STRUCTURAL AND BIOCHEMICAL STUDIES ON CHROMATIN-PROTEIN COMPLEXES IN TRANSCRIPTION INITIATION

Dissertation

for the award of the degree
"Doctor rerum naturalium"
by the Georg-August-Universität Göttingen

within the graduate program
"International Max Planck Research School Molecular Biology"
of the Georg-August University School of Science (GAUSS)

submitted by

Julio Abril Garrido

from Córdoba, Spain

Göttingen 2023

Members of the Thesis Committee

Prof. Dr. Patrick Cramer
Department of Molecular Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Prof. Dr. Kai Tittmann
Department of Molecular Enzymology
Georg-August University of Göttingen

Prof. Dr. Argyris Papantonis
Department of Translational Epigenetics
University Medical Center, Göttingen

Members of the Examination Board

Prof. Dr. Patrick Cramer (1st Referee)
Department of Molecular Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Prof. Dr. Kai Tittmann (2nd Referee)
Department of Molecular Enzymology
Georg-August University of Göttingen

Further Members of the Examination Board

Prof. Dr. Argyris Papantonis
Department of Translational Epigenetics
University Medical Center, Göttingen

Dr. Marieke Oudelaar
Department of Genome Organization and Regulation
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Prof. Dr. Hauke Hillen
Department of Cellular Biochemistry
University Medical Center, Göttingen

Dr. Juliane Liepe
Department of Quantitative and Systems Biology
Max Planck Institute for Multidisciplinary Sciences, Göttingen

Date of oral examination: 8th December 2023

Acknowledgements

I would like to start by thanking Patrick Cramer for hosting me in his laboratory during my doctoral studies. Not only did he provide a fantastic and friendly working atmosphere, but also a motivating space where I could learn from world experts in transcription. Through his trust and inspiration, I have learnt how to solve problems with independence, creativity, and simplicity. His mentoring and support will stay with me through my career.

Secondly, I would like to thank my TAC members, Kai Tittmann and Argyris Papantonis, for their mentoring through my studies. It was comforting to know they ensured I had enough supervision to learn and that my projects stayed on track. Also, I am thankful to Steffen Burkhardt and Kerstin Grüniger (i.e. Molecular Biology Program) for easing all organisational and bureaucratic matters during my studies.

Special thanks to Haibo Wang, as what I learnt from him has proved invaluable. Much of this journey would not have been possible without his input, mentoring and support. Additionally, I am grateful to have collaborated with Christian Dienemann, Michael Lidschreiber and Taras Velychko, whose input to the project was critical. Of course, many thanks to current and former members of the laboratory, in particular room 208 and annex, for sharing knowledge, fun moments and helpful advice. Special mention goes to Shintaro Aibara, Isaac Fianu, Shen Han, Elisa Oberbeckmann, Sara Osman, Paulina Seweryn, Sandra Schilbach, Felix Wagner, James Walshe, and Yumeng Zhan. I would also like to express my gratitude to Frauke Grabbe for her protein purification support, but also Ute Neef, Kerstin Maier, Kirsten Backs, Petra Rus, Thomas Schulz, Ulrich Steuerwald and Janine Blümel for organisational support.

Despite the distance, thanks to my family, to my "*Patos macizones*" friends: Alejandro, Rafa, Pablo and Prada for their support all these years. In addition, thanks to my "*Göttinger*" friends for helping me push through the hardships of a PhD: Debojit, Κωνσταντίνα, Πάνος, Sakshi, and Σοφία.

I cannot express enough gratitude to thank Κατερίνα who gave me relentless support in my ups and downs, but also has been my biggest critic – your input, advice and guidance helped me grew out of this experience. Göttingen and I would definitely not have been the same without you. Thank you, because your love and patience was more I could have ever asked for. Lastly, Λόρη taught us parenthood sticks with you like a PhD, the bond and attachment we created with you in 2 months is unique. Thanks for that, for your goofiness, happiness, and woofing support.

Publications

Part of this work has been published in a peer-reviewed journal:

Abril-Garrido, J., Dienemann, C., Grabbe, F., Velychko, T., Lidschreiber, M., Wang, H., Cramer, P. (2023). Structural basis of transcription reduction by a promoter-proximal +1 nucleosome. Molecular Cell 83, 1798-1809. https://doi.org/10.1016/j.molcel.2023.04.011.

Author contributions: J.A.G. carried out all experiments and data analysis unless stated otherwise. J.A.G. and C.D. collected cryo-EM data. J.A.G. purified Pol II, histones, and prepared nucleosomes and F.G. assisted with protein purification of initiation factors. T.V. collected and M.L. analysed TT-seq, RNA-seq and MNase-seq data. H.W. and P.C. supervised research. J.A.G., H.W. and P.C. interpreted the data and wrote the manuscript, with input from all authors.

– The following sections were extracted from Abril-Garrido *et al.* and modified to adapt them to the format of this work:

2. Materials
3. Methods
- Sections **3.2-3.6** were taken from the mentioned manuscript and section **3.1** was written for this dissertation.

4. Results
- A few sentences were added to section **4.2** on this dissertation, with the aim to describe preliminary work that was not included in the peer-reviewed publication (related to **Figure 6**).

– As an exception to the above, abstract, introduction, discussion and outlook sections have been written entirely for the purpose of this dissertation.

– Main figures, supplemental figures and tables were extracted from Abril-Garrido *et al*. and adapted to this dissertation. Therefore, the numbering of these elements deviates from the published work. The following figures have been hereby newly added and are not part of the peer-reviewed publication:

Figure 1. Nucleosome structure.
Figure 2. Gene promoter architecture and accessibility.
Figure 3. Pol II transcription cycle.
Figure 4. Pol II transcription initiation at a glance.

Figure 6. Cryo-EM screening of most stable PIC-nucleosome complex for high-resolution determination.
Figure S1. Histone octamer assembly.
Table 1. Reagents, resources, and data used or generated in this study. **This table, now included as a main table, belongs to the key resources table extracted from the published manuscript.**
Table 2. H2A, H2B, H3 and H4 histones purification buffers.
Table 3. Histone octamer and nucleosome reconstitutions buffers.

The parts adapted from the manuscript, together with the remaining of this written dissertation, were written in British English.

Table of contents

List of figures

List of tables

List of abbreviations

Å	Ångstrom
AdMLP	Adenoviral major late promoter
ADP	Adenosine triphosphate
AIEX	Anion exchange
ATP	Adenosine triphosphate
BioID	Proximity-dependent biotinylation
β-ME	β-Mercaptoethanol
bp	Base pairs
BSA	Bovine serum albumin
BSD	DOS-like domain
CAK	CDK-activating kinase
CC	Closed PIC
CDK7/8/9	Cyclin-dependent kinase 7/8/9
CEX	Cation exchange
CHD	Chromodomain helicase DNA binding
CKM	CDK8 kinase module
CPA	Cleavage and polyadenylation complex
cPIC	Core PIC
Cryo-EM	Electron cryo-microscopy
Cryo-ET	Electron cryo-tomography
CTCF	CCCTC-binding factor
CTF	Contrast transfer function
CycH	Cyclin H
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DSIF	DRB-sensitivity inducing factor

DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ETS	Erythroblast transformation specific
eWH	Extended winged helix
GraFix	Gradient fixation
GTFs	General transcription factors
GTP	Guanosine triphosphate
H2A/H2B	Histone H2A/H2B
H3/H4	Histone 3/4
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC	Intermediate complex
IDRs	Intrinsically disordered regions
IEX	Ion exchange
IMR	Initially melted region
INO80	Inositol requiring mutant 80
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ISWI	Imitation switch
ITC	Initially transcribing complex
KCl	Potassium chloride
KDa	Kilodalton
KOH	Potassium hydroxide
MAT1	Menage-a-trois 1/CAK assembly factor MAT1
MDa	Megadalton
$MgCl_2$	Magnesium chloride
miRNA	Micro RNA
MNase-seq	Micrococcal nuclease-sequencing
mRNA	Messenger RNA
NaCl	Sodium chloride

NCP	Nucleosome core particle
NDR	Nucleosome-depleted region
NELF	Negative elongation factor
NER	Nucleotide excision repair
NFR	Nucleosome-free region
$(NH_4)SO_4$	Ammonium sulphate
NMR	Nuclear magnetic resonance
nt	Ribonucleotides
NTPs	Nucleoside triphosphate
NTT	N-terminal tail
OC	Open PIC
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PAS	Polyadenylation signal
PCR	Polymerase chain reaction
PEC	Paused elongation complex
PHD	Plekstrin homology domain
Pi	Phosphate ion
PI	Protease inhibitors
PMSF	Phenylmethylsulphonyl fluoride
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol II CTD	RNA polymerase II C-terminal domain
Pol III	RNA polymerase III
Pol IV	RNA polymerase IV
Pol V	RNA polymerase V
Pols	DNA-dependent RNA polymerases
PP1	Protein phosphatase 1
PP2A	Protein phosphatase 2A

Pre-mRNA	Messenger RNA precursor
P-TEFb	Positive transcription elongation factor b
PVA	Polyvinyl alcohol
RNA	Ribonucleic acid
RPKM	Reads per kilobase per million reads
rRNA	Ribosomal RNA
S. cerevisiae	*Saccharomyces cerevisiae*
S. scrofa	*Sus scrofa*
SAGA	Spt-Ada-Gcn5 acetyltransferase complex
SDS	Sodium dodecyl sulphate
SGD	Salt-gradient dialysis
SHL	Superhelical location
siRNA	Small-interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
ssDNA	Single-stranded DNA
Swi/snf	Switch/sucrose non-fermentable
TADs	Topologically associated domains
TAFs	TBP-associated factors
TBE	Tris-borate EDTA
TBP	TATA-box binding protein
TCEP	Tris(2-carboxyethyl)phosphine
TE	Tris-EDTA
TEV	Tobacco etch virus
TFs	Transcription factors
TFIIA/B/D/E/F/H/S	Transcription factor IIA/IIB/IID/IIE/IIF/IIH/IIS
tRNA	Transfer RNA
TSS	Transcription start site
TT-seq	Transient transcriptome sequencing

U1 snRNP	U1 small nuclear ribonucleoprotein
WH	Winged helix
X. laevis	*Xenopus laevis*
XPA	Xeroderma pigmentosum group A-complementing protein
XPB	Xeroderma pigmentosum group B-complementing protein
XPD	Xeroderma pigmentosum group D-complementing protein
$ZnCl_2$	Zinc chloride

Abstract

Gene expression is an essential process in the cell for the development and homeostasis of any organism. The eukaryotic genome is tightly packed with nucleosomes, the basic unit of chromatin, which protects DNA integrity. Chromatin compaction regulates the DNA accessibility of genes, thus having an effect on transcription and gene regulation. Gene promoters are flanked by the -1 and +1 nucleosome upstream and downstream of the transcription start site (TSS), respectively, while the promoter region is usually nucleosome-depleted to facilitate the assembly of the transcription initiation machinery. In eukaryotes, the DNA-dependent RNA polymerase II (Pol II) transcribes protein-coding genes, which can initiate transcription at promoters only in complex with the general transcription factors (GTFs) TBP, TFIIA, TFIIB, TFIIF, TFIIE, and TFIIH. Their recruitment nucleates the assembly of the RNA polymerase II pre-initiation complex (PIC), known to be one of the major transcription regulation checkpoints, so that RNA synthesis can begin. Previously, it was shown that the nucleosome is a key repressor of transcription initiation, yet the regulatory mechanism is unclear. To study the regulatory role of the +1 nucleosome on transcription initiation, I performed functional studies and investigated the inherent mechanism by electron cryo-microscopy (cryo-EM). First, we correlated RNA synthesis (TT-seq) with the position of the +1 nucleosome (MNase-seq) in HEK293 cells. We describe that promoter proximity of the +1 nucleosome correlates with a reduction of transcription *in vivo* in a gradual manner. I corroborated these results *in vitro* and determined two cryo-EM structures of a mammalian PIC engaged with a promoter-proximal +1 nucleosome. The PIC assembles productively when the edge of the +1 nucleosome is located ~20 bp downstream of the TSS. However, when the nucleosome shifts its position ~10 bp upstream, the PIC forms in an inactive conformation. The transcription factor IIH (TFIIH) is found in its closed state where only one lobe of its ATPase subunit XPB contacts DNA, incompatible with DNA opening. Therefore, I provide here a regulation mechanism by which a promoter-proximal +1 nucleosome can impair productive PIC assembly and thereby reduce transcription. Together with recent structural studies, it emerges that the effect of the +1 nucleosome on transcription depends on its relative position to the gene promoter and PIC.

1. Introduction

1.1 Gene expression

The nucleus is a heavily dense compartment within the cell and the information contained within must be decoded effectively for proper cell development and maintenance (Gerace and Burke, 1988). This information is encoded within deoxyribonucleic acid (DNA) molecules which first must be transcribed into ribonucleic acid (RNA) by RNA polymerases (Pols). This rate-limiting step, coined as transcription (Crick, 1970; Orphanides and Reinberg, 2002), is key so that all necessary building blocks are produced for protein synthesis, also called translation. Whereas prokaryotes perform RNA and protein synthesis within the cytoplasm, eukaryotes acquired an additional layer of regulation through evolution (Orphanides and Reinberg, 2002; Werner and Grohmann, 2011). Synthesis of RNA occurs in the nucleus (transcription) while production of proteins (translation) takes place in the cytoplasm (Woldringh, 2002). These two processes are tightly regulated within their compartments, and the spatial separation acquired through evolution introduced another important process within the cell – nuclear-cytoplasmic transport through the nuclear pore complex (Gorlich and Kutay, 1999).

1.2 Eukaryotic genome organisation

The entire DNA information contained within the nucleus comprise what is called the genome. The human genome consists of 3 billion base pairs distributed over 23 pairs of chromosomes, however, its length goes over 2 m when the DNA is extended. This presents a problem for the cell as the genomic information is stored in the nucleus, which has a diameter of only ~10 µm. Whereas prokaryotes deal with this problem only by twisting the DNA in a process called supercoiling, eukaryotes also require the help of accessory proteins to pack the DNA information into a structure coined as chromatin.

In the late 19th century, W. Flemming observed a phosphorous acidic substance in the eukaryotic nucleus (Flemming, 1882). Subsequent work of F. Miescher, A. Kossel and other colleagues further described the proteins found in this substance, first as nuclein and later as histones (Miescher, 1871; Kossel, 1884). The characterisation of chromatin was only possible in the late 20th century, after it

became clear that DNA carries the genetic information and the double helical DNA structure was determined (Avery et al., 1944; Franklin and Gosling, 1953; Watson and Crick, 1953). Then, histones extraction from the nucleus (Stedman and Stedman, 1951), evidence for dimer and tetramer formation (Kelley, 1973; Kornberg and Thomas, 1974), and the visualisation of chromatin as "beads on a string" through electron microscopy studies (Olins and Olins, 1974), laid the foundations for the characterisation of chromatin. This seminal work, together with digestion and cross-linking data, led to the model that DNA wraps around an octamer of histones forming the nucleosome (Hewish and Burgoyne, 1973; Kornberg, 1974; Oudet et al., 1975).

1.3 Chromatin structure

1.3.1 Nucleosomes: the basic unit of chromatin

Nucleosomes therefore represent the basic repetitive unit of chromatin where they regulate multiple processes in the cell, such as DNA replication, repair, and transcription (Phillips and Johns, 1965; Bram and Ris, 1971; Huberman, 1973; Kornberg, 1974; Kornberg and Thomas, 1974; Olins and Olins, 1974; Finch et al., 1975; Germond et al., 1975; Oudet et al., 1975; Woodcock et al., 1976; Kornberg and Lorch, 1999). The term "nucleosome core particle" (NCP) was traditionally given to the DNA-bound octamer of histones, whereas "nucleosome" is used to refer to the NCP with linker DNA of variable lengths (van Holde and van Holde, 1989). To ease readability through this dissertation, we will only employ the term "nucleosome".

Core histones H2A, H2B, H3 and H4 are very small basic proteins, consisting of 100-130 amino acids (~11-15 KDa). All four histones are composed of three α-helices connected by linkers and make up what is called the "histone fold". H2A-H2B and H3-H4 dimerise in an antiparallel fashion through their α2 helix (Arents et al., 1991) (**Figure 1A**). In the nucleosome, two H3-H4 dimers form a tetramer through a four-helix bundle mediated by H3, and the two H2A-H2B dimers interact with the H3-H4 tetramer through another four-helix bundle mediated by the histone folds of H2B and H4 (**Figure 1B**). The group of A. Klug first described at low-resolution that ~1.7 turns of DNA wrap around the histone octamer in a left-handed

manner (Richmond et al., 1984). Subsequently, a 2.8 Å crystal structure from recombinantly expressed *Xenopus laevis* (*X. laevis*) core histones determined the structure of the nucleosome at high-resolution (Luger et al., 1997). This groundbreaking study represented the first reliable evidence for the structure of the nucleosome, showing a pseudo two-fold symmetric arrangement around its central axis, also called the nucleosome dyad. The histone-facing major groove at the dyad is defined as the superhelical location (SHL) 0 and serves as the reference for the rotational orientation of the DNA on the nucleosome. Using the dyad as a reference, all histone-facing major grooves of the nucleosome are defined from SHL -7 to +7. This structure determination was essential to decipher that the structure of the nucleosome is stabilised through a combination of direct hydrogen bonds and with water molecules, non-polar and charge-based interactions. This model also shed light on the importance of the core histones flexible N-termini unstructured regions, known as the “histone tails”. These unstructured tails emanate from the nucleosome core, interact with the wrapped DNA, and further extend into the solvent. Importantly, these histone tails are long known to interact with other nucleosomes and linker DNA (Davey et al., 2002; Dorigo et al., 2003), and to be signalling platforms where different “writer” enzymes deposit post-translational modifications to regulate downstream processes, such as methylation, acetylation, phosphorylation, and ubiquitination (Jenuwein and Allis, 2001; Allis and Jenuwein, 2016).

Core histones are highly conserved across metazoan species and, therefore, the nucleosome structure from *X. laevis* (Luger et al., 1997; Davey et al., 2002), chicken (Harp et al., 2000), *Saccharomyces cerevisiae* (*S. cerevisiae*) (White et al., 2001) and human proved to be very similar (Tsunaka et al., 2005). H2A-H2B bear a highly acidic region called the “acidic patch”, mainly composed of acidic amino acids (e.g. aspartates and glutamates), which has also been described to be conserved. The acidic patch has been reported to interact with other nucleosomes (Kalashnikova et al., 2013), as well as comprising a hub for recruiting chromatin-binding factors such as chromatin modifiers, remodellers and pioneer factors (McGinty and Tan, 2016, 2021). To this end, histones variants have been described to be important in regulating the signalling occurring on the nucleosome. For example, the chromatin remodeller SWR1 replaces H2A for H2A.Z, changes the octamer surface by

introducing a metal ion and destabilises the interaction previously established between H2A and H2B. Such histone variant is thought to potentially influence the interactions with chromatin-binding factors and, therefore, regulate their recruitment on different nucleosomes (Redon et al., 2002).

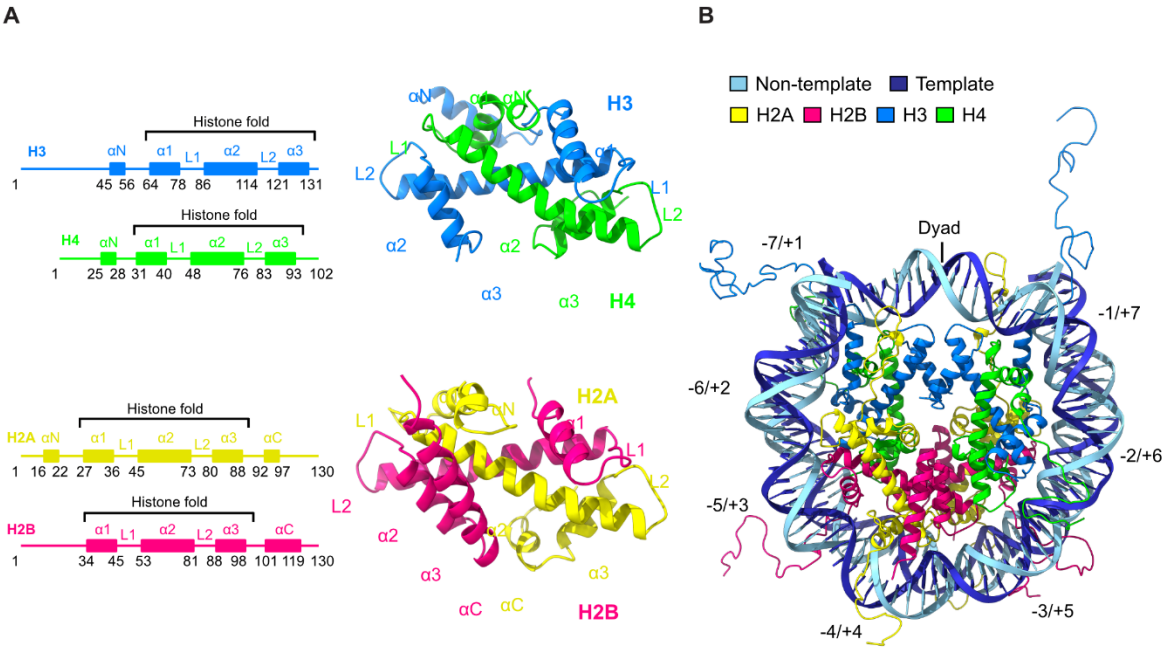


Figure 1. Nucleosome structure.

(A) Core histones architecture domain (left most), H3-H4 and H2A-H2B dimers structures (right most).

(B) Crystal structure of the nucleosome where the dyad and SHL positions are indicated. The models depicted here were modified after Davey et al. (2002) (PDB ID: 1KX5). The colour code is provided at the top of panel (B).

1.3.2 Higher-order chromatin structures

For higher-order DNA compaction, nucleosomes have been described to arrange in repeats forming different types of chromatin fibres. Whereas the existence of the so-called 10-nm fibre is well-established (Maeshima et al., 2014), it is unclear of whether nucleosomes arrange in 30-nm fibres *in vivo* (Maeshima et al., 2010; Fussner et al., 2011; Maeshima et al., 2019). 30-nm fibres have only been observed in *in vitro* early electron microscope studies and in specific cell types (Finch and Klug, 1976), where nucleosomes were proposed to organise consecutively in a regular fashion. Similarly, recently X-ray crystallography and cryo-EM studies

support the presence of highly ordered and compacted nucleosomes *in vitro* (Schalch et al., 2005; Song et al., 2014; Ekundayo et al., 2017; Garcia-Saez et al., 2018; Adhireksan et al., 2020; Dombrowski et al., 2022; Zhou et al., 2022). However, early electron microscopy studies, together with recent fluorescence microscopy and *in vivo* studies (McDowall et al., 1986; Dekker, 2008; Nishino et al., 2012; Ricci et al., 2015), suggest that nucleosomes do not organise in an ordered manner and challenge the hierarchical folding model previously proposed (Sedat and Manuelidis, 1978). Instead, a model emerged that suggests nucleosomes fold irregularly in small clusters, forming 10-nm fibres, which is sustained by a condensin scaffold that keeps these genomic regions associated to each other (Mirny et al., 2019). These regions have also been recently coined as topologically associated domains (TADs), which associate between each other and constitute chromosomal territories (Dixon et al., 2016; Lupianez et al., 2016; Hansen et al., 2018; Szabo et al., 2019; Zhou et al., 2019). This model supports chromatin as a dynamic entity that organises and folds in a loose fashion. Here, a mechanism called loop extrusion can readily modify the size and distribution of different chromatin loops and thus facilitate chromatin accessibility.

1.4 Chromatin accessibility

1.4.1 Nucleosome architecture at gene promoters

A highly regulated step in the cell is therefore the ability to access the DNA information occluded by chromatin (**Figure 2**). Although this is essential for many processes such as DNA repair, replication, and transcription, we will only discuss herein its implications in transcription. The gene promoter is a DNA region responsible for recruiting RNA polymerase II (Pol II) to protein-coding genes so that messenger RNA (mRNA) can be produced (Sainsbury et al., 2015). In response to different signalling cascades, the promoter region constitutes a hub that organises whether transcription of a given gene must undergo transcription initiation. Here, nucleosomes play a critical regulatory role by occluding the access of the transcription initiation machinery to the promoter DNA (Kornberg and Lorch, 1999; Li et al., 2007). As previously mentioned, nucleosomes arrays *in vivo* acquire an irregular and loose fold, which also stands true for inactive gene promoters (Schones et al., 2008; Jiang and Pugh, 2009a; Valouev et al., 2011). Promoter-

flanking nucleosomes, also known as -1 and +1 nucleosomes, can cover the promoter region to different extents and therefore display an irregular arrangement, also called “fuzzy” nucleosomes. In line with this, whereas nucleosomes at promoters of housekeeping genes are often phased in a regular manner to facilitate transcription, developmental or inducible genes show fuzzy nucleosome patterns (Lee et al., 2007; Mavrich et al., 2008; Tirosh and Barkai, 2008). This evidence elegantly illustrate how nucleosomes position and distribution on specific genes are key to regulate development and cell homeostasis. At promoter regions, positioning nucleosomes is a key step to generate ~150-200 bp of nucleosome-free DNA, called nucleosome-depleted or nucleosome-free regions (NDRs, NFRs). Only then, Pol II and the transcription machinery can be recruited to begin transcription (**Figure 2**) (Workman, 2006; Li et al., 2007; Fuda et al., 2009).

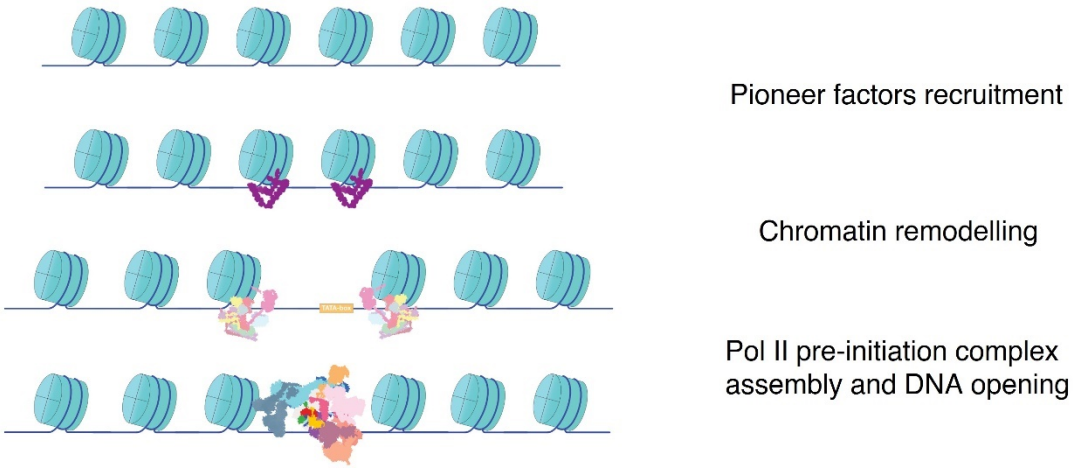


Figure 2. Gene promoter architecture and accessibility.

Gene promoters are occluded by the presence of nucleosomes, where the combination of multiple binding and remodelling events precede the generation of NDRs. First, transcription factors (in purple) bind nucleosomes to recruit ATP-dependent chromatin remodellers (in pastel colours), and thereby generate NDRs. This process removes and remodels nucleosomes so that gene promoters are accessible for the recruitment of the transcription initiation machinery (dark pastel colours).

1.4.2 Remodelling events of promoter-flanking nucleosomes

To generate and position promoter-flanking nucleosomes around NDRs, nucleosomes must be actively positioned, removed and/or modified (Li et al., 2007; Zhou et al., 2019). Whereas DNA sequences, especially AT-rich sequences, can be nucleosome repelling (Segal et al., 2006; Kaplan et al., 2009), additional factors are required to remodel chromatin (Klemm et al., 2019). Transcription factors called "pioneer factors" can directly bind nucleosomes, and together with histone modifications and variants (e.g. methylation, acetylation, ubiquitination, H2A.Z, H3.3), can ultimately recruit ATP-dependent chromatin remodellers (**Figure 2**) (Zaret and Carroll, 2011; Iwafuchi-Doi and Zaret, 2014; Balsalobre and Drouin, 2022). Chromatin remodellers are ATP-dependent enzymes able to slide, unwrap or evict nucleosomes to make the underlying DNA accessible (Flaus and Owen-Hughes, 2004; Smith and Peterson, 2005; Saha et al., 2006).

There are 4 main families of chromatin remodellers: SWI/SNF (switch/sucrose non-fermentable), ISWI (Imitation Switch), INO80 (Inositol requiring mutant 80) and CHD (Chromodomain helicase DNA binding). In *S. cerevisiae*, RSC and Swi/Snf (PBAF and BAF complexes in human, respectively) are multi-subunit complexes that have been extensively shown to slide promoter-occluding nucleosomes, evict them at AT-rich regions and ultimately generate NDRs (Clapier and Cairns, 2009; Lorch et al., 2014; Krietenstein et al., 2016; Kubik et al., 2019). On the other hand, yeast chromatin remodellers like CHD1, INO80, ISW1 and ISW2 are important for sliding and spacing nucleosomes (Gamarra and Narlikar, 2021). Two distinct pathways have been proposed for the positioning of the +1 and -1 nucleosomes in yeast. Whereas a transcription factor hub formed at the NDR can recruit the remodellers ISW2 or ISW1a (Krietenstein et al., 2016), INO80 can recognise DNA topology and nucleosome-proximal sequences to position these nucleosomes (Oberbeckmann et al., 2021a). While this occurs at the gene promoter, remodellers such as ISW1b regularly phase nucleosomes present in the gene body (Krietenstein et al., 2016).

Transcription can be productive only through a prior phased spacing of nucleosomes, and therefore remodellers of CHD and ISWI families cooperate to generate NDRs of variable length with phased nucleosomes (Gamarra and Narlikar, 2021). Structural studies have reported that CHD remodellers help Pol II

transcribing through a nucleosome (Farnung et al., 2017; Farnung et al., 2021). Additionally, numerous evidence describe how Pol II nucleosome passage generates hexasomes, a nucleosome which has lost a H2A-H2B dimer copy (Kireeva et al., 2002; Bintu et al., 2011; Bevington and Boyes, 2013; Hsieh et al., 2013; Cole et al., 2014). This is compelling evidence as hexasomes were recently described as the preferred substrate of INO80 (Hsieh et al., 2022). At the same time, whereas INO80 replaces H2A.Z for the canonical H2A at the +1 nucleosome (Papamichos-Chronakis et al., 2011; Wang et al., 2016; Brahma et al., 2017; Ranjan et al., 2020), another chromatin remodeller, the SWR1 complex exchanges H2A for H2A.Z at this nucleosome (Krogan et al., 2003; Ranjan et al., 2013; Yen et al., 2013). Once more, this clearly exemplifies how different chromatin remodellers cooperate to fine-tune the promoter-flanking nucleosomes composition, the generation of NDRs and, therefore, transcription.

1.5 Chromatin and gene regulation

Through different cell stages, the distinct spacing between nucleosomes modifies their packing order within chromatin fibres, which correlates with different levels of transcription (Guenther et al., 2007; Li et al., 2007; Schones et al., 2008; Valouev et al., 2011; Schwartz et al., 2019; Serizay et al., 2020). Chromatin can be highly or loosely compacted in the nucleus which forms heterochromatin and euchromatin, respectively. These two different chromatin states have been associated with inactive (heterochromatin) and active transcription (euchromatin) (Henikoff, 2000; Richards and Elgin, 2002; Allshire and Madhani, 2018; Morrison and Thakur, 2021). Nevertheless, this canonical view has been challenged recently as RNA synthesis can occur in heterochromatin compartments, albeit in low levels (Penagos-Puig and Furlan-Magaril, 2020). In addition, new technologies allowed us to understand that different genomic regions associate in chromatin loops, also called TADs, which ultimately can interact with each other comprising chromosomal compartments (Dixon et al., 2016; Lupianez et al., 2016; Hansen et al., 2018; Szabo et al., 2019; Zhou et al., 2019).

As mentioned above, promoter NDRs are flanked by the -1 and +1 nucleosomes, located immediately upstream and downstream of the TSS, respectively (Jiang and Pugh, 2009b; Struhl and Segal, 2013; Hughes and Rando, 2014; Lai and Pugh,

2017; Kornberg and Lorch, 2020; Serizay et al., 2020). The +1 nucleosome is thought to regulate transcription initiation since the PIC assembles near to its position (Jiang and Pugh, 2009b; Li et al., 2010; Rhee and Pugh, 2012; Tramantano et al., 2016; Shao and Zeitlinger, 2017; Serizay et al., 2020; Rossi et al., 2021; Santana et al., 2022). Different studies show that the +1 nucleosome position changes with gene activity (Guenther et al., 2007; Schones et al., 2008; Valouev et al., 2011; Schwartz et al., 2019; Serizay et al., 2020), where its edge locates ~40-60 base pairs (bp) downstream of the TSS in active genes (Albert et al., 2007; Mavrich et al., 2008; Schones et al., 2008; Weber et al., 2014; Zumer et al., 2021), and closer to the TSS in inactive ones (Valouev et al., 2011), or with stalled transcription activity (Schones et al., 2008). This agrees with *in vivo* studies in yeast which show that the presence of nucleosomes reduces transcription activity (Han and Grunstein, 1988; Korber and Barbaric, 2014; Klein-Brill et al., 2019; Kubik et al., 2019), as well as with the long-standing biochemical observations that transcription initiation *in vitro* is inhibited by a nucleosome located at the promoter (Knezetic and Luse, 1986; Lorch et al., 1987; Workman and Roeder, 1987; Li et al., 2010). Additionally, numerous evidence suggest that the repressive nature of the +1 nucleosome serves as a tool for TSS selection, which avoids cryptic transcription events by preventing the production of aberrant mRNA products (Whitehouse et al., 2007; Hennig et al., 2012; Pointner et al., 2012; Smolle and Workman, 2013). Through these studies, it becomes evident that nucleosomes are key regulators of gene activity. Therefore, the dynamic role of the nucleosome must be further investigated to better understand how chromatin regulates transcription.

1.6 Transcription

The simplest transcriptional systems has been described for phages and viruses (Cheetham and Steitz, 2000). A single Pol, generally made up of one subunit, performs this process aided by regulatory promoter sequences (Rosa, 1979; Sousa et al., 1993; Cheetham et al., 1999). In prokaryotes, a single RNA polymerase holoenzyme, made up of four core subunits and the sigma transcription factor, is solely responsible for transcribing DNA into all types of RNA (Kang et al., 1997; Zhang et al., 1999; Woldringh, 2002; Feklistov et al., 2014; Murakami, 2015). However, different sigma factors can selectively regulate which genes to transcribe

In eukaryotes, the complexity of transcription was also complemented by the presence of different DNA-dependent RNA polymerases (Pols) (Roeder and Rutter, 1969; Kedinger et al., 1970), numerous transcription factors (TFs), activators (Engelke et al., 1980; Matsui et al., 1980; Payvar et al., 1981; Dynan and Tjian, 1983; Buratowski et al., 1989; Flores et al., 1992), and coactivators (Flanagan et al., 1991; Meisterernst et al., 1991; Thompson et al., 1993; Verrijzer et al., 1995; Grant et al., 1997; Ryu and Tjian, 1999; Ryu et al., 1999; Malik et al., 2000; Larschan and Winston, 2001).

The main eukaryotic RNA polymerases were first isolated ~50 years ago from sea urchin embryos by separating them with anion exchange chromatography (Roeder and Rutter, 1969). Today, it is known that eukaryotic organisms employ up to 5 different multi-subunit Pols to transcribe DNA into different types of RNA in the nucleus (Roeder and Rutter, 1969; Herr et al., 2005; Kanno et al., 2005; Haag and Pikaard, 2011; Matzke and Mosher, 2014). Other Pols are responsible for transcribing the chloroplast and mitochondrial genome, which are generally comprised by one single-subunit, reminiscent of phages polymerases (Masters et al., 1987; Asin-Cayuela and Gustafsson, 2007; Yagi and Shiina, 2014; Hillen et al., 2018). All Pols show a high conservation of the core and catalytic subunits and only differ in a few regulatory subunits (Vannini and Cramer, 2012; Roeder, 2019).

RNA polymerase I (Pol I) transcribes ribosomal RNA (rRNA), which are precursors of 28S, 18S and 5.8S rRNA (Hannan et al., 1998; Goodfellow and Zomerdijk, 2013; Neyer et al., 2016; Misiaszek et al., 2021). Whereas RNA polymerase III (Pol III) also generates rRNA and some small nuclear RNA (snRNA), it mainly produces transfer RNA (tRNA) (Schramm and Hernandez, 2002; Dieci et al., 2007; Vannini et al., 2010; Arimbasseri and Maraia, 2016; Abascal-Palacios et al., 2018; Ramsay et al., 2020; Girbig et al., 2021). Plant-specific RNA polymerase IV (Pol IV) and RNA polymerase V (Pol V) catalyse the production of small and long-noncoding RNAs, respectively, to regulate DNA methylation-dependent gene silencing (Herr et al., 2005; Kanno et al., 2005; Haag and Pikaard, 2011; Matzke and Mosher, 2014; Huang et al., 2021; Xie et al., 2023). So far, it is obvious how indispensable eukaryotic transcription is – both rRNA and tRNA are essential players for protein synthesis to take place. Nevertheless, the synthesis of the messenger RNA

precursor (pre-mRNA) is the remaining keystone to be laid. This is catalysed by RNA polymerase II (Pol II), which also produces small nucleolar RNA (snoRNA), microRNA (miRNA), small-interfering RNA (siRNA) and most snRNA (Cramer et al., 2000; Cramer et al., 2001; Cramer, 2004; Liu et al., 2013; Osman and Cramer, 2020). Pol II, therefore, indirectly regulates downstream processes: co- and post-transcriptional maturation of the pre-mRNA into mRNA, their nuclear export and eventual translation into protein (Moore and Proudfoot, 2009; Livingstone et al., 2010; Schmid and Jensen, 2010). The correct functioning of Pol II in the cell is the remaining cornerstone which, together with Pol I and Pol III, transcribe all necessary RNA molecules to complete the assembly of the ribosome machinery and, overall, permit the synthesis of polypeptide chains.

1.7 RNA polymerase II transcription cycle

RNA polymerase II carries out eukaryotic transcription in cycles divided in 3 major stages: initiation, elongation, and termination/recycling (**Figure 3**) (Shandilya and Roberts, 2012; Hantsche and Cramer, 2016; Osman and Cramer, 2020).

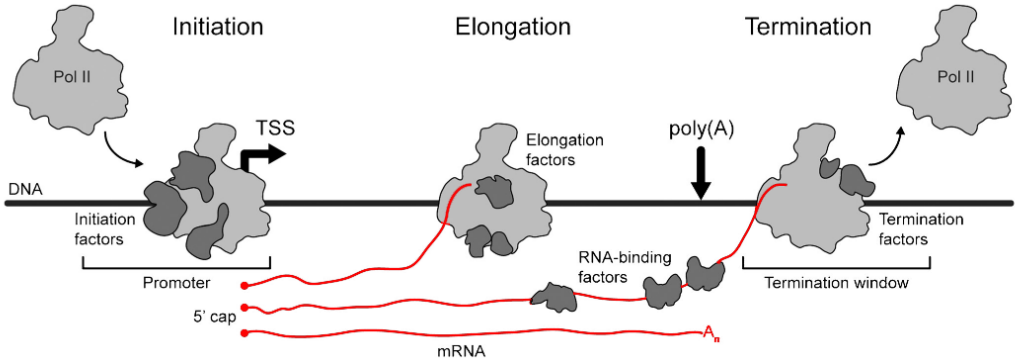


Figure 3. Pol II transcription cycle.

GTFs (dark grey) aid in the recruitment of Pol II (grey) to the gene promoter. Once initiation has begun, the nascent RNA chain is synthesised, and new ribonucleotides are added through the process of elongation. For this transition to take place, GTFs must dissociate from Pol II to allow for the recruitment of elongation factors (dark grey in scheme). In this transition, as soon as the 5'-end of the RNA goes out of the Pol II exit tunnel, capping enzymes attach a GTP molecule to the RNA 5'-end to enhance its stability. Afterwards, when Pol II reaches the polyadenylation signal, a long stretch of adenines (poly(A) tail) gets attached to the 3'-end of the transcript and the RNA gets cleaved, thanks to the action of termination factors (dark grey). Finally, termination and elongation factors dissociate with the help of exonucleases

so that Pol II molecules can get recycled for a new round of transcription. Modified after Hantsche and Cramer (2016).

