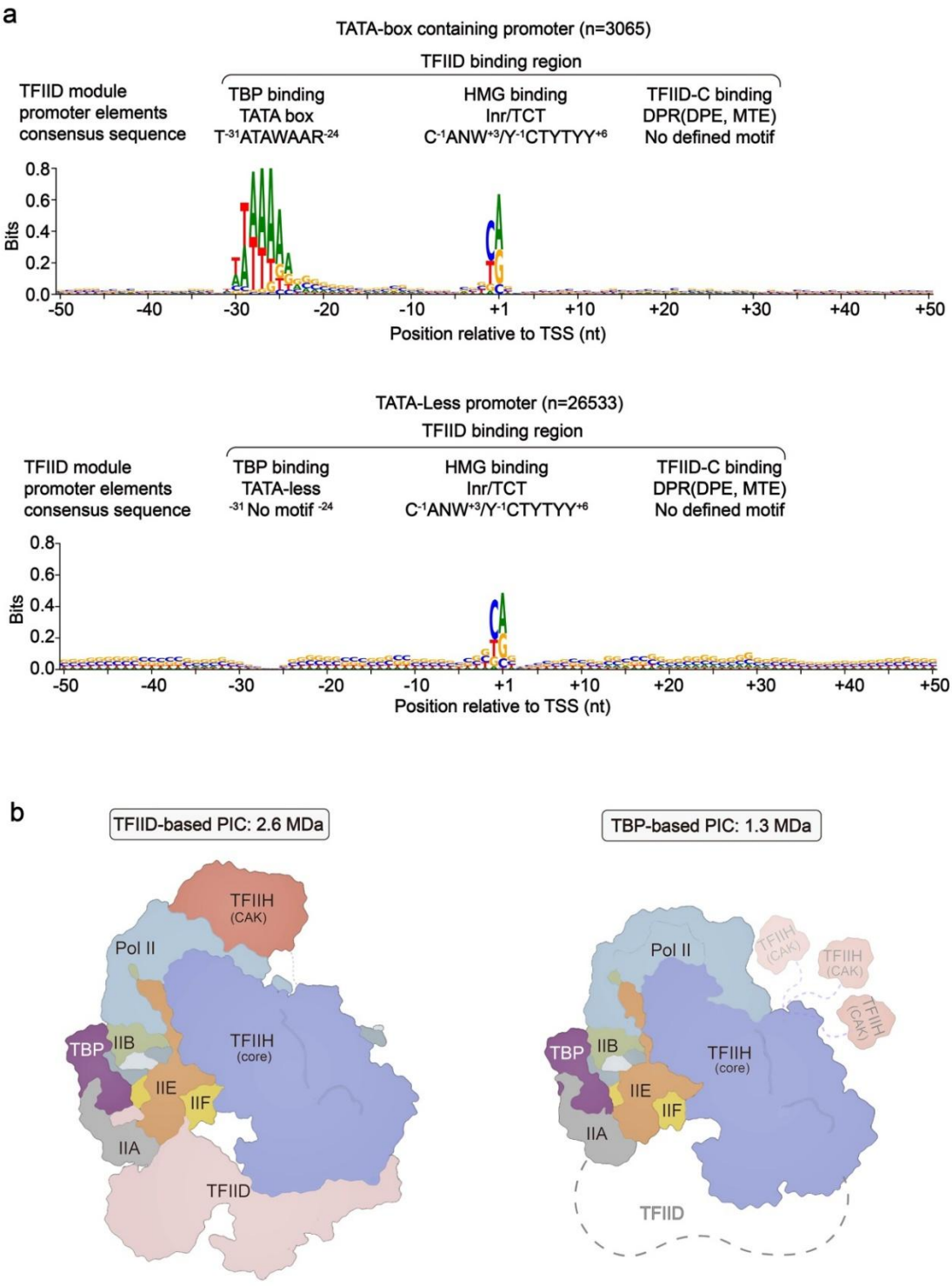
nature reviews molecular cell biology

https://doi.org/10.1038/s41580-026-01019-2

Supplementary information

The molecular basis of transcription initiation by RNA polymerase II

In the format provided by the authors and unedited



Supplementary Figure 1 | TFIID-mediated promoter recognition and PIC assembly.

a | Sequence logos and TFIID binding modules across TATA-box and TATA-less promoters. TSS positions for 29,598 human promoters, at single-base resolution, were obtained from the EPDnew database[1]. Despite the presence or absence of a consensus TATA box (at ~-31 to -24 bp), TFIID

