Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structure of the closed Pol II transcription initiation complex and implications for promoter opening

Merle Aminata Hantsche
aus
Göttingen, Deutschland

2016

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Structure of the closed Pol II transcription initiation complex and implications for promoter opening

Merle Aminata Hantsche
aus
Göttingen, Deutschland

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Göttingen, den 19.09.2016

Merle Hantsche

Dissertation eingereicht am 23.09.2016

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 14.11.2016

"Our heads are round so our thoughts can change direction."

- Francis Picabia

Acknowledgements

I am deeply grateful to Patrick Cramer for giving me the opportunity to work on a challenging and exiting project in an outstanding scientific environment. I very enjoyed the mix of invaluable advice and freedom he gave me to shape the project and my scientific development.

I thank all past and present members of the Cramer group from Munich and Göttingen for the warm and motivating atmosphere that makes the lab so special.

Several people from the Cramer group were crucial for the success of this thesis. Special thanks to Clemens Plaschka for the fruitful collaboration on promoter opening, Kerstin Kinkelin for initial Pol II-TFIIF crystallisation experiments and for getting me started with lab work, Alan Cheung and Sarah Sainsbury for introducing me to crystallography, Carrie Bernecky and Clemens Plaschka for teaching me everything about cryo-EM and figure preparation using Chimera, Simon Neyer, Carrie Bernecky and Clemens Plaschka for help with cryo-EM data collection, Carina Burzinski for protein purification of Pol II and the transcription factors and Sarah Sainsbury and Christian Dienemann for many discussions and sharing ideas about transcription initiation and experimental details.

I want to thank the Baumeister Department of the MPI in Martinsried, especially Jürgen Plitzko and Julia Mahamid, for access to the Titon Krios and technical support. I am also grateful to Achilleas Frangakis from the Buchmann Institute in Frankfurt for access to the Titon Krios and to the Loewe Cluster. Data processing on the Loewe Cluster was only possible with the support of Uwe Tangen who helped me getting the software running. I want to thank the members of the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), especially Christian Böhme and Tim Ehlers, for setting up and maintaining the computer cluster of the Cramer group. I also thank my internship students Franziska Boneberg and Sebastian Künert for their help. Many thanks to Martin Grininger, Margaux Michel and Sarah Sainsbury for critical reading of this thesis.

I am thankful to the members of my thesis committee: Dr. Dietmar Martin, Prof. Dr. Andreas Ladurner, Prof. Dr. Roland Beckmann, Dr. Franz Herzog and Prof. Dr. Klaus Förstemann for their time and support.

I want to thank my parents for their continuous support and advice and for always encouraging me to think outside the box. Ultimately, I thank Martin for his love.

Summary

Transcription is the first step of gene expression in all living cells. Regulatory mechanisms of transcription are fundamental for cell differentiation, organism development and adaption to environmental changes. One key event is transcription initiation. In eukaryotic cells RNA polymerase II (Pol II) transcribes messenger RNA and assembles with the general transcription factors (TF) -IIA, -IIB, -IID, -IIF, -IIE and -IIH on promoter DNA. DNA is opened and a 'transcription bubble' is formed that allows Pol II to synthesise a complementary copy of the genetic information. The molecular mechanisms of promoter assembly and opening remain poorly understood owing to the limited resolution of previous structural studies, the large size of the complex and the dynamics of the process.

We report the cryo-electron microscopy structure of a closed transcription initiation complex containing *Saccharomyces cerevisiae* Pol II and the general transcription factors except TFIIH on double-stranded promoter DNA at 8.8 Å resolution. Additionally, we show in a separate crystallographic study that the yeast-specific N-terminus of TFIIF subunit Tfg1 binds to the Pol II external 1 region. A high-resolution structure of the respective open complex at 3.6 Å resolution containing a 15 nucleotide mismatch transcription bubble served for model building of the closed complex. The open complex structure reveals detailed information on the intricate interactions of the general transcription factors with each other, with Pol II and promoter DNA and it suggests a mechanism of DNA template strand loading into the Pol II active centre cleft. In the transition from closed to open complex formation we identify movements mainly in TFIIE. DNA opening occurs around the tip of the Pol II clamp and the TFIIE extended winged helix domain. In functional assays we show that the TFIIE extended winged helix domain and recruitment of TFIIE through its E-ribbon domain are important for transcription *in vitro* and *in vivo*.

Moreover, we report that promoter opening can occur in the absence of the ATP-dependent factor TFIIH. Analysis of the closed complex data reveals that most of the particles (72 %) resemble the open complex structure and contain open promoter DNA. Based on our data we propose a general model how promoter opening can be achieved by the use of binding energy alone. Finally, our results underline the high structural conservation between the human and yeast transcription initiation systems.

Publications

Part of this work has been published:

M. Hantsche and P. Cramer. (2016) The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angewandte Chemie International Edition.* doi: 10.1002/anie.201608066R1 and 10.1002/ange.201608066R1

Author contributions: MH and PC prepared the manuscript.

C. Plaschka*, **M. Hantsche***, C. Dienemann, C. Burzinski, J. Plitzko, P. Cramer. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**(7603): 353-358. doi: 10.1038/nature17990 * These authors contributed equally to this work.

Author contributions: CP designed and carried out high-resolution cryo-EM structure determinations of OC1-OC4. MH designed and carried out Pol II-TFIIF crystallographic analysis, and cryo-EM structure determinations of OC5 and CC. CP and MH designed and carried out functional assays. CD cloned and purified full-length TBP and TFIIA. CD and CB assisted with protein purification. JP supervised EM data collection. PC designed and supervised research. CP, MH, and PC prepared the manuscript.

W. Mühlbacher, S. Sainsbury, M. Hemann, **M. Hantsche**, S. Neyer, F. Herzog, P. Cramer. (2014) Conserved architecture of the core RNA polymerase II initiation complex. *Nature Communications* **5**:4310. doi: 10.1038/ncomms5310

Author contributions: W.M. performed experiments. S.S. and W.M. analysed data. M.He and F.H. collected and validated MS data. M.Ha modelled the yeast TFIIF dimerisation and winged helix modules and the yeast TFIIB C-terminal cyclin domain. S.N. programmed the MATLAB script used to generate Fig. 1d. P.C. designed research and wrote the manuscript, with help from all authors.

Additional publication:

S. Neyer, M. Kunz, C. Geiss, **M. Hantsche**, V.-V. Hodirnau, A. Seybert, C. Engel, M. P. Scheffer, P. Cramer, A. S. Frangakis. (2016) Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature* (advanced online publication). doi: 10.1038/nature20561

> *Author contributions:* S.N. planned and carried out the single particle sample preparation, data collection and data analysis. M.K. planned and carried out the tomographic data analysis. C.G. carried out the sample preparation for tomography. M.H. advised on structure determination procedures. V.V.H. advised on and carried out sample preparation for tomography. A.S. advised on sample preparation for tomography. C.E. advised on biochemical procedures. M.P.S. advised on tomographic data analysis. P.C. designed and supervised research, and supervised single particle structure determination. A.S.F. designed and supervised research, supervised single particle data collection and performed tomographic data collection and analysis. S.N., P.C. and A.S.F. prepared the manuscript, with contributions from all authors.

Contents

Part I

Introduction

Parts of this section have been published:

M. Hantsche and P. Cramer. (2016) The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angewandte Chemie International Edition.* doi: 10.1002/anie.201608066R1

Ten years ago, in 2006, the Nobel Prize in Chemistry was awarded to Roger Kornberg for "… studies of the molecular basis of eukaryotic transcription" (Kornberg, 2007). Kornberg and colleagues solved the first structure of a eukaryotic RNA polymerase, the ten-subunit RNA polymerase II (Pol II) core enzyme from the baker's yeast *Saccharomyces cerevisiae* (Cramer et al., 2000, 2001; Gnatt et al., 2001). Pol II is the central enzyme that transcribes protein-coding genes and synthesises messenger RNA (mRNA), which in turn serves as a template to direct protein synthesis. With a molecular weight of around 500,000, the Pol II core structure was the largest fully asymmetric protein complex solved by X-ray crystallography.

The atomic details obtained from the yeast Pol II structure (Cramer et al., 2000, 2001; Gnatt et al., 2001) and a related structure of a bacterial RNA polymerase determined by the laboratory of Seth Darst around the same time (Zhang et al., 1999) provided first insight into the mechanisms used by these molecular machines, and were the starting point for a detailed structural characterisation of gene transcription in the years to follow. These developments were possible due to technological advances in structural biology. While X-ray crystallography and biochemical probing were the structural methods of choice for a long time, cryo-electron microscopy (EM) recently developed to a stage where it allows one to study large assemblies at near-atomic resolution (Kuhlbrandt, 2014; Nogales and Scheres, 2015).

1 Transcription initiation

To initiate transcription, RNA polymerase must locate the promoter DNA sequence at the beginning of a gene, open the DNA duplex, and use the DNA template single strand to initiate RNA synthesis (Conaway and Conaway, 1993; Roeder, 1996; Grünberg and

Hahn, 2013; Sainsbury et al., 2015). Transcription initiation is a highly regulated process; it preferentially occurs at those genes that are expressed in a certain cell type. Pol II initiation requires five general transcription factors that position Pol II on promoter DNA, assist Pol II in finding the transcription start site, and facilitate DNA opening and initial RNA synthesis. In the classical model (Buratowski et al., 1989), transcription factor (TF) IID or its subunit TATA box-binding protein (TBP) bind promoter DNA and TFIIB, which can then recruit the Pol II-TFIIF complex. Then TFIIE and TFIIH bind to the assembly to form a pre-initiation complex (PIC) on closed, double-stranded promoter DNA. Next, promoter DNA is opened in an ATP-dependent manner, resulting in a DNA 'bubble' and the formation of an open complex (OC). RNA synthesis then leads to an initially transcribing complex (ITC). When the RNA grows to a critical length, Pol II escapes the promoter and forms a stable elongation complex (EC) and exchanges general initiation factors with elongation factors (Figure 1). Initiation requires the complete 12-subunit form of Pol II that includes the subcomplex Rpb4-Rpb7 (Armache et al., 2003; Bushnell and Kornberg, 2003) that lacked from the initial analysis of the Pol II structure (Cramer et al., 2001).

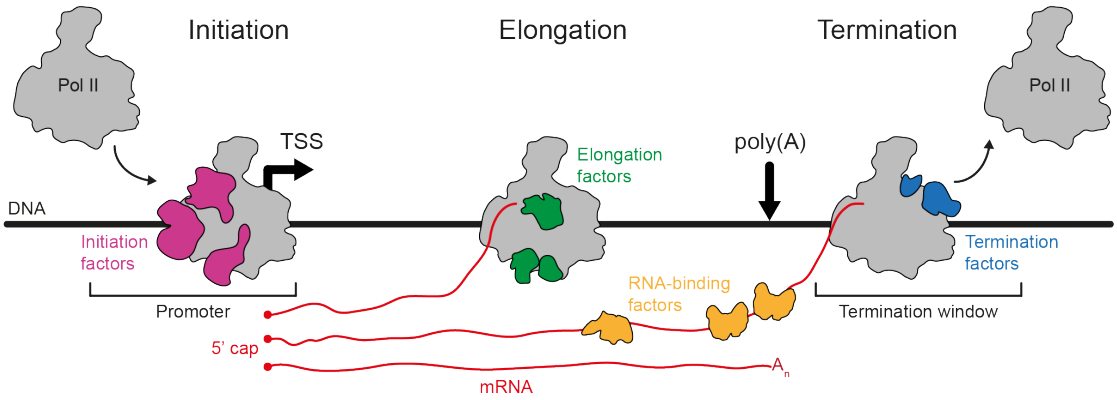

Figure 1: Schematic representation of the transcription cycle. Pol II binds the promoter of a gene close to the transcription start site (TSS) with the help of initiation factors. During elongation, the nascent mRNA chain is prolonged. The polyadenylation (poly(A)) site marks the end of the gene where the mRNA is cleaved. Further downstream, Pol II is displaced from the DNA template and freed for a new round of transcription. The mRNA is processed by addition of a 5′ cap (red dot) and a poly(A) tail (A_n) at the 3′ end.

1.1 Recognition of promoter DNA

Starting in the 1990s, several studies provided insights into how promoter DNA is recognised by general transcription factors (Sainsbury et al., 2015). As a result, we learned how such factors bind to DNA and to each other to form a promoter assembly that marks the beginning of a gene. Eukaryotic promoters contain various sequence elements that interact with components of the general transcription factors and are located around the transcription start site (Smale and Kadonaga, 2003; Kadonaga, 2012) (Figure 1). The first identified eukaryotic promoter element was the TATA box, an AT-rich sequence of eight base pairs (Basehoar et al., 2004), which is specifically recognised by TBP. Co-crystal structures revealed that TBP induces DNA bending by almost 90 degrees (Kim et al., 1993a,b; Kim and Burley, 1994; Nikolov et al., 1996; Juo et al., 1996; Patikoglou et al., 1999). Adjacent binding of TFIIA stabilises the TBP-DNA complex (Tan et al., 1996; Geiger et al., 1996; Bleichenbacher et al., 2003; Anandapadamanaban et al., 2013).

The pseudo-symmetric structure of TBP raised an important question, namely how TBP binding could confer directionality to transcription, i.e. that the polymerase moves in the right direction. Although this is still not fully understood, it appears that TBP has some preference for binding only in one orientation, and additionally the adjacent factor TFIIB determines directionality. TFIIB binds the TBP-DNA complex such that it specifically contacts DNA elements on either side of the TATA box (Littlefield et al., 1999; Tsai and Sigler, 2000; Lagrange et al., 1998; Deng and Roberts, 2005). Since TFIIB bridges to Pol II, it can load Pol II in a unidirectional manner.

Biochemical and structural studies were generally carried out on TATA box-containing promoters, although this motif is only present at ~20 % of eukaryotic promoters (Basehoar et al., 2004). However, TBP is bound on both TATA-containing and TATA-less promoters *in vivo* (Basehoar et al., 2004; Rhee and Pugh, 2012). TBP is a subunit of the multi-protein complex TFIID that contains 13-14 TBP-associated factors (TAFs) (Tora, 2002). TAFs contribute to promoter recognition and mediate specific interactions to different promoter elements around the transcription start site (Chalkley and Verrijzer, 1999; Theisen et al., 2010; Burke and Kadonaga, 1997; Lee et al., 2005). Low resolution EM studies revealed a horseshoe-like density that adopted two different conformations (Andel et al., 1999; Grob et al., 2006; Elmlund et al., 2009; Bieniossek et al., 2013; Cianfrocco et al., 2013). Recently, the cryo-EM structure of human TFIID bound to promoter DNA provided a more detailed view of TFIID-promoter interactions (Louder et al., 2016).

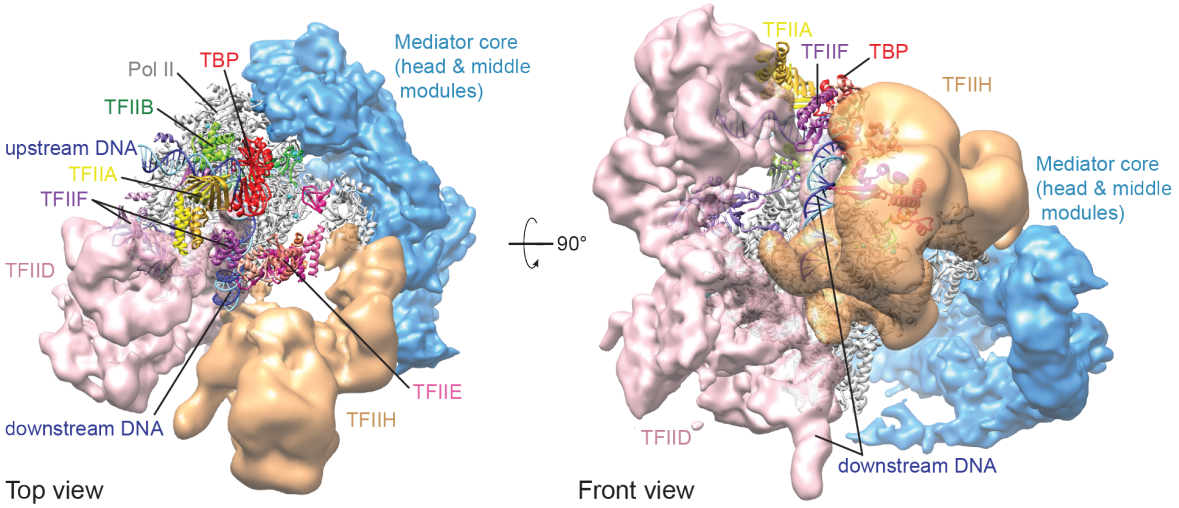

Figure 2: Composite topological model of the Pol II initiation complex. Cryo-EM densities of TFIID (EMD-3305) (Louder et al., 2016), TFIIH (EMD-3307) (Louder et al., 2016), and Mediator head and middle modules (EMD-2789) (Plaschka et al. 2015) are superposed on the structure of the core PIC (PDB 5FZ5) (Plaschka, Hantsche et al., 2016).

TFIID forms two major contacts to promoter DNA, first via TBP to the TATA element and second via TAF1 to downstream promoter elements (Figure 2). Thus TFIID may act as a molecular ruler to position TBP upstream of the transcription start site also at TATA-less promoters.

1.2 Opening of promoter DNA

Based on the progress in understanding promoter recognition, the next step was to investigate how the promoter assembly can recruit and position the polymerase at the beginning of a gene. Since TFIIB is the central bridge between promoter DNA and the polymerase, the location of TFIIB on the polymerase surface had to be determined. Early studies positioned the TFIIB N-terminal domain on Pol II (Chen and Hahn, 2003; Bushnell et al., 2004), but this did not enable for conclusive modelling of the promoter assembly on Pol II. Localisation of the C-terminal region of TFIIB using biochemical assays led to a topological model for the Pol II-promoter assembly complex and suggested that promoter DNA runs over the polymerase active centre cleft (Chen and Hahn, 2004; Miller and Hahn, 2006). Crystal structures of Pol II-TFIIB complexes confirmed and refined this model (Kostrewa et al., 2009; Liu et al., 2010a). These studies led to models of the closed and open complexes and revealed that around 30 nucleotides of DNA are

required to connect the TATA element with the active centre of Pol II (Kostrewa et al., 2009), explaining the minimal distance between the TATA box and the transcription start site (Giardina and Lis, 1993). Further functions of TFIIB were revealed when the Pol II-TFIIB complex was crystallised in the presence of a DNA scaffold with a short RNA (Sainsbury et al., 2013). Binding of TFIIB allosterically rearranged the catalytic site of Pol II, stimulating RNA synthesis.

Protein cross-linking and biochemical probing were used to locate the general transcription factors TFIIF, TFIIE and TFIIH on the Pol II surface (Kim et al., 2000; Chen et al., 2007, 2010b; Eichner et al., 2010; Grünberg and Hahn, 2013; Fishburn and Hahn, 2012; Mühlbacher et al., 2014). Visualisation of the topology of a complete pre-initiation complex was achieved by cryo-EM in 2013 (He et al., 2013). The architecture of the human complex was consistent with the suggested models for the yeast complex derived from biochemical studies (Chen et al., 2007; Grünberg and Hahn, 2013; Mühlbacher et al., 2014). Subsequent cryo-EM structures of human and yeast Pol II complexes underlined the high conservation between both systems (Plaschka et al., 2015; Murakami et al., 2015; Bernecky et al., 2016). A limited resolution of these initiation complex densities allowed docking of known structures of transcription factors into the densities (Tan et al., 1996; Sainsbury et al., 2013; Kim et al., 1993a; Meinhart et al., 2003; Okuda et al., 2000, 2004; Gaiser et al., 2000; Groft et al., 1998; Kamada et al., 2001; Kilpatrick et al., 2012; Fan et al., 2006; Miller and Hahn, 2006). However, for mechanistic insights into Pol II initiation, higher resolution is required.

Recent biochemical studies gave insight how TFIIH achieves promoter DNA opening and loading into the Pol II active centre cleft (Kim et al., 2000; Grünberg and Hahn, 2013; Fishburn et al., 2015). Structural information supports the functional findings (Kostrewa et al., 2009; He et al., 2013; Murakami et al., 2015; Louder et al., 2016; He et al., 2016). TFIIH binds DNA downstream and translocates along DNA away from Pol II (Figure 2). Since both the promoter and TFIIH are held in the pre-initiation complex, this translocase action creates torsional stress on DNA, facilitating its melting and pushing downstream DNA into the active centre cleft. The emerging DNA template strand is threaded into the active centre, where it is bound by an element of TFIIB, the B-reader, which functions in recognising the transcription start site (Kostrewa et al., 2009).

1.3 Regulation of transcription initiation

Promoter elements are located in the vicinity of the transcription start site of a gene (Figure 1). There are however multiple additional DNA sequence elements outside of the promoter that can bind gene-specific proteins that regulate transcription. These regulatory elements are located near the promoter, but in multi-cellular organisms additional regulatory elements are found very distant from the promoter, called enhancers. Enhancers bind DNA sequence-specific transcription factors in response to environmental or cellular signals, for example during organism development and cellular differentiation. In order to convey a signal from the DNA-bound transcription factor to the polymerase, a co-activator is required that forms a bridge between regulatory transcription factors and the general transcription machinery.

Mediator is a central co-activator needed for transcription of most protein-coding genes (Kornberg, 2005; Malik and Roeder, 2010; Conaway and Conaway, 2011; Allen and Taatjes, 2015). Though discovered over two decades ago (Flanagan et al., 1991; Kim et al., 1994), the molecular mechanisms Mediator uses are not well understood because structural studies were impeded by the size, flexibility and modularity of the large complex. In yeast, Mediator has a molecular weight of 1.4 MDa and is composed of 25 proteins, and metazoan Mediator even contains several additional subunits. Based on early studies, the Mediator structure was divided into four modules, the head, middle, tail and kinase modules (Asturias et al., 1999). A large number of EM studies of Mediator and several of its complexes were reported (Davis et al., 2002; Elmlund et al., 2006; Cai et al., 2009, 2010; Bernecky et al., 2011; Cai et al., 2012; Tsai et al., 2013, 2014; Wang et al., 2014), but these remained at low resolution, rendering the relative location of complexes and mechanistic proposals uncertain. The head module interacts with the general Pol II initiation machinery, the tail binds gene-specific transcription factors, and the kinase module can modify the flexible C-terminal domain (CTD) of the large Pol II subunit. The CTD is a unique Pol II element serving as a platform recruiting proteins for co-transcriptional processes, and changes its phosphorylation pattern during the transcription cycle (Meinhart et al., 2005; Heidemann et al., 2013; Jeronimo et al., 2013).

Recent advances in the structural biology of Mediator elucidated its interaction with the Pol II initiation machinery (Larivière et al., 2012b; Plaschka et al., 2016b). Co-expression enabled preparation of the 7-subunit Mediator head module (Takagi et al., 2006; Larivière et al., 2012a), the 7-subunit middle module (Koschubs et al., 2010) and later of the core Mediator comprising the head and middle modules and a subunit that

forms an architectural backbone, Med14 (Cevher et al., 2014; Plaschka et al., 2015). The crystal structure of the head module could be solved (Imasaki et al., 2011; Larivière et al., 2012a; Robinson et al., 2012), based on structures of several subcomplexes (Larivière et al., 2012b). The structure of a core initiation complex bound by the core Mediator was determined by cryo-EM and showed that the Mediator head module is located on one side of Pol II where it can stabilise TFIIB and the polymerase stalk, which is involved in initiation (Plaschka et al., 2015). Modelling shows that Mediator and TFIID bind to opposite sides of Pol II with TFIIH likely in between (Louder et al., 2016) (Figure 2). Such embedding of Mediator in the initiation assembly may explain its functions in stabilising the PIC. The tail module of Mediator communicates with a plethora of transcription factors (Borggrefe and Yue, 2011). Activator-binding triggers specific conformational changes in Mediator (Taatjes et al., 2002; Ebmeier and Taatjes, 2010; Meyer et al., 2010), but how these changes are transmitted and how they influence the interaction with Pol II and the initiation complex is unknown.

2 RNA chain elongation and termination

During RNA chain elongation, the polymerase repeats the so-called nucleotide addition cycle. In each cycle, Pol II presents the template DNA base, selects and binds a complementary RNA nucleoside triphosphate (NTP) substrate, catalyses phosphodiester bond formation to add a nucleotide to the growing RNA chain, and translocates to the next template position, thereby freeing the NTP-binding site. A wealth of crystal structures of Pol II with various DNA-RNA scaffolds elucidated the molecular mechanisms of the nucleotide addition cycle and many other aspects of elongation (Cheung and Cramer, 2012; Martinez-Rucobo and Cramer, 2013). Studies in the bacterial system contributed substantially to our understanding of elongation, because all polymerases share a conserved catalytic centre and transcription mechanism (Cramer et al., 2000, 2001; Zhang et al., 1999; Vassylyev et al., 2002; Hirata et al., 2008; Korkhin et al., 2009; Svetlov and Nudler, 2013).

2.1 Nucleotide addition cycle

In the first step of the nucleotide addition cycle, the correct NTP is selected by the Pol II active centre in a sampling process (Westover et al., 2004; Kettenberger et al., 2004; Wang et al., 2006; Sydow et al., 2009; Cheung et al., 2011). The NTP binds transiently to

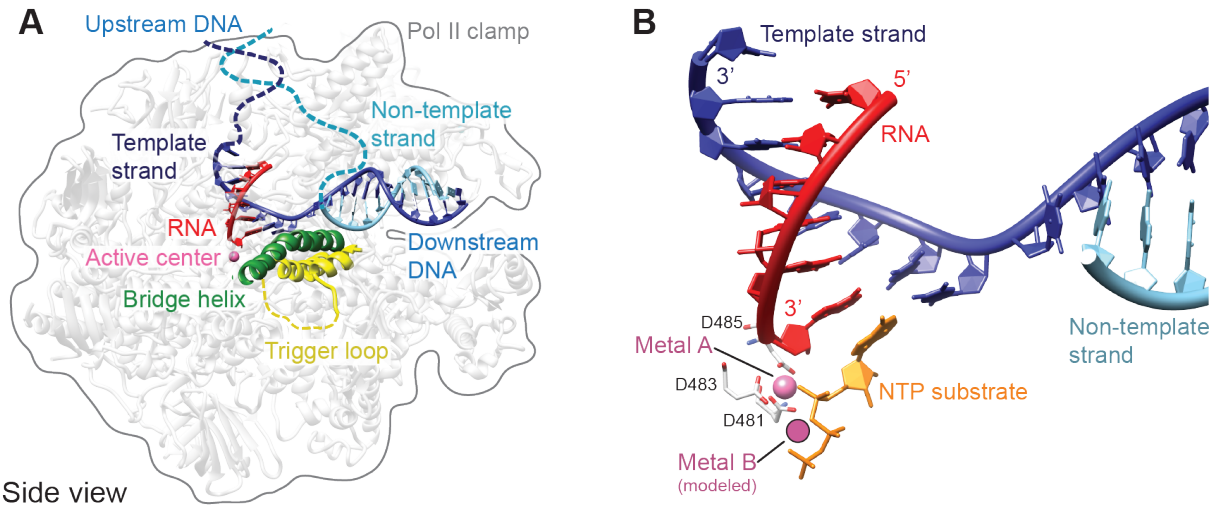

Figure 3: Polymerase active centre and nucleic acid interactions. A, Central slice through the Pol II elongation complex indicating the path of DNA (template strand, blue; non-template strand, cyan) and newly synthesised RNA (red). Important elements of the active centre are highlighted. (PDB 4A3F) (Cheung et al., 2011) **B**, Detailed view of nucleic acids and the active centre. The NTP substrate is a non-hydrolysable nucleoside triphosphate (AMPCPP) (Cheung et al., 2011). Metal A is permanently bound and coordinated by three conserved aspartate residues (D481, D483 and D485). For catalysis, a second magnesium ion (metal B) is recruited, which was modelled according to the bacterial EC (PDB 2O5J) (Vassylyev et al., 2007b).

an open and catalytically inactive pre-insertion conformation. Upon Watson-Crick base-pairing of the cognate NTP with the template base, the active site closes by folding of the so-called trigger loop and the NTP is moved to the insertion site, where it is positioned for catalysis (Figure 3b). This two-step mechanism of NTP selection is also used by the bacterial RNA polymerase (Vassylyev et al., 2007a,b). The mechanism for catalytic nucleotide addition was proposed based on structural studies of DNA polymerases (Steitz et al., 1994) before the structure of an RNA polymerase was available. The mechanism involves two metal ions, one (metal A) persistently bound in the active centre, and the other (metal B) being mobile (Cramer et al., 2001). Both metal ions are coordinated by conserved negatively charged amino acid residues. Metal A binds the RNA 3'-end and metal B positions the NTP moiety (Figure 3b). Catalysis occurs through a nucleophilic substitution (SN_2) mechanism, in which the RNA 3'-OH group acts as the nucleophile attacking the NTP α-phosphate. Release of pyrophosphate may induce opening of the active centre.

2.2 Translocation mechanism

Before the next NTP can be bound, Pol II has to translocate by one step along the DNA to free the NTP-binding site that is occupied by the newly added RNA 3'-nucleotide. Structural studies of Pol II bound by the mushroom toxin α-amanitin helped to reveal the translocation mechanism (Bushnell et al., 2002; Kaplan et al., 2008; Brueckner and Cramer, 2008). Amanitin binds beneath two Pol II elements in the active centre, the trigger loop and bridge helix, which form a Brownian ratchet that rapidly switches between a pre- and a post-translocation state (Figure 3a). NTP binding to the active centre acts like a pawl of the ratchet. Only after addition of the nucleotide to the growing RNA, the oscillations resume, moving the nascent hybrid base-pair out of and the next template base into the active site. Directionality is achieved as NTPs can only bind in the post-translocation state of the ratchet. The toxin α-amanitin traps the ratchet in an intermediary state, stabilising a translocation intermediate (Brueckner and Cramer, 2008). From these studies the basic DNA template-directed RNA synthesis mechanism was derived.

2.3 Transcription elongation

The elongation complex is very stable mainly because of the stability of the DNA-RNA hybrid in the active centre (Kireeva et al., 2008). Additionally, elongation factors repress pausing and enhance Pol II processivity, i.e. the ability of Pol II to remain associated with the template DNA until transcription is terminated. Spt5 (bacterial NusG) is a highly conserved and ubiquitous elongation factor (Werner, 2012; Hartzog and Fu, 2013). Structural studies of the factor alone (Guo et al., 2008; Hirtreiter et al., 2010; Wenzel et al., 2010) and in complex with polymerase (Klein et al., 2011; Martinez-Rucobo et al., 2011; Bernecky et al., 2016) illustrated that a conserved domain of Spt5 spans the active centre cleft, locking nucleic acids in the cleft and thus preventing elongation complex dissociation. Elongation factors function also as a recruitment platform for co-transcriptional processes, like 5' RNA capping, splicing or chromatin remodelling (Werner, 2012; Hartzog and Fu, 2013). Many additional multiprotein elongation complexes are involved in these events. The molecular basis and coordination of these processes represent important areas for future studies.

