EUKARYOTIC TRANSCRIPTION

Everything at once: cryo-EM yields remarkable insights into human RNA polymerase II transcription

After years of only low-resolution and partial assemblies, the entire human preinitiation complex (PIC), including the large and flexible Mediator and TFIID complexes, has come into focus. Five recent papers from three different research groups have transformed our understanding of transcription initiation by RNA polymerase II.

Allison C. Schier and Dylan J. Taatjes

Since the discovery of RNA polymerase II (Pol II)[1], structural data lagged behind the biochemical and cellular studies that established fundamental aspects of Pol II function and its regulation by associated factors. Among the myriad Pol II–associated factors, a core set is required to direct it to transcription start sites on genomic DNA and to control Pol II activity once it is bound to these sites. This core set of factors is known as the preinitiation complex (PIC), comprising TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Pol II and Mediator[2]. The PIC is approximately 4 MDa in size and its Mediator, TFIID and TFIIH components are both large and structurally dynamic. This presents many challenges for structural biologists, but these challenges have been overcome by the He[3], Cramer[4,5] and Xu[6,7] groups, who report high-resolution structures of the complete human PIC for the first time.

While important prior work established PIC structural models in yeast[8–10], Pol II transcription is more complex in humans, as human genomes possess diverse combinations of core-promoter-sequence motifs, and Mediator and TFIID contain subunits and domains not present in yeast. The recent structural elucidation of human PICs reveals that yeast PIC structures served as a good basic model, but structural differences were evident in each study. The most extensive structural changes between yeast and human Mediator-containing PICs were characterized by Chen and coworkers[7]. This may result from their use of PICs containing TFIID instead of TATA-binding protein (TBP; as in Abdella et al.[3] and Rengachari et al.[5]) and their analysis of a conformationally distinct Mediator complex.

Because the structure of the human Mediator complex remained poorly defined, each of the three new cryo-EM studies of Mediator-containing human PICs yield new high-resolution data for Mediator that are in agreement with a recent structure of the mouse Mediator complex[11], with a notable exception (next paragraph). As expected from yeast and mouse Mediator structural data, human MED14 and MED17 are major structural scaffolds. Data from the Xu lab[7], however, revealed that the metazoan-specific MED26 subunit provides another potential structural hub within human Mediator because, on the basis of crosslinking-mass spectrometry data, it interacts with up to nine other Mediator subunits.

The Xu lab[7] also reports a Mediator conformational state that is markedly distinct from that in the structures reported by Abdella et al.[3] and Rengachari et al.[5]. This alternative structure, called MED^{B} because of a bent conformation for the tail region of Mediator, showed large-scale reorganization of the tail, with rotation and movements of 25–55 Å among domains within the MED16, MED23 and MED24 subunits. Furthermore, portions of the tail subunits MED25 and MED16 were disordered in the MED^{B} conformational state such that they could not be resolved by cryo-EM. By contrast, MED16 and a portion of MED25 were structured in the MED^{E} structural state (E = extended tail domain). Structural data from the He lab[3] represent the MED^{E} structural state and are consistent with data from Rengachari et al.[5], although the latter study did not include the distal tail subunits MED15, MED16 or MED23–25.

The Xu lab separately determined the complete MED^{E} structure and linked the MED^{B} structural state to the expression of isoform 1 of MED16. MED^{E} was found to contain MED16 isoform 2. The structural changes between MED^{B} and MED^{E} center around the MED16 subunit, but the differences between MED16 isoform 1 and MED16 isoform 2 are minor. MED16 is an 877-residue protein (isoform 1), and isoform 2 (841 residues) is identical except for about 40 residues at the C terminus. Further verification is needed but, if the structural differences between MED^{B} and MED^{E} derive entirely from the distinct MED16 isoforms, this would represent an intriguing example

