Structure of the Pol II initiation complex with TFIIH and core Mediator and mechanistic implications for transcription

Dissertation

for the award of the degree

'Doctor rerum naturalium'

of the Georg-August-Universität Göttingen

within the graduate program

'Biomolecules – Structure, Function and Dynamics'

of the Georg-August University School of Science (GAUSS)

submitted by

Sandra Schilbach

from Oelsnitz i.Vogtland, Germany

Göttingen 2017

Members of the Thesis Committee

Prof. Dr. Patrick Cramer
      Department of Molecular Biology
      Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Holger Stark
      Department of Structural Dynamics
      Max Planck Institute for Biophysical Chemistry, Göttingen

Dr. Vladimir Pena
      Research Group Macromolecular Crystallography
      Max Planck Institute for Biophysical Chemistry, Göttingen

Members of the Examination Board

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

Prof. Dr. Holger Stark   (2nd Referee)
      Department of Molecular Biology
      Max Planck Institute for Biophysical Chemistry, Göttingen

Further members of the Examination Board

Prof. Dr. Ralf Ficner
      Department for Molecular Structural Biology
      Georg-August-University, Göttingen

Dr. Vladimir Pena
      Research Group Macromolecular Crystallography
      Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Marina Rodnina
      Department of Physical Biochemistry
      Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Henning Urlaub
      Research Group Bioanalytical Mass Spectrometry
      Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 19th of February 2018

Affidavit

I, Sandra Schilbach, hereby declare that my dissertation entitled 'Structure of the Pol II initiation complex with TFIIH and core Mediator and mechanistic implications for transcription' has been written independently and with no other sources and aids than quoted. This dissertation or parts thereof have not been submitted elsewhere for any academic award or qualification. The electronic version of this dissertation is congruent to the printed versions both in content and in format.

Göttingen, 31st of December 2017

....

(Sandra Schilbach)

Acknowledgements

Many persons who I met during my time as a PhD student have become not only colleagues but also friends to me. I would like to thank all of them all for brightening this experience with countless conversations and by providing personal support, entertainment during dull times, and company on the frequent long evenings in the lab. I hope that our friendships will continue when this PhD will have ended.

Looking back onto the path that I have come, I want to express my deepest gratitude towards my family and in particular to my parents, who always gave me the opportunity to follow my heart any my interest. I would like to thank them for their unconditional support, their understanding of my passion for science and for their interest in my work. I thoroughly enjoyed our discussions, which frequently helped me to divert my thoughts and change perspective.

Finally, I would like to thank Michael for being at my side in those last years, for sharing so many joyful and sad moments with me, for his patience with my frequent absences and for his belief in me.

Publications

Part of this work has been published:

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H. & Cramer, P. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551,** 204-209 (2017).

The following chapters and sections of this work are excerpted from Schilbach et *al.*:

Chapter 4.5   Project-specific techniques and experimental setups

- Generation of poly-promoter expression constructs for core-TFIIH and the TFIIH kinase module
- Expression and purification of *S. cerevisiae* core-TFIIH
- Expression and purification of the *S. cerevisiae* TFIIH kinase module
- Preparation of the PIC-cMed complex for cryo-EM
- Chemical crosslinking and crosslink analysis by mass spectrometry
- Cryo-EM data collection
- Image processing
- Structural modeling

Chapter 5.3   Reconstitution and structural analysis of transcription pre-initiation complexes with TFIIH and mediator

- Section 5.3.1   Structures of the PIC and PIC-cMed complex
- Section 5.3.2   TFIIH structure
- Section 5.3.3   TFIIH interactions with cPIC
- Section 5.3.4   TFIIH and DNA opening
- Section 5.3.5   TFIIH and Pol II phosphorylation

Chapter 6   Conclusion

The following items of this work are excerpted from Schilbach et *al.*:

Section 5.1.4 Preparation of recombinant core-TFIIH complexes

-   Figure 5.5 (Preparation of recombinant TFIIH); corresponds to Extended Data Figure 1 a in Schilbach *et al.*

Section 5.3.1 Structures of the PIC and PIC-cMed complex

-   Figure 5.9 a-c (Preparation of the PIC-cMed complex); corresponds to Extended Data Figure 1 b-d in Schilbach *et al.*
-   Figure 5.10 a-d (Cryo-EM data processing and quality of reconstruction); corresponds to Extended Data Figure 2 a-d in Schilbach *et al.*
-   Figure 5.11 (Structure of the Pol II PIC); corresponds to Figure 1 in Schilbach *et al.*
-   Figure 5.12 a-d (EDC crosslinking analysis of PIC-cMed); corresponds to Extended Data Figure 3 a-d in Schilbach *et al.*
-   Figure 5.13 (Structure of the PIC-cMed complex); corresponds to Figure 2 in Schilbach *et al.*

Section 5.3.2 TFIIH structure

-   Figure 5.14 a, b (Structure of TFIIH); corresponds to Figure 3 a, b in Schilbach *et al.*
-   Figure 5.15 a-m (TFIIH structure and quality of the cryo-EM density); corresponds to Extended Data Figure 4 a-m in Schilbach *et al.*
-   Figure 5.16 a, b (Location of essential regions in TFIIH and sites mutated in disease); corresponds to Extended Data Figure 5 a, b in Schilbach *et al.*

Section 5.3.3 TFIIH interactions with cPIC

-   Figure 5.17 a, b (Interactions of TFIIH with cPIC); corresponds to Figure 4 a, b in Schilbach *et al.*
-   Figure 5.18 a-g (TFIIE-TFIIH interactions); corresponds to Extended Data Figure 6 a-g in Schilbach *et al.*

Section 5.3.4 TFIIH and DNA opening

-   Figure 5.19 a, b (TFIIH and DNA opening); corresponds to Figure 5 a,b in Schilbach *et al.*
-   Figure 5.20 a-f (Detailed analysis of Ssl2 ATPase conformation and implications for translocase activity); corresponds to Extended Data Figure 7 a-f in Schilbach *et al.*

Section 5.3.5 TFIIH and Pol II phosphorylation

-   Figure 5.21 a, b (TFIIH and phosphorylation of Pol II); corresponds to Figure 6 a, b in Schilbach *et al.*
-   Figure 5.22 a-e (Structure and conformational changes of cMed); corresponds to Extended Data Figure 8 a-e in Schilbach *et al.*

Section 8.4.1 Composition of the PIC-cMed complex

- Table 8.1 (Components of the PIC-cMed complex); corresponds to Extended Data Table 1 in Schilbach *et al.*

Section 8.4.2 Modeling of TFIIH

- Table 8.2 (Modeling of yeast TFIIH subunits, domains and regions); corresponds to Supplementary Table 1 in Schilbach *et al.*

Section 8.4.3 Modeling of cPIC and cMed

- Table 8.3 (Adaptation and extension of yeast cPIC and cMed models); corresponds to Supplementary Table 4 in Schilbach *et al.*

Section 8.4.4 Cryo-EM data collection and model statistics

- Table 8.4 (Cryo-EM data collection and model statistics for the PIC and the PIC-cMed complex structures); corresponds to Extended Data Table 2 in Schilbach *et al.*

Section 8.4.5 Crosslinking information

- Table 8.5 a, b (Compilation of published BS3- and SBAT- crosslinking information); corresponds to Supplementary Table 2 a, b in Schilbach *et al.*
- Table 8.6 a, b (Newly obtained EDC-crosslinks); corresponds to Supplementary Table 3 a, b in Schilbach *et al.*

The design of Tables 8.5 and 8.6 has been adapted to this work.

The following elements have been adapted to the format of this work and may deviate from Schilbach *et al.*:

- Numbering of figures and tables
- References to figures and tables
- References to literature
- Cross-references to figures or tables within figure legends

References to supplementary videos and supplementary data were removed.

The excerpted text was adapted to American English.

Table of Contents

1   Summary

The transcription of protein coding genes is a key step in the conversion of genomic information into the content of a cell's proteome. In eukaryotes, this process is performed by the multi-subunit DNA-dependent RNA polymerase (Pol) II and subject to intricate regulatory systems. During the consecutive phases of transcription, initiation, elongation and termination, Pol II associates with various factors that modulate its activity and thereby aid in the coordination of gene expression. Transcriptional control at the level of initiation is dominated by the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. At the core-promoter Pol II assembles with the GTFs in an ordered manner to form the pre-initiation complex (PIC), which is competent to induce the opening of promoter DNA and to commence RNA synthesis. The subsequent transition from initiation to the elongation phase is preluded by poorly understood events that result in detachment of Pol II from the promoter ('promoter escape') and disassembly of the initiation machinery. Both promoter opening and escape mark pivotal initiation checkpoints that are susceptible to regulation, and for both processes an ATP-dependent involvement of the TFIIH complex has been demonstrated. TFIIH equals Pol II in size and intricacy, consists of two dissociable 'core' and 'kinase' modules, and comprises three subunits with conserved enzymatic function. Two of these, a DNA-translocase and a kinase, act at the distinct stages of initiation, respectively. In addition to the GTFs, promoter-located Pol II associates with the global co-activator Mediator. This multi-protein complex bridges between the initiation machinery and transcription factors bound at distant enhancer DNA elements and can adapt transcription in response to environmental and developmental stimuli. The 'head' and 'middle' module segments of Mediator (core Mediator, cMed) are essential and directly contact the initiation complex. Mediator promotes initiation by cooperatively enhancing the recruitment of Pol II and the GTFs to the core-promoter, stabilizing the assembled PIC and stimulating TFIIH kinase activity. Owing to the sheer size and complexity of Mediator and TFIIH, however, the underlying molecular mechanisms that govern the dynamic progression of the initiation machinery through the stages of PIC formation, promoter opening and promoter escape remain largely elusive. An extensive structural characterization of initiation complexes containing these factors may contribute to deduce their essential functions in detail but to date has been impeded by their flexibility and poor biochemical stability, as well as by the lack of highly pure samples for analysis.

In an effort to overcome such limitations, the first protocol for the large-scale preparation of recombinant full-length TFIIH from the yeast *Saccharomyces cerevisiae* was established. This work describes effective co-expression approaches in *Escherichia coli* and insect cell systems that yielded various TFIIH subcomplexes. Using purified TFIIH core and kinase modules, the assembly of complete 10-subunit TFIIH was demonstrated. In addition, this work reports the *in vitro* reconstitution and cryo-electron microscopy (EM) analysis of the yeast PIC-cMed complex, a macromolecular 46-subunit assembly of ~2 MDa, which encompasses all initiation-related proteins essential for cell viability in yeast. Reconstructions of the TFIIH-comprising PIC and PIC-cMed complex were derived at nominal resolutions of 4.7 Å and 5.8 Å, respectively. The

obtained cryo-EM maps reveal secondary structure throughout and provide the first visualization of TFIIH within the initiation machinery at high resolution. To facilitate model building and placement into the density, crosslinking experiments were conducted, which in particular aided to identify interactions between TFIIE and TFIIH. Based on previously determined models for the cMed complex and the PIC without TFIIH (core PIC, cPIC), on newly generated homology and *ab initio* models for multiple domains in TFIIE and TFIIH, and on *de-novo* built segments, the to date most complete and accurate structure of a transcription initiation complex was compiled. The remaining, unassigned protein sequences largely comprise regions that are predicted to be disordered and thus likely adopt flexible conformations within the structure.

The PIC and PIC-cMed models reveal interactions between cMed, the cPIC and TFIIH, and in particular demonstrate how TFIIE anchors TFIIH to the cPIC. The subunit arrangement within the cPIC and cMed, as well as previously proposed interfaces between cMed, the Rpb4/7 stalk and the foot domain of Pol II, and TFIIB were confirmed and further explicated. Upon binding to the PIC, the cMed middle module undergoes significant rearrangements, which result in partial loss of its contacts to the head module and in 'opening' of the cMed structure. The previously observed 'cradle' that is formed by the cPIC and cMed and may accommodate the C-terminal domain (CTD) of Pol II is further defined by TFIIH. The core-TFIIH module emanates from the cPIC in a crescent-like shape, with its ATPases Ssl2 (XPB in human) and Rad3 (XPD) at the ends of the lobes and in proximity to the cPIC. Rad3 is anchored to the cPIC by a TFIIH kinase module subunit, Tfb3 (MAT1), and Ssl2 engages with downstream promoter DNA. The ATPases are connected via the scaffolding subunits Ssl1 (p44), Tfb2 (p52), Tfb4 (p34) and Tb5 (p8). The previously uncharacterized subunit Tfb1 (p62) serves as a flexible tether between various parts of core-TFIIH. Tfb3 directly contacts Pol II by binding in a groove between the Pol II stalk and the large TFIIE subunit Tfa1 (TFIIEα). To accommodate Tfb3, the stalk of Pol II is shifted in both structures. Additional interactions between TFIIH and the cPIC involve the C-terminal section of Tfa1. Several novel Tfa1 elements that bind the TFIIH subunits Tfb1 and Ssl2 were detected, thus indicating a mechanism for TFIIE-mediated TFIIH recruitment to the PIC. The structures also provide further information on the function of the catalytic TFIIH subunits during promoter opening and escape. The ATPase Ssl2, which is implicated in the unwinding of promoter DNA, is contacted by the newly assigned 'E-bridge' helix in Tfa1, suggesting a concept for TFIIE-stimulated DNA opening. Comparison of the Ssl2 ATPase domains to those of enzymes, which translocate on double-stranded DNA (dsDNA), revealed striking similarity and a respective Ssl2 translocation mechanism was proposed, consistent with biochemical evidence. The kinase module of TFIIH comprises the kinase Kin28 (CDK7), which phosphorylates the Pol II CTD and stimulates promoter escape. Except for the N-terminus of Tfb3, which anchors it to Rad3 and the cPIC, this module is flexible in the PIC structure. However, it adopts a preferred position outside of the cradle and in proximity to Mediator in the PIC-cMed complex. Moreover, it is located adjacently to one of two openings in the cradle that emerge after conformational rearrangement of cMed upon its PIC incorporation, thereby implying how Kin28 may access the CTD and how Mediator may enhance its kinase activity.

In summary, the derived structures pose a major advance towards the elucidation of the molecular mechanisms governing transcription initiation and may be used as a framework and guidance for their further biochemical, functional and structural characterization.

2 Introduction

2.1 Overview of transcription

The central dogma of molecular biology, first formulated by Francis Crick in the 1950's, outlines the unidirectional conversion of a cell's genomic information into the content of its proteome[1]. Interpreted in a simplistic way, this information transfer is divided in two major processes that are coupled in prokaryotes but spatially and temporally segregated in eukaryotic cells. While the transcription of genomic DNA into RNA occurs in the nucleus, the subsequent translation is performed in the cytoplasm by ribosomes, which interpret the generated RNA transcripts and synthesize corresponding polypeptide chains[2,3]. *In vivo*, however, translation is not an exclusive or inevitable consequence of transcription. Instead it requires a particular form of RNA molecules, protein-coding messenger RNAs (mRNAs) that carry specific sequence features and modifications, as input templates. Further transcription products, which are generally referred to as non-coding RNAs, are highly variable and abundant within the cell, have a wide spectrum of regulatory functions and serve as structural components of intricate ribonucleoprotein (RNP) machineries[4]. Despite their diverse roles and features, all RNA transcripts are synthesized by one class of enzymes, the DNA-dependent RNA polymerases (Pols).

Although these macromolecular multi-subunit complexes are highly conserved throughout the three kingdoms of life, archaea and bacteria comprise only one type of Pol, whereas a set of slightly divergent Pols has evolved in eukaryotes[5]. With the exception of chloroplasts and mitochondria, which contain bacterial- and phage-derived polymerases[6,7], eukaryotic cells utilize up to five distinct nuclear Pols, referred to as Pol I – Pol V. They share a conserved core and active center and catalyze the same basic reaction but target different DNA templates and generate a variety of RNA classes[8]. Pol IV and Pol V, which synthesize non-coding RNAs involved in gene silencing, are plant-specific[9] while Pol I – Pol III are common to all eukaryotic cells. Pol I and Pol III produce the precursor of the 5.8S, 18S and 28S ribosomal RNA (rRNA)[10] and other essential non-translated transcripts like transfer RNA (tRNA), the ribsosomal 5S rRNA, small nuclear RNA (snRNA) or small nucleolar RNA (snoRNA)[11], respectively. Pol II, although also transcribing a variety of non-coding RNA classes, mainly generates pre-mRNAs, the mRNA precursors. Pre-mRNAs undergo co- and post-transcriptional processing in the nucleus and are exported to the cytoplasm as mature mRNAs for ribosomal translation[12,13]. Pol II mediated transcription is the cell's sole source of mRNA transcripts and therefore a key determinant of its proteome. Owing to the pivotal role of Pol II and the high diversity of its DNA templates, which are located throughout the entire genome, its regulation requires a particularly sophisticated system of control mechanisms that have to engage effectively at various stages of the Pol II transcription cycle.

2.2   RNA Polymerase II transcription cycle and role of the CTD

The transcription process can be divided into three discrete phases termed initiation, elongation and termination. After termination, the majority of polymerases undergo recycling and re-initiation, thereby completing the transcription cycle (Fig. 1.1). Each stage of transcription is subject to various regulatory mechanisms that involve specific sets of auxiliary proteins, which directly or indirectly associate with Pol II, transcribed DNA, adjacent histones, or the nascent RNA chain, and are collectively termed transcription factors (TFs)[14,15]. In addition, the distinct phases are characterized by explicit modification patterns of the C-terminal domain (CTD) of Rpb1, the largest Pol II subunit. The CTD comprises a series of tandem heptapeptide repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$, which constitute an unstructured, low complexity domain that protrudes from the Pol II core. The length of the CTD, i.e. the number of its repeats, and their individual divergence from the consensus sequence are species-specific[16,17]. During transcription, diverse post-translational modifications (PTMs) of the CTD, such as phosphorylation, proline-isomerization or methylation of non-canonical lysine and arginine residues, are dynamically introduced, interpreted and removed. These establish the 'CTD code', a key factor in the coordination of transcription-coupled signaling and transcriptional regulation[18,19]. While many types of PTMs have been reported, the effects of CTD phosphorylation at serine and tyrosine residues have been studied most extensively. The phosphorylation pattern of the CTD essentially impacts its propensity to serve as a docking platform for specific classes of accessory proteins throughout the transcription cycle and thus governs transitions between its consecutive (sub-)phases. For example, components of co-transcriptionally acting machineries, such as capping enzymes or splicing factors, are recruited to the transcribing Pol II in the early and productive stages of elongation by CTD residues that were phosphorylated in the respectively preceding initiation and early elongation phases[20]. The CTD is a unique feature of Pol II and homologous domains have not been observed for the remaining nuclear Pols, emphasizing the requirement of an intricate control system for this particular type of enzyme.

The first stage of the transcription cycle, transcription initiation, begins with formation of the pre-initiation complex (PIC) when Pol II is recruited to promoters and assembles with the general transcription factors (GTFs) on closed DNA in proximity to the transcription start site (TSS)[21,22]. During transition from a closed to an open state of the PIC unwinding of a short promoter-specific DNA region is induced, resulting in generation of an initial 'transcription bubble'. DNA opening permits positioning of the template DNA strand in the active center of Pol II and promotes subsequent synthesis of a complementary RNA strand[23]. Initiation complexes containing short RNA transcripts, referred to as initially transcribing complexes (ITCs), are unstable and can undergo abortive initiation[24]. When the nascent RNA exceeds a critical length of 13-14 nucleotides, however, the initiation machinery disassembles in a poorly understood event termed promoter clearance or escape and the GTFs are exchanged with elongation factors[25]. In metazoans, the early elongation phase is frequently interrupted by promoter-proximal pausing of Pol II, particularly at genes that are part of stimulus-responsive pathways. Affected polymerases are sequestered by pausing factors in response to poorly characterized cellular signaling

mechanisms and stall after synthesis of 20-60 RNA nucleotides until the transcription block is relieved. The recruitment of pausing factors also correlates with a specific CTD phosphorylation pattern that is gradually transformed into that of an elongation-competent Pol II until pause release is triggered[26,27]. Subsequently, Pol II enters the phase of productive elongation, in which RNA transcripts are synthesized with rates of up to 4 kb/min[28,29]. The binding of elongation factors and the incorporation of the rigid DNA-RNA hybrid at the center of the elongation complex contribute to its enhanced stability, speed and processivity and assist in coordinating the action of co-transcriptional mRNA processing machineries and chromatin remodelers[30-32]. Although the architecture of an extended Pol II elongation complex, which contains multi-protein accessory factors like the polymerase associated factor 1 (Paf1)[33], has only been determined on a topological level[34], high-resolution models of the core elongation complex[35,36] have been derived. In addition, details of the RNA nucleotide addition cycle, including RNA-base selection, its incorporation into the growing nucleotide chain and translocation of Pol II, have been revealed[37].
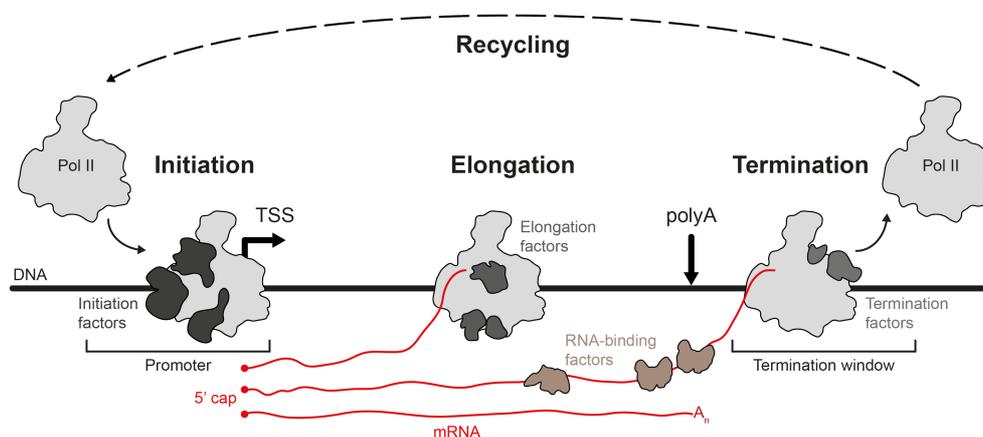
Figure 1.1 | The Pol II transcription cycle. To form the initiation complex Pol II assembles with the general transcription factors at the core-promoter in close proximity to the transcription start site (TSS). Following promoter opening, initial RNA synthesis and promoter escape, the nascent mRNA transcript is extended in the elongation phase. The end of a transcription unit (gene) is marked by the polyadenylation (polyA) site, which provides the signal for mRNA cleavage. Pol II dissociates from the DNA template downstream of the polyA site and may undergo recycling to begin a new round of transcription. Produced mRNAs are co- and post-transcriptionally processed by addition of a 5' cap (red dot) and a 3' poly(A) tail (A_n), respectively. Adapted from[37].

Once Pol II reaches the end of a transcription unit, marked by a polyadenylation (polyA) signal in the DNA sequence, the newly synthesized and partially processed mRNA transcript is released and transcription is terminated by dissociation of Pol II from the template DNA[37]. Amongst the distinct phases of the Pol II transcription cycle, termination remains most ambiguous and alternate concepts are still discussed. The allosteric model is based on an indirect destabilization and consequent disassembly of the elongation complex upon transcribing into the polyA site and binding of 3' RNA-processing factors[38], whereas the torpedo model suggests an active displacement of Pol II from the complex by a RNA-exonuclease that is recruited after release of the mRNA transcript[39]. However, prevalence of one of the proposed models, as well as the identity and function of further prospective termination factors remain yet elusive. Structural

information on termination intermediates is not available although conformational and compositional rearrangements in the elongation complex upon its transition into a pre-termination complex have been indicated[40]. Following Pol II dissociation from the DNA, the phosphorylation marks on the CTD are removed and its original hypo-phosphorylated state is restored. Regenerated, transcription-competent Pol II may then relocate to initiation complexes to restart a new transcription cycle. In a chromatin context this process is facilitated by gene-looping and auxiliary factors that mediate interactions between promoters and terminators and position them in close spatial proximity[15].

2.3   Outline of transcription initiation and its regulation by the general transcription factors

The accurate yet dynamic regulation of Pol II is pivotal as it underlies developmental processes like cell fate determination and differentiation, but is simultaneously required for the maintenance of homeostasis within diverse cell types. Consequently, transcription undergoes constant and rapid adaptation in response to environmental signals or to changes in cellular or tissue context[41,42]. To sustain such a system, the Pol II machinery recruits a plethora of accessory proteins that coordinate transcriptional activity by conferring specificity to the Pol II enzyme, influencing its processivity and coupling pre-mRNA synthesis to co-transcriptional events. Although control mechanisms may take effect at all phases of the transcription cycle, they usually target the first stages of gene expression. Promoter-proximal pausing is typically observed as an early elongation checkpoint in metazoans[26,43], but the precise regulation of transcription at the level of initiation is most common in all eukaryotic species[44,45]. In this phase the transcription machinery passes several key events that are prone to intervention, such as PIC assembly, DNA melting or promoter escape. To identify the respectively involved transcriptional regulators, numerous structural, biochemical, and system-wide studies aimed to dissect the intricate processes that govern the transitions between the intermediate states of initiation[42,46-48].

According to the classical model, the first major initiation intermediate, the PIC, is formed on promoter DNA by coordinated stepwise assembly of Pol II with the general (or basal) TFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH[21]. These subsequently aid Pol II during TSS selection, DNA melting, initial RNA synthesis, and promoter escape (Fig. 1.2, Table 1.1). Eukaryotic promoters comprise various target sites for DNA-binding transcription factors that are distributed throughout the core-promoter and adjacent promoter-proximal sequences. Metazoan cells additionally feature enhancer elements in distal regulatory DNA regions[41,49]. Upstream activating sequences (UAS), which were originally identified in the yeast *Saccharomyces cerevisiae*, are located more closely to core-promoters but are functionally related to enhancers[50,51]. Whereas the integration of signals from UAS- or enhancer-associated transcription factors requires bridging adapters like the co-activator Mediator[52], the basal Pol II initiation machinery interacts with DNA directly at core-promoters. These encompass ~40-45 nucleotides in metazoans, extend asymmetrically over the TSS, and contain specific subsets of respective

core-promoter motifs[53,54]. Most biochemical and structural studies of Pol II initiation complexes were performed on core-promoter sequences with TATA-box motifs[55] residing ~25-30 nucleotides upstream of the TSS, although canonical TATA-box elements are present only at a minority (20-30%) of promoters and may be located further upstream, as observed in *S. cerevisiae*[56,57]. Core-promoters mark the site of DNA interaction for the multifunctional transcription factor TFIID, which comprises the TATA-box binding protein (TBP) and 13-14 TBP-associated factors (TAFs). The TAFs contribute to recognition of distinct core-promoter elements, modulate TBP specificity and supposedly assist in its correct positioning on promoter DNA in absence of canonical TATA-box motifs. However, they are still subject to intensive analysis[48,54,58-60]. In contrast, the structure and function of TBP, which is essential for canonical transcription and universally conserved, has been elucidated and extensively studied in various initiation complexes[8]. TBP binds promoter DNA upstream of the TSS, preferentially at TATA-box elements, and bends the DNA by ~90°, thus providing a pseudo-symmetric platform for the recruitment of further GTFs[61-63]. Owing to the complexity and poor biochemical stability of holo-TFIID, it is mostly substituted by TBP for *in vitro* experiments.

TBP-DNA complexes are stabilized by the accessory factor TFIIA, which sequesters promoter DNA upstream of the TATA-box motif, inhibits attachment of regulatory proteins and stimulates basal and activated transcription[64-67]. Guided by its association with the TATA-box flanking regions that may contain up- or downstream positioned TFIIB recognition elements (BREs), TFIIB subsequently binds in a unidirectional mode, implying a mechanism for oriented PIC assembly despite the pseudo-symmetry of the TBP-DNA complex[68-71]. In addition to conferring directionality to the upstream promoter assembly, TFIIB is essential for the recruitment of Pol II to the TBP-DNA-TFIIA-TFIIB complex[72,73] and each of its functional domains engages in extensive interactions with various regions of Pol II[74-76]. Moreover, it facilitates DNA bending by TBP[77] and correct initial positioning of promoter DNA

8
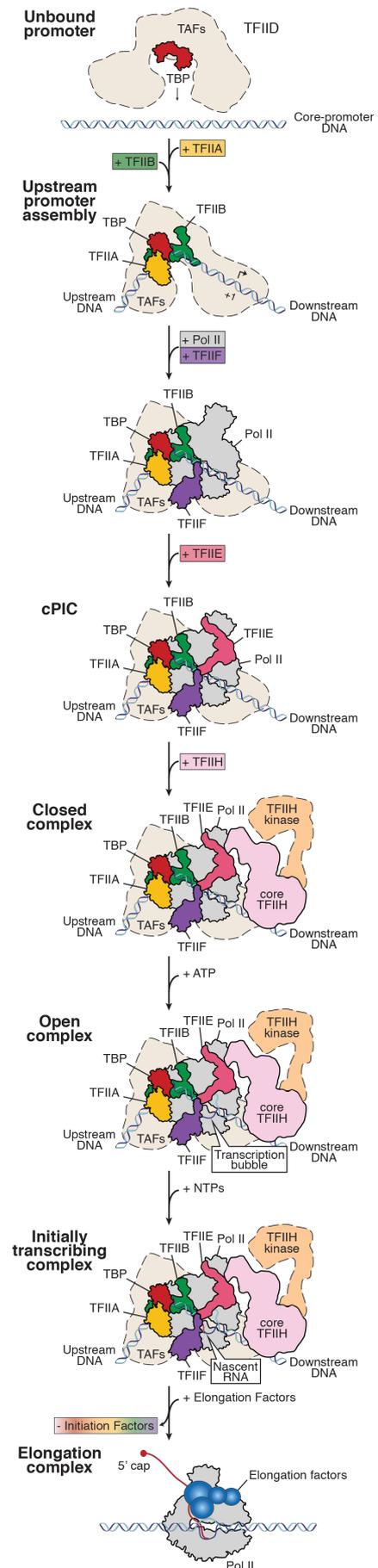
◄ **Figure 1.2 | Schematic of Pol II initiation.** Schematic representation of the classical stepwise model for PIC assembly and of the subsequent promoter opening and escape stages. Descriptions of the respective intermediate states are indicated on the left. PIC assembly begins with promoter recognition by TFIID (light brown) and binding to DNA (shades of blue). The TFIID subunit TBP (red) induces a kink in promoter DNA, which facilitates association of TFIIA (yellow) and TFIIB (green). TFIIA and TFIIB flank the TBP-DNA complex on both sides, thereby stabilizing the resulting upstream promoter assembly. TFIIB aids to recruit the Pol II-TFIIF complex (grey, purple). This intermediate is joined by TFIIE (magenta), forming the cPIC. TFIIE and TFIIF traverse the active center cleft of Pol II and interact with each other, thus further stabilizing the complex. PIC assembly is completed upon binding of TFIIH (light pink/orange). In an ATP-dependent mechanism, TFIIH induces melting of the DNA double strand and generates the initial 'transcription bubble'. After the closed-to-open transition of the PIC, RNA synthesis in the Pol II active center begins. Once a critical length of the nascent transcript in the initially transcribing complex is reached, the initiation machinery disassembles and the general transcription factors on Pol II are exchanged for elongation factors (blue). The dissociable kinase module of TFIIH (orange) is displayed as a separate element. ATP, adenosine triphosphate; NTP, nucleoside triphosphate; PIC, pre-initiation complex; TAF, TBP-associated factor; TBP, TATA-box binding protein. Adapted from[42].

over the Pol II active center cleft[75,78]. Pol II enters the PIC together with TFIIF (Fig. 1.2), an accessory factor that inhibits its unspecific binding to DNA[42]. TFIIF further enhances stability of the formed TBP-DNA-TFIIA-TFIIB-TFIIF-Pol II PIC intermediate[79], in particular by supporting TFIIB[80,81], and aids in TSS selection[81,82]. Similarly to TFIIB, TFIIF interactions also involve numerous contacts to promoter DNA and Pol II subunits[83-89]. Structural studies in the human system revealed that docking of TFIIF results in partial opening of the Pol II clamp and further sequestration and stabilization of the DNA strand within the PIC[48,85,86], which may facilitate its accommodation and access to the active site cleft of Pol II. Consistently, *in vitro* promoter melting can be induced for specific DNA templates without the lastly recruited factors TFIIE and TFIIH, but not without TFIIB and TFIIF[81,90-92].

The final steps of PIC assembly encompass the integration of the GTFs TFIIE and TFIIH, which are primarily involved in promoter opening and the formation of an initial 'transcription bubble' (Fig. 1.2). *In vitro* experiments with pre-melted DNA heteroduplexes demonstrated successful transcription initiation in absence of these factors[92-94]. In the PIC, TFIIE is located between the Pol II clamp and stalk domains and is, similarly to TFIIB and TFIIF, tightly associated with the polymerase. TFIIE further contributes to PIC stabilization, partially by binding a TFIIF domain that traverses the Pol II active center cleft and thereby further encasing loaded promoter DNA[86,89,95]. In the *S. cerevisiae* system, an additional direct contact between TFIIE and the DNA strand at a position close to the upstream edge of the subsequently generated transcription bubble was observed[89,96-98]. This interaction presumably favors formation of an open promoter state[89] and may relate to the previously reported ability of TFIIE to bind single-stranded DNA (ssDNA)[99]. Initiation complexes that contain TBP, DNA, TFIIA, TFIIB, TFIIE and TFIIF, but lack TFIIH, are referred to as core PICs (cPICs). As the GTFs TFIIB, TFIIE and TFIIF are highly modular, they assume their final conformations only upon incorporation into the cPIC and in presence of their interaction partners. Consequently, cPICs consist of an elaborate network of protein-protein and protein-DNA contacts that envelopes the TATA-box and adjacent core-promoter sequences and is required for correct positioning of the initiation machinery on the DNA prior to promoter melting[37]. PIC formation is completed with the recruitment and integration of

the multi-subunit factor TFIIH, which is facilitated by TFIIE[94,100]. Although TFIIH has been structurally characterized within the context of the PIC only on a topological level[85-87,101], an interaction between the large subunit of TFIIE and one of the TFIIH domains has been described in detail[102]. While the illustrated classical scheme for stepwise PIC formation is predominant, potential alternative pathways involving pre-assembled intermediates may be present *in vivo*.

According to the canonical interpretation of initiation, TFIIH essentially promotes generation of the initial transcription bubble and TFIIH activity is indispensable for transition of the PIC from a closed to an open promoter state (Fig. 1.2). Structural studies, crosslinking experiments and biochemical probing have demonstrated that TFIIH comprises an ATP-dependent double-stranded DNA (dsDNA) translocase, Ssl2/XPB (yeast, *S. cerevisiae* / *H. sapiens*)[103-107], which is positioned on DNA downstream of the initially melted region[85-87,96,98,108] and presumably directed away from Pol II and the cPIC[108,109]. Within the restraints of the PIC, Ssl2/XPB translocase action on downstream DNA is assumed to orchestrate threading of DNA into the active center cleft of Pol II, which, in combination with a rotational movement and the generation of torque, promotes DNA unwinding[96,108,109]. The ATPase activity of Ssl2/XPB is stimulated in presence of TFIIE[110,111]. Pol II is the only nuclear polymerase that requires an ATP-consuming co-factor for transcription initiation, reflecting its meticulous regulation. However, the exact mechanism of promoter DNA melting, the underlying principle of its TFIIE-mediated stimulation or the intermediate states adopted by the PIC machinery during initial bubble formation remain elusive. In addition, non-canonical DNA opening events in initiation complexes lacking TFIIH or Ssl2/XPB have been reported *in vitro*[89] and *in vivo*[112], suggesting the existence of further, yet unidentified processes that may be engaged in promoter melting.

Emphasizing the cooperative and synergistic nature of the PIC and its components, TFIIB, TFIIE and TFIIF have been implicated in contributing to efficient DNA unwinding and in maintenance of the generated transcription bubble[37,42]. Consistently, structural analysis of yeast initiation complexes containing closed and open DNA scaffolds indicated significant rearrangements in the TFIIE domain that contacts promoter DNA close to the upstream bubble edge[89]. Both TFIIB and TFIIF fulfill additional post-opening roles in TSS selection[76,82,113,114] and stimulation of initial RNA synthesis[114-116]. TFIIB assists in correct positioning of the DNA in the Pol II active site by interacting with the unwound template strand, and in stabilization of the early ITC[74]. Moreover it may be involved in separation of the DNA- and RNA strands and in directing the newly synthesized RNA chain to the Pol II exit tunnel[76]. TFIIF has been inferred to enhance Pol II processivity by reducing or suppressing pausing events[116,117].

In addition to its pivotal function in melting the DNA double strand during promoter opening, TFIIH also governs the transition from transcription initiation to the early elongation stage after incorporation of the first RNA nucleotides into the nascent pre-mRNA transcript. It comprises a highly active cyclin-dependent kinase, Kin28/CDK7, which targets the Ser5 and, to a lesser extent, Ser7 residues of the Pol II CTD[118-122], as well as TFIIE[110], the elongation factor p-TEFb[123] and transcriptional activators[124,125]. While transcription initiation experiments in reconstituted systems can be successfully performed in the absence of Kin28/CDK7[126] a

requirement of TFIIH-dependent phosphorylation of Pol II for transcription has been demonstrated *in vivo*[127-129]. The Kin28/CDK7-mediated CTD phosphorylation pattern is associated with completion of the transcription initiation phase and subsequent disassembly of the ITC[130,131], although the principles that underlie these events have not yet been resolved. Moreover, CTD phosphorylation by TFIIH has been correlated with dissociation of the co-activator Mediator from the basal transcription machinery in yeast[132-134] consistent with the obligatory release of Pol II from promoter-associated auxiliary factors and its consequent escape from the promoter prior to elongation. Several studies also implied a stimulating function of the Ssl2/XPB ATPase for promoter clearance and subsequent productive transcription, albeit without proposing a coherent mechanism[135-137]. Another critical aspect of the initiation-to-elongation transition is the RNA-induced displacement of TFIIB. Upon reaching a length of ~13-14 nucleotides and its emergence from the RNA exit tunnel, the nascent transcript is expected to sterically clash with the B-ribbon element of TFIIB, resulting in its destabilization and the dismissal of TFIIB from the ITC[74,76]. In summary, distinct processes cooperate to interfere with the stability of the initiation complex and promote the release of Pol II, which subsequently is able to re-associate with elongation factors and to proceed to the next transcription stage (Fig. 1.2). Nevertheless, promoter escape continues to be one of the most poorly understood checkpoints of the transcription cycle.

Table 1.1 | Basal transcription factors of the Pol II initiation complex. List of the general transcription factors and their subunits from *S. cerevisiae*. The prevalent functions of the respective factors in transcription initiation are indicated. Based on[42].

Factor	Subunit(s)	Function
TFIIA	Toa1, Toa2	stabilization of TBP/TFIID-DNA complex; suppression of negative regulatory factors
TFIIB	TFIIB	stabilization of initiation complex; determination of transcriptional directionality; recruitment of Pol II-TFIIF complex; DNA positioning; TSS selection; stimulation of initial RNA synthesis
TFIID	TBP, TAF1-14	initial promoter recognition; interaction with enhancer-bound transcriptional activators and nucleosomes; PIC-assembly platform
TFIIE	Tfa1, Tfa2	stabilization of initiation complex; TFIIH recruitment; stimulation of TFIIH ATPase and kinase activity; stabilization of transcription bubble
TFIIF	Tfg1, Tfg2, Tfg3 (TAF14)†	stabilization of initiation complex (particularly TFIIB); DNA positioning‡; TSS selection; stimulation of initial RNA synthesis
TFIIH	Ccl1, Kin28, Rad3, Ssl1, Ssl2, Tfb1-5	DNA unwinding during promoter opening; CTD-phosphorylation; stimulation of promoter escape

† Yeast-specific subunit; component of TFIID, TFIIF and chromatin remodelers; non-essential in TFIID and TFIIF[42].

‡ Evidence from structural analysis of initiation complexes in *H. sapiens*[48,85,86].

CTD, C-terminal domain; Pol, DNA-dependent RNA polymerase; TAF, TBP-associated factor; TBP, TATA-box binding protein; TSS, transcription start site.

While initiation complex disassembly involves the detachment of Pol II and TFIIF, it does not imply a complete disintegration of the initiation machinery. Several lines of evidence suggest that TFIIA, TFIID, TFIIE and TFIIH remain bound to DNA and co-activators at the promoter[138,139]. This scaffold complex[15] may facilitate re-initiation of Pol II after its recycling via gene looping and TFIIB-mediated promoter-terminator interactions[140-144], which in turn may be subject to regulation by the TFIIH kinase Kin28/CDK7[143].

2.4 The general transcription factor TFIIH

The multi-protein initiation factor TFIIH (Table 1.2) is highly conserved in eukaryotes, comprises ten subunits and exceeds a molecular mass of 0.5 MDa in yeast, thereby resembling Pol II in size and complexity[145]. TFIIH is the largest factor required for constitutive transcription from core-promoter sequences, since TFIID can be substituted by its subunit TBP if TATA-box motifs are present[146]. As a part of the basal initiation machinery TFIIH fulfills central roles in Pol II mediated transcription, including the unwinding of the DNA double helix to commence promoter melting and the phosphorylation of the Pol II CTD to facilitate promoter escape[145,147,148]. It is further involved in the coordination of transcriptional regulation, in transactivator signaling[149,150], and, possibly, in RNA synthesis by Pol I and Pol III[151-153]. In addition, TFIIH is an essential component of the nucleotide excision repair (NER) pathway that unwinds dsDNA around damaged sequences and exposes the single strands. TFIIH is recruited to lesions by DNA damage sensors like Rad4/XPC or stalled Pol II elongation complexes and in turn provides access for downstream NER factors that protect ssDNA strands, perform strand incisions and resynthesize the removed fragments[145,148,154]. A functional link between DNA transcription and DNA repair was first established when TFIIH was found to participate in both of these fundamental processes[155,156]. Individual subunits or subassemblies of TFIIH can be independently engaged in mechanisms outside of transcription or DNA repair, such as in control of the metazoan cell cycle[157,158] or in chromosome segregation[159], reflecting the dynamic and comprehensive nature of the complex. Owing to its pivotal character, mutations in TFIIH underlie severe clinical diseases like xeroderma pigmentosum (XP), trichothiodystrophy (TTD) or Cockayne syndrome (CS)[160-162] and multiple disease phenotypes can be linked to various of its functions[147,148].

TFIIH consists of two distinct subcomplexes, a three-subunit kinase module and a seven-subunit core (Table 1.2). The interaction between both modules is sensitive to high salt conditions and their stability in isolation has been validated *in vitro*[163-165], in accordance with the presence of holo-TFIIH, core-TFIIH and CAK (cyclin-dependent kinase (CDK)-activating kinase; metazoan terminology for the kinase module) complexes in the cell[148]. Whereas 10-subunit TFIIH is required for transcriptional processes[119,166], the kinase module impedes NER activity[167] and is displaced from the repair complex after initial bubble opening at DNA lesions, possibly by allosteric rearrangements upon incorporation of further NER factors[168]. The remaining core-TFIIH is sufficient for DNA bubble expansion during the subsequent NER stages[148].

Table 1.2 │ Composition of the TFIIH complex. Assignment of subunits to core and kinase modules of TFIIH. Yeast subunits, their molecular mass and their most prevalent functions are indicated, respectively. In addition, corresponding human homologues are listed. Based on[145].

TFIIH Module	Subunit in yeast (*S. cerevisiae*)	Mass (kDa)	Corresponding subunit in *H. sapiens*	Function
Kinase	Ccl1	45.2	CycH	regulation of kinase activity of Kin28/CDK7
	Kin28	35.2	CDK7	kinase (primary target: Pol II CTD)
	Tfb3	38.1	MAT1	structural function; stabilization of kinase module; interaction with XPD
Core-TFIIH	Rad3	89.8	XPD	5′ to 3′ ATP-dependent helicase; bridge between the kinase and core module
	Ssl1	52.3	p44	structural function; E3 ubiquitin ligase (yeast only); stimulation of helicase activity of XPD/Rad3
	Ssl2	95.3	XPB	3′ to 5′ ATP-dependent helicase; translocase
	Tfb1	72.9	p62	structural function; interaction with TFIIE and NER components
	Tfb2	58.5	p52	structural function; regulation of ATPase activity of XPB/Ssl2
	Tfb4	37.5	p34	structural function; strong interaction with Ssl1/p44
	Tfb5	8.2	p8/TTDA	regulation of ATPase activity of XPB/Ssl2

CDK, cyclin-dependent kinase; CTD, C-terminal domain; MAT1, ménage à trois 1 (protein); TTD, trichothiodystrophy; NER, nucleotide excision repair; XPB, xeroderma pigmentosum group B protein; XPD, xeroderma pigmentosum group D protein.

Three subunits of the holo-TFIIH complex possess intrinsic and conserved catalytic activity. The cyclin-dependent kinase Kin28/CDK7 (yeast, *S. cerevisiae / H. sapiens*) assembles with its regulatory cyclin Ccl1/CycH and the scaffolding factor Tfb3/MAT1 to form the TFIIH kinase module with essential roles in transcription and metazoan cell cycle control[120,163,169,170]. Core-TFIIH contains the subunits Rad3/XPD, Ssl1/p44, Ssl2/XPB, Tfb1/p62, Tfb2/p52, Tfb4/p34 and Tfb5/p8. The two largest components of the core, Rad3/XPD and Ssl2/XPB, exhibit ATPase activity and are categorized as members of the DNA helicase superfamily 2 (SF2)[171] but with opposite polarities[147,172,173]. This classification is, however, under debate. Rad3/XPD presumably acts as a conventional ATP-dependent helicase to unwind dsDNA[173-175], whereas Ssl2/XPB displays weaker helicase activity[172,173] and has been proposed to function rather as a dsDNA translocase[108,109]. While Ssl2/XPB suffices for promoter melting[103,105,176] and escape[135,136] during transcription initiation, both ATPases are actively involved in DNA opening during NER, although the exact mechanism of their cooperation remains controversial[105,145]. In addition to its catalytic role Rad3/XPD bridges between the core and kinase modules of TFIIH by interacting

with Ssl1/p44 and Tfb3/MAT1, respectively[173,177]. Associated Ssl1/p44, in turn, stimulates the activity of Rad3/XPD[178,179]. In yeast but not in metazoans, Ssl1 also appears to operate as an E3-ubiqitin ligase, albeit with unknown cellular targets[180]. Ssl1 and the remaining non-enzymatic core-TFIIH subunits comprise folds that frequently participate in protein-protein interactions, such as von Willebrand factor A (vWA)[179,181] or Really Important New Gene (RING)[182] domains and Zn-fingers. Secondary structure predictions furthermore reveal a remarkably large proportion of α-helical segments within the least characterized core-TFIIH subunits Tfb1/p62 and Tfb2/p52, which are highly indicative of a structural role for these proteins. The pleckstrin homology domain (PHD) of Tfb1/p62 has been implicated in binding the large subunit of TFIIE, Tfa1/TFIIEα[102], but likewise associates with distinct partners such as transcriptional activators[183-187] and NER factors[188,189]. The vWA-containing Ssl1/p44 and Tfb4/p34 proteins interact intimately and display strong structural homology, suggesting a common origin and evolution via gene duplication[190]. The smallest TFIIH subunit Tfb5/p8 and the C-terminus of Tfb2/p52 fold into a compact heterodimer that enhances Ssl2/XPB ATPase activity in NER[191,192]. In summary, the available data suggest that many core-TFIIH subunits primarily serve as scaffolding elements to provide a stable framework for the correct positioning and the stimulation of their catalytic counterparts Rad3/XPD and Ssl2/XPB within higher-order complexes in transcription and NER.

2.5 Structural studies of TFIIH and its components

Despite extensive efforts to determine the structure of TFIIH, its precise architecture remains elusive to date. The isolation and analysis of TFIIH are challenged by its complexity, poor stability and, in particular, biochemical heterogeneity. Only in the course of a century after its discovery as an essential *in vitro* transcription initiation factor[193], researchers succeeded in assigning the DNA-dependent ATPase activity of TFIIH to the putative DNA repair helicases Rad3/XPD and Ssl2/XPB[156,172,194,195], in identifying the remaining components of core-TFIIH[196-200] and in recognizing the association of core-TFIIH with a cell cycle kinase and its regulatory factors[120,163]. Whereas its subunit composition is now well-established, the preparation of TFIIH, which is still predominantly extracted from endogenous sources, remains laborious and ineffective[85,201]. Recent technical advances significantly increased the quality and yields of 10-subunit TFIIH from yeast[202] but further complications, such as latent sample heterogeneity, persist. Although alternative procedures involving recombinant production methods have been established in principle[203,204], the more robust and sophisticated classical protocols prevail for large-scale approaches.

Early analyses of endogenously prepared TFIIH by negative stain electron microscopy (EM) had revealed a rigid, ring-like shape with a protrusion that presumably encompassed the kinase module and provided indications for a modular arrangement of the complex[205,206]. Subsequent studies thus aimed to improve the resolution of the initial structures and to position TFIIH within the context of the transcription machinery[85-87,101,165,201]. Negative stain EM experiments with distinct yeast TFIIH subcomplexes, such as core-TFIIH, which was devoid of

one or both ATPases, yielded various reconstructions at resolutions of ~25-30 Å[165]. These confirmed the initially observed location of the kinase module and inferred the approximate positions of Rad3 and Ssl2 within TFIIH. The ATPases were situated adjacent to each other and to the kinase module while the scaffolding core-TFIIH subunits formed a semicircular connection between them. Subsequent, better resolved cryo-EM reconstructions of TFIIH-containing initiation complexes revealed an overall crescent-like shape of TFIIH and its position on downstream promoter DNA. Moreover, they indicated an interaction interface between TFIIH and the stalk of Pol II[85-87]. Inconsistencies between studies performed with PICs from yeast[201] and *H. sapiens*[85] were reconciled[87], resulting in a unified topological model for the transcription initiation machinery and the location of TFIIH. The recent investigation of a PIC-Mediator complex from yeast[101] confirmed these results at low resolution. Surprisingly, however, a weak additional density, which had been not been observed in previous PIC studies, was detected in proximity to the proposed position of Rad3 and assigned to the TFIIH kinase module based on crosslinking information. The findings suggest that the high mobility of this TFIIH subcomplex within the PIC context[85,86] is reduced in presence of Mediator[101]. While the described EM-analyses provided multiple novel insights, none of the derived reconstructions exceeded a resolution of ~15 Å in the TFIIH segments, possibly due to insufficient sample quality. In combination with biochemical data, the respective EM maps thus permitted an approximate placement of homology models or atomic structures of TFIIH subunits, which had been determined by other methods[86,87,101]. Except for the ATPases Rad3/XPD and Ssl2/XPB, however, these models could not be unambiguously fitted and may be misoriented[86]. Therefore, although an evidently conserved position of TFIIH within the PIC has been established, the architecture of TFIIH has been revealed merely on a topological level and essential details like intermolecular interaction surfaces remain uncharacterized.

Whereas high-resolution models of TFIIH have not yet been obtained, atomic structures of various subunits and domains, which have been derived by NMR or X-ray crystallography, are available. Structures of distinct archaeal homologues of the ATPases Rad3/XPD[174,175,207,208] and Ssl2/XPD[107,209] unanimously demonstrated that the enzymes are monomeric and comprise two ATPase domains, consistent with their allocation to the SF2 family of helicases. Rad3/XPD contains two insertions in ATPase lobe 1, a 4Fe4S cluster and the 'ARCH' domain, which have been implicated in its function as a NER helicase[174,175,207,208]. Ssl2/XPB resembles SWI/SNF-type ATPases[86,107] and includes non-canonical N- and C-terminal extensions that may be involved in the recognition of damaged DNA and in DNA unwinding, thus legitimating its observed weak helicase activity[107]. Analyses of specific TFIIH domains, including the RING fingers at the Tfb3/MAT1 N-terminus[210] and the Ssl1/p44 C-terminus[182] or the vWA folds of Ssl1/p44[179] and Tfb4/p34[181], revealed the expected canonical motifs, albeit with unique features. To date, the model of a heterodimer between Tfb5/p8 and the C-terminus of Tfb2/p52 represents the only structural characterization of a protein-protein interaction within TFIIH. Both domains adopt equal folds and associate with each other into an uncommon, compact U-shaped 'dimerization module' via β-strand addition[192]. Structures of the Tfb1/p62 PHD in complex with numerous

peptide substrates, originating from NER components[188,189], transcriptional activators[183,185-187] or the acidic C-terminus of Tfa1/TFIIEα[102], illustrated a variety of binding modes to this domain, which is frequently engaged in protein targeting. The *H. sapiens* TFIIH kinase CDK7[211] and the associated CycH[212] feature canonical folds with high resemblance to other cell cycle kinases and to the Mediator kinase-cyclin pair CDK8/CycC[213]. However, as most of the structures were derived from isolated fragments, they do not provide information on intermolecular arrangements or the overall assembly framework within core-TFIIH or the 10-subunit complex. Numerous biochemical, functional[105,174,181,182,203,214-216], and crosslinking[101,190,201] analyses, on the other hand, revealed an abundance of interactions within TFIIH, thereby permitting the approximate assignment of its subunit architecture. TFIIH may be conceptually divided into four distinct regions; a Ssl2/XPB-Tfb2/p52-Tfb5/p8 complex that resides on one lobe of the crescent, Rad3/XPD, which is located on the other lobe and anchors the kinase module via Tfb3/MAT1, and the Ssl1/p44-Tfb4/p34 dimer that constitutes the TFIIH 'backbone' and connects the lobes. Nevertheless, numerous cross-connections between the four modules were discovered, suggesting the presence of flexible domains and a partially intertwined organization of the complex, which specifically involves Tfb1/p62[101,190,201]. Moreover, the studies identified well-conserved sections with predicted secondary structure elements (SSEs) that were essential for TFIIH integrity within each subunit of the core module[216].

A plethora of functional and structural data on TFIIH has been gathered since its first description more than 25 years ago[193]. However, intriguing questions about its precise architecture and the mechanisms underlying its cooperation with the transcription and NER machineries remain, in particular as highly-resolved models of initiation complexes with TFIIH are still lacking and structural information on its role in NER has not yet been derived.

2.6 The Mediator co-activator complex

Constitutive transcription is primarily coordinated by auxiliary proteins that reside at core-promoter regions or in their direct vicinity, such as the GTFs. The expression of a large proportion of genes, however, is also subject to regulation by transcription factors that bind to distant UAS- or enhancer motifs. These DNA sequences are of particular relevance for activator-dependent transcription, which commences in response to developmental or environmental signals. Although enhancers are specific for their respective promoters, they may be located thousands of nucleotides from their target TSSs in metazoans. Functional links between these promoters and enhancers are established by regulatory multi-protein co-activators, which operate as signal transduction hubs[217]. Via the simultaneous integration of enhancer-associated transcription factors and the basal transcription machinery at the core-promoter, such co-activators provide "a drawbridge across the enhancer-promoter divide"[218] and mediate the crosstalk between the distinct transcriptional elements.