2.4 Transcription termination

At the end of the transcription cycle Pol II must release the RNA transcript and dissociate from the DNA template (Figure 1). Correct transcription termination is important to prevent transcription interference at downstream genes and to enable recycling of Pol II (Kuehner et al., 2011; Mischo and Proudfoot, 2013; Proudfoot, 2016). When the polymerase runs over the polyadenylation (poly(A)) signal, which marks the end of a gene, the nascent RNA chain is cleaved and a poly(A) tail is added to the 3'-end of the RNA. Pol II however terminates further downstream at multiple positions. Recent work showed that human genes contain on average 4 termination sites within a window of several thousand base pairs downstream of the poly(A) site (Schwalb et al., 2016). Two different models for Pol II termination were suggested. In the allosteric model, binding of RNA 3'-processing factors induces structural rearrangements in the Pol II elongation complex that elicit termination (Logan et al., 1987). In the torpedo model, a nuclease degrades the newly synthesised RNA after cleavage, catches up with the elongation complex and thereby dislodges Pol II from DNA (Connelly and Manley, 1988; Proudfoot, 1989). The protein Rat1 in yeast (Xrn2 in humans) is the torpedo nuclease (West et al., 2004; Kim et al., 2004). There is also evidence for a unified mechanism that includes aspects from both models (Luo et al., 2006; Schreieck et al., 2014). The structural mechanism of transcription termination still remains enigmatic. There is evidence that the Pol II elongation complex slows down and changes its factor composition and its phosphorylation state when Pol II runs over the poly(A) site (Ahn et al., 2004; Mayer et al., 2012; Schreieck et al., 2014). However, the nature of the pre-termination complex is discussed (Zhang et al., 2015; Fong et al., 2015). The transient nature of the pre-termination complex has thus far prevented its structural analysis, but it is likely that termination involves re-positioning of Spt5 and opening of the Pol II clamp, to facilitate release of the nucleic acids.

3 Polymerase conservation and gene class specificity

The eukaryotic genome is transcribed not only by Pol II, which produces mRNA, but also by Pol I and Pol III, which synthesise mainly ribosomal RNA (rRNA) and transfer RNA (tRNA), respectively. Whereas Pol II makes thousands of different mRNAs, most RNA in the cell is made by Pol I and Pol III, which synthesise around 80 % of the total RNA transcripts in yeast (Paule and White, 2000). All three polymerases share a 10-subunit

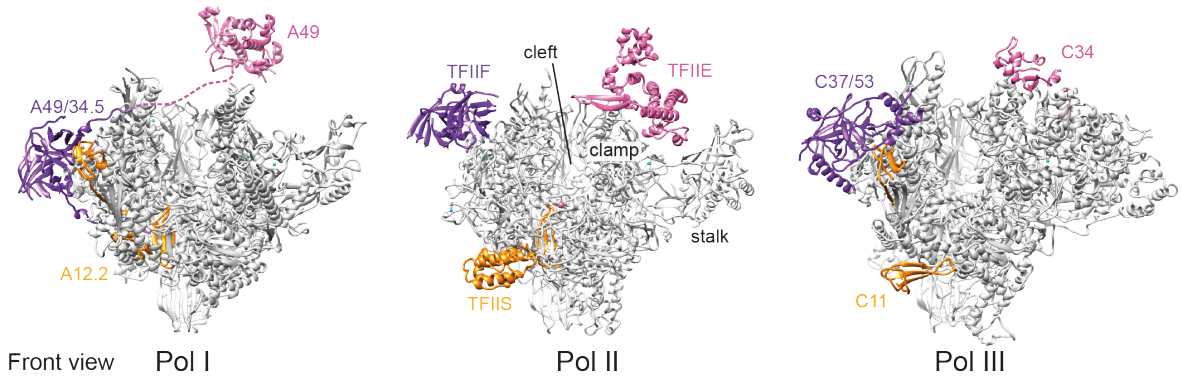

Figure 4: Conservation of eukaryotic polymerases. Structures of Pol I (PDB 4C2M) (Engel et al., 2013), Pol II (PDB 1WCM) (Armache et al., 2005), and Pol III (PDB 5FJ9) (Hoffmann et al., 2015) are shown in a front view (Cramer et al., 2001). Pol II is shown together with a part of TFIIE (PDB 5FYW) (Plaschka, Hantsche et al., 2016), the dimerization domain of TFIIF (PDB 5FYW) (Plaschka, Hantsche et al., 2016) and TFIIS (PDB 3PO3) (Cheung et al., 2011). Subunits of Pol I and Pol III that are structurally and functionally similar to TFIIE, TFIIF and TFIIS, are colored according to Figure 2. The C-terminal part of A49 (PDB 3NFH) (Geiger et al., 2010) and the subunit C34 show some homology with TFIIE (pink), the subcomplex A49/34.5 and C37/53 are homologous to the dimerisation domain of TFIIF (purple), and the subunits A12.2 and C11 are homologs of TFIIS (orange). The Pol II stalk, cleft and clamp are labelled.

core and a conserved two-subunit stalk (Vannini and Cramer, 2012) (Figure 4). Recent high-resolution structures of yeast Pol I (Engel et al., 2013; Fernández-Tornero et al., 2013) and Pol III (Hoffmann et al., 2015) revealed that the polymerases also share the same active site and apparently apply the same catalytic mechanism for the nucleotide addition cycle. However, the polymerases differ in their molecular size and number of subunits, with Pol II being the smallest and Pol III being the largest of the three polymerases. Compared to Pol II, Pol I and Pol III contain additional subunits on their periphery that resemble domains of the general transcription factors TFIIE and TFIIF, and the cleavage-stimulatory factor TFIIS (Kuhn et al., 2007; Geiger et al., 2010; Vannini et al., 2010; Fernández-Tornero et al., 2010; Ruan et al., 2011; Lefèvre et al., 2011; Wu et al., 2011, 2012) (Figure 4).

The differences on the polymerase surfaces reflect gene class-specific requirements for transcribing and regulating different RNA transcripts. Like Pol II, Pol I and Pol III rely on a specific set of initiation factors that control recruitment and positioning of the polymerases to their distinct promoter types (Vannini, 2013). Universally required are apparently TBP and a TFIIB-like factor (Knutson and Hahn, 2011; Naidu et al., 2011; Lopez-De-Leon et al., 1992; Colbert and Hahn, 1992). For Pol II, initiation requires

the ATP-consuming factor TFIIH, whereas the two other polymerases can open DNA without ATP consumption. Thus, Pol I and Pol III contain subdomains as integral parts that are important for transcription initiation and processivity. In the Pol II system these domains form separate factors, probably due to an extended need for regulatory control. The structural characterisation of Pol I and Pol III initiation complexes will be an important next step in understanding promoter specificity of the three polymerases, and eventually their distinct modes of regulation.

4 Aims and scope of this work

Transcription initiation is a key event in the regulation of gene expression. Structural characterisation of the interplay of Pol II with the general transcription factors and promoter DNA is a crucial step in understanding the molecular mechanism behind promoter DNA recognition and opening. Three-dimensional structural information of individual proteins and subcomplexes was acquired for many years. Due to the limited resolution of available complex structures, however, many questions on the molecular basis of initiation remain open. Recent advances in the development of cryo-EM opened the path to study large macromolecular complexes at high resolution.

We determined the architecture of a closed transcription initiation complex consisting of Pol II, TFIIA, TFIIB, TBP, TFIIE, TFIIF and closed promoter DNA at 8.8 Å. We observed that most complexes had spontaneously opened the promoter DNA in the absence of TFIIH. Model building was guided by the recently determined structure of the respective open complex at 3.6 Å resolution. Comparison of the two structures revealed how the general transcription factors interact with Pol II and DNA and which protein domains are involved in the transition from closed to open complex formation. Complementary functional assays were performed to investigate the role of these domains for transcription *in vivo* and *in vitro*. In a separate crystallographic study we showed that the N-terminus of TFIIF subunit Tfg1 binds to the Pol II external 1 region.

Part II

Materials and Methods

5 Materials

5.1 Bacterial strains

Table 1: *E. coli* strains used in this study

Strain	Genotype	Source
XL1-Blue	*rec1A endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* $[F'\ proAB\ lacl^q Z\Delta M15\ Tn10\ (Tet^r)]$	Stratagene
BL21-Codon Plus(DE3)-RIL	B; F^- *ompT hsdS* $(r_B^-\ m_B^-)$ dcm^+ Tet^r *gal*_(DE3); *endA* Hte $[argU, ileY, LeuW, Cam^r]$	Stratagene
Rosetta B834(DE3)	F^- *ompT hsdS* $(r_B^-\ m_B^-)$ dcm^+ met^B	Novagen

5.2 Yeast strains

Table 2: *S. cerevisiae* strains used or generated in this study

Strain	Genotype	Source
BJ5464 Rpb3 His-Bio	BJ5464 Rpb3 His-Bio tag introduced at 5′ end of Rpb3 gene, use of URA3 selection marker	(Kireeva et al., 2000)
BJ5464 Rpb3 His-Bio Δ*rpb9*	BJ5464 Rpb3 His-Bio *rpb9*::natNT2	E. Lehmann
SHY734 Tfa1	MATα *deltaade2::hisG his3delta200 leu2delta0 lys2delta0 met15delta0 trp1delta63 ura3delta0 tfa1*delta::kanMx pSH633 (Tfa1 *ars cen ura3*)	S. Hahn
SHY734 3xFlag-Tfa1	MATα *deltaade2::hisG his3delta200 leu2delta0 lys2delta0 met15delta0 trp1delta63 ura3delta0 tfa1*delta::kanMx pSH810 (3xFlag-Tfa1 *ars cen leu2*)	this study

5.3 Plasmids and oligonucleotides

Table 3: Plasmids used in this study

Vector	Insert	Type	Source
pAHS3C	N-terminal 10x histidine-8x arginine-SUMO-tag, cleavable with 3C protease	pETDuet	S. Bilakovic
TFIIF	*Sc* Tfg1, Tfg2	pAHS3C	this study
TFIIFΔins	*Sc* Tfg1(Δ169-299), Tfg2(Δ138-210)	pAHS3C	this study
TFIIFΔins-TLD	*Sc* Tfg1(Δ169-299)-(Bye1 TLD 225-370), Tfg2	pAHS3C	this study
S6-TFIIFΔins-TLD	*Sc* S6-Tfg1(Δ169-299)-(Bye1 TLD 225-370), Tfg2 (Δ138-210)	pAHS3C	this study
TFIIFΔins-Rpb9	*Sc* Tfg1(Δ169-299, Δ511-735)-Rpb9, Tfg2 (Δ138-210)	pAHS3C	this study
TFIIE	*Sc* Tfa1, Tfa2	pET21	L. Larivière
TFIIE M1	*Sc* Tfa1 (N50E/K51E/T52E), Tfa2	pET21	C. Plaschka
TFIIE M2	*Sc* Tfa1 (N50A/K51A/T52A), Tfa2	pET21	C. Plaschka
TFIIE M3	*Sc* Tfa1 (P56A/A59E/R62E), Tfa2	pET21	C. Plaschka
TFIIE M4	*Sc* Tfa1 (ΔE-wing), Tfa2	pET21	C. Plaschka
TFIIE M5	*Sc* Tfa1 (Ala E-wing), Tfa2	pET21	C. Plaschka
TFIIE M6	*Sc* Tfa1 (L134E/V137E/L140E), Tfa2	pET21	C. Plaschka
TFIIE M7	*Sc* Tfa1 (L134A/V137EA/L140A), Tfa2	pET21	C. Plaschka
HIS4	native *HIS4* promoter (-428 to +24 respective to the A in the start codon)	pBluescript II KS+	S. Hahn

Table 4: Oligonucleotides used for EM, crystallizsation and assays

Type	Sequence (5′-3′)	Application
template DNA	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	cryo-EM
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 A	
RT primer	Cy5-TTCACCAGTGAGACGGGCAACAGCCAAGCT	transcription assay

5.4 Peptides

Table 5: Peptides used for crystallisation

Type	Sequence	Source
Tfg1 N-term (19 - 41 aa)	SPFIKRDRMRRNFLRMRMGQNGS	Peptide Speciality Laboratories
SeMet Tfg1 N-term	SPFIKRDR(**SeM**)RRNFLRMRMGQNGS	

5.5 Antibodies

Table 6: Antibodies used in this study

Antibody	Dilution	Host	Source
anti-Rpb3	1:1,000	mouse	Neoclone
anti-TFIIB	1:4,000	rabbit	Abcam
anti-Histone H3	1:4,000	HPR-coupled	Abcam
anti-Flag tag	1:5,000	HPR-coupled	Miltenyi Biotec
anti-mouse IgG-HRP	1:3,000	goat	GE Healthcare
anti-rabbit IgG-HRP	1:3,000	donkey	GE Healthcare

5.6 Growth media and additives

Table 7: Growth media for *E. coli* and *S. cerevisiae*

Media	Application	Description
LB	*E. coli* culture	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.5 % (w/v) NaCl
SeMet	*E. coli* culture	21.6 g/l SeMet minimal medium base, 5.1 g/l nutrient mix, 100 mg/ml SeMet (all components from AthenES)
YPD	Yeast culture	2 % (w/v) peptone, 2 % (w/v) glucose, 1 % (w/v) yeast extract
YPDS	Yeast nuclear extract preparation	2 % (w/v) peptone, 2.2 % (w/v) glucose, 1 % (w/v) yeast extract, 18.25 % (w/v) sorbitol
SC	Yeast culture	0.69 % (w/v) nitrogen base, 0.6 % (w/v) CSM amino acid drop-out mix, 2 % (w/v) glucose

Table 8: Media additives for *E. coli* and *S. cerevisiae*

Additive	**Description**	**Stock solution**	**Applied concentration**
IPTG	*E. coli* induction	1 M in H_2O	0.25 - 0.5 mM
50x 5052	*E. coli* autoinduction medium	25 % (w/v) glycerol, 2.5 % (w/v) glucose, 10 % (w/v) α-lactose	0.5 % (w/v) glycerol, 0.05 % (w/v) glucose, 0.2 % (w/v) α-lactose
20x NPS	*E. coli* autoinduction medium	0.5 M $(NH_4)_2 SO_4$, 1 M KH_2PO_4, 1 M Na_2HPO_4	25 mM $(NH_4)_2 SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4
Ampicillin	Antibiotic	100 mg/ml in H_2O	100 µg/ml
Chloramphenicol	Antibiotic	50 mg/ml in EtOH	50 µg/ml
Kanamycin	Antibiotic	30 mg/ml in H_2O	530 µg/ml
clonNAT	Antibiotic	200 mg/ml in H_2O	100 µg/ml
5-FOA	Yeast plasmid shuffling	100 mg/ml in DMSO	1 mg/ml

5.7 Buffers and solutions

Table 9: General buffers, dyes and solutions

Name	Description/Source	Application
Electrophoresis running buffer	NuPAGE MES/MOPS SDS running buffer, Invitrogen	SDS-PAGE
5x SDS-PAGE sample buffer	250 mM Tris pH 6.8 at 20 °C, 50 % (v/v) glycerol, 0.03 % (w/v) bromophenol blue, 7.5 % (w/v) SDS, 100 mM DTT	SDS-PAGE
Native PAGE running buffer	Blue native PAGE, Invitrogen	Native PAGE
Native PAGE Dark Blue Cathode Buffer	Blue native PAGE, Invitrogen	Native PAGE
Gel staining solution	InstantBlue, Expedion	Coomassie staining
Silver nitrate solution	0.1 % (w/v) $AgNO_3$, 0.01 % (v/v) formaldehyde	Silver staining
Developing solution	3 % (w/v) Na_2CO_3, 0.05 % (v/v) formaldehyde	Silver staining
Tris-glycin running buffer	Trans-Blot Turbo Blotting System, Bio-Rad	Western blotting
PBS-T buffer	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4 pH 7.4, 0.1 % Tween-20	Western blotting
100x PI	0.028 mg/ml leupeptin, 0.137 mg/ml pepstatin A, 0.017 mg/ml PMSF, 0.33 mg/ml benzamidine, in 100 % EtOH p.a.	Protease inhibitor mix
10x TAE	50 mM EDTA pH 8.0, 2.5 M Tris-acetate	Agarose gels
TE	1 mM Tris pH 8.0 at 20 °C, 0.1 mM EDTA	dissolving DNA
10x TBE	900 mM Tris, 900 mM boric acid, 20 mM EDTA pH 8.0	transcription assay
2x urea loading buffer	20 % (v/v) 10× TBE, 8 M RNase-free urea	Urea PAGE
2x urea loading dye	20 % (v/v) 10× TBE, 8 M RNase-free urea; 0.03 % (w/v) bromophenol blue; 0.03 % (w/v) xylene cyanol	Urea PAGE

Table 10: Buffers used for preparation of competent *E. coli* cells

Name	Description
TFB-I buffer	30 mM KAc, 50 mM Mn_2Cl, 100 mM RbCl, 10 mM $CaCl_2$, 15 % (v/v) glycerol
TFB-II buffer	10 mM MOPS pH 7.0, 75 mM $CaCl_2$, 10 mM RbCl, 15 % (v/v) glycerol

Table 11: Buffers used for preparation of competent yeast cells

Name	Description
LiAc-TE	1.1 M LiAc, 10 mM Tris pH 7.5, 1 mM EDTA
LiAc-TE-PEG	1.1 M LiAc, 10 mM Tris pH 7.5, 1 mM EDTA, 40 % (w/v) PEG 4000

Table 12: Buffers used for protein extraction from yeast cells

Name	Description
Treatment solution	7.5 M NaOH, 4 mM β-ME
Sample buffer	5 % (w/v) SDS, 0.2 M Tris-HCl pH 6.8, 1 mM EDTA, 215 mM β-ME, 8 M urea, 0.01 % (w/v) bromophenol blue

Table 13: Buffers used for Pol II purification

Name	Description
3x Freezing buffer	150 mM Tris pH 7.9 at 4 °C, 3 mM EDTA, 30 µM $ZnCl_2$, 30 % (v/v) glycerol, 3 % (v/v) DMSO, 30 mM DTT, 3x PI
HSB150	50 mM Tris pH 7.9 at 4 °C, 150 mM KCl, 1 mM EDTA, 10 µM $ZnCl_2$, 10 % (v/v) glycerol, 10 mM DTT, 1x PI
HSB1000/7	50 mM Tris pH 7.9 at 4 °C, 1000 mM KCl, 7 mM imidazole, 1 mM EDTA, 10 µM $ZnCl_2$, 10 % (v/v) glycerol, 2.5 mM DTT, 1x PI
Ni buffer	20 mM Tris pH 7.9 at 4 °C, 150 mM KCl, 7/50/100 mM imidazole
MonoQ 150	20 mM Tris-acetate pH 7.9 at 4 °C, 150 mM KAc, 10 % (v/v) glycerol, 0.5 mM EDTA pH 7.9, 10 µM $ZnCl_2$, 10 mM DTT
MonoQ 2000	20 mM Tris-acetate pH 7.9 at 4 °C, 2000 mM KAc, 10 % (v/v) glycerol, 0.5 mM EDTA pH 7.9, 10 µM $ZnCl_2$, 10 mM DTT
Pol II buffer	10 mM HEPES pH 7.0 at 20 °C, 200 mM KCl, 5 % glycerol, 2 mM DTT

Rpb4-Rpb7 freezing buffer	50 mM Tris pH 7.0 at 4 °C, 150 mM NaCl, 10 % glycerol, 10 mM DTT, 1x PI
Ni buffer	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 0/10/20/50/200 mM imidazole, 10 mM DTT, 1x PI
Salt buffer	50 mM Tris pH 7.5 at 4 °C, 2 M NaCl, 10 mM DTT, 1x PI
MonoQ 0 buffer	20 mM Tris pH 7.5 at 4 °C, 1 mM EDTA, 10 mM DTT
MonoQ 2000 buffer	20 mM Tris pH 7.5 at 4 °C, 2 M NaCl, 1 mM EDTA, 10 mM DTT
Complex buffer	25 mM HEPES pH 7.5, 150 mM KAc, 5 % glycerol, 1 mM $MgCl_2$, 5 mM DTT

Table 14: Buffers used for TFIIE purification

Name	Description
Lysis buffer	50 mM Tris-HCl pH 8.0 at 4 °C, 300 mM NaCl, 0.02 % Tween-20, 5 mM DTT, 1x PI
HisTrap buffer	50 mM Tris-HCl pH 8.0 at 4 °C, 300 mM NaCl, 10 mM imidazole, 5 mM DTT
HeparinTrap buffer	50 mM Tris-HCl pH 8.0 at 4 °C, 100 mM NaCl, 2 % glycerol, 5 mM DTT
Gel filtration buffer	5 mM HEPES pH 7.25 at 4 °C, 40 mM $(NH_4)_2 SO_4$, 10 µM $ZnCl_2$, 10 mM DTT

Table 15: Buffers used for TFIIF purification

Name	Description
Lysis buffer	50 mM HEPES pH 7.0 at 4 °C, 350 mM KCl, 50 mM imidazole, 10 % glycerol, 2 mM DTT, 1x PI
HisTrap buffer	50 mM HEPES pH 7.0 at 4 °C, 250 mM KCl, 800 mM imidazole, 10 % glycerol, 2 mM DTT
HiTrap SP buffer	50 mM HEPES pH 7.0 at 4 °C, 150 mM KCl, 10 % glycerol, 2 mM DTT
Gel filtration buffer	10 mM MES pH 6.2 at 4 °C, 150 mM KCl, 10 % glycerol, 2 mM DTT

Table 16: Buffers used for nuclear extract preparation

Name	Description
Resuspension buffer	50 mM Tris pH 7.5 at 4 °C, 20 mM EDTA, 30 mM DTT
Lysis buffer	18 % (w/v) polysucrose 400, 10 mM Tris pH 7.5 at 4 °C, 20 mM KAc, 5 mM MgAc, 1 mM EDTA, 0.5 mM spermidine, 0.15 mM spermine, 3 mM DTT, 1x PI
Centrifugation buffer	100 mM Tris pH 7.9 at 4 °C, 50 mM KAc, 10 mM $MgSO_4$, 20 % (v/v) glycerol, 2 mM EDTA
Dialysis buffer	20 mM HEPES pH 7.6 at 4 °C, 10 mM $MgSO_4$, 1 mM EGTA, 20 % (v/v) glycerol

Table 17: Buffers used for *in vitro* transcription assay

Name	Description
5x Acetate transcription buffer	100 mM HEPES pH 7.6, 500 mM KAc, 5 mM EDTA, 25 mM MgAc
5x Primer annealing buffer	25 mM Tris pH 8.3 at 20 °C, 375 mM KCl, 5 mM EDTA pH 8
5x Synthesis buffer	250 mM Tris pH 8.3 at 20 °C, 375 mM KCl, 22.5 mM $MgCl_2$, 75 mM DTT

6 Methods

6.1 Molecular cloning

6.1.1 *E. coli*

Construct design

Primers for InFusion cloning contained an overhang of 15 nucleotides at the 5′ end, followed by a stretch of nucleotides complementary to the gene of interest yielding an annealing temperature of 65 °C to 70 °C. For insertions and deletions, the InFusion strategy was used except that the plasmid was amplified in two parts with overlapping regions in the deletion/insertion site and the resistance cassette.

Polymerase Chain Reaction (PCR)

PCR reactions were carried out in 50 µl volumes containing Phusion High Fidelity PCR Master Mix (New England Biolabs), 0.5 µM of each primer, 3 % (v/v) DMSO and 1 -5 ng plasmid DNA or 100 ng genomic DNA. Annealing temperatures of the primer pair was calculated by the New England Biolabs web-page (tmcalculator.neb.com) and elongation times according to the length of the product. Thermocycling programs comprised 30 cycles. PCR products were visualised by agarose gel electrophoresis and purified using the QIAquick gel extraction kit (Qiagen).

Electrophoretic separation of DNA

DNA separation was carried out by agarose gel electrophoresis using a 0.5 - 1.5 % agarose gel containing 1 µg/ml Sybr Safe (Invitrogen). DNA samples were mixed with 6x loading dye (Thermo Scientific) and loaded on the gel together with a size marker (GeneRuler DNA Ladder Mix, Thermo Scientific). Gel was run in 1x TAE buffer for 30 min at 200 V. DNA was visualised and documented using a gel imaging system (Intas).

Enzymatic restriction cleavage

Plasmids were linearised using restriction endonucleases (New England Biolabs) according to the manufacturer's instructions. 2 µg plasmid was incubated with the restriction endonucleases in the adequate reaction buffer at 37 °C for 3 h. Digested vectors were purified using the QIAquick PCR purification kit (Qiagen).

InFusion cloning

InFusion cloning (Clontech) is based on homologous recombination. A 10 µl reaction mix contained 50 -100 ng linearised vector, fourfold molar excess of insert and 1x InFusion reaction mixture and was incubated for 15 min at 37 °C, followed by 15 min at 50 °C. 5 - 10 µl of the reaction were transformed into *E. coli* cells (see Table 1).

Preparation of chemically competent *E.coli* cells

400 ml LB medium supplemented with tetracycline (see Table 8) was inoculated 1:100 with an overnight pre-culture and grown to an optical density at 600 nm (OD_{600}) of 0.5. Cells were harvested by centrifugation for 10 min at 4,000 rpm and 4 °C. Cell pellet was resuspended in 10 ml TFB-I buffer (see Table 10) and again centrifuged as before. Pellet was resuspended in 8 ml TFB-II buffer (see Table 10) and aliquoted into pre-cooled Eppendorf tubes. Aliquotes were flash-frozen in liquid nitrogen and stored at -80 °C.

Transformation of chemically competent *E. coli* cells

5 - 10 µl of the InFusion reaction or 50 - 100 ng plasmid DNA was added to 50 µl competent cells and incubated on ice for 10 min. Suspension was heat-shocked at 42 °C for 45 sec and then cooled down on ice for 2 min. 450 µl LB medium was added and cells were incubated for 30 - 60 min at 37 °C at 350 rpm shaking. Cells were harvested by centrifugation for 5 min at 4,000 rpm, resuspended in 150 µl LB medium and plated on an agar plate with the appropriate antibiotic resistance. Plate was incubated overnight at 37 °C.

Isolation and verification of plasmid DNA

Single colonies from selective plates were picked and used to inoculate 5 ml LB medium containing the appropriate antibiotic resistance. Cultures were grown overnight at 37 °C and plasmid DNA was isolated using a Miniprep purification kit (Qiagen). Presence of the insert was verified by PCR and positive clones were send for sequencing (GATC, Seqlab).

6.1.2 *S. cerevisiae*

Preparation of competent *S. cerevisiae* cells and transformation

50 ml YPD medium supplemented with ampicillin and tetracycline (see Table 8) was inoculated with an overnight pre-culture of the yeast strain and grown at 30 °C and 155 rpm shaking to an OD_{600} of 0.3 to 0.8. Per transformation 10^8 cells were used. Cells were harvested by centrifugation for 5 min at 2,500 rpm. Cells were washed with sterile water, centrifuged, resuspended in 1 ml sterile water and subsequently washed with 1 ml LiAc-TE buffer (see Table 11). After centrifugation as before cells were resuspended in LiAc-TE buffer at a concentration of 10^9 cells/ml. 2.5 µl boiled carrier DNA (herring sperm at 10 mg/ml, Invitrogen) and 200 ng plasmid DNA was added to each transformation, mixed gently and incubated for 10 min at RT. 260 µl LiAc-TE-PEG buffer (see Table 11) was added, mixed and incubated for 30 min at RT. 43 µl DMSO was added and a heat shock at 42 °C for 5 min was carried out. Subsequently, 1 ml YPD medium was added to the transformation mix and cells were recovered for 1 h at 30 °C and 350 rpm shaking. Cells were centrifuged as before, resuspended in 100 µl sterile water and plated. Plates were incubated for 3 days at 30 °C.

Plasmid shuffling

The yeast strain used for TFA1 genetic assays was transformed with a plasmid containing 3xFlag-tagged TFA1 (pSH810, ars cen LEU2 3xFlag) (see Table 2). Transformants were streaked once onto –Ura –Leu plates, once onto –Leu plates, twice onto 5-FOA plates, and subsequently onto YEPD plates. Successful plasmid shuffling was verified by immunostaining (see Section 6.2).

Preparation of glycerol stocks

Glycerol stocks of yeast strains were prepared for storage. A match-sized amount of yeast colonies was scraped from a plate and resuspended in 750 ml of a sterile 30 % glycerol solution. The stock was stored at -80 °C (no flash-freezing in liquid nitrogen).

Protein extraction from yeast cells for electrophoretic analysis

A small amount of yeast cells from plate was resuspended in 1 ml cold water. 150 µl treatment solution (see Table 12) was added and the sample was mixed. After an incubation time of 15 min on ice, 150 µl of 55 % trichloroacetic acid (TCA) was added. The sample was mixed, again incubated on ice for 10 min and centrifuged for 10 min at 4 °C and 14,000 rpm in a table centrifuge. The supernatant was discarded and the pellet resuspended in 50 µl sample buffer (see Table 12). The sample was neutralised with 1 M Tris (pH 8) and incubated at 65 °C for 10 min. After another centrifugation step for 5 min at RT and 14,000 rpm the supernatant was transferred to a new Eppendorf tube and 10 µl were loaded on a denaturing Bis-Tris acrylamide gel for western blot analysis (see Section 6.2).

6.2 General protein methods

Protein expression, purification and storage

All proteins were expressed in *E. coli* BL21(DE3)RIL cells (see Table 1). The identity of all purified proteins was confirmed by mass spectrometry. All purified proteins and complexes were flash-frozen in liquid nitrogen and stored at -80 °C.

Determination of protein concentration

Total protein concentrations were determined by measuring the absorption at 280 nm using a NanoDrop-2000 spectrophotometer (Thermo Scientific). Individual molar absorption coefficients of proteins and nucleic acid scaffolds were calculated using Expasy ProtParam (Gasteiger et al., 2005). Alternatively, protein concentrations were determined by Bradford assay (Bradford, 1976). The Bio-Rad Protein Assay Dye reagent (Bio-Rad)

was used according to the manufacturer's instructions and absorption of the samples was measured at a wavelength of 595 nm. Protein concentrations were calculated based on reference curves of bovine serum albumin (Fraction V, Roth).

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins were visualised by SDS-PAGE analysis. Protein samples were mixed with 5 x SDS-PAGE loading dye (see Table 9), incubated at 95 °C for 2 - 5 min and loaded on a precast 4 - 12 % NuPAGE Bis-Tris gradient gel (Invitrogen) together with a protein size marker (PageRuler Prestained Protein Ladder, Thermo Scientific). Gels were run using either 1x MOPS or 1x MES SDS running buffer (Invitrogen) at 200 V for 45 - 60 min. Gels were stained with InstantBlue (Expedeon) and imaged.

Native PAGE

Native PAGE was used for visualisation of cross-linked protein complexes. All steps were carried out at 4 °C. A precast 3 - 12 % NuPAGE Bis-Tris gel (Invitrogen) was pre-run using 1x NativePAGE running buffer (Invitrogen) at 70 V for 20 min. Protein samples were mixed with sample buffer supplemented with glycerol. Wells were rinsed with 1 x NativePAGE Dark Blue buffer (Invitrogen) and protein sample as well as protein size marker (NativeMark Unstained Protein Standard, Invitrogen) were loaded beneath the Dark Blue buffer. The gel was run first at 70 V then at 100 V for 4 - 5 h. For visualisation of nucleic acids the gel was incubated in a 1:10,000 solution of SYBR Gold Nucleic Acid Gel Stain (Thermo Scientific) for 15 min and imaged using a gel imaging system (Intas). Afterwards the gel was stained with InstantBlue (Expedeon) and imaged.

Silver staining

SDS-PAGE acrylamide gel was first soaked in 50 % (v/v) ethanol for 20 min, then in 5 % (v/v) ethanol for 20 min, then in 35 µM DTT for 5 min and finally in silver nitrate solution (see Table 9) for 10 min. The gel was rinsed twice with water and developed in developing solution (see Table 9) to the desired darkness of the protein bands. Solid citric acid monohydrate was added to neutralised the pH and stop the reaction. The gel

was transferred to water and imaged.