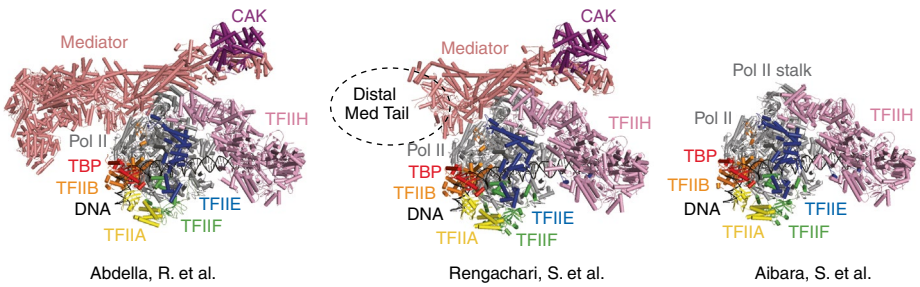

Fig. 1 | Comparison of recent PIC and Mediator–PIC structures. The left structure was reported by Abdella et al.[3] (PDB 7LBM), the middle by Rengachari et al.[5] (PDB 7NVR) and the right by Aibara et al.[4] (PDB 7NVY). Labels are shaded according to the color of the factor. The dotted oval represents the missing Mediator tail subunits in the Rengachari et al.[5] study. The most recently published TFIID-containing Mediator–PIC structure from Chen et al.[7] was not yet available for comparison here.

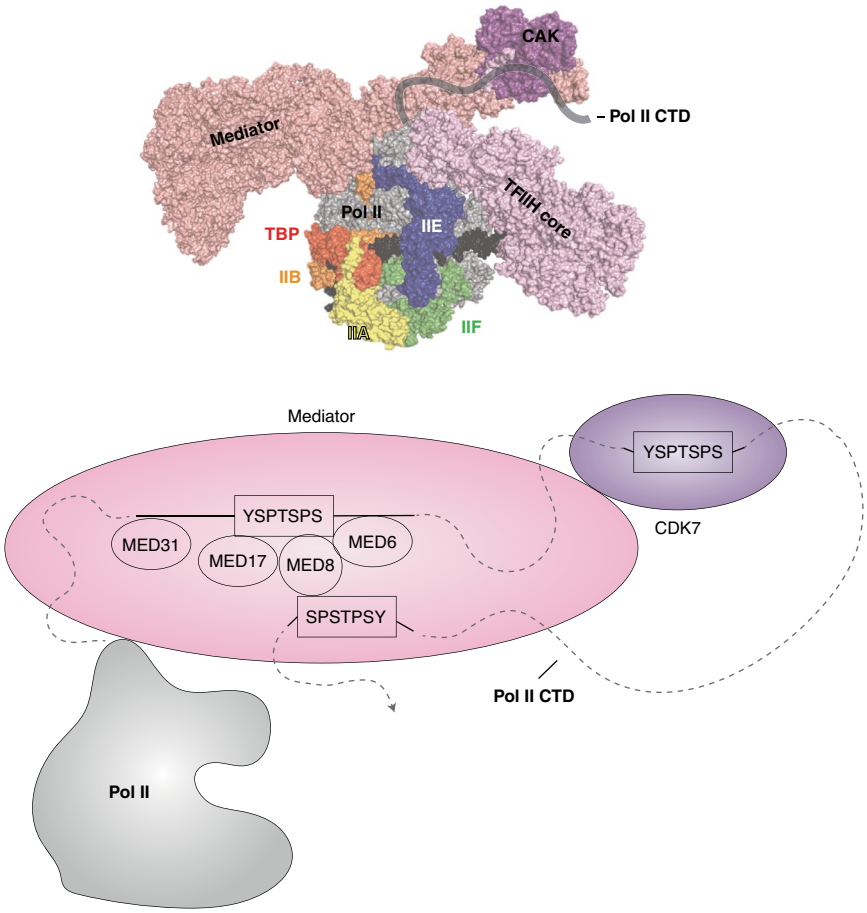

Fig. 2 | Model of the Pol II CTD path to Mediator and CDK7. The top panel shows an overview of the Mediator–PIC, with an approximate path of the Pol II CTD drawn in. The orientation and color scheme are identical to those of Abdella et al.[3] (PDB 7LBM) shown in Fig. 1. The bottom panel shows a schematic of the Pol II CTD peptides mapped onto the Mediator–PIC structure, based upon data from Abdella et al.[3] and Chen et al.[7]. It is not clear which of the 52 Pol II CTD heptad repeats (YSPTSPS consensus) bind Mediator and CDK7 in the PIC complexes. The Mediator subunits that interact with the CTD are also indicated. Structured and modeled portions of the CTD are represented by solid lines, and unstructured portions are represented by dashed lines.

of an alternatively spliced isoform that impacts the structure, and presumably the function, of Mediator.