Firstly, a gene promoter must be recognised in order to nucleate the assembly of the RNA polymerase II pre-initiation complex (PIC), known as the first checkpoint of the cycle (Matsui et al., 1980; Buratowski et al., 1989; Flores et al., 1992; Conaway and Conaway, 1993; Roeder, 1996; Grunberg and Hahn, 2013; Sainsbury et al., 2015). To correctly identify the transcript to be produced, Pol II recruitment and PIC assembly is tightly coordinated by the presence of general transcription factors (GTFs), activators and coactivators. Only then, Pol II can rapidly identify the transcription start site (TSS), which corresponds to the first 5' ribonucleotide that will be added on the initiating RNA chain. Once the right promoter has been recognised, a productive PIC translocates the double-stranded DNA molecule, unwinds its two strands by melting 11-15 base pairs (bp) ('transcription bubble'), and positions it into the cleft of Pol II so that initial RNA synthesis can begin (Kim et al., 2000; Hahn and Young, 2011; He et al., 2013; Fishburn et al., 2015; Sainsbury et al., 2015; He et al., 2016; Nogales et al., 2017a; Schilbach et al., 2017; Dienemann et al., 2019; Aibara et al., 2021; Schilbach et al., 2021). In the first place, a DNA-RNA hybrid forms after the addition of the first ribonucleotide (nt), which is translocated towards the Pol II RNA exit tunnel as ribonucleotides are added sequentially to the RNA molecule (Brueckner et al., 2009). The addition of these ribonucleotides is catalysed by a two-metal ion mechanism in a process called the nucleotide addition cycle. Here, aspartate residues of the active site stabilise two magnesium ions, one present at the active site and another one provided by the incoming ribonucleotide. The active site's magnesium ion coordinates a nucleophilic attack performed from the hydroxyl group of the last added 3' ribonucleotide on the α-phosphate bond of the new ribonucleotide triphosphate, which will then join to the growing RNA chain. A pyrophosphate is produced as a result of this reaction, which is stabilised by the magnesium ion that the newly added ribonucleotide provided. Pol II subsequently translocates the DNA-RNA hybrid in a 3'-5' direction for the cycle to progress.

After a 20-25 nt-long RNA chain has been successfully produced, Pol II escapes the gene promoter: the longer length of the RNA molecule and the position of Pol II along DNA ejects the GTFs and other auxiliary factors (Luse and Jacob, 1987; Holstege et al., 1997; Luse, 2013). Whereas the precise mechanism of this transition

is unclear, it is of utmost importance since Pol II will be then in optimal conditions to speed up and elongate the RNA chain. However, before Pol II can processively elongate the RNA molecule, it must undergo the second checkpoint of this process – promoter-proximal pausing. Promoted by a combination of GC-rich sequences and so-called promoter-proximal pausing factors, Pol II slows down after 25-150 nt have been added on the growing RNA chain to prevent early termination events in metazoans (Gilmour and Lis, 1986; Strobl and Eick, 1992; Adelman and Lis, 2012; Liu et al., 2015). The binding of DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) to Pol II generates the paused elongation complex (PEC) (Wada et al., 1998; Yamaguchi et al., 1999; Narita et al., 2003; Landick, 2006; Vos et al., 2018a). Here, NELF is a key player since it both occludes the Pol II trigger loop, which is essential for translocation of the DNA-RNA hybrid, and the binding of transcription elongation factor IIS (TFIIS), known to promote pause release in those regions where RNA backtracking occurs (Reines et al., 1992; Adelman et al., 2005; Nechaev et al., 2010). Pol II pause release takes place by the recruitment of the positive transcription elongation factor b (P-TEFb) which harbours the cyclin-dependent kinase 9 (CDK9) (Marshall and Price, 1995; Wei et al., 1998). This kinase phosphorylates NELF, thereby promoting its release from Pol II, which effectively brings about the completion and release of promoter-proximal pausing (Marshall et al., 1996; Fujinaga et al., 2004; Yamada et al., 2006). This post-translational modification indirectly facilitates the maintenance of DSIF (Spt4/5 in yeast), recruitment of the elongation factor PAF1 complex (PAF) and SPT6, and stimulatory factor RTF1 (Vos et al., 2018b; Vos et al., 2020). At this stage, Pol II enters processive elongation, during which it can synthesise RNA at rates of ~2 kb/min on average.

Multiple co-transcriptional RNA processes occur in the cell nucleus such as 5’ end pre-mRNA capping, splicing, and 3’ end processing (Neugebauer, 2002). When the RNA chain is 20-30 nt-long between Pol II initiation and pausing, the 5’ end that comes out of the Pol II exit tunnel needs to be capped to stabilise the pre-mRNA molecule (Shatkin, 1976; Furuichi et al., 1977; Rasmussen and Lis, 1993; Shuman, 2001; Cowling, 2009; Ramanathan et al., 2016). Three capping enzymes catalyse the addition of a guanosine 5’-triphosphate (GTP) molecule to the very first ribonucleotide of the RNA, an essential process in protecting the RNA from

exonuclease activity and essential for mRNA cytoplasmic export (Gorlich et al., 1996; Visa et al., 1996; Lewis and Izaurralde, 1997). For pre-mRNA maturation into mRNA, non-coding intronic regions must be spliced out so that coding exons can be joined together as a second step in preparing a mature mRNA for translation, process catalysed by the spliceosome (Wilkinson et al., 2020). However, the beginning of this process occurs predominantly in a co-transcriptional manner (Wuarin and Schibler, 1994; Tennyson et al., 1995; Kotovic et al., 2003; Bird et al., 2004; Lacadie and Rosbash, 2005; Listerman et al., 2006; Wallace and Beggs, 2017). In this sense, pioneering structural work has elucidated how the first step in splicing is coupled to transcription, where the U1 small nuclear ribonucleoprotein (U1 snRNP) recognises and retains the 5' splice site close to the Pol II exit tunnel (Zhang et al., 2021a). While the subsequent steps of the process are unclear, this model suggests that co-transcriptional splicing occurs with the formation of an intron loop that grows as Pol II transcribes, facilitating the recognition of downstream splicing motifs.

Another important co-transcriptional process involved at different stages of transcription is the post-translational modification of the C-terminal domain of the Pol II RPB1 subunit (Pol II CTD), such as proline isomerisation, methylation of non-canonical lysine and arginine, and phosphorylation (Bird et al., 2004; Buratowski, 2009; Hsin and Manley, 2012; Eick and Geyer, 2013; Zaborowska et al., 2016; Harlen and Churchman, 2017). The Pol II CTD is a largely disordered region composed of Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 ($Y_1S_2P_3T_4S_5P_6S_7$) heptapeptide repeats, with 26 and 52 repeats in yeast and human, respectively. Specifically, phosphorylation has been shown to be a switch-like modification that helps in the recruitment and release of numerous transcription factors, coordinating the cycling of Pol II through different stages of transcription. For example, the cycling-dependent kinase 7 (CDK7) and 8 (CDK8) are part of the transcription factor IIH (TFIIH) and the Mediator co-activator complex, respectively (Ebmeier et al., 2017; Fant and Taatjes, 2019; Fisher, 2019; Luyties and Taatjes, 2022; Richter et al., 2022). These kinases catalyse the phosphorylation of Pol II CTD Ser5 but have shown to mediate different functions in transcription initiation. Whereas CDK7 is involved in promoter escape and recruitment of the capping enzymes (Fabrega et al., 2003; Sogaard and Svejstrup, 2007), CDK8 can act both as a transcriptional

activator and repressor by modulating its association with the Mediator complex (Hengartner et al., 1998; Liu et al., 2004; Osman et al., 2021). On the other hand, CDK9 is part of the P-TEFb elongation complex which orchestrates Pol II pause release by phosphorylating Ser2, pausing factor NELF and elongation factors (Marshall and Price, 1995; Wei et al., 1998; Fujinaga et al., 2004; Peterlin and Price, 2006; Yamada et al., 2006; Viladevall et al., 2009).

As a last step in the transcription cycle, Pol II must dissociate and get recycled so that the mRNA can be further processed in downstream events. At this stage, Pol II undergoes what is called 3'-end processing of the transcript, where the final form of the RNA is matured so that it can get exported out of the nucleus (Proudfoot, 1989; Proudfoot, 1996; Proudfoot, 2011, 2016). Nowadays, the consensus of how Pol II terminates and recycles is a unification of the allosteric and torpedo models (Eaton et al., 2020; Eaton and West, 2020). Here, a plethora of termination factors are first recruited upon the recognition of the polyadenylation signal (PAS) that is transcribed by Pol II (Mandel et al., 2006; Shi et al., 2009; Chan et al., 2014; Schonemann et al., 2014; Clerici et al., 2017). The multi-subunit cleavage and polyadenylation complex (CPA) recognises the PAS signal, stimulates its cleavage and, in turn, catalyses the addition of a long stretch of adenine monophosphates (poly(A) tail). The addition of the poly(A) tail stabilises further the transcript and constitutes, therefore, the last maturation step of the mRNA product. In a second step, the RNA that had been transcribed beyond the PAS needs to be degraded and Pol II dissociated from the DNA template. The CPA complex harbours the protein phosphatase 1 (PP1) which dephosphorylates elongation factors, promoting their dissociation and eventually the transcription deceleration of Pol II (Kecman et al., 2018; Parua et al., 2018; Cortazar et al., 2019). This creates the perfect environment for XRN2, a 5'-3' exonuclease, to degrade uncapped RNA, collide with Pol II, promote its release from the DNA template and ultimately its recycling (West et al., 2004; Brannan et al., 2012; Eaton et al., 2018; Eaton and West, 2018). As an extra layer of regulation, in certain cellular contexts Pol II needs to dissociate at the pausing step on snRNA, non-coding and protein-coding genes in a process called pre-mature termination or transcriptional attenuation (Baillat et al., 2005; Gardini et al., 2014; Lai et al., 2015; Skaar et al., 2015; Elrod et al., 2019; Kamieniarz-Gdula and Proudfoot, 2019; Rubtsova et al., 2019; Tatomer et al., 2019; Beckedorff et al.,

2020; Lykke-Andersen et al., 2021). This process is regulated by the recruitment of the ~1.5 MDa 14-subunit Integrator complex which interacts with Pol II, DSIF and NELF (Stadelmayer et al., 2014; Yamamoto et al., 2014; Fianu et al., 2021). First, Integrator recruits the protein phosphatase 2A (PP2A) to dephosphorylate pausing factors for their retention, counteracting the activity of P-TEFb (Stadelmayer et al., 2014; Huang et al., 2020; Zheng et al., 2020; Fianu et al., 2021). Then, the cleavage module of Integrator, formed by INTS11-INTS9-INTS4, cleaves the RNA owing to its endonuclease activity which leads to Pol II pre-mature termination (Chen et al., 2012; Albrecht et al., 2018; Fianu et al., 2021; Pfleiderer and Galej, 2021).

1.8 RNA polymerase II transcription initiation: a functional perspective

Transcription initiation of protein-coding genes in eukaryotes is a tightly regulated process in which Pol II must be timely recruited to the right gene promoter. Expectedly, kicking off transcription of a gene must transit through distinctly regulated steps: from promoter recognition to promoter escape (**Figure 4**) (Sainsbury et al., 2015; Hantsche and Cramer, 2016).

1.8.1 Promoter recognition

At the heart of transcription initiation, the cell needs to ensure that Pol II assembles efficiently at a given gene, before transcribing the DNA information into mRNA. For effective Pol II recruitment, a combination of proximal and distal enhancer elements, core promoter motifs and, therefore, activators, coactivators, and general transcription factors is employed. For promoter recognition, the recruitment of the TATA-box binding protein (TBP) is a key step in the nucleation of the Pol II pre-initiation complex (PIC) (Matsui et al., 1980; Reinberg et al., 1987; Van Dyke et al., 1988; Buratowski et al., 1989; Chasman et al., 1993; Zawel and Reinberg, 1993; Roeder, 1996; Thomas and Chiang, 2006; Grunberg and Hahn, 2013; Levine et al., 2014). TBP is a saddle-shaped single-subunit protein which binds TATA or TATA-like consensus sequences, generally found ~30 bp upstream of the TSS in humans, and induces a ~90º kink in the promoter DNA (Nikolov et al., 1992; Chasman et al., 1993; Kim et al., 1993a; Kim et al., 1993b; Geiger et al., 1996; Tan et al., 1996; Bleichenbacher et al., 2003; Ponjavic et al., 2006). However, most eukaryotic promoters do not contain a TATA-box element, but rather CpG islands, which raises

the question what the cue for TSS selection is (Carninci et al., 2006; Sandelin et al., 2007; Yang et al., 2007; Vo Ngoc et al., 2017a; Vo Ngoc et al., 2017b; Haberle and Stark, 2018). The transcription factor IID (TFIID) is a large ~1.3 MDa complex composed by 13-14 TBP-associated factors (TAFs) and TBP itself, which forms a trilobular three-dimensional structure that shows extensive flexibility throughout different stages of promoter recognition (Dynlacht et al., 1991; Andel et al., 1999; Brand et al., 1999; Leurent et al., 2002; Sanders et al., 2002a; Tora, 2002; Leurent et al., 2004; Grob et al., 2006; Papai et al., 2009; Bieniossek et al., 2013; Cianfrocco et al., 2013; Cianfrocco and Nogales, 2013; Louder et al., 2016; Nogales et al., 2017b; Nogales et al., 2017c; Kolesnikova et al., 2018; Patel et al., 2018). This multi-subunit complex has been shown to deliver TBP to the gene promoter, where DNA promoter deposition can be direct or stepwise depending on the promoter architecture (Louder et al., 2016; Patel et al., 2018; Chen et al., 2021a). Most promoters, which lack either downstream elements recognised by TFIID and/or upstream elements (e.g. TATA box) bound by TBP, directly deposit promoter DNA in the Pol II cleft. In this type of promoters, the spatial arrangement of TFIID serves a molecular ruler, however, in less than ~10% of promoters, which present both upstream and downstream elements, these DNA sequences constitute the molecular ruler, favouring a stepwise deposition.

In addition, even though yeast TBP has been shown to be essential for basal transcription *in vitro* and *in vivo* of TATA-box-containing promoters (Hahn et al., 1989; Horikoshi et al., 1989; Goodrich et al., 1996; Roeder, 1996; Kuras et al., 2000), in TFIID-depleted yeast nuclear extracts, the presence of TFIID enhances transcription efficiency compared to the addition of only TBP (Sanders et al., 2002b). This observation is consistent with the role of TFIID in promoter recognition, but also with biochemical and mass spectrometry data that suggests TFIID is also a coactivator which mediates interactions with transcriptional activators (Thut et al., 1995; Rojo-Niersbach et al., 1999; Munz et al., 2003; Li et al., 2004; Lively et al., 2004; Hilton et al., 2005; Wang et al., 2007; Liu et al., 2009; Papai et al., 2010; Goos et al., 2022).

1.8.2 Upstream complex assembly

Despite not being an essential GTF, the transcription factor IIA (TFIIA) binds to the convex surface of TBP, N-terminal region of TAF5 and N-terminal helix of TAF4 which ultimately stabilises the TBP-DNA complex and, on the other hand, situates the lobe A of TFIID in post-translocated state, which had harboured TBP before promoter recognition (Cortes et al., 1992; Imbalzano et al., 1994; Geiger et al., 1996; Tan et al., 1996; Cianfrocco et al., 2013; Louder et al., 2016; Patel et al., 2018; Chen et al., 2021a).

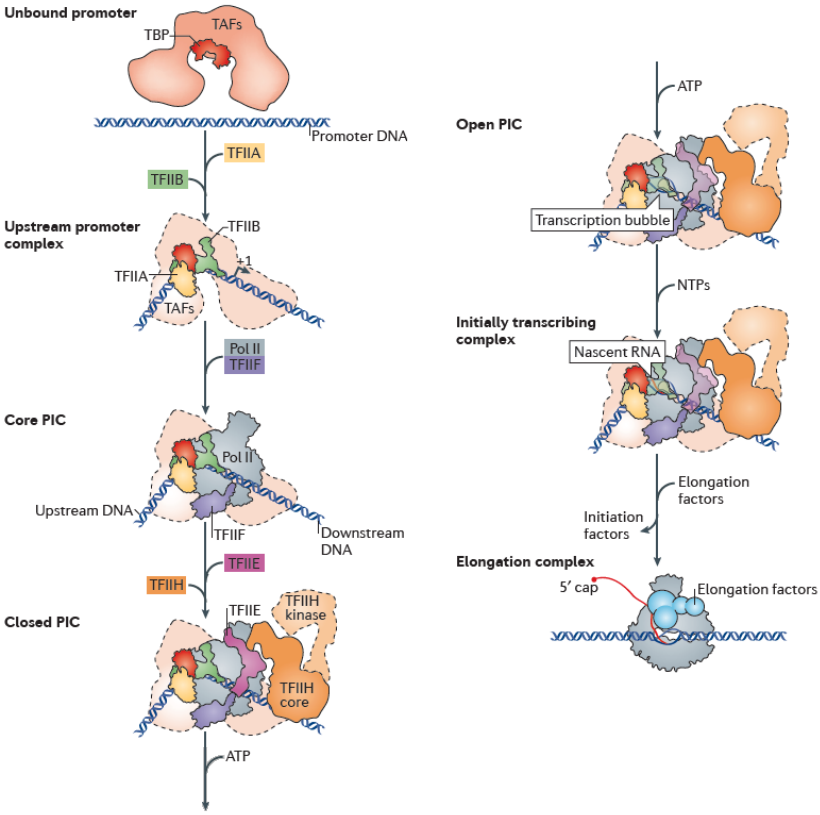


Figure 4. Pol II transcription initiation at a glance.

Assembly of eukaryotic PIC on promoter DNA and initiation of Pol II transcription is depicted on the diagram. TFIID or its TBP subunit (light and dark red, respectively) complex recognise the promoter consensus sequence, cause a 90º kink on DNA which is stabilised thereafter by TFIIA (yellow) and TFIIB (green). The resulting upstream promoter complex recruits TFIIF (purple) and Pol II (grey) to assemble the core PIC. Subsequent binding of TFIIE (magenta) and TFIIH (dark orange) finalises the formation of a productive PIC on a closed promoter DNA. In the presence of ATP, DNA can be opened, and pre-mRNA synthesis can begin. Finally, dissociation of initiation factors and binding of elongation ones (blue) can give rise to the

beginning of elongation. TAFs, TBP-associated factors. Modified after Sainsbury et al. (2015).

Subsequently, binding of transcription factor IIB (TFIIB) engulfs a C-terminal beta sheet of TBP and binds DNA sequences flanking the promoter motif bound by TBP, which stabilises the DNA kink created by TBP and confers transcription directionality (Van Dyke et al., 1988; Sawadogo and Sentenac, 1990; Buratowski and Zhou, 1993; Imbalzano et al., 1994; Lagrange et al., 1998; Littlefield et al., 1999; Tsai and Sigler, 2000; Zhao and Herr, 2002; Deng and Roberts, 2005). Whereas the C-terminal region of TFIIB, the B-core or cyclins domain, stabilises TBP-DNA binding and completes the formation of the upstream initiation complex assembly (TBP-DNA-TFIIA-TFIIB), together with its N-terminal part, the B-ribbon domain, contact Pol II (Bushnell et al., 2004; Kostrewa et al., 2009; Liu et al., 2010; Sainsbury et al., 2013). Not only do these domains interact with Pol II, but also does a long linker region connecting the B-core and B-ribbon domains does and, therefore, TFIIB represents an essential factor in the recruitment of Pol II to promoter DNA. Importantly, this linker region threads into the Pol II cleft and is divided in two subdomains: the B-linker and B-reader elements. The B-linker helps to keep DNA opened in the Pol II cleft by contacting the rewound upstream DNA, whereas the B-reader domain recognises the TSS, interacts with the DNA template strand hybridised with the growing RNA, and ultimately aids in the subsequent release of RNA from the DNA-RNA hybrid by directing it into the exit tunnel of Pol II.

1.8.3 Core pre-initiation complex assembly

Pol II is co-recruited with the transcription factor IIF (TFIIF) which initiates the formation of the core RNA polymerase II pre-initiation complex (cPIC) (Flores et al., 1991; Sainsbury et al., 2015; Nogales et al., 2017a). TFIIF is composed of 2 and 3 subunits in human and yeast, respectively, where the third subunit in yeast is not essential (Tfg3). Both N-terminal regions of TFIIF subunit α (Tfg1 in yeast) and β (Tfg2 in yeast) contain dimerization regions that dimerise with each other on the jaw of Pol II to constitute the dimerization domain of TFIIF, contributing to the stabilisation of Pol II (Chen et al., 2010a; Eichner et al., 2010; Fishburn and Hahn, 2012; He et al., 2013; Murakami et al., 2015a; Plaschka et al., 2015; He et al., 2016; Plaschka et al., 2016). The largest subunit TFIIF-α forms two α-helices in its C-

terminus, also called TFIIF charged region, which contact DNA along the Pol II cleft, aids in TSS selection and stabilises the melting of the first few DNA base pairs (Ghazy et al., 2004). On the other hand, the C-terminal linker region of TFIIF-β threads along the Pol II protrusion and contacts TFIIB cyclin domains, which further stabilises the binding of TFIIB on promoter DNA (Cabart et al., 2011). This linker region extends into the C-terminus of TFIIF-β, comprised by the winged helix domain (WH) which binds upstream promoter DNA. This binding event is important for locking and preparing upstream promoter DNA for subsequent DNA translocation into Pol II cleft and opening (Tan et al., 1994; He et al., 2016; Plaschka et al., 2016).

The assembly of the core PIC is only finalised by the recruitment of the transcription factor IIE (TFIIE) which primes DNA for opening (Forget et al., 2004; Miller and Hahn, 2006). TFIIE locates between the Pol II clamp and stalk and is composed of 2 subunits, TFIIE subunit α and β (Tfa1 and Tfa2 in yeast, respectively) which are involved in DNA stabilisation and the recruitment of other GTFs, respectively (Kuldell and Buratowski, 1997; Chen et al., 2007; He et al., 2016; Plaschka et al., 2016). The winged helix domain 1 (WH1) of TFIIE-β interacts with the WH domain of TFIIF-β and, together with the E-wing and extended winged helix (eWH) domains of TFIIE-α (which bind to the Pol II clamp and the upstream edge of the transcription bubble), they readily lock upstream DNA for promoter melting. Interestingly, a helix, located N-terminally of the WH1 of TFIIE-β, contacts TBP, and contributes to the anchor of TFIIE to promoter DNA, and the stabilisation of the upstream complex (Maxon et al., 1994; Schilbach et al., 2021). The WH2 and C-terminal region of TFIIE-β, called the E-tether, represent the core domains attaching the beta subunit to TFIIE-α. While WH2 binds to the eWH domain of the alpha subunit, the E-tether interacts with the E-linker region of TFIIE-α, which is a helical region of TFIIE-α linking its N-terminus (mainly composed by the E-wing and eWH) and the C-terminus (comprised by the E-ribbon domain) (Schilbach et al., 2017; Schilbach et al., 2021).

1.8.4 Pre-initiation complex assembly

TFIIE has a dual role in transcription initiation: its beta subunit contacts the core PIC formed by TBP-DNA-TFIIA-TFIIB-TFIIF while its alpha subunit recruits the

transcription factor IIH (TFIIH) to this complex (Cortes et al., 1992; Flores et al., 1992; Maxon et al., 1994; Holstege et al., 1995; Kuldell and Buratowski, 1997; Okuda et al., 2008). It has been shown that the interaction of E-linker helices of TFIIE-α with the RING domain of the TFIIH kinase subunit MAT1 (Tfb3 in yeast) is conserved from yeast to human (He et al., 2016; Schilbach et al., 2017; Aibara et al., 2021; Schilbach et al., 2021). However, the situation in yeast differs to the human system and structural studies have elucidated more extensive TFIIE-TFIIH interactions in *S. cerevisiae* (Schilbach et al., 2017; Schilbach et al., 2021). For example, the RING domain of MAT1 not only contacts the E-linker helices, but also RPB7. In addition, the flexibly connected C-terminal helical regions E-dock, E-bridge, and E-floater have been visualised in close contact with TFIIH in yeast. The E-dock helix, as the name suggests, allows the docking of the plekstrin homology domain (PHD) of Tfb1 (p62 in humans) to the eWH domain, located adjacent to the Pol II clamp. The E-bridge anchors the DOS-like domain 2 (BSD2) of Tfb1 (p62 in humans) to the lobe 2 of the TFIIH ATPase subunit Ssl2 (XPB in humans). Finally, the E-floater binds to the BSD1 domain of Tfb1 (p62 in humans). Therefore, not only TFIIE completes the formation of the core PIC, but bridges the recruitment of TFIIH to Pol II, thus assembling the RNA polymerase II pre-initiation complex (PIC) with closed promoter DNA (closed PIC, CC) (Sainsbury et al., 2015).

TFIIH is a 10-subunit complex with diverse activities due to its translocase XPB subunit (Ssl2 in yeast), helicase XPD subunit (Rad3 in yeast) and CDK-activating kinase module (CAK) formed by CDK7-MAT1-CYCH (Kin28-Tfb3-Ccl1 in yeast) (Conaway and Conaway, 1989; Weber et al., 1990; Schaeffer et al., 1993; Roy et al., 1994; Schaeffer et al., 1994; Adamczewski et al., 1996; Feaver et al., 1997; Marinoni et al., 1997a; Marinoni et al., 1997b; Takagi et al., 2003; Giglia-Mari et al., 2004; Ranish et al., 2004). The dissociable CAK module allows TFIIH to not only work in transcription initiation, but also in nucleotide excision repair which will not be discussed herein (Svejstrup et al., 1995; Coin et al., 2008; Kokic et al., 2019).

1.8.5 Promoter DNA melting

Groundbreaking studies have described in detail for linear DNA how TFIIH is responsible for opening DNA by moving it towards the active site of Pol II as a prelude for transcription initiation (Wang et al., 1992; Giardina and Lis, 1993;

Ohkuma and Roeder, 1994; Holstege et al., 1996; Kim et al., 2000; Lin et al., 2005; Miller and Hahn, 2006; Grunberg et al., 2012; He et al., 2013; Fishburn et al., 2015; Murakami et al., 2015a; He et al., 2016; Schilbach et al., 2017; Tomko et al., 2017; Aibara et al., 2021; Schilbach et al., 2021). Here, the upstream complex keeps the DNA rotationally fixed, while the binding of both XPB ATPase lobes to DNA (~20 bp downstream of the TSS) causes a local twist in the base-pairing of the DNA template strand. This binding event also distorts the promoter DNA region where DNA will be initially opened (initially melted region, IMR) (Dienemann et al., 2019). Upon binding of adenosine triphosphate (ATP), XPB switches into a post-translocated state where this twist is propagated upstream, and the DNA returns to its B-DNA conformation. The release of ADP and Pi resets the ATPase subunits into a pre-translocated state which causes the downstream DNA to be twisted again and an overall upstream DNA shift by 1 bp. In this translocation process, a two-step DNA-opening mechanism has been suggested (Holstege et al., 1996; Aibara et al., 2021; Schilbach et al., 2021). TFIIH first creates a DNA bubble of half a turn of DNA (~6 bp) ~30 bp downstream of the TATA box, which has been coined as an intermediate complex (IC) (Tomko et al., 2017; Schilbach et al., 2021). Pre-opening half a turn of DNA has been shown to make transcription TFIIH-independent, which led to the proposal of the second step of this process. In this last step, the instability of the IC might be enough to cause spontaneous promoter opening and the transition to open promoter DNA be completed (open PIC, OC) (Sainsbury et al., 2015).

Unlike yeast, DNA opening in humans induces the closure of the Pol II clamp which is required to accommodate DNA in its opened state (He et al., 2016; Plaschka et al., 2016; Hantsche and Cramer, 2017; Schilbach et al., 2017; Li et al., 2019; Aibara et al., 2021). This also stimulates the dissociation of MAT1 from TFIIE, which argues in favour of the proposed two-step opening mechanism (Aibara et al., 2021). This DNA opening mechanism, however, has been suggested for yeast as well. Therefore, the differences observed for the clamp closure, and MAT1 dissociation, likely reflect that yeast Pol II scans DNA for the TSS before promoter opening, whereas human Pol II undergoes promoter-proximal pausing (Rougvie and Lis, 1988; Giardina and Lis, 1993; Core and Lis, 2008; Murakami et al., 2015b; Fishburn et al., 2016; Qiu et al., 2020). In addition, for the human PIC, ordering of the B-reader and linker elements has only been observed in the OC (He et al., 2016;

Aibara et al., 2021). These elements have been shown crucial in DNA opening for archaeal organisms, which lack TFIIH, explaining why TFIIH-independent spontaneous promoter opening does not occur in human, unlike yeast (Kostrewa et al., 2009; Plaschka et al., 2016; Schilbach et al., 2017; Dienemann et al., 2019). Interestingly, relying on the ATP-dependent translocase TFIIH makes Pol II promoter opening unique. Unlike all other Pols, which open DNA in an upstream-to-downstream fashion (Naryshkin et al., 2000; Feklistov and Darst, 2011; Alekseev et al., 2017; Abascal-Palacios et al., 2018; Glyde et al., 2018; Han et al., 2018; Vorlander et al., 2018; Sadian et al., 2019; Chen et al., 2020; Pilsl and Engel, 2020), TFIIH-dependent Pol II opening does it in the opposite manner: in a downstream-to-upstream fashion (He et al., 2016; Aibara et al., 2021; Schilbach et al., 2021).

1.8.6 Initial RNA synthesis and promoter escape

After initial promoter melting and DNA opening, the open PIC complex transitions into the formation of the initially transcribing complex (ITC) (Luse and Jacob, 1987; Pal et al., 2005). At this stage, the PIC remains intact and starts the nucleotide addition cycle to produce the initial RNA chain (He et al., 2013; Sainsbury et al., 2013; He et al., 2016). TFIIB plays an important role in the stabilisation of the ITC complex since the B-ribbon interacts with the unwound DNA template strand and has been associated with the separation of the DNA-RNA hybrid, leading to the positioning of the growing RNA into the Pol II exit tunnel (Bushnell et al., 2004; Zhang and Dietrich, 2005; Liu et al., 2010; Kaplan et al., 2012; Sainsbury et al., 2013). The disassembly of the ITC would signify the start of the transition between transcription initiation and elongation (Goodrich and Tjian, 1994; Sainsbury et al., 2015). Although the mechanism of this transition is unclear, biochemical studies have demonstrated that an RNA beyond ~12-13 nt would clash with the B-ribbon domain, suggesting this could be one of the initial catalytic steps in the replacement of GTFs with elongation factors (Cabart et al., 2011). Additionally, the TFIIH kinase CDK7 plays a role in this process by phosphorylating the Pol II CTD residues Ser5 and Ser7, as well as GTFs like TFIIE and elongation factors. For example, cellular and structural studies have shown that phosphorylation of the Pol II CTD would induce the dissociation of the co-activator Mediator in yeast, leading to a

destabilisation of the PIC (Wong et al., 2014; Schilbach et al., 2017; Schilbach et al., 2023).

A model, where a PIC becomes unstable, and the RNA length would eject TFIIB from the promoter, suggests that Pol II would no longer be attached to the upstream complex. This would leave behind a TFIID/TBP-TFIIA-promoter DNA complex, which has been proposed as a scaffold that would facilitate new rounds of transcription – the re-initiation complex (Zawel et al., 1995; Yudkovsky et al., 2000). Whereas it is known that TFIIF remains associated with Pol II stimulating elongation (Zawel et al., 1995; Elmendorf et al., 2001; Zhang et al., 2005; Cabart et al., 2011), the post-initiation fate of TFIIE and TFIIH is unclear. A recent study suggested that the dissociation of these GTFs might happen modularly in two different steps: TFIIE-α and CAK are released first from Pol II, while TFIIE-β and TFIIH core module dissociate last (Compe et al., 2019).

1.8.7 Activating transcription: the Mediator complex

The Mediator is a ~1.5 MDa complex conserved in eukaryotes, with 25 subunits in yeast and 30 in human, that controls the functioning of RNA polymerase II by stabilising PIC formation and stimulating TFIIH kinase activity (Kim et al., 1994; Jiang et al., 1998; Meyer et al., 2010; Plaschka et al., 2015; Schilbach et al., 2017; Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021). In its co-activator role, it has been suggested that this complex mediates enhancer-promoter interactions (Allen and Taatjes, 2015; Soutourina, 2018; Osman and Cramer, 2020; Richter et al., 2022). Enhancers are generally located kilobases apart from the promoter regions for which they mediate their activation (Lai et al., 2013; Hsieh et al., 2014; Aranda-Orgilles et al., 2016). Here, transcriptional activators first bind enhancer regions by recognising their cognate DNA motifs. CCCTC-binding factor (CTCF) and cohesin then establish enhancer-promoter boundaries and loop extensive DNA regions in the genome to bring them in spatial proximity (Finn and Misteli, 2019; Szabo et al., 2019; Misteli, 2020; Oudelaar and Higgs, 2021). Genome architecture can therefore aid the Mediator complex in relaying activation signals to Pol II by controlling the stability of PIC formation and thus transcription initiation efficiency (Richter et al., 2022). Although the mechanism is yet unclear, different variants for these models have been proposed, where ultimately a TF-enhancer-

bound Mediator confers the activation signals to a promoter-bound Pol II. Studies in yeast suggest that Mediator binds transcription factors on upstream enhancer regions, recruits and assembles with Pol II on them, and through gene-looping can deliver Pol II to the appropriate gene promoter, likely dictated by the assembly of the PIC upstream complex (El Khattabi et al., 2019; Baek et al., 2021; Sun et al., 2021). Another model, although compatible, proposes that Mediator assembles exclusively with TFs on distal enhancers, whereas Pol II is directly recruited to the gene promoter. Thus, Mediator and Pol II would assemble independently and come into contact through gene-looping maintained by CTCF and cohesin (Kagey et al., 2010; Jeronimo and Robert, 2014; Wong et al., 2014; Jeronimo et al., 2016). This crosstalk would allow the recruitment of Mediator to the gene promoter, where Mediator and/or the enhancer would be bound with TFs (Xiao et al., 2021). This aims at keeping the local concentration of the coactivator high at gene promoters or, in other words, at facilitating the formation of a stabler productive PIC (Palacio and Taatjes, 2022; Richter et al., 2022).

In addition, the Mediator complex bears a transcriptional kinase module called CDK8 kinase module (CKM) (Taatjes et al., 2002; Andrau et al., 2006). The association of CKM with Mediator has been shown to sterically impede the binding of this co-activator with Pol II, which constitutes an additional level of regulation in Mediator-dependent activation of transcription (Elmlund et al., 2006; Tsai et al., 2013). While *in vivo* experiments have found that this kinase is required for transcriptional activation during a heat-shock response, biochemical studies have elucidated that CDK8-dependent phosphorylation of Mediator might regulate the dissociation of CKM (Osman et al., 2021). Altogether, it is compelling to assume that, upon activation stimuli, Mediator regulates its own assembly on enhancers to tune PIC formation and, therefore, Pol II transcription.

1.9 Structural overview of transcription initiation complexes

The determination of the 12-subunit Pol II structure from *S. cerevisiae* (Cramer et al., 2001; Armache et al., 2003; Bushnell and Kornberg, 2003; Armache et al., 2005; Meyer et al., 2006; Meyer et al., 2009) enabled the structural characterisation of transcription initiation-related complexes. Subsequently, the determination of Pol II complexes bound to short RNAs (Westover et al., 2004; Cheung et al., 2011; Liu et

al., 2011), and/or TFIIB (Bushnell et al., 2004; Kostrewa et al., 2009; Liu et al., 2010; Sainsbury et al., 2013), paved the way for beginning to understand how Pol II transcription starts. However, Pol II recruitment and transcription initiation can only be understood with the assembly of GTFs on the gene promoter that constitutes the core PIC. Several groups elucidated structurally how the yeast and human basal machinery comprised by TBP, TFIIA, TFIIB, Pol II, TFIIF, and TFIIE assembles on promoter DNA and prepares it for opening (He et al., 2013; Murakami et al., 2015a; He et al., 2016; Plaschka et al., 2016; Dienemann et al., 2019). Together with studies that investigated the architecture of Pol II-Mediator complexes (Bernecky et al., 2011; Cai et al., 2012; Robinson et al., 2012), yeast TFIIH-containing PIC-Mediator structures could then be determined (Robinson et al., 2016; Schilbach et al., 2017; Tsai et al., 2017). Additionally, further understanding could be recently obtained into DNA promoter opening from high-resolution structural studies of yeast and human PIC (Aibara et al., 2021; Schilbach et al., 2021).

The low-resolution structural studies did not allow to build a tail-containing atomic model of Mediator in complex with Pol II. Recent cryo-EM investigations of Mediator bound to Pol II in yeast (Zhang et al., 2021b; Gorbea Colon et al., 2023), and PIC-Mediator complexes in human (Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021), allowed to build complete atomic models of yeast and human Pol II-bound Mediator, respectively. This groundbreaking work gave insights into how transcription begins and functions, but did not explain how other co-activators, such as SAGA or TFIID, might work with GTFs and Pol II to orchestrate transcription initiation. Different laboratories could determine the structure of SAGA (Brand et al., 1999; Wu et al., 2004; Setiaputra et al., 2015; Papai et al., 2020; Wang et al., 2020a; Herbst et al., 2021) and TFIID alone (Dynlacht et al., 1991; Andel et al., 1999; Brand et al., 1999; Leurent et al., 2002; Sanders et al., 2002a; Tora, 2002; Leurent et al., 2004; Grob et al., 2006; Papai et al., 2009; Bieniossek et al., 2013), and TFIID in complex with promoter DNA (Cianfrocco et al., 2013; Cianfrocco and Nogales, 2013; Louder et al., 2016; Kolesnikova et al., 2018; Patel et al., 2018). Even though we still lack structural information on PIC-SAGA complexes, one cryo-EM study reported the structures of TFIID in complex with the remaining GTFs and Pol II on different gene promoters (Chen et al., 2021a). This study explains decades-long research on how TBP delivery by TFIID to the promoter nucleates PIC assembly

(Matsui et al., 1980; Reinberg et al., 1987; Van Dyke et al., 1988; Buratowski et al., 1989; Chasman et al., 1993; Zawel and Reinberg, 1993; Roeder, 1996; Thomas and Chiang, 2006; Grunberg and Hahn, 2013; Levine et al., 2014).

To date, most structural information has been obtained using naked DNA templates. In cells, however, DNA is wrapped around an octamer of histones which forms the nucleosome – the fundamental unit that makes up chromatin. The first structural studies of PICs in complex with a +1 nucleosome, located in a position representative of active genes, were recently reported (Chen et al., 2022; Wang et al., 2022). By positioning this nucleosome ~40-50 bp from the TSS, human TFIIH and Mediator formed weak interactions with modified and unmodified nucleosomes (Chen et al., 2022). For the *S. cerevisiae*, a high-resolution cryo-EM structure of a PIC-nucleosome complex was determined. The authors found that TFIIH engages with the +1 nucleosome, and that the ATP-dependent activity of TFIIH translocase subunit Ssl2 (XPB in human) induces the rotation of the +1 nucleosome, leading to a partial unwrapping of the nucleosome (Wang et al., 2022).

1.10 Scope of this work

Transcription can only be fully understood if we analyse its function considering chromatin organisation. We begin to structurally grasp now how nucleosomes constitute a regulatory mechanism for Pol II-RNA elongation. Even though transcription initiation is the initial checkpoint of this process, most structural studies have focused on understanding the process by using chromatin-free DNA. Only recently, as mentioned above, the first structural insights were gained into the interactions established by yeast and human PICs with the +1 nucleosome. Nevertheless, these studies did not provide specific mechanisms by which the position of a promoter-proximal +1 nucleosome regulates transcription. These mechanisms must be investigated as compaction of chromatin regulates gene activity, known to be critical in biological processes such as mitosis and heterochromatin maintenance.

To investigate the regulatory role of the +1 nucleosome in transcription initiation, we first performed an *in vivo* analysis of transient transcriptome (TT-seq) and micrococcal nuclease (MNase) sequencing data. This allowed us to correlate gene activity with the position of the +1 nucleosome. From this *in vivo* analysis, we could also infer the median distances of these promoter-proximal +1 nucleosomes to the TSS, which we used for all *in vitro* subsequent analyses. I performed *in vitro* transcription assays with highly purified recombinant GTFs and endogenous Pol II, and corroborated the *in vivo* analysis performed. Subsequently, I unravelled its mechanism by determining the structure of key PIC-nucleosome intermediate states by cryo-EM. By using cryo-EM, I had to first screen the TSS-nucleosome *in vivo* distances to find the most stable intermediates. I could then obtain high-resolution states with which I deciphered the mechanism for the observed effect.