Western blotting and immunostaining

For western blotting the Trans-Blot Turbo Blotting system (Bio-Rad) was used. Bio-Rad Mini-Protean TGX 4 - 15 % gels (Bio-Rad) were run in 1 x Tris-Glycin running buffer at 200 V for 30 - 45 min. Using the Trans-Blot Turbo Transfer Pack (Bio-Rad) the sandwich for blotting was assembled according to the manufacturer's instructions. The blotting program for high molecular weight (MW) proteins was run. After transfer membrane was blocked with 2 % milk in PBS-T (see Table 9) for 1 h at RT. 6 - 10 ml of primary antibody (see Table 6) diluted in 2 % milk in PBS-T were added and incubated for 1 h at RT or at 4 °C overnight. The membrane was washed three times with 1 x PBS-T for 15 min for each wash. In case of a HPR-coupled primary antibody (see Table 6), the membrane could be developed directly (see below). In case a secondary antibody was required, 10 ml secondary antibody, diluted in 2 % milk in PBS-T, was used and incubated for 1 h at RT. The membrane was washed 3 times with 1x PBS-T for 15 min for each wash and developed using the Super Signal West Pico Chemiluminescent Substrate (Thermo Scientific) and imaged with an Advanced Fluorescent Imager (Intas).

Trichloroacetic acid (TCA) protein precipitation

Trichloroacetic acid precipitation was used to concentrate dilute protein solutions for visualisation by SDS-PAGE. TCA was added to a final concentration of 10 % to the protein sample and incubated for 15 min on ice. The sample was centrifuged for 10 min at 14,000 rpm and 4 °C and supernatant was carefully removed. The pellet was washed with 1 ml ice-cold acetone, centrifuged as before and supernatant removed. Residual acetone was evaporated by incubating the sample for 5 min at 95 °C. Pellet was resuspended in 15 µl 1x SDS loading dye and analysed by SDS-PAGE.

Mass spectrometry

Protein identification was carried out by mass spectrometry by Thomas Fröhlich at the Genecenter Munich or by Monika Raabe and Annika Kühn from the Urlaub lab at the

MPI Göttingen.

Dynamic light scattering

Dynamic light scattering was used to determine the size distribution in protein solutions. By this the stoichometry of a protein complex can be evaluated. 70 µl protein sample at a concentration of 2 mg/ml was transferred into a quartz cuvette and measured with a Viscotek 802 DLS (Malvern Instruments).

Thermal shift assay

Buffer conditions for TFIIF were optimised using a thermal shift assay. The fluorophore SYPRO Orange (Invitrogen) binds to hydrophobic protein regions which become accessible upon protein denaturation. In the assay protein stability is measured under various conditions over a temperature gradient where the turning point between folded and denatured protein state is defined as T_m. The screen was carried out in 50 µl reactions containing 4 µg protein, 50 mM buffer and 5x SYPRO Orange (Invitrogen). A pH range from 4.5 to 9.0 (sodium acetate pH 4.5, sodium citrate pH 5.5, sodium phosphate pH 6.2, MES pH 6.2, Bis-Tris pH 6.5, MES pH 6.5, MOPS pH 7.0, HEPES pH 7.0, ammonium acetate pH 7.3, Tris pH 7.5, HEPES pH 8.0, Tris pH 8.0, bicine pH 9.0) was covered and salt concentrations from 0 to 500 mM sodium chloride. Samples were mixed, sealed and analysed in a Real-Time PCR Thermal Cycler (Bio-Rad). Fluorescent signal was measured at 472 nm for each temperature step ranging from 20 °C to 95 °C.

Protein crystallisation

Initial crystallisation screening was performed by the crystallisation facilities at the MPI of Biochemistry, Martinsried and the MPI of Biophysical Chemistry, Göttingen. Various commercial screens and in-house produced screens were used and sitting-drops of 100 nl were set up at 20 °C. Initial crystal hits were refined manually in 15-well hanging drop plates (Qiagen).

Bioinformatic tools

Protein and gene sequences were downloaded from the NCBI or *S. cerevisiae* genome (SGD) databases. Sequence data was visualised and processed in Ape (Davis et al., 2002). Bioinformatic analysis was performed with the help of the Bioinformatics Toolkit (Alva et al., 2016). Multiple sequence alignments were generated using Clustal Omega (Sievers et al., 2011), MSAProbs (Liu et al., 2010b), MUSCLE (Edgar, 2004) and Aline (Bond and Schüttelkopf, 2009). Protein secondary structures were predicted by Quick2D (Bioinformatics Toolkit).

6.3 Specific methods

Only methods developed and performed as part of this thesis are described in this section. Methods carried out by other coauthors regarding Section 7 are described in the methods part of the publication (Plaschka, Hantsche et al., 2016). Parts of the methods described here have been published:

C. Plaschka*, **M. Hantsche***, C. Dienemann, C. Burzinski, J. Plitzko, P. Cramer. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**(7603): 353-358. doi: 10.1038/nature17990

W. Mühlbacher, S. Sainsbury, M. Hemann, **M. Hantsche**, S. Neyer, F. Herzog, P. Cramer. (2014) Conserved architecture of the core RNA polymerase II initiation complex. *Nature Communications* **5**:4310. doi: 10.1038/ncomms5310
For details on author contributions see page x.

6.3.1 Protein expression and purification

Purification of endogenous RNA Polymerase II

Fermentation 0.5 l YPD medium, supplemented with ampicillin and tetracycline (see Table 8), was inoculated with the respective yeast strain (see Table 2) from a fresh plate and grown overnight at 30 °C and 155 rpm shaking. The overnight cultures were used to inoculate 6 l YPD medium, containing ampicillin and tetracycline, and cultures were grown to an OD_{600} of 3 - 4. A 250 l-fermenter was inoculated with this culture to a starting OD_{600} of 0.1 - 0.2 and grown to a final OD_{600} of 9 - 10. Cells were harvested and 1 kg of cell pellet was resuspended in 330 ml 3x freezing buffer (see Table 13) at 4 °C,

aliquoted in 200 ml, flash-frozen in liquid nitrogen and stored at -80 °C.

Protein purification Cell suspension was lysed using 200 ml glass beads (Roth) in a beat beater (Hamilton Beach). Lysis was carried out for 90 min at 4 °C with duty cycles of 30 sec on and 90 sec off. Cell lysate was centrifuged twice for 30 min at 14,000 rpm at 4 °C and subsequently ultracentrifuged for 2 h at 42,000 rpm in a Ti45 rotor (Beckmann Coulter). Aqueous phase was pooled and proteins were precipitated with 50 % (w/v) solid ammonium sulphate stirring overnight at 4 °C. After centrifugation twice for 45 min at 14,000 rpm at 4 °C, the pellet was resuspended in 140 ml HSB 0/7 buffer (see Table 13) per 100 g pellet. The conductivity was set to that of buffer HSB 1000/7 (see Table 13) by adding HSB 0/7 buffer. The sample was incubated for 60 min at 4 °C with 12 ml Ni-NTA agarose (Qiagen), equilibrated in HSB 0/7 buffer. After washing with 5 column volumes (CV) HSB 1000/7 and 3 CV Ni 7 buffer, proteins were eluted with 3 CV Ni 50 buffer and 3 CV Ni 100 buffer (see Table 13). Conductivity of pooled fractions was adjusted to that of buffer MonoQ 150 by addition of MonoQ 0 buffer (see Table 13). Sample was loaded to a MonoQ 10/100 column (GE Healthcare), equilibrated with MonoQ 150 buffer, and eluted with a linear gradient from 150 -1500 mM potassium acetate over 12 CV. Fractions containing Pol II were pooled and twofold molar excess of recombinant Rpb4-Rpb7 (see below) was added and incubated on ice for 45 min. A Superose 6 10/300 column (GE Healthcare) was equilibrated with Pol II buffer (see Table 13) and sample was applied in 3 -5 runs. Fractions containing Pol II were concentrated to 3 - 4 mg/ml.

Expression and purification of Rpb4-Rpb7 Recombinant Rpb4-Rpb7 was co-expressed in *E. coli* cells and expression was carried out in autoinduction medium. 2 l LB medium supplemented with ampicillin, 1x 5052 and 1x NPS (see Table 8) was inoculated and grown at 37 °C until OD_{600} reached 0.6 - 0.8. The cultures were cooled on ice and subsequently incubated for 24 h at 20 °C. Cells were harvested by centrifugation for 10 min at 4,000 rpm. Pellets of 2 l culture were resuspended in 50 ml Rpb4-Rpb7 freezing buffer (see Table 13) and lysed by sonication. The lysate was cleared by centrifugation and loaded twice on equilibrated 3 ml Ni-NTA agarose (Qiagen). The resin was washed with 5 CV Ni 0 buffer, then with each 3 CV of salt buffer, Ni 10 buffer and Ni 20 buffer (see Table 13). Proteins were eluted by 3 CV Ni 50 buffer and 6 CV Ni 200 buffer. Pooled elution fractions were diluted with MonoQ 0 buffer matching the conductivity of MonoQ

100 buffer and applied to a MonoQ 10/100 column (GE Healthcare), equilibrated in MonoQ 100 buffer (see Table 13). After washing with 2 CV, elution was carried out with a linear salt gradient from 100 - 2000 mM sodium chloride over 10 CV. Rpb4-Rpb7 was further purified by gel filtration using a Superdex 75 10/300 column (GE Healthcare) equilibrated in Pol II buffer (see Table 13) and concentrated to 6 mg/ml.

Expression and purification of recombinant TFIIE

Recombinant TFIIE was obtained by co-expression of its subunits Tfa1 and Tfa2 in *E. coli*. Cells were transformed and grown in LB medium at 37 °C to an OD_{600} of 0.6, and expression was induced by 0.5 mM IPTG for 18 h at 18 °C. Cells were lysed by sonication in lysis buffer (see Table 14). The lysate was cleared by centrifugation and applied to a 5 ml HisTrap HP column, equilibrated in HisTrap buffer (see Table 14). The column was washed with 10 CV HisTrap buffer and eluted with HisTrap buffer containing 250 mM imidazole. TFIIE was then subjected to anion exchange chromatography using a 5 ml HiTrap Heparin column (GE Healthcare), equilibrated in HeparinTrap buffer (see Table 14). The complex was eluted with a linear gradient of HeparinTrap buffer from 100 - 2000 mM sodium chloride. To improve purity, TFIIE was further applied to Superose 12 10/300 size exclusion column, in gel filtration buffer (see Table 14). TFIIE containing fractions were pooled and concentrated to 9.6 mg/ml. TFIIE mutants were purified according to the same protocol.

Expression and purification of recombinant TFIIF

Owning to difficulties with cloning and expression of the *S. cerevisiae TFG1* gene in *E. coli* cells, the closely related gene encoding *S. mikatae TFG1* was used for a long time for studies with recombinant yeast TFIIF (Chen et al., 2007; Eichner et al., 2010; Mühlbacher et al., 2014; Plaschka et al., 2015). *S. mikatae* and *S. cerevisiae* Tfg1 share a sequence identity of 86 % and *S. mikatae* Tfg1 complements *TFG1* function in *S. cerevisiae* (Chen et al., 2007). A functional *E. coli* promoter and an internal translation initiation site in the N-terminal coding region of the *TFG1* gene were reported to be the cause for the toxicity of Tfg1 (Yang et al., 2010). However, our efforts to clone the *TFG1* gene with the proposed mutations (Yang et al., 2010) into *E. coli* cells were unsuccessful. In order to obtain recombinant *S. cerevisiae* Tfg1 we used a codon-optimised sequence for expression

in *E. coli* (Life Technologies). TFIIF subunits Tfg1 and Tfg2 were co-expressed in *E. coli* and cells were grown in LB medium at 37 °C to an OD_{600} of 0.8. Expression was induced with 0.2 mM IPTG for 3 h at 37 °C. Cells were lysed by sonication in lysis buffer (see Table 15). Cleared lysate was applied to a 5 ml HisTrap HP column equilibrated in lysis buffer. The column was washed with 8 CV lysis buffer containing 1000 mM KCl, and eluted with a linear gradient from lysis buffer to HisTrap buffer (see Table 15). The conductivity of the eluate was adjusted to match that of HiTrap SP buffer and 3C protease cleavage was carried out for 2 h. The complex was then applied to cation exchange chromatography using a 1 ml HiTrap SP HP column (GE Healthcare), equilibrated in HiTrap SP buffer (see Table 15), and eluted in a linear gradient from 150 - 1000 mM potassium chloride. TFIIF was further purified by size exclusion chromatography using a Superdex 200 10/300 Increase column (GE Healthcare), in gel filtration buffer (see Table 15). Purified TFIIF was concentrated to 3.9 mg/ml. TFIIF mutants were purified according to the same protocol.

Expression and purification of selenomethionine-substituted TFIIF

TFIIF plasmid was transformed into the methionine auxotrophic *E. coli* strain B834 (see Table 1) as described (see Section 6.1). Transformed cells were grown in LB medium, supplemented with ampicillin and chloramphenicol (see Table 8), to an OD_{600} of 0.5 at 37 °C. Cultures were centrifuged for 10 min at 4,000 rpm and pellets resuspended in SeMet minimal medium (see Table 7). 1 l SeMet minimal medium, supplemented with ampicillin and chloramphenicol, was inoculated and SeMet added to a final concentration of 100 mg/l. Cultures were grown at 37 °C until OD_{600} increased by 0.2. Then cultures were cooled down, protein expression was induced by addition of 0.1 mM IPTG and cultures were incubated overnight at 18 °C. Cells were harvested by centrifugation as before. Protein purification was carried out as described above.

6.3.2 Cryo-EM structure of the closed complex

Formation and isolation of the closed complex

The closed complex (CC) contained a nucleic acid scaffold based on the yeast *HIS4* promoter (see Table 4). The initiation complex was prepared as follows. Pol II (200 µg at

3 mg/ml) was incubated with a fourfold molar excess of TFIIF. A twofold molar excess of nucleic acid scaffold over Pol II, tenfold molar excess of TFIIA, fourfold molar excess of TBP and TFIIB were added to Complex buffer and incubated with preformed Pol II-TFIIF complex for 8 min at 25 °C. TFIIE was added in a tenfold molar excess over Pol II and incubated for 5 min at 25 °C. cMed-Med1 was added in a 1.2-fold molar excess over Pol II and incubated for 50 min at 25 °C. The complex was purified using a Superose 6 3.2/300 size exclusion column (GE Healthcare), equilibrated in Complex buffer (see Table 13). Fraction containing the complex were pooled (0.4 - 0.8 mg/ml) and additionally incubated with equimolar amounts of nucleic acid scaffold.

Protein cross-linking

Closed complex samples were cross-linked for 30 min on ice using 0.1 % glutaraldehyde (Electron Microscopy Sciences), and the reaction was quenched with 50 mM lysine (Sigma-Aldrich). The cross-linked sample was repurified in a second size exclusion step using a Superose 6 3.2/300 column, equilibrated in Complex buffer lacking glycerol (see Table 13). Fractions containing initiation complexes were pooled (0.2 - 0.6 mg/ml) and used for EM grid preparation.

EM-grid preparation and data collection

Negative stain EM Protein samples between 40 - 100 µg/ml were applied on carbon-coated mesh grids (Plano) for quality control. Grids were charged for 20 sec and 4 µl protein sample was applied and incubated for 2 min before excess sample was removed with filter paper. Grids were washed twice with each a 4 µl water drop. Finally, a 4 µl drop of freshly prepared 2 % (w/v) uranyl formate was added and incubated for 1 min. Afterwards, excess uranyl formate was removed and grids were stored on filter paper in a petri dish in the dark.

Negative stain data was acquired on a CM200 (Phillips) at 200 keV equipped with a 4,000 x 4,000 CCD camera. Images were acquired at a nominal magnification of 88,000x (2.51 Å/pixel) and with an exposure time of 500 ms.

cryo-EM Closed complex samples were applied to R3.5/1 holey carbon grids (Quantifoil). Grids were glow-discharged for 15 s before deposition of 4 µl complex, and subsequently blotted and vitrified by plunging into liquid ethane with a Vitrobot Mark IV (FEI) operated at 4 °C and 100 % humidity. Cryo-EM data was acquired on a FEI Titan Krios operated in EFTEM mode at 300 keV, and equipped with a K2 Summit direct detector (Gatan). Automated data collection was carried out using the TOM^2 toolbox (Korinek et al., 2011) to acquire 959 movies with a range of defocus values (from –0.8 µm to –5 µm) at a nominal magnification of 37,000x (1.35 Å/pixel). The camera was operated in 'super-resolution' mode (0.675 Å/pixel), with a total exposure time of 6 s fractionated into 20 frames, a dose rate of $\sim 8\,e^-/(\text{pixel}\,s)$, and total dose of $40\,e^-/Å^2$ per movie.

Data processing and model building

For single particle analysis of the closed complex, an initial set of 10,591 particles was selected semi-automatically using e2boxer.py from EMAN2 (Tang et al., 2007). CTF parameters were estimated using CTFFIND (Rohou and Grigorieff, 2015). CTF correction and subsequent image processing was done with RELION 1.3 (Scheres, 2012), unless otherwise noted. Resolution was reported based on the gold-standard FSC (0.143 criterion) (Chen et al., 2013) and temperature factors were determined and applied automatically in RELION (Scheres, 2012). Selected particles were extracted with a 300^2 pixel box and preprocessed to normalise images. Reference-free 2D class averages were calculated, and three representative classes were low-pass filtered to 30 Å resolution and used as templates for automated picking of all micrographs (Scheres, 2015). The resulting 155,079 particle images were screened manually and by reference-free 2D classification, yielding 111,625 particle images that were used for subsequent processing. The 7.8 Å cryo-EM map of the yeast cITC (EMD-2785) (Plaschka et al., 2015) was low-pass filtered to 50 Å and used as initial model for 3D refinement of the 10,591 particle set. This revealed an closed complex density to an estimated resolution of 14 Å. This density was low-pass filtered to 50 Å and used for processing of the complete closed complex cryo-EM data set. A 3D reconstruction of all particles was calculated to 7.5 Å, and subjected to particle polishing using RELION 1.4 beta (Scheres, 2012). This lead to an improved density at 6.5 Å resolution.

Hierarchical 3D classification without image alignment using a soft mask encompassing the complete CC resulted in two populations, CC and OC5, respectively (see

Figure 9). Both classes were refined using the 3D auto-refine procedure against the respective particles within that class with a soft mask in the shape of the CC. The CC soft mask was also suitable for the OC since both complexes have a similar shape. The CC reconstruction was determined from 7,527 particles to a resolution of 8.2 Å. The OC5 reconstruction was determined from 79,797 particles to a nominal resolution of 6.1 Å and a temperature factor of -176 $Å^2$ was applied. To improve the densities of TFIIE, Tfg2 WH and downstream DNA in the CC, we carried out focused classification of the Pol II stalk and subsequently TFIIE, Tfg2 WH and downstream DNA using soft masks encompassing these regions. The individual classes were refined as before and yielded a CC at 8.8 Å comprising 5,690 particles and a temperature factor of -300 $Å^2$ was applied. A similar focused classification scheme for OC5 revealed moderately improved density for TFIIE (not shown). Local resolution estimates were determined using a sliding window of 40^3 voxels as previously described except that a single pair of two half-maps was used and resolution estimates were not capped at the nominal resolution as no local filters were applied.

To generate the CC model, the high-resolution OC model (PDB 5FYW) of Pol II, TFIIA, TFIIB, TBP, TFIIE E-ribbon and TFIIF dimerisation domain was rigid-body fitted into the CC density using an automated global 6D correlation search in Situs (Wriggers, 2010). The remaining protein part was divided into two rigid-bodies containing (i) Tfa1 and Tfa2 WH2, and (ii) Tfa2 WH1 and Tfg2 WH, which were independently fitted into the CC density using Situs. To model closed promoter DNA, upstream DNA was extended with canonical duplex B-form DNA in COOT (Emsley and Cowtan, 2004) and rigid-body fitted in UCSF Chimera (Pettersen et al., 2004) to reflect the density. The model was validated with the FSC of map versus model (see Figure 9). The model is available under the accession number PDB 5FZ5.

6.3.3 Functional assays

Nuclear extract preparation

3 l YPD medium, supplemented with ampicillin and tetracycline (see Table 8), was inoculated with a 3xFlag-Tfa1 yeast pre-culture (see Table 2) to a starting OD_{600} of 0.2 and grown over night to a final OD_{600} of 7 - 8 at 30 °C and 155 rpm shaking. Cells were harvested by centrifugation at 4,000 rpm for 10 min at 18 °C using a F9 rotor (Thermo Scientific). The supernatant was discarded and pellets was resuspended in

35 ml resuspension buffer (see Table 16) by gentle manual shaking, followed by 15 min incubation in a water bath shaker at 30 °C and 100 rpm shaking. After centrifugation at 4,000 rpm for 10 min at 18 °C, pellets were resuspended in 20 ml YPDS (see Table 7) by gentle manual shaking. For cell wall digestion and spheroplasting, 3 ml of a 2 M sorbitol solution and 3 ml resuspension buffer supplemented with 18 mg zymolyase (Siekagaku) and 1x PI was added and incubated for 30 min in a water bath shaker at 30 °C and 100 rpm shaking. The progress of spheroplasting was monitored by measuring OD_{600} (2 µl cell suspension in 1 ml water), the final OD_{600} should be 1/10 of the initial OD_{600}. To stop the reaction, 100 ml YPDS was added and cells were centrifuged as before. The pellet was resuspended in 250 ml YPDS and incubated for 1 h in a water bath shaker at 30 °C and 100 rpm shaking. After centrifugation, the pellet was resuspended in 200 ml ice-cold YPDS. The following steps were carried out on ice and ice-cold buffers were used. Cells were centrifuged as before but at 4 °C and resuspended again in 200 ml YPDS. This washing step was repeated and after centrifugation the pellet was resuspended in 250 ml 1 M sorbitol and again centrifuged as before.

For cell membrane disruption, the pellet was resuspended in 100 ml lysis buffer (see Table 16) and the cell suspension was passed three times through a pre-chilled Dounce glass homogeniser (Kontes). The lysate was transferred to 200 ml centrifugation bottles and centrifuged at 5,000 rpm for 10 min at 4 °C using a F14 rotor (Thermo Scientific). The step was repeated twice each time using fresh centrifugation bottles. The supernatant was then transferred into two 50 ml centrifugation tubes and centrifuged at 13,000 rpm for 30 min at 4 °C using a A27 rotor (Thermo Scientific). Supernatant was discarded and crude nuclei were resuspended in 15 ml centrifugation buffer (see Table 16) using a sterile inoculation loop.

For nuclear membrane disruption, 3 M ammonium sulphate solution (pH 7.5) was added to a final concentration of 0.5 M and incubated at 4 °C for 30 min on a spinning wheel. Nuclear extract was separated from nuclear membrane components by an ulta-centrifugation step at 28,400 rpm for 90 min at 4 °C using a SureSpin 630 rotor (Thermo Scientific). To precipitate nuclear extract, 0.35 g fine ground ammonium sulphate was added per ml supernatant, incubated at 4 °C for 30 min on a spinning wheel and ultra-centrifuged twice at 10,200 rpm for 20 and 4 min, respectively, at 4 °C, discarding the supernatant. Nuclear extract was resuspended in 250 - 350 µl dialysis buffer (see Table 16) using a sterile inoculation loop and dialysed against 1.5 l dialysis buffer containing 75 mM ammonium sulphate for for 4.5 h. 50 µl aliquotes were flash-frozen in liquid nitrogen and stored at -80 °C.

Immunodepletion of nuclear extract

Nuclear extract was depleted of TFIIE by immunodepletion of 3xFlag-Tfa1. 150 µl anti-Flag M2 agarose (Sigma) was washed twice with dialysis buffer (see Table 16) containing 75 mM ammonium sulphate and then incubated with 1 mg/ml bovine serum albumin (BSA) protein (Sigma) for 1 h at 4 °C on a spinning wheel. After three wash steps conducted as before, the suspension was divided into two Eppendorf tubes. M2 agarose bead suspension was centrifuged for 5 min at 3,000 rpm at 4 °C and supernatant was discarded. 70 µl nuclear extract was added to the beads and incubated for 1 h at 4 °C on a spinning wheel. Immunodepleted nuclear extract was separated from beads by Micro Bio-Spin chromatography columns (Bio-Rad) and added to the second Eppendorf tube containing fresh M2 beads. Another round of immunodepletion was carried out as before. Specificity of the depletion was confirmed by western blotting and immunostaining (see Section 6.2).

In vitro transcription assay

Pol II *in vitro* transcription assay was carried out on a yeast *HIS4* promoter which was inserted into a pBluescript II KS+ plasmid (see Table 3). The amount of Pol II in the nuclear extract was estimated by a standard curve using immunostaining. A reaction volume of 25 µl contained 3-5 µl (equivalent to 2 pmol Pol II) depleted yeast nuclear extract, 5 pmol TFIIE (wild type or mutant), 150 ng template plasmid, 1x acetate transcription buffer (see Table 17), 192 µg phosphocreatine (Sigma), 0.2 µg creatine phosphokinase (Sigma), 10 U RiboLock RNase inhibitor (Thermo Scientific), 100 µM NTPs (Thermo Scientific) and 150 ng recombinant Gcn4 activator. The reaction mix was incubated for 1 h at 18 °C. The synthesised RNA was isolated using the RNeasy MinElute Cleanup Kit (Qiagen) according to the manufacturer's instructions. RNA was eluted with 14 µl RNase-free H_2O and transcripts were analysed by primer extension followed by a reverse transcriptase step. 20 µl reaction mix contained 12 µl RNA, 1x annealing buffer (see Table 17) and 0.125 pmol fluorescently labelled reverse transcriptase primer (see Table 4). The sample was boiled for 2 min at 95 °C and then primers were annealed for 45 min at 48 °C in the dark. After addition of 40 µl synthesis mix containing 1x synthesis buffer (see Table 17), 2 µg actinomycin D (Sigma-Aldrich), 0.15 mM dNTPs (Thermo

scientific) and 50 U M-MuLV reverse transcriptase (New England Biolabs), the sample was incubated for 30 min at 37 °C. The resulting cDNA was ethanol precipitated by addition of 1/10 of the volume of sodium acetate, 3 volumes ice-cold ethanol and 2 µl glycogen (Thermo Scientific) and incubation for 20 min on dry ice. After centrifugation for 10 min at maximum speed at 4 °C, the supernatant was removed and the pellet was washed with 180 µl ice-cold 80 % ethanol. The sample was centrifuged as before, the supernatant discarded and residual liquid removed using SpeedVac (Concentrator plus, Eppendorf). The pellet was resuspended in 4 µl RNase (40 µg/ml, Qiagen), incubated for 3 min at RT. 4 µl urea loading buffer (see Table 9) was added and the sample boiled for 1 min. Transcripts were separated on a denaturing 8 % polyacrylamide TBE gel in 1x TBE, visualised with a Typhoon 9500 scanner (GE Healthcare) and quantified with ImageQuant (GE Healthcare). For quantification the relative activity of each variant compared to TFIIE was calculated for each replicate. The mean intensity and standard deviation of three replicates was calculated from their relative activities.

6.3.4 X-ray crystallographic analysis of the Pol II-TFIIF complex

Structural modelling of TFIIF

Models for the yeast TFIIF dimerisation module and WH domains were generated from known crystal structures (dimerisation module, chains A and F in PDB 1F3U; Tfg1 WH domain, PDB 1I27; Tfg2 WH domain, PDB 1BBY) using MODELLER (Eswar et al., 2006). Residues 92–153 and 324–417 in *S. cerevisiae* Tfg1 correspond to residues 5–62 and 73–168 in human RAP74. Residues 54–138 and 208–227 of *S. cerevisiae* Tfg2 align to residues 2–119 of human RAP30, respectively. Residues 678–736 of *S. cerevisiae* Tfg1 WH domain align with residues 454–517 in human RAP74 and residues 292–350 in *S. cerevisiae* Tfg2 WH domain align to residues 176–243 in human RAP30.

Fluorescent labelling of TFIIF

A S6-tag (Zhou et al., 2007) was cloned to the Tfg1 N-terminus to allow specific enzyme-mediated labelling of TFIIF with a fluorophore. The S6-tag consists of 12 amino acids and is specifically modified by SFP Synthase attaching a modified CoA moiety. The Starter Kit for *in vitro* labelling of proteins (New England Biolabs) was used according

to the manufacturer's instructions. The reaction mix was incubated for 30 min at 20 °C. Excess fluorophore was removed by gel filtration or desalting columns (Bio-Rad).

Formation of the Pol II-TFIIF complex

Pol II (at 3.5 mg/ml) was incubated with a twofold molar excess of TFIIF for 30 min at 20 °C in Pol II buffer (see Table 13). The complex was isolated using a Superose 6 10/300 increase column (GE Healthcare), equilibrated in Pol II buffer. Fractions containing stoichiometric complex were pooled and concentrated to the desired concentration.

Crystallisation of the Pol II-TFIIF complex

12-subunit Pol II was prepared (see Section 6.3.1) and crystallised by vapour diffusion with 2 - 7 % PEG 6,000, 300 mM sodium acetate, 200 mM ammonium acetate, 50 mM HEPES pH 7.0, 5 mM TCEP as reservoir solution. Crystals were grown for 3 - 5 days at 20 °C and were cryo-protected in mother solution containing 22 % glycerol. TFIIF was added to the cryo-protectant at 1 mg/ml and crystals were incubated overnight at 8 °C. Subsequently, crystals were fished and frozen in liquid nitrogen.

Identification of observed TFIIF density

Initial experiments performed by Kerstin Kinkelin revealed an additional electron difference density on the Pol II surface after soaking Pol II crystals with TFIIF (see above). To assign the observed density to a TFIIF region, several approaches were pursued. First, Pol II crystals were soaked with SeMet-labelled TFIIF (see Section 6.3.1), but no anomalous signal for selenium was detected. Second, analysis of cross-linking data of the endogenous Pol II-TFIIF complex revealed a high-confidence cross-link from the Pol II surface (Rpb2 K1057) directly next to the observed difference density to the N-terminus of Tfg1 (K23) (Chen et al., 2010b). Secondary structure predictions of Tfg1 suggested the presence of an α-helix from residues 25 to 36 matching the observed difference density. Subsequently, Pol II crystals were soaked with a peptide comprising the Tfg1 N-terminus (S19 - S41) (see Table 5) which resulted in a similar difference electron density as obtained by soaking with TFIIF. Soaking was repeated with a SeMet-labelled

peptide (see Table 5) to allow model building.

Data collection and X-ray structure determination of the Tfg1 N-terminal region bound to Pol II

Diffraction data were collected on a PILATUS 6M detector at the X06SA beamline (SLS, Villigen, Switzerland) at 100 K. Data for the Pol II–Tfg1 SeMet peptide crystal was collected at a wavelength of 0.97972 Å and for the Pol II–TFIIF crystal at 0.91889 Å. Data were processed with XDS and XSCALE (Kabsch, 2010). The structure was phased with the crystal structure of the 12-subunit Pol II (PDB code 3PO2) (Cheung and Cramer, 2011) lacking nucleic acid and refined using BUSTER (Blanc et al., 2004). Model building of the TFIIF N-terminal residues in COOT (Emsley and Cowtan, 2004) was guided by a selenium anomalous difference Fourier peak (M27). Subsequent refinement in BUSTER used secondary structure restraints for the Tfg1 peptide helix. The final structures had an R_{free} − factor of 18.0 % and 19.4 % for the Pol II–Tfg1 SeMet peptide crystal and the Pol II–TFIIF crystal, respectively, and showed good stereochemistry (Chen et al., 2010a). In the Pol II-Tfg1 SeMet peptide structure 90 % of the residues fall in favoured regions of the Ramachandran plot and 3 % in disallowed regions. For the Pol II-TFIIF structure 89 % of the residues fall in favoured regions of the Ramachandran plot and 3 % in disallowed regions. Coordinates and structure factors of the Pol II-Tfg1 peptide and the Pol II-TFIIF crystals have been deposited at the Protein Data Bank under the accession numbers 5IP7 and 5IP9.