Cryo-EM structures of Mediator-containing PICs published by Abdella et al.[3], Chen et al.[7] and Rengachari et al.[5] are generally consistent, despite the different purification protocols and PIC assembly methods used for sample preparation. Some differences that were apparent probably reflect differences in PIC composition (see paragraph beginning "Perhaps the biggest gap..."). Each lab reported a similar location and orientation of Mediator relative to Pol II and the rest of the PIC (Fig. 1). Mediator interactions with the TFIIH-associated CDK7 kinase module (called the CAK, for CDK-activating kinase) were shown to involve the MED6 subunit and the hook domain, with each binding on opposite sides of the CAK. Each study showed similar, but not identical, structural interfaces between Mediator and the Pol II stalk. Mediator–Pol II interactions were also observed with the Pol II dock domain within RPB1, the Pol II RPB3/11 subunits, RPB8 (ref. [7]), and with the C-terminal domain (CTD) of the largest Pol II subunit, RPB1 (refs. [3,7]). Mediator interactions with the B-ribbon of TFIIB, which connects directly to the Pol II active site, were also observed[5].

The Pol II CTD is intrinsically disordered and consists of 52 heptad repeats of the general consensus sequence YSPTSPS. Because of its disordered nature, structural data for the CTD is sparse. The He/Tjian and Xu labs were each able to resolve and model two CTD fragments into their Mediator–PIC cryo-EM structures, and the He/Tjian team was able to determine the directionality (that is, N terminus to C terminus) of the repetitive CTD sequence. As shown in Fig. 2, the Pol II CTD binds Mediator, and CTD heptad repeats C-terminal to the Mediator binding site are then stably positioned to bind the CDK7 active site for phosphorylation. The Mediator-dependent juxtaposition of the Pol II CTD and the CAK domain of TFIIH provides a straightforward mechanism by which Mediator promotes Pol II CTD phosphorylation by CDK7. Separately, the He/Tjian group[3] showed that the CDK7 kinase was in an active conformation in the PIC. CTD phosphorylation is fundamentally important to regulate RNA processing (for example, capping, splicing and polyadenylation), which ensures that Pol II mRNA transcripts are stable and can be translated into proteins. Mediator is known to bind the unphosphorylated CTD, whereas CTD phosphorylation blocks Mediator binding[12]. TFIIH undergoes structural changes upon binding the PIC (Fig. 3a), including a Mediator-dependent stabilization of the CAK.

Perhaps the biggest gap in our understanding of PIC structure involves the 1.3 MDa TFIID complex. TFIID initiates PIC assembly by recognizing promoter DNA elements downstream of the transcription start site via its TAF1, TAF2 and TAF7 subunits, followed by structural reorganization to deposit TBP at a DNA sequence upstream of the transcription start site[6,13]. The two recent papers from Chen and coworkers[6,7] provide key insights into how TFIID orchestrates Pol II transcription initiation. Here, we focus on the Mediator-bound PIC work, which includes TFIID[7]. TFIID was shown to bind TFIIH through two different sites. First, the TFIID subunit TAF1 contacts XPB and promotes its interaction with the Pol II jaw (RPB5 subunit); second, the TFIID subunit TAF2 binds the p8 and p52 subunits of TFIIH. These interactions appear to direct the p8 and p52 subunits to move in a concerted fashion, together with XPB, and these movements are coordinated with a structural shift in the Pol II stalk that coincides with structural changes in the E-ribbon domain of TFIIE and the Mediator subunit MED8 around the Pol II stalk[7]. Collectively, this TFIID-induced reorganization alters Mediator orientation with respect to Pol II and establishes a new contact between the Mediator hook domain and XPB. As a result, XPB becomes an interaction hub, simultaneously contacting downstream promoter DNA, Pol II, TFIID and Mediator. This XPB contact network probably sets up subsequent structural transitions required