The Mediator complex is a prominent and well-conserved member of the co-activator family. It is involved in transcription of nearly all protein-coding genes[52,217] and thus considered

one of its global regulators[219]. An association of Mediator with Pol II and the initiation machinery (Fig. 1.3a) was first described more than two decades ago[220]. The molecular mechanisms by which Mediator acts on transcription, however, have been scarcely explored, owing to its sheer size and its conformational and compositional flexibility, which entail the substantial structural dynamicity of this complex. In yeast, Mediator comprises 1.4 MDa and 25 subunits, while metazoan Mediator complexes encompass additional subunits and may exist as cell-type specific variants that lack distinct components[37,221]. Early EM studies revealed a modular architecture of Mediator[222,223] and accounted for its division in four topological segments: the tail, the head and the middle module, which are connected via the Med14 scaffold, and a reversibly associated kinase[224]. Whereas nearly all head and middle module subunits are essential for cell viability in yeast, Mediator kinase and Mediator tail can be ablated completely. The complex of the head and middle modules is therefore also termed 'core Mediator' (cMed)[224,225]. Mediator tail interacts with activation domains of gene-specific transcription factors and is thus the major determinant of Mediator recruitment to enhancers[52,224]. However, since several activators, such as p53, are able to contact additional subunits in the head and middle modules[226], Mediator may undergo alternative modes of enhancer recruitment or cooperatively integrate signals from multiple transcription factors[227]. The head and middle modules interact intimately and are critical for the association of Mediator with Pol II and the GTFs. Moreover, they have been implicated in binding the unphosphorylated CTD of Pol II and thus in the presumed Mediator-dependent promoter recruitment mechanism for the initiation machinery[220,228]. Complex formation between Mediator and Pol II is impeded in presence of the kinase module since their interactions are mutually exclusive[229,230]. As the Mediator kinase antagonizes PIC assembly at Mediator-regulated promoters, it is generally considered as a transcriptional repressor[128,223], although it performs also divergent, activating functions. It targets very distinct phosphorylation substrates, including the Pol II CTD, transcriptional activators or histone tails[231,232], suggesting a versatile, yet poorly understood, role in transcriptional control.

EM analyses have captured multiple conformational states of Mediator[222,223,233], indicating a pronounced intrinsic flexibility of the complex. Consistent with the reported dynamic transitions and the various functionalities of Mediator, structural shifts upon its interaction with distinct binding partners have been observed[52]. A recent study demonstrated significant rearrangements in the middle module upon Mediator association with Pol II[234]. Similarly, the docking of transcriptional activators appears to induce conformational changes that propagate from their binding sites in the tail module to the Pol II interaction surface in the head- and middle modules[223,235] and, in turn, stimulate Pol II transcriptional activity[226]. However, the nature of the involved rearrangements, the mechanism of their transmission throughout Mediator architecture, and their ultimate effect on initiation complex conformation remain subject to further investigation. The synergy between Mediator and Pol II enhances the efficiency of GTF recruitment during PIC assembly at core-promoters, as well as the activity of the engaged factors[236,237]. Genetic screens identified an interaction between the TFIIH ATPase Rad3/XPD and the middle module subunit Med11, suggesting a putative model for Mediator-stimulated TFIIH

recruitment[238]. In contrast to its assistance in PIC formation and stabilization, Mediator also contributes to the destabilization of the initiation machinery after promoter opening. It stimulates the activity of the TFIIH kinase Kin28/CDK7, which in turn hyper-phosphorylates the CTD of Pol II and induces disruption of the Mediator-Pol II interface, thereby ultimately facilitating initiation complex disassembly and promoter clearance[19,133,239-242]. The underlying principles of these events, however, are debated. While Mediator is an essential co-factor for transcription initiation, its capacities also extend to further phases of the transcription cycle. It has been implicated in elongation and promoter proximal pausing[243] and supposedly promotes Pol II recycling and transcription re-initiation by stabilizing the promoter-retained scaffold complex[138]. In addition, Mediator assumes regulatory functions in a plethora of processes that range from genome organization or the reshaping of chromatin architecture[52], to DNA repair[244] and the integration of developmental cues via various signaling cascades[52,245].
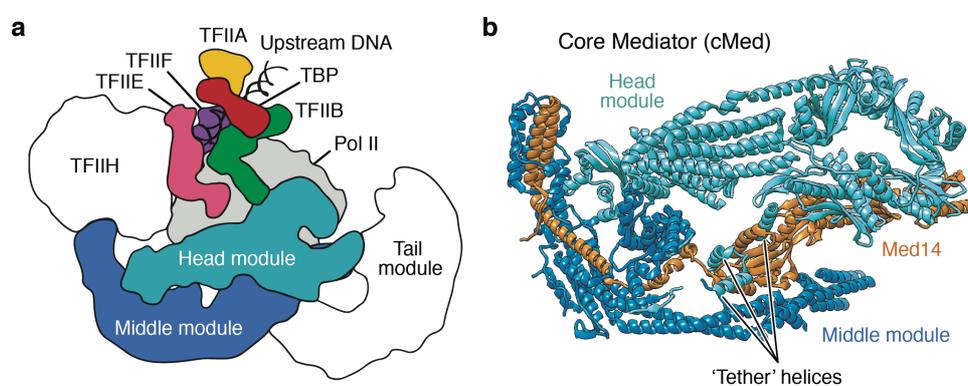
Figure 1.3 | Structure of core Mediator. a, Schematic depiction of the PIC-Med complex. Modules of cMed were positioned on the cPIC[89] based on data of an ITC-cMed complex[88]. The approximate position of the tail module was inferred from cryo-EM studies of free and PIC-bound endogenous Mediator[101,246] at lower resolution. The approximate position of TFIIH was inferred from cryo-EM studies of the PIC[86,87]. Components, for which high-resolution information is available, are colored. Color code as in **Fig. 1.1**. The Mediator head and middle modules are depicted in cyan and in blue, respectively. **b,** Ribbon model of cMed. The model is based on the atomic *Schizosaccharomyces pombe* cMed structure[247] (PDB code 5N9J). The scaffold subunit Med14, which is assigned to the middle module, is highlighted (orange) and the tether helices of the head module are indicated. Color code and view as in **a.** The view corresponds to the 'left' side view described previously[248]. Adapted from[247].

The modular and dynamic nature of Mediator became apparent already from initial low-resolution EM reconstructions of endogenous complexes[222,223]. These studies indicated that a complete model of Mediator might only be derived by a divide-and-conquer approach utilizing smaller subassemblies. Consequently, numerous structures of isolated segments with various complexity, originating from all four Mediator modules, have been obtained to date[249]. However, only recent developments in the biochemical preparation and structural analysis of Mediator permitted the elucidation of higher order arrangements within the holo-complex. The first major breakthrough was marked by the successful isolation of the recombinant Mediator head module, which constitutes the best-conserved and least flexible region of the complex, and the determination of its atomic structure by X-ray crystallography[248]. Subsequent advances in cryo-EM and

crosslinking techniques furthermore facilitated the generation of more accurate topological models for Mediator architecture[246,250-252] and, eventually, the positioning of the Mediator head module structure within the reconstruction of a minimal yeast ITC-cMed complex[88]. This model revealed a repositioning of the Pol II stalk upon Mediator binding and three interfaces between cMed, Pol II and TFIIB, consistent with the enhanced recruitment of TFIIB to Mediator-Pol II complexes[237]. A recent study additionally presented high-resolution cryo-EM reconstructions of Mediator from *Schizosaccharomyces pombe* without the kinase module in a free conformation and in association with Pol II, suggesting large rearrangements[234]. The to date best-resolved and most completely modeled structure, however, was derived by X-ray crystallography[247] (Fig. 1.3b). It encompasses a recombinant cMed complex from *S. pombe*, which includes all head and middle module subunits, as well as parts of the Med14 scaffold, and identified four main head-middle contact interfaces that are well-conserved. Med14 is strongly involved in the intricate interaction between both modules and three 'tether helices' of Mediator head subunits undergo a domain swap to bridge to the middle module. Similar structures for the more flexible Mediator tail and kinase modules have not been determined. However, a low-resolution reconstruction of a PIC-Mediator complex from yeast is available[101] and confirms the previously inferred location of the tail in free Mediator[246].

2.7   Structural perspectives on transcription initiation

The coordinated stepwise process of initiation and the option to trap or mimic several of the assumed intermediate states strongly facilitates the structural dissection of this phase of transcription. In order to reconstruct the stages of PIC assembly and to recapitulate conformational changes in the complex during promoter opening or initial RNA synthesis, multiple initiation states have been studied, frequently in combination with biochemical probing.

The era of the high-resolution analysis of protein-coding gene transcription began with the determination of X-ray structures of TBP and the components of the upstream promoter assembly. These studies examined the DNA binding modes of the GTFs TFIIA[65-67] and TFIIB[253,254] and reported the characteristic 90° bend which the DNA undergoes upon TBP binding[61,62]. Soon after, structures of the 10-[255] and the initiation-competent 12-subunit[256] Pol II enzymes from yeast revealed the ultimate "basis of transcription"[255]. The investigation of Pol II-GTF complexes, however, proved to be challenging. Following analysis of Pol II-TFIIB assemblies by X-ray crystallography and crosslinking-mapping experiments[74,75,78,257], the atomic model of an ITC-like Pol II-TFIIB-DNA complex eventually demonstrated the exact interaction of TFIIB with Pol II and identified the B-linker and B-reader elements, which contact and stabilize the template DNA strand in the Pol II active site[76]. Still, despite significant efforts, further co-crystal structures of initiation factors with Pol II could not be obtained. Instead, studies of the macromolecular transcription initiation machinery were promoted by the subsequent "resolution revolution"[258] of cryo-EM. Initial analyses of various *H. sapiens* initiation intermediates implied conformational changes between the closed and open PIC states, such as closure of the Pol II clamp, and

positioned the GTFs TFIIE and TFIIF, which flank the promoter DNA on both sides of the Pol II active center cleft[85]. The EM-reconstructions were generally consistent with previously obtained models for the upstream promoter assembly and with locations of TFIIE and the TFIIF dimerization domain that had been inferred by biochemical and crosslinking studies[81,83,84,95,97]. Similar experiments performed with initiation complexes from yeast agreed well with the *H. sapiens* model once initial discrepancies[201,259] were reconciled[87,88]. Recently, structures of cPICs from yeast and *H. sapiens* lacking TFIIH were independently derived at resolutions of 3.9 Å and 3.6 Å, respectively. They revealed further details about the DNA contacts and the peripheral domains of the included GTFs, as well as about their rearrangements upon promoter melting[86,89]. Owing to the modular composition of the GTFs, however, several flexible regions are missing from the models. Whereas most of these are non-essential, also the C-terminal half of TFIIEα/Tfa1, which presumably recruits TFIIH[94,100], has not been localized yet. Its stabilization may depend on the presence of TFIIH in the complex, consistent with the observed "context-dependent folding transitions"[47] of the other GTFs upon their incorporation into the PIC. A precise comparison of the available structures from the human and yeast systems indicates an overall striking similarity and suggests a well-conserved initiation complex architecture and a generally unified DNA opening mechanism, although minor, species-specific discrepancies may exist[47,48].

Several cryo-EM studies of initiation complexes exceeding the cPIC subunit composition have been reported but resulted in reconstructions of significantly lower quality and resolution. Analyses of complete PICs including TFIIH were performed with reconstituted complexes from yeast and *H. sapiens* and ultimately inferred an apparently conserved position of TFIIH on the cPIC and on downstream DNA[85-87,101]. However, as TFIIH is highly flexible, neither global reconstructions nor TFIIH-focused strategies revealed SSEs within TFIIH or details about its interactions with the cPIC. Consequently, previously obtained structures or homology models of TFIIH subunits were assigned with low orientational confidence[86], resulting in a merely topological model of the TFIIH segments in the PIC. In a different approach a minimal yeast cITC was reconstituted with cMed, and the complex architecture was derived at a resolution of 9.7 Å. The cryo-EM density permitted fitting of atomic structures of Pol II, the GTFs and the Mediator head module and implied three contact interfaces between the head module, Pol II and TFIIB, as well as a movement of the Pol II stalk upon Mediator binding. Due to the lack of a complete reference model, however, it did not implicate details of the Mediator middle module and its interaction with the head[88]. The to date largest *in vitro* reconstituted initiation complex included Mediator, Pol II and all GTFs except for the TAFs of TFIID, which are usually omitted from structural studies due to their poor biochemical stability. Its cryo-EM reconstruction comprised two unaccounted additional density regions, which were attributed to the tail of Mediator and the TFIIH kinase and established their position in the initiation machinery[101]. The tail module was located at the base of Mediator as suggested[246,252], and the TFIIH kinase module appeared to reside in close proximity to Mediator and the putative position of Rad3/XPD[101]. However, while this analysis provided an overview of the completely assembled initiation machinery and

confirmed previous findings of the cPIC and cITC-cMed structures[86,89], it revealed only the general complex topology owing to limited resolution.

In summary, while the core of the transcription initiation complex has been studied at atomic resolution in multiple states and in distinct organisms, similar information on the complete initiation machinery is still lacking, as the size and flexibility of components like TFIIH and Mediator pose major challenges to their structural characterization.

2.8   Aims and scope of this work

In eukaryotes, Pol II mediated transcription is a key process in the coordination of gene expression. Transcriptional regulation, and in particular events governing the stage of initiation, have thus been at the focus of research for decades. Although a plethora of biochemical and functional studies have investigated the mechanisms underlying initiation complex assembly, promoter opening and promoter escape, the comprehensive interpretation of the acquired data relies on their complementation with three-dimensional structural information. In the past, numerous detailed models of isolated components, subcomplexes and submodules of the initiation machinery have been obtained. In addition, the analysis of reconstituted PICs[86,87,89] and ITCs[88] was facilitated by recent developments in cryo-EM techniques[258], which permitted the structural dissection of respectively large macromolecular assemblies. Despite these efforts, however, high-resolution information is limited to the core elements of the initiation complex to date. Incorporation of the global co-activator Mediator[88,101] or the basal transcription factor TFIIH[86,87,101] into cPICs was attempted but their architecture could be assessed merely on a topological level. Owing to its pivotal roles in promoter opening and escape[145,147,148], in particular the structure of TFIIH will be relevant for the elucidation of the molecular basis of initiation and its analysis is thus of utmost priority. Detailed studies of TFIIH are impeded by its distinct flexibility but also by the lack of highly pure protein samples, since commonly derived endogenous TFIIH is of low abundance and frequently of heterogeneous quality. Therefore, to arrive at a well-resolved three-dimensional model for TFIIH or its complexes in transcription, a preceding optimization of the existing protocols for TFIIH preparation is required.

This work aimed to establish methods for the routine large-scale recombinant production of TFIIH from the yeast *S. cerevisiae* and to determine its subunit structure at high resolution. Extensive screening of expression constructs and diverse purification approaches ultimately resulted in strategies for the preparation of stable TFIIH core- and kinase modules, as well as of derivatives of core-TFIIH that lacked one or both of its ATPases, respectively. Complete 10-subunit TFIIH was reconstituted and isolated in a sucrose gradient. When crystallographic analyses of TFIIH subcomplexes were futile and 10-subunit TFIIH became available, the focus of this work was shifted. It subsequently aimed at the incorporation of purified TFIIH into an initiation – co-activator complex and at the extension of previously performed structural studies of cPIC[89] and cMed[247] to the cryo-EM analysis of the yeast PIC-cMed complex. This macromolecular assembly has a molecular mass of ~2 MDa and comprises 46 peptides that were,

with the exception of the Pol II subunits, produced recombinantly. During EM data processing, particle classes with diverse occupancy of cMed and TFIIH were detected and reconstructions of the yeast PIC and the PIC-cMed complex at resolutions of 4.7 Å and 5.8 Å, respectively, were obtained. The reconstruction of cryo-EM maps that revealed continuous secondary structure throughout the inherently flexible PIC and PIC-cMed complexes was facilitated by WarpCraft, a novel computational tool developed by D. Tegunov[260]. Modeling was performed based on available atomic structures and homology models of the cPIC and cMed, which were fitted and subsequently adapted and extended to the PIC-cMed complex. The resulting structure indicated details about the interfaces between Pol II, Mediator and TFIIB as well as rearrangements in the stalk region of Pol II and in the cMed middle module. The PIC and PIC-cMed structures furthermore revealed the exact subunit assignment and the modular architecture of TFIIH, as well as its interactions with Pol II, TFIIE, promoter DNA and cMed. To facilitate TFIIH model building, EDC-based crosslinking experiments were conducted, which in particular aided to the assignment of its interactions with TFIIE. In an elaborate homology and *ab initio* modeling approach, structures for most regions of core-TFIIH were derived and placed in the EM map together with published atomic models for single domains from yeast. Respective intra-/ and inter-subunit linkers were subsequently traced in the density. In addition, the TFIIH kinase module was located in proximity to cMed in the PIC-cMed reconstruction. The obtained structures explain a multitude of functional data on the recruitment of TFIIH, its interaction with the initiation machinery and its stimulation by Mediator, and provide further guidance on how TFIIH may engage in promoter opening, CTD phosphorylation and the facilitation of promoter escape. In conclusion, the presented work is a further advance towards the complete understanding of the intricate mechanisms that coordinate the successive stages of initiation and provides an accurate basis for their further biochemical and functional characterization.

3 Materials

3.1 Bacterial strains

Table 3.1 | Bacterial strains used in this study.

Species	Strain	Genotype	Supplier
E. coli	XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]*	Agilent
E. coli	BL21-Codon Plus(DE3)-RIL	*E. coli B F$^-$ ompT hsdS(r$_B^-$ m$_B^-$) dcm+ Tet^r E. coli gal λ (DE3) endA Hte [argU ileY leuW Cam^r]*	Agilent
E. coli	Rosetta B834(DE3)	*E. coli B F$^-$ ompT hsdS(r$_B^-$ m$_B^-$) dcm+ metB (DE3)*	Novagen
E. coli	DH10EMBacY	*F$^-$ mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK λ- rpsL nupG / bMON14272‡ yfp+/ pMON7124*	Geneva Biotech

‡ Modified version of bMON14272 baculovirus shuttle vector (bacmid) containing a YFP reporter gene and gene deletions (chiA and v-cath). For more detailed explanations refer to section 4.2.

3.2 Yeast strains

Table 3.2 | Yeast strains used in this study.

Species	Strain	Genotype	Supplier
S. cerevisiae	BJ5464 Rpb3-His$_6$-Bio	*MATa; ura3-52 trp1 leu2D1 his3D200 pep4::HIS3 prb1D1.6R can1 GAL; insertion of 6×His-biotin tag 5' of Rpb3 gene (selection marker URA3)*	M. Kashlev[260]

3.3 Eukaryotic cell lines

Table 3.3 | Eukaryotic (insect) cell lines used in this study.

Species	Cell line	Origin	Supplier
Spodoptera frugiperda	Sf21 *(IPLB-Sf21-AE)*	immortalized pupal ovarian tissue cells[261]	Thermo Fisher Scientific
Spodoptera frugiperda	Sf9	immortalized pupal ovarian tissue cells, clonal isolate of parental cell line *IPLB-Sf21-AE*[261]	Thermo Fisher Scientific
Trichoplusia ni	Hi5 (High Five) *(BTI-TN-5B1-4)*	immortalized pupal ovarian tissue cells[262]	Expression Systems

3.4   Plasmids and Oligonucleotides

Table 3.4 | Plasmids used in this study. A detailed list of precursor plasmids obtained during generation of the final Series-438-based constructs (438-A/C) can be provided upon request. Complete sequences of primers used for generation and validation of plasmids described in this study can be provided upon request.

Gene	Residue range	Affinity tag	Vector backbone	Resistance	Source
*Sc*TBP	1-240	6×His (C$_{term}$)	pOPINE	Amp	C. Plaschka[89]
*Sc*Toa1[†]	1-94, 210-286	untagged	pOPINE	Amp	C. Plaschka[89]
*Sc*Toa2[†]	1-122	6×His (C$_{term}$)			
*Sc*Tfg1[†]	1-735	10×His-8×Arg-SUMO-3C (N$_{term}$)	pAHS3C	Amp	C. Plaschka[89]
*Sc*Tfg2	1-400	untagged			
*Sc*TFIIB	1-345	6×His (C$_{term}$)	pOPINE	Amp	S. Sainsbury
*Sc*Tfa2	1-482	untagged	pET21	Amp	C. Plaschka[89]
*Sc*Tfa1	1-328	6×His (C$_{term}$)			
*Sc*Tfb1	1-642	6×His-SUMO (N$_{term}$)	pET-MCN	Kan	this study
*Sc*Tfb1	165-642	6×His-SUMO (N$_{term}$)	pET-MCN	Kan	this study
*Sc*Ssl1	1-461	6×His-SUMO (N$_{term}$)	pET-MCN	Amp	this study
Sc Tfb2	1-513	untagged	pET-MCN	Kan	this study
*Sc*Ssl1	1-461	6×His-SUMO (N$_{term}$)	pET-MCN (multi-ORF)	Amp	this study
*Sc*Tfb4	1-338	untagged			
*Sc*Tfb4	1-338	6×His-SUMO (N$_{term}$)	pET-MCN	Amp	this study
*Sc*Tfb2	1-513	6×His-SUMO (N$_{term}$)	pET-MCN (multi-ORF)	Amp	this study
*Sc*Tfb5	1-72	untagged			
*Sc*Ssl1	1-461	6×His-SUMO (N$_{term}$)	pET-MCN (multi-ORF)	Amp	this study
*Sc*Tfb2	1-513	6×His-SUMO (N$_{term}$)			
*Sc*Tfb4	1-338	6×His-SUMO (N$_{term}$)			
*Sc*Tfb5	1-72	6×His-SUMO (N$_{term}$)			
*Sc*Ssl1	92-461	6×His-SUMO (N$_{term}$)	pET-MCN (multi-ORF)	Amp	this study
*Sc*Tfb2	1-513	6×His-SUMO (N$_{term}$)			
*Sc*Tfb4	18-322	6×His-SUMO (N$_{term}$)			
*Sc*Tfb5	1-72	6×His-SUMO (N$_{term}$)			
*Sc*Tfb1	1-642	6×His-SUMO (N$_{term}$)	438-A (multi-ORF)	Amp Gen	this study
*Sc*Ssl1	1-461	6×His-SUMO (N$_{term}$)			
*Sc*Tfb4	1-338	6×His-MBP-10×Arg-TEV (N$_{term}$)			
*Sc*Tfb5	1-72	6×His-SUMO (N$_{term}$)			
*Sc*Tfb2	1-513	6×His-SUMO (N$_{term}$)			
*Sc*Rad3	1-778	untagged			

[†] The genes were codon optimized for protein expression in *E. coli*.

Amp, ampicillin; Kan, kanamycin; Gen, gentamycin; Strep, streptomycin; ORF, open reading frame; TEV, tobacco etch virus

Gene	Residue range	Affinity tag	Vector backbone	Resistance	Source
*Sc*Tfb1	1-642	untagged			
*Sc*Ssl1	1-461	MBP-3C (N_{term})			
*Sc*Tfb4	1-338	untagged	438-A (multi-ORF)	Amp Gen	this study
*Sc*Tfb5	1-72	untagged			
*Sc*Tfb2	1-513	untagged			
*Sc*Ssl2	1-778	6×His-MBP-10×Arg-TEV (N_{term})			
*Sc*Tfb1	1-642	6×His-SUMO (N_{term})			
*Sc*Ssl1	1-461	6×His-SUMO (N_{term})			
*Sc*Tfb4	1-338	6×His-MBP-10×Arg-TEV (N_{term})	438-A (multi-ORF)	Amp Gen	this study
*Sc*Tfb5	1-72	6×His-3C (N_{term})			
*Sc*Tfb2	1-513	6×His-3C (N_{term})			
*Sc*Rad3[+]	1-778	untagged			
*Sc*Ssl2[+]	1-843	untagged			
*Sc*Kin28	1-306	6×His-MBP-10×Arg-TEV (N_{term})	438-C (multi-ORF)	Amp Gen	this study
*Sc*Ccl1	1-393	6×His-SUMO (N_{term})			
*Sc*Tfb3	1-321	6×His-3C (N_{term})			
*Sc*Med1	1-566	10×His-8×Arg-SUMO-3C (N_{term})			
*Sc*Med4	1-284	untagged	pCDFDuet	Strep	C. Plaschka[89]
*Sc*Med9	1-149	untagged			
*Sc*Med17	1-687	untagged			
*Sc*Med22	1-121	untagged			
*Sc*Med11	1-115	untagged			
*Sc*Med6	1-295	untagged	pCOLADuet	Kan	C. Plaschka[88]
*Sc*Med8	1-223	untagged			
*Sc*Med20	1-210	untagged			
*Sc*Med18	1-307	untagged			
*Sc*Med31	1-127	untagged			
*Sc*Med10	1-157	untagged	pCDFDuet	Strep	C. Plaschka[89]
*Sc*Med7	1-222	untagged			
*Sc*Med21	1-140	untagged			
*Sc*Med19	1-220	untagged	pETDuet	Amp	C. Plaschka[88]
*Sc*Med14	1-745	10×His (N_{term})			

[+] The genes were codon optimized for protein expression in *S. frugiperda*.

Amp, ampicillin; Kan, kanamycin; Gen, gentamycin; Strep, streptomycin; ORF, open reading frame; TEV, tobacco etch virus

Table 3.5 | Oligonucleotides used for assembly of the PIC-cMed complex.

Type	Sequence (5'-3')	Length [nt]
template DNA	TGA CAC AGC GCA GTT GTG CTA TGA TAT TTT TAT GTA TGT ACA ACA CAC ATC GGA GGT GAA TCG AAC GTT CCA TAG CTA TTA TAT ACA CAG CGT GCT ACT GTT CTC G	106
non-template DNA	CGA GAA CAG TAG CAC GCT GTG TAT ATA ATA GCT ATG GAA CGT TCG ATT CAC CTC CGA TGT GTG TTG TAC ATA CAT AAA AAT ATC ATA GCA CAA CTG CGC TGT GTC A	106

nt: nucleotides

Table 3.6 | General design of primers used for ligation-independent-cloning into Series-438 vectors.
Ligation-independent-cloning (LIC) compatible overhangs ('tags') are indicated. Complete sequences of primers used for generation and validation of plasmids described in this study can be provided upon request.

Tag	Sequence (5'-3')	Vector backbone	ORF directionality	5' ATG in ORF required	3' stop codon in ORF required
vBac F	TAC TTC CAA TCC AAT CG — 5'-ORF	438-A	5'-3' (fw)	yes	-
v1 rv	TTA TCC ACT TCC AAT GTT ATT A — 3'-ORF	438-A 438-C	3'-5' (rv)	-	yes
v1 F	TAC TTC CAA TCC AAT GCA — 5'-ORF	438-C	5'-3' (fw)	no	-

F, forward; fw, forward; rv, reverse; ORF, open reading frame

3.5 Thermal cycler programs

Table 3.7 | Overview of thermal cycler programs.

PCR type	Program parameters	Application
standard PCR	98°C, 60 s; (98°C, 10 s; MT-4°C, 30 s; 72°C, 1-3 min) x30; 72°C, 10 min; 4°C, ∞	cloning, colony PCR
Round-The-Horn PCR	98°C, 60 s; (98°C, 30 s; MT-4°C, 30 s; 72°C, 8-10 min) x20; 10°C, 1 min; 4°C, ∞	cloning
QuikChange PCR	98°C, 30 s; (95°C, 20 s; MT-4°C, 30 s; 72°C, 8-10 min) x15; 72°C, 10 min; 4°C, ∞	cloning

PCR, polymerase chain reaction; MT, melting temperature of PCR primers (reference for setting: primer with lowest MT)

3.6 Antibodies

Table 3.8 | Antibodies used in this study.

Antibody	Specificity	Dilution	Source organism	Supplier / ID
α-Rpb3	Rpb3 subunit of Pol II	1 : 1000	mouse	Neoclone / WP012
α-Tyr1-P	phosphorylated Tyr1 residues of Rpb1-CTD of Pol II	1 : 7	rat	D. Eick / 3D12[263]
α-Ser2-P	phosphorylated Ser2 residues of Rpb1-CTD of Pol II	1 : 60	rat	D. Eick / 3E10[264]
α-Ser5-P	phosphorylated Ser5 residues of Rpb1-CTD of Pol II	1 : 60	rat	D. Eick / 3E8[264]
α-Ser7-P	phosphorylated Ser7 residues of Rpb1-CTD of Pol II	1 : 10	rat	D. Eick / 4E12[264]
α-GST-HRP	GST	1 : 5000	-	GE Healthcare / RPN1236
α-mouse-IgG-HRP	mouse IgG	1 : 3000	goat	Abcam / ab5870
α-rat-IgG-HRP	rat IgG	1 : 5000	goat	Sigma-Aldrich (A9037)

CTD, C-terminal domain; GST, glutathione S-transferase; IgG, immunoglobulin G; HRP, horseradish peroxidase; P, phospho

3.7 Growth media and additives

Table 3.9 | Growth media for *E. coli*, *S. cerevisiae* and insect cell culture.

Media	Application	Composition / Supplier
LB	*E. coli* culture / plates	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl (1.5% (w/v) agar for solid plates)
ZY	*E. coli* auto-induction culture	2% (w/v) peptone, 1.0% (w/v) yeast extract
YPD	*S. cerevisiae* culture	2% (w/v) peptone, 2% (w/v) glucose, 1.5% (w/v) yeast extract (1.8% (w/v) agar for solid plates)
Gibco® Sf-900™ III SFM	Sf9 / Sf21 culture (growth and maintenance of suspension and monolayer cultures; baculovirus production and propagation)	low-hydrolysate, serum-free, protein-free, animal origin-free insect cell culture medium / Thermo Fisher Scientific
ESF921™	Hi5 culture (growth and maintenance of suspension cultures; large scale protein expression)	serum-free, protein-free insect cell culture media, supplemented with L-glutamine and Kolliphor® P188 / Expression Systems

Table 3.10 | Additives for *E. coli*, *S. cerevisiae* and insect cell culture.

Additive	Application	Stock	Effective applied concentration
ampicillin	antibiotic (*E. coli* culture / plates)	100 mg/mL in ddH$_2$O	100 µg/mL
kanamycin	antibiotic (*E. coli* culture / plates)	50 mg/mL in ddH$_2$O	50 µg/mL
chloramphenicol	antibiotic (*E. coli* culture / plates)	30 mg/mL in 96% EtOH	30 µg/mL
gentamycin	antibiotic (*E. coli* culture / plates)	10 mg/mL in ddH$_2$O	10 µg/mL
streptomycin	antibiotic (*E. coli* culture / plates)	30 mg/mL in ddH$_2$O	30 µg/mL
tetracycline	antibiotic (*E. coli* culture / plates)	10 mg/mL in ddH$_2$O	10 µg/mL
IPTG	expression induction / blue-white selection)	1 M in ddH$_2$O	0.25 – 0.7 mM / 1 mM
X-Gal	blue-white selection (*E. coli* plates)	150 mg/mL in DMSO	150 µg/mL
20*x* NPS	auto-induction additive (*E. coli* culture)	0.5 M (NH4)$_2$SO$_4$, 1 M KH$_2$PO$_4$, 1 M Na$_2$HPO$_4$	25 mM (NH4)$_2$SO$_4$, 50 mM KH$_2$PO$_4$, 50 mM Na$_2$HPO$_4$
50*x* 5052	auto-induction additive (*E. coli* culture)	20% (w/v) glycerol, 2.5% (w/v) glucose, 10% (w/v) α-lactose monohydrate	0.4% (w/v) glycerol, 0.05% (w/v) glucose, 0.2% (w/v) α-lactose monohydrate
MgSO$_4$	auto-induction additive (*E. coli* culture)	0.5 M in ddH$_2$O	1 mM
X-tremeGENE™ 9	transfection agent (insect cell culture)	supplied in 80% ethanol	1.5 µL/mL

IPTG, isopropyl-β-D-1-thiogalactopyranoside; X-Gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; DMSO, dimethyl sulfoxide

3.8   Chemicals, kits and consumables

Table 3.11 | Overview of suppliers for chemicals, kits, reagents, consumables and technical equipment used in this study. Suppliers of specific or unique reagents, materials or technical equipment are indicated in the text flow.

Material type	Supplier(s)
chemicals	Merck, Roth, Sigma-Aldrich
enzymes, buffers, additives and reagents for molecular cloning	Fermentas, New England Biolabs (NEB), Promgea
kits for DNA preparation (QIAquick and QIAprep series)	QIAGEN
laboratory materials and consumables	Eppendorf, Greiner Bio-One, Merck Millipore, Rainin, Roth, Sarstedt, Thermo Fisher Scientific
technical equipment	Beckman Coulter, BINDER, Biometra, Bio-Rad, Eppendorf, General Electrics (GE), Merck Millipore, New Brunswick Scientific, Sartorius, Tecan, Thermo Fisher Scientific, VWR

3.9   Buffers and solutions

Table 3.12 | Standard buffers, dyes and solutions.

Buffer	Composition/Description (Supplier)	Application
4x SDS-PAGE loading dye	45% (v/v) glycerol, 280 mM Tris pH 6.8 at 20°C, 8% (w/v) SDS, 10% (v/v) β-mercaptoethanole, 0.03% (w/v) bromophenol blue	SDS-PAGE
gel electrophoresis running buffer	20x NuPAGE™ MES/MOPS SDS running buffer (Invitrogen)	SDS-PAGE
gel stain	InstantBlue (Expedeon)	Coomassie staining
tris-glycine transfer buffer	Trans-Blot® Turbo™ Transfer System buffer (Bio-Rad)	western blotting
PBS-T	137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$ pH 7.4, 1.76 mM KH$_2$PO$_4$ pH 7.4, 0.1% Tween-20	western blotting
PCR master mix	2x Phusion® High-Fidelity PCR Master Mix (NEB)	PCR
6x DNA loading dye	Gel Loading Dye, Purple (NEB)	agarose gel electrophoresis
10x TAE	50 mM EDTA pH 8.0 at 20°C, 2.5 M Tris-acetate	agarose gel electrophoresis
NEBuffer™ 3.1	50 mM Tris-HCl pH 7.9 at 25°C, 100 mM NaCl, 10 mM MgCl$_2$, 0.1 mg/mL BSA (NEB)	restriction endonuclease digest
CutSmart® buffer	20 mM Tris-acetate pH 7.9 at 25°C, 50 mM potassium acetate, 10 mM magnesium acetate, 0.1 mg/mL BSA (NEB)	restriction endonuclease digest
Fast AP™ buffer	10 mM Tris-Hcl pH 8.0 at 37°C, 5 mM MgCl$_2$, 100 mM KCl, 0.02% Triton X-100, 0.1 mg/mL BSA (Thermo Fisher Scientific)	restriction endonuclease digest

T4 DNA Ligase buffer	50 mM Tris-HCl pH 7.5 at 25°C, 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT (NEB)	ligation
T4 polymerase buffer	100 mM Tris-HCl pH 8.0 at 25°C, 500 mM NaCl, 100 mM MgCl$_2$, 10 mM DTT	LIC cloning
P1	50 mM Tris-HCl pH 8.0 at 25°C, 10 mM EDTA, 100 µg/mL RNase A (QIAGEN)	bacmid isolation
P2	200 mM NaOH, 1% SDS (QIAGEN)	bacmid isolation
N3	4.2 M Gu-HCl, 0.9 M potassium acetate pH 4.8 (QIAGEN)	bacmid isolation
DPBS	138 mM NaCl, 2.7 mM KCl, 8.1 mM Na$_2$HPO$_4$ pH 6.9, 1.47 mM KH$_2$PO$_4$ pH 6.9	insect cell culture
100x PI	0.028 mg/mL leupeptin, 0.137 mg/mL pepstatin A, 17 mg/mL PMSF, 33 mg/mL benzamidine (in 100% EtOH)	protease inhibitor mix
negative stain solution	2% (w/v) uranyl formate in dddH$_2$O	negative stain EM

AP, alkaline phosphatase; ATP, adenosine triphosphate; BSA, bovine serum albumin; DTT, dithiothreitol; EDTA, (ethylenedinitrilo)tetraacetic acid; EM, electron microscopy; EtOH, ethanol; LIC, ligation independent cloning; MES, 2-(N-morpholino)ethanesulfonic acid; MOPS, 3-(N-morpholino)propanesulfonic acid; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate buffered saline; PCR, polymerase chain reaction PMSF, phenylmethylsulfonyl fluoride; SDS, sodium dodecyl sulfate; TAE, tris-acetate-EDTA; Tris, tris(hydroxymethyl)aminomethane

Buffers utilized for protein purifications are referenced and described in the text flow. Buffers required for purification of Pol II, TFIIA, TFIIB, TBP, TFIIE, TFIIF and cMed were reported previously[76,88,89,260] and are not listed.

4   Methods

4.1   Standard cloning

Experimental procedures described in Section 4.1 were performed at 25°C unless indicated otherwise.

Cloning strategy

Owing to the intricate architecture of TFIIH, various co-expression strategies for the production of soluble protein were established. Vectors used for protein expression in *E. coli* were based on the pET-MCN (Multi-Cloning and expressioN) series[265,266]. After initial insertion of isolated open reading frames (ORFs) into pET-MCN vectors, this system permits a sequential fusion of ORFs from distinct plasmids with a 'cut-and-paste' strategy. For ORF-combination, the backbones of 'acceptor' pET-MCN vectors are opened by SpeI digest, whereas 'donor' vectors are treated with SpeI and XbaI, which results in excision of DNA fragments containing the previously inserted ORF and a ribosomal binding site (RBS). Ligation of such insert fragments with 'acceptor' vectors produces fusion constructs comprising the 'acceptor' backbone and genes from both 'donor' and 'acceptor' vectors (Figure 8.2). Combination of SpeI and XbaI sites by ligation is unidirectional and a fused SpeI/XbaI site will not be cleaved by either enzyme, thereby preserving a multi-ORF vector's capacity for further recombination. Repetition of the recombination procedure results in pET-MCN vectors that contain multiple ORFs within a long multi-cistronic coding region, which is preceded by a single inducible T7 promoter. Each gene retains a separate RBS. Stable pET-MCN vectors encoding ORFs of up to five genes and encompassing > 15,000 nucleotides (nt) were generated. In addition, co-expression strategies involving two pET-MCN vectors with orthogonal antibiotic resistances (kanamycin and ampicillin) but encompassing fewer ORFs were applied to enhance the variability of the system.

A related system was utilized for the production of vectors with multiple ORFs that could be inserted into a baculoviral genome and subsequently transfected into insect cells for heterologous protein expression. A more detailed explanation of the underlying ligation independent cloning (LIC) and assembly strategy is provided in section 4.2.

Standard Polymerase Chain Reaction

This method was used for the amplification of linear DNA fragments from various intra- and extra-genomic templates. The 5' and 3' ends of the generated Polymerase-Chain-Reaction (PCR) products were defined by specific PCR primers. These were designed for melting temperatures in the range of 58-65°C and optimally contained a 15-18 nt GC-rich region complimentary to the respective template gene. Overhangs of up to 30 nt length were optionally added to the 5' ends of PCR primers, for example if a template was extended by Round-The-Horn PCR or to enhance enzymatic activity during cleavage of the PCR products with restriction endonucleases.

The setup for a standard PCR included ~50 ng (plasmids/PCR products) or ~250 ng (genomic DNA) of DNA template, forward and reverse PCR primers at concentrations of 0.4 pmol/µL and 25 µL of the 2x Phusion® High-Fidelity PCR Master Mix (NEB). Reactions were adjusted to a volume of 50 µL with ddH$_2$O. PCR programs (Table 3.7) comprised 30-40 thermocycles and were executed in a TProfessional TRIO Thermocycler® (Biometra). Settings such as the primer annealing temperature and the time for primer extension were adapted to the PCR type and to the melting temperature of the utilized primers. Under standard conditions an average synthesis rate of 1,500 nt/min was assumed for Phusion® polymerase (NEB), and primer annealing temperatures were expected to be 4°C below their respective melting temperatures as determined with ApE (Section 4.4). PCR products were separated according to their length by agarose gel electrophoresis, excised from the gel and purified using the QIAquick PCR gel extraction kit according to the manufacturer's instructions.

Restriction endonuclease digest and dephosphorylation

The generation of linearized DNA inserts and vectors with complimentary overhangs for their combination involved cleavage with respective restriction endonuclease enzymes. Digest reactions comprised 5 µg of template DNA, 1.5 µL of restriction enzyme(s) (NEB), and 5 µL of the recommended 10x reaction buffer (NEB), and were adjusted to a volume of 50 µL with ddH$_2$O. Samples were incubated 4-5 h or overnight (O/N) at a temperature optimized for the used restriction enzyme(s), mostly 37°C or 25°C. Traditional ligation-dependent cloning methods additionally required the treatment of linearized vector backbones with FastAP™ (Thermo Scientific) alkaline phosphatase to remove the terminal phosphate groups. Digest reactions were supplemented with 20 µL of 10x FastAP™ buffer and 1 µL of enzyme (Thermo Scientific), adjusted to a reaction volume of 200 µL with ddH$_2$O and incubated at 37°C for 1.5 h. Linearized insert fragments were purified via agarose gel electrophoresis and gel extraction whereas linearized and optionally dephosphorylated vectors were purified using the QIAquick PCR Purification kit according to the manufacturer's instructions.

Agarose Gel Electrophoresis

Linear DNA fragments were separated according to their size by electrophoresis in 1% (w/v) agarose gels. For short fragments (< 500 nt) 1.5 % (w/v) agarose gels with a better resolution in the smaller size range were used. The respective amount of agarose was dissolved in 150 mL 1x TAE buffer by repeated boiling and the solution polymerized into a gel upon cooling. Addition of 1 µL/mL of the DNA intercalating chemical SYBR™Safe DNA Gel Stain (Invitrogen) into the liquid agarose solution permitted subsequent visualization of DNA samples under ultraviolet (UV) light. Polymerized gels were covered with 1x TAE buffer. Samples were prepared by mixing with an appropriate amount of 6x DNA Loading Dye (NEB) and loaded onto the gel together with a 1 kb DNA Ladder size standard (NEB). A voltage of 120 V was applied for 30-45 min until a sufficient separation of the fragments was observed. DNA bands were visualized with the GEL iX20 Imager system (Intas) for documentation.

Gel Extraction

Agarose gels were UV-illuminated with the BST-20G-D2E BlueLED BioTransilluminator (Biostep) to detect the position of specific DNA fragments within the gels. Correct DNA bands were identified by comparison with a size standard and excised and removed from the gel with a scalpel. Gel pieces were transferred into tubes to determine their weight. Extraction and purification of DNA from gel slices was performed using the QIAquick Gel Extraction kit according to the manufacturer's instructions and including optional steps.

Ligation

In a ligation reaction the formation of a covalent chemical bond between the terminal 5' phosphate and 3' hydroxyl groups of linear DNA fragments is catalyzed. This method was used to combine linearized dephosphorylated vector backbones with linearized ORF-containing DNA inserts or to regenerate functional plasmids from the linear products of a Round-The-Horn PCR. Ligase reactions were performed with 50 ng of vector DNA and, for traditional cloning methods, 5-7x molar excess of the DNA insert. Additionally, the reactions comprised 2 µL of 10x T4 DNA Ligase buffer (Thermo Fisher Scientific) and 5-7 units (U) of T4 DNA Ligase (Thermo Fisher Scientific). They were adjusted to a volume of 20 µL with ddH$_2$O and incubated for 1 h at 25°C. Subsequently, the complete reaction volume was used for transformation in chemically competent XL1-Blue cells without further purification.

Round-The-Horn PCR-based site-directed mutagenesis[†]

This method encompasses the PCR-based amplification of both strands of a template vector and a subsequent step for their re-ligation. The obtained linear PCR products are complimentary to each other and anneal to form a dsDNA fragment but the ligation of their 5' and 3' ends is required for the regeneration of a stable circular plasmid. The Round-The-Horn (RTH) setup permits the introduction of substantial modifications into a DNA template. Dependent on the sequence of the 5' overhangs of the PCR primers and the design of their complimentary regions, i.e. their annealing sites to the gene of interest, the original DNA sequence can be extended, subjected to site-directed multi-basepair mutagenesis or truncated. RTH primers require a 5' phosphate group as a chemical modification for the ligation reaction.

RTH-PCR reactions were performed with RTH-specific 5'-modified primers and an increased time for primer extension (Table 3.7) in order to permit amplification of the entire template vector sequence. PCR products were directly subjected to restriction endonuclease digest with DpnI (NEB) for 2 h at 37°C to degrade the original template, and purified by agarose gel electrophoresis and gel extraction. Purified DNA was ligated in a standard reaction with 70 ng input DNA and transformed in chemically competent XL1-Blue cells without further purification.

[†] Based on the OpenWetWare protocol by Sean Moore.

Transformation in chemically competent cells, preparation of vector DNA and sequencing

Frozen aliquots (100 µL) of chemically competent XL1-Blue-, BL21-Codon Plus(DE3)-RIL- or Rosetta B834(DE3) cells (Table 3.1) were thawed on ice and mixed with pre-cooled DNA samples comprising either 100 ng of purified plasmid, the complete volume of a ligation reaction or the complete volume of a LIC annealing reaction (Section 4.2). Cells were incubated with DNA for 25 min on ice, heat-shocked for 45 s at 42°C and recovered on ice for 2 min. Per reaction, 500 µL LB media (Table 3.9) were added and the transformed cells were incubated in a ThermoMixer® C (Eppendorf) for 1-1.5 h (37°C, 900 rpm). After short centrifugation (2 min, 2,300g), the media supernatant was removed except for 100 µL LB, in which the cells were resuspended. Cells were plated on LB agar plates (Table 3.9) supplemented with the required antibiotic(s) (Table 3.10) and incubated O/N at 37°C.

For the preparation of vector DNA, single cell colonies were picked from the plates and used to inoculate 25 mL of LB media supplemented with the respective antibiotic(s). Liquid cultures were incubated O/N (37°C, 160 rpm). Cells were harvested by centrifugation (3,000g, 15 min, 25°C) and plasmids were extracted from the cell pellets using the QIAprep Spin Minprep kit according to the manufacturer's instructions. Purified plasmids were eluted with ddH$_2$O.

DNA sequences of plasmids and inserts were verified by Sanger sequencing (Microsynth). Sequencing reactions contained 500 ng of DNA and the respective sequencing primers at a concentration of 2 pmol/µL, and were adjusted to a volume of 15 µL with ddH$_2$O.

4.2   Generation of baculovirus-compatible ORFs and insertion into baculovirus shuttle vectors

Insertion of ORFs into MacroBac Series-438 vectors by ligation-independent cloning

Ligation-independent cloning (LIC) provides an alternative strategy to introduce DNA fragments into a vector backbone. In contrast to traditional cloning methods a final ligation reaction is not required, owing to a specific design of the 5' and 3' ends of the LIC components. Instead, complimentary 15-18 nt ssDNA overhangs are generated and annealed, and remaining nicks or gaps in the DNA strands are repaired in *E.coli* cells after transformation. The recently described Series-438 MacroBac system[267] uses a LIC-based scheme for the assembly of multi-gene baculovirus-compatible expression vectors (Fig. 8.3).

Series-438 plasmids (modified pFastBac derivatives, Addgene #55218 and #55220) comprise a LIC-compatible site for the insertion of ORFs, which is exposed after cleavage with SspI. Vectors were digested with SspI (NEB) without subsequent phosphatase treatment and purified by agarose gel electrophoresis and gel extraction. To obtain and amplify ORF-containing insert fragments with matching 5' end 3' ends, standard PCRs with Series-438-derived, LIC-compatible primers[267] were performed (Tables 3.6 and 3.7). PCR products were subjected to agarose gel electrophoresis and gel extraction.

Complimentary ssDNA overhangs for annealing were generated from the blunt 5' and 3' DNA ends obtained after SspI digest or PCR by treatment with T4 DNA Polymerase (LIC-qualified, Novagen). The T4 enzyme possesses both 5' – 3' polymerase and 3' – 5' exonuclease activities. In the absence of dNTPs the exonuclease creates 5' ssDNA overhangs on a dsDNA template. Addition of specific dNTPs, however, restricts the 3' – 5' exonuclease processivity to the site of the first matching DNA base on the complimentary strand. At this site, further exonuclease cleavage is impeded and the added dNTP is incorporated by the DNA polymerase activity of the T4 enzyme. This mechanism permits formation of defined ssDNA overhangs on double-stranded vectors and inserts. In order to produce matching ssDNA overhangs, Series-438 vectors have to be treated with T4 polymerase in combination with dGTPs whereas PCR products require dCTPs (Fig. 8.3). The complimentary overhangs permit the annealing of vectors and inserts but prevent internal annealing events. The linearized vector and insert fragments obtained by SspI digest and PCR were treated with T4 polymerase in separate reactions containing dGTP (vector samples) or dCTP (insert samples), respectively. Reactions comprised 150 ng of linearized DNA (vector or insert), 2.5 mM of the respective dNTP type, 5 mM DTT, 2 µL of 10x T4 DNA Polymerase buffer (NEB) and 2 U of T4 polymerase (Novagen), and were adjusted to a volume of 20 µL with ddH$_2$O. After incubation for 60 min at 25°C, T4 polymerase was heat-inactivated (75°C, 20 min) and reactions were stored at -20°C. For annealing after T4 polymerase treatment, 2 µL of vector and 2 µL of insert DNA were mixed and incubated at 25°C for 30 min. Reactions were stopped by addition of 1.3 µL mM 25 mM EDTA and incubated for 10 min at 25°C. The complete reaction volume was directly transformed into XL1-Blue cells (Section 4.1).

Generation of poly-promoter MacroBac Series-438 vectors containing multiple ORFs

For co-expression of multiple genes, ORFs cloned into Series-438 vectors were assembled in a sequential manner using a Biobrick™-based system[267,268]. This approach ultimately yielded large poly-promoter fusion plasmids.

Combination of two Series-438 vectors requires SwaI digest of an 'acceptor' vector and PmeI digest of a 'donor' vector. Whereas SwaI cleavage simply opens the vector at a defined site, PmeI cleavage results in excision of a fragment containing the ORF and its adjacent regions, including the associated polyhedrin (polH) promoter and the SV40 polyadenylation (polyA) termination sequence (Fig. 8.3). After its isolation this fragment constitutes the new insert. The specific LIC-compatible design of the DNA sequences flanking the SwaI and PmeI restriction sites permits fusion of the 'acceptor' vector backbone with the insert in a standard LIC reaction.

'Acceptor' and 'donor' vectors were subjected to endonuclease cleavage with SwaI and PmeI (NEB), respectively, without subsequent phosphatase treatment. Both samples were purified by agarose gel electrophoresis and gel extraction to retrieve linearized vector and insert fragments. LIC-compatible ssDNA overhangs were generated by T4 polymerase treatment in reactions supplemented with dGTP for SwaI-treated samples (vectors) and with dCTP for PmeI-treated samples (inserts). Vectors and inserts were annealed, resulting in fusion plasmids with the backbone of SwaI-digested 'acceptor' vectors but with a complete set of ORFs from both the

respective 'acceptor' and 'donor' vectors. Each ORF was preceded by a polH promoter and followed by a SV40 polyA termination site, permitting independent transcription of each gene.

Repetition of this procedure with Series-438 vectors carrying two and more genes resulted in gradual accumulation of ORFs within one vector. This strategy ultimately permitted the generation of stable poly-promoter Series-438 vectors (438-A/C multi-ORF) with up to 12 inserted ORFs and lengths of $> 30,000$ nt.

Strategy for the introduction of ORFs into baculovirus shuttle vectors (bacmids)

ORFs for baculovirus-induced expression are not directly fused with baculovirus genomes *in vitro*. Instead they are co-transformed in *E. coli* cells, which feature the respective viral genome on a bacmid vector, and then transferred by gene transposition. The DH10EMBacY *E. coli* strain used in this study thus contains two extra-genomic pieces of DNA, which permit a regulated ORF transposition. The first is the transposition helper plasmid pMON7124[269] that carries, among other elements, a tetracycline resistance marker and the bacterial Tn7 transposon. This transposon comprises a series of genes (TnsA-E), which are required for transposition of DNA sequences flanked by Tn7 recognition sites (Tn7L and Tn7R) into a specific target element, the Tn7 transposon attachment site (attTn7)[270]. The second piece of DNA is a modified version of the baculovirus shuttle vector bMON14272[271]. This bacmid contains a kanamycin resistance marker, a LacZα gene with an inserted mini Tn7 transposon attachment site (mini-attTn7), all genomic regions of the *Autographa californica nuclear polyhedrosis virus* (*AcNPV*, a species of baculoviridae) that are important for viral DNA stability and propagation in host cells, and all elements that are required for replication of the bacmid in bacteria. The bMON14272[‡] bacmid in DH10EMBacY cells was altered by disruption of two viral genes encoding a chitinase (chiA) and a cysteinase (v-cath), which improves the levels of protein production in infected host cells, and by introduction of an EYFP reporter gene under control of the polH promoter. Consequently, the EYFP production of host cells may be used to monitor their infection status[272,273].

Series-438 MacroBac vectors[267] contain Tn7L and Tn7R sites that flank a region, which encompasses the respective ORF, its promoter and terminator elements and a gentamycin resistance marker, thereby generating a transposable element. If a Series-438 vector is successfully transformed into DH10EMBacY cells, the transposase (encoded on pMON7124) mediates transposition of the ORF and its associated elements into the mini-attTn7 site on the bacmid, thereby disrupting the LacZα gene by a large insertion. This mechanism permits the identification of positive colonies by simultaneous selection for the gentamycin resistance and blue-white screening on plates containing both the antibiotic and the metabolites IPTG and X-gal[274].

Preparation of electrocompetent DH10EMBacY cells

Cells from a commercially obtained glycerol stock of DH10EMBacY cells (Table 3.1) were plated on a LB agar plate (Table 3.9) supplemented with ampicillin, kanamycin and tetracycline (Table 3.10) and incubated O/N at 37°C to select cells with the correct bMON14272[‡]/

pMON7124 genotype. A pre-culture supplemented with the respective antibiotics was inoculated with a single colony from the plate and grown O/N (37°C, 150 rpm). The pre-culture was used to inoculate 250 mL of pre-warmed LB media supplemented with the respective antibiotics in a 1:250 (v/v) ratio. The main culture was grown (37°C, 160 rpm) to an optical density at 600 nm (OD_{600}) of 0.6. Cells were transferred into 50 mL Falcon tubes, incubated 20 min at 4°C, harvested by centrifugation (10 min, 3,000g) and resuspended in 1 mL sterile pre-cooled ddH_2O per Falcon tube. Cell pellets in each Falcon tube were washed once with 30 mL sterile pre-cooled ddH_2O, once with 1 mL sterile pre-cooled 10% (v/v) glycerol, and finally resuspended in 1 mL of sterile pre-cooled 10% (v/v) glycerol. The suspension was aliquoted (50 µL) in pre-cooled Eppendorf tubes, flash-cooled in liquid N_2 and stored at -80°C.

Transformation of electrocompetent DH10EMBacY cells

Frozen aliquots (50 µL) of electrocompetent DH10EMBacY cells (Table 3.1) were thawed on ice and incubated with 0.5-1 µL of purified, pre-cooled vector DNA (Section 4.1) for 15 min. DNA/cell suspensions were transferred to pre-cooled BIORAD Gene Pulser®/Micropulser™ electroporation cuvettes (0.1 cm gap) (BioRad). For electroporation, one pulse (25 µF, 1.8 kV) was applied. Transformed cells were immediately recovered in 1 mL LB media (Table 3.9) and incubated in 13 mL tubes (Sarstedt) at 37°C (4-5 h, 150 rpm). Small volumes of recovered cells (10-20 µL) were diluted to 100 µL with LB media and plated on LB agar plates (Table 3.9) supplemented with gentamycin, X-Gal and IPTG (Table 3.10) for blue-white selection. Plates were incubated until a clear distinction of white and blue colonies became possible (37°C, 1.5-2 days). Several single white colonies were picked and restreaked on fresh LB plates with gentamycin, X-Gal and IPTG. If restreaked cell colonies remained white after one day of incubation at 37°C, the respective clones were considered to be positive for an insertion into the mini-attTn7 site in the bMON14272[‡] bacmid and were used for the preparation of bacmid DNA.

