selected 2D class averages from negative stained electron micrographs. TFIIIC lobe τB is indicated with white arrow. Each experimental condition shows either TFIIIC alone, bound to asparagine tRNA gene (DNA*), or 5S rRNA gene (DNA).

The TFIIIA-TFIIIC-Brf1-TBP state that we capture in this study is not compatible with Pol III binding. Alignment of Brf1-TBP from Pol III PIC structures with Brf1-TBP in our model results in severe clashes between Pol III and TFIIIC (Figure 4.4D). Therefore, it is suggested that direct interaction between Pol III and TFIIIC may be not necessary for Pol III recruitment to promoter. Previous study has shown that once TFIIIB is assembled on DNA, TFIIIC is dispensable for in vitro transcription (*43*). Additionally, yeast TFIIIC and Pol III occupancy on DNA are inversely correlated, as shown by chromatin immunoprecipitation (*63*). The interaction between Brf1, B" and τ131 also increases during transcription repression (*63*). Furthermore, smFRET data demonstrates that the DNA within the TFIIIA-TFIIIC-Brf1-TBP complex is not static and undergoes spontaneous transitions on slow time scales (10s of seconds) to a partially unwrapped state (Figure 3.11C). This state could be crucial for allowing Pol III to bind productively.

Finally, the TFIIIC complex needs to be displaced from DNA to allow Pol III transcription. The B" binding and completion of TFIIIB may trigger a transition to a new state where the contact between τ131 and Brf1-TBP is broken, resulting in the detachment of TFIIIC τA lobe from DNA (Figure 4.2, state 4). At the same time, the stable TFIIIB-DNA complex is assembled and can recruit Pol III (*43, 155*). This "Pol III loading' state may resemble yeast and human Pol III PIC structures (*51-53*). Once the PIC is formed, Pol III may initiate transcription. For transcription to occur over the full length of the gene, both TFIIIC and TFIIIA need to be displaced from the DNA.

We would like to discuss the role of the TFIIIA-TFIIIC complex in the context of chromatin. The 5S rRNA gene contains a strong nucleosome positioning sequence, and the nucleosome dyad position overlaps with TSS (*165, 166*). Untimely 5S nucleosome assembly can impede transcription initiation by Pol III. The TFIIIA-TFIIIC complex may play an important role in displacing the nucleosome, allowing for the formation of the transcription initiation complex. Once the nucleosome is removed and TFIIIA-TFIIIC is bound to the promoter region, TFIIIC assists in assembling TFIIIB. In contrast to the 5S nucleosome, DNA-bound TFIIIC enables TFIIIB and Pol III to access the DNA and initiate transcription promptly. We suggest that the DNA is wrapped around the TFIIIA-TFIIIC complex to ensure fast access to the promoter region and protect the DNA from 5S nucleosome reassembly. Therefore, the sharp bending of DNA and the extensive contact between TFIIIA-TFIIIC and DNA in our structure can serve two purposes simultaneously: guiding TBP towards its binding sequence and preventing the reassembly of the 5S nucleosome.

4.5 Functional comparison of TFIIIC and TFIID

TFIIIC in Pol III system and TFIID in Pol II system play similar roles in transcription initiation (Figure 4.4). TFIIIC and TFIID are large protein complexes with multiple lobes. TFIIIC is split into two lobes τA and τB, while TFIID has three lobes: A, B, and C (Figure 4.4, B and E). Both TFIIIC and TFIID extensively bind promoter DNA in the intragenic regions using several binding surfaces (Figure 4.4, B, C and F). Both factors can position TBP near its DNA binding site (Figure 4.4, C and F). This TBP positioning requires large structural rearrangement of τ131 N-terminal TPR in TFIIIC (Figure 4.4, compare B and C). Similarly, Lobe A moves relative to the rest of TFIID to load TBP on DNA (Figure 4.4, compare E and F). However, PIC assembly will likely look different between Pol III and Pol II. Superposition of the Pol III PIC on TFIIIA-TFIIIC-Brf1-TBP showed severe clashes between Pol III and TFIIIC (Figure 4.4D). This suggests that TFIIIC must be displaced from promoter to allow Pol III transcription initiation. TFIID, on the contrary, can accommodate Pol II PIC assembly with minimal rearrangements (Figure 4.4G). This difference between TFIIIC-assisted and TFIID-assisted PIC assembly may reflect functional differences between Pol III and Pol II. While Pol III is recognized for its fast and high-throughput transcription, the initiation of Pol II transcription requires precise control.