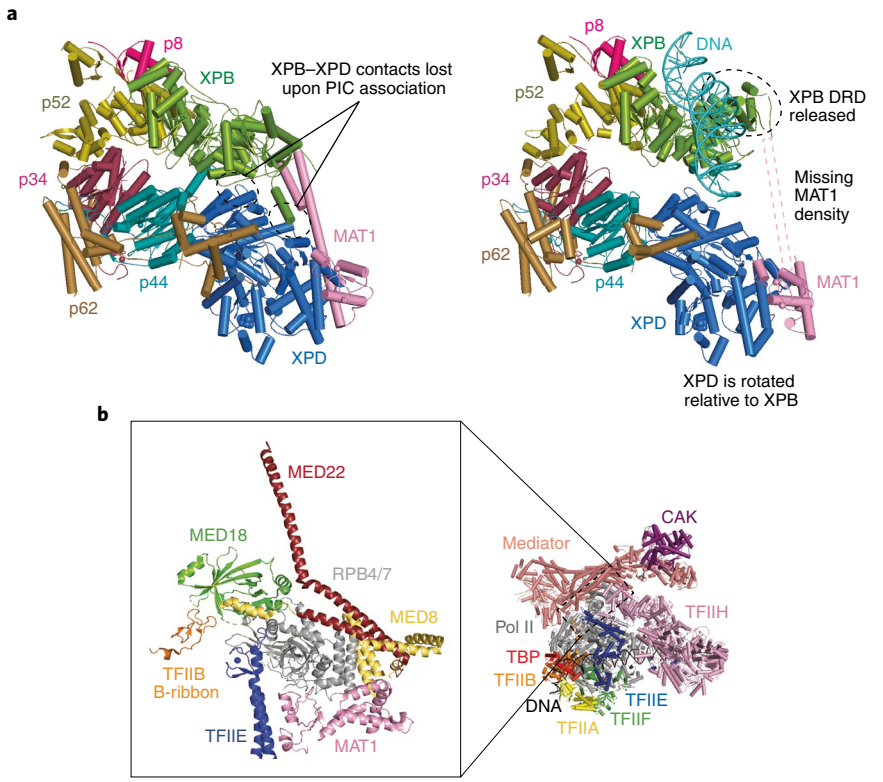

Fig. 3 | Selected structural details important for Pol II transcription initiation. a, Comparison of free core TFIIH[25] (left, PDB 6NMI) and PIC-bound TFIIH[4] (right, PDB 7NVY) highlights structural transitions that occur upon DNA binding. DRD, damage-recognition domain. **b**, A zoomed-in view of structural interactions centered on the Pol II stalk (RPB4/7; PDB 7NVR), which undergoes structural changes upon promoter opening. Labels are colored according to the colors for each domain.

for transcription initiation and Pol II promoter escape.

In an article[4] that accompanies their TBP-based Mediator-bound PIC structure[5], the Cramer group determined high-resolution PIC structures with both closed and open promoter DNA. Although these PIC complexes lacked TFIID and Mediator, the data reveal structural changes that are probably essential for remodeling PIC to allow Pol II to "escape" the promoter and elongate a transcript. The open promoter complex was generated simply by adding the ATP analogue ADP·BeF_3, which bound the TFIIH XPB subunit that possesses ATPase and DNA translocase activity[14]. Unlike ATP, ADP·BeF_3 cannot be hydrolyzed by XPB, yet its binding was sufficient to partially open the DNA template near the transcription start site[4]. This suggests that a single ATP hydrolysis/translocation step is sufficient to open human Pol II promoters, in agreement with recent biochemical results[15]. Interestingly, structural data for the open complex revealed that the Pol II clamp closes, and this coincides with a shift in the position of the Pol II stalk and disruption of the interaction of MAT1 with the stalk, releasing the MAT1 contact.