Isolation of bacmid DNA by alkaline lysis and isopropanol precipitation

Cells from white colonies, which were positive for DNA transposition into bMON14272[‡], were used to inoculate 5 mL of LB media (Table 3.9) supplemented with gentamycin (Table 3.10). Cultures were incubated O/N (37°C, 180 rpm). Cells were harvested (5 min, 3,000g) and cell pellets were resuspended in 250 µL buffer P1[†]. Suspensions were mixed with 250 µL alkaline buffer P2[†] for cell lysis and subsequently with 350 µL acidic buffer N3[†] for neutralization. Debris and precipitated proteins were pelleted by centrifugation (10 min, 21,000g) and the lysate supernatant of each sample was transferred to a fresh Eppendorf tube. The centrifugation and transfer steps were repeated and the remaining supernatant was mixed with 700 µL of isopropanol. DNA was precipitated by incubation at -20°C for 8-12 h and pelleted by centrifugation (4°C, 30 min, 21,000g). The supernatant was removed completely and the DNA pellets were washed with 500 µL of 70% (v/v) ethanol at -20°C. Ethanol was removed after centrifugation (4°C, 10 min, 21,000g) and DNA pellets were covered with ~30 µL of 70% (v/v) ethanol. Samples with precipitated bacmid DNA were stored at -20°C until used for transfection

into insect cells (Section 4.3). To minimize the probability of failure during one of the subsequent insect-cell-dependent steps, bacmids were prepared in duplicates. Respectively, the isolation procedure was performed with cells from two distinct white colonies that had originated from the same transformation in DH10EMBacY cells.

[†] Buffers P1, P2 and N3 were supplemented with the QIAprep Spin Minprep kit (Qiagen).

4.3   Insect cell culture

Insect cell culture maintenance and monitoring

Insect cell lines (Table 3.3) were maintained in suspension, without light exposure, at constant temperature and agitation rates (27°C, 60 rpm), and in the appropriate type of medium (Table 3.9) unless indicated otherwise. Cells were incubated in pre-sterilized glass flasks of various sizes (250-3,000 mL volume) at target cell densities of $0.5-0.7\times10^6$ cells/mL. The filling volume of the flasks was limited to 1/10 and 1/5 of the flask size for Sf9/Sf21- and Hi5-cultures, respectively. Regular semi-weekly transfer of cell cultures into fresh flasks was required to counteract a gradual decrease in cell viability. Actions involving the opening of sterile containers such as cell culture flasks or bottles with media and viruses and the transfer of their contents were performed under a sterile laminar flow hood (Biowizard Golden Line, Kojair) after disinfection with 70% (v/v) ethanol and with sterile disposable materials unless indicated otherwise.

Average values for cell density, cell viability and cell diameter were assessed with a CASY® Modell TT Cell Counter and Analyzer System equipped with a 150 mM capillary (OMNI Life Science) according to the manufacturers instructions and using 50 µL of insect cell culture per sample. Fluorescence measurements were performed with an Infinite® M1000 Pro micro-plate reader (Tecan) using excitation (25 pulses, 400 Hz) and emission wavelengths of 514 and 527 nm, respectively. Samples were prepared from 1 mL of insect cell culture. For measurement, cells were pelleted by centrifugation (5 min, 250g), resuspended in 100 µL or 200 µL DPBS (Table 3.12) for Sf9/Sf21- or Hi5-cultures, respectively, and transferred into 96-well plates (Greiner).

Transfection of Sf9 cells with bacmid DNA and production of V_0 viruses

Tubes containing precipitated bacmid DNA were transferred under a sterile hood (Biowizard Golden Line, Kojair) and opened. Ethanol was removed and pellets were dried for 20 min. To re-dissolve DNA, pellets were incubated with ~20 µL ddH$_2$O for 20 min without further resuspension. In the meantime, a mix of 10 µL of X-tremeGENE™ 9 transfection agent (Table 3.10) and 100 µL of Sf-900™ III SFM medium (Table 3.9) was prepared. Dissolved bacmid DNA was mixed with 200 µL of Sf-900™ III SFM medium and with 100 µL of the transfection agent-containing solution by tube inversion. The resulting samples ('bacmid-TAs') were incubated for 1 h with occasional agitation.

A culture of Sf9 cells with a density of 1.0×10^6 cells/mL was prepared and 3 mL of cells were added to five out of the six wells in sterile 6-well plates (Greiner). The remaining well of the plates was filled with 3 mL of sterile Sf-900™ III SFM medium. Plates were kept at 27°C until incubation of the bacmid-TAs was completed. Each bacmid-TA sample was split in half and 150 µL of the mix were added dropwise to two wells with Sf9 cells, respectively. Four Sf9-containing wells of each 6-well plate were supplemented with bacmid-TA mix. One well with Sf-900™ III SFM medium and one well with Sf9 cells but without bacmid-TA mix served as controls. As bacmids were generally prepared in duplicates (Section 4.2), each 6-well plate ultimately comprised viruses originating from one generated cloning construct.

Plates were incubated without agitation and light exposure (27°C, 60-80 h). Beginning 48 h after transfection, Sf9 cells were analyzed for expression of EYFP with a fluorescence microscope (Leica) in 12 h intervals. Since DH10EMBacY-derived bacmids contain the EYFP gene under control of a viral promoter, appearance of fluorescent insect cells indicated successful transfection of bacmid DNA and production and release of viral particles by infected cells. If more than ten yellow-fluorescent Sf9 cells were observed in a well, the virus-containing media was removed with a pipet, transferred in sterile Falcon tubes and stored at 4°C without light exposure. These initial V_0 viruses were harvested at the latest 72-80 h after transfection.

Virus propagation and production of V_1 viruses

V_0 viruses were used to infect larger numbers of cells in order to produce more virus particles (V_1) and enhance the strength of the initial viruses but not for the purpose of protein expression. Sf9 cultures containing 25 mL of cells with a density of 1.0×10^6 cells/mL were transferred in 500 mL flasks and supplemented with sufficient amounts of respective V_0 supernatants for infection. Pipetted volumes varied between 0.3 mL and 5 mL and were dependent on the number of fluorescent cells observed during V_0 production. Infected cell cultures were incubated at standard conditions for up to 144 h. Changes in density, viability, diameter and EYFP production of the cells were monitored in 24 h intervals. Cultures with cell densities exceeding the standard value were adjusted to 1.0×10^6 cells/mL by dilution, if necessary. Culture volumes were limited to 50 mL due to flask size. Under ideal conditions, Sf9 cells approximately doubled in numbers within 48 h after infection and reached a stationary phase thereafter. The first time point without further cell division was referred to as 'day of proliferation arrest' (DPA). Cultures were incubated and monitored for additional 48-96 h after DPA. Fluorescence values and cell diameter of the cultures gradually increased in this timespan, whereas cell viability remained constant for 24-48 h before dropping quickly. V_1 viruses were harvested once cell viabilities of 80-85% were reached. Cells were transferred to sterile Falcon tubes and pelleted by centrifugation (20 min, 250g). V_1-containing supernatants were decanted into new sterile Falcon tubes and stored at 4°C without light exposure.

In non-ideal cases, overly potent viruses inhibited cell division too strongly, resulting in a <1.5x increase in total cell numbers after 48 h, or ineffective viruses failed to infect cells and neither proliferation arrest nor an increase in cell diameter was observed. To generate V_1 viruses

of consistently high quality, the initial titers of such V_0 viruses were adapted or fresh V_0 viruses were prepared before experiments for V_1 production were repeated. Sf9 cells were occasionally replaced by Sf21 cells, which had an increased doubling rate and required shorter time spans for V_1 production.

4.4   Protein biochemistry and molecular biology

Strategies for protein expression, purification and storage

S. cerevisiae proteins were either produced by recombinant over-expression in *E. coli* and insect cells or prepared endogenously from *S. cerevisiae* strains carrying specific genomic tags. TFIIA, TBP, TFIIE, TFIIF and cMed subunits were expressed essentially as described[89] in *E. coli* BL21-Codon Plus(DE3)-RIL or Rosetta B834(DE3) strains (Table 3.1) using an IPTG-inducible system. The original protocol for TFIIB expression in BL21-Codon Plus(DE3)-RIL *E. coli* cells[76] was slightly altered by substitution of LB media with ZY auto-induction media supplemented with NPS, 5052 and $MgSO_4$ (Tables 3.9 and 3.10) and consequent omission of IPTG addition[275] but otherwise remained unchanged. TFIIH core and kinase modules were produced in *Trichoplusia ni* High Five (Hi5) cells (Table 3.3) after baculovirus infection. Further information about the expression of TFIIH subunits and subcomplexes is provided below.

Part of this study was the establishment of novel strategies for purification and assembly of 10-subunit TFIIH and its subcomplexes. Respective methods are described in detail in section 4.5 and generally involved affinity chromatography, ion exchange chromatography and size exclusion chromatography (SEC) steps. The cMed complex and the remaining transcription factors used for reconstitution of the PIC-cMed complex (TFIIA, TFIIB, TBP, TFIIE, TFIIF) were purified according to previous reports[76,88,89] with minor alterations. Pol II with a His_6-Biotin-tag on Rpb3 was prepared from the *S. cerevisiae* strain BJ5464 (Table 3.2) as described[260]. Purified proteins and protein complexes were aliquoted, flash-cooled in liquid nitrogen and stored at -80°C.

Protein expression in *E. coli* cells

TFIIH subunits and subcomplexes were initially expressed in BL21-Codon Plus(DE3)-RIL cells (Table 3.1) according to a standardized protocol. *E. coli* cells were (co-)transformed (Section 4.1) with the respective expression plasmids (Table 3.4). Single colonies from the transformation plates were used to inoculate 150 mL of LB media (Table 3.9) supplemented with the respective antibiotics (Table 3.10). The pre-cultures were incubated O/N (37°C, 160 rpm) and added in a 1:100 (v/v) ratio to 12 L of LB media supplemented with the respective antibiotics. Cells were grown (37°C, 150 rpm) to an OD_{600} of 0.5 and shifted to 18°C with continued agitation. Protein expression was induced with 0.5-0.7 mM IPTG at an OD_{600} of 0.8 and carried out for 18-20 h at 18°C. Cells were harvested by centrifugation (7,900g, 4°C, 8 min) and supernatants were

discarded. Pellets were resuspended in the respective lysis buffers (20 mL lysis buffer / 1 L cell culture), flash-cooled in liquid nitrogen and stored at -80°C.

Protein expression in Hi5 insect cells

TFIIH subcomplexes comprising catalytic subunits were expressed in Hi5 insect cells (Table 3.3) since this cell line very efficiently generated crucial post-translational modifications (PTMs) like phosphorylations (analyzed by S. Vos, unpublished data). Moreover, Hi5 cells were the only tested expression system that permitted production of Kin28, Rad3 and Ssl2 in sufficient amounts and with enzymatic activity. Protein expression was induced by infection of Hi5 cells with viruses of the V_1 stage (Section 4.3) in small- and large-scale cultures. Cell cultures with volumes of either 50 mL (small scale) or 300 mL (large scale) were prepared at densities of 1.0×10^6 cells/mL in 500 mL or 3 L flasks. Infectious V_1 supernatant was added in dilutions of 1:1500 – 1:2000 (v/v). Applied virus titers depended on the strength and lifetime of the utilized V_1 viruses and were experimentally determined and adapted for each respective virus. After infection, Hi5 cultures were incubated at standard conditions (27°C, 60 rpm). Changes in density, viability, diameter and EYFP production of the cells were monitored in 24 h intervals. Cell densities surpassing the standard value were re-adjusted to 1.0×10^6 cells/mL by dilution with fresh media. If cultures exceeded the filling limit of the used flasks (100 or 600 mL, respectively) after dilution, surplus cells were transferred into appropriately sized fresh flasks and further incubated. Ideally, Hi5 cells divided once within 24 h after infection and then entered a stationary phase. As gene expression was controlled by the late viral polH promoter, proteins were mainly produced in the post-DPA phase. Therefore, cultures were incubated and monitored for an additional 48-72 h after DPA for optimal expression results. A gradual accumulation of protein (unpublished data) was accompanied by an increase in fluorescence levels and cell diameter of the infected cultures. Cell viability remained unchanged for ~24 h and then quickly dropped. Hi5 expression cultures ('V_2 cultures') were harvested at cell viability values of 85-88%. Harvesting was performed under non-sterile conditions. Cells were collected by centrifugation (238g, 45 min, 4°C) and supernatants were discarded. Pellets were resuspended in respective lysis buffers (17 mL lysis buffer / 100 mL cell culture), flash-cooled in liquid nitrogen and stored at -80°C.

Protein concentration and quantification

Concentration of protein samples was performed to reduce the sample volume before SEC runs and before final storage at -80°C. Depending on the characteristics of the concentrated protein or complex, either AMICON® Ultra Centrifugal Filter Units (Millipore) or Vivaspin® Centrifugal Concentrators (GE Healthcare) with distinct sample volumes (0.5-15 mL) were used. Molecular weight cutoffs were generally chosen to be two-fold lower than the predicted molecular mass of the applied sample. Protein concentration was stopped if the target volume for injection onto SEC columns or the solubility limit of a sample was reached.

Molar concentration and DNA content of protein samples were assessed by measurement of their absorbance at wavelengths of 280 nm (UV_{280}) and 260 nm (UV_{260}) with a NanoDrop-

2000 spectrophotometer (ThermoScientific). UV_{260}/UV_{280} ratios of 0.5-0.7 indicated DNA-free samples or tolerable low levels of DNA contamination, whereas ratios > 1 were observed only for DNA-bound complexes. Sample concentration c was calculated from weighted extinction coefficients E (i.e. absorbance of a 0.1% solution) of each analyzed protein or complex as determined by Expasy ProtParam[276] and the measured UV_{280} absorbance value A (setting: *1 Abs = 1 mg/mL*) according to a simplified version of Beer's Law ($c = A / E$).

Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Single components of protein samples were separated according to their molecular mass by SDS-PAGE analysis. Samples were mixed with an appropriate amount of 4x SDS-PAGE loading dye (Table 3.12), incubated for 1-10 min at 95°C and loaded onto pre-cast gradient NuPAGE 4-12% Bis-Tris Protein Gels (Invitrogen) together with the PageRuler™ Prestained Protein Ladder molecular weight standard (Thermo Fisher Scientific). SDS-PAGE gels were run for 60-90 min at 130-150 V using either 1x MOPS or 1x MES running buffer (Invitrogen) (Table 3.12). To fixate and visualize resolved peptide bands, gels were stained with InstantBlue™ (Expedeon) for 2 h. After destaining with ddH$_2$O, gels were scanned for documentation with an Epson Perfection V800 flatbed scanner (Epson). If gels were used for western blot experiments, they were directly transferred into ddH$_2$O without prior staining.

Western blot analysis and immunostaining

Targeted analysis of protein samples by western blot and antibody staining was performed to quickly detect specific protein subunits, domains or modifications, such as phosphorylations. Proteins were transferred from gels onto polyvinylidene difluoride (PVDF) membranes using the Trans-Blot® Turbo™ Blotting System (Bio-Rad). Unstained SDS-PAGE gels were washed in ddH$_2$O for 1-2 min to remove residual SDS. Gels were stacked with ready-to-use pre-assembled Trans-Blot® Turbo™ Mini PVDF Transfer Packs (Bio-Rad) into blotting sandwiches according to the manufacturer's instructions and placed in the blotting chamber. The pre-set transfer program for high molecular weight (MW) proteins (10 min, 25 V, 2.5 A) was usually selected since the molecular mass of several analyzed proteins exceeded 100 kDa. After transfer completion the membranes were blocked with 2.5% (w/v) milk powder solution in 1x PBS-T (Table 3.12) (2 h, 25°C, 60 r.p.m.). Subsequently the blocking solution was exchanged with 2.5% (w/v) milk powder / 1x PBS-T solution containing respective primary antibodies in appropriate dilutions (Table 3.8) and membranes were incubated (2-3 h, 25°C, 60 r.p.m. or O/N, 4°C 60 r.p.m.). To remove residual antibodies, membranes were washed (5x, 5 min) with fresh 1x PBS-T. If primary antibodies were HRP-coupled (Table 3.8), membranes were analyzed directly. If incubation with HRP-coupled secondary antibodies was required, membranes were incubated with 2.5% (w/v) milk powder / PBS-T solution containing secondary antibodies in appropriate dilutions (Table 3.8) (2-3 h, 25°C, 60 r.p.m. or O/N, 4°C 60 r.p.m.) and the washing procedure with 1x PBS-T was repeated. HRP-coupled antibodies were detected and visualized on the membrane using the Pierce™ Enhanced Chemi-Luminescence Western Blotting Substrate kit

(Thermo Fisher Scientific). Membranes were imaged with the Advanced Fluorescent Imager (Intas) for documentation.

Protein identification by mass spectrometry

The presence and identity of specific proteins within a purified sample was confirmed by mass spectrometry. Experiments and analysis were performed by M. Raabe and A. Kühn as part of the facility service provided by the group of H. Urlaub (Max Planck Institute for Biophysical Chemistry, Göttingen). Sample components were separated on polyacrylamide gels by SDS-PAGE and stained as described. Bands corresponding to proteins or peptides of interest were excised, digested in-gel with trypsin and extracted according to standard protocols[277]. Extracted peptides were enriched by SEC and analyzed on LTQ-Orbitrap Velos or Q-Exactive mass spectrometers (Thermo Fisher Scientific). Data sets were analyzed with Mascot[278] against the NCBI non-redundant protein database for *S. cerevisiae*. Upon request, samples were initially enriched for phosphorylated peptides and phosphorylation sites were mapped during analysis.

Pulldowns and binary interaction assays

The quality of newly generated and validated expression constructs was assessed with small-scale bead-based pulldown assays. This method permitted fast screening of numerous samples and simultaneously provided information about protein expression levels, protein solubility and protein or complex stability during initial affinity purification steps. Pulldowns were performed with Ni-NTA Agarose resin (QIAGEN) or Amylose Resin (NEB). Owing to its lower capacity, 700 µL of amylose resin were required for each pulldown, whereas 400 µL of Ni-NTA agarose resin were sufficient. The described protocol was applied for pulldowns with constructs expressed both in *E. coli* or insect cells. Frozen cell pellets from small-scale *E. coli* or insect cell expression cultures (1 L or 100 mL, respectively) were thawed at 25°C, supplemented with catalytic quantities of DNaseI (Sigma-Aldrich) and lysed with an EmulsiFlex-C5 cell disruptor (Avestin) (3-5 passages, 83,000 kPa). Cell lysates were cleared by centrifugation (79,000g, 60 min). The soluble protein-containing fractions were added to the respective amounts of Ni-NTA agarose or amylose resin, which had been pre-equilibrated by successive washes with ddH$_2$O and lysis buffer. Bead-containing suspensions were incubated for 1 h at 4°C with agitation. Beads were pelleted (4°C, 30g, 3 min), transferred into empty Micro Bio-Spin™ chromatography columns (Bio-Rad) and washed on-column with 15 mL of lysis buffer in consecutive steps, using 1 mL of buffer per wash. The supernatant and wash fractions were discarded. Bound protein was eluted twice using 250 µL of lysis buffer supplemented with 500 mM imidazole or 50 mM maltose for Ni-NTA agarose or amylose resin, respectively.

To test binary protein-protein interactions or the stability of multimeric assemblies under pulldown conditions, specific co-expression constructs with tags at only one or two of the target proteins were generated. Samples containing co-expressed protein complexes were prepared and subjected to bead-based pulldown assays as described and elution fractions were analyzed for the presence of untagged proteins that had been co-eluted as part of the complex.

Preparation of nucleic acid scaffolds

Oligonucleotides used for the assembly of PIC-cMed complexes for EM analysis or for activity assays (Table 3.5) were ordered separately (IDT) and dissolved in UltraPure™ DNAse/RNAse-Free Water (Thermo Fisher Scientific) to a concentration of 200 µM. To yield a final scaffold concentration if 100 µM, template and non-template strands were mixed in equimolar amounts (usually 15+15 µL) and annealed in a TProfessional TRIO Thermocycler® (Biometra). The annealing program comprised a phase of initial heating to 95°C followed by stepwise cooling of the sample in 1°C / 30 s increments to a temperature of 10°C. Annealed scaffolds were stored at -20°C and could be repeatedly re-thawed.

Sucrose gradient centrifugation

Centrifugation in a sucrose (or glycerol) gradient is a gentle approach for the separation of fragile or sensitive protein assemblies according to their molecular mass. This method was used to prepare the PIC-cMed complex and further PIC samples for analytical experiments and for EM.

Sucrose gradients were generated from a 'sucrose light solution', which usually comprised 15% (v/v) sucrose, and a 'sucrose heavy solution', which comprised 30% or 40% (v/v) sucrose. Both sucrose solutions additionally contained all respective buffer components. For standard gradients, 4 mL Thinwall Ultra-Clear™ centrifugation tubes (Beckman Coulter) were utilized. Tubes were filled with sucrose light solution slightly above their middle mark (2.2-2.4 mL). Subsequently sucrose heavy solution was added from the bottom of the tubes such that the ultimate volume ratio of both solutions was 1:1. Tubes were sealed and gradients were prepared by mixing with a BioComp Gradient Master 108 (BioComp Instruments) according to the pre-set tilt and rotation parameters of a 15-30% or a 15-40% sucrose gradient program (BioComp Instruments). Approximately 10-15 min after gradient formation, samples were applied to the top of the tubes, i.e. to the least dense part of the gradient. Sample volumes of up to 300 µL could be loaded but required careful removal of the topmost 150-200 µL of the gradient solution. Gradient centrifugation was performed at 175,000g for 16 h at 4°C (SW60 Ti rotor, Beckman Coulter). Protein complexes sedimented in the gradient according to their molecular mass without perturbation by their hydrodynamic (Stokes) radii. After centrifugation, 200 µL fractions were collected from the top of the gradient and fractions were analyzed by SDS-PAGE. All steps were performed at 4°C.

For experiments that required higher stability of the prepared sample, such as EM, a specific type of gradient centrifugation, 'gradient fixation' ('GraFix')[279], was applied. GraFix combines the standard sedimentation method with simultaneous chemical sample fixation by crosslinking and thus requires supplementation of the high-density gradient solution with a crosslinking agent. For standard experiments, 0.13% (v/v) glutaraldehyde crosslinker were added to the sucrose heavy solution. No further alterations to the gradient centrifugation protocol were introduced. Following centrifugation, however, the collected 200 µL fractions were quenched with a mix of 10 mM aspartate and 30 mM lysine (10 min) to saturate excess crosslinker. In order to remove sucrose, crosslinker and quencher from the protein samples, the respective fractions

were dialyzed for 10 h in Slide-A-Lyzer MINI Dialysis Devices (2 ml, MWCO 20,000) (Thermo Fisher Scientific).

Crystallization screening

Initial protein crystallization screening was performed by J. Wawrzinek and T. Schulz as part of the facility service provided by the group of P. Cramer (Max Planck Institute for Biophysical Chemistry, Göttingen). Purified proteins were concentrated to 10-30 µM, centrifuged (4°C, 21,000*g*, 10 min) to remove aggregates, transferred into fresh tubes and supplied to the facility on ice. Crystallization screens were performed at 20°C with drops of 200 nL volume (100 nL protein solution, 100 nL reservoir buffer solution). Screens were set up using a Crystal Gryphon robot (Art Robbins Instruments) on 96-well INTELLI® (Art Robbins Instruments) or MRC (Molecular Dimensions) sitting-drop crystallization plates. The facility offered a variety of commonly used commercial screens (Hampton Research, QIAGEN, Molecular Dimensions) and several customized screens that were produced in-house. Facility services included storage and regular imaging of crystallization plates in a Rock Imager 1000 (Formulatrix) within a 30-day period.

Basic bioinformatics tools

Gene and protein sequences were obtained from the gene database of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/gene), the Saccharomyces Genome Database (SGD; http://www.yeastgenome.org) and the UniProt Database[280]. Analysis of cloning data, such as the evaluation of sequencing results or the generation of maps for newly assembled expression plasmids, was performed with 'A plasmid Editor' (ApE) (http://biologylabs.utah.edu/jorgensen/wayned/ape). Homologues of conserved *S. cerevisiae* proteins were identified and confirmed in various species using the NCBI BLAST suite[281]. Multiple sequence alignments (MSAs) were carried out with Clustal Omega[282] and MUSCLE[283]. Secondary structure elements (SSEs) of proteins were predicted with Quick2D[†], PSIPRED[284], J-Pred[285] and Phyre2[286]. In order to visualize the level of conservation for specific SSEs during homology modeling[ø], PDB structures of confirmed homologues were mapped onto MSAs with ALINE[287] and ESPript 3.0[288]. Furthermore, the HHpred program suite[†289] was used to predict SSEs, folds and functions of specific poorly defined or conserved domains.

[†] Quick2D and HHpred are part of the MPI Bioinformatics Toolkit[290].

[ø] For a detailed report about the homology modeling strategies that were applied to model structures within TFIIE and TFIIH refer to Section 4.5 and Tables 8.2 and 8.3.

BLAST: Basic Local Alignment Search Tool; PDB: Protein Data Base (http://www.rcsb.org)

4.5   Project-specific techniques and experimental setups

This section focuses on customized methods and experimental schemes that were developed or specifically adapted for this study. These include particular strategies for protein expression, purification and assembly, for collection of EM data, for data processing and for model building, as well as specific functional assays. Several protocols described here have been published in:

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H. & Cramer, P. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551,** 204-209 (2017).

A detailed list of published text and items is provided in 'Publications' on page V. In addition, published protocols are marked with an asterisk (*) within this section. Protocol captions may deviate from the publication for clarity. Contributions of co-authors of the publication are indicated in the text flow.

The design of expression constructs for subunits of the *S. cerevisiae* cPIC and cMed complexes, as well as methods for their preparation and assembly into subcomplexes cMed, Pol II, TFIIA, TFIIE and TFIIF have been reported[76,88,89,260] and are not discussed.

Generation of poly-promoter expression constructs for core-TFIIH and the TFIIH kinase module*

Full-length subunits of *S. cerevisiae* TFIIH with the exception of Rad3 and Ssl2 were amplified from purified genomic DNA by PCR and transferred into modified pFastBac vectors[†] (derivatives of 438-A and 438-C; Addgene 55218 and 55220) by ligation independent cloning (LIC). The intron in Kin28 was removed by quick-change mutagenesis PCR after initial vector assembly. DNA sequences encoding full-length Rad3 and Ssl2 were obtained as *Spodoptera frugiperda* codon-optimized constructs from GeneArt (ThermoFisher Scientific), amplified from the vectors by PCR, and transferred into modified pFastBac vectors[†] by LIC. Within the vectors of the 438-series, the TFIIH subunits contain either N-terminal 6×His- or 6×His-MBP-tags or remain untagged. N-terminal 6×His-tags are followed by cleavage sites for either Ulp1 or the rhinovirus protease (3C) whereas the N-terminal 6×His-MBP-tags are followed by a modified cleavage site for tobacco etch virus (TEV) protease. After separate transfer of each gene into a 438-vector, the single vectors were combined by successive rounds of LIC to generate a 7-subunit construct encoding the genes for core-TFIIH (Rad3, Ssl1, Ssl2, Tfb1, Tfb2, Tfb4 and Tfb5) and a 3-subunit construct encoding the genes for the TFIIH kinase module (Ccl1, Kin28 and Tfb3). Each subunit is preceded by a PolH promoter and followed by a SV40 termination site. Within these constructs, the 6×His-MBP-tags are placed on Tfb4 and Kin28. Plasmid sequences are available upon request. Preparation of bacmids, production of insect cell virus of the V_0 and V_1 stage and protein expression in insect cells were performed essentially as described[291ø].

[†] 'Modified pFastBac vectors' correspond to Series-438 vectors described in Section 4.2.

[ø] A detailed protocol is also provided in Sections 4.3 and 4.4.

Generation of poly-promoter expression constructs for further TFIIH subcomplexes

Combination of single-gene Series-438 vectors by successive rounds of LIC (Section 4.2, Fig. 8.3) was also performed to generate 6-subunit constructs encoding full-length genes for two 'reduced' core-TFIIH variants (Ssl1, Tfb1, Tfb2, Tfb4, Tfb5 and Rad3 or Ssl2), which lacked one of the ATPases, respectively (Table 3.4). In contrast to the core-TFIIH expression plasmid, however, the Rad3 and Ssl2 sequences in 'reduced' core-TFIIH vectors correspond to conventional genomic sequences of *S. cerevisiae*. Codon-optimized sequences for Rad3 and Ssl2 were introduced to increase expression yields only for core-TFIIH. As described in the previous paragraph, each subunit in the resulting poly-promoter, multi-ORF plasmids is preceded by a PolH promoter and followed by a SV40 termination site to minimize effects of the respective gene order on protein expression levels. Within the 'reduced' core-TFIIH vectors, the primarily used (6×His-)MBP-tags were ultimately placed on Ssl1 and Tfb4, whereas intermediates and various test constructs were also tagged on Rad3, Tfb1, Tfb2 and Ssl2 according to different strategies. Although various additional test vectors with up to 12 distinct genes were produced using the LIC-based assembly system, these were ultimately not of relevance for this study.

Multi-gene vectors of the pET-MCN series were prepared similarly to Series-438 vectors, with the following exceptions. After amplification of genes from purified genomic DNA by standard PCR they were inserted into pet-MCN plasmids by conventional restriction endonuclease cleavage and ligation methods (Section 4.1). Genes containing restriction sites targeted by SpeI and XbaI were modified by RTH PCR-based mutagenesis. Within the pET-MCN vectors, the TFIIH subunits contain either N-terminal 6×His-tags followed by Ulp1 cleavage sites ('SUMO') or remain untagged (Table 3.4). After separate transfer of each gene into pET-MCN plasmids, up to five distinct genes of minimal core-TFIIH (Ssl1, Tfb1, Tfb2, Tfb4 and Tfb5) were assembled into multi-cistronic constructs by a recombination strategy based on restriction digest and ligation as described (Section 4.1, Fig. 8.2). In addition, pET-MCN vectors with orthogonal selective markers (kanamycin and ampicillin) were co-transformed in *E. coli* cells, further increasing the number and variability of proteins obtained by co-expression. While various multi-gene vectors with distinctly tagged and truncated versions of minimal core-TFIIH subunits were produced, only selected constructs are discussed in this study in detail (Table 3.4).

Plasmid sequences and a complete list of generated expression vectors are available upon request. Preparation of bacmids and production of insect cell viruses of the V_0 and V_1 stage were performed as described in Section 4.3. Protein expression in *E. coli* and Hi5 insect cells was performed as described in Section 4.4.

Expression and purification of *S. cerevisiae* minimal core-TFIIH

Recombinant *S. cerevisiae* minimal core-TFIIH was prepared in several variants, comprising either full-length subunits or distinctly truncated versions of Ssl1, Tfb1 and Tfb4 in diverse combinations. The applied purification scheme, however, remained essentially unchanged.

The complexes were expressed in *E. coli* according to the standard protocol (Section 4.4) and subjected to affinity, ion exchange and size exclusion chromatography steps. All purification

procedures were performed at 4°C unless stated otherwise. Resuspended (700 mM NaCl, 25 mM Na-HEPES pH 7.5, 10% (v/v) glycerol, 20 mM imidazole, 2 mM TCEP, 1*x* PI), frozen *E. coli* pellets were thawed at 25°C and supplemented with catalytic amounts of DNaseI (Sigma-Aldrich). Cells were lysed with an EmulsiFlex-C5 cell disruptor (Avestin) (5 passages, 83,000 kPa) and the cell lysate was cleared by centrifugation (79,000g, 1 h). The supernatant was filtered through 0.5 µM syringe filters (Merck Millipore) and applied to two coupled GE HisTrap HP (5 mL) columns (GE Healthcare) with a total bed volume of 10 ml, pre-equilibrated in buffer N-700 (700 mM NaCl, 25 mM Na-HEPES pH 7.5, 10% (v/v) glycerol, 20 mM imidazole, 5 mM β-mercaptoethanole). The columns were washed with 12 column volumes (CV) of buffer N-700 and the protein was eluted with a linear gradient of 0-100% buffer NE-500 (500 mM NaCl, 25 mM Na-HEPES pH 7.5, 10% (v/v) glycerol, 500 mM imidazole, 5 mM β-mercaptoethanole) in 8 CV. Protein-containing fractions were pooled, supplemented with 1 mg 6×His-Ulp1 protease (lab stock, prepared by S. Schilbach, unpublished data) and dialyzed for 12 h against buffer AE-100 (100 mM NaCl, 25 mM Na-HEPES pH 7.5, 5% (v/v) glycerol, 5 mM β-mercaptoethanole) at 4°C. Dialyzed sample was filtered through a 0.5 µM syringe filter (Merck Millipore) and subjected to anion exchange chromatography on a GE Mono Q 5/50 GL column (GE Healthcare) pre-equilibrated in buffer AE-100. After washing the column with 5 CV buffer AE-100, the protein was eluted with a step to 2% buffer AE-2000 (2 M NaCl, 25 mM Na-HEPES pH 7.5, 5% (v/v) glycerol, 5 mM β-mercaptoethanole) followed by a linear gradient from 2–9% buffer AE-2000 in 105 CV. Fractions containing stoichiometric minimal core-TFIIH were pooled, concentrated using an AMICON® Ultra, MWCO 10 kDa (Amicon) centrifugal device and applied to a GE Superdex200 Increase 10/300 GL SEC column (GE Healthcare) pre-equilibrated in gel filtration buffer (100 mM NaCl, 25 mM Na-HEPES pH 7.5, 2 mM TCEP). To remove subunits present in excess, peak fractions were pooled conservatively and then concentrated to ~10 mg/mL using an AMICON® Ultra, MWCO 10 kDa (Amicon) centrifugal device. In addition to increasing sample purity, the SEC step was used to alter buffer conditions by variation of the salt and buffer components in the gel filtration buffer or by its supplementation with 2 µM ZnCl$_2$. Purified and buffer-exchanged minimal core-TFIIH variants were directly subjected to crystallization trials (Section 4.4) at concentrations of 5-10 mg/mL (10-30 µM). Typical yields were in the range of 0.4-0.5 mg per liter of *E. coli* culture.

Expression and purification of *S. cerevisiae* core-TFIIH*

The complex was expressed in Hi5 insect cells after baculovirus infection (Section 4.4)[†]. Cells were collected by centrifugation (238g, 45 min, 4°C) and resuspended in lysis buffer (400 mM potassium acetate, 25 mM HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole, 0.284 µg ml^{-1} leupeptin, 1.37 µg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF, 0.33 mg ml^{-1} benzamidine). The cell suspension was flash cooled in liquid nitrogen and stored at -80°C. Recombinant *S. cerevisiae* core-TFIIH was purified by consecutive steps of affinity chromatography, ion exchange chromatography and size exclusion chromatography. All purification procedures were performed at 4°C unless stated otherwise. Frozen insect cell pellets were thawed at 25°C,

supplemented with catalytic amounts of DNaseI and lysed with an EmulsiFlex-C5 cell disruptor (Avestin) (3 passages, 83,000 kPa). The cell lysate was cleared by centrifugation (79,000g; 60 min) and the protein-containing soluble fraction was filtered through 0.8 μM syringe filters (Merck Millipore). The supernatant was then applied to a GE XK 16-20 column (GE Healthcare) containing a bed volume of 25 ml amylose resin (New England Biolabs) and pre-equilibrated in buffer M-300 (300 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole). After application of core-TFIIH-containing lysate supernatant, the column was washed with 3 column volumes (CV) of buffer M-300 and the protein was eluted with 2 CV ME buffer (350 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 10% glycerol (v/v), 50 mM maltose and 5 mM β-mercaptoethanole) onto a GE HiTrap Heparin HP (5 ml) column pre-equilibrated in buffer M-350 (350 mM KOAc, 25 mM K-HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole). The column was washed with 3 CV of buffer M-350 and the protein was eluted with a linear gradient of 0-30% buffer M-2000 (2 M potassium acetate, 25 mM K-HEPES, pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole) in 20 CV. Peak fractions were pooled, supplemented with 1 mg 6×His-TEV protease, 0.5 mg 6×His-3C protease and 0.5 mg 6×His-Ulp1 protease and kept at 4°C for 6 h. The cleaved sample was subjected to anion exchange chromatography using a GE HiTrap Q HP (1 ml) column pre-equilibrated in buffer A-400 (400 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole). After sample application the column was washed with 10 CV buffer A-400 and the protein was eluted with a linear gradient from 0-30% buffer A-2000 (2 M potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole) in 80 CV. Fractions containing stoichiometric 7-subunit core-TFIIH were pooled, concentrated using a Vivaspin 6 MWCO 50 000 (GE Healthcare) centrifugal device and applied to a GE Superose12 10/300 GL size exclusion column pre-equilibrated in gel filtration buffer (600 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 2 mM TCEP). Peak fractions were pooled, concentrated to 4 mg ml^{-1} using a Vivaspin 500 MWCO 50,000 (GE Healthcare) centrifugal device, aliquoted, flash-cooled in liquid nitrogen and stored at -80 °C. Typical yields were in the range of 0.3-0.4 mg per liter of insect cell culture.

[†] This sentence was added to the published protocol for clarity of the text flow.

Expression and purification of *S. cerevisiae* 'reduced' core-TFIIH

Recombinant *S. cerevisiae* core-TFIIH was prepared in three variants, comprising either seven or six full-length subunits. The 6-subunit 'reduced' core-TFIIH versions lacked either Rad3 or Ssl2 and were differently tagged (Table 3.4). However, as both constructs contained at least one MBP-tag derivative, the expression and purification scheme applied for core-TFIIH remained essentially unchanged for their preparation. As an exception, core-TFIIH$^{\Delta Rad3}$ was stable in low salt conditions (300 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 2 mM TCEP) during the final SEC step.

Expression and purification of the *S. cerevisiae* TFIIH kinase module*

Recombinant TFIIH kinase module was expressed in Hi5 insect cells after baculovirus infection (Section 4.4) and subjected to consecutive affinity, ion exchange and size exclusion chromatography steps. All purification procedures were performed at 4°C unless stated otherwise. Resuspended (400 mM potassium acetate, 25 mM HEPES pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanole, 0.284 µg ml^{-1} leupeptin, 1.37 µg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF, 0.33 mg ml^{-1} benzamidine), frozen insect cell pellets were thawed at 25°C, supplemented with catalytic amounts of DNaseI (Sigma-Aldrich) and lysed with an EmulsiFlex-C5 cell disruptor (Avestin) (3 passages, 83,000 kPa). The cell lysate was cleared by centrifugation (79,000*g*, 60 min) and the supernatant was filtered through 0.8 µM syringe filters (Merck Millipore)[†]. After cell lysis and lysate clearance, the sample was loaded onto a GE XK 16-20 column containing a bed volume of 25 ml amylose resin pre-equilibrated in buffer M-200 (200 mM potassium acetate, 25 mM K-HEPES pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole). The column was washed with 3 CV buffer M-200 and the protein was eluted with 2 CV buffer ME (200 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 50 mM maltose and 5 mM β-mercaptoethanole). Peak fractions were pooled, supplemented with 1 mg 6×His-TEV protease and 0.5 mg 6×His-3C protease and kept at 4°C for 6 h. The cleaved protein sample was subjected to anion exchange chromatography using a GE HiTrap Q HP (1 ml) column pre-equilibrated in buffer M-200. After sample application the column was washed with 10 CV of buffer M-200 and the protein was eluted with a linear gradient from 0-30% buffer A-2000 (2 M potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 5 mM β-mercaptoethanole) in 80 CV. Fractions containing stoichiometric kinase trimer were pooled, concentrated using a Vivaspin 6 MWCO 10,000 (GE Healthcare) centrifugal device and applied to a GE Superdex200 10/300 GL size exclusion column pre-equilibrated in gel filtration buffer (150 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 5% glycerol (v/v), 2 mM TCEP). Peak fractions were concentrated to 7 mg ml^{-1} using a Vivaspin 500 MWCO 10,000 (GE Healthcare) centrifugal device, aliquoted, flash-cooled in liquid nitrogen and stored at -80 °C. Typical yields were in the range of 1.0 mg per 500 ml of insect cell culture.

[†] The preceding text passage replaced the sentence "The TFIIH kinase module was prepared similarly."[292].

ATPase assay

In order to test the ATPase activities of core-TFIIH in separation, the 6-subunit 'reduced' core-TFIIH variants were subjected to a regenerative coupled ATPase assay that provided stable levels of ATP. In this system, hydrolyzed ATP is continuously regenerated by pyruvate kinase (PK), which in turn converts phosphoenolpyruvate (PEP) into pyruvate. Pyruvate on the other hand is converted into lactate by L-lactate dehydrogenase (LDH), which simultaneously oxidizes NADH into NAD$^+$. Thus, ATP hydrolysis is ultimately directly proportional to the oxidation of NADH. To quantify levels of NADH oxidation, absorption of the samples at 340 nm (A$_{340}$) was monitored. Whereas both NADH and NAD$^+$ absorb at 260 nm, only NADH absorbs at 340 nm and values for A$_{340}$ gradually decrease upon its oxidation. Although this setup in principle permits

quantification of the ATP-hydrolysis rate and thus the ATPase activity of a tested enzyme, in this study a simplified version of the assay was utilized. To determine whether ATP hydrolysis had occurred, merely the initial and final levels of NADH were compared.

ATPase assay mastermixes containing ssDNA or dsDNA (Table 3.5, Section 4.4) substrates (0.05-2 µM), MgCl$_2$ (4 mM), NADH (0.25 mM), PEP (2 mM), 2.0 U LDH (Sigma-Aldrich) and 1.5 U PK (Sigma-Aldrich) in buffer AA (150 mM potassium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 2 mM magnesium acetate, 5 mM β-mercaptoethanole) were prepared and 50 µL aliquots were transferred into wells of 384-well micro-plates (Greiner). Catalytic amounts of 6-subunit 'reduced' core-TFIIH complexes (0.1 µM) were added and samples were pre-incubated for 2 min at 30°C. Reactions were initiated by addition of ATP (2 mM). Using an Infinite® M1000 Pro micro-plate reader (Tecan), A$_{340}$ of the samples was monitored during incubation (1 h, 30°C) at allocated time points, usually every 60 s. Data of the single measurements were not used for quantification but to visualize and confirm the gradual decrease of NADH concentration in the samples. Wells without ATP, substrate or enzyme components were prepared as negative controls.

CTD phosphorylation assay

A simplified *in vitro* Pol II CTD phosphorylation assay was performed to test the activity and specificity of recombinant TFIIH kinase module. To produce completely dephosphorylated substrate for the assay, Pol II was treated for 1 h at 4°C with λ-phosphatase (lab stock, prepared by S. Vos, unpublished data) prior to the final SEC step during Pol II purification[260].

CTD-phosphorylation reactions containing dephosphorylated Pol II (8.8 µM) and MgCl$_2$ (4 mM) in buffer P (150 mM potassium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 2 mM magnesium acetate, 5 mM β-mercaptoethanole) were prepared. Catalytic amounts of the TFIIH kinase trimer (0.04 µM) were added and samples were pre-incubated for 2 min at 25°C. Reactions were started by addition of 4 mM ATP (Thermo Fisher Scientific) and monitored for 3 h at allocated time points. At each time point, 2 µL samples were retrieved. Samples were analyzed in 1:10 dilution by SDS-PAGE and western blot with primary α-Tyr1-P, α-Ser2-P, α-Ser5-P and α-Ser7-P antibodies and secondary α-rat-IgG-HRP antibodies (Table 3.8) as described (Section 4.4).

Reconstitution and analysis of 10-subunit TFIIH

Complete 10-subunit TFIIH was reconstituted from the 7-subunit core and 3-subunit kinase modules. Prior to assembly, core-TFIIH was diluted with ddH$_2$O to lower the salt content to 500 mM potassium acetate. Diluted protein was then mixed with 1.3*x* molar excess of kinase trimer and incubated (2-3 h, 5-10°C). The sample was centrifuged (21,000*g*, 10 min) to remove potential aggregates and subjected to sucrose-gradient centrifugation (Section 4.4.). The sucrose gradient was generated from a 10% sucrose light solution (10% (w/v) sucrose, 250 mM potassium acetate, 2 mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 0.75 mM AMP-PNP, 1 mM TCEP) and a 25% sucrose heavy solution (25% (w/v) sucrose, 250 mM potassium acetate, 2

mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 0.75 mM AMP-PNP, 1 mM TCEP). After gradient centrifugation, 200 μL fractions were collected from the top of the gradient and samples were analyzed by SDS-PAGE (Section 4.4).

Reconstitution and analysis of PIC variants

The cPIC, cPIC-cMed and PIC-cMed complexes, as well as a PIC devoid of Rad3 and kinase trimer ('PICΔRad3Δ3mer'), were reconstituted and subjected to sucrose-gradient centrifugation for initial analytical experiments and for negative stain EM analysis. Whereas the following scheme recapitulates the complete PIC-cMed assembly, redundant steps, such as incubation with TFIIH or cMed, were omitted if the respective proteins were not included in the final complex. Also, AMP-PNP was added only if complete 10-subunit TFIIH was present in a sample although all samples were diluted accordingly with buffer S. Unless stated otherwise, all incubation steps were performed at 25 °C.

10-subunit TFIIH was reconstituted from equimolar amounts of the 7-subunit core and the 3-subunit kinase modules (4°C, 10 min), prior to formation of the respective initiation complexes. PICs were assembled according to Table 8.1[†], starting with preparation of Pol II-TFIIF and DNA-TFIIA-TBP-TFIIB complexes (25°C, 10 min), which were combined to generate a Pol II/IIA–IIB–TBP–IIF–DNA assembly (cPIC). The cPIC sample was kept at 25°C for 5 min. TFIIE was added to previously reconstituted 10-subunit TFIIH and incubated for 5 min. Subsequently, the TFIIE-TFIIH complex was mixed with the cPIC. After incubation for 5 min, cMed and buffer S (2 mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 1 mM TCEP), which optionally included an appropriate amount of AMP-PNP for a final concentration of 0.75 mM in the sample, were added. Since buffer S also lowered salt concentration of the sample and thus facilitated complex formation, it was also provided if no AMP-PNP was required. Reconstituted complexes were incubated for 120 min with slight agitation.

Samples were centrifuged (21,000g, 10 min) to remove potential aggregates and subjected to sucrose-gradient centrifugation (Section 4.4). Standard gradients were generated from a 15% sucrose light solution (15% (w/v) sucrose, 150 mM potassium acetate, 2 mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 1 mM TCEP, optionally 0.75 mM AMP-PNP) and a 40%, sucrose heavy solution (40% (w/v) sucrose, 150 mM potassium acetate, 2 mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 1 mM TCEP, optionally 0.75 mM AMP-PNP). Several complexes, such as the PICΔRad3Δ3mer, were isolated in a gradient generated from 15% and 30% (30% (w/v) sucrose, 150 mM potassium acetate, 2 mM magnesium acetate, 25 mM K-HEPES pH 7.5, 2.5% (v/v) glycerol, 1 mM TCEP, optionally 0.75 mM AMP-PNP) sucrose solutions. For reconstitution of samples that were subsequently analyzed by negative stain EM, the sucrose heavy solution was supplemented with 0.13% (v/v) glutaraldehyde, whereas no crosslinker was added for analytical trials. After centrifugation, 200 μL fractions were collected from the top and complex formation was assessed by SDS-PAGE. Crosslinked complexes were quenched and dialyzed (150 mM potassium acetate, 2 mM magnesium acetate, 25 mM K-HEPES

pH 7.5, 1 mM TCEP) as described (Section 4.4). Dialyzed samples were applied to negative stain EM grids.

[†] Table 8.1 also indicates molar ratios of PIC-cMed complex components for assembly.

Preparation of the PIC-cMed complex for cryo-EM*

Closed PIC-cMed complex was prepared according to a protocol adapted from the previously reported assembly scheme[88], but with a slightly altered nucleic acid scaffold. The 106 nucleotide scaffold is based on the HIS4-promoter sequence (template[ø]: 5′-TGACACAGCG CAGTTGTGCTATGATATTTTTATGTATGTACAACACACATCGGAGGTGAATCGAACGT TCCATAGCTATTATATACACAGCGTGCTACTGTTCTCG-3′; nontemplate[ø]: 5′-CGAGAAC AGTAGCACGCTGTGTATATAATAGCTATGGAACGTTCGATTCACCTCCGATGTGTGTT GTACATACATAAAAATATCATAGCACAACTGCGCTGTGTCA-3′) and contains additional downstream DNA. Complete 10-subunit TFIIH was reconstituted from the 7-subunit core and the kinase trimer at 4°C before formation of the PIC-cMed complex. The PIC-cMed complex was assembled for cryo-EM according to the order in Table 8.1[†]. Beginning with the formation of a Pol II-IIF complex, the other initiation factors were added to generate a Pol II/IIA-IIB-TBP-IIF-DNA complex. TFIIE was incubated with previously assembled 10-subunit TFIIH for several minutes before being added to the Pol II-containing complex. After incubation for 5 min, buffer S (25 mM K-HEPES, pH 7.5, 2 mM magnesium acetate, 2.5% glycerol (v/v), 1 mM TCEP), with an appropriate amount of AMP-PNP to reach a final concentration of 0.75 mM, and cMed were added. The PIC-cMed complex was incubated for another 120 min shaking gently at 400 r.p.m. Unless stated otherwise, all incubation steps were performed at 25°C.

The PIC-cMed sample was centrifuged at 21,000g for 10 min and subjected to sucrose-gradient centrifugation in a 5 ml centrifugation tube. The gradient was generated from a 15% sucrose light solution (15% (w/v) sucrose, 150 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 2 mM magnesium acetate, 2.5% glycerol (v/v), 1 mM TCEP, 0.75 mM AMP-PNP) and a 40% sucrose heavy solution (40% (w/v) sucrose, 150 mM potassium acetate, 25 mM K-HEPES pH 7.5, 2 mM magnesium acetate, 2.5% glycerol (v/v), 1 mM TCEP, 0.75 mM AMP-PNP) containing 0.13% (v/v) glutaraldehyde crosslinker with a BioComp Gradient Master 108 (BioComp Instruments). Centrifugation was performed at 175,000g for 16 h at 4°C. Subsequently, 200 µl fractions were collected and quenched with a mix of 10 mM aspartate and 30 mM lysine for 10 min. Fractions containing crosslinked PIC-cMed complex were dialysed for 10 h in dialysis buffer (150 mM potassium acetate, 25 mM K-HEPES, pH 7.5, 2 mM magnesium acetate, 1 mM TCEP) in Slide-A-Lyzer MINI Dialysis Devices (2 ml, 20,000 MWCO) (ThermoFisher Scientific) to remove sucrose and glycerol. The dialyzed sample was concentrated to 0.7 mg ml^{-1} using a Vivaspin 500 MWCO 100,000 (GE Healthcare) centrifugal device and applied to cryo-EM grids.

[ø] Sequences of template and non-template strands are also listed in Table 3.5. Scaffold formation from single nucleic acid strands is described in Section 4.4.

[†] Table 8.1 also indicates molar ratios of PIC-cMed complex components for assembly.

Chemical crosslinking and crosslink analysis by mass spectrometry*[†]

PIC-cMed sample was crosslinked with a final concentration of 200 mM EDC (ThermoFisher Scientifc) in the sucrose heavy solution during gradient centrifugation. Fractions from the sucrose gradient were quenched with 50 mM ammonium bicarbonate. Fractions were dialyzed as before to remove sucrose and pooled for precipitation. Precipitated sample was dissolved in 50 μl buffer containing 8 M urea and 50 mM ammonium bicarbonate. Crosslinked sample was digested 1:20 (w/w) with trypsin and peptides were enriched by peptide size-exclusion chromatography and analysed in duplicate on an Dionex UltiMate 3000 RSLCnano HPLC system (Thermo Fisher Scientific) coupled to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific). MS acquisition was performed as described[293] with the exception that peptides were separated on the analytical column using a 63-min linear gradient. The data sets were analyzed with pLink 1.23[294] against a database containing the sequences of the protein components in the complex. Database search parameters included mass accuracies of MS1 < 10 p.p.m. and MS2 < 20 p.p.m., carbamidomethylation on cysteine as a fixed modification and oxidation on methionine as a variable modification. The number of residues of each peptide on a cross-link pair was set between 5 and 40 amino acids. A maximum of two trypsin-missed cleavage sites was allowed. An initial false discovery rate (FDR) cutoff of 1% was set. For simplicity, the crosslink score was represented as a negative logarithm value of the original pLink score and identified spectra with a score larger than three were considered. Results were visualized using the xiNET online server[295] and the XLink Analyzer Plugin[296] for UCSF Chimera[297]. New crosslinks are summarized in Fig. 5.12.

[†] Experiments were performed in collaboration with M. Hantsche, C.-T. Lee and I. Parfentev.

Negative stain EM data collection and analysis

Negative stain data were acquired to initially assess the quality of a protein sample on EM grids. Data collection was performed on carbon coated copper mesh grids (S160-4; Plano). Grids were glow-discharged (45 s, 25 mA, 0.39 mbar) with a PELCO easiGlow™ System (Ted Pella) before 3-5 µL sample were applied. After incubation for 60-90 s, excess sample was removed with filer paper (Whatman) and grids were washed twice for 30 s by floating them on the surface of 1 mL drops of dddH$_2$O. Grids were stained by floating them on 20 µL drops of freshly prepared 2% (w/v) uranyl formate solution for 30 s. The staining procedure was performed thrice with different drops of staining solution. Excess uranyl formate was removed with filter paper and grids were stored on filter paper without light exposure.

Negative stain EM data were acquired on a CM200 transmission electron microscope (Phillips) operated at 200 kV and equipped with a field emission gun (FEG) and a 2k x 2k CCD camera (TVIPS). Images were collected with an in-house TCL-based software in spot-scanning mode, at a nominal magnification of 88,000x (2.51 Å / pixel) and with a defocus range from -0.5 to -2.5 µm. Average datasets comprised 200-300 images with 15,000-20,000 particles. Particle coordinates were determined with the e2boxer.py tool implemented in EMAN2[298]. CTF correction and subsequent image processing, such as particle picking, reference-free 2D- or template-guided

3D-classification, were performed with the RELION 2.0.4 package[299,300].

Cryo-EM data collection*

Cryo-EM data collection was performed on R1.2/1.3 gold grids (Quantifoil). Grids were glow-discharged[†] for 45 s before application of 5 µl concentrated PIC-cMed sample, blotted for 5 s and vitrified by plunging into liquid ethane with a Vitrobot Mark IV (FEI) operated at 4°C and 100% humidity. Cryo-EM data were acquired on a FEI Titan Krios G2 transmission electron microscope (FEI) operated in EFTEM mode at 300 kV and equipped with a K2 Summit direct detector (Gatan). Automated data acquisition was carried out using the FEI EPU software package at a nominal magnification of 105,000x (1.37 Å per pixel). A total of 14,000 image stacks were collected at a defocus range from -0.5 µm to -5.0 µm. Each stack contained 40 frames that were acquired over a 10 s exposure time window in the counting mode of the camera. A dose rate of 4.2 e- Å^{-2} was applied, resulting in a total dose of 42 e- Å^{-2}.

[†] Glow discharging (45 s, 15 mA, 0.39 mbar) was performed with a PELCO easiGlow™ System (Ted Pella).

Image processing*

Cryo-EM image frames were stacked and processed with MotionCor2[301] and CTF parameter estimation was performed with Gctf[302]. CTF correction and subsequent image processing were performed with the RELION 2.0.4 package[299,300] unless indicated otherwise. Post-processing of refined models was performed with automatic *B*-factor determination in RELION and resolution was reported based on the gold-standard Fourier shell correlation (FSC) (0.143 criterion) as described[303] unless indicated otherwise. Local resolution estimates were determined using a sliding window of 40^3 voxels as described [89].