2. Materials

Table 1. Reagents, resources, and data used or generated in this study. This table contains a list with the strains, chemicals, vectors, oligonucleotides, proteins, and other resources used for this dissertation.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E.coli BL21 CodonPlus (DE3) RIL	Agilent	Cat#230245
E. coli LOBSTR-BL21(DE3)-RIL	Kerafast	Cat# EC1002
E.coli DH10EMBacY	Geneva Biotech	N/A
E.coli XL-1 Blue	Agilent	Cat#200249
Biological *samples*		
Sus scrofa thymus	Locally sourced	N/A
Chemicals, peptides, and recombinant proteins		
Sus scrofa RNA polymerase II	Vos et al. (2018a)	N/A
Homo sapiens TBP	Aibara et al. (2021)	N/A
Homo sapiens TFIIB	Aibara et al. (2021)	N/A
Homo sapiens TFIIA	Aibara et al. (2021)	N/A
Homo sapiens TFIIF	Aibara et al. (2021)	N/A
Homo sapiens TFIIE	Aibara et al. (2021)	N/A
Homo sapiens TFIIH-core	This study	N/A
Homo sapiens CAK	Kokic et al. (2019)	N/A
Xenopus laevis histones H3, H4, H2A, H2B	Dyer et al. (2004)	N/A
8WG16 (αRPB1 CTD) antibody	Hu et al. (2006)	N/A
Glutaraldehyde 25%	EMS	Cat#16200
Poly(vinyl alcohol)	Sigma-Aldrich	Cat#P8136
RNasin Plus Ribonuclease Inhibitor	Promega	Cat#N2611
Proteinase K	New England Biolabs	Cat#P8107S
Invitrogen UltraPure 0.5M EDTA, pH 8.0	Thermo Fisher Scientific	Cat#15575020
SDS 10%	Thermo Fisher Scientific	Cat#AM9822
DNAseI (RNase-free)	New England Biolabs	Cat#M0303S
Invitrogen Ambion Sodium Acetate (3M), pH 5.5	Thermo Fisher Scientific	Cat#AM9740
NTP Set, 100 mM Solution	Thermo Fisher Scientific	Cat#R0481
SYBR™ Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific	Cat#S11494
BSA-Molecular Biology Grade	New England Biolabs	Cat#B9000S
Urea (RNase-free)	Panreac AppliChem	Cat#A1049
2x RNA Loading Dye	New England Biolabs	Cat#B0363S
DL-Dithiothreitol solution, 1M	Sigma-Aldrich	Cat#43816
40% Acrylamide/bis-acrylamide 19:1	Sigma-Aldrich	Cat#A9926
TRIS borate-EDTA buffer solution (10x)	Sigma-Aldrich	Cat#93290
Phusion DNA Polymerase	House sourced	N/A
Phusion® HF Buffer Pack	New England Biolabs	Cat#B0518S

dNTP Set, 100 mM Solutions	Thermo Fisher Scientific	Cat#R0186
DMSO	Sigma-Aldrich	Cat#D8418
Deposited data		
cPIC-nucleosome^{10W} cryo-EM globally refined map	This study	EMD-16335
Core PIC$^{cPIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16336
Nucleosome$^{cPIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16337
cPIC-nucleosome^{10W} cryo-EM composite map	This study	EMD-16338
cPIC-nucleosome^{10W} model	This study	PDB ID: 8BZ1
PIC-nucleosome^{10W} cryo-EM globally refined map	This study	EMD-16331
Core PIC$^{PIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16339
Nucleosome$^{PIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16340
TFIIH$^{PIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16342
CAK$^{PIC\text{-}Nuc10W}$ cryo-EM focused refined map	This study	EMD-16341
PIC-nucleosome^{10W} cryo-EM composite map	This study	EMD-16343
PIC-nucleosome^{10W} model	This study	PDB ID: 8BYQ
PIC-nucleosome^{18W} cryo-EM globally refined map	This study	EMD-16274
Core PIC$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16365
Nucleosome$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16366
TFIIH$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16367
XPB-containing TFIIH$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16367, add.
XPD-containing TFIIH$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16367, add.
CAK$^{PIC\text{-}Nuc18W}$ cryo-EM focused refined map	This study	EMD-16368
PIC-nucleosome^{18W} cryo-EM composite map	This study	EMD-16369
PIC-nucleosome^{18W} model	This study	PDB ID: 8BVW
Experimental models: cell lines		
Sf9 Cells	Thermo Fisher Scientific	Cat#11496015
High Five Cells	Thermo Fisher Scientific	Cat#B85502
Oligonucleotides		
Widom 601 template: 5' – ATC GGA TGT ATA TAT CTG ACA CGT GCC TGG AGA CTA GGG AGT AAT CCC CTT GGC GGT TAA AAC GCG GGG GAC AGC GCG TAC GTG CGT TTA AGC GGT GCT AGA GCT GTC TAC GAC CAA TTG AGC GGC CTC GGC ACC GGG ATT CTC GAT – 3'	This work	Integrated DNA Technologies
Widom 601 non-template 5' – ATC GAG AAT CCC GGT GCC GAG GCC GCT CAA TTG GTC GTA GAC AGC TCT AGC ACC GCT TAA ACG CAC GTA CGC GCT GTC CCC CGC GTT TTA ACC GCC AAG GGG ATT ACT CCC TAG TCT CCA GGC ACG TGT CAG ATA TAT ACA TCC GAT – 3'	This work	Integrated DNA Technologies
AdML0W cloning forward primer 5' – TCG AGG TAC CGG ATC CGA TAT CCG GGT GTT CCT GAA GGG GGG CTA TAA AAG GGG GTG GGG GCG CGT TCG TCC TCA ATC GAG AAT CCC GGT GCC GAG G – 3'	This work	Sigma-Aldrich

AdML10W cloning forward primer 5' – TCG AGG TAC CGG ATC CGA TAT CCG GGT GTT CCT GAA GGG GGG CTA TAA AAG GGG GTG GGG GCG CGT TCG TCC TCA CTC TCT TCC GAT CGA GAA TCC CGG TGC CGA GG – 3'	This work	Sigma-Aldrich
AdML18W cloning forward primer 5' – TCG AGG TAC CGG ATC CGA TAT CCG GGT GTT CCT GAA GGG GGG CTA TAA AAG GGG GTG GGG GCG CGT TCG TCC TCA CTC TCT TCC GCA TCG CTG ATC GAG AAT CCC GGT GCC GAG G – 3'	This work	Sigma-Aldrich
Widom 601 cloning reverse primer 5' – CGA AGA TCT GAT ATC ATC GGA TGT ATA TAT CTG ACA CGT GCC TGG AGAC – 3'	This work	Sigma-Aldrich
AdMLW PCR forward primer 5' – CGG GTG TTC CTG AAG GGG GGC TAT AAA AGG GGG TG – 3'	This work	Sigma-Aldrich
Widom 601 PCR reverse primer 5' – ATC GGA TGT ATA TAT CTG ACA CGT GCC TGG AGA CTA GGG AG – 3'	This work	Sigma-Aldrich
Recombinant DNA		
438A-hTBP	Aibara et al. (2021)	N/A
pOPINF-hTFIIB	Aibara et al. (2021)	N/A
438A-hTFIIA	Aibara et al. (2021)	N/A
pETDuet-1-hTFIIE	Aibara et al. (2021)	N/A
pAHS3C-hTFIIF	Aibara et al. (2021)	N/A
438C-XPD-p52-p34-p8-p62-p44-XPD (hTFIIH-core)	This study	N/A
438B-CCNH-CDK7-MAT1 (CAK)	Kokic et al. (2019)	N/A
pUC119-AdML0W	This study	N/A
pUC119-AdML10W	This study	N/A
pUC119-AdML18W	This study	N/A
Software and algorithms		
SerialEM 4.0	Mastronarde (2005)	https://bio3d.colorado.edu/SerialEM/#Source
cryoSPARC 3.2.0	Punjani et al. (2017)	https://cryosparc.com/
RELION 3.1	Scheres (2020); Zivanov et al. (2018)	https://github.com/3dem/relion
Warp 1.0.9	Tegunov and Cramer (2019)	http://www.warpem.com
PHENIX 1.19.2	Afonine et al. (2018)	http://www.phenix-online.org
PyMol 2.5.0	Schrödinger and Delano (2020)	http://www.pymol.org
UCSF Chimera	Pettersen et al. (2004)	https://www.cgl.ucsf.edu/chimera/
UCSF Chimera X-1.4	Goddard et al. (2018)	https://www.cgl.ucsf.edu/chimerax/
ISOLDE 1.3	Croll (2018)	https://isolde.cimr.cam.ac.uk/

Coot 0.9.6	Emsley et al. (2010)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ImageJ 2.1.0	Schindelin et al. (2012)	https://imagej.nih.gov/ij/index.html
Prism 9.1.0	GraphPad Software Inc (California, USA)	https://www.graphpad.com/
Other		
Glacios	Thermo Fisher Scientific	N/A
Falcon-III Direct Electron Detector	Thermo Fisher Scientific	N/A
Titan Krios G2	FEI/Thermo Fisher Scientific	N/A
QuantumLS energy filter	Gatan	N/A
K3 Summit Direct Electron Detector	Gatan	N/A
Typhoon™ 9500 FLA imager	GE Healthcare Life Sciences	N/A
BioComp Gradient Master 108	BioComp Instruments	N/A
Model 491 Prep Cell	Bio-Rad	Cat#1702927
Slide-A-Lyzer™ MINI Dialysis Devices (3.5 KDa MWCO)	Thermo Fisher Scientific	Cat# 69552
Slide-A-Lyzer™ MINI Dialysis Devices (20 KDa MWCO)	Thermo Fisher Scientific	Cat# 69590
Amicon Millipore 15 ml (50 KDa MWCO)	MERCK Milipore	Cat# UFC9050
Quantifoil™ R3.5/1, copper, mesh 200	Quantifoil	N/A

3. Methods

3.1 Nucleosome reconstitution

Histones preparation and nucleosome reconstitution was performed as described (Dyer et al., 2004; Wang et al., 2020b).

3.1.1 Histones expression

X. laevis histones sequences were inserted by restriction digest cloning in the expression vector pET3a (Wagner et al., 2020). These plasmids were transformed by heat-shock into a BL21 CodonPlus (DE3) RIL *E. coli* expression strain, using ampicillin-chloramphenicol resistance LB-agar plates. After transformation, precultures were grown at 37ºC, 180 rpm until an optical density at 600 nm wavelength (OD_{600}) of 0.4 was reached. A larger culture of 12 L was inoculated with the precultures, complemented with a final working concentration of 50 µg/ml ampicillin and 17 µg/ml chloramphenicol. Cultures were grown under the same conditions until an OD_{600} of 0.35 – 0.45 was reached. Since the expression of histones in these vectors are controlled by a T7 promoter, the cultures were induced with 0.5 mM IPTG for 3 hours at 37ºC, 140 rpm. Cells were harvested at 4ºC, 6000 rpm for 20 minutes (Sorvall Lynx 6000 Centrifuge (ThermoFisher Scientific), F9-6x1000 LEX rotor), resuspended in washing buffer (**Table 2**), flash-frozen and stored at -80ºC until use.

Table 2. H2A, H2B, H3 and H4 histones purification buffers.

	Washing buffer	Unfolding buffer	Urea dialysis buffer	Buffer A	Buffer B	Dialysis buffer
Tris-HCl	50 mM	50 mM	50 mM	50 mM	50 mM	-
NaCl	200 mM	200 mM	200 mM	200 mM	600 mM	-
Urea	-	-	8 M	8 M	8 M	-
Guanidinium-HCl	-	6 M	-	-	-	-
β-ME	1 mM	1 mM	2.5 mM	1 mM	1 mM	5 mM
PI	1x	1x	-	-	-	-

The final concentration used for every component of the buffer is shown in the table, while on the left are shown the stock components used for preparing these buffers. Tris: Tris(hydroxymethyl)aminomethane; NaCl: sodium chloride; β-ME: β-Mercaptoethanol; PI: protease inhibitors cocktail; HCl: hydrochloride acid. It should be noted that before adding salts or adjusting pH of the urea dialysis buffer, the urea stock solution had to be deionised with "Ion exchanger Amberlite® MB-3" (EMD Millipore Corporation).

3.1.2 Histones purification

All histones were purified following the same procedure described in this section. Bacterial cells were lysed by sonication on ice for 10 minutes with an amplitude of 40% (0.4 seconds on, 0.6 seconds off) by using the BRANSON Digital Sonifier® 450 (Branson Ultrasonic Corporation). To remove the cell debris, the sample was subsequently centrifuged at 15000 rpm, 4ºC for 20 minutes (Sorvall Lynx 6000 Centrifuge (ThermoFisher Scientific, A27-8x50 rotor). As histones accumulate in inclusion bodies, the pelleted fractions were manually washed using a Dounce tissue grinder (Sigma-Aldrich) with washing buffer on ice (**Table *2***), and centrifuged in three cycles. Prior to the first cycle, bovine pancreas 10 µg/mL DNase I (1 mg/ml stock, Sigma-Aldrich) and 2 mM $MgCl_2$ were added into the suspension to digest histone-unbound DNA. To extract the histones from inclusion bodies, the grinder, 2 mL of DMSO and 50 mL of unfolding buffer (UB) (**Table 2**) were employed at room temperature to dissolve and unfold the histones, contained in the pellet fraction. Once dissolved, inclusion bodies were further incubated at room temperature for 1 hour. The debris of inclusion bodies were then removed by spinning the solution at 25000 rpm, 4ºC for 20 minutes (Sorvall Lynx 6000 Centrifuge (ThermoFisher Scientific, A27-8x50 rotor). Histones, contained in the supernatant, were then subjected to dialysis at 4ºC against urea dialysis buffer (**Table 2**) three times in Spectra/Por®3 Dialysis Membrane MWCO 3.5 KDa (Spectrum Labs) to maintain the chaotropic agent, as well as decrease the pH of the buffer. The first step consists in dialysing overnight against pH 7.5, the second for 11 hours to shift from pH 7.5 to 7.0, and finally an overnight shift to pH 6.7.

For their purification, an anion exchange (AIEX) chromatography column (HiTrap Q HP, 5 ml column, 1 CV = 5.027 ml) was connected in line with a cation exchange (CEX) chromatography column (HiTrap SP HP, 5 ml column, 1 CV = 5.027 ml). Firstly, ethanol was rinsed out with water and columns equilibrated with 6 CV of 2 M NaCl and 6 CV of Buffer A (**Table 2**). The dialysed sample was subsequently applied on the columns; however, the AIEX column was removed before starting the washing step. Proteins bound to the CEX column were washed off with 6 CV of Buffer A and histones eluted using a 20 CV gradient from Buffer A to Buffer B (**Table 2**). A final wash with 2 CV of Buffer B was performed, as well as both columns were

recovered with 2 M NaCl and 20% ethanol. Lastly, purified histones were dialysed two times against dialysis buffer (**Table 2**) in Spectra/Por®3 Dialysis Membrane MWCO 3.5 KDa (Spectrum Labs), aliquoted to yield 2 mg/aliquot, flash-frozen, lyophilized (Alpha 1-2 LDplus, Christ) and stored at -80ºC before use.

3.1.3 Octamer assembly

Lyophilized histones were resuspended in Octamer UB (**Table 3**) for 30 minutes at 4ºC to have a final concentration of ~3 mg/ml. After resuspension, H2A, H2B, H3 and H4 were mixed at a 1.2:1.2:1:1 molar ratio, dialysed three times for 6 hours, overnight and 6 hours, in Spectra/Por®3 Dialysis Membrane MWCO 3.5 KDa (Spectrum Labs) against refolding buffer (RB) (**Table 3**). The dialysed sample was recovered and concentrated in 10K cut-off Amicon® Ultra Centrifugal Filter to reach ~3 mg/ml at 8000 rpm, 4ºC (Centrifuge 5424 R, Eppendorf). To purify the assembled histone octamer, the sample was applied on a HiLoad 16/600 Superdex 200 pg column (Cytiva), pre-equilibrated with refolding buffer (RB) (**Table 3**). Elution fractions were checked in 15-wells Novex™ 16% Tris-Glycine gels (Invitrogen™, ThermoFisher Scientific) in 1x Tris-glycine buffer (25 mM Tris, 250 mM glycine, 0.1% SDS), using a PageRuler™ Prestained Protein Ladder (ThermoFisher Scientific). Denatured conditions were conferred by the running buffer and the 4x loading buffer (280 mM Tris-HCl pH 6.8, 8% SDS, 10% β-mercaptoethanol, 0.1% bromophenol blue, 45% glycerol). Gels were stained with Coomassie blue stain InstantBlue™ (Expedeon Ltd.). Peak fractions containing the histone octamer were pooled and supplemented with a buffer containing 50% glycerol and 1.5 M NaCl to reach a final glycerol concentration of 40%. This allowed to store the histone octamer at -20ºC before use without compromising its integrity.

Table 3. Histone octamer and nucleosome reconstitutions buffers.

	Octamer UB	Refolding buffer (RB)	RB high salt	RB low salt
HEPES pH 7.5	20 mM	10 mM	20 mM	20 mM
NaCl	-	2 M	2 M	20 mM
EDTA	-	2.5 mM	2 mM	2 mM
Guanidinium-HCl	7 M	-	-	-
DTT	10 mM	2.5 mM	1 mM	1 mM

3.1.4 Nucleosomal DNA preparation

Adenoviral major late promoter (AdMLP) DNA scaffolds containing a 147 bp Widom-601 sequence, which was located at different distances from the TSS, were inserted into pUC119 vectors through restriction digest cloning (**Materials**). DNA templates were amplified by PCR from these vectors using different primers depending on the scaffold of interest (**Materials**). Amplifications were performed based on the respective protocol for Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific). The PCR program utilised was the following: polymerase activation at 98ºC, 30 s; 35 cycles of denaturation, annealing and extension (98ºC, 10 s; 55º, 30 s and 72ºC, 3 minutes, respectively); final extension at 72ºC, 5 minutes. Size of the amplicons were checked in a 1% agarose gel to confirm that the products of interest were correctly amplified, where 100 bp DNA ladder (New England BioLabs Inc.), SERVA DNA Stain Clear G (SERVA) and Gel Loading Dye Purple 6x (New England BioLabs Inc.) were used for the electrophoresis. When scaling up, a total PCR volume of 10-20 ml was amplified in 96-well PCR plates. PCR products were purified through Resource Q 6 ml (GE Healthcare), using a gradient of 0-50% TE high-salt buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 2 M NaCl). Eluates where ethanol-precipitated and resuspended in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0).

3.1.5 Salt-gradient dialysis

The above-mentioned DNA templates were used for nucleosome reconstitution using salt-gradient dialysis (SGD). A series of molar titrations of histone octamer: DNA, ranging from 0.9:1, 1:1, 1.1:1 to 1.2:1, was first performed. The mentioned reactions were mixed on ice and incubated for 30 minutes at 4ºC, after which they were transferred into Slide-A-Lyzer™ MINI Dialysis Units 3.5K MWCO (ThermoFisher Scientific). The samples were then gradient dialysed with a peristaltic pump to exchange RB high salt with RB low salt (**Table 3**), at a flow rate of 1.8 ml/min for 16 – 24 hours at 4ºC. Thereafter, samples were recovered from the dialysis devices and spun down. Before assembling the PIC-nucleosome complexes, these nucleosomes were further purified over 4% polyacrylamide gels (0.2x TBE) using a Model 491 PrepCell (Bio-Rad), and subsequently dialysed into PIC-buffer (20 mM HEPES pH 7.5, 100 mM KCl, 2.5% glycerol, 2 mM $MgCl_2$, 1 mM TCEP) overnight at 4°C. Reconstitutions were checked on 1% agarose gels (0.5x

TBE running buffer) and gels were stained using SYBR™ Gold (Invitrogen) for 15-30 minutes. Nucleosomes were concentrated on Amicon Millipore 15 ml 50,000 MWCO centrifugal concentrator to a final concentration of 7-9 µM and their concentration was monitored by measuring their absorbance at 260 nm using NanoDrop® 2000 Spectrophotometer (ThermoFisher Scientific)

3.2 Biochemical preparation of mammalian PIC-nucleosome complexes

3.2.1 Expression and purification of human GTFs

Human TBP, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH were expressed and purified as described previously (Kokic et al., 2019; Aibara et al., 2021). Whereas TFIIA and TBP were independently expressed in insect cells, TFIIB, TFIIF and TFIIE were expressed in LOBSTR-BL21(DE3)-RIL *E. coli* and BL21-Codon Plus(DE3)-RIL *E. coli* cells, respectively. All these GTFs were purified first using GE HisTrap HP (5 ml), subsequently followed by ion exchange and size exclusion chromatography steps.

TFIIH was cloned in two different insect cell expression vectors: XPD, p8, p52, p44, p62, p34 and XPB in one (7-core TFIIH), and the CAK module subunits MAT1, CDK7 and Cyclin H together in a second vector. While the first vector contained N-terminal 6xHis-TEV tags on p44 and p62 and a N-terminal 6xHis-MBP-TEV tag on XPD, all subunits of the CAK module were tagged N-terminally with 6xHis-TEV tags. Hi5 cells expressing the 7-core TFIIH were lysed with an EmulsiFlex-C5 cell disruptor (Avestin) supplemented with DNase I in lysis buffer (25 mM KOH-HEPES pH 7.6, 400 mM KCl, 20% glycerol, 5 mM TCEP, 0.284 µg ml−1 leupeptin, 1.37 µg ml−1 pepstatin A, 0.17 mg ml−1 PMSF and 0.33 mg ml−1 benzamidine). Lysate was loaded on a self-packed XK16/20 column (Cytiva) with 25 ml of amylose resin (New England Biolabs) in buffer A1 (25 mM KOH-HEPES pH 7.6, 400 mM KCl, 10% glycerol and 2 mM DTT) and eluted with buffer B1 (25 mM KOH-HEPES pH 7.6, 400 mM KCl, 10% glycerol, 2 mM DTT and 100 mM maltose) into a GE HiTrap Q HP (5 ml) column, pre-equilibrated with buffer A1. Elution of the anion exchange step was performed from 0-100% buffer HB1 (25 mM KOH-HEPES pH 7.6, 2000 mM KCl, 10% glycerol and 2 mM DTT), flow-through fractions collected, cleaved with 2.5 mg TEV protease for 8 hr at 4°C and loaded into a GE HiTrap

Heparin HP (1 ml) column, pre-equilibrated with buffer HA1 (25 mM KOH-HEPES pH 7.6, 300 mM KCl, 10% glycerol and 2 mM DTT). Elution was performed using 0-100% buffer B1, fractions collected and further purified using a Superose 6 Increase 10/300 GL in buffer GF (25 mM KOH-HEPES pH 7.6, 300 mM KCl, 10% glycerol and 3 mM TCEP). Stochiometric 7-core TFIIH was concentrated with Vivaspin 6 50,000 MWCO (GE Healthcare), flash-frozen in liquid nitrogen and stored at -70°C. The kinase module of TFIIH was purified as described with minor modifications (Boehning et al., 2018). In brief, Hi5 cells expressing the kinase were lysed with an EmulsiFlex-C5 cell disruptor (Avestin) supplemented with DNase I in lysis buffer (25 mM KOH-HEPES pH 7.6, 400 mM KCl, 15% glycerol, 1 mM $MgCl_2$, 10 µM $ZnCl_2$, 30 mM imidazole pH 8.0, 3 mM TCEP, 0.284 µg ml−1 leupeptin, 1.37 µg ml−1 pepstatin A, 0.17 mg ml−1 PMSF and 0.33 mg ml−1 benzamidine). The clarified lysate was loaded onto a GE HisTrap HP (5 ml) column, pre-equilibrated with lysis buffer, subsequently washed with buffer A1 HisTrap (25 mM KOH-HEPES pH 7.6, 100 mM KCl, 10% glycerol, 30 mM imidazole pH 8.0 and 1 mM DTT) and eluted with a linear gradient of 0-100% of buffer B1 HisTrap (25 mM K-HEPES, pH 7.6, 100 mM KCl, 10% glycerol, 500 mM imidazole pH 8.0 and 1 mM DTT) in 12 CV. Peak fractions were pooled, diluted with buffer A1 IEX (25 mM K-HEPES, pH 7.6, 100 mM KCl, 5% glycerol and 1 mM DTT), incubated with 2.5 mg of TEV protease at 4°C for 8 hr and loaded onto a GE HiTrap Q HP (5 ml), pre-equilibrated with buffer A1 IEX. The column was washed with 10 CV of buffer A1 IEX and eluted with a linear gradient 0-30% buffer B1 IEX (25 mM K-HEPES, pH 7.6, 2000 mM KCl, 5% glycerol and 1 mM DTT) for 80 CV, a step elution with 50% buffer B1 IEX for 2 CV and a final step with 100% buffer B1 IEX for 2 CV. Stochiometric TFIIH kinase trimer was pooled and concentrated using Vivaspin 20 10,000 MWCO (GE Healthcare) and loaded onto a HiLoad 16/600 Superdex 200 pg (Cytiva) which was eluted with buffer GF2 (25 mM K-HEPES, pH 7.6, 100 mM KCl, 5% glycerol and 2 mM TCEP). Stochiometric fractions were pooled, concentrated with Vivaspin 20 10,000 MWCO (GE Healthcare), flash-frozen in liquid nitrogen and stored at -70°C.

3.2.2 Purification of endogenous *Sus scrofa* Pol II

The 12-subunit Pol II was purified endogenously from *Sus scrofa* thymus tissue as previously reported (Hu et al., 2006; Bernecky et al., 2016; Vos et al., 2018a). In

summary, *S. scrofa* thymus tissue was homogenized in a 2 L blender (Waring) for 3 minutes at 4°C. The homogenized tissue was filtered through two layers of Miracloth, mixed with polyethyleneimine (final concentration 0.04%) and stirred for 30 minutes at 4°C. The solution was centrifuged at maximum speed and pellets resuspended in 0.4 M HepR buffer (25 mM Tris-HCl pH 7.9, 400 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 0.284 µg ml−1 leupeptin, 1.37 µg ml−1 pepstatin A, 0.17 mg ml−1 PMSF and 0.33 mg ml−1 benzamidine), centrifuged once more and supernatant adjusted to the conductivity of 0.2 M HepR buffer (25 mM Tris-HCl pH 7.9, 200 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 0.284 µg ml−1 leupeptin, 1.37 µg ml−1 pepstatin A, 0.17 mg ml−1 PMSF and 0.33 mg ml−1 benzamidine) with 0 M HepR buffer (25 mM Tris-HCl pH 7.9, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 0.284 µg ml−1 leupeptin, 1.37 µg ml−1 pepstatin A, 0.17 mg ml−1 PMSF and 0.33 mg ml−1 benzamidine). This lysate was loaded on a 225-ml MacroPrepQ column, pre-equilibrated in 0.2 M HepR buffer, the column was washed with 0.2 M HepR buffer (supplemented with 1 mM DTT) and eluted with 0.4 M HepR buffer (supplemented with 1 mM DTT). Eluates were pooled, precipitated by addition of $(NH_4)_2SO_4$ to 50% saturation, stirred at 4°C for 1 hr, centrifuged and pellets resuspended in 0 M HepR2 buffer (25 mM Tris-HCl pH 7.9, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 1 mM sodium metabisulfite, 0.25 mM PMSF and 1 mM benzamidine). The conductivity of the solution was adjusted on ice to that of 0.15 M HepR buffer (25 mM Tris-HCl pH 7.9, 150 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 1 mM sodium metabisulfite, 0.25 mM PMSF and 1 mM benzamidine), loaded at 4°C on a 5-ml gravity flow column of 8WG16 (αRPB1 CTD) antibody-coupled sepharose, pre-equilibrated in 0.15 M HepR buffer. The antibody column was washed with five column volumes of 0.5 M HepR buffer (25 mM Tris-HCl pH 7.9, 500 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$, 10% glycerol, 1 mM sodium metabisulfite, 0.25 mM PMSF and 1 mM benzamidine) at 4°C, and eluted at room temperature with 0.5 M HepR2 buffer (25 mM Tris-HCl pH 7.9, 500 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$, 50% glycerol, 1 mM sodium metabisulfite, 0.25 mM PMSF and 1 mM benzamidine). Eluted fractions were immediately 5-fold diluted with Pol II dilution buffer (25 mM Tris-HCl pH 7.9, 1 mM EDTA, 10 µM $ZnCl_2$ and 2 mM DTT). Diluted fractions were pooled, centrifuged and the supernatant loaded to a UnoQ1 column (Bio-Rad), pre-equilibrated with 0.1 M

HepR buffer (25 mM Tris-HCl pH 7.9, 100 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$ and 2 mM DTT), the column washed and eluted with a 20 CV linear gradient 20-100% from 0.1 M HepR buffer to 0.5 M HepR3 buffer (25 mM Tris-HCl pH 7.9, 500 mM $(NH_4)_2SO_4$, 1 mM EDTA, 10 µM $ZnCl_2$ and 2 mM DTT). Gdown1-free Pol II fractions were pooled, concentrated using an Amicon 100,000 MWCO Ultra Centrifugal Filter (Merck Millipore), buffer exchanged with Pol II final buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 10 µM $ZnCl_2$ and 1 mM DTT), flash-frozen in liquid nitrogen and stored at −70 °C.

3.2.3 Assembly of PIC-nucleosome complexes for cryo-EM

The PIC-nucleosome complexes were prepared identically for both AdMLP templates (PIC-Nuc^{10W} and PIC-Nuc^{18W}), following the previously established protocol (Plaschka et al., 2016; Schilbach et al., 2017; Aibara et al., 2021). In short, the 7-subunit core TFIIH (480 pmol) was mixed with the 3-subunit kinase module (480 pmol) to reconstitute the complete 10-subunit TFIIH at 25°C for 10 min. At the same time, Pol II (240 pmol) was pre-incubated with TFIIF (1.2 nmol) at 25°C for 10 min. Subsequently, TFIIH was incubated with TFIIE (480 pmol), the KCl concentration was immediately adjusted to 150 mM with 0-salt buffer (20 mM HEPES pH 7.5, 2.5% glycerol, 2 mM $MgCl_2$, 1 mM TCEP) and the subcomplex incubated at 25°C for 5 min. In the meantime, the PIC upstream complex was formed by adding TBP (1.2 nmol), TFIIA (2.4 nmol) and TFIIB (1.2 nmol) to the nucleosomal scaffolds (300 pmol) and incubating it at 25°C for 5 min. Afterwards, both the upstream complex and TFIIH-TFIIE were combined with Pol II-TFIIF, and sample salt concentration was decreased to 100 mM KCl by adding 0-salt buffer. This reaction was incubated at ~400 rpm, 25°C for 90 – 120 min. Once the reaction was finished, the sample was centrifuged at 21,130*g* for 10-15 min and further purified by gradient ultracentrifugation. 30% of the sample was purified in a gradient for analytical purposes, whereas the remaining 70% was purified by GraFix (Kastner et al., 2008), and used for structural studies. The gradient was prepared from 15% to 40% sucrose in a buffer with 20 mM HEPES pH 7.5, 100 mM KCl, 2.5% glycerol, 2 mM $MgCl_2$, 1 mM TCEP, via a BioComp Gradient Master 108 (BioComp Instruments). Sample preparations for cryo-EM were complemented with 0.2% glutaraldehyde in the 40% sucrose solution. The ultracentrifugation step was carried

out at 175,000*g* for 16 h at 4°C. Subsequently, the gradient was fractionated in 200 µl aliquots, checked by SDS-PAGE and Coomassie staining (analytical gradient) or Native-PAGE and SYBR Gold and Coomassie staining. GraFix samples were immediately quenched after fractionation with a cocktail of pH-adjusted 10 mM lysine and 40 mM aspartate. The stochiometric crosslinked PIC-nucleosome complexes were dialysed for 6-7 hr at 4°C into PIC-dialysis buffer (20 mM HEPES pH 7.5, 75 mM KCl, 1% glycerol, 2 mM $MgCl_2$, 1 mM TCEP) in Slide-A-Lyzer MINI Dialysis Devices (0.1 ml, 20 KDa MWCO) (Thermo Fisher Scientific) for sucrose removal.

3.3 Electron cryo-microscopy

3.3.1 Sample screening and data collection

Both PIC-nucleosome complexes (~130 µl) were incubated on a floating ~3.0 nm continuous carbon support for 7 min, after which the carbon film was attached to a holey carbon grid (Quantifoil R3.5/1, copper, mesh 200), washed with 4 µl of PIC-dialysis buffer and placed in a Vitrobot Mark IV (FEI/Thermo Fisher Scientific) under 100% humidity at 4°C. Under these conditions, samples were blotted with force 5 for 2 s and plunged frozen into liquid ethane. Optimal samples were identified using a Glacios transmission-electron microscope (Thermo Fisher Scientific) operated at 200 keV and equipped with a Falcon-III direct-electron detector (Thermo Fisher Scientific). Data was then collected using SerialEM 4.0 (Mastronarde, 2005) on a Titan Krios G2 transmission-electron microscope (FEI/Thermo Fisher Scientific) operated at 300 keV, with 20 eV slit width of a QuantumLS energy filter (Gatan), and equipped with a K3 summit direct detector. Imaging was performed at a nominal magnification of 81,000x (corresponding to a pixel size of 1.05 Å/pixel), with 3 s exposure in counting mode and a total dose of 41.58 and 50.45 e^- per $Å^2$, over 40 and 50 frames for PIC-Nuc^{18W} and PIC-Nuc^{10W}, respectively, at a defocus range from 0.5-1.5 µm. A total of 41517 and 36478 micrographs were collected for PIC-$nucleosome^{18W}$ and PIC-$nucleosome^{10W}$, respectively.

3.3.2 Data pre-processing

Motion correction, CTF-estimation, dose-weighting and particle-picking was performed in Warp 1.0.9 (Tegunov and Cramer, 2019). Micrographs were filtered

by resolution and motion estimation, yielding a total of 39399 and 29668, on which Warp auto-picking resulted in 4,667,603 and 4,606,320 initial particles for PIC-nucleosome^{18W} and PIC-nucleosome^{10W}, respectively.

3.3.3 Data processing

For PIC-nucleosome^{18W}, 4,667,603 particles were extracted with a binning factor of 4. The data was initially classified in cryoSPARC 3.2.0 (Punjani et al., 2017) through 4 rounds of 2D and 3D classification where, for the latter, an *ab initio* model was generated in order to sort out falsely picked particles, ice contamination and aggregated particles. After the initial cleaning of the datasets, all subsequent image processing steps were performed using RELION 3.1.0 (**Figure S2**) (Zivanov et al., 2018; Scheres, 2020). 1,725,420 particles containing cPIC were merged and unbinned, after which 2 rounds of CTF refinement, masked 3D refinement and Bayesian polishing on cPIC were done to reconstruct the cPIC at 2.6 Å. From these particles, after a series of signal subtraction where a spherical mask was initially applied to keep the signal coming between the Pol II stalk and foot, and focused masked 3D focused classifications with or without image alignment, we identified a final set of 147,341 particles which contained the CAK module at 3.3 Å. Subsequently, global 3D classifications with alignment were carried out to keep particles containing both TFIIH and the nucleosome. After signal subtracting and classifying these 1,462,564 particles with a spherical mask covering TFIIH and the nucleosome, the processing was split in order to yield the highest achievable resolution for both TFIIH and nucleosome. Regarding TFIIH, signal subtraction and 2 rounds of masked 3D focused classification yielded a reconstruction of TFIIH at 4.3 Å (188,832 particles). Masked 3D focused refinements led us to obtain XPD-containing TFIIH at 4.7 Å and XPB-containing TFIIH at 3.9 Å. In addition, reverting signal subtraction on these particles and performing either a masked or global refinement on cPIC or PIC-Nuc^{18W} generated maps at 3.0 Å and 4.0 Å, respectively. On the other hand, this same type of classification procedure was employed in order to identify those particles of highest resolution for the nucleosome in 3 subsequent rounds. 246,363 particles containing the nucleosome were then refined to 3.6 Å.

Data processing for PIC-nucleosome^{10W} was performed similarly to PIC-nucleosome^{18W} (**Figure S5**). Briefly, after initial cleaning of data, and performing 2

rounds of CTF refinement, masked 3D refinement and Bayesian polishing, we obtained 1,415,094 particles containing cPIC at 2.4 Å. The CAK module could be resolved at 3.8 Å by following a strategy similar as described above for PIC-Nuc^{18W}. Secondly, a global 3D classification with alignment was carried out to sort particles that did not contain TFIIH (cPIC-Nuc) from the ones that did (PIC-Nuc and PIC-like). The 705,995 particles containing TFIIH or TFIIH and nucleosome were exhaustively 3D classified, yielding a set of 376,063 particles, whereas the cPIC-Nuc reconstruction was generated from 668,443 particles. For these sets of particles, the cPIC signal was subtracted by applying a spherical mask on the TFIIH-Nuc region, subsequently performing a masked 3D focused classification. In the cPIC-Nuc scheme, sorting was aimed at keeping particles with highest resolution details for the nucleosome, whereas for PIC-Nuc we classified for those having the highest occupancy for TFIIH-Nuc. A last round of masked 3D focused classification without alignment was performed on both schemes and the TFIIH and the nucleosome of highest resolution were selected for further processing. Lastly, signal subtraction was reverted on the final set of particles for both reconstructions, and masked global 3D refinement was applied on cPIC-Nuc^{10W} (3.8 Å) and PIC-Nuc^{10W} (4.1 Å). For the former, the focused maps of cPIC and the nucleosome were reconstructed at 3.1 Å and 3.2 Å, respectively. As for PIC-Nuc^{10W}, focused maps of cPIC, TFIIH and the nucleosome were obtained at 3.2 Å, 4.5 Å and 3.5 Å.

3.3.4 Data post-processing

The resolution of the reconstructions was determined following the gold-standard Fourier shell correlation (cut-off at 0.143). Sharpening of maps was performed with the postprocessing tool of RELION 3.1.0, which automatically calculated the reported *B*-factors (**Table S1**). Local resolution was estimated in RELION 3.1.0 using the previously calculated *B*-factors. For the overall and cPIC maps, however, local resolution maps with *B*-factor of 0 Å were determined in RELION 3.1.0 and subsequently used in PHENIX 1.19.2 for map-sharpening (Afonine et al., 2018). Density map figures were made in UCSF ChimeraX-1.4 (Goddard et al., 2018).

3.3.5 Model building and refinement

Previously built and published structural models (PDB ID 7NVS, 6NMI, 7NVW, 7OHC, 6XBZ, 7EGB, 7ZSB) (Greber et al., 2019; Greber et al., 2020; Aibara et al., 2021; Chen et al., 2021a; Wang et al., 2021a; Wang et al., 2022) were rigid-body fitted into the cryo-EM density maps obtained with highest resolution using UCSF Chimera (Pettersen et al., 2004). Iterative rounds of real-space refinement and manual adjustments were performed using ISOLDE 1.3 (Croll, 2018) and PHENIX 1.19.2 (Afonine et al., 2018), whereas *de novo* building was performed in COOT 0.9.6 (Emsley et al., 2010). Merging of the refined structural models was done in COOT 0.9.6 (Emsley et al., 2010), and ISOLDE 1.3 (Croll, 2018) was used to flexibly fit the linker DNA between cPIC and the nucleosome. Validation statistics from MolProbity (Chen et al., 2010b) showed good geometry and stereochemistry for the final refined models (**Table S1**). Atomic model figures were made in PyMOL 2.5.0 (Schrödinger and Delano, 2020) and UCSF ChimeraX-1.4 (Goddard et al., 2018), where the colour assigned for every component is consistent throughout the manuscript.

3.4 *In vitro* transcription assay

Transcription initiation assays were performed *in vitro* with reconstituted components as described previously (Aibara et al., 2021), albeit with minor modifications. DNA templates and nucleosomes were prepared as generated for cryo-EM studies (**Nucleosomal DNA preparation**). DNA was stored at -20°C in TE buffer and nucleosomes were used right after reconstitution.

Briefly, we assembled the PIC stepwise at 25°C on both nucleosome free- and nucleosome-reconstituted DNA templates, as described above. Per reaction replicate, 3.7 pmol of DNA or nucleosome, 4.6 pmol Pol II, 23 pmol TFIIF and TFIIA, 6.9 pmol TFIIE, TFIIH and CAK, 11.5 pmol TBP and TFIIB were used. Replicates were performed in a final volume of 23.8 µl, with final buffer conditions of 3 mM HEPES pH 7.9, 20 mM Tris-HCl pH 7.9, 60 mM KCl, 8 mM $MgCl_2$, 2% (w/v) PVA, 3% glycerol, 0.5 mM DTT, 0.5 mg/ml BSA and 20 units RNase inhibitor. After assembling the PIC for 30 min, 1.25 µl of 10 mM NTP solution (Thermo Fisher Scientific) was added to each reaction (final concentration 0.5 mM/NTP) and

incubated at 30°C for 60 min. Transcription reactions were stopped with 116 μl Stop buffer (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.5 mM EDTA, 1% SDS and 4 μg proteinase K (New England Biolabs)) and incubated for 30-60 minutes at 37 °C. Nucleic acids were then precipitated with isopropanol, in presence of 300 mM sodium acetate and 0.5 mg/ml GlycoBlue (Thermo Fisher Scientific), on ice for 60 minutes. After resuspending nucleic acids, they were immediately supplemented with 1 unit DNase I (New England Biolabs) and incubated at 37°C for 60 min to digest the DNA template. A second nucleic acid isolation was performed by precipitating with isopropanol overnight at -20°C. Samples were then resuspended in 10 μl of water. RNA samples were diluted with 2x RNA Loading Dye (New England Biolabs), loaded into urea gels (2 M urea, 1x TBE, 6% acrylamide: bis-acrylamide 19:1) and separated by electrophoresis in 1x TBE buffer running buffer for 33 minutes at 180 V. Low Range ssRNA Ladder (New England Biolabs) was used for size reference, gels were stained for 10 min with SYBR™ Gold (Invitrogen) and RNA was visualized with a Typhoon 9500 FLA imager (GE Healthcare Life Sciences).