Many interesting new hypotheses have been raised by this outstanding set of papers[3–7]. Below, we briefly touch upon a few that pull together some common themes among the studies and speculate about potential implications for Pol II transcription. Abdella et al.[3] and Chen et al.[7] each propose a CTD-tracking or "gating" mechanism by which Mediator can bind and release the unphosphorylated CTD to enable iterative CDK7-catalyzed CTD phosphorylation. This mechanism is supported by the observation of two different CTD-bound structural states in the TBP- versus TFIID-containing PICs. With TFIID-containing PICs, the Mediator head and middle modules undergo a more substantial structural shift that forms a "sandwich" around the CTD, whereas these structural transitions are incomplete in the TBP-based PICs. The distinct CTD-bound structures may represent functional intermediates that provide a means to thread the CTD through the Mediator "gateway" to CDK7. Two different Mediator-bound segments (that is, non-sequential; Fig. 2) of the 52-heptad-repeat CTD sequence were resolved by Chen et al.[7], and they reasonably speculate that threading may involve transient release of one or both segments. Alternatively, the binding of multiple CTD heptad repeats represented by the "sandwiched" Mediator structural state may simply act to localize and properly orient the large, disordered CTD for efficient phosphorylation by CDK7. The Mediator–CTD binding affinity (K_D, measured with yeast Mediator) is about 1 nM, with off rates of approximately 8×10^{-5} s^{-1} (ref. [8]), which corresponds to a residence time of 3+ hours. Consistent with these results, Chen et al.[7] showed that Mediator engulfs the CTD to form the sandwiched interaction through structural changes centered around the hook and knob region. This effectively buries at least four CTD repeats in a hydrophobic environment, which is typical for high-affinity, stable protein–protein interactions.

Regardless of the precise mechanism(s) by which Mediator promotes Pol II CTD phosphorylation, it is apparent that Mediator releases the CTD at some point during transcription initiation. How may this happen? Some possible clues are provided by the open-promoter PIC structure from Aibara et al.[4], in which the Pol II clamp and stalk undergo conformational changes. The Pol II stalk is a tightly packed interaction hub in the Mediator-containing PIC (Fig. 3b), and structural shifts at the stalk could trigger conformational changes that promote opening of the "CTD sandwich" formed by the Mediator hook and knob domains. Coincident with this, structural shifts in XPB are required to translocate DNA to form the open complex[4]. Data from Chen et al.[7] show that XPB is another interaction hub in the PIC, and conformational changes in XPB could affect not only the DNA structure but also impact the structures of TFIID and Mediator as well.

An important yet challenging future direction is to assess whether distinct PIC structural intermediates are formed during activated transcription; that is, transcription controlled by sequence-specific, DNA-binding transcription factors (TFs). TFs were not explicitly included or resolved in the present PIC structures, but this may reflect structural disorder. For the PIC factors, potential distinct TF-dependent activation mechanisms may be especially sensitive to Mediator because it is stably bound by many different TFs, and TFIID may be less frequently targeted[16]. Mediator also appears to alter its structural state upon TF binding[17], but this remains

controversial because high-resolution data are generally lacking[18].

In cells, accumulating evidence supports TF-dependent "transcriptional bursting," in which multiple Pol II complexes initiate multiple rounds of transcription from the same promoter[19]. This mechanism requires rapid Pol II promoter escape, promoter-proximal pause release, and reinitiation[20,21], which implies that Mediator–CTD interactions will be transient and XPB-dependent promoter opening will be accelerated. Structural data from Rengachari et al.[5] revealed an interaction between the MED18 Mediator subunit and the B-ribbon of TFIIB (Fig. 3b). Adjacent motifs in TFIIB reside in the Pol II active site and help to stabilize the open complex via interactions with single-stranded DNA. Prior work suggested that transcriptional bursting was Mediator dependent[22]. We speculate that one way for Mediator to promote rapid Pol II reinitiation would be to act through TFIIB to maintain an open, pre-melted DNA template at the transcription start site.