To obtain an initial particle set, coordinates of approximately 15,000 particles were determined semi-automatically with the e2boxer.py tool implemented in EMAN2[298]. The coordinates were imported into RELION and the respective particles were extracted with a 380^2 pixel box and normalized. Reference-free 2D class-averages were calculated and 20 representative 2D classes were selected. These were low-pass filtered to 20 Å and used as templates for automated particle picking on the first 700 micrographs, resulting in approximately 200,000 particles. Particles were extracted with a 380^2 pixel box size, normalized and screened by a combination of manual inspection and iterative rounds of reference-free 2D-classification. From the obtained improved 2D class-averages, 20 representative 2D classes were selected, low-pass filtered to 20 Å and used as templates for automated picking on the remaining micrographs with RELION. Initially, around 1.6 million particle images were obtained. Particles were extracted with a box size of 350^2 pixel, normalized and screened using a combination of iterative rounds of reference-free 2D- and template-guided 3D-classification with image alignment combined with manual inspection of the images in specific classes. An initial reference (ModelI) for the screening 3D-classifications had been obtained by performing one pre-3D-classification with the initial 200,000 particles using a 60 Å low-pass filtered EM map of the cPIC-cMed complex (EMDB accession EMD-2786)[88] as reference. Calculation of five 3D classes resulted in one class

with the complete PIC-cMed complex. This class was used as 'Model I' for the screening 3D-classifications after low-pass filtering to 60 Å. During the screening process, approximately 60% of the initial 1.6 million particles were discarded, resulting in 650,000 input particles. Using 'Model I' as the initial reference, iterative rounds of hierarchical 3D-classification with image alignment were performed as outlined in Fig. 5.10. After the first round of classification, classes with clearly visible density for cMed were selected. The same procedure was applied for classes with clear TFIIH density but no density for cMed, resulting in a separation of the classification tree in one branch for the PIC-cMed particles and one branch for PIC particles that lacked cMed. Before the second round of 3D-classification, new reference models ('Model PIC' and 'Model PIC-cMed') were generated from the best classes of the first round of 3D-classification and low-pass filtered to 60 Å. The second round of template-guided 3D-classification for the PIC-cMed branch was consequently performed with 'Model PIC-cMed' as a reference, whereas for the PIC branch 'Model PIC' served as reference. Subjecting the best 3D class of the PIC branch to a focused 3D-refinement with a local mask encompassing only TFIIH resulted in a reconstruction[†] with a resolution of 7.4 Å (after post-processing) from 32,000 particles.

[†] Not deposited.

Flexible Refinement ('WarpCraft')

The flexible refinement tool WarpCraft was developed and applied by D. Tegunov as reported[292]. Flexible refinement combined with local filtering of classes with apparently highly resolved PIC and PIC-cMed particles resulted in reconstructions with a nominal resolution of 4.7 Å from 58,000 particles for the PIC and a nominal resolution of 5.8 Å from 16,000 particles for the PIC-cMed complex (Fig. 5.10).

Structural Modeling*

For structural modeling we used both the continuous EM maps obtained by WarpCraft and EM maps with focus on specific regions in TFIIH. Model placement and docking of rigid bodies into the EM maps was performed with UCSF Chimera[297]. The I-TASSER[304,305], SWISS-Model[306,307] and Rosetta[308,309] tools were used for the generation of homology models of various PIC-cMed components as indicated in Tables 8.2 and 8.3. Manual modification of models and *de-novo* model building procedures were performed with COOT[310]. The model of the *S. cerevisiae* cPIC[89] was placed into the EM map and the Pol II clamp and stalk regions, as well as TFIIA, TFIIF[†] and peripheral regions in Rpb3, Rpb6, Rpb8, Rpb9 and Rpb12 were adjusted as rigid bodies. The model of TFIIB was extended in the B-linker and B-reader regions based on the Pol II-TFIIB crystal structure[ø76] (PDB code 4BBR). Homology models for TFIIE subunits Tfa1 and Tfa2 were generated based on the *H. sapiens* crystal structures of TFIIE[311] (PDB code 5GPY) and flexibly fitted into the TFIIE density, replacing the previous TFIIE model[ø]. The *S. cerevisiae* cMed homology model was adapted from the previously generated homology model of the *S. pombe* cMed crystal structure[247] (PDB code 5N9J). To improve the fit to the EM map, cMed was divided into seven rigid bodies (head module, knob, hook-connector, plank, beam RWD1-UBC1, beam

RWD2 and beam UBC2) that were placed in the density individually[†]. Downstream DNA was generated by placing three pieces of ideal B-DNA into the density, connecting these in COOT and performing alternating rounds of real space refinement with secondary structure restraints and geometry optimization in PHENIX[312]. For a summary on structural modeling of proteins, see Tables 8.2 and 8.3.

We generated a conservative model of *S. cerevisiae* TFIIH with the use of available structural information. Models of domains were first derived based on structures of TFIIH homologues from different species and on other structures with regions of partially related sequences. Homology models were generated for the Tfb1 BSD1 and BSD2 domains, for the three Tfb2 helix-turn-helix motifs, for the Tfb3 RING-finger, for the Tfb4 vWA-fold, for the eZnF domains in Tfb4 and Ssl1, and for the Ssl1 RING-finger. These models were derived from the *H. sapiens* NMR structure of the BSD1 domain (PDB code 2DII), the *Staphylococcus aureus* CadC crystal structure[313] (PDB code 1U2W), the *H. sapiens* MUS81 NMR structure[314] (PDB code 2MC3), the *Pyrococcus furiosus* TrmBL2 crystal structure[315] (PDB code 5BOX), the NMR structure of the *H. sapiens* Mat1 RING-finger[210] (PDB code 1G25), the crystal structure of the *H. sapiens* p34 vWA-fold[181] (PDB code 4PN7), the crystal structure of a *H. sapiens* E3 ubiquitin ligase (PDB code 3LRQ), the crystal structure of *P. furiosus* rubrerythrin[316] (PDB code 1NNQ) and the NMR structure of the *H. sapiens* p44 RING-finger[182] (PDB code 1Z60), respectively. In addition, two 3-helix bundle domains, one located at the C-terminus of Tfb1 (residues 543-639) and one located C-terminally of the Tfb3 RING-finger (residues 71-145) were modeled ab initio using the QUARK server[317]. Together with the crystal and NMR structures of the Tfb1 PHD[183] (PDB code 1Y5O), the Tfb2-Tfb5 dimerization domains[192] (PDB code 3DGP) and the vWA-fold of Ssl1[179] (PDB code 4WFQ), the homology and *ab initio* models listed above were placed into the density and rigid-body adjusted. If the correct position of models could not be deduced from the electron density directly, placement was performed on the basis of BS3- and SBAT-derived crosslinks that had been published[101,190,201] or EDC-derived crosslinks obtained in this study (Fig. 8.12).

Several homology models were subjected to conservative modifications, in particular minor truncations, short α-helical extensions and positional corrections, to improve their fit to the electron density manually. The *S. cerevisiae* crystal and NMR structures exhibited a good fit to the electron density and did not require modification with the exception of the Tfb1 PHD that was C-terminally extended (residues 115-121). Homology models for the ATPases Rad3 and Ssl2 were generated from crystal structures of their *Thermoplasma acidophilum*[208] (PDB code 2VSF), *Archaeglobus fulgidis*[107] (PDB code 2FWR) and *H. sapiens*[209] (PDB code 4ERN) homologues. The models were split into their domains (Rad3: lobe 1, FeS-cluster, ARCH and lobe 2; Ssl2: lobe 1 and lobe 2) and placed individually into the electron density. Lobe 1 and lobe 2 of Rad3 did not require further adaptation. The FeS-cluster was placed by superpositioning the *T. acidophilum* Rad3 structure onto the TFIIH model in COOT and extracting the coordinates of the Fe and S atoms. A backbone model of the ARCH domain was generated with Gorgon[318,319] and used as an additional input to calculate a second homology model, which then was adjusted to the density. It

accounted for an evolutionary difference between *S. cerevisiae* and *T. acidophilum* that had resulted in an extension of two α-helices and an insertion of one α-helix and a loop (residues 255-347). In the Ssl2 homology model an additional α-helix (residues 468-481) was placed into well-defined density substituting for an initially unstructured stretch of residues. Two loops (residues 426-451, 692-702) with significant deviation from the EM density were manually adjusted. The location of four Tfb1 α-helices (residues 308-330, 369-394, 465-483 and 495-519) and two TFIIE α-helices (residues 267-289 and 349-373) was confirmed by XL-MS analysis as described above and the respective α-helices were placed into the corresponding density. The TFIIE acidic region (residues 407-417) interacting with the PHD of Tfb1 was modeled based on the NMR structure of the *H. sapiens* TFIIE C-terminal region bound by the PHD of p62[102] (PDB code 2RNR). Additionally, a few linkers and α-helical regions within TFIIH subunits Tfb1 (residues 219-251, 295-307, 331-353 and 484-494), Tfb2 (residues 3-40, 113-159, 195-214, 380-419 and 433-450), Tfb4 (residues 89-97, 103-114 and 257-273), and Ssl1 (residues 308-324 and 373-386) which could be clearly traced in the EM maps and assigned respectively were built *de novo*. Lastly, one of the clutch domains, consisting of four β-strands and one α-helix, was built *de novo*. Owing to the limited resolution of the EM maps in this region, however, the secondary structure elements in the domain could not be assigned to the two adjacent TFIIH subunits Tfb2 and Ssl2 with confidence[‡]. To address these concerns the modeled clutch domain was assigned to a newly introduced and unrelated chain Z. Residues with any probability to deviate from the indicated sequence register, i.e. all residues in the *de-novo* built elements, were denoted as UNK in the deposited PDB model.

The model fit to the EM maps was further optimized by iterative rounds of flexible fitting with vmd[320] and MDFF[321]. Each flexible fitting procedure was divided in three simulation steps, starting with a simulation at room temperature, followed by a cooling step to 0 K and a third step in which the simulation was performed at 0 K. Flexible fitting was performed without domain restraints for small units and with domain restraints once models had been combined into larger entities.

Density-adjusted PIC and PIC-cMed models were refined using the geometry minimization routine in PHENIX[312] with applied secondary structure and rotamer restraints. A brief overview of EM data collection, data processing and model statistics for the final PIC and PIC-cMed models is provided in Table 8.4. Figures were generated using UCSF Chimera[297].

[†] Modeling was performed by M. Hantsche.
[ø] Modeling was performed by C. Dienemann.
[‡] The possibility of a domain swap by a β-strand exchange between Tfb2 and Ssl2 could not be excluded.

5 Results

5.1 Preparation and assembly of recombinant TFIIH

Since the discovery of the TFIIH complex more than two decades ago[193] its preparation has remained a tedious and inefficient process despite a plethora of biochemical and biophysical studies[145,148] aiming to elucidate its structural and functional properties. To date, TFIIH is obtained primarily from endogenous sources, such as HeLA and yeast cells, and purified by affinity and immuno-affinity chromatography techniques[85,165]. These standard approaches require large quantities of input cell material and yield comparatively little and partially inhomogeneous protein. Although efforts to express *H. sapiens* TFIIH recombinantly in insect cells were successful in principle[203,204], large-scale production schemes, which would permit further extension of the structural and *in-vitro* analyses of the complex, have not been established. As a consequence, the first step towards a more profound, high-resolution structural characterization of TFIIH and its assemblies, like the transcription initiation machinery, must be the development of an efficient strategy for the recombinant production of high-quality TFIIH samples.

5.1.1 Dissection of the protein network in core-TFIIH

Initial experiments were focused on the analysis of interactions within TFIIH and on the identification of targets for crystallization experiments, whereas the assembly of complete 10-subunit TFIIH and its incorporation into the initiation complex was defined as a future goal.

The recombinant expression of TFIIH subunits and small subcomplexes was first tested in *E. coli* after transformation with the respective pET-MCN plasmids (Methods, Table 3.4). Experiments revealed that all subunits except for those with enzymatic function (Kin28, Rad3, Ssl2) could be produced in bacteria. With the exception of Tfb1, the proteins were obtained in soluble form either untagged or after introduction of N-terminal Ulp1-cleavable 6×His-tags (6×His-SUMO). The solubility of Tfb1 was dependent on the presence of Ssl1, confirming the intimate physical interaction between these subunits[155,322] (Fig. 5.1 a). Pairwise co-expression coupled with subsequent pulldown interaction assays further confirmed the previously characterized interactions between Ssl1 and Tfb1, and between Tfb2 and Tfb5[192,322]. In addition, the formation of a predicted Ssl1-Tfb4 dimer[323] was observed and Tfb2 was newly identified as a direct binding partner of Ssl1 and Tfb4 (Fig. 5.1 a, b). These findings provided better insight into the protein network within core-TFIIH and, in combination with prior analyses[121,165,198,203,205], suggested a strategy for reconstitution of the holo-complex. While 'minimal core-TFIIH', encompassing Ssl1, Tfb1, Tfb2, Tfb4 and Tfb5, was to be co-expressed in *E. coli* cells and purified as one complex, the catalytic subunits Rad3 and Ssl2 would be produced in a different expression system, purified separately and assembled with minimal core-TFIIH into full, 7-subunit core-TFIIH. The TFIIH kinase module was to be prepared as an isolated complex based on previously reported strategies[324].
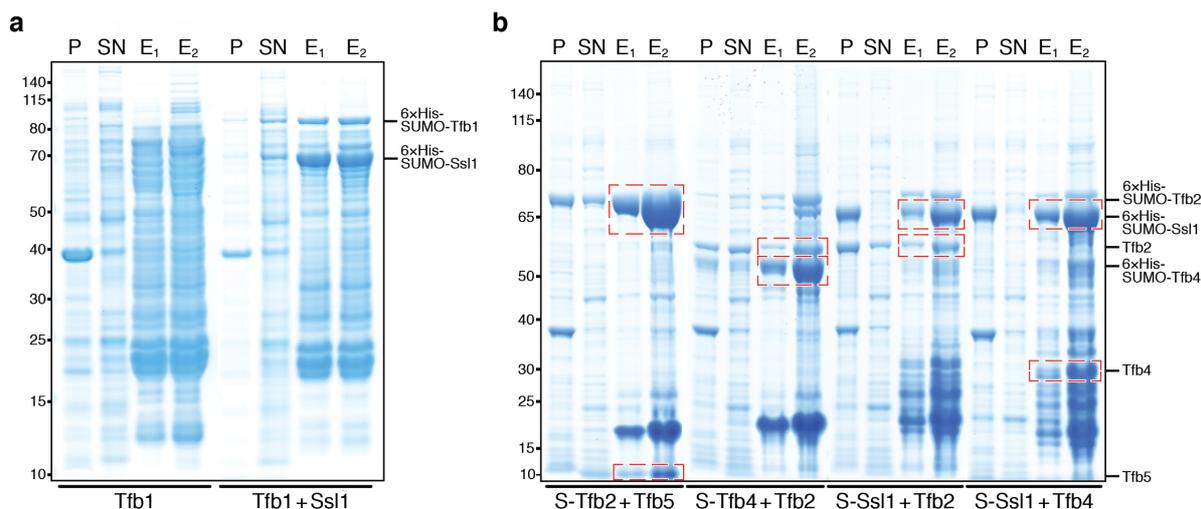
Figure 5.1 | Dimeric complexes within core-TFIIH. a, Ssl1 co-expression enhances solubility of Tfb1. Comparison of protein levels during expression and pulldown of 6×His-SUMO-Tfb1 in isolation (left) and in complex with Ssl1 (right) by SDS-PAGE analysis. Tfb1 is detected only in presence of Ssl1. Pulldowns were performed with Ni-NTA agarose resin. Sample fractions are indicated (P, pellet; SN, supernatant; E, elution) and a molecular weight standard is depicted on the left. **b,** Binary interaction assays identify prominent dimers within core-TFIIH. Various pairs of TFIIH subunits comprising one 6×His-SUMO-tagged and one untagged subunit were co-expressed. Dimer complexes obtained by affinity pulldown with Ni-NTA agarose resin are marked with dashed red outlines. Protein levels were analyzed by SDS-PAGE and protein identity was confirmed by mass spectrometry (MS). Labels as described in **a**. S, 6×His-SUMO.

5.1.2 Preparation of recombinant minimal core-TFIIH and initial crystallization trials

In addition to its proposed role for the assembly of 7-subunit core-TFIIH, minimal core-TFIIH appeared to be a reasonable target for crystallization. The pentameric complex was ultimately produced from *E. coli* cells that had been co-transformed with a pET-MCN vector comprising the genes for Ssl1, Tfb2, Tfb4, and Tfb5 and a second pET-MCN vector carrying the Tfb1 gene (Table 3.4). Incorporation of the Tfb1 gene in the multi-cistronic construct and co-expression of all five proteins from one plasmid had resulted in poor yields for Tfb1, regardless of its position within the transcript. All proteins were expressed as SUMO-tagged versions with enhanced solubility. A three-step purification scheme for preparation of minimal core-TFIIH was established. The purified complex was stable during size exclusion chromatography (SEC) in low-salt conditions and contained all subunits in apparently stoichiometric amounts (Methods, Fig. 5.2). Truncations of minimal core-TFIIH were generated based on reported domain assignments[190,280], the prediction of secondary structure elements (SSEs) and on limited proteolysis experiments. These suggested the removal of N- and C-terminal stretches in Ssl1, Tfb1 and Tfb4 as well as deletion of internal loops in Ssl1, Tfb1, Tfb4 and Tfb2. The most truncated minimal core-TFIIH version to retain complex stability during purification and final SEC encompassed Ssl1Δ1-91, Tfb1Δ1-164, Tfb2, Tfb4Δ1-17Δ323-338 and Tfb5 (Methods, Fig. 5.2).

Extensive crystallization trials (Methods) remained unsuccessful despite testing of distinctly truncated core-TFIIH variants in various buffer conditions (TRIS, HEPES, pH range 6.8-8.5) and with different salt (NaCl, KCl, KOAc) concentrations (50-300 mM). Although initial crystals were yielded in rare cases, these were irreproducible or identified as salt crystals during

in-situ diffraction screening on a PX scanner (Rigaku). Additional internal loop deletions were attempted but never experimentally executed as the project's focus was shifted to the production of complete 10-subunit TFIIH and its incorporation into the transcription initiation complex.
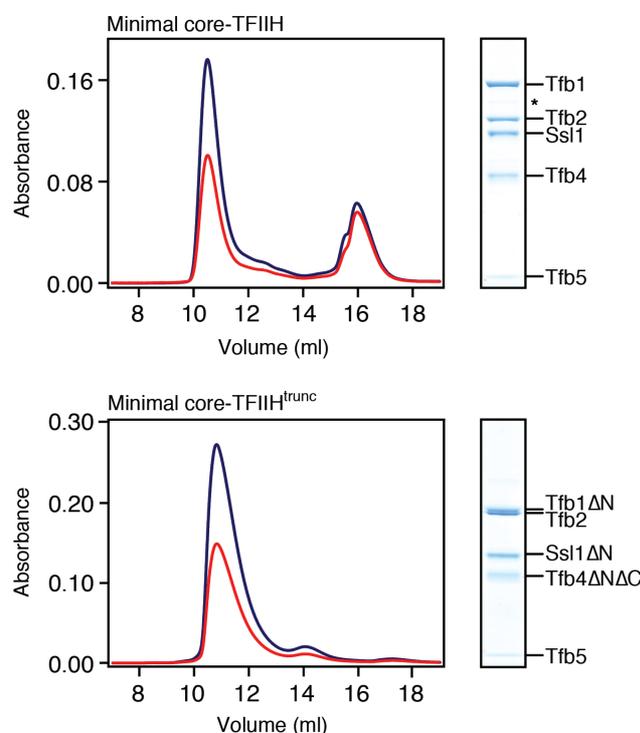
Figure 5.2 | Preparation of recombinant minimal core-TFIIH. Analysis of purified minimal core-TFIIH complexes by SEC and SDS-PAGE revealed formation of homogeneous complexes with apparently stoichiometric subunits. An impurity in the full-length complex (top), which was identified as an N-terminal degradation product of Tfb1 by MS, is indicated by an asterisk. The most truncated stable minimal core-TFIIH version (bottom, truncations abbreviated as ΔN or ΔC) contains highly pure subunits. Protein identity was confirmed by MS.

5.1.3 Preparation of Rad3 and Ssl2

In addition to minimal core-TFIIH, which had been obtained successfully, the ATPases Ssl2 and Rad3 were required for the assembly of 7-subunit core-TFIIH. Previous studies involving their archaeal and *H. sapiens* homologues[107,175,204,208,209] suggested that recombinant production of the *S. cerevisiae* proteins in insect cells was feasible. Multiple Series-438-based Ssl2- and Rad3-constructs were generated and proteins were expressed in Hi5 cells after baculovirus infection. Various tags and solubility enhancers, such as 6/8×His-3C/SUMO/TEV-, 6×His-MBP- or GST-tags, as well as N- and C-terminal tag positions were tested. However, although most variants were well expressed, the isolated proteins remained essentially insoluble or were present in solution only in form of multivariate aggregates despite the usage of *S. frugiperda* codon-optimized constructs and the performance of extensive buffer screens. These results indicated that Rad3 and Ssl2 required their natural binding partners as chaperones for solubility in yeast. Consequently, complete 7-subunit core-TFIIH had to be obtained as a singe complex by co-expression in insect cells, thereby rendering the production of minimal core-TFIIH in *E. coli* cells obsolete.

5.1.4 Preparation of recombinant core-TFIIH complexes

The basic trials for baculovirus-mediated expression of TFIIH subcomplexes in insect cells were performed with constructs for a 6-subunit 'reduced' core-TFIIH assembly, which comprised Ssl1, Ssl2, Tfb1, Tfb2, Tfb4 and Tfb5 but lacked Rad3 ('core-TFIIH$^{\Delta Rad3}$'). This subunit was initially excluded to decrease the number of respectively co-expressed subunits and thus the variables for construct generation. Moreover, a strong affinity of Rad3 for the TFIIH kinase module had been reported[105,164] and it remained elusive whether the absence of this interaction in core-TFIIH would affect recombinant protein production or complex solubility. Co-expression vectors were compiled using a parallelized semi-high-throughput LIC-based cloning strategy (Methods). Literature[165], SSE predictions and previous results suggested the placement of affinity tags on Ssl1, Tfb1 or Tfb4. Series-438 plasmids, which contained 3C- and SUMO-tags or TEV-cleavable (6×His-)MBP- and 6×His-tags on the genes for these subunits, were analyzed by co-expression tests in Hi5 insect cells and subsequent pulldown and interaction assays (Methods). In these experiments intact though partially heterogeneous complexes were obtained, indicating that core-TFIIH$^{\Delta Rad3}$ formed a stable assembly.
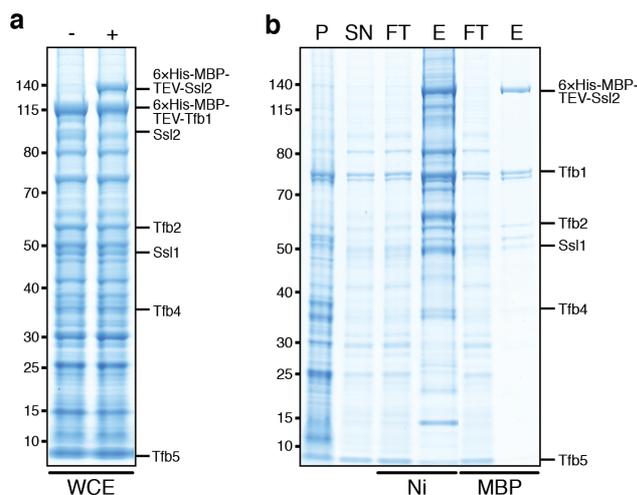
Figure 5.3 | Beneficial effects of 6×His-MBP-tags on core-TFIIH$^{\Delta Rad3}$ expression and purity. a, Addition of a 6×His-MBP-tag enhances Ssl2 expression. Comparison of Ssl2 expression levels of core-TFIIH$^{\Delta Rad3}$ constructs comprising 6×His-MBP-tags on either Tfb1 (left lane) or on Tfb1 and Ssl2 (right lane). A significantly larger amount of Ssl2 was produced if it was additionally tagged. Whole-cell-extract (WCE) samples were analyzed by SDS-PAGE. A molecular weight standard is provided on the left. **b,** Use of MBP-tags in affinity pulldowns yields near-pure protein sample. Affinity assays with a core-TFIIH$^{\Delta Rad3}$ construct comprising a 6×His-MBP-tag on Ssl2 were performed with equal input sample (SN) but utilizing either Ni-NTA agarose (Ni; lanes 3+4) or amylose (MBP; lanes 5+6) resin. A small amount of near-pure core-TFIIH$^{\Delta Rad3}$ complex was obtained after elution from amylose resin. Proteins in the Ni-NTA agarose elution fraction were more abundant but the core-TFIIH$^{\Delta Rad3}$ complex was significantly contaminated. Overall low yields of the assay indicate poor accessibility of the affinity tag on Ssl2. Protein levels were analyzed by SDS-PAGE. Labels as described in Fig. 5.1.

As Ssl2 was produced in significantly smaller amounts than other components of core-TFIIH$^{\Delta Rad3}$, the expression ratio between Ssl2 and the remaining subunits was optimized by introduction of a 6×His-MBP-tag (Fig. 5.3 a), which additionally enhanced protein solubility. Although the use of MBP-tags yielded almost pure protein in one affinity step, the 6×His-MBP-tag on Ssl2 was not sufficient to retain the complex efficiently in affinity assays, possibly due to poor accessibility of the maltose-binding site (Fig. 5.3 b). Thus, ultimately a co-expression construct comprising (6×His-)MBP-tags on the Ssl1 and Ssl2 genes was

selected for large-scale preparation (Fig 5.4 a, Table 3.4). Proteins were purified by a four-step strategy, yielding an apparently stoichiometric 6-subunit complex that was stable during SEC in low-salt conditions (Methods, Fig. 5.4 b).
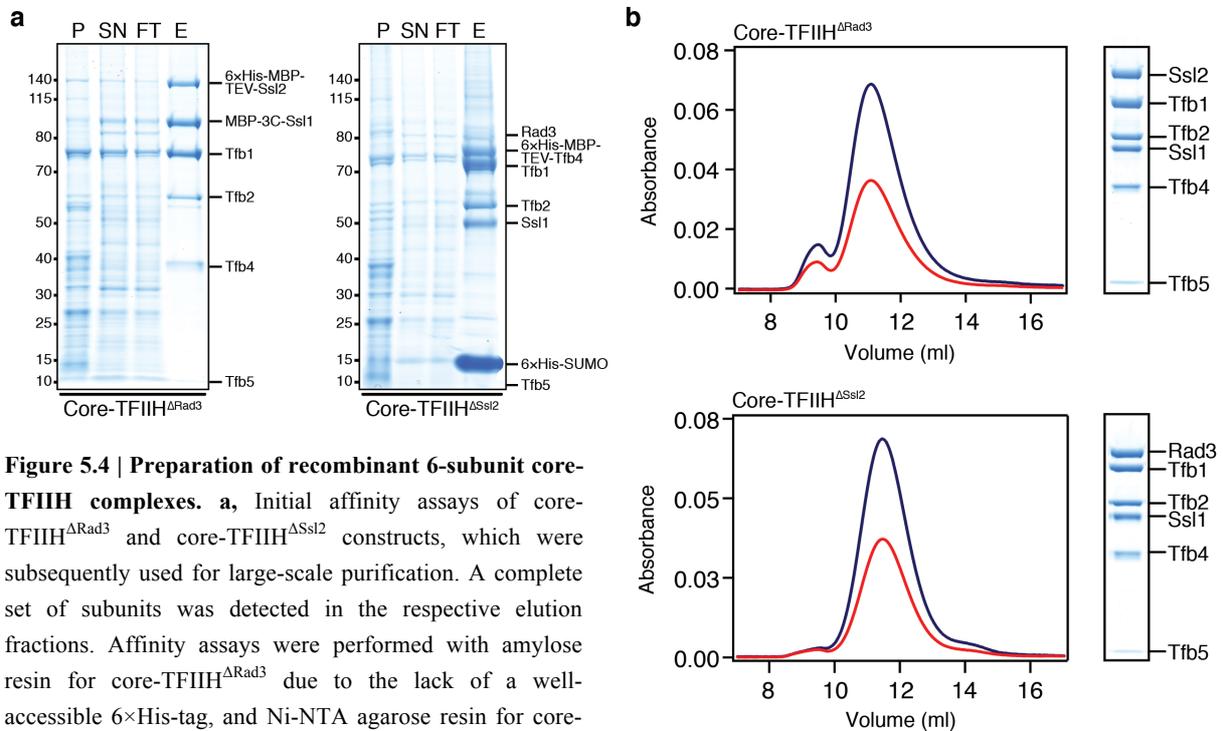
Figure 5.4 | Preparation of recombinant 6-subunit core-TFIIH complexes. a, Initial affinity assays of core-TFIIH$^{\Delta Rad3}$ and core-TFIIH$^{\Delta Ssl2}$ constructs, which were subsequently used for large-scale purification. A complete set of subunits was detected in the respective elution fractions. Affinity assays were performed with amylose resin for core-TFIIH$^{\Delta Rad3}$ due to the lack of a well-accessible 6×His-tag, and Ni-NTA agarose resin for core-TFIIH$^{\Delta Ssl2}$. Samples were analyzed by SDS-PAGE. Elution fractions were treated with Ulp1 (15 min, 4°C) before analysis, resulting in cleavage of residual 6×His-SUMO tags. Labels as described in Fig. 5.1. **b,** Purified core-TFIIH$^{\Delta Rad3}$ and core-TFIIH$^{\Delta Ssl2}$ complexes. Analysis of recombinant TFIIH assemblies by SEC and SDS-PAGE revealed formation of pure and homogeneous complexes with apparently stoichiometric subunits. Protein identity was confirmed by MS.

Initial constructs for the preparation of 7-subunit core-TFIIH were generated based on the core-TFIIH$^{\Delta Rad3}$ expression vector, which was extended by addition of 6×His-MBP-Rad3 by LIC (Methods). However, application of the established protocols resulted in poor expression and low yields of core-TFIIH (< 20 µg / 1 L cell culture), suggesting adverse effects of the MBP-tags on Rad3, Ssl1 or Ssl2 in context of the 7-subunit complex. Consequently, the semi-high-throughput screening approach, which had permitted identification of an optimized core-TFIIH$^{\Delta Rad3}$ expression plasmid, was also applied to evaluate prospective 7-subunit core-TFIIH constructs. The analyses revealed that the 6×His-MBP-tag on Rad3 was poorly accessible and, more importantly, partially resistant to removal by TEV-cleavage. These observations indicated unanticipated co-/ or mis-folding events between MBP and Rad3. Indeed, the removal of tags from both Rad3 and Ssl2 ultimately proved to be a key step for the production of high-quality core-TFIIH. Moreover, tags on the Ssl1 subunit appeared to be incompatible with presence of Rad3 in the 7-subunit complex, possibly as a side effect of the intricate association between the ATPase and Ssl1[173]. Thus, the finally selected co-expression vector comprised a single 6×His-MBP-tag on Tfb4. In addition, *S. frugiperda* codon-optimized versions of the Rad3 and Ssl2 genes were introduced, which provided increased expression rates of the ATPases with respect to

the affinity-tagged subunit (Table 3.4). The components of core-TFIIH were co-expressed and purified as described for core-TFIIH$^{\Delta Rad3}$ (Methods) and the 7-subunit complex was stable during SEC in buffers containing 5% glycerol (Fig. 5.5).

To obtain a counterpart to core-TFIIH$^{\Delta Rad3}$ for activity tests, a construct for 'core-TFIIH$^{\Delta Ssl2}$', a 6-subunit complex which lacked Ssl2, was generated. This co-expression vector was identical to the one assembled for core-TFIIH but lacked the Ssl2 gene (Fig 5.4 a, Table 3.4). The complex was prepared according to previously established strategies. During purification, core-TFIIH$^{\Delta Ssl2}$ displayed characteristics that resembled 7-subunit core-TFIIH more than core-TFIIH$^{\Delta Rad3}$ (Fig 5.4 b).
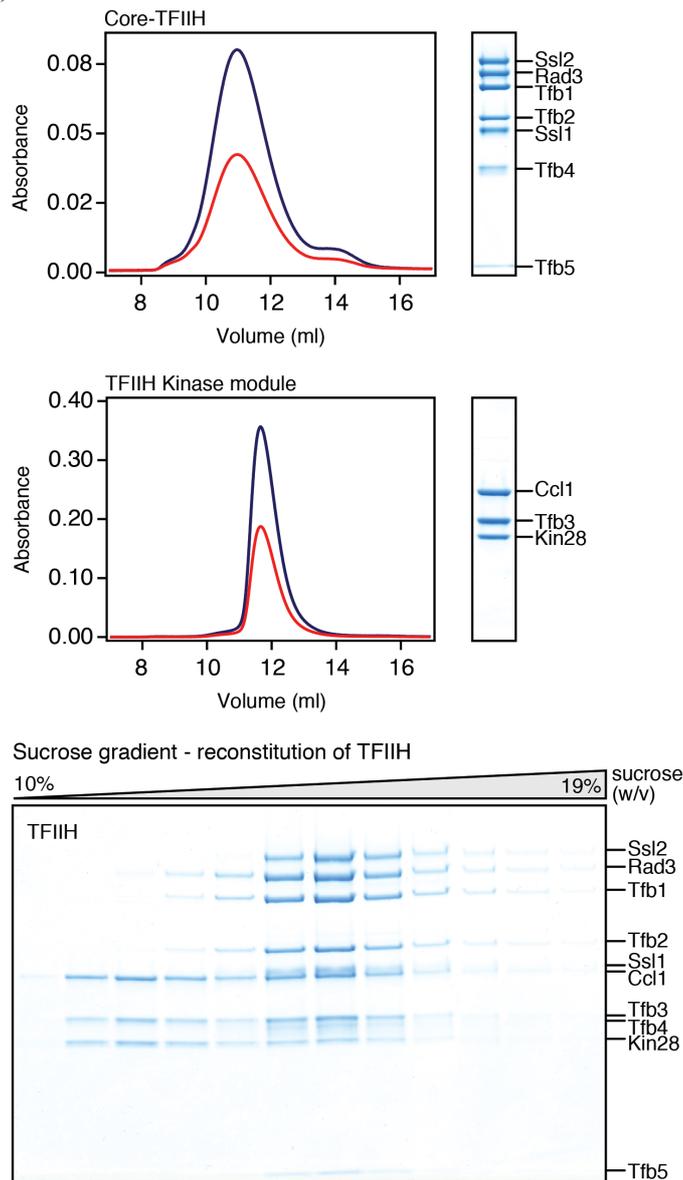
Figure 5.5 | Preparation of recombinant TFIIH. Analysis of purified TFIIH core and kinase modules by size-exclusion chromatography and SDS-PAGE revealed high purity and homogeneity of the complexes with apparently stoichiometric subunits. SDS-PAGE analysis of fractions 1-13 of a sucrose gradient centrifugation after reconstitution of TFIIH from purified core and kinase modules. A shift in the bands originating from the subunits of the kinase module (Ccl1, Kin28 and Tfb3) by four fractions was detected, indicating formation of complete TFIIH. This experiment was repeated three times with equivalent results. Figure 5.5 and the corresponding figure legend are excerpts from Schilbach et al.[292], as indicated on page VI.

In addition to core-TFIIH, which had been successfully retrieved, the TFIIH kinase module was required for the assembly of 10-subunit TFIIH. According to literature[324], co-expression of the three subunits Ccl1, Kin28 and Tfb3 was a well-established process. Since the catalytic subunit was likely to be a limiting factor during preparation, the Kin28 gene was fused with a 6×His-MBP-tag when expression vectors were generated by LIC (Methods). Applying the previously described screening approach, distinctly tagged versions of Tfb3 and Ccl1 were co-expressed with 6×His-MBP-Kin28 and assessed in pulldown interaction assays. The construct for large-scale production was selected based on the final ratio between kinase, cyclin and the accessory factor Tfb3. It comprised cleavable 6×His-(3C/SUMO)-tags on Tfb3 and Ccl1, which, however, were not used during purification. The applied purification scheme was adapted from the four-step strategy for core-TFIIH preparation and yielded a stoichiometric complex that was stable during SEC in very low salt conditions (Methods, Fig. 5.5).

10-subunit holo-TFIIH was reconstituted from the 7-subunit core and the 3-subunit kinase modules and isolated in a sucrose gradient (Methods, Fig. 5.5). Similar approaches involving size exclusion chromatography were ineffective. A 1.3x excess of kinase module was sufficient for TFIIH complex formation.

5.1.5 Activity assays

To test the catalytic activity of the recombinant TFIIH core and kinase modules, enzymatic assays were performed. A CTD-phosphorylation assay (Methods) with a dephosphorylated Pol II substrate was applied to test the functionality and specificity of the isolated kinase module (Fig. 5.6). An increase in phosphorylation was detected for all analyzed residues (Tyr1, Ser2, Ser5 and Ser7) of the Pol II CTD, indicating enzymatic activity of the trimeric complex. In addition, the kinase module exhibited substrate specificity and phosphorylated primarily Ser5- and, to a lesser extent, Ser7 residues. Ser2 and Tyr1 residues displayed an unexpected, albeit significantly weaker, increase in phosphorylation levels, which may be attributed to unspecific phosphorylation events or antibody cross-reactivity. Although a direct comparison between phosphorylation levels of different residues is problematic since no quantification was performed, the obtained results appear to be generally consistent with literature[19]. However, it cannot be excluded that the specificity of the kinase module is altered in context of 10-subunit TFIIH or the complete initiation machinery.

The ATPase activities of Ssl2 and Rad3 were separately assayed utilizing the core-TFIIH$^{\Delta Rad3}$ and core-TFIIH$^{\Delta Ssl2}$ constructs (Methods). Reflecting their distinct substrate specificities, dsDNA and ssDNA substrates were provided for core-TFIIH$^{\Delta Rad3}$ and core-TFIIH$^{\Delta Ssl2}$, respectively. Consistent with literature, both constructs exhibited enzymatic activity and hydrolyzed ATP[109,325], suggesting a correct fold of the ATPases within the 6-subunit complexes. ATP hydrolysis rates were not determined.
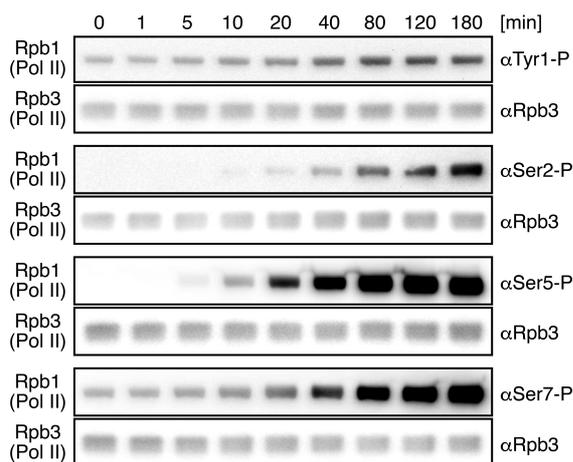
Figure 5.6 | Recombinant TFIIH kinase module is catalytically active and selectively phosphorylates the Pol II CTD. Analysis of the Pol II CTD phosphorylation state by western blot. Specific antibodies against the Tyr1, Ser2, Ser5 and Ser7 residues in the CTD of the Pol II subunit Rpb1 were utilized to monitor relative levels of CTD phosphorylation over a 180 min period. Although phosphorylation levels gradually increased for each analyzed residue, the TFIIH kinase module preferentially targeted Ser5- and Ser7 residues. The Pol II subunit Rpb3 was loaded as a control.

5.2 Optimization of initiation complex assembly and initial negative stain EM analysis

Once protocols for the preparation of recombinant TFIIH were being established and the complex was becoming available for structural investigation, appropriate targets and techniques for its analysis were specified. Although originally defined as a long-term goal, ultimately the incorporation of 10-subunit TFIIH into the transcription initiation machinery was attempted. To this effect a cMed-containing pre-initiation (PIC-cMed) complex was to be formed and its structure was to be determined by cryo-EM. Related transcription initiation complexes had been subjected to cryo-EM previously[86,88,89,101], thereby providing initial guidance on sample preparation. Due to the complexity and sheer size of such an assembly, however, the existing reconstitution approaches had to be re-evaluated and optimized.

5.2.1 Complex preparation by gradient centrifugation with optimized TFIIE excess

Owing to the fragility and sensitivity of macromolecular initiation complexes, an altered protocol for their assembly and isolation, which involved their sedimentation in a sucrose gradient, was established (Methods)[†]. Using this technique, stable and apparently perfectly stoichiometric samples were reproducibly obtained. In addition, this method permitted a significant reduction of the molar excess of TFIIE over the limiting factor Pol II during initiation complex reconstitution, which was demonstrated for the cPIC-cMed complex (Fig. 5.7). Instead of a 10x molar excess of TFIIE, as claimed in the original SEC-based approach[89], only a 2x molar excess was required for assembly of a stoichiometric cPIC-cMed complex in a 15-30% sucrose gradient. Since TFIIH is recruited to the PIC primarily by TFIIE[100] equimolar amounts of TFIIH were found to be sufficient as input for formation of the PIC-cMed complex, thus greatly facilitating sample preparation.

[†] Initial trials of sucrose gradient centrifugation with cPIC complexes were performed by C. Dienemann.
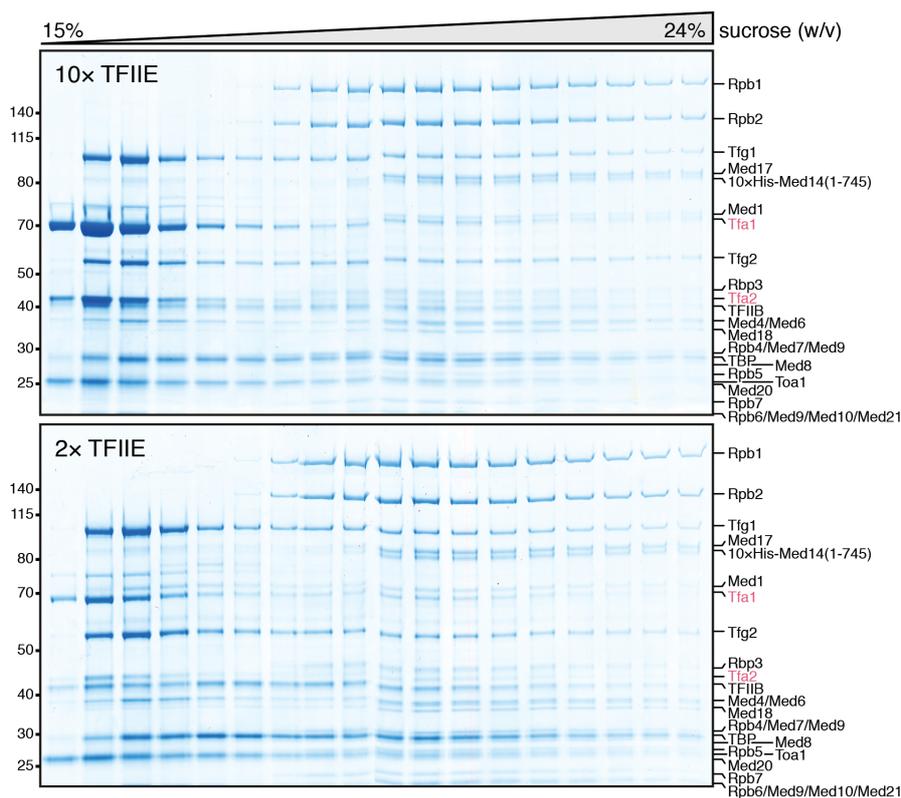
Figure 5.7 | Reconstitution of cPIC-cMed complexes with 10*x* and 2*x* molar excess of TFIIE. SDS-PAGE analysis of fractions 1-18 of 15-30% sucrose gradient centrifugations. Stoichiometric cPIC-cMed complexes were successfully assembled with 10*x* (top) and 2*x* (bottom) molar excess of TFIIE. Other cPIC-cMed components were added as described (Table 8.1). TFIIE subunits (Tfa1, Tfa2) are highlighted. Tfa1 and Tfa2 are present in apparently identical amounts in both reconstituted cPIC-cMed complexes.

5.2.2 Initial analysis of a PIC-cMed complex variant by negative stain EM

The optimized sucrose gradient protocol was initially tested with a TFIIH-containing sample during assembly of an incomplete PIC-cMed complex that comprised core-TFIIH$^{\Delta Rad3}$, the first recombinant construct to become available. The components were reconstituted according to the scheme that was designed for formation of the complete PIC-cMed complex (Table 8.1) but a 15-30% sucrose gradient was applied for sedimentation. The obtained 'PICΔRad3ΔTrimer' complex contained subunits in stoichiometric amounts except for core-TFIIH$^{\Delta Rad3}$, which appeared to be present in excess (Fig. 5.8 a), possibly owing to unspecific interactions with DNA. To determine whether the TFIIH subcomplex remained stably associated with cPIC-cMed, the sample was subjected to negative stain EM (Methods). The collected micrographs revealed well-distributed particles with distinct shapes, indicating a large variety of orientations and a good angular distribution (Fig. 5.8 b). A small dataset was collected and processed, resulting in a 3D-reconstruction of the complex at ~30 Å resolution. Comparison with the cryo-EM map of the cPIC-cMed complex (EMDB accession EMD-2786)[88] revealed the presence of additional density that presumably corresponded to a monomer of the 320 kDa core-TFIIH$^{\Delta Rad3}$ complex (Fig. 5.8 c) with several flexible regions. Based on these positive results, the reconstitution of the complete

PIC-cMed complex and its structural analysis by cryo-EM were performed immediately upon retrieval of the recombinant TFIIH core and kinase modules.
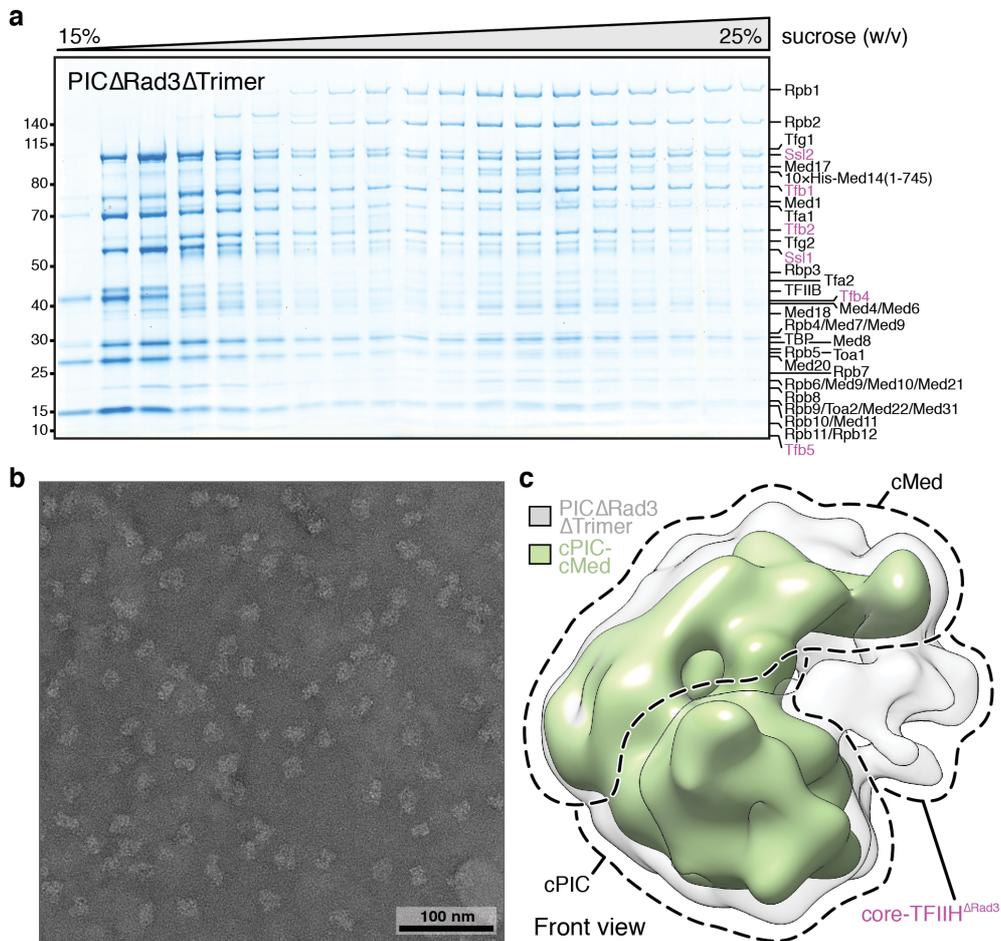
Figure 5.8 | Negative stain EM analysis of the PICΔRad3ΔTrimer complex. a, Complex formation. SDS-PAGE analysis of fractions 1-19 of a 15-30% sucrose gradient centrifugation after reconstitution of the PICΔRad3ΔTrimer complex. The gel indicates successful assembly of a largely stoichiometric complex, containing 1.5-2x excess core-TFIIH$^{\Delta Rad3}$. TFIIH subunits (Ssl1, Ssl2, Tfb1, Tfb2, Tfb4, Tfb5) are highlighted. **b,** Representative negative stain EM micrograph. A scale bar is provided. **c,** 3D-reconstruction of the PICΔRad3ΔTrimer complex. Comparison of the obtained EM density (grey) with the cryo-EM map of the cPIC-cMed complex (green, filtered to 30 Å) reveals additional density. Complexes are viewed from the front[255].

5.3 Reconstitution and structural characterization of transcription pre-initiation complexes with TFIIH and Mediator

The results presented in this section were published in:

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H. & Cramer, P. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551,** 204-209 (2017).

The following text, as well as figures and figure legends reflect the contents of the publication in unaltered form with the exception of literature references, figure and table numbers and their respective references. These elements have been adapted to the format of this work.

Extended Data Fig. 1 Panel a) of the publication was re-assigned to Section 5.1, Fig. 5.5 for clarity of the text flow, as indicated. This item is therefore missing from Section 5.3. Extended Data Fig. 1 Panel b-d) of the publication were re-assigned to Fig. 5.9 a-c.

A detailed list of published items is also provided on page V.

References to supplementary videos and supplementary data were removed, as the respective files are not available with the printed version of this work.

5.3.1 Structures of the PIC and PIC-cMed complex

Thus far, TFIIH was purified in small quantities from natural sources. To overcome this limitation, we prepared both TFIIH modules in recombinant form after co-expressing their subunits (Methods, Fig. 5.5). The two modules contained TFIIH subunits in apparently stoichiometric amounts and could be assembled into the complete 10-subunit TFIIH. Reconstituted TFIIH formed a stable complex with cPIC and cMed. The resulting 46-subunit PIC-cMed complex was subjected to cryo-EM data collection (Methods, Fig. 5.9 a, b). Unsupervised particle sorting led to cryo-EM reconstructions of the PIC and PIC-cMed complex at nominal resolutions of 4.7 Å and 5.8 Å, respectively (Fig. 5.10 a-d).