3.5 TT-seq and MNase-seq data analysis

TT-seq labelled and total RNA data (raw and processed) and MNase-seq data (DANPOS3 called nucleosome dyad positions) were taken from Velychko et al. (manuscript submitted). Protein-coding genes (RefSeq GCF_000001405.39, NM) were split into groups based on their RNA synthesis levels as follows: active genes were defined as the genes that are contained in the major transcript isoform annotation taken from Velychko et al., which is based on the total RNA expression data. Active genes were further split deciles q1-q10 based on their RNA synthesis level (TT-seq labelled RNA RPKM). Moreover, active genes were selected to have labelled RNA RPKM $\geq$ 0.01 and inactive (off) genes were defined by having RPKM < 0.01 for both labelled and total RNA samples. This cutoff was determined by plotting densities of replicate-averaged $\log_2$ (TT-seq labelled RNA RPKM) values over all genes and selecting a suitable cutoff in the valley between the two peaks of the bimodal distribution. To determine the correct TSS for each of the inactive genes we selected genes with only one RefSeq annotated isoform. For each gene, the +1 nucleosome dyad position was defined by considering nucleosome dyad position(s)

falling within the region from TSS to 200 bp downstream. If two nucleosome dyads overlapped this region, the one closer to the TSS was defined as +1 nucleosome, and if no nucleosome dyad was called in this region, the gene was excluded from further analysis. The final gene sets contained 2680 inactive genes and 9970 active genes (997 per RNA synthesis decile), of which 2261 (off), 892 (q1), 812 (q2), 854 (q3), 837 (q4), 875 (q5), 898 (q6), 903 (q7), 913 (q8), 933 (q9) and 956 (q10) had an annotated +1 nucleosome. TSS to +1 nucleosome edge distances were calculated by subtracting 73 bp from the TSS-to-dyad distances.

3.6 Quantification and Statistical Analysis

For *in vivo* data analysis, significance between deciles was determined by calculating *p*-values by two-sided Mann–Whitney U-test.

For the *in vitro* transcription assays, intensity values of gels were quantified using ImageJ 2.1.0 (Schindelin et al., 2012), subtracted against the background and normalized to the signal of the corresponding reaction of DNA templates without nucleosomes reconstituted. To facilitate comparisons between different nucleosome distances, signals were scaled to the normalized intensity of Nuc^{18W}. Statistical analysis was performed using one-way ANOVA tests with Welch‘s correction to obtain statistical significance (*p* values). All statistical analysis and diagrams were generated using RStudio or GraphPad Prism 9.1.0.

4. Results

4.1 Promoter proximity of the +1 nucleosome reduces transcription

Various studies have related the proximal location to the TSS of the +1 nucleosome to low gene expression levels (Schones et al., 2008; Valouev et al., 2011), but it was thus far not investigated whether genes that contain promoter-proximal nucleosomes indeed show low RNA synthesis rates. To address this, we analysed transient transcriptome sequencing (TT-seq) and micrococcal nuclease sequencing (MNase-seq) data obtained from HEK293 cells (Velychko et al., manuscript submitted). We split genes into ten groups based on their RNA synthesis levels (inactive and RNA synthesis deciles q1-q10) and calculated the distance from the TSS to the +1 nucleosome edge for genes in each group. We observed that a reduction of RNA synthesis correlates with an upstream shift of the +1 nucleosome (**Figure 5A**). The median distance from the nucleosome edge to the TSS for moderate-to-highly (q3-q10), weakly (q1-q2) active and inactive (off) genes was 44-35, 30-20 and 12 bp, respectively. This suggests that gene activity is reduced when the nucleosome is proximally located to the promoter region *in vivo*.

To corroborate these findings, we carried out *in vitro* transcription assays with highly purified, recombinant *X. laevis* histones assembled into a histone octamer (**Figure S1**), human initiation factors and endogenous *Sus scrofa* Pol II, which is 99.9% identical to human Pol II (**Figure S2A**). For these assays, we used DNA templates containing a +1 nucleosome positioned at increasing distances from the TSS (**Figure 5B-5D and Figure 7A**). Promoter-dependent *de novo* transcription of nucleosome-containing DNA templates required TFIIH, providing a positive control (**Figure 5B-5C**). We found that transcription was reduced ~2.5-fold or ~4.5-fold when templates were used that contained a nucleosome with its edge located either 18 or 10 base pairs from the TSS (PIC-Nuc^{18W} and PIC-Nuc^{10W} templates, respectively), when compared to the corresponding nucleosome-free DNA templates (**Figure 5C**). We further observed that positioning a nucleosome with its edge directly at the TSS (PIC-Nuc^{0W}) fully inhibited transcription (**Figure 5D**), consistent with previous studies (Knezetic and Luse, 1986; Lorch et al., 1987; Workman and Roeder, 1987; Li et al., 2010). In summary, shifting a +1 nucleosome

from a downstream location to a more promoter-proximal position closer to the TSS reduces transcription activity *in vitro* and *in vivo*.

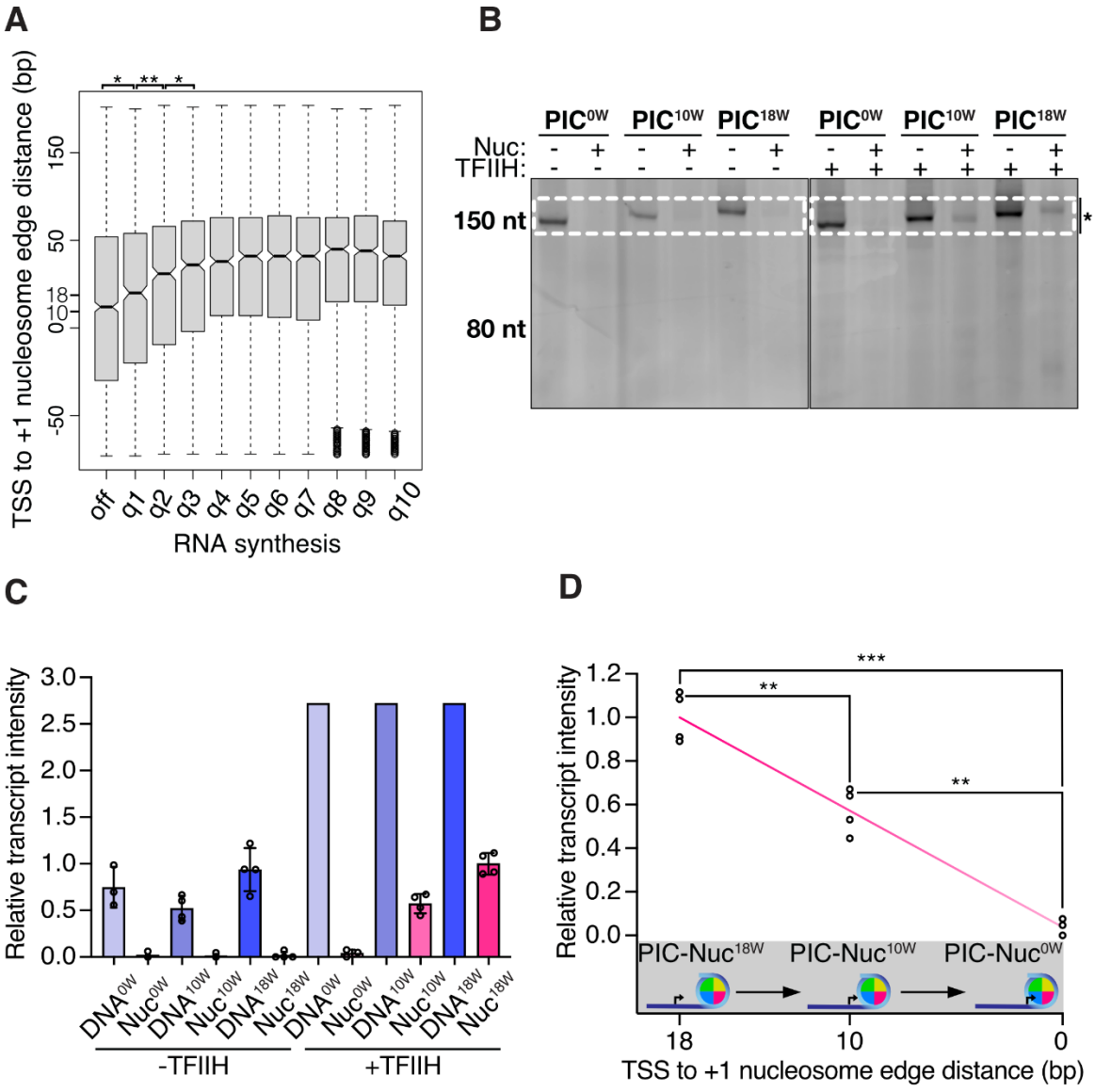


Figure 5. Promoter proximity of the +1 nucleosome reduces transcription.

(A) Boxplots showing TSS to +1 nucleosome edge distances for gene groups with different RNA synthesis levels in HEK293 cells. Genes were split into inactive and deciles of RNA synthesis q1-q10 (low-high, see **3.5**). Box limits are the first and last deciles, and the band inside the box is the median. The ends of the whiskers extend the box by 1.5 times the interquartile range. Notches represent 95% confidence intervals for the median values. Statistical significance is denoted above the respective boxplots (**, *p*-value < 0.01; *, *p*-value < 0.05). *P*-values for off-q1, q1-q2 and q2-q3 are 3.2E-02, 4.0E-3 and 2.6E-02, respectively.

(B) Urea-PAGE of *in vitro* transcription assays where the edge of the nucleosome-positioning sequence is adjacent to the TSS, or 10 bp and 18 bp downstream of the TSS (PIC^{0W}, PIC^{10W} and PIC^{18W}, respectively). Transcription reactions were performed with and without TFIIH and the nucleosome. The expected full-length RNA product is indicated with an asterisk. Dashed rectangles denote the area used for quantifications.

(C) Quantifications of the transcription assays shown in B (see **3.6**). Data is represented as mean over replicates (spheres).

(D) TFIIH-dependent transcription of nucleosome-reconstituted DNA templates shows that the proximity of the nucleosome regulates gene activity. Plotted intensities correspond to transcribing PIC complexes in the presence of a nucleosome located at different distances from the TSS. Statistical significance is shown (***, *p*-value < 0.001; **, *p*-value < 0.01). *P*-values for PIC-Nuc^{18W}-PIC-Nuc^{10W}, PIC-Nuc^{18W}-PIC-Nuc^{0W} and PIC-Nuc^{10W}-PIC-Nuc^{0W} are 4.2E-03, 3.0E-4 and 1.8E-03, respectively (see **3.6**).

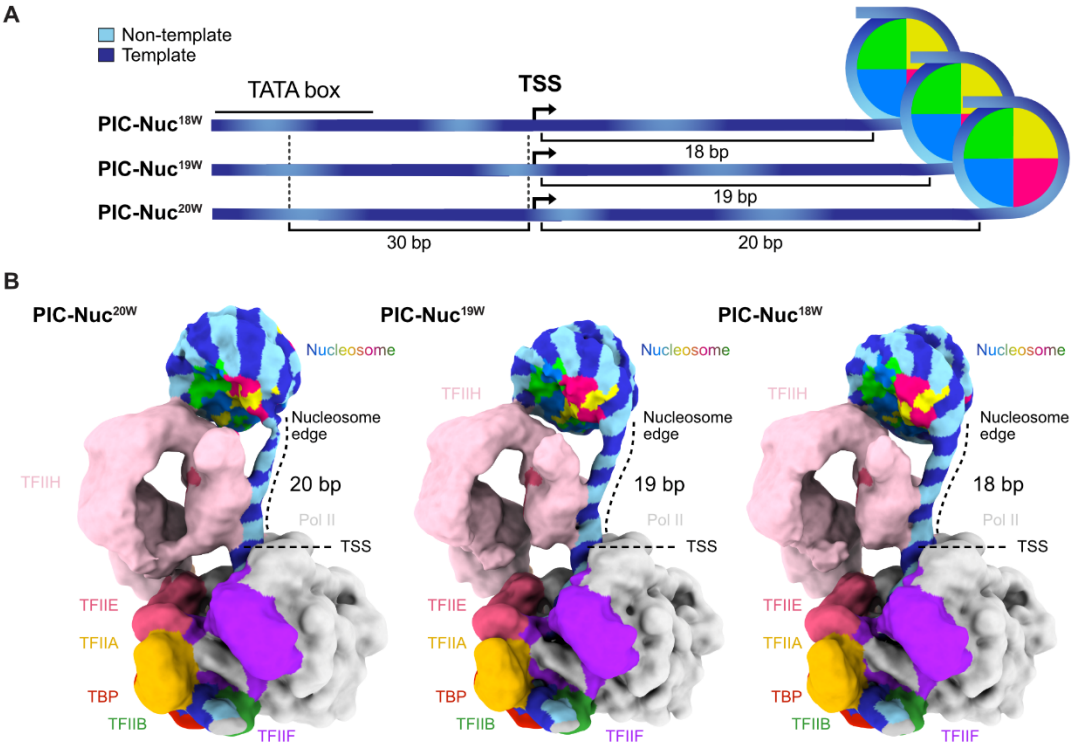


Figure 6. Cryo-EM screening of most stable PIC-nucleosome complex for high-resolution determination.

(A) Nucleosome-containing DNA templates used for the cryo-EM maps shown in B. Distances between the TSS and the edge of the nucleosome are indicated below each DNA template. The TSS and the TATA box are highlighted with arrows and solid lines, respectively.

(B) Cryo-EM maps of PIC-nucleosome complexes where the TSS-nucleosome distances are 18-20 bp. TSS and TSS-nucleosome distances are indicated by a dashed curved line. Colours are highlighted next to the respective protein subunits.

4.2 Mammalian PIC-nucleosome structure

To investigate how the position of the +1 nucleosome influences transcription initiation, we performed cryo-EM analyses of reconstituted PIC-nucleosome complexes. We first designed three DNA promoters containing the edge of the +1 nucleosome 18-20 bp downstream of the TSS (**Figure 6A**). This design is consistent with TSS-nucleosome distances we and others found *in vivo* for weakly expressed genes (**Figure 5**) (Schones et al., 2008; Valouev et al., 2011). After structure inspection at low-resolution, we did not observe PIC large conformational changes amongst all distances employed (**Figure 6B**). However, PIC-Nuc^{18W} showed stronger cryo-EM density for the +1 nucleosome edge (**Figure 6B**), and thus we proceeded to determine its high-resolution structure. We reconstituted the PIC on this nucleosome template (**Figure 7A**) and determined the structure by cryo-EM analysis at an overall resolution of 4.0 Å (PIC-Nuc^{18W}) (**Figure 7A and 7B, Figure S2A-S2D**). Signal subtraction and focused refinement strategies improved the resolution of the XPB-containing part of TFIIH, the XPD-containing part of TFIIH, the TFIIH subcomplex CDK-activating kinase module (CAK), the core PIC (cPIC) and the nucleosome to 3.9 Å, 4.7 Å, 3.3 Å, 3.0 Å and 3.6 Å, respectively (**Figure S2D and Figure S3**).

In the resulting structure, we observed the canonical conformation of the PIC (He et al., 2016; Aibara et al., 2021), and only minor movements in the upstream complex containing TBP, TFIIA and TFIIB (**Figure 7B and Figure S4A**), consistent with its previously reported flexibility (Aibara et al., 2021). TFIIH adopts an open conformation, as observed in all previous PIC structures (**Figure 8A and Figure S4A**) (He et al., 2016; Abdella et al., 2021; Aibara et al., 2021; Chen et al., 2021a; Chen et al., 2021b; Rengachari et al., 2021; Chen et al., 2022), allowing for the complete engagement of both ATPase lobes with DNA (**Figure 9A**). This is facilitated by the distal position of the nucleosome, which renders the downstream DNA-binding region of XPB nucleosome-free (**Figure S4B**). This downstream region is slightly bent when compared to the structure of the PIC on a nucleosome-

free DNA (**Figure S4A**) (Aibara et al., 2021). In this PIC-nucleosome structure, about two turns of nucleosomal DNA are detached from the histone octamer at superhelical location (SHL) –5 to SHL –7 (**Figure 9B**).

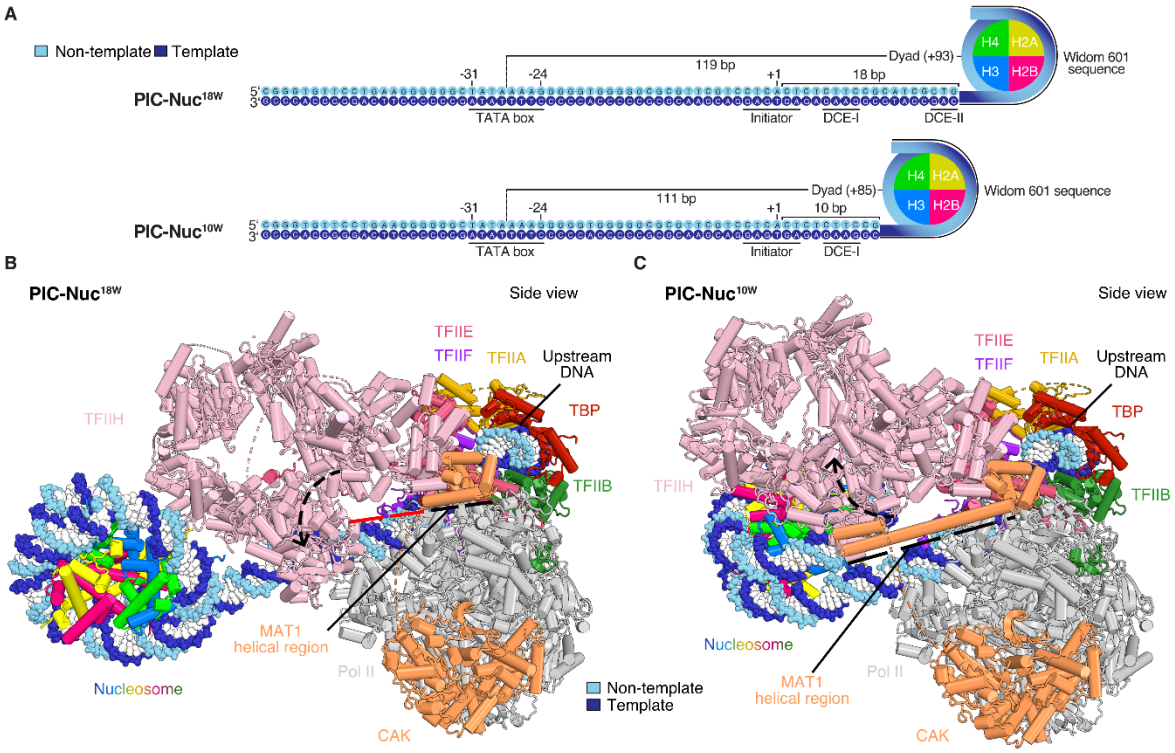


Figure 7. Structure of PIC-nucleosome complexes.

(A) Scheme of the nucleosome-containing DNA templates used for the structures shown in B and C. Distances between the TSS and the edge of the nucleosome as well as from the TATA box midpoint to the nucleosome dyad are indicated. Core promoter motifs and the nucleosome positioning sequences are highlighted with solid lines and curves, respectively.

(B) Model of the mammalian PIC-nucleosome^{18W} complex (PIC-Nuc^{18W}) shown as cartoon-sphere in side view.

(C) Model of the mammalian PIC-nucleosome^{10W} complex (PIC-Nuc^{10W}) shown as cartoon-sphere in side view.

Major conformational changes between PIC-Nuc^{18W} and PIC-Nuc^{10W} are indicated with arrowheads. The MAT1 helical region is highlighted with a dashed line (black colour denotes ordered and red disordered).

4.3 Nucleosome proximity alters PIC conformation

To investigate the structural basis of transcription reduction by a promoter-proximal +1 nucleosome, we also determined the structure of a PIC-nucleosome complex where the +1 nucleosome was positioned at a more proximal location, with the edge of the nucleosome located 10 bp downstream of the TSS (PIC-Nuc^{10W}) (**Figure 7A and 7C**). We obtained an overall resolution of 4.1 Å for this complex (**Figure S5A-S5D**). With the use of signal subtraction and focused refinement, we improved the resolution of the reconstructions for the cPIC to 3.1-3.2 Å, for TFIIH to 4.5 Å, for the nucleosome to 3.2-3.5 Å and for the CAK to 3.8 Å (**Figure S5D and Figure S6**).

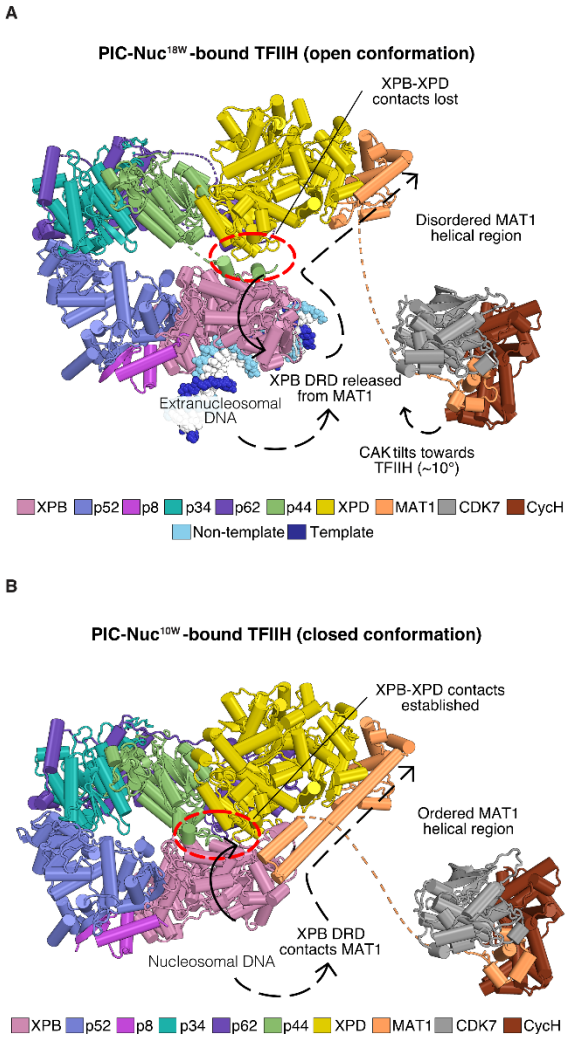


Figure 8. Nucleosome proximity alters TFIIH conformation.

(A) A distal nucleosome (PIC-Nuc^{18W}) induces an open state of TFIIH. Solid and dashed arrowheads denote rearrangements from the closed state, the red dashed oval indicates the region of XPB-XPD contacts.

(B) A proximal nucleosome (PIC-Nuc^{10W}) induces a closed state of TFIIH. Solid and dashed arrowheads denote rearrangements from the open state, the red dashed oval indicates the region of XPB-XPD contacts.

In the obtained structure, the conformation of the core PIC (cPIC, lacking TFIIH) does not deviate from that observed in the PIC-Nuc^{18W} structure and is also similar to that observed in the absence of a nucleosome (**Figure S7A**) (Aibara et al., 2021). However, in contrast to all known PIC structures, TFIIH adopts a closed conformation that had only been previously observed for free TFIIH (**Figure 8B and Figure S7B**) (He et al., 2016; Greber et al., 2017; Greber et al., 2019; Abdella et al., 2021; Aibara et al., 2021; Chen et al., 2021a; Chen et al., 2021b; Rengachari et al., 2021). In this closed conformation, the two ATPase subunits, XPB and XPD, contact each other. The TFIIH subunit MAT1 favours this arrangement by contacting the XPB damage recognition domain (DRD) and its helical region stabilizes the closed TFIIH conformation. In summary, these observations show that the more TSS-proximal location of the +1 nucleosome led to an alternative PIC conformation with a closed state of TFIIH.

4.4 Closed state of TFIIH is incompatible with DNA opening

In the observed closed state of TFIIH, the XPB translocase subunit, which is essential for opening DNA, cannot fully engage the promoter. Instead, the ATPase lobe 1 binds nucleosomal DNA approximately 88 bp downstream of the TSS, at SHL +0.5, whereas lobe 2 is located ~30 Å away from the DNA (**Figure 9C**). The observed PIC conformation and PIC-DNA contacts are thus incompatible with promoter DNA opening and transcription initiation. The closed TFIIH conformation was induced by the proximal location of the nucleosome because it is the only difference in the two structure determination experiments.

During 3D classifications, we could also resolve the cPIC-nucleosome complex lacking TFIIE and TFIIH (cPIC-Nuc^{10W}) and refined its structure at 3.8 Å resolution (**Figure 10A, Figure S5D, and Figure S6**). In this structure, the nucleosome overlaps with the location of TFIIH in a PIC complex as it is shifted by ~20 Å and tilted by ~25°, compared to the complete PIC-Nuc^{10W} structure (**Figure 10B**). This is consistent with a proximally positioned +1 nucleosome interfering with complete

promoter DNA engagement by TFIIH. We therefore suggest that the PIC with the closed TFIIH conformation represents an inhibited state.

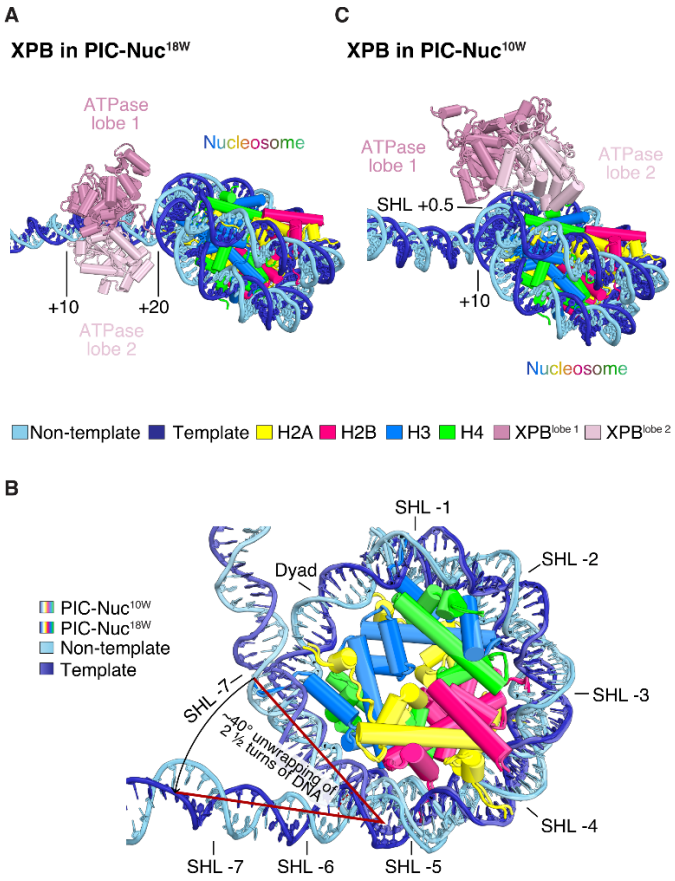


Figure 9. Nucleosome position alters PIC-DNA contacts.

Comparison of XPB binding to **(A)** a distal (PIC-Nuc^{18W}) and **(C)** proximal nucleosome (PIC-Nuc^{10W}) by superimposition on XPB and the nucleosome. The position of the nucleosome alters DNA binding of the ATPase lobes of TFIIH subunit XPB. The colour code is provided at the bottom legend. Numbers on DNA denote the distance from the TSS.

(B) Cartoon representation of the different nucleosomes determined (PIC-bound), showing distinct nucleosomal DNA wrapping states. Top-left legend specifies the colour used for the models. Model superimposition was carried out by aligning on the nucleosome. SHL, superhelical location.

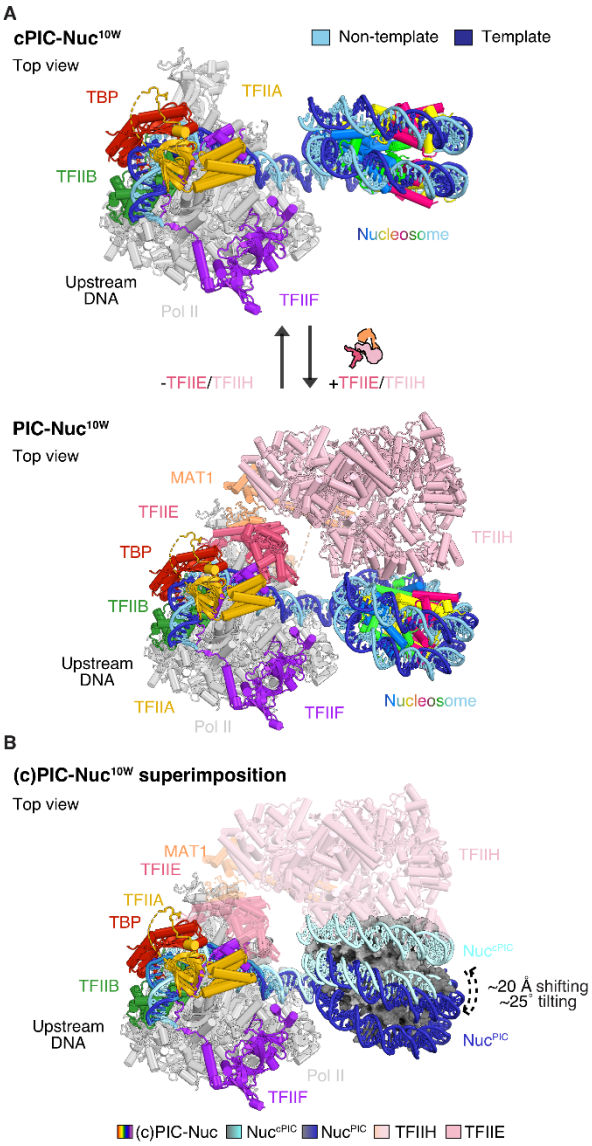


Figure 10. Structural transition of the nucleosome upon binding or release of TFIIE and TFIIH on the cPIC-Nuc10W.

(A) Top views of the cPIC-Nuc^{10W} (lacking TFIIE and TFIIH, top panel) and PIC-Nuc^{10W} (bottom panel).

(B) Overlay of both structural models shown in **(A)**, showing a conformational change of the nucleosome depending on the presence of TFIIE and TFIIH. Arrowheads describe the direction of such movements. Superimposition of the models was performed by aligning on Pol II. Histone octamers are depicted in different shades of grey surface. Colours of different protein subunits or DNA strands are depicted in the figure or at the bottom legend.

4.5 Altered TFIIH-nucleosome contacts

TFIIH bridges the cPIC to the +1 nucleosome in both structures we determined. The comparison of these two structures shows that the rotational position of the nucleosome with respect to TFIIH changes (**Figure 7B and 7C**). As a consequence, TFIIH engages differently with promoter DNA and its contacts with the nucleosome are altered. In the PIC-Nuc^{18W} structure (**Figure 11A**), the TFIIH subunit p52 establishes most of the nucleosome contacts, with mostly basic residues within the α13–α14 loop (residues 274-278) binding nucleosomal DNA at SHL –1, and acidic residues of the p52 middle domain interacting with basic amino acids of the C-terminus of H2A and H3 α1. The N-terminal region of XPB contacts the N-terminal region of histone H3. Conversely, in the PIC-Nuc^{10W} structure (**Figure 11B**), the p8 subunit contacts the H2B acidic patch of the nucleosome and the N-terminal regions of H3 and H4. Basic residues of XPB ATPase lobe 1 bind to nucleosomal DNA at SHL +0.5, forming a major TFIIH-nucleosome interface.

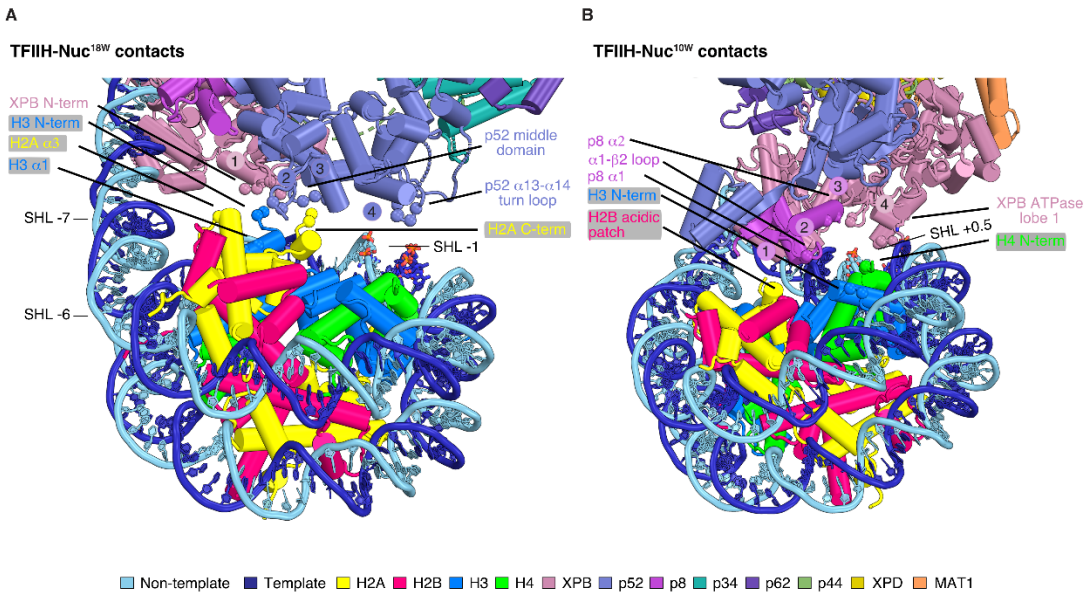


Figure 11. TFIIH-nucleosome contacts.

(A) Cartoon representation showing interactions between TFIIH and the nucleosome in PIC-Nuc^{18W}. Residues T51, K52, D54 and Q64 of the XPB N-terminus (interface 1) interact mainly electrostatically with the N-terminal tail (NTT) of H3 (residues 37-41). p52 acidic residues (E189, E192, interface 2) contact basic residues from both H2A α3 (K74) and H3 α1 (R52, K56), and its residues 168-170 (interface 3) establish backbone interactions with residues 118-119 of H2A C-terminus. In addition, mostly basic residues of p52 α13-α14 loop (274-278, interface

4) bind to the minor groove adjacent to the nucleosome dyad (SHL –1). SHL, superhelical location.

(B) Cartoon representation showing interactions between TFIIH and the nucleosome in PIC-Nuc^{10W}. Residues K31-N27 (interface 1), R56 (interface 2) and Q63 (interface 3) of the p8 subunit make electrostatic contacts to the H2B acidic patch, and the N-terminal regions of H3 (D81) and H4 (R23), respectively. XPB ATPase lobe 1 residues 416-420 and K449 (interface 4) engulf the DNA major groove adjacent to the nucleosome dyad (SHL +0.5). SHL, superhelical location.

The positions of the C-α atoms of interacting residues are shown as spheres.

4.6 TFIIH kinase module and RPB6 NTT

We also observed the TFIIH kinase module (CAK) in our structures (**Figure 8 and Figure S8A**). The CAK interacts with MED6 and the hook of Mediator in previously determined Mediator-PIC complexes, positioning CAK far from the Pol II surface (Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021). In our structures, however, the CAK docks between the RPB1 foot and the Pol II stalk (RPB4-RPB7), using its subunit Cyclin H (CycH) to form a wedge between them. The N-terminal region of CAK CycH subunit contacts the RPB1 linker helix connecting to the C-terminal domain and the RPB7 α1-α2 loop, whereas the C-terminal helix of CycH mostly establishes charge-based interactions with the RPB1 foot helix α28 and α31 (**Figure S8B**). This location had been previously observed with different levels of confidence in yeast (Tsai et al., 2017) and human (**Figure S8C**) (Yan et al., 2019a; Chen et al., 2021a).

Finally, in our structures we observed an extra density in the Pol II cleft (**Figure S8D**). Based on the similar location found for the yeast N-terminal tail (NTT) of the Pol II subunit Rpb6 (Wang et al., 2022), we suggest that this density is due to the mammalian counterpart in RPB6 (**Figure S8E and Figure S8F**). Whereas the *S. cerevisiae* NTT is located over the Pol II bridge helix, the density of the putative mammalian NTT locates closer to RPB1 helix α37 (**Figure S8F**). Consistent with findings in yeast (Wang et al., 2022), the cryo-EM density shows clashes between the putative NTT and a modelled DNA template in the active centre (**Figure S8G**), indicating that the NTT must be released from the cleft upon formation of an open complex.

5. Discussion

5.1 Nucleosome position: a regulator of transcription

In the work presented here, we observe *in vivo* that an upstream shift of the +1 nucleosome coincides with a reduction of RNA synthesis. This can be conceptualised in the opposite direction – a distally-located +1 nucleosome effectively reduces the barrier for transcription. This observation is consistent with many yeast and human genome-wide studies, which measure total RNA and point out that genes with different levels of transcription change their chromatin landscape (Han and Grunstein, 1988; Guenther et al., 2007; Li et al., 2007; Schones et al., 2008; Valouev et al., 2011; Korber and Barbaric, 2014; Klein-Brill et al., 2019; Kubik et al., 2019; Schwartz et al., 2019; Serizay et al., 2020). For example, the +1 nucleosome position shifts upstream closer to the promoter region in genes with low transcription activity (Schones et al., 2008; Valouev et al., 2011) and therefore rises the barrier. In this work, we analysed this effect to a deeper level since we could detect the rate of RNA synthesis with TT-seq, and split genes accordingly to correlate them with the position of the +1 nucleosome. The mentioned published studies only detected total RNA levels, and therefore the sensitivity of TT-seq allows us to further understand the role of the +1 nucleosome in transcription *in vivo*. In fact, our analysis shows that the +1 nucleosome position changes gradually closer to the promoter as the reduction of transcription is more pronounced. This suggests that the +1 nucleosome does not completely repress but rather fine-tunes transcription.

The emergence of single-molecule assays and genomic technologies helped establishing this model. Single-molecule studies describe that indeed the nucleosome slows down transcription elongation but does not inhibit it (Dangkulwanich et al., 2013; Chen et al., 2019). Genome-wide studies also show that Pol II pauses on the +1 nucleosome after the promoter-proximal pause site (Churchman and Weissman, 2011; Kwak et al., 2013; Weber et al., 2014; Ramachandran et al., 2017), for which different elongation factors help overcome these barriers differently (Zumer et al., 2021). To this end, position of nucleosomes within a promoter matter. This is elegantly exemplified by the fact that the +2, +3, and subsequent downstream nucleosomes, represent less of a barrier to

transcription than the +1 nucleosome (Teves et al., 2014). Expectedly, this also applies to a single nucleosome, which is supported by our *in vivo* findings and from others (Schones et al., 2008; Valouev et al., 2011). The barrier effect of a nucleosome fits within the framework of gene expression, as chromatin compaction could comprise an essential tool to tune transcription levels.