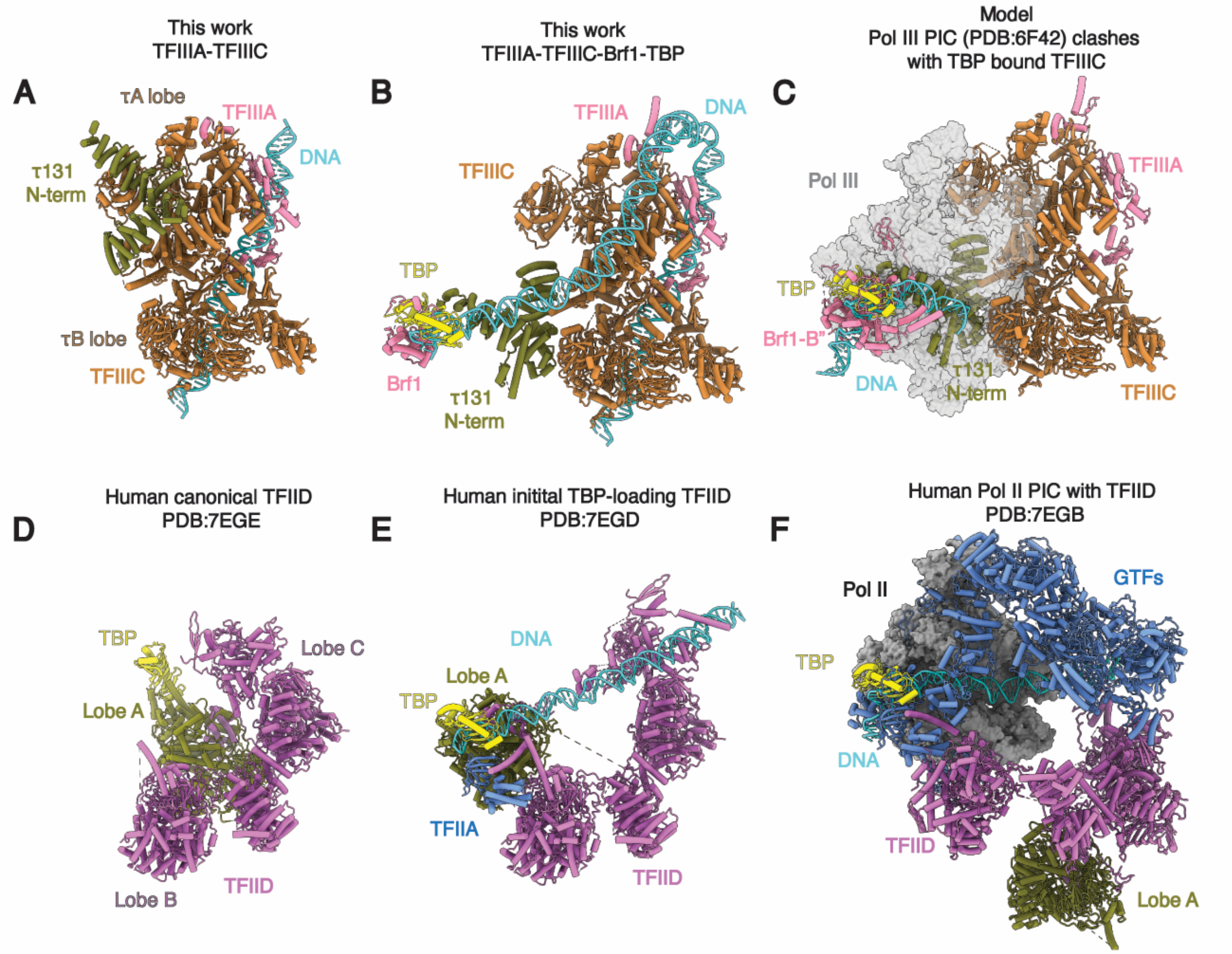

Figure 4.4. Functional comparison of TFIIIC and TFIID. A) Selected 2D class averages from negative stained electron micrographs. TFIIIC lobe τB is indicated with white arrow. Each

experimental condition shows either TFIIIC alone, bound to asparagine tRNA gene (DNA*), or 5S rRNA gene (DNA). (B-G) Comparison of TFIIIC in Pol III system and TFIID in Pol II system and their role in transcription initiation. TFIIIC (B) and TFIID (E) are large protein complexes with multiple lobes (τA and τB in TFIIIC; Lobes A, B, C in TFIID). Both TFIIIC (B, C) and TFIID (F) extensively bind promoter DNA in the intragenic regions using several binding surfaces. Both TFIIIC (C) and TFIID (F) play key role in locating TATA box sequence and positioning TBP near its DNA binding site. This TBP positioning requires large structural rearrangement in TFIIIC (movement of τ131 N-terminal TPR, colored in olive green, compare B and C) and in TFIID (Lobe A, colored in olive green, compare E and F). Superposition of the Pol III PIC on TFIIIA-TFIIIC-Brf1-TBP showed severe clashes between Pol III (grey transparent surface) and TFIIIC (D). TFIID, on the contrary, can accommodate Pol II PIC assembly with minimal rearrangements (G).