Secondary structure was visible in maps obtained with RELION[299] after focused refinement on cPIC, TFIIH or cMed. To reconstruct continuous cryo-EM maps from particles with such flexible regions, we developed a computational tool, 'WarpCraft' (Methods). WarpCraft represents maps as pseudo-atomic models and simulates restrained motions between flexible map regions. This avoids the spatial divergence of separate focused refinements, and can make the construction of composite maps obsolete. Thus we obtained cryo-EM maps that revealed highly defined secondary structure throughout the PIC and PIC-cMed complex.
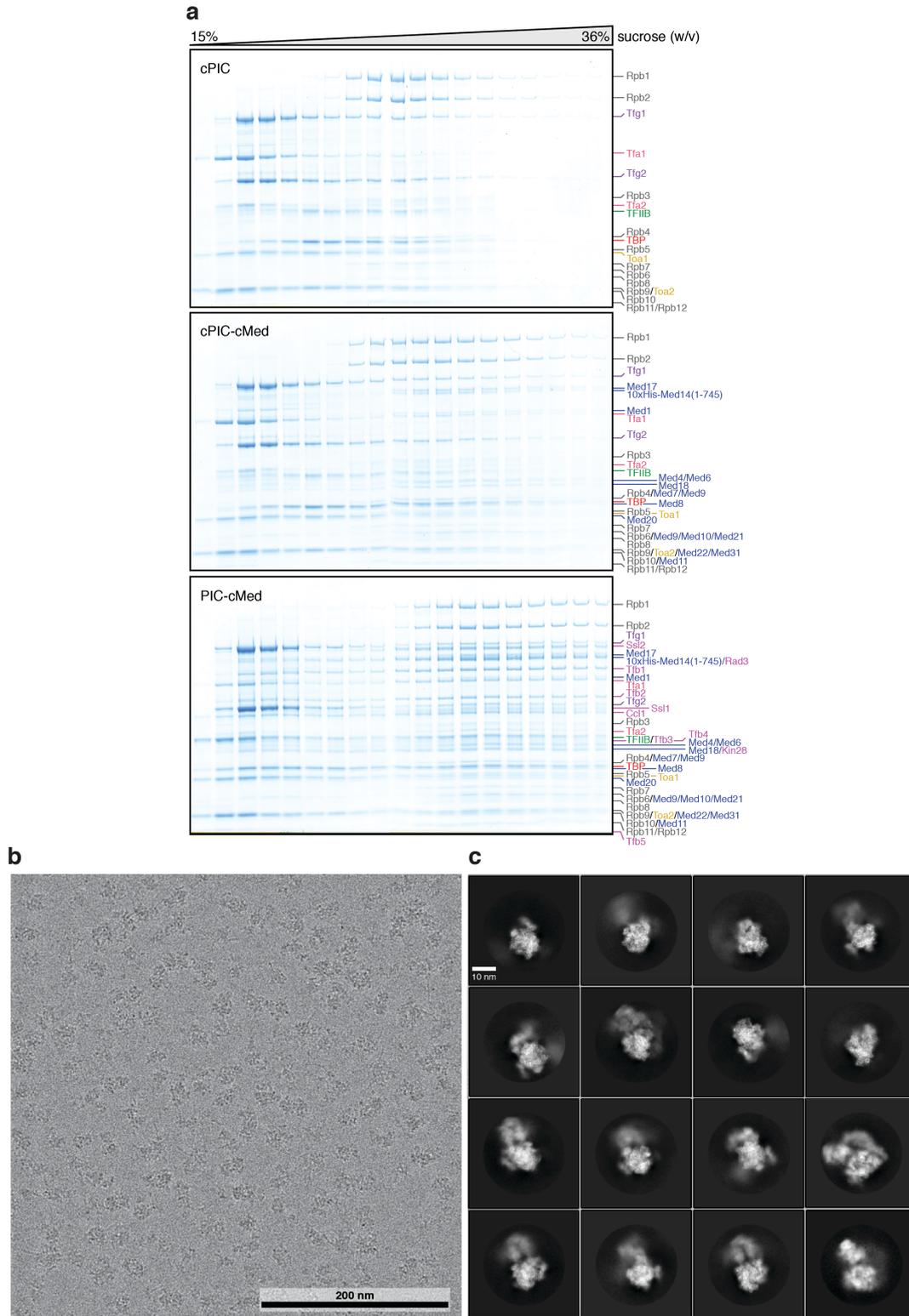
Figure 5.9 | Preparation of the PIC-cMed complex. a, Assembly of complexes. SDS-PAGE analysis of fractions 1-19 of 15-40% sucrose gradient centrifugations (Methods). Labeling of protein subunits according to the color scheme in Figs 5.11 and 5.13. The analysis demonstrates successful formation of the cPIC, cPIC-cMed and PIC-cMed complexes (top to bottom). Bands originating from Pol II, cMed and TFIIH are shifted by several fractions, indicating formation of higher-order complexes. Subunits are present in apparently stoichiometric amounts. This experiment was repeated three times with equivalent results. **b,** Representative cryo-EM micrograph of the PIC-cMed complex. A scale bar is provided. **c,** 2D-class averages reveal 2D reconstructions from particles with clear signal for TFIIH and/or cMed adjacent to the centrally located Pol II density. A scale bar is provided.
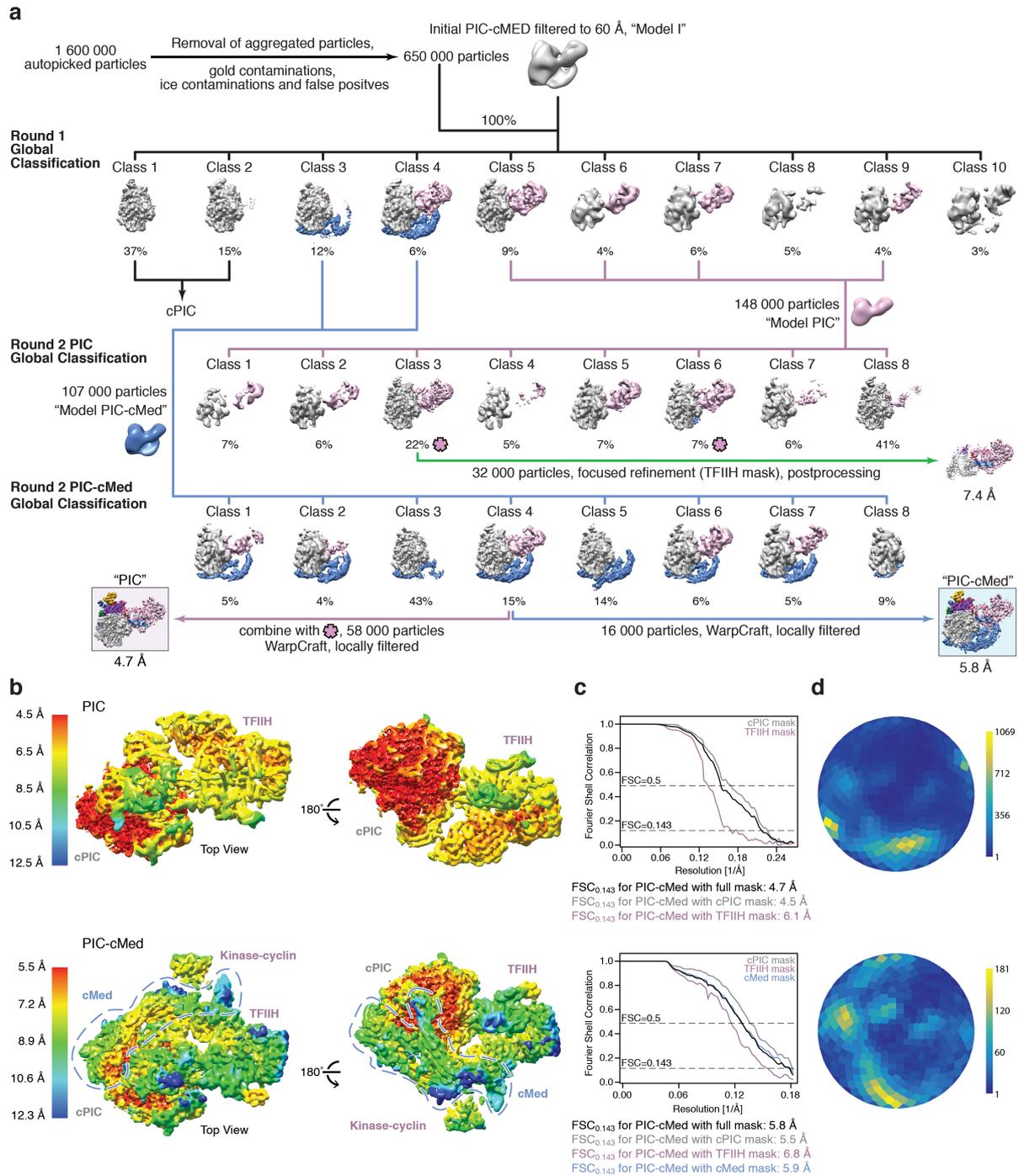
Figure 5.10 | Cryo-EM data processing and quality of reconstructions. a, Particle sorting and classification tree used for 3D reconstruction of the PIC and PIC-cMed complex at nominal resolutions of 4.7 Å and 5.8 Å, respectively. The distinct branches of the classification tree (Methods) are highlighted in pink (PIC) and blue (PIC-cMed). In a conventional focused refinement approach in RELION[299,300], the best-resolved PIC class was reconstructed with a local TFIIH mask, resulting in a focused map with a nominal resolution of 7.4 Å (green branch) that was not deposited. **b,** Two views of the final reconstructions of PIC and PIC-cMed colored according to local resolution[89]. The colour scheme is indicated. **c,** Fourier shell correlation (FSC) between half maps of the final reconstructions of PIC and PIC-cMed. Resolutions for the gold-standard FSC 0.143 criterion are listed. For comparison of distinct regions within PIC and PIC-cMed reconstructions, FSC 0.143 was additionally calculated using local masks. **d.** Angular distribution plot for all particles in the final reconstructions of PIC and PIC-cMed. Color shading from blue to yellow correlates with the number of particles at a specific orientation as indicated.

To solve the PIC structure (Fig. 5.11), we first fitted our cPIC structure[89] to the density and made minor adjustments to TFIIB, the TFIIE subunits Tfa1 and Tfa2, and the Pol II clamp. The PIC adopts the open promoter state with unwound DNA in the active center as before[89]. Structures and models for 22 TFIIH domains were unambiguously fitted to the remaining density (Table 8.2). Eleven connections within and between TFIIH domains were traced and the obtained model was refined by flexible real space fitting (Methods). The TFIIH structure is consistent with 153 known protein-protein crosslinks obtained with bis(sulfosuccinimidyl) suberate (BS3) and 1,1′-(suberoyldioxy)bisazabenzotriazol) (SBAT)[101,190,201], and with additional 55 crosslinks obtained with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (Fig. 5.12 a-d, Tables 8.5, 8.6).
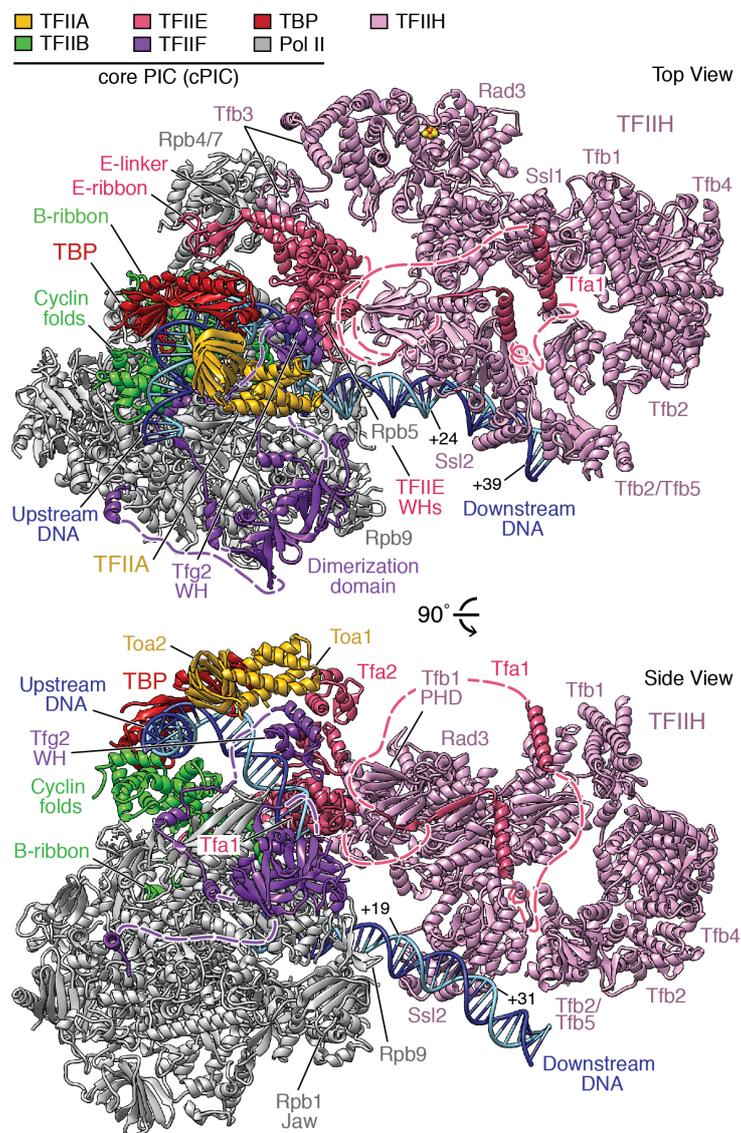
Figure 5.11 | Structure of the Pol II PIC. Two views[255] of the yeast PIC cryo-EM structure. The DNA template and non-template strands are in dark and light blue, respectively. Positions of TFIIH subunits are indicated. Dashed lines represent flexible linkers in TFIIE and TFIIF. The color code is used throughout.
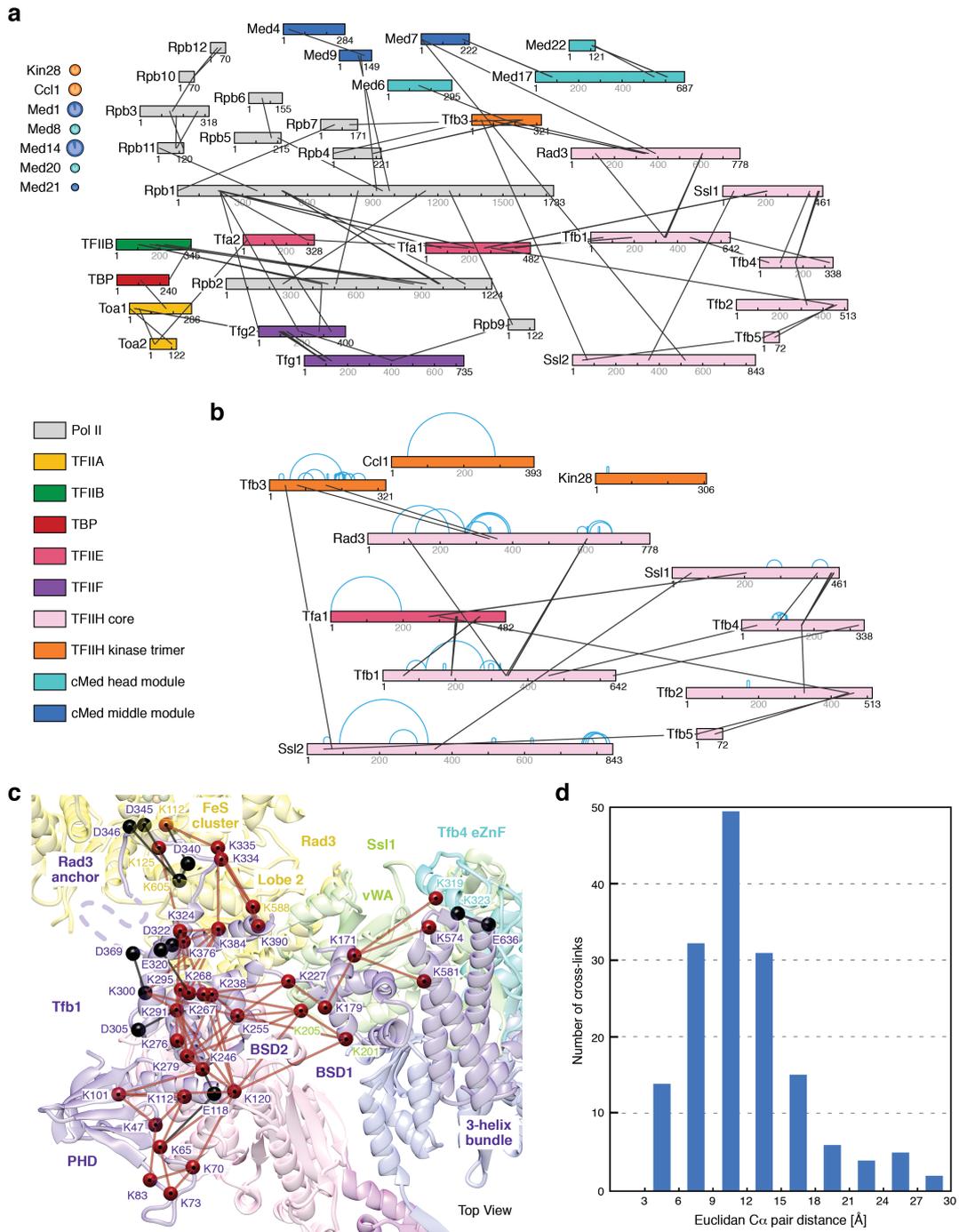
Figure 5.12 | EDC crosslinking analysis of PIC-cMed. a, EDC-derived inter-subunit crosslinks between selected subunits in the PIC-cMed complex. Observed crosslinks are consistent with the structure of the cPIC and with positions of previously reported BS3- and SBAT-crosslinks. Color code as indicated. **b,** EDC-crosslinks observed in TFIIH and between TFIIH and cPIC. Intra- and inter-subunit crosslinks are depicted as blue and black lines, respectively. Crosslinks between the TFIIE Tfa1 C-terminal region and Tfb1, Tfb2 and Ssl1 confirm interactions between TFIIE elements and TFIIH. **c,** Crosslinking hub of the Tfb1 N-terminal region. Ribbon representation of Tfb1 (residues 1-353, 369-394 and 544-639) and the surrounding domains of Rad3, Ssl1 and Tfb4. BS3-/SBAT- and EDC-derived crosslinks are depicted in red and black, respectively. The displayed crosslinks aided modeling of the Tfb1 PHD, BSD1, BSD2 and Rad3 anchor domains into the cryo-EM density. **d,** Statistical analysis of EDC-derived crosslinks. Most observed crosslinks are within a cutoff Cα distance of 16 Å. Cα distances of up to 21 Å may be attributed to flexibility of the involved residues and the coordinate error of the model. Some outliers with Cα distances of 22-30 Å were observed for the well-defined cPIC and Rad3 structures and may have originated from over-crosslinking of particles.

To solve the PIC-cMed structure (Fig. 5.13), we placed the generated PIC model into the PIC–cMed cryo-EM map. We then fitted the remaining density with the *S. cerevisiae* cMed model obtained from the *S. pombe* crystal structure[247]. We obtained a model for the PIC-cMed complex after flexible real space fitting of seven rigid bodies in cMed and manual adjustments (Methods, Table 8.3). The DNA path is virtually identical in both new structures, highly similar in the yeast open cPIC[89], and similar in the human open PIC[86]. The obtained PIC and PIC-cMed structures consist of atomic models where high-resolution structures were available (76% and 73%, respectively), and of backbone models for other parts of TFIIH and cMed.
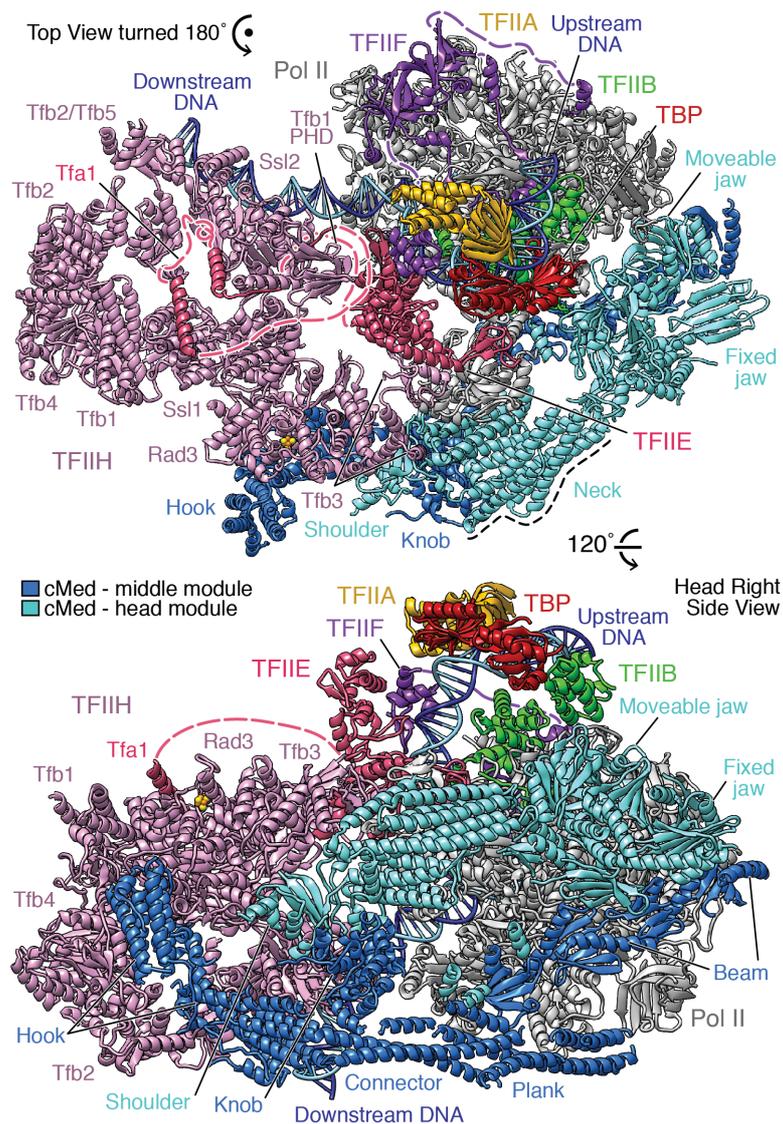
Figure 5.13 | Structure of the PIC-cMed complex. Two views of the PIC-cMed cryo-EM structure. The first view is rotated by 180° compared to the top view in Fig. 5.11. The second view is obtained by a 120° rotation around a horizontal axis. Mediator submodules within the head (blue) and middle (cyan) modules are indicated.

5.3.2 TFIIH structure

The PIC structure reveals that the TFIIH core forms a crescent-shaped complex spanning from Ssl2 to Rad3 (Fig. 5.14 a, b, Fig. 5.15 a-k). Ssl2 binds downstream DNA as previously observed[86,87,109], consistent with its role in DNA opening[103]. Rad3 is located approximately 40 Å away from DNA, in agreement with its ATPase activity being dispensable for transcription[105]. The TFIIH subunits Tfb5, Tfb2, Tfb4 and Ssl1 are arrayed in between the two ATPases. The Tfb1 subunit meanders along Tfb4, Ssl1 and Rad3 and its plekstrin homology domain (PHD) protrudes from the crescent towards the Pol II clamp.
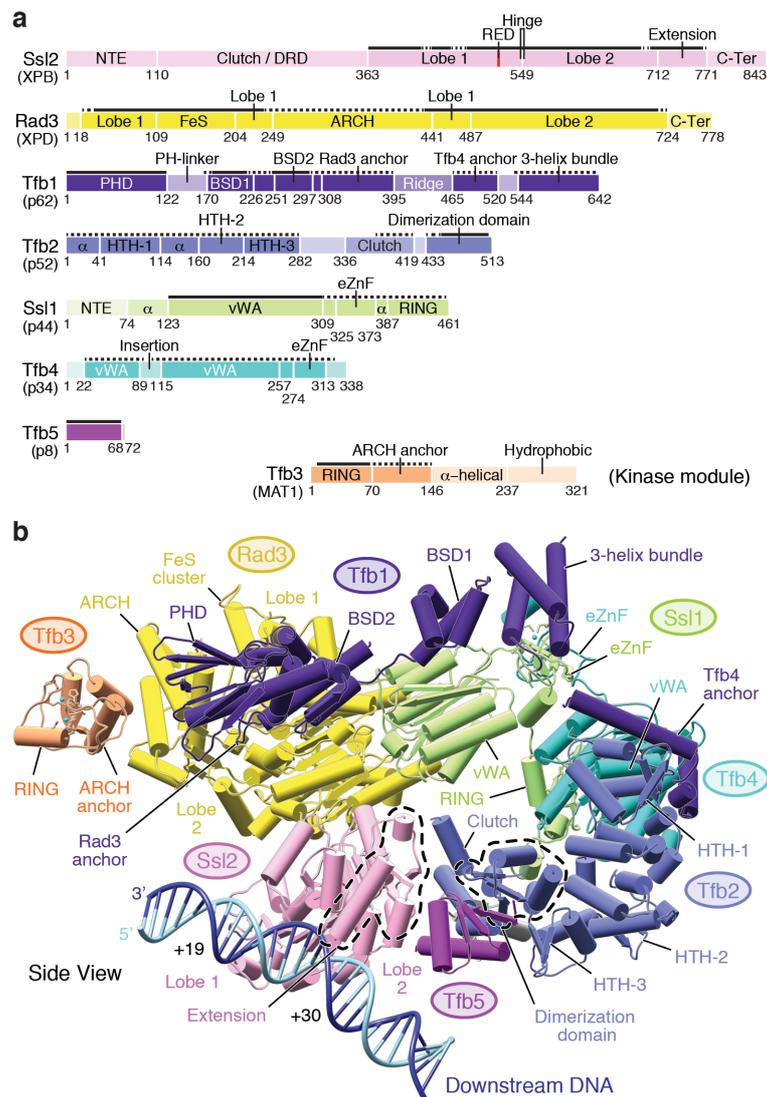
Figure 5.14 | Structure of TFIIH. a, Domain organization of yeast TFIIH subunits except Kin28 and Ccl1. Names of corresponding human subunits are in parentheses. Residue numbers are given for domain borders. Color saturation scales with the percentage of residues modeled as atomic or backbone structures (solid and dashed black bars, respectively). The highlighted RED motif is essential and strictly conserved throughout the XPB family. DRD, damage recognition domain; HTH, helix-turn-helix; NTE, N-terminal extension. **b,** TFIIH structure in cylindrical representation viewed from the side (Fig. 5.11). The DNA register with respect to the putative TSS +1 is indicated.
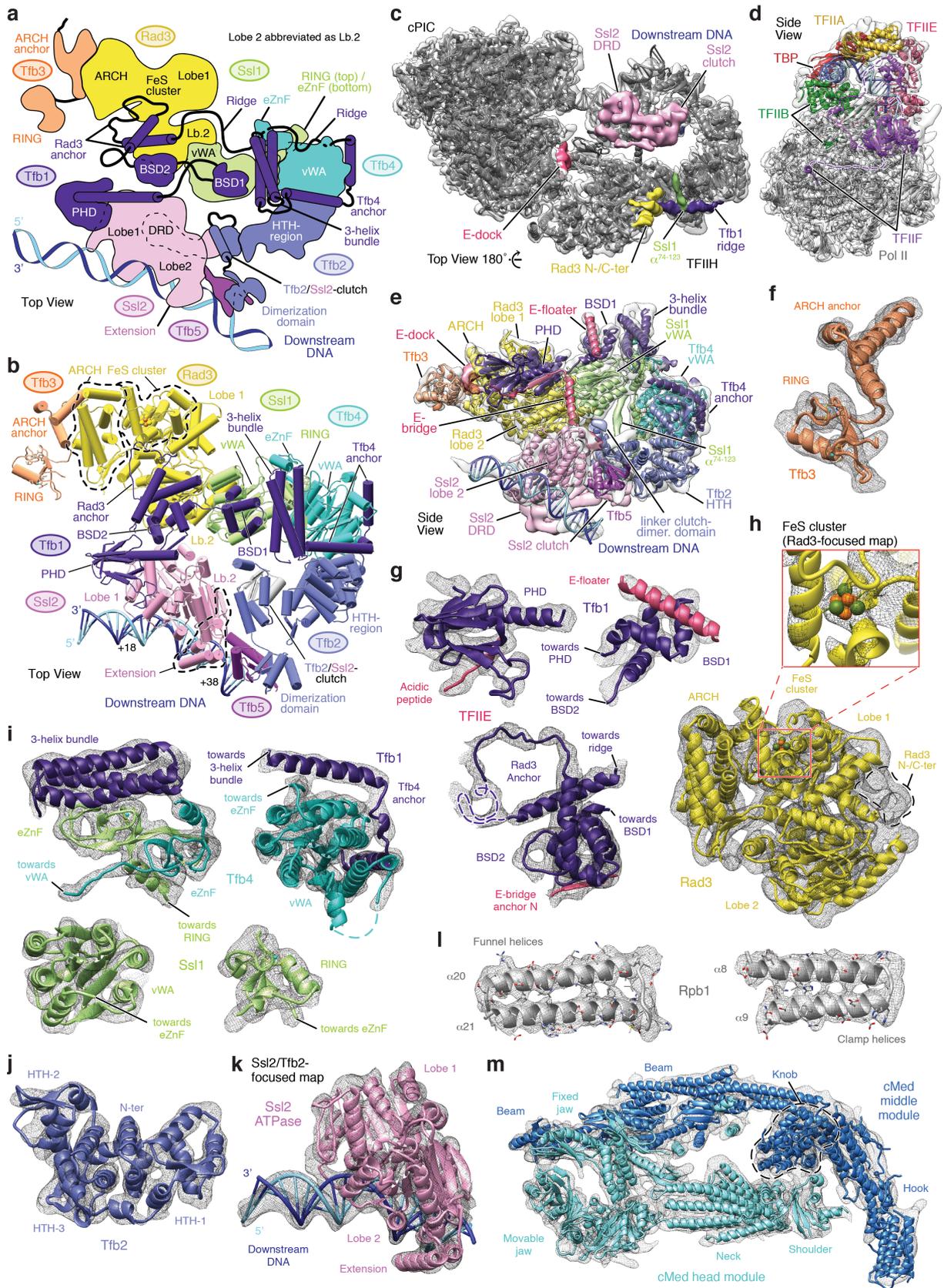
Figure 5.15 | TFIIH structure and quality of the cryo-EM density. See next page for legend.

◄ **Figure 5.15 | TFIIH structure and quality of the cryo-EM density. a,** Schematic of TFIIH subunit and domain architecture with bound double-stranded DNA (dsDNA) using the top view. Flexible linkers are depicted as black lines. Prominent helices within the folds of the tethering subunit Tfb1 and in Tfb2 are highlighted. **b,** Top view of the TFIIH structure in cylindrical representation. Prominent domains are labeled. The DNA register with respect to the putative transcription start site +1 is indicated. **c,** Overall fit of PIC structure into final WarpCraft PIC reconstruction. Observed density for a few remaining regions that could be clearly assigned but were not modeled are highlighted as indicated in Table 8.2. **d,** Fit of cPIC structure into final WarpCraft PIC reconstruction at a higher contour level than in **c** shows the high resolution of the map in this region. **e,** Fit of TFIIH model into final WarpCraft PIC reconstruction. EM map reveals secondary structure throughout. Observed density for regions that could be clearly assigned but were not modeled are highlighted (compare with Table 8.2). **f-k,** EM density (black mesh) for domains and subunits of TFIIH reveals secondary structure throughout. Loops and linkers were traced when continuous density between unambiguously placed models was observed. Depicted density is part of either the WarpCraft PIC reconstruction or a focused reconstruction with a local mask on TFIIH core unless indicated otherwise. **l,** Cryo-EM reconstruction of the PIC reveals side-chain density in well-ordered regions. Depicted are helical regions in the large Pol II subunit Rpb1. **m,** Fit of the PIC-cMed model into the final WarpCraft PIC-cMed reconstruction. Structures of cMed head and middle modules account for density within this region.

The TFIIH core structure shows that the bilobal Ssl2 ATPase contains a C-terminal extension in lobe 2 that contacts Tfb5 in the Tfb2-Tfb5 dimerization module[192]. Ssl2 and Tfb2 interact via newly observed and partially modeled 'clutch' domains. Tfb2 further contains a region with three helix-turn-helix subdomains that binds Tfb4, which comprises a van Willebrandt (vWA) domain with an insertion and an extended zinc-finger (eZnF) domain. Like Tfb4, Ssl1 contains vWA and eZnF domains, and an additional RING domain. Tfb4 and Ssl1 interact intimately and form the backbone of TFIIH. Ssl1 also binds Rad3, a bilobal ATPase with two insertions in lobe 1, an iron-sulfur (FeS) cluster, and an ARCH domain. Whereas the FeS cluster resembles that in homologous archaeal structures[174,175,208], the ARCH domain contains an additional helix and two helix extensions. Tfb1 comprises an N-terminal PHD, two BTF2-like, synapse-associated and DOS2-like (BSD) domains[326], helical regions that anchor Rad3 and Tfb4 (Rad3 anchor and Tfb4 anchor, respectively), and a C-terminal 3-helix bundle that binds the two eZnF domains.

Our TFIIH structure defines the orientation of eight domains in TFIIH subunits that were inferred by previous studies of the PIC[86,87,101]. It also reveals 15 additional domains, numerous connections, and details of domain interactions. Regions in TFIIH subunits that are essential for cell viability in yeast[216] tend to be ordered in our structure (Fig. 5.16 a). The TFIIH structure also suggests the effect of mutations in human TFIIH subunits p8, XPB and XPD that are associated with the human diseases xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome[148,160,161]. Many of the mutated sites are predicted to destabilize the TFIIH core structure (Fig. 5.16 b).
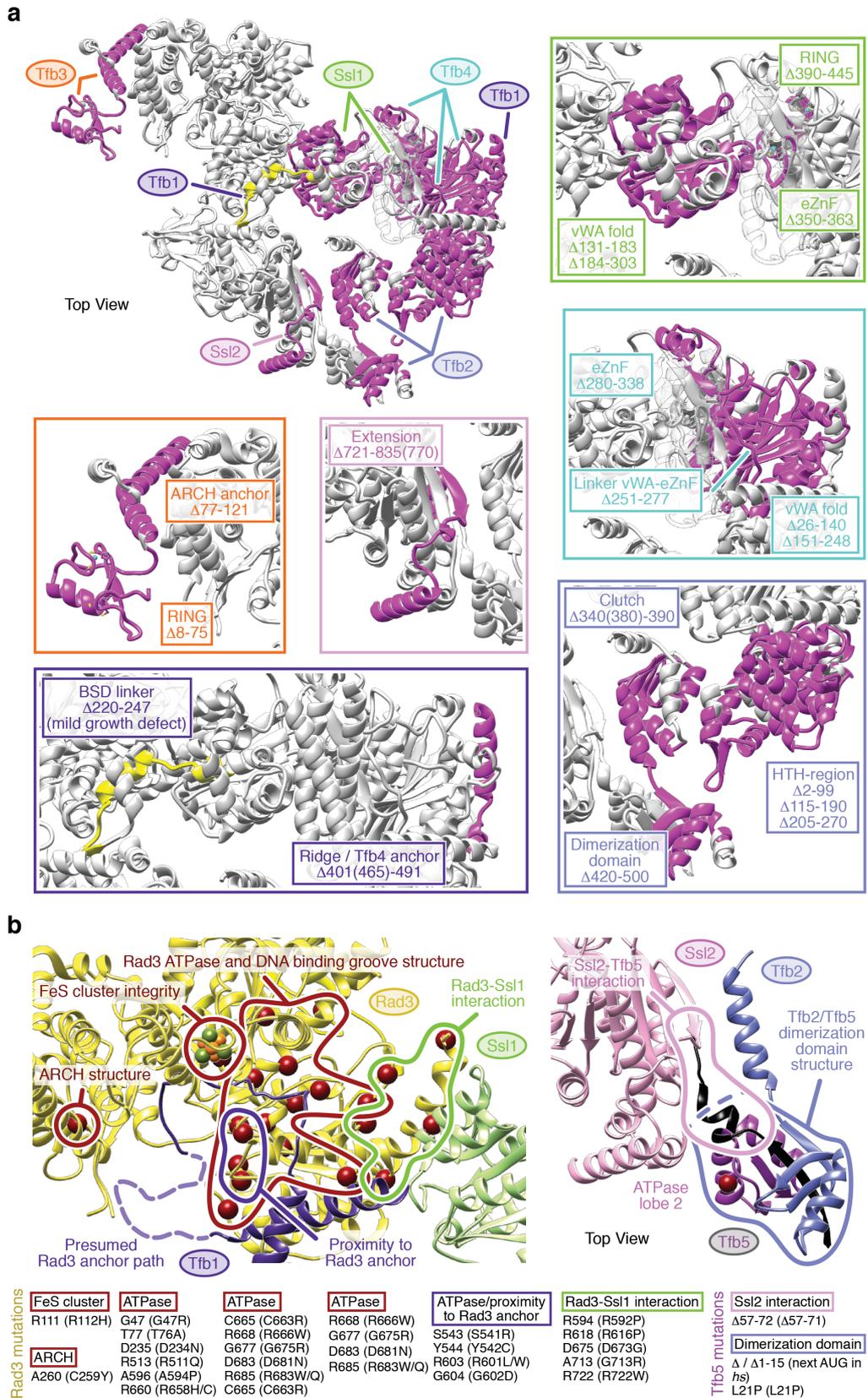
Figure 5.16 | Location of essential regions in TFIIH and sites mutated in disease. See next page for legend.

◄ **Figure 5.16 | Location of essential regions in TFIIH and sites mutated in disease. a,** TFIIH regions essential for cell viability in yeast. Mapping of TFIIH regions identified to be essential in *S. cerevisiae* by *in vivo* deletion analysis[216] on the PIC structure revealed that they are generally forming well-ordered regions of the TFIIH core. Structures are viewed from the top (Fig. 5.11) with regions colored in magenta or yellow if their removal caused cell lethality or growth defects, respectively. Affected TFIIH subunits and ranges of deleted residues are highlighted in colors according to Fig. 5.14. For deletions exceeding the modeled residue range, the last modeled residue is indicated in parentheses. **b,** Mapping of human disease mutations onto the structures of Rad3 (human XPD) and Tfb5 (human p8). Reported mutations in xeroderma pigmentosum, trichothiodystrophy or Cockayne syndrome[148,160,161] were included. The sites of point mutations are depicted as red spheres, and Tfb5 truncations are colored in black. Color coding of TFIIH subunits as in Fig. 5.14. A list of yeast residues highlighted in the PIC structure is provided together with the corresponding human mutations in parentheses. Mutation sites are conserved. Rad3 mutations apparently interfere either with the stability and/or the function of the ATPase core or with the Rad3-Ssl1 interaction. Only few mutations target the FeS cluster or ARCH domain. Newly available data on the Rad3 anchor in Tfb1 suggest close proximity to at least four mutation sites that may affect the Rad3-Tfb1 interaction in this region. Tfb5 mutations either abolish Ssl2 binding or the formation of the dimerization domain with the Tfb2 C-terminus, resulting in destabilization of the Ssl2/Tfb2 region. If the clutch domains remain intact, however, a complete disruption of the Ssl2/Tfb2 interaction seems unlikely. We omitted Ssl2 from analysis as our structure does not cover the region in which reported mutations occur.

5.3.3 TFIIH interactions with cPIC

The PIC structure reveals four sites of interaction between TFIIH and cPIC (Fig. 5.17 a, b). First, the TFIIH kinase module subunit Tfb3 bridges between the Pol II stalk subcomplex Rpb4-Rpb7, TFIIE and Rad3 (Fig. 5.18 a). In particular, the Tfb3 RING domain binds between the Rpb7 OB domain and the TFIIE E-linker helices, and the Tfb3 'ARCH anchor' contacts the Rad3 ARCH domain. This is consistent with the known interaction between the TFIIH kinase module and Rad3[105,164] and the initiation function of Rpb4-Rpb7[327], which also binds TFIIE[89] and cMed[88]. The Tfb3 contact with Pol II further explains why the PIC recruits TFIIH that contains the kinase module, rather than only core TFIIH[166].
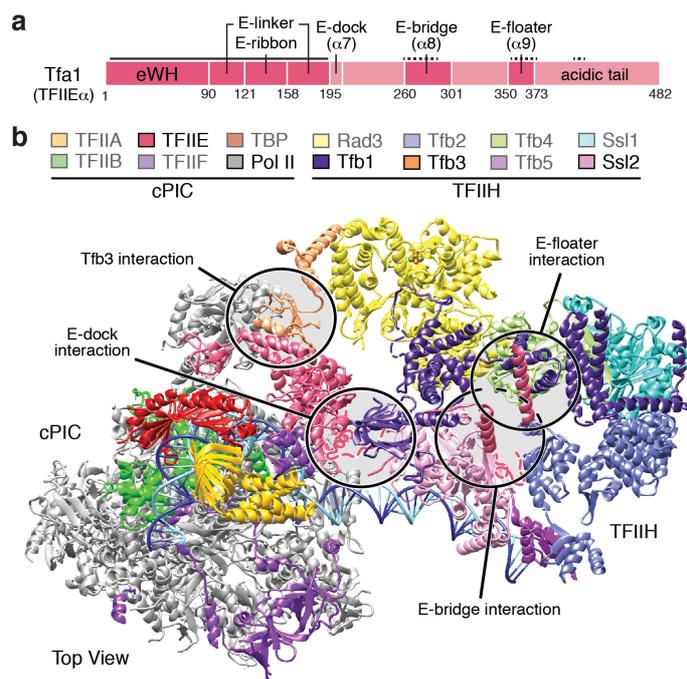
Figure 5.17 | Interactions of TFIIH with cPIC. a, Domain organization of TFIIE subunit Tfa1 (human TFIIEα) including the previously unassigned helices α7 (E-dock), α8 (E-bridge) and α9 (E-floater). Solid or dashed bars refer to protein residues modeled as atomic or backbone structures, respectively.
b, TFIIH-cPIC interactions. PIC is viewed from the top (Fig. 5.11). Regions involved in the formation of the four interfaces are encircled. The color code of cPIC and TFIIH subunits highlights components that participate in the interaction.
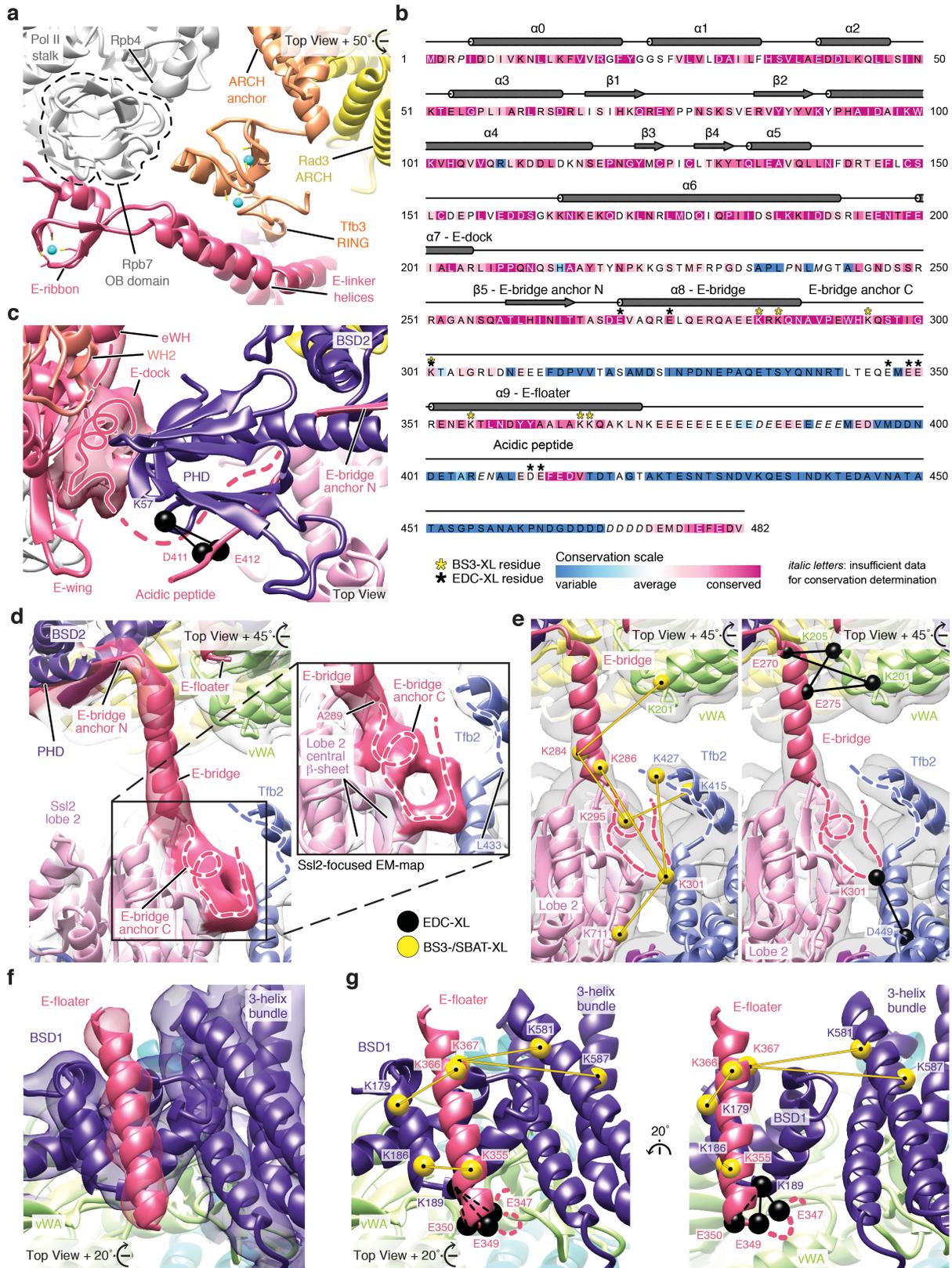
Figure 5.18 | TFIIE-TFIIH interactions. See next page for legend.

◄ **Figure 5.18 | TFIIE-TFIIH interactions. a,** Tfb3-Pol II interaction. The TFIIH kinase module subunit Tfb3 (human MAT1) tethers Pol II and the TFIIH core together. Ribbon representation of the Tfb3 N-terminal RING-finger binding in a groove between the Pol II stalk subunit Rpb7 and the TFIIE E-linker helices. The RING-finger is linked to the ARCH anchor which binds the ARCH domain of Rad3. **b,** Secondary structure and conservation of TFIIE subunit Tfa1 as determined with CONSURF[329]. Regions observed in the PIC and PIC-cMed structures are exceptionally well conserved throughout evolution. C-terminal residues with used crosslinks are indicated. **c,** E-dock. The predicted Tfa1 helix α7 is wedged between the TFIIE extended winged helix (eWH) domain situated on the Pol II clamp and the PHD of Tfb1 in the TFIIH core. α7 was not modeled owing to weak density at the interface of the two major mobile parts of the PIC structure (cPIC and TFIIH) and owing to the absence of crosslinks (Methods). The Tfb1 PHD is additionally contacted by the Tfa1 C-terminal acidic region. The identity and directionality of this acidic peptide were unambiguously established by crosslinking (Methods). **d, e,** E-bridge. This helix (α8) extends from the Tfb1 BSD2 domain at the center of the TFIIH crescent to the central β-sheet of the Ssl2 ATPase lobe 2. The C-terminal anchor peptide (dashed line) was not modeled into the density due to limited resolution. The identity and directionality of the E-bridge was unambiguously established by independent crosslinking experiments (Methods). **f, g,** E-floater. The Tfa1 helix α9 is positioned by the BSD1 domain of Tfb1 and located adjacent to the 3-helix bundle at the center of the TFIIH crescent. The identity and directionality of the E-floater was unambiguously established by independent crosslinking experiments (Methods).

A role for Tfb3 in TFIIH recruitment can also explain why the kinase module is required for transcription initiation in a reconstituted system[119] although its kinase activity is not[328]. The C-terminal part of Tfb3 is disordered and connects to the kinase-cyclin pair, which is also mobile in the PIC structure.

The three additional interactions between TFIIH and cPIC involve the mobile C-terminal region of TFIIE subunit Tfa1 (human TFIIEα). This TFIIE region forms three previously unobserved helices that are flexibly connected and named here E-dock (α7), E-bridge (α8) and E-floater (α9) (Fig. 5.18 b). The E-dock apparently enables docking of the Tfb1 PHD to the TFIIE extended winged helix domain that is located on the Pol II clamp (Fig. 5.18 c). The E-bridge extends from Tfb1 domain BSD2 to the Ssl2 lobe 2 (Fig. 5.18 d, e). The E-floater binds the BSD1 domain in Tfb1 (Fig. 5.18 f, g). Taken together, these contact sites explain why TFIIE is required for TFIIH recruitment to the PIC[100].

5.3.4 TFIIH and DNA opening

The PIC structure shows that the Ssl2 ATPase engages with promoter DNA approximately 25-30 base pairs (bp) downstream of the putative transcription start site (TSS) +1 (Fig. 5.19 a, b, Fig. 5.20 a, b, f). This location is consistent with crosslinking data[96] and previous cryo-EM studies[86,87], and with the translocase model for ATP-dependent DNA opening[108,109]. According to this model, Ssl2 uses ATP hydrolysis to translocate on DNA away from Pol II. If the Ssl2 location is fixed, Ssl2 action results in a reeling of DNA into the active center. The PIC structure supports a fixed location of Ssl2 and the proposed directionality of translocation. The two ATPase lobes bind the DNA backbones on both sides of the minor groove, similar to the ATPase in the chromatin remodeling enzyme Chd1[291]. Comparisons with Chd1 and with ATPase structures of NS3 and Rad3 (Fig. 5.20 d, e) indicate that Ssl2 tracks along the DNA template strand in the 3′-5′ direction, consistent with biochemical studies[111,172,330]. One study suggested that tracking occurs on the non-template strand in 5′-3′ direction[109], but this would result in the same overall movement.
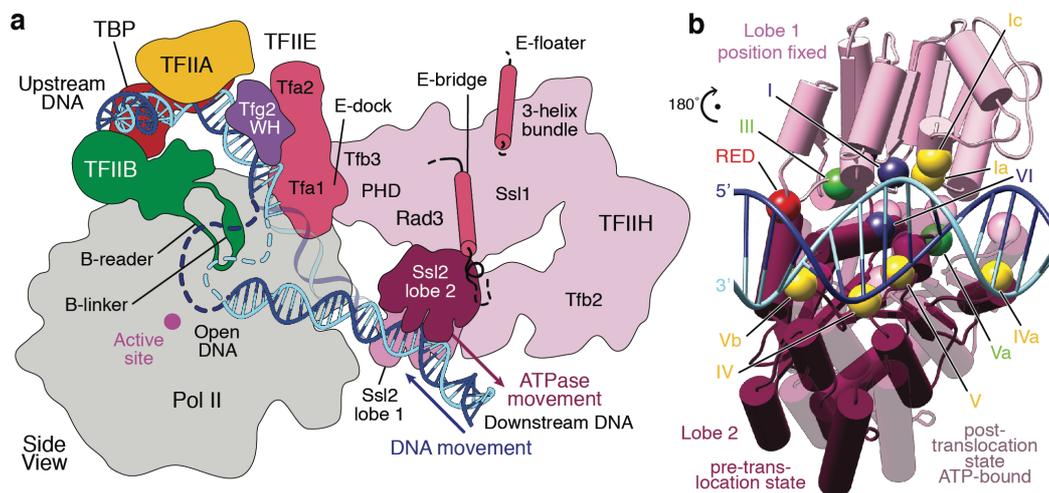
Figure 5.19 | TFIIH and DNA opening. a, Schematic cross-section of the PIC with open and closed DNA viewed from the side. PIC elements involved in DNA opening are depicted. Color coding as in Fig. 5.11 except for Ssl2 lobe 1 (pink) and lobe 2 (burgundy). The Ssl2 ATPase translocates to the right and DNA moves to the left during DNA opening. **b,** Putative ratcheting of lobe 2 in the Ssl2 ATPase with respect to lobe 1. The PIC structure reveals the pre-translocation state (no ATP bound). The post-translocation state of lobe 2 was modeled by superposition of Chd1 (PDB code 5O9G). Helicase motifs are indicated (Fig. 5.20).

The PIC structure also suggests how TFIIE may stimulate the ATPase activity of TFIIH[110]. According to the current model for ATPase translocation[291,331], ATP binding induces a ratcheting movement of lobe 2 with respect to lobe 1, and a DNA translocation by one base pair. In our structure, we trapped the pre-translocation state of Ssl2 with an empty ATPase active site (Fig. 5.19 b). The C-terminal end of the TFIIE E-bridge contacts the Ssl2 lobe 2, suggesting that the E-bridge can influence the conformational ratcheting in the Ssl2 ATPase that occurs during DNA translocation.
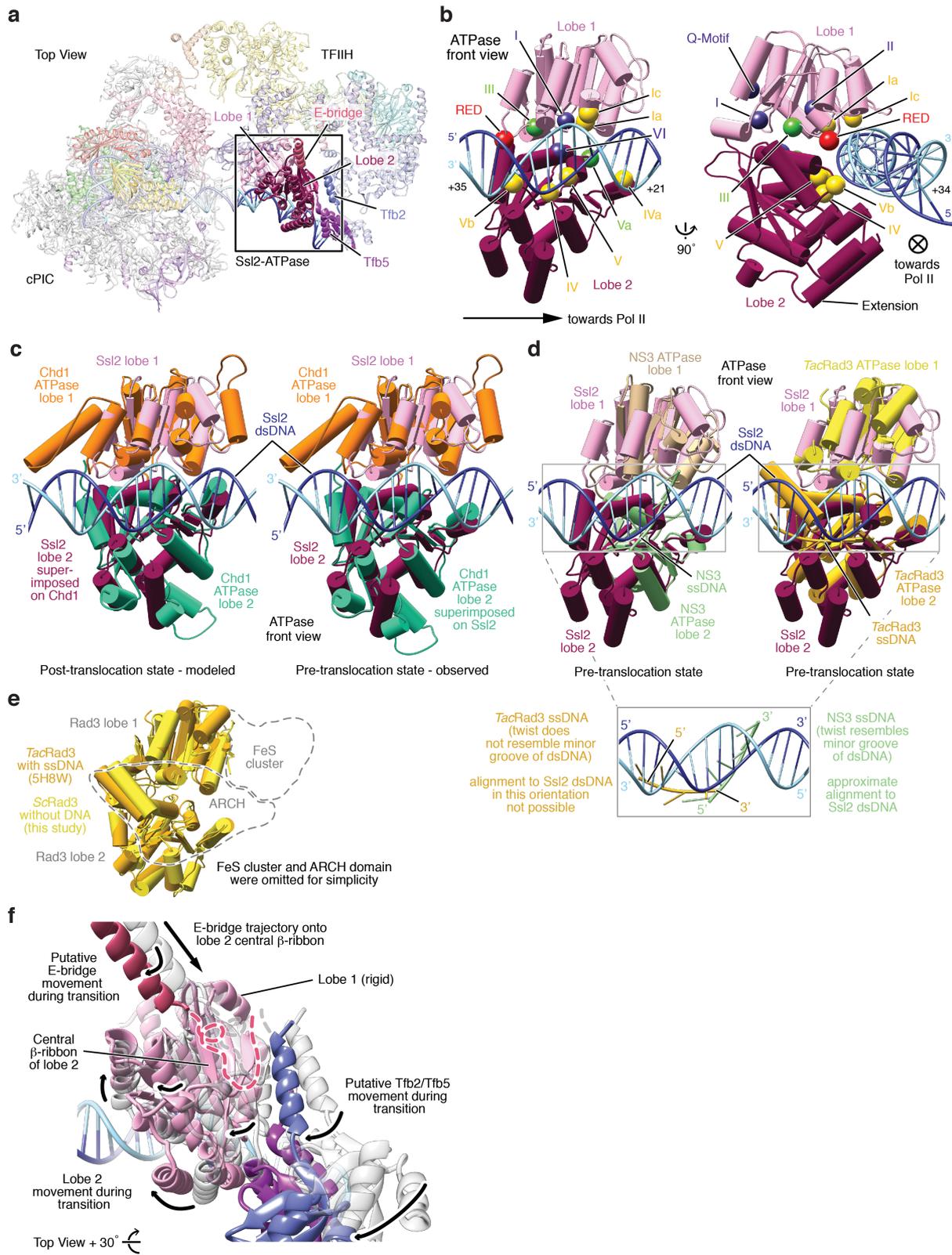
Figure 5.20 | Detailed analysis of Ssl2 ATPase conformation and implications for translocase activity. See next page for legend.

◄ **Figure 5.20 | Detailed analysis of Ssl2 ATPase conformation and implications for translocase activity. a,** Overview of PIC complex with highlighted Ssl2 (human XPB) ATPase lobes 1 and 2 (in pink and burgundy, respectively) and interacting domains of Tfb2, Tfb5 and Tfa1. **b,** Detailed view on Ssl2 positioned on dsDNA in the presumed pre-translocation state. The ATP analogue AMP-PNP was present in the buffer but was not observed in the active site of the Ssl2 ATPase, supporting the model that we trapped the structure in the pre-translocation state. Register of covered nucleotides with respect to the putative TSS +1 is indicated. Highlighted helicase motifs were identified and assigned as described[171]. Yellow colored motifs are involved in the DNA interaction, purple motifs participate in NTP binding and hydrolysis, and green motifs are involved in coupling of ATP hydrolysis to DNA binding. Both lobes of the ATPase contact both nucleic acid strands. **c,** Chd1 and Ssl2 ATPases are closely related on a structural level and share the same fold. The presumed post-translocation state of Ssl2 was modeled by separate alignment of ATPase lobe 1 and 2 to the respective lobes in the structure of Chd1 bound to an ATP analogue (PDB code 5O9G); the presumed pre-translocation state was modeled vice versa using the Ssl2 structure as reference model. In both states the structures overlap to a high degree. **d,** The Ssl2-DNA arrangement observed in the PIC structure resembles that of 3′-5′-directed rather than 5′-3′-directed members of the SF2 family. Superposition of the Ssl2-dsDNA structure with models of the NS3 (PDB code 3KQK)[332] and *T. acidophilum* (Tac) Rad3 (PDB code 5H8W)[175] ATPase domains reveals a closer resemblance of Ssl2 to the 3′-5′-helicase NS3. Additionally, the bound single-stranded (ss) DNA fragment in the NS3 model aligned well to the dsDNA in the Ssl2 structure whereas the bound fragment in the *Tac*Rad3 structure was positioned differently and did not exhibit a minor groove twist as observed for NS3 and Ssl2 in the respective position. **e,** Superposition of structures of *Tac*Rad3 and *Sc*Rad3 ATPase domains indicates very high level of structural homology. ATPase lobes 1 and 2 were superimposed separately to account for the absence of bound DNA in the *Sc*Rad3 structure. **f,** Putative movement of E-bridge and the Tfb2-Tfb5 dimerization domain upon Ssl2 transition from the presumed pre- to the presumed post-translocation state (grey and color, respectively). Upon movement of lobe 2, the E-bridge may undergo a rotation-translation movement towards Pol II and against its own trajectory onto the central β-ribbon of the Ssl2 ATPase lobe 2. The flexible Tfb2-Tfb5 dimerization domain would swing towards Pol II.

5.3.5   TFIIH and Pol II phosphorylation

The PIC-cMed structure provides details on the previously described PIC-Mediator interfaces[88], and suggests conformational changes in Mediator upon PIC binding (Fig. 5.21 a, b, Fig. 5.22 a-e). The Mediator head module is largely unchanged[248], but the conformation of the middle module differs from that in the cMed structure[247] (Fig. 5.22 c). The submodules in the middle module apparently undergo concerted movements. Whereas the plank rotates to bind the Pol II foot, the hook and knob undergo swinging motions and the beam moves towards the head module jaws. Comparison with the cMed cryo-EM structure[234] also suggests conformational changes in Mediator upon PIC binding.