Finally, we note that substantial portions of the TFIID- and Mediator-containing PICs remain unresolved due to conformational flexibility or intrinsic disorder; such unstructured regions may become ordered upon binding other proteins (for example, TFs) or may help to regulate the partitioning of PIC into molecular condensates[23]. Crosslinking mass spectrometry data from Chen et al.[7] suggest that disordered regions (that is, not resolved in these studies) mediate additional protein–protein contacts within Mediator. For example, MED26 formed crosslinks with nine different subunits, two of which (MED6 and MED8) are separated from the structured C-terminal MED26 residues; similarly, MED1 formed crosslinks with MED29, which is distant from the structured N-terminal MED1 domain. These data probably reflect stable structural states, given that mass spectrometry detection requires perhaps 100 million crosslinking events between the same two residues, which is not likely with random sampling. As noted by Abdella et al.[3] and Chen et al.[7], many sites bound by TFs remain disordered in the PIC structures (for example, MED1, MED15 and MED25), and unresolved regions on MED26 interact with Pol II elongation factors[24].

These "everything-at-once" structural insights have answered many long-standing questions in the transcription field. Although important new questions will continue to drive the field forward, future research will benefit from this more solid foundation and will be based upon a far more detailed and accurate understanding of Pol II transcription initiation. ❐

Allison C. Schier and Dylan J. Taatjes ✉
Dept. of Biochemistry, University of Colorado, Boulder, CO, USA.
✉*e-mail: taatjes@colorado.edu*

Published online: 15 June 2021
https://doi.org/10.1038/s41594-021-00613-6

References

1. Roeder, R. G. *Nat. Struct. Mol. Biol.* **26**, 783–791 (2019).
2. Schier, A. C. & Taatjes, D. J. *Genes Dev.* **34**, 465–488 (2020).
3. Abdella, R. et al. *Science* **372**, 52–56 (2021).
4. Aibara, S., Schilbach, S. & Cramer, P. *Nature* https://doi.org/10.1038/s41586-021-03554-8 (2021).
5. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C. & Cramer, P. *Nature* https://doi.org/10.1038/s41586-021-03555-7 (2021).
6. Chen, X. et al. *Science* **372**, eaba8490 (2021).
7. Chen, X. et al. *Science* https://doi.org/10.1126/science.abg0635 (2021).
8. Robinson, P. J. et al. *Cell* **166**, 1411–1422.e16 (2016).
9. Schilbach, S. et al. *Nature* **551**, 204–209 (2017).
10. Tsai, K.-L. et al. *Nature* **544**, 196–201 (2017).
11. Zhao, H. et al. *Nat. Commun.* **12**, 1355 (2021).
12. Søgaard, T. M. M. & Svejstrup, J. Q. *J. Biol. Chem.* **282**, 14113–14120 (2007).
13. Patel, A. B. et al. *Science* **362**, eaau8872 (2018).
14. Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. *Proc. Natl Acad. Sci. USA* **112**, 3961–3966 (2015).
15. Tomko, E. J. et al. *J. Mol. Biol.* https://doi.org/10.1016/j.jmb.2021.166813 (2021).
16. Rossi, M. J. et al. *Nature* **592**, 309–314 (2021).
17. Taatjes, D. J., Näär, A. M., Andel, F. III, Nogales, E. & Tjian, R. *Science* **295**, 1058–1062 (2002).
18. Henley, M. J. et al. *Proc. Natl Acad. Sci. USA* **117**, 27346–27353 (2020).
19. Rodriguez, J. & Larson, D. R. *Annu. Rev. Biochem.* **89**, 189–212 (2020).
20. Gressel, S. et al. *Elife* **6**, e29736 (2017).
21. Shao, W. & Zeitlinger, J. *Nat. Genet.* **49**, 1045–1051 (2017).
22. Tantale, K. et al. *Nat. Commun.* **7**, 12248 (2016).
23. Sabari, B. R., Dall'Agnese, A. & Young, R. A. *Trends Biochem. Sci.* **45**, 961–977 (2020).
24. Takahashi, H. et al. *Cell* **146**, 92–104 (2011).
25. Greber, B. J., Toso, D. B., Fang, J. & Nogales, E. *Elife* **8**, e44771 (2019).

Competing interests

D.J.T. is a member of the scientific advisory board at Dewpoint Therapeutics.