4.6 Future Directions

The future studies of Mediator-dependent transcription hold great promise for unraveling the mechanisms underlying transcriptional regulation. Currently available structural information on the Mediator and its interactions with the Pol II PIC and the +1 nucleosome, establishes a solid foundation for further investigations. One aspect that requires deeper exploration is the functional interplay between Mediator and transcription activators. Details of this process remain largely unknown. It is possible that the presence of transcription activators could stabilize specific regions of the Mediator, enabling the visualization of previously elusive parts of Mediator domains. This line of research would shed light on the dynamic nature of transcriptional regulation and provide insights into the molecular mechanisms underlying gene expression control.

In addition to the interactions with the +1 nucleosome, another exciting avenue for future studies lies in the exploration of transcription machinery interactions with chromatin beyond the +1 nucleosome. Understanding the role of broader chromatin landscape in transcription regulation could elucidate the role of Mediator in the coordination of transcriptional events across the genome.

Furthermore, the regulatory effect of the interaction between the Mediator kinase module and the Mediator body remains a crucial question. The precise mechanisms by which the kinase module influences transcriptional activity and modulates Mediator function are yet to be fully elucidated. Investigating the crosstalk between these two modules and deciphering their roles in the regulation of transcriptional initiation, elongation, and termination processes would enhance our understanding of the Mediator functional versatility.

The future directions of Pol III transcription initiation studies hold great potential for unraveling the mechanistic details of this process. While we have made significant progress in understanding certain steps of Pol III PIC assembly and transcription initiation, our current knowledge remains limited. To further our understanding, direct visualization of additional steps in the Pol III transcription initiation process is essential. One crucial area of investigation involves exploring the interactions between DNA, TFIIIC, and full-length TFIIIB during Pol III PIC assembly. We have proposed a model for TFIIIC-assisted TFIIIB loading based on structural information and previous functional and structural studies. However, directly visualizing these interactions would provide invaluable insights into the molecular mechanisms governing Pol III transcription initiation. Understanding the dynamic interplay between these factors and their spatial organization within the PIC would greatly enhance our understanding of the assembly process.

In addition to full-length TFIIIB, visualizing the interactions between Pol III, and DNA-TFIIIA-TFIIIC complex would greatly advance our understanding of the molecular events that drive this process. Obtaining structural information on the Pol III machinery in complex with DNA templates at various stages of transcription initiation would provide unprecedented insights into the molecular mechanism that underlies Pol III transcription.

Furthermore, investigating the interactions between TFIIIC and different types of Pol III promoters is of great interest. In this work we have studied 5S rRNA promoter, that belongs to Type I. However, TFIIIC interactions with differently organized Type II promoters, present in all tRNA genes, remain elusive. TFIIIC plays a pivotal role in recognizing and binding various promoters, accommodating templates of different lengths and containing diverse regulatory elements. Visualizing TFIIIC engagement with promoters of different architectures and understanding the structural basis for its versatility would provide crucial insights into the regulatory mechanisms underlying Pol III transcription initiation.

REFERENCES

1. R. G. Roeder, W. J. Rutter, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224**, 234-237 (1969).
2. A. F. Palazzo, E. S. Lee, Non-coding RNA: what is functional and what is junk? *Front Genet* **6**, 2 (2015).
3. R. D. Kornberg, Eukaryotic transcriptional control. *Trends Cell Biol* **9**, M46-49 (1999).
4. E. P. Geiduschek, G. P. Tocchini-Valentini, Transcription by RNA polymerase III. *Annu Rev Biochem* **57**, 873-914 (1988).
5. P. Cramer *et al.*, Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**, 337-352 (2008).
6. H. Khatter, M. K. Vorlander, C. W. Muller, RNA polymerase I and III: similar yet unique. *Curr Opin Struct Biol* **47**, 88-94 (2017).
7. C. Blattner *et al.*, Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev* **25**, 2093-2105 (2011).
8. Y. Han *et al.*, Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *Elife* **6**, (2017).
9. Y. Sadian *et al.*, Molecular insight into RNA polymerase I promoter recognition and promoter melting. *Nat Commun* **10**, 5543 (2019).
10. B. A. Knutson, M. L. Smith, A. E. Belkevich, A. M. Fakhouri, Molecular Topology of RNA Polymerase I Upstream Activation Factor. *Mol Cell Biol* **40**, (2020).
11. F. Baudin *et al.*, Mechanism of RNA polymerase I selection by transcription factor UAF. *Sci Adv* **8**, eabn5725 (2022).
12. M. C. Thomas, C. M. Chiang, The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**, 105-178 (2006).
13. S. Buratowski, S. Hahn, L. Guarente, P. A. Sharp, Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).
14. G. Dieci, M. C. Bosio, B. Fermi, R. Ferrari, Transcription reinitiation by RNA polymerase III. *Biochim Biophys Acta* **1829**, 331-341 (2013).
15. J. Soutourina, Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* **19**, 262-274 (2018).
16. E. Nogales, R. K. Louder, Y. He, Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **46**, 59-83 (2017).
17. Y. J. Kim, S. Bjorklund, Y. Li, M. H. Sayre, R. D. Kornberg, A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).
18. C. Bernecky, P. Grob, C. C. Ebmeier, E. Nogales, D. J. Taatjes, Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLoS Biol* **9**, e1000603 (2011).
19. L. El Khattabi *et al.*, A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **178**, 1145-1158 e1120 (2019).
20. K. L. Tsai *et al.*, Mediator structure and rearrangements required for holoenzyme formation. *Nature* **544**, 196-201 (2017).