The PIC-cMed structure further reveals an additional density for the Kin28-Ccl1 kinase-cyclin pair on the outer surface of cMed (Fig. 5.21 a, b). This density is located above one of two openings that flank the knob at the Mediator head-middle interface. The kinase-cyclin pair resides between the Mediator hook, knob and shoulder, roughly consistent with its previously reported position[101]. The density for Kin28-Ccl1 is weaker than the density for cMed or TFIIH, indicating that the kinase-cyclin pair retains some mobility.

How the TFIIH kinase reaches its phosphorylation substrate, the Pol II CTD, was unclear. The linker to the mobile CTD extends from Pol II towards the inner surface of Mediator that lines a previously described cradle formed between Mediator and Pol II[88] (Fig. 5.21 a, b). To reach the kinase, the CTD may exit the cradle and extend around Mediator or through Mediator[101]. However, the CTD crosslinks to the inner surface of the cradle[247], suggesting that it resides in the cradle, where it can be accommodated if it adopts a compact globular shape[255].
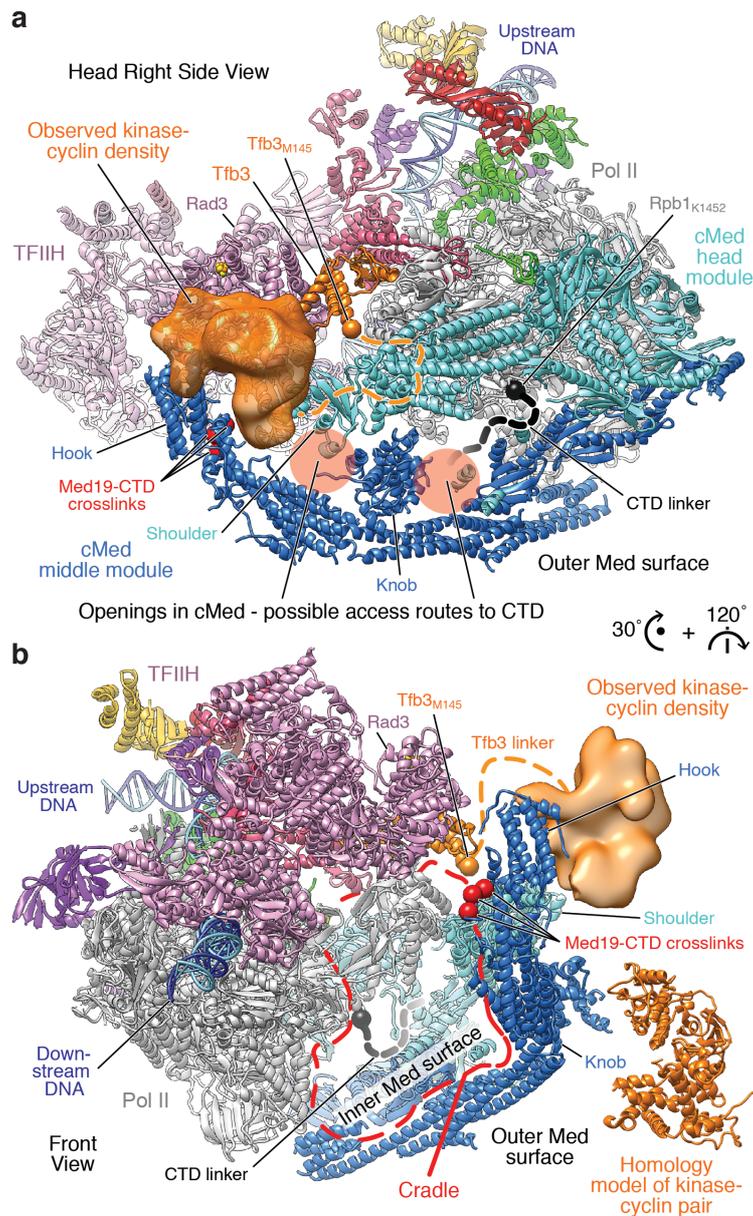
Figure 5.21 | TFIIH and phosphorylation of Pol II. a, PIC-cMed structure as in Fig. 5.13 but with additional cryo-EM density for the mobile TFIIH Kin28-Ccl1 kinase-cyclin pair (orange, filtered to 15 Å). An orange sphere depicts the last modeled residue in the Tfb3 linker to the kinase-cyclin pair (Met145). A black sphere depicts the last ordered residue in the Rpb1 linker to the CTD (Lys1452). Red spheres depict Med19 residues that crosslink to the CTD C-terminal end. Filled red circles indicate two openings at the Mediator head-middle interface. **b,** The same structure viewed from the front into the cradle between Pol II and Mediator (red outline). A model for the kinase-cyclin pair is shown for size comparison in an arbitrary position.

The TFIIH kinase may access the CTD through the openings at the head-middle interface. Phosphorylation of CTD regions would then lead to repulsion between accumulating negative charges, expansion of the CTD globule in the cradle, a weakening of the Pol II-Mediator interaction and Mediator dissociation. Loss of Mediator destabilizes the PIC and would facilitate Pol II escape from the promoter.
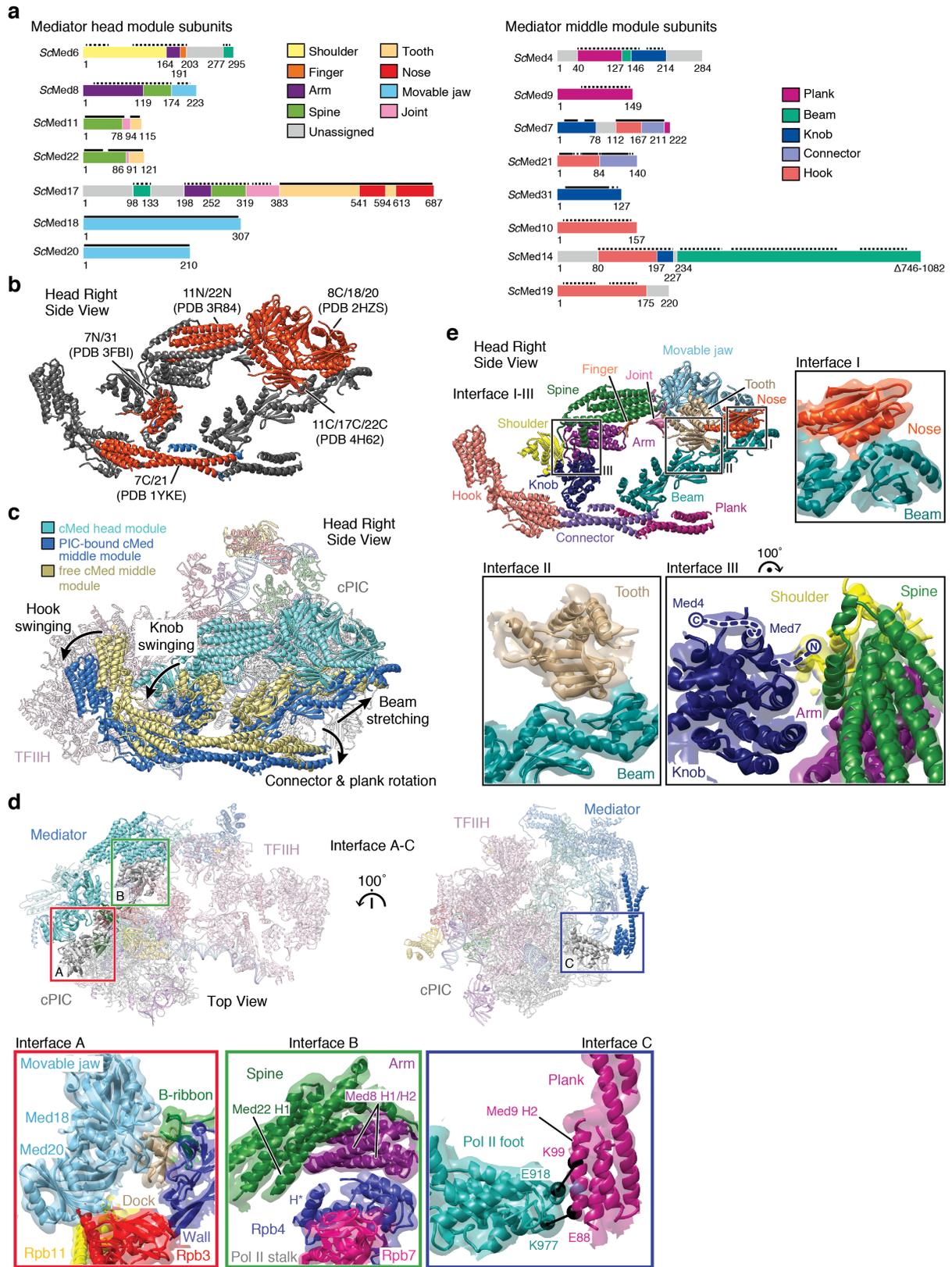
Figure 5.22 | Structure and conformational changes of cMed. See next page for legend.

◄ **Figure 5.22 | Structure and conformational changes of cMed. a,** Schematic representation of cMed subunits. Regions contributing to submodules are colored as in the *S. pombe* cMed crystal structure[247]. Solid and dashed black lines refer to protein regions that were modeled as atomic or backbone models, respectively. **b,** Ribbon model of cMed colored by type of structural model used for interpreting the cryo-EM density. Regions with backbone models based on the *S. pombe* cMed structure[247], regions with atomic models inclusive of the PDB code, and *de novo* modeled regions are indicated in grey, orange and blue, respectively. **c,** Repositioning of the cMed middle module upon PIC binding. The structures of unbound cMed (khaki, PDB code 5N9J) and PIC-cMed complex (blue, this study) were superimposed on the cMed head module. The positions of the cMed middle module domains hook, knob, connector, plank and beam apparently undergo conformational changes upon PIC binding, as indicated by arrows. This may cause or enlarge two observed openings at the head-middle interface. **d,** PIC-cMed interactions. Structure of the PIC-cMed complex in two views. The three previously identified interfaces[88] between cPIC and cMed are indicated. In interface A, the Mediator movable jaw (light blue) contacts the Pol II Rpb3-Rpb11 heterodimer (red/yellow), the dock domain (beige) and the TFIIB β-ribbon (green). In interface B, the Mediator spine domain (green) contacts helix H* of the Pol II stalk subunit Rpb4 (blue) with its Med22 helix H1, and the Mediator arm domain (violet) contacts Rpb4 with its Med8 helices H1 and H2. In interface C, the Mediator plank domain (pink) contacts the Pol II foot region (cyan) with its Med9 helix H2. Two newly observed EDC-crosslinks between Med9 helix H2 and the Pol II foot domain are indicated by black spheres. **e.** Mediator head-middle module interfaces. In the unbound *S. pombe* cMed X-ray structure, four interfaces (I–IV) were observed between the head and middle modules[247]. Owing to stretching of the beam, interfaces I and II are altered in the PIC-bound cMed structure. In the new conformation, the Med4 C-terminal region in the Mediator knob is flexible and does not contact the spine region (interface III). Interface IV between the shoulder and hook domains is lost. Mediator domains are colored as in **a**.

6   Conclusion

This section was published in:

The following text reflects the contents of the publication in unaltered form with the exception of literature references. These elements have been adapted to the format of this work. A detailed list of published items is also provided on page V.

We have been aiming to achieve detailed structures of the yeast Pol II PIC and its complex with Mediator ever since the structure of the core Pol II enzyme was determined[255]. Important steps towards this goal included the Pol II-TFIIB crystal structure, which led to minimal models of the closed and open promoter complexes[75], and our recent structures of cPIC[89] and cMed[247]. The crucial step reported here was to prepare recombinant TFIIH, to derive its structure, and to arrive at structures of the PIC and the PIC-cMed complex. The PIC-cMed complex lacks TFIID and the Mediator tail module, but their location on the PIC is known from work by others in the human[58] and yeast[101] systems, respectively.

The structures presented here define interactions of TFIIH within the PIC and interactions of cMed with the PIC, and provide unexpected insights. First, anchoring of TFIIH to the cPIC involves a subunit of the TFIIH kinase module, ensuring that complete TFIIH is incorporated into the PIC. Second, a mobile extension of TFIIE tethers several parts of TFIIH, including the Ssl2 ATPase. Third, the TFIIH kinase is mobile in the PIC, but adopts a preferred location on Mediator when cMed binds the PIC. Finally, PIC-bound Mediator contains two openings at its head-middle interface that may allow access of the TFIIH kinase to the Pol II CTD residing in the cradle. The structures thus provide the basis for future mechanistic studies of TFIIE-stimulated and TFIIH-dependent promoter opening, Mediator-stimulated CTD phosphorylation and promoter escape, and gene regulation during transcription initiation.

7 Discussion and Outlook

7.1 A paradigm for the structural dissection of macromolecular multi-subunit assemblies

Ultimately, a comprehensive approach, which included sophisticated techniques in the fields of biochemistry, biophysics and bioinformatics, was pivotal for the determination of the PIC and PIC-cMed structures at high resolution. Recent advances in routine cryo-EM analysis and the respective processing and reconstruction software[258,300-302] were successfully exploited in combination with an optimized sample preparation strategy and powerful modeling suites, which also permitted identification of genetically distant homologues in template searches. Limitations such as the intrinsic flexibility between the TFIIH and cPIC entities in the complex were overcome by the development of innovative computational tools like WarpCraft to obtain a well-resolved EM map for TFIIH. Other challenges such as the derivation and correct positioning of models for single domains within TFIIH, however, would not have been met without the availability of existing structural or crosslink data. In addition, structural information on many regions of the cPIC and cMed complex had been compiled by numerous studies in the preceding years (exemplary in[85-87,101,247,248]. The present work now provided the architecture of the PIC-cMed complex as an ultimate framework, in which this previously gathered information was coalesced and extended with novel findings to generate the complete model of a transcription complex that comprises all proteins essential for productive initiation in an *in-vitro* system[25,94,135] and its most prevalent co-activator.

The versatile strategy applied to generate the PIC and PIC-cMed structures may serve as a paradigm for the implementation of future projects that involve similar, highly intricate samples. In addition, quickly evolving cryo-electron tomography (cryo-ET) techniques may further extend the system to study the respective macromolecular complexes in their native cellular environment. Indeed, related schemes were used to arrive at models for many distinct multi-subunit assemblies involved in transcription and transcriptional regulation, as isolated structures of polymerases and numerous accessory factors are widely available. Future application targets in the field of Pol II encompass not only further complexes that govern initiation, pausing, elongation or termination but also co-transcriptional events such as transcription-coupled repair (TCR) or co-transcriptional splicing, possibly even in the context of chromatin.

7.2 Comparison of free and PIC-bound TFIIH and implications of structural rearrangements upon PIC-association

While the obtained structures of the PIC and PIC-cMed complex were the first to reveal details on TFIIH in the context of the transcription initiation machinery, simultaneously also a high-resolution cryo-EM structure of TFIIH in its free conformation was reported[333]. Although the models derived in the studies describe TFIIH complexes originating from yeast and *H. sapiens*,

respectively, TFIIH subunits are sufficiently well conserved between these organisms to permit their comparison. Owing to the previous lack of detailed structural information, these models provide the first opportunity for a discussion of putative conformational changes in TFIIH during the transition between its free and its cPIC-associated state.

Whereas the global shape of TFIIH deviates significantly in the two models, direct superposition of isolated subunits indicates a high degree of overlap within most regions (Fig. 7.1 a). Minor discrepancies, such as slight displacements or divergent lengths of SSEs and variable linker regions, are presumably a result of species-specific differences. The general domain- and module conformations, however, appear to be well conserved. Consistently, all subunits that were confidently assigned were positioned and oriented equally in both studies, which additionally reflects the high quality of the independently generated TFIIH models. The ATPase lobes of Rad3/XPD and Ssl2/XPB, as well as the Rad3/XPD FeS-cluster and ARCH domains are analogously arranged with respect to each other, in pronounced contrast to the folds observed in the crystal structures of their archaeal homologues[174,175,207,208]. Seemingly, incorporation of the ATPases into TFIIH in the adopted conformations is favored or even exclusive, probably due to their requirement for correct assembly and stability of the holo-complex. In comparison to their pre-translocated state in the PIC model, the ATPase lobes of Ssl2/XPB are differently positioned in the absence of DNA in free TFIIH (Fig. 7.1 b). They apparently undergo a closure upon binding to DNA, which is likely to affect the conformation of the ATP binding pocket and the active center of Ssl2/XPB. The unengaged state therefore may represent a relaxed Ssl2/XPB conformation that is less active and less competent for ATP-binding, consistent with the observed stimulation of the Ssl2/XPB ATPase in presence of dsDNA substrates[109].

The scaffolding factors Ssl1/p44, Tfb2/p52 and Tfb4/p34, which constitute the backbone of core-TFIIH and form the middle of its crescent-like shape, contain the same vWA (Ssl1/p44, Tfb4/p34) and HTH/α-helical solenoid motifs (Tfb2/p52) in yeast and *H. sapiens* TFIIH. However, the free TFIIH model lacks several regions of these subunits, such as parts of the eZnF- and RING-finger domains in Tfb4/p34 and Ssl1/p44 (Fig 7.2). They appear to be more robust and better defined within the PIC architecture and could therefore be assigned with confidence in the PIC and PIC-cMed models. Their flexibility in the free TFIIH structure is possibly a secondary effect of missing sequestration of the Tfb1/p62 PHD by the dock helix in Tfa1/TFIIEα. In the PIC, this interaction seems to have a stabilizing effect on the adjacent BSD1, BSD2 and the 3-helical bundle of Tfb1/p62, which in turn binds to the eZnF domains of Ssl1/p44 and Tfb4/p34, and consequently results in higher rigidity of the entire region. Consistently, also the Tfb1/p62 PHD and BSD1, which in addition lacks its contact with the E-floater helix of Tfa1/TFIIEα, were not detected in the cryo-EM map of free TFIIH. The Tfb1/p62 BSD2 and 3-helical bundle were, albeit observed and partially modeled, not confidently assigned to Tfb1/p62 due to this missing information[333]. In summary, the mapping and model building for large segments of Tfb1/p62 and adjacent small domains of Ssl1/p44 and Tfb4/p34 was feasible for cPIC-bound TFIIH but not for the free conformation owing to a strongly increased flexibility in these peripheral regions (Fig. 7.2). The HTH/α-helical motifs in Tfb2/p52 in the structure of free TFIIH were modeled

incompletely without β-turns and were thus interpreted as a purely solenoid-like arrangement[333] although the corresponding EM-map indicates a similar HTH-fold as observed for the cPIC-associated state. In contrast, the *H. sapiens* p52/p8 dimerization domain, which was modeled based on the yeast crystal structure[192], overlaps well with the template. The EM-maps of free and cPIC-bound TFIIH suggest that the clutch domains of Ssl2/XPB and Tfb2/p52 mediate the primary interaction between these proteins and adopt a fold that resembles the Tfb2/Tfb5 dimerization domain, involving two opposing β-sheets. Due to low resolution in this map region, however, the structure of cPIC-bound yeast TFIIH includes only part of the clutch domains and lacks the N-terminal portion and the DRD-domain of Ssl2, which were modeled, although not completely assigned, for XPB in free TFIIH[333]. Seemingly, cPIC-bound TFIIH misses stabilizing contacts between the Ssl2- and Rad3-halves of the core-TFIIH crescent after its incorporation into the initiation complex (Fig 7.2).
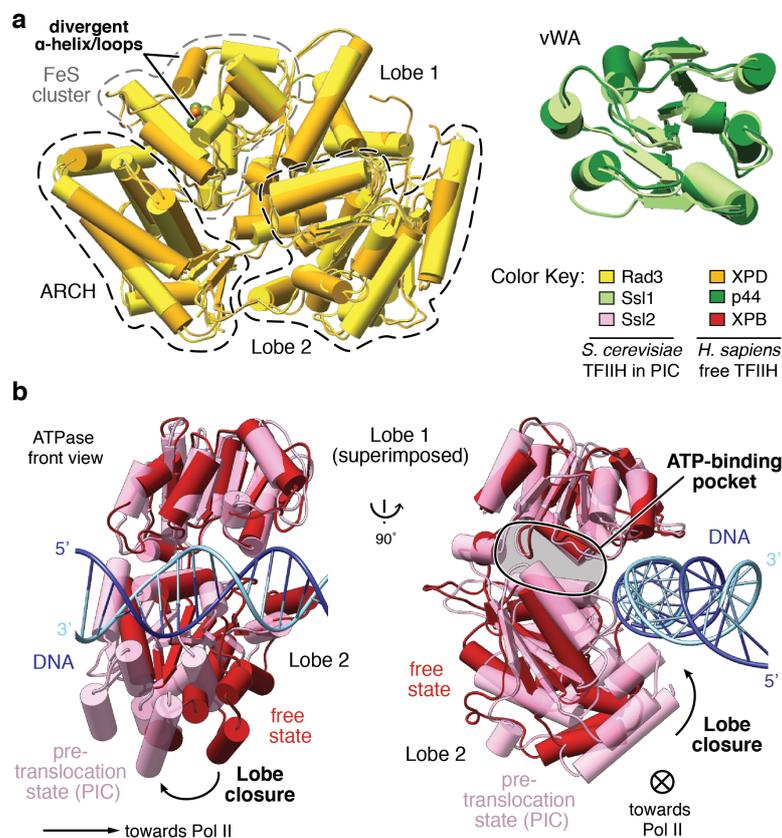
Figure 7.1 | Superposition of subunits within free and PIC-bound TFIIH. a, Subunits of TFIH from *H. sapiens* (PDB code 5OF4) and yeast exhibit highly conserved domain folds. Exemplary superposition of the TFIIH subunit Rad3/XPD and the vWA domain of Ssl1/p44 reveal close resemblance. Structural elements and local folds appear highly similar in both species. Minor species-specific differences between structures are primarily observed in loops or peripheral regions, for example in parts of the Rad3 FeS cluster. Individual folds and domains do not undergo rearrangements during transition from the free to the PIC-bound state of TFIIH. The color code is indicated. **b,** Comparison of the Ssl2/XPB ATPase domains in the free and DNA-engaged state. Structure of XPB (not DNA-associated) in free TFIIH (PDB code 5OF4) was superimposed on Ssl2 in the pre-translocation state using ATPase lobe 1 as reference. ATPase lobe 2 of XPB is positioned at greater distance from the DNA than lobe 2 of Ssl2, which interacts with both DNA strands (**Figs. 5.19** and **5.20**). Structural shifts between the open and closed state of lobe 2 transmit to the Ssl2/XPB ATP-binding pocket situated between lobes 1 and 2. The color code is indicated. Views as in **Figs. 5.19** and **5.20**.

The kinase module of TFIIH appears to be the most mobile segment in free and cPIC-bound TFIIH and was not observed in 3D-reconstructions of human TFIIH and the yeast PIC. However, while 2D-classifications of free TFIIH define an approximate positional range for the kinase-cyclin pair[333], indicating an intrinsic stabilization, its location in the PIC structure is revealed only in presence of cMed. This suggests that Mediator, albeit without detected direct interactions, prompts the TFIIH kinase module to adopt a position in close proximity to the initiation complex, in agreement with a stimulation of its CTD kinase activity[133,240,241] and other PIC-Med studies[101]. Mediator may direct partial immobilization of the kinase module without the involvement of rigid and structured elements but instead by tethering via flexible linkers or low-affinity interactions. Such a scenario would explain the observed EM-density, which seems to correspond rather to a preferred, not exclusive, location of the kinase-cyclin pair and indicates that the sub-complex retains some mobility. Nevertheless, the elucidation of the underlying mechanisms for TFIIH kinase module recruitment and stimulation will ultimately require very highly resolved structures of PIC-cMed or PIC-Med complexes in combination with biochemical mapping and mutational analyses.

The decreased flexibility of the kinase module in free TFIIH is directly associated with a marked discrepancy between free and cPIC-bound TFIIH. Although structures of both models correlate strongly on the subunit level (Fig. 7.1 a, b), the global crescent- or horseshoe-like shape of core-TFIIH apparently undergoes a major rearrangement upon its incorporation into the initiation machinery (Fig. 7.2). While TFIIH adopts a rather open conformation in the cPIC-bound state, in which the ATPases Ssl2 and Rad3 are located at the terminal and most distant points of the core-TFIIH crescent, free TFIIH is much more compacted and the ATPases XPB and XPD are in direct contact with each other. Although species-specific differences may contribute to this contraction, it appears to primarily result from a robust linkage of both ATPases by a long α-helix in MAT1 (Fig. 7.2). This α-helix extends from the *H. sapiens* equivalent of the Tfb3 Rad3-anchor, which interacts with XPD, to the DRD domain of XPB, which is associated with ATPase lobe 1[333]. Whereas its existence is strongly predicted by various SSE identification tools for *S. cerevisiae* as well, a respective Tfb3 α-helix was not observed in the yeast PIC structure. Thus, the N-terminal segment of XPB/Ssl2 lacks its stabilizing effect, which provides an explanation for the decreased resolution of the PIC's EM-map in this region. In the PIC, the contact between a putative Tfb3/MAT1 α-helix and the Ssl2/XPB DRD may not be established due to the more expanded shape of core-TFIIH. The α-helix may simply be unable to traverse the correspondingly increased gap between Rad3/XPD and Ssl2/XPB. Intriguingly, the Tfb3/MAT1 α-helix also represents the only reported physical connection between the kinase-cyclin pair and core-TFIIH since it is embedded between the Rad3/XPD-anchor of Tfb3/MAT1 and its C-terminus, which anchors Kin28/CDK7 and Ccl1/CycH[214]. Thus, while the α-helix is engaged between the ATPases, the kinase-cyclin pair is kept in relative proximity to core-TFIIH, separated only by small spacer region between the C-terminal end of the α-helix and the C-terminus of Tfb3/MAT1. Accordingly, the kinase-cyclin pair is observed as a blurry shape in 2D-classifications of free TFIIH[333]. However, if the MAT1 α-helix is conserved between *H. sapiens* and yeast, it appears to

become a mobile element, tethered only to the Rad3/XPD-anchor of Tfb3/MAT1, upon recruitment of TFIIH into the PIC, as demonstrated in this study. In consequence, the length and flexibility of the 'spacer' that connects the kinase-cyclin pair with core-TFIIH is significantly enhanced, which implies a sudden increase of the activity range of the TFIIH kinase and would permit the phosphorylation of distant targets. Moreover, it remains uncertain whether the MAT1 α-helix will retain its conformation after dissociation from the DRD of Ssl2/XPB or will become a disordered linker. The described putative rearrangements of the kinase module subunits would explain why the kinase-cyclin pair was not observed in 2D-classes of the yeast PIC but of *H. sapiens* TFIIH. Cryo-EM analyses of the human PIC seem to confirm such a theory[86,333] although the obtained EM-maps have to be interpreted cautiously owing to the low resolution.
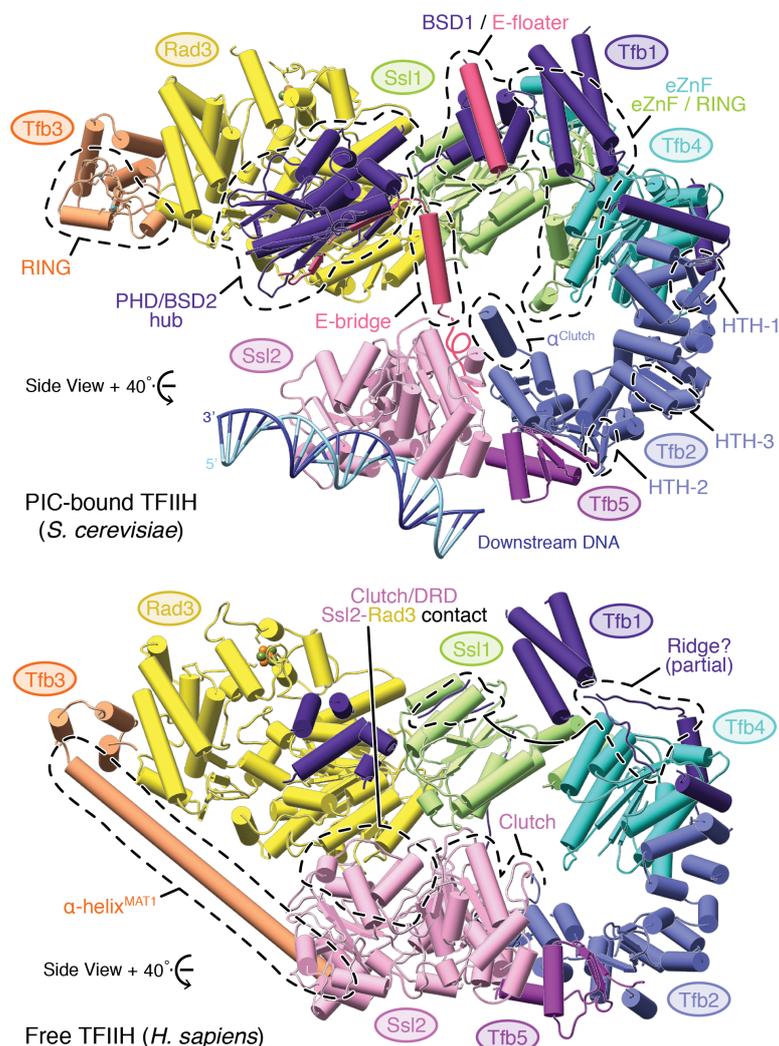
Figure 7.2 | Comparison of free and PIC-bound TFIIH. TFIIH adopts very distinct conformations in the free and PIC-associated states. Structures are viewed from the side (adapted from **Fig 5.14**). XPB and XPD in free TFIIH contact each other directly while Ssl2 and Rad3 are physically separated in the PIC-derived structure. Elements specifically observed/modeled in only one of the structures are indicated with dashed black lines, respectively. The free TFIIH model comprises peptides that may be part of the Tfb1/p62 Ridge domain, which were not modeled in the yeast structure owing to poor resolution. Alternatively they may be (partially) attributed to the eZnF or RING folds of Ssl1/p44 and Tfb4/p34. The α^{Clutch}-helix of Tfb2 is a yeast-specific element and not predicted in the sequence of p52 in *H. sapiens*. Color code as in **Fig. 5.14**. The color code of free TFIIH was adapted to the subunit assignment of PIC-associated TFIIH.

The comparison of the two recently obtained structural studies that revealed TFIIH in its free and cPIC-bound state demonstrates that TFIIH undergoes major conformational changes on a global level during its incorporation into the initiation machinery, while most of its central modules retain their original local fold. However, several peripheral mobile domains in free TFIIH become more rigid and ordered within the PIC, presumably due to stabilizing interactions with the newly identified helical elements in the C-terminus of Tfa1/TFIIEα. The notable expansion of the crescent shape of core-TFIIH appears to be directly induced by its association with the PIC and is probably based on three pivotal elements. First, Rad3/XPD is fixed in one position by the N-terminal Rad3-anchor and RING-finger of Tfb3/MAT1, which bind the ARCH domain of Rad3/XPD and intercalate between the stalk of Pol II and the E-linker helices, respectively. Correspondingly, the RING-finger of Tfb3/MAT1 was clearly assignable only within the PIC[333]. Second, Ssl2/XPB with a reported high affinity for dsDNA[109] presumably shifts to engulf the DNA downstream of Pol II with its ATPase domains, thereby opening the core-TFIIH crescent and losing its contact with the Tfb3/MAT1 α-helix. The described relocation of the ATPases, and, consequently, the tips of the crescent, is facilitated by the newly discovered bridge helix of Tfa1/TFIIEα. The E-bridge is wedged between the Tfb1/p62 BSD2, which in turn is anchored to ATPase lobe 2 of Rad3, and ATPase lobe 2 of Ssl2/XPB (Fig 7.2). Although atomic details of the interactions between the E-bridge and core-TFIIH remain yet elusive, they may provide sufficient stability for this element to serve as a rigid spacer between the two halves of the TFIIH crescent. As discussed, the expansion of core-TFIIH and the presumed resulting release of the Tfb3/MAT1 α-helix present novel implications on the location and mobility of the TFIIH kinase-cyclin pair. However, although the proposed hypotheses appear to be plausible, the analysis of structural rearrangements in TFIIH using protein complexes derived from one species and prepared by a uniform strategy will be an important future goal. At present, yeast TFIIH was obtained from a recombinant expression system whereas human TFIIH was purified endogenously. In addition, the introduction of mutations in distinct domains of TFIIH and the associated elements of Tfa1 and Pol II may provide further detailed knowledge about their interaction network and its impact on stability and functionality of TFIIH during transcription initiation. In the past, transcription assays were successfully used to characterize variants of TFIIE or cMed[88,89]. An adapted setup or, as an alternative, promoter opening assays may prove suitable for the analysis of TFIIH. For such an approach, the recombinant production scheme established in this study would be clearly beneficial. Similar procedures may have to be applied to dissect the mechanism by which Mediator coordinates the kinase-cyclin pair.

7.3   Towards an atomic model of transcription initiation

The structures of the PIC and PIC-cMed complex are in agreement with numerous functional data and offer valid explanations for previous observations like the intimate association of TFIIE and TFIIH[100], TFIIE-dependent ATPase stimulation of TFIIH[110], or the kinase module's requirement for *in-vitro* initiation experiments[119]. However, as indicated before, many aspects of the underlying molecular mechanisms that ultimately drive transcription initiation remain yet elusive. A better understanding of these processes at the structural level demands a plethora of atomic structures that reflect the various states adopted by the entirety of the transcription initiation machinery during the distinct phases of initiation. While the PIC and PIC-cMed structures represent snapshots of the initiation stage after DNA melting and formation of an initial transcription bubble, their resolution is not sufficient to reveal all molecular details, particularly not in the newly described TFIIH and cMed regions.

Although the current method for PIC-cMed preparation is superior to any other tested approach, the system is not completely optimized, as indicated by the identification of heterogeneous particles in the cryo-EM data. Whereas SDS-PAGE analysis indicated full occupancy of the cMed and TFIIH sub-modules, as well as homogeneous crosslinking of the PIC-cMed complex, only 6% of the visualized particles contained both cMed and TFIIH and more than 50% contained neither of them. Moreover, many particles were discarded owing to strong conformational heterogeneity, which may have been caused by mechanical destabilization of the complex rather than by its intrinsic flexibility. Therefore, as a first step to arrive at PIC-cMed structures of high- or even atomic resolution, the sample preparation protocol has to be adapted to achieve higher occupancy rates for cMed and TFIIH while retaining a similar particle concentration for cryo-EM experiments. The observed partial decomposition of the cPIC-cMed complex is most likely attributable to concentration of the sample after crosslinking and dialysis, or to the plunge-freezing process. Thus, alternative methods, which increase concentration but exert less stress on macromolecules than conventional ultrafiltration, need to be assayed. For example, proteins may be dialyzed against a buffer containing high molecular weight PEGs or Ficoll-400 (Sigma-Aldrich) prior to application to cryo-EM grids. Another approach, frequently used for preparation of endogenous protein samples with low concentration, would involve sample adsorption to cryo-EM grids with a continuous carbon support film[58,85,86,333]. These grids would, albeit at the cost of a slightly decreased contrast and an increased DQE, permit *in-situ* sample concentration by adjustment of the incubation time and may additionally stabilize fragile regions of the complex and prevent decomposition during plunge freezing. Once the current biochemical limitations of sample preparation have been overcome and a final robust strategy for the assembly of the PIC-cMed complex is established, cryo-EM experiments are expected to result in better resolved reconstructions, simply due to the presence of a larger number of intact, rigid particles. Subsequently, the PIC-cMed analysis may be extended to further states of the initiation complex such as closed PIC-cMed or initially transcribing complexes (ITCs) with RNA.

While the determination of the PIC-cMed structure marks a major breakthrough, the complex still does not represent the entire initiation machinery. The macromolecular factors

Mediator tail and TFIID, or more precisely its TAF subunits, are additionally required for completion of the canonically defined initiation complex[334]. The incorporation of these missing elements into the present model is highly anticipated but may introduce significant rearrangements. Except for the stalk region of Pol II, the architecture of the cPIC remained almost unaltered upon addition of TFIIH and cMed but both cMed and TFIIH underwent drastic conformational changes compared to their free states[234,247,333]. A future assembly with Mediator tail and activators may similarly affect peripheral regions in the PIC-cMed complex. The tail segment of Mediator comprises a similar size as cMed in yeast and associates with transcriptional activators at DNA enhancer elements[52], thereby mediating physical long-range interactions between them and the cMed-bound initiation machinery. Thus, it is vital to the regulation of initiation complex assembly as well as to the transmission of signals to tune its activity. Taking the versatile nature of Mediator[222,223,233] and the capability of cMed to undergo significant structural reorganization (Fig. 5.22 c) into account, local and global rearrangements of Mediator, and possibly its interaction partners, are likely a crucial component of the crosstalk between activators and the Pol II-associated complex[52,246]. Although a cryo-EM analysis of a yeast PIC-Med complex including Mediator tail has been performed, the resolution in most parts of the obtained EM-map was not sufficient to distinguish conformational changes[101]. Consequently the structural impact of an association of the PIC with complete Mediator and transcriptional activators remains elusive. Moreover, structural cooperativity between Mediator and TFIID, the largest general transcription factor with a molecular mass of 1.2 MDa in yeast[42], during promoter recruitment has been reported[335]. TFIID recognizes core promoter sequences and further modulates promoter-enhancer interactions by binding to multiple transcriptional activators as well as acetylated nucleosomes[42,58]. Thus, the complex is generally considered as the second base component of the PIC assembly platform. The structure of a TFIIA-TFIID-promoter complex has been revealed at medium resolution[58] and provides insights into TFIID-mediated promoter recognition. Moreover it offers a model for correct positioning of TBP and the priming of promoter DNA for PIC recruitment independently of TATA-box element availability[336], To date, however, the structure of a PIC-(cMed)-TFIID complex has not yet been obtained. Superposition of EM-maps and models of PIC, TFIID and cMed by using common elements suggests only minor clashes of TFIID with the other factors[37,336] (Fig 7.3), indicating that the holo-complex may be stable. In addition, it has been demonstrated that TFIID, similarly to Mediator, is highly modular, undergoes significant rearrangements and exists in cell-type specific alternative states in metazoans[42,58,336]. Consequently, the determination of initiation complex structures containing canonical TFIID or conformational variants may reveal distinct subunit organizations and additional effects on the remainder of the initiation machinery.

Once the discussed optimization of the PIC-cMed reconstitution strategy has reached a stage of routine application for cryo-EM analysis, the current complex may be extended by incorporation of TFIID, Mediator tail and transcriptional activators. Similarly to TFIIH, these factors may be produced recombinantly in insect cells, permitting variations in subunit composition and the introduction of mutations. In a long-term effort, it thus may not only be

possible to ultimately arrive at an atomic model of the canonical transcription initiation complex but also to structurally dissect its entirety during the various stages of initiation and track putative conformational changes. Based on the derived structures, additional targeted mutational analyses and biochemical assays may then contribute to the elucidation of further key determinants and molecular mechanisms that govern the initial phase of transcription.
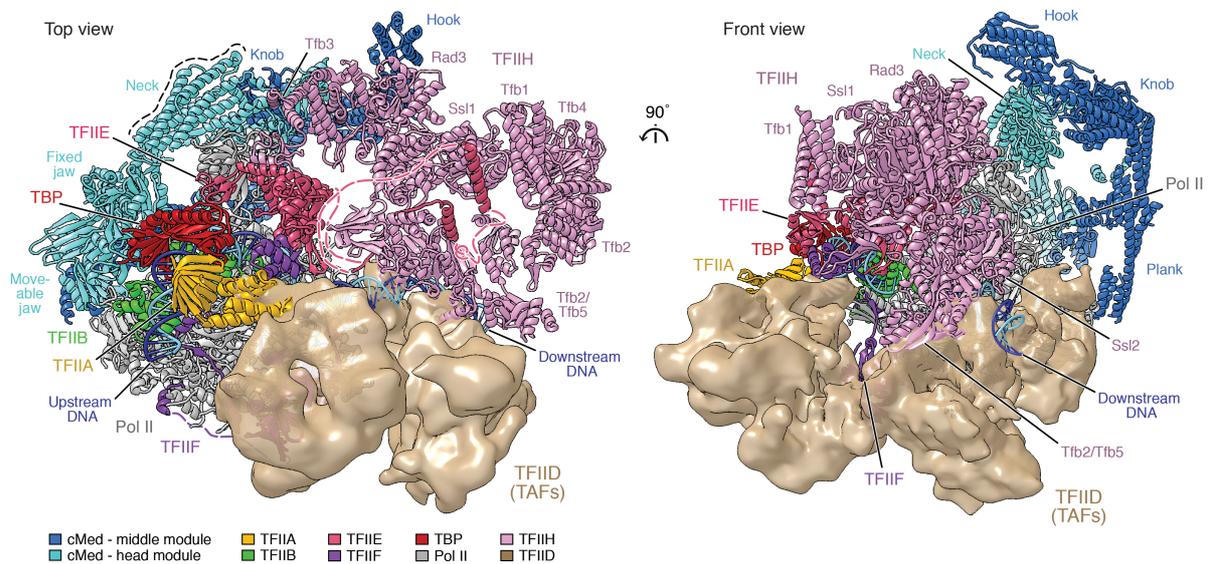
Figure 7.3 | Composite topological model of a TFIID-containing Pol II initiation complex. Cryo-EM density of TFIID (brown) (analysis of a TFIIA-TFIID-promoter complex, EMDB accession EMD-3305)[58] was superposed on the structure of the PIC-cMed complex using the common elements of the upstream promoter assembly. Only the TAF subunits of TFIID are depicted. Overlaps and clashes between the TAFs, cPIC and TFIIH are negligible. Initiation complexes are viewed from the side and the front (Figs. 5.11, 5.21). The color code is indicated.

7.4 Novel concepts to transcription initiation

Very recently several studies have presented novel experimental data, which, although not being in direct opposition to the canonical interpretation of transcription initiation events, suggest that an extension or revision of current, commonly accepted models may be advisable. These analyses especially affect the fields of promoter recruitment, promoter opening and promoter escape as well as the regulatory mechanisms involved in these processes. Although they provide a complementation to available structural information such as the models of the PIC and PIC-cMed complex, also new questions are being raised.

For a long time, the existence of two differently regulated gene classes utilizing distinct platforms for PIC assembly has been implied. Genes containing promoters with poorly defined TATA box elements ('TATA-less genes') were considered to be dependent on TFIID, consistent with a model of TFIID serving as a 'molecular ruler' for correct TBP positioning[58]. Promoter recruitment to genes with canonical TATA boxes, however, was supposedly controlled by another transcriptional co-activator complex, SAGA, which shares several subunits with TFIID, has a

molecular mass of ~ 2 MDa in yeast and is a histone modifier[337]. Overlaps between SAGA- and TFIID-based regulation were assumed to be minimal, especially since genes of the respectively dominated classes were located in distinct chromatin environments with divergent histone modifications and +1 nucleosome positioning[56,57]. However, applying the recently advanced 'chromatin endogenous cleavage coupled with high throughput sequencing' (ChEC-seq) technique, a genome-wide recruitment of both TFIID and SAGA to both TATA-less and TATA-containing genes has been demonstrated in yeast[338,339]. Consequently, a novel model suggests co-localization of these co-activators at essentially all promoters of Pol II-transcribed genes and thus a global role for SAGA in transcription initiation, similar to that of Mediator[338]. Although SAGA does not bind DNA and interactions with Pol II or the remainder of the transcription initiation machinery have not been reported, the existence of promoter assemblies that integrate the three macromolecular co-activators, Mediator, SAGA and TFIID, now seems plausible. However, it remains elusive whether SAGA is able to form transient or stable complexes with the other co-activators or further PIC components. Cooperation may also occur primarily on a functional level, regulated by recruitment of Mediator and SAGA to specific, co-localizing transcriptional activators bound to adjacent enhancer DNA regions[340]. Thus, the characterization of promoter architectures containing the Med-SAGA-TFIID triad and their association with PIC assemblies will require an advanced set of structure-function analyses, particularly since chromatin context may be a determinant for complex properties, composition and activity. Nevertheless, if the biochemical reconstitution or isolation of a SAGA-containing transcription initiation complex would be realized, the canonical model of the initiation machinery[334] would have to be extended by an additional macromolecular factor. As such a ~5 MDa complex would likely also have to be positioned in the correct chromatin environment, its analysis may be one of the greatest challenges for structural biology in the near future.

Despite extensive biochemical and *in-vivo* studies, the exact mechanism by which unwinding of promoter DNA occurs in the Pol II system has remained enigmatic. In contrast to other multi-subunit RNA polymerases, Pol II requires the presence of ATP and activity of the TFIIH subunit Ssl2/XPB for formation of an open promoter complex[90,104,106]. The presented PIC and PIC-cMed structures visualized the ATPase lobes of Ssl2 bound to DNA, consistent with models implying an involvement of Ssl2 in promoter opening[103,105,176]. They unambiguously demonstrate that Ssl2 binds distant from the site of initial DNA opening and confirm that it does not act as a conventional DNA-helicase[85-87,96,98,108,109]. Ssl2 rather functions as a dsDNA translocase that is directed away from the cPIC, thereby feeding promoter DNA into the active center cleft of Pol II and inducing promoter unwinding through a right-handed screwing rotation of the DNA.

The reports of 'spontaneous', Ssl2/TFIIH-independent DNA opening and transcription events in structural and functional analyses[89,112] have challenged the concept of Ssl2/XPB as an indispensable factor for the 'mechanical' promoter opening process. Instead, a 'built-in-block' model[112] was suggested, assigning a regulatory, rather suppressive role to DNA-bound, but inactive, Ssl2/XPB. However, also in the structures of the PIC and PIC-cMed complexes, which

contained TFIIH and a non-hydrolysable ATP-analogue, spontaneous DNA opening was observed, indicating that the expected stabilization of closed DNA by Ssl2 had not occurred. As spontaneous DNA unwinding would require the absence of interactions with the ATPase domains of Ssl2, this effect may be explained either by slow kinetics of the *in-vitro* PIC formation, which permitted DNA opening before TFIIH and Ssl2 were correctly positioned, or by inherently dynamic properties of the Ssl2-mediated DNA sequestration. These dynamics would imply a series of poor binding events permitting the DNA to 'slip' through the enzyme's ATPase domains. Recent investigations of the activity of yeast Ssl2/TFIIH may support such a theory. E. Tomko and colleagues studied *in-vitro* promoter opening events in real-time and at single base pair resolution, thereby identifying an initial promoter opening intermediate encompassing 5-6 bp of heteroduplex DNA[341]. This intermediate, which was identified as a direct product of ATP-dependent Ssl2 action, underwent both extension into a fully established 12-13 bp transcription bubble and collapse into a closed state. Moreover, analyzed DNA fragments apparently participated in multiple cycles of DNA opening and closing events. The observed reversibility of the transition between closed DNA and the 5-6 bp intermediate bubble suggested that the Ssl2 enzyme might not constantly remain tightly bound to DNA but dissociate and re-associate, thereby allowing DNA to 'slip' and initial bubbles to collapse. This mechanism may have contributed to the unexpected spontaneous opening of DNA in the PIC and PIC-cMed complexes but without further information, kinetic effects cannot be completely excluded as an underlying cause. Thus the proposed explanation remains speculative.

The identification of a Ssl2-generated 5-6 bp promoter opening intermediate[341] moreover implies that DNA unwinding is dependent on further processes in addition to the ATP-consuming action of Ssl2/TFIIH, since a fully established transcription bubble encompasses 12-13 bp[86,341]. The idea of a corresponding two-step mechanism, which potentially limits ATP-expense per opening event, was first introduced more than two decades ago. Holstege et al. demonstrated TFIIH-independence of a second phase of promoter opening *in-vitro* by using 5-9 bp-comprising heteroduplex scaffolds as minimal initial templates[104]. Their observations may relate to the processes underlying the spontaneous opening of promoter DNA that have been described lately. Remarkably, structural evidence for spontaneous opening was reported only for *HIS4* promoter DNA sequences[89], which require little energy input for DNA strand separation. *In-vitro* transcription experiments have demonstrated that specific promoters are more resistant to opening and suggested that the intrinsic biophysical properties of the DNA, especially its meltability and supercoiling state, as well as environmental conditions significantly affect promoter opening efficiency in context of the PIC[94]. For the AT-rich *HIS4*-promoter and in specific assembly buffers, this equilibrium may be shifted towards open DNA so severely that spontaneous opening occurs *in-vitro* when long sample incubation times apply, such as for preparation of the cPIC[89] and PIC-cMed complex. As discussed, the Ssl2-induced 'block' may be temporarily released, thereby permitting the transition of *HIS4*-DNA into a thermodynamically favored open state in TFIIH-containing assemblies as well.

Some of the discussed theories will likely contribute to a final explanation for spontaneous DNA opening observed in initially closed cPIC[89], PIC and PIC-cMed complexes in absence of ATP. Still, it remains elusive how a proposed two-step mechanism is consolidated and regulated at different types of promoters inside a living cell. In order to develop a reliable model, the structural basis of the intermediate states of promoter opening will have to be analyzed and reconciled with the plethora of available biochemical data. This involves investigation of closed PICs comprising DNA templates with high meltability and, in particular, PIC intermediates with various small heteroduplex scaffolds. These may aid to recapitulate the transition from initial promoter unwinding, facilitated by the catalyst-like function of TFIIH, to ATP-independent transcription bubble extension. Deciphering the precise roles of TFIIH and factors like TFIIE and TFIIF, which may contact and reposition DNA during promoter melting[42,47,89], will moreover promote the investigation of potential parallels to the ATP-independent initiation processes in the Pol I and III systems. After all, a unified mechanism of promoter opening in eukaryotic multi-subunit RNA polymerases may still be in part existent.

When the EM-map of the PIC-cMed complex was first derived and interpreted, it was thoroughly analyzed for traces of the Pol II CTD in the density. In yeast, the CTD comprises 26 hepta-peptide repeats of the $Y_1S_2P_3T_4S_5P_6S_7$ consensus sequence, of which few are slightly degenerate. The CTD may be altered by introduction of various PTMs that define the universal CTD code. Specific patterns of phosphorylation, for example, primarily regulate the progression through the transcription cycle[19]. Thus, while fulfilling no known catalytic or structural function, the CTD is a pivotal component for the integration and transmission of regulatory transcriptional signals and is essential for the organization of the distinct transcription phases. At the onset of transcription initiation, the unmodified CTD presumably aids in recruitment of Pol II to promoters by serving as an anchor that binds the co-activator Mediator. Although high affinity of the CTD for endogenous Mediator and, in particular, its isolated head and middle modules has been demonstrated *in vitro* [250,342], to date only little structural evidence has been provided to elucidate the nature of this interaction. A single X-ray structure of a 4-heptapeptide repeat of the CTD bound to the Mediator head module was published[343] but has been challenged by novel findings, which indicate potential steric clashes between the proposed path of the CTD peptide and the interfaces between the Mediator head and middle modules[247]. On the contrary, recent experiments suggest that the CTD in its unmodified state may not participate in defined, sequence-specific interactions. Instead, its incorporation into liquid droplets formed by transcription-associated proteins that comprise intrinsically disordered regions has been repeatedly demonstrated[19]. Indeed, owing to its repetitive character, many properties of the CTD resemble that of low-complexity domains and favor the formation of multivalent intra- and intermolecular interactions, potentially resulting in liquid-liquid phase separation. Consistent with such a hypothesis, a defined electron density that may be attributed to the CTD was absent from the EM-reconstruction of the PIC-cMed complex. However, EM-maps with significantly higher resolution will be required to confirm this observation and, possibly permit the detection of a compact, but

weakly ordered, hydrogel-like CTD entity in the density. As proposed, the CTD is expected to be enveloped and shielded by the cradle formed by Pol II, Mediator and TFIIH. The theory of a compacted CTD[250] that assumes a molten-globule shape was formulated early on[255] and may imply a function in CTD interaction for intrinsically disordered regions in Mediator, which are predicted within numerous conserved subunit segments[52,344]. Moreover, it agrees with reported crosslinks between the CTD C-terminus and Med19[88,252] that face towards the inside of the cradle. In contrast, the model of an elongated CTD-peptide meandering along Mediator, as advocated by Robinson *et. al.*[101], mis-positioned these crosslinks on the cradle outside, likely due to a lack of information on the detailed Mediator middle module structure[247]. A more global and precise positional analysis of the Pol II CTD may be permitted by replacement of single residues in the CTD sequence by artificial lysines and subsequent XL-MS experiments. Reconstituted PIC-cMed complexes comprising such mutant CTDs could be investigated to derive information about the domains of PIC-cMed subunits located adjacently to the respective heptad repeats and may reveal interactions of putative low-complexity regions.

During the last step of transcription initiation, promoter escape, Mediator and most general transcription factors are released from Pol II, which subsequently proceeds to the elongation stage. Though in tandem with further processes, such as emergence of the nascent transcript from the mRNA exit tunnel and consecutive displacement of TFIIB[42], the dissociation of Mediator strongly promotes the destabilization and disassembly of the initiation complex. Numerous studies have suggested that phosphorylation of the CTD by the TFIIH kinase Kin28/CDK7 is a prerequisite for promoter clearance since it weakens the affinity between the PIC and Mediator[132-134]. However, to date is has remained elusive by which mechanism the observed destabilization occurs or whether CTD phosphorylation by Kin28/CDK7 is sufficient to induce disintegration of the complete initiation machinery *in vitro*. Having established protocols for the preparation of active, recombinant TFIIH and the formation of PIC-cMed complexes, which constitute the majority of the initiation machinery, the required activity and pulldown assays can now be performed with purified factors and artificially dephosphorylated Pol II. In addition, such analyses may provide further data on the putative arrangement of the CTD in a low-complexity globular domain that resides in the central cradle of the initiation complex. CTD peptides lose their ability for incorporation into hydrogels upon phosphorylation[19], suggesting that the addition of phosphate groups modifies relevant physiochemical properties of the CTD, for example its hydrophobicity and charge distribution. The decreased hydrophobicity of the CTD and the accumulation of negative charges following heavy phosphorylation at physiological conditions would result in internal electrostatic repulsion of diverse regions, thus disrupting multivalent interaction networks. Moreover, the accompanying spatial expansion of the CTD might infer steric clashes with the surface of the surrounding cradle and thus lead to dissociation of Mediator and, consequently, other PIC components. According to this theory, any kind of hyper-phosphorylation, independent of the identity of the modified residues and independent of the involved kinase, would suffice for Mediator release from the PIC. This has been demonstrated in previous experiments performed with endogenous Mediator, Pol II and distinct kinases[241]. To

prove that the destabilization of the initiation complex is a primarily stochastic event also in a larger context, the PIC-cMed complex could be subjected to corresponding time-course experiments with Kin28 and other kinases. Alternatively, the Mediator affinity of recombinant Pol II variants that mimic specific CTD phosphorylations could be analyzed.

The promotion of the CTD from a simple platform for protein recruitment to a semi-structural component, which may be actively involved in controlling initiation complex shape and the definition of micro-compartments for transcription, implies intriguing functional and experimental perspectives. However, the physiological relevance of the Pol II CTD's ability to participate in liquid-liquid phase separation and the possible formation of transient 'transcription compartments' is currently heavily debated[19]. Although numerous observations agree with this hypothesis, there is yet too little direct evidence for a partial regulation of transcription by low-complexity interactions between Pol II and its associated factors. Structural biology may only be able to provide limited information on the underlying systems, as the currently applied methods are not ideal for the analysis of amorphous, fluid-like domain arrangements. However, a fundamental explanation of the mechanisms that govern initial promoter recruitment and promoter escape will only be derived once the respective biological states of the Pol II CTD have been completely elucidated. Thus, the CTD will have to be characterized by hybrid approaches, involving single-molecule techniques or additional indirect biochemical experiments, as discussed.