Part III

Results and Discussion

7 Transcription initiation complex structures elucidate DNA opening

Results presented in this section were obtained in collaboration with Clemens Plaschka and have been published:

C. Plaschka*, **M. Hantsche***, C. Dienemann, C. Burzinski, J. Plitzko, P. Cramer. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**(7603): 353-358. doi: 10.1038/nature17990
For details on author contributions see page x.

The main results of this thesis have been published together with the high-resolution structure of the open complex in Plaschka, Hantsche et al., 2016. This results section is based on the publication which starts with the description of the open complex for a comprehensive understanding of all results. Supplemental data figures of the open complex are found in Sections 12 and 13.

7.1 Open transcription initiation complex at 3.6 Å

We extended our previous cryo-EM analysis of a Pol II core initiation complex containing TBP, TFIIB, and TFIIF (Plaschka et al., 2015) by adding TFIIA and TFIIE. Formation of a stable and stoichiometric complex required the presence of core Mediator, which showed, however, low occupancy and high flexibility under cryo-EM conditions, as observed before (Plaschka et al., 2015), and was excluded from further analysis (Supp. Figure 21a-c). We acquired 257,259 cryo-EM single particle images using a K2 direct electron detector (Supp. Figure 21b, c). Unsupervised particle sorting led to an OC structure at an overall resolution of 3.6 Å, revealing the Pol II core at up to 3.1 Å (OC1) (Figure 5a, b, Supp. Figure 21d-g). Further particle sorting revealed improved density for TFIIB and TFIIF at around 4 Å resolution (OC2 and OC4), and for TFIIE at 4.4 Å resolution (OC3) (Supp. Figure 21i, k, l). The final structure mainly consists of atomic models (90 %) and contains backbone models for parts of the basal factors (Supp. Figure 21f-l).

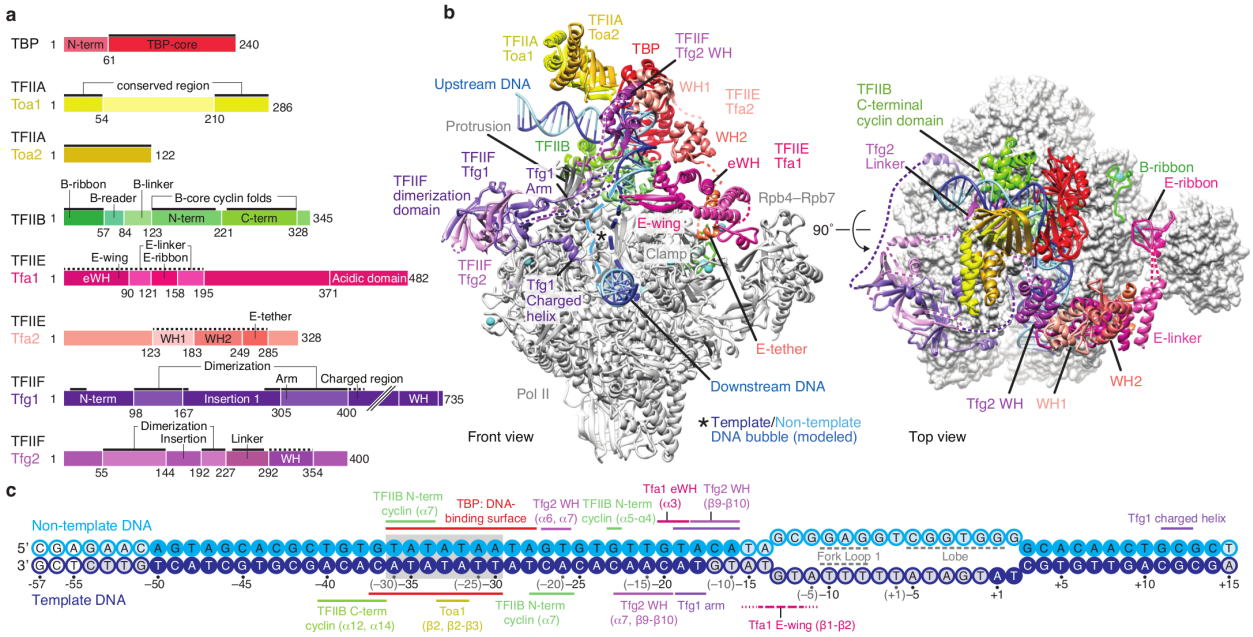

Figure 5: Open complex structure at 3.6 Å resolution. **a**, Domain organisation of yeast basal transcription factors TBP, TFIIA, -B, -E, and -E. Solid and dashed black bars indicate protein regions that are present in the OC structure as atomic and backbone models, respectively. Colour code used throughout. **b**, Two views (Cramer et al., 2001) of the yeast OC structure. Pol II is in grey. DNA template and non-template strands are in dark blue and cyan, respectively. On the right, Pol II is shown as a surface representation, all other proteins are shown as ribbon models. **c**, Protein-DNA contacts. Promoter DNA nucleotides are depicted with solid, shaded, and empty circles when they were included in the structure, excluded owing to weak density, or excluded owing to a lack of density, respectively. Solid and dashed lines indicate observed and putative protein interactions, respectively. A magenta dashed line indicates the contact between closed DNA and the TFIIE E-wing. The register of promoter DNA is given for analogous yeast (black) and human (grey) positions with respect to the transcription start site (TSS, +1). TATA box is indicated by a grey box.

The structure reveals upstream DNA above the Pol II wall and downstream DNA in the active centre cleft (Plaschka et al., 2015). We observed only fragmented density for single strands within the DNA bubble (Figure 5b, c, Supp. Figure 21m), and no density for RNA, which was not retained during cryo-EM sample preparation (not shown). The bubble contains 15 mismatches, begins at the natural distance of ~20 bp downstream of the TATA box (Giardina and Lis, 1993), and extends further downstream than when it initially forms, resembling the situation during transcription start site scanning. For the basal factors, regions that are essential for cell viability are generally observed, whereas non-essential, non-conserved regions (Grünberg et al., 2012; Nikolov et al., 1995; Geiger et al., 1996; Tan et al., 1996; Eichner et al., 2010; Deng and Roberts, 2007) are often mobile

(Figure 5a, Supp. Figure 21j-l). TFIIA is located near upstream DNA and TBP as expected (Geiger et al., 1996) (Figure 5b, Supp. Figure 21j). For TFIIB, the B-ribbon and B-core domains are well defined, including the newly modelled C-terminal cyclin domain from yeast (Figure 6), whereas the B-linker shows weak density, and the B-reader (Sainsbury et al., 2013; Plaschka et al., 2015) is mobile (Supp. Figure 21i).

7.2 DNA position and retention

7.2.1 Architecture of TFIIF

TFIIF adopts an intricate fold within the OC. Its dimerisation module and charged helix are located on the Pol II lobe domain as in the ITC (Mühlbacher et al., 2014; Plaschka

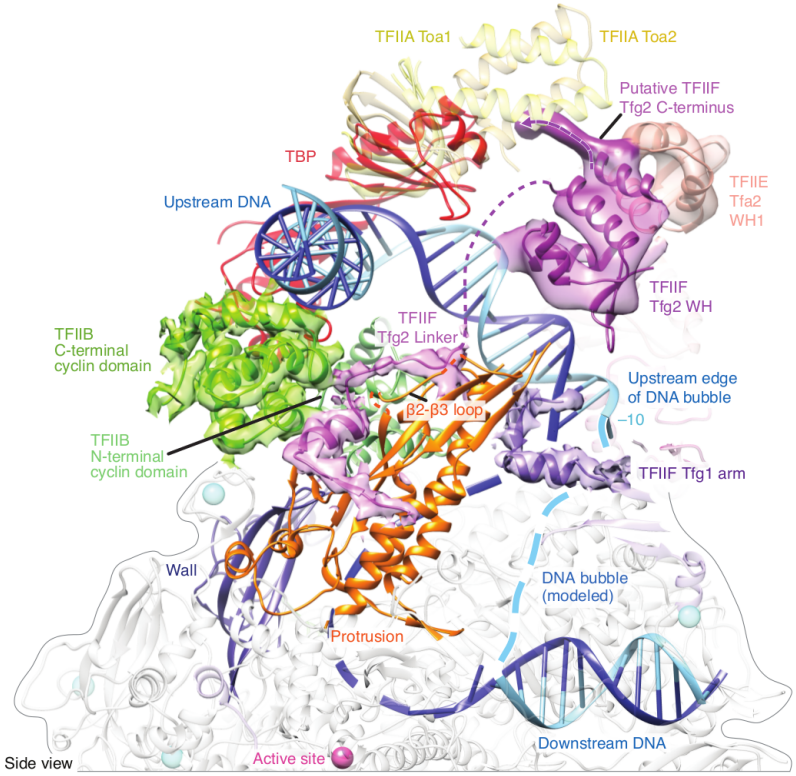

Figure 6: Basal factors position and retain DNA. Details of the upstream DNA assembly viewed from the side (Cramer et al., 2001). Highlighted are the locations of the Pol II wall, protrusion, TBP, TFIIB, TFIIF arm, Tfg2 linker and winged helix, TFIIE Tfa2 WH1, DNA (template, blue; non-template, cyan), and the active site. TFIIA is transparent. Cryo-EM densities for newly modelled regions of TFIIB, TFIIF, and TFIIE are superimposed on their structural models. Interactions with the Pol II protrusion and upstream edge of the DNA bubble are indicated. TBP contacts a density assigned to the Tfg2 C-terminal region, consistent with interaction of their human counterparts (Robert et al., 1996).

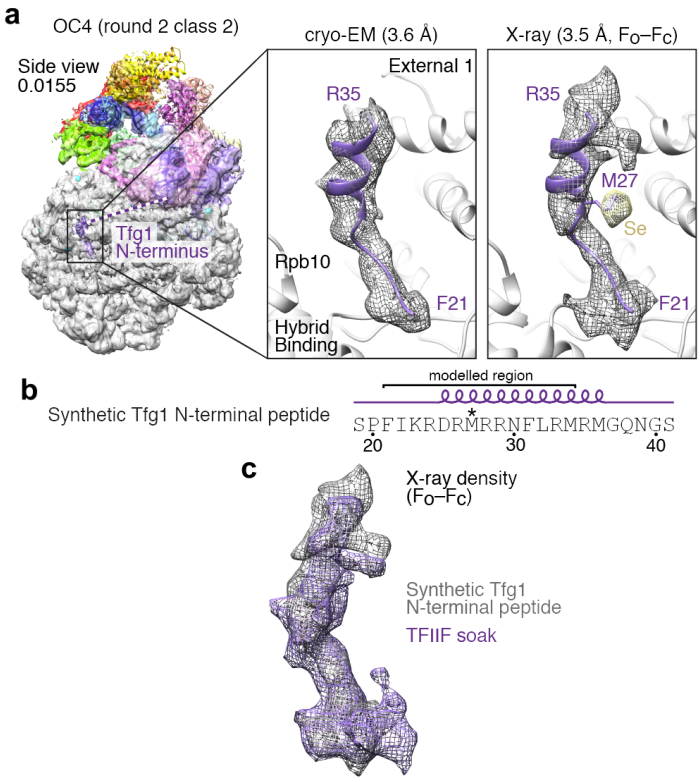

Figure 7: Tfg1 N-terminus binds to Pol II external 1 domain. a, Crystallographic analysis of the yeast-specific Tfg1 N-terminal region. Weak density for the Tfg1 N-terminal region was observed by cryo-EM (OC4 round 2 class 2) at low contour level (0.0155) close to Pol II elements external 1 and the hybrid binding region (left). X-ray analysis (right) of the corresponding peptide (Tfg1 W21-R35) enabled modelling and assignment of residue M27 due to the anomalous signal. The F_o-F_c electron density map (grey, contour level 2.5σ), seleno-methionine anomalous difference Fourier (yellow, contour level 5σ), and final model in ribbon presentation (purple) are shown. **b**, The sequence of the synthetic peptide used for soaking into Pol II crystals is shown. The modified methionine residue and predicted secondary structure are indicated. **c**, The F_o-F_c electron density maps obtained from soaking Pol II crystals with TFIIF (purple) and seleno-methionine labelled peptide (grey), respectively, show similar density in the same location on Pol II.

et al., 2015) (Supp. Figure 22a, c). The 'arm' in the large TFIIF subunit Tfg1 (human RAP74) adds a β-strand to the Pol II protrusion, and projects into the cleft, where it may stabilise the DNA bubble (He et al., 2013; Mühlbacher et al., 2014; Plaschka et al., 2015) (Figure 6). The linker in TFIIF subunit Tfg2 (human RAP30) emanates from the dimerisation module, and winds along the base of the protrusion, where it binds a hydrophobic pocket. The Tfg2 linker continues between the protrusion and the TFIIB cyclin domains, and connects to the Tfg2 C-terminal winged-helix (WH) domain on top of the cleft (Figure 6, Supp. Figure 22d). The Tfg2 linker stabilises TFIIB on the wall of

Table 18: X-ray crystallographic data collection and refinement statistics

	Pol II-Tfg1 SeMet	Pol II-TFIIF
Data collection		
Space group	C222(1)	C222(1)
Cell dimensions δ		
a, *b*, *c* (Å)	221.2, 392.7, 282.2	222.8, 392.7, 283.4
Resolution (Å)	50-3.5 (3.6-3.5)*	50-3.9 (4.0-3.9)*
R_{sym} (%)	11.3 (105.4)	13.6 (91.3)
$I/\sigma I$	15.7 (2.0)	12.8 (2.4)
Completeness (%)	100 (100)	99.8 (100)
Redundancy	7.1 (6.9)	7.5 (7.6)
$CC_{1/2}$ (%)	99.8 (78.9)	99.8 (82.6)
Refinement		
Resolution (Å)	49-3.5	48.9-3.9
No. reflections	151,192	112,717
R_{work}/R_{free} (%)	15.4/18.0	16.2/19.4
No. atoms		
Protein	31,330	31,330
Ligand/ion	9	9
B-factors (A^2)		
Protein	127.9	143.7
Ligand/ion	125.7	124.4
R.m.s deviations		
Bond length (Å)	0.01	0.01
Bond angles (°)	1.30	1.26

* Highest resolution shell is shown in parenthesis.

Pol II (Fishburn and Hahn, 2012; Cabart et al., 2011). The yeast-specific N-terminal region of Tfg1 binds near the Pol II external 1 domain, according to a separate crystallographic analysis (Figure 7, Table 18).

7.2.2 Architecture of TFIIE

TFIIE is located between the clamp and the Rpb4–Rpb7 stalk (Figure 8a, b, Supp. Figure 23), consistent with previous topological placement of TFIIE (Grünberg et al., 2012; He et al., 2013; Robert et al., 1996; Forget et al., 2004; Chen et al., 2007) and its archaeal counterpart (Grohmann et al., 2011). The TFIIE structure differs from a recent model obtained at 6 Å resolution (Murakami et al., 2015) (see legend in Supp. Figure 21e for details, Supp. Figure 23, and references (Grünberg et al., 2012; Murakami et al., 2013)).

The large TFIIE subunit Tfa1 (human TFIIEα) contains an extended winged helix ('eWH') domain (Meinhart et al., 2003) and a zinc ribbon domain ('E-ribbon') (Okuda et al., 2004) that are connected by α-helices (called here 'E-linker') (Figure 8b). The eWH domain uses its helix α3 to contact both the tip of the Pol II clamp helices and

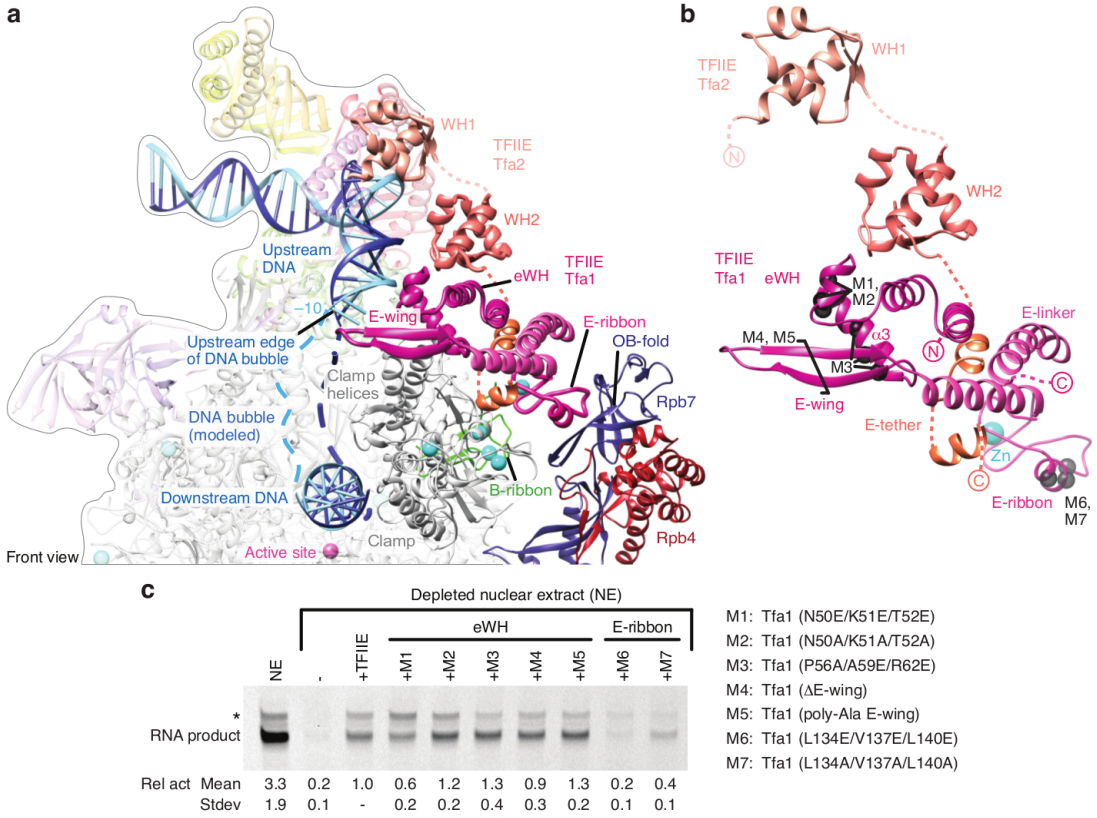

Figure 8: TFIIE architecture and interactions. **a**, TFIIE interactions within the OC. Depicted are interactions of the TFIIE E-ribbon with the Pol II clamp, stalk subunit Rpb7, and the TFIIB B-ribbon, and interactions of the TFIIE eWH domain with the Pol II clamp helices and upstream DNA. The eWH E-wing lies close to the upstream DNA edge, similar to WH domains involved in DNA strand separation. Colours as in Figure 5, except for the Pol II stalk (Rpb4, dark red; Rpb7, dark blue). **b**, TFIIE domain architecture. The TFIIE variants used for functional assays are indicated as Cα spheres for point mutations, and with a black bracket for E-wing alterations. Connectivity of the Tfa2 E-tether helices is uncertain. **c**, Selected TFIIE variants impair transcription from a *HIS4* promoter. TFIIE-depleted nuclear extract (NE) was reconstituted with recombinant TFIIE and TFIIE variants carrying mutations in the Tfa1 eWH (M1, Tfa1(N50E/K51E/T52E); M2, Tfa1(N50A/K51A/T52A); M3, Tfa1(P56A/A59E/R62E); M4, Tfa1(Δ E-wing); M5, Tfa1(poly-Ala E-wing) and the Tfa1 E-ribbon (M6, Tfa1(L134E/V137E/L140E); M7, Tfa1(L134A/V137A/L140A). RNA products were visualised by primer extension and the mean intensity and standard deviation (s.d.) from triplicate experiments are provided, relative (rel.) to the activity of wild-type TFIIE. An asterisk marks RNA products resulting from an alternative upstream transcription start site.

the DNA backbone at positions –13/–14 upstream of the transcription start site (TSS, position +1) (Figures 5c, 8a). The E-ribbon binds between the clamp, the Rpb7 oligonucleotide binding (OB) domain, and the B-ribbon (Figure 8a). The E-linker (He et al., 2013; Murakami et al., 2015) and the mobile C-terminal domain of Tfa1 contact TFIIH (Maxon et al., 1994; Okuda et al., 2008), which may alter TFIIE conformation. The small TFIIE subunit Tfa2 (human TFIIEβ) contains two WH domains ('WH1' (Okuda et al., 2000), 'WH2'), and two conserved α-helices (called here 'E-tether') that bind the E-linker (Figure 8b). Consistent with the structure, the E-tether is essential for TFIIE function (Grünberg et al., 2012; Okamoto et al., 1998) and subunit dimerisation (Supp. Figure 23a, references (Grünberg et al., 2012; Blombach et al., 2015)). The structure further indicates that TFIIE must be displaced or at least moved before the elongation factor Spt4/5 (human DSIF) can bind to polymerase (Grohmann et al., 2011; Martinez-Rucobo et al., 2011).

7.2.3 TFIIE and TFIIF position upstream DNA

The OC structure thus reveals how TFIIF and TFIIE bind open promoter DNA from opposite sides of the Pol II cleft. First, TFIIF adopts an extended induced structure that allows it to retain the upstream DNA–TBP–TFIIB assembly on the wall and to bind the DNA bubble and downstream duplex in the cleft (Figures 5b, 6). Second, the TFIIF Tfg2 WH domain and the TFIIE Tfa2 WH1 domain contact each other above upstream DNA to encircle and retain DNA. Third, the eWH domain of TFIIE binds DNA in the region of initial DNA opening and its long $\beta1-\beta2$ hairpin (Meinhart et al., 2003) (called here the 'E-wing') projects to the upstream edge of the bubble (Figure 8a), suggesting that the eWH domain stabilises open DNA. Taken together, the highly modular and flexible basal factors TFIIF and TFIIE undergo substantial induced folding transitions to engage in multiple protein-DNA and protein-protein interactions to stabilise the OC.

7.3 DNA opening and loading

7.3.1 Closed complex opens spontaneously in the absence of TFIIH

Modelling of a closed DNA promoter onto the OC structure shows that closed DNA would clash with the TFIIE eWH domain. This suggests that the eWH domain adopts a different position before DNA opening. To investigate this, and to provide insights into the transition from the CC to the OC, we repeated structure determination with closed

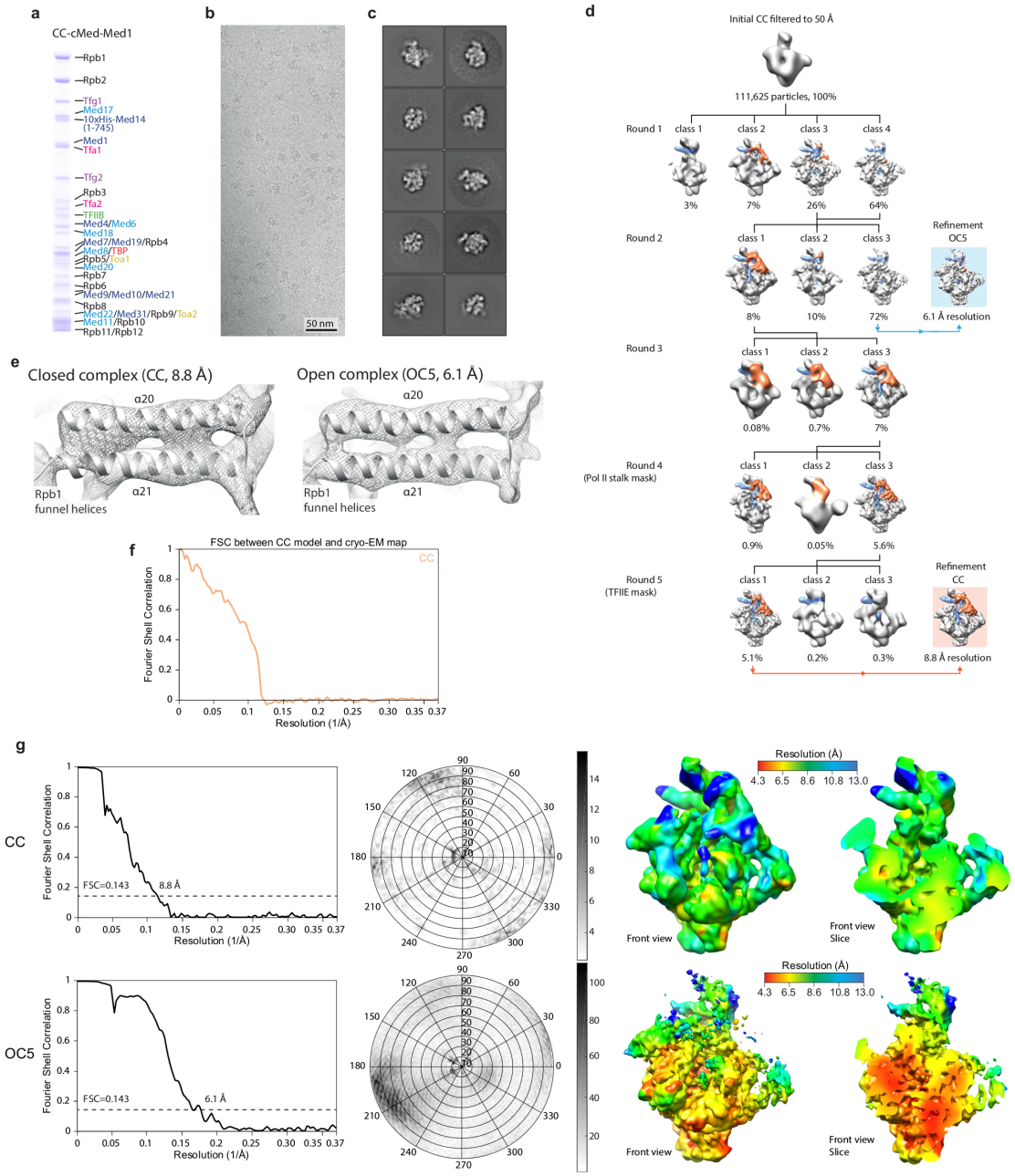

Figure 9: Cryo-EM analysis of the closed complex. **a**, SDS-PAGE analysis of CC-cMed-Med1 complex after size-exclusion chromatography. **b**, Cryo-EM micrograph. Scale bar, 50 nm. **c**, Ten representative reference-free 2D class averages. **d**, 3D classification of the CC cryo-EM data set. Mobile regions in the reconstructions are highlighted: promoter DNA (blue), TFIIE (except E-ribbon), and Tfg2 WH (orange). **e**, Detailed view of the Pol II funnel helices in the CC (left) and OC5 (right) densities. **f**, Gold-standard FSC between the CC coordinate model and the cryo-EM map. **g**, Gold-standard FSC (left) of the CC cryo-EM single particle reconstruction (FSC = 0.143). Orientation distribution plot of all particles contributng to the CC reconstruction (middle). The CC cryo-EM map is shown coloured by local resolution. Below the same for OC5.

DNA instead of pre-opened DNA (Section 6.3.2, Figure 9a-c). Surprisingly, cryo-EM analysis revealed that about 3 out of 4 particles contained open DNA, although closed DNA was used for complex preparation. DNA opening occurred in the absence of TFIIH (Figures 9d, g, 11d, e). From these particles we obtained an independent reconstruction of the spontaneously formed OC at 6.1 Å resolution (OC5, Figure 11d). Weak density for upstream and downstream DNA segments indicates that DNA bubbles of various sizes formed during DNA opening (Figure 11e). The reconstruction resembles the high-resolution OC structure, suggesting that the latter was not perturbed by the use of pre-opened DNA (Figure 11d).

7.3.2 Movements during the transition from closed to open complex

From the remaining particle images we obtained a reconstruction of the CC at 8.8 Å resolution (Figure 10a). Comparison of this CC reconstruction with the OC structure reveals movements mainly in TFIIE (Figures 10b, 11a-c). In the CC, the E-wing lies on

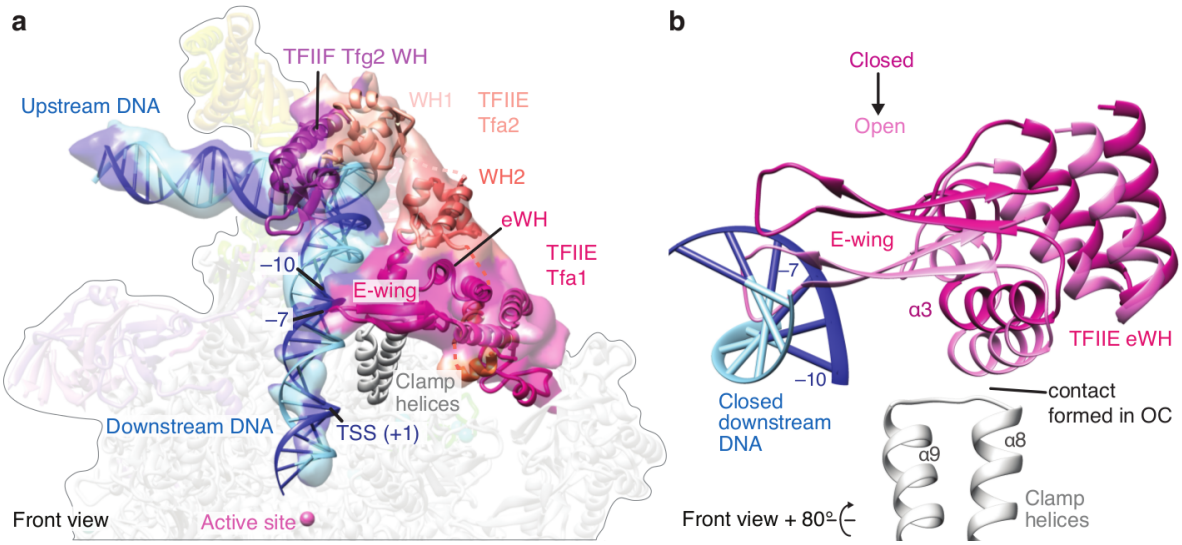

Figure 10: Closed complex cryo-EM structure. **a**, Details of the closed complex viewed from the front (Cramer et al., 2001). Highlighted are TFIIE, TFIIF Tfg2 WH, and DNA, superimposed on their density. The promoter DNA displays increased flexibility downstream of the E-wing contact at position -7 upstream of the TSS (+1). **b** Different position of the TFIIE eWH in closed (dark magenta) and open (light magenta) complexes, viewed from the top (Cramer et al., 2001).