21. S. Schilbach *et al.*, Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209 (2017).
22. P. J. Robinson *et al.*, Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422 e1416 (2016).
23. C. Plaschka *et al.*, Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376-380 (2015).
24. A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, P. Cramer, A structural perspective of CTD function. *Genes Dev* **19**, 1401-1415 (2005).
25. A. Ghosh, C. Lima, Wiley Interdiscip. Rev. *RNA* **1**, 152-172 (2010).
26. Y. He *et al.*, Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365 (2016).
27. D. Nair, Y. Kim, L. C. Myers, Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts. *J Biol Chem* **280**, 33739-33748 (2005).
28. C. Esnault *et al.*, Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol Cell* **31**, 337-346 (2008).
29. K. L. Tsai *et al.*, Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* **157**, 1430-1444 (2014).
30. N. Petrenko, Y. Jin, K. H. Wong, K. Struhl, Mediator Undergoes a Compositional Change during Transcriptional Activation. *Mol Cell* **64**, 443-454 (2016).
31. P. J. Robinson, D. A. Bushnell, M. J. Trnka, A. L. Burlingame, R. D. Kornberg, Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A* **109**, 17931-17935 (2012).
32. C. Plaschka, K. Nozawa, P. Cramer, Mediator Architecture and RNA Polymerase II Interaction. *J Mol Biol* **428**, 2569-2574 (2016).
33. M. A. Cevher *et al.*, Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nat Struct Mol Biol* **21**, 1028-1034 (2014).
34. G. Dieci, G. Fiorino, M. Castelnuovo, M. Teichmann, A. Pagano, The expanding RNA polymerase III transcriptome. *Trends Genet* **23**, 614-622 (2007).
35. A. G. Arimbasseri, R. J. Maraia, RNA Polymerase III Advances: Structural and tRNA Functional Views. *Trends Biochem Sci* **41**, 546-559 (2016).
36. Q. Wang, J. L. Daiss, Y. Xu, C. Engel, Snapshots of RNA polymerase III in action - A mini review. *Gene* **821**, 146282 (2022).
37. D. R. Engelke, S. Y. Ng, B. S. Shastry, R. G. Roeder, Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**, 717-728 (1980).
38. E. Guffanti *et al.*, A minimal promoter for TFIIIC-dependent in vitro transcription of snoRNA and tRNA genes by RNA polymerase III. *J Biol Chem* **281**, 23945-23957 (2006).
39. R. Ferrari, C. Rivetti, J. Acker, G. Dieci, Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polymerase III transcription reinitiation. *Proc Natl Acad Sci U S A* **101**, 13442-13447 (2004).
40. S. Camier, A. M. Dechampesme, A. Sentenac, The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. *Proc Natl Acad Sci U S A* **92**, 9338-9342 (1995).