In vivo studies, however, only reflect global effects and correlations that might not explain a certain process or mechanism. For this reason, we used nucleosome-reconstituted DNA templates, highly purified GTFs and Pol II in an *in vitro* transcription assay and corroborated that proximity of the +1 nucleosome to the promoter reduces transcription. Early *in vitro* studies had already observed an inhibition of transcription initiation and elongation of chromatinised simian virus DNA, when using *Escherichia coli* (*E. coli*) RNA polymerase and eukaryotic polymerases (Wasylyk and Chambon, 1979; Wasylyk et al., 1979). Inhibition of initiation was later confirmed by transcription assays performed with HeLa nuclear extracts on chromatinised circular DNA templates (Knezetic and Luse, 1986), although efficient initiation can occur if the PIC is preassembled on DNA before chromatin reconstitution (Workman and Roeder, 1987; Knezetic et al., 1988). In parallel, another laboratory reconstituted single nucleosomes using an *E. coli* nucleosome positioning sequence and rat liver histones. They described that this nucleosome inhibits transcription initiation, however, it allowed RNA elongation with purified SP6 RNA polymerase and crudely purified rat Pol II once transcription had been previously initiated (Lorch et al., 1987). Complementary biochemical studies showed that the nucleosome also constitutes a barrier for transcription elongation (Shaw et al., 1978; Izban and Luse, 1992). Overall, our results are consistent with these long-standing biochemical observations, although we did not observe a complete inhibition in all cases. Only when a nucleosome positions over the promoter, we could recapitulate a complete inhibition of transcription. This suggests that nucleosomes reconstituted in these studies might not have had a fixed position, but rather a fuzzy one. It could easily be explained by the fact that none used a strong positioning sequence, but a native or poor positioning sequence. In our case, we used a strong positioning sequence to precisely control the position of the +1 nucleosome (Lowary and Widom, 1998). We describe a reduction of transcription with promoter proximity of the +1 nucleosome, however, we only observe a

complete inhibition of transcription when the nucleosome covered the TSS. We employed highly purified GTFs, nucleosomes and Pol II, unlike the cited studies, which relied on crude extracts that do not allow to control every variable in an *in vitro* experiment. For example, we could show that TFIIH is essential for transcribing nucleosomal templates, yet it is unclear whether these early studies had TFIIH present in the used extracts. A recent biochemical work also confirmed our *in vivo* and *in vitro* findings by using strongly-positioned nucleosomes and nuclear extracts – proximal nucleosomes have an inhibitory effect on transcription (Fisher and Luse, 2023). Taken together, our *in vivo* and *in vitro* findings suggest that the +1 nucleosome regulates transcription gradually, depending on its position relative to the gene promoter.

Even though it is of great interest to use native nucleosome sequences in such experiments, it is probable that transcription would be blocked due to the occlusion of the promoter. For a controlled experiment, chromatin remodellers would have to be used since they are essential to generate NDRs, as shown for *S. cerevisiae* (Kubik et al., 2019; Oberbeckmann et al., 2021b). It is also unclear whether the extracts used by these researchers contain the right remodelling enzymes to generate an environment of well-positioned nucleosomes for transcription (Knezetic and Luse, 1986; Lorch et al., 1987).

Taking advantage of our highly characterised *in vitro* system, we aimed to dissect the mechanism for the transcription reduction caused by a promoter-proximal +1 nucleosome. For this, we subjected our biochemical preparations to cryo-EM and visualised, for the first time, the effect of a proximal nucleosome on transcription initiation (**Figure 12**). Our structural analysis shows that when a nucleosome, located ~20 bp downstream of the TSS, shifts 10 bp upstream closer to the promoter, it occludes the promoter DNA binding region of XPB (Ssl2 in yeast). XPB is the translocase subunit of TFIIH, which pumps DNA into the active site of Pol II and comprises the first step of DNA opening. The presence of a nucleosome only allows one of the ATPase lobes of XPB to bind naked DNA, inconsistent with the processivity of XPB and thereby DNA opening. In addition, a fraction of our particles showed dissociation of TFIIE and TFIIH from the PIC, which led us to suggest that the Pol II initiation machinery adopts an inhibited state in the presence of a promoter-proximal +1 nucleosome. When we superimpose our structural PIC-

nucleosome models with a PIC model with opened promoter DNA (He et al., 2016; Aibara et al., 2021), we show that, upon TFIIH-independent DNA opening, the nucleosome would sterically clash with the Pol II clamp and the TFIIE winged helix domain (**Figure S9**). This explains the requirement of TFIIH in our transcription assays, as well as with the fact that promoter-proximal nucleosomes interfere with productive transcription initiation. Although in an inactive form, PIC remains assembled on the promoter. One could speculate that this could be an efficient way for the cell to 1) tune down transcription and 2) be readily prepared for the next round of initiation. A chromatin remodeller like Chd1, shown to unwrap ~2 turns of DNA (Farnung et al., 2017; Sundaramoorthy et al., 2018), would be enough to change the conformation of the nucleosome and allow TFIIH to bind naked DNA with both ATPase lobes. This agrees well with our data since when we position the nucleosome 18 bp downstream of the TSS, the assembly of a productive PIC induces a detachment of ~2 turns of DNA off the nucleosome, making it compatible with DNA opening.

5.2 Conformational changes of TFIIH underlie its function

TFIIH is a large multi-subunit flexible complex, since its subunits dynamically change their conformation to fit its diverse function. Two distinct conformations have been described for this complex in a context-specific manner – open and closed states. TFIIH adopts its closed state in its free form (apo-TFIIH) and for nucleotide excision repair (NER) (Greber et al., 2017; Greber et al., 2019; Kokic et al., 2019; Topolska-Wos et al., 2020; van Eeuwen et al., 2021; Kim et al., 2023), whereas it is predominantly found in its open state during transcription initiation (Murakami et al., 2015a; He et al., 2016; Robinson et al., 2016; Schilbach et al., 2017; Aibara et al., 2021; Schilbach et al., 2021). Here, the association or dissociation of both TFIIH ATPase subunits XPB and XPD dictates whether this complex adopts a closed or open state, respectively. Even though apo-TFIIH and NER-TFIIH structures are highly similar, there are two key differences induced by the functional context. The XPD anchor domain of p62 locates in the DNA binding pocket of XPD in both apo-TFIIH and PIC-TFIIH, which inhibits the single-stranded DNA (ssDNA) translocase activity of XPD. Secondly, in NER-TFIIH both XPD-bound ssDNA and the presence of XPA promote the displacement of the TFIIH CAK module. This is best exemplified

by the visualisation of the kinase subunit MAT1 in apo- and PIC-TFIIH complexes. In the NER complex, the replacement of MAT1 by XPA is essential to ensure that RecA lobes of both XPB and XPD are brought in proximity to cooperate for DNA repair. It has been recently suggested that XPB-XPD contacts influence the stability of TFIIH, and their proximity regulate function (Yu et al., 2023).

Our work, and the one performed by others, unequivocally shows that XPB-XPD association comprises a separation of function tool of TFIIH. When the nucleosome edge locates 18 bp downstream of the TSS, we observe that the PIC adopts its canonical productive conformation previously described by others (He et al., 2016; Aibara et al., 2021). Here, the availability of naked DNA enables the binding of XPB, TFIIH adopts an open state where XPB-XPD dissociate, and thereby TFIIH is specifically competent for transcription initiation. Furthermore, when we shift the nucleosome upstream by ~10 bp, we describe an assembled PIC in which TFIIH adopts a closed state, reminiscent of the apo-TFIIH conformation (Greber et al., 2017; Greber et al., 2019). Therefore, as described in the previous section, the proximity of the +1 nucleosome disables XPB DNA binding, causes the re-association of XPB and XPD subunits, switches TFIIH into a closed state and makes DNA opening incompatible. Previous structural work modelled the MAT1 long helical region, which connects it with the XPB DRD domain, for both Mediator-free and -containing PIC complexes (Abdella et al., 2021; Chen et al., 2021a; Chen et al., 2021b). However, this is not supported by the resolution of their cryo-EM densities since it becomes fuzzy at the interface with the DRD domain. For the Mediator-containing PIC, the modelling of a large part of this helix is correct as this coactivator has been shown to interact directly with MAT1, however, the density at the interface with DRD is not convincing for structural modelling (Abdella et al., 2021; Chen et al., 2021b). In our work, we confidently built a complete MAT1 helical region which contributes to the stabilisation of TFIIH in closed state by contacting the XPB DRD domain. Previous models proposed that TFIIH open state is critical for the formation of a PIC complex (Greber et al., 2017; Greber et al., 2019; Aibara et al., 2021), however, we show for the first time that a PIC complex can assemble with different TFIIH conformational states to fine-tune its DNA opening function. Therefore, our findings provide an additional layer of regulation to this model, where

the chromatin environment modulates TFIIH structure, function and, therefore, transcription initiation.

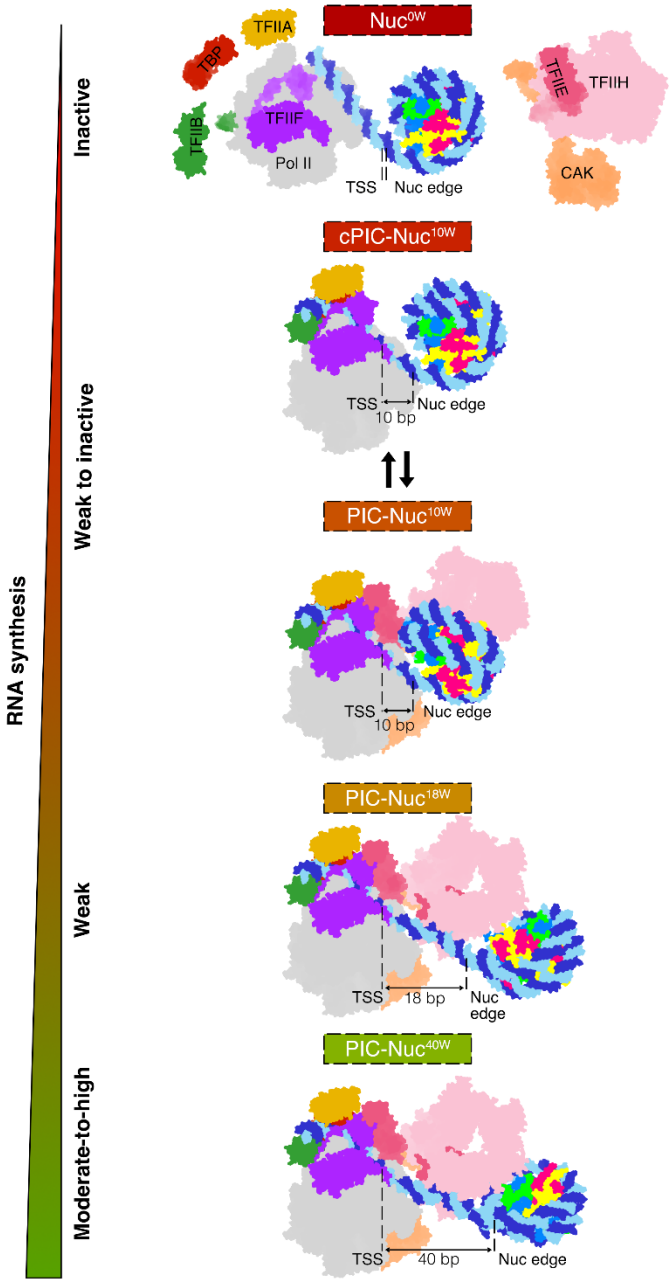


Figure 12. Model of transcription reduction by a +1 nucleosome.

The proximity of a downstream nucleosome to the promoter region of a gene distinctly reduces Pol II-mediated RNA synthesis by changing the binding of TFIIH to promoter DNA. The model of PIC-Nuc^{40W} belongs to PDB ID 8GXS (Chen et al., 2022). Different gene categories and a gradient showing decreasing RNA synthesis are denoted on the left (green, high synthesis; red, no synthesis).

5.3 Alternative TFIIH kinase position within a Pol II initiation complex

As mentioned previously, the 3-subunit kinase module that contains CDK7 is recruited to the promoter as part of TFIIH. However, its recruitment does not suffice for phosphorylation of the Pol II CTD, it also requires loading on the Pol II machinery. Namely, structural studies have observed that the Mediator complex positions the CAK module on top of the middle module, locking it through interactions with the hook region, MED14 α1 and α2 helices and MED6 shoulder domain (Schilbach et al., 2017; Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021; Schilbach et al., 2023). In our study, we describe an alternative location for the TFIIH kinase module, between the RPB1 foot and RPB4-RPB7 subunits (Pol II stalk). A previous low-resolution cryo-EM study in yeast identified a similar location (~15 Å resolution) (Tsai et al., 2017). Later, researchers employed Molecular Dynamics simulations for the human TFIID-free PIC and proposed the position we describe (Yan et al., 2019b). A recent investigation also reported a high-resolution structure of human CAK at this alternative location in a TFIID-PIC complex (Chen et al., 2021a). Therefore, our study comprises the first reliable proof for an alternative location of CAK in a TBP-PIC complex.

PIC studies that reported the location of CAK bound to the coactivator Mediator also visualised different parts of the Pol II CTD threading through the Mediator complex (Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021; Schilbach et al., 2023). For human PICs, two fragments of CTD were described to contact the Mediator head and middle modules (Abdella et al., 2021; Chen et al., 2021b; Rengachari et al., 2021). While a yeast PIC study reported 11 CTD repeats at the same location as in human, they also observed a third fragment contacting the hook domain (Schilbach et al., 2023). Unlike the yeast study, the CTD was located at the active site of CDK7 for the human complexes. This led to the model that coordinated movements of the head-middle axis of Mediator orchestrate the directionality of the CTD movement. Since we lack Mediator in our biochemical preparation, we do not observe these intermediates with Mediator-loaded CAK. On the contrary, we report the TFIIH kinase establishing extensive interactions with Pol II. Unlike the cited studies, we do not observe any Pol II CTD cryo-EM density at the active site of CDK7. Therefore, we propose that the state we determined could be a CAK pre-

loaded intermediate, where Pol II can initiate transcription but cannot escape the promoter. Altogether, the evidence reflect that Mediator is a key component of the initiation machinery in relocating CAK and loading the CTD in CDK7 for its phosphorylation. The switch from a pre-loaded to a loaded CAK state might comprise a turning point, where Pol II can start RNA synthesis and subsequently escape the promoter. This agrees well with studies suggestive of PIC-Mediator dissociation after Ser5 CTD phosphorylation, which would destabilise the PIC and prime promoter escape (Wong et al., 2014; Schilbach et al., 2017; Schilbach et al., 2023).

As mentioned previously, a TFIID-PIC study reported the same CAK location as we describe here, suggested that TFIID supports such position and enhances Ser5 phosphorylation (Chen et al., 2021a). Our results disagree with such observations, as the CAK position in our TBP-containing structures and interactions with Pol II are virtually identical as in the cited study. They also performed a TBP-PIC *in vitro* kinase assay and showed stimulation of Ser5 phosphorylation. They only observed the TFIID-dependent stimulation of phosphorylation at high TFIID/TBP concentrations, whereas at low TFIID/TBP concentrations, the enhancement was highest only with TBP. Therefore, these findings and ours indicate TFIID does not support CAK positioning, and that CTD phosphorylation stimulation is dependent on the concentration of different GTFs. In a broad context, our evidence proves that the specific location of the CAK module is conserved from yeast to human, is TFIID-independent and Mediator-dependent.

5.4 RPB6 N-terminal tail function

A previous study reported the presence of the yeast Rpb6 NTT in the cleft of Pol II through cryo-EM and cross-linking mass spectrometry analyses (Wang et al., 2022). Given the similarity between the structures determined in this study and ours, we examined whether the mammalian RPB6 NTT could also be located. We describe a weak cryo-EM density in a similar location to the yeast counterpart, and tentatively speculate this could be the mammalian putative RPB6 NTT. Such densities were never reported for any yeast or human PIC complexes before (He et al., 2013; Murakami et al., 2015a; He et al., 2016; Plaschka et al., 2016; Robinson et al., 2016; Schilbach et al., 2017; Abdella et al., 2021; Aibara et al., 2021; Chen et al., 2021a;

Chen et al., 2021b; Schilbach et al., 2021). It is compelling to assume that the presence of the nucleosome stabilises the PIC with closed promoter DNA and induces the ordering of the NTT. Our structural data supports this speculation, since we do not observe PIC states with opened DNA and our biochemical assays show a reduction of transcription in the presence of a nucleosome. Similarly, authors of the yeast PIC-nucleosome study did not observe PIC complexes with opened promoter DNA and reported a reduction of transcription (Wang et al., 2022). Even though the NTT has not been described for the RPB6 homologs in most bacterial and archaeal organisms (Minakhin et al., 2001) and is not highly conserved from yeast to humans (Wang et al., 2022), the physicochemical properties of the amino acid substitutions are mostly maintained. The position described for the yeast NTT and in our study suggest that this loop would prevent the loading of promoter DNA in the Pol II cleft (He et al., 2016; Aibara et al., 2021). Importantly, whereas *in vivo* studies proved the NTT dispensable but temperature-sensitive in yeast (Nouraini et al., 1996; Ishiguro et al., 2000), RPB6 NTT deletion in HeLa cells slowed down cell growth (Okuda et al., 2022). In this last study, this deletion also resulted in a reduction of transcription for all three Pols (Pol I, II and III), consistent with the similar loop regions identified for Pol I and Pol III (Engel et al., 2013; Fernandez-Tornero et al., 2013; Wang et al., 2021b; Wang et al., 2022). We speculate the NTT could comprise a regulatory mechanism conserved in metazoans across Pols to sense the loading of promoter DNA at the active site of Pol II.

In addition, a recent nuclear magnetic resonance (NMR) study investigated the interactions between the RPB6 NTT region and the p62 PHD domain (Okuda et al., 2022). Unexpectedly, they found that the NTT threads into the hydrophobic pocket of the PHD domain. The PHD domain of TFIIH subunit p62 has been previously shown to interact with the eWH domain of TFIIE in yeast (Schilbach et al., 2017; Schilbach et al., 2021). Human TFIIE-TFIIH interactions are known to be weaker as compared to the yeast system, and therefore this interaction has not been visualised for the human PIC (He et al., 2016; Aibara et al., 2021). Given that the crosstalk between TFIIE and TFIIH is crucial for Pol II DNA opening in eukaryotes, an interaction between RPB6 NTT and p62 PHD domain could comprise further evidence for the DNA opening sensing mechanism of RPB6 NTT. However, this awaits further structural characterisation in the human system

5.5 Chromatin-bound Pol II transcription initiation structures

Two different laboratories recently reported the first structures of PIC assembled on promoters with the +1 nucleosome, where the TSS-nucleosome distances used are representative of active genes (Chen et al., 2022; Wang et al., 2022). Whereas these studies shed light into PIC-nucleosome interactions, they did not use proximal nucleosome in their systems.

The +1 nucleosome locates closer to TFIIH in all our structures, when comparing them to the *S. cerevisiae* PIC-nucleosome models (**Figure S10**) (Wang et al., 2022). However, when NTPs were added in this study, authors observed that XPB (Ssl2 in yeast) drives a 75º rotation of the nucleosome towards TFIIH. This study elegantly demonstrates how the ATPase activity of XPB can also change the established PIC-nucleosome interactions. Whereas only the p52-p8 dimerization domain interacts with the +1 nucleosome, XPB, p52, p44 and p34 (Ssl2, Tfb2, Ssl1 and Tfb4 in yeast) contact the nucleosome after NTPs addition. In our study, the nucleosome positions closer to TFIIH without the addition of any NTP in the productive PIC-nucleosome structure, but highly resembles the orientation of the yeast PIC-nucleosome when NTPs were added (Wang et al., 2022). Our work comprises further evidence for TFIIH mediating a preferred orientation of the +1 nucleosome with respect to the PIC machinery. Nevertheless, while we observe minimal contacts between XPB and histones, most of the interactions we describe are mediated by different domains of p52 in the productive PIC-nucleosome model. This agrees well with reports that indicate these TFIIH subunits are not highly conserved (Rimel and Taatjes, 2018). In addition, we had another unexpected observation from comparing our study with the *S. cerevisiae* one (Wang et al., 2022). In both studies, a fraction of DNA is detached from the +1 nucleosome upon PIC assembly, albeit to different extents. Whereas we describe ~2 turns of unwrapped nucleosomal DNA, the cited study reported only ~1 turn of detached DNA before addition of NTPs, and ~2 DNA turns after their addition (Wang et al., 2022). We speculate that the different original orientation of the +1 nucleosome might lead to different degrees of detached nucleosomal DNA. At the same time, the different initial positioning and nucleosome wrapping state might suggest that the yeast PIC must undergo TSS scanning before

transcription initiates, unlike the human PIC (Giardina and Lis, 1993; Fishburn et al., 2016; Qiu et al., 2020).

A human study reported TFIID-containing PIC-Mediator structures assembled on +1 nucleosome-containing promoters, with TSS-nucleosome distances found for highly active genes *in vivo* (Chen et al., 2022). Whereas we mentioned above that the TFIIH-nucleosome interactions are not highly conserved from yeast to human, the interactions found in this human study are highly similar to the ones we observe. Authors reported that when the +1 nucleosome locates 70 bp downstream of the TSS, the PIC is not in contact with the nucleosome, however, when it positions ~40-50 bp downstream of the TSS, TFIIH contacts the +1 nucleosome (Chen et al., 2022). By comparing both of our studies, the +1 nucleosome orientation is conserved, favoured by the fact that mostly XPB and p52 interact with the +1 nucleosome. In this cited study, these interactions were further proved by mutating the interacting domains and determining the structures, where the +1 nucleosome subsequently dissociated from the mutated regions. In addition, they reported that both TFIIH and Mediator establish contacts with the +1 nucleosome but only when H2A.Z is incorporated into the histone octamer. This finding is consistent with *Drosophila* and mammalian investigations describing that a H2A.Z +1 nucleosome might be less stable and provide a platform for transcription initiation (Weber et al., 2014; Day et al., 2016). This human PIC-nucleosome study reports that TFIID and Mediator orchestrate PIC assembly on a chromatinised promoter and suggested an enhancement of transcription (Chen et al., 2022). However, this enhancement was only described when mutating the TATA-box element to a TATA-like one (i.e. two base pair mutations). This finding is interesting in the light of different reports suggesting the +1 nucleosome dictates TSS selection of TATA-less promoters (Jiang and Pugh, 2009b; Radman-Livaja and Rando, 2010; Rhee and Pugh, 2012; Dreos et al., 2016; Lam et al., 2019; Rossi et al., 2021; Tora and Vincent, 2021). However, we still lack structural information on the assembly of PIC in chromatinised TATA-less promoters. In addition, they did not observe an increase of transcription when the nucleosomes were positioned in a promoter-distal location (Chen et al., 2022), unlike what we describe *in vivo* and *in vitro* which agrees with what has been reported by the cellular and biochemical work of others (Knezetic and Luse, 1986; Lorch et al., 1987; Han and Grunstein, 1988; Guenther et al., 2007; Schones et al.,

2008; Valouev et al., 2011; Korber and Barbaric, 2014; Klein-Brill et al., 2019; Kubik et al., 2019; Schwartz et al., 2019; Serizay et al., 2020; Fisher and Luse, 2023). Additionally, whereas we observe a detachment of nucleosomal DNA in our productive PIC-nucleosome structure, the cited human study claimed their nucleosomes are found in the canonical fully wrapped state (Chen et al., 2022). This is likely to be a wrong interpretation since in both of our models TFIIH contacts DNA, histones and the +1 nucleosome in virtually identical manners. They determined nucleosomes at 8-15 Å resolution, from which the wrapping state of a nucleosome cannot be ascertained reliably. Since we determine the +1 nucleosome at 3-5 Å resolution, our study stands as the only human reliable evidence for nucleosomal DNA detachment in a productive PIC complex.

The yeast and human studies proposed the +1 nucleosome is a barrier and activator of transcription, respectively. The mechanism proposed for the yeast system was that the nucleosome stability reduces transcription, which was shown by performing a transcription assay with a native nucleosome positioning sequence (Wang et al., 2022), instead of the artificial Widom 601 (Lowary and Widom, 1998). Although the reduction in transcription observed was lower than with the artificial sequence, RNA production was reduced. On the other hand, Chen et al. (2022) performed transcription assays where they mutated subunits of the hook domain of Mediator, which they allegedly found interacting with H2A.Z-containing +1 nucleosome. Only when they carried out a severe truncation of MED26, by deleting ~470 amino acids from a 600 amino acids-long subunit, a strong reduction of transcription was described. Based on this data, they concluded that Mediator and TFIID orchestrate the assembly of PIC on the +1 nucleosome and enhance transcription. However, even though MED26 and TFIID have been reported to interact (Takahashi et al., 2011), their structural data does not explain such observed effect. Neither a TFIID-MED26 contact was determined, nor their resolution was high enough to determine whether MED26 is bound to the +1 nucleosome (Chen et al., 2022). In addition, reaching a conclusion from a transcription assay where ~80% of the highly disordered protein MED26 is deleted can be misleading. Instead, performing conservative substitutions to, for example, disrupt charge-based contacts, can better dissect the function of MED26 in this context.

With the work presented here, we fill in missing gaps of the mentioned studies (Chen et al., 2022; Wang et al., 2022). They did not investigate the role and mechanism of a +1 nucleosome closely located to the promoter – a promoter-proximal +1 nucleosome. We describe a reduction of transcription, as extensively researched *in vitro* (Wasylyk and Chambon, 1979; Wasylyk et al., 1979; Knezetic and Luse, 1986; Lorch et al., 1987; Workman and Roeder, 1987; Knezetic et al., 1988; Fisher and Luse, 2023) and *in vivo* (Han and Grunstein, 1988; Guenther et al., 2007; Schones et al., 2008; Valouev et al., 2011; Korber and Barbaric, 2014; Klein-Brill et al., 2019; Kubik et al., 2019; Schwartz et al., 2019; Serizay et al., 2020), but we aimed to additionally dissect the mechanism for such transcriptional downregulation. Here we provide evidence that not only the high stability of this nucleosome impedes transcription, as shown for the yeast study (Wang et al., 2022), but also triggers conformational changes on the PIC machinery. A promoter-proximal +1 nucleosome changes its relative orientation to the PIC, makes PIC-DNA contacts incompatible, switches TFIIH into an inhibited state, interferes with productive PIC assembly and translates in a reduction of transcription. The presence of such nucleosome changes the interactions established with TFIIH, now mostly dominated by the TFIIH subunit p8. Therefore, TFIIH not only adopts different conformations but also contacts the nucleosome differently, depending on its position both in yeast (Wang et al., 2022) and mammals (Chen et al., 2022). Accordingly, all contacts described are long-range weak interactions, which agrees well with the fact that PIC is a short-lived complex (de Graaf et al., 2010; Sainsbury et al., 2015; Nogales et al., 2016; Zhang et al., 2016; Nguyen et al., 2021). We speculate that these transient, changeable interactions and states, are required for the +1 nucleosome to fine-tune transcription initiation. Overall, we propose a model where the +1 nucleosome distance to the promoter influences its relative orientation to the PIC and, therefore, the Pol II-chromatin crosstalk. As such, it now emerges that the +1 nucleosome can regulate transcription in numerous ways (**Figure 12**).

5.6 Transcribing through the +1 nucleosome

As discussed, nucleosomes comprise roadblocks for Pol II to overcome so that transcription takes place. *In vivo* studies suggest that Pol II can transcribe nucleosomes at a speed comparable to naked DNA (Li et al., 2007; Petesch and

Lis, 2012; Teves et al., 2014). *In vitro* reconstitutions, however, have struggled so far to dissect how Pol II can achieve this in cells (Dangkulwanich et al., 2013; Farnung et al., 2018; Kujirai et al., 2018; Chen et al., 2019). As *in vitro* studies suggest that the +1 nucleosome comprises a barrier for Pol II transcription, efforts have been aimed at elucidating the mechanism for this effect. Several laboratories have described the structural basis for Pol II stalling through the nucleosome, how elongation factors help relieve the barriers (Farnung et al., 2018; Kujirai et al., 2018; Ehara et al., 2019; Farnung et al., 2021) and, recently, how nucleosomes are maintained on DNA through Pol II transcription (Ehara et al., 2022; Filipovski et al., 2022). All these studies rely on a pre-formed DNA-RNA hybrid that allows Pol II to start transcribing from a fixed location. Nevertheless, for Pol II nucleosome passage in cells, the PIC must assemble on chromatin and transcribe through the +1 nucleosome first.

Transcription occurs in promoters with a +1 nucleosome position at different locations, as we and others show (Workman and Roeder, 1987; Knezetic et al., 1988; Schones et al., 2008; Valouev et al., 2011; Fisher and Luse, 2023). It is unclear what the fate of a proximal or distal +1 nucleosome is after the burst of the first round of transcription initiation. We show that TFIIH is essential for nucleosome transcription initiation and, as discussed previously, TFIIH-independent transcription of a nucleosome template would cause clashes with the Pol II machinery. It is therefore evident from our superimposition that, even in the presence of TFIIH, the +1 nucleosome must undergo remodelling for initial RNA synthesis. Whereas we cannot rule out that transcription factors and/or chromatin remodellers might evict the +1 nucleosome, we hypothesise that the formation of a hexasome could help in surpassing the initial nucleosome barrier. Hexasome formation implies the loss of one H2A-H2B dimer copy and has been widely described in Pol II nucleosome passage (Kireeva et al., 2002; Bintu et al., 2011; Bevington and Boyes, 2013; Hsieh et al., 2013; Cole et al., 2014). Similarly, we lack structural insights into how the +1 nucleosome regulates all stages of Pol II transcription initiation – PIC assembly, DNA opening, ITC formation, initial RNA synthesis and promoter escape. In the future, to dissect transcription initiation through chromatin, it would therefore be required to determine transition states of the PIC transcribing into +1 nucleosome-containing promoter DNA templates.

6. Future directions and open questions

The fate of the +1 nucleosome after transcription initiation is the immediate question that arises after the observations we made, however, many other enigmas remain to be answered.

6.1 PIC-activators structural studies

Extensive research has been conducted on the role of Mediator as a messenger protein to activate transcription and, to a lesser extent, the co-activator TFIID (Johnson et al., 2002; Malik and Roeder, 2023). A great deal of evidence has showed that the Mediator complex contains several hot spots for transcription factors. Predominantly, the tail module has been described as one of the main hot spots for transcription factors, as it bears subunits such as MED15, MED25 and MED23 (Soutourina, 2018; Richter et al., 2022). For example, proteins of the ETS transcription factors family, yeast Gcn5 and the osteoblast differentiation-driving factor RUNX2 dynamically interact with MED25, MED15 and MED23, respectively (Liu et al., 2016; Henley et al., 2020; Tuttle et al., 2021). Not only the tail module of Mediator can bridge these interactions, but also subunits of the middle module like MED1 (Fondell et al., 1996; Hittelman et al., 1999; Wang et al., 2002; Chu et al., 2020). MED1 is the Mediator subunit described as the main protein mediating interactions with activators such as nuclear receptors and B cell-specific transcription factors (Belorusova et al., 2020; Chu et al., 2020). Although TFIID-activator interactions have not been studied as extensively as for the Mediator complex, a growing body of evidence indicates TFIID might also comprise an important bridge for transcription factors. To this end, researchers have observed that TAFs contribute to activator-dependent transcription (Grunberg et al., 2016; Warfield et al., 2017) and identified TAF4 to interact with several activators such as ETO activating domains and E proteins (Zhang et al., 2004; Wright and Tjian, 2009; Chen et al., 2013). Recently, a high-throughput affinity-purification mass spectrometry analysis, combined with proximity-dependent biotinylation (BioID), determined a plethora of transcription factors binding to TAF4, 5, 6, 9 and 12, which are present in two copies within TFIID (Goos et al., 2022).

Additionally, many of these coactivators' subunits, which interact with transcription factors, contain large intrinsically disordered regions (IDRs). In fact, roughly ~70-90% of MED15, MED1 and TAF4 are predicted to fold in a disordered fashion (Patel et al., 2018; Nayak and Taatjes, 2022). For most of their folding, transcription factors also present many IDRs (Liu et al., 2006). Even though these regions might potentially become ordered upon coactivator-activator interactions, this represents a major challenge to tackle in the near-future. To date, we lack fundamental structural insights on how these large assemblies work together in three-dimensional space. Recently, a yeast structure described how Gal4-VP16 homodimer induces the dimerization of a Mediator-Pol II complex on an upstream activating sequence (UAS) (Gorbea Colon et al., 2023). The authors suggested this dimer might have an activator-dependent role in the control of divergent transcription. Also, even though Mediator was co-purified with VP16 in a human Mediator-PIC cryo-EM structural study, no visible density was determined during data processing (Abdella et al., 2021). Besides, another cryo-EM study observed density for DNA-bound p53 on a TFIID-PIC complex (Chen et al., 2021a). However, no interactions were observed with any of the TAFs, likely due to the flexibility of the complex and/or the low resolution obtained for the p53-bound region. In line with this, cohesin and CTCF have been described to mediate the formation and maintenance of enhancer-promoter chromatin loops (Finn and Misteli, 2019; Szabo et al., 2019; Misteli, 2020; Oudelaar and Higgs, 2021). This affects how the Mediator complex engages distal enhancers and is ultimately recruited to the promoter (Richter et al., 2022). To understand the enhancer-promoter interplay, it would therefore be important to structurally characterise an enhancer-promoter chromatinised assembly. This is essential to elucidate 1) how the chromatin genome organisation-transcription crosstalk functions and 2) how transcription factors and chromatin contribute to transcription activation. It is probable that employing both integrated structural biology approaches, such as the combination of cryo-EM, NMR, and cross-linking mass spectrometry, with single-molecule and live imaging studies will be required to answer these questions. Only if we combine all these efforts, we will better understand how promoter-specific PIC recruitment enhancement and, thus, activation of transcription occurs.

6.2 Chromatin transcription intersections

Our work and the study from others begin to dissect the regulatory role of the +1 nucleosome in transcription initiation (Chen et al., 2022; Wang et al., 2022). However, as already mentioned, the Pol II transcription cycle can only be completely understood when we consider the promoter chromatin environment. To begin with, the current PIC-nucleosome structures and our work do not explain how PIC assembles on different chromatinised promoters. Do the promoter-flanking nucleosomes recruit the PIC machinery? Are these nucleosomes key in promoter recognition? If so, how, and which GTFs are involved?

To answer these questions, the role of the +1 nucleosome should be further studied, as its histone tails are enriched in post-translational modifications. For example, H3 lysine 4 is trimethylated (H3K4me3), lysine residues 5, 8, 12, and 16 of histone H4 get acetylated (H4K5ac, H4K8ac, H4K12ac, and H4K16ac) and 9, 14 and 18 are acetylated on histone H3 (H3K9ac, H3K14ac, and H3K18ac) (Kim et al., 2005; Heintzman et al., 2007; Schones et al., 2008). TFIID human TAF3 and TAF1 subunits bear a plant homeodomain (PHD) and a double bromodomain which recognise H3K4me3 and multiply acetylated H4 K5/K8/K12/K16, respectively (Jacobson et al., 2000; Vermeulen et al., 2007; van Ingen et al., 2008). Structural studies have not visualised the dynamic interactions between TFIID and the +1 nucleosome yet and, therefore, awaits further investigation (Chen et al., 2022). Together with the role of the nucleosome located immediately upstream of the PIC, the -1 nucleosome (Jiang and Pugh, 2009b; Struhl and Segal, 2013; Hughes and Rando, 2014; Lai and Pugh, 2017; Serizay et al., 2020; Chen et al., 2022), must be investigated. Such investigations would help us improve our understanding of how different promoters are recognised and Pol II recruited to them. In this context, it would be of great interest to study it in combination with the presence of transcriptional activators, given that these are commonly positioned upstream of the promoter region. It is unclear how the trajectory of the upstream DNA would be in an activator-bound PIC complex, as some transcription factors are known to induce DNA bending (van der Vliet and Verrijzer, 1993). It is evident that a kink occurring upstream of the TBP-DNA promoter region would affect the folding and crosstalk of the -1 nucleosome with the PIC machinery. Therefore, structural efforts should be

aimed at determining TFIID-PIC-Mediator-activator complexes in the presence of the -1 and +1 nucleosome. This would go a long way towards gaining insights into the combinatorial role promoter-flanking nucleosomes might have on PIC assembly and, thus, transcription initiation. Similarly, including further downstream nucleosomes on these assemblies, such as the +2, +3 and so forth, would likely give us a much better understanding of chromatin transcription initiation. To achieve this in the long term, it would be important to establish a system where a large chromatinised assembly (e.g. a whole genome) is transcribed *in vitro* or *in vivo* and subjected to a combination of single-particle cryo-EM and electron cryotomography (cryo-ET). In fact, it might change the conceptual advances already made, as we know that, for example, the transcription barrier is lower for downstream nucleosomes *in vivo* (Teves et al., 2014). Would it be possible that the whole chromatin environment affects how the transcription machinery functions and its working speed?

7. References

1. Aibara, S., Schilbach, S., and Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. Nature *594*, 124-128.
2. Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W.*, et al.* (2021a). Structural insights into preinitiation complex assembly on core promoters. Science *372*.
3. Greber, B.J., Toso, D.B., Fang, J., and Nogales, E. (2019). The complete structure of the human TFIIH core complex. Elife *8*.
4. Wang, H., Schilbach, S., Ninov, M., Urlaub, H., and Cramer, P. (2022). Structures of transcription preinitiation complex engaged with the +1 nucleosome. Nat Struct Mol Biol.
5. Chen, X., Wang, X., Liu, W., Ren, Y., Qu, X., Li, J., Yin, X., and Xu, Y. (2022). Structures of +1 nucleosome-bound PIC-Mediator complex. Science *378*, 62-68.
6. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.
7. He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature *533*, 359-365.
8. Gerace, L., and Burke, B. (1988). Functional organization of the nuclear envelope. Annu Rev Cell Biol *4*, 335-374.
9. Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.
10. Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.
11. Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol *9*, 85-98.
12. Woldringh, C.L. (2002). The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol Microbiol *45*, 17-29.
13. Gorlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol *15*, 607-660.
14. Flemming, W. (1882). Zellsubstanz, kern und zelltheilung (Vogel).
15. Miescher, J.F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen.
16. Kossel, A. (1884). Ueber einen peptonartigen Bestandtheil des Zellkerns.
17. Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. J Exp Med *79*, 137-158.
18. Franklin, R.E., and Gosling, R.G. (1953). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. Nature *172*, 156-157.
19. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.
20. Stedman, E., and Stedman, E. (1951). The basic proteins of cell nuclei. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences *235*, 565-595.
21. Kelley, R.I. (1973). Isolation of a histone IIb1-IIb2 complex. Biochem Biophys Res Commun *54*, 1588-1594.
22. Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. Science *184*, 865-868.
23. Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.
24. Hewish, D.R., and Burgoyne, L.A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem Biophys Res Commun *52*, 504-510.
25. Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.
26. Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell *4*, 281-300.
27. Phillips, D.M., and Johns, E.W. (1965). A Fractionation of the Histones of Group F2a from Calf Thymus. Biochem J *94*, 127-130.
28. Bram, S., and Ris, H. (1971). On the structure of nucleohistone. J Mol Biol *55*, 325-336.
29. Huberman, J.A. (1973). Structure of chromosome fibers and chromosomes. Annu Rev Biochem *42*, 355-378.
30. Finch, J.T., Noll, M., and Kornberg, R.D. (1975). Electron microscopy of defined lengths of chromatin. Proc Natl Acad Sci U S A *72*, 3320-3322.

31. Germond, J.E., Hirt, B., Oudet, P., Gross-Bellark, M., and Chambon, P. (1975). Folding of the DNA double helix in chromatin-like structures from simian virus 40. Proc Natl Acad Sci U S A *72*, 1843-1847.
32. Woodcock, C.L., Frado, L.L., Hatch, C.L., and Ricciardiello, L. (1976). Fine structure of active ribosomal genes. Chromosoma *58*, 33-39.
33. Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.
34. van Holde, K.E., and van Holde, K.E. (1989). The proteins of chromatin. I. Histones. Chromatin, 69-180.
35. Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.
36. Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. Nature *311*, 532-537.
37. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
38. Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.
39. Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J Mol Biol *327*, 85-96.
40. Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.
41. Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.
42. Harp, J.M., Hanson, B.L., Timm, D.E., and Bunick, G.J. (2000). Asymmetries in the nucleosome core particle at 2.5 A resolution. Acta Crystallogr D Biol Crystallogr *56*, 1513-1534.
43. White, C.L., Suto, R.K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. EMBO J *20*, 5207-5218.
44. Tsunaka, Y., Kajimura, N., Tate, S., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res *33*, 3424-3434.
45. Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. J R Soc Interface *10*, 20121022.
46. McGinty, R.K., and Tan, S. (2016). Recognition of the nucleosome by chromatin factors and enzymes. Curr Opin Struct Biol *37*, 54-61.
47. McGinty, R.K., and Tan, S. (2021). Principles of nucleosome recognition by chromatin factors and enzymes. Curr Opin Struct Biol *71*, 16-26.
48. Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Curr Opin Genet Dev *12*, 162-169.
49. Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014). Chromatin as dynamic 10-nm fibers. Chromosoma *123*, 225-237.
50. Maeshima, K., Hihara, S., and Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? Curr Opin Cell Biol *22*, 291-297.
51. Fussner, E., Ching, R.W., and Bazett-Jones, D.P. (2011). Living without 30nm chromatin fibers. Trends Biochem Sci *36*, 1-6.
52. Maeshima, K., Ide, S., and Babokhov, M. (2019). Dynamic chromatin organization without the 30-nm fiber. Curr Opin Cell Biol *58*, 95-104.
53. Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897-1901.
54. Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.
55. Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science *344*, 376-380.
56. Ekundayo, B., Richmond, T.J., and Schalch, T. (2017). Capturing Structural Heterogeneity in Chromatin Fibers. J Mol Biol *429*, 3031-3042.
57. Garcia-Saez, I., Menoni, H., Boopathi, R., Shukla, M.S., Soueidan, L., Noirclerc-Savoye, M., Le Roy, A., Skoufias, D.A., Bednar, J., Hamiche, A.*, et al.* (2018). Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. Mol Cell *72*, 902-915 e907.