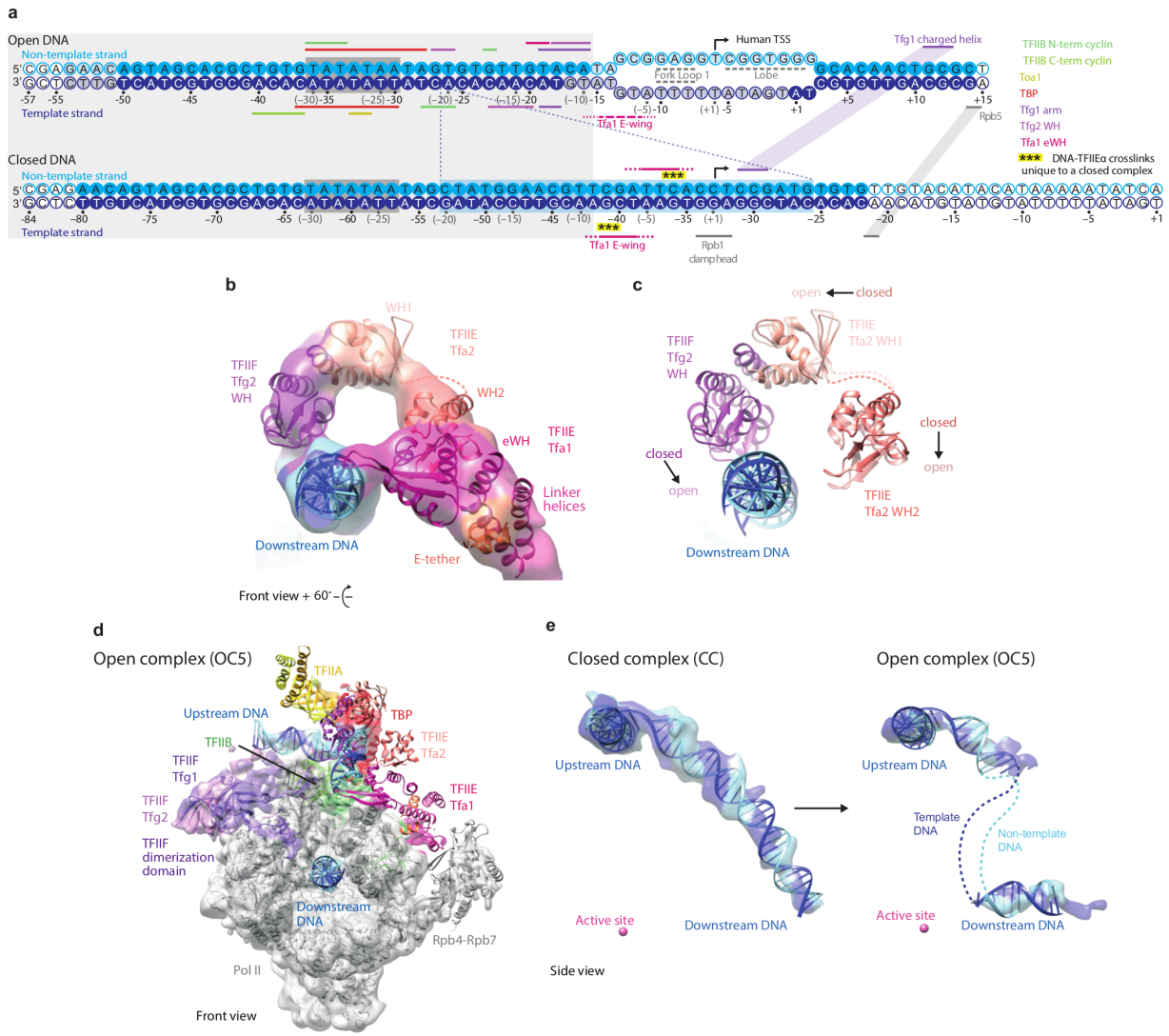

Figure 11: Transition from closed to open complex. **a**, Promoter sequences and difference in protein-DNA interactions are shown for the two distinct nucleic acid scaffolds used for preparation of closed and open complexes (compare Figure 5c). The *HIS4*-promoter sequence absent in the modified OC nucleic acid scaffold (grey box) is indicated. Protein-DNA interactions in the region covered by the light grey box are unchanged between CC and OC, and shown only for the OC for clarity. Unique and altered interactions are shown for each complex. DNA-TFIIEα photo-crosslinks, indicated by black asterisks, were observed in a closed but not open promoter state (Kim et al., 2000) and are consistent with the CC model. **b**, Fit of TFIIE, Tfg2 WH and downstream DNA into CC density. Although the overall fit reflects density well, the eWH domain and its E-wing may be rotated further away from promoter DNA. **c**, Details on the location of downstream DNA, Tfg2 WH, and Tfa2 WH1 and WH2 in the closed (dark colours) and open (light colours) complexes in the same view as in **b**. **d**, Cryo-EM density of OC5 and the OC ribbon model are shown in a front view (Cramer et al., 2001). The OC5 map shows weak density in regions of upstream assembly, TFIIE, and DNA that may be caused by increased flexibility owing to the heterogeneous population of spontaneously opened DNA. Colours as in Figure 5. **e**, Fit of promoter DNA to cryo-EM densities of CC and OC5.

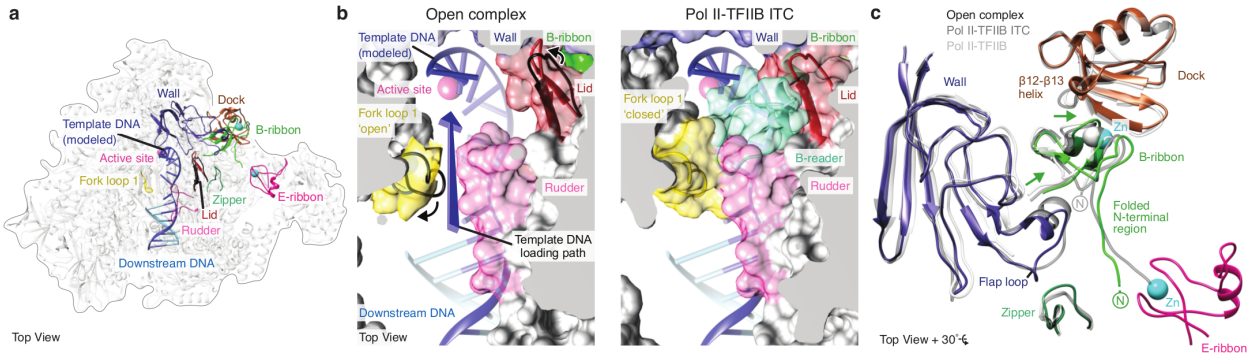

Figure 12: Cleft clearance and DNA template loading. **a**, OC structure viewed from the top (Cramer et al., 2001). Highlighted are the Pol II active site, fork loop 1, lid, rudder, wall, dock, zipper, TFIIB B-ribbon, TFIIE E-ribbon and downstream DNA (template, dark blue; non-template, cyan). The template single-strand was modelled using the Pol II-TFIIB ITC crystal structure (Sainsbury et al., 2013). **b**, Fork loop 1 and lid assume new positions in the OC compared to the ITC (Sainsbury et al., 2013) and this opens a path (arrow) for loading of the template DNA strand (blue) into the active site. Surface representations of Pol II cleft, and cleft elements fork loop 1, lid, and rudder in the OC (left), and in a Pol II-TFIIB ITC (PDB: 4BBS, right) (Sainsbury et al., 2013). Movement of the Pol II lid (left, black to dark red) leads to a steric clash with the B-reader. **c**, Allosteric binding of the TFIIE E-ribbon may lead to an altered position of the TFIIB B-ribbon. Movements in Pol II wall and flap loop, dock, zipper, and B-ribbon are observed in presence of TFIIE compared to the crystal structures of the binary Pol II-TFIIB complex (dark grey, PDB: 3K1F) (Kostrewa et al., 2009) and Pol II-TFIIB ITC (light grey, PDB: 4BBS) (Sainsbury et al., 2013). The altered B-ribbon position may be stabilised by binding to a short helix formed in loop β12-β13 of the dock domain.

top of the DNA around position -7, in the region where DNA opening begins (Maxon et al., 1994) (Figure 11a). Consistent with this contact, Tfa1 (or the human counterpart TFIIEα) cross-links to DNA near this point in the CC (Forget et al., 2004; Kim et al., 2000; Miller and Hahn, 2006) (Figure 11a). Conversion of the CC to the OC involves movement of upstream DNA and the DNA-associated domains Tfa2 WH1 and Tfg2 WH towards the cleft. DNA opening allows the TFIIE eWH domain to bind the tip of the clamp, and enables the E-wing to move near the upstream edge of the DNA bubble (Figures 10b, 11c). Consistent with these changes, cross-links between upstream DNA and the large TFIIE subunit are altered when the CC is converted to the OC (Kim et al., 2000). These observations indicate that DNA opening involves TFIIE, and in particular the eWH domain.

7.3.3 TFIIE may be involved in DNA loading

TFIIE may additionally help in loading of the DNA template strand into the active centre during the transition from the CC to the OC. In previous structures, the path for loading the template was obstructed by the TFIIB B-reader and the Pol II fork loop 1 and lid. However, the B-reader is mobile in the OC structure, and fork loop 1 and the lid are moved to provide a path for template-strand loading (Figure 12b, Supp. Figure 24a-c). These movements are apparently triggered by TFIIE because binding of the E-ribbon leads to a shift in the B-ribbon that partially withdraws the B-reader from the cleft (Figure 12c). Thus allosteric binding of the E-ribbon apparently induces 'clearance' of the Pol II cleft that may facilitate template-strand loading into the active centre, and transcription start site scanning in yeast.

7.3.4 Functional data underlines role of TFIIE eWH and E-ribbon

To support the proposed functions of structural elements in TFIIE, we prepared recombinant TFIIE variants and tested them for binding to the CC and for promoter-dependent transcription activity in yeast nuclear extract (Section 6.3.3, Figures 8c, 13a, c). Mutation of only three surface residues in the E-ribbon that contact the Pol II subunit Rpb7 strongly impair both binding to the CC and transcription activity, and lead to a severe growth defect in yeast (Figures 8c, 13a, b). Further, the eWH domain is required for TFIIE function (Grünberg et al., 2012), and disruption of the eWH contact with DNA by introducing negatively charged glutamate residues in the eWH helix α3 leads to a transcription defect and impairs binding to the CC (Figures 8c, 13a). Other mutations in the eWH domain did not show functional defects in our assays. Deletion of the E-wing results in a mild growth phenotype (Figure 13b), and does not impair *in vitro* transcription (Figure 8c), maybe because TFIIH compensates for the loss of E-wing function in these assays.

7.4 Model of transcription initiation

Structural comparisons of the yeast CC, OC, and ITC (Plaschka et al., 2015) with each other and with the highly conserved human CC (He et al., 2013) reveal differences in the positions of DNA, the clamp, and TFIIE, and lead to an extended model of DNA opening (Figures 14, 15). In this model, promoter DNA is initially bent away from the active site by $\sim 20°$ near position -10 at the tip of the clamp helices (yeast CC). The

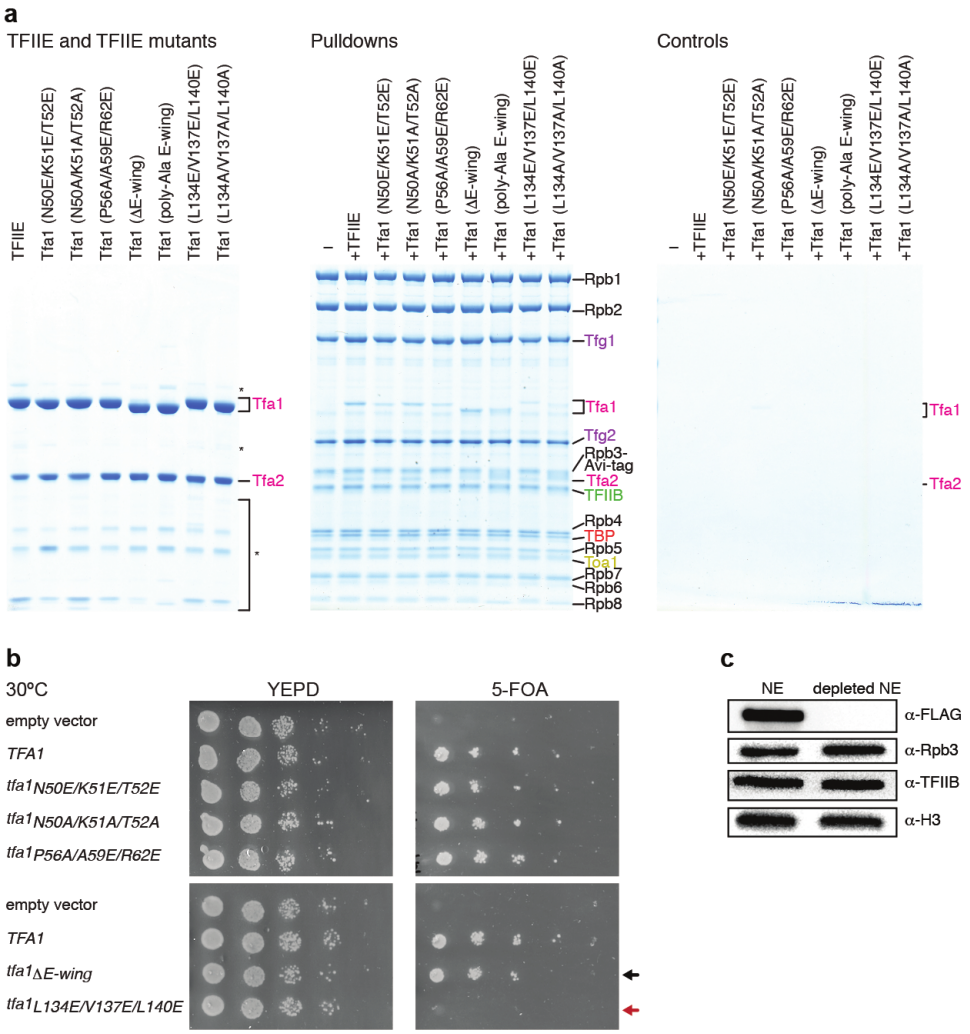

Figure 13: Structure-function analysis of TFIIE. a, Pulldown assays (analysed by SDS-PAGE) with recombinant TFIIE variants carrying mutations at the TFIIE-CC interface revealed that the E-ribbon is essential for TFIIE recruitment. To confirm the integrity of the purified TFIIE variants, 2 μg were analysed (left). Some minor contaminant and degradation bands of TFIIE are indicated by an asterisk. The bead elution from the pulldown assay is shown (middle), providing negative (no TFIIE) and positive control (TFIIE) controls in the two leftmost lanes. The binding of all TFIIE variants to the CC was impaired compared to the wild-type protein, with the exception of the Tfa1(Δ E-wing) mutant, suggesting that all other interfaces contribute to TFIIE binding affinity. The bead-only control (right) indicated that TFIIE and TFIIE variants did not show unspecific binding to the beads. **b**, Yeast complementation assays were performed in triplicate experiments with wild-type *TFA1*, an empty vector, and *TFA1* variants with mutations as indicated. A mild growth defect was observed for the Tfa1(Δ E-wing) mutant (indicated by a black arrow) and a strong defect for the Tfa1(L134E/V137E/L140E) mutant (red arrow). **c**, Western blot analysis of the 3xFlag-tagged Tfa1 confirms specific immunedepletion of Tfa1 in the nuclear extract (NE), whereas levels of Pol II (Rpb3), TFIIB, and Histone H3 were unaffected.

clamp then opens slightly (Figure 15b), allowing promoter DNA to bend in the opposite direction and to enter the upper part of the cleft ('human CC'). This frees the site at the tip of the clamp helices that can now bind the TFIIE eWH domain. DNA can then no longer swing back to its initial position, because this would result in a steric clash with the repositioned eWH domain. However, when DNA opening occurs around position -10, the upstream DNA can swing back to its original position and this stabilises the DNA duplex single-strand junction. TFIIH then rotates downstream DNA and pushes the template single strand into the cleft using ATP-dependent translocase activity, as suggested (Grünberg et al., 2012). When the bubble extends downstream (Giardina and Lis, 1993), the template strand is loaded into the active centre cleft via a path cleared by allosteric binding of TFIIE (yeast OC). Template-strand loading allows the clamp to close again and to trap downstream DNA in the cleft (Figure 15b). The B-reader then

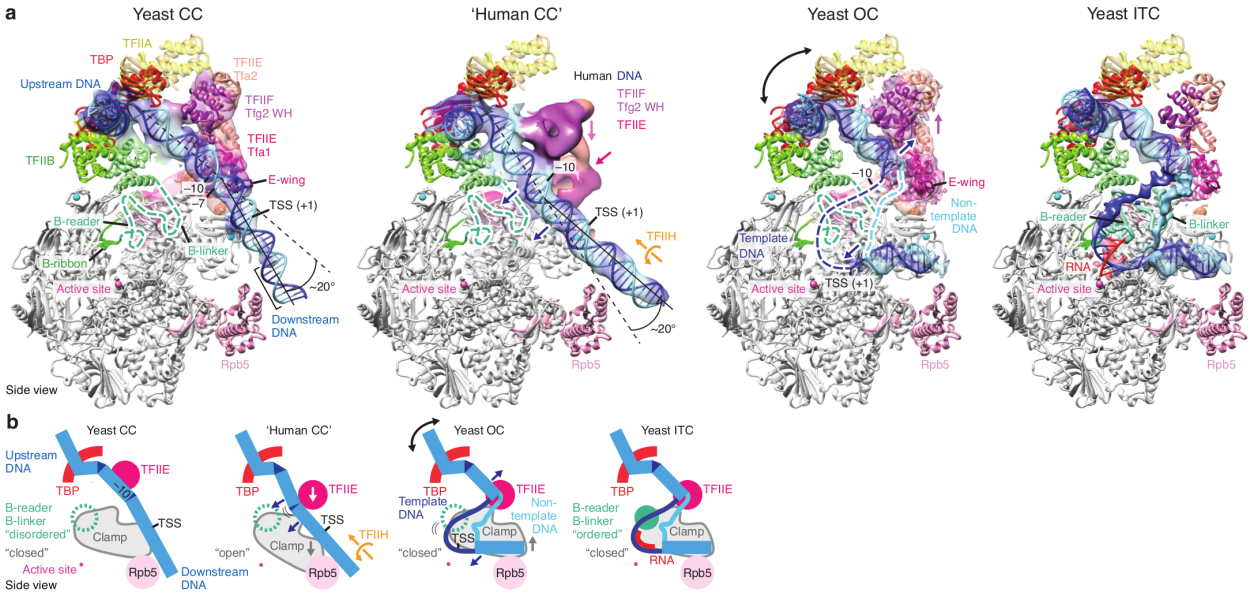

Figure 14: Model for DNA opening during transcription initiation. **a**, Gallery of initiation complexes depicting proposed movements (arrows) of DNA and basal factors during the transition from the CC to OC to ITC, from left to right, viewed from the side (Cramer et al., 2001). Yeast CC and OC structures (this work) were complemented with our previous yeast ITC structure (EMD-2785) (Plaschka et al., 2015) and an alternative model of the CC ('human CC'), which was obtained by replacing the DNA with that in the human CC (EMD-2306) (He et al., 2013), and adjusting the clamp to the position observed in the human CC. Shown are cryo-EM densities for DNA, Tfg2 WH, and TFIIE. DNA positions -10, -7 (yeast CC) and +1 (TSS) are labelled. DNA was extended by one turn for the yeast CC (black bracket). The locations of TFIIA and TFIIE in the ITC were inferred from the yeast OC. Obstructing Pol II and TFIIF regions were removed for clarity. **b**, Schematic representation of **a**. Key elements for DNA opening are indicated.

covers the template strand, and helps to detect the transcription start site (Kostrewa et al., 2009), triggering RNA synthesis (yeast ITC).

This model explains how DNA opening can be achieved with the use of binding energy alone, at least at some promoters (Buratowski et al., 1991; Holstege et al., 1995). DNA opening allows for new protein interactions at the upstream DNA duplex single-strand junction involving the eWH domain, the TFIIF arm, and the Pol II clamp. DNA loading into the cleft also enables new interactions of the downstream DNA duplex with the TFIIF charged helix and the Pol II cleft and clamp. These additional contacts may

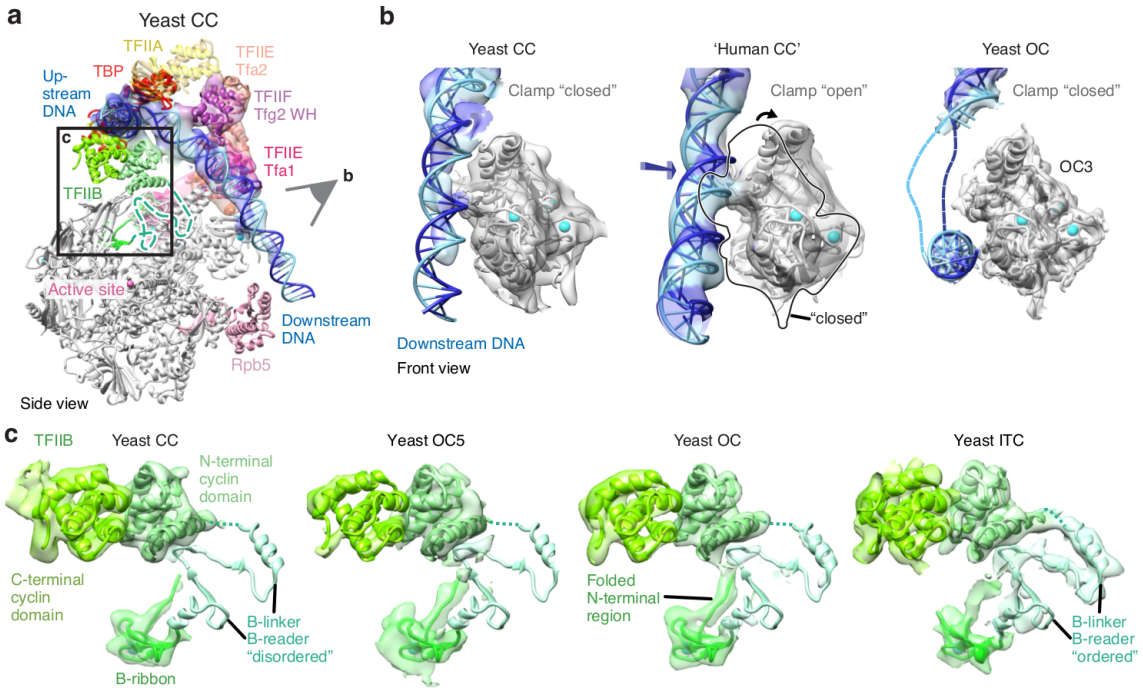

Figure 15: Pol II clamp positions and TFIIB B-reader mobility during DNA opening. **a**, The yeast CC is shown from a side view (Cramer et al., 2001), indicating the path of DNA and location of TFIIE. The eye symbol (grey) indicates the point of view in **b**. **b**, The Pol II clamp may undergo transitions during DNA opening as indicated. The OC model of the Pol II clamp is shown superimposed on yeast CC (this study), and yeast OC (this study). The OC model Pol II clamp was rigid-body fitted to the human CC cryo-EM density (EMD-2306) (He et al., 2013) and is superimposed. The view is from the front (Cramer et al., 2001). **c**, The TFIIB B-reader element shows strong density only in the ITC state, suggesting that its mobility in earlier states may be important for maintaining a cleared path for template DNA loading into the Pol II cleft. Ordering of the B-reader may further lead to stabilisation of the upstream promoter assembly that is flexible in the OC. Cryo-EM densities for yeast CC (this work), OC5 (this work), OC (this work), and ITC (EMD-2785) complexes are superimposed on the TFIIB model (PDB: 4BBS for the B-linker and B-reader). As secondary structure elements could not be resolved in the human CC (He et al., 2013), we excluded this cryo-EM density from comparison.

compensate for the energy needed for DNA melting and help to trap open DNA and to prevent its re-closure.

A similar mechanism for ATP-independent DNA opening may be used in other transcription systems. Proteins with homologies to TFIIB, TFIIE, and TFIIF, but not TFIIH, are present in the Pol I and Pol III systems (Vannini and Cramer, 2012), and counterparts of TBP, TFIIB, and TFIIE are found in archaea (Blombach et al., 2015). The bacterial transcription initiation system is structurally unrelated, but conceptually similar. DNA opening occurs spontaneously above the cleft, the open DNA is trapped with the use of binding energy, and this requires clamp opening and closure (Chakraborty et al., 2012; Feklistov and Darst, 2011; Zhang et al., 2012; Zuo and Steitz, 2015; Bae et al., 2015). DNA opening at a subset of bacterial genes however requires the ATP-dependent initiation factor σ^{54}, and this enables further regulation (Yang et al., 2015). Similarly, Pol II regulation can occur at the level of DNA opening (Kouzine et al., 2013), and TFIIH is required to keep DNA open at least at selected promoters (Holstege et al., 1996). This suggests that the Pol II initiation system evolved to depend on the ATP-consuming factor TFIIH for DNA opening, probably in response to an expanded need for gene regulation.

Part IV
Conclusions and Outlook

8 Conclusions

Transcription initiation is a key step in regulating gene expression. To understand the molecular mechanism of this process detailed structural information of various states is required. In this study, we present three-dimensional structural information of a closed and a spontaneously open transcription initiation complex comprising Pol II, TBP, TFIIA, -IIB, -IIE, -IIF and *HIS4* promoter DNA. Models for both complexes were generated by rigid-body fitting of the previously determined high-resolution structure of the open complex containing a mismatch DNA bubble. This high-resolution structure reveals for the first time detailed interactions of the general transcription factors with each other, with Pol II and promoter DNA. The transcription factors TFIIB, TFIIE and TFIIF are highly modular with different structured domains being connected by flexible linkers, which undergo induced folding upon assembly.

Comparison of closed and open initiation complexes shows movements mainly in TFIIE, indicating a central role for TFIIE in promoter opening or stabilisation of open DNA. In functional assays we show that recruitment of TFIIE to the initiation complex is crucial for transcription. However, the involvement of the E-wing in promoter opening did not become apparent in the applied assays, possibly because of the experimental setup, that did not allow for monitoring promoter opening in the absence of TFIIH.

A central result of this work is that promoter opening at the *HIS4* promoter can occur in the absence of TFIIH. This observation suggests a redefined and more differentiated view on promoter opening in the Pol II system and needs further investigation.

Comparison with the recent human model (He et al., 2016) illustrates the high conservation between the two species regarding the architecture of the initiation complex (Figure 16) and underlines the importance and relevance of yeast as a model system.

9 Towards a high-resolution closed complex structure

The recent technical developments in cryo-EM and data processing software allow not only visualisation of large molecular complexes at high resolution but also the study of

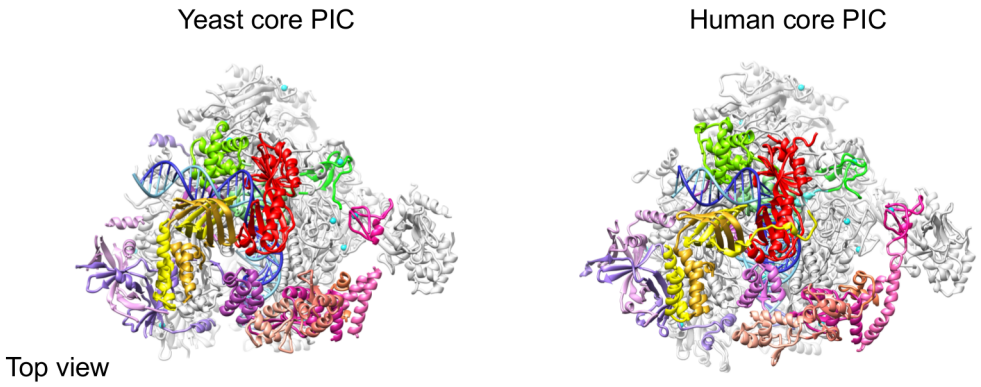

Figure 16: Conservation of the yeast and human transcription initiation complex. Shown are the yeast (PDB 5FYW) (Plaschka, Hantsche et al., 2016) and the human (PDB 5IYB) (He et al., 2016) open complexes in a top view (Cramer et al., 2001). Colour scheme as in Figure 5.

their dynamic conformational properties. Different complex conformations that exist in solution can be trapped upon freezing and resolved by 2D and 3D classification strategies. Often they provide important biological insight of the rearrangements and movements within a protein complex. Analysis of the closed complex data showed that only 8 % of the particles contained a closed double-stranded DNA and 72 % contained open DNA. It was possible to differentiate between the two states in two rounds of classification (Figure 9). However, conformational analysis with cryo-EM is at its limit, when positional variation can not be represented by classification. This is seen in the spontaneously open state, where DNA opening is not uniform and a large population of different bubble sizes exists, which finally decreases local resolution. One approach to overcome this problem is to increase the number of particles used for classification, yet in the case of continuous and merging states this may not be sufficient. Another strategy is to foster or to lock a certain complex conformation by the addition of inhibitors or binding partners as it was done for the open complex by using a DNA scaffold with a mismatch bubble. At a nominal resolution of 8.8 Å, the interpretation of the received electron density for the closed complex was only possible by rigid-body fitting of the open complex model (see Section 6.3.2). A future high-resolution structure would provide more detailed information on TFIIE conformation and its interaction with promoter DNA. This would enable a more precise comparison with the open promoter state and thus a better understanding of the transition from closed to open state.

In order to obtain a high resolution structure of the closed complex, the ratio between closed and spontaneously open state has to be shifted. A first step would be to use a

different native promoter DNA sequence. Inspection of the *HIS4* promoter sequence shows that the region downstream of the TATA-box is very AT-rich (Figure 11a) decreasing the energy required to open the promoter. A GC-rich sequence, on the other side, is not so readily opened as shown in *in vitro* transcription experiments (Holstege et al., 1995). Indeed, the super core promoter (SCP) (Juven-Gershon et al., 2006) used in cryo-EM studies of the human initiation complex (He et al., 2013, 2016) is GC-rich and no spontaneous DNA opening was observed. Interestingly, in a cryo-EM study of the closed *S. cerevisiae* initiation complex using the *HIS4* promoter no DNA opening was reported (Murakami et al., 2015), either. Murakami et al. describe variability of structure and location of the G-lobe (part of the complex containing the GTFs and DNA), though. A possible reason for this discrepancy may be the presence of TFIIH having a repressing effect on promoter opening. Since TFIIH binds to downstream DNA it may stabilise and retain its position.

Core Mediator was present in the sample preparation of the closed complex but it showed very low occupancy under cryo-EM conditions. SDS-PAGE analysis of the isolated complex illustrates a stoichiometric complex (Figure 9). But native PAGE analysis after cross-linking shows that several subcomplexes were cross-linked (Appendix Figure 27) which could not be isolated by size exclusion chromatography in the final step. It is therefore likely that in the sample, even before cryo-grid preparation, only few complexes were present with Mediator bound. To increase Mediator occupancy the protocol for complex formation needs to be optimised. An alternative to size exclusion chromatography is gradient centrifugation which can be also coupled with cross-linking (Kastner et al., 2008).

10 Extended investigations on the role of TFIIH in promoter opening

The result that the *HIS4* promoter can be opened by the initiation complex in the absence of TFIIH was unexpected. In the canonical model of transcription initiation the TFIIH subunit Rad25 (human XPB) opens promoter DNA in an ATP-depending step (Conaway and Conaway, 1993; Grünberg and Hahn, 2013; Sainsbury et al., 2015). But TFIIH was not present in the sample preparation. Our observation suggests that a more differentiated view on promoter opening is required.