41. A. M. Ginsberg, B. O. King, R. G. Roeder, Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. *Cell* **39**, 479-489 (1984).
42. J. Miller, A. D. McLachlan, A. Klug, Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**, 1609-1614 (1985).
43. G. A. Kassavetis, B. R. Braun, L. H. Nguyen, E. P. Geiduschek, S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell* **60**, 235-245 (1990).
44. T. E. Cloutier, M. D. Librizzi, A. K. Mollah, M. Brenowitz, I. M. Willis, Kinetic trapping of DNA by transcription factor IIIB. *Proc Natl Acad Sci U S A* **98**, 9581-9586 (2001).
45. A. B. Lassar, P. L. Martin, R. G. Roeder, Transcription of class III genes: formation of preinitiation complexes. *Science* **222**, 740-748 (1983).
46. G. Male *et al.*, Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat Commun* **6**, 7387 (2015).
47. P. Schultz *et al.*, The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. *EMBO J* **8**, 3815-3824 (1989).
48. M. K. Vorlander *et al.*, Structure of the TFIIIC subcomplex tauA provides insights into RNA polymerase III pre-initiation complex formation. *Nat Commun* **11**, 4905 (2020).
49. N. M. Taylor *et al.*, Structural and functional characterization of a phosphatase domain within yeast general transcription factor IIIC. *J Biol Chem* **288**, 15110-15120 (2013).
50. A. Mylona *et al.*, Structure of the tau60/Delta tau91 subcomplex of yeast transcription factor IIIC: insights into preinitiation complex assembly. *Mol Cell* **24**, 221-232 (2006).
51. Y. Han, C. Yan, S. Fishbain, I. Ivanov, Y. He, Structural visualization of RNA polymerase III transcription machineries. *Cell Discov* **4**, 40 (2018).
52. G. Abascal-Palacios, E. P. Ramsay, F. Beuron, E. Morris, A. Vannini, Structural basis of RNA polymerase III transcription initiation. *Nature* **553**, 301-306 (2018).
53. M. K. Vorlander, H. Khatter, R. Wetzel, W. J. H. Hagen, C. W. Muller, Molecular mechanism of promoter opening by RNA polymerase III. *Nature* **553**, 295-300 (2018).
54. M. P. Foster *et al.*, Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat Struct Biol* **4**, 605-608 (1997).
55. B. M. Lee *et al.*, Induced fit and "lock and key" recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J Mol Biol* **357**, 275-291 (2006).
56. R. T. Nolte, R. M. Conlin, S. C. Harrison, R. S. Brown, Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc Natl Acad Sci U S A* **95**, 2938-2943 (1998).
57. R. J. White, RNA polymerase III transcription and cancer. *Oncogene* **23**, 3208-3216 (2004).
58. R. J. White, RNA polymerases I and III, non-coding RNAs and cancer. *Trends Genet* **24**, 622-629 (2008).
59. S. A. Johnson, L. Dubeau, D. L. Johnson, Enhanced RNA polymerase III-dependent transcription is required for oncogenic transformation. *J Biol Chem* **283**, 19184-19191 (2008).
60. H. Goodarzi *et al.*, Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* **165**, 1416-1427 (2016).

61. D. Graczyk, M. Ciesla, M. Boguta, Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. *Biochim Biophys Acta Gene Regul Mech* **1861**, 320-329 (2018).
62. A. G. Winter *et al.*, RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. *Proc Natl Acad Sci U S A* **97**, 12619-12624 (2000).
63. M. Ciesla, E. Skowronek, M. Boguta, Function of TFIIIC, RNA polymerase III initiation factor, in activation and repression of tRNA gene transcription. *Nucleic Acids Res* **46**, 9444-9455 (2018).
64. R. Ferrari *et al.*, TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Mol Cell* **77**, 475-487 e411 (2020).
65. Z. Moqtaderi *et al.*, Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat Struct Mol Biol* **17**, 635-640 (2010).
66. Z. Moqtaderi, K. Struhl, Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. *Mol Cell Biol* **24**, 4118-4127 (2004).
67. J. Hanske, Y. Sadian, C. W. Muller, The cryo-EM resolution revolution and transcription complexes. *Curr Opin Struct Biol* **52**, 8-15 (2018).
68. W. Kuhlbrandt, Biochemistry. The resolution revolution. *Science* **343**, 1443-1444 (2014).
69. Y. He, J. Fang, D. J. Taatjes, E. Nogales, Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-486 (2013).
70. D. J. Taatjes, R. Tjian, Structure and function of CRSP/Med2; a promoter-selective transcriptional coactivator complex. *Mol Cell* **14**, 675-683 (2004).
71. C. Suloway *et al.*, Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60 (2005).
72. 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).
73. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
74. T. Ogura, K. Iwasaki, C. Sato, Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *J Struct Biol* **143**, 185-200 (2003).
75. G. C. Lander *et al.*, Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* **166**, 95-102 (2009).
76. 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. *J Struct Biol* **166**, 205-213 (2009).
77. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
78. G. Tang *et al.*, EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46 (2007).
79. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).