58. Adhireksan, Z., Sharma, D., Lee, P.L., and Davey, C.A. (2020). Near-atomic resolution structures of interdigitated nucleosome fibres. Nat Commun *11*, 4747.
59. Dombrowski, M., Engeholm, M., Dienemann, C., Dodonova, S., and Cramer, P. (2022). Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. Nat Struct Mol Biol *29*, 493-501.
60. Zhou, K., Gebala, M., Woods, D., Sundararajan, K., Edwards, G., Krzizike, D., Wereszczynski, J., Straight, A.F., and Luger, K. (2022). CENP-N promotes the compaction of centromeric chromatin. Nat Struct Mol Biol *29*, 403-413.
61. McDowall, A.W., Smith, J.M., and Dubochet, J. (1986). Cryo-electron microscopy of vitrified chromosomes in situ. EMBO J *5*, 1395-1402.
62. Dekker, J. (2008). Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. J Biol Chem *283*, 34532-34540.
63. Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T.*, et al.* (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO J *31*, 1644-1653.
64. Ricci, M.A., Manzo, C., Garcia-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell *160*, 1145-1158.
65. Sedat, J., and Manuelidis, L. (1978). A direct approach to the structure of eukaryotic chromosomes. Cold Spring Harb Symp Quant Biol *42 Pt 1*, 331-350.
66. Mirny, L.A., Imakaev, M., and Abdennur, N. (2019). Two major mechanisms of chromosome organization. Curr Opin Cell Biol *58*, 142-152.
67. Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol Cell *62*, 668-680.
68. Lupianez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet *32*, 225-237.
69. Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20-32.
70. Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. Sci Adv *5*, eaaw1668.
71. Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. Nat Struct Mol Biol *26*, 3-13.
72. Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 129-143.
73. Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.
74. Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.
75. Jiang, C., and Pugh, B.F. (2009a). A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. Genome Biol *10*, R109.
76. Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature *474*, 516-520.
77. Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.
78. Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.
79. Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome Res *18*, 1084-1091.
80. Workman, J.L. (2006). Nucleosome displacement in transcription. Genes Dev *20*, 2009-2017.
81. Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186-192.
82. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.
83. Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.
84. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat Rev Genet *20*, 207-220.
85. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev *25*, 2227-2241.

86. Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev *28*, 2679-2692.
87. Balsalobre, A., and Drouin, J. (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol *23*, 449-464.
88. Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr Opin Genet Dev *14*, 165-173.
89. Smith, C.L., and Peterson, C.L. (2005). ATP-dependent chromatin remodeling. Curr Top Dev Biol *65*, 115-148.
90. Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol *7*, 437-447.
91. Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.
92. Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2014). Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. Genes Dev *28*, 2492-2497.
93. Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. Cell *167*, 709-721 e712.
94. Kubik, S., Bruzzone, M.J., Challal, D., Dreos, R., Mattarocci, S., Bucher, P., Libri, D., and Shore, D. (2019). Opposing chromatin remodelers control transcription initiation frequency and start site selection. Nat Struct Mol Biol *26*, 744-754.
95. Gamarra, N., and Narlikar, G.J. (2021). Collaboration through chromatin: motors of transcription and chromatin structure. J Mol Biol *433*, 166876.
96. Oberbeckmann, E., Krietenstein, N., Niebauer, V., Wang, Y., Schall, K., Moldt, M., Straub, T., Rohs, R., Hopfner, K.P., Korber, P.*, et al.* (2021a). Genome information processing by the INO80 chromatin remodeler positions nucleosomes. Nat Commun *12*, 3231.
97. Farnung, L., Vos, S.M., Wigge, C., and Cramer, P. (2017). Nucleosome-Chd1 structure and implications for chromatin remodelling. Nature *550*, 539-542.
98. Farnung, L., Ochmann, M., Engeholm, M., and Cramer, P. (2021). Structural basis of nucleosome transcription mediated by Chd1 and FACT. Nat Struct Mol Biol *28*, 382-387.
99. Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell *9*, 541-552.
100. Bintu, L., Kopaczynska, M., Hodges, C., Lubkowska, L., Kashlev, M., and Bustamante, C. (2011). The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. Nat Struct Mol Biol *18*, 1394-1399.
101. Bevington, S., and Boyes, J. (2013). Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J *32*, 1381-1392.
102. Hsieh, F.K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M. (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. Proc Natl Acad Sci U S A *110*, 7654-7659.
103. Cole, H.A., Ocampo, J., Iben, J.R., Chereji, R.V., and Clark, D.J. (2014). Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases. Nucleic Acids Res *42*, 12512-12522.
104. Hsieh, L.J., Gourdet, M.A., Moore, C.M., Munoz, E.N., Gamarra, N., Ramani, V., and Narlikar, G.J. (2022). A hexasome is the preferred substrate for the INO80 chromatin remodeling complex, allowing versatility of function. Mol Cell *82*, 2098-2112 e2094.
105. Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200-213.
106. Wang, F., Ranjan, A., Wei, D., and Wu, C. (2016). Comment on "A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme". Science *353*, 358.
107. Brahma, S., Udugama, M.I., Kim, J., Hada, A., Bhardwaj, S.K., Hailu, S.G., Lee, T.H., and Bartholomew, B. (2017). INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. Nat Commun *8*, 15616.
108. Ranjan, A., Nguyen, V.Q., Liu, S., Wisniewski, J., Kim, J.M., Tang, X., Mizuguchi, G., Elalaoui, E., Nickels, T.J., Jou, V.*, et al.* (2020). Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. Elife *9*.
109. Krogan, N.J., Keogh, M.C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V.*, et al.* (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell *12*, 1565-1576.

110. Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell *154*, 1232-1245.
111. Yen, K., Vinayachandran, V., and Pugh, B.F. (2013). SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. Cell *154*, 1246-1256.
112. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
113. Schwartz, U., Nemeth, A., Diermeier, S., Exler, J.H., Hansch, S., Maldonado, R., Heizinger, L., Merkl, R., and Langst, G. (2019). Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. Nucleic Acids Res *47*, 1239-1254.
114. Serizay, J., Dong, Y., Janes, J., Chesney, M., Cerrato, C., and Ahringer, J. (2020). Distinctive regulatory architectures of germline-active and somatic genes in C. elegans. Genome Res *30*, 1752-1765.
115. Henikoff, S. (2000). Heterochromatin function in complex genomes. Biochim Biophys Acta *1470*, O1-8.
116. Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.
117. Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol *19*, 229-244.
118. Morrison, O., and Thakur, J. (2021). Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. Int J Mol Sci *22*.
119. Penagos-Puig, A., and Furlan-Magaril, M. (2020). Heterochromatin as an Important Driver of Genome Organization. Frontiers in Cell and Developmental Biology *8*.
120. Jiang, C., and Pugh, B.F. (2009b). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.
121. Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.
122. Hughes, A.L., and Rando, O.J. (2014). Mechanisms underlying nucleosome positioning in vivo. Annu Rev Biophys *43*, 41-63.
123. Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat Rev Mol Cell Biol *18*, 548-562.
124. Kornberg, R.D., and Lorch, Y. (2020). Primary Role of the Nucleosome. Mol Cell *79*, 371-375.
125. Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y.H., and Reinberg, D. (2010). Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. Mol Cell *38*, 41-53.
126. Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *483*, 295-301.
127. Tramantano, M., Sun, L., Au, C., Labuz, D., Liu, Z., Chou, M., Shen, C., and Luk, E. (2016). Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. Elife *5*.
128. Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet *49*, 1045-1051.
129. Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R*., et al.* (2021). A high-resolution protein architecture of the budding yeast genome. Nature *592*, 309-314.
130. Santana, J.F., Collins, G.S., Parida, M., Luse, D.S., and Price, D.H. (2022). Differential dependencies of human RNA polymerase II promoters on TBP, TAF1, TFIIB and XPB. Nucleic Acids Res.
131. Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.
132. Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell *53*, 819-830.
133. Zumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., and Cramer, P. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. Mol Cell *81*, 3096-3109 e3098.
134. Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. Cell *55*, 1137-1145.
135. Korber, P., and Barbaric, S. (2014). The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. Nucleic Acids Res *42*, 10888-10902.

136. Klein-Brill, A., Joseph-Strauss, D., Appleboim, A., and Friedman, N. (2019). Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. Cell Rep *26*, 279-292 e275.
137. Knezetic, J.A., and Luse, D.S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. Cell *45*, 95-104.
138. Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell *49*, 203-210.
139. Workman, J.L., and Roeder, R.G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell *51*, 613-622.
140. Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature *450*, 1031-1035.
141. Hennig, B.P., Bendrin, K., Zhou, Y., and Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. EMBO Rep *13*, 997-1003.
142. Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Stralfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. EMBO J *31*, 4388-4403.
143. Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. Biochim Biophys Acta *1829*, 84-97.
144. Cheetham, G.M., and Steitz, T.A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Curr Opin Struct Biol *10*, 117-123.
145. Rosa, M.D. (1979). Four T7 RNA polymerase promoters contain an identical 23 bp sequence. Cell *16*, 815-825.
146. Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. Nature *364*, 593-599.
147. Cheetham, G.M., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature *399*, 80-83.
148. Kang, J.G., Hahn, M.Y., Ishihama, A., and Roe, J.H. (1997). Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from Streptomyces coelicolor A3(2). Nucleic Acids Res *25*, 2566-2573.
149. Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811-824.
150. Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol *68*, 357-376.
151. Murakami, K.S. (2015). Structural biology of bacterial RNA polymerase. Biomolecules *5*, 848-864.
152. Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.
153. Kedinger, C., Gniazdowski, M., Mandel, J.L., Jr., Gissinger, F., and Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochem Biophys Res Commun *38*, 165-171.
154. Engelke, D.R., Ng, S.Y., Shastry, B.S., and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell *19*, 717-728.
155. Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem *255*, 11992-11996.
156. Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.A., and Yamamoto, K.R. (1981). Purified glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo. Proc Natl Acad Sci U S A *78*, 6628-6632.
157. Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell *35*, 79-87.
158. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.
159. Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. J Biol Chem *267*, 2786-2793.
160. Flanagan, P.M., Kelleher, R.J., 3rd, Sayre, M.H., Tschochner, H., and Kornberg, R.D. (1991). A mediator required for activation of RNA polymerase II transcription in vitro. Nature *350*, 436-438.
161. Meisterernst, M., Roy, A.L., Lieu, H.M., and Roeder, R.G. (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell *66*, 981-993.

162. Thompson, C.M., Koleske, A.J., Chao, D.M., and Young, R.A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. Cell *73*, 1361-1375.
163. Verrijzer, C.P., Chen, J.L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. Cell *81*, 1115-1125.
164. Grant, P.A., Duggan, L., Cote, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F.*, et al.* (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev *11*, 1640-1650.
165. Ryu, S., and Tjian, R. (1999). Purification of transcription cofactor complex CRSP. Proc Natl Acad Sci U S A *96*, 7137-7142.
166. Ryu, S., Zhou, S., Ladurner, A.G., and Tjian, R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. Nature *397*, 446-450.
167. Malik, S., Gu, W., Wu, W., Qin, J., and Roeder, R.G. (2000). The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. Mol Cell *5*, 753-760.
168. Larschan, E., and Winston, F. (2001). The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4. Genes Dev *15*, 1946-1956.
169. Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.
170. Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet *37*, 761-765.
171. Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol *12*, 483-492.
172. Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet *15*, 394-408.
173. Masters, B.S., Stohl, L.L., and Clayton, D.A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell *51*, 89-99.
174. Asin-Cayuela, J., and Gustafsson, C.M. (2007). Mitochondrial transcription and its regulation in mammalian cells. Trends Biochem Sci *32*, 111-117.
175. Yagi, Y., and Shiina, T. (2014). Recent advances in the study of chloroplast gene expression and its evolution. Front Plant Sci *5*, 61.
176. Hillen, H.S., Temiakov, D., and Cramer, P. (2018). Structural basis of mitochondrial transcription. Nat Struct Mol Biol *25*, 754-765.
177. Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell *45*, 439-446.
178. Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nat Struct Mol Biol *26*, 783-791.
179. Hannan, K.M., Hannan, R.D., and Rothblum, L.I. (1998). Transcription by RNA polymerase I. Front Biosci *3*, d376-398.
180. Goodfellow, S.J., and Zomerdijk, J.C. (2013). Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. Subcell Biochem *61*, 211-236.
181. Neyer, S., Kunz, M., Geiss, C., Hantsche, M., Hodirnau, V.V., Seybert, A., Engel, C., Scheffer, M.P., Cramer, P., and Frangakis, A.S. (2016). Structure of RNA polymerase I transcribing ribosomal DNA genes. Nature *540*, 607-610.
182. Misiaszek, A.D., Girbig, M., Grotsch, H., Baudin, F., Murciano, B., Lafita, A., and Muller, C.W. (2021). Cryo-EM structures of human RNA polymerase I. Nat Struct Mol Biol *28*, 997-1008.
183. Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev *16*, 2593-2620.
184. Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. Trends Genet *23*, 614-622.
185. Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassavetis, G.A., and Cramer, P. (2010). Molecular basis of RNA polymerase III transcription repression by Maf1. Cell *143*, 59-70.
186. Arimbasseri, A.G., and Maraia, R.J. (2016). RNA Polymerase III Advances: Structural and tRNA Functional Views. Trends Biochem Sci *41*, 546-559.
187. Abascal-Palacios, G., Ramsay, E.P., Beuron, F., Morris, E., and Vannini, A. (2018). Structural basis of RNA polymerase III transcription initiation. Nature *553*, 301-306.
188. Ramsay, E.P., Abascal-Palacios, G., Daiss, J.L., King, H., Gouge, J., Pilsl, M., Beuron, F., Morris, E., Gunkel, P., Engel, C.*, et al.* (2020). Structure of human RNA polymerase III. Nat Commun *11*, 6409.

189. Girbig, M., Misiaszek, A.D., Vorlander, M.K., Lafita, A., Grotsch, H., Baudin, F., Bateman, A., and Muller, C.W. (2021). Cryo-EM structures of human RNA polymerase III in its unbound and transcribing states. Nat Struct Mol Biol *28*, 210-219.
190. Huang, K., Wu, X.X., Fang, C.L., Xu, Z.G., Zhang, H.W., Gao, J., Zhou, C.M., You, L.L., Gu, Z.X., Mu, W.H.*, et al.* (2021). Pol IV and RDR2: A two-RNA-polymerase machine that produces double-stranded RNA. Science *374*, 1579-1586.
191. Xie, G., Du, X., Hu, H., Li, S., Cao, X., Jacobsen, S.E., and Du, J. (2023). Structure and mechanism of the plant RNA polymerase V. Science *379*, 1209-1213.
192. Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640-649.
193. Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.
194. Cramer, P. (2004). Structure and function of RNA polymerase II. Adv Protein Chem *67*, 1-42.
195. Liu, X., Bushnell, D.A., and Kornberg, R.D. (2013). RNA polymerase II transcription: structure and mechanism. Biochim Biophys Acta *1829*, 2-8.
196. Osman, S., and Cramer, P. (2020). Structural Biology of RNA Polymerase II Transcription: 20 Years On. Annu Rev Cell Dev Biol *36*, 1-34.
197. Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.
198. Livingstone, M., Atas, E., Meller, A., and Sonenberg, N. (2010). Mechanisms governing the control of mRNA translation. Phys Biol *7*, 021001.
199. Schmid, M., and Jensen, T.H. (2010). Nuclear quality control of RNA polymerase II transcripts. Wiley Interdiscip Rev RNA *1*, 474-485.
200. Shandilya, J., and Roberts, S.G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. Biochim Biophys Acta *1819*, 391-400.
201. Hantsche, M., and Cramer, P. (2016). The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. Angew Chem Int Ed Engl *55*, 15972-15981.
202. Conaway, R.C., and Conaway, J.W. (1993). General initiation factors for RNA polymerase II. Annu Rev Biochem *62*, 161-190.
203. Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci *21*, 327-335.
204. Grunberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci *38*, 603-611.
205. Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science *288*, 1418-1422.
206. Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.
207. He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481-486.
208. Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc Natl Acad Sci U S A *112*, 3961-3966.
209. Nogales, E., Louder, R.K., and He, Y. (2017a). Structural Insights into the Eukaryotic Transcription Initiation Machinery. Annu Rev Biophys *46*, 59-83.
210. Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature *551*, 204-209.
211. Dienemann, C., Schwalb, B., Schilbach, S., and Cramer, P. (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. Mol Cell *73*, 97-106 e104.
212. Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F., and Cramer, P. (2021). Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. Cell *184*, 4064-4072 e4028.
213. Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. Curr Opin Struct Biol *19*, 294-299.
214. Luse, D.S., and Jacob, G.A. (1987). Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. J Biol Chem *262*, 14990-14997.
215. Holstege, F.C., Fiedler, U., and Timmers, H.T. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J *16*, 7468-7480.

216. Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim Biophys Acta *1829*, 63-68.
217. Gilmour, D.S., and Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Mol Cell Biol *6*, 3984-3989.
218. Strobl, L.J., and Eick, D. (1992). Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. EMBO J *11*, 3307-3314.
219. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.
220. Liu, X., Kraus, W.L., and Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. Trends Biochem Sci *40*, 516-525.
221. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F*., et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev *12*, 343-356.
222. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.
223. Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.K., Hasegawa, J., Omori, M., Inukai, N*., et al.* (2003). Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol *23*, 1863-1873.
224. Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062-1066.
225. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018a). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601-606.
226. Reines, D., Ghanouni, P., Li, Q.Q., and Mote, J., Jr. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. J Biol Chem *267*, 15516-15522.
227. Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell *17*, 103-112.
228. Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.
229. Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem *270*, 12335-12338.
230. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451-462.
231. Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem *271*, 27176-27183.
232. Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Biol *24*, 787-795.
233. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell *21*, 227-237.
234. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018b). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607-612.
235. Vos, S.M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. Nat Struct Mol Biol *27*, 668-677.
236. Neugebauer, K.M. (2002). On the importance of being co-transcriptional. J Cell Sci *115*, 3865-3871.
237. Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645-653.
238. Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5'-Terminal structure and mRNA stability. Nature *266*, 235-239.
239. Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.

240. Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol *66*, 1-40.
241. Cowling, V.H. (2009). Regulation of mRNA cap methylation. Biochem J *425*, 295-302.
242. Ramanathan, A., Robb, G.B., and Chan, S.H. (2016). mRNA capping: biological functions and applications. Nucleic Acids Res *44*, 7511-7526.
243. Gorlich, D., Kraft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R.A., Mattaj, I.W., and Izaurralde, E. (1996). Importin provides a link between nuclear protein import and U snRNA export. Cell *87*, 21-32.
244. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., and Mattaj, I.W. (1996). A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J Cell Biol *133*, 5-14.
245. Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur J Biochem *247*, 461-469.
246. Wilkinson, M.E., Charenton, C., and Nagai, K. (2020). RNA Splicing by the Spliceosome. Annu Rev Biochem *89*, 359-388.
247. Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol *14*, 7219-7225.
248. Tennyson, C.N., Klamut, H.J., and Worton, R.G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet *9*, 184-190.
249. Kotovic, K.M., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol Cell Biol *23*, 5768-5779.
250. Bird, G., Zorio, D.A., and Bentley, D.L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Mol Cell Biol *24*, 8963-8969.
251. Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell *19*, 65-75.
252. Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol *13*, 815-822.
253. Wallace, E.W.J., and Beggs, J.D. (2017). Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. RNA *23*, 601-610.
254. Zhang, S., Aibara, S., Vos, S.M., Agafonov, D.E., Luhrmann, R., and Cramer, P. (2021a). Structure of a transcribing RNA polymerase II-U1 snRNP complex. Science *371*, 305-309.
255. Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol Cell *36*, 541-546.
256. Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev *26*, 2119-2137.
257. Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.
258. Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. Nat Struct Mol Biol *23*, 771-777.
259. Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Biol *18*, 263-273.
260. Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D*., et al.* (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep *20*, 1173-1186.
261. Fant, C.B., and Taatjes, D.J. (2019). Regulatory functions of the Mediator kinases CDK8 and CDK19. Transcription *10*, 76-90.
262. Fisher, R.P. (2019). Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery. Transcription *10*, 47-56.
263. Luyties, O., and Taatjes, D.J. (2022). The Mediator kinase module: an interface between cell signaling and transcription. Trends Biochem Sci *47*, 314-327.
264. Richter, W.F., Nayak, S., Iwasa, J., and Taatjes, D.J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. Nat Rev Mol Cell Biol *23*, 732-749.
265. Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell *11*, 1549-1561.
266. Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem *282*, 14113-14120.

267. Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell *2*, 43-53.
268. Liu, Y., Kung, C., Fishburn, J., Ansari, A.Z., Shokat, K.M., and Hahn, S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol Cell Biol *24*, 1721-1735.
269. Osman, S., Mohammad, E., Lidschreiber, M., Stuetzer, A., Bazso, F.L., Maier, K.C., Urlaub, H., and Cramer, P. (2021). The Cdk8 kinase module regulates interaction of the mediator complex with RNA polymerase II. J Biol Chem *296*, 100734.
270. Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell *23*, 297-305.
271. Viladevall, L., St Amour, C.V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J.J., Shokat, K.M., Schwer, B., Leatherwood, J.K., and Fisher, R.P. (2009). TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. Mol Cell *33*, 738-751.
272. Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem Sci *14*, 105-110.
273. Proudfoot, N. (1996). Ending the message is not so simple. Cell *87*, 779-781.
274. Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.
275. Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.
276. Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev *34*, 132-145.
277. Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet *36*, 664-675.
278. Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature *444*, 953-956.
279. Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell *33*, 365-376.
280. Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 3rd, Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. Genes Dev *28*, 2370-2380.
281. Schonemann, L., Kuhn, U., Martin, G., Schafer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev *28*, 2381-2393.
282. Clerici, M., Faini, M., Aebersold, R., and Jinek, M. (2017). Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. Elife *6*.
283. Kecman, T., Kus, K., Heo, D.H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C.V., and Vasiljeva, L. (2018). Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. Cell Rep *25*, 259-269 e255.
284. Parua, P.K., Booth, G.T., Sanso, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. Nature *558*, 460-464.
285. Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. Mol Cell *76*, 896-908 e894.
286. West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.
287. Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J.*, et al.* (2012). mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol Cell *46*, 311-324.
288. Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev *32*, 127-139.
289. Eaton, J.D., and West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. Transcription *9*, 321-326.

290. Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.
291. Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. Mol Cell *56*, 128-139.
292. Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. Nature *525*, 399-403.
293. Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. Cell Res *25*, 288-305.
294. Elrod, N.D., Henriques, T., Huang, K.L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. Mol Cell *76*, 738-752 e737.
295. Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends Genet *35*, 553-564.
296. Rubtsova, M.P., Vasilkova, D.P., Moshareva, M.A., Malyavko, A.N., Meerson, M.B., Zatsepin, T.S., Naraykina, Y.V., Beletsky, A.V., Ravin, N.V., and Dontsova, O.A. (2019). Integrator is a key component of human telomerase RNA biogenesis. Sci Rep *9*, 1701.
297. Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.S., Jiang, J.Z., Jonathan, M., Huang, K.L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. Genes Dev *33*, 1525-1538.
298. Beckedorff, F., Blumenthal, E., daSilva, L.F., Aoi, Y., Cingaram, P.R., Yue, J., Zhang, A., Dokaneheifard, S., Valencia, M.G., Gaidosh, G.*, et al.* (2020). The Human Integrator Complex Facilitates Transcriptional Elongation by Endonucleolytic Cleavage of Nascent Transcripts. Cell Rep *32*, 107917.
299. Lykke-Andersen, S., Zumer, K., Molska, E.S., Rouviere, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., and Jensen, T.H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. Mol Cell *81*, 514-529 e516.
300. Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S.*, et al.* (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nat Commun *5*, 5531.
301. Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., and Yamaguchi, Y. (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. Nat Commun *5*, 4263.
302. Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., and Cramer, P. (2021). Structural basis of Integrator-mediated transcription regulation. Science *374*, 883-887.
303. Huang, K.L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K., and Wagner, E.J. (2020). Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. Mol Cell *80*, 345-358 e349.
304. Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J.*, et al.* (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. Science *370*.
305. Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F., and Wagner, E.J. (2012). An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. RNA *18*, 2148-2156.
306. Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dundr, M., and Wagner, E.J. (2018). Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module. Nucleic Acids Res *46*, 4241-4255.
307. Pfleiderer, M.M., and Galej, W.P. (2021). Structure of the catalytic core of the Integrator complex. Mol Cell *81*, 1246-1259 e1248.
308. Reinberg, D., Horikoshi, M., and Roeder, R.G. (1987). Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. J Biol Chem *262*, 3322-3330.
309. Van Dyke, M.W., Roeder, R.G., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science *241*, 1335-1338.
310. Chasman, D.I., Flaherty, K.M., Sharp, P.A., and Kornberg, R.D. (1993). Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc Natl Acad Sci U S A *90*, 8174-8178.

311. Zawel, L., and Reinberg, D. (1993). Initiation of transcription by RNA polymerase II: a multi-step process. Prog Nucleic Acid Res Mol Biol *44*, 67-108.
312. Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.
313. Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13-25.
314. Nikolov, D.B., Hu, S.H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. Nature *360*, 40-46.
315. Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520-527.
316. Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.
317. Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast TFIIA/TBP/DNA complex. Science *272*, 830-836.
318. Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. Nature *381*, 127-151.
319. Bleichenbacher, M., Tan, S., and Richmond, T.J. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J Mol Biol *332*, 783-793.
320. Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol *7*, R78.
321. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C*., et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.
322. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436.
323. Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52-65.
324. Vo Ngoc, L., Cassidy, C.J., Huang, C.Y., Duttke, S.H., and Kadonaga, J.T. (2017a). The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. Genes Dev *31*, 6-11.
325. Vo Ngoc, L., Wang, Y.L., Kassavetis, G.A., and Kadonaga, J.T. (2017b). The punctilious RNA polymerase II core promoter. Genes Dev *31*, 1289-1301.
326. Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol *19*, 621-637.
327. Dynlacht, B.D., Hoey, T., and Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell *66*, 563-576.
328. Andel, F., 3rd, Ladurner, A.G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-dimensional structure of the human TFIID-IIA-IIB complex. Science *286*, 2153-2156.
329. Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. Science *286*, 2151-2153.
330. Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., and Schultz, P. (2002). Mapping histone fold TAFs within yeast TFIID. EMBO J *21*, 3424-3433.
331. Sanders, S.L., Garbett, K.A., and Weil, P.A. (2002a). Molecular characterization of Saccharomyces cerevisiae TFIID. Mol Cell Biol *22*, 6000-6013.
332. Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev *16*, 673-675.
333. Leurent, C., Sanders, S.L., Demeny, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., and Schultz, P. (2004). Mapping key functional sites within yeast TFIID. EMBO J *23*, 719-727.
334. Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. Structure *14*, 511-520.
335. Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., and Schultz, P. (2009). Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. Structure *17*, 363-373.

336. Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. Nature *493*, 699-702.
337. Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. Cell *152*, 120-131.
338. Cianfrocco, M.A., and Nogales, E. (2013). Regulatory interplay between TFIID's conformational transitions and its modular interaction with core promoter DNA. Transcription *4*, 120-126.
339. Louder, R.K., He, Y., Lopez-Blanco, J.R., Fang, J., Chacon, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604-609.
340. Nogales, E., Fang, J., and Louder, R.K. (2017b). Structural dynamics and DNA interaction of human TFIID. Transcription *8*, 55-60.
341. Nogales, E., Patel, A.B., and Louder, R.K. (2017c). Towards a mechanistic understanding of core promoter recognition from cryo-EM studies of human TFIID. Curr Opin Struct Biol *47*, 60-66.
342. Kolesnikova, O., Ben-Shem, A., Luo, J., Ranish, J., Schultz, P., and Papai, G. (2018). Molecular structure of promoter-bound yeast TFIID. Nat Commun *9*, 4666.
343. Patel, A.B., Louder, R.K., Greber, B.J., Grunberg, S., Luo, J., Fang, J., Liu, Y., Ranish, J., Hahn, S., and Nogales, E. (2018). Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science *362*.
344. Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L. (1989). Identification of a yeast protein homologous in function to the mammalian general transcription factor, TFIIA. EMBO J *8*, 3379-3382.
345. Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A., and Roeder, R.G. (1989). Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature *341*, 299-303.
346. Goodrich, J.A., Cutler, G., and Tjian, R. (1996). Contacts in context: promoter specificity and macromolecular interactions in transcription. Cell *84*, 825-830.
347. Kuras, L., Kosa, P., Mencia, M., and Struhl, K. (2000). TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo. Science *288*, 1244-1248.
348. Sanders, S.L., Jennings, J., Canutescu, A., Link, A.J., and Weil, P.A. (2002b). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Mol Cell Biol *22*, 4723-4738.
349. Thut, C.J., Chen, J.L., Klemm, R., and Tjian, R. (1995). p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. Science *267*, 100-104.
350. Rojo-Niersbach, E., Furukawa, T., and Tanese, N. (1999). Genetic dissection of hTAF(II)130 defines a hydrophobic surface required for interaction with glutamine-rich activators. J Biol Chem *274*, 33778-33784.
351. Munz, C., Psichari, E., Mandilis, D., Lavigne, A.C., Spiliotaki, M., Oehler, T., Davidson, I., Tora, L., Angel, P., and Pintzas, A. (2003). TAF7 (TAFII55) plays a role in the transcription activation by c-Jun. J Biol Chem *278*, 21510-21516.
352. Li, H.H., Li, A.G., Sheppard, H.M., and Liu, X. (2004). Phosphorylation on Thr-55 by TAF1 mediates degradation of p53: a role for TAF1 in cell G1 progression. Mol Cell *13*, 867-878.
353. Lively, T.N., Nguyen, T.N., Galasinski, S.K., and Goodrich, J.A. (2004). The basic leucine zipper domain of c-Jun functions in transcriptional activation through interaction with the N terminus of human TATA-binding protein-associated factor-1 (human TAF(II)250). J Biol Chem *279*, 26257-26265.
354. Hilton, T.L., Li, Y., Dunphy, E.L., and Wang, E.H. (2005). TAF1 histone acetyltransferase activity in Sp1 activation of the cyclin D1 promoter. Mol Cell Biol *25*, 4321-4332.
355. Wang, X., Truckses, D.M., Takada, S., Matsumura, T., Tanese, N., and Jacobson, R.H. (2007). Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators. Proc Natl Acad Sci U S A *104*, 7839-7844.
356. Liu, W.L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R. (2009). Structures of three distinct activator-TFIID complexes. Genes Dev *23*, 1510-1521.
357. Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. Nature *465*, 956-960.
358. Goos, H., Kinnunen, M., Salokas, K., Tan, Z., Liu, X., Yadav, L., Zhang, Q., Wei, G.H., and Varjosalo, M. (2022). Human transcription factor protein interaction networks. Nat Commun *13*, 766.

359. Cortes, P., Flores, O., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. Mol Cell Biol *12*, 413-421.
360. Imbalzano, A.N., Zaret, K.S., and Kingston, R.E. (1994). Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. J Biol Chem *269*, 8280-8286.
361. Sawadogo, M., and Sentenac, A. (1990). RNA polymerase B (II) and general transcription factors. Annu Rev Biochem *59*, 711-754.
362. Buratowski, S., and Zhou, H. (1993). Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A *90*, 5633-5637.
363. Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev *12*, 34-44.
364. Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A *96*, 13668-13673.
365. Tsai, F.T., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. EMBO J *19*, 25-36.
366. Zhao, X., and Herr, W. (2002). A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. Cell *108*, 615-627.
367. Deng, W., and Roberts, S.G. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev *19*, 2418-2423.
368. Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*, 983-988.
369. Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323-330.
370. Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science *327*, 206-209.
371. Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*, 437-440.
372. Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A *88*, 9999-10003.
373. Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M.*, et al.* (2010a). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J *29*, 717-726.
374. Eichner, J., Chen, H.T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J *29*, 706-716.
375. Fishburn, J., and Hahn, S. (2012). Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. Mol Cell Biol *32*, 12-25.
376. Murakami, K., Tsai, K.L., Kalisman, N., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2015a). Structure of an RNA polymerase II preinitiation complex. Proc Natl Acad Sci U S A *112*, 13543-13548.
377. Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F.*, et al.* (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature *518*, 376-380.
378. Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. Nature *533*, 353-358.
379. Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., and Ponticelli, A.S. (2004). Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. Mol Cell Biol *24*, 10975-10985.
380. Cabart, P., Ujvari, A., Pal, M., and Luse, D.S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. Proc Natl Acad Sci U S A *108*, 15786-15791.
381. Tan, S., Aso, T., Conaway, R.C., and Conaway, J.W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J Biol Chem *269*, 25684-25691.
382. Forget, D., Langelier, M.F., Therien, C., Trinh, V., and Coulombe, B. (2004). Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. Mol Cell Biol *24*, 1122-1131.

383. Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. Nat Struct Mol Biol *13*, 603-610.
384. Kuldell, N.H., and Buratowski, S. (1997). Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. Mol Cell Biol *17*, 5288-5298.
385. Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol *14*, 696-703.
386. Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.
387. Holstege, F.C., Tantin, D., Carey, M., van der Vliet, P.C., and Timmers, H.T. (1995). The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. EMBO J *14*, 810-819.
388. Okuda, M., Tanaka, A., Satoh, M., Mizuta, S., Takazawa, M., Ohkuma, Y., and Nishimura, Y. (2008). Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. EMBO J *27*, 1161-1171.
389. Conaway, R.C., and Conaway, J.W. (1989). An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. Proc Natl Acad Sci U S A *86*, 7356-7360.
390. Weber, C.A., Salazar, E.P., Stewart, S.A., and Thompson, L.H. (1990). ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. EMBO J *9*, 1437-1447.
391. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H., Chambon, P., and Egly, J.M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science *260*, 58-63.
392. Roy, R., Adamczewski, J.P., Seroz, T., Vermeulen, W., Tassan, J.P., Schaeffer, L., Nigg, E.A., Hoeijmakers, J.H., and Egly, J.M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. Cell *79*, 1093-1101.
393. Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J.H., and Egly, J.M. (1994). The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. EMBO J *13*, 2388-2392.
394. Adamczewski, J.P., Rossignol, M., Tassan, J.P., Nigg, E.A., Moncollin, V., and Egly, J.M. (1996). MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH. EMBO J *15*, 1877-1884.
395. Feaver, W.J., Henry, N.L., Wang, Z., Wu, X., Svejstrup, J.Q., Bushnell, D.A., Friedberg, E.C., and Kornberg, R.D. (1997). Genes for Tfb2, Tfb3, and Tfb4 subunits of yeast transcription/repair factor IIH. Homology to human cyclin-dependent kinase activating kinase and IIH subunits. J Biol Chem *272*, 19319-19327.
396. Marinoni, J.C., Rossignol, M., and Egly, J.M. (1997a). Purification of the transcription/repair factor TFIIH and evaluation of its associated activities in vitro. Methods *12*, 235-253.
397. Marinoni, J.C., Roy, R., Vermeulen, W., Miniou, P., Lutz, Y., Weeda, G., Seroz, T., Gomez, D.M., Hoeijmakers, J.H., and Egly, J.M. (1997b). Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH. EMBO J *16*, 1093-1102.
398. Takagi, Y., Komori, H., Chang, W.H., Hudmon, A., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (2003). Revised subunit structure of yeast transcription factor IIH (TFIIH) and reconciliation with human TFIIH. J Biol Chem *278*, 43897-43900.
399. Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J.*, et al.* (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat Genet *36*, 714-719.
400. Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X.J., Eng, J., and Aebersold, R. (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIH. Nat Genet *36*, 707-713.
401. Svejstrup, J.Q., Wang, Z., Feaver, W.J., Wu, X., Bushnell, D.A., Donahue, T.F., Friedberg, E.C., and Kornberg, R.D. (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell *80*, 21-28.
402. Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C., and Egly, J.M. (2008). Nucleotide excision repair driven by the dissociation of CAK from TFIIH. Mol Cell *31*, 9-20.
403. Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., and Cramer, P. (2019). Structural basis of TFIIH activation for nucleotide excision repair. Nat Commun *10*, 2885.
404. Wang, W., Carey, M., and Gralla, J.D. (1992). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science *255*, 450-453.

405. Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. Science *261*, 759-762.
406. Ohkuma, Y., and Roeder, R.G. (1994). Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. Nature *368*, 160-163.
407. Holstege, F.C., van der Vliet, P.C., and Timmers, H.T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. EMBO J *15*, 1666-1677.
408. Lin, Y.C., Choi, W.S., and Gralla, J.D. (2005). TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nat Struct Mol Biol *12*, 603-607.
409. Grunberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nat Struct Mol Biol *19*, 788-796.
410. Tomko, E.J., Fishburn, J., Hahn, S., and Galburt, E.A. (2017). TFIIH generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning. Nat Struct Mol Biol *24*, 1139-1145.
411. Hantsche, M., and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. Curr Opin Struct Biol *47*, 17-22.
412. Li, M., Xia, X., Tian, Y., Jia, Q., Liu, X., Lu, Y., Li, M., Li, X., and Chen, Z. (2019). Mechanism of DNA translocation underlying chromatin remodelling by Snf2. Nature *567*, 409-413.
413. Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795-804.
414. Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791-1792.
415. Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015b). Uncoupling Promoter Opening from Start-Site Scanning. Mol Cell *59*, 133-138.
416. Fishburn, J., Galburt, E., and Hahn, S. (2016). Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II. J Biol Chem *291*, 13040-13047.
417. Qiu, C., Jin, H., Vvedenskaya, I., Llenas, J.A., Zhao, T., Malik, I., Visbisky, A.M., Schwartz, S.L., Cui, P., Cabart, P*., et al.* (2020). Universal promoter scanning by Pol II during transcription initiation in Saccharomyces cerevisiae. Genome Biol *21*, 132.
418. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R.H. (2000). Structural organization of the RNA polymerase-promoter open complex. Cell *101*, 601-611.
419. Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell *147*, 1257-1269.
420. Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.M., Le May, N., and Coin, F. (2017). Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Mol Cell *65*, 504-514 e504.
421. Glyde, R., Ye, F., Jovanovic, M., Kotta-Loizou, I., Buck, M., and Zhang, X. (2018). Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. Mol Cell *70*, 1111-1120 e1113.
422. Han, Y., Yan, C., Fishbain, S., Ivanov, I., and He, Y. (2018). Structural visualization of RNA polymerase III transcription machineries. Cell Discov *4*, 40.
423. Vorlander, M.K., Khatter, H., Wetzel, R., Hagen, W.J.H., and Muller, C.W. (2018). Molecular mechanism of promoter opening by RNA polymerase III. Nature *553*, 295-300.
424. Sadian, Y., Baudin, F., Tafur, L., Murciano, B., Wetzel, R., Weis, F., and Muller, C.W. (2019). Molecular insight into RNA polymerase I promoter recognition and promoter melting. Nat Commun *10*, 5543.
425. Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A.Y., Olinares, P.D.B., Saecker, R.M., Winkelman, J.T., Maloney, M.F., Chait, B.T., Ross, W*., et al.* (2020). Stepwise Promoter Melting by Bacterial RNA Polymerase. Mol Cell *78*, 275-288 e276.
426. Pilsl, M., and Engel, C. (2020). Structural basis of RNA polymerase I pre-initiation complex formation and promoter melting. Nat Commun *11*, 1206.
427. Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol Cell *19*, 101-110.
428. Zhang, Z., and Dietrich, F.S. (2005). Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. Nucleic Acids Res *33*, 2838-2851.
429. Kaplan, C.D., Jin, H., Zhang, I.L., and Belyanin, A. (2012). Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. PLoS Genet *8*, e1002627.