In the 1990s several *in vitro* experiments indicated that TFIIH is required for promoter

opening and that it is an ATP-dependent process (Parvin and Sharp, 1993; Holstege et al., 1996; Wang et al., 1991; Jiang and Gralla, 1995). Later *in vivo* studies came to similar results (Guzmán and Lis, 1999; Ostapenko and Gileadi, 2000). However, only very few promoters were studied, like the adenovirus major late promoter (MLP) or the adenovirus E4 promoter. Both promoters are GC-rich in the region around of the transcription start site. Holstege et al. already showed in 1995 that the sequence in this region determines if TFIIE is sufficient for promoter opening (Holstege et al., 1995). Moreover, not all proteins used for *in vitro* assays were recombinantly expressed and purified (Parvin and Sharp, 1993; Holstege et al., 1995, 1996) so that cross-contamination could not be excluded.

In the future, the dependency of TFIIH for DNA opening needs to be re-investigated and studied on a global scale. To analyse how many promoters in the yeast genome depend on the presence of TFIIH for transcription initiation, the nascent mRNA in a Rad25-depleted nucleus (anchor-away strain) could be monitored by 4tU-Seq (Schulz et al., 2013). With this approach it is not possible to differentiate between promoter opening and transcription initiation since the readout is a RNA product. This is an important issue since TFIIH is required at a later point of transcription initiation: Kin28 phosphorylates the CTD of Rpb1 and thus triggers promoter escape. For decoupling promoter opening and transcription, *in vitro* DNA opening assays with purified factors need to be established. For this approach it is also important to assay different kinds of promoter sequences.

Pol I and Pol III can open DNA alone without the help of an ATP-dependent factor. Since all three polymerase structures show a conserved polymerase core it can be speculated if the Pol II system is really so different. Maybe promoter opening and closing is a transient and dynamic process and the transcription initiation factors including TFIIH are required to shift the equilibrium to the open state by stabilising this conformation. This hypothesis is supported by single molecule fluorescence resonance energy transfer (sm-FRET) studies of *S. cerevisiae* mitochondrial RNA polymerase Rpo41 and its transcription factor Mtf1 (Kim et al., 2012) showing that Rpo41 alone can open promoter DNA, but only transiently and that Mtf1 and ATP enhance the opening transition and prolong its lifetime. For Pol II the need to regulate and suppress spontaneous promoter melting is much higher than for Pol I and Pol III because it transcribes thousands of different and specific genes at different time points. Promoter escape may also be a suitable step for a tight regulation of transcription initiation. Here the Pol II CTD, which is a unique Pol II element, plays a central role as well as TFIIH.

11 Towards a complete picture of transcription initiation

Determination of the atomic structure of *S. cerevisiae* Pol II around the millennium marked a breakthrough (Cramer et al., 2000, 2001; Gnatt et al., 2001). Since then the architecture larger subcomplexes of the initiation machinery were solved (Figure 17). The molecular mechanisms of transcription initiation and elongation are emerging but many questions remain. What are the structures of the multisubunit factors TFIID and TFIIH? Mid-resolution cryo-EM densities provide topological information (Murakami et al., 2015; Louder et al., 2016; He et al., 2016) (Figure 2) but high-resolution details are missing. How exactly do the TAFs contact promoter DNA and which rearrangements occur in TFIID during Pol II loading? How does TFIIE recruit TFIIH? So far structural

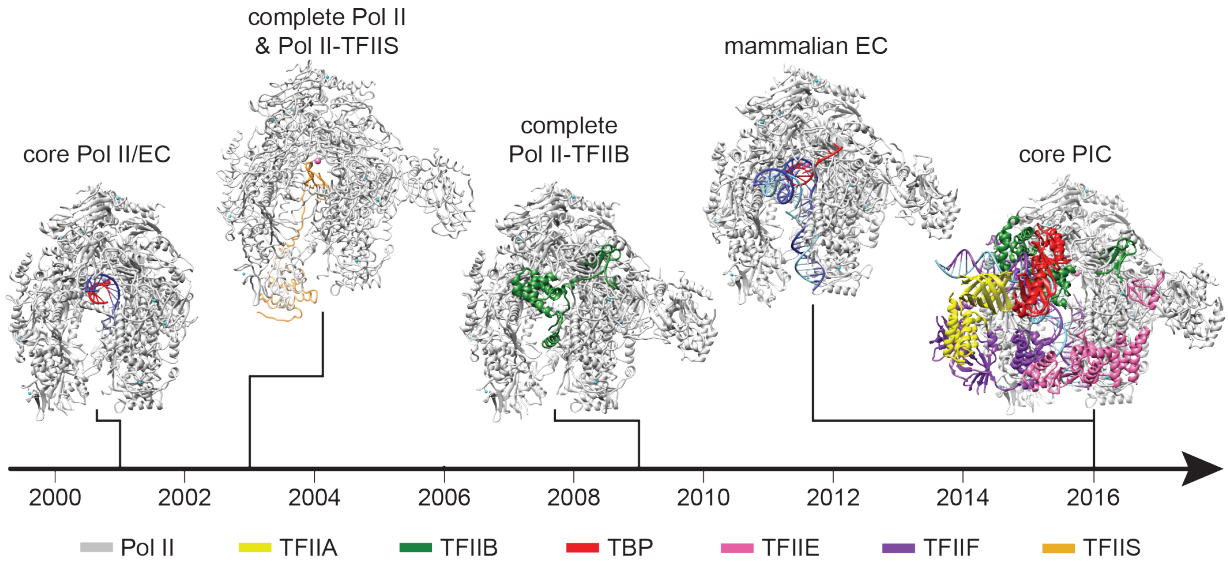

Figure 17: Selected atomic models of Pol II transcription complexes obtained by structural biology since the millennium. The following structures are shown in a previously defined top view (Cramer et al., 2001): core Pol II and core Pol II elongation complex (EC) (PDB 1I6H) (Gnatt et al., 2001), complete Pol II and Pol II-TFIIS complex (PDB 1PQV) (Kettenberger et al., 2003), complete Pol II-TFIIB complex (PDB 3K1F) (Kostrewa et al., 2009), mammalian Pol II EC (PDB 5FLM), and core pre-initiation complex (PIC) (PDB 5FYW) (Plaschka, Hantsche et al., 2016). The protein factor color scheme is shown at the bottom.

studies used endogenously purified TFIID and TFIIH. Recombinant expression methods, e.g. in insect cells, may result in more homogeneous and stoichiometric complexes and thus may increase resolution. Many flexible regions of TFIIE and TFIIF could not be traced in the cryo-EM study (Figure 5a). Extending the core initiation complex with additional binding partners like TFIIH may lead to folding or stabilisation of these

domains.

What are the mechanisms of transcription initiation and regulation in the context of chromatin? For transcription initiation of many eukaryotic genes the co-activator Mediator is necessary. How are signals from activators transmitted to the initiation complex? How does Mediator stimulate CTD-phosphorylation by TFIIH? Answers to these questions require high-resolution structural information on the complete Mediator, its interactions with the Pol II initiation complex, on one side, and with activator-bound DNA, on the other side.

To understand transcription initiation on a global level, structural, biochemical and biophysical studies need to be extended to TATA-less promoters which account for most of the promoters (Basehoar et al., 2004). Also, the structural characterization of the initiation complexes of Pol I and Pol III will substantially contribute to the understanding of eukaryotic transcription initiation and gene regulation.

Taken together, the structures presented in this work are an important step towards the molecular understanding of transcription initiation. Additional three-dimensional structures of complete initiation complexes at different stages of initiation together with functional studies will eventually provide the molecular basis for gene regulation in eukaryotes.

Part V

Supplemental data

12 Cryo-EM data statistics of the open complex

Data presented here and in Section 13 was generated by Clemens Plaschka and has been published:

C. Plaschka*, **M. Hantsche***, C. Dienemann, C. Burzinski, J. Plitzko, P. Cramer. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**(7603): 353-358. doi: 10.1038/nature17990

a **Cryo-EM data collection and refinement statistics of the OC structure**

	Pol II–TFIIB ribbon (OC1)	Upstream assembly* (OC2)	TFIIF dimerization (OC4)
Data collection			
Particles	102,876	17,282	29,455
Pixel Size (Å)	1.35	1.35	1.35
Defocus range (μm)	–0.7 to –4.2	–0.7 to –4.2	–0.7 to –4.2
Voltage (kV)	300	300	300
Electron dose (e^{-} Å^{-2})	33	33	33
Reconstruction (RELION)			
Accuracy of rotations (°)	1.58	1.11	0.96
Accuracy of translations (pixel)	0.71	0.59	0.50
Resolution (Å)	3.58	3.97	3.89
Map sharpening B-factor (Å^2)	111	95	79
Model composition			
Non-hydrogen atoms	27844	3488	1839
Protein residues	3506	444	221
Refinement (PHENIX)			
Map CC (around atoms)	0.768	0.631	0.539
Rms deviations			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	0.94	1.11	1.25
Validation			
Molprobity score	1.99	2.23	2.51
All-atom clashscore	12.43	18.87	31.30
Rotamer outliers	0.3%	0%	0%
C-beta deviations	0	0	0
Ramachandran plot			
Outliers	0.3%	0.5%	1.0%
Allowed	5.4%	6.8%	8.4%
Favoured	94.3%	92.7%	90.6%

*A model of Upstream assembly components was refined into the OC2 cryo-EM density: Pol II protrusion (residues 44-218 and 405-465), Tfg2 Linker (residues 249-280), TFIIB N-and C-terminal cyclin domains (residues 124-326), and the Tfg1 arm (residues 327-335).

b

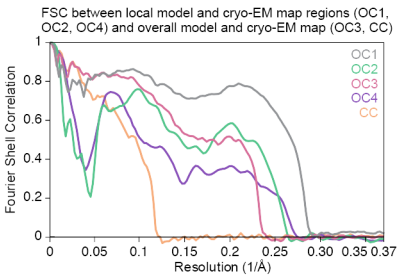

Figure 18: Data collection, refinement statistics, and structure validation of OC cryo-EM data. **a**, Cryo-EM data collection and refinement statistics of the OC structure. Different regions of the composite OC structure were refined into OC1, OC2 and OC4-focused maps to obtain an atomic model for 90 % of the structure. **b**, Gold-standard FSC between the respective coordinate models and local regions of the OC1, OC2, and OC4 cryo-EM maps used for model refinement and between overall OC and CC models compared to OC3 (best TFIIE density) and CC cryo-EM maps.

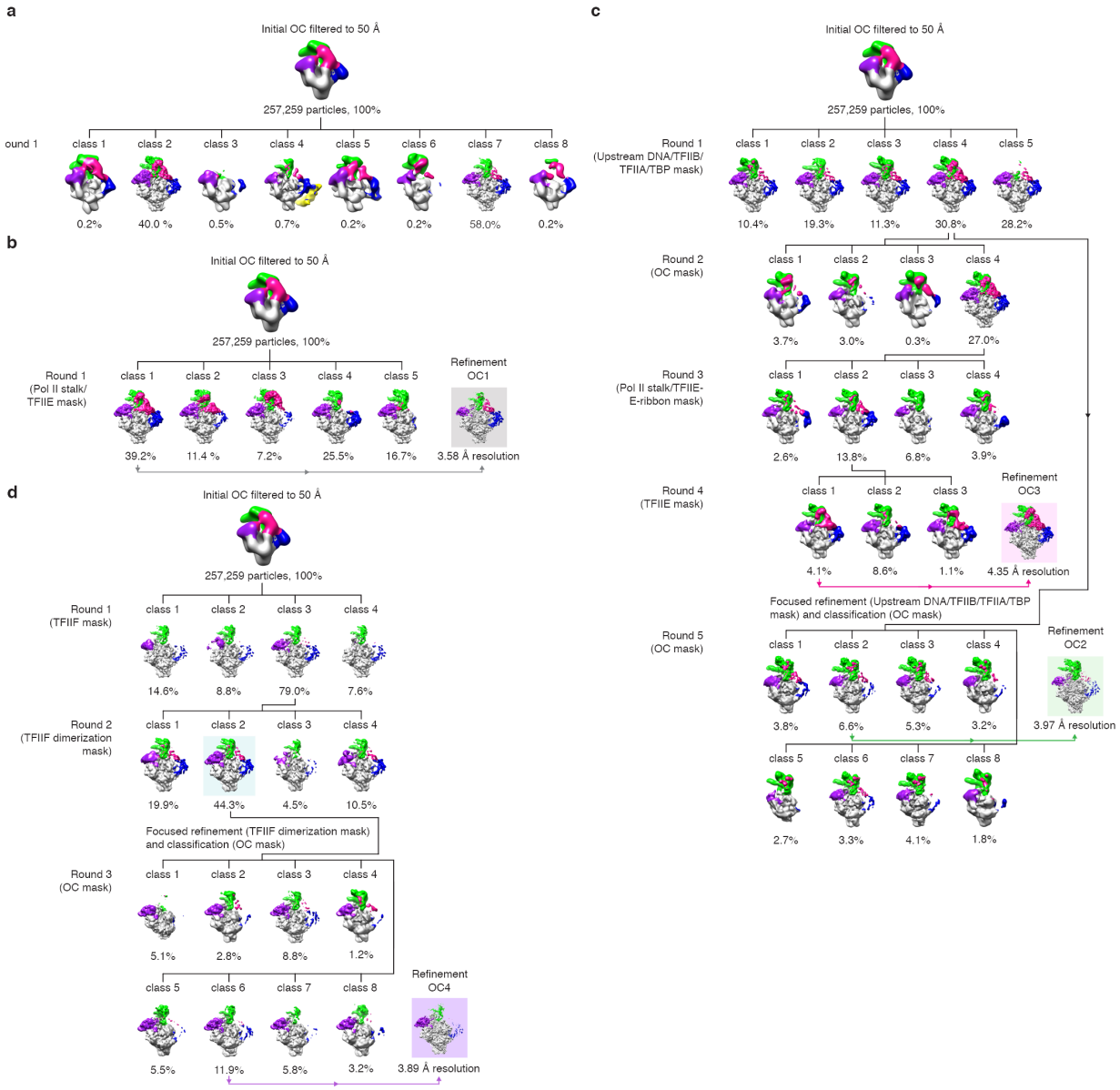

Figure 19: Three-dimensional classification of OC cryo-EM data. a, Classification of the cryo-EM data set using an initial OC reconstruction as the reference model, revealed heterogeneity. The percentage of single particles contributing to each class is provided. To help visualize structural differences, 3D reconstructions of the OC are coloured according to mobile regions: Pol II core, TFIIB B-ribbon (grey); upstream DNA, TFIIA, TBP, TFIIB cyclin domains, Tfg2 linker (green); TFIIF dimerization domain (purple); TFIIE except E-ribbon, Tfg2 WH (magenta); Pol II Rpb4-Rpb7 stalk and E-ribbon (blue); cMed-Med1 (yellow). **b**, Focused classification using a mask covering the Pol II stalk and E-ribbon. **c**, Improvement of densities for Tfg2 linker, TFIIB, and TFIIE, through rounds of focused 3D classification using various masks. **d**, Improvement of densities of TFIIF dimerization domain and the Tfg1 arm through three rounds of classification using various masks.

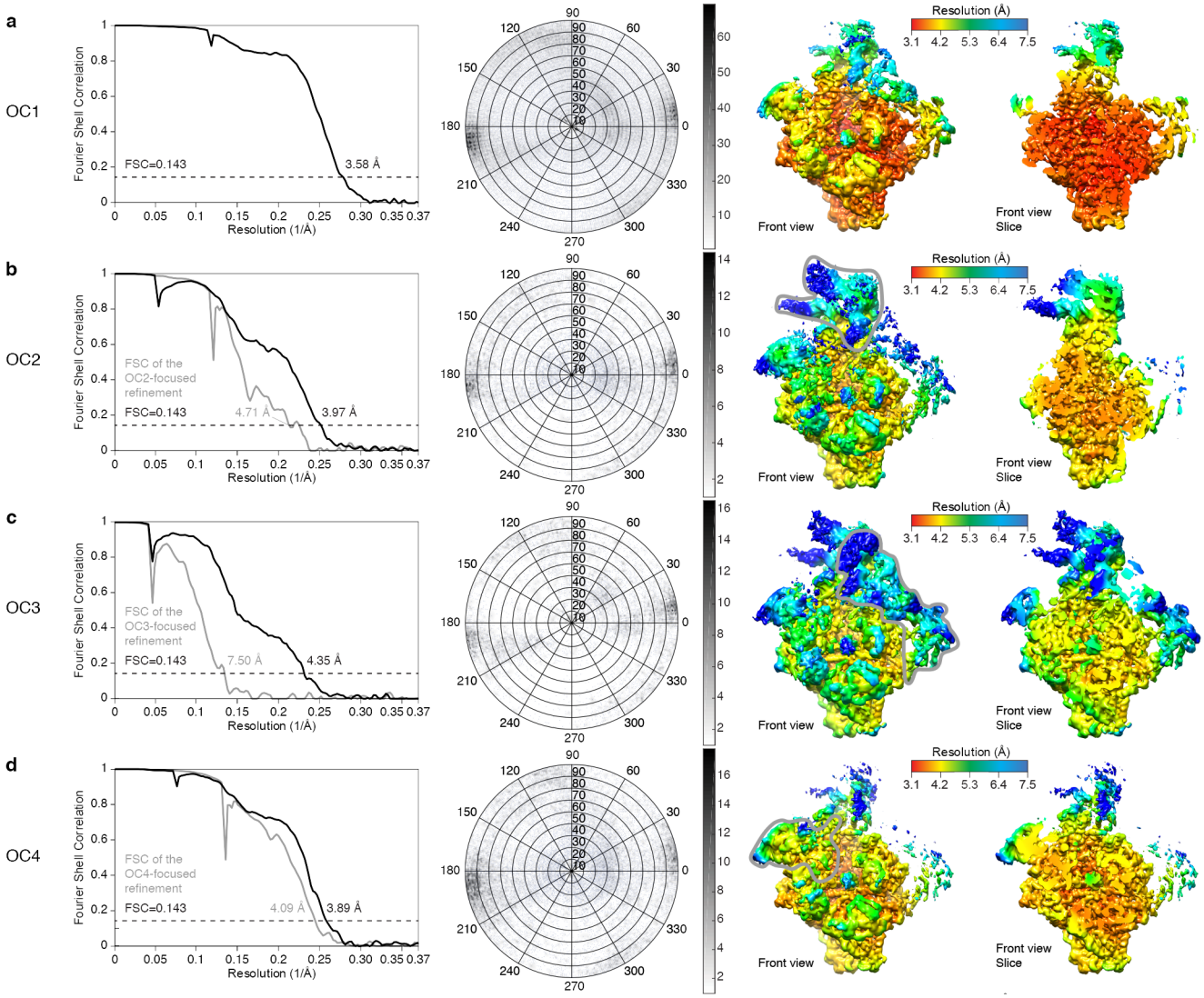

Figure 20: Resolution of OC cryo-EM reconstructions. a, Gold-standard FSC (left) of the OC1 cryo-EM single particle reconstruction (FSC = 0.143). Orientation distribution plot of all particles that contribute to the OC1 reconstruction (middle). The OC1 cryo-EM map is shown coloured by local resolution as described (Plaschka et al. 2015). **b**, As in **a** but for the OC2 reconstruction. The gold-standard FSC for the density obtained from focused refinement (OC2-focused) with a soft mask around the upstream DNA assembly is indicated in grey. The region masked for focused refinement is indicated with a grey outline on the cyro-EM map coloured by local resolution (right). **c**, As in **b** but for the OC3 and OC3 focus-refined reconstructions. **d**, As in **b** but for the OC4 and OC4 focus-refined reconstructions.

13 Structural details of the open complex

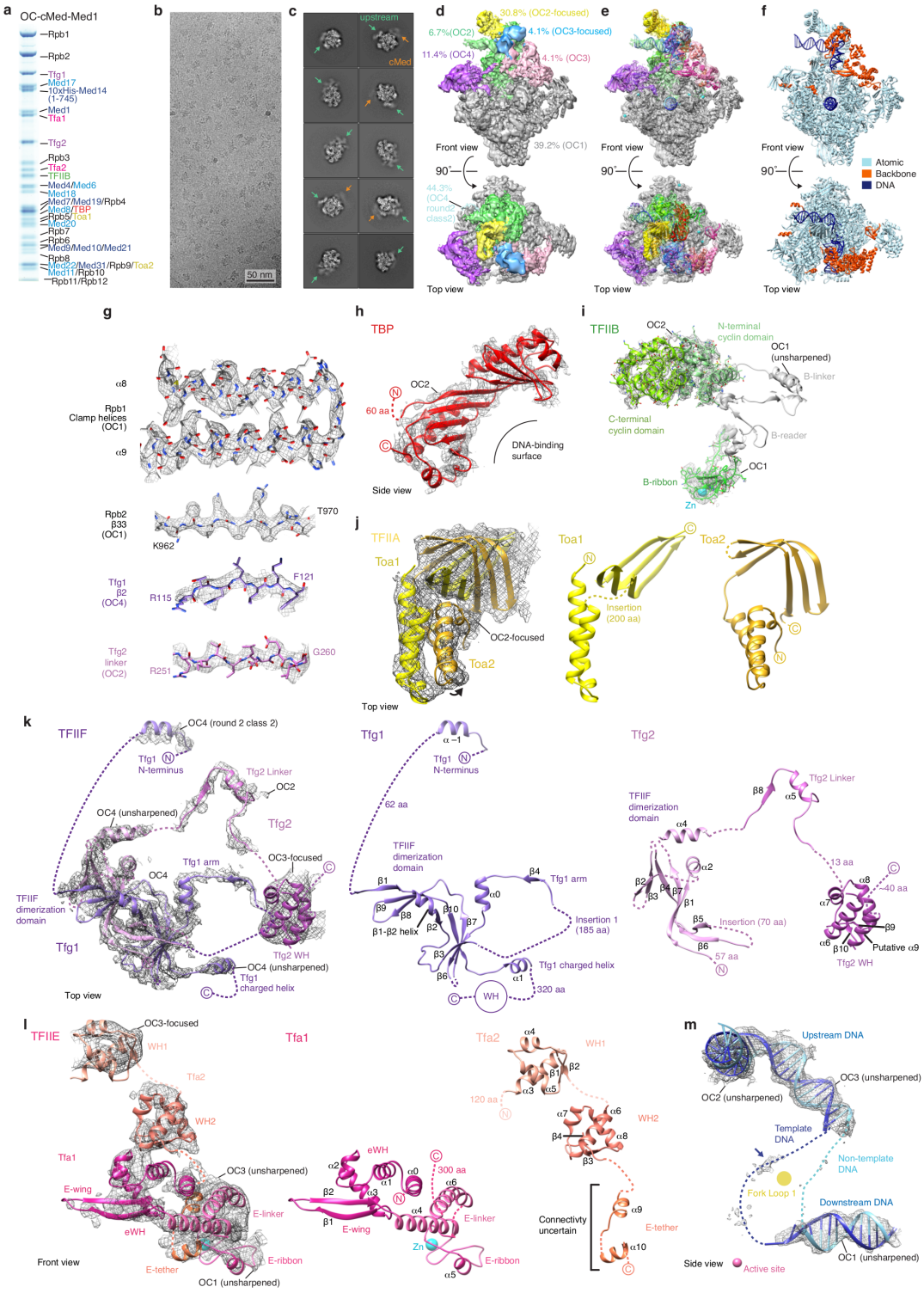

Figure 21: Modelling of open complex cryo-EM densities. See next page for caption.

Figure 21: Modelling of open complex cryo-EM densities. a, SDS-PAGE analysis of OC-cMed-Med1 complex after size exclusion chromatography. Protein colours as in Figure 5. Although core Mediator was required for stable association of TFIIE, it largely dissociated under cryo-EM conditions as observed previously (Plaschka et al., 2015). Some remaining core Mediator was flexible and located as described (Plaschka et al., 2015), but could not be included in further high-resolution analysis. **b**, Cryo-EM micrograph of the OC-cMed-Med1 complex. The scale bar is 50 nm. **c**, Ten representative reference-free 2D class averages of OC-cMed-Med1 reveal flexibility of the upstream DNA assembly including TFIIE (green arrow) and very weak density for core Mediator (orange arrow). Compare Supp. Figure 19a, c. **d**, Composite cryo-EM density of the OC shown in front and top views (Cramer et al., 2001). Colors indicated the parts of the six cryo-EM densities used for modeling of the open complex (OC1, grey; OC2, green; OC2-focused, yellow; OC3, salmon; OC3-focused, blue; OC4 purple; OC4 round 2 class 2, light blue). Shown are the unsharpened cryo-EM densities. The percentage of particles from the full set of 257,259 that was used for the respective reconstruction is indicated. **e**, Composite cryo-EM density of the OC superimposed on a ribbon model of the OC, colored as in Figure 5. The composite cryo-EM density enabled modeling of the initiation factors and DNA. Our structure also enabled correction of the revised yeast initiation complex model obtained by Murakami et al. from cryo-EM at 6 Åresolution (Murakami et al., 2015). We note the following differences between our structure and the model of Murakami et al.: (1) The TFIIF Tfg2 WH domain is rotated by $\sim$180°, which is further inconsistent with NMR data on the TFg2 WH-DNA interface (Kilpatrick et al., 2012) and fits comparatively worse to protein-protein crosslinking data between the Tfg2 WH and Tfa2 WH1 (Murakami et al., 2013). (2) Domains of TFIIE except Tfa2 WH1 were placed incorrectly: Tfa1 eWH (rotation and translation into the E-linker density; 17 Ådistance for helix α3 in our CC), Tfa1 E-ribbon (rotation and translation into E-linker density; 35 Ådistance between the Zn atoms), and Tfa2 WH2 ($\sim$180°rotation). Further, the Tfa2 E-tether region was incorrectly assigned to density belonging to the Tfa1 eWH. The Tfa1 E-linker was not modeled. (3) The TFIIF Tfg1 arm was modeled into an empty space lacking density, and the Tfg1 helix α0 was absent. Our models of the TFIIF dimerization domain, Tfg2 linker, Tfg1 N-terminus, and Tfg1 arm fit perfectly into densities from the recent study (Murakami et al., 2015), indicating the electron microscopic reconstruction is correct, but that the modeling was premature at the available resolution. **f**, Ribbon model of the OC colored according to how different parts of the OC were modeled into the OC cryo-EM densities (see panel **d**). Regions with atomic (light blue) and backbone models (orange), and DNA (dark blue) are indicated. Views as in **d**. **g**, Representative regions of the sharpened cryo-EM densities OC1 (3.6 Å), OC2 (4.0 Å), and OC4 (3.9 Å) are shown with the underlying refined coordinate model. The OC1 density shows clear side-chain features for Rpb1 clamp helices α8 and α9 and Rpb2 β33, the OC4 density for Tfg1 β2 that is part of the dimerization domain, and the OC2 density for part of the Tfg2 linker. For OC nomenclature see Supp. Figure 19. **h**, Fit of the TBP crystal structure (PDB code 1YTB) (Kim et al., 1993a) to the OC2 cryo-EM density show in a Pol II side view (Cramer et al., 2001). **i**, Fit of TFIIB N- and C-terminal cyclin domains, B-linker and B-reader, and B-ribbon elements to OC1 and OC2 cryo-EM densities. The B-linker element displays weak density, and the B-reader is not observed.

Figure 21: Modelling of open complex cryo-EM densities. j, Fit of the TFIIA crystal structure (PDB code 1YTF) (Geiger et al., 1996) to OC2-focused cryo-EM density in a Pol II top view (Cramer et al., 2001) (left). The four-helix bundle undergoes a minor rotation towards the β-barrel, and is apparently flexible (compare Supp. Figure 24e). Toa1 (middle) and Toa2 (right) subunit structures are shown. A large non-conserved insertion in Toa1 (Δ95-209), lacking in recombinant TFIIA, may affect the relative positioning of the four-helix bundle to the β-barrel. **k**, Fit of the TFIIF model to OC cryo-EM densities viewed from the top (Cramer et al., 2001). TFIIF dimerization domain and Tfg1 N-terminal region, arm, and charged helix elements are superimposed on the OC4 cryo-EM density. Tfg2 linker and winged helix (WH) domains are superimposed on OC2 and OC3-focused cryo-EM densities, respectively. Subunit architectures for Tfg1 (middle) and Tfg2 (right) subunits are shown, indicating disordered regions. Secondary structure elements were labeled according to the crystallographic model of the human RAP30-RAP74 heterodimer (Gaiser et al., 2000). **l**, Fit of the TFIIE model to OC cryo-EM densities shown from the front (Cramer et al., 2001) (left). Models for Tfa1 eWH, E-linker and E-ribbon are superimposed onto OC1 and OC3 densities. Models for Tfa2 WH1 domain, Tfa2 WH2 and E-tether were fitted into OC3-focused and OC3 densities. Tfa1 (middle) and Tfa2 (right) subunits are shown, indicating disordered regions. The connectivity of the E-tether helices remains uncertain. **m**, Fit of promoter DNA to OC cryo-EM densities is shown in a side view (Cramer et al., 2001). A weak density for single-stranded template DNA contacts the Pol II fork loop 1, and is indicated by a blue arrow. Upstream and downstream DNA models are superimposed with OC3 and OC1 densities, respectively. The location of the Pol II active site magnesium ion is indicated.

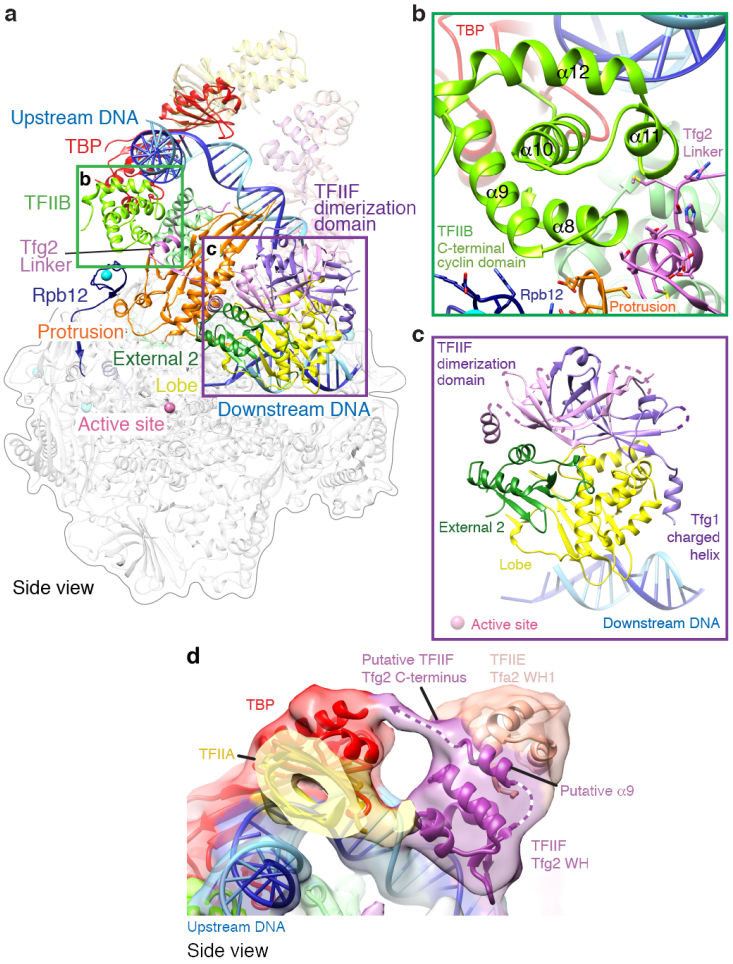

Figure 22: Details of TFIIB and TFIIF in the upstream assembly. a, View of the open complex from the side (Cramer et al., 2001). Pol II elements external 2, lobe, protrusion, Pol II subunit Rpb12 and basal fractor TBP, TFIIB, and -F are highlighted. The remainder of the open complex is transparent. Green and purple boxes indicate the locations of TFIIB C-terminal cyclin and TFIIF dimerization domains, respectively. **b**, Interactions of TFIIB C-terminal cyclin domain with Pol II protrusion, Rpb12, Tfg2 linker and DNA. **c**, Details of the TFIIF dimerization domain interactions with Pol II external 2 and lobe. **d**, The putative Tfg2 C-terminus contacts TBP. The view is from the side (Cramer et al., 2001). A tubular cryo-EM density from the OC3 map, low-pass filtered to 8 Å, emanates from the TFIIF Tfg2 WH-TFIIE Tfa2 WH1 density, and was tentatively assigned to the Tfg2 C-terminal region. The putative Tfg2 density reaches the TBP subunit, consistent with their suggested interaction (Robert et al. 1996, Ha et al., 1993).