80. R. Henderson *et al.*, Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205-214 (2012).
81. R. Sanchez-Garcia *et al.*, DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874 (2021).
82. 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).
83. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).
84. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
85. K. Miwa *et al.*, Crystal Structure of Human General Transcription Factor TFIIE at Atomic Resolution. *J Mol Biol* **428**, 4258-4266 (2016).
86. 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).
87. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
88. L. Zimmermann *et al.*, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
89. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394 (2015).
90. B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* **86**, 2 9 1-2 9 37 (2016).
91. L. Lariviere *et al.*, Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol* **13**, 895-901 (2006).
92. D. Monte *et al.*, Crystal structure of human Mediator subunit MED23. *Nat Commun* **9**, 3389 (2018).
93. 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).
94. 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).
95. B. J. Greber *et al.*, The cryoelectron microscopy structure of the human CDK-activating kinase. *Proc Natl Acad Sci U S A* **117**, 22849-22857 (2020).
96. N. R. Brown, M. E. Noble, J. A. Endicott, L. N. Johnson, The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat Cell Biol* **1**, 438-443 (1999).
97. B. J. Greber, D. B. Toso, J. Fang, E. Nogales, The complete structure of the human TFIIH core complex. *Elife* **8**, (2019).
98. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, (2018).

99. R. K. Louder *et al.*, Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604-609 (2016).
100. H. Takahashi *et al.*, Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell* **146**, 92-104 (2011).
101. C. Jeronimo, F. Robert, The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. *Trends Cell Biol* **27**, 765-783 (2017).
102. A. Y. Belorusova *et al.*, Molecular determinants of MED1 interaction with the DNA bound VDR-RXR heterodimer. *Nucleic Acids Res* **48**, 11199-11213 (2020).
103. A. B. Hittelman, D. Burakov, J. A. Iniguez-Lluhi, L. P. Freedman, M. J. Garabedian, Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. *EMBO J* **18**, 5380-5388 (1999).
104. S. Malik, A. E. Wallberg, Y. K. Kang, R. G. Roeder, TRAP/SMCC/mediator-dependent transcriptional activation from DNA and chromatin templates by orphan nuclear receptor hepatocyte nuclear factor 4. *Mol Cell Biol* **22**, 5626-5637 (2002).
105. L. Grontved, M. S. Madsen, M. Boergesen, R. G. Roeder, S. Mandrup, MED14 tethers mediator to the N-terminal domain of peroxisome proliferator-activated receptor gamma and is required for full transcriptional activity and adipogenesis. *Mol Cell Biol* **30**, 2155-2169 (2010).
106. E. Vojnic *et al.*, Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol* **18**, 404-409 (2011).
107. A. G. Milbradt *et al.*, Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol* **18**, 410-415 (2011).
108. K. D. Meyer, S. C. Lin, C. Bernecky, Y. Gao, D. J. Taatjes, p53 activates transcription by directing structural shifts in Mediator. *Nat Struct Mol Biol* **17**, 753-760 (2010).
109. D. J. Taatjes, A. M. Naar, F. Andel, 3rd, E. Nogales, R. Tjian, Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* **295**, 1058-1062 (2002).
110. C. Bernecky, D. J. Taatjes, Activator-mediator binding stabilizes RNA polymerase II orientation within the human mediator-RNA polymerase II-TFIIF assembly. *J Mol Biol* **417**, 387-394 (2012).
111. H. Zhao *et al.*, Structure of mammalian Mediator complex reveals Tail module architecture and interaction with a conserved core. *Nat Commun* **12**, 1355 (2021).
112. H. Zhang *et al.*, Mediator structure and conformation change. *Molecular Cell* **81**, 1781-1788. e1784 (2021).
113. O. Jasnovidova, R. Stefl, Wiley Interdiscip. Rev. *RNA* **4**, (2013).
114. G. A. Kassavetis, E. Soragni, R. Driscoll, E. P. Geiduschek, Reconfiguring the connectivity of a multiprotein complex: fusions of yeast TATA-binding protein with Brf1, and the function of transcription factor IIIB. *Proc Natl Acad Sci U S A* **102**, 15406-15411 (2005).
115. J. Zivanov, T. Nakane, S. H. W. Scheres, Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253-267 (2020).
116. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