430. Goodrich, J.A., and Tjian, R. (1994). Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. Cell *77*, 145-156.
431. Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell *54*, 601-612.
432. Schilbach, S., Wang, H., Dienemann, C., and Cramer, P. (2023). Yeast PIC-Mediator structure with RNA polymerase II C-terminal domain. Proc Natl Acad Sci U S A *120*, e2220542120.
433. Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev *9*, 1479-1490.
434. Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature *408*, 225-229.
435. Elmendorf, B.J., Shilatifard, A., Yan, Q., Conaway, J.W., and Conaway, R.C. (2001). Transcription factors TFIIF, ELL, and Elongin negatively regulate SII-induced nascent transcript cleavage by non-arrested RNA polymerase II elongation intermediates. J Biol Chem *276*, 23109-23114.
436. Zhang, C., Zobeck, K.L., and Burton, Z.F. (2005). Human RNA polymerase II elongation in slow motion: role of the TFIIF RAP74 alpha1 helix in nucleoside triphosphate-driven translocation. Mol Cell Biol *25*, 3583-3595.
437. Compe, E., Genes, C.M., Braun, C., Coin, F., and Egly, J.M. (2019). TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. Nat Commun *10*, 2084.
438. Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599-608.
439. Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. Proc Natl Acad Sci U S A *95*, 8538-8543.
440. Meyer, K.D., Lin, S.C., Bernecky, C., Gao, Y., and Taatjes, D.J. (2010). p53 activates transcription by directing structural shifts in Mediator. Nat Struct Mol Biol *17*, 753-760.
441. Abdella, R., Talyzina, A., Chen, S., Inouye, C.J., Tjian, R., and He, Y. (2021). Structure of the human Mediator-bound transcription preinitiation complex. Science *372*, 52-56.
442. Chen, X., Yin, X., Li, J., Wu, Z., Qi, Y., Wang, X., Liu, W., and Xu, Y. (2021b). Structures of the human Mediator and Mediator-bound preinitiation complex. Science *372*.
443. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., and Cramer, P. (2021). Structure of the human Mediator-RNA polymerase II pre-initiation complex. Nature *594*, 129-133.
444. Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.
445. Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat Rev Mol Cell Biol *19*, 262-274.
446. Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.
447. Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S.*, et al.* (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc Natl Acad Sci U S A *111*, 7319-7324.
448. Aranda-Orgilles, B., Saldana-Meyer, R., Wang, E., Trompouki, E., Fassl, A., Lau, S., Mullenders, J., Rocha, P.P., Raviram, R., Guillamot, M.*, et al.* (2016). MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. Cell Stem Cell *19*, 784-799.
449. Finn, E.H., and Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. Science *365*.
450. Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell *183*, 28-45.
451. Oudelaar, A.M., and Higgs, D.R. (2021). The relationship between genome structure and function. Nat Rev Genet *22*, 154-168.
452. El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X.*, et al.* (2019). A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. Cell *178*, 1145-1158 e1120.

453. Baek, I., Friedman, L.J., Gelles, J., and Buratowski, S. (2021). Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. Mol Cell *81*, 3576-3588 e3576.
454. Sun, F., Sun, T., Kronenberg, M., Tan, X., Huang, C., and Carey, M.F. (2021). The Pol II preinitiation complex (PIC) influences Mediator binding but not promoter-enhancer looping. Genes Dev *35*, 1175-1189.
455. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.
456. Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nat Struct Mol Biol *21*, 449-455.
457. Jeronimo, C., Langelier, M.F., Bataille, A.R., Pascal, J.M., Pugh, B.F., and Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Mol Cell *64*, 455-466.
458. Xiao, J.Y., Hafner, A., and Boettiger, A.N. (2021). How subtle changes in 3D structure can create large changes in transcription. Elife *10*.
459. Palacio, M., and Taatjes, D.J. (2022). Merging Established Mechanisms with New Insights: Condensates, Hubs, and the Regulation of RNA Polymerase II Transcription. J Mol Biol *434*, 167216.
460. Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. Science *295*, 1058-1062.
461. Andrau, J.C., van de Pasch, L., Lijnzaad, P., Bijma, T., Koerkamp, M.G., van de Peppel, J., Werner, M., and Holstege, F.C. (2006). Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. Mol Cell *22*, 179-192.
462. Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C.O., Koeck, P.J., Holmberg, S., Hebert, H., and Gustafsson, C.M. (2006). The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. Proc Natl Acad Sci U S A *103*, 15788-15793.
463. Tsai, K.L., Sato, S., Tomomori-Sato, C., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2013). A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. Nat Struct Mol Biol *20*, 611-619.
464. Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc Natl Acad Sci U S A *100*, 6964-6968.
465. Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. Proc Natl Acad Sci U S A *100*, 6969-6973.
466. Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem *280*, 7131-7134.
467. Meyer, P.A., Ye, P., Zhang, M., Suh, M.H., and Fu, J. (2006). Phasing RNA polymerase II using intrinsically bound Zn atoms: an updated structural model. Structure *14*, 973-982.
468. Meyer, P.A., Ye, P., Suh, M.H., Zhang, M., and Fu, J. (2009). Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data. J Biol Chem *284*, 12933-12939.
469. Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481-489.
470. Cheung, A.C., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. EMBO J *30*, 4755-4763.
471. Liu, X., Bushnell, D.A., Silva, D.A., Huang, X., and Kornberg, R.D. (2011). Initiation complex structure and promoter proofreading. Science *333*, 633-637.
472. Bernecky, C., Grob, P., Ebmeier, C.C., Nogales, E., and Taatjes, D.J. (2011). Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. PLoS Biol *9*, e1000603.
473. Cai, G., Chaban, Y.L., Imasaki, T., Kovacs, J.A., Calero, G., Penczek, P.A., Takagi, Y., and Asturias, F.J. (2012). Interaction of the mediator head module with RNA polymerase II. Structure *20*, 899-910.
474. Robinson, P.J., Bushnell, D.A., Trnka, M.J., Burlingame, A.L., and Kornberg, R.D. (2012). 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.
475. Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1422 e1416.
476. Tsai, K.L., Yu, X.D., Gopalan, S., Chao, T.C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W.*, et al.* (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature *544*, 196-+.

477. Zhang, H., Chen, D.H., Mattoo, R.U.H., Bushnell, D.A., Wang, Y., Yuan, C., Wang, L., Wang, C., Davis, R.E., Nie, Y.*, et al.* (2021b). Mediator structure and conformation change. Mol Cell *81*, 1781-1788 e1784.
478. Gorbea Colon, J.J., Palao, L., 3rd, Chen, S.F., Kim, H.J., Snyder, L., Chang, Y.W., Tsai, K.L., and Murakami, K. (2023). Structural basis of a transcription pre-initiation complex on a divergent promoter. Mol Cell *83*, 574-588 e511.
479. Wu, P.Y., Ruhlmann, C., Winston, F., and Schultz, P. (2004). Molecular architecture of the S. cerevisiae SAGA complex. Mol Cell *15*, 199-208.
480. Setiaputra, D., Ross, J.D., Lu, S., Cheng, D.T., Dong, M.Q., and Yip, C.K. (2015). Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. J Biol Chem *290*, 10057-10070.
481. Papai, G., Frechard, A., Kolesnikova, O., Crucifix, C., Schultz, P., and Ben-Shem, A. (2020). Structure of SAGA and mechanism of TBP deposition on gene promoters. Nature *577*, 711-716.
482. Wang, H., Dienemann, C., Stutzer, A., Urlaub, H., Cheung, A.C.M., and Cramer, P. (2020a). Structure of the transcription coactivator SAGA. Nature *577*, 717-720.
483. Herbst, D.A., Esbin, M.N., Louder, R.K., Dugast-Darzacq, C., Dailey, G.M., Fang, Q., Darzacq, X., Tjian, R., and Nogales, E. (2021). Structure of the human SAGA coactivator complex. Nat Struct Mol Biol *28*, 989-996.
484. Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol *375*, 23-44.
485. Hu, X., Malik, S., Negroiu, C.C., Hubbard, K., Velalar, C.N., Hampton, B., Grosu, D., Catalano, J., Roeder, R.G., and Gnatt, A. (2006). A Mediator-responsive form of metazoan RNA polymerase II. Proc Natl Acad Sci U S A *103*, 9506-9511.
486. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol *152*, 36-51.
487. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods *14*, 290-296.
488. Scheres, S.H.W. (2020). Amyloid structure determination in RELION-3.1. Acta Crystallogr D Struct Biol *76*, 94-101.
489. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*.
490. Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat Methods *16*, 1146-1152.
491. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol *74*, 531-544.
492. Schrödinger, L., and Delano, W. (2020). PyMOL v2.5.0.
493. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.
494. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci *27*, 14-25.
495. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr D Struct Biol *74*, 519-530.
496. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.
497. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.
498. Wang, H., Farnung, L., Dienemann, C., and Cramer, P. (2020b). Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. Nat Struct Mol Biol *27*, 8-13.
499. Wagner, F.R., Dienemann, C., Wang, H., Stutzer, A., Tegunov, D., Urlaub, H., and Cramer, P. (2020). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. Nature *579*, 448-451.
500. Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P.*, et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol *25*, 833-+.

501. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551-554.
502. Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E.*, et al.* (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods *5*, 53-55.
503. Greber, B.J., Perez-Bertoldi, J.M., Lim, K., Iavarone, A.T., Toso, D.B., and Nogales, E. (2020). The cryoelectron microscopy structure of the human CDK-activating kinase. Proc Natl Acad Sci U S A *117*, 22849-22857.
504. Wang, H., Xiong, L., and Cramer, P. (2021a). Structures and implications of TBP-nucleosome complexes. Proc Natl Acad Sci U S A *118*.
505. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010b). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.
506. Greber, B.J., Nguyen, T.H.D., Fang, J., Afonine, P.V., Adams, P.D., and Nogales, E. (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature *549*, 414-417.
507. Yan, C.L., Dodd, T., He, Y., Tainer, J.A., Tsutakawa, S.E., and Ivanov, I. (2019a). Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. Nature Structural & Molecular Biology *26*, 397-+.
508. Dangkulwanich, M., Ishibashi, T., Liu, S., Kireeva, M.L., Lubkowska, L., Kashlev, M., and Bustamante, C.J. (2013). Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. Elife *2*, e00971.
509. Chen, Z., Gabizon, R., Brown, A.I., Lee, A., Song, A., Diaz-Celis, C., Kaplan, C.D., Koslover, E.F., Yao, T., and Bustamante, C. (2019). High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. Elife *8*.
510. Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-373.
511. Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950-953.
512. Ramachandran, S., Ahmad, K., and Henikoff, S. (2017). Transcription and Remodeling Produce Asymmetrically Unwrapped Nucleosomal Intermediates. Mol Cell *68*, 1038-1053 e1034.
513. Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem Sci *39*, 577-586.
514. Wasylyk, B., and Chambon, P. (1979). Transcription by eukaryotic RNA polymerases A and B of chromatin assembled in vitro. Eur J Biochem *98*, 317-327.
515. Wasylyk, B., Thevenin, G., Oudet, P., and Chambon, P. (1979). Transcription of in vitro assembled chromatin by Escherichia coli RNA polymerase. J Mol Biol *128*, 411-440.
516. Knezetic, J.A., Jacob, G.A., and Luse, D.S. (1988). Assembly of RNA polymerase II preinitiation complexes before assembly of nucleosomes allows efficient initiation of transcription on nucleosomal templates. Mol Cell Biol *8*, 3114-3121.
517. Shaw, P.A., Sahasrabuddhe, C.G., Hodo, H.G., 3rd, and Saunders, G.F. (1978). Transcription of nucleosomes from human chromatin. Nucleic Acids Res *5*, 2999-3012.
518. Izban, M.G., and Luse, D.S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J Biol Chem *267*, 13647-13655.
519. Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.
520. Fisher, M.J., and Luse, D.S. (2023). Promoter-proximal nucleosomes attenuate RNA polymerase II transcription through TFIID. J Biol Chem *299*, 104928.
521. Oberbeckmann, E., Niebauer, V., Watanabe, S., Farnung, L., Moldt, M., Schmid, A., Cramer, P., Peterson, C.L., Eustermann, S., Hopfner, K.P.*, et al.* (2021b). Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. Nat Commun *12*, 3232.
522. Sundaramoorthy, R., Hughes, A.L., El-Mkami, H., Norman, D.G., Ferreira, H., and Owen-Hughes, T. (2018). Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome. Elife *7*.
523. Topolska-Wos, A.M., Sugitani, N., Cordoba, J.J., Le Meur, K.V., Le Meur, R.A., Kim, H.S., Yeo, J.E., Rosenberg, D., Hammel, M., Scharer, O.D.*, et al.* (2020). A key interaction with RPA orients XPA in NER complexes. Nucleic Acids Res *48*, 2173-2188.
524. van Eeuwen, T., Shim, Y., Kim, H.J., Zhao, T., Basu, S., Garcia, B.A., Kaplan, C.D., Min, J.H., and Murakami, K. (2021). Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision repair. Nat Commun *12*, 3338.

525. Kim, J., Li, C.L., Chen, X., Cui, Y., Golebiowski, F.M., Wang, H., Hanaoka, F., Sugasawa, K., and Yang, W. (2023). Lesion recognition by XPC, TFIIH and XPA in DNA excision repair. Nature *617*, 170-175.
526. Yu, J., Yan, C., Dodd, T., Tsai, C.L., Tainer, J.A., Tsutakawa, S.E., and Ivanov, I. (2023). Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. Nat Commun *14*, 2758.
527. Yan, C., Dodd, T., He, Y., Tainer, J.A., Tsutakawa, S.E., and Ivanov, I. (2019b). Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. Nat Struct Mol Biol *26*, 397-406.
528. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.
529. Nouraini, S., Archambault, J., and Friesen, J.D. (1996). Rpo26p, a subunit common to yeast RNA polymerases, is essential for the assembly of RNA polymerases I and II and for the stability of the largest subunits of these enzymes. Mol Cell Biol *16*, 5985-5996.
530. Ishiguro, A., Nogi, Y., Hisatake, K., Muramatsu, M., and Ishihama, A. (2000). The Rpb6 subunit of fission yeast RNA polymerase II is a contact target of the transcription elongation factor TFIIS. Mol Cell Biol *20*, 1263-1270.
531. Okuda, M., Suwa, T., Suzuki, H., Yamaguchi, Y., and Nishimura, Y. (2022). Three human RNA polymerases interact with TFIIH via a common RPB6 subunit. Nucleic Acids Res *50*, 1-16.
532. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. Nature *502*, 650-655.
533. Fernandez-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M., Ruiz, F.M., Gruene, T., Legrand, P., Steuerwald, U., and Muller, C.W. (2013). Crystal structure of the 14-subunit RNA polymerase I. Nature *502*, 644-649.
534. Wang, Q., Li, S., Wan, F., Xu, Y., Wu, Z., Cao, M., Lan, P., Lei, M., and Wu, J. (2021b). Structural insights into transcriptional regulation of human RNA polymerase III. Nat Struct Mol Biol *28*, 220-227.
535. Rimel, J.K., and Taatjes, D.J. (2018). The essential and multifunctional TFIIH complex. Protein Sci *27*, 1018-1037.
536. Day, D.S., Zhang, B., Stevens, S.M., Ferrari, F., Larschan, E.N., Park, P.J., and Pu, W.T. (2016). Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. Genome Biol *17*, 120.
537. Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? Dev Biol *339*, 258-266.
538. Dreos, R., Ambrosini, G., and Bucher, P. (2016). Influence of Rotational Nucleosome Positioning on Transcription Start Site Selection in Animal Promoters. PLoS Comput Biol *12*, e1005144.
539. Lam, K.C., Chung, H.R., Semplicio, G., Iyer, S.S., Gaub, A., Bhardwaj, V., Holz, H., Georgiev, P., and Akhtar, A. (2019). The NSL complex-mediated nucleosome landscape is required to maintain transcription fidelity and suppression of transcription noise. Genes Dev *33*, 452-465.
540. Tora, L., and Vincent, S.D. (2021). What defines the maternal transcriptome? Biochem Soc Trans *49*, 2051-2062.
541. Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P.*, et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.
542. de Graaf, P., Mousson, F., Geverts, B., Scheer, E., Tora, L., Houtsmuller, A.B., and Timmers, H.T. (2010). Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. J Cell Sci *123*, 2663-2671.
543. Nogales, E., Louder, R.K., and He, Y. (2016). Cryo-EM in the study of challenging systems: the human transcription pre-initiation complex. Curr Opin Struct Biol *40*, 120-127.
544. Zhang, Z., English, B.P., Grimm, J.B., Kazane, S.A., Hu, W., Tsai, A., Inouye, C., You, C., Piehler, J., Schultz, P.G.*, et al.* (2016). Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. Genes Dev *30*, 2106-2118.
545. Nguyen, V.Q., Ranjan, A., Liu, S., Tang, X., Ling, Y.H., Wisniewski, J., Mizuguchi, G., Li, K.Y., Jou, V., Zheng, Q.*, et al.* (2021). Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol Cell *81*, 3560-3575 e3566.

546. Petesch, S.J., and Lis, J.T. (2012). Overcoming the nucleosome barrier during transcript elongation. Trends Genet *28*, 285-294.
547. Farnung, L., Vos, S.M., and Cramer, P. (2018). Structure of transcribing RNA polymerase II-nucleosome complex. Nat Commun *9*, 5432.
548. Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.I., and Kurumizaka, H. (2018). Structural basis of the nucleosome transition during RNA polymerase II passage. Science *362*, 595-598.
549. Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H., and Sekine, S.I. (2019). Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. Science *363*, 744-747.
550. Ehara, H., Kujirai, T., Shirouzu, M., Kurumizaka, H., and Sekine, S.I. (2022). Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. Science *377*, eabp9466.
551. Filipovski, M., Soffers, J.H.M., Vos, S.M., and Farnung, L. (2022). Structural basis of nucleosome retention during transcription elongation. Science *376*, 1313-1316.
552. Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., and Carey, M. (2002). TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. Genes Dev *16*, 1852-1863.
553. Malik, S., and Roeder, R.G. (2023). Regulation of the RNA polymerase II pre-initiation complex by its associated coactivators. Nat Rev Genet.
554. Liu, Z., Yao, X., Yan, G., Xu, Y., Yan, J., Zou, W., and Wang, G. (2016). Mediator MED23 cooperates with RUNX2 to drive osteoblast differentiation and bone development. Nat Commun *7*, 11149.
555. Henley, M.J., Linhares, B.M., Morgan, B.S., Cierpicki, T., Fierke, C.A., and Mapp, A.K. (2020). Unexpected specificity within dynamic transcriptional protein-protein complexes. Proc Natl Acad Sci U S A *117*, 27346-27353.
556. Tuttle, L.M., Pacheco, D., Warfield, L., Wilburn, D.B., Hahn, S., and Klevit, R.E. (2021). Mediator subunit Med15 dictates the conserved "fuzzy" binding mechanism of yeast transcription activators Gal4 and Gcn4. Nat Commun *12*, 2220.
557. Fondell, J.D., Ge, H., and Roeder, R.G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. Proc Natl Acad Sci U S A *93*, 8329-8333.
558. Hittelman, A.B., Burakov, D., Iniguez-Lluhi, J.A., Freedman, L.P., and Garabedian, M.J. (1999). Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins. EMBO J *18*, 5380-5388.
559. Wang, Q., Sharma, D., Ren, Y., and Fondell, J.D. (2002). A coregulatory role for the TRAP-mediator complex in androgen receptor-mediated gene expression. J Biol Chem *277*, 42852-42858.
560. Chu, C.S., Hellmuth, J.C., Singh, R., Ying, H.Y., Skrabanek, L., Teater, M.R., Doane, A.S., Elemento, O., Melnick, A.M., and Roeder, R.G. (2020). Unique Immune Cell Coactivators Specify Locus Control Region Function and Cell Stage. Mol Cell *80*, 845-861 e810.
561. Belorusova, A.Y., Bourguet, M., Hessmann, S., Chalhoub, S., Kieffer, B., Cianferani, S., and Rochel, N. (2020). Molecular determinants of MED1 interaction with the DNA bound VDR-RXR heterodimer. Nucleic Acids Res *48*, 11199-11213.
562. Grunberg, S., Henikoff, S., Hahn, S., and Zentner, G.E. (2016). Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. EMBO J *35*, 2435-2446.
563. Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. Mol Cell *68*, 118-129 e115.
564. Zhang, J., Kalkum, M., Yamamura, S., Chait, B.T., and Roeder, R.G. (2004). E protein silencing by the leukemogenic AML1-ETO fusion protein. Science *305*, 1286-1289.
565. Wright, K.J., and Tjian, R. (2009). Wnt signaling targets ETO coactivation domain of TAF4/TFIID in vivo. Proc Natl Acad Sci U S A *106*, 55-60.
566. Chen, W.Y., Zhang, J., Geng, H., Du, Z., Nakadai, T., and Roeder, R.G. (2013). A TAF4 coactivator function for E proteins that involves enhanced TFIID binding. Genes Dev *27*, 1596-1609.
567. Nayak, S., and Taatjes, D.J. (2022). SnapShot: Mediator complex structure. Cell *185*, 3458-3458 e3451.
568. Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic disorder in transcription factors. Biochemistry *45*, 6873-6888.

569. Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.
570. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.
571. Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.
572. Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.
573. van Ingen, H., van Schaik, F.M., Wienk, H., Ballering, J., Rehmann, H., Dechesne, A.C., Kruijzer, J.A., Liskamp, R.M., Timmers, H.T., and Boelens, R. (2008). Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure *16*, 1245-1256.
574. van der Vliet, P.C., and Verrijzer, C.P. (1993). Bending of DNA by transcription factors. Bioessays *15*, 25-32.

Appendix

Supplemental figures

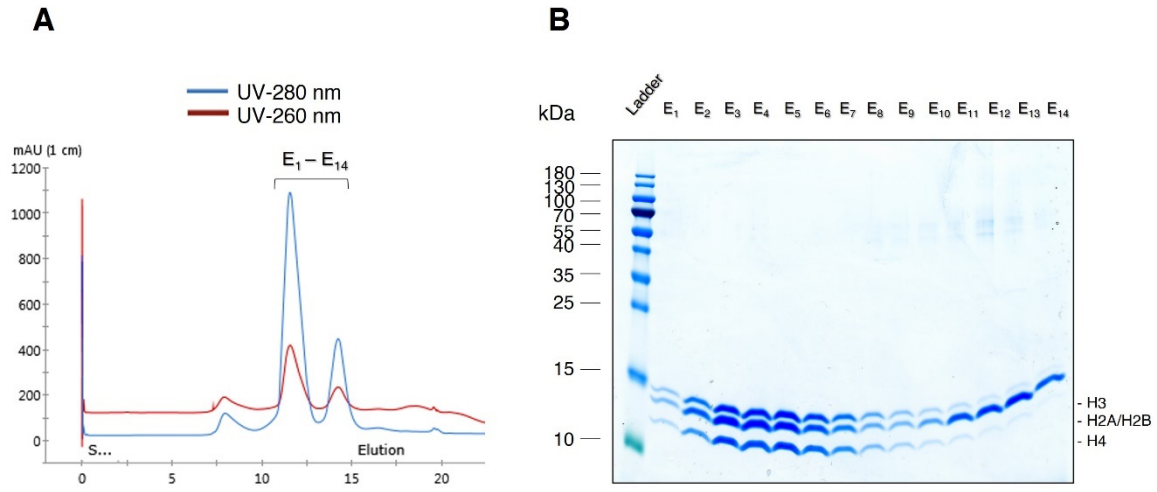


Figure S1. Histone octamer assembly.

(A) Chromatographic separation of *X. laevis* histone octamer from subspecies after gel filtration. Marked elution fractions (E_1-E_{14}) were analytically checked with denaturing electrophoresis **(B)**.

(**B)** 16% Tris-glycine-SDS gel. Molecular weight markers are shown before lane 1, and histones marked with arrows at their corresponding molecular weight. Elution 2 – elution 6 were pooled together to have a stochiometric histone octamer free of H3-H4 tetramers and H2A-H2B dimers. E, elution fraction.

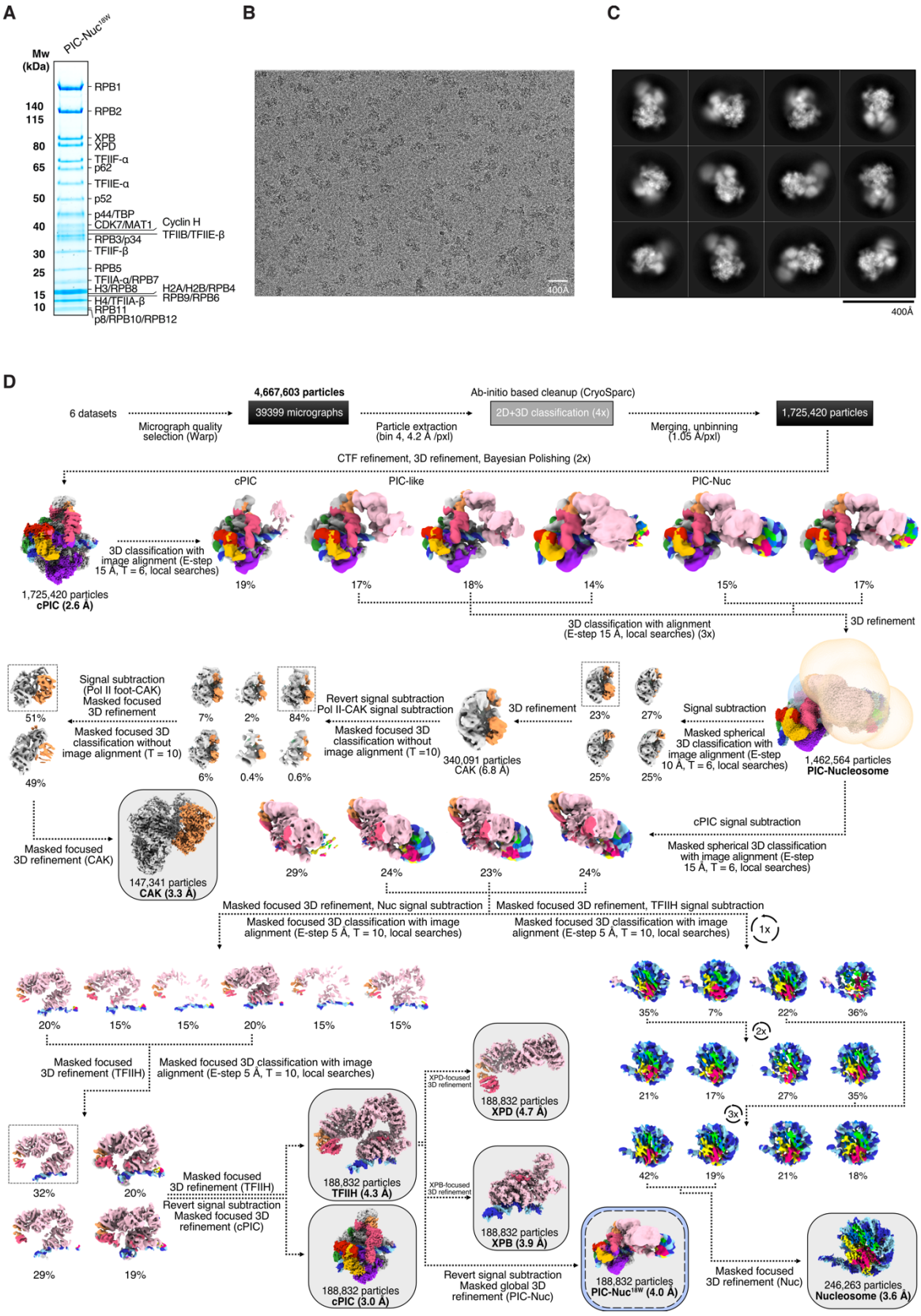

A
PIC-Nuc18W
Mw (kDa)
140
115
80
65
50
40
30
25
15
10
RPB1
RPB2
XPB
XPD
TFIIF-α
p62
TFIIE-α
p52
p44/TBP
CDK7/MAT1
RPB3/p34
TFIIF-β
RPB5
TFIIA-α/RPB7
H3/RPB8
H4/TFIIA-β
RPB11
p8/RPB10/RPB12
Cyclin H
TFIIB/TFIIE-β
H2A/H2B/RPB4
RPB9/RPB6
B
400Å
C
400Å
D
6 datasets
Micrograph quality selection (Warp)
4,667,603 particles
39399 micrographs
Particle extraction (bin 4, 4.2 Å /pxl)
Ab-initio based cleanup (CryoSparc)
2D+3D classification (4x)
Merging, unbinning (1.05 Å/pxl)
1,725,420 particles
CTF refinement, 3D refinement, Bayesian Polishing (2x)
cPIC
PIC-like
PIC-Nuc
3D classification with image alignment (E-step 15 Å, T = 6, local searches)
1,725,420 particles
cPIC (2.6 Å)
19%
17%
18%
14%
15%
17%
3D classification with alignment (E-step 15 Å, local searches) (3x)
3D refinement
Signal subtraction (Pol II foot-CAK) Masked focused 3D refinement
Masked focused 3D classification without image alignment (T = 10)
51%
49%
7%
2%
84%
6%
0.4%
0.6%
Revert signal subtraction Pol II-CAK signal subtraction
Masked focused 3D classification without image alignment (T =10)
3D refinement
340,091 particles
CAK (6.8 Å)
23%
25%
27%
25%
Signal subtraction
Masked spherical 3D classification with image alignment (E-step 10 Å, T = 6, local searches)
1,462,564 particles
PIC-Nucleosome
Masked focused 3D refinement (CAK)
147,341 particles
CAK (3.3 Å)
cPIC signal subtraction
Masked spherical 3D classification with image alignment (E-step 15 Å, T = 6, local searches)
29%
24%
23%
24%
Masked focused 3D refinement, Nuc signal subtraction
Masked focused 3D refinement, TFIIH signal subtraction
Masked focused 3D classification with image alignment (E-step 5 Å, T = 10, local searches)
Masked focused 3D classification with image alignment (E-step 5 Å, T = 10, local searches)
1x
20%
15%
15%
20%
15%
15%
Masked focused 3D refinement (TFIIH)
Masked focused 3D classification with image alignment (E-step 5 Å, T = 10, local searches)
35%
7%
22%
36%
2x
21%
17%
27%
35%
3x
42%
19%
21%
18%
32%
20%
29%
19%
Masked focused 3D refinement (TFIIH)
Revert signal subtraction Masked focused 3D refinement (cPIC)
188,832 particles
TFIIH (4.3 Å)
188,832 particles
cPIC (3.0 Å)
XPD-focused 3D refinement
XPB-focused 3D refinement
188,832 particles
XPD (4.7 Å)
188,832 particles
XPB (3.9 Å)
Revert signal subtraction Masked global 3D refinement (PIC-Nuc)
188,832 particles
PIC-Nuc18W (4.0 Å)
Masked focused 3D refinement (Nuc)
246,263 particles
Nucleosome (3.6 Å)

Figure S2. Sample preparation and cryo-EM processing analysis of PIC-Nucleosome^{18W}, Related to Figure 7.

(A) Sucrose gradient fraction used for cryo-EM sample preparation (analysed by SDS-PAGE). Molecular weight is described on the left side of the gel and assembly components on the right side.

(B) Exemplary cryo-EM micrograph. A scale bar is provided at the bottom right of the figure.

(C) Representative 2D class averages depicting mammalian PIC-nucleosome^{18W}. Scale bar is placed at the bottom right of the figure.

(D) Processing classification strategy employed to sort out particles and for structure determination of PIC-Nuc^{18W}. Each step of the tree is described on the diagram and the reported resolutions are shown under the respective cryo-EM maps. Final maps used for model building are located within grey boxes, whereas consensus refinements are enclosed in grey-blue boxes.

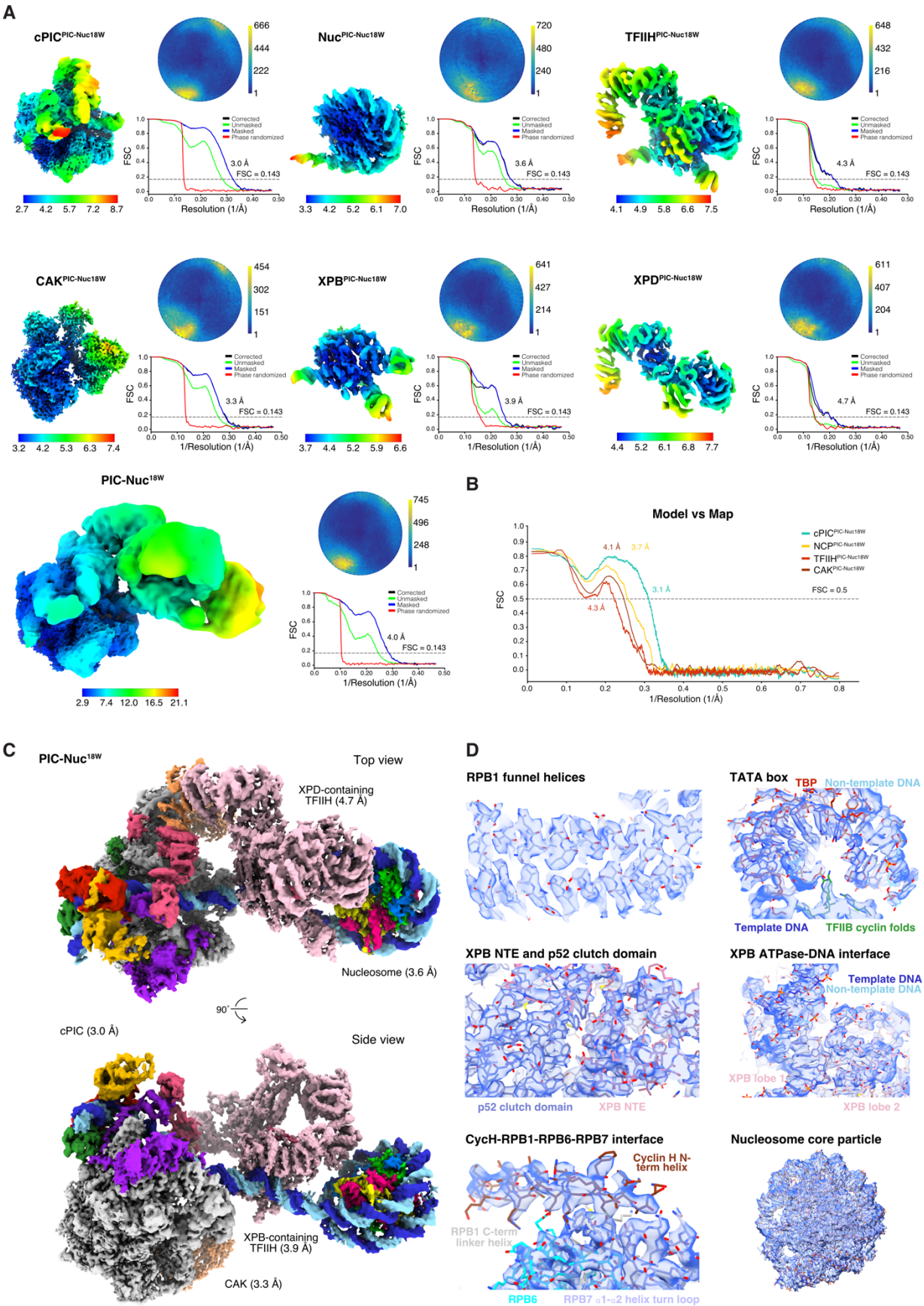

A
cPIC PIC-Nuc18W
Nuc PIC-Nuc18W
TFIIH PIC-Nuc18W
CAK PIC-Nuc18W
XPB PIC-Nuc18W
XPD PIC-Nuc18W
PIC-Nuc18W
Corrected
Unmasked
Masked
Phase randomized
FSC
Resolution (1/Å)
1/Resolution (1/Å)
FSC = 0.143
3.0 Å
3.6 Å
4.3 Å
3.3 Å
3.9 Å
4.7 Å
4.0 Å
B
Model vs Map
cPIC PIC-Nuc18W
NCP PIC-Nuc18W
TFIIH PIC-Nuc18W
CAK PIC-Nuc18W
FSC = 0.5
4.1 Å
3.7 Å
3.1 Å
4.3 Å
1/Resolution (1/Å)
C
PIC-Nuc18W
Top view
XPD-containing TFIIH (4.7 Å)
Nucleosome (3.6 Å)
90°
cPIC (3.0 Å)
Side view
XPB-containing TFIIH (3.9 Å)
CAK (3.3 Å)
D
RPB1 funnel helices
TATA box
TBP
Non-template DNA
Template DNA
TFIIB cyclin folds
XPB NTE and p52 clutch domain
p52 clutch domain
XPB NTE
XPB ATPase-DNA interface
Template DNA
Non-template DNA
XPB lobe 1
XPB lobe 2
CycH-RPB1-RPB6-RPB7 interface
Cyclin H N-term helix
RPB1 C-term linker helix
RPB6
RPB7 α1-α2 helix turn loop
Nucleosome core particle

Figure S3. Map quality assessment of the PIC-nucleosome^{18W} structure, Related to Figure 7.

(A) Focused refinements and overall maps filtered and coloured by estimated local resolution, shown as top views. The superscript typography of each map is used to differentiate it from the ones shown in **Figure S6**. Their angular distributions and resolution (following the FSC 0.143 cut-off criterion) are provided on the right side of each map.

(B) Model-map correlation between the individual focused refined maps, used for model building, and the respective built atomic models. Resolution following the FSC 0.5 cut-off criterion is marked with a dashed line.

(C) Composite maps showing individual local refinements in distinct views for PIC-Nuc^{18W}. The different protein and DNA components are coloured according to the colours used in their structural models, depicted in **Figure 7B**.

(D) Cryo-EM density maps (shown in blue surface) fitted to the corresponding region of the modelled structures.

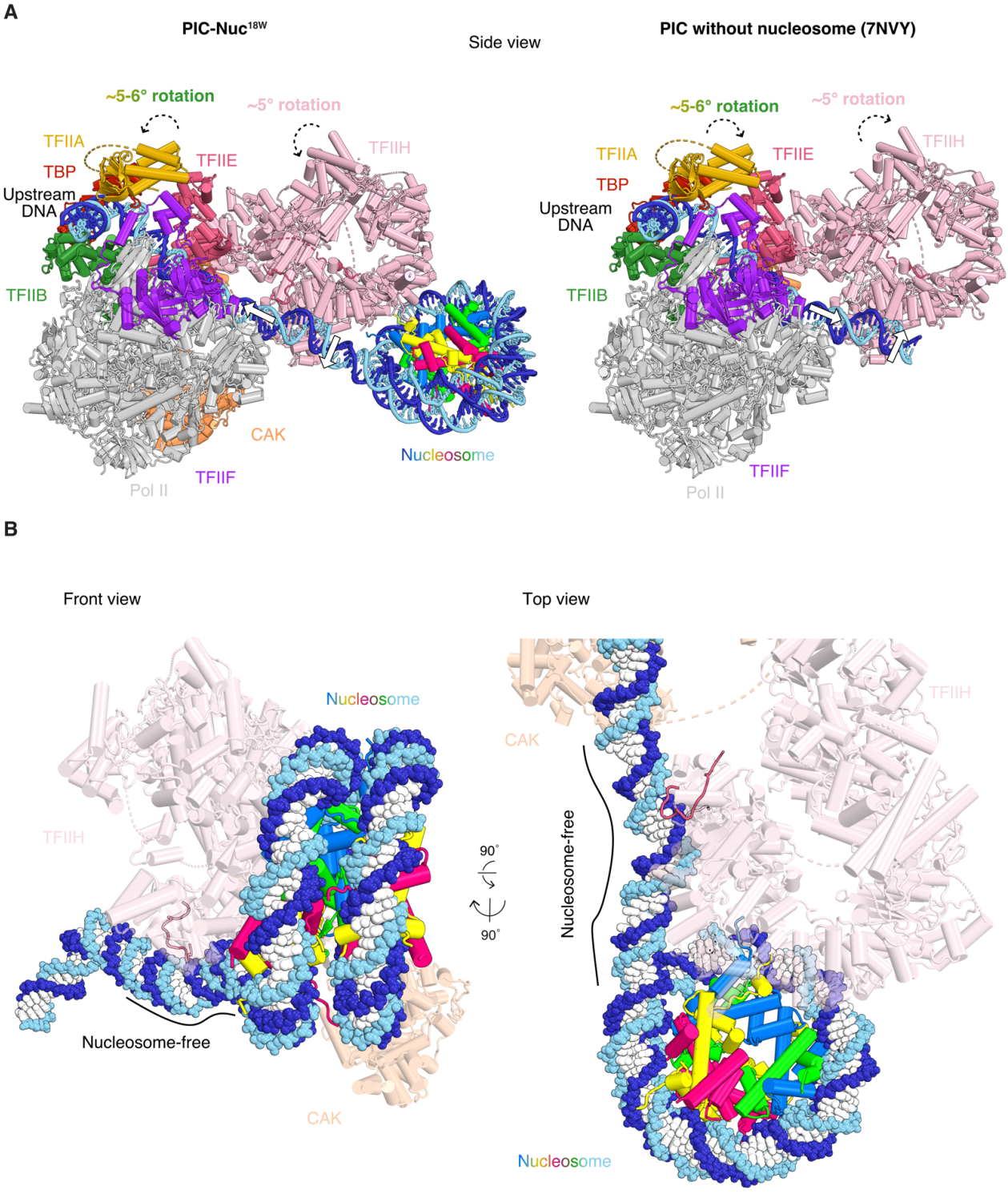

A
PIC-Nuc18W
Side view
PIC without nucleosome (7NVY)
~5-6° rotation
~5° rotation
TFIIA
TBP
TFIIE
TFIIH
Upstream DNA
TFIIB
CAK
Nucleosome
TFIIF
Pol II
B
Front view
Top view
Nucleosome
TFIIH
CAK
Nucleosome-free
90°
90°

Figure S4. Downstream DNA is free for TFIIH binding in PIC-Nuc^{18W} and mammalian PIC without nucleosome, Related to Figure 8.