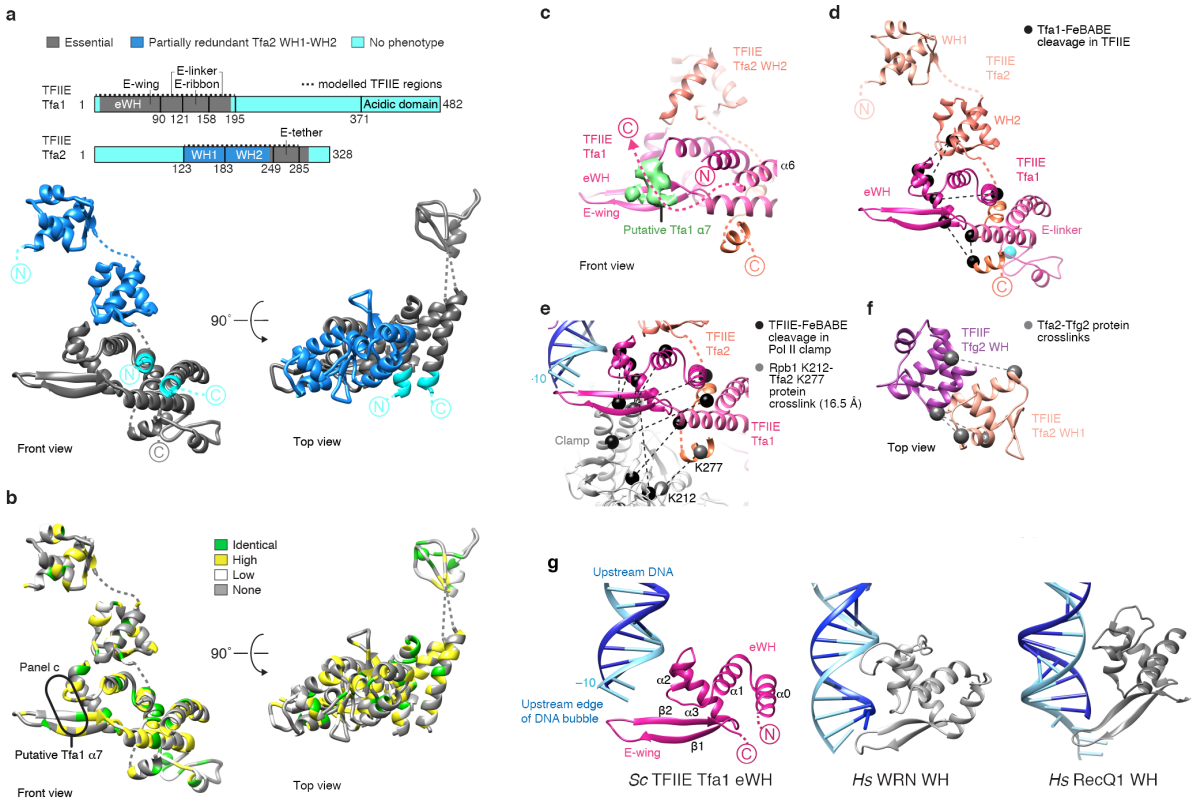

Figure 23: Modelling of TFIIE. a, The architectural model of TFIIE contains all regions required for viability in yeast (Grünberg et al., 2012). A domain schematic (top) indicates the excellent overlap between modeled (dashed line) and functionally essential regions. Essential (grey), partially redundant Tfa2 winged helix 1 (WH1) and WH2 domains (blue), and non-essential elements (cyan) are indicated on the TFIIE model, shown in previously defined front and top views (Cramer et al., 2001) of Pol II. **b**, TFIIE sequence conservation. The sequence conservation among *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Gallus gallus*, and *Homo sapiens* was mapped onto a ribbon representation of the TFIIE model. Highly, strongly, weakly and non-conserved residues are colored in green, yellow, white, and gray respectively. The location of a non-modeled helical density in the OC3 cryo-EM map, which may correspond to Tfa1 helix $\alpha 7$, is indicated. Views as in **a**. **c**, An additional density (green) in the OC3 cryo-EM on top of the Tfa1 E-wing was tentatively assigned to Tfa1 helix $\alpha 7$ and this may stabilize the long beta-hairpin. Shown is a front view (Cramer et al., 2001). **d**, Tfa1-FeBABE cleavage sites in TFIIE (Grünberg et al., 2012) are consistent with the TFIIE architecture. **e**, Tfa1- and Tfa2-FeBABE cleavage sites in the Pol II clamp (Grünberg et al., 2012) and a protein-protein crosslink between Rpb1 K212 (Pol II clamp)-Tfa2 K277 (TFIIE E-tether) (Murakami et al., 2013) are consistent with the location of eWH and E-tether. **f**, Tfa2-Tfg2 protein-protein crosslinks (Murakami et al., 2013) are consistent with the Tfg2 WH-Tfa2 WH1 architecture. **g**, The long E-wing in the TFIIE subunit Tfa1 eWH is characteristic of WH domains involved in DNA strand separation (Harami et al., 2013). The upstream edge of the transcription bubble and eWH domain are shown in a front view (Cramer et al., 2001) rotated by $\sim 20^{\circ}$ in the horizontal axis. Corresponding regions of human (*Hs*) WRN WH (PDB code 2WWY) and RecQ1 WH (PDB code 3AAF) domains are shown.

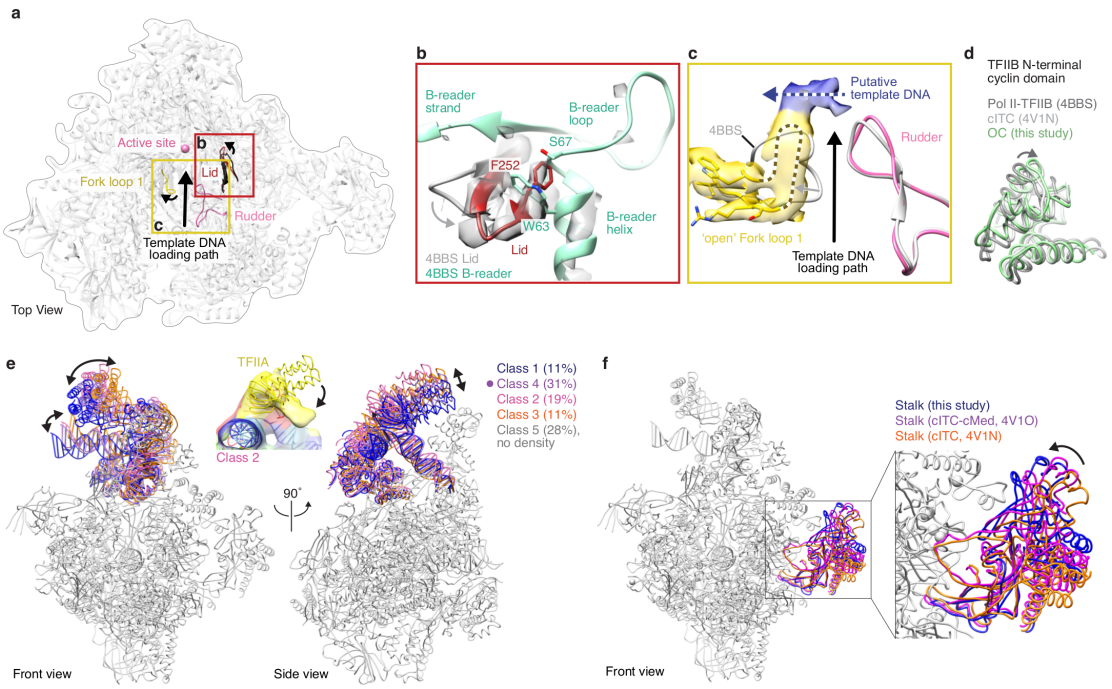

Figure 24: Pol II cleft clearance, structural flexibility and rearrangements in the OC. a, Pol II lid and fork loop 1 assume new conformations in the OC, clearing the Pol II cleft for loading of single-stranded template DNA. Arrows indicate the direction of movement of the two Pol II elements, and the template DNA loading path. The lid (dark red) in the open complex is moved in comparison to the lid of a Pol II-TFIIB ITC crystallographic study (PDB code 4BBS). Yellow and red boxes indicate zoom-ins in panels **b** and **c**, respectively. **b**, The movement in the Pol II lid leads to a steric clash with the TFIIB B-reader, observed in a Pol II-TFIIB ITC crystal (PDB code 4BBS), and facilitates its withdrawal in the open complex. In particular the lid residue F252 clashes with W63 and S67 of the B-reader. The OC1 cryo-EM density is shown for both lid and B-reader elements. **c**, The cryo-EM density of the OC1 reveals an 'open' Pol II fork loop 1 and a stably associated fragment of putative template DNA. The 'open' state of fork loop 1 provides additional space for loading of single-stranded template DNA past the Pol II rudder, towards the active site cleft. **d**, The position of the TFIIB N-terminal cyclin domain (light green) is altered in comparison to a Pol II-TFIIB ITC crystal structure (Sainsbury et al., 2013) (dark grey), but similar to its location in a cITC (Plaschka et al., 2015) (light grey), likely due to the presence of DNA. **e**, Flexibility of the upstream DNA assembly. The cryo-EM data of the OC was sorted based on structural differences using an upstream assembly mask that included upstream DNA, TFIIA, TBP, and TFIIB cyclin domains (OC2 round 1, compare Appendix Figure 19c). Four of five resultant classes revealed different positions of the upstream complex, indicated here by fitted ribbon models of the OC. Previously defined front and side views (Cramer et al., 2001) are shown. Class 2 (middle) revealed the TFIIA four-helix bundle rotated by 85°, consistent with a high degree of flexibility. Class 4 represents the largest fraction of the data (31 %), and gave a more defined density for the upstream complex, which was improved by further classification (Supp. Figure 19c).

Figure 24: Pol II cleft clearance, structural flexibility and rearrangements in the OC. e, continued: Class 5 presented with no density for the upstream complex or the Tfg2 linker, but did show density for the TFIIB B-ribbon and the TFIIF dimerization domain, suggesting that TFIIB and TFIIF remained bound to the complex. This is consistent with TFIIF-dependent association of the TFIIB-core domain with the Pol II wall (Fishburn et al., 2012), and this apparently requires an ordered Tfg2 linker. **f**, The Rpb4-Rpb7 stalk adopts different positions in cITC, cITC-cMed, and OC. This suggests that Mediator and TFIIE may bind co-operatively. This is consistent with previous findings (Esnault et al., 2008) and with pulldowns (Figure 13a), which suggest that the TFIIE E-ribbon-stalk interface, which is important for TFIIE recruitment, is stabilized in the presence of Mediator.

Part VI

Appendix

14 Supplemental data for closed complex sample preparation

Supplemental data shown in this section illustrates optimisation steps and details of the closed complex formation for cryo-EM analysis.

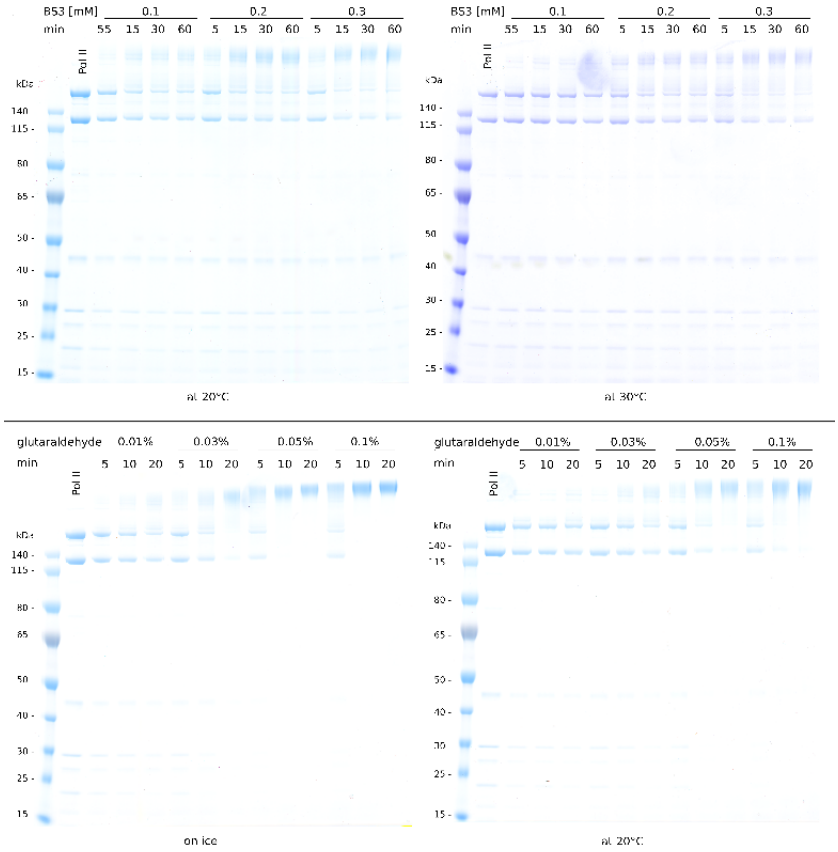

Figure 25: Cross-linking titration of Pol II using BS3 and glutaraldehyde. For optimisation of the cross-linking step Pol II was cross-linked by BS3 and glutaraldehyde at different cross-linker concentrations, temperatures and for different time spans. Results were analsed by SDS-PAGE. Glutaraldehyde cross-links Pol II more efficiently than BS3. Cross-linking on ice by glutaraldehyde has the advantage that the protein complex is not exposed to drastic temperature shifts which may negatively influence complex stability.

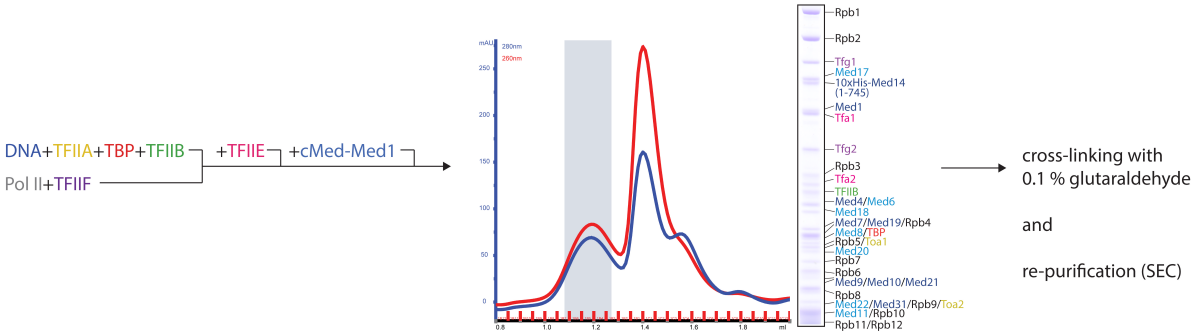

Figure 26: Sample preparation scheme of the closed complex for cryo-EM. Pol II, closed DNA, transcription factors and core Mediator+Med1 were assembled in an ordered manner and incubated for 50 min at 25 °C. A stoichiometric complex was isolated via size exclusion chromatography. The grey box in the chromatogram indicates the pooled fractions which were analysed by SDS-PAGE. After cross-linking with 0.1 % glutaraldehyde on ice, the complex was re-purified by size exclusion chromatography and used for grid preparation.

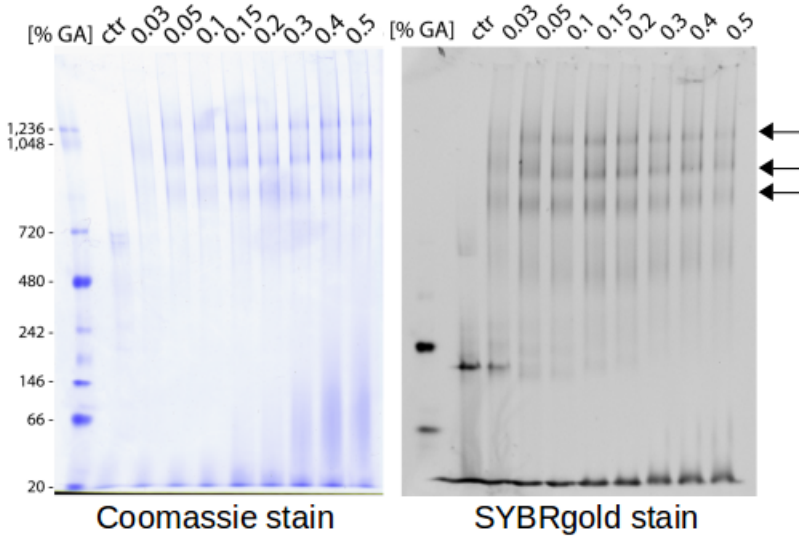

Figure 27: Native PAGE analysis of cross-linked closed complex. Shown is a cross-linking titration of the closed complex using 0.03-0.5 % glutaraldehyde (GA) and a control (ctr, no cross-linking) analysed by native PAGE with a coomassie stain for proteins (left) and a SYBRgold stain for nucleic acids (right). A molecular size marker is indicated on the left. The closed complex containing core Mediator-Med1 has a molecular size of approximately 1.4 MDa. After cross-linking 3 different complex species appear (marked by the arrows), indicating that 2 additional cross-linked subcomplexes were present in the sample.

15 Homology model of *S. cerevisiae* TFIIF

Results presented in this section have been published:

W. Mühlbacher, S. Sainsbury, M. Hemann, **M. Hantsche**, S. Neyer, F. Herzog, P. Cramer. (2014) Conserved architecture of the core RNA polymerase II initiation complex. *Nature Communications* **5**:4310. doi: 10.1038/ncomms5310
For details on author contributions see page x.

A homology model of the TFIIF dimerisation domain was generated based on a human crystal structure (PDB code 1F3U) (Gaiser et al., 2000). In *S. cerevisiae* both subunits of TFIIF contain insertions inside the dimerisation domain fold compared to the human sequences (Figures 28 and 29).

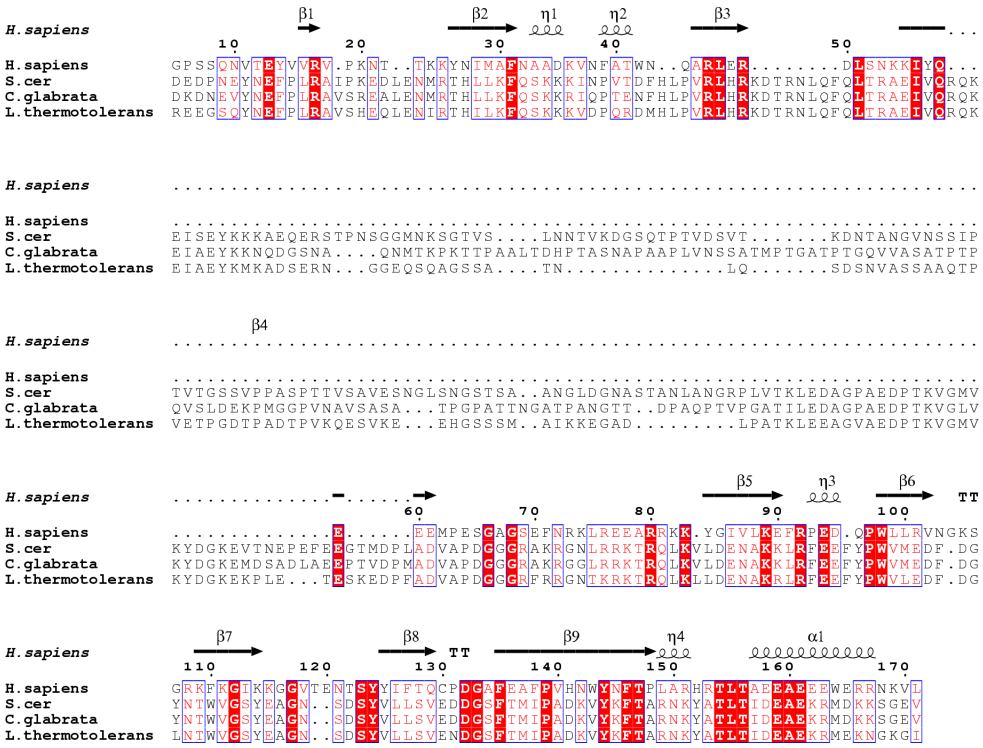

Figure 28: Sequence alignment of Tfg1 dimerisation domain. The sequence of *S. cerevisiae* Tfg1 dimerisation domain ranging from residues 92-417 was aligned to *H. sapiens* (residues 5-168), *C. glabrata* (residues 88-394) and *L. thermotolerans* (residues 72-342) by MSAProbs (Liu et al., 2010b). Secondary structure information is based on the crystal structure of the human dimerisation domain (Gaiser et al., 2000).

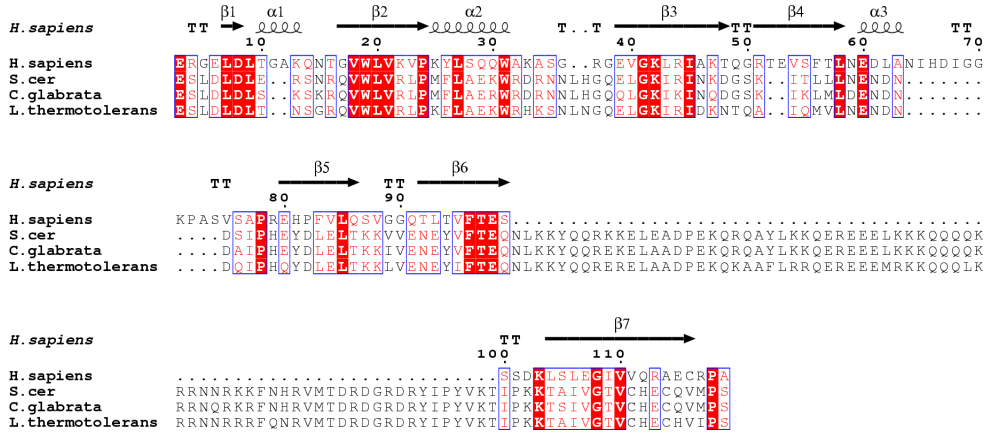

Figure 29: Sequence alignment of Tfg2 dimerisation domain. The sequence of *S. cerevisiae* Tfg2 dimerisation domain ranging from residues 54-227 was aligned to *H. sapiens* (residues 3-119), *C. glabrata* (residues 49-221) and *L. thermotolerans* (residues 41-213) by MSAProbs (Liu et al., 2010b). Secondary structure information is based on the crystal structure of the human dimerisation domain (Gaiser et al., 2000).

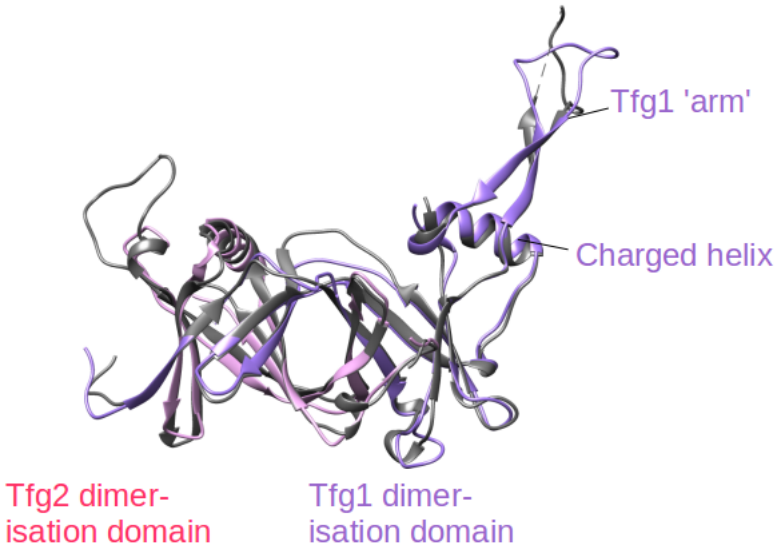

Figure 30: Homology model of *S. cerevisiae* TFIIF dimiersation domain. Shown is the model of *S. cerevisiae* Tfg1 (purple) and Tfg2 (plum) dimerisation domain generated by MODELLER (Eswar et al., 2006) superposed to the human crystal structure (dark grey) (Gaiser et al., 2000).

Initial sequence alignments generated by MSAProbs (Liu et al., 2010b) and MUSCLE (Edgar, 2004) were manually refined regarding the start and end of the yeast-specific insertions. In the final alignment Tfg1 *S. cerevisiae* residues 92-153 and 324-417 align to *H. sapiens* residues 5-62 and 73-168, and Tfg2 *S. cerevisiae* residues 54-138 and 208-227 align to *H. sapiens* residues 2-119, respectively. The alignments for Tfg1 and Tfg2 were fused by

a poly-alanine linker in order to generate a homology model of the dimerisation domain with MODELLER (Eswar et al., 2006) (Figure 30). Analysis of the cross-linking/MS data revealed later that the location of the charged helix differ in the crystal packing (Gaiser et al., 2000) and in solution (Mühlbacher et al., 2014).

16 Additional unpublished data

The TFIIF Tfg1 N-terminus was co-crystallised with Pol II (see Section 7.2.1). Further attempts to co-crystallise Pol II with TFIIF were unsuccessful. Different co-crystallisation strategies were performed and are briefly summarised in this section.

A new TFIIF construct was cloned, lacking the yeast-specific insertions in the dimerisation domain (TFIIFΔins, see Table 3) in order to reduce flexibility. Crystallisation screens of the Pol II-TFIIFΔins complex were performed in parallel to Pol II alone with no promising results.

In a next step, a Pol II-binding domain was fused to TFIIF to increase its local concentration on Pol II (TFIIFΔins-TLD, see Table 3, Figure 31b). The TFIIS-like domain (TLD) of Bye1 binds to the Pol II jaw, close to the location where the TFIIF dimerisation domain binds, without affecting Pol II function (Kinkelin et al., 2013). There was just one condition in which crystals were solely observed for the protein complex (0.1 M HEPES pH 7.5, 0.2 M lithium sulphate, 25 % PEG 3350) (Figure 31a, c). The hit was manually reproduced and refined. SDS-PAGE and MS analysis suggested that Pol II and TFIIF was present in the crystals. Diffraction data analysis revealed the same space group and cell dimensions as typical Pol II crystals (C222(1)). As crystal packing in Pol II crystals does not provide enough space for accepting the TFIIF dimerisation domain, data were not further processed as well as crystallisation conditions not further refined.

To be able to differentiate between crystals containing only Pol II or Pol II-TFIIF at an early stage, a S6-tag was cloned to to the Tfg1 N-terminus allowing for specific labelling of TFIIF by a fluorophore (S6-TFIIFΔins-TLD, see Table 3, Section 6.3.4 and Figure 31d). Crystallisation trials were set up as before and drops containing crystals were monitored with a fluorescent microscope. In all conditions a defined fluorescent signal was observed from the crystals (Figure 31e). Based on the previous results this suggested that TFIIF was located inside the crystals, possibly bound to Pol II via the Tfg1 N-terminal helix but the rest of TFIIF flexible in the solvent channels.

In a following step, TFIIF was fused to a subunit of Pol II to prevent dissociation of

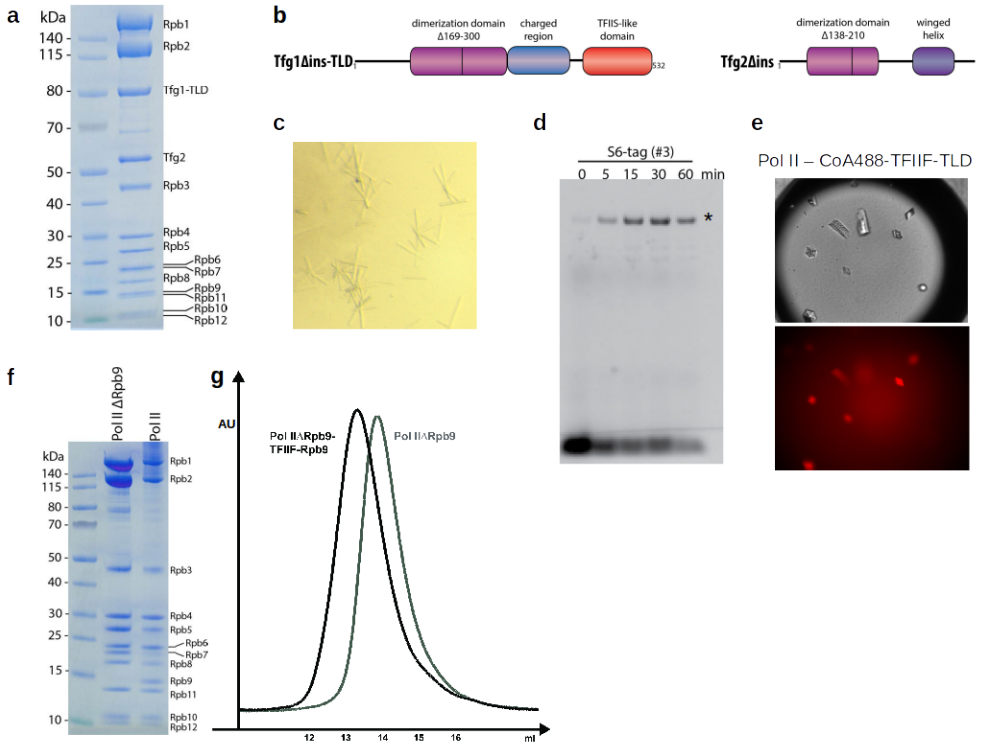

Figure 31: Additional Pol II-TFIIF co-crystallisation trials. **a**, SDS-PAGE analysis (coomassie staining) shows a pure and stoichiometric Pol II-TFIIF-TLD complex. **b**, Domain organisation of the TFIIFΔins-TLD construct. **c**, Initial crystal hit of the Pol II-TFIIFΔins-TLD complex. **d**, SDS-PAGE analysis of an enzyme-mediated labelling titration of the S6-tag with a CoA-488 substrate. Labelled S6-Tfg1-TLD is marked by an asterisk. Fluorescent signal was visualised by a LEDblue Cy2 filter. **e**, Exemplary crystal hit for the Pol II-CoA488-TFIIFΔins-TLD complex visualised by bright field (upper picture) and fluorescent filter (lower picture). **f**, SDS-PAGE analysis (coomassie staining) of Pol II and Pol IIΔRpb9. **g**, Reconstituion of Pol IIΔRpb9 with TFIIFΔins-Rpb9 by size exclusion chromatography.