117. T. I. Croll, ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530 (2018).
118. H. Ashkenazy *et al.*, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350 (2016).
119. C. Banerjee, B. W. Liauw, R. Vafabakhsh, Visualizing the Conformational Dynamics of Membrane Receptors Using Single-Molecule FRET. *J Vis Exp*, (2022).
120. J. W. van de Meent, J. E. Bronson, C. H. Wiggins, R. L. Gonzalez, Jr., Empirical Bayes methods enable advanced population-level analyses of single-molecule FRET experiments. *Biophys J* **106**, 1327-1337 (2014).
121. S. Rozenfeld, P. Thuriaux, Genetic interactions within TFIIIC, the promoter-binding factor of yeast RNA polymerase III. *Mol Genet Genomics* **265**, 705-710 (2001).
122. V. Nagarajavel, J. R. Iben, B. H. Howard, R. J. Maraia, D. J. Clark, Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex in vivo. *Nucleic Acids Res* **41**, 8135-8143 (2013).
123. R. Arrebola *et al.*, Tau91, an essential subunit of yeast transcription factor IIIC, cooperates with tau138 in DNA binding. *Mol Cell Biol* **18**, 1-9 (1998).
124. H. Dumay-Odelot, J. Acker, R. Arrebola, A. Sentenac, C. Marck, Multiple roles of the tau131 subunit of yeast transcription factor IIIC (TFIIIC) in TFIIIB assembly. *Mol Cell Biol* **22**, 298-308 (2002).
125. Y. J. Hsieh, Z. Wang, R. Kovelman, R. G. Roeder, Cloning and characterization of two evolutionarily conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with TFIIIB and RNA polymerase III. *Mol Cell Biol* **19**, 4944-4952 (1999).
126. T. D. Petes, L. M. Hereford, K. G. Skryabin, Characterization of two types of yeast ribosomal DNA genes. *J Bacteriol* **134**, 295-305 (1978).
127. B. R. Braun, D. L. Riggs, G. A. Kassavetis, E. P. Geiduschek, Multiple states of protein-DNA interaction in the assembly of transcription complexes on Saccharomyces cerevisiae 5S ribosomal RNA genes. *Proc Natl Acad Sci U S A* **86**, 2530-2534 (1989).
128. T. Leveillard, G. A. Kassavetis, E. P. Geiduschek, Saccharomyces cerevisiae transcription factors IIIB and IIIC bend the DNA of a tRNA(Gln) gene. *J Biol Chem* **266**, 5162-5168 (1991).
129. M. L. Brown, G. P. Schroth, J. M. Gottesfeld, D. P. Bazett-Jones, Protein and DNA requirements for the transcription factor IIIA-induced distortion of the 5 S rRNA gene promoter. *J Mol Biol* **262**, 600-614 (1996).
130. B. R. Braun, G. A. Kassavetis, E. P. Geiduschek, Bending of the Saccharomyces cerevisiae 5S rRNA gene in transcription factor complexes. *J Biol Chem* **267**, 22562-22569 (1992).
131. D. P. Bazett-Jones, M. L. Brown, Electron microscopy reveals that transcription factor TFIIIA bends 5S DNA. *Mol Cell Biol* **9**, 336-341 (1989).
132. R. Sreenivasan *et al.*, Fluorescence-Detected Conformational Changes in Duplex DNA in Open Complex Formation by Escherichia coli RNA Polymerase: Upstream Wrapping and Downstream Bending Precede Clamp Opening and Insertion of the Downstream Duplex. *Biochemistry* **59**, 1565-1581 (2020).

133. B. R. Braun, B. Bartholomew, G. A. Kassavetis, E. P. Geiduschek, Topography of transcription factor complexes on the Saccharomyces cerevisiae 5 S RNA gene. *J Mol Biol* **228**, 1063-1077 (1992).
134. B. Bartholomew, G. A. Kassavetis, E. P. Geiduschek, Two components of Saccharomyces cerevisiae transcription factor IIIB (TFIIIB) are stereospecifically located upstream of a tRNA gene and interact with the second-largest subunit of TFIIIC. *Mol Cell Biol* **11**, 5181-5189 (1991).
135. Y. Huang, M. Hamada, R. J. Maraia, Isolation and cloning of four subunits of a fission yeast TFIIIC complex that includes an ortholog of the human regulatory protein TFIIICbeta. *J Biol Chem* **275**, 31480-31487 (2000).
136. O. Rowland, J. Segall, Interaction of wild-type and truncated forms of transcription factor IIIA from Saccharomyces cerevisiae with the 5 S RNA gene. *J Biol Chem* **271**, 12103-12110 (1996).
137. C. A. Milne, J. Segall, Mapping regions of yeast transcription factor IIIA required for DNA binding, interaction with transcription factor IIIC, and transcription activity. *J Biol Chem* **268**, 11364-11371 (1993).
138. S. Del Rio, D. R. Setzer, The role of zinc fingers in transcriptional activation by transcription factor IIIA. *Proc Natl Acad Sci U S A* **90**, 168-172 (1993).
139. D. S. Wuttke, M. P. Foster, D. A. Case, J. M. Gottesfeld, P. E. Wright, Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol* **273**, 183-206 (1997).
140. J. Hayes, T. D. Tullius, A. P. Wolffe, A protein-protein interaction is essential for stable complex formation on a 5 S RNA gene. *J Biol Chem* **264**, 6009-6012 (1989).
141. K. Rothfels, O. Rowland, J. Segall, Zinc fingers 1 and 7 of yeast TFIIIA are essential for assembly of a functional transcription complex on the 5 S RNA gene. *Nucleic Acids Res* **35**, 4869-4881 (2007).
142. O. Rowland, J. Segall, A hydrophobic segment within the 81-amino-acid domain of TFIIIA from Saccharomyces cerevisiae is essential for its transcription factor activity. *Mol Cell Biol* **18**, 420-432 (1998).
143. H. Suh *et al.*, Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol Cell* **61**, 297-304 (2016).
144. R. Schuller *et al.*, Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol Cell* **61**, 305-314 (2016).
145. T. M. Sogaard, J. Q. Svejstrup, Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J Biol Chem* **282**, 14113-14120 (2007).
146. X. Chen *et al.*, Structures of the human Mediator and Mediator-bound preinitiation complex. *Science* **372**, (2021).
147. S. Rengachari, S. Schilbach, S. Aibara, C. Dienemann, P. Cramer, Structure of the human Mediator-RNA polymerase II pre-initiation complex. *Nature* **594**, 129-133 (2021).
148. W. K. Cho *et al.*, Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415 (2018).
149. A. Boija *et al.*, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816 (2018).