The determination of a model for the PIC-cMed complex marks a milestone in one of the most ambitious projects of structural biology, which was started more than two decades ago when researchers sought to obtain the first structures of the upstream promoter complex. Naturally, ever more sophisticated methods for the reconstitution of transient macromolecular assemblies and their high-resolution analysis will have to be developed to permit a structural dissection of the complete Pol II-associated machinery throughout the various stages of transcription. The PIC-cMed model, however, may become a crucial resource for future developments in the field of transcription initiation, similarly to the first structures of the cPIC. Thus, the PIC-cMed model may not only serve as a framework and paradigm for the expansion of structural studies to the architecture of the entire initiation machinery and its chromatin environment but may also provide a basis for the establish- and improvement of further advanced biochemical initiation assays.

Nowadays, a wide selection of structural biology tools is available for *in-vitro* analyses of asymmetric particles ranging from small peptides[345] to gigantic macromolecular machineries[346] and hybrid methods can be applied to arrive at reliable models even for flexible, highly modular assemblies like the PIC-cMed complex. The structural assessment of large transcription complexes will remain at the focus of approaches aiming to decipher the molecular details of processes governing the transcription cycle. However, the astute incorporation of findings from other fields will become essential for future research progress. As discussed, some of the established models describing transcription and, specifically, initiation, may require extension by novel unprecedented concepts. Theories such as the attribution of low-complexity liquid-like domain characteristics to the CTD have the potential to revolutionize the present interpretation of transcription mechanisms. Consequently, the current concepts of integrated structural biology will have to be adapted to the ever more complex and dynamic systems to be analyzed in the future. This will require an even more pervasive integration of complementary techniques such as cryo-ET, super resolution microscopy or molecular dynamics simulations, but also system-wide *in-vivo* mapping or single molecule assays. Yet, in a collaborative, community-wide effort the analysis of fundamental processes such as transcription in both their native cellular context and on the single molecule level may eventually be achieved.

8 Supplementary Materials

8.1 WarpCraft

WarpCraft[292] was utilized to simultaneously refine the cPIC, TFIIH and cMed regions, which were flexible with respect to each other in the conventionally obtained cryo-EM reconstructions. In particular, the resolution of the TFIIH segment was significantly improved over globally refined maps (Fig. 8.1 a, b). Focused 3D-refinement approaches with local masks on TFIIH or specific regions therein yielded reconstructions of similar resolution as the global 3D-refinement with WarpCraft.
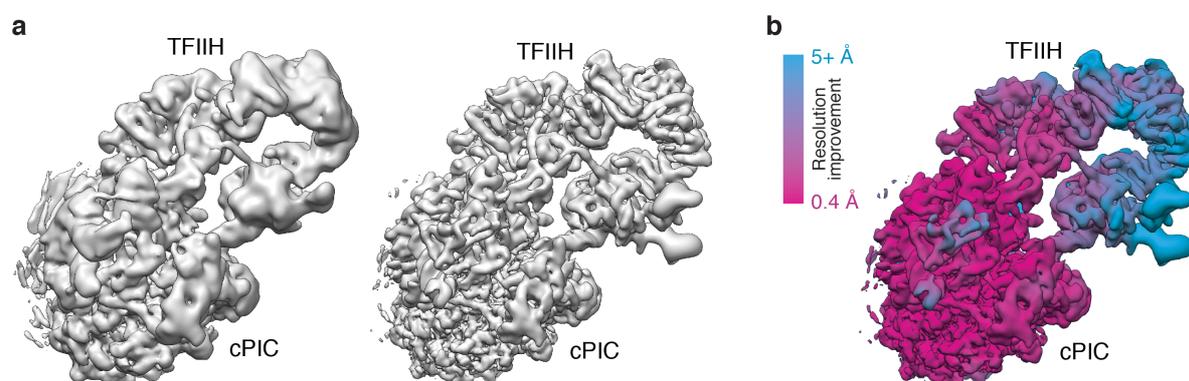
Figure 8.1 | Resolution improvement by WarpCraft. a, Cryo-EM maps of the PIC obtained from 58,000 particles after refinement in RELION (left) and with WarpCraft (right). For PIC reconstructions with WarpCraft, the previously obtained globally refined map was divided in 20 segments for local filtering and calculations were constrained to the first 20 normal modes. TFIIH and cPIC regions are indicated. **b,** Final EM map of the PIC, colored according to resolution improvement by WarpCraft. The color scheme is indicated. Maps were generously provided by D. Tegunov.

8.2   The pET-MCN system

The pET-MCN system[265,266] was utilized to generate *E. coli* expression vectors comprising multiple ORFs (Methods, Fig. 8.2). The concept is based on repeated restriction endonuclease cleavage and subsequent ligation reactions. Ligation of overhangs produced by two specific distinct endonucleases (SpeI and XbaI) results in a fused restriction site, which may not be cleaved by either of the originally used enzymes. Thus the capability of a pET-MCN vector to undergo multiple rounds of assembly is preserved. Since combined genes are under control of a single preceding lac operon (lacO) and T7 promoter, the order of their assembly may affect their expression efficiency, in particular for genes located at the 3' end of the resulting multi-cistronic construct. The removal of single genes from such a vector is not possible. In this study, up to five genes were combined with the pET-MCN system.
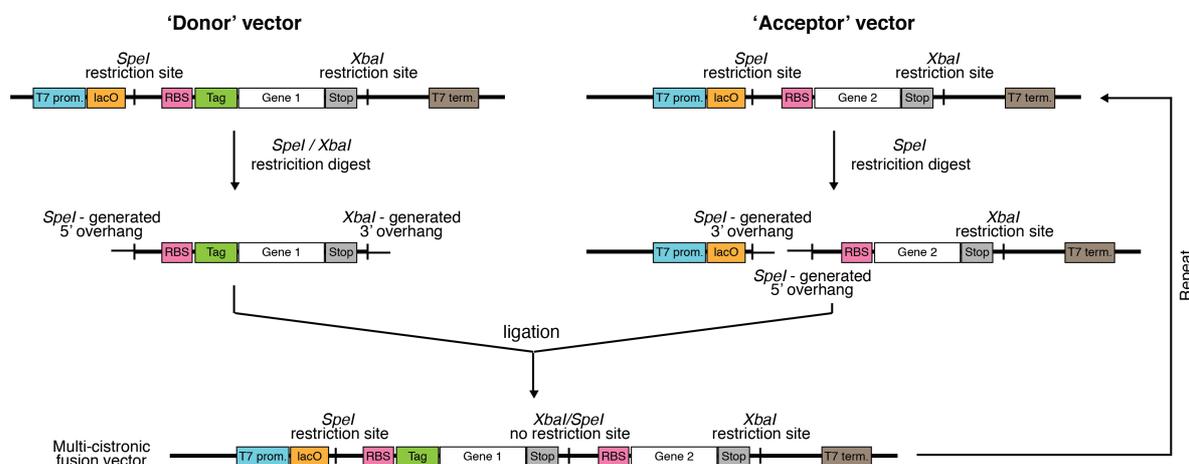
Figure 8.2 | Assembly of multi-cistronic vectors with the pET-MCN system. Schematic depiction of the combination of two distinct pET-MCN vectors, containing ORFs for 'Gene 1' and 'Gene 2', respectively. Genes are initially introduced into pET-MCN vectors by conventional restriction endonuclease cleavage and ligation utilizing a multiple cloning site (MCS). The multi-cistronic fusion product comprises the backbone of the original 'acceptor' vector and the genes of both the 'acceptor' and 'donor' vectors. Multi-cistronic constructs may serve as templates for further steps of vector assembly. The 5' and 3' designations refer to dsDNA positions relative to the ORF, not to ssDNA strands. T7 prom, T7 promoter; lacO, lac operon; RBS, ribosomal binding site; Tag, affinity tag; Stop, stop codon; T7 term, T7 terminator.

8.3 The MacroBac system

Similarly to the pET-MCN system, the MacroBac[267] system was used to assemble large multi-gene expression constructs. These were transferred into baculoviral shuttle vectors and utilized for subsequent infection of insect cells to induce recombinant protein expression. The assembly of MacroBac vectors is based on LIC methods and the Biobricks™ system (Methods). The vector design preserves the viral promoter and a SV40 polyA termination signal for each gene after vector combination. Consequently, distinct genes are expressed independently of each other and the order of assembly is irrelevant for protein production levels, contrary to the pET-MCN system. In this study, vectors comprising up to 12 inserts could be assembled by repeated rounds of LIC.
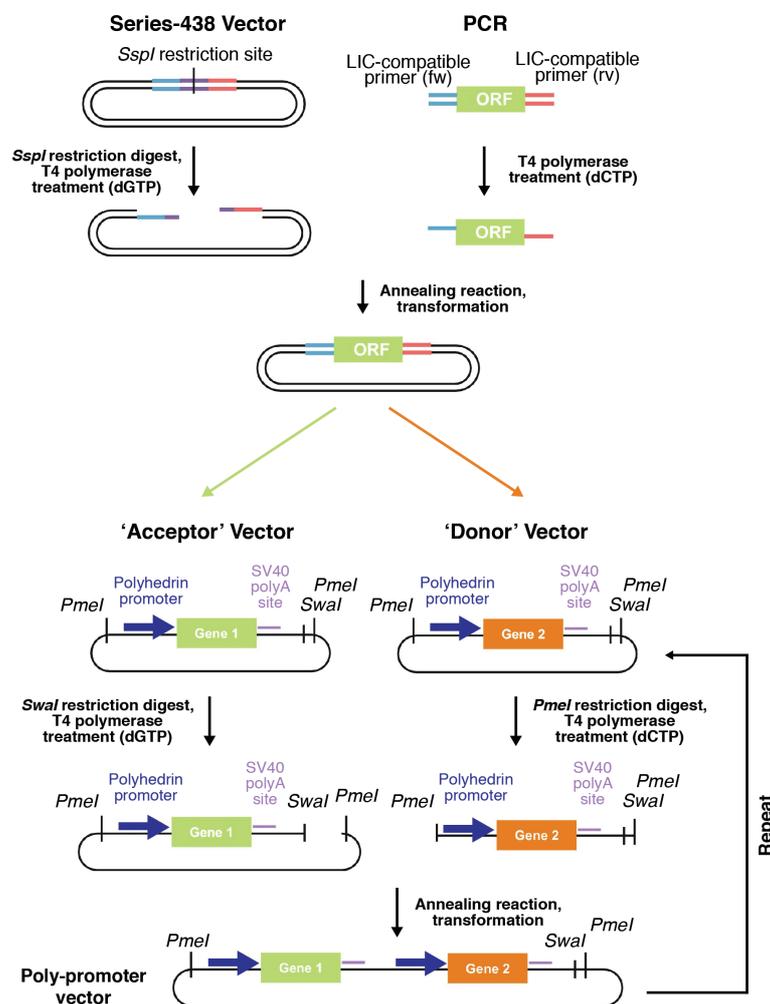
Figure 8.3 | Assembly of poly-promoter vectors with the MacroBac system. Schematic representation of the generation of initial MacroBac vectors and vector combination. Endonuclease cleavage with *SspI* exposes a LIC-compatible insertion site in Series-438 vector backbones. Sequences for LIC-compatible PCR primers are listed in Table 3.6. The poly-promoter fusion product comprises the backbone of the original 'acceptor' vector and the genes of both the 'acceptor' and 'donor' vectors. Poly-promoter constructs may serve as a template for further steps of vector assembly. LIC, ligation independent cloning; fw, forward; ORF, open reading frame; rv, reverse. Adapted from[267].

8.4 Supplementary Tables

8.4.1 Composition of the PIC-cMed complex

Table 8.1 │ Components of the PIC-cMed complex. Names of the human homologues of the yeast subunits are provided. For details about complex assembly and composition, see the main text and Methods.

Component	Subunit (yeast)	Corresponding human subunit	Construct residues (aa) / scaffold length (nt)	Mass (kDa)	Molarity
Pol II	Rpb1	RPB1	1-1733	191.6	1
	Rpb2	RPB2	1-1224	138.7	1
	Rpb3‡	RPB3	1-318	35.3	1
	Rpb4	RPB4	1-221	25.4	1
	Rpb5	RPB5	1-215	25.1	1
	Rpb6	RPB6	1-155	17.9	1
	Rpb7	RPB7	1-171	19.1	1
	Rpb8	RPB8	1-146	16.5	1
	Rpb9	RPB9	1-122	14.3	1
	Rpb10	RPB10	1-70	8.3	1
	Rpb11	RPB11	1-120	13.6	1
	Rpb12	RPB12	1-70	7.7	1
TFIIF	Tfg1‡	RAP74	1-735	82.2	5
	Tfg2	RAP30	1-400	46.6	5
Nucleic acid strands	Template	-	106	32.6	1.5
	Non-template	-	106	32.6	1.5
TFIIA	Toa1	TFIIAα	1-94, 210-286	19.4	10
	Toa2‡	TFIIAβ	1-122	13.4	10
TBP	TBP‡	TBP	1-240	27.0	5
TFIIB	TFIIB‡	TFIIB	1-345	38.2	5
TFIIE	Tfa1‡	TFIIEα	1-482	54.7	2.5
	Tfa2	TFIIEβ	1-328	37.0	2.5
core-TFIIH	Rad3	XPD	1-778	89.8	2.5
	Ssl1	p44	1-461	52.3	2.5
	Ssl2	XPB	1-843	95.3	2.5
	Tfb1	p62	1-642	72.9	2.5
	Tfb2	p52	1-513	58.5	2.5
	Tfb4	p34	1-338	37.5	2.5
	Tfb5	p8/TTDA	1-72	8.2	2.5
TFIIH kinase module	Ccl1	CycH	1-393	45.2	2.5
	Kin28	CDK7	1-306	35.2	2.5
	Tfb3	MAT1	1-321	38.1	2.5
cMed	Med1‡	Med1	1-566	64.2	1.5
	Med4	Med4	1-284	32.2	1.5
	Med6	Med6	1-295	32.8	1.5
	Med7	Med7	1-222	25.6	1.5
	Med8	Med8	1-223	25.3	1.5
	Med9	Med9	1-149	17.4	1.5
	Med10	Med10	1-157	17.9	1.5
	Med11	Med11	1-115	13.3	1.5
	Med14‡	Med14	1-745	84.6	1.5
	Med17	Med17	1-687	78.5	1.5
	Med18	Med18	1-307	34.3	1.5
	Med19	Med19	1-220	24.9	1.5
	Med20	Med20	1-210	22.9	1.5
	Med21	Med21	1-140	16.1	1.5
	Med22	Med22	1-121	13.8	1.5
	Med31	Med31	1-127	14.7	1.5
Final	PIC-cMed	46 subunits	16,622 aa	1948.7	-

aa: amino acids, nt: nucleotides, kDa: kilodalton

‡ constructs contain N- or C-terminal 6xHis or 10xHis tags as described

8.4.2 Modeling of TFIIH

Table 8.2 | Modeling of yeast TFIIH subunits, domains and regions.

For details refer to main text and Methods.

Subunit / Chain ID	Domain	Residue range (yeast)	Initial model	PDB-template for initial model	Modeling algorithm	Changes to initial model	Density assigned in Figure 5.15c / Color
Rad3 / 0	N-terminus	1-17	not modeled	-	-	-	yes / yellow
Rad3 / 0	Lobe 1	18-108, 204-248, 441-484	homology-/ *ab-initio* model	2VSF:A	I-Tasser		-
Rad3 / 0	FeS cluster	109-203	homology-/ *ab-initio* model	2VSF:A	I-Tasser		-
Rad3 / 0	ARCH	249-293, 327-425, 429-440	(1) *de-novo* modeling of α-helices	-	Gorgon	MDFF and manual corrections, PHENIX geometry minimization	-
			(2) homology-/ *ab-initio* model	α-helix PDB from Gorgon, 5IVW:W, 5IVW:A, 5IY9:W	I-Tasser		-
Rad3 / 0	ARCH linker	294-326	poly-alanine model	-	-		-
Rad3 / 0	Lobe 2	485-723	homology-/ *ab-initio* model	2VSF:A	I-Tasser		-
Rad3 / 0	C-terminus	724-778	not modeled	-	-	-	yes / yellow
Tfb1 / 1	PHD	1-121	yeast NMR-structure	1Y5O:A	-	PHENIX geometry minimization	-
Tfb1 / 1	PH-linker	122-175	not modeled	-	-	-	no
Tfb1 / 1	BSD1	176-218	homology-/ *ab-initio* model	2DII:A	I-Tasser		-
Tfb1 / 1	Linker	219-251	poly-alanine model loop	-	-		-
Tfb1 / 1	BSD2	252-294	homology-/ *ab-initio* model	2DII:A	I-Tasser	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb1 / 1	Linker	295-307	poly-alanine model loop	-	-		-
Tfb1 / 1	Rad3 anchor (α-Helix 1)	308-330	poly-alanine model α-helix	-	-		-
Tfb1 / 1	Linker	331-353	poly-alanine model loop	-	-		-
Tfb1 / 1	Rad3 anchor (α-Helix 2)	369-394	poly-alanine model α-helix	-	-		-
Tfb1 / 1	Ridge	395-464	not modeled	-	-	-	partially / purple
Tfb1 / 1	Tfb4 anchor (α-Helix 1)	465-483	poly-alanine model α-helix	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb1 / 1	Linker	484-494	poly-alanine model loop	-	-		-

Tfb1 / 1	Tfb4 anchor (α-Helix 2)	495-519	poly-alanine model α-helix	-	-		-
Tfb1 / 1	Linker	520-543	not modeled	-	-	-	no
Tfb1 / 1	3-helix bundle	544-639	(1) *ab-initio* model	-	QUARK	MDFF and manual corrections, PHENIX geometry minimization	-
			(2) homology-/ *ab-initio* model	PDB output from QUARK 2W6D:A, 1WRD:A, 4M70:A, 5F8P:A,	I-Tasser		
Tfb1 / 1	C-terminus	640-643	not modeled	-	-	-	no
Tfb2 / 2	N-terminus	1-2	not modeled	-	-	-	no
Tfb2 / 2	α	3-40	poly-alanine model α-helix	-	-		-
Tfb2 / 2	HTH-1	41-113	homology-/ *ab-initio* model	1U2W:A	SWISS		-
Tfb2 / 2	α	114-131	poly-alanine model loop	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb2 / 2	α	132-159	poly-alanine model α-helix/loop	-	-		-
Tfb2 / 2	HTH-2	160-194	homology-/ *ab-initio* model	2MC3:A	SWISS		-
Tfb2 / 2	HTH-2	195-213	poly-alanine model α-helix	-	-		-
Tfb2 / 2	HTH-3	214-281	homology-/ *ab-initio* model	5BOX:A	SWISS		-
Tfb2 / 2	Linker	282-336	not modeled	-	-	-	no
Tfb2 / 2	Clutch	337-419	poly-alanine model α-helix/β-sheet (partially)	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb2 / 2	Linker	420-432	not modeled	-	-	-	yes / blue
Tfb2 / 2	Dimerization domain	433-450	poly-alanine model α-helix	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb2 / 2	Dimerization domain	451-508	yeast crystal structure	3DGP:A	-	PHENIX geometry minimization	-
Tfb2 / 2	C-terminus	509-513	not modeled	-	-	-	no
Tfb3 / 3	N-terminus	1-7	not modeled	-	-	-	no

Tfb3 / 3	RING	8-70	homology-/ *ab-initio* model	1G25:A, 3M62:A	I-Tasser		-
Tfb3 / 3	ARCH anchor	71-145	(1) *ab-initio* model	-	QUARK	MDFF and manual corrections, PHENIX geometry minimization	-
			(2) homology/ *ab-initio* model	PDB output from QUARK; 2W6D:A, 3NIX:A, 5EQZ:A, 2EWF:A	I-Tasser		-
Tfb3 / 3	C-terminus	146-321	not modeled	-	-	-	no
Tfb4 / 4	N-terminus	1-21	not modeled	-	-	-	no
Tfb4 / 4	vWA	22-88	homology-/ *ab-initio* model	4PN7:A, 4CRN:X, 4PLA:A, 4OKU:A, 5IY9:3	I-Tasser	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb4 / 4	vWA insertion	89-97	poly-alanine model loop	-	-		-
Tfb4 / 4	vWA insertion	98-102	not modeled	-	-	-	no
Tfb4 / 4	vWA insertion	103-114	poly-alanine model loop	-	-		-
Tfb4 / 4	vWA	115-256	homology-/ *ab-initio* model	4PN7:A, 4CRN:X, 4PLA:A, 4OKU:A, 5IY9:3	I-Tasser	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb4 / 4	Linker	257-273	poly-alanine model loop	-	-		-
Tfb4 / 4	eZnF / C-terminus	274-323	homology-/ *ab-initio* model	3LRQ:D	Robetta		-
Tfb4 / 4	C-terminus	324-338	not modeled	-	-	-	no
Tfb5 / 5	N-terminus	1	not modeled	-	-	-	no
Tfb5 / 5	Dimerization domain	2-64	yeast crystal structure	3DGP:B	-	PHENIX geometry minimization	-
Tfb5 / 5	C-terminus	65-68	poly-alanine model loop	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Tfb5 / 5	C-terminus	69-72	not modeled	-	-	-	no
Ssl1 / 6	unstructured NTE	1-73	not modeled	-	-	-	no
Ssl1 / 6	α	74-122	not modeled	-	-	-	yes / green
Ssl1 / 6	vWA	123-308	yeast crystal structure	4WFQ:A	-	PHENIX geometry minimization	-
Ssl1 / 6	Linker	309-324	poly-alanine model loop	-	-	MDFF and manual corrections,	-

Ssl1 / 6	eZnF	325-372	homology-/ *ab-initio* model	1NNQ:A	Robetta	PHENIX geometry minimization	-
Ssl1 / 6	α	373-386	poly-alanine model α-helix	-	-	-	-
Ssl1 / 6	RING	387-457	homology-/ *ab-initio* model	1Z60:A	Robetta		-
				1Z60:A	SWISS		
Ssl1 / 6	C-terminus	458-461	not modeled	-	-	-	no
Ssl2 / 7	NTE	1-110	-	-	-	-	no
Ssl2 / 7	Clutch and DRD	111-362	not modeled	-	-	-	yes / pink
Ssl2 / 7	Lobe 1	363-425	homology-/ *ab-initio* model	2FWR:A	SWISS		-
Ssl2 / 7	Lobe 1	426-451	homology model / poly-alanine model	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Ssl2 / 7	Lobe 1	452-462	homology-/ *ab-initio* model	2FWR:A	SWISS		-
Ssl2 / 7	Lobe 1	463-481	poly-alanine model α-helix/loop	-	-		-
Ssl2 / 7	Lobe 1	482-548	homology-/ *ab-initio* model	2FWR:A	SWISS		-
Ssl2 / 7	Lobe 2	549-691	homology-/ *ab-initio* model	4ERN:A	Robetta		-
Ssl2 / 7	Lobe 2	692-702	poly-alanine model α-helix/loop	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
Ssl2 / 7	Lobe 2	703-712	homology-/ *ab-initio* model	4ERN:A	Robetta		-
Ssl2 / 7	Extension	713-770	homology-/ *ab-initio* model	4ERN:A	Robetta		-
Ssl2 / 7	C-Terminus	771-843	not modeled	-	-	-	no
TFIIE / W	E-bridge	259-266	poly-alanine model β-strand	-	-	MDFF and manual corrections, PHENIX geometry minimization	-
TFIIE / W	E-bridge	267-289	poly-alanine model α-helix	-	-		-
TFIIE / W	E-floater	349-373	poly-alanine model α-helix	-	-		-
TFIIE / W	Acidic peptide	407-417	*homo sapiens* NMR structure	2RNR:A	-	amino acid replacement MDFF, PHENIX geometry minimization	-

8.4.3 Modeling of cPIC and cMed

Table 8.3 | Adaptation and extension of yeast cPIC and cMed models.

For details refer to the main text and Methods.

Factor/Subunit / Chain ID	Initial Model	PDB-template for initial model	Modeling algorithm	Changes to initial model
Pol II / A-L	yeast high resolution EM structure	5FYW:A-L	-	rigid body fit of clamp region (residues 1-346 in chain A) and peripheral regions in Rpb3 (residues 22-44, 73-97), Rpb6 (residues 72-154), Rpb8 (residues 2-146), Rpb9 (residues 20-43, 55-117) and Rpb12 (residues 26-65); MDFF correction of stalk (chain D, G), PHENIX geometry minimization
TFIIA / U-V	yeast high resolution EM structure	5FYW:U-V	-	rigid body fit, PHENIX geometry minimization
TFIIB / M	yeast high resolution EM structure	5FYW:M	-	extension by residues 59-123 based on PDB 4BBR:M; MDFF correction of this range, PHENIX geometry minimization
TBP / O	yeast high resolution EM structure	5FYW:O	-	PHENIX geometry minimization
TFIIE / W-X	homology model	5GPY:A-B	I-Tasser SWISS	MDFF correction of homology model, PHENIX geometry minimization
TFIIF / Q-R	yeast high resolution EM structure	5FYW:Q-R	-	rigid body fit of residues 329-415 in chain Q and residues 58-244 in chain R; extension of linker (residues 245-250) in chain R, PHENIX geometry minimization

cMed model originates from *S. cerevisiae* homology model[247] of PDB 5N9J and was subjected to alterations listed below; complete cMed model was subjected to PHENIX geometry minimization routine

Med4 / h	homology model	5N9J:H	-	extension of α-helix in residue range 118-126; *de-novo* modeling of α-helix in residue range 130-145; MDFF correction of homology model
Med7 / i	yeast crystal structure	1YKE:A	-	extension of α-helix in residue range 206-211; MDFF correction of homology model
Med14 / l	homology model	5N9J:L	-	truncation of residues 341-345; MDFF correction of homology model
Med21 / j	yeast crystal structure	1YKE:B	-	extension of α-helix in residue range 128-138; MDFF correction of homology model
Med31 / o	yeast crystal structure	3FBI:B	-	truncation of residues 94-110; replacement by *de-novo* modeled α-helix (SSE predicted); MDFF correction of homology model

8.4.4 Cryo-EM data collection and model statistics

Table 8.4 | Cryo-EM data collection and model statistics for the PIC and the PIC-cMed complex structures. For details about EM data collection, data processing and model building, see the main text and Methods.

	PIC (EMD-3846) (PDB 5OQJ)	PIC-cMed (EMD-3850) (PDB 5OQM)
Data collection		
Magnification	105 000	105 000
Particles	58 000	16 000
Pixel size (Å)	1.37	1.37
Defocus range (μm)	-0.5 to -5.0	-0.5 to -5.0
Voltage (kV)	300	300
Electron exposure ($e^-/Å^2$)	42	42
Symmetry imposed	C1	C1
Reconstruction		
Map resolution (Å)	4.70	5.84
Map sharpening B-factor ($Å^2$)	-201	-334
FSC threshold	0.143	0.143
Model composition		
Non-hydrogen atoms	62,931	79,757
Protein residues	8,188	10,825
DNA bases	148	148
Ligand atoms	25	25
Model validation		
MolProbity Score	1.92	1.89
Clashscore	7.83	7.53
Rotamer outliers (%)	0.30	0.37
Cβ-deviations	0	0
Ramachandran Plot		
Favored (%)	91.95	92.21
Allowed (%)	6.66	6.45
Disallowed (%)	1.39	1.34

8.4.5 Crosslinking information

Table 8.5 | Compilation of published BS3- and SBAT- crosslinking information.

a. TFIIH crosslinks from previous studies[101,190,201] utilizing BS3 and SBAT and respective C_α-distances within the PIC-cMed model obtained in this study. Crosslinks used to verify various TFIIH regions are indicated. For details also refer to the main text and Methods.

Study	Link type	Protein 1	Protein 2	Residue 1	Residue 2	Distance (Å)	Used for modeling	Remark
Robinson 2016	intra	Rad3	Rad3	30	476	16.4	x	-
Murakami 2013, Luo 2016, Robinson 2016	intra	Rad3	Rad3	30	481	14.8	x	-
Luo 2015	intra	Rad3	Rad3	81	673	19.5	x	-
Robinson 2016	intra	Rad3	Rad3	95	125	39.0	-	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Rad3	Rad3	112	125	5.9	x	-
Robinson 2016	intra	Rad3	Rad3	112	132	10.7	x	-
Robinson 2016	intra	Rad3	Rad3	112	142	24.0	x	-
Robinson 2016	intra	Rad3	Rad3	112	180	16.0	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Rad3	Rad3	112	605	19.6	x	-
Luo 2015	intra	Rad3	Rad3	122	125	8.6	x	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	122	132	9.4	x	-
Luo 2015	intra	Rad3	Rad3	125	588	26.2	x	-
Luo 2015	intra	Rad3	Rad3	144	276	27.1	x	-
Luo 2015	intra	Rad3	Rad3	144	281	18.5	x	-
Luo 2015	intra	Rad3	Rad3	173	68	17.3	x	-
Robinson 2016	intra	Rad3	Rad3	183	112	15.0	x	-
Robinson 2016	intra	Rad3	Rad3	188	180	11.6	x	-
Luo 2015	intra	Rad3	Rad3	228	68	9.7	x	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	336	343	10.5	x	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	347	343	6.2	x	-
Luo 2015	intra	Rad3	Rad3	357	372	13.7	x	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	364	372	11.7	x	-
Luo 2015	intra	Rad3	Rad3	447	476	11.9	x	-
Luo 2015	intra	Rad3	Rad3	447	636	24.9	x	-
Luo 2015	intra	Rad3	Rad3	476	30	16.3	x	-
Luo 2015	intra	Rad3	Rad3	476	481	16.4	x	-
Robinson 2016	intra	Rad3	Rad3	476	636	28.4	x	-
Robinson 2016	intra	Rad3	Rad3	490	476	37.6	-	low score
Robinson 2016	intra	Rad3	Rad3	490	636	43.2	-	-
Luo 2015	intra	Rad3	Rad3	499	489	24.5	x	-
Robinson 2016	intra	Rad3	Rad3	588	636	44.9	-	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	605	125	18.4	x	-
Luo 2015	intra	Rad3	Rad3	605	588	21.4	x	-
Luo 2015, Robinson 2016	intra	Rad3	Rad3	605	636	24.3	x	-
Luo 2015	intra	Rad3	Rad3	673	112	38.0	-	-
Luo 2015	intra	Rad3	Rad3	673	125	39.6	-	low score
Luo 2015	intra	Rad3	Rad3	673	489	10.2	x	-
Robinson 2016	intra	Rad3	Rad3	673	490	7.1	x	-
Luo 2015	intra	Rad3	Rad3	673	588	24.4	x	-
Luo 2015	intra	Rad3	Rad3	673	636	42.8	-	-
Luo 2015	intra	Rad3	Rad3	695	481	16.0	x	-
Luo 2015	intra	Rad3	Rad3	695	490	12.0	x	-
Luo 2015	intra	Rad3	Rad3	695	673	15.5	x	-
Luo 2015	inter	Rad3	Ssl2	499	372	29.2	x	-
Robinson 2016	inter	Rad3	Tfb1	180	334	14.3	x	-
Luo 2015	intra	Ssl1	Ssl1	139	201	17.4	x	-
Luo 2015	intra	Ssl1	Ssl1	139	205	14.7	x	-

Luo 2015	intra	Ssl1	Ssl1	197	201	7.2	x	-
Robinson 2016	intra	Ssl1	Ssl1	312	321	24.9	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Ssl1	Ssl1	397	414	17.4	x	-
Murakami 2013, Luo 2015	intra	Ssl1	Ssl1	397	420	15.3	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Ssl1	Ssl1	414	420	4.6	x	-
Luo 2015	intra	Ssl1	Ssl1	433	397	10.6	x	-
Robinson 2016	inter	Ssl1	Tfa1	201	284	21.9	x	-
Luo 2015	inter	Ssl1	Tfb1	201	101	54.8	-	-
Robinson 2016	inter	Ssl1	Tfb1	201	238	33.2	-	positional variants
Luo 2015	inter	Ssl1	Tfb2	197	415	29.2	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Ssl1	Tfb2	201	415	27.4	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Ssl1	Tfb2	201	418	22.1	x	-
Luo 2015, Robinson 2016	inter	Ssl1	Tfb2	201	419	22.4	x	-
Luo 2015	inter	Ssl1	Tfb4	197	173	34.0	-	-
Luo 2015	inter	Ssl1	Tfb4	197	207	27.4	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Ssl1	Tfb4	315	319	17.8	x	-
Robinson 2016	inter	Ssl1	Tfb4	321	319	11.4	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Ssl1	Tfb4	321	323	10.5	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Ssl1	Tfb4	397	95	20.8	x	-
Murakami 2013, Luo 2015	inter	Ssl1	Tfb4	414	95	23.2	x	-
Robinson 2016	inter	Ssl2	Rpb5	472	94	18.1	x	-
Luo 2015	intra	Ssl2	Ssl2	523	372	25.7	x	-
Luo 2015	intra	Ssl2	Ssl2	766	624	22.3	x	-
Robinson 2016	inter	Ssl2	Tfa2	472	277	62.7	-	only 1 spectral count
Luo 2015	inter	Ssl2	Tfb2	734	415	19.6	x	-
Robinson 2016	inter	Tfa1	Tfb1	355	186	13.2	x	-
Murakami 2013, Robinson 2016	inter	Tfa1	Tfb1	366	179	12.7	x	-
Murakami 2013, Robinson 2016	inter	Tfa1	Tfb1	367	581	19.2	x	-
Luo 2015	inter	Tfb1	Rad3	73	125	83.1	-	low score
Luo 2015	inter	Tfb1	Rad3	101	95	87.3	-	-
Luo 2015	inter	Tfb1	Rad3	120	125	60.0	-	low score
Robinson 2016	inter	Tfb1	Rad3	334	183	14.4	x	-
Robinson 2016	inter	Tfb1	Rad3	334	588	13.7	x	-
Luo 2015	inter	Tfb1	Rad3	335	112	13.4	x	-
Luo 2015	inter	Tfb1	Rad3	335	588	16.1	x	-
Luo 2015	inter	Tfb1	Rad3	376	125	24.0	x	-
Luo 2015	inter	Tfb1	Rad3	390	588	10.3	x	-
Robinson 2016	inter	Tfb1	Rpb1	73	1093	116.7	-	only 1 spectral count, low score
Luo 2015	inter	Tfb1	Ssl1	120	201	33.2	-	-
Luo 2015, Robinson 2016	inter	Tfb1	Ssl1	238	205	21.4	x	-
Luo 2015	inter	Tfb1	Ssl1	246	205	26.7	x	-
Robinson 2016	inter	Tfb1	Ssl1	255	201	28.7	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Tfb1	Ssl1	255	205	18.6	x	-
Robinson 2016	inter	Tfb1	Tfa1	587	367	24.9	x	-
Luo 2015	intra	Tfb1	Tfb1	47	101	11.5	x	-
Robinson 2016	intra	Tfb1	Tfb1	47	112	9.8	x	-
Robinson 2016	intra	Tfb1	Tfb1	57	65	24.3	x	-
Robinson 2016	intra	Tfb1	Tfb1	73	65	15.8	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	73	70	9.0	x	-
Robinson 2016	intra	Tfb1	Tfb1	73	71	6.5	x	-
Luo 2015	intra	Tfb1	Tfb1	73	120	30.5	x	-

Robinson 2016	intra	Tfb1	Tfb1	83	47	15.8	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	83	70	11.7	x	-
Luo 2015	intra	Tfb1	Tfb1	83	71	13.0	x	-
Luo 2015	intra	Tfb1	Tfb1	83	291	41.3	-	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	101	112	16.1	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	101	120	28.4	x	-
Luo 2015	intra	Tfb1	Tfb1	101	335	62.1	-	-
Robinson 2016	intra	Tfb1	Tfb1	112	65	14.0	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	120	65	21.6	x	-
Luo 2015	intra	Tfb1	Tfb1	120	70	21.7	x	-
Murakami 2013, Robinson 2016	intra	Tfb1	Tfb1	120	246	16.2	x	-
Luo 2015	intra	Tfb1	Tfb1	120	279	14.8	x	-
Luo 2015	intra	Tfb1	Tfb1	120	291	24.3	x	-
Murakami 2013, Robinson 2016	intra	Tfb1	Tfb1	171	574	23.1	x	-
Murakami 2013	intra	Tfb1	Tfb1	171	581	22.1	x	-
Robinson 2016	intra	Tfb1	Tfb1	227	179	20.7	x	-
Robinson 2016	intra	Tfb1	Tfb1	227	255	14.4	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	227	267	20.6	x	-
Luo 2015	intra	Tfb1	Tfb1	238	101	31.7	-	low score
Robinson 2016	intra	Tfb1	Tfb1	238	120	26.3	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	238	255	11.2	x	-
Robinson 2016	intra	Tfb1	Tfb1	238	279	23.5	x	-
Luo 2015	intra	Tfb1	Tfb1	246	101	20.7	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	246	255	17.6	x	-
Luo 2015	intra	Tfb1	Tfb1	246	276	26.1	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	246	279	19.5	x	-
Luo 2015	intra	Tfb1	Tfb1	246	291	16.3	x	-
Robinson 2016	intra	Tfb1	Tfb1	255	186	28.9	x	-
Robinson 2016	intra	Tfb1	Tfb1	255	300	23.9	x	-
Robinson 2016	intra	Tfb1	Tfb1	268	279	14.8	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	268	324	15.2	x	-
Robinson 2016	intra	Tfb1	Tfb1	268	334	35.2	-	only 1 spectral count
Luo 2015	intra	Tfb1	Tfb1	276	291	19.5	x	-
Luo 2015	intra	Tfb1	Tfb1	291	70	39.9	-	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	291	295	5.9	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	291	300	10.4	x	-
Robinson 2016	intra	Tfb1	Tfb1	291	324	20.3	x	-
Luo 2015	intra	Tfb1	Tfb1	300	335	37.9	-	low score
Robinson 2016	intra	Tfb1	Tfb1	300	384	23.6	x	-
Robinson 2016	intra	Tfb1	Tfb1	324	267	15.9	x	-
Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	324	384	11.9	x	-
Luo 2015	intra	Tfb1	Tfb1	335	384	19.6	x	-
Luo 2015	intra	Tfb1	Tfb1	335	390	22.3	x	-
Luo 2015	intra	Tfb1	Tfb1	376	384	12.3	x	-
Robinson 2016	intra	Tfb1	Tfb1	384	268	20.4	x	-
Luo 2015	intra	Tfb1	Tfb1	384	295	17.1	x	-
Luo 2015	intra	Tfb1	Tfb1	498	508	15.0	x	-
Luo 2015	intra	Tfb1	Tfb1	508	501	10.2	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb1	Tfb1	508	515	9.9	x	-
Luo 2015	inter	Tfb1	Tfb3	291	10	51.1	-	low score
Luo 2015	inter	Tfb1	Tfb4	83	323	95.4	-	-
Robinson 2016	inter	Tfb1	Tfb4	101	323	92.4	-	only 1 spectral count, low score
Luo 2015	inter	Tfb1	Tfb4	120	323	68.0	-	-

Murakami 2013, Robinson 2016	intra	Tfb1	Tfb4	171	319	23.2	x	-
Luo 2015	inter	Tfb1	Tfb4	483	108	6.8	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb2	Tfb2	238	262	12.4	x	-
Luo 2015, Robinson 2016	intra	Tfb2	Tfb2	415	419	6.5	x	-
Murakami 2013, Luo 2015, Robinson 2016	inter	Tfb2	Tfb5	495	6	8.9	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb2	Tfb5	506	6	18.1	x	-
Luo 2015	inter	Tfb3	Rad3	10	125	44.7	-	-
Luo 2015	inter	Tfb3	Rad3	17	125	50.0	-	low score
Robinson 2016	inter	Tfb3	Rad3	119	276	17.7	x	-
Luo 2015, Robinson 2016	intra	Tfb3	Tfb3	17	57	11.8	x	-
Luo 2015	intra	Tfb3	Tfb3	27	65	16.4	x	-
Luo 2015	intra	Tfb3	Tfb3	61	67	14.9	x	-
Luo 2015	intra	Tfb3	Tfb3	86	67	16.9	x	-
Murakami 2013	intra	Tfb3	Tfb3	94	132	17.2	x	-
Murakami 2013, Luo 2015	intra	Tfb3	Tfb3	119	132	8.7	x	-
Luo 2015	intra	Tfb3	Tfb3	141	65	29.5	x	-
Luo 2015	intra	Tfb3	Tfb3	141	67	26.0	x	-
Luo 2015	intra	Tfb3	Tfb3	141	132	14.0	x	-
Luo 2015	inter	Tfb4	Ssl1	108	397	42.3	-	-
Luo 2015, Robinson 2016	inter	Tfb4	Tfb2	108	238	24.6	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb4	Tfb4	84	138	19.4	x	-
Luo 2015	intra	Tfb4	Tfb4	95	108	23.6	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb4	Tfb4	95	138	29.4	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb4	Tfb4	131	138	15.9	x	-
Luo 2015, Robinson 2016	intra	Tfb4	Tfb4	165	207	14.8	x	-
Murakami 2013, Luo 2015, Robinson 2016	intra	Tfb4	Tfb4	173	207	10.1	x	-
Luo 2015	intra	Tfb4	Tfb4	319	323	6.6	x	-
Luo 2015	intra	Tfb4	Tfb4	321	323	5.7	x	-
Luo 2015	inter	Tfb5	Ssl1	6	201	73.2	-	-
Luo 2015	inter	Tfb5	Ssl2	60	721	17.0	x	-
Robinson 2016	intra	Tfb5	Tfb5	46	6	13.7	x	-
Robinson 2016	intra	Tfb5	Tfb5	60	51	14.4	x	-

b. TFIIH crosslinks from previous studies[101,190,201] that were located up to 12 residues from the next residue modeled in this study. These crosslinks were additionally used to confirm TFIIH regions. For details also refer to Fig. 5.18.

Study	Link type	Protein 1	Protein 2	Residue 1	Residue 2
Robinson 2016	inter	Tfa1	Ssl2	301	711
Robinson 2016	intra	Tfa1	Tfa1	284	295
Robinson 2016	intra	Tfa1	Tfa1	286	301
Robinson 2016	inter	Tfa1	Tfb2	295	415
Murakami 2013	inter	Tfa1	Tfb2	301	427

Table 8.6 | Newly obtained EDC-crosslinks.

a. Summary and statistics of EDC-crosslinks in the PIC-cMed complex obtained in this study.

PIC-cMed components	Total number of crosslinks	Number of crosslinks mappable on the PIC-cMed structure
Total crosslinks	262	158
Inter-protein crosslinks		
Total	109	65
Pol II – Pol II	18	15
Pol II – TFIIB	8	4
Pol II – TFIIE	8	3
Pol II – TFIIF	4	2
Pol II – TFIIH	4	0
Pol II – cMed	2	2
TFIIA – TFIIA	2	2
TFIIA – TFIIE	1	0
TFIIA – TBP	1	0
TFIIB – TBP	1	0
TFIIA – TFIIF	3	0
TFIIE – TFIIE	1	0
TFIIE – TFIIF	2	1
TFIIF – TFIIF	14	9
TFIIE – TFIIH	11	10
TFIIH – TFIIH	19	13
TFIIH – cMed	3	0
cMed – cMed	7	4
Intra-protein crosslinks		
Total	153	93
Pol II	35	34
TFIIA	4	3
TFIIB	5	4
TBP	8	0
TFIIE	3	3
TFIIF	22	6
TFIIH	59	34
cMed	17	9

b. List of single EDC-crosslinks. 'Total Count' refers to sum of spectral counts originating from both replicates and 'Score' refers to highest detected score value (Methods). C_α distances are indicated if structural information was available within a range of 4 residues from the cross-linked residue. Several crosslinks were used to guide TFIIE density interpretation and to confirm TFIIE-TFIIH interactions. For details also refer to the main text and Fig. 5.18.

Link type	Protein 1	Protein 2	Residue 1	Residue 2	Total Count	Score	Distance (Å)	Alternative residue
Intra	Ccl1	Ccl1	45	286	2	3,93	NA	-
Intra	Kin28	Kin28	32	37	2	6,72	NA	-
Intra	Med1	Med1	375	389	1	8,88	NA	-
Intra	Med1	Med1	377	389	1	8,01	NA	-
Intra	Med14	Med14	420	544	2	3,84	7.8	418
Intra	Med14	Med14	539	544	2	10,84	8.0	-
Intra	Med17	Med17	386	416	1	9,73	9.5	413
Intra	Med17	Med17	386	418	2	13,55	NA	-
Intra	Med17	Med17	418	427	3	11,05	NA	-
Intra	Med17	Med17	419	427	2	15,78	NA	-
Intra	Med17	Med17	421	427	2	23,03	5.2	424

Intra	Med17	Med17	579	608	1	3,74	12.6	-
Intra	Med17	Med17	581	608	12	18,05	13.1	580
Inter	Med17	Med22	548	118	1	5,51	9.6	-
Inter	Med17	Med22	608	105	4	11,67	5.4	-
Inter	Med17	Med22	608	106	1	15,05	8.6	-
Inter	Med17	Med22	608	109	3	11,46	7.4	-
Inter	Med17	Med7	75	191	1	4,83	NA	-
Inter	Med17	Med7	78	191	1	7,28	NA	-
Intra	Med20	Med20	2	187	1	3,69	13.2	-
Intra	Med20	Med20	10	183	3	4,31	8.0	-
Intra	Med21	Med21	107	113	1	8,04	10.0	-
Intra	Med4	Med4	34	39	2	14,65	NA	-
Intra	Med7	Med7	93	103	3	8,82	NA	-
Intra	Med8	Med8	189	210	2	3,31	NA	-
Inter	Med9	Med4	117	26	1	4,75	NA	-
Inter	Med9	Rpb1	88	977	3	4,77	9.9	-
Inter	Med9	Rpb1	99	918	6	5,74	15.9	-
Inter	Rad3	Med7	389	1	2	9,7	NA	-
Intra	Rad3	Rad3	68	225	8	10,59	13.3	-
Intra	Rad3	Rad3	132	269	1	12,46	23.4	-
Intra	Rad3	Rad3	268	336	2	3,21	10.1	-
Intra	Rad3	Rad3	269	336	2	8,01	13.7	-
Intra	Rad3	Rad3	276	389	3	7,67	15.8	-
Intra	Rad3	Rad3	276	394	1	3,91	24.3	-
Intra	Rad3	Rad3	281	389	35	14,94	12.8	-
Intra	Rad3	Rad3	336	340	6	18,42	5.9	-
Intra	Rad3	Rad3	578	605	2	5,13	11.8	-
Intra	Rad3	Rad3	605	608	1	12,22	8.9	-
Intra	Rad3	Rad3	605	674	1	3,3	28.7	-
Intra	Rad3	Rad3	613	673	3	6,58	11.7	-
Intra	Rad3	Rad3	636	640	4	4,69	6.0	-
Inter	Rad3	Tfb1	112	340	2	7,25	13.6	-
Inter	Rad3	Tfb1	605	340	1	4,94	7.0	-
Inter	Rad3	Tfb1	605	345	3	10,12	16.7	-
Inter	Rad3	Tfb1	605	346	1	5,3	17.8	-
Inter	Rad3	Tfb3	336	77	18	15,55	9.8	-
Inter	Rad3	Tfb3	343	74	3	8,51	8.9	-
Intra	Rpb1	Rpb1	39	49	11	10,7	7.0	-
Intra	Rpb1	Rpb1	186	193	4	9,02	4.3	197
Intra	Rpb1	Rpb1	186	195	5	8,77	4.3	197
Intra	Rpb1	Rpb1	368	398	5	5,71	10.2	-
Intra	Rpb1	Rpb1	724	728	3	3,73	6.0	-
Intra	Rpb1	Rpb1	833	1102	13	13	9.0	-
Intra	Rpb1	Rpb1	951	1290	1	9,74	9.9	-
Intra	Rpb1	Rpb1	1093	1074	13	6,87	24.9	-
Intra	Rpb1	Rpb1	1093	1309	3	6,16	10.0	-
Intra	Rpb1	Rpb1	1132	1206	12	10,42	10.2	-
Inter	Rpb1	Rpb11	368	8	6	15,99	16.8	-
Inter	Rpb1	Rpb2	481	987	1	6,02	10.7	-
Inter	Rpb1	Rpb2	483	987	8	5,53	6.4	-
Inter	Rpb1	Rpb2	486	979	6	5,71	12.3	-
Inter	Rpb1	Rpb2	486	1102	6	6,34	8.3	-
Inter	Rpb1	Rpb2	1144	262	47	6,27	13.0	-
Inter	Rpb1	Rpb5	945	201	5	9,17	8.4	-
Inter	Rpb1	Rpb7	2	65	1	12,31	15.6	3
Inter	Rpb1	Rpb9	1253	20	1	8,65	12.6	1254
Inter	Rpb1	Tfa1	193	71	4	6,95	19.2	197

Inter	Rpb1	Tfa1	195	71	11	7,88	19.2	197
Inter	Rpb1	Tfa1	196	71	10	4,21	19.2	197
Inter	Rpb10	Rpb12	68	33	2	7,14	9.1	65
Inter	Rpb10	Rpb12	68	50	8	10,54	18.8	65
Intra	Rpb11	Rpb11	14	20	2	12,22	14.6	-
Intra	Rpb11	Rpb11	16	20	1	10,23	9.1	-
Intra	Rpb11	Rpb11	20	36	3	16,04	8.4	-
Intra	Rpb11	Rpb11	26	38	3	3,29	24.9	-
Intra	Rpb12	Rpb12	58	68	1	4,44	17.2	65
Inter	Rpb2	Rpb1	507	833	3	13,1	NA	-
Inter	Rpb2	Rpb1	979	485	3	4,68	10.4	-
Intra	Rpb2	Rpb2	99	183	5	7,89	12.6	-
Intra	Rpb2	Rpb2	131	134	3	15,45	9.9	-
Intra	Rpb2	Rpb2	133	138	1	6,14	17.0	-
Intra	Rpb2	Rpb2	227	262	2	6,32	11.7	-
Intra	Rpb2	Rpb2	239	507	1	9,19	NA	-
Intra	Rpb2	Rpb2	246	468	2	4,57	19.2	467
Intra	Rpb2	Rpb2	391	393	1	5,15	5.5	-
Intra	Rpb2	Rpb2	393	621	1	4,57	9.9	-
Intra	Rpb2	Rpb2	886	908	14	8,94	8.8	-
Intra	Rpb2	Rpb2	886	909	1	6,4	9.3	-
Intra	Rpb2	Rpb2	923	934	4	6,1	12.2	-
Intra	Rpb2	Rpb2	924	934	1	4,61	12.1	-
Intra	Rpb2	Rpb2	1057	1061	1	5,25	6.2	-
Inter	Rpb2	Tfg2	441	279	4	7,24	4.7	-
Inter	Rpb2	TFIIB	438	108	2	3,19	27.9	-
Inter	Rpb2	TFIIB	438	112	2	5,73	25.8	-
Inter	Rpb2	TFIIB	921	224	3	8,51	NA	-
Inter	Rpb2	TFIIB	922	224	8	12,19	NA	-
Inter	Rpb2	TFIIB	923	155	3	7,88	23.9	-
Inter	Rpb3	Rpb10	149	67	1	3,23	10.3	65
Inter	Rpb3	Rpb11	137	118	51	15,25	NA	-
Inter	Rpb3	Rpb11	266	84	2	11,79	9.9	265
Inter	Rpb3	Rpb11	266	88	1	8,84	12.3	265
Intra	Rpb3	Rpb3	90	160	4	12,77	11.5	-
Intra	Rpb4	Rpb4	127	142	2	9,93	6.1	-
Intra	Rpb4	Rpb4	131	142	13	17,27	8.2	-
Intra	Rpb5	Rpb5	161	172	5	8,86	9.5	-
Intra	Rpb5	Rpb5	191	194	1	16,46	5.9	-
Intra	Rpb5	Rpb5	194	201	6	7,2	23.0	-
Inter	Rpb6	Rpb5	62	171	4	11,2	NA	-
Intra	Rpb6	Rpb6	112	123	20	6,75	9.6	-
Inter	Rpb9	Tfg1	9	400	6	16,92	8.4	-
Intra	Ssl1	Ssl1	262	303	3	10,3	10.7	-
Intra	Ssl1	Ssl1	387	430	8	4,45	9.5	-
Inter	Ssl1	Tfa1	201	270	2	5,16	19.0	-
Inter	Ssl1	Tfa1	201	275	2	4,32	14.3	-
Inter	Ssl1	Tfa1	205	270	1	5,04	14.6	-
Inter	Ssl1	Tfa1	205	275	2	9,84	15.2	-
Inter	Ssl1	Tfb4	401	95	9	25,87	13.1	-
Inter	Ssl1	Tfb4	439	165	16	7,22	9.8	-
Inter	Ssl1	Tfb4	441	165	7	8,88	14.7	-
Inter	Ssl1	Tfb4	446	165	2	4,76	15.2	-
Inter	Ssl2	Med7	520	1	4	18,53	NA	-
Inter	Ssl2	Ssl1	351	52	1	3,64	NA	-
Intra	Ssl2	Ssl2	46	65	2	6,51	NA	-
Intra	Ssl2	Ssl2	90	334	6	13,37	NA	-

Intra	Ssl2	Ssl2	228	304	4	4,61	NA	-
Intra	Ssl2	Ssl2	510	520	2	10,53	25.1	-
Intra	Ssl2	Ssl2	510	521	2	3,71	23.7	-
Intra	Ssl2	Ssl2	617	624	1	6,12	10.7	-
Intra	Ssl2	Ssl2	758	827	2	3,71	NA	-
Intra	Ssl2	Ssl2	762	791	3	5,79	NA	-
Intra	Ssl2	Ssl2	762	827	2	3,82	NA	-
Intra	Ssl2	Ssl2	768	827	1	3,28	NA	-
Intra	Ssl2	Ssl2	774	835	1	3,53	NA	-
Intra	TBP	TBP	2	27	1	6,94	NA	-
Intra	TBP	TBP	2	44	1	7,29	NA	-
Intra	TBP	TBP	2	188	2	6,77	NA	-
Intra	TBP	TBP	27	35	3	4,88	NA	-
Intra	TBP	TBP	27	47	2	10,36	NA	-
Intra	TBP	TBP	47	52	1	3,83	NA	-
Intra	TBP	TBP	47	54	2	4,22	NA	-
Intra	TBP	TBP	47	108	1	9,2	NA	-
Inter	Tfa1	Rpb1	222	188	10	3,21	NA	-
Inter	Tfa1	Rpb1	222	193	18	11,9	NA	-
Inter	Tfa1	Rpb1	222	195	16	9,64	NA	-
Inter	Tfa1	Rpb1	222	196	47	9,75	NA	-
Intra	Tfa1	Tfa1	1	195	2	11,53	10.7	4,194
Inter	Tfa1	Tfb1	345	189	2	6,08	11.7	349
Inter	Tfa1	Tfb1	347	189	1	6,74	11.7	349
Inter	Tfa1	Tfb1	349	189	14	9,55	11.7	-
Inter	Tfa1	Tfb1	350	189	11	9,59	12.6	-
Inter	Tfa2	Rpb1	294	193	1	4,4	NA	-
Inter	Tfa2	Tfa1	294	195	3	9,73	NA	-
Intra	Tfa2	Tfa2	140	149	5	7,23	15.7	-
Intra	Tfa2	Tfa2	273	277	1	6,41	6.1	-
Inter	Tfa2	Tfg2	133	335	2	6,42	8.3	-
Inter	Tfb1	Tfa1	57	411	1	3,4	8.3	-
Inter	Tfb1	Tfa1	57	412	1	3,44	9.4	-
Intra	Tfb1	Tfb1	65	118	2	11,74	17.7	-
Intra	Tfb1	Tfb1	118	279	3	8,85	11.2	-
Intra	Tfb1	Tfb1	166	173	2	3,86	6.3	168
Intra	Tfb1	Tfb1	268	320	7	15,09	12.4	-
Intra	Tfb1	Tfb1	291	305	2	5,17	10.0	-
Intra	Tfb1	Tfb1	322	324	7	11,97	5.3	-
Inter	Tfb1	Tfb4	458	43	1	3,13	NA	-
Inter	Tfb1	Tfb4	636	323	8	15,83	14.4	-
Inter	Tfb2	Tfa1	449	301	7	11,53	NA	-
Intra	Tfb2	Tfb2	168	175	19	21,6	7.0	-
Inter	Tfb2	Tfb4	326	166	1	7,8	NA	-
Inter	Tfb2	Tfb5	449	51	9	12,01	12.0	-
Inter	Tfb2	Tfb5	463	1	2	7,28	7.5	2
Inter	Tfb3	Med6	154	143	4	8,37	NA	-
Inter	Tfb3	Rad3	156	357	5	6,45	NA	-
Inter	Tfb3	Rpb4	25	1	1	5,51	NA	-
Inter	Tfb3	Rpb4	226	186	32	13,88	NA	-
Inter	Tfb3	Rpb4	235	186	1	4,72	NA	-
Inter	Tfb3	Rpb7	220	134	1	3,57	NA	-
Inter	Tfb3	Ssl2	44	69	7	6,68	NA	-
Intra	Tfb3	Tfb3	27	40	1	4,1	5.8	-
Intra	Tfb3	Tfb3	57	204	1	4,11	NA	-
Intra	Tfb3	Tfb3	80	121	11	17,24	7.0	-
Intra	Tfb3	Tfb3	94	154	2	5,27	NA	-

Intra	Tfb3	Tfb3	94	156	1	3,39	NA	-
Intra	Tfb3	Tfb3	159	163	2	7,5	NA	-
Intra	Tfb3	Tfb3	180	192	4	7,51	NA	-
Intra	Tfb3	Tfb3	180	226	1	5,66	NA	-
Intra	Tfb3	Tfb3	181	192	1	6,38	NA	-
Intra	Tfb3	Tfb3	182	192	3	6,44	NA	-
Intra	Tfb3	Tfb3	182	226	3	8,35	NA	-
Intra	Tfb3	Tfb3	192	204	2	8,81	NA	-
Intra	Tfb3	Tfb3	192	207	5	9,96	NA	-
Intra	Tfb3	Tfb3	192	210	1	8,02	NA	-
Intra	Tfb3	Tfb3	204	235	1	5,41	NA	-
Intra	Tfb3	Tfb3	207	235	2	8,48	NA	-
Intra	Tfb3	Tfb3	226	265	1	5,4	NA	-
Intra	Tfb4	Tfb4	84	127	5	5	12.1	-
Intra	Tfb4	Tfb4	84	128	16	5,86	8.4	-
Intra	Tfb4	Tfb4	89	102	1	9,98	15.0	103
Intra	Tfb4	Tfb4	102	123	2	10,97	9.1	103
Intra	Tfb4	Tfb4	102	127	1	7,97	10.9	103
Intra	Tfb4	Tfb4	108	123	7	14,01	10.8	-
Intra	Tfb4	Tfb4	108	127	10	11,52	14.4	-
Inter	Tfb5	Ssl2	6	46	2	3,24	NA	-
Inter	Tfg1	Tfa1	91	269	1	3,89	NA	-
Intra	Tfg1	Tfg1	60	68	4	16,07	NA	-
Intra	Tfg1	Tfg1	60	69	3	12,16	NA	-
Intra	Tfg1	Tfg1	89	97	2	5,16	NA	-
Intra	Tfg1	Tfg1	274	411	1	12,18	NA	-
Intra	Tfg1	Tfg1	274	416	1	9,12	NA	-
Intra	Tfg1	Tfg1	280	289	8	15,2	NA	-
Intra	Tfg1	Tfg1	281	289	1	9,22	NA	-
Intra	Tfg1	Tfg1	411	419	2	9,31	7.6	415
Intra	Tfg1	Tfg1	658	671	5	14,2	NA	-
Intra	Tfg1	Tfg1	658	672	3	11,72	NA	-
Intra	Tfg1	Tfg1	671	704	2	8,68	NA	-
Inter	Tfg1	Tfg2	93	99	1	4,99	5.8	97
Inter	Tfg1	Tfg2	93	103	7	13,91	10.7	97
Inter	Tfg1	Tfg2	94	103	3	6,24	10.7	97
Inter	Tfg1	Tfg2	97	94	1	5,98	15.9	-
Inter	Tfg1	Tfg2	97	99	16	6,18	5.8	-
Inter	Tfg1	Tfg2	100	94	2	3,47	6.4	-
Inter	Tfg1	Tfg2	125	130	5	14,97	9.1	-
Inter	Tfg1	Tfg2	126	130	35	17,99	6.1	-
Inter	Tfg1	Tfg2	126	132	5	14,32	10.0	-
Inter	Tfg2	Rpb1	2	196	1	6,84	NA	-
Inter	Tfg2	Rpb1	186	196	3	8,63	NA	-
Inter	Tfg2	Tfg1	100	91	1	3,15	NA	-
Inter	Tfg2	Tfg1	103	92	7	15,12	NA	-
Inter	Tfg2	Tfg1	110	91	1	4,9	NA	-
Inter	Tfg2	Tfg1	114	91	1	3,2	NA	-
Inter	Tfg2	Tfg1	194	411	1	4,8	NA	-
Intra	Tfg2	Tfg2	23	54	1	12,75	NA	-
Intra	Tfg2	Tfg2	31	206	1	3,28	NA	-
Intra	Tfg2	Tfg2	38	142	2	5,73	NA	-
Intra	Tfg2	Tfg2	38	206	1	6,03	NA	-
Intra	Tfg2	Tfg2	54	141	6	15,33	11.5	58,138
Intra	Tfg2	Tfg2	54	142	5	9,35	11.5	58,138
Intra	Tfg2	Tfg2	79	245	1	5,75	15.5	244
Intra	Tfg2	Tfg2	80	114	1	3,59	13.6	110

Intra	Tfg2	Tfg2	126	130	1	4,38	10.1	-
Intra	Tfg2	Tfg2	308	357	1	9,57	NA	-
Intra	Tfg2	Tfg2	356	360	1	4,04	NA	-
Inter	TFIIB	Rpb2	100	470	2	9,2	NA	
Inter	TFIIB	Rpb2	184	865	8	3,85	14.2	-
Inter	TFIIB	Rpb2	224	923	4	5,8	NA	-
Inter	TFIIB	TBP	343	241	2	3,82	NA	-
Intra	TFIIB	TFIIB	75	147	1	3,12	NA	-
Intra	TFIIB	TFIIB	88	155	5	11,85	9.8	-
Intra	TFIIB	TFIIB	103	108	1	5,13	9.3	-
Intra	TFIIB	TFIIB	155	219	4	6,87	14.8	218
Intra	TFIIB	TFIIB	160	217	1	12,22	12.3	
Inter	Toa1	TBP	169	110	2	3,83	NA	-
Inter	Toa1	Tfg2	23	2	2	17,06	NA	-
Inter	Toa1	Tfg2	24	2	1	12,39	NA	-
Inter	Toa1	Tfg2	26	2	1	15,32	NA	-
Intra	Toa1	Toa1	44	48	1	5,18	5.0	47
Intra	Toa1	Toa1	144	167	2	4,02	NA	-
Inter	Toa1	Toa2	2	101	7	11,21	12.1	104
Inter	Toa1	Toa2	45	21	17	10,62	8.7	-
Inter	Toa2	Tfa2	24	20	2	7,93	NA	-
Intra	Toa2	Toa2	2	42	2	5,4	9.6	5
Intra	Toa2	Toa2	59	88	9	15,94	4.8	-

9 References

1 Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).