TFIIF during crystallisation. Rpb9 is one of the few Pol II subunits which can be deleted in yeast. Pol IIΔRpb9 (see Table 2) was purified as described (see Section 6.3.1, Figure 31f) (Ruan et al., 2011). Rpb9 was fused to the Tfg1 C-terminus (TFIIFΔins-Rpb9, see Table 3) and Pol IIΔRpb9 was reconstituted with the TFIIF-Rpb9 construct (Figure 31g). Initial crystallisation screens in parallel with Pol II and Pol IIΔRpb9 provided no crystal hits for the fusion complex. This result may indicate that co-crystallisation of the Pol II-TFIIF complex in the Pol II crystal form is prevented.

References

Ahn, S. H., Kim, M., and Buratowski, S. (2004). Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. *Molecular Cell*, 13(1):67–76.

Allen, B. L. and Taatjes, D. J. (2015). The Mediator complex: a central integrator of transcription. *Nature reviews. Molecular cell biology*, 16(3):155–166.

Alva, V., Nam, S.-z., Söding, J., and Lupas, A. N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research*, 44:410–415.

Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M. I., Lundström, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nature Structural & Molecular Biology*, 20(8):1008–14.

Andel, F., Ladurner, A. G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-Dimensional Structure of the Human TFIID-IIA-IIB Complex. *Science*, 286(5447):2153–2156.

Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):6964–8.

Asturias, F. J., Jian, Y. W., Myers, L. C., Gustafsson, C. M., and Kornberg, R. D. (1999). Conserved Structures of Mediator and RNA Polymerase II Holoenzyme. *Science*, 283(5404):985–987.

Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R., and Darst, S. A. (2015). Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife*, 4(e08504):1–23.

Basehoar, A. D., Zanton, S. J., and Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116(5):699–709.

Bernecky, C., Grob, P., Ebmeier, C. C., Nogales, E., and Taatjes, D. J. (2011). Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLoS Biology*, 9(3):e1000603.

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. *Nature*, 529(7587):551–4.

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. *Nature*, 493(7434):699–702.

Blanc, E., Vonrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D: Biological Crystallography*, D60:2210–2221.

Bleichenbacher, M., Tan, S., and Richmond, T. J. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *Journal of Molecular Biology*, 332(4):783–793.

Blombach, F., Salvadori, E., Fouqueau, T., Yan, J., Reimann, J., Sheppard, C., Smollett, K. L., Albers, S. V., Kay, C. W. M., Thalassinos, K., and Werner, F. (2015). Archaeal TFEα/β is a hybrid of TFIIE and the RNA polymerase III subcomplex hRPC62/39. *eLife*, 4:e08378.

Bond, C. S. and Schüttelkopf, A. W. (2009). ALINE: a protein-sequence alignment editor for publication-quality alignments. *Acta Crystallographica Section D: Biological Crystallography*, 65(5):510–512.

Borggrefe, T. and Yue, X. (2011). Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Seminars in Cell and Developmental Biology*, 22(7):759–768.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72:248–254.

Brueckner, F. and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature Structural & Molecular Biology*, 15(8):811–8.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 56(4):549–561.

Buratowski, S., Sopta, M., Greenblatt, J., and Sharp, P. A. (1991). RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *Proceedings of the National Academy of Sciences of the United States of America*, 88(17):7509–7513.

Burke, T. W. and Kadonaga, J. T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAF(II)60 of Drosophila. *Genes and Development*, 11(22):3020–3031.

Bushnell, D. A., Cramer, P., and Kornberg, R. D. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):1218–1222.

Bushnell, D. A. and Kornberg, R. D. (2003). Complete, 12-subunit RNA polymerase II at 4.1 A resolution: Implications for the initiation of transcription. *Proc. Natl. Acad. Sci. USA*, 100(12):6969–6973.

Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. (2004). Structural Basis of Transcription: An RNA Polymerase II-TFIIB Cocrystal at 4.5 Angstroms. *Science*, 303(5660):983–988.

Cabart, P., Ujvari, A., Pal, M., and Luse, D. S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38):15786–15791.

Cai, G., Chaban, Y. L., Imasaki, T., Kovacs, J. A., Calero, G., Penczek, P. A., Takagi, Y., and Asturias, F. J. (2012). Interaction of the mediator head module with RNA polymerase II. *Structure*, 20(5):899–910.

Cai, G., Imasaki, T., Takagi, Y., and Asturias, F. J. (2009). Mediator Structural Conservation and Implications for the Regulation Mechanism. *Structure*, 17(4):559–567.

Cai, G., Imasaki, T., Yamada, K., Cardelli, F., Takagi, Y., and Asturias, F. J. (2010). Mediator Head module structure and functional interactions. *Nature Structural & Molecular Biology*, 17(3):273–279.

Cevher, M. A., Shi, Y., Li, D., Chait, B. T., Malik, S., and Roeder, R. G. (2014). Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature Structural & Molecular biology*, 21(12):1028–34.

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Nixon, B. T., Knight, J., Weiss, S., and Ebright, R. H. (2012). Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science*, 337(6094):591–595.

Chalkley, G. E. and Verrijzer, C. P. (1999). DNA binding site selection by RNA polymerase II TAFs: A TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO Journal*, 18(17):4835–4845.

Chen, H. T. and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Molecular Cell*, 12(2):437–447.

Chen, H. T. and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: A model for the structure of the PIC. *Cell*, 119(2):169–180.

Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature Structural & Molecular Biology*, 14(8):696–703.

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., and Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, 135:24–35.

Chen, V. B., Bryan, W., Headd, J. J., Keedy, D. A., Robert, M., Kapral, G. J., Murray, L. W., Jane, S., and David, C. (2010a). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, D66:12–21.

Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.-C., Nilges, M., Cramer, P., and Rappsilber, J. (2010b). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO Journal*, 29(4):717–726.

Cheung, A. C. M. and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471(7337):249–53.

Cheung, A. C. M. and Cramer, P. (2012). A movie of RNA polymerase II transcription. *Cell*, 149(7):1431–1437.

Cheung, A. C. M., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. *EMBO Journal*, 30(23):4755–63.

Cianfrocco, M. A., Kassavetis, G. A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J. T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell*, 152(1-2):120–131.

Colbert, T. and Hahn, S. (1992). A yeast TFIIB-related factor involved in RNA polymerase III transcription. *Genes and Development*, 6(10):1940–1949.

Conaway, R. C. and Conaway, J. W. (1993). General initiation factors for RNA polymerase II. *Annual Review of Biochemistry*, 62:161–190.

Conaway, R. C. and Conaway, J. W. (2011). Origins and activity of the Mediator complex. *Seminars in Cell and Developmental Biology*, 22(7):729–734.

Connelly, S. and Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes and Development*, 2(4):440–452.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., THompson, N. E., Burgess, R. R., Edwards, A. M., Davis, P. R., and Kornberg, R. D. (2000). Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. *Science*, 288(5466):640–649.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292(5523):1863–1876.

Davis, J. A., Takagi, Y., Kornberg, R. D., and Asturias, F. J. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular Cell*, 10(2):409–415.

Deng, W. and Roberts, S. G. E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes and Development*, 19(20):2418–2423.

Deng, W. and Roberts, S. G. E. (2007). TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma*, 116(5):417–429.

Ebmeier, C. C. and Taatjes, D. J. (2010). Activator-Mediator binding regulates Mediator-cofactor interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25):11283–11288.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797.

Eichner, J., Chen, H.-T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO Journal*, 29(4):706–716.

Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C. O., Koeck, P. J. B., Holmberg, S., Hebert, H., and Gustafsson, C. M. (2006). The cyclin-dependent kinase 8 module sterically blocks Mediator interactions with RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43):15788–93.

Elmlund, H., Baraznenok, V., Linder, T., Szilagyi, Z., Rofougaran, R., Hofer, A., Hebert, H., Lindahl, M., and Gustafsson, C. M. (2009). Cryo-EM Reveals Promoter DNA Binding and Conformational Flexibility of the General Transcription Factor TFIID. *Structure*, 17(11):1442–1452.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, D60:2126–2132.

Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. *Nature*, 502(7473):650–5.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Molecular Cell*, 31(3):337–346.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, 0 5:Unit 5.6.

Fan, L., Arvai, A. S., Cooper, P. K., Iwai, S., Hanaoka, F., and Tainer, J. A. (2006). Conserved XPB Core Structure and Motifs for DNA Unwinding: Implications for Pathway Selection of Transcription or Excision Repair. *Molecular Cell*, 22(1):27–37.

Feklistov, A. and Darst, S. A. (2011). Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, 147(6):1257–1269.

Fernández-Tornero, C., Böttcher, B., Rashid, U. J., Steuerwald, U., Flörchinger, B., Devos, D. P., Lindner, D., and Müller, C. W. (2010). Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO Journal*, 29(22):3762–3772.

Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U. J., Taylor, N. M. I., Ruiz, F. M., Gruene, T., Legrand, P., Steuerwald, U., and Müller, C. W. (2013). Crystal structure of the 14-subunit RNA polymerase I. *Nature*, 502(7473):644–649.

Fishburn, J. and Hahn, S. (2012). Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF. *Molecular and Cellular Biology*, 32(1):12–25.

Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(13):3961–6.

Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H., and Kornberg, R. D. (1991). A mediator required for activation of RNA polymerase II transcription in vitro. *Nature*, 350(6317):436–438.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., Nguyen, T., Karp, S., and Bentley, D. L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell*, 60(2):256–267.

Forget, D., Langelier, M.-f., The, C., Trinh, V., and Coulombe, B. (2004). Photo-Cross-Linking of a Purified Preinitiation Complex Reveals Central Roles for the RNA Polymerase II Mobile Clamp and TFIIE in Initiation Mechanisms. *Molecular and Cellular Biology*, 24(3):1122–1131.

Gaiser, F., Tan, S., and Richmond, T. J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. *Journal of Molecular Biology*, 302(5):1119–1127.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Balroch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *John M. Walker: The Proteomics Protocols handbook, Humana Press*, pages 571–607.

Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996). Crystal Structure of the Yeast TFIIA/TBP/DNA Complex. *Science*, 272(5263):830–6.

Geiger, S. R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A. J. R., and Cramer, P. (2010). RNA Polymerase I Contains a TFIIF-Related DNA-Binding Subcomplex. *Molecular Cell*, 39(4):583–594.

Giardina, C. and Lis, J. T. (1993). DNA melting on yeast RNA polymerase II promoters. *Science*, 261(14):759–762.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. a., and Kornberg, R. D. (2001). Polymerase II Elongation Complex at Structural Basis of Transcription : An RNA Polymerase II Elongation Complex at 3 . 3 Å Resolution. *Science*, 292(5523):1876.

Grob, P., Cruse, M. J., Inouye, C., Peris, M., Penczek, P. A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIID: Conformational breathing in the integration of gene regulatory cues. *Structure*, 14(3):511–520.

Groft, C. M., Uljon, S. N., Wang, R., and Werner, M. H. (1998). Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 95(16):9117–9122.

Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R. H., Michaelis, J., and Werner, F. (2011). The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Molecular Cell*, 43(2):263–274.

Grünberg, S. and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. *Trends in Biochemical Sciences*, 38(12):603–611.

Grünberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Structural & Molecular Biology*, 19(8):788–796.

Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., Teng, M., and Niu, L. (2008). Core Structure of the Yeast Spt4-Spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. *Structure*, 16(11):1649–1658.

Guzmán, E. and Lis, J. T. (1999). Transcription factor TFIIH is required for promoter melting in vivo. *Molecular and Cellular Biology*, 19(8):5652–8.

Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L.-U., Green, M., and Reinberg, D. (1993). Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription Factors and RNA polymerase II. *Genes and Development*, 7(6):1021–1032.

Harami, G. M., Gyimesi, M., and Kovacs, M. (2013). From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends in Biochemical Sciences*, 38(7):364–371.

Hartzog, G. A. and Fu, J. (2013). The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):105–115.

He, Y., Fang, J., Taatjes, D. J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442):481–6.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*, 533(7603):359–365.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):55–62.

Hirata, A., Klein, B. J., and Murakami, K. S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature*, 451(7180):851–4.

Hirtreiter, A., Damsma, G. E., Cheung, A. C. M., Klose, D., Grohmann, D., Vojnic, E., Martin, A. C. R., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Research*, 38(12):4040–4051.

Hoffmann, N. A., Jakobi, A. J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W. J. H., Sachse, C., and Müller, C. W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, 528(7581):231–236.

Holstege, F. C., Tantin, D., Carey, M., der van V, and Timmers, H. T. (1995). The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO Journal*, 14(4):810–819.

Holstege, F. C., van der Vliet, P. C., and Timmers, H. T. (1996). Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO Journal*, 15(7):1666–1677.

Imasaki, T., Calero, G., Cai, G., Tsai, K.-L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G. L., Asturias, F. J., Kornberg, R. D., and Takagi, Y. (2011). Architecture of the Mediator head module. *Nature*, 475(7355):240–3.

Jeronimo, C., Bataille, A. R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical Reviews*, 113(11):8491–8522.

Jiang, Y. and Gralla, J. D. (1995). Nucleotide requirements for activated RNA polymerase II open complex formation in vitro. *Journal of Biological Chemistry*, 270(3):1277–1281.

Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996). How proteins recognize the TATA box. *Journal of Molecular Biology*, 261(2):239–254.

Juven-Gershon, T., Cheng, S., and Kadonaga, J. T. (2006). Rational design of a super core promoter that enhances gene expression. *Nat Methods*, 3(11):917–922.

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D: Biological Crystallography*, D66:125–132.

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

Kamada, K., De Angelis, J., Roeder, R. G., and Burley, S. K. (2001). Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proceedings of the National Academy of Sciencesof the United States of America*, 98(6):3115–3120.

Kaplan, C. D., Larsson, K. M., and Kornberg, R. D. (2008). The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by Alpha-Amanitin. *Molecular Cell*, 30(5):547–556.

Kastner, B., Fischer, N., Golas, M. M., Boehringer, D., Hartmut, K., Deckert, J., Hauer, F., Wolf, E., Unterhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Lührmann, R., and Stark, H. (2008). GraFix: sample preparation for single- particle electron cryomicroscopy. *Nature Methods*, 5(1):53–55.

Kettenberger, H., Armache, K. J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular Cell*, 16(6):955–965.

Kilpatrick, A. M., Koharudin, L. M. I., Calero, G. A., and Gronenborn, A. M. (2012). Structural and binding studies of the C-terminal domains of yeast TFIIF subunits Tfg1 and Tfg2. *Proteins: Structure, Function and Bioinformatics*, 80(2):519–529.

Kim, H., Tang, G. Q., Patel, S. S., and Ha, T. (2012). Opening-closing dynamics of the mitochondrial transcription pre-initiation complex. *Nucleic Acids Research*, 40(1):371–380.

Kim, J. L. and Burley, S. K. (1994). 1.9 A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Structural Biology*, 1(9):638–653.

Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365(6446):520–527.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–22.

Kim, T., Ebright, R. H., and Reinberg, D. (2000). Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. *Science*, 288(5470):1418–1421.

Kim, Y., Geiger, J., Hahn, S., and Sigler, P. (1993b). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365(6446):512–520.

Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77(4):599–608.

Kinkelin, K., Wozniak, G. G., Rothbart, S. B., Lidschreiber, M., Strahl, B. D., and Cramer, P. (2013). Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38):15277–15282.

Kireeva, M. L., Komissarova, N., Waugh, D. S., and Kashlev, M. (2000). The 8-Nucleotide-long RNA:DNA Hybrid Is a Primary Stability Determinant of the RNA Polymerase II Elongation Complex. *Journal of Biological Chemistry*, 275(9):6530–6536.

Kireeva, M. L., Nedialkov, Y. A., Cremona, G. H., Purtov, Y. A., Lubkowska, L., Malagon, F., Burton, Z. F., Strathern, J. N., and Kashlev, M. (2008). Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Molecular Cell*, 30(5):557–566.

Klein, B. J. B., Bose, D., Baker, K. J. K., Yusoff, Z. M., Zhang, X., and Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America*, 108(2):546–550.

Knutson, B. A. and Hahn, S. (2011). Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. *Science*, 333(6049):1637–40.

Korinek, A., Beck, F., Baumeister, W., Nickell, S., and Plitzko, J. M. (2011). Computer controlled cryo-electron microscopy - TOM2 a software package for high-throughput applications. *Journal of Structural Biology*, 175(3):394–405.

Korkhin, Y., Unligil, U. M., Littlefield, O., Nelson, P. J., Stuart, D. I., Sigler, P. B., Bell, S. D., and Abrescia, N. G. A. (2009). Evolution of complex RNA polymerases: The complete archaeal RNA polymerase structure. *PLoS Biology*, 7(5):e1000102.

Kornberg, R. D. (2005). Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences*, 30(5):235–239.

Kornberg, R. D. (2007). The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 104(32):12955–12961.

Koschubs, T., Lorenzen, K., Baumli, S., Sandstrom, S., Heck, A. J. R., and Cramer, P. (2010). Preparation and topology of the Mediator middle module. *Nucleic Acids Research*, 38(10):3186–3195.

Kostrewa, D., Zeller, M. E., Armache, K. J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271):323–330.

Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K. R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., Menoni, H., Hoeijmakers, J., Vermeulen, W., Ge, H., Przytycka, T. M., Levens, D., and Casellas, R. (2013). Global regulation of promoter melting in naive lymphocytes. *Cell*, 153(5):988–999.

Kuehner, J. N., Pearson, E. L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews. Molecular Cell Biology*, 12(5):283–294.

Kuhlbrandt, W. (2014). Cryo-EM enters a new era. *eLife*, 3:e03678.

Kuhn, C. D., Geiger, S. R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell*, 131(7):1260–1272.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998). New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes and Development*, 12(1):34–44.

Larivière, L., Plaschka, C., Seizl, M., Wenzeck, L., Kurth, F., and Cramer, P. (2012a). Structure of the Mediator head module. *Nature*, 492(7429):448–51.

Larivière, L., Seizl, M., and Cramer, P. (2012b). A structural perspective on Mediator function. *Current Opinion in Cell Biology*, 24(3):305–313.

Lee, D.-H., Gershenzon, N., Gupta, M., Ilya, P., Reinberg, D., Lewis, B. a., and Ioshikhes, I. P. (2005). Functional Characterization of Core Promoter Elements : the Downstream Core Element Is Recognized by TAF1. *Molecular and Cellular Biology*, 25(21):9674–9686.

Lefèvre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., Pinaud, N., Teichmann, M., and Fribourg, S. (2011). Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nature Structural & Molecular Biology*, 18(3):352–358.

Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24):13668–73.

Liu, X., Bushnell, D. A., Wang, D., Calero, G., and Kornberg, R. D. (2010a). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science*, 327(5962):206–9.

Liu, Y., Schmidt, B., and Maskell, D. L. (2010b). MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics*, 26(16):1958–1964.

Logan, J., Falck-Pedersen, E., Darnell, J. E., and Shenk, T. (1987). 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 of the United States of America*, 84(23):8306–10.

Lopez-De-Leon, A., Librizzi, M., Puglia, K., and Willis, I. M. (1992). PCF4 encodes an RNA polymerase III transcription factor with homology to TFIIB. *Cell*, 71(2):211–220.

Louder, R. K., He, Y., López-Blanco, J. R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*, 531(7596):604–9.

Luo, W., Johnson, A. W., and Bentley, D. L. (2006). The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: Implications for a unified allosteric-torpedo model. *Genes and Development*, 20(8):954–965.

Malik, S. and Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews. Genetics*, 11(11):761–72.

Martinez-Rucobo, F. W. and Cramer, P. (2013). Structural basis of transcription elongation. *Biochimica et Biophysica Acta*, 1829(1):9–19.

Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO Journal*, 30(7):1302–1310.

Maxon, M. E., Goodrich, J. A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: A model for promoter clearance. *Genes and Development*, 8(5):515–524.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science*, 336(6089):1723–1725.

Meinhart, A., Blobel, J., and Cramer, P. (2003). An Extended Winged Helix Domain in General Transcription Factor E/IIEa. *Journal of Biological Chemistry*, 278(48):48267–48274.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. *Genes and Development*, 19(12):1401–1415.

Meyer, K. D., Lin, S.-C., Bernecky, C., Gao, Y., and Taatjes, D. J. (2010). p53 activates transcription by directing structural shifts in Mediator. *Nature Structural & Molecular Biology*, 17(6):753–60.

Miller, G. and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. 13(7):603–610.

Mischo, H. E. and Proudfoot, N. J. (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1):174–185.

Mühlbacher, W., Sainsbury, S., Hemann, M., Hantsche, M., Neyer, S., Herzog, F., and Cramer, P. (2014). Conserved architecture of the core RNA polymerase II initiation complex. *Nature communications*, 5:4310.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D. A., Adams, C. M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B. J., Levitt, M., and Kornberg, R. D. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science*, 342(6159):1238724.

Murakami, K., Tsai, K.-L., Kalisman, N., Bushnell, D. a., Asturias, F. J., and Kornberg, R. D. (2015). Structure of an RNA polymerase II preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2):13543–13548.

Naidu, S., Friedrich, J. K., Russell, J., and Zomerdijk, J. C. B. M. (2011). TAF1B is a TFIIB-like component of the basal transcription machinery for RNA polymerase I. *Science*, 333(6049):1640–2.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, a., Roeder, R. G., and Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10):4862–4867.

Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545):119–28.

Nogales, E. and Scheres, S. H. W. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular Cell*, 58(4):677–689.

Okamoto, T., Yamamoto, S., Watanabe, Y., Ohta, T., Hanaoka, F., Roeder, R. G., and Ohkuma, Y. (1998). Analysis of the role of TFIIE in transcriptional regulation through structure-function studies of the TFIIEbeta subunit. *Journal of Biological Chemistry*, 273(31):19866–19876.

Okuda, M., Tanaka, A., Arai, Y., Satoh, M., Okamura, H., Nagadoi, A., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2004). A novel zinc finger structure in the large subunit of human general transcription factor TFIIE. *Journal of Biological Chemistry*, 279(49):51395–51403.

Okuda, M., Tanaka, A., Satoh, M., Mizuta, S., Takazawa, M., Ohkuma, Y., and Nishimura, Y. (2008). Structural insight into the TFIIE-TFIIH interaction: TFIIE and p53 share the binding region on TFIIH. *EMBO Journal*, 27(7):1161–1171.

Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., and Nishimura, Y. (2000). Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface. *EMBO Journal*, 19(6):1346–1356.

Ostapenko, D. and Gileadi, O. (2000). Rad25p, a DNA helicase subunit of yeast transcription factor TFIIH, is required for promoter escape in vivo. *Gene*, 245(1):109–117.

Parvin, J. D. and Sharp, P. A. (1993). DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*, 73(3):533–540.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., and Burley, S. K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes and Development*, 13(24):3217–3230.

Paule, M. R. and White, R. J. (2000). Transcription by RNA polymerases I and III. *Nucleic Acids Research*, 28(6):1283–1298.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, 25:1605–1612.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016a). Transcription initiation complex structures elucidate DNA opening. *Nature*, 533(7603):353–8.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C. H., Baumeister, W., Herzog, F., Villa, E., and Cramer, P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, 518(7539):376–380.

Plaschka, C., Nozawa, K., and Cramer, P. (2016b). Mediator architecture and RNA polymerase II interaction. *Journal of Molecular Biology*, 428(12):2569–2574.

Proudfoot, N. J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences*, 14(3):105–110.

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291):aad99261–9.

Rhee, H. S. and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 487(7405):128–128.

Robert, F., Forget, D., Li, J., Greendblatt, J., and Coulombe, B. (1996). Localization of subunits of transcription factor IIE and IIF immediately upstream of the transcriptional initiation site of the adenovirus major late promoter. *Journal of Biological Chemistry*, 271(15):8517–8520.

Robinson, P. J. J., Bushnell, D. a., Trnka, M. J., Burlingame, a. L., and Kornberg, R. D. (2012). Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17931–17935.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, 21(9):327–335.

Rohou, A. and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192:216–221.

Ruan, W., Lehmann, E., Thomm, M., Kostrewa, D., and Cramer, P. (2011). Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *Journal of Biological Chemistry*, 286(21):18701–18707.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews. Molecular Cell Biology*, 16(3):129–143.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432):437–40.

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180:519–530.

Scheres, S. H. W. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of Structural Biology*, 189:114–122.

Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nature Structural & Molecular Biology*, 21(2):175–9.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell*, 155(5):1075–1087.

Schulz, S., Gietl, A., Smollett, K., Tinnefeld, P., Werner, F., and Grohmann, D. (2016). TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 113(13):E1816–25.

Schwalb, B., Michel, M., Zacher, B., Demel, C., Tresch, A., and Gagneur, J. (2016). TT-seq maps the human transient transcriptome. *Science*, 352(6290):1225–1228.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., Thompson, J. D., Higgins, D. G., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(539).

Smale, S. T. and Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual Review of Biochemistry*, 72:449–479.

Steitz, T. A., Smerdon, S. J., Jäger, J., and Joyce, C. M. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science*, 266(5193):2022–2025.

Svetlov, V. and Nudler, E. (2013). Basic mechanism of transcription by RNA polymerase II. *Biochimica et biophysica acta*, 1829(1):20–8.

Sydow, J. F., Brueckner, F., Cheung, A. C. M., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Molecular Cell*, 34(6):710–721.

Taatjes, D. J., Na, A. M., Nogales, E., and Tjian, R. (2002). Structure , Function , and Conformations of the CRSP Coactivator. *Science*, 295(5557):1058–1062.

Takagi, Y., Calero, G., Komori, H., Brown, J. A., Ehrensberger, A. H., Hudmon, A., Asturias, F., and Kornberg, R. D. (2006). Head Module Control of Mediator Interactions. *Molecular Cell*, 23(3):355–364.

Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J. (1996). Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*, 381(6578):127–151.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157:38–46.

Theisen, J. W. M., Lim, C. Y., and Kadonaga, J. T. (2010). Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Molecular and cellular biology*, 30(14):3471–3479.

Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II. *Genes and Development*, 16(6):673–675.

Tsai, F. T. and Sigler, P. B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. *EMBO Journal*, 19(1):25–36.

Tsai, K.-L., Sato, S., Tomomori-Sato, C., Conaway, R. C., Conaway, J. W., and Asturias, F. J. (2013). A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature Structural & Molecular Biology*, 20(5):611–9.

Tsai, K. L., Tomomori-Sato, C., Sato, S., Conaway, R. C., Conaway, J. W., and Asturias, F. J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell*, 157(6):1430–1444.

Vannini, A. (2013). A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. *Biochimica et Biophysica Acta*, 1829(3-4):258–264.

Vannini, A. and Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Molecular Cell*, 45(4):439–446.

Vannini, A., Ringel, R., Kusser, A. G., Berninghausen, O., Kassavetis, G. A., and Cramer, P. (2010). Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell*, 143(1):59–70.

Vassylyev, D. G., Sekine, S.-i., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, 417(6890):712–719.

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448(7150):157–162.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150):163–8.

Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. *Cell*, 127(5):941–954.

Wang, W., Carey, M., and Gralla, J. A. Y. D. (1991). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. *Science*, 255(5043):450–453.

Wang, X., Sun, Q., Ding, Z., Ji, J., Wang, J., Kong, X., Yang, J., and Cai, G. (2014). Redefining the modular organization of the core Mediator complex. *Cell research*, 24(7):796–808.

Wenzel, S., Martins, B. M., Rösch, P., and Wöhrl, B. M. (2010). Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochemical Journal*, 425(2):373–380.

Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of Molecular Biology*, 417(1-2):13–27.

West, S., Gromak, N., and Proudfoot, N. J. (2004). Human 5′ to 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525.

Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004). Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 119(4):481–489.

Wriggers, W. (2010). Using Situs for the integration of multi-resolution structures. *Biophysical Reviews*, 2:21–27.

Wu, C. C., Herzog, F., Jennebach, S., Lin, Y. C., Pai, C. Y., Aebersold, R., Cramer, P., and Chen, H. T. (2012). RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47):19232–19237.

Wu, C. C., Lin, Y. C., and Chen, H. T. (2011). The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIIC, Bdp1, and the RNA polymerase III active center. *Molecular and Cellular Biology*, 31(13):2715–2728.

Yang, C., Khaperskyy, D. A., Hou, M., and Ponticelli, A. S. (2010). Improved methods for experssion and purification of S.cerevisiae TFIIF and TFIIH; Identification of a functional E.coli promoter and internal transclation initiation within the N-terminal coding region of the TFIIF TFG1 subunit. *Protein Expression and Purification*, 70(2):172–178.

Yang, Y., Darbari, V. C., Zhang, N., Lu, D., Wang, Y.-p., Winkelman, J. T., Gourse, R. L., Murakami, K. S., Buck, M., and Zhang, X. (2015). Structures of the RNA polymerase- sigma54 reveal new and conserved regulatory strategies. *Science*, 349(6250):882–885.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell*, 98(6):811–824.

Zhang, H., Rigo, F., and Martinson, H. G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Molecular Cell*, 59(3):437–448.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., and Ebright, R. H. (2012). Structural basis of transcription initiation. *Science*, 338(6110):1076–80.

Zhou, Z., Cironi, P., Lin, A. J., Xu, Y., Golan, D. E., Silver, P. A., Walsh, C. T., and Yin, J. (2007). Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. *ACS Chemical Biology*, 2(5):337–346.

Zuo, Y. and Steitz, T. (2015). Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Molecular Cell*, 58(3):534–540.

List of abbreviations

5-FOA	5-fluoroorotic acid
Δ	genomic knockout
°C	degree celsius
Å	Ångstrom
aa	amino acids
AMPCPP	α,β-Methyleneadenosine 5´-triphosphate
ATP	adenosine-5′-triphosphate
bp	base pair(s)
fi-ME	fi-mercaptoethanol
BSA	bovine serum albumine
CC	closed complex
clonNAT	nourseothricin resistance cassette
CTD	C-terminal domain
CV	column volumes
Da	Dalton
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	2′-deoxyribonucleoside triphosphate
DTT	dithiothreitol
EC	elongation complex
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetate
et al.	and others [Latin: *er alii*]
g	gram or gravitational acceleration on earth's surface
h	hour(s)
HCl	hydrochloric acid
HEPES	2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid
His	Histidine
ITC	initial transcribing complex
IPTG	isopropyl ß-D-1-thiogalactopyranoside
l	liter
LB	lysogeny broth
Leu	leucine
M	molar (mol/liter)
min	minute(s)

MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
MW	molecular weight
NCBI	National Center for Biotechnology Information
Ni-NTA	nickel-nitrilotriacetic acid
nt	nucleotide
NTP	nucleoside triphosphate
OC	open complex
OD_{600}	optical density at 600 nm
ORF	open reading frame
pA, polyA	polyadenylation
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	poly(ethylene glycol)
PI	protease inhibitor
PIC	pre-initiation complex
PMSF	Phenylmethylsulfonylfluorid
Pol	DNA-dependent RNA polymerase
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
s	second(s)
SC	synthetic complete
Sc, S. cerevisiae	*Saccharomyces cerevisiae*
SeMet	Selenomethionine
SDS	sodium dodecyl sulphate
SLS	Swiss Light Source
T	temperature
TAE	Tris-acetate-EDTA
TBE	tris borate EDTA buffer
TCEP	tris(2-carboxyethyl) phosphine hydrochloride
TE	tris-EDTA
TEMED	N,N,N′,N′-tetramethylethylenediamine
Tris	tris-(hydroxymethyl)-aminomethane

tRNA	transfer RNA
TSS	transcription start site
U	unit
Ura	uracil
V	volt
v/v	volume per volume
wt	wildtype
w/v	weight per volume
YPD	yeast extract peptone dextrose
YPDS	yeast extract peptone dextrose sorbitol

Nucleic acids basis:

A	adenine
C	cytosine
G	guanine
T	thymine

List of Figures

List of Tables