150. A. V. Zamudio *et al.*, Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol Cell* **76**, 753-766 e756 (2019).

151. S. Schilbach, H. Wang, C. Dienemann, P. Cramer, Yeast PIC-Mediator structure with RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A* **120**, e2220542120 (2023).

152. X. Chen *et al.*, Structures of +1 nucleosome-bound PIC-Mediator complex. *Science* **378**, 62-68 (2022).

153. W. Seifert-Davila *et al.*, Structural insights into human TFIIIC promoter recognition. *bioRxiv*, 2023.2005.2016.540835 (2023).

154. J. J. Bieker, P. L. Martin, R. G. Roeder, Formation of a rate-limiting intermediate in 5S RNA gene transcription. *Cell* **40**, 119-127 (1985).

155. M. F. Carey, S. P. Gerrard, N. R. Cozzarelli, Analysis of RNA polymerase III transcription complexes by gel filtration. *J Biol Chem* **261**, 4309-4317 (1986).

156. K. R. Clemens, X. Liao, V. Wolf, P. E. Wright, J. M. Gottesfeld, Definition of the binding sites of individual zinc fingers in the transcription factor IIIA-5S RNA gene complex. *Proc Natl Acad Sci U S A* **89**, 10822-10826 (1992).

157. J. J. Hayes, K. R. Clemens, Locations of contacts between individual zinc fingers of Xenopus laevis transcription factor IIIA and the internal control region of a 5S RNA gene. *Biochemistry* **31**, 11600-11605 (1992).

158. C. A. Joazeiro, G. A. Kassavetis, E. P. Geiduschek, Alternative outcomes in assembly of promoter complexes: the roles of TBP and a flexible linker in placing TFIIIB on tRNA genes. *Genes Dev* **10**, 725-739 (1996).

159. G. A. Kassavetis *et al.*, The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell* **71**, 1055-1064 (1992).

160. E. P. Geiduschek, G. A. Kassavetis, The RNA polymerase III transcription apparatus. *J Mol Biol* **310**, 1-26 (2001).

161. N. Chaussivert, C. Conesa, S. Shaaban, A. Sentenac, Complex interactions between yeast TFIIIB and TFIIIC. *J Biol Chem* **270**, 15353-15358 (1995).

162. R. D. Moir, K. V. Puglia, I. M. Willis, Interactions between the tetratricopeptide repeat-containing transcription factor TFIIIC131 and its ligand, TFIIIB70. Evidence for a conformational change in the complex. *J Biol Chem* **275**, 26591-26598 (2000).

163. Y. Liao, R. D. Moir, I. M. Willis, Interactions of Brf1 peptides with the tetratricopeptide repeat-containing subunit of TFIIIC inhibit and promote preinitiation complex assembly. *Mol Cell Biol* **26**, 5946-5956 (2006).

164. G. A. Kassavetis, D. L. Riggs, R. Negri, L. H. Nguyen, E. P. Geiduschek, Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Mol Cell Biol* **9**, 2551-2566 (1989).

165. D. Rhodes, Structural analysis of a triple complex between the histone octamer, a Xenopus gene for 5S RNA and transcription factor IIIA. *EMBO J* **4**, 3473-3482 (1985).

166. J. M. Gottesfeld, DNA sequence-directed nucleosome reconstitution on 5S RNA genes of Xenopus laevis. *Mol Cell Biol* **7**, 1612-1622 (1987).