2 Gerace, L. & Burke, B. Functional organization of the nuclear envelope. *Annual review of cell biology* **4**, 335-374 (1988).

3 Orphanides, G. & Reinberg, D. A unified theory of gene expression. *Cell* **108**, 439-451 (2002).

4 Hombach, S. & Kretz, M. in *Non-coding RNAs in Colorectal Cancer* pp. 3-17 (Springer, 2016).

5 Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology* **9**, 85-98 (2011).

6 Asin-Cayuela, J. & Gustafsson, C. M. Mitochondrial transcription and its regulation in mammalian cells. *Trends in biochemical sciences* **32**, 111-117 (2007).

7 Yagi, Y. & Shiina, T. Recent advances in the study of chloroplast gene expression and its evolution. *Frontiers in plant science* **5** (2014).

8 Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular cell* **45**, 439-446 (2012).

9 Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15**, 394-408 (2014).

10 Hannan, K. M., Hannan, R. D. & Rothblum, L. I. Transcription by RNA polymerase I. *Proteins* **4**, 3 (1998).

11 Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. & Pagano, A. The expanding RNA polymerase III transcriptome. *TRENDS in Genetics* **23**, 614-622 (2007).

12 Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back totranscription and ahead to translation. *Cell* **136**, 688-700 (2009).

13 Schmid, M. & Jensen, T. H. Nuclear quality control of RNA polymerase II transcripts. *Wiley Interdisciplinary Reviews: RNA* **1**, 474-485 (2010).

14 Svejstrup, J. Q. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica Et Biophysica Acta (BBA)-Gene Structure and Expression* **1677**, 64-73 (2004).

15 Shandilya, J. & Roberts, S. G. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1819**, 391-400 (2012).

16 Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development* **20**, 2922-2936 (2006).

17 Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Molecular cell* **36**, 541-546 (2009).

18 Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews* **113**, 8456-8490 (2013).

19 Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature reviews Molecular cell biology* **18**, 263-273 (2017).

20   Jeronimo, C. l., Bataille, A. R. & Robert, F. o. The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical reviews* **113**, 8491-8522 (2013).

21   Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).

22   Conaway, R. C. & Conaway, J. W. General initiation factors for RNA polymerase II. *Annual review of biochemistry* **62**, 161-190 (1993).

23   Wade, J. T. & Struhl, K. The transition from transcriptional initiation to elongation. *Current opinion in genetics & development* **18**, 130-136 (2008).

24   Luse, D. & Jacob, G. A. Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. *Journal of Biological Chemistry* **262**, 14990-14997 (1987).

25   Holstege, F. C., Fiedler, U. & Timmers, H. T. M. Three transitions in the RNA polymerase II transcription complex during initiation. *The EMBO journal* **16**, 7468-7480 (1997).

26   Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics* **13**, 720-731 (2012).

27   Liu, X., Kraus, W. L. & Bai, X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends in biochemical sciences* **40**, 516-525 (2015).

28   Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nature structural & molecular biology* **14**, 796-806 (2007).

29   Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nature structural & molecular biology* **16**, 1128-1133 (2009).

30   Kireeva, M. L. *et al.* Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Molecular cell* **30**, 557-566 (2008).

31   Werner, F. A nexus for gene expression—molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of molecular biology* **417**, 13-27 (2012).

32   Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annual review of genetics* **47**, 483-508 (2013).

33   Guo, J. & Price, D. H. RNA polymerase II transcription elongation control. *Chemical reviews* **113**, 8583-8603 (2013).

34   Xu, Y. *et al.* Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nature communications* **8**, 15741 (2017).

35   Martinez–Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal* **30**, 1302-1310 (2011).

36   Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *nAture structurAl & moleculAr biology* **24**, 809-815 (2017).

37   Hantsche, M. & Cramer, P. Strukturelle Grundlage der Transkription: 10 Jahre nach dem Chemie–Nobelpreis. *Angewandte Chemie* **128**, 16204-16214 (2016).

38   Logan, J., Falck-Pedersen, E., Darnell, J. E. & Shenk, T. A poly (A) addition site and a downstream termination region are required for efficient cessation of transcription by

RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences* **84**, 8306-8310 (1987).

39   Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development* **2**, 440-452 (1988).

40   Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews Molecular cell biology* **12**, 283-294 (2011).

41   Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-192 (2009).

42   Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nature reviews Molecular cell biology* **16**, 129-144 (2015).

43   Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319**, 1791-1792 (2008).

44   Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-335 (1996).

45   Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes & development* **10**, 2657-2683 (1996).

46   Grunberg, S. & Hahn, S. Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* **38**, 603-611 (2013).

47   Hantsche, M. & Cramer, P. Conserved RNA polymerase II initiation complex structure. *Curr Opin Struct Biol* **47**, 17-22 (2017).

48   Nogales, E., Louder, R. K. & He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annual review of biophysics* **46**, 59-83 (2017).

49   Butler, J. E. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development* **16**, 2583-2592 (2002).

50   Guarente, L. UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. *Cell* **52**, 303-305 (1988).

51   Dobi, K. C. & Winston, F. Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. *Molecular and cellular biology* **27**, 5575-5586 (2007).

52   Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature reviews Molecular cell biology* **16**, 155-166 (2015).

53   Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W. & Kadonaga, J. T. The RNA polymerase II core promoter—the gateway to transcription. *Current opinion in cell biology* **20**, 253-259 (2008).

54   Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology* **1**, 40-51 (2012).

55   Lifton, R., Goldberg, M., Karp, R. & Hogness, D. in *Cold Spring Harbor symposia on quantitative biology.* 1047-1051 (Cold Spring Harbor Laboratory Press).

56   Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699-709 (2004).

57   Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301 (2012).

58   Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604-609 (2016).

59   Verrijzer, C. P., Chen, J.-L., Yokomori, K. & Tjian, R. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115-1125 (1995).

60   Burley, S. & Roeder, R. Biochemistry and structural biology of transcription factor IID (TFIID). *Annual review of biochemistry* **65**, 769-799 (1996).

61   Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).

62   Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527 (1993).

63   Nikolov, D. B. *et al.* Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences* **93**, 4862-4867 (1996).

64   Imbalzano, A. N., Kwon, H., Green, M. R. & Kingston, R. E. Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* **370**, 481-485 (1994).

65   Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *SCIENCE-NEW YORK THEN WASHINGTON-*, 830-836 (1996).

66   Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127-134 (1996).

67   Bleichenbacher, M., Tan, S. & Richmond, T. J. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *Journal of molecular biology* **332**, 783-793 (2003).

68   Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. & Ebright, R. H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & development* **12**, 34-44 (1998).

69   Littlefield, O., Korkhin, Y. & Sigler, P. B. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proceedings of the National Academy of Sciences* **96**, 13668-13673 (1999).

70   Tsai, F. T. & Sigler, P. B. Structural basis of preinitiation complex assembly on human pol II promoters. *The EMBO journal* **19**, 25-36 (2000).

71   Deng, W. & Roberts, S. G. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & development* **19**, 2418-2423 (2005).

72   Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241**, 1335 (1988).

73   Sawadogo, M. & Sentenac, A. RNA polymerase B (II) and general transcription factors. *Annual review of biochemistry* **59**, 711-754 (1990).

74   Bushnell, D. A., Westover, K. D., Davis, R. E. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**, 983-988 (2004).

75   Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323-330 (2009).

76   Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* **493**, 437-440 (2013).

77   Zhao, X. & Herr, W. A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. *Cell* **108**, 615-627 (2002).

78   Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206-209 (2010).

79   Tan, S., Aso, T., Conaway, R. C. & Conaway, J. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *Journal of Biological Chemistry* **269**, 25684-25691 (1994).

80   Čabart, P., Újvári, A., Pal, M. & Luse, D. S. 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. *Proceedings of the National Academy of Sciences* **108**, 15786-15791 (2011).

81   Fishburn, J. & Hahn, S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Molecular and cellular biology* **32**, 12-25 (2012).

82   Ghazy, M. A., Brodie, S. A., Ammerman, M. L., Ziegler, L. M. & Ponticelli, A. S. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Molecular and cellular biology* **24**, 10975-10985 (2004).

83   Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross–linking and mass spectrometry. *The EMBO journal* **29**, 717-726 (2010).

84   Eichner, J., Chen, H. T., Warfield, L. & Hahn, S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *The EMBO journal* **29**, 706-716 (2010).

85   He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481-486 (2013).

86   He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359-365 (2016).

87   Murakami, K. *et al.* Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci U S A* **112**, 13543-13548 (2015).

88   Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518**, 376-380 (2015).

89   Plaschka, C. *et al.* Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353-358 (2016).

90   Parvin, J. D. & Sharp, P. A. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* **73**, 533-540 (1993).

91   Tyree, C. M. *et al.* Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes & development* **7**, 1254-1265 (1993).

92   Pan, G. & Greenblatt, J. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *Journal of Biological Chemistry* **269**, 30101-30104 (1994).

93   Tantin, D. & Carey, M. A heteroduplex template circumvents the energetic requirement for ATP during activated transcription by RNA polymerase II. *Journal of Biological Chemistry* **269**, 17397-17400 (1994).

94   Holstege, F., Tantin, D., Carey, M., Van der Vliet, P. & Timmers, H. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *The EMBO Journal* **14**, 810 (1995).

95   Chen, H.-T., Warfield, L. & Hahn, S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature structural & molecular biology* **14**, 696-703 (2007).

96   Kim, T.-K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418-1421 (2000).

97   Forget, D., Langelier, M.-F., Thérien, C., Trinh, V. & Coulombe, B. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Molecular and cellular biology* **24**, 1122-1131 (2004).

98   Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature structural & molecular biology* **13**, 603-610 (2006).

99   Kuldell, N. H. & Buratowski, S. Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. *Molecular and cellular biology* **17**, 5288-5298 (1997).

100   Maxon, M. E., Goodrich, J. A. & Tjian, R. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev* **8**, 515-524 (1994).

101   Robinson, P. J. *et al.* Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**, 1411-1422 e1416 (2016).

102   Okuda, M. *et al.* Structural insight into the TFIIE–TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *The EMBO journal* **27**, 1161-1171 (2008).

103   Guzder, S. N., Sung, P., Bailly, V., Prakash, L. & Prakash, S. RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* **369**, 578-581 (1994).

104   Holstege, F., Van der Vliet, P. & Timmers, H. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *The EMBO journal* **15**, 1666 (1996).

105   Tirode, F., Busso, D., Coin, F. & Egly, J. M. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* **3**, 87-95 (1999).

106   Guzmán, E. & Lis, J. T. Transcription factor TFIIH is required for promoter melting in vivo. *Molecular and cellular biology* **19**, 5652-5658 (1999).

107   Fan, L. *et al.* Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Molecular cell* **22**, 27-37 (2006).

108   Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology* **19**, 788-796 (2012).

109   Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. 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 (2015).

110   Ohkuma, Y. & Roeder, R. G. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368**, 160 (1994).

111   Lin, Y. C., Choi, W. S. & Gralla, J. D. TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nature structural & molecular biology* **12**, 603-607 (2005).

112   Alekseev, S. *et al.* Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. *Mol Cell* **65**, 504-514 e504 (2017).

113   Pinto, I., Ware, D. E. & Hampsey, M. The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo. *Cell* **68**, 977-988 (1992).

114   Zhang, C., Zobeck, K. L. & Burton, Z. F. Human RNA polymerase II elongation in slow motion: role of the TFIIF RAP74 α1 helix in nucleoside triphosphate-driven translocation. *Molecular and cellular biology* **25**, 3583-3595 (2005).

115   Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *Journal of Biological Chemistry* **274**, 35668-35675 (1999).

116   Khaperskyy, D. A., Ammerman, M. L., Majovski, R. C. & Ponticelli, A. S. Functions of Saccharomyces cerevisiae TFIIF during transcription start site utilization. *Molecular and cellular biology* **28**, 3757-3766 (2008).

117   Price, D., Sluder, A. & Greenleaf, A. Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Molecular and Cellular Biology* **9**, 1465-1475 (1989).

118   Lu, H., Zawel, L., Fisher, L., Egly, J.-M. & Reinberg, D. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358**, 641-645 (1992).

119   Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell* **79**, 1103-1109 (1994).

120   Roy, R. *et al.* The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* **79**, 1093-1101 (1994).

121   Keogh, M.-C., Cho, E.-J., Podolny, V. & Buratowski, S. Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. *Molecular and cellular biology* **22**, 1288-1297 (2002).

122   Glover-Cutter, K. *et al.* TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology* **29**, 5455-5464 (2009).

123   Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature structural & molecular biology* **19**, 1108-1115 (2012).

124   Lu, H., Fisher, R. P., Bailey, P. & Levine, A. J. The CDK7-cycH-p36 complex of transcription factor IIH phosphorylates p53, enhancing its sequence-specific DNA binding activity in vitro. *Molecular and cellular biology* **17**, 5923-5934 (1997).

125   Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J.-M. & Chambon, P. Stimulation of RARα activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. *Cell* **90**, 97-107 (1997).

126   Mäkelä, T. *et al.* A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription. *Proceedings of the National Academy of Sciences* **92**, 5174-5178 (1995).

127   Valay, J.-G. *et al.* TheKIN28Gene is Required both for RNA Polymerase II Mediated Transcription and Phosphorylation of the Rpb1p CTD. *Journal of molecular biology* **249**, 535-544 (1995).

128   Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-728 (1998).

129   Liu, Y. *et al.* Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Molecular and cellular biology* **24**, 1721-1735 (2004).

130   Kim, H. *et al.* Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology* **17**, 1279-1286 (2010).

131   Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology* **17**, 1272-1278 (2010).

132   Dvir, A., Conaway, R. C. & Conaway, J. W. A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. *Proceedings of the National Academy of Sciences* **94**, 9006-9010 (1997).

133   Wong, K. H., Jin, Y. & Struhl, K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Mol Cell* **54**, 601-612 (2014).

134   Jeronimo, C. & Robert, F. Kin28 regulates the transient association of Mediator with core promoters. *Nature structural & molecular biology* **21**, 449-455 (2014).

135   Goodrich, J. A. & Tjian, R. Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**, 145-156 (1994).

136   Moreland, R. J. *et al.* A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *J Biol Chem* **274**, 22127-22130 (1999).

137   Zhou, M. *et al.* TFIIH inhibits CDK9 phosphorylation during human immunodeficiency virus type 1 transcription. *Journal of Biological Chemistry* **276**, 44633-44640 (2001).

138   Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225-229 (2000).

139   El Kaderi, B., Medler, S., Raghunayakula, S. & Ansari, A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *Journal of Biological Chemistry* **284**, 25015-25025 (2009).

140   O'Sullivan, J. M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nature genetics* **36** (2004).

141   Sarge, K. D. & Park-Sarge, O.-K. Gene bookmarking: keeping the pages open. *Trends in biochemical sciences* **30**, 605-610 (2005).

142   Singh, B. N. & Hampsey, M. A transcription-independent role for TFIIB in gene looping. *Molecular cell* **27**, 806-816 (2007).

143   Wang, Y., Fairley, J. A. & Roberts, S. G. Phosphorylation of TFIIB links transcription initiation and termination. *Current Biology* **20**, 548-553 (2010).

144   Goel, S., Krishnamurthy, S. & Hampsey, M. Mechanism of Start Site Selection by RNA Polymerase II INTERPLAY BETWEEN TFIIB AND Ssl2/XPB HELICASE SUBUNIT OF TFIIH. *Journal of Biological Chemistry* **287**, 557-567 (2012).

145   Compe, E. & Egly, J.-M. Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annual review of biochemistry* **85**, 265-290 (2016).

146   Pugh, B. F. & Tjian, R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187-1197 (1990).

147   Egly, J.-M. & Coin, F. A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. *DNA repair* **10**, 714-721 (2011).

148   Compe, E. & Egly, J. M. TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* **13**, 343-354 (2012).

149   Keriel, A., Stary, A., Sarasin, A., Rochette-Egly, C. & Egly, J.-M. XPD mutations prevent TFIIH-dependent transactivation by nuclear receptors and phosphorylation of RARα. *Cell* **109**, 125-135 (2002).

150   Ito, S. *et al.* XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Molecular cell* **26**, 231-243 (2007).

151   Iben, S. *et al.* TFIIH plays an essential role in RNA polymerase I transcription. *Cell* **109**, 297-306 (2002).

152   Oler, A. J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature structural & molecular biology* **17**, 620-628 (2010).

153   Assfalg, R. *et al.* TFIIH is an elongation factor of RNA polymerase I. *Nucleic acids research* **40**, 650-659 (2011).

154   Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes & development* **15**, 507-521 (2001).

155   Feaver, W. J. *et al.* Dual roles of a multiprotein complex from S. cerevisiae in transcription and DNA repair. *Cell* **75**, 1379-1387 (1993).

156   Schaeffer, L. *et al.* DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *SCIENCE-NEW YORK THEN WASHINGTON-* **260**, 58-58 (1993).

157   Larochelle, S., Pandur, J., Fisher, R. P., Salz, H. K. & Suter, B. Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. *Genes & development* **12**, 370-381 (1998).

158   Ganuza, M. *et al.* Genetic inactivation of Cdk7 leads to cell cycle arrest and induces premature aging due to adult stem cell exhaustion. *The EMBO journal* **31**, 2498-2510 (2012).

159   Ito, S. *et al.* MMXD, a TFIIH-independent XPD-MMS19 protein complex involved in chromosome segregation. *Molecular cell* **39**, 632-640 (2010).

160   Oh, K. S. *et al.* Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Human mutation* **27**, 1092-1103 (2006).

161   Stefanini, M., Botta, E., Lanzafame, M. & Orioli, D. Trichothiodystrophy: from basic mechanisms to clinical implications. *DNA repair* **9**, 2-10 (2010).

162   DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *Journal of Investigative Dermatology* **132**, 785-796 (2012).

163   Adamczewski, J. *et al.* MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH. *The EMBO journal* **15**, 1877 (1996).

164   Rossignol, M., Kolb-Cheynel, I. & Egly, J. M. Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. *EMBO J* **16**, 1628-1637 (1997).

165   Gibbons, B. J. *et al.* Subunit architecture of general transcription factor TFIIH. *Proc Natl Acad Sci U S A* **109**, 1949-1954 (2012).

166   Svejstrup, J. Q. *et al.* Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. *Cell* **80**, 21-28 (1995).

167   Araújo, S. J. *et al.* Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes & development* **14**, 349-359 (2000).

168   Coin, F. *et al.* Nucleotide excision repair driven by the dissociation of CAK from TFIIH. *Molecular cell* **31**, 9-20 (2008).

169   Mäkelä, T. P. *et al.* A cyclin associated with the CDK-activating kinase MO15. *Nature* **371**, 254-257 (1994).

170   Fisher, R. P. & Morgan, D. O. A novel cyclin associates with M015/CDK7 to form the CDK-activating kinase. *Cell* **78**, 713-724 (1994).

171   Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Current opinion in structural biology* **20**, 313-324 (2010).

172   Schaeffer, L. *et al.* The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *The EMBO journal* **13**, 2388 (1994).

173   Coin, F. *et al.* Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nature genetics* **20** (1998).

174   Kuper, J., Wolski, S. C., Michels, G. & Kisker, C. Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. *The EMBO journal* **31**, 494-502 (2012).

175   Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Penedo, J. C., White, M. F. & Naismith, J. H. Mechanism of DNA loading by the DNA repair helicase XPD. *Nucleic acids research* **44**, 2806-2815 (2016).

176   Svejstrup, J. Q., Vichi, P. & Egly, J.-M. The multiple roles of transcription/repair factor TFIIH. *Trends in biochemical sciences* **21**, 346-350 (1996).

177   Chen, J. & Suter, B. Xpd, a structural bridge and a functional link. *Cell Cycle* **2**, 503-506 (2003).

178   Seroz, T., Perez, C., Bergmann, E., Bradsher, J. & Egly, J.-M. p44/SSL1, the regulatory subunit of the XPD/RAD3 helicase, plays a crucial role in the transcriptional activity of TFIIH. *Journal of Biological Chemistry* **275**, 33260-33266 (2000).

179   Kim, J. S. *et al.* Crystal structure of the Rad3/XPD regulatory domain of Ssl1/p44. *Journal of Biological Chemistry* **290**, 8321-8330 (2015).

180   Takagi, Y. *et al.* Ubiquitin ligase activity of TFIIH and the transcriptional response to DNA damage. *Molecular cell* **18**, 237-243 (2005).

181   Schmitt, D. R., Kuper, J., Elias, A. & Kisker, C. The structure of the TFIIH p34 subunit reveals a von Willebrand factor A like fold. *PloS one* **9**, e102389 (2014).

182   Kellenberger, E. *et al.* Solution structure of the C-terminal domain of TFIIH P44 subunit reveals a novel type of C4C4 ring domain involved in protein-protein interactions. *Journal of Biological Chemistry* **280**, 20785-20792 (2005).

183   Di Lello, P. *et al.* NMR structure of the amino-terminal domain from the Tfb1 subunit of TFIIH and characterization of its phosphoinositide and VP16 binding sites. *Biochemistry* **44**, 7678-7686 (2005).

184   Chabot, P. R. *et al.* Structural and functional characterization of a complex between the acidic transactivation domain of EBNA2 and the Tfb1/p62 subunit of TFIIH. *PLoS pathogens* **10**, e1004042 (2014).

185   Okuda, M. & Nishimura, Y. Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. *Journal of the American Chemical Society* **136**, 14143-14152 (2014).

186   Okuda, M., Araki, K., Ohtani, K. & Nishimura, Y. The Interaction Mode of the Acidic Region of the Cell Cycle Transcription Factor DP1 with TFIIH. *Journal of molecular biology* **428**, 4993-5006 (2016).

187   Lecoq, L. *et al.* Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF-κB and transcription regulatory factors. *Nucleic acids research* **45**, 5564-5576 (2017).

188   Okuda, M., Kinoshita, M., Kakumu, E., Sugasawa, K. & Nishimura, Y. Structural insight into the mechanism of TFIIH recognition by the acidic string of the nucleotide excision repair factor XPC. *Structure* **23**, 1827-1837 (2015).

189   Okuda, M., Nakazawa, Y., Guo, C., Ogi, T. & Nishimura, Y. Common TFIIH recruitment mechanism in global genome and transcription-coupled repair subpathways. *Nucleic acids research* (2017).

190   Luo, J. *et al.* Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Molecular cell* **59**, 794-806 (2015).

191   Coin, F. *et al.* p8/TTD-A as a repair-specific TFIIH subunit. *Molecular cell* **21**, 215-226 (2006).

192   Kainov, D. E., Vitorino, M., Cavarelli, J., Poterszman, A. & Egly, J.-M. Structural basis for group A trichothiodystrophy. *Nature structural & molecular biology* **15**, 980-984 (2008).

193   Conaway, R. C. & Conaway, J. W. An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *Proceedings of the National Academy of Sciences* **86**, 7356-7360 (1989).

194   Weber, C. A., Salazar, E. P., Stewart, S. A. & Thompson, L. H. ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. *The EMBO journal* **9**, 1437 (1990).

195   Weeda, G., Ma, L., van Ham, R. C., Van der Eb, A. J. & Hoeijmakers, J. H. Sturcture and expression of the human XPBC/ERCC-3 gene involved in DNA repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Nucleic acids research* **19**, 6301-6308 (1991).

196   Feaver, W. J. *et al.* Genes For Tfb2, Tfb3, and Tfb4 subunits of yeast transcription/repair factor IIH homology to human cyclin-dependent kinase activating kinase and IIH subunits. *Journal of Biological Chemistry* **272**, 19319-19327 (1997).

197   Marinoni, J. C. *et al.* Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH. *The EMBO Journal* **16**, 1093-1102 (1997).

198 Takagi, Y. *et al.* Revised subunit structure of yeast transcription factor IIH (TFIIH) and reconciliation with human TFIIH. *Journal of Biological Chemistry* **278**, 43897-43900 (2003).

199 Giglia-Mari, G. *et al.* A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature genetics* **36** (2004).

200 Ranish, J. A. *et al.* Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nature genetics* **36** (2004).

201 Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).

202 Murakami, K. *et al.* Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proceedings of the National Academy of Sciences* **109**, 4816-4821 (2012).

203 Jawhari, A. *et al.* p52 Mediates XPB function within the transcription/repair factor TFIIH. *Journal of Biological Chemistry* **277**, 31761-31767 (2002).

204 Abdulrahman, W. *et al.* ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities. *Proceedings of the National Academy of Sciences* **110**, E633-E642 (2013).

205 Schultz, P. *et al.* Molecular structure of human TFIIH. *Cell* **102**, 599-607 (2000).

206 Chang, W. H. & Kornberg, R. D. Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell* **102**, 609-613 (2000).

207 Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *cell* **133**, 789-800 (2008).

208 Wolski, S. C. *et al.* Crystal structure of the FeS cluster–containing nucleotide excision repair helicase XPD. *PLoS biology* **6**, e149 (2008).

209 Hilario, E., Li, Y., Nobumori, Y., Liu, X. & Fan, L. Structure of the C-terminal half of human XPB helicase and the impact of the disease-causing mutation XP11BE. *Acta Crystallographica Section D: Biological Crystallography* **69**, 237-246 (2013).

210 Gervais, V. *et al.* Solution Structure of the N-terminal Domain of the Human TFIIH MAT1 Subunit NEW INSIGHTS INTO THE RING FINGER FAMILY. *Journal of Biological Chemistry* **276**, 7457-7464 (2001).

211 Lolli, G., Lowe, E. D., Brown, N. R. & Johnson, L. N. The crystal structure of human CDK7 and its protein recognition properties. *Structure* **12**, 2067-2079 (2004).

212 Kim, K. K., Chamberlin, H. M., Morgan, D. O. & Kim, S.-H. Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase. *Nature Structural & Molecular Biology* **3**, 849-855 (1996).

213 Schneider, E. *et al.* The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder. *Journal of molecular biology* **412**, 251-266 (2011).

214 Busso, D. *et al.* Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. *Journal of Biological Chemistry* **275**, 22815-22823 (2000).

215 Coin, F., Oksenych, V. & Egly, J.-M. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Molecular cell* **26**, 245-256 (2007).

216 Warfield, L., Luo, J., Ranish, J. & Hahn, S. Function of Conserved Topological Regions within the Saccharomyces cerevisiae Basal Transcription Factor TFIIH. *Molecular and cellular biology* **36**, 2464-2475 (2016).

217 Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics* **11**, 761-772 (2010).

218 Malik, S. & Roeder, R. G. Mediator: a drawbridge across the enhancer-promoter divide. *Molecular cell* **64**, 433-434 (2016).

219 Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *J Biol Chem* **281**, 80-89 (2006).

220 Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).

221 Bourbon, H.-M. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic acids research* **36**, 3993-4008 (2008).

222 Asturias, F. J., Jiang, Y. W., Myers, L. C., Gustafsson, C. M. & Kornberg, R. D. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* **283**, 985-987 (1999).

223 Taatjes, D. J., Näär, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* **295**, 1058-1062 (2002).

224 Jeronimo, C. & Robert, F. The Mediator Complex: At the Nexus of RNA Polymerase II Transcription. *Trends in cell biology* (2017).

225 Cevher, M. A. *et al.* Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature structural & molecular biology* **21**, 1028-1034 (2014).

226 Meyer, K. D., Lin, S.-c., Bernecky, C., Gao, Y. & Taatjes, D. J. p53 activates transcription by directing structural shifts in Mediator. *Nature structural & molecular biology* **17**, 753-760 (2010).

227 Taatjes, D. J. The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in biochemical sciences* **35**, 315-322 (2010).

228 Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361-1375 (1993).

229 Elmlund, H. *et al.* The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proceedings of the National Academy of Sciences* **103**, 15788-15793 (2006).

230 Knuesel, M. T., Meyer, K. D., Bernecky, C. & Taatjes, D. J. The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes & development* **23**, 439-451 (2009).

231 Fryer, C. J., White, J. B. & Jones, K. A. Mastermind recruits CycC: CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Molecular cell* **16**, 509-520 (2004).

232 Meyer, K. D. *et al.* Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *The EMBO journal* **27**, 1447-1457 (2008).

233   Cai, G., Imasaki, T., Takagi, Y. & Asturias, F. J. Mediator structural conservation and implications for the regulation mechanism. *Structure* **17**, 559-567 (2009).

234   Tsai, K.-L. *et al.* Mediator structure and rearrangements required for holoenzyme formation. *Nature* (2017).

235   Ebmeier, C. C. & Taatjes, D. J. Activator-Mediator binding regulates Mediator-cofactor interactions. *Proceedings of the National Academy of Sciences* **107**, 11283-11288 (2010).

236   Johnson, K. M. & Carey, M. Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. *Current biology* **13**, 772-777 (2003).

237   Baek, H. J., Kang, Y. K. & Roeder, R. G. Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *Journal of biological chemistry* **281**, 15172-15181 (2006).

238   Esnault, C. *et al.* Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Molecular cell* **31**, 337-346 (2008).

239   Svejstrup, J. Q. *et al.* Evidence for a mediator cycle at the initiation of transcription. *Proceedings of the National Academy of Sciences* **94**, 6075-6078 (1997).

240   Nair, D., Kim, Y. & Myers, L. C. Mediator and TFIIH govern carboxyl-terminal domain-dependent transcription in yeast extracts. *Journal of biological chemistry* **280**, 33739-33748 (2005).

241   Max, T., Søgaard, M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *Journal of Biological Chemistry* **282**, 14113-14120 (2007).

242   Jeronimo, C., Collin, P. & Robert, F. The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *Journal of molecular biology* **428**, 2607-2622 (2016).

243   Conaway, R. C. & Conaway, J. W. The Mediator complex and transcription elongation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1829**, 69-75 (2013).

244   Eyboulet, F. *et al.* Mediator links transcription and DNA repair by facilitating Rad2/XPG recruitment. *Genes & development* **27**, 2549-2562 (2013).

245   Yin, J.-w. & Wang, G. The Mediator complex: a master coordinator of transcription and cell lineage development. *Development* **141**, 977-987 (2014).

246   Tsai, K.-L. *et al.* Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* **157**, 1430-1444 (2014).

247   Nozawa, K., Schneider, T. R. & Cramer, P. Core Mediator structure at 3.4 A extends model of transcription initiation complex. *Nature* **545**, 248-251 (2017).

248   Larivière, L. *et al.* Structure of the Mediator head module. *Nature* **492**, 448-451 (2012).

249   Larivière, L., Seizl, M. & Cramer, P. A structural perspective on Mediator function. *Current opinion in cell biology* **24**, 305-313 (2012).

250   Tsai, K.-L. *et al.* A conserved Mediator–CDK8 kinase module association regulates Mediator–RNA polymerase II interaction. *Nature structural & molecular biology* **20**, 611-619 (2013).

251   Larivière, L. *et al.* Model of the Mediator middle module based on protein cross-linking. *Nucleic acids research* **41**, 9266-9273 (2013).

252   Robinson, P. J. *et al.* Molecular architecture of the yeast Mediator complex. *Elife* **4**, e08719 (2015).

253   Nikolov, D. B. *et al.* Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119-128 (1995).

254   Bagby, S. *et al.* Solution structure of the C-terminal core domain of human TFIIB: similarity to cyclin A and interaction with TATA-binding protein. *Cell* **82**, 857-867 (1995).

255   Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* **292**, 1863-1876 (2001).

256   Armache, K.-J., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *Journal of Biological Chemistry* **280**, 7131-7134 (2005).

257   Chen, H.-T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169-180 (2004).

258   Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443-1444 (2014).

259   Cai, G. *et al.* Interaction of the mediator head module with RNA polymerase II. *Structure* **20**, 899-910 (2012).

260   Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell* **34**, 710-721 (2009).

261   Vaughn, J., Goodwin, R., Tompkins, G. & McCawley, P. The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro Cellular & Developmental Biology-Plant* **13**, 213-217 (1977).

262   Wickham, T., Davis, T., Granados, R., Shuler, M. & Wood, H. Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnology progress* **8**, 391-396 (1992).

263   Mayer, A. *et al.* CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**, 1723-1725 (2012).

264   Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780-1782 (2007).

265   Romier, C. *et al.* Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallographica Section D: Biological Crystallography* **62**, 1232-1242 (2006).

266   Perrakis, A. & Romier, C. Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview. *Structural Proteomics: High-Throughput Methods*, 247-256 (2008).

267   Gradia, S. D. *et al.* MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein Complexes. *Methods in Enzymology* (2017).

268   Knight, T. Idempotent vector design for standard assembly of biobricks. (MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 2003).

269   Barry, G. F. A broad-host-range shuttle system for gene insertion into the chromosomes of Gram-negative bacteria. *Gene* **71**, 75-84 (1988).

270   Waddell, C. S. & Craig, N. L. Tn7 transposition: recognition of the attTn7 target sequence. *Proceedings of the National Academy of Sciences* **86**, 3958-3962 (1989).

271   Luckow, V. A., Lee, S., Barry, G. & Olins, P. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *Journal of virology* **67**, 4566-4579 (1993).

272   Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nature biotechnology* **22**, 1583-1587 (2004).

273   Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus–based eukaryotic protein complex production. *Current protocols in protein science*, 5.20. 21-25.20. 26 (2008).

274   Rohrmann, G. F. Baculovirus molecular biology. (2013).

275   Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification* **41**, 207-234 (2005).

276   Gasteiger, E. *et al. Protein identification and analysis tools on the ExPASy server*. (Springer, 2005).

277   Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical chemistry* **68**, 850-858 (1996).

278   Cottrell, J. S. & London, U. Probability-based protein identification by searching sequence databases using mass spectrometry data. *electrophoresis* **20**, 3551-3567 (1999).

279   Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods* **5**, 53-55 (2008).

280   Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic acids research* **45**, D158-D169 (2017).

281   Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic acids research* **41**, W29-W33 (2013).

282   Sievers, F. *et al.* Fast, scalable generation of high–quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).

283   Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).

284   Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* **41**, W349-W357 (2013).

285   Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic acids research* **43**, W389-W394 (2015).

286   Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).

287   Bond, C. S. & Schuettelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallographica Section D: Biological Crystallography* **65**, 510-512 (2009).

288   Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-W324 (2014).

289   Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins: Structure, Function, and Bioinformatics* **77**, 128-132 (2009).

290   Alva, V., Nam, S.-Z., Söding, J. & Lupas, A. N. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic acids research* **44**, W410-W415 (2016).

291   Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome–Chd1 structure and implications for chromatin remodelling. *Nature* **550**, 539-542 (2017).

292   Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204-209 (2017).

293   Jakhanwal, S., Lee, C. T., Urlaub, H. & Jahn, R. An activated Q-SNARE/SM protein complex as a possible intermediate in SNARE assembly. *The EMBO Journal* **36**, 1788-1802 (2017).

294   Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nature methods* **9**, 904-906 (2012).

295   Combe, C. W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Molecular & Cellular Proteomics* **14**, 1137-1147 (2015).

296   Kosinski, J. *et al.* Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *Journal of structural biology* **189**, 177-183 (2015).

297   Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612 (2004).

298   Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**, 38-46 (2007).

299   Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

300   Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, e18722 (2016).

301   Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* (2017).

302   Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* **193**, 1-12 (2016).

303   Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35 (2013).

304   Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738 (2010).

305   Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**, 7-8 (2015).

306   Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42**, W252-W258 (2014).

307   Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols* **4**, 1-13 (2009).

308   Raman, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function, and Bioinformatics* **77**, 89-99 (2009).

309   Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-1742 (2013).

310   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).

311   Miwa, K. *et al.* Crystal Structure of Human General Transcription Factor TFIIE at Atomic Resolution. *J Mol Biol* **428**, 4258-4266 (2016).

312   Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221 (2010).

313   Ye, J., Kandegedara, A., Martin, P. & Rosen, B. P. Crystal structure of the Staphylococcus aureus pI258 CadC Cd (II)/Pb (II)/Zn (II)-responsive repressor. *Journal of bacteriology* **187**, 4214-4221 (2005).

314   Fadden, A. J. *et al.* A winged helix domain in human MUS81 binds DNA and modulates the endonuclease activity of MUS81 complexes. *Nucleic acids research* **41**, 9741-9752 (2013).

315   Ahmad, M. U. D. *et al.* Structural insights into nonspecific binding of DNA by TrmBL2, an archaeal chromatin protein. *Journal of molecular biology* **427**, 3216-3229 (2015).

316   Tempel, W. *et al.* Structural genomics of Pyrococcus furiosus: X-ray crystallography reveals 3D domain swapping in rubrerythrin. *PROTEINS: Structure, Function, and Bioinformatics* **57**, 878-882 (2004).

317   Xu, D. & Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge–based force field. *Proteins: Structure, Function, and Bioinformatics* **80**, 1715-1735 (2012).

318   Baker, M. L. *et al.* Modeling protein structure at near atomic resolutions with Gorgon. *Journal of structural biology* **174**, 360-373 (2011).

319   Baker, M. L., Baker, M. R., Hryc, C. F., Ju, T. & Chiu, W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* **97**, 655-668 (2012).

320   Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38 (1996).

321   Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673-683 (2008).

322   Bardwell, A. J. *et al.* Yeast nucleotide excision repair proteins Rad2 and Rad4 interact with RNA polymerase II basal transcription factor b (TFIIH). *Mol Cell Biol* **14**, 3569-3576 (1994).

323   Feaver, W. J. *et al.* Subunit Interactions in Yeast Transcription/Repair Factor TFIIH REQUIREMENT FOR Tfb3 SUBUNIT IN NUCLEOTIDE EXCISION REPAIR. *Journal of Biological Chemistry* **275**, 5941-5946 (2000).

324   Pinhero, R., Liaw, P. & Yankulov, K. A uniform procedure for the purification of CDK7/CycH/MAT1, CDK8/CycC and CDK9/CycT1. *Biological procedures online* **6**, 163-172 (2004).

325   Kuper, J. *et al.* In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS biology* **12**, e1001954 (2014).

326   Doerks, T., Huber, S., Buchner, E. & Bork, P. BSD: a novel domain in transcription factors and synapse-associated proteins. *Trends in biochemical sciences* **27**, 168-170 (2002).

327   Edwards, A. M., Kane, C., Young, R. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *Journal of Biological Chemistry* **266**, 71-75 (1991).

328   Serizawa, H., Conaway, J. W. & Conaway, R. C. Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. *Nature* **363**, 371 (1993).

329   Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research* **38**, W529-W533 (2010).

330   Hwang, J. R. *et al.* A 3′→ 5′ XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription. *Journal of Biological Chemistry* **271**, 15898-15904 (1996).

331   Wigley, D. B. & Bowman, G. D. A glimpse into chromatin remodeling. *Nature Structural & Molecular Biology* **24**, 498-500 (2017).

332   Gu, M. & Rice, C. M. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proceedings of the National Academy of Sciences* **107**, 521-528 (2010).

333   Greber, B. J. *et al.* The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414-417 (2017).

334   Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences* **104**, 12955-12961 (2007).

335   Johnson, K. M., Wang, J., Smallwood, A., Arayata, C. & Carey, M. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & Development* **16**, 1852-1863 (2002).

336   Nogales, E., Patel, A. B. & Louder, R. K. Towards a mechanistic understanding of core promoter recognition from cryo-EM studies of human TFIID. *Current Opinion in Structural Biology* **47**, 60-66 (2017).

337   Koutelou, E., Hirsch, C. L. & Dent, S. Y. Multiple faces of the SAGA complex. *Current opinion in cell biology* **22**, 374-382 (2010).

338   Baptista, T. *et al.* SAGA is a general cofactor for RNA polymerase II transcription. *Molecular cell* **68**, 130-143. e135 (2017).

339   Warfield, L. *et al.* Transcription of nearly all yeast RNA polymerase II-transcribed genes is dependent on transcription factor TFIID. *Molecular cell* **68**, 118-129. e115 (2017).

340   Taatjes, D. J. The Continuing SAGA of TFIID and RNA Polymerase II Transcription. *Molecular cell* **68**, 1-2 (2017).

341   Tomko, E. J., Fishburn, J., Hahn, S. & Galburt, E. A. TFIIH generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning. *Nature Structural and Molecular Biology* **24**, 1139 (2017).

342   Myers, L. C. *et al.* The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes & development* **12**, 45-54 (1998).

343   Robinson, P. J., Bushnell, D. A., Trnka, M. J., Burlingame, A. L. & Kornberg, R. D. Structure of the mediator head module bound to the carboxy-terminal domain of RNA

polymerase II. *Proceedings of the National Academy of Sciences* **109**, 17931-17935 (2012).

344 Tóth-Petróczy, Á. *et al.* Malleable Machines in Transcription Regulation: The Mediator Complex. *PLoS Computational Biology* **4** (2008).

345 Cameron, A. J., Squire, C. J., Edwards, P. J., Harjes, E. & Sarojini, V. Crystal and NMR Structures of a Peptidomimetic β–Turn That Provides Facile Synthesis of 13–Membered Cyclic Tetrapeptides. *Chemistry–An Asian Journal*.

346 Zhang, J. *et al.* Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature* **551**, 57-63 (2017).

List of Figures

List of Tables

List of Abbreviations

3C	rhinovirus protease		dNTP	deoxynucleoside triphosphate
°C	degree celsius		DNA	deoxyribonucleic acid
Å	Ångstrom		DNase	deoxyribonuclease
β-Me	β-mercaptoethanole		DPA	day of proliferation arrest
A	absorbance value; ampere		DPBS	Dulbecco's phosphate buffered saline
aa	amino acid(s)		DRD	(DNA) damage recognition domain
AcNPV	*Autographa californica nuclear polyhedrosis virus*		DTT	dithiothreitol
A. fulgidis, Af	*Archaeoglobus fulgidis*		dsDNA	double stranded deoxyribonucleic acid
AMP-PNP	adenosine 5'(β,γ-imido) triphosphate (lithium salt hydrate)		e⁻	electron(s)
AP	alkaline phosphatase		E	weighted extinction coefficient; elution
ApE	A plasmid Editor		*E. coli*	*Escherichia coli*
Arg	arginine		EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
Amp	ampicillin		EDTA	(ethylenedinitrilo)tetraacetic acid
ATP	adenosine triphosphate		EM	electron microscopy
attTn7	Tn7 transposon attachment site		EMDB	electron microscopy data bank
bp	base pair(s)		et al.	et alii
BLAST	Basic Local Alignment Search Tool		EtOH	ethanol
BRE	TFIIB recognition element		EYFP	enhanced yellow fluorescent protein
BS3	bis(sulfosuccinimidyl)suberate		eZnF	extended zinc finger
BSA	bovine serum albumin		F, fw	forward
BSD	BTF2-like, synapse associated and DOS2-like		FeS	iron-sulfur (cluster)
c	centi; concentration		FDR	false discovery rate
CAK	cyclin-dependent kinase activating kinase		FSC	Fourier shell correlation
CC	closed complex		g	gram(s)
CCD	charge-coupled device		*g*	standard gravity
CDK	cyclin-dependent kinase		Gen	gentamycin
cMed	core Mediator		GTF	general transcription factor
cPIC	core pre-initiation complex		GraFix	gradient fixation
CS	Cockayne syndrome		GST	glutathione S-transferase
CTD	C-terminal domain		h	hour(s)
C-terminus	carboxy-terminus		H₂O	water
CTF	contrast transfer function		HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
CV	column volume(s)		Hi5	High Five
CycC	cyclin C		His	histidine
CycH	cyclin H		*H. sapiens, Hs*	*Homo sapiens*
Da	Dalton		HRP	horseradish peroxidase
dATP	deoxyadenosine triphosphate		HTH	helix-turn-helix
dCTP	deoxycytidine triphosphate		i.e.	id est
dd	double distilled		IgG	immunoglobulin G
ddd	triple distilled		IPTG	isopropyl-β-D-1-thiogalactopyranoside
dGTP	deoxyguanosine triphosphate			
DMSO	dimethyl sulfoxide			

ITC	initially transcribing complex		pH	potential of hydrogen
k	kilo		PHD	pleckstrin homology domain
Kan	kanamycin		PI	protease inhibitor
kb	kilobases		PIC	pre-initiation complex
L, l	liter(s)		PMSF	phenylmethylsulfonyl fluoride
lacO	lac operon		Pol	DNA-dependent RNA polymerase
LB	lysogeny broth/Luria-Bertani broth		PolH	polyhedrin
LIC	ligation independent cloning		polyA	polyadenylation
m	milli		PTM	post-translational modification
M	mega; molar (mol/liter)		PVDF	polyvinylidene difluoride
μ	micro		r, rv	reverse
MAT1	ménage-à-trois 1		RBS	ribosomal binding site
MBP	maltose-binding protein		RING	really important new gene
MCN	Multi-Cloning and expressioN		RNA	ribonucleic acid
MCS	multiple cloning site		RNAse	ribonuclease
MES	2-(N-morpholino)ethanesulfonic acid		RNP	ribonucleoprotein
min	minute(s)		r.p.m.	revolutions per minute
MOPS	3-(N-morpholino)propanesulfonic acid		rRNA	ribosomal ribonucleic acid
mRNA	messenger ribonucleic acid		RTH	Round-The-Horn
MS	mass spectrometry		s	second(s)
MSA	multiple sequence alignments		*S. aureus*	*Staphylococcus aureus*
MT	melting temperature		*S. cerevisiae, Sc*	*Saccharomyces cerevisiae*
MW	molecular weight		SBAT	1,1'-(suberoyldioxy) bisazabenzotriazol
MWCO	molecular weight cutoff		SDS	sodium dodecyl sulfate
n	nano		SEC	size exclusion chromatography
N_2	nitrogen		Ser	serine
NCBI	National Center for Biotechnology Information		SF2	superfamily 2
Ni-NTA	nickel-nitrilotriacetic acid		*S. frugiperda, Sf*	*Spodoptera frugiperda*
NER	nucleotide excision repair			
NMR	nuclear magnetic resonance		SGD	Saccharomyces Genome Database
nt	nucleotide(s)		SN	supernatant
NTE	N-terminal extension		snRNA	small nuclear ribonucleic acid
N-terminus	amino-terminus		snoRNA	small nucleolar ribonucleic acid
NTP	nucleoside triphosphate		*S. pombe, Sp*	*Schizosaccharomyces pombe*
OC	open complex		ssDNA	single stranded deoxyribonucleic acid
OD_{600}	optical density at 600 nm		SSE	seconday structure element
O/N	overnight		Ssl	suppressor of stem-loop mutation
ORF	open reading frame		Strep	streptomycin
p	pico		SV40	simian-virus 40
P	phosphor(ylation); pellet		SWI/SNF	SWItch/sucrose non-fermentable
Paf1	polymerase associated factor 1		T	temperature
PAGE	polyacrylamide gel electrophoresis		*T. acidophilum, Tac*	*Thermoplasma acidophilum*
PBS	phosphate buffered saline			
PCR	polymerase chain reaction		TAE	tris-acetate-EDTA
PDB	protein data bank		TAF	TBP-associated factor
P. furiosus, Pf	*Pyrococcus furiosus*		TBP	TATA-box binding protein

TCEP	tris(2-carboxyethyl)phosphine hydrochloride
term	terminal
TEV	tobacco etch virus
TF	transcription factor
Tfa	transcription factor A
Tfb	transcription factor B
T. ni, Tn	*Trichoplusia ni*
Tn	transposon
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TSS	transcription start site
TTD	trichothiodystrophy
Tyr	tyrosine
U	unit(s)
UAS	upstream activating sequence
UV	ultraviolet
v	volume
V	volt(s)

v/v	volume per volume
vWA	von Willebrand factor A
w	weight
WCE	whole cell extract
WH	winged helix
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XL	crosslink(ing)
XP	xeroderma pigmentosum
XPB	xeroderma pigmentosum group B
XPC	xeroderma pigmentosum group C (complementing) protein
XPD	xeroderma pigmentosum group D
YFP	yellow fluorescent protein
YPD	yeast extract peptone dextrose
Zn	zinc
ZnF	zinc finger