(A) Comparison of PIC-Nuc^{18W} (left panel) with PIC without nucleosome (right panel) (PDB ID 7NVY (Aibara et al., 2021)) in side view. PIC is found in its canonical conformation, and only minor movements in the upstream complex were observed. The downstream DNA region is slightly bent when compared to the structure of the nucleosome-free PIC (white arrow). Arrowheads indicate the local motion of different PIC subunits.

(B) Front view and top close-up view of TFIIH-Nuc^{18W}. TFIIH binds downstream DNA when the nucleosome is located at a distal position to the promoter region. Nucleosomes are shown in sphere representation and TFIIH in transparent cartoon representation. Solid lines denote nucleosome-free DNA.

Superimposition of models was performed by aligning on Pol II. Colour of the different components is denoted adjacent to them.

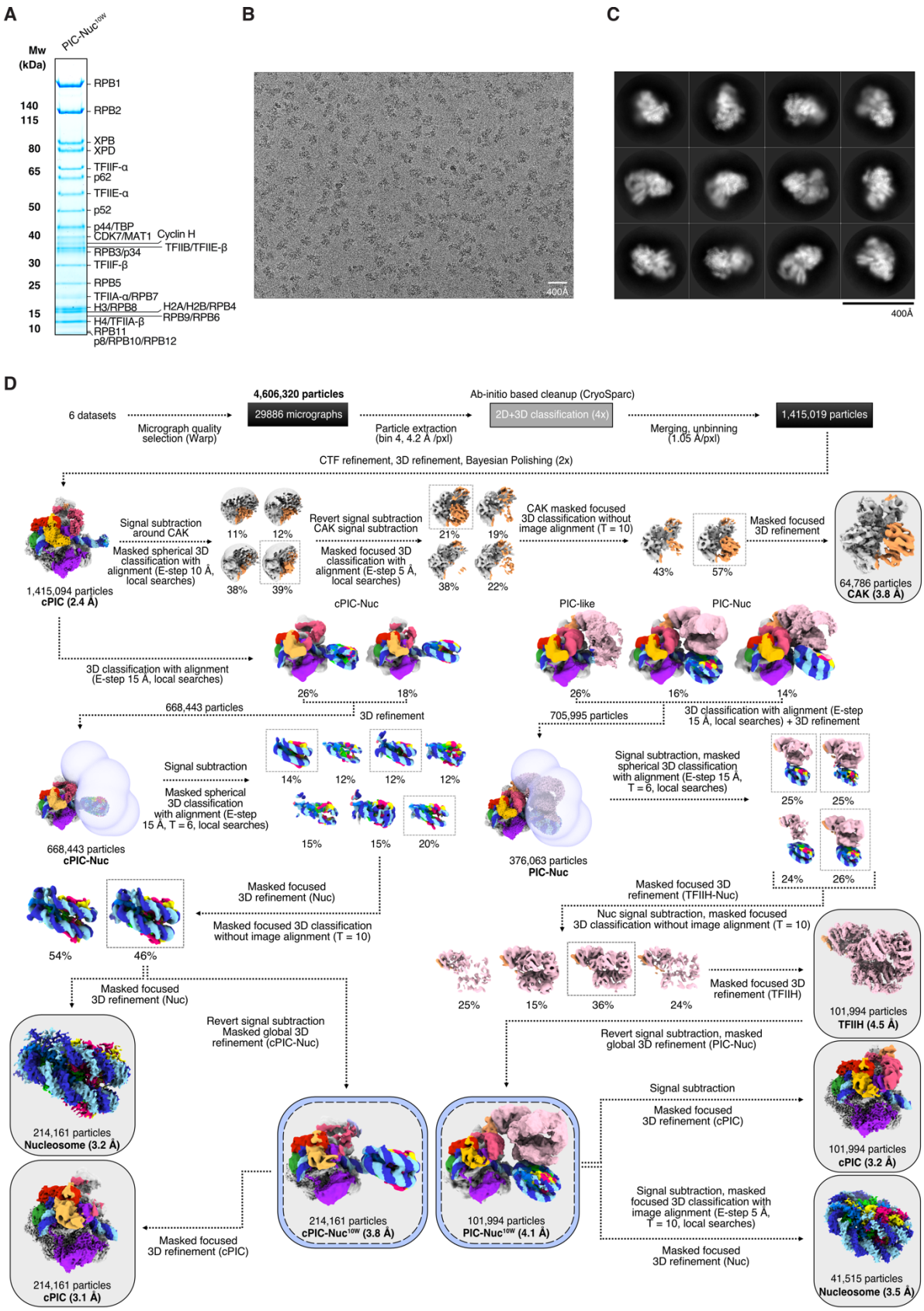

A
PIC-Nuc10W
Mw (kDa)
140
115
80
65
50
40
30
25
15
10
RPB1
RPB2
XPB
XPD
TFIIF-α
p62
TFIIE-α
p52
p44/TBP
CDK7/MAT1
Cyclin H
TFIIB/TFIIE-β
RPB3/p34
TFIIF-β
RPB5
TFIIA-α/RPB7
H3/RPB8
H2A/H2B/RPB4
RPB9/RPB6
H4/TFIIA-β
RPB11
p8/RPB10/RPB12
B
400Å
C
400Å
D
4,606,320 particles
6 datasets
Micrograph quality selection (Warp)
29886 micrographs
Particle extraction (bin 4, 4.2 Å /pxl)
Ab-initio based cleanup (CryoSparc)
2D+3D classification (4x)
Merging, unbinning (1.05 Å/pxl)
1,415,019 particles
CTF refinement, 3D refinement, Bayesian Polishing (2x)
Signal subtraction around CAK
Masked spherical 3D classification with alignment (E-step 10 Å, local searches)
11%
12%
38%
39%
Revert signal subtraction CAK signal subtraction
Masked focused 3D classification with alignment (E-step 5 Å, local searches)
21%
19%
38%
22%
CAK masked focused 3D classification without image alignment (T = 10)
43%
57%
Masked focused 3D refinement
64,786 particles
CAK (3.8 Å)
1,415,094 particles
cPIC (2.4 Å)
3D classification with alignment (E-step 15 Å, local searches)
cPIC-Nuc
26%
18%
668,443 particles
3D refinement
PIC-like
PIC-Nuc
26%
16%
14%
705,995 particles
3D classification with alignment (E-step 15 Å, local searches) + 3D refinement
668,443 particles
cPIC-Nuc
Signal subtraction
Masked spherical 3D classification with alignment (E-step 15 Å, T = 6, local searches)
14%
12%
12%
12%
15%
15%
20%
376,063 particles
PIC-Nuc
Signal subtraction, masked spherical 3D classification with alignment (E-step 15 Å, T = 6, local searches)
25%
25%
24%
26%
Masked focused 3D refinement (Nuc)
Masked focused 3D classification without image alignment (T = 10)
54%
46%
Masked focused 3D refinement (TFIIH-Nuc)
Nuc signal subtraction, masked focused 3D classification without image alignment (T = 10)
25%
15%
36%
24%
Masked focused 3D refinement (TFIIH)
101,994 particles
TFIIH (4.5 Å)
Masked focused 3D refinement (Nuc)
Revert signal subtraction Masked global 3D refinement (cPIC-Nuc)
Revert signal subtraction, masked global 3D refinement (PIC-Nuc)
214,161 particles
Nucleosome (3.2 Å)
214,161 particles
cPIC-Nuc10W (3.8 Å)
101,994 particles
PIC-Nuc10W (4.1 Å)
Signal subtraction
Masked focused 3D refinement (cPIC)
101,994 particles
cPIC (3.2 Å)
Signal subtraction, masked focused 3D classification with image alignment (E-step 5 Å, T = 10, local searches)
Masked focused 3D refinement (Nuc)
41,515 particles
Nucleosome (3.5 Å)
Masked focused 3D refinement (cPIC)
214,161 particles
cPIC (3.1 Å)

Figure S5. Sample preparation and cryo-EM processing analysis of PIC-nucleosome^{10W}, Related to Figure 7.

(A) Sucrose gradient fraction used for cryo-EM sample preparation (analysed by SDS-PAGE). Molecular weight is described on the left side of the gel and assembly components on the right side.

(B) Exemplary cryo-EM micrograph. A scale bar is provided at the bottom right of the figure.

(C) Representative 2D class averages depicting mammalian PIC-nucleosome^{10W}. Scale bar is placed at the bottom right of the figure.

(D) Processing classification strategy employed to sort out particles and for structure determination of PIC-Nuc^{10W}. Each step of the tree is described on the diagram and the reported resolutions are shown under the respective cryo-EM maps. Final maps used for model building are located within grey boxes, whereas consensus refinements are enclosed in grey-blue boxes.

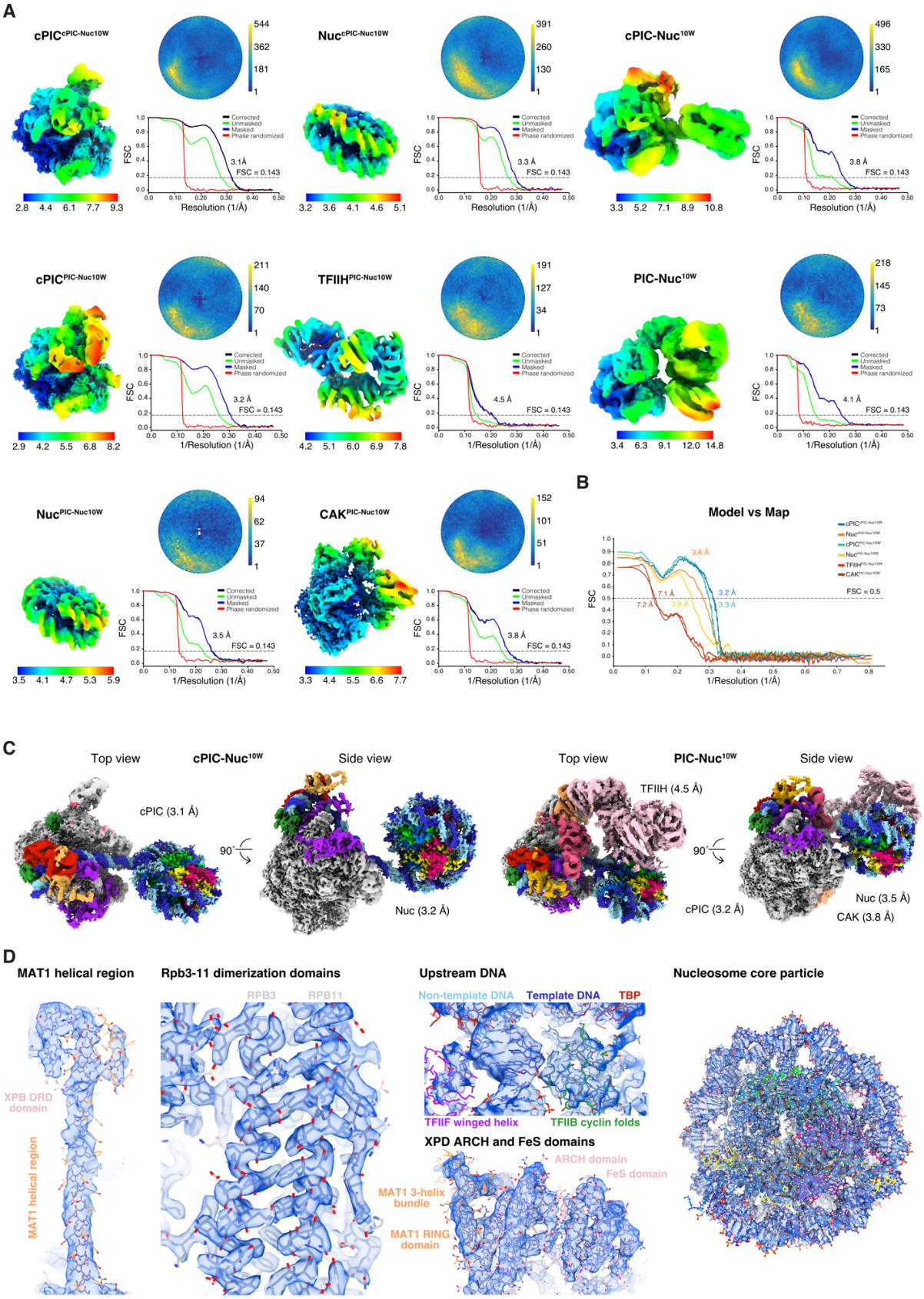

A
cPIC cPIC-Nuc10W
Nuc cPIC-Nuc10W
cPIC-Nuc 10W
cPIC PIC-Nuc10W
TFIIH PIC-Nuc10W
PIC-Nuc 10W
Nuc PIC-Nuc10W
CAK PIC-Nuc10W
Corrected
Unmasked
Masked
Phase randomized
FSC
Resolution (1/Å)
1/Resolution (1/Å)
FSC = 0.143
B
Model vs Map
FSC = 0.5
C
Top view
Side view
cPIC-Nuc10W
PIC-Nuc10W
cPIC (3.1 Å)
Nuc (3.2 Å)
TFIIH (4.5 Å)
cPIC (3.2 Å)
Nuc (3.5 Å)
CAK (3.8 Å)
90°
D
MAT1 helical region
XPB DRD domain
MAT1 helical region
Rpb3-11 dimerization domains
RPB3
RPB11
Upstream DNA
Non-template DNA
Template DNA
TBP
TFIIF winged helix
TFIIB cyclin folds
XPD ARCH and FeS domains
ARCH domain
FeS domain
MAT1 3-helix bundle
MAT1 RING domain
Nucleosome core particle

Figure S6. Map quality assessment of the PIC-nucleosome^{10W} structure, Related to Figure 7.

(A) Focused refinements and overall maps filtered and coloured by estimated local resolution, shown as top views. The superscript typography of each map indicates whether those particles contained TFIIH. Their angular distributions and resolution (following the FSC 0.143 cut-off criterion) are provided on the right side of each map.

(B) Model-map correlation between the individual focused refinement maps, used for model building, and the respective built atomic models. Resolution following the FSC 0.5 cut-off criterion is marked with a dashed line.

(C) Composite maps showing individual local refinements in distinct views for both cPIC-Nuc^{10W} (lacking TFIIE and TFIIH) and PIC-Nuc^{10W}. The different protein and DNA components are coloured according to the colours used in their structural models, depicted in **Figure 7C** and **S8**.

(D) Cryo-EM density maps (shown in blue surface) fitted to the corresponding region of the modelled structures.

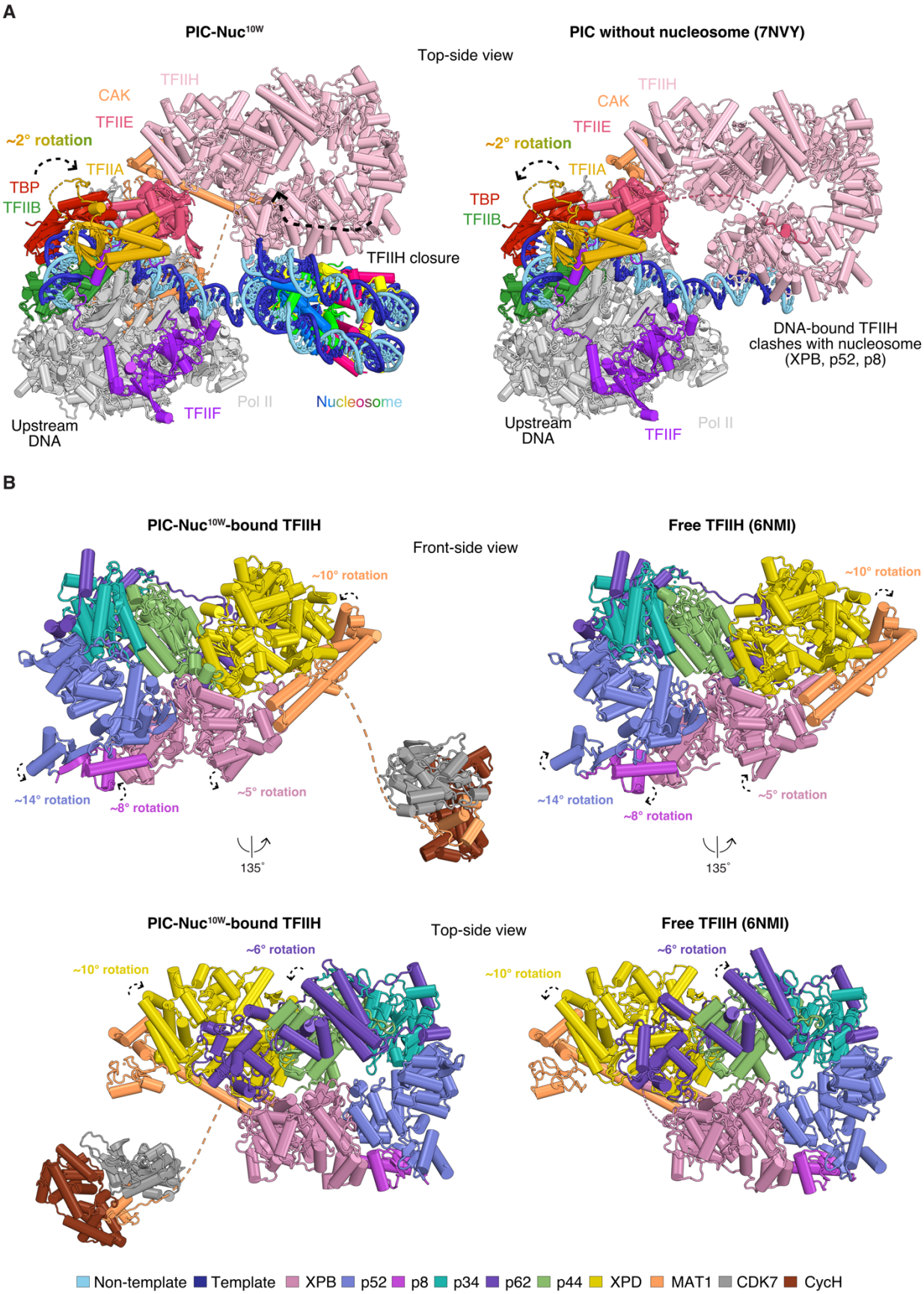

A
PIC-Nuc10W
Top-side view
PIC without nucleosome (7NVY)
TFIIH
CAK
TFIIE
~2° rotation
TFIIA
TBP
TFIIB
TFIIH closure
Pol II
Nucleosome
TFIIF
Upstream DNA
DNA-bound TFIIH clashes with nucleosome (XPB, p52, p8)
B
PIC-Nuc10W-bound TFIIH
Front-side view
Free TFIIH (6NMI)
~10° rotation
~14° rotation
~8° rotation
~5° rotation
135˚
PIC-Nuc10W-bound TFIIH
Top-side view
Free TFIIH (6NMI)
~6° rotation
Non-template
Template
XPB
p52
p8
p34
p62
p44
XPD
MAT1
CDK7
CycH

Figure S7. Comparison of PIC-Nuc^{10W} and PIC-Nuc^{10W}-bound TFIIH with previously reported structures without the nucleosome, Related to Figure 8 and 9.

(A) Comparison of PIC-nucleosome^{10W} on a previously reported nucleosome-free PIC structure (PDB ID 7NVY (Aibara et al., 2021)) in top-side view. TFIIH now adopts a closed conformation due to the presence of the nucleosome at the region where XPB normally engages free DNA. Arrowheads on TFIIH indicate the closure of TFIIH. Superimposition of the models was performed by aligning on Pol II. Colour of the different components is denoted adjacent to them.

(B) Comparison of TFIIH from PIC-Nuc^{10W} (PIC-Nuc^{10W}) (left panel) with free TFIIH (right panel) (PDB ID 6NMI (Greber et al., 2019)) in different views. TFIIH adopts different conformations due to the different position of the nucleosome. Arrowheads indicate the local motion of TFIIH subunits. Superimposition of the models was performed by aligning on TFIIH. TFIIH is depicted in cartoon representation and the colour of the different components is denoted exclusively at the bottom legend.

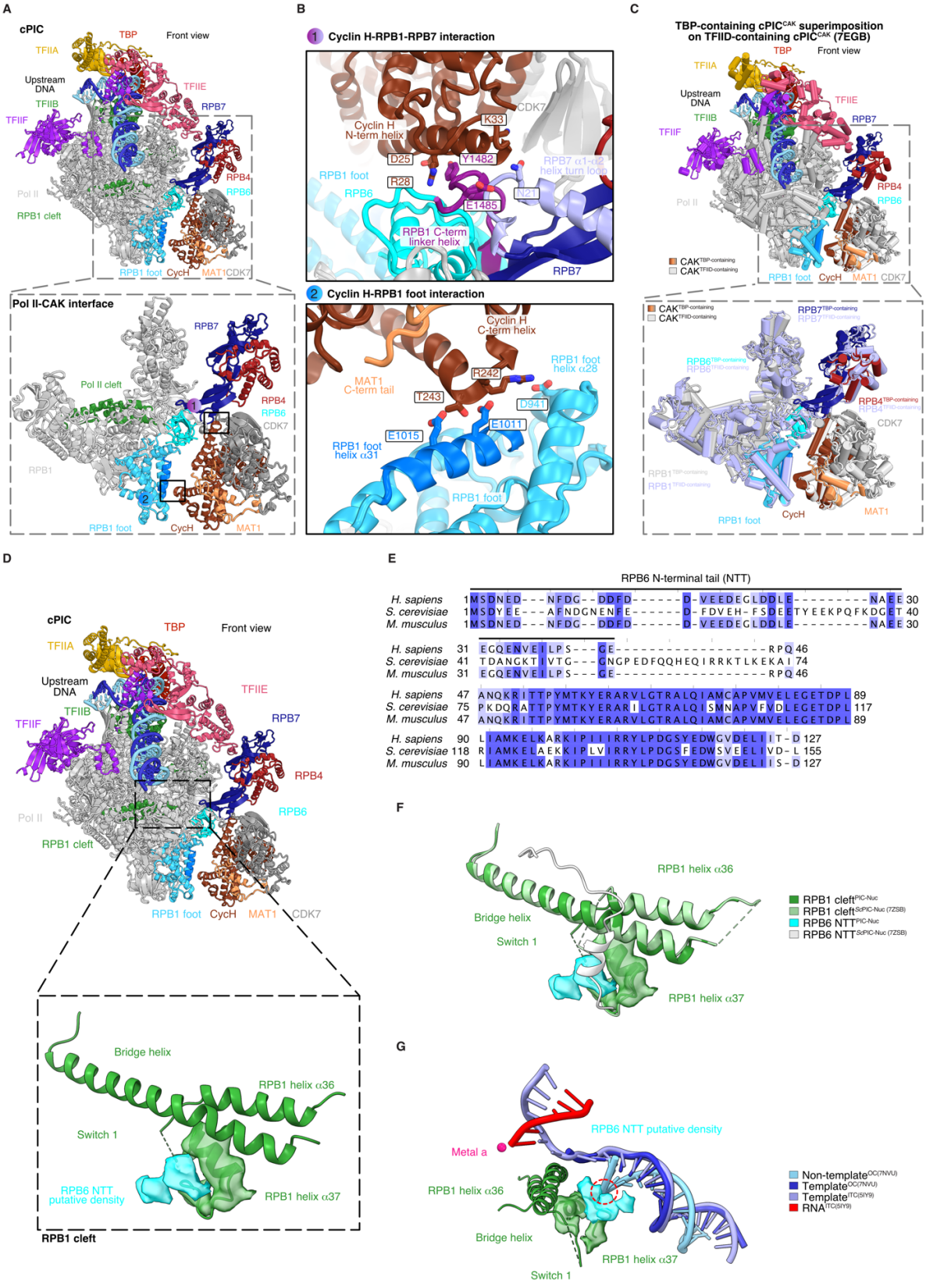

A
cPIC
TBP
Front view
TFIIA
Upstream DNA
TFIIB
TFIIE
TFIIF
RPB7
RPB4
RPB6
Pol II
RPB1 cleft
RPB1 foot
CycH
MAT1
CDK7
Pol II-CAK interface
RPB1
B
1 Cyclin H-RPB1-RPB7 interaction
Cyclin H N-term helix
K33
D25
Y1482
R28
E1485
N21
RPB7 α1-α2 helix turn loop
RPB1 C-term linker helix
2 Cyclin H-RPB1 foot interaction
Cyclin H C-term helix
MAT1 C-term tail
R242
T243
D941
E1011
E1015
RPB1 foot helix α28
RPB1 foot helix α31
C
TBP-containing cPIC^CAK superimposition on TFIID-containing cPIC^CAK (7EGB)
CAK^TBP-containing
CAK^TFIID-containing
RPB7^TBP-containing
RPB7^TFIID-containing
RPB6^TBP-containing
RPB6^TFIID-containing
RPB4^TBP-containing
RPB4^TFIID-containing
RPB1^TBP-containing
RPB1^TFIID-containing
D
Bridge helix
RPB1 helix α36
Switch 1
RPB6 NTT putative density
RPB1 helix α37
RPB1 cleft
E
RPB6 N-terminal tail (NTT)
H. sapiens
S. cerevisiae
M. musculus
F
RPB1 helix α36
Bridge helix
Switch 1
RPB1 helix α37
RPB1 cleft^PIC-Nuc
RPB1 cleft^ScPIC-Nuc (7ZSB)
RPB6 NTT^PIC-Nuc
RPB6 NTT^ScPIC-Nuc (7ZSB)
G
RPB6 NTT putative density
Metal a
RPB1 helix α36
Bridge helix
RPB1 helix α37
Switch 1
Non-template^OC(7NVU)
Template^OC(7NVU)
Template^ITC(5IY9)
RNA^ITC(5IY9)

Figure S8. A PIC-bound nucleosome stabilises TFIIH CAK and favours ordering of RPB6 NTT, Related to Figure 7.

(A) Enlarged view of the Pol II-CAK interaction interface, shown in front view and within a dashed box. CAK engages via two contact points between the Pol II stalk (RPB4/RPB7) and foot (RPB1) via its subunit Cyclin H. TFIIH and the nucleosome are omitted to facilitate visualisation.

(B) Cyclin H interacts with Pol II through its N- and C-terminal regions. The N-terminal region contacts electrostatically both RPB1 C-terminus linker helix and RPB7 α1-α2 loop, whereas its C-terminus interacts with RPB1 foot helices α28 and α31. Interacting residues are enclosed in rectangular boxes.

(C) Comparison of CAK between TBP-containing (this study) and TFIID-containing PIC complexes (PDB ID 7EGB (Chen et al., 2021a)), shown in front view within a dashed box. There are no perceivable differences between these complexes, except for inherent protein local motion.

(D) Magnification of the Pol II cleft, shown in front view and within a dashed box. Putative RPB6 NTT density locates above the helix α31 of Pol II cleft. TFIIH and the nucleosome are omitted to facilitate visualisation.

(E) Protein sequence alignment of RPB6 orthologs in *H. sapiens*, *S. cerevisiae* and *M. musculus*. The alignment was performed with Jalview 2.11.2.4 (Waterhouse et al., 2009) using the T-Coffee algorithm. The RPB6 NTT is denoted with a solid black line. The colour code indicates the conservation scoring across different organisms.

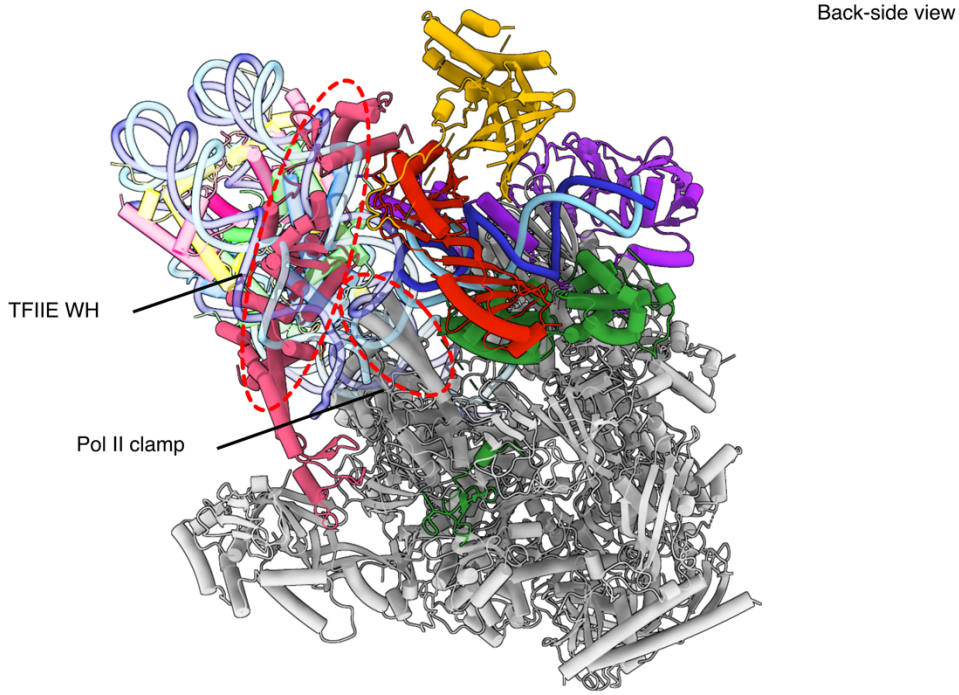
Overlay of the putative RPB6 NTT density of mammalian PIC-nucleosome (this study) on a yeast PIC-nucleosome complex (PDB ID 7ZSB (Wang et al., 2022)) (**F**), a nucleosome-free mammalian PIC complex with opened promoter DNA (PDB ID 7NVU (Aibara et al., 2021)) and a human PIC-initial transcribing complex (ITC, PDB ID 5IY9 (He et al., 2016)) (**G**). The putative NTT clashes with the loaded DNA of opened promoter DNA complexes, suggesting the inhibitory function of the nucleosome favours the ordering of such tail.

Colours of the different components in all panels is denoted adjacent to their corresponding subunits and by legends. Superimposition of models shown was performed by aligning on Pol II.

.

A

cPIC-Nuc^{10W} superimposition on the open promoter complex (7NVU)



B

cPIC-Nuc^{18W} superimposition on the open promoter complex (7NVU)

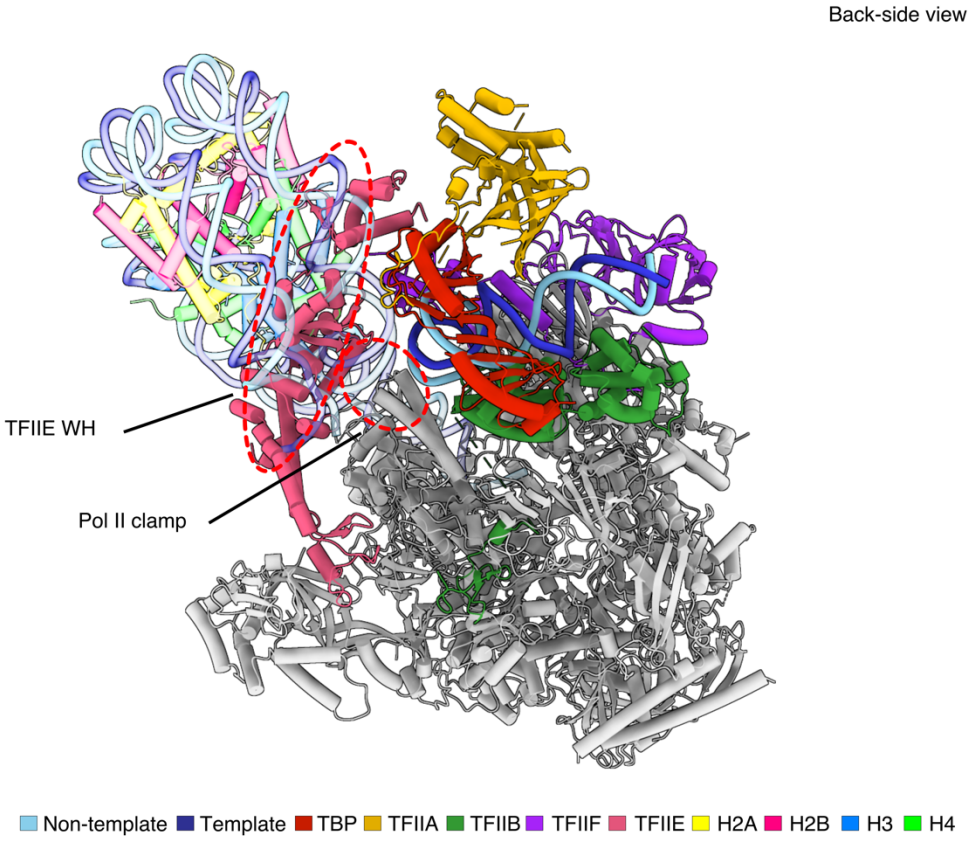


Figure S9. Core PIC-nucleosome complexes are incompatible with DNA opening, Related to Figure 5 and 7.

(A) **(B**) Modelling of our PIC-nucleosome structures missing TFIIH (cPIC) on an open promoter complex (OC, PDB ID 7NVU (Aibara et al., 2021)), shown in back-side view. TFIIH-independent spontaneous promoter opening would not be compatible on a nucleosomal template since the nucleosome would clash with the Pol II clamp and TFIIE winged helix (WH). Superimposition of the models was performed by aligning on Pol II and the DNA. Steric clashes are indicated with arrowheads and red dashed ovals, proteins are depicted in cartoon representation and the colour of the different components is denoted exclusively at the bottom legend.

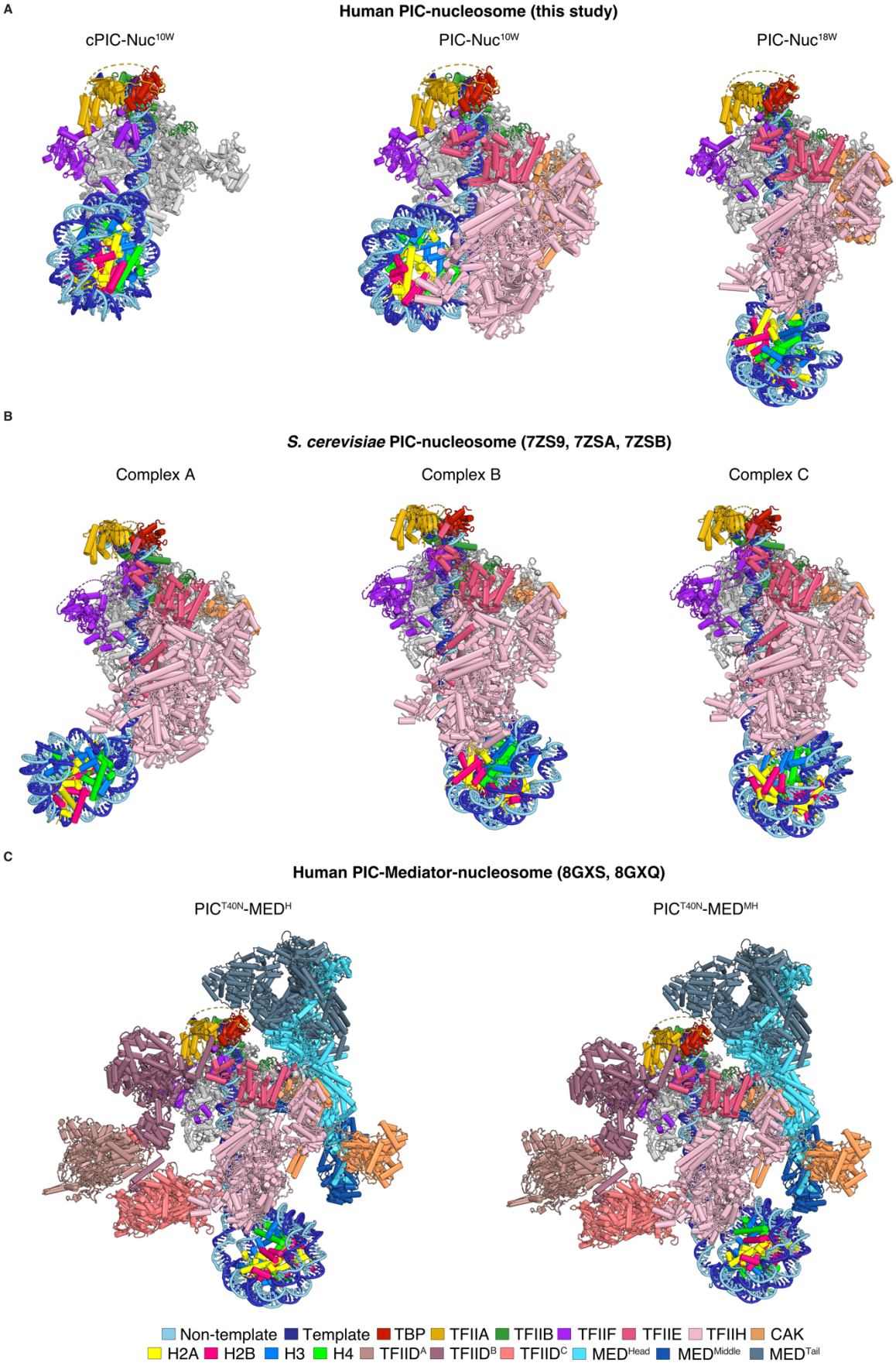

A
Human PIC-nucleosome (this study)
cPIC-Nuc10W
PIC-Nuc10W
PIC-Nuc18W
B
S. cerevisiae PIC-nucleosome (7ZS9, 7ZSA, 7ZSB)
Complex A
Complex B
Complex C
C
Human PIC-Mediator-nucleosome (8GXS, 8GXQ)
PICT40N-MEDH
PICT40N-MEDMH
Non-template
Template
TBP
TFIIA
TFIIB
TFIIF
TFIIE
TFIIH
CAK
H2A
H2B
H3
H4
TFIIDA
TFIIDB
TFIIDC
MEDHead
MEDMiddle
MEDTail

Figure S10. Comparison of human and yeast PIC-nucleosome complexes, related to Figure 12.

Models of the PIC-nucleosome complexes determined in (**A**) this study, (**B**) in yeast (PDB IDs 7ZS9, 7ZSA, 7ZSB (Wang et al., 2022)) and (**C**) in human (PDB IDs 8GXS, 8GXQ (Chen et al., 2022)) Superimposition of the models was performed by aligning on Pol II and the representations shown in the same view. Colour of the different components is denoted exclusively at the bottom legend.

Supplemental tables

Table S1. Cryo-EM data acquisition, processing, and refinement statistics, Related to Figure 7.

Model name	cPIC-nucleosome^{10W}	PIC-nucleosome^{10W}	PIC-nucleosome^{18W}
PDB code	(PDB 8BZ1)	(PDB 8BYQ)	(PDB 8BVW)
Map code	(EMD-16335)	(EMD-16331)	(EMD-16274)
Data collection and processing			
Magnification	81,000x	81,000x	81,000x
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	50.45	50.45	41.58
Defocus range (μm)	0.6 – 1.3	0.6 – 1.3	0.5 – 1.5
Pixel size (Å)	1.05	1.05	1.05
Initial particle images (no.)	4,606,320	4,606,320	4,667,603
Symmetry imposed	C1	C1	C1
Final particle images (no.)	214,161	101,994	188,832
Map resolution (Å) at FSC = 0-143	3.8	4.1	4.0
Map resolution range (Å)	2.8 – 7.7	2.9 – 8.2	2.7 – 8.7
Map sharpening *B* factor (Å^2)	-90	-150	-114
Refinement			
Initial models used (PDB code)	7NVS, 7OHC	7NVS, 7OHC, 6NMI, 6XBZ, 7EGB	7NVS, 7OHC, 7ZSB, 7NVW, 6XBZ, 7EGB
Model composition			
Non-hydrogen atoms	53411	86010	84811
Protein residues	5667	9687	9510
Nucleotides	396	396	412
Ligands	ZN: 9	ZN: 16	ZN: 17
	MG: 1	MG: 1	MG: 1
		SF4: 1	SF4: 1
		UNK: 154	UNK: 27

Mean B factors (Å^2)			
Protein	117.33	176.57	125.41
Nucleotides	145.93	212.46	212.92
Ligand	146.29	223.85	136.37
R.m.s. deviations			
Bond lengths (Å)	0.004	0.005	0.006
Bond angles (°)	0.737	0.860	0.896
Validation			
MolProbity score	1.22	1.47	1.46
Clashscore	4.46	6.52	5.48
Poor rotamers (%)	0.00	0.06	0.04
Ramachandran plot			
Favoured (%)	98.19	97.44	96.93
Allowed (%)	1.81	2.56	3.07
Disallowed (%)	0.00	0.00	0.00

PDB: Protein Data Bank; EMD: Electron Microscopy Data Bank; FSC: Fourier Shell Correlation; R.m.s.: root-mean-square