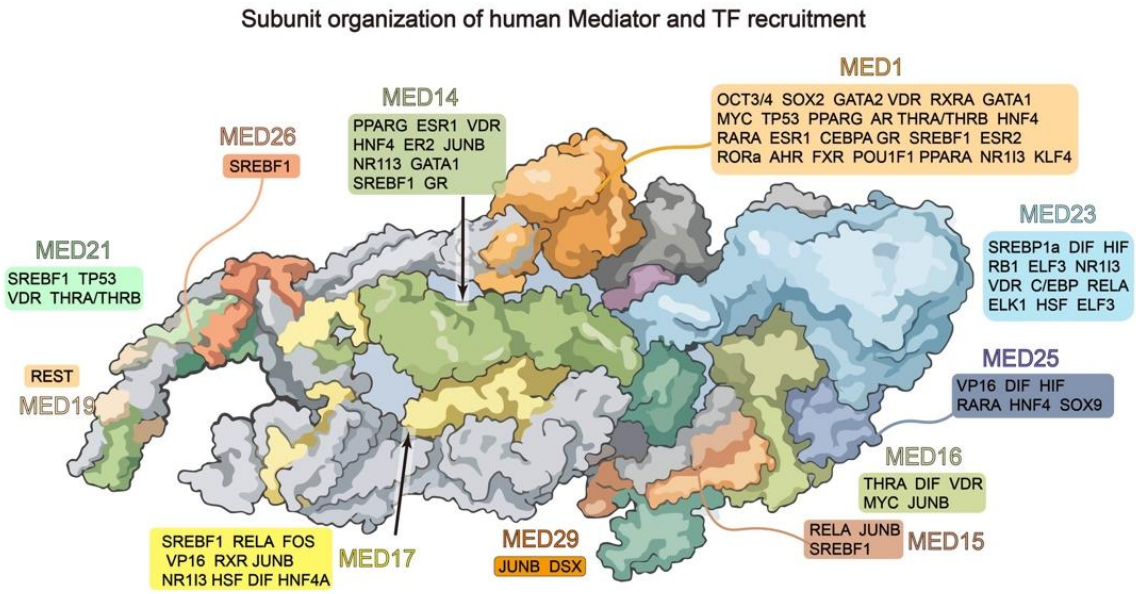
utilizes a unified binding mode and acts as a molecular ruler to bridge the upstream and downstream core promoter regions. Specific TFIID DNA-binding modules are mapped to their respective footprints. Notably, a specialized class of TATA-less promoters, primarily those of ribosomal protein genes, utilizes a TCT motif (polypyrimidine initiator) for transcription initiation.

Classification into TATA-box (n = 3,065) and TATA-less (n = 26,533) groups follows the pre-defined EPDnew annotation: a promoter is classified as TATA-box-containing if the motif is identified at position −28 (±3 bp) relative to the TSS using the program FindM, with an optimized cut-off value of −8.16 (97% true-positive rate). Sequences spanning −50 to +50 bp relative to the TSS (GRCh38/hg38 reference genome) were extracted for each group; the resulting position frequency matrices were converted to information content (in bits) and displayed as sequence logos using the Python package Logomaker[2].

b | Structural comparison between TFIID-based and TBP-based PICs. Molecular models illustrate the stark difference in size and complexity between the complete TFIID-based PIC (2.6 MDa) and the traditionally studied TBP-based PIC (1.3 MDa). The TFIID-based PIC provides a massive, rigid scaffold that not only positions TBP onto DNA (even at TATA-less promoters) but also rigidly anchors the TFIIH CAK module near the Pol II CTD, explaining the enhanced kinase efficiency of the holo-TFIID assembly compared to TBP-only systems.

The shown structures are TFIID-based PIC (PDB: 7EGB) and TBP-based PIC (PDB: 7NVY).

Abbreviations: CTD, carboxy-terminal domain; DPE, downstream promoter element; DPR, downstream core promoter region; HMG, high-mobility group; Inr, initiator; MTE, motif ten element; PIC, pre-initiation complex; Pol II, RNA polymerase II; TAF, TBP-associated factor; TBP, TATA-box binding protein; TCT, polypyrimidine initiator; TSS, transcription start site.



Supplementary Figure 2 | Structural modularity of Mediator and the CTD gating model.

The subunit organization of the tripartite human Mediator complex, comprising the Head, Middle, and Tail modules (PDB: 7EMF). Individual subunits serve as specialized docking sites for a vast repertoire of transcription factors (TFs), bridging regulatory signals to the core transcription machinery. The complex is extensively decorated with intrinsically disordered regions (IDRs), which facilitate multivalent and dynamic interactions with TF activation domains[3,4].

References

1. Dreos, R., Ambrosini, G., Cavin Perier, R. & Bucher, P. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Res* **41**, D157-164, doi:10.1093/nar/gks1233 (2013).
2. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272-2274, doi:10.1093/bioinformatics/btz921 (2020).
3. Poss, Z. C., Ebmeier, C. C. & Taatjes, D. J. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* **48**, 575-608, doi:10.3109/10409238.2013.840259 (2013).
4. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816, doi:10.1016/j.cell.2018.10.042 (2018).